Introduction to linear mixed effects models in R

26th January 2022

Nura Sidarus

ROYAL HOLLOWAY UNIVERSITY OF LONDON

@rhulpsychology

Overview

What?

Why?

When?

How?

What are LMEMs?

Aka:

- Multilevel Models
- Hierarchical Models
- Random effects Models
- Variance components Models

➢ An extension of multiple (linear) regression models, to account for hierarchy / structure in the data

Cities

Schools

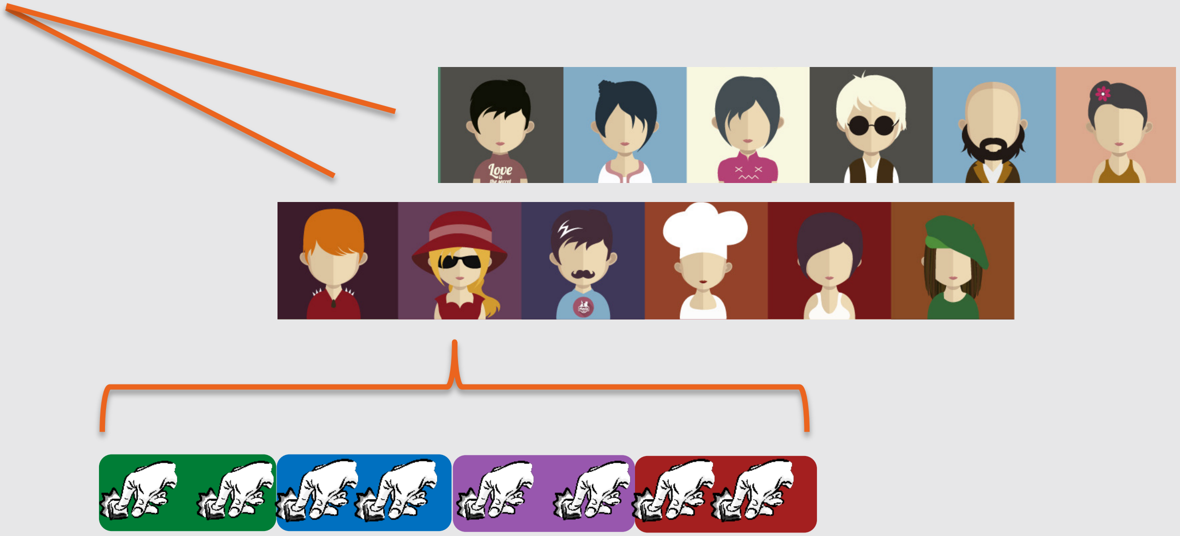
Classrooms

Kids

Group-level manipulation

