

# SCRATCH

Kevin O'Brien

February 3, 2017

# Contents

0.1	MCS Data Sets . . . . .	8
0.2	Introduction . . . . .	8
0.2.1	Overview of R implementations . . . . .	9
0.3	Computation and Notation . . . . .	11
0.4	Liao Shaio . . . . .	11
0.4.1	Limits of agreement for Carstensen's data . . . . .	11
0.5	Hamlett and Lam . . . . .	13
0.5.1	Inappropriate Techniques for MCS . . . . .	14
0.5.2	Links and Papers . . . . .	14
0.6	Paired T tests . . . . .	16
0.7	Methods of assessing agreement . . . . .	16
0.7.1	Equivalence and Interchangeability . . . . .	17
<b>1</b>	<b>Appendix</b>	<b>19</b>
1.1	Escaramis . . . . .	21
1.1.1	Background . . . . .	21
1.1.2	Methods . . . . .	21
1.1.3	Results . . . . .	21
1.1.4	Conclusions . . . . .	22
1.2	Schabenberger . . . . .	23

1.3	Hawkins : Diagnostics for conformity of paired quantitative measurements	24
1.4	Profile Function with "lmer"	28
1.5	Turkan's LMEs	28
1.5.1	Ordinary Least Product Regression	29
1.5.2	A regression based approach based on Bland Altman Analysis	29
1.6	Measurement Error Models	30
1.7	Case Deletion Diagnostics for LME models	30
1.7.1	Case Deletion Diagnostics	33
1.7.2	Effects on fitted and predicted values	33
1.7.3	Case Deletion Diagnostics for Mixed Models	33
1.7.4	Methods and Measures	34
1.7.5	Matrix Notation for Case Deletion	35
1.7.6	Case deletion notation	35
1.7.7	Partitioning Matrices	35
1.7.8	Case Deletion Diagnostics	35
1.7.9	Case Deletion Diagnostics for Mixed Models	35
1.7.10	Case Deletion Diagnostics	36
1.7.11	Deletion Diagnostics	36
1.7.12	Terminology for Case Deletion diagnostics	37
1.8	Work List	37
1.9	Diagnostics	38
1.9.1	Identifying outliers with a LME model object	38
1.9.2	Diagnostics for Random Effects	38
1.10	Iterative and non-iterative influence analysis	39
1.10.1	Iterative Influence Analysis	39
1.10.2	Iterative vs Non-Iterative Influence Analysis	39

1.11	Generalized Least Squares . . . . .	41
1.11.1	Introduction to Generalized Least Squares . . . . .	41
1.11.2	Introduction . . . . .	42
1.11.3	Predictors and Estimators . . . . .	44
1.11.4	Two Options . . . . .	46
1.11.5	Profile Likelihood Confidence Intervals . . . . .	46
1.12	Two-tailed testing . . . . .	47
1.13	One Tailed Testing . . . . .	47
1.14	Enabling One Tailed Testing . . . . .	47
1.15	Profile Likelihood . . . . .	48
1.16	Implementation of PL Confidence Intervals . . . . .	48
<b>2</b>	<b>Quiroz J1, Burdick RK.</b>	<b>50</b>
<b>3</b>	<b>Model Diagnostics</b>	<b>51</b>
3.1	Introduction . . . . .	51
3.1.1	Checking model assumptions . . . . .	51
3.1.2	Influence Diagnostics: Basic Idea and Statistics . . . . .	52
3.1.3	Introduction . . . . .	52
3.2	Outline of Thesis . . . . .	53
3.3	What is Influence . . . . .	54
3.3.1	Quantifying Influence . . . . .	54
3.4	Residual diagnostics . . . . .	55
3.4.1	Residuals . . . . .	55
3.4.2	Residuals diagnostics in mixed models . . . . .	55
3.5	Extension of technique to LME Models . . . . .	57
3.6	Marginal and Conditional Residuals . . . . .	58

3.6.1	Marginal and Conditional Residuals . . . . .	58
3.6.2	Marginal Residuals . . . . .	58
3.7	Standardized and studentized residuals . . . . .	59
3.7.1	Studentization . . . . .	59
3.8	Case Deletion Diagnostics . . . . .	60
3.8.1	Case Deletion Diagnostics . . . . .	60
3.9	Effects on fitted and predicted values . . . . .	60
3.9.1	Deletion Diagnostics . . . . .	60
3.9.2	Case Deletion Diagnostics for Mixed Models . . . . .	61
3.9.3	Methods and Measures . . . . .	61
3.10	Influence analysis . . . . .	62
3.10.1	Cook's 1986 paper on Local Influence . . . . .	62
3.10.2	Overall Influence . . . . .	62
3.11	Likelihood Distance . . . . .	63
3.11.1	Likelihood Distance . . . . .	63
3.12	Haslett's Analysis . . . . .	64
3.13	Iterative and non-iterative influence analysis . . . . .	64
3.13.1	Iterative Influence Analysis . . . . .	64
3.14	Matrix Notation for Case Delection . . . . .	65
3.14.1	Case deletion notation . . . . .	65
3.14.2	Partitioning Matrices . . . . .	65
3.15	CPJ's Three Propositions . . . . .	66
3.15.1	Proposition 2 . . . . .	66
3.15.2	Proposition 3 . . . . .	66
3.16	Augmented GLMs . . . . .	67
3.17	Covariance Parameters . . . . .	68

3.18	Terminology for Case Deletion diagnostics . . . . .	69
3.19	The CPJ Paper . . . . .	70
3.19.1	Case-Deletion results for Variance components . . . . .	70
3.19.2	CPJ Notation . . . . .	70
<b>4</b>	<b>Roy2013</b>	<b>72</b>
<b>5</b>	<b>Appendices</b>	<b>73</b>
5.1	Sherman Morrison Woodbury Formula . . . . .	73
5.2	The Hat Matrix . . . . .	75
5.2.1	Hat Values for MCS regression . . . . .	75
5.3	Cross Validation . . . . .	76
5.3.1	Cross Validation: Updating standard deviation . . . . .	76
5.4	Updating Estimates . . . . .	78
5.4.1	Updating of Regression Estimates . . . . .	78
5.4.2	Updating Standard deviation . . . . .	78
5.4.3	Updating of Regression Estimates . . . . .	78
5.4.4	Updating Standard deviation . . . . .	79
5.4.5	Updating of Regression Estimates . . . . .	79
5.4.6	Updating Regression Estimates . . . . .	80
5.4.7	Inference on intercept and slope . . . . .	80
5.5	Lesaffre's paper. . . . .	82
1.	Agreement and Method Comparison Studies	
(a)	What is Agreement?	
(b)	Repeatability	
(c)		

(d)

(e)

## 2. Bland Altman Single Observations

(a)

(b)

## 3. Alternative Methods

(a) Deming Regression

(b) Mountain Plot

(c) Bartko's Ellipse

(d) Formal Tests and Procedures

## 4. Replicate Observations

## 5. LME models

## 6. Estimation and Algorithms

(a) ML and REML estimation

(b) MINQUE

(c)

## 7. Residual Diagnostics

(a) Marginal and Conditional Diagnostics

(b) Scaled Residuals

## 8. Influence Diagnostics

- (a) Underlying Concepts
- (b) Managing the Covariance Parameters
- (c) Predicted Values, PRESS Residual and the PRESS Statistic
- (d) Leverage
- (e) Internally and Externally Studentized Residuals
- (f) DFFITs and MDFFITs
- (g) Covariance Ratio and Trace
- (h) Likelihood Distance
- (i) Non-iterative Update Procedures



## 0.1 MCS Data Sets

1. Blood Data
  2. Cardiac Data
  3. Nadler Hurley
- Introduction to Method Comparison Studies
    - Accuracy and Precision
    - Repeatability (Bland Altman 1999)
    - Barnharts Paper
    -
  - Bland and Altman Plot
    - Bland and Altman 1983 and 86
    - Limits of Agreement
    - 
    -

## 0.2 Introduction

Outliers and detection of influent observations is an important step in the analysis of a data set. There are several ways of evaluating the influence of perturbations in the data set and in the model given the parameter estimates.

### 0.2.1 Overview of R implementations

Further to previous material, an appraisal of the current state of development (or lack thereof) for current implemenations for LME models, particularly for `nlme` and `lme4` fitted models.

