

Residual and Influence Diagnostics

Kevin O'Brien

April 18, 2012

Contents

1	Residual and Influence Diagnostics	4
1.1	Chapter Overview	4
1.2	Residual Diagnostics	5
1.2.1	Introduction	5
1.2.2	Marginal and Conditional Diagnostics	5
1.2.3	Scaled Residuals	6
1.3	Influence Diagnostics	7
1.3.1	Underlying Concepts	8
1.3.2	Cook's Distance	8
1.3.3	Managing the Covariance Parameters	8
1.3.4	Predicted Values, PRESS Residual and the PRESS Statistic . .	8
1.3.5	Leverage	8
1.3.6	Internally and Externally Studentized Residuals	9
1.3.7	DFFITs and MDFFITs	9
1.3.8	Covariance Ratio and Trace	9
1.3.9	Likelihood Distance	9
1.3.10	Non-iterative Update Procedures	11
1.3.11	Miscellaneous	11
2	Case Deletion Diagnostics for Mixed Models	12
2.1	Overview	12
2.2	Case Deletion Diagnostics for Mixed Models	12
2.2.1	Extending deletion diagnostics to LMEs	13
2.2.2	Influence on measure component ratios	13
2.3	Christensen et al	15

2.4	Haslett Hayes	16
2.5	Schabenberger	17
2.5.1	Iterative and non-iterative influence analysis	17
3	Deletion Diagnostics	18
3.1	Deletion Diagnostics	18
3.1.1	Terminology for Case Deletion diagnostics	18
3.1.2	Case-Deletion results for Variance components	18
3.2	Efficient updating	19
3.2.1	The Hat Matrix	19
3.3	Efficient updating theorem	19
4	Updating Techniques and Cross Validation	21
4.1	Updating of Regression Estimates	21
4.2	Updating of Regression Estimates	21
4.2.1	Updating Standard deviation	22
4.2.2	Updating Regression Estimates	22
4.2.3	Inference on intercept and slope	23
4.3	Sherman Morrison Woodbury Formula	24
4.4	Cross Validation	24
4.5	Model Validation	25
5	Software Implementation	26
5.1	The Hat Matrix	26
5.2	Hat Values for MCS regression	26
5.3	Grubbs' data	27
5.3.1	Grubb's example	28
5.3.2	Influence measures using R	28
6	Generalized linearmodels	30
6.1	Generalized Linear model	30
6.2	Generalized Model(GzLM)	30
6.2.1	What is a GzLM	30
6.2.2	GzLM Structure	31
6.2.3	Link Function	31

6.2.4	Canonical parameter	31
6.2.5	Dispersion parameter	31
6.2.6	Iteratively weighted least square	31
6.2.7	Residual Components	32
6.3	Generalized linear mixed models	33
7	Updating for Classical linear models	34
7.1	Updating of Regression Estimates	34
7.1.1	Updating Standard deviation	34
7.1.2	Updating Regression Estimates	35
7.1.3	Inference on intercept and slope	35
8	Lesaffre's paper.	37
8.1	Lesaffre's paper.	37
9	LME Likelihood	38
9.1	One Way ANOVA	38
9.1.1	Page 448	38
9.1.2	Page 448- simple example	38
9.1.3	Extention to several random effects	40
	Bibliography	41

Chapter 1

Residual and Influence Diagnostics

1.1 Chapter Overview

1. Residual Diagnostics
 - (a) Marginal and Conditional Diagnostics
 - (b) Scaled Residuals
2. Influence Diagnostics
 - (a) Underlying Concepts
 - (b) Managing the Covariance Parameters
 - (c) Predicted Values, PRESS Residual and the PRESS Statistic
 - (d) Leverage
 - (e) Internally and Externally Studentized Residuals
 - (f) DFFITs and MDFFITs
 - (g) Covariance Ratio and Trace
 - (h) Likelihood Distance
 - (i) Non-iterative Update Procedures

