

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

Rense Nieuwenhuis^{1,*}, Ben Pelzer¹, Manfred te Grotenhuis¹

1. Radboud University, The Netherlands

* Corresponding author. E-mail: contact@rensenieuwenhuis.nl

Keywords: social sciences, influential cases, diagnostics, mixed effects models, lme4

Mixed effects regression models tend to become common practice in the field of Social Sciences. However, diagnostic tools to evaluate these models lag behind. For instance there is no general applicable tool to check whether all units (or cases) roughly have the same influence on the regression parameters. It is however commonly accepted that tests for influential cases should be performed, especially when estimates are based on a relatively small number of cases. Testing for influence with mixed effects models is especially important in Social Science applications, for two reasons. First, models in the Social Sciences are frequently based on large numbers of individuals while the number of higher level units is often relatively small. Secondly, often the higher level units are remarkably similar, for instance in the case of neighboring countries. **Influence.ME** is a new package for R which provides two innovations for evaluating influential cases: it extends existing procedures for use with mixed effects models, and it allows to not only search for single influential cases, but for combinations of cases that as a combination exert too much influence.

The basic rationale behind measuring influential cases 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 observation is, several measures of influence are common practice. First, DFBETAS is a standardized measure of the absolute difference between the estimate *with* a particular case included and the estimate *without* that particular case. Second, Cook's distance provides an overall measurement of the change in *all* parameter estimates, or a selection thereof.

To apply the same logic to mixed effects models one has to measure the influence of a particular higher level unit on the estimates of a higher level predictor. This means that the mixed effect model has to be adjusted to neutralize the unit's influence on that estimate, while at the same time allowing the unit's lower-level cases to help estimate the effects of the lower-level predictors in the model. This procedure is based on a modification of the intercept and the addition of a dummy variable for the cases that might be influential. **Influence.ME** provides several measures of influential cases, and is specifically designed for use with mixed effects regression models using the afore mentioned modified intercept and dummy approach. Using both 'real' and simulated data from Social Science applications of mixed effects models, five tools to detect influential cases which are available in the package will be discussed:

- Cook's Distance
- DFBETAS
- Percent change of the estimated parameter magnitude
- Changes in statistical significance of parameter estimates
- Changes in the sign of parameter estimates

In contrast with other algorithms for detecting influential cases, influence.ME is capable to uncover groups of cases that are influential. Since this rapidly becomes computationally highly intensive, additional script functions are provided that assist in manually dividing the computation into multiple sessions, or to possibly to share the computations between different computers.

References

- Belsley, D.A., Kuh, E. & Welsch, R.E. (1980). *Regression Diagnostics. Identifying Influential Data and Source of Collinearity*. Wiley.
- Snijders, T.A. & Bosker, R.J. (1999). *Multilevel Analysis, an introduction to basic and advanced multilevel modelling*. Sage.
- Van Der Meer, T., Te Grotenhuis, M. & Pelzer, B. *Influential cases in multi-level modeling. A methodological comment on 'National context, religiosity, and volunteering' by Ruiter and De Graaf*. Current status: resubmitted to the American Sociological Review.