

# Contents

0.1	Overall Influence . . . . .	4
0.2	Cook's 1986 paper on Local Influence . . . . .	4
<b>1</b>	<b>Measures of Influence</b>	<b>5</b>
1.1	DFFITS . . . . .	5
1.2	PRESS . . . . .	5
<b>2</b>	<b>Influence in LME Models</b>	<b>5</b>
<b>3</b>	<b>Influence analysis for LME Models</b>	<b>6</b>
3.1	Influence Diagnostics: Basic Idea and Statistics . . . . .	6
3.2	Influence Analysis for LME Models . . . . .	6
3.3	Influence Statistics for LME models . . . . .	7
3.4	Influence Analysis for LME Models . . . . .	10
3.5	Influence Statistics for LME models . . . . .	10
3.6	What is Influence . . . . .	10
3.7	Quantifying Influence . . . . .	11
<b>4</b>	<b>Extension of techniques to LME Models</b>	<b>12</b>
<b>5</b>	<b>Influence analysis</b>	<b>13</b>
5.1	Cook's 1986 paper on Local Influence . . . . .	13
5.2	Overall Influence . . . . .	13
<b>6</b>	<b>Terminology for Case Deletion diagnostics</b>	<b>14</b>
<b>7</b>	<b>Cook's Distance</b>	<b>15</b>
7.1	Cook's Distance . . . . .	15
7.2	Cooks's Distance . . . . .	15
<b>8</b>	<b>Efficient Updating Theorem</b>	<b>16</b>
8.0.1	Random Effects . . . . .	16
8.0.2	linear functions . . . . .	17
<b>9</b>	<b>Efficient Updating Theorem</b>	<b>17</b>

<b>10 Zewotir Measures of Influence in LME Models</b>	<b>17</b>
<b>11 Zewotir Measures of Influence in LME Models</b>	<b>18</b>
11.1 Cook's Distance . . . . .	18
11.1.1 Random Effects . . . . .	19
11.1.2 linear functions . . . . .	19
<b>12 Computation and Notation</b>	<b>20</b>
<b>13 Demidenko's I Influence</b>	<b>20</b>
<b>14 Computation and Notation</b>	<b>20</b>
<b>15 Measures 2</b>	<b>21</b>
15.1 Cook's Distance . . . . .	21
15.2 Variance Ratio . . . . .	21
15.3 Cook-Weisberg statistic . . . . .	21
15.4 Andrews-Pregibon statistic . . . . .	21
<b>16 Haslett's Analysis</b>	<b>22</b>
<b>17 Computation and Notation</b>	<b>23</b>
<b>18 Haslett's Analysis</b>	<b>24</b>
<b>19 Demidenko's I Influence</b>	<b>24</b>
19.1 Extension of techniques to LME Models . . . . .	26
19.2 Influence Statistics for LME models . . . . .	27
19.3 Extension of techniques to LME Models . . . . .	28
19.4 Standardized and studentized residuals . . . . .	29
19.5 Residual Analysis for Linear Models, LME models and GLMs . . . . .	29
19.6 Identifying outliers with a LME model object . . . . .	30
19.7 Diagnostics for Random Effects . . . . .	30
19.8 Influence Diagnostics: Basic Idea and Statistics . . . . .	31
19.9 Case Deletion Diagnostics for Mixed Models . . . . .	31
19.10Methods and Measures . . . . .	31
19.11Cook's 1986 paper on Local Influence . . . . .	31

19.12Cook's Distance . . . . .	32
19.12.1 Random Effects . . . . .	33
19.12.2 linear functions . . . . .	33
<b>20 Computation and Notation</b>	<b>34</b>
<b>21 Cook's Distance for LMEs</b>	<b>35</b>
21.1 Change in the precision of estimates . . . . .	35
21.2 Cook's Distance . . . . .	35
21.2.1 Random Effects . . . . .	36
21.2.2 linear functions . . . . .	36
21.3 Cook's Distance . . . . .	36
21.4 Cooks's Distance . . . . .	37
<b>22 Cook's Distance for LMEs</b>	<b>38</b>
22.1 Cook's Distance . . . . .	38
22.2 Change in the precision of estimates . . . . .	39
22.3 Cook's Distance . . . . .	39
22.4 Cooks's Distance . . . . .	40
22.5 Cook's Distance . . . . .	40
22.5.1 Interpretation . . . . .	40
22.6 Cook's Distance . . . . .	41
<b>23 The Hat Matrix</b>	<b>43</b>
<b>24 Sherman Morrison Woodbury Formula</b>	<b>45</b>
24.1 Hat Values for MCS regression . . . . .	45
<b>25 Cross Validation</b>	<b>46</b>
25.1 Cross Validation: Updating standard deviation . . . . .	49
<b>26 Updating Estimates</b>	<b>50</b>
26.1 Updating Standard deviation . . . . .	50
26.2 Updating of Regression Estimates . . . . .	50
26.3 Updating Regression Estimates . . . . .	50
26.4 Inference on intercept and slope . . . . .	51

26.4.1 Inference on correlation coefficient . . . . .	51
<b>27 Lesaffre's paper.</b>	<b>52</b>
<b>28 Cross Validation</b>	<b>54</b>
<b>29 Updating Estimates</b>	<b>54</b>
29.1 Updating of Regression Estimates . . . . .	54
29.2 Updating Standard deviation . . . . .	54
29.3 Updating of Regression Estimates . . . . .	55
29.4 Updating Regression Estimates . . . . .	55
29.5 Inference on intercept and slope . . . . .	56
29.5.1 Inference on correlation coefficient . . . . .	56
29.6 Updating Regression Estimates . . . . .	57
29.7 Updating of Regression Estimates . . . . .	57
29.8 Updating Standard deviation . . . . .	58
29.9 Inference on intercept and slope . . . . .	58
29.9.1 Inference on correlation coefficient . . . . .	58
Influence	

## 0.1 Overall Influence

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

## 0.2 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 quite different from the case deletion approach, comparisons are of interest.