1.2 Residual Diagnostics

Consider a residual vector of the form $\hat{e} = \mathbf{P}\mathbf{Y}$, where \mathbf{P} is a projection matrix, possibly an oblique projector. External studentization uses an estimate of Var that does not involve the i th observation. Externally studentized residuals are often preferred over studentized residuals because they have well known distributional properties in the standard linear models for independent data. Residuals that are scaled by the estimated variances of the responses are referred to as Pearson-type residuals. Standardization:

$$\frac{\hat{e}_i}{\sqrt{v_i}}$$

Studentization

$$\frac{\hat{e}_i}{\sqrt{\hat{v}_i}}$$

1.2.1 Introduction

1.2.2 Marginal and Conditional Diagnostics

[Gregoire, Schabenberger, Barrett (1995)]

$$\mathbf{r}_m = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}$$

$$\mathbf{r}_c = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{Z}\hat{\boldsymbol{\gamma}}$$

For the individual observation the raw studentized and pearson type residuals are computed as follows:

$$r_{mi} = Y_i - \mathbf{X}'_i\hat{\boldsymbol{\beta}}$$

$$r_{ci} = r_{mi} - Y_i - \mathbf{z}'_i\hat{\boldsymbol{\gamma}}$$

Marginal Residuals

The marginal residuals are defined according to

$$\hat{\xi} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{M}^{-1}\mathbf{Q}\mathbf{y}.$$

Plots of the elements of the marginal residual vector versus the explanatory variables in X can be used to check the linearity of \mathbf{y} in a similar manner to the residual plots used in linear models.

Conditional Residuals

A conditional probability is the difference between the observed value and the predicted value of the dependent variable.

$$\hat{\epsilon}_i = y_i - X_i\hat{\beta} + Z_i\hat{b}_i$$

In general conditional residuals are not well suited for verifying model assumptions and detecting outliers. Even if the true model residuals are uncorrelated and have equal variance, conditional variances will tend to be correlated and their variances may be different for different subgroups of individuals (?).

1.2.3 Scaled Residuals

Pinheiro and Bates (1994) describes three types of residual that describe the variabilities present in LME models

1. marginal residuals, $\hat{\xi}$, which predict marginal errors,
2. conditional residuals, $\hat{\epsilon}$, which predict conditional errors,
3. the BLUP, $Z\hat{b}$, that predicts random effects.

Each type of residual is useful to evaluate some assumption of the model.

According to Hilton-Minton [1995], a residual is considered pure for a specific type of error if it depends only on the fixed components and on the error that it is supposed to predict. Residuals that depend on other types of error are known as ‘confounded errors’.

Studentized Residuals

Standardization is not possible in practice. Studentized residuals are residuals divided by the estimated standard estimation.

1.3 Influence Diagnostics

- a Overall Measures that compare changes in objectives functions; (restricted) maximum likelihood (Cook Weisberg, 1982)
- b Influence on parameter estimates: Cook's Distance, MDFFITs
- c Influence on precision of estimates: CovTrace and CovRatio
- d Influence on fitted and predicted values: PRESS residuals, PRESS statistics, DFFITs
- e Outlier properties : internally and externally studentized residuals, leverage

Cook 1986 introduced methods for local influence assessment. These methods provide a powerful tool for examining perturbations in the assumption of a model, particularly the effects of local perturbations of parameters of observations. The local-influence approach to influence assessment is quite different from the case deletion approach, comparisons are of interest.

Christensen et al. (1992) developed their global influences for the deletion of single observations in two steps: a one-step estimate for the REML (or ML) estimate of the variance components, and an ordinary case-deletion diagnostic for a weighted regression problem (conditional on the estimated covariance matrix) for fixed effects.

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.

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} .

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.

1.3.1 Underlying Concepts

1.3.2 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.

1.3.3 Managing the Covariance Parameters

1.3.4 Predicted Values, PRESS Residual and the PRESS Statistic

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

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.

$$PRESS = \sum (y - y^{(k)})^2 \quad (1.1)$$

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.

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

1.3.5 Leverage

Leverage can be defined through the projection matrix that results from a transformation of the model with the inverse of the Cholesky decomposition of \mathbf{V} , or an oblique

projector.

