

Transfer Report

Kevin O'Brien

August 16, 2010

Contents

1	Introduction	2
1.1	Introduction	2
2	Proposal of Research	3
2.1	The Linear Mixed Effects Model	3
2.2	LME models in method comparison studies	5
2.3	Roy’s LME methodology for assessing agreement	6
2.4	Lai Shiao	9
2.5	Carstensen’s Mixed Models	12
2.6	Case Deletion Diagnostics for LME models	15
	Bibliography	17

Chapter 1

Introduction

1.1 Introduction

Chapter 2

Proposal of Research

2.1 The Linear Mixed Effects Model

The linear mixed effects model is given by

$$Y = X\beta + Zu + \epsilon \quad (2.1)$$

\mathbf{Y} is the vector of n observations, with dimension $n \times 1$. \mathbf{b} is a vector of fixed p effects, and has dimension $p \times 1$. It is composed of coefficients, with the first element being the population mean. \mathbf{X} is known as the design ‘matrix’, model matrix for fixed effects, and comprises 0s or 1s, depending on whether the relevant fixed effects have any effect on the observation is question. \mathbf{X} has dimension $n \times p$. \mathbf{e} is the vector of residuals with dimension $n \times 1$.

The random effects models can be specified similarly. \mathbf{Z} is known as the ‘model matrix for random effects’, and also comprises 0s or 1s. It has dimension $n \times q$. \mathbf{u} is a vector of random q effects, and has dimension $q \times 1$.

\mathbf{V} , the variance matrix of \mathbf{Y} , can be expressed as follows;

$$\mathbf{V} = \text{var}(\mathbf{Xb} + \mathbf{Zu} + \mathbf{e}) \quad (2.2)$$

$$\mathbf{V} = \text{var}(\mathbf{Xb}) + \text{var}(\mathbf{Zu}) + \text{var}(\mathbf{e}) \quad (2.3)$$

$\text{var}(\mathbf{X}\mathbf{b})$ is known to be zero. The variance of the random effects $\text{var}(\mathbf{Z}\mathbf{u})$ can be written as $Z\text{var}(\mathbf{u})Z^T$.

By letting $\text{var}(u) = G$ (i.e $\mathbf{u} \sim N(0, \mathbf{G})$), this becomes ZGZ^T . This specifies the covariance due to random effects. The residual covariance matrix $\text{var}(e)$ is denoted as R , ($\mathbf{e} \sim N(0, \mathbf{R})$). Residual are uncorrelated, hence \mathbf{R} is equivalent to $\sigma^2\mathbf{I}$, where \mathbf{I} is the identity matrix. The variance matrix \mathbf{V} can therefore be written as;

$$\mathbf{V} = ZGZ^T + \mathbf{R} \quad (2.4)$$

The best linear unbiased predictor (BLUP) is used to estimating random effects, i.e to derive \mathbf{u} . The best linear unbiased estimator (BLUE) is used to estimate the fixed effects, \mathbf{b} . They were formulated in a paper by Henderson et al. (1959), which provides the derivations of both. Inferences about fixed effects have come to be called ‘estimates’, whereas inferences about random effects have come to be called ‘predictions’. hence the naming of BLUP is to reinforce distinction between the two, but it is essentially the same principal involved in both cases (GK, 1991). The BLUE of \mathbf{b} , and the BLUP of \mathbf{u} can be shown to be;

$$\hat{\mathbf{b}} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y} \quad (2.5)$$

$$\hat{\mathbf{u}} = GZ^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \quad (2.6)$$

The practical application of both expressions requires that the variance components be known. An estimate for the variance components must be derived to either maximum likelihood (ML) or more commonly restricted maximum likelihood (REML).

Importantly calculations based on the above formulae require the calculation of the inverse of \mathbf{V} . In simple examples \mathbf{V}^{-1} is a straightforward calculation, but with higher dimensions it becomes a very complex calculation.

2.2 LME models in method comparison studies

Linear mixed effects (LME) models can facilitate greater understanding of the potential causes of bias and differences in precision between two sets of measurement. Lai and Shiao (2005) views the uses of linear mixed effects models as an expansion on the Bland-Altman methodology, rather than as a replacement. Carstensen et al. (2008) remarks that modern statistical computation, such as that used for LME models, greatly improve the efficiency of calculation compared to previous ‘by-hand’ methods. In this chapter various LME approaches to method comparison studies shall be examined.

2.3 Roy's LME methodology for assessing agreement

Roy (2009) proposes the use of LME models to perform a test on two methods of agreement to determine whether they can be used interchangeably.

