# Chapter 1

# Introduction to Influence Diagnostics

## 1.1 Introduction to Influence Diagnostics

The process of model validation is a vital part of the statistical modelling process, yet it is too often overlooked. Using a small handful of simple measures and methods, such as the AIC and $R^2$ measures, is insufficient to properly assess the usefulness of a fitted model. A full and comprehensive analysis that tests of all of the assumptions, as far as possible, should be carried out.

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.

Following model specification and estimation, it is of interest to explore the model-data agreement by raising pertinent questions. Pinheiro and Bates provide some insight into how to compute and interpret model diagnostic plots for LME models. Unfortunately this aspect of LME theory is not as expansive as the corresponding body of work for Linear Models. Their particular observations will be reverted to shortly. Further to the analysis of residuals, Schabenberger (2004) recommends the examination of the following questions:

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

The last of these three questions, regarding influential points, is of particular interest in the context of Method Comparison. After fitting an LME model, 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.

Influential points have a large influence on the fit of the model. By influential points, we mean one or more observations whose removal would cause a different conclusion in the analysis, e.g. substantially changes the estimate of the regression coefficients. West et al. (2007) remarks that influence diagnostics play an important role in the interpretation of results, because influential data can negatively influence the statistical fit and generalizability of the model. Schabenberger (2004) 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.

The process 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 can be thought of as consequence of leverage and outlierness. 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. They can point to a model breakdown and lead to development of a better model

While linear models and GLMS can be studied with a wide range of well-established diagnostic technqiues, the choice of methodology is much more restricted for the case of LMEs. However Influence diagnostics for LME Models is an area of active research. Research on diagnostic analyses of LME models are presented in Beckman et al. (1987), Christensen et al. (1992), Hilden-Minton (1995), Lesaffre and Verbeke (1998), Banerjee and Frees (1997), Fung et al. (2002), Demidenko (2004), Zewotir and Galpin (2005), Zewotir (2008) and Nobre and Singer (2007, 2011).

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.

Schabenberger (2004) states that 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 and Leverage

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. Christensen et al. (1992) advises that identification of outliers is necessary before conclusions may be drawn from the fitted model. The leverage of an observation is a further consideration.

Likelihood based estimation techniques, such as ML and REML, are sensitive to outliers. The question of whether or not a point should be considered an outlier must therefore be addressed. An outlier is an observation whose true value is unusual given its value on the predictor variables.

An observation with an extreme, but not unusual, value on a predictor variable is a point with high leverage. High leverage points can have a great amount of effect on the estimate of regression coefficients. In general. a high leverage point means a extreme value for the one or more of the independent variables, and a greater potential of overly influencing the final fitted model. However, if a case has extreme values for the independent variables but is fitted very well by a regression model, this case is not necessarily overly influential.

In classical linear models, leverages are the diagonal elements $h_{ii}$ of the Projection matrix, also known as the Hat Matrix $\boldsymbol{H}$. Schabenberger (2004) describes two analogues of the Hat matrix for LME models.

### 1.1.1 Cook's 1986 paper on Local Influence

Cook (1977) greatly expands the study of residuals and influence measures. 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 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.

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

Cook (1986) introduces methods for local influence assessment for classical linear models. 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.

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).

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

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

### 1.1.2 Comparison with Residual Analysis

Nieuwenhuis et al. (2012) compares residual analysis and influence analysis. Cases with high residuals (defined as the difference between the observed and the predicted scores on the dependent variable) or with high standardized residuals (defined as the residual divided by the standard deviation of the residuals) are indicated as outliers.

However, an influential case is not necessarily an outlying residual. On the contrary: a strongly influential case dominates the regression model in such a way, that the estimated regression line lies closely to this case. The analysis of residuals cannot be used for the detection of influential cases (Crawley, 2012).

### 1.1.3 Deletion Diagnostics(Christensen et al)

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. Christensen et al. (1992) notes the case deletion diagnostics techniques have not been applied to linear mixed effects models and seeks to develop methodologies in that respect. Christensen et al. (1992) develops these techniques in the context of REML.

Christensen et al. (1992) develops case deletion diagnostics, in particular the equivalent of Cook's distance, a well-known metric, for diagnosing influential observations when estimating the fixed effect parameters and variance components.

