

# Chapter 1

## Linear Mixed effects Models

### 1.1 Linear Mixed effects Models

A linear mixed effects (LME) model is a statistical model containing both fixed effects and random effects (random effects are also known as variance components). LME models are a generalization of the classical linear model, which contain fixed effects only. When the levels of factors are considered to be sampled from a population, and each level is not of particular interest, they are considered random quantities with associated variances. The effects of the levels, as described, are known as random effects. Random effects are represented by unobservable normally distributed random variables. Conversely fixed effects are considered non-random and the levels of each factor are of specific interest.

Fisher (1918) introduced variance components models for use in genetical studies. Whereas an estimate for variance must take a non-negative value, an individual variance component, i.e. a component of the overall variance, may be negative.

The framework has developed since, including contributions from Tippett (1931), who extend the use of variance components into linear models, and Eisenhart (1947),

who introduced the ‘mixed model’ terminology and formally distinguished between mixed and random effects models. Henderson (1950) devised a framework for deriving estimates for both the fixed effects and the random effects, using a set of equations that would become known as ‘mixed model equations’ or ‘Henderson’s equations’. LME methodology is further enhanced by Henderson’s later works (Henderson, 1953; Henderson et al., 1959, 1963, 1973, 1984). The key features of Henderson’s work provide the basis for the estimation techniques.

Hartley and Rao (1967) demonstrated that unique estimates of the variance components could be obtained using maximum likelihood methods. However these estimates are known to be biased ‘downwards’ (i.e. underestimated) , because of the assumption that the fixed estimates are known, rather than being estimated from the data. Patterson and Thompson (1971) produced an alternative set of estimates, known as the restricted maximum likelihood (REML) estimates, that do not require the fixed effects to be known. Thusly there is a distinction the REML estimates and the original estimates, now commonly referred to as ML estimates.

Laird and Ware (1982) provides a form of notation for notation for LME models that has since become the standard form, or the basis for more complex formulations. Due to computation complexity, linear mixed effects models have not seen widespread use until many well known statistical software applications began facilitating them. SAS Institute added PROC MIXED to its software suite in 1992 (Singer, 1998). Pinheiro and Bates (1994) described how to compute LME models in the **S-plus** environment.

Using Laird-Ware form, the LME model is commonly described in matrix form,

$$y = X\beta + Zb + \epsilon \tag{1.1}$$

where  $y$  is a vector of  $N$  observable random variables,  $\beta$  is a vector of  $p$  fixed effects,  $X$  and  $Z$  are  $N \times p$  and  $N \times q$  known matrices, and  $b$  and  $\epsilon$  are vectors of  $q$  and  $N$ ,

respectively, random effects such that  $E(b) = 0$ ,  $E(\epsilon) = 0$  and

$$\text{var} \begin{pmatrix} b \\ \epsilon \end{pmatrix} = \begin{pmatrix} D & 0 \\ 0 & \Sigma \end{pmatrix}$$

where  $D$  and  $\Sigma$  are positive definite matrices parameterized by an unknown variance component parameter vector  $\theta$ . The variance-covariance matrix for the vector of observations  $y$  is given by  $V = ZDZ' + \Sigma$ . This implies  $y \sim (X\beta, V) = (X\beta, ZDZ' + \Sigma)$ . It is worth noting that  $V$  is an  $n \times n$  matrix, as the dimensionality becomes relevant later on. The notation provided here is generic, and will be adapted to accord with complex formulations that will be encountered in due course.

## 1.2 Linear mixed effects models

These models are used when there are both fixed and random effects that need to be incorporated into a model.

Fixed effects usually correspond to experimental treatments for which one has data for the entire population of samples corresponding to that treatment.

Random effects, on the other hand, are assigned in the case where we have measurements on a group of samples, and those samples are taken from some larger sample pool, and are presumed to be representative.

As such, linear mixed effects models treat the error for fixed effects differently than the error for random effects.

## 1.3 Limits of agreement in LME models

Limits of agreement are used extensively for assessing agreement, because they are intuitive and easy to use. Necessarily their prevalence in literature has meant that

they are now the best known measurement for agreement, and therefore any newer methodology would benefit by making reference to them.

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  $\lambda_m^2$ . Carstensen et al. (2008) 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

$$\text{var}(y_A - y_B) = 2d^2 + \lambda_A^2 + \lambda_B^2. \quad (1.2)$$

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

Carstensen et al. (2008) presents a data set ‘fat’, which is a comparison of measurements of subcutaneous fat by two observers at the Steno Diabetes Center, Copenhagen. Measurements are in millimeters (mm). Each person is measured three times by each observer. The observations are considered to be ‘true’ replicates.

