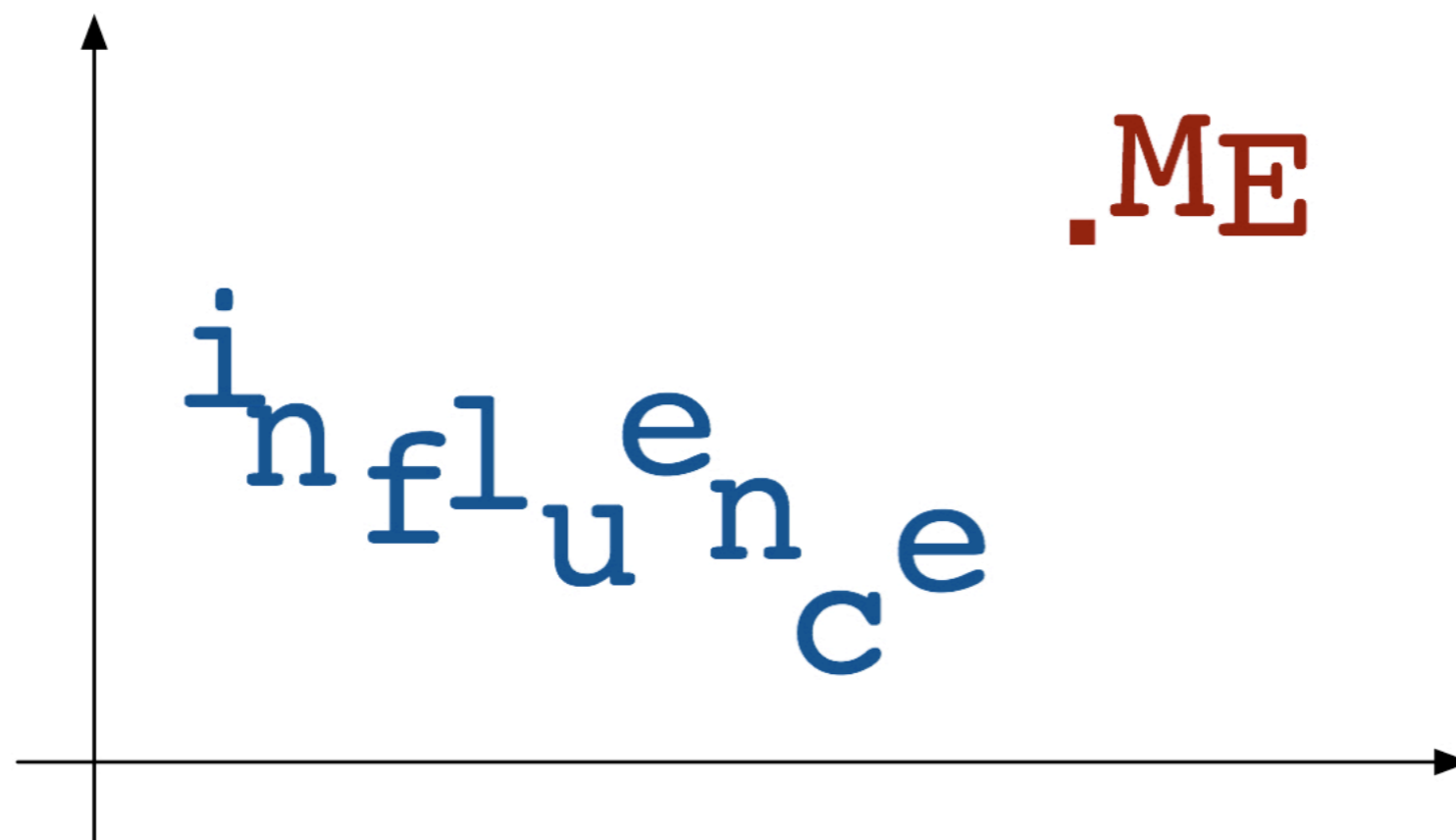


Influence.ME:

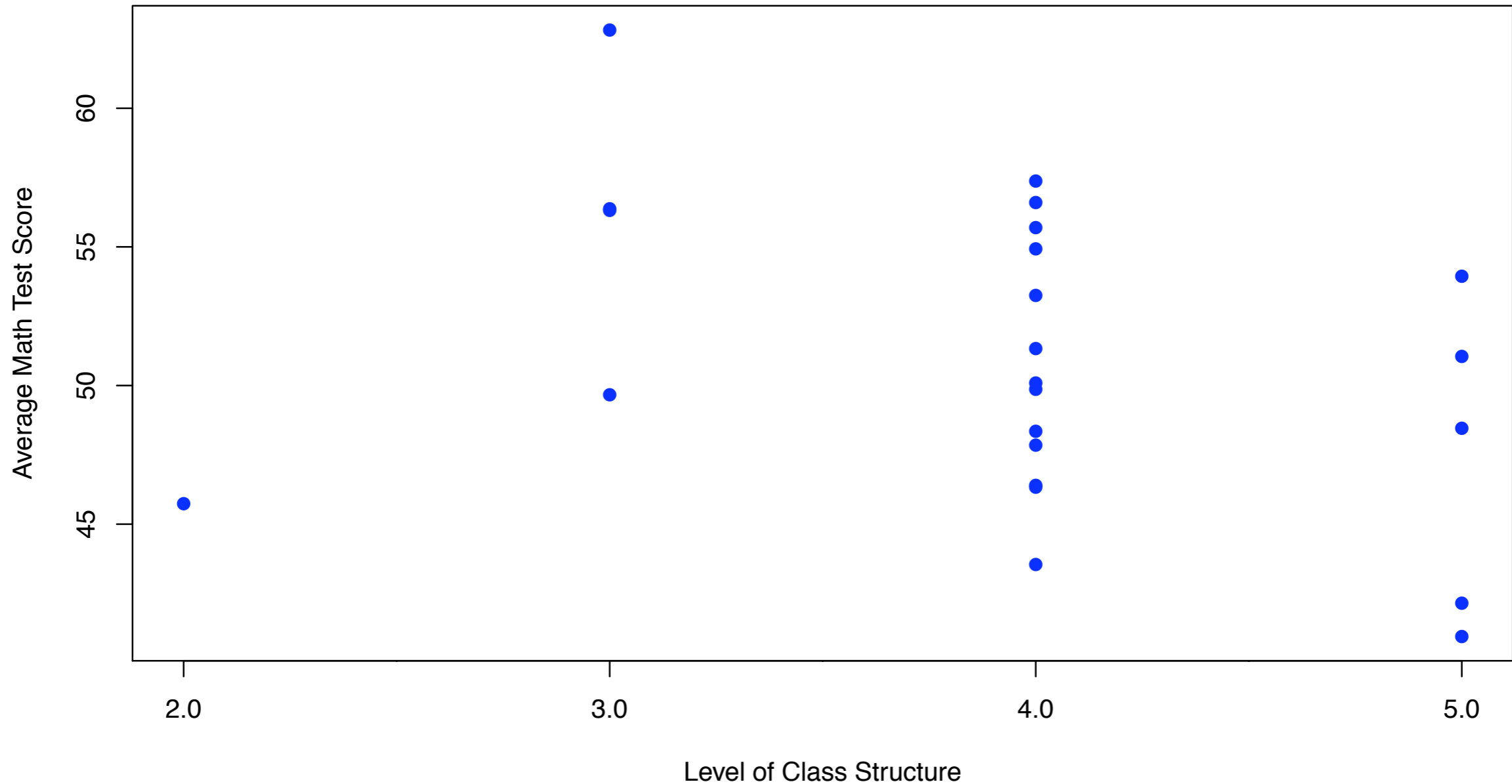
Tools for detecting influential data in mixed models