# 1 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.

## 1.1 DFFITS

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

$$DFFITS = \frac{\hat{y}_i - \widehat{y}_{i(k)}}{s_{(k)}\sqrt{h_{ii}}}$$

## 1.2 PRESS

The prediction residual sum of squares (PRESS) is a 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.

$$PRESS = \sum (y - y^{(k)})^2 \quad (1)$$

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

## DFBETA

$$DFBETA_a = \hat{\beta} - \hat{\beta}_{(a)} \quad (2)$$

$$= B(Y - Y_a) \quad (3)$$

# 2 Influence in LME Models

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

### 3 Influence analysis for LME Models

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  $\beta$  and  $\theta$ . 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’.

#### 3.1 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.

#### 3.2 Influence Analysis for LME Models

The linear mixed effects model is a useful methodology for fitting a wide range of models. However, linear mixed effects models are known to be sensitive to outliers. Christensen et al. (1992) advises that identification of outliers is necessary before conclusions may be drawn from the fitted model.

Standard statistical packages concentrate on calculating and testing parameter estimates without considering the diagnostics of the model. The assessment of the effects of perturbations in data, on the outcome of the analysis, is known as statistical influence analysis. Influence analysis examines the robustness of the model. Influence analysis methodologies have been used extensively in classical linear models, and provided the basis for methodologies for use with LME models. Computationally inexpensive diagnostics tools have been developed to examine the issue of influence (Zewotir and Galpin, 2005). Studentized residuals, error contrast matrices and the inverse of the response variance covariance matrix are regular components of these tools.

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

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

## Influence Analysis for LME Models

The linear mixed effects model is a useful methodology for fitting a wide range of models. However, linear mixed effects models are known to be sensitive to outliers. Christensen et al advises that identification of outliers is necessary before conclusions may be drawn from the fitted model.

Standard statistical packages concentrate on calculating and testing parameter estimates without considering the diagnostics of the model. The assessment of the effects of perturbations in data, on the outcome of the analysis, is known as statistical influence analysis. Influence analysis examines the robustness of the model. Influence analysis methodologies have been used extensively in classical linear models, and provided the basis for methodologies for use with LME models. Computationally inexpensive diagnostics tools have been developed to examine the issue of influence (Zewotir). Studentized residuals, error contrast matrices and the inverse of the response variance covariance matrix are regular components of these tools.

### 3.3 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

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

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

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

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



## 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

### 3.4 Influence Analysis for LME Models

The linear mixed effects model is a useful methodology for fitting a wide range of models. However, linear mixed effects models are known to be sensitive to outliers. Christensen et al. (1992) advises that identification of outliers is necessary before conclusions may be drawn from the fitted model.

Standard statistical packages concentrate on calculating and testing parameter estimates without considering the diagnostics of the model. The assessment of the effects of perturbations in data, on the outcome of the analysis, is known as statistical influence analysis. Influence analysis examines the robustness of the model. Influence analysis methodologies have been used extensively in classical linear models, and provided the basis for methodologies for use with LME models. Computationally inexpensive diagnostics tools have been developed to examine the issue of influence (Zewotir and Galpin, 2005). Studentized residuals, error contrast matrices and the inverse of the response variance covariance matrix are regular components of these tools.

### 3.5 Influence Statistics for LME models

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

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

### 3.6 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).

### 3.7 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.

## 4 Extension of techniques to LME Models

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

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

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

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

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

## 5 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  $\beta$  and  $\theta$ . 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’.

### 5.1 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 quite different from the case deletion approach, comparisons are of interest.

### 5.2 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 ].

## 6 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.

## 7 Cook's Distance

### 7.1 Cook's Distance

Cooks Distance ( $D_i$ ) is an overall measure of the combined impact of the  $i$ th 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  $k$ th case is deleted.  $D_{(k)}$  can be calculated without fitting a new regression coefficient each time an observation is deleted.

### 7.2 Cook's Distance

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

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

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

Cook's Distance How to extract/compute leverage and Cook's distances for linear mixed effects models

Does anyone know how to compute (or extract) leverage and Cook's distances for a mer class object (obtained through lme4 package)? I'd like to plot these for a residuals analysis.

You should have a look at the R package influence.ME. It allows you to compute measures of influential data for mixed effects models generated by lme4.

An example model:

```
library(lme4) model <- lmer(mpg ~ disp + (1 | cyl), mtcars)
```

The function influence is the basis for all further steps:

```
library(influence.ME) infl <- influence(model, obs = TRUE)
```

Calculate Cook's distance:

```
cooks.distance(infl)
```

Plot Cook's distance:

```
plot(infl, which = "cook")
```

enter image description here

How to extract/compute leverage and Cook's distances for linear mixed effects models

Does anyone know how to compute (or extract) leverage and Cook's distances for a mer class object (obtained through lme4 package)? I'd like to plot these for a residuals analysis.

You should have a look at the R package influence.ME. It allows you to compute measures of influential data for mixed effects models generated by lme4.