Bivariate correlation coefficients have been shown to be of limited use in method comparison studies (Bland and Altman, 1986). However, recently correlation analysis has been developed to cope with repeated measurements, enhancing their potential usefulness. Roy incorporates the use of correlation into his methodology.

Roy's method considers two methods to be in agreement if three conditions are met.

- no significant bias, i.e. the difference between the two mean readings is not "statistically significant",
- high overall correlation coefficient,
- the agreement between the two methods by testing their repeatability coefficients.

The methodology uses a linear mixed effects regression fit using compound symmetry (CS) correlation structure on \mathbf{V} .

$$\Lambda = \frac{\max_{H_0} L}{\max_{H_1} L}$$

Roy (2009) considers the problem of assessing the agreement between two methods with replicate observations in a doubly multivariate set-up using linear mixed effects models.

Roy (2009) uses examples from Bland and Altman (1986) to be able to compare both types of analysis.

Roy (2009) proposes a LME based approach with Kronecker product covariance structure with doubly multivariate setup to assess the agreement between two methods. This method is designed such that the data may be unbalanced and with unequal numbers of replications for each subject.

The maximum likelihood estimate of the between-subject variance covariance matrix of two methods is given as D . The estimate for the within-subject variance covariance matrix is $\hat{\Sigma}$. The estimated overall variance covariance matrix ‘Block Ω_i ’ is the addition of \hat{D} and $\hat{\Sigma}$.

$$\text{Block } \Omega_i = \hat{D} + \hat{\Sigma} \quad (2.7)$$

For the the RV-IC comparison, \hat{D} is given by

$$\hat{D} = \begin{bmatrix} 1.6323 & 1.1427 \\ 1.1427 & 1.4498 \end{bmatrix} \quad (2.8)$$

The estimate for the within-subject variance covariance matrix is given by

$$\hat{\Sigma} = \begin{bmatrix} 0.1072 & 0.0372 \\ 0.0372 & 0.1379 \end{bmatrix} \quad (2.9)$$

The estimated overall variance covariance matrix for the the ‘RV vs IC’ comparison is given by

$$\text{Block}\Omega_i = \begin{bmatrix} 1.7396 & 1.1799 \\ 1.1799 & 1.5877 \end{bmatrix}. \quad (2.10)$$

The power of the likelihood ratio test may depends on specific sample size and the specific number of replications, and the author proposes simulation studies to examine this further.

2.4 Lai Shiao

Lai and Shiao (2005) use mixed models to determine the factors that affect the difference of two methods of measurement using the conventional formulation of linear mixed effects models.

If the parameter \mathbf{b} , and the variance components are not significantly different from zero, the conclusion that there is no inter-method bias can be drawn. If the fixed effects component contains only the intercept, and a simple correlation coefficient is used, then the estimate of the intercept in the model is the inter-method bias. Conversely the estimates for the fixed effects factors can advise the respective influences each factor has on the differences. It is possible to pre-specify different correlation structures of the variance components \mathbf{G} and \mathbf{R} .

Oxygen saturation is one of the most frequently measured variables in clinical nursing studies. ‘Fractional saturation’ (HbO_2) is considered to be the gold standard method of measurement, with ‘functional saturation’ (SO_2) being an alternative method. The method of examining the causes of differences between these two methods is applied to a clinical study conducted by ?. This experiment was conducted by 8 lab practitioners on blood samples, with varying levels of haemoglobin, from two donors. The samples have been in storage for varying periods (described by the variable ‘Bloodage’) and are categorized according to haemoglobin percentages(i.e 0%,20%,40%,60%,80%,100%). There are 625 observations in all.

Lai and Shiao (2005) fits two models on this data, with the lab technicians and the replicate measurements as the random effects in both models. The first model uses haemoglobin level as a fixed effects component. For the second model, blood age is added as a second fixed factor.

Single fixed effect

The first model fitted by Lai and Shiao (2005) takes the blood level as the sole fixed effect to be analyzed. The following coefficient estimates are estimated by ‘Proc Mixed’;

$$\text{fixed effects : } 2.5056 - 0.0263\text{Fhbperct}_{ijtl} \quad (2.11)$$

$$(\text{p-values : } = 0.0054, < 0.0001, < 0.0001)$$

$$\text{random effects : } u(\sigma^2 = 3.1826) + e_{ijtl}(\sigma_e^2 = 0.1525, \rho = 0.6978)$$

$$(\text{p-values : } = 0.8113, < 0.0001, < 0.0001)$$

With the intercept estimate being both non-zero and statistically significant ($p = 0.0054$), this models supports the presence inter-method bias is 2.5% in favour of SO_2 . Also, the negative value of the haemoglobin level coefficient indicate that differences will decrease by 0.0263% for every percentage increase in the haemoglobin .

