

Chapter 1

Appendix C

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.

For these tests, four candidate LME models are constructed. The differences in the models are specifically in how the \mathbf{G} and Λ 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 $a_{11} = a_{22}$.

1.1 Roy's Variability Tests

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.

Variability tests proposed by Roy (2009) affords the opportunity to expand upon Carstensen's approach.

The first test allows of the comparison the within-subject variability of two methods. Similarly, the second test assesses the between-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.

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.

Other important aspects of the method comparison study are consequent. The limits of agreement are computed using the results of the first model.

1.2 Reference Model

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

```
> 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")
```

1.2.1 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.

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

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

```
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

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.

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 `NMW.fit` 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.

This is a simple model, this time with an interaction effect. There is a fixed effect for each method. This model has random effects at two levels b_i for the subject, and another, b_{ij} , for the respective method within each subject.

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

$$b_i \sim \mathcal{N}(0, \sigma_1^2), \quad b_{ij} \sim \mathcal{N}(0, \sigma_2^2), \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

In this model, the random interaction terms all have the same variance σ_2^2 . These terms are assumed to be independent of each other, even within the same subject.

1.2.2 Inter-method Bias

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.

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.

1.3 Likelihood Ratio Tests

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

1.4 Variability Test 1: Within –item Variability

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.

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

These matrices are determined to be as follows;

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

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.

1.4.1 Variability Test 1

This is a test on whether both methods A and B have the same within-subject variability or not.

$$H_0 : \sigma_A = \sigma_B \tag{1.1}$$

$$H_A : \sigma_A = \sigma_B \tag{1.2}$$

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

1.4.2 Test 1

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

For the null model the MLE of the within-subject variance covariance matrix is given below.

$$\hat{\mathbf{\Sigma}}_{Symm} = \begin{pmatrix} 37.40 & 16.06 \\ 16.06 & 83.14 \end{pmatrix} \quad (1.3)$$

With the alternative model the MLE is as follows:

$$\hat{\mathbf{\Sigma}}_{CS} = \begin{pmatrix} 60.27 & 16.06 \\ 16.06 & 60.27 \end{pmatrix} \quad (1.4)$$

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

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 outcome of this test is that it can be assumed that they have equal The test statistic is the

difference of the $-2 \log$ likelihoods; 28.617. The p -value is less than 0.0001. In this case we reject the null hypothesis that both models have the same within-subject variabilities.

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.

```
> 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 test statistic is the difference of the $-2 \log$ likelihoods; 0.153. The p -value is 0.696. Therefore we fail to reject the hypothesis that both have the same between-subject variabilities.

1.5 Variability Test 2: Between Subject Variances

This is a test on whether both methods A and B have the same between-subject variability or not.

$$H_0 : g_A = g_B \quad (1.5)$$

$$H_A : g_A \neq g_B \quad (1.6)$$

When implemented using **R**, this test is facilitated by constructing a model specifying a symmetric form for \mathbf{G} (i.e. the alternative model) and comparing it with a model that has compound symmetric form for \mathbf{G} (i.e. the null model). For this test $\hat{\Sigma}$ has a symmetric form for both models, and will be the same for both.

For example, the first test model ‘`NMB.fit`’ is implemented with the same code as `ref.nlme`, except for the term `pdCompSymm` in the second line, rather than `pdSymm`.

1.5.1 Nested Model (Between-Item Variability)

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

With the alternative model, the MLE of the between-subject variance covariance matrix \mathbf{G} is given by

$$\hat{\mathbf{G}}_{Symm} = \begin{pmatrix} 923.98 & 785.24 \\ 785.24 & 971.30 \end{pmatrix} \quad (1.7)$$

With the reference model the MLE is as follows:

$$\hat{\mathbf{G}}_{CS} = \begin{pmatrix} 946.50 & 784.32 \\ 784.32 & 946.50 \end{pmatrix} \quad (1.8)$$

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.

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.

1.6 Variability Test 3: Overall Variability

Additionally there is a third nested model, that can be used to test overall variability, substantively a joint test for between-item and within-item variability. The motivation for including such a test in the suite is not clear, although it does circumvent the need for multiple comparison procedures in certain circumstances, hence providing a simplified procedure for non-statisticians.

```
> NMO.fit = lme(y ~ meth-1, data = dat,    #CS , CS#  
+   random = list(item=pdCompSymm(~ meth-1)),  
+   correlation = corCompSymm(form=~1 | item/repl),  
+   method="ML")
```

1.6.1 Model Fit 3

This model is a more general model, compared to 'model fit 2'. This model treats the random interactions for each subject as a vector and allows the variance-covariance matrix for that vector to be estimated from the set of all positive-definite matrices. \mathbf{y}_i is the entire response vector for the i th subject. \mathbf{X}_i and \mathbf{Z}_i are the fixed- and random-effects design matrices respectively.

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, 85$$

$$\mathbf{Z}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}), \quad \boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{\Lambda})$$

1.6.2 Test 3

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;

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

The log-likelihood of the alternative 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.

This is a test on whether both methods A and B have the same overall variability or not.

$$H_0 : \sigma_A = \sigma_B \quad (1.9)$$

$$H_A : \sigma_A \neq \sigma_B \quad (1.10)$$

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

With the null model the MLE of the within-subject variance covariance matrix is given below.

$$\hat{\Sigma}_{Symm} = \begin{pmatrix} 961.38 & 801.40 \\ 801.40 & 1054.43 \end{pmatrix} \quad (1.11)$$

With the alternative model the MLE is as follows:

$$\hat{\Sigma}_{CS} = \begin{pmatrix} 1007.92 & 801.65 \\ 801.65 & 1007.92 \end{pmatrix} \quad (1.12)$$

The test statistic is the difference of the $-2 \log$ likelihoods; 28.884 . The p -value is less than 0.0001 . We again reject the null hypothesis. Each model has a different overall variability, a foregone conclusion from the second variability test.

Bibliography

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.