

1 Modelling Agreement with LME Models

Roy (2009) considers the matter of comparing the agreement of two methods using a linear mixed effects (LME) model. An LME model used to describe replicate measurements made by two methods on a number of individuals is presented.

Let y_{mir} be the r -th replicate measurement of the i th item (e.g. patient) by the m -th method. It is assumed that only two methods to be compared, hence $m = \{1, 2\}$. The fixed effect for each method, that gives rise to inter-method bias, is denoted β_m , with the difference, i.e. the inter-method bias, therefore as $\beta_1 - \beta_2$. Additionally there is an intercept term β_0 , essentially the true mean of the population of items. This being pertinent when comparing this model with others.

$$y_{mir} = \beta_0 + \beta_m + b_{mi} + \epsilon_{mir} \quad (1)$$

The random effect for item i , which is associated with method m , is denoted b_{mi} . The residual is denoted ϵ_{mir} . A detailed discussion of the variances of both shall be provided in due course.

1.1 Response Vector

Consider the response vector \mathbf{y}_i comprises the $2n_i$ observations of the item, as measured by two methods, taking n_i measurements each. Each response on item i is stacked into a response vector.

$$\mathbf{y}_i = (y_{1i1}, y_{2i1}, y_{1i2}, \dots, y_{mir}, \dots, y_{1in_i}, y_{2in_i})' \quad (2)$$

To formulate a model for the response vector \mathbf{y}_i , the fixed effects for both methods are given as β_1 and β_2 , in addition to the true value effect β_i , respectively, while the random effect terms are given as b_1 and b_2 . Two matrices of indicator variables (later referred to as \mathbf{X}_i and \mathbf{Z}_i) enable the correct effects for each response. For expository purposes, we choose $n_i = 3$. Hence a 6×1 random vector corresponding to the i th

subject. Thus, for two methods and each with three replicates, the response vector would be formulated as:

$$\mathbf{y}_i = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} b_{1i} \\ b_{2i} \end{pmatrix} + \begin{pmatrix} \epsilon_{1i1} \\ \epsilon_{2i1} \\ \epsilon_{1i2} \\ \epsilon_{2i2} \\ \epsilon_{1i3} \\ \epsilon_{2i3} \end{pmatrix} \quad (3)$$

This model can be conveniently presented in matrix form as follows;

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, I \quad (4)$$

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. Assumptions made on the structures of \mathbf{G} and \mathbf{R}_i will be discussed in due course.

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. $\max(n_i) = 2p$.
- \mathbf{y}_i is the $2n_i \times 1$ response vector for measurements on the i -th item.
- \mathbf{X}_i is the $2n_i \times 3$ model matrix for the fixed effects for observations on item i .
- $\boldsymbol{\beta}$ is the 3×1 vector of fixed-effect coefficients, one for the true value for item i , and one effect each for both methods.
- 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).

2 Variance Matrices

\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 .

Hamlett et al. (2004) shows that \mathbf{R}_i can be expressed as $\mathbf{R}_i = \mathbf{I}_{n_i} \otimes \mathbf{\Sigma}$. The partial within-item variancecovariance matrix of two methods at any replicate is denoted $\mathbf{\Sigma}$, where σ_1^2 and σ_2^2 are the within-subject variances of the respective methods, and σ_{12} is the within-item covariance between the two methods. It is assumed that the within-item variancecovariance matrix $\mathbf{\Sigma}$ is the same for all replications. Again it is important to note that no special assumptions are made about the structure of the matrix.

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \quad (5)$$

2.1 Coefficient of Repeatability

The coefficient of repeatability is a measure of how well a measurement method agrees with itself over replicate measurements (Bland and Altman, 1999). Once the within-item variability is known, the computation of the coefficients of repeatability for both methods is straightforward.

2.2 Overall Variability

The overall variability between the two methods is the sum of between-item variability \mathbf{G} and within-item variability $\mathbf{\Sigma}$. Roy (2009) denotes the overall variability as Block - $\mathbf{\Omega}_i$. 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 the disagreement in either between-item variabilities or within-item variabilities, or both. Roy (2009) allows for a formal test of each.

2.3 Hypothesis Testing

The formulation presented above usefully facilitates a series of significance tests that advise as to how well the two methods agree. These tests are as follows:

- A formal test for the equality of between-item variances,
- A formal test for the equality of within-item variances,
- A formal test for the equality of overall variances.

These tests are complemented by the ability to consider the inter-method bias and the overall correlation coefficient. Two methods can be considered to be in agreement if criteria based upon these methodologies are met. Additionally Roy makes reference to the overall correlation coefficient of the two methods, which is determinable from variance estimates.

3 Carstensen's Limits of agreement

Carstensen et al. (2008) presents a methodology to compute the limits of agreement based on LME models. Importantly, Carstensen's underlying model differs from Roy's model in some key respects, and therefore a prior discussion of Carstensen's model is required.

3.1 Carstensen's Model

Carstensen (2004) presents a model 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;

$$y_{mi} = \alpha_m + \mu_i + e_{mi} \quad e_{mi} \sim \mathcal{N}(0, \sigma_m^2) \quad (6)$$

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.

$$y_{mir} = \alpha_m + \mu_i + c_{mi} + e_{mir}, \quad e_{mi} \sim \mathcal{N}(0, \sigma_m^2), \quad c_{mi} \sim \mathcal{N}(0, \tau_m^2). \quad (7)$$

Of particular importance is terms of the model, a true value for item i (μ_i). The fixed effect of Roy's model comprise of an intercept term and fixed effect terms for both methods, with no reference to the true value of any individual item. A distinction can be made between the two models: Roy's model is a standard LME model, whereas Carstensen's model is a more complex additive model.

3.2 Assumptions on Variability

Aside from the fixed effects, another important difference is that Carstensen's model requires that particular assumptions be applied, specifically that the off-diagonal elements of the between-item and within-item variability matrices are zero. By extension the overall variability off diagonal elements are also zero.

Also, implementation requires that the between-item variances are estimated as the same value: $g_1^2 = g_2^2 = g^2$. Necessarily Carstensen's method does not allow for a formal test of the between-item variability.

$$\begin{pmatrix} \omega_2^1 & 0 \\ 0 & \omega_2^2 \end{pmatrix} = \begin{pmatrix} g^2 & 0 \\ 0 & g^2 \end{pmatrix} + \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$$

In cases where the off-diagonal terms in the overall variability matrix are close to zero, the limits of agreement due to Carstensen et al. (2008) are very similar to the limits of agreement that follow from the general model.

References

- Bland, J. and D. Altman (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research* 8(2), 135–160.
- Carstensen, B. (2004). Comparing and predicting between several methods of measurement. *Biostatistics* 5(3), 399–413.
- Carstensen, B., J. Simpson, and L. C. Gurrin (2008). Statistical models for assessing agreement in method comparison studies with replicate measurements. *The International Journal of Biostatistics* 4(1).
- 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.
- 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.