Contents

- `R` command and `R` object - Typewriter Font
- `R` Package name - Italics
- Selected Acronyms and Proper Nouns - Italics

This chapter is broken into two parts. The first part is a review of diagnostics methods for linear models, intended to acquaint the reader with the subject, and also to provide a basis for material covered in the second part. Particular attention is drawn to graphical methods.

The second part of the chapter looks at diagnostics techniques for LME models, firsly covering the theory, then proceeding to a discussion on implementing these using `R` code. While a substantial body of work has been developed in this area, ther are still area worth exploring. In particular the development of graphical techniques pertinent to LME models should be looked at.

0.1 Introduction

In classical linear models model diagnostics have been become a required part of any statistical analysis, and the methods are commonly available in statistical packages and standard textbooks on applied regression. However it has been noted by several papers that model diagnostics do not often accompany LME model analyses. Model diagnostic techniques determine whether or not the distributional assumptions are satisfied, and to assess the influence of unusual observations.

0.1.1 What is Influence

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

0.1.2 Quantifying Influence

The basic procedure for quantifying influence is simple as follows:

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

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.

0.1.3 Model Data Agreement

Schabenberger(20XX) describes the examination of model-data agreement as comprising several elements;

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

0.1.4 Influence Diagnostics: Basic Idea and Statistics

The general idea of quantifying the influence of one or more observations relies on computing parameter estimates based on all data points, removing the cases in question from the data, refitting the model, and computing statistics based on the change between full-data and reduced-data estimation.

0.1.5 Quantifying Influence

The basic procedure for quantifying influence is simple as follows:

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

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.

0.1.6 Quantifying Influence

The basic procedure for quantifying influence is simple as follows:

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

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.

0.1.7 Residual diagnostics

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

Extension of techniques to LME Models

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

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

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

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

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

Influence Diagnostics Basic Idea and Statistics

The general idea of quantifying the influence of one or more observations relies on computing parameter estimates based on all data points, removing the cases in question from the data, refitting the model, and computing statistics based on the change between full-data and reduced-data estimation.

Influence statistics can be coarsely grouped by the aspect of estimation that is their primary target:

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

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.

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

0.1.8 Residual Plots

A residual plot is a graph that shows the residuals on the vertical axis and the independent variable on the horizontal axis. If the points in a residual plot are randomly dispersed around the horizontal axis, a linear regression model is appropriate for the data; otherwise, a non-linear model is more appropriate.

Below the table on the left shows inputs and outputs from a simple linear regression analysis, and the chart on the right displays the residual (e) and independent variable (X) as a residual plot.

```
x & 60 & 70 & 80 & 85 & 95 \\ \hline
y & 70 & 65 & 70 & 95 & 85 \\ \hline
y.hat 65.411 71.849 78.288 81.507 87.945
e 4.589 -6.849 -8.288 13.493 -2.945
```

The residual plot shows a fairly random pattern - the first residual is positive, the next two are negative, the fourth is positive, and the last residual is negative. This random pattern indicates that a linear model provides a decent fit to the data.

Below, the residual plots show three typical patterns. The first plot shows a random pattern, indicating a good fit for a linear model. The other plot patterns are non-random (U-shaped and inverted U), suggesting a better fit for a non-linear model.

Bibliography

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

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

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

Haslett, J. (1999). A simple derivation of deletion diagnostic results for the general linear model with correlated errors. *Journal of the Royal Statistical Society (Series B) 61*, 603–609.

Haslett, J. and D. Dillane (2004). Application of 'delete = replace' to deletion diagnostics for variance component estimation. *Journal of the Royal Statistical Society (Series B) 66*, 131–143.

Haslett, J. and K. Hayes (1998). Residuals for the linear model with general covariance structure. *Journal of the Royal Statistical Society (Series B) 60*, 201–215.

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

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

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

0.1.9 Introduction

In statistics and optimization, statistical errors and residuals are two closely related and easily confused measures of the deviation of an observed value of an element of a statistical sample from its "theoretical value". The error (or disturbance) of an observed value is the deviation of the observed value from the (unobservable) true function value, while the residual of an observed value is the difference between the observed value and the estimated function value.

The distinction is most important in regression analysis, where it leads to the concept of studentized residuals.

0.1.10 Residual

Residual (or error) represents unexplained (or residual) variation after fitting a regression model. It is the difference (or left over) between the observed value of the variable and the value suggested by the regression model.

The difference between the observed value of the dependent variable (y) and the predicted value () is called the residual (e). Each data point has one residual.

Residual = Observed value - Predicted value

$$e = y - \hat{y}$$

Both the sum and the mean of the residuals are equal to zero. That is, e = 0 and e = 0.

0.1.11 Residual

A residual (or fitting error), on the other hand, is an observable estimate of the unobservable statistical error. Consider the previous example with men's heights and suppose we have a random sample of n people. The sample mean could serve as a good estimator of the population mean. Then we have:

The difference between the height of each man in the sample and the unobservable population mean is a statistical error, whereas The difference between the height of each man in the sample and the observable sample mean is a residual. Note that the sum of the residuals within a random sample is necessarily zero, and thus the residuals are necessarily not independent. The statistical errors on the other hand are independent, and their sum within the random sample is almost surely not zero.

