mixed models: introduction

Ben Bolker

18 November 2025

Table of contents

Overview

Mixed models

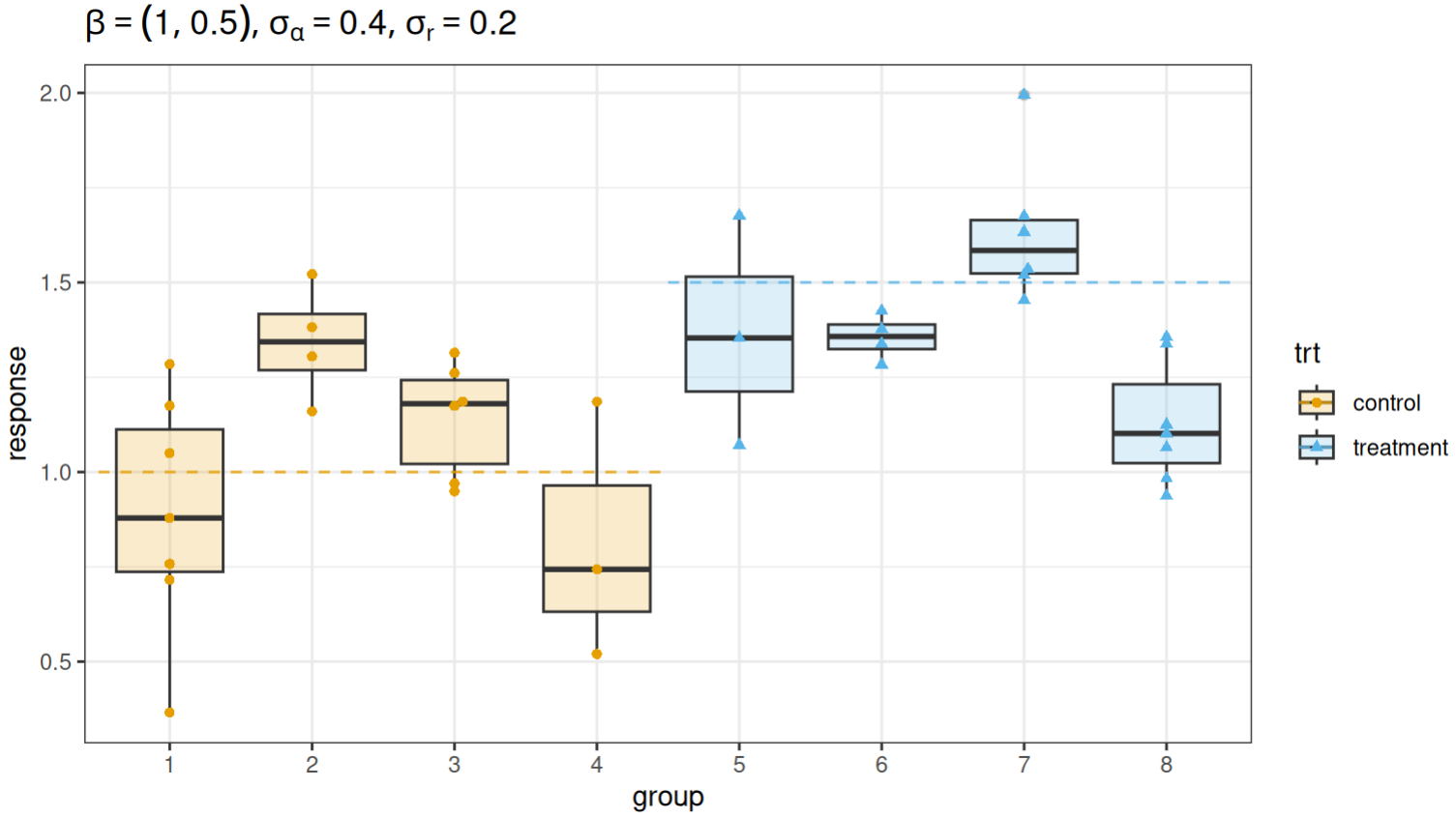
- *tabular* data (row = observation, column = variable)
- *regression* models
  - continuous and categorical *predictor* variables
  - single (univariate) numeric *response* variable
- with one or more *grouping variables* ('blocks', 'clusters', …)
- for prediction or inference
- typically for 'medium-sized' data (e.g. 10^6 rows, 10^3 clusters, 10-20 predictors)

Mixed models (easy version)

- take a standard linear-type model (e.g. $y_i = a + bx_i + \epsilon_i$)
- … but observations fall in *groups*
- breaks assumption of independence (*pseudoreplication*)
- add *between-group* (random) variation:

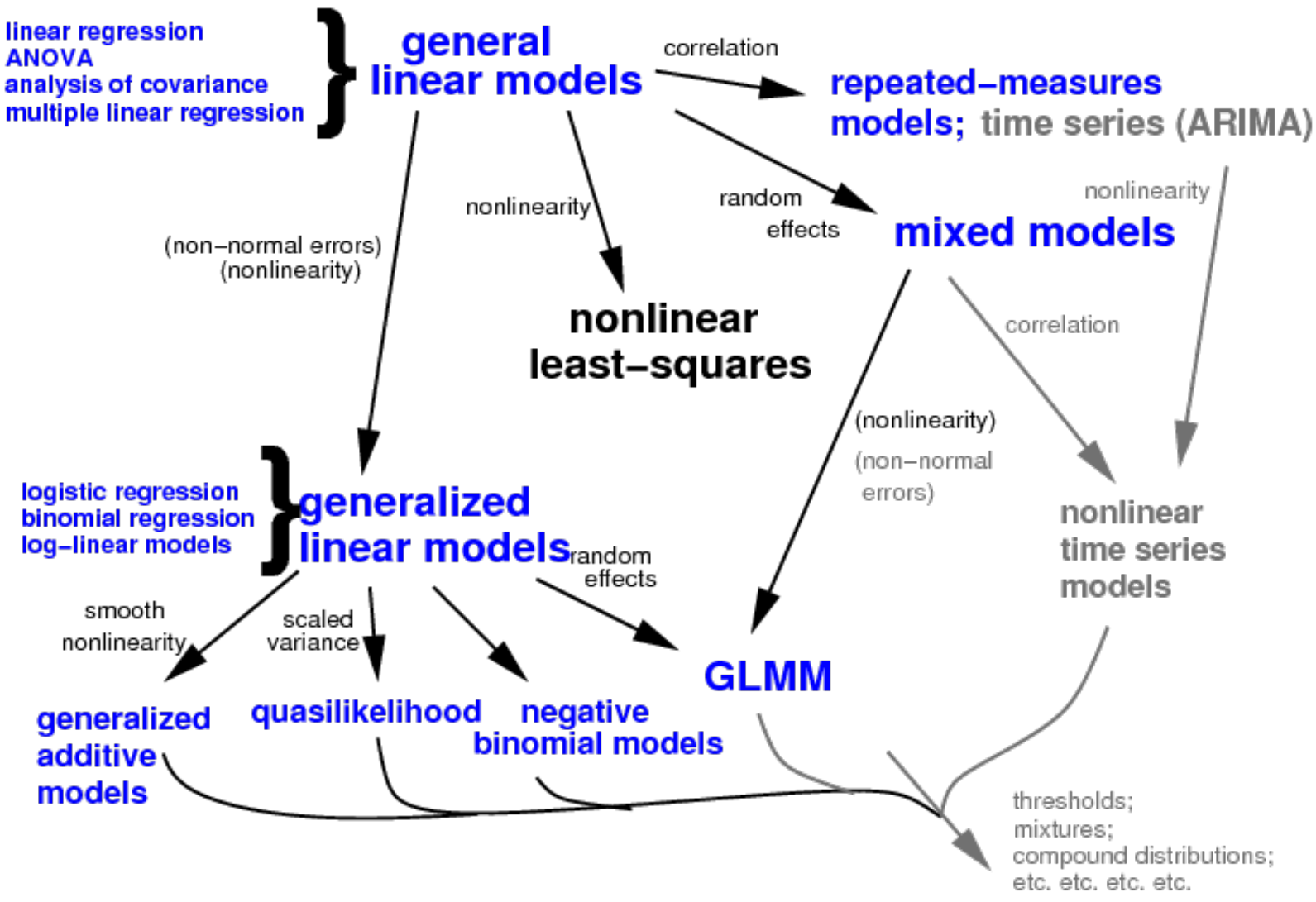
$$y_{ij} = a + bx_{ij} + \alpha_i + \epsilon_{ij}$$

single group, categorical predictor (= one-way ANOVA)



