Contents

1   Residual

A residual (or fitting error), on the other hand, is an observable estimate of the unobservable statistical error. Residual (or error) represents unexplained (or residual) variation after fitting a regression model. It is the difference (or left over) between the observed value of the variable and the value suggested by the regression model. Consider the previous example with men's heights and suppose we have a random sample of n people. The sample mean could serve as a good estimator of the population mean. Then we have:

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

$$\text{Residual} = \text{Observed value} - \text{Predicted value}$$

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

Both the sum and the mean of the residuals are equal to zero. .

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

1.1   Other uses of the word "error" in statistics

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

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

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

1.2   Residual Plots

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

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

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

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

In the next lesson, we will work on a problem, where the residual plot shows a non-random pattern. And we will show how to "transform" the data to use a linear model with nonlinear data.

2   Studentization

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

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

3   Residual Analysis for LME Models

In classical linear models model diagnostics have been become a required part of any statistical analysis, and the methods are commonly available in statistical packages and standard textbooks on applied regression. However it has been noted by several papers that model diagnostics do not often accompany LME model analyses.

Cite:Zewotir lists several established methods of analyzing influence in LME models. These methods include

- Cook's distance for LME models,

- likelihood distance,

- the variance (information) ration,

- the Cook-Weisberg statistic,

- the Andrews-Prebigon statistic.

3.1 LME REsiduals

Cox and Snell (1968, JRSS-B): general definition of residuals for models with single source of variability Hilden-Minton (1995, PhD thesis UCLA), Verbeke and Lesaffre (1997, CSDA) or Pinheiro and Bates (2000, Springer): extension to define three types of residuals that accommodate the extra source of variability present in linear mixed models, namely:

i) Marginal residuals,

predictors of marginal errors,

ii) Conditional residuals,

$$be = yX\hat{\beta}Zbb = \hat{\sigma}Q\hat{y}$$

, predictors of conditional errors

$$e = yE[y|b] = yX\beta Zb$$

iii) BLUP, Zbb, predictors of random effects,

$$Zb = E[y|b]E[y]$$

4 Residual Diagnostics

Consider a residual vector of the form $\hat{e} = \boldsymbol{PY}$, where $\boldsymbol{P}$ is a projection matrix, possibly an oblique projector. External studentization uses an estimate of Var that does not involve the ith 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

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

5 residuals.lme nlme- Extract lme Residuals

The residuals at level i are obtained by subtracting the fitted levels at that level from the response vector (and dividing by the estimated within-group standard error, if `type="pearson"`).

The fitted values at level i are obtained by adding together the population fitted values (based only on the fixed effects estimates) and the estimated contributions of the random effects to the fitted values at grouping levels less or equal to i.

```
fm1 <- lme(distance ~ age + Sex,
data = Orthodont, random = ~ 1)
```

```
head(residuals(fm1, level = 0:1))
summary(residuals(fm1) /
residuals(fm1, type = "p"))


# constant scaling factor 1.432
```

6   Diagnostic Plots for Linear Models with R

Plot Diagnostics for an `lm` Object
    Six plots (selectable by `which`) are currently available:

1. a plot of residuals against fitted values,

2. a Scale-Location plot of *sqrt(— residuals —)* against fitted values,

3. a Normal Q-Q plot,

4. a plot of Cook's distances versus row labels,

5. a plot of residuals against leverages,

6. a plot of Cook's distances against leverage/(1-leverage).

By default, the first three and 5 are provided.

6.0.1   Residuals plots

lme allows to plot the residuals in the following ways:

```
res_lme=residuals(model_lme)
plot(res_lme)
qqnorm(res_lme)
qqline(res_lme)
plot(model_lme)
```

When the `plot` function calls the model object, the residual plot is produced.

```
plot(JS.roy1, which=c(1) )
```

