Multilevel Logistic Models

And MLM for Categorical Outcomes

October 24 2020 (updated: 6 November 2022)

Learning Objectives

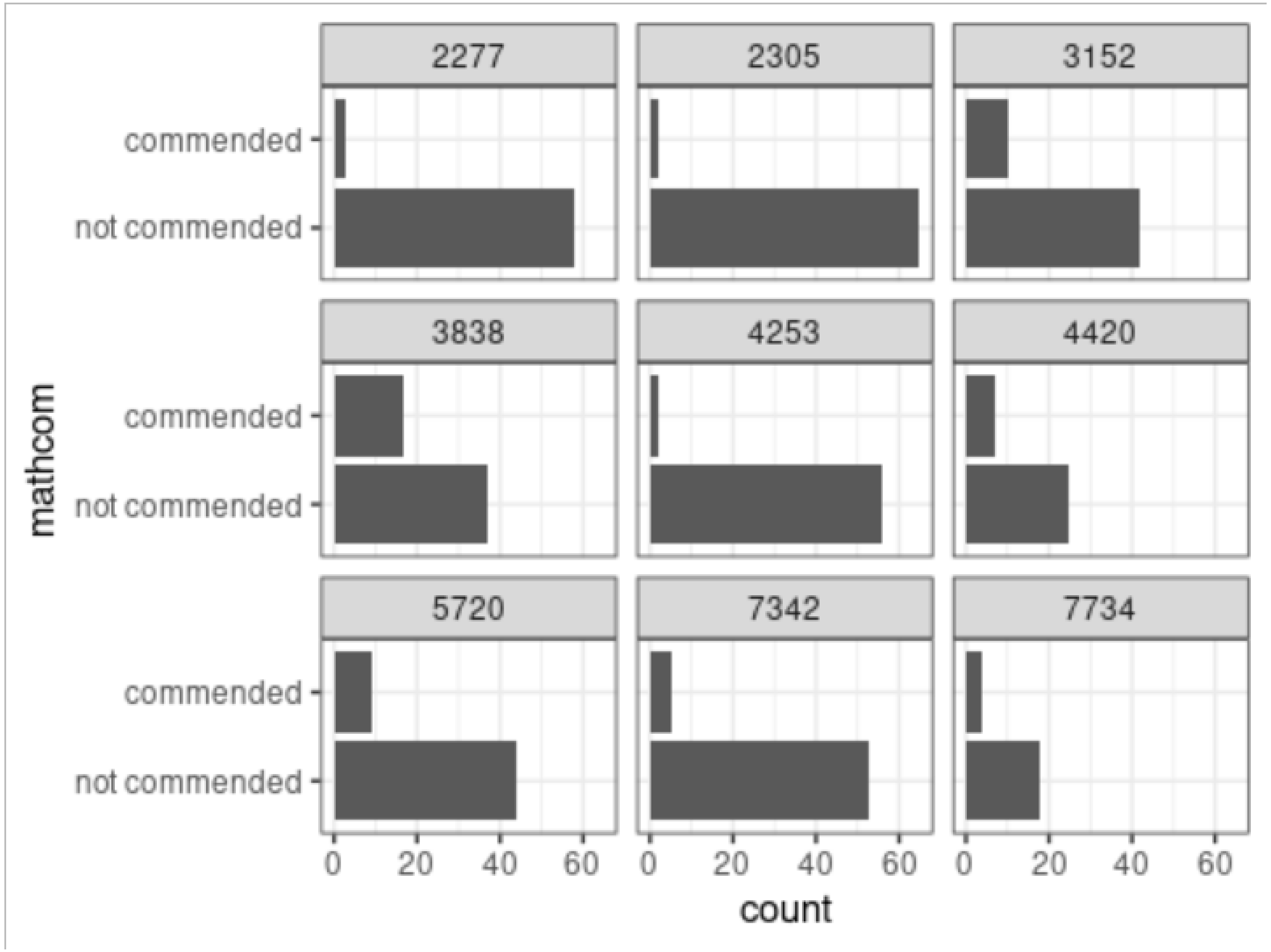
- Describe the problems of using a regular multilevel model for a binary outcome variable
- Write model equations for multilevel logistic regression
- Estimate intraclass correlations for binary outcomes
- Plot model predictions in probability unit

