

# Variance Components (VARCOMP)

---

The Variance Components procedure provides estimates for variances of random effects under a general linear model framework. Four types of estimation methods are available in this procedure.

## Notation

The following notation is used throughout this chapter. Unless otherwise stated, all vectors are column vectors and all quantities are known.

$n$	Number of observations, $n \geq 1$
$k$	Number of random effects, $k \geq 0$
$m_0$	Number of parameters in the fixed effects, $m_0 \geq 0$
$m_i$	Number of parameters in the $i$ th random effect, $m_i \geq 0$ , $i = 1, \dots, k$
$m$	Total number of parameters, $m = m_0 + m_1 + \dots + m_k$
$\sigma_i^2$	Unknown variance of the $i$ th random effect, $\sigma_i^2 \geq 0$ , $i = 1, \dots, k$
$\sigma_e^2$	Unknown variance of the residual term, same as $\sigma_{k+1}^2$ , $\sigma_e^2 > 0$
$\gamma_i^2$	Unknown variance ratio of the $i$ th random effect, $\gamma_i^2 = \sigma_i^2 / \sigma_e^2$ , $\gamma_i^2 \geq 0$ , $i = 1, \dots, k$ , and $\gamma_{k+1}^2 = 1$
$\mathbf{y}$	The length $n$ vector of observations
$\mathbf{e}$	The length $n$ vector of residuals
$\mathbf{X}_i$	The $n \times m_i$ design matrix, $i = 0, 1, \dots, k$
$\beta_0$	The length $m_0$ vector of parameters of the fixed effects
$\beta_i$	The length $m_i$ vector of parameters of the $i$ th random effect, $i = 1, \dots, k$

Unless otherwise stated, a  $p \times p$  identity matrix is denoted as  $\mathbf{I}_p$ , a  $p \times q$  zero matrix is denoted as  $\mathbf{0}_{p \times q}$ , and a zero vector of length  $p$  is denoted as  $\mathbf{0}_p$ .

## Weights

For the sake of clarity and simplicity, the algorithms described in this chapter assume unit frequency weight and unit regression weight for all cases. Weights can be applied as described in the following two sections.

### Frequency Weight

The WEIGHT command specifies frequency weights.

- Cases with nonpositive frequency are excluded from all calculations in the procedure.
- Non-integral frequency weight is rounded to the nearest integer.
- The total sample size is equal to the sum of positive rounded frequency weights.

### Regression Weight

The REGWGT subcommand specifies regression weights. Suppose the  $l$ th ( $l = 1, \dots, n$ ) case has a regression weight  $w_l > 0$  (cases with nonpositive regression weights are excluded from all calculations in the procedure). Let  $\mathbf{W} = \text{diag}(w_1, \dots, w_n)$  be the  $n \times n$  diagonal weight matrix. Then the VARCOMP procedure will perform all calculations as if  $\mathbf{y}$  is physically transformed to  $\mathbf{W}^{\frac{1}{2}}\mathbf{y}$  and  $\mathbf{X}_i$  to  $\mathbf{W}^{\frac{1}{2}}\mathbf{X}_i$ ,  $i = 0, 1, \dots, k$ ; and then the pertinent algorithm is applied to the transformed data.

## Model

The mixed model is represented, following Rao (1973), as

$$\mathbf{y} = \mathbf{X}_0\beta_0 + \sum_{i=1}^k \mathbf{X}_i\beta_i + \mathbf{e}$$

The random vectors  $\beta_1, \dots, \beta_k$  and  $\mathbf{e}$  are assumed to be jointly independent. Moreover, the random vector  $\beta_i$  is distributed as  $N_{m_i}(\mathbf{0}, \sigma_i^2 \mathbf{I}_{m_i})$  for  $i = 1, \dots, k$  and the residual vector  $\mathbf{e}$  is distributed as  $N_n(\mathbf{0}, \sigma_e^2 \mathbf{I}_n)$ . It follows from these assumptions that  $\mathbf{y}$  is distributed as  $N_n(\mathbf{X}_0 \beta_0, \sigma_e^2 \mathbf{V})$  where

$$\mathbf{V} = \sum_{i=1}^k \gamma_i^2 \mathbf{X}_i \mathbf{X}_i' + \mathbf{I}_n = \sum_{i=1}^{k+1} \gamma_i^2 \mathbf{V}_i$$

where  $\mathbf{V}_i = \mathbf{X}_i \mathbf{X}_i'$ ,  $i = 1, \dots, k$ , and  $\mathbf{V}_i = \mathbf{I}_n$ .

## Minimum Norm Quadratic Unbiased Estimate (MINQUE)

Given the initial guess or the prior values  $\gamma_i^2 = \alpha_i$  ( $\alpha_i \geq 0$ ),  $i = 1, \dots, k+1$ , the MINQUE of  $\sigma$  are obtained as a solution of the linear system of equations:

$$\mathbf{S}\sigma = \mathbf{q}$$

where  $\mathbf{S} = \{s_{ij}\}$  is a  $(k+1) \times (k+1)$  symmetric matrix,  $\mathbf{q} = \{q_i\}$  is a  $(k+1)$ -vector, and  $\sigma' = (\sigma_1^2, \dots, \sigma_{k+1}^2)$ . Define

$$\mathbf{R} = \mathbf{V}^{-1} - \mathbf{V}^{-1} \mathbf{X}_0 (\mathbf{X}_0' \mathbf{V}^{-1} \mathbf{X}_0)^{-1} \mathbf{X}_0' \mathbf{V}^{-1}$$

The elements of  $\mathbf{S}$  and  $\mathbf{q}$  are

$$s_{ij} = \begin{cases} \text{SSQ}(\mathbf{X}_i' \mathbf{R} \mathbf{X}_j) & i = 1, \dots, k, \quad j = 1, \dots, k \\ \text{SSQ}(\mathbf{X}_i' \mathbf{R}) & i = 1, \dots, k, \quad j = k+1 \\ \text{SSQ}(\mathbf{R} \mathbf{X}_j) & i = k+1, \quad j = 1, \dots, k \\ \text{SSQ}(\mathbf{R}) & i = k+1, \quad j = k+1 \end{cases}$$

#### 4 Variance Components (VARCOMP)

and

$$q_i = \begin{cases} \text{SSQ}(\mathbf{X}'_i \mathbf{Ry}) & i = 1, \dots, k \\ \text{SSQ}(\mathbf{Ry}) & i = k + 1 \end{cases}$$

where  $\text{SSQ}(\mathbf{A})$  is the sum of squares of all elements of a matrix  $\mathbf{A}$ .

