Agenda

Multilevel GLM & models of time

1. Administrative
2. Mulilevel generalized linear models (MLGLMs)
3. More on R formulas for ML models
4. Modeling time (part 1)

Slides are licensed under CC BY-NC-SA 4.0
Uploading to generative AI tools or use for training machine learning models is prohibited

ADMINISTRATIVE

Strike update

- Tuesday's stike was a pain, but went smoothly
- Thank you for your patience and understanding!

Plan

- ***Today***: cover MLGLMs and introduce various temporal models
- ***Tuesday***: talk about hazard/duration models and have student presentations

ADMINISTRATIVE

Presentation format

- 20 slides
- Automatically advance every 20 seconds
- Slides can (and probably should) repeat to keep them on screen longer
- This is a short format! focus on big ideas and punchline

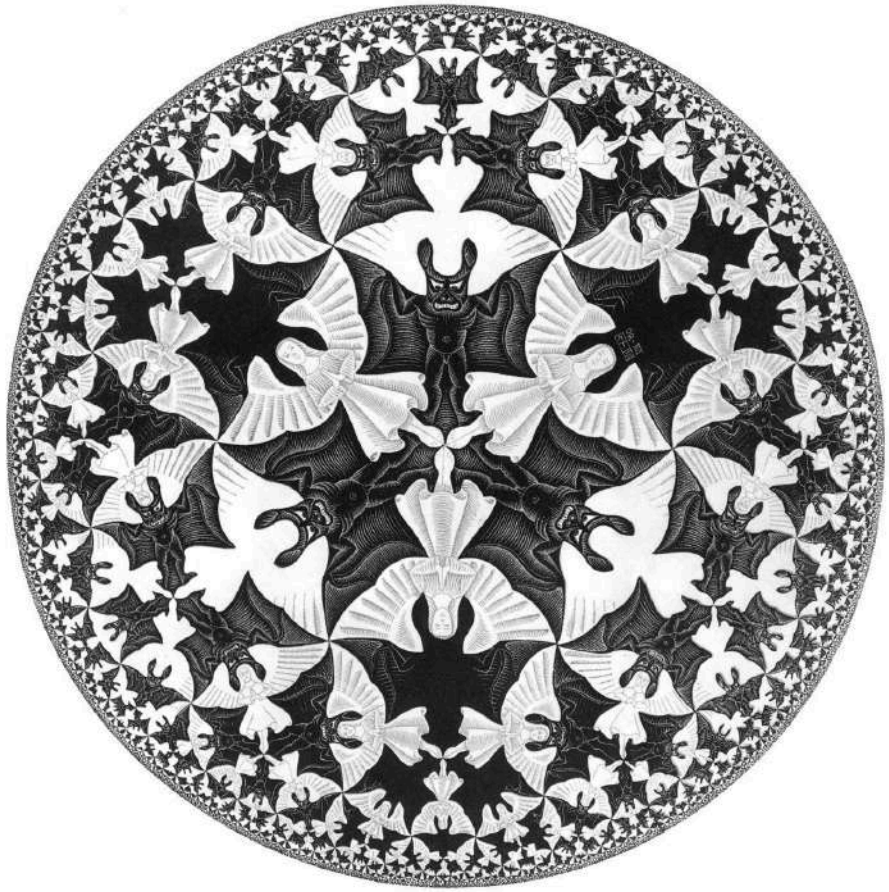
Requirements

- Send me a PDF with *exactly* 20 slides (pages) by 11h59 on Monday, Nov 30
- *Rehearse!*

Provisional order

- Su-Wei
- Sol
- Zixuan
- Yuxin
- Sarah

Multilevel generalized linear models



Generalized multilevel models

Generalized multilevel linear models

Simply add a link function and change the outcome distribution.

E.g. modeling whether a student did better on the math test than the reading test (M_{ik}).

$$M_{ik} \sim \text{Binomial}(1, p_{ik})$$

$$\text{logit}(p_{ik}) = \beta_{0k} + \beta_{1k} Age_i + \beta_{2k} Female_i$$

$$\beta_{0k} = \gamma_{00} + \gamma_{01} TExp_k + \gamma_{02} TFemale_k + \eta_{0k}$$

$$\beta_{1k} = \gamma_{10} + \eta_{1k}$$

$$\beta_{2k} = \gamma_{20} + \gamma_{21} TFemale_k + \eta_{2k}$$

Generalized multilevel models

This is where the 'generalized' part of the model lives

$$M_{ik} \sim \text{Binomial}(1, p_{ik})$$
$$\text{logit}(p_{ik}) = \beta_{0k} + \beta_{1k} Age_i + \beta_{2k} Female_i$$

$$\beta_{0k} = \gamma_{00} + \gamma_{01} TExp_k + \gamma_{02} TFemale_k + \eta_{0k}$$
$$\beta_{1k} = \gamma_{10} + \eta_{1k}$$
$$\beta_{2k} = \gamma_{20} + \gamma_{21} TFemale_k + \eta_{2k}$$