LME models assume that the residuals of the model are normally distributed. A Normal probability plot can be constructed to check this assumption. Commonly used `R` commands can be used to construct the plot.

```
qqnorm(resid(JS.roy1),pch="*",col="red")
qqline(resid(JS.roy1),col="blue")
```

```
table(dat$method[1:255])
##
##   J   S
## 255   0
table(dat$method[256:510])
##
##   J   S
##   0 255
```

```
library(predictMeans)
CookD(model, group=method, plot=TRUE, idn=5, newwd=FALSE)
```

```
> shapiro.test(resid(JS.roy1)[256:510])

Shapiro-Wilk normality test

data:  resid(JS.roy1)[256:510]
W = 0.9395, p-value = 9.503e-09
```

```
plot(roy.NLME, resid(., type = "p") ~ fitted(.) | method,
abline = 0, id=.05)
```

```
library(predictMeans)
CookD(model, group=method, plot=TRUE, idn=5, newwd=FALSE)
```

```
blood.red <- blood[!(blood$subject %in% c(68,78,80)),]
dim(blood.red)
# 27 observations should be removed.

blood.NLME.red <-lme(BP ~ method-1 , random=~1|subject,data = blood.red)
plot(blood.NLME.red, resid(., type = "p") ~ fitted(.) | method, abline = 0, id=.05)
```

```
> shapiro.test(resid(JS.roy1)[1:255])

Shapiro-Wilk normality test

data:  resid(JS.roy1)[1:255]
W = 0.9931, p-value = 0.2852
```

```
> shapiro.test(resid(JS.roy1)[256:510])

Shapiro-Wilk normality test

data:  resid(JS.roy1)[256:510]
W = 0.9395, p-value = 9.503e-09
```

```
data.frame( response = resid(JS.ARoy20091, type = "response"),
pearson  = resid(JS.ARoy20091, type = "pearson"),
normalized = resid(JS.ARoy20091, type = "normalized") )
```

```
response        pearson      normalized
1    -4.65805902 -0.761587227 -0.7615872269
2    -0.88701342 -0.145025661  0.0776238081
3    -5.16580898 -0.844603753 -0.8446037530
4     2.29041830  0.374480726  0.6450898404
5     7.87508366  1.287567009  1.2875670086
6    -6.57048659 -1.074266908 -1.5090772378
......................................
```

For the J observations, the variance is 6.116252 whereas for the S observations, the denominator is 9.118144. (with the expected ratio of 1.490806)

```
> pearson %>%
+    as.numeric %>%
+    matrix(nrow=85) %>%
+    round(4)
[,1]     [,2]     [,3]     [,4]     [,5]     [,6]
[1,] -0.7616  0.2194  0.3829 -0.2983  0.3597 -0.0790
[2,] -0.1450  0.1820 -0.1450 -0.5014  0.1567  0.2663
[3,] -0.8446  0.4634  0.1364 -0.1630 -0.2727  0.1660
[4,]  0.3745 -0.2795 -0.2795 -0.2658 -0.2658  0.6115
[5,]  1.2876 -0.6744 -0.6744  0.8935 -0.0935 -0.8612
[6,] -1.0743  1.8687 -0.7473 -0.0383  0.2908 -0.3673
.............................................
```

We can plot the residuals against the fitted values, to assess the assumption of constant variance.

```
# standardized residuals versus fitted values
plot(JS.ARoy20091, resid(., type = "pearson") ~ fitted(.) ,
abline = 0, id = 0.05)
```

```
par(mfrow=c(1,2))
qqnorm((resid(JS.ARoy20091)[1:255]),
pch="*",col="red",
ylim=c(-40,40),
main="Method J")
qqline(resid(JS.ARoy20091)[1:255],col="blue")
qqnorm((resid(JS.ARoy20091)[256:510]),
pch="*",col="red",
ylim=c(-40,40),
main="Method S")
qqline(resid(JS.ARoy20091)[256:510],col="blue")
par(mfrow=c(1,1))
```

6.0.2   Residuals plots

When the `plot` function calls the model object, the residual plot is produced.

```
plot(JS.roy1, which=c(1) )
```