Rense Nieuwenhuis // Ben Pelzer // Manfred te Grotenhuis



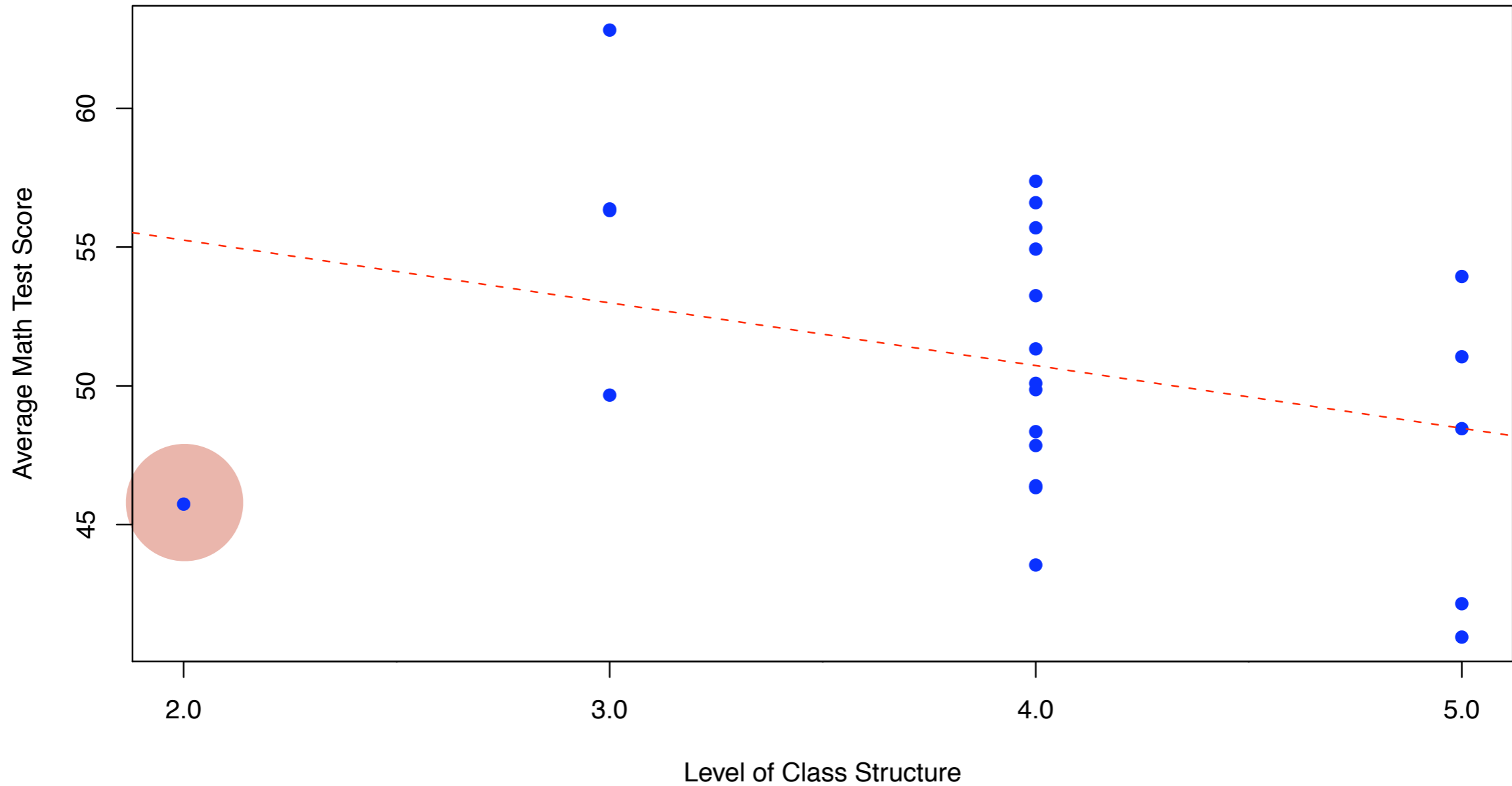
A first indication something may go wrong ...