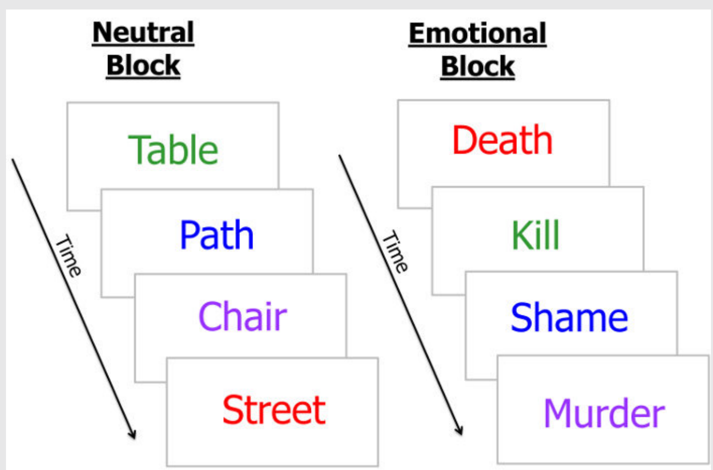
Participants

Repeated Measures



Covariate

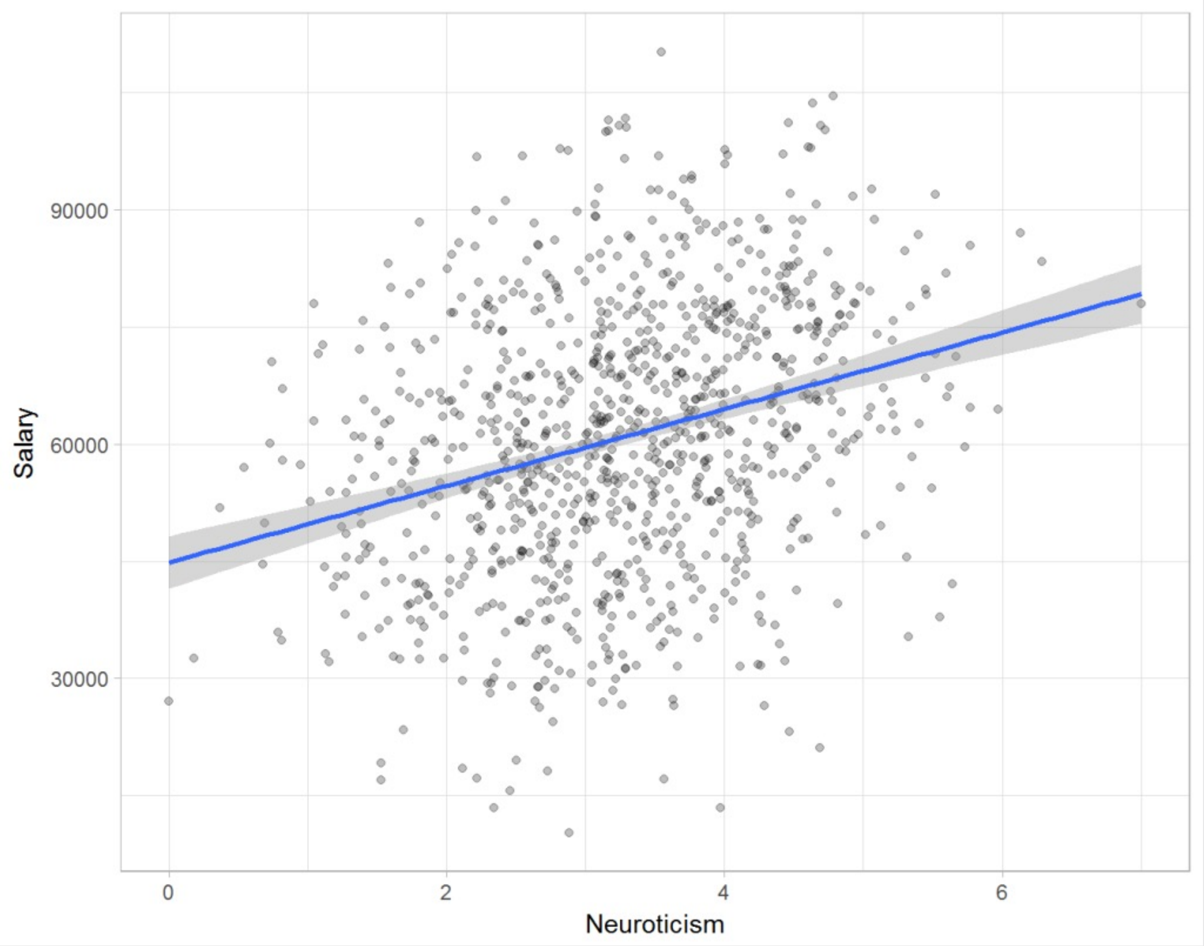
Stimuli



Why?

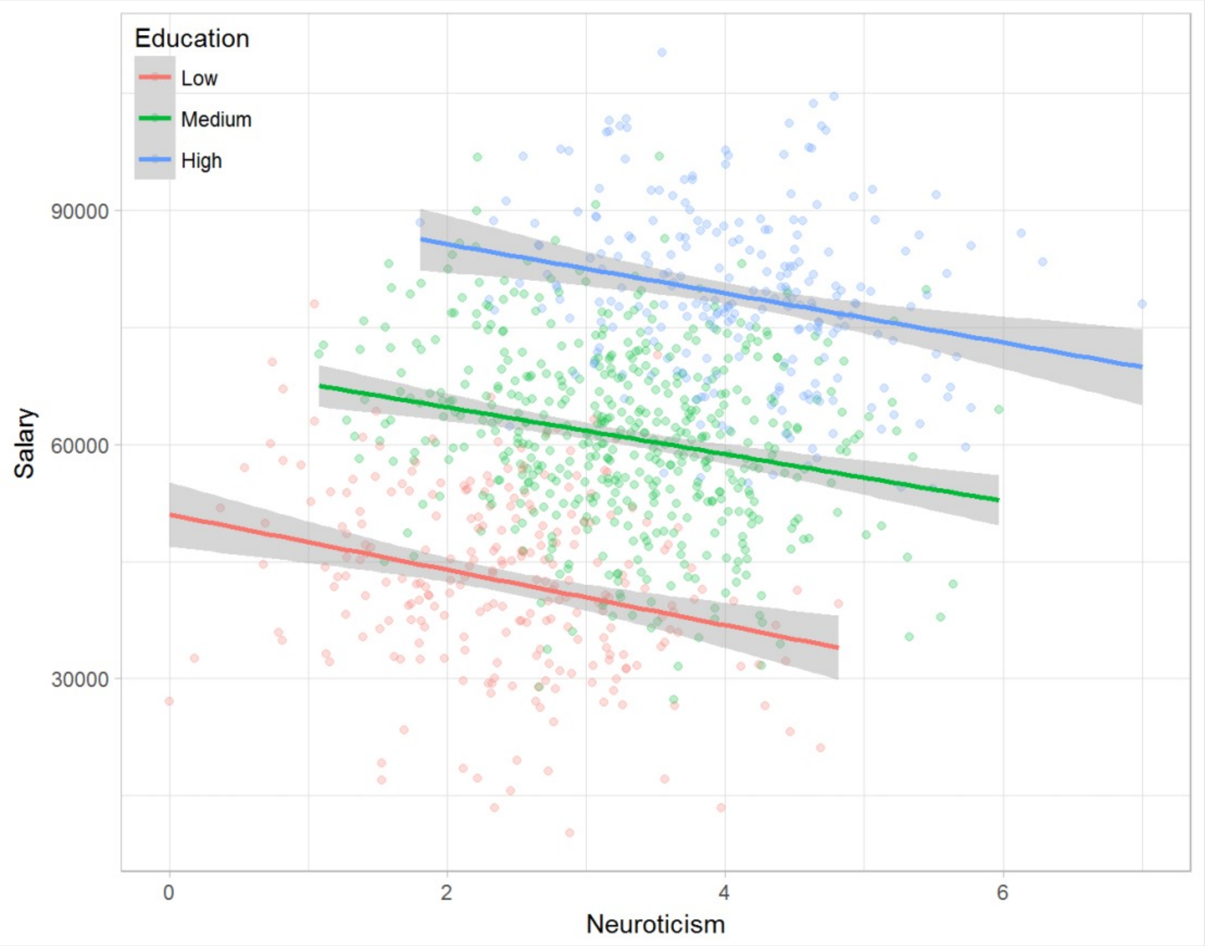
- More accurate inference at population level, better generalisation

Simpson’s Paradox



https://paulvanderlaken.com/2017/09/27/simpsons-paradox-two-hr-examples-with-r-code/

Simpson's Paradox



https://paulvanderlaken.com/2017/09/27/simpsons-paradox-two-hr-examples-with-r-code/

Why?

