NOTES ON
ESTIMATING VARIANCE COMPONENTS FROM UNBALANCED DATA IN
MIXED MODELS OF THE ANALYSIS OF VARIANCE

by

S. R. Searle

Biometrics Unit, Cornell University, Ithaca, New York

Abstract

An outline is given of 6 available methods for estimating variance components from unbalanced data in mixed models of the analysis of variance.

April, 1974

NOTES ON
ESTIMATING VARIANCE COMPONENTS FROM UNBALANCED DATA IN
MIXED MODELS OF THE ANALYSIS OF VARIANCE

by

S. R. Searle

Biometrics Unit, Cornell University, Ithaca, New York

April, 1974

Introduction

Confine attention to the 2-way crossed classification model:

\[ y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} \]

\[ i = 1 \cdots a \quad j = 1 \cdots b \quad k = 1 \cdots n_{ij} \]

\[ \sum_{ij} n_{ij} = N \]

Fixed effects model

Balanced data

All \( n_{ij} = n \): the analysis of variance is familiar.

<table>
<thead>
<tr>
<th></th>
<th>( \sum )</th>
<th>( abn )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>1</td>
<td>( \sum \gamma^2 \cdots )</td>
</tr>
<tr>
<td>Rows</td>
<td>( a-1 )</td>
<td>( \text{SSA} = \sum \gamma^2_{ij} \cdots - \sum \gamma \cdots )</td>
</tr>
<tr>
<td>Columns</td>
<td>( b-1 )</td>
<td>( \text{SSB} = \sum \gamma^2_j \cdots - \sum \gamma \cdots )</td>
</tr>
<tr>
<td>Interaction ((a-1)(b-1))</td>
<td>( \text{SSAB} = \sum \gamma^2_{ij} - \sum \gamma^2_{i \cdots} - \sum \gamma^2_{j \cdots} + \sum \gamma \cdots )</td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>( ab(n-1) )</td>
<td>( \text{SSE} = \sum \sum \gamma^2_{ijk} - \sum \gamma^2_{ij} \cdots )</td>
</tr>
<tr>
<td>Total</td>
<td>( abn )</td>
<td>( \sum \sum \gamma^2_{ijk} )</td>
</tr>
</tbody>
</table>

Unbalanced data: \( s \) cells containing data.

2 partitionings of sums of squares.

<table>
<thead>
<tr>
<th>Rows before columns</th>
<th>OR</th>
<th>Columns before rows</th>
</tr>
</thead>
<tbody>
<tr>
<td>( R(\mu) )</td>
<td>1</td>
<td>( R(\mu) )</td>
</tr>
<tr>
<td>( R(\alpha</td>
<td>\mu) )</td>
<td>( a-1 )</td>
</tr>
<tr>
<td>( R(\beta</td>
<td>\mu,\alpha) )</td>
<td>( b-1 )</td>
</tr>
<tr>
<td>( R(\gamma</td>
<td>\mu,\alpha,\beta) )</td>
<td>( s-a-b+1 )</td>
</tr>
<tr>
<td>( \text{SSE} )</td>
<td>( N-s )</td>
<td>( \text{SSE} )</td>
</tr>
<tr>
<td>Total</td>
<td>( N )</td>
<td>Total</td>
</tr>
</tbody>
</table>
Mixed model:

\[ \beta_j's \text{ remain as fixed effects} \quad \alpha_i's \text{ random} \quad \gamma_{ij}'s \text{ random} \]

\[ E(\alpha_i) = 0 \quad E(\gamma_{ij}) = 0 \]

\[ \text{var}(\alpha) = \sigma^2_{\alpha} \quad \text{var}(\gamma) = \sigma^2_{\gamma} \]

Want to estimate: \( \mu, \beta's, \sigma^2_{\alpha}, \sigma^2_{\gamma} \text{ and } \sigma^2_e \)

Balanced data:

Use part of analysis of variance table for fixed effects model

\[ E(\text{SSA}) = (a-1)(bna^2_\alpha + n\sigma^2_{\gamma} + \sigma^2_e) \]

\[ E(\text{SSAB}) = (a-1)(b-1)(n\sigma^2_{\gamma} + \sigma^2_e) \]

\[ E(\text{SSE}) = ab(n-1)\sigma^2_e \]

