Contents

Chapter 1

Using LME Models for Method Comparison Problems

1.1 Overview

This chapter discusses the use of LME models in Method Comparison problems. In this section, we provide a brief discussion of each section in this chapter.

Introduction to LME Models, Fitting LME Models to MCS Data

In this section, we introduce the LME model, discusss how it can be applied to MCS problems, and how it is desirable in the case of replicate measurements, giving some examples from previous work (i.e. Carstensen et Al, Lai & Shaio, and Roy).

Carstensen et al recommends that replicate measurements for each method, but recognizes that resulting data are more difficult to analyze. To this end, LME models are advocated.

Further to that, there will be a discussion on fitting various types LME models using freely available software. To fully understand the complexities, a comparison of the **nlme** and **LME4** `R` Packages is required.

While the MCS problem is conventionally poised in the context of two methods of measurements, LME models allow for a straightforward analysis whereby several methods of measurement can be measured simulataneously. However Simple models only can only indicate agreement of lack thereof, and the presence of inter-method bias. To consider more complex questions, more complex LME models are required. Useful approaches, such as Roy (2009) will be introduced in a later section.

Residual Analysis for LME, Applications to MCS Data

This short section will look at residual analysis for LME models. The underlying assumptions for LME models are similar to those of classical linear mdoels. There are two key techniques: a residual plot and the normal probability plot. Using the nlme package it is possible to create plots specific to each method. This is useful in determine which methods 'disagree' with the rest. Analysis of the residuals would determine if the methods of measurement disagree systematically, or whether or not erroneous measurements associated with a subset of the cases are the cause of disagreement. Erroneous measurements are incorrect measurements that indicate disagreement between methods that would otherwise be in agreement.

Case Deletion Diagnostics for LME Data: Cooks Distance, DF-Betas

In this section we introduce influence analysis and case deletion diagnosics. A full overview of the topic will be provided although there are specific tools that are particularly useful in the case of MCS problems.

A discussion of how leave-k-out diagnostics would work in the context of MCS problems is required. There are several scenaros. Suppose we have two methods of measurement X and Y, each with three measurements for a specific case: $(x1, x2, x3, y1, y2, y3)$

- Leave One Out - requires we omit one observation (e.g. x1)

- Leave Pair Out - we would omit one pair of observation (e.g. x1 and y1)
- Leave Case Out - we omit all observations associated with a particular case. (e.g. x1,x2,x3,y1,y2,y3)

Using DFBETAs to assess agreement

Suppose an LME model was formulated to model agreement for various (i.e. 2or more) methods of measurement, with replicate measurements. If the methods are to be agreement, the DFBetas for each case would be the same for both methods.

As such, agreement between any two methods can be determined by a simple scatterplot of the DFBetas. If the points align along the line of equality, then both methods can be said to be in agreement. If they appear not to be agreement, asubsequent analysis using a technique proposed by Roy(2009) can be used to identify the specific cause for this lack of agreement.

The Pearson Correlation coefficient of the DFBetas can be used in conjection with this analysis. A high correlation confirms good agreement. The Bonferroni Outlier Test and Cook's Distance values can be used to identify unusual cases, when the relationship is modelled as a classical linear model. A test for non-constant variance may also be applied.

Deming Regression can be used to corroborate equality. Significance test for Deming regression estimates are not available, but 95% bootstrap confidence intervals for the slope estimate and intercept estimates can be computed.

Additionally a mean difference plot can be used to identify outliers. This mean-difference plot differs from the Bland-Altman plot in that the plot is denominated in terms of dfbeta values, and not in measurement units.

If lack of agreement is indicated between methods of measurement, use of Roy's Testing is advised.(next section)

Using Roy's Test to Identify cause of Lack of agreement

Barnhart specifies three conditions for method of measurement that are required for two methods of measurement to be considered in agreement.

(i) No Significant Inter-method bias

(ii) No significant Difference in Within-Subject Variance

(iii) No significant Difference in Within-Subject Variance

Roy(2009) demonstrates a LME model specification, and a series of tests that look at each of the criteria individually. If two methods of measuement lack agreement, the specific reason for this lack of agreement can be pinpointed.