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.

### 1.1.4 Extension of Diagnostic Methods to LME models

When similar concepts of statistical influence are applied to mixed models, things are more complicated.

Demidenko (2004) extends several regression diagnostic techniques commonly used in linear regression, such as leverage, infinitesimal influence, case deletion diagnostics, Cook's distance, and local influence to the linear mixed-effects model. In each case, the proposed new measure has a direct interpretation in terms of the effects on a parameter of interest, and reduces to the familiar linear regression measure when there are no random effects.

The new measures that are proposed by Demidenko (2004) are explicitly defined functions and do not require re-estimation of the model, especially for cluster deletion diagnostics. The basis for both the cluster deletion diagnostics and Cook's distance is a generalization of Miller's simple update formula for case deletion for linear models. Furthermore Demidenko (2004) shows how Pregibon's infinitesimal case deletion diagnostics is adapted to the linear mixed-effects model. Schabenberger (2004) notes that removing observations or sets of observations affects fixed effects and covariance parameter estimates.

### 1.1.5 A Procedure for Quantifying Influence

Schabenberger (2004) 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* or `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. 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.

### 1.1.6 Influence Diagnostics: Closed Form Expressions

Schabenberger (2004) considers several important aspects of the use and implementation of influence measures in LME models, noting that it is not always possible to derive influence statistics necessary for comparing full- and reduced-data parameter estimates.

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.

The parameter vector $\theta$ denotes all unknown parameters in the $R$ and $D$ 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 .

### 1.1.7 Computation Matters

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

Schabenberger (2004) examines the use and implementation of influence measures in LME models.

Schabenberger (2004) notes that it is not always possible to derive influence statistics necessary for comparing full- and reduced-data parameter estimates.

On occasion, quantification is not possible. Assume, for example, that a data point is removed and the new estimate of the $D$ 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.

### 1.1.8 Iterative and Non-Iterative Influence Analysis

Schabenberger (2004) highlights some of the issue regarding implementing LME model diagnostics, describing the choice between iterative influence analysis and non-iterative influence analysis.

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

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

Calculation of Influence diagnostics can be performed using non-iterative or iterative methods.

For linear models, the implementation of influence analysis is straightforward, but for LME models the process is more complex.

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. However, this doesnt increase the procedures execution time by the same degree.

Schabenberger makes several comments about iterative procedures. Despite the addition execution time of these brute forces approaches, they are less reliant on assumptions.

Non-iterative methods are computationally efficient, but require the rather strong assumption that all covariance parameters are known, and thus are not updated, with the exception of the profiled residual variance.

Iterative Influence diagnostics requiring fitting the model without the observations in question. Computation execution time is substantially longer, although this is balanced by algorithmic simplicity, with no assumptions beyond those used for the original model. Additionally iterative approaches facilitate several complementary analyses to be carried out concurrently.

The execution times for iterative procedures are longer relative to iterative procedures, measured in seconds rather than microseconds.

### Computational Limitations for Cook's Distance

Application of Cooks Distances are limited by computation tractability.

Application of case-deletion diagnostics offer some interested for Method Comparison Studies

Care must be given when interpreting these plots. For example the position of case 68 on the BSVR indicates that that case 68

Any diagnostic plot may constructed using Overall variability and intermethod bias.

### 1.1.9 Analyzing Influence in 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.

For example, you are not only concerned with capturing the fixed and random components of the model. The LME model structure presents unique and interesting challenges that prompt us to reexamine the traditional ideas of influence and residual analysis.

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.

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

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 $\beta$ and $u$, given $\hat{V}$.

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.

### 1.1.10 Cook's Distance

Cooks Distance ($D_i$) is a well known diagnostic technique used in classical linear models, that functions an overall measure of the combined impact of the $i$th case of all estimated regression coefficients. Cook's Distance measures the influence of an observation that is a measure of aggregate impact of each observation on the group of regression coefficients, as well as the group of fitted values. Observations, or sets of observations, that have high Cook's distance usually have high residuals, although this is not necessarily the case.

