Contents

Chapter 1

Residual and Influence Diagnostics

1.1 Chapter Overview

1. Residual Diagnostics
    (a) Marginal and Conditional Diagnostics
    (b) Scaled Residuals
2. 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

1.2 Residual Diagnostics

Consider a residual vector of the form $\hat{e} = \boldsymbol{PY}$, where $\boldsymbol{P}$ is a projection matrix, possibly an oblique projector. External studentization uses an estimate of Var that does not involve the ith observation. Externally studentized residuals are often preferred over studentized residuals because they have well known distributional properties in the standard linear models for independent data. Residuals that are scaled by the estimated variances of the responses are referred to as Pearson-type residuals. Standardization:

$$\frac{\hat{e}_i}{\sqrt{v_i}}$$

Studentization

$$\frac{\hat{e}_i}{\sqrt{\hat{v}_i}}$$

1.2.1 Marginal and Conditional Diagnostics

[Gregoire,Schabenberger, Barrett (1995)]

$$\boldsymbol{r}_m = \boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}$$

$$\boldsymbol{r}_c = \boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{Z}\hat{\boldsymbol{\gamma}}$$

For the individual observation the raw studentized and pearson type residuals are computed as follows:

$$r_{mi} = Y_i - X'\hat{\boldsymbol{\beta}}$$

$$r_{ci} = r_{mi} - Y_i - z_i'\hat{\boldsymbol{\gamma}}$$

Marginal Residuals

The marginal residuals are defined according to

$$\hat{\xi} = y - X\hat{\beta} = M^{-1}Qy.$$

Plots of the elements of the marginal residual vector versus the explanatory variables in X can be used to check the linearity of $\boldsymbol{y}$ in a similar manner to the residual plots used in linear models.

Conditional Residuals

A conditional probability is the difference between the observed value and the predicted value of the dependent variable.

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

In general conditional residuals are not welel suited for verifying model assumptions and detecting outliers. Even if the true model residuals are uncorrelated and have equal variance, conditional variances will tend to be correlated and their variances may be different for different subgroups of individuals (West et al., 2007).

1.2.2 Scaled Residuals

Pinheiro and Bates (1994) describes three types of residual that describe the variabilities present in LME models

1. marginal residuals, $\hat{\xi}$, which predict marginal errors,
2. conditional residuals, $\hat{\epsilon}$, which predict conditional errors,
3. the BLUP,$\boldsymbol{Z\hat{b}}$, that predicts random effects.

Each type of residual is useful to evaluates some assumption of the model.

According to hilton-minton [1995], a residual is considered pure for a specfic type fo error if it depends only on the fixed components and on the error that it is supposed to predict. Residuals that depend on other types of error are known as 'confounded errors'.

Studentized Residuals

Standardization is not possible in practice. Studentized residuals are residuals divided by the estimated standard estimation.

1.3 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.

1.4 A Procedure for Quantifying Influence

The basic procedure for quantifying influence is simple:

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.

We use the subscript (U) to denote quantities obtained without the observations in the set U. For example, (U) denotes the fixed-effects ***leave-U-out*** estimates. Note that the set U can contain multiple observations.

If the global measure suggests that the points in U are influential, you should next determine the nature of that influence. In particular, the points 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

It is important to further decompose the initial finding to determine whether data points are actually troublesome. Simply because they are influential somehow, should not trigger their removal from the analysis or a change in the model. For example, if points primarily affect the precision of the covariance parameters without exerting much influence on the fixed effects, then their presence in the data may not distort hypothesis tests or confidence intervals about β.

1.4.1 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.

1.4.2 Influence Diagnostics: Basic Idea and Statistics

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. The goal is rather to determine which cases are influential and the manner in which they are important to the analysis. Outliers, for example, may be the most noteworthy data points in an analysis. They can point to a model breakdown and lead to development of a better model.

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.

1.4.3 Diagnostic Methods for OLS models

Influence diagnostics are formal techniques allowing for the identification of observations that exert substantial influence on the estimates of fixed effects and variance covariance parameters.

The idea of influence diagnostics for a given observation is to quantify the effect of omission of this observation from the data on the results of the model fit. To this aim, the concept of likelihood displacement is used.

Cook (1977) greatly expanded the study of residuals and influence measures. Cook's Distance , denoted as$D_{(i)}$, is a well known diagnostic technique used in classical linear models, used as an overall measure of the combined impact of the ith case of all estimated regression coefficients. Cook's key observation was the effects of deleting each observation in turn could be calculated with little additional computation. That is to say, $D_{(i)}$ can be calculated without fitting a new regression coefficient each time an observation is deleted. Consequently deletion diagnostics have become an integral part of assessing linear models.

The focus of this analysis is related to the estimation of point estimates (i.e. regression coefficients). It must be pointed out that the effect on the precision of estimates is separate from the effect on the point estimates. Data points that have a small Cook's distance, for example, can still greatly affect hypothesis tests and confidence intervals, if their influence on the precision of the estimates is large.

As well as individual observations, Cook's distance can be used to analyse the influence of observations in subset U on a vector of parameter estimates (Cook, 1977).

$$\hat{e}_{i(U)} = y_i - x\hat{\beta}_{(U)} \tag{1.1}$$

$$\delta_{(U)} = \hat{\beta} - \hat{\beta}_{(U)} \tag{1.2}$$

1.4.4 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.

Cook (1977) greatly expanded the study of residuals and influence measures. Cook's key observation was the effects of deleting each observation in turn could be calculated with little additional computation. That is to say, $D_{(i)}$ can be calculated without fitting a new regression coefficient each time an observation is deleted. Consequently deletion diagnostics have become an integral part of assessing linear models. Cook proposed a measure that combines the information of leverage and residual of the observation, now known simply as the Cook's Distance. Cook's Distance , denoted as$D_{(i)}$, is a well known diagnostic technique used in classical linear models, used as an overall measure of the combined impact of the $i-$th case of all estimated regression coefficients.

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

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

1.5 Influence Statistics for LME models

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

1.5.1 Cook's Distance

- For variance components γ

Diagnostic tool for variance components

$$C_{\theta i} = ((\hat{\theta})_{[i]} - (\hat{\theta}))^T \mathrm{cov}((\hat{\theta}))^{-1}((\hat{\theta})_{[i]} - (\hat{\theta}))$$

1.5.2 Variance Ratio

- For fixed effect parameters β.

1.5.3 Cook-Weisberg statistic

- For fixed effect parameters β.

1.5.4 Zewotir Measures of Influence in LME Models

? describes a number of approaches to model diagnostics, investigating each of the following;

- Variance components
- Fixed effects parameters
- Prediction of the response variable and of random effects
- likelihood function

? 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.

1.5.5 Andrews-Pregibon statistic

- For fixed effect parameters β.

The Andrews-Pregibon statistic AP_i is a measure of influence based on the volume of the confidence ellipsoid. The larger this statistic is for observation i, the stronger the influence that observation will have on the model fit.

Random Effects

A large value for $CD(u)_i$ indicates that the $i-$th observation is influential in predicting random effects.

1.5.6 Computation and Notation

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

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

1.5.7 Cook's Distance

- For variance components γ: $CD(\gamma)_i$,
- For fixed effect parameters β: $CD(\beta)_i$,
- For random effect parameters $\boldsymbol{u}$: $CD(u)_i$,
- For linear functions of $\hat{beta}$: $CD(\psi)_i$

Diagnostic tool for variance components

$$C_{\theta i} = (\hat{(\theta)}_{[i]} - \hat{(\theta)})^T \mathrm{cov}(\hat{(\theta)})^{-1}(\hat{(\theta)}_{[i]} - \hat{(\theta)})$$

Random Effects

A large value for $CD(u)_i$ indicates that the $i-$th observation is influential in predicting random effects.

linear functions

$CD(\psi)_i$ does not have to be calculated unless $CD(\beta)_i$ is large.

1.5.8 Information Ratio

Chapter 2

Influence Diagnostics

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.

2.1 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 (Schabenberger, 2004).

2.2 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: Cooks (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

Furthermore, closed-form expressions for computing the change in important model quantities might not be available. This section provides background material for the various influence diagnostics available with the MIXED procedure. See the section Mixed Models Theory for relevant expressions and definitions. The parameter vector denotes all unknown parameters in the and matrix. The observations whose influence is being ascertained are represented by the set and referred to simply as "the observations in ." The estimate of a parameter vector, such as , obtained from all observations except those in the set is denoted . In case of a matrix , the notation represents the matrix with the rows in removed; these rows are collected in . If is symmetric, then notation implies removal of rows and columns. The vector comprises the responses of the data points being removed, and is the variance-covariance matrix of the remaining observations. When , lowercase notation emphasizes that single points are removed, such as .

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.

2.3 Overall Influence and Iterative Influence Analysis

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].

2.3.1 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].

2.3.2 Iterative Influence Analysis

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

2.3.3 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 doesnt increase the procedures execution time by the same degree.

2.3.4 Local Influence

Beckman et al. (1987) applied the local influence method of Cook (1986) to the analysis of the LME 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.

2.4 Influence analysis for 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.

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.

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'.

2.4.1 Influence Analysis for LME Models

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.

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

2.4.2 Computation Matters

Key to the implementations of influence diagnostics in the MIXED procedure is the attempt to quantify influence, where possible, by drawing on the basic definitions of the various statistics in the classical linear model.

