Contents

0.1 Influence Diagnostics

0.1.1 Introduction

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

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

What is Influence

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, 2004).

Influence can be thought of as the product of leverage and outlierness. An observation is said to be influential if removing the observation substantially changes the estimate of the regression coefficients. The `R` programming language has a variety of methods used to study each of the aspects for a linear 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.

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. 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.

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 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.

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.

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 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. 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 (Schabenberger, 2004).

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. 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.

Outliers and Leverage

The question of whether or not a point should be considered an outlier must also be addressed. An outlier is an observation whose true value is unusual given its value on the predictor variables. The leverage of an observation is a further consideration. Leverage describes an observation with an extreme 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.

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 $\boldsymbol{H}$.

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

Cook's 1986 paper on Local Influence

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 classical OLS 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. 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.

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.

Diagnostic Methods for OLS models

Cook (1977) greatly expanded 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.

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).

$$\hat{e}_{i(U)} = y_i - x\hat{\beta}_{(U)} \tag{1}$$

$$\delta_{(U)} = \hat{\beta} - \hat{\beta}_{(U)} \tag{2}$$

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. Christensen et al. (1992) 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 (Zewotir and Galpin, 2005).

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.

Computation Matters

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.

Key to the implementations of influence diagnostics for LME Models 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.

Influence Diagnostics: Closed Form Expressions

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 θ denotes all unknown parameters in the R and G 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 .

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.

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

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.

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. 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.

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

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.

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 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.

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.

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

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 $\hat{V}$.

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.

Methods and Measures

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

Zewotir and Galpin (2005) 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

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.

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

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

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].

Extension of Diagnostic Methods to LME models

When similar notions of statistical influence are applied to mixed models, things are more complicated. Removing data points affects fixed effects and covariance parameter estimates. Update formulas for *leave-one-out* estimates typically fail to account for changes in covariance parameters.

? noted 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. ======= **?** noted the case deletion diagnostics techniques had not been applied to linear mixed effects models and seeks to develop methodologies in that respect. **?** develops these techniques in the context of REML.

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.

Demidenko (2004) proposes two likelihood-based diagnostics for identifying individuals that can influence the choice between two competing models.

Iterative Influence Analysis

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

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.

The local-influence approach to influence assessment is quite different 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.

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.

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

$$\begin{aligned} \hat{\beta} &= X^T & (3) \\ \hat{\gamma} &= G(\hat{\theta})Z^T & (4) \end{aligned}$$

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.

Non-iterative Update Procedures

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

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

Residuals

Studentized residuals, error contrast matrices and the inverse of the response variance covariance matrix are regular components of these tools.

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]

0.2 Deletion Diagnostics

0.2.1 Christensen et al

Christensen et al. (1992) studied case deletion diagnostics, in particular the analog of Cooks distance, 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.

Christensen et al. (1992) 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.

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) provides an overview of case deletion diagnostics for fixed effect 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

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

$$X'_{[i]}V_{[i]}^{-1}X_{[i]} = X'V^{-1}X - \frac{\hat{x}_i\hat{x}'_i}{s_i} \quad (5)$$

$$(X'_{[i]}V_{[i]}^{-1}X_{[i]})^{-1} = (X'V^{-1}X)^{-1}\frac{(X'V^{-1}X)^{-1}\hat{x}_i\hat{x}'_i(X'V^{-1}X)^{-1}}{s_i - \bar{h}_i} \quad (6)$$

$$X'_{[i]}V_{[i]}^{-1}Y_{[i]} = X\prime V^{-1}Y - \frac{\hat{x}_i\hat{y}'_i}{s_i} \quad (7)$$

0.2.2 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.

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. **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 $\boldsymbol{A}$.

0.2.3 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.

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.

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 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.

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.

0.2.4 Cook's Distance

Cook's Distance is a good measure of the influence of an observation that is a measure of aggregate impact of each observation on the group of regression coefficients, as well as the group of fitted values. If the predictions are the same with or without the observation in question, then the observation has no influence on the regression model. If the predictions differ greatly when the observation is not included in the analysis, then the observation is influential.

Cook's Distance

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.