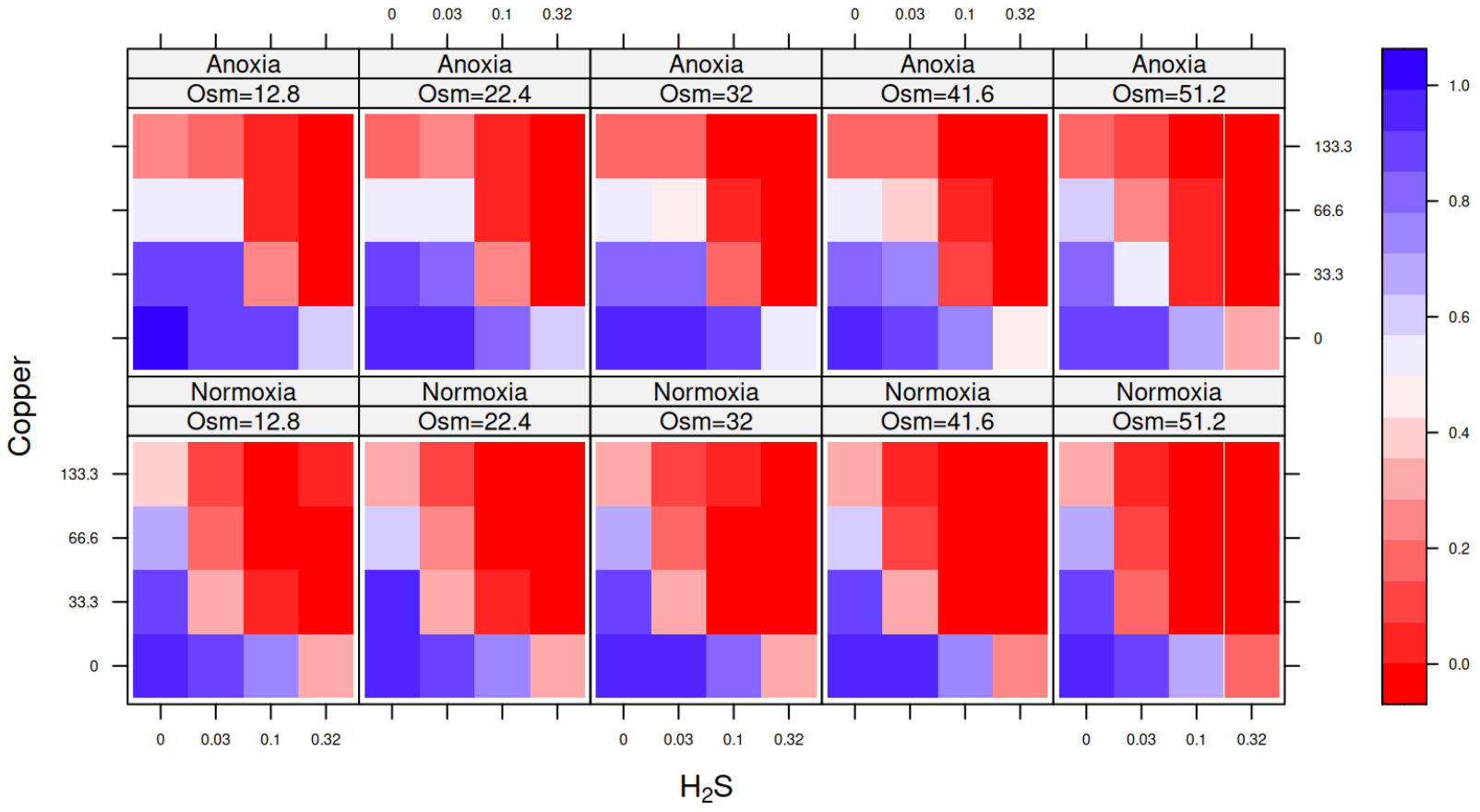
The mixed model universe

- more complex *fixed effects* components (multivariate regression, multiway ANOVA, ANCOVA, …)
- multiple grouping variables (random effects *terms*)
- multiple effects varying across groups (*random slopes models* etc.)
- non-Normal *conditional distributions*
  - exponential dispersion family (à la GLMs): binomial/Poisson/etc.
  - weirder: Tweedie, *t* distribution, Beta, …
- nonlinearity: *link functions* (log, logit, etc.)
- structured covariances: autoregressive, compound symmetric, regression splines …


linear regression
ANOVA
analysis of covariance
multiple linear regression
general linear models
correlation
repeated–measures models; time series (ARIMA)
nonlinearity
random effects
mixed models
nonlinearity
(non–normal errors)
(nonlinearity)
nonlinear least–squares
correlation
(nonlinearity)
(non–normal errors)
logistic regression
binomial regression
log–linear models
generalized linear models
random effects
nonlinear time series models
smooth nonlinearity
scaled variance
GLMM
generalized additive models
quasilikelihood
negative binomial models
thresholds;
mixtures;
compound distributions;
etc. etc. etc. etc.

