Chapter 1

A Simplified LME Framework for Method Comparison

1.1 A Review of Criteria for Agreement in Roy's Framework

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 an LME framework. For this framework, Roy (2009) recalls three criteria from Barnhart et al. (2007) for two methods to be considered in agreement (which are presented here in a different order);

1. There be no significant inter-method bias between the two methods, i.e. there is no persistent tendency for one method to give higher values than the other.

2. There is no difference in the within-subject variabilities, i.e. no significant difference in the random error among the replications taken by the same method on the same subject.

3. There is no significant difference in the between-subject variabilities, the variances of the mean measurements for each case from both methods are not significantly different.

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 measuement lack agreement, the specific reason or reasons for this lack of agreement can be identified. 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.

Roy (2009) provides an adjusted set of criteria, considering two methods to be in agreement if three conditions are met. Firstly that there is no significant bias, i.e. the difference between the two mean readings is not "statistically significant". Secondly there is agreement between the two methods by testing their repeatability coefficients. Finally, that there is a high overall correlation coefficient.

Criteria 1 and 2 are extensions of key statistical properties of Accuracy and Precision, which are discussed by Barnhart et al. (2007) in the context of Method Comparison, and can be considered as essential when comparing methods of measurement.

1.2   Critiquing the Relevance of Third Criterion

The third criteria concerns the equality of between-method variabilities, and is the basis of one of Roy's three tests. Between-subject variability is needed when one is interested in the true or real difference between the two methods giving different measurements on the same subject. Between-subject variabilities are crucial when one is concerned about the true difference between the two methods (Roy, 2009). Constructively the null hypothesis of this test calls for the equality of variances of the mean measurements for both sets of measurements. From a theoretical point of view, this criterion is a valid requirement for method agreement.

However, one should consider how a violation of this criterion could occur. When items are measured on the same scale by both methods, the resulting measurements should be numerically near to each other on a consistent basis, even in the absence of agreement.

Roy (2009) points out that sometimes there can be significant difference in between-subject variabilities of the two methods. The presence of proportional bias will give rise to a significant difference in between-subject variance. If proportional bias is not present, there must be a catastrophic failure of one or both of the measurement methods, i.e. they are systematically unable to yield a pair of measurements for an item that are numerically close together.

If this was the case, this phenomenon would be immediately identifiable in the exploratory phase of the analysis, i.e. from a visual inspection of the scatter-plot, or from the correlation coefficient.

A complicated model framework, such as an LME model, would not be required to determine this outcome.

As such, the test of this third criterion should not be conducted simultaneously with tests for the other two. Instead it should be carried out the model building phase. Violations of criteria 1 and 2 may exist in the context of two functioning measurement systems, as demonstrated in the examples provided by Roy (2009).

1.3   Towards A Simplified Model Framework

Bland and Altman (1999) recommended simultaneous estimation of repeatability and agreement for replicate measurements. When no longer required to simultaneously test for a violation of the third criteria, we are afforded the opportunity to test the first two criteria with a simpler LME model framework (simpler when compared to Roy's model). A violation of the first criterion can be detected by inspection of the estimate for the fixed effects. A violation of the second criterion can be detected by an inspection of the residuals.

An advantage of using a simplified LME framework is that it affords the opportunity to test criteria 1 and 2 for for more than two methods of measurement simultaneously. Both Bland and Altman (1999) and Roy (2009) use as an example the simultaneous measurements of systolic blood pressure were made by each of the two experienced observers (denoted by J and R) using a sphygmomanometer, and by a semi-automatic blood pressure monitor (denoted by S). Three sets (p = 3) of readings were made in quick succession on 85 subjects. The following approach shows that it can be quickly determined which pair of methods are in agreement.

Step 1 A LME model is fitted to the entire data set. A fixed effect is used to represent the measurement methods. The random error is collected in a random effect. For the sake of simplicity, one of the measurement methods can be chosen as the base case.

Step 2 A test of criterion 1 can be implemented by an examination of the fixed effects. While the creators of the LME4 R package advised against the use of p-values, a set of confidence intervals can be determined.

Step 3 A test of criterion 2 can be implemented by an examination of the residuals. An omnibus test for all p methods together can be implemented as a Bartlett Test of homogeneity of variances Pairwise comparison may be implemented using the standard $F-$test for equality of variance. Both tests are facilitated by most statistical software packages.

The approach proposed by Roy (2009) should be retained in the overall method comparison framework, as it is still strongly preferred for computing limits of agreement and the repeatability coefficient. Roy (2009) states the importance of reporting repeatability when assessing measurement, because it measures the purest random error that is not influenced by any other factors (Barnhart et al., 2007). Roy (2009) advises against the use any kind of correlation coefficient for assessing agreement for data with replications, even though that framework allows for the computation of the overall correlation coefficient.

```
## Step 1. Load R packages
library(dplyr);  library(magrittr);  library(broom)
library(lme4)
library(GGally)


## Step 2. Get the Blood Pressure Data
source("https://raw.githubusercontent.com/DragonflyStats/MCS-code/master/blooddata.R")


## Step 3. Fit the Simplified LME Model
MCSModel <- lmer(BP ~ method + (1|subject),data=blood)


## Step 4. Test the inter-method bias criterion, i.e. criterion 1.
summary(MCSModel)
ggcoef(MCSModel)  # visual analogue


## Step 5. Augment the Blood data set with additional estimates
blood <- augment(MCSModel,blood)


## Step 6. Compare Estimate for Residal Variances
blood %>% group_by(method) %>% summarize(Sigma.sq.W = var(.resid))


## Step 7. Omnibus Test for Equality of Variances
bartlett.test(.resid~method,data=blood)


## Step 8. Pairwise Test for Equality of Variances
blood.js <- blood %>% filter(method %in% c("J","S"))
var.test(.resid~method,data=blood.js)
```

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 (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research 8*(2), 135–160.

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.