Binary Outcomes

- Pass/fail
- Agree/disagree
- Choosing stimulus A/B
- Diagnosis/no diagnosis

Example Data

- HSB data
- `mathcom`
  - 0 (not commended) if `mathach` < 20
  - 1 (commended) if `mathach` ≥ 20

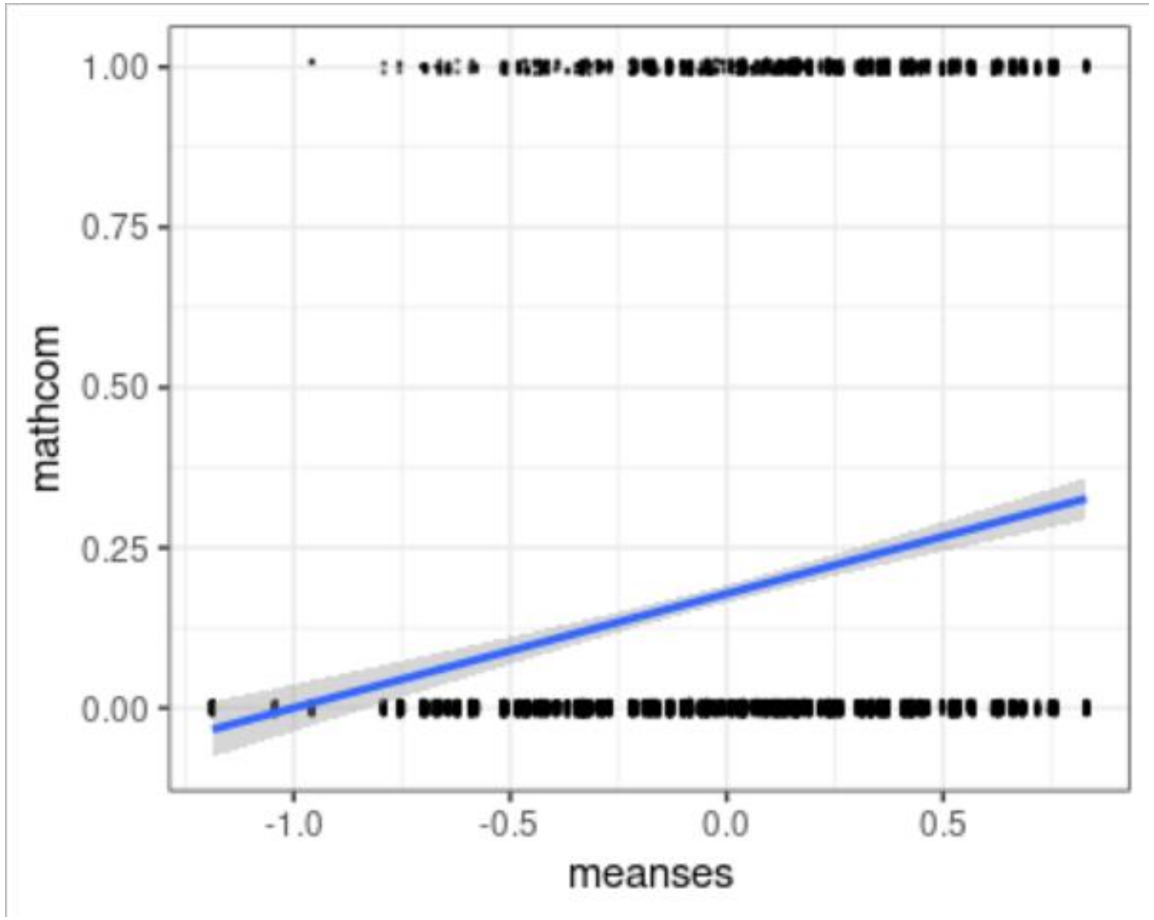


2277
2305
3152
3838
4253
4420
5720
7342
7734
commended
not commended
mathcom
0
20
40
60
count

