Understanding Variability in Multilevel Models for Categorical Responses

Sophia Rabe-Hesketh

University of California, Berkeley

& Institute of Education, University of London

Joint work with Anders Skrondal

Hierarchical Linear Modeling SIG
AERA Annual Meeting, Vancouver, April 2012

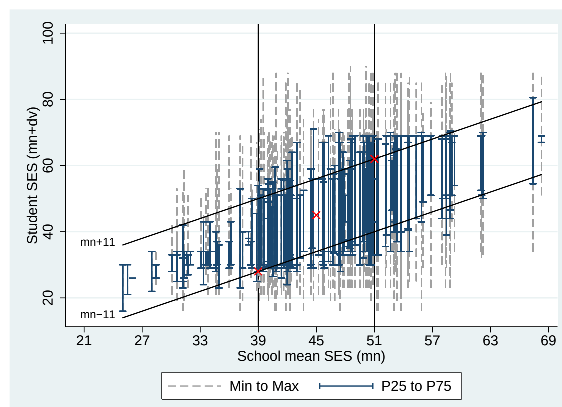
Outline

- Two-level logistic random-intercept model (1: students, 2: schools)
- Measures of residual variability and dependence
  - Median odds ratio
  - Intraclass correlation of latent responses
  - Intraclass correlation of observed responses
  - Standard deviation of probabilities
  - Variance partition coefficient
- Graphical displays of variability
  - Densities of probabilities
  - Percentiles of probabilities
  - Probabilities for schools in the data
- Ordinal responses and three-level data (1: students, 2: teachers, 3: schools)

Example: PISA (Programme for International Student Assessment)

- PISA study assesses reading, math and science achievement among 15-year-old students in tens of countries every 3 years
  - Here consider U.S. data from 2000
- Schools j were randomly sampled and then students i were randomly sampled from the selected schools
- Variables:
  - Reading proficiency y_{ij} (1=yes, 0=no)
  - Student SES x_{ij}
    - ◇ `mn`: School mean SES $\overline{x}_{\cdot j}$
    - ◇ `dv`: Deviation of student SES from school mean $x_{ij} - \overline{x}_{\cdot j}$
- Sample size (listwise)
  - 2069 students in 148 schools
  - Number of students per school: 1 to 28, mean/median 14

PISA: Distribution of SES



- For schools, `mn`:
  - 25th & 75th %tiles: 39 & 51
  - Mean & Median: 45
- For students, `dv`:
  - 25th & 75th %tiles: -11 & 11
  - Mean & Median: 0
- Three covariate patterns:

$\mathbf{x}_{ij}$	`mn`	`dv`	`ses`
lo	39	-11	28
me	45	0	45
hi	51	11	62

Two-level logistic random-intercept model

- Two-level logistic random-intercept model for unit i (level 1) nested in cluster j (level 2)

$$\text{logit}[\Pr(y_{ij}=1|\mathbf{x}_{ij},\zeta_j)] = \log \underbrace{\left[\frac{\Pr(y_{ij}=1|\mathbf{x}_{ij},\zeta_j)}{\Pr(y_{ij}=0|\mathbf{x}_{ij},\zeta_j)}\right]}_{\text{Odds}}$$
$$= \beta_1 + \beta_2 \mathtt{dv}_{ij} + \beta_3 \mathtt{mn}_j + \zeta_j$$

  - $\mathbf{x}_{ij}$ is vector of covariates $(\mathtt{dv}_{ij}, \mathtt{mn}_j)'$
  - β_1 is the fixed intercept
  - β_2 and β_3 are fixed regression coefficients
  - ζ_j is a random intercept $\zeta_j \sim N(0,\psi)$
- Sometimes write $\mathbf{x}_{ij}'\boldsymbol{\beta} \equiv \beta_1 + \beta_2 \mathtt{dv}_{ij} + \beta_3 \mathtt{mn}_j$

Maximum likelihood estimates of model parameters

		Null model	Full model	
Param.	Covariate	Est (SE)	Est (SE)	OR (95% CI)
β_1		−0.71 (0.10)	−4.79 (0.43)	
$10\beta_2$	`dv/10`		0.18 (0.03)	1.2 (1.1,1.3)
$10\beta_3$	`mn/10`		0.89 (0.09)	2.4 (2.1,2.9)
ψ		0.82	0.28	

$$\log[\text{Odds}(y_{ij}=1|\mathbf{x}_{ij},\zeta_j)] = \beta_1 + \beta_2 \mathtt{dv}_{ij} + \beta_3 \mathtt{mn}_j + \zeta_j$$
$$\text{Odds}(y_{ij}=1|\mathbf{x}_{ij},\zeta_j) = \exp(\beta_1)\exp(\beta_2)^{\mathtt{dv}_{ij}}\exp(\beta_3)^{\mathtt{mn}_j}\exp(\zeta_j)$$

