

Contents

1	Introduction	3
1.1	Robinson's (1991) review	3
1.2	Predictors and Estimators	5
2	Analysis of Influence	6
2.1	Further Assumptions of Linear Models	6
2.2	Stating the LME Model	6
2.3	Summary of Schabenberger's Paper	7
2.4	Summary of Paper	8
3	Case Deletion Diagnostics for LME models	9
4	Haslett's Analysis	9
4.1	Residual Diagnostics in LME models	11
4.2	Residual Diagnostics in LME models	12
5	Case Deletion Diagnostics for LME models	12
5.1	Remarks on the Multivariate Normal Distribution	15
6	Modelling Agreement with LME Models	15
7	Cook's Distance	16
7.1	Matrix Notation for Case Deletion	17
8	Cook's Distance	17
9	Exention of Cook's Distance methodology to LME models	18
10	Exention of Cook's Distance methodology to LME models	18
11	Cook's Distance	19
12	Extension of Diagnostic Methods to LME models	19
13	Cook's Distance for LMEs	21
14	Cook's Distance for LMEs	21
14.0.1	Random Effects	21
14.0.2	linear functions	21
14.0.3	linear functions	22
14.1	Change in the precision of estimates	22
15	Robinson's (1991) review	22

16 Robinson's (1991) review	24
17 Predictors and Estimators	24
18 Predictors and Estimators	25
19 Simplifying GLS	26
20 The extended likelihood	27
21 Cook's Distance	27
22 Repeated measurements in LME models	30

- R command and R object - Typewriter Font
- R Package name - Italics
- Selected Acronyms and Proper Nouns - Italics

- This chapter is broken into two parts. The first part is a review of diagnostics methods for linear models, intended to acquaint the reader with the subject, and also to provide a basis for material covered in the second part. Particular attention is drawn to graphical methods.
- The second part of the chapter looks at diagnostics techniques for LME models, firstly covering the theory, then proceeding to a discussion on implementing these using R code.
- While a substantial body of work has been developed in this area, there are still areas worth exploring. In particular the development of graphical techniques pertinent to LME models should be looked at.

1 Introduction

1.1 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 $\text{var}(\varepsilon) = V$. 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 $\hat{\beta} = (X^t V^{-l} X)^{-l} X^t V^{-l} Y = B Y$ and at the minimum the sum of squares is $Y^t (V^{-l} - V^{-l} (X^t V^{-l} X)^{-l} X^t V^{-l}) Y = Y^t Q Y$.

*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 $var(Z - \hat{Z}(Y))$. 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 .

1.2 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.

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

If the inverse exists we have that

$$\begin{pmatrix} \lambda_z \\ \gamma_z \end{pmatrix} = \begin{pmatrix} V & X \\ X^t & 0 \end{pmatrix}^{-1} \begin{pmatrix} cov(Y, Z) \\ A^t \end{pmatrix} \quad (2)$$

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.

Model Diagnostics

Abstract

This chapter is broken into two parts. The first part is a review of diagnostics methods for linear models, intended to acquaint the reader with the subject, and also to provide a basis for material covered in the second part. Particular attention is drawn to graphical methods.

2 Analysis of Influence

2.1 Further Assumptions of Linear Models

As with fitted models, the assumption of normality of residuals and homogeneity of variance is applicable to LMEs also.

Homoscedascity is the technical term to describe the variance of the residuals being constant across the range of predicted values. Heteroscedascity is the converse scenario : the variance differs along the range of values.

In recent years, mixed models have become invaluable tools in the analysis of experimental and observational data. In these models, more than one term can be subject to random variation. Mixed model technology enables you to analyze complex experimental data with hierarchical random processes, temporal, longitudinal, and spatial data, to name just a few important applications.

2.2 Stating the LME Model

The general linear mixed model is

$$Y = X\beta + Zu + \varepsilon$$

where Y is a $(n \times 1)$ vector of observed data, X is an $(n \times p)$ fixed-effects design or regressor matrix of rank k , Z is a $(n \times g)$ random-effects design or regressor matrix, u is a $(g \times 1)$ vector of random effects, and ε is an $(n \times 1)$ vector of model errors (also random effects). The distributional assumptions made by the MIXED procedure are as follows: u is normal with mean 0 and variance G ; ε is normal with mean 0 and variance R ; the random components u and ε are independent. Parameters of this model are the fixed-effects and all unknowns in the variance matrices G and R . The unknown variance elements are referred to as the covariance parameters and collected in the vector *theta*.

The concept of critiquing the model-data agreement applies in mixed models in the same way as in linear fixed-effects models. In fact, because of the more complex model structure, you can argue that model and data diagnostics are even more important. For example, you are not only concerned with capturing the important variables in the model. You are also concerned with distributing them correctly between the fixed and random components of the model. The mixed model structure presents unique and interesting challenges that prompt us to reexamine the traditional ideas of influence and residual analysis. This paper presents the extension of traditional tools and statistical measures for influence and residual analysis to the linear mixed model and demonstrates their implementation in the MIXED procedure (experimental features in SAS 9.1). The remainder of this paper is organized as follows. The Background section briefly discusses some mixed model estimation theory and the challenges to model diagnosis that result from it.

