Introduction to Multilevel Models

Paul Bürkner

Multilevel Models

- Modeling data measured on different levels within one model
- Account for the dependency structure of the data
- Estimate variation on all levels of the model

Synonyms:

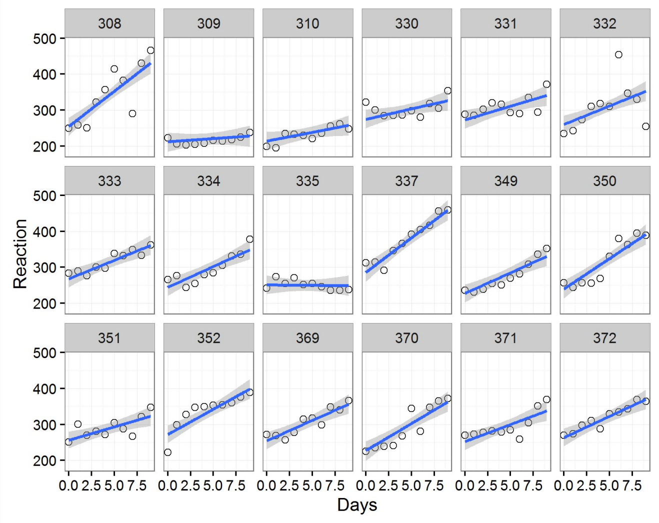
- Hierachical models
- Random effects models
- Mixed (effects) models

A sample Data Set: Sleepstudy

```r
data("sleepstudy", package = "lme4")
head(sleepstudy, 10)
```

Reaction	Days	Subject
249.5600	0	308
258.7047	1	308
250.8006	2	308
321.4398	3	308
356.8519	4	308
414.6901	5	308
382.2038	6	308
290.1486	7	308
430.5853	8	308
466.3535	9	308

Sleepstudy: Visualization



Linear Regression: Complete Pooling

Basic Linear Regression

$$y_n \sim \mathcal{N}(\eta_n, \sigma)$$

$$\eta_n = b_0 + b_1 x_n$$

We can write the first equation equivalently as

$$y_n = \eta_n + e_n$$

$$e_n \sim \mathcal{N}(0, \sigma)$$

Assumptions of Linear Regression (Gelman & Hill 2007)

Validity of the data

- There is no substitute for good data

Statistical assumptions:

- Additivity and linearity
- Independence of errors
- Equal variance of errors
- Normality of errors

Additivity and Linearity

Assume that the predictor term η is a linear combination of the predictor variables multiplied by the regression coefficents:

$$\eta_n = \sum_{i=1}^{K} b_i x_{ni}$$

How to handle violations:

- User specified non-linear predictor terms
- Semiparametric non-linear approaches such as splines or Gaussian processes

Independence of errors

Assume that errors of different observations are uncorrelated:

$$cov(e_n, e_m) = 0$$

for $n \neq m \in \{1, ..., N\}$

How to handle violations:

- Model the dependency

Equal variance and normality of errors

Assume that all errors are normally distributed and share the same variance (or standard deviation):

$$e_n \sim \mathcal{N}(0, \sigma)$$

How to handle violations (equal variances):

- Find variables explaining the unequal variances
- Model unequal variances

How to handle violations (normality of errors):

- Use transformationsn or generalized linear models

Linear Regression: Partial Pooling

Varying intercept model

$$y_n \sim \mathcal{N}(\mu_n, \sigma)$$

$$\mu_n = b_{0j[n]} + b_1 x_n$$

$$b_{0j} \sim \mathcal{N}(b_0, \sigma_{b_0})$$

- Apply shrinkage to the varying (*random*) intercepts
- Assume the slopes to be the same (*fixed*) across participants

Linear Regression: Partial Pooling

Varying intercept, varying slope model

$$y_n \sim \mathcal{N}(\mu_n, \sigma)$$

$$\mu_n = b_{0j[n]} + b_{1j[n]} x_n$$

$$(b_{0j}, b_{1j}) \sim \mathcal{MN}((b_0, b_1), \Sigma_b)$$

$$\Sigma_b = \begin{pmatrix} \sigma_{b_0}^2 & \sigma_{b_0}\sigma_{b_1}\rho_{b_0 b_1} \\ \sigma_{b_0}\sigma_{b_1}\rho_{b_0 b_1} & \sigma_{b_1}^2 \end{pmatrix}$$

- Apply shrinkage to the varying intercepts and slopes
- Model the correlation between intercepts and slopes

