Theor. Appl. Genet. 49 (1977) 273-282

Prediction of the Merits of Single Crosses

C.R. Henderson
Department of Animal Science, Cornell University, Ithaca, New York (USA)

Summary. Best linear unbiased prediction of single crosses is described for a model that is somewhat more complete than those previously published. The method applies to unequal subclass numbers with unequal means of the observations. The lines are assumed to be a random sample from some population. Also, methods for unbiased estimation of the variances and covariances from such data are presented.

1. Introduction

Diallel crosses are used for two purposes: first as an aid in interpreting the genetic mechanism, and second to test the relative merits of different single crosses for some practical applications such as the production of corn. For the first purpose, the lines used in the diallel crosses must be regarded as a random sample from some population. Then the analysis of the results usually involves estimation of components of variance and covariance. For the purposes of evaluating crosses, most workers have estimated by least squares, presumably assuming that the lines are fixed. If in fact they are, this is a logical estimation method. If, however, the lines are actually random, the problem becomes one of prediction, and least squares is not the best method. If all observations have a common mean, μ, selection index techniques have certain optimum properties provided certain required variances and covariances are known. In contrast, if the means are different and unknown, selection index methods are not appropriate. If one can assume known variances and covariances, it is possible to find predictions that are unbiased and have minimum prediction error variance. It is the purpose of this paper to describe these methods and to present methods for estimating variances and covariances needed in these prediction methods. In addition, a model is used that has one more parameter than has been used in most discussions of diallel crosses.

2. Diallel Cross Design and Model

There are available q lines for use in a diallel cross design. Each of these q lines is used as both line of sire and line of dam. The resulting $q \times q$ two-way table of numbers of observations on each cross has all diagonal elements = 0. Some of the off-diagonal elements may also be 0, but the most efficient design would ordinarily be one in which the number of observations would be equal except for the diagonals. We assume that the purpose of the design is to select the best crosses for use in a future breeding program.

The model assumed for $\underline{y}$, the observation vector of length n, is

$$\underline{y} = \underline{X\beta} + \underline{Z}_1\underline{s} + \underline{Z}_2\underline{d} + \underline{Z}_3\underline{r} + \underline{e}$$

where $\underline{X}$ is a known $n \times p$ matrix, possibly with both dummy variates and covariates; $\underline{\beta}$ is an unknown fixed vector; $\underline{Z}_1$, $\underline{Z}_2$, and $\underline{Z}_3$ are known matrices of order $n \times q$, $n \times q$, and $n \times q(q-1)$, respectively, and with all elements = 1 or 0; $\underline{s}' = [s_1 \ldots s_q]$, corresponding to line used as line of sire; $\underline{d}' = [d_1 \ldots d_q]$, corresponding to line used as line of dam; and $\underline{r}$ is a $q(q-1) \times 1$ vector of line of sire by line of dam interactions with elements, r_{ij}, ordered j within i and including those r_{ij} $(i \neq j)$ corresponding to $n_{ij} = 0$. The reason for including these terms is that they can be predicted in the random lines model. $E(\underline{e}) = \underline{0}$, and $Var(\underline{e}) = \underline{I}\sigma_e^2$. The distributional properties of $\underline{s}$, $\underline{d}$, and $\underline{r}$ in the random lines model are described later. An underscored upper case letter denotes a matrix, an underscored lower case letter a column vector, and an italicized lower case letter a scalar.

It is assumed that the i subscript denotes line of sire and the j subscript denotes line of dam. In scalar notation, the model is

$$y_{ijk} = \sum_{h=1}^{p} x_{hijk}\beta_h + s_i + d_j + r_{ij} + e_{ijk} .$$

3. Fixed Lines

If lines are regarded as fixed, then $\underline{s}$, $\underline{d}$, and $\underline{r}$ are fixed. If all n_{ij} $(i \neq j) > 0$, the least squares equations are

$$\begin{pmatrix} \underline{C}_{00} & \underline{C}_{01} & \underline{C}_{02} & \underline{C}_{03} \\ \underline{C}'_{01} & \underline{C}_{11} & \underline{C}_{12} & \underline{C}_{13} \\ \underline{C}'_{02} & \underline{C}'_{12} & \underline{C}_{22} & \underline{C}_{23} \\ \underline{C}'_{03} & \underline{C}'_{13} & \underline{C}'_{23} & \underline{C}_{33} \end{pmatrix} \begin{pmatrix} \hat{\underline{\beta}} \\ \hat{\underline{s}} \\ \hat{\underline{d}} \\ \hat{\underline{r}} \end{pmatrix} = \begin{pmatrix} \underline{t}_0 \\ \underline{t}_1 \\ \underline{t}_2 \\ \underline{t}_3 \end{pmatrix} \qquad (1)$$

where $\underline{C}_{00} = \underline{X}'\underline{X}$; $\underline{C}_{01}$ is a $p \times q$ matrix with elements $[x_{hi..}]$ ordered i within h with each row corresponding to an h; $\underline{C}_{02}$ is a $p \times q$ matrix with elements $[x_{h.j.}]$ ordered j within h with each row corresponding to an element of h; $\underline{C}_{03}$ is a $p \times q(q-1)$ matrix with elements $[x_{hij.}]$ ordered j within i with each row corresponding to an h; $\underline{C}_{11}$ is a q^2 diagonal matrix with diagonal elements = $n_{i.}$; $\underline{C}_{12}$ is the q^2 matrix of line of sire × line of dam subclass numbers; $\underline{C}_{13}$ is a $q(q-1)$ matrix with the ith row containing $n_{i1} \ldots n_{iq}$ and 0's, the positions of the n_{ij} corresponding to the positions of r_{ij} in the $\underline{r}$ vector; $\underline{C}_{22}$ is a q^2 diagonal matrix with diagonal elements = $n_{.j}$; $\underline{C}_{23}$ is a $q \times q(q-1)$ matrix with the jth row containing $n_{ij} \ldots n_{qj}$ and 0's, the positions of the n_{ij} corresponding to the positions of r_{ij} in the $\underline{r}$ vector; $\underline{C}_{33}$ is a diagonal matrix of order $q(q-1)$ with elements = n_{ij}; $\underline{t}_0 = \underline{X}'\underline{y}$; $\underline{t}'_1 = [y_{1..} \ldots y_{q..}]$; $\underline{t}'_2 = [y_{.1.} \ldots y_{.q.}]$; and $\underline{t}_3$ is a $q(q-1)$ vector with elements $y_{ij.}$ $(i \neq j)$ and ordered as in the elements of $\underline{r}$. If $n_{ij}(i \neq j) = 0$, the corresponding value of $y_{ij.} = 0$.

