

## 0.1 Demonstration of Roy's testing

Roy 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.

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.

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. These matrices are determined to be as follows;

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

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.

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

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

Again, 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. The coefficient of repeatability for methods  $J$  and  $S$  are found to be 16.95 mmHg and 25.28 mmHg respectively.

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;

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

The log-likelihood of the alternative model 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.

Lastly, Roy considers the overall correlation coefficient. The diagonal blocks  $\hat{r}_{\Omega ii}$  of the correlation matrix indicate an overall coefficient of 0.7959. This is less than the

threshold of 0.82 that Roy recommends.

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

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.

## 0.2 Standard Deviation of Differences

In computing limits of agreement, it is first necessary to have an estimate for the standard deviations of the differences. When the agreement of two methods is analyzed using LME models, a clear method of how to compute the standard deviation is required. As the estimate for inter-method bias and the quantile would be the same for both methodologies, the focus hereon is solely on the variance of differences.

The standard deviation of the differences of methods  $x$  and  $y$  is computed using values from the overall VC matrix.

$$\text{Var}(x - y) = \text{Var}(x) + \text{Var}(y) - 2\text{Cov}(x, y)$$

## 0.3 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 ('MCS1') fits an LME model on the data set 'dat'. The variable 'method' is assigned as the fixed effect, with the response variable 'BP' (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 methodology, 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”.

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

```
MCS1 = 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`.

```
MCS2 = 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")
```

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.

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 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.



# Bibliography

Bland, J. and D. Altman (1999). Measuring agreement in method comparison studies.

*Statistical Methods in Medical Research* 8(2), 135–160.

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.

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 linear mixed effects model to assess the agreement between two methods with replicated observations. *Journal of Biopharmaceutical Statistics* 19, 150–173.