

Estimation in the General Incomplete Block Design with Applications to the Analysis of Mating Structures

THOMAS R. BEMENT, V. JAYARAMA KRISHNA, GEORGE A. MILLIKEN and ROBERT R. SCHALLES

Kansas State University, Manhattan, Kansas (USA)

Summary. Models for analysing data for two commonly used sampling schemes involving inbreeding and environmental effects is considered. The general incomplete block model is utilized where mating types are equated to incomplete blocks. The models allow for varying degrees of inbreeding within mating types and between mating types. The analysis is such as to remove some of the confounding of inbreeding and environmental effects. The models are formulated in matrix notation utilizing pseudoinverses of matrices. Intrablock and interblock estimates of environmental effects or treatment effects are obtained with a suggestion as how one might combine the two estimates. An example is presented to demonstrate the computations.

1. Introduction

Examination of the breeding structure in a livestock population reveals different forms of departure from the random mating scheme, perhaps because of planned consanguineous matings or extensive use of proven sires in small populations. The type of sampling scheme adapted to select progeny for performance testing will influence the choice of mating design and thus, the model to be fitted. Also, as the environment changes, inbreeding increases; therefore inbreeding and year or environmental effects are confounded. That phenomenon must also be accounted for by choice of the model. We here examined the characteristics of two commonly used sampling schemes.

1.1. One Offspring from each Mating

In this situation a given combination of ♂ × ♀ (male by female), a mating type, yields only one progeny. The progeny are identified as $M_1P \dots M_nP$,

Contribution No. 221 of the Department of Statistics, Statistical Laboratory, and contribution No. 463 of the Department of Animal Science and Industry, Kansas Agricultural Experiment Station, Kansas State University, Manhattan, Kansas 66506. Contributing Regional Project NC-1.

where M_1P denotes the progeny from mating 1, ..., and M_nP denotes the progeny from mating n . The variance of a resulting progeny is weighted by $1 - f_i^2$ where f_i denotes the inbreeding coefficient of the i th progeny computed from Wright's path coefficients. That weighting procedure unconfounds some of the environmental and inbreeding effects and enables a less-biased estimate of σ^2 to be obtained. Thus, the variance of a progeny is $\sigma^2(1 - f_i^2)$ where σ^2 is the variance of noninbred progeny and the covariance between two progeny is denoted by $\text{Cov } M_iM_j$, which is computed from the genetic relationship between mating types M_i and M_j (Table 1). The resulting covariance structure of the progeny depends on which type of mating scheme is used. We discuss two mating schemes.

Case 1: Random Mating Population

If the progeny are derived from a random mating population, then $f_i = 0$, $\text{Cov } M_iM_j = 0$, and the covariance structure of the progeny (Table 1) has the form of an identity matrix denoted by I .

Case 2: Nonrandom Mating Population

If the progeny are from a nonrandom mating population, then the mates have co-ancestry and the pro-

Table 1. Relationship between progeny

Progeny Title	M_1P	M_2P	... M_jP	... M_nP
M_1P	$\sigma^2(1 - f_1^2)$	$\text{Cov } M_1M_2 \dots$	$\text{Cov } M_1M_j \dots$	$\text{Cov } M_1M_n$
M_2P	$\text{Cov } M_2M_1$	$\sigma^2(1 - f_2^2) \dots$	$\text{Cov } M_2M_j \dots$	$\text{Cov } M_2M_n$
\vdots	\vdots	\vdots	\vdots	\vdots
M_jP	$\text{Cov } M_jM_1$	$\text{Cov } M_jM_2 \dots$	$\sigma^2(1 - f_j^2) \dots$	$\text{Cov } M_jM_n$
\vdots	\vdots	\vdots	\vdots	\vdots
M_nP	$\text{Cov } M_nM_1$	$\text{Cov } M_nM_2 \dots$	$\text{Cov } M_nM_j \dots$	$\sigma^2(1 - f_n^2)$

geny are inbred. That reduces the genetic variance of an individual, and also some of the mating types may be related; thus, some $Cov M_i M_j \neq 0$.

A linear model for describing the progeny performance can be written

$$y = X_2 \mu + e, \tag{4.0}$$

where y is a $n \times 1$ vector of observable random variables (i.e., values of the trait measured), X_2 is a $n \times t$ design matrix of rank t , denoting the presence or absence of a treatment (such a year born, sex, breed, etc.), μ is a $t \times 1$ vector of unknown treatment means, and e is a $n \times 1$ vector of unobservable random variables with mean O and dispersion matrix $\sigma_e^2 C$.