2.3 Summary of Schabenberger's Paper

On occasion, quantification is not possible. Assume, for example, that a data point is removed and the new estimate of the G matrix is not positive definite. This may occur if a variance component estimate now falls on the boundary of the parameter space. Thus, it may not be possible to compute certain influence statistics comparing the full-data and reduced-data parameter estimates. However, knowing that a new singularity was encountered is important qualitative information about the data points influence on the analysis.

The basic procedure for quantifying influence is simple:

1. Fit the model to the data and obtain estimates of all parameters.
2. Remove one or more data points from the analysis and compute updated estimates of model parameters.
3. Based on full- and reduced-data estimates, contrast quantities of interest to determine how the absence of the observations changes the analysis.

We use the subscript (U) to denote quantities obtained without the observations in the set U. For example, (U) denotes the fixed-effects *leave-U-out* estimates. Note that the set U can contain multiple observations.

If the global measure suggests that the points in U are influential, you should next determine the nature of that influence. In particular, the points can affect

- the estimates of fixed effects
- the estimates of the precision of the fixed effects
- the estimates of the covariance parameters
- the estimates of the precision of the covariance parameters
- fitted and predicted values

It is important to further decompose the initial finding to determine whether data points are actually troublesome. Simply because they are influential somehow, should not trigger their removal from the analysis or a change in the model. For example, if points primarily affect the precision of the covariance parameters without exerting much influence on the fixed effects, then their presence in the data may not distort hypothesis tests or confidence intervals about β .

2.4 Summary of Paper

Standard residual and influence diagnostics for linear models can be extended to LME models. The dependence of the fixed effects solutions on the covariance parameters has important ramifications on the perturbation analysis. Calculating the studentized residuals-And influence statistics whereas each software procedure can calculate both conditional and marginal raw residuals, only SAS Proc Mixed is currently the only program that provide studentized residuals Which are preferred for model diagnostics. The conditional Raw residuals are not well suited to detecting outliers as are the studentized conditional residuals. (schabenbege r)

LME are flexible tools for the analysis of clustered and repeated measurement data. LME extend the capabilities of standard linear models by allowing unbalanced and missing data, as long as the missing data are MAR. Structured covariance matrices for both the random effects G and the residuals R . missing at Random.

A conditional residual is the difference between the observed value and the predicted value of a dependent variable- Influence diagnostics are formal techniques that allow the identification observation that heavily influence estimates of parameters. To alleviate the problems with the interpretation of conditional residuals that may have unequal variances, we consider scaling. Residuals obtained in this manner are called studentized residuals.

- Standard residual and influence diagnostics for linear models can be extended to linear mixed models. The dependence of fixed-effects solutions on the covariance parameter estimates has important ramifications in perturbation analysis.
- To gauge the full impact of a set of observations on the analysis, covariance parameters need to be updated, which requires refitting of the model.
- The conditional (subject-specific) and marginal (population-averaged) formulations in the linear mixed model enable you to consider conditional residuals that use the estimated BLUPs of the random effects, and marginal residuals which are deviations from the overall mean.
- Residuals using the BLUPs are useful to diagnose whether the random effects components in the model are specified correctly, marginal residuals are useful to diagnose the fixed-effects components.
- Both types of residuals are available in SAS 9.1 as an experimental option of the MODEL statement in the MIXED procedure.
- It is important to note that influence analyses are performed under the assumption that the chosen model is correct. Changing the model structure can alter the conclusions. Many other variance models have been fit to the data presented in the repeated measures example. You need to see the conclusions about which model component is affected in light of the model being fit.

3 Case Deletion Diagnostics for LME models

Haslett & Dillane (19XX) remark that linear mixed effects models didn't experience a corresponding growth in the use of deletion diagnostics, adding that McCullough and Searle (2001) makes no mention of diagnostics whatsoever.

Christensen (19XX) describes three propositions that are required for efficient case-deletion in LME models. The first proposition describes how to efficiently update V when the i th element is deleted.

$$V_{[i]}^{-1} = \Lambda_{[i]} - \frac{\lambda\lambda'}{\nu ii} \quad (3)$$

The second of Christensen's propositions is the following set of equations, which are variants of the Sherman Wood bury updating formula.

$$X'_{[i]} V_{[i]}^{-1} X_{[i]} = X' V^{-1} X - \frac{\hat{x}_i \hat{x}_i'}{s_i} \quad (4)$$

$$(X'_{[i]} V_{[i]}^{-1} X_{[i]})^{-1} = (X' V^{-1} X)^{-1} + \frac{(X' V^{-1} X)^{-1} \hat{x}_i \hat{x}_i' (X' V^{-1} X)^{-1}}{s_i - \bar{h}_i} \quad (5)$$

$$X'_{[i]} V_{[i]}^{-1} Y_{[i]} = X' V^{-1} Y - \frac{\hat{x}_i \hat{y}_i'}{s_i} \quad (6)$$

