1 General Linear model

Mixed Effects Models are seen as especially robust in the analysis of unbalanced data when compared to similar analyses done under the General Linear Model framework (Pinheiro and Bates, 2000).

A Mixed Effects Model is an extension of the General Linear Model that can specify additional random effects terms

1.1 Equivalence of LME model

Henderson's mixed model equations are presented on page 147 of Youngjo et al. Youngjo et al demonstrate that this formulation is equivalent to an augmented general linear model.

Youngjo et al show that the linear mixed effects model can be shown to be the augmented classical linear model involving fixed effects parameters only.

2 The LME model as a general linear model

Henderson's equations in (**??**) can be rewritten $(T'W^{-1}T)\delta = T'W^{-1}y_a$ using

$$\delta = \begin{pmatrix} \beta \\ b \end{pmatrix}, \; y_a = \begin{pmatrix} y \\ \psi \end{pmatrix}, \; T = \begin{pmatrix} X & Z \\ 0 & I \end{pmatrix}, \text{ and } W = \begin{pmatrix} \Sigma & 0 \\ 0 & D \end{pmatrix},$$

where **?** describe $\psi = 0$ as quasi-data with mean $\mathrm{E}(\psi) = b$. Their formulation suggests that the joint estimation of the coefficients β and b of the linear mixed effects model can be derived via a classical augmented general linear model $y_a = T\delta + \varepsilon$ where $\mathrm{E}(\varepsilon) = 0$ and $\mathrm{var}(\varepsilon) = W$, with *both* β and b appearing as fixed parameters. The usefulness of this reformulation of an LME as a general linear model will be revisited.

3 The LME model as a general linear model

Henderson's equations in can be rewritten $(T'W^{-1}T)\delta = T'W^{-1}y_a$ using

```
\[
\delta = \pmatrix{\beta \cr b},
\ y_{a} = \pmatrix{
y \cr \psi
},
\ T = \pmatrix{
X & Z  \cr
```

```
0 & I
},
\ \textrm{and} \ W = \pmatrix{
\Sigma & 0  \cr
0 &  D },
\]
```

where **cite[Lee:Neld:Pawi:2006]** describe $\psi = 0$ as quasi-data with mean $\mathrm{E}(\psi) = b$. Their formulation suggests that the joint estimation of the coefficients β and b of the linear mixed effects model can be derived via a classical augmented general linear model $y_a = T\delta + \varepsilon$ where $\mathrm{E}(\varepsilon) = 0$ and $\mathrm{var}(\varepsilon) = W$, with *both* β and b appearing as fixed parameters. The usefulness of this reformulation of an LME as a general linear model will be revisited.

Generalized linear models are a generalization of classical linear models.

4 Simplifying GLS (K Hayes)

4.1 Introduction

Hayes and Haslett (1998) present an approach to the problem of **general least squares** estimation of the general linear model in terms of constrained optimization, which is in turn solved via Lagrange multipliers. The crux of the proposed approach is that one system of equations is sufficiently versatile, and provides for

- the estimation of new observations,
- estimation of fixed parameters in regression
- estimation of fixed and random effects in mixed models,
- the diagnostics associated with conditional and marginal residuals
- and of subset deletion.

4.2 Overview

Hayes and Haslett (1998) have demonstrated how the problem of best linear unbiased estimation can be posed in terms of Lagrange multipliers. Both BLUE and BLUP can be treated as distinct estimation problems from the following equation.

$$\begin{pmatrix} V & X \\ X^t & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{\lambda}_z \\ \boldsymbol{\gamma}_z \end{pmatrix} = \begin{pmatrix} \mathrm{cov}(Y, Z) \\ A^t \end{pmatrix} \tag{1}$$

Hence BLUE and BLUP can be considered as the estimation of two different variables from Y. This equation has a natural role in the derivation of *leave- k-out* residuals

and diagnostic measures, and replaces the traditional approach of using a variety of clumsy updating formulas. Note that this approach may be used to determine the impact of deletion on any quantity computed from Y.