Now, if some $n_{ij}(i \neq j) = 0$, delete the corresponding rows of $\underline{C}'_{03}$, $\underline{C}'_{13}$, $\underline{C}'_{23}$, and $\underline{C}_{33}$ in (1). Also delete the corresponding columns of the coefficient matrix in (1) and similarly for the right-hand side. Then it is easy to see that the rank of the reduced matrix,

$$\begin{pmatrix} \underline{C}_{11} & \underline{C}_{12} & \underline{C}_{13} \\ \underline{C}'_{12} & \underline{C}_{22} & \underline{C}_{23} \\ \underline{C}'_{13} & \underline{C}'_{23} & \underline{C}_{33} \end{pmatrix},$$

is equal to the number of filled subclasses. A solution is $\hat{\underline{s}} = \hat{\underline{d}} = \underline{0}$ and $\hat{\underline{r}}$ = some solution to (2) where $\underline{C}_{03}$ and $\underline{C}_{33}$ are reduced if there are missing subclasses;

$$\begin{pmatrix} \underline{C}_{00} & \underline{C}_{03} \\ \underline{C}'_{03} & \underline{C}_{33} \end{pmatrix} \begin{pmatrix} \hat{\underline{\beta}} \\ \hat{\underline{r}} \end{pmatrix} = \begin{pmatrix} \underline{t}_0 \\ \underline{t}_3 \end{pmatrix}. \qquad (2)$$

Assuming that there are no interactions of factors in $\underline{\beta}$ with $\underline{s}$, $\underline{d}$, and $\underline{r}$, it is clear that the only estimable functions of $\underline{s}$, $\underline{d}$, and $\underline{r}$ are linear functions of $s_i + d_j + r_{ij}$ for which $n_{ij} > 0$. What, then, is the meaning of differences among sire lines? If all subclasses were filled (including diagonals), we might define the difference between lines i and k as

$$\frac{1}{q}\sum_{j}(s_i + r_{ij} - s_k - r_{kj}).$$

This, of course, is not possible because of the missing diagonals and possibly other subclasses as well. A not very satisfactory comparison might be

$$\frac{1}{n_f}\sum(s_i + r_{ij} - s_k - r_{kj})$$

where summation is over the subclasses for which n_{ij} and n_{kj} both are greater than 0, and there are n_f such pairs. Realistically, all we can do in the fixed case is to compare the relative merits of the crosses for which $n_{ij} > 0$. Of course, if it were known that all $r_{ij} = 0$, we could compare lines by solving

$$\begin{pmatrix} \underline{C}_{00} & \underline{C}_{01} & \underline{C}_{02} \\ \underline{C}_{01}' & \underline{C}_{11} & \underline{C}_{12} \\ \underline{C}_{11}' & \underline{C}_{12}' & \underline{C}_{22} \end{pmatrix} \begin{pmatrix} \hat{\underline{\beta}} \\ \hat{\underline{s}} \\ \hat{\underline{d}} \end{pmatrix} = \begin{pmatrix} \underline{t}_0 \\ \underline{t}_1 \\ \underline{t}_2 \end{pmatrix} .$$

The assumption of no interaction is probably untenable since the breeder uses diallel crosses because he wants to capitalize on interaction, sometimes called "nicking."

3.1. Special case in which $s_i = d_i$ and $r_{ij} = r_{ji}$

In some species, it is logical to assume that the ijth cross is identical to the jith cross. That is, $s_i = d_i$ and $r_{ij} = r_{ji}$. In that case, let the model be

$$\underline{y} = \underline{X}\underline{\beta} + \underline{Z}_4\underline{g} + \underline{Z}_5\underline{h} + \underline{e}$$

where $\underline{g}' = [g_1 \ \dots \ g_q]$ and $\underline{h}' = [h_{12} \ h_{13} \dots h_{q-1,q}]$ in which the first subscript of h_{ij} is less than the second. $\underline{g}$ corresponds to both $\underline{s}$ and $\underline{d}$ of the previous model, and $\underline{h}$ corresponds to $\underline{r}$. The least squares equations when all $n_{ij} + n_{ji} > 0$ except for diagonals are

$$\begin{pmatrix} \underline{C}_{00} & \underline{C}_{04} & \underline{C}_{05} \\ \underline{C}_{04}' & \underline{C}_{44} & \underline{C}_{45} \\ \underline{C}_{05}' & \underline{C}_{45}' & \underline{C}_{55} \end{pmatrix} \begin{pmatrix} \hat{\underline{\beta}} \\ \hat{\underline{g}} \\ \hat{\underline{h}} \end{pmatrix} = \begin{pmatrix} \underline{t}_0 \\ \underline{t}_4 \\ \underline{t}_5 \end{pmatrix} \qquad (3)$$

where $\underline{C}_{04} = \underline{C}_{01} + \underline{C}_{02}$; $\underline{C}_{05}$ is formed from $\underline{C}_{03}$ by adding the pairs of columns corresponding to r_{ij} and r_{ji}; $\underline{C}_{44} = \underline{C}_{11} + \underline{C}_{12} + \underline{C}_{12}' + \underline{C}_{22}$; $\underline{C}_{45}$ is formed by adding $\underline{C}_{13}$ to $\underline{C}_{23}$ and then adding the pairs of columns corresponding to r_{ij} and r_{ji}; $\underline{C}_{55}$ is a diagonal matrix of order $q(q-1)/2$ with diagonal elements = $n_{ij} + n_{ji}$; $\underline{t}_4 = \underline{t}_1 + \underline{t}_2$; and $\underline{t}_5$ is formed from $\underline{t}_3$ by adding pairs of elements corresponding to n_{ij} and n_{ji}.

