Introduction to multilevel modeling

Thom Baguley and Mark Andrews
Nottingham Trent University

Overview

1. Nested models for repeated measures or clustered data
2. Estimation and inference
3. Random slopes
4. Fully crossed models
5. Fitting maximal random effects structures

1. Nested models for repeated measures or clustered data

Repeated measures ANOVA

Usual practice in psychology is to analyse repeated measures data using ANOVA:

Oneway independent measures ANOVA

$$Y_{ij} = \mu + \tau_j + \varepsilon_{ij} \quad \sum \tau_j = 0 \quad \varepsilon_{ij} \sim N\left(0, \sigma_\varepsilon^2\right)$$

Repeated measures ANOVA

Usual practice in psychology is to analyse repeated measures data using ANOVA:

Oneway independent measures ANOVA

$$Y_{ij} = \mu + \tau_j + \varepsilon_{ij} \quad \sum \tau_j = 0 \quad \varepsilon_{ij} \sim N\left(0, \sigma_\varepsilon^2\right)$$

Oneway repeated measures ANOVA

$$Y_{ij} = \mu + \pi_i + \tau_j + \varepsilon_{ij} \quad \sum \tau_j = 0 \quad \sum \pi_i = 0 \quad \varepsilon_{ij} \sim N\left(0, \sigma_\varepsilon^2\right)$$

Limitations of standard approaches

- sphericity or multi-sample sphericity assumed

Limitations of standard approaches

- sphericity or multi-sample sphericity assumed
- predictors must be orthogonal

Limitations of standard approaches

- sphericity or multi-sample sphericity assumed
- predictors must be orthogonal

  (e.g., no time-varying covariates)

Limitations of standard approaches

- sphericity or multi-sample sphericity assumed
- predictors must be orthogonal
  (e.g., no time-varying covariates)
- missing values not handled well

Limitations of standard approaches

- sphericity or multi-sample sphericity assumed
- predictors must be orthogonal
  (e.g., no time-varying covariates)
- missing values not handled well
- stimuli treated as fixed effect (see Clark, 1973)

Limitations of standard approaches

- sphericity or multi-sample sphericity assumed
- predictors must be orthogonal
  (e.g., no time-varying covariates)
- missing values not handled well
- stimuli treated as fixed effect (see Clark, 1973)
- aggregation or disaggregation of effects at higher or level

Multilevel models with random intercepts

Single level regression model (i.e., with fixed intercept)

$$Y_i = b_0 + b_1 X_{1i} + e_i \qquad e_i \sim N\left(0, \sigma_e^2\right)$$

Multilevel models with random intercepts

Single level regression model (i.e., with fixed intercept)

$$Y_i = b_0 + b_1 X_{1i} + e_i \qquad e_i \sim N\left(0, \sigma_e^2\right)$$

Random intercept model

$$Y_{ij} = b_{0j} + b_1 X_{1i} + u_j + e_{ij} \quad u_j \sim N\left(0, \sigma_u^2\right) \;\; e_{ij} \sim N\left(0, \sigma_e^2\right)$$

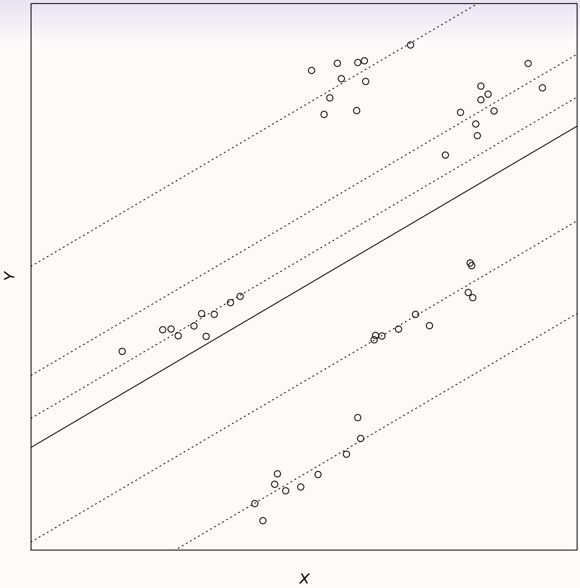
Multilevel models with random intercepts

Single level regression model (i.e., with fixed intercept)

$$Y_i = b_0 + b_1 X_{1i} + e_i \quad e_i \sim N\left(0, \sigma_e^2\right)$$

Random intercept model

$$Y_{ij} = b_{0j} + b_1 X_{1i} + u_j + e_{ij} \quad u_j \sim N\left(0, \sigma_u^2\right) \; e_{ij} \sim N\left(0, \sigma_e^2\right)$$



Random intercept model

Multilevel models with random intercepts

Single level regression model (i.e., with fixed intercept)

$$Y_i = b_0 + b_1 X_{1i} + e_i \quad e_i \sim N\left(0, \sigma_e^2\right)$$

Random intercept model

$$Y_{ij} = b_{0j} + b_1 X_{1i} + u_j + e_{ij} \quad u_j \sim N\left(0, \sigma_u^2\right) \; e_{ij} \sim N\left(0, \sigma_e^2\right)$$

Example: Voice and face processing

Does male (n=30) voice pitch *rise* or *fall* when rating an attractive face?