Crucially, a review of internet resources indicates that almost all of the progress in this regard has been done for `lme4` fitted models, specifically the *Influence.ME* R package. (Nieuwenhuis et al 2014) Conversely there is very little for `nlme` models. One would immediately look at the current development workflow for both packages.

As an aside, Douglas Bates was arguably the most prominent R developer working in the LME area. However Bates has now prioritised the development of LME models in another computing environment , i.e Julia.

With regards to `nlme`, the package is now maintained by the R core development team. The most recent major text is by Galecki & Burzykowski, who have published *Linear Mixed Effects Models using R*. Also, the accompanying R package, nlmeU package is under current development, with a version being released 0.70 – 3.

The `lme4` pacakge is used to fit linear and generalized linear mixed-effects models in the R environment. The `lme4` package is also under active development, under the leadership of Ben Bolker (McMaster Uni., Canada).

### Important Consideration for MCS

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 identified shortly. To advance the ideas that eminate from Roys’ 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 `textbfinfluence` can accomplish is merited.

## 0.3 Computation and Notation

with  $\mathbf{V}$  unknown, a standard practice for estimating  $\mathbf{X}\boldsymbol{\beta}$  is to estimate the variance components  $\sigma_j^2$ , compute an estimate for  $\mathbf{V}$  and then compute the projector matrix  $A$ ,  $\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{A}\mathbf{Y}$ .

Zewotir remarks that  $\mathbf{D}$  is a block diagonal with the  $i$ -th block being  $u\mathbf{I}$

## 0.4 Liao Shiao

Lai et Shiao is interesting in that it extends the usual method comparison study question. It correctly identifies LME models as a methodology that can be used to make such questions tractable. The Data Set used in their examples is unavailable for independent use. Therefore, for the sake of consistency, a data set will be simulated based on the Blood Data that will allow for extra variables.

A Study of the Bland-Altman Plot and its Associated Methodology

Joseph G. Voelkel Bruce E. Siskowski

## 0.5 Limits of agreement for Carstensen's data

Carstensen et al. (2008) describes the calculation of the limits of agreement (with the inter-method bias implicit) for both data sets, based on his formulation;

$$\hat{\alpha}_1 - \hat{\alpha}_2 \pm 2\sqrt{2\hat{\tau}^2 + \hat{\sigma}_1^2 + \hat{\sigma}_2^2}.$$

For the 'Fat' data set, the inter-method bias is shown to be 0.045. The limits of agreement are  $(-0.23, 0.32)$

Carstensen demonstrates the use of the interaction term when computing the limits of agreement for the 'Oximetry' data set. When the interaction term is omitted,

the limits of agreement are  $(-9.97, 14.81)$ . Carstensen advises the inclusion of the interaction term for linked replicates, and hence the limits of agreement are recomputed as  $(-12.18, 17.12)$ .

## 0.6 Hamlett and Lam

The methodology proposed by ? is largely based on Hamlett et al. (2004), which in turn follows on from Lam et al. (1999).

The desired outcome of this research is to

- Formulate a methodology that represents Best practice in Method Comparison Studies. Indeed the methodology is envisaged to advance what is considered best practice, inter alia, by making diagnostics procedures a standard part of MCS.
- Provide for ease of use such that non-statisticians can master and implement the method, with a level of training that one would expect as part of a Professional CPD programme.

Apropos of the matter of ease-of-use, certain assumptions must be made.

The user has a reasonable amount of computer literacy. The user would have a reasonable understanding of statistics, consistent with an undergraduate statistics module. That is to say, that the user is acquainted with the idea of  $p$ -values.

Easy to follow set of instructions to properly implement the method.

Linear Mixed Effects Models can be implemented by using one of the following R packages. lme4 nlme

The first package to be introduced was nlme, developed by Jose Pinheiro and Douglas Bates ( Authors of the the companion textbook, NAME)

As this package has been under ongoing development for quite a long time, it is now allows for a lot of complex LME implementations. Furthermore, nlme is one of the base R packages. That is to say, when one downloads and installs R, nlme is automatically installed also, and can be called immediately.

Having said that, the authors have pointed to several limitations of the overall methodology through R. The original developers have both left the project, but other

statisticians have taken over the development, and indeed a new version of nlme was released.

LME4 is a more recent package. at a glance, the syntax is easier, but the development is less advanced. There are several functionalities that can not be implemented with lme4 yet. As an example - CHAP5 in PB - has no equivalent in LME4. Indeed no textbook exists to co-incide with LME4.

The main author, Douglas Bates, has turned his attention to development of LME models in the Julia programming language.

The nlmeU package is described by its authors as an extesntion of the nlme package, and indeed provides for additionally functionality. The package is also useful as it serves as a companion piece to the book by Galecki and Burzwhatski.

The nlme package also allows for the specification of GLS models.

## **Objects and Classes**

The main nlme object is an `nlme` model.

The main lme4 object is called an `lmer` model

The lattice package is used for graphical methods.

Model Diagnostics with `nlme`

### **0.6.1 Inappropriate Techniques for MCS**

### **0.6.2 Links and Papers**

Westgard Statistics - <http://www.westgard.com/lesson23.htm>

## Measurement Systems Analysis

The topic of measurement sensitivity analysis (MSA, also known as Gauge R&R) is prevalent in industrial statistics (i.e Six Sigma).

There is extensive literature that covers the area. For the sake of brevity, we will use Cano et al.

For sake of clarity, Cano's definitions of repeatability and reproducibility are listed, with added emphasis.

Reproducibility is rarely, if ever, discussed in the domain of Method Comparison Studies. This may be due to the fact that prevalent methodologies can be used for the problem. However the methodologies proposed by this research can easily be extended.

## Leave-One-Out Diagnostics with lmeU

Galecki et al provide a brief the matter of LME influence diagnostics in their book.

The command `lmeU` fits a model with a particular subject removed. The identifier of the subject to be removed is passed as the only argument

A plot of the per-observation diagnostics individual subject log-likelihood contributions can be rendered.

### The addition of an extra factor

Interaction terms are featured in ANOVA designs.

My search just now found no mention of Cook's distance or influence measures.

The closest I found was an unanswered question on this from April 2003 (<http://finzi.psych.upenn.edu>)

Beyond that, there is an excellent discussion of "Examining a Fitted Model" in Sec. 4.3 (pp. 174-197) of Pinheiro and Bates (2000) *Mixed-Effects Models in S and S-Plus* (Springer).



Pinheiro and Bates decided NOT to include plots of Cook's distance among the many diagnostics they did provide. However, `plot(fit.lme)` plots 'standardized residuals' vs. predicted or 'fitted values'. Wouldn't points with large influence stand apart from the crowd in terms of 'fitted value'?

Of course, there are many things other one could do to get at related information, including reading the code for 'influence' and 'lme', and figure out from that how to write an 'influence' method for an 'lme' object.

## 0.7 Paired T tests

This method can be applied to test for statistically significant deviations in bias. This method can be potentially misused for method comparison studies.

It is a poor measure of agreement when the rater's measurements are perpendicular to the line of equality[Hutson et al]. In this context, an average difference of zero between the two raters, yet the scatter plot displays strong negative correlation.

### Components in assessing agreement

1. The degree of linear relationship between the two sets
2. The amount of bias as represented by the difference in the means
3. The Differences in the two variances.

## 0.8 Methods of assessing agreement

1. Pearson's Correlation Coefficient
2. Intraclass correlation coefficient

3. Bland Altman Plot
4. Bartko's Ellipse (1994)
5. Blackwood Bradley Test
6. Lin's Reproducibility Index
7. Luiz Step function

Bland and Altman attend to the issue of repeated measures in 1996.

Repeated measurements on several subjects can be used to quantify measurement error, the variation between measurements of the same quantity on the same individual.

Bland and Altman discuss two metrics for measurement error; the within-subject standard deviation ,and the correlation coefficient.

The above plot incorporates both the conventional limits of agreement ( the inner pair of dashed lines), the 't' limits of agreement ( the outer pair of dashed lines) centred around the inter-method bias (indicated by the full line). This plot is intended for expository purposes only, as the sample size is small.

### **0.8.1 Equivalence and Interchangeability**

Limits of agreement are intended to analyse equivalence. How this is assessed is the considered judgement of the practitioner. In Bland and Altman (1986) an example of good agreement is cited. For two methods of measuring 'oxygen saturation', the limits of agreement are calculated as (-2.0,2.8).A practitioner would ostensibly find this to be sufficiently narrow.