#### MINQUE(0)

The prior values are  $\alpha_i = 0$ ,  $i = 1, \dots, k$ , and  $\alpha_{k+1} = 1$ . Under this set of prior values,  $\mathbf{V} = \mathbf{I}_n$  and  $\mathbf{R} = \mathbf{I}_n - \mathbf{X}_0(\mathbf{X}'_0 \mathbf{X}_0)^- \mathbf{X}'_0$ . Since this  $\mathbf{R}$  is an idempotent matrix, some of the elements of  $\mathbf{S}$  and  $\mathbf{q}$  can be simplified to

$$s_{i,k+1} = \text{trace}(\mathbf{X}'_i \mathbf{R} \mathbf{X}_i) \quad i = 1, \dots, k;$$

$$s_{k+1,j} = \text{trace}(\mathbf{X}'_j \mathbf{R} \mathbf{X}_j) \quad j = 1, \dots, k;$$

$$s_{k+1,k+1} = n - \text{rank}(\mathbf{X}_0)$$

$$q_{k+1} = \mathbf{y}' \mathbf{R} \mathbf{y}$$

Using the algorithm by Goodnight (1978), the elements of  $\mathbf{S}$  and  $\mathbf{q}$  are obtained without explicitly computing  $\mathbf{R}$ . The steps are described as follows:

**Step 1.** Form the symmetric matrix:

$$\begin{bmatrix} \mathbf{X}'_0 \mathbf{X}_0 & \mathbf{X}'_0 \mathbf{X}_1 & \cdots & \mathbf{X}'_0 \mathbf{X}_k & \mathbf{X}'_0 \mathbf{y} \\ \mathbf{X}'_1 \mathbf{X}_0 & \mathbf{X}'_1 \mathbf{X}_1 & \cdots & \mathbf{X}'_1 \mathbf{X}_k & \mathbf{X}'_1 \mathbf{y} \\ \vdots & \vdots & & \vdots & \vdots \\ \mathbf{X}'_k \mathbf{X}_0 & \mathbf{X}'_k \mathbf{X}_1 & \cdots & \mathbf{X}'_k \mathbf{X}_k & \mathbf{X}'_k \mathbf{y} \\ \mathbf{y}' \mathbf{X}_0 & \mathbf{y}' \mathbf{X}_1 & \cdots & \mathbf{y}' \mathbf{X}_k & \mathbf{y}' \mathbf{y} \end{bmatrix}$$

**Step 2.** Sweep the above matrix by pivoting on each diagonal of  $\mathbf{X}'_0\mathbf{X}_0$ . This produces the following matrix:

$$\begin{bmatrix} \mathbf{G} & \mathbf{GX}'_0\mathbf{X}_1 & \cdots & \mathbf{GX}'_0\mathbf{X}_k & \mathbf{GX}'_0\mathbf{y} \\ \mathbf{X}'_1\mathbf{X}_0\mathbf{G} & \mathbf{X}'_1\mathbf{RX}_1 & \cdots & \mathbf{X}'_1\mathbf{RX}_k & \mathbf{X}'_1\mathbf{Ry} \\ \vdots & \vdots & & \vdots & \vdots \\ \mathbf{X}'_k\mathbf{X}_0\mathbf{G} & \mathbf{X}'_k\mathbf{RX}_1 & \cdots & \mathbf{X}'_k\mathbf{RX}_k & \mathbf{X}'_k\mathbf{Ry} \\ \mathbf{y}'\mathbf{X}_0\mathbf{G} & \mathbf{y}'\mathbf{RX}_1 & \cdots & \mathbf{y}'\mathbf{RX}_k & \mathbf{y}'\mathbf{Ry} \end{bmatrix}$$

where  $\mathbf{G} = (\mathbf{X}'_0\mathbf{X}_0)^-$ . In the process of computing the above matrix, the rank of  $\mathbf{X}_0$  is obtained as the number of nonzero pivots found.

**Step 3.** Form  $\mathbf{S}$  and  $\mathbf{q}$ . The MINQUE(0) of  $\sigma$  are  $\hat{\sigma} = \mathbf{S}^- \mathbf{q}$ .

## MINQUE(1)

The prior values are  $\alpha_i = 1$ ,  $i = 1, \dots, k+1$ . Under this set of prior values,  $\mathbf{V} = \sum_{i=1}^{k+1} \mathbf{X}'_i\mathbf{X}_i$ . Using Giesbrecht (1983), the matrix  $\mathbf{S}$  and the vector  $\mathbf{q}$  are obtained through an iterative procedure. The steps are described as follows:

**Step 1.** Construct the augmented matrix  $\mathbf{A} = [\mathbf{X}_0 | \mathbf{X}_1 | \cdots | \mathbf{X}_k | \mathbf{y}]$ . Then compute the  $(m+1) \times (m+1)$  matrix  $\mathbf{T}_{(k+1)} = \mathbf{A}'\mathbf{A}$ .

