

## 0.1 Statement of the LME model

Further to a paper published by Laird and Ware in 1982, a linear mixed effects model is a linear model that combined fixed and random effect terms formulated as follows;

$$Y_i = X_i\beta + Z_ib_i + \epsilon_i$$

- $Y_i$  is the  $n \times 1$  response vector
- $X_i$  is the  $n \times p$  Model matrix for fixed effects
- $\beta$  is the  $p \times 1$  vector of fixed effects coefficients
- $Z_i$  is the  $n \times q$  Model matrix for random effects
- $b_i$  is the  $q \times 1$  vector of random effects coefficients, sometimes denoted as  $u_i$
- $\epsilon$  is the  $n \times 1$  vector of observation errors

## 0.2 The Linear Mixed Effects Model

The linear mixed effects model is given by

$$Y = X\beta + Zu + \epsilon \tag{1}$$

$\mathbf{Y}$  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 in 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;

$$\mathbf{V} = \text{var}(\mathbf{Xb} + \mathbf{Zu} + \mathbf{e}) \quad (2)$$

$$\mathbf{V} = \text{var}(\mathbf{Xb}) + \text{var}(\mathbf{Zu}) + \text{var}(\mathbf{e}) \quad (3)$$

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

By letting  $\text{var}(u) = G$  (i.e  $\mathbf{u} \sim N(0, \mathbf{G})$ ), this becomes  $\mathbf{ZGZ}^T$ . This specifies the covariance due to random effects. The residual covariance matrix  $\text{var}(e)$  is denoted as  $\mathbf{R}$ , ( $\mathbf{e} \sim N(0, \mathbf{R})$ ). 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;

$$\mathbf{V} = \mathbf{ZGZ}^T + \mathbf{R} \quad (4)$$

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;

$$\hat{\mathbf{b}} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y} \quad (5)$$

$$\hat{\mathbf{u}} = \mathbf{GZ}^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \quad (6)$$

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.

### 0.2.1 T

he classical model is based on measurements  $y_{mi}$  by method  $m = 1, 2$  on item  $i = 1, 2, \dots$

$$y_{mi} = \alpha_m + \mu_i + e_{mi}$$

$$e_{mi} \sim N(0, \sigma_m^2)$$

Even though the separate variances can not be identified, their sum can be estimated by the empirical variance of the differences.

Like wise the separate  $\alpha$  can not be estimated, only their difference can be estimated as  $\bar{D}$

# Chapter 1

## LME models for MCS

### 1.1 Statement of the LME model

Further to a paper published by Laird and Ware in 1982, a linear mixed effects model is a linear model that combined fixed and random effect terms formulated as follows;

$$Y_i = X_i\beta + Z_ib_i + \epsilon_i$$

- $Y_i$  is the  $n \times 1$  response vector
- $X_i$  is the  $n \times p$  Model matrix for fixed effects
- $\beta$  is the  $p \times 1$  vector of fixed effects coefficients
- $Z_i$  is the  $n \times q$  Model matrix for random effects
- $b_i$  is the  $q \times 1$  vector of random effects coefficients, sometimes denoted as  $u_i$
- $\epsilon$  is the  $n \times 1$  vector of observation errors

### 1.1.1 Bendix Carstensen’s data sets

Carstensen et al. (2008) describes the sampling method when discussing of a motivating example. Diabetes patients attending an outpatient clinic in Denmark have their  $HbA_{1c}$  levels routinely measured at every visit. Venous and Capillary blood samples were obtained from all patients appearing at the clinic over two days.

Samples were measured on four consecutive days on each machines, hence there are five analysis days. Carstensen notes that every machine was calibrated every day to the manufacturers guidelines.

