BU-1445-M

A MATHEMATICAL SUPPLEMENT

TO C.R. HENDERSON'S BOOK "APPLICATIONS OF LINEAR MODELS IN ANIMAL BREEDING"

Shayle R. Searle

Professor Emeritus
Department of Biometrics,
College of Agriculture & Life Sciences,
Cornell University,
Ithaca, N.Y., 14853

University of Guelph

1998

PREFACE

Professor Charles R. Henderson (1911–1989) of Cornell University took his only sabbatical leave in New Zealand during 1955-6. At that time I was Research Statistician in the Herd Improvement Department of the New Zealand Dairy Board, which had sponsored Henderson's coming to New Zealand. As a result, I was lucky enough to have him as an office mate for nine months. That was a great opportunity to get to know him well before coming to Cornell, where he was my Ph.D. advisor, 1956-8. It therefore gives me great pleasure to offer these notes on his book "Applications of Linear Models in Animal Breeding", published by the University of Guelph, 1984. It is referenced in this Supplement as CRH.

Those well acquainted with Professor Henderson's lectures and writings would agree that he was an enormous source of great ideas – but sometimes his conveying of them, either in lecturing or writing, was not at the same high level as the originality of those ideas. I believe that to be true of his book, too. My reaction to a first reading of it was that it could do with a little tidying up, especially with respect to mathematical clarity and detailed derivation of many of the formulae quoted and used in applications. A number of professional animal breeders have told me they feel the same way.

Their encouragement flamed my own interests and this Supplement is the result. And supplement it truly is: it is *not* a re-writing of the book. But it *is* designed to be read solely in conjunction with the book. As such it pays scant attention to CRH's many arithmetical examples, except in the last dozen or so chapters, where development is given of some of the numerical equations and their solutions. This is in concert with the overall objective of this Supplement, to provide mathematical fullness for the development of many of the algebraic results which are quoted and used, often with meager back-up. Stemming from this objective are ideas in the book which I do not like (e.g., MIVQUE and approximations thereto), or do not understand and/or which I think are wrong. At these places I have not hesitated to make personal (opinionated!) comment and to pose questions I cannot answer and problems I have been unable to solve. Hopefully, the clarity of such reactions will prompt others to provide solutions.

To Norma Phalen I extend my sincere thanks for her typing all this algebra. Her patience is incredible.

Finally, my heartfelt thanks go to Larry Schaeffer of the University of Guelph for supporting preparation and distribution of this Supplement, and for his help in correcting what were some of my blatant mistakes. Others undoubtedly remain. They are all mine. Corrections are eagerly sought.

Shayle R. Searle
505 The Parkway
Ithaca, N.Y., 14850

October 1998

NOTATION and LAYOUT

Chapters, paragraphs and page numbers As much as possible the notation follows that of the book. Chapters correspond to those of CRH; and paragraphs, which are numbered, for example, as 1.1, 1.2, 1.3, $\cdots$, often coincide with those of the book. Page numbers are shown, for example, as [3] for page 3 of the book, [3, 1.2] for paragraph 1.2 on page 3 of the book, and [3, (1.3)] for equation (1.3) on page 3 of the book.

Equation numbers Equations with decimal numbers are those of the book. Equations without decimal numbers are mine – they are numbered consecutively, starting with (1) in each chapter.

Bold Face font To conserve time and effort, bold font has *not* been used for matrices and vectors, except in places where distinction of vectors from scalars might otherwise be too confusing.

Consistency Every attempt has been made to be consistent in both notation and cross references. But, in view of the considerable effort required for complete consistency, no excruciating endeavour has been made in this connection.

Books Referenced by Acronym Back-up of many topics in CRH is detailed in one or more of the following four books which are therefore frequently referenced (by acronym) in this Supplement.

LM:	"Linear Models", Searle, Wiley, 1971.
MAUFS:	"Matrix Algebra Useful for Statistics", Searle, Wiley, 1982.
LMFUD:	"Linear Models for Unbalanced Data", Searle, Wiley, 1987.
VC:	"Variance Components", Searle, Casella and McCulloch, Wiley, 1992.

TABLE OF CONTENTS

Chapter 1

Constructing a Linear Model

The starting point is the familiar model equation

$$y = X\beta + Zu + e. \tag{1.1}$$

$$X_{n\times p},\quad Z_{n\times q}, \text{ known, } r(X) = r \leq p \leq n.$$

$$\beta_{p\times 1}, \text{ fixed effects; usually unknown.}$$

$$u_{q\times 1} \sim (0,\ G) \qquad e_{N\times 1} \sim (0,\ R) \qquad \mathrm{cov}(u, e') = 0$$

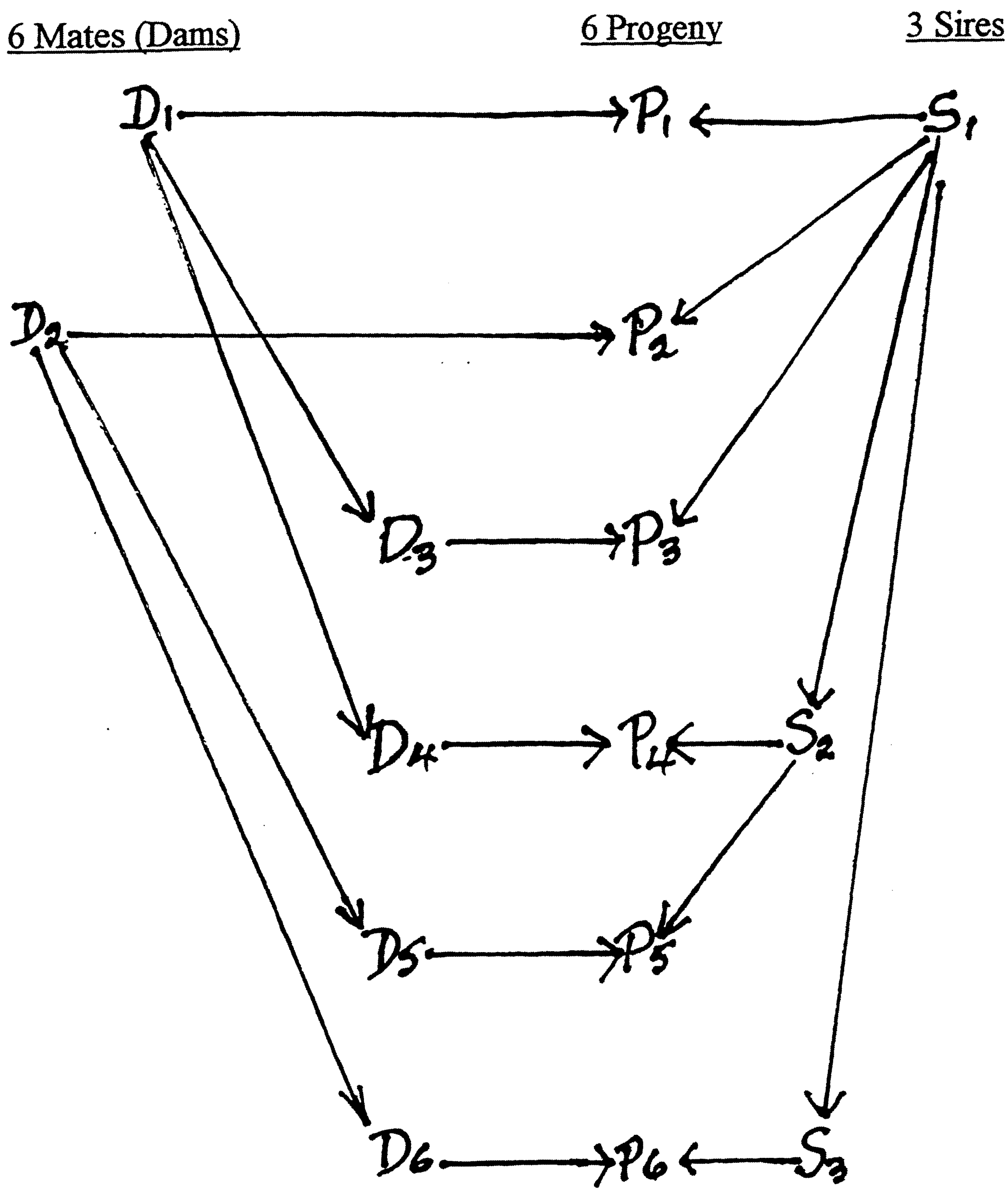
$$y \sim (X\beta,\ V = ZGZ' + R)$$

G and R are usually non-singular.

1.1 (Example) Simple regression [3, 1.1]

$$y_i = \mu + x_i\alpha + e_i$$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{bmatrix}; \quad X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_i \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \beta = \begin{bmatrix} \mu \\ \alpha \end{bmatrix}, \quad e = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_i \\ \vdots \\ e_n \end{bmatrix}.$$



Genetic Relationships for the Matrices on [4] and on pages 2 and 3.

and from considering the variance of any single record

$$\sigma_\epsilon^2 = \sigma_y^2 - \sigma_a^2 = (1-h^2)\sigma_y^2$$

and

$$\sigma_e^2 = \sigma_y^2 - \sigma_s^2 = (1-\frac{1}{4}h^2)\sigma_y^2$$

Therefore

$$R\sigma_e^2 = A_p h^2\sigma_y^2 + I(1-h^2)\sigma_y^2 - Z'A_sZ\frac{1}{4}h^2\sigma_y^2.$$

Thus

$$R = [A_p h^2 + I(1-h^2) - Z'A_sZ\frac{1}{4}h^2]\sigma_y^2/\sigma_e^2.$$

On using $h^2 = \frac{1}{4}$, and the scaling factor of $\sigma_y^2/\sigma_e^2 = 16/15$, along with A_p, Z and A_s we get

$$\frac{15}{16}R = \frac{1}{64}\begin{bmatrix} 16 & 4 & 6 & 4 & 2 & 2 \\ & 16 & 4 & 2 & 4 & 4 \\ & & 16 & 3 & 2 & 2 \\ & & & 16 & 4 & 1 \\ & \text{Sym} & & & 16 & 2 \\ & & & & & 16 \end{bmatrix} + \frac{3}{4}I - \frac{1}{16}\begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{2} \\ 1 & \frac{1}{2} & \frac{1}{2} \\ 1 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 1 & \frac{1}{4} \\ \frac{1}{2} & 1 & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{4} & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

and so

$$R = \frac{1}{60}\begin{bmatrix} 16 & 4 & 6 & 4 & 2 & 2 \\ & 16 & 4 & 2 & 4 & 4 \\ & & 16 & 3 & 2 & 2 \\ & & & 16 & 4 & 1 \\ & \text{Sym} & & & 16 & 2 \\ & & & & & 16 \end{bmatrix} + \frac{4}{5}I - \frac{1}{15}\begin{bmatrix} 1 & 1 & 1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 1 & 1 & 1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 1 & 1 & 1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & 1 & 1 & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & 1 & 1 & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{4} & \frac{1}{4} & 1 \end{bmatrix}$$

which, as $R\sigma_e^2$, is the third 6×6 matrix on [4].

As on [5]

$$G = \mathrm{var}(u) = \begin{bmatrix} Ag_{11} & Ag_{12} \\ Ag_{12} & Ag_{22} \end{bmatrix} \qquad R = \mathrm{var}(e) = \begin{bmatrix} Ir_{11} & Ir_{12} \\ Ir_{12} & Ir_{22} \end{bmatrix}.$$

1.4 Two-way mixed model [5, 1.4]

n_{ij} values			
Sires	Treatments		$n_{i.}$
	$j=1$	$j=2$	
$i=1$	2	1	3
$i=2$	0	2	2
$i=3$	3	0	3
$n_{.j}$	5	3	8

$$\begin{aligned} y_{ijk} &= \mu + t_j + s_i + (st)_{ij} + e_{ijk} \\ y &= X\beta + Zu + e \end{aligned}$$

$$\begin{bmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{221} \\ y_{222} \\ y_{311} \\ y_{312} \\ y_{313} \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdot \\ 1 & 1 & \cdot \\ 1 & \cdot & 1 \\ 1 & \cdot & 1 \\ 1 & \cdot & 1 \\ 1 & 1 & \cdot \\ 1 & 1 & \cdot \\ 1 & 1 & \cdot \end{bmatrix} \begin{bmatrix} \mu \\ t_1 \\ t_2 \end{bmatrix} + \begin{bmatrix} 1 & \cdot & \cdot & 1 & \cdot & \cdot & \cdot \\ 1 & \cdot & \cdot & 1 & \cdot & \cdot & \cdot \\ 1 & \cdot & \cdot & \cdot & 1 & \cdot & \cdot \\ \cdot & 1 & \cdot & \cdot & \cdot & 1 & \cdot \\ \cdot & 1 & \cdot & \cdot & \cdot & 1 & \cdot \\ \cdot & \cdot & 1 & \cdot & \cdot & \cdot & 1 \\ \cdot & \cdot & 1 & \cdot & \cdot & \cdot & 1 \\ \cdot & \cdot & 1 & \cdot & \cdot & \cdot & 1 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ st_{11} \\ st_{12} \\ st_{22} \\ st_{31} \end{bmatrix} + e.$$

Comments

(i) Easiest to have fixed effects (e.g., treatments) indexed by i.

(ii) To develop X and Z, first write down the vectors of fixed effects, β, and random effects, u.

For $i = 1, 2, 3$ and $n_i = 2 \; \forall \; i$

$$\mathrm{X}\beta = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ t_1 \\ t_2 \\ t_3 \end{bmatrix} \qquad \mathrm{X}_*\beta_* = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}$$

$$E(y) = \mathrm{X}\beta = \mathrm{X}_*\beta_* \quad \text{if} \quad \alpha_i = \mu + \mathrm{t}_i.$$

1.6 Example of $\mathbf{ZGZ' = Z_*G_*Z'_*}$ [7]

Three unrelated cows with 3, 2 and 1 records:

$$\begin{aligned} \mathrm{y}_{ij} &= \mu + \mathrm{c}_i + \mathrm{e}_{ij} \quad i = 1,2,3, \quad n_1 = 3, \quad n_2 = 2, \quad n_3 = 1 \\ \operatorname{cov}\left(y_{ij},\, y_{ij'}\right) &= \sigma_c^2 \;\text{ and }\; \sigma_c^2/\sigma_y^2 = r, \text{ repeatability} \\ \sigma_y^2 &= \sigma_c^2 + \sigma_e^2 \;\Rightarrow\; \sigma_e^2 = (1-r)\,\sigma_y^2 \\ \operatorname{var}(y) &= ZGZ' + R \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \sigma_c^2 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} + \sigma_e^2 I_6 \\ &= \sigma_c^2 \begin{bmatrix} J_3 & 0 & 0 \\ 0 & J_2 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \sigma_e^2 I_6 \end{aligned}$$

The third glibly makes the very true statement that "the most important and most difficult aspect" is modelling; but nothing more is said.

1.9 Comments on the Chapter

In the title "Constructing a Linear Model" the all-important word is "Constructing" – and practically nothing is said about this. What *is* shown is how several standard statistical models fit into the characterization $y = X\beta + Zu + e$.

Chapter 2

Linear Unbiased Estimation

In [11, line 1] "linear functions of β, say $k'\beta$" should be "a linear function of elements of β, say $k'\beta$".

For $E(y) = X\beta$,

$$E(a'y) = a'X\beta$$

and iff $a'X\beta = k'\beta$ then $a'y$ is said to be an unbiased estimator of (unbiased for) $k'\beta$.

Note: For k' and y, there are usually many vectors a'. An example of this is the following. Suppose

$$E(y) = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ t_1 \\ t_2 \end{bmatrix}$$

with $k'\beta$ being $t_1 - t_2$. Then

$$E[1\ \ 0\ \ -1\ \ 0\ \ 0] \quad \text{and} \quad E\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & -5 & 2 & 2 \end{bmatrix} y = t_1 - t_2.$$

Generally speaking $a'y$ is an unbiased estimator of $k'\beta$ iff $a'X = k'$. The sufficiency part of this (if $a'X = k'$ then $a'y$ is unbiased for $k'\beta$) is always true. There are safely ignorable situations when the necessity part is not true (see McCulloch and Searle, 1995).

2.1.3 Fourth Method [13, 2.1.3]

For $(X'X)^-$ being a generalized inverse, $k'\beta$ is estimable if $k'(X'X)^-X'X = k'$. This is very practical because it does not involve rank, nor does it require finding an L or a C as in paragraphs 2.1.2 and 2.1.2, respectively.

Proof That $k'(X'X)^-X'X = k' \Rightarrow k'\beta$ estimable.

$$\begin{aligned} k' &= k'(X'X)^-X'X \\ &= a'X \text{ for } a' = k'(X'X)^-X'. \end{aligned}$$

Hence

$$k'\beta = a'X\beta = E(a'y).$$

2.2 When is $k'\beta$ estimable?

$k'\beta$ is always estimable for $k' = t'X$ for any t. This is the same algebraic relationship as $E(a'y) = k'\beta$ but reworded in a manner that has a different emphasis; namely, for any t'. Whatever, using $k' = t'X$ always makes $k'\beta$ estimable; i.e., $t'X\beta$ is always estimable. This is a very useful fact because it means that whenever the concern is to estimate β, we can avoid considerations of estimability simply by concentrating on $t'X\beta$ – and by doing this for whatever values of t' we desire. In particular, by letting t' be the rows of I we have every element of $X\beta$ as being etimable, a situation which is often described as $X\beta$ being estimable.

Chapter 3

Best Linear Unbiased Estimation (BLUE)

3.1 Introduction

If $a'y$ is to estimate $k'\beta$ unbiasedly, we want $a'X = k'$; and since "best" means minimum variance among unbiased estimators, we want to minimize $a'Va$ subject to $a'X = k'$. Then we set out to minimize

$$\varphi = a'Va + 2\theta\,(k - X'a)$$

where 2θ is a vector of Lagrange multipliers. 2θ, not just θ is used, with benefit of hindsight, to simplify arithmetic.

$$\partial\varphi/\partial a = 0 \quad \Rightarrow \quad 2Va + 2X\theta = 0. \tag{1}$$

$$\partial\varphi/\partial\theta = 0 \quad \Rightarrow \quad X'a \quad = k. \tag{2}$$

These two equations constitute (3.1). From (1) get $a = -V^{-1}X\theta$. Using this in (2) gives

$$-X'V^{-1}X\theta = k \Rightarrow \theta = -(X'V^{-1}X)^{-}k.$$

Hence

$$a = V^{-1}X(X'V^{-1}X)^{-}k.$$

Proof: of $VV_* = I$.

$$\begin{aligned}
VV_* &= (ZGZ' + R)V_* \\
&= RR^{-1} + ZGZ'R^{-1} - (ZGZ' + R)R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1} \\
&= I + ZGZ'R^{-1} - ZG(Z'R^{-1}Z + G^{-1})(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1} \\
&= I \quad \Longrightarrow \quad V^{-1} = V_*.
\end{aligned}$$

Solving (3.4)

The second equation in (3.4) yields

$$\hat{u} = (Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1}(y - X\beta^0).$$

Using this in the first equation of (3.4) gives

$$X'R^{-1}X\beta^0 + X'R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1}(y - X\beta) = X'R^{-1}y$$

which is, from $V^{-1} = V_*$,

$$X'R^{-1}X\beta^0 + X'(R^{-1} - V^{-1})(y - X\beta^0) = X'R^{-1}y;$$

and this reduces, as on [17], to

$$\begin{aligned}
X'V^{-1}X\beta^0 &= X'V^{-1}y \\
\beta^0 &= (X'V^{-1}X)^{-}X'V^{-1}y.
\end{aligned}$$

3.3 Variance of BLUE [18, 3.2]

Taking $K'\beta$ estimable $\Rightarrow K' = T'X$ for some T.

$$\begin{aligned}
\text{BLUE}(K'\beta) &= K'\beta^0 \\
&= T'X(X'V^{-1}X)^{-1}X'V^{-1}y \\
\text{var}[\text{BLUE}(K'\beta)] &= T'X(X'V^{-1}X)^{-}X'V^{-1}VV^{-1}X[(X'V^{-1}X)^{-}]'X'T.
\end{aligned}$$

$$T_1 = \begin{bmatrix} A^- & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -A^-B \\ I \end{bmatrix} (D - CA^-B)^-[-CA^- \quad I] \tag{5a}$$

$$T_2 = \begin{bmatrix} 0 & 0 \\ 0 & D^- \end{bmatrix} = \begin{bmatrix} I \\ -D^-C \end{bmatrix} (A - BD^-C)^-[I \quad -BD^-]. \tag{5b}$$

These apply only when Q has rank equal to the sum of the ranks of A and of $D - CA^-B$ (for T_1); or equal to the sum of ranks of D and of $A - BD^-C$ (for T_2) – see MAUFS, Section 10.5. This rank condition is met when Q is symmetric.

Using T_2 on

$$\begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} = \begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix}^-$$

gives

$$\begin{aligned} C_{11} &= [X'R^{-1}X - X'R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1}X]^- \\ &= [X'\{R^{-1} - R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1}\}X]^- \\ &= (X'V^{-1}X)^-. \end{aligned}$$

Hence

$$\operatorname{var}(K'\beta^0) = K'C_{11}K. \tag{3.6}$$

The rank condition is satisfied because $r(Q) = r(X) + q$ where q is the number of random effects, and $r(D) + r(A - BD^{-1}C) = r(X) +$ order of $G^{-1} = r(X) + q = r(Q)$.

3.5 Generalized Inverses and Mixed Model Equations [19, 3.3]

$AA^-A = A \Rightarrow Ap = z$ has solution $p = A^-z$. Equations $Ap = z$ must be consistent. A more general solution is $\tilde{p} = A^-z + (I - A^-A)t$ for any t.

$$S_1 = [(Z'R^{-1}Z + G^{-1}) - Z'R^{-1}X_1(X_1'R^{-1}X_1)^{-1}X_1'R^{-1}Z]^{-1}. \tag{6}$$

Also,

$$\begin{bmatrix} C_{00} & C_{02} \\ C_{02}' & C_{22} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & (Z'R^{-1}Z + G^{-1}) \end{bmatrix} + \begin{bmatrix} I \\ -(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1}X_1 \end{bmatrix}$$
$$\times S_2[I \quad -X_1'R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}]$$

where using T_2 from (5b) gives

$$\begin{aligned} S_2 &= [X_1'R^{-1}X_1 - X_1'R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1}X_1]^{-1} \qquad (7) \\ &= [X_1'\{R^{-1} - R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1}\}X_1]^{-1} \\ &= (X_1'V^{-1}X_1)^{-1}. \end{aligned}$$

It is a standard result that these two expressions for inverting a paritioned matrix are equal. To demonstrate but one term we show that $S_2 = C_{00}$ for

$$C_{00} = (X_1'R^{-1}X_1)^{-1} + (X_1'R^{-1}X_1)^{-1}X_1'R^{-1}Z\ S_1\ Z'R^{-1}X_1(X_1'R^{-1}X_1)^{-1} \tag{8}$$

To show this, use (7) for S_2^{-1} to get

$$\begin{aligned} &S_2^{-1}(X_1'R^{-1}X_1)^{-1}X'R^{-1}Z \\ &= [X_1'R^{-1}X_1 - X_1'R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1}X_1](X_1'R^{-1}X_1)^{-1}X_1'R^{-1}Z \\ &= X_1'R^{-1}Z[I - (Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1}X_1(X_1'R^{-1}X_1)^{-1}X_1'R^{-1}Z] \\ &= X_1'R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}[Z'R^{-1}Z + G^{-1} - Z'R^{-1}X_1(X_1'R^{-1}X_1)^{-1}X_1'R^{-1}Z] \\ &= X_1'R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}S_1^{-1}, \quad \text{from (6).} \qquad (8a) \end{aligned}$$

Therefore (8) pre-multiplied by S_2^{-1} from (7), followed by using (8a), is

$$\begin{aligned} &S_2^{-1}(X_1'R^{-1}X_1)^{-1} + S_2^{-1}(X_1'R^{-1}X_1)^{-1}X_1'R^{-1}Z\ S_1\ Z'R^{-1}X_1(X_1'R^{-1}X_1)^{-1} \\ &= S_2^{-1}(X_1'R^{-1}X_1)^{-1} + X_1'R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1}X_1(X_1'R^{-1}X_1)^{-1} \\ &= [S_2^{-1} + X_1'R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1}X_1](X_1'R^{-1}X_1)^{-1} \\ &= X_1'R^{-1}X_1(X_1'R^{-1}X_1)^{-1} = I \end{aligned}$$

and so $C_{00} = S_2$. More easily, using regular rather than generalized inverses in T_1 and T_2, we show that

$$A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} = (A - BD^{-1}C)^{-1},$$

Thus arises the need for (9) — provided it exists. This is established using $Y = LX$, for L non-singular.

Notation Because V is positive definite there is a non-singular L such that $V^{-1} = L'L$; and on defining $Y = LX$ we could write

$$X'V^{-1}X = Y'Y, \tag{12}$$

with $Y_{n \times p}$ of rank r with $r < p < n$, as with X. And likewise we could write $X'V^{-1}y = Y'z$ for $z = Ly$. But to avoid this additional notation we use just $X'X$ to represent $X'X$ or $X'V^{-1}X$, whichever is appropriate, and likewise $X'y$ for $X'y$ or $X'V^{-1}y$.

3.5.2.1 Properties of M′

$M'\beta$ not estimable $\Rightarrow$ rows of M' are LIN of rows of Y.

M' shall have full rank, so that no elements of $M'\beta$ are linear combinations of others: rows of M' are LIN.

M' shall have maximum full row rank, $p - r$.

Theorem:

$$\text{The matrix } T = \begin{bmatrix} X'X & M \\ M' & 0 \end{bmatrix} \text{ is non-singular.} \tag{13}$$

Proof: In $[X'X \;\; M]$ the $p-r$ LIN columns of M are LIN of the r LIN columns of X' and hence of $X'X$. Therefore, $[X'X \;\; M]$ has $p - r + r = p$ LIN columns. Moreover, its p LIN rows are LIN of the $p - r$ LIN rows of $[M' \;\; 0]$. Therefore, the matrix T has $p + p - r = 2p - r$ LIN rows. But $X'X$ has p rows and M' has $p - r$ rows. Hence its rank is $2p - r$; and so T is non-singular. Q.E.D.

To establish that C_{11} in (9) is a generalized inverse of $X'V^{-1}X$, that is, of $X'X$ in (9), we first establish some properties connecting M' and X.

3.5.2.3 C_{11} as a generalized inverse

With $X'V^{-1}X$ denoted by $X'X$ as explained earlier, (9) gives

$$\begin{bmatrix} X'X & M \\ M' & 0 \end{bmatrix} \begin{bmatrix} C_{11} & C_{12} \\ C_{12}' & C_{22} \end{bmatrix} = I = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}. \tag{18}$$

Therefore

$$X'X\,C_{11} + MC_{12}' = I \qquad \text{and} \qquad X'X\,C_{12} + MC_{22} = 0, \tag{19}$$

$$M'C_{11} = 0 \qquad \text{and} \qquad M'C_{12} = I. \tag{20}$$

Pre-multiply (19) by D' and use $D'X' = 0$ from (17) to get

$$D'M\,C_{12}' = D' \qquad \text{and} \qquad D'M\,C_{22} = 0.$$

But with $D'M$ being non-singular (because $M'D$ is)

$$C_{12}' = (D'M)^{-1}D' \qquad \text{and} \qquad C_{22} = 0. \tag{21}$$

Then, because from (19)

$$X'XC_{11} = I - MC_{12}' \tag{22}$$

we have

$$\begin{aligned} X'X\,C_{11}\,X'X &= X'X - MC_{12}'X'X \\ &= X'X - M(D'M)^{-1}D'X'X, \text{ from (21)} \\ &= X'X \text{ because, from (17), } D'X' = 0. \end{aligned}$$

Thus C_{11} is a generalized inverse of $X'X$. Moreover, it is a symmetric reflexive generalized inverse: symmetric because the matrix (being inverted) on the left-hand side of (9) is, and reflexive because, from (22),

$$C_{11}'\,X'X = I - C_{12}M',$$

and so

$$C_{11}'X'XC_{11} = C_{11} - C_{12}M'C_{11} = C_{11}, \tag{23}$$

since $M'C_{11} = 0$ from (20).

And similarly

$$Z'XC_{11} + AC'_{12} = 0, \quad Z'XC_{12} + AC_{22} = I, \quad \text{and} \quad Z'XC_{13} + AC'_{23} = 0. \tag{29}$$

Also, as in (20)

$$M'C_{11} = 0, \qquad M'C_{12} = 0, \qquad \text{and now} \qquad M'C_{13} = I. \tag{30}$$

Then, just as in deriving (21), pre-multiply each equation in (28) by D (which is symmetric because it is a covariance matrix) and use $XD = 0 = (DX')'$ to get

$$DMC'_{13} = D \qquad DMC'_{23} = 0 \qquad DMC_{33} = 0. \tag{31}$$

But $DM = D'M = (MD)'$ is non-singular and so

$$C'_{13} = (DM)^{-1}D \qquad C_{23} = 0 \qquad \text{and} \qquad C_{33} = 0. \tag{32}$$

From the first result in (32) we see that the third equation in (30) is satisfied. And using (32) in (28) gives

$$X'XC_{11} + X'ZC'_{12} + MC'_{13} = I, \text{ unchanged,} \tag{33}$$

$$X'XC_{12} + X'ZC_{22} = 0, \text{ and} \tag{34}$$

$$X'XC_{13} = 0, \tag{35}$$

all with C'_{13} as in (32).

We now show that (35) is true by showing that

$$\begin{bmatrix} X'X & X'Z \\ Z'X & A \end{bmatrix} \begin{bmatrix} C_{11} & C_{12} \\ C'_{12} & C_{22} \end{bmatrix} \begin{bmatrix} X'X & X'Z \\ Z'X & A \end{bmatrix} = \begin{bmatrix} X'X & X'Z \\ Z'X & A \end{bmatrix}. \tag{36}$$

To do so, consider each submatrix in the product on the left-hand side of (36). First, the (1,1) term is

$$\begin{aligned} &(X'XC_{11} + X'ZC'_{12})X'X + (X'XC_{12} + X'ZC_{22})Z'X \\ &= X'X\ C_{11}X'X + X'ZC_{12}X'X, \text{ using (34),} \\ &= (I - X'ZC'_{12} - MC'_{13})X'X + X'ZC'_{12}X'X, \text{ using (33),} \\ &= X'X \text{ using (35).} \end{aligned}$$

$$C_{12} = -C_{11}X'ZA^{-1} \tag{38}$$

$$C_{22} = A^{-1} + A^{-1}Z'XC_{11}X'ZA^{-1} = A^{-1} + C_{12}'C_{11}^{-}C_{12} \tag{39}$$

These expressions can also be obtained from the middle equation of (28) after using $C_{23} = 0$ of (32), and the first two equations of (29):

$$X'XC_{12} + X'ZC_{22} = 0 \tag{40}$$

$$Z'XC_{11} + AC_{12}' = 0 \tag{41}$$

$$Z'XC_{12} + AC_{22} = I. \tag{42}$$

Then from (42)

$$C_{22} = A^{-1} - A^{-1}Z'XC_{12} \tag{43}$$

so that

$$X'XC_{12} + X'Z(A^{-1} - A^{-1}Z'XC_{12}) = 0,$$

giving

$$C_{12} = -(X'X - X'ZA^{-1}Z'X)^{-}X'ZA^{-1} = -C_{11}X'ZA^{-1}. \tag{44}$$

Hence from (43)

$$C_{22} = A^{-1} + A^{-1}Z'X(X'X - X'ZA^{-1}\,Z'X)^{-}X'ZA^{-1}. \tag{45}$$

Then in (41)

$$Z'XC_{11} - AA^{-1}Z'X(X'X - X'ZA^{-1}Z'X)^{-} = 0$$

$$Z'X[C_{11} - (X'X - X'ZA^{-1}Z'X)^{-}] = 0. \tag{46}$$

It is easily seen that (46) is satisfied by taking $C_{11} = (X'X - X'Z\ A^{-1}\ Z'X)^{-}$, which is (37), whereupon (44) is (38), and (45) is (39). Thus solutions to (40)–(42) are (37)–(39).

3.5.2.6 Example (not from CRH)

Consider the model equation

$$y_{ijk} = \mu + \alpha_1 + \gamma_j + \epsilon_{ijk} \tag{47}$$

$$C_{12} = -\begin{bmatrix} 0 & 0 \\ 0 & 45/38 \end{bmatrix}\begin{bmatrix} 2 & 4 \\ 2 & 4 \end{bmatrix}\begin{bmatrix} 1/2.5 & 0 \\ 0 & 1/4.5 \end{bmatrix} = -\begin{bmatrix} 0 & 0 \\ \frac{45(2)}{38(2.5)} & \frac{45(4)}{38(4.5)} \end{bmatrix} = \frac{1}{9.5}\begin{bmatrix} 0 & 0 \\ -0 & -10 \end{bmatrix};$$

$$C_{22} = \begin{bmatrix} \frac{1}{2.5} & 0 \\ 0 & \frac{1}{4.5} \end{bmatrix} + \frac{1}{9.5}\begin{bmatrix} 0 & 9 \\ 0 & 10 \end{bmatrix}\frac{38}{45}\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\frac{1}{9.5}\begin{bmatrix} 0 & 0 \\ 9 & 10 \end{bmatrix} = \frac{1}{9.5}\begin{bmatrix} 11 & 8 \\ 8 & 11 \end{bmatrix},$$

after some simple arithmetic. And these results are evident in (48).

3.5.3 A third type of g-inverse [22, 3.3.3]

Because $M'C_{11} = 0$ we can add $MM'C_{11}$ to the first equation in (19) to have, after also using (21) for C'_{12},

$$(X'X + MM')\,C_{11} + M(D'M)^{-1}D' = I.$$

This is also

$$(X'X + MM')\,C_{11} + M(M'D)(M'D)^{-1}(D'M)^{-1}D' = I$$

or

$$(X'X + MM')\,C_{11} + MM'D(D'MM'D)^{-1}D' = I,$$

i.e.,

$$(X'X + MM')[C_{11} + D(D'MM'D)^{-1}D'] = I, \tag{50}$$

because $XD = 0$. But $\begin{bmatrix} X \\ M' \end{bmatrix}$ has full column rank. Therefore

$$[X' \;\; M]\begin{bmatrix} X \\ M' \end{bmatrix} = (X'X + MM') \text{ is non-singular.}$$

Therefore in (50)

$$C_{11} = (X'X + MM')^{-1} - D(D'MM'D)^{-1}D'. \tag{51}$$

Thus we have a formula for calculating C_{11} from M' and X (the latter leading to D).