Using Roy's Model to Compute LoAs and CR

In this short section, a demonstration of how Roy's technique can be used to compute two common MCS metrics: Limits of Agreement and the Coefficient of Repeatabilty. While Limits of Agreement are not used in the analysis proposed here, there are ubiquituous in literature, and a demonstration on how to compute them with the Roy Model would assist the adoption of this proposed method.

Model Diagnostics for Roy's Models

Further to previous work, this section revisits case-deletion and residual diagnostics, and explores how approaches devised by Galecki and Burzykowski can be used to appraise Roy's model.

Permutation and Power Tests

This section explores topics such as dependent variable simulation and Power Analysis (Galecki and Burzykowski) and Permutation testing. Using the ***predictmeans*** `R` package, it is possible to perform permutation t-tests for coefficients of (fixed) effects and permutation F-tests.

1.2 Diagnostic Tools for the nlme package

With the nlme package, the generic function `lme()` fits a linear mixed-effects model in the formulation described in Laird and Ware (1982) but allowing for nested random effects.

The within-group errors are allowed to be correlated and/or have unequal variances, which is very important in fitting the models for Roy's Tests

The nlme package has a limited set of diagnostic tools that can be used to assess the model fit. A review of the package manual is sufficient to get a sense of the package's capability in that regard.

1.2.1 Conditional and Marginal Residuals

A residual is the difference between an observed quantity and its estimated or predicted value. For LME models, **?** describes two types of residuals, marginal residuals and conditional residuals.

- 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, $r_{ci} = y_i - x_i'\hat{b} - z_i'\hat{\gamma}$

In a model without random effects, both sets of residuals coincide (**?**) . We shall revert to this matter in due course.

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 $\hat{y}_i$.

$$\hat{\epsilon}_i = y_i - \hat{y}_i = y_i - (X_i\hat{\beta} + Z_i\hat{b}_i)$$

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.

1.2.2 Residual Analysis for MCS

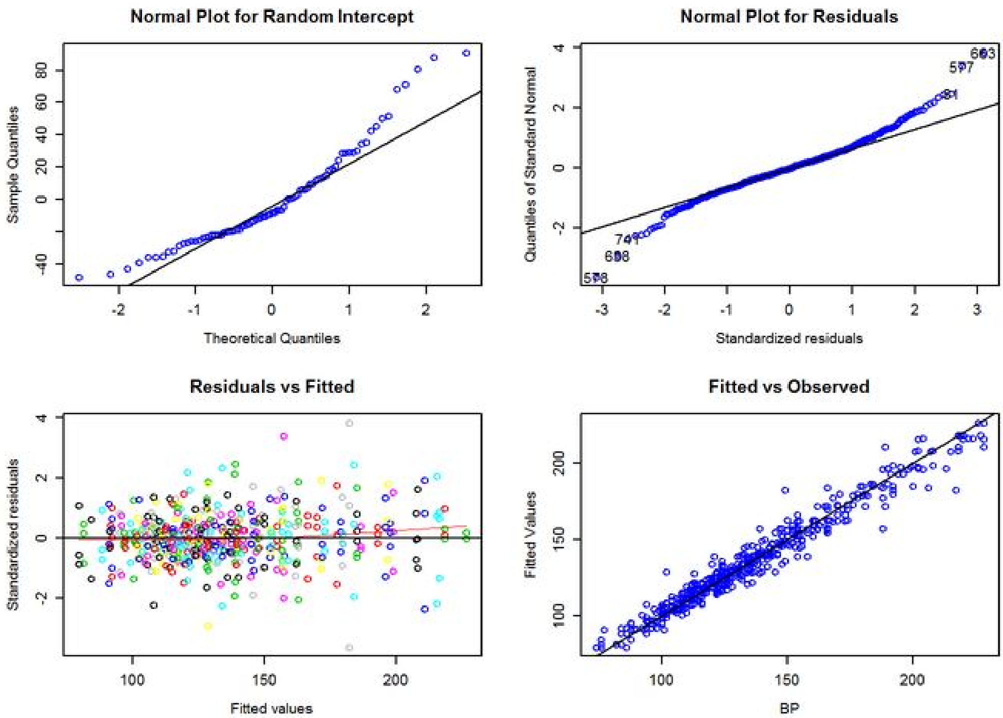


Figure 1.2.1:

1.2.3 PRESS Residuals and PRESS Statistic

In statistics, the predicted residual sum of squares (PRESS) statistic is a form of cross-validation used in regression analysis to provide a summary measure of the fit of a model to a sample of observations that were not themselves used to estimate the model. It is calculated as the sums of squares of the prediction residuals for those observations.