### 1.1.2 Limits of agreement for Carstensen’s data

Carstensen et al. (2008) describes the calculation of the limits of agreement (with the inter-method bias implicit) for both data sets, based on his formulation;

$$\hat{\alpha}_1 - \hat{\alpha}_2 \pm 2\sqrt{2\hat{\tau}^2 + \hat{\sigma}_1^2 + \hat{\sigma}_2^2}.$$

For the ‘Fat’ data set, the inter-method bias is shown to be 0.045. The limits of agreement are  $(-0.23, 0.32)$

Carstensen demonstrates the use of the interaction term when computing the limits of agreement for the ‘Oximetry’ data set. When the interaction term is omitted, the limits of agreement are  $(-9.97, 14.81)$ . Carstensen advises the inclusion of the interaction term for linked replicates, and hence the limits of agreement are recomputed as  $(-12.18, 17.12)$ .

### 1.1.3 Limits of Agreement in LME models

Carstensen et al. (2008) uses LME models to determine the limits of agreement. Between-subject variation for method  $m$  is given by  $d_m^2$  and within-subject variation is given by  $\lambda_m^2$ . ? remarks that for two methods  $A$  and  $B$ , separate values of  $d_A^2$  and  $d_B^2$  cannot be estimated, only their average. Hence the assumption that  $d_x = d_y = d$  is necessary. The between-subject variability  $\mathbf{D}$  and within-subject variability  $\mathbf{\Lambda}$  can be presented in matrix form,

$$\mathbf{D} = \begin{pmatrix} d_A^2 & 0 \\ 0 & d_B^2 \end{pmatrix} = \begin{pmatrix} d^2 & 0 \\ 0 & d^2 \end{pmatrix}, \quad \mathbf{\Lambda} = \begin{pmatrix} \lambda_A^2 & 0 \\ 0 & \lambda_B^2 \end{pmatrix}.$$

The variance for method  $m$  is  $d_m^2 + \lambda_m^2$ . Limits of agreement are determined using the standard deviation of the case-wise differences between the sets of measurements by two methods  $A$  and  $B$ , given by

$$\text{var}(y_A - y_B) = 2d^2 + \lambda_A^2 + \lambda_B^2. \quad (1.1)$$

Importantly the covariance terms in both variability matrices are zero, and no covariance component is present.

Roy (2009) has demonstrated a methodology whereby  $d_A^2$  and  $d_B^2$  can be estimated separately. Also covariance terms are present in both  $\mathbf{D}$  and  $\mathbf{\Lambda}$ . Using Roy's methodology, the variance of the differences is

$$\text{var}(y_{iA} - y_{iB}) = d_A^2 + \lambda_B^2 + d_B^2 + \lambda_A^2 - 2(d_{AB} + \lambda_{AB}) \quad (1.2)$$

All of these terms are given or determinable in computer output. The limits of agreement can therefore be evaluated using

$$\bar{y}_A - \bar{y}_B \pm 1.96 \times \sqrt{\sigma_A^2 + \sigma_B^2 - 2(\sigma_{AB})}. \quad (1.3)$$

For Carstensen's 'fat' data, the limits of agreement computed using Roy's method are consistent with the estimates given by ?;  $0.044884 \pm 1.96 \times 0.1373979 = (-0.224, 0.314)$ .

### 1.1.4 Repeatability

Barnhart emphasizes the importance of repeatability as part of an overall method comparison study. Before there can be good agreement between two methods, a method must have good agreement with itself. The coefficient of repeatability, as proposed by Bland and Altman (1999) is an important feature of both Carstensen's and Roy's methodologies. The coefficient is calculated from the residual standard deviation (i.e.  $1.96 \times \sqrt{2} \times \sigma_m = 2.83\sigma_m$ ).

## 1.2 Hamlett and Lam

The methodology proposed by ? is largely based on Hamlett et al. (2004), which in turn follows on from Lam et al. (1999).



### 1.2.1 Roy's variability tests

Variability tests proposed by ? affords the opportunity to expand upon Carstensen's approach.