Then, using the first matrix on [24] as K', namely

$$K' = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix},$$

doing the arithmetic yields

$$K'K = 2I_3 \qquad \text{and} \qquad K(K'K)^{-1} = \frac{1}{2}K;$$

and

$$K'X'V^{-1}X = \begin{bmatrix} 22 & 10 & 12 & 6 & 16 \\ -1 & 5 & -6 & 1 & -2 \\ -5 & -1 & -4 & 3 & -8 \end{bmatrix} \qquad \text{and} \qquad K'X'V^{-1}y = \begin{bmatrix} 24 \\ 2 \\ 4 \end{bmatrix}.$$

Then $(K'K)^{-1}K'X'V^{-1}XK(K'K)^{-}\hat{\alpha} = (K'K)^{-1}K'X'V^{-1}y$ which is [21, (3.12)] is

$$\begin{bmatrix} 11 & -.5 & -2.5 \\ -.5 & 2.75 & .75 \\ -2.5 & .75 & 2.75 \end{bmatrix} \hat{\alpha} = \begin{bmatrix} 12 \\ 1 \\ 2 \end{bmatrix}$$

as in [24].

Chapter 4

Hypotheses Concerning β

4.1 Introduction

Hypotheses are described as follows.

$$\left.\begin{array}{lll} \text{Null} & : & H_0'\beta = c_0 \quad r(H_0') = m \\ \text{Alternative} & : & H_a'\beta = c_a \quad r(H_a) = a \end{array}\right\} \text{Full row rank } r(X) > m > a.$$

Note that $H_*' = Pc_0$ can be considered a hypothesis only if H_*' has full row rank, but also only if the equations $H_*'\beta = c_0$ are consistent; which they will be, of course, if H_*' has full row rank.

The last three sentences of [25] are confusing:

(i) "$\cdots$ the null hypothesis must be contained in the alternative hypothesis." What *does* this mean?

(ii) "$\cdots$ if the null is true the alternative must be true." This seems to be quite wrong. If it were correct and the null were true, then why have the alternative if it was going to be true too?

(iii) "$\cdots$ so, we require" $H_a' = MH_0'$ and $c_a = Mc_0$. This makes no sense to me.

The first of these sums of squares, $R(a|\mu)$, is described as testing

$$\begin{bmatrix} 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \mu \\ a_1 \\ a_2 \\ a_3 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = 0. \tag{1}$$

This is wrong: it has 5 degrees of freedom and $R(a|\mu)$ has only 2 degrees of freedom because there are three rows (factor A).

The correct hypothesis is just the first two terms in (1), namely

$$H : \begin{matrix} a_1 = a_3 \\ a_2 = a_3 \end{matrix}, \quad \text{i.e., } H : a_1 = a_2 = a_3. \tag{2}$$

Why describe the alternative hypothesis as

$$H : \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix} \beta = 0? \tag{3}$$

This is

$$H : b_1 = b_2 = b_3 = b_4,$$

and as such is no alternative to $H : a_1 = a_2 = a_3$. And in terms of the last three sentences on [25] the hypothesis of (2), taken as a null hypothesis, certainly cannot be described as "contained in" (3) thought of as an alternative hypothesis.

Note that nothing on [26] is said about how many observations there are in each i, j cell. For $k = 1, 2, \cdots, n_{ij}$ it is only when $n_{ij} = n \; \forall \; i, j$ (i.e., only for balanced data) that (2) is the hypothesis for $R(a|\mu)$. In contrast, (3) is the hypothesis for $R(b|\mu, a)$ for both balanced and unbalanced data – so long as the no-interaction model is used.

solutions to GLS equations "subject to restrictions $H_0'\beta_0 = c_0$." And here is the second confusion: $H_0'\beta_0 = c_0$ starts off as being called a hypothesis and then gets called a restriction.

It seems easier to retain H as a symbol for labelling a hypothesis and to write a hypothesis as

$$H\colon K'\beta = c,$$

using subscripts to H, K and c (but not β) when more than one hypothesis is being considered; e.g.,

$$H_0\colon K_0'\beta = c_0 \qquad \text{and} \qquad H_a\colon K_a'\beta = c_a.$$

Then one can still use β^0 to represent solutions to equations; in particular (with known V)

$$\beta^0 = (X'V^{-1}X)^-X'V^{-1}y \tag{11}$$

when no hypothesis is being considered, and β_0^0 and β_a^0 are solutions under hypotheses H_0 and H_a, respectively.

4.4.2 The general case

For the general hypothesis $H : K'\beta = c$ we calculate β_H^0 as that value of β which minimizes $(y - X\beta)'V^{-1}(y - X\beta)$ subject to $H : K'\beta = c$, i.e., which minimizes

$$(y - X\beta)'V^{-1}(y - X\beta) + 2\theta'(K'\beta - c). \tag{12}$$

This, as may easily be shown, leads to equations

$$\begin{bmatrix} X'V^{-1}X & K \\ K' & 0 \end{bmatrix} \begin{bmatrix} \beta_H^0 \\ \theta_H \end{bmatrix} = \begin{bmatrix} X'V^{-1}y \\ c \end{bmatrix}. \tag{13}$$

These are (4.4) with K in place of H_0, and β_H^0 in place of β_0. These notation changes help clarify the procedures. β always represents unknown parameters, except in (12) where β is viewed as a mathematical variable for which one chooses as β_H^0 that value of β which minimizes (12). Thus β_H^0 is the solution of the GLS equations under the hypothesis $H : K'\beta = c$, and it is different from β^0 of (11) which applies when there is no hypothesis.

4.4.4 With the hypothesis $H\colon K'\beta = c$

Under $H\colon K'\beta = c$, we have $K' = T'X$ for some T' and hence $K'(X'V^{-1}X)^{-}X'V^{-1}X = K'$, a result that is used repeatedly; and then the residual sum of squares is

$$\begin{aligned}\mathrm{SSE}_H &= (y - X\beta_H^0)'V^{-1}(y - X\beta_H^0)\\ &= y'V^{-1}y - 2\beta_H^{0\prime}X'V^{-1}y + \beta_H^{0\prime}X'V^{-1}X\beta_H^0 \qquad (18)\\ &= y'V^{-1}y - 2(\beta^0 - \gamma^0)'X'V^{-1}y + (\beta^0 - \gamma^0)'X'V^{-1}X(\beta^0 - \gamma^0) \qquad (19)\end{aligned}$$

after writing $\beta_H^0 = \beta^0 - \gamma^0$ from (16) with

$$\gamma^0 = (X'V^{-1}X)^{-}K\theta_H. \qquad (20)$$

Thus, from (19)

$$\begin{aligned}\mathrm{SSE}_H &= y'V^{-1}y - 2\beta^{0\prime}X'V^{-1}y + 2\gamma^{0\prime}X'V^{-1}y\\ &\quad + \beta^{0\prime}X'V^{-1}X\beta^0 + \gamma^{0\prime}X'V^{-1}X\gamma^0 - 2\gamma_0'X'V^{-1}X\beta^0\\ &= \mathrm{SSE} + \gamma^{0\prime}X'V^{-1}X\gamma^0, \quad \text{using} \quad X'V^{-1}X\beta^0 = X'V^{-1}y\\ &= \mathrm{SSE} + \theta_H'K'(X'V^{-1}X)^{-}X'V^{-1}X(X'V^{-1}X)^{-}K\theta_H\\ &= \mathrm{SSE} + \theta_H'K'(X'V^{-1}X)^{-}K\theta_H.\end{aligned}$$

Then, on using θ_H of (15)

$$\mathrm{SSE}_H = \mathrm{SSE} + (K'\beta^0 - c)'[K'(X'V^{-1}X)^{-}K]^{-1}(K'\beta^0 - c). \qquad (21)$$

Thus

$$\mathrm{SSE}_H - \mathrm{SSE} = (K'\beta^0 - c)'[K'(X'V^{-1}X)^{-}K]^{-1}(K'\beta^0 - c),$$

i.e.,

$$\begin{aligned}\mathrm{SS(H)} &= \mathrm{SSE}_H - \mathrm{SSE}, \qquad (22)\\ &= (K'\beta^0 - c)'[K'(X'V^{-1}X)^{-}K]^{-1}(K'\beta^0 - c),\end{aligned}$$

akin to (9).

and $V = R = 5I_9$. For testing $H : t_1 = t_2 = t_3$ the calculations on [28] are

$$\begin{aligned}\text{SSE} &= (y - X\beta^0)'V^{-1}(y - X\beta^0), \quad \text{written as} \quad (y - X\beta_0)'V^{-1}(y - X\beta_0)\\ &= 9/4\end{aligned}$$

and

$$\begin{aligned}\text{SSE}_H &= (y - X\beta^0_H)'V^{-1}(y - X\beta^0_H), \quad \text{written as} \quad (y - X\beta_a)'V^{-1}(y - X\beta_a)\\ &= 146/45.\end{aligned}$$

and hence the sum of squares due to H is, using (22)

$$\text{SS(H)} = \text{SSE}_H - \text{ SSE } = \frac{146}{45} - \frac{9}{4} = \frac{179}{180}. \tag{26}$$

From the calculations in [28] we also get

$$\begin{aligned}\text{SSR}_H = \beta^{0'}_H X'V^{-1}y &= [49/9 \;\; 0 \;\; 0 \;\; 0](.2)[49 \;\; 25 \;\; 15 \;\; 9]' = 49^2/45 = 53\,\frac{16}{45} \text{ and}\\ \text{SSR} = \beta^{0'}X'V^{-1}y &= [0 \;\; 25/4 \;\; 15/3 \;\; 9/2](.2)[49 \;\; 25 \;\; 15 \;\; 9]' = 125/4 + 15 + 8.1 + 54\,\frac{7}{20}.\end{aligned}$$

Thus from (25)

$$\text{SS(H)} \; = 54\,\frac{7}{20} - 53\,\frac{16}{45} = 1\,\frac{63-64}{180} = \frac{179}{180}, \tag{27}$$

as in (26).

4.5.3 Analysis of variance calculations

An alternative procedure is to use analysis of variance arithmetic (when $V = \lambda I$ for some scalar λ). This is done for two models: the full model, which has no hypothesis, and the reduced model which is the full model reduced on incorporating the hypothesis. The model equations and

Same example, but $V \neq \lambda I$

$$V = \left\{_d \; (2I_4 + 2J_4) \quad (2I_3 + 4J_3) \quad \begin{bmatrix} 2 & 5 \\ 5 & 13 \end{bmatrix} \right\}$$

$$V^{-1} = \left\{_d \; \left(\tfrac{1}{2}I_4 - \tfrac{1}{10}J_4\right) \quad \left(\tfrac{1}{2}I_3 - \tfrac{1}{7}J_3\right) \quad \begin{bmatrix} 13 & -5 \\ -5 & 2 \end{bmatrix} \right\}$$

$$X' = \begin{bmatrix} 1'_4 & 1'_3 & 1'_2 \\ 1'_4 & \cdot & \cdot \\ \cdot & 1'_3 & \cdot \\ \cdot & \cdot & 1'_2 \end{bmatrix}$$

$$X'V^{-1}X = \begin{bmatrix} \frac{1}{10}1'_4 & \frac{1}{14}1'_3 & 8 & -3 \\ \frac{1}{10}1'_4 & \cdot & \cdot & \cdot \\ \cdot & \frac{1}{14}1'_3 & \cdot & \cdot \\ \cdot & \cdot & 8 & -3 \end{bmatrix} X = \begin{bmatrix} \frac{4}{10} + \frac{3}{14} + 5 & \frac{4}{10} & \frac{3}{14} & 5 \\ \frac{4}{10} & \frac{4}{10} & \cdot & \cdot \\ \frac{3}{14} & \cdot & \frac{3}{14} & \cdot \\ 5 & \cdot & \cdot & 5 \end{bmatrix}$$

$$(X'V^{-1}X)^- = \left\{_d \; 0 \;\; 2\tfrac{1}{2} \;\; 4\tfrac{2}{3} \;\; \tfrac{1}{5} \right\} \quad \text{and} \quad \mathbf{y} = [6 \;\; 7 \;\; 8 \;\; 4 \;\; 5 \;\; 6 \;\; 5 \;\; 4]$$

$$X'V^{-1}y = \begin{bmatrix} \frac{25}{10} + \frac{15}{14} + 40 - 12 \\ \frac{25}{10} \\ \frac{15}{14} \\ 28 \end{bmatrix}$$

$$\beta^0 = \begin{bmatrix} \mu^0 \\ t_1^0 \\ t_2^0 \\ t_3^0 \end{bmatrix} = (X'V^{-1}X)^- X'V^{-1}y = \begin{bmatrix} 0 \\ 6\frac{1}{4} \\ 5 \\ 5\frac{3}{5} \end{bmatrix}$$

$$\text{SSE} = (y - X\beta^0)'(y - X\beta^0)$$

4.5.4 A warning on reductions in sums of squares

Equation (25) is a case where SS(H) can be calculated as the difference between two reductions in sums of squares. But, as in [LMFUD, Sec. 8.8e], this difference cannot always be used. In fact, whereas the difference between residual sums of squares, $\text{SSE}_H - \text{SSE}$, can always be used, the difference between reductions in sums of squares, $\text{SSR} - \text{SSR}_H$, can be used only when c is null, i.e., $c = 0$. We illustrate for $H: t_1 - t_2 = 4$ in the preceding example. For then, under the hypothesis,

$$y = X\begin{bmatrix}\mu\\ t_1\\ t_2\\ t_3\end{bmatrix} + e \qquad \text{becomes} \qquad y = X\begin{bmatrix}\mu\\ t_1\\ t_1-4\\ t_3\end{bmatrix} + e.$$

This leads to adding 4 to each y_{2j}-value so that y becomes

$$y^* = \begin{bmatrix}6\\7\\8\\4\\4+4\\5+4\\6+4\\5\\4\end{bmatrix} = \begin{bmatrix}6\\7\\8\\4\\8\\9\\10\\5\\4\end{bmatrix} = \begin{bmatrix}1&1&\cdot\\1&1&\cdot\\1&1&\cdot\\1&1&\cdot\\1&1&\cdot\\1&1&\cdot\\1&1&\cdot\\1&\cdot&1\\1&\cdot&1\end{bmatrix}\begin{bmatrix}\mu\\ t_1\\ t_3\end{bmatrix} + e = X_*\beta_* + e.$$

Thus $y_{*1\cdot} = 52$ (with 7 data) and $y_{*3\cdot} = 9$ with 2 data. Hence, using analysis of variance calculations

$$\text{R(model)} = \Sigma n_{*i\cdot}\bar{y}^2_{*i\cdot} = 52^2/7 + 9^2/2 = 426\,\frac{11}{14}.$$

Then, because y has become y_*, the value of SST changes from $\Sigma y^2_{ij} = 283$ to $\Sigma y^2_{*ij} = 451$. Therefore

$$\text{SSE}_H = 451 - \text{R(model)} = 451 - 426\,\frac{11}{14} = 24\,\frac{3}{14}.$$

And only K' and c depend on the hypothesis being tested.

In the example,

$$X = \begin{bmatrix} 1 & 1 & \cdot & \cdot \\ 1 & 1 & \cdot & \cdot \\ 1 & 1 & \cdot & \cdot \\ 1 & 1 & \cdot & \cdot \\ 1 & \cdot & 1 & \cdot \\ 1 & \cdot & 1 & \cdot \\ 1 & \cdot & 1 & \cdot \\ 1 & \cdot & 1 & \cdot \\ 1 & \cdot & \cdot & 1 \\ 1 & \cdot & \cdot & 1 \end{bmatrix} \qquad X'X = \begin{bmatrix} 9 & 4 & 3 & 2 \\ 4 & 4 & \cdot & \cdot \\ 3 & \cdot & 3 & \cdot \\ 2 & \cdot & \cdot & 2 \end{bmatrix}, \qquad (X'X)^{-} = \begin{bmatrix} 0 & & & \\ & \frac{1}{4} & & \\ & & \frac{1}{3} & \\ & & & \frac{1}{2} \end{bmatrix}.$$

And for $H : t_1 = t_2 = t_3$,

$$K'\beta = c \quad \text{is} \quad \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} \mu \\ t_1 \\ t_2 \\ t_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Then the normal equations $X'V^{-1}X\beta^0 = X'V^{-1}y$, namely

$$\begin{bmatrix} 9 & 4 & 3 & 2 \\ 4 & 4 & \cdot & \cdot \\ 3 & \cdot & 3 & \cdot \\ 2 & \cdot & \cdot & 2 \end{bmatrix} \begin{bmatrix} \mu^0 \\ t_1^0 \\ t_2^0 \\ t_3^0 \end{bmatrix} = \begin{bmatrix} 49 \\ 25 \\ 15 \\ 9 \end{bmatrix}, \quad \text{give} \quad \beta^0 = \begin{bmatrix} \mu^0 \\ t_1^0 \\ t_2^0 \\ t_3^0 \end{bmatrix} = \begin{bmatrix} 0 \\ 25/4 \\ 15/3 \\ 9/2 \end{bmatrix}. \tag{28}$$

Thus

$$K'\beta^0 = \begin{bmatrix} 25/4 - 15/3 \\ 25/4 - 9/2 \end{bmatrix} = \begin{bmatrix} 5/4 \\ 7/4 \end{bmatrix},$$

We calculate this as SS(H) with

$$(X'X)^{-} = \text{diag}\left\{0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ \frac{1}{3}\ \ \frac{1}{2}\ \ 1\ \ 1\ \ \frac{1}{2}\ \ \frac{1}{5}\right\}, \tag{30}$$

$$\beta^0 = [0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 2\ \ 1\ \ 2\ \ 3\ \ 2.5\ \ 1.8]'. \tag{31}$$

Then writing the hypothesis as

$$H\colon \begin{bmatrix} \mu_{11} - \mu_{12} - \mu_{21} + \mu_{22} \\ \mu_{11} - \mu_{13} - \mu_{21} + \mu_{23} \end{bmatrix} = 0 \text{ being } K'\beta = 0$$

gives

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

and so

$$K'\beta^0 = \begin{bmatrix} 2 - 1 - 3 + 2.5 \\ 2 - 2 - 3 + 1.8 \end{bmatrix} = \begin{bmatrix} .5 \\ -1.2 \end{bmatrix}$$

and

$$\begin{aligned} \left[K'(X'X)^{-}K\right]^{-1} &= \begin{bmatrix} \frac{1}{3} + \frac{1}{2} + 1 + \frac{1}{2} & \frac{1}{3} + 1 \\ \frac{1}{3} + 1 & \frac{1}{3} + 1 + 1 + \frac{1}{5} \end{bmatrix}^{-1} = 15 \begin{bmatrix} 35 & 20 \\ 20 & 38 \end{bmatrix}^{-1} \\ &= \frac{1}{62} \begin{bmatrix} 38 & -20 \\ -20 & 35 \end{bmatrix}. \end{aligned}$$

Hence

$$\begin{aligned} \text{SS(H)} = (K'\beta^0)' \left[K'(X'X)^{-}K\right]^{-1} K'\beta^0 &= \frac{1}{62} [.25(38) + 1.44(35) + 2(.5)(-1.2)(-20)] \\ &= \frac{83.9}{62} = 1.3532, \quad \text{as in} \quad (29). \end{aligned} \tag{32}$$

LM p.278

$$\begin{aligned}
R(\mu) &= N\bar{y}^2_{...} &&= 14(27/14)^2 &&= 52.0714 \\
R(\mu, r) &= \Sigma_i n_{i.}\bar{y}^2_{i..} &&= \frac{10^2}{6} + \frac{17^2}{8} &&= 52.7917 \\
R(\mu, c) &= \Sigma_j n_{.j}\bar{y}^2_{.j.} &&= \frac{92}{4} + \frac{72}{4} + \frac{11^2}{6} &&= 52.6667
\end{aligned}$$

LM p. 297

$$\begin{aligned}
R(\mu, r, c) &= \Sigma_j n_{.j}\bar{y}^2_{.j.} + \frac{u_1^2}{t_{11}} \\
&= 52.6667 + \frac{\left\{10 - \left[3\left(\frac{9}{4}\right) + 2\left(\frac{7}{4}\right) + 1\left(\frac{11}{6}\right)\right]\right\}^2}{6 - \left(\frac{3^2}{4} + \frac{2^2}{4} + \frac{1^2}{6}\right)} \\
&= 52.6667 + \frac{(-2.0833)^2}{2.5833} = 52.6667 + 1.6801 = 54.3468
\end{aligned}$$

LM p. 275

$$\begin{aligned}
R(\mu, r, c, rc) &= \Sigma_{ij} n_{ij}\bar{y}^2_{ij.} \\
&= \frac{6^2}{3} + \frac{2^2}{2} + 4 + 9 + \frac{5^2}{2} + \frac{9^2}{5} = 55.7000.
\end{aligned}$$

Then the sum of squares for testing interactions is

$$R(\mu, r, c, rc) - R(\mu, r, c) = 55.7 - 54.3468 = 1.3532$$

as in (32); and that for testing equality of row effects in the absence of interactions is

$$R(r|\mu, c) = R(\mu, r, c) - R(\mu, c) = 54.3468 - 52.6667 = 1.6801$$

as in (33). Similarly, of course, for testing equality of column effects one needs

$$R(c|\mu, r) = R(\mu, r, c) - R(\mu, r) = 54.3468 - 52.7917 = 1.5551.$$

With $\beta^{0\prime} = [0\ 0\ 0\ 0\ 0\ 0\ 2\ 1\ 2\ 3\ 2.5\ 1.8]$, from (31) and writing (34) as $H : K'\beta = 0$, we have

$$K'\beta^0 = \begin{bmatrix} 2+1+2-3-2.5-1.8 \\ 2-2-3+1.8 \\ 1-2-2.5+1.8 \end{bmatrix} = \begin{bmatrix} -2.3 \\ -1.2 \\ -1.7 \end{bmatrix}$$

and with $(X'X)^- = \text{diag}\left\{0\ 0\ 0\ 0\ 0\ 0\ \tfrac{1}{3}\ \tfrac{1}{2}\ 1\ 1\ \tfrac{1}{2}\ \tfrac{1}{5}\right\}$ from (30),

$$K'(X'X)^-K = \begin{bmatrix} 1 & 1 & 1 & -1 & -1 & -1 \\ 1 & 0 & -1 & -1 & 0 & 1 \\ 0 & 1 & -1 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 1 & -1 & -1 \\ -1 & -1 & 0 \\ -\frac{1}{2} & 0 & -\frac{1}{2} \\ -\frac{1}{5} & \frac{1}{5} & \frac{1}{5} \end{bmatrix} = \begin{bmatrix} 3\frac{8}{15} & \frac{2}{15} & -\frac{1}{5} \\ \frac{2}{15} & 2\frac{8}{15} & 1\frac{1}{5} \\ -\frac{1}{5} & 1\frac{1}{5} & 2\frac{1}{5} \end{bmatrix},$$

$$\left[K'(X'X)^-K\right]^{-1} = \left(\frac{1}{15}\begin{bmatrix} 53 & 2 & -3 \\ 2 & 38 & 18 \\ -3 & 18 & 33 \end{bmatrix}\right)^{-1} = \frac{1}{108}\begin{bmatrix} 31 & -4 & 5 \\ -4 & 58 & -32 \\ 5 & -32 & 67 \end{bmatrix},$$

Chapter 5

Prediction of Random Variables

Many of the numerous results in this chapter are stated without derivation, probably because their details are quite lengthy. Also, they pertain more to statistics than to animal breeding. For these notes there are therefore two alternatives: (i) to include all those details, or (ii) to refer the reader to the VC reference wherein Chapter 7 sets out the details in full array. Because (i) would add considerable, solely mathematical, length to these notes and would entail little more than copying from VC, alternative (ii) has been chosen: to give the reader specific references to VC, at the same time emphasizing important concepts as is deemed necessary.

Notation Since in this section confusion between vectors and scalars is all too easy, bold face font is sometimes used for vectors and matrices.

5.1 Best Prediction (BP) [33, 5.1]

Equation (5.1) gives the best predictor $\hat{w} = f(\mathbf{y}) = E(w|\mathbf{y})$ of w, a scalar random variable that is simply the univariate case of the general result for a

$$\text{vector } \mathbf{u}\text{: best predictor } \hat{\mathbf{u}} = \text{BP}(\mathbf{u}) = E(\mathbf{u}|\mathbf{y}). \tag{1}$$

This is VC 261, (3). Its derivation is shown on VC 262, based on minimizing not just $E(\hat{w} - w)^2$ of [33, line 2 of 5.1] but the more general quadratic $E\left\{(\tilde{\mathbf{u}} - \mathbf{u})'\mathbf{A}(\tilde{\mathbf{u}} - \mathbf{u})\right\}$ for some matrix $\mathbf{A}$.

The result is derived on VC 264.

4. Ranking predictors

A fourth result, described on [34], but not included there in the listed properties of BLP(u) is the following. Ranking predictors of $u_1, \cdots, u_N$ from largest to smallest, and selecting the highest α-percentage of those predictors, maximizes $E(u)$ for that α-percentage, if BLP(u) is used as the predictor. VC 264-5 shows a derivation.

5.2 Best linear prediction (BLP) [34, 5.2]

Reconciling the derivation of BLP in [34-5] with that in VC Sec. 7.3 is a little tricky. The end result is the same in both places.

The derivation in [34-5] deals with w, starts by defining $\hat{w} = \mathbf{a}'\mathbf{y} + b$ (linear in y), and derives $\mathbf{a}$ and b by minimizing $E(\hat{w} - w)^2$. After defining

$$E(w) = \gamma, \qquad E(\mathbf{y}) = \boldsymbol{\alpha}, \qquad \text{Cov}(\mathbf{y},\ w) = \mathbf{c} \quad \text{and} \quad \text{var}(\mathbf{y}) = \mathbf{V}$$

this leads to

$$\text{BLP}(w) = \gamma + \mathbf{c}'\mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\alpha}) = E(w) + \text{Cov}(w,\ \mathbf{y}')\mathbf{V}^{-1}[\mathbf{y} - E(\mathbf{y})]. \tag{3}$$

VC Sec. 7.3 uses $\mathbf{u}$, starting with $\tilde{\mathbf{u}} = \mathbf{a} + \mathbf{B}\mathbf{y}$ (note here the use of the letters $\mathbf{a}$ and $\mathbf{B}$ from that of $\mathbf{a}$ and b in $\hat{w}$). Then $\mathbf{a}$ and $\mathbf{B}$ are derived by minimizing $E(\tilde{\mathbf{u}} - \mathbf{u})'\mathbf{A}(\tilde{\mathbf{u}} - \mathbf{u})$ for positive definite $\mathbf{A}$. With definitions

$$E(\mathbf{u}) = \boldsymbol{\mu}_U, \qquad E(y) = \boldsymbol{\mu}_Y, \qquad \text{cov}(\mathbf{u},\ \mathbf{y}') = \mathbf{C} \quad \text{and} \quad \text{var}(\mathbf{y}) = \mathbf{V}$$

this yields

$$\tilde{\mathbf{u}} = \text{BLP}(\mathbf{u}) = \boldsymbol{\mu}_U + \mathbf{C}\mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\mu}_Y). \tag{4}$$

This, which is (23) of VC 268, is simply the vector form of BLP(w) of (3).

Variance-covariance properties of BLP(u) come from (4) very easily. First

$$\text{var}(\tilde{\mathbf{u}}) \quad = \quad \text{var}\left[\mathbf{C}\mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\mu}_Y)\right] = \mathbf{C}\mathbf{V}^{-1}\mathbf{V}\mathbf{V}^{-1}\mathbf{C}' = \mathbf{C}\mathbf{V}^{-1}\mathbf{C}'$$

as in the last line of [35]. Now we earlier derived

$$\mathrm{var}(\tilde{\mathbf{u}}) = \mathrm{cov}(\tilde{\mathbf{u}},\ \mathbf{u}') = CV^{-1}C'.$$

Hence, for w being an element of $\mathbf{u}$, the ratio $\mathrm{var}(\hat{w})/\mathrm{cov}(\hat{w}, w)$ in (5) is unity. Thus (5) gives

$$\mathbf{a} = \mathbf{V}^{-1}\mathbf{c}.$$

Then unbiasedness of $\hat{w} = \mathbf{a}'\mathbf{y} + b$ gives

$$E(w) = E(\hat{w}) = \mathbf{a}'E(\mathbf{y}) + b$$

and so

$$\begin{aligned} \hat{w} = \mathbf{a}'\mathbf{y} + b &= \mathbf{c}'\mathbf{V}^{-1}\mathbf{y} + E(w) - a'E(\mathbf{y}) \\ &= E(w) + \mathrm{cov}(w, \mathbf{y}')V^{-1}[\mathbf{y} - E(\mathbf{y})] \end{aligned}$$

which is BLP(w) of (3). Thus the BLP maximizes the correlation between a random variable, w, and its BLP.

5.3 Ranking

Following (5.11) is a remark about ranking. It relates to a salient problem concerning the use of predicted values. How does the ranking on predicted values compare with the ranking on true (realized but often unobservable) values? Henderson (1963) has shown, under certain conditions (including normality), that the probability of predictors of u_i and u_j having the same pairwise ranking as u_i and u_j is maximized. And Portnoy (1982) extends this to the usual components of variance model for which ranking the u_is in the same order as the $\tilde{u}_i$s rank themselves does maximize the probability of correctly ranking all the u_is. This is, of course, important in animal genetics where predicting genetic merit is vital to the breeding of successive generations.

5.4 BP equals BLP under normality

<u>Notation</u> We revert to the norm for these notes of not using bold face for matrices and vectors.

so that y of (7) has variance

$$V = \text{var}(y) = ZDZ' + R. \tag{9}$$

Define the function we wish to estimate (or predict, whichever word one prefers) as

$$f = t'X\beta + h'u \tag{10}$$

for any $[t' \;\; h'] \neq 0$. To have an estimator of f that is unbiased, linear (in elements of y) and "best" we want the estimator to be

(i) linear in y : $\lambda' y$ for $\lambda' \neq 0$;

(ii) unbiased: $E(\lambda' y - f) = 0$;

(iii) best: we want the error mean square, $E[\lambda' y - (t'X\beta + h'u)]^2$ subject to $E(\lambda' y - f) = 0$ to be a minimum.

In the model equation (7) we take $E(u) = 0$, giving $E(y) = X\beta$ so that (ii) above gives $\lambda' X\beta - t'X\beta = 0$. We want this to be true for all β, and so need $\lambda' X = t'X$, or

$$X'\lambda = X't. \tag{11}$$

Then, subject to (11), we want from (iii) above to minimize

$$\begin{aligned} E(\lambda' y - t'X\beta - h'u)^2 &= E(\lambda' X\beta + \lambda' Zu + \lambda' e - t'X\beta - h'u)^2 \\ &= E\left[\lambda'(Zu + e) - h'u\right]^2 \\ &= \lambda' V\lambda + h'Dh - 2\lambda' ZDh, \end{aligned} \tag{12}$$

after using (8). To do this we minimize

$$\theta = \lambda' V\lambda + h'Dh - 2\lambda' ZDh + 2m'(X'\lambda - X't) \tag{13}$$

where m' is a vector of Lagrange multipliers. Thus

$$\partial\theta/\partial\lambda = 0 \quad \text{gives} \quad 2V\lambda - 2ZDh + 2Xm = 0 \quad \Rightarrow \quad \lambda = -V^{-1}Xm + V^{-1}ZDh, \tag{14}$$

and for this to be true for all k we must have $\lambda' X = 0$. This is equivalent to having $t = 0$ in (11); and using this in (16) gives

$$\lambda' y = h' D Z' V^{-1}(y - X\beta^0).$$

This, from (10) with $t = 0$, is $\text{BLUP}(h'u) = h'DZ'V^{-1}(y - X\beta^0)$, which, as in [38, line 7], is $m'C'V^{-1}(y - X\beta^0)$ with m' being h' and $C' = DZ'$.

5.7 Using functions of y having zero expectation [38, 5.4.2]

For $\beta_* = L'y$

$$E(X\beta_*) = E(XL'y) = XL'E(y) = XL'X\beta$$

and if

$$E(X\beta_*) = XL'X\beta \quad \text{is to be} \quad X\beta \ \forall \ \beta$$

then

$$XL'X = X.$$

Equations (5.19) through (5.23) of [38] are quite straightforward. The line below (5.23) deserves support.

Proof: $\hat{w}$ is invariant to T and to $(T'VT)^-$.

$$\begin{aligned} \hat{w} &= C_*'V_*^- y_*, \quad && \text{from} \quad (5.21) \\ &= (T'C)'(T'VT)^- T'y, \quad && \text{from earlier equations} \\ &= C'T(T'VT)^- T'y. && (19) \end{aligned}$$

This is invariant to T because $XL'X = X$ indicates that L' is a generalized inverse of X, say $(X'X)^- X'$; and then $T' = I - XL'$ is $T' = I - X(X'X)^- X'$, invariant to $(X'X)^-$. Then in (19)

$$T(T'VT)^- T' = V^{-1}VT(T'VT)^- T'VV^{-1} = V^{-1}Q'QT(T'Q'QT)^- T'Q'QV^{-1}$$

for $V = Q'Q$ and non-singular; and $QT(T'Q'QT)^- T'Q'$ is invariant to QT. Thus $\hat{w}$ is invariant to T and $(T'VT)^-$.

$$\begin{aligned}
\mathrm{Cov}(\hat{w}, w') &= (K' - C'V^{-1}X)\mathrm{Cov}(\beta^0, u') + C'V^{-1}\mathrm{Cov}(y, u') \\
&= (K' - C'V^{-1}X)(X'V^{-1}X)^{-}X'V^{-1}ZG + C'V^{-1}ZG \\
&= K'(X'V^{-1}X)^{-}X'V^{-1}C + C'V^{-1}C - C'V^{-1}X(X'V^{-1}X)^{-}X'V^{-1}C \\
&= K'(X'V^{-1}X)^{-}X'V^{-1}C + C'PC \quad \text{for} \quad P = V^{-1} - V^{-1}X(X'V^{-1}X)^{-}X'V^{-1}
\end{aligned}$$

which is (5.26).

$$\begin{aligned}
\mathrm{var}(\hat{w}) &= \mathrm{var}[K'\beta^0 + C'V^{-1}(y - X\beta^0)] \\
\mathrm{var}(\beta^0) &= (X'V^{-1}X)^{-}X'V^{-1}X(X'V^{-1}X)^{-'} = (X'V^{-1}X)^{-} \text{ say} \\
\mathrm{cov}(y,\ \beta^{0'}) &= VV^{-1}X(X'V^{-1}X)^{-} = X(X'V^{-1}X)^{-}
\end{aligned}$$

and on writing $A \equiv (X'V^{-1}X)^{-}$,

$$\begin{aligned}
\mathrm{var}(\hat{w}) &= \mathrm{var}\left[(K' - C'V^{-1}X)\beta^0 + C'V^{-1}y\right] \\
&= (K' - C'V^{-1}X)A(K - X'V^{-1}C) + C'V^{-1}C + C'V^{-1}XA(K - X'V^{-1}C) \\
&\qquad + (K' - C'V^{-1}X)AX'V^{-1}C \\
&= K'(X'V^{-1}X)^{-}K + C'V^{-1}C - C'V^{-1}X(X'V^{-1}X)^{-}X'V^{-1}C.
\end{aligned}$$

This is (5.28). Finally

$$\begin{aligned}
\mathrm{var}(w - \hat{w}) &= \mathrm{var}(w) - \mathrm{cov}(\hat{w}, w') - \mathrm{cov}(w, \hat{w}') + \mathrm{var}(\hat{w}) \\
&= \mathrm{var}(w) - K'(X'V^{-1}X)^{-}X'V^{-1}C - C'PC \\
&\qquad - C'V^{-1}X(X'V^{-1}X)^{-}K - C'PC + K'(X'V^{-1}X)^{-}K + C'PC \\
&= G + K'(X'V^{-1}X)^{-}K - K'(X'V^{-1}X)^{-}X'V^{-1}C - C'V^{-1}X(X'V^{-1}X)^{-}K - C'PC
\end{aligned}$$

which is (5.29).

