PREDICTION, MIXED MODELS, AND VARIANCE COMPONENTS

S. R. Searle
Biometrics Unit, Cornell University, Ithaca, N. Y.

BU-468-M June 1973

ABSTRACT

Three methods are described for predicting a random vector that cannot be observed from a realized value of one that can: best prediction, best linear prediction, and Henderson's mixed model prediction. Derivation and properties of the latter are given, and a relationship to Bayes estimation is shown. The need for estimating variance components is emphasized and a summary account given of 8 methods of estimation.

PREDICTION, MIXED MODELS, AND VARIANCE COMPONENTS

S. R. Searle
Biometrics Unit, Cornell University, Ithaca, N. Y.

BU-468-M June 1973

1. INTRODUCTION

There are many situations in biology of having a vector (or scalar value) of observations on some random variables from which we wish to predict the value of some other random variable or variables that cannot be observed. Similar situations also occur outside of biology. A biological example is that of predicting the genetic merit of a dairy bull from the milk yields of his daughters and female records. A non-biological example is that of predicting instrument bias in a micrometer selected randomly out of a manufacturer's lot, using measurements made on a number of objects. And an example in psychology is the one of predicting a person's intelligence from his scores on a battery of tests.

A general statement of the problem is easy. Suppose $\underset{\sim}{U}$ and $\underset{\sim}{Y}$ are jointly distributed vectors of random variables with those in $\underset{\sim}{Y}$ being observable, but those in $\underset{\sim}{U}$ not being observable. The problem is to predict $\underset{\sim}{U}$ from some realized, observed value of $\underset{\sim}{Y}$, say $\underset{\sim}{y}$. Usually $\underset{\sim}{Y}$ contains more elements than $\underset{\sim}{U}$, and indeed $\underset{\sim}{U}$ is often scalar. In the I.Q. example, $\underset{\sim}{U}$ is the scalar, unknowable true value of a person's intelligence, and $\underset{\sim}{y}$ is the vector of his test scores.

2. PREDICTION

Three methods of prediction are of interest: best prediction, best linear prediction, and mixed model prediction. The description which follows draws heavily on the work of C. R. Henderson, who for more than 18 years has sustained

my interest in the prediction problem in the context of animal breeding. Numerous discussions and occasional papers during that time, on mixed models (Henderson et al. [1959]), on variance components (Searle and Henderson [1961], and Henderson et al. [1973]) and on dairy breeding problems themselves (Searle and Henderson [1959, 1960]) have been of invaluable assistance to me, for which I am most grateful. In particular, the opening paragraphs of Henderson [1973a] have been of especial assistance in preparing this account of prediction.

2.1 Best prediction

Suppose for the moment that $\underline{U}$ is scalar. The criterion of the predictor being "best" is taken to be that of minimizing the mean squared error of prediction. When $f(u, \underline{y})$ is the joint density function of the random variables $\underline{U}$ and $\underline{Y}$ at the point $u, \underline{y}$ then with the predictor being denoted by $\tilde{u}$, the mean square error of prediction is

$$E(\tilde{u} - u)^2 = \int \int (\tilde{u} - u)^2 f(u, \underline{y}) d\underline{y} du \qquad (1)$$

where E represents expectation. A generalization of this to a vector of random variables is

$$E(\tilde{\underline{u}} - \underline{u})'\underline{A}(\tilde{\underline{u}} - \underline{u}) = \int \int (\tilde{\underline{u}} - \underline{u})'\underline{A}(\tilde{\underline{u}} - \underline{u}) f(\underline{u}, \underline{y}) d\underline{y} d\underline{u} \qquad (2)$$

where $\underline{A}$ is any positive definite symmetric matrix. Clearly, for $\underline{A}$ being scalar and unity (2) is identical to (1).

The best predictor $\tilde{\underline{u}}$ is that which minimizes (2). As shown in the appendix it turns out to be

$$\text{best predictor} \quad : \quad \tilde{\underline{u}} = E(\underline{u}|\underline{y}) \; ; \qquad (3)$$

i.e., the best predictor of $\underline{u}$ is the conditional mean of $\underline{u}$ given $\underline{y}$. Two

features of this result are worthy of note: first, it holds for all density functions $f(\underline{u}, \underline{y})$, and second, as noted in Solomon [1971], it does not depend on $\underline{A}$ of (2).

Certain properties of this predictor are important and apply to other predictors. They are discussed in Cochran [1951] and in Rao [1965, pp. 79 and 220-222] for the case of scalar $\underline{U}$. First, the predictor is unbiased for sampling over Y : for E_Y representing expectation over $\underline{Y}$

$$E_Y(\tilde{\underline{u}}) = E(\underline{u}) \quad . \tag{4}$$

Second, prediction errors $\tilde{u}$ - u have a covariance matrix that is the mean value, over sampling on $\underline{Y}$, of that of $\underline{u}|\underline{y}$:

$$\text{var}(\tilde{\underline{u}} - \underline{u}) = E_Y[\text{var}(\underline{u}|\underline{y})] \ , \tag{5}$$

Also,

$$\text{cov}(\tilde{\underline{u}}, \underline{u}') = \text{var}(\tilde{\underline{u}}) \tag{6}$$

and

$$\text{cov}(\tilde{\underline{u}}, \underline{y}') = \text{cov}(\underline{u}, \underline{y}') \quad . \tag{7}$$

Derivation of these results is given in the appendix.

For scalar $\underline{u}$ there are 2 further properties of interest. The first is that the correlation between any element u of $\underline{u}$ and any predictor of it that is a function of $\underline{y}$ is maximum for the best predictor, that maximum value being

$$\rho(\tilde{u}, u) = \sigma_{\tilde{u}} / \sigma_{u} \quad . \tag{8}$$

Second, selecting any upper fraction of the population on the basis of values of $\tilde{u}$ insures that for that selected proportion

$$E(u) \text{ is maximized} \quad . \tag{9}$$

Rao [1965] suggests a proof.

It is to be emphasized that $\tilde{\underline{u}} = E(\underline{u}|\underline{y})$ is a random variable, a function of $\underline{y}$. Thus the problem of estimating the predictor remains, and demands some knowledge of the joint density $f(\underline{u}, \underline{y})$. Should this be normal,

$$\begin{bmatrix} \underline{u} \\ \underline{y} \end{bmatrix} \sim N\left\{ \begin{bmatrix} \underline{\mu}_U \\ \underline{\mu}_Y \end{bmatrix}, \begin{bmatrix} \underline{V}_U & \underline{C} \\ \underline{C}' & \underline{V} \end{bmatrix} \right\}, \tag{10}$$

then, as is well known,

$$\tilde{\underline{u}} = E(\underline{u}|\underline{y}) = \underline{\mu}_U + \underline{C}\underline{V}^{-1}(\underline{y} - \underline{\mu}_U) \quad . \tag{11}$$

Properties (5)—(9) of $\tilde{\underline{u}}$ still hold. In (5), we now have from (10) that $\text{var}(\underline{u}|\underline{y}) = \underline{V}_U - \underline{C}\underline{V}^{-1}\underline{C}'$, so that in (5) itself

$$\text{var}(\tilde{\underline{u}} - \underline{u}) = \underline{V}_U - \underline{C}\underline{V}^{-1}\underline{C}' \quad . \tag{12}$$

And using (11) in (6) gives

$$\text{cov}(\tilde{\underline{u}}, \underline{u}') = \text{var}(\tilde{\underline{u}}) = \underline{C}\underline{V}^{-1}\underline{C}' \quad . \tag{13}$$

Then in (8)

$$\rho(\tilde{u}_i, u_i) = \sqrt{\frac{\underline{c}_i'\underline{V}^{-1}\underline{c}_i}{\sigma^2_{u_i}}} \tag{14}$$

where $\underline{c}_i'$ is the i^{th} row of $\underline{C}$.

The estimation problem is clearly visible in these results. The predictor is given in (11) but it and its succeeding properties cannot be estimated until the 4 parameters $\underline{\mu}_U$, $\underline{\mu}_Y$, $\underline{C}$ and $\underline{V}$ have been estimated.

2.2 Best linear prediction

The best predictor (3) is not necessarily linear in $\underline{y}$. Suppose attention is now confined to predictors of $\underline{u}$ that <u>are</u> linear in $\underline{y}$, of the form

$$\tilde{\underline{u}} = \underline{a} + \underline{B}\underline{y} \tag{15}$$

for some vector $\underline{a}$ and matrix $\underline{B}$. Minimizing (2) for $\tilde{\underline{u}}$ of (15), in order to obtain the next linear predictor, leads (see appendix) to

$$\tilde{\underline{u}} = \underline{\mu}_U + \underline{C}\underline{V}^{-1}(\underline{y} - \underline{\mu}_Y) \ , \tag{16}$$

where $\underline{\mu}_U$, $\underline{\mu}_Y$, $\underline{C}$ and $\underline{V}$ are as defined in (10) but without assuming normality as there.

An immediate observation on (16) is that it is identical to (11). This shows that the best linear predictor (16), derivation of which demands no knowledge of the form of $f(u, \underline{y})$, is identical to the best predictor under normality, (11). Properties (12)—(14) therefore apply equally to (16) as they do to (11). Problems of estimation still remain.

2.3 Pairwise ranking

In establishing (9), that selection on the basis of the best predictor $\tilde{u}$ maximizes $E(u)$ of the selected proportion of the population, Cochran's [1951] development implicitly relies on each scalar $\tilde{u}$ having the same variance and being derived from a $\underline{y}$ that is independent of other $\underline{y}$'s . Sampling is over repeated samples of u (scalar) and $\underline{y}$. However, these conditions are not met for the elements of $\tilde{\underline{u}}$ derived in (11). Each such element is derived from the whole vector $\underline{y}$, their variances are not equal, and the elements of $\underline{y}$ used in one element of $\tilde{\underline{u}}$ are not necessarily independent of those used for another element of $\tilde{\underline{u}}$. Maximization of $E(u)$ for individuals selected on the basis of elements in $\tilde{\underline{u}}$ is therefore not assured. In place of this we have a property about pairwise ranking.

In the language of dairy cattle breeding a salient problem is this: having

predicted the genetic merit of several bulls from available records (on daughters and/or female ancestors), and ranked the bulls from highest to lowest according to those predictions, what is the probability that that is the correct ranking? (By correct ranking is meant the ranking according to the bulls' true genetic merits.) When the predicted values $\tilde{u}$ have the properties of the Cochran development, namely equal variances and independent $\underline{y}$'s, this question is answered by the property of maximized E(u) for the selected fraction. But for $\tilde{u}$'s that are elements of $\underline{\tilde{u}}$ the question in this form cannot be answered. What can be said is this: under certain assumptions which shall be specified, using the elements of $\underline{\tilde{u}} = E(\underline{u}|\underline{y})$ for ranking maximizes the probability that pairwise rankings utilizing $\underline{y}$ are correct. A proof of this, based on Henderson [1963] follows.[1/]

Consider predicting two elements of $\underline{u}$, u_1 and u_2, from a vector of observations $\underline{y}$, using predictors $\hat{u}_1$ and $\hat{u}_2$ respectively. Write

$$\hat{d} \equiv \hat{u}_1 - \hat{u}_2 \quad \text{and} \quad d = u_1 - u_2 \quad . \tag{17}$$

Then ranking on $\hat{u}_1$ and $\hat{u}_2$ will be correct if $\hat{d} > 0$ when $d > 0$ and if $\hat{d} < 0$ when $d < 0$. The probabilities we seek to maximize are therefore

$$\Pr(\hat{d} > 0|d > 0) \quad \text{and} \quad \Pr(\hat{d} < 0|d < 0) \quad . \tag{18}$$

Consider the first of these and note that it can be expressed as

$$\Pr(\hat{d} > 0|d > 0) = \Pr(\hat{d} > 0,\ d > 0)/\Pr(d > 0)$$

$$= \int_0^\infty \Pr(\hat{d} > 0|d = k)g(k)dk/\Pr(d > 0)$$

where $g(d)$ is the marginal density of d . From this it is clear that $\Pr(\hat{d} > 0|d > 0)$ can be maximized if we can maximize

---

1/ Discussions with R. R. Davidson are gratefully acknowledged.

$$\Pr(\hat{d} > 0 | d = k) \quad \text{for all } k > 0 \tag{19}$$

using the same rule for all $k > 0$.