Generalized multilevel models

This is the same as a standard (Gaussian) multilevel model

$$M_{ik} \sim \text{Binomial}(1, p_{ik})$$

$$\text{logit}(p_{ik}) = \beta_{0k} + \beta_{1k} Age_i + \beta_{2k} Female_i$$

$$\beta_{0k} = \gamma_{00} + \gamma_{01} TExp_k + \gamma_{02} TFemale_k + \eta_{0k}$$

$$\beta_{1k} = \gamma_{10} + \eta_{1k}$$

$$\beta_{2k} = \gamma_{20} + \gamma_{21} TFemale_k + \eta_{2k}$$

Generalized multilevel models

Generalized multilevel linear models

Simply add a link function and change the outcome distribution.

E.g. modeling whether a student did better on the math test than the reading test (M_{ik}).

$$M_{ik} \sim \text{Binomial}(1, p_{ik})$$

$$\text{logit}(p_{ik}) = \beta_{0k} + \beta_{1k} Age_i + \beta_{2k} Female_i$$

$$\beta_{0k} = \gamma_{00} + \gamma_{01} TExp_k + \gamma_{02} TFemale_k + \eta_{0k}$$

$$\beta_{1k} = \gamma_{10} + \eta_{1k}$$

$$\beta_{2k} = \gamma_{20} + \gamma_{21} TFemale_k + \eta_{2k}$$

All coefficients and parameters are affected by link function

Interpretation requires careful thinking about the ways that coefficients affect p_{ik}.

Interpreting *direction* of effect (positive vs. negative) is still straightforward. E.g. a strong positive estimate on γ_{02} would suggest that female teachers do a relatively better job of teaching mathematics.

Generalized multilevel models

Gamma–Poisson `family = negbinomial`

$$
\begin{aligned}
Y_{ik} &\sim \text{Pois}(\lambda_{ik}) \\
\lambda_{ik} &\sim \text{Gamma}(\mu_{ik}, \theta) \\
\log(\mu_{ik}) &= \beta_{0k} + \beta_{1k} Age_i + \beta_{2k} Female_i \\
\\
\beta_{0k} &= \gamma_{00} + \gamma_{01} TExp_k + \gamma_{02} TFemale_k + \eta_{0k} \\
\beta_{1k} &= \gamma_{10} + \eta_{1k} \\
\beta_{2k} &= \gamma_{20} + \gamma_{21} TFemale_k + \eta_{2k}
\end{aligned}
$$

Multinomial `family = categorical`

$$
\begin{aligned}
Y_{ik} &\sim \text{Cat}\left(\text{softmax}(s_{0ik}, \ldots, s_{jik})\right) \\
s_{0ik} &= 0 \\
s_{jik} &= \beta_{0k} + \beta_{1k} Age_i + \beta_{2k} Female_i \\
\\
\beta_{0k} &= \gamma_{00} + \gamma_{01} TExp_k + \gamma_{02} TFemale_k + \eta_{0k} \\
\beta_{1k} &= \gamma_{10} + \eta_{1k} \\
\beta_{2k} &= \gamma_{20} + \gamma_{21} TFemale_k + \eta_{2k}
\end{aligned}
$$

Ordinal `family = cumulative`

$$
\begin{aligned}
Y_{ik} &\sim \text{Categorical}(\mathbf{p_{ik}}) \\
p_{jik} &= q_{jik} - q_{(j-1)ik} \\
\text{logit}(q_{jik}) &= \alpha_{jk} - \phi_{ik} \\
\phi_i &= \beta_{1k} Age_i + \beta_{2k} Female_i \\
\\
\alpha_{jk} &= \gamma_{00} + \gamma_{01} TExp_k + \gamma_{02} TFemale_k + \eta_{0k} \\
\beta_{1k} &= \gamma_{10} + \eta_{1k} \\
\beta_{2k} &= \gamma_{20} + \gamma_{21} TFemale_k + \eta_{2k}
\end{aligned}
$$

Same model structure can be applied to any of the GLMs we've looked at.

Specifying these can be a pain, but normally you don't need to mess with the details — they're well implemented in `brms`. Simply specify the correct `family` parameter and multiple equations (multinomial) and intercepts (ordinal) will be created.

See the help for `brmsfamily` in R for the exhaustive list of model families available.