In the random effects estimates, the variance due to the practitioners is 3.1826, indicating that there is a significant variation due to technicians ($p = 0.0311$) affecting the differences. The variance for the estimates is given as 0.1525, ($p < 0.0001$).

Two fixed effects

Blood age is added as a second fixed factor to the model, whereupon new estimates are calculated;

$$\text{fixed effects : } -0.2866 + 0.1072\text{Bloodage}_{ijtl} - 0.0264\text{Fhbperct}_{ijtl}$$

$$(\text{p-values : } = 0.8113, < 0.0001, < 0.0001)$$

$$\text{random effects : } u(\sigma^2 = 10.2346) + e_{ijtl}(\sigma_e^2 = 0.0920, \rho = 0.5577)$$

$$(\text{p-values : } = 0.0446, < 0.0001, < 0.0001) \quad (2.12)$$

With this extra fixed effect added to the model, the intercept term is no longer statistically significant. Therefore, with the presence of the second fixed factor, the model is no longer supporting the presence of inter-method bias. Furthermore, the second coefficient indicates that the blood age of the observation has a significant bearing on the size of the difference between both methods ($p < 0.0001$). Longer storage times for blood will lead to higher levels of particular blood factors such as MetHb and HbCO (due to the breakdown and oxidisation of the haemoglobin). Increased levels of MetHb and HbCO are concluded to be the cause of the differences. The coefficient for the haemoglobin level doesn't differ greatly from the single fixed factor model, and has a much smaller effect on the differences. The random effects estimates also indicate significant variation for the various technicians; 10.2346 with $p = 0.0446$.

Lai and Shiao (2005) demonstrates how that linear mixed effects models can be used to provide greater insight into the cause of the differences. Naturally the addition of further factors to the model provides for more insight into the behavior of the data.

2.5 Carstensen's Mixed Models

Carstensen (2004) proposes linear mixed effects models for deriving conversion calculations similar to Deming's regression, and for estimating variance components for measurements by different methods. The following model (in the authors own notation) is formulated as follows, where y_{mir} is the r th replicate measurement on subject i with method m .

$$y_{mir} = \alpha_m + \beta_m \mu_i + c_{mi} + e_{mir} \quad (e_{mi} \sim N(0, \sigma_m^2), c_{mi} \sim N(0, \tau_m^2)) \quad (2.13)$$

The intercept term α and the $\beta_m \mu_i$ term follow from Dunn (2002), expressing constant and proportional bias respectively , in the presence of a real value μ_i . c_{mi} is a interaction term to account for replicate, and e_{mir} is the residual associated with each observation. Since variances are specific to each method, this model can be fitted separately for each method.

The above formulation doesn't require the data set to be balanced. However, it does require a sufficient large number of replicates and measurements to overcome the problem of identifiability. The import of which is that more than two methods of measurement may be required to carry out the analysis. There is also the assumptions that observations of measurements by particular methods are exchangeable within subjects. (Exchangeability means that future samples from a population behaves like earlier samples).

Carstensen (2004) uses the above formula to predict observations for a specific individual i by method m ;

$$BLUP_{mir} = \hat{\alpha}_m + \hat{\beta}_m \mu_i + c_{mi} \quad (2.14)$$

. Under the assumption that the μ s are the true item values, this would be sufficient

to estimate parameters. When that assumption doesn't hold, regression techniques (known as updating techniques) can be used additionally to determine the estimates. The assumption of exchangeability can be unrealistic in certain situations. Carstensen (2004) provides an amended formulation which includes an extra interaction term ($d_{mr} \sim N(0, \omega_m^2)$) to account for this.

Carstensen et al. (2008) sets out a methodology of computing the limits of agreement based upon variance component estimates derived using linear mixed effects models. Measures of repeatability, a characteristic of individual methods of measurements, are also derived using this method.