An example model:

```
library(lme4)
model <- lmer(mpg ~ disp + (1 | cyl), mtcars)
```

The function influence is the basis for all further steps:

```
library(influence.ME)
infl <- influence(model, obs = TRUE)
```

Calculate Cook's distance:

```
cooks.distance(infl)
```

Plot Cook's distance:

```
plot(infl, which = "cook")
```

Zewotir

## 8 Efficient Updating Theorem

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

- 

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

### 8.0.1 Random Effects

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



### 8.0.2 linear functions

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

## 9 Efficient Updating Theorem

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

- 

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

## 10 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

## 11 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

### 11.1 Cook's Distance

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

### 11.1.1 Random Effects

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

### 11.1.2 linear functions

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

## 12 Computation and Notation

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

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

## 13 Demidenko's I Influence

The concept of I Influence is generalized to the non linear regression model. Zewotir's Paper

## 14 Computation and Notation

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

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

## 15 Measures 2

### 15.1 Cook's Distance

- For variance components  $\gamma$

Diagnostic tool for variance components

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

### 15.2 Variance Ratio

- For fixed effect parameters  $\beta$ .

### 15.3 Cook-Weisberg statistic

- For fixed effect parameters  $\beta$ .

### 15.4 Andrews-Pregibon statistic

- For fixed effect parameters  $\beta$ .

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.

## 16 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.

## 17 Computation and Notation

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

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

## 18 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) develop some important dualities which have simple implications for diagnostics.

## 19 Demidenko's I Influence

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

## References

- Beckman, R., C. Nachtsheim, and R. Cook (1987). Diagnostics for mixed-model analysis of variance. *Technometrics* 29(4), 413–426.
- 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. and K. Hayes (1998). Residuals for the linear model with general covariance structure. *Journal of the Royal Statistical Society (Series B)* 60, 201–215.



- 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.

## 19.1 Extension of techniques to LME Models

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

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

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

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

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

## 19.2 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

### 19.3 Extension of techniques to LME Models

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

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

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

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

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

## 19.4 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’.

## 19.5 Residual Analysis for Linear Models, LME models and GLMs

### Keywords:

- Residuals (*Beginners*),
- Testing the Assumption of Normality (*Beginners*)
- Diagnostic Plots with the `plot` function
- Cook’s Distance
- DFFits and DFBeta
- Standardized and Studentized Residuals
- Influence Leverage and Outlierness

## 19.6 Identifying outliers with a LME model object

The process is slightly different than with standard LME model objects, since the *influence* function does not work on lme model objects. Given *mod.lme*, we can use the plot function to identify outliers.

## 19.7 Diagnostics for Random Effects

Empirical best linear unbiased predictors EBLUPS provide the a useful way of diagnosing random effects.

EBLUPs are also known as “shrinkage estimators” because they tend to be smaller than the estimated effects would be if they were computed by treating a random factor as if it was fixed (West et al )

## 19.8 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.

## 19.9 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

## 19.10 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-Prebignon statistic.

## 19.11 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 quite different from the case deletion approach, comparisons are of interest.

## 19.12 Cook's Distance

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



### **19.12.1 Random Effects**

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

### **19.12.2 linear functions**

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

## 20 Computation and Notation

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

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

## 21 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  $\beta$  or  $\theta$ .

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.

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

For random effect estimates, the Cook's distance is

$$CD_i(b) = g'_{(i)}(I_r + \text{var}(\hat{b})D)^{-2}\text{var}(\hat{b})g_{(i)}.$$

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

### 21.1 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.

### 21.2 Cook's Distance

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

### 21.2.1 Random Effects

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

### 21.2.2 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.

## 21.3 Cook's Distance

In classical linear regression, a commonly used measure 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 defining.

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

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

### Cook's Distance

Cook's Distance ( $D_i$ ) is an overall measure of the combined impact of the  $i$ th 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.

## 21.4 Cook's Distance

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

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

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  $\mathbf{X}$ .

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

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

## 22 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.

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

For random effect estimates, the Cook's distance is

$$CD_i(b) = g'_{(i)}(I_r + \text{var}(\hat{b})D)^{-2}\text{var}(\hat{b})g_{(i)}.$$

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

### 22.1 Cook's Distance

Cooks Distance ( $D_i$ ) is an overall measure of the combined impact of the  $i$ th 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  $k$ th 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  $\beta$  or  $\theta$ .

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  $\mathbf{X}$  (?).

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

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

For random effect estimates, the Cook's distance is

$$CD_i(b) = g_{(i)}(I_r + \text{var}(\hat{b})D)^{-2}\text{var}(\hat{b})g_{(i)}.$$

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

## 22.2 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.

## 22.3 Cook's Distance

Cooks Distance ( $D_i$ ) is an overall measure of the combined impact of the  $i$ th 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  $k$ th 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  $\beta$  or  $\theta$ .

## 22.4 Cook's Distance

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

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

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

## 22.5 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.

### 22.5.1 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.



## 22.6 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. Appendices

## sectionThe Hat Matrix

The projection matrix  $H$  (also known as the hat matrix), is a well known identity that maps the fitted values  $\hat{Y}$  to the observed values  $Y$ , i.e.  $\hat{Y} = HY$ .

$$H = X(X^T X)^{-1} X^T \quad (4)$$

$H$  describes the influence each observed value has on each fitted value. The diagonal elements of the  $H$  are the ‘leverages’, which describe the influence each observed value has on the fitted value for that same observation. The residuals ( $R$ ) are related to the observed values by the following formula:

$$R = (I - H)Y \quad (5)$$

The variances of  $Y$  and  $R$  can be expressed as:

$$\begin{aligned} \text{var}(Y) &= H\sigma^2 \\ \text{var}(R) &= (I - H)\sigma^2 \end{aligned} \quad (6)$$

Updating techniques allow an economic approach to recalculating the projection matrix,  $H$ , by removing the necessity to refit the model each time it is updated. However this approach is known for numerical instability in the case of down-dating.