If the predictions are the same with or without the observation in question, then the observation has no influence on the regression model. If the predictions differ greatly when the observation is not included in the analysis, then the observation is influential.

#### Interpretating Cook's Distance

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

Use of threshold values for Cook's Distance is discouraged (Fox, 1997). However, informal heuristics do exist for OLS models; Obervations for which Cook's distance is higher than 1 are usually considered as influential. Another informal threshold of $4/n$ or $4/(n-k-1)$, where $n$ is the number of observations and $k$ the number of explanatory variables. Fox (1997) advises the use of diagnostic plotting and to examine in closer details the points with "*values of D that are substantially larger than the rest*", and that thresholds should feature only to enhance graphical displays.

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 for LMEs**

Christensen et al. (1992) would later adapt the Cook's Distance measure for the analysis of 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 $\beta$ or $\theta$.

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

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

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

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.

If the predictions are the same with or without the observation in question, then the observation has no influence on the regression model. If the predictions differ greatly when the observation is not included in the analysis, then the observation is influential.

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

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 $\boldsymbol{X}$ (**?**).

**Taxonomy of Cook's Distances for LMEs**

Zewotir and Galpin (2005) discusses a taxonomy of Cook's distance when applied to LME models.

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

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

### 1.1.11 Assumptions for Update Formulas

While influence analysis concepts are statistically straightforward, implementation for LME models is more complex. 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. However, in LME 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.

Update formulas for "leave-U-out" estimates typically fail to account for changes in covariance parameters. Update formulae for fixed effects models are available only when the covariance parameters are assumed to be known. As the influence that each item would have on the variance estimate of a method comparison model is crucial, this negates their usefulness for Roy's Model.

Iterative methods retain the potential for usefulness, if applied at different stage of the modelling process. Diagnostic measures, specifically the DFBETA, have characteristics that would make them very useful at the exploratory stage of the method comparison process. Implicitly various assumptions about variance are used, but simultaneously an approach based on DFBETA can be used to assess if these assumptions are valid.

### 1.1.12 Deletion Diagnostics

Influence diagnostics are formal techniques that assess the influence of observations on parameter estimates for $\beta$ and $\theta$. A common technique is to refit the model with an observation or group of observations omitted.

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

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.

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.

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. 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.

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.

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

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.

The natural sampling unit is the item or subject, similar to the example provided by Schabenberger (2004). Schabenberger notation $U$ to denote quantities conmputed from data with subset of cases U omitted.

### 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 $\boldsymbol{A}$.

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.

### 1.1.13 (Zewotir) Methods and Measures

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 (Zewotir and Galpin, 2005).

Zewotir and Galpin (2005) remarks the development of efficient computational formulas is crucial making deletion diagnostics useable, allowing one to obtain the case deletion diagnostics by making use of basic building blocks, computed only once for the full model. Zewotir and Galpin (2005) describes a number of approaches to model diagnostics, including variance components, dixed effects parameters, prediction of the response variable and of random effects, and the likelihood function

Zewotir and Galpin (2005) lists several established methods of analyzing influence in LME models. These methods include Cook's distance for LME models, likelihood distance, the variance (information) ration, the Cook-Weisberg statistic, and the Andrews-Prebigon statistic.

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

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 $\beta$.

### 1.1.14 Overall Influence- Likelihood Distance

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, 1983).

The likelihood distance is a global summary measure that expresses the joint influence of the subsets of observations, $U$, on all parameters in $\phi$ that were subject to updating. For classical linear models, the implementation of influence analysis is straightforward.

Schabenberger (2004) points out the likelihood distance gives the amount by which the log-likelihood of the model fitted from the full data changes if one were to estimate the model from a reduced-data estimates.

.

Importantly $LD(\psi_{(U)})$ 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.

However, for LME models, the problem 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 or cases, then refitting the model. This is a very simplistic approach, and computationally expensive.

In LME models, fitted by either ML or REML, an important overall influence measure is the likelihood distance. The procedure requires the calculation of the full data estimates $\hat{\psi}$ and estimates based on the reduced data set $\hat{\psi}_{(U)}$. The likelihood distance is given by determining

$$LD_{(U)} = 2\{l(\hat{\psi}) - l(\hat{\psi}_{(U)})\} \quad (1.3)$$

$$RLD_{(U)} = 2\{l_R(\hat{\psi}) - l_R(\hat{\psi}_{(U)})\} \quad (1.4)$$

Large values indicate that $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$ differ considerably.