- Conditional odds ratio associated with 10-point increase in $\mathtt{dv}_{ij}$ for given $\mathtt{mn}_j$ and ζ_j:

$$\frac{\exp(\beta_1)\exp(\beta_2)^{a+10}\exp(\beta_3)^{\mathtt{mn}_j}\exp(\zeta_j)}{\exp(\beta_1)\exp(\beta_2)^{a}\exp(\beta_3)^{\mathtt{mn}_j}\exp(\zeta_j)} = \exp(\beta_2)^{10} = \exp(10\beta_2)$$

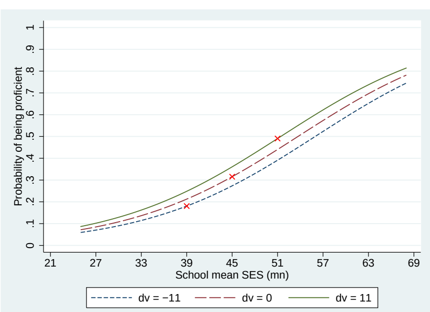
  - Comparing two students from same school (given $\mathtt{mn}_j$ and ζ_j) whose SES differs by 10 points, student with higher SES has 1.2 times the odds of being proficient as other student – odds are 20% greater
- Conditional odds ratio associated with 10-point increase in $\mathtt{mn}_j$ for given $\mathtt{dv}_{ij}$ and ζ_j:
  - Comparing two schools that differ in their mean SES by 10 points and have the same random intercept, odds of a student (with SES at the school mean) being proficient are 2.4 times as great for higher-SES school

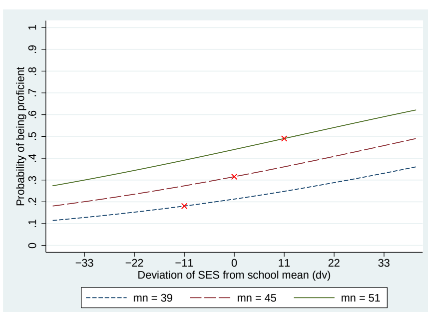
Predicted conditional probabilities $\widehat{p}(\mathbf{x}_{ij},\zeta_j)$

- Conditional probability with estimates $\widehat{\beta}_1$, $\widehat{\beta}_2$, $\widehat{\beta}_3$ plugged in

$$\widehat{p}(\mathbf{x}_{ij},\zeta_j) \equiv \widehat{\Pr}(y_{ij}=1|\mathbf{x}_{ij},\zeta_j) = \frac{\exp(\widehat{\beta}_1 + \widehat{\beta}_2 \mathtt{dv}_{ij} + \widehat{\beta}_3 \mathtt{mn}_j + \zeta_j)}{1+\exp(\widehat{\beta}_1 + \widehat{\beta}_2 \mathtt{dv}_{ij} + \widehat{\beta}_3 \mathtt{mn}_j + \zeta_j)}$$

- Plug in interesting values of $\mathbf{x}_{ij}$ and ζ_j, e.g., $\zeta_j = 0$ (mean & median)
  - $\widehat{p}(\text{lo},0) = 0.18$, $\widehat{p}(\text{me},0) = 0.32$, $\widehat{p}(\text{hi},0) = 0.49$





Median odds ratio

- Model for conditional odds

$$\text{Odds}(y_{ij}=1|\mathbf{x}_{ij},\zeta_j) = \exp(\beta_1)\exp(\beta_2)^{\mathtt{dv}_{ij}}\exp(\beta_3)^{\mathtt{mn}_j}\exp(\zeta_j)$$

- Randomly sample schools, then students (for given $\mathbf{x}$)
- Odds ratio, for two students i and i' from different schools j and j':

$$\frac{\text{Odds}(y_{ij}=1|\mathbf{x},\zeta_j)}{\text{Odds}(y_{i'j'}=1|\mathbf{x},\zeta_{j'})} = \exp(\zeta_j - \zeta_{j'})$$

- Odds ratio, comparing student with greater odds to other student:

$$\text{OR} = \exp(|\zeta_j - \zeta_{j'}|), \quad (\zeta_j - \zeta_{j'}) \sim N(0, 2\psi)$$

- Odds ratio $\exp(|\zeta_j - \zeta_{j'}|)$ varies randomly and has estimated median

$$\text{OR}_{\text{median}} = \exp\{\sqrt{2\widehat{\psi}}\,\Phi^{-1}(3/4)\}$$

[Larsen et al., 2000]

Median odds ratio (cont'd)