If the limits of agreement are not clinically important, which is to say that the differences tend not to be substantial, the two methods may be used interchangeably. Dunn

(2002) takes issue with the notion of ‘equivalence’, remarking that while agreement indicated equivalence, equivalence does not reflect agreement.

# Chapter 1

## Appendix

### Bayesian BA - Philip J Schluter

Bayesian Bland Altman Approaches A multivariate hierarchical Bayesian approach to measuring agreement in repeated measurement method comparison studies

\*<http://www.biomedcentral.com/1471-2288/9/6>\*

### Background

Assessing agreement in method comparison studies depends on two fundamentally important components; validity (the between method agreement) and reproducibility (the within method agreement).

The Bland-Altman limits of agreement technique is one of the favoured approaches in medical literature for assessing between method validity. However, few researchers have adopted this approach for the assessment of both validity and reproducibility.

This may be partly due to a lack of a flexible, easily implemented and readily available statistical machinery to analyse repeated measurement method comparison data.

## Methods

Adopting the Bland-Altman framework, but using Bayesian methods, we present this statistical machinery. Two multivariate hierarchical Bayesian models are advocated, one which assumes that the underlying values for subjects remain static (exchangeable replicates) and one which assumes that the underlying values can change between repeated measurements (non-exchangeable replicates).

## Results

We illustrate the salient advantages of these models using two separate datasets that have been previously analysed and presented; (i) assuming static underlying values analysed using both multivariate hierarchical Bayesian models, (ii) assuming each subject's underlying value is continually changing quantity and analysed using the non-exchangeable replicate multivariate hierarchical Bayesian model.

**Conclusion** These easily implemented models allow for full parameter uncertainty, simultaneous method comparison, handle unbalanced or missing data, and provide estimates and credible regions for all the parameters of interest. Computer code for the analyses is also presented, provided in the freely available and currently cost free software package WinBUGS. [http://www.mrc-biostat.cam.ac.uk/winbugs/](#)

## Bayesian Approach

A multivariate hierarchical Bayesian approach to measuring agreement in repeated measurement method comparison studies PJ Schluter - BMC medical research methodology, 2009 - [biomedcentral.com](http://biomedcentral.com)

- Assessing agreement in method comparison studies depends on two fundamentally important components; validity (the between method agreement) and reproducibility (the within method agreement).

- The Bland-Altman limits of agreement technique is one of the f

## **1.1 Escaramis**

### **1.1.1 Background**

In an agreement assay, it is of interest to evaluate the degree of agreement between the different methods (devices, instruments or observers) used to measure the same characteristic. We propose in this study a technical simplification for inference about the total deviation index (TDI) estimate to assess agreement between two devices of normally-distributed measurements and describe its utility to evaluate inter- and intra-rater agreement if more than one reading per subject is available for each device.

### **1.1.2 Methods**

We propose to estimate the TDI by constructing a probability interval of the difference in paired measurements between devices, and thereafter, we derive a tolerance interval (TI) procedure as a natural way to make inferences about probability limit estimates. We also describe how the proposed method can be used to compute bounds of the coverage probability.

### **1.1.3 Results**

The approach is illustrated in a real case example where the agreement between two instruments, a handle mercury sphygmomanometer device and an OMRON 711 automatic device, is assessed in a sample of 384 subjects where measures of systolic blood pressure were taken twice by each device. A simulation study procedure is implemented to evaluate and compare the accuracy of the approach to two already established meth-

ods, showing that the TI approximation produces accurate empirical confidence levels which are reasonably close to the nominal confidence level.

#### 1.1.4 Conclusions

The method proposed is straightforward since the TDI estimate is derived directly from a probability interval of a normally-distributed variable in its original scale, without further transformations. Thereafter, a natural way of making inferences about this estimate is to derive the appropriate TI. Constructions of TI based on normal populations are implemented in most standard statistical packages, thus making it simpler for any practitioner to implement our proposal to assess agreement.

Lin defined the TDI as the boundary,  $\kappa_P$  which captures a large proportion  $p$  of paired based differences from two devices or observers within the boundary.

The value of  $\kappa_P$  that yields  $P(|D| < \kappa_p) = p$  where  $D$  is the paired-difference variate.

$$\kappa_P = F^{-1}(p) = \sigma_D \sqrt{\chi^2(p, 1, \mu_D^2/\sigma_d^2)}$$

$$\kappa_P = Z_{\frac{1+p}{2}} \|\varepsilon\|$$

Tolerance Interval around the TDI estimate

$$\hat{\kappa}_p = \hat{\mu}_D = Z_{p_i} \sigma_d$$

Coverage Probability is another user friendly measure of agreement which is related to the computation of the TDI.

## 1.2 Schabenberger

*schab* examines the use and implementation of influence measures in LME models.

Influence is understood to be the ability of a single or multiple data points, through their presences or absence in the data, to alter important aspects of the analysis, yield qualitatively different inferences, or violate assumptions of the statistical model (*schabenberger*).

Outliers are the most noteworthy data points in an analysis, and an objective of influence analysis is how influential they are, and the manner in which they are influential.

*schab* describes a simple procedure for quantifying influence. Firstly a model should be fitted to the data, and estimates of the parameters should be obtained. The second step is that either single or multiple data points, specifically outliers, should be omitted from the analysis, with the original parameter estimates being updated.

This is known as ‘*leave one out* *leave k out*’ analysis. The final step of the procedure is comparing the sets of estimates computed from the entire and reduced data sets to determine whether the absence of observations changed the analysis.

*schabenberger* notes that it is not always possible to derive influence statistics necessary for comparing full- and reduced-data parameter estimates.

In recent years, mixed models have become invaluable tools in the analysis of experimental and observational data. In these models, more than one term can be subject to random variation. Mixed model technology enables you to analyze complex experimental data with hierarchical random processes, temporal, longitudinal, and spatial data, to name just a few important applications.

*schab* remarks that the concept of critiquing the model-data agreement applies in mixed models in the same way as in linear fixed-effects models. In fact, because of



the more complex model structure, you can argue that model and data diagnostics are even more important. For example, you are not only concerned with capturing the important variables in the model. You are also concerned with “distributing them correctly between the fixed and random components of the model. The mixed model structure presents unique and interesting challenges that prompt us to reexamine the traditional ideas of influence and residual analysis.

### **1.3 Hawkins : Diagnostics for conformity of paired quantitative measurements**

- Matched pairs data arise in many contexts in case-control clinical trials, for example, and from cross-over designs. They also arise in experiments to verify the equivalence of quantitative assays. This latter use (which is the main focus of this paper) raises difficulties not always seen in other matched pairs applications.
- Since the designs deliberately vary the analyte levels over a wide range, issues of variance dependent on mean, calibrations of differing slopes, and curvature all need to be added to the usual model assumptions such as normality.
- Violations in any of these assumptions invalidate the conventional matched pairs analysis.
- A graphical method, due to Bland and Altman, of looking at the relationship between the average and the difference of the members of the pairs is shown to correspond to a formal testable regression model.
- Using standard regression diagnostics, one may detect and diagnose departures from the model assumptions and remedy them for example using variable trans-

formations. Examples of different common scenarios and possible approaches to handling them are shown.

A multi-Rate nonparametric test of agreement and corresponding agreement plot

- Published in: Computational Statistics and Data Analysis 54(2010)109-119 - Author: Alan D. Hutson, University of Buffalo

This approach takes advantage of readily available tests of uniformity found in most statistical software packages. Such tests include the KS d statistic, the Anderson Darling Statistic and the Cramer-Von Mises statistical test for univariate data.

An important aspect of this approach is the "Agreement Region".

## Roy Test

Roys Tests (Roy 2009) Roy 2009 devised an LME based Testing approach to the MCS problem, based on earlier work by Hamlett et al. Roy 2009 presents a series of three formal hypothesis tests for assessing agreement between two methods of measurement. Roy also alludes to some of the current shortcomings of the approach.

Comparing different model specifications with LRT tests

- Roy 2007 - Roy 2009 - Hamlett et al. - Roy Leiva 2011

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

```
> 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. There three formulations share the same structure, and can be specified by making slight alterations of the code for the Reference Model.

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")
```

Nested Model (Within item Variability)

```
> 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")
```

Nested Model (Overall Variability) Additionally there is a third nested model, that can be used to test overall variability, substantively a 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")
```

ANOVAs for Original Fits 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 mode as arguments to the command `anova()`. The figure below displays the three tests described by Roy (2009).