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*.

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

## 1.2 Using DFBETAs from LME Models to Assess Agreement

### DFBETA and DFFITS

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. DFBETAS is a standardized measure of the absolute difference between the estimate with a particular case included and the estimate without that particular case (Belsley et al., 2005).

The DFBETA is a measure that standardizes the absolute difference in parameter estimates between an LME model based on a full set of data, and a model from which a subset of data is removed, thus measuring the impact each observation has on a particular predictor

In general, large values of DFBETAS indicate observations that are influential in estimating a given parameter. Belsley et al. (2005) recommend 2 as a general cutoff value to indicate influential observations and as a size-adjusted cutoff. There is no agreement as to the critical threshold for DFBETAs. 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".

DFFITS is a statistical measured designed to a show how influential an observation is in a statistical model. DFFITS is a diagnostic meant to show how influential a point is in a statistical regression. It is defined as the change, in the predicted value for a point, obtained when that point is left out of the regression, divided by the estimated standard deviation of the fit at that point:

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

Emphasis shall be placed on DFBETA, but a discussion of DFFITS is merited as it potentially provides for useful techniques in method comparison.

### DFBETAs for Method Comparison

For LME Models, a value for DFBETAS is calculated for each of the $k$ fixed effects, and for each of the $n$ item. Correctly there will be $p+1$ DFBETAs (the intercept, $\beta_0$, and one $\beta$ for each covariate). When the LME model is specified without an intercept term, as in Roy's Model, there is a set of DFBETAs corresponding to each measurement method, hence an $n \times p$ matrix.

In the case of method comparison studies, a series of scatterplots can be constructed to compare each pair of measurement methods. Furthermore 95% confidence ellipse can be constructed around these scatterplots.

The LME approach proposed by Roy (2009) is constrained by computational tractability. Consequently a simpler LME formulation is used, one similar to that of Carstensen et al. (2008). However one constraint that can be dispensed with is the restriction to two methods of measurement: we can now use any number of methods. The benefit of using this model is that metrics such as Cook's Distance and DFBETAs can be computed also.

Furthermore, these measures form the basis of the analysis, rather than the estimates derived from the model. In the context of method comparison, these variables are the methods of measurement. Agreement will considered in the context of inter-method bias and the within-item variance ratio. Between-item variance ratio is not considered for this analysis.

For a Method Comparison study, DFBETAs can be used as a proxy measurement, allowing simple techniques to be used for assessing agreement. Suppose an LME model was formulated to model agreement for two or more methods of measurement, specifically 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 lack of agreement is caused, in part or in full, by differing within-item variances, there would be differing DFBETAs for each pair of methods. If the points align along the line of equality, then both methods can be said to be in agreement for within item variance. However DFBETAs are not useful for determining inter-method bias. If there is good agreement between methods, or if lack of agreement is caused by inter-method bias only, the DFBETA values will be almost identical for each subject in the data set.

Following the idea proposed by Bland and Altman (1986), an identity plot to visually inspect this relationship between sets of DFBETAs. Modern statistical software usually allows for the creation of co-plots, so a grid of identity plots may be easily rendered for comparing each pair of methods. Used in conjection with a Bland-Altman plot, this co-plot can quickly determine agreement and indicate the source of lack of agreement.

For an LME model fitted to the Blood data, the results tabulated below can be produced. Cases can be ranked by the Cook's Distance, such that the most divergent DFBETA are highlighted, with the top 6 being presented below). The remaining columns are the DFBeta for each of the fixed effects, for each of the 85 subject.

| Subject | Cook's D | Method J | Method R | Method S |
|---|---|---|---|---|
| 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 |