In LME models, fitted by either ML or REML, an important overall influence measure is the likelihood distance (?). The procedure requires the calculation of the full data estimates $\hat{\psi}$ and estimates based on the reduced data set $\hat{\psi}_{(U)}$. The likelihood distance is given by determining

$$LD_{(U)} = 2\{l(\hat{\psi}) - l(\hat{\psi}_{(U)})\} \quad (7)$$

$$RLD_{(U)} = 2\{l_R(\hat{\psi}) - l_R(\hat{\psi}_{(U)})\} \quad (8)$$

Haslett & Dillane (199X) offers an procedure to assess the influences for the variance components within the linear model, complementing the existing methods for the fixed components.

The essential problem is that there is no useful updating procedures for \hat{V} , or for \hat{V}^{-1} . Haslett & Dillane (199X) propose an alternative, and computationally inexpensive approach, making use of the 'delete=replace' identity.

Haslett (1999) considers the effect of 'leave k out' calculations on the parameters β and σ^2 , using several key results from Haslett and Hayes (1998) on partitioned matrices.

4 Haslett's Analysis

For fixed effect linear models with correlated error structure Haslett (1999) showed that the effects on the fixed effects estimate of deleting each observation in turn could be cheaply computed from the fixed effects model predicted residuals.

A general theory is presented for residuals from the general linear model with correlated errors. It is demonstrated that there are two fundamental types of residual associated with this model, referred to here as the marginal and the conditional residual.

These measure respectively the distance to the global aspects of the model as represented by the expected value and the local aspects as represented by the conditional expected value.

These residuals may be multivariate.

Haslett and Hayes (1998) developes some important dualities which have simple implications for diagnostics.

Haslett and Hayes - Residuals

Haslett and Hayes (1998) and Haslett (1999) considered the case of an LME model with correlated covariance structure.

4.1 Residual Diagnostics in LME models

- A **residual** is the difference between the observed quantity and the predicted value. In LME models a distinction is made between marginal residuals and conditional residuals.
- A **Marginal residual** is the difference between the observed data and the estimated marginal mean (Schabenberger pg3) The computation of case deletion diagnostics in the classical model is made simple by the fact that important estimates can be computed without refitting the model.
- Such update formulae are available in the mixed model only if you assume that the covariance parameters are not affected by the removal of the observation in question. Schabenberger remarks that this is not a reasonable assumption.

Basic procedure for quantifying influence is simple

1. Fit the model to the data
2. Remove one or more data points from the analysis and compute updated estimates of model parameters
3. Based on the full and reduced data estimates, contrast quantities of interest to determine how the absence of the observations changed the analysis.

The likelihood distance is a global summary measure expressing the joint influence of the observations in the set U on all parameters in Ψ that were subject to updating.

4.2 Residual Diagnostics in LME models

- A **residual** is the difference between the observed quantity and the predicted value. In LME models a distinction is made between marginal residuals and conditional residuals.
- A **Marginal residual** is the difference between the observed data and the estimated marginal mean (Schabenberger pg3) The computation of case deletion diagnostics in the classical model is made simple by the fact that important estimates can be computed without refitting the model.
- Such update formulae are available in the mixed model only if you assume that the covariance parameters are not affected by the removal of the observation in question. Schabenberger remarks that this is not a reasonable assumption.

Basic procedure for quantifying influence is simple

1. Fit the model to the data
2. Remove one or more data points from the analysis and compute updated estimates of model parameters
3. Based on the full and reduced data estimates, contrast quantities of interest to determine how the absence of the observations changed the analysis.

The likelihood distance is a global summary measure expressing the joint influence of the observations in the set U on all parameters in Ψ that were subject to updating.

5 Case Deletion Diagnostics for LME models

Haslett and Dillane (2004) remark that linear mixed effects models didn't experience a corresponding growth in the use of deletion diagnostics, adding that McCullough and Searle (2001) makes no mention of diagnostics whatsoever.