```
> testB    = anova(Ref.Fit,NMB.fit)                # Between-Subject Vari
> testW    = anova(Ref.Fit,NMW.fit)                # Within-Subject Variabil
> test0     = anova(Ref.Fit,NM0.fit)                # Overall Variabilities
```

## 1.4 Profile Function with "lmer"

The `profile()` function for lmer models is now available in the latest version of lme4, to be installed by typing:

```
install.packages("lme4",repos="http://r-forge.r-project.org")  
  
also
```

The `mle` function from the stats4 package is a wrapper of `optim`, which makes it quite easy to produce profile likelihood computations.

See `help("profile,mle-method", package = "stats4")` for more information.

<http://people.upei.ca/hstryhn/stryhn208.pdf>

The profile likelihood (or likelihood or likelihood ratio) method is applicable to all likelihood based statistical analysis and is generally less sensitive to the difficulties encountered by Wald-Tyoe CIs.

## 1.5 Turkan's LMEs

The linear mixed model is considerably sensitive to outliers and influential observations. It is known that outliers and influential observations affect substantially the results of analysis. So it is very important to be aware of these observations.

Some diagnostics which are analogue of diagnostics in multiple linear regression were developed to detect outliers and influential observations in the linear mixed model.

*In this paper, the new diagnostic measure which is analogue of the Pena's influence statistic is developed for the linear mixed model.*

Estimation and Building blocks in LME models

$$\hat{u} = DZ^T H^{-1}(y - X\hat{\beta})$$

$$\hat{y} = (I_n - H^{-1})y + H^{-1}X\hat{\beta}$$

The proposed diagnostic Measure.

### **1.5.1 Ordinary Least Product Regression**

Ludbrook (1997) states that the grouping structure can be straightforward, but there are more complex data sets that have a hierarchical(nested) model.

Observations between groups are independent, but observations within each groups are dependent because they belong to the same subpopulation. Therefore there are two sources of variation: between-group and within-group variance. Mean correction is a method of reducing bias.

### **1.5.2 A regression based approach based on Bland Altman Analysis**

Lu et al used such a technique in their comparison of DXA scanners. They also used the Blackwood Bradley test. However it was shown that, for particular comparisons, agreement between methods was indicated according to one test, but lack of agreement was indicated by the other.

## 1.6 Measurement Error Models

Dunn (2002) proposes a measurement error model for use in method comparison studies. Consider  $n$  pairs of measurements  $X_i$  and  $Y_i$  for  $i = 1, 2, \dots, n$ .

$$X_i = \tau_i + \delta_i \quad (1.1)$$

$$Y_i = \alpha + \beta\tau_i + \epsilon_i$$

In the above formulation is in the form of a linear structural relationship, with  $\tau_i$  and  $\beta\tau_i$  as the true values, and  $\delta_i$  and  $\epsilon_i$  as the corresponding measurement errors. In the case where the units of measurement are the same, then  $\beta = 1$ .

$$E(X_i) = \tau_i \quad (1.2)$$

$$E(Y_i) = \alpha + \beta\tau_i$$

$$E(\delta_i) = E(\epsilon_i) = 0$$

The value  $\alpha$  is the inter-method bias between the two methods.

$$z_0 = d = 0 \quad (1.3)$$

$$z_{n+1} = z_n^2 + c \quad (1.4)$$

## 1.7 Work List

1. ML v REML
2. Nested Models and LRTs
3. Generalized Least Squares

4. Diagnostics
5. Simplifying GLS
6. Paper progression



## 1.8 Diagnostics

### 1.8.1 Identifying outliers with a LME model object

The process is slightly different than with standard LME model objects, since the *influence* function does not work on lme model objects. Given *mod.lme*, we can use the plot function to identify outliers.

### 1.8.2 Diagnostics for Random Effects

Empirical best linear unbiased predictors EBLUPS provide the a useful way of diagnosing random effects.

EBLUPs are also known as “shrinkage estimators” because they tend to be smaller than the estimated effects would be if they were computed by treating a random factor as if it was fixed (West etal )

## 1.9 Iterative and non-iterative influence analysis

Schabenberger (2005) highlights some of the issue regarding implementing mixed model diagnostics.

### 1.9.1 Iterative Influence Analysis

For linear models, the implementation of influence analysis is straightforward. However, for LME models, the process is more complex. Update formulas for the fixed effects are available only when the covariance parameters are assumed to be known. A measure of total influence requires updates of all model parameters. This can only be achieved in general is by omitting observations, then refitting the model.

Schabenberger (2005) describes the choice between iterative influence analysis and non-iterative influence analysis.

### 1.9.2 Iterative vs Non-Iterative Influence Analysis

While the basic idea of influence analysis is straightforward, the implementation in mixed models can be tricky. For example, update formulas for the fixed effects are available only when the covariance parameters are assumed to be known. At most the profiled residual variance can be updated without refitting the model.

A measure of total influence requires updates of all model parameters, and the only way that this can be achieved in general is by removing the observations in question and refitting the model.

Because this **bruteforce** method involves iterative reestimation of the covariance parameters, it is termed *iterative influence analysis*. Reliance on closed-form update formulas for the fixed effects without updating the (un-profiled) covariance parameters is termed a noniterative influence analysis.

An iterative analysis seems like a costly, computationally intensive enterprise. If you compute iterative influence diagnostics for all  $n$  observations, then a total of  $n + 1$  mixed models are fit iteratively. This does not imply, of course, that the procedures execution time increases  $n$ -fold. Keep in mind that

- iterative reestimation always starts at the converged full-data estimates. If a data point is not influential, then its removal will have little effect on the objective function and parameter estimates. Within one or two iterations, the process should arrive at the reduced-data estimates.
- if complete reestimation does require many iterations, then this is important information in itself. The likelihood surface has probably changed drastically, and the reduced-data estimates are moving away

from the full-data estimates.

## 1.10 Two-tailed testing

A test for equality of variances, based on the likelihood Ratio test, is very simple to implement using existing methodologies. All that is required is to specify the reference model and the relevant nested model as arguments to the command `anova()`. The output can be interpreted in the usual way.

## 1.11 One Tailed Testing

The approach proposed by Roy deals with the question of agreement, and indeed interchangeability, as developed by Bland and Altman's corpus of work. In the view of Dunn, a question relevant to many practitioners is which of the two methods is more precise.

The relationship between precision and the within-item and between-item variability must be established. Roy establishes the equivalence of repeatability and within-item variability, and hence precision. The method with the smaller within-item variability can be deemed to be the more precise.

## 1.12 Enabling One Tailed Testing

A useful approach is to compute the confidence intervals for the ratio of within-item standard deviations (equivalent to the ratio of repeatability coefficients), which can be interpreted in the usual manner (or alternatively, the ratio of the variances). In fact, the ratio of within-item standard deviations, with the attendant confidence interval, can be determined using a single R command: `intervals()`.

Pinheiro and Bates (pg 93-95) give a description of how confidence intervals for the variance components are computed. Furthermore a complete set of confidence

intervals can be computed to complement the variance component estimates. However, to facilitate one tailed testing, What is required is the computation of the variance ratios of within-item and between-item standard deviations.