If some $n_{ij} + n_{ji}$ $(i > j) = 0$, condense equation (3) by deleting the rows from $[\underline{C}_{05}' \ \underline{C}_{45}' \ \underline{C}_{55}]$ and similarly for the corresponding columns of

$$\begin{pmatrix} \underline{C}_{05} \\ \underline{C}_{45} \\ \underline{C}_{55} \end{pmatrix}$$

and elements of $\underline{t}_5$.

Now, the only estimable functions of $\underline{g}$ and $\underline{f}$ are linear functions of $g_i + h_{ij}$ for which $n_{ij} + n_{ji} > 0$.

4. Lines Random

Now we assume that a random set of lines has been derived from the same original population, that these lines all have the same inbreeding coefficients, and a random set of progeny is obtained from single crosses among these lines.

4.1. Equivalent models

A number of different models could be written, all of them equivalent to a basic model as follows:

$$y_{ijk} = \sum_{h=1}^{p} x_{hijk}\beta_h + u_{ij} + e_{ijk} .$$

The u_{ij} all have means = 0 and the following logical covariance structure, provided the lines have the sam inbreeding coefficients and are unrelated:

Var $(u_{ij}) = c_1$ for all i,j;

Cov $(u_{ij}, u_{ji}) = c_2$
= covariance between reciprocals;

Cov $(u_{ij}, u_{ih}) = c_3$ for $j \neq h$
= covariance between crosses with the same line of sire but different line of dam;

Cov $(u_{ij}, u_{hj}) = c_4$ for $i \neq h$
= covariance between crosses with the same line of dam but different line of sire;

Cov $(u_{ij}, u_{hi}) = c_5$ for $j \neq h$
= covariance between crosses with the line of sire of one the same as the line of dam of the other;

$\text{Cov}\,(u_{ij}, u_{hl}) = 0$ for $i \neq h, l$ and $j \neq h, l$.

Now to correspond with the elements of the basic model, let

$$u_{ij} + e_{ijk} = s_i + d_j + r_{ij} + e_{ijk}\,.$$

Assume that these have zero means and the following variance-covariance matrix:

$$\text{Var}\begin{pmatrix}\underline{s}\\ \underline{d}\\ \underline{r}\end{pmatrix} = \begin{pmatrix}\underline{I}\sigma_s^2 & \underline{I}\sigma_{sd} & \underline{0} & \underline{0}\\ \underline{I}\sigma_{sd} & \underline{I}\sigma_d^2 & \underline{0} & \underline{0}\\ \underline{0} & \underline{0} & \underline{V} & \underline{0}\\ \underline{0} & \underline{0} & \underline{0} & \underline{I}\sigma_e^2\end{pmatrix}$$

where $\underline{V}$ has diagonal elements = σ_r^2, and all off-diagonal elements are 0 except for those pertaining to the intersection of r_{ij} and r_{ji}, which have the value $\sigma_{rr'}$. That is, $\text{Cov}\,(r_{ij}, r_{ji}) = \sigma_{rr'}$. Then, $c_1 = \sigma_s^2 + \sigma_d^2 + \sigma_r^2$; $c_2 = 2\sigma_{sd} + \sigma_{rr'}$; $c_3 = \sigma_s^2$; $c_4 = \sigma_d^2$; and $c_5 = \sigma_{sd}$.

An equivalent model that has been used more commonly (e.g., Henderson 1952) is

$$u_{ij} + e_{ijk} = g_i + g_j + m_j + c_{ij} + h_{ij} + e_{ijk}$$

where $i < j$ and

$$u_{ji} + e_{jik} = g_j + g_i + m_i + c_{ij} + h_{ji} + e_{jik}$$

where $j > i$. That is, the first subscript on c is always less than the second. All of the variables are assumed to have zero means and variance-covariance matrix as follows:

$$\text{Var}\begin{pmatrix}\underline{g}\\ \underline{m}\\ \underline{c}\\ \underline{h}\\ \underline{e}\end{pmatrix} = \begin{pmatrix}\underline{I}\sigma_g^2 & \underline{I}\sigma_{gm} & \underline{0} & \underline{0} & \underline{0}\\ \underline{I}\sigma_{gm} & \underline{I}\sigma_g^2 & \underline{0} & \underline{0} & \underline{0}\\ \underline{0} & \underline{0} & \underline{I}\sigma_s^2 & \underline{0} & \underline{0}\\ \underline{0} & \underline{0} & \underline{0} & \underline{I}\sigma_r^2 & \underline{0}\\ \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{I}\sigma_e^2\end{pmatrix}$$

So far as I am aware, the author is the only one who has suggested a σ_{gm} term (Henderson 1952). It is needed, however, if this model is to be equivalent to the basic u_{ij} model.

