Influence.ME:
Tools for detecting influential data in mixed models

Rense Nieuwenhuis  //  Ben Pelzer  //  Manfred te Grotenhuis
A first indication something may go wrong ...
A first indication something may go wrong ...

Math score by Class Structure, by school

Average Math Test Score

Level of Class Structure
A first indication something may go wrong ...
A first indication something may go wrong ...

Math score by Class Structure, by school

Average Math Test Score

Level of Class Structure

Radboud University Nijmegen
A first indication something may go wrong ...
Mixed models in Social Sciences
Mixed models in Social Sciences

- Mixed, Multilevel, or Hierarchical Models
  - Observations nested within “groups”
  - Explanatory variables at all “levels”
Mixed models in Social Sciences

• Mixed, Multilevel, or Hierarchical Models
  • Observations nested within “groups”
  • Explanatory variables at all “levels”

• High-N Surveys
  • General Social Survey (n = 51,020)
  • World Value Survey (n = 267,870)
Mixed models in Social Sciences

- Mixed, Multilevel, or Hierarchical Models
  - Observations nested within “groups”
  - Explanatory variables at all “levels”

- High-N Surveys
  - General Social Survey (n = 51,020)
  - World Value Survey (n = 267,870)

- Small number of “groups” (van der Meer et al. 2009)
  - No country-comparative study exceeds 54 countries
  - Re-evaluation of risk for influential data
Measures of Influential Data
Measures of Influential Data

• Compare estimates *including* a particular case to the estimates *without* that particular case
  • In multilevel regression: case=group
Measures of Influential Data

- Compare estimates *including* a particular case to the estimates *without* that particular case
  - In multilevel regression: case=group

- **DFbetaS**: standardized difference in magnitude of single parameter estimate (Belsley et al., 1980)
Measures of Influential Data

• Compare estimates *including* a particular case to the estimates *without* that particular case
  • In multilevel regression: case=group

• **DFbetaS**: standardized difference in magnitude of single parameter estimate (Belsley et al., 1980)

• **Cook’s Distance**: standardized summary measure of influence on (one or) multiple parameter estimates (Cook 1977, Belsley et al., 1980)
Measures of Influential Data

- Compare estimates including a particular case to the estimates without that particular case
  - In multilevel regression: case=group

- **DFbetaS**: standardized difference in magnitude of single parameter estimate
  (Belsley et al., 1980)

- **Cook’s Distance**: standardized summary measure of influence on (one or) multiple parameter estimates
  (Cook 1977, Belsley et al., 1980)

- Improvement in influence.ME: cases not deleted, but influence neutralized by altered intercept + dummy variable
  (Langford & Lewis, 1998)
Influence.ME: Analytical Steps
Influence.ME: Analytical Steps

Original model
Influence.ME: Analytical Steps

Original model → estex()

Estimates without influence group 'j'
Influence.ME: Analytical Steps

Original model

estex()

Estimates without influence group 'j'

ME.cook()
ME.dfbetas()

No influential data? Correct(ed) model

Identification of influential data
Influence.ME: Analytical Steps

Original model → estex() → ME.cook() / ME.dfbetas() → exclude.influence()

Estimates without influence group 'j'

Corrected model to re-check → Identification of influential data

No influential data? Correct(ed) model
Again, a first indication something is wrong ...
Example: School 23  (Kreft & De Leeuw, 1998)

Linear mixed model fit by REML
Formula: math ~ structure + (1 | school.ID)

Number of obs: 519, groups: school.ID, 23

Fixed effects:

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>60.002</td>
<td>5.853</td>
</tr>
<tr>
<td>structure</td>
<td>-2.343</td>
<td>1.456</td>
</tr>
</tbody>
</table>
Example: School 23  (Kreft & De Leeuw, 1998)

Linear mixed model fit by REML
Formula: math ~ structure + (1 | school.ID)

Number of obs: 519, groups: school.ID, 23

Fixed effects:

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>60.002</td>
<td>5.853</td>
<td>10.252</td>
<td></td>
</tr>
<tr>
<td>structure</td>
<td>-2.343</td>
<td>1.456</td>
<td>-1.609</td>
<td></td>
</tr>
</tbody>
</table>
Example: School 23  (Kreft & De Leeuw, 1998)

Linear mixed model fit by REML
Formula: math ~ structure + (1 | school.ID)

Number of obs: 519, groups: school.ID, 23

Fixed effects:

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>60.002</td>
<td>5.853</td>
<td>10.252</td>
</tr>
<tr>
<td>structure</td>
<td>-2.343</td>
<td>1.456</td>
<td>-1.609</td>
</tr>
</tbody>
</table>
Adjusted Model
Adjusted Model

> model.7472 <- exclude.influence(model.simple,
+   "school.ID",
+   "7472")
> model.7472 <- exclude.influence(model.simple, 
+   "school.ID",
+   "7472")

> model.62821 <- exclude.influence(model.7472, 
+   "school.ID",
+   "62821")
Adjusted Model