On occasion, quantification is not possible. Assume, for example, that a data point is removed and the new estimate of the G matrix is not positive definite. This may occur if a variance component estimate now falls on the boundary of the parameter space. Thus, it may not be possible to compute certain influence statistics comparing the full-data and reduced-data parameter estimates. However, knowing that a new singularity was encountered is important qualitative information about the data points influence on the analysis.

2.4.3 Extension of techniques 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) Beckman et al. (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.

2.4.4 Analyzing Influence in LME models

"*Influence* is defined by **?** 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 rather to identify influential cases and the manner in which they are important to the analysis. A consequence of this that cases may be to mark data points for deletion so that a better model fit can be achieved for the reduced data (**?**).

? considers several important aspects of the use and implementation of influence measures in LME models. *schabenberger* notes that it is not always possible to derive influence statistics necessary for comparing full- and reduced-data parameter estimates.

? 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.

2.4.5 Influence in LME models (schab)

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.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'.

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 of 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.

2.5 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.

2.6 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.

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 β and σ^2, which exclude the ith 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.

3. Case Deletion Diagnostics for LME Data: Cooks Distance, DFBetas

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: specifically the Cook's Distance and the DFBeta.

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: $(x_1, x_2, x_3, y_1, y_2, y_3)$

- Leave One Out - one observation is omitted (e.g. x_1)
- Leave Pair Out - one pair of observation is omitted (e.g. x_1 and y_1)
- Leave Case (or Subject) Out - All observations associated with a particular case or subject are omitted. (e.g. $\{x_1, x_2, x_3, y_1, y_2, y_3\}$)

Other metrics, such as the likelihood distance, will also be introduced, and revisited in a later section.

2.6.1 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 β and σ^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.

CPJ develops case deletion diagnostics, in particular the equivalent of Cook's distance, for diagnosing influential observations when estimating the fixed effect parameters and variance components.

2.7 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

2.8 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.

Effects on fitted and predicted values

$$\hat{e}_{i(U)} = y_i - x\hat{\beta}_{(U)} \tag{2.1}$$

2.9 Case 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. 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 β and σ^2, which exclude the ith 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.

2.10 Case Deletion Diagnostics

? develops case deletion diagnostics, in particular the equivalent of Cook's distance, for diagnosing influential observations when estimating the fixed effect parameters and variance components.

2.11 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 β and σ^2, which exclude the ith 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.

Linear models for uncorrelated data have well established measures to gauge the influence of one or more observations on the analysis. For such models, closed-form update expressions allow efficient computations without refitting the model.

Since the pioneering work of Cook in 1977, deletion measures have been applied to many statistical models for identifying influential observations. Case-deletion diagnostics provide a useful tool for identifying influential observations and outliers.

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.

The computation of case deletion diagnostics in the classical model is made simple by the fact that estimates of β and σ^2, which exclude the $i-$th observation, can be computed without re-fitting the model.

2.12 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. When applied to LME models, such update formulas are available only if one assumes that the covariance parameters are not affected by the removal of the observation in question. However, this is rarely a reasonable assumption.

2.13 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].

2.14 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 influence. 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.

$$\begin{aligned} DFBETA_a &= \hat{\beta} - \hat{\beta}_{(a)} \quad (2.2) \\ &= B(Y - Y_{\bar{a}} \quad (2.3) \end{aligned}$$

$$DFFITS = \frac{\widehat{y_i} - \widehat{y_{i(k)}}}{s_{(k)}\sqrt{h_{ii}}} \quad (2.4)$$

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.

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

2.15 Measures of Influence

The impact of an observation, or a case with multiple obverations, 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. The measure DFBETA is the studentized value of this difference.

2.16 DFBETAs

The measure that measures how much impact each observation has on a particular predictor is DFBETAs The DFBETA for a predictor and for a particular observation is the difference between the regression coefficient calculated for all of the data and the regression coefficient calculated with the observation deleted, scaled by the standard error calculated with the observation deleted.

DFBETA is a measure found for each observation in a dataset. The DFBETA for a particular observation is the difference between the regression coefficient for an included variable calculated for all of the data and the regression coefficient calculated with the observation deleted, scaled by the standard error calculated with the observation deleted.

The cut-off value for DFBETAs is $\frac{2}{\sqrt{n}}$, where n is the number of observations. However, another cut-off is to look for observations with a value greater than 1.00. Here cutoff means, ?this observation could be overly influential on the estimated coefficient.?

2.17 DFBETAs

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. A value for DFBETAS is calculated for each parameter in the model separately.

This function computes the DFBETAS based on the information returned by the estex() function.

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.

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.

$$\begin{aligned} DFBETA_a &= \hat{\beta} - \hat{\beta}_{(a)} && (2.5) \\ &= B(Y - Y_{\bar{a}} && (2.6) \end{aligned}$$

In the case of method comparison studies, there are two covariates, and one can construct 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.

2.18 DFFITS

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

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

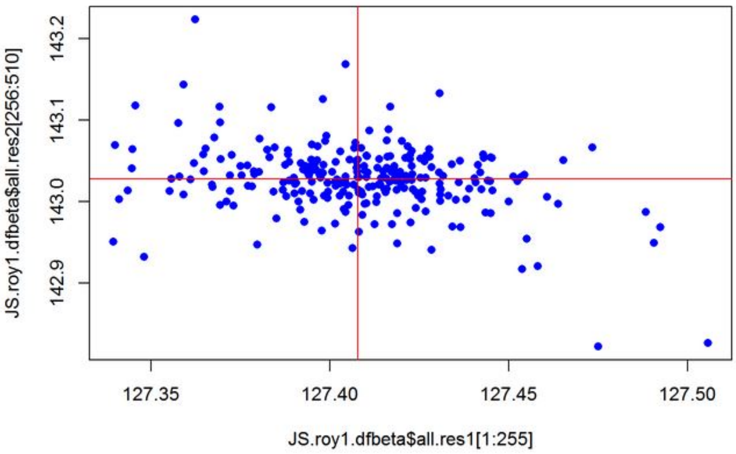


Figure 2.18.1:

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

DFFITS is a diagnostic meant to show how influential a point is in a statistical regression. It was proposed in 1980. It is defined as the change (”DFFIT”), in the predicted value for a point, obtained when that point is left out of the regression, ”Studentized” by dividing by the estimated standard deviation of the fit at that point:

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

DFbetas for Blood Data

```
plot(JS.roy1.dfbeta$all.res1[1:255],JS.roy1.dfbeta$all.res2[256:510],
pch=16,col="blue")
abline(v=JS.roy1.dfbeta$all.res1[256],col="red")
abline(h=JS.roy1.dfbeta$all.res2[1],col="red")
```

2.19 Influence Diagnostics

a Overall Measures that compare changes in objectives functions; (restricted) maximum likelihood (Cook Weisberg, 1982)

b Influence on parameter estimates: Cook's Distance, MDFFITs

c Influence on precision of estimates: CovTrace and CovRatio

d Influence on fitted and predicted values: PRESS residuals, PRESS statistics, DFFITs

e Outlier properties : internally and externally studentized residuals, leverage

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.

Christensen et al. (1992) 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 resgression problem (conditional on the estimated covariance matrix) for fixed effects.

Cook (1977) greatly expanded the study of residuals and influence measures. Cook's key observation was the effects of deleting each observation in turn could be computed without undue additional computational expense. Consequently deletion diagnostics have become an integral part of assessing linear models.

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

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. The measure DFBETA is the studentized value of this difference.

2.20 Cook's Distance

Cooks Distance (D_i) is an overall measure of the combined impact of the ith case of all estimated regression coefficients. It uses the same structure for measuring the combined impact of the differences in the estimated regression coefficients when the kth case is deleted. $D_{(k)}$ can be calculated without fitting a new regression coefficient each time an observation is deleted.

2.20.1 Predicted Values, PRESS Residual and the PRESS Statistic

The PRESS statistic is the sum of the squared PRESS residuals PRESS $= \sum \hat{\varepsilon}^2_{i(U)}$

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.

$$PRESS = \sum (y - y^{(k)})^2 \tag{2.7}$$

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.

$$\begin{aligned} e_{-Q} &= y_Q - x_Q \hat{\beta}^{-Q} \\ PRESS &= \sum (y - y^{-Q})^2 \\ PRESS_{(U)} &= y_i - x \hat{\beta}_{(U)} \end{aligned}$$

2.20.2 Leverage

Leverage can be defined through the projection matrix that results from a transformation of the model with the inverse of the Cholesky decomposition of $\boldsymbol{V}$, or an oblique projector.

$\boldsymbol{Y} = \boldsymbol{H}\hat{\boldsymbol{Y}}$ While H is idempotent, it is generally not symmetric and thus not a projection matrix in the narrow sense.

$$h_{ii} = x_i'(X'X)^{-1}x_i$$

The trace of $\boldsymbol{H}$ equals the rank of $\boldsymbol{X}$. If V_{ij} denotes the element in row i, column j of $\boldsymbol{V}^{-1}$, then for a model containing only an intercept the diagonal elements of $\boldsymbol{H}$.

$$h_{ii} = \frac{\sum v_{ij}}{\sum\sum v_{ij}}$$

2.20.3 Internally and Externally Studentized Residuals

The computation of internally studentized residuals relies on the diagonal values of $\boldsymbol{V}(\hat{\boldsymbol{\theta}}) - \boldsymbol{Q}(\hat{\boldsymbol{\theta}})$ Externally studentized residuals require iterative influece analysis or a profiled residual variance.