Math score by Class Structure, by school



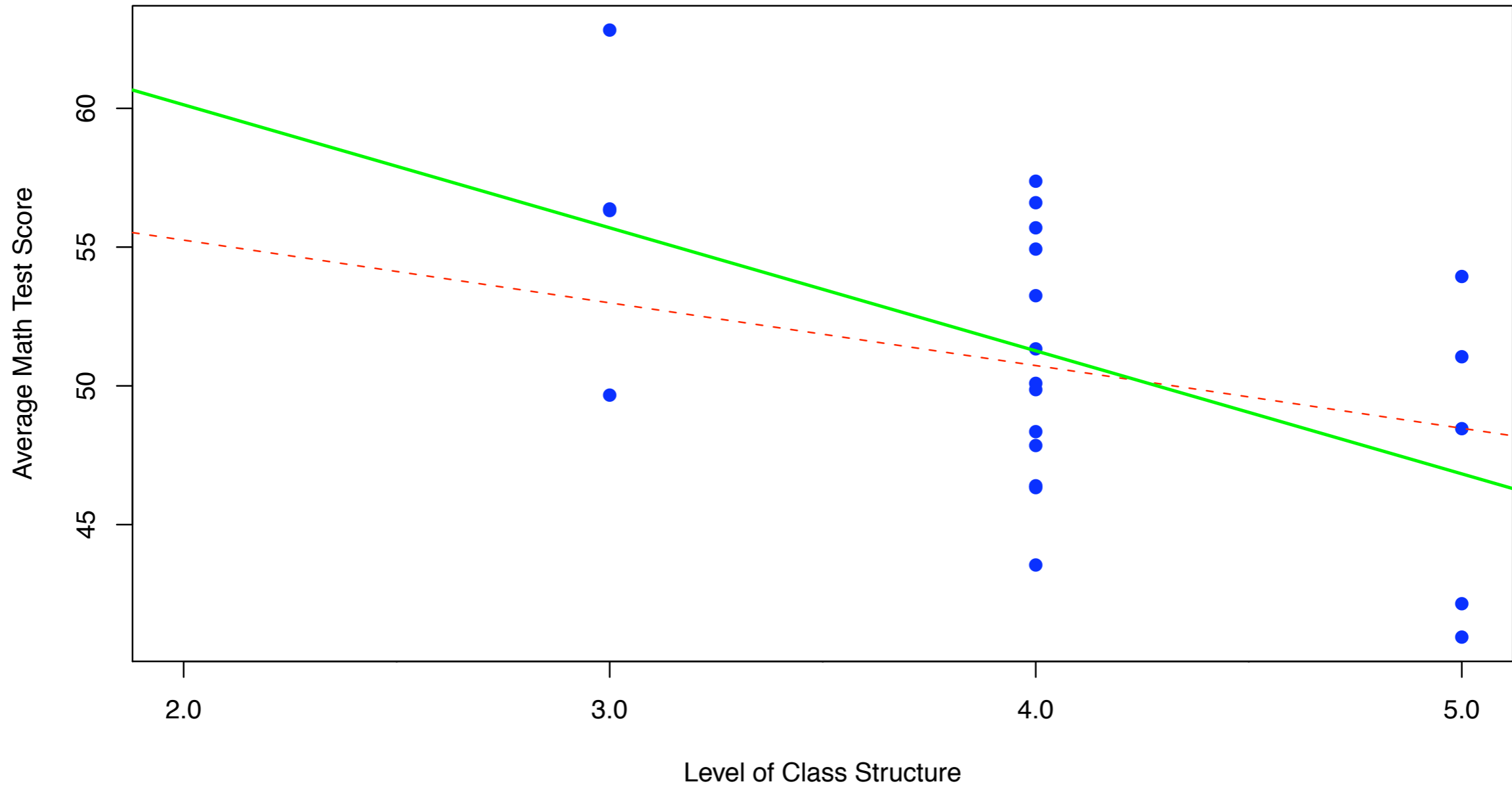
A first indication something may go wrong ...

Math score by Class Structure, by school



A first indication something may go wrong ...

