SAS® Technical Report R-102

Computing Expected Means Squares

5981

SAS® SAS Institute Inc.

Computing Expected Mean Squares

by

J. H. Goodnight F. M. Speed

SAS® Technical Report R-102

ABSTRACT

This evaluation of the expected mean squares arising from the analysis of unbalanced mixed ANOVA models has long been an analytically intractable problem. This paper presents a theoretically and computationally simple technique.

SAS Institute Inc.
SAS Circle, Box 8000
Cary, NC 27512-8000

The correct bibliographic citation for this technical report is as follows: SAS Institute Inc., SAS® Technical Report R-102, Computing Expected Mean Squares, Cary, NC: SAS Institute Inc., 1978.

SAS® Technical Report R-102, Computing Expected Mean Squares

1. INTRODUCTION

Hartley's [1967] method of "synthesis" was one of the first general methods developed for computing the coefficients in the expected values of mean squares for the mixed model. Gaylor et al. [1970] described the computation of expected values arising from the set of mean squares of the Forward Doolittle.

Speed and Hocking [1974] developed the most computationally simple technique to date. The method presented here may be viewed as minor modification of the Speed and Hocking technique.

2. THE MODEL

Following Searle [1971] the mixed model is represented as:

$$Y = X_o\beta_o + \sum_{i=1}^{k} X_i\beta_i + e \qquad (1)$$

where,

(1) Y is an n-vector of observations

(2) $X_o, X_1, \ldots, X_k$ are nxm_i known matrices

(3) β_o is a vector of fixed effects

(4) The vectors $\beta_i (i=1,\ldots,k)$ are assumed independent of each other and e and are distributed $N(0, \sigma_i^2 I_{m_i})$

(5) e is an n-vector assumed $N(0, \sigma_e^2 I)$

On making the above normality assumptions,

$$E(Y) = X_o \beta_o$$

$$Var(Y) = \sum_{i=1}^{k} X_i X_i' \sigma_i^2 + I \sigma_e^2$$

and any symmetric quadratic form Y'QY has

$$E(Y'QY) = \beta_o' X_o' Q X_o \beta_o + \sum_{i=1}^{k} tr(X_i' Q X_i)\sigma_i^2 + tr(Q)\sigma_e^2 \quad (2)$$

Letting $X = [X_o|X_1|\ldots|X_k]$, note that all of the matrices except Q, needed in (2) are submatrices of X'QX.

3. COMPUTING SUMS OF SQUARES

A common practice for computing the sums of squares in mixed models is to formulate testable hypotheses as if all effects were fixed and then compute the corresponding SS.

Letting $\beta = \begin{bmatrix} \beta_o \\ \hline \beta_1 \\ \hline \vdots \\ \hline \beta_k \end{bmatrix}$ and treating β as if it were fixed

and letting $(X'X)^-$ be any generalized inverse of X'X, then

$$SS_L = SS(H_o: L\beta=0) = (Lb)'(L(X'X)^-L')^{-1}(Lb) \quad (3)$$

where $b = (X'X)^-X'Y$ and L is any matrix of full row rank in the row space of X.

Equation (3) is a quadratic form in Y, namely $Y'Q_LY$, where

$$Q_L = X(X'X)^- L' (L(X'X)^- L')^{-1} L(X'X)^- X'$$

Only the elements of $X'Q_LX$ are needed to evaluate the expected value of (3) when (1) is assumed to be the true model. Since L is in the row space of X, $L = L(X'X)^- X'X$.

Therefore

$$X'Q_L X = L'(L(X'X)^- L')^{-1} L \tag{4}$$

Although (4) reduces the computational burden of computing expected values of SS's, a further reduction is developed in the next section.

4. EXPECTED VALUES OF SS

Theorem: If L is of full row rank and in the row space of X, and SS_L is computed using (3) then there exists a matrix $C = [C_o|C_1|\ldots|C_k]$ of the same dimensions of L such that:

(i) C = ML and

(ii) $E(SS_L) = \beta_o' C_o' C_o \beta_o + \sum_{i=1}^{k} SSQ(C_i)\, \sigma_i^2 + n_L \sigma_e^2$

where $SSQ(C_i)$ = sum of squares of the elements of the C_i submatrices, and

n_L = the number of rows in L

Proof: If the matrix

$$[L(X'X)^- L' \mid L] \tag{5}$$

is formed and Cholesky adjustments are performed on the diagonals of the left hand matrix then the following matrix results:

$$[U \mid C] \tag{6}$$

Here U is the Cholesky decomposition (upper) of $L(X'X)^- L'$ and $C = (U')^{-1}L$. Thus $X'Q_L X = L'(L(X'X)^- L')^{-1}L = L'(U'U)^{-1}L = C'C$

and the expected value of SS_L may be computed from the elements of

$$C'C = \begin{bmatrix} C_o' C_o & C_o' C_1 & \cdots & C_o' C_k \\ C_1' C_o & C_1' C_1 & \cdots & C_1' C_k \\ \vdots & & & \\ C_k' C_o & C_k' C_1 & \cdots & C_k' C_k \end{bmatrix}$$

Therefore

$$E(SS_L) = \beta_o C_o' C_o \beta_o + \sum_{i=1}^{k} tr(C_i' C_i)\ \sigma_i^2 + n_L\ \sigma_e^2$$

but $tr(C_i'C_i) = SSQ(C_i)$, thus C'C need not be computed and point (ii) of the theorem is proven. Since

$$C = (U')^{-1} L$$
$$= ML$$

point (i) is also proven.

Several minor lemmas are a consequence of the above theorem.

<u>Lemma 1</u>: If any submatrix (L_i) of

$$L = [L_o | L_1 | \ldots | L_k]$$

is zero, then the expected value of SS_L does not involve the i^{th} effect.

<u>Proof</u>: $C = [ML_o | ML_1 | \ldots | ML_k]$

<u>Lemma 2</u>: If L consists of any number of the non-zero rows of the Cholesky matrix of X'X, then C = L.

<u>Proof</u>: This is a direct consequence of the method of "synthesis."

5. CONCLUSION

When working with an un-reparameterized model, computing the expected values of mean squares is no more difficult than computing the mean squares themselves.

REFERENCES

Gaylor, D. W., Lucas, H. L., and Anderson, R. L. [1970]. Calculation of expected mean squares by the abbreviated Doolittle methods. Biometrics 26, 641-55.

Hartley, H. O. [1967]. Expectation, variances and covariances of ANOVA mean squares by "synthesis." Biometrics 23, 105-14. Correction 23, 853.

Searle, S. R. [1971]. Topics in variance component estimation. Biometrics 27, 1-72.

Speed, F. M. and Hocking, R. R. [1974]. Computation of expectations, variances and covariances of ANOVA mean squares. Biometrics 30, 157-169.

Index

SAS Institute Inc.
SAS Campus Drive
Cary, NC 27513

ISBN 1-55544-952-2

9 781555 449520