$\mathbf{Y} = \mathbf{H}\hat{\mathbf{Y}}$ While \mathbf{H} is idempotent, it is generally not symmetric and thus not a projection matrix in the narrow sense.

$$h_{ii} = \mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i$$

The trace of \mathbf{H} equals the rank of \mathbf{X} . If V_{ij} denotes the element in row i , column j of \mathbf{V}^{-1} , then for a model containing only an intercept the diagonal elements of \mathbf{H} .

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

1.3.6 Internally and Externally Studentized Residuals

The computation of internally studentized residuals relies on the diagonal values of $\mathbf{V}(\hat{\boldsymbol{\theta}}) - \mathbf{Q}(\hat{\boldsymbol{\theta}})$ Externally studentized residuals require iterative influence analysis or a profiled residual variance.

Cook's Distance

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

A DFFIT measures the change in predicted values due to the removal of data points. (Belsey, Kuh and Welsch (1980))

$\mathbf{D}(\boldsymbol{\beta}) = \boldsymbol{\delta}_{(U)}'\boldsymbol{\delta}_{(U)}/rank(\mathbf{X})$ Cook's D can be calibrated according to a chi-square distribution with degrees of freedom equal to the rank of \mathbf{X} ?.

CovTrace($\boldsymbol{\beta}$)

1.3.7 DFFITs and MDFFITs

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

1.3.8 Covariance Ratio and Trace

1.3.9 Likelihood Distance

The log-likelihood function l and restricted log-likelihood l_R of the LME model. $\boldsymbol{\psi}$ is the collection of all parameters (i.e. the fixed effects $\boldsymbol{\beta}$ and the covariance parameters $\boldsymbol{\theta}$).

Reduced data estimates ($\boldsymbol{\psi}_{(U)}$)

$$RLD_{(U)} = 2\{l_R(\boldsymbol{\psi}) - l_R(\boldsymbol{\psi}_{(U)})\}$$

$$LD_{(U)} = 2\{l(\boldsymbol{\psi}) - l(\boldsymbol{\psi}_{(U)})\}$$

Likelihood distance, known as likelihood displacements. The likelihood distance gives twice the amount by which the log likelihood of the full data changes if one were to use an estimate based on fewer data points. The likelihood distance is the a global summary measure of the influence of the observations in U jointly on all parameters.

An overall influence statistic measures the change in the objective function being minimized. In linear mixed models cook and Weisberg devised the Likelihood distance, known elsewhere as likelihood displacement.

The likelihood distance gives the amount of data by which the log-likelihood found when using the full data sets would change when the data set is reduced.

Importantly the value $l(\hat{\phi})_{(U)}$ is not the log-likelihood obtained from the reduced set, but the determining the likelihood function based on the full set at the reduced data estimates.

The approach can be applied to ML and REML models. Schabenberger (2004) uses in notation the subscript R to specify that REML models are under consideration.

The likelihood distance LD of observation group U is given by

$$LD_U = 2l(\hat{\phi}) - 2l(\hat{\phi}_{(U)}) \quad (1.2)$$

where l is the log likelihood function. If LD_U is large then the observation group U is influential on the likelihood function. The likelihood distance is a global summary measure, expressing the joint influence of the observation group U on all parameters ϕ subject to updating.

That U is influential is not grounds for deletion or changing the model. Should U be found to be influential, Schabenberger (2004) advises that the nature of that influence be determined. Estimates of the coefficients and precision of fixed effects, the coefficients and precision of covariance parameters, and fitted and predicted values should all be examined in light of determining that U is influential. Influence may be exerted by U on covariance parameters without affecting the fixed effects.

The likelihood distance has been widely used to detect outlying observations in data analysis. [cite: Cook and Weisberg] suggested that the likelihood distance may be compared to a χ^2 distribution for large samples.