- For two randomly drawn students from two randomly drawn schools, estimated median odds ratio, comparing student from school with larger random intercept to other student, is

$$\begin{aligned} \mathsf{OR}_{\mathsf{median}} &= 2.4 \text{ for null model} \\ \mathsf{OR}_{\mathsf{median}} &= 1.7 \text{ for full model} \end{aligned}$$

- In null model, odds ratio due to random intercept exceeds 2.4 half the time
  - In full model, estimated odds ratio for 10-point increase in school-mean SES is 2.4

Latent response formulation

- Imagine latent (unobserved) continuous response y^*_{ij} (e.g., reading ability)
- Observed response is 1 if latent response is greater than 0 (e.g., proficient if ability greater than 0):

$$y_{ij} = \begin{cases} 1 & \text{if } y^*_{ij} > 0 \\ 0 & \text{otherwise} \end{cases}$$

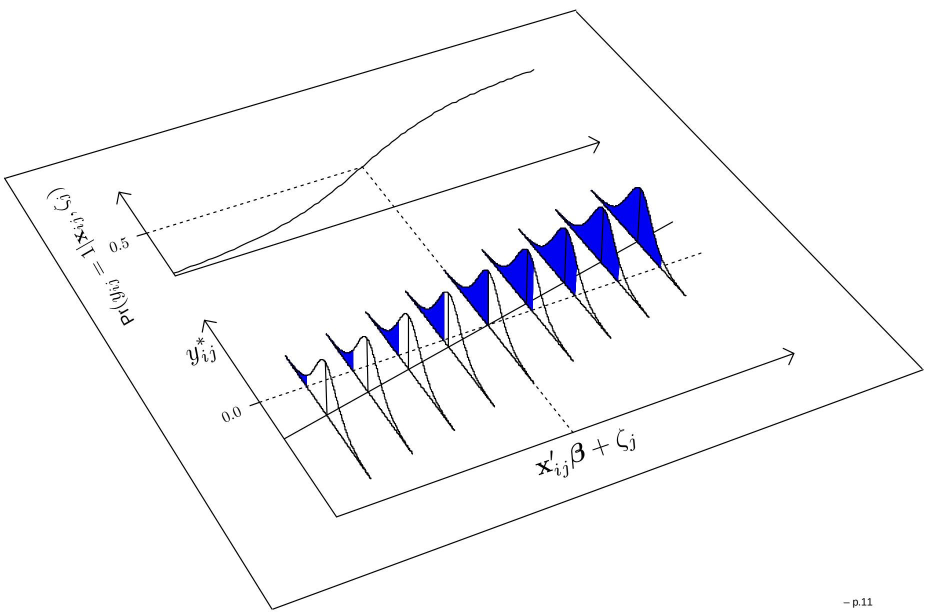
- Model for latent response:

$$y^*_{ij} \;=\; \underbrace{\beta_1 + \beta_2 \mathtt{dv}_{ij} + \beta_3 \mathtt{mn}_j}_{\mathbf{x}'_{ij}\boldsymbol{\beta}} + \zeta_j + \epsilon_{ij}, \quad \zeta_j \sim N(0, \psi),\; \epsilon_{ij} \sim \text{logistic}$$

  - Logistic distribution has mean 0 and variance $\pi^2/3$ and is similar to normal distribution
- Model for observed response:

$$\text{logit}[\Pr(y_{ij} = 1|\mathbf{x}_{ij}, \zeta_j)] = \beta_1 + \beta_2 \mathtt{dv}_{ij} + \beta_3 \mathtt{mn}_j + \zeta_j, \quad \zeta_j \sim N(0, \psi)$$

Equivalence of formulations



Intraclass correlation of latent responses

- Model for y^*_{ij} is standard multilevel/hierarchical linear model, therefore use standard ICC

$$\mathsf{ICC}_{\mathsf{lat}} = \frac{\mathsf{Var}(\zeta_j)}{\mathsf{Var}(\zeta_j) + \mathsf{Var}(\epsilon_{ij})} = \frac{\psi}{\psi + \pi^2/3}$$

  - Correlation between students i and i' in same school j

$$\mathsf{Cor}(y^*_{ij}, y^*_{i'j}|\mathbf{x}_{ij}, \mathbf{x}_{i'j}) = \mathsf{Cor}(\zeta_j + \epsilon_{ij}, \zeta_j + \epsilon_{i'j}) = \frac{\psi}{\psi + \pi^2/3}$$

  - Proportion of variance that is between schools

$$\mathsf{Var}(y^*_{ij}|\mathbf{x}_{ij}) = \mathsf{Var}(\zeta_j + \epsilon_{ij}) = \mathsf{Var}(\zeta_j) + \mathsf{Var}(\epsilon_{ij}) = \psi + \pi^2/3$$