Example: Voice and face processing

Does male (n=30) voice pitch *rise* or *fall* when rating an attractive face?

Time-varying covariate (baseline pitch for speaking each rating)

Exercise 1

- load the pitch data into R
- fit a two-level random intercept model using `lmer()`
- fit a three-level nested random intercept model using `lmer()`

2. Estimation and inference

Estimation in multilevel models

Estimation is iterative and usually uses maximum likelihood based approaches:

- Full maximum likelihood (ML)
- Restricted maximum likelihood (REML)
- Markov chain Monte Carlo methods (MCMC)

Comparing models

Confidence intervals and tests:

- deviance (likelihood ratio) tests

Comparing models

Confidence intervals and tests:

- deviance (likelihood ratio) tests
  (change in -2LL has an approximate χ^2 distribution)
- Wald tests and CIs

Comparing models

Confidence intervals and tests:

- deviance (likelihood ratio) tests

  (change in -2LL has an approximate χ^2 distribution)
- Wald tests and CIs

  (estimate/SE has approximate *z* distribution)

Comparing models

Confidence intervals and tests:

- deviance (likelihood ratio) tests

  (change in -2LL has an approximate χ^2 distribution)
- Wald tests and CIs

  (estimate/SE has approximate z distribution)
- bootstrapping

Comparing models

Confidence intervals and tests:

- deviance (likelihood ratio) tests
  (change in -2LL has an approximate χ^2 distribution)
- Wald tests and CIs
  (estimate/SE has approximate *z* distribution)
- bootstrapping
- MCMC methods (e.g., HPD intervals)

Comparing models

Confidence intervals and tests:

- deviance (likelihood ratio) tests
  (change in -2LL has an approximate χ^2 distribution)
- Wald tests and CIs
  (estimate/SE has approximate z distribution)
- bootstrapping
- MCMC methods (e.g., HPD intervals)

Information criteria

- AIC, BIC or (MCMC derived DIC and WAIC)

Comparing models

Confidence intervals and tests:

- deviance (likelihood ratio) tests
  (change in -2LL has an approximate χ^2 distribution)
- Wald tests and CIs
  (estimate/SE has approximate *z* distribution)
- bootstrapping
- MCMC methods (e.g., HPD intervals)

Information criteria

- AIC, BIC or (MCMC derived DIC and WAIC)
  (-2LL with a penalty for number of parameters)

Accurate inference ...

- for standard repeated measures ANOVA models it is possible to use t and F statistics
- if a complex covariance structure or unbalanced model this may be problematic owing to:
  a) difficulty estimating the error *df*
  b) boundary effects for variance estimates

Possible solutions

- asymptotic approximations (in large samples)

Possible solutions

- asymptotic approximations (in large samples)
- corrections such as the Kenwood-Rogers approximation

Possible solutions

- asymptotic approximations (in large samples)
- corrections such as the Kenwood-Rogers approximation

(e.g., using `pbkrtest`)

Possible solutions

- asymptotic approximations (in large samples)
- corrections such as the Kenwood-Rogers approximation (e.g., using `pbkrtest`)
- bootstrapping

Possible solutions

- asymptotic approximations (in large samples)
- corrections such as the Kenwood-Rogers approximation (e.g., using `pbkrtest`)
- bootstrapping
- MCMC estimation

Possible solutions

- asymptotic approximations (in large samples)
- corrections such as the Kenwood-Rogers approximation (e.g., using `pbkrtest`)
- bootstrapping
- MCMC estimation (e.g., using `MCMCglmm`)

... with MCMC methods being the preferred approach (being generally both safe and versatile)

Exercise 2

- again using the pitch data
- get inferences for the attractiveness effect using `lmer()`
- get confidence intervals for the attractiveness effect using `MCMCglmm()`