In case 1, $C = I$ and the best linear unbiased estimator (b.l.u.e.) of μ can always be obtained by ordinary least squares as

$$\hat{\mu} = (X_2' X_2)^{-1} X_2' y.$$

In case 2, $C \neq I$. Thus, once the elements of C have been obtained μ can be computed via weighted least squares as $\mu = (X_2' C^{-1} X_2)^{-1} X_2' C^{-1} y$, which is the b.l.u.e. of μ for case 2.

1.2. Two or more Progeny from each Mating

A mating type that occurs more than once in successive breeding seasons or years is designated as a repeat mating and it yields more than one progeny. The progeny from a repeat mating are fullsibs. When different matings give raise to such groups of full sibs, the aggregate of these full sib groups can be arranged as in Table 2, where $M_j P_i$ denotes the i th progeny from the j th mating type. If the progeny are derived from a nonrandom mating population, the inbreeding tends to reduce the variance of individuals within a full sib group. There can be a relationship between mating types and when it occurs it must also be accounted for in the model. Let $M_1 P_1, M_1 P_2, \dots, M_1 P_{n_1}$ denote the n_1 full sibs from mating type 1. The genetic correlations among the full sibs

for mating type 1 are represented in this matrix (where $M_1 P_1, \dots, M_1 P_{n_1}$ are the row and column denotations, meaning mating type 1 — progeny 1, ..., mating type 1 — progeny n_1);

$$\begin{matrix} & M_1 P_1 & M_1 P_2 & \dots & M_1 P_{n_1} \\ \begin{matrix} M_1 P_1 \\ M_1 P_2 \\ \vdots \\ M_1 P_{n_1} \end{matrix} & \begin{bmatrix} 1 & .5(1 + f_1) & \dots & .5(1 + f_1) \\ .5(1 + f_1) & 1 & \dots & .5(1 + f_1) \\ \vdots & \vdots & \ddots & \vdots \\ .5(1 + f_1) & .5(1 + f_1) & \dots & 1 \end{bmatrix} \end{matrix}$$

To reduce the confounding between environment and inbreeding, this matrix is weighted by $\sigma_b^2(1 - f_1^2)$, and the matrix in Table 2 is used in the analysis to account for the within-full-sib-group relationships. Also possible relationships between mating types must be included in the model. The matrix of variances and covariances for mating types is:

$$\begin{matrix} & M_1 & M_2 & \dots & M_n \\ \begin{matrix} M_1 \\ M_2 \\ \vdots \\ M_n \end{matrix} & \begin{bmatrix} \sigma_b^2 & Cov M_1 M_2 & \dots & Cov M_1 M_n \\ Cov M_1 M_2 & \sigma_b^2 & \dots & Cov M_2 M_n \\ \vdots & \vdots & \ddots & \vdots \\ Cov M_n M_1 & Cov M_n M_2 & \dots & \sigma_b^2 \end{bmatrix} \end{matrix} \tag{4.1}$$

where σ_b^2 denotes the variance between unrelated mating types and $Cov M_i M_j$ denotes the genetic relationship between mating types M_i and M_j .

Case 3: If the progeny are derived from a random mating population, then $f_i = 0$ and $Cov M_i M_j = 0$. The diagonal elements of Table 2 will be 1.0 and the off-diagonal elements .50, which are coefficients of relationships among fullsibs. The between mating type covariance structure will be $\sigma_b^2 I_n$ since there is no c-ancestry among mates.

Case 4: If the progeny are derived from a nonrandom mating population, then $f_i > 0$ and $Cov M_j M_1 > 0$, depending on the degree of relationship. Thus, the structure of the matrix in Table 2, as well as the