**Step 2.** Define  $\mathbf{H}_{(l)} = \sum_{i=l}^{k+1} \mathbf{X}'_i\mathbf{X}_i$ , and  $\mathbf{T}_{(l)} = \mathbf{A}'\mathbf{H}_{(l)}^{-1}\mathbf{A}$ ,  $l = 1, \dots, k$ . Update  $\mathbf{T}_{(l+1)}$  to  $\mathbf{T}_{(l)}$  using the  $W$  Transform given in Goodnight and Hemmerle (1979). The updating formula is

$$\mathbf{T}_{(l)} = \mathbf{T}_{(l+1)} - \mathbf{A}'\mathbf{H}_{(l+1)}^{-1}\mathbf{X}_l \left( \mathbf{I}_{m_l} + \mathbf{X}'_l\mathbf{H}_{(l+1)}^{-1}\mathbf{X}_l \right)^- \mathbf{X}'_l\mathbf{H}_{(l+1)}^{-1}\mathbf{A}$$

## 6 Variance Components (VARCOMP)

**Step 3.** Once  $\mathbf{T}_{(1)} = \mathbf{A}'\mathbf{H}_{(1)}^{-1}\mathbf{A} = \mathbf{A}'\mathbf{V}^{-1}\mathbf{A}$  is obtained, apply the Sweep operation to the diagonal elements of upper left  $m_0 \times m_0$  submatrix of  $\mathbf{T}_{(1)}$ . The resulting matrix will contain the quadratic form  $\mathbf{y}'\mathbf{R}\mathbf{y}$ , the vectors  $\mathbf{y}'\mathbf{R}\mathbf{X}_j$ ,  $j = 1, \dots, k$ , and the matrices  $\mathbf{X}_i'\mathbf{R}\mathbf{X}_j$ ,  $i = 1, \dots, k$ ,  $j = 1, \dots, k$ .

**Step 4.** Compute the elements of  $\mathbf{S}$  and  $\mathbf{q}$ . Since  $\mathbf{R}\mathbf{V}\mathbf{R} = \mathbf{R}$ , then

$$SSQ(\mathbf{R}\mathbf{X}_j) = \text{tr}(\mathbf{X}_j'\mathbf{R}\mathbf{X}_j) - \sum_{i=1}^k SSQ(\mathbf{X}_i'\mathbf{R}\mathbf{X}_i) \quad j = 1, \dots, k$$

$$SSQ(\mathbf{R}) = n - \text{rank}(\mathbf{X}_0) - \sum_{i=1}^k \text{tr}(\mathbf{X}_i'\mathbf{R}\mathbf{X}_i) - \sum_{j=1}^k SSQ(\mathbf{R}\mathbf{X}_j)$$

$$SSQ(\mathbf{R}\mathbf{y}) = \mathbf{y}'\mathbf{R}\mathbf{y} - \sum_{j=1}^k SSQ(\mathbf{y}'\mathbf{R}\mathbf{X}_j)$$

The MINQUE(1) of  $\boldsymbol{\sigma}$  are  $\hat{\boldsymbol{\sigma}} = \mathbf{S}^{-1}\mathbf{q}$ .

## Maximum Likelihood Estimate (MLE)

The maximum likelihood estimates are obtained using the algorithm by Jennrich and Sampson (1976). The algorithm is an iterative procedure that combines Newton-Raphson steps and Fisher scoring steps.

### Parameters

The parameter vector is  $\boldsymbol{\theta} = \begin{bmatrix} \beta_0 \\ \gamma^2 \\ \sigma_e^2 \end{bmatrix}$  where  $\gamma^2 = \begin{bmatrix} \gamma_1^2 \\ \vdots \\ \gamma_k^2 \end{bmatrix}$ .

## Likelihood Function

The likelihood function is

$$L \equiv L(\theta) = (2\pi)^{-n/2} |\sigma_e^2 \mathbf{V}|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{X}_0\beta_0)' \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}_0\beta_0) / \sigma_e^2\right).$$

The log-likelihood function is

$$l = \log L = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma_e^2) - \frac{1}{2} \log|\mathbf{V}| - \frac{1}{2\sigma_e^2} (\mathbf{y} - \mathbf{X}_0\beta_0)' \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}_0\beta_0).$$

## The Gradient Vector

$$\begin{aligned} \frac{\partial l}{\partial \beta_0} &= \frac{1}{\sigma_e^2} \mathbf{X}_0' \mathbf{V}^{-1} \mathbf{r}, \\ \frac{\partial l}{\partial \gamma_i^2} &= \frac{1}{2\sigma_e^2} \mathbf{r}' \mathbf{V}^{-1} \mathbf{X}_i \mathbf{X}_i' \mathbf{V}^{-1} \mathbf{r} - \frac{1}{2} \text{tr}(\mathbf{X}_i' \mathbf{V}^{-1} \mathbf{X}_i), \quad i = 1, \dots, k, \\ \frac{\partial l}{\partial \sigma_e^2} &= \frac{1}{2\sigma_e^4} \mathbf{r}' \mathbf{V}^{-1} \mathbf{r} - \frac{n}{2\sigma_e^2}. \end{aligned}$$

where  $\mathbf{r} = \mathbf{y} - \mathbf{X}_0\beta_0$ . The gradient vector is

$$\frac{dl}{d\theta} = \begin{bmatrix} \frac{\partial l}{\partial \beta_0} \\ \frac{\partial l}{\partial \gamma^2} \\ \frac{\partial l}{\partial \sigma_e^2} \end{bmatrix}$$

## 8 Variance Components (VARCOMP)

### The Hessian Matrix

$$\frac{\partial^2 l}{\partial \beta_0 \partial \beta_0} = -\frac{1}{\sigma_e^2} \mathbf{X}_0' \mathbf{V}^{-1} \mathbf{X}_0$$

$$\frac{\partial^2 l}{\partial \gamma_i^2 \partial \beta_0} = -\frac{1}{\sigma_e^2} \mathbf{r}' \mathbf{V}^{-1} \mathbf{X}_i \mathbf{X}_i' \mathbf{V}^{-1} \mathbf{X}_0 \quad i = 1, \dots, k,$$