Now assume that $\hat{d}$ and d have a bivariate normal distribution

$$\begin{bmatrix} \hat{d} \\ d \end{bmatrix} \sim N\left\{ \begin{bmatrix} \tau_1 \\ \tau_2 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \right\} . \tag{20}$$

Then we know that

$$\hat{d} | d = k \sim N\left[\tau_1 + \frac{\rho\sigma_1}{\sigma_2}(k - \tau_2),\ \sigma_1^2(1 - \rho^2)\right] . \tag{21}$$

If the mean in (21) is positive, maximizing $\Pr(\hat{d} > 0 | d = k)$ is achieved by maximizing

$$\theta = \frac{\tau_1 + \frac{\rho\sigma_1}{\sigma_2}(k - \tau_2)}{\sigma_1\sqrt{1 - \rho^2}} ; \tag{22}$$

but if the mean is negative $-\theta$ has to be minimized. For all positive k it is clearly impossible to simultaneously achieve this maximizing and minimizing. However, if $E(\hat{d}) = \tau_1 = 0$ and $E(d) = \tau_2 = 0$ then the mean in (21) is $\rho\sigma_1 k/\sigma_2$. Since ρ is the correlation between d and its predictor it can be taken as positive so that for positive k the mean $\rho\sigma_1 k/\sigma_2$ is positive. Then θ of (22) becomes

$$\varphi = \frac{\rho k}{\sigma_2\sqrt{1 - \rho^2}} , \tag{23}$$

which has to be maximized. Because σ_2^2 is the variance of d it is constant so far as φ is concerned. Hence, for each positive k, φ is maximized by making $\rho/\sqrt{1 - \rho^2}$ as large as possible, i.e., by maximizing ρ . This is the common rule for all $k > 0$. A converse argument for maximizing the second probability

in (18) by considering $\Pr(\hat{d} < 0 | d = k)$ for negative k leads to the same result. Hence, on assuming $\hat{d}$ and d to have a bivariate normal distribution with zero means, we have shown that the probability of the ranking by $\hat{u}_1$, $\hat{u}_2$ being correct is maximized when $\hat{d}$ is chosen so as to maximize ρ, the correlation of $\hat{d}$ and d . But $\hat{d}$ is the predictor of d, and by sections 2.1 and 2.2, particularly equation (8), we know that $\hat{d} = E(d|\underline{y})$ maximizes $\rho_{\hat{d}d}$. Under the normality assumption, and with the zero means already referred to, (11) and equivalently (16) then give

$$\hat{d} = \underline{c}'\underline{V}^{-1}(\underline{y} - \underline{\mu}_Y) \tag{24}$$

where $\underline{c}' = \text{cov}(d,\underline{y}')$. But by (17) $d = u_1 - u_2$ and so $\underline{c}' = \underline{c}_1' - \underline{c}_2'$ where $\underline{c}_1'$ and $\underline{c}_2'$ are the rows of $\underline{C}$ corresponding to u_1 and u_2 in $\underline{C} = \text{cov}(\underline{u},\underline{y}')$ of (10). Hence (24) is

$$\hat{u}_1 - \hat{u}_2 = (\underline{c}_1' - \underline{c}_2')\underline{V}^{-1}(\underline{y} - \underline{\mu}_Y)$$

equivalent to

$$\tilde{u}_i = \lambda + \underline{c}_i'\underline{V}^{-1}(\underline{y} - \underline{\mu}_Y) \quad \text{for } i = 1,2 \ . \tag{25}$$

This is identical to elements of (11), the best predictor under normality, with the only proviso that $\underline{\mu}_U = \lambda\underline{1}$ for some constant λ, i.e., all elements of $\underline{u}$ must have the same mean, λ say. With this not very restrictive condition we therefore see that under normality the best predictors (which are then also best linear predictors) maximize the probability of correct pairwise rankings.

2.4 Mixed model prediction

The preceding discussion is concerned with the prediction of random variables. Through maximizing the probability of correct pairwise rankings the predictors are appropriate values upon which to base selection; e.g., in genetics, selecting the animals with highest predictions to be parents of the next

generation. (One apparently unanswered problem, though, is to find conditions under which maximizing the probability of correct pairwise rankings for all pairs also maximizes the probability of a correct overall ranking.) Since we are concerned here with the prediction (and selection) of random variables, the procedure might be called Model II prediction corresponding to Model II, the random effects model, in analysis of variance. In this connection Lehman [1961] has discussed Model I prediction, corresponding to the fixed effects model. Consideration is now given to mixed model prediction, corresponding to mixed models in analysis of variance in which some factors are of fixed effects and others are of random effects.

The model we initially use for $\underline{y}$ is the familiar

$$E(\underline{y}) = \underline{X}\underline{\beta} \tag{26}$$

for $\underline{\beta}$ being some vector of unknown constants (fixed effects); and to retain the ranking property we take

$$E(\underline{u}) = \underline{0} \quad . \tag{27}$$

Then we consider the problem of predicting

$$\underline{w} = \underline{K}'\underline{\beta} + \underline{u} \tag{28}$$

for some known matrix $\underline{K}'$. Since $\underline{w}$ involves both fixed effects and random variables there might be debate as to whether we should 'estimate' $\underline{w}$ or 'predict' $\underline{w}$. We will 'predict' $\underline{w}$, and will choose $\tilde{\underline{w}}$ as a predictor to have 3 properties:

$$\begin{aligned} &\text{"best" in the sense of (2): minimizing } E(\tilde{\underline{w}} - \underline{w})'A(\tilde{\underline{w}} - \underline{w}) \\ &\text{linear in y: } \tilde{\underline{w}} = \underline{a} + \underline{B}\underline{y} \\ &\text{unbiased: } E(\tilde{\underline{w}}) = E(\underline{w}) \end{aligned} \tag{29}$$

The resulting predictor is a best linear unbiased prediction. Note that unbias-

edness is now a criterion of the prediction procedure and not just a byproduct of it as in section 2. Introducing it as a criterion arises from the presence of $\underline{\beta}$.

It is clear from (27) and (28) that $E(\underline{w}) = \underline{K}'\underline{\beta}$. We then have

$$\begin{bmatrix} \underline{w} \\ \underline{y} \end{bmatrix} \sim \text{with mean} \begin{bmatrix} \underline{K}'\underline{\beta} \\ \underline{X}\underline{\beta} \end{bmatrix} \text{ and covariance matrix } \begin{bmatrix} \underline{V}_U & \underline{C} \\ \underline{C}' & \underline{V} \end{bmatrix} \tag{30}$$

similar to (10), although without yet assuming normality. The unbiasedness required of $\tilde{\underline{w}}$ in (30) demands that $\underline{a} + \underline{B}\underline{X}\underline{y} = \underline{K}'\underline{\beta}$ for all $\underline{\beta}$ and so $\underline{a} = \underline{0}$ and $\underline{B}\underline{X} = \underline{K}'$. Consequently the predictor is $\tilde{\underline{w}} = \underline{B}\underline{y}$, and in $\underline{w} = \underline{K}'\underline{\beta} + \underline{u}$ the term $\underline{K}'\underline{\beta}$ is an estimable function of $\underline{\beta}$ in the model $E(\underline{y}) = \underline{X}\underline{\beta}$. This limits the form of $\underline{K}'$ in $\underline{w}$, but it is obviously a reasonable limitation.

Details of the derivation of $\underline{B}$ for $\tilde{\underline{w}} = \underline{B}\underline{y}$ satisfying all 3 criteria of (29) are shown in the appendix. The result is that

$$\underline{B} = \underline{C}\underline{V}^{-1} + (\underline{K}' - \underline{C}\underline{V}^{-1}\underline{X})(\underline{X}'\underline{V}^{-1}\underline{X})^{-}\underline{X}'\underline{V}^{-1} \tag{31}$$

so that

$$\tilde{\underline{w}} = \underline{K}'(\underline{X}'\underline{V}^{-1}\underline{X})^{-}\underline{X}'\underline{V}^{-1}\underline{y} + \underline{C}\underline{V}^{-1}[\underline{y} - \underline{X}(\underline{X}'\underline{V}^{-1}\underline{X})^{-}\underline{X}'\underline{V}^{-1}\underline{y}] \tag{32}$$

where $(\underline{X}'\underline{V}^{-1}\underline{X})^{-}$ is a generalized inverse of $\underline{X}'\underline{V}^{-1}\underline{X}$ satisfying $\underline{X}'\underline{V}^{-1}\underline{X}(\underline{X}'\underline{V}^{-1}\underline{X})^{-}\underline{X}'\underline{V}^{-1}\underline{X}$ $= \underline{X}'\underline{V}^{-1}\underline{X}$. Note in (32) the occurrence of $(\underline{X}'\underline{V}^{-1}\underline{X})^{-}\underline{X}'\underline{V}^{-1}\underline{y}$, and observe that for the linear model $E(\underline{y}) = \underline{X}\underline{\beta}$ with $\text{var}(\underline{y}) = \underline{V}$, as in (26) and (30), this is a solution to the generalized least squares equations

$$\underline{X}'\underline{V}^{-1}\underline{X}\underline{\beta}^{\circ} = \underline{X}'\underline{V}^{-1}\underline{y}$$

with

$$\underline{\beta}^{\circ} = (\underline{X}'\underline{V}^{-1}\underline{X})^{-}\underline{X}'\underline{V}^{-1}\underline{y} \quad .$$

Hence the predictor is

$$\tilde{\underline{w}} = \underline{K}'\underline{\beta}^{\circ} + \underline{C}\underline{V}^{-1}(\underline{y} - \underline{X}\underline{\beta}^{\circ}) \tag{33}$$

The form of this predictor is of interest. It is the sum of two parts: (i) $\underline{K}'\underline{\beta}^{\circ}$ the best linear unbiased estimator of the estimable function $\underline{K}'\underline{\beta}$ in the model $E(\underline{y}) = \underline{X}\underline{\beta}$, for $\text{var}(\underline{y}) = \underline{V}$ known, and (ii) $\tilde{\underline{u}} = \underline{C}\underline{V}^{-1}(\underline{y} - \underline{X}\underline{\beta}^{\circ})$ of (16), the best linear predictor of $\underline{u}$, with $\underline{\mu}_U = 0$ and with $\underline{\mu}_Y = \underline{X}\underline{\beta}$ replaced by its best linear unbiased estimator $\underline{X}\underline{\beta}^{\circ}$. To emphasize this we rewrite (33) as

$$\tilde{\underline{w}} = \underline{K}'\underline{\beta}^{\circ} + \tilde{\underline{u}}^{\circ} \qquad (34)$$
$$\text{for } \tilde{\underline{u}}^{\circ} = \underline{C}\underline{V}^{-1}(\underline{y} - \underline{X}\underline{\beta}^{\circ})$$

$\tilde{\underline{w}}$ is thus the sum of what one might call the Model I predictor of $\underline{K}'\underline{\beta}$ and the Model II predictor of $\underline{u}$, using $\underline{\beta}^{\circ}$. Result (30) is given in Henderson [1973a] and that part of it not involving $\underline{K}'\underline{\beta}$ is also in Henderson [1963] in a slightly different context.