More on specifying MLMs in R



Building a two-level model

A simple two-level model

$$S_{ik} \sim \text{Norm}(\mu_{ik}, \sigma)$$

$$\mu_{ik} = \beta_{0k} + \beta_1 Age_i + \beta_2 Female_i + \beta_{3k} Black_i$$

$$\beta_{0k} = \gamma_{00} + \gamma_{01} TeacherBlack_k + \gamma_{02} PropBlack_k + \eta_{0k}$$

$$\beta_{3k} = \gamma_{30} + \gamma_{31} TeacherBlack_k + \gamma_{32} PropBlack_k + \eta_{3k}$$

Expanded notation

$$\mu_{ik} = \beta_{0k} + \beta_1 Age_i + \beta_2 Female_i + \beta_{3k} Black_i$$

$$\beta_{0k} = \gamma_{00} + \gamma_{11} TeacherBlack_k + \gamma_{051} PropBlack_k + \eta_{0k}$$

$$\beta_{3k} = \gamma_{30} + \gamma_{31} TeacherBlack_k + \gamma_{32} PropBlack_k + \eta_{3k}$$

$$\begin{aligned}\mu_{ik} = {} & (\gamma_{00} + \gamma_{01} TeacherBlack_k + \gamma_{02} PropBlack_k + \eta_{0k}) \\ & + \beta_1 Age_i \\ & + \beta_2 Female_i \\ & + (\gamma_{30} + \gamma_{31} TeacherBlack_k + \gamma_{32} PropBlack_k + \eta_{0k}) Black_i\end{aligned}$$

Expanded notation

$$\mu_{ik} = (\gamma_{00} + \gamma_{01} TeacherBlack_k + \gamma_{02} PropBlack_k + \eta_{0k})$$
$$+ \beta_1 Age_i$$
$$+ \beta_2 Female_i$$
$$+ (\gamma_{30} + \gamma_{31} TeacherBlack_k + \gamma_{32} PropBlack_k + \eta_{0k}) Black_i$$

$$s_{ik} = \gamma_{00} + \gamma_{01} TeacherBlack_k + \gamma_{02} PropBlack_k$$
$$+ \beta_1 Age_i + \beta_2 Female_i + \gamma_{30} Black_i$$
$$+ \gamma_{31} TeacherBlack_k Black_i + \gamma_{32} PropBlack_k Black_i$$
$$+ \eta_{0k} + \eta_{3k} Black_i + \varepsilon_i$$

Expanded notation

Fixed effects

Explained variation in outcome variable.

Describes the way that outcome and predictor variables co-vary.

$$s_{ik} = \gamma_{00} + \gamma_{01} TeacherBlack_k + \gamma_{02} PropBlack_k + \beta_1 Age_i + \beta_2 Female_i + \gamma_{30} Black_i + \gamma_{31} TeacherBlack_k Black_i + \gamma_{32} PropBlack_k Black_i + \eta_{0k} + \eta_{3k} Black_i + \varepsilon_i$$

Random effects

Unexplained variation in outcome variable.

Described in terms of individual variability and different types of group variability.

Building an R formula

With level-two covariates, interactions need to be specified

```
student_reading_score ~
        student_age_s + student_female
        student_re_black*teacher_re_black +
        student_re_black*class_prop_black +
        (1 + student_re_black | teacher_id)
```

$$
\begin{aligned}
s_{ik} = {} & \gamma_{00} + \gamma_{01} TeacherBlack_k + \gamma_{02} PropBlack_k \\
& + \beta_1 Age_i + \beta_2 Female_i + \gamma_{30} Black_i \\
& + \gamma_{31} TeacherBlack_k Black_i + \gamma_{32} PropBlack_k Black_i \\
& + \eta_{0k} + \eta_{3k} Black_i + \varepsilon_i
\end{aligned}
$$

Building an R formula

Outcome variable

```
student_reading_score ~
        student_age_s + student_female
        student_re_black*teacher_re_black +
        student_re_black*class_prop_black +
        (1 + student_re_black | teacher_id)
```

$$
\begin{aligned}
s_{ik} = {} & \gamma_{00} + \gamma_{01} TeacherBlack_k + \gamma_{02} PropBlack_k \\
& + \beta_1 Age_i + \beta_2 Female_i + \gamma_{30} Black_i \\
& + \gamma_{31} TeacherBlack_k Black_i + \gamma_{32} PropBlack_k Black_i \\
& + \eta_{0k} + \eta_{3k} Black_i + \varepsilon_i
\end{aligned}
$$

Building an R formula

Global intercept (included automatically)

```
student_reading_score ~
        student_age_s + student_female
        student_re_black*teacher_re_black +
        student_re_black*class_prop_black +
        (1 + student_re_black | teacher_id)
```

$$
\begin{aligned}
s_{ik} = & \boxed{\gamma_{00}} + \gamma_{01} TeacherBlack_k + \gamma_{02} PropBlack_k \\
& + \beta_1 Age_i + \beta_2 Female_i + \gamma_{30} Black_i \\
& + \gamma_{31} TeacherBlack_k Black_i + \gamma_{32} PropBlack_k Black_i \\
& + \eta_{0k} + \eta_{3k} Black_i + \varepsilon_i
\end{aligned}
$$