$$\frac{\partial^2 l}{\partial \gamma_i^2 \partial \gamma_j^2} = \frac{1}{2} \text{tr}(\mathbf{X}_i' \mathbf{V}^{-1} \mathbf{X}_j \mathbf{X}_j' \mathbf{V}^{-1} \mathbf{X}_i) - \frac{1}{\sigma_e^2} \mathbf{r}' \mathbf{V}^{-1} \mathbf{X}_i \mathbf{X}_i' \mathbf{V}^{-1} \mathbf{X}_j \mathbf{X}_j' \mathbf{V}^{-1} \mathbf{r} \quad i = 1, \dots, k; j = 1, \dots, k,$$

$$\frac{\partial^2 l}{\partial \sigma_e^2 \partial \beta_0} = -\frac{1}{\sigma_e^4} \mathbf{r}' \mathbf{V}^{-1} \mathbf{X}_0$$

$$\frac{\partial^2 l}{\partial \sigma_e^2 \partial \gamma_j^2} = -\frac{1}{2\sigma_e^4} \mathbf{r}' \mathbf{V}^{-1} \mathbf{X}_j \mathbf{X}_j' \mathbf{V}^{-1} \mathbf{r} \quad j = 1, \dots, k$$

$$\frac{\partial^2 l}{\partial \sigma_e^2 \partial \sigma_e^2} = \frac{n}{2\sigma_e^4} - \frac{1}{\sigma_e^6} \mathbf{r}' \mathbf{V}^{-1} \mathbf{r}$$

The Hessian matrix is

$$\frac{d^2 l}{d\theta d\theta} = \begin{bmatrix} \frac{\partial^2 l}{\partial \beta_0 \partial \beta_0} & \frac{\partial^2 l}{\partial \beta_0 \partial \gamma^2} & \frac{\partial^2 l}{\partial \beta_0 \partial \sigma_e^2} \\ \frac{\partial^2 l}{\partial \gamma^2 \partial \beta_0} & \frac{\partial^2 l}{\partial \gamma^2 \partial \gamma^2} & \frac{\partial^2 l}{\partial \gamma^2 \partial \sigma_e^2} \\ \frac{\partial^2 l}{\partial \sigma_e^2 \partial \beta_0} & \frac{\partial^2 l}{\partial \sigma_e^2 \partial \gamma^2} & \frac{\partial^2 l}{\partial \sigma_e^2 \partial \sigma_e^2} \end{bmatrix}$$



where

$$\frac{\partial^2 l}{\partial \gamma^2 \partial \beta_0} = \begin{bmatrix} \frac{\partial^2 l}{\partial \gamma_1^2 \partial \beta_0} \\ \vdots \\ \frac{\partial^2 l}{\partial \gamma_k^2 \partial \beta_0} \end{bmatrix}, \quad \frac{\partial^2 l}{\partial \gamma^2 \partial \gamma^2} = \begin{bmatrix} \frac{\partial^2 l}{\partial \gamma_1^2 \partial \gamma_1^2} & \cdots & \frac{\partial^2 l}{\partial \gamma_1^2 \partial \gamma_k^2} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 l}{\partial \gamma_k^2 \partial \gamma_1^2} & \cdots & \frac{\partial^2 l}{\partial \gamma_k^2 \partial \gamma_k^2} \end{bmatrix}$$

and

$$\frac{\partial^2 l}{\partial \sigma_e^2 \partial \gamma^2} = \begin{bmatrix} \frac{\partial^2 l}{\partial \sigma_e^2 \partial \gamma_1^2} \\ \vdots \\ \frac{\partial^2 l}{\partial \sigma_e^2 \partial \gamma_k^2} \end{bmatrix}$$

### The Fisher Information Matrix

As  $E(\mathbf{r}) = \mathbf{0}_n$  and  $E(\mathbf{r}' \mathbf{V}^{-1} \mathbf{r}) = n\sigma_e^2$ , the expected second derivatives are

$$E\left(\frac{\partial^2 l}{\partial \beta_0 \partial \beta_0}\right) = -\frac{1}{\sigma_e^2} \mathbf{X}_0' \mathbf{V}^{-1} \mathbf{X}_0$$

$$E\left(\frac{\partial^2 l}{\partial \gamma_i^2 \partial \beta_0}\right) = \mathbf{0}'_{m_0} \quad i = 1, \dots, k$$

$$E\left(\frac{\partial^2 l}{\partial \gamma_i^2 \partial \gamma_j^2}\right) = -\frac{1}{2} \text{tr}(\mathbf{X}_i' \mathbf{V}^{-1} \mathbf{X}_j \mathbf{X}_j' \mathbf{V}^{-1} \mathbf{X}_i) \quad i = 1, \dots, k, j = 1, \dots, k$$

10 Variance Components (VARCOMP)

$$E\left(\frac{\partial^2 l}{\partial \sigma_e^2 \partial \beta_0}\right) = \mathbf{0}'_{m_0}$$

$$E\left(\frac{\partial^2 l}{\partial \sigma_e^2 \partial \gamma_j^2}\right) = -\frac{1}{2\sigma_e^2} \text{tr}(\mathbf{X}'_j \mathbf{V}^{-1} \mathbf{X}_j) \quad j = 1, \dots, k,$$

$$E\left(\frac{\partial^2 l}{\partial \sigma_e^2 \partial \sigma_e^2}\right) = -\frac{n}{2\sigma_e^4}$$