A fitted model having been produced, each observation in turn is removed and the model is refitted using the remaining observations. The out-of-sample predicted value is calculated for the omitted observation in each case, and the PRESS statistic is calculated as the sum of the squares of all the resulting prediction errors:[4]

$$\text{PRESS} = \sum_{i=1}^{n} (y_i - \hat{y}_{i,-i})^2$$

Given this procedure, the PRESS statistic can be calculated for a number of candidate model structures for the same dataset, with the lowest values of PRESS indicating the best structures. Models that are over-parameterised (over-fitted) would tend to give small residuals for observations included in the model-fitting but large residuals for observations that are excluded.

An (unconditional) predicted value is $\hat{y}_i = x_i'\hat{\boldsymbol{\beta}}$, where the vector x_i is the ith row of $\boldsymbol{X}$. For an `lme` object, such as our fitted model `JS.roy1`, the predicted values for each subject can be determined using the `coef.lme` function.

```
> JS.roy1 %>% coef %>% head(5)
methodJ   methodS
74      84.31724  91.08404
36      91.54994  97.05548
3       81.16581  96.48653
62      92.09493  90.89073
31      88.41411 103.38802
```

The (raw) residual is given as $\varepsilon_i = y_i - \hat{y}_i$. The PRESS residual is similarly constructed, using the predicted value for observation i with a model fitted from reduced data.

$$\varepsilon_{i(U)} = y_i - x_i'\hat{\boldsymbol{\beta}}_{(U)}$$

The PRESS statistic is the sum of the squared PRESS residuals:

$$PRESS = \sum \varepsilon^2_{i(U)}$$

where the sum is over the observations in $\boldsymbol{U}$.

Pinheiro and Bates provide some insight into how to compute and interpret model diagnostic plots for lme models. Unfortunately this aspect of LME theory is not as expansive as the corresponding body of work for Linear Models.

1.3 Checking the Assumption by Method

1.3.1 Residual Analysis

qqnorm.lme Normal Plot of Residuals or Random Effects from an lme Object

Description

Diagnostic plots for assessing the normality of residuals and random effects in the linear mixed-effects fit are obtained. The form argument gives considerable flexibility in the type of plot specification. A conditioning expression (on the right side of a — operator) always implies that different panels are used for each level of the conditioning factor, according to a Trellis display.

Residuals plots

lme allows to plot the residuals in the following ways:

```
res_lme=residuals(model_lme)
plot(res_lme)
qqnorm(res_lme)
qqline(res_lme)
plot(model_lme)
```

1.3.2 Diagnostic Plots for LME models

When the `plot` function calls the model object, the residual plot is produced.

```
plot(JS.roy1, which=c(1) )
```