5 Generalized Least Squares

generalized least squares (GLS) is a technique for estimating the unknown parameters in a linear regression model. The GLS is applied when the variances of the observations are unequal (heteroscedasticity), or when there is a certain degree of correlation between the observations. In these cases ordinary least squares can be statistically inefficient, or even give misleading inferences.

$$Y = X\beta + \varepsilon, \qquad \mathrm{E}[\varepsilon|X] = 0,\ \mathrm{Var}[\varepsilon|X] = \Omega.$$

5.1 Introduction to Generalized Least Squares

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{\epsilon}_i \tag{2}$$

Estimation under this model has been studied extensively in the linear regression model.

6 Introduction

6.0.1 Robinson's (1991) review

Robinson's (1991) review of best linear unbiased prediction (BLUP), together with the subsequent discussion, has emphasized the very considerable range of models that may be addressed via the general least squares (GLS) solution to the general linear model $Y = X\beta + \varepsilon$, where $E(\varepsilon) = 0$ and $var(\varepsilon) = V$. These include linear mixed models, geostatistics, time series and multivariate regression.

The texts by Christensen (1996, 1991) and the connections to modern topics of image analysis, quality analysis, Bayesian methods, and splines (all in Robinson and discussion) make it an eminently suitable topic for teaching in any course concerning statistical linear models.

Nevertheless some of the matrix algebra that results from solving the normal equations for individual specifications of the general linear model will be daunting, and far from intuitive for many students, even those who are at home in linear space. The conventional approach to prediction and estimation from data Y associated with covariates X via the general linear model $Y = X\beta + \varepsilon$ is essentially a two-stage process.

The first stage is to determine the best,in the GLS sense, estimator $\hat{\beta}$ of β and subsequently to determine everything else from this.

The estimator is said to be best if it minimizes the generalization of the sum of squares $\hat{e}^t V^{-1} \hat{e}$, where $\hat{e} = Y - X\hat{\beta}$

It is straightforward to show that $\hat{\beta} = (X^t V^{-l} X)^{-l} X^t V^{-l} Y = BY$ and at the minimum the sum of squares is $Y^t(V^{-l} - V^{-l}(X^t V^{-l} X)^{-l} X^t V^{-l})Y = Y^t QY$.

The purpose of this note is to give emphasis to one derivation, based on Lagrange multipliers, which leads to a system of equations that is very intuitive and lends itself readily to specialization. This approach is in fact standard in the geostatistical treatment of ***kriging*** *(see Matheron 1962; Journel and Huijbregts 1981; Ripley 1981; Cressie 1993). In the genetics literature it is associated with the name of Henderson (1983); or in the classical statistical literature Hocking (1996, p. 73) is a suitable reference.*

The approach based on Lagrange multipliers deemphasizes the explicit determination of $\hat{\beta}$ and leads to a clearer understanding of the complementary (but for some confusing) tasks known as best linear unbiased estimation (BLUE) and best linear unbiased prediction (BLUP). Regrettably, Robinson-despite offering four derivations, and having as his main concern the interplay of BLUP and BLUE-gives it little prominence.

It has recently been discussed by Searle (1997, p; 278) who said that it makes another approach (Searle, Casella, and McCulloch 1992, p. 271) seem "obtuse and unnecessarily complicated." By contrast, our treatment emphasizes the fact that it leads to a single set of equations whose solution sheds simplifying light on very many issues in general least squares.

The American Statistician's Teacher's Corner (e.g., McLean, Sanders, and Stroup 1991; Puntanen and Styan 1989) has already played host to previous attempts to simplify the explanation of such topics. Various authors (CPJ, Haslett Hayes ,Martin) have visited the more specialized area of diagnostics and have developed ***down-dating*** (leave-k-out) formulas.

The conventional approach here is via tricky identities based on the inverses of partitioned matrices. Here again the Lagrange system of equations leads to a much simplified and-we claim-much more intuitive derivation of these more technical results.

