

Mixed Models for Multilevel Data Analysis: An Applied Introduction

Don Hedeker
University of Illinois at Chicago

This work was supported by National Institute of Mental Health Contract N44MH32056.

What are Multilevel Data?

- Data that are hierarchically structured, nested, clustered
- Data collected from units organized or observed within units at a higher level (from which data are also obtained)

<i>data collected on</i>	<i>who are clustered within</i>
students	classrooms
siblings	families
repeated observations	individuals

==> these are examples of two-level data

level 1 - (students) - measurement of primary outcome and important mediating variables

level 2 - (classrooms) - provides context or organization of level-1 units which may influence outcome; other mediating variables

What is Multilevel Data Analysis?

“any set of analytical procedures that involve data gathered from individuals and from the social structure in which they are embedded and are analyzed in a manner that models the multilevel structure”

L. Burstein, *Units of Analysis*, 1985, Int Ency of Educ

- analysis that *models the multilevel structure*
- recognizes influence of structure on individual outcome

<i>structure</i>	<i>may influence response from</i>
classroom	students
family	siblings
individual	repeated observations

Why do Multilevel Data Analysis?

- assess amount of variability due to each level (*e.g.*, family variance and individual variance)

- model level 1 outcome in terms of effects at both levels

$$\textit{individual var.} = fn(\textit{individual var.} + \textit{family var.})$$

- assess interaction between level effects (*e.g.*, individual outcome influenced by family SES for males, not females)

- Responses are not independent - individuals within clusters share influencing factors

⇒ Multilevel analysis - another example of *Golden Rule of Statistics*: “one person’s error term is another person’s (or many persons’) career”

cluster	subject	<i>cluster variables</i>		<i>subject variables</i>		
		tx group	size	outcome	sex	age
1	1
	\vdots
	n_1
2	1
	\vdots
	n_2
.	1
	\vdots
	$n_{.}$
N	1
	\vdots
	n_N

$i = 1 \dots N$ clusters

$j = 1 \dots n_i$ subjects in cluster i

		<i>time-invariant variables</i>			<i>time-varying variables</i>	
subject	time	tx group	sex	age	outcome	dose
1	1
	\vdots
	n_1
2	1
	\vdots
	n_2
.	1
	\vdots
	n_i
N	1
	\vdots
	n_N

$i = 1 \dots N$ subjects

$j = 1 \dots n_i$ timepoints for subject i

Mixed models aka

- random-effects models
- random-coefficient models
- multilevel models
- hierarchical linear models

Useful for analyzing

- Clustered data
 - subjects (level-1) within clusters (level-2)
 - * e.g., clinics, hospitals, families, worksites, schools, classrooms, city wards
- Longitudinal data
 - repeated obs. (level-1) within subjects (level-2)

2-level Model for Clustered Data

Consider the model with p covariates for the $n_i \times 1$ response vector \mathbf{y} for cluster i ($i = 1, 2, \dots, N$):

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + v_i + \boldsymbol{\varepsilon}_i$$

$\mathbf{y}_i = n_i \times 1$ vector of responses for cluster i

$\mathbf{X}_i = n_i \times (p + 1)$ covariate matrix

$\boldsymbol{\beta} = (p + 1) \times 1$ vector of regression coefficients

$v_i =$ cluster effects $\sim \mathcal{NID}(0, \sigma_v^2)$

$\boldsymbol{\varepsilon}_i = n_i \times 1$ vector of residuals $\sim \mathcal{NID}(0, \sigma^2 \mathbf{I}_{n_i})$

- as cluster subscript i is present for \mathbf{y} and \mathbf{X} , cluster sample size can vary
- the covariate matrix \mathbf{X} can include
 - covariates measured at subject-level
 - covariates measured at cluster-level
 - cross-level interactions
- the total number of covariates = p
- the number of columns in \mathbf{X} is $p + 1$ to include intercept (first column of \mathbf{X} consists only of ones)

v_i - random parameter distributed $\mathcal{NID}(0, \sigma_v^2)$

- separates model from usual (fixed-effects) multiple regression model
- represent effect of subject clustering (one for every cluster)
- if subject clustering has little effect
 - estimates of $v_i \approx 0$
 - σ_v^2 will approach 0
- if subject clustering has strong effect
 - estimates of $v_i \neq 0$
 - σ_v^2 will increase from 0

$$\mathbf{y}_i \sim \mathcal{NID}(\mathbf{X}_i \boldsymbol{\beta}, \sigma_v^2 \mathbf{1}_i \mathbf{1}_i' + \sigma^2 \mathbf{I}_{n_i})$$