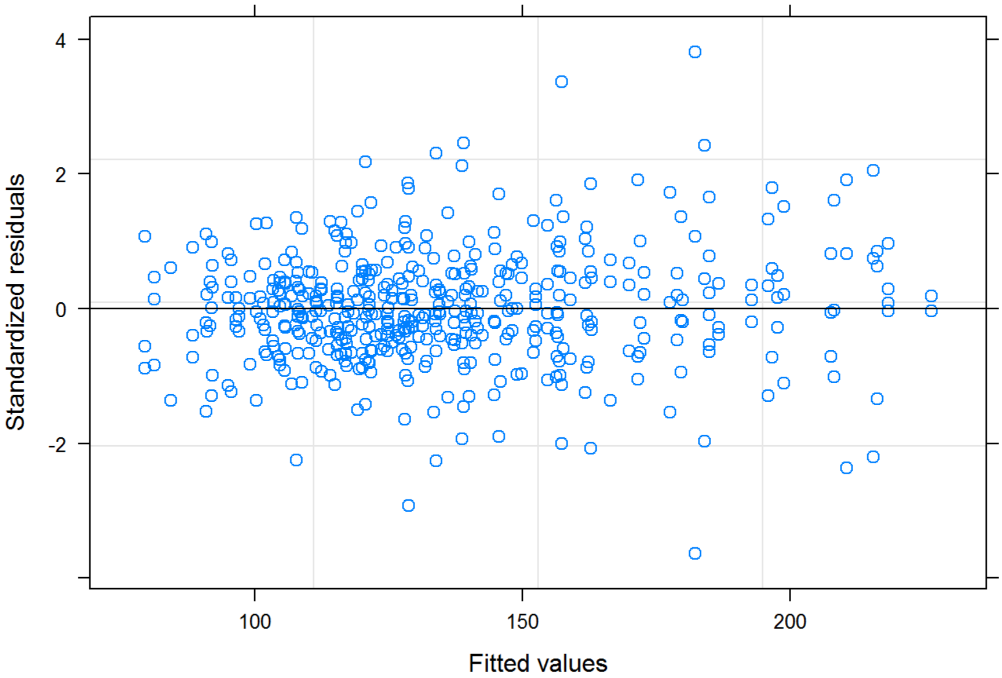


Figure 1.3.2:

LME models assume that the residuals of the model are normally distributed. A Normal probability plot can be constructed to check this assumption. Commonly used `R` commands can be used to construct the plot.

```
qqnorm(resid(JS.roy1),pch="*",col="red")
qqline(resid(JS.roy1),col="blue")
```

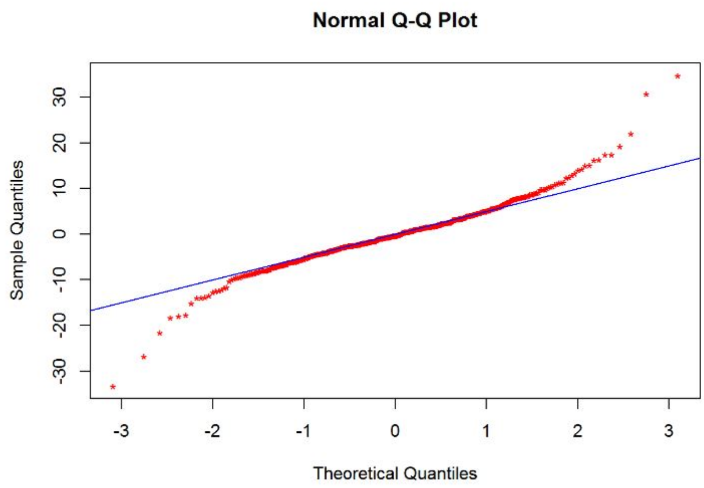


Figure 1.3.3:

```
table(dat$method[1:255])
##
##   J   S
## 255   0
table(dat$method[256:510])
##
##   J   S
##   0 255
```