1.3.10 Non-iterative Update Procedures

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

$$\hat{\beta} - \hat{\beta}_{(U)} = \Omega X V (U P U)$$

Residual variance

When σ^2 is profiled out of the marginal variance-covariance matrix, a closed-form estimate of σ^2 that is only based on only the remaining observation an be computed as follows, provided $V = V(\theta)$ [cite: Hurtado 1993]

Likelihood Distances

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

1.3.11 Miscellaneous

DFBETA

$$DFBETA_a = \hat{\beta} - \hat{\beta}_{(a)} \quad (1.3)$$

$$= B(Y - Y_{\bar{a}}) \quad (1.4)$$

Mean Square Prediction Error

$$MSPR = \frac{\sum (y_i - \hat{y}_i)^2}{n^*} \quad (1.5)$$

Effects on parameter estimate

Cook's Distance. CD .

Effects on the fitted and predicted values

Schabenberger (2004) describes the use of the *PRESS* and *DFFITs* in determining influence.

The *PRESS* residual is the difference between the observed value and the predicted (marginal) value.

$$\hat{e}_{i(U)} = y_i - x_i \hat{\beta}_{(U)} \quad (1.6)$$

Chapter 2

Case Deletion Diagnostics for Mixed Models

2.1 Overview

1. Extending deletion diagnostics to LMEs
2. Christensen et al
3. Haslett hayes
4. Schabenberger
5. Tewomir

2.2 Case Deletion Diagnostics for 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.

2.2.1 Extending deletion diagnostics to LMEs

after fitting a mixed, it is important to carry put model diagnostics to check whether distributional assumptions for the residuals as satisfied and whether the fit the model is sensitive to unusual assumptions. The process of carrying out model diagnostic involves several informal and formal techniques.

$$X = \begin{bmatrix} x'_i \\ X(i) \end{bmatrix}, Z = \begin{bmatrix} z'_{ij} \\ Z_{j(i)} \end{bmatrix}, Z = \begin{bmatrix} z'_{ij} \\ Z_{j(i)} \end{bmatrix},$$

$$y = \begin{bmatrix} y'_{ij} \\ y_{j(i)} \end{bmatrix} \text{ and } H = \begin{bmatrix} h_{ii} & h \\ h_{j(i)} & h \end{bmatrix}$$

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

$\mathbf{a}_{(i)}$ denotes a vector \mathbf{a} with the i -th element, a_i , removed.

$$\check{a}_i = \mathbf{a}_i - \mathbf{A}_{(i)} \mathbf{H}_{[i]} \mathbf{h}_i \quad (2.1)$$

2.2.2 Influence on measure component ratios

The general diagnostic tools for variance component ratios are the analogues of the Cook's distance and the information ratio.

The analogue of Cooks distance measure for variance components γ is denoted $CD(\gamma)$.

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

Large values of $CD(\gamma)$ highlight observation groups for closer attentions

The analogue of the information ratio measures the change in the determinant of the maximum likelihood estimates information matrix

$$IR_\gamma = \frac{\det(\mathbf{Q} - \mathbf{G})}{\det(\mathbf{Q})}$$

Ideally when all observations have the same influence on the information matrix IR_γ is approximately one. Deviations from one indicate the group U is influential.

Since $\text{var}(\hat{\gamma})$ and \mathbf{I}_r are fixed for all observations, $IR\gamma$ is a function of \mathbf{G} , in turn a function of \mathbf{C}_i and c_{ii} .

2.3 Christensen et al

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

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

- ? 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.

- ? 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

2.4 Haslett Hayes

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.

In contrast to classical linear models, diagnostics for LME are difficult to perform and interpret, because of the increased complexity of the model

2.5 Schabenberger

Standard residual and influence diagnostics for linear models can be extended to linear mixed models. The dependence of fixed-effects solutions on the covariance parameter estimates has important ramifications in perturbation analysis. To gauge the full impact of a set of observations on the analysis, covariance parameters need to be updated, which requires refitting of the model.

The conditional (subject-specific) and marginal (population-averaged) formulations in the linear mixed model enable you to consider conditional residuals that use the estimated BLUPs of the random effects, and marginal residuals which are deviations from the overall mean. Residuals using the BLUPs are useful to diagnose whether the random effects components in the model are specified correctly, marginal residuals are useful to diagnose the fixed-effects components.

2.5.1 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 doesn't increase the procedures execution time by the same degree.

estimation

$$\hat{\beta} = X^T \tag{2.2}$$

$$\hat{\gamma} = G(\hat{\theta})Z^T \tag{2.3}$$

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.