ROYAL HOLLOWAY UNIVERSITY OF LONDON

- More accurate inference at population level, better generalisation
  - e.g. in crossed designs, accounting for the effect of different "stimuli" allows you to generalise beyond the instances used
- Make better use of your data
  - Use all the data, rather than rely on summary statistics (e.g. mean)
  - Assume properties of clusters of data come from a population
    - Shrinkage / partial pooling
- More flexible:
  - Can easily be extended to add predictors at higher levels, data with more hierarchical levels or cross-classified, multivariate data…
  - Make less assumptions (e.g. sphericity, homogeneity of variance, etc)

When?

- Complex hierarchical structures in data
  - Multiple levels of hierarchy
  - Cross-classified
  - Time series, i.e. longitudinal studies
- Imbalanced or missing data
- Categorical and continuous predictors
  - Where the continuous predictor is of interest
  - And/or to estimate the interaction between categorical and continuous predictor (whereas ANCOVA assumes no interaction)

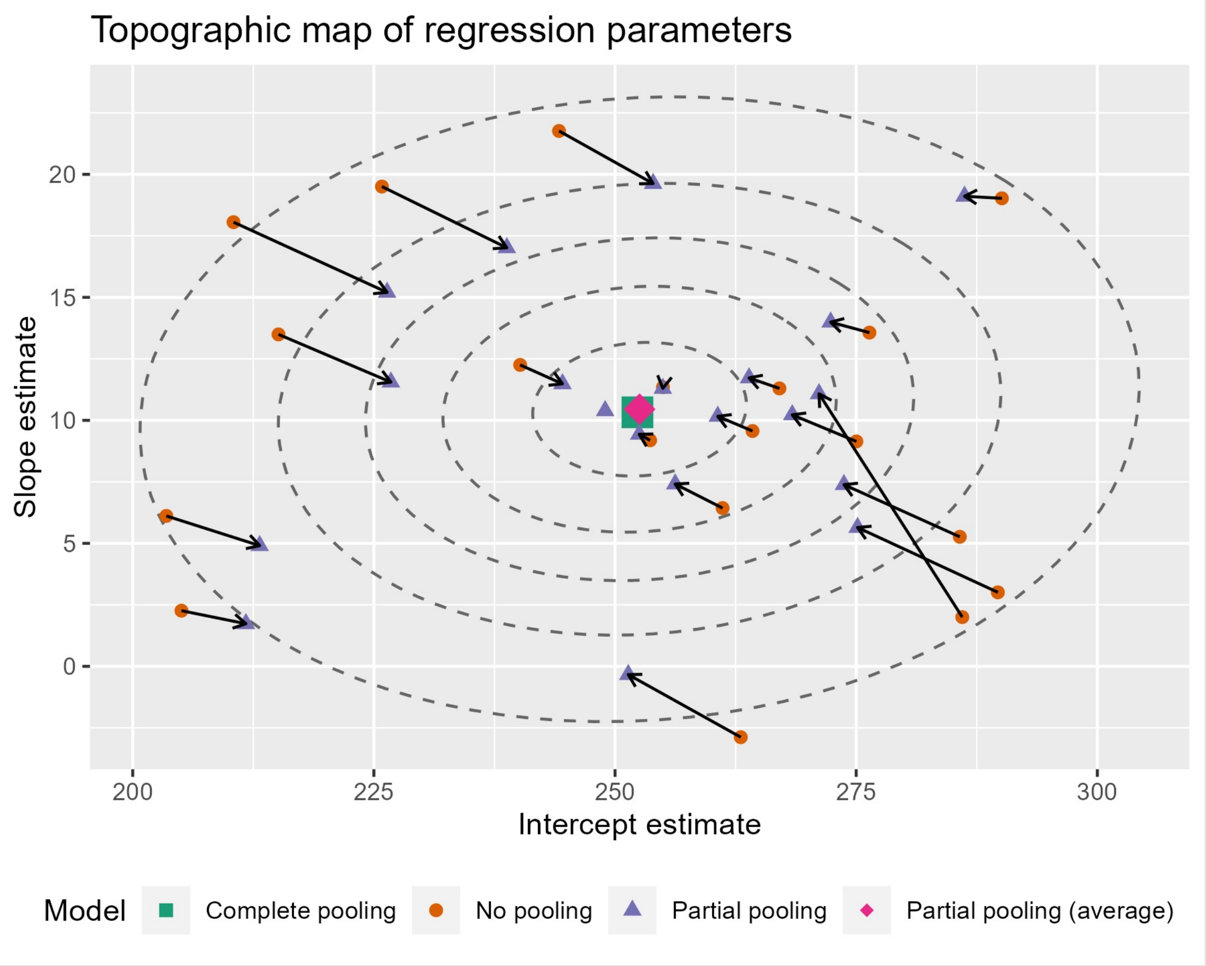
How?

- Observations from a cluster give some information about what might be observed in another cluster drawn from the same population
- But also need to generalise the “right” amount from each cluster, according to what you’ve already learned from the population
- Shrinkage / partial pooling
  - Clusters of data inform estimates of other clusters
  - But, “weight” of each cluster depends on its reliability, i.e. the amount of data in it, and the variation among clusters

For example

- 10 Subjects doing an RT task
  - due to errors, they end up with different number of trials
- Estimate the parameters of the underlying population from which Subjects were drawn
  - Each subject should give some information
  - But should generalise less from unreliable Ss:
    - Ss with fewer trials, as more sampling error
    - Ss with more extreme values ("outliers")
- Hence, model estimates of each participant's RTs (random effects) will be "pulled" towards the population estimates (average RT in task, fixed effects)

Shrinkage / partial pooling



https://www.r-bloggers.com/2017/06/plotting-partial-pooling-in-mixed-effects-models/

Examples

ROYAL HOLLOWAY UNIVERSITY OF LONDON

- Test scores, predicted by previous test scores, across multiple schools

Estimate coefficients (β_0, β_1…)