Estimators

\[ \text{SSA} = (a-1)(bna^2_\alpha + n\sigma^2_{\gamma} + \sigma^2_e) \]

\[ \text{SSAB} = (a-1)(b-1)(n\sigma^2_{\gamma} + \sigma^2_e) \]

\[ \text{SSE} = ab(n-1)\sigma^2_e \]

Properties of estimators: unbiased

minimum variance quadratic unbiased

under normality, minimum variance unbiased

Unbalanced data:

Variety of methods available, several based on same principle as preceding:

Develop \( q \) as a vector of quadratic forms in \( y \)

Derive \( E(q) \); each element will be a linear combination of variance components, elements of \( \sigma^2 \).

\[ E(q) = \zeta \sigma^2 \text{ for some } \zeta \]

Estimation: \( \hat{\sigma}^2 = \zeta^{-1} q \)

Question: What quadratics are used as elements for \( q \)?
O. **Analysis of variance method (Henderson's [1953] Method 1)**

This method uses quadratic forms analogous to sums of squares of balanced data ANOVA, e.g.,

\[
SSA^* = \sum_{i} \bar{y}_{i}^2 - N\bar{y}^2
\]

\[
SSAB^* = \sum_{ij} \bar{y}_{ij}^2 - \sum_{i} \bar{y}_{i}^2 - \sum_{j} \bar{y}_{j}^2 + N\bar{y}^2
\]

Note: **SSAB^* is not positive definite; it is not a sum of squares.**

Estimation: equate SS^*’s to expectations

Properties: easy to compute

- unbiased for random models
- sampling variances available for 1, 2 and 3-way classifications
- not unbiased for mixed models, because the fixed effects, $\beta_j$’s, occur in E(SS^*)’s.
1. **Henderson's [1953] Method 2**

Designed to overcome biasedness of Method 1 for mixed models.

Retains relative ease of computing.

**Principle:** "Correct" data for fixed effects

Use Method 1 on corrected data

Make slight adjustments.

\[ y = Xb + Zu + e \]

\( \overset{\text{fixed}}{\overset{\text{random}}{1}} \)

Use normal equations as if \( u \) were fixed:

\[
\begin{bmatrix}
X'X & X'Z \\
Z'X & Z'Z
\end{bmatrix}
\begin{bmatrix}
b' \\
u'
\end{bmatrix}
=
\begin{bmatrix}
x'y \\
z'y
\end{bmatrix}
\]

Correct for \( b' \):

\[ z = y - Xb' = \mu + Zu + Ke, \text{ for some } K. \]

Use Method 1 on \( z \) just as if it were \( y \) without fixed effects.

**Adjustments:** to coefficients of \( \sigma_e^2 \) in \( E(SS's) \), to account for \( K \).

**Condition:** no interactions, fixed-by-random

**History:**

Henderson [1953]: first described, and not clear.

Searle [1968]: generalized, clarified, decried as not invariant.

Henderson, Searle and Schaeffer [1974]: invariance established, and computing procedure described.

Use R(·)'s of fitting constants for fixed effects models

\[
ER(\alpha, \gamma | \mu, \beta) = c_1 \sigma^2 + c_1 \gamma + (s-b) \sigma^2_e \\
ER(\gamma | \mu, \alpha, \beta) = c_2 \gamma + (s-a-b+1) \sigma^2_e \\
E \text{SSE} = (N-s) \sigma^2_e
\]

or, if no interaction

\[
ER(\alpha | \mu, \beta) = c_1 \sigma^2 + (a-1) \sigma^2_e \\
E \text{SSE} = (N-a-b+1) \sigma^2_e
\]

Properties: unbiased
reduce to ANOVA for balanced

Difficulties: can be difficult to compute (i.e., inverting large matrices)
can have more equations than variance components
e.g., for 2-way random model, can use

\[
R(\alpha | \mu) \quad R(\beta | \mu) \quad R(\beta | \mu, \alpha) \\
R(\beta | \mu, \alpha) \quad \text{OR} \quad R(\alpha | \mu, \beta) \quad \text{OR} \quad R(\alpha | \mu, \beta) \\
R(\gamma | \mu, \alpha, \beta) \quad R(\gamma | \mu, \alpha, \beta) \quad R(\gamma | \mu, \alpha, \beta) \\
\text{SSE} \quad \text{SSE} \quad \text{SSE} \\
y'y - \bar{y}^2 \quad y'y - \bar{y}^2
\]
As a preliminary to other methods consider the general model:

\[ y = Xb + Zu + e \]