LME models assume that the residuals of the model are normally distributed. A Normal probability plot can be constructed to check this assumption. Commonly used `R` commands can be used to construct the plot.

```
qqnorm(resid(JS.roy1),pch="*",col="red")
qqline(resid(JS.roy1),col="blue")
```

```
table(dat$method[1:255])
##
##   J   S
## 255   0
table(dat$method[256:510])
##
##   J   S
##   0 255
```

```
plot(roy.NLME, resid(., type = "p") ~ fitted(.) | method,
abline = 0, id=.05)
```

```
data.frame( response = resid(JS.ARoy20091, type = "response"),
pearson  = resid(JS.ARoy20091, type = "pearson"),
normalized = resid(JS.ARoy20091, type = "normalized") )
```

```
response        pearson     normalized
1    -4.65805902 -0.761587227 -0.7615872269
2    -0.88701342 -0.145025661  0.0776238081
3    -5.16580898 -0.844603753 -0.8446037530
4     2.29041830  0.374480726  0.6450898404
5     7.87508366  1.287567009  1.2875670086
6    -6.57048659 -1.074266908 -1.5090772378
.......................................
```

For the J observations, the variance is 6.116252 whereas for the S observations, the denominator is 9.118144. (with the expected ratio of 1.490806)

```
> pearson %>%
+    as.numeric %>%
+    matrix(nrow=85) %>%
+    round(4)
 [,1]    [,2]    [,3]    [,4]    [,5]    [,6]
[1,] -0.7616  0.2194  0.3829 -0.2983  0.3597 -0.0790
[2,] -0.1450  0.1820 -0.1450 -0.5014  0.1567  0.2663
[3,] -0.8446  0.4634  0.1364 -0.1630 -0.2727  0.1660
[4,]  0.3745 -0.2795 -0.2795 -0.2658 -0.2658  0.6115
[5,]  1.2876 -0.6744 -0.6744  0.8935 -0.0935 -0.8612
[6,] -1.0743  1.8687 -0.7473 -0.0383  0.2908 -0.3673
......................................
```

We can plot the residuals against the fitted values, to assess the assumption of constant variance.

```
# standardized residuals versus fitted values
plot(JS.ARoy20091, resid(., type = "pearson") ~ fitted(.) ,
abline = 0, id = 0.05)
```

```
par(mfrow=c(1,2))
qqnorm((resid(JS.ARoy20091)[1:255]),
pch="*",col="red",
ylim=c(-40,40),
main="Method J")
qqline(resid(JS.ARoy20091)[1:255],col="blue")
qqnorm((resid(JS.ARoy20091)[256:510]),
pch="*",col="red",
ylim=c(-40,40),
main="Method S")
qqline(resid(JS.ARoy20091)[256:510],col="blue")
par(mfrow=c(1,1))
```

This code will allow you to make QQ plots for each level of the random effects. LME models assume that not only the within-cluster residuals are normally distributed, but that each level

of the random effects are as well. Depending on the model, you can vary the level from 0, 1, 2 and so on

```
qqnorm(JS.ARoy20091, ~ranef(.))

#  qqnorm(JS.ARoy20091, ~ranef(.,levels=1)
```
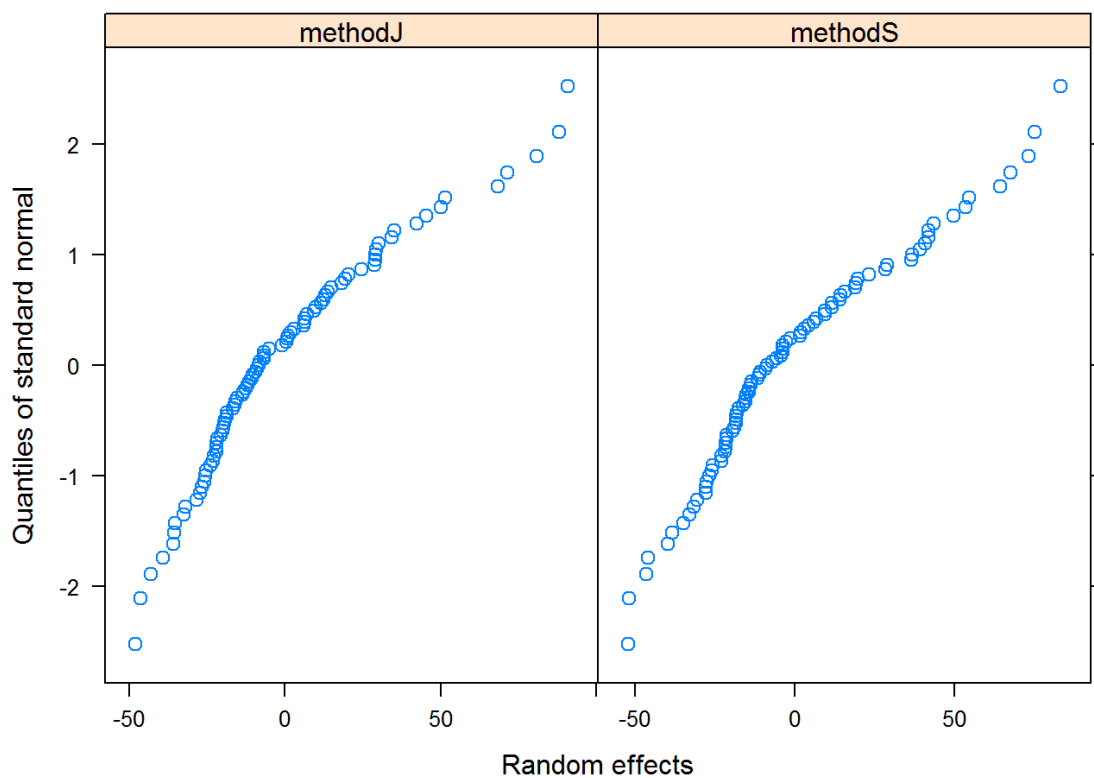

