MAXIMUM LIKELIHOOD ESTIMATION OF VARIANCE COMPONENTS

A thesis submitted for the degree of Master of Statistics of the Australian National University November 1991

Alice Marion Richardson
I, Alice Richardson, declare that all work in this thesis is my own.
All references to the work of others have been duly acknowledged.

Alice Richardson
ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr. Alan Welsh, for his helpful comments throughout both the research and writing stages of this thesis. Thanks also to Ms. Michelle Ross who provided help with Macintosh word processing, and to Professor Charles Heathcote for his general support. Finally, thanks are due to my hall of residence, Ursula College, where the in-college computing facilities are second to none.
ABSTRACT

In this thesis, the Maximum Likelihood and Restricted Maximum Likelihood methods of estimating variance components are investigated for the one-way model. Expressions for the estimators and their variances are obtained, and algorithms for finding the estimates are tested by means of a Monte Carlo study. The quantitative effects of non-normality on the variability of estimates are discussed. Finally, diagnostic tests for identifying outliers and non-normality are proposed, and illustrated with data concerning soybean plant growth.
1. INTRODUCTION

The complex covariance structure of the variance components model can be used to describe data arising in fields as diverse as designed experiments in agriculture and observational studies in the social sciences. This combination of complexity and wide applicability has meant that the estimation of variance components has been a rich source of research problems over the last century. Even so, there still remain unsolved problems pertaining to various facets of the variance components model, in particular, which estimation method is to be preferred in a particular situation. Many kinds of estimator have been proposed, with very few guidelines for choosing between them.

The objective of this thesis is to investigate the classical methods of estimating variance components, concentrating on Maximum Likelihood (ML) and Restricted Maximum Likelihood (REML) for the one–way mixed model, in both the balanced and unbalanced case.

1.1 DESCRIPTION OF THE MODEL

Consider the general linear model:

\[ y = X\alpha + \sum_{i=1}^{c} Z_i \beta_i + \epsilon \]  

where \( y \) is a \( n \)-vector of observations, \( X \) and \( Z_i \) are known \( n \times q \) and \( n \times p_i \) design matrices respectively; \( \alpha \) is a \( q \)-vector of fixed effects; the \( \beta_i \) are \( p_i \)-vectors of random effects, \( 1 \leq i \leq c \); and \( \epsilon \) is an \( n \)-vector of errors. The \( p_i \) levels of each random effect are assumed to be a random sample from an infinite population of levels. By putting \( \text{Var}(\beta_i) = D_i \) and \( \text{Var}(\epsilon) = R \), we obtain

\[ \text{Var}(y) = R + \sum_{i=1}^{c} Z_i D_i Z_i^T = V. \]

The problem is to estimate \( R \) and \( D_1, \ldots, D_c \).

The standard variance components model is a special case of (1.1). Often \( \alpha \) reduces to a single parameter, the overall mean, which will be the situation throughout this thesis. The levels of each factor \( \beta_i \) are assumed to be independently normally distributed with mean zero and variance \( \sigma_i^2 \). Similarly, each random error \( \epsilon_i \) is assumed to be independently normally distributed with mean zero, and variance \( \sigma_0^2 = \sigma^2 \). Finally, \( \beta_i \)
and $\varepsilon$ are assumed to be uncorrelated. Thus $D_i = \sigma_i^2 I_k$ where $I_k$ is a $k \times k$ identity matrix, $R = \sigma_\varepsilon^2 I_n$, and

$$V = \sigma_\varepsilon^2 I_n + \sum_{i=1}^{c} \sigma_i^2 Z_i Z_i^T = \sigma_\varepsilon^2 I_n + \sum_{i=1}^{c} \sigma_i^2 G_i$$

where $G_i = Z_i Z_i^T$. The problem reduces to estimating $\sigma_\varepsilon^2$ and $\sigma_1^2, \ldots, \sigma_c^2$.

Some authors use a slightly different parameterisation in terms of $\gamma$:

$$\gamma_i = \sigma_i^2 / \sigma_\varepsilon^2, \quad i = 1, \ldots, c \text{ and } \gamma_0 = \sigma_\varepsilon^2.$$  

The $\sigma^2$ formulation is used in this thesis because it is usually easier to work with the actual parameters requiring estimation. It is however a simple matter to move between parameterisations using the Jacobian of the transformation.

### 1.2 Extant Research

According to Rao and Kleffe (1988), the first use of variance components was in the work of astronomers Airy (1861) and Chauvenet (1863), who used a one-way random model. Fisher (1918) was the first to use the term “analysis of variance” (ANOVA) and Eisenhart (1947) introduced the term “Model II” for the random (or mixed) effects model that lends itself to variance component analysis.

But it is Henderson’s (1953) paper that could best be described as the “landmark” piece of research in this field, for Henderson cites only four earlier papers, but most subsequent researchers cite him. (Those four earlier papers are Crump (1946, 1951), Eisenhart (1947) and Hazel and Terril (1945).) Essentially all the papers before Henderson dealt with specific cases: Henderson was the first to present methods in general. Henderson proposed three methods of estimating variance components: they are all variations on ANOVA estimates which involve equating various quadratic forms of the observations to their expected values, then solving for the variance components. His methods are still popular as a first attempt at estimating variance components.

Other methods involving equating quadratic forms to their expectations have since been proposed e.g. symmetric sums estimators. Searle’s (1981) book contain lists of formulae for this particular method for models with up to three crossed factors.

However, such methods have their drawbacks, in particular the large number of times that negative estimates of variance are obtained. Also, for all but the simplest models, there is a choice of quadratic forms to equate to their respective expected values. Thus
there exists a wide variety of ANOVA methods, but no clear criteria for choosing between them.

Hartley and Rao (1967) took a step away from this confusing situation when they developed a procedure for the maximum-likelihood estimation of components of variance based on the steepest-ascent method of solving equations. They also discussed asymptotic efficiency of the estimates by looking at the second derivative of the log likelihood. Searle (1970) also derived formulae for variance matrices in the same way.

However ML estimates (MLEs) are not perfect either – their main drawback is bias. The simplest example of this is also a very well-known one – that of estimating the single “variance component” $\sigma^2$ of a $N(\mu, \sigma^2)$ distribution, when $\mu$ is unknown. If a random sample $x_1, \ldots, x_n$ is taken from the distribution, the MLE of $\sigma^2$ is $\frac{1}{n} \sum (x_i - \bar{x})^2$

which has expected value $\frac{(n-1)\sigma^2}{n}$. The unbiased estimator, which effectively takes into account the fact that the mean is unknown and must be estimated too, is therefore $\frac{1}{n-1} \sum (x_i - \bar{x})^2$.

Patterson and Thompson (1971) incorporated this notion of allowing for the estimation of fixed effects by maximising the likelihood, not of all the data, but of a selected set of "error contrasts". (An error contrast is defined as a linear function $S$ of $y$ whose expectation is 0.) Letting $\text{Var}(y) = V$, the set of error contrasts they chose has a singular covariance matrix, $S^T V S$, even though the inverse of this matrix is required in the derivative of the likelihood. Patterson and Thompson overcame the difficulty by substituting the spectral form of $S^T V S$ into the likelihood, however the variances to be estimated are then embedded in the eigenvalues, so differentiating with respect to them is still awkward.

Corbeil and Searle (1976a) avoided the singularity issue by taking the error contrasts to be a set of $(n - \text{rank}(X))$ rows of $S$, and they called their new estimators REML estimators. In their paper they cover estimation by a version of the Newton–Raphson method and the variance matrix of the estimates. The same authors (1976b) also compare the variance of ML and REML estimators for balanced designs up to two-way, and give a numerical example of an unbalanced two-way crossed model without interaction.

Finally, Harville (1977) showed that there is no need for error contrasts, but simply a need to insert into the likelihood a term allowing for the fixed effects being estimated as well. He converted the REML equations from an $n \times n$ system of equations into a $p \times p$ system, where $p < n$, but the system still requires iterative techniques to solve.
More recently, authors such as Fellner (1986) and Hocking, Green and Bremer (1989) have been addressing the problem of estimation for non–normal data, and diagnostic tests to detect non–normality. Non–normality is a problem with real–life data, and it can occur in many ways because of the complex nature of experiments analysed by variance components, often involving several crossed or nested factors. There appear to be two basic approaches to non–normality as represented by these two papers:

1) derive new estimators that also identify outlying observations (Hocking et al.)
2) robustify current estimators to reduce the effect of outlying observations on the estimate (Fellner).

Parallelling this series of estimators was the development of MINQEs and MIVQEs. MIVQES are Minimum Variance translation–invariant Quadratic Unbiased Estimators. In this context, the quadratic form $y^T Ay$ is translation–invariant if $(y - Xa)^T A (y - Xa) = y^T Ay$ i.e. $AX = 0$. For the standard variance components model following (1.1), where $\beta_i \sim N(0, \sigma^2_i I_n)$ and $\varepsilon \sim N(0, \sigma^2_0 I_n)$, the MIVQUE of $\sum_{i=0}^{c} \lambda_i \sigma_i^2$ is $y^T Ay$, where $A$ is chosen to minimise $\text{Var}(y^T Ay)$, subject to $E(y^T Ay) = \sum_{i=0}^{c} \lambda_i \sigma_i^2$. The main drawback of the procedure is that it sometimes produces negative estimates of variance.

Minimum Norm Quadratic Estimators are a whole family of estimators: MINQEs (Unbiased), MINQIEs (Invariant), MINQUIEs (Unbiased and Invariant), MINQUNNDEs (Unbiased and Non–Negative Definite) and MINQENNDEs (Invariant and Non–Negative Definite). Briefly, the derivation of MINQEs is as follows (see Rao and Kleffe (1988)):

A natural estimator of $\sum_{i=0}^{c} \lambda_i \sigma_i^2$ is $\hat{\beta}^T L \hat{\beta}$ (for some suitable $L$.) We estimate $\hat{\beta}^T L \hat{\beta}$ by

$$y^T Ay = (\alpha \beta) \begin{pmatrix} X^T A X & X^T A Z \\ Z^T A X & Z^T A Z \end{pmatrix} (\alpha \beta)^T$$

The aim is to minimise the difference between $L$ and $A$ i.e. minimise

$$\left\| \begin{pmatrix} X^T A X & X^T A Z \\ Z^T A X & Z^T A Z - L \end{pmatrix} \right\|.$$
Harville (1977) states that MIVQUE assuming normality is equivalent to MINQUE assuming $|| ||$ is the Euclidean norm. Under these conditions also MIVQUE (that is, MINQUE) is equivalent to the first iteration of REML. Rao and Kleffe (1988) also state that MINQE is equivalent to the first iteration of ML. Thus MINQE and MIVQUE are contained within ML and REML estimation respectively.

These last two techniques, ML and REML, are the focus of this thesis.
2. ML AND REML ESTIMATION OF VARIANCE COMPONENTS
FOR THE ONE-WAY MODEL

2.1 AN EXAMPLE OF A ONE-WAY EXPERIMENT

Several of the published articles about variance components include data sets to illustrate the points being made. The data set used to illustrate this thesis, given in Table 2.1, is quoted by Hocking (1985) and comes from Snedecor and Cochran (1967). They explain that the data are from "an experiment in which four seed treatments were compared with no treatment (Check) on soybean seeds. The data are the number of plants which failed to emerge out of 100 in each plot."

Table 2.1. Number of soybean plants (out of 100) failing to germinate

<table>
<thead>
<tr>
<th>treatment</th>
<th>replications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Check</td>
<td>8 10 12 13 11</td>
</tr>
<tr>
<td>Arasan</td>
<td>2 6 7 11 5</td>
</tr>
<tr>
<td>Spergon</td>
<td>4 10 9 8 10</td>
</tr>
<tr>
<td>Semesan Jr</td>
<td>3 5 9 10 6</td>
</tr>
<tr>
<td>Fermate</td>
<td>9 7 5 5 3</td>
</tr>
</tbody>
</table>

2.2 THE GENERAL ONE-WAY MODEL

The soybean experiment in Table 2.1 can be expressed in general terms as follows:

\[
\begin{align*}
\text{Total} & \\
\text{Treatment 1} & \quad y_{11} \ldots y_{1n1} & T_1 \\
\vdots & \vdots & \vdots \\
\text{Treatment p} & \quad y_{p1} \ldots y_{pn_p} & T_p \\
\text{Total} & \quad y_{1n1} \ldots y_{pn_p} & G
\end{align*}
\]

or in matrix notation as

\[
y = X\alpha + Z\beta + \varepsilon
\]

(2.1)

\[
\begin{pmatrix}
y_{11} \\
\vdots \\
y_{p1} \\
\vdots \\
y_{pn_p}
\end{pmatrix} = \begin{pmatrix}
1 \\
\vdots \\
1 \\
\vdots \\
1
\end{pmatrix} \alpha + \begin{pmatrix}
1_{n1} \\
\vdots \\
1_{n2} \\
\vdots \\
1_{np}
\end{pmatrix} \beta + \varepsilon,
\]

6
where $1_{ni}$ is a $n_i \times 1$ vector of ones. If the experiment is balanced, there are the same number of observations in each treatment i.e. $n_i = m$ for $1 \leq i \leq p$. Otherwise, the experiment is unbalanced. In terms of the model (1.1), $c = 1$ and $p$ denotes the number of levels of $\beta$.

To simplify notation, let $\sigma^2_e = e$ and $\sigma^2_{\beta} = b$. Therefore

$$\text{Var} (y) = V = R + ZDZ^T$$

$$= eIn + b \begin{pmatrix} J_{ni} \\ J_{np} \end{pmatrix}$$

where $J_m$ is a $m \times m$ matrix of ones

$$= eIn + bG,$$ where $G = ZZ^T$.

The matrix $V$ is non-zero only in blocks down the diagonal, and we may thus refer to the $i$th diagonal block of $V$ as $eIn_i + bJ_{ni}$, $i = 1, \ldots, p$. Then using the fact that

$$(eIn_i + bJ_{ni})^{-1} = \frac{1}{e} I_{ni} - \left( \frac{b}{e(e+n_i b)} \right) J_{ni} \text{ (Graybill (1983) p. 190),}$$

$V^{-1}$ is block diagonal with blocks of

$$\begin{pmatrix} e+(n_i-1)b & -b \\ e(e+n_ib) & e(e+n_i b) \end{pmatrix} \begin{pmatrix} -b & e+(n_i-1)b \\ e(e+n_i b) & e(e+n_i b) \end{pmatrix}$$

The block diagonal structure of $V$ and $V^{-1}$ greatly simplifies the calculation of the likelihood in the one-way case, because it carries through to most of the matrices involved. The expression "ith block" next to a matrix means that the matrix is block diagonal, and that only the ith block has been written down.

2.3 ESTIMATION OF EFFECTS

Even though the effects $\beta$ and $\epsilon$ are not parameters, it is useful to estimate them because they can in turn be used to estimate the parameters of interest, the variance components. An estimate of the parameter $\alpha$ is required for the same reason.

The MLE of $\alpha$ is easily shown to be $\hat{\alpha} = (X'V^{-1}X)^{-1}X'V^{-1}y$. For the one-way model
\[
\hat{\alpha} = (1 \ldots 1) \begin{pmatrix}
\frac{e+(n_i-1)b}{e(e+n_i b)} & \frac{-b}{e(e+n_i b)} \\
\frac{-b}{e(e+n_i b)} & \frac{e+(n_i-1)b}{e(e+n_i b)}
\end{pmatrix}
\begin{pmatrix}
y_{i1} \\
\vdots \\
y_{i1} \\
y_{ini}
\end{pmatrix}
+ \frac{1}{\sum_{i,j}[V^{-1}]_{ij}} \text{ith block}
\]

\[
= \sum_{i=1}^{p} \frac{T_i}{e+n_i b} \sum_{i=1}^{p} \frac{n_i}{e+n_i b}
\]

which reduces to the overall mean \( \bar{y} \) if the experiment is balanced.

No MLE of \( \beta \) exists because \( \beta \) does not feature in the likelihood. The best (minimum MSE) predictor of \( \beta \), derived in the Appendix at (A1.1), is \( \hat{\beta} = DZ'V^{-1}(y - X\hat{\alpha}) \), and it follows immediately that the best predictor of \( e \) is \( \hat{e} = y - X\hat{\alpha} - Z\hat{\beta} \).

The expression for \( \hat{\beta} \) in the one-way model is

\[
\hat{\beta} = \begin{pmatrix}
\frac{b}{e+n_1 b} & \frac{b}{e+n_1 b} \\
\frac{b}{e+n_p b} & \frac{b}{e+n_p b}
\end{pmatrix}
\begin{pmatrix}
y_{11} - \hat{\alpha} \\
\vdots \\
y_{pn_p} - \hat{\alpha}
\end{pmatrix}
\]

\[
= \begin{pmatrix}
\frac{n_1 b}{e+n_1 b} & \bar{y}_1 - \hat{\alpha} \\
\vdots \\
\frac{n_p b}{e+n_p b} & \bar{y}_p - \hat{\alpha}
\end{pmatrix}
\]

\[
= DZ'Py
\]

where \( P = V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1} \)

\[
= V^{-1} - \left( \sum_{i=1}^{p} \frac{n_i}{e+n_i b} \right)^{-1} \begin{pmatrix}
\frac{1}{(e+n_1 b)^2} J_{n_1} & \frac{1}{(e+n_1 b)(e+n_j b)} J_{n_1 \times n_j} \\
\frac{1}{(e+n_1 b)(e+n_j b)} J_{n_1 \times n_j} & \frac{1}{(e+n_p b)^2} J_{n_p}
\end{pmatrix}
\]

where \( J_{n_i \times n_j} \) is a \( n_i \times n_j \) matrix of ones.

Therefore \( P \) has diagonal blocks of
The matrix \( P \) will be seen in the next section to be a very important quantity in REML estimation, because it is the inverted variance matrix with a correction allowing for the fact that \( \alpha \) has been estimated. The expression \( e + n_ib\) that is a feature of the elements of \( P \) in the specific context of the one-way model will also be seen again and again.