- fixed
- random

\[ E(u) = 0 \quad E(e) = 0 \]
\[ \text{var}(u) = D \quad \text{var}(e) = R \]

\[ E(y) = Xb \]
\[ \text{var}(y) = ZDZ' + R = V \]

**GLS for** \( \widetilde{b} \):

\[
X'V^{-1}Xb^0 \overset{\sim}{=} X'V^{-1}Y
\]

Difficulty: \( V^{-1} \) of order \( N \).

**GLS for** \( b \) and \( u \), assuming \( u \) fixed

\[
\begin{bmatrix}
X' \sim R^{-1}X & X' \sim R^{-1}Z \\
Z' \sim R^{-1}X & Z' \sim R^{-1}Z
\end{bmatrix}
\begin{bmatrix}
\sim b \\
\sim u
\end{bmatrix}
= \begin{bmatrix}
X' \sim R^{-1}Y \\
Z' \sim R^{-1}Y
\end{bmatrix}
\]

Amend equations by adding \( D^{-1} \) to \( Z' \sim R^{-1}Z \):

\[
\begin{bmatrix}
X' \sim R^{-1}X & X' \sim R^{-1}Z \\
X' \sim R^{-1}Z & Z' \sim R^{-1}Z + D^{-1}
\end{bmatrix}
\begin{bmatrix}
\sim b^* \\
\sim u^*
\end{bmatrix}
= \begin{bmatrix}
X' \sim R^{-1}Y \\
Z' \sim R^{-1}Y
\end{bmatrix}
\]

These are sometimes called the "mixed model equations".

Can show that \( b^* \) is the GLS of \( b \): i.e., \( b^* \overset{\sim}{=} b^0 \)

Special case: \( u = \alpha \), a single random factor, \( \sigma^2_\alpha \) and \( \text{var}(e) = \sigma^2_e \)

Define \( \lambda = \sigma^2_e / \sigma^2_\alpha \) and \( P = Z'Z + \lambda I \)

\[
\begin{bmatrix}
X'X & X'Z \\
Z'X & P
\end{bmatrix}
\begin{bmatrix}
\sim b^* \\
\sim u^*
\end{bmatrix}
= \begin{bmatrix}
X'Y \\
Z'Y
\end{bmatrix}
\]
3. Thompson's iterative method

Models with only 1 random factor; e.g., 2-way classification without interaction

\[ y_{ijk} = \mu + \beta_j + \alpha_i + e_{ijk} \]

\[ \bar{y} = X\beta + Z\alpha + e \]

Fitting constants method: Based on

\[
\begin{bmatrix}
X'X & X'Z \\
Z'X & Z'Z
\end{bmatrix}
\begin{bmatrix}
b_0^* \\
a^*
\end{bmatrix} =
\begin{bmatrix}
X'Y \\
Z'Y
\end{bmatrix}
\text{and } R(\mu, \alpha, \beta) = (b_0'^*, a^*) \begin{bmatrix} X'Y \\ Z'Y \end{bmatrix}
\]

\[ \hat{\sigma}^2_e = \frac{y'y - R(\mu, \alpha, \beta)}{N-a-b+1} \]

\[ \hat{\sigma}^2_\alpha = \frac{R(\alpha|\mu, \beta) - (a-1)\sigma^2_e}{N - \Sigma \sum_{ij}/n_i}. \]

Cunningham and Henderson [1968]: Used mixed model equations

\[
\begin{bmatrix}
X'X & X'Z \\
Z'X & P
\end{bmatrix}
\begin{bmatrix}
b^* \\
a^*
\end{bmatrix} =
\begin{bmatrix}
X'Y \\
Z'Y
\end{bmatrix}
\text{and } R^*(\mu, \alpha, \beta) = (b^*, a^*) \begin{bmatrix} X'Y \\ Z'Y \end{bmatrix}
\]

and got

\[ \hat{\sigma}^2_e = \frac{y'y - R^*(\mu, \alpha, \beta)}{N-a-b+1} \]

\[ \hat{\sigma}^2_\alpha = \frac{R^*(\mu, \alpha, \beta) - R(\mu, \beta) - (a-1)\hat{\sigma}^2_e}{N + a\lambda - \Sigma \sum_{ij}/n_i}. \]

Iterate on \( \lambda = \sigma^2_e/\sigma^2_\alpha \) with \( P = Z'Z + \lambda I \).