Cook's Distance

$$\boldsymbol{\delta}_{(U)} = \hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(U)}$$

A DFFIT measures the change in predicted values due to the removal of data points. (Belsey, Kuh and Welsch (1980))

$\boldsymbol{D}(\boldsymbol{\beta}) = \boldsymbol{\delta}'_{(U)}\boldsymbol{\delta}_{(U)}/rank(\boldsymbol{X})$ Cook's D can be calibrated according to a chi-square distribution with degress of freedom equal to the rank of $\boldsymbol{X}$ **?**.

2.20.4 DFFITs and MDFFITs

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

2.20.5 Likelihood Distance

The log-likelihood function l and restricted log-likelihood l_R ofthe LME model. $\boldsymbol{\psi}$ is the collection of all parameters (i.e. the fixed effects $\boldsymbol{\beta}$ and the covariance parameters $\boldsymbol{\theta}$).

Reduced data estimates ($\boldsymbol{\psi}_{(U)}$)

$$RLD_{(U)} = 2\{l_R(\boldsymbol{\psi}) - l_R(\boldsymbol{\psi}_{(U)})\}$$

$$LD_{(U)} = 2\{l(\boldsymbol{\psi}) - l(\boldsymbol{\psi}_{(U)})\}$$

Likelihood distance, known as likelihood displacements. The likelihood distance gives twice the amount by which the log likelihood of the full data changes if one were to use an estimate based on fewer data points. The likelihood distance is the a global summary measure of the influence of the observations in U jointly on all parameters.

An overall influence statistic measures the change in the objective function being minimized. In linear mixed models cook and Weisberg devised the Likelihood distance, known elsewhere as likelihood displacement.

The likelihood distance gives the amount of data by which the log-likelihood found when using the full data sets woudld change when the data set is reduced.

Importantly the value $l(\hat{\phi})_{(U)}$ is not the log-likelihood obtained from the reduced set, but the determining the likelihood function based on the full set at the reduced data estimates.

The approach can be applied to ML and REML models. Schabenberger (2004) uses in notation the subscript R to specify that REML models are under consideration.

The likelihood distance LD of observation group U is given by

$$LD_U = 2l(\hat{\phi}) - 2l(\hat{\phi}_{(U)}) \tag{2.8}$$

where l is the log likelihood function. If LD_U is large then the observation group U is influential on the likelihood function. The likelihood distance is a global summary measure, expressing the joint influence of the observation group U on all parameters ϕ subject to updating.

That U is influential is not grounds for deletion or changing the model. Should U be found to be influential, Schabenberger (2004) advises that the nature of that influence be determined. Estimates of the coefficients and precision of fixed effects, the coefficients and precision of covariance parameters, and fitted and predicted values should all be examined in light of determining that U is influential. Influence may be exerted by U on covariance parameters without affecting the fixed effects.

The likelihood distance has been widely used to detect outlying observations in data analysis. [cite: Cook and Weisberg] suggested that the likelihood distance may be compared to a χ^2 distribution for large samples.

2.20.6 Non-iterative Update Procedures

The change in the fixed-effects estimates following removal of the observations in U is

$$\hat{\beta} - \hat{\beta}_{(U)} = \boldsymbol{\Omega X V}\,(\boldsymbol{UPU})$$

Residual variance

When σ^2 is profiled out of the marginal variance-covariance matrix, a closed-form estimate of σ^2 that is only based on only the remaining observation an be computed as follows, provided $\boldsymbol{V} = \boldsymbol{V}(\boldsymbol{\theta})$ [cite: Hurtado 1993]

Likelihood Distances

For noniterative methods the following computational devices are used to compute (restricted) likelihood distances provided that the residual variance σ^2 is profiled.

2.20.7 Miscellaneous

DFBETA

$$\begin{aligned} DFBETA_a &= \hat{\beta} - \hat{\beta}_{(a)} \qquad (2.9) \\ &= B(Y - Y_{\bar{a}} \qquad (2.10) \end{aligned}$$

Mean Square Prediction Error

$$MSPR = \frac{\sum(y_i - \hat{y}_i)^2}{n^*} \tag{2.11}$$

Effects on parameter estimate

Cook's Distance. CD.

Effects on the fitted and predicted values

Schabenberger (2004) descibes the use of the $PRESS$ and $DFFITS$ in determining influence.

The $PRESS$ residual is the difference between the observed value and the predicted (marginal)value.

$$\hat{e}_{i(U)} = y_i - x\hat{\beta}_{(U)} \tag{2.12}$$

2.21 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 ϕ that were subject to updating.

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.

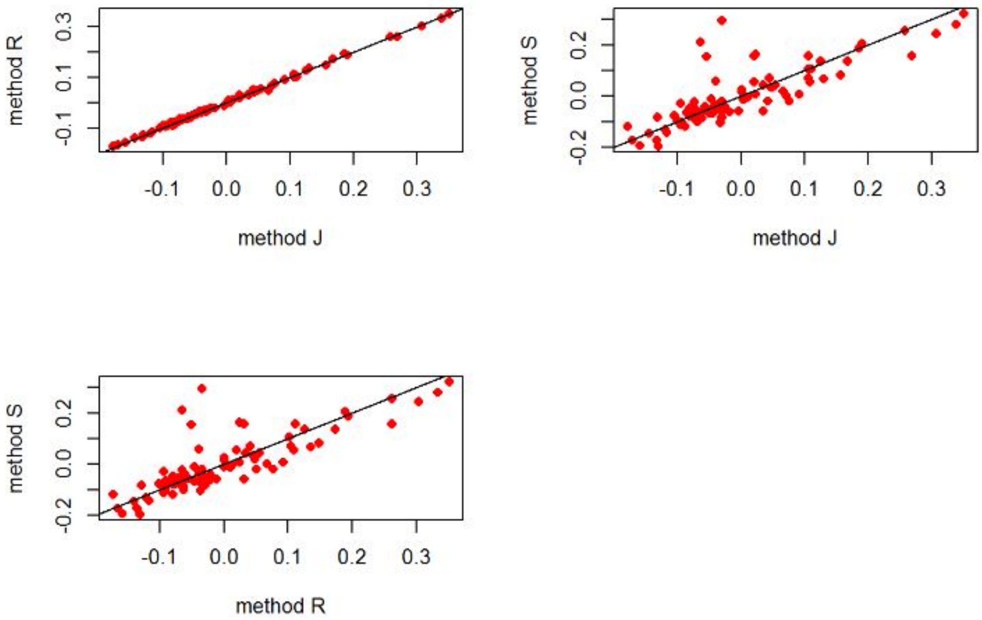
It is obtained by evaluating the likelihood function based on the full data set (containing all n observations) at the reduced-data estimates.

4. Using DFBETAs to Assess Agreement

Suppose an LME model was formulated to model agreement for various (i.e. 2 or 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.**

For the model fitted to the blood data with the lme4 R package, the results tabulated below can be produced. All 85 subjects are ranked by Cook's Distance (with only the top 6 being presented here). The remaining columns are the DFBeta for each of the fixed effects, for each of the 85 subject.

Subject	Cook's D	methodJ	methodR	methodS
78	0.61557407	-0.02934556	-0.03387780	0.2954937
80	0.41590973	-0.06305026	-0.06515241	0.2123881
68	0.22536651	-0.05334867	-0.05062375	0.1555187
72	0.09348500	0.02388626	0.02419887	0.1617474
48	0.08706988	0.02147541	0.03145273	0.1581591
30	0.07118415	0.26925807	0.26215970	0.1581569



In the first of the three plots (*Top Right*), strong agreement between method J and method R is indicated. The other plots indicate lack of agreement of methods J and R with method S.

If lack of agreement is indicated, a subsequent analysis using a technique proposed by Roy(2009) can be used to identify the specific cause for this lack of agreement (see next section).

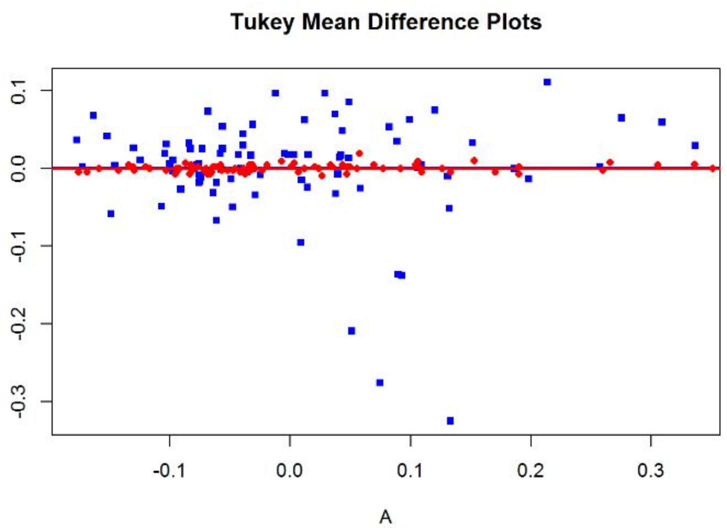
The Pearson Correlation coefficient of the DFBetas can be used in conjection with this analysis. A high correlation confirms good agreement. No threshold value for agreement is suggested, and analysts are advised to perform model diagnostics regardless of the correlation coeffient.