> model.7472 <- exclude.influence(model.simple, +  "school.ID", +  "7472")

> model.62821 <- exclude.influence(model.7472, +  "school.ID", +  "62821")

Fixed effects:

<table>
<thead>
<tr>
<th>Term</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept.alt</td>
<td>64.285</td>
<td>6.353</td>
<td>10.119</td>
<td></td>
</tr>
<tr>
<td>estex.62821</td>
<td>73.069</td>
<td>4.735</td>
<td>15.432</td>
<td></td>
</tr>
<tr>
<td>estex.7472</td>
<td>52.571</td>
<td>3.600</td>
<td>14.602</td>
<td></td>
</tr>
<tr>
<td>structure</td>
<td>-3.416</td>
<td>1.535</td>
<td>-2.226</td>
<td></td>
</tr>
</tbody>
</table>
## Adjusted Model

```r
> model.7472 <- exclude.influence(model.simple,
+   "school.ID",
+   "7472")

> model.62821 <- exclude.influence(model.7472,
+   "school.ID",
+   "62821")
```

### Fixed effects:

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept.alt</td>
<td>64.285</td>
<td>6.353</td>
<td>10.119</td>
<td></td>
</tr>
<tr>
<td>estex.62821</td>
<td>73.069</td>
<td>4.735</td>
<td>15.432</td>
<td></td>
</tr>
<tr>
<td>estex.7472</td>
<td>52.571</td>
<td>3.600</td>
<td>14.602</td>
<td></td>
</tr>
<tr>
<td>structure</td>
<td>-3.416</td>
<td>1.535</td>
<td>-2.226</td>
<td></td>
</tr>
</tbody>
</table>
Adjusted Model

> model.7472 <- exclude.influence(model.simple, +  "school.ID", +  "7472")

> model.62821 <- exclude.influence(model.7472, +  "school.ID", +  "62821")

Fixed effects:

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept.alt</td>
<td>64.285</td>
<td>6.353</td>
<td>10.119</td>
<td></td>
</tr>
<tr>
<td>estex.62821</td>
<td>73.069</td>
<td>4.735</td>
<td>15.432</td>
<td></td>
</tr>
<tr>
<td>estex.7472</td>
<td>52.571</td>
<td>3.600</td>
<td>14.602</td>
<td></td>
</tr>
<tr>
<td>structure</td>
<td>-3.416</td>
<td>1.535</td>
<td>-2.226</td>
<td></td>
</tr>
</tbody>
</table>
Adjusted Model

```r
> model.7472 <- exclude.influence(model.simple,
+   "school.ID",
+   "7472")

> model.62821 <- exclude.influence(model.7472,
+   "school.ID",
+   "62821")
```

Fixed effects:

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept.alt</td>
<td>64.285</td>
<td>6.353</td>
</tr>
<tr>
<td>estex.62821</td>
<td>73.069</td>
<td>4.735</td>
</tr>
<tr>
<td>estex.7472</td>
<td>52.571</td>
<td>3.600</td>
</tr>
<tr>
<td>structure</td>
<td>-3.416</td>
<td>1.535</td>
</tr>
</tbody>
</table>
Known Issues & Future Development
Known Issues & Future Development

- Modification of intercept
  - More difficult to converge
  - Fails with factor-variables in model
  - Solution: use delete=TRUE in estex()
Known Issues & Future Development

- Modification of intercept
  - More difficult to converge
  - Fails with factor-variables in model
  - Solution: use `delete=TRUE` in `estex()`

- Currently, only fixed effects
  - Measures of influence for random effects available
Known Issues & Future Development

- Modification of intercept
  - More difficult to converge
  - Fails with factor-variables in model
  - Solution: use `delete=TRUE` in `estex()`

- Currently, only fixed effects
  - Measures of influence for random effects available

- Can be highly computational intensive
  - split over multiple sessions / computers
Known Issues & Future Development

• Modification of intercept
  • More difficult to converge
  • Fails with factor-variables in model
  • Solution: use `delete=TRUE` in `estex()`

• Currently, only fixed effects
  • Measures of influence for random effects available

• Can be highly computational intensive
  • split over multiple sessions / computers

• Development continues in Rennes ...
  • Partial residual plots
http://www.rensenieuwenhuis.nl/r-project/influenceme/
Discussion on Influential Data in Sociology

• Original Article:

• Research Note:

• Response to Research Note:
References


DFBETAS: (Belsley et al., 1980)

$$dfbetas_{ij} = \frac{\hat{\gamma}_i - \gamma_i(\hat{-j})}{se(\gamma_i(\hat{-j}))}$$

Cutoff: \( \frac{2}{\sqrt{n}} \)

Cook’s distance: (Snijders & Berkhof, 2008)

$$C^0_{ij} = \frac{1}{r+1}(\hat{\gamma} - \gamma(\hat{-j}))' \sum_{F}^{-1}(\hat{\gamma} - \hat{\gamma}(\hat{-j}))$$

Cutoff: \( \frac{4}{n} \)