? describes three propositions that are required for efficient case-deletion in LME models. The first proposition describes how to efficiently update V when the i th element is deleted.

$$V_{[i]}^{-1} = \Lambda_{[i]} - \frac{\lambda\lambda'}{\nu_{ii}} \quad (9)$$

The second of Christensen's propositions is the following set of equations, which are variants of the Sherman Woodbury updating formula.

$$X'_{[i]} V_{[i]}^{-1} X_{[i]} = X' V^{-1} X - \frac{\hat{x}_i \hat{x}_i'}{s_i} \quad (10)$$

$$(X'_{[i]} V_{[i]}^{-1} X_{[i]})^{-1} = (X' V^{-1} X)^{-1} + \frac{(X' V^{-1} X)^{-1} \hat{x}_i \hat{x}_i' (X' V^{-1} X)^{-1}}{s_i - \bar{h}_i} \quad (11)$$

$$X'_{[i]} V_{[i]}^{-1} Y_{[i]} = X' V^{-1} Y - \frac{\hat{x}_i \hat{y}_i'}{s_i} \quad (12)$$

In LME models, fitted by either ML or REML, an important overall influence measure is the likelihood distance (?). The procedure requires the calculation of the full data estimates $\hat{\psi}$ and estimates based on the reduced data set $\hat{\psi}_{(U)}$. The likelihood distance is given by determining

$$LD_{(U)} = 2\{l(\hat{\psi}) - l(\hat{\psi}_{(U)})\} \quad (13)$$

$$RLD_{(U)} = 2\{l_R(\hat{\psi}) - l_R(\hat{\psi}_{(U)})\} \quad (14)$$

Let y_{mir} be the r th replicate measurement on the i th item by the m th method, where $m = 1, 2$, $i = 1, \dots, N$, and $r = 1, \dots, n_i$. When the design is balanced and there is no ambiguity we can set $n_i = n$. The LME model underpinning Roy's approach can be written

$$y_{mir} = \beta_0 + \beta_m + b_{mi} + \epsilon_{mir}. \quad (15)$$

Here β_0 and β_m are fixed-effect terms representing, respectively, a model intercept and an overall effect for method m . The β terms can be gathered together into (fixed effect) intercept terms $\alpha_m = \beta_0 + \beta_m$. The b_{1i} and b_{2i} terms are correlated random effect parameters having $E(b_{mi}) = 0$ with $\text{Var}(b_{mi}) = g_m^2$ and $\text{Cov}(b_{mi}, b_{m'i}) = g_{12}$. The random error term for each response is denoted ϵ_{mir} having $E(\epsilon_{mir}) = 0$, $\text{Var}(\epsilon_{mir}) = \sigma_m^2$, $\text{Cov}(b_{mir}, b_{m'ir}) = \sigma_{12}$, $\text{Cov}(\epsilon_{mir}, \epsilon_{m'ir'}) = 0$ and $\text{Cov}(\epsilon_{mir}, \epsilon_{m'ir'}) = 0$. Two methods of measurement are in complete agreement if the null hypotheses $H_1: \beta_1 = \beta_2$ and $H_2: \sigma_1^2 = \sigma_2^2$ and $H_3: g_1^2 = g_2^2$ hold simultaneously. Roy (2009) proposes a Bonferroni correction to control the familywise error rate for tests of $\{H_1, H_2, H_3\}$ and account for difficulties arising due to multiple testing. Let $\omega_m^2 = \sigma_m^2 + g_m^2$ represent the overall variability of method m . Roy also integrates H_2 and H_3 into a single testable hypothesis $H_4: \omega_1^2 = \omega_2^2$. CONCERNS?

? demonstrates how to implement a method comparison study further to model (1) using the SAS proc mixed package. Carstensen et al. (2008) demonstrates how to construct limits of agreement using SAS, STATA and R. In the case of SAS, the PROC MIXED procedure is used. Implementation in R is performed using the nlme package (?).

Carstensen et al. (2008) remarks that the implementation using R is quite "arcane".

As R is freely available, this paper demonstrates an implementation of Roy's model using R.

The R statistical software package is freely available.

The LME model is very easy to implement using PROC MIXED of SAS and the results are also easy to interpret. The SAS proc mixed procedure has very simple syntax.

As the required code to fit the models is complex, R code necessary to fit the models is provided.

A demonstration is provided on how to use the output to perform the tests, and to compute limits of agreement.

We assume the data are formatted as a dataset with four columns named:
meth, method of measurement, the number of methods being M , item, items (persons, samples) measured by each method, of which there are I , repl, replicate indicating repeated measurement of the same item by the same method, and y , the measurement.

5.1 Remarks on the Multivariate Normal Distribution

Diligence is required when considering the models. Carstensen specifies his models in terms of the univariate normal distribution. Roy's model is specified using the bivariate normal distribution. This gives rise to a key difference between the two models, in that a bivariate model accounts for covariance between the variables of interest. The multivariate normal distribution of a k -dimensional random vector $X = [X_1, X_2, \dots, X_k]$ can be written in the following notation:

$$X \sim \mathcal{N}(\mu, \Sigma),$$

or to make it explicitly known that X is k -dimensional,

$$X \sim \mathcal{N}_k(\mu, \Sigma).$$

with k -dimensional mean vector

$$\mu = [E[X_1], E[X_2], \dots, E[X_k]]$$

and $k \times k$ covariance matrix

$$\Sigma = [\text{Cov}[X_i, X_j]], \quad i = 1, 2, \dots, k; \quad j = 1, 2, \dots, k$$

1. Univariate Normal Distribution

$$X \sim \mathcal{N}(\mu, \sigma^2),$$

2. Bivariate Normal Distribution

(a)

$$X \sim \mathcal{N}_2(\mu, \Sigma),$$

(b)

$$\mu = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}.$$

6 Modelling Agreement with LME Models

Roy's uses an LME model approach to provide a set of formal tests for method comparison studies.