A variety of variances and covariances can be derived:

$$\text{var}(\underline{K}'\underline{\beta}^{\circ}) = \underline{K}'(\underline{X}'\underline{V}^{-1}\underline{X})^{-}\underline{K} \qquad (35)$$

$$\text{var}(\underline{u}^{\circ}) = \underline{C}\underline{V}^{-1}\underline{C}' - \underline{C}\underline{V}^{-1}\underline{X}(\underline{X}'\underline{V}^{-1}\underline{X})^{-}\underline{X}'\underline{V}^{-1}\underline{C}' \qquad (36)$$

$$\text{cov}(\underline{K}'\underline{\beta}^{\circ}, \tilde{\underline{u}}^{\circ\prime}) = \underline{0} \qquad (37)$$

$$\text{var}(\tilde{\underline{w}}) = \text{var}(\underline{K}'\underline{\beta}^{\circ}) + \text{var}(\underline{u}^{\circ}) \qquad (38)$$

$$\text{cov}(\tilde{\underline{u}}^{\circ}, \underline{u}') = \text{var}(\underline{u}^{\circ}) \qquad (39)$$

$$\text{var}(\underline{u}^{\circ} - \underline{u}) = \underline{V}_U - \text{var}(\tilde{\underline{u}}^{\circ}) \qquad (40)$$

$$\text{cov}(\underline{K}\underline{\beta}^{\circ}, \underline{u}') = \underline{B}\underline{X}(\underline{X}'\underline{V}^{-1}\underline{X})^{-}\underline{X}'\underline{V}^{-1}\underline{C} \qquad (41)$$

$$\text{var}(\tilde{\underline{w}} - \underline{w}) = \text{var}(\underline{K}'\underline{\beta}^{\circ}) + \text{var}(\underline{u}^{\circ} - \underline{u}) - \text{cov}(\underline{K}'\underline{\beta}^{\circ}, \underline{u}) - \text{cov}(\underline{u}', \underline{\beta}^{\circ\prime}\underline{K}) \qquad (42)$$

All of the preceding results involve no assumption of normality. On introducing that assumption, as in (10) with $\underline{\mu}_U = \underline{0}$ and $\underline{\mu}_Y = \underline{X}\underline{\beta}$, we have

$$\begin{bmatrix} \underline{w} = \underline{K}'\underline{\beta} + \underline{u} \\ \underline{y} \end{bmatrix} \sim N\left\{ \begin{bmatrix} \underline{K}'\underline{\beta} \\ \underline{X}\underline{\beta} \end{bmatrix}, \begin{bmatrix} \underline{V}_U & \underline{C} \\ \underline{C}' & \underline{V} \end{bmatrix} \right\} \tag{43}$$

Then $\underline{X}\underline{\beta}^{\circ}$ is the maximum likelihood (as well as the best linear unbiased) estimator of $\underline{X}\underline{\beta}$, for $\underline{V}$ assumed known, and since from (43)

$$E(\underline{w}|\underline{y}) = \underline{K}'\underline{\beta} + \underline{C}\underline{V}^{-1}(\underline{y} - \underline{X}\underline{\beta}) ,$$

it follows that for $\underline{V}$ known, $\tilde{\underline{w}}$ of (34) is the maximum likelihood estimator of $E(\underline{w}|\underline{y})$. Furthermore, with $\tilde{\underline{u}}^{\circ} = \underline{C}\underline{V}^{-1}(\underline{y} - \underline{X}\underline{\beta}^{\circ})$, $\underline{u}$ and $\underline{u}^{\circ}$ are normally distributed with zero means and because of (39)

$$E(\underline{u}|\tilde{\underline{u}}^{\circ}) = \text{cov}(\tilde{\underline{u}}^{\circ},\underline{u}')[\text{var}(\underline{u}^{\circ})]^{-1}\tilde{\underline{u}}^{\circ} = \tilde{\underline{u}}^{\circ}$$

and

$$\begin{aligned} \text{var}(\underline{u}|\tilde{\underline{u}}^{\circ}) &= \underline{V}_U - \text{cov}(\tilde{\underline{u}}^{\circ},\underline{u}')[\text{var}(\underline{u}^{\circ})]^{-1}\text{cov}(\tilde{\underline{u}}^{\circ},\underline{u}') \\ &= \underline{V}_U - \text{var}(\underline{u}^{\circ}) \\ &= \text{var}(\underline{u}^{\circ} - \underline{u}) \quad \text{as in (40).} \end{aligned}$$

And, of course, as has already been shown, the elements of $\tilde{\underline{u}}^{\circ}$ have the property of maximizing the probability of correct pairwise rankings. But this property does not hold for elements of $\tilde{\underline{w}}$, unless $E(\underline{w}) = \underline{K}'\underline{\beta}$ is of the form $\lambda\underline{i}$.

3. THE MIXED MODEL

Consonant with mixed model prediction just discussed we now consider the mixed model of analysis of variance, namely a linear model involving both fixed effects and random effects. It can be typified as

$$\underline{y} = \underline{X}\underline{\beta} + \underline{X}\underline{u} + \underline{e} \tag{44}$$

where $\underline{y}$ is a vector of observations, $\underline{\beta}$ is a vector whose elements are the effects of one or more fixed effects factors (including a general mean), and $\underline{u}$ is a vector of the effects of the random effects factors. $\underline{X}$ and $\underline{Z}$ are known matrices, often design or incidence matrices with elements 0 and 1, and $\underline{e}$ is a vector of random error terms. Although $\underline{X}$ is usually a design matrix it can include regressor variables, in which case the corresponding elements of $\underline{\beta}$ are regression coefficients. $\underline{X}$ and $\underline{Z}$ are generally of less than full column rank.

Distributional properties of $\underline{u}$ and $\underline{e}$ are assumed to be as follows:

$$\begin{gathered} E(\underline{u}) = \underline{0} \quad \text{and} \quad E(\underline{e}) = \underline{0} \\ \text{var}(\underline{u}) = \underline{D}, \quad \text{var}(\underline{e}) = \underline{R} \quad \text{and} \quad \text{cov}(\underline{u}, \underline{e}') = \underline{0} \, . \end{gathered} \tag{45}$$

In this we are rewriting $\underline{V}_U$ of the earlier discussion as $\underline{D}$,

$$\text{var}(\underline{u}) = \underline{V}_U \equiv D, \tag{46}$$

and from (44) we then have

$$\text{var}(\underline{y}) = \underline{V} = \underline{Z}\underline{D}\underline{Z}' + \underline{R} \, . \tag{47}$$

Also, from (44)

$$\text{cov}(\underline{u}, \underline{y}') = \underline{C} \equiv \underline{D}\underline{Z}' \, . \tag{48}$$

Hence the predictor $\tilde{\underline{w}}$ of (34) is

$$\tilde{\underline{w}} = \underline{K}'\underline{\beta}^{o} + \tilde{\underline{u}}^{o} \tag{49}$$

where for

$$\underline{\beta}^{o} = (\underline{X}'\underline{V}^{-1}\underline{X})^{-}\underline{X}'\underline{V}^{-1}\underline{y} \tag{50}$$

we have

$$\underline{u}^{o} = \underline{DZ}'\underline{V}^{-1}(\underline{y} - \underline{X}\underline{\beta}^{o}) \quad . \tag{51}$$

Similar small changes involving $\underline{C} = \underline{DZ}$ occur in the variances and covariances of (35) - (42).

3.1 Calculating the predictor

On the assumption that $\underline{D}$ and $\underline{R}$ of (45) are known, calculation of $\underline{\beta}^{o}$ and $\underline{u}^{o}$ of (50) and (51) involves $\underline{V}^{-1} = (\underline{ZDZ}' + \underline{R})^{-1}$. Since $\underline{V}$ has order equal to the number of observations, which in many applications is very large because the model includes random effects, calculation of $\underline{V}^{-1}$ can be a mean task even for to-day's computers; for example, $\underline{V}$ of order 10,000! However, this inversion can be avoided.

Assume temporarily that $\underline{u}$ in the mixed model (44) represents fixed and not random effects. Then var($\underline{y}$) would be $\underline{R}$ and not $\underline{V}$ and the generalised least squares equations would be

$$\begin{bmatrix} \underline{X}'\underline{R}^{-1}\underline{X} & \underline{X}'\underline{R}^{-1}\underline{Z} \\ \underline{Z}'\underline{R}^{-1}\underline{X} & \underline{Z}'\underline{R}^{-1}\underline{Z} \end{bmatrix} \begin{bmatrix} \hat{\underline{\beta}} \\ \hat{\underline{u}} \end{bmatrix} = \begin{bmatrix} \underline{X}'\underline{R}^{-1}\underline{y} \\ \underline{Z}'\underline{R}^{-1}\underline{y} \end{bmatrix} \quad . \tag{52}$$

Suppose these equations are amended by adding $\underline{D}^{-1}$ to the lower right-hand sub-matrix $\underline{Z}'\underline{R}^{-1}\underline{Z}$ on the left-hand side, to give

$$\begin{bmatrix} \underline{X}'\underline{R}^{-1}\underline{X} & \underline{X}'\underline{R}^{-1}\underline{Z} \\ \underline{Z}'\underline{R}^{-1}\underline{Z} & \underline{Z}'\underline{R}^{-1}\underline{Z} + \underline{D}^{-1} \end{bmatrix} \begin{bmatrix} \underline{\beta}^{o} \\ \underline{u}^{o} \end{bmatrix} = \begin{bmatrix} \underline{X}'\underline{R}^{-1}\underline{y} \\ \underline{Z}'\underline{R}^{-1}\underline{y} \end{bmatrix} \quad . \tag{53}$$

Then $\underline{\beta}^{o}$ and $\underline{u}^{o}$ of these equations are exactly $\underline{\beta}^{o}$ and $\underline{u}^{o}$ of (50) and (51). Proof of this for $\underline{X}$ of full column rank was first given in Henderson et al [1959] and

is also available in Searle [1971a, p. 460]. Although the proof for $\underline{X}$ of less than full rank involves only minor changes it is given in the appendix for the sake of completeness. The proof involves verifying that

$$\underline{V}^{-1} = \underline{W} = \underline{R}^{-1} - \underline{R}^{-1}\underline{Z}(\underline{Z}'\underline{R}^{-1}\underline{Z} + \underline{D}^{-1})^{-1}\underline{Z}'\underline{R}^{-1}, \tag{54}$$

the identity which Kempthorne [1972] refers to in discussing Lindley and Smith [1972].

It is interesting to note that although there are many solutions to (53), all of them have the same $\underline{u}^{\circ}$. This is not obvious from the form of (53) but it is, of course, evident from (51) because $\underline{X}\underline{\beta}^{\circ}$, being the best linear unbiased estimator of $\underline{X}\underline{\beta}$, is invariant to $\underline{\beta}^{\circ}$.

The advantage of (53) is a computational one, that the matrix on the left has order equal to the total number of fixed and random effects in the model, which is usually very much less than the total number of observations, which is the order of $\underline{V}$ whose inverse is needed in (50) and (51).

In most applications $\underline{R} = \text{var}(\underline{e}) = \sigma^2\underline{I}$ so that (54) becomes

$$\underline{V}^{-1} = \underline{W} = [\underline{I} - \underline{Z}(\underline{Z}'\underline{Z} + \sigma^2\underline{D}^{-1})^{-1}\underline{Z}']/\sigma^2 . \tag{55}$$

This involves inverses of order equal to just the number of random effects in the model, and since $\underline{D} = \text{var}(\underline{u})$ is often diagonal of the form $\begin{bmatrix} \sigma^2_{\alpha}\underline{I}_a & 0 \\ 0 & \sigma^2_{\beta}\underline{I}_b \end{bmatrix}$, for example, $\underline{D}^{-1}$ is easily calculated and the only computed inverse required is $(\underline{Z}'\underline{Z} + \sigma^2\underline{D}^{-1})^{-1}$. Through (55), (50) and (51) the values $\underline{\beta}^{\circ}$ and $\underline{u}^{\circ}$ are then available.