- For PISA data, plug in $\widehat{\psi}$

$$\begin{aligned} \mathsf{ICC}_{\mathsf{lat}} &= \frac{0.82}{0.82 + 3.29} = 0.20 \text{ for null model} \\ \mathsf{ICC}_{\mathsf{lat}} &= \frac{0.28}{0.28 + 3.29} = 0.08 \text{ for full model} \end{aligned}$$

Intraclass correlation of observed responses

- Correlation $\text{Cor}(y_{ij}, y_{i'j}|\mathbf{x}_{ij}, \mathbf{x}_{i'j})$ of observed responses for students i and i' in same school j depends on covariates $\mathbf{x}_{ij}$, $\mathbf{x}_{i'j}$
- For simplicity, assume $\mathbf{x}_{ij} = \mathbf{x}_{i'j} = \mathbf{x}$

$$\text{ICC}_{\text{obs}} = \widehat{\text{Cor}}(y_{ij}, y_{i'j}|\mathbf{x}) = \frac{\overline{p}_{11}(\mathbf{x}) - \overline{p}(\mathbf{x})^2}{\overline{p}(\mathbf{x})[1-\overline{p}(\mathbf{x})]}$$

  - Among schools and students with covariates $\mathbf{x}$, randomly choose a school and then randomly choose students from the school
    - $\overline{p}(\mathbf{x})$ is probability that a student is proficient
    - $\overline{p}_{11}(\mathbf{x})$ is probability that two students are both proficient
    - Population averaged or marginal probability

$$\overline{p}(\mathbf{x}) = \widehat{\text{Pr}}(y_{ij}=1|\mathbf{x}) = \int \widehat{p}(\mathbf{x}, \zeta_j) g(\zeta_j; 0, \widehat{\psi})\, d\zeta_j$$

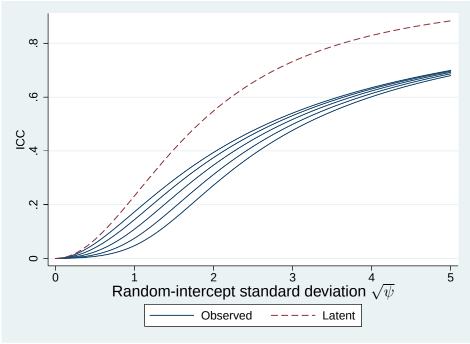
Average over random effects with Gaussian density $g(\zeta_j; 0, \widehat{\psi})$

[Rodríguez & Elo, 2003]

Intraclass correlation of observed responses (cont'd)

- PISA:

	Null model	Full model		
		lo	me	hi
ICC_{obs}	0.14	0.04	0.06	0.06
ICC_{lat}	0.20	0.08	0.08	0.08



- $\text{logit}[\text{Pr}(y_{ij}=1|\zeta_j)] = \beta_1 + \zeta_j$
- $\zeta_j \sim N(0, \psi)$
- ICC_{obs} for $\beta_1 = 0$ to $\beta_1 = 3.7$
- $\text{ICC}_{\text{obs}} < \text{ICC}_{\text{lat}}$
- ICC_{obs} largest for $\beta_1 = 0$

[Rodríguez & Elo, 2003]

Kappa coefficient of observed responses

- Recall

$$\text{ICC}_{\text{obs}} = \widehat{\text{Cor}}(y_{ij}, y_{i'j}|\mathbf{x}) = \frac{\overline{p}_{11}(\mathbf{x}) - \overline{p}(\mathbf{x})^2}{\overline{p}(\mathbf{x})[1-\overline{p}(\mathbf{x})]}$$

- Probabilities of agreement for students in same school:

$$\text{Pr}(y_{ij} = y_{i'j} = 1|\mathbf{x}) \equiv \overline{p}_{11}(\mathbf{x}) = \text{ICC}_{\text{obs}}\, \overline{p}(\mathbf{x})[1-\overline{p}(\mathbf{x})] + \overline{p}(\mathbf{x})^2$$
$$\text{Pr}(y_{ij} = y_{i'j} = 0|\mathbf{x}) \equiv \overline{p}_{00}(\mathbf{x}) = \text{ICC}_{\text{obs}}\, \overline{p}(\mathbf{x})[1-\overline{p}(\mathbf{x})] + [1-\overline{p}(\mathbf{x})]^2$$