### 2.4 ESTIMATION OF VARIANCE COMPONENTS

#### 2.4.1 MAXIMUM LIKELIHOOD

To estimate \( e \) and \( b \), we start with the log likelihood, \( L \). Ignoring terms free of \( e \) and \( b \)

\[
L = \left( -\frac{1}{2} \right) \log |V| - \left( \frac{1}{2} \right) (y - X\alpha)^T V^{-1} (y - X\alpha).
\]

(2.4)

Using (A1.5) to differentiate (2.4) with respect to \( e \)

\[
\frac{\partial L}{\partial e} = \left( -\frac{1}{2} \right) \text{tr} \left[ V^{-1} (\partial V / \partial e) \right] + \left( \frac{1}{2} \right) (y - X\alpha)^T V^{-1} (\partial V / \partial e) V^{-1} (y - X\alpha)
\]

\[
= \left( -\frac{1}{2} \right) \text{tr} \left[ V^{-1} \right] + \left( \frac{1}{2} \right) (y - X\tilde{\alpha})^T V^{-1} V^{-1} (y - X\tilde{\alpha})
\]

replacing \( \alpha \) by \( \tilde{\alpha} \), and noting that \( \partial V / \partial e = I_n \).

\[
\frac{\partial L}{\partial e} = \left( \begin{array}{cc}
\frac{e+(n_i-1)b}{e(e+n_ib)} & -b \\
-b & \frac{e+(n_i-1)b}{e(e+n_ib)}
\end{array} \right)
\]

ith block

\[
(y_{i1} - \tilde{\alpha}, \ldots, y_{im} - \tilde{\alpha})
\]

\[
\left( \begin{array}{cc}
\frac{(e+(n_i-1)b)^2+(n_i-1)b^2}{e^2(e+n_ib)^2} & -\frac{2be-n_i b^2}{e^2(e+n_ib)^2} \\
-\frac{2be-n_i b^2}{e^2(e+n_ib)^2} & \frac{(e+(n_i-1)b)^2+(n_i-1)b^2}{e^2(e+n_ib)^2}
\end{array} \right)
\]

ith block
\[ \frac{\partial L}{\partial e} = \frac{n-p}{-2e} - \frac{1}{2} \sum_{i=1}^{p} \frac{1}{e+n_i b} + \frac{1}{2e^2} \sum_{i=1}^{p} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 + \frac{1}{2} \sum_{i=1}^{p} \frac{n_i (\bar{y}_i - \bar{\alpha})^2}{(e+n_i b)^2}. \] (2.5)

Similarly, differentiating (2.4) with respect to \( b \),

\[ \frac{\partial L}{\partial b} = (-1/2) \text{tr}[V^{-1}(\partial V/\partial b)] + (1/2) (y - X\alpha)^T V^{-1}(\partial V/\partial b)V^{-1}(y - X\alpha) \]

and after replacing \( \alpha \) by \( \bar{\alpha} \) and recalling that \( \partial V/\partial b = ZZ^T = G \),

\[ \frac{\partial L}{\partial b} = (-1/2) \text{tr}[V^{-1}G] + (1/2) (y - X\bar{\alpha})^T V^{-1}G V^{-1}(y - X\bar{\alpha}) \]

\[ = (-1/2) \text{tr} \begin{pmatrix} \frac{1}{e+n_i b} & \cdots & \frac{1}{e+n_i b} \\ \vdots & \ddots & \vdots \\ \frac{1}{e+n_i b} & \cdots & \frac{1}{e+n_i b} \end{pmatrix} \text{ith block} \]

\[ + \left(y_{i1} - \bar{\alpha}, \ldots, y_{im} - \bar{\alpha}\right) \begin{pmatrix} \frac{n_i}{e+n_i b} & \cdots & \frac{n_i}{e+n_i b} \\ \vdots & \ddots & \vdots \\ \frac{n_i}{e+n_i b} & \cdots & \frac{n_i}{e+n_i b} \end{pmatrix} \begin{pmatrix} y_{i1} - \bar{\alpha} \\ \vdots \\ y_{im} - \bar{\alpha} \end{pmatrix} \text{ith block} \]

\[ = -\frac{1}{2} \sum_{i=1}^{p} \frac{n_i}{e+n_i b} + \frac{1}{2} \sum_{i=1}^{p} \frac{n_i^2 (\bar{y}_i - \bar{\alpha})^2}{(e+n_i b)^2}. \] (2.6)

2.4.2 RESTRICTED MAXIMUM LIKELIHOOD

Now the likelihood equation for finding the REML estimates differs from (2.4) by only one term, as below:

\[ L_i = (-1/2) \log|V| - (1/2) \log|X^TV^{-1}X| - (1/2) (y - X\alpha)^T V^{-1}(y - X\alpha) \] (2.7)

so that the derivatives are (using (A 1.6)):

\[ \frac{\partial L_i}{\partial e} = (-1/2) \text{tr}[P] + (1/2) (y - X\bar{\alpha})^T V^{-1}V^{-1}(y - X\bar{\alpha}) \]

\[ = -\frac{1}{2} \sum_{i=1}^{p} \frac{n_i e + (n_i - 1)b(\Sigma n_i/e+n_i b) - n_ie}{e(e+n_i b)^2(\Sigma n_i/e+n_i b)} \]

\[ + \frac{1}{2e^2} \sum_{i=1}^{p} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 + \frac{1}{2} \sum_{i=1}^{p} \frac{n_i (\bar{y}_i - \bar{\alpha})^2}{(e+n_i b)^2}. \] (2.8)
\[ \frac{\partial L_1}{\partial b} = (-1/2) \text{tr}[\Pi G] + (1/2) (y - X\hat{\alpha})^T V^a G V^a (y - X\hat{\alpha}) \]

\[ = (-1/2) \text{tr} \begin{pmatrix} 
\frac{(e+nib)(\Sigma n_i/e+nib) - n_i}{(e+nib)^2(\Sigma n_i/e+nib)} & \frac{-n_i}{(e+nib)(e+njb)(\Sigma n_i/e+nib)} \\
\frac{-n_i}{(e+nib)(e+njb)(\Sigma n_i/e+nib)} & \frac{J_{njxni}}{J_{njxni}} \\
\end{pmatrix} \]

\[ + (1/2) (y - X\hat{\alpha})^T V^a G V^a (y - X\hat{\alpha}) \]

\[ = -\frac{1}{2} \sum_{i=1}^{p} \frac{(e+nib)(\Sigma n_i/e+nib) - n_i}{(e+nib)^2(\Sigma n_i/e+nib)} + \frac{1}{2} \sum_{i=1}^{p} \frac{n_i^2(\bar{y}_i - \hat{\alpha})^2}{(e+nib)^2}. \] (2.9)

### 2.5 A Special Case of Explicit Solutions

If the experiment is balanced and the ANOVA estimates are positive, then (2.8) and (2.9) can be used to show that REMLEs are equivalent to ANOVA estimates, as follows. The ANOVA table for a one-way balanced experiment is

<table>
<thead>
<tr>
<th>source of error</th>
<th>dof</th>
<th>SS</th>
<th>MS</th>
<th>E(MS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>treatments</td>
<td>p-1</td>
<td>mΣ(\bar{y}_i - \bar{y})^2</td>
<td>SS/p-1 = MSB</td>
<td>e + mb</td>
</tr>
<tr>
<td>error</td>
<td>n-p</td>
<td>ΣΣ(y_{ij} - \bar{y}_i)^2</td>
<td>SS/n-p = MSE</td>
<td>e</td>
</tr>
<tr>
<td>total</td>
<td>n-1</td>
<td>ΣΣ(y_{ij} - \bar{y})^2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

and equating mean squares to expected mean squares we obtain the ANOVA estimates of the variance components: \( \hat{\epsilon} = \text{MSE} \) and \( \hat{b} = \frac{\text{MSB} - \text{MSE}}{m} \).

Writing (2.9) for balanced data:

\[ \frac{\partial L_1}{\partial b} = 0 \Rightarrow \frac{-n(m - b)}{2(e + mb)} + \frac{m \text{SSB}}{2(e + mb)^2} = 0 \Rightarrow \hat{\epsilon} + \hat{mb} = \frac{\text{SSB}}{p - 1} = \text{MSB} \] (2.10)

and (2.8) can be written

\[ \frac{\partial L}{\partial \epsilon} = 0 \Rightarrow \frac{-n(\hat{\epsilon} + (m-1)b) + \hat{\epsilon}}{-2\epsilon(e + mb)} + \frac{\text{SSE}}{2\epsilon^2} + \frac{\text{SSB}}{2(e + mb)^2} = 0 \]
\[ \Rightarrow \frac{-mn(\hat{e} + (m-1)\hat{b}) + m\hat{e}}{-me(\hat{e} + mb)} + \frac{SSE}{\hat{e}^2} + \frac{\hat{e}(n-m)}{m\hat{e}(\hat{e} + mb)} = 0 \] using (2.10)

\[ \Rightarrow \hat{e} = \frac{SSE}{n-p} = MSE, \text{ the ANOVA estimate of } e. \]

Substituting back in (2.10), \( \hat{b} = \frac{MSB - MSE}{m} \), the ANOVA estimate of \( b \).

Similarly, explicit MLEs can be calculated if the data is balanced. Writing (2.6) for balanced data yields

\[ \frac{\partial L}{\partial b} = 0 \Rightarrow \frac{-n}{2(\hat{e} + mb)} + \frac{mSSB}{2(\hat{e} + mb)^2} = 0 \Rightarrow \hat{e} = \frac{SSB}{p} \]

while (2.5) for balanced data reads

\[ \frac{\partial L}{\partial e} = 0 \Rightarrow \frac{n-p}{2\hat{e}} - \frac{p}{2(\hat{e} + mb)} + \frac{SSE}{2\hat{e}^2} + \frac{SSB}{2(\hat{e} + mb)^2} = 0 \]

\[ \Rightarrow \frac{n-p}{\hat{e}} - \frac{p}{\hat{e} + mb} + \frac{SSE}{\hat{e}^2} + \frac{p}{\hat{e} + mb} = 0 \]

using (2.11)

\[ \Rightarrow \hat{e} = \frac{SSE}{n-p} = MSE. \]

Substituting back in (2.11), \( \hat{b} = \frac{(SSB/p) - MSE}{m} \).

These exact results will be used to calculate measures of efficiency in the balanced case (Chapter 4).

2.6 DISCUSSION

The amount of work required to solve either the ML or REML equations is virtually the same because the equations differ only in one term. As in §2.5, the obvious way to proceed is to set (2.5), (2.6), (2.8) and (2.9) to zero and solve for the unknown variance components. However explicit solutions are not generally possible because the nature of the equations is such that iterative techniques need to be employed. Two useful algorithms to do this are discussed in the next chapter.
3. NUMERICAL PROCEDURES FOR MAXIMUM LIKELIHOOD ESTIMATION

This chapter examines two procedures for obtaining MLEs and REMLEs by iteration. They are not the only methods by any means, but their advantages are  
1) they are both conceptually clear  
2) they both lend themselves to be written in the form of S-plus functions  
3) they have not been thoroughly investigated in the published literature.

3.1 ANDERSON'S ALGORITHM

Differentiating the likelihood (2.4) with respect to $\theta_i$ (where $\theta_i$, $i = 1, 2$ are the elements of the vector of parameters $\theta = (\alpha, \beta)^T$) and setting the result equal to zero we obtain  
$$ \frac{\partial L}{\partial \theta_i} = 0 \Rightarrow \text{tr}[V^i(\partial V/\partial \theta_i)] = (y - X\alpha)^T V^i(\partial V/\partial \theta_i) V^i(y - X\alpha). $$

The left hand side can be rewritten as  
$$ \text{tr}[V^i(\partial V/\partial \theta_i)] = \text{tr} \left[ \sum \text{tr} \left( \frac{\partial^2 V}{\partial \theta_i \partial \theta_j} \right) \theta_j \right] = \sum \text{tr}[V^i(\partial V/\partial \theta_i) V^i(\partial V/\partial \theta_j)] \theta_j $$

so that  
$$ \text{tr}[V^i(\partial V/\partial \theta_i) V^i(\partial V/\partial \theta_j)] \theta = (y - X\alpha)^T V^i(\partial V/\partial \theta_i) V^i(y - X\alpha) $$

$$ B\theta = d $$

$$ \hat{\theta} = B^{-1}d. \quad (3.1) $$

If we start with (2.7), the restricted maximum likelihood $L_1$, the algebra is very similar.  
$$ \frac{\partial L_1}{\partial \theta} = 0 \Rightarrow \text{tr}[P(\partial V/\partial \theta_i)] = (y - X\tilde{\alpha})^T V^i(\partial V/\partial \theta_i) V^i(y - X\tilde{\alpha}) $$

But the left hand side equals $\text{tr}[P(\partial V/\partial \theta_i) P(\partial V/\partial \theta_j)] \theta$ so that  
$$ \text{tr}[P(\partial V/\partial \theta_i) P(\partial V/\partial \theta_j)] \theta = (y - X\tilde{\alpha})^T V^i(\partial V/\partial \theta_i) V^i(y - X\tilde{\alpha}) $$

$$ B_1 \theta = d $$

$$ \hat{\theta} = B_1^{-1}d \quad (3.2) $$

where $B$, $B_1$ and $d$ depend on $\theta$.

The matrix $B = \begin{pmatrix} \text{tr}[V^{-2}] & \text{tr}[V^iGV^i] \\ \text{tr}[V^iGV^i] & \text{tr}[V^iGV^iG] \end{pmatrix}$

and for the one--way model the individual traces are as follows:
\[
\text{tr}[V^{-2}] = \text{tr} \begin{pmatrix}
\frac{(e+(n_i-1)b)^2 + (n_i-1)b^2}{e^2(e+n_i b)^2} & \frac{-2be-n_i b^2}{e^2(e+n_i b)^2} \\
\frac{-2be-n_i b^2}{e^2(e+n_i b)^2} & \frac{(e+(n_i-1)b)^2 + (n_i-1)b^2}{e^2(e+n_i b)^2}
\end{pmatrix}
\text{ith block}
\]

\[
= \sum_{i=1}^{p} n_i \left( \frac{e^2+2(n_i-1)eb+n_i(n_i-1)b^2}{e^2(e+n_i b)^2} \right)
\]

(3.3)

\[
\text{tr}[V^{-1}G^{-1}] = \text{tr} \begin{pmatrix}
1/(e+n_1b)^2 J_{n_1} \\
1/(e+n_p b)^2 J_{n_p}
\end{pmatrix}
= \sum_{i=1}^{p} \frac{n_i}{(e+n_i b)^2}
\]

(3.4)

\[
\text{tr}[V^{-1}G^{-1}G] = \text{tr} \begin{pmatrix}
n_1/(e+n_1b)^2 J_{n_1} \\
n_p/(e+n_p b)^2 J_{n_p}
\end{pmatrix}
= \sum_{i=1}^{p} \frac{n_i^2}{(e+n_i b)^2}
\]

(3.5)

Similarly for REML, the matrix \(B_1 = \begin{pmatrix} \text{tr}[P^2] & \text{tr}[PGP] \\ \text{tr}[PGP] & \text{tr}[PGPG] \end{pmatrix} \)

and each trace is

\[
\text{tr}[P^2] = \text{tr} \begin{pmatrix}
\frac{(e+n_1b)(e+(n_i-1)b)(\Sigma n_i/e+n_1b) - e}{e(e+n_1 b)^2(\Sigma n_i/e+n_1b)} & \frac{-b(e+n_1b)(\Sigma n_i/e+n_1b) - e}{e(e+n_1 b)^2(\Sigma n_i/e+n_1b)} \\
\frac{-b(e+n_1b)(\Sigma n_i/e+n_1b) - e}{e(e+n_1 b)^2(\Sigma n_i/e+n_1b)} & \frac{(e+n_1b)(e+(n_i-1)b)(\Sigma n_i/e+n_1b) - e}{e(e+n_1 b)^2(\Sigma n_i/e+n_1b)}
\end{pmatrix}
\]

(ith diagonal block; offdiagonal blocks are \(\frac{-1}{(e+n_1b)(e+n_1b)(\Sigma n_i/e+n_1b)} J_{n_1 \times n_1} \))

14
\[
\sum_{i=1}^{p} n_i \left[ \frac{(e+n_1b)(e+(n_1-1)b)(\Sigma n_i/e+n_1b) - e}{e(n_1b)^2(\Sigma n_i/e+n_1b)} \right]^2 \\
+ (n_1-1) \left[ \frac{-b(e+n_1b)(\Sigma n_i/e+n_1b) - e}{e+n_1b)^2(\Sigma n_i/e+n_1b)} \right]^2 + \sum_{j \neq i} \frac{n_j}{((e+n_1b)(e+n_1b)(\Sigma n_i/e+n_1b))^2} \\
= \frac{n(e^2+2(m-1)e+b+m(m-1)b^2)}{e^2(e+mb)^2} \text{ if the experiment is balanced.}
\]

(3.6)

\[
\text{tr}[PGP] \\
\begin{pmatrix}
\frac{e(e+n_1b)(\Sigma n_i/e+n_1b) - n_1e}{e(e+n_1b)^2(\Sigma n_i/e+n_1b)} & \frac{-n_i}{(e+n_1b)(e+n_1b)(\Sigma n_i/e+n_1b)} \\
\frac{e(n_1b)(\Sigma n_i/e+n_1b) - n_1e}{e(n_1b)^2(\Sigma n_i/e+n_1b)} & \frac{e+n_1b)(\Sigma n_i/e+n_1b) - n_1e}{e+n_1b)^2(\Sigma n_i/e+n_1b)}
\end{pmatrix}^p
\]

\[
= \sum_{i=1}^{p} n_i \left[ \frac{(e+n_1b)(\Sigma n_i/e+n_1b) - n_1e(n_1-1)(e+n_1b)(\Sigma n_i/e+n_1b) - e}{e(e+n_1b)^2(\Sigma n_i/e+n_1b)^2} \\
+ (n_1-1) \frac{-b(e+n_1b)(\Sigma n_i/e+n_1b) - e)(e+n_1b)(\Sigma n_i/e+n_1b) - n_1e}{e(e+n_1b)^2(\Sigma n_i/e+n_1b)^2} \\
+ \sum_{j \neq i} \frac{n_j}{((e+n_1b)(e+n_1b)(\Sigma n_i/e+n_1b))^2} \\
= \frac{n-m}{(e+mb)^2} \text{ for balanced data}
\]

(3.7)

\[
\text{tr}[PGPG] \\
\begin{pmatrix}
\frac{e(e+n_1b)(\Sigma n_i/e+n_1b) - n_1e}{e(e+n_1b)^2(\Sigma n_i/e+n_1b)} & \frac{-n_i}{(e+n_1b)(e+n_1b)(\Sigma n_i/e+n_1b)} \\
\frac{-(e+n_1b)(\Sigma n_i/e+n_1b) - n_1e}{e(n_1b)^2(\Sigma n_i/e+n_1b)} & \frac{e+n_1b)(\Sigma n_i/e+n_1b) - n_1e}{e+n_1b)^2(\Sigma n_i/e+n_1b)}
\end{pmatrix}^p
\]

\[
= \sum_{i=1}^{p} n_i \left[ \frac{(e+n_1b)(\Sigma n_i/e+n_1b) - n_1e}{(e+n_1b)^2(\Sigma n_i/e+n_1b)} \right]^2 + \sum_{j \neq i} \frac{n_j}{((e+n_1b)(e+n_1b)(\Sigma n_i/e+n_1b))^2}
\]
\[ \frac{m(n-m)}{(e+mb)^2} \] for balanced data. (3.8)

These traces show again how important the expression \( e + nb \) is to ML and REML estimation. Furthermore, they will be re-used in calculating measures of efficiency (Chapter 4).

Finally, for both ML and REML, the vector \( d \) comes from (2.8) and (2.9) and is made up of the two elements:

\[
(y - X\hat{\alpha})^T V^{-2} (y - X\hat{\alpha}) = \frac{1}{e^2} \sum_{i=1}^{p} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 + \sum_{i=1}^{p} \frac{n_i(y_i - \hat{\alpha})^2}{(e + nb)^2}
\]

\[
(y - X\hat{\alpha})^T V^{-4} GV^{-4} (y - X\hat{\alpha}) = \sum_{i=1}^{p} \frac{n_i(y_i - \hat{\alpha})^2}{(e + nb)^2}.
\]

The iterative part of the algorithm consists of choosing starting values for \( e \) and \( b \), substituting them into \( B \) or \( B_1 \) and \( d \), and solving for the new values of \( e \) and \( b \). The algorithm will always converge, because \( B \) and \( B_1 \) are always positive semidefinite. The proof for \( B \) is given below:

For a matrix to be positive semidefinite, we require that \( \delta^T B \delta \geq 0 \) for any vector \( \delta \). Now

\[
\delta^T B \delta = \sum_{j,k} [B]_{jk} \delta_j \delta_k
\]

\[
= \sum_j \text{tr}[V^{-1}(\partial V/\partial \theta_j) V^{-1}(\partial V/\partial \theta_k)] \delta_j \delta_k
\]

\[
= \text{tr} \left[ V^{-2} \sum_j \partial V/\partial \theta_j \delta_j V^{-1} \sum_k \partial V/\partial \theta_k \delta_k \right]
\]

\[
= \text{tr} \left[ \left( V^{-1} \sum_j \partial V/\partial \theta_j \delta_j \right)^2 \right]
\]

which is non-negative unless \( V^{-1} \sum_j \partial V/\partial \theta_j \delta_j = 0 \) which is impossible since \( V^{-1} \) is non-singular and the \( \partial V/\partial \theta_j \) are linearly independent.

