

# Henderson Method 1

L. R. Schaeffer, April 27, 1999

Method 1 was derived by Dr. C. R. Henderson in 1949 as a result of his Ph.D. thesis work at Iowa State University. The method was not published until 1953 along with Methods 2 and 3 in Biometrics in one of Henderson's landmark papers. Method 1 applies the analysis of variance(AOV) procedure for balanced data to unbalanced data. Method 1 is only applicable to models that are completely random, i.e,  $\mathbf{Xb} = \mathbf{1}\mu$ . Method 1 is theoretically derived to be unbiased and translation invariant but only under a model of completely random mating and no selection. Method 1 is probably the easiest of all methods to calculate.

## Definition of Quadratic Forms

Three types of quadratic forms are defined for Henderson's Method 1. They are

$$\begin{aligned}\mathbf{Q}_1 &= \mathbf{I}, \text{ which gives} \\ \mathbf{y}'\mathbf{Q}_1\mathbf{y} &= \mathbf{y}'\mathbf{y} \\ &= \text{Total Sum of Squares,} \\ \mathbf{Q}_2 &= (1/N)\mathbf{J} = (1/N)\mathbf{1}\mathbf{1}' \text{ which gives} \\ \mathbf{y}'\mathbf{Q}_2\mathbf{y} &= (1/N)\mathbf{y}'\mathbf{1}\mathbf{1}'\mathbf{y}, \\ &= \text{Sum of Squares due to Mean,} \\ \mathbf{Q}_{i+2} &= \mathbf{Z}_i(\mathbf{Z}_i'\mathbf{Z}_i)^{-1}\mathbf{Z}_i'\end{aligned}$$

for  $i = 1$  to  $s$  is the number of random factors in the model. The quadratic form is

$$\mathbf{y}'\mathbf{Z}_i(\mathbf{Z}_i'\mathbf{Z}_i)^{-1}\mathbf{Z}_i'\mathbf{y}.$$

Note that  $\mathbf{Z}_i'\mathbf{Z}_i$  is a diagonal matrix whose elements are the number of observations in each level of the  $i^{th}$  random factor, and that elements of  $\mathbf{Z}_i'\mathbf{y}$  are the sums of observations within each level of the  $i^{th}$  random factor. Let  $n_{ik}$  represent the  $k^{th}$  diagonal element of  $\mathbf{Z}_i'\mathbf{Z}_i$ , then  $\mathbf{y}'\mathbf{Q}_{i+2}\mathbf{y} = \sum_{k=1}^{q_i} (y_{ik}^2/n_{ik})$  where  $q_i$  is the number of levels of the  $i^{th}$  random factor.

Each of the  $\mathbf{Q}$ -matrices in Method 1 has the following properties;

1.

$$\mathbf{Q}_m\mathbf{1} = \mathbf{1}$$

for  $m = 1$  to  $s + 2$ . That is, each  $\mathbf{Q}$ -matrix has elements which add to 1 within each row, and

2.

$$\mathbf{Q}_m\mathbf{Q}_m = \mathbf{Q}_m$$

for  $m = 1$  to  $s + 2$ . That is, each  $\mathbf{Q}$ -matrix is idempotent.

The first property is necessary for Method 1 to give translation invariant estimates of the variance components, and the second property allows the expectations of the quadratic forms to simplify to easy computing formulas.

### Expectations of Quadratic Forms

The expectations of the Method 1 quadratic forms are easy to derive. Recall that

$$\begin{aligned} \text{Var}(\mathbf{y}) &= \mathbf{V} = \sum_{i=0}^s \mathbf{V}_i \sigma_i^2, \text{ and} \\ E(\mathbf{y}'\mathbf{Q}\mathbf{y}) &= \text{tr}\mathbf{Q}\mathbf{V} + \mathbf{b}'\mathbf{X}'\mathbf{Q}\mathbf{X}\mathbf{b}. \end{aligned}$$

For Method 1 we have  $\mathbf{X}\mathbf{b} = \mathbf{1}\mu$  so that

$$\mathbf{b}'\mathbf{X}'\mathbf{Q}\mathbf{X}\mathbf{b} = \mu^2 \mathbf{1}'\mathbf{Q}\mathbf{1}$$

and because  $\mathbf{Q}\mathbf{1} = \mathbf{1}$  for all of the Method 1 quadratic forms, then

$$\mu^2 \mathbf{1}'\mathbf{Q}\mathbf{1} = \mu^2 \mathbf{1}'\mathbf{1} = N\mu^2.$$

Hence, all quadratic forms in Method 1 have  $N\mu^2$  as part of their expectations. Consequently, when differences are taken among these quadratic forms, then  $N\mu^2$  is eliminated, and the estimators are translation invariant.

Now concentrate on  $\text{tr}\mathbf{Q}\mathbf{V}$  for each of the three types of quadratic forms. For the Total Sum of Squares,

$$\begin{aligned} E(\mathbf{y}'\mathbf{y}) &= \text{tr}\mathbf{V} + N\mu^2 \\ \text{tr}\mathbf{V} &= \text{tr}\left(\sum_{i=1}^s \mathbf{Z}_i \mathbf{Z}_i' \sigma_i^2 + \mathbf{I}\sigma_0^2\right) \\ &= \sum_{i=1}^s \text{tr}\mathbf{Z}_i \mathbf{Z}_i' \sigma_i^2 + \text{tr}\mathbf{I}\sigma_0^2 \\ &= \sum_{i=1}^s \text{tr}\mathbf{Z}_i' \mathbf{Z}_i \sigma_i^2 + N\sigma_0^2 \\ &= \sum_{i=1}^s N\sigma_i^2 + N\sigma_0^2 = \sum_{i=0}^s N\sigma_i^2. \end{aligned}$$

Trace of  $\mathbf{Z}_i' \mathbf{Z}_i$  is the sum of the  $n_{ik}$  for the  $q_i$  levels of factor  $i$  with  $\sum_{k=1}^{q_i} n_{ik} = N$ . Hence, the expectation of  $\mathbf{y}'\mathbf{y}$  in Method 1 is always

$$E(\mathbf{y}'\mathbf{y}) = N\mu^2 + \sum_{i=0}^s N\sigma_i^2.$$

For the Sum of Squares due to the Mean,

$$\text{tr}\mathbf{Q}_2\mathbf{V} = \sum_{i=1}^s \text{tr}\mathbf{Q}_2\mathbf{Z}_i \mathbf{Z}_i' \sigma_i^2 + \text{tr}\mathbf{Q}_2\sigma_0^2$$

$$\begin{aligned}
&= \sum_{i=1}^s (1/N) \operatorname{tr} \mathbf{1} \mathbf{1}' \mathbf{Z}_i \mathbf{Z}_i' \sigma_i^2 + (1/N) \operatorname{tr} \mathbf{1} \mathbf{1}' \sigma_0^2 \\
&= \sum_{i=1}^s (1/N) (\mathbf{1}' \mathbf{Z}_i) (\mathbf{Z}_i' \mathbf{1}) \sigma_i^2 + \sigma_0^2 \\
&= \sum_{i=1}^s (1/N) \left( \sum_{k=1}^{q_i} n_{ik}^2 \right) \sigma_i^2 + \sigma_0^2
\end{aligned}$$

Thus, the expectation of the sum of squares due to the mean in Method 1 is always

$$E[(1/N) \mathbf{y}' \mathbf{1} \mathbf{1}' \mathbf{y}] = N \mu^2 + \sum_{i=1}^s (1/N) \left( \sum_{k=1}^{q_i} n_{ik}^2 \right) \sigma_i^2 + \sigma_0^2.$$

For the Sum of Squares for each random factor,

$$\begin{aligned}
\operatorname{tr} \mathbf{Q}_m \mathbf{V} &= \sum_{i=1}^s \operatorname{tr} \mathbf{Q}_m \mathbf{Z}_i \mathbf{Z}_i' \sigma_i^2 + \operatorname{tr} \mathbf{Q}_m \sigma_0^2 \\
&= \sum_{i=1}^s \operatorname{tr} [\mathbf{Z}_i' \mathbf{Z}_j (\mathbf{Z}_j' \mathbf{Z}_j)^{-1} \mathbf{Z}_j' \mathbf{Z}_i] \sigma_i^2 + \operatorname{tr} [\mathbf{Z}_j' \mathbf{Z}_j (\mathbf{Z}_j' \mathbf{Z}_j)^{-1}] \sigma_0^2 \\
&= \sum_{i=1}^s \left[ \sum_{k=1}^{q_j} \frac{\sum_{h=1}^{q_i} n_{kh}^2}{n_k} \right] \sigma_i^2 + \operatorname{tr} \mathbf{I}_{q_j} \sigma_0^2
\end{aligned}$$

where  $n_{kh}$  is the number of observations within the  $kh^{th}$  subclass between factors  $i$  and  $j$ . When  $i$  equals  $j$  in the summations, then  $n_{k1} = n_k$ , and the entire coefficient simplifies to  $N$ . The coefficient of  $\sigma_0^2$  simplifies to  $q_j$ .

### Obtaining Estimates

Let  $\mathbf{w}$  be the vector of quadratic forms, i.e.

$$\mathbf{w}' = [ \mathbf{y}' \mathbf{y} \quad (1/N) \mathbf{y}' \mathbf{1} \mathbf{1}' \mathbf{y} \quad \mathbf{y}' \mathbf{Z}_1 (\mathbf{Z}_1' \mathbf{Z}_1)^{-1} \mathbf{Z}_1' \mathbf{y} \quad \dots \quad \mathbf{y}' \mathbf{Z}_s (\mathbf{Z}_s' \mathbf{Z}_s)^{-1} \mathbf{Z}_s' \mathbf{y} ]$$

and let  $\boldsymbol{\sigma}' = [ \sigma_0^2 \quad \sigma_1^2 \quad \dots \quad \sigma_s^2 ]$ , then

$$E(\mathbf{w}) = \mathbf{F} \boldsymbol{\sigma} + \mathbf{1} N \mu^2$$

where a typical element of  $\mathbf{F}$  is  $\operatorname{tr} \mathbf{Q}_j \mathbf{V}_i$  for the  $j^{th}$  quadratic form and the  $i^{th}$  component.

Traditionally with balanced AOV tables, the sum of squares due to the mean was usually subtracted from the other sum of squares. Similarly, find any matrix  $\mathbf{K}'$  such that  $\mathbf{K}' \mathbf{1} = 0$  and  $\mathbf{K}'$  has full row rank, then

$$\begin{aligned}
E(\mathbf{K}' \mathbf{w}) &= \mathbf{K}' \mathbf{F} \boldsymbol{\sigma} + \mathbf{K}' \mathbf{1} N \mu^2 \\
&= \mathbf{K}' \mathbf{F} \boldsymbol{\sigma} \\
&\text{because } \mathbf{K}' \mathbf{1} = 0.
\end{aligned}$$

To estimate  $\sigma$  then equate  $\mathbf{K}'\mathbf{w}$  to its expectation so that

$$\begin{aligned}\mathbf{K}'\mathbf{F} \hat{\sigma} &= \mathbf{K}'\mathbf{w} \\ \hat{\sigma} &= (\mathbf{K}'\mathbf{F})^{-1}\mathbf{K}'\mathbf{w}.\end{aligned}$$

However, we could have obtained estimates without using  $\mathbf{K}'$  by

$$\begin{pmatrix} \hat{\sigma} \\ \hat{\mu}^2 \end{pmatrix} = \begin{bmatrix} \mathbf{F} & \mathbf{1}N \end{bmatrix}^{-1} \mathbf{w}.$$

### An Example

Consider the case of milk yields by daughters of dairy bulls in several different herds. Let  $s$ , the number of random factors in the model, be two (herds and sires), and thus, there are three components of variance to estimate. Data for analysis appears in the table below. Milk yields are given in percentage units relative to a breed mean for cows of a certain age and month of calving. The total sum of squares is 390,729.

Example dairy data for Method 1 analysis.

Herd	Sire	Daughter Milk Yields				Subclass totals	
1	1	157,	160,	138		455	(3)
	2	96,	110,	115,	120	441	(4)
	3	82,	65,			147	(2)
	4	120,	130,	110		360	(3)
2	1	140,	142,	145		427	(3)
	2	122,	117,	98		337	(3)
	3	70,	94,			164	(2)
3	2	112,	125,	105		342	(3)
	3	100,	92,			192	(2)
	4	116,	129,	133		378	(3)

	Totals	Number	Sum
By Sire	1	6	882
	2	10	1120
	3	6	503
	4	6	738
By Herd	1	12	1403
	2	8	928
	3	8	912
Overall		28	3243

### Quadratic Forms

$$\mathbf{y}'\mathbf{y} = 390,729.$$

$$(1/N)\mathbf{y}'\mathbf{1}\mathbf{1}'\mathbf{y} = (3243)^2/28 = 375,608.8929$$

For herds :

$$\mathbf{y}'\mathbf{Z}_1(\mathbf{Z}'_1\mathbf{Z}_1)^{-1}\mathbf{Z}'_1\mathbf{y} = (1403)^2/12 + (928)^2/8 + (912)^2/8$$

$$= 375,650.0834$$

For sires :

$$\mathbf{y}'\mathbf{Z}_2(\mathbf{Z}'_2\mathbf{Z}_2)^{-1}\mathbf{Z}'_2\mathbf{y} = (882)^2/6 + (1120)^2/10 + (503)^2/6$$

$$+ (738)^2/6$$

$$= 388,036.166667$$

### Expectations

$$E(\mathbf{y}'\mathbf{y}) = N \mu^2 + N \sigma_0^2 + N \sigma_1^2 + N \sigma_2^2$$

$$= 28 \mu^2 + 28 \sigma_0^2 + 28 \sigma_1^2 + 28 \sigma_2^2$$

$$E[(1/N)\mathbf{y}'\mathbf{1}\mathbf{1}'\mathbf{y}] = N \mu^2 + 1 \sigma_0^2 + k_1 \sigma_1^2 + k_2 \sigma_2^2$$

$$= 28 \mu^2 + 1 \sigma_0^2 + 9.714286 \sigma_1^2 + 7.428571 \sigma_2^2$$

where

$$k_1 = [ (12)^2 + (8)^2 + (8)^2 ]/28$$

$$\begin{aligned}
&= 9.714286 \\
k_2 &= [ (6)^2 + (10)^2 + (6)^2 + (6)^2 ]/28 \\
&= 7.428571
\end{aligned}$$

For Herds :

$$\begin{aligned}
E[\mathbf{y}'\mathbf{Z}_1(\mathbf{Z}'_1\mathbf{Z}_1)^{-1}\mathbf{Z}'_1\mathbf{y}] &= N \mu^2 + q_1\sigma_0^2 + N \sigma_1^2 + k_3\sigma_2^2 \\
&= 28 \mu^2 + 3 \sigma_0^2 + 28 \sigma_1^2 + 8.666667 \sigma_2^2
\end{aligned}$$

where

$$\begin{aligned}
k_3 &= [ 3^2 + 4^2 + 2^2 + 3^2 ]/12 + [ 3^2 + 3^2 + 2^2 ]/8 \\
&\quad + [ (3^2 + 2^2 + 3^2) ]/8 \\
&= 38/12 + 22/8 + 22/8 \\
&= 8.666667
\end{aligned}$$

For sires :

$$\begin{aligned}
E[\mathbf{y}'\mathbf{Z}_2(\mathbf{Z}'_2\mathbf{Z}_2)^{-1}\mathbf{Z}'_2\mathbf{y}] &= N \mu^2 + q_2\sigma_0^2 + k_4\sigma_1^2 + N \sigma_2^2 \\
&= 28 \mu^2 + 4 \sigma_0^2 + 11.4 \sigma_1^2 + 28 \sigma_2^2
\end{aligned}$$

where

$$\begin{aligned}
k_4 &= [ 3^2 + 3^2 ]/6 + [ 4^2 + 3^2 + 3^2 ]/10 \\
&\quad + [ 2^2 + 2^2 + 2^2 ]/6 + [ 3^2 + 3^2 ]/6 \\
&= 18/6 + 34/10 + 12/6 + 18/6 \\
&= 11.4.
\end{aligned}$$

Solving for Estimates

$$\begin{pmatrix} 28 & 28 & 28 & 28 \\ 28 & 1 & 9.714286 & 7.428571 \\ 28 & 3 & 28 & 8.666667 \\ 28 & 4 & 11.4 & 28 \end{pmatrix} \begin{pmatrix} \hat{\mu}^2 \\ \hat{\sigma}_0^2 \\ \hat{\sigma}_1^2 \\ \hat{\sigma}_2^2 \end{pmatrix} = \begin{pmatrix} 390,729. \\ 375,608.8929 \\ 375,650.0834 \\ 388,036.1667 \end{pmatrix}$$

Premultiply both sides of the above by  $\mathbf{K}'$  where

$$\mathbf{K}' = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{pmatrix}$$

Ignoring the column for  $\mu^2$  which becomes all zeros, then we have

$$\begin{pmatrix} 27 & 18.285714 & 20.571429 \\ 25 & 0 & 19.333333 \\ 24 & 16.6 & 0 \end{pmatrix} \begin{pmatrix} \hat{\sigma}_0^2 \\ \hat{\sigma}_1^2 \\ \hat{\sigma}_2^2 \end{pmatrix} = \begin{pmatrix} 15,120.1071 \\ 15,078.9166 \\ 2,692.8334 \end{pmatrix}$$

The resulting estimates are

$$\begin{aligned}\hat{\sigma}_0^2 &= 149.4246, \\ \hat{\sigma}_1^2 &= -53.8167, \\ \hat{\sigma}_2^2 &= 586.7225.\end{aligned}$$

### Sampling Variances of Estimates

Let TSS, MSS, HSS, and SSS represent the quadratic forms for the total sum of squares, mean, herds, and sires, respectively, then the coefficients of the unknown parameters for computing the variances and covariances of the corresponding quadratic forms are in the following table.

	$\sigma_0^4$	$\sigma_0^2\sigma_1^2$	$\sigma_0^2\sigma_2^2$	$\sigma_1^4$	$\sigma_1^2\sigma_2^2$	$\sigma_2^4$
TSS	56.	112.	112.	544.	328.	416.
TSS,MSS	2.	38.8571	29.7143	196.5714	288.	117.7143
TSS,HSS	6.	112.	34.6667	544.	328.	132.6667
TSS,SSS	8.	45.6	112.	224.5333	328.	416.
MSS	2.	38.8571	29.7143	188.7347	288.6531	110.3673
MSS,HSS	2.	38.8571	29.7143	196.5714	288.	110.3810
MSS,SSS	2.	38.8571	29.7143	188.8762	288.	117.7143
HSS	6.	112.	34.6667	544.	328.	112.9514
HSS,SSS	2.4111	45.6	34.6667	224.5333	328.	132.6667
SSS	8.	45.6	112.	193.5867	328.	416.

Each line above would also have a function,  $4 \mathbf{b}'\mathbf{X}'\mathbf{Q}\mathbf{V}\mathbf{Q}\mathbf{X}\mathbf{b}$ , associated with it which will be identical in each term (variance or covariance). These terms have been ignored since they will disappear when differences between pairs of quadratic forms are taken.

Now assume the true components of variance have the following values;  $\sigma_0^2 = 5$ ,  $\sigma_1^2 = .1$ , and  $\sigma_2^2 = 1$ . Post-multiply the table of coefficients above by the vector (25 .5 5 .01 .1 1) which correspond to the columns of the table above, and store the results in matrix form as

$$Var(\mathbf{w}) = \begin{pmatrix} 2470.2400 & 366.4800 & 550.2400 & 1233.8453 \\ 366.4800 & 359.1200 & 359.1467 & 366.4030 \\ 550.2400 & 359.1467 & 530.5247 & 424.1231 \\ 1233.8453 & 366.4030 & 424.1231 & 1233.5359 \end{pmatrix}.$$

Using the same  $\mathbf{K}'$  as before for taking differences in quadratic forms, then

$$Var(\mathbf{K}'\mathbf{w}) = \begin{pmatrix} 2096.4000 & 1912.6667 & 1236.3177 \\ 1912.6667 & 1900.2847 & 1110.2778 \\ 1236.3177 & 1110.2778 & 1236.0852 \end{pmatrix},$$

and if we let

$$\mathbf{P} = (\mathbf{K}'\mathbf{F})^{-1} = \begin{pmatrix} 27 & 18.285714 & 20.571429 \\ 25 & 0. & 19.333333 \\ 24 & 16.6 & 0. \end{pmatrix}^{-1}$$

then  $\hat{\sigma} = \mathbf{PK}'\mathbf{w}$  and

$$Var(\mathbf{PK}'\mathbf{w}) = \begin{pmatrix} 2.2956 & -.2959 & -.3218 \\ -.2959 & .5428 & .0157 \\ -.3218 & .0157 & 2.0778 \end{pmatrix}.$$

The variance of  $\hat{\sigma}_0^2$  is then 2.2956 provided the true variances are  $\sigma_0^2 = 5$ ,  $\sigma_1^2 = .1$ , and  $\sigma_2^2 = 1$ . The covariance between  $\hat{\sigma}_0^2$  and  $\hat{\sigma}_2^2$  is -.3218, etc. For this small example, the variances of the estimates are quite large. Another set of true values could be used to derive  $Var(\hat{\sigma})$  if desired. Few researchers ever compute the sampling variances of their estimates due to the complexity and number of the traces that must be computed.