Other uses of the word "error" in statistics:

The use of the term "error" as discussed in the sections above is in the sense of a deviation of a value from a hypothetical unobserved value. At least two other uses also occur in statistics, both referring to observable prediction errors:

- Mean square error or mean squared error (abbreviated MSE) and root mean square error (RMSE) refer to the amount by which the values predicted by an estimator differ from the quantities being estimated (typically outside the sample from which the model was estimated).

- Sum of squared errors, typically abbreviated SSE or SSe, refers to the residual sum of squares (the sum of squared residuals) of a regression; this is the sum of the squares of the deviations of the actual values from the predicted values, within the sample used for estimation. Likewise, the sum of absolute errors (SAE) refers to the sum of the absolute values of the residuals, which is minimized in the least absolute deviations approach to regression.

0.1.12 Influence Diagnostics: Basic Idea and Statistics

The general idea of quantifying the influence of one or more observations relies on computing parameter estimates based on all data points, removing the cases in question from the data, refitting the model, and computing statistics based on the change between full-data and reduced-data estimation.

0.1.13 Standardization

A random variable is said to be standardized if the difference from its mean is scaled by its standard deviation. The residuals above have mean zero but their variance is unknown, it depends on the true values of θ. Standardization is thus not possible in practice.

0.1.14 Studentization

Instead, you can compute studentized residuals by dividing a residual by an estimate of its standard deviation.

0.1.15 Internal and External Studentization

If that estimate is independent of the $i-$th observation, the process is termed 'external studentization'. This is usually accomplished by excluding the $i-$th observation when computing the estimate of its standard error. If the observation contributes to the standard error computation, the residual is said to be internally studentized.

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

0.1.16 Computation

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

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

0.1.17 Standardized and studentized residuals

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

0.1.18 Internal and External Studentization

If that estimate is independent of the $i-$th observation, the process is termed 'external studentization'. This is usually accomplished by excluding the $i-$th observation when computing the estimate of its standard error. If the observation contributes to the standard error computation, the residual is said to be internally studentized.

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

0.1.19 Computation

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

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

0.1.20 Pearson Residual

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

0.1.21 Standardized and studentized residuals

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

0.1.22 Computation

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

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

0.1.23 Pearson Residual

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

0.1.24 Covariance Parameters

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

0.1.25 Computation

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

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

0.1.26 Studentization

In statistics, a studentized residual is the quotient resulting from the division of a residual by an estimate of its standard deviation. Typically the standard deviations of residuals in a sample vary greatly from one data point to another even when the errors all have the same standard deviation, particularly in regression analysis; thus it does not make sense to compare residuals at different data points without first studentizing. It is a form of a Student's t-statistic, with the estimate of error varying between points.

This is an important technique in the detection of outliers. It is named in honor of William Sealey Gosset, who wrote under the pseudonym Student, and dividing by an estimate of scale is called studentizing, in analogy with standardizing and normalizing: see Studentization.

0.2 Covariance Parameters

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

0.2.1 Pearson Residual

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

Bibliography

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

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

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

Haslett, J. (1999). A simple derivation of deletion diagnostic results for the general linear model with correlated errors. *Journal of the Royal Statistical Society (Series B) 61*, 603–609.

Haslett, J. and D. Dillane (2004). Application of 'delete = replace' to deletion diagnostics for variance component estimation. *Journal of the Royal Statistical Society (Series B) 66*, 131–143.

Haslett, J. and K. Hayes (1998). Residuals for the linear model with general covariance structure. *Journal of the Royal Statistical Society (Series B) 60*, 201–215.

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

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

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

Diagnostic Methods for OLS models

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

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

Influence Diagnostics: Basic Idea and Statistics

The general idea of quantifying the influence of one or more observations relies on computing parameter estimates based on all data points, removing the cases in question from the data, refitting the model, and computing statistics based on the change between full-data and reduced-data estimation.

0.3 Case Deletion Diagnostics

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

0.3.1 Deletion Diagnostics

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

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

Data from single individuals, or a small group of subjects may influence non-linear mixed effects model selection. Diagnostics routinely applied in model building may identify such individuals, but these methods are not specifically designed for that purpose and are, therefore, not optimal. We describe two likelihood-based diagnostics for identifying individuals that can influence the choice between two competing models.

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

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

0.4 Effects on fitted and predicted values

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

0.5 Leverage and Influence

0.5.1 Influence

The influence of an observation can be thought of in terms of how much the predicted scores for other observations would differ if the observation in question were not included.

Cook's D is a good measure of the influence of an observation and is proportional to the sum of the squared differences between predictions made with all observations in the analysis and predictions made leaving out the observation in question. 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.

0.5.2 Interpreting Cook's Distance

A common rule of thumb is that an observation with a value of Cook's D over 1.0 has too much influence. As with all rules of thumb, this rule should be applied judiciously and not thoughtlessly.

0.5.3 Leverage

The leverage of an observation is based on how much the observation's value on the predictor variable differs from the mean of the predictor variable. The greater an observation's leverage, the more potential it has to be an influential observation.

For example, an observation with a value equal to the mean on the predictor variable has no influence on the slope of the regression line regardless of its value on the criterion variable. On the other hand, an observation that is extreme on the predictor variable has the potential to affect the slope greatly.