Figure 1:

```
data.frame( response = resid(JS.roy1, type = "response"),
pearson  = resid(JS.roy1, type = "pearson"),
normalized = resid(JS.roy1, type = "normalized") )
```

```
response        pearson     normalized
1     -4.65805902 -0.761587227 -0.7615872269
2     -0.88701342 -0.145025661  0.0776238081
3     -5.16580898 -0.844603753 -0.8446037530
```

```
4      2.29041830  0.374480726   0.6450898404
5      7.87508366  1.287567009   1.2875670086
6     -6.57048659 -1.074266908  -1.5090772378
```
....

For the J observations, the variance is 6.116252 whereas for the S observations, the denominator is 9.118144. (with the expected ratio of 1.490806)

```
> pearson %>%
+   as.numeric %>%
+   matrix(nrow=85) %>%
+   round(4)
[,1]    [,2]    [,3]    [,4]    [,5]    [,6]
[1,] -0.7616  0.2194  0.3829 -0.2983  0.3597 -0.0790
[2,] -0.1450  0.1820 -0.1450 -0.5014  0.1567  0.2663
[3,] -0.8446  0.4634  0.1364 -0.1630 -0.2727  0.1660
[4,]  0.3745 -0.2795 -0.2795 -0.2658 -0.2658  0.6115
[5,]  1.2876 -0.6744 -0.6744  0.8935 -0.0935 -0.8612
[6,] -1.0743  1.8687 -0.7473 -0.0383  0.2908 -0.3673
...........................................
```

We can plot the residuals against the fitted values, to assess the assumption of constant variance.

```
# standardized residuals versus fitted values
plot(JS.roy1, resid(., type = "pearson") ~ fitted(.) ,
abline = 0, id = 0.05)
```

- The **Scale-Location** plot, also called Spread-Location or S-L plot, takes the square root of the absolute residuals in order to diminish skewness (sqrt(—E—)) is much less skewed than — E — for Gaussian zero-mean E).

