

# Contents

<b>1</b>	<b>Fitting MCS Models with R</b>	<b>3</b>
1.0.1	Criteria for Agreement . . . . .	3
1.1	LME models in method comparison studies . . . . .	4
1.1.1	Tests . . . . .	4
1.2	Roy's Framework . . . . .	5
1.3	Systolic Blood Pressure Data Set . . . . .	6
1.4	Implementation in R . . . . .	6
1.5	Computation of LMEs using R . . . . .	8
1.6	Important Consideration for MCS . . . . .	8
1.7	Roy's Model . . . . .	10
1.8	Fitting Models with the LME4 R package . . . . .	11
1.8.1	Demonstration of Roy's testing . . . . .	12
1.9	Roy's Variability Tests . . . . .	12
1.9.1	Correlation . . . . .	13
1.10	Roy's Model . . . . .	13
1.11	Model Terms for Roy's Techniques . . . . .	13
1.12	Roy's Candidate Models . . . . .	14
1.13	Model terms . . . . .	15
1.13.1	Roy's Reference Model . . . . .	15
1.13.2	Model Fit 1 . . . . .	16
1.13.3	Model Fit 2 . . . . .	19

1.14 Likelihood Ratio Test . . . . .	20
1.14.1 LRTs with <b>R</b> . . . . .	21
1.14.2 Worked Examples : LikelihoodRatio Tests . . . . .	23
1.14.3 Nested Model (Between-Item Variability) . . . . .	24

# Chapter 1

## Fitting MCS Models with R

### 1.0.1 Criteria for Agreement

Roy (2009) proposes a suite of hypothesis tests for assessing the agreement of two methods of measurement, when replicate measurements are obtained for each item, using a LME approach. (An item would commonly be a patient).

Two methods of measurement can be said to be in agreement if there is no significant difference between in three key respects. Firstly, there is no inter-method bias between the two methods, i.e. there is no persistent tendency for one method to give higher values than the other.

Secondly, both methods of measurement have the same within-subject variability. In such a case the variance of the replicate measurements would consistent for both methods. Lastly, the methods have equal between-subject variability.

For the mean measurements for each case, the variances of the mean measurements from both methods are equal.

Roy (2009) 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 (2009) 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.

Barnhart et al. (2007) describes the sources of disagreement as differing population means, different between-subject variances, different within-subject variances between two methods and poor correlation between measurements of two methods.

Roy (2009) considers two methods to be in agreement if three conditions are met.

1. no significant bias, i.e. the difference between the two mean readings is not "statistically significant",
2. high overall correlation coefficient,
3. the agreement between the two methods by testing their repeatability coefficients.

Roy (2009) demonstrates a LME model specification, and a series of tests that look at each of these agreement criteria individually. If two methods of measurement lack agreement, the specific reason or reasons for this lack of agreement can be identified.

## **1.1 LME models in method comparison studies**

Linear mixed effects (LME) models can facilitate greater understanding of the potential causes of bias and differences in precision between two sets of measurement. Lai and Shiao (2005) views the uses of linear mixed effects models as an expansion on the Bland-Altman methodology, rather than as a replacement. Carstensen et al. (2008) remarks that modern statistical computation, such as that used for LME models, greatly improve the efficiency of calculation compared to previous 'by-hand' methods.

Roy provides three case studies, using data sets well known in method comparison studies, to demonstrate how the methodology should be used.

### **1.1.1 Tests**

Roy (2009) considers four independent hypothesis tests.

- Testing of hypotheses of differences between the means of two methods
- Testing of hypotheses in between subject variabilities in two methods,

- Testing of hypotheses of differences in within-subject variability of the two methods,
- Testing of hypotheses in differences in overall variability of the two methods.

Bivariate correlation coefficients have been shown to be of limited use in method comparison studies (Bland and Altman, 1986). However, recently correlation analysis has been developed to cope with repeated measurements, enhancing their potential usefulness. Roy incorporates the use of correlation into this methodology.

## 1.2 Roy's Framework

Roy (2009) uses an approach based on linear mixed effects (LME) models for the purpose of comparing the agreement between two methods of measurement, where replicate measurements on items, typically individuals, by both methods are available.

This framework provides three tests of hypothesis appropriate for evaluating the agreement between the two methods of measurement under this sampling scheme. These tests consider null hypotheses that assume: absence of inter-method bias; equality of between-subject variabilities of the two methods; equality of within-subject variabilities of the two methods.

