Mixing it up with random effects

Joshua Loftus

What is a mixed model?

For simplicity we'll only talk about linear models.

Mixed GLS

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{b} + \epsilon, \quad \mathrm{Cov}(\mathbf{y}) = \Sigma$$

- β, $\mathbf{b}$, and ϵ are all unobserved
- β is a vector of *parameters*
- $\mathbf{b}$ is a vector of *random variables*
- ϵ error with $\mathrm{E}(\epsilon) = 0$, $\mathrm{Cov}(\mathbf{b}, \epsilon) = 0$
- Inference about (β, Σ) from conditional distribution $\mathbf{y}|\mathbf{b}$

Examples

Mixed GLS

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{b} + \epsilon, \quad \mathrm{Cov}(\mathbf{y}) = \Sigma$$

- "Random slopes and intercepts"
- Error is not i.i.d. / Clustered errors
- Test scores of students, school effect, teacher effect
  Assume $\mathbf{b} \sim N(0, \sigma_T^2 I)$. What if σ_T^2 is large? Small?
  What if there are only a handful of teachers in the study?
- Repeated measures / Longitudinal, e.g. gene $\sim$ drug * time

Fitting the model

- If $\text{Var}(\mathbf{b}) = \mathbf{D}$ and $\text{Var}(\epsilon) = \mathbf{R}$ then $\text{Var}(\mathbf{y}) = \mathbf{R} + \mathbf{Z}\mathbf{D}\mathbf{Z}^T$
- $\mathbf{R}$, $\mathbf{D}$, and maybe even $\mathbf{Z}$ are functions of another parameter θ ("variance components")
- Often reasonable to assume multivariate normality of $\mathbf{y}|\mathbf{b}$
- Maximum likelihood estimation of θ based on $L(\theta, \beta; \mathbf{y})$ does not account for loss in degrees of freedom caused by estimating β. Analogous to $\hat{\sigma}/n$ vs. $\hat{\sigma}/(n-p)$
- REML based on "residual" of $\mathbf{y}$ (residual contrasts)
- REML coincides with ANOVA for balanced designs

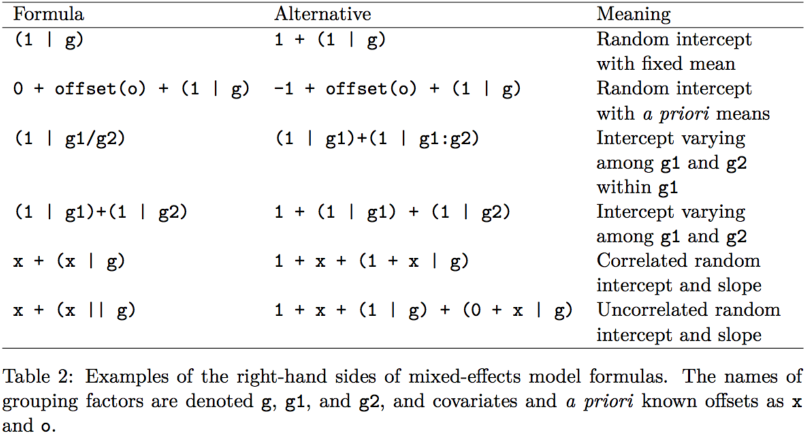
Fitting mixed models in *R* with lme4

Examples using the lme4 package in *R*

- pitch $\sim$ gender $+$ (1|subject) $+$ (1|scenario)
- price $\sim$ time $+$ (time|product)
- participation $\sim$ extroversion $+$ (1|school/class)

Read more (these links were also in the email I sent earlier)
`http://cran.r-project.org/web/packages/lme4/vignettes/lmer.pdf`
`http://cran.r-project.org/web/packages/lme4/lme4.pdf`

Formulas in lme4



Formula	Alternative	Meaning
`(1 | g)`	`1 + (1 | g)`	Random intercept with fixed mean
`0 + offset(o) + (1 | g)`	`-1 + offset(o) + (1 | g)`	Random intercept with *a priori* means
`(1 | g1/g2)`	`(1 | g1)+(1 | g1:g2)`	Intercept varying among `g1` and `g2` within `g1`
`(1 | g1)+(1 | g2)`	`1 + (1 | g1) + (1 | g2)`	Intercept varying among `g1` and `g2`
`x + (x | g)`	`1 + x + (1 + x | g)`	Correlated random intercept and slope
`x + (x || g)`	`1 + x + (1 | g) + (0 + x | g)`	Uncorrelated random intercept and slope

Table 2: Examples of the right-hand sides of mixed-effects model formulas. The names of grouping factors are denoted `g`, `g1`, and `g2`, and covariates and *a priori* known offsets as `x` and `o`.

Discussion

- Questions?
- More examples: fixed effects vs. random effects
- Next topic?
  Time series
  Bootstrap
  Multiple comparisons + selective inference
  Causal inference
  Missingness / data cleaning / etc
  Bonus session on basic stats?