3.2 Examples

2-way mixed model

Consider a 2-way classification, with a rows and b columns, and one observation per cell, for which the model is

$$y_{ij} = \mu + \alpha_i + \beta_j + e_{ij}, \quad \text{for } i = 1,2,\ldots,a \text{ and } j = 1,2,\ldots,b \ . \tag{56}$$

If the α_i's are taken as random variables, independently distributed with zero means and variance σ^2_α, and the error terms similarly distributed with $E(e_{ij}) = 0$ and $v(e_{ij}) = \sigma^2_e$, then in the mixed model notation of (44) and (45)

$$\underset{\sim}{X} = \left\{ \underset{\sim}{1}_{ab} \begin{bmatrix} \underset{\sim}{I}_b \\ \underset{\sim}{I}_b \\ \vdots \\ \underset{\sim}{I}_b \end{bmatrix} \right\}, \quad \underset{\sim}{Z} = \overset{a}{\underset{1}{\Sigma^+}} \underset{\sim}{1}_b, \quad \underset{\sim}{R} = \sigma^2 \underset{\sim}{I}_{ab}, \text{ and } \underset{\sim}{D} = \sigma^2_\alpha \underset{\sim}{I}_a \ ,$$

where $\underset{\sim}{1}_m$ is a vector of m unities, and Σ^+ represents the operation of a Kronecker (direct) sum of matrices. Making these substitutions in (55), (50) and (51) leads after a little simplification to

$$\underset{\sim}{\beta}^\circ = \begin{bmatrix} \mu^\circ \\ \{\beta_j\} \end{bmatrix} = \begin{bmatrix} 0 \\ \{\bar{y}_{.j}\} \end{bmatrix}$$

for $j = 1,2,\ldots,b$

and

$$\underset{\sim}{\tilde{u}} = \{\tilde{\alpha}^\circ_i\} = \frac{b\sigma^2_\alpha}{\sigma^2_e + b\sigma^2_\alpha} (\bar{y}_{i.} - \bar{y}_{..}) \ . \tag{57}$$

2-way, random model

When the β_j's are also taken as random variables, independently distributed with zero means and variances σ^2_β, then μ is the only fixed effect in the model and we have for (44) and (45)

$$\underset{\sim}{X} = \underset{\sim}{1}_{ab}, \quad \underset{\sim}{Z} = \left\{ \begin{matrix} a \\ \Sigma \\ 1 \end{matrix} \underset{\sim}{1}_b \quad \left[\begin{matrix} \underset{\sim}{I}_b \\ \vdots \\ \underset{\sim}{I}_b \end{matrix} \right] \right\}, \quad \underset{\sim}{R} = \sigma^2_e \underset{\sim}{I}_{ab} \text{ and } \underset{\sim}{D} = \begin{bmatrix} \sigma^2_\alpha \underset{\sim}{I}_a & 0 \\ 0 & \sigma^2_\beta \underset{\sim}{I}_b \end{bmatrix}.$$

Substitution in and simplification of (50) and (51) then give

$$\mu^o = \bar{y}_{..}, \quad \alpha_i^o = \frac{b\sigma^2_\alpha}{\sigma^2_e + b\sigma^2_\alpha} (\bar{y}_{i.} - \bar{y}_{..}) \text{ and } \beta_j^o = \frac{a\sigma^2_\beta}{\sigma^2_e + a\sigma^2_\beta} (\bar{y}_{.j} - \bar{y}_{..}) \quad (58)$$

Although this model is not of as much practical interest as a mixed model as is that in which the β_j's are fixed, it is of theoretical interest because the results (58) are identical to the Bayes estimates of Lindley and Smith [1972]. One difference is that no assumption has been made here about the form of distribution of the α's and β's whereas Lindley and Smith's results demand normality.

1-way random model

Results of the form (57) and (58) have been familiar to animal breeders for many years. A simple use of them is in the case of the 1-way classification with unequal numbers of observations in the subclasses:

$$\begin{aligned} y_{ij} &= \mu + \alpha_i + e_{ij} \qquad i = 1,2,\ldots,a \\ &\qquad\qquad\qquad\quad j = 1,2,\ldots,n_i . \end{aligned} \quad (59)$$

Treating the α_i's as random the terms of (44) and (45) are

$$\underset{\sim}{X} = \underset{\sim}{1}'_n, \quad \underset{\sim}{Z} = \sum_{i=1}^{a} \underset{\sim}{1}_{n_i}, \quad \underset{\sim}{D} = \sigma^2_\alpha \underset{\sim}{I}_a \text{ and } \underset{\sim}{R} = \sigma^2_e \underset{\sim}{I}_n .$$

After substitution in (50) and (51) simplification gives

$$\mu^o = \sum_{i=1}^{a} \frac{n_i \bar{y}_{i.}}{\sigma^2_e + n_i\sigma^2_\alpha} \Big/ \sum_{i=1}^{a} \frac{n_i}{\sigma^2_e + n_i\sigma^2_\alpha} \quad (60)$$

and

$$\tilde{\alpha}_i^o = \frac{n_i\sigma_\alpha^2}{\sigma_e^2 + n_i\sigma_\alpha^2}(\bar{y}_{i.} - \mu^o) . \qquad (61)$$

It is noticeable that μ^o is the generalized least squares estimator of μ and, under normality, the maximum likelihood estimator. (61), which is akin to similar expressions in (57) and (58), is more recognizable to animal breeders in forms such as

$$\tilde{\alpha}_i^o = \frac{n_i r}{1 + (n_i - 1)r}(\bar{y}_{i.} - \mu^o)$$

where in certain contexts the ratio $r = \sigma_\alpha^2/(\sigma_\alpha^2 + \sigma_e^2)$ is the animal breeding parameter repeatability; or as

$$\alpha_i^o = \frac{n_i h}{4 + (n_i - 1)h}(\bar{y}_{i.} - \mu^o)$$

where in other contexts $h = 4\sigma_\alpha^2/(\sigma_\alpha^2 + \sigma_e^2)$ is the parameter heritability. Practical uses of these kind of formulae, used as "estimated producing ability" and as "estimated breeding value" respectively, or of precursors of them, are to be found in dairy science literature in such early references as Lush [1931, 1933, 1948] and Wright [1931]. In this setting Lush was an enthusiastic promoter of their use in developing breeding plans for the improvement of agricultural livestock.

In practice, deriving (57) and (60) through the use of (55) is unnecessarily tedious because in both these cases $\underline{V}$ has the form $\sum_{i=1}^{a}{}^{+}(p_i\underline{I}_{n_i} + q_i\underline{J}_{n_i})$, a form of whose inverse is well known. ($\underline{I}_{n_i}$ is an identity matrix of order n_i, and $\underline{J}_{n_i}$ is a square matrix of order n_i with every element unity.)

4. ESTIMATING VARIANCE COMPONENTS

The different predictors in Section 2 demand knowledge of different things. The best predictor $\tilde{\underline{u}} = E(\underline{u}|\underline{y})$ requires $f(\underline{u}|\underline{y})$, whereas the best linear predictor

(best under normality) $\tilde{\underline{u}} = \underline{\mu}_U + \underline{C}\underline{V}^{-1}(\underline{y} - \underline{\mu}_Y)$ requires just first and second moments. The mixed model predictor $\tilde{\underline{w}} = \underline{K}'\underline{\beta}^o + \underline{C}\underline{V}^{-1}(\underline{y} - \underline{X}\underline{\beta}^o)$ for $\underline{\beta}^o = (\underline{X}'\underline{V}^{-1}\underline{X})^-\underline{X}'\underline{V}^{-1}\underline{y}$ demands knowing second moments, $\underline{V} = \text{var}(\underline{y})$ and $\underline{C} = \text{cov}(\underline{u},\underline{y}')$; and when the mixed model $\underline{y} = \underline{X}\underline{\beta} + \underline{Z}\underline{u} + \underline{e}$ is used this requires knowledge of $\text{var}(\underline{u}) = \underline{D}$ and $\text{var}(\underline{e}) = \underline{R}$ with $\underline{V} = \underline{Z}\underline{D}\underline{Z}' + \underline{R}$ and $\underline{C} = \underline{D}\underline{Z}'$. It is clear, therefore, in the face of not knowing true values of the needed second moments that we need to estimate them, although ideally it might be preferable to estimate predictors directly in some optimum manner. Nevertheless, the usual practice is to estimate the components of variance that make up the elements of $\underline{V}$, and in $\underline{V}$ replace those components by their estimates to derive an estimate $\hat{\underline{V}}$. In the predictor, $\hat{\underline{V}}$ is then used in place of $\underline{V}$. For example, in the 1-way classification model of (59)

$$\underline{V} = \sum_{i=1}^{a} + \left(\sigma^2_e \underline{I}_{n_i} + \sigma^2_\alpha \underline{J}_{n_i}\right)$$

and after estimates $\hat{\sigma}^2_e$ and $\hat{\sigma}^2_\alpha$ have been obtained

$$\hat{\underline{V}} = \sum_{i=1}^{a} + \left(\hat{\sigma}^2_e \underline{I}_{n_i} + \hat{\sigma}^2_\alpha \underline{J}_{n_i}\right)$$

replaces $\underline{V}$ in the predictor.

It is customary to estimate the variance components needed for $\hat{\underline{V}}$ from data different from those from which predictors are to be derived. In some applications repeated estimates have been gathered so often over the years that subject-matter research workers are prepared to give *a priori* values to $\underline{C}$ and $\underline{V}$ and use them as if they were population values. The variances and covariances of (35)—(42) then apply, whereas they would be considerably more complicated were the sampling nature of $\hat{\underline{C}}$ and $\hat{\underline{V}}$ taken more correctly into account.

Estimating variance components is therefore a problem very pertinent to prediction. Since lengthy reviews of this subject have recently appeared, e.g., Searle

[1971a, b], only a thumbnail outline is given here in the form of comments on several estimation methods available, together with some updating. For detailed accounts of most of the methods and extensive literature references thereto the reader is referred to the sources just cited.

4.0 Balanced and unbalanced data

Balanced data means data in which there are the same number of observations in every sub-most cell of the data. Unbalanced data are those having unequal numbers of observations in such cells, including the possibility of none at all in some cells (i.e., empty cells). Survey data, biological and otherwise, are often unbalanced with many empty cells; e.g., in certain dairy breeding data as many as 90% of the cells of a 2-way crossed classification may be empty.

There is one universally accepted method of estimating variance components from balanced data. It involves calculating an analysis of variance as if the model were a fixed effects model. Each mean square is then equated to its expected value under the mixed model appropriate to the data. Since the resulting expectations are linear combinations of the unknown variance components, the equations so formed can be solved for these components. The solutions are the estimated components. Generally speaking they are known as analysis of variance estimators. They have several desirable properties:

(i) They are unbiased.

(ii) Under normality, sampling variances are available.

(iii) They are easy to compute.

(iv) They have minimum variance among all quadratic unbiased estimators.

(v) Under normality they have minimum variance among all unbiased estimators.

Apart from the estimated error variance (which, under normality, is distributed as

a multiple of a chi-square distribution) there is no closed form for the sampling distribution of these estimators. In some cases, the distributions can be expressed as an infinite sum of weighted chi-square variables, although the weights involve the unknown components.

Estimating variance components from unbalanced data is considerably more complex than from balanced data, because there is no universal method of estimation. Consider, for example, trying to adapt the method just described for balanced data. Immediately there arises the question of "what analysis of variance?". For example in fitting a rows-and-columns model, which analysis of variance is to be used: that for fitting rows before columns, or the one for fitting columns before rows? Even apart from this problem the resulting estimators have only the first two of the properties (i)—(v) listed above, and of those not always the second. Nevertheless, there has been widespread adoption of the underlying technique of that method, namely of equating calculated mean squares to their expected values. The difficulty is in the choice of what are to be used as mean squares, or more generally as quadratic forms. It is the wide choice of quadratic forms available that has given rise to there being a number of methods of estimation.

We may note in passing that for any method of estimating variance components by quadratic forms, the corresponding bilinear form estimator of a covariance component can be obtained by applying the familiar formula expressing a covariance in terms of variances, namely $\sigma_{xy} = \frac{1}{2}(\sigma^2_{x+y} - \sigma^2_x - \sigma^2_y)$. This is discussed in Searle and Rounsaville [1973].