- The **Residual-Leverage** plot shows contours of equal Cook's distance, for values of cook.levels (by default 0.5 and 1) and omits cases with leverage one with a warning. If the leverages are constant (as is typically the case in a balanced aov situation) the plot uses factor level combinations instead of the leverages for the x-axis. (The factor levels are ordered by mean fitted value.)
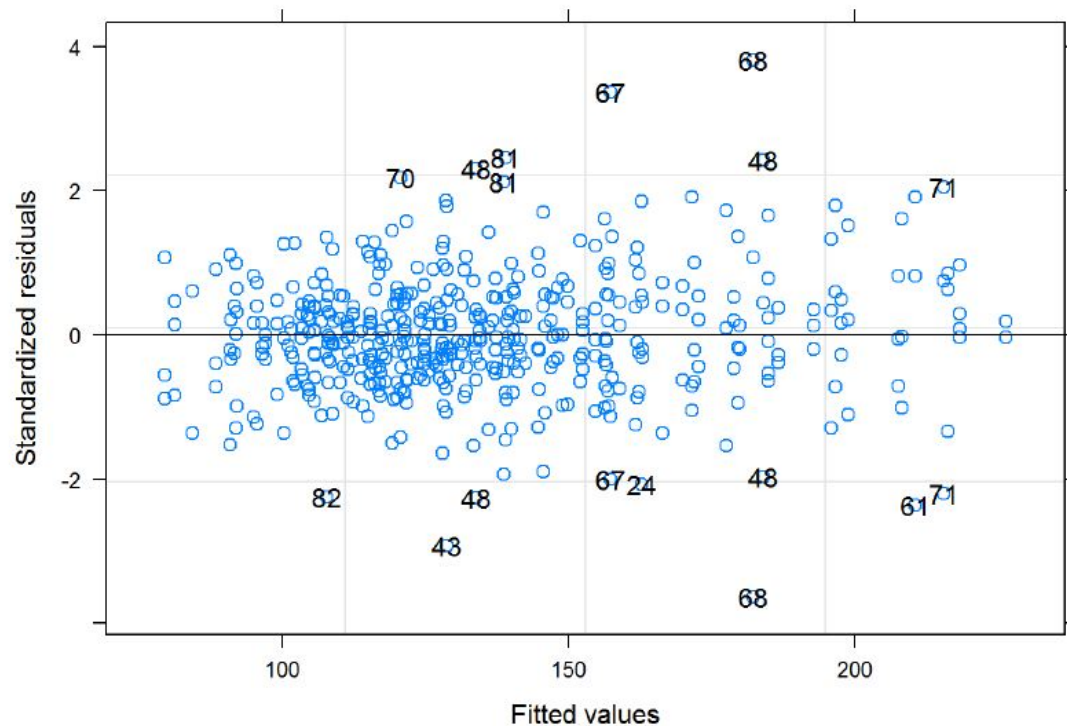
Figure 2:

```
par(mfrow=c(4,1))
plot(fittedmodel)
par(opar)
```

Residuals for LME Models

7 Diagnostic Tools for the nlme package

With the nlme package, the generic function `lme()` fits a linear mixed-effects model in the formulation described in Laird and Ware (1982) but allowing for nested random effects.

The within-group errors are allowed to be correlated and/or have unequal variances, which is very important in fitting the models for Roy's Tests

The nlme package has a limited set of diagnostic tools that can be used to assess the model fit. A review of the package manual is sufficient to get a sense of the package's capability in that regard.

8   Computation and Notation

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

   Model Diagnostics

9   Introduction

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

9.1   Model Data Agreement

schabenberger describes the examination of model-data agreement as comprising several elements; residual analysis, goodness of fit, collinearity diagnostics and influence analysis.

9.2   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.

9.3   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. CPJ 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.

9.4   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

9.5   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.

9.6   Quantifying Influence

The basic procedure for quantifying influence is simple as follows:

- Fit the model to the data and obtain estimates of all parameters.

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

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

10   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.

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.

11   Residual diagnostics

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

11.1   Residuals diagnostics in mixed models

The marginal and conditional means in the linear mixed model are $E[\boldsymbol{Y}] = \boldsymbol{X\beta}$ and $E[\boldsymbol{Y}|\boldsymbol{u}] = \boldsymbol{X\beta} + \boldsymbol{Zu}$, respectively.

A residual is the difference between an observed quantity and its estimated or predicted value. In the mixed model you can distinguish marginal residuals r_m and conditional residuals r_c.

11.2   Marginal and Conditional Residuals

A marginal residual is the difference between the observed data and the estimated (marginal) mean, $r_{mi} = y_i - x'_0\hat{b}$ A conditional residual is the difference between the observed data and the predicted value of the observation, $r_{ci} = y_i - x'_i\hat{b} - z'_i\hat{\gamma}$

In linear mixed effects models, diagnostic techniques may consider 'conditional' residuals. A conditional residual is the difference between an observed value y_i and the conditional predicted value $\hat{y}_i$.