A linear mixed effects model is formulated, and implementation through several software packages is demonstrated. All of the necessary terms are presented in the computer output. The limits of agreement are therefore,

$$0.0449 \pm 1.96 \times \sqrt{2 \times 0.0596^2 + 0.0772^2 + 0.0724^2} = (-0.220, 0.309). \quad (1.3)$$

Roy (2009a) 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

$$\text{var}(y_{iA} - y_{iB}) = d_A^2 + \lambda_B^2 + d_A^2 + \lambda_B^2 - 2(d_{AB} + \lambda_{AB}) \quad (1.4)$$

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

$$\bar{y}_A - \bar{y}_B \pm 1.96 \times \sqrt{\sigma_A^2 + \sigma_B^2 - 2(\sigma_{AB})}. \quad (1.5)$$

For Carstensen's 'fat' data, the limits of agreement computed using Roy's method are consistent with the estimates given by Carstensen et al. (2008);  $0.044884 \pm 1.96 \times 0.1373979 = (-0.224, 0.314)$ .

### 1.3.1 Linked replicates

Carstensen et al. (2008) proposes the addition of an random effects term to their model when the replicates are linked. This term is used to describe the 'item by replicate' interaction, which is independent of the methods. This interaction is a source of variability independent of the methods. Therefore failure to account for it will result in variability being wrongly attributed to the methods.

Carstensen et al. (2008) introduces a second data set; the oximetry study. This study done at the Royal Childrens Hospital in Melbourne to assess the agreement between co-oximetry and pulse oximetry in small babies.

In most cases, measurements were taken by both method at three different times. In some cases there are either one or two pairs of measurements, hence the data is unbalanced. Carstensen et al. (2008) describes many of the children as being very sick, and with very low oxygen saturations levels. Therefore it must be assumed that a biological change can occur in interim periods, and measurements are not true replicates.

Carstensen et al. (2008) demonstrate the necessity of accounting for linked replicates by comparing the limits of agreement from the ‘oximetry’ data set using a model with the additional term, and one without. When the interaction is accounted for the limits of agreement are (-9.62,14.56). When the interaction is not accounted for, the limits of agreement are (-11.88,16.83). It is shown that the failure to include this additional term results in an over-estimation of the standard deviations of differences.

Limits of agreement are determined using Roy’s methodology, without adding any additional terms, are found to be consistent with the ‘interaction’ model;  $(-9.562, 14.504)$ . Roy’s methodology assumes that replicates are linked. However, following Carstensen’s example, an additional interaction term is added to the implementation of Roy’s model to assess the effect, the limits of agreement estimates do not change. However there is a conspicuous difference in within-subject matrices of Roy’s model and the modified model (denoted 1 and 2 respectively);

$$\hat{\mathbf{\Lambda}}_1 = \begin{pmatrix} 16.61 & 11.67 \\ 11.67 & 27.65 \end{pmatrix} \quad \hat{\mathbf{\Lambda}}_2 = \begin{pmatrix} 7.55 & 2.60 \\ 2.60 & 18.59 \end{pmatrix}. \quad (1.6)$$

(The variance of the additional random effect in model 2 is 3.01.)

Akaike (1974) introduces the Akaike information criterion (*AIC*), a model selection tool based on the likelihood function. Given a data set, candidate models are ranked according to their AIC values, with the model having the lowest AIC being considered the best fit. Two candidate models can be said to be equally good if there is a difference of less than 2 in their AIC values.

The Akaike information criterion (AIC) for both models are  $AIC_1 = 2304.226$  and  $AIC_2 = 2306.226$ , indicating little difference in models. The AIC values for the Carstensen ‘unlinked’ and ‘linked’ models are 1994.66 and 1955.48 respectively, indicating an improvement by adding the interaction term.

The  $\hat{\mathbf{\Lambda}}$  matrices are informative as to the difference between Carstensen’s unlinked

and linked models. For the oximetry data, the covariance terms (given above as 11.67 and 2.6 respectively ) are of similar magnitudes to the variance terms. Conversely for the ‘fat’ data the covariance term ( $-0.00032$ ) is negligible. When the interaction term is added to the model, the covariance term remains negligible. (For the ‘fat’ data, the difference in AIC values is also approximately 2).

To conclude, Carstensen’s models provided a rigorous way to determine limits of agreement, but don’t provide for the computation of  $\hat{D}$  and  $\hat{\Lambda}$ . Therefore the test’s proposed by Roy (2009a) can not be implemented. Conversely, accurate limits of agreement as determined by Carstensen’s model may also be found using Roy’s method. Addition of the interaction term erodes the capability of Roy’s methodology to compare candidate models, and therefore shall not be adopted.

Finally, to complement the blood pressure (i.e.‘J vs S’) method comparison from the previous section (i.e.‘J vs S’), the limits of agreement are  $15.62 \pm 1.96 \times 20.33 = (-24.22, 55.46)$ .)

## 1.4 Introduction to Mixed Models

All models are characterized by the mean  $\alpha$  and the error terms. In addition to these terms, any model described so far will have either random effects terms or fixed effects terms and accordingly are referred to as random or fixed models. Models that have both fixed effects terms and random effects terms are known as 'mixed effects models'. Once the theory underlying fixed and random effects models has been fully understood, the progression to understanding mixed models is very simple.

Elaborating on the original mice litter example, the six litters by each mouse were fed according to three different dietary treatments (Searle, 1997). Therefore a fixed effect  $\phi_j$  has been added to the model, which is now formulated as follows;

$$y_{ij} = \mu + \delta_i + \phi_j + \gamma_{ij} + \epsilon_{ijk} \quad (1.7)$$

As before, an interaction effect  $\gamma_{ij}$  must also be added to the model. In cases where the interaction term describes the combined effect of fixed and random components, it should be treated as random effect. The variance of the above model is composed of the  $\sigma_\delta^2$ ,  $\sigma_\gamma^2$  and  $\sigma_\epsilon^2$ .