The Bonferroni Outlier Test and Cook's Distance values can be used to identify unusual cases, when the relationship between sets of dfbeta is modelled as a (classical) linear model. In this model, the covariates should be homoskedastic. A test for non-constant variance may be used to verify this. These diagnostic procedures are implementable using the ***car*** R package.

Deming Regression can be used to verify the line of 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 (This is the subject of the next section).



5. 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 these agreement criteria individually. If two methods of measuement lack agreement, the specific reason or reasons for this lack of agreement can be identified.

Roy proposes an LME model with Kronecker product covariance structure in a doubly multivariate setup. Response for ith subject can be written as

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + b_{1i} z_{i1} + b_{2i} z_{i2} + \epsilon_i$$

- β_1 and β_2 are fixed effects corresponding to both methods. (β_0 is the intercept.)
- b_{1i} and b_{2i} are random effects corresponding to both methods.

Overall variability between the two methods (Ω) is sum of between-subject (D) and within-subject variability (Σ),

$$\text{Block } \boldsymbol{\Omega}_i = \begin{bmatrix} d_1^2 & d_{12} \\ d_{12} & d_2^2 \end{bmatrix} + \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}.$$

6. 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, they are ubiquituous in literature, and a demonstration on how to compute them with the Roy Model would assist the adoption of this proposed method.

The coefficient of repeatability is encountered in Gage R & R analysis. *(A future exploration of how LME models can be used in that field would be of interest. This is something to include in the Conclusions Section).*

7. 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 & Burzykowski (2013) can be used to appraise Roy's model. These authors specifically look at Cook's Distances and Likelihood Distances. For the Roy Model, Cook's Distances may also be generated using the ***predictmeans***

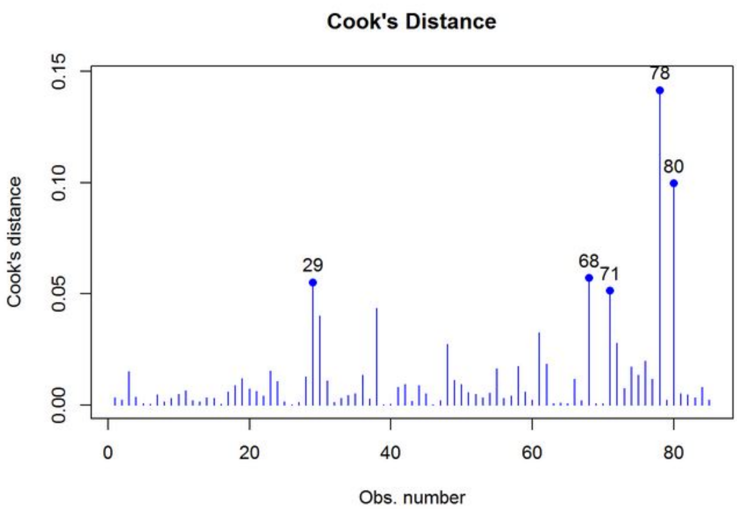


Figure 2.21.2:

As the model is structurally different from the models discussed in the earlier sections, Residual analysis will be briefly revisited.

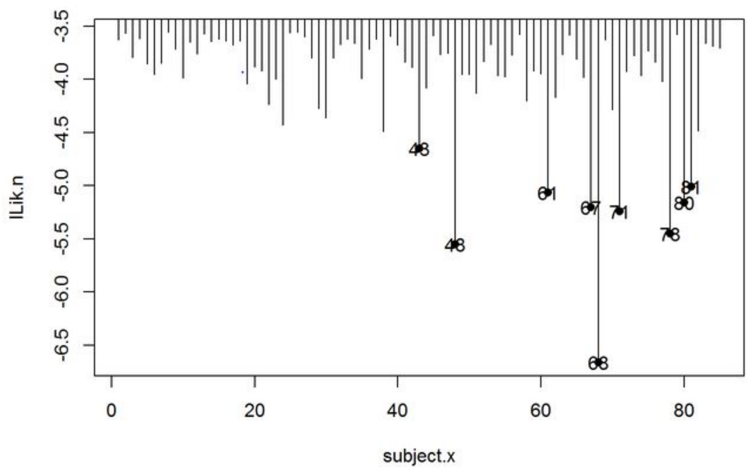


Figure 2.21.3:

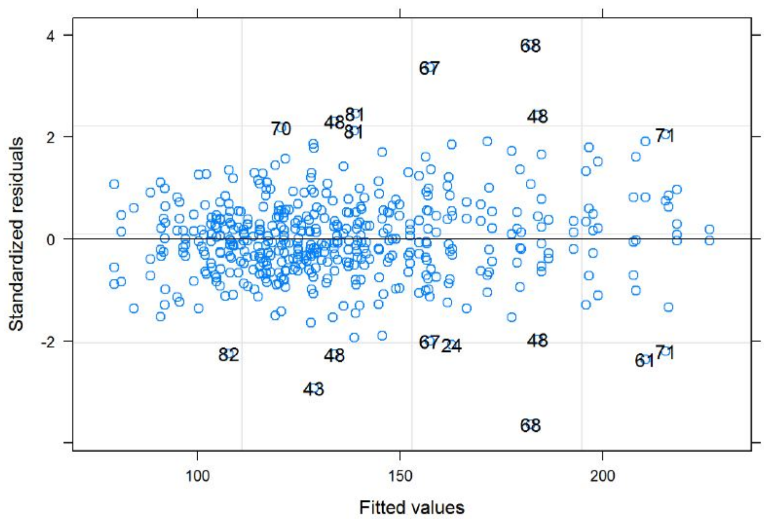


Figure 2.21.4:

8. Case Deletion Diagnostics for the Variance Ratios

Schabenberger advises on the use of deletion diagnostics for variance components of an LME model. Taking the core principals of his methods, and applying them to the Method Comparison problem, case deletion diagnostics are used on the variance components of the Roy model., specifically the ratio of between subject variances and the within subject covariances respecitvely.

$$\text{BSVR} = \frac{\sigma_2^2}{\sigma_2^2} \qquad\qquad \text{WSVR} = \frac{d_2^2}{d_2^2}$$

These variance ratios are re-computed for each case removed, and may be analysed seperately or jointly for outliers.

The Grubbs' Test for Outliers is a commonly used technique for assessing outlier in a univariate data set. As there may be several outliers (i.e. influential cases) present, the Grubbs test is not practical. However outlier detection using to Tukey's specification for boxplots (i.e. greater than $Q_3 + 1.5IQR$ or less than $Q_1 - 1.5IQR$), will suffice. Ranking the asbolute values of the standardizaed scores can also can be used to identify influential cases, even if the data is not normally distributed.

Bivariate Analyses may be applied jointly to the both sets of data sets, e.g Mahalanobis distances. The Mahalanobis distance, while not an intuitive measure in the context of the data, can be used to rank highly influential cases.

9. Permutation Test, Power Tests and Missing Data

This section explores topics such as dependent variable simulation and power analysis, introduced by Galecki & Burzykowski (2013), and implementable with their ***nlmeU*** `R` package. Using the ***predictmeans*** `R` package, it is possible to perform permutation t-tests for coefficients of (fixed) effects and permutation F-tests.

The matter of missing data has not been commonly encountered in either Method Comparison Studies or Linear Mixed Effects Modelling. However Roy (2009) deals with the relevant assumptions regrading missing data. Galecki & Burzykowski (2013) approaches the subject of missing data in LME Modelling. The ***nlmeU*** package includes the `patMiss` function, which "*allows to compactly present pattern of missing data in a given vector/matrix/data frame or combination of thereof*".

2.21.1 Influential Observations : DFBeta and DFBetas

2.22 Effects on fitted and predicted values

$$\hat{e}_{i(U)} = y_i - x\hat{\beta}_{(U)} \tag{2.13}$$

2.23 Cooks's Distance

Cook's Distance is a measure indicating to what extent model parameters are influenced by (a set of) influential data on which the model is based. This function computes the Cook's distance based on the information returned by the estex() function.

2.23.1 Cook's Distance

- For variance components γ: $CD(\gamma)_i$,
- For fixed effect parameters β: $CD(\beta)_i$,
- For random effect parameters $\boldsymbol{u}$: $CD(u)_i$,
- For linear functions of $\hat{beta}$: $CD(\psi)_i$

Random Effects

A large value for $CD(u)_i$ indicates that the $i-$th observation is influential in predicting random effects.

linear functions

$CD(\psi)_i$ does not have to be calculated unless $CD(\beta)_i$ is large.

Cook (1977) greatly expanded the study of residuals and influence measures. Cook's key observation was the effects of deleting each observation in turn could be computed without undue additional computational expense. Consequently deletion diagnostics have become an integral part of assessing linear models.

Cook (1986) gave a completely general method for assessing influence of local departures from assumptions in statistical models.

2.23.2 Cook's Distance

In classical linear regression, a commonly used meausre of influence is Cook's distance. It is used as a measure of influence on the regression coefficients.

For linear mixed effects models, Cook's distance can be extended to model influence diagnostics by definining.

$$C_{\beta i} = \frac{(\hat{\beta} - \hat{\beta}_{[i]})^T (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})(\hat{\beta} - \hat{\beta}_{[i]})}{p}$$

It is also desirable to measure the influence of the case deletions on the covariance matrix of $\hat{\beta}$.