- Probabilities of agreement for students in different schools:

$$\text{Pr}(y_{ij} = y_{i'j'} = 1|\mathbf{x}) = \overline{p}(\mathbf{x})^2$$
$$\text{Pr}(y_{ij} = y_{i'j'} = 0|\mathbf{x}) = [1-\overline{p}(\mathbf{x})]^2 = 1 - 2\overline{p}(\mathbf{x}) + \overline{p}(\mathbf{x})^2$$
$$\text{Pr}(y_{ij} = y_{i'j'}|\mathbf{x}) = 1 - 2\overline{p}(\mathbf{x})[1-\overline{p}(\mathbf{x})]$$

$$\kappa = \frac{\text{Pr}(y_{ij} = y_{i'j}|\mathbf{x}) - \text{Pr}(y_{ij} = y_{i'j'}|\mathbf{x})}{1 - \text{Pr}(y_{ij} = y_{i'j'}|\mathbf{x})} = \frac{\text{ICC}_{\text{obs}}\, 2\overline{p}(\mathbf{x})[1-\overline{p}(\mathbf{x})]}{2\overline{p}(\mathbf{x})[1-\overline{p}(\mathbf{x})]} = \text{ICC}_{\text{obs}}$$

[Eldridge et al., 2009]

Standard deviation of probabilities

- Already considered median of $\widehat{p}(\mathbf{x}, \zeta_j)$

$$\text{Median}[\widehat{p}(\mathbf{x}, \zeta_j)] = \widehat{p}(\mathbf{x}, 0)$$

- Already considered mean of $\widehat{p}(\mathbf{x}, \zeta_j)$

$$\overline{p}(\mathbf{x}) = \widehat{\text{Pr}}(y_{ij}=1|\mathbf{x}) = \int \widehat{p}(\mathbf{x}, \zeta_j) g(\zeta_j; 0, \widehat{\psi})\, d\zeta_j$$

- Get standard deviation of $\widehat{p}(\mathbf{x}, \zeta_j)$ by integration or by simulation

$$\text{sd}[\widehat{p}(\mathbf{x}, \zeta_j)] = \sqrt{\text{Var}[\widehat{p}(\mathbf{x}, \zeta_j)]}$$

Null model	Full model		
	lo	me	hi
0.18	0.08	0.11	0.12

Variance partition coefficient

- General result

$$\underbrace{\text{Var}(\text{responses})}_{\text{Total variance}} = \underbrace{\text{Var}(\text{School-mean response})}_{\text{Between-school variance } v_2} + \underbrace{\text{Mean}(\text{Within-school variance})}_{\text{Within-school variance } v_1}$$

- For school j:
  - School-mean response: $\widehat{\text{Mean}}(y_{ij}|\mathbf{x},\zeta_j) = \widehat{p}(\mathbf{x},\zeta_j)$
    $\Rightarrow$ Find variance over distribution of ζ_j
  - Within-school variance: $\widehat{\text{Var}}(y_{ij}|\mathbf{x},\zeta_j) = \widehat{p}(\mathbf{x},\zeta_j)[1-\widehat{p}(\mathbf{x},\zeta_j)]$
    $\Rightarrow$ Find mean over distribution of ζ_j
- Total variance:

$$\widehat{\text{Var}}(y_{ij}|\mathbf{x}_{ij}) = v_2 + v_1 = \overline{p}(\mathbf{x}_{ij})[1-\overline{p}(\mathbf{x}_{ij})]$$

- Variance partition coefficient [Goldstein et al., 2002, Simulation, Method B]

$$\text{VPC} = \frac{v_2}{v_2+v_1} = \text{ICC}_{\text{obs}} \neq \text{ICC}_{\text{lat}}$$

All results for PISA

$\mathbf{x}$	mn	dv	$\widehat{p}(\mathbf{x},0)$	$\overline{p}(\mathbf{x})$	sd$[\widehat{p}(\mathbf{x},\zeta_j)]$	ICC_{lat}	ICC_{obs}	VPC
Null model								
–	–	–	0.331	0.354	0.180	0.200	0.143	0.143
Full model								
lo	39	-11	0.181	0.193	0.081	0.078	0.042	0.042
me	45	0	0.315	0.333	0.111	0.078	0.056	0.056
hi	51	11	0.491	0.491	0.124	0.078	0.062	0.062

Density of probabilities

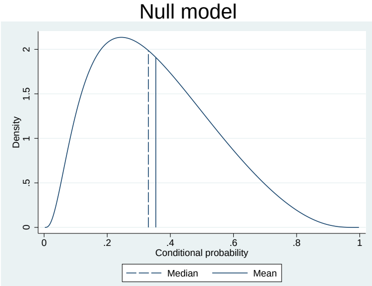
- Conditional probabilities, conditional on ζ_j, depend on ζ_j

$$\widehat{p}(\mathbf{x},\zeta_j) \equiv \Pr(y_{ij}=1|\mathbf{x}_{ij}=\mathbf{x},\zeta_j)$$