- usual mean from multiple regression model
- var-covar structure accounts for clustering
 - within a cluster, variance = $\sigma^2 + \sigma_v^2$ and covariance = σ_v^2
 - “compound symmetry” structure
 - ratio of the cluster variance σ_v^2 to the total variance $\sigma^2 + \sigma_v^2$ is the *intraclass correlation*

Intra-“class” correlation $r = \sigma_v^2 / (\sigma_v^2 + \sigma^2)$

- “class” is bad term, since in education “class” has meaning
- Goldstein suggests “intra-unit” correlation, replacing “unit” with appropriate term (clinic, school, family, firm *etc.*,)
- takes on values between 0 (when $\sigma_v^2 = 0$) and 1 (when $\sigma^2 = 0$)
- degree of similarity of measurements within a cluster
- ratio of variability attributable to cluster over total variability
- proportion of total (unexplained) variability of y_{ij} which is accounted for the clusters
- tends larger for smaller clusters (Kish, 1965; Donner, 1982)
 - 0.05 to 0.12 for spouse pairs, 0.0016 to 0.0126 for physician practices, 0.0005 to 0.0085 for counties
- can change depending on the dependent variable

Anorexic Women Study (Casper) - 63 sisters in 26 families
 Maximum Likelihood (ML) estimates

	Height	Psych Factor	BMI
intercept	64.166	0.568	0.352
family variance	2.743	0.031	0.000
residual variance	2.895	0.055	0.005
intra-family correlation	0.487	0.362	0.000
<i>descriptive statistics</i>			
<i>mean</i>	64.16	0.57	0.35
<i>variance</i>	5.66	0.084	0.005

Random-effects Regression Models for Clustered Data: With an Example from Smoking Prevention Research

Hedeker, Gibbons, and Flay

Journal of Consulting and Clinical Psychology, 1994,
62:757-765

The Television School and Family Smoking Prevention and Cessation Project (Flay, *et al.*, 1988); a subsample of this project was chosen with the characteristics:

- *sample* - 1600 7th-graders - 135 classrooms - 28 LA schools
 - between 1 to 13 classrooms per school
 - between 2 to 28 students per classroom
- *outcome* - knowledge of the effects of tobacco use
- *timing* - students tested at pre and post-intervention
- *design* - schools randomized to
 - a social-resistance classroom curriculum (CC)
 - a media (television) intervention (TV)
 - CC combined with TV
 - a no-treatment control group

Tobacco and Health Knowledge Scale

Subgroup Descriptive Statistics at Pretest and Post-Intervention

	CC = no		CC = yes	
	TV = no	TV = yes	TV = no	TV = yes
<i>n</i>	421	416	380	383
Pretest mean	2.152	2.087	2.050	1.979
sd	1.182	1.288	1.285	1.286
Post-Int mean	2.361	2.539	2.968	2.823
sd	1.296	1.437	1.405	1.312
Difference	0.209	0.452	0.918	0.844

Within-Cluster / Between-Cluster representation

Within-clusters model - level 1 ($j = 1, \dots, n_i$)

$$PostTHKS_{ij} = b_{0i} + [b_{1i}PreTHKS_{ij}] + \varepsilon_{ij}$$

Between-clusters model - level 2 ($i = 1, \dots, N$)

$$b_{0i} = \beta_0 + [\beta_2CC_i] + v_{0i}$$

$$b_{1i} = \beta_1$$

$$\varepsilon_{ij} \sim NID(0, \sigma^2) \quad \text{level-1 residuals}$$

$$v_{0i} \sim NID(0, \sigma_v^2) \quad \text{level-2 residuals}$$

TVSFP Study (Flay *et al.*, 1988): Tobacco and Health Knowledge *Posttest* Scores
 1600 students in 135 classrooms in 28 schools: ML estimates (and standard errors)