4.1 Henderson's Method 1. (Analysis of variance)

This uses sums of squares that are unbalanced-data analogies of those used

in the analysis of variance of balanced data. In some cases these analogies turn out not to be positive semi-definite, since they are not actually sums of squares. Nevertheless, being quadratic forms, they constitute a legitimate basis for estimating variance components. The resulting estimators are relatively easy to calculate and are unbiased, and for many random effects models their sampling variances, under normaltiy, are known.

Method 1 cannot be used for mixed models. The only way its use for mixed models can be forced is either by ignoring the fixed effects or by assuming they are random. Either way, the resulting estimators for the variances of the random effects of the mixed model are biased.

4.2 Henderson's Method 2

The inappropriateness of Method 1 for mixed models is the motive for Method 2. Capitalizing on the easy computations of Method 1, the procedure of Method 2 is to correct the data (of a mixed model) according to some estimates of the fixed effects and then use Method 1 on the data so corrected. Some minor adjustments to Method 1 are needed. The method cannot be used when the model contains interactions between fixed and random effects.

First proposed in Henderson [1953], this method has undoubtedly been wrongly used in succeeding years because of its subtleties. In re-describing it in matrix terminology (and proving the impossibility of using it when there are interactions between fixed and random effects), Searle [1968] strongly asserted that the Method was not uniquely specifiable, an assertion repeated in Searle [1971a, b]. It is a pleasure to report that this assertion is false: Method 2 is well-defined. Proof of this is given in Henderson et al. [1973]. The limitation on not being able to use the method in the presence of interactions between fixed and random

effects does, however, still stand. Computationally it is in most cases a viable method provided the number of fixed effects in the model is not large.

4.3 Henderson's Method 3 (Fitting constants)

This method uses the sums of squares due to fitting the model as if it were a fixed effects model. Expectations are taken over the true model for the data. The method is particularly suited to mixed models because it yields variance components estimators unaffected by the fixed effects. Furthermore, it is not subject to the limitation of Method 2 concerning interactions between fixed and random effects. However, with large data sets it may be impractical or exceedingly expensive to compute the needed sums of squares, due to the necessity of inverting matrices that may be of very large order. Also, the method can yield more equations to solve than there are components to estimate, which presents a problem. Nevertheless, the resulting estimators are unbiased and in some cases their sampling variances and covariances, under normality, are known. A recent application of this method to models that include covariates is made by Mount and Searle [1972]. It is interesting that their results reduce to calculating sums of sums of squares of residuals.

4.4 Analysis of means

When all cells of the model contain at least one datum, sums of squares arising in analyses of variance of cell means can be used for estimating variance components. The calculations are easy, the estimators are unbiased and certain other properties are available.

4.5 Symmetric sums

This is a method which has not received much attention since its development

by Koch [1968]. The quadratic forms it uses are based on symmetric sums of squares of differences between observations rather than analysis-of-variance style quadratics. For example, in the 1-way classification model of (59) it uses

$$E \sum_{i=1}^{a} \sum_{i' \neq i}^{a} \sum_{j=1}^{n_i} \sum_{j'=1}^{n_{i'}} \left(y_{ij} - y_{i'j'}\right)^2 = \sum_{i=1}^{} \sum_{i' \neq i}^{} n_i n_{i'} \left(\sigma_\alpha^2 + \sigma_e^2\right) .$$

4.6 Synthesis

The method of synthesis provided by Hartley [1967] gives no new methodology for estimating variance components. It is a computational procedure that avoids much of the algebra involved in the previous methods. It requires using, in turn, each column of the design matrix of the model as the $\underline{y}$ vector in each quadratic form $\underline{y}'\underline{Q}\underline{y}$ on which any particular method is based. Even though these columns contain many zeros, there may be a large number of columns involved, and the method can thus be very consuming of computer time.

4.7 Maximum likelihood

An iterative procedure for solving the maximum likelihood equations for the mixed model under normality is given in Hartley and Rao [1967], and a computer program for the procedure is discussed in Hartley and Vaughan [1973]. Algorithms for improving this program are available in Hemmerle [1972].

Alternative suggestions for solving the maximum likelihood equations are made by Patterson and Thompson [1971], and similarly by Henderson [1973b]. General expressions for sampling variances of the large sample maximum likelihood estimators are given in Searle [1970], with specific applications to the 2-way nested classification in the same paper, and to the 3-way nested classification

in Rudan and Searle [1971]. These expressions utilize an analytical (as distinct from numerical) form of $\underline{V}^{-1}$ in each case. The apparent impossibility of deriving this for the 2-way crossed classification random model, with unbalanced data, is discussed by Searle and Rudan [1973].

4.8 An iterative procedure

Reductions in sums of squares arising in a natural way from the mixed model equations (53) are the basis for an iterative estimation procedure suggested by Cunningham and Henderson [1968]. Corrected by Thompson [1969], computing formulae for this method for the 2-way crossed classification mixed model, unbalanced data are now available in Searle [1973] for the no interaction case and in Corbeil and Searle [1973] for the interaction case.

4.9 MINQUE and BQUE procedures

Best quadratic unbiased estimators (BQUE's) summarized in Townsend and Searle [1971] have been generalized by Rao [1970, 1971a, b and 1972] in procedures he calls MINQUE and MIVQUE, To summarize them we rewrite the mixed model (44) as

$$\underline{y} = \underline{X}\underline{\beta} + \sum_{j=1}^{k} \underline{Z}_j \underline{u}_j$$

where $\underline{u}_j$ is the vector of N_j effects for the j'th random effects factor in the model (a main effects or interaction factor). Also, the k'th of such factors is defined as the error terms, $\underline{u}_k = \underline{e}$ and $\underline{Z}_k = \underline{I}$. Then with $E(\underline{u}_j) = \underline{0}$, $E(\underline{u}_j \underline{u}'_{j'}) = \underline{0}$ for $j \neq j'$, and $E(\underline{u}_j \underline{u}'_j) = \sigma^2_j \underline{I}_{N_j}$, except that $\sigma^2_k \equiv \sigma^2_e$ and $N_j \equiv N$ for N being the number of observations, we define the following terms.

$$\underline{V}_j = \underline{Z}_j \underline{Z}'_j, \quad \underline{V}^* = \sum_j \underline{V}_j \quad \text{and} \quad \underline{V} = \text{var}(\underline{y}) = \sum_j \sigma^2_j \underline{V}_j \quad .$$

[Rao uses the symbol $\underline{V}$ for $\underline{V}^{*}$ and $\underline{V}^{*}$ for $\underline{V}$, but the above notation is more compatible with using $\underline{V}$ for var($\underline{y}$).]. Rao then estimates a variance component by $\underline{y}'\underline{A}\underline{y}$ choosing $\underline{A}$, symmetric, so that the estimator is both unbiased and invariant to changes in $\underline{\beta}$. He suggests two different estimators. One minimizes the Euclidian norm $\text{tr}(\underline{V}^{*}\underline{A})^2$ and is called the MInimum Norm Quadratic Unbiased Estimator, or MINQUE. (See e.g. Rao [1971a, p. 268 and 1972 pp. 112-3].) Swallow and Searle [1973] have named this Basic MINQUE to distinguish it from what they call Alternative MINQUE, which minimizes $\text{tr}(\underline{V}\underline{A})^2$. This too is suggested by Rao [1971a, p. 268 and 1972, p. 113] although he does not use distinguishing names. The second estimator suggested by Rao [1972, pp. 447, 453] is one which has minimum variance, the MInimum Variance Quadratic Unbiased Estimator, MIVQUE, which is derived as $\underline{y}'\underline{A}\underline{y}$ by minimizing

$$\text{var}(\underline{y}'\underline{A}\underline{y}) = 2\text{tr}(\underline{V}\underline{A})^2 + \text{a term in } \underline{A} \text{ and kurtosis parameters.}$$

Under normality assumptions kurtosis parameters are zero and MIVQUE is then equivalent to alternative MINQUE. Rao's papers show that for

$$\underline{T} = \underline{V}^{-1} - \underline{V}^{-1}\underline{X}(\underline{X}'\underline{V}^{-1}\underline{X})^{-1}\underline{X}'\underline{V}^{-1}$$

and

$$\underline{S} = \{s_{ij}\} = \{\text{tr}(\underline{V}_i\underline{R}\underline{V}_j\underline{R})\} \text{ and } \underline{w} = \{w_j\} = \{\underline{y}'\underline{T}\underline{V}_j\underline{T}\underline{y}\}$$

for $i,j = 1,2,\cdots,k+1$, the vector of MIVQUE's under normality (alternative MINQUE's) is $\hat{\underline{\sigma}}^2 = \underline{S}'\underline{w}$. The same procedure used with $\underline{V}$ of $\underline{R}$ replaced by $\underline{V}^{*}$ gives the basic MINQUE's. This summary is given in Swallow and Searle [1973], whose main results are explicit expressions for these estimators and their sampling variances for unbalanced data in the 1-way classification, one model with non-zero μ and the other with $\mu \equiv 0$. In the latter case the MIVQUE procedure under normality gives, as it should, the BQUE (best quadratic unbiased estimation) results of Townsend and Searle [1971].

Suggestions by Henderson [1973c] connect the MINQUE procedure to expressions $\tilde{u}_j^o \tilde{u}_j^o$ derived from the mixed model predictor $\tilde{u}^o$ of (51). A similar expression is also seen in Patterson and Thompson [1971]. Additional estimators of a similar nature are given in LaMotte [1973a, b].

ACKNOWLEDGEMENTS

As indicated in Section 2, grateful acknowledgment is made to C. R. Henderson for his inspiration over many years in the topics of this paper. Preparation of this paper was partially supported by Grant GJ 31746 from the National Science Foundation, Washington, D. C.

REFERENCES

Cochran, W. G. [1951]. Improvement by means of selection. Proc. 2nd Berkeley Symposium. pp. 449-470.

Corbeil, R. R. and Searle, S. R. [1973]. An iterative procedure for estimating variance components in the 2-way crossed classification, mixed model, with interaction, using unbalanced data. Paper No. BU-460-M in the Biometrics Unit, Cornell University, Ithaca, New York.

Cunningham, E. P. and Henderson, C. R. [1968]. An iterative procedure for estimating fixed effects and variance components in mixed model situations. Biometrics 24, 13-25. Correction 25, 777-8.

Hartley, H. O. [1967]. Expectation, variances and covariances of ANOVA mean squares by 'synthesis.' Biometrics 23, 105-14. Correction 23, 853.

Hartley, H. O. and Rao, J. N. K. [1967]. Maximum likelihood estimation for the mixed analysis of variance model. Biometrika 54, 93-108.

Hartley, H. O. and Vaughan, W. K. [1972]. A computer program for the mixed analysis of variance model based on maximum likelihood. Chapter 8 in Statistical Papers in Honor of George W. Snedecor, T. A. Bancroft, Editor, Iowa State University Press.

Hemmerle, W. J. [1972]. Maximum likelihood algorithms for linear models with unequal variances. ARO Technical Report No. 11, Institute of Statistics, Texas A & M University, College Station, Texas. (Also available from the University of Rhode Island.)

Henderson, C. R. [1953]. Estimation of variance and covariance components. Biometrics 9, 226-52.

Henderson, C. R. [1963]. Selection index and expected genetic advance. Statistical Genetics in Plant Breeding, NAS-NRC publication No. 982.

Henderson, C. R. [1973a]. Sire evaluation and genetic trends. To appear in Animal Breeding and Genetics Symposium in Honor of Dr. Jay L. Lush. Am. Soc. of Animal Science (in press).

Henderson, C. R. [1973b]. Maximum likelihood estimation of variance components. Mimeo, Animal Science Department, Cornell University, Ithaca, New York, May 1973.