### Least Squares Equations

The quadratic forms(except for TSS) and their expectations can also be obtained from Ordinary Least Squares equations. For the example data, the OLS equations are

$$\begin{pmatrix} 28 & 6 & 10 & 6 & 6 & 12 & 8 & 8 \\ 6 & 6 & 0 & 0 & 0 & 3 & 3 & 0 \\ 10 & 0 & 10 & 0 & 0 & 4 & 3 & 3 \\ 6 & 0 & 0 & 6 & 0 & 2 & 2 & 2 \\ 6 & 0 & 0 & 0 & 6 & 3 & 0 & 3 \\ 12 & 3 & 4 & 2 & 3 & 12 & 0 & 0 \\ 8 & 3 & 3 & 2 & 0 & 0 & 8 & 0 \\ 8 & 0 & 3 & 2 & 3 & 0 & 0 & 8 \end{pmatrix} \begin{pmatrix} \hat{\mu} \\ \hat{s}_1 \\ \hat{s}_2 \\ \hat{s}_3 \\ \hat{s}_4 \\ \hat{h}_1 \\ \hat{h}_2 \\ \hat{h}_3 \end{pmatrix} = \begin{pmatrix} 3243 \\ 882 \\ 1120 \\ 503 \\ 738 \\ 1403 \\ 928 \\ 912 \end{pmatrix}.$$

The first step is to calculate the variance-covariance matrix of the right hand sides. Note that, for herds,

$$\mathbf{W}'\mathbf{Z}_1 = \begin{pmatrix} 12 & 8 & 8 \\ 3 & 3 & 0 \\ 4 & 3 & 3 \\ 2 & 2 & 2 \\ 3 & 0 & 3 \\ 12 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 8 \end{pmatrix},$$

so that

$$\mathbf{U}_1 = \mathbf{W}'\mathbf{Z}_1\mathbf{Z}_1'\mathbf{W}$$

$$= \begin{pmatrix} 272 & 60 & 96 & 56 & 60 & 144 & 64 & 64 \\ 60 & 18 & 21 & 12 & 9 & 36 & 24 & 0 \\ 96 & 21 & 34 & 20 & 21 & 48 & 24 & 24 \\ 56 & 12 & 20 & 12 & 12 & 24 & 16 & 16 \\ 60 & 9 & 21 & 12 & 18 & 36 & 0 & 24 \\ 144 & 36 & 48 & 24 & 36 & 144 & 0 & 0 \\ 64 & 24 & 24 & 16 & 0 & 0 & 64 & 0 \\ 64 & 0 & 24 & 16 & 24 & 0 & 0 & 64 \end{pmatrix},$$

and similarly for sires,

$$\mathbf{W}'\mathbf{Z}_2 = \begin{pmatrix} 6 & 10 & 6 & 6 \\ 6 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 \\ 0 & 0 & 6 & 0 \\ 0 & 0 & 0 & 6 \\ 3 & 4 & 2 & 3 \\ 3 & 3 & 2 & 0 \\ 0 & 3 & 2 & 3 \end{pmatrix},$$

so that

$$\begin{aligned} \mathbf{U}_2 &= \mathbf{W}'\mathbf{Z}_2\mathbf{Z}_2'\mathbf{W} \\ &= \begin{pmatrix} 208 & 36 & 100 & 36 & 36 & 88 & 60 & 60 \\ 36 & 36 & 0 & 0 & 0 & 18 & 18 & 0 \\ 100 & 0 & 100 & 0 & 0 & 40 & 30 & 30 \\ 36 & 0 & 0 & 36 & 0 & 12 & 12 & 12 \\ 36 & 0 & 0 & 0 & 36 & 18 & 0 & 18 \\ 88 & 18 & 30 & 12 & 18 & 38 & 25 & 25 \\ 60 & 18 & 30 & 12 & 0 & 25 & 22 & 13 \\ 60 & 0 & 30 & 12 & 18 & 25 & 13 & 22 \end{pmatrix}. \end{aligned}$$

Then

$$Var(\mathbf{W}'\mathbf{y}) = \mathbf{U}_1\sigma_1^2 + \mathbf{U}_2\sigma_2^2 + \mathbf{W}'\mathbf{W}\sigma_0^2.$$

The  $\mathbf{Q}$ -matrices for  $\mathbf{W}'\mathbf{y}$  are diagonal matrices. For the mean sum of squares, MSS,

$$\mathbf{Q}_2 = \text{diag} \left( \frac{1}{28} \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \right),$$

and for the herd and sire sum of squares,

$$\mathbf{Q}_3 = \text{diag} \left( 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ \frac{1}{12} \ \frac{1}{8} \ \frac{1}{8} \right),$$

and

$$\mathbf{Q}_4 = \text{diag} \left( 0 \ \frac{1}{6} \ \frac{1}{10} \ \frac{1}{6} \ \frac{1}{6} \ 0 \ 0 \ 0 \right).$$

The quadratic forms are

$$\begin{aligned} \mathbf{y}'\mathbf{W}\mathbf{Q}_2\mathbf{W}'\mathbf{y} &= 375,608.8929 \\ \mathbf{y}'\mathbf{W}\mathbf{Q}_3\mathbf{W}'\mathbf{y} &= 375,650.0834 \\ \mathbf{y}'\mathbf{W}\mathbf{Q}_4\mathbf{W}'\mathbf{y} &= 388,036.1667 \end{aligned}$$

The expectations of these quadratic forms are given by

$$\begin{pmatrix} \text{tr}\mathbf{Q}_2\mathbf{U}_1 & \text{tr}\mathbf{Q}_2\mathbf{U}_2 & \text{tr}\mathbf{Q}_2\mathbf{W}'\mathbf{W} \\ \text{tr}\mathbf{Q}_3\mathbf{U}_1 & \text{tr}\mathbf{Q}_3\mathbf{U}_2 & \text{tr}\mathbf{Q}_3\mathbf{W}'\mathbf{W} \\ \text{tr}\mathbf{Q}_4\mathbf{U}_1 & \text{tr}\mathbf{Q}_4\mathbf{U}_2 & \text{tr}\mathbf{Q}_4\mathbf{W}'\mathbf{W} \end{pmatrix} \begin{pmatrix} \sigma_1^2 \\ \sigma_2^2 \\ \sigma_0^2 \end{pmatrix} + N\mu^2\mathbf{1}.$$

Then you need to add the total sum of squares and its expectation and proceed as before.

### Problem

Apply Method 1 to the following data using a model with three random factors. Compute the sampling variances of the estimates.

$$\mathbf{y} = \begin{pmatrix} 32 \\ 44 \\ 116 \\ 94 \\ 80 \\ 52 \\ 66 \\ 110 \\ 86 \\ 76 \\ 35 \\ 41 \\ 121 \\ 92 \\ 82 \\ 49 \\ 69 \\ 107 \\ 90 \\ 73 \\ 30 \end{pmatrix}, \quad \mathbf{Z}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$$\mathbf{Z}_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{Z}_3 = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

# Henderson Method 2

L. R. Schaeffer, April 27, 1999

The main drawback of Method 1 was that only completely random models could be handled. Methods 2 and 3 accommodate mixed models. Method 2 is an unbiased, translation invariant procedure, and can be used with mixed models provided that interactions between fixed and random factors, or nesting of random factors within fixed factors, do not exist in the model.

Method 2 requires the least squares (LS) equations for all factors in the model. A solution vector is obtained by taking a generalized inverse of the coefficient matrix of the LS equations. Because there are an infinite number of possible solution vectors to the LS equations, Searle(1968) maintained for many years that Method 2 was not uniquely defined. That is, different solution vectors could lead to different estimates of the variance components. This was later proven to be incorrect, and Method 2 was proven to be unique if Henderson's directions are followed exactly (Henderson, Searle, and Schaeffer 1974).

## The Calculations

Henderson's Method 2 involves the following steps:

**Step 1.** Construct LS equations for all factors (fixed and random) in the model. Let  $\mathbf{W} = ( \mathbf{X} \ \mathbf{Z} )$  and  $\boldsymbol{\beta}' = ( \mathbf{b}' \ \mathbf{u}' )$ , then solve

$$\mathbf{W}'\mathbf{W} \hat{\boldsymbol{\beta}} = \mathbf{W}'\mathbf{y} \text{ or } \hat{\boldsymbol{\beta}} = (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{y}.$$

**Step 2.** Adjust  $\mathbf{y}$  for  $\hat{\mathbf{b}}$  from the OLS equations in step 1. Let

$$\mathbf{y}^0 = \mathbf{y} - \mathbf{X}\hat{\mathbf{b}}.$$

**Step 3.** Apply Method 1 to  $\mathbf{y}^0$  and adjust the coefficients of  $\hat{\sigma}_0^2$  in the expectations of all quadratic forms to allow for the change in  $V(\mathbf{y}^0)$  from  $V(\mathbf{y})$ .

## Theoretical Development

Let

$$\mathbf{W} = ( \mathbf{X} \ \mathbf{Z} ) = ( \mathbf{X}_b \ \mathbf{X}_a \ \mathbf{Z}_a \ \mathbf{Z}_b )$$

such that the following rank requirements are met,

$$\begin{aligned} r( \mathbf{X}_a \ \mathbf{Z}_a ) &= r( \mathbf{X} \ \mathbf{Z} ), \text{ and} \\ r( \mathbf{Z}_a ) &= r( \mathbf{Z} ). \end{aligned}$$

Thus,  $r(\mathbf{X} \ \mathbf{Z}) = r(\mathbf{X}) + r(\mathbf{Z}) - 1$ . This implies that the fewest possible rows and columns of  $\mathbf{Z}$  are set to zero in obtaining a generalized inverse of the OLS coefficient matrix. Then

$$\mathbf{W}'\mathbf{W} = \begin{pmatrix} \mathbf{X}'_b\mathbf{X}_b & \mathbf{X}'_b\mathbf{X}_a & \mathbf{X}'_b\mathbf{Z}_a & \mathbf{X}'_b\mathbf{Z}_b \\ \mathbf{X}'_a\mathbf{X}_b & \mathbf{X}'_a\mathbf{X}_a & \mathbf{X}'_a\mathbf{Z}_a & \mathbf{X}'_a\mathbf{Z}_b \\ \mathbf{Z}'_a\mathbf{X}_b & \mathbf{Z}'_a\mathbf{X}_a & \mathbf{Z}'_a\mathbf{Z}_a & \mathbf{Z}'_a\mathbf{Z}_b \\ \mathbf{Z}'_b\mathbf{X}_b & \mathbf{Z}'_b\mathbf{X}_a & \mathbf{Z}'_b\mathbf{Z}_a & \mathbf{Z}'_b\mathbf{Z}_b \end{pmatrix}.$$

A generalized inverse is obtained by setting the rows and columns corresponding to  $\mathbf{X}_b$  and  $\mathbf{Z}_b$  to zero and inverting the remaining elements. Let

$$(\mathbf{W}'\mathbf{W})^- = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{P} & \mathbf{S} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}' & \mathbf{M} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix},$$

then

$$\begin{pmatrix} \mathbf{P} & \mathbf{S} \\ \mathbf{S}' & \mathbf{M} \end{pmatrix} \begin{pmatrix} \mathbf{X}'_a\mathbf{X}_a & \mathbf{X}'_a\mathbf{Z}_a \\ \mathbf{Z}'_a\mathbf{X}_a & \mathbf{Z}'_a\mathbf{Z}_a \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}.$$

The solutions are

$$\begin{pmatrix} \hat{\mathbf{b}}_b \\ \hat{\mathbf{b}}_a \\ \hat{\mathbf{u}}_a \\ \hat{\mathbf{u}}_b \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{P}\mathbf{X}'_a\mathbf{y} + \mathbf{S}\mathbf{Z}'_a\mathbf{y} \\ \mathbf{S}'\mathbf{X}'_a\mathbf{y} + \mathbf{M}\mathbf{Z}'_a\mathbf{y} \\ \mathbf{0} \end{pmatrix}.$$

An unbiased estimate of the residual variance must be obtained, such as

$$\hat{\sigma}_0^2 = (\mathbf{y}'\mathbf{y} - \hat{\mathbf{b}}'_a\mathbf{X}'_a\mathbf{y} - \hat{\mathbf{u}}'_a\mathbf{Z}'_a\mathbf{y}) / (N - r(\mathbf{W})).$$

Each observation must now be adjusted for  $\hat{\mathbf{b}}$ , as

$$\begin{aligned} \mathbf{y}^0 &= \mathbf{y} - \mathbf{X}\hat{\mathbf{b}} \\ &= \mathbf{y} - \mathbf{X}_b\hat{\mathbf{b}}_b - \mathbf{X}_a\hat{\mathbf{b}}_a \\ &= \mathbf{y} - \mathbf{X}_a\hat{\mathbf{b}}_a. \end{aligned}$$

Each observation does not need to be adjusted separately. Note that

$$\mathbf{Z}'_j\mathbf{y}^0 = \mathbf{Z}'_j\mathbf{y} - \mathbf{Z}'_j\mathbf{X}_a\hat{\mathbf{b}}_a,$$

for the  $j^{\text{th}}$  factor in the model, and that  $\mathbf{Z}'_j\mathbf{y}$  and  $\mathbf{Z}'_j\mathbf{X}_a$  are available from the LS equations. Hence, only the right hand sides of the LS equations need be adjusted.

To apply Method 1 to  $\mathbf{y}^0$  requires that the model for  $\mathbf{y}^0$  be completely random except for a common mean effect. The proof of this is given by Henderson et al.(1974). The model equation for  $\mathbf{y}^0$  is

$$\mathbf{y}^0 = \mathbf{y} - \mathbf{X}_a\hat{\mathbf{b}}_a = \mathbf{T}\mathbf{y}$$

where

$$\mathbf{T} = (\mathbf{I} - \mathbf{X}_a \mathbf{P} \mathbf{X}_a' - \mathbf{X}_a \mathbf{S} \mathbf{Z}_a').$$

Fortunately,

$$\begin{aligned} \mathbf{T} \mathbf{Z} &= (\mathbf{I} - \mathbf{X}_a \mathbf{P} \mathbf{X}_a' - \mathbf{X}_a \mathbf{S} \mathbf{Z}_a') (\mathbf{Z}_a \ \mathbf{Z}_b) \\ &= (\mathbf{Z}_a \ \mathbf{Z}_b) - \mathbf{X}_a (\mathbf{P} \mathbf{X}_a' - \mathbf{S} \mathbf{Z}_a') (\mathbf{Z}_a \ \mathbf{Z}_b) \\ &= \mathbf{Z} - \mathbf{X}_a \mathbf{0} \\ &= \mathbf{Z}, \\ \mathbf{T} \mathbf{X}_a &= \mathbf{0}, \quad \text{and} \\ \mathbf{T} \mathbf{X}_b &= c \mathbf{1} \end{aligned}$$

So that the model for  $\mathbf{y}^0$  is

$$\begin{aligned} \mathbf{y}^0 &= \mathbf{T} \mathbf{y} = \mathbf{T} \mathbf{X} \mathbf{b} + \mathbf{T} \mathbf{Z} \mathbf{u} + \mathbf{T} \mathbf{e} \\ &= c \mathbf{1} + \mathbf{Z} \mathbf{u} + \mathbf{T} \mathbf{e}. \end{aligned}$$

Proof that  $\mathbf{T} \mathbf{X} \mathbf{b} = c \mathbf{1}$  is also given in Henderson et al.(1974), and this is the key to showing that Method 2 gives unique estimates of variance components regardless of the generalized inverse used to solve the OLS equations.

Consequently,

$$\begin{aligned} \text{Var}(\mathbf{y}^0) &= \text{Var}(\mathbf{Z} \mathbf{u} + \mathbf{T} \mathbf{e}) \\ &= \sum_{i=1}^s \mathbf{z}_i \mathbf{z}_i' \sigma_i^2 + \mathbf{T} \mathbf{T}' \sigma_0^2. \end{aligned}$$

The variance-covariance matrix of  $\mathbf{y}^0$  is slightly different from that of the original observation vector  $\mathbf{y}$ . This difference causes changes to the expectations of the Method 1 quadratic forms where  $\sigma_0^2$  is involved.

The coefficient of  $\sigma_0^2$  in the expectation of  $(1/N) \mathbf{y}^{0'} \mathbf{1} \mathbf{1}' \mathbf{y}^0$  is

$$\begin{aligned} (1/N) \text{tr}(\mathbf{1} \mathbf{1}' \mathbf{T} \mathbf{T}') &= (1/N) \text{tr}(\mathbf{1}' \mathbf{T} \mathbf{T}' \mathbf{1}) \\ &= (1/N) \text{tr}(\mathbf{1}' \mathbf{1}) + (1/N) \text{tr}(\mathbf{1}' \mathbf{X}_a \mathbf{P} \mathbf{X}_a' \mathbf{1}) \\ &= 1 + (1/N) \text{tr}[\mathbf{P} (\mathbf{X}_a' \mathbf{1}) (\mathbf{1}' \mathbf{X}_a)] \end{aligned}$$

Thus, we have the usual coefficient plus an additional term. Similarly, the coefficient of  $\sigma_0^2$  in the expectation of  $\mathbf{y}^{0'} \mathbf{Z}_j (\mathbf{Z}_j' \mathbf{Z}_j)^{-1} \mathbf{Z}_j' \mathbf{y}^0$  where  $\mathbf{Z}_j$  is the design matrix of the  $j^{\text{th}}$  random factor of the model, is

$$\begin{aligned} \text{tr}[\mathbf{Z}_j (\mathbf{Z}_j' \mathbf{Z}_j)^{-1} \mathbf{Z}_j' \mathbf{T} \mathbf{T}'] &= \text{tr}[(\mathbf{Z}_j' \mathbf{Z}_j)^{-1} \mathbf{Z}_j' \mathbf{T} \mathbf{T}' \mathbf{Z}_j] \\ &= \text{tr}[(\mathbf{Z}_j' \mathbf{Z}_j)^{-1} (\mathbf{Z}_j' \mathbf{Z}_j + \mathbf{Z}_j' \mathbf{X}_a \mathbf{P} \mathbf{X}_a' \mathbf{Z}_j)] \\ &= \text{tr}[(\mathbf{Z}_j' \mathbf{Z}_j)^{-1} \mathbf{Z}_j' \mathbf{Z}_j] + \text{tr}[\mathbf{P} \mathbf{X}_a' \mathbf{Z}_j (\mathbf{Z}_j' \mathbf{Z}_j)^{-1} \mathbf{Z}_j' \mathbf{X}_a] \\ &= q_j + \text{tr}[\mathbf{P} \mathbf{X}_a' \mathbf{Z}_j (\mathbf{Z}_j' \mathbf{Z}_j)^{-1} \mathbf{Z}_j' \mathbf{X}_a] \end{aligned}$$

The coefficients of the other variance components in the expectations are the same as in the usual Method 1.

### An Example

Below are the LS equations for a two-way classification model with one fixed factor and one random factor. Let factor A be the fixed factor and let factor B be random. The total sum of squares of the observations was 1,979.

$$\left( \begin{array}{ccc|ccc} 16 & 4 & 6 & 6 & 3 & 4 & 5 & 4 \\ \hline 4 & 4 & 0 & 0 & 1 & 1 & 2 & 0 \\ 6 & 0 & 6 & 0 & 1 & 1 & 1 & 3 \\ 6 & 0 & 0 & 6 & 1 & 2 & 2 & 1 \\ \hline 3 & 1 & 1 & 1 & 3 & 0 & 0 & 0 \\ 4 & 1 & 1 & 2 & 0 & 4 & 0 & 0 \\ 5 & 2 & 1 & 2 & 0 & 0 & 5 & 0 \\ 4 & 0 & 3 & 1 & 0 & 0 & 0 & 4 \end{array} \right) \begin{pmatrix} \hat{\mu} \\ \hat{A}_1 \\ \hat{A}_2 \\ \hat{A}_3 \\ \hat{B}_1 \\ \hat{B}_2 \\ \hat{B}_3 \\ \hat{B}_4 \end{pmatrix} = \begin{pmatrix} 173 \\ 49 \\ 66 \\ 58 \\ 32 \\ 51 \\ 51 \\ 39 \end{pmatrix}$$

Using the partitioning given in the previous section, then let

$$\begin{aligned} \hat{\mathbf{b}}'_b &= [\hat{\mu} \quad \hat{A}_1], \\ \hat{\mathbf{b}}'_a &= [\hat{A}_2 \quad \hat{A}_3], \\ \hat{\mathbf{u}}'_b &= \text{does not exist in this example,} \\ \hat{\mathbf{u}}'_a &= [\hat{B}_1 \quad \hat{B}_2 \quad \hat{B}_3 \quad \hat{B}_4] \end{aligned}$$

and similarly for the design matrices  $\mathbf{X}$  and  $\mathbf{Z}$ . The solutions are

$$\begin{pmatrix} \hat{\mu} \\ \hat{A}_1 \\ \hat{A}_2 \\ \hat{A}_3 \end{pmatrix} = \begin{pmatrix} 0. \\ 0. \\ -0.7597 \\ -2.7161 \end{pmatrix}, \quad \text{and} \quad \begin{pmatrix} \hat{B}_1 \\ \hat{B}_2 \\ \hat{B}_3 \\ \hat{B}_4 \end{pmatrix} = \begin{pmatrix} 11.8253 \\ 14.2980 \\ 11.4384 \\ 10.9988 \end{pmatrix}$$

The solutions times their corresponding right hand side elements gives

$$\hat{\beta}'\mathbf{W}'\mathbf{y} = 1,912.2422,$$

and an estimate of the residual variance component is

$$\begin{aligned} \hat{\sigma}_0^2 &= (\mathbf{y}'\mathbf{y} - \hat{\beta}'\mathbf{W}'\mathbf{y}) / (N - r(\mathbf{W})) \\ &= (1,979 - 1,912.2422) / (16 - 6) \\ &= 6.675782. \end{aligned}$$

This estimate of  $\hat{\sigma}_0^2$  is the same for any generalized inverse of  $\mathbf{W}'\mathbf{W}$ .

Now adjust  $\mathbf{W}'\mathbf{y}$  for the estimates of the fixed effects to give

$$\begin{aligned}\mathbf{W}'\mathbf{y}^0 &= \mathbf{W}'\mathbf{y} - \mathbf{W}'\mathbf{X}\hat{\mathbf{b}} \\ &= \begin{pmatrix} 173 \\ 49 \\ 66 \\ 58 \\ 32 \\ 51 \\ 51 \\ 39 \end{pmatrix} - \begin{pmatrix} 16 & 4 & 6 & 6 \\ 4 & 4 & 0 & 0 \\ 6 & 0 & 6 & 0 \\ 6 & 0 & 0 & 6 \\ 3 & 1 & 1 & 1 \\ 4 & 1 & 1 & 2 \\ 5 & 2 & 1 & 2 \\ 4 & 0 & 3 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ -0.7597 \\ -2.7161 \end{pmatrix} \\ &= \begin{pmatrix} 193.8547 \\ 49.0000 \\ 70.5579 \\ 74.2968 \\ 35.4758 \\ 57.1919 \\ 57.1919 \\ 43.9951 \end{pmatrix}\end{aligned}$$

The necessary quadratic forms are

$$\begin{aligned}\hat{\sigma}_0^2 &= 6.675782, \\ (1/N)\mathbf{y}^{0'}\mathbf{1}\mathbf{1}'\mathbf{y}^0 &= (1/16)(193.8547)^2 = 2,348.7276, \\ \mathbf{y}^{0'}\mathbf{Z}_B(\mathbf{Z}'_B\mathbf{Z}_B)^{-1}\mathbf{Z}'_B\mathbf{y}^0 &= (35.4758)^2/3 + (57.1919)^2/4 \\ &\quad + (57.1919)^2/5 + (43.9951)^2/4 = 2,375.3139,\end{aligned}$$

where  $\mathbf{Z}_B$  is the design matrix for factor B. The expectations of the next two quadratic forms are

$$\begin{aligned}E[(1/N)\mathbf{y}^{0'}\mathbf{1}\mathbf{1}'\mathbf{y}^0] &= 16c^2 + 4.125\sigma_B^2 + (1+k_1)\sigma_0^2 \\ E[\mathbf{y}^{0'}\mathbf{Z}_B(\mathbf{Z}'_B\mathbf{Z}_B)^{-1}\mathbf{Z}'_B\mathbf{y}^0] &= 16c^2 + 16\sigma_B^2 + (4+k_2)\sigma_0^2\end{aligned}$$

where

$$\begin{aligned}k_1 &= (1/N)\text{tr}(\mathbf{P}\mathbf{X}'_a\mathbf{1}\mathbf{1}'\mathbf{X}_a), \text{ and} \\ k_2 &= \text{tr}(\mathbf{P}\mathbf{X}'_a\mathbf{Z}_B(\mathbf{Z}'_B\mathbf{Z}_B)^{-1}\mathbf{Z}'_B\mathbf{X}_a)\end{aligned}$$

In this example,

$$\mathbf{P} = \begin{pmatrix} .532189 & .291845 \\ .291845 & .436542 \end{pmatrix},$$

and  $\mathbf{X}'_a\mathbf{1} = (6 \ 6)$ . Therefore,

$$k_1 = \frac{1}{16} \text{tr} \left( \mathbf{P} \begin{pmatrix} 6 \\ 6 \end{pmatrix} (6 \ 6) \right) = 3.49295, \quad \text{and}$$

$$k_2 = \text{tr} \left( \mathbf{P} \begin{pmatrix} 1 & 1 & 1 & 3 \\ 1 & 2 & 2 & 1 \end{pmatrix} \begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 2 \\ 3 & 1 \end{pmatrix} \right)$$

$$= 3.81239.$$

The resulting equations to solve are

$$\begin{pmatrix} 0 & 0.0000 & 1.0000 \\ 16 & 4.1250 & 4.4929 \\ 16 & 16.0000 & 7.8124 \end{pmatrix} \begin{pmatrix} \hat{\sigma}^2 \\ \hat{\sigma}_B^2 \\ \hat{\sigma}_0^2 \end{pmatrix} = \begin{pmatrix} 6.6758 \\ 2,348.7276 \\ 2,375.3139 \end{pmatrix}$$

Let

$$\mathbf{K}' = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 1 \end{pmatrix},$$

then  $\mathbf{K}'\mathbf{F}\hat{\boldsymbol{\sigma}} = \mathbf{K}'\mathbf{w}$ , which gives

$$\begin{pmatrix} 0.0000 & 1.0000 \\ 11.8750 & 3.3194 \end{pmatrix} \begin{pmatrix} \hat{\sigma}_1^2 \\ \hat{\sigma}_0^2 \end{pmatrix} = \begin{pmatrix} 6.6758 \\ 26.5863 \end{pmatrix},$$

giving estimates of  $\hat{\sigma}_0^2 = 6.6758$ , and  $\hat{\sigma}_B^2 = .37275$ .