5.10 Variances from Mixed Model Equations [40, 5.7]

From (21) let

$$\begin{bmatrix} \beta^0 \\ \hat{u} \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} \\ C'_{12} & C_{22} \end{bmatrix} \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}$$

where

$$\begin{aligned} \begin{bmatrix} C_{11} & C_{12} \\ C'_{12} & C_{22} \end{bmatrix} &= \begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z+G^{-1} \end{bmatrix}^{-} \\ &= \begin{bmatrix} 0 & 0 \\ 0 & U^{-} \end{bmatrix} + \begin{bmatrix} I \\ -U^{-}Z'R^{-1}X \end{bmatrix} T^{-} \begin{bmatrix} I & -X'R^{-1}ZU^{-} \end{bmatrix} \end{aligned}$$

where

$$\begin{aligned} U &= Z'R^{-1}Z+G^{-1} \\ T &= X'R^{-1}X - X'R^{-1}Z(Z'R^{-1}Z+G^{-1})^{-1}Z'R^{-1}X \\ &= X'[R^{-1} - R^{-1}Z(Z'R^{-1}Z+G^{-1})^{-1}Z'R^{-1}]X. \\ &= X'V^{-1}X, \end{aligned}$$

using $VV_* = VV^{-1} = I$ of these notes for Section [3.2]. Hence

$$T = X'V^{-1}X$$

and

$$\begin{bmatrix} C_{11} & C_{12} \\ C'_{12} & C_{22} \end{bmatrix} = \begin{bmatrix} T^{-} & -T^{-}X'R^{-1}ZU^{-} \\ -U^{-}Z'R^{-1}XT^{-} & U^{-}+U^{-}Z'R^{-1}X(X'V^{-1}X)^{-}X'R^{-1}ZU^{-} \end{bmatrix}.$$

But from below (22)

$$U^{-1}Z'R^{-1} = (Z'R^{-1}Z+G^{-1})^{-1}Z'R^{-1} = C'V^{-1} = GZ'V^{-1}.$$

Then

$$\begin{bmatrix} C_{11} & C_{12} \\ C'_{12} & C_{22} \end{bmatrix} = \begin{bmatrix} (X'V^{-1}X)^{-} & -(X'V^{-1}X)^{-}X'V^{-1}ZG \\ -GZ'V^{-1}X(X'V^{-1}X)^{-} & (Z'R^{-1}Z+G^{-1})^{-1}+GZ'V^{-1}X(X'V^{-1}X)^{-}X'V^{-1}ZG \end{bmatrix}. \tag{22a}$$

$\text{var}(\hat{u}) = G - C_{22}(5.38)$

$\text{cov}(\hat{u}, u') = C'V^{-1}\text{cov}[(y - X\beta^0), u']$

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

$= GZ'V^{-1}[I - X(X'V^{-1}X)^{-}X'V^{-1}]ZG$

$= GZ'V^{-1}ZG + (Z'R^{-1}Z + G^{-1})^{-1} - C_{22}$

$=$ G - $C_{22}(5.39)$

$\text{var}(\hat{u} - u) = G - C_{22} + G - 2(G - C_{22}) = C_{22}(5.40)$

$\text{var}(\hat{w} - w) = v[K'(\beta^0 - \beta) + \hat{u} - u]$

$= \text{v}(K'\beta^0) + v(\hat{u} - u) + \text{cov}(K'\beta^0, \hat{u}' - u') + \text{cov}(\hat{u} - u, \beta^{0'}K)$

$= K'C_{11}K + C_{22} + K'C_{12} + C_{12}'K.(5.41)$

5.11 Prediction of errors [41, 5.8]

Equation (5.18) is for *scalar* w with $E(w) = k'\beta$, $\text{var}(w) = v$ and $\text{cov}(w, y') = c'$, giving

$$\tilde{w} = k'\beta^0 + c'V^{-1}(y - X\beta^0).$$

Adapted to *vector* w, k' becomes K', and c' becomes C so that

$$\tilde{w} = K'\beta^0 + C'V^{-1}(y - X\beta^0).$$

Thus the special case

$$w = \epsilon = y - X\beta, \qquad E(w) = 0 \Rightarrow K' = 0$$

gives

$$\tilde{\epsilon} = CV^{-1}(y - X\beta^0) \qquad [41, \text{line } 4]$$

and then, because

of (5.33). Then

$$\begin{aligned}
\hat{e}_p &= (I - WCW'R^{-1})y \\
\text{var}(\hat{e}_p - e_p) &= \text{var}(y - X\beta^0 - Z\tilde{u} - e_p) \\
&= \text{var}(X\beta + Zu + e_p - X\beta^0 - Z\tilde{u} - e_p) \\
&= \text{var}[-X(\beta^0 - \beta) - Z(\tilde{u} - u)] \\
&= X\text{var}(\beta^0 - \beta)X' + Z\ \text{var}(\tilde{u} - u)Z' + 2X\text{cov}[\beta^0 - \beta, (\tilde{u} - u)']Z' \\
&= XC_{11}X' + ZC_{22}Z' + 2XC_{12}Z', \text{ from (5.40) and (5.37), respectively.} \\
&= WCW'. \qquad [42,\ \text{line 2}]
\end{aligned}$$

$$\begin{aligned}
\text{cov}[(\hat{e}_p - e_p),\ (K'\beta^0)'] &= \text{cov}[\{-X(\beta^0 - \beta) - Z(\tilde{u} - u)\},\ (K'\beta^0)'] \\
&= -X\text{var}(\beta^0)K - Z\text{cov}[(\tilde{u} - u),\ \beta^{0'}]K \\
&= -XC_{11}K - ZC_{12}K \text{ from (5.34), (5.37)} \\
&= -WC'K. \qquad [42,\ \text{line 7}] \\
\text{cov}[(\hat{e}_p - e_p,\ (\hat{u} - u)'] &= \text{cov}[\{-X(\beta^0 - \beta) - Z(\hat{u} - u)\},\ (\hat{u} - u)'] \\
&= -XC_{12} - ZC_{22} \\
&= -WC_2' \qquad [42,\ \text{line 8}] \\
\text{cov}[(\hat{e}_p - e_p,\ (\tilde{e}_m - e_m)'] &= \text{cov}[(\hat{e}_p - e_p,\ \{R'_{pm}R_{pp}^{-1}(\hat{e}_p - e_p)\}'] \\
&= \text{var}(\hat{e}_p - e_p)R_{pp}^{-1}R_{pm} \\
&= WCW'R_{pp}^{-1}R_{pm} \qquad [42,\ \text{line 9}] \\
\text{var}(\hat{e}_m - e_m) &= \text{var}(\hat{e}_m) + \text{var}(e_m) - 2\text{cov}(\hat{e}_m, e'_m) \\
&= \text{var}(R'_{pm}R_{pp}^{-1}\hat{e}_p) + R_{mm} - 0, \qquad (23)
\end{aligned}$$

the zero because $\hat{e}_m$ is a function of $\hat{e}_p$, and hence of y; and e_m does not occur in y. We now need $\text{var}(\hat{e}_p)$. By definition

$$\text{var}(e_p) = R_{pp}.$$

This simplifies by using

$$\begin{aligned} V^{-1} - V^{-1}X(X'V^{-1}X)^{-}X'V^{-1} &= P \\ ZGZ' &= V - R \end{aligned}$$

and

$$\begin{aligned} Z(Z'R^{-1}Z + G^{-1})^{-1} &= ZG\left[Z'R^{-1}Z + G^{-1} - Z'R^{-1}Z\right](Z'R^{-1}Z + G^{-1})^{-1}Z' \\ &= ZG\left[I - Z'R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}\right]Z' \\ &= ZGZ'\left[R^{-1} - R^{-1}Z(Z'R^{-1}Z + G^{-1})Z'R^{-1}\right]R \\ &= ZGZ'V^{-1}R = (V - R)V^{-1}R \\ &= R - RV^{-1}R. \end{aligned}$$

Therefore

$$\begin{aligned} WCW' &= V(V^{-1} - P)V + V(P - V^{-1})(V - R) + (V - R)(P - V^{-1})V \\ &\quad + (R - RV^{-1}R) + (V - R)(V^{-1} - P)(V - R) \\ &= V - VPV + VPV - V - VPR + R + VPV - RPV - V + R \\ &\quad + R - RV^{-1}R + V - R - R + RV^{-1}R - VPV - RPV + VPR - RPR, \end{aligned}$$

and from this everything cancels except $R - RPR$, so leaving $WCW' = R - RPR$. Hence (25) is

$$\begin{aligned} \mathrm{var}(\hat{e}_p - e_p) &= [I - (R - RPR)R^{-1}]V[I - R^{-1}(R - RPR)] + R - 2[I - (R - RPR)R^{-1}]R \\ &= RPVPR + R - 2RPR \\ &= RPR + R - 2RPR, \quad \text{because} \quad PVP = P \\ &= R - RPR \\ &= WCW. \end{aligned}$$

Now consider the last two results preceding [42, 5.9]. From [41, last line] has

$$\text{var}\begin{bmatrix} y \\ 0 \end{bmatrix} = \begin{bmatrix} V & 0 \\ 0 & 0 \end{bmatrix}.$$

An alternative to (28) is

$$y = X\beta + [Z \quad 0]\begin{bmatrix} u \\ u_n \end{bmatrix} + e.$$

Applying the formula for β^0 in (27) to this set-up gives the estimator β^* as

$$\begin{aligned}
\beta^* &= \left\{[X' \quad X_n']\begin{bmatrix} V^{-1} & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} X \\ X^n \end{bmatrix}\right\}^- [X' \quad X_n']\begin{bmatrix} V^{-1} & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} y \\ 0 \end{bmatrix} \\
&= (X'V^{-1}X)^- X'V^{-1}y \\
&= \beta^0.
\end{aligned}$$

Likewise, applying $\hat{u}$ of (27) gives

$$\begin{aligned}
\begin{bmatrix} \tilde{u} \\ \tilde{u}_n \end{bmatrix} &= \text{cov}\left\{\begin{bmatrix} u \\ u_n \end{bmatrix}, \begin{bmatrix} y \\ 0 \end{bmatrix}'\right\}\begin{bmatrix} V & 0 \\ 0 & 0 \end{bmatrix}^- \begin{bmatrix} y - X\beta^* \\ 0 - X\beta^* \end{bmatrix} \\
&= \begin{bmatrix} GZ' & 0 \\ B' & 0 \end{bmatrix}\begin{bmatrix} V^{-1} & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} y - X\beta^0 \\ 0 - X\beta^0 \end{bmatrix} \\
&= \begin{bmatrix} GZ'V^{-1}(y - X\beta^0) \\ B'V^{-1}(y - X\beta^0) \end{bmatrix}.
\end{aligned}$$

Thus we get

$$\hat{u} = GZ'V^{-1}(y - X\beta^0), \quad \text{as before}$$

and

$$\begin{aligned}
\hat{u}_n &= B'V^{-1}(y - X\beta^0) &\quad (5.47) \\
&= C'Z'V^{-1}(y - X\beta^0) = C'G^{-1}\hat{u}. &\quad (5.48)
\end{aligned}$$

for

$$T = (G_{11} - C'G^{-1}C)^{-1},$$

giving $W_{11} = G^{-1} + G^{-1}CTC'G^{-1}$, $W_{12} = -G^{-1}CT$ and $W_{22} = T$.

It is stated that equations (5.49) have the same solutions as (5.48) and $\hat{u}$ preceding (5.47). We show this.

First, from the last equation of (5.49)

$$\begin{aligned} W_{22}\hat{u}_n &= -(W_{12})'\hat{u} \\ \hat{u}_n &= T^{-1}TC'G^{-1}\hat{u} = C'G^{-1}\hat{u} \end{aligned}$$

which is (5.48). Then, part of the second equation of (5.49) is

$$\begin{aligned} W_{11}\hat{u} + W_{12}\hat{u}_n &= W_{11}\hat{u} + W_{12}C'G^{-1}\hat{u} \\ &= (G^{-1} + G^{-1}CTC'G^{-1} - G^{-1}CTC'G^{-1})\hat{u} \\ &= G^{-1}\hat{u}. \end{aligned}$$

Thus that second equation is

$$Z'R^{-1}X\beta^0 + (Z'R^{-1}Z + G^{-1})\hat{u} = Z'R^{-1}y$$

which, with the first equation of (5.49) constitutes the MMEs (29).

5.13 Prediction When G is Singular [43, 5.10]

Let H be a matrix we do not like, e.g., the matrix of the MMEs when G is singular. Then the matrix in (5.50) is

$$\begin{bmatrix} I & 0 \\ 0 & G \end{bmatrix} H = L, \text{ say.}$$

Now compute C, a generalized inverse of L:

$$LCL = L.$$

where there is only one record on each animal: var(a) $= A\sigma_a^2$. The MMEs have order $p+n$

Nevertheless, under these conditions it is suggested that equations

$$\begin{bmatrix} V & X \\ X' & 0 \end{bmatrix} \begin{bmatrix} s \\ \beta^0 \end{bmatrix} = \begin{bmatrix} y \\ 0 \end{bmatrix} \tag{5.57}$$

be used. No indication is given as to the origin of these equations, nor as to why they, of order $n+p$, should be used (only?) when $p+q>n$; i.e., $n+p>2n-q$. The equations are easily solved

$$\begin{aligned} Vs + X\beta^0 = y \Rightarrow s &= V^{-1}(y - X\beta^0) \\ X's &= 0 = X'V^{-1}y - X'V^{-1}X\beta^0 \\ (X'V^{-1}X)\beta^0 &= X'V^{-1}y \\ \hat{u} &= GZ's = GZ'V^{-1}(y - X\beta^0). \end{aligned}$$

Define

$$\begin{aligned} \begin{bmatrix} C_{11} & C_{12} \\ C_{12}' & C_{22} \end{bmatrix} &= \begin{bmatrix} V & X \\ X' & 0 \end{bmatrix}^- = \begin{bmatrix} V^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -V^{-1}X \\ I \end{bmatrix} [0 - X'V^{-1}X]^- [-X'V^{-1} \quad I] \\ &= \begin{bmatrix} V^{-1} - V^{-1}X(X'V^{-1}X)^- X'V^{-1} & V^{-1}X(X'V^{-1}X)^- \\ (X'V^{-1}X)^- X'V^{-1} & -(X'V^{-1}X)^- \end{bmatrix}. \end{aligned}$$

Hence $C_{11} = P$ and so

$$\begin{aligned} \text{var}(K'\beta^0) &= \text{var}[K'(X'V^{-1}X)^- X'V^{-1}y] \\ &= K'(X'V^{-1}X)^- X'V^{-1}VV^{-1}X(X'V^{-1}X)^- K \\ &= -K'C_{22}K \end{aligned} \tag{5.59}$$

$$\begin{aligned} \text{var}(\hat{u}) &= \text{var}[GZ'V^{-1}(y - X\beta^0)] \\ &= \text{var}(GZ'Py) \\ &= GZ'PZG \quad \text{because} \quad PVP = P \\ &= GZ'C_{11}ZG \quad \text{because} \quad C_{11} = P. \end{aligned} \tag{5.60}$$

$$\text{cov}(K'\beta^0, \hat{u}') = \text{cov}\{K'(X'V^{-1}X)^- X'V^{-1}y,\ y'P\}$$

$$\begin{bmatrix} 2 & 1 & 4 & 1 & 5 \\ 1 & 7 & -11 & -6 & -17 \\ 4 & -11 & 34 & 15 & 49 \\ 1 & -6 & 15 & 7 & 22 \\ 5 & -17 & 49 & 22 & 71 \end{bmatrix} = \begin{bmatrix} I \\ L' \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 7 \end{bmatrix} [I \quad L]$$

for

$$L = \begin{bmatrix} 3 & 1 & 4 \\ -2 & -1 & -3 \end{bmatrix}, \text{ of order } 2 \times 3 \text{ has rank 2, not 3.}$$

Finally, even after writing (30) I see no reason why X and Z being linearly dependent on R leads to $X = \begin{bmatrix} X_1 \\ L'X_1 \end{bmatrix}$; and $Z = \begin{bmatrix} Z_1 \\ LZ_1 \end{bmatrix}$. True, CRH says "if" X and Z are of this nature. Then, of course

$$V = ZGZ' + R = \begin{bmatrix} I \\ L \end{bmatrix} (Z_1G_1' + R_1)[I \quad L']$$

is singular.

5.19 Another Example: Numeric [59, 5.15]

5.20 Prediction When u and e Are Correlated [61, 5.17]

Derivation of (5.82) is straightforward. Verification of (5.81) is easy:

$$\begin{aligned} \text{var}(\epsilon) &= \text{var}[Zu + e - Tu] = \text{var}(e - SG^{-1}u) \\ &= R + SG^{-1}GG^{-1}S' - 2SG^{-1}S' = R - SG^{-1}S' \\ &= B \end{aligned}$$

$$\begin{aligned} \text{cov}(Tu,\ \epsilon') &= \text{cov}[Tu, e' - u'G^{-1}S'] = TS' - TGG^{-1}S' \\ &= TS' - TS' = 0. \end{aligned}$$

Chapter 6

G and R Known to Proportionality

6.1 Defining Proportionality

It is assumed that

$$G = G_* \sigma_e^2 \qquad \text{and} \qquad R = R_* \sigma_e^2 \tag{6.1}$$

where G_* and R_* are taken as *known*, but σ_e^2 is unknown.

6.2 BLUE and BLUP [70, 6.2]

With $V = V_* \sigma_e^2$, the equations for β^0 and $\hat{u}$ are precisely as previously, but with V replaced by V_*. To show that (6.6) is the same as (6.7) note that the numerator of (6.6) is

$$\begin{aligned}
& y'V^{-1}y - \beta^{0'}X'V^{-1}y \\
&\quad = y'V^{-1}(y - X\beta^0) \\
&\quad = y'R^{-1}[I - Z(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1}](y - X\beta^0) \\
&\quad = y'R^{-1}[(y - X\beta^0) - Z\hat{u}], \qquad \text{after using [41, (5.44)]} \\
&\quad = y'R^{-1}y - y'R^{-1}X\beta^0 - y'R^{-1}Z\hat{u} \\
&\quad = y'R^{-1}y - \beta^{0'}X'R^{-1}y - \hat{u}'Z'R^{-1}y
\end{aligned}$$

which is essentially the numerator of [71, (6.7)]. The remainder of Chapter 6 concerning tests of hypotheses seems straightforward.

Chapter 7

Known Functions of Fixed Effects

7.1 Tests of Estimability [75, 7.1]

For $T'\beta$ non-estimable, T' of full row rank $t < p - r$, it is stated that there is always a matrix C, of order $p \times (r - t)$ and full column rank, such that

$$\begin{bmatrix} X \\ T' \end{bmatrix} C = 0. \tag{7.1}$$

And then $K'\beta$ is estimable if and only if

$$K'C = 0. \tag{7.2}$$

<u>Proof</u> (i): If $K'\beta$ is estimable then $K'C = 0$.

Estimable $K'\beta$ means $K' = Q'X$ for some Q'. Therefore

$$K'C = Q'XC = Q'0 = 0, \quad \text{because} \quad XC = 0 \quad \text{from (7.1)}.$$

<u>Proof</u> (ii): If $K'C = 0$ then $K' = Q'X$ for some Q'.

From (7.1) $XC = 0 \Rightarrow C = (I - X^-X)z$, for arbitrary z. Therefore, if $K'C = 0$, we have $K'(I - X^-X)z = 0$; and letting z be in turn the columns of I gives $K' = K'X^-X = Q'X$ for $Q' = K'X^-$.

7.3 Sampling Variances [79, 7.3]

For (7.11)

$$\begin{aligned}
\begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} &= \begin{bmatrix} X'V^{-1}X & T \\ T' & 0 \end{bmatrix}^{-} \\
&= \begin{bmatrix} (X'V^{-1}X)^{-} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -(X'V^{-1}X)^{-}T \\ I \end{bmatrix} [-T'(X'V^{-1}X)^{-}T]^{-}[-T'(X'V^{-1}X)^{-} \quad I] \\
C_{11} &= (X'V^{-1}X)^{-} - (X'V^{-1}X)^{-}T[T'(X'V^{-1}X)^{-}T]'T'(X'V^{-1}X)^{-}
\end{aligned}$$

and

$$\mathrm{var}(K'\beta^0) = K'C_{11}K. \tag{7.6}$$

From (7.12) – (7.14) when $c = 0$

$$\mathrm{var}(K'\beta^0) = \mathrm{var}\left[(K_1' \quad K_2')\begin{pmatrix} \beta_1^0 \\ \beta_2^0 \end{pmatrix}\right] = \mathrm{var}[K_1'\beta_1^0 + K_2'(-T_2'^{-1}T_1')\beta_1^0].$$

Write

$$S' \text{ for } T_2'^{-1}T_1' \text{ and } M \text{ for } [I \quad -S]$$

$$\begin{aligned}
\mathrm{var}(K'\beta^0) &= \mathrm{var}\{([I \quad -S]K)'\beta_1^0\} \\
&= K'M'(W'V^{-1}W)^{-}MK \\
&= K'M'(MX'V^{-1}XM')^{-}MK.
\end{aligned}$$

Question How can this be shown equal to (7.11), which is $K'C_{11}K$?

7.4 Hypothesis Testing [80, 7.4]

This seems straightforward.

Chapter 8

Methods for G and R Unknown

8.1 Unbiased Estimators [83, 8.1]

The last line of [83] and the first of [84] refer to $\tilde{G}$ and $\tilde{R}$, as defined in items 2 and 3 prior to [83, 8.1].

The first line of [83, 8.1] indicates that there are many unbiased estimators of $K'\beta$ – for which $K'\beta$ is usually considered estimable, i.e.

$$K' = T'X \tag{1}$$

for some T'. On [84-5] at least six such estimators are suggested. We discuss these six, using the symbol

$$\mathrm{Var}(y) = V = ZGZ' + R \tag{2}$$

more than does [85-6].

8.1.1 Ordinary Least Squares (OLS) [8.4, (8.1) and (8.2)]

Solve

$$X'X\beta^0 = X'y. \tag{8.1}$$

Then

$$E(K'\beta^0) = K'(X'X)^- X'E(y) = T'X(X'X)^- X'X\beta = X\beta$$

and

$$\mathrm{var}(K'\beta^0) = K'(X'X)^- X'VX(X'X)^- K. \tag{8.2}$$

Comment

(i) No reason is given for defining D as the diagonal matrix of the diagonal elements of V. That definition of D is not customary in statistics.

(ii) In place of D^{-1} in [84, 8.3] one usually finds V^{-1} with the result

$$\beta^0 = (X'V^{-1}X)^{-}X'V^{-1}y.$$

Then for estimable $K'\beta$, the best linear unbiased estimator (BLUE) is

$$\text{BLUE}(K'\beta) = K'\beta^0 \quad \text{for} \quad K' = T'X.$$

This is often referred to as the generalized least squares estimator (GLSE) or weighted least squares estimator (WLSE). An even more general form is $K'(X'WX)^{-}X'Wy$ for any symmetric, non-negative definite matrix W. This is discussed in Searle (1995) where, for example, it is shown to be an unbiased estimator of estimable $K'\beta$ if and only if $X = CWX$ (with $WX \neq 0$) for some C.

8.1.3 GLSE using $\tilde{R}^{-1}$ [84 , (8.5) and (8.6)]

Solve

$$X'\tilde{R}^{-1}X\beta^0 = X'\tilde{R}^{-1}y \tag{8.5}$$

giving

$$\text{var}(K'\beta^0) = K'(X'\tilde{R}^{-1}X)^{-}X'\tilde{R}^{-1}V\tilde{R}^{-1}X(X'\tilde{R}^{-1}X)^{-}K. \tag{8.6}$$

Comment (from L.R. Schaeffer)

In animal breeding situations the customary forms of G and R are $G = A\sigma_a^2$ and $R = \sigma_e^2 I$, usually with $\sigma_e^2 \gg \sigma_a^2$ and hence $1/\sigma_a^2 > 1/\sigma_e^2$. This is the basis for the sentence which follows [84 (8.6)]. On the other hand, in the MMEs the $G^{-1}\sigma_e^2 = A^{-1}\sigma_e^2/\sigma_a^2 \to 0$ as $\sigma_a^2 \to \infty$ (or if $\sigma_a^2 \gg \sigma_e^2$) and then the MMEs $\to$ OLS, as in [84, (8.7)].

Then

$$\begin{aligned}
\mathrm{var}(K'\beta^0 + Mu^0) &= \mathrm{var}[K' \quad M']CW'y] \\
&= [K' \quad M']CW'VWC' \begin{bmatrix} K \\ M \end{bmatrix} \qquad (8.8) \\
&= [K' \quad M']CW'[R + ZGZ']WC' \begin{bmatrix} K \\ M \end{bmatrix} \\
&= [K' \quad M']CW'RWC' \begin{bmatrix} K \\ M \end{bmatrix} + [K' \quad M']CW'ZGZ'WC' \begin{bmatrix} K \\ M \end{bmatrix}. \qquad (5)
\end{aligned}$$

If the second term is to simplify to $M'GM$ as in (8.9), we must consider

$$\begin{aligned}
&(K' \quad M')CW'Z \\
&= [T'X \quad M'] \left\{ \begin{bmatrix} (X'X)^- & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -(X'X)^- X'Z \\ I \end{bmatrix} (Z'PZ)^-[-Z'X(X'X)^- \quad I] \right\} \begin{bmatrix} X'Z \\ Z'Z \end{bmatrix} \qquad (6) \\
&= T'X(X'X)^- X'Z + [-T'X(X'X)^- X'Z + M'](Z'PZ)^-(Z'PZ) \\
&= T'X(X'X)^- X'Z[I - (Z'PZ)^- Z'PZ] + M'(Z'PZ)^- Z'PZ \\
&= M' \quad \text{if} \quad (Z'PZ)^- Z'PZ = I.
\end{aligned}$$

Then the second term in (5) is $M'GM$ and (8.9) is established.

If $R = \sigma_e^2 I$ the first term of (8.9) is

$$(K' \quad M')CW'WC \begin{bmatrix} K \\ M \end{bmatrix} = [K' \quad M']C \begin{bmatrix} K \\ M \end{bmatrix} \qquad (8.10)$$

because C is a generalized inverse of $W'W$ and to get (8.10) we take C to be symmetric and reflexive.

$$Z'PZu^0 = Z'PZ(Z'PZ)^-[Z' - Z'X(X'X)^-X]y = Z'Py. \tag{8.17}$$

Then, since $E(y) = X\beta$ and $PX = 0$, and for $(Z'PZ)^- \equiv C$

$$E(u^0) = CZ'PX\beta = 0.$$

This is often described as u^0 being unbiased; but note that that is not the usual statistical meaning of unbiased. The statistical meaning is that the expected value of a parameter estimator equals the parameter; e.g., $E(\hat{\beta}) = \beta$. But in $E(u^0) = 0$ the 0 is not a parameter. Maybe, if the model includes $E(u) = 0$, one could call the 0 a parameter – but that is stretching things a bit.

Clearly, from (8.17)

$$\begin{aligned} u^0 &= (Z'PZ)^-Z'Py = CZ'Py \\ \text{var}(u^0) &= CZ'PVPZC \qquad (8.18) \\ \text{cov}(u^0, u') &= CZ'PZG. \qquad (8.19) \end{aligned}$$

Question

Derivation of BLUP(u) as TS^-u^0 of (8.21) is as follows, with $P = I - X(X'X)^-X'$ and, as in [88, line preceding (5.18)], $C = (Z'PZ)^-$. Hence, taking $V = I$,

$$\begin{aligned} TS^-u^0 &= GZ'PZC(CZ'PVPZC)^-CZ'Py. \qquad (7) \\ &= GC^-C(CC^-C)^-CZ'Py \\ &= GC^-C(C^-)CZ'Py \\ &= GC^-CZ'Py \\ &= GZ'PZ(Z'PZ)^-Z'Py \\ &= GZ'Py, \quad \text{because} \quad Z'PZ(Z'PZ)^-Z'P = Z'P(Z'P)'[Z'P(Z'P)']^-Z'P = Z'P \\ &= GZ'[y - X(X'X)^-X'y] \\ &= GZ'(y - X\beta^0) \\ &= \text{BLUP}(u) \quad \text{with} \quad V = I. \end{aligned}$$

Note: (8.20) is for an individual u_i, whereas (8.21) is for all of the u_i together and so is (Schaeffer) optimal; but (8.20) is not.

Chapter 9

Biased Estimation and Prediction

9.1 Derivation of BLBE and BLBP [93, 9.1]

Acronyms: BLBP: best linear biased predictor
BLBE: best linear biased estimator.

For predicting $k_1'\beta_1 + k_2'\beta_2 + m'u$ with $a'y$ the mean square error of prediction is given as

$$\text{MSE} = a'Ra + (a'X_2 - k_2')\beta_2\beta_2'(X_2'a - k_2) + (a'Z - m')G(Z'a - m). \tag{9.1}$$

It seems as if β_2 is here being treated as known, although that is never explicitly stated. In other words, β_2 seems to be getting treated as a prior value of β_2: see item 1 on [99].

(9.1) is not quite correct. It is, in a sense, after reading the two lines below [93, (9.1)]; i.e., after using $a'X_1 = k_1$. Explanation follows.

Derivation starts with $\text{MSE} = E(a'y - k_1'\beta_1 - k_2'\beta_2 - m'u)^2$. For convenience write

$$s_1 = k_1'\beta_1 \quad \text{and} \quad s_2 = k_2'\beta_2,$$

noting that each is a scalar. Then

$$\begin{aligned}
\text{MSE} &= E(a'y - s_1 - s_2 - m'u)^2 \\
&= E[(a'y)^2 + s_1^2 + s_2^2 + (m'u)^2 - 2(s_1 + s_2)a'y - 2a'yu'm + 2s_1s_2 + 2(s_1 + s_2)m'u] \\
&= E(a'yy'a) + s_1^2 + s_2^2 + E(m'uu'm) - 2(s_1 + s_2)a'(X_1\beta_1 + X_2\beta_2) \\
&\qquad - 2a'ZGm + 2s_1s_2 + 2(s_1 + s_2)m'0 \\
&= a'[V + (X_1\beta_1 + X_2\beta_2)(X_1\beta_1 + X_2\beta_2)']a + s_1^2 + s_2^2 + 2s_1s_2 + m'(G + 0)m \\
&\qquad - 2(s_1 + s_2)a'(X_1\beta_1 + X_2\beta_2) - 2a'ZGm.
\end{aligned}$$

The feature of interest is therefore $\partial\text{MSE}/\partial a$. Let us label (9.1) as

$$\text{MSE}_1 = a'Ra + [(a'X_2 - k_2')\beta_2]^2 + (a'Z - m)G(a'Z - m)'$$

and then using MSE_2 for (1)

$$\text{MSE}_2 = \text{MSE}_1 + [(a'X_1 - k_1')\beta_1]^2 + 2(a'X_2 - k_2')\beta_2(a'X_1 - k_1')\beta_1.$$

Then

$$\begin{aligned}
\frac{1}{2}\frac{\partial}{\partial a}\text{MSE}_1 &= Ra + (a'X_2 - k_2')\beta_2 X_2\beta_2 + (ZGZ'a - ZGm) \\
&= (R + ZGZ')a + a'X_2\beta_2 X_2\beta_2 - k_2'\beta_2 X_2\beta_2 - ZGm \\
&= Va + X_2\beta_2(a'X_2\beta_2)' - X_2\beta_2(k_2'\beta_2)' - ZGm \\
&= (V + X_2\beta_2\beta_2'X_2')a - (X_2\beta_2\beta_2'k_2 + ZGm).
\end{aligned}$$

Therefore equation (2) for MSE_1 is

$$\begin{bmatrix} V + X_2\beta_2\beta_2'X_2 & X_1 \\ X_1' & 0 \end{bmatrix} \begin{bmatrix} a \\ \theta \end{bmatrix} = \begin{bmatrix} X_2\beta_2\beta_2'k_2 + ZGm \\ k_1 \end{bmatrix}. \tag{9.2}$$

In contrast to this

$$\begin{aligned}
\frac{1}{2}\frac{\partial}{\partial a}(\text{MSE}_2) &= \frac{1}{2}\frac{\partial}{\partial a}(\text{MSE}_1) + (a'X_1 - k_1')\beta_1 X_1\beta_1 + X_2\beta_2(a'X_1 - k_1')\beta_1 \\
&\quad + (a'X_2 - k_2')\beta_2 Z_1\beta_1.
\end{aligned}$$

Thus

$$\begin{aligned}
&\frac{1}{2}\frac{\partial}{\partial a}(\text{MSE}_2) - \frac{1}{2}\frac{\partial}{\partial a}(\text{MSE}_1) \\
&\quad = [(a'X_1 - k_1')\beta_1](X_1\beta_1 + X_2\beta_2) + (a'X_2 - k_2')\beta_2 X_1\beta_1 \\
&\quad = X_1\beta_1\beta_2'X_2'a - X_1\beta_1\beta_2'k_2, \quad \text{on using } X_1'a = k_1.
\end{aligned}$$

Therefore for MSE_2 used in (2) the equations are

$$\begin{bmatrix} V + X_2\beta_2\beta_2'X_2' + X_1\beta_1\beta_2'X_2' & X_1 \\ X_1' & 0 \end{bmatrix} \begin{bmatrix} a \\ \theta \end{bmatrix} = \begin{bmatrix} X_2\beta_2\beta_2'k_2 + ZGm + X_1\beta_1\beta_2'k_2 \\ k_1 \end{bmatrix}$$

9.3 Assumed Pattern of Values of β [96, 9.3]

The connection of β to the average values in (9.13) – (9.16) is not clear. It seems as if, given

$$\text{average } \alpha_{ij}^2 = \gamma; \tag{9.13}$$

then, because it is being assumed that

$$\sum_{j=1}^{c} \alpha_{ij} = 0$$

we have

$$\left(\sum_{j=1}^{c} \alpha_{ij}\right)^2 = \sum_{j=1}^{c} \alpha_{ij}^2 + \sum_{j \neq j'}\sum \alpha_{ij}\alpha_{ij'} = 0.$$

Hence

$$\frac{\sum\sum_{j \neq j'} \alpha_{ij}\alpha_{ij'}}{c(c-1)} = \frac{\sum_{j=1}^{c} \alpha_{ij}^2}{c(c-1)} = \frac{-\gamma}{c-1}. \tag{9.14}$$

(9.15) follows similarly from $\sum_{i=1}^{r} \alpha_{ij} = 0$. And from

$$\left(\sum_{i=1}^{r}\sum_{j=1}^{c} \alpha_{ij}\right)^2 = \sum_{i=1}^{r}\sum_{j=1}^{c} \alpha_{ij}^2 + \sum_{i \neq i'}\sum\sum_{j \neq j'}\sum \alpha_{ij}\alpha_{i'j} = 0$$

dividing by $rc(r-1)(c-1)$ gives [96, (9.16)] – only without the minus sign. HOW COME?