By inter-method bias we mean that a systematic difference exists between observations recorded by the two methods. Differences in between-subject variabilities of the two methods arise when one method is yielding average response levels for individuals that are more variable than the average response levels for the same sample of individuals taken by the other method.

Differences in within-subject variabilities of the two methods arise when one method is yielding responses for an individual that are more variable than the responses for this same individual taken by the other method. The two methods of measurement can be considered to agree, and subsequently can be used interchangeably, if all three null hypotheses are true.

This bias is specified as a fixed effect in the LME model. For a practitioner who has a reasonable level of competency in **R** and undergraduate statistics (in particular simple linear regression model) this is a straight-forward procedure.

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 (2009) 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  $\mathbf{D}$  and  $\mathbf{\Lambda}$  matrices are constructed, using either an unstructured form or a compound symmetry form. To illustrate these differences, consider a generic matrix  $\mathbf{A}$ ,

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

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  $a_{11} = a_{22}$ .

### 1.3 Systolic Blood Pressure Data Set

Roy (2009) 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.

## 1.4 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 (`'ref.nlme'`) fits an LME model on the data set 'Blood'. The variable `'meth'` is assigned as the fixed effect, with the response variable `'y'` (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 models, 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".

`y` : Response variable

`meth` : Method of Measurement

`item` : Subject

`repl` : Replicate index

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.

## 1.5 Computation of LMEs using R

Pinheiro and Bates (1994) advises how to implement LME models in statistical software (ostensibly for S and S PLUS, but R is very similar). When tackling linear mixed effects models using the R language, a statistician can use the `lme()` function, from the `nlme` package.

The first two arguments for `lme` are `fixed` and `data`, which give the model for the expected responses (i.e. the fixed part of the model), and the data that the model should be fitted from. The next argument is `random`, a one-sided formula which describes the random effects, and the grouping structure for the model.

The `lme()` command fits a LME model to the data set using either Maximum Likelihood (ML) or Restricted Maximum Likelihood (REML). The `method` argument can specify whether to use 'REML', the default setting, or 'ML'.

## 1.6 Important Consideration for Method Comparison

The key issue is that `nlme` allows for the particular specification of Roy's Model, specifically direct specification of the VC matrices for within subject and between subject residuals.

The `lme4` package does not allow for Roy's Model, for reasons that will be identified shortly. To advance the ideas that emanate from Roy's paper, one is required to use the `nlme` context. However, to take advantage of the infrastructure already provided for `lme4` models, one may change the research question away from that of Roy's paper. To this end, an exploration of what ***influence.ME*** can accomplish is merited. The first of Roy's candidate model can be implemented using the following code;



```
ref.nlme <- 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 (2009), 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`.

```
test1.nlme <- 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")
```

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 `test1.nlme` 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 differences is not explicitly given. Again, a simple R function can be written to calculate the limits of agreement directly.

## 1.7 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

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

- $\beta_1$  and  $\beta_2$  are fixed effects corresponding to both methods, with  $\beta_0$  as the intercept.
- $b_{1i}$  and  $b_{2i}$  are random effects corresponding to both methods.

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  $\mathbf{D}$ . The estimate for the within-subject variance covariance matrix is  $\hat{\Sigma}$ . The estimated overall variance covariance matrix 'Block  $\Omega_i$ ' is the addition of  $\hat{D}$  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  $\Sigma$ . Roy (2009) denotes the overall variability as Block -  $\Omega_i$ . 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 -  $\Omega_i$  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.

## 1.8 Fitting Models with the LME4 R package

Two LME models are fitted to the data, one using the `nlme` package, one with the `lme4` package. These models shall be called “`ref.nlme`” and “`ref.lme4`” respectively.

In both cases the method is characterized by a fixed effect, while there is a random effect for each subject. This random effect accounts for the replicate measurements associated with each subject. The differences between the estimate provided by the respective models are negligible, due to the simplicity of the model specification.

Maximum likelihood or restricted maximum likelihood (REML) estimates of the parameters in linear mixed-effects models can be determined using the `lmer` function in the ***lme4*** package for R. As for most model-fitting functions in R, the model is described in an `lmer` call by a formula, in this case including both fixed- and random-effects terms.

The formula and data together determine a numerical representation of the model from which the profiled deviance or the profiled REML criterion can be evaluated as a function of some of the model parameters. The appropriate criterion is optimized, using one of the constrained optimization functions in R, to provide the parameter estimates. We describe the structure of the model, the steps in evaluating the profiled deviance or REML criterion, and the structure of classes or types that represents such a model.

Sufficient detail is included to allow specialization of these structures by users who wish to write functions to fit specialized linear mixed models, such as models incorporating pedigrees or smoothing splines, that are not easily expressible in the formula language used by `lmer()`.

### 1.8.1 Demonstration of Roy’s testing

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.

## 1.9 Roy's Variability Tests

Variability tests proposed by Roy (2009) 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.

### 1.9.1 Correlation

Lastly, ? 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.

$$\hat{\mathbf{r}}_{\Omega_{ii}} = \begin{pmatrix} 1 & 0.7959 \\ 0.7959 & 1 \end{pmatrix} \quad (1.2)$$

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.

## 1.10 Roy's Model

## 1.11 Model Terms for Roy's Techniques

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

$$\mathbf{b}_i = \begin{pmatrix} b_{1i} \\ b_{2i} \end{pmatrix} \quad (1.3)$$

$\mathbf{V}$  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.

## 1.12 Roy's Candidate Models