Henderson, C. R. [1973c]. MINQUE of variance components. Mimeo, Animal Science Department, Cornell University, Ithaca, New York, May 1973.

Henderson, C. R., Kempthorne, O., Searle, S. R., and Von Krosigk, C. N. [1959]. Estimation of environmental and genetic trends from records subject to culling. Biometrics 15, 192-218.

Henderson, C. R., Searle, S. R., and Schaeffer, L. R. [1973]. The uniqueness of Method 2 of estimating variance components. Submitted to Biometrics.

Kempthorne, O. [1972]. Discussion of Lindley and Smith [1972]. J. Roy. Stat. Soc. (B), 34, 33-37.

Koch, G. G. [1968]. Some further remarks on "A general approach to the estimation of variance components". Technometrics 10, 551-558.

LaMotte, L. R. [1973a]. Quadratic estimation of variance components. Biometrics 29, 311-330.

LaMotte, L. R. [1973b]. On non-negative quadratic unbiased estimation of variance components. J. Am. Stat. Assoc. (in press).

Lehmann, E. L. [1961]. Some Model I problems of selection. Ann. Math. Stat. 32, 990.

Lindley, D. V. and Smith, A. F. M. [1972]. Bayes estimates for the linear model. J. Roy. Stat. Soc. (B), 34, 1-41.

Lush, J. L. [1931]. The number of daughters necessary to prove a sire. J. Dairy Sci. 14, 209.

Lush, J. L. [1933]. The bull index problem in the light of modern genetics. J. Dairy Sci. 16, 501.

Lush, J. L. [1948]. The genetics of populations. Mimeo, Iowa State University, Ames, Iowa.

Mount, T. D. and Searle, S. R. [1972]. Estimating variance components in covariance models. Paper No. BU-403-M in the Biometrics Unit, Cornell University, Ithaca, New York.

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

Rao, C. R. [1965]. Linear Statistical Inference and its Applications. Wiley, New York.

Rao, C. R. [1970]. Estimation of heteroscedastic variances in linear models. J. Am. Stat. Assoc. 65, 161-172.

Rao, C. R. [1971a]. Estimation of variance and covariance components - MINQUE theory. J. Multivariate Anal. 1, 257-275.

Rao, C. R. [1971b]. Minimum variance quadratic unbiased estimation of variance components. J. Multivariate Anal. 1, 445-456.

Rao, C. R. [1972]. Estimation of variance and covariance components in linear models. J. Am. Stat. Assoc. 67, 112-115.

Rudan, J. W. and Searle, S. R. [1971]. Large sample variances of maximum likelihood estimators of variance components in the 3-way nested classification. Biometrics, 27, 1087-1091.

Searle, S. R. [1968]. Another look at Henderson's method of estimating variance components. Biometrics 24, 749-788.

Searle, S. R. [1970]. Large sample variances of maximum likelihood estimators of variance components. Biometrics 26, 505-24.

Searle, S. R. [1971a]. Linear Models. Wiley, New York.

Searle, S. R. [1971b]. Topics in variance components estimation. Biometrics 27, 1-76.

Searle, S. R. [1973]. Computing procedures for estimating variance components from unbalanced data in the 2-way crossed classification, no interaction mixed model. Paper No. BU-450-M in the Biometrics Unit, Cornell University, Ithaca, New York.

Searle, S. R. and Henderson, C. R. [1959]. Establishing age correction factors related to the level of herd production. J. Dairy Sci. 42, 824.

Searle, S. R. and Henderson, C. R. [1960]. Judging the effectiveness of age correction factors. J. Dairy Sci. 43, 966.

Searle, S. R. and Henderson, C. R. [1961]. Computing procedures for estimating components of variance in the two-way classification, mixed model. Biometrics 17, 607-16. Correction 23, 852.

Searle, S. R. and Rounsaville, T. R. [1973]. A remark on estimating covariance components. Paper No. BU-429-M in the Biometrics Unit, Cornell University, Ithaca, New York.

Searle, S. R. and Rudan, J. W. [1973]. Wanted: an inverse matrix. Communications in Statistics II, No. 2, August 1973.

Solomon, D. L. [1971]. A Bayesian interpretation and extension of the genetic selection index. Paper No. BU-362-M in the Biometrics Unit, Cornell University, Ithaca, New York.

Swallow, W. H. and Searle, S. R. [1973]. Minimum norm and minimum variance quadratic unbiased estimators of variance components from unbalanced data in the 1-way classification. Paper No. BU-447-M in the Biometrics Unit, Cornell University, Ithaca, New York.

Townsend, E. C. and Searle, S. R. [1971]. Best quadratic unbiased estimation of variance components from unbalanced data in the 1-way classification. Biometrics 27, 643-657.

Thompson, R. [1969]. Iterative estimation of variance components for non-orthogonal data. Biometrics 25, 767-773.

Wright, Sewall [1931]. On the evaluation of Dairy Sires. Proc. Am. Soc. Animal Prod. 1931,

APPENDIX

Detailed derivations are given here of certain results stated in Sections 2 and 3. Although several are well known, they are given for the sake of completeness.

2.1A Best prediction

Derivation

Minimize, for $\underline{A}$ positive definite and symmetric,

$$E(\tilde{\underline{u}} - \underline{u})'\underline{A}(\tilde{\underline{u}} - \underline{u}) = \iint(\tilde{\underline{u}} - \underline{u})'\underline{A}(\tilde{\underline{u}} - \underline{u})f(\underline{u},\underline{y})\, d\underline{u}\, d\underline{y}$$

$$= \int \left[\int(\tilde{\underline{u}} - \underline{u})'\underline{A}(\tilde{\underline{u}} - \underline{u})f(\underline{u}|\underline{y})\, d\underline{u} \right] f(\underline{y})\, d\underline{y}$$

where $f(\underline{u}|\underline{y})$ and $f(\underline{y})$ are conditional and marginal densities respectively. Minimizing with respect to $\tilde{\underline{u}}$ only requires minimizing of the integral over $\underline{u}$ and gives

$$\int(2\underline{A}\tilde{\underline{u}} - 2\underline{A}\underline{u})f(\underline{u}|\underline{y})\, d\underline{u} = 0 .$$

Hence, since $\underline{A}$ is positive definite,

$$\tilde{\underline{u}} = \frac{\int \underline{u} f(\underline{u}|\underline{y})\, d\underline{u}}{\int f(\underline{u}|\underline{y})\, d\underline{u}} = \int \underline{u} f(\underline{u}|\underline{y})\, d\underline{u} = E(\underline{u}|\underline{y}) . \tag{3}$$

Expectation

$$E(\tilde{\underline{u}}) = E_Y E_{U|Y}\,[E(\underline{u}|\underline{y})] = E_Y[E(\underline{u}|\underline{y})]$$

$$= \int \left[\int \underline{u} f(\underline{u}|\underline{y})\, d\underline{u} \right] f(\underline{y})\, d\underline{y}$$

$$= \iint \underline{u} f(\underline{u},\underline{y})\, d\underline{u}\, d\underline{y} = E(\underline{u}) . \tag{4}$$

Variances and covariances

$$\text{var}(\tilde{\underline{u}} - \underline{u}) = E(\tilde{\underline{u}} - \underline{u})(\tilde{\underline{u}} - \underline{u})', \text{ because } E(\tilde{\underline{u}} - \underline{u}) = \underline{0} \text{ from (4)},$$

$$= E_Y E_{U|Y}[E(\underline{u}|\underline{y})E(\underline{u}|\underline{y})' + \underline{u}\underline{u}' - \underline{u}E(\underline{u}|\underline{y})' - E(\underline{u}|\underline{y})\underline{u}']$$

$$= E_Y[E(\underline{u}|\underline{y})E(\underline{u}|\underline{y})' + E(\underline{u}\underline{u}'|\underline{y}) - 2E(\underline{u}|\underline{y})E(\underline{u}|\underline{y})']$$

$$= E_Y[E(\underline{u}\underline{u}'|\underline{y}) - E(\underline{u}|\underline{y})E(\underline{u}'\underline{y})']$$

$$= E_Y[\text{var}(\underline{u}|\underline{y})] \; . \tag{5}$$

$$\text{cov}(\tilde{\underline{u}}, \underline{u}') = E(\tilde{\underline{u}}\underline{u}') - E(\tilde{\underline{u}})E(\underline{u}')$$

$$= E_Y E_{U|Y}[E(\underline{u}|\underline{y})\underline{u}'] - E(\underline{u})E(\underline{u}')$$

$$= E_Y[E(\underline{u}|\underline{y})E(\underline{u}|\underline{y})'] - [E_Y E(\underline{u}|\underline{y})][E_Y E(\underline{u}|\underline{y})]'$$

$$= \text{var}[E(\underline{u}|\underline{y})] = \text{var}(\tilde{\underline{u}}) \tag{6}$$

$$\text{cov}(\tilde{\underline{u}}, \underline{y}') = E(\tilde{\underline{u}}\underline{y}') - E(\tilde{\underline{u}})E(\underline{y}')$$

$$= E_Y E_{U|Y}[E(\underline{u}|\underline{y})\underline{y}'] - E(\underline{u})E(\underline{y}')$$

$$= E_Y E_{U|Y}(\underline{u}\underline{y}') - E(\underline{u})E(\underline{y}')$$

$$= E(\underline{u}\underline{y}') - E(\underline{u})E(\underline{y}')$$

$$= \text{cov}(\underline{u}, \underline{y}') \tag{7}$$

Maximum correlation

As a function of $\underline{y}$ let p be any predictor of u, an element of $\underline{u}$. Then

$$\text{cov}(p,u) = E\{[p - E(p)][u - E(u)]\}$$

$$= E\{[p - E(p)]u\}$$

$$= E_Y E_{U|Y}\{[p - E(p)]u\}$$

$$= E_Y\{[p - E(p)]E(u|\underset{\sim}{y})\} \text{ because } p \text{ is a function of } \underset{\sim}{y}$$

$$= E_Y\{[p - E(p)]\tilde{u}\}$$

$$= \text{cov}(p,\tilde{u}).$$

When $p = \tilde{u}$, $\text{cov}(\tilde{u},u) = \text{cov}(\tilde{u},\tilde{u}) = \sigma^2_{\tilde{u}}$

and

$$\rho(\tilde{u},u) = \frac{\text{cov}(\tilde{u},u)}{\sigma_{\tilde{u}}\sigma_u} = \frac{\sigma_{\tilde{u}}}{\sigma_u} \quad . \tag{8}$$

Hence in general

$$\rho^2(p,u) = \frac{\text{cov}^2(p,u)}{\sigma^2_p\sigma^2_u} = \frac{\text{cov}^2(p,\tilde{u})}{\sigma^2_p\sigma^2_{\tilde{u}}}\frac{\sigma^2_{\tilde{u}}}{\sigma^2_u} = \rho^2(p,\tilde{u})\rho^2(\tilde{u},u) \ .$$

For choice of p this is maximum when $\rho^2(p,\tilde{u}) = 1$, i.e., $p = \tilde{u}$. Hence (8) is maximum $\rho(\tilde{u},u)$. This proof follows Rao [1965, p. 221].