Fixed / Population Effects

Random / Varying Effects

Estimate variances:

$$\underbrace{y_{ij}}_{\text{Test score}} = \underbrace{\beta_0}_{\text{Overall Intercept}} + \underbrace{\beta_1 x_{ij}}_{\text{Overall Effect of Previous Score}} + \underbrace{u_j}_{\text{Varying Intercept by School}} + \underbrace{e_{ij}}_{\text{Error}}$$

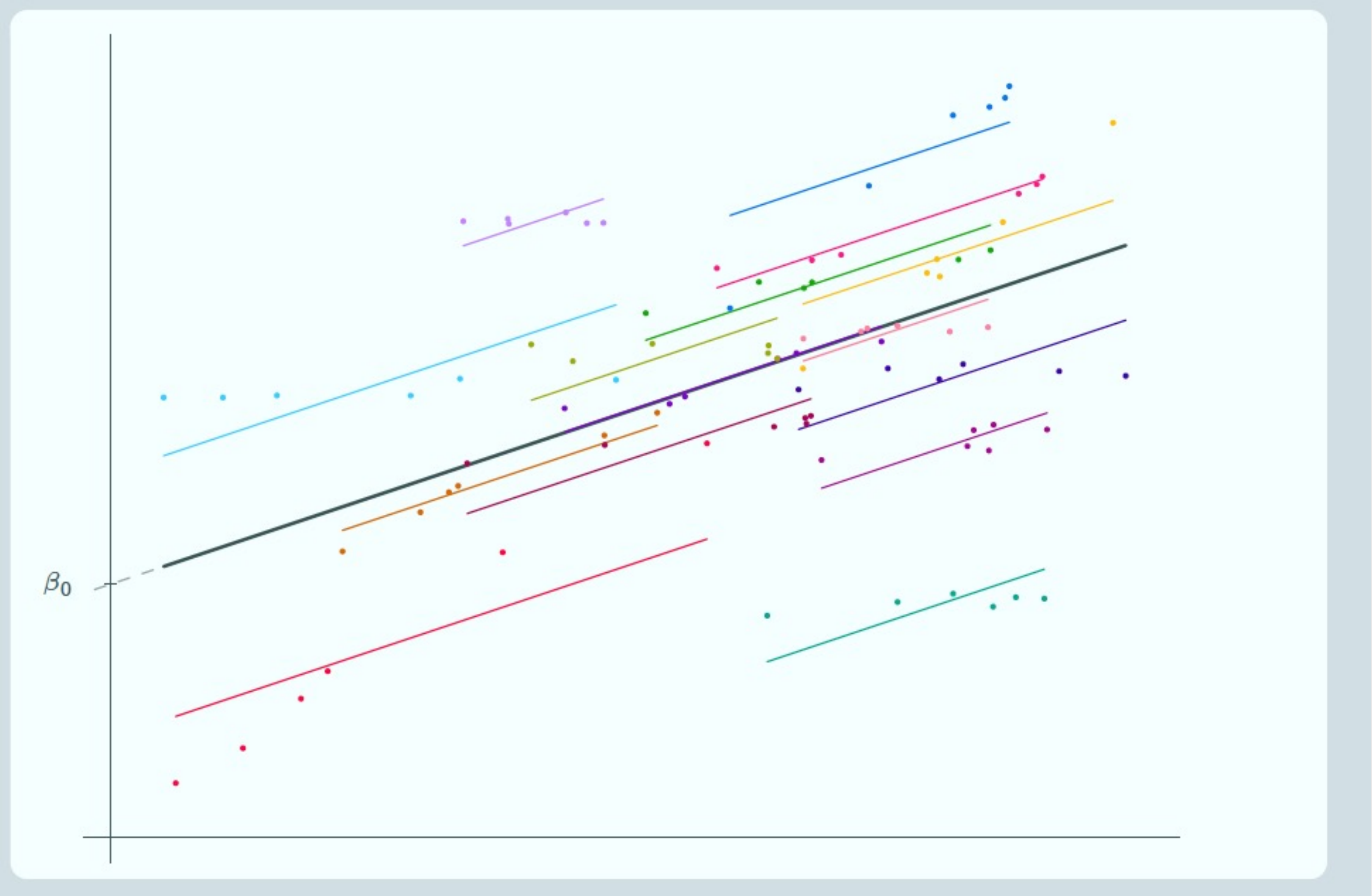
$$e_{ij} \sim \mathrm{N}(0, \sigma_e^2)$$

$$u_j \sim \mathrm{N}(0, \sigma_u^2)$$

http://www.bristol.ac.uk/cmm/software/support/workshops/materials/multilevel-m.html

Examples

Varying Intercepts Model



http://www.bristol.ac.uk/cmm/software/support/workshops/materials/multilevel-m.html

Examples

- Varying Intercepts & Slopes

Varying Slope of Previous Score by School

$$y_{ij} = \beta_0 + \beta_1 \boxed{x_1}_{ij} + u_{0j} + u_{1j} \boxed{x_1}_{ij} + e_{0ij}$$