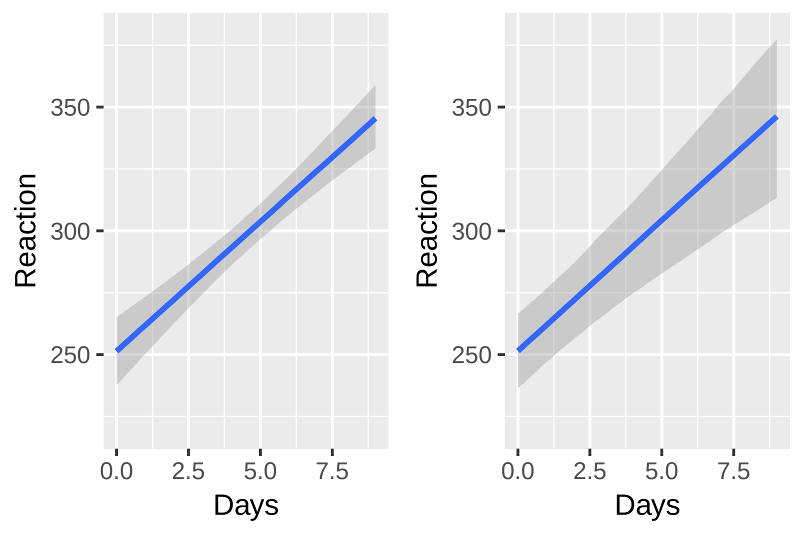
Linear Regression: No Pooling

$$y_n \sim \mathcal{N}(\mu_n, \sigma)$$

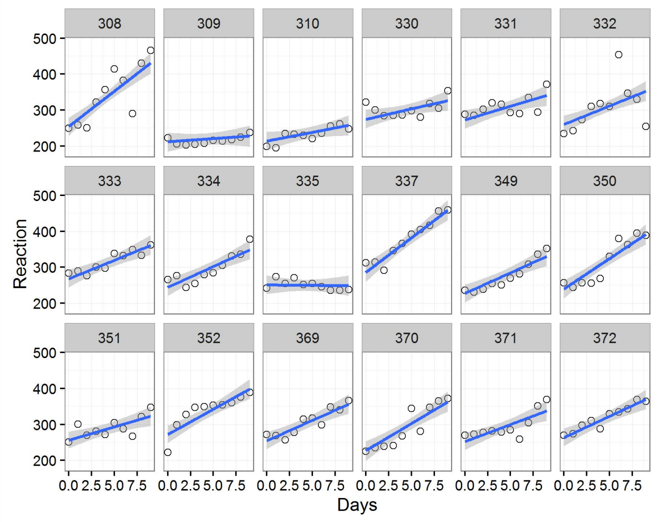
$$\mu_n = b_{0j[n]} + b_{1j[n]} x_n$$

- No multilevel framework needed
- No shrinkage on the regression coefficients
- Less stable estimates

Complete vs. Partial Pooling



Individual Estimates Based on Partial Pooling



Higher level predictors

- Suppose that a predictor z is assessed on a higher level of the data (e.g. on class level)

Varying intercept model

$$y_n \sim \mathcal{N}(\mu_n, \sigma)$$

$$\mu_n = b_{0j[n]} + b_1 x_n$$

$$b_{0j} \sim \mathcal{N}(b_{00} + b_{01} z_j, \sigma_{b_0})$$

Rewriting multilevel models

We can write the model above equivalently as follows:

$$y_n \sim \mathcal{N}(\mu_n, \sigma)$$

$$\mu_n = b_{00} + b_{01} z_j + \tilde{b}_{0j[n]} + b_1 x_n$$

$$\tilde{b}_{0j} \sim \mathcal{N}(0, \sigma_{b_0})$$

Why We Don't Need Multiple Data Sets

Data in long format:

Grade	IQ	Class	Teacher_ability
1	108	a	1
4	91	a	1
4	92	a	1
4	92	a	1
5	87	a	1
4	93	b	5
1	85	b	5
2	88	b	5
4	101	b	5
3	114	b	5

Some Advantages of Multilevel Models

- Conveniently **estimate variation** on different levels of the data
- **Account for** all sources of **uncertainty** in the population-level effects
- **Increase precision** of group-level estimates especially of those with small available information
- **Predict values** of new groups not originally present in the data

Advantages of Bayesian Multilevel Models

- Improve partial pooling by defining reasonable priors on hyperparameters
- Allow to fit much more group-level effects as compared to frequentist MLM implementations
- Really estimate group-level effects (i.e, b_{0j}, b_{1j} etc.)
- Get full posteriors of hierachical parameters, which often have skewed distributions

Why I Don't Like Other Names for MLMs

Hierachical models:

- Crossed group-level effects (e.g., observations nested within country and year) do not follow a hierachy

Random effects models:

- Implies a frequentist understanding and way of estimating group-level effects

Mixed (i.e., fixed and random) effects models:

- The term "fixed effects" is not used consistently in the literature
- There are better frameworks (e.g., the type of pooling) to classify effects than classifying after "fixed" and "random" effects.

Additional Things to Keep in Mind

Three Goals for Inference from Data:

- Estimation of model parameters
- Prediction of new data
- Model comparison

Predictors may improve model fit / predictions even if not significant (or vice versa)

Always try to understand your data

Time for exercise 3

Further Reading

- Gelman, A., & Hill, J. (2007). Data analysis using regression and multilevel/hierarchical models. Cambridge University Press.
- Zuur, A. F., Hilbe, J. M., & Ieno, E. N. (2013). A Beginner's Guide to GLM and GLMM with R: A Frequentist and Bayesian Perspective for Ecologists. Newburgh: Highland Statistics.
- McElreath, R. (2016). Statistical rethinking: A Bayesian course with examples in R and Stan. CRC Press.