2.2A Best linear prediction

Matrix results

When $\text{tr}(\underset{\sim}{X}\underset{\sim}{P})$ exists its value is $\Sigma\Sigma x_{ij}p_{ji}$. Hence $\frac{\partial}{\partial x_{ij}}\text{tr}(\underset{\sim}{X}\underset{\sim}{P}) = p_{ji}$

and so $$\frac{\partial}{\partial \underset{\sim}{X}}\text{tr}(\underset{\sim}{X}\underset{\sim}{P}) \overset{\text{def}}{\equiv} \left\{\frac{\partial}{\partial x_{ij}}\text{tr}(\underset{\sim}{X}\underset{\sim}{P})\right\} = \{p_{ji}\} = \underset{\sim}{P}' \ .$$

Also, because $\text{tr}(\underset{\sim}{X}\underset{\sim}{P}) = \text{tr}(\underset{\sim}{P}\underset{\sim}{X})$

$$\frac{\partial}{\partial \underset{\sim}{X}}\text{tr}(\underset{\sim}{X}\underset{\sim}{P}) = \frac{\partial}{\partial \underset{\sim}{X}}\text{tr}(\underset{\sim}{P}\underset{\sim}{X}) = \underset{\sim}{P}' \ . \tag{A1}$$

And since $\text{tr}(\underset{\sim}{X}'\underset{\sim}{P}) = \text{tr}(\underset{\sim}{P}\underset{\sim}{X}')$

$$\frac{\partial}{\partial \underset{\sim}{X}}\text{tr}(\underset{\sim}{X}'\underset{\sim}{P}) = \frac{\partial}{\partial \underset{\sim}{X}}\text{tr}(\underset{\sim}{P}\underset{\sim}{X}') = \underset{\sim}{P} \ . \tag{A2}$$

Hence

$$\frac{\partial}{\partial \underline{X}} tr(\underline{P}_1 \underline{X}' \underline{Q} \underline{X} \underline{P}_2) = \underline{Q} \underline{X} \underline{P}_2 \underline{P}_1 + \underline{Q}' \underline{X} \underline{P}_1' \underline{P}_2' \quad . \tag{A3}$$

Also

$$\frac{\partial}{\partial \underline{X}} tr(\underline{P}_1 \underline{P}_2 \underline{X} \underline{Q}_1 \underline{Q}_2) = \frac{\partial}{\partial \underline{X}} tr(\underline{X} \underline{Q}_1 \underline{Q}_2 \underline{P}_1 \underline{P}_2) = \underline{P}_2' \underline{P}_1' \underline{Q}_2' \underline{Q}_1' \tag{A4}$$

and

$$\frac{\partial}{\partial \underline{X}} (\underline{r}' \underline{P} \underline{X} \underline{Q} \underline{s}) = \frac{\partial}{\partial \underline{X}} tr(\underline{r}' \underline{P} \underline{X} \underline{Q} \underline{s}) = \underline{P}' \underline{r} \underline{s}' \underline{Q} \quad . \tag{A5}$$

Minimization

For $\tilde{\underline{u}} = \underline{a} + \underline{B}\underline{y}$ we minimize, for positive definite symmetric $\underline{A}$,

$$E(\tilde{\underline{u}} - \underline{u})' \underline{A} (\tilde{\underline{u}} - \underline{u}) = E(\underline{a} + \underline{B}\underline{y} - \underline{u})' \underline{A} (\underline{a} + \underline{B}\underline{y} - \underline{u})$$

$$= E \left\{ \underline{a}' \underline{A} \underline{a} + 2\underline{a}' \underline{A} \underline{B} \underline{y} - 2\underline{a}' \underline{A} \underline{u} + (\underline{u}' \ \underline{y}') \begin{bmatrix} -\underline{I} \\ \underline{B}' \end{bmatrix} \underline{A} (-\underline{I} \ \underline{B}) \begin{bmatrix} \underline{u} \\ \underline{y} \end{bmatrix} \right\}$$

$$= \underline{a}' \underline{A} \underline{a} + 2\underline{a}' \underline{A} \underline{B} \underline{\mu}_Y - 2\underline{a}' \underline{A} \underline{\mu}_U$$

$$+ (\underline{\mu}_U' \ \underline{\mu}_Y') \begin{bmatrix} -\underline{A} & -\underline{A}\underline{B} \\ -\underline{B}'\underline{A} & \underline{B}'\underline{A}\underline{B} \end{bmatrix} \begin{bmatrix} \underline{\mu}_U \\ \underline{\mu}_Y \end{bmatrix} + tr \begin{bmatrix} -\underline{A} & -\underline{A}\underline{B} \\ -\underline{B}'\underline{A} & \underline{B}'\underline{A}\underline{B} \end{bmatrix} \begin{bmatrix} \underline{V}_U & \underline{C} \\ \underline{C}' & \underline{V} \end{bmatrix}$$

$$= \underline{a}' \underline{A} \underline{a} + 2\underline{a}' \underline{A} \underline{B} \underline{\mu}_Y - 2\underline{a}' \underline{A} \underline{\mu}_U$$

$$- \underline{\mu}_U' \underline{A} \underline{\mu}_U - 2\underline{\mu}_U' \underline{A} \underline{B} \underline{\mu}_Y + \underline{\mu}_Y' \underline{B}' \underline{A} \underline{B} \underline{\mu}_Y$$

$$+ tr(-\underline{A}\underline{V}_U - \underline{A}\underline{B}\underline{C}' - \underline{B}'\underline{A}\underline{C} + \underline{B}'\underline{A}\underline{B}\underline{V}) \quad . \tag{A6}$$

Using (A1) - (A5) to differentiate this with respect to the elements of $\underline{B}$ gives

$$2\underline{A}'\underline{a}\underline{\mu}_Y' - 2\underline{A}'\underline{\mu}_U\underline{\mu}_Y' + \underline{A}\underline{B}\underline{\mu}_Y\underline{\mu}_Y' + \underline{A}'\underline{B}\underline{\mu}_Y'\underline{\mu}_Y - \underline{A}'\underline{C} - \underline{A}\underline{C} + \underline{A}\underline{B}\underline{V} + \underline{A}'\underline{B}\underline{V}' = \underline{0} .$$

The symmetry of $\underline{A}$ and $\underline{V}$ and the non-singularity of $\underline{A}$ reduce this to

$$(\underline{a} - \underline{\mu}_U + \underline{B}\underline{\mu}_Y)\underline{\mu}_Y' + \underline{B}\underline{V} = \underline{C} . \tag{A7}$$

Differentiating (A6) with respect to $\underline{a}$ gives $\underline{a} = \underline{\mu}_U - \underline{B}\underline{\mu}_Y$ so that (A7) gives $B = \underline{C}\underline{V}^{-1}$ and hence

$$\tilde{\underline{u}} = \underline{a} + \underline{B}\underline{y} = \underline{\mu}_Y + \underline{C}\underline{V}^{-1}(\underline{y} - \underline{\mu}_Y) \tag{16}$$

2.4A Mixed model prediction

With $\underline{w} = \underline{K}'\underline{\beta} + \underline{u}$ and $\tilde{\underline{w}} = \underline{B}\underline{y}$ for $\underline{B}\underline{X} = \underline{K}'$ we have

$$E\begin{bmatrix} \underline{w} \\ \underline{y} \end{bmatrix} = \begin{bmatrix} \underline{K}'\underline{\beta} \\ \underline{X}\underline{\beta} \end{bmatrix} \quad \text{and} \quad \text{var}\begin{bmatrix} \underline{w} \\ \underline{y} \end{bmatrix} = \begin{bmatrix} \underline{V}_U & \underline{C} \\ \underline{C}' & \underline{V} \end{bmatrix} ,$$

and seek to minimize $E(\tilde{\underline{w}} - \underline{w})'\underline{A}(\tilde{\underline{w}} - \underline{w})$ for $\underline{A}$ being positive definite symmetric. We minimize

$$\lambda = E(\tilde{\underline{w}} - \underline{w})'\underline{A}(\tilde{\underline{w}} - \underline{w}) + \text{tr}[\underline{T}(\underline{B}\underline{X} - \underline{K}')]$$

where $\underline{T}$ is a matrix of Lagrange multipliers. Since $\underline{A}$ is positive definite there is no loss of generality in writing $\underline{T} = 2\underline{M}\underline{A}$ for some matrix $\underline{M}$. Then

$$\lambda = E(\underline{B}\underline{y} - \underline{w})'\underline{A}(\underline{B}\underline{y} - \underline{w}) + 2\text{tr}[\underline{M}\underline{A}(\underline{B}\underline{X} - \underline{K}')]$$

$$= E(\underline{w}' \; \underline{y}') \begin{bmatrix} -\underline{I} \\ \underline{B} \end{bmatrix} \underline{A}(-\underline{I} \; \underline{B}) \begin{bmatrix} \underline{w} \\ \underline{y} \end{bmatrix} + 2\mathrm{tr}[\underline{M}\underline{A}(\underline{B}\underline{X} - \underline{K}')]$$

$$= (\underline{\beta}'\underline{K} \quad \underline{\beta}'\underline{X}') \begin{bmatrix} -\underline{A} & -\underline{A}\underline{B} \\ -\underline{B}'\underline{A} & \underline{B}'\underline{A}\underline{B} \end{bmatrix} \begin{bmatrix} \underline{K}'\underline{\beta} \\ \underline{X}\underline{\beta} \end{bmatrix} + \mathrm{tr} \begin{bmatrix} -\underline{A} & -\underline{A}\underline{B} \\ -\underline{B}'\underline{A} & \underline{B}'\underline{A}\underline{B} \end{bmatrix} \begin{bmatrix} \underline{V}_U & \underline{C} \\ \underline{C}' & \underline{V} \end{bmatrix} + 2\mathrm{tr}[\underline{M}\underline{A}(\underline{B}\underline{X} - \underline{K}')]$$

$$= -\underline{\beta}'\underline{K}\underline{A}\underline{K}'\underline{\beta} - 2\underline{\beta}'\underline{K}\underline{A}\underline{B}\underline{X}\underline{\beta} + \underline{\beta}'\underline{X}'\underline{B}'\underline{A}\underline{B}\underline{X}\underline{\beta}$$

$$+ \mathrm{tr}(-\underline{A}\underline{V}_U - \underline{A}\underline{B}\underline{C}' - -\underline{B}'\underline{A}\underline{C} + \underline{B}'\underline{A}\underline{B}\underline{V}) + 2\mathrm{tr}[\underline{M}\underline{A}(\underline{B}\underline{X} - \underline{K}')] \; .$$

Using the matrix results (A1) through (A5) gives $\partial\lambda/\partial \underline{B} = \underline{0}$ as (using $\underline{A} = \underline{A}'$)

$$- 2\underline{A}\underline{K}'\underline{\beta}\underline{\beta}'\underline{X}' + \underline{A}\underline{B}\underline{X}\underline{\beta}\underline{\beta}'\underline{X}' + \underline{A}\underline{B}\underline{X}\underline{\beta}\underline{\beta}'\underline{X}' - \underline{A}\underline{C} - \underline{A}\underline{C} + \underline{A}\underline{B}\underline{V} + \underline{A}\underline{B}\underline{V} + 2\underline{A}\underline{M}'\underline{X}' = \underline{0}. \qquad (A8)$$

And $\partial\lambda/\partial \underline{M} = \underline{0}$ gives $\underline{B}\underline{X} = \underline{K}'$ because $\underline{A}$ is non-singular. Using these results in (A7) gives

$$- \underline{B}\underline{X}\underline{\beta}\underline{\beta}'\underline{X}' + \underline{B}\underline{X}\underline{\beta}\underline{\beta}'\underline{X}' - \underline{C} + \underline{B}\underline{V} + \underline{M}'\underline{X}' = \underline{0}$$

i.e.
$$\underline{B}\underline{V} + \underline{M}'\underline{X}' = \underline{C} \; . \qquad (A9)$$

This and
$$\underline{B}\underline{X} = \underline{K}' \qquad (A10)$$

are the equations to be solved for $\underline{B}$. From (A9)

$$\underline{B} = (\underline{C} - \underline{M}'\underline{X}')\underline{V}^{-1} \qquad (A11)$$

and substitution in (A10) gives

$$\underline{M}' = (\underline{C}\underline{V}^{-1}\underline{X} - \underline{K}')(\underline{X}'\underline{V}^{-1}\underline{X})^{-}$$

so that in (A11)

$$\underline{B} = \underline{C}\underline{V}^{-1} + (\underline{K}' - \underline{C}\underline{V}^{-1}\underline{X})(\underline{X}'\underline{V}^{-1}\underline{X})^{-}\underline{X}'\underline{V}^{-1} \; . \qquad (31)$$