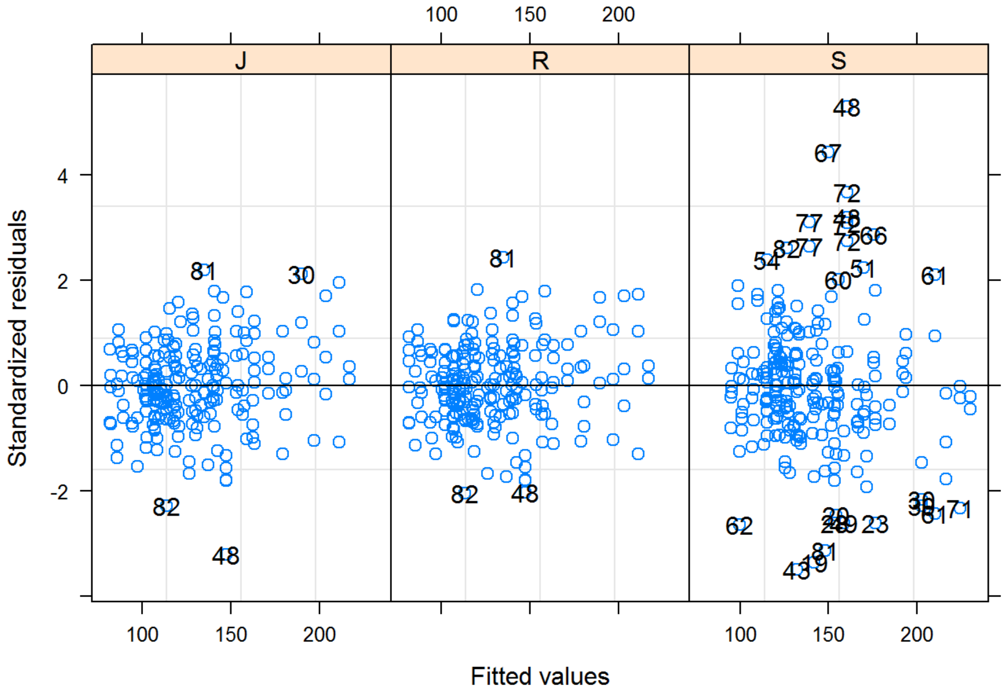


Figure 1.3.4:

```
plot(roy.NLME, resid(., type = "p") ~ fitted(.) | method,
   abline = 0, id=.05)
```

Cooks distance- Predict means thing here

Cook's Distance is a model diagnostic measure of an observation that is a measure of aggregate impact of each observation on the group of regression coefficients. Observations, or sets of observations, that have high Cook's distance usually have high resdiauls. We will revisit Cook's distance fully in due cource.

Cook's Distance is a good measure of the influence of an observation that is a measure of aggregate impact of each observation on the group of regression coefficients, as well as the group of fitted values.

The `CookD` fucntion , from the predictmeans R package, produces Cooks distance plots for an LME model (predictmeans)

```
library(predictMeans)
CookD(model, group=method, plot=TRUE, idn=5, newwd=FALSE)
```

```
## Cook's Distance
```

The particular cases that we will omit for the subsequent analysis are subjects 68,78 and 80.

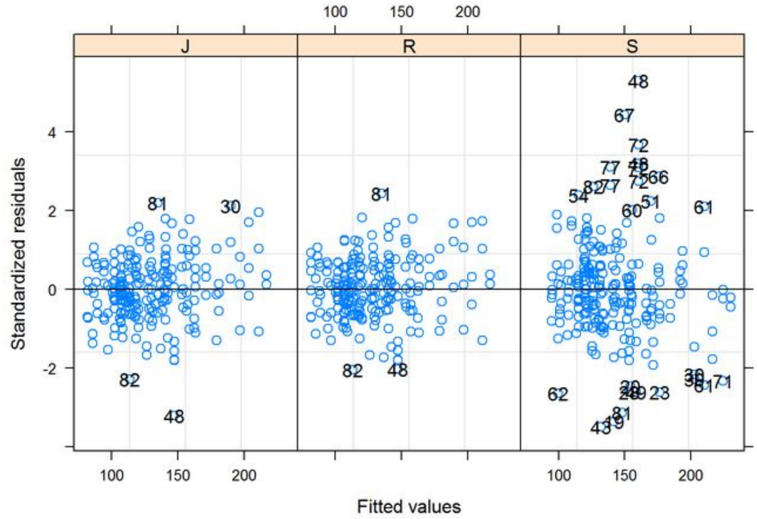
Reduced Data Set

It is important to determine if a specific group of cases or subjects give rise to the lack of agreement in the methods. If one were to examine fitted model if these cases were removed.

```
blood.red <- blood[!(blood$subject %in% c(68,78,80)),]
dim(blood.red)
# 27 observations should be removed.

roy.NLME.red <-lme(BP ~ method-1 , random=~1|subject,data = blood.red)
plot(roy.NLME.red, resid(., type = "p") ~ fitted(.) | method, abline = 0, id=.05)
```

In this instance, we conclude that there is a systemic diagreement between method S and the other two methods, and that lack of agreement can not be sourced to a handful of cases.