Table 2. Within mating type genetic relationships

	$M_1 P_1$...	$M_1 P_{n_1}$...	$M_j P_1$...	$M_j P_{n_j}$...	$M_n P_1$...	$M_n P_{n_n}$
$M_1 P_1$	$1 - f_1^2$...	$.5(1 - f_1^2)(1 + f_1)$		0		0		0		0
	\vdots		\vdots		\vdots		\vdots		\vdots		\vdots
$M_1 P_{n_1}$	$.5(1 - f_1^2)(1 + f_1)$...	$1 - f_1^2$		0		0		0		0
	\vdots		\vdots		\vdots		\vdots		\vdots		\vdots
$M_j P_1$	0		0		$1 - f_j^2$...	$.5(1 - f_j^2)(1 + f_j)$		0		0
	\vdots		\vdots		\vdots		\vdots		\vdots		\vdots
$M_j P_{n_j}$	0		0		$.5(1 - f_j^2)(1 + f_j)$...	$1 - f_j^2$		0		0
	\vdots		\vdots		\vdots		\vdots		\vdots		\vdots
$M_n P_1$	0		0		0		0		$1 - f_n^2$...	$.5(1 - f_n^2)(1 + f_n)$
	\vdots		\vdots		\vdots		\vdots		\vdots		\vdots
$M_n P_{n_n}$	0		0		0		0		$.5(1 - f_n^2)(1 + f_n)$...	$1 - f_n^2$

between mating type covariance structure, is changed. As inbreeding increases, the genetic variance within a block or full-sib group shifts proportionately to between blocks or matings. In a completely inbred population there will be no genetic variation within a full-sib group. Because that phenomenon tends to decrease environmental variation, we can use the matrix of within mating type relationships in a weighted analysis to unconfound some of the environmental and inbreeding effects.

A linear model for the general incomplete block design with random block (mating types and blocks are synonymous) effects to describe the progeny performance can be written

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta} + \mathbf{X}_2\boldsymbol{\mu} + \mathbf{e}, \quad (1.2)$$

where

\mathbf{X}_1 is a $n \times b$ design matrix of rank b , denoting the presence of a block.

$\boldsymbol{\beta}$ is a $b \times 1$ vector of unobservable random variables denoting block errors with mean \mathbf{O} and variance $\sigma_b^2 \mathbf{V}_b$, and

$\mathbf{X}_2, \boldsymbol{\mu}, \mathbf{e}$ are as defined in model (1.0).

In this case, \mathbf{V}_b denotes relationships between mating types as exhibited in (1.2). The matrix \mathbf{C} , which consists of the within mating type relationships, is the block diagonal matrix exhibited in Table 2. In addition, the random vectors \mathbf{e} and $\boldsymbol{\beta}$ are assumed to be independently distributed and often it is assumed that \mathbf{e} and $\boldsymbol{\beta}$ are multivariate normal random vectors. We also assume there are k_i experimental units in the i th block.

The problem of interest is to estimate the vector $\boldsymbol{\mu}$. It is well known (see, for example, John 1971 or Graybill 1961) that under conditions imposed on model (1.2) two independent estimators of $\boldsymbol{\mu}$ exist: intrablock and interblock. The intrablock estimator of $\boldsymbol{\mu}$ can be considered to have arisen from a linear model obtained by equating within block deviations from block means to their expectations. Under those conditions and where $\mathbf{C} = \mathbf{I}$, the b.l.u.e. of $\boldsymbol{\mu}$ based on the intrablock model is always available. When $\mathbf{C} \neq \mathbf{I}$, this intrablock model using weighted least squares, (i.e., accounting for the covariance matrix \mathbf{C}), yields the estimators of $\boldsymbol{\mu}$ which are b.l.u.e.

The interblock estimator of $\boldsymbol{\mu}$ is derived from the linear model by equating block totals to their expectations. In Case 3, where $\text{Cov } M_j M_i = 0$ implying that $\mathbf{V}_b = \mathbf{I}$, the b.l.u.e. of $\boldsymbol{\mu}$ can be obtained using this model. In case 4, where $\text{Cov } M_j M_i$ is greater than 0, implying a $\mathbf{V}_b \neq \mathbf{I}$, a straightforward solution for obtaining $\hat{\boldsymbol{\mu}}$ does not exist. This interblock estimator will have greater variance and in general may not be b.l.u.e. When the number of blocks is quite large, such an estimator could be a valuable source of information and should be considered. This paper presents a new matrix formulation method for estimating $\boldsymbol{\mu}$ in (1.2) when \mathbf{V}_b and \mathbf{C} are of the form just

described and also suggests a way to combine the intra and interblock estimators.

