

# 1 Carstensen's Model

? proposes an approach for comparing two or more methods of measurement based on linear mixed effects models. This approach extends the well established Bland-Altman methodology for the case of replicate measurements on each item. Carstensen considers the matter of computing an appropriate estimate for the standard deviation of case-wise differences, so as to determine the limits of agreement. As the interest lies in extending the Bland-Altman methodology, other formal tests are not described.

Using Carstensen's notation, a measurement  $y_{mi}$  by method  $m$  on individual  $i$  the measurement  $y_{mir}$  is the  $r$ th replicate measurement on the  $i$ th item by the  $m$ th method, where  $m = 1, 2, \dots, M$   $i = 1, \dots, N$ , and  $r = 1, \dots, n_i$  is formulated as follows;

$$y_{mir} = \alpha_m + \mu_i + c_{mi} + a_{ir} + \epsilon_{mir}, \quad c_{mi} \sim \mathcal{N}(0, \tau_m^2), a_{ir} \sim \mathcal{N}(0, \varsigma^2), \epsilon_{mi} \sim \mathcal{N}(0, \varphi_m^2). \quad (1)$$

Here the terms  $\alpha_m$  and  $\mu_i$  represent the fixed effect for method  $m$  and a true value for item  $i$  respectively. The random effect terms comprise an interaction term  $c_{mi}$  and the residuals  $\epsilon_{mir}$ . The  $c_{mi}$  term represent random effect parameters corresponding to the two methods, having  $E(c_{mi}) = 0$  with  $\text{Var}(c_{mi}) = \tau_m^2$ .

The random error term for each response is denoted  $\epsilon_{mir}$  having  $E(\epsilon_{mir}) = 0$ ,  $\text{Var}(\epsilon_{mir}) = \varphi_m^2$ . All the random effects are assumed independent, and that all replicate measurements are assumed to be exchangeable within each method.

When only two methods are to be compared, separate estimates of  $\tau_m^2$  can not be obtained. Instead the average value  $\tau^2$  is obtained and used.

Carstensen's approach is that of a standard two-way mixed effects ANOVA with replicate measurements. With regards to the specification of the variance terms, Carstensen remarks that using his approach is common, remarking that *The only slightly non-standard (meaning "not often used") feature is the differing residual variances between methods (?)*.

In contrast to Roy's model, Carstensen's model requires that commonly used 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:  $\tau_1^2 = \tau_2^2 = \tau^2$ .

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

The presence of the true value term  $\mu_i$  gives rise to an important difference between Carstensen's and Roy's models. 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. In other words, Roy considers the group of items being measured as a sample taken from a population. Therefore 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.

## 2 Carstensen's Mixed Models

? 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  $m$ .

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

The intercept term  $\alpha$  and the  $\beta_m \mu_i$  term follow from ?, expressing constant and proportional bias respectively , in the presence of a real value  $\mu_i$ .  $c_{mi}$  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).

? uses the above formula to predict observations for a specific individual  $i$  by method  $m$ ;

$$BLUP_{mir} = \hat{\alpha}_m + \hat{\beta}_m \mu_i + c_{mi} \quad (3)$$

. 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. ? 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.

### 3 Carstensen's Mixed Models

? 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  $m$ .

$$y_{mir} = \alpha_m + \beta_m \mu_i + c_{mi} + e_{mir} \quad (e_{mi} \sim N(0, \sigma_m^2), c_{mi} \sim N(0, \tau_m^2)) \quad (4)$$

The intercept term  $\alpha$  and the  $\beta_m \mu_i$  term follow from ?, expressing constant and proportional bias respectively , in the presence of a real value  $\mu_i$ .  $c_{mi}$  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).

? uses the above formula to predict observations for a specific individual  $i$  by method  $m$ ;

$$BLUP_{mir} = \hat{\alpha}_m + \hat{\beta}_m \mu_i + c_{mi} \quad (5)$$

. 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. ? provides an amended formulation which includes an extra interaction term ( $d_{mr} \sim N(0, \omega_m^2)$ ) 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.

? 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;

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

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.

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

? 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:

$$var(y_{1j} - y_{2j}) \quad (8)$$

### 3.1 Bendix Carstensen's data sets

?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.