Calculation of Leverage (h)

The first step is to standardize the predictor variable so that it has a mean of 0 and a standard deviation of 1. Then, the leverage (h) is computed by squaring the observation's value on the standardized predictor variable, adding 1, and dividing by the number of observations.

0.5.4 Summary of Influence Statistics

- **Studentized Residuals** Residuals divided by their estimated standard errors (like t-statistics). Observations with values larger than 3 in absolute value are considered outliers.

- **Leverage Values (Hat Diag)** Measure of how far an observation is from the others in terms of the levels of the independent variables (not the dependent variable). Observations with values larger than $2(k+1)/n$ are considered to be potentially highly influential, where k is the number of predictors and n is the sample size.

- **DFFITS** Measure of how much an observation has effected its fitted value from the regression model. Values larger than $2\sqrt{(k+1)/n}$ in absolute value are considered highly influential.

- **DFBETAS** Measure of how much an observation has effected the estimate of a regression coefficient (there is one DFBETA for each regression coefficient, including the intercept). Values larger than 2/sqrt(n) in absolute value are considered highly influential.

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

- **Cooks D** Measure of aggregate impact of each observation on the group of regression coefficients, as well as the group of fitted values. Values larger than 4/n are considered highly influential.

0.6 Influence analysis

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

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

0.7 Iterative and non-iterative influence analysis

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

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

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

0.7.1 Iterative Influence Analysis

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

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

0.8 Influence analysis

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

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

0.8.1 Cook's 1986 paper on Local Influence

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

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

0.8.2 Overall Influence

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

Bibliography

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

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

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

Haslett, J. (1999). A simple derivation of deletion diagnostic results for the general linear model with correlated errors. *Journal of the Royal Statistical Society (Series B) 61*, 603–609.

Haslett, J. and D. Dillane (2004). Application of 'delete = replace' to deletion diagnostics for variance component estimation. *Journal of the Royal Statistical Society (Series B) 66*, 131–143.

Haslett, J. and K. Hayes (1998). Residuals for the linear model with general covariance structure. *Journal of the Royal Statistical Society (Series B) 60*, 201–215.

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

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

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

0.8.3 Quantifying Influence

The basic procedure for quantifying influence is simple as follows:

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

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.

0.9 Measures 2

0.9.1 Cook's Distance

- For variance components γ

Diagnostic tool for variance components

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

0.9.2 Variance Ratio

- For fixed effect parameters β.

0.9.3 Variance Ratio

- For fixed effect parameters β.

0.9.4 Cook-Weisberg statistic

- For fixed effect parameters β.

0.9.5 Andrews-Pregibon statistic

- For fixed effect parameters β.

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

0.10 Measures 2

0.10.1 Cook's Distance

- For variance components γ

Diagnostic tool for variance components

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

0.11 Zewotir Measures of Influence in LME Models

Zewotir 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

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.

0.11.1 Information Ratio

0.12 Computation and Notation

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

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

0.13 Measures of Influence

The impact of an observation on a regression fitting can be determined by the difference between the estimated regression coefficient of a model with all observations and the estimated coefficient when the particular observation is deleted. The measure DFBETA is the studentized value of this difference.

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 $\hat{V}$.

0.13.1 DFFITS

DFFITS is a statistical measured designed to a show how influential an observation is in a statistical model. It is closely related to the studentized residual.

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

0.13.2 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

0.14 Measures of Influence

The impact of an observation on a regression fitting can be determined by the difference between the estimated regression coefficient of a model with all observations and the estimated coefficient when the particular observation is deleted. The measure DFBETA is the studentized value of this difference.

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 $\hat{V}$.

0.14.1 DFBETA

$$\begin{aligned} DFBETA_a &= \hat{\beta} - \hat{\beta}_{(a)} && (2) \\ &= B(Y - Y_{\bar{a}} && (3) \end{aligned}$$

0.14.2 DFFITS

DFFITS is a statistical measured designed to a show how influential an observation is in a statistical model. It is closely related to the studentized residual.

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

0.14.3 PRESS

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

$$PRESS = \sum (y - y^{(k)})^2 \tag{4}$$

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

0.14.4 DFBETA

$$\begin{aligned} DFBETA_a &= \hat{\beta} - \hat{\beta}_{(a)} && (5) \\ &= B(Y - Y_{\bar{a}} && (6) \end{aligned}$$

0.14.5 Influential Observations : DFBeta and DFBetas

Cook's distance refers to how far, on average, predicted y-values will move if the observation in question is dropped from the data set. dfbeta refers to how much a parameter estimate changes if the observation in question is dropped from the data set. Note that with k covariates, there will be k+1 dfbetas (the intercept,β_0, and 1 β for each covariate). Cook's distance is presumably more important to you if you are doing predictive modeling, whereas dfbeta is more important in explanatory modeling.

0.15 Efficient Updating Theorem

Zewotir and Galpin (2005) describes the basic theorem of efficient updating.

- $$m_i = \frac{1}{c_{ii}}$$

0.16 Zewotir Measures of Influence in LME Models

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

Chapter 1

Zewotir's Paper

1.1 Efficient Updating Theorem

Zewotir and Galpin (2005) describes the basic theorem of efficient updating.

- $$m_i = \frac{1}{c_{ii}}$$

1.2 Zewotir Measures of Influence in LME Models

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

1.2.1 Cook's Distance

- For variance components γ: $CD(\gamma)_i$,

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

Random Effects

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

linear functions

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

1.2.2 Information Ratio

1.3 Computation and Notation

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

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

1.4 Demidenko's I Influence

The concept of I Influence is generalized to the non linea regression model.