The essence of the approach is to seek that linear combination of the available data Y which is best for the estimation of Z among those linear estimators which are constrained to be unbiased. We adopt therefore a constrained minimization approach, using Lagrange multipliers. By best we mean that combination $\hat{Z}(Y) = \lambda_z^t Y$ which has least mean square error $E(Z-\lambda_z^t Y)^2$, and by unbiased we mean $E(Z-\lambda_z^t Y)) = 0$. Here Z denotes that scalar which is to be the objective of the estimation. This estimator is written as $\hat{Z}(Y)$ to make its dependence on Y explicit. Note that the term "best" is applied in the context of minimizing the prediction variance $var(Z - Z(Y))$. We shall see that Z may be used to denote either a random variable or an unknown parameter, and that it will be sufficient to specify Z via $E[Z]$ and $cov(Z, Y)$. If Z is not a random variable then of course the latter is zero and $E[Z] = Z$. We establish-very simply, as below-a general solution in terms of A and cov(Z, Y) and achieve particular tasks by identification of these. Our presentation is for a scalar Z, but the notation facilitates generalization to vector Z.

6.1 Predictors and Estimators

We note that Robinson (1991) stated "A convention has somehow developed that estimators of random effects are called predictors while estimators of fixed effects are called estimators." We agree that this distinction is confusing and indeed unnecessary.

We seek $\hat{Z}(Y) = \lambda_z^t Y$, where λ_z^t, is an $n \times 1$ vector of estimation coefficients. It is convenient to specify $E[Z] = A\beta$ for known A. In this context A denotes a row vector, but we generalize this in the following. The constraint requiring $\hat{Z}(Y)$ to be unbiased now reduces to $(A - \lambda_z^t X) = 0$. A solution is found by minimizing $var(Z - \lambda_z^t Y) + \gamma_z^t(X^t\lambda_z - A^t)$, where γ_z is a $p \times 1$ vector of Lagrange multipliers, where p is the length of the parameter vector β. Setting to zero the derivatives with respect to λ_z and γ_z yields the system.

$$\begin{pmatrix} V & X \\ X^t & 0 \end{pmatrix} \begin{pmatrix} \lambda_z \\ \gamma_z \end{pmatrix} = \begin{pmatrix} \text{cov}(Y, Z) \\ A^t \end{pmatrix} \tag{3}$$

If the inverse exists we have that

$$\begin{pmatrix} \lambda_z \\ \gamma_z \end{pmatrix} = \begin{pmatrix} V & X \\ X^t & 0 \end{pmatrix}^{-1} \begin{pmatrix} \text{cov}(Y, Z) \\ A^t \end{pmatrix} \tag{4}$$

so that

$$\hat{Z}(Y) = \begin{pmatrix} \lambda_z^t & \gamma_z^t \end{pmatrix} = \begin{pmatrix} Y \\ 0 \end{pmatrix}$$

In terms of the estimation problem being considered the square matrix on the left-hand side of (1) concerns "what we have," namely, the data plus constraints.

The matrix does not depend on Z and consequently need only be constructed once before application to a range of problems. The right- hand side contains the term $cov(Z, Y)$ and can be specified for whatever Z is being considered.

It is this feature of system (1) that makes a generic approach to estimation possible.

7 Hierarchical likelihood

Inferential method was developed for the mixed linear model via Lee and Nelder's (1996) hierarchical-likelihood (h-likelihood).

8 Importance-Weighted Least-Squares (IWLS)

9 Augmented GLMs

With the use of h-likihood, a random effected model of the form can be viewed as an 'augmented GLM' with the response varaibkes $(y^t, \phi_m^t)^t$, (with $\mu = E(y)$,$u = E(\phi)$, $var(y) = \theta V(\mu)$. The augmented linear predictor is

$$\eta_{ma} = (\eta^t, \eta_m^t)^t) = T\omega.$$

.

The subscript M is a label referring to the mean model.

$$\begin{pmatrix} Y \\ \psi_M \end{pmatrix} = \begin{pmatrix} X & Z \\ 0 & I \end{pmatrix} \begin{pmatrix} \beta \\ \nu \end{pmatrix} + e^* \tag{5}$$

The error term e^* is normal with mean zero. The variance matrix of the error term is given by

$$\Sigma_a = \begin{pmatrix} \Sigma & 0 \\ 0 & D \end{pmatrix}. \tag{6}$$

$$y_a = T\delta + e^* \tag{7}$$

Weighted least squares equation

10 Augmented linear model

The subscript M is a label referring to the mean model.