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

However, using conditional residuals for diagnostics presents difficulties, as they tend to be correlated and their variances may be different for different subgroups, which can lead to erroneous conclusions.

$$r_{mi} = x_i^T\hat{\beta} \tag{1}$$

11.3   Marginal Residuals

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

12   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.

12.1   Standardization

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

12.2   Studentization

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

12.3   Internal and External Studentization

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

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

12.4   Computation

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

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

12.5   Pearson Residual

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

13   Covariance Parameters

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

14   Case Deletion Diagnostics

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

14.1   Deletion Diagnostics

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

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

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

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

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

15   Effects on fitted and predicted values

$$\hat{e}_{i(U)} = y_i - x\hat{\beta}_{(U)} \tag{2}$$

15.1   Case Deletion Diagnostics for Mixed Models

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

Christensen develops these techniques in the context of REML

15.2   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 lists several established methods of analyzing influence in LME models. These methods include

- Cook's distance for LME models,

- likelihood distance,

- the variance (information) ration,

- the Cook-Weisberg statistic,

- the Andrews-Prebigon statistic.

16   Influence analysis

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

west 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'.

16.1   Cook's 1986 paper on Local Influence

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

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

16.2   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].

17 Terminology for Case Deletion diagnostics

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.

18   Likelihood Distance

The likelihood distance gives the amount by which the log-likelihood of the full data changes if one were to evaluate it at the reduced-data estimates. The important point is that $l(\psi_{(U)})$ is not the log-likelihood obtained by fitting the model to the reduced data set.

It is obtained by evaluating the likelihood function based on the full data set (containing all n observations) at the reduced-data estimates.

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

18.1   Likelihood Distance

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

19   Iterative and non-iterative influence analysis

schabenberger 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.

19.1   Iterative Influence Analysis

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

schabenberger describes the choice between iterative influence analysis and non-iterative influence analysis.

20　The CPJ Paper

20.1　Case-Deletion results for Variance components

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

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

20.2　CPJ Notation

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

CPJ noted the following identity:

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

CPJ use the following as building blocks for case deletion statistics.

- $\breve{x}_i$

- $\breve{z}_i$

- $\breve{z}_i j$

- $\breve{y}_i$

- $p_i i$

- m_i

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

21   Matrix Notation for Case Deletion

21.1   Case deletion notation

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

21.2   Partitioning Matrices

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

22   CPJ's Three Propositions

22.0.1   Proposition 1

$$\boldsymbol{V}^{-1} = \left[\begin{array}{cc} \nu^{ii} & \lambda'_i \\ \lambda_i & \Lambda_{[i]} \end{array} \right]$$

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

22.1   Proposition 2

(i) $\boldsymbol{X}^T_{[i]} \boldsymbol{V}^{-1}_{[i]} \boldsymbol{X}_{[i]} = \boldsymbol{X}' \boldsymbol{V}^{-1} \boldsymbol{X}$

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

(iii) $\boldsymbol{X}^T_{[i]} \boldsymbol{V}^{-1}_{[i]} \boldsymbol{Y}_{[i]} = \boldsymbol{X}' \boldsymbol{V}^{-1} \boldsymbol{Y}$

22.2   Proposition 3

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

23   Measures of Influence

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

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

23.1   DFFITS

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

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

23.2   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.

$$PRESS = \sum (y - y^{(k)})^2 \tag{3}$$

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

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

23.3   DFBETA

$$
\begin{aligned}
DFBETA_a &= \hat{\beta} - \hat{\beta}_{(a)} \tag{4}\\
&= B(Y - Y_{\bar{a}} \tag{5}
\end{aligned}
$$

23.4   Residuals diagnostics in LME Models

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

- A marginal residual is the difference between the observed data and the estimated (marginal) mean, $r_{mi} = y_i - x_0'\hat{b}$

- A conditional residual is the difference between an observed value y_i and the conditional predicted value $\hat{y}_i$,
$$r_{ci} = y_i - x_i'\hat{b} - z_i'\hat{\gamma}$$

The marginal and conditional means in the linear mixed model are $E[\boldsymbol{Y}] = \boldsymbol{X\beta}$ and $E[\boldsymbol{Y}|\boldsymbol{u}] = \boldsymbol{X\beta} + \boldsymbol{Zu}$, respectively.