Chapter 3

Deletion Diagnostics

3.1 Deletion Diagnostics

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

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

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

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

3.1.1 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.

3.1.2 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 ex-

aming case influence. The method describes focuses on REML estimation, but can easily be adapted to ML or other methods.

3.2 Efficient updating

[cite:tewomir]

3.2.1 The Hat Matrix

The hat matrix, also known as the projection matrix, is well known in classical linear models. The diagonal elements h_{ii} are known as ‘leverages’. The properties of \mathbf{H} , such as symmetry and idempotency, are well known.

$$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

$$\mathbf{H} = \begin{bmatrix} h_{ii} & \mathbf{h}'_i \\ \mathbf{h}_i & \mathbf{H}_{(i)} \end{bmatrix}$$

$\mathbf{H}_{(i)}$ is an $(n-1) \times (n-1)$ matrix. It’s inversion for each i is computationally expensive.

$$\mathbf{C} = \mathbf{H}^{-1} = \begin{bmatrix} c_{ii} & \mathbf{h}'_c \\ \mathbf{c}_i & \mathbf{C}_{(i)} \end{bmatrix}$$

3.3 Efficient updating theorem

It is convenient to write partitioned matrices in which the i -th case is isolated. The partitioned matrix is written as $i = 1$, but the results apply in general.

If $\mathbf{C}'_i = [c_{ii}, \mathbf{c}'_i]$, such that \mathbf{C}_i is the i -th column of \mathbf{H}^{-1} then

- $m_i = \frac{1}{c_{ii}}$
- $\check{x}_i = \frac{1}{c_{ii}}\mathbf{X}'\mathbf{C}_i$

- $\check{z}_{ji} = \frac{1}{c_{ii}} \mathbf{Z}'_j \mathbf{C}_i$

- $\check{y}_i = \frac{1}{c_{ii}} \mathbf{y}' \mathbf{C}_i$

Once \mathbf{H}^{-1} is determined, an efficient updating formula can be applied.

$$\mathbf{H}^{-1} = \mathbf{I} - \mathbf{Z}(\mathbf{D}^{-1} + \mathbf{Z}\mathbf{Z})^{-1} \mathbf{Z}' \quad (3.1)$$

Chapter 4

Updating Techniques and Cross Validation

4.1 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.

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.

4.2 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.

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.

4.2.1 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.

$$S^2 = \frac{\sum_{i=1}^n (x_i^2) - \frac{(\sum_{i=1}^n x_i)^2}{n}}{n - 1} \quad (4.1)$$

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

$$Sxy = \frac{\sum_{i=1}^n (x_i y_i) - \frac{(\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n}}{n - 1}. \quad (4.2)$$

4.2.2 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

$$Sxx^{(j)} = \frac{\sum_{i=1}^n (x_i^2) - (x_j)^2 - \frac{((\sum_{i=1}^n x_i) - x_j)^2}{n-1}}{n - 2} \quad (4.3)$$

$$Syy^{(j)} = \frac{\sum_{i=1}^n (y_i^2) - (y_j)^2 - \frac{((\sum_{i=1}^n y_i) - y_j)^2}{n-1}}{n - 2} \quad (4.4)$$

$$Sxy^{(j)} = \frac{\sum_{i=1}^n (x_i y_i) - (y_j x_j) - \frac{((\sum_{i=1}^n x_i) - x_j)((\sum_{i=1}^n y_i) - y_j)}{n-1}}{n - 2} \quad (4.5)$$

The updated estimate for the slope is therefore

$$\hat{\beta}_1^{(j)} = \frac{Sxy^{(j)}}{Sxx^{(j)}} \quad (4.6)$$

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

$$\bar{x}^{(j)} = \frac{(\sum_{i=1}^n x_i) - (x_j)}{n - 1}, \quad (4.7)$$

$$\bar{y}^{(j)} = \frac{(\sum_{i=1}^n y_i) - (y_j)}{n - 1}. \quad (4.8)$$

The updated intercept estimate is therefore

$$\hat{\beta}_0^{(j)} = \bar{y}^{(j)} - \hat{\beta}_1^{(j)} \bar{x}^{(j)}. \quad (4.9)$$