Reduced Data Set

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

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

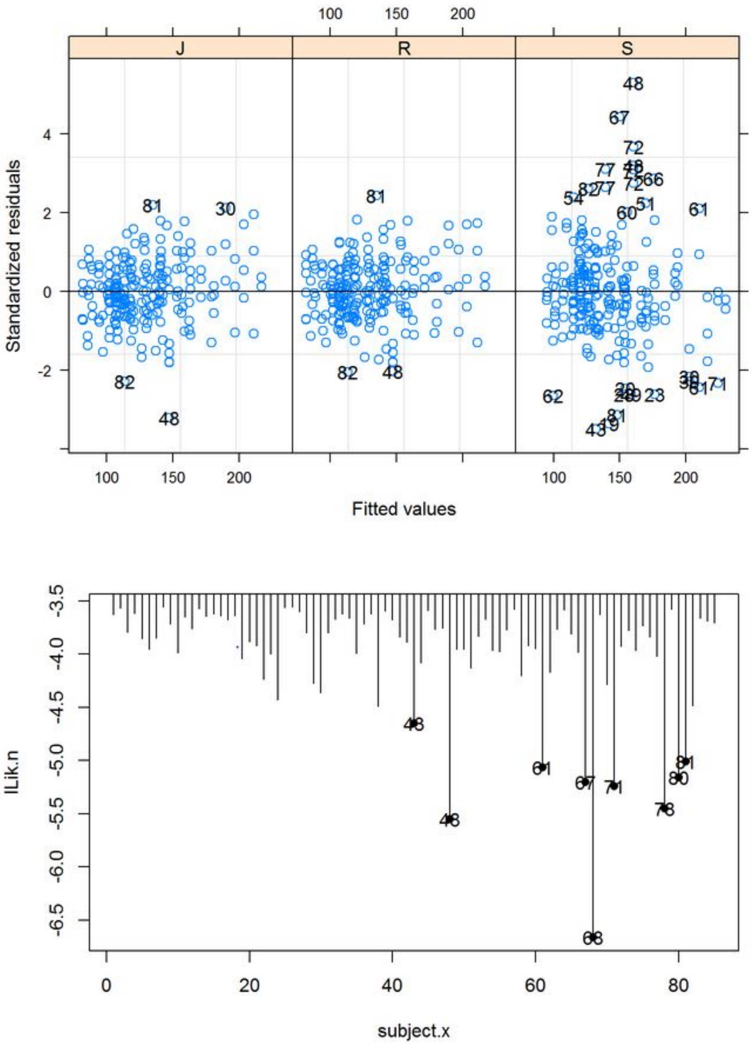


Figure 0.2.1:

Cook's Distance for Blood Data

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

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

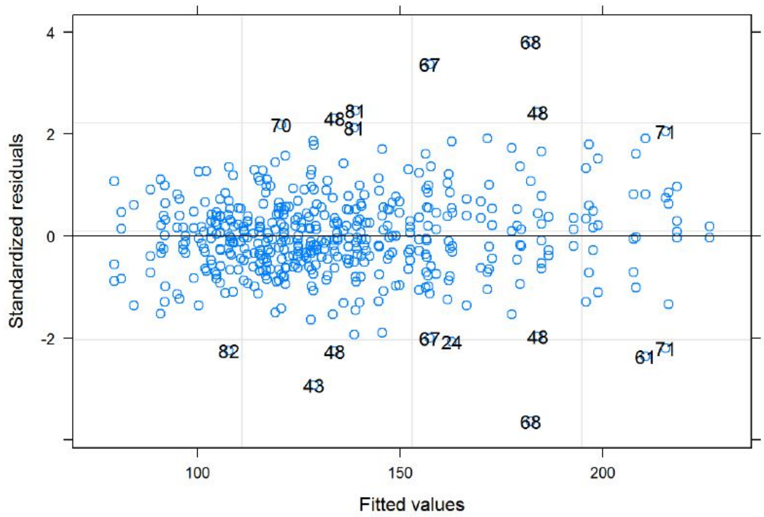


Figure 0.2.2:

Cook's Distance is a good measure of the influence of an observation that is a measure of aggregate impact of each observation on the group of regression coefficients, as well as the group of fitted values. The particular cases that we will omit for the subsequent analysis are subjects 68,78 and 80.

Cook's Distance for LMEs

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

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

Use of threshold value 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 to be considered as influential. Alternatively there is an informal threshold of $4/N$ or $4/(Nk1)$, 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 just be used 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.

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 β or θ.

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

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

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

Taxonomy of Cook's Distances for LMEs

Schabenberger (2004) discusses a taxonomy of Cook's distance when applied to LME models.

- 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 $\hat{\beta}$.

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 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.

A large value for $CD(u)_i$ indicates that the $i-$th observation is influential in predicting random effects. For linear functions, $CD(\psi)_i$ does not have to be calculated unless $CD(\beta)_i$ is large.

Cook's Distance

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.

linear functions

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

0.2.5 Computational Limitations for Cook's Distance

Application of Cooks Distances are limited by compuation 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

A similar plot may be constructed using the

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

0.2.6 Case Deletion Diagnostics for LME models

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

$$V_{[i]}^{-1} = \Lambda_{[i]} - \frac{\lambda\lambda\prime}{\nu ii} \tag{8}$$

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

$$X'_{[i]}V^{-1}_{[i]}X_{[i]} = X'V^{-1}X - \frac{\hat{x}_i\hat{x}'_i}{s_i} \tag{9}$$

$$(X'_{[i]}V^{-1}_{[i]}X_{[i]})^{-1} = (X'V^{-1}X)^{-1}\frac{(X'V^{-1}X)^{-1}\hat{x}_i\hat{x}'_i(X'V^{-1}X)^{-1}}{s_i - \bar{h}_i} \tag{10}$$

$$X'_{[i]}V^{-1}_{[i]}Y_{[i]} = X\prime V^{-1}Y - \frac{\hat{x}_i\hat{y}'_i}{s_i} \tag{11}$$

In LME models, fitted by either ML or REML, an important overall influence measure is the likelihood distance (Cook, 1986). 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

$$LD_{(U)} = 2\{l(\hat{\psi}) - l(\hat{\psi}_{(U)})\} \tag{12}$$

$$RLD_{(U)} = 2\{l_R(\hat{\psi}) - l_R(\hat{\psi}_{(U)})\} \tag{13}$$

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.

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.

$$\breve{a}_i = \boldsymbol{a}_i - \boldsymbol{A}_{(i)}\boldsymbol{H}_{[i]}\boldsymbol{h}_i \tag{14}$$

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 $CD(\gamma)$.

$$CD_U(\gamma) = (\hat{\gamma}_{(U)} - \hat{\gamma})'[\mathrm{var}(\hat{\gamma})]^{-1}(\hat{\gamma}_{(U)} - \hat{\gamma})$$
$$= -\boldsymbol{g'}_{(U)}(\boldsymbol{Q} - \boldsymbol{G})^{-1}\boldsymbol{Q}(\boldsymbol{Q} - \boldsymbol{G})\boldsymbol{g}_{(U)}$$
$$= \boldsymbol{g'}_{(U)}(\boldsymbol{I}_r \mathrm{var}(\hat{\gamma})\boldsymbol{G})^{-2}\mathrm{var}(\hat{\gamma})\boldsymbol{g}_{(U)}$$

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}.

0.2.7 Likelihood Distances

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

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.

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.

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.