Chapter 2

Zewotir's Paper

2.1 Efficient Updating Theorem

Zewotir and Galpin (2005) describes the basic theorem of efficient updating.

- $$m_i = \frac{1}{c_{ii}}$$

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.

2.1.1 Information Ratio

2.2 Computation and Notation

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

Zewotir and Galpin (2005) remarks that $\boldsymbol{D}$ is a block diagonal with the $i-$th block being $u\boldsymbol{I}$

2.3 Measures 2

2.3.1 Cook's Distance

- For variance components γ

Diagnostic tool for variance components

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

2.3.2 Variance Ratio

- For fixed effect parameters β.

2.3.3 Cook-Weisberg statistic

- For fixed effect parameters β.

2.3.4 Andrews-Pregibon statistic

- For fixed effect parameters β.

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

2.4 Haslett's Analysis

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

Chapter 3

Zewotir's Paper

3.1 Efficient Updating Theorem

Zewotir and Galpin (2005) describes the basic theorem of efficient updating.

- $$m_i = \frac{1}{c_{ii}}$$

Random Effects

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

linear functions

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

3.1.1 Information Ratio

3.2 Computation and Notation

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

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

3.3 Haslett's Analysis

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

A general theory is presented for residuals from the general linear model with correlated errors. It is demonstrated that there are two fundamental types of residual associated with this model, referred to here as the marginal and the conditional residual.

These measure respectively the distance to the global aspects of the model as represented by the expected value and the local aspects as represented by the conditional expected value.

These residuals may be multivariate.

Haslett and Hayes (1998) developes some important dualities which have simple implications for diagnostics.

3.4 Demidenko's I Influence

The concept of I Influence is generalized to the non linea regression model.

Bibliography

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

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

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

Haslett, J. (1999). A simple derivation of deletion diagnostic results for the general linear model with correlated errors. *Journal of the Royal Statistical Society (Series B) 61*, 603–609.

Haslett, J. and D. Dillane (2004). Application of 'delete = replace' to deletion diagnostics for variance component estimation. *Journal of the Royal Statistical Society (Series B) 66*, 131–143.

Haslett, J. and K. Hayes (1998). Residuals for the linear model with general covariance structure. *Journal of the Royal Statistical Society (Series B) 60*, 201–215.

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

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

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

Diagnostics for repeated measurements in LME models.

Most currently available methods for detecting discordant subjects and observations in linear mixed effects model fits adapt existing methods for single-level regression data. The most common methods are generalizations of deletion-based approaches, primarily Cook's distance. This article describes the limitations of modifications to Cook's distance and local influence, and suggests a new nondeletion subject-level method, studentized residual sum of squares (TRSS) plots. We also suggest a new observation-level deletion method that detects discordant observations as an application of TRSS plots.

TRSS Plots

The proposed method provides greater information on repeated measurements by utilizing revised residuals and efficiently evaluating the effect of discordant subjects and observations on the estimation of parameters including variance components. We compare the performance of the proposed methods with current methods by using the orthodontic growth data: a longitudinal dataset with 27 subjects each observed four times. TRSS plots successfully identified discordant subjects that were missed by modified Cook's distance methods and the local influence approach.

- Diagnostics for repeated measurements in linear mixed effects models
- Most currently available methods for detecting discordant subjects and observations in linear mixed effects model fits adapt existing methods for single-level regression data.
- The most common methods are generalizations of *deletion-based* approaches, primarily Cook's distance.
- This article describes the limitations of modifications to Cook's distance and local influence, and suggests a new nondeletion subject-level method, studentized residual sum of squares (TRSS) plots.
- We also suggest a new observation-level deletion method that detects discordant observations as an application of TRSS plots. The proposed method provides greater information on repeated measurements by utilizing revised residuals and efficiently evaluating the effect of discordant subjects and observations on the estimation of parameters including variance components.
- We compare the performance of the proposed methods with current methods by using the orthodontic growth data: a longitudinal dataset with 27 subjects each observed four times.
- TRSS plots successfully identified discordant subjects that were missed by modified Cook's distance methods and the local influence approach.

3.5 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.

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

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

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

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

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

Influence is understood to be the ability of a single or multiple data points, through their presences or absence in the data, to alter important aspects of the analysis, yield qualitatively different inferences, or violate assumptions of the statistical model (Schabenberger, 2004).

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.

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.

A residual is the difference between an observed quantity and its estimated or predicted value. In LME models, there are two types of residuals, marginal residuals and conditional residuals. A marginal residual is the difference between the observed data and the estimated marginal mean. A conditional residual is the difference between the observed data and the predicted value of the observation. In a model without random effects, both sets of residuals coincide.

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.

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

$$\begin{aligned} LD_{(U)} &= 2\{l(\hat{\psi}) - l(\hat{\psi}_{(U)})\} && (3.5) \\ RLD_{(U)} &= 2\{l_R(\hat{\psi}) - l_R(\hat{\psi}_{(U)})\} && (3.6) \end{aligned}$$

3.5.1 Case Deletion Diagnostics

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.