Now,

$$c_1 = 2\sigma_g^2 + \sigma_m^2 + 2\sigma_{gm} + \sigma_c^2 + \sigma_h^2\,;$$

$$c_2 = 2\sigma_g^2 + 2\sigma_{gm} + \sigma_c^2\,;$$

$$c_3 = \sigma_g^2\,;$$

$$c_4 = \sigma_g^2 + \sigma_m^2 + 2\sigma_{gm}\,;$$

$$c_5 = \sigma_g^2 + \sigma_{gm}\,.$$

The relationships between the parameters of the model of this paper and the parameters of the 1952 paper are as follows:

<u>Parameters of this Paper</u> <u>Parameters of 1952 Model</u>

$$\begin{pmatrix}\sigma_s^2\\ \sigma_d^2\\ \sigma_{sd}\\ \sigma_r^2\\ \sigma_{rr'}\end{pmatrix} = \begin{pmatrix}\sigma_g^2\\ \sigma_g^2 + 2\sigma_{gm} + \sigma_m^2\\ \sigma_g^2 + \sigma_{gm}\\ \sigma_c^2 + \sigma_h^2\\ \sigma_c^2\end{pmatrix}$$

Certain variances of these models can be expressed as linear functions of the basic genetic parameters of the population. Let σ_{ij}^2 be the variance contributed to some trait by the interaction of i nonallelic genes and j gene pairs in a noninbred population and ignore linkage. Then, the genetic variance for the trait in a noninbred population is

$$\sum_{i=0}^{n}\sum_{j=0}^{n}\sigma_{ij}^2\,,$$

excluding $i = j = 0$ and n = the total number of loci contributing to the variance of the trait. For further discussion of this partitioning, see Cockerham (1963).

Then, σ_s^2 (of this paper) = σ_g^2 (1952 model) =

$$\sum_{i=1}^{n}\frac{1}{2^i}f^i\sigma_{i0}^2$$

where f is the inbreeding coefficient of the population. $\sigma_{rr'}$ (of this paper) = σ_c^2 (1952 model) =

$$\sum_{i=2}^{n}\frac{2^{i-1}-1}{2^{i-1}}f^i\sigma_{i0}^2 + \sum_{i=0}^{n}\sum_{j=1}^{n}f^{i+2p}\sigma_{ij}^2\,.$$

4.2. Prediction

Now we can find the best linear unbiased predictors of $\underline{d}$, $\underline{s}$, and $\underline{r}$ or any linear function of these variables. For example, the BLUP (best linear unbiased predictor) of the cross of line of sire i by line of dam j is $\hat{s}_i + \hat{d}_j + \hat{r}_{ij}$. The BLUP of the relative merits of sire lines in crosses with a random sample of the population from which the lines were derived is $\hat{s}_i$.

The predictors are obtained by modifying equation (1) as follows:

$$\text{to} \begin{pmatrix} \underline{C}_{11} & \underline{C}_{12} \\ \underline{C}'_{12} & \underline{C}_{22} \end{pmatrix},$$

$$\text{add} \begin{pmatrix} \underline{I}\sigma_s^2 & \underline{I}\sigma_{sd} \\ \underline{I}\sigma_{sd} & \underline{I}\sigma_d^2 \end{pmatrix}^{-1} \sigma_e^2 .$$

The derivation of BLUP is described in Henderson (1963).

If $n_{ij} = n_{ji} = 0$, the predictors of r_{ij} and r_{ji} are 0. If, however, either n_{ij} or $n_{ji} > 0$, nontrivial predictors of r_{ij} and r_{ji} are obtained.

Prediction error variances can be obtained from a g-inverse of the coefficient matrix (Henderson 1975).

4.3. Special case in which $s_i = d_i$ and $r_{ij} = r_{ji}$

Let the model be described as in Section 3.1. The distributional properties of the variables are

$$\text{Var} \begin{pmatrix} \underline{g} \\ \underline{h} \\ \underline{e} \end{pmatrix} = \begin{pmatrix} \underline{I}\sigma_g^2 & \underline{0} & \underline{0} \\ \underline{0} & \underline{I}\sigma_h^2 & \underline{0} \\ \underline{0} & \underline{0} & \underline{I}\sigma_e^2 \end{pmatrix}.$$

Then BLUP of $\underline{g}$ and $\underline{h}$ or any linear functions of them is obtained by solving equations (3) modified as follows:

to $\underline{C}_{44}$, add $\underline{I}\sigma_e^2/\sigma_g^2$;

to $\underline{C}_{55}$, add $\underline{I}\sigma_e^2/\sigma_h^2$.

5. Estimation of Variances and Covariances

The logical estimator of σ_e^2 is

$$\hat{\sigma}_e^2 = (\underline{y}'\underline{y} - \hat{\underline{\beta}}'\underline{t}_0 - \hat{\underline{r}}'\underline{t}_3)/[n\text{-rank of the coefficient matrix of (2)}]$$

where $\hat{\underline{\beta}}$ and $\hat{\underline{r}}$ are any solutions to (2) and n is the total number of observations.

To estimate the other variances and covariances, we compute the following reductions in sums of squares and equate to their expectations: R($\underline{\beta}$, $\underline{s}$, $\underline{d}$, $\underline{r}$); R($\underline{\beta}$, $\underline{s}$, $\underline{d}$); R($\underline{\beta}$, $\underline{s}$); R($\underline{\beta}$, $\underline{d}$); R($\underline{\beta}$, $\underline{g}$, $\underline{h}$); and R($\underline{\beta}$, $\underline{g}$). One would use R($\underline{\beta}$) rather than R($\underline{\beta}$, $\underline{s}$, $\underline{d}$). The expectations of each of these reductions contain $\underline{\beta}'\ \underline{X}'\ \underline{X}\underline{\beta}$.