Thompson's [1969] method:

Located error in expectations of Cunningham and Henderson; correction yields

\[ \tilde{\sigma}^2_e = \frac{y'y - R^*(\mu, \alpha, \beta)}{N-b} \]

\[ \tilde{\sigma}^2_\alpha = \frac{R^*(\mu, \alpha, \beta) - R(\mu, \beta)}{N - \Sigma \sum_{ij}/n_i}. \]

Iterate on \( \lambda = \sigma^2_e/\sigma^2_\alpha \).

Computing formulae for 2-way, no interaction: Searle [1973].

Extension to 2-way, with interaction: Corbeil and Searle [1973].

(This is an extension from 1 to 2 random factors.)
4. MINQUE (Four papers by C. R. Rao)

\[ y = Xb + \sum_{\theta=A}^{K+1} Z_{\theta} u_{\theta} \]

\[ \theta = \text{factors A, B, \ldots, K, including interaction} \]

\[ u_{K+1} = \varepsilon, \quad Z_{K+1} = \frac{1}{N} \]

\[ E(u_{\theta}) = 0; \quad \text{var}(u_{\theta}) = c^2 I_{\theta N_{\theta}}; \quad \text{and cov}(u_{\theta}, u_{\phi}) = 0 \text{ for } \theta \neq \phi. \]

Notation: \[ \eta_{\theta} = Z_{\theta} Z_{\theta}'; \quad W = \sum_{\theta=A}^{K+1} \eta_{\theta}; \quad (\text{Rao's } \eta) \quad V = \sum_{\theta=A}^{K+1} \sigma_{\theta}^2 \eta_{\theta} \quad (\text{Rao's } \eta^*) \]

Estimation: by quadratics \[ y' Ay \text{ with } AX = 0 \text{ choosing } A: \]

MINQUE: to minimize \[ 2\text{tr}(VA)^2 + \text{term in } \eta \text{ and kurtosis parameters} \]

[1971a, p. 268; 1972, p. 113]

Under normality minimize \[ \text{tr}(VA)^2: \]

[1971b, pp. 447, 453].

\[ R = \eta^{-1} - \eta^{-1} x (x'y^{-1}x')^{-1} x'y^{-1} \]

\[ S = \{ s_{\theta \phi} \} = \{ \text{tr}(V_{\theta} R V_{\phi}) \} \text{ for } \theta, \phi = A, B, \ldots, K + 1 \]

\[ u = \{ u_{\theta} \} = \{ y' R_{\theta} y \} \text{ for } \theta = A, B, \ldots, K + 1 \]

\[ \sigma^2 = s^{-1} u. \]

Iterate on \[ \sigma^2: = (\sigma_A^2, \sigma_B^2, \ldots, \sigma_K^2, \sigma_e^2). \]

MINQUE: Use \[ \sigma^2 = 1 \text{ without iterating; i.e., use } \eta \text{ for } \eta \text{ in } R. \]

[1971a, p. 268; 1972, p. 113]

Example of use: Maddala and Mount [1973]

Generalizations: LaMotte [1973]
5. **Maximum likelihood**

Use normality and same model as MINQUE:

\[
\gamma = \mathbf{Xb} + \sum_{\theta=A}^{K} \mathbf{Z}_\theta \mathbf{u}_\theta + \mathbf{e}
\]

**Notation:**

\[
\gamma_\theta = \frac{\sigma_\theta^2}{\sigma_e^2} \\
\mathbf{H} = \mathbf{I}_N + \sum_{\theta=A}^{K} \gamma_\theta \mathbf{Z}_\theta \mathbf{Z}_\theta'
\]

\[
\text{var}(\gamma) = \sigma_e^2 \mathbf{H} = \mathbf{V}.
\]