But notice: the book gives no details of the subscripts: presumably it is $i \neq i'$ and $j \neq j'$, but nothing is said on this score.

9.4 Evaluation of Bias [96, 9.4]

It is convenient for this section and the next to use H of (9.26):

$$H = \begin{bmatrix} X_1'R^{-1} \\ PX_2'R^{-1} \\ Z'R^{-1} \end{bmatrix} \tag{4}$$

and to observe that for (9.24) and (9.25)

$$T = HX_2 \qquad \text{and} \qquad S = HZ. \tag{5}$$

Comment I find all this to be unrealistic. Nowhere does there seem to be a statement of re-estimating β_2 starting from some pre-assigned value of it. And the text has some mystifying statements: [95, line 2] has "If P were non-singular". That is impossible. P is $\beta_2\beta_2'$, the outer product of a vector with itself; that is *always* singular. And [95, lines 1-2 of the paragraph preceding Section 9.2] has "β_2^* has a peculiar and seemingly undesirable property, namely $\beta_2^* = k\beta_2$ where k is some constant".

This does not seem to be good statistical practice.

9.5 Evaluation of Mean Squared Errors [97, 9.5]

This would seem to require evaluation of

$$\Delta = E\left\{(M_1' \;\; M_2' \;\; M_3')\left[CHy - \begin{pmatrix}\beta_1\\ \beta_2\\ u\end{pmatrix}\right]\left[CHy - \begin{pmatrix}\beta_1\\ \beta_2\\ u\end{pmatrix}\right]'\begin{bmatrix}M_1\\ M_2\\ M_3\end{bmatrix}\right\}. \tag{9}$$

Problem I cannot reduce Δ to be (9.28). To begin, consider

$$\begin{aligned}\Delta_1 &= E[CHy(CHy)'] = CH\,E(yy')H'C' \\ &= CH[V + E(y)E(y')]H'C' \\ &= CH[R + ZGZ' + (X_1\beta_1 + X_2\beta_2)(X_1\beta_1 + X_2\beta_2)']H'C' \qquad (10) \\ &= CHRH'C' + CSGSC' + \left[\begin{pmatrix}\beta_1\\ 0\\ 0\end{pmatrix} + CT\beta_2\right]\left[\begin{pmatrix}\beta_1\\ 0\\ 0\end{pmatrix} + CT\beta_2\right]'\end{aligned}$$

after using (5) and (8). Now as part of B of (9.27) $CHRH'C'$ is the last of the three expressions prior to the equal signs. And $CSGSC'$ in Δ_1 is very like the second of those three expressions except it has $C_3S - I$ whereas Δ_1 has C_3S. Likewise, the last term of Δ_1 has $CT\beta_2\beta_2'T'C'$ wherein CT includes C_2T but in the text, the first term in (9.27) has $C_2T - I$; and, of course, there are other terms in Δ_1 coming from that final product; e.g., $\begin{pmatrix}\beta_1\beta_2'T'C\\ 0\\ 0\end{pmatrix}$.

Problem Where do these terms $C_3S - I$ and $C_2T - I$ come from?

$$= \begin{bmatrix} 0 & -C_1T\beta_2\beta_2' & -C_1SG \\ -\beta_2\beta_2'T'C_1' & \beta_2\beta_2' - \beta_2\beta_2'T'C_2' - C_2T\beta_2\beta_2' & -\beta_2\beta_2'T'C_3' - C_2SG \\ -GS'C_1' & -C_3T\beta_2\beta_2' - GS'C_2' & -GS'C_3' - C_3SG \end{bmatrix}$$

– which is nowhere near part of B!

9.6 Estimability in Biased Estimation [99, 9.6]

Lines 3-4 of [99, 9.6] suggest that if "we relax the requirement of unbiasedness is the above an appropriate definition of estimability?"

Comment Surely if unbiasedness is relaxed then in the context of estimation there is no linear function (i.e., linear combination of elements) of y that has expectation $K'\beta$. That being so, estimability becomes disconnected from unbiasedness.

[99, item 1] seems to be the first clear statement of intending to use an *á priori* value of β_2 for getting a better estimate. What a pity that was not stated on [93].

At [100, lines 3-4], if t_3^0.is the *á priori* for t_3 why not estimate μ as $\hat{\mu} = \bar{y}_{3\cdot} - t_3^0$? And at [100, bottom] why not estimate $\mu + a_2 + b_3$ as $\bar{y}_{23\cdot}$?

9.7 Tests of Hypotheses [101, 9.7]

Comment At the bottom of [101] it seems confusing to have a C partitioned in 2×2 form when it applies to a matrix that is a 3×3 form. But presumably C_{11}, of order $p \times p$ corresponds to the $(X_1 \quad X_2)'(X_1 \quad X_2)$ parts of (9.32) and (9.33) and C_{22} to the $Z'Z$ part.

Typo At [101, 4 lines up] the second β^* needs no "hat".

9.8 Estimation of P [102, 9.8]

Comment I don't like $P = \beta_2\beta_2'$ as part of an estimation procedure.

The determinant term is

$$\begin{aligned}
\left(\frac{|V|}{|R||C|}\right)^{\frac{1}{2}} &= \left(\frac{|WCW'+R|}{|R||C|}\right)^{\frac{1}{2}} = \left(\frac{|R||WCW'R^{-1}+I|}{|R||C|}\right)^{\frac{1}{2}} \\
&= \left(\frac{|CW'R^{-1}W+I|}{|C|}\right)^{\frac{1}{2}}, \text{ because } |AB+I| = |BA+I| \\
&= |W'R^{-1}W + C^{-1}|^{\frac{1}{2}}. \qquad (15)
\end{aligned}$$

And the exponential term is

$$\exp -\frac{1}{2}\{\gamma'(W'R^{-1}W+C^{-1})\gamma - 2\gamma'(W'R^{-1}y+C^{-1}\mu) + y'(R^{-1}-V^{-1})y$$

$$+\mu'C^{-1}\mu - \beta'X'V^{-1}X\beta + 2\beta'X'V^{-1}y\}.$$

Now use

$$\begin{aligned}
V^{-1} &= (WCW'+R)^{-1} \\
&= R^{-1} - R^{-1}W(W'R^{-1}W+C^{-1})^{-1}W'R^{-1}
\end{aligned}$$

and for any symmetric A and vector t

$$\gamma'A\gamma - 2\gamma't = (\gamma - A^{-1}t)'A(\gamma - A^{-1}t) - t'A^{-1}t.$$

Thus for the exponential term we get

$$\begin{aligned}
\exp -\frac{1}{2}\Big\{&[\gamma - (W'R^{-1}W+C^{-1})^{-1}(W'R^{-1}y+C^{-1}\mu)]'(W'R^{-1}W+C^{-1}) \\
&\times[\gamma - (W'R^{-1}W+C^{-1})^{-1}(W'R^{-1}y+C^{-1}\mu)] \\
&+ y'R^{-1}W(W'R^{-1}W+C^{-1})W'R^{-1}y \\
&+ \mu'C^{-1}\mu - \beta'X'V^{-1}X\beta + 2\beta'X'V^{-1}y\Big\}. \qquad (16)
\end{aligned}$$

Hence by multiplying (14), (15) and (16) together we get (13) as

$$\pi(\gamma|y) = \frac{\exp[-\frac{1}{2}(\gamma - A^{-1}t)'A(\gamma - A^{-1}t) + s]}{(2\pi)^{\frac{1}{2}(p+q)}|A|^{\frac{1}{2}}}$$

9.10.2 Minimum Mean Squared Error Estimation [111, 9.10.2]

Let Ay be the desired estimate. Then the mean squared error is (with $A = A'$)

$$\begin{aligned}
&E(Ay-\gamma)(Ay-\gamma)' \\
&\quad = E(Ayy'A - \gamma y'A - Ay\gamma' + \gamma\gamma') \\
&\quad = E[A(W\gamma+e)(W\gamma+e)'A - \gamma(W\gamma+e)'A - A(W\gamma+e)'\gamma' + \gamma\gamma'] \\
&\quad = E[AW\gamma\gamma'W'A + 2AW\gamma e'A + Aee'A - \gamma\gamma'W'A - \gamma e'A - AW\gamma\gamma' - Ae\gamma' + \gamma\gamma'] \\
&\quad = AW(C+\mu\mu')W'A + 0 + ARA - (C+\mu\mu')W'A - 0 - AW(C+\mu\mu') - 0 + (C+\mu\mu') \\
&\qquad\qquad \text{Write } Q = C + \mu\mu' = Q' \qquad [\text{Recall: } C = \mathrm{var}(\gamma)] \\
&\quad = AWQW'A + ARA - QW'A - AWQ + Q \\
&\quad = A(WQW'+R)A - QW'A - AWQ + Q \\
&\quad = [A - (WQW'+R)^{-1}WQ]'(WQW'+R)[A-(WQW'+R)^{-1}WQ] \\
&\qquad + [Q - Q'W'(WQW'+R)^{-1}WQ]. \qquad (17)
\end{aligned}$$

The second term is $(W'R^{-1}W + Q^{-1})^{-1}$ – which is positive definite. Therefore (17) is minimized by letting

$$A - (WQW'+R)^{-1}WQ = 0, \qquad (18)$$

i.e.,

$$\begin{aligned}
A &= [R^{-1} - R^{-1}W(W'R^{-1}W+Q^{-1})^{-1}W'R^{-1}]WQ \\
&= R^{-1}WQ - R^{-1}W(W'R^{-1}W+Q^{-1})^{-1}(W'R^{-1}W+Q^{-1}-Q^{-1})Q \\
&= R^{-1}WQ - R^{-1}WQ + R^{-1}W(W'R^{-1}W+Q^{-1})^{-1} \\
&= R^{-1}W(W'R^{-1}W+Q^{-1})^{-1}. \qquad (19)
\end{aligned}$$

Therefore

$$(W'R^{-1}W + Q^{-1})A' = W'R^{-1}.$$

This development began with defining A as symmetric. Yet neither (18) nor (19) display this property. Nevertheless, using it, namely $A = A'$ gives

Chapter 10

Quadratic Estimation of Variances

10.1 A general model for variances and covariances [113, 10.1]

The general mixed model as already considered has model equation

$$y = X\beta + Zu + e$$

with

$$\begin{array}{ll} y_{n\times 1} & \text{a vector of data,} \\ \beta_{p\times 1} & \text{a vector of fixed effects,} \\ u_{q\times 1} & \text{a vector of random effects,} \\ X_{n\times p} \text{ and } Z_{n\times q} & \text{known matrices} \end{array}$$

and

$$e_{n\times 1} \quad \text{a vector of random (residual) error terms.}$$

Stochastic properties usually attributed to y, u and e are

$$\begin{array}{lcccl} E(y) = X\beta & E(u) & = & 0 & E(e) = 0 \\ \text{var}(y) = V & \text{var}(u) & = & G & \text{var}(e) = R \end{array}$$

and

$$\text{cov}(u,\ e') = 0.$$

This gives

$$V = ZGZ' + R.$$

10.1.2 Generalizing R

In (10.2) and (10.3) G is generalized through taking $\mathbf{u}' = [\mathbf{u}'_1, \ldots, \mathbf{u}'_b]$, as in (1), with b being the number of random factors. And in the generalization of R in (10.4) and (10.5) similar to that of G, namely as

$$R = \{_m \ \mathbf{R}_{ij} r_{ij}\}_{i,j=1}^{c}, \tag{2}$$

and c is the number of e-vectors. And note that i and j in (2) are not necessarily the same as i and j in (10.2) and (10.3). They cannot be. G has order $q = \sum q_i$ whereas R has order n.

10.1.3 Examples

The first example, starting at [114, bottom] is totally straightforward, except for its last line [115, third line up]. It is not true that "$G_{12}g_{12}$ does not exist." It does exist; it is null, of order 3×5; i.e., $\mathbf{0}_{3\times 5}$.

For the second example (the table at the bottom of [115]), $\mathbf{u}_1$ and $\mathbf{u}_2$ are the sire effects for traits 1 and 2, respectively. So

$$\mathbf{u}' = [\mathbf{u}'_1 \ \ \mathbf{u}'_2] = [u_{11} \ \ u_{12} \ \ u_{21} \ \ u_{22}].$$

$Z_1 u_1$ and $Z_2 u_2$ are as shown on [116]. But we are told that sire 2 is a son of sire 1. Therefore

$$\text{var}(\mathbf{u}) = \begin{bmatrix} G_{11}g_{11} & G_{12}g_{12} \\ G_{21}g_{21} & G_{22}g_{22} \end{bmatrix}$$

where $G_{11} = G_{22} = G_{12} = \begin{bmatrix} 1 & .5 \\ .5 & 1 \end{bmatrix}$, as shown.

The variance of $\mathbf{e}$ is given as

$$R = \begin{bmatrix} I_5 r^*_{11} & R_{12} r^*_{12} \\ R_{21} r^*_{21} & I_3 r^*_{22} \end{bmatrix}$$

where $\begin{bmatrix} r^*_{11} & r^*_{12} \\ r^*_{21} & r^*_{22} \end{bmatrix}$ is described as the variance-covariance matrix for the error terms of the two traits. What *is* this exactly? For a trait 1 observation on animal k and a trait 2 observation on the same animal let the error terms be e_{ik} and e_{2k}, respectively. Then

$$\text{var} \begin{bmatrix} e_{1k} \\ e_{2k} \end{bmatrix} = \begin{bmatrix} r^*_{11} & r^*_{12} \\ r^*_{21} & r^*_{22} \end{bmatrix}.$$

Question On [116] the g_{ij} and r_{ij}-terms have an asterisk. Why? Maybe as an attempt at distinguishing between true parameters and *á priori* values of them.

are other random effects the situation will be more difficult. Also, "adjusted progeny mean" is undefined, but may mean

$$(Z'R^{-1}Z)^{-1}\left[X'R^{-1}(y - X\beta - \text{other random effects})\right].$$

10.5 Form of Quadratics [119, 10.5]

This section is somewhat vague. First, "full model" is undefined; apparently it is $E(y) = W\alpha$. Second, no reason is given for wanting to use OLS (ordinary least squares) for estimating β and u. Third, the definition of W_i in (10.16) is unclear; and finally "reduced model" is also undefined: it appears to be $E(y) = W_1\alpha_1$. The only hint of (10.15) or (10.17) being pertinent to estimating variance components is the line under (10.16), that the reduced model always includes $X\beta$; i.e., it is reduced only by dropping some (or none) of the u_i.

10.6 Expectations of Quadratics [120, 10.6]

Matrix notes Recall that $tr(ABC) = tr(CAB) = tr(BCA)$; and that if $A^2 = A$ it is described as idempotent and its rank and trace are equal.

From $E(y'Qy)$ at the bottom of [117] and top of [118], putting $Q = I$ gives

$$E(y'y) = \sum_{i=1}^{b}\sum_{j=1}^{b}\left[\text{tr}(Z_iG_{ij}Z_j')g_{ij} + \text{tr}(R_{ij}r_{ij})\right] + \beta'X'X\beta. \tag{6}$$

This leads to $E(y'y)$ of [120, (10.20)] only when

$$\begin{aligned} G_{ij} &= 0 \quad \text{or} \quad g_{ij} = 0 \; \forall \; i \neq j \\ R_{ij} &= 0 \quad \text{or} \quad r_{ij} = 0 \; \forall \; i \neq j \\ R_{ii} &= I \quad \text{and} \quad r_{ii} = \sigma_e^2 \; \forall \; i. \end{aligned} \tag{7}$$

Then (6) becomes

$$E(y'y) = \sum_{i=1}^{b}\text{tr}(Z_iG_{ii}Z_i')g_{ii} + n\sigma_e^2 + \beta'X'X\beta. \tag{10.20}$$

And, in traditional variance components models, where $G_{ii} = I_{q_i}$ this becomes

$$E(y'y) = \sum_{i=1}^{b}\text{tr}(Z_iZ_i')g_{ii} + n\sigma_e^2 + \beta'X'X\beta. \tag{10.21}$$

Therefore, using the standard results $X(X'X)^{-}X'X = X$ and $M_X X = 0$,

$$X'W(W'W)^{-}W' = X' \tag{10}$$

and so

$$Z'W(W'W)^{-}W' = Z'X(X'X)^{-}X' + Z'M_X = Z'. \tag{11}$$

Hence in (8) and (10.23)

$$\begin{aligned} &E[y'W(W'W)^{-}W'y] \\ &\quad = \operatorname{tr}[Z'W(W'W)^{-}W'ZG] + r(W)\sigma_e^2 + \beta'X'X\beta \\ &\quad = \operatorname{tr}(Z'ZG) + r(W)\sigma_e^2 + \beta'X'X\beta \end{aligned}$$

and for $G = \{_d \ G_{ii}g_{ii}\}$ this is

$$\sum_{i=1}^{b} \operatorname{tr}(Z_i'Z_i)g_{ii} + r(W)\sigma_e^2 + \beta'X'X\beta = n\sum_{i=1}^{b} g_{ii} + r(W)\sigma_e^2 + \beta'X'X\beta. \tag{10.24}$$

Note in passing that (10) and (11) easily confirm $W'W(W'W)^{-}W'W = W'W$.

For the reduction for the reduced model (10.18) is $(\alpha_1^0)'W_1'y = y'W_1(W_1'W_1)^{-}W_1'y$. Hence from (10.23)

$$\begin{aligned} &E[y'W_1(W_1'W_1)^{-}W_1y] \\ &= \sum_{i=1}^{b} \operatorname{tr}[(W_1'W_1)^{-}W_1'Z_iG_{ii}Z_i'W_1]g_{ii} + r(W_1)\sigma_e^2 + \beta'X'W_1(W_1'W_1)^{-}W_1'X\beta. \end{aligned} \tag{10.25}$$

Following (10.25) we see that X "is included in W_1", meaning that X is a submatrix of W_1; thus for some W_0

$$W_1 = [X \quad W_0]$$

and so from (9)

$$W_1(W_1'W_1)^{-}W_1' = X(X'X)^{-}X' + M_XW_0(W_0'M_XW_0)^{-}W_0'M_X$$

and hence

$$X'W_1(W_1'W_1)^{-}W_1' = X'.$$

This and $G_{ii} = I$ reduces (10.25) to

$$\sum_{i=1}^{b} \operatorname{tr}[(W_1'W_1)^{-}W_1'Z_iZ_i'W_1]g_{ii} + r(W_1)\sigma_e^2 + \beta'X'X\beta. \tag{12}$$

10.8 Henderson's Method 1 [122, 10.8]

Correction In (10.41) the $(Z_i Z_i)^{-1}$ should be $(Z_i' Z_i)^{-1}$.

Clarification In the fourth line of the second paragraph after (10.43) one must presume that the comment "coefficient of σ_i^2" is implicitly referring to the coefficient in (10.41).

It seems to me in the 2-way crossed classification example on pages 123-129 that it is a pity that there is no reference to Henderson's earlier writings (e.g., *Biometrics*, 1953) nor to other people's treatment of this example. For instance in the lower part of [124] the notation Red(ts), Red(t) and so on is not at all clear. It is well known that these calculations are, for example,

$$\text{Red}(ts) = \sum_i \sum_j y_{ij\cdot}^2 / n_{ij}$$

and

$$\text{Red}(t) = \sum_i y_{i\cdot\cdot}^2 / n_{i\cdot}$$

Moreover the more informative notation, based on the model equation

$$y_{ijk} = \mu + t_i + s_j + (ts)_{ij} + e_{ijk} \tag{13}$$

is

$$R(\mu, t, s, ts) = \sum_i \sum_j y_{ij\cdot}^2 / n_{ij} = \sum_i \sum_j n_{ij} \bar{y}_{ij\cdot}^2 = 2037.56$$

and

$$R(\mu, t) = \sum_i y_{i\cdot\cdot}^2 / n_{i\cdot} = \sum_i n_i \bar{y}_{i\cdot\cdot}^2 = 20212.83$$

and then, for example

$$\text{SSAB}^* = \sum_i \sum_j n_{ij} \bar{y}_{ij\cdot}^2 - \sum_i n_{i\cdot} \bar{y}_{\cdot j\cdot}^2 - \sum_j n_{\cdot j} \bar{y}_{\cdot j\cdot}^2 + n_{\cdot\cdot} \bar{y}_{\cdot\cdot\cdot}^2$$

and

$$\text{SSA} = \sum_i n_{i\cdot} (\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot\cdot\cdot})^2.$$

good idea based on $P = I - X(X'X)^{-}X'$ is to form the equations

$$\{_m \; Z_i'PZ_j\}_{i,j=1}^{\;b} \{_c \; \mathbf{u}_i\}_{i=1}^{b} = \{_c \; Z_i'Py\}_{i=1}^{\;b}. \tag{10.67}$$

However, the second line after these equations suggested computing b "reductions from (10.67), and this would be Method 3." This statement gives no hint as to *how* the reductions would be calculated. And it pays no heed to the kind of problem that arises in the 2-way classification: use $R(t|\mu)$ or $R(t|\mu, s)$? Using either (10.67) for calculating reductions in sums of squares, or the D_i-idea in (10.68) really has no appeal. Each is just an example of arbitrarily picking some quadratics for using in the $E(q) = F\sigma^2$ algorithm without any statistical criterion being applied towards determining what quadratics to use. VC 222 addresses this serious weakness of the ANOVA method of estimating variance components.

10.11 Henderson's Method 2 [137, 10.11]

The description given here of Method 2 is considerably different from that given in Henderson (1953) and the extension thereof in VC 190-201.

First, notice the following omissions, presumably taken as accepted.

$$\begin{aligned} y &= X\beta + \sum_i Z_i u_i + e \\ V = \mathrm{var}(y) &= \sum_i Z_i Z_i' \sigma_i^2 + \sigma_e^2 I. \end{aligned}$$

Also, at [137, mid-page], the $(Z_a) =$ rank(Z) should be rank$(Z_a) =$ rank(Z).

To involve

$$P_* = X_a'X_a - X_a'Z_a(Z_a'Z_a)^{-1}Z_a'X_a = X_a'MX_a \quad \text{for} \quad M = I - Z_a(Z_a'Z_a)^{-1}Z_a \tag{14}$$

the inverse coming from (10.79) must be

$$\begin{bmatrix} X_a'X_a & X_a'Z_a \\ Z_a'X_a & Z_a'Z_a \end{bmatrix}^{-1} = \begin{bmatrix} 0 & 0 \\ 0 & (Z_a'Z_a)^{-1} \end{bmatrix} + \begin{bmatrix} I \\ -(Z_a'Z_a)^{-1}Z_a'X_a \end{bmatrix} P_*^{-1}[I \quad -X_a'Z_a(Z_a'Z_a)^{-1}]. \tag{15}$$

Then equations (10.79) yield

$$\begin{bmatrix} \beta_a \\ u_a \end{bmatrix} = \begin{bmatrix} P_*^{-1}X_a'My \\ (Z_a'Z_a)^{-1}Z_a'(y - X_a\beta_a) \end{bmatrix}. \tag{16}$$

and

$$MZ_a = 0 \qquad \text{and} \qquad MZ_b = MZ_aK = 0,$$

from using (14) and (19). Thus

$$y - X_a\beta_a = (I - X_aP_*^{-1}X_a'M)(X_b\beta_b + e) + Z_au_a + Z_bu_b.$$

Hence for $\beta' = [\beta_a' \;\; \beta_b']$

$$\begin{aligned} y - X\beta &= y - X_a\beta_a - X_b\beta_b \\ &= -X_aP_*^{-1}X_a'MX_b\beta_b + Zu + (I - X_aP_*^{-1}X_a'M)e \end{aligned} \tag{21}$$

If the first term of (21) can be written as μ^*1 for some μ^* then (21) has the correct form; it has Zu for the random effects, the same as y, and it has e multiplied by some factor other than I. But does μ^* exist? And is the multiplying factor of e correct? [138] has no comment whatsoever about the model for $y - X\hat{\beta}$ needing a term μ^*1, in contrast to equation (44) of VC 192.

The coefficient of e in (21) certainly does not seem to be in line with (10.86) of [138]. From (21) the coefficient of σ_e^2 in $E(y - X\hat{\beta})'Z_i(Z_i'Z_i)Z_i'(y - X\hat{\beta})$ would be

$$\text{tr}[(I - X_aP_*^{-1}X_a'M)Z_i(Z_i'Z_i)^{-1}Z_i'(I - X_aP_*^{-1}X_a'M)]$$

and there seems to be no way of reducing this to (10.86); but see Henderson, Searle and Schaeffer (1974).

10.12 An Unweighted Means ANOVA [139, 10.12]

A description of this method, more detailed than that on [139-141], is available in VC 219-20. Also available there are details of using the Yates (1934) weighted means analysis of variance.

Both of these Yates' sets of calculations were designed for hypothesis testing for fixed effects models. Using them for estimating variance components in mixed models is just another example of using $E(\mathbf{q}) = F\sigma^2$ to get $\hat{\sigma}^2 = \mathbf{F}^{-1}\mathbf{q}$ without having any substantive statistical reason for using Yates' sums of (or, equivalently mean) squares as elements of $\mathbf{q}$. As already mentioned at the end of Section 10.10, the weaknesses of this kind of ANOVA approach are discussed at VC 222.

Chapter 11

MIVQUE of Variances and Covariances

Warning To me (and others, e.g., VC 398) MIVQUE is not a legitimate estimation procedure. This is because MIVQUE estimators are functions of prior values of ratios σ_i^2/σ_e^2 of the variance components being estimated. Thus people with different prior values will, from the same data, get different estimates. This does not seem reasonable. Also, as with ANOVA estimation, there is no protection against negative estimates.

[143, last line] might seem to imply that (11.1) yields variance components estimates. Not so, of course. Equations (11.1) are the MMEs with solutions

$$\text{BLUE}(\beta) = \beta^0 \quad \text{and} \quad \text{BLUP}(u) = \hat{u} = GZ'\tilde{V}^{-1}(y - X\beta^0). \tag{1}$$

The thrust of this chapter is that parts of the MMEs, notably BLUP(u), can be used for calculating MIVQUE estimates of variances and covariances of subvectors u_i of u.

11.1 The LaMotte Result for MIVQUE [144, 11.1]

The five different classes of estimators discussed by LaMotte (1973) are summarized in VC 393-4. The estimate referred to in (11.5) is Class C_4 on VC 394, described as translation invariant and unbiased. The sentence following (11.5) indicates that the quadratic forms represented there are used just by equating them to their expected values. That is true; but the derivation of this fact, and of (11.5) itself, is not given. This we now do.

which is the i'th term on the left-hand side of (2). Thus (2) can be described as equating the quadratics $y'\tilde{P}Z_iZ_i'\tilde{P}y$ to their expected values. With this in mind [144, 11.2] and [145, 11.3] show how $\hat{u}$ from the MMEs can be used in calculating $y'\tilde{P}Z_iZ_i'\tilde{P}y$. Details of this are developed in Section 11.3.

11.2 Alternatives to LaMotte quadratics [144, 11.2]

This is simple. Representing (2) as $B\hat{\sigma}^2 = q$ with $E(q) = B\sigma^2$, then $\hat{\sigma}^2 = B^{-1}q = (HB)^{-1}HQ$ for any non-singular H. By clever choice of H it may be easier to compute $(HB)^{-1}Hq$ than $B^{-1}q$, and this is the underlying idea for introducing $\hat{u}$.

11.3 Quadratics equal to LaMotte's [145, 11.3]

This shows how (11.5) can be reduced to the form $\hat{u}'Q\hat{u}$ which is used repeatedly in the rest of the chapter. The clue to this is the generalization of $V = \sum_{i=0}^{r} V_i\sigma_i^2$ of (6) to $V = \sum_{i=1}^{k} V_t\theta_t$ for the θ_ts being not just variances as in (16) but covariances also. To use this, recall that

$$\mathbf{u}' = [\mathbf{u}_1' \ldots \mathbf{u}_i' \ldots \mathbf{u}_b'] \qquad \text{and} \qquad \mathbf{e} = [\mathbf{e}_1' \ldots \mathbf{e}_j' \ldots \mathbf{e}_c'].$$

Then $G = \text{var}(\mathbf{u})$ and $R = \text{var}(\mathbf{e})$ can be partitioned respectively into b^2 and c^2 submatrices as in (11.10):

$$G = \{_m \;\; G_{ij}g_{ij}\}_{i,j=1}^{b} \quad \text{and} \quad R = \{_m \;\; R_{ij}r_{ij}\}_{i,j=1}^{c}, \tag{8}$$

with $g_{ji} = g_{ij}$, $r_{ji} = r_{ij}$ and, for $j < i$, $G_{ij} = G_{ji}'$ and $R_{ij} = R_{ji}'$. Now define

$$G_{ij}^* \text{ (and } R_{ij}^*\text{) as } G \text{ (and } R\text{) with all submatrices null except } G_{ij},\ G_{ij}',\ \text{and } R_{ij} \text{ and } R_{ij}'. \tag{9}$$

For example

$$G_{11}^* = \begin{bmatrix} G_{11} & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad G_{12}^* = \begin{bmatrix} 0 & G_{12} & 0 \\ G_{12}' & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{9a}$$

Then

Next, on defining C_i as the i'th column of $\tilde{G}^{-1}$ (see [146], bottom) consider $\tilde{G}^{-1}G^*_{ii}\tilde{G}^{-1}$ for $b = 3$ and $i = 2$, remembering that G is symmetric;

$$\tilde{G}^{-1}G^*_{22}\tilde{G}^{-1} = [C_1 \;\; C_2 \;\; C_3]\begin{bmatrix} 0 & 0 & 0 \\ 0 & G_{22} & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} C_1' \\ C_2' \\ C_3' \end{bmatrix} = [0 \;\; C_2G_{22} \;\; 0]\begin{bmatrix} C_1' \\ C_2' \\ C_3' \end{bmatrix} = C_2G_{22}C_2'.$$

This exemplifies (11.18) for $i = 2$. Similarly, for $i = 2$ and $j = 3$

$$\begin{aligned} \tilde{G}^{-1}G^*_{23}\tilde{G}^{-1} &= [C_1 \;\; C_2 \;\; C_3]\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & G_{23} \\ 0 & G_{23}' & 0 \end{bmatrix}\begin{bmatrix} C_1' \\ C_2' \\ C_3' \end{bmatrix} \\ &= [0 \;\; C_3G_{23}' \;\; C_2G_{23}']\begin{bmatrix} C_1' \\ C_2' \\ C_3' \end{bmatrix} \\ &= C_3G_{23}'C_2' + C_2G_{23}'C_3' \\ &= C_2G_{23}C_3' + C_3G_{23}'C_2', \end{aligned}$$

which exemplifies (11.19) after the correction of adding a prime to the final C_i. The same kind of algebra applies for R.

11.3.1 A simple example

When every $g_{ij} = 0$ for $i \neq j$

$$G = \operatorname{diag}(G_{ii}g_{ii}) \qquad \text{and} \qquad G^{-1} = \operatorname{diag}(G_{ii}^{-1}g_{ii}^{-1}).$$

Then (11.16) is

$$\begin{aligned} &\hat{\mathbf{u}}'[\operatorname{diag}(G_{ii}^{-1}g_{ii}^{-1})]\, G^*_{ii}\, [\operatorname{diag}(G_{ii}^{-1}g_{ii}^{-1})]\mathbf{u} \\ &\quad= \hat{\mathbf{u}}'[\text{a null matrix except for } G_{ii}^{-2}G_{ii}g_{ii}^{-2} \text{ as the } i\text{'th block in the diagonal}]\mathbf{u} \\ &\quad= \hat{\mathbf{u}}_i'G_{ii}^{-1}g_{ii}^{-2}\hat{\mathbf{u}}_i \qquad (12) \end{aligned}$$

as in the 4'th line above (11.20) after correcting the latter to have a subscript i on the final $\hat{u}$.

After (12) the next displayed expression on [147] is introduced as "an alternative is obviously" $\hat{u}_i'G_{ii}^{-1}\hat{u}_i$. One may well wonder why "obviously"? The reason is, as discussed earlier, the usage of these quadratics is to equate them to their expected values, so that multiplication by a scalar does

$$\begin{aligned}
&= y'y - y'X\beta^0 - y'(V-R)V^{-1}(y - X\beta^0) \\
&= y'y - y'X\beta^0 - y'ZGZ'V^{-1}(y - X\beta^0) \\
&= y'y - y'X\beta^0 - y'Z\hat{u}.
\end{aligned}$$

Since $\hat{\sigma}_e^2$ can be estimated as part of MIVQUE why would one want to use the OLS residual mean square atop [148]?

In (11.22) and (11.23) the matrix A is undefined. Presumably it is a genetic relationship matrix, as in [5, 1.2].

11.3.3 Another computation method

A simplification of the MIVQUE equations in (2) leads to a useful computational method which requires only the summing of squared elements of matrices. It is based on

$$\text{tr}(AB) = \text{tr}(BA)$$

and

$$\begin{aligned}
\text{tr}(AA') &= \sum_i \sum_j a_{ij}^2 = \sum (\text{every element of } A)^2 \\
&= \text{sesq}(A),
\end{aligned}$$

so defining "sesq" as "sum of elements squared". Then (2) is

$$\left\{_m \ \text{sesq}(Z_i'\tilde{P}Z_j)\right\}_{i,j} \hat{\sigma}^2 = \left\{_c \ \text{sesq}(Z_i'\tilde{P}y)\right\}_i. \tag{13}$$

11.4 Computation of missing u [149, 11.4]

This consists of but six lines of text. The reference to Chapter 5 is to [48, 5.11]

11.5 Quadratics in $\hat{e}$ with missing observations [150, 11.5]

In [150-1] note that P is neither $I - X(X'X)^-X'$ nor $V^{-1} - V^{-1}X(X'V^{-1}X)^-X'V^{-1}$ as it has been earlier.

11.6 Expectations of quadratics in $\hat{u}$ and $\hat{e}$ [151, 11.6]

The trace results in (11.27)-(11.29) are quite standard.

11.11 Sampling variances [156, 11.11]

Typo: In (11.49), the left-hand Q should be Q_i.

Comment: A basic difficulty with this presentation is that the specific forms of the Q_i and Q_j in (11.49) and (11.50) are not given. This is also true of all the B_i, F_j and H_i matrices; and, of course, the P introduced in [157, line 2] is neither $I - X(X'X)^- X$ nor $V^{-1} - V^{-1}X(X'V^{-1}X)^- X'V^{-1}$. Similar comments also apply to [157, 11.11.1].

In [156, lines 2 and 3 of 11.11] **g** and **r** seem to be introduced without any specific definition. Presumably g is the vector of scalars g_{ij} in $G = \{_m\ G_{ij}g_{ij}\}$ and **r** has the scalars r_{ij} of $R = \{_m\ R_{ij}r_{ij}\}$. However, at [158, line before (11.59)] **r** is defined as the right-hand sides of the mixed model equations.

11.12 Illustrations and simplified models [158-75, 11.12-11.16]

Much of the remainder of the chapter consists of numerical illustrations and simplifications.