2.23.3 Cooks's Distance

Cook's D statistics (i.e. colloquially Cook's Distance) is a measure of the influence of observations in subset U on a vector of parameter estimates (Cook, 1977).

$$\delta_{(U)} = \hat{\beta} - \hat{\beta}_{(U)}$$

If V is known, Cook's D can be calibrated according to a chi-square distribution with degrees of freedom equal to the rank of $\boldsymbol{X}$ (**?**).

For LME models, Cook's distance can be extended to model influence diagnostics by defining.

It is also desirable to measure the influence of the case deletions on the covariance matrix of $\hat{\beta}$.

2.24 Cook's Distance for LMEs

Diagnostic methods for fixed effects are generally analogues of methods used in classical linear models. Diagnostic methods for variance components are based on 'one-step' methods. Cook (1986) gives a completely general method for assessing the influence of local departures from assumptions in statistical models.

For fixed effects parameter estimates in LME models, the Cook's distance can be extended to measure influence on these fixed effects.

$$\mathrm{CD}_i(\beta) = \frac{(c_{ii} - r_{ii}) \times t_i^2}{r_{ii} \times p}$$

For random effect estimates, the Cook's distance is

$$\mathrm{CD}_i(b) = g\prime_{(i)}(I_r + \mathrm{var}(\hat{b})D)^{-2}\mathrm{var}(\hat{b})g_{(i)}.$$

Large values for Cook's distance indicate observations for special attention.

Cook's Distance

Cooks Distance (D_i) is an overall measure of the combined impact of the ith case of all estimated regression coefficients. It uses the same structure for measuring the combined impact of the differences in the estimated regression coefficients when the $i-$th case is deleted.

Importantly, $D_{(i)}$ can be calculated without fitting a new regression coefficient each time an observation is deleted.

Cook (1977) greatly expanded the study of residuals and influence measures. Cook's key observation was the effects of deleting each observation in turn could be computed without undue additional computational expense. Consequently deletion diagnostics have become an integral part of assessing linear models.

Cook's Distance is a well known diagnostic technique used in classical linear models, extended to LME models. For LME models, two formulations exist; a Cook's distance that examines the change in fixed fixed parameter estimates, and another that examines the change in random effects parameter estimates. The outcome of either Cook's distance is a scaled change in either β or θ.

2.24.1 Cook's Distance

In statistics, Cook's Distance or Cook's D is a commonly used estimate of the influence of a data point when performing least squares regression analysis.[1] In a practical ordinary least squares analysis, Cook's distance can be used in several ways: to indicate data points that are particularly worth checking for validity; to indicate regions of the design space where it would be good to be able to obtain more data points. It is named after the American statistician R. Dennis Cook, who introduced the concept in 1977.

2.24.2 Change in the precision of estimates

The effect on the precision of estimates is separate from the effect on the point estimates. Data points that have a small Cook's distance, for example, can still greatly affect hypothesis tests and confidence intervals, if their influence on the precision of the estimates is large.

Interpretation

Specifically D_i can be interpreted as the distance one's estimates move within the confidence ellipsoid that represents a region of plausible values for the parameters.[clarification needed] This is shown by an alternative but equivalent representation of Cook's distance in terms of changes to the estimates of the regression parameters between the cases where the particular observation is either included or excluded from the regression analysis.

2.24.3 Cook's Distance

Some texts tell you that points for which Cook's distance is higher than 1 are to be considered as influential. Other texts give you a threshold of $4/N$ or $4/(Nk1)$, where N is the number of observations and k the number of explanatory variables. In your case the latter formula should yield a threshold around 0.1 .

John Fox (1), in his booklet on regression diagnostics is rather cautious when it comes to giving numerical thresholds. He advises the use of graphics and to examine in closer details the points with "values of D that are substantially larger than the rest". According to Fox, thresholds should just be used to enhance graphical displays.

In your case the observations 7 and 16 could be considered as influential. Well, I would at least have a closer look at them. The observation 29 is not substantially different from a couple of other observations.

(1) Fox, John. (1991). Regression Diagnostics: An Introduction. Sage Publications.

Chapter 3

Case Deletion Diagnostics for Mixed Models

3.1 Case Deletion Diagnostics for LME models

Haslett and Dillane (2004) remark that linear mixed effects models didn't experience a corresponding growth in the use of deletion diagnostics, adding that McCullough and Searle (2001) makes no mention of diagnostics whatsoever.

Christensen et al. (1992) describes three propositions that are required for efficient case-deletion in LME models. The first proposition decribes how to efficiently update V when the ith element is deleted.

$$V_{[i]}^{-1} = \Lambda_{[i]} - \frac{\lambda\lambda\prime}{\nu ii} \tag{3.1}$$

The second of christensen's propostions is the following set of equations, which are variants of the Sherman Wood bury updating formula.

$$\begin{aligned}
X'_{[i]}V_{[i]}^{-1}X_{[i]} &= X'V^{-1}X - \frac{\hat{x}_i\hat{x}'_i}{s_i} && (3.2)\\
(X'_{[i]}V_{[i]}^{-1}X_{[i]})^{-1} &= (X'V^{-1}X)^{-1} + \frac{(X'V^{-1}X)^{-1}\hat{x}_i\hat{x}'_i(X'V^{-1}X)^{-1}}{s_i - \bar{h}_i} && (3.3)\\
X'_{[i]}V_{[i]}^{-1}Y_{[i]} &= X\prime V^{-1}Y - \frac{\hat{x}_i\hat{y}'_i}{s_i} && (3.4)
\end{aligned}$$

Schabenberger (2004) notes that it is not always possible to derive influence statistics necessary for comparing full- and reduced-data parameter estimates. Haslett and Dillane (2004) offers an procedure to assess the influences for the variance components within the linear model, complementing the existing methods for the fixed components. The essential problem is that there is no useful updating procedures for $\hat{V}$, or for $\hat{V}^{-1}$. Haslett and Dillane (2004) propose an alternative , and computationally inexpensive approach, making use of the 'delete=replace' identity.

Haslett (1999) considers the effect of 'leave k out' calculations on the parameters β and σ^2, using several key results from Haslett and Hayes (1998) on partioned matrices.

3.1.1 Case Deletion Diagnostics

? develops case deletion diagnostics, in particular the equivalent of Cook's distance, for diagnosing influential observations when estimating the fixed effect parameters and variance components.

3.1.2 Effects on fitted and predicted values

$$\hat{e}_{i(U)} = y_i - x\hat{\beta}_{(U)} \tag{3.5}$$

3.1.3 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

A general method for comparing nested models fit by maximum liklihood is the liklihood ratio test. This test can be used for models fit by REML (restricted maximum liklihood), but only if the fixed terms in the two models are invariant, and both models have been fit by REML. Otherwise, the argument: method=ML must be employed (ML = maximum liklihood).

Example of a liklihood ratio test used to compare two models:

The output will contain a p-value, and this should be used in conjunction with the AIC scores to judge which model is preferred. Lower AIC scores are better.

Generally, liklihood ratio tests should be used to evaluate the significance of terms on the random effects portion of two nested models, and should not be used to determine the significance of the fixed effects.

A simple way to more reliably test for the significance of fixed effects in an LME model is to use conditional F-tests, as implemented with the simple anova function.

Example:

will give the most reliable test of the fixed effects included in model1.

3.1.4 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.

? 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.

3.1.5 Case deletion notation

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

3.1.6 Partitioning Matrices

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

3.1.7 Case Deletion Diagnostics

? develops case deletion diagnostics, in particular the equivalent of Cook's distance, for diagnosing influential observations when estimating the fixed effect parameters and variance components.

3.1.8 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

3.2 Overview

1. Extending deletion diagnostics to LMEs
2. Christensen et al
3. Haslett hayes
4. Schabenberger
5. Tewomir

3.3 Case Deletion Diagnostics for Mixed Models

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.

An iterative analysis may seem computationally expensive. computing iterative influence diagnostics for n observations requires $n+1$ mixed models to be fitted iteratively.

3.3.1 Extending deletion diagnostics to LMEs

after fitting a mixed, it is important to carry put model diagnostics to check whether distributional assumptions for the residuals as satisfied and whether the fit the model is sensitive to unusual assumptions. The process of carrying out model diagnostic involves several informal and formal techniques.

$$X = \begin{bmatrix} x_i' \\ X(i) \end{bmatrix}, Z = \begin{bmatrix} z_{ij}' \\ Z_{j(i)} \end{bmatrix}, Z = \begin{bmatrix} z_{ij}' \\ Z_{j(i)} \end{bmatrix},$$

$$y = \begin{bmatrix} y_{ij}' \\ y_{j(i)} \end{bmatrix} \text{ and } H = \begin{bmatrix} h_{ii} & h \\ h_{j(i)} & h \end{bmatrix}$$

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

$\boldsymbol{a}_{(i)}$ denotes a vector $\boldsymbol{a}$ with the i-th element, a_i, removed.

$$\breve{a}_i = \boldsymbol{a}_i - \boldsymbol{A}_{(i)} \boldsymbol{H}_{[i]} \boldsymbol{h}_i \tag{3.6}$$

3.3.2 Influence on measure component ratios

The general diagnostic tools for variance component ratios are the analogues of the Cook's distance and the information ratio.