The Fisher Information matrix is

$$E\left(\frac{d^2 l}{d\theta d\theta}\right) = \begin{bmatrix} \frac{\partial^2 l}{\partial \beta_0 \partial \beta_0} & \mathbf{0}_{m_0 \times (m-m_0)} & \mathbf{0}_{m_0} \\ \mathbf{0}_{(m-m_0) \times m_0} & E\left(\frac{\partial^2 l}{\partial \gamma^2 \partial \gamma^2}\right) & E\left(\frac{\partial^2 l}{\partial \gamma^2 \partial \sigma_e^2}\right) \\ \mathbf{0}'_{m_0} & E\left(\frac{\partial^2 l}{\partial \sigma_e^2 \partial \gamma^2}\right) & E\left(\frac{\partial^2 l}{\partial \sigma_e^2 \partial \sigma_e^2}\right) \end{bmatrix}$$

where

$$E\left(\frac{\partial^2 l}{\partial \gamma^2 \partial \gamma^2}\right) = \begin{bmatrix} E\left(\frac{\partial^2 l}{\partial \gamma_1^2 \partial \gamma_1^2}\right) & \dots & E\left(\frac{\partial^2 l}{\partial \gamma_1^2 \partial \gamma_k^2}\right) \\ \vdots & \ddots & \vdots \\ E\left(\frac{\partial^2 l}{\partial \gamma_k^2 \partial \gamma_1^2}\right) & \dots & E\left(\frac{\partial^2 l}{\partial \gamma_k^2 \partial \gamma_k^2}\right) \end{bmatrix}$$

and

$$E\left(\frac{\partial^2 l}{\partial \sigma_e^2 \partial \gamma^2}\right) = \begin{bmatrix} E\left(\frac{\partial^2 l}{\partial \sigma_e^2 \partial \gamma_1^2}\right) \\ \vdots \\ E\left(\frac{\partial^2 l}{\partial \sigma_e^2 \partial \gamma_k^2}\right) \end{bmatrix}$$

## Iteration Procedure

### Initial Values

- *Fixed Effect Parameters:*  $\hat{\beta}_0 = (\mathbf{X}'_0 \mathbf{X}_0)^{-1} \mathbf{X}'_0 \mathbf{y}$ .
- *Random Effect Variance Components:* For the  $i$ th ( $i = 1, \dots, k$ ) random effect, compute  $\hat{\beta}_i = (\mathbf{X}'_i \mathbf{X}_i)^{-1} \mathbf{X}'_i \mathbf{y}$ . Then assign the variance of the  $m_i$  elements of  $\hat{\beta}_i$  using divisor  $(m_i - 1)$  to the estimate  $\hat{\sigma}_i^2$  if  $m_i \geq 2$ ; otherwise  $\hat{\sigma}_i^2 = 0$ .
- *Residual Variance:*  $\hat{\sigma}_e^2 = \mathbf{r}' \mathbf{r} / n$  where  $\mathbf{X} = [\mathbf{X}_0 | \mathbf{X}_1 | \dots | \mathbf{X}_k]$  and  $\mathbf{r} = \mathbf{y} - (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y}$ . If  $\hat{\sigma}_e^2 = 0$  but  $k \geq 1$  then reset  $\hat{\sigma}_e^2 = 10^{-8}$  so that the iteration can continue.

The variance ratios are then computed as  $\hat{\gamma}_i^2 = \hat{\sigma}_i^2 / \hat{\sigma}_e^2$ ,  $i = 1, \dots, k$ . Following the same method in which the residual variance is initialized,  $\hat{\sigma}_e^2 > 0$  for  $k \geq 1$ .

### Updating

At the  $s$ th iteration ( $s = 0, 1, \dots$ ), the parameter vector is updated as

$$\hat{\theta}_{(s+1)} = \hat{\theta}_{(s)} + \rho \Delta \hat{\theta}_{(s)}$$

where  $\Delta \hat{\theta}_{(s)}$  is the value of increment  $\Delta \theta$  evaluated at  $\theta = \hat{\theta}_{(s)}$ , and  $\rho > 0$  is a step size such that  $l(\hat{\theta}_{(s+1)}) \geq l(\hat{\theta}_{(s)})$ . The increment vector depends on the choice of step type—Newton-Raphson versus Fisher scoring. The step size is determined by the step-halving technique with  $\rho = 1$  initially and a maximum of 10 halvings.

## 12 Variance Components (VARCOMP)

### Choice of Step

Following Jennrich and Sampson (1976), the first iteration is always the Fisher scoring step because it is more robust to poor initial values. For subsequent iteration the Newton-Raphson step is used if:

1. The Hessian matrix is nonnegative definite, and
2. The increment in the log-likelihood function of step 1 is less than or equal to one.

Otherwise the Fisher scoring step is used. The increment vector for each type of step is:

- *Newton-Raphson Step*:  $\Delta\theta = \left( -\frac{d^2l}{d\theta d\theta} \right)^{-1} \frac{dl}{d\theta}$ .
- *Fisher Scoring Step*:  $\Delta\theta = \left( -E \left( \frac{d^2l}{d\theta d\theta} \right) \right)^{-1} \frac{dl}{d\theta}$ .

### Convergence Criteria

Given the convergence criterion  $\varepsilon > 0$ , the iteration is considered converged when the following criteria are satisfied:

1.  $\left| l(\hat{\theta}_{(s+1)}) - l(\hat{\theta}_{(s)}) \right| < \varepsilon \times \max\left(1, \left| l(\hat{\theta}_{(s)}) \right| \right)$ , and
2.  $\left\langle \rho(\hat{\theta}_{(s+1)} - \hat{\theta}_{(s)}) \right\rangle < \varepsilon \times \max\left(1, \left\langle \hat{\theta}_{(s)} \right\rangle \right)$  where  $\langle \mathbf{a} \rangle$  is the sum of absolute values of elements of the vector  $\mathbf{a}$ .

### Negative Variance Estimates

Negative variance estimates can occur at the end of an iteration. An *ad hoc* method is to set those estimates to zero before the next iteration.

## Covariance Matrix

Let  $\hat{\theta}$  be the vector of maximum likelihood estimates. Their covariance matrix is given by

$$\text{cov}(\hat{\theta}) = \left( -E \left( \frac{d^2 l}{d\theta d\theta} \right) \Big|_{\theta=\hat{\theta}} \right)^{-1}$$

Let

$$\psi = \begin{bmatrix} \beta_0 \\ \sigma_1^2 \\ \vdots \\ \sigma_k^2 \\ \sigma_e^2 \end{bmatrix}$$

be the original parameters. Their maximum likelihood estimates are given by

$$\hat{\psi} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\sigma}_e^2 \hat{\gamma}_1^2 \\ \vdots \\ \hat{\sigma}_e^2 \hat{\gamma}_k^2 \\ \hat{\sigma}_e^2 \end{bmatrix}$$

and their covariance matrix is estimated by

$$\text{cov}(\hat{\psi}) = \mathbf{J} \text{cov}(\hat{\theta}) \mathbf{J}'$$

## 14 Variance Components (VARCOMP)

where

$$\mathbf{J} = \begin{bmatrix} \mathbf{I}_{m_0} & \mathbf{0}_{m_0 \times k} & \mathbf{0}_{m_0} \\ \mathbf{0}_{k \times m_0} & \sigma_e^2 \mathbf{I}_k & \gamma \\ 0 & 0 & 1 \end{bmatrix}$$

which is the  $(m_0 + k + 1) \times (m_0 + k + 1)$  Jacobian matrix of transforming  $\theta$  to  $\psi$ .