Carstensen (2004) also advocates the use of linear mixed models in the study of method comparisons. The model is constructed to describe the relationship between a value of measurement and its real value. The non-replicate case is considered first, as it is the context of the Bland-Altman plots. This model assumes that inter-method bias is the only difference between the two methods. A measurement y_{mi} by method m on individual i is formulated as follows;

$$y_{mi} = \alpha_m + \mu_i + e_{mi} \quad (e_{mi} \sim N(0, \sigma_m^2)) \quad (2.15)$$

The differences are expressed as $d_i = y_{1i} - y_{2i}$. For the replicate case, an interaction term c is added to the model, with an associated variance component. All the random effects are assumed independent, and that all replicate measurements are assumed to be exchangeable within each method.

$$y_{mir} = \alpha_m + \mu_i + c_{mi} + e_{mir} \quad (e_{mi} \sim N(0, \sigma_m^2), c_{mi} \sim N(0, \tau_m^2)) \quad (2.16)$$

Carstensen et al. (2008) proposes a methodology to calculate prediction intervals in the presence of replicate measurements, overcoming problems associated with Bland-Altman methodology in this regard. It is not possible to estimate the interaction variance components τ_1^2 and τ_2^2 separately. Therefore it must be assumed that they are equal. The variance of the difference can be estimated as follows:

$$var(y_{1j} - y_{2j}) \quad (2.17)$$

2.6 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.

Christensen et al. (1992) 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.

$$V_{[i]}^{-1} = \Lambda_{[i]} - \frac{\lambda\lambda'}{\nu ii} \quad (2.18)$$

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

$$X'_{[i]} V_{[i]}^{-1} X_{[i]} = X' V^{-1} X - \frac{\hat{x}_i \hat{x}_i'}{s_i} \quad (2.19)$$

$$(X'_{[i]} V_{[i]}^{-1} X_{[i]})^{-1} = (X' V^{-1} X)^{-1} + \frac{(X' V^{-1} X)^{-1} \hat{x}_i \hat{x}_i' (X' V^{-1} X)^{-1}}{s_i - \bar{h}_i} \quad (2.20)$$

$$X'_{[i]} V_{[i]}^{-1} Y_{[i]} = X' V^{-1} Y - \frac{\hat{x}_i \hat{y}_i'}{s_i} \quad (2.21)$$

Schabenberger (2004) examines the use and implementation of influence measures in LME models.

Influence is understood to be the ability of a single or multiple data points, through their presences or absence in the data, to alter important aspects of the analysis, yield qualitatively different inferences, or violate assumptions of the statistical model (Schabenberger, 2004).

Outliers are the most noteworthy data points in an analysis, and an objective of influence analysis is how influential they are, and the manner in which they are influential.

Schabenberger (2004) describes a simple procedure for quantifying influence. Firstly a model should be fitted to the data, and estimates of the parameters should be ob-

tained. The second step is that either single or multiple data points, specifically outliers, should be omitted from the analysis, with the original parameter estimates being updated. This is known as ‘leave one out’ or ‘leave k out’ analysis. The final step of the procedure is comparing the sets of estimates computed from the entire and reduced data sets to determine whether the absence of observations changed the analysis.

A residual is the difference between an observed quantity and its estimated or predicted value. In LME models, there are two types of residuals, marginal residuals and conditional residuals. A marginal residual is the difference between the observed data and the estimated marginal mean. A conditional residual is the difference between the observed data and the predicted value of the observation. In a model without random effects, both sets of residuals coincide.

Schabenberger (2004) notes that it is not always possible to derive influence statistics necessary for comparing full- and reduced-data parameter estimates. Haslett and Dillane (2004) offers a 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 and Dillane (2004) 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.

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

$$LD_{(U)} = 2\{l(\hat{\psi}) - l(\hat{\psi}_{(U)})\} \quad (2.22)$$

$$RLD_{(U)} = 2\{l_R(\hat{\psi}) - l_R(\hat{\psi}_{(U)})\} \quad (2.23)$$

Bibliography

- Bland, J. and D. Altman (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet* *i*, 307–310.
- 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.
- Dunn, G. (2002). *Statistical Evaluation of Measurement Error* (Second ed.). Stanford: American Mathematical Society and Digital Press.
- GK, R. (1991). That blups are a good thing: The estimation of random effects. *Statistical Science* *6*(1), 15–32.
- 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.
- Henderson, C., O. Kempthorne, S. Searle, and C. von Krosigk (1959). The estimation of environmental and genetic trends from records subject to culling. *Biometrics* 15, 192–218.
- Lai, D. and S.-Y. P. K. Shiao (2005). Comparing two clinical measurements: a linear mixed model approach. *Journal of Applied Statistics* 32(8), 855–860.
- McCullough, C. and S. Searle (2001). *Generalized , Linear and Mixed Models*. Wiley Interscience.
- Roy, A. (2009). An application of the linear mixed effects model to ass the agreement between two methods with replicated observations. *Journal of Biopharmaceutical Statistics* 19, 150–173.
- Schabenberger, O. (2004). Mixed model influence diagnostics. 18929.