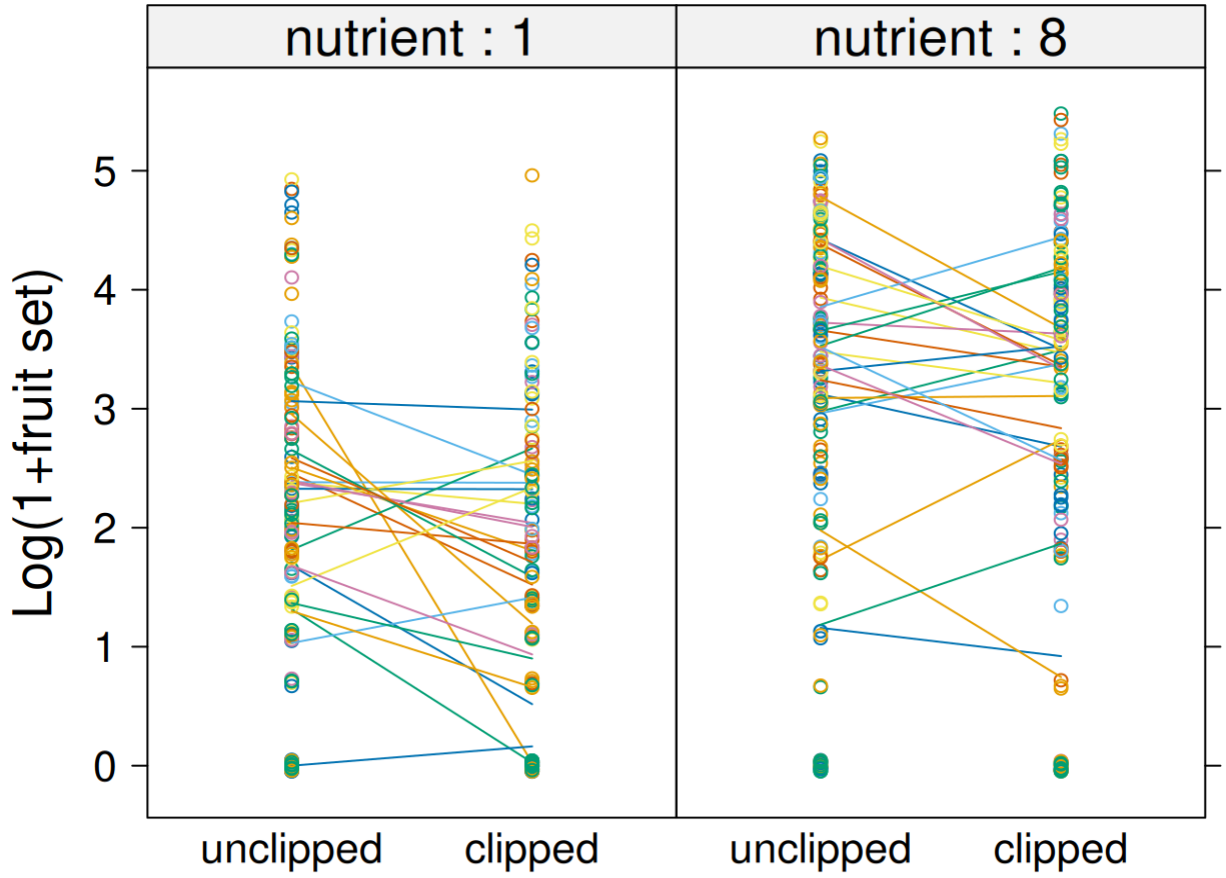
Examples

Experimental: bloodworm cell survival



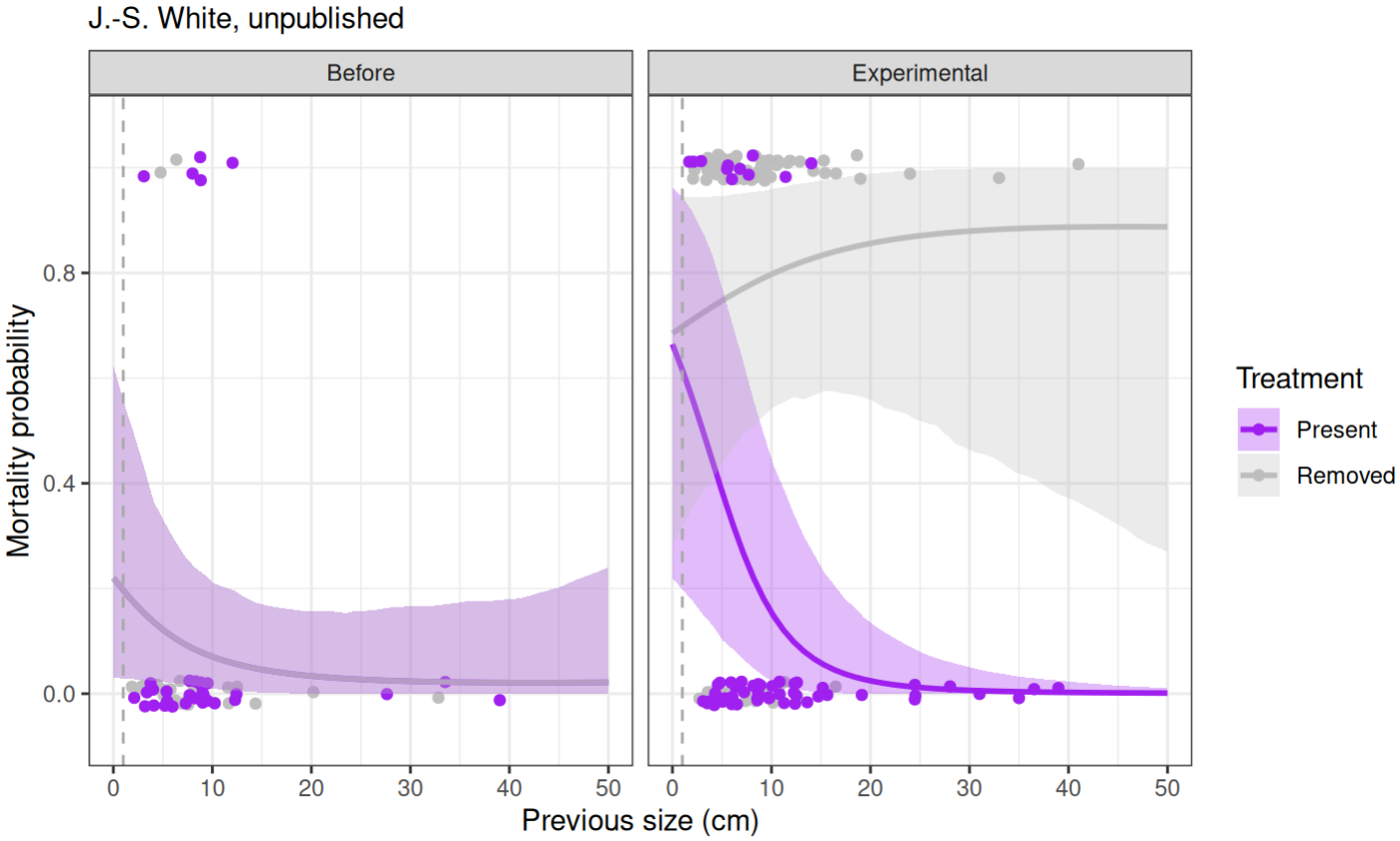
D. Julian, unpublished data

Arabidopsis: fertilization & herbivory



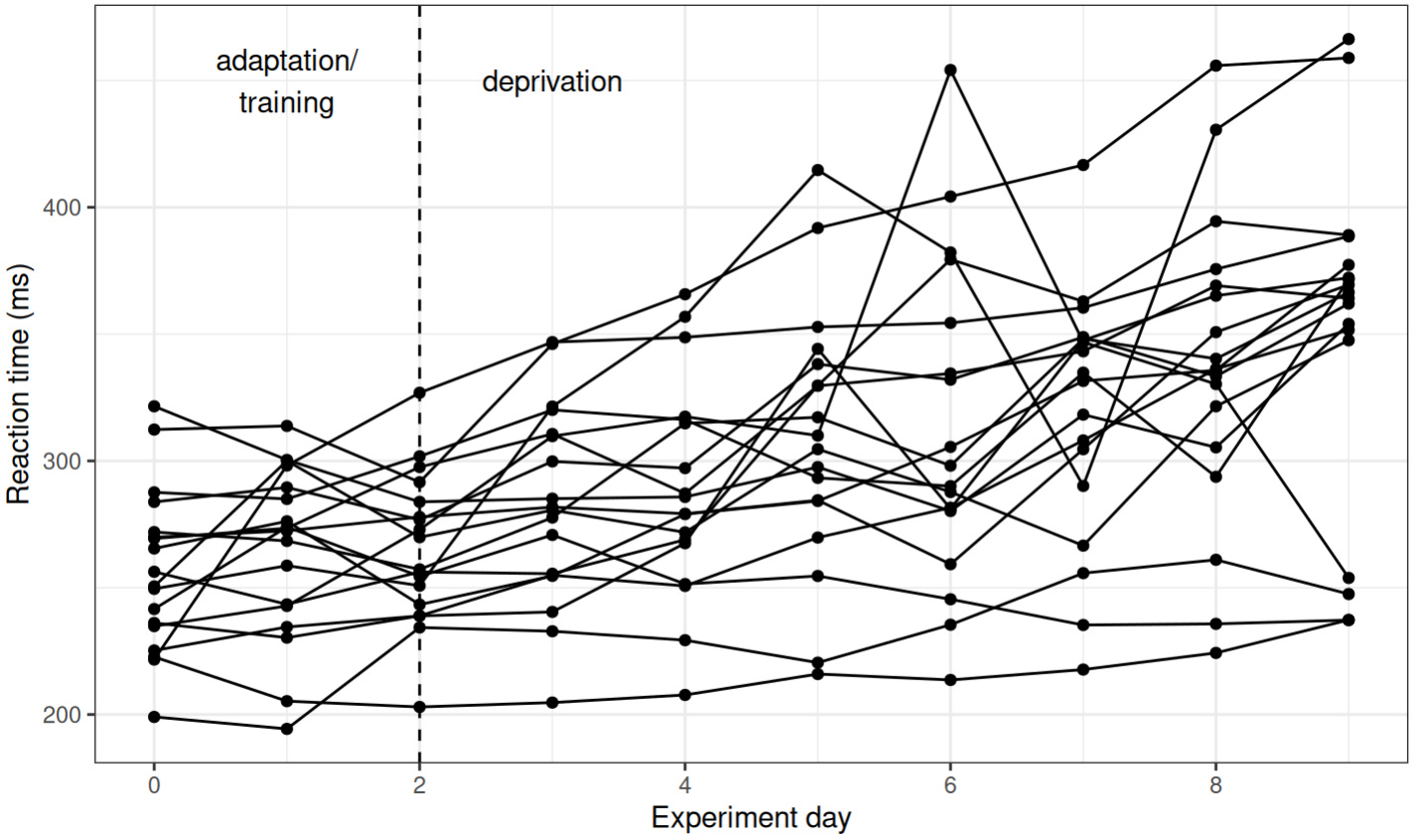
Banta et al. 2010

Coral demography



Sleep deprivation study

Belenky et al. (2003)