4.2.3 Inference on intercept and slope

$$\hat{\beta}_1 \pm t_{(\alpha, n-2)} \sqrt{\frac{S^2}{(n-1)S_x^2}} \quad (4.10)$$

$$\frac{\hat{\beta}_0 - \beta_0}{SE(\hat{\beta}_0)} \quad (4.11)$$

$$\frac{\hat{\beta}_1 - \beta_1}{SE(\hat{\beta}_1)} \quad (4.12)$$

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 .

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

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

(4.13)

4.3 Sherman Morrison Woodbury Formula

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

$$(A + a^T B)^{-1} = A^{-1} - A^{-1} a^T (I - b A^{-1} a^T)^{-1} b A^{-1} \quad (4.14)$$

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

$$(X^T X \pm x_i x_i^T)^{-1} = (X^T X)^{-1} \mp \frac{(X^T X)^{-1} (x_i x_i^T (X^T X)^{-1})}{1 - x_i^T (X^T X)^{-1} x_i} \quad (4.15)$$

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. $\hat{Y} = HY$.

$$H = X(X^T X)^{-1} X^T \quad (4.16)$$

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:

$$R = (I - H)Y \quad (4.17)$$

The variances of Y and R can be expressed as:

$$\begin{aligned} \text{var}(Y) &= H\sigma^2 \\ \text{var}(R) &= (I - H)\sigma^2 \end{aligned} \quad (4.18)$$

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.

4.4 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.

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.

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.

4.5 Model Validation

Three basic approaches are described by Neter et al

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

Chapter 5

Software Implementation

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

$$H = X(X^T X)^{-1} X^T \quad (5.1)$$

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:

$$R = (I - H)Y \quad (5.2)$$

The variances of Y and R can be expressed as:

$$\begin{aligned} \text{var}(Y) &= H\sigma^2 \\ \text{var}(R) &= (I - H)\sigma^2 \end{aligned} \quad (5.3)$$

5.2 Hat Values for MCS regression

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

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

5.3 Grubbs' data

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 $\hat{\beta}_1^{-4}$

$$Y^{(k)} = \hat{\beta}^{(k)} X^{(k)} \quad (5.4)$$

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

$$D^{-Q} = \hat{\beta}^{-Q} A^{-Q} \quad (5.5)$$

5.3.1 Grubb's example

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 $\hat{\beta}_1^{-4}$

$$Y^{-Q} = \hat{\beta}^{-Q} X^{-Q} \quad (5.6)$$

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

$$D^{-Q} = \hat{\beta}^{-Q} A^{-Q} \quad (5.7)$$

5.3.2 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 6

Generalized linearmodels

6.1 Generalized Linear model

In statistics, the generalized linear model (GzLM) is a flexible generalization of ordinary least squares regression. The GzLM generalizes linear regression by allowing the linear model to be related to the response variable via a link function and by allowing the magnitude of the variance of each measurement to be a function of its predicted value.

Mixed Effects Models offer a flexible framework by which to model the sources of variation and correlation that arise from grouped data. This grouping can arise when data collection is undertaken in a hierarchical manner, when a number of observations are taken on the same observational unit over time, or when observational units are in some other way related, violating assumptions of independence.

6.2 Generalized Model(GzLM)

Nelder and Wedderburn (1972) integrated the previously disparate and separate approaches to models for non-normal cases in a framework called "generalized linear models." The key elements of their approach is to describe any given model in terms of it's link function and it's variance function.

6.2.1 What is a GzLM

$$E(\mathbf{Y}) = \boldsymbol{\mu} = g^{-1}(\mathbf{X}\boldsymbol{\beta}) \tag{6.1}$$

where $E(Y)$ is the expected value of Y , $X\beta$ is the linear predictor, a linear combination of unknown parameters, β and g is the link function.

$$\text{Var}(\mathbf{Y}) = \text{V}(\boldsymbol{\mu}) = \text{V}(g^{-1}(\mathbf{X}\boldsymbol{\beta}))$$

6.2.2 GzLM Structure

The GzLM consists of three elements.