R($\underline{\beta}$, $\underline{s}$, $\underline{d}$, $\underline{r}$) can be found by solving

$$\begin{pmatrix} \underline{C}_{00} & \underline{C}_{03} \\ \underline{C}'_{03} & \underline{C}_{33} \end{pmatrix} \begin{pmatrix} \hat{\underline{\beta}} \\ \hat{\underline{r}} \end{pmatrix} = \begin{pmatrix} \underline{t}_0 \\ \underline{t}_3 \end{pmatrix}$$

and then computing $\hat{\underline{\beta}}'\underline{t}_0 + \hat{\underline{r}}'\underline{t}_3$. R($\underline{\beta}$, $\underline{s}$, $\underline{d}$) can be found by solving

$$\begin{pmatrix} \underline{C}_{00} & \underline{C}_{01} & \underline{C}_{02} \\ \underline{C}'_{01} & \underline{C}_{11} & \underline{C}_{12} \\ \underline{C}'_{02} & \underline{C}'_{12} & \underline{C}_{22} \end{pmatrix} \begin{pmatrix} \hat{\underline{\beta}} \\ \hat{\underline{s}} \\ \hat{\underline{d}} \end{pmatrix} = \begin{pmatrix} \underline{t}_0 \\ \underline{t}_1 \\ \underline{t}_2 \end{pmatrix}$$

and computing $\hat{\underline{\beta}}'\underline{t}_0 + \hat{\underline{s}}'\underline{t}_1 + \hat{\underline{d}}'\underline{t}_2$. R($\underline{\beta}$, $\underline{s}$) can be found by solving

$$\begin{pmatrix} \underline{C}_{00} & \underline{C}_{01} \\ \underline{C}'_{01} & \underline{C}_{11} \end{pmatrix} \begin{pmatrix} \hat{\underline{\beta}} \\ \hat{\underline{s}} \end{pmatrix} = \begin{pmatrix} \underline{t}_0 \\ \underline{t}_1 \end{pmatrix}$$

and computing $\hat{\underline{\beta}}'\underline{t}_0 + \hat{\underline{s}}'\underline{t}_1$. R($\underline{\beta}$, $\underline{d}$) can be found by solving

$$\begin{pmatrix} \underline{C}_{00} & \underline{C}_{02} \\ \underline{C}'_{02} & \underline{C}_{22} \end{pmatrix} \begin{pmatrix} \hat{\underline{\beta}} \\ \hat{\underline{d}} \end{pmatrix} = \begin{pmatrix} \underline{t}_0 \\ \underline{t}_2 \end{pmatrix}$$

and computing $\hat{\underline{\beta}}'\underline{t}_0 + \hat{\underline{d}}'\underline{t}_2$. All of the $\underline{C}_{ij}$ above are as defined in (1) but modified for missing observations.

$R(\underline{\beta}, \underline{g}, \underline{h})$ can be found by solving

$$\begin{pmatrix} \underline{C}_{00} & \underline{C}_{05} \\ \underline{C}'_{05} & \underline{C}_{55} \end{pmatrix} \begin{pmatrix} \hat{\underline{\beta}} \\ \hat{\underline{h}} \end{pmatrix} = \begin{pmatrix} \underline{t}_0 \\ \underline{t}_5 \end{pmatrix}$$

and computing $\hat{\underline{\beta}}'\underline{t}_0 + \hat{\underline{h}}'\underline{t}_5$. $R(\underline{\beta}, \underline{g})$ can be found by solving

$$\begin{pmatrix} \underline{C}_{00} & \underline{C}_{04} \\ \underline{C}'_{04} & \underline{C}_{44} \end{pmatrix} \begin{pmatrix} \hat{\underline{\beta}} \\ \hat{\underline{g}} \end{pmatrix} = \begin{pmatrix} \underline{t}_0 \\ \underline{t}_4 \end{pmatrix}$$

and computing $\hat{\underline{\beta}}'\underline{t}_0 + \hat{\underline{g}}'\underline{t}_4$. The $\underline{C}_{ij}$ in the two sets of equations above are defined as in (3) but modified for cases of $n_{ij} + n_{ji} = 0$.

As was stated before, $\underline{\beta}'\underline{X}'\underline{X}\underline{\beta}$ appears in the expectation of each of these reductions. The coefficient of σ_e^2 is the rank of the coefficient matrix used in computing the reduction. For example, the coefficient of σ_e^2 in $R(\underline{\beta}, \underline{s})$ is

$$\text{rank} \begin{pmatrix} \underline{C}_{00} & \underline{C}_{01} \\ \underline{C}'_{01} & \underline{C}_{11} \end{pmatrix} .$$

To find the coefficients of σ_s^2, σ_d^2, σ_{sd}, σ_r^2, $\sigma_{rr'}$ in any reduction, denote the equations for computing the reductions as $\underline{C}\underline{\alpha} = \underline{t}$. Now compute the coefficients of $\underline{s}$, $\underline{d}$, and $\underline{r}$ in $E(\underline{t})$, regarding $\underline{s}$, $\underline{d}$, and $\underline{r}$ as fixed for purposes of taking expectations. Denote this expectation by $\underline{B}_1\underline{s} + \underline{B}_2\underline{d} + \underline{B}_3\underline{r}$. Let $\underline{C}^-$ be some g-inverse of $\underline{C}$. Then the coefficients in the expectations of the reductions are

$\text{tr}\underline{C}^-\underline{B}_1\underline{B}'_1$ for σ_s^2 ;

$\text{tr}\underline{C}^-\underline{B}_2\underline{B}'_2$ for σ_d^2 ;

$2\text{tr}\underline{C}^-\underline{B}_1\underline{B}'_2$ for σ_{sd} ;

$\text{tr}\underline{C}^-\underline{B}_3\underline{B}'_3$ for σ_r^2 ;

$\text{tr}\underline{C}^-\underline{B}_3\underline{Q}\underline{B}'_3$ for $\sigma_{rr'}$.