3.5.2 Effects on fitted and predicted values

$$\hat{e}_{i(U)} = y_i - x\hat{\beta}_{(U)} \tag{3.7}$$

3.5.3 Case Deletion Diagnostics for Mixed Models

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

? develops these techniques in the context of REML

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

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

!"

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

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

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

Example: " !"

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

3.5.4 Methods and Measures

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

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.

3.5.5 Matrix Notation for Case Deletion

3.5.6 Case deletion notation

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

3.5.7 Partitioning Matrices

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

3.5.8 Case Deletion Diagnostics

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.

3.5.9 Case Deletion Diagnostics for Mixed Models

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

? develops these techniques in the context of REML

3.5.10 Terminology for Case Deletion diagnostics

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

3.5.11 Case Deletion Diagnostics

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.

3.5.12 Deletion Diagnostics

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

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

Data from single individuals, or a small group of subjects may influence non-linear mixed effects model selection. Diagnostics routinely applied in model building may identify such individuals, but these methods are not specifically designed for that purpose and are, therefore, not optimal. We describe two likelihood-based diagnostics for identifying individuals that can influence the choice between two competing models.

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

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

3.5.13 Terminology for Case Deletion diagnostics

Preisser(19XX) describes two type of diagnostics. When the set consists of only one observation, the type is called 'observation-diagnostics'. For multiple observations, Preisser describes the diagnostics as 'cluster-deletion' diagnostics.

3.5.14 Cook's Distance

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

Random Effects

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

linear functions

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

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

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

3.5.15 Cook's Distance

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

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

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

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

Cook's Distance

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

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

3.5.16 Cooks's Distance

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

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

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 definining.

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

3.6 Cook's Distance for LMEs

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

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

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

For random effect estimates, the Cook's distance is

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

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

3.6.1 Cook's Distance

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

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

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

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

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

For random effect estimates, the Cook's distance is

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

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

3.6.2 Change in the precision of estimates

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

3.6.3 Cook's Distance

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

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

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

3.6.4 Cooks's Distance

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

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

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

3.6.5 Cook's Distance

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

Interpretation

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

3.6.6 Cook's Distance

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

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

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

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

Bibliography

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

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

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

Haslett, J. (1999). A simple derivation of deletion diagnostic results for the general linear model with correlated errors. *Journal of the Royal Statistical Society (Series B) 61*, 603–609.

Haslett, J. and D. Dillane (2004). Application of 'delete = replace' to deletion diagnostics for variance component estimation. *Journal of the Royal Statistical Society (Series B) 66*, 131–143.

Haslett, J. and K. Hayes (1998). Residuals for the linear model with general covariance structure. *Journal of the Royal Statistical Society (Series B) 60*, 201–215.

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

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

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

3.7 Influence analysis

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

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

3.7.1 Cook's 1986 paper on Local Influence

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

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

3.7.2 Overall Influence

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

3.8 Terminology for Case Deletion diagnostics

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

3.9 Cook's Distance

3.9.1 Cook's Distance

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

3.9.2 Cooks's Distance

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

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

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

Bibliography

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

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

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

Haslett, J. (1999). A simple derivation of deletion diagnostic results for the general linear model with correlated errors. *Journal of the Royal Statistical Society (Series B) 61*, 603–609.

Haslett, J. and D. Dillane (2004). Application of 'delete = replace' to deletion diagnostics for variance component estimation. *Journal of the Royal Statistical Society (Series B) 66*, 131–143.

Haslett, J. and K. Hayes (1998). Residuals for the linear model with general covariance structure. *Journal of the Royal Statistical Society (Series B) 60*, 201–215.

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

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

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

3.9.3 Cook's Distance

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

Random Effects

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

linear functions

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

3.9.4 Information Ratio

3.10 Computation and Notation

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

Zewotir and Galpin (2005) remarks that $\boldsymbol{D}$ is a block diagonal with the $i-$th block being $u\boldsymbol{I}$

3.11 Cook's Distance for LMEs

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

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

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

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

For random effect estimates, the Cook's distance is

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

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

3.11.1 Change in the precision of estimates

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

3.12 The CPJ Paper

3.12.1 Case-Deletion results for Variance components

Christensen et al. (1992) examines case deletion results for estimates of the variance components, proposing the use of one-step estimates of variance components for examining case influence. The method describes focuses on REML estimation, but can easily be adapted to ML or other methods.

This paper develops their global influences for the deletion of single observations in two steps: a one-step estimate for the REML (or ML) estimate of the variance components, and an ordinary case-deletion diagnostic for a weighted regression problem (conditional on the estimated covariance matrix) for fixed effects.

3.12.2 CPJ Notation

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

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

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

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

- $\breve{x}_i$
- $\breve{z}_i$
- $\breve{z}_i j$
- $\breve{y}_i$
- $p_i i$

- m_i

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

3.13 Matrix Notation for Case Deletion

3.13.1 Case deletion notation

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

3.13.2 Partitioning Matrices

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

3.14 The CPJ Paper

3.14.1 Case-Deletion results for Variance components

Christensen et al. (1992) examines case deletion results for estimates of the variance components, proposing the use of one-step estimates of variance components for examining case influence. The method describes focuses on REML estimation, but can easily be adapted to ML or other methods.

This paper develops their global influences for the deletion of single observations in two steps: a one-step estimate for the REML (or ML) estimate of the variance components, and an ordinary case-deletion diagnostic for a weighted regression problem (conditional on the estimated covariance matrix) for fixed effects.

