

Chapter 1

Updating Formulae

1.1 subset deletion

Haslett and Dillane (2004) said this

1.2 Marginal Residuals

$$\begin{aligned}\hat{\beta} &= (X^T R^{-1} X)^{-1} X^T R^{-1} Y \\ &= BY\end{aligned}$$

1.3 DFBETA

$$DFBETA_a = \hat{\beta} - \hat{\beta}_{(a)} \tag{1.1}$$

$$= B(Y - Y_{\bar{a}}) \tag{1.2}$$

1.4 Deletion Diagnostics

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

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

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

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

Chapter 2

Updating Techniques and Cross Validation

2.1 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.

2.1.1 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.

$$S^2 = \frac{\sum_{i=1}^n (x_i^2) - \frac{(\sum_{i=1}^n x_i)^2}{n}}{n - 1} \quad (2.1)$$

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

$$Sxy = \frac{\sum_{i=1}^n (x_i y_i) - \frac{(\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n}}{n-1}. \quad (2.2)$$

2.1.2 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

$$Sxx^{(j)} = \frac{\sum_{i=1}^n (x_i^2) - (x_j)^2 - \frac{((\sum_{i=1}^n x_i) - x_j)^2}{n-1}}{n-2} \quad (2.3)$$

$$Syy^{(j)} = \frac{\sum_{i=1}^n (y_i^2) - (y_j)^2 - \frac{((\sum_{i=1}^n y_i) - y_j)^2}{n-1}}{n-2} \quad (2.4)$$

$$Sxy^{(j)} = \frac{\sum_{i=1}^n (x_i y_i) - (y_j x_j) - \frac{((\sum_{i=1}^n x_i) - x_j)((\sum_{i=1}^n y_i) - y_j)}{n-1}}{n-2} \quad (2.5)$$

The updated estimate for the slope is therefore

$$\hat{\beta}_1^{(j)} = \frac{Sxy^{(j)}}{Sxx^{(j)}} \quad (2.6)$$

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

$$\bar{x}^{(j)} = \frac{(\sum_{i=1}^n x_i) - (x_j)}{n-1}, \quad (2.7)$$

$$\bar{y}^{(j)} = \frac{(\sum_{i=1}^n y_i) - (y_j)}{n-1}. \quad (2.8)$$

The updated intercept estimate is therefore

$$\hat{\beta}_0^{(j)} = \bar{y}^{(j)} - \hat{\beta}_1^{(j)} \bar{x}^{(j)}. \quad (2.9)$$

2.1.3 Inference on intercept and slope

$$\hat{\beta}_1 \pm t_{(\alpha, n-2)} \sqrt{\frac{S^2}{(n-1)S_x^2}} \quad (2.10)$$

$$\frac{\hat{\beta}_0 - \beta_0}{SE(\hat{\beta}_0)} \quad (2.11)$$

$$\frac{\hat{\beta}_1 - \beta_1}{SE(\hat{\beta}_1)} \quad (2.12)$$

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 .

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

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

(2.13)

2.2 Sherman Morrison Woodbury Formula

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

$$(A + a^T B)^{-1} = A^{-1} - A^{-1} a^T (I - b A^{-1} a^T)^{-1} b A^{-1} \quad (2.14)$$

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

$$(X^T X \pm x_i x_i^T)^{-1} = (X^T X)^{-1} \mp \frac{(X^T X)^{-1} (x_i x_i^T (X^T X)^{-1})}{1 - x_i^T (X^T X)^{-1} x_i} \quad (2.15)$$

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. $\hat{Y} = HY$.

$$H = X(X^T X)^{-1} X^T \quad (2.16)$$

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:

$$R = (I - H)Y \quad (2.17)$$

The variances of Y and R can be expressed as:

$$\begin{aligned} \text{var}(Y) &= H\sigma^2 \\ \text{var}(R) &= (I - H)\sigma^2 \end{aligned} \quad (2.18)$$

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.

2.3 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.

2.3.1 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.

2.3.2 DFFITS

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

2.3.3 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.

$$PRESS = \sum (y - y^{(k)})^2 \tag{2.19}$$

2.3.4 Delete = Replace

2.4 Application to MCS

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

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

$$Y^{(k)} = \hat{\beta}^{(k)} X^{(k)} \quad (2.20)$$

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

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

```
Coefficients: (Intercept)          A
-37.51896      0.04656
```

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

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

$$D^{-Q} = \hat{\beta}^{-Q} A^{-Q} \quad (2.21)$$

2.5 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. $\hat{Y} = HY$.

$$H = X(X^T X)^{-1} X^T \quad (2.22)$$

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:

$$R = (I - H)Y \quad (2.23)$$

The variances of Y and R can be expressed as:

$$\begin{aligned} \text{var}(Y) &= H\sigma^2 \\ \text{var}(R) &= (I - H)\sigma^2 \end{aligned} \tag{2.24}$$

2.6 Hat Values for MCS regression

```
fit = lm(D~A)
```

$$H = A \left(A^\top A \right)^{-1} A^\top,$$

2.6.1 Influence measures using R

R provides the following influence measures of each observation.