3. Random slope models

A simple random slope model

Random intercept model

$$Y_{ij} = b_{0j} + b_1 X_{1i} + u_j + e_{ij}$$

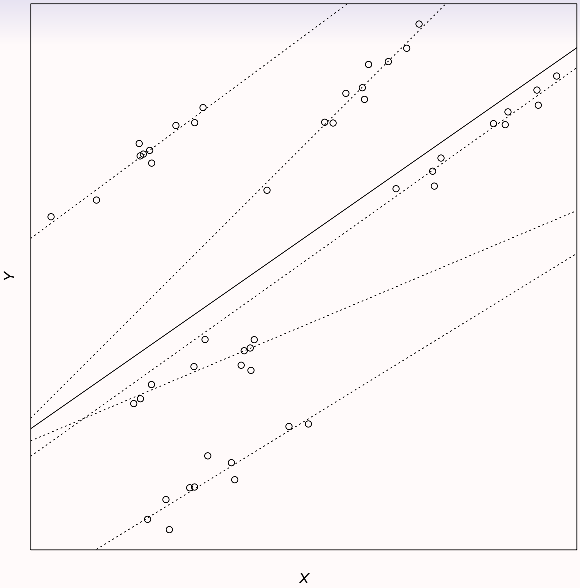
A simple random slope model

Random intercept model

$$Y_{ij} = b_{0j} + b_1 X_{1i} + u_j + e_{ij}$$

Random slope model

$$Y_{ij} = b_{0j} + b_1 X_{1ij} + u_{0j} + u_{1j} X_{1ij} + e_{ij}$$



Random slope model

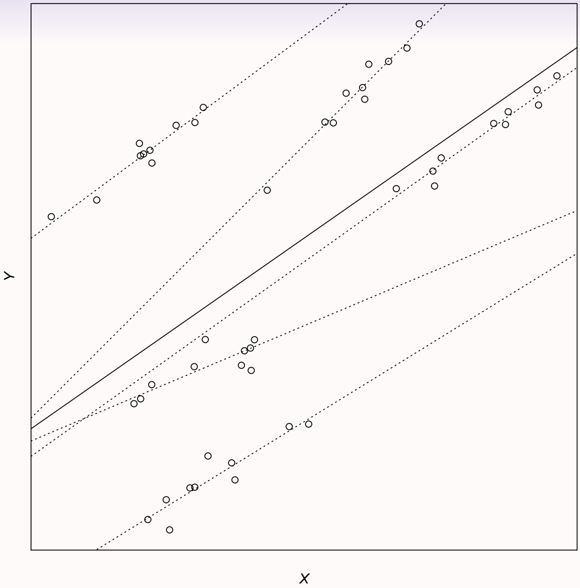
A simple random slope model

Random intercept model

$$Y_{ij} = b_{0j} + b_1 X_{1i} + u_j + e_{ij}$$

Random slope model

$$Y_{ij} = b_{0j} + b_1 X_{1ij} + u_{0j} + u_{1j} X_{1ij} + e_{ij}$$



Random intercept model

A simple random slope model

Random intercept model

$$Y_{ij} = b_{0j} + b_1 X_{1i} + u_j + e_{ij}$$

Random slope model

$$Y_{ij} = b_{0j} + b_1 X_{1ij} + u_{0j} + u_{1j} X_{1ij} + e_{ij}$$

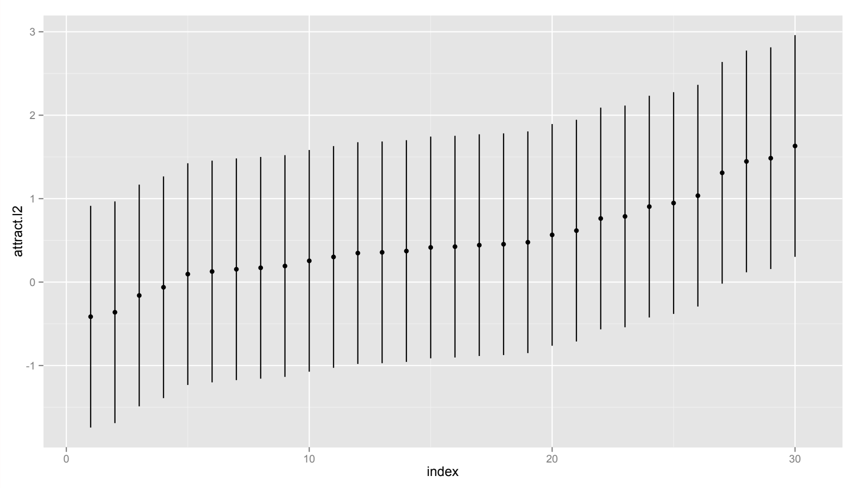
Covariance matrix for a random slope model