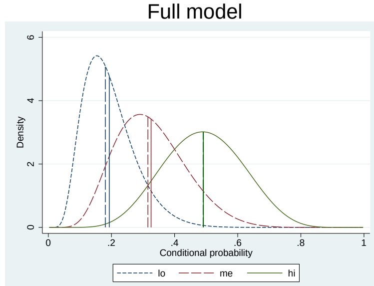
- When ζ_j varies, corresponding probability $\widehat{p} \equiv \widehat{p}(\mathbf{x},\zeta_j)$ varies with density

$$f(\widehat{p}) = g(\log[\widehat{p}/(1-\widehat{p})]; \mathbf{x}'\widehat{\boldsymbol{\beta}}, \psi) \times \frac{1}{\widehat{p}(1-\widehat{p})}$$

Null model



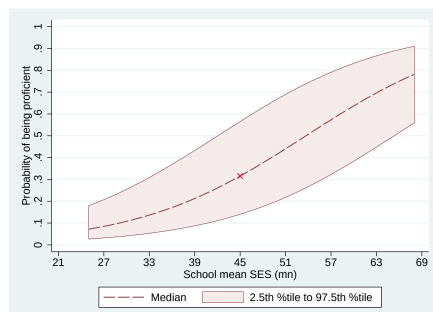
Full model

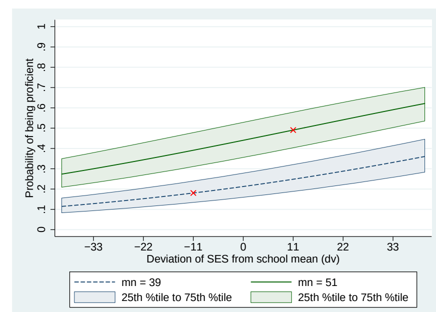


[Duchateau & Janssen, 2005]

Percentiles of probabilities

- A given percentile of $\widehat{p}(\mathbf{x},\zeta_j)$, for given $\mathbf{x}$, can be found by plugging in corresponding percentile of $\zeta_j \sim N(0,\widehat{\psi})$
- **Left figure:** $\widehat{p}(\mathbf{x},\pm 1.96\sqrt{\widehat{\psi}})$ and $\widehat{p}(\mathbf{x},0)$ versus mn with dv$=0$
  - Randomly sample schools with a given mn:
    Interval is 95% range of probabilities for students with dv$=0$





Probabilities for schools in the data

- Probability $\widetilde{p}(\mathbf{x})$ of proficiency for a randomly sampled student with covariates $\mathbf{x}$ in an existing school j
- Data on students in school j:

$$\mathbf{y}_j = (y_{1j}, \ldots, y_{n_j j})' \text{ and } \mathbf{X}_j = \begin{bmatrix} \mathbf{x}'_{1j} \\ \vdots \\ \mathbf{x}'_{n_j j} \end{bmatrix}$$

- Information about ζ_j (Bayes theorem):

$$\text{Posterior}(\zeta_j|\mathbf{y}_j, \mathbf{X}_j) = \frac{g(\zeta_j; 0, \psi)\text{Pr}(\mathbf{y}_j|\mathbf{X}_j, \zeta_j)}{\text{Pr}(\mathbf{y}_j|\mathbf{X}_j)}$$

- Best prediction of probability (in terms of mean-squared error) is

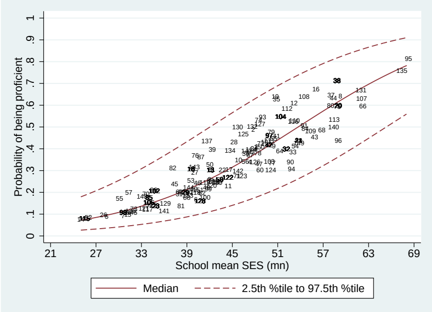
$$\begin{aligned} \widetilde{p}(\mathbf{x}) &= \text{Posterior mean}[\widehat{p}(\mathbf{x}, \zeta_j)] \\ &= \int \widehat{p}(\mathbf{x}, \zeta_j)\text{Posterior}(\zeta_j|\mathbf{y}_j, \mathbf{X}_j)\, d\zeta_j \end{aligned}$$

  - Not equal to $\widehat{p}(\mathbf{x}, \widetilde{\zeta}_j)$, where $\widetilde{\zeta}_j$ is posterior mean of ζ_j

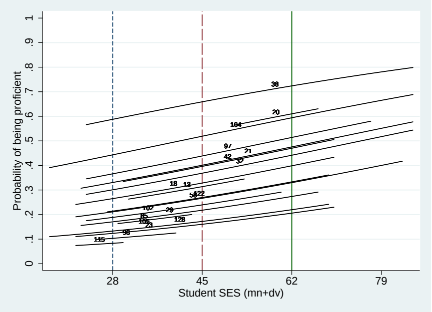
[Skrondal & Rabe-Hesketh, 2009]

Probabilities for schools in the data (cont'd)