3.14.2 CPJ Notation

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

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

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

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

- $\breve{x}_i$
- $\breve{z}_i$
- $\breve{z}_i j$
- $\breve{y}_i$
- $p_i i$
- m_i

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

3.15 Matrix Notation for Case Deletion

3.15.1 Case deletion notation

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

3.15.2 Partitioning Matrices

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

3.16 CPJ's Three Propositions

Proposition 1

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

$$\boldsymbol{V}_{[i]}^{-1} = \boldsymbol{\Lambda}_{[i]} - \frac{\lambda_i \lambda_i'}{\lambda_i}$$

3.16.1 Proposition 2

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

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

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

3.16.2 Proposition 3

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

3.17 CPJ's Three Propositions

Proposition 1

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

$$\boldsymbol{V}_{[i]}^{-1} = \boldsymbol{\Lambda}_{[i]} - \frac{\lambda_i \lambda_i'}{\lambda_i}$$

Proposition 1

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

$$\boldsymbol{V}_{[i]}^{-1} = \boldsymbol{\Lambda}_{[i]} - \frac{\lambda_i \lambda_i'}{\lambda_i}$$

3.17.1 Proposition 2

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

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

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

3.18 CPJ's Three Propositions

3.18.1 Proposition 2

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

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

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

Bibliography

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

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

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

Haslett, J. (1999). A simple derivation of deletion diagnostic results for the general linear model with correlated errors. *Journal of the Royal Statistical Society (Series B) 61*, 603–609.

Haslett, J. and D. Dillane (2004). Application of 'delete = replace' to deletion diagnostics for variance component estimation. *Journal of the Royal Statistical Society (Series B) 66*, 131–143.

Haslett, J. and K. Hayes (1998). Residuals for the linear model with general covariance structure. *Journal of the Royal Statistical Society (Series B) 60*, 201–215.

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

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

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

3.18.2 Key Definitions

Residual The difference between the predicted value (based on the regression equation) and the actual, observed value.

Outlier In linear regression, an outlier is an observation with large residual. In other words, it is an observation whose dependent-variable value is unusual given its value on the predictor variables. An outlier may indicate a sample peculiarity or may indicate a data entry error or other problem.

Leverage An observation with an extreme value on a predictor variable is a point with high leverage. Leverage is a measure of how far an independent variable deviates from its mean. High leverage points can have a great amount of effect on the estimate of regression coefficients.

Influence An observation is said to be influential if removing the observation substantially changes the estimate of the regression coefficients. Influence can be thought of as the product of leverage and outlierness.

Cook's distance A measure that combines the information of leverage and residual of the observation.

3.18.3 Leverage

In statistics, leverage is a term used in connection with regression analysis and, in particular, in analyses aimed at identifying those observations that are far away from corresponding average predictor values. Leverage points do not necessarily have a large effect on the outcome of fitting regression models.

Leverage points are those observations, if any, made at extreme or outlying values of the independent variables such that the lack of neighboring observations means that the fitted regression model will pass close to that particular observation.

Modern computer packages for statistical analysis include, as part of their facilities for regression analysis, various quantitative measures for identifying influential observations: among these measures is partial leverage, a measure of how a variable contributes to the leverage of a datum.

3.18.4 Leverage in LME models

For the general mixed model, leverage can be defined through the projection matrix that results from a transformation of the model with the inverse of the Cholesky decomposition of , or through an oblique projector (Schabenberger, 2004). The MIXED procedure follows the latter path in the computation of influence diagnostics.

The leverage value reported for the th observation is the th diagonal entry of the matrix

which is the weight of the observation in contributing to its own predicted value, . While is idempotent, it is generally not symmetric and thus not a projection matrix in the narrow sense.

The properties of these leverages are generalizations of the properties in models with diagonal variance-covariance matrices. For example, , and in a model with intercept and , the leverage values

are and . The lower bound for is achieved in an intercept-only model, and the upper bound is achieved in a saturated model.

The trace of equals the rank of . If denotes the element in row , column of , then for a model containing only an intercept the diagonal elements of are

Because is a sum of elements in the th row of the inverse variance-covariance matrix, can be negative, even if the correlations among data points are nonnegative. In case of a saturated model with , .

Nobre Singer : Mixed Model Residuals

Usually one assumes

- $b_i \sim N_q(0, G) i = 1, ..., m$
- $e_i \sim N_{n_i}(0, \sigma_i)$
- b_i and e_i independent

- G and σ_i are $(q \times q)$ and $(n_i \times n_i)$ positive denite matrices with elements expressed as functions of a vector of covariance parameters θ not functionally related to β

- If $\sigma_i = I_{n_i}\sigma^2$: homoskedastic conditional independence model

$$\begin{bmatrix} \boldsymbol{b} \\ \boldsymbol{e} \end{bmatrix} \sim \mathcal{N}_{qm+n}$$

$$\boldsymbol{Q} = \boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}$$

Sensitivity and residual analysis of the underlying assumptions constitute important tools for evaluating the fit of any model to given data.