Building an R formula

Interactions (*) automatically include standalone terms

```
student_reading_score ~
        student_age_s + student_female
        student_re_black*teacher_re_black +
        student_re_black*class_prop_black +
        (1 + student_re_black | teacher_id)
```

$$
\begin{aligned}
s_{ik} = \gamma_{00} &+ \boxed{\gamma_{01} TeacherBlack_k} + \gamma_{02} PropBlack_k \\
&+ \beta_1 Age_i + \beta_2 Female_i + \boxed{\gamma_{30} Black_i} \\
&+ \boxed{\gamma_{31} TeacherBlack_k Black_i} + \gamma_{32} PropBlack_k Black_i \\
&+ \eta_{0k} + \eta_{3k} Black_i + \varepsilon_i
\end{aligned}
$$

Building an R formula

Interactions (*) automatically include standalone terms

```
student_reading_score ~
        student_age_s + student_female
        student_re_black*teacher_re_black +
        student_re_black*class_prop_black +
        (1 + student_re_black | teacher_id)
```

$$
\begin{aligned}
s_{ik} = {} & \gamma_{00} + \gamma_{01} TeacherBlack_k + \boxed{\gamma_{02} PropBlack_k} \\
& + \beta_1 Age_i + \beta_2 Female_i + \boxed{\gamma_{30} Black_i} \\
& + \gamma_{31} TeacherBlack_k Black_i + \boxed{\gamma_{32} PropBlack_k Black_i} \\
& + \eta_{0k} + \eta_{3k} Black_i + \varepsilon_i
\end{aligned}
$$

Building an R formula

Redundant terms (`student_re_black`) are not added twice

```
student_reading_score ~
        student_age_s + student_female
        student_re_black*teacher_re_black +
        student_re_black*class_prop_black +
        (1 + student_re_black | teacher_id)
```

$$
\begin{aligned}
s_{ik} = {} & \gamma_{00} + \gamma_{01} TeacherBlack_k + \gamma_{02} PropBlack_k \\
& + \beta_1 Age_i + \beta_2 Female_i + \gamma_{30} Black_i \\
& + \gamma_{31} TeacherBlack_k Black_i + \gamma_{32} PropBlack_k Black_i \\
& + \eta_{0k} + \eta_{3k} Black_i + \varepsilon_i
\end{aligned}
$$

Building an R formula

Random effects use pipe notation (|)

```
student_reading_score ~
        student_age_s + student_female
        student_re_black*teacher_re_black +
        student_re_black*class_prop_black +
        (1 + student_re_black | teacher_id)
```

$$
\begin{aligned}
s_{ik} = {} & \gamma_{00} + \gamma_{01} TeacherBlack_k + \gamma_{02} PropBlack_k \\
& + \beta_1 Age_i + \beta_2 Female_i + \gamma_{30} Black_i \\
& + \gamma_{31} TeacherBlack_k Black_i + \gamma_{32} PropBlack_k Black_i \\
& + \eta_{0k} + \eta_{3k} Black_i + \varepsilon_i
\end{aligned}
$$

Building an R formula

Grouping elements after the pipe

```
student_reading_score ~
        student_age_s + student_female
        student_re_black*teacher_re_black +
        student_re_black*class_prop_black +
        (1 + student_re_black | teacher_id)
```

$$
\begin{aligned}
s_{ik} = {} & \gamma_{00} + \gamma_{01} TeacherBlack_k + \gamma_{02} PropBlack_k \\
& + \beta_1 Age_i + \beta_2 Female_i + \gamma_{30} Black_i \\
& + \gamma_{31} TeacherBlack_k Black_i + \gamma_{32} PropBlack_k Black_i \\
& + \eta_{0k} + \eta_{3k} Black_i + \varepsilon_i
\end{aligned}
$$

Building an R formula

Random intercepts indicated with constant (1)

```
student_reading_score ~
        student_age_s + student_female
        student_re_black*teacher_re_black +
        student_re_black*class_prop_black +
        (1 + student_re_black | teacher_id)
```

$$
\begin{aligned}
s_{ik} = {} & \gamma_{00} + \gamma_{01} TeacherBlack_k + \gamma_{02} PropBlack_k \\
& + \beta_1 Age_i + \beta_2 Female_i + \gamma_{30} Black_i \\
& + \gamma_{31} TeacherBlack_k Black_i + \gamma_{32} PropBlack_k Black_i \\
& + \boxed{\eta_{0k}} + \eta_{3k} Black_i + \varepsilon_i
\end{aligned}
$$