2. The Intrablock and Interblock Estimators

When the matrix \mathbf{C} is the identity matrix, the usual intrablock solution for $\boldsymbol{\mu}$ in (1.2) is obtained by solving the intrablock or reduced normal equations for $\boldsymbol{\mu}$. These reduced normal equations are obtained by treating $\boldsymbol{\beta}$ as a vector of fixed effects, using the $\hat{\boldsymbol{\beta}}$ normal equations to solve for $\hat{\boldsymbol{\beta}}$ in terms of $\hat{\boldsymbol{\mu}}$ and substituting this solution into the $\hat{\boldsymbol{\mu}}$ equations. The following theorem (Bement 1972) shows that the technique is equivalent to obtaining $\hat{\boldsymbol{\mu}}$ from the model

$$(\mathbf{I} - \mathbf{X}_1\mathbf{X}_1^{-})\mathbf{y} = (\mathbf{I} - \mathbf{X}_1\mathbf{X}_1^{-})\mathbf{X}_2\boldsymbol{\mu} + \mathbf{e}_1, \quad (2.1)$$

where

$$E(\mathbf{e}_1) = \mathbf{O} \text{ and } \text{Var}(\mathbf{e}_1) = \sigma_e^2(\mathbf{I} - \mathbf{X}_1\mathbf{X}_1^{-}).$$

The $(-)$ superscript on a matrix denotes the Moore-Penrose (see Graybill 1969) generalized inverse of that matrix.

Definition 2.1. Model (2.1) will be called the intrablock model.

Theorem 2.1. The reduced normal equations are the same as the equations for the intrablock model.

Proof: Applying the technique mentioned above for obtaining the reduced normal equations, we write model (1.1) in the form

$$\mathbf{y} = (\mathbf{X}_1, \mathbf{X}_2) \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\mu} \end{bmatrix} + \mathbf{e}.$$

The normal equations are

$$\begin{bmatrix} \mathbf{X}'_1\mathbf{y} \\ \mathbf{X}'_2\mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'_1\mathbf{X}_1 & \mathbf{X}'_1\mathbf{X}_2 \\ \mathbf{X}'_2\mathbf{X}_1 & \mathbf{X}'_2\mathbf{X}_2 \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\mu}} \end{bmatrix}. \quad (2.2)$$

From (2.2), the $\hat{\boldsymbol{\beta}}$ equations are

$$\mathbf{X}'_1\mathbf{X}_1\hat{\boldsymbol{\beta}} + \mathbf{X}'_1\mathbf{X}_2\hat{\boldsymbol{\mu}} = \mathbf{X}'_1\mathbf{y} \quad (2.3)$$

and the $\hat{\boldsymbol{\mu}}$ equations are

$$\mathbf{X}'_2\mathbf{X}_1\hat{\boldsymbol{\beta}} + \mathbf{X}'_2\mathbf{X}_2\hat{\boldsymbol{\mu}} = \mathbf{X}'_2\mathbf{y}. \quad (2.4)$$

Solving for $\hat{\boldsymbol{\beta}}$ in (2.3) we obtain the solution

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'_1\mathbf{X}_1)^{-} \mathbf{X}'_1(\mathbf{y} - \mathbf{X}_2\hat{\boldsymbol{\mu}})$$

or

$$\hat{\boldsymbol{\beta}} = \mathbf{X}_1^{-}(\mathbf{y} - \mathbf{X}_2\hat{\boldsymbol{\mu}}). \quad (2.5)$$

Substituting (2.5) into (2.4) we obtain

$$\mathbf{X}'_2\mathbf{X}_1\mathbf{X}_1^{-}(\mathbf{y} - \mathbf{X}_2\hat{\boldsymbol{\mu}}) + \mathbf{X}'_2\mathbf{X}_2\hat{\boldsymbol{\mu}} = \mathbf{X}'_2\mathbf{y}$$

or

$$\mathbf{X}'_2(\mathbf{I} - \mathbf{X}_1\mathbf{X}_1^{-})\mathbf{X}_2\hat{\boldsymbol{\mu}} = \mathbf{X}'_2(\mathbf{I} - \mathbf{X}_1\mathbf{X}_1^{-})\mathbf{y}. \quad (2.6)$$

But (2.6) is the set of normal equations for the intrablock model, so the proof is complete.

If $\boldsymbol{\lambda}'\boldsymbol{\mu}$ is estimable in (2.1), then it is well known that the b.l.u.e. of $\boldsymbol{\lambda}'\boldsymbol{\mu}$ in that model is $\boldsymbol{\lambda}'\hat{\boldsymbol{\mu}}$ where

$$\begin{aligned} \hat{\boldsymbol{\mu}} &= [(\mathbf{I} - \mathbf{X}_1\mathbf{X}_1^{-})\mathbf{X}_2]^{-}(\mathbf{I} - \mathbf{X}_1\mathbf{X}_1^{-})\mathbf{y} = \\ &= [(\mathbf{I} - \mathbf{X}_1\mathbf{X}_1^{-})\mathbf{X}_2]^{-}\mathbf{y}. \end{aligned} \quad (2.7)$$

The b.l.u.e. of $\lambda'\mu$ can also be expressed as

$$\lambda'\hat{\mu} = \alpha'X_2'(I - X_1X_1^{-1})y,$$

where α is any solution to

$$\lambda = X_2'(I - X_1X_1^{-1})X_2\alpha.$$

From the preceding discussion it is clear that the intrablock estimate of μ in no way depends on the dispersion structure of β .