$$y_{ij} = \beta_0 + (\beta_1 + u_{1ij}) \boxed{x_1}_{ij} + u_{0j} + e_{0ij}$$

Random/varying effects, estimate variances and covariances

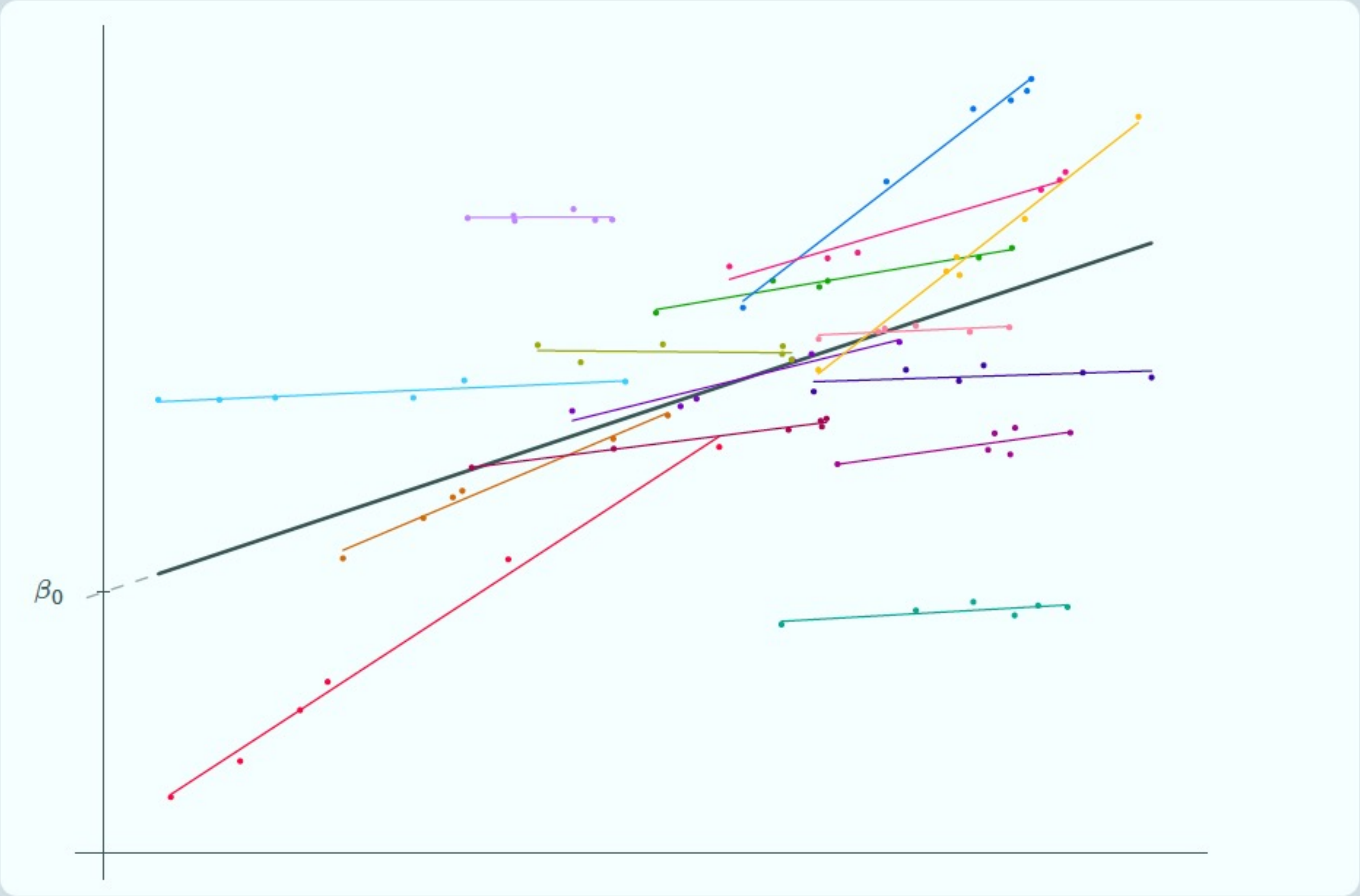
$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim \mathrm{N}(0, \Omega_u), \quad \Omega_u = \begin{bmatrix} \sigma^2_{u0} & \\ \sigma_{u01} & \sigma^2_{u1} \end{bmatrix}$$

$$e_{0ij} \sim \mathrm{N}(0, \sigma^2_{e0})$$

http://www.bristol.ac.uk/cmm/software/support/workshops/materials/multilevel-m.html

Examples

Varying Intercepts & Slopes Model

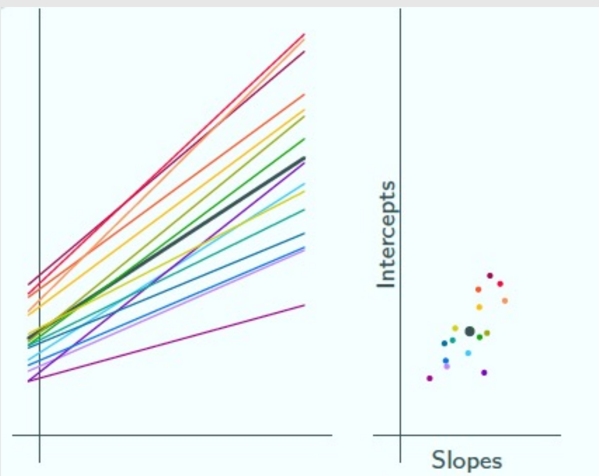


http://www.bristol.ac.uk/cmm/software/support/workshops/materials/multilevel-m.html