Definitions and descriptions

Simple mixed models (scalar, intercept-only)

$$
\begin{aligned}
y_{ij} &= \beta_0 + \beta_1 x_{ij} + \epsilon_{0,ij} + \epsilon_{1,j} \\
&= (\beta_0 + \epsilon_{1,j}) + \beta_1 x_{ij} + \epsilon_{1,j} \\
\epsilon_{0,ij} &\sim \text{Normal}(0, \sigma_0^2) \\
\epsilon_{1,j} &\sim \text{Normal}(0, \sigma_1^2)
\end{aligned}
$$

- Could have multiple, nested levels of random effects (genotype within population within region …), or *crossed* REs

Random-slopes model

$$
\begin{aligned}
y_{ij} &= \beta_0 + \beta_1 x_{ij} + \epsilon_{0,ij} + \epsilon_{1,j} + \epsilon_{2,j} x_{ij} \\
&= (\beta_0 + \epsilon_{1,j}) + (\beta_1 + \epsilon_{2,j}) x_{ij}) \\
\epsilon_{0,ij} &\sim \text{Normal}(0, \sigma_0^2) \\
\{\epsilon_{1,j}, \epsilon_{2,j}\} &\sim \text{MVN}(0, \Sigma)
\end{aligned}
$$

- variation in the *effect* of a treatment or covariate across groups
- group intercept, slope offsets are correlated

Terminology

- **grouping variables** define sets of clusters (e.g. school, year, country …)
- **varying effects** are predictors whose effects that vary across clusters (e.g. intercept, herbivory, time)
- a random effects **term** is the combination of a clustering variable and a set of varying effects
- a **conditional mode** is the predicted value of the *difference* of a cluster effect from the population mean (slope difference of a subject) (also “BLUPs”)
- the **conditional distribution** is the probability distribution of the conditional modes (Gaussian for LMMs, or Poisson, Beta, Tweedie, …)

General definition

$$\underbrace{Y_i}_{\text{response}} \sim \overbrace{\text{Distr}}^{\text{conditional distribution}}\left(\underbrace{g^{-1}(\eta_i)}_{\substack{\text{inverse}\\ \text{link}\\ \text{function}}}, \underbrace{\phi}_{\substack{\text{scale}\\ \text{parameter}}}\right)$$

$$\underbrace{\boldsymbol{\eta}}_{\substack{\text{linear}\\ \text{predictor}}} = \underbrace{\boldsymbol{X\beta}}_{\substack{\text{fixed}\\ \text{effects}}} + \underbrace{\boldsymbol{Zb}}_{\substack{\text{random}\\ \text{effects}}}$$

$$\underbrace{\boldsymbol{b}}_{\substack{\text{conditional}\\ \text{modes}}} \sim \text{MVN}(\mathbf{0}, \underbrace{\Sigma(\boldsymbol{\theta})}_{\substack{\text{variance-}\\ \text{covariance}\\ \text{matrix}}})$$

What are random effects?

A method for …

- accounting for among-individual variation, within-cluster correlation
- compromising between
  complete pooling (no among-cluster variance)
  and *fixed effects* (large among-cluster variance)
- handling levels selected at random from a larger population
- sharing information among clusters (*shrinkage estimation*)
- estimating among-cluster variability
- allowing predictions for unobserved clusters
- modeling **exchangeable** groups

Random-effect myths

- clusters must always be sampled at random
- a complete sample cannot be treated as a random effect
- random effects are always a *nuisance variable*
- nothing can be said about the predictions of a random effect
- you should always use a random effect no matter how few clusters you have

Reasons for random effects (inferential/ philosophical)

- **don't** want to
  - test hypotheses about differences between responses at particular levels of the grouping variable;
- **do** want to
  - quantify variation among groups
  - make predictions about unobserved groups
- have levels that are randomly sampled from/representative of a larger population (*exchangeable*)

Reasons for random effects (practical)

- want to combine information across groups
- have variation in information per level (number of samples or noise level)
- have a categorical predictor that is a nuisance variable (i.e., not of direct interest, but needs to be controlled for)
- have more than 5-6 groups that could be considered exchangeable

See also Crawley (2002); Gelman (2005)

Avoiding mixed models

- average: for *nested* (balanced) LMMs (Murtaugh 2007)
- use fixed effects (or *two-stage models*) instead of random effects when there are
  - many observations per cluster (shrinkage unimportant)
  - few clusters (little advantage; bad variance estimates)

Model setup

Formulas

- random effects specified with | (“pipe”/“bar”)
  as `(1+a|g1) + (1+b|g2) + ...`
- right-hand side (`g1`, `g2`) is the *grouping variable* (always treated as categorical, usually a factor)
- left-hand side describes the *varying terms* (1 [intercept], a, b)
- + separates independent terms
- formulas: empower but conceal details (McElreath 2015)

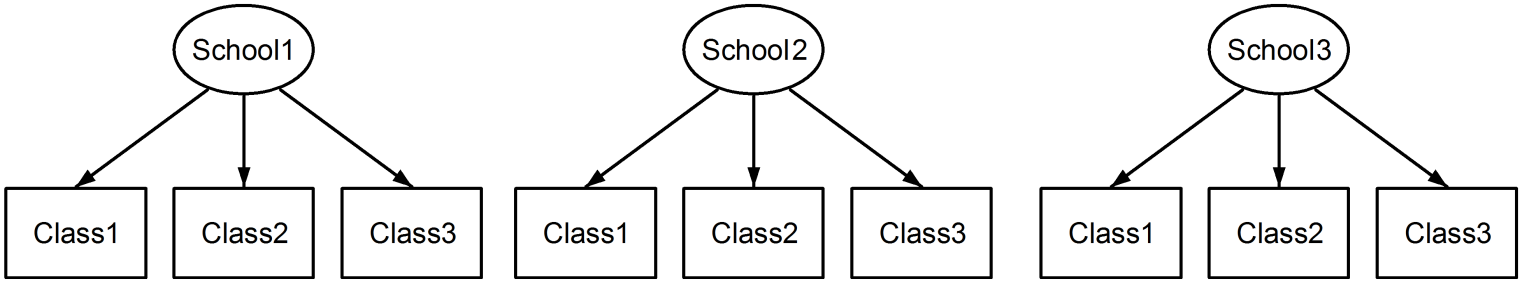
Formula definitions

equation	formula
$\beta_0 + \beta_1 X_i + e_{si}$	n/a (Not a mixed-effects model)
$(\beta_0 + b_{S,0s}) + \beta_1 X_i + e_{si}$	`~ X + (1|Subject)`
$(\beta_0 + b_{S,0s}) + (\beta_1 + b_{S,1s}) X_i + e_{si}$	`~ X + (1 + X|Subject)`
$(\beta_0 + b_{S,0s} + b_{I,0i}) + (\beta_1 + b_{S,1s}) X_i + e_{si}$	`~ X + (1 + X|Subject) + (1|Item)`
As above, but S_{0s}, S_{1s} independent	`~ X + (1||Subject) + (1|Item)`
$(\beta_0 + b_{S,0s} + b_{I,0i}) + \beta_1 X_i + e_{si}$	`~ X + (1|Subject) + (1|Item)`
$(\beta_0 + b_{I,0i}) + (\beta_1 + b_{S,1s}) X_i + e_{si}$	`~ X + (0 + X|Subject) + (1|Item)`

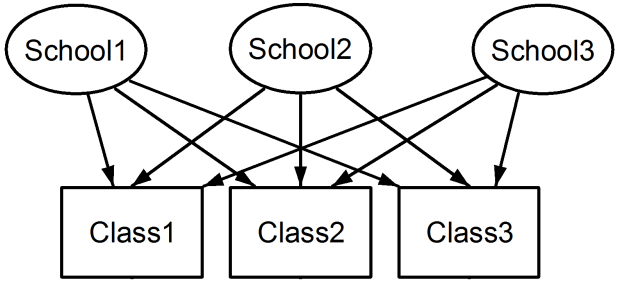
Modified from: http://stats.stackexchange.com/questions/13166/rs-lmer-cheat-sheet?lq=1 (Livius). Subscripts $\{S, I\}$ refer to Subject vs Item effects. Lower-case $\{s, i\}$ indicate particular subjects/items. $\{0, 1\}$ refer to intercept vs slope effects.