The proof for \( B_1 \) cannot follow the same route because \( P \), which replaces \( V^{-1} \) throughout the proof, is definitely singular! But “for fixed θ with V nonsingular, \( B_1 \) is positive semidefinite and the linear system \( B_1 \theta = d \) is consistent for θ (LaMotte 1973, pp. 316 and 327–8). The matrix \( B_1 \) is nonsingular if and only if \( \theta_i \) is estimable in the class of quadratic translation-invariant estimators for \( i = 1, \ldots, m \) (again see LaMotte 1973), in which case \( B_1 \theta = d \) is equivalent to \( \theta = B_1^{-1}d \).” (Harville (1977), p. 328).
3.2 HENDERSON’S ALGORITHM

The derivation of Henderson’s algorithm uses the lengthy manipulation of the log likelihood given in Lemmas A1.5 – A1.8. From Lemma A1.7
\[ \varepsilon = y^T(y - X\hat{\alpha} - Z\hat{\beta})/n \] (3.9)
and from Lemma A1.6
\[ \partial L/\partial \theta_i = (-1/2) \text{tr}[(I - (I + Z^T R^{-1}ZD)^{-1})D^{-1}(\partial D/\partial \theta_i)] + (1/2)\hat{\beta}^T D^{-1}(\partial D/\partial \theta_i)D^{-1}\hat{\beta}. \]
Recalling that \( \theta = (\varepsilon, b)^T \), we can substitute \( b \) for \( \theta_i \) and obtain
\[ \partial L/\partial b = (-1/2) (1/b) (p - \text{tr}[(I + Z^T R^{-1}ZD)^{-1}]) + (1/2b^2)\hat{\beta}^T \beta = 0 \text{ at a stationary point} \]
\[ \Rightarrow \hat{b} = \hat{\beta}^T \beta/(p - \text{tr}[(I + Z^T R^{-1}ZD)^{-1}]). \] (3.10)

Similarly for REML estimation, Lemma A1.8 proves that
\[ \varepsilon = y^T(y - X\hat{\alpha} - Z\hat{\beta})/(n - 1) \] (3.11)
and Lemma A1.6 proves that
\[ \partial L_1/\partial \theta_i = (-1/2) \text{tr}[(I - (I + Z^T S ZD)^{-1})D^{-1}(\partial D/\partial \theta_i)] + (1/2)\hat{\beta}^T D^{-1}(\partial D/\partial \theta_i)D^{-1}\hat{\beta}. \]
Since \( \theta = (\varepsilon, b)^T \):
\[ \partial L_1/\partial b = (-1/2) (1/b) (p - \text{tr}[(I + Z^T S ZD)^{-1}]) + (1/2b^2)\hat{\beta}^T \beta = 0 \text{ at a stationary point} \]
\[ \Rightarrow \hat{b} = \hat{\beta}^T \beta/(p - \text{tr}[(I + Z^T S ZD)^{-1}]). \] (3.12)

Expressions for \( \hat{\alpha}, \hat{\beta}, (I + Z^T R^{-1}ZD) \) and \( (I + Z^T S ZD) \) in the one–way case can be found at (2.2), (2.3), Lemma A1.2 and §A1.5. They all depend on \( \varepsilon \) and \( b \). The iterative part of the algorithm consists of choosing starting values for \( \varepsilon \) and \( b \); substituting them in the right hand side of (3.9) and (3.10), or (3.11) and (3.12); thus obtaining the updated values of \( \varepsilon \) and \( b \).

3.3 ILLUSTRATION WITH REAL DATA

The soybean data used to illustrate these procedures is documented in §2.1, and Splus functions to calculate the MLEs and REMLEs are at (A2.5) – (A2.8). The estimates and number of iterations required to find them were:

<table>
<thead>
<tr>
<th>method</th>
<th>( \hat{\varepsilon} )</th>
<th>( \hat{b} )</th>
<th># iterations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anderson ML</td>
<td>6.82</td>
<td>1.9896</td>
<td>2</td>
</tr>
<tr>
<td>Henderson ML</td>
<td>6.82</td>
<td>1.99</td>
<td>11</td>
</tr>
<tr>
<td>Anderson REML</td>
<td>6.82</td>
<td>2.828</td>
<td>2</td>
</tr>
<tr>
<td>Henderson REML</td>
<td>6.82</td>
<td>2.83</td>
<td>9</td>
</tr>
</tbody>
</table>
Anderson’s algorithm is to be preferred in this case because it converges more quickly. The REMLEs are equal to ANOVA estimates because the data set is balanced, and as such could be calculated in one iteration.

Now that estimates of the variance components have been found, the next step is to calculate the variance of the estimates in order to compare the efficiency of ML and REML.
4. MEASURES OF EFFICIENCY

4.1 GENERAL THEORY

A first-order approximation to the variance of the MLE $\hat{\theta}$ of $\theta$ is

$$[E(\partial^2 L/\partial \theta \partial \theta^T)]^{-1} [E(\partial L/\partial \theta) (\partial L/\partial \theta)^T] [E(\partial^2 L/\partial \theta \partial \theta^T)]^{-1},$$

which is found by expanding $L = \log f(y, \theta)$ in a Taylor series around $\hat{\theta}$. But $-E(\partial L/\partial \theta)(\partial L/\partial \theta)^T = E(\partial^2 L/\partial \theta \partial \theta^T)$ when the distribution of $y$ truly is $f$ since

$$E(\partial^2 L/\partial \theta \partial \theta^T) = E(\partial^2 \log f/\partial \theta \partial \theta^T)$$

$$= \int \left( \frac{1}{f} \right)^2 \frac{\partial f}{\partial \theta} \frac{\partial f}{\partial \theta^T} f \, dy$$

$$= -E(\partial \log f/\partial \theta)(\partial \log f/\partial \theta)^T$$

Thus the information matrix $\mathbf{F} = -E(\partial^2 L/\partial \theta \partial \theta^T)$ and the variance matrix $\mathbf{F}^{-1}$. The diagonal elements of $\mathbf{F}^{-1}$ are of course the variances of the individual estimates, and their value is a lower bound on the variance of any estimate of $\theta$.

4.2 ONE-WAY VARIANCE COMPONENTS MODEL

4.2.1 MAXIMUM LIKELIHOOD

The general second derivatives of the ML equation, and their expected values are at (A1.6). Letting $\theta = (e, b)^T$, the elements of the information matrix $\mathbf{F}$ are:

$$E(\partial^2 L/\partial e^2) = -E(\partial L/\partial e)^2 = (-1/2) \text{tr}[V^{-2}]$$

$$E(\partial^2 L/\partial e \partial b) = -E(\partial L/\partial e)(\partial L/\partial b) = (-1/2) \text{tr}[V^{-1}GV^{-1}]$$

$$E(\partial^2 L/\partial b^2) = -E(\partial L/\partial b)^2 = (1/2) \text{tr}[V^{-1}GV^{-1}].$$

There is no need to evaluate information and variances for $\alpha$ because fixed effects are not the focus of this thesis, and the co-information between $\alpha$ and the variance components, $-E(\partial^2 L/\partial \alpha \partial \theta_i) = 0$ i.e. the information matrix including $\alpha$ is block diagonal and information about $\alpha$ does not affect information about the variance components.

For the one-way model, (3.3) – (3.5) are used to calculate each piece of information as follows:

$$\text{info}(\hat{e}) = (1/2) \text{tr}[V^{-2}] = \frac{1}{2} \sum_{i=1}^{p} n_i \frac{(e^2 + 2(n_i - 1)eb + n_i(n_i - 1)b^2)}{e^2(e + n_i b)^2}$$

(4.1)
co-info(e, b) = (1/2) tr[V^tGV^{-1}] = \frac{1}{2} \sum_{i=1}^{p} \frac{n_i}{(e+n_ib)^2} \tag{4.2}

info(b) = (1/2) tr[V^tGV] = \frac{1}{2} \sum_{i=1}^{p} \frac{n_i^2}{(e+n_ib)^2} \tag{4.3}

Thus the information matrix, F, is

$$
\begin{pmatrix}
\frac{1}{2} \sum_{i=1}^{p} \frac{n_i}{(e+n_ib)^2} & \frac{1}{2} \sum_{i=1}^{p} \frac{n_i}{(e+n_ib)^2} \\
\frac{1}{2} \sum_{i=1}^{p} \frac{n_i^2}{(e+n_ib)^2} & \frac{1}{2} \sum_{i=1}^{p} \frac{n_i^2}{(e+n_ib)^2}
\end{pmatrix}
$$

(4.4)

and its inverse, \( Var(\hat{\theta}) = \begin{pmatrix} \text{var}(\hat{e}) & \text{cov}(\hat{e}, \hat{b}) \\ \text{cov}(\hat{e}, \hat{b}) & \text{var}(\hat{b}) \end{pmatrix} \)

$$
\begin{pmatrix}
\frac{2e^2(\Sigma n_i e/e+n_ib)^2}{n(\Sigma n_i e/e+n_ib)^2 - (\Sigma n_i e/e+n_ib)^2} & \frac{-2e^2(\Sigma n_i e^2/(e+n_ib)^2)}{n(\Sigma n_i e^2/(e+n_ib)^2) - (\Sigma n_i e^2/(e+n_ib)^2)} \\
\frac{-2e^2(\Sigma n_i e^2/(e+n_ib)^2)}{n(\Sigma n_i e^2/(e+n_ib)^2) - (\Sigma n_i e^2/(e+n_ib)^2)} & \frac{2(e^2+2(n_i-1)e+n_i(n_i-1)b^2)}{n(\Sigma n_i e^2/(e+n_ib)^2) - (\Sigma n_i e^2/(e+n_ib)^2)}
\end{pmatrix}
$$

(4.5)

Searle (1970) rewrites info(\( \hat{e} \)) as \( \frac{1}{2e^2} \sum_{i=1}^{p} \frac{1}{(e+n_ib)^2} + n - p \) but this is equivalent to (4.1) only if the data is balanced. If that is the case, the variance matrix of the variance components is

$$
\text{Var}(\hat{\theta}) = \begin{pmatrix}
\frac{2e^2}{n-p} & \frac{-2e^2}{m(n-p)} \\
\frac{-2e^2}{m(n-p)} & \frac{2(e^2+2(n_i-1)e+n_i(n_i-1)b^2)}{m^2(n-p) + \frac{(e+mb)^2}{p}}
\end{pmatrix}
$$

(4.6)
4.2.2 RESTRICTED MAXIMUM LIKELIHOOD

As given in (A1.6), the elements of the information matrix are

\[-E(\partial^2 L_1 / \partial \theta \partial \theta^T) = \frac{1}{2} \text{tr}[P(\partial V / \partial \theta)P(\partial V / \partial \theta)^T].\]

Using (3.6) – (3.8) to evaluate each piece of information:

2 info(\(\hat{e}\)) = \text{tr}[P^2]

\[
= \sum_{i=1}^{p} n_i \left[ \frac{((e+n_i b)(e+(n_i-1)b)(\Sigma n_i/e+n_i b) - e)}{e(e+n_i b)^2(\Sigma n_i/e+n_i b)} \right]^2
+ (n_i-1) \left( \frac{-b(e+n_i b)(\Sigma n_i/e+n_i b) - e}{e(e+n_i b)^2(\Sigma n_i/e+n_i b)} \right)^2
+ \sum_{j \neq i} \frac{n_j}{((e+n_i b)(e+n_j b)(\Sigma n_i/e+n_j b))^2} \right] \tag{4.7}

= \frac{n(e^2+2(m-1)eb+m(m-1)b^2)}{e^2(e+mb)^2} \text{ if the experiment is balanced}

2 co-info(\(\hat{e}, \hat{b}\)) = \text{tr}[PGP]

\[
= \sum_{i=1}^{p} n_i \left[ (e+n_i b)(\Sigma n_i/e+n_i b) - n_i e)((e+n_i b)(e+(n_i-1)b)(\Sigma n_i/e+n_i b) - e) \right]
\left( e(e+n_i b)^2(\Sigma n_i/e+n_i b) \right)^2
+ (n_i-1) \left( \frac{-b(e+n_i b)(\Sigma n_i/e+n_i b) - e)(e+n_i b)(\Sigma n_i/e+n_i b) - n_i e}{e(e+n_i b)^2(\Sigma n_i/e+n_i b)} \right)
+ \sum_{j \neq i} \frac{n_j n_i}{((e+n_i b)(e+n_j b)(\Sigma n_i/e+n_i b))^2} \right] \tag{4.8}

= \frac{n-m}{(e+mb)^2} \text{ for balanced data}

2 info(\(\hat{b}\)) = \text{tr}[PGPG]

\[
= \sum_{i=1}^{p} n_i \left[ n_i \left( \frac{(e+n_i b)(\Sigma n_i/e+n_i b) - n_i}{e+mb)^2(\Sigma n_i/e+n_i b)} \right)^2
+ \sum_{j \neq i} \frac{n_i^2}{((e+n_i b)(e+n_j b)(\Sigma n_i/e+n_j b))^2} \right] \tag{4.9}

= \frac{m(n-m)}{(e+mb)^2} \text{ for balanced data.}
Thus in the case of balanced data the information matrix, $F_1$, is

$$
\begin{bmatrix}
\frac{n(e^2 + 2(m-1)eb + m(m-1)b^2)}{2e^2(e+mb)^2} & \frac{n-m}{2(e+mb)^2} \\
\frac{n-m}{2e^2(e+mb)^2} & \frac{m(n-m)}{2(e+mb)^2}
\end{bmatrix}
$$

(4.10)

which has as inverse

$$
\begin{bmatrix}
\frac{2e^2}{n-p} & -\frac{2e^2}{m(n-p)} \\
-\frac{2e^2}{m(n-p)} & \frac{2}{m^2}\left(\frac{e^2}{n-p} + \frac{(e+mb)^2}{p-1}\right)
\end{bmatrix}
$$

(4.11)

4.3 DISCUSSION
4.3.1 BALANCED DATA

Exact results are available since ANOVA theory (§2.5) shows that certain linear combinations of the MLEs follow certain $\chi^2$ distributions, namely:

$$
\frac{(n-p)e}{e} \sim \chi^2 \text{ on } n-p \text{ degrees of freedom, and}
\frac{p(e + mb)}{(e + mb)} \sim \chi^2 \text{ on } p-1 \text{ degrees of freedom.}
$$

It follows that $\text{Var}(\hat{e}) = \frac{2e^2}{n-p}$ and $\text{Var}(\hat{b}) = \frac{2}{m^2}\left(\frac{e^2}{n-p} + \frac{(p-1)(e+mb)^2}{p^2}\right)$

Linear combinations of the REMLEs follow these $\chi^2$ distributions:

$$
\frac{(n-p)e}{e} \sim \chi^2 \text{ on } n-p \text{ degrees of freedom, and}
\frac{(p-1)(e + mb)}{(e + mb)} \sim \chi^2 \text{ on } p-1 \text{ degrees of freedom.}
$$

It follows that $\text{Var}(\hat{e}) = \frac{2e^2}{n-p}$ and $\text{Var}(\hat{b}) = \frac{2}{m^2}\left(\frac{e^2}{n-p} + \frac{(e+mb)^2}{p-1}\right)$

Thus the variance of the REMLE of $e$ equals the variance of the MLE, but the variance of the REMLE of $b$ is larger than that of the MLE because $1/(p-1) > (p-1)/p^2$. 22
The large-sample variances, which are the diagonal elements of (4.6) and (4.11), display similar behaviour. We see again that the variances of MLEs and REMLEs for \( e \) are equal, so that REML is efficient for \( e \). For \( \hat{b} \), although \( 1/(p - 1) > 1/p \), the two are indistinguishable since the variances are arrived at using first-order Taylor series approximations. Letting \( p \to \infty \), REML estimation of \( b \) is asymptotically efficient.

Comparison of the variance of MLEs and REMLEs can also be found in published articles e.g. Corbeil and Searle (1976b).

4.3.2 UNBALANCED DATA

No exact results are available, and although the information matrices can be constructed in a straightforward manner from (4.4) and (4.7) – (4.9), they are not illuminating due to the complex nature of the traces involved. It is not possible to compare the information matrices and thereby comment on the relative efficiency of the two types of estimate either, because the action of inverting the matrix causes its elements to interact in a complex way that prevents its use as a substitute for its inverse.

Nevertheless it is possible to make some comments on the asymptotic behaviour for both small and large \( e \) and \( b \) by examining the variances of \( \hat{e} \) and \( \hat{b} \) for \( e \) and \( b = 1, 2, 5, \) and 10 and the following unbalanced design matrices:

\[
Z = \begin{pmatrix}
15 \\
14 \\
15 \\
15
\end{pmatrix}
\]

two replicates; \[
\begin{pmatrix}
Z \\
Z
\end{pmatrix}
\]
and five replicates:

\[
\begin{pmatrix}
Z \\
Z \\
Z
\end{pmatrix}
\]

The variances of \( \hat{e} \) are equal under ML and REML so only the values of Var(\( \hat{b} \)) are given in Table 4.1.
Table 4.1 Var(\hat{b}) for unbalanced experiments

<table>
<thead>
<tr>
<th>one replicate</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ML</td>
<td>REML</td>
<td>REML</td>
<td>REML</td>
<td>REML</td>
<td>REML</td>
<td>REML</td>
<td>REML</td>
<td>REML</td>
</tr>
<tr>
<td>e = 1</td>
<td>b = 1</td>
<td>2</td>
<td>5</td>
<td>10</td>
<td>b = 1</td>
<td>2</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>1</td>
<td>0.59</td>
<td>1.96</td>
<td>10.86</td>
<td>41.70</td>
<td>0.74</td>
<td>2.45</td>
<td>13.58</td>
<td>52.13</td>
</tr>
<tr>
<td>2</td>
<td>0.82</td>
<td>2.36</td>
<td>11.77</td>
<td>43.45</td>
<td>1.02</td>
<td>2.94</td>
<td>14.70</td>
<td>54.30</td>
</tr>
<tr>
<td>5</td>
<td>1.78</td>
<td>3.82</td>
<td>14.74</td>
<td>48.95</td>
<td>2.21</td>
<td>4.76</td>
<td>18.41</td>
<td>61.16</td>
</tr>
<tr>
<td>10</td>
<td>4.26</td>
<td>7.14</td>
<td>20.58</td>
<td>59.00</td>
<td>5.22</td>
<td>8.82</td>
<td>25.62</td>
<td>73.62</td>
</tr>
<tr>
<td>two replicates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ML</td>
<td>REML</td>
<td>REML</td>
<td>REML</td>
<td>REML</td>
<td>REML</td>
<td>REML</td>
<td>REML</td>
<td>REML</td>
</tr>
<tr>
<td>e = 1</td>
<td>b = 1</td>
<td>2</td>
<td>5</td>
<td>10</td>
<td>b = 1</td>
<td>2</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>1</td>
<td>0.29</td>
<td>0.98</td>
<td>5.43</td>
<td>20.85</td>
<td>0.33</td>
<td>1.09</td>
<td>6.03</td>
<td>23.17</td>
</tr>
<tr>
<td>2</td>
<td>0.41</td>
<td>1.18</td>
<td>5.88</td>
<td>21.72</td>
<td>0.46</td>
<td>1.31</td>
<td>6.54</td>
<td>24.14</td>
</tr>
<tr>
<td>5</td>
<td>0.89</td>
<td>1.91</td>
<td>7.37</td>
<td>24.47</td>
<td>0.99</td>
<td>2.11</td>
<td>8.19</td>
<td>27.19</td>
</tr>
<tr>
<td>10</td>
<td>2.12</td>
<td>3.56</td>
<td>10.29</td>
<td>29.49</td>
<td>2.34</td>
<td>3.94</td>
<td>11.41</td>
<td>32.74</td>
</tr>
<tr>
<td>five replicates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ML</td>
<td>REML</td>
<td>REML</td>
<td>REML</td>
<td>REML</td>
<td>REML</td>
<td>REML</td>
<td>REML</td>
<td>REML</td>
</tr>
<tr>
<td>e = 1</td>
<td>b = 1</td>
<td>2</td>
<td>5</td>
<td>10</td>
<td>b = 1</td>
<td>2</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>1</td>
<td>0.12</td>
<td>0.39</td>
<td>2.17</td>
<td>8.34</td>
<td>0.12</td>
<td>0.41</td>
<td>2.62</td>
<td>8.69</td>
</tr>
<tr>
<td>2</td>
<td>0.16</td>
<td>0.47</td>
<td>2.35</td>
<td>8.69</td>
<td>0.17</td>
<td>0.49</td>
<td>2.45</td>
<td>9.05</td>
</tr>
<tr>
<td>5</td>
<td>0.36</td>
<td>0.76</td>
<td>2.95</td>
<td>9.79</td>
<td>0.37</td>
<td>0.79</td>
<td>3.07</td>
<td>10.19</td>
</tr>
<tr>
<td>10</td>
<td>0.85</td>
<td>1.43</td>
<td>4.11</td>
<td>11.80</td>
<td>0.88</td>
<td>1.48</td>
<td>4.28</td>
<td>12.28</td>
</tr>
</tbody>
</table>

As the experimental design gets larger, the variance of REMLEs approaches that of MLEs. Convergence appears to be faster for small values of the variance components. Furthermore, since MLEs of b are biased but REMLEs are not, this convergence is even faster if MSE rather than variance is tabulated, since MSE is larger than the corresponding variance. In summary, this study points to REMLEs being asymptotically efficient for unbalanced designs as well as balanced.
4.4 ILLUSTRATION WITH REAL DATA

Splus functions to calculate the variance of estimates of components of variance are at (A2.9) – (A2.10). They are written in terms of matrices rather than using, for example, (4.5), because they look more elegant and matrices are a more general way of expressing the calculations involved.