## sectionThe Hat Matrix

The projection matrix  $H$  (also known as the hat matrix), is a well known identity that maps the fitted values  $\hat{Y}$  to the observed values  $Y$ , i.e.  $\hat{Y} = HY$ .

$$H = X(X^T X)^{-1} X^T \quad (7)$$

$H$  describes the influence each observed value has on each fitted value. The diagonal elements of the  $H$  are the ‘leverages’, which describe the influence each observed value has on the fitted value for that same observation. The residuals ( $R$ ) are related to the observed values by the following formula:

$$R = (I - H)Y \quad (8)$$

The variances of  $Y$  and  $R$  can be expressed as:

$$\begin{aligned} \text{var}(Y) &= H\sigma^2 \\ \text{var}(R) &= (I - H)\sigma^2 \end{aligned} \quad (9)$$

Updating techniques allow an economic approach to recalculating the projection matrix,  $H$ , by removing the necessity to refit the model each time it is updated. However this approach is known for numerical instability in the case of down-dating.

## 23 The Hat Matrix

The projection matrix  $H$  (also known as the hat matrix), is a well known identity that maps the fitted values  $\hat{Y}$  to the observed values  $Y$ , i.e.  $\hat{Y} = HY$ .

$$H = X(X^T X)^{-1} X^T \quad (10)$$

$H$  describes the influence each observed value has on each fitted value. The diagonal elements of the  $H$  are the ‘leverages’, which describe the influence each observed value has on the fitted value for that same observation. The residuals ( $R$ ) are related to the observed values by the following formula:

$$R = (I - H)Y \quad (11)$$

The variances of  $Y$  and  $R$  can be expressed as:

$$\begin{aligned} \text{var}(Y) &= H\sigma^2 \\ \text{var}(R) &= (I - H)\sigma^2 \end{aligned} \quad (12)$$

Updating techniques allow an economic approach to recalculating the projection matrix,  $H$ , by removing the necessity to refit the model each time it is updated. However this approach is known for numerical instability in the case of down-dating.

sectionThe Hat Matrix

The projection matrix  $H$  (also known as the hat matrix), is a well known identity that maps the fitted values  $\hat{Y}$  to the observed values  $Y$ , i.e.  $\hat{Y} = HY$ .

$$H = X(X^T X)^{-1} X^T \quad (13)$$

$H$  describes the influence each observed value has on each fitted value. The diagonal elements of the  $H$  are the ‘leverages’, which describe the influence each observed value has on the fitted value for that same observation. The residuals ( $R$ ) are related to the observed values by the following formula:

$$R = (I - H)Y \quad (14)$$

The variances of  $Y$  and  $R$  can be expressed as:

$$\begin{aligned}\text{var}(Y) &= H\sigma^2 \\ \text{var}(R) &= (I - H)\sigma^2\end{aligned}\tag{15}$$

Updating techniques allow an economic approach to recalculating the projection matrix,  $H$ , by removing the necessity to refit the model each time it is updated. However this approach is known for numerical instability in the case of down-dating.

## 24 Sherman Morrison Woodbury Formula

The ‘Sherman Morrison Woodbury’ Formula is a well known result in linear algebra;

$$(A + a^T B)^{-1} = A^{-1} - A^{-1} a^T (I - b A^{-1} a^T)^{-1} b A^{-1} \quad (16)$$

This result is highly useful for analyzing regression diagnostics, and for matrices inverses in general. Consider a  $p \times p$  matrix  $X$ , from which a row  $x_i^T$  is to be added or deleted. ? sets  $A = X^T X$ ,  $a = -x_i^T$  and  $b = x_i^T$ , and writes the above equation as

$$(X^T X \pm x_i x_i^T)^{-1} = (X^T X)^{-1} \mp \frac{(X^T X)^{-1} (x_i x_i^T (X^T X)^{-1})}{1 - x_i^T (X^T X)^{-1} x_i} \quad (17)$$

The projection matrix  $H$  (also known as the hat matrix), is a well known identity that maps the fitted values  $\hat{Y}$  to the observed values  $Y$ , i.e.  $\hat{Y} = HY$ .

$$H = X(X^T X)^{-1} X^T \quad (18)$$

$H$  describes the influence each observed value has on each fitted value. The diagonal elements of the  $H$  are the ‘leverages’, which describe the influence each observed value has on the fitted value for that same observation. The residuals ( $R$ ) are related to the observed values by the following formula:

$$R = (I - H)Y \quad (19)$$

The variances of  $Y$  and  $R$  can be expressed as:

$$\begin{aligned} \text{var}(Y) &= H\sigma^2 \\ \text{var}(R) &= (I - H)\sigma^2 \end{aligned} \quad (20)$$

Updating techniques allow an economic approach to recalculating the projection matrix,  $H$ , by removing the necessity to refit the model each time it is updated. However this approach is known for numerical instability in the case of down-dating.

### 24.1 Hat Values for MCS regression

With  $A$  as the averages and  $D$  as the casewise differences.

`fit = lm(D~A)`

$$H = A (A^T A)^{-1} A^T,$$

## 25 Cross Validation

Cross validation techniques for linear regression employ the use ‘leave one out’ recalculations. In such procedures the regression coefficients are estimated for  $n - 1$  covariates, with the  $Q^{th}$  observation omitted.

Let  $\hat{\beta}$  denote the least square estimate of  $\beta$  based upon the full set of observations, and let  $\hat{\beta}^{-Q}$  denoted the estimate with the  $Q^{th}$  case excluded.

In leave-one-out cross validation, each observation is omitted in turn, and a regression model is fitted on the rest of the data. Cross validation is used to estimate the generalization error of a given model. alternatively it can be used for model selection by determining the candidate model that has the smallest generalization error.