$$\begin{pmatrix} Y \\ \psi_M \end{pmatrix} = \begin{pmatrix} X & Z \\ 0 & I \end{pmatrix} \begin{pmatrix} \beta \\ \nu \end{pmatrix} + e^* \tag{8}$$

The error term e^* is normal with mean zero. The variance matrix of the error term is given by

$$\Sigma_a = \begin{pmatrix} \Sigma & 0 \\ 0 & D \end{pmatrix}. \tag{9}$$

$$X = \begin{pmatrix} T & Z \\ 0 & I \end{pmatrix} \delta = \begin{pmatrix} \beta \\ \nu \end{pmatrix} \tag{10}$$

$$y_a = T\delta + e^* \tag{11}$$

Weighted least squares equation

11 Augmented GLMs

With the use of h-likelihood, a random effected model of the form can be viewed as an 'augmented GLM' with the response variables $(y^t, \phi_m^t)^t$, (with $\mu = E(y)$,$u = E(\phi)$, $var(y) = \theta V(\mu)$. The augmented linear predictor is

$$\eta_{ma} = (\eta^t, \eta_m^t)^t) = T\omega.$$

.

The subscript M is a label referring to the mean model.

$$\begin{pmatrix} Y \\ \psi_M \end{pmatrix} = \begin{pmatrix} X & Z \\ 0 & I \end{pmatrix} \begin{pmatrix} \beta \\ \nu \end{pmatrix} + e^* \tag{12}$$

The error term e^* is normal with mean zero. The variance matrix of the error term is given by

$$\Sigma_a = \begin{pmatrix} \Sigma & 0 \\ 0 & D \end{pmatrix}. \tag{13}$$

$y_a = T\delta + e^*$

Weighted least squares equation

Generalized linear models are a generalization of classical linear models.

12 The Augmented Model Matrix

$$X = \begin{pmatrix} T & Z \\ 0 & I \end{pmatrix} \delta = \begin{pmatrix} \beta \\ \nu \end{pmatrix} \tag{14}$$

13 Augmented GLMs

With the use of h-likihood, a random effected model of the form can be viewed as an 'augmented GLM' with the response varaibkes $(y^t, \phi_m^t)^t$, (with $\mu = E(y)$,$u = E(\phi)$, $var(y) = \theta V(\mu)$. The augmented linear predictor is

$$\eta_{ma} = (\eta^t, \eta_m^t)^t) = T\omega.$$

.

The subscript M is a label referring to the mean model.

$$\begin{pmatrix} Y \\ \psi_M \end{pmatrix} = \begin{pmatrix} X & Z \\ 0 & I \end{pmatrix} \begin{pmatrix} \beta \\ \nu \end{pmatrix} + e^* \tag{15}$$

The error term e^* is normal with mean zero. The variance matrix of the error term is given by

$$\Sigma_a = \begin{pmatrix} \Sigma & 0 \\ 0 & D \end{pmatrix}. \tag{16}$$

Weighted least squares equation

14 Grubbs' Data

For the Grubbs data the $\hat{\beta}$ estimated are $\hat{\beta}_0$ and $\hat{\beta}_1$ respectively. Leaving the fourth case out, i.e. $k = 4$ the corresponding estimates are $\hat{\beta}_0^{-4}$ and $\hat{\beta}_1^{-4}$

$$Y^{-Q} = \hat{\beta}^{-Q} X^{-Q} \tag{17}$$

When considering the regression of case-wise differences and averages, we write $D^{-Q} = \hat{\beta}^{-Q} A^{-Q}$

	F	C	D	A
1	793.80	794.60	-0.80	794.20
2	793.10	793.90	-0.80	793.50
3	792.40	793.20	-0.80	792.80
4	794.00	794.00	0.00	794.00
5	791.40	792.20	-0.80	791.80
6	792.40	793.10	-0.70	792.75
7	791.70	792.40	-0.70	792.05
8	792.30	792.80	-0.50	792.55
9	789.60	790.20	-0.60	789.90
10	794.40	795.00	-0.60	794.70
11	790.90	791.60	-0.70	791.25
12	793.50	793.80	-0.30	793.65

$$Y^{(k)} = \hat{\beta}^{(k)} X^{(k)} \tag{18}$$

Consider two sets of measurements , in this case F and C , with the vectors of case-wise averages A and case-wise differences D respectively. A regression model of differences on averages can be fitted with the view to exploring some characteristics of the data.

