Small samples and messy data: Time for a new approach?

Andrew R. Johnson[1], Andrea M. Loftus[1], Blake J. Lawrence[2,3], Natalie Gasson[1]

[1] School of Psychology, Curtin University, WA, Australia

[2] Ear Science Institute Australia, Subiaco, WA, Australia

[3] Ear Sciences Centre, Medical School, The University of Western Australia, Crawley, WA, Australia

Introduction

Limited samples available to be recruited are a common reality for experimental researchers. While this may not unduly influence testing, any resulting inference could be compromised if not accounted for in the analysis.[1] Because maximum-likelihood (ML) analyses rely parameters being normally distributed in large samples they tend to perform poorly in smaller, non-normal, samples.[2] This is especially a concern for pre-/post-test or otherwise clustered designs where correlated observations require more complex analyses.[2]

A promising alternative is the use of Bayesian estimation, which does not rely on large-sample normality. An increasingly popular approach is the use of weakly-informative priors, which provide equivalent estimates to ML-based analyses but perform far better in smaller samples.[3] The present study uses Montecarlo simulations and a real dataset to demonstrate this.

Materials and methods

Simulation

A Montecarlo simulation was used to compare the precision of a linear mixed model (LMM) when estimated with restricted maximum-likelihood (REML) and with a Bayesian estimator.

A dataset of 10 individuals, each observed five times was generated. The data were generated according the model of a single predictor affecting all in the same way (i.e. a fixed effect), but where each individual had a different intercept. These data were generated 1000 times, with the LMM (REML and Bayes) fitted to each.

The distribution of standard errors (SE) for the estimates of the intercept will indicate the level of precision for each method. The larger the SE of a given method, the larger the confidence (or credibility) intervals will be and the lower the precision.

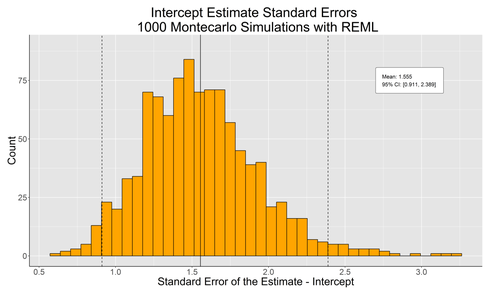
Real Data

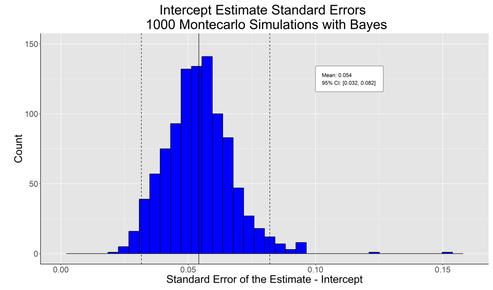
To demonstrate the utility of Bayesian methods, the same analyses were applied to the data collected in a randomized controlled trial of transcranial direct current stimulation (tDCS) and cognitive training in Parkinson's Disease (PD). [4]

The study comprised 42 individuals with PD who were randomized to one of six intervention groups. Individuals were assessed on executive functioning (as in the Stockings of Cambridge) both before and after the intervention, and at a 3-month follow-up. The full study has been published previously.[4]

All analyses were conducted with the 'lme4' and 'rstanarm' packages in R 3.4.4. Code for reproducing simulation results is available in the first author's GitHub.

Results





Simulation

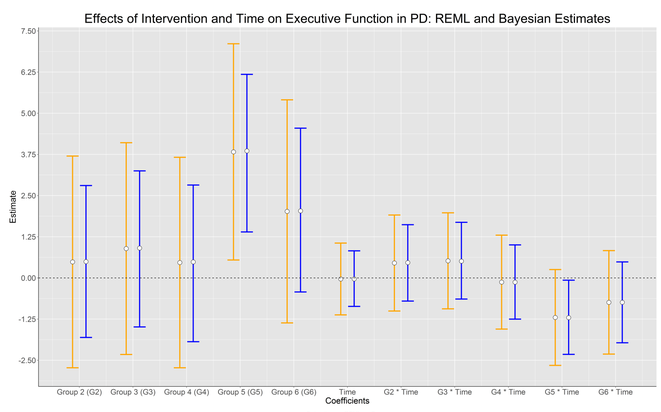
As depicted, the REML-estimated mixed model showed quite poor precision, on average, in the 1000 Montecarlo simulations when compared to the Bayesian-estimated equivalent. The REML model had mean SE of 1.555 across the simulations, markedly larger than the .054 mean SE of the Bayesian estimator. This indicates that, on average, the Bayesian-estimated mixed-model provided more precise estimates than a REML-estimated model.

The Bayesian-estimated model was also far more consistent in its level of precision. In the Bayesian model, 95% of the estimates fell in the range [.032, .082]. This is clearly smaller than the range of estimated SE in the REML model: [.911, 2.389]. This indicates that the Bayesian mixed model had a similar level of precision in each of the 1000 datasets, whereas precision of the REML-estimated model varied greatly throughout.

Real Data

The fixed-effect, random-intercept model was then applied to the tDCS and cognitive training RCT data. The estimates for the effect of each intervention group (with the first used as a reference group), and their confidence or credibility intervals are plotted adjacent.

It is clear that while the actual estimates are near-identical, the Bayesian credibility intervals are much narrower than those of the REML estimator. This is best evidenced by the coefficient for 'G5*Time' (i.e. the progression of Intervention Group 5 over time as compared to that of Group 1). The Bayesian credibility intervals do not cross 0, indicating the presence of some effect. However, the reduced precision of the REML model did not provide the same result.



Conclusions

The study aimed to compare the precision of a linear mixed model when estimated with either restricted maximum-likelihood or Bayes in a small sample. The study identified that the Bayesian estimator was both more precise on average, as well as more consistent in its level of precision than the REML estimator.

When applying these findings to data collected from a RCT in Parkinson's Disease, the results identified a relationship that was not able to be seen when the REML estimator was used.

Overall, the findings of this study suggest that a Bayesian analysis would be the most appropriate for experimental researchers when faced with a limited sample size. Given the increasing ubiquity and accessibility of Bayesian methods, it is becoming far easier for researchers without advanced statistical training to be able to use these types of models.

References

1. Maas CJM, Hox JJ. Sufficient Sample Sizes for Multilevel Modeling. Methodology: European Journal of Research Methods for the Behavioral and Social Sciences 2005;1:86-92.

2. McNeish DM, Stapleton LM. The Effect of Small Sample Size on Two-Level Model Estimates: A Review and Illustration. Educational Psychology Review 2016;28:295-314.

3. McNeish D. On Using Bayesian Methods to Address Small Sample Problems. Structural Equation Modeling: A Multidisciplinary Journal 2016;23:750-773.

4. Lawrence BJ, Gasson N, Johnson AR, Booth L, Loftus AM. Cognitive Training and Transcranial Direct Current Stimulation for Mild Cognitive Impairment in Parkinson's Disease: A Randomized Controlled Trial. Parkinson's Disease 2018;2018:12.