All schools, $\mathbf{x} = (0, \text{mn}_j)$

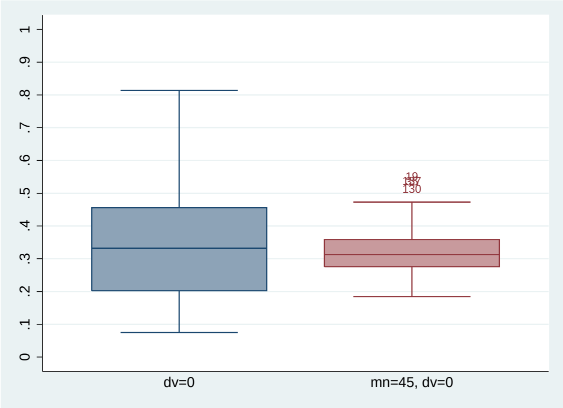


19 randomly chosen schools, $\mathbf{x} = \mathbf{x}_{ij}$



Probabilities for schools in the data (cont'd)

- In previous graphs, $\widetilde{p}(\mathbf{x})$ varies between schools for given value of $\texttt{dv}$ because $\texttt{mn}_j$ and ζ_j vary
- Isolate variability due to ζ_j by setting $\texttt{mn}_j = 45$ and $\texttt{dv}_{ij} = 0$ (boxplot on right)



Three-level ordinal cumulative logit model

- Tennessee class-size experiment
- In 4th grade, teachers assessed student participation
- 2217 students i of 262 teachers j in 75 schools k
- Ordinal response variable:
  - Teacher's rating of how frequently student pays attention y_{ijk}
  - Rated as: 1 (never), 2, 3 (sometimes), 4, 5 (always)
- No covariates for simplicity

Three-level ordinal cumulative logit model: $\text{OR}_{\text{median}}$

- Model probability of exceeding a category $s = 1, 2, 3, 4$

$$\text{logit}[\Pr(y_{ijk} > s|\zeta_{jk}^{(2)}, \zeta_k^{(3)})] = -\kappa_s + \zeta_{jk}^{(2)} + \zeta_k^{(3)}$$

  - Teacher random intercept $\zeta_{jk}^{(2)} \sim N(0, \psi^{(2)})$, $\widehat{\psi}^{(2)} = 0.44$
  - School random intercept $\zeta_k^{(3)} \sim N(0, \psi^{(3)})$, $\widehat{\psi}^{(3)} = 0.12$
- Odds ratio, comparing students of teachers j and j' in school k

$$\frac{\Pr(y_{ijk} > s|\zeta_{jk}^{(2)}, \zeta_k^{(3)})}{\Pr(y_{ij'k} > s|\zeta_{j'k}^{(2)}, \zeta_k^{(3)})} = \exp(\zeta_{jk}^{(2)} - \zeta_{j'k}^{(2)})$$

- Median of odds ratio, comparing student with larger odds to other student:

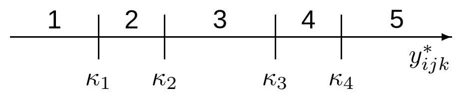
	$\text{OR}_{\text{median}}$
Different teacher, same school	1.88
Different teacher, different school	2.04

Three-level ordinal cumulative logit model: ICC_{lat}

- Latent response y_{ijk}^*

$$y_{ijk}^* = \zeta_{jk}^{(2)} + \zeta_k^{(3)} + \epsilon_{ijk}, \quad \epsilon_{ijk} \sim \text{logistic}$$

- Observed response y_{ijk}
  - results from cutting up y_{ijk}^* at cut-points or thresholds κ_1 to κ_4:



- Intraclass correlations of latent responses
  - Same school, different teacher:

$$\text{ICC}_{\text{lat}}(\text{school}) = \frac{\widehat{\psi}^{(2)}}{\widehat{\psi}^{(2)} + \widehat{\psi}^{(3)} + \pi^2/3} = 0.031$$

  - Same school, same teacher:

$$\text{ICC}_{\text{lat}}(\text{teacher}, \text{school}) = \frac{\widehat{\psi}^{(2)} + \widehat{\psi}^{(3)}}{\widehat{\psi}^{(2)} + \widehat{\psi}^{(3)} + \pi^2/3} = 0.145$$

Three-level ordinal cumulative logit model: ICC_{obs}

- Two randomly chosen students i, i' from same school (either different or same teachers)
- Pearson correlation, $\text{Cor}(y_{ijk}, y_{i'j'k})$, $\text{Cor}(y_{ijk}, y_{i'jk})$, or Kendall's τ_b:

	ICC_{lat}	ICC_{obs}	
		Cor	τ_b
Same school, different teacher	0.031	0.029	0.025
Same school, same teacher	0.145	0.136	0.118