When considering the regression of case-wise differences and averages, we write

$$D^{-Q} = \hat{\beta}^{-Q} A^{-Q} \tag{19}$$

Let $\hat{\beta}$ denote the least square estimate of β based upon the full set of observations, and let $\hat{\beta}^{(k)}$ denoted the estimate with the k^{th} case excluded.

For the Grubbs data the $\hat{\beta}$ estimated are $\hat{\beta}_0$ and $\hat{\beta}_1$ respectively. Leaving the fourth case out, i.e. $k = 4$ the corresponding estimates are $\hat{\beta}_0^{-4}$ and $\hat{\beta}_1^{-4}$

$$Y^{(k)} = \hat{\beta}^{(k)} X^{(k)} \tag{20}$$

Consider two sets of measurements , in this case F and C , with the vectors of case-wise averages A and case-wise differences D respectively. A regression model of

differences on averages can be fitted with the view to exploring some characteristics of the data.

```
Call: lm(formula = D ~ A)

Coefficients: (Intercept)            A
-37.51896      0.04656
```

When considering the regression of case-wise differences and averages, we write

$$D^{-Q} = \hat{\beta}^{-Q} A^{-Q} \tag{21}$$

subsectionInfluence measures using R R provides the following influence measures of each observation.

	dfb.1_	dfb.A	dffit	cov.r	cook.d	hat
1	0.42	-0.42	-0.56	1.13	0.15	0.18
2	0.17	-0.17	-0.34	1.14	0.06	0.11
3	0.01	-0.01	-0.24	1.17	0.03	0.08
4	-1.08	1.08	1.57	0.24	0.56	0.16
5	-0.14	0.14	-0.24	1.30	0.03	0.13
6	-0.00	0.00	-0.11	1.31	0.01	0.08
7	-0.04	0.04	-0.08	1.37	0.00	0.11
8	0.02	-0.02	0.15	1.28	0.01	0.09
9	0.69	-0.68	0.75	2.08	0.29	0.48
10	0.18	-0.18	-0.22	1.63	0.03	0.27
11	-0.03	0.03	-0.04	1.53	0.00	0.19
12	-0.25	0.25	0.44	1.05	0.09	0.12

15 Influence measures using R

R provides the following influence measures of each observation.

16 Augmented GLMs

Generalized linear models are a generalization of classical linear models.

With the use of h-likihood, a random effected model of the form can be viewed as an 'augmented GLM' with the response varaibkes $(y^t, \phi_m^t)^t$, (with $\mu = E(y)$,$u = E(\phi)$, $var(y) = \theta V(\mu)$. The augmented linear predictor is

$$\eta_{ma} = (\eta^t, \eta_m^t)^t) = T\omega.$$

	dfb.1_	dfb.A	dffit	cov.r	cook.d	hat
1	0.42	-0.42	-0.56	1.13	0.15	0.18
2	0.17	-0.17	-0.34	1.14	0.06	0.11
3	0.01	-0.01	-0.24	1.17	0.03	0.08
4	-1.08	1.08	1.57	0.24	0.56	0.16
5	-0.14	0.14	-0.24	1.30	0.03	0.13
6	-0.00	0.00	-0.11	1.31	0.01	0.08
7	-0.04	0.04	-0.08	1.37	0.00	0.11
8	0.02	-0.02	0.15	1.28	0.01	0.09
9	0.69	-0.68	0.75	2.08	0.29	0.48
10	0.18	-0.18	-0.22	1.63	0.03	0.27
11	-0.03	0.03	-0.04	1.53	0.00	0.19
12	-0.25	0.25	0.44	1.05	0.09	0.12

.

The subscript M is a label referring to the mean model.

$$\begin{pmatrix} Y \\ \psi_M \end{pmatrix} = \begin{pmatrix} X & Z \\ 0 & I \end{pmatrix} \begin{pmatrix} \beta \\ \nu \end{pmatrix} + e^* \tag{22}$$

The error term e^* is normal with mean zero. The variance matrix of the error term is given by

$$\Sigma_a = \begin{pmatrix} \Sigma & 0 \\ 0 & D \end{pmatrix}. \tag{23}$$

Weighted least squares equation

17 Application to MCS

Let $\hat{\beta}$ denote the least square estimate of β based upon the full set of observations, and let $\hat{\beta}^{(k)}$ denoted the estimate with the k^{th} case excluded.