If the matrix, C , is not the identity matrix, then

$$\text{Var}(e_1) = \sigma_e^2(I - X_1X_1^{-1})C(I - X_1X_1^{-1}) = G,$$

say, and the normal equations from (2.1) will not in general lead to the intrablock b.l.u.e.'s of estimable linear combinations. In this case the structure of G must be taken into account. If $\lambda'\mu$ is estimable in (2.1), its b.l.u.e. in that model is $\lambda'\hat{\mu}$ where

$$\hat{\mu} = [X_2'F'(FCF')^{-1}FX_2]^{-1}X_2'F'(FCF')^{-1}Fy$$

and where the $n - b \times n$ matrix F is defined by

$$F'F = I - X_1X_1^{-1}$$

and

$$FF' = I_{n-b}.$$

The matrix F is of the form

$$F = \begin{bmatrix} F_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ \vdots & F_2 & \vdots \\ \vdots & \vdots & \ddots \\ 0 & \dots & F_b \end{bmatrix}$$

where F_i is $k_i - 1 \times k_i$ and

$$F_i = \begin{bmatrix} 1 & & & & \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & & & \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & & \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \dots & (k-1) \\ \hline \frac{1}{\sqrt{k_i(k_i-1)}} & \frac{1}{\sqrt{k_i(k_i-1)}} & \frac{1}{\sqrt{k_i(k_i-1)}} & \dots & \frac{1}{\sqrt{k_i(k_i-1)}} \end{bmatrix}.$$

Recall that k_i is the number of experimental units in block i or progeny in mating i .

The estimation of the variance component, σ_e^2 will now be considered. We first treat the case where $C = I$.

The error sum of squares for the model (2.1) is

$$Q = y'(I - X_1X_1^{-1})\{I - (I - X_1X_1^{-1})X_2 \times [(I - X_1X_1^{-1})X_2]^{-1}\}(I - X_1X_1^{-1})y = y'(I - XX^{-1})y$$

where

$$X = (X_1, X_2).$$

The expected value of Q is

$$E(Q) = \sigma_e^2 \text{tr}(I - XX^{-1}) = \sigma_e^2(n - R(X)).$$

If the design is connected, then (see John 1971 or Graybill 1961)

$$R(X) = b + t - 1.$$

It follows that an estimate of σ_e^2 is

$$\hat{\sigma}_e^2 = \frac{Q}{n - R(X)}$$

and in case of a connected design,

$$\hat{\sigma}_e^2 = \frac{Q}{n - t - b + 1}.$$

That estimator is simply the error mean square obtained in the usual intrablock analysis.

Next consider the case where $C \neq I$. As before, let

$$Q = y'(I - XX^{-1})y.$$

In this case, the expected value of Q is

$$E(Q) = \sigma_e^2 \text{tr}[(I - XX^{-1})C] = \sigma_e^2\{\text{tr}(C) - \text{tr}(XX^{-1}C)\}.$$

The estimate of σ_e^2 is then

$$\hat{\sigma}_e^2 = \frac{Q}{\text{tr}(C) - \text{tr}(XX^{-1}C)}.$$

Now, let us consider the interblock estimation of μ . The usual interblock estimate comes from the normal equations derived from the model obtained by equating the vector of block totals, $X_1'y$, to its expectation. This model, which we shall call the interblock model, is

$$X_1'y = X_1'X_2\mu + e_2 \tag{2.8}$$

or

$$z = N\mu + e_2$$

where

$$z = X_1'y, \quad N = X_1'X_2, \quad E(e_2) = 0$$

and

$$\text{Var}(e_2) = \Sigma = \sigma_b^2 \text{diag}(k_i) V_b \text{diag}(k_i) + \sigma_e^2 X_1'CX_1.$$