Linear, Normal MLM

```
Random effects:

Conditional model:
 Groups   Name        Variance Std.Dev.
 id       (Intercept) 0.005148 0.07175
 Residual             0.136664 0.36968
Number of obs: 7185, groups:  id, 160

Dispersion estimate for gaussian family (sigma^2): 0.137

Conditional model:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) 0.178404   0.007222   24.70   <2e-16 ***
meanses     0.178190   0.017483   10.19   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Prediction Out of Range



Problems

- Out of range prediction
  - E.g., predicted value = -0.18 when meanses = -2
- Non-normality
  - The outcome can only take two values, and clearly not normal
- Nonconstant error variance/heteroscedasticity

Multilevel Logistic Model

For binary responses

Logistic Model

- A special case of the *Generalized Linear Mixed Model* (GLMM)
- Modify the linear, normal model in two ways:

1. Outcome distribution: ~~Normal~~ → Bernoulli
2. Predicted value
   - Mean of binary outcome (i.e., probability with range 0 to 1)

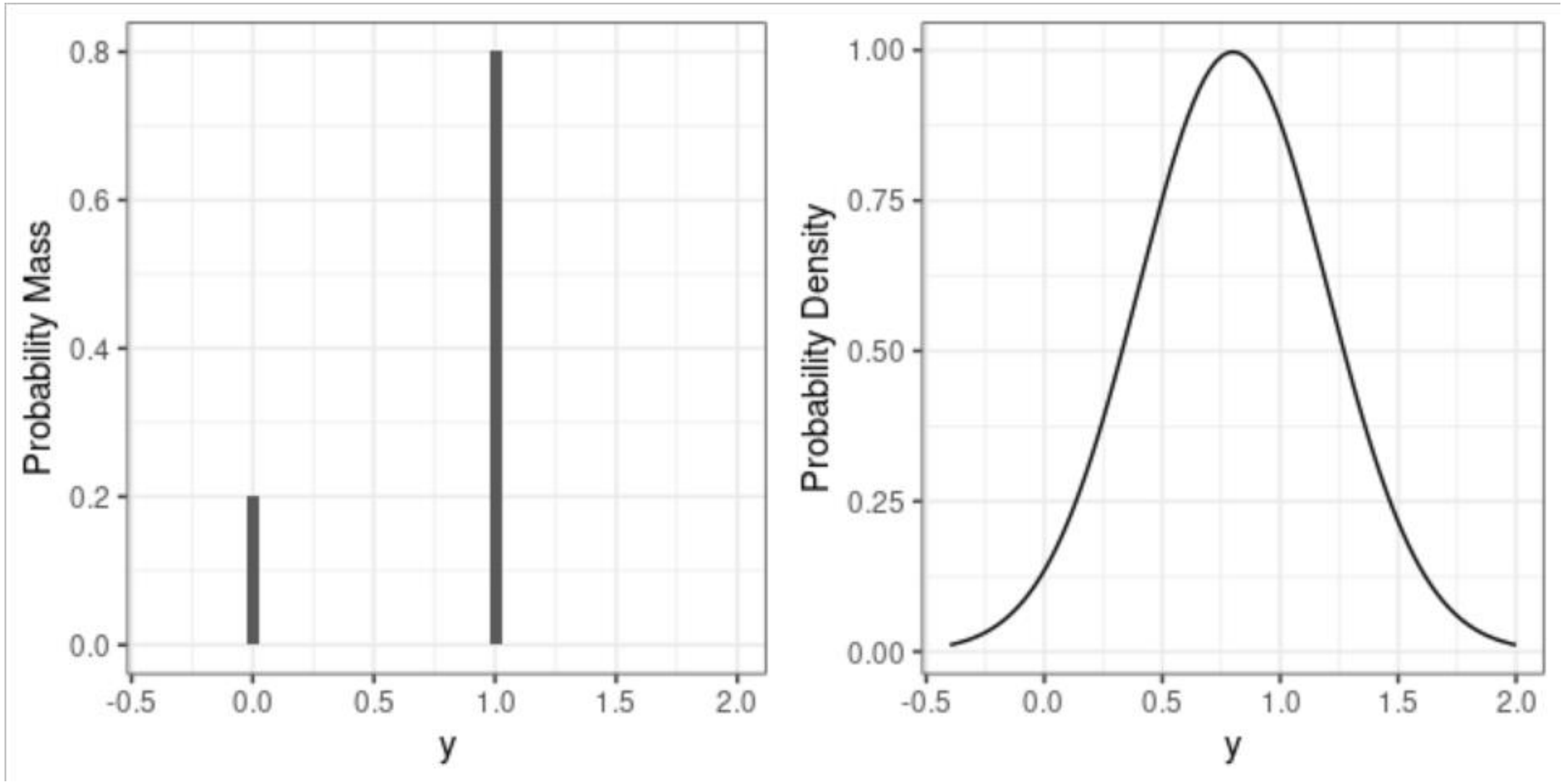
Logistic Model

- A special case of the *Generalized Linear Mixed Model* (GLMM)
- Modify the linear, normal model in two ways:

1. Outcome distribution: ~~Normal~~ → Bernoulli