- Method:
  - 5× 5 table of response for student i (rows) versus response for student i' (columns)

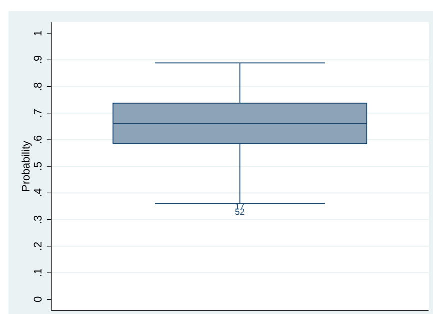
$$\overline{\pi}_{st}(\text{school}) = \Pr(y_{ijk} = s, y_{i'j'k} = t)$$
$$\overline{\pi}_{st}(\text{teacher}, \text{school}) = \Pr(y_{ijk} = s, y_{i'jk} = t)$$

  - Use these probabilities as frequency weights for Cor and τ_b

Probabilities for teachers and schools in the data

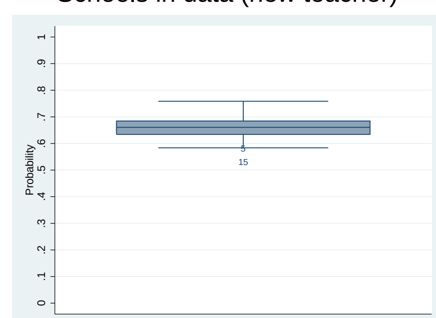
- Probability that student pays attention more than "sometimes" ($y > 3$)
- Randomly chosen student, for:

Teachers in data



Integrate over posterior of $\zeta_{jk}^{(2)}$, $\zeta_k^{(3)}$

Schools in data (new teacher)



Integrate over prior of $\zeta_{jk}^{(2)}$ and posterior of $\zeta_k^{(3)}$

Software

- Stata was used for the results and graphs
  - Datasets and Stata commands available from

    http://www.gllamm.org/pres.html
- Models estimated by ML with adaptive quadrature [Rabe-Hesketh et al., 2005]
- Commands for binary logit models: `xtlogit`, `xtmelogit`
  - `xtrho` [Rodríguez & Elo, 2003] for $\text{ICC}_{\text{obs}} = \text{VPC}$
- Command for binary, ordinal, or multinomial logit models (and other response types): `gllamm` [Rabe-Hesketh et al., 2005]
  - `gllapred` used to obtain $\overline{p}(\mathbf{x})$, $\widetilde{p}(\mathbf{x})$, $\overline{p}_{st}(\mathbf{x})$, etc. [Rabe-Hesketh & Skrondal, 2009]

Discussion

- Have discussed measures and graphs for interpreting variability implied by model with estimated parameters
- Ignored parameter uncertainty (no SEs or CIs)
- Did not consider variance explained by covariates
- Extensions
  - Include random coefficient of x_{ij}
    - for $\overline{p}(\mathbf{x})$, $\widetilde{p}(\mathbf{x})$, and ICC_{obs}, integrate over all random effects
    - ICC_{lat} and $\text{OR}_{\text{median}}$ now depend on x_{ij}
  - More levels: Straightforward
  - Relax normality assumption for random effects: 'non-parametric maximum likelihood estimation' (NPMLE)
  - Other response types
    - Nominal responses: Similar to binary responses
    - Counts: No ICC_{lat}, but can obtain $\text{IRR}_{\text{median}}$ and ICC_{obs}

References

- Duchateau, L. & Janssen, P. (2005). Understanding heterogeneity in generalized linear mixed and frailty models. *The American Statistician* 59, 143-146.
- Eldridge, S. M, et al. (2009). The intra-cluster correlation coefficient in cluster randomized trials: A review of definitions. *International Statistical Review* 77, 378-394.
- Goldstein, H., et al. (2002). Partitioning variation in multilevel models. *Understanding Statistics* 1, 223-331.
- Larsen, K., et al. (2000). Interpreting parameters in the logistic regression model with random effects. *Biometrics* 56, 909-914.
-  Rabe-Hesketh, S. & Skrondal, A. (2012). *Multilevel and Longitudinal Modeling Using Stata (Third Edition)*. College Station, TX: Stata Press.
- Rabe-Hesketh, S. & Skrondal, A. (in prep). Understanding variability in multilevel models for categorical responses.
- Rabe-Hesketh, S., et al. (2005). Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics* 128, 301-323.
- Skrondal, A. & Rabe-Hesketh, S. (2009). Prediction in multilevel generalized linear models. *Journal of the Royal Statistical Society, Series A* 172, 659-687.
- Rodríguez, G. & Elo, I. (2003). Intra-class correlation in random-effects models for binary data. *The Stata Journal* 3, 32-46.