It may be shown that the interaction factors make no contribution to the outcome, i.e.  $\gamma_{ij}$  is consistently calculated as zero. Considering the skin tumour example, a person's age would bear no relation to their gender and hence there would be plausible interaction between the two factors. Indeed, in keeping with the 'Law of Parsimony', factors should be specified such that each would convey separate information. However, interaction terms are extant when the model specifies repeated observations, as there is necessarily a relationship between observations from the same subject. Importantly, interaction effects, being random effects, are attended by variance component terms and therefore also contribute to the overall variance of the model.

Searle (1997) gives a mixed effects model formulation for the Grubbs artillery study.



$y_{ij}$  is the muzzle velocity of the  $i$ th shell, as measured by the  $j$ th chronometer.

$$y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij} \quad (1.8)$$

In this formulation  $\alpha_i$  is the random effect of round  $i$ , and the fixed effect component  $\beta_j$  is the bias in chronometer  $j$ . (Also, no interaction term is used).

## 1.5 Linear Mixed effects Models

A linear mixed effects (LME) model is a statistical model containing both fixed effects and random effects (random effects are also known as variance components). LME models are a generalization of the classical linear model, which contain fixed effects only. When the levels of factors are considered to be sampled from a population, and each level is not of particular interest, they are considered random quantities with associated variances. The effects of the levels, as described, are known as random effects. Random effects are represented by unobservable normally distributed random variables. Conversely fixed effects are considered non-random and the levels of each factor are of specific interest.

Fisher (1918) introduced variance components models for use in genetical studies. Whereas an estimate for variance must take a non-negative value, an individual variance component, i.e. a component of the overall variance, may be negative.

The methodology has developed since, including contributions from Tippett (1931), who extend the use of variance components into linear models, and Eisenhart (1947), who introduced the ‘mixed model’ terminology and formally distinguished between mixed and random effects models. Henderson (1950) devised a methodology for deriving estimates for both the fixed effects and the random effects, using a set of equations that would become known as ‘mixed model equations’ or ‘Henderson’s equations’. LME

methodology is further enhanced by Henderson's later works (Henderson, 1953; Henderson et al., 1959, 1963, 1973, 1984). The key features of Henderson's work provide the basis for the estimation techniques.

Hartley and Rao (1967) demonstrated that unique estimates of the variance components could be obtained using maximum likelihood methods. However these estimates are known to be biased 'downwards' (i.e. underestimated) , because of the assumption that the fixed estimates are known, rather than being estimated from the data. Patterson and Thompson (1971) produced an alternative set of estimates, known as the restricted maximum likelihood (REML) estimates, that do not require the fixed effects to be known. Thusly there is a distinction the REML estimates and the original estimates, now commonly referred to as ML estimates.

Laird and Ware (1982) provides a form of notation for notation for LME models that has since become the standard form, or the basis for more complex formulations. Due to computation complexity, linear mixed effects models have not seen widespread use until many well known statistical software applications began facilitating them. SAS Institute added PROC MIXED to its software suite in 1992 (Singer, 1998). Pinheiro and Bates (1994) described how to compute LME models in the **S-plus** environment.

Using Laird-Ware form, the LME model is commonly described in matrix form,

$$y = X\beta + Zb + \epsilon \tag{1.9}$$

where  $y$  is a vector of  $N$  observable random variables,  $\beta$  is a vector of  $p$  fixed effects,  $X$  and  $Z$  are  $N \times p$  and  $N \times q$  known matrices, and  $b$  and  $\epsilon$  are vectors of  $q$  and  $N$ , respectively, random effects such that  $E(b) = 0$ ,  $E(\epsilon) = 0$  and where  $D$  and  $\Sigma$  are positive definite matrices parameterized by an unknown variance component parameter vector  $\theta$ . The variance-covariance matrix for the vector of observations  $y$  is given by  $V = ZDZ' + \Sigma$ . This implies  $y \sim (X\beta, V) = (X\beta, ZDZ' + \Sigma)$ . It is worth noting that

$V$  is an  $n \times n$  matrix, as the dimensionality becomes relevant later on. The notation provided here is generic, and will be adapted to accord with complex formulations that will be encountered in due course.

### 1.5.1 Formulation of the response vector

Information of individual  $i$  is recorded in a response vector  $\mathbf{y}_i$ . The response vector is constructed by stacking the response of the 2 responses at the first instance, then the 2 responses at the second instance, and so on. Therefore the response vector is a  $2n_i \times 1$  column vector. The covariance matrix of  $\mathbf{y}_i$  is a  $2n_i \times 2n_i$  positive definite matrix  $\mathbf{\Omega}_i$ .

Consider the case where three measurements are taken by both methods  $A$  and  $B$ ,  $\mathbf{y}_i$  is a  $6 \times 1$  random vector describing the  $i$ th subject.

$$\mathbf{y}_i = (y_i^{A1}, y_i^{B1}, y_i^{A2}, y_i^{B2}, y_i^{A3}, y_i^{B3})'$$

The response vector  $\mathbf{y}_i$  can be formulated as an LME model according to Laird-Ware form.