## Restricted Maximum Likelihood Estimate (REML)

The restricted maximum likelihood method finds a linear transformation on  $\mathbf{y}$  such that the resulting vector does not involve the fixed effect parameter vector  $\beta_0$  regardless of their values. It has been shown that these linear combinations are the residuals obtained after a linear regression on the fixed effects. Suppose  $r$  is the rank of  $\mathbf{X}_0$ ; then there are at most  $n - r$  linearly independent combinations. Let  $\mathbf{K}$  be an  $n \times (n - r)$  matrix whose columns are these linearly independent combinations. Then the properties of  $\mathbf{K}$  are (Searle et al., 1992, Chapter 6):

$$\mathbf{K}'\mathbf{X}_0 = \mathbf{0}_{(n-r) \times m_0}$$

$$\mathbf{K}' = \mathbf{T}\mathbf{M}$$

where  $\mathbf{T}$  is a  $(n - r) \times n$  matrix with linearly independent rows and

$$\mathbf{M} = \mathbf{I}_n - \mathbf{X}_0(\mathbf{X}_0'\mathbf{X}_0)^{-}\mathbf{X}_0'$$

It can be shown that REML estimation is invariant to  $\mathbf{K}$  (Searle et al., 1992, Chapter 6); thus, we can choose  $\mathbf{K}$  such that  $\mathbf{K}'\mathbf{K} = \mathbf{I}_{n-r}$  to simplify calculations. It follows that the distribution of  $\mathbf{K}'\mathbf{y}$  is  $N_{n-r}(\mathbf{0}, \sigma_e^2 \mathbf{K}'\mathbf{V}\mathbf{K})$ .

### Parameters

The parameter vector is  $\theta = \begin{bmatrix} \gamma^2 \\ \sigma_e^2 \end{bmatrix}$  where  $\gamma^2 = \begin{bmatrix} \gamma_1^2 \\ \vdots \\ \gamma_k^2 \end{bmatrix}$ .

## Likelihood Function

The likelihood function of  $\mathbf{K}'\mathbf{y}$  is

$$L \equiv L(\theta) = (2\pi)^{-(n-r)/2} |\sigma_e^2 \mathbf{K}'\mathbf{V}\mathbf{K}|^{-1/2} \exp\left(-\frac{1}{2} \mathbf{y}'\mathbf{K}(\mathbf{K}'\mathbf{V}\mathbf{K})^{-1} \mathbf{K}'\mathbf{y} / \sigma_e^2\right).$$

It can be shown (Searle et al., 1992) that

$$\mathbf{R} \equiv \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}_0(\mathbf{X}_0'\mathbf{V}^{-1}\mathbf{X}_0)^{-1}\mathbf{X}_0'\mathbf{V}^{-1} = \mathbf{K}(\mathbf{K}'\mathbf{V}\mathbf{K})^{-1}\mathbf{K}'$$

Thus, the log-likelihood function is

$$l = \log L = -\frac{n-r}{2} \log(2\pi) - \frac{n-r}{2} \log(\sigma_e^2) - \frac{1}{2} \log|\mathbf{K}'\mathbf{V}\mathbf{K}| - \frac{1}{2\sigma_e^2} \mathbf{y}'\mathbf{R}\mathbf{y}.$$

## The Gradient Vector

$$\frac{\partial l}{\partial \gamma_i^2} = \frac{1}{2\sigma_e^2} \mathbf{y}'\mathbf{R}\mathbf{X}_i\mathbf{X}_i'\mathbf{R}\mathbf{y} - \frac{1}{2} \text{tr}(\mathbf{X}_i'\mathbf{R}\mathbf{X}_i) \quad i = 1, \dots, k$$

$$\frac{\partial l}{\partial \sigma_e^2} = \frac{1}{2\sigma_e^4} \mathbf{y}'\mathbf{R}\mathbf{y} - \frac{(n-r)}{2\sigma_e^2}$$

The gradient vector is

$$\frac{dl}{d\theta} = \begin{bmatrix} \frac{\partial l}{\partial \gamma^2} \\ \frac{\partial l}{\partial \sigma_e^2} \end{bmatrix}$$

### The Hessian Matrix

$$\frac{\partial^2 l}{\partial \gamma_i^2 \partial \gamma_j^2} = -\frac{1}{\sigma_e^2} \mathbf{y}' \mathbf{R} \mathbf{X}_i \mathbf{X}_i' \mathbf{R} \mathbf{X}_j \mathbf{X}_j' \mathbf{R} \mathbf{y} + \frac{1}{2} \text{tr}(\mathbf{X}_i' \mathbf{R} \mathbf{X}_j \mathbf{X}_j' \mathbf{R} \mathbf{X}_i) \quad i = 1, \dots, k; j = 1, \dots, k$$

$$\frac{\partial^2 l}{\partial \sigma_e^2 \partial \gamma_j^2} = -\frac{1}{2\sigma_e^4} \mathbf{y}' \mathbf{R} \mathbf{X}_j \mathbf{X}_j' \mathbf{R} \mathbf{y} \quad j = 1, \dots, k$$

$$\frac{\partial^2 l}{\partial \sigma_e^2 \partial \sigma_e^2} = -\frac{1}{\sigma_e^6} \mathbf{y}' \mathbf{R} \mathbf{y} + \frac{n-r}{2\sigma_e^4}$$

