# 1 Residual Analysis for LME Models

## 1.1 Introduction

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

**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.

## 1.2 Zewotir Measures of Influence in LME Models

**?** describes a number of approaches to model diagnostics, investigating each of the following;

- Variance components
- Fixed effects parameters
- Prediction of the response variable and of random effects
- likelihood function

## 1.3 Cook's Distance applied to LMEs

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

## 1.4 Iterative and non-iterative influence analysis for LMEs

**Cite: 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.

## 1.5 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.**Cite: Schabenberger** describes the choice between iterative influence analysis and non-iterative influence analysis.

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

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

**Information Ratio**

# 2 Measures 2

## 2.1 Cook's Distance

- For variance components $\gamma$

Diagnostic tool for variance components

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

## 2.2 Variance Ratio

- For fixed effect parameters $\beta$.

## 2.3 Cook-Weisberg statistic

- For fixed effect parameters $\beta$.

## 2.4 Andrews-Pregibon statistic

- For fixed effect parameters $\beta$.

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

## 2.5 Likelihood Distance

?

## 2.6 Diagnostics for LMEs with `R`

influence.ME: Tools for detecting influential data in mixed effects models

influence.ME provides a collection of tools for detecting influential cases in generalized mixed effects models. It analyses models that were estimated using lme4. The basic rationale behind identifying influential data is that when iteratively single units are omitted from the data, models based on these data should not produce substantially different estimates. To standardize the assessment of how influential a (single group of) observation(s) is, several measures of influence are common practice, such as DFBETAS and Cook's Distance. In addition, we provide a measure of percentage change of the fixed point estimates and a simple procedure to detect changing levels of significance.

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

An example model:

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

The function `influence` is the basis for all further steps:

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

Calculate Cook's distance:

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

Plot Cook's distance:

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

# 3 Case-Deletion Diagnostics for LMEs The CPJ Paper

## 3.1 Case-Deletion results for Variance components

**Cite: 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.

## 3.2 CPJ Notation

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

**Cite: CPJ** noted the following identity:

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

**Cite: 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 $\boldsymbol{H}_{[i]}^{-1}$