The notation, $\text{diag}(\cdot)$, indicates a diagonal matrix with diagonal elements as given in the argument. N is the transpose of the incidence matrix. The form of Σ follows directly from the fact that $X_1'X_1 = \text{diag}(k_i)$. If $\lambda'\mu$ is estimable in (2.8), its b.l.u.e. is $\lambda'\tilde{\mu}$ where

$$\left. \begin{aligned} \mu &= (\Sigma^{-1/2} X_1'X_2)^{-1} X_2'X_1 \Sigma^{-1} y \\ &= (X_2'X_1 \Sigma^{-1} X_1'X_2)^{-1} X_2'X_1 \Sigma^{-1} X_1'y \\ &= (N' \Sigma^{-1} N)^{-1} N' \Sigma^{-1} z. \end{aligned} \right\} \tag{2.9}$$

If N is a full rank matrix, then $\tilde{\mu}$ can be expressed as

$$\tilde{\mu} = (N' \Sigma^{-1} N)^{-1} N' \Sigma^{-1} z. \tag{2.10}$$

The b.l.u.e. of $\lambda'\mu$ from (2.8) can also be written as

$$\lambda'\tilde{\mu} = \varrho'N \Sigma^{-1} z$$

where ϱ is any solution to

$$\lambda = N' \Sigma^{-1} N\varrho.$$

In general, the interblock b.l.u.e. of μ cannot be calculated because Σ is unknown. In this case we may obtain a weighted interblock estimate of μ by using an estimated Σ . To estimate Σ we must first estimate σ_e^2 and σ_b^2 . The estimate of σ_b^2 obtained in the intrablock analysis can be used here, and the component σ_b^2 can be estimated as follows.

Let the matrix A be defined

$$A = I - NN^{-}$$

Then, with respect to the interblock model,

$$\begin{aligned} E(z'Az) &= \sigma_b^2 \text{tr} [A \text{diag} (k_i) V_b \text{diag} (k_i)] + \\ &+ \sigma_e^2 \text{tr} [AX_1'CX_1] = \\ &= \sigma_b^2 \sum_{i=1}^n \sum_{j=1}^n a_{ij} v_{ij} k_i k_j + \sigma_e^2 \text{tr} [AX_1'CX_1] \end{aligned}$$

where $A = [(a_{ij})]$ and $V_b = [(v_{ij})]$. An estimate of σ_b^2 is then

$$\hat{\sigma}_b^2 = \frac{z'Az - \hat{\sigma}_e^2 \text{tr} [AX_1'CX_1]}{\sum_{i=1}^n \sum_{j=1}^n a_{ij} v_{ij} k_i k_j}$$

where $\hat{\sigma}_e^2$ is obtained from the intrablock analysis. Thus, an estimate of Σ is

$$\hat{\Sigma} = \hat{\sigma}_b^2 \text{diag} (k_i) V_b \text{diag} (k_i) + \hat{\sigma}_e^2 X_1'CX_1$$

A weighted least squares interblock estimator of μ is then

$$\begin{aligned} \tilde{\mu}_{\hat{\Sigma}} &= (\hat{\Sigma}^{-1/2} X_1'X_2)^{-1} X_2'X_1 \hat{\Sigma}^{-1} X_1'y \\ &= (X_2'X_1 \hat{\Sigma}^{-1} X_1'X_2)^{-1} X_2'X_1 \hat{\Sigma}^{-1} X_1'y \end{aligned}$$

Since two estimators of μ are available, they can be combined into a single estimator, which in some cases may have smaller variance than either of the two. Let the covariance matrix of $\hat{\mu}$ be estimated by \hat{M}_1 and the covariance matrix of $\tilde{\mu}_{\hat{\Sigma}}$ be estimated by \tilde{M}_2 . Then, the best combined estimator for μ based upon those estimated weights, is

$$\tilde{\mu}_c = [\hat{M}_1^{-1} + \tilde{M}_2^{-1}]^{-1} (\hat{M}_1^{-1} \hat{\mu} + \tilde{M}_2^{-1} \tilde{\mu}_{\hat{\Sigma}})$$

and has an estimated covariance matrix

$$\widehat{\text{Var}}(\tilde{\mu}_c) = (\hat{M}_1^{-1} + \tilde{M}_2^{-1})^{-1}$$

This particular estimator should contain more information than either $\hat{\mu}$ or $\tilde{\mu}_{\hat{\Sigma}}$, but its exact distributional properties are not known. The combined estimator is presented for completeness.

