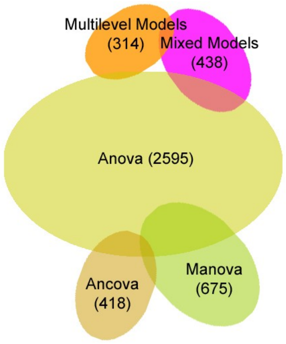
Why mixed models are so few used?

Massimo Borelli, Sara Cervai, Piero Paolo Battaglini

psiqu.units.it – Psychology and Quality – B.R.A.I.N. – University of Trieste – Italy

Abstract. A simple research in the most known databases of scientific literature in psychology (PsycINFO, PsycLit) shows that just the 4% of the empirical studies uses a mixed model/multilevel approach. On the other side more than 60% still uses a linear regression approach, often in an uncritical way. Despite the Generalized Linear Models (GLM) have been published in the early '70 and their strengths have repeatedly been demonstrated in different fields, their use is still neglected. The present paper aims to clarify their effectiveness in the psychological field and how they can overcome linear regression limits.

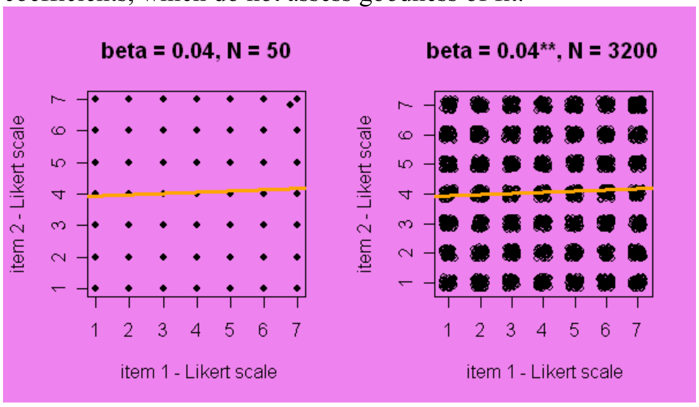
Studies rarely use mixed model / multilevel approach



By querying, for instance, the database PsycINFO and focusing the attention on research articles published on peer reviewed journals during the period going from January 2000 to December 2007 (approxim. 0.61×10^6 files) the odds to find a mixed model / multilevel analysis versus the classical multivariate approach (Anova, Manova, Ancova, ...) is still about one to five.

Very often, the linear regression approach is uncritical

Ordinary least square linear regression requires the residuals to be normally distributed with zero mean and constant variance (homoskedasticity). Despite the fact that software packages offer user-friendly diagnostic tools (residual plot, QQplot, Cook distance,...), scholars often neglect those information, focusing only on "starred" regression coefficients, which do not assess goodness of fit.



left	Estimate	Std. Error	*t* value	Pr(>|*t*|)
(Intercept)	3.89	0.65	5.97	<2.7 e-07
item 1	0.04	0.14	0.26	**0.80**

right	Estimate	Std. Error	*t* value	Pr(>|*t*|)
(Intercept)	3.89	0.08	48.48	<2 e-16
item 1	0.04	0.02	2.44	**0.01**

Both those artificial datasets have not normally distributed residuals and they share the same negligeable effect size ($R^2 = 0.002$), but in the right panel the "starred" beta would lead the unaware researcher to claim that item #1 is a very high significant (linear) predictor for item #2.

Anova requires heavy mathematical hypotheses

Anova analysis requires some *a priori* hypotheses: the normality, the homoskedasticity and the indipendence between samples. Violation to one or more of those assumptions can lead to wrong conclusions. Further difficulties are also connected with mathematical aspects:

- Sphericity condition (structure of variance-covariance matrix)
- Design effect (intra-class correlation)
- Missing data

Anova can not model more than two levels

Anova analysis can not manage more than two levels in data clusters. Choosing to melt clusters or subjects (*aggregated analysis*) or to ignore the existent hierarchy (*disaggregated analysis*) can lead to:

- Shift of meaning
- Ecological fallacy (aggregation bias)
- Interaction between levels
- Inference on the clusters
- Wrong sample size for the macro variables
- Dependence (biased standard errors)

Mixed model / multilevel approach can provide a solution

Item response theory nowadays can be conceptualized in the framework of nonlinear mixed models, and state-of-art computational algorithms can provide correct (although time consuming) solutions in estimating multilevel factor models for ordinal variables. Anyway, method implementation is not straightforward, and some "road show" seems to be necessary to enlarge scholars knowledge. Our future aim will be to realize an *ad hoc* R library to fully analyze ordinal variables through a mixed model approach.

Selected references

- Geert Verbeke, Geert Molenberghs, *Linear Mixed Models for Longitudinal Data*, Springer, 2001.
- Frank Rijmen, Francis Tuerlinckx, Paul De Boeck, Peter Kuppens (2003). *A Nonlinear Mixed Model Framework for Item Response Theory*, Psychological Methods, **8**(2), 185–205.
- Geert Molenberghs, Geert Verbeke, *Models for Discrete Longitudinal Data*, Springer, 2005.
- Michael J. Crawley, *The R Book*, John Wiley & Sons, Chichester, 2007.
- Leonardo Grilli, Carla Rampichini (2007). *Multilevel Factor Models for Ordinal Variables*, Structural Equation Modeling, **14**(1), 1–25.