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i$$

$$\mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{D})$$

$$\boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_i)$$

Information on the fixed effects are contained in a three dimensional vector  $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$ . For computational purposes  $\beta_2$  is conventionally set to zero. Consequently  $\boldsymbol{\beta}$  is the solutions of the means of the two methods, i.e.  $E(\mathbf{y}_i) = \mathbf{X}_i\boldsymbol{\beta}$ . The variance covariance matrix  $\mathbf{D}$  is a general  $2 \times 2$  matrix, while  $\mathbf{R}_i$  is a  $2n_i \times 2n_i$  matrix.

### 1.5.2 Decomposition of the response covariance matrix

The variance covariance structure can be re-expressed in the following form,

$$\text{Cov}(y_i) = \mathbf{\Omega}_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i' + \mathbf{R}_i.$$

$\mathbf{R}_i$  can be shown to be the Kronecker product of a correlation matrix  $\mathbf{V}$  and  $\mathbf{\Lambda}$ . The correlation matrix  $\mathbf{V}$  of the repeated measures on a given response variable is assumed to be the same for all response variables. Both Hamlett et al. (2004) and Lam et al. (1999) use the identity matrix, with dimensions  $n_i \times n_i$  as the formulation for  $\mathbf{V}$ . Roy (2009b) remarks that, with repeated measures, the response for each subject is correlated for each variable, and that such correlation must be taken into account in order to produce a valid inference on correlation estimates. ? proposes various correlation structures may be assumed for repeated measure correlations, such as the compound symmetry and autoregressive structures, as alternative to the identity matrix.

However, for the purposes of method comparison studies, the necessary estimates are currently only determinable when the identity matrix is specified, and the results in Roy (2009b) indicate its use.

For the response vector described, Hamlett et al. (2004) presents a detailed covariance matrix. A brief summary shall be presented here only. The overall variance matrix is a  $6 \times 6$  matrix composed of two types of  $2 \times 2$  blocks. Each block represents one separate time of measurement.

$$\mathbf{\Omega}_i = \begin{pmatrix} \mathbf{\Sigma} & \mathbf{D} & \mathbf{D} \\ \mathbf{D} & \mathbf{\Sigma} & \mathbf{D} \\ \mathbf{D} & \mathbf{D} & \mathbf{\Sigma} \end{pmatrix}$$

The diagonal blocks are  $\mathbf{\Sigma}$ , as described previously. The  $2 \times 2$  block diagonal matrix

in  $\Omega$  gives  $\Sigma$ .  $\Sigma$  is the sum of the between-subject variability  $D$  and the within subject variability  $\Lambda$ .

$\Omega_i$  can be expressed as

$$\Omega_i = Z_i D Z_i' + (I_{n_i} \otimes \Lambda).$$

The notation  $\text{dim}_{n_i}$  means an  $n_i \times n_i$  diagonal block.

## 1.6 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.

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

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

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

$\sigma_A^2$  is the variance of variable  $A$ ,  $\sigma_B^2$  is the variance of variable  $B$  and  $\sigma_{AB}$  is the

covariance of the two variable. It is assumed that  $\Sigma$  does not depend on a particular time point, and is the same over all time points.

### 1.6.1 Formulation of the Response Vector

Information of individual  $i$  is recorded in a response vector  $\mathbf{y}_i$ . The response vector is constructed by stacking the response of the 2 responses at the first instance, then the 2 responses at the second instance, and so on. Therefore the response vector is a  $2n_i \times 1$  column vector. The covariance matrix of  $\mathbf{y}_i$  is a  $2n_i \times 2n_i$  positive definite matrix  $\Omega_i$ .

Consider the case where three measurements are taken by both methods  $A$  and  $B$ ,  $\mathbf{y}_i$  is a  $6 \times 1$  random vector describing the  $i$ th subject.

$$\mathbf{y}_i = (y_i^{A1}, y_i^{B1}, y_i^{A2}, y_i^{B2}, y_i^{A3}, y_i^{B3})'$$

The response vector  $\mathbf{y}_i$  can be formulated as an LME model according to Laird-Ware form.

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i$$

$$\mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{D})$$

$$\boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_i)$$

Information on the fixed effects are contained in a three dimensional vector  $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$ . For computational purposes  $\beta_2$  is conventionally set to zero. Consequently  $\boldsymbol{\beta}$  is the solutions of the means of the two methods, i.e.  $E(\mathbf{y}_i) = \mathbf{X}_i\boldsymbol{\beta}$ . The variance covariance matrix  $\mathbf{D}$  is a general  $2 \times 2$  matrix, while  $\mathbf{R}_i$  is a  $2n_i \times 2n_i$  matrix.