The analogue of Cooks distance measure for variance components γ is denoted

$CD(\gamma)$.

$$
\begin{aligned}
CD_U(\gamma) = (\hat{\gamma}_{(U)} - \hat{\gamma})'[\text{var}(\hat{\gamma})]^{-1}(\hat{\gamma}_{(U)} - \hat{\gamma}) \\
&= -\boldsymbol{g'}_{(U)}(\boldsymbol{Q}-\boldsymbol{G})^{-1}\boldsymbol{Q}(\boldsymbol{Q}-\boldsymbol{G})\boldsymbol{g}_{(U)} \\
&= \boldsymbol{g'}_{(U)}(\boldsymbol{I}_r + \text{var}(\hat{\gamma})\boldsymbol{G})^{-2}\text{var}(\hat{\gamma})\boldsymbol{g}_{(U)}
\end{aligned}
$$

Large values of $CD(\gamma)$ highlight observation groups for closer attentions

The analogue of the information ratio measures the change in the determinant of the maximum likelihood estimates information matrix

$$IR\gamma = \frac{\det(\boldsymbol{Q}-\boldsymbol{G})}{\det(\boldsymbol{Q})}$$

Ideally when all observations have the same influence on the information matrix $IR\gamma$ is approximately one. Deviations from one indicate the group U is influential. Since $\text{var}(\hat{\gamma})$ and $\boldsymbol{I}_r$ are fixed for all observations, $IR\gamma$ is a function of $\boldsymbol{G}$, in turn a function of $\boldsymbol{C}_i$ and c_{ii}.

3.4 Christensen et al

Christensen, Pearson and Johnson (1992) (hereafter CPJ) studied case deletion diagnostics, in particular the analog of Cooks distance, for diagnosing influential observations when estimating the fixed effect parameters and variance components.

?provides an overview of case deletion diagnostics for fixed effect models.

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.

? 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

3.5 Haslett Hayes

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.

A general theory is presented for residuals from the general linear model with correlated errors. It is demonstrated that there are two fundamental types of residual associated with this model, referred to here as the marginal and the conditional residual. These measure respectively the distance to the global aspects of the model as represented by the expected value and the local aspects as represented by the conditional expected value. These residuals may be multivariate.

In contrast to classical linear models, diagnostics for LME are difficult to perform and interpret, because of the increased complexity of the model

3.6 Schabenberger

Standard residual and influence diagnostics for linear models can be extended to linear mixed models. The dependence of fixed-effects solutions on the covariance parameter estimates has important ramifications in perturbation analysis. To gauge the full impact of a set of observations on the analysis, covariance parameters need to be updated, which requires refitting of the model.

The conditional (subject-specific) and marginal (population-averaged) formulations in the linear mixed model enable you to consider conditional residuals that use the estimated BLUPs of the random effects, and marginal residuals which are deviations from the overall mean. Residuals using the BLUPs are useful to diagnose whether the random effects components in the model are specified correctly, marginal residuals are useful to diagnose the fixed-effects components.

3.7 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 doesnt increase the procedures execution time by the same degree.

estimation

$$\hat{\beta} = X^T \quad (3.7)$$

$$\hat{\gamma} = G(\hat{\theta})Z^T \quad (3.8)$$

The difference between perturbation and residual analysis between the linear and LME models. The estimates of the fixed effects β depend on the estimates of the covariance parameters.

3.8 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.

Deletion diagnostics are not commonly used with the LME models, as yet.

3.9 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.

3.9.1 Case-Deletion results for Variance components

Christensen et al. (1992)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.

3.10 Cook's Distance for LMEs

Cook's Distance is a well known diagnostic technique used in classical linear models, extended to LME models. For LME models, two formulations exist; a Cook's distance that examines the change in fixed fixed parameter estimates, and another that examines the change in random effects parameter estimates. The outcome of either Cook's distance is a scaled change in either β or θ.

Diagnostic methods for fixed effects are generally analogues of methods used in classical linear models. Diagnostic methods for variance components are based on 'one-step' methods. *Cook (1986)* gives a completely general method for assessing the influence of local departures from assumptions in statistical models.

For fixed effects parameter estimates in LME models, the Cook's distance can be extended to measure influence on these fixed effects.

$$\mathrm{CD}_i(\beta) = \frac{(c_{ii} - r_{ii}) \times t_i^2}{r_{ii} \times p}$$

For random effect estimates, the Cook's distance is

$$\mathrm{CD}_i(b) = g\prime_{(i)}(I_r + \mathrm{var}(\hat{b})D)^{-2}\mathrm{var}(\hat{b})g_{(i)}.$$

Large values for Cook's distance indicate observations for special attention.

3.10.1 Change in the precision of estimates

The effect on the precision of estimates is separate from the effect on the point estimates. Data points that have a small Cook's distance, for example, can still greatly affect hypothesis tests and confidence intervals, if their influence on the precision of the estimates is large.

Cook's distance

In the study of Linear model diagnostics, Cook proposed a measure that combines the information of leverage and residual of the observation, now known simply as the Cook's Distance. **?** would later adapt the Cook's distance measure for the analysis of LME models.

3.11 Cook's Distance

Cooks Distance (D_i) is an overall measure of the combined impact of the ith case of all estimated regression coefficients. It uses the same structure for measuring the combined impact of the differences in the estimated regression coefficients when the kth case is deleted. $D_{(k)}$ can be calculated without fitting a new regression coefficient each time an observation is deleted. Cook's D statistics (i.e. colloquially Cook's Distance) is a measure of the influence of observations in subset U on a vector of parameter estimates (Cook, 1977).

$$\delta_{(U)} = \hat{\beta} - \hat{\beta}_{(U)}$$

If V is known, Cook's D can be calibrated according to a chi-square distribution with degrees of freedom equal to the rank of $\boldsymbol{X}$ (**?**).

3.12 Cook's Distance

- For variance components γ: $CD(\gamma)_i$,
- For fixed effect parameters β: $CD(\beta)_i$,
- For random effect parameters $\boldsymbol{u}$: $CD(u)_i$,
- For linear functions of $\hat{beta}$: $CD(\psi)_i$

It is also desirable to measure the influence of the case deletions on the covariance matrix of $\hat{\beta}$.

Random Effects

A large value for $CD(u)_i$ indicates that the $i-$th observation is influential in predicting random effects.

linear functions

$CD(\psi)_i$ does not have to be calculated unless $CD(\beta)_i$ is large.

3.13 Exention of Cook's Distance methodology to LME models

Cook's Distance is extended to LME models. For LME models, two formulations exist; a Cook's distance that examines the change in fixed fixed parameter estimates, and another that examines the change in random effects parameter estimates. The outcome of either Cook's distance is a scaled change in either β or θ.

Diagnostic methods for variance components are based on 'one-step' methods. Cook (1986) gives a completely general method for assessing the influence of local departures from assumptions in statistical models. For fixed effects parameter estimates in LME models, the Cook's distance can be extended to measure influence on these fixed effects.

$$\mathrm{CD}_i(\beta) = \frac{(c_{ii} - r_{ii}) \times t_i^2}{r_{ii} \times p}$$

For random effect estimates, the Cook's distance is

$$\mathrm{CD}_i(b) = g\prime_{(i)}(I_r + \mathrm{var}(\hat{b})D)^{-2}\mathrm{var}(\hat{b})g_{(i)}.$$

Large values for Cook's distance indicate observations for special attention.

Cook's Distance was extended from classical linear models to LME models. For linear mixed effects models, Cook's distance can be extended to model influence diagnostics by definining.

$$CD_{\beta i} = \frac{(\hat{\beta} - \hat{\beta}_{[i]})^T (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})(\hat{\beta} - \hat{\beta}_{[i]})}{p}$$

It is also desirable to measure the influence of the case deletions on the covariance matrix of $\hat{\beta}$.

3.14 Cook's Distance for LMEs

Diagnostic methods for fixed effects are generally analogues of methods used in classical linear models. Diagnostic methods for variance components are based on 'one-step' methods. Cook (1986) gives a completely general method for assessing the influence of local departures from assumptions in statistical models.

For fixed effects parameter estimates in LME models, the Cook's distance can be extended to measure influence on these fixed effects.

$$\mathrm{CD}_i(\beta) = \frac{(c_{ii} - r_{ii}) \times t_i^2}{r_{ii} \times p}$$

For random effect estimates, the Cook's distance is

$$\mathrm{CD}_i(b) = g\prime_{(i)}(I_r + \mathrm{var}(\hat{b})D)^{-2}\mathrm{var}(\hat{b})g_{(i)}.$$

Large values for Cook's distance indicate observations for special attention.

Random Effects

A large value for $CD(u)_i$ indicates that the $i-$th observation is influential in predicting random effects.

linear functions

$CD(\psi)_i$ does not have to be calculated unless $CD(\beta)_i$ is large.

For LME models, two formulations exist; a Cook's distance that examines the change in fixed fixed parameter estimates, and another that examines the change in random effects parameter estimates. The outcome of either Cook's distance is a scaled change in either β or θ.

3.15 Cook's Distance for LMEs

Diagnostic methods for fixed effects are generally analogues of methods used in classical linear models. Diagnostic methods for variance components are based on 'one-step' methods. Cook (1986) gives a completely general method for assessing the influence of local departures from assumptions in statistical models.

For fixed effects parameter estimates in LME models, the Cook's distance can be extended to measure influence on these fixed effects.

$$\mathrm{CD}_i(\beta) = \frac{(c_{ii} - r_{ii}) \times t_i^2}{r_{ii} \times p}$$

For random effect estimates, the Cook's distance is

$$\mathrm{CD}_i(b) = g\prime_{(i)}(I_r + \mathrm{var}(\hat{b})D)^{-2}\mathrm{var}(\hat{b})g_{(i)}.$$

Large values for Cook's distance indicate observations for special attention.