	<i>students in classrooms</i>			<i>students in schools</i>		
Intercept	2.618 (0.052)	2.007 (0.072)	1.757 (0.080)	2.682 (0.078)	2.047 (0.089)	1.800 (0.092)
Pretest score		0.302 (0.026)	0.310 (0.026)		0.303 (0.026)	0.310 (0.026)
Classroom curriculum			0.497 (0.086)			0.470 (0.106)
Cluster var	0.194 (0.043)	0.157 (0.037)	0.096 (0.029)	0.130 (0.045)	0.101 (0.036)	0.044 (0.020)
Residual var	1.725 (0.064)	1.601 (0.060)	1.601 (0.059)	1.788 (0.064)	1.653 (0.059)	1.653 (0.059)
ICC	0.101	0.090	0.057	0.068	0.057	0.026
$\log L$	-2760.9	-2696.4	-2681.3	-2756.8	-2692.0	-2684.7
χ^2_1		129.0	30.2		129.6	14.6

Within-Cluster / Between-Cluster representation

Within-clusters model - level 1 ($j = 1, \dots, n_i$)

$$PostTHKS_{ij} = b_{0i} + \varepsilon_{ij}$$

Between-clusters model - level 2 ($i = 1, \dots, N$)

$$b_{0i} = \beta_0 + \beta_1 CC_i + \beta_2 TV_i + \beta_3(CC_i \times TV_i) + v_{0i}$$

$$\varepsilon_{ij} \sim NID(0, \sigma^2) \quad \text{level-1 residuals}$$

$$v_{0i} \sim NID(0, \sigma_v^2) \quad \text{level-2 residuals}$$

- If cluster effect is completely explained by condition, then
 - $v_{0i} = 0$ for all i (thus $\sigma_v^2 = 0$)
 - model is same as ordinary regression (individual-level analysis)
- If $n_i = n$ for all clusters (and no level-1 covariates), then
 - model is same as ordinary regression of cluster means (cluster-level analysis)

THKS post-intervention scores - Regression estimates (se)

	<i>Ordinary Regression</i>		<i>Mixed Model</i>
	Class-level	Student-level	Students in classes
Intercept	2.342 (.117)	2.361 (.066)	2.341 (.092)
classroom curriculum (CC)	.507 (.166)	.607 (.096)	.589 (.133)
television (TV)	-.082 (.150)	.177 (.094)	.120 (.131)
interaction (CC by TV)	.011 (.236)	-.323 (.137)	-.247 (.189)
residual variance	.468	1.860	1.727 (.064)
class variance			.134 (.037)
<i>p</i> < .05	p < .01		<i>ICC</i> = .072

Within-Cluster / Between-Cluster representation

Within-clusters model - level 1 ($j = 1, \dots, n_i$)

$$PostTHKS_{ij} = b_{0i} + b_{1i}PreTHKS_{ij} + \varepsilon_{ij}$$

Between-clusters model - level 2 ($i = 1, \dots, N$)

$$b_{0i} = \beta_0 + \beta_2CC_i + \beta_3TV_i + \beta_4(CC_i \times TV_i) + v_{0i}$$

$$b_{1i} = \beta_1$$

$$\varepsilon_{ij} \sim NID(0, \sigma^2) \quad \text{level-1 residuals}$$

$$v_{0i} \sim NID(0, \sigma_v^2) \quad \text{level-2 residuals}$$

THKS Post-Intervention Scores - Regression Estimates (se)

	<i>Ordinary Regression Models</i>				<i>Mixed Models</i>					
	Class-level		Student-level		Stu in classes		Stu in schools		Three-level	
Intercept	1.3087	***	1.6613	***	1.6769	***	1.6952	***	1.6970	***
	(0.229)		(0.084)		(0.100)		(0.114)		(0.117)	
pretest THKS	0.4962	***	0.3252	***	0.3117	***	0.3103	***	0.3072	***
	(0.097)		(0.026)		(0.026)		(0.026)		(0.026)	
classroom curriculum	0.5749	***	0.6406	***	0.6334	***	0.6601	***	0.6392	***
	(0.153)		(0.092)		(0.120)		(0.144)		(0.147)	
television	-0.0150		0.1987	**	0.1597		0.2024		0.1781	
	(0.150)		(0.090)		(0.118)		(0.140)		(0.144)	
interaction	-0.0485		-0.3216	**	-0.2747		-0.3697	*	-0.3204	
	(0.216)		(0.130)		(0.169)		(0.201)		(0.206)	
error variance	0.3924		1.6929		1.6025	***	1.6523	***	1.6020	***
					(0.059)		(0.059)		(0.059)	
class variance					0.0881	***			0.0636	**
					(0.028)				(0.028)	
school variance							0.0372	**	0.0258	
							(0.018)		(0.020)	

*** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$

Results

- conclusions about CC by TV interaction differ
 - non-significant by class-level analysis, significant by student-level analysis, marginally significant by multilevel
- student-level results close to multilevel, but estimates are more similar than standard errors → underestimation of standard errors by ordinary regression analysis is expected since assumption of independence of observations is violated
- students more homogeneous within classrooms than schools
 - students within classrooms model, $r = 0.052$
 - students within schools model, $r = 0.022$
- 3-level model close to students within classrooms model
 - based on 3-level model, classroom and school effects accounted for 3.8% and 1.5% of total variance, respectively

3-level ICCs

From the three-level model:

error var = 1.6020, class var = 0.0636, school var = 0.0258

Similarity of students within the same school

$$ICC = \frac{0.0258}{1.6020 + 0.0636 + 0.0258} = .0153$$

Similarity of students within the same classrooms (and schools)

$$ICC = \frac{0.0636 + 0.0258}{1.6020 + 0.0636 + 0.0258} = .0529$$

Similarity of classes within the same school

$$ICC = \frac{0.0258}{0.0636 + 0.0258} = .289$$

Explained Variance (Hox, *Multilevel Analysis*, 2002)

$$\text{level-1 } R_1^2 = 1 - \frac{\hat{\sigma}_p^2}{\hat{\sigma}_0^2} \qquad \text{level-2 } R_2^2 = 1 - \frac{\hat{\sigma}_{v_p}^2}{\hat{\sigma}_{v_0}^2}$$

subscript 0 refers to a model with no covariates (*i.e.*, null model),
 subscript p refers to a model with p covariates (*i.e.*, full model)

e.g., students in classrooms models

level	variance	models		R^2
		null	full	
1 (students)	$\hat{\sigma}^2$	1.725	1.602	.071
2 (classrooms)	$\hat{\sigma}_v^2$.194	.088	.546

Explained Variance: 3-level model

$$R_1^2 = 1 - \frac{\hat{\sigma}_p^2}{\hat{\sigma}_0^2} \quad R_2^2 = 1 - \frac{\hat{\sigma}_{v_{(2)p}}^2}{\hat{\sigma}_{v_{(2)0}}^2} \quad R_3^2 = 1 - \frac{\hat{\sigma}_{v_{(3)p}}^2}{\hat{\sigma}_{v_{(3)0}}^2}$$

subscript 0 refers to a model with no covariates (*i.e.*, null model),
subscript p refers to a model with p covariates (*i.e.*, full model)

e.g., students in classrooms in schools models

level	variance	null	full	R^2
1 (students)	$\hat{\sigma}^2$	1.724	1.602	.071
2 (classrooms)	$\hat{\sigma}_{v_{(2)}}^2$.085	.064	.247
3 (schools)	$\hat{\sigma}_{v_{(3)}}^2$.110	.026	.764

Likelihood-ratio tests:

suppose Model I is nested within Model II

$$2 \times \log(L_{\text{II}} / L_{\text{I}}) = 2 \times (\log L_{\text{II}} - \log L_{\text{I}}) \sim \chi_q^2$$

where q = number of additional parameters in Model II

$-2 \log L$ is called the *deviance* (the higher the deviance the poorer the model fit)

$$D_{\text{I}} - D_{\text{II}} \sim \chi_q^2$$

to evaluate the null hypothesis that the additional parameters in Model II jointly equal 0

Comparison of models using LR tests

Model	deviance	CM	χ^2	df	$p <$	halved $p <$
1. student-level	5377.90					
2a. students in classes	5359.83	1	18.07	1	.001	.001
2b. students in schools	5366.01	1	11.89	1	.001	.001
3. three-level	5357.36	1	20.54	2	.001	.001
		2a	2.47	1	.116	.058

LR tests with halved p -values (akin to one-tailed p -values) for tests of variance and covariance parameters is recommended (see Snijders & Bosker, *Multilevel Analysis*, 1999, pps. 90-91)