Linear Mixed-Effects Models

Dale Barr

PS Statistics 2018

Multilevel Data

- All statistical models assume *conditional independence*
- Data sets often have 'clusters', due to
  - Natural clustering
  - Multistage sampling

Random and Fixed Effects

$$Y_{ij} = \mu + u_i + e_{ij}$$

where:

μ	grand mean
u_i	offset associated with unit i
e_{ij}	residual associated with obs j on unit i

- Random effects are characterized by probability distributions
  - $u \sim N(0, \sigma_u^2)$; $e \sim N(0, \sigma_e^2)$
- Fixed effects are characterized by a single population value

Strategies for handling multilevel data

- "Aggregate up" to the highest sampling unit
  - *pros:*
    - easy
  - *cons:*
    - inferences only permitted at highest level
    - info lost about variability at lower levels
    - not possible with categorical/count data
    - poor behavior with unbalanced data
- Estimate variability for sampling units
  - *pros:*
    - possible for any kind of DV
    - permits full generalization
    - allows exploration of individual diffs
    - good behavior with unbalanced data
  - *cons:*
    - complex, difficult, often done wrong!
    - not guaranteed to 'converge'

Nested vs. Crossed Random Effects

- Level j is *nested* within level i if j uniquely occurs within i
  - Examples:
    - children within classrooms, classrooms within schools, schools within districts...
    - observations within individual subjects
    - residents within neighborhoods
- Level j is *crossed* with level i if j is associated with multiple is
  - Examples:
    - subjects and stimulus materials
    - customers and restaurants

The "Language-As-Fixed-Effect" Fallacy

Clark (1973)

- Psycholinguistic experiments sample *language materials* as well as *subjects*
- Language materials should be treated as *random effects*
  - otherwise, results may not generalize
- NB: same goes for most other kinds of stimulus materials!
- Clark's suggestion: F', min-F'
- Linear-mixed effects solution: include subjects and items as *crossed* random effects (Baayen, Davidson, & Bates, 2008)
  - over 1000 citations to date! (Google Scholar)

Generalizing over encounters

(Barr, in press)

The target of inference in much of psychology and related fields has been misidentified as a population of *subjects* or *stimuli*, when the actual target of inference is a population of events: encounters

- readers encountering particular types of words
- male participants judging female faces
- gamers encountering particular types of violent games
- audience members encountering particular types of dance movements
- insomniacs (versus controls) encountering emotional expressions
- birds hearing particular types of birdsongs

Specifying random effects structure: Maximal models
(Barr, Levy, Scheepers, Tily, 2013)

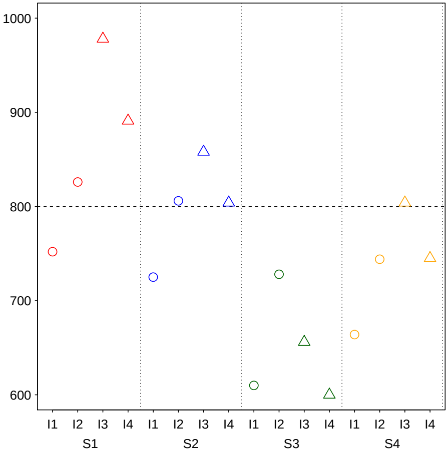
- Typical experimental studies have multiple observations *per cell per subject (or item)*
- Subjects (items) vary in sensitivity to experimental manipulations
- This dependency must be accounted for using *random slopes*
- For experimental hypothesis testing, researchers should use the *maximal random effects structure justified by the design*

A hypothetical experiment

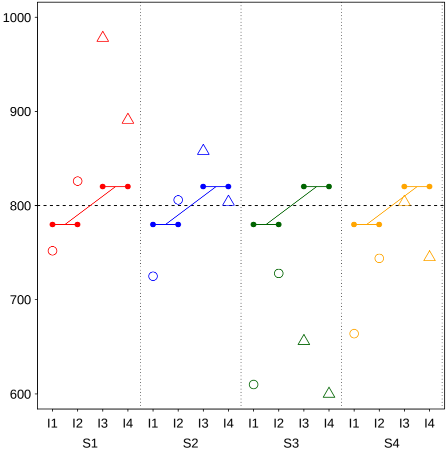
(Barr, Levy, Scheepers, Tily, 2013)

- Subjects perform lexical decision (is `PINT` a word or nonword?)
- DV = response time
- IV = word type, Type A and Type B
- Four subjects, four words

Choosing the right model



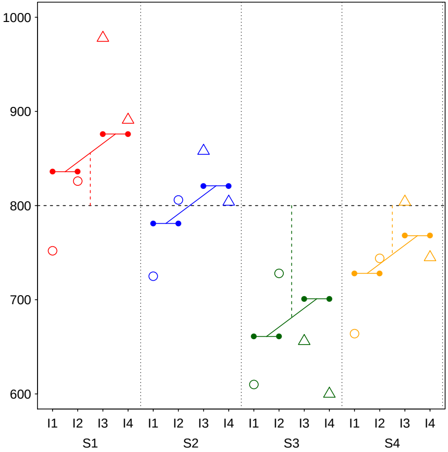
Choosing the right model



$$Y_{ij} = \beta_0 + \beta_1 X_j + e_{ij}$$

$$e_{ij} \sim N(0, \sigma^2)$$

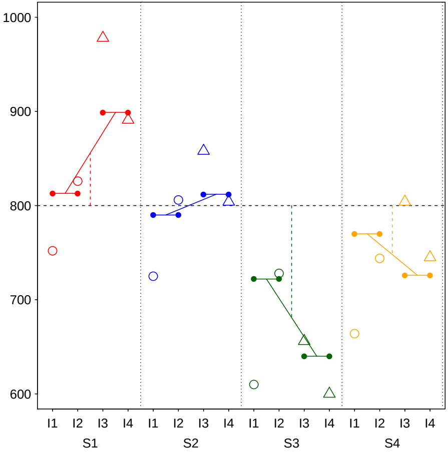
Choosing the right model



$$Y_{ij} = \beta_0 + \beta_1 X_j + e_{ij}$$
$$Y_{ij} = \beta_0 + S_{0i} + \beta_1 X_j + e_{ij}$$