3.16 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 ϕ that were subject to updating.

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 ψ that were subject to updating.

3.17 Grubbs' data

Let $\hat{\beta}$ denote the least square estimate of β based upon the full set of observations, and let $\hat{\beta}^{(k)}$ denoted the estimate with the k^{th} case excluded. For the Grubbs data the $\hat{\beta}$ estimated are $\hat{\beta}_0$ and $\hat{\beta}_1$ respectively. Leaving the fourth case out, i.e. $k = 4$ the corresponding estimates are $\hat{\beta}_0^{-4}$ and $\hat{\beta}_1^{-4}$

$$Y^{(k)} = \hat{\beta}^{(k)} X^{(k)} \tag{3.9}$$

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

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


Coefficients: (Intercept)            A

-37.51896      0.04656
```

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

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

$$D^{-Q} = \hat{\beta}^{-Q} A^{-Q} \tag{3.10}$$

3.17.1 Grubb's example

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

$$Y^{-Q} = \hat{\beta}^{-Q} X^{-Q} \tag{3.11}$$

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

$$D^{-Q} = \hat{\beta}^{-Q} A^{-Q} \tag{3.12}$$

3.17.2 Influence measures using R

R provides the following influence measures of each observation.

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

Chapter 4

LME Likelihood

4.1 One Way ANOVA

4.1.1 Page 448

Computing the variance of $\hat{\beta}$

$$\text{var}(\hat{\beta}) = (X'V^{-1}X)^{-}1 \tag{4.1}$$

It is not necessary to compute V^{-1} explicitly.

$$\begin{aligned} V^{-1}X &= \Sigma^{1}X - Z()Z'\Sigma^{-1}X & (4.2) \\ &= \Sigma^{-1}(X - Zb_x) & (4.3) \end{aligned}$$

The estimate b_x is the same term obtained from the random effects model; $X = Zb_x + e$, using X as an outcome variable. This formula is convenient in applications where b_x can be easily computed. Since X is a matrix of p columns, b_x can simple be computed column by column. according to the columns of X.

4.1.2 Page 448- simple example

Consider a simple model of the form;

$$y_{ij} = \mu + \beta_i + \epsilon_{ij}.$$

The iterative procedure is as follows Evaluate the individual group mean $\bar{y}_i$ and variance $\hat{Sigma^2}_i$. Then use the variance of the group means as an estimate of the σ^2_b. The average of the the variances of the groups is the initial estimate of the σ^2_e.

Iterative procedure

The iterative procedure comprises two steps, with 0 as the first approximation of b_i.

The first step is to compute λ, the ratio of variabilities,

$$\lambda = \frac{\sigma^2_b}{\sigma^2_e}$$

$$\mu = \frac{1}{N}\sum_{ij}(y_{ij} - b_i)$$

$$b_i = \frac{n(\bar{y}_i - \mu)}{n + \lambda}$$

The second step is to updat $sigma^2_e$

$$\sigma^2_e = \frac{e'e}{N - df} \tag{4.4}$$

where e is the vector of $e_{ij} = y_{ij} - \mu - b_i$ and $df = qn/n + \lambda$ and

$$\sigma^2_b = \frac{1}{q}\sum_{i=1}^{q} b^2_1 + (\frac{n}{\sigma^2_e} + \frac{1}{\sigma^2_b})^{-1} \tag{4.5}$$

Worked Example

Further to [pawitan 17.1] the initial estimates for variability are $\sigma_b^2 = 1.7698$ and $\sigma_e^2 = 0.3254$. At convergence the following results are obtained.
n=16, q=5

$$\hat{\mu} = \bar{y} = 14.175$$
$$\hat{\sigma}^2 = 0.325$$
$$\hat{\sigma}_b^2 = 1.395$$
$$\sigma = 0.986$$

At convergene the following estimates are obtained,

$$\hat{\mu} = 14.1751$$
$$\hat{b} = (-0.6211, 0.2683, 1.4389, -1.914, 0.8279)$$
$$\hat{\sigma}_b^2 = 1.3955$$
$$\hat{\sigma}_e^2 = 0.3254$$

4.1.3 Extention to several random effects

[pawitan section 17.7]

4.2 Classical model for single measurements

The classical model is based on measurements y_{mi} by method $m = 1, 2$ on item $i = 1, 2 \ldots$

$$y_{mi} + \alpha_m + \mu_i + e_{mi}$$

$$e_{mi} \sim \backslash(0, \sigma^2_m)$$

Even though the separate variances can not be identified, their sum can be estimated by the empirical variance of the differences.

Like wise the separate α can not be estimated, only theiir difference can be estimated as $\bar{D}$

In the first instance, we require a simple model to describe a measurement by method m. We use the term *item* to denote an individual, subject or sample, to be measured, being randomly sampled from a population. Let y_{mi} be the measurement for item i made by method m.

$$y_{mi} = \alpha_m + \mu_i + e_{mi}$$

- α_m is the fixed effect associated with method m,
- μ_i is the true value for subject i (fixed effect),
- e_{mi} is a random effect term for errors with $e_{mi} \sim \mathcal{N}(0, \sigma^2_m)$.

.

This model implies that the difference between the paired measurements can be expressed as

$$d_i = y_{1i} - y_{2i} \sim \mathcal{N}(\alpha_1 - \alpha_2, \sigma^2_1 - \sigma^2_2).$$

Importantly, this is independent of the item levels μ_i. As the case-wise differences are of interest, the parameters of interest are the fixed effects for methods α_m.

$$y_{mi} = \alpha_m + \mu_i + e_{mi}$$

Importantly these variance covariance structures are central to Roy methodology.

Roy (2009) proposes a series of hypothesis tests based on these matrices as part of her methodology. These tests shall be reverted to in due course.

The standard deviation of the differences of variables a and b is computed as

$$\mathrm{var}(a-b) = \mathrm{var}(a) + \mathrm{var}(b) - 2\mathrm{cov}(a,b)$$

Hence the variance of the difference of two methods, that allows for the calculation of the limits of agreement, can be calculated as

$$\mathrm{var}(d) = \omega_1^2 + \omega_2^2 - 2 \times \omega_1 2$$

4.3 Sampling

One important feature of replicate observations is that they should be independent of each other. In essence, this is achieved by ensuring that the observer makes each measurement independent of knowledge of the previous value(s). This may be difficult to achieve in practice. (Check who said this)

4.4 Limits of agreement in LME models

Limits of agreement are used extensively for assessing agreement, due to they're being intuitive and easy to use. Necessarily their prevalence in literature has meant that they are now the best known measurement for agreement, and that any newer methodology would benefit by making reference to them.

Carstensen et al. (2008) uses LME models to determine the limits of agreement. Between-subject variation for method m is given by d_m^2 and within-subject variation is given by λ_m^2. Carstensen et al. (2008) remarks that for two methods A and B, separate values of d_A^2 and d_B^2 cannot be estimated, only their average. Hence the assumption that $d_x = d_y = d$ is necessary. The between-subject variability $\boldsymbol{D}$ and within-subject variability $\boldsymbol{\Lambda}$ can be presented in matrix form,

$$\boldsymbol{D} = \begin{pmatrix} d_A^2 & 0 \\ 0 & d_B^2 \end{pmatrix} = \begin{pmatrix} d^2 & 0 \\ 0 & d^2 \end{pmatrix}, \qquad \boldsymbol{\Lambda} = \begin{pmatrix} \lambda_A^2 & 0 \\ 0 & \lambda_B^2 \end{pmatrix}. \tag{4.6}$$

The variance for method m is $d_m^2 + \lambda_m^2$. Limits of agreement are determined using the standard deviation of the case-wise differences between the sets of measurements by two methods A and B, given by

$$\mathrm{var}(y_A - y_B) = 2d^2 + \lambda_A^2 + \lambda_B^2. \tag{4.7}$$

Importantly the covariance terms in both variability matrices are zero, and no covariance component is present.

Carstensen et al. (2008) presents a data set 'fat', which is a comparison of measurements of subcutaneous fat by two observers at the Steno Diabetes Center, Copenhagen. Measurements are in millimeters (mm). Each person is measured three times by each observer. The observations are considered to be 'true' replicates.

A linear mixed effects model is formulated, and implementation through several software packages is demonstrated. All of the necessary terms are presented in the computer output. The limits of agreement are therefore,

$$0.0449 \pm 1.96 \times \sqrt{2 \times 0.0596^2 + 0.0772^2 + 0.0724^2} = (-0.220, 0.309). \tag{4.8}$$

? has demonstrated a methodology whereby d_A^2 and d_B^2 can be estimated separately. Also covariance terms are present in both $\boldsymbol{D}$ and $\boldsymbol{\Lambda}$. Using Roy's methodology, the variance of the differences is

$$\mathrm{var}(y_{iA} - y_{iB}) = d_A^2 + \lambda_B^2 + d_A^2 + \lambda_B^2 - 2(d_{AB} + \lambda_{AB}) \tag{4.9}$$

All of these terms are given or determinable in computer output. The limits of agreement can therefore be evaluated using

$$\bar{y_A} - \bar{y_B} \pm 1.96 \times \sqrt{\sigma_A^2 + \sigma_B^2 - 2(\sigma_{AB})}. \tag{4.10}$$

For Carstensen's 'fat' data, the limits of agreement computed using Roy's method are consistent with the estimates given by Carstensen et al. (2008); $0.044884 \pm 1.96 \times 0.1373979 = (-0.224, 0.314)$.

4.4.1 Linked replicates

Carstensen et al. (2008) proposes the addition of an random effects term to their model when the replicates are linked. This term is used to describe the 'item by replicate' interaction, which is independent of the methods. This interaction is a source of variability independent of the methods. Therefore failure to account for it will result in variability being wrongly attributed to the methods.

Carstensen et al. (2008) introduces a second data set; the oximetry study. This study done at the Royal Children?s Hospital in Melbourne to assess the agreement between co-oximetry and pulse oximetry in small babies.

In most cases, measurements were taken by both method at three different times. In some cases there are either one or two pairs of measurements, hence the data is unbalanced. Carstensen et al. (2008) describes many of the children as being very sick, and with very low oxygen saturations levels. Therefore it must be assumed that a biological change can occur in interim periods, and measurements are not true replicates.

Carstensen et al. (2008) demonstrate the necessity of accounting for linked replicated by comparing the limits of agreement from the 'oximetry' data set using a model with the additional term, and one without. When the interaction is accounted for the limits of agreement are (-9.62,14.56). When the interaction is not accounted for, the limits of agreement are (-11.88,16.83). It is shown that the failure to include this additional term results in an over-estimation of the standard deviations of differences.

Limits of agreement are determined using Roy's methodology, without adding any additional terms, are are found to be consistent with the 'interaction' model; $(-9.562, 14.504)$. Roy's methodology assumes that replicates are linked. However, following Carstensen's example, an addition interaction term is added to the implementation of Roy's model to assess the effect, the limits of agreement estimates do not change. However there is a conspicuous difference in within-subject matrices of both models (denoted 1 and 2 respectively);