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 (?).

$$LD((\boldsymbol{U})) = 2[l(\hat{\boldsymbol{\phi}}) - l\hat{\boldsymbol{\phi}}_{\boldsymbol{\omega}}]$$

$$RLD((\boldsymbol{U})) = 2[l_R(\hat{\boldsymbol{\phi}}) - l_R(\hat{\boldsymbol{\phi}})_{\boldsymbol{\omega}}]$$

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.*

Likelihood Distances

In LME models, fitted by either ML or REML, an important overall influence measure is the likelihood distance (Cook, 1986). 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

$$\begin{aligned} LD_{(U)} &= 2\{l(\hat{\psi}) - l(\hat{\psi}_{(U)})\} && (15) \\ RLD_{(U)} &= 2\{l_R(\hat{\psi}) - l_R(\hat{\psi}_{(U)})\} && (16) \end{aligned}$$

Likelihood Distances

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'.

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

$$
\begin{aligned}
LD_{(U)} &= 2\{l(\hat{\psi}) - l(\hat{\psi}_{(U)})\} && (17)\\
RLD_{(U)} &= 2\{l_R(\hat{\psi}) - l_R(\hat{\psi}_{(U)})\} && (18)
\end{aligned}
$$

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

0.2.8 Deletion Diagnostics

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

Christensen et al

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

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. itetChristensen 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.

Christensen et al. (1992) provides an overview of case deletion diagnostics for fixed effect 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

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

$$X'_{[i]}V_{[i]}^{-1}X_{[i]} = X'V^{-1}X - \frac{\hat{x}_i\hat{x}'_i}{s_i} \tag{19}$$

$$(X'_{[i]}V_{[i]}^{-1}X_{[i]})^{-1} = (X'V^{-1}X)^{-1} + \frac{(X'V^{-1}X)^{-1}\hat{x}_i\hat{x}'_i(X'V^{-1}X)^{-1}}{s_i - \bar{h}_i} \tag{20}$$

$$X'_{[i]}V_{[i]}^{-1}Y_{[i]} = X\prime V^{-1}Y - \frac{\hat{x}_i\hat{y}'_i}{s_i} \tag{21}$$

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.

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. **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 $\boldsymbol{A}$.

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.

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 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.

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.

Cook's Distance

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. Deletion diagnostics provide a means of assessing the influence of an observation (or groups of observations) on inference on the estimated parameters of LME models. Cook's Distance is a good measure of the influence of an observation that is a measure of aggregate impact of each observation on the group of regression coefficients, as well as the group of fitted values. If the predictions are the same with or without the observation in question, then the observation has no influence on the regression model. If the predictions differ greatly when the observation is not included in the analysis, then the observation is influential.

Cook's Distance

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).

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

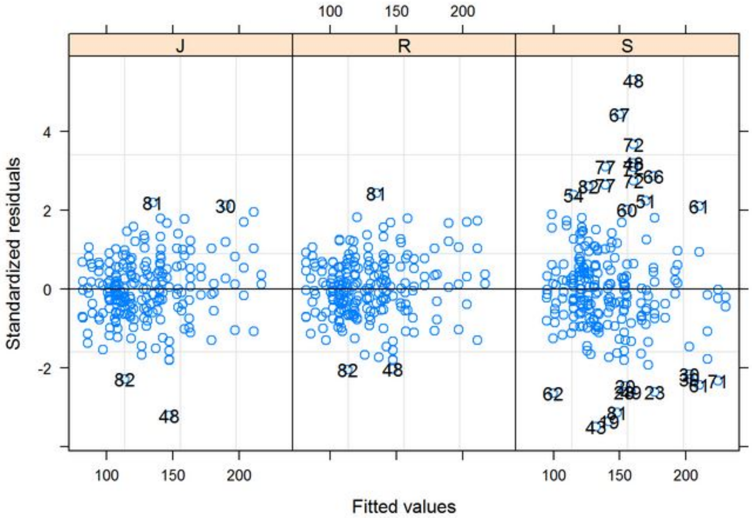
If V is known, Cook's D can be calibrated according to a chi-square distribution with degrees of freedom equal to the rank of $\boldsymbol{X}$ (**?**).

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

Reduced Data Set

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

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



Cook's Distance for Blood Data

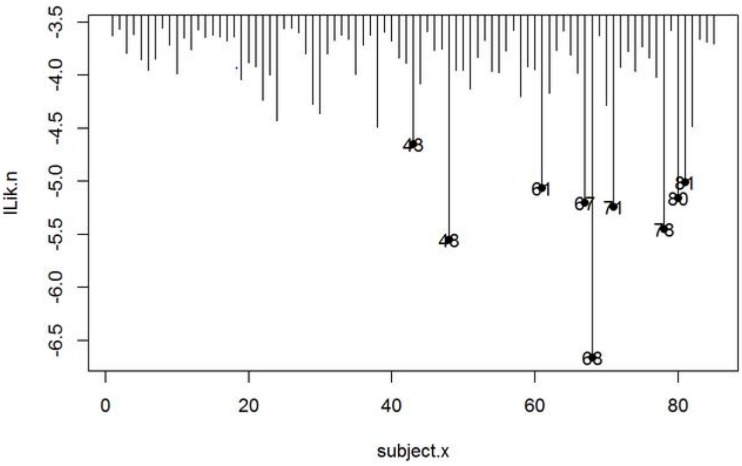


Figure 0.2.3:

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

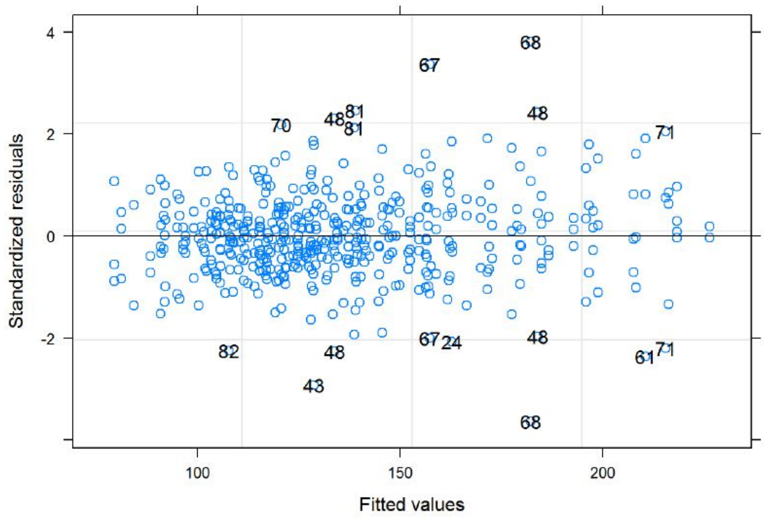


Figure 0.2.4:

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

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

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

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

Cook's Distance for LMEs

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

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

Use of threshold value for Cook's Distance is discouraged (**cite: JohnFox**). However, informal heuristics do exist for OLS models; Obervations for which Cook's distance is higher than 1 are to be considered as influential. Alternatively there is an informal threshold of $4/N$ or $4/(Nk1)$, where N is the number of observations and k the number of explanatory variables.

cite: JohnFox 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 just be used 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.

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 β or θ.

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

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

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

Taxonomy of Cook's Distances for LMEs

Schabenberger (2004) discusses a taxonomy of Cook's distance when applied to LME models.

- 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 $\hat{\beta}$.

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 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.

A large value for $CD(u)_i$ indicates that the $i-$th observation is influential in predicting random effects. For linear functions, $CD(\psi)_i$ does not have to be calculated unless $CD(\beta)_i$ is large.

Cook's Distance

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.

linear functions

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

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.

$$V_{[i]}^{-1} = \Lambda_{[i]} - \frac{\lambda\lambda\prime}{\nu ii} \tag{22}$$

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

$$\begin{aligned} X'_{[i]}V_{[i]}^{-1}X_{[i]} &= X'V^{-1}X - \frac{\hat{x}_i\hat{x}'_i}{s_i} && (23) \\ (X'_{[i]}V_{[i]}^{-1}X_{[i]})^{-1} &= (X'V^{-1}X)^{-1} + \frac{(X'V^{-1}X)^{-1}\hat{x}_i\hat{x}'_i(X'V^{-1}X)^{-1}}{s_i - \bar{h}_i} && (24) \\ X'_{[i]}V_{[i]}^{-1}Y_{[i]} &= X\prime V^{-1}Y - \frac{\hat{x}_i\hat{y}'_i}{s_i} && (25) \end{aligned}$$

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.

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.

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.

? 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

$$X = \begin{bmatrix} x_i' \\ X(i) \end{bmatrix}, Z = \begin{bmatrix} z_{ij}' \\ Z_{j(i)} \end{bmatrix}, Z = \begin{bmatrix} z_{ij}' \\ Z_{j(i)} \end{bmatrix},$$
$$y = \begin{bmatrix} y_{ij}' \\ y_{j(i)} \end{bmatrix} \text{ and } H = \begin{bmatrix} h_{ii} & h \\ h_{j(i)} & h \end{bmatrix}$$

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.

$$\breve{a}_i = \boldsymbol{a}_i - \boldsymbol{A}_{(i)} \boldsymbol{H}_{[i]} \boldsymbol{h}_i \tag{26}$$

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 $CD(\gamma)$.

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

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 $\mathrm{var}(\hat{\gamma})$ 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}.

Computational Limitations for Cook's Distance

Application of Cooks Distances are limited by compuation 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

A similar plot may be constructed using the

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

0.3 Using DFBETAs from LME Models to Assess Agreement

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

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

DFBETA for MCS

The LME approach proposed by Roy (2009) is constrained by computational tractability. Consequently a simpler LME formulation is required, one similar to that of Carstensen et al. (2008). However one constraint that can be dispensed with is that restriction to two methods of measurement: we can use any number of methods. The benefit of using this model is that diagnostics measures 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.

Recalling the definition of DFBETAs (in the context of LME models) (DEFINITION)

In the context of method comparison, these variables are the methods of measurement. If the methods are in agreement, the DFBETA values will be almost identical for each subject in the data set.

Here cases are ranked by the Cook's Distance, such that the most divergent DFBETA are highlighted.

Following the idea proposed by Bland and Altman (1986), an identity to visually inspect this relationship. Modern statistical software usually allows for the creation of co-plots, and as such as a grid of identity plots may be created for each pair of methods in the data set.

DFFITS

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 ("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:

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

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

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

Mean Square Prediction Error

$$MSPR = \frac{\sum(y_i - \hat{y}_i)^2}{n^*} \tag{27}$$

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.

$$\hat{e}_{i(U)} = y_i - x\hat{\beta}_{(U)} \tag{28}$$

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.

$$PRESS = \sum (y - y^{(k)})^2 \tag{29}$$

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

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

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

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

When fitting linear models, PRESS can be used as a criterion for model selection, with smaller values indicating better model fits.

$$PRESS = \sum (y - y^{(k)})^2 \tag{30}$$

PRESS Residuals and PRESS Statistic

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

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

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

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

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

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

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

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

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.

0.3.1 DFBETA, DFFITS and PRESS

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.

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.

DFBETAs

The measure that measures how much impact each observation has on a particular predictor is 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.

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.

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.