A naive approach would be to compute the variance ratios by relevant F distribution quantiles. However, the question arises as to the appropriate degrees of freedom. However, Douglas Bates has stated that an alternative approach is required (i.e. Profile Likelihoods)

”The omission of standard errors on variance components is intentional. The distribution of an estimator of a variance component is highly skewed and obtaining an estimate of the standard deviation of a skewed distribution is not very useful. A much better approach is based on profiling the objective function.” (Douglas Bates May 2012)

## 1.13 Profile Likelihood

Normal-based confidence intervals for a parameter of interest are inaccurate when the sampling distribution of the estimate is skewed. The technique known as profile likelihood can produce confidence intervals with better coverage. It may be used when the model includes only the variable of interest or several other variables in addition. Profile-likelihood confidence intervals are particularly useful in nonlinear models.

Profile likelihood confidence intervals are based on the log-likelihood function.

## 1.14 Implementation of PL Confidence Intervals

The suitable calculation of confidence limits for this variance ratio are to be computed using the profile likelihood approach. The R package `profilelikelihood` will

be assessed for feasibility, particularly the command `profilelikelihood.lme()`

# Chapter 2

## Model Diagnostics

### 2.1 Introduction

In classical linear models model diagnostics have been become a required part of any statistical analysis, and the methods are commonly available in statistical packages and standard textbooks on applied regression. However it has been noted by several papers that model diagnostics do not often accompany LME model analyses.

#### 2.1.1 Checking model assumptions

In classical linear regression, it is important to carry out model diagnostic techniques to determine whether or not the distributional assumptions are satisfied. Model diagnostics are also used to determine the influence of unusual observations.

Schabenberger (2004) describes the examination of model-data agreement as comprising several elements; residual analysis, goodness of fit, collinearity diagnostics and influence analysis.

### 2.1.2 Influence Diagnostics: Basic Idea and Statistics

The general idea of quantifying the influence of one or more observations relies on computing parameter estimates based on all data points, removing the cases in question from the data, refitting the model, and computing statistics based on the change between full-data and reduced-data estimation. Influence statistics can be coarsely grouped by the aspect of estimation that is their primary target:

- overall measures compare changes in objective functions: (restricted) likelihood distance (Cook and Weisberg 1982, Ch. 5.2)
- influence on parameter estimates: Cook's (Cook 1977, 1979), MDFFITS (Belsley, Kuh, and Welsch 1980, p. 32)
- influence on precision of estimates: CovRatio and CovTrace
- influence on fitted and predicted values: PRESS residual, PRESS statistic (Allen 1974), DFFITS (Belsley, Kuh, and Welsch 1980, p. 15)
- outlier properties: internally and externally studentized residuals, leverage

### 2.1.3 Introduction

The linear mixed effects model is a useful methodology for fitting a wide range of models. However, linear mixed effects models are known to be sensitive to outliers. ? advises that identification of outliers is necessary before conclusions may be drawn from the fitted model.

Standard statistical packages concentrate on calculating and testing parameter estimates without considering the diagnostics of the model.

The assessment of the effects of perturbations in data, on the outcome of the analysis, is known as statistical influence analysis. Influence analysis examines the robustness



of the model.

Influence analysis methodologies have been used extensively in classical linear models, and provided the basis for methodologies for use with LME models.

Computationally inexpensive diagnostics tools have been developed to examine the issue of influence (?). Studentized residuals, error contrast matrices and the inverse of the response variance covariance matrix are regular components of these tools.

## 2.2 Outline of Thesis

Thus the study of method comparison is introduced. The intention of this thesis is to progress the study of method comparison studies, using a statistical method known as Linear mixed effects models. Chapter two shall describe linear mixed effects models, and how the use of the linear mixed effects models have so far extended to method comparison studies. Implementations of important existing work shall be presented, using the R programming language.

Model diagnostics are an integral component of a complete statistical analysis. In chapter three model diagnostics shall be described in depth, with particular emphasis on linear mixed effects models, further to chapter two.

For the fourth chapter, important linear mixed effects model diagnostic methods shall be extended to method comparison studies, and proposed methods shall be demonstrated on data sets that have become well known in literature on method comparison. The purpose is to both calibrate these methods and to demonstrate applications for them. The last chapter shall focus on robust measures of important parameters such as agreement.

## 2.3 What is Influence

Broadly defined, “influence” is understood as the ability of a single or multiple data points, through their presence or absence in the data, to alter important aspects of the analysis, yield qualitatively different inferences, or violate assumptions of the statistical model. The goal of influence analysis is not primarily to mark data points for deletion so that a better model fit can be achieved for the reduced data, although this might be a result of influence analysis (Schabenberger, 2004).

Influence is defined as the ‘ability of a single or multiple data points, through their presence or absence

### 2.3.1 Quantifying Influence

The basic procedure for quantifying influence is simple as follows:

- Fit the model to the data and obtain estimates of all parameters.
- Remove one or more data points from the analysis and compute updated estimates of model parameters.
- Based on full- and reduced-data estimates, contrast quantities of interest to determine how the absence of the observations changes the analysis.

## 2.4 Residual diagnostics

For classical linear models, residual diagnostics are typically implemented as a plot of the observed residuals and the predicted values. A visual inspection for the presence of trends inform the analyst on the validity of distributional assumptions, and to detect outliers and influential observations.

In linear mixed effects models, diagnostic techniques may consider ‘conditional’ residuals. A conditional residual is the difference between an observed value  $y_i$  and the conditional predicted value  $\hat{y}_i$ .

$$\epsilon_{i|j} = y_i - \hat{y}_i = y_i - (X_i \hat{\beta} + Z_i \hat{b}_i)$$

However, using conditional residuals for diagnostics presents difficulties, as they tend to be correlated and their variances may be different for different subgroups, which can lead to erroneous conclusions.

### 2.4.1 Residuals

The computation of internally studentized residuals relies on the diagonal entries of  $V(\hat{\theta}) - Q(\hat{\theta})$ , where  $Q(\hat{\theta})$  is computed as

$$Q(\hat{\theta}) = X(X'Q(\hat{\theta})^{-1}X)X^{-1}$$

Externally studentized residual require iterative influence analysis or a profiled residuals variance.

### 2.4.2 Residuals diagnostics in mixed models

A residual is the difference between an observed quantity and its estimated or predicted value. In the mixed model you can distinguish marginal residuals  $rm$  and conditional

residuals  $rc$ . A marginal residual is the difference between the observed data and the estimated (marginal) mean.

Cook (1986) introduces powerful tools for local-influence assessment and examining perturbations in the assumptions of a model. In particular the effect of local perturbations of parameters or observations are examined.

## 2.5 Extension of technique to LME Models

Model diagnostic techniques, well established for classical models, have since been adapted for use with linear mixed effects models. Diagnostic techniques for LME models are inevitably more difficult to implement, due to the increased complexity.