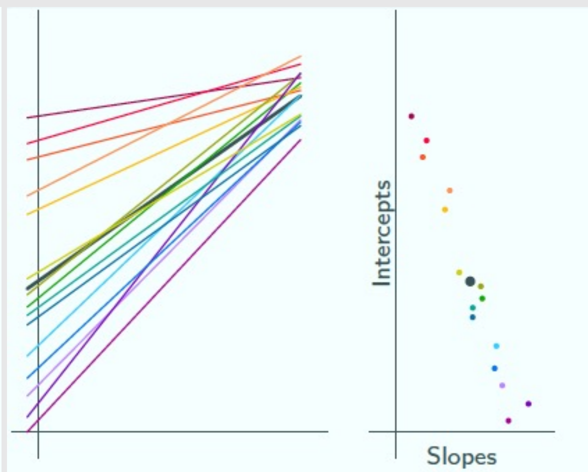
Examples

- Covariance Intercepts & Slopes

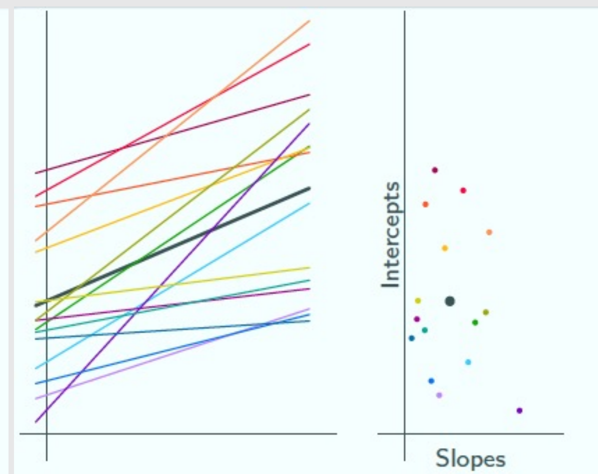
σ_{u01} is positive



σ_{u01} is negative



$\sigma_{u01} = 0$



http://www.bristol.ac.uk/cmm/software/support/workshops/materials/multilevel-m.html

Practicalities – fixed or random effect?

- “Fixed” effects are for the variables for which you want to estimate the overall effect, at the population level
  - E.g. when values of a variable are directly manipulated by the experimenter (e.g. 4 experimental conditions)
  - Or effects you want to account for overall, e.g. a covariate

- ”Random” effects are for those variables where you want to generalise from to the population level
  - Or when the values are not possible to specify / not interesting / infinite (e.g. participants, items)

Practicalities – random effect structure?

- Decide based on theory and study design
  - Some argue for data driven approach, but not considering random effects (e.g. slopes) can lead to over/misestimating the fixed effects (cf. Barr et al 2013 *J Memory Lang*)
- Yet, can be constrained by data and need to prioritise
  - Warning: "model does not converge"
  - May not be able to estimate all varying effects
    - Try removing the correlations between varying effects e.g. (a + b || subj) (cf. Barr et al 2013 *J Memory Lang*)
    - Consider the coding of the predictors (dummy, orthogonal…)
    - Select based on theoretical importance
    - And/or in a data-driven way - model comparison

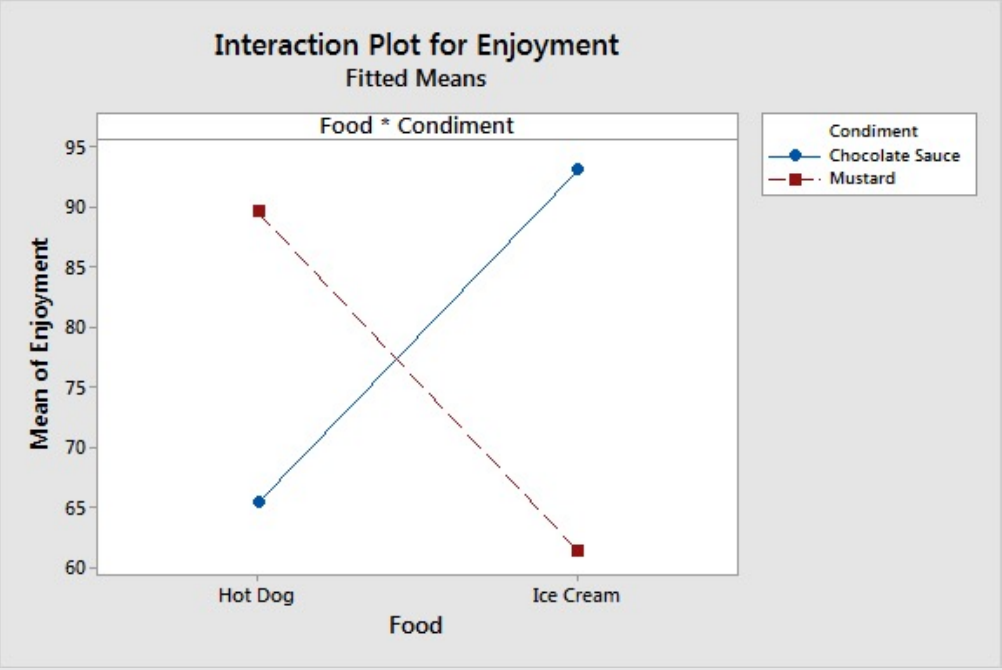
Practicalities – coding of predictors

- How predictors are coded impacts the interpretation of the effects (coefficients)
- But can also affect the model fitting
- Categorical variables:
  - Dummy coding: one group is the reference group, the other is compared to it
  - Sum coding: compare effects to the grand mean
    - Esp. important when there are multiple predictors and interactions
  - NB: if predictor has more than 2 levels, may need other schemes
- Continuous variables:
  - Best to centre the predictors
  - And possibly scale them, to then interpret effects in SD units

Go to code

Interactions

- Adding chocolate increases enjoyment of ice cream, but not of hot dogs
- A drug has an effect on children but not in adults



Specifying an interaction

- Factor_1 * Factor_2
- Interaction + main effects
- Factor_1 : Factor_2
- Interaction only

Output for the interaction

ROYAL HOLLOWAY UNIVERSITY OF LONDON

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: RT ~ luminance * wordness_sum + (1 + wordness_sum | subject) +      (1 | item)
   Data: d
Control: lmerControl(optimizer = "bobyqa")

REML criterion at convergence: 119444.2

Scaled residuals: 
    Min      1Q  Median      3Q     Max 
-5.2704 -0.5477 -0.1467  0.3427 13.1829 

Random effects:
 Groups   Name          Variance Std.Dev. Corr
 item     (Intercept)     867.6   29.45       
 subject  (Intercept)    3464.1   58.86       
          wordness_sum1   172.8   13.15   0.06
 Residual               10316.4  101.57       
Number of obs: 9840, groups:  item, 240; subject, 41