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. A value for DFBETAS is calculated for each covariate, and for each case, in the model separately.

$$DFBETA_a = \hat{\beta} - \hat{\beta}_{(a)} \tag{31}$$

$$= B(Y - Y_{\bar{a}} \tag{32}$$

In the case of method comparison studies, there are two covariates, and one can construct scatterplots 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 $k1$ dfbetas (the intercept,β_0, and one β for each covariate). When the model is specified without an intercept term, as in the last chapter, there is a set of DFBETAs corresponding to each measurement method.

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".

DFBETA, DFFITS and PRESS

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.

DFBETAs

The measure that measures how much impact each observation has on a particular predictor is 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.

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.

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.

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. A value for DFBETAS is calculated for each covariate, and for each case, in the model separately.

$$
\begin{aligned}
DFBETA_a &= \hat{\beta} - \hat{\beta}_{(a)} && (33) \\
&= B(Y - Y_{\bar{a}} && (34)
\end{aligned}
$$

In the case of method comparison studies, there are two covariates, and one can construct scatterplots 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 one β for each covariate). When the model is specified without an intercept term, as in the last chapter, there is a set of DFBETAs corresponding to each measurement method.

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".

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.

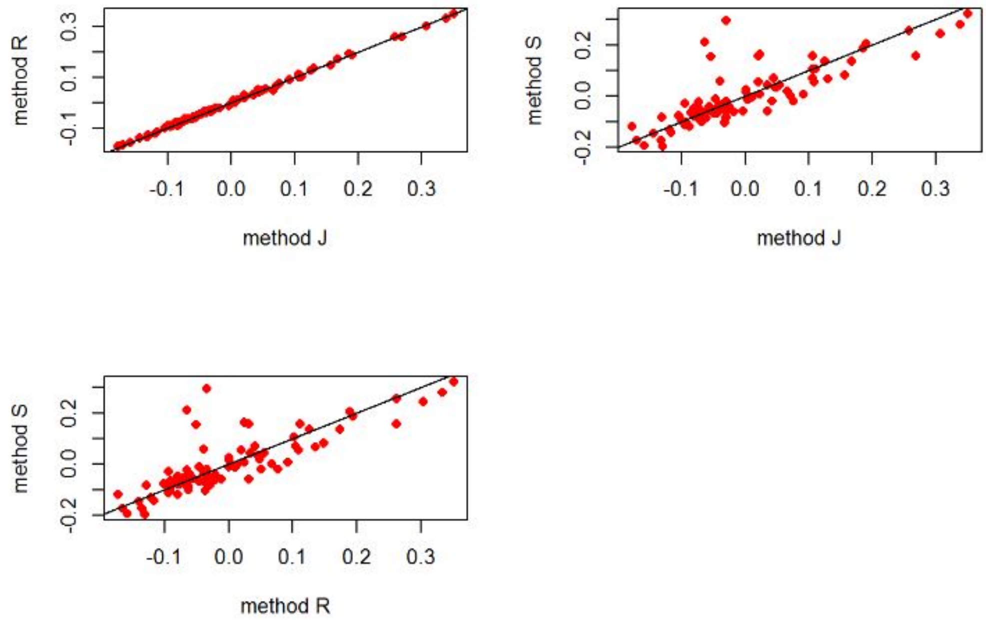
Application of DFBETAs in MCS Analysis

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

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).

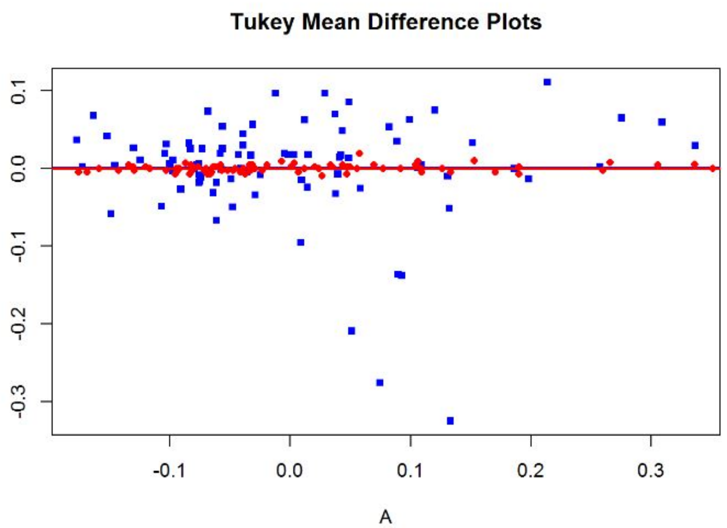
The Pearson Correlation coefficient of the DFBetas can be used in conjection with this analysis. A high correlation confirms good agreement. No threshold value for agreement is suggested, and analysts are advised to perform model diagnostics regardless of the correlation coeffient.

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).



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

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

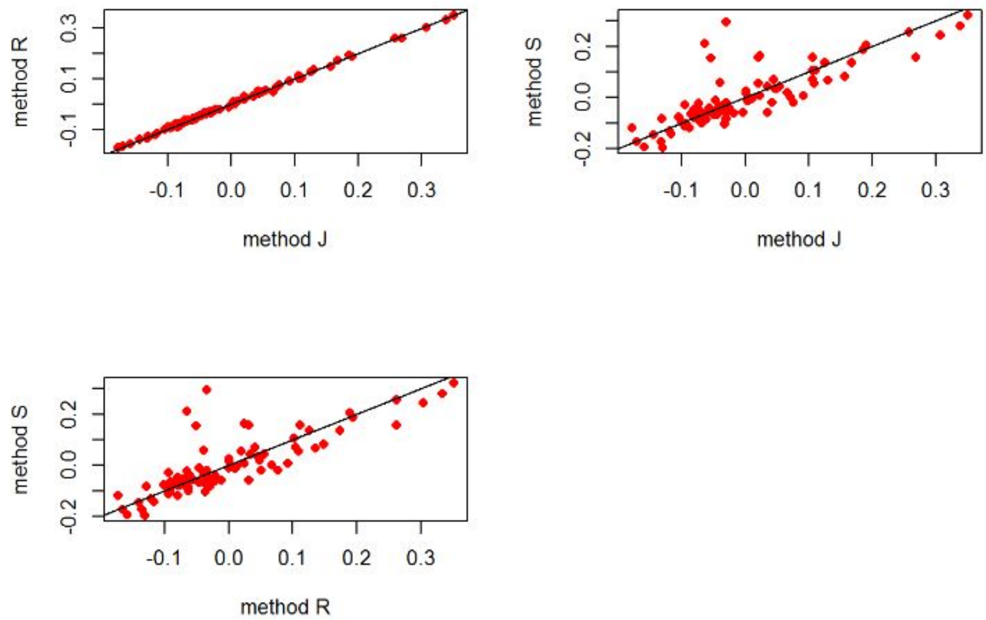
0.3.2 Application of DFBETAs in MCS Analysis