The Hessian matrix is

$$\frac{d^2 l}{d\theta d\theta} = \begin{bmatrix} \frac{\partial^2 l}{\partial \gamma^2 \partial \gamma^2} & \frac{\partial^2 l}{\partial \gamma^2 \partial \sigma_e^2} \\ \frac{\partial^2 l}{\partial \sigma_e^2 \partial \gamma^2} & \frac{\partial^2 l}{\partial \sigma_e^2 \partial \sigma_e^2} \end{bmatrix}$$

where

$$\frac{\partial^2 l}{\partial \gamma^2 \partial \gamma^2} = \begin{bmatrix} \frac{\partial^2 l}{\partial \gamma_1^2 \partial \gamma_1^2} & \dots & \frac{\partial^2 l}{\partial \gamma_1^2 \partial \gamma_k^2} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 l}{\partial \gamma_k^2 \partial \gamma_1^2} & \dots & \frac{\partial^2 l}{\partial \gamma_k^2 \partial \gamma_k^2} \end{bmatrix} \quad \text{and} \quad \frac{\partial^2 l}{\partial \sigma_e^2 \partial \gamma^2} = \begin{bmatrix} \frac{\partial^2 l}{\partial \sigma_e^2 \partial \gamma_1^2} \\ \vdots \\ \frac{\partial^2 l}{\partial \sigma_e^2 \partial \gamma_k^2} \end{bmatrix}$$

### The Fisher Information Matrix

Since  $\mathbf{K}'\mathbf{X}_0 = \mathbf{0}_{(n-r) \times m_0}$  and  $\text{trace}(\mathbf{R}\mathbf{V}) = n - r$ , the expected second derivatives are

$$E\left(\frac{\partial^2 l}{\partial \gamma_i^2 \partial \gamma_j^2}\right) = -\frac{1}{2} \text{tr}(\mathbf{X}_i' \mathbf{R} \mathbf{X}_j \mathbf{X}_j' \mathbf{R} \mathbf{X}_i) \quad i = 1, \dots, k, j = 1, \dots, k$$



$$E\left(\frac{\partial^2 l}{\partial \sigma_e^2 \partial \gamma_j^2}\right) = -\frac{1}{2\sigma_e^2} \text{tr}(\mathbf{X}'_j \mathbf{R} \mathbf{X}_j) \quad j = 1, \dots, k$$

$$E\left(\frac{\partial^2 l}{\partial \sigma_e^2 \partial \sigma_e^2}\right) = -\frac{n-r}{2\sigma_e^4}$$

The Fisher Information matrix is

$$E\left(\frac{d^2 l}{d\theta d\theta}\right) = \begin{bmatrix} E\left(\frac{\partial^2 l}{\partial \gamma^2 \partial \gamma^2}\right) & E\left(\frac{\partial^2 l}{\partial \gamma^2 \partial \sigma_e^2}\right) \\ E\left(\frac{\partial^2 l}{\partial \sigma_e^2 \partial \gamma^2}\right) & E\left(\frac{\partial^2 l}{\partial \sigma_e^2 \partial \sigma_e^2}\right) \end{bmatrix}$$

where

$$E\left(\frac{\partial^2 l}{\partial \gamma^2 \partial \gamma^2}\right) = \begin{bmatrix} E\left(\frac{\partial^2 l}{\partial \gamma_1^2 \partial \gamma_1^2}\right) & \cdots & E\left(\frac{\partial^2 l}{\partial \gamma_1^2 \partial \gamma_k^2}\right) \\ \vdots & \ddots & \vdots \\ E\left(\frac{\partial^2 l}{\partial \gamma_k^2 \partial \gamma_1^2}\right) & \cdots & E\left(\frac{\partial^2 l}{\partial \gamma_k^2 \partial \gamma_k^2}\right) \end{bmatrix}$$

and

$$E\left(\frac{\partial^2 l}{\partial \sigma_e^2 \partial \gamma^2}\right) = \begin{bmatrix} E\left(\frac{\partial^2 l}{\partial \sigma_e^2 \partial \gamma_1^2}\right) \\ \vdots \\ E\left(\frac{\partial^2 l}{\partial \sigma_e^2 \partial \gamma_k^2}\right) \end{bmatrix}$$

## Iteration Procedure

### Initial Values

- *Random Effect Variance Components:* For the  $i$ th ( $i = 1, \dots, k$ ) random effect, compute  $\hat{\beta}_i = (\mathbf{X}'_i \mathbf{X}_i)^{-1} \mathbf{X}'_i \mathbf{y}$ . Then assign the variance of the  $m_i$  elements of  $\hat{\beta}_i$  using divisor  $(m_i - 1)$  to the estimate  $\hat{\sigma}_i^2$  if  $m_i \geq 2$ , otherwise  $\hat{\sigma}_i^2 = 0$ .
- *Residual Variance:*  $\hat{\sigma}_e^2 = \mathbf{r}'\mathbf{r}/(n-r)$  where  $\mathbf{X} = [\mathbf{X}_0 | \mathbf{X}_1 | \dots | \mathbf{X}_k]$  and  $\mathbf{r} = \mathbf{y} - (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$ . If  $\hat{\sigma}_e^2 = 0$  but  $k \geq 1$  then reset  $\hat{\sigma}_e^2 = 10^{-8}$  so that the iteration can continue.

The variance ratios are then computed as  $\hat{\gamma}_i^2 = \hat{\sigma}_i^2 / \hat{\sigma}_e^2$ ,  $i = 1, \dots, k$ . Following the way the residual variance is initialized,  $\hat{\sigma}_e^2 > 0$  for  $k \geq 1$ .