Evidently leave-one-out cross validation has similarities with ‘jackknifing’, a well known statistical technique. However cross validation is used to estimate generalization error, whereas the jackknife technique is used to estimate bias.

subsectionCross Validation: Updating standard deviation

The variance of a data set can be calculated using the following formula.

$$S^2 = \frac{\sum_{i=1}^n (x_i^2) - \frac{(\sum_{i=1}^n x_i)^2}{n}}{n - 1} \quad (21)$$

While using bivariate data, the notation  $Sxx$  and  $Syy$  shall apply to the variance of  $x$  and of  $y$  respectively. The covariance term  $Sxy$  is given by

$$Sxy = \frac{\sum_{i=1}^n (x_i y_i) - \frac{(\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n}}{n - 1} \quad (22)$$

Let the observation  $j$  be omitted from the data set. The estimates for the variance identities can be updating using minor adjustments to the full sample estimates. Where  $(j)$  denotes that the  $j$ th has been omitted, these identities are

$$Sxx^{(j)} = \frac{\sum_{i=1}^n (x_i^2) - (x_j)^2 - \frac{((\sum_{i=1}^n x_i) - x_j)^2}{n-1}}{n - 2} \quad (23)$$

$$Syy^{(j)} = \frac{\sum_{i=1}^n (y_i^2) - (y_j)^2 - \frac{((\sum_{i=1}^n y_i) - y_j)^2}{n-1}}{n - 2} \quad (24)$$

$$Sxy^{(j)} = \frac{\sum_{i=1}^n (x_i y_i) - (y_j x_j) - \frac{((\sum_{i=1}^n x_i) - x_j)((\sum_{i=1}^n y_i) - y_j)}{n-1}}{n - 2} \quad (25)$$

The updated estimate for the slope is therefore

$$\hat{\beta}_1^{(j)} = \frac{Sxy^{(j)}}{Sxx^{(j)}} \quad (26)$$

It is necessary to determine the mean for  $x$  and  $y$  of the remaining  $n - 1$  terms

$$\bar{x}^{(j)} = \frac{(\sum_{i=1}^n x_i) - (x_j)}{n - 1}, \quad (27)$$

$$\bar{y}^{(j)} = \frac{(\sum_{i=1}^n y_i) - (y_j)}{n - 1}. \quad (28)$$

The updated intercept estimate is therefore

$$\hat{\beta}_0^{(j)} = \bar{y}^{(j)} - \hat{\beta}_1^{(j)} \bar{x}^{(j)}. \quad (29)$$

subsectionCross Validation: Updating standard deviation

The variance of a data set can be calculated using the following formula.

$$S^2 = \frac{\sum_{i=1}^n (x_i^2) - \frac{(\sum_{i=1}^n x_i)^2}{n}}{n-1} \quad (30)$$

While using bivariate data, the notation  $Sxx$  and  $Syy$  shall apply to the variance of  $x$  and of  $y$  respectively. The covariance term  $Sxy$  is given by

$$Sxy = \frac{\sum_{i=1}^n (x_i y_i) - \frac{(\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n}}{n-1} \quad (31)$$

Let the observation  $j$  be omitted from the data set. The estimates for the variance identities can be updating using minor adjustments to the full sample estimates. Where  $(j)$  denotes that the  $j$ th has been omitted, these identities are

$$Sxx^{(j)} = \frac{\sum_{i=1}^n (x_i^2) - (x_j)^2 - \frac{((\sum_{i=1}^n x_i) - x_j)^2}{n-1}}{n-2} \quad (32)$$

$$Syy^{(j)} = \frac{\sum_{i=1}^n (y_i^2) - (y_j)^2 - \frac{((\sum_{i=1}^n y_i) - y_j)^2}{n-1}}{n-2} \quad (33)$$

$$Sxy^{(j)} = \frac{\sum_{i=1}^n (x_i y_i) - (y_j x_j) - \frac{((\sum_{i=1}^n x_i) - x_j)((\sum_{i=1}^n y_i) - y_j)}{n-1}}{n-2} \quad (34)$$

The updated estimate for the slope is therefore

$$\hat{\beta}_1^{(j)} = \frac{Sxy^{(j)}}{Sxx^{(j)}} \quad (35)$$

It is necessary to determine the mean for  $x$  and  $y$  of the remaining  $n-1$  terms

$$\bar{x}^{(j)} = \frac{(\sum_{i=1}^n x_i) - (x_j)}{n-1}, \quad (36)$$

$$\bar{y}^{(j)} = \frac{(\sum_{i=1}^n y_i) - (y_j)}{n-1}. \quad (37)$$

The updated intercept estimate is therefore

$$\hat{\beta}_0^{(j)} = \bar{y}^{(j)} - \hat{\beta}_1^{(j)} \bar{x}^{(j)}. \quad (38)$$



## 25.1 Cross Validation: Updating standard deviation

The variance of a data set can be calculated using the following formula.

$$S^2 = \frac{\sum_{i=1}^n (x_i^2) - \frac{(\sum_{i=1}^n x_i)^2}{n}}{n-1} \quad (39)$$

While using bivariate data, the notation  $Sxx$  and  $Syy$  shall apply to the variance of  $x$  and of  $y$  respectively. The covariance term  $Sxy$  is given by

$$Sxy = \frac{\sum_{i=1}^n (x_i y_i) - \frac{(\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n}}{n-1} \quad (40)$$

Let the observation  $j$  be omitted from the data set. The estimates for the variance identities can be updating using minor adjustments to the full sample estimates. Where  $(j)$  denotes that the  $j$ th has been omitted, these identities are

$$Sxx^{(j)} = \frac{\sum_{i=1}^n (x_i^2) - (x_j)^2 - \frac{((\sum_{i=1}^n x_i) - x_j)^2}{n-1}}{n-2} \quad (41)$$

$$Syy^{(j)} = \frac{\sum_{i=1}^n (y_i^2) - (y_j)^2 - \frac{((\sum_{i=1}^n y_i) - y_j)^2}{n-1}}{n-2} \quad (42)$$

$$Sxy^{(j)} = \frac{\sum_{i=1}^n (x_i y_i) - (y_j x_j) - \frac{((\sum_{i=1}^n x_i) - x_j)((\sum_{i=1}^n y_i) - y_j)}{n-1}}{n-2} \quad (43)$$