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

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.

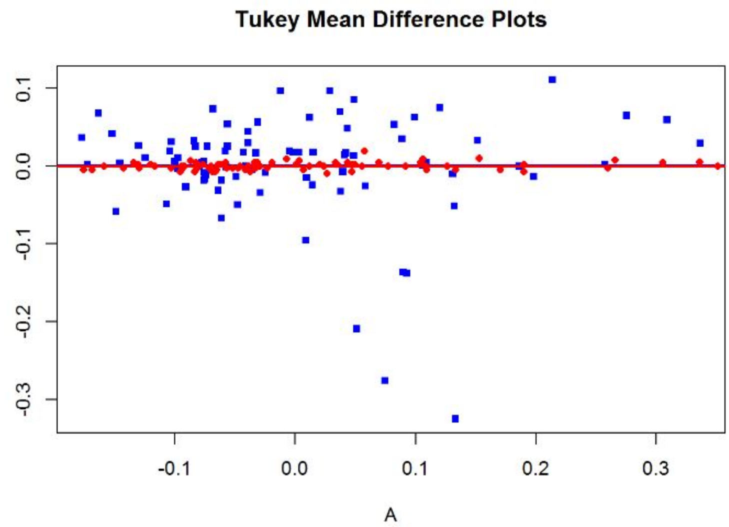
Deming Regression can be used to verify the line of equality. Significance test for Deming regression estimates are not available, but 95% bootstrap confidence intervals for the slope estimate and intercept estimates can be computed.

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).

0.3.3 DFBETA for MCS

The LME approach proposed by Roy (2009) is constrained by computational tractability. Consequently a simpler LME formulation is required, one similar to that of Carstensen et al. (2008). However one constraint that can be dispensed with is that restriction to two methods of measurement: we can use any number of methods. The benefit of using this model is that diagnostics measures 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.



Recalling the definition of DFBETAs (in the context of LME models) (DEFINITION)

In the context of method comparison, these variables are the methods of measurement. If the methods are in agreement, the DFBETA values will be almost identical for each subject in the data set.

Here cases are ranked by the Cook's Distance, such that the most divergent DFBETA are highlighted.

Following the idea proposed by Bland and Altman (1986), an identity to visually inspect this relationship. Modern statistical software usually allows for the creation of co-plots, and as such as a grid of identity plots may be created for each pair of methods in the data set.

DFFITS

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 ("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:

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

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

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

Mean Square Prediction Error

$$MSPR = \frac{\sum(y_i - \hat{y}_i)^2}{n^*} \tag{35}$$

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.

$$\hat{e}_{i(U)} = y_i - x\hat{\beta}_{(U)} \tag{36}$$

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.

$$PRESS = \sum(y - y^{(k)})^2 \tag{37}$$

$$e_{-Q} = y_Q - x_Q\hat{\beta}^{-Q}$$

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

$$PRESS_{(U)} = y_i - x\hat{\beta}_{(U)}$$

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

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

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

PRESS Residuals and PRESS Statistic

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

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

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

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

An (unconditional) predicted value is $\hat{y}_i = x_i'\hat{\boldsymbol{\beta}}$, where the vector x_i is the ith row of $\boldsymbol{X}$. The (raw) residual is given as $\varepsilon_i = y_i - \hat{y}_i$. The PRESS residual is similarly constructed, using the predicted value for observation i with a model fitted from reduced data.

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

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.

0.3.4 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***

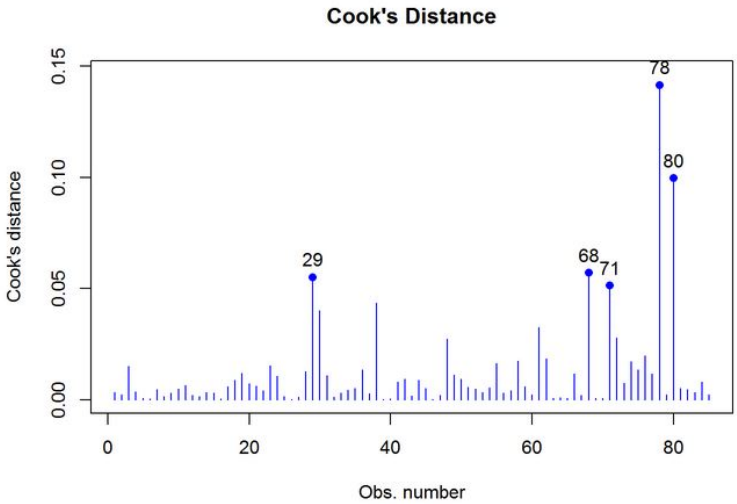


Figure 0.3.5:

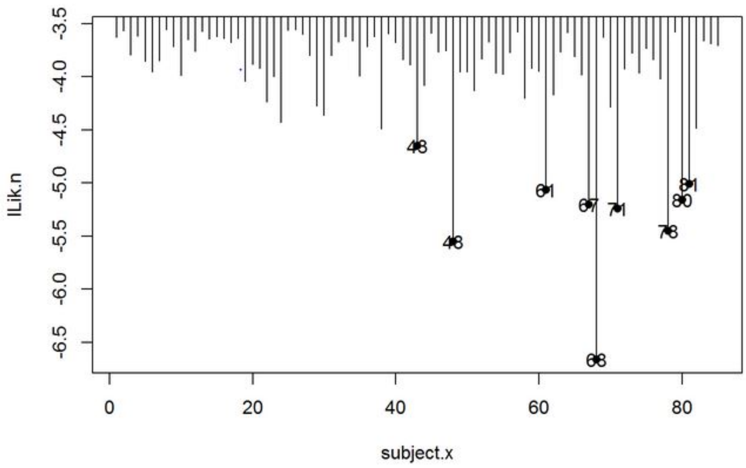


Figure 0.3.6:

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

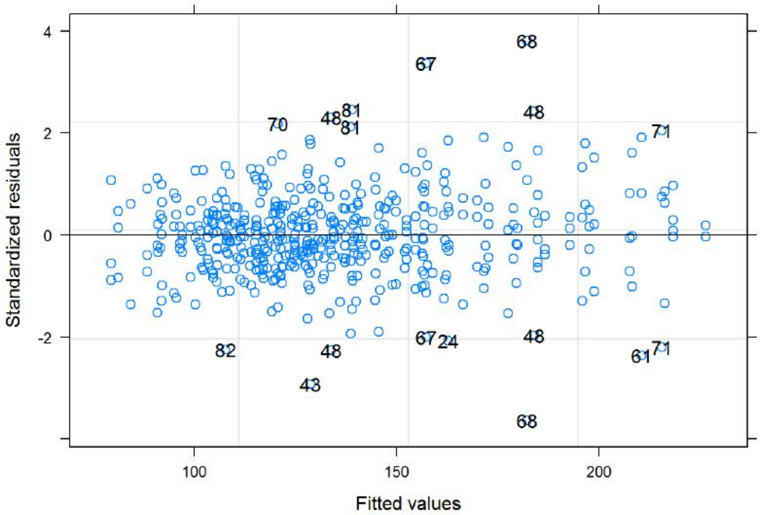


Figure 0.3.7:

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***

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

0.3.5 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.

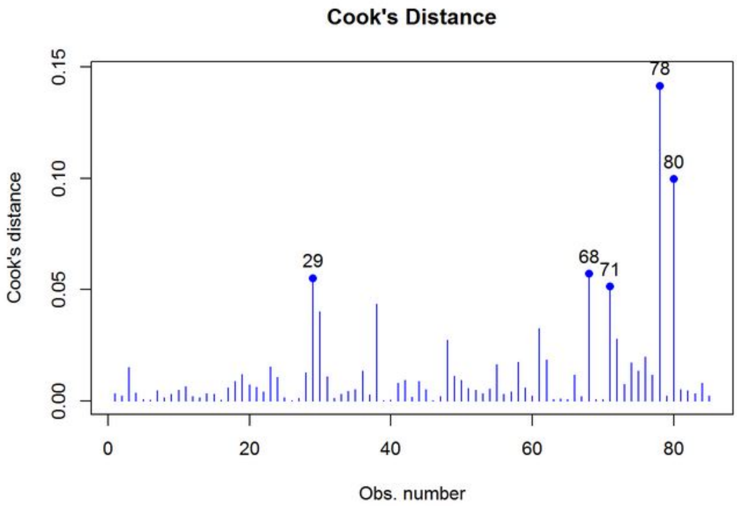


Figure 0.3.8:

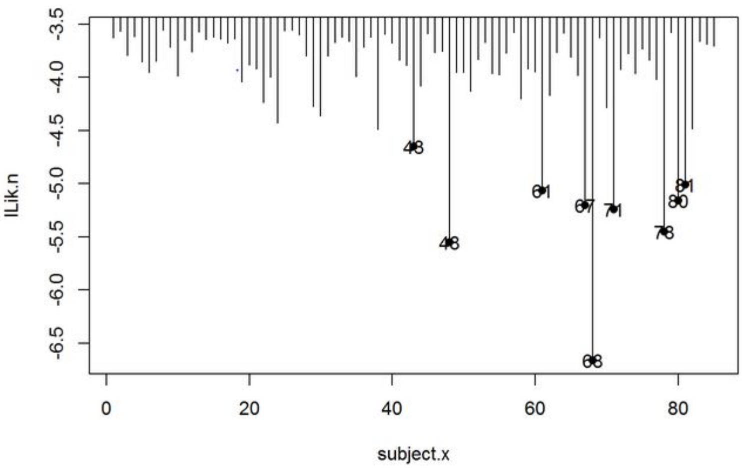


Figure 0.3.9:

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

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

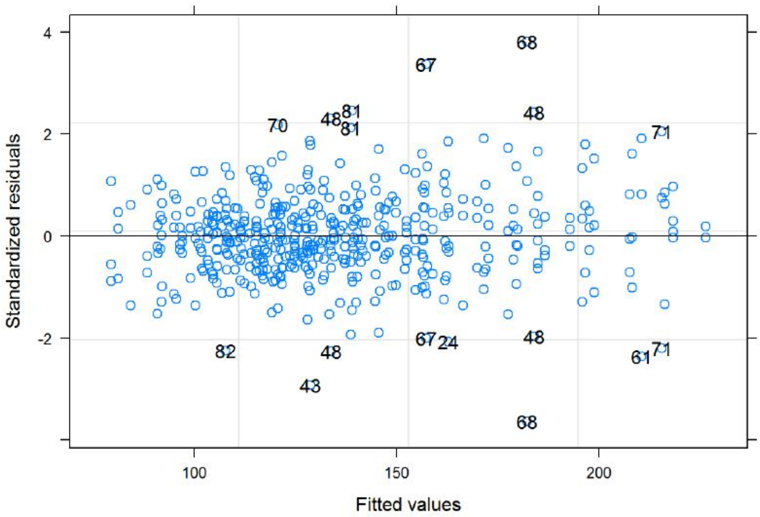


Figure 0.3.10:

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.