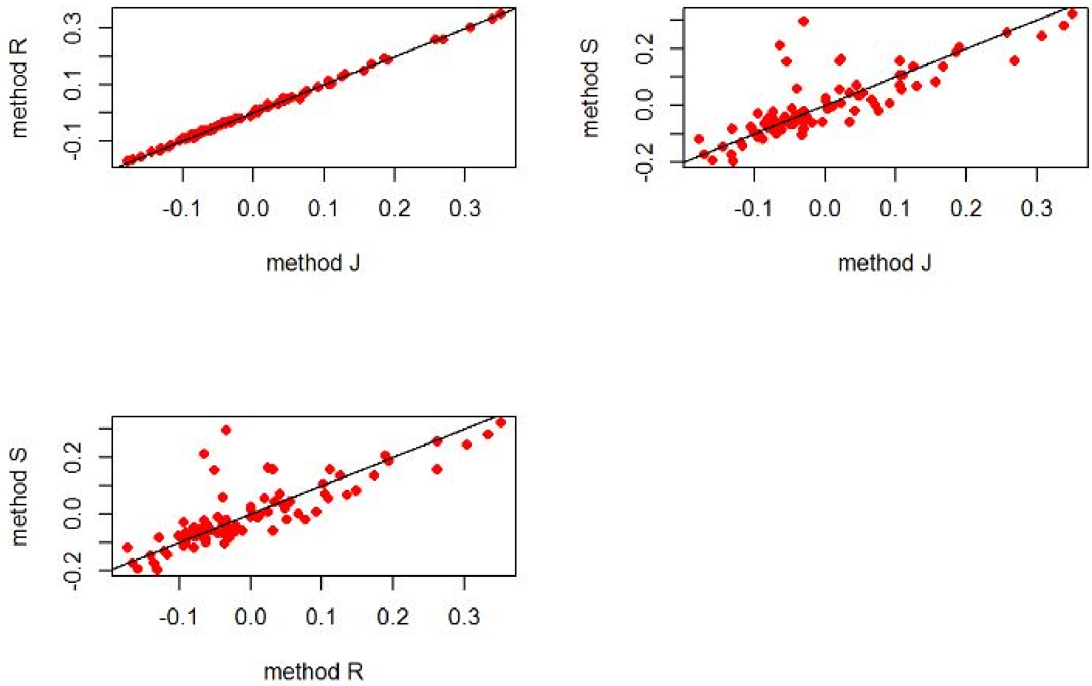
For DFBETA identity plots are presented below. This set of plots indicate agreement between methods J and R in terms of within-item variance, while severe lack of agreement exists between these methods and the third method S, as is the conclusion of Roy (2009).

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.

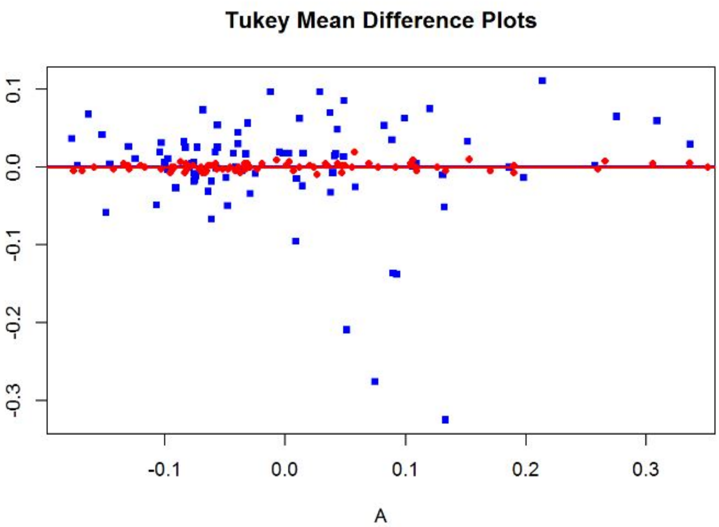
Other analyses may be used to complement these plots. The Pearson Correlation coefficient of the DFBETAs can be used in conjection with this analysis. A high correlation confirms good agreement, thouhg no threshold value for agreement is suggested.

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.

As as alternative to scatterplots, a mean difference plot could be used to assess agreement of with-item variance. 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. Here two of the three pairs of methods are compared on the same plot, red points indicate the J-R comparison while blue points are for the J-S comparison.