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

Chapter 3

More on Linear Mixed Effects Models

3.1 Augmented GLMs

Generalized linear models are a generalization of classical linear models.

3.1.1 Augmented linear model

The subscript M is a label referring to the mean model.

$$\begin{pmatrix} Y \\ \psi_M \end{pmatrix} = \begin{pmatrix} X & Z \\ 0 & I \end{pmatrix} \begin{pmatrix} \beta \\ \nu \end{pmatrix} + e^* \quad (3.1)$$

The error term e^* is normal with mean zero. The variance matrix of the error term is given by

$$\Sigma_a = \begin{pmatrix} \Sigma & 0 \\ 0 & D \end{pmatrix}. \quad (3.2)$$

$$X = \begin{pmatrix} T & Z \\ 0 & I \end{pmatrix} \delta = \begin{pmatrix} \beta \\ \nu \end{pmatrix} \quad (3.3)$$

$$y_a = T\delta + e^* \tag{3.4}$$

Weighted least squares equation

3.2 Model Validation

Three basic approaches are described by Neter et al

1. Collection of new data to check the model
2. Comparison of results with theoretical expectations
3. use of a ‘hold out sample’ to check the model and its predictive capability.

3.2.1 Mean Square Prediction Error

$$MSPR = \frac{\sum (y_i - \hat{y}_i)^2}{n^*} \quad (3.5)$$

3.3 Leave one out

$\hat{y}_{\bar{i}}$ is the fitted value for the i th observation based on all y values but the i th observation.

$$\beta_{(-i)}^{\wedge}$$

$$\alpha_{(-i)}^{\wedge}$$

3.4 Leave k out

3.5 Cross Validation Techniques

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

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

$$Y^{-Q} = \hat{\beta}^{-Q} X^{-Q} \quad (3.6)$$

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

$$D^{-Q} = \hat{\beta}^{-Q} A^{-Q} \quad (3.7)$$

3.5.1 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.

$$PRESS = \sum (y - y^{-Q})^2 \quad (3.8)$$

3.5.2 PRESS

$$e_{-Q} = y_Q - x_Q \hat{\beta}^{-Q} \quad (3.9)$$

3.6 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.

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

$$(A + a^T B)^{-1} = A^{-1} - A^{-1} a^T (I - b A^{-1} a^T)^{-1} b A^{-1} \quad (3.10)$$

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

$$(X^T X \pm x_i x_i^T)^{-1} = (X^T X)^{-1} \mp \frac{(X^T X)^{-1} (x_i x_i^T (X^T X)^{-1})}{1 - x_i^T (X^T X)^{-1} x_i} \quad (3.11)$$

The projection matrix v is a

This approach allows an economic approach to recalculating the projection matrix, V , by removing the necessity to refit the model each time it is updated.

This approach is known for numerical instability in the case of downdating.

3.7 Influence Measures

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

3.7.1 Subset Deletion

3.8 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 k^{th} observation omitted.

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.

3.9 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

$$C_i = 2|\Delta_i' L^{-1} \Delta_i|. \quad (3.12)$$

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

$$IF_i = -L^{-1} \Delta_i \quad (3.13)$$

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.

3.10 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.

3.10.1 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.

$$S^2 = \frac{\sum_{i=1}^n (x_i^2) - \frac{(\sum_{i=1}^n x_i)^2}{n}}{n - 1} \quad (3.14)$$

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

$$Sxy = \frac{\sum_{i=1}^n (x_i y_i) - \frac{(\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n}}{n - 1}. \quad (3.15)$$

This

3.10.2 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

$$Sxx^{(j)} = \frac{\sum_{i=1}^n (x_i^2) - (x_j)^2 - \frac{((\sum_{i=1}^n x_i) - x_j)^2}{n-1}}{n-2} \quad (3.16)$$

$$Syy^{(j)} = \frac{\sum_{i=1}^n (y_i^2) - (y_j)^2 - \frac{((\sum_{i=1}^n y_i) - y_j)^2}{n-1}}{n-2} \quad (3.17)$$

$$Sxy^{(j)} = \frac{\sum_{i=1}^n (x_i y_i) - (y_j x_j) - \frac{((\sum_{i=1}^n x_i) - x_j)((\sum_{i=1}^n y_i) - y_j)}{n-1}}{n-2} \quad (3.18)$$

The updated estimate for the slope is therefore

$$\hat{\beta}_1^{(j)} = \frac{Sxy^{(j)}}{Sxx^{(j)}} \quad (3.19)$$

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

$$\bar{x}^{(j)} = \frac{(\sum_{i=1}^n x_i) - (x_j)}{n-1}, \quad (3.20)$$

$$\bar{y}^{(j)} = \frac{(\sum_{i=1}^n y_i) - (y_j)}{n-1}. \quad (3.21)$$

The updated intercept estimate is therefore

$$\hat{\beta}_0^{(j)} = \bar{y}^{(j)} - \hat{\beta}_1^{(j)} \bar{x}^{(j)}. \quad (3.22)$$