Suppose that  $\hat{\mu}$  and  $\hat{A}_3$  were set to zero rather than  $\hat{\mu}$  and  $\hat{A}_1$  in obtaining solutions to the OLS equations. The reader should verify the following results.

$$\mathbf{P} = \begin{pmatrix} .43654 & .14470 \\ .14470 & .38504 \end{pmatrix}$$

$$\hat{\mathbf{b}}' = (0. \ 2.7161 \ 1.9565 \ 0.)$$

$$(1/N)\mathbf{y}^{0'}\mathbf{1}\mathbf{1}'\mathbf{y}^0 = 1,413.6978,$$

$$\mathbf{y}^{0'}\mathbf{Z}_B(\mathbf{Z}_B'\mathbf{Z}_B)^{-1}\mathbf{Z}_B'\mathbf{y}^0 = 1,440.2840,$$

$$k_1 = 1.73697 \text{ and } k_2 = 2.05641, \text{ and}$$

$$\hat{\sigma}_0^2 = 6.6758, \quad \hat{\sigma}_B^2 = .37275.$$

## Problems

1. Below are the OLS equations for a 2-way model with factors A(4 levels) and B(5 levels). Consider factor A as fixed and factor B as random. The total sum of squares

of the observations was 43,000. Estimate the variance components by Method 2.

$$\begin{pmatrix} 202 & 46 & 34 & 80 & 42 & 26 & 22 & 46 & 54 & 54 \\ 46 & 46 & 0 & 0 & 0 & 7 & 3 & 4 & 14 & 18 \\ 34 & 0 & 34 & 0 & 0 & 3 & 5 & 14 & 8 & 4 \\ 80 & 0 & 0 & 80 & 0 & 10 & 8 & 18 & 22 & 22 \\ 42 & 0 & 0 & 0 & 42 & 6 & 6 & 10 & 10 & 10 \\ 26 & 7 & 3 & 10 & 6 & 26 & 0 & 0 & 0 & 0 \\ 22 & 3 & 5 & 8 & 6 & 0 & 22 & 0 & 0 & 0 \\ 46 & 4 & 14 & 18 & 10 & 0 & 0 & 46 & 0 & 0 \\ 54 & 14 & 8 & 22 & 10 & 0 & 0 & 0 & 54 & 0 \\ 54 & 18 & 4 & 22 & 10 & 0 & 0 & 0 & 0 & 54 \end{pmatrix} \begin{pmatrix} \hat{\mu} \\ \hat{A}_1 \\ \hat{A}_2 \\ \hat{A}_3 \\ \hat{A}_4 \\ \hat{B}_1 \\ \hat{B}_2 \\ \hat{B}_3 \\ \hat{B}_4 \\ \hat{B}_5 \end{pmatrix} = \begin{pmatrix} 2796 \\ 714 \\ 505 \\ 1067 \\ 510 \\ 398 \\ 334 \\ 553 \\ 739 \\ 772 \end{pmatrix}.$$

2. Prove that  $\mathbf{TXb} = c\mathbf{1}$ . (Not Easy)

3. Method 2 has often been done incorrectly in the literature from 1953 to 1970. The researchers would estimate the fixed effects as

$$\mathbf{b}^o = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y},$$

and used this estimate to adjust the observation vector,  $\mathbf{y}$ . Derive the correct expectations of  $(1/N)\mathbf{y}^{a'}\mathbf{1}\mathbf{1}'\mathbf{y}^a$  and  $\mathbf{y}^{a'}\mathbf{Z}_1(\mathbf{Z}_1'\mathbf{Z}_1)^{-1}\mathbf{Z}_1'\mathbf{y}^a$  where

$$\mathbf{y}^a = \mathbf{y} - \mathbf{Xb}^o,$$

and  $s = 1$ .

# Henderson Method 3

L. R. Schaeffer, April 27, 1999

Method 2 does not allow models to contain interactions between fixed and random factors or nesting of random factors within fixed factors. Henderson's Method 3 is capable of handling general mixed models, but computationally the method is more demanding than Methods 1 or 2. Method 3 has been called '*the fitting constants*' method because reductions in sums of squares due to fitting submodels of the full model are used. The method is unbiased and translation invariant, but is not uniquely defined in the sense that more reductions can be computed than are necessary to estimate the variance components.

Method 3 has been shown in some cases to provide more accurate estimates of variance components than Methods 1 or 2. In special design and model situations, Method 3 is equivalent to Methods 1 or 2. The reason that Method 3 gives greater accuracy seems to be due to fitting submodels such that two or more random factors are simultaneously considered rather than singly as would be the case in Methods 1 or 2.

## The Calculations

For a general model with  $s$  random factors there are  $2^s$  possible submodels and reductions that could be computed, but only  $s + 2$  reductions are needed to estimate the unknown variances. Specific guidelines for choosing the best (most accurate) set of  $s + 2$  reductions out of  $2^s$  reductions are not available. The following rules are recommended to uniquely define Method 3 in all situations and to hopefully provide the most accurate set of reductions. The necessary quadratic forms are a)  $\mathbf{y}'\mathbf{y}$ , the total sum of squares, b) the reduction due to the full model, where the least squares equations are

$$\mathbf{W}'\mathbf{W} \hat{\boldsymbol{\beta}}' = \mathbf{W}'\mathbf{y},$$

then

$$\begin{aligned} R(\mathbf{b}, \mathbf{u}_1, \dots, \mathbf{u}_s) &= R(\boldsymbol{\beta}) = \hat{\boldsymbol{\beta}}' \mathbf{W}'\mathbf{y}, \text{ and} \\ \hat{\sigma}_0^2 &= (\mathbf{y}'\mathbf{y} - R(\boldsymbol{\beta})) / (N - r(\mathbf{W})), \end{aligned}$$

as in Method 2, and c) reductions due to omitting one random factor at a time from the model. Let

$$\boldsymbol{\beta}_j = (\mathbf{b}, \mathbf{u}_1, \dots, \mathbf{u}_{j-1}, \mathbf{u}_{j+1}, \dots, \mathbf{u}_s)$$

with  $\mathbf{u}_j$  omitted from the model and let  $\mathbf{W}_j$  be the corresponding design matrix with  $\mathbf{Z}_j$  omitted. Then the reduction due to this submodel is

$$R(\boldsymbol{\beta}_j) = \hat{\boldsymbol{\beta}}_j' \mathbf{W}_j' \mathbf{y}.$$

There are  $s$  such reductions with one random factor omitted at a time.

To illustrate, assume a model with three random factors A, B, and C, then the following reductions should be used. Notice that each reduction contains all of the fixed factors in the model.

$$\begin{aligned} & \mathbf{y}'\mathbf{y}, \\ & R(\mathbf{b}, A, B, C), \\ & R(\mathbf{b}, A, B), \\ & R(\mathbf{b}, A, C), \\ & R(\mathbf{b}, B, C). \end{aligned}$$

Now assume a model with two random factors and interaction, then use

$$\begin{aligned} & \mathbf{y}'\mathbf{y}, \\ & R(\mathbf{b}, A, B, AB), \\ & R(\mathbf{b}, A, B), \\ & R(\mathbf{b}, A), \\ & R(\mathbf{b}, B). \end{aligned}$$

Note that when interactions are in the model then

$$\begin{aligned} R(\mathbf{b}, A, AB) &= R(\mathbf{b}, B, AB) \\ &= R(\mathbf{b}, A, B, AB) \\ &= R(\mathbf{b}, AB), \end{aligned}$$

and therefore some slight differences in choice of reductions is necessary when interactions are present in the model.

Suppose that in a two random factor model, factor D is a fixed factor, but that there is an interaction between D and random factor A, but not with random factor B. Then use

$$\begin{aligned} & \mathbf{y}'\mathbf{y}, \\ & R(\mathbf{b}, A, B, AD), \\ & R(\mathbf{b}, A, B), \\ & R(\mathbf{b}, A, AD), \\ & R(\mathbf{b}, B). \end{aligned}$$

In this case

$$R(\mathbf{b}, B, AD) = R(\mathbf{b}, A, B, AD)$$

and

$$R(\mathbf{b}, A, AD) = R(\mathbf{b}, AD).$$

## Expectations of Reductions

Let

$$\begin{aligned}\mathbf{C}_j &= (\mathbf{W}'_j \mathbf{W}_j)^-, \text{ then} \\ \hat{\boldsymbol{\beta}}_j &= \mathbf{C}_j \mathbf{W}'_j \mathbf{y}, \text{ and} \\ R(\boldsymbol{\beta}_j) &= \hat{\boldsymbol{\beta}}'_j \mathbf{W}'_j \mathbf{y} \\ &= \mathbf{y}' \mathbf{W}_j \mathbf{C}_j \mathbf{W}'_j \mathbf{y}.\end{aligned}$$

The expectation is then

$$E[R(\boldsymbol{\beta}_j)] = \mathbf{b}' \mathbf{X}' \mathbf{W}_j \mathbf{C}_j \mathbf{W}'_j \mathbf{X} \mathbf{b} + \sum_{i=0}^s tr(\mathbf{W}_j \mathbf{C}_j \mathbf{W}'_j \mathbf{V}_i) \sigma_i^2.$$

From the theory of generalized inverses,

$$\mathbf{W}_j \mathbf{C}_j \mathbf{W}'_j \mathbf{W}_j = \mathbf{W}_j$$

and consequently,

$$\mathbf{b}' \mathbf{X}' \mathbf{W}_j \mathbf{C}_j \mathbf{W}'_j \mathbf{X} \mathbf{b} = \mathbf{b}' \mathbf{X}' \mathbf{X} \mathbf{b}$$

and

$$\begin{aligned}tr(\mathbf{W}_j \mathbf{C}_j \mathbf{W}'_j \mathbf{Z}_i \mathbf{Z}'_i) &= tr(\mathbf{Z}_i \mathbf{Z}'_i) \text{ provided that } \mathbf{Z}_i \text{ was included in } \mathbf{W}_j, \\ &\text{or} \\ tr(\mathbf{W}_j \mathbf{C}_j \mathbf{W}'_j \mathbf{Z}_i \mathbf{Z}'_i) &= tr[\mathbf{C}_j (\mathbf{W}'_j \mathbf{Z}_i) (\mathbf{Z}'_i \mathbf{W}_j)] \text{ if } \mathbf{Z}_i \text{ was not in } \mathbf{W}_j.\end{aligned}$$

Also, the coefficient of  $\sigma_0^2$  is

$$tr(\mathbf{W}_j \mathbf{C}_j \mathbf{W}'_j) = tr(\mathbf{C}_j \mathbf{W}'_j \mathbf{W}_j),$$

but we know that  $\mathbf{C}_j \mathbf{W}'_j \mathbf{W}_j$  is idempotent and the trace of an idempotent matrix is equal to the rank of the matrix, so that  $tr(\mathbf{C}_j \mathbf{W}'_j \mathbf{W}_j) = r(\mathbf{W}_j)$ . Therefore,

$$E[R(\boldsymbol{\beta}_j)] = \mathbf{b}' \mathbf{X}' \mathbf{X} \mathbf{b} + r(\mathbf{W}_j) \sigma_0^2 + \sum_{i=1}^s tr[\mathbf{C}_j (\mathbf{W}'_j \mathbf{Z}_i) (\mathbf{Z}'_i \mathbf{W}_j)] \sigma_i^2.$$

## An Example

Consider a three factor model with factors A, B, and C where A is a fixed factor and

B and C are random factors, and the LS equations are as given below:

$$\left( \begin{array}{ccc|ccc|ccccc} 71 & 40 & 31 & 28 & 21 & 22 & 16 & 14 & 13 & 16 & 12 \\ \hline 40 & 40 & 0 & 18 & 10 & 12 & 13 & 0 & 9 & 14 & 4 \\ 31 & 0 & 31 & 10 & 11 & 10 & 3 & 14 & 4 & 2 & 8 \\ \hline 28 & 18 & 10 & 28 & 0 & 0 & 6 & 6 & 4 & 8 & 4 \\ 21 & 10 & 11 & 0 & 21 & 0 & 3 & 4 & 4 & 6 & 4 \\ 22 & 12 & 10 & 0 & 0 & 22 & 7 & 4 & 5 & 2 & 4 \\ \hline 16 & 13 & 3 & 6 & 3 & 7 & 16 & 0 & 0 & 0 & 0 \\ 14 & 0 & 14 & 6 & 4 & 4 & 0 & 14 & 0 & 0 & 0 \\ 13 & 9 & 4 & 4 & 4 & 5 & 0 & 0 & 13 & 0 & 0 \\ 16 & 14 & 2 & 8 & 6 & 2 & 0 & 0 & 0 & 16 & 0 \\ 12 & 4 & 8 & 4 & 4 & 4 & 0 & 0 & 0 & 0 & 12 \end{array} \right) \begin{pmatrix} \hat{\mu} \\ \hat{A}_1 \\ \hat{A}_2 \\ \hat{B}_1 \\ \hat{B}_2 \\ \hat{B}_3 \\ \hat{C}_1 \\ \hat{C}_2 \\ \hat{C}_3 \\ \hat{C}_4 \\ \hat{C}_5 \end{pmatrix} = \begin{pmatrix} 294 \\ 195 \\ 99 \\ 90 \\ 89 \\ 115 \\ 67 \\ 36 \\ 62 \\ 73 \\ 56 \end{pmatrix}$$

The reductions and their expectations are

$$\begin{aligned} \mathbf{y}'\mathbf{y} &= 2,834, \\ E(\mathbf{y}'\mathbf{y}) &= \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b} + N\sigma_B^2 + N\sigma_C^2 + N\sigma_0^2 \\ R(\mathbf{b}, B, C) &= 1,345.119648, \\ E[R(\mathbf{b}, B, C)] &= \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b} + N\sigma_B^2 + N\sigma_C^2 + \text{r}(\mathbf{X}\mathbf{Z}_B\mathbf{Z}_C)\sigma_0^2, \\ R(\mathbf{b}, B) &= 1,327.550930, \\ E[R(\mathbf{b}, B)] &= \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b} + N\sigma_B^2 + \text{tr}(\mathbf{C}_C(\mathbf{W}'_C\mathbf{Z}_C)(\mathbf{Z}'_C\mathbf{W}_C))\sigma_C^2 \\ &\quad + \text{r}(\mathbf{X}\mathbf{Z}_B)\sigma_0^2, \\ R(\mathbf{b}, C) &= 1,286.908325, \\ E[R(\mathbf{b}, C)] &= \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b} + \text{tr}(\mathbf{C}_B(\mathbf{W}'_B\mathbf{Z}_B)(\mathbf{Z}'_B\mathbf{W}_B))\sigma_B^2 + N\sigma_C^2 \\ &\quad + \text{r}(\mathbf{X}\mathbf{Z}_C)\sigma_0^2. \end{aligned}$$

Summarizing the above,

$$E \begin{pmatrix} \mathbf{y}'\mathbf{y} \\ R(\mathbf{b}, B, C) \\ R(\mathbf{b}, B) \\ R(\mathbf{b}, C) \end{pmatrix} = \begin{pmatrix} 1 & 71. & 71. & 71. \\ 1 & 71. & 71. & 8. \\ 1 & 71. & 22.0095 & 4. \\ 1 & 26.3804 & 71. & 6. \end{pmatrix} \begin{pmatrix} \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b} \\ \sigma_B^2 \\ \sigma_C^2 \\ \sigma_0^2 \end{pmatrix}$$

Taking differences from  $\mathbf{y}'\mathbf{y}$  eliminates the  $\mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b}$  term from all quadratic forms and gives

$$\begin{pmatrix} 0. & 0. & 63. \\ 0. & 48.9905 & 67. \\ 44.6196 & 0. & 65. \end{pmatrix} \begin{pmatrix} \hat{\sigma}_B^2 \\ \hat{\sigma}_C^2 \\ \hat{\sigma}_0^2 \end{pmatrix} = \begin{pmatrix} 1,488.880352 \\ 1,506.449070 \\ 1,547.091675 \end{pmatrix},$$

and the solutions are  $\hat{\sigma}_B^2 = 0.245302$ ,  $\hat{\sigma}_C^2 = -1.570986$ , and  $\hat{\sigma}_0^2 = 23.633021$ . Notice the zeros in the above expectation matrix as a result of taking differences from  $\mathbf{y}'\mathbf{y}$ . This helps to reduce the sampling variances of the estimated variance components over other possible sets of reductions that could be used in Method 3.

## Problems

1. Below are feed consumption data on progeny of 3 sires and 10 dams over a ten day period after weaning. Age at weaning is known to influence feed consumption during this period.

Table of feed consumption data.

Sire	Dam	Sex	Age at Weaning(d)	Feed Intake (mouthfuls)
1	1	1	185	775
1	1	1	185	690
1	2	2	190	618
1	2	1	190	631
1	3	2	200	615
1	3	1	200	704
1	4	2	210	573
2	5	1	215	662
2	5	1	215	600
2	6	1	170	526
2	6	2	170	485
2	6	2	170	581
2	7	2	175	494
3	8	1	206	660
3	8	2	206	511
3	9	2	180	500
3	9	1	180	450
3	10	2	194	470
3	10	1	194	505
3	10	1	194	516
3	10	2	194	448

Assume the model is

$$y_{ijkl} = \mu + b(\text{Age}) + X_i + S_j + D_{jk} + e_{ijkl}$$

where  $y_{ijkl}$  is feed intake,  $\mu$  is an overall mean,  $b(\text{Age})$  is a covariate of feed intake on age at weaning,  $X_i$  is a sex of offspring effect (fixed),  $S_j$  is a random sire effect,  $D_{jk}$  is a random dam effect, and  $e_{ijkl}$  is a random residual effect.

Estimate the variance components by Henderson's Method 3. You might want to take deviations of age at weaning from the overall average weaning age of 200-d, to avoid rounding errors.

2. Suppose we have data to be analyzed by the following model

$$y_{ijklm} = \mu + A_i + B_{ij} + C_k + D_{kl} + e_{ijklm},$$

where the fixed factors of the model are  $\mu$  and  $A_i$ . Write all of the possible reductions for Method 3 that could be used with this model, and mark those that you would use to estimate the variances.

# Other Simple Unbiased Methods

L. R. Schaeffer, April 27, 1999

## Absorption of Fixed Effects

The absorption of all fixed effects in the LS equations into the equations for the random factors essentially adjusts the right hand sides of the random effects for the fixed effects, and consequently any quadratic forms involving the absorbed right hand sides are translation invariant. Numerous methods may therefore be defined by using quadratic forms of the absorbed right hand sides. The LS equations are

$$\begin{pmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} \end{pmatrix} \begin{pmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{pmatrix}.$$

Let

$$\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}',$$

and partition  $\mathbf{Z}$  into  $[\mathbf{Z}_1 \ \mathbf{Z}_2 \ \dots \ \mathbf{Z}_s]$ , then the absorbed random effects equations are

$$\begin{pmatrix} \mathbf{Z}'_1\mathbf{M}\mathbf{Z}_1 & \mathbf{Z}'_1\mathbf{M}\mathbf{Z}_2 & \dots & \mathbf{Z}'_1\mathbf{M}\mathbf{Z}_s \\ \mathbf{Z}'_2\mathbf{M}\mathbf{Z}_1 & \mathbf{Z}'_2\mathbf{M}\mathbf{Z}_2 & \dots & \mathbf{Z}'_2\mathbf{M}\mathbf{Z}_s \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{Z}'_s\mathbf{M}\mathbf{Z}_1 & \mathbf{Z}'_s\mathbf{M}\mathbf{Z}_2 & \dots & \mathbf{Z}'_s\mathbf{M}\mathbf{Z}_s \end{pmatrix} \begin{pmatrix} \hat{\mathbf{u}}_1 \\ \hat{\mathbf{u}}_2 \\ \vdots \\ \hat{\mathbf{u}}_s \end{pmatrix} = \begin{pmatrix} \mathbf{Z}'_1\mathbf{M}\mathbf{y} \\ \mathbf{Z}'_2\mathbf{M}\mathbf{y} \\ \vdots \\ \mathbf{Z}'_s\mathbf{M}\mathbf{y} \end{pmatrix}.$$

To illustrate various methods, suppose we have the following LS equations for a model with two random factors (C and D) and two fixed factors (A and B).

$$\begin{pmatrix} 200 & 150 & 50 & 75 & 125 & 20 & 35 & 50 & 65 & 30 & 65 & 70 & 35 & 30 \\ 150 & 150 & 0 & 50 & 100 & 15 & 25 & 30 & 55 & 25 & 55 & 60 & 15 & 20 \\ 50 & 0 & 50 & 25 & 25 & 5 & 10 & 20 & 10 & 5 & 10 & 10 & 20 & 10 \\ 75 & 50 & 25 & 75 & 0 & 10 & 15 & 15 & 20 & 15 & 20 & 20 & 25 & 10 \\ 125 & 100 & 25 & 0 & 125 & 10 & 20 & 35 & 45 & 15 & 45 & 50 & 10 & 20 \\ \hline 20 & 15 & 5 & 10 & 10 & 20 & 0 & 0 & 0 & 0 & 5 & 6 & 4 & 5 \\ 35 & 25 & 10 & 15 & 20 & 0 & 35 & 0 & 0 & 0 & 9 & 12 & 6 & 8 \\ 50 & 30 & 20 & 15 & 35 & 0 & 0 & 50 & 0 & 0 & 16 & 21 & 8 & 5 \\ 65 & 55 & 10 & 20 & 45 & 0 & 0 & 0 & 65 & 0 & 25 & 27 & 7 & 6 \\ 30 & 25 & 5 & 15 & 15 & 0 & 0 & 0 & 0 & 30 & 10 & 4 & 10 & 6 \\ 65 & 55 & 10 & 20 & 45 & 5 & 9 & 16 & 25 & 10 & 65 & 0 & 0 & 0 \\ 70 & 60 & 10 & 20 & 50 & 6 & 12 & 21 & 27 & 4 & 0 & 70 & 0 & 0 \\ 35 & 15 & 20 & 25 & 10 & 4 & 6 & 8 & 7 & 10 & 0 & 0 & 35 & 0 \\ 30 & 20 & 10 & 10 & 20 & 5 & 8 & 5 & 6 & 6 & 0 & 0 & 0 & 30 \end{pmatrix} \begin{pmatrix} \hat{\mu} \\ \hat{A}_1 \\ \hat{A}_2 \\ \hat{B}_1 \\ \hat{B}_2 \\ \hline \hat{C}_1 \\ \hat{C}_2 \\ \hat{C}_3 \\ \hat{C}_4 \\ \hat{C}_5 \\ \hat{D}_1 \\ \hat{D}_2 \\ \hat{D}_3 \\ \hat{D}_4 \end{pmatrix} = \begin{pmatrix} 21125 \\ 16200 \\ 4925 \\ 7850 \\ 13275 \\ \hline 2505 \\ 3960 \\ 5710 \\ 6620 \\ 2230 \\ 9040 \\ 8135 \\ 2405 \\ 1545 \end{pmatrix}.$$

Absorbing the fixed effects ( $\mu A_1 A_2 B_1 B_2$ ) gives (each element is multiplied by 11)

$$\begin{pmatrix} 196.5 & -39.5 & -52.0 & -69.5 & -35.5 & -14.5 & -8.0 & -0.5 & 23.0 \\ -39.5 & 316.5 & -97.0 & -121.5 & -58.5 & -22.5 & 2.0 & -9.5 & 30.0 \\ -52.0 & -97.0 & 390.0 & -169.0 & -72.0 & 7.0 & 49.0 & -21.0 & -35.0 \\ -69.5 & -121.5 & -169.0 & 468.5 & -108.5 & 28.5 & 29.0 & -19.5 & -38.0 \\ -35.5 & -58.5 & -72.0 & -108.5 & 274.5 & 1.5 & -72.0 & 50.5 & 20.0 \\ -14.5 & -22.5 & 7.0 & 28.5 & 1.5 & 468.5 & -268.0 & -96.5 & -104.0 \\ -8.0 & 2.0 & 49.0 & 29.0 & -72.0 & -268.0 & 478.0 & -98.0 & -112.0 \\ -0.5 & -9.5 & -21.0 & -19.5 & 50.5 & -96.5 & -98.0 & 256.5 & -62.0 \\ 23.0 & 30.0 & -35.0 & -38.0 & 20.0 & -104.0 & -112.0 & -62.0 & 278.0 \end{pmatrix} \begin{pmatrix} \hat{C}_1 \\ \hat{C}_2 \\ \hat{C}_3 \\ \hat{C}_4 \\ \hat{C}_5 \\ \hat{D}_1 \\ \hat{D}_2 \\ \hat{D}_3 \\ \hat{D}_4 \end{pmatrix} = \begin{pmatrix} 4325 \\ 3030 \\ 5485 \\ -3365 \\ -10575 \\ 23255 \\ 7355 \\ -13005 \\ -17605 \end{pmatrix}.$$

### Quadratic Forms Of Absorbed RHS

Let

$$\begin{aligned} \mathbf{r}_j &= \mathbf{Z}'_j \mathbf{M} \mathbf{y}, \\ \text{then } E(\mathbf{r}_j) &= \mathbf{Z}'_j \mathbf{M} E(\mathbf{y}) \\ &= \mathbf{Z}'_j \mathbf{M} \mathbf{X} \mathbf{b}, \\ \text{but } \mathbf{M} \mathbf{X} &= \mathbf{0}, \\ \text{so } E(\mathbf{r}_j) &= 0, \text{ for } j = 1 \text{ to } s. \end{aligned}$$

The variance-covariance matrix of  $\mathbf{Z}' \mathbf{M} \mathbf{y}$  is

$$\begin{aligned} \text{Var}(\mathbf{Z}' \mathbf{M} \mathbf{y}) &= \sum_{i=0}^s \mathbf{Z}' \mathbf{M} \mathbf{V}_i \mathbf{M} \mathbf{Z} \sigma_i^2 \\ &= \sum_{i=1}^s (\mathbf{Z}' \mathbf{M} \mathbf{Z}_i) (\mathbf{Z}'_i \mathbf{M} \mathbf{Z}) \sigma_i^2 + \mathbf{Z}' \mathbf{M} \mathbf{Z} \sigma_0^2. \end{aligned}$$

Let  $\mathbf{r}' = [ \mathbf{r}'_1 \ \mathbf{r}'_2 \ \dots \ \mathbf{r}'_s ]$ , then a quadratic form of  $\mathbf{r}$  has expectation

$$E(\mathbf{r}' \mathbf{Q} \mathbf{r}) = \sum_{i=1}^s \text{tr}[\mathbf{Q} (\mathbf{Z}' \mathbf{M} \mathbf{Z}_i) (\mathbf{Z}'_i \mathbf{M} \mathbf{Z})] \sigma_i^2 + \text{tr}[\mathbf{Q} \mathbf{Z}' \mathbf{M} \mathbf{Z}] \sigma_0^2.$$

These expectations are easy to calculate when  $\mathbf{Q}$  is a diagonal matrix, in which case only the diagonals of  $\mathbf{Z}'\mathbf{M}\mathbf{Z}$  and  $(\mathbf{Z}'\mathbf{M}\mathbf{Z}_i)(\mathbf{Z}'_i\mathbf{M}\mathbf{Z})$  are needed, and these are multiplied times the diagonal elements of  $\mathbf{Q}$ . Below are the diagonal elements of the matrices needed to compute the expectations of quadratic forms of the absorbed right hand sides for the example data.