The first test allows of the comparison the begin-subject variability of two methods. Similarly, the second test assesses the within-subject variability of two methods. A third test is a test that compares the overall variability of the two methods.

The tests are implemented by fitting a specific LME model, and three variations thereof, to the data. These three variant models introduce equality constraints that act null hypothesis cases.

Other important aspects of the method comparison study are consequent. The limits of agreement are computed using the results of the first model.

### 1.2.2 Repeated Measurements

In cases where there are repeated measurements by each of the two methods on the same subjects , Bland Altman suggest calculating the mean for each method on each subject and use these pairs of means to compare the two methods.

The estimate of bias will be unaffected using this approach, but the estimate of the standard deviation of the differences will be too small, because of the reduction of the effect of repeated measurement error. Bland Altman propose a correction for this.

Carstensen attends to this issue also, adding that another approach would be to treat each repeated measurement separately.

### 1.3 Repeated Measurements

In cases where there are repeated measurements by each of the two methods on the same subjects, Bland Altman suggest calculating the mean for each method on each subject and use these pairs of means to compare the two methods. The estimate of bias will be unaffected using this approach, but the estimate of the standard deviation of the differences will be too small, because of the reduction of the effect of repeated measurement error. Bland Altman propose a correction for this. Carstensen attends to this issue also, adding that another approach would be to treat each repeated measurement separately.

### 1.4 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  $m$ .

$$y_{mir} = \alpha_m + \beta_m \mu_i + c_{mi} + e_{mir} \quad (e_{mi} \sim N(0, \sigma_m^2), c_{mi} \sim N(0, \tau_m^2)) \quad (1.4)$$

The intercept term  $\alpha$  and the  $\beta_m \mu_i$  term follow from Dunn (2002), expressing constant and proportional bias respectively, in the presence of a real value  $\mu_i$ .  $c_{mi}$  is an 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 are also 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  $m$ ;

$$BLUP_{mir} = \hat{\alpha}_m + \hat{\beta}_m \mu_i + c_{mi} \quad (1.5)$$

. Under the assumption that the  $\mu$ 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 ( $d_{mr}d_{mr} \sim N(0, \omega_m^2)$ ) to account for this.

? 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.

### 1.4.1 Using LME models to create Prediction Intervals

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;

$$y_{mi} = \alpha_m + \mu_i + e_{mi} \quad (e_{mi} \sim N(0, \sigma_m^2)) \quad (1.6)$$

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.

$$y_{mir} = \alpha_m + \mu_i + c_{mi} + e_{mir} \quad (e_{mi} \sim N(0, \sigma_m^2), c_{mi} \sim N(0, \tau_m^2)) \quad (1.7)$$

? 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  $\tau_1^2$  and  $\tau_2^2$  separately. Therefore it must be assumed that they are equal. The variance of the difference can be estimated as follows:

$$var(y_{1j} - y_{2j}) \quad (1.8)$$

### 1.4.2 Computation

Modern software packages can be used to fit models accordingly. The best linear unbiased predictor (BLUP) for a specific subject  $i$  measured with method  $m$  has the form  $BLUP_{mir} = \hat{\alpha}_m + \hat{\beta}_m \mu_i + c_{mi}$ , under the assumption that the  $\mu$ s are the true item values.

### 1.4.3 Repeated Measurements

In cases where there are repeated measurements by each of the two methods on the same subjects, Bland Altman suggest calculating the mean for each method on each subject and use these pairs of means to compare the two methods.

The estimate of bias will be unaffected using this approach, but the estimate of the standard deviation of the differences will be too small, because of the reduction of the effect of repeated measurement error. Bland Altman propose a correction for this.

Carstensen attends to this issue also, adding that another approach would be to treat each repeated measurement separately.

In this model, the variances of the random effects must depend on  $m$ , since the different methods do not necessarily measure on the same scale, and different methods

naturally must be assumed to have different variances. Carstensen (2004) attends to the issue of comparative variances.