**Equations:**

\[
\mathbf{X}'\mathbf{H}^{-1}\mathbf{Xb} = \mathbf{X}'\mathbf{H}^{-1}\gamma
\]

\[
\sigma_e^2 = (\gamma - \mathbf{Xb}')\mathbf{H}^{-1}(\gamma - \mathbf{Xb})/N
\]

\[
\text{tr}(\mathbf{H}^{-1}\mathbf{Z}_\theta \mathbf{Z}_\theta') = (\gamma - \mathbf{Xb}')\mathbf{H}^{-1}\mathbf{Z}_\theta \mathbf{Z}_\theta'\mathbf{H}^{-1}(\gamma - \mathbf{Xb})/\sigma_e^2, \quad \text{for} \ \theta = A, \ldots, K
\]

For unbalanced data these equations have no solution; neither do they for some balanced data situations (e.g. 2-way crossed classification, random model, with interaction). Solutions must be confined to positive values.

**History:**

Hartley and Rao [1967]: Established equations, and solved (numerically) by steepest descent.

Hartley and Vaughn [1972]: Computer program, and small examples.

Harville [1975]: A comprehensive review.

Hemmerle and Hartley [1973]: Newton-Raphson, and a transformation.

Jennrich and Sampson [1976]: Discusses several algorithms.

Miller [1973]: Improved iterative procedure.
6. **REML: Restricted Maximum Likelihood**

Use normality and same model as ML and MINQUE:

\[
\gamma = Xb + \sum_{\theta=A}^{K} \eta_{\theta} y_{\theta} + \varepsilon
\]

Define \( b \) so that \( \gamma \) has full column rank. An easy definition is \( b \equiv \) vector of population means of the filled sub-most cells of the fixed effects factors.

\[
k = \text{number of filled cells, } n_t \text{ observations in } t^{th}, t = 1 \ldots k
\]

\[
\gamma = \sum_{t=1}^{k} \frac{1}{n_t} n_t, \text{ a direct (Kronecker) sum of } 1 \text{-vectors}
\]

\[
S = I - \gamma (X'X)^{-1}X' = \sum_{t=1}^{k} \left( I_{n_t} - \frac{1}{n_t} J_{n_t} \right)
\]

\[
T = S \text{ after deleting rows } n_1, (n_1 + n_2), \ldots, (n_1 + n_2 + \ldots + n_k).
\]

\[
Z = \begin{bmatrix} T^2 \\ X'H^{-1}Y \end{bmatrix} \sim N \left[ \begin{bmatrix} 0 & \text{TH}'\sigma^2_e \\ X'H^{-1}Xb & 0 \end{bmatrix}, \begin{bmatrix} \text{TH}'\sigma^2_e & 0 \\ 0 & X'H^{-1}X\sigma^2_e \end{bmatrix} \right]
\]

To estimate \( \sigma^2 \), maximize the likelihood of \( T^2 \), which does not involve \( b \).

**History:**

Patterson and Thompson [1971]: Initial ideas, confined to b.i.b. designs.

Hocking and Kutner [1975]: Simulations on a b.i.b. design.

Harville [1975]: Comprehensive review.

Corbeil and Searle [1976]: Generalization, and computing procedures using Hemmerle and Hartley [1973].

Corbeil and Searle [1977]: Analytic comparisons for balanced data and numeric comparisons for unbalanced data.
7. Relationships among Methods

(1) ANOVA = Henderson 1 (Definition).

(2) ML estimators are ML solutions subject to non-negativity conditions.

Balanced Data

(3) ANOVA = Henderson 2 = Henderson 3 = REML = MINQUE (MIVQUE).

(4) Some ML equations have no closed form solution. When solutions do exist, some (but not all) = ANOVA. (Differences occur in "degrees of freedom").

Unbalanced Data

ML and MINQUE (MIVQUE) model:

\[ y_{N\times1} = Xb + \sum_{\theta=A}^{K} Z_{\theta}u_{\theta} + \varepsilon, \]  
with \( u_{\theta} \) order \( n_{\theta} \times 1 \).

