

Chapter 1

MCS Data sets

1.1 Linked replicates

In the first example provided by Carstensen et al. (2008), it was assumed that the replicates were exchangeable within each method by item stratum. Sometimes, however, replicates are taken in parallel by each of the methods, which means that the values are linked by a common environment; typically time or sampling occasion.

1.2 MCS Datasets

To properly compare both Carstensen's and Roy's methodologies, we will apply both to data sets used by both authors, and to data sets that are well known in method comparison studies literature. A discussion of each dataset is provided.

1.2.1 Data sets

Subcutaneous fat (unlinked)

Carstensen et al. (2008) presents data from a comparison of measurements of subcutaneous fat by two observers at the Steno Diabetes Center in Copenhagen. Measurements are taken on each patient three times by both methods. **The sequence of measure-**

ments is not considered to be of importance, so the replicate measurements are exchangeable within patient (item) and method.

Oximetry Example (linked)

An example of this is the oximetry study, done at the Royal Childrens Hospital in Melbourne to examine the agreement between pulse oximetry and co-oximetry in small babies. **Each patient was measured three times by each method; performed at three different times for each infant.** There were 61 babies in the study, of these, four had only measurements on two occasions, and one on only one occasion.

Blood Data

Bland and Altman (1999) presents a set of blood pressure data from a study in which simultaneous measurement were made by two experienced observers (denoted J and R) using a sphygmomanometer and by a semi-automatic BP monitor (denoted S). **Three readings were recorded in quick succession.**

1.2.2 Comparison of methodologies

1.2.3 datasets

The two LME approaches are assessed as to whether the respective estimates for the standard deviation of differences accord. Method comparison data sets presented in either of the relevant papers are analyzed using both methods. Carstensen et al. (2008) demonstrates an approach using the “fat” and “ox” data sets. Roy (2009) uses the “Blood” dataset, which featured in Bland and Altman (1999) , a key paper that first addressed method comparison study in the context of replicate measurements. As this dataset features replicate measurements from three methods, three method comparison studies are possible. A third data set “hamlett” is presented in Hamlett et al. (2004) a paper that Roy’s methodology draws from.

1.3 Using other data sets

Two further data sets applied to both methodologies are the “Cardiac” and “PEFR” , which are both contained on Carstensen’s MethComp package. Importantly a difference emerges in the estimates provided by both samples.

1.3.1 Peak Expiry Flow Rate Data

The PEFR data describes the comparison of two measurements of peak expiratory flow rate (PEFR). One of these measurements uses a “Large” meter and the other a “Mini” meter.

Two measurements were made with a Wright peak flow meter and two with a mini Wright meter, in **random order**. All measurements were taken by the same observer, using the same two instruments. (These data were collected to demonstrate the statistical method and provide no evidence on the comparability of these two instruments.)

1.3.2 Cardiac data

For each subject cardiac output is measured repeatedly (three to six times) by impedance cardiography (IC) and radionuclide ventriculography (RV).

It is not entirely clear from the source whether the replicates are exchangeable within (method,item) or whether they represent pairs of measurements.

From the description it looks as if replicates are linked between methods, but in the paper they are treated as if they were not. Source The dataset is adapted from table 4 in Bland and Altman (1999). To apply the same logic to mixed effects models one has to measure the influence of a particular higher level unit on the estimates of a higher level predictor.

This means that the mixed effect model has to be adjusted to neutralize the unit’s influence on that estimate, while at the same time allowing the unit’s lower-level cases to help estimate the effects of the lower-level predictors in the model.

This procedure is based on a modification of the intercept and the addition of a dummy variable for the cases that might be influential.

Influence.ME provides several measures of influential cases, and is specifically designed for use with mixed effects regression models using the afore mentioned modified intercept and dummy approach.

Using both ?real? and simulated data from Social Science applications of mixed effects models, five tools to detect influential cases which are available in the package will be discussed:

- Cook?s Distance
- DFBETAS
- Percent change of the estimated parameter magnitude
- Changes in statistical significance of parameter estimates
- Changes in the sign of parameter estimates

In contrast with other algorithms for detecting influential cases, influence.ME is capable to uncover groups of cases that are influential. Since this rapidly becomes computationally highly intensive, additional script functions are provided that assist in manually dividing the computation into multiple sessions, or to possibly to share the computations between different computers.

influence.ME is an R package for detecting influential data in multilevel regression models (or, mixed effects models as they are referred to in the R community). The application of multilevel models has become common practice, but the development of diagnostic tools has lagged behind. Hence, we developed influence.ME, which calculates standardized measures of influential data for the point estimates of generalized multilevel models, such as DFBETAS, Cook?s distance, as well as percentile change and a test for changing levels of significance. influence.ME calculates these measures of influence while accounting for the nesting structure of the data. A paper detailing

this package was published in the R Journal (available from the R Journal (.PDF) and my researchgate.net profile).

influence.ME depends on lme4. As the authors of lme4 have completely revised the inner workings of lme4 and are currently releasing version 1.0,

1.4 influence.ME: Tools for detecting influential data in mixed effects models

- influence.ME provides a collection of tools for detecting influential cases in generalized mixed effects models.
- It analyses models that were estimated using lme4.
- The basic rationale behind identifying influential data is that when iteratively single units are omitted from the data, models based on these data should not produce substantially different estimates.
- To standardize the assessment of how influential a (single group of) observation(s) is, several measures of influence are common practice, such as DFBETAS and Cook's Distance. In addition, we provide a measure of percentage change of the fixed point estimates and a simple procedure to detect changing levels of significance.
- Despite the increasing popularity of multilevel regression models, the development of diagnostic tools lagged behind. Typically, in the social sciences multilevel regression models are used to account for the nesting structure of the data, such as students in classes, migrants from origin-countries, and individuals in countries. The strength of multilevel models lies in analyzing data on a large number of groups with only a couple of observations within each group, such as for instance students in classes.