18 Grubbs' Data

For the Grubbs data the $\hat{\beta}$ estimated are $\hat{\beta}_0$ and $\hat{\beta}_1$ respectively. Leaving the fourth case out, i.e. $k = 4$ the corresponding estimates are $\hat{\beta}_0^{-4}$ and $\hat{\beta}_1^{-4}$

$$Y^{-Q} = \hat{\beta}^{-Q} X^{-Q} \tag{24}$$

When considering the regression of case-wise differences and averages, we write $D^{-Q} = \hat{\beta}^{-Q} A^{-Q}$

	F	C	D	A
1	793.80	794.60	-0.80	794.20
2	793.10	793.90	-0.80	793.50
3	792.40	793.20	-0.80	792.80
4	794.00	794.00	0.00	794.00
5	791.40	792.20	-0.80	791.80
6	792.40	793.10	-0.70	792.75
7	791.70	792.40	-0.70	792.05
8	792.30	792.80	-0.50	792.55
9	789.60	790.20	-0.60	789.90
10	794.40	795.00	-0.60	794.70
11	790.90	791.60	-0.70	791.25
12	793.50	793.80	-0.30	793.65

$$Y^{(k)} = \hat{\beta}^{(k)} X^{(k)} \tag{25}$$

Consider two sets of measurements , in this case F and C , with the vectors of case-wise averages A and case-wise differences D respectively. A regression model of differences on averages can be fitted with the view to exploring some characteristics of the data.

When considering the regression of case-wise differences and averages, we write

$$D^{-Q} = \hat{\beta}^{-Q} A^{-Q} \tag{26}$$

Let $\hat{\beta}$ denote the least square estimate of β based upon the full set of observations, and let $\hat{\beta}^{(k)}$ denoted the estimate with the k^{th} case excluded.

For the Grubbs data the $\hat{\beta}$ estimated are $\hat{\beta}_0$ and $\hat{\beta}_1$ respectively. Leaving the fourth case out, i.e. $k = 4$ the corresponding estimates are $\hat{\beta}_0^{-4}$ and $\hat{\beta}_1^{-4}$

$$Y^{(k)} = \hat{\beta}^{(k)} X^{(k)} \tag{27}$$

Consider two sets of measurements , in this case F and C , with the vectors of case-wise averages A and case-wise differences D respectively. A regression model of differences on averages can be fitted with the view to exploring some characteristics of the data.

```
Call: lm(formula = D ~ A)

Coefficients: (Intercept)            A
-37.51896      0.04656
```

When considering the regression of case-wise differences and averages, we write

$$D^{-Q} = \hat{\beta}^{-Q} A^{-Q} \tag{28}$$

19 Augmented GLMs

With the use of h-likihood, a random effected model of the form can be viewed as an 'augmented GLM' with the response varaibkes $(y^t, \phi_m^t)^t$, (with $\mu = E(y)$,$u = E(\phi)$, $var(y) = \theta V(\mu)$. The augmented linear predictor is

$$\eta_{ma} = (\eta^t, \eta_m^t)^t) = T\omega.$$

.

The subscript M is a label referring to the mean model.

$$\begin{pmatrix} Y \\ \psi_M \end{pmatrix} = \begin{pmatrix} X & Z \\ 0 & I \end{pmatrix} \begin{pmatrix} \beta \\ \nu \end{pmatrix} + e^* \tag{29}$$

The error term e^* is normal with mean zero. The variance matrix of the error term is given by

$$\Sigma_a = \begin{pmatrix} \Sigma & 0 \\ 0 & D \end{pmatrix}. \tag{30}$$

$y_a = T\delta + e^*$
Weighted least squares equation

20 Application to MCS

Let $\hat{\beta}$ denote the least square estimate of β based upon the full set of observations, and let $\hat{\beta}^{(k)}$ denoted the estimate with the k^{th} case excluded.