Building an R formula

Random-slope variables included in grouping expression

```
student_reading_score ~
        student_age_s + student_female
        student_re_black*teacher_re_black +
        student_re_black*class_prop_black +
        (1 + student_re_black | teacher_id)
```

$$
\begin{aligned}
s_{ik} = {} & \gamma_{00} + \gamma_{01} TeacherBlack_k + \gamma_{02} PropBlack_k \\
& + \beta_1 Age_i + \beta_2 Female_i + \gamma_{30} Black_i \\
& + \gamma_{31} TeacherBlack_k Black_i + \gamma_{32} PropBlack_k Black_i \\
& + \eta_{0k} + \eta_{3k} Black_i + \varepsilon_i
\end{aligned}
$$

Models of time



COMMON STRATEGIES FOR TIME

Autoregression Models
(and Gaussian processes)

Model outcome at time *t* as a function of covariates and outcome at time *t* – 1

$$y_t = y_{t-1} + \beta \mathbf{X}_{t-1} + \varepsilon_t$$

$$y_{\Delta t} = \beta \mathbf{X}_{t-1} + \varepsilon_t$$

Survival / hazard / duration models

Model the timing of a one-time event (graduation, job acquisition, death)

More on this next class

$$\lambda(t|\mathbf{X}) = \lambda_0 \exp(\beta \mathbf{X})$$

Embedding time within observations
(or observations within time)

Fit temporal structure into a multilevel model

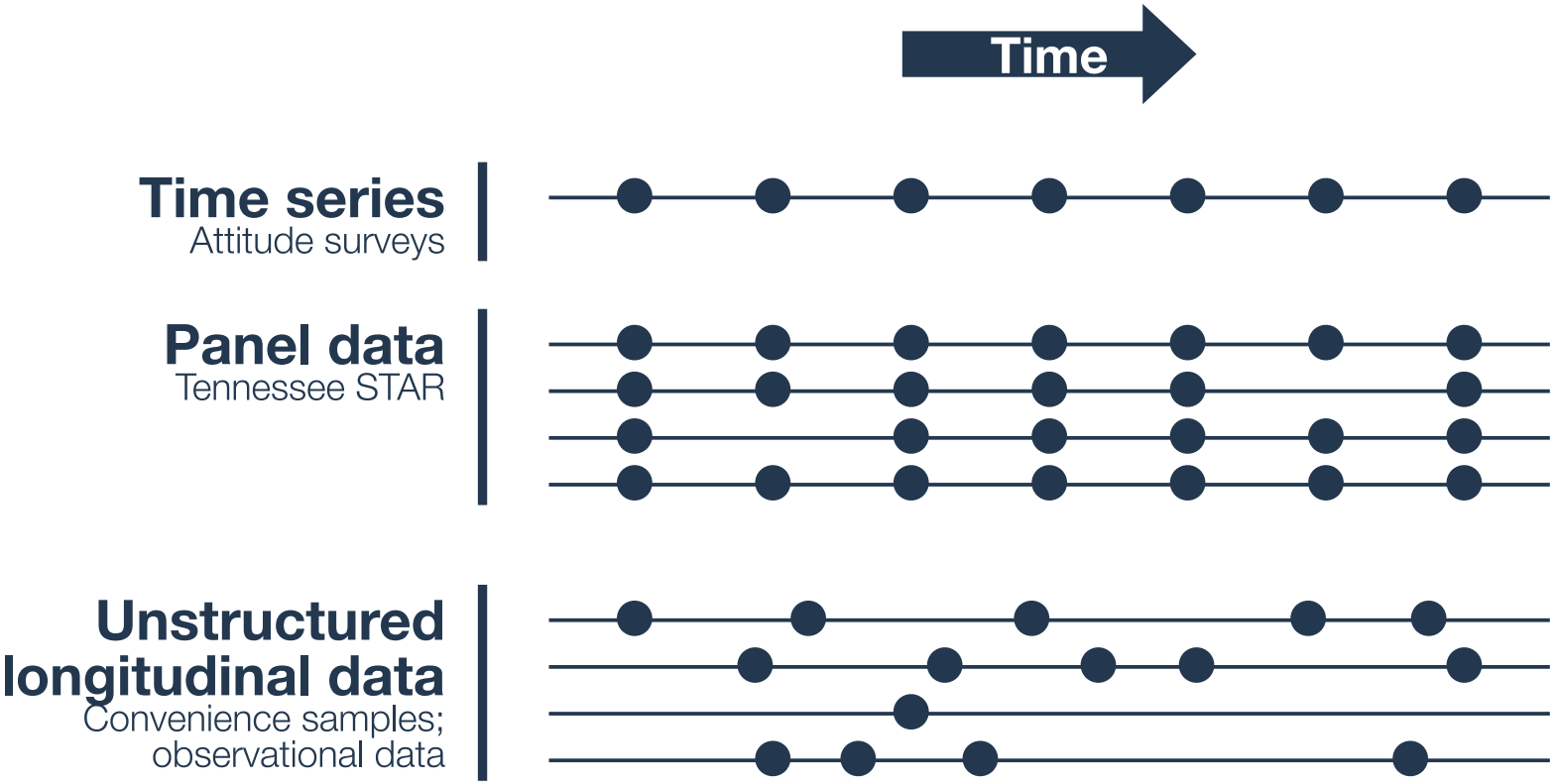
More on this in a moment

Ad hoc models

Countless context-specific ways to model a randomly varying or functionally defined effect of time on outcomes