23.5   Marginal and Conditional Residuals

A marginal residual is the difference between the observed data and the estimated (marginal) mean, $r_{mi} = y_i - x_0'\hat{b}$ A conditional residual is the difference between the observed data and the predicted value of the observation, $r_{ci} = y_i - x_i'\hat{b} - z_i'\hat{\gamma}$

In linear mixed effects models, diagnostic techniques may consider 'conditional' residuals. A conditional residual is the difference between an observed value y_i and the conditional predicted value $\hat{y}_i$.

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

However, using conditional residuals for diagnostics presents difficulties, as they tend to be correlated and their variances may be different for different subgroups, which can lead to erroneous conclusions.

$$r_{mi} = x_i^T \hat{\beta} \tag{6}$$

23.6   Marginal Residuals

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

23.7   Marginal and Conditional Residuals

A marginal residual is the difference between the observed data and the estimated (marginal) mean, $r_{mi} = y_i - x_0'\hat{b}$ A conditional residual is the difference between the observed data and the predicted value of the observation, $r_{ci} = y_i - x_i'\hat{b} - z_i'\hat{\gamma}$

In linear mixed effects models, diagnostic techniques may consider 'conditional' residuals. A conditional residual is the difference between an observed value y_i and the conditional predicted value $\hat{y}_i$.

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

However, using conditional residuals for diagnostics presents difficulties, as they tend to be correlated and their variances may be different for different subgroups, which can lead to erroneous conclusions.

23.8   Residuals diagnostics in LME Models

The marginal and conditional means in the linear mixed model are $E[\boldsymbol{Y}] = \boldsymbol{X\beta}$ and $E[\boldsymbol{Y}|\boldsymbol{u}] = \boldsymbol{X\beta} + \boldsymbol{Zu}$, respectively.

$$r_{mi} = x_i^T \hat{\beta} \tag{7}$$

23.9   Marginal Residuals

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

Residuals

Residuals are used to examine model assumptions and to detect outliers and potentially influential data point. The raw residuals r_{mi} and r_{ci} are usually not well suited for these purposes.

- Conditional Residuals r_{ci}

- Marginal Residuals r_{mi}

-

Marginal Residuals

Distinction From Linear Models

- The differences between perturbation and residual analysis in the linear model and the linear mixed model are connected to the important facts that b

  and b depend on the estimates of the covariance parameters, that b

  has the form of an (estimated) generalized least squares (GLS) estimator, and that is a random vector.

- In a mixed model, you can consider the data in a conditional and an unconditional sense. If you imagine a particular realization of the random effects, then you are considering the conditional distribution Y—

- If you are interested in quantities averaged over all possible values of the random effects, then you are interested in Y; this is called the marginal formulation. In a clinical trial, for example, you may be interested in drug efficacy for a particular patient. If random effects vary by patient, that is a conditional problem. If you are interested in the drug efficacy in the population of all patients, you are using a marginal formulation. Correspondingly, there will be conditional and marginal residuals, for example.

- The estimates of the fixed effects

  depend on the estimates of the covariance parameters. If you are interested in determining the influence of an observation on the analysis, you must determine whether this is influence on the fixed effects for a given value of the covariance parameters, influence on the covariance parameters, or influence on both.

- Mixed models are often used to analyze repeated measures and longitudinal data. The natural experimental or sampling unit in those studies is the entity that is repeatedly observed, rather than each individual repeated observation. For example, you may be analyzing monthly purchase records by customer.

- An influential data point is then not necessarily a single purchase. You are probably more interested in determining the influential customer. This requires that you can measure the influence of sets of observations on the analysis, not just influence of individual observations.

- The computation of case deletion diagnostics in the classical model is made simple by the fact that model. Such update formulas are available in the mixed model only if you assume that the covariance parameters are not affected by the removal of the observation in question. This is rarely a reasonable assumption.

- The application of well-known concepts in model-data diagnostics to the mixed model can produce results that are at first counter-intuitive, since our understanding is steeped in the ordinary least squares (OLS) framework. As a consequence, we need to revisit these important concepts, ask whether they are portable to the mixed model, and gain new appreciation for their changed properties. An important example is the ostensibly simple concept of leverage.