For the soybean data (see §2.1) \( F^{-1} = \begin{pmatrix} 4.65 & -0.93 \\ -0.93 & 4.68 \end{pmatrix} \) and \( F_1^{-1} = \begin{pmatrix} 4.65 & -0.93 \\ -0.93 & 8.97 \end{pmatrix} \)

The relationships described in §4.3.1 clearly hold.

The next chapter completes the discussion of ML and REML estimation for normally-distributed data by examining distributions of estimates using a Monte Carlo experiment.
For this Monte Carlo study, two variance component models were considered with

\[
Z = \begin{pmatrix}
15 & 15 \\
15 & 15 \\
15 & 15 \\
\end{pmatrix}
\] (balanced) or \(Z = \begin{pmatrix}
15 & 14 \\
15 & 15 \\
\end{pmatrix}
\) (unbalanced). \(5.1\)

The variance components were chosen to be \(e = 4\) and \(b = 16\). By estimating \(e\) and \(b\) and examining the expected value, variance and number of iterations required to find the estimates, we can compare the performance of Anderson’s and Henderson’s algorithm. We can also examine the distributional properties of the estimates.

### 5.1 EXPECTED SUMMARY STATISTICS

Firstly, we use the fact that all MLEs (including REMLEs) \(\hat{\theta}\) are asymptotically normally distributed with mean \(E(\hat{\theta})\) and variance \([-E(\partial^2 L/\partial \theta \partial \theta^T)]^{-1}\) to produce a table of expected results (Table 5.1) to compare with the simulated results.

The variances of MLEs and REMLEs from both balanced and unbalanced experiments are calculated using Splus functions (A2.9) and (A2.10), which were introduced in Chapter 4.

To find the expected value of MLEs, start with (3.1) and take expected values. Thus

\[
E(\hat{\theta}) = B^{-1} E(d) = \begin{pmatrix}
\text{tr}[V^{-2}] & \text{tr}[V^{-1}G^{-1}] \\
\text{tr}[V^{-1}G^{-1}] & \text{tr}[V^{-1}G^{-1}G] \\
\end{pmatrix}^{-1} \begin{pmatrix}
\text{tr}[P] \\
\text{tr}[PG] \\
\end{pmatrix}
\]

Expressions in \(e, m\) and \(b\) can be substituted for the traces so that for balanced data

\[
E\begin{pmatrix}
\hat{e} \\
\hat{b} \\
\end{pmatrix} = \begin{pmatrix}
\frac{e^2}{n-p} & \frac{-e^2}{m(n-p)} \\
\frac{-e^2}{m(n-p)} & \frac{1}{m^2 n - p} + \frac{(e + mb)^2}{p} \\
\end{pmatrix} \begin{pmatrix}
\frac{n(e + (m - 1)b) - e}{e(e + mb)} \\
\frac{n - m}{e + mb} \\
\end{pmatrix}
\]

Expanding each row of the system:
\[
E(\hat{e}) = \frac{e^2}{n-p} \left( \frac{(n-1)e + n(m-1)b}{e(e+mb)} \right) - \frac{e^2(n-m)}{m(n-p)(e+mb)} \\
= \frac{me((n-1)e - e(p-1) + n(m-1)b)}{m(n-p)(e+mb)} \\
= \frac{e((n-p)e + m(n-p)b)}{(n-p)(e+mb)} \\
= e \text{ (unbiased)}
\]

\[
E(\hat{b}) = \frac{-ne((n-1)e + n(m-1)b)}{mn(n-p)(e+mb)} + \frac{(n-m)(ne^2 + 2(n-p)meb + (n-p)m^2b^2)}{m(e+mb)n(n-p)} \\
= \frac{(nb - mb - e)(n-p)(e+mb)}{n(n-p)(e+mb)} \\
= b - \frac{e + mb}{n} \text{ (negatively biased)}. 
\]

The expectations for unbalanced MLEs follow a similar pattern and are calculated using the Splus function (A2.11). All REMLEs are unbiased, which can be derived from (3.2) and checked using (A2.12).

For \( e = 4 \) and \( b = 16 \), the above results are summarised in Table 5.1.

<table>
<thead>
<tr>
<th>method, data</th>
<th>( E(\hat{e}) )</th>
<th>( \text{Var}(\hat{e}) )</th>
<th>( E(\hat{b}) )</th>
<th>( \text{Var}(\hat{b}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>ML balanced</td>
<td>4</td>
<td>1.6</td>
<td>12.64</td>
<td>112.96</td>
</tr>
<tr>
<td>REML balanced</td>
<td>4</td>
<td>1.6</td>
<td>16</td>
<td>141.18</td>
</tr>
<tr>
<td>ML unbalanced</td>
<td>4</td>
<td>1.68</td>
<td>12.63</td>
<td>113.50</td>
</tr>
<tr>
<td>REML unbalanced</td>
<td>4</td>
<td>1.68</td>
<td>16</td>
<td>141.86</td>
</tr>
</tbody>
</table>

This table presents an ideal that the Monte Carlo experiment is aiming to reproduce. The actual results of the experiment are presented in the next section.

5.3 OBSERVED SUMMARY STATISTICS

Table 5.2 lists the observed expected values and variances of 200 estimates of \( e \) and \( b \) for the two experimental designs and two algorithms under consideration, obtained using the Splus functions (A2.13) – (A2.15). The distributions of the estimates are
displayed in Figures 5.1 and 5.2. In Figure 5.1, the estimates are compared with the normal distribution; in Figure 5.2, linear combinations of the estimates are compared with $\chi^2$ distributions on the number of degrees of freedom given in §2.5.

Table 5.2. Observed Summary Statistics (200 trials)

<table>
<thead>
<tr>
<th>method, data</th>
<th>$E(\hat{e})$</th>
<th>var($\hat{e}$)</th>
<th>$E(\hat{b})$</th>
<th>var($\hat{b}$)</th>
<th>$-ve$ estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anderson ML balanced</td>
<td>4.15</td>
<td>2.05</td>
<td>13.89</td>
<td>90.64</td>
<td>none</td>
</tr>
<tr>
<td>Henderson ML balanced</td>
<td>4.15</td>
<td>2.05</td>
<td>13.89</td>
<td>90.64</td>
<td>none</td>
</tr>
<tr>
<td>Anderson REML balanced</td>
<td>4.15</td>
<td>2.05</td>
<td>17.57</td>
<td>141.54</td>
<td>none</td>
</tr>
<tr>
<td>Henderson REML balanced</td>
<td>4.15</td>
<td>2.05</td>
<td>17.57</td>
<td>141.54</td>
<td>none</td>
</tr>
<tr>
<td>Anderson ML unbalanced</td>
<td>4.00</td>
<td>1.70</td>
<td>11.88</td>
<td>83.95</td>
<td>3 of $b$</td>
</tr>
<tr>
<td>Henderson ML unbalanced</td>
<td>4.00</td>
<td>1.70</td>
<td>11.88</td>
<td>83.89</td>
<td>none</td>
</tr>
<tr>
<td>Anderson REML unbalanced</td>
<td>4.00</td>
<td>1.70</td>
<td>15.06</td>
<td>130.93</td>
<td>1 of $b$</td>
</tr>
<tr>
<td>Henderson REML unbalanced</td>
<td>4.00</td>
<td>1.70</td>
<td>15.06</td>
<td>130.89</td>
<td>none</td>
</tr>
</tbody>
</table>

A smaller Monte Carlo study was also undertaken to determine the average number of iterations (to the nearest integer) required for convergence of the two algorithms in the four estimation situations (balanced or unbalanced data; ML or REML estimation). Table 5.3 lists the results.

Table 5.3. Average number of iterations required for convergence (5 trials)

<table>
<thead>
<tr>
<th>data</th>
<th>method</th>
<th># iterations</th>
<th>data</th>
<th>method</th>
<th># iterations</th>
</tr>
</thead>
<tbody>
<tr>
<td>balanced</td>
<td>Anderson ML</td>
<td>2</td>
<td>unbalanced</td>
<td>Anderson ML</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Henderson ML</td>
<td>8</td>
<td></td>
<td>Henderson ML</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Anderson REML</td>
<td>2</td>
<td></td>
<td>Anderson REML</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Henderson REML</td>
<td>7</td>
<td></td>
<td>Henderson REML</td>
<td>7</td>
</tr>
</tbody>
</table>

5.3 DISCUSSION

Comparing the performance of the algorithms as summarised in Table 5.2, the algorithms are almost the same. However, as noted by Searle (1987), the solution to (3.1) is not necessarily a MLE because MLEs are by definition in their parameter space i.e. Anderson’s algorithm occasionally produces negative estimates. (The number of negative estimates here is however very small.) Estimates by Henderson’s algorithm are always in their parameter space, partly because (as proven by Harville (1975, 1977)), the denominator of equations (3.10) and (3.12) is strictly positive. On the other hand,
Figure 5.1 Normal probability plots of MLEs and REMLEs
Figure 5.2 Chi-square probability plots of MLEs and REMLEs
Table 5.3 shows that Anderson's algorithm always converges more quickly, so that neither algorithm is clearly preferable over the other.

Comparing Table 5.1 with Table 5.2, we see that the observed summary statistics are within about 20% of their theoretical values, and some e.g. unbalanced $E(\hat{\varepsilon})$, are exact. However, the Normal probability plots (Figure 5.1) display distinct curvature, indicating non-normal distributions. The plots of sorted linear combinations of observed estimates against $\chi^2$ quantiles (Figure 5.2) are, in general, much straighter, indicating that the quantities plotted are distributed as $\chi^2$ variables. Furthermore, as shown in Figure 5.3, taking the cube root of the estimates induces near-normality in the distribution of the estimates. Thus it may be more appropriate for inferential purposes to work with the transformed estimates, rather than the raw ones.

The smallness of the experimental designs in (5.1) is probably the reason why the large-sample normal distributions do not seem to hold, but the small-sample $\chi^2$ distributions do hold quite well. The effect of having such a small design is that $e$ is being estimated with fewer than 30 degrees of freedom (observations) and $b$ with, in effect, $p - 1 = 4$ degrees of freedom.

Thus this part of the thesis, dealing with estimation for normally distributed data, concludes with a lesson in experimental design. If components of variance are the parameters requiring estimation, it is advisable to have as many treatments as possible, even if this entails reducing the number of observations per treatment.
Figure 5.3 Normal probability plots of cube root of MLEs and REMLEs.
6. NON-NORMALITY

So far we have always assumed that the random factors $\beta$ and $\epsilon$ come from a $N(0, \sigma_1^2)$ distribution. This normality assumption allows the development of the variance matrices in Chapter 4, where it was found that for the one-way model, REML is asymptotically efficient for both balanced and unbalanced designs.

This chapter will address the questions: what happens to the size of the variances of the estimates when the data is non-normally distributed; does the MSE of the estimates behave in a similar way; and what is the asymptotic behaviour of the MSEs in the presence of non-normality?

6.1 VARIANCES FOR NON-NORMAL DISTRIBUTIONS: GENERAL CASE

The assumed distribution of the data is $f(y, \theta)$, but the actual distribution is $g(y, \theta)$. Then

$$E(\partial L/\partial \theta) = E(\partial \log f/\partial \theta) = \int (\partial g/\partial \theta)/f(\theta) \ g \ dy$$

and

$$E(\partial^2 L/\partial \theta \partial \theta^T) = \int (-1/f)^2(\partial g/\partial \theta)(\partial g/\partial \theta)^T + (1/f)(\partial^2 f/\partial \theta \partial \theta^T) \ g \ dy.$$ 

So

$$\text{var}(\hat{\theta}) = [E(\partial^2 L/\partial \theta \partial \theta^T)]^{-1}[E(\partial L/\partial \theta)(\partial L/\partial \theta)^T][E(\partial^2 L/\partial \theta \partial \theta^T)]^{-1}$$

(6.1)

using the same Taylor series argument as §4.1. Calculation of the matrix (6.1) is the aim of the next section.

The particular non-normal distribution, $g$, considered in this thesis is the contaminated normal i.e. with probability $1 - \delta_0$, $\epsilon \sim N(0, \sigma)$ and with probability $\delta_0$, $\epsilon \sim N(0, k_0 \sigma)$, where $k_0 > 1$. Thus

$$\text{Var}(\epsilon) = (1 - \delta_0 + \delta_0 k_0) \sigma = \sigma^*.$$ 

Similarly

$$\text{Var}(\beta) = (1 - \delta_1 + \delta_1 k_1) \beta = \beta^*.$$ 

(6.2)

Clearly observations from this contaminated normal have variance

$$E(y - X\alpha)(y - X\alpha) = V^* = \begin{pmatrix} e^* + b^* & b^* \\ b^* & e^* + b^* \end{pmatrix}$$

So long as $\beta$ and $\epsilon$ are independent, $V^*$ will have the same structure as $V$ whatever contaminated normal distribution is considered: simply denote the actual variances of $\epsilon$ and $\beta$ by $e^*$ and $b^*$.  

33
6.2 ONE-WAY UNBALANCED MODEL VARIANCE MATRIX

6.2.1 ML

The variance matrix as a whole is

\[
\begin{pmatrix}
E(\partial^2 L/\partial e^2) & E(\partial^2 L/\partial e \partial b) \\
E(\partial^2 L/\partial e \partial b) & E(\partial^2 L/\partial b^2)
\end{pmatrix}^{-1}
\begin{pmatrix}
E(\partial L/\partial e)^2 \\
(E(\partial L/\partial e))(\partial L/\partial b)
\end{pmatrix}
\begin{pmatrix}
E(\partial^2 L/\partial e^2) & E(\partial L/\partial e)(\partial L/\partial b) \\
E(\partial^2 L/\partial e \partial b) & E(\partial^2 L/\partial b^2)
\end{pmatrix}^{-1}.
\]

(6.3)

The expectations that constitute (6.3) are

\[
E(\partial^2 L/\partial e^2) = \mathbb{E}\left[(l/2)\text{tr}(V'(\partial^2 V/\partial e^2)) + (l/2)(y - X\alpha)^T V^\top(\partial^2 V/\partial e \partial e) V^\top(y - X\alpha)\right]
= (1/2)\text{tr}[V^{-2}] - \text{tr}[V^{-3}V^*] 
\]

(6.4)

\[
E(\partial^2 L/\partial e \partial b) = \mathbb{E}\left[(l/2)\text{tr}(V'(\partial^2 V/\partial e \partial b)) - (l/2)(y - X\alpha)^T V^\top(\partial^2 V/\partial e \partial b) V^\top(y - X\alpha)\right]
= (1/2)\text{tr}[V^{-2}G V^*] - \text{tr}[V^{-2}G^2V^*] 
\]

(6.5)

\[
E(\partial^2 L/\partial b^2) = \mathbb{E}\left[(l/2)\text{tr}(V'(\partial^2 V/\partial b^2)) - (l/2)(y - X\alpha)^T V^\top(\partial^2 V/\partial b \partial b) V^\top(y - X\alpha)\right]
= (1/2)\text{tr}[V^{-2}G^2V^*] - \text{tr}[V^{-2}G^2V^*] 
\]

(6.6)

\[
E(\partial L/\partial e)^2 = \mathbb{E}\left[(l/2)\text{tr}(V'^2(\partial^2 V/\partial e)) + (l/2)(y - X\alpha)^T V^\top(\partial^2 V/\partial e) V^\top(y - X\alpha)^2\right]
= (1/4)(\text{tr}[V'^2])^2 + (1/4)\text{tr}[V^{-2}V^*] \text{tr}[V^{-2}V^*] + (1/2)\text{tr}[V^{-2}V^2V*V^*] - (1/2)\text{tr}[V'^2] \text{tr}[V^{-2}V^*] 
\]

(6.7)

\[
E(\partial L/\partial e)(\partial L/\partial b) = \mathbb{E}\left[(l/2)\text{tr}(V'(\partial^2 V/\partial e)) + (l/2)(y - X\alpha)^T V^\top(\partial^2 V/\partial e) V^\top(y - X\alpha)\right]
= (1/4)\text{tr}[V'^2]\text{tr}[V^{-2}G] - (1/4)\text{tr}[V'^2]\text{tr}[V^{-2}G^2V^*] 
\]

(6.8)

\[
E(\partial L/\partial b)^2 = \mathbb{E}\left[(l/2)\text{tr}(V'(\partial^2 V/\partial b)) + (l/2)(y - X\alpha)^T V^\top(\partial^2 V/\partial b) V^\top(y - X\alpha)^2\right]
= (1/4)(\text{tr}[V'^2G])^2 - (1/2)\text{tr}[V'^2G]\text{tr}[V^{-2}G^2V^*] 
\]

(6.9)
It is no longer true that, for example, $E(\frac{\partial^2 L}{\partial \varepsilon^2}) = -E(\frac{\partial L}{\partial \varepsilon})^2$. However, it is true that letting $V^* = V$ in (6.4) – (6.9), the expected values equate to the expectations in the normally distributed case (§4.2.1).