Generalized Leverage

3.19 Case Deletion Diagnostics for LME models

Haslett & Dillane (19XX) 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 (19XX) describes three propositions that are required for efficient case-deletion in LME models. The first proposition decribes how to efficiently update V when the ith element is deleted.

$$V_{[i]}^{-1} = \Lambda_{[i]} - \frac{\lambda\lambda\prime}{\nu ii} \tag{3.8}$$

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

$$\begin{aligned} X'_{[i]}V_{[i]}^{-1}X_{[i]} &= X'V^{-1}X - \frac{\hat{x}_i\hat{x}'_i}{s_i} && (3.9) \\ (X'_{[i]}V_{[i]}^{-1}X_{[i]})^{-1} &= (X'V^{-1}X)^{-1} + \frac{(X'V^{-1}X)^{-1}\hat{x}_i\hat{x}'_i(X'V^{-1}X)^{-1}}{s_i - \bar{h}_i} && (3.10) \\ X'_{[i]}V_{[i]}^{-1}Y_{[i]} &= X\prime V^{-1}Y - \frac{\hat{x}_i\hat{y}'_i}{s_i} && (3.11) \end{aligned}$$

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

$$LD_{(U)} = 2\{l(\hat{\psi}) - l(\hat{\psi}_{(U)})\} \tag{3.12}$$

$$RLD_{(U)} = 2\{l_R(\hat{\psi}) - l_R(\hat{\psi}_{(U)})\} \tag{3.13}$$

Haslett & Dillane (199X) 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 & Dillane (199X) propose an alternative , and computationally inexpensive approach, making use of the '`delete=replace`' identity.

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

3.20 Haslett's Analysis

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

Bibliography

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

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

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

Haslett, J. (1999). A simple derivation of deletion diagnostic results for the general linear model with correlated errors. *Journal of the Royal Statistical Society (Series B) 61*, 603–609.

Haslett, J. and D. Dillane (2004). Application of 'delete = replace' to deletion diagnostics for variance component estimation. *Journal of the Royal Statistical Society (Series B) 66*, 131–143.

Haslett, J. and K. Hayes (1998). Residuals for the linear model with general covariance structure. *Journal of the Royal Statistical Society (Series B) 60*, 201–215.

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

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

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

Haslett and Hayes - Residuals

Haslett and Hayes (1998) and Haslett (1999) considered the case of an LME model with correlated covariance structure.

3.20.1 Residual Diagnostics in LME models

- A **residual** is the difference between the observed quantity and the predicted value. In LME models a distinction is made between marginal residuals and conditional residuals.
- A **Marginal residual** is the difference between the observed data and the estimated marginal mean (Schabenberger pg3) The computation of case deletion diagnostics in the classical model is made simple by the fact that important estimates can be computed without refitting the model.
- Such update formulae are available in the mixed model only if you assume that the covariance parameters are not affect by the removal of the observation in question. Schabenberger remarks that this is not a reasonable assumption.

Basic procedure for quantifying influence is simple

1. Fit the model to the data

2. Remove one or more data points from the analysis and compute updated estimates of model parameters

3. Based on the full and reduced data estimates, contrast quantities of interest to determine how the absence of the observations changed the analysis.

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

3.21 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.

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

$$V_{[i]}^{-1} = \Lambda_{[i]} - \frac{\lambda\lambda\prime}{\nu ii} \tag{3.14}$$

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

$$\begin{aligned} X'_{[i]}V_{[i]}^{-1}X_{[i]} &= X'V^{-1}X - \frac{\hat{x}_i\hat{x}'_i}{s_i} && (3.15) \\ (X'_{[i]}V_{[i]}^{-1}X_{[i]})^{-1} &= (X'V^{-1}X)^{-1} + \frac{(X'V^{-1}X)^{-1}\hat{x}_i\hat{x}'_i(X'V^{-1}X)^{-1}}{s_i - \bar{h}_i} && (3.16) \\ X'_{[i]}V_{[i]}^{-1}Y_{[i]} &= X\prime V^{-1}Y - \frac{\hat{x}_i\hat{y}'_i}{s_i} && (3.17) \end{aligned}$$

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

$$
\begin{aligned}
LD_{(U)} &= 2\{l(\hat{\psi}) - l(\hat{\psi}_{(U)})\} & (3.18)\\
RLD_{(U)} &= 2\{l_R(\hat{\psi}) - l_R(\hat{\psi}_{(U)})\} & (3.19)
\end{aligned}
$$

3.22 Turkan's LMEs

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

Some diagnostics which are analogue of diagnostics in multiple linear regression were developed to detect outliers and influential observations in the linear mixed model. *In this paper, the new diagnostic measure which is analogue of the Pena's influence statistic is developed for the linear mixed model.*

Estimation and Building blacks in LME models

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

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

The proposed diagnostic Measure.

Bibliography

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

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

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

Haslett, J. (1999). A simple derivation of deletion diagnostic results for the general linear model with correlated errors. *Journal of the Royal Statistical Society (Series B) 61*, 603–609.

Haslett, J. and D. Dillane (2004). Application of 'delete = replace' to deletion diagnostics for variance component estimation. *Journal of the Royal Statistical Society (Series B) 66*, 131–143.

Haslett, J. and K. Hayes (1998). Residuals for the linear model with general covariance structure. *Journal of the Royal Statistical Society (Series B) 60*, 201–215.