$$e_{ij} \sim N(0, \sigma^2)$$
$$S_{0i} \sim N(0, \tau_{00})$$

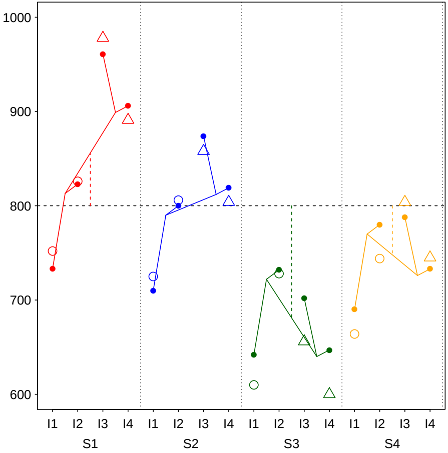
Choosing the right model



$$Y_{ij} = \beta_0 + \beta_1 X_j + e_{ij}$$
$$Y_{ij} = \beta_0 + S_{0i} + \beta_1 X_j + e_{ij}$$
$$Y_{ij} = \beta_0 + S_{0i} + (\beta_1 + S_{1i}) X_j + e_{ij}$$

$$e_{ij} \sim N(0, \sigma^2)$$
$$(S_{0i}, S_{1i}) \sim N\left[(0,0), \Sigma_\tau\right]$$
$$\Sigma_\tau = \begin{pmatrix} {\tau_{00}}^2 & \rho\tau_{00}\tau_{11} \\ \rho\tau_{00}\tau_{11} & {\tau_{11}}^2 \end{pmatrix}$$

Choosing the right model



$$Y_{ij} = \beta_0 + \beta_1 X_j + e_{ij}$$
$$Y_{ij} = \beta_0 + S_{0i} + \beta_1 X_j + e_{ij}$$
$$Y_{ij} = \beta_0 + S_{0i} + (\beta_1 + S_{1i})X_j + e_{ij}$$
$$Y_{ij} = \beta_0 + S_{0i} + I_{0j} + (\beta_1 + S_{1i})X_j + e_{ij}$$

$$e_{ij} \sim N(0, \sigma^2)$$
$$(S_{0i}, S_{1i}) \sim N\left[(0,0), \Sigma_\tau\right]$$
$$\Sigma_\tau = \begin{pmatrix} {\tau_{00}}^2 & \rho\tau_{00}\tau_{11} \\ \rho\tau_{00}\tau_{11} & {\tau_{11}}^2 \end{pmatrix}$$
$$I_{0j} \sim N(0, {\omega_{00}}^2)$$

What about a by-item random slope?

- A by-item random slope does not make sense when items *are* the experimental manipulation
- e.g., most words are either a noun or a verb; such words cannot vary in the effect of noun vs. verb!

Maximal random effects justified *by the data* or *by the design*?

“keep it parsimonious” versus “keep it maximal”

Principles for Choosing Maximal Random Effects

- All sampling units get a random intercept
- Any factor gets a by-unit random slope if it is both:
  1. within-unit, and
  2. has multiple observations per level per unit
- Any interaction term gets a by-unit random slope where:
  1. all factors are within-unit
  2. there are multiple observations per unit/cell
     - where 'cell' is a combination of factor levels
- For more tips see Barr et al. (2013); Barr (2013)

`lme4::lmer()` model syntax

- by-subjects random-intercepts-only model

```
y ~ x1 * x2 + (1 | subject_id)
```

- by-subjects random slopes for main effects

```
y ~ x1 * x2 + (1 + x1 + x2 | subject_id)
```

- maximal model: by-subjects random slopes for all within-unit effects with multiple observations per cell

```
y ~ x1 * x2 + (1 + x1 + x2 + x1:x2 | subject_id)
y ~ x1 * x2 + (1 + x1 * x2 | subject_id)
```

- maximal zero covariances model

```
y ~ x1 * x2 + (1 + x1 * x2 || subject_id)
```

Dealing with nonconvergence

- Misspecification of random effects
  - *unidentifiable* parameters in the model
- Using old/unstable version of `lme4` (<1.1-7)
- Using suboptimal optimizer for `glmer`
  - use argument `glmerControl(optimizer='bobyqa')`
- Uncentered/unscaled predictors
- Too few subjects/items
- Distributional assumptions not satisfied
- Null effects

What to do?

- Make sure effects are identifiable
- Increase iterations

```
mod <- lmer(y ~ x1 * x2 + (1 + x1 * x2 | subject_id), dat,
            REML = FALSE,
            control = lmerControl(maxfun = 20000))
```

- Check distributional assumptions
- Fit zero-correlation model

```
mod <- lmer(y ~ x1 * x2 + (1 + x1 * x2 || subject_id), dat,
            REML = FALSE)
```

- Start removing random effects