Nesting vs crossing

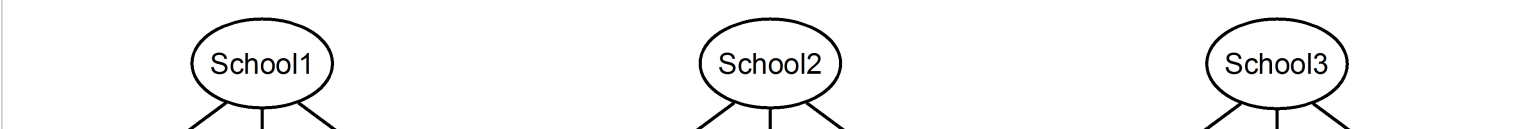
Nested: sub-unit IDs only measured within a single larger unit. e.g.: Class1 in School1 independent of Class1 in School2



Crossed: sub-unit IDs can be measured in multiple larger units. e.g.



Unique coding: removes ambiguity



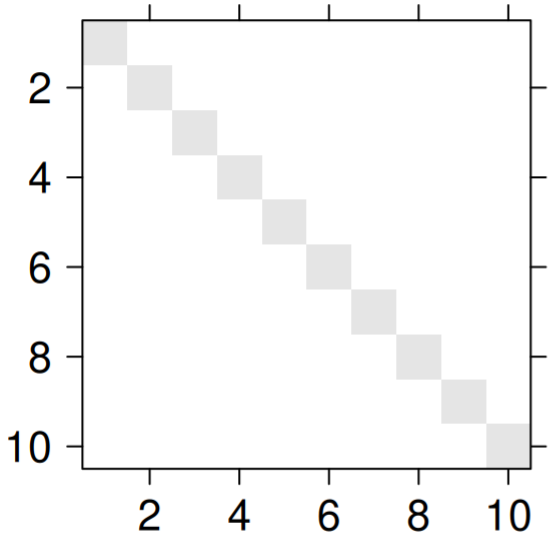
Formulas, interactions, nesting etc.

- Nested: `(1|f/g)` ≡ `(1|f) + (1|f:g)`. Subplots vary within plots, no commonality across plots
- Crossed: `(1|f) + (1|g)`. Years vary; plots vary independently
- Crossed+: `(1|f*g)` ≡ `(1|f) + (1|g) + (1|f:g)`. Years vary, plots vary independently, plots vary within years (need >1 observation per plot/year combination).

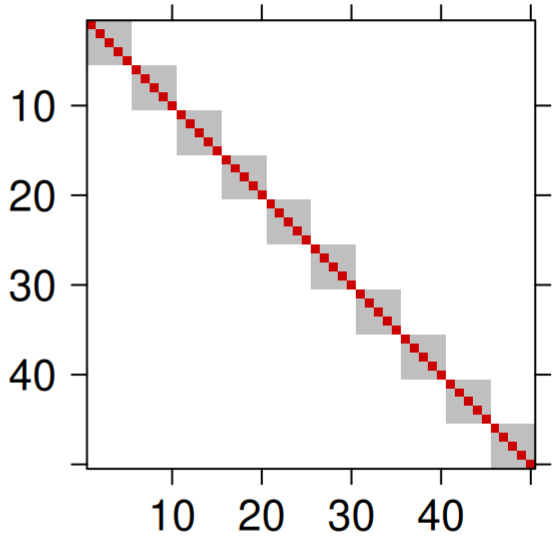
Don't need explicit nesting if sub-groups are uniquely labeled (i.e. `A1`, `A2`, …, `B1`, `B2`, …)

example covariance matrices

- 10 groups, 5 obs/group, intercept RE



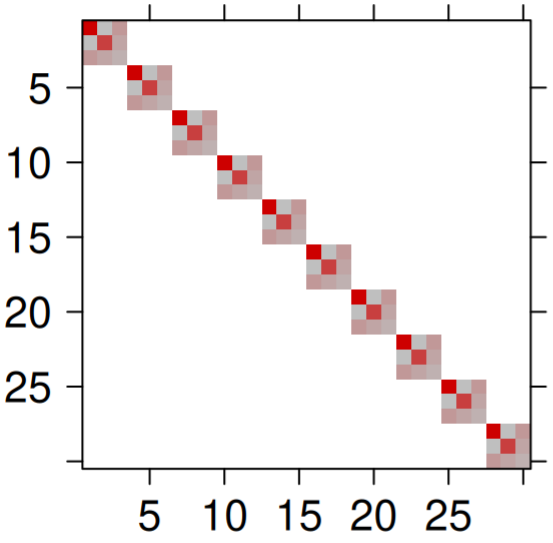
random effects Σ



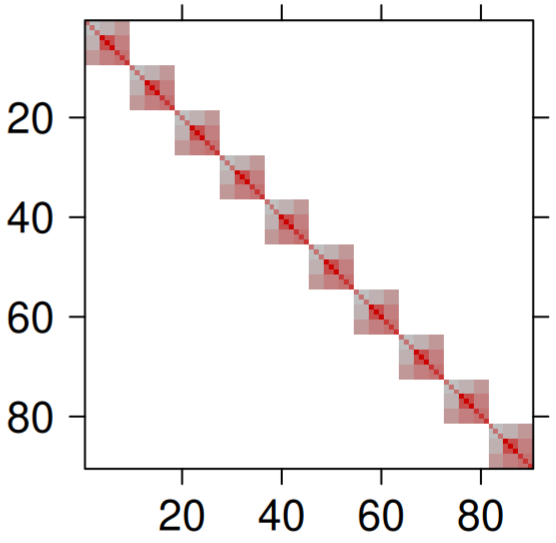
observation Σ

example covariance matrices

- 10 groups, 5 obs/group, 3 effects/group



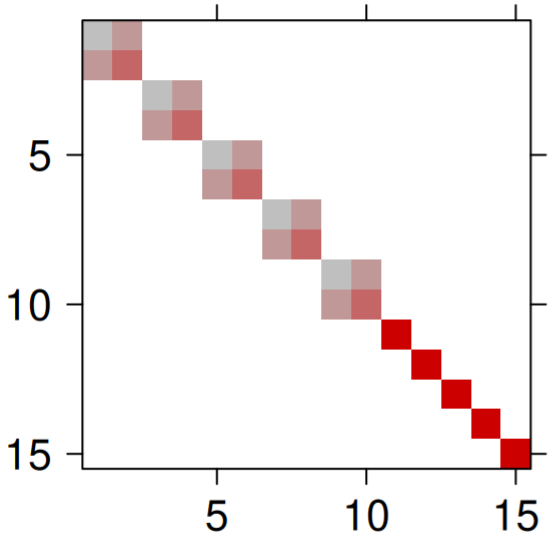
random effects Σ



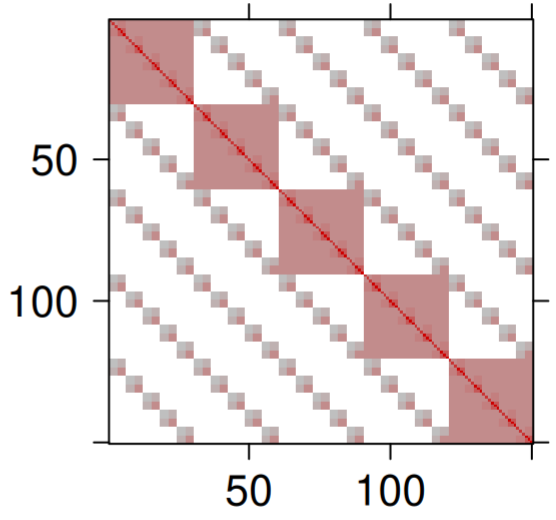
observation Σ

example covariance matrices

- 5 groups, 2 effects (v1) × 5 groups, 1 effect (v1)