- The definition of leverage adopted by the MIXED procedure can, in some instances, produce negative values, which are mathematically impossible in OLS. Other measures that have been proposed may be non-negative, but trade other advantages. Another example are properties of residuals. While OLS residuals necessarily sum to zero in any model (with intercept), this not true of the residuals in many mixed models.

24   Residual diagnostics

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

24.1   Residuals diagnostics in mixed models

The marginal and conditional means in the linear mixed model are $E[\boldsymbol{Y}] = \boldsymbol{X\beta}$ and $E[\boldsymbol{Y}|\boldsymbol{u}] = \boldsymbol{X\beta} + \boldsymbol{Zu}$, respectively.

    A residual is the difference between an observed quantity and its estimated or predicted value. In the mixed model you can distinguish marginal residuals r_m and conditional residuals r_c.

$$r_{mi} = x_i^T \hat{\beta} \tag{8}$$

24.2   Marginal Residuals

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

25   Conditional and Marginal Residuals

Conditional residuals include contributions from both fixed and random effects, whereas marginal residuals include contribution from only fixed effects.

    Suppose the linear mixed-effects model lme has an $n \times p$ fixed-effects design matrix $\boldsymbol{X}$ and an $n \times q$ random-effects design matrix $\boldsymbol{Z}$.

    Also, suppose the p-by-1 estimated fixed-effects vector is $\hat{\beta}$, and the q-by-1 estimated best linear unbiased predictor (BLUP) vector of random effects is $\hat{b}$. The fitted conditional response is

$$\hat{y}_{Cond} = X\hat{\beta} + Z\hat{b}$$

and the fitted marginal response is

$$\hat{y}_{Mar} = X\hat{\beta}$$

residuals can return three types of residuals:

- raw,

- Pearson, and

- standardized.

For any type, you can compute the conditional or the marginal residuals. For example, the conditional raw residual is

$$r_{Cond} = y - X\hat{\beta} - Z\hat{b}$$

and the marginal raw residual is

$$r_{Mar} = y - X\hat{\beta}$$

Cox and Snell (1968, JRSS-B): general definition of residuals for models with single source of variability Hilden-Minton (1995, PhD thesis UCLA), Verbeke and Lesaffre (1997, CSDA) or Pinheiro and Bates (2000, Springer): extension to define three types of residuals that accommodate the extra source of variability present in linear mixed models, namely:

i) Marginal residuals,

predictors of marginal errors,

ii) Conditional residuals,

$$be = yX\hat{\beta}Zbb = \hat{\sigma}Q\hat{y}$$

, predictors of conditional errors

$$e = yE[y|b] = yX\beta Zb$$

iii) BLUP, Zbb, predictors of random effects,

$$Zb = E[y|b]E[y]$$

Marginal residuals

$$y - X\beta = Z\eta + \epsilon$$

- Should be mean 0, but may show grouping structure

- May not be homoskedastic.

- Good for checking fixed effects, just like linear regr.

Conditional residuals

$$y - X\beta - Z\eta = \epsilon$$

- Should be mean zero with no grouping structure

- Should be homoscedastic.

- Good for checking normality of outliers

Random effects

$$y - X\beta - \epsilon = Z\eta$$

- Should be mean zero with no grouping structure

- May not be be homoscedastic.

26 Residual diagnostics

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

26.1 Residuals diagnostics in mixed models

The marginal and conditional means in the linear mixed model are $E[\boldsymbol{Y}] = \boldsymbol{X\beta}$ and $E[\boldsymbol{Y}|\boldsymbol{u}] = \boldsymbol{X\beta} + \boldsymbol{Zu}$, respectively.

A residual is the difference between an observed quantity and its estimated or predicted value. In the mixed model you can distinguish marginal residuals r_m and conditional residuals r_c.

$$r_{mi} = x_i^T \hat{\beta} \tag{9}$$

26.2 Marginal Residuals

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

26.3   Confounded Residuals

Hilden-Minton (1995, PhD thesis, UCLA): residual is 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 errors are called ***confounded residuals***