Variances and covariances

$$\underline{X}\underline{\beta}^{\circ} = \underline{X}(\underline{X}'\underline{V}^{-1}\underline{X})^{-}\underline{X}'\underline{V}^{-1}\underline{y}$$

$$\text{var}(\underline{X}\underline{\beta}^{\circ}) = \underline{X}(\underline{X}'\underline{V}^{-1}\underline{X})^{-}\underline{X}'\underline{V}^{-1}\underline{X}(\underline{X}'\underline{V}^{-1}\underline{X})^{-'}\underline{X}'$$

$$= \underline{X}(\underline{X}'\underline{V}^{-1}\underline{X})^{-}\underline{X}'\underline{V}'\underline{X}(\underline{X}'\underline{V}^{-1}\underline{X})^{-'}\underline{X}'\underline{V}^{-\frac{1}{2}}\underline{V}^{\frac{1}{2}} \text{ because } \underline{V} \text{ is positive definite,}$$

$$= \underline{X}(\underline{X}'\underline{V}^{-1}\underline{X})^{-}(\underline{X}'\underline{V}^{-\frac{1}{2}})\underline{V}^{\frac{1}{2}}, \text{ because } \underline{Q}'\underline{Q}(\underline{Q}'\underline{Q})^{-}\underline{Q}' = \underline{Q}' \text{ for any } \underline{Q},$$

$$= \underline{X}(\underline{X}'\underline{V}^{-1}\underline{X})^{-}\underline{X}' .$$

$$\underline{K}' = \underline{B}\underline{X}$$

$$\text{var}(\underline{K}'\underline{\beta}^{\circ}) = \underline{B}\underline{X}(\underline{X}'\underline{V}^{-1}\underline{X})^{-}\underline{X}'\underline{B}'$$

$$= \underline{K}'(\underline{X}'\underline{V}^{-1}\underline{X})\underline{K} \qquad (35)$$

$$\text{cov}(\underline{X}\underline{\beta}^{\circ},\underline{y}') = \underline{X}(\underline{X}'\underline{V}^{-1}\underline{X})^{-}\underline{X}'\underline{V}^{-1}\underline{V}$$

$$= \underline{X}(\underline{X}'\underline{V}^{-1}\underline{X})^{-}\underline{X}'$$

$$= \text{var}(\underline{X}\underline{\beta}^{\circ})$$

$$\tilde{\underline{u}}^{\circ} = \underline{C}\underline{V}^{-1}(\underline{y} - \underline{X}\underline{\beta}^{\circ})$$

$$\text{var}(\tilde{\underline{u}}^{\circ}) = \underline{C}\underline{V}^{-1}\underline{V}\underline{V}^{-1}\underline{C}' + \underline{C}\underline{V}^{-1}[-\text{var}(\underline{X}\underline{\beta}^{\circ})]\underline{V}^{-1}\underline{C}'$$

$$= \underline{C}\underline{V}^{-1}\underline{C}' - \underline{C}\underline{V}^{-1}\underline{X}(\underline{X}'\underline{V}^{-1}\underline{X})^{-}\underline{X}'\underline{V}^{-1}\underline{C} \qquad (36)$$

$$\text{cov}(\underline{X}\underline{\beta}^{\circ},\tilde{\underline{u}}^{\circ}') = \text{cov}[\underline{X}\underline{\beta}^{\circ}, (\underline{y} - \underline{X}\underline{\beta}^{\circ})'\underline{V}^{-1}\underline{C}']$$

$$= [\text{var}(\underline{X}\underline{\beta}^{\circ}) - \text{var}(\underline{X}\underline{\beta}^{\circ})]\underline{V}^{-1}\underline{C}$$

$$= \underline{0} \qquad (37)$$

$$\tilde{w} = K\beta^{o} + u^{o}$$

$$\operatorname{var}(\tilde{w}) = \operatorname{var}(K\beta^{o}) + \operatorname{var}(u^{o}) \tag{38}$$

$$\begin{aligned}\operatorname{cov}(\tilde{u}^{o}, u') &= \operatorname{cov}\{CV^{-1}[I - X(X'V^{-1}X)^{-}X'V^{-1}]y, u'\} \\ &= CV^{-1}[I - X(X'V^{-1}X)^{-}X'V^{-1}]C' \\ &= \operatorname{var}(u^{o}) \qquad (39)\end{aligned}$$

$$\begin{aligned}\operatorname{var}(\tilde{u}^{o} - u) &= \operatorname{var}(u) - \operatorname{var}(u^{o}) \\ &= V_{U} - \operatorname{var}(u^{o}) \qquad (40)\end{aligned}$$

$$\begin{aligned}\operatorname{cov}(K'\beta^{o}, u') &= \operatorname{cov}(BX\beta^{o}, u') \\ &= BX(X'V^{-1}X)^{-}X'V^{-1}C' \qquad (41)\end{aligned}$$

$$\begin{aligned}\operatorname{var}(\tilde{w} - w) &= \operatorname{var}(K'\beta^{o} + \tilde{u}^{o} - K'\beta - u) \\ &= \operatorname{var}(K'\beta^{o}) + \operatorname{var}(u^{o} - u) + \operatorname{cov}[K'\beta^{o}, (u^{o} - u)'] \\ &\qquad + \operatorname{cov}[(u^{o} - u), (K'\beta^{o})'] \\ &= \operatorname{var}(K'\beta^{o}) + \operatorname{var}(u^{o} - u) - \operatorname{cov}(K'\beta^{o}, u') - \operatorname{cov}(u, \beta^{o}'K) \qquad (42)\end{aligned}$$

3.1A Calculating the predictor (in the mixed model)

From the second equation of

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}D \\ Z'R^{-1}X & Z'R^{-1}Z + D^{-1} \end{bmatrix} \begin{bmatrix} \beta^{o} \\ u^{o} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix} \tag{A12}$$

we get

$$u^{o} = (Z'R^{-1}Z + D^{-1})^{-1}Z'R^{-1}(y - X\beta^{o}) \tag{A13}$$

So long as $\underset{\sim}{D} = \text{var}(\underset{\sim}{u})$ is non-singular $(\underset{\sim}{Z}'\underset{\sim}{R}^{-1}\underset{\sim}{Z} + \underset{\sim}{D}^{-1})^{-1}$ always exists, because $\underset{\sim}{R}^{-1}$ and $\underset{\sim}{D}^{-1}$ are symmetric, and equal to $\underset{\sim}{P}'\underset{\sim}{P}$ and $\underset{\sim}{Q}^{-1'}\underset{\sim}{Q}^{-1}$ say, and so

$$\underset{\sim}{Z}'\underset{\sim}{R}^{-1}\underset{\sim}{Z} + \underset{\sim}{D}^{-1} = \underset{\sim}{Q}^{-1'}(\underset{\sim}{K}'\underset{\sim}{K} + \underset{\sim}{I})\underset{\sim}{Q}^{-1} \text{ for } \underset{\sim}{K} = \underset{\sim}{PZQ}$$

is non-singular (Searle [1971, p. 24, lemma 8]). (A13) always holds, therefore, and substituting it into the first equation of (A12) gives, after a little reduction,

$$\underset{\sim}{X}'\underset{\sim}{W}\underset{\sim}{X}\beta^{\circ} = \underset{\sim}{X}'\underset{\sim}{W}\underset{\sim}{y} \tag{A14}$$

where

$$\underset{\sim}{W} = \underset{\sim}{R}^{-1} - \underset{\sim}{R}^{-1}\underset{\sim}{Z}(\underset{\sim}{Z}'\underset{\sim}{R}^{-1}\underset{\sim}{Z} + \underset{\sim}{D}^{-1})^{-1}\underset{\sim}{Z}'\underset{\sim}{R}^{-1} .$$

It remains to show that $\underset{\sim}{WV} = \underset{\sim}{I}$ which it does:

$$\begin{aligned}
\underset{\sim}{WV} &= [\underset{\sim}{R}^{-1} - \underset{\sim}{R}^{-1}\underset{\sim}{Z}(\underset{\sim}{Z}'\underset{\sim}{R}^{-1}\underset{\sim}{Z} + \underset{\sim}{D}^{-1})^{-1}\underset{\sim}{Z}'\underset{\sim}{R}^{-1}](\underset{\sim}{ZDZ}' + \underset{\sim}{R}) \\
&= \underset{\sim}{R}^{-1}\underset{\sim}{ZDZ}' + \underset{\sim}{I} - \underset{\sim}{R}^{-1}\underset{\sim}{Z}(\underset{\sim}{Z}'\underset{\sim}{R}^{-1}\underset{\sim}{Z} + \underset{\sim}{D}^{-1})^{-1}(\underset{\sim}{Z}'\underset{\sim}{R}^{-1}\underset{\sim}{ZDZ}' + \underset{\sim}{Z}') \\
&= \underset{\sim}{R}^{-1}\underset{\sim}{ZDZ}' + \underset{\sim}{I} - \underset{\sim}{R}^{-1}\underset{\sim}{Z}(\underset{\sim}{Z}'\underset{\sim}{R}^{-1}\underset{\sim}{Z} + \underset{\sim}{D}^{-1})^{-1}(\underset{\sim}{Z}'\underset{\sim}{R}^{-1}\underset{\sim}{Z} + \underset{\sim}{D}^{-1})\underset{\sim}{DZ}' \\
&= \underset{\sim}{R}^{-1}\underset{\sim}{ZDZ}' + \underset{\sim}{I} - \underset{\sim}{R}^{-1}\underset{\sim}{ZDZ}' \\
&= \underset{\sim}{I} .
\end{aligned}$$

With $\underset{\sim}{W}$ and $\underset{\sim}{V}$ both being symmetric this implies $\underset{\sim}{W} = \underset{\sim}{V}^{-1}$. Hence (A14) is $\underset{\sim}{X}'\underset{\sim}{V}^{-1}\underset{\sim}{X}\beta^{\circ} = \underset{\sim}{X}'\underset{\sim}{V}^{-1}\underset{\sim}{y}$ for which $\beta^{\circ} = (\underset{\sim}{X}'\underset{\sim}{V}^{-1}\underset{\sim}{X})^{-}\underset{\sim}{X}'\underset{\sim}{V}^{-1}\underset{\sim}{y}$ is a familiar solution, as in (50). It remains to show that $\underset{\sim}{u}^{\circ}$ of (A13) is $\underset{\sim}{u}^{\circ} = \underset{\sim}{DZ}'\underset{\sim}{V}^{-1}(\underset{\sim}{y} - \underset{\sim}{X}\beta^{\circ})$ of (51). It is, because in (A13)

$$\begin{aligned}
(\underset{\sim}{Z}'\underset{\sim}{R}^{-1}\underset{\sim}{Z} + \underset{\sim}{D}^{-1})^{-1}\underset{\sim}{Z}'\underset{\sim}{R}^{-1} &= (\underset{\sim}{Z}'\underset{\sim}{R}^{-1}\underset{\sim}{Z} + \underset{\sim}{D}^{-1})^{-1}\underset{\sim}{Z}'\underset{\sim}{R}^{-1}\underset{\sim}{VV}^{-1} \\
&= (\underset{\sim}{Z}'\underset{\sim}{R}^{-1}\underset{\sim}{Z} + \underset{\sim}{D}^{-1})^{-1}\underset{\sim}{Z}'\underset{\sim}{R}^{-1}(\underset{\sim}{ZDZ}' + \underset{\sim}{R})\underset{\sim}{V}^{-1} \\
&= (\underset{\sim}{Z}'\underset{\sim}{R}^{-1}\underset{\sim}{Z} + \underset{\sim}{D}^{-1})^{-1}(\underset{\sim}{Z}'\underset{\sim}{R}^{-1}\underset{\sim}{Z} + \underset{\sim}{D}^{-1})\underset{\sim}{DZ}'\underset{\sim}{V}^{-1} \\
&= \underset{\sim}{DZ}'\underset{\sim}{V}^{-1} .
\end{aligned}$$