1.4 Preliminaries: Useful R commands

`vcov` returns the variance-covariance matrix of the main parameters of a fitted model object.

```
> vcov(JS.roy1)
methodJ  methodS
methodJ 11.01701  9.30105
methodS  9.30105 11.75301
```

`getVarCov` extract the variance-covariance matrix from a fitted model, such as a mixed-effects model.

```
> getVarCov(JS.roy1)
Random effects variance covariance matrix
methodJ methodS
methodJ  923.98  785.23
```

```
methodS  785.23  971.29
Standard Deviations: 30.397 31.166
```

1.5 coef.lme

The command `coef.lme` extract lmes coefficients

The estimated coefficients at level i are obtained by adding together the fixed effects estimates and the corresponding random effects estimates at grouping levels less or equal to i. The resulting estimates are returned as a data frame, with rows corresponding to groups and columns to coefficients. Optionally, the returned data frame may be augmented with covariates summarized over groups.

```
> JS.roy1 %>% coef %>% head(10)
methodJ   methodS
74  84.31724  91.08404
36  91.54994  97.05548
3   81.16581  96.48653
62  92.09493  90.89073
31  88.41411 103.38802
42  95.56881 104.70922
11 103.46092 111.33625
41  94.97700 108.08384
55  79.44762 110.02055
17 101.53470 112.46964
```

Suppose we refit the model, with a reduced data set (for example, with subject 68 missing)

```
x=68
datred <- dat[dat$subject !=x,]

# datred is other 84 cases (504 observations)

JS.roy1.red = lme(BP ~ method-1,
data = datred,
random = list(subject=pdSymm(~ method-1)),
weights=varIdent(form=~1|method),
correlation = corSymm(form=~1 | subject/repl),
method="ML")
```

```
> JS.roy1.red %>% coef %>% dim
[1] 84  2
>
> JS.roy1.red %>% coef %>% head(10)
methodJ   methodS
74  84.31425  90.89550
36  91.54526  96.86813
3   81.17663  96.36665
62  92.07933  90.64629
31  88.42473 103.27726
42  95.57027 104.56000
11 103.46071 111.18930
41  94.98491 107.96791
55  79.48324 110.02963
```

```
17 101.53939 112.34591
```

Chapter 2

Using DFBETAs from LME Models to Assess Agreement

? 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 (**?**).

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.

? 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.
- 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.

This is known as '*leave one out* or `leave k out' analysis`.

2.0.1 Influence

Influence arises at two stages of the LME model. Firstly when $\mathbf{V}$ is estimated by $\hat{V}$, and subsequent estimations of the fixed and random regression coefficients β and u, given $\hat{\mathbf{V}}$.

4 It is a multi-dimensional generalization of the idea of measuring how many standard deviations away P is from the mean of D.

This distance is zero if P is at the mean of D, and grows as P moves away from the mean: Along each principal component axis, it measures the number of standard deviations from P to the mean of D. If each of these axes is rescaled to have unit variance, then Mahalanobis distance corresponds to standard Euclidean distance in the transformed space. Mahalanobis distance is thus unitless and scale-invariant, and takes into account the correlations of the data set.

Influence() - Description

`influence()` is the workhorse function of the `influence.ME` package.

Based on a priorly estimated mixed effects regression model (estimated using lme4), the `influence()` function iteratively

modifies the mixed effects model to neutralize the effect a grouped set of data has on the parameters, and which

returns returns the fixed parameters of these iteratively modified models.

These are used to compute measures of influential data.

Usage

```
influence(model, group=NULL, select=NULL, obs=FALSE,
```

```
gf="single", count = FALSE, delete=TRUE, ...)
\
```

The `influence()` function was known as the `estex()` command in previous versions of the influence.ME pacakge

2.0.2 Influential Observations : DFBeta and DFBetas

The dfbeta refers to how much a parameter estimate changes if the observation or case in question is dropped from the data set. Cook's distance is presumably more important to you if you are doing predictive modeling, whereas dfbeta is more important in explanatory modeling.