Models of time

Common temporal data structures



Nesting time and students



Student scores over time

Correlation between scores from the same student.

	1985	1986	1987	1988
Student 1	540	531	564	563
Student 2	479	487	505	510
Student 3	503	505	501	503
Student 4	461	471	—	—
Student 5	525	—	506	541
…	…	…	…	…

Student scores over time

	1985	1986	1987	1988
Student 1	540	531	564	563
Student 2	479	487	505	510
Student 3	503	505	501	503
Student 4	461	471	—	—
Student 5	525	—	506	541
…	…	…	…	…

Scores tend to increase over time.

Student scores over time

“Wide” vs “tall” tabular data:

	1985	1986	1987	1988
Student 1	540	531	564	563
Student 2	479	487	505	510
Student 3	503	505	501	503
Student 4	461	471	—	—
Student 5	525	—	506	541
	…	…	…	…

Wide



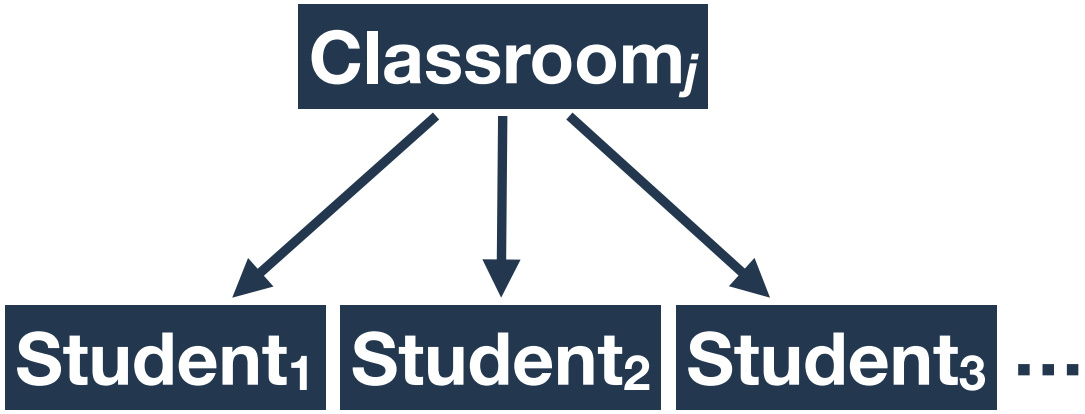
Student	Year	Score
1	1985	540
1	1986	531
1	1987	564
1	1988	563
2	1985	479
2	1986	487
2	1987	505
2	1988	510
3	1985	503
…	…	…

Tall

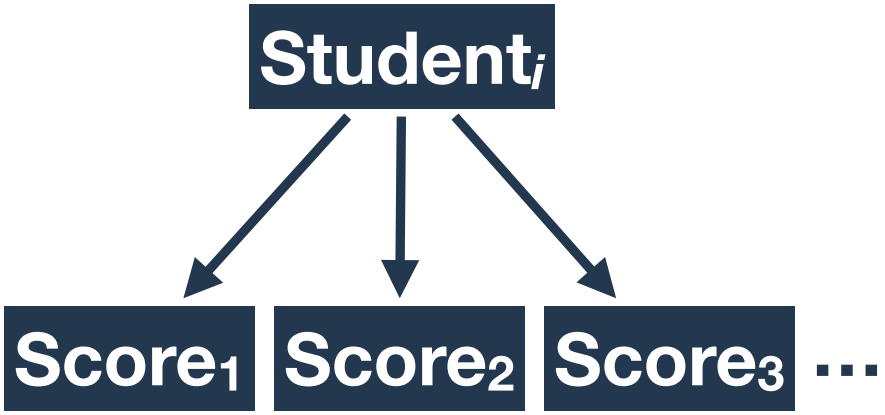
Student random effects

Students groups into classrooms

Single year



Test scores grouped by student



Student random effects

Score for student *i* at time *t*.

$$S_{ti} \sim \text{Norm}(\mu_{ti}, \sigma)$$

Average score for student *i*.