### 1.6.2 Decomposition of the response covariance matrix

The variance covariance structure can be re-expressed in the following form,

$$\text{Cov}(\mathbf{y}_i) = \Omega_i = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i' + \mathbf{R}_i.$$

$\mathbf{R}_i$  can be shown to be the Kronecker product of a correlation matrix  $\mathbf{V}$  and  $\mathbf{\Lambda}$ . The correlation matrix  $\mathbf{V}$  of the repeated measures on a given response variable is assumed to be the same for all response variables. Both Hamlett et al. (2004) and Lam et al. (1999) use the identity matrix, with dimensions  $n_i \times n_i$  as the formulation for  $\mathbf{V}$ . Roy (2009a) remarks that, with repeated measures, the response for each subject is correlated for each variable, and that such correlation must be taken into account in order to produce a valid inference on correlation estimates. Roy (2006) proposes various correlation structures may be assumed for repeated measure correlations, such as the compound symmetry and autoregressive structures, as alternative to the identity matrix.

However, for the purposes of method comparison studies, the necessary estimates are currently only determinable when the identity matrix is specified, and the results in Roy (2009a) indicate its use.

For the response vector described, Hamlett et al. (2004) presents a detailed covariance matrix. A brief summary shall be presented here only. The overall variance matrix is a  $6 \times 6$  matrix composed of two types of  $2 \times 2$  blocks. Each block represents one separate time of measurement.

$$\mathbf{\Omega}_i = \begin{pmatrix} \mathbf{\Sigma} & \mathbf{D} & \mathbf{D} \\ \mathbf{D} & \mathbf{\Sigma} & \mathbf{D} \\ \mathbf{D} & \mathbf{D} & \mathbf{\Sigma} \end{pmatrix}$$

The diagonal blocks are  $\mathbf{\Sigma}$ , as described previously. The  $2 \times 2$  block diagonal matrix in  $\mathbf{\Omega}$  gives  $\mathbf{\Sigma}$ .  $\mathbf{\Sigma}$  is the sum of the between-subject variability  $\mathbf{D}$  and the within subject variability  $\mathbf{\Lambda}$ .

$\mathbf{\Omega}_i$  can be expressed as

$$\mathbf{\Omega}_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i' + (\mathbf{I}_{n_i} \otimes \mathbf{\Lambda}).$$

The notation  $\text{dim}_{n_i}$  means an  $n_i \times n_i$  diagonal block.

### 1.6.3 Correlation terms

Hamlett et al. (2004) demonstrated how the between-subject and within subject variabilities can be expressed in terms of correlation terms.

$$\mathbf{D} = \begin{pmatrix} \sigma_A^2 \rho_A & \sigma_A \sigma_B \rho_{AB} \delta \\ \sigma_A \sigma_B \rho_{AB} \delta & \sigma_B^2 \rho_B \end{pmatrix}$$

$$\mathbf{\Lambda} = \begin{pmatrix} \sigma_A^2 (1 - \rho_A) & \sigma_{AB} (1 - \delta) \\ \sigma_{AB} (1 - \delta) & \sigma_B^2 (1 - \rho_B) \end{pmatrix}.$$

$\rho_A$  describe the correlations of measurements made by the method  $A$  at different times. Similarly  $\rho_B$  describe the correlation of measurements made by the method  $B$  at different times. Correlations among repeated measures within the same method are known as intra-class correlation coefficients.  $\rho_{AB}$  describes the correlation of measurements taken at the same same time by both methods. The coefficient  $\delta$  is added for when the measurements are taken at different times, and is a constant of less than 1 for linked replicates. This is based on the assumption that linked replicates measurements taken at the same time would have greater correlation than those taken at different times. For unlinked replicates  $\delta$  is simply 1. Hamlett et al. (2004) provides a useful graphical depiction of the role of each correlation coefficients.



## 1.7 Extended LME model

The extended single level LME model relaxes the independence assumption, allowing heteroscedastic and correlated within group errors.

$$\epsilon_i = \mathcal{N}(0, \sigma^2 \Lambda_i) \quad (1.10)$$

$\Lambda_i$  are positive definite matrices.  $\sigma^2$  is factored out of the matrix for computational reasons.

## 1.8 Using LME for method comparison

Due to the prevalence of modern statistical software, Carstensen et al. (2008) advocates the adoption of computer based approaches, such as LME models, to method comparison studies. Carstensen et al. (2008) remarks upon ‘by-hand’ approaches advocated in Bland and Altman (1999) discouragingly, describing them as tedious, unnecessary and ‘outdated’. Rather than using the ‘by hand’ methods, estimates for required LME parameters can be read directly from program output. Furthermore, using computer approaches removes constraints associated with ‘by-hand’ approaches, such as the need for the design to be perfectly balanced.

### 1.8.1 Roy’s Approach

For the purposes of comparing two methods of measurement, Roy (2009a) presents a framework that utilizes linear mixed effects model. This methodology provides for the formal testing of inter-method bias, between-subject variability and within-subject variability of two methods. The formulation contains a Kronecker product covariance structure in a doubly multivariate setup. By doubly multivariate set up, Roy means

that the information on each patient or item is multivariate at two levels, the number of methods and number of replicated measurements. Further to Lam et al. (1999), it is assumed that the replicates are linked over time. However it is easy to modify to the unlinked case.