Element of Model	$(\mathbf{Z}'\mathbf{M}\mathbf{Z}_C)(\mathbf{Z}'_C\mathbf{M}\mathbf{Z})$	$(\mathbf{Z}'\mathbf{M}\mathbf{Z}_D)(\mathbf{Z}'_D\mathbf{M}\mathbf{Z})$	$\mathbf{Z}'\mathbf{M}\mathbf{Z}$
$C_1$	404.685950	6.640496	17.863636
$C_2$	1,068.809917	12.400826	28.772727
$C_3$	1,636.016529	34.016529	35.454545
$C_4$	2,309.239669	28.739669	42.590909
$C_5$	801.561984	67.243802	24.954545
$D_1$	13.057851	2,573.921488	42.590909
$D_2$	70.198347	2,664.925620	43.454545
$D_3$	28.611570	731.838843	23.318182
$D_4$	37.173554	863.537190	25.272727

### MIVQUE-0

The most simple set of quadratic forms of  $\mathbf{r}$  that could be used are  $\mathbf{r}'\mathbf{Q}_1\mathbf{r}$  and  $\mathbf{r}'\mathbf{Q}_2\mathbf{r}$

$$\mathbf{Q}_1 = \mathbf{diag}(1\ 1\ 1\ 1\ 1\ 0\ 0\ 0\ 0),$$

$$\text{and } \mathbf{Q}_2 = \mathbf{diag}(0\ 0\ 0\ 0\ 0\ 1\ 1\ 1\ 1),$$

where  $\mathbf{diag}$  indicates a diagonal matrix whose diagonals are the elements within the parentheses.

Assume that  $\mathbf{y}'\mathbf{M}\mathbf{y} = 668,160.62$  for this example with  $r(\mathbf{M}) = 200 - 3 = 197$ , degrees of freedom, then

$$\begin{aligned} \mathbf{r}'\mathbf{Q}_1\mathbf{r} &= (1/121) [ 4325^2 + 3030^2 + 5485^2 + 3365^2 + 10575^2 ] \\ &= 1,496,905.8, \quad \text{and} \\ E(\mathbf{r}'\mathbf{Q}_1\mathbf{r}) &= (404.7 + 1068.8 + 1636.0 + 2309.2 + 801.6)\sigma_C^2 \\ &\quad + (6.64 + 12.40 + 34.02 + 28.74 + 67.24)\sigma_D^2 \\ &\quad + (17.86 + 28.77 + 35.45 + 42.59 + 24.95)\sigma_0^2 \\ &= 6220.314\sigma_C^2 + 149.041\sigma_D^2 + 149.636\sigma_0^2, \\ \mathbf{r}'\mathbf{Q}_2\mathbf{r} &= (1/121) [ 23255^2 + 7355^2 + 13005^2 + 17605^2 ] \\ &= 8,875,678.512, \quad \text{and} \\ E(\mathbf{r}'\mathbf{Q}_2\mathbf{r}) &= (13.06 + 70.20 + 28.61 + 37.17)\sigma_C^2 + (2573.9 \\ &\quad + 2664.9 + 731.8 + 863.5)\sigma_D^2 + (42.59 + 23.32 \end{aligned}$$

$$\begin{aligned}
& +43.45 + 25.27) \sigma_0^2 \\
& = 149.041 \sigma_C^2 + 6834.223 \sigma_D^2 + 134.636 \sigma_0^2, \quad \text{and} \\
E(\mathbf{y}'\mathbf{M}\mathbf{y}) & = \text{tr}(\mathbf{Z}'_C\mathbf{M}\mathbf{Z}_C) \sigma_C^2 + \text{tr}(\mathbf{Z}'_D\mathbf{M}\mathbf{Z}_D) \sigma_D^2 + r(\mathbf{M}) \sigma_0^2 \\
& = 149.636 \sigma_C^2 + 134.636 \sigma_D^2 + 197 \sigma_0^2,
\end{aligned}$$

The resulting equations are

$$\begin{pmatrix} 6220.314 & 149.041 & 149.636 \\ 149.041 & 6834.223 & 134.636 \\ 149.636 & 134.636 & 197. \end{pmatrix} \begin{pmatrix} \hat{\sigma}_C^2 \\ \hat{\sigma}_D^2 \\ \hat{\sigma}_0^2 \end{pmatrix} = \begin{pmatrix} 1,496,905.8 \\ 8,875,678.5 \\ 668,160.62 \end{pmatrix}$$

which gives  $\hat{\sigma}_C^2 = 152.46$ ,  $\hat{\sigma}_D^2 = 1247.65$ , and  $\hat{\sigma}_0^2 = 2423.19$ .

The above procedure is known as MIVQUE-0. MIVQUE-0 stands for Minimum Variance Quadratic Unbiased Estimation using zeros for prior values of the variance components except  $\sigma_0^2$ . Calling this procedure MIVQUE is very misleading because the estimates from MIVQUE-0 have been shown to have very large sampling variances. However, because the method is very easy to apply, MIVQUE-0 estimates are frequently used as starting values for other methods of estimation. The estimates from MIVQUE-0 are unbiased and translation invariant, but are not minimum variance.

### Quadratics Using Diagonals Of LS Equations

Many methods could be described by simply changing the definition of  $\mathbf{Q}$ . For example, define  $\mathbf{Q}$  as a diagonal matrix with corresponding elements equal to the diagonals of  $(\mathbf{Z}'_j\mathbf{Z}_j)^{-1}$  and the rest to zero. For the example data, let

$$\mathbf{Q}_3 = \text{diag}( .05 \ .028571 \ .02 \ .015385 \ .033333 \ 0 \ 0 \ 0 \ 0 )$$

and

$$\mathbf{Q}_4 = \text{diag}( 0 \ 0 \ 0 \ 0 \ 0 \ .015385 \ .014286 \ .028571 \ .033333 )$$

then

$$\begin{aligned}
\mathbf{r}'\mathbf{Q}_3\mathbf{r} & = (1/121)[ (.05)4325^2 + (.028571)3030^2 + (.02)5485^2 \\
& \quad + (.015385)3365^2 + (.033333)10575^2 ] \\
& = 47,116.953,
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{r}'\mathbf{Q}_4\mathbf{r} & = (1/121)[ (.015385)23255^2 + (.014286)7355^2 \\
& \quad + (.028571)13005^2 + (.033333)17605^2 ] \\
& = 200,464.95.
\end{aligned}$$

The expectations are

$$E(\mathbf{r}'\mathbf{Q}_3\mathbf{r}) = [ (.05)404.69 + (.028571)1068.81 + (.02)1636.02$$

$$\begin{aligned}
& +(.0153585)2309.24 + (.033333)801.56 ]\sigma_C^2 \\
& + [ (.05)6.64 + (.028571)12.40 + (.02)34.02 \\
& +(.015385)28.74 + (.033333)67.24 ]\sigma_D^2 \\
& + [ (.05)17.86 + (.028571)28.77 + (.02)35.45 \\
& +(.015385)42.59 + (.033333)24.95 ]\sigma_0^2 \\
= & 145.738 \sigma_C^2 + 4.050\sigma_D^2 + 3.911 \sigma_0^2
\end{aligned}$$

and

$$\begin{aligned}
E(\mathbf{r}'\mathbf{Q}_4\mathbf{r}) & = [ (.015385)13.06 + (.014286)70.20 \\
& +(.028571)28.61 + (.033333)37.17 ]\sigma_C^2 \\
& + [ (.015385)2573.92 + (.014286)2664.93 \\
& +(.028571)731.84 + (.033333)863.54 ]\sigma_D^2 \\
& + [ (.015385)42.59 + (.014286)43.45 \\
& +(.028571)23.32 + (.033333)25.27 ]\sigma_0^2 \\
= & 3.260 \sigma_C^2 + 127.363 \sigma_D^2 + 2.785 \sigma_0^2
\end{aligned}$$

The resulting estimates using the above quadratics with  $\mathbf{y}'\mathbf{M}\mathbf{y}$  are

$$\hat{\sigma}_C^2 = 191.38, \quad \hat{\sigma}_D^2 = 1495.44, \quad \text{and} \quad \hat{\sigma}_0^2 = 3366.74.$$

#### Henderson's Method 4

In 1980 Henderson presented a particular set of  $\mathbf{Q}$  matrices for the preceding methodology, which was formally described in his book (1984). This method has been given various names such as Henderson's New Method, Diagonal MIVQUE, and Henderson's Method 4. Dempfle et al. (1983) showed that the sampling variances of Method 4 estimates were similar to and sometimes better than those of Method 3 estimates.

Method 4 uses a specific set of quadratic forms. The motivation for these particular forms is that they more closely approximate the quadratic forms used with MIVQUE (which will be discussed later). The specific quadratic forms are, for the example data:

$\mathbf{Q}_5 =$  diagonals of  $(\mathbf{Z}'_C\mathbf{M}\mathbf{Z}_C + \mathbf{I}k_C)^{-2}$ , plus zeros for the rows corresponding to other random factors, where  $k_C =$  a guess of the value of  $\sigma_0^2/\sigma_C^2$ , and

$\mathbf{Q}_6 =$  diagonals of  $(\mathbf{Z}'_D\mathbf{M}\mathbf{Z}_D + \mathbf{I}k_D)^{-2}$ , plus zeros for the rows corresponding to other random factors, where  $k_D =$  a guess of the value of  $\sigma_0^2/\sigma_D^2$ .

Assume that  $k_C = 16$  and that  $k_D = 2$ , then

$$\begin{aligned}
\mathbf{Q}_5 & = (10^{-4})[8.720 \ 4.989 \ 3.777 \ 2.913 \ 5.962 \ 0 \ 0 \ 0 \ 0], \\
\text{and} \\
\mathbf{Q}_6 & = (10^{-4})[0 \ 0 \ 0 \ 0 \ 0 \ 5.029 \ 4.84 \ 15.6 \ 13.444].
\end{aligned}$$

Completing the calculations,

$$\begin{aligned} \mathbf{r}'\mathbf{Q}_5\mathbf{r} &= 844.85 \\ \text{and } \mathbf{r}'\mathbf{Q}_6\mathbf{r} &= 8088.17 \\ E(\mathbf{r}'\mathbf{Q}_5\mathbf{r}) &= 2.6546 \sigma_C^2 + .0733 \sigma_D^2 + .0706 \sigma_0^2, \text{ and} \\ E(\mathbf{r}'\mathbf{Q}_6\mathbf{r}) &= .1352 \sigma_C^2 + 4.8869 \sigma_D^2 + .1128 \sigma_0^2. \end{aligned}$$

The estimates, using  $\mathbf{y}'\mathbf{M}\mathbf{y}$  and the above quadratics, are

$$\hat{\sigma}_C^2 = 184.68, \quad \hat{\sigma}_D^2 = 1571.71, \quad \text{and} \quad \hat{\sigma}_0^2 = 3390.71.$$

In all of the methods in this section,  $\mathbf{y}'\mathbf{M}\mathbf{y}$  can be replaced by any unbiased estimate of  $\sigma_0^2$ .

### Problem

Below are parts of the OLS equations for a model with five factors. The first three factors are fixed and the last two factors are random (each with 6 levels). The total sum of squares was 3,143,224 based on 41 observations.

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} 11 & 0 & 0 & 8 & 3 & 1 & 1 & 3 & 4 & 2 \\ 0 & 14 & 0 & 6 & 8 & 3 & 2 & 2 & 2 & 5 \\ 0 & 0 & 16 & 7 & 9 & 5 & 5 & 3 & 2 & 1 \\ 8 & 6 & 7 & 21 & 0 & 4 & 3 & 6 & 4 & 4 \\ 3 & 8 & 9 & 0 & 20 & 5 & 5 & 2 & 4 & 4 \\ 1 & 3 & 5 & 4 & 5 & 9 & 0 & 0 & 0 & 0 \\ 1 & 2 & 5 & 3 & 5 & 0 & 8 & 0 & 0 & 0 \\ 3 & 2 & 3 & 6 & 2 & 0 & 0 & 8 & 0 & 0 \\ 4 & 2 & 2 & 4 & 4 & 0 & 0 & 0 & 8 & 0 \\ 2 & 5 & 1 & 4 & 4 & 0 & 0 & 0 & 0 & 8 \end{pmatrix},$$

$$\mathbf{X}'\mathbf{Z} = \begin{pmatrix} 4 & 4 & 2 & 0 & 1 & 0 & 4 & 5 & 0 & 1 & 1 & 0 \\ 1 & 2 & 2 & 3 & 3 & 3 & 0 & 1 & 6 & 5 & 2 & 0 \\ 2 & 1 & 3 & 4 & 3 & 3 & 2 & 1 & 0 & 1 & 5 & 7 \\ 7 & 3 & 6 & 2 & 3 & 0 & 6 & 4 & 4 & 3 & 4 & 0 \\ 0 & 4 & 1 & 5 & 4 & 6 & 0 & 3 & 2 & 4 & 4 & 7 \\ 1 & 0 & 3 & 1 & 2 & 2 & 1 & 1 & 0 & 3 & 1 & 3 \\ 2 & 0 & 0 & 2 & 0 & 4 & 2 & 0 & 1 & 1 & 2 & 2 \\ 4 & 1 & 0 & 3 & 0 & 0 & 2 & 2 & 1 & 0 & 1 & 2 \\ 0 & 5 & 1 & 0 & 2 & 0 & 1 & 2 & 3 & 1 & 1 & 0 \\ 0 & 1 & 3 & 1 & 3 & 0 & 0 & 2 & 1 & 2 & 3 & 0 \end{pmatrix},$$

$$\mathbf{Z}'_1\mathbf{Z}_1 = \text{diag}(7 \ 7 \ 7 \ 7 \ 7 \ 6),$$

$$\mathbf{Z}'_2\mathbf{Z}_2 = \text{diag}(6 \ 7 \ 6 \ 7 \ 8 \ 7),$$

$$\mathbf{Z}'_1 \mathbf{Z}_2 = \begin{pmatrix} 4 & 2 & 0 & 0 & 1 & 0 \\ 1 & 2 & 3 & 0 & 1 & 0 \\ 0 & 2 & 0 & 3 & 2 & 0 \\ 0 & 0 & 2 & 0 & 2 & 3 \\ 1 & 1 & 1 & 1 & 1 & 2 \\ 0 & 0 & 0 & 3 & 1 & 2 \end{pmatrix},$$

and the right hand sides are

$$\mathbf{X}'\mathbf{y} = \begin{pmatrix} 3338 \\ 3365 \\ 4091 \\ 5855 \\ 4939 \\ 2441 \\ 2087 \\ 2295 \\ 2195 \\ 1776 \end{pmatrix},$$

$$\mathbf{Z}'_1 \mathbf{y} = \begin{pmatrix} 2541 \\ 1954 \\ 1775 \\ 1511 \\ 1688 \\ 1325 \end{pmatrix}, \quad \text{and} \quad \mathbf{Z}'_2 \mathbf{y} = \begin{pmatrix} 2201 \\ 2146 \\ 1674 \\ 1669 \\ 1624 \\ 1480 \end{pmatrix}.$$

Estimate the variances using MIVQUE-0, Henderson's Method 4, and another set of quadratic forms of your own choice.

# MIVQUE

L. R. Schaeffer, April 27, 1999

During the late 1960's, statisticians began to develop methods of variance component estimation that had more desirable properties than simply unbiasedness and translation invariance. Two researchers, working independently, derived minimum variance quadratic unbiased estimators (MIVQUE) in 1970. C.R. Rao published four papers in his newly formed journal on MIVQUE for  $\mathbf{y}$  being normally distributed. Rao also derived a method called MINQUE (minimum norm quadratic unbiased estimation) for cases when  $\mathbf{y}$  does not have a normal distribution, but when  $\mathbf{y}$  is normally distributed then minimizing the Euclidean norm, as Rao does, is equivalent to minimizing the variance of the quadratic form which gives MIVQUE. In order to minimize the variance of quadratic forms, the true unknown variances must be known.

At the same time, L.R. LaMotte, at the University of Kentucky, also derived MIVQUE in bulletins which were not formally published in a statistical journal. Henderson(1973) took the MIVQUE method and derived easier computing formulas based upon solutions to the mixed model equations (MME) and upon the elements of the generalized inverse of the coefficient matrix of the MME.

MIVQUE assumes that  $\mathbf{V}$ , the variance-covariance matrix of  $\mathbf{y}$ , is known. The variances of the quadratic forms are minimized for this  $\mathbf{V}$ . Since  $\mathbf{V}$  is not known in reality, some prior information about  $\mathbf{V}$  is utilized, and consequently the variance of quadratic forms is only minimized for this prior  $\mathbf{V}$ . MIVQUE is unbiased and translation invariant and if the prior  $\mathbf{V}$  is the same as the true  $\mathbf{V}$ , then it also has minimum variance. Thus, MIVQUE in general, is not minimum variance in practice, but is only as good as the prior values that are used.

## Quadratic Forms

Assume that  $\mathbf{y}$  has a multivariate normal distribution with mean vector  $\mathbf{Xb}$ , and variance-covariance matrix  $\mathbf{V}$ . Let  $\theta_i$  be the prior information about  $\sigma_i^2$  for  $i = 0$  to  $s$ , then let  $\mathbf{H}$  be our prior  $\mathbf{V}$  which is

$$\mathbf{H} = \sum_{j=1}^s \mathbf{Z}_j \mathbf{Z}_j' \theta_j + \mathbf{I} \theta_0.$$

A projection matrix is constructed as

$$\mathbf{P} = \mathbf{H}^{-1} - \mathbf{H}^{-1} \mathbf{X} (\mathbf{X}' \mathbf{H}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{H}^{-1}.$$

A projection matrix is simply a way of absorbing the fixed effects and at the same time accounting for information about  $\mathbf{V}$ . Note that  $\mathbf{PX} = \mathbf{0}$  and that  $\mathbf{PH}$  is idempotent. The MIVQUE quadratic forms are then

$$\mathbf{y}' \mathbf{P} \mathbf{Z}_i \mathbf{Z}_i' \mathbf{P} \mathbf{y} \text{ for } i = 1 \text{ to } s \text{ and } \mathbf{y}' \mathbf{P} \mathbf{P} \mathbf{y}.$$

Note that these are quadratic forms of  $\mathbf{P}\mathbf{y}$  and since  $E(\mathbf{P}\mathbf{y}) = 0$ , then the quadratic forms are translation invariant. Also,

$$\begin{aligned}\sum_{i=1}^s \mathbf{y}'\mathbf{P}\mathbf{Z}_i\mathbf{Z}_i'\mathbf{P}\mathbf{y} \theta_i + \mathbf{y}'\mathbf{P}\mathbf{P}\mathbf{y} \theta_0 &= \mathbf{y}'\mathbf{P}\mathbf{H}\mathbf{P}\mathbf{y} \\ &= \mathbf{y}'\mathbf{P}\mathbf{y}.\end{aligned}$$

To simplify these quadratic forms notice that

$$\begin{aligned}\mathbf{P}\mathbf{y} &= (\mathbf{H}^{-1} - \mathbf{H}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{H}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{H}^{-1})\mathbf{y} \\ &= \mathbf{H}^{-1}(\mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{H}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{H}^{-1}\mathbf{y}) \\ &= \mathbf{H}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \text{ and} \\ \hat{\mathbf{b}} &= (\mathbf{X}'\mathbf{H}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{H}^{-1}\mathbf{y}.\end{aligned}$$

Therefore,

$$\mathbf{y}'\mathbf{P}\mathbf{Z}_i\mathbf{Z}_i'\mathbf{P}\mathbf{y} = (\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})'\mathbf{H}^{-1}\mathbf{Z}_i\mathbf{Z}_i'\mathbf{H}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}).$$

Henderson (1973) noticed that

$$\hat{\mathbf{u}}_i = \mathbf{G}_i\mathbf{Z}_i'\mathbf{P}\mathbf{y}$$

where  $\mathbf{G}_i$  is the assumed variance covariance matrix of  $\mathbf{u}_i$  (in this case  $\mathbf{G}_i = \mathbf{I} \theta_i$ ), and  $\hat{\mathbf{u}}_i$  is BLUP of  $\mathbf{u}_i$  and the solution of  $\hat{\mathbf{u}}_i$  from the MME, then

$$\begin{aligned}\mathbf{y}'\mathbf{P}\mathbf{Z}_i\mathbf{Z}_i'\mathbf{P}\mathbf{y} &= \mathbf{y}'\mathbf{P}\mathbf{Z}_i[\mathbf{G}_i\mathbf{G}_i^{-1}\mathbf{G}_i^{-1}\mathbf{G}_i]\mathbf{Z}_i'\mathbf{P}\mathbf{y} \\ &= (\mathbf{y}'\mathbf{P}\mathbf{Z}_i\mathbf{G}_i)\mathbf{G}_i^{-2}(\mathbf{G}_i\mathbf{Z}_i'\mathbf{P}\mathbf{y}) \\ &= \hat{\mathbf{u}}_i'\mathbf{G}_i^{-2}\hat{\mathbf{u}}_i \\ &= \hat{\mathbf{u}}_i'\hat{\mathbf{u}}_i/\theta_i^2 \text{ when } \mathbf{G}_i = \mathbf{I} \theta_i.\end{aligned}$$

Similarly, Henderson showed that

$$\mathbf{y}'\mathbf{P}\mathbf{P}\mathbf{y} = (\mathbf{y}'\mathbf{y} - \hat{\mathbf{b}}'\mathbf{X}'\mathbf{y} - \sum_{i=1}^s \hat{\mathbf{u}}_i'\mathbf{Z}_i'\mathbf{y} - \sum_{i=1}^s \hat{\mathbf{u}}_i'\hat{\mathbf{u}}_i\alpha_i)/\theta_0$$

where  $\alpha_i = \theta_0/\theta_i$ .