2. Predicted value
   - ~~Mean of binary outcome (i.e., probability with range 0 to 1)~~
   - Transformed mean (i.e., log odds with range $-\infty$ to ∞)

Outcome Distribution

Bernoulli **Normal**



Transformation (Step 1): Odds

- Odds: Probability / (1 – Probability)
- Example:
  - 80% chance of being commended
  - = 4 to 1 odds in favor of being commended
  - Odds = 4 = 80% / (1 – 80%)
- Range of odds: 0 to ∞

Transformation (Step 2): Log-Odds

- Instead of predicting the probability, we predict the log odds
  - Solve the out of range problem

$$\text{Log Odds} = \log \frac{\text{Probability}}{1 - \text{Probability}}$$

- E.g., Probability = 0.8, odds = 4, log odds = 1.39
- E.g., Probability = 0.1, odds = 0.11, log odds = -2.20
- Range of log-odds: $-\infty$ to ∞

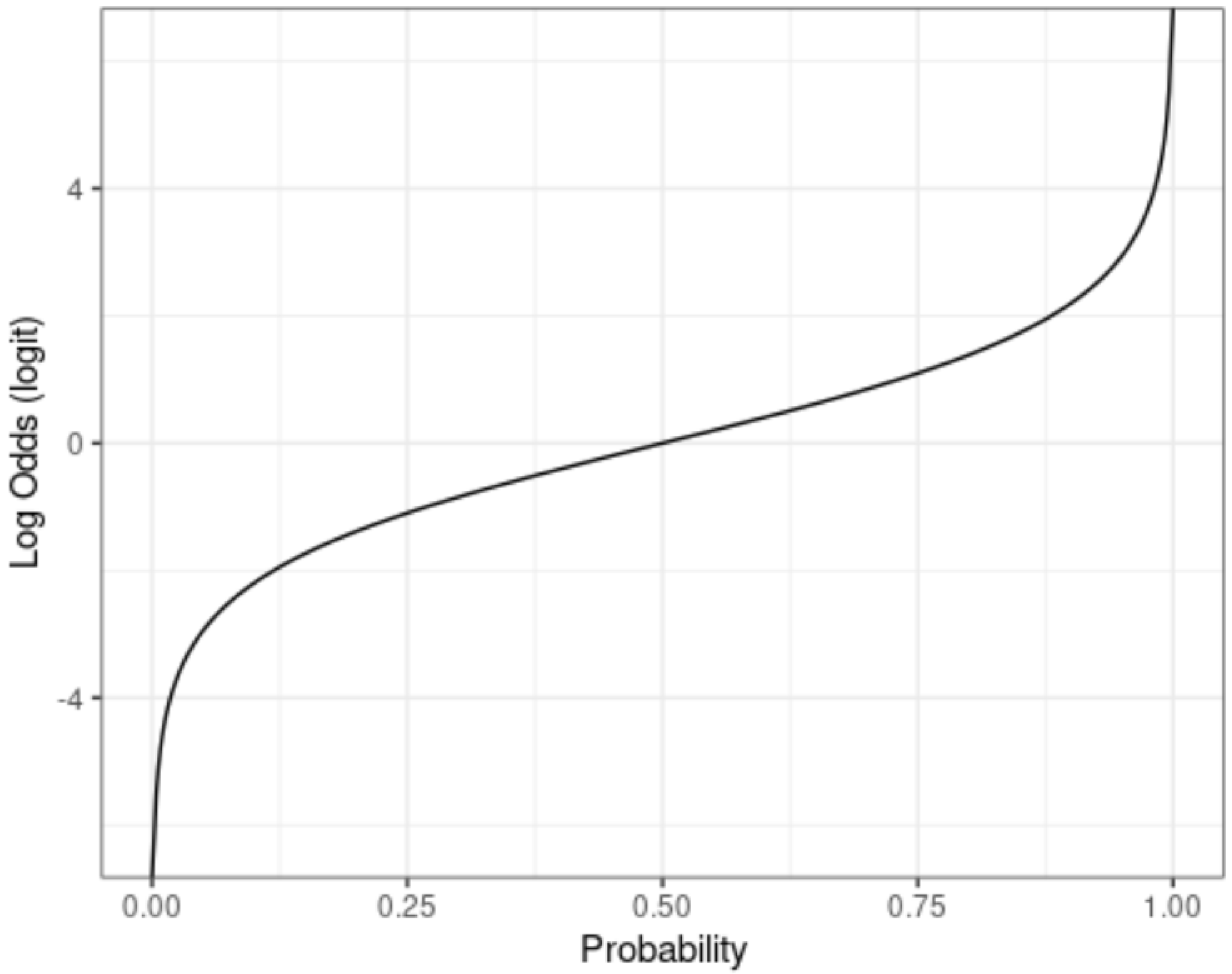
No longer needs to worry about out-of-range prediction

In logistic models, the coefficients are in the unit of log-odds

The transformation is called the [link function]

Interpretation less straight forward

Graph is preferred



Equations for Logistic MLM

Unconditional Model

Linear, Normal Model

- Lv 1: $\text{mathcom}_{ij} = \beta_{0j} + e_{ij}$

  $e_{ij} \sim N(0, \sigma)$

- Lv 2: $\beta_{0j} = \gamma_{00} + u_{0j}$