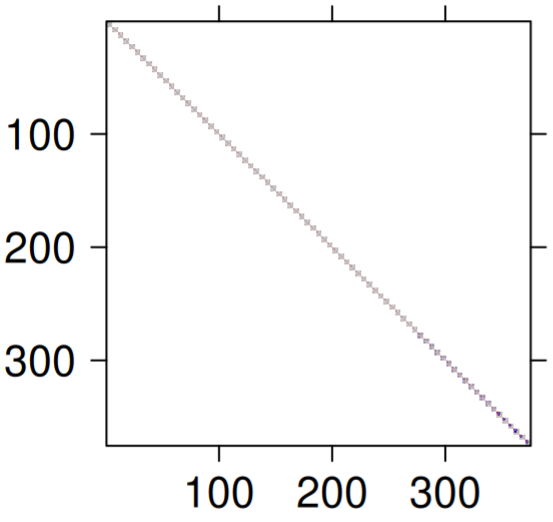
random effects Σ



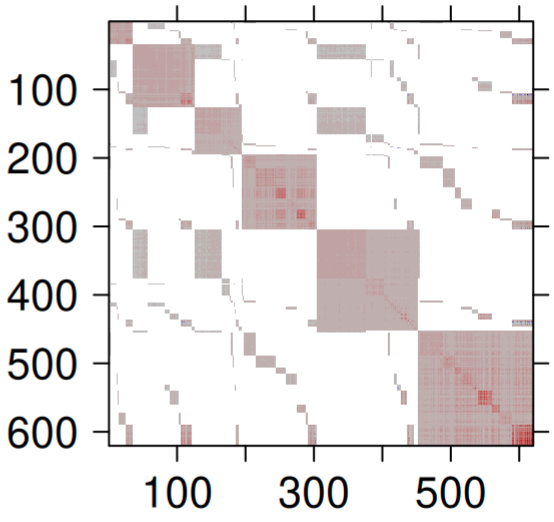
observation Σ

example covariance matrices

Real example (fire/diversity; blocks are geographic regions)



random effects Σ



observation Σ

Factors varying across groups

Can generate large variance-covariance matrices. For a four-level factor, we have β_0 (intercept), β_1 (level 2 - level 1), β_2 (level 3 - level 1), β_3 (level 4 - level 1).

$$\Sigma = \begin{bmatrix} \sigma^2_{\{b|1\}} & \cdot & \cdot & \cdot \\ \sigma_{\{b|1\},\{b|a_{21}\}} & \sigma^2_{\{b|a_{21}\}} & \cdot & \cdot \\ \sigma_{\{b|1\},\{b|a_{31}\}} & \sigma_{\{b|a_{21}\},\{b|a_{31}\}} & \sigma^2_{\{b|a_{31}\}} & \cdot \\ \sigma_{\{b|1\},\{b|a_{41}\}} & \sigma_{\{b|a_{21}\},\{b|a_{41}\}} & \sigma_{\{b|a_{31}\},\{b|a_{41}\}} & \sigma^2_{\{b|a_{41}\}} \end{bmatrix}$$

What is the maximal model?

- Which effects vary *within* which groups?
- If effects don't vary within groups, then we *can't* estimate among-group variation in the effect
  - e.g. in typical clinical trials each patient gets only one treatment (placebo or drug)
  - convenient
  - maybe less powerful (among-group variation is lumped into residual variation)
  - can't evaluate variation in effects

Maximal model is often impractical

- for a RE where n effects vary across clusters we need to estimate $p = n(n+1)/2$ covariance parameters
- e.g. bloodworm example: 5×4×4×2 factorial design (160 combinations), all measured in each of 10 individuals.
  - Treating covariates as numeric, allow all interactions: 16 effects → $p = 136$
  - Categorical predictors, no interactions: 12 effects → $p = 78$
  - Numeric predictors, no interactions: 4 effects → $p = 10$
- **BUT** ignoring within-cluster effects is also dangerous (Schielzeth et al. 2009; Heisig et al. 2019; Oberauer 2022) (!)

Estimation

Maximum likelihood estimation

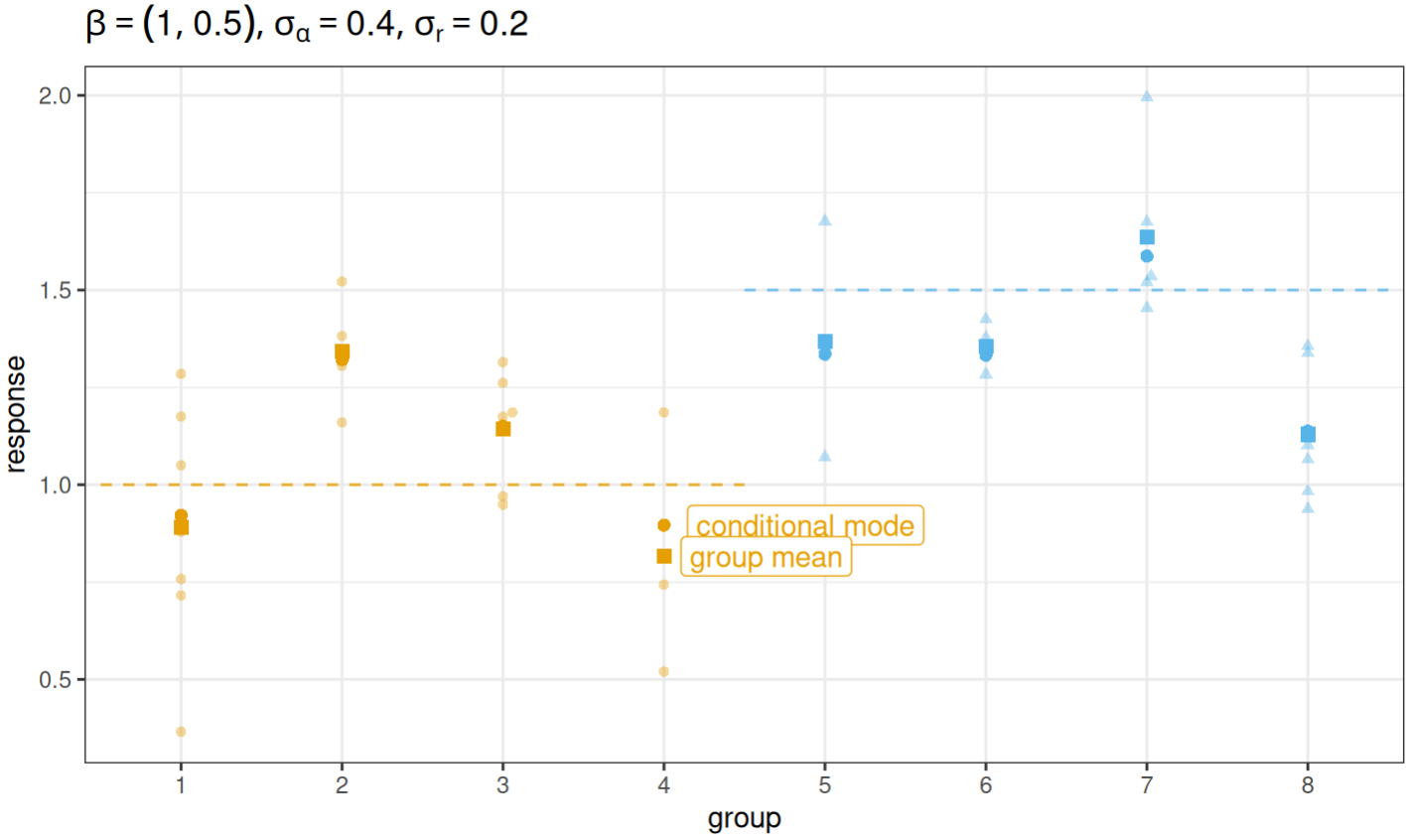
- Best fit is a compromise between
- consistency of data with fixed effects + conditional modes
- consistency of random effect with RE distribution
- Goodness-of-fit *integrates* over the RE distribution

$$\mathcal{L}(x|\beta,\theta) = \iint \mathcal{L}(x|\beta,\boldsymbol{b}) \cdot \mathcal{L}(\boldsymbol{b}|\Sigma(\theta))\, d\boldsymbol{b}$$