Thus, under the basic linear model for variance component estimation the necessary quadratic forms for MIVQUE are

$$\hat{\mathbf{u}}_i'\hat{\mathbf{u}}_i \text{ and } \mathbf{y}'\mathbf{y} - \hat{\mathbf{b}}'\mathbf{X}'\mathbf{y} - \sum_{i=1}^s \hat{\mathbf{u}}_i'\mathbf{Z}_i'\mathbf{y}.$$

## Expectations

The MME are

$$\begin{pmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z}_1 & \mathbf{X}'\mathbf{Z}_2 & \dots & \mathbf{X}'\mathbf{Z}_s \\ \mathbf{Z}'_1\mathbf{X} & \mathbf{Z}'_1\mathbf{Z}_1 + \mathbf{I}\alpha_1 & \mathbf{Z}'_1\mathbf{Z}_2 & \dots & \mathbf{Z}'_1\mathbf{Z}_s \\ \mathbf{Z}'_2\mathbf{X} & \mathbf{Z}'_2\mathbf{Z}_1 & \mathbf{Z}'_2\mathbf{Z}_2 + \mathbf{I}\alpha_2 & \dots & \mathbf{Z}'_2\mathbf{Z}_s \\ \vdots & \vdots & \vdots & & \vdots \\ \mathbf{Z}'_s\mathbf{X} & \mathbf{Z}'_s\mathbf{Z}_1 & \mathbf{Z}'_s\mathbf{Z}_2 & \dots & \mathbf{Z}'_s\mathbf{Z}_s + \mathbf{I}\alpha_s \end{pmatrix} \begin{pmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}}_1 \\ \hat{\mathbf{u}}_2 \\ \vdots \\ \hat{\mathbf{u}}_s \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'_1\mathbf{y} \\ \mathbf{Z}'_2\mathbf{y} \\ \vdots \\ \mathbf{Z}'_s\mathbf{y} \end{pmatrix},$$

and let a generalized inverse of the coefficient matrix be represented as

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_{xx} & \mathbf{C}_{x1} & \mathbf{C}_{x2} & \dots & \mathbf{C}_{xs} \\ \mathbf{C}_{1x} & \mathbf{C}_{11} & \mathbf{C}_{12} & \dots & \mathbf{C}_{1s} \\ \mathbf{C}_{2x} & \mathbf{C}_{21} & \mathbf{C}_{22} & \dots & \mathbf{C}_{2s} \\ \vdots & \vdots & \vdots & & \vdots \\ \mathbf{C}_{sx} & \mathbf{C}_{s1} & \mathbf{C}_{s2} & \dots & \mathbf{C}_{ss} \end{pmatrix} = \begin{pmatrix} \mathbf{C}_x \\ \mathbf{C}_1 \\ \mathbf{C}_2 \\ \vdots \\ \mathbf{C}_s \end{pmatrix}.$$

Now let  $\mathbf{r} = \mathbf{W}'\mathbf{y}$ , the right hand sides of the MME, then

$$\begin{aligned} \text{Var}(\mathbf{r}) &= \mathbf{W}'\text{Var}(\mathbf{y})\mathbf{W} \\ &= \sum_{j=1}^s \mathbf{W}'\mathbf{Z}_j\mathbf{Z}'_j\mathbf{W} \sigma_j^2 + \mathbf{W}'\mathbf{W} \sigma_0^2 \\ &= \sum_{j=0}^s \mathbf{U}_j\sigma_j^2. \end{aligned}$$

Also,

$$\hat{\mathbf{u}}_i = \mathbf{C}_i\mathbf{r},$$

and,

$$\hat{\mathbf{u}}_i'\hat{\mathbf{u}}_i = \mathbf{r}'\mathbf{C}'_i\mathbf{C}_i\mathbf{r}.$$

The expectation of  $\hat{\mathbf{u}}_i'\hat{\mathbf{u}}_i$  is

$$E(\hat{\mathbf{u}}_i'\hat{\mathbf{u}}_i) = \sum_{j=0}^s \text{tr}(\mathbf{C}'_i\mathbf{C}_i\mathbf{U}_j) \sigma_j^2.$$

This expectation can be simplified by noting that (for  $\mathbf{X}$  having full rank)

$$\mathbf{C}\mathbf{W}'\mathbf{W} = \begin{pmatrix} \mathbf{I}_x & -\mathbf{C}_{x1}\alpha_1 & -\mathbf{C}_{x2}\alpha_2 & \dots & -\mathbf{C}_{xs}\alpha_s \\ \mathbf{0} & \mathbf{I} - \mathbf{C}_{11}\alpha_1 & -\mathbf{C}_{12}\alpha_2 & \dots & -\mathbf{C}_{1s}\alpha_s \\ \mathbf{0} & -\mathbf{C}_{21}\alpha_1 & \mathbf{I} - \mathbf{C}_{22}\alpha_2 & \dots & -\mathbf{C}_{2s}\alpha_s \\ \vdots & \vdots & \vdots & & \vdots \\ \mathbf{0} & -\mathbf{C}_{s1}\alpha_1 & -\mathbf{C}_{s2}\alpha_2 & \dots & \mathbf{I} - \mathbf{C}_{ss}\alpha_s \end{pmatrix}$$

then, for example, for  $i$  not equal to  $j$ ,

$$\begin{aligned}
tr(\mathbf{C}'_i \mathbf{C}_i \mathbf{U}_j) &= tr(\mathbf{C}'_i \mathbf{C}_i \mathbf{W}' \mathbf{Z}_j \mathbf{Z}'_j \mathbf{W}) \\
&= tr[ (\mathbf{Z}'_j \mathbf{W} \mathbf{C}'_i) (\mathbf{C}_i \mathbf{W}' \mathbf{Z}_j) ] \\
&= tr[ (-\mathbf{C}'_{ij} \alpha_j) (-\mathbf{C}_{ij} \alpha_j) ] \\
&= tr(\mathbf{C}_{ji} \mathbf{C}_{ij}) \alpha_j^2.
\end{aligned}$$

The expected value of  $\hat{\mathbf{u}}'_i \hat{\mathbf{u}}_i$  can be written as

$$\begin{aligned}
E(\hat{\mathbf{u}}'_i \hat{\mathbf{u}}_i) &= (q_i - 2 tr \mathbf{C}_{ii} \alpha_i + tr \mathbf{C}_{ii}^2 \alpha_i^2) \sigma_i^2 \\
&\quad + \sum_{j \neq i}^s tr(\mathbf{C}'_{ij} \mathbf{C}_{ij} \alpha_j^2) \sigma_j^2 \\
&\quad + (tr \mathbf{C}_{ii} - tr \mathbf{C}_{ii}^2 \alpha_i - \sum_{j \neq i}^s tr \mathbf{C}'_{ij} \mathbf{C}_{ij} \alpha_j) \sigma_0^2
\end{aligned}$$

In the same way,

$$E(\mathbf{y}'\mathbf{y}) = \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b} + N \sum_{j=0}^s \sigma_j^2,$$

and

$$\hat{\mathbf{b}}'\mathbf{X}'\mathbf{y} + \sum_{i=1}^s \hat{\mathbf{u}}'_i \mathbf{Z}'_i \mathbf{y} = \mathbf{r}'\mathbf{C}\mathbf{r}.$$

Then

$$E(\mathbf{r}'\mathbf{C}\mathbf{r}) = \mathbf{b}'\mathbf{X}'\mathbf{W}\mathbf{C}\mathbf{W}'\mathbf{X}\mathbf{b} + \sum_{j=0}^s tr \mathbf{C}\mathbf{U}_j \sigma_j^2,$$

but

$$\mathbf{X}'\mathbf{W}\mathbf{C}\mathbf{W}'\mathbf{X} = \mathbf{X}'\mathbf{X}$$

so that

$$\begin{aligned}
E(\mathbf{y}'\mathbf{y} - \mathbf{r}'\mathbf{C}\mathbf{r}) &= \sum_{j=0}^s (N - tr \mathbf{C}\mathbf{U}_j) \sigma_j^2 \\
&= \sum_{j=1}^s (q_j \alpha_j - tr \mathbf{C}_{jj} \alpha_j^2) \sigma_j^2 \\
&\quad + (N - r(\mathbf{X}) - \sum_{i=1}^s (q_i - tr \mathbf{C}_{ii} \alpha_i)) \sigma_0^2.
\end{aligned}$$

MIVQUE requires the inverse elements of the MME in order to calculate the expectations of the quadratic forms. In most animal breeding applications, obtaining an inverse of the coefficient matrix could be impossible due to the large order of the equations.

**Example**

Assume a model with one fixed factor and one random factor. The least squares equations are given below, and  $\mathbf{y}'\mathbf{y} = 770$ .

$$\begin{pmatrix} 12 & 0 & 4 & 3 & 5 \\ 0 & 8 & 0 & 6 & 2 \\ 4 & 0 & 4 & 0 & 0 \\ 3 & 6 & 0 & 9 & 0 \\ 5 & 2 & 0 & 0 & 7 \end{pmatrix} \begin{pmatrix} \hat{A}_1 \\ \hat{A}_2 \\ \hat{D}_1 \\ \hat{D}_2 \\ \hat{D}_3 \end{pmatrix} = \begin{pmatrix} 63 \\ 52 \\ 28 \\ 57 \\ 30 \end{pmatrix}.$$

Let factor D be random and assume that  $\alpha_D = \theta_0/\theta_D = 2$ , then to form the MME add 2 to the diagonals of the equations for factor D. The variance-covariance matrix of the right hand sides is

$$V(\mathbf{r}) = \mathbf{U}_0\sigma_0^2 + \mathbf{U}_D\sigma_D^2$$

where  $\mathbf{U}_0 = \mathbf{W}'\mathbf{W}$ , the LS equations, and  $\mathbf{U}_D = \mathbf{W}'\mathbf{Z}_D\mathbf{Z}_D'\mathbf{W}$ , that is

$$\begin{aligned} \mathbf{U}_D &= \begin{pmatrix} 4 & 3 & 5 \\ 0 & 6 & 2 \\ 4 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 7 \end{pmatrix} \begin{pmatrix} 4 & 0 & 4 & 0 & 0 \\ 3 & 6 & 0 & 9 & 0 \\ 5 & 2 & 0 & 0 & 7 \end{pmatrix} \\ &= \begin{pmatrix} 50 & 28 & 16 & 27 & 35 \\ 28 & 40 & 0 & 54 & 14 \\ 16 & 0 & 16 & 0 & 0 \\ 27 & 54 & 0 & 81 & 0 \\ 35 & 14 & 0 & 0 & 49 \end{pmatrix}. \end{aligned}$$

With  $\alpha_D = 2$  then

$$\mathbf{C} = \begin{pmatrix} .25158 & .16139 & -.16772 & -.15665 & -.17563 \\ .16139 & .33703 & -.10759 & -.22785 & -.16456 \\ -.16772 & -.10759 & .27848 & .10443 & .11709 \\ -.15665 & -.22785 & .10443 & .25791 & .13766 \\ -.17563 & -.16456 & .11709 & .13766 & .24525 \end{pmatrix}$$

and the solutions to the mixed model equations are

$$\mathbf{C}\mathbf{r} = \begin{pmatrix} \hat{A}_1 \\ \hat{A}_2 \\ \hat{D}_1 \\ \hat{D}_2 \\ \hat{D}_3 \end{pmatrix} = \begin{pmatrix} 5.3481 \\ 6.7563 \\ 1.1013 \\ .0380 \\ -1.1392 \end{pmatrix}.$$

The quadratic forms are

$$\hat{\mathbf{u}}_D'\hat{\mathbf{u}}_D = \hat{D}_1^2 + \hat{D}_2^2 + \hat{D}_3^2$$

$$\begin{aligned}
&= (1.1013)^2 + (.0380)^2 + (-1.1392)^2 \\
&= 2.5121 \\
&= \mathbf{r}'\mathbf{C}'_D\mathbf{C}_D\mathbf{r}
\end{aligned}$$

where

$$\mathbf{C}_D = \begin{pmatrix} -.16772 & -.10759 & .27848 & .10443 & .11709 \\ -.15665 & -.22785 & .10443 & .25791 & .13766 \\ -.17563 & -.16456 & .11709 & .13766 & .24525 \end{pmatrix}$$

$$\mathbf{C}'_D\mathbf{C}_D = \begin{pmatrix} .08352 & .08264 & -.08363 & -.08209 & -.08428 \\ .08264 & .09057 & -.07303 & -.09265 & -.08432 \\ -.08363 & -.07303 & .10217 & .07213 & .07570 \\ -.08209 & -.09265 & .07213 & .09637 & .08149 \\ -.08428 & -.08432 & .07570 & .08149 & .09281 \end{pmatrix}.$$

The other quadratic form is

$$\begin{aligned}
\hat{\mathbf{b}}'\mathbf{X}'\mathbf{y} + \hat{\mathbf{u}}'_D\mathbf{Z}'_D\mathbf{y} &= 687.0823 \\
&= 5.3481(63) + 6.7563(52) + 1.1013(28) \\
&\quad + .0380(57) - 1.1392(30) \\
&= \mathbf{r}'\mathbf{C}\mathbf{r}.
\end{aligned}$$

Thus,

$$\mathbf{y}'\mathbf{y} - \mathbf{r}'\mathbf{C}\mathbf{r} = 770 - 687.0823 = 82.9177.$$

The expectations of these quadratic forms are as follows:

$$\begin{aligned}
E(\mathbf{r}'\mathbf{C}'_D\mathbf{C}_D\mathbf{r}) &= E(\hat{\mathbf{u}}'_D\hat{\mathbf{u}}_D) \\
&= \text{tr}(\mathbf{C}'_D\mathbf{C}_D\mathbf{U}_D)\sigma_D^2 + \text{tr}(\mathbf{C}'_D\mathbf{C}_D\mathbf{U}_0)\sigma_0^2 \\
&= 1.038816 \sigma_D^2 + .198946 \sigma_0^2
\end{aligned}$$

also note that

$$\begin{aligned}
\text{tr}(\mathbf{C}'_D\mathbf{C}_D\mathbf{U}_D) &= (q_D - 2 \text{tr}\mathbf{C}_{DD}\alpha_D + \text{tr}\mathbf{C}_{DD}^2\alpha_D^2) \\
&= 3 - 2(.78164)(2) + (.29135)(4) \\
&= 1.038816
\end{aligned}$$

and

$$\begin{aligned}
\text{tr}(\mathbf{C}'_D\mathbf{C}_D\mathbf{U}_0) &= (\text{tr}\mathbf{C}_{DD} - \text{tr}\mathbf{C}_{DD}^2\alpha_D) \\
&= (.78164) - (.29135)(2) = .198946.
\end{aligned}$$

Similarly,

$$\begin{aligned}
E(\mathbf{y}'\mathbf{y} - \mathbf{r}'\mathbf{C}\mathbf{r}) &= (N - \text{tr}\mathbf{C}\mathbf{U}_D) \sigma_D^2 + (N - \text{tr}\mathbf{C}\mathbf{U}_0) \sigma_0^2 \\
&= 2.873418 \sigma_D^2 + 16.563291 \sigma_0^2
\end{aligned}$$

where

$$\begin{aligned} N - tr\mathbf{C}\mathbf{U}_D &= q_D\alpha_D - tr\mathbf{C}_{DD}\alpha_D^2 \\ &= 3(2) - (.78164)(4) \\ &= 2.873418, \end{aligned}$$

and

$$\begin{aligned} N - tr\mathbf{C}\mathbf{U}_0 &= N - r(\mathbf{X}) - (q_D - tr\mathbf{C}_{DD}\alpha_D) \\ &= 20 - 2 - (3 - .78164(2)) \\ &= 16.563291. \end{aligned}$$

Equate the quadratic forms to their expectations and solve for the estimates of variance components.

$$\begin{aligned} \begin{pmatrix} \hat{\sigma}_D^2 \\ \hat{\sigma}_0^2 \end{pmatrix} &= \begin{pmatrix} 1.038816 & .198946 \\ 2.873418 & 16.563291 \end{pmatrix}^{-1} \begin{pmatrix} 2.5121 \\ 82.9177 \end{pmatrix} \\ &= \begin{pmatrix} 1.509653 \\ 4.744218 \end{pmatrix}. \end{aligned}$$

## Problem

Below are the OLS equations for some data for a model with two random factors (B and C) and one fixed factor (A). The total number of observations was 200, and the total sum of squares was 1,000,000. Assume that the prior variance ratios were

$$\alpha_B = 9, \text{ and } \alpha_C = 5.$$

Apply MIVQUE and one of the previous methods to this example data.

$$\begin{pmatrix} 25 & 0 & 0 & 5 & 10 & 10 & 0 & 10 & 0 & 3 & 12 & 0 \\ 0 & 100 & 0 & 80 & 5 & 0 & 15 & 23 & 17 & 7 & 40 & 13 \\ 0 & 0 & 75 & 15 & 35 & 20 & 5 & 47 & 3 & 0 & 8 & 17 \\ 5 & 80 & 15 & 100 & 0 & 0 & 0 & 49 & 9 & 3 & 18 & 21 \\ 10 & 5 & 35 & 0 & 50 & 0 & 0 & 22 & 0 & 4 & 24 & 0 \\ 10 & 0 & 20 & 0 & 0 & 30 & 0 & 9 & 7 & 0 & 5 & 9 \\ 0 & 15 & 5 & 0 & 0 & 0 & 20 & 0 & 4 & 3 & 13 & 0 \\ 10 & 23 & 47 & 49 & 22 & 9 & 0 & 80 & 0 & 0 & 0 & 0 \\ 0 & 17 & 3 & 9 & 0 & 7 & 4 & 0 & 20 & 0 & 0 & 0 \\ 3 & 7 & 0 & 3 & 4 & 0 & 3 & 0 & 0 & 10 & 0 & 0 \\ 12 & 40 & 8 & 18 & 24 & 5 & 13 & 0 & 0 & 0 & 60 & 0 \\ 0 & 13 & 17 & 21 & 0 & 9 & 0 & 0 & 0 & 0 & 0 & 30 \end{pmatrix} \begin{pmatrix} \hat{A}_1 \\ \hat{A}_2 \\ \hat{A}_3 \\ \hat{B}_1 \\ \hat{B}_2 \\ \hat{B}_3 \\ \hat{B}_4 \\ \hat{C}_1 \\ \hat{C}_2 \\ \hat{C}_3 \\ \hat{C}_4 \\ \hat{C}_5 \end{pmatrix} = \begin{pmatrix} 875 \\ 3000 \\ 1125 \\ 2000 \\ 1500 \\ 1200 \\ 300 \\ 2200 \\ 800 \\ 50 \\ 1600 \\ 350 \end{pmatrix}.$$

# REML

L. R. Schaeffer, April 28, 1999

Restricted maximum likelihood (REML), was first suggested by Thompson (1962), but was described formally by Patterson and Thompson (1971).

1.  $\mathbf{y}$  must have a multivariate normal distribution.
2. The method is translation invariant.
3. The estimator is restricted to the allowable parameter space.
4. REML is a biased procedure.
5. The exact same quadratic forms as MIVQUE are utilized.

There are several methods of obtaining REML estimators. One method is to take the MIVQUE quadratic forms and derive new expectations assuming that the prior values are equal to the true parameter values. REML estimates can also be obtained by iterating on the MIVQUE quadratic forms and their expectations. The MIVQUE estimates are used to form new prior values of the variance ratios and apply MIVQUE over and over until the estimates and prior values are equal. The final estimates are REML if the variance estimates remain positive. If any of the estimates become negative, then the iteration process must be stopped and REML estimates can not be obtained in that manner.

## Derivation of REML from MIVQUE

From MIVQUE, the expectation of quadratic forms of solutions to mixed model equations were as follows:

$$\begin{aligned} E(\hat{\mathbf{u}}_i' \hat{\mathbf{u}}_i) &= (q_i - 2 \operatorname{tr} \mathbf{C}_{ii} \alpha_i + \operatorname{tr} \mathbf{C}_{ii}^2 \alpha_i^2) \sigma_i^2 \\ &\quad + \sum_{j \neq i}^s \operatorname{tr} \mathbf{C}'_{ij} \mathbf{C}_{ij} \alpha_j^2 \sigma_j^2 \\ &\quad + (\operatorname{tr} \mathbf{C}_{ii} - \sum_{j=1}^s \operatorname{tr} \mathbf{C}'_{ij} \mathbf{C}_{ij} \alpha_j) \sigma_o^2 \end{aligned}$$

where  $\alpha_i = \theta_o / \theta_i$  and  $\theta_o$  and  $\theta_i$  are assumed prior values of  $\sigma_o^2$  and  $\sigma_i^2$ . To derive REML assume that  $\alpha_i = \sigma_o^2 / \sigma_i^2$  and substitute into the above expectation to obtain

$$E(\hat{\mathbf{u}}_i' \hat{\mathbf{u}}_i) = q_i \sigma_i^2 - \operatorname{tr} \mathbf{C}_{ii} \sigma_o^2,$$

which gives

$$\hat{\sigma}_i^2 = (\hat{\mathbf{u}}_i' \hat{\mathbf{u}}_i + \operatorname{tr} \mathbf{C}_{ii} \hat{\sigma}_o^2) / q_i$$

Notice that  $\hat{\mathbf{u}}_i' \hat{\mathbf{u}}_i$  is a sum of squares and therefore always positive, and that  $\operatorname{tr} \mathbf{C}_{ii}$  is always positive because the diagonals of  $\mathbf{C}_{ii}$  represent the variances of prediction error

(i.e.  $Var(\hat{\mathbf{u}}_i - \mathbf{u}_i)$ ). Hence the estimator,  $\hat{\sigma}_i^2$  is always positive, i.e. within the parameter space.

Likewise,

$$E(\mathbf{y}'\mathbf{y} - \hat{\mathbf{b}}'\mathbf{X}'\mathbf{y} - \sum_{i=1}^s \hat{\mathbf{u}}_i'\mathbf{Z}'_i\mathbf{y}) = (N - r(\mathbf{X}))\sigma_o^2$$

under the assumption that  $\alpha_i = \sigma_o^2/\sigma_i^2$ . Then

$$\hat{\sigma}_o^2 = (\mathbf{y}'\mathbf{y} - \hat{\mathbf{b}}'\mathbf{X}'\mathbf{y} - \sum_{i=1}^s \hat{\mathbf{u}}_i'\mathbf{Z}'_i\mathbf{y})/(N - r(\mathbf{X})).$$

The above pseudo expectations give formulas for  $\hat{\sigma}_i^2$  and  $\hat{\sigma}_o^2$  that are known as the EM algorithm. EM stands for Expectation Maximization, a procedure described by Dempster et al. (1977). REML estimates are obtained by using  $\hat{\sigma}_o^2$  and  $\hat{\sigma}_i^2$  to form new ratios,  $\hat{\alpha}_i$ , then the new ratios are used as priors to calculate a new  $\hat{\mathbf{u}}_i$ , until the ratio used as a prior and the estimated ratio are equal. With the REML EM algorithm,

1. The user must iterate until convergence.
2. Convergence is not guaranteed.
3. The final estimate may not maximize the likelihood function over the entire range of parameter values. This would be convergence to a local maximum rather than a global maximum.
4. Convergence will normally occur if the number of observations is large.
5. With few observations and several random factors there may be problems in reaching convergence.