  $u_{0j} \sim N(0, \tau_0)$

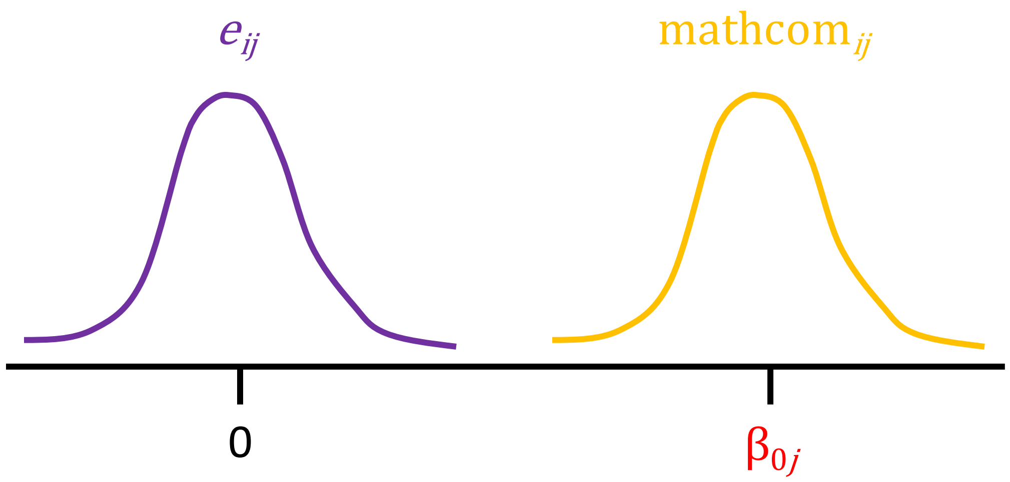
Another Way to Write the Model

- Lv 1: $mathcom_{ij} \sim N(\mu_{ij}, \sigma)$

  $\mu_{ij} = \beta_{0j}$

- Lv 2: $\beta_{0j} = \gamma_{00} + u_{0j}$

  $u_{0j} \sim N(0, \tau_0)$



Replace the Distribution

- Lv 1: $\text{mathcom}_{ij} \sim \text{Bernoulli}(\mu_{ij})$

  $$\mu_{ij} = \beta_{0j}$$

- Lv 2: $\beta_{0j} = \gamma_{00} + u_{0j}$

  $$u_{0j} \sim N(0, \tau_0)$$

Note: The Bernoulli distribution does not have a scale parameter

Transformation/Link Function

- Lv 1: $\text{mathcom}_{ij} \sim \text{Bernoulli}(\mu_{ij})$

  $\eta_{ij} = \text{logit}(\mu_{ij}) = \log[\mu_{ij} / (1 - \mu_{ij})]$

  $\eta_{ij} = \beta_{0j}$

- Lv 2: $\beta_{0j} = \gamma_{00} + u_{0j}$

  $u_{0j} \sim N(0, \tau_0)$

Transform probability to log-odds

Model log-odds
η_{ij} = linear predictor

Multilevel Logistic Model

- Lv 1: $\text{mathcom}_{ij} \sim \text{Bernoulli}(\mu_{ij})$

$$\eta_{ij} = \text{logit}(\mu_{ij}) = \log[\mu_{ij} / (1 - \mu_{ij})]$$
$$\eta_{ij} = \beta_{0j}$$

- Lv 2: $\beta_{0j} = \gamma_{00} + u_{0j}$

$$u_{0j} \sim N(0, \tau_0)$$

u_{0j} = School *j*'s deviation in log-odds

γ_{00} = log-odds for an average school

β_{0j} = Mean log-odds for school *j*

brms output