```
\begin{equation}
```

```
\hat{\boldsymbol{\Lambda}}_{1}= \pmatrix{
16.61 & 11.67\cr
11.67 & 27.65 }\qquad
\boldsymbol{\hat{\Lambda}}_{2}= \pmatrix{
7.55 & 2.60 \cr
2.60 & 18.59}
\end{equation}
```

The variance of the additional random effect in model 2 is 3.01.

The Akaike information criterion (AIC) for both of models are $AIC_1 = 2304.226$ and $AIC_2 = 2306.226$. Having a difference of AIC values of 2 is equivalent to both models being equally as good as the other. The AIC values for the Carstensen 'unlinked' and 'linked' models are 1994.66 and 1955.48 respectively.

The $\hat{\boldsymbol{\Lambda}}$ matrices are informative as to the difference between Carstensen's unlinked and linked models. For the oximetry data, the covariance terms (given above as 11.67 and 2.6 respectively) are of similar magnitudes to the variance terms. Conversely for the 'fat' data the covariance term (-0.00032) is negligible. When the interaction term is added to the model, the covariance term remains negligible. (For the 'fat' data, the difference in AIC values is also 2).

To conclude, Carstensen's models provided a rigorous way to determine limits of agreement, but don't provide for the computation of $\hat{\boldsymbol{D}}$ and $\hat{\boldsymbol{\Lambda}}$. Therefore the test's proposed by **?** can not be implemented. Conversely, accurate limits of agreement can be found using Roy's method.

Addition of the interaction term erodes the capability of Roy's methodology to compare candidate models, and therefore shall not be adopted.

(N.B. To complement the blood pressure 'J vs S' analysis, the limits of agreement are $15.62 \pm 1.96 \times 20.33 = (-24.22, 55.46)$.)

4.5 Conclusion

Carstensen et al. (2008) and **?** highlight the need for method comparison methodologies suitable for use in the presence of replicate measurements. **?** presents a comprehensive methodology for assessing the agreement of two methods, for replicate measurements. This methodology has the added benefit of overcoming the problems of unbalanced data and unequal numbers of replicates. Implementation of the methodology, and interpretation of the results, is relatively easy for practitioners who have only basic statistical training. Furthermore, it can be shown that widely used existing methodologies, such as the limits of agreement, can be incorporated into Roy's methodology.

Chapter 5

Appendices 1

5.1 Model Terms (Roy 2009)

- Let y_{mir} be the response of method m on the ith subject at the $r-$th replicate.
- Let $\boldsymbol{y}_{ir}$ be the 2×1 vector of measurements corresponding to the $i-$th subject at the $r-$th replicate.
- Let $\boldsymbol{y}_{i}$ be the $R_i \times 1$ vector of measurements corresponding to the $i-$th subject, where R_i is number of replicate measurements taken on item i.
- Let $\alpha_m i$ be the fixed effect parameter for method for subject i.
- Formally Roy uses a separate fixed effect parameter to describe the true value μ_i, but later combines it with the other fixed effects when implementing the model.
- Let u_{1i} and u_{2i} be the random effects corresponding to methods for item i.
- $\boldsymbol{\epsilon}_i$ is a n_i-dimensional vector comprised of residual components. For the blood pressure data $n_i = 85$.
- $\boldsymbol{\beta}$ is the solutions of the means of the two methods. In the LME output, the bias ad corresponding t-value and p-values are presented. This is relevant to Roy's first test.

5.2 Algorithms : ML v REML

Maximum likelihood estimation is a method of obtaining estimates of unknown parameters by optimizing a likelihood function. The ML parameter estimates are the values of the argument that maximise the likelihood function, i.e. the estimates that make the observed values of the dependent variable most likely, given the distributional assumptions

The most common iterative algorithms used for the optimization problem in the context of LMEs are the EM algoritm, fisher scoring algorithm and NR algorithm, which [cite:West] commends as the preferred method.

A mixed model is an extension of the general linear models that can specify additional random effects terms.

Parameter of the mixed model can be estimated using either ML or REML, while the AIC and the BIC can be used as measures of "goodness of fit" for particular models, where smaller values are considered preferable.

(***Wikipedia***)The restricted (or residual, or reduced) maximum likelihood (REML) approach is a particular form of maximum likelihood estimation which does not base estimates on a maximum likelihood fit of all the information, but instead uses a likelihood function calculated from a transformed set of data, so that nuisance parameters have no effect.

In contrast to the earlier maximum likelihood estimation, REML can produce unbiased estimates of variance and covariance parameters.

ML procedures for LME

The maximum likelihood procedure of Hartley and Rao yields simultaneous estimates for both the fixed effects and the random effect, by maximising the likelihood of $\boldsymbol{y}$ with respect to each element of $\boldsymbol{\beta}$ and $\boldsymbol{b}$.

5.3 Estimation of random effects

Estimation of random effects for LME models in the NLME package is accomplished through use of both EM (Expectation-Maximization) algorithms and Newton-Raphson algorithms.

- EM iterations bring estimates of the parameters into the region of the optimum very quickly, but convergence to the optimum is slow when near the optimum.
- Newton-Raphson iterations are computationally intensive and can be unstable when far from the optimum. However, close to the optimum they converge quickly.
- The LME function implements a hybrid approach, using 25 EM iterations to quickly get near the optimum, then switching to Newton-Raphson iterations to quickly converge to the optimum.
- If convergence problems occur, the "controlargument in LME can be used to change the way the model arrives at the optimum.

5.4 Covariance Parameters

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

5.4.1 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.

? 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.

5.5 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.

5.6 Measures 2

5.6.1 Cook's Distance

- For variance components γ

Diagnostic tool for variance components

$$C_{\theta i} = ((\hat{\theta})_{[i]} - (\hat{\theta}))^T \mathrm{cov}((\hat{\theta}))^{-1}((\hat{\theta})_{[i]} - (\hat{\theta}))$$

5.6.2 Variance Ratio

- For fixed effect parameters β.

5.6.3 Cook-Weisberg statistic

- For fixed effect parameters β.

5.6.4 Andrews-Pregibon statistic

- For fixed effect parameters β.

The Andrews-Pregibon statistic AP_i is a measure of influence based on the volume of the confidence ellipsoid. The larger this statistic is for observation i, the stronger the influence that observation will have on the model fit.

5.7 Computation and Notation

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

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

Bibliography

Beckman, R., C. Nachtsheim, and R. Cook (1987). Diagnostics for mixed-model analysis of variance. *Technometrics 29*(4), 413–426.

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. (1977). Detection of influential observations in linear regression. *Technometrics 19*, 15–18.

Cook, R. (1986). Assessment of local influence. *Journal of the Royal Statistical Society. Series B (Methodological) 48*(2), 133–169.

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.

McCullough, C. and S. Searle (2001). *Generalized , Linear and Mixed Models*. Wiley Interscience.

Pinheiro, J. and D. Bates (1994). *Mixed Effects Models in S and S plus* (2nd ed.). Reading, Massachusetts: Springer.

Preisser, J. S. (1996). Deletion diagnostics for generalised estimating equations. *Biometrika 83*(3), 551–5562.

Roy, A. (2009). An application of linear mixed effects model to assess the agreement between two methods with replicated observations. *Journal of Biopharmaceutical Statistics 19*, 150–173.

Schabenberger, O. (2004). Mixed model influence diagnostics. 18929.

West, B., K. Welch, and A. Galecki (2007). *Linear Mixed Models: a Practical Guide Using Statistical Software*. Chapman and Hall CRC.