6.2.2 REML

The calculations for REMLEs follow a similar pattern to those just given for MLEs. The variance matrix as a whole is:

$$
\begin{pmatrix}
E(\frac{\partial^2 L_1}{\partial \varepsilon^2}) & E(\frac{\partial^2 L_1}{\partial \varepsilon \partial b}) \\
E(\frac{\partial^2 L_1}{\partial \varepsilon \partial b}) & E(\frac{\partial^2 L_1}{\partial b^2})
\end{pmatrix}
\begin{pmatrix}
E(\frac{\partial L_1}{\partial \varepsilon})^2 & E(\frac{\partial L_1}{\partial \varepsilon})(\frac{\partial L_1}{\partial b}) \\
E(\frac{\partial L_1}{\partial \varepsilon})(\frac{\partial L_1}{\partial b}) & E(\frac{\partial^2 L_1}{\partial b^2})
\end{pmatrix}
-1
\begin{pmatrix}
E(\frac{\partial L_1}{\partial \varepsilon})^2 & E(\frac{\partial L_1}{\partial \varepsilon})(\frac{\partial L_1}{\partial b}) \\
E(\frac{\partial L_1}{\partial \varepsilon})(\frac{\partial L_1}{\partial b}) & E(\frac{\partial^2 L_1}{\partial b^2})
\end{pmatrix}
-1
$$

and the elements of the matrices are:

$$
E(\frac{\partial^2 L_1}{\partial \varepsilon^2}) = E[-\frac{1}{2}\text{tr}(P(\frac{\partial^2 V}{\partial \varepsilon^2}) - (P(\frac{\partial V}{\partial \varepsilon}))^2]
+ \frac{1}{2}(y - X\alpha)^T V^{-1}(\frac{\partial^2 V}{\partial \varepsilon^2}) - 2(\frac{\partial V}{\partial \varepsilon})P(\frac{\partial V}{\partial \varepsilon}))V^{-1}(y - X\alpha)]
= \frac{1}{2}\text{tr}(P^2) - \text{tr}(V^{-1}PV^{-1}(V^* - X(X^TV^*X)^{-1}X^T))
$$

(6.10)

$$
E(\frac{\partial^2 L_1}{\partial \varepsilon \partial b}) = E[-\frac{1}{2}\text{tr}(P(\frac{\partial^2 V}{\partial \varepsilon \partial b}) - P(\frac{\partial V}{\partial b})P(\frac{\partial V}{\partial \varepsilon}))]
+ \frac{1}{2}(y - X\alpha)^T V^{-1}(\frac{\partial^2 V}{\partial \varepsilon \partial b}) - 2(\frac{\partial V}{\partial b})P(\frac{\partial V}{\partial \varepsilon}))V^{-1}(y - X\alpha)]
= \frac{1}{2}\text{tr}(P^2G) - \text{tr}(V^{-1}GPV^{-1}(V^* - X(X^TV^*X)^{-1}X^T))
$$

(6.11)

$$
E(\frac{\partial^2 L_1}{\partial b^2}) = E[-\frac{1}{2}\text{tr}(P(\frac{\partial V}{\partial b}))^2]
+ \frac{1}{2}(y - X\alpha)^T V^{-1}(\frac{\partial^2 V}{\partial b^2}) - 2(\frac{\partial V}{\partial b})P(\frac{\partial V}{\partial b}))V^{-1}(y - X\alpha)]
= \frac{1}{2}\text{tr}(P^2GP) - \text{tr}(V^{-1} GPV^{-1}(V^* - X(X^TV^*X)^{-1}X^T)).
$$

(6.12)

$$
E(\frac{\partial L_1}{\partial \varepsilon})^2 = E[-\frac{1}{2}\text{tr}(P(\frac{\partial V}{\partial \varepsilon})) + \frac{1}{2}(y - X\alpha)^T V^{-1}(\frac{\partial V}{\partial \varepsilon})V^{-1}(y - X\alpha)]^2
= \frac{1}{4}\text{tr}(P^2) - \frac{1}{2}\text{tr}(P)\text{tr}(V^{-2}(V^* - X(X^TV^*X)^{-1}X^T)
+ \frac{1}{4}\text{tr}(V^{-2}(V^* - X(X^TV^*X)^{-1}X^T))^2
+ \frac{1}{2}\text{tr}(V^{-2}(V^* - X(X^TV^*X)^{-1}X^T))^2
$$

(6.13)

$$
E(\frac{\partial L_1}{\partial \varepsilon})(\frac{\partial L_1}{\partial b}) = E[-\frac{1}{2}\text{tr}(P(\frac{\partial V}{\partial \varepsilon})) + \frac{1}{2}(y - X\alpha)^T V^{-1}(\frac{\partial V}{\partial \varepsilon})V^{-1}(y - X\alpha)]
[(\frac{1}{2}\text{tr}(P(\frac{\partial V}{\partial b})) + \frac{1}{2}(y - X\alpha)^T V^{-1}(\frac{\partial V}{\partial b})V^{-1}(y - X\alpha)]]
$$

35
\[
E(\partial L_1/\partial b)^2 = E((-1/2)\text{tr}[P(\partial V/\partial b)] + (1/2)(y - X\delta)^\top V^{-1}(\partial V/\partial b)V^{-1}(y - X\delta))^2
= \frac{1}{4}(\text{tr}[PG])^2 - \frac{1}{2}\text{tr}[PG]\text{tr}[V^{-1}GV'(V^\ast - X(X^TV^\ast''X)'X')]
+ \frac{1}{4}\text{tr}[V^{-1}GV''(V^\ast - X(X^TV^\ast''X)'X')^2]
+ \frac{1}{2}\text{tr}[(V^{-1}GV'(V^\ast - X(X^TV^\ast''X)'X'))^2].
\]

As before, letting \( V^* = V \) in (6.10) – (6.15) reduces these equations to the equations for normally distributed data in §4.2.2.

6.3 NUMERICAL RESULTS

Consider the variance of \( \hat{e} \) and \( \hat{b} \) (estimates of \( e = b = 1 \)) when the contaminated variances \( e^* = (1 - \delta_0 + \kappa_0\delta_0) \) and \( b^* = (1 - \delta_1 + \kappa_1\delta_1) = 1, 1.5, 2, 3 \), using the experimental designs (5.1). Thus we are assuming that the data is normally distributed with \( e = b = 1 \), but in reality it follows a contaminated normal distribution with \( e^* \) and \( b^* \) taking the range of values given above.

It would be possible to write down the expected derivatives (6.4) – (6.15) in terms of \( n_1, \ldots, n_p, e, b, e^* \) and \( b^* \); gather the expressions into matrices and multiply to get the variance matrices. But the derivatives take a complex form and are not very illuminating. The matrix equations are more elegant, and they can easily be converted into computer code (see (A2.16) – (A2.17)) so that the quantitative effects of non-normality can now be studied. Output from these functions for the values of \( e^* \) and \( b^* \) described above is presented in Tables 6.1 and 6.2.

**Table 6.1. Var(\( \hat{e} \)) for the given non-normal distributions**

<table>
<thead>
<tr>
<th></th>
<th>( b^* = 1 )</th>
<th>1.5</th>
<th>2</th>
<th>3</th>
<th>( e^* = 1 )</th>
<th>0.10</th>
<th>0.11</th>
<th>0.11</th>
<th>0.11</th>
</tr>
</thead>
<tbody>
<tr>
<td>( e^* = 1 )</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
<td>1</td>
<td>0.11</td>
<td>0.11</td>
<td>0.11</td>
<td>0.11</td>
</tr>
<tr>
<td>1.5</td>
<td>0.12</td>
<td>0.12</td>
<td>0.12</td>
<td>0.12</td>
<td>1.5</td>
<td>0.12</td>
<td>0.12</td>
<td>0.12</td>
<td>0.12</td>
</tr>
<tr>
<td>2</td>
<td>0.16</td>
<td>0.16</td>
<td>0.16</td>
<td>0.16</td>
<td>2</td>
<td>0.16</td>
<td>0.16</td>
<td>0.16</td>
<td>0.16</td>
</tr>
<tr>
<td>3</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
<td>3</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
</tr>
</tbody>
</table>

Since the ML and REML variances of \( \hat{e} \) are equal, only the balanced and unbalanced cases need be considered. In both cases the variance of \( \hat{e} \) under non-normality is equal to or exceeds the variance of \( \hat{e} \) under normality.
Table 6.2. Var(\(\hat{b}\)) for the given non–normal distributions

Balanced experiment

<table>
<thead>
<tr>
<th></th>
<th>e* = 1</th>
<th>1.5</th>
<th>2</th>
<th>3</th>
<th></th>
<th>e* = 1</th>
<th>1.5</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>ML</td>
<td>0.58</td>
<td>0.42</td>
<td>0.42</td>
<td>0.44</td>
<td>REML</td>
<td>0.72</td>
<td>0.51</td>
<td>0.48</td>
<td>0.49</td>
</tr>
<tr>
<td>1</td>
<td>0.50</td>
<td>0.39</td>
<td>0.38</td>
<td>0.39</td>
<td></td>
<td>1.5</td>
<td>0.62</td>
<td>0.47</td>
<td>0.45</td>
</tr>
<tr>
<td>2</td>
<td>0.45</td>
<td>0.38</td>
<td>0.37</td>
<td>0.38</td>
<td></td>
<td>2</td>
<td>0.56</td>
<td>0.45</td>
<td>0.43</td>
</tr>
<tr>
<td>3</td>
<td>0.40</td>
<td>0.36</td>
<td>0.36</td>
<td>0.37</td>
<td></td>
<td>3</td>
<td>0.49</td>
<td>0.43</td>
<td>0.42</td>
</tr>
</tbody>
</table>

Unbalanced experiment

<table>
<thead>
<tr>
<th></th>
<th>e* = 1</th>
<th>1.5</th>
<th>2</th>
<th>3</th>
<th></th>
<th>e* = 1</th>
<th>1.5</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>ML</td>
<td>0.59</td>
<td>0.43</td>
<td>0.42</td>
<td>0.44</td>
<td>REML</td>
<td>0.74</td>
<td>0.52</td>
<td>0.49</td>
<td>0.50</td>
</tr>
<tr>
<td>1</td>
<td>0.51</td>
<td>0.39</td>
<td>0.39</td>
<td>0.40</td>
<td></td>
<td>1.5</td>
<td>0.63</td>
<td>0.48</td>
<td>0.45</td>
</tr>
<tr>
<td>2</td>
<td>0.45</td>
<td>0.38</td>
<td>0.37</td>
<td>0.38</td>
<td></td>
<td>2</td>
<td>0.57</td>
<td>0.46</td>
<td>0.44</td>
</tr>
<tr>
<td>3</td>
<td>0.40</td>
<td>0.37</td>
<td>0.36</td>
<td>0.37</td>
<td></td>
<td>3</td>
<td>0.49</td>
<td>0.44</td>
<td>0.43</td>
</tr>
</tbody>
</table>

For small experiments, we prefer MLEs on the basis of efficiency since for both balanced and unbalanced data, \(\text{Var}(\text{MLE of } b) < \text{Var}(\text{REML of } b)\).

It is interesting to note that the variances under non–normality are often smaller than under normality. This may be a result of the chosen values of e, b, e*, b* and the small size of the experiment, or it may be a manifestation of a global conservativeness of variance component estimates in the presence of non–normality.

Consider as an alternative to the variance, the MSE of the estimates. To do so, we require the bias of each estimate, which for \(\hat{e}\), is \(E(\hat{e}) - 1 = e* - 1\). For REMLEs of b, the bias = \(b* - 1\). On the other hand, small–sample MLEs of b are further biased so that for the experimental designs (5.1) and the chosen values of e* and b*, the bias is

<table>
<thead>
<tr>
<th></th>
<th>e* = 1</th>
<th>1.5</th>
<th>2</th>
<th>3</th>
<th></th>
<th>e* = 1</th>
<th>1.5</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>ML</td>
<td>-0.24</td>
<td>0.16</td>
<td>0.56</td>
<td>1.36</td>
<td>REML</td>
<td>-0.2419</td>
<td>0.1581</td>
<td>0.5580</td>
<td>1.3580</td>
</tr>
<tr>
<td>1</td>
<td>-0.26</td>
<td>0.14</td>
<td>0.54</td>
<td>1.34</td>
<td></td>
<td>-0.2629</td>
<td>0.1371</td>
<td>0.5371</td>
<td>1.3371</td>
</tr>
<tr>
<td>2</td>
<td>-0.28</td>
<td>0.12</td>
<td>0.52</td>
<td>1.32</td>
<td></td>
<td>-0.2838</td>
<td>0.1162</td>
<td>0.5161</td>
<td>1.3161</td>
</tr>
<tr>
<td>3</td>
<td>-0.32</td>
<td>0.08</td>
<td>0.48</td>
<td>1.28</td>
<td></td>
<td>-0.3256</td>
<td>0.0743</td>
<td>0.4743</td>
<td>1.2742</td>
</tr>
</tbody>
</table>

Adding the squared bias to the variances, the MSE is as shown in Table 6.3. MSE(\(\hat{e}\)) need not be considered because MSE(MLE of e) = MSE(REMLE of e) in the presence of non–normality (cf. Table 6.1).
Table 6.3. MSE(b) for the given non–normal distributions

Balanced experiment

<table>
<thead>
<tr>
<th>ML</th>
<th>b* = 1</th>
<th>1.5</th>
<th>2</th>
<th>3</th>
<th>REML</th>
<th>b* = 1</th>
<th>1.5</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>e* = 1</td>
<td>0.64</td>
<td>0.45</td>
<td>0.73</td>
<td>2.28</td>
<td>e* = 1</td>
<td>0.72</td>
<td>0.76</td>
<td>1.48</td>
<td>4.49</td>
</tr>
<tr>
<td>1.5</td>
<td>0.57</td>
<td>0.41</td>
<td>0.67</td>
<td>2.19</td>
<td>1.5</td>
<td>0.62</td>
<td>0.72</td>
<td>1.45</td>
<td>4.44</td>
</tr>
<tr>
<td>2</td>
<td>0.53</td>
<td>0.39</td>
<td>0.64</td>
<td>2.12</td>
<td>2</td>
<td>0.56</td>
<td>0.70</td>
<td>1.43</td>
<td>4.43</td>
</tr>
<tr>
<td>3</td>
<td>0.50</td>
<td>0.37</td>
<td>0.59</td>
<td>2.01</td>
<td>3</td>
<td>0.49</td>
<td>0.68</td>
<td>1.42</td>
<td>4.42</td>
</tr>
</tbody>
</table>

Unbalanced experiment

<table>
<thead>
<tr>
<th>ML</th>
<th>b* = 1</th>
<th>1.5</th>
<th>2</th>
<th>3</th>
<th>REML</th>
<th>b* = 1</th>
<th>1.5</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>e* = 1</td>
<td>0.65</td>
<td>0.45</td>
<td>0.74</td>
<td>2.29</td>
<td>e* = 1</td>
<td>0.74</td>
<td>0.77</td>
<td>1.49</td>
<td>4.50</td>
</tr>
<tr>
<td>1.5</td>
<td>0.58</td>
<td>0.41</td>
<td>0.67</td>
<td>2.18</td>
<td>1.5</td>
<td>0.63</td>
<td>0.73</td>
<td>1.45</td>
<td>4.45</td>
</tr>
<tr>
<td>2</td>
<td>0.53</td>
<td>0.39</td>
<td>0.64</td>
<td>2.11</td>
<td>2</td>
<td>0.57</td>
<td>0.71</td>
<td>1.44</td>
<td>4.44</td>
</tr>
<tr>
<td>3</td>
<td>0.50</td>
<td>0.37</td>
<td>0.59</td>
<td>2.00</td>
<td>3</td>
<td>0.49</td>
<td>0.69</td>
<td>1.43</td>
<td>4.43</td>
</tr>
</tbody>
</table>

The overall pattern of Table 6.2 is retained: MSE decreases as e* increases for a given b*, while a generally increasing pattern is observed as b* increases for a given e*.

6.4 ASYMPTOTIC BEHAVIOUR

By replicating the experimental design as was done in §4.3.2, we may gain some insight into the asymptotic behaviour of MLEs and REMLEs of b and may see if the asymptotic efficiency of REMLEs carries over to the non–normal situation. Given the various biases involved, it is probably better to continue comparing MSEs.

Table 6.4 Asymptotic behaviour of MSE(b)

<table>
<thead>
<tr>
<th>ML</th>
<th>b* = 1</th>
<th>1.5</th>
<th>2</th>
<th>3</th>
<th>REML</th>
<th>b* = 1</th>
<th>1.5</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>two replicates</td>
<td>e* = 1</td>
<td>0.30</td>
<td>0.36</td>
<td>0.89</td>
<td>3.15</td>
<td>e* = 1</td>
<td>0.32</td>
<td>0.52</td>
<td>1.29</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>0.27</td>
<td>0.33</td>
<td>0.84</td>
<td>3.07</td>
<td>1.5</td>
<td>0.28</td>
<td>0.49</td>
<td>1.26</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.25</td>
<td>0.32</td>
<td>0.82</td>
<td>3.03</td>
<td>2</td>
<td>0.25</td>
<td>0.48</td>
<td>1.25</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.24</td>
<td>0.30</td>
<td>0.79</td>
<td>2.95</td>
<td>3</td>
<td>0.23</td>
<td>0.49</td>
<td>1.25</td>
</tr>
<tr>
<td>five replicates</td>
<td>e* = 1</td>
<td>0.12</td>
<td>0.33</td>
<td>1.02</td>
<td>3.75</td>
<td>e* = 1</td>
<td>0.13</td>
<td>0.40</td>
<td>1.20</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>0.10</td>
<td>0.31</td>
<td>0.99</td>
<td>3.71</td>
<td>1.5</td>
<td>0.11</td>
<td>0.38</td>
<td>1.17</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.10</td>
<td>0.31</td>
<td>0.98</td>
<td>3.68</td>
<td>2</td>
<td>0.10</td>
<td>0.38</td>
<td>1.16</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.10</td>
<td>0.31</td>
<td>0.97</td>
<td>3.65</td>
<td>3</td>
<td>0.10</td>
<td>0.39</td>
<td>1.17</td>
</tr>
</tbody>
</table>
two replicates of unbalanced design

<table>
<thead>
<tr>
<th>ML</th>
<th>e* = 1</th>
<th>0.31</th>
<th>0.36</th>
<th>0.89</th>
<th>3.15</th>
</tr>
</thead>
<tbody>
<tr>
<td>e*</td>
<td>1.5</td>
<td>0.27</td>
<td>0.33</td>
<td>0.84</td>
<td>3.07</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.25</td>
<td>0.32</td>
<td>0.82</td>
<td>3.02</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.24</td>
<td>0.30</td>
<td>0.78</td>
<td>2.94</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>REML</th>
<th>b* = 1</th>
<th>0.32</th>
<th>0.52</th>
<th>1.30</th>
<th>4.34</th>
</tr>
</thead>
<tbody>
<tr>
<td>e*</td>
<td>1.5</td>
<td>0.28</td>
<td>0.50</td>
<td>1.26</td>
<td>4.30</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.26</td>
<td>0.49</td>
<td>1.26</td>
<td>4.28</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.23</td>
<td>0.49</td>
<td>1.25</td>
<td>4.28</td>
</tr>
</tbody>
</table>

five replicates

<table>
<thead>
<tr>
<th>ML</th>
<th>e* = 1</th>
<th>0.12</th>
<th>0.33</th>
<th>1.03</th>
<th>3.76</th>
</tr>
</thead>
<tbody>
<tr>
<td>e*</td>
<td>1.5</td>
<td>0.11</td>
<td>0.31</td>
<td>0.99</td>
<td>3.71</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.10</td>
<td>0.31</td>
<td>0.98</td>
<td>3.68</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.11</td>
<td>0.31</td>
<td>0.97</td>
<td>3.64</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>REML</th>
<th>b* = 1</th>
<th>0.13</th>
<th>0.40</th>
<th>1.20</th>
<th>4.26</th>
</tr>
</thead>
<tbody>
<tr>
<td>e*</td>
<td>1.5</td>
<td>0.11</td>
<td>0.38</td>
<td>1.17</td>
<td>4.22</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.10</td>
<td>0.38</td>
<td>1.17</td>
<td>4.21</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.10</td>
<td>0.39</td>
<td>1.17</td>
<td>4.20</td>
</tr>
</tbody>
</table>

For a given value of b*, as e* increases MSE still decreases. No explanation for this phenomenon can be found, except that it corresponds to a situation where the random variation is swamping the systematic variation — not an ideal experimental situation in which to be. At least we can say that for small values of e* and b*, the MSE of REMLEs in the presence of non-normality approaches the MSE of MLEs in the presence of non-normality, for both balanced and unbalanced designs.

In order to allow for this effect of non-normality when interpreting the variance of estimates, we need ways to identify non-normality in data. The diagnostic tests discussed in Chapter 7 are some suitable solutions.
7. DIAGNOSTIC TESTS

Diagnostic tests cover a wide range of techniques, some graphical, some numerical. In all cases the aim is to detect outlying or influential observations in the data. To date, very few articles have appeared specifically devoted to diagnostic tests for the variance component model. Therefore this chapter contains a brief review of current diagnostic tests, and the extension of these tests to the variance components model. Splus functions (A2.18) – (A2.25) and the soybean data are used to illustrate each test.

7.1 SCATTERPLOT OF OBSERVATIONS

The observations are plotted on the y-axis (Figure 7.1). The ordering along the x-axis is not important so long as points from the same treatment are grouped together. From this very simple plot we can try to see whether the spread is similar across the treatments i.e. whether the $\beta_i$ are all from a $N(0, b)$ distribution, and whether the within-treatments spread is larger than the between-treatments spread i.e. whether $e > b$. It does indeed seem to be the case here that $e > b$, but it is hard to tell if the spread is similar over all five treatments.

7.2 Q–Q PLOTS

For variance components models, where we assume $\beta \sim N(0, b)$, it will be of interest to compare each treatment's observations with all the others. If all the treatment effects are from the same distribution, each Q–Q plot will lie close to a straight line. If none of the plots associated with the $i$th treatment follow this pattern, then the $i$th treatment effect may not come from the same distribution as the others.

This is not the standard format for Q–Q plots, where the order statistics of one sample of independent observations are plotted against the order statistics of a second independent sample. The observations for each treatment in a variance components model are not independent, but (in theory) they have the same dependence structure, thus the interpretation of the Q–Q plots remains as above.