### 1.2.1 Model Diagnostics for Roy's Models

Further to previous work, this section revisits case-deletion and residual diagnostics, and explores how approaches devised by Gałecki and Burzykowski (2013) can be used to appraise Roy's model. These authors specifically look at Cook's Distances and Likelihood Distances.

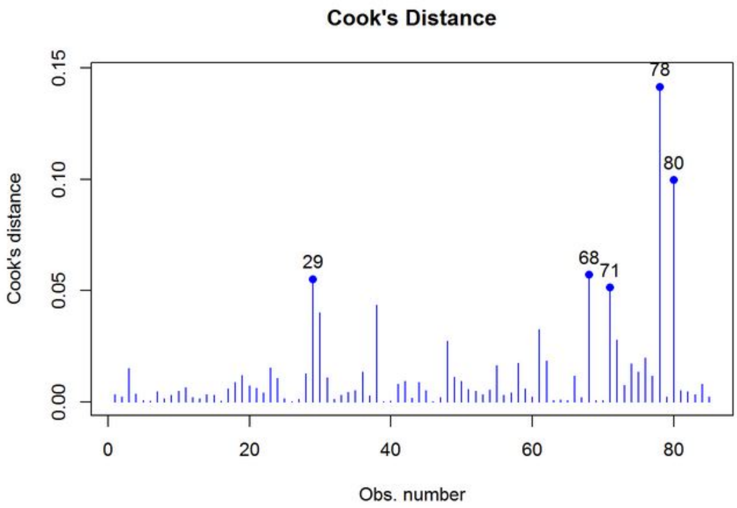



Figure 1.2.1:

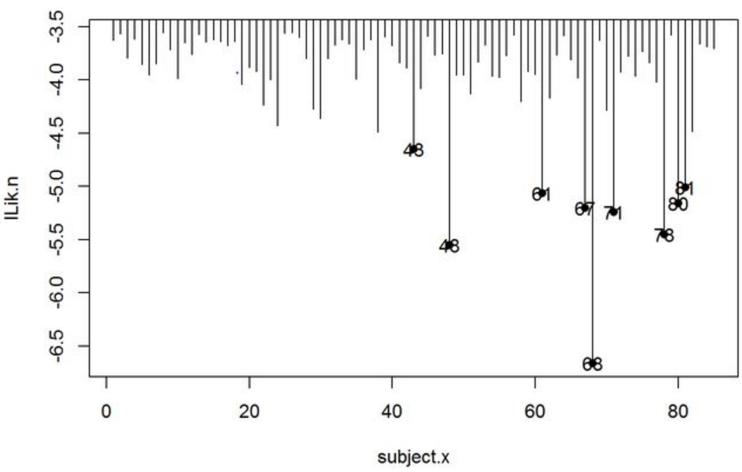



Figure 1.2.2:

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

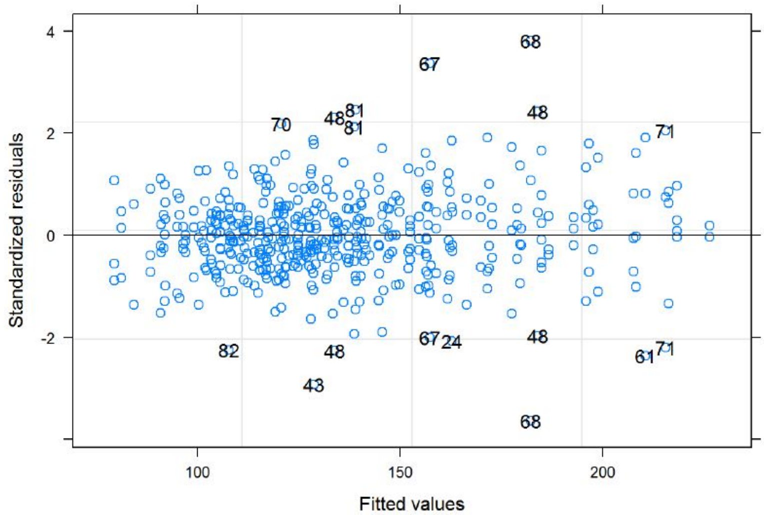



Figure 1.2.3:

### Case Deletion Diagnostics for Variance Ratios

Schabenberger (2004) 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.