$$\mu_{ti} = \beta_{0i} + \beta_1 CSize_{ti}$$

$$\beta_{0i} = \gamma_{00} + \eta_{0i}$$

Average score across all students.

$$\eta_{0i} \sim \text{Norm}(0, \phi_0)$$

Student random effects

$$S_{ti} \sim \text{Norm}(\mu_{ti}, \sigma)$$
$$\mu_{ti} = \beta_{0i} + \beta_1 CSize_{ti}$$
$$\beta_{0i} = \gamma_{00} + \eta_{0i}$$

$$\eta_{0i} \sim \text{Norm}(0, \phi_0)$$

$$\gamma_{00} \sim \text{Norm}(500, 100)$$
$$\beta_1 \sim \text{Norm}(0, 50)$$
$$\sigma \sim \text{HalfCauchy}(0, 50)$$
$$\phi_0 \sim \text{HalfCauchy}(0, 50)$$

	Mean	*90% credible interval*	
γ_{00}	547.06	544.05	549.94
β_1	1.23	0.58	1.87
σ	56.94	54.97	58.93
ϕ_0	33.91	30.16	37.60

Linear time trend

$$
\begin{aligned}
S_{ti} &\sim \text{Norm}(\mu_{ti}, \sigma) \\
\mu_{ti} &= \beta_{0i} + \beta_{1i} Year_{ti} + \beta_2 CSize_{ti} \\
\beta_{0i} &= \gamma_{00} + \eta_{0i} \\
\beta_{1i} &= \gamma_{10} + \eta_{1i}
\end{aligned}
$$

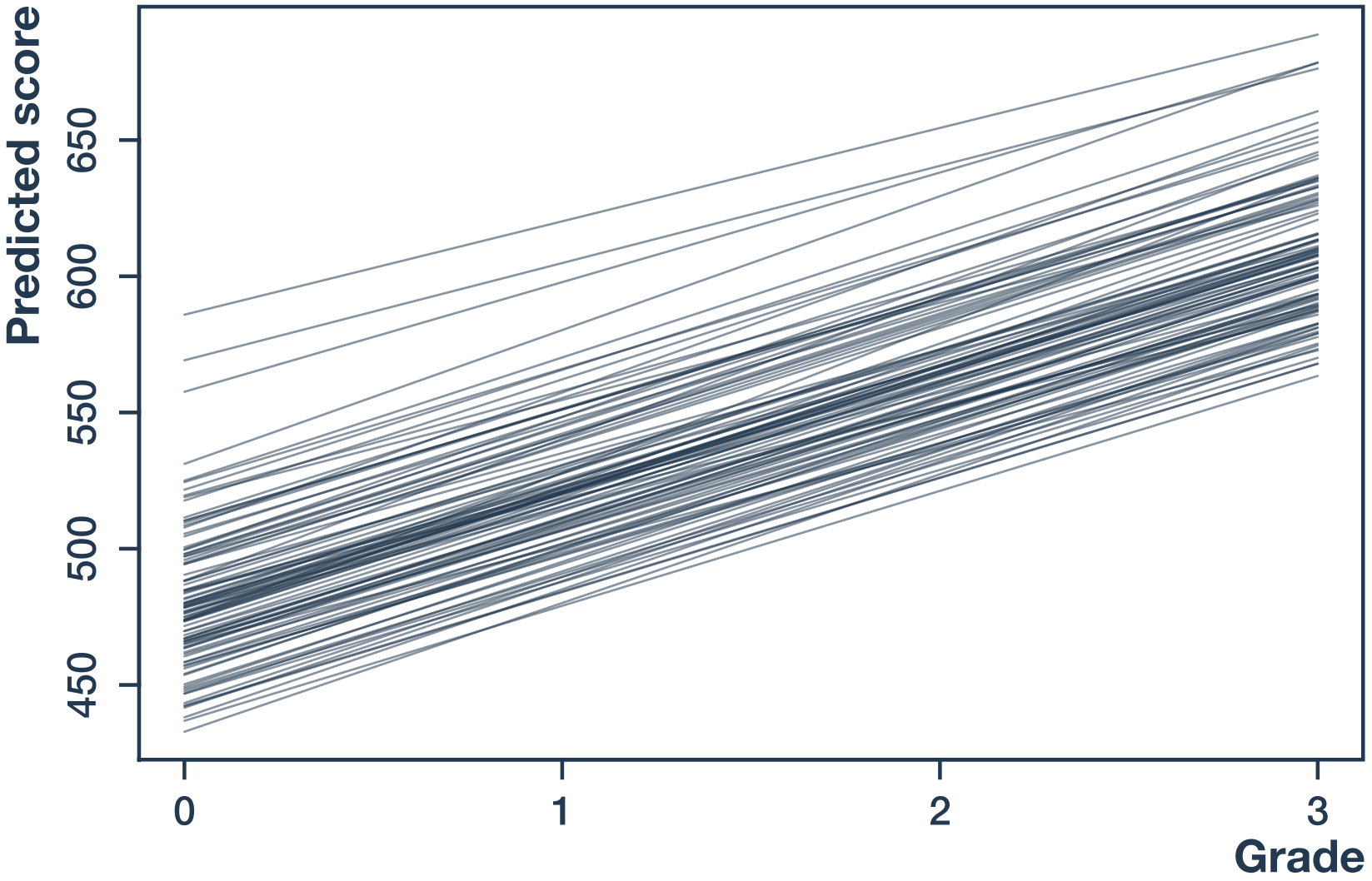
$$
[\eta_{0i}, \eta_{1i}] \sim \text{MVNorm}([0, 0], \Phi, R)
$$