Four candidate models are fitted to the data.

These models are similar to one another, but for the imposition of equality constraints.

These tests are the pairwise comparison of candidate models, one formulated without constraints, the other with a constraint.

Roy's model uses fixed effects $\beta_0 + \beta_1$ and $\beta_0 + \beta_1$ to specify the mean of all observations by methods 1 and 2 respectively.

Roy adheres to Random Effect ideas in ANOVA

Roy treats items as a sample from a population.

Allocation of fixed effects and random effects are very different in each model

Carstensen's interest lies in the difference between the population from which they were drawn.

Carstensen's model is a mixed effects ANOVA.

$$Y_{mir} = \alpha_m + \mu_i + c_{mi} + e_{mir}, \quad c_{mi} \sim \tau_{\updownarrow}^{\Xi}, \quad e_{mir} \sim \sigma_{\updownarrow}^{\Xi},$$

This model includes a method by item interaction term.

Carstensen presents two models. One for the case where the replicates, and a second for when they are linked.

Carstensen's model does not take into account either between-item or within-item covariance between methods.

In the presented example, it is shown that Roy's LoAs are lower than those of Carstensen. Carstensen makes some interesting remarks in this regard.

The only slightly non-standard (meaning "not often used") feature is the differing residual variances between methods.

7 Cook's Distance

- For variance components γ : $CD(\gamma)_i$,
- For fixed effect parameters β : $CD(\beta)_i$,
- For random effect parameters \mathbf{u} : $CD(u)_i$,

- For linear functions of $\hat{\beta}$: $CD(\psi)_i$

It is also desirable to measure the influence of the case deletions on the covariance matrix of $\hat{\beta}$.

8 Cook's Distance

Cook (1977) greatly expanded the study of residuals and influence measures. Cook's Distance, denoted as $D_{(i)}$, is a well known diagnostic technique used in classical linear models, used as an overall measure of the combined impact of the i th case of all estimated regression coefficients. Cook's key observation was the effects of deleting each observation in turn could be calculated with little additional computation. That is to say, $D_{(i)}$ can be calculated without fitting a new regression coefficient each time an observation is deleted. Consequently deletion diagnostics have become an integral part of assessing linear models.

The focus of this analysis is related to the estimation of point estimates (i.e. regression coefficients). It must be pointed out that the effect on the precision of estimates is separate from the effect on the point estimates. Data points that have a small Cook's distance, for example, can still greatly affect hypothesis tests and confidence intervals, if their influence on the precision of the estimates is large.

As well as individual observations, Cook's distance can be used to analyse the influence of observations in subset U on a vector of parameter estimates (Cook, 1977).

$$\hat{e}_{i(U)} = y_i - x\hat{\beta}_{(U)} \quad (16)$$

$$\delta_{(U)} = \hat{\beta} - \hat{\beta}_{(U)} \quad (17)$$

- *The previous Section (Section 4) is a literary review of residual diagnostics and influence procedures for Linear Mixed Effects Models, drawing heavily on Schabenberger and Zewotir.*
- *Section 4 begins with an introduction to key topics in residual diagnostics, such as influence, leverage, outliers and Cook's distance. Other concepts such as DF-FITS and DFBETAs will be introduced briefly, mostly to explain why they are not particularly useful for the Method Comparison context, and therefore are not elaborated upon.*
- *In brief, Variable Selection is not applicable to Method Comparison Studies, in the commonly used context. Testing a rather simplistic specified model against one with more random effects terms is tractable, but this research question is of secondary importance.*

8.1 Matrix Notation for Case Deletion

For notational simplicity, $\mathbf{A}(i)$ denotes an $n \times m$ matrix \mathbf{A} with the i -th row removed, a_i denotes the i -th row of \mathbf{A} , and a_{ij} denotes the (i, j) -th element of \mathbf{A} .

9 Cook's Distance

- For variance components γ : $CD(\gamma)_i$,
- For fixed effect parameters β : $CD(\beta)_i$,
- For random effect parameters \mathbf{u} : $CD(u)_i$,
- For linear functions of $\hat{\beta}$: $CD(\psi)_i$

It is also desirable to measure the influence of the case deletions on the covariance matrix of $\hat{\beta}$.

10 Extension of Cook's Distance methodology to LME models

Cook's Distance is extended to LME models. For LME models, two formulations exist; a Cook's distance that examines the change in fixed fixed parameter estimates, and another that examines the change in random effects parameter estimates. The outcome of either Cook's distance is a scaled change in either β or θ .