### Methods for Identifying Outliers

The Grubbs' Test for Outliers is a commonly used technique for assessing outlier in a univariate data set, of which there are several variants. The first variant of Grubb's test is used to detect if the sample dataset contains one outlier, statistically different than the other values. The test statistic is based by calculating score of this outlier $G$ (outlier minus mean and divided by sd) and comparing it to appropriate critical values. Alternative method is calculating ratio of variances of two datasets - full dataset and dataset without outlier. The second variant is used to check if lowest and highest value are two outliers on opposite tails of sample. It is based on calculation of ratio of range to standard deviation of the sample.

The third variant calculates ratio of variance of full sample and sample without two extreme observations. It is used to detect if dataset contains two outliers on the same tail.

As there may be several outliers present, the Grubbs test is not practical. However an indication that a point being beyond the fences according to Tukey's specification for boxplots, ( i.e. greater than $Q_3 1.5\text{IQR}$ or less than $Q_1 - 1.5\text{IQR}$), will suffice.

### Mahalanbis Distance

Bivariate Analyses may be applied jointly to the both sets of data sets, e.g Mahalanobis distances.

The WSVR values are plotted against the corresponding BSVR values. Confidence Ellipses can be superimposed over the plot with minimal effort. Two ellipses are generated by this technique, a 50 % and 97.5% confidence ellipse respectively. Outlying cases are idenified by the plot. Subject 68 is evident

The subjects were ranked by Mahalanobis distance, with the top 10 being presented in the following table. Both sets of ratio are addtionally expressed as a ratio of the full model variance ratios.

| Subject (u) | MD | $WSVR_{(u)}$ | WSVR (%) | $BSVR_{(u)}$ | BSVR (%) |
|---|---|---|---|---|---|
| 68 | 44.7284 | 1.3615 | 0.9132 | 1.0353 | 0.9849 |
| 30 | 16.7228 | 1.5045 | 1.0092 | 1.1024 | 1.0487 |
| 71 | 11.5887 | 1.5210 | 1.0202 | 1.0932 | 1.0400 |
| 80 | 11.0326 | 1.4796 | 0.9925 | 1.0114 | 0.9621 |
| 38 | 10.3671 | 1.5011 | 1.0069 | 1.0917 | 1.0385 |
| 67 | 10.1940 | 1.4308 | 0.9598 | 1.0514 | 1.0002 |
| 43 | 7.6932 | 1.4385 | 0.9649 | 1.0511 | 0.9999 |
| 72 | 4.7350 | 1.4900 | 0.9995 | 1.0262 | 0.9762 |
| 48 | 4.4321 | 1.4950 | 1.0028 | 1.0280 | 0.9779 |
| 29 | 4.3005 | 1.4910 | 1.0001 | 1.0769 | 1.0244 |

From this table one may conclude that subjects 72, 48 and 29 are not particularly influential. Interestingly Subject 78, which was noticeable in the case deletion diagnostics for fixed effects, does not feature in this table.