1. A probability distribution from the exponential family.
2. A linear predictor $\eta = X\beta$.
3. A link function g such that $E(Y) = \mu = g^{-1}(\eta)$.

6.2.3 Link Function

Definition 1 : The link function provides the relationship between the linear predictor and the mean of the distribution function. There are many commonly used link functions, and their choice can be somewhat arbitrary. It can be convenient to match the domain of the link function to the range of the distribution function's mean.

Definition 2 : A link function is the function that links the linear model specified in the design matrix, where columns represent the beta parameters and rows the real parameters.

6.2.4 Canonical parameter

θ , called the dispersion parameter,

6.2.5 Dispersion parameter

τ , called the dispersion parameter, typically is known and is usually related to the variance of the distribution.

6.2.6 Iteratively weighted least square

IWLS is used to find the maximum likelihood estimates of a generalized linear model.

Definition: An iterative algorithm for fitting a linear model in the case where the data may contain outliers that would distort the parameter estimates if other estimation

procedures were used. The procedure uses weighted least squares, the influence of an outlier being reduced by giving that observation a small weight. The weights chosen in one iteration are related to the magnitudes of the residuals in the previous iteration with a large residual earning a small weight.

6.2.7 Residual Components

In GzLMS the deviance is the sum of the deviance components

$$D = \sum d_i \tag{6.2}$$

In GzLMS the deviance is the sum of the deviance components

6.3 Generalized linear mixed models

[pawitan section 17.8]

The Generalized linear mixed model (GLMM) extend classical mixed models to non-normal outcome data.

In statistics, a generalized linear mixed model (GLMM) is a particular type of mixed model. It is an extension to the generalized linear model in which the linear predictor contains random effects in addition to the usual fixed effects. These random effects are usually assumed to have a normal distribution.

Fitting such models by maximum likelihood involves integrating over these random effects.

Chapter 7

Updating for Classical linear models

7.1 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.

7.1.1 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.

$$S^2 = \frac{\sum_{i=1}^n (x_i^2) - \frac{(\sum_{i=1}^n x_i)^2}{n}}{n - 1} \quad (7.1)$$

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

$$Sxy = \frac{\sum_{i=1}^n (x_i y_i) - \frac{(\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n}}{n-1}. \quad (7.2)$$

7.1.2 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

$$Sxx^{(j)} = \frac{\sum_{i=1}^n (x_i^2) - (x_j)^2 - \frac{((\sum_{i=1}^n x_i) - x_j)^2}{n-1}}{n-2} \quad (7.3)$$

$$Syy^{(j)} = \frac{\sum_{i=1}^n (y_i^2) - (y_j)^2 - \frac{((\sum_{i=1}^n y_i) - y_j)^2}{n-1}}{n-2} \quad (7.4)$$

$$Sxy^{(j)} = \frac{\sum_{i=1}^n (x_i y_i) - (y_j x_j) - \frac{((\sum_{i=1}^n x_i) - x_j)((\sum_{i=1}^n y_i) - y_j)}{n-1}}{n-2} \quad (7.5)$$

The updated estimate for the slope is therefore

$$\hat{\beta}_1^{(j)} = \frac{Sxy^{(j)}}{Sxx^{(j)}} \quad (7.6)$$

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

$$\bar{x}^{(j)} = \frac{(\sum_{i=1}^n x_i) - (x_j)}{n-1}, \quad (7.7)$$

$$\bar{y}^{(j)} = \frac{(\sum_{i=1}^n y_i) - (y_j)}{n-1}. \quad (7.8)$$

The updated intercept estimate is therefore

$$\hat{\beta}_0^{(j)} = \bar{y}^{(j)} - \hat{\beta}_1^{(j)} \bar{x}^{(j)}. \quad (7.9)$$

7.1.3 Inference on intercept and slope

$$\hat{\beta}_1 \pm t_{(\alpha, n-2)} \sqrt{\frac{S^2}{(n-1)S_x^2}} \quad (7.10)$$

$$\frac{\hat{\beta}_0 - \beta_0}{SE(\hat{\beta}_0)} \quad (7.11)$$

$$\frac{\hat{\beta}_1 - \beta_1}{SE(\hat{\beta}_1)} \quad (7.12)$$

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 .