Beckman, Nachtsheim and Cook (1987) applied the local influence method of Cook (1986) to the analysis of the linear mixed model.

While the concept of influence analysis is straightforward, implementation in mixed models is more complex. Update formulae for fixed effects models are available only when the covariance parameters are assumed to be known.

If the global measure suggests that the points in  $U$  are influential, the nature of that influence should be determined. In particular, the points in  $U$  can affect

- the estimates of fixed effects
- the estimates of the precision of the fixed effects
- the estimates of the covariance parameters
- the estimates of the precision of the covariance parameters
- fitted and predicted values

## 2.6 Marginal and Conditional Residuals

The marginal and conditional means in the linear mixed model are  $E[\mathbf{Y}] = \mathbf{X}\boldsymbol{\beta}$  and  $E[\mathbf{Y}|\mathbf{u}] = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}$ , respectively.

A residual is the difference between an observed quantity and its estimated or predicted value. In the mixed model you can distinguish marginal residuals  $r_m$  and conditional residuals  $r_c$ . A marginal residual is the difference between the observed data and the estimated (marginal) mean,  $r_{mi} = y_i - x'_0\hat{b}$ . A conditional residual is the difference between the observed data and the predicted value of the observation,  $r_{ci} = y_i - x'_i\hat{b} - z'_i\hat{\gamma}$ .

### 2.6.1 Marginal and Conditional Residuals

$$r_{mi} = x_i^T \hat{\beta} \tag{2.1}$$

### 2.6.2 Marginal Residuals

$$\begin{aligned} \hat{\beta} &= (X^T R^{-1} X)^{-1} X^T R^{-1} Y \\ &= BY \end{aligned}$$

## 2.7 Standardized and studentized residuals

To alleviate the problem caused by inconstant variance, the residuals are scaled (i.e. divided) by their standard deviations. This results in a ‘standardized residual’. Because true standard deviations are frequently unknown, one can instead divide a residual by the estimated standard deviation to obtain the ‘studentized residual’.

Another possible scaled residual is the ‘Pearson residual’ whereby a residual is divided by the standard deviation of the dependent variable. The Pearson residual can be used when the variability of  $\hat{\beta}$  is disregarded in the underlying assumptions.

### 2.7.1 Studentization

A random variable is said to be standardized if the difference from its mean is scaled by its standard deviation. The residuals above have mean zero but their variance is unknown, it depends on the true values of  $\theta$ . Standardization is thus not possible in practice. Instead, you can compute studentized residuals by dividing a residual by an estimate of its standard deviation. If that estimate is independent of the  $i$ th observation, the process is termed external studentization. This is usually accomplished by excluding the  $i$ –th observation when computing the estimate of its standard error. If the observation contributes to the standard error computation, the residual is said to be *internally studentized*.

## 2.8 Case Deletion Diagnostics

Christensen, Pearson and Johnson (1992) studied case deletion diagnostics, in particular the equivalent of Cook's distance, for diagnosing influential observations when estimating the fixed effect parameters and variance components.

### 2.8.1 Case Deletion Diagnostics

Case-deletion diagnostics provide a useful tool for identifying influential observations and outliers.

The computation of case deletion diagnostics in the classical model is made simple by the fact that estimates of  $\beta$  and  $\sigma^2$ , which exclude the  $i$ th observation, can be computed without re-fitting the model. Such update formulas are available in the mixed model only if you assume that the covariance parameters are not affected by the removal of the observation in question. This is rarely a reasonable assumption.

## 2.9 Effects on fitted and predicted values

$$\hat{e}_{i(U)} = y_i - x\hat{\beta}_{(U)} \quad (2.2)$$

### 2.9.1 Deletion Diagnostics

Since the pioneering work of Cook in 1977, deletion measures have been applied to many statistical models for identifying influential observations.

Deletion diagnostics provide a means of assessing the influence of an observation (or groups of observations) on inference on the estimated parameters of LME models.

Data from single individuals, or a small group of subjects may influence non-linear mixed effects model selection. Diagnostics routinely applied in model building may



identify such individuals, but these methods are not specifically designed for that purpose and are, therefore, not optimal. We describe two likelihood-based diagnostics for identifying individuals that can influence the choice between two competing models.

### **2.9.2 Case Deletion Diagnostics for Mixed Models**

? notes the case deletion diagnostics techniques have not been applied to linear mixed effects models and seeks to develop methodologies in that respect.

? develops these techniques in the context of REML

### **2.9.3 Methods and Measures**

The key to making deletion diagnostics useable is the development of efficient computational formulas, allowing one to obtain the case deletion diagnostics by making use of basic building blocks, computed only once for the full model.

Zewotir and Galpin (2005) lists several established methods of analyzing influence in LME models. These methods include

- Cook's distance for LME models,
- likelihood distance,
- the variance (information) ration,
- the Cook-Weisberg statistic,
- the Andrews-Prebigon statistic.

## 2.10 Influence analysis

Likelihood based estimation methods, such as ML and REML, are sensitive to unusual observations. Influence diagnostics are formal techniques that assess the influence of observations on parameter estimates for  $\beta$  and  $\theta$ . A common technique is to refit the model with an observation or group of observations omitted.

West et al. (2007) examines a group of methods that examine various aspects of influence diagnostics for LME models. For overall influence, the most common approaches are the ‘likelihood distance’ and the ‘restricted likelihood distance’.

### 2.10.1 Cook’s 1986 paper on Local Influence

Cook 1986 introduced methods for local influence assessment. These methods provide a powerful tool for examining perturbations in the assumption of a model, particularly the effects of local perturbations of parameters of observations.

The local-influence approach to influence assessment is quite different from the case deletion approach, comparisons are of interest.

### 2.10.2 Overall Influence

An overall influence statistic measures the change in the objective function being minimized. For example, in OLS regression, the residual sums of squares serves that purpose. In linear mixed models fit by maximum likelihood (ML) or restricted maximum likelihood (REML), an overall influence measure is the likelihood distance [Cook and Weisberg ].

## 2.11 Likelihood Distance

The likelihood distance gives the amount by which the log-likelihood of the full data changes if one were to evaluate it at the reduced-data estimates. The important point is that  $l(\psi_U)$  is not the log-likelihood obtained by fitting the model to the reduced data set.

It is obtained by evaluating the likelihood function based on the full data set (containing all  $n$  observations) at the reduced-data estimates.

The likelihood distance is a global, summary measure, expressing the joint influence of the observations in the set  $U$  on all parameters in  $\psi$  that were subject to updating.

### 2.11.1 Likelihood Distance

The likelihood distance is a global, summary measure, expressing the joint influence of the observations in the set  $U$  on all parameters in  $\phi$  that were subject to updating.

## 2.12 Haslett's Analysis

For fixed effect linear models with correlated error structure Haslett (1999) showed that the effects on the fixed effects estimate of deleting each observation in turn could be cheaply computed from the fixed effects model predicted residuals.

## 2.13 Iterative and non-iterative influence analysis

Schabenberger (2004) highlights some of the issue regarding implementing mixed model diagnostics.

A measure of total influence requires updates of all model parameters.