### An Example

Below are the mixed model equations for an example with three factors, two of which are random.

$$\begin{pmatrix} 50 & 0 & 5 & 15 & 30 & 5 & 10 & 20 & 15 \\ 0 & 40 & 5 & 15 & 20 & 5 & 10 & 20 & 5 \\ 5 & 5 & 10 & 0 & 0 & 2 & 3 & 4 & 1 \\ 15 & 15 & 0 & 30 & 0 & 5 & 7 & 11 & 7 \\ 30 & 20 & 0 & 0 & 50 & 3 & 10 & 25 & 12 \\ 5 & 5 & 2 & 5 & 3 & 10 & 0 & 0 & 0 \\ 10 & 10 & 3 & 7 & 10 & 0 & 20 & 0 & 0 \\ 20 & 20 & 4 & 11 & 25 & 0 & 0 & 40 & 0 \\ 15 & 5 & 1 & 7 & 12 & 0 & 0 & 0 & 20 \end{pmatrix} \begin{pmatrix} F_1 \\ F_2 \\ A_1 \\ A_2 \\ A_3 \\ B_1 \\ B_2 \\ B_3 \\ B_4 \end{pmatrix} = \begin{pmatrix} 3200 \\ 2380 \\ 580 \\ 1860 \\ 3140 \\ 700 \\ 1320 \\ 2400 \\ 1160 \end{pmatrix}$$

The total sum of squares is  $\mathbf{y}'\mathbf{y} = 356,000$ .

To demonstrate EM REML, let  $\alpha_A = 10$  and  $\alpha_B = 5$  be the starting values of the ratios for factors A and B, respectively. The solution vector is

$$\begin{pmatrix} F_1 \\ F_2 \\ A_1 \\ A_2 \\ A_3 \\ B_1 \\ B_2 \\ B_3 \\ B_4 \end{pmatrix} = \begin{pmatrix} 64.6313 \\ 59.4225 \\ -2.1363 \\ .4955 \\ 1.6368 \\ 5.1064 \\ 2.6402 \\ -2.6433 \\ -5.1034 \end{pmatrix}.$$

Then

$$\mathbf{y}'(\mathbf{X}\hat{\mathbf{b}} + \mathbf{Z}\hat{\mathbf{u}}) = 347,871.2661$$

and from the inverse of the coefficient matrix,

$$tr\mathbf{C}_{AA} = .16493, \text{ and } tr\mathbf{C}_{BB} = .3309886$$

which give rise to the following estimates,

$$\begin{aligned} \hat{\sigma}_0^2 &= (356,000 - 347,871.2661)/88 \\ &= 92.371976, \\ \hat{\sigma}_A^2 &= (7.4925463 + .16493(92.371976))/3 \\ &= 7.575855, \\ \hat{\sigma}_B^2 &= (66.0771576 + .3309886(92.371976))/4 \\ &= 24.16280774. \end{aligned}$$

New ratios are formed as

$$\alpha_A = 92.371976/7.575855 = 12.192944,$$

and

$$\alpha_B = 92.371976/24.16280774 = 3.822899$$

and these are used to form the mixed model equations again, new solutions and traces are calculated, and so on, until the estimated ratios and the prior values of the ratios are equal. The estimates in this example converge to

$$\begin{aligned} \hat{\sigma}_0^2 &= 91.8639, \\ \hat{\sigma}_A^2 &= 2.5692, \end{aligned}$$

and

$$\hat{\sigma}_B^2 = 30.5190.$$

Derivation of REML from the likelihood function and a description of derivative free REML are given in the regular notes for the course 10-638 Animal Models.

# Maximum Likelihood

L. R. Schaeffer, April 28, 1999

Hartley and Rao (1967) described the maximum likelihood approach for the estimation of variance components.

1. ML uses the log likelihood function of  $\mathbf{y}$ .
2. ML uses the exact same quadratic forms as REML and MIVQUE.
3. ML is translation invariant, but biased.

Henderson(1973) derived ML estimates from mixed model equations, and ML is a more biased method than REML especially the estimator of the residual variance. Estimates of the residual variance are generally smaller with ML than with REML due to the difference in degrees of freedom. REML was proposed in order to account for the loss in degrees of freedom in ML due to estimating the fixed effects. The main disadvantages of ML are the same as for REML in that the procedure is iterative and convergence is not guaranteed. An advantage of ML over REML is that in some models the traces that need to be calculated are easier than those needed for REML (Schaeffer 1976).

## Derivation

Assume that  $\mathbf{y}$  is normally distributed with mean  $\mathbf{Xb}$  and variance-covariance matrix  $\mathbf{V}$  and follows the basic linear mixed model. Except for a constant, the log likelihood function of  $\mathbf{y}$  is

$$L(\mathbf{y}) = -.5 \ln |\mathbf{V}| - .5 (\mathbf{y} - \mathbf{Xb})' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{Xb})$$

The derivatives of  $L(\mathbf{y})$  with respect to  $\mathbf{b}$  and to  $\sigma_i^2$  for  $i = 0, 1, \dots, s$  are

$$\begin{aligned} \partial L(\mathbf{y}) / \partial \mathbf{b} &= \mathbf{X}' \mathbf{V}^{-1} \mathbf{Xb} - \mathbf{X}' \mathbf{V}^{-1} \mathbf{y} \\ \partial L(\mathbf{y}) / \partial \sigma_i^2 &= -.5 \operatorname{tr}[\mathbf{V}^{-1} (\partial \mathbf{V} / \partial \sigma_i^2)] \\ &\quad + .5 (\mathbf{y} - \mathbf{Xb})' \mathbf{V}^{-1} (\partial \mathbf{V} / \partial \sigma_i^2) \mathbf{V}^{-1} (\mathbf{y} - \mathbf{Xb}) \\ &= -.5 \operatorname{tr}[\mathbf{V}^{-1} \mathbf{V}_i] + .5 (\mathbf{y} - \mathbf{Xb})' \mathbf{V}^{-1} \mathbf{V}_i \mathbf{V}^{-1} (\mathbf{y} - \mathbf{Xb}) \end{aligned}$$

Equating the derivatives to zero gives

$$\begin{aligned} \hat{\mathbf{b}} &= (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1} \mathbf{y}, \\ \text{and } \operatorname{tr}[\mathbf{V}^{-1} \mathbf{V}_i] &= (\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})' \mathbf{V}^{-1} \mathbf{V}_i \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}). \end{aligned}$$

Recall that  $\mathbf{Py} = \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})$ , where  $\mathbf{P}$  is the projection matrix, and that  $\mathbf{V}_i = \mathbf{Z}_i \mathbf{Z}_i'$ , then the latter equation becomes

$$\begin{aligned} \operatorname{tr}[\mathbf{V}^{-1} \mathbf{Z}_i \mathbf{Z}_i'] &= \mathbf{y}' \mathbf{P} \mathbf{V}_i \mathbf{P} \mathbf{y} \\ &= \hat{\mathbf{u}}_i' \hat{\mathbf{u}}_i / \sigma_i^4. \end{aligned}$$

Notice that the right hand side of the above equation is the same as in the REML and MIVQUE derivations. However, the difference among the methods is in the left hand side of this equation. For REML, we have  $tr[\mathbf{P}\mathbf{V}_i]$  and for MIVQUE it is  $tr[\mathbf{PVP}\mathbf{V}_i]$ , and for ML it is  $tr[\mathbf{V}^{-1}\mathbf{V}_i]$ . For ML, the left hand can be manipulated as

$$tr[\mathbf{V}^{-1}\mathbf{V}_i] = tr[(\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1})\mathbf{Z}_i\mathbf{Z}'_i]$$

and if

$$\begin{aligned}\mathbf{T} &= (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}, \\ \mathbf{R} &= \mathbf{I}\sigma_0^2, \\ \text{and } \mathbf{G} &= \sum^+ \mathbf{I}\sigma_i^2,\end{aligned}$$

then

$$tr[\mathbf{V}^{-1}\mathbf{V}_i] = tr(\mathbf{Z}'_i\mathbf{Z}_i)\sigma_0^{-2} - tr(\mathbf{Z}'_i\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{Z}_i)\sigma_0^{-4}.$$

If  $\mathbf{T}$  can be partitioned into submatrices for each random factor, then

$$\mathbf{T}\sigma_0^{-2}(\mathbf{Z}'\mathbf{Z} + \sum^+ \mathbf{I}\alpha_i) = \mathbf{I},$$

and

$$\mathbf{T}\mathbf{Z}'\mathbf{Z}\sigma_0^{-2} = \mathbf{I} - \mathbf{T}(\sum^+ \mathbf{I}\sigma_i^{-2}).$$

then

$$tr(\mathbf{Z}'_i\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{Z}_i)\sigma_0^{-4} = tr(\mathbf{Z}'_i\mathbf{Z}_i)\sigma_0^{-2} - tr(\mathbf{I} - \mathbf{T}_{ii}\sigma_i^{-2})\sigma_i^{-2}.$$

Finally,

$$\begin{aligned}tr[\mathbf{V}^{-1}\mathbf{V}_i] &= tr(\mathbf{I} - \mathbf{T}_{ii}\sigma_i^{-2})\sigma_i^{-2} \\ &= tr\mathbf{I}\sigma_i^{-2} - tr\mathbf{T}_{ii}\sigma_i^{-4} \\ &= q_i\sigma_i^{-2} - tr\mathbf{T}_{ii}\sigma_i^{-4}.\end{aligned}$$

Combining results gives

$$\hat{\sigma}_i^2 = (\hat{\mathbf{u}}'_i\hat{\mathbf{u}}_i + tr\mathbf{T}_{ii}\hat{\sigma}_0^2)/q_i$$

for  $i = 1, 2, \dots, s$ , and for  $i = 0$  gives

$$\hat{\sigma}_0^2 = (\mathbf{y}'\mathbf{y} - \hat{\mathbf{b}}'\mathbf{X}'\mathbf{y} - \hat{\mathbf{u}}'\mathbf{Z}'\mathbf{y})/N.$$

There are two major differences in these formulas from those for REML, namely,

1.  $tr\mathbf{T}_{ii}\hat{\sigma}_0^2$  in ML versus  $tr\mathbf{C}_{ii}\hat{\sigma}_0^2$  in REML in  $\hat{\sigma}_i^2$  formula.

2.  $N$  in ML versus  $(N - r(\mathbf{X}))$  in REML as denominator of  $(\mathbf{y}'\mathbf{y} - \hat{\mathbf{b}}'\mathbf{X}'\mathbf{y} - \hat{\mathbf{u}}'\mathbf{Z}'\mathbf{y})$  in  $\hat{\sigma}_0^2$  formula.

Note that

$$\mathbf{T} = (\mathbf{Z}'\mathbf{Z} + \sum^+ \mathbf{I}\alpha_i)^{-1}$$

and

$$\mathbf{C} = (\mathbf{Z}'\mathbf{M}\mathbf{Z} + \sum^+ \mathbf{I}\alpha_i)^{-1}.$$

ML, therefore, does not account for the degrees of freedom needed to estimate  $\mathbf{b}$ , as reflected by  $\mathbf{T}$  and by  $N$  in the above formulas.

### Example With One Random Factor

The computations for ML become fairly simple with only one random factor in the model. In that case,  $\mathbf{T}$  is a diagonal matrix and its elements can be readily computed. To illustrate, consider the least squares equations below for a model with fixed herd effects,  $H$ , and random sire effects,  $S$ , in an analysis of milking speed observations.

$$\begin{pmatrix} 8 & 0 & 2 & 3 & 1 & 2 & 0 \\ 0 & 14 & 3 & 3 & 2 & 4 & 2 \\ 2 & 3 & 5 & 0 & 0 & 0 & 0 \\ 3 & 3 & 0 & 6 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 & 3 & 0 & 0 \\ 2 & 4 & 0 & 0 & 0 & 6 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} \hat{H}_1 \\ \hat{H}_2 \\ \hat{S}_1 \\ \hat{S}_2 \\ \hat{S}_3 \\ \hat{S}_4 \\ \hat{S}_5 \end{pmatrix} = \begin{pmatrix} 36.4 \\ 67.2 \\ 26.5 \\ 30.0 \\ 14.1 \\ 26.2 \\ 6.8 \end{pmatrix}$$

Also let  $\mathbf{y}'\mathbf{y} = 500$  and  $\alpha_s = 6$ , then the solutions are

$$\begin{aligned} \hat{H}_1 &= 4.44037, & \hat{H}_2 &= 4.78680, & \hat{S}_1 &= .29626, \\ \hat{S}_2 &= .19321, & \hat{S}_3 &= .009559, & \hat{S}_4 &= -.15233, \\ \text{and } \hat{S}_5 &= -.34670, \end{aligned}$$

The following quantities are easily calculated.

$$\begin{aligned} \mathbf{y}'\mathbf{y} - \hat{\mathbf{b}}'\mathbf{X}'\mathbf{y} - \hat{\mathbf{u}}'\mathbf{Z}'\mathbf{y} &= 500 - 490.7357068 = 9.264293, \\ \hat{\mathbf{u}}'_s\hat{\mathbf{u}}_s &= .268595, \\ \hat{\sigma}_0^2 &= 9.264293 / 22 = .421104, \\ tr\mathbf{T}_{ss} &= \sum_{i=1}^{q_s} (n_i + \alpha_s)^{-1}, \\ &= (5 + 6)^{-1} + (6 + 6)^{-1} + (3 + 6)^{-1} \\ &\quad + (6 + 6)^{-1} + (2 + 6)^{-1} \\ &= .568493, \text{ and} \\ \hat{\sigma}_s^2 &= (.268595 + .568493 (.421104))/5 = .101598. \end{aligned}$$

In this example,  $\mathbf{T}_{ss} = (\mathbf{Z}'_s \mathbf{Z}_s + \mathbf{I}\alpha_s)^{-1}$ , and therefore must be multiplied by  $\hat{\sigma}_0^2$  in the computation of  $\hat{\sigma}_s^2$ . This gives a new variance ratio of  $\hat{\alpha}_s = 4.144807$ . These steps would be repeated, many times, until the estimates converge to the values that maximize the likelihood function. ML calculations for a model with only one random factor are very simple even when the random factor has many levels.

### One Random Factor Nested Within Another

In a two random factor model we have

$$\mathbf{T} = \begin{pmatrix} \mathbf{Z}'_1 \mathbf{Z}_1 + \mathbf{I}\alpha_1 & \mathbf{Z}'_1 \mathbf{Z}_2 \\ \mathbf{Z}'_2 \mathbf{Z}_1 & \mathbf{Z}'_2 \mathbf{Z}_2 + \mathbf{I}\alpha_2 \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{T}_{11} & \mathbf{T}_{12} \\ \mathbf{T}_{21} & \mathbf{T}_{22} \end{pmatrix}$$

and both  $tr\mathbf{T}_{11}$  and  $tr\mathbf{T}_{22}$  are needed by ML. However, if factor 2 is nested within factor 1, then we know that the columns of  $\mathbf{Z}_1$  can be formed from adding together appropriate columns of  $\mathbf{Z}_2$ . Thus,  $\mathbf{T}$  can be written entirely in terms of  $\mathbf{Z}_2$ , and the matrix that describes which columns of  $\mathbf{Z}_2$  to add together to give  $\mathbf{Z}_1$ . Call this matrix  $\mathbf{F}$  where  $\mathbf{Z}_1 = \mathbf{Z}_2 \mathbf{F}$ . Now  $\mathbf{T}$  can be rewritten as

$$\mathbf{T} = \begin{pmatrix} \mathbf{F}' \mathbf{Z}'_2 \mathbf{Z}_2 \mathbf{F} + \mathbf{I}\alpha_1 & \mathbf{F}' \mathbf{Z}'_2 \mathbf{Z}_2 \\ \mathbf{Z}'_2 \mathbf{Z}_2 \mathbf{F} & \mathbf{Z}'_2 \mathbf{Z}_2 + \mathbf{I}\alpha_2 \end{pmatrix}^{-1}$$

and by using partitioned matrix techniques, then

$$\mathbf{T}_{11} = \left( \mathbf{F}' (\mathbf{Z}'_2 \mathbf{Z}_2 - \mathbf{Z}'_2 \mathbf{Z}_2 (\mathbf{Z}'_2 \mathbf{Z}_2 + \mathbf{I}\alpha_2)^{-1} \mathbf{Z}'_2 \mathbf{Z}_2) \mathbf{F} + \mathbf{I}\alpha_1 \right)^{-1}$$

and

$$\begin{aligned} tr\mathbf{T}_{11} &= \sum_i \left( \sum_j (n_{ij} - n_{ij}^2 / (n_{ij} + \alpha_2)) + \alpha_1 \right)^{-1} \\ &= \sum_i p_i, \end{aligned}$$

where the  $i$  subscript is summing over factor 1 and the  $j$  subscript is summing over levels of factor 2, and  $n_{ij}$  is the number of observations for a particular subclass. Likewise,

$$\mathbf{T}_{22} = \mathbf{D} + \mathbf{D} \mathbf{Z}'_2 \mathbf{Z}_2 \mathbf{F} \mathbf{T}_{11} \mathbf{F}' \mathbf{Z}'_2 \mathbf{Z}_2 \mathbf{D}$$

where

$$\mathbf{D} = (\mathbf{Z}'_2 \mathbf{Z}_2 + \mathbf{I}\alpha_2)^{-1},$$

then

$$tr\mathbf{T}_{22} = \sum_i \left( \sum_j (n_{ij} + \alpha_2)^{-1} + p_i \sum_j (n_{ij}^2 / (n_{ij} + \alpha_2)^2) \right)$$

Both traces can be computed readily by proper sorting of the data files by levels of factor 2 within levels of factor 1. Schaeffer (1976) gives the details for these calculations.

To illustrate ML for this type of model, consider the previous milking speed example expanded to include cow effects (i.e. daughters within sires in the table below). Also, some cows are allowed to have more than one observation.

Table 2.2 Example 2-minute milk yield data on dairy cows

Cow	Sire	Herd	Number of Observations	Total 2-min yield(kg)
A	1	1	2	9.9
B	1	1	3	20.0
C	2	1	2	9.5
D	2	1	2	8.8
E	2	1	1	5.6
F	3	1	1	4.7
G	4	1	3	12.4
H	4	1	2	8.9
I	1	2	1	5.5
J	1	2	2	10.1
K	1	2	1	2.2
L	2	2	2	10.2
M	2	2	2	10.3
N	2	2	1	5.6
O	3	2	2	8.6
P	3	2	2	10.2
Q	4	2	2	7.8
R	4	2	2	8.1
S	4	2	2	8.2
T	4	2	1	4.3
U	5	2	2	7.0
V	5	2	2	8.6

The least squares equations for this example are of order 29. Assume that  $\mathbf{y}'\mathbf{y} = 1,000$  and that  $\alpha_s = 5$  and  $\alpha_c = 2$ , for sires and cows, respectively. To compute the traces that are needed, for each sire we have the following:

$$p_i = [ \sum_j (n_{ij} - n_{ij}^2 / (n_{ij} + \alpha_c)) + \alpha_s ]^{-1}$$

$$p_1 = [ (2 - 4/4) + (3 - 9/5) + (1 - 1/3) + 1 + 2/3 + 5 ]^{-1}$$

$$= .104895,$$

$$p_2 = [ 5.33333 + 5. ]^{-1} = .096774,$$

$$p_3 = .130435,$$

$$p_4 = .092025,$$

$$\text{and } p_5 = .142857.$$

$$\text{Thus, } tr\mathbf{T}_{11} = .104895 + .096774 + .130435 + .092025 + .142857$$

$$= .566986.$$

Now for each sire, compute

$$\begin{aligned}\lambda_i &= \sum_j (n_{ij} + \alpha_c)^{-1} + p_i \sum_j (n_{ij}^2 / (n_{ij} + \alpha_c)^2), \\ \lambda_1 &= (4^{-1} + 5^{-1} + 3^{-1} + 4^{-1} + 3^{-1}) + .104895 \\ &\quad (4/16 + 9/25 + 1/9 + 4/16 + 1/9) \\ &= 1.480187, \\ \lambda_2 &= 1.784946, \\ \lambda_3 &= .913044, \\ \lambda_4 &= 1.668712, \\ \text{and } \lambda_5 &= .571429.\end{aligned}$$

Summing gives

$$\begin{aligned}tr\mathbf{T}_{22} &= 1.480187 + 1.784946 + .913044 + 1.668712 + .571429 \\ &= 6.418318.\end{aligned}$$

The solutions for herds and sires from the mixed model equations were

$$\begin{aligned}\hat{H}_1 &= 4.914415, \quad \hat{H}_2 = 4.463194, \quad \hat{S}_1 = .2000877, \\ \hat{S}_2 &= .1799721, \quad \hat{S}_3 = .0431308, \quad \hat{S}_4 = -.262278, \\ \text{and } \hat{S}_5 &= -.160912.\end{aligned}$$

Solutions for cows are not shown. Other calculations gave

$$\begin{aligned}\mathbf{y}'\mathbf{y} - \hat{\mathbf{b}}'\mathbf{X}'\mathbf{y} - \hat{\mathbf{u}}'\mathbf{Z}'\mathbf{y} &= 1,000 - 886.97832 = 113.02168, \\ \hat{\mathbf{u}}'_s \hat{\mathbf{u}}_s &= .1689679, \\ \hat{\mathbf{u}}'_c \hat{\mathbf{u}}_c &= 2.4662926, \\ \hat{\sigma}_0^2 &= 113.02168 / 40 = 2.8255421 \\ \hat{\sigma}_s^2 &= (.1689679 + .566986 (2.8255421))/5 \\ &= .354202, \quad \text{and} \\ \hat{\sigma}_c^2 &= (2.4662926 + 6.418318 (2.8255421))/22 \\ &= .936433.\end{aligned}$$

The new variance ratios were then  $\alpha_s = 7.97721$  and  $\alpha_c = 3.017346$ . These new ratios would be used to begin another iteration.

### The EM Algorithm

EM stands for Expectation Maximization. From Searle, Casella, and McCulloch (1992) the following explanation is given. The procedure alternates between calculating conditional expected values and maximizing simplified likelihoods. The actual data  $\mathbf{y}$  are called the incomplete data in the EM algorithm, and the complete data are considered to be

$\mathbf{y}$  and the unobservable random effects,  $\mathbf{u}_i$ . If we knew the realized values of the unobservable random effects then their variance would be the average of their squared values, i.e.,

$$\hat{\sigma}_i^2 = \mathbf{u}_i' \mathbf{u}_i / q_i.$$

However, in real life we do not know the realized values of the random effects.

The steps of the EM algorithm are as follows:

**Step 0.** Decide on starting values for the variances and set  $m = 0$ .

**Step 1.(E-step)** Calculate the conditional expectation of the sufficient statistics, conditional on the incomplete data.

$$\begin{aligned} E(\mathbf{u}_i' \mathbf{u}_i \mid \mathbf{y}) &= \sigma_i^{4(m)} \mathbf{y}' \mathbf{P}^{(m)} \mathbf{Z}_i \mathbf{Z}_i' \mathbf{P}^{(m)} \mathbf{y} \\ &\quad + \text{tr}(\sigma_i^{2(m)} \mathbf{I} - \sigma_i^{4(m)} \mathbf{Z}_i' (\mathbf{V}^{(m)})^{-1} \mathbf{Z}_i) \\ &= \hat{t}_i^{(m)} \end{aligned}$$

**Step 2.(M-step)** Maximize the likelihood of the complete data,

$$\sigma_i^{2(m+1)} = \hat{t}_i^{(m)} / q_i, \quad i = 0, 1, 2, \dots, s.$$

**Step 3.** If convergence is reached, set  $\hat{\boldsymbol{\sigma}} = \boldsymbol{\sigma}^{(m+1)}$ , otherwise increase  $m$  by one and return to Step 1.