Fixed effects:
                        Estimate Std. Error       df t value Pr(>|t|)    
(Intercept)              587.747      9.789   50.162  60.044  < 2e-16 ***
luminance                  4.743      1.291 9734.485   3.675 0.000239 ***
wordness_sum1            -28.307      3.942  305.058  -7.181 5.33e-12 ***
luminance:wordness_sum1    3.890      1.291 9734.485   3.015 0.002580 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) lumnnc wrdn_1
luminance   -0.264              
wordnss_sm1  0.030  0.000       
lmnnc:wrd_1  0.000  0.000 -0.655
```

Main effects

Output for the interaction

ROYAL HOLLOWAY UNIVERSITY OF LONDON

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: RT ~ luminance * wordness_sum + (1 + wordness_sum | subject) +      (1 | item)
   Data: d
Control: lmerControl(optimizer = "bobyqa")

REML criterion at convergence: 119444.2

Scaled residuals:
    Min      1Q  Median      3Q     Max
-5.2704 -0.5477 -0.1467  0.3427 13.1829

Random effects:
 Groups   Name          Variance Std.Dev. Corr
 item     (Intercept)     867.6   29.45
 subject  (Intercept)    3464.1   58.86
          wordness_sum1   172.8   13.15   0.06
 Residual               10316.4  101.57
Number of obs: 9840, groups:  item, 240; subject, 41

Fixed effects:
                        Estimate Std. Error       df t value Pr(>|t|)
(Intercept)              587.747      9.789   50.162  60.044  < 2e-16 ***
luminance                  4.743      1.291 9734.485   3.675 0.000239 ***
wordness_sum1            -28.307      3.942  305.058  -7.181 5.33e-12 ***
luminance:wordness_sum1    3.890      1.291 9734.485   3.015 0.002580 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) lumnnc wrdn_1
luminance   -0.264
wordnss_sm1  0.030  0.000
lmnnc:wrd_1  0.000  0.000 -0.655
```

Interaction effect

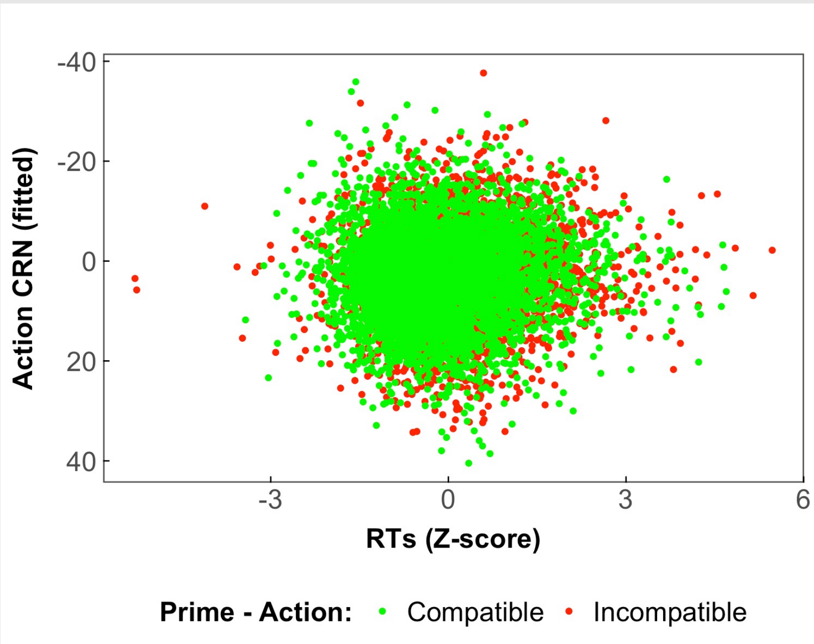
But what does it **mean**?

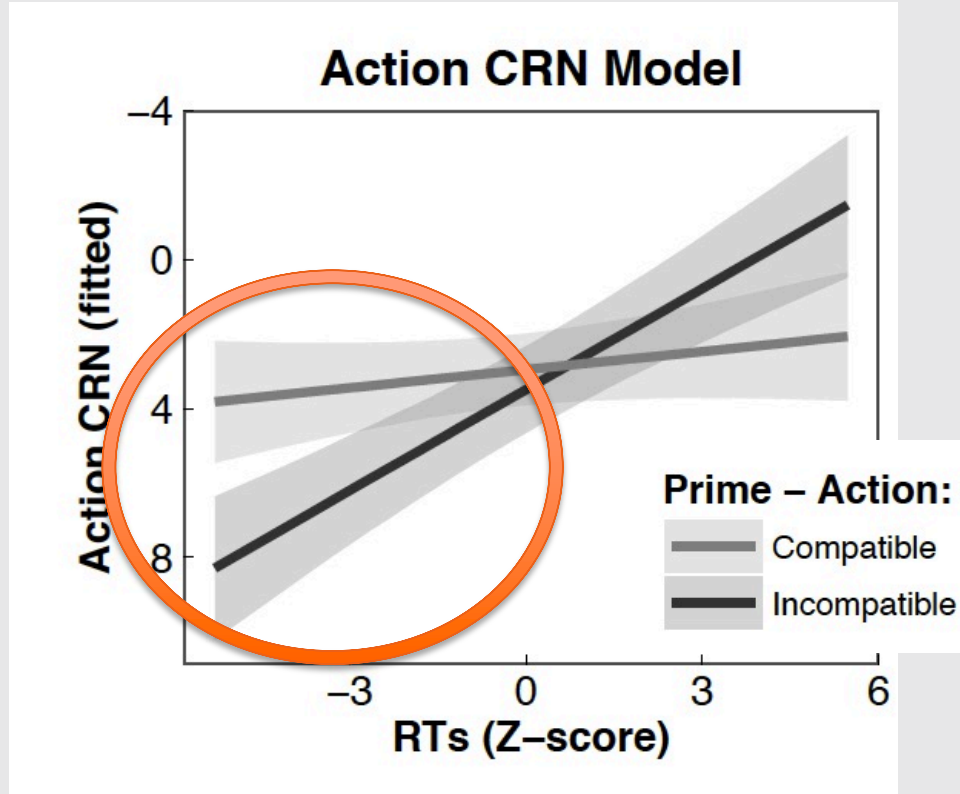
Exploring Interactions – what does it mean?