however, this doesn't increase the procedures execution time by the same degree.

### 2.13.1 Iterative Influence Analysis

For linear models, the implementation of influence analysis is straightforward. However, for LME models, the process is more complex. Update formulas for the fixed effects are available only when the covariance parameters are assumed to be known. A measure of total influence requires updates of all model parameters. This can only be achieved in general is by omitting observations, then refitting the model.

Schabenberger (2004) describes the choice between iterative influence analysis and non-iterative influence analysis.

## 2.14 Matrix Notation for Case Delection

### 2.14.1 Case deletion notation

For notational simplicity,  $\mathbf{A}(i)$  denotes an  $n \times m$  matrix  $\mathbf{A}$  with the  $i$ -th row removed,  $a_i$  denotes the  $i$ -th row of  $\mathbf{A}$ , and  $a_{ij}$  denotes the  $(i, j)$ -th element of  $\mathbf{A}$ .

### 2.14.2 Partitioning Matrices

Without loss of generality, matrices can be partitioned as if the  $i$ -th omitted observation is the first row; i.e.  $i = 1$ .

## 2.15 CPJ's Three Propositions

### Proposition 1

$$\mathbf{V}^{-1} = \begin{bmatrix} \nu^{ii} & \lambda'_i \\ \lambda_i & \Lambda_{[i]} \end{bmatrix}$$

$$\mathbf{V}_{[i]}^{-1} = \Lambda_{[i]} - \frac{\lambda_i \lambda'_i}{\lambda_i}$$

### 2.15.1 Proposition 2

$$(i) \quad \mathbf{X}_{[i]}^T \mathbf{V}_{[i]}^{-1} \mathbf{X}_{[i]} = \mathbf{X}' \mathbf{V}^{-1} \mathbf{X}$$

$$(ii) \quad = (\mathbf{X}' \mathbf{V}^{-1} \mathbf{Y})^{-1}$$

$$(iii) \quad \mathbf{X}_{[i]}^T \mathbf{V}_{[i]}^{-1} \mathbf{Y}_{[i]} = \mathbf{X}' \mathbf{V}^{-1} \mathbf{Y}$$

### 2.15.2 Proposition 3

This proposition is similar to the formula for the one-step Newtown Raphson estimate of the logistic regression coefficients given by pregibon (1981) and discussed in Cook Weisberg.

## 2.16 Augmented GLMs

Generalized linear models are a generalization of classical linear models.

The subscript  $M$  is a label referring to the mean model.

$$\begin{pmatrix} Y \\ \psi_M \end{pmatrix} = \begin{pmatrix} X & Z \\ 0 & I \end{pmatrix} \begin{pmatrix} \beta \\ \nu \end{pmatrix} + e^* \quad (2.3)$$

The error term  $e^*$  is normal with mean zero. The variance matrix of the error term is given by

$$\Sigma_a = \begin{pmatrix} \Sigma & 0 \\ 0 & D \end{pmatrix}. \quad (2.4)$$

$$X = \begin{pmatrix} T & Z \\ 0 & I \end{pmatrix} \delta = \begin{pmatrix} \beta \\ \nu \end{pmatrix} \quad (2.5)$$

$$y_a = T\delta + e^* \quad (2.6)$$

Weighted least squares equation

## 2.17 Covariance Parameters

The unknown variance elements are referred to as the covariance parameters and collected in the vector  $\theta$ .



## 2.18 Terminology for Case Deletion diagnostics

Preisser (1996) describes two type of diagnostics. When the set consists of only one observation, the type is called 'observation-diagnostics'. For multiple observations, Preisser describes the diagnostics as 'cluster-deletion' diagnostics.

## 2.19 The CPJ Paper

### 2.19.1 Case-Deletion results for Variance components

examines case deletion results for estimates of the variance components, proposing the use of one-step estimates of variance components for examining case influence. The method describes focuses on REML estimation, but can easily be adapted to ML or other methods.

Christensen developed their global influences for the deletion of single observations in two steps: a one-step estimate for the REML (or ML) estimate of the variance components, and an ordinary case-deletion diagnostic for a weighted regression problem (conditional on the estimated covariance matrix) for fixed effects. Lesaffre's approach accords with that proposed by Christensen et al when applied in a repeated measurement context, with a large sample size.

### 2.19.2 CPJ Notation

$$\mathbf{C} = \mathbf{H}^{-1} = \begin{bmatrix} c_{ii} & \mathbf{c}'_i \\ \mathbf{c}_i & \mathbf{C}_{[i]} \end{bmatrix}$$

Christensen et al. (1992) noted the following identity:

$$\mathbf{H}^{-1}_{[i]} = \mathbf{C}_{[i]} - \frac{1}{c_{ii}} \mathbf{c}_{[i]} \mathbf{c}'_{[i]}$$

Christensen et al. (1992) use the following as building blocks for case deletion statistics.

- $\check{x}_i$
- $\check{z}_i$
- $\check{z}_{ij}$

- $\check{y}_i$
- $p_i i$
- $m_i$

All of these terms are a function of a row (or column) of  $\mathbf{H}$  and  $\mathbf{H}_{[i]}^{-1}$

# Chapter 3

## Roy2013

<http://business.utsa.edu/wps/MSS/0017MSS-253-2013.pdf>

Testing the Equality of Mean Vectors for Paired Doubly Multivariate Observations

Example 2. (Mineral Data): This data set is taken from Johnson and Wichern (2007, p. 43). An investigator measured the mineral content of bones (radius, humerus and ulna) by photon absorptiometry to examine whether dietary supplements would slow bone loss in 25 older women. Measurements were recorded for three bones on the dominant and nondominant sides. Thus, the data is doubly multivariate and clearly  $u = 2$  and  $q = 3$ . The bone mineral contents for the rst 24 women one year after their participation in an experimental program is given in Johnson and Wichern (2007, p. 353).

Thus, for our analysis we take only rst 24 women in the rst data set. We test whether there has been a bone loss considering the data as doubly multivariate and has BCS structure. We rearrange the variables in the data set by grouping together the mineral content of the dominant sides of radius, humerus and ulna as the rst three variables, that is, the variables in the rst location ( $u = 1$ ) and then the mineral contents for the non-dominant side of the same bones ( $u = 2$ )

# Chapter 4

## Appendices

### 4.1 Sherman Morrison Woodbury Formula

The ‘Sherman Morrison Woodbury’ Formula is a well known result in linear algebra;

$$(A + a^T B)^{-1} = A^{-1} - A^{-1} a^T (I - b A^{-1} a^T)^{-1} b A^{-1} \quad (4.1)$$

This result is highly useful for analyzing regression diagnostics, and for matrices inverses in general. Consider a  $p \times p$  matrix  $X$ , from which a row  $x_i^T$  is to be added or deleted. ? sets  $A = X^T X$ ,  $a = -x_i^T$  and  $b = x_i^T$ , and writes the above equation as

$$(X^T X \pm x_i x_i^T)^{-1} = (X^T X)^{-1} \mp \frac{(X^T X)^{-1} (x_i x_i^T (X^T X)^{-1})}{1 - x_i^T (X^T X)^{-1} x_i} \quad (4.2)$$

The projection matrix  $H$  (also known as the hat matrix), is a well known identity that maps the fitted values  $\hat{Y}$  to the observed values  $Y$ , i.e.  $\hat{Y} = HY$ .

$$H = X(X^T X)^{-1} X^T \quad (4.3)$$

$H$  describes the influence each observed value has on each fitted value. The diagonal elements of the  $H$  are the ‘leverages’, which describe the influence each observed value

has on the fitted value for that same observation. The residuals ( $R$ ) are related to the observed values by the following formula:

$$R = (I - H)Y \tag{4.4}$$

The variances of  $Y$  and  $R$  can be expressed as:

$$\begin{aligned} \text{var}(Y) &= H\sigma^2 \\ \text{var}(R) &= (I - H)\sigma^2 \end{aligned} \tag{4.5}$$

Updating techniques allow an economic approach to recalculating the projection matrix,  $H$ , by removing the necessity to refit the model each time it is updated. However this approach is known for numerical instability in the case of down-dating.

## 4.2 The Hat Matrix

The projection matrix  $H$  (also known as the hat matrix), is a well known identity that maps the fitted values  $\hat{Y}$  to the observed values  $Y$ , i.e.  $\hat{Y} = HY$ .

$$H = X(X^T X)^{-1} X^T \quad (4.6)$$

$H$  describes the influence each observed value has on each fitted value. The diagonal elements of the  $H$  are the ‘leverages’, which describe the influence each observed value has on the fitted value for that same observation. The residuals ( $R$ ) are related to the observed values by the following formula:

$$R = (I - H)Y \quad (4.7)$$

The variances of  $Y$  and  $R$  can be expressed as:

$$\begin{aligned} \text{var}(Y) &= H\sigma^2 \\ \text{var}(R) &= (I - H)\sigma^2 \end{aligned} \quad (4.8)$$

Updating techniques allow an economic approach to recalculating the projection matrix,  $H$ , by removing the necessity to refit the model each time it is updated. However this approach is known for numerical instability in the case of down-dating.

### 4.2.1 Hat Values for MCS regression

With  $A$  as the averages and  $D$  as the casewise differences.