[158, 11.12.1]: MIVQUE with $\hat{\sigma}_e^2$ = OLS residual

Then, for approximate MIVQUE [152, 11.7], with three approximate g-inverses

[161, 11.12.2]: diagonal matrix

[162, 11.12.3]: block diagonal matrix

[163, 11.12.4]: triangular block diagonal matrix

[164, 11.13]: See Section 11.13 which follows

[165, 11.14]: illustrates a multivariate model

Two other types of MIVQUE are as follows:

[173, 11.15.1]: an estimator described as not translation invariant and (not un)biased; but the equation given for it does not seem to coincide with any of the LaMotte equations in VC 393-4.

[174, 11.15.2]: a translation invariant and (not un)biased estimator which is Class C_3 and hence C_2 on VC 394. The equivalence of the two expressions is due to

$$(y - X\beta^0)'V^{-1}(y - X\beta^0) = (VPy)'Py = y'Py \quad \text{because} \quad PVP = P.$$

Also from (16)

$$\begin{aligned}\hat{u} &= (Z'R^{-1}Z + G^{-1})^{-1}(Z'R^{-1}y - Z'R^{-1}X\beta^0) \\ &= (Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1}(y - X\beta^0) \qquad (19)\end{aligned}$$

Therefore from (17)

$$\begin{aligned}&\hat{e}'R^{-1}\hat{e} + \sum \alpha_i u_i' G_i^{-1} u_i \\ &= \sigma_e^2 (y - X\beta^0)'V^{-1}(y - X\beta^0) \\ &= \sigma_e^2 (y - X\beta^0)'\frac{1}{\sigma_e^2}\left[R^{-1} - R^{-1}Z(Z'R^{-1}Z + G^{-1}\sigma_e^2)^{-1}Z'R^{-1}\right](y - X\beta^0), \text{ from (17)} \\ &= (y - X\beta^0)'R^{-1}(y - X\beta^0) - (y - X\beta^0)'R^{-1}Z\hat{u}, \text{ from (19)} \\ &= y'R^{-1}y - \beta^{0\prime}X'R^{-1}y - y'R^{-1}Z\hat{u} - y'R^{-1}X\beta^0 + \beta^{0\prime}X'R^{-1}X\beta^0 + \beta^0 X'R^{-1}Z\hat{u} \\ &= y'R^{-1}y - \beta^{0\prime}X'R^{-1}y - \hat{u}'Z'R^{-1}y - \beta^{0\prime}(X'R^{-1}y - X'R^{-1}X\beta^0 - X'R^{-1}Z\hat{u}) \\ &= y'R^{-1}y - [\beta^{0\prime} \quad \hat{u}']\begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix} + 0, \text{ from the first equation in (16).}\end{aligned}$$

Hence

$$\hat{e}'R^{-1}\hat{e} = y'R^{-1}y - (\text{solution vector})'[\text{r.h.s. vector} - \text{of (16)}] - \sum \alpha_i \hat{u}_i' G_i^{-1}\hat{u}_i$$

as preceding [165, 11.14].

Chapter 12

REML and ML Estimation

12.1 An Introduction: ML

The chapter's first sentence is essentially "REML $\cdots$ can be obtained by iterating on MIVQUE". Nothing is said about what REML is (other than what the acronym stands for), nor about its origin and derivation (other than the standard literature reference, Patterson and Thompson, 1971). This is an awkward start for the reader who is not conversant with at least the main underpinnings of maximum likelihood (ML) and restricted maximum likelihood (REML) estimation of variance components in the traditional mixed model. Some of these underpinnings are now presented, with references to VC Chapter 6 which consists of more than twenty pages of detail about these methods of estimation.

12.1.1 A General Model

The starting point for data vector y of order n is

$$y = X\beta + Zu + e, \tag{1}$$

as has already been used, with β being a vector of fixed effects, X and Z known, u a vector of random effects, and e a residual random error. The most general first and second moments attributed to u and e are

$$u \sim (0,\ G), \qquad e \sim (0,\ R) \qquad \text{and} \qquad \operatorname{cov}(u,\ e') = S'. \tag{2}$$

Equating to zero expression (4), and expression (5) for t being in turn each functionally different element of G, R and S yields what are known as the ML equations. Their solutions are the ML solutions; and these are all estimators if they lie in the parameter space, as discussed in [182, 12.7].

It is not difficult to appreciate that equating to zero (4) and all cases of (5) usually results in equations that are not easy to solve. Indeed, they almost always have to be solved by numerical methods, often in the form of iterative procedures.

12.1.3 The Traditional Mixed Model

12.1.3.1 *The model*

The traditional mixed model is typified by its special forms of G, R and S, namely

$$G = \left\{_d\ \sigma_i^2 I_{q_i}\right\}_{i=1}^{r}, \quad R = \sigma_e^2 I_n, \quad \text{and} \quad S = 0.$$

The form of G comes from $\mathbf{u} = \{_c\ \mathbf{u}_i\}_{i=1}^{r}$ where $\mathbf{u}_i$ is a $q_i \times 1$ vector of random effects having $\text{var}(\mathbf{u}_i) = \sigma_i^2 I_{q_i}$. Then with $Z = \{_r\ Z_i\}_{i=1}^{r}$ conformable for Zu (r being the number of random effects factors)

$$V = ZGZ' + \sigma_e^2 I_n = \sum_{i=1} Z_i Z_i' \sigma_i^2 + \sigma_e^2 I_n;$$

and on defining

$$\sigma_0^2 = \sigma_e^2, \qquad Z_0 = I_n \qquad \text{and} \qquad q_0 = n, \tag{6}$$

$$V = \sum_{i=0}^{r} Z_i Z_i' \sigma_i^2. \tag{7}$$

12.1.3.2 *Estimation*

V of (7) means that t of (5) takes just the values σ_i^2; and $\partial V/\partial \sigma_i^2 = Z_i Z_i'$ for $i = 0, 1, \ldots, r$. Using this, the ML equations from (4) and (5) are

$$X'\dot{V}^{-1}X\beta = X'\dot{V}^{-1}y \tag{8}$$

and

$$\text{tr}(\dot{V}^{-1}Z_iZ_i') = (y - X\dot{\beta})'\dot{V}^{-1}Z_iZ_i'\dot{V}^{-1}(y - X\dot{\beta}) \tag{9}$$

for $i = 0, 1, \ldots, r$, as in (20) and (21) of VC 236. The notation of a dot above β and V emphasizes that this vector and matrix have elements for which the ML equations (8) and (9) have to be solved.

12.1.3.3 *Sampling variances*

The large-sample asymptotic dispersion matrix of the ML estimators is

$$\text{var}(\sigma^2_{\text{ML}}) = 2\left[\left\{_m \text{ tr}(V^{-1}Z_iZ_i'V^{-1}Z_jZ_j')\right\}_{i,j=0}^{r}\right]^{-1}. \tag{14}$$

Note that this has V where the matrix on the left of (11) has $\dot{V}$. Derivation of (14) can be found in VC Section 6.3a.

12.2 REML

12.2.1 The Concept

Restricted maximum likelihood (REML) estimation can be described in several different ways. The simplest is to think of it as maximum likelihood on linear combinations $k'y$ of the observations in y, with k' being chosen so that $k'y$ contains no β. This means that k' is such that $k'X = 0$. Since there are many vectors k' of this nature the method is, in fact, based on $K'X = 0$, where the rows of K' are a collection of such vectors k'. Those rows are chosen to be linearly independent, and there are as many of them as possible, namely $n-\rho$ for $\rho = \text{rank}(X)$. Then REML estimation is estimation from applying maximum likelihood to $K'y$, where y is taken as being normally distributed, exactly as with ML in Section 12.1.

The name 'error contrast' is sometimes associated with each row of $K'y$ arising from the fact that $y = X\beta + Zu + e$ along with $K'X = 0$ gives $K'y = K'Zu + K'e$ having expectation zero (and not involving β).

12.2.2 REML for the General Model

With

$$y \sim \mathcal{N}(X\beta,\ V)$$

having $K'X = 0$ gives

$$K'y \sim \mathcal{N}(0,\ K'VK).$$

Therefore the likelihood function for $K'y$ is

$$L(V|K'y) = \exp -\frac{1}{2}y'K(K'VK)^{-1}K'y/(2\pi)^{\frac{1}{2}(N-\rho)}|K'VK|^{\frac{1}{2}}. \tag{15}$$

which can also be expressed as

$$\left\{_m \operatorname{sesq}(Z_i'\dot{P}Z_j)\right\}_{i,j=0}^{r} \hat{\sigma}^2_{\text{REML}} = \left\{_c \operatorname{sesq}(Z_i'\dot{P}y\right\}_{i=0}^{r}. \tag{21}$$

And finally the large-sample asymptotic dispersion matrix is

$$\operatorname{var}(\hat{\sigma}^2_{\text{REML}}) = 2\left[\left\{_m \operatorname{tr}(Z_i'PZ_jZ_j'PZ_i)\right\}_{i,j=0}^{r}\right]^{-1}. \tag{22}$$

12.2.4 Points of interest

12.2.4.1 *Differences from ML* The three equivalent forms of the estimation equations for REML, namely (19), (20) and (21), differ from those for ML, (10), (12) and (13) only by the left-hand side of the equation having a P for REML where there is a V^{-1} for ML. The right-hand sides are the same for REML and ML. And for the dispersion matrix the P in the REML case, (22), replaces the V^{-1} in the ML case, (14).

12.2.4.2 *No matrix K* The easiest understanding of REML stems from the concept of applying maximum likelihood to $K'y$ for a K' such that $K'X = 0$; and there are many such matrices K'. Despite this, it is a noticeable feature of the expressions (19), (20) or (21) for calculating REML that none of them specifically involve a K. This is because whenever K occurs it is only in the form $K(K'VK)^{-1}K'$ which, as in (17), is $P = V^{-1} - V^{-1}X(X'V^{-1}X)^{-}X'V^{-1}$.

12.2.4.3 *Balanced data* An interesting feature of REML for all cases of balanced data from mixed models is that REML solutions [of equation (19) or, equivalently, (20) or (21)] are the same as ANOVA estimators – and this is so whether normality is assumed or not (see Anderson, 1978, pages 97-104). This is an appealing result because ANOVA estimators from balanced data have optimal minimum variance properties. Thus there is some comfort in knowing that REML solutions reduce to having these properties for balanced data. But this result is only for REML solutions and not for REML estimators. The estimators are never negative whereas the solutions can be – as can ANOVA estimators.

12.2.4.4 *Degrees of freedom* Consider data that are a simple random sample $x_1, \cdots, x_n$ identically and independently distributed $\mathcal{N}(\mu, \sigma^2)$. With $\bar{x} = \sum_i x_i/n$, the ML estimator of σ^2 is $\sum_i(x_i - \bar{x})^2/n$, whereas the REML estimator is $\sum_i(x_i - \bar{x})^2/(n-1)$. This is the simplest example

Writing computer programs for ML or REML estimation is not, in my opinion, a task for the amateur programmer. Some difficult questions which need to be addressed are the following, as listed in VC Section 6.4.

(i) What method of iteration is best?

(ii) Does the choice of iterative method depend on the form of the equations used?

(iii) Is the most succinct and easily understood form of the estimation equations the best for computational purposes?

(iv) Is convergence of the iteration always assured?

(v) If convergence is achieved, is it always at a global maximum of the likelihood and not just a local maximum?

(vi) Do initial starting values for the iteration affect the value at which convergence is achieved?

(vii) If so, is there any particular set of starting values that always yields convergence at the global maximum of the likelihood?

(viii) What is the cost, in terms of computer time and/or money to do the necessary computing?

(ix) The matrix V is, by definition, always non-negative definite; and usually positive definite. The latter has been assumed. What, therefore, is to be done numerically if, at some step in the iteration, the calculated V is not positive definite?

(x) More seriously, what is to be done if the calculated V is singular? [Harville (1977) addresses this concern.]

(xi) Since ML estimators, as distinct from just solutions to the estimation equations, must satisfy $\tilde{\sigma}_e^2 > 0$ and $\tilde{\sigma}_i^2 \geq 0$ for $i = 1, \cdots, r$, these conditions must be taken into account in computer programs that are used for solving the ML equations to obtain ML estimators. Customarily, any $\tilde{\sigma}_i^2$ that is computed as a negative value is put equal to zero – an action which can sometimes be interpreted as altering the model being used. It also raises the further difficulty

In the traditional mixed model G and R are diagonal, leading to (7). In the general model they are not diagonal as at the bottom of [177]. But presumably the adaptation of V to (9) in Section 11.3 could be invoked to utilize the connection of REML to I-MIVQUE (iterative MIVQUE).

The paragraph preceding [175, 12.2] is a little misleading in its discussion of unbiasedness because neither ML nor REML estimators are unbiased, and least of all when estimated values have been calculated by iteration or other numerical procedure.

12.5 An Alternative Algorithm for REML [178, 12.2]

Equation (12.1) is (5.38) – derived in this Supplement in Section 5.10. And (12.2) is derived in Section 5.11. Note that

$$C = \begin{bmatrix} C_{00} & C_{01} \\ C_{10} & C_{11} \end{bmatrix} \text{ here is } C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \text{ in (5.33).}$$

Typo: In [178, last line] "locks" should be "blocks".

A typical one of the first b equations of (12.5) is

$$\text{tr}(Q_i G) = \hat{u}' Q_i \hat{u} + \text{tr}(Q_i C_{11}),$$

which is

$$\hat{u}' Q_i \hat{u} = \text{tr}[(Q_i(G - C_{11})] = \text{tr}[Q_i \,\text{var}(\hat{u})] = E(\hat{u}' Q_i \hat{u})$$

because $E(\hat{u}) = 0$. Thus (12.5) represents equating these quadratic forms to their expected values, in exactly the manner described at the end of Section 11.1.

Much of this attention to computing algorithms, at least for the traditional mixed model, seems redundant now that SAS Proc MIXED is available.

12.6 ML Estimation [179, 12.3]

A presentation of BLUP(u) as a Bayes estimator of u is in VC Section 7.6d.

[179, penultimate line] describes Q_j as quadratics. Surely, they are matrices from quadratic forms, not the forms themselves.

which is (12.7) without σ_e^2. Probably the phrase "if $\alpha = \sigma_e^2/\sigma_i^2$" in the line below (12.6) explains the occurrence of σ_e^2 in (12.7) – though I doubt it.

12.9 Biased Estimation with Few Iterations [180, 12.6]

A "small simulation" cannot illustrate anything except itself.

12.10 The Problem of Finding Permissible Estimators [182, 12.7]

This is an excellent discussion. My only criticism is in [184, line 2]: why does it "make no sense" to add a negative value to a diagonal element of $Z'R^{-1}Z$ in the MMEs? The MMEs add $(\sigma_e^2/\sigma_\alpha^2)I$ to $Z'R^{-1}Z$ and neither σ_e^2 nor σ_α^2, the true population values, are negative. But if those values are unknown what does one do? Estimate them. $\hat{\sigma}_e^2$ cannot be negative, but $\hat{\sigma}_\alpha^2$ from ANOVA (or MIVQUE) can be. So does one use that negative value in $(\hat{\sigma}_e^2/\hat{\sigma}_\alpha^2)I$? No. It may give V as not n.n.d. And putting $\hat{\sigma}_\alpha^2$ to zero makes $(\hat{\sigma}^2/\hat{\sigma}_\alpha^2)I$ nonsense when added to $Z'R^{-1}Z$. Maybe (suggests L. Schaeffer) using $|\hat{\sigma}_\alpha^2|$ is appropriate. I doubt it, because for negative $\hat{\sigma}_\alpha^2$ we have $|\hat{\sigma}_\alpha^2| = -\hat{\sigma}_\alpha^2$, and that is not very reasonable.

12.11 Method for Singular G [184, 12.8]

This seems incomplete.

Chapter 13

Effects of Selection

13.1 Introduction [185, 13.1]

Easy reading.

13.2 An Example of Selection [186, 13.2]

The first 6×6 matrix is the variance-covariance matrix of the vector of means, the $\bar{y}_{ij\cdot}$ terms. The variances are four terms of $1+15/10 = 2.5$, one of $1+15/500 = 1.03$ and the last is $1+15/100 = 1.15$. The two covariances are $\text{cov}(\bar{y}_{11\cdot}, \bar{y}_{12\cdot}) = \text{cov}(\bar{y}_{21\cdot}, \bar{y}_{22\cdot}) = \sigma_s^2 = 1$. Note that the $\bar{y}_{ij\cdot}$ terms are not in lexicon order.

On [186], I have no idea how the numbers in the last two displays were derived. Scanning the two references in the first line of [187] did not help.

The numbers in the first display of [187] are simply those on the left side of the second-to-last display of [186].

The numbers in the second display on [187] are as follows,

$$\begin{bmatrix} n_{\cdot 1} = 40 & 0 & n_{11} = 10 & n_{21} = 10 & n_{31} = 10 & n_{41} = 10 \\ & n_{\cdot 2} = 600 & n_{12} = 500 & n_{22} = 100 & 0 & 0 \\ & & n_{1\cdot} + 15 = 525 & 0 & 0 & 0 \\ & & & n_{2\cdot} + 15 = 125 & 0 & 0 \\ & & & & n_{3\cdot} + 15 = 25 & 0 \\ & & & & & n_{4\cdot} + 15 = 25 \end{bmatrix},$$

Now suppose selection on w is such that

$$w \text{ becomes } \sim \mathcal{N}(s,\ H_s) \sim \mathcal{N}[d + (s - d),\ H_s]. \tag{7}$$

Then, on (6) and (7) being special cases of (1) and (2), respectively, with

$$v_1 = \begin{bmatrix} y \\ u \end{bmatrix} \qquad \text{and} \qquad v_2 = w,$$

we find from (3)

$$E\begin{bmatrix} y \\ u \\ w \end{bmatrix} = E\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} \mu_1 + C_{12}C_{22}^{-1}k \\ \mu_2 + k \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} X\beta \\ 0 \end{bmatrix} + \begin{bmatrix} B \\ B_u \end{bmatrix} H^{-1}[s-d] \\ d + (s-d) \end{bmatrix} \tag{8}$$

with

$$H_0 = H^{-1}(H - H_s)H^{-1} \tag{9}$$

comparable to (4). Thus

$$E\begin{bmatrix} y \\ u \\ w \end{bmatrix} = \begin{bmatrix} X\beta + Bt \\ B_u t \\ s \end{bmatrix} \tag{13.6}$$

for

$$t = H^{-1}(s - d). \tag{10}$$

In similar manner (3) yields

$$\begin{aligned} \text{Var}\begin{bmatrix} y \\ u \\ w \end{bmatrix} &= \begin{bmatrix} \begin{bmatrix} V & ZG \\ GZ' & G \end{bmatrix} - \begin{bmatrix} B \\ B_u \end{bmatrix} H_0[B' \quad B_u'] & \begin{bmatrix} B \\ B_u \end{bmatrix} H^{-1}H_s \\ H_sH^{-1}[B' \quad B_u'] & H_s \end{bmatrix} \\ &= \begin{bmatrix} V - BH_0B' & ZG - BH_0B_u' & BH^{-1}H_s \\ GZ' - B_uH_0B' & G - B_uH_0B_u' & B_uH^{-1}H_s \\ H_sH^{-1}B' & H_sH^{-1}B_u' & H_s \end{bmatrix}. \end{aligned} \tag{13.7}$$

Corrections. B_u' in the (1,2) submatrix, and B_u in the (2,1) submatrix are shown in (13.7) of [188] as B' and B, respectively.

Minimizing (14) w.r.t. the Lagrange multipliers θ and φ leads to (13); and w.r.t. b yields

$$\begin{aligned} (V - BH_0B')b - (ZG - BH_0B'_u)m &- BH^{-1}H_sf + X\theta + B\varphi = 0 \\ Vb - BH_0(B'b - B'_um) - ZGm &- BH^{-1}H_sf + X\theta + B\varphi = 0. \end{aligned}$$

Substituting for $B'b$ from (13), and for H_0 from (9) gives

$$Vb - BH^{-1}(H - H_s)H^{-1}Hf - ZGm - BH^{-1}H_sf + X\theta + B\varphi = 0$$

and this reduces to

$$Vb + X\theta + B\varphi = ZGm + Bf. \tag{15}$$

This and the two equations in (13) have to be solved for b. They can be arrayed in matrix-vector form as

$$\begin{bmatrix} V & X & B \\ X' & 0 & 0 \\ B' & 0 & 0 \end{bmatrix} \begin{bmatrix} b \\ \theta \\ \varphi \end{bmatrix} = \begin{bmatrix} ZGm + Bf \\ k \\ B'_um + H'f \end{bmatrix}, \tag{16}$$

which is (23) of Henderson (1975a).

To solve (16) observe that

$$\begin{aligned} \begin{bmatrix} V & X & B \\ X' & 0 & 0 \\ B' & 0 & 0 \end{bmatrix}^- &= \begin{bmatrix} V^{-1} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} -V^{-1}(X\ \ B) \\ I \end{bmatrix} S^- \left[-\binom{X'}{B'} V^{-1} \quad I \right] \\ &= \begin{bmatrix} V^{-1} - V^{-1}(X\ \ B)S^-\binom{X'}{B'}V^{-1} & V^{-1}(X\ \ B)S^- \\ S^-\binom{X'}{B'}V^{-1} & -S^- \end{bmatrix} \end{aligned} \tag{17}$$

for

$$S = \begin{bmatrix} X'V^{-1}X & X'V^{-1}B \\ B'V^{-1}X & B'V^{-1}B \end{bmatrix}. \tag{18}$$

Therefore, with b being the first row of (17) post-multiplied by the right-hand side of (16) we get

$$\begin{aligned} b'y &= \left[\left\{ V^{-1} - V^{-1}[X\ \ B]S^-\binom{X'}{B'}V^{-1} \right\} (ZGm + Bf) + V^{-1}[X\ \ B]S^- \begin{pmatrix} k \\ B'_um + H'f \end{pmatrix} \right]' y \\ &= (m'GZ' + f'B') \left[V^{-1}y - V^{-1}(X\ \ B)S^-\binom{X'}{B'}V^{-1}y \right] + [k' \quad m'B_u + f'H]S^- \begin{bmatrix} X' \\ B' \end{bmatrix} V^{-1}y. \end{aligned}$$

But with the definitions of X_* and β_* given in (23), equations (25) are precisely those of (13.8). Thus β^0 and t^0 in (19) and (21) are the same as $\hat{\beta}$ and $\hat{t}$ of (22). As a result, from the third equation of (22)

$$Z'R^{-1}X\beta^0 + Z'R^{-1}Bt^0 + (Z'R^{-1}Z + G^{-1})v^0 = Z'R^{-1}y,$$

so that

$$Z'R^{-1}(y - X\beta^0 - Bt^0) = (Z'R^{-1}Z + G^{-1})v^0.$$

Therefore in (21)

$$m'GZ'V^{-1}(y - X\beta^0 - Bt^0) = m'v^0. \tag{26}$$

Consequently, on substituting (20) and (26) into (19)

$$b'y = k'\beta^0 + (m'B_u + f'H)t^0 + m'v^0, \tag{27}$$

for

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z & X'R^{-1}B \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} & Z'R^{-1}B \\ B'R^{-1}X & B'R^{-1}Z & B'R^{-1}B \end{bmatrix} \begin{bmatrix} \beta^0 \\ v^0 \\ t^0 \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \\ B'R^{-1}y \end{bmatrix}, \tag{28}$$

which is exactly (22) with second and third rows (and columns) interchanged, and with β^0 and t^0 replacing (but equal to) $\hat{\beta}$ and $\hat{t}$, respectively.

Now pre-multiply each side of (28) by

$$P = \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & -B_u' & I \end{bmatrix}$$

and use $B = ZB_u + B_e$ given below (5), i.e.,

$$B' - B_u'Z' = B_e'.$$

This changes (28) to be

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z & X'R^{-1}B \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} & Z'R^{-1}B \\ B_e'R^{-1}X & B_e'R^{-1}Z - B_u'G^{-1} & B_e'R^{-1}B \end{bmatrix} \begin{bmatrix} \beta^0 \\ v^0 \\ t^0 \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \\ B_e'R^{-1}y \end{bmatrix}. \tag{29}$$

Then, with

$$P' = \begin{bmatrix} I & 0 & 0 \\ 0 & I & -B_u \\ 0 & 0 & I \end{bmatrix} \quad \text{and} \quad P'^{-1} = \begin{bmatrix} I & 0 & 0 \\ 0 & I & B_u \\ 0 & 0 & I \end{bmatrix}$$

$$
\begin{aligned}
E_{11} &= T = (X'V^{-1}X)^{-} \\
E_{12} &= -(X'V^{-1}X)^{-}X'R^{-1}RV^{-1}ZG = -(X'V^{-1}X)^{-}X'V^{-1}ZG \\
E_{22} &= W^{-1} + W^{-1}Z'R^{-1}X(X'V^{-1}X)^{-}X'R^{-1}ZW^{-1} \\
&= G - GZ'V^{-1}ZG + GZ'V^{-1}RR^{-1}X(X'V^{-1}X)^{-}X'R^{-1}RV^{-1}ZG \\
&= G - GZ'[V^{-1} - V^{-1}X(X'V^{-1}X)^{-}X'V^{-1}]ZG \\
&= G - GZ'PZG
\end{aligned}
$$

for

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

$$
\begin{aligned}
[X \quad Z]E &= [XE_{11} + ZE_{12}' \quad XE_{12} + ZE_{22}] \\
&= [A_1 \quad A_2]
\end{aligned}
$$

for

$$
\begin{aligned}
A_1 &= X(X'V^{-1}X)^{-} - ZGZ'V^{-1}X(X'V^{-1}X)^{-} \\
&= (V - ZGZ')V^{-1}X(X'V^{-1}X)^{-} \\
&= RV^{-1}X(X'V^{-1}X)^{-}
\end{aligned}
$$

and

$$
\begin{aligned}
A_2 &= -X(X'V^{-1}X)^{-}X'V^{-1}ZG + ZG - ZGZ'PZG \\
&= \{-X(X'V^{-1}X)^{-}X'V^{-1} + I - (V - R)P\}ZG \\
&= RPZG.
\end{aligned}
$$

Since $(X'V^{-1}X)^{-}$ occurs so often we use T, as in (33), so then

$$\begin{aligned}
A_3 &= B_e'R^{-1}(XE_{12}') - B_u'G^{-1}E_{12}' \\
&= B_e'R^{-1}A_1 + B_u'G^{-1}GZ'V^{-1}XT \\
&= B_e'R^{-1}RV^{-1}XT + B_u'Z'V^{-1}XT \\
&= (B_e' + B_u'Z')V^{-1}XT \\
&= B'V^{-1}XT, \text{ from (30);}
\end{aligned}$$

and

$$\begin{aligned}
A_4 &= B_e'R^{-1}(XE_{12} + ZE_{22}) - B_u'G^{-1}E_{22} \\
&= B_e'R^{-1}A_2 - B_u'G^{-1}(G - GZ'PZG) \\
&= B_e'R^{-1}RPZG - B_u'(I - Z'PZG) \\
&= B_e'PZG - B_u' + B_u'Z'PZG \\
&= B'PZG - B_u', \text{ using (30).}
\end{aligned}$$

Hence

$$M_{12}'M_{11}^{-} = [B'V^{-1}XT \quad (B'PZG - B_u')]. \tag{43}$$

Then

$$M_{12}'M_{11}^{-}M_{12} = [B'V^{-1}XT \quad (B'PZG - B_u')] \begin{bmatrix} X'R^{-1}B_e \\ Z'R^{-1}B_e - G^{-1}B_u \end{bmatrix}.$$

In making this product use

$$XTX' = V - VPV \qquad \text{and} \qquad ZGZ' = V - R.$$

This gives

$$\begin{aligned}
M_{12}'M_{11}^{-}M_{12} &= B'V^{-1}(V - VPV)R^{-1}B_e + B'P(V - R)R^{-1}B_e - B'PZB_u \\
&\quad - B_u'Z'R^{-1}B_e + B_u'G^{-1}B_u \\
&= (B' - B_u'Z')R^{-1}B_e + B_u'G^{-1}B_u - B'P(B_e + ZB_u) \\
&= B_e'R^{-1}B_e + B_u'G^{-1}B_u - B'PB. \qquad (44)
\end{aligned}$$

Now note that (39) gives

$$\begin{aligned}(B'_u - B'PZG)Z' &= B'_uZ' - B'P(V-R)\\ &= B'_uZ' - B' + B'V^{-1}XTX' + B'PR\end{aligned}$$

and so

$$\begin{aligned}\beta^0 &= \Big\{TX'[I - V^{-1}(V-R)]\\ &\quad + TX'V^{-1}B(B'PB)^-[B'V^{-1}(V-VPV) - B'_uZ' + B'P(V-R) - B'_e]\Big\}R^{-1}y\\ &= \{TX'V^{-1}R + TX'V^{-1}B(B'PB)^-(-B'PR)\}R^{-1}y, \qquad (49)\end{aligned}$$

since $B = ZB_u + B_e$. Thus

$$\beta^0 = TX'V^{-1}y - TX'V^{-1}B(B'PB)^-B'Py.$$

Therefore, by making use of $PVP = P$ and $X'P = 0$

$$\begin{aligned}\text{var}(\beta^0) &= TX'V^{-1}VV^{-1}XT + TX'V^{-1}B(B'PB)^-B'PVPB(B'PB)^-B'V^{-1}XT\\ &\quad - TX'V^{-1}VPB(B'PB)^-B'V^{-1}XT - TX'V^{-1}B(B'PB)^-B'PVV^{-1}XT\\ &= TX'V^{-1}XT + TX'V^{-1}B(B'PB)^-B'PB(B'PB)^-B'V^{-1}XT\\ &= T + TX'V^{-1}B(B'PB)^-B'V^{-1}XT\\ &= C_{11},\ \text{from (47a) and as in (13.11).} \qquad (50)\end{aligned}$$

Similarly, from (13.9)

$$\begin{aligned}u^0 &= (C'_{12}X' + C_{22}Z' + C_{23}B'_e)R^{-1}y\\ &= \{-GZ'V^{-1}XTX' - (B_u - GZ'PB)(B'PB)^-B'V^{-1}XTX'\\ &\quad + (G - GZ'PZG)Z' + (B_u - GZ'PB)(B'PB)^-(B'_u - B'PZG)Z'\\ &\quad + (B_u - GZ'PB)(B'PB)^-B'_e\}R^{-1}y\\ &= \{-GZ'V^{-1}(V - VPV) + GZ'[I - P(V-R)]\\ &\quad + (B_u - GZ'PB)(B'PB)^-[-B'V^{-1}(V-VPV) + B'_uZ' - B'P(V-R) + B'_e]\}R^{-1}y\\ &= \{GZ'PR + (B_u - GZ'PB)(B'PB)^-B'PR\}R^{-1}y\\ &= GZ'Py - (B_u - GZ'PB)(B'PB)^-B'Py. \qquad (51)\end{aligned}$$

Therefore

$$\text{cov}[\beta^0,\ (u^0-u)'] = -TX'V^{-1}ZG + TX'V^{-1}B(B'PB)^{-}(B_u' + B'PZG) = C_{12} \tag{54}$$

as in (13.12). But

$$\begin{aligned}\text{var}(u^0-u) &= \text{var}(u^0) + G - 2\text{cov}(u^0,\ u') \\ &= C_{22} + 2GZ'PZG - G + 2C_{23}B'PZG + G \\ &\quad -2[GZ'PZG + (B_u - GZ'PB)(B'PB)^{-}B'PZG],\end{aligned} \tag{55}$$

after using (52) and $\text{cov}(y,\ u') = ZG$ with (51). Then (55), with the help of (47e), reduces to C_{22} of (13.13).

13.4.6 Summary

So, in summary we have derived

$$\beta^0 = (X'V^{-1}X)^{-}X'V^{-1}y - (X'V^{-1}X)^{-}X'V^{-1}B(B'PB)^{-}B'Py, \tag{49}$$

$$u^0 = GZ'Py + (B_u - GZ'PB)(B'PB)^{-}B'Py \tag{51}$$

and

$$\text{var}(\beta^0) = C_{11}, \quad \text{from (50)}, \tag{13.11}$$

$$\text{cov}[\beta^0,\ (u^0-u)'] = C_{12}, \quad \text{from (54)}, \tag{13.12}$$

$$\text{var}(u^0-u) = C_{22} \tag{13.13}$$

$$\text{cov}(\beta^0,\ u^0) = C_{13}B_u', \quad \text{from (53)}, \tag{13.14}$$

but for

$$\text{var}(u^0) = G - C_{22} + C_{23}B_u' + B_uC_{23}' - B_uH_oB_u' \tag{13.15}$$

we have (13.15) without its $-B_uH_oB_u'$ term, from (52).

Then it is a standard result that

$$E(y_2|y_1) = X_2\beta + V_{21}V_{11}^{-1}(y_1 - X_1\beta).$$

But in place of y_1 we are dealing with $M'y_1$ where [194, line 3]

$$E(M'y_1) - M'X_1\beta = t.$$

Thus

$$E(y_2|M'y_1) = X_2\beta + \text{cov}[y_2,\ (M'y_1)'](M'V_{11}M)^{-1}(M'y - M'X_1\beta).$$

But

$$\text{cov}(y_2,\ y_1') = \text{cov}(Z_2u,\ u'Z_1') + \text{cov}(e_2,\ e_1') = Z_2GZ_1' + R_{12}'.$$

Hence, for $R_{12} = 0$,

$$\begin{aligned} E(y_2|M'y) &= X_2\beta + Z_2GZ_1'M(M'V_{11}M)^{-1}t \\ &= X_2\beta + Z_2GZ_1'Mk \end{aligned}$$

for k on [194, line 2].

The sentence below [194, (13.24)] also deserves verification. With $M = I$ and Z_1 non-singular the third equation in (13.24) is

$$Z_1GZ_1'\theta = Z_1u^0.$$

Hence the second equation of (13.24) becomes

$$Z_2'R_{22}^{-1}X_2\beta^0 + (Z_2'R_{22}^{-1}Z_2 + G^{-1})u^0 - Z_1'(Z_1GZ_1')^{-1}Z_1u^0 = Z_2'R_{22}^{-1}y$$

which, very simply for Z_1^{-1} existing (which makes no sense), reduces to

$$Z_2'R_{22}^{-1}X_2\beta^0 + Z_2'R_{22}^{-1}Z_2u^0 = Z_2'R_{22}^{-1}y,$$

which is the second equation in (13.23). But Z_1^{-1} existing is nonsense.