Figure 7.2 displays the Q–Q plots for all pairs of treatments for the soybean data. There is too much curvature in too many of the graphs to believe that all the treatment effects are from the same distribution.
Figure 7.2 Q-Q plot of all pairs of treatments: soybean data
7.3 FITTED EFFECT PLOTS

Whereas Q–Q plots require only the raw data, fitted effect plots are drawn after $e$ and $b$ have been estimated. The $p$ values of $\hat{\beta} = \hat{D}Z'V^{-1}(y - X\hat{\alpha})$ plotted against the quantiles of a $N(0,1)$ distribution should lie in a straight line, as should the $n$ values of $\hat{\epsilon} = y - X\hat{\alpha} - Z\hat{\beta}$. The observed errors $\hat{\epsilon}$ are in fact the residuals for the variance components model.

Figure 7.3 shows the Normal probability plots of the estimated effects for the soybean data. Plots for both ML and REML are given, but they are very similar. The residuals include a number of outliers in the lower tail, but otherwise follow a reasonably normal distribution. It is rather hard to comment on the Q–Q plots of just five treatment effects because the number of points is so small: nevertheless, these graphs seem to display curvature i.e. non-normality.

7.4 REGRESSION–TYPE DIAGNOSTICS

Regression diagnostics come in two varieties – ones that assess influence (often based on deleting one observation at a time) and ones that assess outlyingness (various sorts of residuals). For the ordinary regression model $y = X\beta + e$, where $y$ is an $n \times 1$ vector of observations and $\beta$ is a $p \times 1$ vector of parameters, some regression diagnostics are:

1) Hat matrix diagonals $h_i = [X(X'X)^{-1}X']_{ii}$.
   The $i$th data value is influential if $h_i$ is close to one. "Large" $h_i$ are greater than $2p/n$.

2) DFBETAS$_{ij} = $ scaled change in $j$th parameter estimate on deleting $i$th observation
   \[\frac{\hat{\beta}_i - \hat{\beta}_{i(0)}}{\hat{\sigma}_{(0)} \sqrt{(X'X)^{-1}}_{jj}}\]
   As is usual with statistics scaled by their variance, if $|DFBETAS_{ij}| > 2$ then the $i$th observation is influential. DFBETAS is similar to another well–known measure of influence, Cook’s Distance.

3) COVRATIO = \[\frac{\det(\hat{\sigma}^{-2}_0 (X'_{(0)}X_{(0)})^{-1})}{\det(\hat{\sigma}_0^2 (X'X)^{-1})}\]
   \[= \text{ratio of determinants of covariance matrices before and after deleting ith observation.}\]
   The $i$th observation is influential if its covratio lies outside the interval $[1 - 3p/n, 1 + 3p/n]$.  

43
Figure 7.3 Normal probability plots of observed effects

Q-Q plot of ML residuals
i.e. epsilon-hat

Q-Q plot of ML treatment effects
i.e. beta-hat

Q-Q plot of REML residuals

Q-Q plot of REML treatment effects
7.4.1 REGRESSION-TYPE DIAGNOSTICS ADAPTED FOR THE VARIANCE COMPONENTS MODEL

1) Hat matrix diagonals need to be redefined for the variance components model, using the fact that \( \hat{y} = Hy \).

Since \( y = X\alpha + Z\beta + \epsilon \) for the variance components model

\[
\hat{y} = X\hat{\alpha} + Z\hat{\beta}
\]

\[
= X(X'V^{-1}X)^{-1}X'V^{-1}y + ZDV^{-1}(y - X\hat{\alpha})
\]

\[
= X(X'V^{-1}X)^{-1}X'V^{-1}y + ZDV^{-1}(I - X(X'V^{-1}X)^{-1}X')y
\]

\[
= (X(X'V^{-1}X)^{-1}X'V^{-1} + ZDV^{-1})y
\]

Thus the new hat matrix \( H_1 = X(X'V^{-1}X)^{-1}X'V^{-1} + ZDV^{-1} \). Ideally \( H_1 \) should have the properties of \( H \), listed below.

a) is \( H_1 \) symmetric? No, because \( X(X'V^{-1}X)^{-1}X'V^{-1} \) is not symmetric.
b) is \( H_1 \) idempotent? No, because

\[
H_1^2 = [X(X'V^{-1}X)^{-1}X'V^{-1} + ZDV^{-1}][X(X'V^{-1}X)^{-1}X'V^{-1} + ZDV^{-1}] \neq H_1.
\]
c) \( [H_1]_{ii} \rightarrow 1 \Rightarrow \) ith observation is influential? This is still true because \( \hat{y} = H_1y \).
d) balanced data \( \Rightarrow [H_1]_{ii} \) all equal? This is still the case for the variance components model, and in fact for the one-way model the hat matrix diagonals are the same for any observations that come from groups with the same \( n_i \).
e) \( \text{Var}(\hat{y}) = kH_1 \)? This is not the case, because \( \text{Var}(\hat{y}) = \text{Var}(X\hat{\alpha} + Z\hat{\beta}) \)

\[
= \text{Var}(X(X'V^{-1}X)^{-1}X'V^{-1}y + ZDV^{-1}Py)
\]

\[
= (X(X'V^{-1}X)^{-1}X'V^{-1} + ZDV^{-1})V(X(X'V^{-1}X)^{-1}X'V^{-1} + ZDV^{-1})V^{-1}
\]

\[
= X(X'V^{-1}X)^{-1}X'V^{-1} + ZDV^{-1}PVDP = P
\]

\( \neq kH_1 \).

Note also that the variance matrix is symmetric by definition, but it is not idempotent.

There is in fact no reason why we should expect to retain the “nice” properties of the hat matrix under OLS, but at least the important diagnostic properties still hold. For the soybean data, the hat matrix diagonals do not tell us anything of interest because the experiment is balanced, and \( [H_1]_{ii} = 0.14 \) for all \( i \). Observations with high leverage would have \( [H_1]_{ii} > 2p/n = 12/25 = 0.48 \).

2) DFBETAS\(_{ij} \) = scaled change in jth parameter estimate on deleting ith observation.

Now the parameters of interest are \( e \) and \( b \), which have been referred to together as \( \theta \).

Thus DFTHETAS_i = \( \frac{\hat{\theta} - \hat{\theta}(i)}{\sqrt{\text{var}(\hat{\theta}(i))}} \),

DFTHETAS_i = \( \frac{\hat{b} - \hat{b}(i)}{\sqrt{\text{var}(\hat{b}(i))}} \)

where the variances are based on the variance matrices of Chapter 4 e.g. for ML.
\begin{align*}
\text{var}(\hat{\theta})_{(i)} &= F^{-1}_{(i)} = \frac{1}{\det F_{(i)}} \begin{pmatrix} 
\text{tr}[V^{-1}_{(i)}G_{(i)}V^{-1}_{(i)}] & -\text{tr}[V^{-1}_{(i)}G_{(i)}V^{-1}_{(i)}] \\
-\text{tr}[V^{-1}_{(i)}G_{(i)}V^{-1}_{(i)}] & \text{tr}[V^{-2}_{(i)}]
\end{pmatrix} \\
\text{where } F_{(i)} &= \begin{pmatrix} 
\text{tr}[V^{-2}_{(i)}] & \text{tr}[V^{-1}_{(i)}G_{(i)}V^{-1}_{(i)}] \\
\text{tr}[V^{-1}_{(i)}G_{(i)}V^{-1}_{(i)}] & \text{tr}[V^{-1}_{(i)}G_{(i)}V^{-1}_{(i)}G_{(i)}]
\end{pmatrix}.
\end{align*}

The DFTHETAS for the soybean data (Figure 7.4) indicate that removing extreme observations affects the diagnostic quantity the most. But the relationship of a particular observation to the others for that treatment also affects the size of DFTHETAS e.g. the 11 in treatment 2 that is so much larger than the other observations in that treatment produces the unusually large DFTHETAS, marked in Figure 7.4. However, given that DFTHETAS is a scaled quantity, none here are significantly large, so that no influential observations can be identified.

3) \text{COVRATIO} = \text{ratio of determinants of covariance matrices (see (4.4) and (4.10))}
\[
\frac{\det F^{-1}_{(i)}}{\det F^{-1}_{(i)}} \text{ (ML)} \quad \text{or} \quad \frac{\det F^{-1}_{1(i)}}{\det F^{-1}_{1}} \text{ (REML)}.
\]

As before, an “extreme” ratio indicates an influential observation.

The COVRATIOS for the soybean data (Figure 7.4) tell us that removing extreme (large or small) observations tends to decrease the new determinant, since covratio < 1 \(\Rightarrow\) \text{new variance matrix} < \text{full variance matrix}. Removing moderately-sized observations tends to increase the new determinant. All covratios for this data lie in the interval \([1 - 18/25, 1 + 18/25]\), indicating no especially influential observations.

\textbf{7.6 FURTHER ILLUSTRATION WITH SOYBEAN DATA}

The value of the ML and REML diagnostics are very similar, indicating that diagnostic tests are not going to offer a strong criterion for choosing one method over the other. Therefore the conclusion to be drawn from either set of diagnostics as a whole is that the data probably does not come from a \(N(\alpha, V)\) distribution, largely because of the treatment effects. The most suspect observation is the 4th data point in treatment 2, the 11. If this observation is deleted, the hat matrix diagonals are:
Figure 7.4 Regression diagnostics for soybean data

- Scaled change in MLE of $e$
- Scaled change in REMLE of $e$
- Scaled change in MLE of $b$
- Scaled change in REMLE of $b$
- Covariance ratio using ML
- Covariance ratio using REML
ML: 0.1563 (n_i = 5) and 0.1851 (n_i = 4)
REML: 0.1652 (n_i = 5) and 0.1977 (n_i = 4), none of which are significant.

The other diagnostic plots given in Figure 7.5 - Figure 7.7 confirm that the data set no longer contains serious outliers. There is still some curvature in the matrix of Q-Q plots (Figure 7.5), especially in the third treatment, but the worst curvature of Figure 7.2 has been removed. All four fitted effect plots (Figure 7.6) display reasonable normality. There are no influential observations identified in the deletion diagnostics (Figure 7.7), although there is a clustering effect visible in the DFTHETASE plots. This effect is caused mostly by the 4 in treatment 3, the observation second in line for removal as an outlier.

Finally in this brief analysis of the reduced data set, the new MLEs of e and b are 5.667 (variance 3.38) and 3.006 (variance 7.18) respectively; and the REMLEs are 5.666 (variance 3.38) and 4.060 (variance 13.91).

Concluding this chapter, it appears that the diagnostic tests proposed have indeed assisted in identifying observations that violate the assumptions of the variance components model. The removal of such observations should improve the quality of the new estimates.
Figure 7.5 Q-Q plot of all pairs of treatments: reduced soybean data
Figure 7.6 Normal probability plots of observed effects: reduced soybean data

Q-Q plot of ML residuals
i.e. epsilon-hat

Q-Q plot of REML residuals

Q-Q plot of ML treatment effects
i.e. beta-hat

Q-Q plot of REML residuals
Figure 7.7 Regression diagnostics for reduced soybean data

- Scaled change in MLE of $e$
- Scaled change in REMLE of $e$
- Scaled change in MLE of $b$
- Scaled change in REMLE of $b$
- Covariance ratio using ML
- Covariance ratio using REML
8. CONCLUSION

8.1 OPPORTUNITIES FOR FURTHER RESEARCH

This section identifies some of the topics in variance component estimation that require further investigation.

The first broad area is diagnostic tests. More work could be done on the behaviour of the tests listed in Chapter 7 in a wider variety of situations e.g. larger designs with more components of variance. Also, the diagnostics proposed by Beckman, Nachtsheim and Cook (1987) need further study. They use likelihood ratios to measure the effect of perturbing error variances, treatment variances or responses.

Secondly, now that some insight has been gained into the quantitative effect of non-normality on the size and variability of variance component estimates, attention may be turned to developing robust variance component estimators. It would also be helpful to know whether the observed conservativeness of estimates under non-normality is a global effect or simply a result of the small sample design and particular values of e and b that were considered. It would be good too to prove some results analytically about the large-sample variance of MLEs and REMLEs in the presence of non-normality.

Finally, it has been noted several times that the expression e + nb plays an important role in the estimation of e and b. Interesting insights might be gained by reparameterising the model (1.1) in terms of e and c = e + nb, and estimating e and c. Furthermore, this new parameter c could be used as a single measure of the performance of algorithms in a Monte Carlo study. However, care would have to be exercised when dealing with unbalanced data.

8.2 SUMMARY OF RESULTS

The main conclusions to be drawn from this thesis are

1) The explicit expressions for likelihood derivatives are relatively easy to write down and interpret for balanced one-way data, but much more difficult for unbalanced data.
2) Under normality, MLEs are to be preferred for small one-way designs on the basis of efficiency, but REMLEs are to be preferred on the basis of unbiasedness.
3) Asymptotically, these MLEs and REMLEs are both unbiased and efficient.
4) In the presence of non-normality in small experiments, REMLEs are unbiased for e* and b*, MLEs of e* are unbiased and MLEs of b* are biased.
5) In the presence of non-normality, the large-sample MSE of REMLEs approaches that of MLEs but this MSE is often smaller than the corresponding MSE in the absence of non-normality.
Searle ((1987) p.508) shows that the predictor of $\beta$ with minimum MSE is $E(\beta|y)$ i.e.

$$\hat{\beta} = E(\beta) + \text{Cov}(\beta, y)\text{Var}(y)^{-1}(y - E(y))$$

$$= 0 + \text{Cov}(\beta, \beta^TZ)^{V^{-1}}(y - X\hat{\alpha})$$

$$= DZ^TV^{-1}(y - X\hat{\alpha})$$

$$= DZ^TV^{-1}(y - X(X^TV^{-1}X)^{-1}X^TV^{-1}y)$$

$$= DZ^TV^{-1}X^TV^{-1}X^TV^{-1}y$$

$$= DZ^TV^{-1}(y - X(X^TV^{-1}X)^{-1}X^TV^{-1}y)$$

Harville (1977) shows there is another way to express $\hat{\beta}$, using the solutions to (A1.2):

$$\begin{pmatrix} X^TR^\dagger X & X^TR^\dagger Z \\ Z^TR^\dagger X & D^\dagger + Z^TR^\dagger Z \end{pmatrix} \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = \begin{pmatrix} X^TR^\dagger y \\ Z^TR^\dagger y \end{pmatrix}$$

(A1.2)

The matrix on the left hand side can be inverted using the following lemma from Rao(1973):

$$\begin{pmatrix} A & B \\ B^T & C \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} + FE^{-1}F^T & FE^{-1} \\ -E^{-1}F & E^{-1} \end{pmatrix}$$

where $E = C - B^TA^{-1}B$ and $F = A^{-1}B$.

Letting $A = X^TR^\dagger X$, $B = X^TR^\dagger Z$ and $C = D^\dagger + Z^TR^\dagger Z$ implies that $E = D^\dagger + Z^TR^\dagger Z - Z^TR^\dagger X(X^TR^\dagger X)^{-1}X^TR^\dagger Z = D^\dagger + Z^TSZ$, where $S = R^\dagger - R^\dagger X(X^TR^\dagger X)^{-1}X^TR^\dagger$ and $F = (X^TR^\dagger X)^{-1}X^TR^\dagger Z$.

Thus $K = \begin{pmatrix} X^TR^\dagger X & X^TR^\dagger Z \\ Z^TR^\dagger X & D^\dagger + Z^TR^\dagger Z \end{pmatrix}^{-1}$ has elements $[K]_{11}$, $[K]_{12} = [K]_{21}$ and $[K]_{22}$ as follows:

$$[K]_{11} = (X^TR^\dagger X)^{-1} + (X^TR^\dagger X)^{-1}X^TR^\dagger ZD(I + Z^TSZ)^{-1}Z^TR^\dagger X(X^TR^\dagger X)^{-1}$$

$$[K]_{12} = -D(I + Z^TSZ)^{-1}Z^TR^\dagger X(X^TR^\dagger X)^{-1}$$

$$[K]_{22} = D(I + Z^TSZ)^{-1}.$$ 

Expanding the second row of the solution to (A1.2) we see

$$\hat{\beta} = -D(I + Z^TSZ)^{-1}Z^TR^\dagger X(X^TR^\dagger X)^{-1}X^TR^\dagger y + D(I + Z^TSZ)^{-1}Z^TR^\dagger y$$

$$= -D(I + Z^TSZ)^{-1}Z^TR^\dagger X(X^TR^\dagger X)^{-1}X^TR^\dagger - R^\dagger y$$

$$= D(I + Z^TSZ)^{-1}Z^TSy.$$  

(A1.3)

This version of $\hat{\beta}$ will be used to show how Henderson estimates $b$. 

53
These derivatives can be found in Graybill (1983).

If $V$, a matrix, is a function of a vector of parameters $\theta$ then

\[
\frac{\partial}{\partial \theta_i} \log(|V|) = \text{tr}[V^{-1}(\partial V/\partial \theta_i)]
\]

\[
\frac{\partial}{\partial \theta_i} V^i = -V^i(\partial V/\partial \theta_i)V^i
\]

\[
\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log(|V|) = \text{tr}[-V^i(\partial V/\partial \theta_i)V^j(\partial V/\partial \theta_j) + V^i(\partial^2 V/\partial \theta_i \partial \theta_j)]
\]

\[
\frac{\partial^2}{\partial \theta_i \partial \theta_j} V^i = V^i(\partial V/\partial \theta_i)V^j(\partial V/\partial \theta_j)V^i - V^i(\partial^2 V/\partial \theta_i \partial \theta_j)V^i
\]

\[
+ V^i(\partial V/\partial \theta_j)V^j(\partial V/\partial \theta_i)V^i.
\]  \((A1.4)\)

Applying these derivatives to the likelihood (2.4) we obtain (as in Harville (1975)): 

\[
\frac{\partial L}{\partial \theta_i} = (-\frac{1}{2}) \text{tr}[V^{-1}(\partial^2 V/\partial \theta_i \partial \theta_j)] + \frac{1}{2} (y - X_0)^T V^{-1}(\partial V/\partial \theta_i)V^{-1}(y - X_0).
\]

The expected value, $E(\partial L/\partial \theta_i)$, is zero. Differentiating the likelihood a second time we obtain

\[
\frac{\partial^2 L}{\partial \theta_i \partial \theta_j} = (-\frac{1}{2}) \text{tr}[V^{-1}(\partial^2 V/\partial \theta_i \partial \theta_j)] - V^{-1}(\partial V/\partial \theta_i)V^{-1}(y - X_0)
\]

\[
+ \frac{1}{2} (y - X_0)^T V^{-1}(\partial^2 V/\partial \theta_i \partial \theta_j)V^{-1}(y - X_0)
\]

with expected value $E(\partial^2 L/\partial \theta_i \partial \theta_j) = (-1/2) \text{tr}[V^{-1}(\partial V/\partial \theta_i)V^{-1}(\partial V/\partial \theta_j)]$. \((A1.5)\)

Similarly for the restricted likelihood (2.7) the first derivative is 

\[
\frac{\partial L_1}{\partial \theta_i} = (-\frac{1}{2}) \text{tr}[P(\partial V/\partial \theta_i)] + \frac{1}{2} (y - X_0)^T V^{-1}(\partial V/\partial \theta_i)V^{-1}(y - X_0)
\]

with expected value zero. The second derivative is

\[
\frac{\partial^2 L_1}{\partial \theta_i \partial \theta_j} = (-\frac{1}{2}) \text{tr}[P(\partial^2 V/\partial \theta_i \partial \theta_j) - P(\partial V/\partial \theta_j)P(\partial V/\partial \theta_i)]
\]

\[
+ \frac{1}{2} (y - X_0)^T V^{-1}(\partial^2 V/\partial \theta_i \partial \theta_j)V^{-1}(y - X_0)
\]

with expected value $E(\partial^2 L_1/\partial \theta_i \partial \theta_j) = (-1/2) \text{tr}[P(\partial V/\partial \theta_i)P(\partial V/\partial \theta_j)]$. \((A1.6)\)

The vector $\theta$ is the vector $(\sigma^2_e, \sigma^2_1, \ldots, \sigma^2_p)$ in the variance components model.