Diagnostic methods for variance components are based on 'one-step' methods. Cook (1986) gives a completely general method for assessing the influence of local departures from assumptions in statistical models. For fixed effects parameter estimates in LME models, the Cook's distance can be extended to measure influence on these fixed effects.

$$CD_i(\beta) = \frac{(c_{ii} - r_{ii}) \times t_i^2}{r_{ii} \times p}$$

For random effect estimates, the Cook's distance is

$$CD_i(b) = g_{(i)}'(I_r + \text{var}(\hat{b})D)^{-2} \text{var}(\hat{b})g_{(i)}.$$

Large values for Cook's distance indicate observations for special attention.

Cook's Distance was extended from classical linear models to LME models. For linear mixed effects models, Cook's distance can be extended to model influence diagnostics by defining.

$$CD_{\beta i} = \frac{(\hat{\beta} - \hat{\beta}_{[i]})^T (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})(\hat{\beta} - \hat{\beta}_{[i]})}{p}$$

It is also desirable to measure the influence of the case deletions on the covariance matrix of $\hat{\beta}$.

11 Exention of Cook's Distance methodology to LME models

Cook's Distance is extended to LME models. For LME models, two formulations exist; a Cook's distance that examines the change in fixed fixed parameter estimates, and another that examines the change in random effects parameter estimates. The outcome of either Cook's distance is a scaled change in either β or θ .

Diagnostic methods for variance components are based on 'one-step' methods. Cook (1986) gives a completely general method for assessing the influence of local departures from assumptions in statistical models. For fixed effects parameter estimates in LME models, the Cook's distance can be extended to measure influence on these fixed effects.

$$CD_i(\beta) = \frac{(c_{ii} - r_{ii}) \times t_i^2}{r_{ii} \times p}$$

For random effect estimates, the Cook's distance is

$$CD_i(b) = g'_{(i)}(I_r + \text{var}(\hat{b})D)^{-2}\text{var}(\hat{b})g_{(i)}.$$

Large values for Cook's distance indicate observations for special attention.

Cook's Distance was extended from classical linear models to LME models. For linear mixed effects models, Cook's distance can be extended to model influence diagnostics by definining.

$$CD_{\beta i} = \frac{(\hat{\beta} - \hat{\beta}_{[i]})^T(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})(\hat{\beta} - \hat{\beta}_{[i]})}{p}$$

It is also desirable to measure the influence of the case deletions on the covariance matrix of $\hat{\beta}$.

12 Extension of Diagnostic Methods to LME models

When similar notions of statistical influence are applied to mixed models, things are more complicated. Removing data points affects fixed effects and covariance parameter estimates. Update formulas for *leave-one-out* estimates typically fail to account for changes in covariance parameters.

? noted the case deletion diagnostics techniques have not been applied to linear mixed effects models and seeks to develop methodologies in that respect. ? develops these techniques in the context of REML.

Christensen et al. (1992) noted the case deletion diagnostics techniques had not been applied to linear mixed effects models and seeks to develop methodologies in that respect. Christensen et al. (1992) develops these techniques in the context of REML.

Demidenko (2004) extends several regression diagnostic techniques commonly used in linear regression, such as leverage, infinitesimal influence, case deletion diagnostics, Cook's distance, and local influence to the linear mixed-effects model. In each case, the proposed new measure has a direct interpretation in terms of the effects on a parameter of interest, and reduces to the familiar linear regression measure when there are no random effects.

The new measures that are proposed by Demidenko (2004) are explicitly defined functions and do not require re-estimation of the model, especially for cluster deletion diagnostics. The basis for both the cluster deletion diagnostics and Cook's distance is a generalization of Miller's simple update formula for case deletion for linear models. Furthermore Demidenko (2004) shows how Pregibon's infinitesimal case deletion diagnostics is adapted to the linear mixed-effects model.

Demidenko (2004) proposes two likelihood-based diagnostics for identifying individuals that can influence the choice between two competing models.

13 Cook's Distance for LMEs

Diagnostic methods for fixed effects are generally analogues of methods used in classical linear models. Diagnostic methods for variance components are based on ‘one-step’ methods. Cook (1986) gives a completely general method for assessing the influence of local departures from assumptions in statistical models.

For fixed effects parameter estimates in LME models, the Cook's distance can be extended to measure influence on these fixed effects.

$$CD_i(\beta) = \frac{(c_{ii} - r_{ii}) \times t_i^2}{r_{ii} \times p}$$

For random effect estimates, the Cook's distance is

$$CD_i(b) = g'_{(i)}(I_r + \text{var}(\hat{b})D)^{-2}\text{var}(\hat{b})g_{(i)}.$$

Large values for Cook's distance indicate observations for special attention.

14 Cook's Distance for LMEs

Diagnostic methods for fixed effects are generally analogues of methods used in classical linear models. Diagnostic methods for variance components are based on ‘one-step’ methods. Cook (1986) gives a completely general method for assessing the influence of local departures from assumptions in statistical models.