The last paragraph of [194] has, for me, little practical value so far as using selection in estimating β and u is concerned.

$$S = \begin{bmatrix} X'V^{-1}X & X'L \\ L'X & L'VL \end{bmatrix}^- = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \tag{57}$$

$$= \begin{bmatrix} (X'V^{-1}X)^{-1} & 0 \\ 0 & (L'VL)^{-1} \end{bmatrix} \text{ if } L'X = 0. \tag{58}$$

Then the matrix is (13.35) is the same as that in (17) but with VL used for B. Thus

$$\begin{bmatrix} V & X & VL \\ X' & 0 & 0 \\ L'V & 0 & 0 \end{bmatrix}^- = \begin{bmatrix} V^{-1} - V^{-1}(X \quad VL)S^- \begin{bmatrix} X' \\ L'V \end{bmatrix} V^{-1} & V^{-1}[X \quad VL]S^- \\ S^- \begin{bmatrix} X' \\ L'V \end{bmatrix} V^{-1} & -S^- \end{bmatrix} \tag{59}$$

for S in (57). Now with V_0 as the leading term let

$$T^- = \begin{bmatrix} V_0 & X & VL \\ X' & 0 & 0 \\ L'V & 0 & 0 \end{bmatrix}^- = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix}. \tag{60}$$

From the definition of generalized inverse we know that

$$\begin{bmatrix} V_0 & X & VL \\ X' & 0 & 0 \\ L'V & 0 & 0 \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} \begin{bmatrix} V_0 & X & VL \\ X' & 0 & 0 \\ L'V & 0 & 0 \end{bmatrix} = \begin{bmatrix} V_0 & X & VL \\ X' & 0 & 0 \\ L'V & 0 & 0 \end{bmatrix}; \tag{61}$$

and so

$$\begin{bmatrix} V_0A_{11} + XA_{21} + VLA_{31} & V_0A_{12} + XA_{22} + VLA_{32} & V_0A_{13} + XA_{23} + VLA_{33} \\ X'A_{11} & X'A_{12} & X'A_{13} \\ L'VA_{11} & L'VA_{12} & L'VA_{33} \end{bmatrix} T = T.$$

Therefore, on equating the six different (apart from transposed) submatrices

$$\begin{aligned}
(V_0A_{11} + XA_{21} + VLA_{31})V_0 &+ (V_0A_{12} + XA_{22} + VLA_{32})X' + (V_0A_{13} + XA_{23} + VLA_{33})L'V = V_0 \\
(V_0A_{11} + XA_{21} + VLA_{31})X &= X \\
(V_0A_{11} + XA_{21} + VLA_{31})VL &= VL \\
X'A_{11}X &= 0 \\
X'A_{11}VL &= 0 \\
L'V_{11}VL &= 0
\end{aligned} \tag{62}$$

Now when $V_0 = V$ the A_{ij}-submatrices in (60) come from (59). But we want to show that using those A_{ij}-submatrices with $V_0 = V_s$ gives (61) with $V_0 = V_s$. Before doing that we note a correction to, and a query about, the expression for $V_s = \text{var}(y|L'y)$ on [199].

In these terms we see that the only A_{ij}-matrices involved are A_{11}, A_{12} and A_{13}. And from comparing (60) and (59), and using (58) for S^- in (59), these are

$$\begin{aligned} A_{11} &= V^{-1} - V^{-1}[X(X'V^{-1}X)^{-}X' + VL(L'VL)^{-}L'V]V^{-1} \\ A_{12} &= V^{-1}X(X'V^{-1}X)^{-1} \\ A_{13} &= L(L'VL)^{-1}. \end{aligned} \tag{66}$$

So now with

$$V_s - V = -VLKL'V = (V_s - V)'$$

from (64), the effect of using V_s rather than V in (59) is to use $V_s - V$ of (64) for V_0 in (65) with the A-matrices of (66). In doing so observe that

$$\begin{aligned} L'VA_{11} &= L' - L'X(X'V^{-1}X)^{-} - L' = 0, \quad \text{because} \quad L'X = 0 \\ L'VA_{12} &= L'X(X'V^{-1}X) = 0 \\ L'VA_{13} &= I \end{aligned}$$

Therefore we take

$$\begin{aligned} V_0A_{11} &= (V_s - V)A_{11} = -VLKL'VA_{11} = 0 \\ V_0A_{12} &= (V_s - V)A_{12} = -VLKL'VA_{12} = 0 \\ V_0A_{13} &= (V_s - V)A_{13} = -VLKL'VA_{13} = -VLK. \end{aligned}$$

Hence taking the Δs in reverse order

$$\begin{aligned} \Delta_3 &= 0 \\ \Delta_2 &= 0 \\ \Delta_1 &= 0 + 0 + VL(-VLK)' + 0 + (-VLK)L'V \\ &= -2VLKL'V \\ &= 2(V_s - V). \end{aligned}$$

Δ_1 should be $V_s - V$ in order to have $V + \Delta_1$ be V_s. So there is an error: either CRH's statement [200. lines 1-2] about C_{11}. C_{12} and C_{13} (our A_{11}, A_{12}, A_{13}) is incorrect or the derivation of Δ_1 is wrong.

Chapter 14

Restricted Best Linear Prediction

14.1 Restricted Selection Index [203, 14.1]

Read the Kempthorne-Nordskog reference.

14.2 Restricted BLUP [204, 14.2]

Derivation of (14.2) almost assuredly proceeds in the same manner as that of (13.9) and (13.36) and other equations in Chapter 13. But I have a question.

Question How does one utilize the restriction "expected value of $C'u$ given $a'y = 0$"? Presumably that is $E(C'u|a'y = 0)$. But, under normality,

$$\begin{pmatrix} u \\ y \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} 0 \\ X\beta \end{pmatrix}, \begin{pmatrix} G & GZ' \\ G & V \end{pmatrix}\right]$$

$$E[C'u|a'y] = C'GZ'a(a'Va)^{-1}a'(y - X\beta)$$

and if this is to be 0 then either we want $Z'a$ or $a'(y - X\beta)$ to be 0, neither of which seem workable. Moreover, if a is to be [204, line 4 of Section 14.2] "chosen so that $a'y$ is invariant to β," then that means having $a'X = 0$; and that would seem to negate the desire of $a'y$ predicting $k'\beta + m'u$ because with $a'X = 0$ there is no β in $a'y$.

So?

Chapter 15

Sampling From Finite Populations

15.1 Finite e

When the e_i-population $e_1, e_2, \cdots, e_t$ is considered finite, the sum $e_1+e_2+\cdots+e_t$ is fixed, because it is the sum of the whole population. Therefore the variance of that sum is zero; i.e.,

$$\operatorname{var}\left(\sum_{i=1}^{t} e_i\right) = 0. \tag{1}$$

Assume

$$\operatorname{var}(e_i) = \sigma^2 \quad \forall \quad i, \quad \text{and} \quad \operatorname{cov}(e_i,\ e_j) = c \quad \forall \quad i \neq j.$$

Then (1) is

$$t\sigma^2 + t(t-1)c = 0 \quad \Rightarrow \quad c = -\sigma^2/(t-1). \tag{2}$$

Therefore for a sample of n ($< t$) drawn from the population of size t the variance-covariance matrix is

$$\begin{bmatrix} 1 & -1/(t-1) & \cdots & -1/(t-1) \\ \vdots & & & \\ -1/(t-1) & & & 1 \end{bmatrix}_{n\times n} \sigma^2 \tag{15.1}$$

which, for J_n, an $n \times n$ matrix of ones can be written as

$$(tI_n - J_n)/(t-1). \tag{3}$$

15.2 Finite u [208, 15.2]

Quite straightforward.

15.3 Infinite-by-Finite Interactions [209, 15.3]

It seems a pity that the controversy mentioned in the first line of this section is not accompanied by literature references. Details and references therein are available as follows.

Interaction effects having variance $I\sigma_\gamma^2$	Interaction effects summing to zero
LM 401, Table 9.9	LM 401-4, Table 9.10
VC 122-3, Table 4.6	VC 123-7, Table 4.7

15.4 Finite-by-Finite Interactions [210, 15.4]

No comment.

15.5 Finite, Factorial, Mixed Models [210, 15.5]

No comment: see Searle and Fawcett (1980).

15.6 Covariance Matrices [211, 15.6]

On [212], equation (15.11) is the same nature as (15.3); (15.12) is the same as the covariances in [210, 15.4], except for the following correction.

Correction In (15.12) the minus sign of the last $-\text{Var}$ should be deleted: see [210, 15.4].

15.7 Estimability and Predictability [213, 15.7]

The main paragraph on [214] concerns having a sample of two sires from a population of five sires, but with records on only the sample of two sires. There is then a discussion of "does $\hat{\mu}$ refer to the mean averaged over the 2 sires in the sample or over the 5" in the population. This is then formulated as predicting

$$\mu + (s_1 + s_2)/2 \qquad \text{or} \qquad \mu + (s_1 + s_2 + s_3 + s_4 + s_5)/5. \tag{4}$$

$$= (5I_5 - J_5)\begin{bmatrix} .4 & 0 \\ 0 & .1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} \bar{y}_{1\cdot} \\ \bar{y}_{2\cdot} \end{bmatrix} = \begin{bmatrix} 1.6 & -.1 \\ -.4 & .4 \\ -.41 1_3 & -.11_3 \end{bmatrix}\begin{bmatrix} \bar{y}_{1\cdot} \\ \bar{y}_{2\cdot} \end{bmatrix}. \tag{12}$$

Assembling (8) - (12) into (6) gives the equations at the top of [215]. Because those equations are of the form

$$\begin{bmatrix} F & 0 \\ L & I \end{bmatrix}\begin{bmatrix} \hat{\mu} \\ \hat{s}_1 \\ \hat{s}_2 \\ \hline \hat{s}_3 \\ \hat{s}_4 \\ \hat{s}_5 \end{bmatrix} = \begin{bmatrix} N_1 \\ \hline N_2 \end{bmatrix}\begin{bmatrix} \bar{y}_{1\cdot} \\ \bar{y}_{2\cdot} \end{bmatrix},$$

where N_1 and N_2 are matrices, it is clear that solutions $\hat{\mu}$, $\hat{s}_1$ and $\hat{s}_2$ will not involve $\hat{s}_3$, $\hat{s}_4$ and $\hat{s}_5$. Indeed since

$$\begin{bmatrix} F & 0 \\ L & I \end{bmatrix}^{-1} = \begin{bmatrix} F^{-1} & 0 \\ -LF^{-1} & I \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \hat{\mu} \\ \hat{s}_1 \\ \hat{s}_2 \end{bmatrix} = F^{-1}N_1\begin{bmatrix} \bar{y}_{1\cdot} \\ \bar{y}_{2\cdot} \end{bmatrix}. \tag{13}$$

And then

$$\begin{bmatrix} \hat{s}_3 \\ \hat{s}_4 \\ \hat{s}_5 \end{bmatrix} = [N_2 - LF^{-1}N_1]\begin{bmatrix} \bar{y}_{1\cdot} \\ \bar{y}_{2\cdot} \end{bmatrix} = N_2\begin{bmatrix} \bar{y}_{1\cdot} \\ \bar{y}_{2\cdot} \end{bmatrix} - L\begin{bmatrix} \hat{\mu} \\ \hat{s}_1 \\ \hat{s}_2 \end{bmatrix}. \tag{14}$$

With there being no records on sires 3, 4, and 5, it seems to me to be quite unreasonable to consider estimating s_3, s_4 and s_5. Their estimate from (14) will be just functions of $\hat{\mu}$, $\hat{s}_1$ and $\hat{s}_2$ – without any records on s_3, s_4 and s_5. That being so, why worry about the prediction error variances of such unobtainable estimates? Schaeffer rightly corrects me as follows.

> "They might be progeny of S_1 and S_2, and therefore replacements for breeding of females – so it is not [always] unreasonable. The prediction error variances would help the decision to replace or not."

Now, on applying (13) to the equations atop [215] we get the solutions in lower [214]:

$$\begin{bmatrix} \hat{\mu} \\ \hat{s}_1 \\ \hat{s}_2 \end{bmatrix} = \begin{bmatrix} .5 & .4 & .1 \\ 1.5 & 2.6 & -.1 \\ 0 & -.4 & 1.4 \end{bmatrix}^{-1}\begin{bmatrix} .4 & .1 \\ 1.6 & -.1 \\ -.4 & .4 \end{bmatrix}\begin{bmatrix} \bar{y}_{1\cdot} \\ \bar{y}_{2\cdot} \end{bmatrix}$$

$$= \frac{1}{9}\begin{bmatrix} 36 & -6 & -3 \\ -21 & 7 & 2 \\ -6 & 2 & 7 \end{bmatrix}\begin{bmatrix} .4 & .1 \\ 1.6 & -.1 \\ -.4 & .4 \end{bmatrix}\begin{bmatrix} \bar{y}_{1\cdot} \\ \bar{y}_{2\cdot} \end{bmatrix} = \frac{1}{9}\begin{bmatrix} 6 & 3 \\ 2 & -2 \\ -2 & 2 \end{bmatrix}\begin{bmatrix} \bar{y}_{1\cdot} \\ \bar{y}_{2\cdot} \end{bmatrix}. \tag{15}$$

as in (15).

Correction The γ_{11} in [216, last line] should be γ_{i1}.

15.8 BLUP When Some u_i Are Finite [217, 15.8]

Comments made earlier about trying to estimate sire effects for sires with no data apply here too, for sire effects and for interaction effects, on which there are no data. For example, in [218, (15.15)] there is no $y_{4.}$ for sire 4. What is worse is that 0 is used for the $y_{4.}$ as if there is an observed $y_{4.}$ of zero. Nonsense, there is no $y_{4.}$. Its non-existence does not mean it is zero. Non-existence and observed zero are *not* the same. If an animal of interest, s_4 say, in [15.15], has no records but is related to animals that do, then that animal can be predicted from its relative's records using its relationship to those animals. For example, if s_4 has no records but its sire is s_2 and its maternal grandsire is s_3, then surely

$$\tilde{s}_4 = \tfrac{1}{2}\tilde{s}_2 + \tfrac{1}{4}\tilde{s}_3.$$

The last three lines of [219] apply to the so-called γ_{21} and γ_{32} equations; and they relate to the reference to γ_{21} and γ_{32} in the lines before [219, (15.17)].

Corrections The second $\hat{t}_1$ in [219, (15.17)] should be $\hat{t}_2$. The γ_{12} in [220, 2nd line] should be γ_{11}.

Comment The mention of BLUPs adding to zero in [219, lines 5-6, and in 7 up] is simply part of some quite general results for the usual mixed model (with G being block diagonal of blocks $\sigma_i^2 I_{q_i}$) that BLUPs of random effects summed over all levels of a factor always add to zero. See Searle (1997), and also these notes at Section [23.1].

15.9 An Easier Computational Method [220, 15.9]

Derivation of (15.18) comes from (6) with

$$X = \mathbf{1}_{10}, \quad Z = \begin{bmatrix} \mathbf{1}_5 & \cdot & \cdot \\ \cdot & \mathbf{1}_3 & \cdot \\ \cdot & \cdot & \mathbf{1}_2 \end{bmatrix}, \quad R = 10I \quad \text{and} \quad G = 3I_3 - J_3.$$

Chapter 16

The One-Way Classification

Following $y_{ij} = \mu + a_i + e_{ij}$ of (16.1), the e_{ij} are simply defined as having mean zero. A better approach is to start by defining $E(y_{ij}) = \mu + a_i$ and then define e_{ij} as $e_{ij} = y_{ij} - E(y_{ij})$. This avoids having to specifically describe what e_{ij} consists of. It does, of course, yield (16.1). Defining $\text{var}(e_{ij}) \equiv \sigma_e^2$, or describing that as a property of e_{ij} is better stated as "attributing" a variance of σ_e^2 to each e_{ij}.

16.1 Estimation and Tests for Fixed a [223, 16.1]

It would help to have at some point $y = X\beta + e$ as the general model equation and $X'X\beta^0 = X'y$ as the resulting OLS equations. Resorting to the MMEs (to which no reference is given on [223], but see [16, (3.4)], for example) is a little cumbersome, but certainly provides uniformity of methodology.

The arithmetic in the lower half of [225] would be easier to follow (and thus be more instructive for beginners) if fractions were retained. For example

$$
\begin{aligned}
\text{var}(K'\beta^0) &= \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} 0 & & & \\ & \frac{1}{8} & & \\ & & \frac{1}{3} & \\ & & & \frac{1}{4} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1 & 1 \\ -1 & 0 \\ 0 & -1 \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{8}+\frac{1}{3} & \frac{1}{8} \\ \frac{1}{8} & \frac{1}{8}+\frac{1}{4} \end{bmatrix} = \frac{1}{24}\begin{bmatrix} 11 & 3 \\ 3 & 9 \end{bmatrix} = \begin{bmatrix} .45833 & .125 \\ .125 & .375 \end{bmatrix} \\
K'\beta^0 &= \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} 0 \\ 49/8 \\ 16/3 \\ 13/4 \end{bmatrix} = \begin{bmatrix} 49/8 - 16/3 \\ 49/8 - 13/4 \end{bmatrix} = \frac{1}{24}\begin{bmatrix} 19 \\ 69 \end{bmatrix} = \begin{bmatrix} .79167 \\ 2.875 \end{bmatrix}.
\end{aligned}
$$

and it can be seen that the calculated values here are part of (16.6). The remainder of (16.6) is derived similarly.

16.2.2 Sums of Squares

The top line of [227] merits expansion in terms of the hypotheses (called 'tests') and reductions in sums of squares in the lower half of [226]. What is being said is as laid out in the following table.

Table A: Sums of Squares

As indicated on [227, line 1]			$R(\cdot\|\cdot)$ Notation
4-5	$R(\beta_1\|\mu)$	=	$R(\mu,\beta_1) - R(\mu)$
3-4	$R(\beta_2\|\mu,\beta_1)$	=	$R(\mu,\beta_1,\beta_2) - R(\mu,\beta_1)$
2-3	$R(\beta_3\|\mu,\beta_1,\beta_2)$	=	$R(\mu,\beta_1,\beta_2,\beta_3) - R(\mu,\beta_1,\beta_2)$
1-2	$R(\beta_4\|\mu,\beta_1,\beta_2,\beta_3)$	=	$R(\mu,\beta_1,\beta_2,\beta_3,\beta_4) - R(\mu,\beta_1,\beta_2,\beta_3)$
	TOTAL	=	$R(\mu,\beta_1,\beta_2,\beta_3,\beta_4) - R(\mu)$

Table B: Hypotheses Tested by the Sums of Squares of Table A

Sum of Squares			Hypothesis
$R(\beta_1\|\mu)$	$H\colon\beta_1$	=	0, adjusted for μ, ignoring β_2, β_3 and β_4
$R(\beta_2\|\mu,\beta_1)$	$H\colon\beta_2$	=	0, adjusted for μ and β_1, ignoring β_3 and β_4
$R(\beta_3\|\mu,\beta_1,\beta_2)$	$H\colon\beta_3$	=	0, adjusted for μ, β_1 and β_2, ignoring β_3
$R(\beta_4\|\mu,\beta_1,\beta_2,\beta_3)$	$H\colon\beta_4$	=	0, adjusted for μ, β_1, β_2 and β_3

The description of these in [227, lines 6-9] as linear, quadratic, cubic and quartic is misleading, because although each sum of squares in Table B is independent of

$$\text{SSE} = y'y - R(\mu,\beta_1,\beta_2,\beta_3,\beta_4) = y'y - [\text{Total } + R(\mu)],$$

those sums of squares are not independent of each other. Independent sums of squares can be achieved by using orthogonal polynomials (see Pearson & Hartley, and Robson, 1959).

16.2.3 Hypotheses and models

The verbal descriptions shown in Table B need to be considered with care, particularly with regard to such phrases as "ignoring β_2, β_3 and β_4" in the first line of the table. That means, for instance, that $R(\beta_1|\mu)$ tests $H\colon\beta_1 = 0$ in the model equation $y_{ij} = \mu + \beta_1 x_i + e_{ij}$. It is *not*

$$\tilde{x} = \Sigma n_i x_i / N \quad \text{with} \quad N = \Sigma n_i.$$

Then

$$M_0 \mathbf{w}_1(\mathbf{M}_0, \mathbf{w}_1)^- = \mathbf{z}\mathbf{z}'/\mathbf{z}'\mathbf{z} = \mathbf{z}\mathbf{z}'/\Sigma n_i(x_i - \tilde{x})^2.$$

Hence, using (5), the hypothesis tested by $R(\beta_1|\mu)$ in (2) is

$$\begin{aligned} H &: \mathbf{z}\beta_1 + (\mathbf{z}\mathbf{z}'/\mathbf{z}'\mathbf{z})\left\{_c\, x_i^2 \mathbf{1}_{n_i}\right\}\beta_2 = 0 \\ H &: \mathbf{z}\left[\beta_1 + \frac{\mathbf{z}_i'\{_c\, x_i^2 \mathbf{1}_{n_i}\}}{\mathbf{z}'\mathbf{z}}\beta_2\right] = 0 \\ H &: \mathbf{z}\left[\beta_1 + \frac{\Sigma n_i(x_i - \tilde{x})x_i^2}{\Sigma n_i(x_i - \tilde{x})^2}\right] = 0 \end{aligned}$$

and this is not $H\colon \beta_1 = 0$.

In contrast, line 3 of LMFUD Table 8.5 is for $\mathbf{y} = X_1\beta_1 + X_2\beta_2 + e$, with $R(\beta_2|\beta_1)$ then testing

$$H\colon M_1 X_2 \beta_2 = 0.$$

Adapted to $\mathbf{y} = \mu\mathbf{w}_0 + \beta_1\mathbf{w}_1 + e$, the hypothesis for $R(\beta_1|\mu)$ is

$$H\colon M_0\mathbf{w}_1\beta_1 = 0 \qquad \text{and} \qquad H\colon \{(x_i - \tilde{x})\mathbf{1}_{n_i}\}\beta_1 = 0,$$

which can only be true when $H\colon \beta_1 = 0$ is true.

Thus $R(\beta_1|\mu)$ for $y_{ij} = \mu + \beta_1 x_i + e_{ij}$ tests $H\colon \beta_1 = 0$, but for $y_{ij} = \mu + \beta_1 x_i + \beta_2 x_i^2 + e_{ij}$ it does not. This principle extends to the other sums of squares in Table B.

16.3 Biased Estimation of $\mu + a_i$ [227, 16.3]

The second line begins "Using the same data as in the previous section...". This seems to be wrong. The "previous section" is 16.2, and its data is for fitting the quartic of (16.5) which has five parameters, μ and four βs. But (16.8) is six equations. Moreover, in Section 16.2 the parameters other than μ are βs and both by its title and the last equation on [227] the parameters are as, not βs. And one might think that "previous section" could apply to the section two back, namely 16.1, because its parameters are as; but only three of them, not five, as is implicit in (16.8).

Query So where does the data come from?

16.7.1 BLUPs add to zero

The equation between (16.12) and (16.13) is $10\Sigma\hat{a}_i = 0$, namely $\Sigma\hat{a}_i = 0$. This is the simplest example of a very general result for the usual mixed model (with G being block diagonal with blocks $\sigma_i^2 I_{q_i}$) that BLUPs, for example of main effects, always add to zero; and so do BLUPs of random interaction effects, including interactions of fixed and random factors, for which they also add to zero over each level of the fixed effects factor. These results are derived and discussed in Searle (1997), and in these notes at Section [23.1].

16.7.2 A property of an inverse matrix

The first line of [233] merits derivation. It concerns the inverse of the matrix on the left of (16.12) which, in general, we write as

$$H = \begin{bmatrix} N & n' \\ n & D \end{bmatrix} \tag{7}$$

where

$$N = \Sigma n_i, \quad \mathbf{n}' = \{_r\, n_i\} \quad \text{and} \quad D = \{_d\, n_i + \lambda\} \quad \text{for} \quad \lambda = \sigma_e^2/\sigma_\alpha^2. \tag{8}$$

Then [233, line 1] states that

$$[0 \quad \mathbf{1}_a']H^{-1} = [-1 \quad \mathbf{1}_a']/\lambda. \tag{9}$$

We proceed to prove this.

From (7)

$$H^{-1} = \begin{bmatrix} 1/N & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -n'/N \\ I \end{bmatrix} (D - nn'/N)^{-1}[-n/N \quad I]. \tag{10}$$

But a general result in matrix algebra

$$(D - CA^{-1}B)^{-1} = D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1}$$

has the special case

$$(D - \mathbf{nn}'/N)^{-1} = D^{-1} + \frac{D^{-1}nn'D^{-1}}{N - n'D^{-1}n}.$$

Therefore

$$[0 \quad 1_a']H^{-1} = 1'\left[D^{-1} + \frac{D^{-1}nn'D^{-1}}{N - n'D^{-1}n}\right]\left[-\frac{n}{N} \quad I\right]. \tag{11}$$

from (12). And since for random a_i the model $y = \mu + a_i + e_{ij}$ has $X = \mathbf{1}$, and using I_i and J_i to represent matrices of order n_i,

$$\begin{aligned}
X'V^{-1}X &= \mathbf{1}'V^{-1}\mathbf{1} = \mathbf{1}'\left\{_d\, \sigma_e^2 I_i + \sigma_\alpha^2 J_i\right\}^{-1}\mathbf{1} \\
&= \mathbf{1}'\left\{_d\, \frac{1}{\sigma_e^2}\left(I_i - \frac{\sigma_\alpha^2}{\sigma_e^2 + n_i\sigma_\alpha^2}J_i\right)\right\}\mathbf{1} \\
&= \frac{1}{\sigma_e^2}\left(\sum n_i - \sum \frac{n_i^2\sigma_\alpha^2}{\sigma_e^2 + n_i\sigma_\alpha^2}\right) \\
&= \frac{1}{\sigma_e^2}\sum n_i\left(1 - \frac{n_i\sigma_\alpha^2}{\sigma_e^2 + n_i\sigma_\alpha^2}\right) = \sum \frac{n_i}{\sigma_e^2 + n_i\sigma_\alpha^2} \\
&= \sum \frac{n_i\sigma_e^2/\sigma_\alpha^2}{\sigma_e^2\left(\sigma_e^2/\sigma_\alpha^2 + n_i\right)} \\
&= \frac{\lambda}{\sigma_e^2}\sum \frac{n_i}{n_i + \lambda}.
\end{aligned}$$

Thus

$$\begin{aligned}
(X'V^{-1}X)^- &= \sigma_e^2\Big/\lambda\sum \frac{n_i}{n_i + \lambda} \\
&= \sigma_e^2 C_{11}, \text{ from (14).}
\end{aligned}$$

This is the basis of the $\text{var}(\hat{\mu}) = .079\sigma_e^2$ result in [233, line 4].

To confirm the Method I estimation of σ_e^2 and σ_α^2 on [233, lines 8-12] we use VC section F.1b, for which, based on the "Suppose this is 2.8" of [233, line 12].

$$\begin{aligned}
T_0 &= y'y = 2.8(19-5) + 210.9583 = 250.1583 \\
T_A &= \Sigma y_{i.}^2/n_i = 10^2/5 + 7^2/2 + 3^2/1 + 8^2/3 + 33^2/8 = 210.9583 \\
T_\mu &= y_{..}^2/n_. = 61^2/19 = 195.8421 \\
S_2 &= 5^2 + 2^2 + 1^2 + 3^2 + 8^2 = 103 \\
\hat{\sigma}_e^2 &= \frac{250.1583 - 210.9583}{19-5} = 2.8 \\
\hat{\sigma}_\alpha^2 &= \frac{210.9583 - 195.8421 - (5-1)2.8}{19 - 103/19} = .288.
\end{aligned}$$

The "Suppose this [i.e., $\hat{\sigma}_e^2$] is 2.8" concerns me, as somewhat concocted in order to satsify the $\sigma_e^2/\sigma_\alpha^2 = 10$ of [232, two lines above (16.12)]: for note, $\hat{\sigma}_e^2/\hat{\sigma}_\alpha^2 = 2.8/.288$ is nearly 10.

The last paragraph of [233] does not appeal to me: it is "approximate MIVQUE". Ugh.

$$
\begin{aligned}
&= 19\mu^2 + \sigma_s^2 \left\{103 + 37\frac{1}{2}\right\} \Big/ 19 + \sigma_e^2 \\
&= 19\mu^2 + 7.395\sigma_s^2 + \sigma_e^2.
\end{aligned}
$$

The term in σ_s^2 is as in [237 two lines below (16.18)]. The a_{ii} and $a_{ii'}$ terms in these calculations are elements of the relationship A-matrix given below [236, (15.14)]. And the a is the number of sires: $i = 1, 2, \cdots, a$.

As usual, I chose to ignore approximate MIVQUE.

Chapter 17

The Two-Way Classification

17.1 The two-way fixed model [239, 17.1]

The first word in the line before (17.3) is very important: equations (17.2) through (17.8) are *definitions*. Also, the symbols $\bar{\mu}_{i\cdot}$ and $\bar{\mu}_{\cdot j}$ in (17.3) and (17.4) are not defined. For example, is $\bar{\mu}_{i\cdot}$ defined as $\Sigma_{j=1}^{c}\mu_{ij}/c$ or as $\Sigma_{j=1}^{c} n_{ij}\mu_{ij}/n_{i\cdot}$? Presumably the former.

Using that definition, $\bar{\mu}_{i\cdot} = \Sigma_{j=1}^{c} \mu_{ij}/c$, and its obvious extensions to $\bar{\mu}_{\cdot j}$ and $\bar{\mu}_{\cdot\cdot}$, shows how the *definitions* (17.5) through (17.8) are related to the familiar overparameterized model of (17.1). For example, (17.5) is

$$\text{Row effect} = \bar{\mu}_{i\cdot} - \bar{\mu}_{\cdot\cdot} = \mu + a_i + \bar{b}_\cdot + \bar{\gamma}_{i\cdot} - (\mu + \bar{a}_\cdot + \bar{b}_\cdot + \bar{\gamma}_{\cdot\cdot}) = a_i - \bar{a} + \bar{\gamma}_{i\cdot} - \bar{\gamma}_{\cdot\cdot} \tag{1}$$

and (17.8) is

$$\begin{aligned} \text{Interaction effect} &= \mu_{ij} - \bar{\mu}_{i\cdot} - \bar{\mu}_{\cdot j} + \bar{\mu}_{\cdot\cdot} = \mu + a_i + b_j + \gamma_{ij} - (\mu + a_i + \bar{b}_\cdot + \bar{\gamma}_{i\cdot}) \\ &\qquad - (\mu + \bar{a}_\cdot + b_j + \bar{\gamma}_{\cdot j}) + (\mu + \bar{a}_\cdot + \bar{b}_\cdot) + \bar{\gamma}_{\cdot\cdot} \\ &= (\gamma_{ij} - \bar{\gamma}_{i\cdot} - \bar{\gamma}_{\cdot j} + \bar{\gamma}_{\cdot\cdot}). \end{aligned} \tag{2}$$

Notice, though, that (2) is *not* the accepted definition of interaction as discussed, for example, in LM page 318 and in LMFUD, page 327, equation (9). On defining (2) as

$$\varphi_{ij} = \mu_{ij} - \bar{\mu}_{i\cdot} - \bar{\mu}_{\cdot j} + \bar{\mu}_{\cdot\cdot} = \gamma_{ij} - \bar{\gamma}_{i\cdot} - \bar{\gamma}_{\cdot j} + \bar{\gamma}_{\cdot\cdot}, \tag{3}$$

the definition of interaction in those references is where W is defined as

$$\mathrm{E}\begin{bmatrix} \{\tilde{y}_{i\cdot\cdot}\} \\ \{\tilde{y}_{\cdot j\cdot}\} \end{bmatrix} = \mathrm{W}\begin{bmatrix} \mathbf{b} \\ \mathbf{t} \end{bmatrix}$$

with $\tilde{y}_{i\cdot\cdot} = \Sigma_j \bar{y}_{ij\cdot}$ and $\tilde{y}_{\cdot j\cdot} = \Sigma_i \bar{y}_{ij\cdot}$ as totals (not means, as stated in [245, line 4]) of cell means. For example, for $i = 3$, from the data on [242]

$$\tilde{y}_{3\cdot\cdot} = 61/5 + 13/1 + 61/4 = 12.2 + 13 + 15.25 = 40.45,$$

as in (17.22).

The logic behind (17.22) is as follows. For the overparameterized model (17.1) applied to the breed-by-treatment data of the table on [242]

$$\begin{aligned} E(y_{ijk}) &= \mu + b_i + t_j + \gamma_{ij} \\ E(\bar{y}_{ij\cdot}) &= \mu + b_i + t_j + \gamma_{ij} \end{aligned}$$

$$\bar{y}_{ij\cdot} = \mu^\circ + b_i^\circ + t_j^\circ + \gamma_{ij}^\circ \tag{4}$$

$$\sum_{j=1}^{c} \bar{y}_{ij\cdot} = c\mu^\circ + cb_i^\circ + t_\cdot^\circ + \gamma_{i\cdot}^\circ \tag{5}$$

$$\sum_{i=1}^{r} \bar{y}_{ij\cdot} = r\mu^\circ + b_i^\circ + rt_j^\circ + \gamma_{\cdot j}^\circ. \tag{6}$$

On deciding to derive solutions with

$$\mu^\circ = 0, \qquad \gamma_{i\cdot}^\circ = 0 \qquad \text{and} \qquad \gamma_{\cdot j}^\circ = 0 \tag{7}$$

(5) and (6) reduce to

$$cb_i^\circ + \Sigma_j t_j^\circ = \Sigma_j \bar{y}_{ij\cdot} \quad \forall \quad j = 1, \cdots, r \tag{8}$$

and

$$\Sigma_i b_i^\circ + rt_j^\circ = \Sigma_i \bar{y}_{ij\cdot} \quad \forall \quad i = 1, \cdots, c.$$

These are precisely (17.22). After solving them use (4) to obtain

$$\gamma_{ij}^\circ = \bar{y}_{ij\cdot} - b_i^\circ - t_j^\circ \tag{9}$$

as in [245, three lines above 17.3]. Very clever. And the values given by (9) will satisfy (7). For example, from (5), with $\mu^\circ = 0$,

$$\gamma_{i\cdot}^\circ = \Sigma_j \bar{y}_{ij\cdot} - cb_i^\circ - \Sigma t_j^\circ = 0. \qquad \text{from (8)}.$$

then they will be found biased. But that is not logical. Biasedness is based on expected values over the model from which estimators are derived.

In the last three lines of [249] the "suggested $\cdots$ reduction in SS" can be expressed as

$$R(a|\mu, b) = R(\mu, a, b) - R(\mu, b). \tag{10}$$

The question is then raised of testing this "against some denominator", and it is suggested that if $\hat{\sigma}_e^2$ is used "the denominator is too small". But $\hat{\sigma}_e^2$ here is not defined; presumably it is

$$\hat{\sigma}_e^2 = \frac{y'y - R(\mu, a, b, \gamma)}{N - s}, \tag{11}$$

where s is the number of filled cells. Alternatively, if "$R \times C$ for MS is used" that would be

$$\mathrm{M} = R(\mu, a, b, \gamma) - R(\mu, a, b) \tag{12}$$

and it is suggested that the "denominator is probably too large". The word "probably" is important because in fact (12) is not necessarily always larger than the numerator of $\hat{\sigma}_e^2$ in (11). Moreover, it seems to me that M is not appropriate anyway. Either $\hat{\sigma}_e^2$ should be used or alternatively

$$\tilde{\sigma}_e^2 = \frac{y'y - R(\mu, a, b)}{N - (a + b - 1)}. \tag{13}$$

Of these two alternatives, $\hat{\sigma}_e^2$ and $\tilde{\sigma}_e^2$, $\tilde{\sigma}_e^2$ is from the no-interaction model. Since the estimators of μ, a and b used implicitly in (13) are from the no-interaction model it seems to me that $\tilde{\sigma}_e^2$ is appropriate. This methodology is then consistent, in the sense of being a no-interaction analysis. And *within that context* the F-statistic based on (10) and (13), namely

$$F = \frac{R(a|\mu, b)}{(r-1)\tilde{\sigma}_e^2}$$

is definitely testing

$$H{:}\, a_1 = a_2 = \ldots = a_v. \tag{14}$$

LMFUD, both at page 106, equation (81) and in Section 9.2f, deals with this in some detail, as does Section 7.1g also.