$\underline{Q}$ is a square matrix of the same order as the number of filled subclasses (number of elements of $\underline{r}$ in the least squares equations). All elements are 0 except for 1's in the intersection of r_{ij} with r_{ji}.

Certain of the coefficients need not be computed as described above but can be written directly. These are

$R(\underline{\beta}, \underline{s}, \underline{d}, \underline{r})$: coefficients of σ_r^2, σ_s^2, $\sigma_d^2 = n$

coefficients of $\sigma_{rr'}$, $\sigma_{sd} = 0$;

$R(\underline{\beta}, \underline{s}, \underline{d})$: coefficients of σ_s^2, $\sigma_d^2 = n$

coefficient of $\sigma_{sd} = 0$;

$R(\underline{\beta}, \underline{s})$: coefficient of $\sigma_s^2 = n$

coefficient of $\sigma_{sd} = 0$;

$R(\underline{\beta}, \underline{d})$: coefficient of $\sigma_d^2 = n$

coefficient of $\sigma_{sd} = 0$.

$R(\underline{\beta}, \underline{g}, \underline{h})$ and $R(\underline{\beta}, g)$: coefficient of σ_s^2, σ_d^2

$\sigma_s^2 + \frac{1}{2}$ coefficient of $\sigma_{sd} = n$.

If $\underline{X\beta}$ contains only a common mean, μ, and if the number of observations per off-diagonal subclass is equal, the estimation of variances is simplified greatly. Let the number of such observations per subclass be k. Then, $n = q(q-1)k$. Now the reduction in sums of squares simplify to:

$$R(\underline{\mu}, \underline{s}, \underline{d}, \underline{r}) = \frac{1}{k}\sum_i \sum_j y_{ij.}^2 \ ;$$

$$R(\underline{\mu}, \underline{s}, \underline{d}) = \frac{(q-1)\left(\sum_i y_{i..}^2 + \sum_j y_{.j.}^2\right) + 2\sum y_{i..}y_{.j.} - 2y_{...}^2}{q(q-2)k} + \frac{y_{...}^2}{q(q-1)k} \ ;$$

$$R(\underline{\mu}, \underline{s}) = \frac{1}{(q-1)k}\sum_i y_{i..}^2 \ ;$$

$$R(\underline{\mu}, \underline{d}) = \frac{1}{(q-1)k}\sum_j y_{.j.}^2 \ ;$$

$$R(\underline{\mu}, \underline{g}, \underline{h}) = \frac{1}{2k}\sum_i \sum_j (y_{ij.} + y_{ji.})^2 \ ;$$

$$R(\underline{\mu}, \underline{g}) = \frac{1}{2k(q-1)(q-2)}\left[(q-1)\sum_i (y_{i..} + y_{.i.})^2 - 2y_{..}^2\right].$$

The expectations of these reductions are

$$R(\underline{\mu},\underline{s},\underline{d},\underline{r}):\; q(q-1)k(\sigma_s^2+\sigma_d^2+\sigma_r^2) + q(q-1)\sigma_e^2\;;$$

$$R(\underline{\mu},\underline{s},\underline{d}):\; q(q-1)k(\sigma_s^2+\sigma_d^2) + 2(q-1)k\sigma_r^2 + \sigma_{rr'} + 2(q-1)\sigma_e^2\;;$$

$$R(\underline{\mu},\underline{s}):\; q(q-1)k\sigma_s^2 + qk(\sigma_d^2+\sigma_r^2) + q\sigma_e^2\;;$$

$$R(\underline{\mu},\underline{d}):\; q(q-1)k\sigma_d^2 + qk(\sigma_s^2+\sigma_r^2) + q\sigma_e^2\;;$$

$$R(\underline{\mu},\underline{g},\underline{h}):\; \frac{q(q-1)k}{2}(\sigma_s^2+\sigma_d^2+2\sigma_{sd}) + \frac{q(q-1)k}{2}(\sigma_r^2+\sigma_{rr'}) + \frac{q(q-1)k}{2}\sigma_e^2\;;$$

$$R(\underline{\mu},\underline{g}):\; \frac{q(q-1)k}{2}(\sigma_s^2+\sigma_d^2+2\sigma_{sd}) + qk(\sigma_r^2+\sigma_{rr'}) + q\sigma_e^2\,.$$

6. Illustration

Suppose we have drawn a sample of four lines from some population and the number of cross-line progeny is as follows:

	Line of Dam			
Sire Line	1	2	3	4
1	0	5	2	9
2	3	0	0	6
3	4	7	0	8
4	8	2	9	0

The sums of the observations for each subclass are

	Line of Dam			
Sire Line	1	2	3	4
1	--	74	25	161
2	40	--	--	71
3	42	64	--	65
4	57	21	143	--

There is a covariate associated with each observation, and the subclass totals for the covariate are

	Line of Dam			
Sire Line	1	2	3	4
1	--	24	7	35
2	6	--	--	34
3	17	20	--	41
4	25	5	53	--