Roy (2009a) sets out three criteria for two methods to be considered in agreement. Firstly that there be no significant bias. Second that there is no difference in the between-subject variabilities, and lastly that there is no significant difference in the within-subject variabilities. Roy further proposes examination of the the overall variability by considering the second and third criteria be examined jointly. Should both the second and third criteria be fulfilled, then the overall variabilities of both methods would be equal.

A formal test for inter-method bias can be implemented by examining the fixed effects of the model. This is common to well known classical linear model methodologies. The null hypotheses, that both methods have the same mean, which is tested against the alternative hypothesis, that both methods have different means. The inter-method bias and necessary  $t$ -value and  $p$ -value are presented in computer output. A decision on whether the first of Roy's criteria is fulfilled can be based on these values.

Importantly Roy (2009a) further proposes a series of three tests on the variance components of an LME model, which allow decisions on the second and third of Roy's criteria. For these tests, four candidate LME models are constructed. The differences in the models are specifically in how the the  $D$  and  $\Lambda$  matrices are constructed, using either an unstructured form or a compound symmetry form. To illustrate these differences, consider a generic matrix  $A$ ,

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

A symmetric matrix allows the diagonal terms  $a_{11}$  and  $a_{22}$  to differ. The compound symmetry structure requires that both of these terms be equal, i.e  $a_{11} = a_{22}$ .

The first model acts as an alternative hypothesis to be compared against each of three other models, acting as null hypothesis models, successively. The models are compared using the likelihood ratio test. Likelihood ratio tests are a class of tests based on the comparison of the values of the likelihood functions of two candidate models.

### 1.8.2 Correlation

In addition to the variability tests, Roy advises that it is preferable that a correlation of greater than 0.82 exist for two methods to be considered interchangeable. However if two methods fulfil all the other conditions for agreement, failure to comply with this one can be overlooked. Indeed Roy demonstrates that placing undue importance to it can lead to incorrect conclusions. Roy (2009a) remarks that current computer implementations only gives overall correlation coefficients, but not their variances. Consequently it is not possible to carry out inferences based on all overall correlation coefficients.

### 1.8.3 Variability test 1

The first test determines whether or not both methods  $A$  and  $B$  have the same between-subject variability, further to the second of Roy's criteria.

$$H_0 : d_A = d_B$$

$$H_A : d_A \neq d_B$$

This test is facilitated by constructing a model specifying a symmetric form for  $D$  (i.e. the alternative model) and comparing it with a model that has compound symmetric

form for  $D$  (i.e. the null model). For this test  $\hat{\mathbf{\Lambda}}$  has a symmetric form for both models, and will be the same for both.

#### 1.8.4 Variability test 2

This test determines whether or not both methods  $A$  and  $B$  have the same within-subject variability, thus enabling a decision on the third of Roy's criteria.

$$H_0 : \lambda_A = \lambda_B$$

$$H_A : \lambda_A = \lambda_B$$

This model is performed in the same manner as the first test, only reversing the roles of  $\hat{\mathbf{D}}$  and  $\hat{\mathbf{\Lambda}}$ . The null model is constructed a symmetric form for  $\hat{\mathbf{\Lambda}}$  while the alternative model uses a compound symmetry form. This time  $\hat{\mathbf{D}}$  has a symmetric form for both models, and will be the same for both.

As the within-subject variabilities are fundamental to the coefficient of repeatability, this variability test likelihood ratio test is equivalent to testing the equality of two coefficients of repeatability of two methods. In presenting the results of this test, Roy (2009a) includes the coefficients of repeatability for both methods.

#### 1.8.5 Variability test 3

The last of the variability test examines whether or not methods  $A$  and  $B$  have the same overall variability. This enables the joint consideration of second and third criteria.

$$H_0 : \sigma_A = \sigma_B$$

$$H_A : \sigma_A = \sigma_B$$

The null model is constructed a symmetric form for both  $\hat{D}$  and  $\hat{\Lambda}$  while the alternative model uses a compound symmetry form for both.

### 1.8.6 Demonstration of Roy's testing

Roy provides three case studies, using data sets well known in method comparison studies, to demonstrate how the methodology should be used. The first two examples used are from the 'blood pressure' data set introduced by Bland and Altman (1999). The data set is a tabulation of simultaneous measurements of systolic blood pressure were made by each of two experienced observers (denoted 'J' and 'R') using a sphygmomanometer and by a semi-automatic blood pressure monitor (denoted 'S'). Three sets of readings were made in quick succession. Roy compares the 'J' and 'S' methods in the first of her examples.

The inter-method bias between the two method is found to be 15.62 , with a  $t$ -value of  $-7.64$ , with a  $p$ -value of less than 0.0001. Consequently there is a significant inter-method bias present between methods  $J$  and  $S$ , and the first of the Roy's three agreement criteria is unfulfilled.