```
Group-Level Effects:
~id (Number of levels: 160)
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)     0.76      0.06     0.64     0.89 1.00     1380     2196

Population-Level Effects:
          Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept    -1.69      0.07    -1.84    -1.55 1.00     1714     2444
```

- For an average school, the estimated log-odds for being commended = -1.69, 95% CI [-1.84, -1.55]
- The estimated school-level standard deviation in log-odds for being commended = 0.76, 95% CI [0.64, 0.89]

Intraclass Correlation

There is no σ parameter

```
Group-Level Effects:
~id (Number of levels: 160)
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)     0.76      0.06     0.64     0.89 1.00     1380     2196
```

- In the unit of log odds, σ^2 is fixed to be π^2 / 3
  - $\pi = 3.14159265\ldots$
- Intraclass correlation:
- $\rho = \frac{\tau_0^2}{\tau_0^2 + \sigma^2} = \frac{0.76^2}{0.76^2 + \pi^2/3} = .15$

Interpretations of Coefficients

Conditional Model

Multilevel Logistic Model

- Lv 1: $\text{mathcom}_{ij} \sim \text{Bernoulli}(\mu_{ij})$

$$\eta_{ij} = \text{logit}(\mu_{ij}) = \log[\mu_{ij} / (1 - \mu_{ij})]$$

$$\eta_{ij} = \beta_{0j}$$

- Lv 2: $\beta_{0j} = \gamma_{00} + \gamma_{01}\,\text{meanses}_j + u_{0j}$