For fixed effects parameter estimates in LME models, the Cook's distance can be extended to measure influence on these fixed effects.

$$CD_i(\beta) = \frac{(c_{ii} - r_{ii}) \times t_i^2}{r_{ii} \times p}$$

For random effect estimates, the Cook's distance is

$$CD_i(b) = g'_{(i)}(I_r + \text{var}(\hat{b})D)^{-2}\text{var}(\hat{b})g_{(i)}.$$

Large values for Cook's distance indicate observations for special attention.

14.0.1 Random Effects

A large value for $CD(u)_i$ indicates that the i -th observation is influential in predicting random effects.

14.0.2 linear functions

$CD(\psi)_i$ does not have to be calculated unless $CD(\beta)_i$ is large.

For LME models, two formulations exist; a Cook's distance that examines the change in fixed fixed parameter estimates, and another that examines the change in

random effects parameter estimates. The outcome of either Cook's distance is a scaled change in either β or θ .

14.0.3 linear functions

$CD(\psi)_i$ does not have to be calculated unless $CD(\beta)_i$ is large.

For LME models, two formulations exist; a Cook's distance that examines the change in fixed fixed parameter estimates, and another that examines the change in random effects parameter estimates. The outcome of either Cook's distance is a scaled change in either β or θ .

14.1 Change in the precision of estimates

The effect on the precision of estimates is separate from the effect on the point estimates. Data points that have a small Cook's distance, for example, can still greatly affect hypothesis tests and confidence intervals, if their influence on the precision of the estimates is large.

15 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 $\text{var}(\varepsilon) = V$. 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 $var(Z - \hat{Z}(Y))$. 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 .

16 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 $\text{var}(\varepsilon) = V$. 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 $\hat{e} = Y - X\hat{\beta}$

17 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 $\text{var}(Z - \lambda_z^t Y) + \gamma_z^t (X^t \lambda_z - A^t)$, 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.

$$\begin{pmatrix} V & X \\ X^t & 0 \end{pmatrix} \begin{pmatrix} \lambda_z \\ \gamma_z \end{pmatrix} = \begin{pmatrix} \text{cov}(Y, Z) \\ A^t \end{pmatrix} \quad (18)$$

If the inverse exists we have that

$$\begin{pmatrix} \lambda_z \\ \gamma_z \end{pmatrix} = \begin{pmatrix} V & X \\ X^t & 0 \end{pmatrix}^{-1} \begin{pmatrix} \text{cov}(Y, Z) \\ A^t \end{pmatrix} \quad (19)$$

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.

18 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.

$$\begin{pmatrix} V & X \\ X^t & 0 \end{pmatrix} \begin{pmatrix} \lambda_z \\ \gamma_z \end{pmatrix} = \begin{pmatrix} cov(Y, Z) \\ A^t \end{pmatrix} \quad (20)$$

If the inverse exists we have that

$$\begin{pmatrix} \lambda_z \\ \gamma_z \end{pmatrix} = \begin{pmatrix} V & X \\ X^t & 0 \end{pmatrix}^{-1} \begin{pmatrix} cov(Y, Z) \\ A^t \end{pmatrix} \quad (21)$$

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.

19 Simplifying GLS

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 .

20 The extended likelihood

The desire to have an entirely likelihood-based justification for estimates of random

```
\begin{eqnarray*}
\ell_h(\beta, \theta, b|y)
&= \displaystyle -\frac{1}{2} \left\{ \log|\Sigma| + (y - X\beta - Zb)' \Sigma^{-1} \right. \\
&\quad \left. + \log|D| + b' D^{-1} b \right\}.
\end{eqnarray*}
```

Given θ , differentiating with respect to β and b returns Henderson's

```
\subsubsection{The LME model as a general linear model}
Henderson's equations in (\ref{Henderson:Equations}) can be rewritten  $(T'W$ 
\[
\begin{bmatrix} \Delta \\ y_a \end{bmatrix} = \begin{bmatrix} \beta \\ b \end{bmatrix}
\begin{bmatrix} y \\ \psi \end{bmatrix}
\begin{bmatrix} T \\ X \text{ \& } Z \\ 0 \text{ \& } I \end{bmatrix}
\text{ \& } W = \begin{bmatrix} \Sigma \text{ \& } 0 \\ 0 \text{ \& } D \end{bmatrix}
\]
```

where \cite{Lee:Neld:Pawi:2006} describe $\psi = 0$ as quasi-data with mean $\mathbf{0}$

21 Cook's Distance

Cook (1977) greatly expanded the study of residuals and influence measures. Cook's Distance, denoted as $D_{(i)}$, is a well known diagnostic technique used in classical linear models, used as an overall measure of the combined impact of the i th case of all estimated regression coefficients. Cook's key observation was the effects of deleting each observation in turn could be calculated with little additional computation. That is to say, $D_{(i)}$ can be calculated without fitting a new regression coefficient each time an observation is deleted. Consequently deletion diagnostics have become an integral

part of assessing linear models.