The updated estimate for the slope is therefore

$$\hat{\beta}_1^{(j)} = \frac{Sxy^{(j)}}{Sxx^{(j)}} \quad (44)$$

It is necessary to determine the mean for  $x$  and  $y$  of the remaining  $n-1$  terms

$$\bar{x}^{(j)} = \frac{(\sum_{i=1}^n x_i) - (x_j)}{n-1}, \quad (45)$$

$$\bar{y}^{(j)} = \frac{(\sum_{i=1}^n y_i) - (y_j)}{n-1}. \quad (46)$$

The updated intercept estimate is therefore

$$\hat{\beta}_0^{(j)} = \bar{y}^{(j)} - \hat{\beta}_1^{(j)} \bar{x}^{(j)}. \quad (47)$$

## 26 Updating Estimates

### 26.1 Updating Standard deviation

A simple, but useful, example of updating is the updating of the standard deviation when an observation is omitted, as practised in statistical process control analyzes. From first principles, the variance of a data set can be calculated using the following formula.

$$S^2 = \frac{\sum_{i=1}^n (x_i^2) - \frac{(\sum_{i=1}^n x_i)^2}{n}}{n-1} \quad (48)$$

While using bivariate data, the notation  $Sxx$  and  $Syy$  shall apply hither to the variance of  $x$  and of  $y$  respectively. The covariance term  $Sxy$  is given by

$$Sxy = \frac{\sum_{i=1}^n (x_i y_i) - \frac{(\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n}}{n-1}. \quad (49)$$

### 26.2 Updating of Regression Estimates

Updating techniques are used in regression analysis to add or delete rows from a model, allowing the analyst the effect of the observation associated with that row. In time series problems, there will be scientific interest in the changing relationship between variables. In cases where there a single row is to be added or deleted, the procedure used is equivalent to a geometric rotation of a plane.

Consider a  $p \times p$  matrix  $X$ , from which a row  $x_i^T$  is to be added or deleted. ? sets  $A = X^T X$ ,  $a = -x_i^T$  and  $b = x_i^T$ , and writes the above equation as

$$(X^T X \pm x_i x_i^T)^{-1} = (X^T X)^{-1} \mp \frac{(X^T X)^{-1} (x_i x_i^T (X^T X)^{-1})}{1 - x_i^T (X^T X)^{-1} x_i} \quad (50)$$

### 26.3 Updating Regression Estimates

Let the observation  $j$  be omitted from the data set. The estimates for the variance identities can be updating using minor adjustments to the full sample estimates. Where  $(j)$  denotes that the  $j$ th has been omitted, these identities are

$$Sxx^{(j)} = \frac{\sum_{i=1}^n (x_i^2) - (x_j)^2 - \frac{((\sum_{i=1}^n x_i) - x_j)^2}{n-1}}{n-2} \quad (51)$$

$$Syy^{(j)} = \frac{\sum_{i=1}^n (y_i^2) - (y_j)^2 - \frac{((\sum_{i=1}^n y_i) - y_j)^2}{n-1}}{n-2} \quad (52)$$

$$Sxy^{(j)} = \frac{\sum_{i=1}^n (x_i y_i) - (y_j x_j) - \frac{((\sum_{i=1}^n x_i) - x_j)((\sum_{i=1}^n y_i) - y_j)}{n-1}}{n-2} \quad (53)$$

The updated estimate for the slope is therefore

$$\hat{\beta}_1^{(j)} = \frac{Sxy^{(j)}}{Sxx^{(j)}} \quad (54)$$

It is necessary to determine the mean for  $x$  and  $y$  of the remaining  $n - 1$  terms

$$\bar{x}^{(j)} = \frac{(\sum_{i=1}^n x_i) - (x_j)}{n-1}, \quad (55)$$

$$\bar{y}^{(j)} = \frac{(\sum_{i=1}^n y_i) - (y_j)}{n-1}. \quad (56)$$

The updated intercept estimate is therefore

$$\hat{\beta}_0^{(j)} = \bar{y}^{(j)} - \hat{\beta}_1^{(j)} \bar{x}^{(j)}. \quad (57)$$

## 26.4 Inference on intercept and slope

$$\hat{\beta}_1 \pm t_{(\alpha, n-2)} \sqrt{\frac{S^2}{(n-1)S_x^2}} \quad (58)$$

$$\frac{\hat{\beta}_0 - \beta_0}{SE(\hat{\beta}_0)} \quad (59)$$

$$\frac{\hat{\beta}_1 - \beta_1}{SE(\hat{\beta}_0)} \quad (60)$$

### 26.4.1 Inference on correlation coefficient

This test of the slope is coincidentally the equivalent of a test of the correlation of the  $n$  observations of  $X$  and  $Y$ .

$$H_0 : \rho_{XY} = 0$$

$$H_A : \rho_{XY} \neq 0$$

(61)

Appendices

sectionThe Hat Matrix

The projection matrix  $H$  (also known as the hat matrix), is a well known identity that maps the fitted values  $\hat{Y}$  to the observed values  $Y$ , i.e.  $\hat{Y} = HY$ .

$$H = X(X^T X)^{-1} X^T \quad (62)$$

$H$  describes the influence each observed value has on each fitted value. The diagonal elements of the  $H$  are the ‘leverages’, which describe the influence each observed value has on the fitted value for that same observation. The residuals ( $R$ ) are related to the observed values by the following formula:

$$R = (I - H)Y \quad (63)$$

The variances of  $Y$  and  $R$  can be expressed as:

$$\begin{aligned} \text{var}(Y) &= H\sigma^2 \\ \text{var}(R) &= (I - H)\sigma^2 \end{aligned} \quad (64)$$

Updating techniques allow an economic approach to recalculating the projection matrix,  $H$ , by removing the necessity to refit the model each time it is updated. However this approach is known for numerical instability in the case of down-dating.