Four useful results for obtaining the expected values in (A1.5), (A1.6) and also (6.4) – (6.15) are:

1) $E(y - X_0)(y - X_0)^T = V$

2) $E(y - X_0)^T W(y - X_0) = \text{tr}[W E(y - X_0)(y - X_0)^T] = \text{tr}[WV]$

3) $E(y - X_0)^T W(y - X_0)(y - X_0)^T Q(y - X_0) = \text{tr}[WVQV] + 2\text{tr}[WVQV]$


4) If $E(y - X_0)(y - X_0)^T = V$

$E(y - X_0)(y - X_0)^T = E(y - X_0 + X_0 - X_0)(y - X_0 + X_0 - X_0)^T$

$= V - 2X(X^TV^{-1}X)^TX^T + X(X^TV^{-1}X)^TX^T$

$= V - X(X^TV^{-1}X)^TX^T$.  

54
A.1.3 Lemmas Involving Determinants

These lemmas are required to manipulate the likelihood into the forms used by Henderson in his algorithms for ML and REML estimation of variance components.

Proof. The proof relies on multiplying a matrix with the required determinant by another matrix with a determinant of one.

\[
\begin{pmatrix}
I & ZD \\
0 & I
\end{pmatrix}
\begin{pmatrix}
R & -ZD \\
Z^t & I
\end{pmatrix}
= \begin{pmatrix}
R + ZDZ^t & 0 \\
Z^t & I
\end{pmatrix}
= \begin{pmatrix}
V & 0 \\
0 & I
\end{pmatrix}
\]

so that $\begin{pmatrix}
R & -ZD \\
Z^t & I
\end{pmatrix}$ = $|R| |I + Z^t R^{-1}ZD|$. Thus $|V| = |R| |I + Z^t R^{-1}ZD|$ as required.

The matrix $(I + Z^t R^{-1}ZD)$ for the one–way model is

\[
\begin{pmatrix}
e + n_1 b \\
e
\end{pmatrix}
\]

Lemma A1.2. $V^t = R^t - R^t ZD(I + Z^t R^{-1}ZD)^{-1} Z^t R^{-1}$.
Proof. Multiply by $V = R + ZDZ^t$ to get $I$.

$(R^t - R^t ZD(I + Z^t R^{-1}ZD)^{-1} Z^t R^{-1})(R + ZDZ^t)
= I - R^t ZD(I + Z^t R^{-1}ZD)^{-1} Z^t R^{-1} - R^t ZD(I + Z^t R^{-1}ZD)^{-1} Z^t R^{-1} ZDZ^t
= I - R^t ZD(I + Z^t R^{-1}ZD)^{-1} - I + (I + Z^t R^{-1}ZD)^{-1} Z^t R^{-1} ZD Z^t
= I - R^t ZD((I + Z^t R^{-1}ZD)^{-1}(I + Z^t R^{-1}ZD) - I) Z^t
= I.$

Lemma A1.3. $|V| |X^t V^{-1}X| = |R| |B| = |R| |X^t R^{-1}X| |I + Z^t SZD|$, where

\[
B = \begin{pmatrix}
X^t R^{-1}X \\
I + Z^t R^{-1}ZD
\end{pmatrix}
\]

Proof. If $B = \begin{pmatrix}
B_{11} & B_{12} \\
B_{21} & B_{22}
\end{pmatrix}$, then $|B| = |B_{11}| |B_{22} - B_{21} B_{11}^{-1} B_{12}|$ and also

$|B| = |B_{22}| |B_{11} - B_{21} B_{11}^{-1} B_{12}|$ (Graybill (1969) p.184). Thus

$|B| = |X^t R^{-1}X| |I + Z^t R^{-1}ZD - Z^t R^{-1}X(X^t R^{-1}X)^{-1} X^t R^{-1}ZD|
= |X^t R^{-1}X| |I + Z^t R^{-1} - R^{-1}X(X^t R^{-1}X)^{-1} X^t R^{-1}ZD|
= |X^t R^{-1}X| |I + Z^t SZD|
and also
\[ |B| = |I + Z^T R^{-2} ZD| |X^T R^{-2} X - X^T R^{-2} ZD (I + Z^T R^{-2} ZD)^{-1} Z^T R^{-2} X| \]
\[ = |I + Z^T R^{-2} ZD| |X^T (R^{-4} - R^{-2} ZD (I + Z^T R^{-2} ZD)^{-1} Z^T R^{-2}) X| \]
\[ = |I + Z^T R^{-2} ZD| |X^T V^{-2} X| \]
by Lemma A1.2.

By Lemma A1.1 |V| = |R| |I + Z^T R^{-2} ZD|, so that

\[ |V| |X^T V^{-2} X| = |R| |I + Z^T R^{-2} ZD| |X^T V^{-2} X| \]
\[ = |R| |B| \]
\[ = |R| |X^T R^{-2} X| |I + Z^T S Z D| \]
as required.

A1.4 ALTERNATIVE EXPRESSIONS FOR THE LIKELIHOOD

Using the results of the previous section, alternative forms of L and L_1 (as used by
Henderson) are now calculated.

Lemma A1.4. \( V^{-1}(y - X\tilde{\alpha}) = R^{-1}(y - X\tilde{\alpha} - Z\tilde{\beta}) = S(y - Z\tilde{\beta}) \)

Proof. \( V^{-1}(y - X\tilde{\alpha}) = V^{-1}(y - X(X^T V^{-1} X)^{-1} X^T V^{-1} y) \]
\[ = (V^{-1} - V^{-1} X(X^T V^{-1} X)^{-1} X^T V^{-1}) y \]
\[ = Py. \]

\( R^{-1}(y - X\tilde{\alpha} - Z\tilde{\beta}) = R^{-1}(y - X(X^T V^{-1} X)^{-1} X^T V^{-1} y - Z D Z^{-1} Py) \]
\[ = R^{-1}(I - X(X^T V^{-1} X)^{-1} X^T V^{-1} - Z D Z^{-1} P) y \]
\[ = R^{-1}(V - Z D^{-1} P) y \]
\[ = Py \text{ also.} \]

Thus \( V^{-1}(y - X\tilde{\alpha}) = R^{-1}(y - X\tilde{\alpha} - Z\tilde{\beta}) \). Finally

\( S(y - Z\tilde{\beta}) = R^{-1}(y - Z\tilde{\beta}) - R^{-1} X(X^T R^{-1} X)^{-1} X^T R^{-1}(y - Z D Z^{-1} P) y \)
\[ = R^{-1}(y - Z\tilde{\beta}) \]
\[ - R^{-1} X(X^T R^{-1} X)^{-1} X^T R^{-1}(I - Z D Z^{-1} (V^{-4} - V^{-1} X(X^T V^{-1} X)^{-1} X^T V^{-1})) y \]
\[ = R^{-1}(y - Z\tilde{\beta}) - R^{-1} X(X^T R^{-1} X)^{-1} X^T R^{-1} y + R^{-1} X(X^T R^{-1} X)^{-1} X^T R^{-1} Z D Z^{-1} V^{-1} y \]
\[ + R^{-1} X(X^T R^{-1} X)^{-1} X^T R^{-1} Z D Z^{-1} V^{-1} (X^T V^{-1} X)^{-1} X^T V^{-1} y \]
\[ - R^{-1} X(X^T R^{-1} X)^{-1} X^T R^{-1} Z D Z^{-1} V^{-1} (X^T V^{-1} X)^{-1} X^T V^{-1} y \]
\[ = R^{-1}(y - Z\tilde{\beta}) - R^{-1} X(X^T R^{-1} X)^{-1} X^T R^{-1} y + R^{-1} X(X^T R^{-1} X)^{-1} X^T R^{-1} (V - R) V^{-1} y \]
\[ + R^{-1} X(X^T R^{-1} X)^{-1} X^T R^{-1} (V - R) V^{-1} y \]
\[ - R^{-1} X(X^T R^{-1} X)^{-1} X^T R^{-1} (V - R) V^{-1} (X^T V^{-1} X)^{-1} X^T V^{-1} y \]
\[ = R^{-1}(y - Z\tilde{\beta}) - R^{-1} X(X^T R^{-1} X)^{-1} X^T R^{-1} y + R^{-1} X(X^T R^{-1} X)^{-1} X^T R^{-1} y \]
\[ - R^{-1} X(X^T R^{-1} X)^{-1} X^T V^{-1} y \]
\[ - R^{-1} X(X^T V^{-1} X)^{-1} X^T V^{-1} y + R^{-1} X(X^T R^{-1} X)^{-1} X^T V^{-1} y \]
\[ = R^{-1}(y - X\tilde{\alpha} - Z\tilde{\beta}) \text{ which completes the proof of the lemma.} \]
Lemma A1.5. \( L_1 = (-1/2) \log |V| - (1/2) \log |X^T V^{-1} X| \)

\[ = (-1/2) \log |R| - (1/2) \log |X^T R^{-1} X| - (1/2) \log |I + Z^T S Z D| \]

\[ - (1/2) y^T S(y - Z \hat{\beta}). \]

This version of the likelihood is used by Henderson to estimate \( b \).

Proof. Apply Lemma A1.3 to the first two terms. For the last term, apply Lemma A1.4 and note that

\[ (y - X\alpha)^T V^{-1} (y - X\alpha) = (y - X\alpha)^T S(y - Z \hat{\beta}) \]

\[ = y^T S(y - Z \hat{\beta}) - \alpha^T X^T S(y - Z \hat{\beta}) \]

\[ = y^T S(y - Z \hat{\beta}) - \alpha^T (X^T R^{-1} - X^T R^{-1} X) X R^{-1} (y - Z \hat{\beta}) \]

Similarly, \( L = (-1/2) \log |V| - (1/2) (y - X\alpha)^T V^{-1} (y - X\alpha) \)

\[ = (-1/2) \log |R| - (1/2) \log |I + Z^T S D| \]

after substituting \( \alpha \) for \( \alpha \) in the likelihood.

Lemma A1.6. If \( \theta_i \) is any variance component apart from \( \sigma^2_e \),

\[ \partial L_1 / \partial \theta_i = (-1/2) \text{tr} [(I + Z^T S D)^{-1} Z^T S (\partial D / \partial \theta_i) + (1/2) \hat{\beta}^T D^{-1} (\partial D / \partial \theta_i) D^{-1} \hat{\beta}]. \]

Proof. This lemma is simply a manipulation of the derivative of the likelihood obtained in Lemma A1.5. From that lemma

\[ L_1 = (-1/2) \log |R| - (1/2) \log |X^T R^{-1} X| - (1/2) \log |I + Z^T S D| \]

\[ - (1/2) y^T S(y - Z \hat{\beta}) \]

\[ = (-1/2) \log |R| - (1/2) \log |X^T R^{-1} X| - (1/2) \log |I + Z^T S D| \]

\[ - (1/2) y^T S y + (1/2) y^T S Z D \hat{\beta} \]

\[ = (-1/2) \log |R| - (1/2) \log |X^T R^{-1} X| - (1/2) \log |I + Z^T S D| \]

\[ - (1/2) y^T S y + (1/2) y^T S Z D \hat{\beta} \]

after substituting \( \alpha \) for \( \alpha \) in the likelihood.

\[ \partial L_1 / \partial \theta_i = (-1/2) \text{tr} [(I + Z^T S D)^{-1} Z^T S (\partial D / \partial \theta_i)] \]

\[ + \text{tr} [(I + Z^T S D)^{-1} Z^T S D (\partial D / \partial \theta_i) (I + Z^T S D)^{-1} Z^T S] \]

\[ = (-1/2) \text{tr} [(I + Z^T S D)^{-1} Z^T S D (I + Z^T S D)^{-1} Z^T S] \]

\[ + (1/2) \text{tr} [(I + Z^T S D)^{-1} (I + Z^T S D - I) D^{-1} (\partial D / \partial \theta_i)] \]

\[ + (1/2) \text{tr} [(I + Z^T S D)^{-1} (I + Z^T S D - I) D^{-1} (\partial D / \partial \theta_i)] \]

\[ = (-1/2) \text{tr} [(I + Z^T S D)^{-1} D^{-1} (\partial D / \partial \theta_i)] \]

\[ + (1/2) \text{tr} [(I + Z^T S D)^{-1} (\partial D / \partial \theta_i) + (1/2) \hat{\beta}^T D^{-1} (\partial D / \partial \theta_i) D^{-1} \hat{\beta}]. \]

Similarly, \( L = (-1/2) \log |V| - (1/2) (y - X\alpha)^T V^{-1} (y - X\alpha) \) after substituting \( \alpha \) for \( \alpha \)

\[ = (-1/2) \log |R| - (1/2) \log |(I + Z^T R^{-1} Z D)^{-1}| - (1/2) y^T (y - S \hat{\beta}) \]
using Lemma 1.1 to rewrite $|V|$ so that
\[\frac{\partial L}{\partial \theta_i} = (-1/2) \text{tr}(I + Z^T R^{-1} Z D)^{-1} (Z^T R^{-1} Z \partial D / \partial \theta_i) + (1/2) \beta^T D^{-1} (\partial D / \partial \theta_i) D^{-1} \beta\]
\[= (-1/2) \text{tr}(I + Z^T R^{-1} Z D)^{-1} (Z^T R^{-1} Z D + I - I) D^{-1} (\partial D / \partial \theta_i)\]
\[+ (1/2) \beta^T D^{-1} (\partial D / \partial \theta_i) D^{-1} \beta\]
\[= (-1/2) \text{tr}(I - (I + Z^T R^{-1} Z D)^{-1}) D^{-1} (\partial D / \partial \theta_i)\]
\[+ (1/2) \beta^T D^{-1} (\partial D / \partial \theta_i) D^{-1} \beta.\]

**Lemma A1.7.** The MLE of $e$, $\hat{e} = y^T (y - X\alpha - Z\beta)/n$.

**Proof.** The likelihood is $L = (-1/2) \log|V| - (1/2) (y - X\alpha)^T V^{-1} (y - X\alpha)$.

Letting $V = eH$, where $H$ is independent of $e$, the likelihood becomes
$L = (-n/2) \log|e| - (1/2) \log|H| - (1/2e) (y - X\alpha)^T H^{-1} (y - X\alpha)$ with derivative
\[\frac{\partial L}{\partial e} = (-n/2e) + (y - X\alpha)^T H^{-1} (y - X\alpha)/2e^2 = 0\] at a stationary point
\[\Rightarrow -ne + (y - X\alpha)^T H^{-1} (y - X\alpha) = 0\]
\[\Rightarrow \hat{e} = \frac{(y - X\alpha)^T H^{-1} (y - X\alpha)}{2e^2}\]
\[= \frac{ny^T S(y - Z\beta)}{n}\]
\[= \frac{ey^T R^{-1} (y - X\alpha - Z\beta)}{n}\]
\[= \frac{y^T (y - X\alpha - Z\beta)}{n}\]
by Lemmas A1.4 and A1.5, after replacing $\alpha$ by $\hat{\alpha}$

by Lemma A1.4

since $R = eI_n$.

**Lemma A1.8.** The REMLE of $e$ is $\hat{e} = y^T (y - X\alpha - Z\beta)/(n - 1)$.

**Proof.** The likelihood is
$L_1 = (-1/2) \log|V| - (1/2) \log|X^T V^{-1} X| - (1/2) (y - X\alpha)^T V^{-1} (y - X\alpha)$.

Setting $\frac{\partial L_1}{\partial e} = 0$ we obtain
\[\text{tr}[P] = (y - X\hat{\alpha})^T V^{-2} (y - X\hat{\alpha})\]
\[\text{tr}[P] = y^T P^2 y\]
\[\text{tr}[PZDZ^T] = y^T PZDZ^T y\]
\[\text{tr}[P(V - R)] = y^T P(V - R) y\]
\[\text{tr}[PV] - e \text{tr}[P] = y^T PV y - e y^T P^2 y\]
Now $\text{tr}[PV] = \text{rank}[PV] = n - 1$ and using the fact that $\text{tr}[P] = y^T P^2 y$, we get
\[n - 1 = y^T PV y\]
\[= (y - X\hat{\alpha})^T V^{-1} (y - X\hat{\alpha}).\]
Therefore
\[
\hat{e} = \frac{(y - X\hat{\alpha})'H'(y - X\hat{\alpha})}{n-1} = \frac{(y - X\hat{\alpha})'e\hat{V}'(y - X\hat{\alpha})}{n-1} = \frac{ey'\bar{S}(y - \bar{Z}\beta)}{n-1} = \frac{ey'\bar{R}^{-1}(y - X\hat{\alpha} - Z\bar{\beta})}{n-1} = \frac{y'(y - X\hat{\alpha} - Z\bar{\beta})}{n-1}
\]
by Lemmas A.1.4 and A.1.5
by Lemma A.1.4
since R = eI_n.

A1.5 INVERSION OF (I + Z'SZD)

For the one–way unbalanced model being considered here:
\[
S = R^{-1} - R^{-1}X(X'^{-1}R^{-1}X)^{-1}X'R^{-1} = \frac{1}{e} \begin{pmatrix} 1 - (1/n) & -1/n \\ -1/n & 1 - (1/n) \end{pmatrix}
\]
so that
\[
I + Z'SZD = \begin{pmatrix} ne+n_1nb-n_1^2b & -n_1njb \\ -n_1njb & ne+n_pnb-n_pb^2 \end{pmatrix}
\]

Graybill ((1969) p.189) proves that if a p × p matrix K = C + ors', where C is diagonal, r and s are p×1 vectors and \( \alpha \neq -\left(\sum_{i=1}^{p} r_i s_i / c_{ii}\right)^{-1} \),
then \( K^{-1} = C^{-1} + \gamma uv' \), where \( u_i = r_i / c_{ii}, v_i = s_i / c_{ii} \) and \( \gamma = -\alpha \left(1 + \alpha \sum_{i=1}^{p} r_i s_i / c_{ii}\right)^{-1} \).