Math score by Class Structure, by school



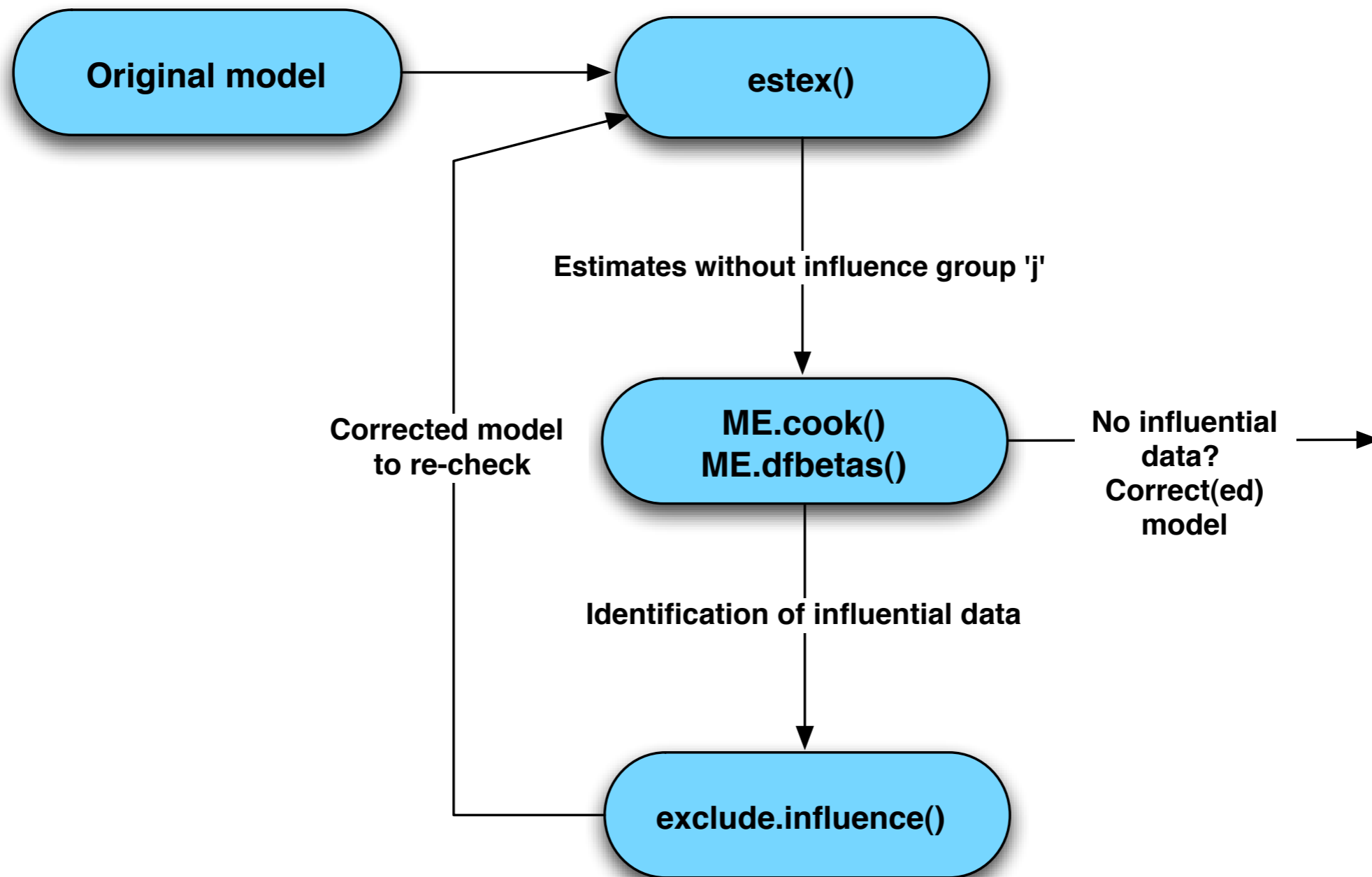
Mixed Models in Social Sciences

- 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

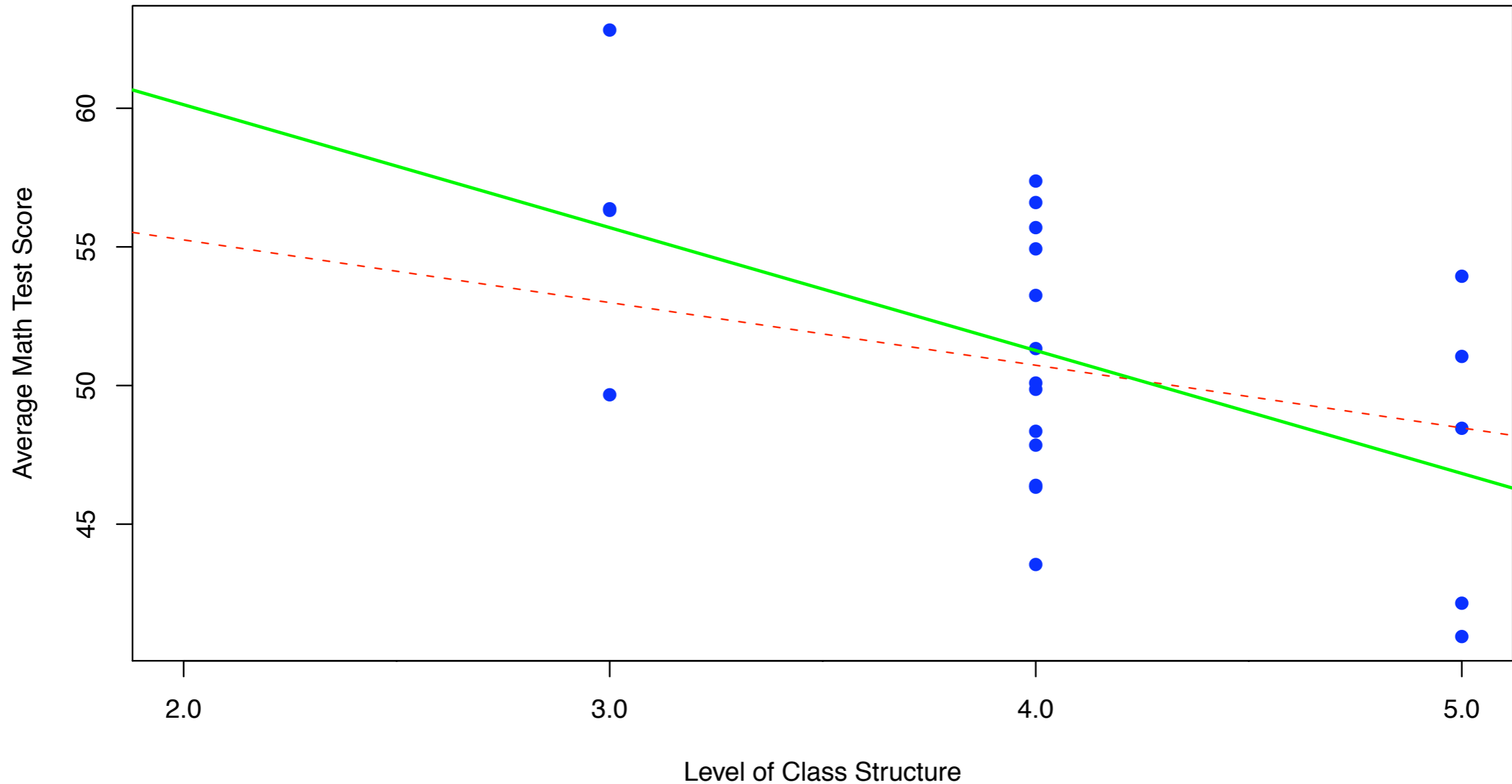
- Compare estimates *including* a particular case to the estimates *without* that particular case
 - In multilevel regression: case=group
- **DFbetaS**: standardized difference in magnitude of single parameter estimate (Belsley et al., 1980)
- **Cook's Distance**: standardized summary measure of influence 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