2.0.3 DFBETA

DFBETAS (standardized difference of the beta) is a measure that standardizes the absolute difference in parameter estimates between a (mixed effects) regression model based on a full set of data, and a model from which a (potentially influential) subset of data is removed. `dfbeta()`

The DFBETAS statistics are the scaled measures of the change in each parameter estimate and are calculated by deleting the th observation:

where is the th element of . In general, large values of DFBETAS indicate observations that are influential in estimating a given parameter. Belsley, Kuh, and Welsch (1980) recommend 2 as a general cutoff value to indicate influential observations and as a size-adjusted cutoff.

$$DFBETA_a = \hat{\beta} - \hat{\beta}_{(a)} \quad (2.1)$$

$$= B(Y - Y_{\bar{a}} \quad (2.2)$$

In the case of method comparison studies, there are two covariates, and one can constru ct catterplots of the pairs of dfbeta values accordingly, both for LOO and LSO calculations. Furthermore 95% confidence ellipse can be constructed around these scatterplots. Note that with k covariates, there will be $k+1$ dfbetas (the intercept,β_0, and 1 β for each covariate). For example there would be 2 sets of of dfbeta, 510 valoues for each in the case of LOO, and 85 for LSO diagnostics.

2.0.4 DFFITS

DFFITS is a statistical measured designed to a show how influential an observation is in a statistical model.

$$DFFITS = \frac{\widehat{y_i} - \widehat{y_{i(k)}}}{s_{(k)}\sqrt{h_{ii}}}$$

It is closely related to the studentized residual. For the sake of brevity, we will concentrate on the Studentized Residuals.

2.0.5 PRESS

The prediction residual sum of squares (PRESS) is an value associated with this calculation. When fitting linear models, PRESS can be used as a criterion for model selection, with smaller values indicating better model fits.

$$PRESS = \sum (y - y^{(k)})^2 \tag{2.3}$$

- $e_{-Q} = y_Q - x_Q\hat{\beta}^{-Q}$
- $PRESS_{(U)} = y_i - x\hat{\beta}_{(U)}$

2.1 Application of DFBETAs in MCS Analysis

When in an MCS study. DFBetas can be used as a proxy measurement, allowing simple techniques to be used for assessing agreement.

Chapter 3

Model Diagnostics for MCS

Abstract

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

? describes the examination of model-data agreement as comprising several elements;

- residual analysis,
- goodness of fit,
- collinearity diagnostics
- influence analysis.

This chapter is comprised of two sections:

1. Residual Diagnostics
2. Influence Diagnostics

3.1 Introduction

In statistical modelling, the process of model validation is a critical step, but also a step that is too often overlooked. A very simple procedure is to examine well known metrics, such as the AIC and R^2 measures. However, using a small handful of simple measures and methods is insufficient to properly assess the quality of a fitted model. To do so properly, a full and comprehensive analysis that tests of all of the assumptions, as far as possible, must be carried out.

A statistical model, whether of the fixed-effects or mixed-effects variety, represents how you think your data were generated. Following model specification and estimation, it is of interest to explore the model-data agreement by raising pertinent questions. Further to the analysis of residuals, **?** recommends the examination of the following questions.

- Does the model-data agreement support the model assumptions?
- Should model components be refined, and if so, which components? For example, should certain explanatory variables be added or removed, and is the covariance of the observations properly specified?
- Are the results sensitive to model and/or data? Are individual data points or groups of cases particularly influential on the analysis?

3.2 Introduction to Influence analysis

Model diagnostic techniques determine whether or not the distributional assumptions are satisfied, and to assess the influence of unusual observations. 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. For linear models for uncorrelated data, it is not necessary to refit the model after removing a data point in order to measure the impact of an observation on the model. The change in fixed effect estimates, residuals, residual sums of squares, and the variance-covariance matrix of the fixed effects can be computed based on the fit to the full data alone. By contrast, in mixed models several important complications arise. Data points can affect not only the fixed effects but also the covariance parameter estimates on which the fixed-effects estimates depend.

3.2.1 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 (**?**).