The focus of this analysis is related to the estimation of point estimates (i.e. regression coefficients). It must be pointed out that the effect on the precision of estimates is separate from the effect on the point estimates. Data points that have a small Cook's distance, for example, can still greatly affect hypothesis tests and confidence intervals, if their influence on the precision of the estimates is large.

As well as individual observations, Cook's distance can be used to analyse the influence of observations in subset U on a vector of parameter estimates (Cook, 1977).

$$\hat{e}_{i(U)} = y_i - x\hat{\beta}_{(U)} \tag{22}$$

$$\delta_{(U)} = \hat{\beta} - \hat{\beta}_{(U)} \tag{23}$$

- *The previous Section (Section 4) is a literary review of residual diagnostics and influence procedures for Linear Mixed Effects Models, drawing heavily on Schabenberger and Zewotir.*
- *Section 4 begins with an introduction to key topics in residual diagnostics, such as influence, leverage, outliers and Cook's distance. Other concepts such as DF-FITS and DFBETAs will be introduced briefly, mostly to explain why they are not particularly useful for the Method Comparison context, and therefore are not elaborated upon.*
- *In brief, Variable Selection is not applicable to Method Comparison Studies, in the commonly used context. Testing a rather simplistic specified model against one with more random effects terms is tractable, but this research question is of secondary importance.*

22 Repeated measurements in LME models

In many statistical analyzes, the need to determine parameter estimates where multiple measurements are available on each of a set of variables often arises. Further to Lam et al. (1999), Hamlett et al. (2004) performs an analysis of the correlation of replicate measurements, for two variables of interest, using LME models.

Let y_{Aij} and y_{Bij} be the j th repeated observations of the variables of interest A and B taken on the i th subject. The number of repeated measurements for each variable may differ for each individual. Both variables are measured on each time points. Let n_i be the number of observations for each variable, hence $2 \times n_i$ observations in total.

It is assumed that the pair y_{Aij} and y_{Bij} follow a bivariate normal distribution.

$$\begin{pmatrix} y_{Aij} \\ y_{Bij} \end{pmatrix} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \text{ where } \boldsymbol{\mu} = \begin{pmatrix} \mu_A \\ \mu_B \end{pmatrix}$$

The matrix $\boldsymbol{\Sigma}$ represents the variance component matrix between response variables at a given time point j .

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_A^2 & \sigma_{AB} \\ \sigma_{AB} & \sigma_B^2 \end{pmatrix}$$

σ_A^2 is the variance of variable A , σ_B^2 is the variance of variable B and σ_{AB} is the covariance of the two variable. It is assumed that $\boldsymbol{\Sigma}$ does not depend on a particular time point, and is the same over all time points.

References

- Bland, J. and D. Altman (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research* 8(2), 135–160.
- Carstensen, B. (2004). Comparing and predicting between several methods of measurement. *Biostatistics* 5(3), 399–413.
- Carstensen, B., J. Simpson, and L. C. Gurrin (2008). Statistical models for assessing agreement in method comparison studies with replicate measurements. *The International Journal of Biostatistics* 4(1).
- Christensen, R., L. M. Pearson, and W. Johnson (1992). Case-deletion diagnostics for mixed models. *Technometrics* 34(1), 38–45.
- Cook, R. (1977). Detection of influential observations in linear regression. *Technometrics* 19, 15–18.
- Cook, R. (1986). Assessment of local influence. *Journal of the Royal Statistical Society. Series B (Methodological)* 48(2), 133–169.

- Demidenko, E. (2004). *Mixed Models: Theory And Application*. Dartmouth College: Wiley Interscience.
- Hamlett, A., L. Ryan, and R. Wolfinger (2004). On the use of PROC MIXED to estimate correlation in the presence of repeated measures. *Proceedings of the Statistics and Data Analysis Section, SAS Users Group International 198-229*, 1–7.
- Haslett, J. (1999). A simple derivation of deletion diagnostic results for the general linear model with correlated errors. *Journal of the Royal Statistical Society (Series B)* 61, 603–609.
- Haslett, J. and D. Dillane (2004). Application of ‘delete = replace’ to deletion diagnostics for variance component estimation. *Journal of the Royal Statistical Society (Series B)* 66, 131–143.
- Haslett, J. and K. Hayes (1998). Residuals for the linear model with general covariance structure. *Journal of the Royal Statistical Society (Series B)* 60, 201–215.
- Lam, M., K. Webb, and D. O’Donnell (1999). Correlation between two variables in repeated measurements. *American Statistical Association, Proceedings of the Biometric Session*, 213–218.
- McCullough, C. and S. Searle (2001). *Generalized , Linear and Mixed Models*. Wiley Interscience.
- Roy, A. (2009). An application of linear mixed effects model to assess the agreement between two methods with replicated observations. *Journal of Biopharmaceutical Statistics* 19, 150–173.