Again, a first indication something is wrong ...

Math score by Class Structure, by school



Example: School 23 (Kreft & De Leeuw, 1998)

Linear mixed model fit by REML

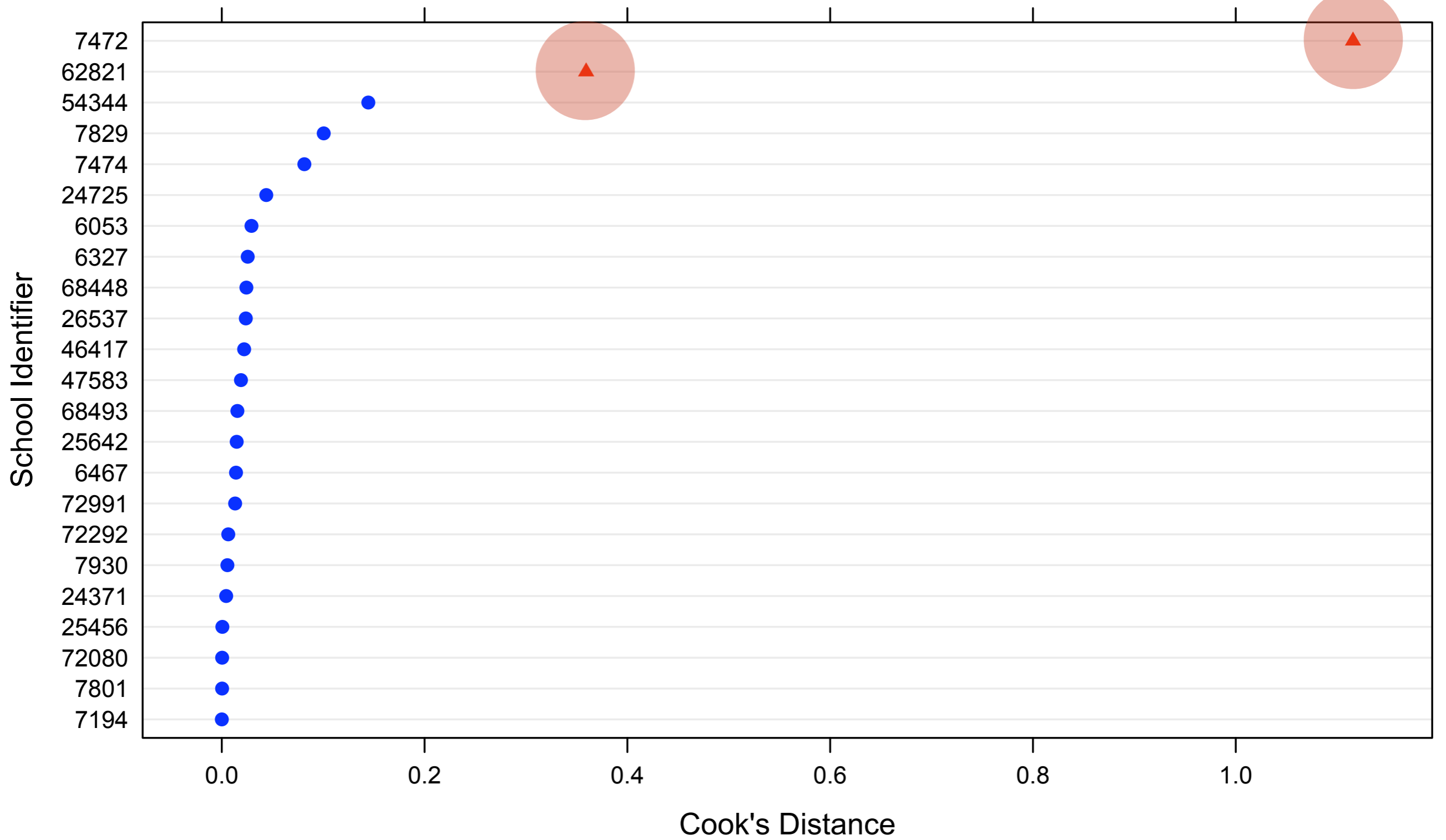
Formula: `math ~ structure + (1 | school.ID)`