### Updating

At the  $s$ th iteration ( $s = 0, 1, \dots$ ), the parameter vector is updated as

$$\hat{\theta}_{(s+1)} = \hat{\theta}_{(s)} + \rho \Delta \hat{\theta}_{(s)}$$

where  $\Delta \hat{\theta}_{(s)}$  is the value of increment  $\Delta \theta$  evaluated at  $\theta = \hat{\theta}_{(s)}$ , and  $\rho > 0$  is a step size such that  $l(\hat{\theta}_{(s+1)}) \geq l(\hat{\theta}_{(s)})$ . The increment vector depends on the choice of step type—Newton-Raphson versus Fisher scoring. The step size is determined by the step-halving technique with  $\rho = 1$  initially and a maximum of 10 halvings.

### Choice of Step

Following Jennrich and Sampson (1976), the first iteration is always the Fisher scoring step because it is more robust to poor initial values. For subsequent iterations, the Newton-Raphson step is used if:

1. The Hessian matrix is non-negative definite, and
2. The increment in the log-likelihood function of previous step is less than or equal to one.

Otherwise, the Fisher scoring step is used instead. The increment vector for each type of step is:

- *Newton-Raphson Step*:  $\Delta\theta = \left( -\frac{d^2l}{d\theta d\theta} \right)^{-1} \frac{dl}{d\theta}$ .
- *Fisher Scoring Step*:  $\Delta\theta = \left( -E \left( \frac{d^2l}{d\theta d\theta} \right) \right)^{-1} \frac{dl}{d\theta}$ .

### Convergence Criteria

Given the convergence criterion  $\varepsilon > 0$ , the iteration is considered converged when the following criteria are satisfied:

1.  $\left| l(\hat{\theta}_{(s+1)}) - l(\hat{\theta}_{(s)}) \right| < \varepsilon \times \max\left(1, \left| l(\hat{\theta}_{(s)}) \right| \right)$ , and
2.  $\left\langle \rho(\hat{\theta}_{(s+1)} - \hat{\theta}_{(s)}) \right\rangle < \varepsilon \times \max\left(1, \left\langle \hat{\theta}_{(s)} \right\rangle \right)$  where  $\langle \mathbf{a} \rangle$  is the sum of absolute values of elements of the vector  $\mathbf{a}$ .

### Negative Variance Estimates

Negative variance estimates can occur at the end of an iteration. An *ad hoc* method is to set those estimates to zero before the next iteration.

### Covariance Matrix

Let  $\hat{\theta}$  be the vector of maximum likelihood estimates. Their covariance matrix is given by

$$\text{cov}(\hat{\theta}) = \left( -E \left( \frac{d^2l}{d\theta d\theta} \right) \Big|_{\theta=\hat{\theta}} \right)^{-1}$$

## 20 Variance Components (VARCOMP)

Let

$$\Psi = \begin{bmatrix} \sigma_1^2 \\ \vdots \\ \sigma_k^2 \\ \sigma_e^2 \end{bmatrix}$$

be the original parameters. Their maximum likelihood estimates are given by

$$\hat{\Psi} = \begin{bmatrix} \hat{\sigma}_e^2 \hat{\gamma}_1^2 \\ \vdots \\ \hat{\sigma}_e^2 \hat{\gamma}_k^2 \\ \hat{\sigma}_e^2 \end{bmatrix}$$

and their covariance matrix is estimated by

$$\text{cov}(\hat{\Psi}) = \mathbf{J} \text{cov}(\hat{\theta}) \mathbf{J}'$$

where

$$\mathbf{J} = \begin{bmatrix} \sigma_e^2 \mathbf{I}_k & \gamma \\ 0 & 1 \end{bmatrix}$$

which is the  $(k+1) \times (k+1)$  Jacobian matrix of transforming  $\theta$  to  $\psi$ .

## ANOVA Estimate

The ANOVA variance component estimates are obtained by equating the expected mean squares of the random effects to their observed mean squares. The VARCOMP procedure offers two types of sum of squares: Type I and Type III (see Appendix 11 for details).

Let

$$\Psi = \begin{bmatrix} \sigma_1^2 \\ \vdots \\ \sigma_k^2 \\ \sigma_e^2 \end{bmatrix}$$

be the vector of variance components.

Let

$$\mathbf{q} = \begin{bmatrix} MS_1 \\ \vdots \\ MS_k \\ MSE \end{bmatrix}$$

where  $MS_i$ ,  $i = 1, \dots, k$  is the observed mean squares of the  $i$ th random effect, and  $MSE$  is the residual mean squares.

Let

$$\mathbf{S} = \begin{bmatrix} \mathbf{s}'_1 \\ \vdots \\ \mathbf{s}'_k \\ \mathbf{s}'_{k+1} \end{bmatrix}$$

be a  $(k+1) \times (k+1)$  matrix whose rows are coefficients for the expected mean squares. For example, the expected mean squares of the  $i$ th random effect is  $\mathbf{s}'_i \Psi$ . Algorithms for computing the expected mean squares can be found in the section "Univariate Mixed Model" in the chapter *GLM Univariate and Multivariate*. The ANOVA variance component estimates are then obtained by solving the system of linear equations:

$$\mathbf{S}\Psi = \mathbf{q}$$

## References

- Giesbrecht, F. G. 1983. An efficient procedure for computing MINQUE of variance components and generalized least squares estimates of fixed effects. *Communications in Statistics, Series A: Theory and Method*, 12: 2169–2177.
- Goodnight, J. H. 1978. *Computing MIVQUE0 Estimates of Variance Components*. SAS Technical Report R-105, Cary, N.C.: SAS Institute.
- Goodnight, J. H., and Hemmerle, W. J. 1979. A simplified algorithm for the  $W$  transformation in variance component estimation. *Technometrics*, 21: 265–267.
- Jennrich, R. I., and Sampson, P. F. 1976. Newton-Raphson and related algorithms for maximum likelihood variance component estimation. *Technometrics*, 18: 11–17.
- Rao, C. R. 1973. *Linear statistical inference and its applications*, 2nd ed. New York: John Wiley & Sons, Inc.
- Searle, S. R., Casella, G., and McCulloch, C. E. 1992. *Variance components*. New York: John Wiley & Sons, Inc.