```
> Ref.Fit = lme(y ~ meth-1, data = dat, #Symm , Symm#
+ random = list(item=pdSymm(~ meth-1)),
+ weights=varIdent(form=~1|meth),
+ correlation = corSymm(form=~1 | item/repl),
+ method="ML")
```

Roy(2009) presents two nested models that specify the condition of equality as required, with a third nested model for an additional test. These three formulations share the same structure, and can be specified by making slight alterations of the code for the Reference Model.

## 1.13 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  $2 \times 1$  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  $2 \times 1$  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  $2 \times 2$  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).

### 1.13.1 Roy's Reference Model

Conventionally LME models can be tested using Likelihood Ratio Tests, wherein a reference model is compared to a nested model.

Roy (2009) presents two nested models that specify the condition of equality as required, with a third nested model for an additional test. There three formulations share the same structure, and can be specified by making slight alterations of the code for the Reference Model.

### 1.13.2 Model Fit 1

This is a simple model with no interactions. There is a fixed effect for each method and a random effect for each subject.

$$y_{ijk} = \beta_j + b_i + \epsilon_{ijk}, \quad i = 1, \dots, 2, j = 1, \dots, 85, k = 1, \dots, 3$$

$$b_i \sim \mathcal{N}(0, \sigma_b^2), \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

```
> Ref.Fit = lme(y ~ meth-1, data = dat, #Symm , Symm#  
+ random = list(item=pdSymm(~ meth-1)),  
+ weights=varIdent(form=~1|meth),  
+ correlation = corSymm(form=~1 | item/repl),  
+ method="ML")
```

Linear mixed-effects model fit by REML

Data: dat

Log-restricted-likelihood: -2155.853

Fixed: BP ~ method

(Intercept)      methodS

127.40784      15.61961

Random effects:

Formula: ~1 | subject

(Intercept) Residual

StdDev:      29.39085 12.44454

Number of Observations: 510

Number of Groups: 85

The following output was obtained.

Linear mixed-effects model fit by REML

Data: dat

Log-restricted-likelihood: -2047.714

Fixed: BP ~ method

(Intercept)      methodS



127.40784      15.61961

Random effects:

Formula: ~1 | subject

(Intercept)

StdDev:      28.28452

Formula: ~1 | method %in% subject

(Intercept) Residual

StdDev:      12.61562 7.763666

Number of Observations: 510

Number of Groups:

subject method %in% subject

85                      170

For the blood pressure data used in Roy (2009), 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.

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.

A likelihood ratio test is performed to determine which model is more suitable. To perform this test, simply use the `anova` command with the names of the candidate models as arguments. The following piece of code implements 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 differences is not explicitly given. Again, A simple R function can be written to calculate the limits of agreement directly.

## 1.14 Likelihood Ratio Test

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.

LRTs can be used to test hypotheses about covariance parameters or fixed effects parameters in the context of LMEs. The test statistic for the likelihood ratio test is the difference of the log-likelihood functions, multiplied by  $-2$ .

The probability distribution of the test statistic is approximated by the  $\chi^2$  distribution with  $(\nu_1 - \nu_2)$  degrees of freedom, where  $\nu_1$  and  $\nu_2$  are the degrees of freedom of models 1 and 2 respectively. Each of these three test shall be examined in more detail shortly.

A general method for comparing nested models fit by maximum likelihood is the likelihood ratio test. This test can be used for models fit by REML (restricted maximum likelihood), but only if the fixed terms in the two models are invariant, and both models have been fit by REML. Otherwise, the argument: `method="ML"` must be employed (ML = maximum likelihood).

The output will contain a p-value, and this should be used in conjunction with the AIC scores to judge which model is preferred. Lower AIC scores are better.

Generally, likelihood ratio tests should be used to evaluate the significance of terms on the random effects portion of two nested models, and should not be used to determine the significance of the fixed effects.

A simple way to more reliably test for the significance of fixed effects in an LME model is to use conditional F-tests, as implemented with the simple `anova()` function.

### 1.14.1 LRTs with R

Likelihood ratio tests are very simple to implement in R, simply use the '`anova()`' commands. Sample output will be given for each variability test. The likelihood ratio test is the procedure used to compare the fit of two models. For each candidate model, the '-2 log likelihood' (M2LL) is computed. The test statistic for each of the three hypothesis tests is the difference of the M2LL for each pair of models. If the p-value in each of the respective tests exceed as significance level chosen by the analyst, then the null model must be rejected.

$$-2\ln\Lambda_d = [\text{M2LL under H0 model}] - [\text{M2LL under HA model}] \quad (1.4)$$

These test statistics follow a chi-square distribution with the degrees of freedom computed as the difference of the LRT degrees of freedom.

$$\nu = [\text{LRT df under H0 model}] - [\text{LRT df under HA model}] \quad (1.5)$$

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