A computer programme to obtain $\hat{\mu}$ from the intrablock model is available at Kansas State University which can handle 120 levels of any combination of variables in the X_2 matrix. The programme has worked with 315 observations, but can be adapted to handle 500 observations. To obtain $\tilde{\mu}_{\hat{\Sigma}}$ from Interblock model the above programme can be easily modified to fit 120 blocks or matings.

3. Example

We extracted some data from Krishna (1973) to demonstrate the analysis. The data (in Table 3)

Table 3. Data for the example

Sire	Cow	f	Calf	Birthdate	Birthweight
501	309	8.1%	123	4/1/61	72
			231	4/21/62	81
			319	3/29/63	80
501	753	20.0%	039	4/7/60	61
			129	4/14/61	63
			251	6/17/62	72
637	601	34.4%	067	5/29/60	67
			155	5/29/61	69
			379	6/4/63	83
637	29	23.7%	071	6/3/60	63
			149	5/21/61	64
969	23	17.9%	230	3/17/62	65
			323	3/29/63	67
969	703	23.0%	211	4/2/62	72
			351	4/27/63	73

consists of six mating types where some have a common sire. The sires and dams used in the analysis were assumed to be unrelated. The f_i for the calves of a mating are presented and birth weight is the variable to be analyzed. The effects to be estimated are the years in which the calves are born. The following matrices are utilized by the analysis. The matrix corresponding to the mating types is:

$$X_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

and the vector of mating type effects is

$$\beta' = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6)$$

The design matrix denoting the presence or absence of a treatment (in this case, year effect) is

$$X_2 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

and the year effects are

$$\boldsymbol{\mu}' = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\mu}_3, \boldsymbol{\mu}_4), \boldsymbol{\mu}_1 \text{ denotes 1960, } \boldsymbol{\mu}_2 \text{ denotes 1961, } \boldsymbol{\mu}_3 \text{ denotes 1962 and } \boldsymbol{\mu}_4 \text{ denotes 1963.}$$

The within mating type error structure is $\sigma_e^2 \mathbf{C}$ where

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_1 & \mathbf{O} & \dots & \mathbf{O} \\ \mathbf{O} & \mathbf{C}_2 & \dots & \mathbf{O} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{O} & \mathbf{O} & \dots & \mathbf{C}_6 \end{pmatrix}$$

and where

$$\mathbf{C}_1 = \begin{pmatrix} 0.9934 & 0.5370 & 0.5370 \\ 0.5370 & 0.9934 & 0.5370 \\ 0.5370 & 0.5370 & 0.9934 \end{pmatrix}$$

$$\mathbf{C}_2 = \begin{pmatrix} 0.9600 & 0.5760 & 0.5760 \\ 0.5760 & 0.9600 & 0.5760 \\ 0.5760 & 0.5760 & 0.9600 \end{pmatrix}$$

$$\mathbf{C}_3 = \begin{pmatrix} 0.8817 & 0.5925 & 0.5925 \\ 0.5925 & 0.8817 & 0.5925 \\ 0.5925 & 0.5925 & 0.8817 \end{pmatrix}$$

$$\mathbf{C}_4 = \begin{pmatrix} 0.9438 & 0.5838 \\ 0.5838 & 0.9438 \end{pmatrix}$$

$$\mathbf{C}_5 = \begin{pmatrix} 0.9680 & 0.5706 \\ 0.5706 & 0.9680 \end{pmatrix}$$

and

$$\mathbf{C}_6 = \begin{pmatrix} 0.9471 & 0.5825 \\ 0.5825 & 0.9471 \end{pmatrix}$$

The between block or mating type error structure is $\sigma_b^2 \mathbf{V}$ where

$$\mathbf{V} = \begin{pmatrix} 1 & 0.25 & 0 & 0 & 0 & 0 \\ 0.25 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0.25 & 0 & 0 \\ 0 & 0 & 0.25 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0.25 \\ 0 & 0 & 0 & 0 & 0.25 & 1 \end{pmatrix}$$

The off diagonal elements are 0.25 because those mating types have a common sire. Using equation (2.7) and following, the intramating type estimate of $\boldsymbol{\mu}$ is

$$\hat{\boldsymbol{\mu}} = \begin{pmatrix} -6.979 \\ -4.905 \\ 5.095 \\ 6.789 \end{pmatrix}$$

The value of $Q = \mathbf{y}'(\mathbf{I} - \mathbf{X}\mathbf{X}')^{-1} \mathbf{y}$ is 11.94, thus

$$\hat{\sigma}_e^2 = \frac{Q}{tr(\mathbf{C}) - tr(\mathbf{X}\mathbf{X}'\mathbf{C})} = 11.94/2.267 = 5.277.$$