? generalize this approach to account for situations where the distributions are not identical, which is commonly the case. The TDI is not consistent and may not preserve its asymptotic nominal level, and that the coverage probability approach of Lin et al. (2002) is overly conservative for moderate sample sizes. This methodology proposed by ? is a regression based approach that models the mean and the variance of differences as functions of observed values of the average of the paired measurements.

Maximum likelihood estimation is used to estimate the parameters. The REML estimation is not considered since it does not lead to a joint distribution of the estimates of fixed effects and random effects parameters, upon which the assessment of agreement is based.

## 1.5 Random Effects and MCS

The methodology comprises two calculations. The second calculation is for the standard deviation of means. Before the modified Bland and Altman method can be applied for repeated measurement data, a check of the assumption that the variance of the repeated measurements for each subject by each method is independent of the mean of the repeated measures. This can be done by plotting the within-subject standard deviation against the mean of each subject by each method. Mean Square deviation measures the total deviation of a

### 1.5.1 Random coefficient growth curve model

(Chincilli 1996) Random coefficient growth curve model, a special type of mixed model have been proposed a single measure of agreement for repeated measurements.

$$\mathbf{d} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e} \tag{1.9}$$

The distributional assumptions also require  $\mathbf{d}$  to  $\mathbf{N}$

## 1.6 Other Approaches

### 1.6.1 Random coefficient growth curve model

(Chincilli 1996) Random coefficient growth curve model, a special type of mixed model have been proposed a single measure of agreement for repeated measurements.

### 1.6.2 Marginal Modelling

(Diggle 2002) proposes the use of marginal models as an alternative to mixed models. Marginal models are appropriate when inferences about the mean response are of specific interest.

## 1.7 KP

Most residual covariance structures are design for one within-subject factor. However two or more may be present. For such cases, an appropriate approach would be the residual covariance structure using Kronecker product of the underlying within-subject factor specific covariances structure.

## 1.8 LME

Consistent with the conventions of mixed models, ? formulates the measurement  $y_{ij}$  from method  $i$  on individual  $j$  as follows;

$$y_{ij} = P_{ij}\theta + W_{ij}v_i + X_{ij}b_j + Z_{ij}u_j + \epsilon_{ij}, (j = 1, 2, i = 1, 2, \dots, n) \quad (1.10)$$

The design matrix  $P_{ij}$ , with its associated column vector  $\theta$ , specifies the fixed effects common to both methods. The fixed effect specific to the  $j$ th method is articulated by the design matrix  $W_{ij}$  and its column vector  $v_i$ . The random effects common to both methods is specified in the design matrix  $X_{ij}$ , with vector  $b_j$  whereas the random effects specific to the  $i$ th subject by the  $j$ th method is expressed by  $Z_{ij}$ , and vector  $u_j$ .

Noticeably this notation is not consistent with that described previously. The design matrices are specified so as to includes a fixed intercept for each method, and a random intercept for each individual. Additional assumptions must also be specified;

$$v_{ij} \sim N(0, \Sigma), \quad (1.11)$$

These vectors are assumed to be independent for different  $i$ s, and are also mutually independent. All Covariance matrices are positive definite. In the above model effects can be classed as those common to both methods, and those that vary with method. When considering differences, the effects common to both effectively cancel each other out. The differences of each pair of measurements can be specified as following;

$$d_{ij} = X_{ij}b_j + Z_{ij}u_j + \epsilon_{ij}, (j = 1, 2, i = 1, 2, \dots, n) \quad (1.12)$$

This formulation has seperate distributional assumption from the model stated previously.

This agreement covariate  $x$  is the key step in how this methodology assesses agreement.



# Bibliography

- 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).
- 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.
- 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.
- 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.
- 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.

Lin, L., A. Hedayat, B. Sinha, and M. Yang (2002). Statistical methods in assessing agreement: Models issues and tools. *Journal of American Statistical Association* 97, 257–270.

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.