A random slope at level 2 gives the following covariance structure:

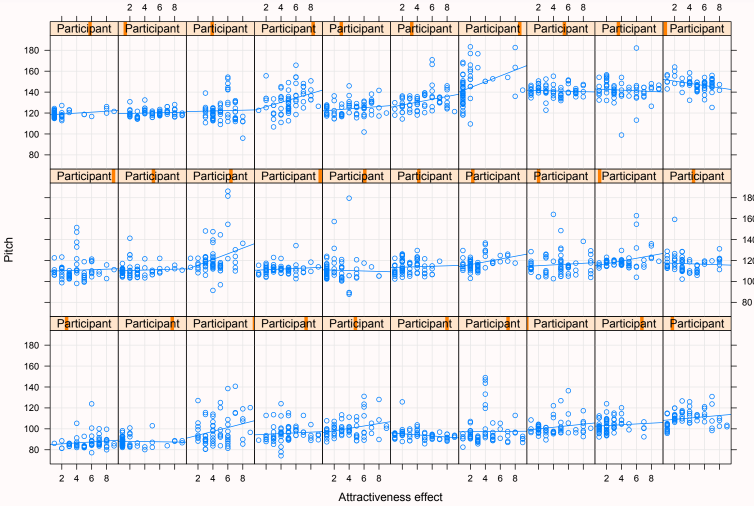
$$\begin{bmatrix} \sigma^2_{u_0} & \\ \sigma_{u01} & \sigma^2_{u_1} \end{bmatrix}$$

$$\begin{bmatrix} \sigma^2_e \end{bmatrix}$$

... adding a random slope adds two parameters: the slope variance estimate and the covariance between intercept and slope



The random slopes for the attractiveness effect



Attractiveness effect plotted for each person

Exercise 3

- again using the pitch data
- fit random slopes and (try to) get inferences for this model

4. Fully crossed models

The stimuli-as-fixed-effect falllacy

- Most psychology research treats stimuli as a *fixed effect*
  (i.e., no generalization beyond the presented stimuli)

The stimuli-as-fixed-effect falllacy

- Most psychology research treats stimuli as a *fixed effect*
  (i.e., no generalization beyond the presented stimuli)
- In contrast, we treat participants as *random effects*
  (i.e., generalisation to an infinite population)

The stimuli-as-fixed-effect falllacy

- Most psychology research treats stimuli as a *fixed effect*
  (i.e., no generalization beyond the presented stimuli)
- In contrast, we treat participants as *random effects*
  (i.e., generalisation to an infinite population)

... but many of our stimuli are from a larger population

e.g., faces, voices, words

The stimuli-as-fixed-effect falllacy

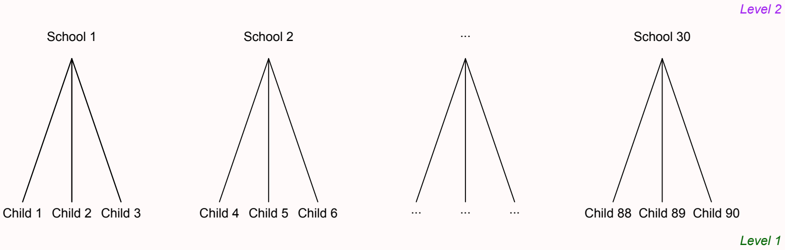
- Most psychology research treats stimuli as a *fixed effect*
  (i.e., no generalization beyond the presented stimuli)
- In contrast, we treat participants as *random effects*
  (i.e., generalisation to an infinite population)

... but many of our stimuli are from a larger population

e.g., faces, voices, words

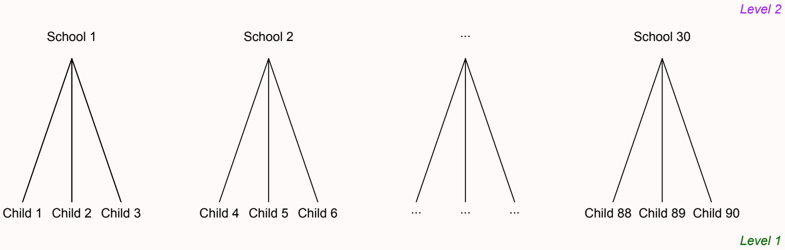
- Ignoring this increases Type I error

Nested

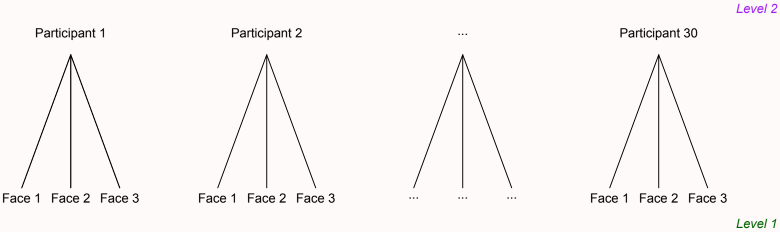


Level 2
School 1
School 2
...
School 30
Child 1
Child 2
Child 3
Child 4
Child 5
Child 6
...
...
...
Child 88
Child 89
Child 90
Level 1