Linear time trend

$$
\begin{aligned}
S_{ti} &\sim \text{Norm}(\mu_{ti}, \sigma) \\
\mu_{ti} &= \beta_{0i} + \beta_{1i} Year_{ti} + \beta_2 CSize_{ti} \\
\beta_{0i} &= \gamma_{00} + \eta_{0i} \\
\beta_{1i} &= \gamma_{10} + \eta_{1i} \\
\\
[\eta_{0i}, \eta_{1i}] &\sim \text{MVNorm}([0, 0], \Phi, R) \\
\\
\gamma_{00} &\sim \text{Norm}(500, 100) \\
\gamma_{10} &\sim \text{Norm}(0, 50) \\
\beta_2 &\sim \text{Norm}(0, 50) \\
\sigma &\sim \text{HalfCauchy}(0, 50) \\
\phi_0, \phi_1 &\sim \text{HalfCauchy}(0, 50) \\
R &\sim LKJ(2, 2)
\end{aligned}
$$

	Mean	*90% credible interval*	
γ_{00}	484.64	481.84	487.39
γ_{10}	42.88	41.75	44.02
β_2	–0.33	–0.73	0.070
σ	24.27	23.27	25.32
ϕ_0	38.25	35.84	40.65
ϕ_1	8.89	7.17	10.56
ρ_{01}	–0.37	–0.48	–0.23

Linear time trend



Quadratic time trend

$$
\begin{aligned}
S_{ti} &\sim \text{Norm}(\mu_{ti}, \sigma) \\
\mu_{ti} &= \beta_{0i} + \beta_{1i} Year_{ti} + \beta_{2i} Year_{ti}^2 + \beta_3 CSize_{ti} \\
\beta_{0i} &= \gamma_{00} + \eta_{0i} \\
\beta_{1i} &= \gamma_{10} + \eta_{1i} \\
\beta_{2i} &= \gamma_{20} + \eta_{2i}
\end{aligned}
$$

$$
[\eta_{0i}, \eta_{1i}, \eta_{2i}] \sim \text{MVNorm}([0, 0, 0], \Phi, R)
$$

Quadratic time trend

$$
\begin{aligned}
S_{ti} &\sim \text{Norm}(\mu_{ti}, \sigma) \\
\mu_{ti} &= \beta_{0i} + \beta_{1i} Year_{ti} + \beta_{2i} Year_{ti}^2 + \beta_3 CSize_{ti} \\
\beta_{0i} &= \gamma_{00} + \eta_{0i} \\
\beta_{1i} &= \gamma_{10} + \eta_{1i} \\
\beta_{2i} &= \gamma_{20} + \eta_{2i} \\
\\
[\eta_{0i}, \eta_{1i}, \eta_{2i}] &\sim \text{MVNorm}([0,0,0], \Phi, R) \\
\\
\gamma_{00} &\sim \text{Norm}(500, 100) \\
\gamma_{10} &\sim \text{Norm}(0, 50) \\
\gamma_{20} &\sim \text{Norm}(0, 50) \\
\beta_3 &\sim \text{Norm}(0, 50) \\
\sigma &\sim \text{HalfCauchy}(0, 50) \\
\phi_0, \phi_1, \phi_2 &\sim \text{HalfCauchy}(0, 50) \\
R &\sim LKJ(2, 3)
\end{aligned}
$$

	Mean	*90% credible interval*	
γ_{00}	482.32	479.37	485.30
γ_{10}	49.28	45.77	52.88
γ_{20}	–2.10	–3.14	–1.09
β_3	–0.36	–0.74	0.02
σ	21.91	20.76	23.04
ϕ_0	40.70	38.16	43.43
ϕ_1	31.48	26.35	36.38
ϕ_2	7.29	5.62	8.96
ρ_{01}	–0.45	–0.55	–0.35
ρ_{02}	0.40	0.25	0.53
ρ_{12}	–0.98	–1.00	–0.96

Quadratic time trend