$$u_{0j} \sim N(0, \tau_0)$$

u_{0j} = School j's deviation in log-odds

β_{0j} = Mean log-odds for school j

γ_{00} = Predicted log-odds when meanses = 0 and u_{0j} = 0
γ_{01} = Predicted difference in log-odds associated with a unit change in meanses = 0

Adding a Level-1 Predictor

- Lv 1: $\text{mathcom}_{ij} \sim \text{Bernoulli}(\mu_{ij})$

$$\eta_{ij} = \text{logit}(\mu_{ij}) = \log[\mu_{ij} / (1 - \mu_{ij})]$$
$$\eta_{ij} = \beta_{0j} + \beta_{1j}\,\text{ses\_cmc}_{ij}$$

- Lv 2: $\beta_{0j} = \gamma_{00} + \gamma_{01}\,\text{meanses}_j + u_{0j}$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_0^2 & \tau_{01} \\ \tau_{01} & \tau_1^2 \end{bmatrix}\right)$$

Same thing: Cluster-mean centering, random slopes; just in log-odds

β_{1j}= Predicted difference in log-odds associated with a unit difference in student-level SES <u>within school *j*</u>

brms Output

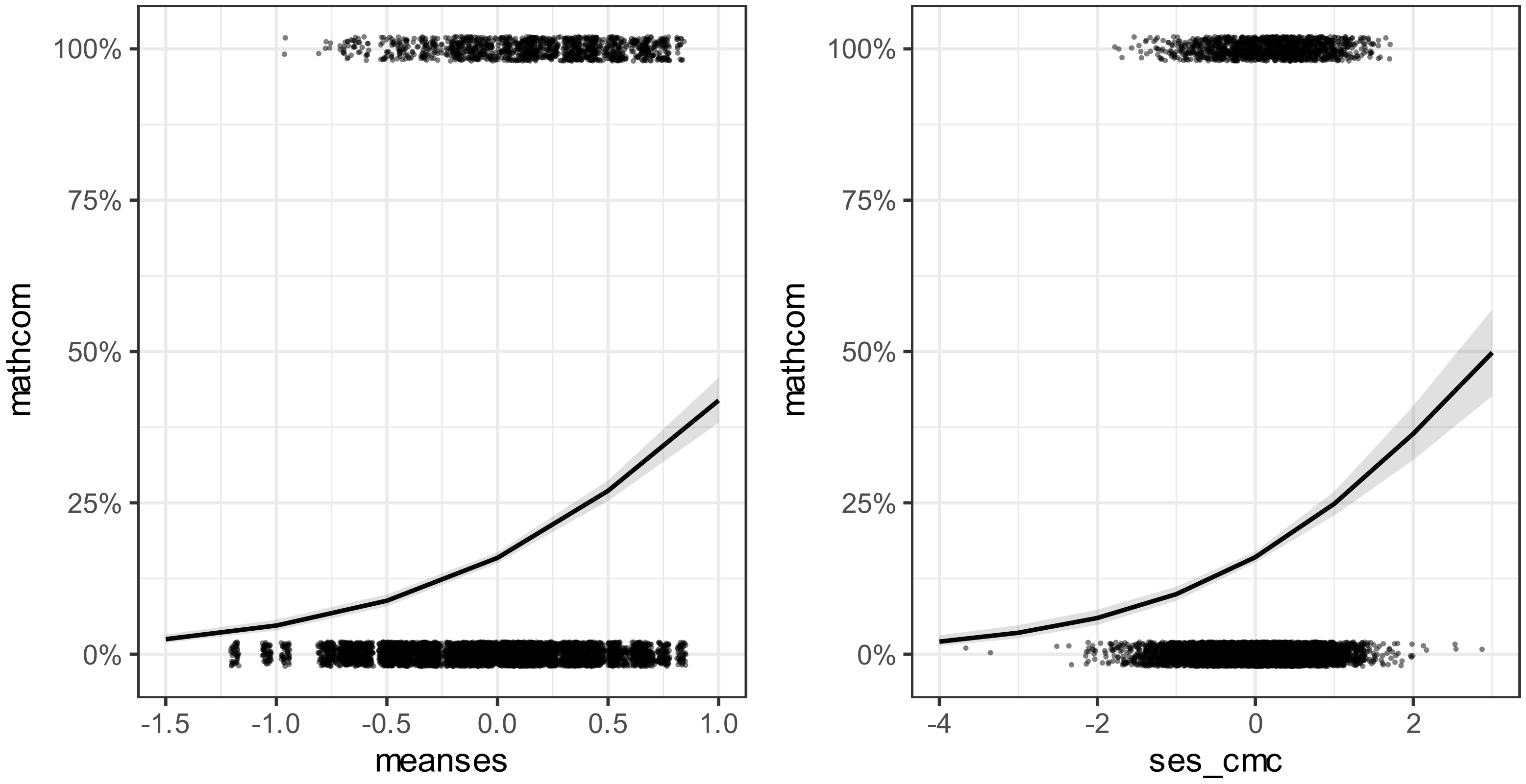
```
Group-Level Effects: 
~id (Number of levels: 160) 
                       Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)              0.52      0.05     0.42     0.64 1.00     1220     2226
sd(ses_cmc)                0.11      0.07     0.01     0.26 1.01     1164     1319
cor(Intercept,ses_cmc)    -0.48      0.44    -0.98     0.72 1.00     2234     2064