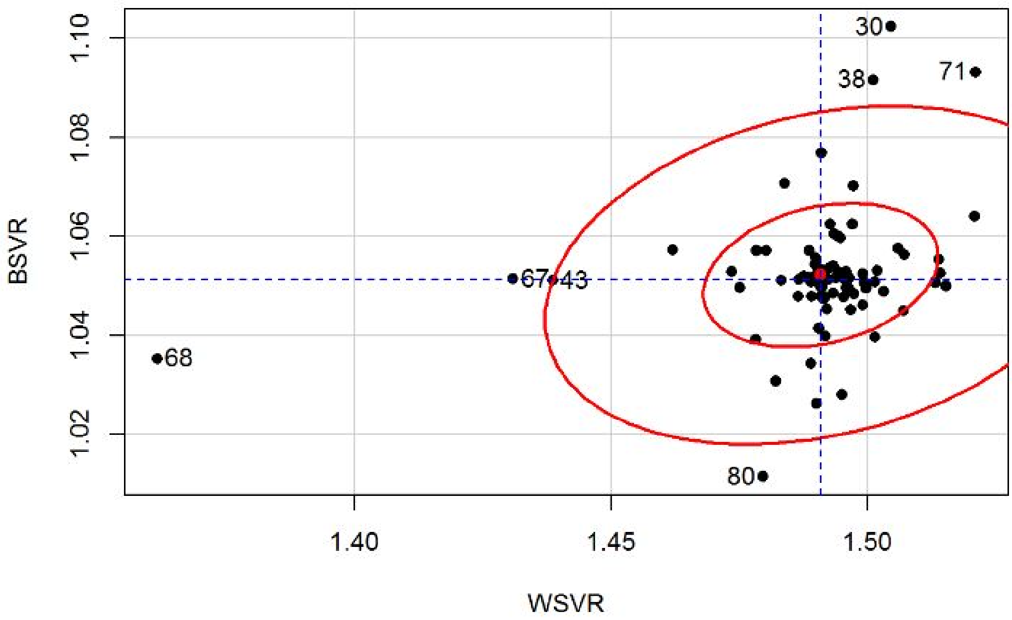



Figure 1.2.4:

### Variance Ratios

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

A useful approach is to compute the confidence intervals for the ratio of within-item standard deviations (equivalent to the ratio of repeatability coefficients), which can be interpreted in the usual manner.

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

What is required is the computation of the variance ratios of within-item and between-item standard deviations.

A nave approach would be to compute the variance ratios by relevant F distribution quantiles. However, the question arises as to the appropriate degrees of freedom. Limits of agreement are easily computable using the LME framework. While we will not be considering this analysis, a demonstration will be provided in the example.

# Bibliography

Banerjee, M. and E. W. Frees (1997). Influence diagnostics for linear longitudinal models. *Journal of the American Statistical Association 92*(439), 999–1005.

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

Belsley, D. A., E. Kuh, and R. E. Welsch (2005). *Regression diagnostics: Identifying influential data and sources of collinearity*, Volume 571. John Wiley & Sons.

Bland, J. and D. Altman (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet i*, 307–310.

Carstensen, B., J. Simpson, and L. C. Gurrin (2008). Statistical models for assessing agreement in method comparison studies with replicate measurements. *The International Journal of Biostatistics 4*(1).

Christensen, R., L. M. Pearson, and W. Johnson (1992). Case-deletion diagnostics for mixed models. *Technometrics 34*(1), 38–45.

Cook, R. (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.

Cook, R. D. and S. Weisberg (1983). Diagnostics for heteroscedasticity in regression. *Biometrika 70*(1), 1–10.

Crawley, M. J. (2012). *The R book.* John Wiley & Sons.

Demidenko, E. (2004). *Mixed Models: Theory And Application.* Dartmouth College: Wiley Interscience.

Fox, J. (1997). *Applied regression analysis, linear models, and related methods.* Sage Publications, Inc.

Fung, W.-K., Z.-Y. Zhu, B.-C. Wei, and X. He (2002). Influence diagnostics and outlier tests for semiparametric mixed models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64*(3), 565–579.

Gałecki, A. and T. Burzykowski (2013). *Linear mixed-effects models using R: A step-by-step approach.* Springer Science & Business Media.

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.

Hilden-Minton, J. A. (1995). *Multilevel diagnostics for mixed and hierarchical linear models.* Ph. D. thesis, University of California Los Angeles.

Lesaffre, E. and G. Verbeke (1998). Local influence in linear mixed models. *Biometrics*, 570–582.

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

Nieuwenhuis, R., H. te Grotenhuis, and B. Pelzer (2012). Influence. me: tools for detecting influential data in mixed effects models.

Nobre, J. S. and J. M. Singer (2007). Residual analysis for linear mixed models. *Biometrical Journal 49*(6), 863–875.

Nobre, J. S. and J. M. Singer (2011). Leverage analysis for linear mixed models. *Journal of Applied Statistics 38*(5), 1063–1072.

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

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

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.

Zewotir, T. (2008). Multiple cases deletion diagnostics for linear mixed models. *Communications in StatisticsTheory and Methods 37*(7), 1071–1084.

Zewotir, T. and J. S. Galpin (2005). Influence diagnostics for linear mixed models. *Journal of Data Science 3*(2), 153–177.