Nested



Fully Crossed



Fully crossed random intercept model (one predictor)

$$Y_{i(j_1 j_2)} = b_0 + b_1 X_{1i(j_1 j_2)} + u_{1j_1} + u_{2j_2} + e_{i(j_1 j_2)}$$

Fully crossed random intercept model (one predictor)

$$Y_{i(j_1 j_2)} = b_0 + b_1 X_{1i(j_1 j_2)} + u_{1j_1} + u_{2j_2} + e_{i(j_1 j_2)}$$

Example: Voice and face processing revisited

Does male voice pitch *rise* or *fall* when rating an attractive face?

Example: Voice and face processing revisited

Does male voice pitch *rise* or *fall* when rating an attractive face?

Important to treat both participants ($n = 30$) and faces ($n = 32$) as random effects ...

Example: Voice and face processing revisited

Does male voice pitch *rise* or *fall* when rating an attractive face?

Important to treat both participants ($n = 30$) and faces ($n = 32$) as random effects ...

Statistical power now depends on the sample size (and variability) of both participants and stimuli ...

Exercise 4

- again use the pitch data
- fit a fully crossed model in `lme4`
- try to get inferences for this model ...

4. Fitting maximal random effects structures

Journal of Memory and Language 68 (2013) 255–278

Contents lists available at SciVerse ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

ELSEVIER

Random effects structure for confirmatory hypothesis testing: Keep it maximal

Dale J. Barr[a,*], Roger Levy[b], Christoph Scheepers[a], Harry J. Tily[c]

[a] Institute of Neuroscience and Psychology, University of Glasgow, 58 Hillhead St., Glasgow G12 8QB, United Kingdom
[b] Department of Linguistics, University of California at San Diego, La Jolla, CA 92093-0108, USA
[c] Department of Brain and Cognitive Sciences, Massachussetts Institute of Technology, Cambridge, MA 02139, USA

ARTICLE INFO

Article history:
Received 5 August 2011
revision received 30 October 2012

Keywords:
Linear mixed-effects models
Generalization
Statistics
Monte Carlo simulation

ABSTRACT

Linear mixed-effects models (LMEMs) have become increasingly prominent in psycholinguistics and related areas. However, many researchers do not seem to appreciate how random effects structures affect the generalizability of an analysis. Here, we argue that researchers using LMEMs for confirmatory hypothesis testing should minimally adhere to the standards that have been in place for many decades. Through theoretical arguments and Monte Carlo simulation, we show that LMEMs generalize best when they include the maximal random effects structure *justified by the design*. The generalization performance of LMEMs including *data-driven* random effects structures strongly depends upon modeling criteria and sample size, yielding reasonable results on moderately-sized samples when conservative criteria are used, but with little or no power advantage over maximal models. Finally, random-intercepts-only LMEMs used on within-subjects and/or within-items data from populations where subjects and/or items vary in their sensitivity to experimental manipulations always generalize worse than separate F_1 and F_2 tests, and in many cases, even worse than F_1 alone. Maximal LMEMs should be the 'gold standard' for confirmatory hypothesis testing in psycholinguistics and beyond.

Possible random structures ...

i) `Y ~X +(1|Subject)`
ii) `Y ~X +(X|Subject)`
iii) `Y ~X +(1|Subject) + (1|Item)`
iv) `Y ~X +(X|Subject) + (1|Item)`
v) `Y ~X +(1|Subject) + (0+ X|Subject) + (1|Item)`
vi) `Y ~X +(0 + X|Subject) + (1|Item)`

Possible random structures ...

i) `Y ~X +(1|Subject)`
ii) `Y ~X +(X|Subject)`
iii) `Y ~X +(1|Subject) + (1|Item)`
iv) `Y ~X +(X|Subject) + (1|Item)`
v) `Y ~X +(1|Subject) + (0+ X|Subject) + (1|Item)`
vi) `Y ~X +(0 + X|Subject) + (1|Item)`

In practice, it makes sense to fit a near maximal structure if the maximal model doesn't converge ...

Optional: Maximal models

- again use the pitch data
- try some of these models out ....