## 27 Lesaffre’s paper.

Lesaffre considers the case-weight perturbation approach.

Cook’s 86 describes a local approach wherein each case is given a weight  $w_i$  and the effect on the parameter estimation is measured by perturbing these weights. Choosing weights close to zero or one corresponds to the global case-deletion approach.

Lesaffre describes the displacement in log-likelihood as a useful metric to evaluate local influence

Lesaffre describes a framework to detect outlying observations that matter in an LME model. Detection should be carried out by evaluating diagnostics  $C_i$ ,  $C_i(\alpha)$  and  $C_i(D, \sigma^2)$ .

Lesaffre defines the total local influence of individual  $i$  as

$$C_i = 2|\Delta \boldsymbol{\mu}_i L^{-1} \Delta \boldsymbol{\mu}_i|. \quad (65)$$

The influence function of the MLEs evaluated at the  $i$ th point  $IF_i$ , given by

$$IF_i = -L^{-1}\Delta_i \tag{66}$$

can indicate how  $\hat{\theta}$  changes as the weight of the  $i$ th subject changes.

The manner by which influential observations distort the estimation process can be determined by inspecting the interpretable components in the decomposition of the above measures of local influence.

Lesaffre comments that there is no clear way of interpreting the information contained in the angles, but that this doesn't mean the information should be ignored.

Appendices

## 28 Cross Validation

Cross validation techniques for linear regression employ the use ‘leave one out’ recalculations. In such procedures the regression coefficients are estimated for  $n - 1$  covariates, with the  $Q^{th}$  observation omitted.

Let  $\hat{\beta}$  denote the least square estimate of  $\beta$  based upon the full set of observations, and let  $\hat{\beta}^{-Q}$  denoted the estimate with the  $Q^{th}$  case excluded.

In leave-one-out cross validation, each observation is omitted in turn, and a regression model is fitted on the rest of the data. Cross validation is used to estimate the generalization error of a given model. alternatively it can be used for model selection by determining the candidate model that has the smallest generalization error.

Evidently leave-one-out cross validation has similarities with ‘jackknifing’, a well known statistical technique. However cross validation is used to estimate generalization error, whereas the jackknife technique is used to estimate bias.

## 29 Updating Estimates

### 29.1 Updating of Regression Estimates

Updating techniques are used in regression analysis to add or delete rows from a model, allowing the analyst the effect of the observation associated with that row. In time series problems, there will be scientific interest in the changing relationship between variables. In cases where there a single row is to be added or deleted, the procedure used is equivalent to a geometric rotation of a plane.

Updating techniques are used in regression analysis to add or delete rows from a model, allowing the analyst the effect of the observation associated with that row.

### 29.2 Updating Standard deviation

A simple, but useful, example of updating is the updating of the standard deviation when an observation is omitted, as practised in statistical process control analyzes. From first principles, the variance of a data set can be calculated using the following formula.

$$S^2 = \frac{\sum_{i=1}^n (x_i^2) - \frac{(\sum_{i=1}^n x_i)^2}{n}}{n - 1} \quad (67)$$

While using bivariate data, the notation  $Sxx$  and  $Syy$  shall apply hither to the variance of  $x$  and of  $y$  respectively. The covariance term  $Sxy$  is given by

$$Sxy = \frac{\sum_{i=1}^n (x_i y_i) - \frac{(\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n}}{n-1}. \quad (68)$$

### 29.3 Updating of Regression Estimates

Updating techniques are used in regression analysis to add or delete rows from a model, allowing the analyst the effect of the observation associated with that row. In time series problems, there will be scientific interest in the changing relationship between variables. In cases where there a single row is to be added or deleted, the procedure used is equivalent to a geometric rotation of a plane.

Consider a  $p \times p$  matrix  $X$ , from which a row  $x_i^T$  is to be added or deleted. ? sets  $A = X^T X$ ,  $a = -x_i^T$  and  $b = x_i^T$ , and writes the above equation as

$$(X^T X \pm x_i x_i^T)^{-1} = (X^T X)^{-1} \mp \frac{(X^T X)^{-1} (x_i x_i^T (X^T X)^{-1})}{1 - x_i^T (X^T X)^{-1} x_i} \quad (69)$$

### 29.4 Updating Regression Estimates

Let the observation  $j$  be omitted from the data set. The estimates for the variance identities can be updating using minor adjustments to the full sample estimates. Where  $(j)$  denotes that the  $j$ th has been omitted, these identities are

$$Sxx^{(j)} = \frac{\sum_{i=1}^n (x_i^2) - (x_j)^2 - \frac{((\sum_{i=1}^n x_i) - x_j)^2}{n-1}}{n-2} \quad (70)$$

$$Syy^{(j)} = \frac{\sum_{i=1}^n (y_i^2) - (y_j)^2 - \frac{((\sum_{i=1}^n y_i) - y_j)^2}{n-1}}{n-2} \quad (71)$$

$$Sxy^{(j)} = \frac{\sum_{i=1}^n (x_i y_i) - (y_j x_j) - \frac{((\sum_{i=1}^n x_i) - x_j)((\sum_{i=1}^n y_i) - y_j)}{n-1}}{n-2} \quad (72)$$