Next, the first variability test is carried out, yielding maximum likelihood estimates of the between-subject variance covariance matrix, for both the null model, in compound symmetry (CS) form, and the alternative model in symmetric (symm) form. These matrices are determined to be as follows;

$$\hat{D}_{CS} = \begin{pmatrix} 946.50 & 784.32 \\ 784.32 & 946.50 \end{pmatrix}, \quad \hat{D}_{Symm} = \begin{pmatrix} 923.98 & 785.24 \\ 785.24 & 971.30 \end{pmatrix}.$$

A likelihood ratio test is perform to compare both candidate models. The log-likelihood of the null model is  $-2030.7$ , and for the alternative model  $-2030.8$ . The test statistic, presented with greater precision than the log-likelihoods, is 0.1592. The  $p$ -value is 0.6958. Consequently we fail to reject the null model, and by extension,

conclude that the hypothesis that methods  $J$  and  $S$  have the same between-subject variability. Thus the second of the criteria is fulfilled.

The second variability test determines maximum likelihood estimates of the within-subject variance covariance matrix, for both the null model, in CS form, and the alternative model in symmetric form.

$$\hat{\Lambda}_{CS} = \begin{pmatrix} 60.27 & 16.06 \\ 16.06 & 60.27 \end{pmatrix}, \quad \hat{\Lambda}_{Symm} = \begin{pmatrix} 37.40 & 16.06 \\ 16.06 & 83.14 \end{pmatrix}.$$

Again, A likelihood ratio test is perform to compare both candidate models. The log-likelihood of the alternative model model is  $-2045.0$ . As before, the null model has a log-likelihood of  $-2030.7$ . The test statistic is computed as 28.617, again presented with greater precision. The  $p$ -value is less than 0.0001. In this case we reject the null hypothesis of equal within-subject variability. Consequently the third of Roy's criteria is unfulfilled. The coefficient of repeatability for methods  $J$  and  $S$  are found to be 16.95 mmHg and 25.28 mmHg respectively.

The last of the three variability tests is carried out to compare the overall variabilities of both methods. With the null model the MLE of the within-subject variance covariance matrix is given below. The overall variabilities for the null and alternative models, respectively, are determined to be as follows;

$$\hat{\Sigma}_{CS} = \begin{pmatrix} 1007.92 & 801.65 \\ 801.65 & 1007.92 \end{pmatrix}, \quad \hat{\Sigma}_{Symm} = \begin{pmatrix} 961.38 & 801.40 \\ 801.40 & 1054.43 \end{pmatrix},$$

The log-likelihood of the alternative model model is  $-2045.2$ , and again, the null model has a log-likelihood of  $-2030.7$ . The test statistic is 28.884, and the  $p$ -value is less than 0.0001. The null hypothesis, that both methods have equal overall variability, is rejected. Further to the second variability test, it is known that this difference is specifically due to the difference of within-subject variabilities.

Lastly, Roy considers the overall correlation coefficient. The diagonal blocks  $\hat{\mathbf{r}}_{\Omega ii}$  of the correlation matrix indicate an overall coefficient of 0.7959. This is less than the threshold of 0.82 that Roy recommends.

$$\hat{\mathbf{r}}_{\Omega ii} = \begin{pmatrix} 1 & 0.7959 \\ 0.7959 & 1 \end{pmatrix}$$

The off-diagonal blocks of the overall correlation matrix  $\hat{\mathbf{r}}_{\Omega ii'}$  present the correlation coefficients further to Hamlett et al. (2004).

$$\hat{\mathbf{r}}_{\Omega ii'} = \begin{pmatrix} 0.9611 & 0.7799 \\ 0.7799 & 0.9212 \end{pmatrix}.$$

The overall conclusion of the procedure is that method  $J$  and  $S$  are not in agreement, specifically due to the within-subject variability, and the inter-method bias. The repeatability coefficients are substantially different, with the coefficient for method  $S$  being 49% larger than for method  $J$ . Additionally the overall correlation coefficient did not exceed the recommended threshold of 0.82.

## 1.9 Standard Deviation of Differences

In computing limits of agreement, it is first necessary to have an estimate for the standard deviations of the differences. When the agreement of two methods is analyzed using LME models, a clear method of how to compute the standard deviation is required. As the estimate for inter-method bias and the quantile would be the same for both methodologies, the focus hereon is solely on the variance of differences.

The standard deviation of the differences of methods  $x$  and  $y$  is computed using values from the overall VC matrix.

$$\text{Var}(x - y) = \text{Var}(x) + \text{Var}(y) - 2\text{Cov}(x, y)$$

## 1.10 Implementation in R

To implement an LME model in R, the `nlme` package is used. This package is loaded into the R environment using the `library` command, (i.e. `library(nlme)`). The `lme` command is used to fit LME models. The first two arguments to the `lme` function specify the fixed effect component of the model, and the data set to which the model is to be fitted. The first candidate model ('MCS1') fits an LME model on the data set 'dat'. The variable 'method' is assigned as the fixed effect, with the response variable 'BP' (i.e. blood pressure).

The third argument contain the random effects component of the formulation, describing the random effects, and their grouping structure. The `nlme` package provides a set of positive-definite matrices, the `pdMat` class, that can be used to specify a structure for the between-subject variance-covariance matrix for the random effects. For Roy's methodology, we will use the `pdSymm` and `pdCompSymm` to specify a symmetric structure and a compound symmetry structure respectively. A full discussion of these



structures can be found in Pinheiro and Bates (1994, pg. 158).