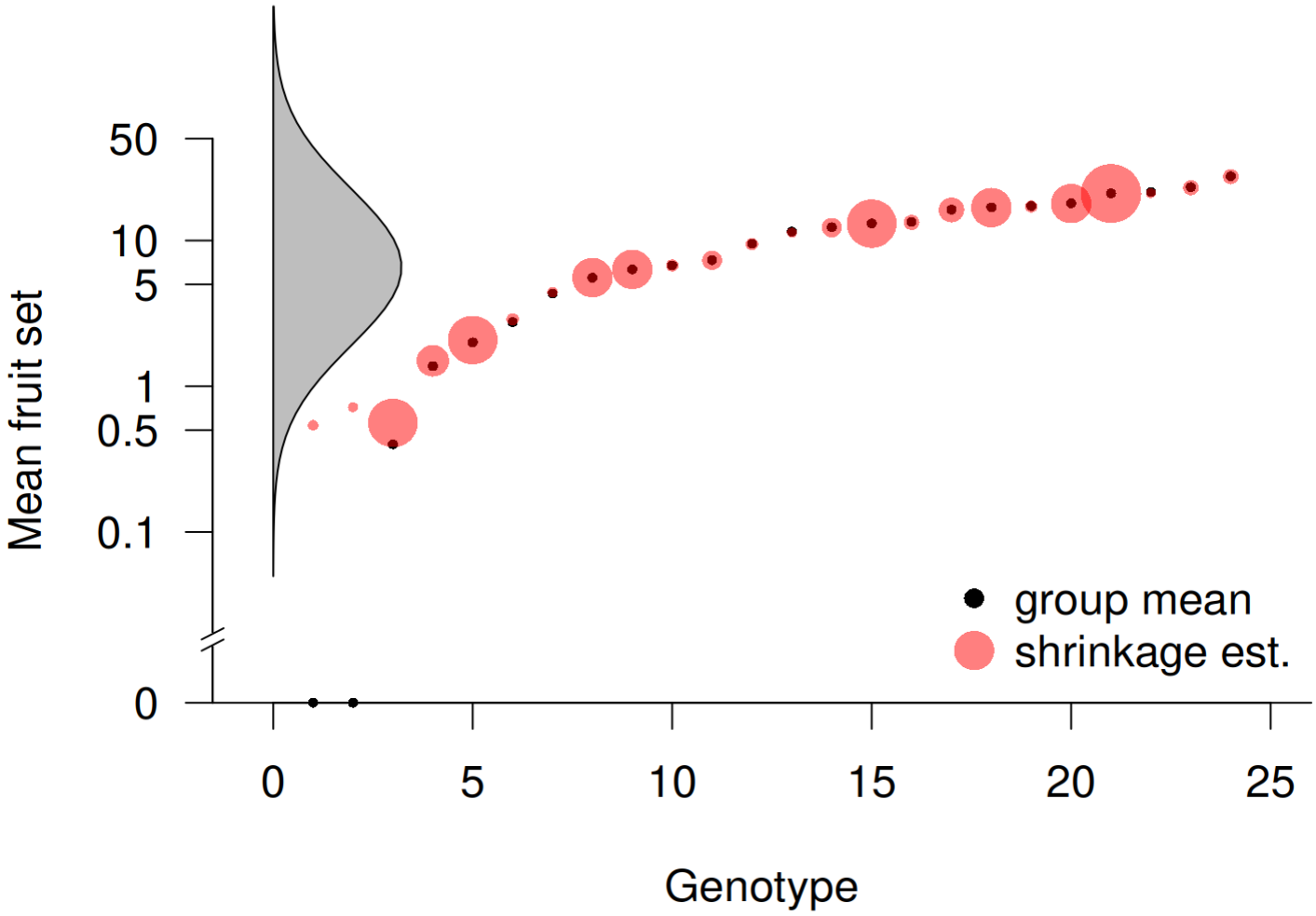
Restricted maximum likelihood (REML)

- factor out fixed effects when estimating variances
- analogous to estimating $s^2 = \frac{\text{SSQ}}{n-1}$ rather than $\frac{\text{SSQ}}{n}$
- typically better variance estimates (unbiased in simple cases)
- defaults differ across packages
- **never** compare models with differing fixed effects fitted by REML 

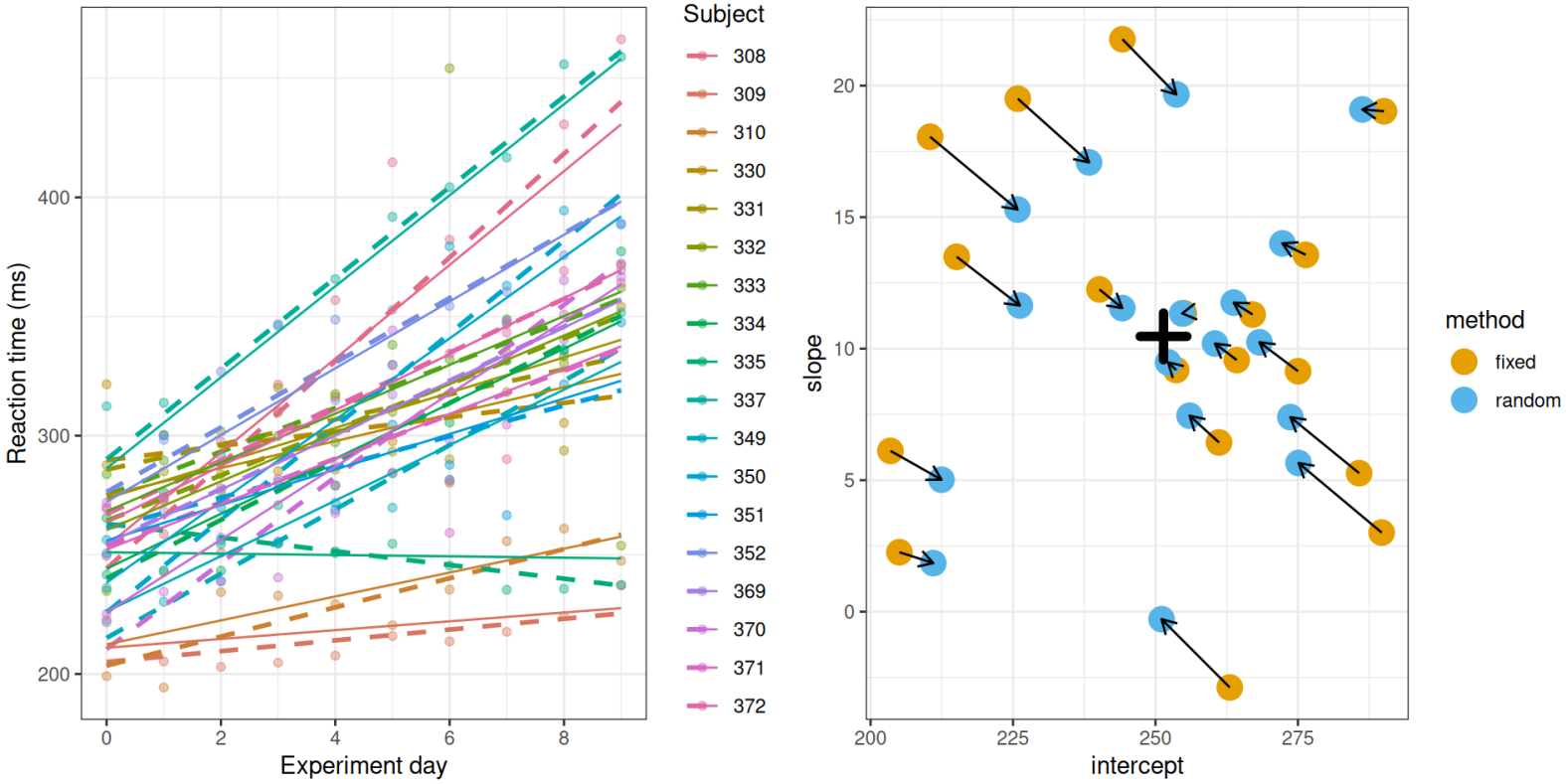
Shrinkage


β = (1, 0.5), σα = 0.4, σr = 0.2
response
group
2.0
1.5
1.0
0.5
1
2
3
4
5
6
7
8
conditional mode
group mean