Note that (14), for the no-interaction model, negates [249, last line] which states that (10) is not providing a test of rows. That statement is true if, as its context seems to imply, it is being

17.8 The two-way mixed model [258, 17.8]

An easy way of appreciating [258, (17.40)] is to think of a small example, one of just three columns, say. Then the terms $\mu + a_i + b_j + \gamma_{ij}$ in the first row are

$$\mu + a_1 + b_1 + \gamma_{11} \qquad \mu + a_1 + b_2 + \gamma_{12} \qquad \mu + a_1 + b_3 + \gamma_{13}\,. \tag{15}$$

For

$$\text{var}(a_i) = \sigma_a^2, \quad \text{var}(\gamma_{ij}) = \sigma_\gamma^2 \quad \text{and } \text{cov}(\gamma_{ij}, \gamma_{ij'}) = \tau.$$

Then the variance-covariance matrix, C, of the three terms in (15) is

$$\mathrm{C} = \begin{bmatrix} \sigma_a^2 + \sigma_\gamma^2 & \sigma_a^2 + \tau & \sigma_a^2 + \tau \\ & \sigma_a^2 + \sigma_\gamma^2 & \sigma_a^2 + \tau \\ \text{symmetric} & & \sigma_a^2 + \sigma_\gamma^2 \end{bmatrix}.$$

The variance (diagonal) elements here are the $\text{var}(\alpha_{ij})$ above [259, (17.42)]; and in (17.42)

$$\tau = \text{cov}(\alpha_{ij}, \alpha_{ij'}) = \text{cov}(\gamma_{ij}, \gamma_{ij'}) = -\sigma_\gamma^2/(q-1). \tag{16}$$

(16) is not (17.42). The latter has $\sigma_a^2 - \sigma_\gamma^2/(q-1)$; it does not have $\text{cov}(\gamma_{ij},\ \ \gamma_{ij'})$. Frankly, (17.42) seems strange.

The result

$$\text{cov}(\gamma_{ij},\ \ \gamma_{ij'} = -\sigma_\gamma^2/(q-1)$$

comes from assuming

$$\Sigma_{j=1}^q \gamma_{ij} = 0 \tag{17}$$

which implies

$$\text{var}\left(\Sigma_{j=1}^q \gamma_{ij}\right) = 0 = q\sigma_\gamma^2 + q(q-1)\tau \;\Rightarrow\; \tau = -\sigma_\gamma^2/(q-1).$$

I dislike (17). It makes no sense for random γs; and it is not functional when some cells have no data. Why not just estimate σ_e^2, σ_a^2 and τ?

Chapter 18

The Three-Way Classification

18.1 The three-way fixed model [265, 15.1]

The definitions in [265-6, (18.3)] are similar to those in [239-240, (17.3)-(17.8)]. Akin to the discussion in Section 17.1 of this supplement, there must be recognition that the *ab*-interaction definition atop [266], namely $\bar{\mu}_{ij\cdot} - \bar{\mu}_{i\cdot\cdot} - \bar{\mu}_{\cdot j\cdot} + \bar{\mu}_{\cdot\cdot\cdot}$, is different from the interaction definition $\bar{\mu}_{ij\cdot} - \bar{\mu}_{i'j\cdot} - \bar{\mu}_{ij'\cdot} + \bar{\mu}_{i'j'\cdot}$ in LMFUD, page 389, equation (17).

18.2 The filled subclass case [266, 18.2]

The whole of LMFUD Section 10.2 is devoted to multiway classifications. Many features are illustrated with a three-factor $2 \times 3 \times 4$ example having much easier arithmetic than the example on [266].

The column product operation at the bottom of [267] is the Hadamard product of two columns: $\{p_i\} \cdot \{q_i\} = \{p_i q_i\}$. See MAUFS Section 2.8n.

The last half of [271 line 5] could be stated more clearly as "... each main effects factor and each interaction factor is deleted in turn ...." In contemplating the whole analysis presented in [266-267] it must be remembered that it is for all-cells-filled data, *and* it defines effects as adding to zero; i.e., the Σ-restrictions are invoked. Without these restrictions, the $2 \times 3 \times 4$ design of the data has the following number of parameters in the overparameterized model of [265, (8.1)].

Section 10.4, which deals with models having not all possible interactions. Procedure 3 does not appeal, because it assumes some interaction effects are zero. And the rest of the section uses prior values for "average sums of squares and products of interaction" – a procedure which has no appeal for me.

18.4 The three-way mixed model [278, 18.4]

Aside from its first dozen words, [279] is somewhat mystifying – and no explanations are given.

Chapter 19

Nested Classifications

19.1 Two-way fixed within fixed [281, 19.1]

$$y_{ijk} = t_i + a_{ij} + e_{ijk}. \tag{1}$$

In [281, third line up] the α_j bears no relation to the αs in (17.18) nor those preceding (17.40). Furthermore, in $\Sigma_j \alpha_j a_{ij}$ the α_j should be α_{ij} because, for example, the α multiplying a_{1j} does not have to be the same as that multiplying a_{2j} (e.g., $.3a_{11} + .7a_{12}$) and because not every t_i is associated with the same number of a_{ij}s.

Note that α_i defined in [282, line 2] is not the just-discussed α_j at the bottom of [281]. Moreover, in that $\alpha_i = t_i + \Sigma_j k_j a_{ij}$ the k_j should be k_{ij}, for the same reasoning as in the preceding paragraph. To consider α_i in general it is necessary to define the number of a_{ij} terms within t_i. Let that number be c_i, so that $j = 1, \cdots, c_i$. Then

$$\alpha_i = t_i + \sum_{j=1}^{c_i} k_{ij} a_{ij} \qquad \text{with} \qquad \sum_{j=1}^{c_i} k_{ij} = 1.$$

Then the ith main effect [282, line 3] is defined as

$$\alpha_i - \bar{\alpha} = t_i + \sum_{j=1}^{c_i} k_{ij} a_{ij} - \frac{1}{a} \sum_{r=1}^{a} \left(t_r + \sum_{j=1}^{c_r} k_{rj} a_{rj} \right).$$

No mention is made of normal equations. They are available in LM, Section 6.4, wherein $\mu + \alpha_i$ plays the part of t_i and β_{ij} the part of a_{ij}. As in LM, page 252, equation (70), a solution (the

And for [282, bottom]

$$\begin{aligned}
q &= [-4.5 \quad -1.5]\left\{\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix}\begin{bmatrix} .1125 & 0 & 0 \\ 0 & .177 & 0 \\ 0 & & .175 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{bmatrix}\right\}^{-1}\begin{bmatrix} -4.5 \\ -1.5 \end{bmatrix} \\
&= [-4.5 \quad -1.5]\begin{bmatrix} .2875 & .175 \\ .175 & .35278 \end{bmatrix}^{-1}\begin{bmatrix} -4.5 \\ -1.5 \end{bmatrix} \\
&= [-4.5 \quad -1.5]\begin{bmatrix} 4.98283 & -2.47180 \\ -2.47180 & 4.06081 \end{bmatrix}\begin{bmatrix} -4.5 \\ -1.5 \end{bmatrix} \\
&= 76.64.
\end{aligned}$$

This is the numerator sum of squares – *not* mean square. The latter is $76.64/2 = 38.32$, which differs from the 26.70 of [282, last line] because $K'\beta^0 = [-4.5 \quad -1.5]'$ and not $[-3.5 \quad -.5]$.

In [283, second line] "differences among a_{ij}" should be "differences among a_{ij} within t_i". Thus the hypothesis is

$$H\colon \begin{bmatrix} a_{11}-a_{12} & 0 & 0 & 0 \\ 0 & a_{21}-a_{23} & 0 & 0 \\ 0 & 0 & a_{22}-a_{23} & 0 \\ 0 & 0 & 0 & a_{31}-a_{32} \end{bmatrix} = 0.$$

The numerator sum of squares (*not* MS – as in [283, mid-page]) is

$$\begin{bmatrix} 5-3 \\ 8-6 \\ 7-6 \\ 9-8 \end{bmatrix}'\begin{bmatrix} 2.2222 & 0 & 0 & 0 \\ 0 & .92308 & -.76923 & 0 \\ 0 & -.76923 & 2.30769 & 0 \\ 0 & 0 & 0 & 1.42557 \end{bmatrix}\begin{bmatrix} 2 \\ 2 \\ 1 \\ 1 \end{bmatrix} = 13.24.$$

This *is* a sum of squares. The corresponding mean square is *not* $13.24/1$; degrees of freedom are clearly 4 (not 1) and so the MS is $13.24/4 = 3.31$.

Since t_i is not estimable, and because $t_i^0 = 0$, testing

$$H\colon \begin{bmatrix} (4a_{11}+5a_{12})/9 & & -(5a_{31}+2a_{32}) \\ & (a_{21}+10a_{22}+2a_{23}) & -(5a_{31}+2a_{32}) \end{bmatrix} = 0$$

is done by the calculation on [283, bottom] and [284, top].

19.2 Two-way random within fixed

Recall that MMEs are of the form

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z+G^{-1} \end{bmatrix}\begin{bmatrix} \beta^0 \\ \tilde{u} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}. \tag{2}$$

where dashed lines in a matrix indicate partitioning to assist readability. Therefore, with $\sigma_e^2 = 40$, and with G^{-1}, X and Z as above, the MMEs of (5) are

$$\begin{bmatrix} \{_d\, 5\cdot\cdot \,|\, 2\cdot\cdot \,|\, \dot{3}\,\cdot \,|\, \cdot 8\cdot \,|\, \cdot\cdot 5\} + 40I_5 \otimes G^{-1} & Z'X \\ X'Z & X'X \end{bmatrix} \begin{bmatrix} \tilde{u} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} Z'y \\ X'y \end{bmatrix} \tag{6}$$

in which

$$X'Z = \left[\begin{array}{ccc|ccc|ccc|ccc|ccc} 5 & \cdot & \cdot & 2 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 3 & \cdot & \cdot & 8 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 5 \end{array}\right] \quad \text{and} \quad X'X = \begin{bmatrix} 7 & \cdot & \cdot \\ \cdot & 11 & \cdot \\ \cdot & \cdot & 5 \end{bmatrix}. \tag{7}$$

And

$$\begin{bmatrix} Z'y \\ X'y \end{bmatrix} = [7\ \cdot\ \cdot\,|6\ \cdot\ \cdot|\cdot\ 7\ \cdot|\cdot\ 4\ \cdot|\cdot\ \cdot\ 8\,|\,13\ \ 16\ \ 8]'. \tag{8}$$

Assembling (4), (7) and (8) into (6) gives the 18×18 set of equations in [285-6, (19.1)]. For example, the leading 3×3 matrix on the left-hand side of (6) is

$$\begin{bmatrix} 5 & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix} + \frac{40}{80}\begin{bmatrix} 40 & -20 & 0 \\ -20 & 35 & -10 \\ 0 & -10 & 20 \end{bmatrix} = \frac{40}{80}\left\{\begin{bmatrix} 10 & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix} + \begin{bmatrix} 40 & -20 & 0 \\ -20 & 35 & -10 \\ 0 & -10 & 20 \end{bmatrix}\right\}$$

$$= 40\left\{\frac{1}{80}\begin{bmatrix} 50 & -20 & 0 \\ -20 & 35 & -10 \\ 0 & -10 & 20 \end{bmatrix}\right\}.$$

The remaining 15 columns of those three rows and the first three elements in $Z'y$ are

$$\left\{ 0_{3\times 12} \quad \begin{bmatrix} 5 & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix} \right\} \quad \text{and} \quad \begin{bmatrix} 7 \\ \cdot \\ \cdot \end{bmatrix}.$$

Hence the first three equations of (5) are as follows with, as on [286], $s_{21}, \cdots, s_{53}$, t_1, t_2, and t_3 following s_{13} in the parameter vector:

$$\left(40\left\{\frac{1}{80}\begin{bmatrix} 50 & -20 & 0 \\ -20 & 35 & -10 \\ 0 & -10 & 20 \end{bmatrix}\right\} \quad 0_{3\times 12} \quad \begin{bmatrix} 5 & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix} \right)\begin{pmatrix} s_{11} \\ s_{12} \\ s_{13} \end{pmatrix} = \begin{bmatrix} 7 \\ \cdot \\ \cdot \end{bmatrix}$$

$$80^{-1}\left(\begin{bmatrix} 50 & -20 & 0 \\ -20 & 35 & -10 \\ 0 & -10 & 20 \end{bmatrix} \quad 0_{3\times 12} \quad \begin{bmatrix} 10 & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix} \right)\begin{pmatrix} s_{11} \\ s_{12} \\ s_{13} \end{pmatrix} = \begin{bmatrix} 7/40 \\ \cdot \\ \cdot \end{bmatrix} = \begin{bmatrix} .175 \\ \cdot \\ \cdot \end{bmatrix}$$

and the three equations

$$\left\{\begin{bmatrix} 5 & 2 & \cdot & \cdot & \cdot \\ \cdot & \cdot & 3 & 8 & \cdot \\ \cdot & \cdot & \cdot & \cdot & 5 \end{bmatrix} \quad \begin{bmatrix} 7 & \cdot & \cdot \\ \cdot & 11 & \cdot \\ \cdot & \cdot & 5 \end{bmatrix}\right\} \begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \\ t_1 \\ t_2 \\ t_3 \end{bmatrix} = \begin{bmatrix} 13 \\ 16 \\ 8 \end{bmatrix}.$$

Now, in reading the preceding sets of five and three equations as a single set of eight equations, notice that the leading 5×5 submatrix on the left-hand side is

$$\begin{aligned} &\{_d\, 5 \;\; 2 \;\; 3 \;\; 8 \;\; 5\} + \{_d\, 40/3 \;\; 40/3 \;\; 10 \;\; 10 \;\; 8\} \\ &= \frac{1}{3}\{_d\, 15+40 \quad 6+40 \quad 9+30 \quad 24+30 \quad 15+14\} \\ &= \frac{1}{3}\{_d\, 45 \;\; 46 \;\; 39 \;\; 54 \;\; 39\}. \end{aligned}$$

Aside from the fraction 1/3 this is the leading submatrix in (19.3), and eliminating that fraction by multiplying every other element in (9) by 3 gives (19.3), without its 1/120 on each side.

19.3 Random within random [287, 19.3]

Now the model is written as

$$y_{ijk} = \mu + s_i + \alpha_{ij} + e_{ijk} \tag{10}$$

where the t_i and a_{ij} of (1) are now $\mu + s_i$ and d_{ij}, respectively. Also, the MMEs (19.5) are written in the usual form (3), not (5). The only fixed effect in (10) is μ; and ratios of variance components for $\sigma_e^2 G^{-1}$ are taken [287, line before (19.5)] as $\sigma_e^2/\sigma_s^2 = 12$ and $\sigma_e^2/\sigma_d^2 = 10$. Then

$$X = \mathbf{1}_{23} \qquad Z = \{_d\, \mathbf{1}_7 \;\; \mathbf{1}_{11} \;\; \mathbf{1}_5\} \qquad \text{and} \qquad Z_d = \{_d\, \mathbf{1}_5 \;\; \mathbf{1}_2 \;\; \mathbf{1}_3 \;\; \mathbf{1}_8 \;\; \mathbf{1}_5\}.$$

Chapter 20

Analysis of Regression Models

This is all very straightforward although it represents only a drop from the sea of books and papers on regression.

For fitting polynomials [293] one should use orthogonal polynomials, a good description of which is to be found in (the old, but detailed) "The Advanced Theory of Statistics" by M. G. Kendall, 1948, Volume II, pages 146-167. See also Pearson and Hartley (1954) and Robson (1959).

Chapter 21

Analysis of Covariance Models [295]

The analysis of covariance for the one-way classification is described with extensive detail (no matrices) in LMFUD, Chapter 6, pages 169-211. This is, of course, not the model of CRH's Chapter 21; he considers only a numerical example of a two-way classification.

LMFUD Section 11.1 (pages 416-418) highlights some deficiencies of the traditional treatment of analysis of covariance, and in Section 11.2 (pages 419-430) shows how the traditional fixed effects model $\mathrm{E}(y) = X\beta$ of main effects and interactions can be usefully and easily extended to $\mathrm{E}(y) = X\beta + Zb$ where columns of Z are columns of observed covariates, and b is the vector of "regression" coefficients (or "slopes") multiplying those covariates. Table 11.4 of LMFUD shows two appropriate analyses of variance based on $\mathrm{E}(y) = X\beta + Zb$ and Table 11.5 shows the hypotheses that can be tested from those analyses of variance. More general hypotheses are also considered; for example, H: $K'b = p$ and H: $K'\beta = m$. The important feature of this approach to analysis of covariance is that it is directly applicable for balanced and for unbalanced data; and for as many covariables as one wishes, necessarily fewer than N minus the rank of X.

21.1 Two-way fixed model with two covariates [295, 21.1]

LMFUD Sections 11.4 through 11.7 deal with a number of special cases of both the one-way and two-way classifications. Section 11.7 does not deal explicitly with the model of [295, 21.1], but LMFUD Section 11.7a-iii can be adapted thereto. First, write $\mu_{ij} = r_i + c_j + \gamma_{ij}$ in the model adapted to $\mathrm{E}(y_{ijk}) = \mu_{ij} + \alpha_1 w_{1ijk} + \alpha_2 w_{2ijk}$ by replacing, for example $\mathrm{E}_{i\cdot zz}$ with $\mathrm{E}_{\cdot\cdot w_1 w_1}$ which we are writing as E_{11}. Thus doing this in **-iii** of LMFUD 452 gives, from (4)

$$\begin{aligned}\begin{bmatrix}\hat{\alpha}_1\\ \hat{\alpha}_2\end{bmatrix} &= \begin{bmatrix}E_{11} & E_{12}\\ E_{21} & E_{22}\end{bmatrix}^{-1}\begin{bmatrix}E_{1y}\\ E_{2y}\end{bmatrix} = \begin{bmatrix}23.08 & 17.5\\ 17.5 & 30.5\end{bmatrix}^{-1}\begin{bmatrix}2\\ -12\end{bmatrix} \qquad (5)\\ &= \frac{1}{397.7}\begin{bmatrix}30.5 & -17.5\\ -17.5 & 23.08\end{bmatrix}\begin{bmatrix}2\\ -12\end{bmatrix}\\ &= \begin{bmatrix}.681\\ -.784\end{bmatrix}, \qquad (6)\end{aligned}$$

exactly as in the last two elements of the solution vector in the two lines under [297, (21.2)]. The other elements (after the zeros) are calculated from (16) of LMFUD page 422, namely

$$\hat{\beta} = (X'X)^{-}X'(y - Z\hat{b}). \qquad (7)$$

In a two-way classification with-interaction model and no covariates, i.e., $\mathrm{E}(y_{ijk}) = \mu_{ij}$, we know that $(X'X)^{-}X'y$ yields $\hat{\mu}_{ij} = \bar{y}_{ij\cdot} = \frac{1}{n_{ij}}\Sigma_k y_{ijk}$. Therefore, for $\mathrm{E}(y_{ijk} = \mu_{ijk} + \alpha_1 w_{1ijk} + \alpha_2 w_{2ijk}$, (7) yields

$$\hat{\mu}_{ij} = \frac{1}{n_{ij}}\Sigma_k(y_{ijk} - \hat{\alpha}_1 w_{1ijk} - \hat{\alpha}_2 w_{2ijk}) = \bar{y}_{ij\cdot} - \hat{\alpha}_1\bar{w}_{1ij\cdot} - \hat{\alpha}_2\bar{w}_{2ij\cdot} \qquad (8)$$

Thus, for example, for $i = 1$ and $j = 1$:

$$\hat{\mu}_{11} = 20/3 - .681(8/3) - (-.784)(12/3) = 7.987,$$

equalling, as it should, the first element after the six zeros in the line below [297, (21.2)].

The clue to this being easier than the lengthy matrix algebra approach of MMEs (a matrix of order 17 in [296, 21.2)] is the R in (4); as explained in LMFUD page 423, it is a matrix of residuals and for the two-way crossed classification with interaction this involves just within-cell sums of squares (and products) as in (2) and (3). For the same model without interaction the residual sums of squares (and products) are more complicated as in the numerator of equation (51) on LMFUD page 154.

Comment No indication is given as to the parameters corresponding to the 17 columns in the matrix of [246, (21.1)]. They are for r_i and c_j (each three in number), γ_{ij} (nine of them) and α_1 and α_2.

Note If the model is devoid of interactions μ_{ij} is estimated not as $\bar{y}_{ij\cdot}$ but as in equations (81) and (82) and using (76) of LMFUD pages 348-9. And $\bar{w}_{1ij\cdot}$ and $\bar{w}_{2ij\cdot}$ have to be replaced by similar calculations.

21.3 Covariates all equal at the same level of a factor [300, 21.3]

The model equation is, with $w_{ij} = w_i$

$$y_{ij} = \mu + t_i + \gamma w_i + e_{ij}.$$

For this (full) model write $t_i + \gamma w_i \equiv \tau_i$ and so have

$$y_{ij} = \mu + \tau_i + e_{ij}.$$

This is a simple one-way classification with

$$R(\mu, \tau) = \Sigma n_i \bar{y}_{i\cdot}^2 = 3(6^2) + 2(5^2) + 4(10^2) = 558.$$

The μ, γ model equation is

$$y_{ij} = \mu + \gamma w_i + e_{ij}$$

and so OLS yields the standard regression results

$$\hat{\gamma} = \frac{\Sigma n_i w_i \bar{y}_{i\cdot} - \frac{1}{N}\Sigma n_i \bar{y}_{i\cdot}\Sigma n_i w_i}{\Sigma n_i w_i^2 - \frac{1}{N}(\Sigma n_i w_i)^2}$$

and

$$\hat{\mu} = \frac{1}{N}(\Sigma n_i \bar{y}_{i\cdot} - \hat{\gamma}\Sigma n_i w_i).$$

And so after a little simplication

$$\begin{aligned} R(\mu, \gamma) &= \hat{\gamma}\Sigma n_i w_i \bar{y}_{i\cdot} + \hat{\mu}\Sigma n_i \bar{y}_{i\cdot} \\ &= \frac{(\Sigma n_i w_i \bar{y}_{i\cdot} - \frac{1}{N}\Sigma n_i \bar{y}_{i\cdot}\Sigma n_i w_i)^2}{\Sigma n_i w_i^2 - \frac{1}{N}(\Sigma n_i w_i)^2} + \frac{(\Sigma n_i \bar{y}_{i\cdot})^2}{N} \\ &= \frac{[276 - 68(34)/9]^2}{144 - 34^2/9} + \frac{68^2}{9} \\ &= 23.48 + 513.77 = 537.25 \end{aligned}$$

Chapter 22

Animal Model, Single Records [303]

The important aspect of the model equation

$$y = X\beta + Zu + Z_a a + e \tag{22.1}$$

is that it represents random effects *other than breeding values* [303, line 9]. This does not affect the treatment of u as being random; it is just a matter of what random effects u represents.

Equations [304, (22.2) – (22.4)] flow very easily from the usual MMEs, e.g., [16, (3.4)].

22.1 Example with daughter-dam pairs [304, (22.1)]

It is a pity that no model equation is given for this example. It is clearly

$$y_{ij} = p_i + a_{ij} + e_{ij} \tag{1}$$

where $i = 1, 2$ for the periods, $j = 1, \ldots, 5$ for each i, a_{1j} is a dam j's record (made in period 1) and a_{2j} is daughter j's record (made in period 2), the daughter of dam j. Thus dam-daughter comparisons are confounded with periods.

Consideration of (1) reveals that X is certainly as at [304, bottom]; and that u and Z do not exist.

Error Therefore in [305, first line] it is not Z which is I, but $Z_a = I$.

The MMEs are therefore [304, (22.4)] without Z, and so have the form

$$\begin{bmatrix} X'X & X' \\ X & I + A^{-1}\sigma_e^2/\sigma_a^2 \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{p}} \\ \tilde{a} \end{bmatrix} = \begin{bmatrix} X'y \\ y \end{bmatrix}. \tag{2}$$

for $\lambda = -10$ and

$$F^{-1} = (23I_5 - .6J_5)^{-1} = \frac{1}{23}\left(I_5 \quad + \frac{.6}{23 - 5(.6)}J_5\right) = \frac{1}{23}(I_5 + .03J_5).$$

Therefore

$$\begin{aligned}
(F - \lambda^2 F^{-1})^{-1} &= \left[23I_5 - .6J_5 - 100\frac{1}{23}(I_5 + .03J_5)\right]^{-1} \\
&= \left[\frac{1}{23}(429I_5 - 16.8J_5)\right]^{-1} \\
&= \frac{23}{429}\left(I_5 + \frac{16.8}{429 - 5(16.8)}J_5\right) \\
&= \frac{23}{429}\left(I_5 + \frac{16.8}{345}J_5\right). \qquad (5)
\end{aligned}$$

From the right-hand side of (3) we see that $(W - .2XX')^{-1}$ is the part of (22.6) corresponding to $\begin{bmatrix} P & Q \\ Q & P \end{bmatrix}$. Thus from (4) we similarly see that $3(F - \lambda^2 F^{-1})^{-1}$ corresponds to P; and (5) gives

$$P = 3(F - \lambda^2 F^{-1})^{-1} = .16084I_5 + .00783J_5,$$

which has diagonal elements .16084 + .00783 = .16867 and off-diagonal elements .00783 as prescribed for P in [305, line below (22.6)]. Similarly we get from (4)

$$\begin{aligned}
Q &= 3\left(-\lambda F^{-1}\right)\left(F - \lambda^2 F^{-1}\right)^{-1} \\
&= 30\frac{1}{23}(I_5 + .03J_5)\frac{23}{429}\left(I_5 + \frac{16.8}{345}J_5\right) \\
&= \frac{10}{143}\left[I_5 + J_5\left(.03 + \frac{.03(5)(16.8)}{345} + \frac{16.8}{345}\right)\right] \\
&= .06993I_5 + .00601J_5
\end{aligned}$$

which has .07594 in diagonals and .00601 in off-diagonals, just as in [305, 3rd line up]. This is further confirmed by looking at the first row of [305, (22.6)]. It comes from (3) as

$$-.2X'\begin{bmatrix} P & Q \\ Q & P \end{bmatrix} = -.2\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} P & Q \\ Q & P \end{bmatrix}$$

and so

$$\begin{aligned}
-.2\mathbf{1}_5'[P \quad Q] &= -.2\mathbf{1}'[.16084I_5 + .00783J_5 \quad .06993I_5 + .00601J + 5] \\
&= -.2\left\{[.16084 + 5(.00783)]\mathbf{1}_5' \quad [.06993 + 5(.00601)]\mathbf{1}_5'\right\} \\
&= [-.04\mathbf{1}_5' \quad -.02\mathbf{1}_5']
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2}\sigma_a^2 + \sigma_e^2 && (8)\\
\text{E(MSE)} &= \frac{1}{8}\text{tr}\left\{\left(\sigma_a^2\begin{bmatrix} I_5 & \frac{1}{2}I_5 \\ \frac{1}{2}I_5 & I_5 \end{bmatrix} + \sigma_e^2 I_{10}\right)\begin{pmatrix} I-\bar{J}_5 & 0 \\ 0 & I-\bar{J}_5 \end{pmatrix}\right\} \\
&= \frac{1}{8}[\sigma_a^2 2(5-1) + \sigma_e^2 2(5-1)] \\
&= \sigma_a^2 + \sigma_e^2. && (9)
\end{aligned}
$$

There is no surprise in this result because in the model equation (1) a_{ij} and e_{ij} are indistinguishable. It is only the $\frac{1}{2}I_5$ submatrices in A that cause E(MSP) to be something other than $\sigma_a^2 + \sigma_e^2$.

Now as to estimation: equating (8) to (6) and (9) to (7) gives

$$\frac{1}{2}\hat{\sigma}_a^2 + \hat{\sigma}_e^2 = 2.5 \quad \text{and} \quad \hat{\sigma}_a^2 + \hat{\sigma}_e^2 = 2.5. \tag{10}$$

These clearly have solution $\hat{\sigma}_e^2 = 2.5$ and $\hat{\sigma}_a^2 = 0$. CRH [306, last line first paragraph] has $\hat{\sigma}_e^2 = 2$ and $\hat{\sigma}_a^2 = .5$. These satisfy the second equation in (10) but not the first. There is no non-zero solution for $\hat{\sigma}_a^2$ only because the two mean squares are equal. Note though that, in general, with

$$
\begin{aligned}
\frac{1}{2}\hat{\sigma}_a^2 + \hat{\sigma}_e^2 = \text{MSP} \qquad &\text{and} \qquad \hat{\sigma}_a^2 + \hat{\sigma}_e^2 = \text{MSE} \\
\hat{\sigma}_a^2 = 2(\text{MSE} - \text{MSP}) \qquad &\text{and} \qquad \hat{\sigma}_e^2 = 2\text{MSP} - \text{MSE}.
\end{aligned}
$$

Thus if

$$
\begin{array}{lllll}
\text{MSP} < \text{MSE} < 2\text{MSP} & \hat{\sigma}_a^2 > 0 & \text{and} & \hat{\sigma}_e^2 > 0 \\
\text{MSP} > \text{MSE} & \hat{\sigma}_a^2 < 0 & \text{and} & \hat{\sigma}_e^2 > 0 \\
\text{MSE} > 2\text{MSP} & \hat{\sigma}_a^2 > 0 & \text{and} & \hat{\sigma}_e^2 < 0.
\end{array}
$$

The calculations for MIVQUE for the unbalanced data would be illustrative if some details were shown and not just computed results.

Chapter 23

Sire Model, Single Records

Equations [310, (23.3)] are the same as [304, (22.2)] except for the sequencing of the parameters: in (22.2) they are β, u, a but in (23.3) they are β, s, u. And (23.4) corresponds to neither (22.3) nor (22.4); (23.4) is (23.3) with $R = \sigma_e^2 I$; (22.4) also has $R = \sigma_e^2 I$ but only after having $Z_a = I$.

23.1 MMEs

For the data of [310, table] there is no u, and hence no Z; and in (23.5) and (23.6) the parameter h is for fixed, herd effects; so after deleting u and Z from (23.4) and putting β (now h) after s and not before it, (23.4) becomes

$$\begin{bmatrix} Z_s'Z_s + A^{-1}\sigma_e^2/\sigma_s^2 & Z_s'X \\ X'Z_s & X'X \end{bmatrix} \begin{bmatrix} \tilde{s} \\ \hat{h} \end{bmatrix} = \begin{bmatrix} Z_s'y \\ X'y \end{bmatrix}$$

with

$$Z_s\mathbf{s} = \begin{bmatrix} 1_8 & \cdot & \cdot \\ \cdot & 1_{12} & \cdot \\ \cdot & \cdot & 1_{20} \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix} \quad \text{and} \quad X\mathbf{h} = \begin{bmatrix} 1_3 & \cdot & \cdot & \cdot \\ \cdot & 1_5 & \cdot & \cdot \\ \cdot & 1_8 & \cdot & \cdot \\ \cdot & \cdot & 1_4 & \cdot \\ 1_4 & \cdot & \cdot & \cdot \\ \cdot & 1_2 & \cdot & \cdot \\ \cdot & \cdot & 1_6 & \cdot \\ \cdot & \cdot & \cdot & 1_8 \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \end{bmatrix}. \tag{1}$$

Thus (23.5) is

$$\begin{bmatrix} Z_s'Z_s & Z_s'X \\ X'Z_s & X'X \end{bmatrix} \begin{bmatrix} \hat{h} \\ \hat{s} \end{bmatrix} = \begin{bmatrix} Z_s'y \\ X'y \end{bmatrix}.$$

This result is given in Searle (1997), wherein it is also pointed out that for $G = \{_d\, \sigma_i^2 I_{q_i}\}$ λ can always be null except for a subvector $\mathbf{1}$, so leading to sums of BLUPs being zero. But when G is not of that form, $ZG\lambda = X\tau$ has to be used explicitly. However, so long as X is an incidence matrix, with no column of covariables, $X\mathbf{1}_{p\times 1} = f\mathbf{1}_{N\times 1}$ when there are f fixed effects factors and so it may often be useful to take $\tau = \mathbf{1}_{p\times 1}$.

