Analyzing Likert-scale data with mixed-effects linear models - a simulation study

Johannes Kizach

University of Aarhus, English Degree Programme, engjk@dac.au.dk

Introduction

Some argue that it is inappropriate to use linear mixed-effects models (LMEMs) to analyze Likert-scale data, but Norman (2010) argues that alternative tests are unnecessary since LMEMs are precise even when the assumptions are violated. Norman's (2010) recommendation is to use parametric tests (e.g. LMEMs) when analyzing Likert-scale data, and Gibson, Piantadosi, and Fedorenko (2011) likewise recommend using these tests when analyzing Likert-scale data.

The current study demonstrates through two simulations that this practice (i.e. analyzing Likert-scale data using LMEMs) does not result in high Type I error rates (cf. Norman 2010).

A second set of simulations shows that a cumulative probit mixed model (CPMM), which is recommended for Likert-scale data (cf. Agresti 2002:283), has a considerably higher Type I error rate when compared to LMEMs.

Simulation procedure

Simulation 1 and 2

First, a realistic data set with 24 items, subjects, two conditions, random noise, intercepts for items and subjects, and slopes for subjects was generated.

Then the data were analyzed using a LMEM with random intercepts for subjects and items, and slopes for subjects following the advice in Barr, Levy, Scheepers, & Tily (2013).

This procedure was repeated (with the number of subjects varying from 10 to 100, in steps of 10). In each case the p-values were saved. The percentage of p-values equal to or lower than 0.05 was then extracted to estimate the precision at alpha-level 0.05, and likewise the p-values equal to or lower than 0.01 were extracted to estimate the precision at the 0.01 alpha-level.

Ideally the models should find an effect of condition in 5% of the cases at the 0.05 alpha-level, and in 1% of the cases at the 0.01 alpha-level.

Simulation 3 and 4

The procedure was very similar to the one described above, but the models used were different.

In simulation 3 a CPMM was used, and since the ordinal package for R (Christensen 2013) has not implemented random slopes, only random intercepts were included in the model.

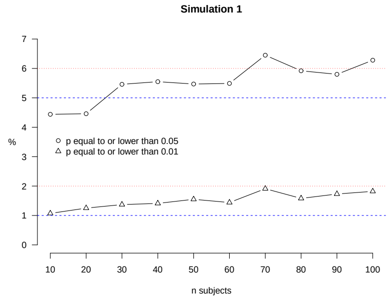
In simulation 4 a LMEM with random intercepts only was used to allow direct comparison with the CPMM.

In all simulations the software R (R Development Core Team 2009) and the packages lme4 (Bates, Maechler, & Bolker 2011) and languageR (Baayen 2011) were used.

Results 1, 2

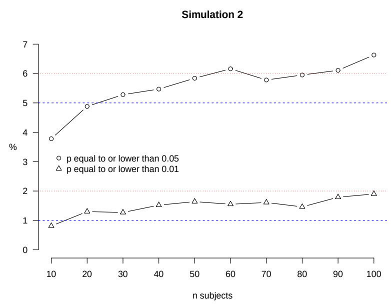
Simulation 1

Here the simulated Likert-scale data ranged from 1 to 7. The simulation of 100.000 datasets showed that from 30 to 60 subjects the LMEM found an effect between 5.5 and 5.6% of the cases at alpha 0.05. At alpha level 0.01 effects were found in 1 to 1.9% of the cases.



Simulation 2

Everything was done exactly as in simulation 1, but now the simulated data was on a 5-point Likert-scale. The results showed that the LMEM does fairly well at alpha 0.05, when there are more than 20 subjects (note that Barr et al. 2013:268 consider a percentage of less than 10 an acceptable result). At alpha 0.01 the LMEM is below the 1% with 10 subjects, but with 20 or more subjects, effects are found in 1.3 to 1.9% of the cases.



Results 3, 4

Simulation 3

6000 datasets with a 7-point Likert-scale response variable were simulated (1000 with 10 subjects, 1000 with 20, etc. up to 60 subjects) and analyzed with a CPMM. The result showed that the model had a higher Type I error rate than the results seen in simulations 1 and 2, ranging from 7.5 to 9.5% at the 0.05 alpha-level, and between 2 and 3.5% at the 0.01 alpha level.