- ~~Look for the sign (+/-) – not enough~~
- Simple slopes test?
- Plot the data?
  - May not be clear enough, or not depict the relevant model
- Generate model predictions
- General linear hypothesis testing
  - i.e. doing t-tests on point estimates of contrasts

- Recall - helps to use:
  - orthogonal (i.e. sum) contrast coding (*categorical*)
  - centred, or standardised predictors (*continuous*)

Examples

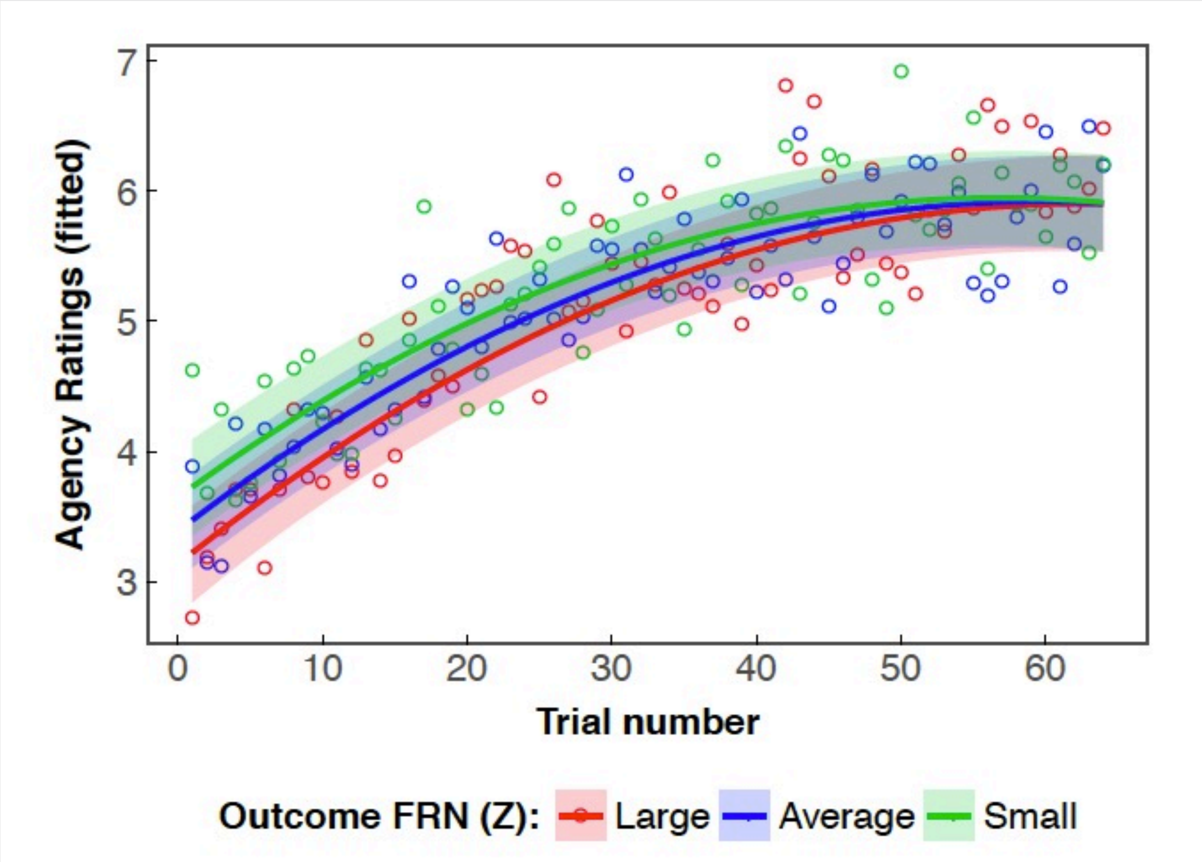




- Effect of priming on Action CRN is limited to average (SD = 0) and fast (SD = 1) RTs

Sidarus, Vuorre & Haggard 2017 *NeuroImage*

Examples



- Outcome FRN is only related to agency ratings at the start of the block

Sidarus, Vuorre & Haggard 2017 *NeuroImage*

Go to code

References & Resources

- DeBruine, L. M., & Barr, D. J. (2021). Understanding Mixed-Effects Models Through Data Simulation. *Advances in Methods and Practices in Psychological Science*, 4(1), 2515245920965119. https://doi.org/10.1177/2515245920965119
  Play with simulations: https://shiny.psy.gla.ac.uk/lmem_sim/

- Richard McElreath – Statistical Rethinking book & lectures on Youtube, e.g. https://www.youtube.com/watch?v=QKYLWSqvtpA&list=PLDcUM9US4XdMdZOhJWJJD4mDBMnbTWw_z&index=17

- Getting confidence intervals for multilevel models (inc. with Bayesian estimation): https://mvuorre.github.io/posts/2016-03-06-multilevel-predictions/

- Gelman & Hill Data 2006 Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press
- Tibon, R., & Levy, D. A. (2015). Striking a balance: analyzing unbalanced event-related potential data. *Quantitative Psychology and Measurement*, *6*, 555.
- http://www.bristol.ac.uk/cmm/software/support/workshops/materials/multilevel-m.html

Thank you for your attention

ROYAL HOLLOWAY UNIVERSITY OF LONDON