Henderson's mixed model equations (HBBE's)

\[ \gamma_{\theta} = \sigma_{\theta}^2 / \sigma_{e}^2 \quad D = \text{diag}\{\gamma_{\theta} I_{n_{\theta}}\} \quad \theta = A, \ldots, K \]

\[
\begin{bmatrix}
X'X & X'Z \\
Z'X & Z'Z + D^{-1}
\end{bmatrix}
\begin{bmatrix}
b^* \\
u^*
\end{bmatrix} =
\begin{bmatrix}
X'Y \\
Z'Y
\end{bmatrix}
\]

with solution

\[
\begin{bmatrix}
b^* \\
u^*
\end{bmatrix} =
\begin{bmatrix}
C_{00} & C_{01} \\
C_{10} & C_{11}
\end{bmatrix}
\begin{bmatrix}
X'Y \\
Z'Y
\end{bmatrix}
\]

where

\[ C_{11} = \{C_{0\phi}\} \quad \theta, \phi = A, \ldots, K \]

\[ u^* = \{u_{\theta}\} \quad \theta = A, \ldots, K \]
MINQUE (MIVQUE) and the HMMEM's

MINQUE equations

\[ S_\theta^2 = q, \text{ i.e. } \{s_\theta^2\}^2 = \{q_\theta\} \text{ for } \theta, \phi = A, \ldots, K, e \]

are given by

\[ s_{\theta \theta} = \frac{\text{tr}(C_{\theta \theta})}{\gamma_\theta^2} - \frac{2\text{tr}(C_{\theta \theta})}{\gamma_\theta + n_\theta} \]

\[ s_{\theta \phi} = \frac{\text{tr}(C_{\theta \phi} C_{\theta \phi})}{\gamma_\phi^2} \]

\[ s_{\theta e} = s_e = \text{tr}[C_{\theta \theta} - \sum_{\phi=A}^{K} \frac{\text{tr}(C_{\theta \phi} C_{\phi \theta})}{\gamma_\phi}/\gamma_\theta] \]

\[ s_{ee} = N - r(\text{HMMEM's}) + \sum_{\theta=A}^{K} \sum_{\phi=A}^{K} \text{tr}(C_{\theta \phi} C_{\phi \theta})/\gamma_\theta \gamma_\phi \]

\[ q_\theta = \frac{u_{\theta}^* u_{\theta}^*}{\gamma_\theta^2} \]

\[ q_e = \gamma' \gamma - b^* x' x - u_{\theta}^* Z y - \sum_{\theta=A}^{K} \gamma_\theta q_\theta \]

ML and the HMMEM's

(6) Iterate ML using

\[ s_e^2 = \frac{y'(y - x b^* - z y^*)}{n} \quad \text{and} \quad s_\theta^2 = \frac{u_{\theta}^* u_{\theta}^*}{n_\theta - \text{tr}(T_{\theta \theta})} \]

for \( (I + Z' Z)^{-1} = \{T_{\theta \phi}\}, \theta, \phi = A, \ldots, K. \)

This iteration always gives positive estimators.

REML and MINQUE (MIVQUE)

(7) REML equations = MINQUE equations.

(8) REML estimators = Iterative MINQUE estimators.

(9) First iterate of REML = A MINQUE estimator.

History:

Patterson and Thompson [1971]: First indication of result for \( q_\theta \), for b.i.b. design.

Henderson [1973]: Extended results, HMMEM's ML and MINQUE.

La Motte [1973]: Indicated results for REML and MINQUE.

Schaeffer [1975]: Published same details HMMEM's and MINQUE.

Harville [1975]: Comprehensive review.
REFERENCES

0. **Analysis of variance method** (Henderson's Method 1)

    (For random models but not mixed models)


1. **Henderson's Method 2** (for models with no fixed-by-random interactions)

    Henderson, C. R. [1953].


2. **Fitting constants method** (Henderson's Method 3)

    Henderson, C. R. [1953].


3. **Thompson's iterative method**


    Searle, S. R. [1971a].


4. **MINQUE**


5. **Maximum Likelihood**


6. **Restricted Maximum Likelihood**


Harville, D. A. [1975].

Patterson, H. D. and Thompson, R. [1971]. Recovery of inter-block information when block sizes are unequal. Biometrika 58, 545-554.

7. **Relationships among ML, REML and MINQUE**

Harville, D. A. [1975].