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

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	60.002	5.853	10.252
structure	-2.343	1.456	-1.609

Cook's Distances



Adjusted Model

- Exclude influence school number 7472:

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

- Run `estex()` and `ME.cook()` again, then evaluate.

We found school 62821 to be overly influential

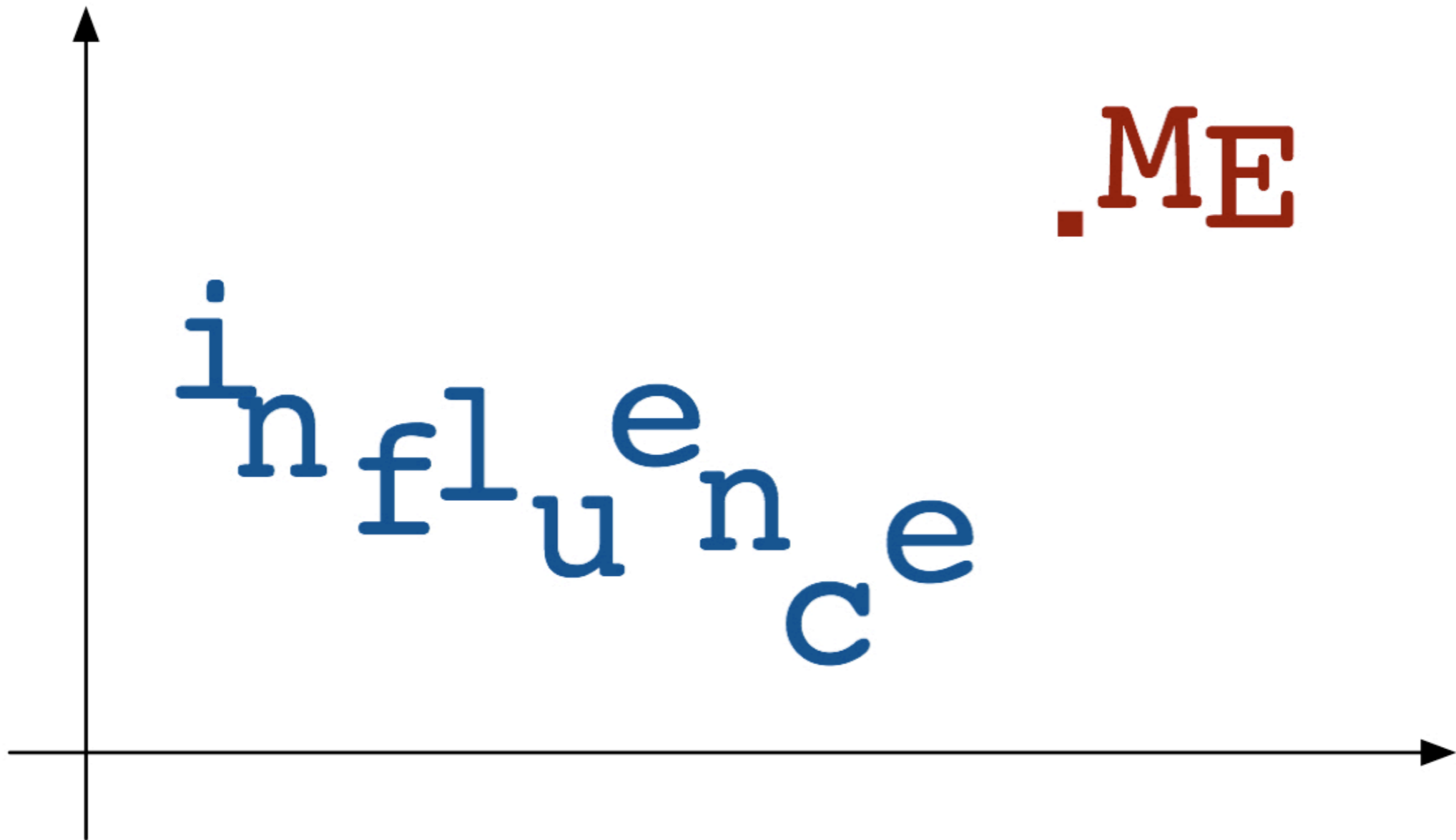
```
> model.62821 <- exclude.influence(model.7472,  
+ "school.ID",  
+ "62821")
```

