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 Simplifying GLS (K Hayes)

3.1 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.

3.2 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.

4 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.

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

4.1 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.

5 Hierarchical likelihood

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

6 Importance-Weighted Least-Squares (IWLS)

7 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.

$$\begin{pmatrix} Y \\ \psi_M \end{pmatrix} = \begin{pmatrix} X & Z \\ 0 & I \end{pmatrix} \begin{pmatrix} \beta \\ \nu \end{pmatrix} + e^* \tag{3}$$

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

$$\Sigma_a = \begin{pmatrix} \Sigma & 0 \\ 0 & D \end{pmatrix}. \tag{4}$$

$$X = \begin{pmatrix} T & Z \\ 0 & I \end{pmatrix} \delta = \begin{pmatrix} \beta \\ \nu \end{pmatrix} \tag{5}$$

$$y_a = T\delta + e^* \tag{6}$$

Weighted least squares equation
$y_a = T\delta + e^*$
Weighted least squares equation
Generalized linear models are a generalization of classical linear models.
The subscript M is a label referring to the mean model.

$$\begin{pmatrix} Y \\ \psi_M \end{pmatrix} = \begin{pmatrix} X & Z \\ 0 & I \end{pmatrix} \begin{pmatrix} \beta \\ \nu \end{pmatrix} + e^* \tag{7}$$

8 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

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

.

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

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

Weighted least squares equation

9 Augmented GLMs

9.1 Augmented linear model

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

$$\begin{pmatrix} Y \\ \psi_M \end{pmatrix} = \begin{pmatrix} X & Z \\ 0 & I \end{pmatrix} \begin{pmatrix} \beta \\ \nu \end{pmatrix} + e^* \tag{10}$$

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

$$\Sigma_a = \begin{pmatrix} \Sigma & 0 \\ 0 & D \end{pmatrix}. \tag{11}$$

$$X = \begin{pmatrix} T & Z \\ 0 & I \end{pmatrix} \delta = \begin{pmatrix} \beta \\ \nu \end{pmatrix} \tag{12}$$

$$y_a = T\delta + e^* \tag{13}$$

Weighted least squares equation

Generalized linear models are a generalization of classical linear models.

10 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 $\hat{\beta}_1^{-4}$

$$Y^{-Q} = \hat{\beta}^{-Q} X^{-Q} \tag{14}$$

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

$$Y^{(k)} = \hat{\beta}^{(k)} X^{(k)} \tag{15}$$

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

$$D^{-Q} = \hat{\beta}^{-Q} A^{-Q} \tag{16}$$

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 $\hat{\beta}_1^{-4}$

$$Y^{(k)} = \hat{\beta}^{(k)} X^{(k)} \tag{17}$$

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

$$D^{-Q} = \hat{\beta}^{-Q} A^{-Q} \tag{18}$$

11 Influence 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

12 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 $\hat{\beta}_1^{-4}$

$$Y^{-Q} = \hat{\beta}^{-Q} X^{-Q} \tag{19}$$

	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

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

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

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.

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

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

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 $\hat{\beta}_1^{-4}$

$$Y^{(k)} = \hat{\beta}^{(k)} X^{(k)} \tag{22}$$

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

$$D^{-Q} = \hat{\beta}^{-Q} A^{-Q} \tag{23}$$

13 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.