In the example of this section, Z and X are given in (1) and, because in (3) any scalar emanating from G can be ignored, we overlook the $\sigma_e^2/12$ in var(s) of [310, three lines above (23.5)], and use

$$G = A = \begin{bmatrix} 1 & .5 & .5 \\ .5 & 1 & .25 \\ .5 & .25 & 1 \end{bmatrix}.$$

Then with

$$X\tau = \begin{bmatrix} \tau_1 \mathbf{1}_3' & \tau_2 \mathbf{1}_{13}' & \tau_3 \mathbf{1}_4' & \tau_1 \mathbf{1}_4' & \tau_2 \mathbf{1}_2' & \tau_3 \mathbf{1}_6' & \tau_4 \mathbf{1}_8' \end{bmatrix}'$$

putting $\tau_1 = \tau_2 = \tau_3 = \tau_4 = 1$ gives (3) as

$$ZG\lambda = ZA\lambda = \begin{bmatrix} \mathbf{1}_8(\lambda_1 + .5\lambda_2 + .5\lambda_3) \\ \mathbf{1}_{12}(.5\lambda_1 + \lambda_2 + .25\lambda_3) \\ \mathbf{1}_{20}(.5\lambda_1 + .25\lambda_2 + \lambda_3) \end{bmatrix} = \mathbf{1}_{40}.$$

Therefore we want

$$\begin{bmatrix} 1 & .5 & .5 \\ .5 & 1 & .25 \\ .5 & .25 & 1 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} = \mathbf{1}_3 \tag{7}$$

which has solution proportional to

$$\lambda' = [1 \quad 2 \quad 2].$$

Thus for BLUP(s) from (2)

$$\lambda'[\text{BLUP}(s)] = 1(-.036661) + 2(.453353) + 2(-.435022) = -.000001,$$

i.e., $\lambda'[\text{BLUP(s)}] = 0$, correct to five decimal places. Note in passing that (7) is

$$\lambda = G^{-1}\mathbf{1}.$$

Note, for example: $c_{11}+c_{12}+c_{13}=0 \quad c_{12}+c_{22}+c_{23}=0$. Also

$$\begin{aligned} r_1 &= 59-[3(48/7)+5(119/15)]=-1.2381 \\ r_2 &= 105-[8(119/15)+4(74/10)]=11.9333 \\ r_3 &= 150-[4(48/7)+2(119/15)+6(74/10)+8(73/8)]=-10.6952; \end{aligned}$$

and note that $r_1+r_2+r_3=0$. Then, as in LM, page 267, equation (16), for C of order 2×2

$$\begin{aligned} \begin{bmatrix} s_1^\circ \\ s_2^\circ \end{bmatrix} = C^{-1}\mathbf{r} &= \begin{bmatrix} 5.0476 & -2.6666 \\ -2.6666 & 6.1333 \end{bmatrix}^{-1} \begin{bmatrix} -1.2381 \\ 11.9333 \end{bmatrix} \\ &= (.0419)\begin{bmatrix} 6.1333 & 2.6666 \\ 2.6666 & 5.0476 \end{bmatrix}\begin{bmatrix} -1.2381 \\ 11.9333 \end{bmatrix} \\ &= \begin{bmatrix} 1.0156 \\ 2.3854 \end{bmatrix}. \end{aligned}$$

And $s_3^\circ=0$. Notation s° rather than $\hat{s}$ is used because the calculated values are only solutions (to OLS equations) not estimates of s. Then, as in LMFUD page 102, equation (68)

$$h_i^\circ = \bar{y}_{i\cdot\cdot} - \Sigma_{j=1}^2 n_{ij}s_j^\circ/n_{i\cdot\cdot}$$

Thus

$$\begin{aligned} h_1^\circ &= 48/7-3(1.0156)/7 &&= 6.4219 \\ h_2^\circ &= 119/15-[5(1.01561+8(2.3874)]/15 &&= 6.3215 \\ h_3^\circ &= 74/10-4(2.3874)/10 &&= 6.4450 \\ h_4^\circ &= 73/8-0 &&= 9.1250. \end{aligned}$$

Arraying these in a vector $[s^{\circ\prime} \;\; h^{\circ\prime}]$, in keeping with the solution [311, line 7 up], gives

$$[s^{\circ\prime} \;\; h^{\circ\prime}] = [1.0156 \quad 2.3874 \quad 0 \quad 6.4219 \quad 6.3215 \quad 6.4450 \quad 9.1250].$$

This looks very different from the [311] solution:

$$[10.14097 \quad 11.51238 \quad 9.125 \quad -2.70328 \quad -2.80359 \quad -2.67995 \quad 0].$$

But this is where estimability comes in. We are dealing with a no-interaction model, and every difference between elements of s is estimable, as is that between elements of **h**. Examples follow (to 3 decimal places).

$$\begin{aligned} \text{BLUE}(s_1-s_2) &= s_2^\circ - s_1^\circ &&= 2.387-1.016 &&= 1.371 &&= 11.512-10.141 \\ \text{BLUE}(s_2-s_3) &= s_2^\circ - s_3^\circ &&= 2.387-0 &&= 2.387 &&= 11.512-9.125 \\ \text{BLUE}(h_1-h_2) &= h_1^\circ - h_2^\circ &&= 6.422-6.322 &&= .100 &&= -2.703-2.801 \\ \text{BLUE}(h_3-h_4) &= h_3^\circ - h_4^\circ &&= 6.4450-9.125 &&= -2.680 &&= -2.680-0. \end{aligned}$$

Chapter 24

Animal Model, Repeated Records

With $\mathbf{c} = \mathbf{a} + \mathbf{p}$ as in [314, (24.3)] it is only the use of A^{-1} corresponding to $\mathbf{a}$ in the MMEs [314, (24.4)] which distinguishes it, in terms of estimation, from $\mathbf{p}$. Indeed, $\tilde{\mathbf{p}}$ and $\tilde{\mathbf{a}}$ are linearly related: $\tilde{\mathbf{p}} = (\sigma_p^2/\sigma_a^2)A^{-1}\tilde{a}$ as in (24.5).

The third equation on [315] is missing I on its left side, so that it is $\sigma_e^2 I = .55 I\sigma_y^2$. By considering diagonal elements only, the first equation gives (because $a_{ii} = 1$ always)

$$\sigma_a^2 = .25\sigma_y^2, \qquad \text{and} \qquad \sigma_p^2 = .2\sigma_y^2 \qquad \text{and} \qquad \sigma_e^2 = .55\sigma_y^2.$$

Therefore

$$h^2 \equiv \sigma_a^2/\sigma_y^2 = .25; \qquad \text{and} \qquad r = \sigma_e^2/\sigma_y^2 = (\sigma_a^2 + \sigma_p^2)/\sigma_y^2 = .25 + .2 = .45.$$

For the example there is no u or Z. For the OLS equations

$$\mathrm{X}' = \begin{bmatrix} 1 & 1 & 1 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 & 1 & 1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & 1 \end{bmatrix} \qquad \text{and} \qquad Z_c' = \begin{bmatrix} 1 & \cdot & \cdot & 1 & \cdot & \cdot & 1 & \cdot \\ \cdot & 1 & \cdot & \cdot & 1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & 1 \\ \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}.$$

For the MMEs of [316, (24.8)] the 2.75 added to diagonal elements of $Z_c'Z_c$ is $\hat{\sigma}_e^2/\hat{\sigma}_p^2$ of [314, (24.4)], its value being $.55\sigma_y^2/.2\sigma_y^2 = 2.75$. And the 2.2 is added to $\{_d\ 3\ \ 2\ \ 2\ \ 1\}$ in the form $2.2A^{-1}$, the 2.2 being $\sigma_e^2/\sigma_a^2 = .55/.25 = 2.2$.

In the solution to the MMEs below [316, (24.9)] the elements of BLUP(p) add to zero but those of BLUP(a) do not. This is because $I\sigma_e^2/\sigma_p^2$ is diagonal but $A^{-1}\sigma_e^2/\sigma_a^2$ is not (see these notes at Section 23.1).

Chapter 25

Sire Model, Repeated Records

Typo In [321, line 4 up] the $X_p'X_p$ should be $Z_p'Z_p$.

For the example on [322]

$$
\begin{aligned}
X &= \{_d\, 1_3 \;\; 1_6 \;\; 1_4 \;\; 1_3 \;\; 1_5 \;\; 1_6 \;\; 1_4\} \\
Z_s' &= \begin{bmatrix} 1_2' & \cdot & 1_3' & \cdot & 1_2' & \cdot & 1_2' & \cdot & 1_3' & \cdot & 1_3' & \cdot & 1_2' & \cdot \\ \cdot & 1_1' & \cdot & 1_3' & \cdot & 1_2' & \cdot & 1_1' & \cdot & 1_2' & \cdot & 1_3' & \cdot & 1_2' \end{bmatrix}.
\end{aligned}
$$

Z_p will have order 31×14, corresponding to the 31 records and the 14 different progeny. To write down Z_p one needs to keep in mind the sequencing of the parameters in the parameter vector, namely $hy_{11} \cdots hy_{24} \;\; s_1 \;\; s_2 \;\; p_1 \cdots p_{14}$. To assist readability Z_p is shown on the next page with row and column numbers, row numbers being the records ordered by progeny within herd-year, and the column numbers are the progeny numbers.

Note in the BLUP solution on [323] that both BLUP(s) and BLUP(p) have elements which sum to zero.

The top of [324] has the appearance of a solution vector; but it is not. It is a vector of unknowns which is to be premultiplied by the 9×9 matrix at the bottom of [323]. The solution vector is the inverse of that matrix premultiplying the 9×1 vector atop [324].

In the final equation of [324] both off-diagonal terms should have two 2s, and on the right-hand side 1.1870 should be 1.1187 (the signs are correct). The solution for s_2 in that equation is then as on [323]:

$$\frac{1.11870(12.26353 - 5.22353)}{12.26353^2 - 5.22353^2} = .06397.$$

Chapter 26

Animal Model, Multiple Traits

The algebra as presented in this chapter is horrendous. Some of it can be abbreviated; and for some of it a small example ($t = 2$ traits) helps understand the general case. We make use of the direct (Kronecker) product operation that $K \otimes L = \{k_{ij}L\}$.

26.1 No missing data [325, (26.1)]

The model equation is, for $i = 1, 2, \cdots, t$

$$\{_c\, \mathbf{y}_i\} = \{_d\, \mathbf{X}_i\}\{_c\, \boldsymbol{\beta}_i\} + \{_d\, \mathbf{I}_n\}\{_c\, \mathbf{a}_i\} + \{_c\, \mathbf{e}_i\}\,. \tag{26.1}$$

Define

$$G_0 = \{_m\, g_{ij}\}_{i,j=1}^{t} \qquad \text{and} \qquad R_0 = \{_m\, r_{ij}\}_{i,j=1}^{t}\,. \tag{1}$$

Then

$$\begin{aligned} \text{var}\,\{_c\, \mathbf{a}_i\} &= G_0 \otimes A = G & (26.3)\\ \text{var}\,\{_c\, \mathbf{e}_i\} &= R_0 \otimes I = R & (26.4)\\ G^{-1} &= G_0^{-1} \otimes A^{-1} & (26.5)\\ R^{-1} &= R_0^{-1} \otimes I & (26.6) \end{aligned}$$

Define

$$R_0^{-1} = \{_m\, r^{ij}\}_{i,j=1}^{t} = \{_m\, \rho_{ij}\}_{i,j=1}^{t}\,. \tag{2}$$

[326] initiates notation r^{ij} for elements of R_0^{-1}; writing $r^{ij} \equiv \rho_{ij}$ makes for easier readability.

26.1.2 Confirming a variance

To find var[BLUP(**a**)] demands knowing the variance-covariance matrix of the right-hand side of (26.7). From (4), and then using (3), (26.3) and (26.4), this is

$$\begin{aligned}
\text{var}(rhs) &= \text{var}\left\{\begin{bmatrix} X_1' & \cdot \\ \cdot & X_2' \\ I & \cdot \\ \cdot & I \end{bmatrix}\begin{bmatrix} \rho_{11}I & \rho_{12}I \\ \rho_{21}I & \rho_{22}I \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\right\} \\
&= \begin{bmatrix} X_1' & \cdot \\ \cdot & X_2' \\ I & \cdot \\ \cdot & I \end{bmatrix}\begin{bmatrix} \rho_{11}I & \rho_{12}I \\ \rho_{21}I & \rho_{22}I \end{bmatrix}\text{var}\left\{\begin{bmatrix} I & \cdot \\ \cdot & I \end{bmatrix}\begin{bmatrix} a_1 \\ a_2 \end{bmatrix}+\begin{bmatrix} e_1 \\ e_2 \end{bmatrix}\right\}\begin{bmatrix} \rho_{11}I & \rho_{12}I \\ \rho_{21}I & \rho_{22}I \end{bmatrix}\begin{bmatrix} X_1 & \cdot & I & \cdot \\ \cdot & X_2 & \cdot & I \end{bmatrix} \\
&= \begin{bmatrix} X_1' & \cdot \\ \cdot & X_2' \\ I & \cdot \\ \cdot & I \end{bmatrix}\begin{bmatrix} \rho_{11}I & \rho_{12}I \\ \rho_{21}I & \rho_{22}I \end{bmatrix}\left\{\begin{bmatrix} I & \cdot \\ \cdot & I \end{bmatrix} G_0 \otimes A \begin{bmatrix} I & \cdot \\ \cdot & I \end{bmatrix} + R_0 \otimes I\right\} \\
&\qquad \times \begin{bmatrix} \rho_{11}I & \rho_{12}I \\ \rho_{21}I & \rho_{22}I \end{bmatrix}\begin{bmatrix} X_1 & \cdot & I & \cdot \\ \cdot & X_2 & \cdot & I \end{bmatrix} \\
&= \begin{bmatrix} X_1' & \cdot \\ \cdot & X_2' \\ I & \cdot \\ \cdot & I \end{bmatrix}(R_0^{-1} \otimes I)\left[(G_0 \otimes A) + (R_0 \otimes I)\right](R_0^{-1} \otimes I)\begin{bmatrix} X_1 & \cdot & I & \cdot \\ \cdot & X_2 & \cdot & I \end{bmatrix} \\
&= \begin{bmatrix} X_1' & \cdot \\ \cdot & X_2' \\ I & \cdot \\ \cdot & I \end{bmatrix}\left[(R_0^{-1}G_0R_0^{-1} \otimes A)\right] + (R_0^{-1} \otimes I)\begin{bmatrix} X_1 & \cdot & I & \cdot \\ \cdot & X_2 & \cdot & I \end{bmatrix}. \qquad (6)
\end{aligned}$$

For this

$$\begin{aligned}
R_0^{-1}G_0R_0^{-1} &= \begin{bmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{bmatrix}\begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix}\begin{bmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{bmatrix} \\
&= \begin{bmatrix} \rho_{11}g_{11}+\rho_{12}g_{21} & \rho_{11}g_{12}+\rho_{12}g_{22} \\ \rho_{21}g_{11}+\rho_{22}g_{21} & \rho_{21}g_{12}+\rho_{22}g_{22} \end{bmatrix}\begin{bmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{bmatrix} \\
&= \begin{bmatrix} \rho_{11}g_{11}\rho_{11}+\rho_{12}g_{21}\rho_{11}+\rho_{11}g_{12}\rho_{21}+\rho_{12}g_{22}\rho_{21} & \rho_{11}g_{11}\rho_{12}+\rho_{12}g_{21}\rho_{12}+\rho_{11}g_{12}\rho_{22}+\rho_{12}g_{22}\rho_{22} \\ \rho_{21}g_{11}\rho_{11}+\rho_{22}g_{21}\rho_{11}+\rho_{21}g_{12}\rho_{21}+\rho_{22}g_{22}\rho_{21} & \rho_{21}g_{11}\rho_{12}+\rho_{22}g_{21}\rho_{12}+\rho_{21}g_{12}\rho_{22}+\rho_{22}g_{22}\rho_{22} \end{bmatrix} \\
&= \begin{bmatrix} \rho_{11}^2g_{11}+2\rho_{11}\rho_{12}g_{12}+\rho_{12}^2g_{22} & \rho_{11}\rho_{21}g_{11}+\rho_{11}\rho_{22}g_{12}+\rho_{12}^2g_{21}+\rho_{12}\rho_{22}g_{22} \\ \rho_{11}\rho_{21}g_{11}+\rho_{11}\rho_{22}g_{21}+\rho_{21}^2g_{12}+\rho_{21}\rho_{22}g_{22} & \rho_{12}^2g_{11}+2\rho_{12}\rho_{22}g_{21}+\rho_{22}^2g_{22} \end{bmatrix} \qquad (7)
\end{aligned}$$

Now in (6) the matrix (7) has to be used in a direct product with A on its right: that means each term in each sum in each element of (7) will multiply A. And then the whole matrix is to be post-multiplied by $W = \begin{bmatrix} X_1 & \cdot & I & \cdot \\ \cdot & X_2 & \cdot & I \end{bmatrix}$ and pre-multiplied by W'. Taking all this into account, inspection of (7) reveals that the matrix multiplying g_{11} is

26.1.3 Confirming a matrix

The paragraph below (26.14) leads to writing the upper right-hand 2×2 submatrix of what we have in (9) as

$$\begin{bmatrix} X_1'X_1\rho_{11}^2 & X_1'X_2\rho_{11}\rho_{21} \\ X_2'X_1\rho_{11}\rho_{21} & X_2'X_2\rho_{22}^2 \end{bmatrix} r_{11} + \begin{bmatrix} 2X_1'X_1\rho_{11}\rho_{12} & X_1'X_2(\rho_{11}\rho_{22} + \rho_{12}\rho_{21}) \\ \text{Sym.} & 2X_2'X_2\rho_{21}\rho_{22} \end{bmatrix} r_{12}$$

$$+ \begin{bmatrix} X_1'X_1\rho_{12}^2 & X_1'X_2\rho_{12}\rho_{22} \\ X_2'X_1\rho_{12}\rho_{22} & X_2'X_2\rho_{22}^2 \end{bmatrix} r_{22}. \tag{10}$$

We show that (10) is the same as that leading 2×2 in (9). To do this consider the coefficient of $X_1'X_1$ in (10):

$$\begin{aligned} c(X_1'X_1) &= \rho_{11}^2 r_{11} + 2\rho_{11}\rho_{12}r_{12} + \rho_{12}^2 r_{22} \\ &= \rho_{11}^2\rho_{22}\Delta + 2\rho_{11}\rho_{12}(-\rho_{12}\Delta) + \rho_{12}^2\rho_{11}\Delta \end{aligned}$$

for $\Delta = r_{11}r_{22} - r_{12}^2 = 1/(\rho_{11}\rho_{22} - \rho_{12}^2)$ coming from

$$R_0^{-1} = \begin{bmatrix} r_{11} & r_{12} \\ r_{12} & r_{22} \end{bmatrix}^{-1} = \frac{1}{\Delta}\begin{bmatrix} r_{22} & -r_{12} \\ -r_{12} & r_{11} \end{bmatrix} = \begin{bmatrix} \rho_{11} & \rho_{12} \\ \rho_{12} & \rho_{22} \end{bmatrix}.$$

Thus

$$c(X_1'X_1) = \rho_{11}\Delta(\rho_{11}\rho_{22} - \rho_{12}^2) = \rho_{11},$$

which is the coefficient of $X_1'X_1$ in (9). Likewise from (10)

$$\begin{aligned} c(X_1'X_2) &= \rho_{11}\rho_{21}r_{11} + (\rho_{11}\rho_{22} + \rho_{12}^2)r_{12} + \rho_{12}\rho_{22}r_{22} \\ &= \Delta[\rho_{11}\rho_{21}\rho_{22} + (\rho_{11}\rho_{22} + \rho_{12}^2)(-\rho_{12}) + \rho_{12}\rho_{22}\rho_{11}] \\ &= \Delta\rho_{12}(\rho_{11}\rho_{22} - \rho_{12}^2) \\ &= \rho_{12}, \end{aligned}$$

and

$$\begin{aligned} c(X_2'X_2) &= \rho_{12}^2 r_{11} + 2\rho_{21}\rho_{22}r_{12} + \rho_{22}^2 r_{22} \\ &= \rho_{22}\Delta(\rho_{11}\rho_{22} - \rho_{12}^2) \\ &= \rho_{22}. \end{aligned}$$

And these *are* the coefficients in (9). Thus [9] agrees with CRH's description below [328. (26.14)].

Also in line three of that paragraph what is meant by "sequential"? Apparently this means 1, or 1,2 or 1,2,3,$\cdots$ and so on. Traits may occur over time: e.g., weights at birth and successive ages.

In [331, (26.31)] the a_{ij}-terms are elements of the relationship matrix A; they are not the a_{ij}s representing animal effects in the model equations on [330]. G_0 in (26.31) is the same as in (26.3) where it occurred in the form $G_i \otimes A$. But now it is

$$\text{var}(\mathbf{a}) = A \otimes G_0. \tag{26.31}$$

And then

$$[\text{var}(\mathbf{a})]^{-1} = A^{-1} \otimes G_0^{-1}. \tag{26.32}$$

For the incidence matrix (26.33) the parameter vector is

$$[\beta_{11}\ \ \beta_{12}\ \ \beta_{21}\ \ \beta_{31}\ \ \beta_{32}\ \ a_{11}\ \ a_{12}\ \ a_{13}\ \ a_{21}\ \ a_{22}\ \ a_{23}\ \ a_{31}\ \ a_{32}\ \ a_{33}]'.$$

And, on ordering the records by traits within animals, with zero for a missing record, the y-vector is

$$y = [5\ \ 3\ \ 6\ \ 2\ \ 5\ \ 7\ \ 0\ \ 3\ \ 4\ \ 2\ \ 0\ \ 0]'$$

The three matrices at the bottom of [332] are, respectively,

$$\begin{bmatrix} 5 & 3 & 1 \\ 3 & 6 & 4 \\ 1 & 4 & 7 \end{bmatrix}^{-1}, \quad \begin{bmatrix} 0 & 0 & 0 \\ 0 & \multicolumn{2}{c}{\begin{pmatrix} 6 & 4 \\ 4 & 7 \end{pmatrix}^{-1}} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 5^{-1} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

[333, (26.34)] is $A \otimes G_0$; and the remainder of the section, namely the right-hand vector, the solution vector and the 17×17 matrix in (26.35) – (26.37) involve too much arithmetic for verification here.

26.3 The EM Algorithm [334, (26.3)]

This is primarily a section on computing, so I offer no comment.

Chapter 27

Sire Model, Multiple Traits

27.1 Only One Trait Observed on a Progeny [341, (27.1)]

Equation [341, (27.1)] is the same form as [325, (26.1)] except for $\{_d\ I\}\{_c\ \mathbf{a}_i\}$ now being $\{_d\ Z_i\}\{_c\ \mathrm{s}_i\}$

Typo The sentence in [342, line 1] is unfathomable.

$$
\begin{aligned}
\mathrm{var(s)} &= B \otimes A = G & (27.2)\\
B = \{_m\ b_{ij} = g_{ij}/4\} &= G_0/4 \\
\mathrm{var(e)} &= D \otimes I = R & (27.3)\\
&= \{_c\ d_i = .75g_{ii} + r_{ii}\} \otimes I.
\end{aligned}
$$

For fixed s GLS equations are

$$
\begin{bmatrix} \{_d\ X_i'X_i/d_i\} & \{_d\ X_i'Z_i/d_i\} \\ \{_d\ Z_i'X_i/d_i\} & \{_d\ Z_i'Z_i/d_i\} \end{bmatrix} \begin{bmatrix} \beta^0 \\ \tilde{s} \end{bmatrix} = \begin{bmatrix} \{_c\ X_i'y_i/d_i\} \\ \{_c\ Z_i'y_i/d_i\} \end{bmatrix}. \tag{27.4}
$$

From (27.2)

$$
G^{-1} = B^{-1} \otimes A^{-1} = 4G^{-1} + 0 \otimes A^{-1} \tag{27.5}
$$

The *raison d'être* for [343. (27.6)] is that "it seems logical to estimate d_i" as

$$
\hat{d}_i = \frac{y'y - \beta_i^{0'} X_i'y_i - u_i^{0'} Z_i'y}{n_i - \mathrm{rank}[X_i\ \ Z_i]}, \tag{27.6}
$$

wherein β_i^0 and u_i^0 are solutions to

$$
\begin{bmatrix} X_i'X_i & X_i'Z_i \\ Z_i'X_i & Z_i'Z_i \end{bmatrix} \begin{bmatrix} \beta_i^0 \\ u_i^0 \end{bmatrix} = \begin{bmatrix} X_i'y_i \\ Z_i'y_i \end{bmatrix}. \tag{27.7}
$$

Then R of (27.10) for these data is

$$R = \begin{bmatrix} r_{11} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} & r_{12} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\ r_{12} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} & r_{22} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{bmatrix}.$$

27.3 Relationship to Sire Model with Repeated Records on Progeny [348, 27.3]

No comment.

Chapter 28

Joint Cow and Sire Evaluation

28.1 Block diagonality of MMEs [349, (28.1)]

A straightforward description.

28.2 Single Record on Single Trait [351, (28.2)]

The mid-page description on [350] applies to the animals of the example mid-page [351]. From the genetic relationships among those animals comes A of (28.1): for example, animal 4 is the progeny of 1 and so the relationship is .5; and animal 11 is a granddaughter of 4 (through 2) and is also a niece of 4 (through 1) and so the relationship is .25 + .125 = .375 as seen in the second row of (28.1). Then A^{-1} is (28.2).

The vector mid-page [352] is sequenced in accord with the description on [350] with the herd effects μ_1 and μ_2 in amongst the a_is. And (28.3) is L, say, with $L = [Z_1 \quad \mathbf{x}_1 \quad Z_2 \quad \mathbf{x}_2 \quad Z_3]$ where $Z_1 = 0$ is for the males a_1, a_4 and a_5 which have no records; $\mathbf{x}_1$ and $\mathbf{x}_2$ are for μ_1 and μ_2, respectively; Z_2 is for the animals 2, 6, 8, 11 which have records in herd 1, and Z_3 is for animals 3, 7, 9 and 10 with records in herd 2.

The matrix in the MMEs of (28.4) is then $L'L$ with $3A^{-1}$ added to all elements pertaining to $Z_i'Z_{i'}$ for $i, i' = 1, 2, 3$. Thus the upper-most left-hand 3×3 is

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + 3 \begin{bmatrix} 2 & -1 & -1 \\ & 3 & 0 \\ & & 3 \end{bmatrix} = \begin{bmatrix} 6 & -3 & -3 \\ & 9 & 0 \\ & & 9 \end{bmatrix}$$

28.6 Gametic model to reduce the number of equations [358, 28.6]

28.6.1 Single record model [359, (28.6.1)]

For the example the model is

$$\begin{bmatrix} 5 \\ 3 \\ 2 \\ 8 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 3 \end{bmatrix} \beta + \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{bmatrix} \mathbf{u} + \mathbf{e}.$$

OLS equations, using $1^2 + 2^2 + 1^2 + 3^2 = 15$ and $1(5) + 2(3) + 1(2) + 3(8) = 37$ are

$$\begin{bmatrix} 15 & 1 & 2 & 1 & 3 \\ 1 & 1 & & & \\ 2 & & 1 & & \\ 1 & & & 1 & \\ 3 & & & & 1 \end{bmatrix} \begin{bmatrix} \beta \\ u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} = \begin{bmatrix} 37 \\ 5 \\ 3 \\ 2 \\ 8 \end{bmatrix} \tag{2}$$

$$A = \begin{bmatrix} 1 & 0 & .5 & .5 \\ 0 & 1 & .5 & 0 \\ .5 & .5 & 1 & .25 \\ .5 & 0 & .25 & 1 \end{bmatrix} \qquad A^{-1} = \frac{1}{6} \begin{bmatrix} 11 & 3 & -6 & -4 \\ 3 & 9 & -6 & 0 \\ -6 & -6 & 12 & 0 \\ -4 & 0 & 0 & 8 \end{bmatrix}. \tag{3}$$

To get the MMEs from (2) we therefore add to its I_4 the matrix $(10/4)A^{-1}$ which, with $\sigma_e^2 = 10$ and $\sigma_a^2 = 4$ involves $(10/4)/6 = 6/12$. Hence we get

$$I_4 + \frac{5}{12} \begin{bmatrix} 11 & 3 & -6 & -4 \\ & 9 & -6 & 0 \\ & & 12 & 0 \\ & & & 8 \end{bmatrix} = \begin{bmatrix} 1 + \frac{55}{12} & \frac{15}{12} & \frac{-30}{12} & \frac{-20}{12} \\ & 1 + \frac{45}{12} & \frac{-30}{12} & 0 \\ & & 1 + \frac{60}{12} & 0 \\ & & & 1 + \frac{40}{12} \end{bmatrix}$$

$$= \begin{bmatrix} 5.5833 & 1.25 & -2.5 & -1.666 \\ & 4.75 & -2.5 & 0 \\ & & 6 & 0 \\ & & & 4.33 \end{bmatrix}. \tag{4}$$

Replacing the I_4 in (2) with (4) and then (for some reason) dividing the whole equation by 10 gives [360, (28.8)].

Question In the paragraph atop [359] there "are b animals with tested progeny": only $c \leq b$ of these b parents are tested, and there are d tested animals with no progeny.

28.6.2 Repeated records model [361, (28.6.2)]

Items (1) – (3) on [362] are the same as (2) – (4) on [359] with σ_p^2 in place of σ_e^2.

The lower four diagonal elements in the matrix of [363, (28.11)] are of the form

$$\frac{1}{.55}\left(n+\frac{.55}{.20}\right)=\frac{1}{.55}\left(n+\frac{\sigma_e^2}{\sigma_p^2}\right),$$

in which n comes from $Z_p'Z_p$ and σ_e^2/σ_p^2 comes from $R[\text{var}(p)]^{-1}=\left(\sigma_e^2|\sigma_p^2\right)I$. And diagonal elements 2 – 5 have the form

$$\frac{1}{.55}\left(n_i+\frac{\sigma_e^2}{\sigma_\alpha^2}a_{ii}\right)\quad\text{for}\quad\frac{\sigma_e^2}{\sigma_\alpha^2}=\frac{.55}{.25}$$

where a^{ii} is the i'th diagonal element of A^{-1}. For example, with $i=1$, $n_1=2$ and, from (3), $a^{11}=11/6$,

$$10.970=\frac{1}{.55}\left(2+\frac{.55}{.25}\frac{11}{6}\right).$$

Thus the lower 8×5 submatrix of the matrix in (28.11) is

$$\frac{1}{.55}\begin{bmatrix}\{_d\, n_i\}+(\sigma_e^2/\sigma_a^2) & \{_d\, n_i\}\\ \{_d\, n_i\} & \left\{_d\,(n_i+\sigma_e^2/\sigma_p^2)\right\}\end{bmatrix}.$$

Comment: The elements of neither BLUP(p) nor BLUP$(\mathbf{a})$ do not add to zero for the reasons given in these notes for Section 23.1.

Chapter 29

Non-Additive Genetic Merit

29.1 Model for genetic components [365, (29.1)]

Very straightforward. The special matrix product symbol # at the bottom of [365] is defined (as on [366]) as $A \# D = \{a_{ij}d_{ij}\}$. It is the Hadamard product, more usually written as $A \cdot D$ or $A \odot D$.

29.2 Single record on every animal [366, (29.2)]

This, too, is straightforward reading. One can observe that an easier-to-read form of (29.6) is

$$\begin{bmatrix} X' \\ I \\ I \end{bmatrix} \left(A\sigma_a^2 + D\sigma_d^2 + I\sigma_e^2\right)[X \quad I \quad I]. \tag{29.6}$$

It is tempting to think that one could achieve some simplification of the algebra on [368] but I've had no luck. In any case, for estimating variance components my preference would be to use ML or REML directly.

29.3 Single or no record on each animal [369, (29.3)]

Again, the model description is straightforward and the arithmetic of the example is fairly hefty.

Typo In the first line, Section 28.2 should be 29.2.

Chapter 30

Line Cross and Breed Cross Analyses

30.1 Genetic Model [381, (30.1)]

This is the same as [365] except that although additive × additive is mentioned in [381, (30.1), lines 2-3] of this section it is overlooked in the subsequent listing:

$$\text{Var(additive} \times \text{additive)} = A\#A\sigma^2_{aa} \equiv A \odot A\sigma^2_{aa}.$$

30.2 Covariances between crosses [382, (30.2)]

My genetics fail me!

30.3 Reciprocal crosses assumed equal [384, (30.3)]

Because line crosses $i \times j$ and $j \times i$ are considered equal there are only six classes with numbers $n_{ij} + n_{ji}$:

$i,j \equiv j,i$					
1,2	1,3	1,4	2,3	2,4	3,5
5	3	2	6	3	5
4	4	2	2	3	9
9	7	4	8	6	14

Thus it is that the matrix of the OLS equations (30.1) contains the sequence of numbers in the last line of the above table.

(3) I do not like to "pretend" about the model; see [389, 6th line up] and again on [390, top line].

(4) What is the significance of the parenthesized subscripts in $r_{(i,j)}$? And what is r, anyway?

Chapter 31

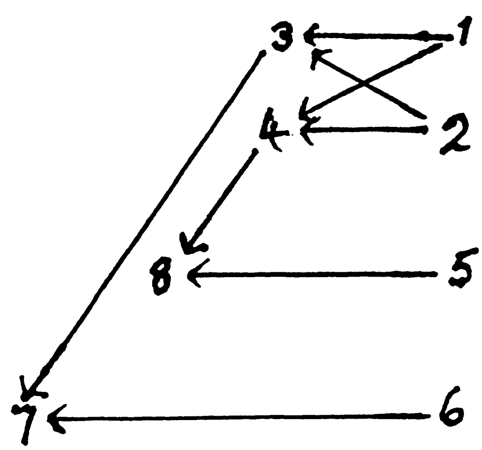
Maternal Effects

31.1 Model for maternal effects [395, (31.1)]

No comment needed.

31.2 Pedigrees used in example [396, (31.2)]

Diagramming pedigrees helps derive A; each line segment represents a multiplicative .5.



31.3 Additive and Dominance Maternal and Direct Effects [398, (31.3)]

The last line, quite rightly, proclaims the example data as being inadequate for estimating variances. But they are nevertheless too voluminous for demonstrating the arithmetic.

Chapter 32

Three-Way Mixed Model

32.1 The Example [399, (32.1)]

In [400, first line] why suggest using "some prior on squares and products of bc_{jk}" and then in the next line do what seems much more practical, utilize a "pseudo σ_{bc}^2"?

32.2 Estimation and prediction [400, 32.2]

Using that pseudo σ_{bc}^2 so that $\sigma_e^2/\sigma_{bc}^2 = 6$, in [401, lines 2-4], the diagonal matrix added to the coefficient matrix is

$$\{_d\, 2I_3 \quad 0I_3 \quad 0I_3 \quad 3I_9 \quad 4I_9 \quad 6I_9 \quad ; 5I_{27}\}$$

corresponding to effects

$$a \quad b \quad c \quad ab \quad ac \quad bc \quad abc.$$

Thus b and c are being treated as fixed but bc is treated as (pseudo) random.

For the solution vectors (to the MMEs) note the cases of BLUPs adding to zero; e.g.,

$$\Sigma\tilde{a}_i = -.54801 + .10555 + .44246 = 0.$$

Likewise for interaction effects, their BLUPs summed over all levels of a random effect add to zero for each level of a fixed effect. For example

$$\widetilde{ab}_{11} + \widetilde{ab}_{21} + \widetilde{ab}_{21} = -1.21520 + .14669 + 1.06850 = -.00001.$$

And on [410]

$$G = \begin{bmatrix} 4 & 8 \\ 8 & 16 \end{bmatrix} \otimes A \qquad \text{and} \qquad R = \begin{bmatrix} 12 & 0 \\ 0 & 48 \end{bmatrix} \otimes I.$$

A and an incidence matrix for the OLS equations are given, but for the MMEs only solutions are given. And we note that BLUPs adding to zero does not occur, for the reasons given in these notes at Section 23.1.

References

(excluding those in CRH)

Anderson, R.D. (1978) *Studies on the estimation of variance components.* Ph.D. Thesis, Cornell University, Ithaca, N. Y.

Hemmerle, W.J. and Hartley, H.O. (1973) Computing maximum likelihood estimates for the mixed A.O.V. model using the W-transformation. *Technometrics*, **15**, 819-831.

Jennrich, R.J. and Sampson, P.F. (1976) Newton-Raphson and related algorithms for maximum likelihood estimation of variance components. *Technometrics*, **18**, 11-17.

McCulloch, C.E. and Searle, S.R. (1995) An identity derived from unbiasedness in linear models. *The American Statistician*, **49**, 39-42.

Pearson, E.C. and Hartley, H.O., eds. (1954) *Biometrika Tables for Statisticians*, Vol. I, 3rd Ed., Cambridge University Press.

Robson, D.S. (1959) A simple method for constructing orthogonal polynomials when the independent variable is unequally spaced. *Biometrics*, **15**, 187-191.

Searle, S.R. (1995) Some follow-up on Aitken's least squares equations. In *Proceedings A.C. Aitken Conference*, University of Otago, Dunedin, New Zealand, 299-309.

Searle, S.R. (1997) Built-in restrictions on best linear unbiased predictors (BLUP) of random effects in mixed models. *The American Statistician*, **51**, 19-21.

Searle, S.R. and Fawcett, R.F. (1980) Expected mean squares in variance components models having finite populations. *Biometrics*, **36**, 243-254.

Searle, S.R., Speed, F.M. and Henderson, H.V. (1982) Some computational and model equivalences in analysis of variance of unequal-subclass-numbers data. *The American Statistician*, **35**, 16-33.

Seber, G.A.F. (1977) *Linear Regression Analysis*, John Wiley & Sons, New York.