Fixed effects:

	Estimate	Std. Error	t value
intercept.alt	64.285	6.353	10.119
estex.62821	73.069	4.735	15.432
estex.7472	52.571	3.600	14.602
structure	-3.416	1.535	-2.226

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 Multilevel Models in Sociology

- **Original Article:**

- **Ruiter**, Stijn and **De Graaf**, Nan Dirk. 2006. National context, religiosity, and volunteering: results from 53 countries. *American Sociological Review* 71: 191-210.

- **Research Note:**

- **Meer**, T. van der, te **Grotenhuis**, M., and **Pelzer**, B. (2010). Influential cases in multilevel modeling. a methodological comment on Ruiter and de Graaf (asr, 2006). *American Sociological Review*, accepted for publication.

- **Response to Research Note:**

- **Ruiter**, Stijn and **De Graaf**, Nan Dirk. (2010). National Religious Context and Volunteering: More Rigorous Tests Supporting the Association. *American Sociological Review*, accepted for publication.

References

- **Bates**, D., **Maechler**, M., and **Dai**, B. (2008). *lme4: Linear mixed-effects models using Eigen and Eigenpack*. R package version 0.999375-28.
- **Belsley**, D. A., **Kuh**, E., and **Welsch**, R. E. (1980). *Regression Diagnostics. Identifying Influential Data and Sources of Collinearity*. Wiley.
- **Cook**, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, 19(1):15–18.
- **Kreft**, I. and **De Leeuw**, J. (1998). *Introducing Multilevel Modelling*. Sage Publications.
- **Langford**, I. H. and **Lewis**, T. (1998). Outliers in multilevel data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 161:121–160.
- **Nieuwenhuis**, R., **Pelzer**, B., and **Te Grotenhuis**, M. (2009). *influence.ME: Tools for detecting influential data in mixed models*. R package version 0.7.
- **Meer**, T. van der, **te Grotenhuis**, M., and **Pelzer**, B. (2010). Influential cases in multilevel modeling. a methodological comment on ruiters and de graaf (asr, 2006). *American Sociological Review*, accepted for publication.
- **Snijders**, T. A. and **Berkhof**, J. (2008). Diagnostic checks for multilevel models. In De Leeuw, J. and Meijer, E., editors, *Handbook of Multilevel Analysis*, chapter 3, pages 141–175. Springer.

Formulae

DFBETAS: (Belsley et al., 1980)

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

Cutoff: $2/\sqrt{n}$

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

$$C_j^{0F} = \frac{1}{r+1} (\hat{\gamma} - \gamma(\hat{-j}))' \hat{\Sigma}_F^{-1} (\hat{\gamma} - \hat{\gamma}(\hat{-j}))$$

Cutoff: $\frac{4}{n}$