The updated estimate for the slope is therefore

$$\hat{\beta}_1^{(j)} = \frac{Sxy^{(j)}}{Sxx^{(j)}} \quad (73)$$

It is necessary to determine the mean for  $x$  and  $y$  of the remaining  $n-1$  terms

$$\bar{x}^{(j)} = \frac{(\sum_{i=1}^n x_i) - (x_j)}{n-1}, \quad (74)$$

$$\bar{y}^{(j)} = \frac{(\sum_{i=1}^n y_i) - (y_j)}{n - 1}. \quad (75)$$

The updated intercept estimate is therefore

$$\hat{\beta}_0^{(j)} = \bar{y}^{(j)} - \hat{\beta}_1^{(j)} \bar{x}^{(j)}. \quad (76)$$

## 29.5 Inference on intercept and slope

$$\hat{\beta}_1 \pm t_{(\alpha, n-2)} \sqrt{\frac{S^2}{(n-1)S_x^2}} \quad (77)$$

$$\frac{\hat{\beta}_0 - \beta_0}{SE(\hat{\beta}_0)} \quad (78)$$

$$\frac{\hat{\beta}_1 - \beta_1}{SE(\hat{\beta}_1)} \quad (79)$$

### 29.5.1 Inference on correlation coefficient

This test of the slope is coincidentally the equivalent of a test of the correlation of the  $n$  observations of  $X$  and  $Y$ .

$$H_0 : \rho_{XY} = 0$$

$$H_A : \rho_{XY} \neq 0$$

(80)



## 29.6 Updating Regression Estimates

Let the observation  $j$  be omitted from the data set. The estimates for the variance identities can be updating using minor adjustments to the full sample estimates. Where  $(j)$  denotes that the  $j$ th has been omitted, these identities are

$$Sxx^{(j)} = \frac{\sum_{i=1}^n (x_i^2) - (x_j)^2 - \frac{((\sum_{i=1}^n x_i) - x_j)^2}{n-1}}{n-2} \quad (81)$$

$$Syy^{(j)} = \frac{\sum_{i=1}^n (y_i^2) - (y_j)^2 - \frac{((\sum_{i=1}^n y_i) - y_j)^2}{n-1}}{n-2} \quad (82)$$

$$Sxy^{(j)} = \frac{\sum_{i=1}^n (x_i y_i) - (y_j x_j) - \frac{((\sum_{i=1}^n x_i) - x_j)((\sum_{i=1}^n y_i) - y_j)}{n-1}}{n-2} \quad (83)$$

The updated estimate for the slope is therefore

$$\hat{\beta}_1^{(j)} = \frac{Sxy^{(j)}}{Sxx^{(j)}} \quad (84)$$

It is necessary to determine the mean for  $x$  and  $y$  of the remaining  $n-1$  terms

$$\bar{x}^{(j)} = \frac{(\sum_{i=1}^n x_i) - (x_j)}{n-1}, \quad (85)$$

$$\bar{y}^{(j)} = \frac{(\sum_{i=1}^n y_i) - (y_j)}{n-1}. \quad (86)$$

The updated intercept estimate is therefore

$$\hat{\beta}_0^{(j)} = \bar{y}^{(j)} - \hat{\beta}_1^{(j)} \bar{x}^{(j)}. \quad (87)$$

## 29.7 Updating of Regression Estimates

Updating techniques are used in regression analysis to add or delete rows from a model, allowing the analyst the effect of the observation associated with that row. In time series problems, there will be scientific interest in the changing relationship between variables. In cases where there a single row is to be added or deleted, the procedure used is equivalent to a geometric rotation of a plane.

Consider a  $p \times p$  matrix  $X$ , from which a row  $x_i^T$  is to be added or deleted. ? sets  $A = X^T X$ ,  $a = -x_i^T$  and  $b = x_i^T$ , and writes the above equation as

$$(X^T X \pm x_i x_i^T)^{-1} = (X^T X)^{-1} \mp \frac{(X^T X)^{-1} (x_i x_i^T (X^T X)^{-1})}{1 - x_i^T (X^T X)^{-1} x_i} \quad (88)$$

## 29.8 Updating Standard deviation

A simple, but useful, example of updating is the updating of the standard deviation when an observation is omitted, as practised in statistical process control analyzes. From first principles, the variance of a data set can be calculated using the following formula.

$$S^2 = \frac{\sum_{i=1}^n (x_i^2) - \frac{(\sum_{i=1}^n x_i)^2}{n}}{n - 1} \quad (89)$$

While using bivariate data, the notation  $Sxx$  and  $Syy$  shall apply hither to the variance of  $x$  and of  $y$  respectively. The covariance term  $Sxy$  is given by

$$Sxy = \frac{\sum_{i=1}^n (x_i y_i) - \frac{(\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n}}{n - 1}. \quad (90)$$

## 29.9 Inference on intercept and slope

$$\hat{\beta}_1 \pm t_{(\alpha, n-2)} \sqrt{\frac{S^2}{(n-1)S_x^2}} \quad (91)$$

$$\frac{\hat{\beta}_0 - \beta_0}{SE(\hat{\beta}_0)} \quad (92)$$

$$\frac{\hat{\beta}_1 - \beta_1}{SE(\hat{\beta}_1)} \quad (93)$$

### 29.9.1 Inference on correlation coefficient

This test of the slope is coincidentally the equivalent of a test of the correlation of the  $n$  observations of  $X$  and  $Y$ .

$$\begin{aligned} H_0 : \rho_{XY} &= 0 \\ H_A : \rho_{XY} &\neq 0 \end{aligned} \quad (94)$$

## References

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

- 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. and K. Hayes (1998). Residuals for the linear model with general covariance structure. *Journal of the Royal Statistical Society (Series B)* 60, 201–215.
- Preisser, J. S. (1996). Deletion diagnostics for generalised estimating equations. *Biometrika* 83(3), 551–556.
- 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.