Population-Level Effects: 
          Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept    -1.76      0.06    -1.87    -1.65 1.00     1840     2199
meanses       1.45      0.14     1.19     1.73 1.00     1801     2381
ses_cmc       0.59      0.06     0.48     0.71 1.00     3791     2775
```

Cluster-/Unit-Specific vs. Population Average

- Coefficients in MLM requires a cluster-specific (CS) interpretation
  - Predicted difference in log-odds for two students in the same school (i.e., conditioned on u_{0j}), one with SES_cmc = 1 and the other with SES_cmc = 0 (so they have the same u_{0j})
- As opposed to population average (PA) coefficients (e.g., GEE)
  - Predicted difference in log-odds for an average student with SES_cmc = 1 and an average student with SES_cmc = 0
- Coefficients are usually smaller with PA than with CS

Interpretation is Hard

- Better approach: Plot the results in probability unit



Notes on Interpretation

- Predicted difference in probability is not constant across different levels of the predictor
- It's useful to get the predicted probabilities for representative values in the data

```
>#   meanses ses_cmc  .fitted
># 1       0    -0.5    0.126
># 2       0     0.5    0.199

>#   meanses ses_cmc  .fitted
># 1    -0.5       0    0.088
># 2     0.5       0    0.270
```

Notes on Interpretation

- Another common practice is to convert the coefficients to odds ratio
  - OR = exp(γ) for average slope
  - OR = exp(β_{1j}) for cluster-specific slope
- It's still hard to understand what a ratio of two odds would mean

Generalized Linear Mixed-Effect Model (GLMM)

For other discrete outcomes

Intrinsically Non-Normal Outcomes

- Counts
  - E.g., # of correct answers, # children, # symptoms, incidence rates
- Rating scales (Ordinal)
  - E.g., Likert scale, ranking
- Nominal
  - E.g., voting in a 3-party election

Generalized Linear Model

- McCullagh & Nelder (1989)
- Generalized linear: linear after some transformation
  - E.g., $\text{logit}(\mu) = b_0 + b_1 X_1 + b_2 X_2$

Generalized Linear Model (cont'd)

- Three elements:
  - Error/conditional distribution of Y (with mean μ and an optional dispersion parameter)
    - E.g., Bernoulli
  - Linear predictor (η)
    - The predicted value (e.g., log odds)
  - Link function ($\eta = g[\mu]$)
    - The transformation

Other Common Types of GLM/GLMM

- Binomial logistic
- Poisson
- Ordinal (not GLMM but highly related)

Binomial Logistic

- For counts (with known number of trials)
  - E.g., number of female hires out of *n* new hires
  - E.g., number of symptoms on a checklist of *n* items
- Multiple Bernoulli trials
- Conditional distribution: Binomial(*n*, μ)
- Link: logit
- Linear predictor: log odds
- R code: `family = binomial("logit")`

Poisson

- For counts (with infinite/vague number of trials)
  - E.g., number of binge drinking episodes
  - E.g., number of spam emails
- Conditional distribution: Poisson(μ)
- Link: log
- Linear predictor: log rate of occurrence
- R code: `family = Poisson("log")`

Ordinal

- For ordinal outcome with less than 5 categories/skewed distribution
  - E.g., Happiness (1-4)
- Conditional distribution: Categorical
- Link: logit
- Linear predictor: log odds of endorsing $k + 1$ or above vs. k or below
  - E.g., choosing 3 or 4 vs. 2 or 1 on the happiness scale
- Check out the R function `ordinal::clmm()`