3.10.3 Inference on intercept and slope

$$\hat{\beta}_1 \pm t_{(\alpha, n-2)} \sqrt{\frac{S^2}{(n-1)S_x^2}} \quad (3.23)$$

$$\frac{\hat{\beta}_0 - \beta_0}{SE(\hat{\beta}_0)} \quad (3.24)$$

$$\frac{\hat{\beta}_1 - \beta_1}{SE(\hat{\beta}_0)} \quad (3.25)$$

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 .

$$\begin{aligned}H_0 : \rho_{XY} &= 0 \\H_A : \rho_{XY} &\neq 0\end{aligned}\tag{3.26}$$

3.11 Sherman Morrison Woodbury Formula

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

$$(A + a^T B)^{-1} = A^{-1} - A^{-1} a^T (I - b A^{-1} a^T)^{-1} b A^{-1}\tag{3.27}$$

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

$$(X^T X \pm x_i x_i^T)^{-1} = (X^T X)^{-1} \mp \frac{(X^T X)^{-1} (x_i x_i^T (X^T X)^{-1})}{1 - x_i^T (X^T X)^{-1} x_i}\tag{3.28}$$

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. $\hat{Y} = HY$.

$$H = X(X^T X)^{-1} X^T\tag{3.29}$$

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:

$$R = (I - H)Y \quad (3.30)$$

The variances of Y and R can be expressed as:

$$\begin{aligned} \text{var}(Y) &= H\sigma^2 \\ \text{var}(R) &= (I - H)\sigma^2 \end{aligned} \quad (3.31)$$

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.

3.12 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.

3.12.1 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.

3.12.2 DFFITS

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

3.12.3 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.

$$PRESS = \sum (y - y^{(k)})^2 \tag{3.32}$$

3.13 Case Deletion Diagnostics

What is a perturbation? PRESS?

3.14 mixed models

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

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

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

Cook (1986) introduces powerful tools for local-influence assessment and examining perturbations in the assumptions of a model. In particular the effect of local perturbations of parameters or observations are examined.

3.15 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

3.15.1 Delete = Replace

3.16 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.

3.17 Case-Deletion results for Fixed effects

3.18 Case-Deletion results for Variance components

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

3.19 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.

Christensen developed their global influences for the deletion of single observations in two steps: a one-step estimate for the REML (or ML) estimate of the variance components, and an ordinary case-deletion diagnostic for a weighted resgression problem (conditional on the estimated covariance matrix) for fixed effects. Lesaffre's approach accords with that proposed by Christensen et al when applied in a repeated measure-

ment context, with a large sample size.

3.20 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.

3.21 estimation

$$\hat{\beta} = X^T \quad (3.33)$$

$$\hat{\gamma} = G(\hat{\theta})Z^T \quad (3.34)$$

The difference between perturbation and residual analysis between the linear and LME models

The estimates of the fixed effects β depend on the estimates of the covariance parameters.

3.22 Marginal and Conditional Residuals

$$r_{mi} = x_i^T \hat{\beta} \quad (3.35)$$

3.23 Likelihood Distance

If the global measure suggests that the points in U are influential

3.24 Effects on fitted and predicted values

$$\hat{e}_{i(U)} = y_i - x\hat{\beta}_{(U)} \quad (3.36)$$

3.25 PRESS

$$PRESS_{(U)} = y_i - x\hat{\beta}_{(U)} \quad (3.37)$$

3.26 Application to MCS

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

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

$$Y^{(k)} = \hat{\beta}^{(k)} X^{(k)} \quad (3.38)$$

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

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

```
Coefficients: (Intercept)          A
-37.51896      0.04656
```

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

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

$$D^{-Q} = \hat{\beta}^{-Q} A^{-Q} \quad (3.39)$$

3.27 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. $\hat{Y} = HY$.

$$H = X(X^T X)^{-1} X^T \quad (3.40)$$

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:

$$R = (I - H)Y \quad (3.41)$$

The variances of Y and R can be expressed as:

$$\begin{aligned} \text{var}(Y) &= H\sigma^2 \\ \text{var}(R) &= (I - H)\sigma^2 \end{aligned} \quad (3.42)$$

3.28 Hat Values for MCS regression

`fit = lm(D~A)`

$$H = A \left(A^{\top} A \right)^{-1} A^{\top},$$

3.28.1 Influence measures using R

R provides the following influence measures of each observation.

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

3.29 Prediction Interval Sum of Squares

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

Bibliography

- Christensen, R., L. M. Pearson, and W. Johnson (1992). Case-deletion diagnostics for mixed models. *Technometrics* 34(1), 38–45.
- Cook, R. (1986). Assessment of local influence. *Journal of the Royal Statistical Society. Series B (Methodological)* 48(2), 133–169.
- 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.
- Preisser, J. S. (1996). Deletion diagnostics for generalised estimating equations. *Biometrika* 83(3), 551–5562.
- Schabenberger, O. (2004). Mixed model influence diagnostics. 18929.