The model is

$$\mu + \gamma x_{ijk} + s_i + d_j + r_{ij} + e_{ijk}\,.$$

$$\sum_i\sum_j\sum_k x_{ijk}^2 = 1527\;;$$

$$\sum_i\sum_j\sum_k x_{ijk}y_{ijk} = 3943\,.$$

6.1. Least squares equations

Then the equations like (1) are as follows:

$$\underline{\beta}' = [\mu\;\;\gamma];$$

$$\underline{s}' = [s_1 \ldots s_4];$$

$$\underline{d}' = [d_1 \ldots d_4];$$

$$\underline{r}' = [r_{12} \ldots r_{34}];$$

$$\underline{C}_{00} = \begin{pmatrix} 63 & 267 \\ 267 & 1527 \end{pmatrix};$$

$$\underline{C}_{01} = \begin{pmatrix} 16 & 9 & 19 & 19 \\ 66 & 40 & 78 & 83 \end{pmatrix};$$

$$\underline{C}_{02} = \begin{pmatrix} 15 & 14 & 11 & 23 \\ 48 & 49 & 60 & 110 \end{pmatrix};$$

$$\underline{C}_{03} = \begin{pmatrix} 5 & 2 & 9 & 3 & 0 & 6 & 4 & 7 & 8 & 8 & 2 & 9 \\ 24 & 7 & 35 & 6 & 0 & 34 & 17 & 20 & 41 & 25 & 5 & 53 \end{pmatrix};$$

$$\underline{C}_{11} = \begin{pmatrix} 16 & 0 & 0 & 0 \\ 0 & 9 & 0 & 0 \\ 0 & 0 & 19 & 0 \\ 0 & 0 & 0 & 19 \end{pmatrix};$$

$$\underline{C}_{12} = \begin{pmatrix} 0 & 5 & 2 & 9 \\ 3 & 0 & 0 & 6 \\ 4 & 7 & 0 & 8 \\ 8 & 2 & 9 & 0 \end{pmatrix};$$

$$\underline{C}_{13} = \begin{pmatrix} 5 & 2 & 9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 6 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 4 & 7 & 8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 8 & 2 & 9 \end{pmatrix};$$

$$\underline{C}_{22} = \begin{pmatrix} 15 & 0 & 0 & 0 \\ 0 & 14 & 0 & 0 \\ 0 & 0 & 11 & 0 \\ 0 & 0 & 0 & 23 \end{pmatrix};$$

$$\underline{C}_{23} = \begin{pmatrix} 0 & 0 & 0 & 3 & 0 & 0 & 4 & 0 & 0 & 8 & 0 & 0 \\ 5 & 0 & 0 & 0 & 0 & 0 & 0 & 7 & 0 & 0 & 2 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 9 \\ 0 & 0 & 9 & 0 & 0 & 6 & 0 & 0 & 8 & 0 & 0 & 0 \end{pmatrix};$$

$\underline{C}_{33}$ = diagonal matrix with diagonals =

[5 2 9 3 0 6 4 7 8 8 2 9];

$\underline{t}_0' = [763 \quad 3943]$;

$\underline{t}_1' = [260 \quad 111 \quad 171 \quad 221]$;

$\underline{t}_2' = [139 \quad 159 \quad 168 \quad 297]$;

$\underline{t}_3' = [74\ 25\ 161\ 40\ 0\ 71\ 42\ 64\ 65\ 57\ 21\ 143]$.

6.2. Prediction

Now suppose the relative values of σ_s^2, σ_d^2, σ_{sd}, σ_r^2, $\sigma_{rr'}$, and σ_e^2 are .5, .6, .4, .2, .1, and 1.0, respectively. Then to

$$\begin{pmatrix} \underline{C}_{11} & \underline{C}_{12} \\ \underline{C}_{12}' & \underline{C}_{22} \end{pmatrix},$$

add

$$\begin{pmatrix} .5\underline{I} & .4\underline{I} \\ .4\underline{I} & .6\underline{I} \end{pmatrix}^{-1} = \begin{pmatrix} 4.286\underline{I} & -2.857\underline{I} \\ -2.857\underline{I} & 3.571\underline{I} \end{pmatrix};$$

to $\underline{C}_{33}$, add

$$\begin{pmatrix} .2 & 0 & 0 & .1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & .2 & 0 & 0 & 0 & 0 & .1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & .2 & 0 & 0 & 0 & 0 & 0 & 0 & .1 & 0 & 0 \\ .1 & 0 & 0 & .2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .2 & 0 & 0 & .1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & .2 & 0 & 0 & 0 & 0 & .1 & 0 \\ 0 & .1 & 0 & 0 & 0 & 0 & .2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .1 & 0 & 0 & .2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .2 & 0 & 0 & .1 \\ 0 & 0 & .1 & 0 & 0 & 0 & 0 & 0 & 0 & .2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & .1 & 0 & 0 & 0 & 0 & .2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .1 & 0 & 0 & .2 \end{pmatrix}^{-1}.$$

The solution is

$[\hat{\mu}, \hat{\gamma}] = [4.215 \quad 1.939]$;

$\hat{\underline{s}}' = [2.774 \quad .433 \quad -2.054 \quad -1.153]$;

$\hat{\underline{d}}' = [.645 \quad .416 \quad .067 \quad -1.128]$;

$\hat{\underline{r}}' = [-.499 \quad -.441 \quad 2.388 \quad 1.120 \quad .300 \quad -1.261 \quad -.342 \quad .599 \quad -1.445 \quad -.898 \quad .017 \quad .464]$.

6.3. Estimation of variances: Red($\underline{\beta}$, $\underline{s}$, $\underline{d}$, $\underline{r}$)

The solution is $\hat{\gamma} = 2.053$, $\hat{\mu} = 0$, and

$\hat{\underline{r}}' = [4.946 \quad 5.315 \quad 9.905 \quad 9.228 \quad .200 \quad 1.775 \quad 3.277 \quad -2.396 \quad .709 \quad 5.368 \quad 3.799]$ with $\hat{r}_{23}$ deleted.

Reduction = 11,396.64 and expectation = $12\sigma_e^2 + 63(\sigma_r^2 + \sigma_s^2 + \sigma_d^2) + \underline{\beta}'\underline{X}'\underline{X}\underline{\beta}$.