The variance-covariance matrix of intramating type estimators is $\hat{\sigma}_e^2[\mathbf{X}_2'\mathbf{F}'(\mathbf{F}\mathbf{C}\mathbf{F}')^{-1}\mathbf{F}\mathbf{X}_2]^{-1}$ or

$$\begin{pmatrix} 0.60190 & -0.03592 & -0.30000 & -0.26600 \\ -0.03592 & 0.45210 & -0.22060 & -0.19560 \\ -0.30000 & -0.22060 & -0.55460 & -0.03397 \\ -0.26600 & -0.19560 & -0.03397 & 0.49550 \end{pmatrix}$$

We next look for the intermating type estimator of $\boldsymbol{\mu}$. The intermating type model is $\mathbf{X}_1'\mathbf{y} = \mathbf{X}_1'\mathbf{X}_2\boldsymbol{\mu} + \mathbf{e}_2$ or

$$\begin{pmatrix} 233 \\ 196 \\ 219 \\ 127 \\ 132 \\ 145 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{pmatrix} + \mathbf{e}_2$$

Before the intermating type estimate is computed, σ_b^2 must be estimated. The estimate is $\hat{\sigma}_b^2 = 16.23$, and the estimate of Σ is

$$\hat{\Sigma} = \begin{pmatrix} 178.8 & 36.52 & 0 & 0 & 0 & 0 \\ 36.52 & 179.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 178.8 & 24.35 & 0 & 0 \\ 0 & 0 & 24.35 & 81.04 & 0 & 0 \\ 0 & 0 & 0 & 0 & 81.16 & 16.23 \\ 0 & 0 & 0 & 0 & 16.23 & 81.06 \end{pmatrix}$$

The weighted intermating type estimate is

$$\hat{\boldsymbol{\mu}}_{\hat{w}} = \begin{pmatrix} -27.61 \\ 23.44 \\ -10.29 \\ 14.46 \end{pmatrix}$$

with estimated covariance matrix

$$\begin{pmatrix} 202.70 & -203.00 & -20.39 & 20.64 \\ -203.00 & 232.50 & 1.39 & -30.94 \\ -20.39 & 1.39 & 100.30 & -81.29 \\ 20.64 & -30.94 & -81.29 & 91.59 \end{pmatrix}$$

The combined estimate is

$$\hat{\boldsymbol{\mu}}_c = \begin{pmatrix} -6.914 \\ -4.802 \\ 4.965 \\ 6.750 \end{pmatrix}$$

with estimated covariance matrix,

$$\begin{pmatrix} 0.5894 & -0.04364 & -0.2897 & -0.2561 \\ -0.04364 & 0.4463 & -0.2136 & -0.1891 \\ -0.2897 & -0.2136 & 0.5448 & -0.0415 \\ -0.2561 & -0.1891 & -0.0415 & 0.4867 \end{pmatrix}$$

Note that the estimated variances of the combined estimate are smaller than the estimated variances of the other two estimates. Thus the combined estimate appears to be the best of the three estimates.

Literature

- Bement, T. R.: Estimation in an incomplete block design when block effects are random and correlated. Ph. D. Dissertation, Kansas State University (1972).
- Graybill, F. A.: An Introduction to Linear Statistical Models. Vol. I, New York: McGraw-Hill 1961.
- Graybill, F. A.: Introduction to Matrices with Applications in Statistics. Belmont, California: Wadsworth 1969.
- John, P. W. M.: Statistical Design and Analysis of Experiments. New York: Macmillan 1971.
- Krishna, V. J.: Effects of inbreeding on cattle performance. Ph. D. Dissertation, Kansas State University (1973).
- Wright, S.: Coefficients of inbreeding and relationship. Amer. Nat. 56, 330 (1922).

Received March 14, 1974

Communicated by W. Seyffert and H. Stubbe

Dr. Thomas R. Bement
Northern Arizona University
Flagstat, Arizona (USA)

Dr. George A. Milliken
Dept. of Statistics
Statistical Laboratory
Kansas State University
Calvin Hall 19
Manhattan, Kansas 66506 (USA)

Dr. V. Jayarama Krishna
Senior Scientist
Guntur, Inda

Dr. Robert R. Schalles
Dept. of Animal Science and Industry
Kansas State University
Manhattan, Kansas 66506 (USA)