Simulation 4

This was parallel to simulation 3 but with a random intercepts only LMEM, to allow direct comparison with the CPMM results. The random intercepts only model does worse than the maximal models in simulation 1 and 2, but it is better than the CPMM, with Type I error rates between 6 to 7.6% at the 0.05 alpha-level, and between 1.1 and 2.8% at the 0.01 alpha level.

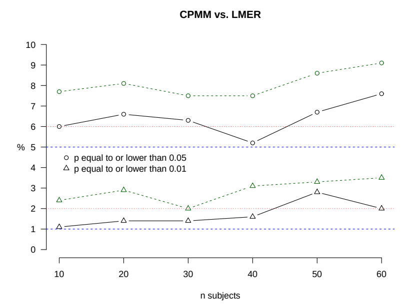
Conclusion

The conclusion is that using LMEMs on Likert-scale data does not increase the Type I error rate substantially and, in fact, it is worse to use a CPMM. In short: Go ahead and use a LMEM on your Likert-scale data!

Results 3, 4

Simulation 3 and 4

The CPMM results are depicted with stippled, green lines, and the LMEM results are shown with solid, black lines:



Example of R-code used

```
Simulation1<-matrix(NA,10,2)

for (j in 1:10) {
tal<-seq(10, 100, 10)
Krypto<-rep(NA, 10000)
  for (i in 1:10000) {

  subject<-as.factor(subject<-rep(1:tal[j],  each=16))
  cond<-rep(c("A","B"),length(subject)/2)
  sub_int<-rnorm(length(unique(subject)),mean=0,sd=2)
  sub_int<-sub_int[subject]
  resp<-sample(1:7,length(subject),replace=T)
  noise<-rnorm(length(subject), mean=0)
  item<-as.factor(item<-rep(1:16, length(unique(subject))))
  item_int<-rnorm(16, mean=0, sd=1)
  item_int<-item_int[item]
  simu<-data.frame(subject, cond, resp, sub_int,item_int, noise, item)
  simu<-simu[order(simu$subject, simu$cond),]
  sub_slo<-rnorm(length(unique(subject)), mean=0, sd=0.4)
  simu$sub_slo<-rep(sub_slo, each=8)
  simu$response<-simu$resp+simu$sub_int+simu$item_int+simu$sub_slo+simu$noise
  simu$response[simu$response>7]<-sample(1:7, 1)
  simu$response[simu$response<1]<-sample(1:7, 1)
  simu$response<-round(simu$response)
  simumod<-lmer(response~cond+(1+cond|subject)+(1|item), simu)
  Krypto[i]<-pvals.fnc(simumod, nsim=0)$fixed[2,2]
  }
  Simulation1 [j,1]<-table(Krypto<=0.01)[2]
  Simulation1 [j,2]<-table(Krypto<=0.05)[2]
 }
```

References

Agresti, A. (2002) *Categorical Data Analysis*. John Wiley & Sons.

Baayen, R.H. (2011) *LanguageR*. R package. Retrieved from http://CRAN.R-project.org/package=languageR

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013) Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278.

Bates, D., Maechler, M., and B. Bolker. (2011) *lme4: Linear mixed-effects models using S4 classes*. R package version 0.999375-42. Retrieved from http://CRAN.R-project.org/package=lme4

Christensen, R.H.B. (2013) *ordinal: Regression Models for Ordinal Data*. http://CRAN.R-project.org/web/packages/ordinal/index.html

Gibson, E., Piantadosi, S., & Fedorenko, K. (2011) Using Mechanical Turk to Obtain and Analyze English Acceptability Judgments: Linguistic Acceptability on Mechanical Turk. *Language and Linguistics Compass*, 5(8), 509–524.

R Development Core Team. (2009) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Retrieved from http://www.R-project.org.

Norman, G. (2010) Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education*, 15(5), 625–632.

Acknowledgments

The study was financed by the Danish Council for Independent Research. Thanks are due to Michael Krabbe Borregaard, Ken Ramshøj Christensen , Doug Saddy, Johanna Wood and two anonymous reviewers for helpful comments, and to Bodo Winter for advice on the coding.