

Chapter 1

A Simplified LME Framework for Method Comparison

1.1 Roy's Model

Roy (2009) proposes a LME based approach with Kronecker product covariance structure with doubly multivariate setup to assess the agreement between two methods. 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. This approach allows for unbalanced data, i.e. unequal numbers of replications for each subject.

Response for i th subject can be written as

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + b_{1i} z_{i1} + b_{2i} z_{i2} + \epsilon_i$$

- β_1 and β_2 are fixed effects corresponding to both methods. (β_0 is the intercept.)
- b_{1i} and b_{2i} are random effects corresponding to both methods.

Overall variability between the two methods (Ω) is sum of between-subject (G) and within-subject

variability (Σ),

$$\text{Block } \boldsymbol{\Omega}_i = \begin{bmatrix} d_1^2 & d_{12} \\ d_{12} & d_2^2 \end{bmatrix} + \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}.$$

In order to express Roy's LME model in matrix notation we gather all $2n_i$ observations specific to item i into a single vector $\mathbf{y}_i = (y_{1i1}, y_{2i1}, y_{1i2}, \dots, y_{mir}, \dots, y_{1in_i}, y_{2in_i})'$. The LME model can be written

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

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$ is a vector of fixed effects, and \mathbf{X}_i is a corresponding $2n_i \times 3$ design matrix for the fixed effects. The random effects are expressed in the vector $\mathbf{b} = (b_1, b_2)'$, with \mathbf{Z}_i the corresponding $2n_i \times 2$ design matrix. The vector $\boldsymbol{\epsilon}_i$ is a $2n_i \times 1$ vector of residual terms.

It is assumed that $\mathbf{b}_i \sim N(0, \mathbf{G})$, $\boldsymbol{\epsilon}_i$ is a matrix of random errors distributed as $N(0, \mathbf{R}_i)$ and that the random effects and residuals are independent of each other.

The maximum likelihood estimate of the between-subject variance covariance matrix of two methods is given as G . 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{G} and $\hat{\Sigma}$.

$$\text{Block } \Omega_i = \hat{G} + \hat{\Sigma} \quad (1.1)$$

\mathbf{G} is the variance covariance matrix for the random effects \mathbf{b} . i.e. between-item sources of variation. The between-item variance covariance matrix \mathbf{G} is constructed as follows:

$$\text{Var} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \mathbf{G} = \begin{pmatrix} g_1^2 & g_{12} \\ g_{12} & g_2^2 \end{pmatrix}$$

It is important to note that no special assumptions about the structure of \mathbf{G} are made. An example of such an assumption would be that \mathbf{G} is the product of a scalar value and the identity matrix.

\mathbf{R}_i is the variance covariance matrix for the residuals, i.e. the within-item sources of variation between both methods. Computational analysis of linear mixed effects models allow for the explicit analysis of both \mathbf{G} and \mathbf{R}_i .

The overall variability between the two methods is the sum of between-item variability \mathbf{G} and within-item variability Σ . Roy (2009) denotes the overall variability as Block - $\boldsymbol{\Omega}_i$. The overall variation for methods 1 and 2 are given by

$$\begin{pmatrix} \omega_1^2 & \omega_{12} \\ \omega_{12} & \omega_2^2 \end{pmatrix} = \begin{pmatrix} g_1^2 & g_{12} \\ g_{12} & g_2^2 \end{pmatrix} + \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$$

The computation of the limits of agreement require that the variance of the difference of measurements. This variance is easily computable from the estimate of the Block - $\mathbf{\Omega}_i$ matrix. Lack of agreement can arise if there is a disagreement in overall variabilities. This may be due to due to the disagreement in either between-item variabilities or within-item variabilities, or both. Roy (2009) allows for a formal test of each.

1.2 Model Terms for Roy's Techniques

\mathbf{b}_i is a m -dimensional vector comprised of the random effects.

$$\mathbf{b}_i = \begin{pmatrix} b_{1i} \\ b_{2i} \end{pmatrix} \quad (1.2)$$

\mathbf{V} represents the correlation matrix of the replicated measurements on a given method. $\mathbf{\Sigma}$ is the within-subject VC matrix.

\mathbf{V} and $\mathbf{\Sigma}$ are positive definite matrices. The dimensions of \mathbf{V} and $\mathbf{\Sigma}$ are $3 \times 3 (= p \times p)$ and $2 \times 2 (= k \times k)$.

It is assumed that \mathbf{V} is the same for both methods and $\mathbf{\Sigma}$ is the same for all replications. $\mathbf{V} \otimes \mathbf{\Sigma}$ creates a $6 \times 6 (= kp \times kp)$ matrix. \mathbf{R}_i is a sub-matrix of this.

1.3 Model terms

It is important to note the following characteristics of this model. Let the number of replicate measurements on each item i for both methods be n_i , hence $2 \times n_i$ responses. However, it is assumed that there may be a different number of replicates made for different items. Let the maximum number of replicates be p . An item will have up to $2p$ measurements, i.e. $\max(n_i) = 2p$.

Later on \mathbf{X}_i will be reduced to a 2×1 matrix, to allow estimation of terms. This is due to a shortage of rank. The fixed effects vector can be modified accordingly. \mathbf{Z}_i is the $2n_i \times 2$ model matrix for the

random effects for measurement methods on item i . \mathbf{b}_i is the 2×1 vector of random-effect coefficients on item i , one for each method. $\boldsymbol{\epsilon}$ is the $2n_i \times 1$ vector of residuals for measurements on item i . \mathbf{G} is the 2×2 covariance matrix for the random effects. \mathbf{R}_i is the $2n_i \times 2n_i$ covariance matrix for the residuals on item i . The expected value is given as $E(\mathbf{y}_i) = \mathbf{X}_i\boldsymbol{\beta}$. (Hamlett et al., 2004) The variance of the response vector is given by $\text{Var}(\mathbf{y}_i) = \mathbf{Z}_i\mathbf{G}\mathbf{Z}_i' + \mathbf{R}_i$ (Hamlett et al., 2004).

Bibliography

- 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.
- 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.
- 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.