This is equivalent to constructing and solving the mixed model equations with a given set of variances,  $\boldsymbol{\sigma}^m$ , and then

$$\begin{aligned} \sigma_0^{2(m+1)} &= (\mathbf{y}' \mathbf{y} - \hat{\mathbf{b}}' \mathbf{X}' \mathbf{y} - \hat{\mathbf{u}}' \mathbf{Z}' \mathbf{y}) / N \\ \text{and } \sigma_i^{2(m+1)} &= (\hat{\mathbf{u}}_i' \hat{\mathbf{u}}_i + \sigma_0^{2(m+1)} \text{tr} \mathbf{T}_{ii}) / q_i \end{aligned}$$

The EM algorithm can be applied to REML estimation as well. Just replace  $\mathbf{y}$  with  $\mathbf{K}' \mathbf{y}$  and replace  $\mathbf{V}$  with  $\mathbf{K}' \mathbf{V} \mathbf{K}$  and note that

$$\begin{aligned} \mathbf{P} &= \mathbf{K}(\mathbf{K}' \mathbf{V} \mathbf{K})^{-1} \mathbf{K}' \\ &= \mathbf{V}^{-1} - \mathbf{V}^{-1} \mathbf{X}(\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1}. \end{aligned}$$

Then the conditional expectation in the E-step becomes

$$\hat{t}_i^{(m)} = \sigma_i^{4(m)} \mathbf{y}' \mathbf{P}^{(m)} \mathbf{Z}_i \mathbf{Z}_i' \mathbf{P}^{(m)} \mathbf{y} + \text{tr}[\sigma_i^{2(m)} \mathbf{I} - \sigma_i^{4(m)} \mathbf{Z}_i' \mathbf{P}^{(m)} \mathbf{Z}_i].$$

# Pseudo Expectation Method

L. R. Schaeffer, April 28, 1999

In the process of making up an exam question for students in my Variance Component Estimation course in 1986 I discovered a quadratic form whose pseudo expectation (following the method of deriving REML from MIVQUE expectations of quadratic forms) did not involve the inverse elements of the mixed model equations. The estimators were also translation invariant because they were based on the absorbed right hand sides of the mixed model equations. The biggest advantage of these estimators was their computational ease. Unfortunately, with more than one random factor in the model, the 'quadratic forms' can be shown to be bilinear forms and hence not always guaranteed to be positive definite. The method, similar to most other methods, was found to give severely biased estimates when data were selected, as in typical animal breeding situations (Ouweltjes et al. 1988).

Van Raden (1986) expanded the Schaeffer (1986) quadratic forms for the case with relationships among animals, which could also be used in the basic linear model for variance components.

## Derivation

Absorb the fixed effects into the equations for the random factors in the mixed model equations. The absorbed right hand side for the  $i^{th}$  random factor is then

$$\mathbf{Z}'_i \mathbf{M} \mathbf{y}$$

where  $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ . The expected value is  $\mathbf{0}$  and the variance-covariance matrix is

$$V(\mathbf{Z}'\mathbf{M}\mathbf{y}) = \sum_{j=1}^s (\mathbf{Z}'\mathbf{M}\mathbf{Z}_j)(\mathbf{Z}'_j\mathbf{M}\mathbf{Z})\sigma_j^2 + \mathbf{Z}'\mathbf{M}\mathbf{Z} \sigma_o^2.$$

The solution vector for the random factors are

$$\begin{pmatrix} \hat{\mathbf{u}}_1 \\ \hat{\mathbf{u}}_2 \\ \vdots \\ \hat{\mathbf{u}}_s \end{pmatrix} = \begin{pmatrix} \mathbf{Z}'_1\mathbf{M}\mathbf{Z}_1 + \mathbf{I}\alpha_1 & \mathbf{Z}'_1\mathbf{M}\mathbf{Z}_2 & \dots & \mathbf{Z}'_1\mathbf{M}\mathbf{Z}_s \\ \mathbf{Z}'_2\mathbf{M}\mathbf{Z}_1 & \mathbf{Z}'_2\mathbf{M}\mathbf{Z}_2 + \mathbf{I}\alpha_2 & \dots & \mathbf{Z}'_2\mathbf{M}\mathbf{Z}_s \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{Z}'_s\mathbf{M}\mathbf{Z}_1 & \mathbf{Z}'_s\mathbf{M}\mathbf{Z}_2 & \dots & \mathbf{Z}'_s\mathbf{M}\mathbf{Z}_s + \mathbf{I}\alpha_s \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{Z}'_1\mathbf{M}\mathbf{y} \\ \mathbf{Z}'_2\mathbf{M}\mathbf{y} \\ \vdots \\ \mathbf{Z}'_s\mathbf{M}\mathbf{y} \end{pmatrix}$$

The proposed quadratic forms are

$$\mathbf{y}'\mathbf{y} - \hat{\mathbf{b}}'\mathbf{X}'\mathbf{y} - \sum_{i=1}^s \hat{\mathbf{u}}'_i\mathbf{Z}'_i\mathbf{y} = \mathbf{y}'\mathbf{M}\mathbf{y} - \sum_{i=1}^s \hat{\mathbf{u}}'_i\mathbf{Z}'_i\mathbf{M}\mathbf{y}.$$

The latter form is the same as in REML and has pseudo expectation equal to

$$\begin{aligned} E(\mathbf{y}'\mathbf{y} - \hat{\mathbf{b}}'\mathbf{X}'\mathbf{y} - \hat{\mathbf{u}}'\mathbf{Z}'\mathbf{y}) &= (N - r(\mathbf{X}))\sigma_o^2 \\ \hat{\sigma}_o^2 &= (\mathbf{y}'\mathbf{y} - \hat{\mathbf{b}}'\mathbf{X}'\mathbf{y} - \hat{\mathbf{u}}'\mathbf{Z}'\mathbf{y}) / (N - r(\mathbf{X})) \\ \text{or } \hat{\sigma}_o^2 &= (\mathbf{y}'\mathbf{M}\mathbf{y} - \hat{\mathbf{u}}'\mathbf{Z}'\mathbf{M}\mathbf{y}) / (N - r(\mathbf{X})). \end{aligned}$$

The usual expectation of  $\hat{\mathbf{u}}_i' \mathbf{Z}_i' \mathbf{M} \mathbf{y}$  is

$$E(\hat{\mathbf{u}}_i' \mathbf{Z}_i' \mathbf{M} \mathbf{y}) = (tr \mathbf{Z}_i' \mathbf{M} \mathbf{Z}_i - tr \mathbf{C}_{ii} \mathbf{Z}_i' \mathbf{M} \mathbf{Z}_i \alpha_i) \sigma_i^2 - \sum_{j \neq i}^s tr(\mathbf{C}_{ij} \mathbf{Z}_j' \mathbf{M} \mathbf{Z}_i \alpha_j) \sigma_j^2 + (q_i - tr \mathbf{C}_{ii} \alpha_i) \sigma_o^2,$$

which when  $\alpha_i$ 's are taken as ratios of the true values, reduces to

$$E(\hat{\mathbf{u}}_i' \mathbf{Z}_i' \mathbf{M} \mathbf{y}) = tr \mathbf{Z}_i' \mathbf{M} \mathbf{Z}_i \sigma_i^2$$

which gives

$$\hat{\sigma}_i^2 = \hat{\mathbf{u}}_i' \mathbf{Z}_i' \mathbf{M} \mathbf{y} / tr \mathbf{Z}_i' \mathbf{M} \mathbf{Z}_i.$$

The above quadratic form is always positive when there is only one random factor in the model, and therefore, the estimator of  $\sigma_i^2$  is always positive. However, with more than one random factor in the model, the quadratic forms are really bi-linear forms and as such can take on negative values. Also,  $tr(\mathbf{Z}_i' \mathbf{M} \mathbf{Z}_i)$  does not depend on  $\alpha_i$  and therefore is a constant that only needs to be computed once. In addition only the diagonals of  $\mathbf{Z}_i' \mathbf{M} \mathbf{Z}_i$  need to be computed. This can be readily accomplished for most models.

### Van Raden's Method

Van Raden (1986) proposed the following quadratic form,  $\mathbf{w}_i' \hat{\mathbf{u}}_i$ , where  $\mathbf{w}_i = \mathbf{D}_i \mathbf{Z}_i' \mathbf{M} \mathbf{y}$  and  $\mathbf{D}_i = diag\{ (\mathbf{Z}_i' \mathbf{M} \mathbf{Z}_i + \mathbf{I} \alpha_i) \}^{-1}$ , then the pseudo expectation is

$$E(\mathbf{w}_i' \hat{\mathbf{u}}_i) = tr(\mathbf{D}_i \mathbf{Z}_i' \mathbf{M} \mathbf{Z}_i) \sigma_i^2$$

which gives

$$\hat{\sigma}_i^2 = \mathbf{w}_i' \hat{\mathbf{u}}_i / tr(\mathbf{D}_i \mathbf{Z}_i' \mathbf{M} \mathbf{Z}_i).$$

Because  $\mathbf{D}_i$  is diagonal, then  $tr(\mathbf{D}_i \mathbf{Z}_i' \mathbf{M} \mathbf{Z}_i)$  is easily obtained, but now depends on  $\alpha_i$  and must be re-computed each iteration. Let  $\lambda_{ik}$  be the  $k^{th}$  diagonal of  $\mathbf{Z}_i' \mathbf{M} \mathbf{Z}_i$  then

$$tr(\mathbf{D}_i \mathbf{Z}_i' \mathbf{M} \mathbf{Z}_i) = \sum_{k=1}^{q_i} [ \lambda_{ik} / (\lambda_{ik} + \alpha_i) ]$$

which is the sum of values that are sometimes used as approximate accuracies of elements of  $\hat{\mathbf{u}}_i$ . Hence  $tr(\mathbf{D}_i \mathbf{Z}_i' \mathbf{M} \mathbf{Z}_i)$  will always be less than or equal to  $q_i$ . Recall that Henderson's Method 4 uses  $\mathbf{w}_i' \mathbf{w}_i$  as the quadratic form, but  $\mathbf{w}_i' \hat{\mathbf{u}}_i$  is likely closer to  $\hat{\mathbf{u}}_i' \hat{\mathbf{u}}_i$  than is  $\hat{\mathbf{u}}_i' \mathbf{Z}_i' \mathbf{M} \mathbf{y}$  or  $\mathbf{w}_i' \mathbf{w}_i$ .

### An Example

Assume a model with two random factors and let  $\mathbf{y}' \mathbf{y} = 401,000$ . As *a priori* values let  $\alpha_1 = 10$  and  $\alpha_2 = 8$ . The rank of the fixed effects matrix,  $\mathbf{X}$ , is 4 and the total number of observations is 18. The first random factor has 3 levels and the second random factor has 6 levels. The absorbed coefficient matrix is

$$\begin{pmatrix} 2.6071 & -1.2143 & -1.3929 & -.2857 & .2500 & .0000 & -.6429 & -.6786 & 1.3571 \\ & 2.4286 & -1.2143 & .5714 & -.5000 & 1.0000 & -.7143 & .3571 & -.7143 \\ & & 2.6071 & -.2857 & .2500 & -1.0000 & 1.3571 & .3214 & -.6429 \\ & & & 2.0952 & .0000 & .0000 & -.2857 & -1.5238 & -.2857 \\ & & & & 1.5000 & -.7500 & -.2500 & -.2500 & -.2500 \\ & & & & & 1.7500 & -.2500 & -.5000 & -.2500 \\ & & & & & & 1.6071 & -.4286 & -.3929 \\ & & & & & & & 3.1310 & -.4286 \\ & & & & & & & & 1.6071 \end{pmatrix}$$

The absorbed right hand sides, solutions, and  $\mathbf{w}_i$  are

Factor	$\mathbf{Z}'_i\mathbf{M}\mathbf{y}$	$\hat{\mathbf{u}}_i$	$\mathbf{w}_i$
1	128.8214	9.6695	10.2182
	155.3571	9.0836	12.5000
	-284.1786	-18.7531	-22.5412
2	6.8905	.0750	.6745
	.5000	1.5094	.0526
	174.2500	14.8447	17.8718
	-157.6786	-12.1682	-16.4127
	49.7976	5.1919	4.4738
	-73.6786	-9.4528	-7.6692

and  $\mathbf{y}'\mathbf{M}\mathbf{y} = 62,796.881$ . The following can be calculated from the above;

$$\begin{aligned} \hat{\mathbf{u}}'_1\mathbf{Z}'_1\mathbf{M}\mathbf{y} &= 7,986.0696, \\ \hat{\mathbf{u}}'_2\mathbf{Z}'_2\mathbf{M}\mathbf{y} &= 5,461.6390, \\ tr\mathbf{Z}'_1\mathbf{M}\mathbf{Z}_1 &= 7.6429, \\ tr\mathbf{Z}'_2\mathbf{M}\mathbf{Z}_2 &= 11.6905, \end{aligned}$$

The estimates are

$$\begin{aligned} \hat{\sigma}_o^2 &= (\mathbf{y}'\mathbf{M}\mathbf{y} - \hat{\mathbf{u}}'\mathbf{Z}'\mathbf{M}\mathbf{y})/(N - r(\mathbf{X})) = 49,349.1724 / 14 \\ &= 3,524.9409, \\ \hat{\sigma}_1^2 &= \hat{\mathbf{u}}'_1\mathbf{Z}'_1\mathbf{M}\mathbf{y}/tr\mathbf{Z}'_1\mathbf{M}\mathbf{Z}_1 = 1,044.9063, \\ \text{and } \hat{\sigma}_2^2 &= \hat{\mathbf{u}}'_2\mathbf{Z}'_2\mathbf{M}\mathbf{y}/tr\mathbf{Z}'_2\mathbf{M}\mathbf{Z}_2 = 467.1870. \end{aligned}$$

Using Van Raden's proposed quadratic forms, the results are

$$\mathbf{w}'_1\hat{\mathbf{u}}_1 = 635.0673,$$

$$\begin{aligned}
\mathbf{w}'_2 \hat{\mathbf{u}}_2 &= 560.8674, \\
tr \mathbf{D}_1 \mathbf{Z}'_1 \mathbf{M} \mathbf{Z}_1 &= .6090, \\
tr \mathbf{D}_2 \mathbf{Z}'_2 \mathbf{M} \mathbf{Z}_2 &= 1.1608, \\
\hat{\sigma}_1^2 &= \mathbf{w}'_1 \hat{\mathbf{u}}_1 / tr \mathbf{D}_1 \mathbf{Z}'_1 \mathbf{M} \mathbf{Z}_1 \\
&= 635.0673 / .6090 = 1,042.8034, \\
\hat{\sigma}_2^2 &= 560.8674 / 1.1608 = 483.1732.
\end{aligned}$$

### Problem

Below are data following a model with one fixed factor(A) and two random factors(T and R). Apply EM REML, DFREML, ML, I-MIVQUE (if they stay positive), Pseudo-Expectation method, and Van Raden's method to these data and compare results.

$A_i$	$T_j$	$R_k$	$y_{ijkl}$	$A_i$	$T_j$	$R_k$	$y_{ijkl}$	$A_i$	$T_j$	$R_k$	$y_{ijkl}$
1	1	1	50	2	1	1	65	3	1	1	80
1	1	1	54	2	1	2	76	3	1	1	83
1	1	2	59	2	1	2	73	3	1	1	78
1	1	2	61	2	1	2	75	3	1	2	95
1	1	3	68	2	1	3	84	3	1	3	102
1	1	4	79	2	1	3	87	3	1	3	97
1	1	4	82	2	1	4	98	3	1	4	109
1	2	1	41	2	1	4	93	3	1	4	112
1	2	2	52	2	2	1	53	3	2	2	81
1	2	3	64	2	2	3	77	3	2	2	79
1	2	4	69	2	2	3	74	3	2	4	101
1	2	4	73	2	2	4	86	3	3	1	63
1	2	4	70	2	3	2	57	3	3	1	57
1	3	3	47	2	3	2	54	3	3	3	75
1	3	3	46	2	3	2	51	3	3	3	84
1	3	4	62	2	3	3	66	3	3	4	93
1	3	4	57					3	3	4	86
1	3	4	66								

# Method R

L. R. Schaeffer, April 28, 1999

This method was first reported through the Animal Geneticist's Discussion Group in 1993. A paper was subsequently published in the Journal of Animal Science, 1994, 72:2247-2253, by A. Reverter, B. L. Golden, R. M. Bourdon, and J. S. Brinks. The method came about as a result of a study on the change in genetic evaluations as animals acquire more progeny as in routine beef cattle evaluations. The method is based on the linear regression of recent solutions (based on all the data) on previous solutions (based on a subset of all data). The expected value of the regression equals 1 regardless of the distribution of observations or predictions. If the wrong ratio of variance components is used in the mixed model equations, then the computed value of the regression deviates from the expected value of one. If the computed value is greater than 1, then the  $h^2$  was too low. The method is iterative, in that the variance ratio is changed until the regression equals one.

As with the pseudo expectation method, Method R is ad hoc in nature and the properties of the method can only be established through simulation or empirical evidence. Schenkel (1998) has shown that Method R also suffers from bias due to non-random mating and when pedigree files are not complete, as does nearly every method.

## Development

To simplify presentation assume only one random factor in the model. Denote the data and model for subset  $j$  by

$$\mathbf{y}_{(j)} = \mathbf{X}_{(j)}\mathbf{b}_{(j)} + \mathbf{Z}_{(j)}\mathbf{u}_{(j)} + e_{(j)},$$

and denote the data and model for all data by the subscript  $(t)$ . Construct and solve MME for  $\mathbf{y}_{(j)}$  and  $\mathbf{y}_{(t)}$ .

$$\begin{pmatrix} \mathbf{X}'_{(k)}\mathbf{X}_{(k)} & \mathbf{X}'_{(k)}\mathbf{Z}_{(k)} \\ \mathbf{Z}'_{(k)}\mathbf{X}_{(k)} & \mathbf{Z}'_{(k)}\mathbf{Z}_{(k)} + \mathbf{G}^{-1}\alpha \end{pmatrix} \begin{pmatrix} \hat{\mathbf{b}}_{(k)} \\ \hat{\mathbf{u}}_{(k)} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'_{(k)}\mathbf{y}_{(k)} \\ \mathbf{Z}'_{(k)}\mathbf{y}_{(k)} \end{pmatrix}$$

for  $k = j$ , and for  $k = t$ . Note that the same variance ratio,  $\alpha$ , is used for both data sets. If an animal model is assumed where  $\mathbf{G} = \mathbf{A}$ , then the dimension of  $\mathbf{G}$  must be the same for both the subset and the total data sets. That is, all animals must be evaluated from each subset and the complete data. Note that  $\mathbf{y}_{(t)}$  contains all of the data in  $\mathbf{y}_{(j)}$  and more.

The  $j^{th}$  subset is constructed by randomly choosing a sample one half of the data. This should probably be a random sample half of the contemporary groups, so that animals in a contemporary group stay intact. The idea of Method R is to obtain a number of subsets from the same complete data set, say 10 or more, and to estimate the variances from each subset. Then the end result would be to average the 10 estimates together and calculate their SD in order to get an idea of the variability of the estimates due to sample size.

Now consider the  $i^{th}$  element of  $\hat{\mathbf{u}}$  from the two analyses above.

$$\begin{aligned} Cov(\hat{u}_{i(j)}, \hat{u}_{i(t)}) &= Cov(\hat{u}_{i(j)}, \hat{u}_{i(j)} + d_i) \\ &= V(\hat{u}_{i(j)}) + Cov(\hat{u}_{i(j)}, d_i) \\ &= V(\hat{u}_{i(j)}) \end{aligned}$$

because the change,  $d_i = \hat{u}_{i(j)} - \hat{u}_{i(t)}$ , from subset  $j$  to complete data  $t$  should be uncorrelated with the solution in subset  $j$ . The regression of  $\hat{u}_{i(t)}$  on  $\hat{u}_{i(j)}$  is then

$$\begin{aligned} \mathfrak{R}_c &= \frac{Cov(\hat{u}_{i(j)}, \hat{u}_{i(t)})}{V(\hat{u}_{i(j)})} \\ &= \hat{\mathbf{u}}'_{(j)} \hat{\mathbf{u}}_{(t)} / \hat{\mathbf{u}}'_{(j)} \hat{\mathbf{u}}_{(j)} \end{aligned}$$

which should equal one.

### An Example

Illustrating Method R really requires a fairly large data set, in order to achieve convergence. This example merely shows the calculations that should be followed, but the example does not converge to a good final estimate. Below are data on progeny of ten sires, for a model with just a mean and random (unrelated) sire and residual effects. The left half of the table represents the data from one subset and the right half represents the complete data.

Sire Number	Subset 1		Complete Data	
	$n_i$	$y_i$	$n_i$	$y_i$
1	85	11,462	213	29,218
2	54	6,819	182	23,745
3	112	15,153	121	16,456
4	90	11,115	191	23,637
5	144	16,625	307	35,666
6	192	26,203	298	40,720
7	136	17,957	237	30,944
8	83	10,151	226	27,899
9	174	20,700	338	39,865
10	161	21,346	274	36,483

The total sum of squares of the observations in Subset 1 was 20,715,797., and for the complete data was 39,977,646.

The solutions for sires within each data set are given below, for a variance ratio of 10.

Sire Number	Subset 1	Complete
1	6.3132	8.2960
2	-1.2768	1.8758
3	6.8886	6.9384
4	-3.8619	-4.4988
5	-11.5384	-11.9239
6	8.2531	7.8913
7	3.9549	1.9931
8	-4.8995	-4.8277
9	-8.3459	-10.2414
10	4.5125	4.4973

From these solutions, then

$$\begin{aligned}\hat{\mathbf{u}}'_{(1)}\hat{\mathbf{u}}_{(1)} &= 434.7651 \\ \hat{\mathbf{u}}'_{(1)}\hat{\mathbf{u}}_{(t)} &= 455.1629 \\ \mathfrak{R} &= 1.04692\end{aligned}$$

which suggests that the ratio for this sample of data should be lower than 10. To get the new ratio, divide 10 by  $\mathfrak{R}$  to give 9.55183. Use the new ratio to construct the MME again for the subset and the complete data sets. Continue until  $\mathfrak{R} = 1$ . The authors suggest that a binary search algorithm could be faster. Start with heritability equal to .5, and depending on the value of  $\mathfrak{R}$  go to heritability of .75 or .25. Continue reducing the interval by one half until  $\mathfrak{R} = 1$ .

The calculations need to be repeated for other subsets of the same data set. The Method R estimate is then the average of the estimates from each subset. Hence there is lots of computing that must be done. However, the only requirement is for a program that computes genetic evaluations for a particular model, and these are readily available. Computationally, Method R is very attractive, and for randomly mating populations Method R will give estimates close to the true values. One question is the number of subsets that should be chosen. This depends on the size of the complete data set and the amount of time to solve equations. The more subsets one has, the better should be the average of the estimates.

The authors claim that the estimates, from their experience, are similar to REML estimates, but the proof is rather loose in a technical sense. The covariances and variances in the regression utilize the same quadratic forms and involve the same inverse elements of the mixed model equations as in EM REML. Method R could be a method for obtaining an initial prior value for use in Bayesian or REML methodology, rather than as a stand-alone method. This is due to its problems with incomplete pedigrees and non random matings, and because the properties can not be proven except empirically.

# Convergence of Iterative Methods

L. R. Schaeffer, April 28, 1999

These notes are terribly outdated now, with the use of DFREML and Bayesian methods, but they may be of interest to those that work with problems of convergence of estimation systems.

Iterative methods of variance component estimation such as REML and ML are known to converge at very slow rates towards a final solution via the EM algorithm. For large datasets, the rate of convergence could affect computing costs substantially. Rates of convergence seem to be very data dependent and even specific to particular traits within a dataset. In REML, iterative MIVQUE is known to converge faster than the REML EM algorithm. Convergence is also adversely affected by the number of random factors in the model. Then there is always the question whether the system will converge or not. Hence the researcher needs to know a) will the system converge, and b) can the rate of convergence be improved, if the system does converge.