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

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

(7.13)

Chapter 8

Lesaffre's paper.

8.1 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

$$C_i = 2|\Delta_i' L^{-1} \Delta_i|. \quad (8.1)$$

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

$$IF_i = -L^{-1} \Delta_i \quad (8.2)$$

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.

Chapter 9

LME Likelihood

9.1 One Way ANOVA

9.1.1 Page 448

Computing the variance of $\hat{\beta}$

$$\text{var}(\hat{\beta}) = (X'V^{-1}X)^{-1} \quad (9.1)$$

It is not necessary to compute V^{-1} explicitly.

$$V^{-1}X = \Sigma^{-1}X - Z(Z'\Sigma^{-1}Z)^{-1}Z'\Sigma^{-1}X \quad (9.2)$$

$$= \Sigma^{-1}(X - Zb_x) \quad (9.3)$$

The estimate b_x is the same term obtained from the random effects model; $X = Zb_x + e$, using X as an outcome variable. This formula is convenient in applications where b_x can be easily computed. Since X is a matrix of p columns, b_x can simple be computed column by column. according to the columns of X .

9.1.2 Page 448- simple example

Consider a simple model of the form;

$$y_{ij} = \mu + \beta_i + \epsilon_{ij}.$$

The iterative procedure is as follows Evaluate the individual group mean \bar{y}_i and variance $\hat{\sigma}_b^2$. Then use the variance of the group means as an estimate of the σ_b^2 . The average of the the variances of the groups is the initial estimate of the σ_e^2 .

Iterative procedure

The iterative procedure comprises two steps, with 0 as the first approximation of b_i .

The first step is to compute λ , the ratio of variabilities,

$$\lambda = \frac{\sigma_b^2}{\sigma_e^2}$$

$$\mu = \frac{1}{N} \sum_{ij} (y_{ij} - b_i)$$
$$b_i = \frac{n(\bar{y}_i - \mu)}{n + \lambda}$$

The second step is to updat σ_e^2

$$\sigma_e^2 = \frac{e'e}{N - df} \quad (9.4)$$

where e is the vector of $e_{ij} = y_{ij} - \mu - b_i$ and $df = qn/n + \lambda$ and

$$\sigma_b^2 = \frac{1}{q} \sum_{i=1}^q b_i^2 + \left(\frac{n}{\sigma_e^2} + \frac{1}{\sigma_b^2} \right)^{-1} \quad (9.5)$$

Worked Example

Further to [pawitan 17.1] the initial estimates for variability are $\sigma_b^2 = 1.7698$ and $\sigma_e^2 = 0.3254$. At convergence the following results are obtained.

n=16, q=5

$$\hat{\mu} = \bar{y} = 14.175$$

$$\hat{\sigma}^2 = 0.325$$

$$\hat{\sigma}_b^2 = 1.395$$

$$\sigma = 0.986$$

At convergence the following estimates are obtained,

$$\hat{\mu} = 14.1751$$

$$\hat{b} = (-0.6211, 0.2683, 1.4389, -1.914, 0.8279)$$

$$\hat{\sigma}_b^2 = 1.3955$$

$$\hat{\sigma}_e^2 = 0.3254$$

9.1.3 Extention to several random effects

[pawitan section 17.7]

Bibliography

- Christensen, R., L. M. Pearson, and W. Johnson (1992). Case-deletion diagnostics for mixed models. *Technometrics* 34(1), 38–45.
- Cook, R. (1977). Detection of influential observations in linear regression. *Technometrics* 19, 15–18.
- Pinheiro, J. and D. Bates (1994). *Mixed Effects Models in S and S plus* (2nd ed.). Reading, Massachusetts: Springer.
- Preisser, J. S. (1996). Deletion diagnostics for generalised estimating equations. *Biometrika* 83(3), 551–5562.
- Schabenberger, O. (2004). Mixed model influence diagnostics. 18929.