```
fit = lm(D~A)
```

$$H = A(A^T A)^{-1} A^T,$$

## 4.3 Cross Validation

Cross validation techniques for linear regression employ the use ‘leave one out’ recalculations. In such procedures the regression coefficients are estimated for  $n - 1$  covariates, with the  $Q^{th}$  observation omitted.

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

In leave-one-out cross validation, each observation is omitted in turn, and a regression model is fitted on the rest of the data.

Cross validation is used to estimate the generalization error of a given model. alternatively it can be used for model selection by determining the candidate model that has the smallest generalization error.

Evidently leave-one-out cross validation has similarities with ‘jackknifing’, a well known statistical technique. However cross validation is used to estimate generalization error, whereas the jackknife technique is used to estimate bias.

### 4.3.1 Cross Validation: Updating standard deviation

The variance of a data set can be calculated using the following formula.

$$S^2 = \frac{\sum_{i=1}^n (x_i^2) - \frac{(\sum_{i=1}^n x_i)^2}{n}}{n - 1} \quad (4.9)$$

While using bivariate data, the notation  $Sxx$  and  $Syy$  shall apply to the variance of  $x$  and of  $y$  respectively. The covariance term  $Sxy$  is given by

$$Sxy = \frac{\sum_{i=1}^n (x_i y_i) - \frac{(\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n}}{n - 1} \quad (4.10)$$

Let the observation  $j$  be omitted from the data set. The estimates for the variance identities can be updating using minor adjustments to the full sample estimates. Where  $(j)$  denotes that the  $j$ th has been omitted, these identities are



$$Sxx^{(j)} = \frac{\sum_{i=1}^n (x_i^2) - (x_j)^2 - \frac{((\sum_{i=1}^n x_i) - x_j)^2}{n-1}}{n-2} \quad (4.11)$$

$$Syy^{(j)} = \frac{\sum_{i=1}^n (y_i^2) - (y_j)^2 - \frac{((\sum_{i=1}^n y_i) - y_j)^2}{n-1}}{n-2} \quad (4.12)$$

$$Sxy^{(j)} = \frac{\sum_{i=1}^n (x_i y_i) - (y_j x_j) - \frac{((\sum_{i=1}^n x_i) - x_j)((\sum_{i=1}^n y_i) - y_k)}{n-1}}{n-2} \quad (4.13)$$

The updated estimate for the slope is therefore

$$\hat{\beta}_1^{(j)} = \frac{Sxy^{(j)}}{Sxx^{(j)}} \quad (4.14)$$

It is necessary to determine the mean for  $x$  and  $y$  of the remaining  $n - 1$  terms

$$\bar{x}^{(j)} = \frac{(\sum_{i=1}^n x_i) - (x_j)}{n-1}, \quad (4.15)$$

$$\bar{y}^{(j)} = \frac{(\sum_{i=1}^n y_i) - (y_j)}{n-1}. \quad (4.16)$$

The updated intercept estimate is therefore

$$\hat{\beta}_0^{(j)} = \bar{y}^{(j)} - \hat{\beta}_1^{(j)} \bar{x}^{(j)}. \quad (4.17)$$

## 4.4 Updating Estimates

### 4.4.1 Updating of Regression Estimates

Updating techniques are used in regression analysis to add or delete rows from a model, allowing the analyst the effect of the observation associated with that row. In time series problems, there will be scientific interest in the changing relationship between variables. In cases where there a single row is to be added or deleted, the procedure used is equivalent to a geometric rotation of a plane.

Updating techniques are used in regression analysis to add or delete rows from a model, allowing the analyst the effect of the observation associated with that row.

### 4.4.2 Updating Standard deviation

A simple, but useful, example of updating is the updating of the standard deviation when an observation is omitted, as practised in statistical process control analyzes. From first principles, the variance of a data set can be calculated using the following formula.

$$S^2 = \frac{\sum_{i=1}^n (x_i^2) - \frac{(\sum_{i=1}^n x_i)^2}{n}}{n - 1} \quad (4.18)$$

While using bivariate data, the notation  $Sxx$  and  $Syy$  shall apply hither to the variance of  $x$  and of  $y$  respectively. The covariance term  $Sxy$  is given by

$$Sxy = \frac{\sum_{i=1}^n (x_i y_i) - \frac{(\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n}}{n - 1}. \quad (4.19)$$

### 4.4.3 Updating of Regression Estimates

Updating techniques are used in regression analysis to add or delete rows from a model, allowing the analyst the effect of the observation associated with that row. In time

series problems, there will be scientific interest in the changing relationship between variables. In cases where there a single row is to be added or deleted, the procedure used is equivalent to a geometric rotation of a plane.

Updating techniques are used in regression analysis to add or delete rows from a model, allowing the analyst the effect of the observation associated with that row.

#### 4.4.4 Updating Standard deviation

A simple, but useful, example of updating is the updating of the standard deviation when an observation is omitted, as practised in statistical process control analyzes. From first principles, the variance of a data set can be calculated using the following formula.

$$S^2 = \frac{\sum_{i=1}^n (x_i^2) - \frac{(\sum_{i=1}^n x_i)^2}{n}}{n - 1} \quad (4.20)$$

While using bivariate data, the notation  $Sxx$  and  $Syy$  shall apply hither to the variance of  $x$  and of  $y$  respectively. The covariance term  $Sxy$  is given by

$$Sxy = \frac{\sum_{i=1}^n (x_i y_i) - \frac{(\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n}}{n - 1}. \quad (4.21)$$

#### 4.4.5 Updating of Regression Estimates

Consider a  $p \times p$  matrix  $X$ , from which a row  $x_i^T$  is to be added or deleted. ? sets  $A = X^T X$ ,  $a = -x_i^T$  and  $b = x_i^T$ , and writes the above equation as

$$(X^T X \pm x_i x_i^T)^{-1} = (X^T X)^{-1} \mp \frac{(X^T X)^{-1} (x_i x_i^T (X^T X)^{-1})}{1 - x_i^T (X^T X)^{-1} x_i} \quad (4.22)$$

#### 4.4.6 Updating Regression Estimates

Let the observation  $j$  be omitted from the data set. The estimates for the variance identities can be updating using minor adjustments to the full sample estimates. Where  $(j)$  denotes that the  $j$ th has been omitted, these identities are

$$Sxx^{(j)} = \frac{\sum_{i=1}^n (x_i^2) - (x_j)^2 - \frac{((\sum_{i=1}^n x_i) - x_j)^2}{n-1}}{n-2} \quad (4.23)$$

$$Syy^{(j)} = \frac{\sum_{i=1}^n (y_i^2) - (y_j)^2 - \frac{((\sum_{i=1}^n y_i) - y_j)^2}{n-1}}{n-2} \quad (4.24)$$

$$Sxy^{(j)} = \frac{\sum_{i=1}^n (x_i y_i) - (y_j x_j) - \frac{((\sum_{i=1}^n x_i) - x_j)((\sum_{i=1}^n y_i) - y_j)}{n-1}}{n-2} \quad (4.25)$$

The updated estimate for the slope is therefore

$$\hat{\beta}_1^{(j)} = \frac{Sxy^{(j)}}{Sxx^{(j)}} \quad (4.26)$$

It is necessary to determine the mean for  $x$  and  $y$  of the remaining  $n - 1$  terms

$$\bar{x}^{(j)} = \frac{(\sum_{i=1}^n x_i) - (x_j)}{n-1}, \quad (4.27)$$

$$\bar{y}^{(j)} = \frac{(\sum_{i=1}^n y_i) - (y_j)}{n-1}. \quad (4.28)$$

The updated intercept estimate is therefore

$$\hat{\beta}_0^{(j)} = \bar{y}^{(j)} - \hat{\beta}_1^{(j)} \bar{x}^{(j)}. \quad (4.29)$$

#### 4.4.7 Inference on intercept and slope

$$\hat{\beta}_1 \pm t_{(\alpha, n-2)} \sqrt{\frac{S^2}{(n-1)S_x^2}} \quad (4.30)$$

$$\frac{\hat{\beta}_0 - \beta_0}{SE(\hat{\beta}_0)} \quad (4.31)$$

$$\frac{\hat{\beta}_1 - \beta_1}{SE(\hat{\beta}_0)} \quad (4.32)$$

### **Inference on correlation coefficient**

This test of the slope is coincidentally the equivalent of a test of the correlation of the  $n$  observations of  $X$  and  $Y$ .

$$H_0 : \rho_{XY} = 0$$

$$H_A : \rho_{XY} \neq 0$$

$$(4.33)$$

## 4.5 Lesaffre's paper.

Lesaffre considers the case-weight perturbation approach.

Cook's 86 describes a local approach wherein each case is given a weight  $w_i$  and the effect on the parameter estimation is measured by perturbing these weights. Choosing weights close to zero or one corresponds to the global case-deletion approach.

Lesaffre describes the displacement in log-likelihood as a useful metric to evaluate local influence

Lesaffre describes a framework to detect outlying observations that matter in an LME model. Detection should be carried out by evaluating diagnostics  $C_i$ ,  $C_i(\alpha)$  and  $C_i(D, \sigma^2)$ .

Lesaffre defines the total local influence of individual  $i$  as

$$C_i = 2|\Delta_i' L^{-1} \Delta_i|. \quad (4.34)$$

The influence function of the MLEs evaluated at the  $i$ th point  $IF_i$ , given by

$$IF_i = -L^{-1} \Delta_i \quad (4.35)$$

can indicate how  $\hat{\theta}$  changes as the weight of the  $i$ th subject changes.

The manner by which influential observations distort the estimation process can be determined by inspecting the interpretable components in the decomposition of the above measures of local influence.

Lesaffre comments that there is no clear way of interpreting the information contained in the angles, but that this doesn't mean the information should be ignored.

# Bibliography

- Bland, J. and D. Altman (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet* *i*, 307–310.
- 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).
- Christensen, R., L. M. Pearson, and W. Johnson (1992). Case-deletion diagnostics for mixed models. *Technometrics* *34*(1), 38–45.
- Cook, R. (1986). Assessment of local influence. *Journal of the Royal Statistical Society. Series B (Methodological)* *48*(2), 133–169.
- Dunn, G. (2002). *Statistical Evaluation of Measurement Error* (Second ed.). Stanford: American Mathematical Society and Digital Press.
- 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.
- Haslett, J. (1999). A simple derivation of deletion diagnostic results for the general linear model with correlated errors. *Journal of the Royal Statistical Society (Series B)* *61*, 603–609.

- Haslett, J. and D. Dillane (2004). Application of ‘delete = replace’ to deletion diagnostics for variance component estimation. *Journal of the Royal Statistical Society (Series B)* 66, 131–143.
- Haslett, J. and K. Hayes (1998). Residuals for the linear model with general covariance structure. *Journal of the Royal Statistical Society (Series B)* 60, 201–215.
- 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.
- Ludbrook, J. (1997). Comparing methods of measurement. *Clinical and Experimental Pharmacology and Physiology* 24, 193–203.
- McCullough, C. and S. Searle (2001). *Generalized , Linear and Mixed Models*. Wiley Interscience.
- Preisser, J. S. (1996). Deletion diagnostics for generalised estimating equations. *Biometrika* 83(3), 551–556.
- Schabenberger, O. (2004). Mixed model influence diagnostics. 18929.
- Schabenberger, O. (2005). Mixed model influence diagnostics. In *SUGI*, Volume 29, pp. 189–29.
- West, B., K. Welch, and A. Galecki (2007). *Linear Mixed Models: a Practical Guide Using Statistical Software*. Chapman and Hall CRC.
- Zewotir, T. and J. Galpin (2005). Influence diagnostics for linear mixed models. *Journal of Data Science* 3, 153–177.