Letting \( c_{ii} = ne + n_i nb, \alpha = -1, r_i = n_i \) and \( s_i = n_i b \), \( (I + Z'SZD)^{-1} \) is
\[
\frac{\left(\frac{n(e+n_1b)(1-\Sigma n_i^2b/n(e+n_1b))+n_1^2b}{n(e+n_1b)^2(n-\Sigma n_i^2b/e+n_1b)}\frac{n_1njb}{n(e+n_1b)(e+n_1b)(n-\Sigma n_i^2b/e+n_1b)}\right)}{\frac{n(e+n_1b)(e+n_1b)(n-\Sigma n_i^2b/e+n_1b)}{n(e+n_1b)(e+n_1b)(n-\Sigma n_i^2b/e+n_1b)}}
\]
\[
\frac{\left(\frac{n(e+n_p^2b/n(e+n_p^2b)+n_p^2b}{n(e+n_p^2b)^2(n-\Sigma n_i^2b/e+n_p^2b)}\frac{n_1njb}{n(e+n_p^2b)(e+n_p^2b)(n-\Sigma n_i^2b/e+n_p^2b)}\right)}{\frac{n(e+n_p^2b)(e+n_p^2b)(n-\Sigma n_i^2b/e+n_p^2b)}{n(e+n_p^2b)(e+n_p^2b)(n-\Sigma n_i^2b/e+n_p^2b)}}
\]

59
A2. SPLUS FUNCTIONS

A2.1 NUMERICAL PROCEDURES FOR VARIANCE COMPONENT ESTIMATION

\[\text{tr} \leftarrow \text{function}(m) \ # \ m \text{ is a square matrix} \quad \text{(A2.1)}\]
\[
\begin{aligned}
\text{V} & \leftarrow \text{function}(Z, e, b) \ # \ V = \text{Var}(y) \text{ where } y, Z, e \text{ and } b \text{ as in §2.1.} \quad \text{(A2.2)} \\
& \quad \{ nn \leftarrow \text{sum}(Z) \\
& \quad \quad E \leftarrow \text{diag}(e, nn) \\
& \quad \quad G \leftarrow (Z \ %*% t(Z)) * b \\
& \quad \quad E + G \}
\end{aligned}
\]

\[\text{T11} \leftarrow \text{function}(Z, e, b) \quad \text{(A2.3)}\]
\[
\begin{aligned}
\text{terms.in.y} & \leftarrow \text{function}(Z, y, e = 1, b = 1) \ # \ y \text{ as in (2.1)} \quad \text{(A2.4)} \\
& \quad \{ \text{expt} \leftarrow y * Z \\
& \quad \quad \text{tot} \leftarrow \text{apply(expt, 2, sum)} \\
& \quad \quad n \leftarrow \text{apply}(Z, 2, \text{sum}) \\
& \quad \quad ybar \leftarrow \text{tot}/n \\
& \quad \quad k \leftarrow e + n * b \\
& \quad \quad \text{alphahat} \leftarrow \text{rep}(\text{sum(tot}/k)/\text{sum}(n/k), \text{length}(y)) \\
& \quad \quad G \leftarrow Z \ %*% t(Z) \\
& \quad \quad \text{vary} \leftarrow \text{V}(Z, e, b) \\
& \quad \quad \text{varyinv} \leftarrow \text{solve}(\text{vary}) \\
& \quad \quad \text{dif} \leftarrow \text{as.matrix}(y - \text{alphahat}) \\
& \quad \quad \text{c(t(dif)} \ %*% \text{varyinv} \ %*% \text{varyinv} \ %*% \text{dif, t(dif)} \ %*% \text{varyinv} \ %*% \text{G} \\
& \quad \quad \quad \ %*% \text{varyinv} \ %*% \text{dif})\}
\end{aligned}
\]

\[\text{anderson.ml} \leftarrow \text{function}(Z, y, e = 1, b = 1) \quad \text{(A2.5)}\]
\[
\begin{aligned}
\text{repeat} \{ \text{vary} \leftarrow \text{V}(Z, e, b) \\
& \quad \text{varyinv} \leftarrow \text{solve}(\text{vary}) \\
& \quad G \leftarrow Z \ %*% t(Z) \\
& \quad \text{st1} \leftarrow \text{tr(varyinv} \ %*% \text{varyinv}) \}
\end{aligned}
\]
\begin{align*}
st_2 & \leftarrow \text{tr}(\text{varyinv} \times G \times \text{varyinv}) \\
st_3 & \leftarrow \text{tr}(\text{varyinv} \times G \times \text{varyinv} \times G) \\
\text{mat} & \leftarrow \text{matrix}(\text{c}(st_1, st_2, st_2, st_3), \text{ncol} = 2) \\
\text{rhs} & \leftarrow \text{terms.in.y}(Z, y, e, b) \\
nxt & \leftarrow \text{solve(mat, rhs)} \\
\text{if}(\text{abs}(\text{nxt} - (e, b)) < \text{rep}(0.001, 2)) \text{ break} \\
e & \leftarrow \text{nxt}[1] \\
b & \leftarrow \text{nxt}[2]
\end{align*}

\text{anderson.reml} \leftarrow \text{function}(Z, y, e = 1, b = 1) 
\{ 
\text{repeat} \{ 
\text{vary} \leftarrow \text{V}(Z, e, b) \\
\text{varyinv} \leftarrow \text{solve(vary)} \\
J \leftarrow \text{matrix}(1, \text{sum}(Z), \text{sum}(Z)) \\
P \leftarrow \text{varyinv} - \frac{(\text{varyinv} \times J \times \text{varyinv})}{\text{sum}(\text{varyinv})} \\
G \leftarrow Z \times G \\
st_1 \leftarrow \text{tr}(P \times P) \\
st_2 \leftarrow \text{tr}(P \times G \times P) \\
st_3 \leftarrow \text{tr}(P \times G \times P \times G) \\
\text{mat} \leftarrow \text{matrix}(\text{c}(st_1, st_2, st_2, st_3), \text{ncol} = 2) \\
\text{rhs} \leftarrow \text{terms.in.y}(Z, y, e, b) \\
nxt \leftarrow \text{solve(mat, rhs)} \\
\text{if}(\text{abs}(\text{nxt} - (e, b)) < \text{rep}(0.001, 2)) \text{ break} \\
e & \leftarrow \text{nxt}[1] \\
b & \leftarrow \text{nxt}[2]
\}

\text{henderson.ml} \leftarrow \text{function}(Z, y, e = 1, b = 1) 
\{ 
\text{expt} \leftarrow y \times Z \\
\text{tot} \leftarrow \text{apply(expt, 2, sum)} \\
\text{n} \leftarrow \text{apply}(Z, 2, \text{sum}) \\
\text{ybar} \leftarrow \text{tot} / \text{n} \\
\text{repeat} \{k \leftarrow e + n \times b \\
\text{alphahat} \leftarrow (\text{sum}(\text{tot} / k)) / (\text{sum}(n / k)) \\
\text{alphahat1} \leftarrow \text{rep(}\text{alphahat}, \text{ncol}(Z)) \\
\text{alphahat2} \leftarrow \text{rep(}\text{alphahat}, \text{length}(y)) \\
\text{betahat} \leftarrow (n \times b \times (\text{ybar} - \text{alphahat1})) / k \\
\text{Zbetahat} \leftarrow \text{rep(}\text{betahat}, n) \\
\text{nxtb} \leftarrow (\text{sum(}\text{betahat} \times 2)) / (\text{ncol}(Z) - e \times \text{sum}(1 / k))
\}

\text{(A2.6)} \\
\text{(A2.7)}
\[
\text{nxt}\text{e} \leftarrow \frac{\sum (y \times (y - \text{alphahat}^2 - Z\text{betahat}))}{\text{length}(y)}
\]

\[
\text{if(abs(c(nxt}\text{e}, \text{nxtb}) - c(e, b)) < \text{rep}(0.001, 2)) \text{ break}
\]

\[
e \leftarrow \text{nxt}\text{e}
\]

\[
b \leftarrow \text{nxtb}
\]

\[
c(n\text{xte}, \text{nxtb})
\]

henderson.reml \leftarrow \text{function}(Z, y, e = 1, b = 1)

\[
\begin{align*}
\text{expt} & \leftarrow y \times Z \\
n\text{tot} & \leftarrow \text{apply(expt, 2, sum)} \\
n & \leftarrow \text{apply}(Z, 2, \text{sum)} \\
y\text{bar} & \leftarrow \text{tot}/n \\
\text{repeat}\{k \leftarrow e + n \times b \\
\text{alphahat} \leftarrow \frac{\sum (\text{tot}/k)}{\sum (n/k)} \\
\text{alphahat}^1 & \leftarrow \text{rep(alphahat, ncol(Z))} \\
\text{alphahat}^2 & \leftarrow \text{rep(alphahat, length(y))} \\
\text{betahat} & \leftarrow \frac{(n \times b \times (y\text{bar} - \text{alphahat}^1))/k}{k} \\
Z\text{betahat} & \leftarrow \text{rep(betahat, n)} \\
\text{term} & \leftarrow \text{T11}(n, e, b) \\
\text{nxtb} & \leftarrow \frac{\sum (\text{betahat}^2)/\text{ncol}(Z) - \text{tr(solve(term))}}{\text{ncol}(Z) - \text{tr(solve(term))}} \\
\text{nxt}\text{e} & \leftarrow \frac{\sum (y \times (y - \text{alphahat}^2 - Z\text{betahat}))/(\text{length}(y) - 1)}{\text{length}(y) - 1} \\
\text{if(abs(c(nxt}\text{e}, \text{nxtb}) - c(e, b)) < \text{rep}(0.001, 2))} \\
\text{break}
\end{align*}
\]

\[
e \leftarrow \text{nxt}\text{e}
\]

\[
b \leftarrow \text{nxtb}
\]

\[
c(n\text{xte}, \text{nxtb})
\]

\section*{A2.2 MEASURES OF EFFICIENCY}

ml.var \leftarrow \text{function}(Z, e, b)

\[
\begin{align*}
\text{vary} & \leftarrow V(Z, e, b) \\
\text{varyinv} & \leftarrow \text{solve(vary)} \\
\text{G} & \leftarrow Z \%\% \text{t}(Z) \\
\text{st1} & \leftarrow \text{tr(varyinv} \%\% \text{varyinv)} \\
\text{st2} & \leftarrow \text{tr(varyinv} \%\% G \%\% \text{varyinv)} \\
\text{st3} & \leftarrow \text{tr(varyinv} \%\% G \%\% \text{varyinv} \%\% G) \\
\text{info} & \leftarrow .5 \times \text{matrix(c(st1, st2, st2, st3), ncol = 2)} \\
\text{solve(info)}
\end{align*}
\]
reml.var <- function(Z, e, b)  
{  
vary <- V(Z, e, b)  
varyinv <- solve(vary)  
J <- matrix(1, sum(Z), sum(Z))  
P <- varyinv - (varyinv %*% J %*% varyinv)/(sum(varyinv))  
G <- Z %*% t(Z)  
st1 <- tr(P %*% P)  
st2 <- tr(P %*% G %*% P)  
st3 <- tr(P %*% G %*% P %*% G)  
info <- .5 * matrix(c(st1, st2, st2, st3), ncol = 2)  
solve(info)}

A2.3 MONTE CARLO STUDY

exp.ml <- function(Z, e, b)  
{  
vary <- V(Z, e, b)  
varyinv <- solve(vary)  
J <- matrix(1, sum(Z), sum(Z))  
P <- varyinv - (varyinv %*% J %*% varyinv)/(sum(varyinv))  
st1 <- tr(varyinv %*% varyinv)  
st2 <- tr(varyinv %*% G %*% varyinv)  
st3 <- tr(varyinv %*% G %*% varyinv %*% G)  
info <- matrix(c(st1, st2, st2, st3), ncol = 2)  
var <- solve(info)  
rhs <- as.matrix(c(tr(P), tr(P %*% G)))  
var %*% rhs}
random.y <- function(n) # n is the number of trials required  
{  y <- matrix(nrow=n, ncol = 25)
  for(i in 1:n)
    y[i, ] <- rnorm(25, 0, 2) + rep(rnorm(5, 0, 4), c(5, 5, 5, 5, 5))
  y}

random.yy <- function(n) 
{  y <- matrix(nrow=n, ncol = 24)
  for(i in 1:n)
    y[i, ] <- rnorm(25, 0, 2) + rep(rnorm(5, 0, 4), c(5, 4, 5, 5, 5))
  y}

simul <- function(Z, out, d, n) 
{  # d is a function e.g. one of (A2.5) – (A2.8).
  # out is the output from (A2.13) or (A2.14)
  save <- matrix(ncol = 2, nrow = n)
  for(i in 1:n)
    save[i, ] <- d(Z, out[i, ])
  save}

A2.4 NON–NORMALITY

nn.ml.var <- function(Z, e = 1, b = 1, estar = 1, bstar = 1) 
{  # estar = e*, bstar = b* (see (6.2))
  vary <- V(Z, e, b)
  varyinv <- solve(vary)
  vstar <- V(Z, estar, bstar)
  J <- matrix(1, sum(Z), sum(Z))
  varyinv2 <- varyinv %*% varyinv
  G <- Z %*% t(Z)
  st1 <- 0.5 * tr(varyinv2) - tr(varyinv2 %*% varyinv %*% vstar)
  st2 <- 0.5 * tr(varyinv2 %*% G) - tr(varyinv %*% G %*% varyinv2 %*% vstar)
  st3 <- 0.5 * tr(varyinv2 %*% G %*% G) - tr(varyinv2 %*% G %*% G %*% varyinv %*% vstar)
  psidash <- matrix(c(st1, st2, st3), ncol = 2)
  st4 <- (tr(varyinv)^2 - 2 * tr(varyinv) * tr(varyinv2 %*% vstar)
          + (tr(varyinv2 %*% vstar))^2 + 2 * tr(varyinv2 %*% varyinv2 %*% varyinv2 %*% vstar)
\[ v_{\text{star}} \]

\[ st5 \leftarrow \text{tr}(\text{varyinv}) \times \text{tr}(\text{varyinv} \times \text{G}) - \text{tr}(\text{varyinv}) \times \text{tr}(\text{varyinv}^2 \times \text{G}) \times \text{tr}(\text{varyinv}) \times \text{tr}(\text{vstar}) + \text{tr}(\text{varyinv}^2 \times \text{vstar}) \times \text{tr}(\text{varyinv}^2 \times \text{G}) \times \text{vstar} + 2 \times \text{tr}(\text{varyinv}^2 \times \text{vstar}) \times \text{tr}(\text{varyinv}^2 \times \text{G}) \times \text{vstar} \]

\[ st6 \leftarrow (\text{tr}(\text{varyinv} \times \text{G}))^2 - 2 \times \text{tr}(\text{varyinv} \times \text{G}) \times \text{tr}(\text{varyinv}^2 \times \text{G}) \times \text{vstar} + (\text{tr}(\text{varyinv}^2 \times \text{G}) \times \text{vstar})^2 + 2 \times \text{tr}(\text{varyinv}^2 \times \text{vstar}) \times \text{tr}(\text{varyinv}^2 \times \text{G}) \times \text{vstar} \]

\[ \text{psipsit} \leftarrow 0.25 \times \text{matrix}(c(st4, st5, st5, st6), ncol = 2) \]

\[ \text{solve}(\text{psidash}) \times \text{psipsit} \times \text{solve}(\text{psidash}) \]

\[ \text{nn.reml.var} \leftarrow \text{function}(Z, e = 1, b = 1, e\_\text{star} = 1, b\_\text{star} = 1) \]

\[ \{ \]
\[ \text{vary} \leftarrow V(Z, e, b) \]
\[ \text{varyinv} \leftarrow \text{solve}(\text{vary}) \]
\[ \text{vstar} \leftarrow V(Z, e\_\text{star}, b\_\text{star}) \]
\[ J \leftarrow \text{matrix}(1, \text{sum}(Z), \text{sum}(Z)) \]
\[ \text{vstarstar} \leftarrow \text{vstar} - J / \text{sum}(\text{solve}(\text{vstar})) \]
\[ \text{varyinv}^2 \leftarrow \text{varyinv} \times \text{varyinv} \]
\[ P \leftarrow \text{varyinv} - (\text{varyinv} \times J \times \text{varyinv}) / (\text{sum}(\text{varyinv})) \]
\[ G \leftarrow Z \times J \times \text{inv}(G) \]
\[ st1 \leftarrow 0.5 \times \text{tr}(P \times P) - \text{tr}(\text{varyinv}^2 \times P \times \text{vstarstar}) \]
\[ st2 \leftarrow 0.5 \times \text{tr}(P \times G \times P) - \text{tr}(\text{varyinv}^2 \times G \times \text{vstarstar}) \]
\[ st3 \leftarrow 0.5 \times \text{tr}(P \times G \times P \times G) - \text{tr}(\text{varyinv}^2 \times G \times P \times P) \]

\[ \text{psidash} \leftarrow \text{matrix}(c(st1, st2, st2, st3), ncol = 2) \]
\[ st4 \leftarrow (\text{tr}(P))^2 - 2 \times \text{tr}(P) \times \text{tr}(\text{varyinv}^2 \times \text{vstarstar}) + (\text{tr}(\text{varyinv}^2 \times \text{vstarstar}))^2 + 2 \times \text{tr}(\text{varyinv}^2 \times \text{vstarstar} \times \text{vstarstar}) \]
\[ st5 \leftarrow \text{tr}(P) \times \text{tr}(P \times G) - \text{tr}(P) \times \text{tr}(\text{varyinv} \times G \times \text{vstarstar}) - \text{tr}(P \times G) \times \text{tr}(\text{varyinv} \times G \times \text{vstarstar}) + \text{tr}(\text{varyinv} \times \text{vstarstar}) \times \text{tr}(\text{varyinv} \times G \times \text{vstarstar}) + 2 \times \text{tr}(\text{varyinv} \times \text{vstarstar} \times \text{vstarstar}) \times \text{tr}(\text{varyinv} \times G \times \text{vstarstar}) \]
\[ st6 \leftarrow (\text{tr}(P \times G))^2 - 2 \times \text{tr}(P \times G) \times \text{tr}(\text{varyinv} \times G \times \text{vstarstar}) + (\text{tr}(\text{varyinv} \times G \times \text{vstarstar}))^2 + 2 \times \text{tr}(\text{varyinv} \times G \times \text{vstarstar} \times \text{vstarstar}) \]

\[ \text{psipsit} \leftarrow .25 \times \text{matrix}(c(st4, st5, st5, st6), ncol = 2) \]

\[ \text{solve}(\text{psidash}) \times \text{psipsit} \times \text{solve}(\text{psidash}) \]
A2.5 DIAGNOSTIC TESTS

qqpairs <- function(yy, 1) {
    # yy is a list, each element being the vector of each treatment's observations
    # 1 is a vector containing the length of each element of y
    yy <- lapply(yy, sort)
    lmin <- min(l)
    if (lmin == 1) stop("pairs does not plot entire graphic if one treatment has
    only one observation")
    out <- matrix(ncol = length(yy), nrow = lmin)
    for(i in 1:length(yy)) {
        if(length(yy[[i]] > lmin)
            yy[[i]] <- as.vector(approx(1:length(yy[[i]]), yy[[i]],
                n = lmin)$y)
        else yy[[i]] <- as.vector(yy[[i]])
        out[,i] <- yy[[i]]
    }
    pairs(out)
}

diag.plot <- function(Z, y, e, b) {
    expt <- y * Z
    tot <- apply(expt, 2, sum)
    n <- apply(Z, 2, sum)
    ybar <- tot/n
    k <- e + n * b
    alphahat <- (sum(tot/k))/sum(n/k))
    alphahat1 <- rep(alphahat, ncol(Z))
    alphahat2 <- rep(alphahat, length(y))
    betahat <- (n * b * (ybar - alphahat1))/k
    Zbetahat <- rep(betahat, n)
    epsilonhat <- y - alphahat2 - Zbetahat
    list(epsilonhat, betahat)
}

myhat <- function(Z,e, b) {
    vary <- V(Z, e, b)
    varyinv <- solve(vary)
    J <- matrix(1, sum(Z), sum(Z))
    P <- varyinv - (varyinv %*% J %*% varyinv)/(sum(varyinv))
    D <- diag(b, ncol(Z))
    hat <- J %*% varyinv/(sum(varyinv)) + Z %*% D %*% t(Z) %*% P
diag(hat)}
det <- function(m)
{
  val <- eigen(m)
  prod(val$values)
}

dfthetas.ml <- function(Z, y, e, b) 
{
  dfthetas <- matrix(ncol = 2, nrow = nrow(Z))
  for(i in 1:nrow(Z)) {
    newZ <- Z[-i, ]
    newy <- y[-i]
    new.est <- anderson.ml(newZ, newy)
    new.var <- ml.var(newZ, new.est[1], new.est[2])
    dfthetas[i, 1] <- (e - new.est[1])/(new.var[1, 1])
    dfthetas[i, 2] <- (b - new.est[2])/(new.var[2, 2])
  }
  dfthetas
}

dfthetas.reml <- function(Z, y, e, b) 
{
  dfthetas <- matrix(ncol = 2, nrow = nrow(Z))
  for(i in 1:nrow(Z)) {
    newZ <- Z[-i, ]
    newy <- y[-i]
    new.est <- anderson.reml(newZ, newy)
    new.var <- reml.var(newZ, new.est[1], new.est[2])
    dfthetas[i, 1] <- (e - new.est[1])/(new.var[1, 1])
    dfthetas[i, 2] <- (b - new.est[2])/(new.var[2, 2])
  }
  dfthetas
}

covratio.ml <- function(Z, y, e, b) 
{
  var <- ml.var(Z, e, b)
  covratio <- vector(length = nrow(Z))
  for(i in 1:nrow(Z)) {
    newZ <- Z[-i, ]
    newy <- y[-i]
    new.est <- anderson.ml(newZ, newy)
    new.var <- ml.var(newZ, new.est[1], new.est[2])
    covratio[i] <- det(new.var)/det(var)
  }
  covratio
}

(A2.21)

(A2.22)

(A2.23)

(A2.24)
covratio.reml <- function(Z, y, e, b) {
  var <- reml.var(Z, e, b)
  covratio <- vector(length = nrow(Z))
  for(i in 1:nrow(Z)) {
    newZ <- Z[ - i, ]
    newy <- y[ - i]
    new.est <- anderson.reml(newZ, newy)
    new.var <- reml.var(newZ, new.est[1], new.est[2])
    covratio[i] <- det(new.var)/det(var)
  }
  covratio
}
REFERENCES


Hazel, L.N. and Terrill, C.E. (1945), "Heritability of weaning weight and staple length in range Rambouillet lambs".


Patterson, H.D. and Thompson, R. (1971), "Recovery of Inter–block Information when Block Sizes are Unequal", *Biometrika* 58, 545 – 554.