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

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.

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.

Mahalanbis Distance

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

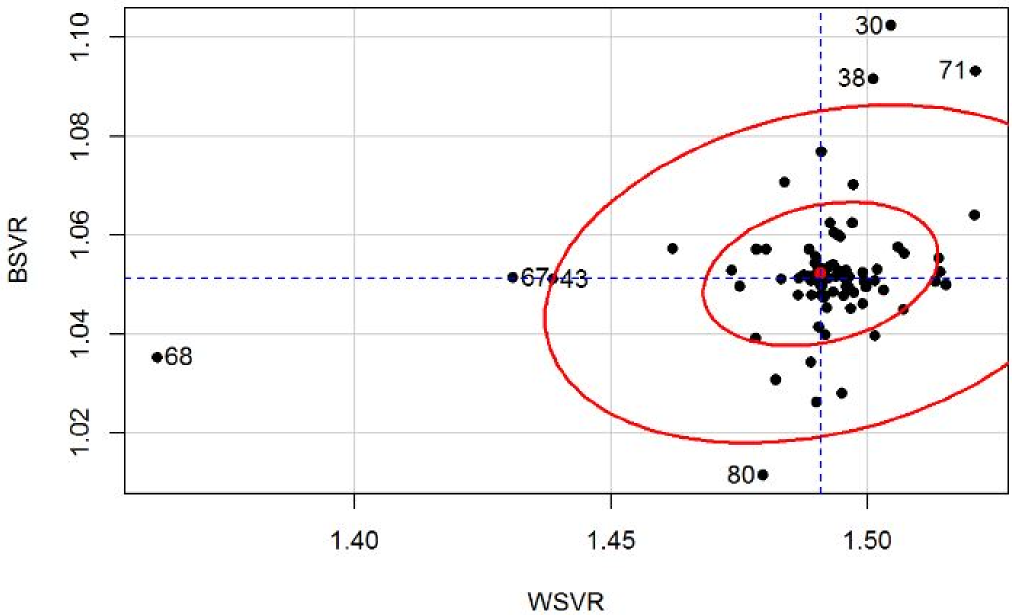


Figure 0.3.11:

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.

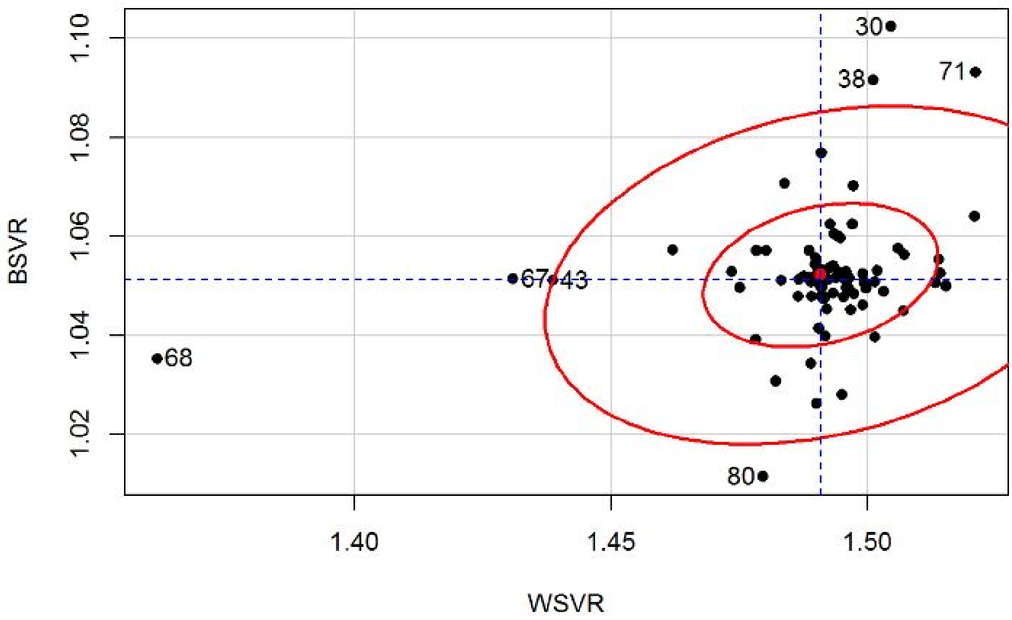


Figure 0.3.12:

Variance Ratios

The approach proposed by Roy deals with the question of agreement, and indeed interchangeability, as developed by Bland and Altmans corpus of work. In the view of Dunn, a question relevant to many practitioners is which of the two methods is more precise. The relationship between precision and the within-item and between-item variability must be established. Roy establishes the equivalence of repeatability and within-item variability, and hence precision. The method with the smaller within-item variability can be deemed to be the more precise.

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.

In fact, the ratio of within-item standard deviations, with the attendant confidence interval, can be determined using a single R command: `intervals()`.

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

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

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

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.

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.

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.

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

Variance Ratios

The approach proposed by Roy deals with the question of agreement, and indeed interchangeability, as developed by Bland and Altmans corpus of work. In the view of

Dunn, a question relevant to many practitioners is which of the two methods is more precise.

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

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.

In fact, the ratio of within-item standard deviations, with the attendant confidence interval, can be determined using a single R command: `intervals()`.

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

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

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

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.

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.

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.

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