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

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

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

Influence.ME: Tools for Detecting Influential Data in Mixed Effects Models

Bibliography

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

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

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

Haslett, J. (1999). A simple derivation of deletion diagnostic results for the general linear model with correlated errors. *Journal of the Royal Statistical Society (Series B) 61*, 603–609.

Haslett, J. and D. Dillane (2004). Application of 'delete = replace' to deletion diagnostics for variance component estimation. *Journal of the Royal Statistical Society (Series B) 66*, 131–143.

Haslett, J. and K. Hayes (1998). Residuals for the linear model with general covariance structure. *Journal of the Royal Statistical Society (Series B) 60*, 201–215.

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

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

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

- *The previous Section (Section 4) is a literary review of residual diagnostics and influence procedures for Linear Mixed Effects Models, drawing heavily on Schabenberger and Zewotir.*

- *Section 4 begins with an introduction to key topics in residual diagnostics, such as influence, leverage, outliers and Cook's distance. Other concepts such as DFFITS and DFBETAs will be introduced briefly, mostly to explain why the are not particularly useful for the Method Comparison context, and therefore are not elaborated upon.*

- *In brief, Variable Selection is not applicable to Method Comparison Studies, in the commonly used used context. Testing a rather simplisticy specificied model against one with more random effects terms is tractable, but this research question is of secondary importance.*

Appendix to Section 4

As an appendix to section 4, an appraisal of the current state of development (or lack thereof) for current implemenations for LME models, particularly for `nlme` and `lme4` fitted models.

Crucially, a review of internet resources indicates that almost all of the progress in this regard has been done for `lme4` fitted models, specifically the *Influence.ME* `R` package. (Nieuwenhuis et 2012)

Conversely there is very little for `nlme` models. To delve into this mor, one would immediately investigate the current development workflow for both packages.

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

The `nlme` package

With regards to `nlme`, the torch has been passed to Galecki Galecki & Burzykowski (UMich. and Hasselt respecitely). Galecki & Burzykowski published *Linear Mixed Effects Models using R*. Also, the accompanying `R` package, nlmeU package is under current development, with a version being released XXXX.

The `lme4` package

The `lme4` package is also under active development, under the leadership of Ben Bolker (McMaster University). According to CRAN, the LME4 package, fits linear and generalized linear mixed-effects models

> The models and their components are represented using S4 classes and methods. The core computational algorithms are implemented using the Eigen C++ library for numerical linear algebra and RcppEigen ”glue”. (CRAN)

The key issue is that `nlme` allows for the particular specification of Roy's Model, speciifically direct spefiication of the VC matrices for within subject and between subject residuals. The `lme4` package does not allow for this. To advance the ideas that eminate from Roys' paper, one is required to use the `nlme` context. However, to take advantage of the infrastructure already provided for `lme4` models, one may change the research question away from that of Roy's paper. To this end, an exploration of what textitinfluence.ME can accomplished is merited. As a complement to this, one can also consider how to properly employ the R^2 measure, in the context of Methoc Comparison Studies, further to the work by Edwards et al, namely "An R^2 statistic for fixed effects in the linear mixed model".

Abstract for "An R^2 statistic for fixed effects in the linear mixed model" Statisticians most often use the linear mixed model to analyze Gaussian longitudinal data.

The value and familiarity of the R2 statistic in the linear univariate model naturally creates great interest in extending it to the linear mixed model. We define and describe how to compute a model R2 statistic for the linear mixed model by using only a single model.

The proposed R2 statistic measures multivariate association between the repeated outcomes and the fixed effects in the linear mixed model. The R2 statistic arises as a 11 function of an appropriate F statistic for testing all fixed effects (except typically the intercept) in a full model.

The statistic compares the full model with a null model with all fixed effects deleted (except typically the intercept) while retaining exactly the same covariance structure.

Furthermore, the R2 statistic leads immediately to a natural definition of a partial R2 statistic. A mixed model in which ethnicity gives a very small p-value as a longitudinal predictor of blood pressure (BP) compellingly illustrates the value of the statistic.

In sharp contrast to the extreme p-value, a very small R^2 , a measure of statistical and scientific importance, indicates that ethnicity has an almost negligible association with the repeated BP outcomes for the study.

Leave-One-Out Diagnostics with `lmeU`

Galecki et al discuss the matter of LME influence diagnostics in their book, although not into great detail.

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

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

Likelihood Displacement

Missing Data in Method Comparison Studies

The matter of missing data has not been commonly encountered in either Method Comparison Studies or Linear Mixed Effects Modelling. However Roy (2009) deals with the relevant assumptions regrading missing data.

Galecki & Burzykowski (2013) tackles the subject of missing data in LME Modelling.

Furthermore the nlmeU package includes the `patMiss` function, which "allows to compactly present pattern of missing data in a given vector/matrix/data frame or combination of thereof".

Bibliography

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

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

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

Haslett, J. (1999). A simple derivation of deletion diagnostic results for the general linear model with correlated errors. *Journal of the Royal Statistical Society (Series B) 61*, 603–609.

Haslett, J. and D. Dillane (2004). Application of 'delete = replace' to deletion diagnostics for variance component estimation. *Journal of the Royal Statistical Society (Series B) 66*, 131–143.

Haslett, J. and K. Hayes (1998). Residuals for the linear model with general covariance structure. *Journal of the Royal Statistical Society (Series B) 60*, 201–215.

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

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

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