```
#ANOVAs
```

```
test1 = anova(fit1,fit2) # Between-Subject Variabilities
```

```
test2 = anova(fit1,fit3) # Within-Subject Variabilities
```

```
test3 = anova(fit1,fit4) # Overall Variabilities
```

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 matter of how well two methods of measurement are said to be in agreement is a frequently posed question in statistical literature. A useful, and broadly consistent, set of definitions of what this agreement entail is put forth by Barnhart et al and Roy (2009). As pointed out by earlier contributors to the subject (commonly referred to as Method Comparison Studies)

Shared with previous contributions (Bland and Altman, Carstensen) is the condition that there should no systematic tendency for one of the methods to consistently provide a value higher than of the other method. If such a tendency did exist, we would refer to it as an inter-method bias.

In earlier literature, the emphasis was placed up on single measurements simultaneously by each of the methods of measurement. Several different approaches, such as the Bland-Altman plot, and Orthogonal Regression (a special case of Deming Regression where the residual variances are assumed to be equal) have been proposed. Arguably, for the single replicate case, the established methodologies are sufficient for assessing agreement between two methods.

In subsequent contributions, the matter of assessing agreement in the presence of replicate measurements was addressed. Some approaches extended already established approaches (Bland-Altman 1999). Other contributions were based on methodologies not seen previously in Method comparison Study Literature (for example, Carstensen et al 2008 and Roy 2009, using LME models).

### 1.14.2 Worked Examples : Likelihood Ratio Tests

The likelihood Ratio test is very simple to implement in R. All that is required is to specify the reference model and the relevant nested model as arguments to the command `anova()`.

The figure below displays the three tests described by Roy (2009).

```
> # Between-Subject Variabilities
> testB      = anova(Ref.Fit,NMB.fit)
>
> # Within-Subject Variabilities
> testW      = anova(Ref.Fit,NMW.fit)
>
> # Overall Variabilities
> testO      = anova(Ref.Fit,NMO.fit)
```

```
> 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
```

### 1.14.3 Nested Model (Between-Item Variability)

```
NMB.fit <- lme(y ~ meth-1,
data = dat,
random = list(item=pdCompSymm(~ meth-1)),
correlation = corSymm(form=~1 | item/repl),
method="ML")
```

# Bibliography

- Barnhart, H., M. Haber, and L. Lin (2007). An overview of assessing agreement with continuous measurements. *Journal of Biopharmaceutical Statistics* 17, 529–569.
- Bland, J. and D. Altman (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet* i, 307–310.
- 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).
- 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.
- Lai, D. and S.-Y. P. K. Shiao (2005). Comparing two clinical measurements: a linear mixed model approach. *Journal of Applied Statistics* 32(8), 855–860.
- 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.
- Pinheiro, J. and D. Bates (1994). *Mixed Effects Models in S and S plus* (2nd ed.). Reading, Massachusetts: Springer.

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.