Below are least squares equations for a model with two fixed factors and one random factor

$$\begin{pmatrix}
 226 & 60 & 72 & 53 & 41 & 100 & 126 & 10 & 86 & 45 & 37 & 48 \\
 & 60 & 0 & 0 & 0 & 14 & 46 & 2 & 10 & 15 & 13 & 20 \\
 & & 72 & 0 & 0 & 53 & 19 & 0 & 21 & 19 & 7 & 25 \\
 & & & 53 & 0 & 22 & 31 & 3 & 32 & 0 & 15 & 3 \\
 & & & & 41 & 11 & 30 & 5 & 23 & 11 & 2 & 0 \\
 & & & & & 100 & 0 & 1 & 79 & 12 & 4 & 4 \\
 & & & & & & 126 & 9 & 7 & 33 & 33 & 44 \\
 & & & & & & & 10 & 0 & 0 & 0 & 0 \\
 & & & & & & & & 86 & 0 & 0 & 0 \\
 & & & & & & & & & 45 & 0 & 0 \\
 & & & & & & & & & & 37 & 0 \\
 & & & & & & & & & & & 48
 \end{pmatrix}
 \begin{pmatrix}
 \hat{\mu} \\
 \hat{A}_1 \\
 \hat{A}_2 \\
 \hat{A}_3 \\
 \hat{A}_4 \\
 \hat{B}_1 \\
 \hat{B}_2 \\
 \hat{C}_1 \\
 \hat{C}_2 \\
 \hat{C}_3 \\
 \hat{C}_4 \\
 \hat{C}_5
 \end{pmatrix}
 =
 \begin{pmatrix}
 6600 \\
 2100 \\
 2160 \\
 1325 \\
 1015 \\
 3000 \\
 3600 \\
 800 \\
 2720 \\
 1450 \\
 630 \\
 1000
 \end{pmatrix}
 .$$

Let factor C be the random factor and assume that  $\mathbf{y}'\mathbf{y} = 2,250,000$ . Let the best guess of the ratio of residual to factor C variances be 15. Both iterative MIVQUE and the REML EM algorithm were applied to these data and the results by iteration are in Tables 1 and 2. Both computing algorithms were slow to converge, but iterative MIVQUE was faster than REML EM.

Table 1. Results of iterative MIVQUE.

Iteration Number(i)	$\hat{\sigma}_o^2$	$\hat{\sigma}_C^2$	$\alpha_C$	$\Delta_{i-1}$	$\phi_{i-1}$
1	9138.59	659.45	13.85793	-1.14207	-----
2	9131.43	709.03	12.87875	-0.97918	1.16635
3	9124.82	757.03	12.05338	-0.82537	1.18635
4	9118.87	802.14	11.36816	-0.68522	1.20453
5	9113.66	843.32	10.80684	-0.56132	1.22073
6	9109.19	879.92	10.35226	-0.45458	1.23481
7	9105.44	911.67	9.98767	-0.36459	1.24683
8	9102.33	938.62	9.69760	-0.29007	1.25690
9	9099.80	961.08	9.46833	-0.22927	1.26519
10	9097.76	979.51	9.28809	-0.18024	1.27203
20	9090.85	1044.20	8.70603	-0.01422	1.29466
30	9090.30	1049.49	8.66166	-0.00106	1.29245
40	9090.26	1049.88	8.65838	-0.00008	1.29750
50	9090.26	1049.91	8.65813		

Table 2. Results for REML.

Iteration Number(i)	$\hat{\sigma}_o^2$	$\hat{\sigma}_C^2$	$\hat{\alpha}_C$	$\Delta_{i-1}$	$\phi_{i-1}$
1	9145.94	622.81	14.68508	-0.31492	-----
2	9143.95	635.96	14.37814	-0.30694	1.02600
3	9141.94	649.31	14.07958	-0.29856	1.02807
4	9139.94	662.81	13.78974	-0.28984	1.03009
5	9137.94	676.44	13.50875	-0.28079	1.03223
6	9135.95	690.16	13.23748	-0.27147	1.03433
7	9133.97	703.94	12.97557	-0.26191	1.03650
8	9132.01	717.73	12.72341	-0.25216	1.03867
9	9130.08	731.51	12.48112	-0.24229	1.04074
10	9128.17	745.23	12.24881	-0.23231	1.04296
20	9111.59	870.89	10.46233		
30	9100.81	958.84	9.49152		
40	9095.07	1007.74	9.02521		
50	9092.60	1027.79	8.8468		

The final estimates were  $\hat{\sigma}_o^2 = 9090.260284$ ,  $\hat{\sigma}_C^2 = 1049.912327$ , and  $\hat{\alpha}_C = 8.65811$ .

### Convergence Curves

The following results are due mainly to Dr. Ignacy Misztal. A plot of the estimated variance ratio against iteration number would show a nonlinear relationship. The curve would appear to be asymptoting to a final value if the system was converging, as in the

example. This curve may be characterized by a model, and therefore, the asymptotic value could be predicted if the parameters of that model were known. Let

$\alpha_o$  = initial starting value of variance ratio,

$\alpha_i$  = value of variance ratio after the  $i^{\text{th}}$  iteration,

$\Delta_i = \alpha_{i+1} - \alpha_i$ , the difference in values of the variance ratio after one iteration, and

$\phi_i = \Delta_{i-1}/\Delta_i$ , ratio of differences in consecutive iterations.

Assume  $\phi_i = \phi$ , a constant for all iterations. For REML this assumption is reasonable since  $\phi_i$  varied from 1.026 to 1.043 over the first ten iterations (Table 2), but for iterative MIVQUE the values ranged from 1.17 to 1.272 in the first ten iterations (Table 1). If a constant value can be assumed, then

$$\begin{aligned}\Delta_1 &= \Delta_o/\phi, \text{ or more generally,} \\ \Delta_2 &= \Delta_1/\phi = (\Delta_o/\phi)/\phi = \Delta_o/\phi^2 \\ \text{or } \Delta_i &= \Delta_o/\phi^i.\end{aligned}$$

Then

$$\begin{aligned}\alpha_1 &= \alpha_o + \Delta_o, \\ \alpha_2 &= \alpha_1 + \Delta_1 = \alpha_o + \Delta_o + \Delta_1 = \alpha_o + \Delta_o \left( \sum_{j=0}^1 \phi^{-j} \right) \\ \text{and } \alpha_i &= \alpha_o + \Delta_o \left( \sum_{j=0}^{i-1} \phi^{-j} \right).\end{aligned}$$

Suppose  $i = \infty$ , then

$$\begin{aligned}\alpha_\infty &= \alpha_o + \Delta_o \left( \sum_{j=0}^{\infty} \phi^{-j} \right), \text{ but notice that} \\ \sum_{j=0}^{\infty} \phi^{-j} &= 1 + \sum_{j=1}^{\infty} \phi^{-j} \\ &= 1 + \phi^{-1} \sum_{j=0}^{\infty} \phi^{-j} \\ &= (1 - \phi^{-1})^{-1}, \text{ and consequently} \\ \alpha_\infty &= \alpha_o + \Delta_o (1 - \phi^{-1})^{-1}.\end{aligned}$$

For REML in Table 2, if  $\phi = 1.035$ ,  $\Delta_o = -.31492$  and  $\alpha_o = 15$ , then

$$\begin{aligned}\alpha_\infty &= 15 + (-.31492) [1 - (1.035)^{-1}]^{-1} \\ &= 15 + (-.31492)/[1 - .96618] \\ &= 15 + (-.31492)/.03382 = 15 - 9.31263 = 5.68737.\end{aligned}$$

Obviously, 5.68 is greatly different from the converged value of 8.65, and this is due to the fact that  $\phi_i$  is not a constant over iterations. However, if  $\phi$  had been 1.05218, then  $\alpha_\infty$  would equal 8.65. Hence a small difference in  $\phi$  can make a big difference in the predicted ratio.

For MIVQUE let

$$\begin{aligned}\phi &= 1.2975, \alpha_o = 15, \text{ and } \Delta_o = 1.14207, \text{ then} \\ \alpha_\infty &= 15 + (-1.14207)/[1 - (1.2975)^{-1}] \\ &= 15 - 1.14207/.22929 = 15 - 4.98096 = 10.01904.\end{aligned}$$

### Common Intercept Approach

The Common Intercept Approach (CIA) was suggested in 1979 by Dr. Lynn Johnson (then a postdoctoral research associate with Dr. Burnside) and has been derived from the nonlinear model of the previous section. The purpose of CIA was to reduce the number of iterations needed to attain convergence in any iterative procedure. In addition, CIA can be used to determine if an iterative procedure will converge for a given data set.

Let  $\alpha_{xo}$  represent one starting value and let  $\alpha_{zo}$  represent a different starting value. The corresponding changes in value after one iteration are  $\Delta_{xo}$  and  $\Delta_{zo}$ . Then

$$\begin{aligned}\alpha_{x\infty} &= \alpha_{xo} + \Delta_{xo}/(1 - \phi^{-1}) \\ \text{or } (1 - \phi^{-1})^{-1} &= (\alpha_{x\infty} - \alpha_{xo})\Delta_{xo}^{-1}\end{aligned}$$

which can be substituted into

$$\begin{aligned}\alpha_{z\infty} &= \alpha_{zo} + \Delta_{zo}/(1 - \phi^{-1})^{-1} \\ &= \alpha_{zo} + \Delta_{zo}(\alpha_{x\infty} - \alpha_{xo})\Delta_{xo}^{-1}\end{aligned}$$

Assuming that

$$\begin{aligned}\alpha_{z\infty} &= \alpha_{x\infty} = \alpha_\infty, \text{ then} \\ \alpha_\infty &= \alpha_{zo} + \Delta_{zo}(\alpha_\infty - \alpha_{xo})\Delta_{xo}^{-1}\end{aligned}$$

which can be re-arranged to give

$$\alpha_\infty = (\Delta_{xo}\alpha_{zo} - \Delta_{zo}\alpha_{xo}) / (\Delta_{xo} - \Delta_{zo}).$$

Using the example data and REML, from Table 2, let

$$\begin{aligned}\alpha_{xo} &= 15, \\ \Delta_{xo} &= -.31492, \\ \alpha_{zo} &= 5, \\ \text{and } \Delta_{zo} &= +.40208,\end{aligned}$$

then CIA gives

$$\begin{aligned}\alpha_{\infty} &= (-.31492(5) - .40208(15))/(-.31492 - .40208) \\ &= -7.60580 / -.717 \\ &= 10.60781.\end{aligned}$$

With MIVQUE, let

$$\begin{aligned}\alpha_{x_0} &= 15, \\ \Delta_{x_0} &= -1.14207, \\ \alpha_{z_0} &= 5, \\ \text{and } \Delta_{z_0} &= .96246,\end{aligned}$$

then

$$\begin{aligned}\alpha_{\infty} &= (-1.14207(5) - .96246(15)) / (-1.14207 - .96246) \\ &= -20.14725 / -2.10453 \\ &= 9.57328.\end{aligned}$$

Thus, two iterations with different starting values are needed to obtain a prediction of the converged value. The closer the two starting values are to the actual converged estimate, then CIA will work better. For example, using REML, with  $\alpha_{x_0} = 9$  and  $\alpha_{z_0} = 8.5$  then

$$\begin{aligned}\Delta_{x_0} &= .02746 \text{ and } \Delta_{z_0} = .013217, \text{ and} \\ \alpha_{\infty} &= (-.02746(8.5) - .013217(9)) / (-.02746 - .013217) \\ &= -.35236 / -.04068 = 8.66233.\end{aligned}$$

With MIVQUE, the result was 8.66014.

If the estimate of  $\alpha_{\infty}$  from CIA is negative, then this indicates that convergence may not be possible. Select new starting ratios that are closer to each other and retry the CIA. If the estimate of the converged value remains negative, then the system will likely not converge. Use of an unbiased, non-iterative procedure on the same data will likely produce negative estimates of one or more variance components. In any event, the negative value of  $\alpha_{\infty}$  should not be used as the next prior value in the iterative procedure.

### Aitken's Approximation

Another method of determining the converged value is Aitken's approximation given by the following formula:

$$\alpha_{\infty} = \frac{\alpha_i \alpha_{i+2} - \alpha_{i+1}^2}{\alpha_i - 2\alpha_{i+1} + \alpha_{i+2}}.$$

To illustrate, take the values from Table 1 for iterations 6, 7, and 8, then

$$\begin{aligned}
 \alpha_{\infty} &= \frac{\alpha_6\alpha_8 - \alpha_7^2}{\alpha_6 - 2\alpha_7 + \alpha_8} \\
 &= [(10.35226)(9.69760) - (9.98767)^2]/[(10.35226) - 2(9.98767) + 9.69760] \\
 &= [0.6385]/[0.07452] \\
 &= 8.56817
 \end{aligned}$$

for MIVQUE, and

$$\begin{aligned}
 \alpha_{\infty} &= [(13.23748)(12.72341) - (12.97557)^2]/[13.23748 - 2(12.97557) + 12.72341] \\
 &= (.06047)/(.00975) \\
 &= 6.20191
 \end{aligned}$$

for REML.

### Relaxation Factors

Relaxation factors attempt to speed convergence. The success of relaxation factors depends upon the magnitude of the factor used. If the factor is too large, then changes between iterations may be too disruptive for the system and convergence may not be achieved any faster, if at all. If the factor is too small, then the increased rate of convergence may not be noticeable. If  $\rho$  is the relaxation factor, then the appropriate starting ratio for the (i+1) iterate is not  $\alpha_i$ , but is

$$\alpha_{(i)} = \alpha_i + \rho [\alpha_i - \alpha_{i-1}].$$

Usually a factor between 0 and 1 is used. Let  $\rho = 1$  in the example, then in REML  $\alpha_1 = 14.68508$  and  $\alpha_o = 15$ , and therefore  $\alpha_{(1)} = 14.37016$ . Thus, only one iteration was skipped, and overall probably half the iterations would be needed. Care needs to be used in choosing a value of  $\rho$  because it could be dependent on the data structure and the observation values themselves. Thus, using  $\rho = .6$  for all situations may not give the best results. Some test runs with different values may be needed.

# Covariances and General Models

L. R. Schaeffer, April 28, 1999

The linear model that has been assumed for variance component estimation to this point has variance-covariance matrices of all random factors of the form  $\mathbf{I}c$ , where  $c$  is the variance to be estimated, and covariances between any two random factors were always null. With animal models and maternal genetic effects, these assumptions are not met, and some of the previous methods cannot be used. Methods applicable to more general models are MIVQUE, REML, ML, Method R and Bayesian methods.

## Covariances Between Two Random Factors

Assume the following model and example problem where

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}_1\mathbf{u}_1 + \mathbf{Z}_2\mathbf{u}_2 + \mathbf{e},$$

and the data are as in the table below. The total sum of squares was  $\mathbf{y}'\mathbf{y} = 325,520$ . Also,  $V(\mathbf{u}_1) = \mathbf{I}\sigma_1^2$ ,  $V(\mathbf{u}_2) = \mathbf{I}\sigma_2^2$ , and letting  $\mathbf{u}_1$  and  $\mathbf{u}_2$  have the same number of elements in the same sequence, then

$$Cov(\mathbf{u}_1, \mathbf{u}_2) = \mathbf{I}\sigma_{12}.$$

This covariance changes the structure of the variance-covariance matrix of  $\mathbf{y}$  to be

$$Var(\mathbf{y}) = \mathbf{Z}_1\mathbf{Z}'_1\sigma_1^2 + \mathbf{Z}_2\mathbf{Z}'_2\sigma_2^2 + (\mathbf{Z}_1\mathbf{Z}'_2 + \mathbf{Z}_2\mathbf{Z}'_1)\sigma_{12} + \mathbf{I}\sigma_0^2.$$

Depending on the method of estimation applied to  $\mathbf{y}$ , the quadratic forms will likely contain  $\sigma_{12}$  in their expectations. If  $\mathbf{W} = [\mathbf{X} \ \mathbf{Z}_1 \ \mathbf{Z}_2]$ , then

$$Var(\mathbf{W}'\mathbf{y}) = \mathbf{W}'\mathbf{Z}_1\mathbf{Z}'_1\mathbf{W}\sigma_1^2 + \mathbf{W}'\mathbf{Z}_2\mathbf{Z}'_2\mathbf{W}\sigma_2^2 + \mathbf{W}'(\mathbf{Z}_1\mathbf{Z}'_2 + \mathbf{Z}_2\mathbf{Z}'_1)\mathbf{W}\sigma_{12} + \mathbf{W}'\mathbf{W}\sigma_0^2.$$

Table: Example data for two random factors with covariance.

<b>y</b>	Level of <b>b</b>	Level of <b>u<sub>1</sub></b>	Level of <b>u<sub>2</sub></b>
145	1	1	2
84	1	2	4
83	1	3	4
75	1	4	5
122	1	1	5
100	2	2	3
172	2	3	2
102	2	4	2
158	2	1	4
156	2	2	5
84	2	3	5
133	3	4	3
139	3	1	3
173	3	2	3
117	3	3	4
187	3	4	2
150	3	1	2
160	3	2	5

The least squares equations,  $\mathbf{W}'\mathbf{W}$  and  $\mathbf{W}'\mathbf{y}$ , are

$$\begin{pmatrix}
 5 & 0 & 0 & 2 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 2 & 2 \\
 0 & 6 & 0 & 1 & 2 & 2 & 1 & 0 & 0 & 2 & 1 & 1 & 2 \\
 0 & 0 & 7 & 2 & 2 & 1 & 2 & 0 & 0 & 2 & 3 & 1 & 1 \\
 2 & 1 & 2 & 5 & 0 & 0 & 0 & 0 & 0 & 2 & 1 & 1 & 1 \\
 1 & 2 & 2 & 0 & 5 & 0 & 0 & 0 & 0 & 0 & 2 & 1 & 2 \\
 1 & 2 & 1 & 0 & 0 & 4 & 0 & 0 & 0 & 1 & 0 & 2 & 1 \\
 1 & 1 & 2 & 0 & 0 & 0 & 4 & 0 & 0 & 2 & 1 & 0 & 1 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 1 & 2 & 2 & 2 & 0 & 1 & 2 & 0 & 0 & 5 & 0 & 0 & 0 \\
 0 & 1 & 3 & 1 & 2 & 0 & 1 & 0 & 0 & 0 & 4 & 0 & 0 \\
 2 & 1 & 1 & 1 & 1 & 2 & 0 & 0 & 0 & 0 & 0 & 4 & 0 \\
 2 & 2 & 1 & 1 & 2 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 5
 \end{pmatrix}
 \begin{pmatrix}
 \hat{b}_1 \\
 \hat{b}_2 \\
 \hat{b}_3 \\
 \hat{u}_{11} \\
 \hat{u}_{12} \\
 \hat{u}_{13} \\
 \hat{u}_{14} \\
 \hat{u}_{15} \\
 \hat{u}_{21} \\
 \hat{u}_{22} \\
 \hat{u}_{23} \\
 \hat{u}_{24} \\
 \hat{u}_{25}
 \end{pmatrix}
 =
 \begin{pmatrix}
 509 \\
 772 \\
 1059 \\
 714 \\
 673 \\
 456 \\
 497 \\
 0 \\
 0 \\
 756 \\
 545 \\
 442 \\
 597
 \end{pmatrix}$$

### MIVQUE

Let

$$\mathbf{G}^{-1}\sigma_0^2 = \begin{pmatrix} 11 \mathbf{I} & -8 \mathbf{I} \\ -8 \mathbf{I} & 11 \mathbf{I} \end{pmatrix},$$

then the solutions to the mixed model equations become

$$\begin{pmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{b}_3 \end{pmatrix} = \begin{pmatrix} 101.2441 \\ 127.9646 \\ 151.2713 \end{pmatrix},$$

$$\begin{pmatrix} \hat{u}_{11} \\ \hat{u}_{12} \\ \hat{u}_{13} \\ \hat{u}_{14} \\ \hat{u}_{15} \end{pmatrix} = \begin{pmatrix} 7.5107 \\ 7.3838 \\ -7.1244 \\ -6.2260 \\ -1.5440 \end{pmatrix}, \quad \begin{pmatrix} \hat{u}_{21} \\ \hat{u}_{22} \\ \hat{u}_{23} \\ \hat{u}_{24} \\ \hat{u}_{25} \end{pmatrix} = \begin{pmatrix} 5.4623 \\ 9.9943 \\ -7.3217 \\ -6.0118 \\ -2.1231 \end{pmatrix}.$$

The quadratic forms and expectations were

1.  $\hat{\mathbf{u}}_1' \hat{\mathbf{u}}_1 = 202.8346$
2.  $\hat{\mathbf{u}}_2' \hat{\mathbf{u}}_2 = 223.9808$
3.  $\hat{\mathbf{u}}_1' \hat{\mathbf{u}}_2 + \hat{\mathbf{u}}_2' \hat{\mathbf{u}}_1 = 415.3844$
4.  $\mathbf{y}'\mathbf{y} - \hat{\mathbf{b}}'\mathbf{X}'\mathbf{y} - \hat{\mathbf{u}}'\mathbf{Z}'\mathbf{y} = 11,372.25$

and

$$E \begin{pmatrix} 1. \\ 2. \\ 3. \\ 4. \end{pmatrix} = \begin{pmatrix} .47264 & .14914 & .13386 & .16285 \\ .15982 & .43500 & .14655 & .16092 \\ .51268 & .47062 & .47776 & .28089 \\ 8.14125 & 7.83617 & -5.75683 & 12.72461 \end{pmatrix} \begin{pmatrix} \sigma_1^2 \\ \sigma_2^2 \\ \sigma_{12} \\ \sigma_0^2 \end{pmatrix}.$$

The estimates were

$$\begin{aligned} \hat{\sigma}_1^2 &= 23.8460, & \hat{\sigma}_2^2 &= 96.8782, \\ \hat{\sigma}_{12} &= 210.9195, & \text{and } \hat{\sigma}_0^2 &= 914.2274. \end{aligned}$$

The correlation between the two random vectors was greater than one indicating that unbiased procedures, in general, may not perform well in estimating the covariance between factors.

## REML

REML is a biased procedure because the estimates are kept within the allowable parameter space. Hence, the estimated correlation between the two random vectors should be between -1 to +1. REML uses the same quadratic forms as MIVQUE, but also

$$tr\mathbf{C}_{11} = .66328, \quad tr\mathbf{C}_{22} = .66863, \quad \text{and} \quad tr\mathbf{C}_{12} = .43290.$$

The estimates were

$$\begin{aligned} \hat{\sigma}_0^2 &= (\mathbf{y}'\mathbf{y} - \hat{\mathbf{b}}'\mathbf{X}'\mathbf{y} - \hat{\mathbf{u}}'\mathbf{Z}'\mathbf{y}) / (N - r(\mathbf{X})) \\ &= 11,372.25 / (18 - 3) = 758.15, \\ \hat{\sigma}_1^2 &= (\hat{\mathbf{u}}_1' \hat{\mathbf{u}}_1 + tr\mathbf{C}_{11} \hat{\sigma}_0^2) / 5 \end{aligned}$$

$$\begin{aligned}
&= (202.8346 + .66328 (758.15))/5 = 141.1401, \\
\hat{\sigma}_2^2 &= (\hat{\mathbf{u}}_2' \hat{\mathbf{u}}_2 + tr \mathbf{C}_{22} \hat{\sigma}_0^2)/5 \\
&= (223.9808 + .66863 (758.15))/5 = 146.1805, \\
\text{and } \hat{\sigma}_{12} &= (\hat{\mathbf{u}}_1' \hat{\mathbf{u}}_2 + tr \mathbf{C}_{12} \hat{\sigma}_0^2)/5 \\
&= (207.6922 + .43290 (758.15))/5 = 107.1791.
\end{aligned}$$

The estimated correlation between the two random vectors with REML was .7462. Of course, the procedure must be iterated many more times before a final solution is obtained, and this could lead to a correlation that converges towards unity.

The course notes have descriptions of methods of variance component estimation for maternal effects and multiple trait models, and will not be reproduced here.