6.4. Estimation of variances: Red($\underline{\beta}$, $\underline{g}$, $\underline{h}$)

Equations to solve are

$$\begin{pmatrix} 1527 & 30 & 24 & 60 & 20 & 39 & 94 \\ 30 & 8 & 0 & 0 & 0 & 0 & 0 \\ 24 & 0 & 6 & 0 & 0 & 0 & 0 \\ 60 & 0 & 0 & 17 & 0 & 0 & 0 \\ 20 & 0 & 0 & 0 & 7 & 0 & 0 \\ 39 & 0 & 0 & 0 & 0 & 8 & 0 \\ 94 & 0 & 0 & 0 & 0 & 0 & 17 \end{pmatrix} \begin{pmatrix} \hat{\gamma} \\ \hat{h}_{12} \\ \hat{h}_{13} \\ \hat{h}_{14} \\ \hat{h}_{23} \\ \hat{h}_{24} \\ \hat{h}_{34} \end{pmatrix} = \begin{pmatrix} 3943 \\ 114 \\ 67 \\ 218 \\ 64 \\ 92 \\ 208 \end{pmatrix}.$$

Table 1. Reductions and their expectations

	Reductions	Coefficients in Expectations						
		σ_e^2	σ_r^2	$\sigma_{rr'}$	σ_s^2	σ_d^2	σ_{sd}	$\underline{\beta}'\underline{X}'\underline{X}\underline{\beta}$
$\underline{y}'\underline{y}$	14,010	63	63	0	63	63	0	1
$(\underline{\beta}, \underline{s}, \underline{d}, \underline{r})$	11,396.64	12	63	0	63	63	0	1
$(\underline{\beta}, \underline{s}, \underline{d})$	11,023.57	8	44.62	3.78	63	63	0	1
$(\underline{\beta}, \underline{s})$	10,981.51	5	27.95	0.41	63	28.27	0	1
$(\underline{\beta}, \underline{d})$	10,544.67	5	27.76	-0.20	27.22	63	0	1
$(\underline{\beta}, \underline{g}, \underline{h})$	10,784.79	7	36.24	25.21	36.89	36.89	52.20	1
$(\underline{\beta}, \underline{g})$	10,746.09	5	26.33	19.21	34.77	34.77	56.45	1

The solution is [2.051 6.560 2.964 5.586 3.284 1.503 .896]. Reduction = 10,784.79. Taking expectations of right-hand sides under a *fixed* model,

$$\underline{B}_1 = \begin{pmatrix} 66 & 40 & 78 & 83 \\ 5 & 3 & 0 & 0 \\ 2 & 0 & 4 & 0 \\ 9 & 0 & 0 & 8 \\ 0 & 0 & 7 & 0 \\ 0 & 6 & 0 & 2 \\ 0 & 0 & 8 & 9 \end{pmatrix};$$

$$\underline{B}_2 = \begin{pmatrix} 48 & 49 & 60 & 110 \\ 3 & 5 & 0 & 0 \\ 4 & 0 & 2 & 0 \\ 8 & 0 & 0 & 9 \\ 0 & 7 & 0 & 0 \\ 0 & 2 & 0 & 6 \\ 0 & 0 & 9 & 8 \end{pmatrix};$$

$$\underline{B}_3 = \begin{pmatrix} 24 & 7 & 35 & 6 & 34 & 17 & 20 & 41 & 25 & 5 & 53 \\ 5 & 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 9 & 0 & 0 & 0 & 0 & 0 & 8 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 7 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 6 & 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 8 & 0 & 0 & 9 \end{pmatrix} \text{ with } r_{23} \text{ deleted;}$$

$$\underline{Q} = \begin{pmatrix} 0&0&0&1&0&0&0&0&0&0&0 \\ 0&0&0&0&0&1&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&1&0&0 \\ 1&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&1&0 \\ 0&1&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&1 \\ 0&0&1&0&0&0&0&0&0&0&0 \\ 0&0&0&0&1&0&0&0&0&0&0 \\ 0&0&0&0&0&0&1&0&0&0&0 \end{pmatrix} \text{ with } r_{23} \text{ deleted.}$$

Then, $E(\text{Reduction}) = 7\sigma_e^2 + 36.24\sigma_r^2 + 25.21\sigma_{rr'} + 36.89\sigma_s^2 + 36.89\sigma_d^2 + 2(26.11)\sigma_{sd}$.

6.5. Estimation of variances: other reductions and expectations

Calculating other reductions and expectations and with $\underline{y}'\underline{y}$ = 14,010, Table 1 results.

Solving for the unknown variances and covariances, we obtain $\hat{\sigma}_e^2 = 51.24$, $\hat{\sigma}_r^2 = 5.61$, $\hat{\sigma}_{rr'} = -16.96$, $\hat{\sigma}_s^2 = 8.31$, $\hat{\sigma}_d^2 = -4.28$, and $\hat{\sigma}_{sd} = 6.29$. Obviously, these are not acceptable values, but this is not surprising considering the small sample of this example. Many more than four lines are needed for variance component estimation.

Acknowledgement

This research was supported in part by the Eastern Artificial Insemination Cooperative, Inc., Ithaca, New York.

Literature

Cockerham, C.C.: Estimation of genetic variance. In: Statistical Genetics and Plant Breeding, pp. 53-93. Washington: National Academy of Sciences-National Research Council (Pub. 982) 1963

Henderson, C.R.: Specific and general combining ability. In: Heterosis, pp. 352-370. Ames: Iowa State College Press 1952

Henderson, C.R.: Selection index and expected genetic advance. In: Statistical Genetics and Plant Breeding. Washington: National Academy of Sciences--National Research Council (Pub. 982) 1963

Henderson, C.R.: Best linear unbiased estimation and prediction under a selection model. Biometrics 31, 423 (1975)

Received October 14, 1976
Communicated by L.D. Van Vleck

C.R. Henderson
Dept. of Animal Science
Cornell University
Ithaca, New York 14850 (U.S.A.)