Arabidopsis shrinkage



Shrinkage in the sleep-study model



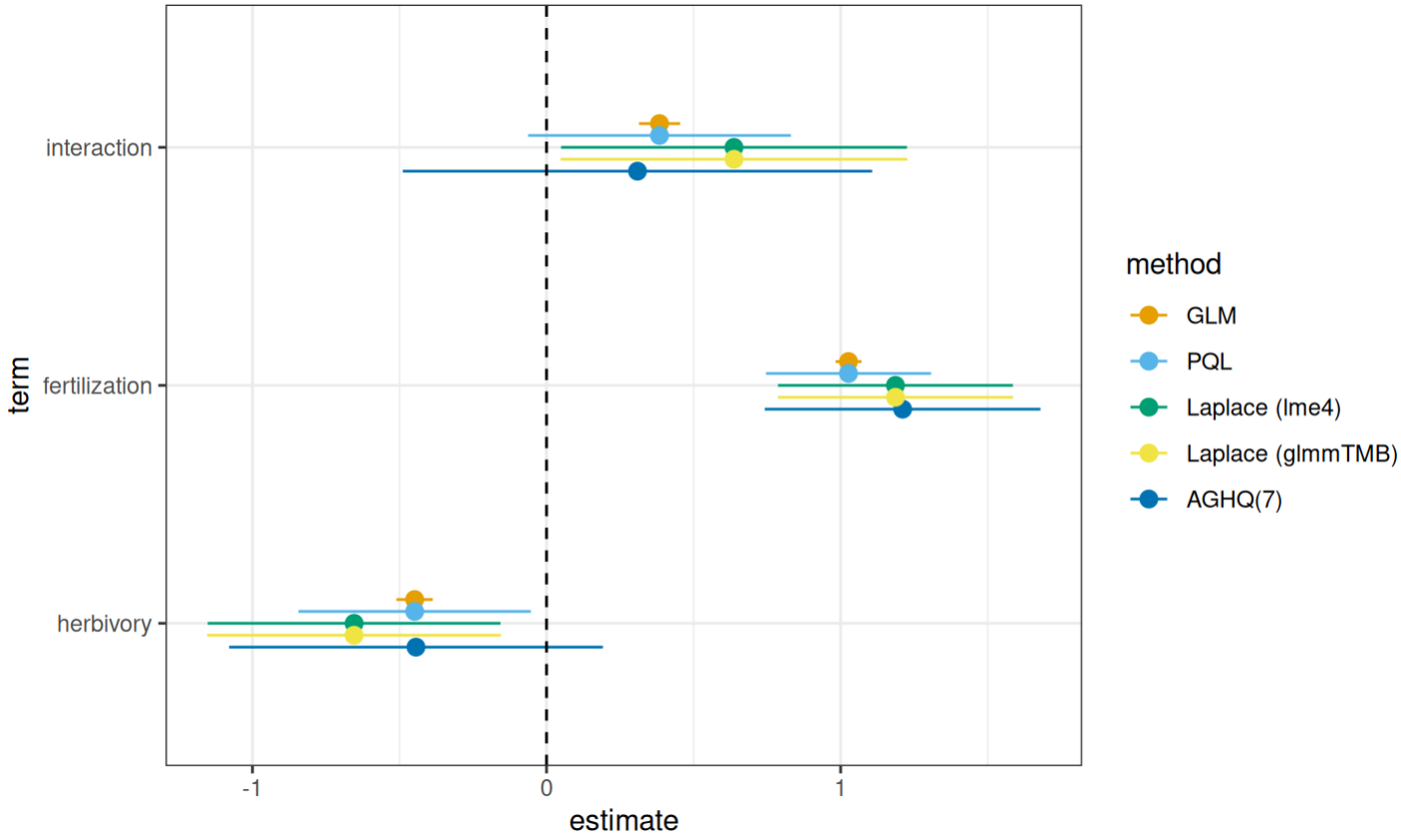
Methods

Estimation methods

Integrating the likelihood is hard for GLMMs

- deterministic approximations:
  - penalized quasi-likelihood (Breslow 2004), Laplace, Gauss-Hermite quadrature, … (Biswas 2015; Bilodeau et al. 2024)
  - best methods needed for small clusters
  - available methods depend on problem structure
  - accuracy (GHQ) vs. flexibility and speed (Laplace, PQL)
- stochastic approximations (Monte Carlo): mostly Bayesian, some frequentist
  - Booth et al. (1999); Gelman et al. (2006); Sung et al. (2007); Ponciano et al. (2009); McElreath (2015)
  - much slower, but flexible and accurate

Estimation: *Arabidopsis*



Inference/post-fitting

Diagnostics

- make sure model is working OK (**singular fits** etc.)
- check model assumptions
  - generalizations of linear/GLM diagnostics: residuals vs fitted, scale-location, etc.
  - conditional modes should be Normally distributed
  - simulation-based checking (`DHARMa`)

Parameter/model interpretation

- fixed effects: same as linear models (population level trends/effects)
- conditional modes/group-level effects
  - deviations from population-level effects
  - can get *conditional standard deviations*, but can't test significance!
- interpreting random effects
  - standard deviations (on same scale as corresponding fixed effects)
  - variances (for population geneticists)

references

Banta, JA et al. 2010. *Oikos* 119 (2) (February): 359–369. doi:10.1111/j.1600-0706.2009.17726.x.

Belenky, G et al. 2003. *Journal of Sleep Research* 12 (1): 1–12. doi:10.1046/j.1365-2869.2003.00337.x.

Bilodeau, B et al. 2024. *Journal of the American Statistical Association* 119 (545) (January): 690–700. doi:10.1080/01621459.2022.2141635.

Biswas, K. 2015. Master's thesis, McMaster University. https://macsphere.mcmaster.ca/bitstream/11375/17272/2/M.Sc_Thesis_final_Keya_Biswas.pdf.

Booth, JG et al. 1999. *Journal of the Royal Statistical Society. Series B* 61 (1): 265–285. doi:10.1111/1467-9868.00176.

Breslow, NE. 2004. In *Proceedings of the second Seattle symposium in biostatistics: Analysis of correlated data*, ed by. Danyu Y. Lin et al., 1–22. Springer.

Crawley, MJ. 2002. *Statistical computing: An introduction to data analysis using S-PLUS*. John Wiley & Sons.

Gelman, A. 2005. *Annals of Statistics* 33 (1): 1–53. doi:doi:10.1214/009053604000001048.

Gelman, A et al. 2006. *Data analysis using regression and Multilevel/Hierarchical models*. Cambridge, England: Cambridge University Press. http://www.stat.columbia.edu/~gelman/arm/.

Heisig, JP et al. 2019. *European Sociological Review* 35 (2) (April): 258–279. doi:10.1093/esr/jcy053.

McElreath, R. 2015. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Boca Raton: Chapman; Hall/CRC.

Murtaugh, PA. 2007. *Ecology* 88 (1): 56–62. doi:10.1890/0012-9658(2007)88[56:SACIED]2.0.CO;2.

Oberauer, K. 2022. *Psychological Science* 33 (4) (April): 648–665. doi:10.1177/09567976211046884.

Ponciano, JM et al. 2009. *Ecology* 90 (2) (February): 356–362. http://www.jstor.org/stable/27650990.

Schielzeth, H et al. 2009. *Behavioral Ecology* 20 (2) (March): 416–420. doi:10.1093/beheco/arn145. http://beheco.oxfordjournals.org/content/20/2/416.

Sung, YJ et al. 2007. *The Annals of Statistics* 35 (3) (July): 990–1011. doi:10.1214/009053606000001389.