Similarly a variety of structures for the with-subject variance-covariance matrix can be implemented using `nlme`. To implement a particular matrix structure, one must specify both a variance function and correlation structure accordingly. Variance functions are used to model the variance structure of the within-subject errors. `varIdent` is a variance function object used to allow different variances according to the levels of a classification factor in the data. A compound symmetry structure is implemented using the `corCompSymm` class, while the symmetric form is specified by `corSymm` class. Finally, the estimation methods is specified as “ML” or “REML”.

The first of Roy's candidate model can be implemented using the following code;

```
MCS1 = lme(BP ~ method-1, data = dat,  
random = list(subject=pdSymm(~ method-1)),  
weights=varIdent(form=~1|method),  
correlation = corSymm(form=~1 | subject/obs), method="ML")
```

For the blood pressure data used in Roy (2009a), all four candidate models are implemented by slight variations of this piece of code, specifying either `pdSymm` or `pdCompSymm` in the second line, and either `corSymm` or `corCompSymm` in the fourth line. For example, the second candidate model 'MCS2' is implemented with the same code as MCS1, except for the term `pdCompSymm` in the second line, rather than `pdSymm`.

```
MCS2 = lme(BP ~ method-1, data = dat,  
random = list(subject=pdCompSymm(~ method-1)),  
weights = varIdent(form=~1|method),  
correlation = corSymm(form=~1 | subject/obs), method="ML")
```

Using this R implementation for other data sets requires that the data set is structured appropriately (i.e. each case of observation records the index, response, method and replicate). Once formatted properly, implementation is simply a case of re-writing the first line of code, and computing the four candidate models accordingly.

To perform a likelihood ratio test for two candidate models, simply use the `anova` command with the names of the candidate models as arguments. The following piece of code implement the first of Roy's variability tests.

```
> anova(MCS1,MCS2)

Model df      AIC      BIC logLik  Test L.Ratio p-value
MCS1   1  8 4077.5 4111.3 -2030.7
MCS2   2  7 4075.6 4105.3 -2030.8 1 vs 2 0.15291 0.6958
>
```

The fixed effects estimates are the same for all four candidate models. The inter-method bias can be easily determined by inspecting a summary of any model. The summary presents estimates for all of the important parameters, but not the complete variance-covariance matrices (although some simple R functions can be written to overcome this). The variance estimates for the random effects for MCS2 is presented below.

```
Random effects:
Formula: ~method - 1 | subject
Structure: Compound Symmetry
StdDev Corr
methodJ 30.765
methodS 30.765 0.829
Residual 6.115
```

Similarly, for computing the limits of agreement the standard deviation of the differ-

ences is not explicitly given. Again, A simple R function can be written to calculate the limits of agreement directly.

## 1.11 Covariance Parameters

The unknown variance elements are referred to as the covariance parameters and collected in the vector  $\theta$ .

# Bibliography

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19(6), 716–723.
- Bland, J. and D. Altman (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research* 8(2), 135–160.
- 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).
- Eisenhart, C. (1947). The assumptions underlying the analysis of variance. *Biometrics* 3(1), 1–21.
- Fisher, R. (1918). The correlation between relatives on the supposition of mendelian inheritance. *Transactions of the Royal Society of Edinburgh* 2, 399–433.
- 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.
- Hartley, H. and J. Rao (1967). Maximum-likelihood estimation for the mixed analysis of variance model. *Biometrika* 54(1/2), 93–108.

- Henderson, C. (1953). Estimation of variance and covariance components. *Biometrics* 9(2), 226–252.
- 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.
- Henderson, C., O. Kempthorne, S. Searle, and C. von Krosigk (1963). The estimation of environmental and genetic trends from records subject to culling. *Biometrics* 15, 192–218.
- Henderson, C., O. Kempthorne, S. Searle, and C. von Krosigk (1973). The estimation of environmental and genetic trends from records subject to culling. *Biometrics* 15, 192–218.
- Henderson, C., O. Kempthorne, S. Searle, and C. von Krosigk (1984). The estimation of environmental and genetic trends from records subject to culling. *Biometrics* 15, 192–218.
- Henderson, C. R. (1950). Estimation of genetic parameters (abstract). *Annals of Mathematical Statistics* 21, 309–310.
- Laird, N. M. and J. H. Ware (1982). Random-effects models for longitudinal data. *Biometrics* 38(4), 963–974.
- 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.
- Patterson, H. and R. Thompson (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika* 58(3), 545–554.

- Pinheiro, J. and D. Bates (1994). *Mixed Effects Models in S and S plus* (2nd ed.). Reading, Massachusetts: Springer.
- Roy, A. (2006). Estimating correlation coefficient between two variables with repeated observations using mixed effects models. *Biometric Journal* 2, 286–301.
- Roy, A. (2009a). An application of linear mixed effects model to assess the agreement between two methods with replicated observations. *Journal of Biopharmaceutical Statistics* 19, 150–173.
- Roy, A. (2009b). 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.
- Searle, S. (1997). *Linear Models*. Wiley classics Library.
- Singer, J. D. (1998). Using sas proc mixed to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics* 24(4), 323–355.
- Tippett, L. (1931). *The Methods of Statistics* (1st ed.). London: Williams and Norgate.