3.2.2 Importance of Influence

The influence of an observation can be thought of in terms of how much the predicted values for other observations would differ if the observation in question were not included in the model fit. 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 β and θ. A common technique is to refit the model with an observation or group of observations omitted. The basic procedure for quantifying influence is simple as follows:

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

3.2.3 Measures of Influence

The impact of an observation on a regression fitting can be determined by the difference between the estimated regression coefficient of a model with all observations and the estimated coefficient when the particular observation is deleted. DFBETA and DFFITS are well known measures of influenc. The measure DFBETA is the studentized value of this difference. DFFITS is a statistical measured designed to a show how influential an observation is in a statistical model. DFFITS is closely related to the studentized residual.

$$DFBETA_a = \hat{\beta} - \hat{\beta}_{(a)} \tag{3.1}$$

$$= B(Y - Y_{\bar{a}} \tag{3.2}$$

$$DFFITS = \frac{\widehat{y_i} - \widehat{y_{i(k)}}}{s_{(k)}\sqrt{h_{ii}}} \tag{3.3}$$

The prediction residual sum of squares (PRESS) is an value associated with this calculation. When fitting linear models, PRESS can be used as a criterion for model selection, with smaller values indicating better model fits.

$$PRESS = \sum (y - y^{(k)})^2$$

3.2.4 Local Influence Analysis

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.

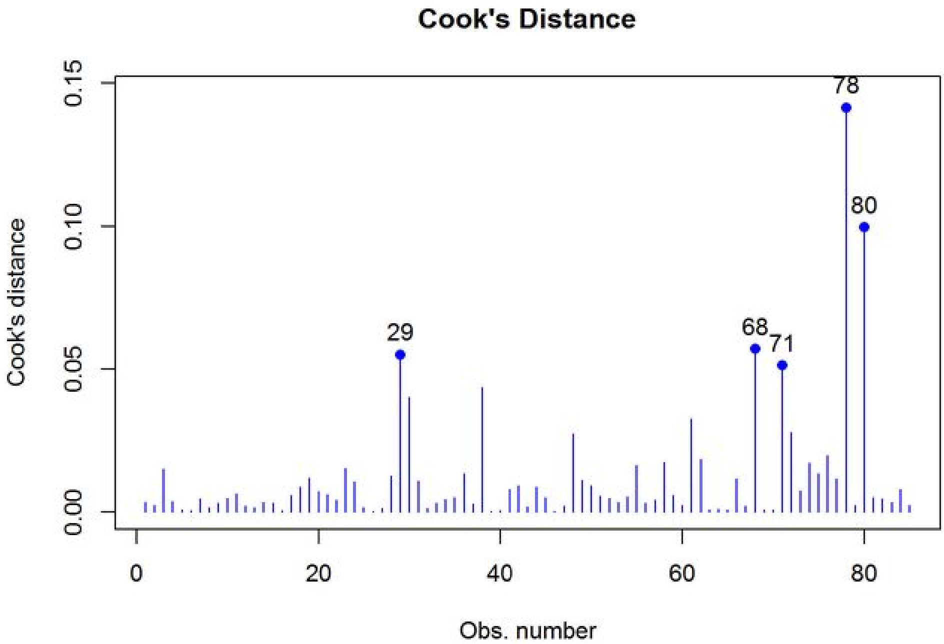
If the global measure suggests that the points in U are influential, the nature of that influence should be determined. In particular, the points in U can affect the following

- 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.

3.2.5 Cook's Distance

? develops case deletion diagnostics, in particular the equivalent of Cook's distance, a well-known metric, for diagnosing influential observations when estimating the fixed effect parameters and variance components. 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. Cook's Distance is a good measure of the influence of an observation that is a measure of aggregate impact of each observation on the group of regression coefficients, as well as the group of fitted values. If the predictions are the same with or without the observation in question, then the observation has no influence on the regression model. If the predictions differ greatly when the observation is not included in the analysis, then the observation is influential.

Cook's Distance for Blood Data



3.2.6 Deviance

In statistics, deviance is a quality of fit statistic for a model that is often used for statistical hypothesis testing. It is a generalization of the idea of using the sum of squares of residuals in ordinary least squares to cases where model-fitting is achieved by maximum likelihood.

Bibliography