- Nevertheless, in the social sciences multilevel models are often used to analyze data on a limited number of groups with per group a large number of observations. A typical example would be the analysis of data on individuals nested within countries. By nature, only a limited number of countries exists. In practice, typical country-comparative analyses are based on about 25 countries. With such a small number of groups (e.g. countries), observations on a single group can easily be overly influential to the outcomes. This means that the conclusions based on the multilevel regression model could no longer hold when a single group is removed from the data.
- In our recent publication in the R Journal, we introduce `influence.ME`, software that provides tools for detecting influential data in multilevel regression models (or: in mixed effects models, as these are commonly referred to in statistics). `influence.ME` is a publically available R package that evaluates multilevel regression models that were estimated with the `lme4.0` package. It calculates standardized measures of influential data for the point estimates of generalized mixed effects models, such as DFBETAS, Cook's distance, as well as percentile change and a test for changing levels of significance. `influence.ME` calculates these measures of influence while accounting for the nesting structure of the data. The package and measures of influential data are introduced, a practical example is given, and strategies for dealing with influential data are suggested.
- With this publication, and of course with the software that was available for quite some time, we hope to contribute to a better usage of multilevel regression models. The provided example and guidelines were geared towards applications in the social sciences, but are applicable in all disciplines.

Chapter 2

Application to Method Comparison Studies

2.1 Application to MCS

Let $\hat{\beta}$ denote the least square estimate of β based upon the full set of observations, and let $\hat{\beta}^{(k)}$ denoted the estimate with the k^{th} case excluded.

2.2 Grubbs' Data

For the Grubbs data the $\hat{\beta}$ estimated are $\hat{\beta}_0$ and $\hat{\beta}_1$ respectively. Leaving the fourth case out, i.e. $k = 4$ the corresponding estimates are $\hat{\beta}_0^{-4}$ and $\hat{\beta}_1^{-4}$

$$Y^{-Q} = \hat{\beta}^{-Q} X^{-Q} \tag{2.1}$$

When considering the regression of case-wise differences and averages, we write $D^{-Q} = \hat{\beta}^{-Q} A^{-Q}$

	F	C	D	A
1	793.80	794.60	-0.80	794.20
2	793.10	793.90	-0.80	793.50
3	792.40	793.20	-0.80	792.80
4	794.00	794.00	0.00	794.00
5	791.40	792.20	-0.80	791.80
6	792.40	793.10	-0.70	792.75
7	791.70	792.40	-0.70	792.05
8	792.30	792.80	-0.50	792.55
9	789.60	790.20	-0.60	789.90
10	794.40	795.00	-0.60	794.70
11	790.90	791.60	-0.70	791.25
12	793.50	793.80	-0.30	793.65

$$Y^{(k)} = \hat{\beta}^{(k)} X^{(k)} \quad (2.2)$$

Consider two sets of measurements , in this case F and C , with the vectors of case-wise averages A and case-wise differences D respectively. A regression model of differences on averages can be fitted with the view to exploring some characteristics of the data.

When considering the regression of case-wise differences and averages, we write

$$D^{-Q} = \hat{\beta}^{-Q} A^{-Q} \quad (2.3)$$

Let $\hat{\beta}$ denote the least square estimate of β based upon the full set of observations, and let $\hat{\beta}^{(k)}$ denoted the estimate with the k^{th} case excluded.

For the Grubbs data the $\hat{\beta}$ estimated are $\hat{\beta}_0$ and $\hat{\beta}_1$ respectively. Leaving the fourth case out, i.e. $k = 4$ the corresponding estimates are $\hat{\beta}_0^{-4}$ and $\hat{\beta}_1^{-4}$

$$Y^{(k)} = \hat{\beta}^{(k)} X^{(k)} \quad (2.4)$$

Consider two sets of measurements , in this case F and C , with the vectors of case-wise averages A and case-wise differences D respectively. A regression model of differences on averages can be fitted with the view to exploring some characteristics of the data.

Call: `lm(formula = D ~ A)`

Coefficients: (Intercept)	A
-37.51896	0.04656

2.2.1 Influence measures using R

R provides the following influence measures of each observation.

	dfb.1_	dfb.A	dffit	cov.r	cook.d	hat
1	0.42	-0.42	-0.56	1.13	0.15	0.18
2	0.17	-0.17	-0.34	1.14	0.06	0.11
3	0.01	-0.01	-0.24	1.17	0.03	0.08
4	-1.08	1.08	1.57	0.24	0.56	0.16
5	-0.14	0.14	-0.24	1.30	0.03	0.13
6	-0.00	0.00	-0.11	1.31	0.01	0.08
7	-0.04	0.04	-0.08	1.37	0.00	0.11
8	0.02	-0.02	0.15	1.28	0.01	0.09
9	0.69	-0.68	0.75	2.08	0.29	0.48
10	0.18	-0.18	-0.22	1.63	0.03	0.27
11	-0.03	0.03	-0.04	1.53	0.00	0.19
12	-0.25	0.25	0.44	1.05	0.09	0.12

Bibliography

- 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.
- 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.