

Multilevel Workshop

Andrew Grogan-Kaylor

2025-07-02

Table of contents

1	Introduction	5
1.1	Background	5
1.2	Simulated Multilevel Data	5
2	Two Level Cross Sectional; And Three Level Longitudinal Models	7
2.1	Cross Sectional Model	7
2.1.1	Get Data	7
2.1.2	The Equation	7
2.1.3	Descriptive Statistics	7
2.1.4	Spaghetti Plot	8
2.1.5	Unconditional Model	8
2.1.6	Conditional Model	10
2.2	Longitudinal Model	11
2.2.1	Get Data	11
2.2.2	The Equation	11
2.2.3	Descriptive Statistics	11
2.2.4	Alternate Plot	12
2.2.5	Unconditional Model	12
2.2.6	Conditional Model	13
2.3	Nice Table of Results	14
2.4	QUESTIONS???	15
3	Cross-Classified Models	16
3.1	Introduction	16
3.2	Get Data	16
3.3	Cross Classified Model	16
3.3.1	Standard (Less Computationally Efficient) Syntax	16
3.3.2	Cross Classified With Computationally Efficient Syntax	17
3.4	Three Level Model	18
3.5	Nice Table of Results of Three Level and Cross Classified Model	20
3.6	Verification of Syntax Equivalence for Cross Classified Model	20
3.7	QUESTIONS???	21

List of Figures

1.1	Simulated Multilevel Data	6
2.1	Spaghetti Plot of Outcome by Warmth by Country	8
2.2	Alternate Plot of Outcome by Time by Individual; First 10 Observations	12

List of Tables

1 Introduction

This site contains materials for a workshop on multilevel modeling.

1.1 Background

Multilevel models are useful when you have data that are nested or clustered inside social units such as schools, neighborhoods, states, or countries.

Multilevel models are also useful when you have longitudinal data where repeated measures are collected for study participants.

1.2 Simulated Multilevel Data

The data used in these workshop materials are *simulated* data on parents, children and families. The data are simulated to come from 30 hypothetical countries around the world. These are the same data used and discussed in my book *Multilevel Thinking: Discovering Variation, Universals, and Particulars in Cross-Cultural Research*.

There are two versions of the data: a *cross-sectional* data set from a single point in time; a *longitudinal* version of the data spanning several time points.

i The Data Can Be Downloaded Here:

- [Cross Sectional Data](#)
- [Longitudinal Data](#)

pos	variable	label
1	country	country id
2	HDI	Human Development Index
3	family	family id
4	id	unique country family id
5	identity	hypothetical identity group variable
6	intervention	recieved intervention
7	physical_punishment	physical punishment in past week
8	warmth	parental warmth in past week
9	outcome	beneficial outcome

Figure 1.1: Simulated Multilevel Data

2 Two Level Cross Sectional; And Three Level Longitudinal Models

2.1 Cross Sectional Model

2.1.1 Get Data

```
use "simulated_multilevel_data.dta", clear
```

2.1.2 The Equation

$$\text{outcome}_{ij} = \beta_0 + \beta_1 \text{parental warmth} + \beta_2 \text{physical punishment} + \beta_3 \text{time} +$$

$$\beta_4 \text{identity}_2 + \beta_5 \text{intervention} + \beta_6 \text{HDI} +$$

$$u_{0j} + u_{1j} \times \text{parental warmth} + e_{ij}$$

2.1.3 Descriptive Statistics

```
summarize // descriptive statistics
```

Variable	Obs	Mean	Std. dev.	Min	Max
country	3,000	15.5	8.656884	1	30
HDI	3,000	64.76667	17.24562	33	87
family	3,000	50.5	28.87088	1	100
id	0				
identity	3,000	.4976667	.5000779	0	1

intervention		3,000	.4843333	.4998378	0	1
physical_p~t		3,000	2.478667	1.360942	0	5
warmth		3,000	3.521667	1.888399	0	7
outcome		3,000	52.43327	6.530996	29.60798	74.83553

2.1.4 Spaghetti Plot

```
spagplot outcome warmth, id(country) scheme(stcolor)
graph export spagplot1.png, width(1000) replace
```

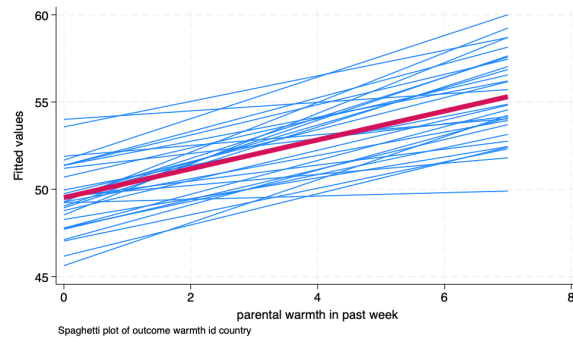


Figure 2.1: Spaghetti Plot of Outcome by Warmth by Country

2.1.5 Unconditional Model

2.1.5.1 Model

```
mixed outcome || country: // unconditional model
```

Performing EM optimization ...

Performing gradient-based optimization:

Iteration 0: Log likelihood = -9802.8371

Iteration 1: Log likelihood = -9802.8371

Computing standard errors ...

Mixed-effects ML regression
Group variable: country

Number of obs = 3,000
Number of groups = 30
Obs per group:
min = 100
avg = 100.0
max = 100
Wald chi2(0) = .
Prob > chi2 = .

Log likelihood = -9802.8371

outcome	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
-----+-----						
_cons	52.43327	.3451217	151.93	0.000	51.75685	53.1097
-----+-----						

Random-effects parameters		Estimate	Std. err.	[95% conf. interval]	
-----+-----					
country: Identity					
	var(_cons)	3.178658	.9226737	1.799552	5.614658
-----+-----					
	var(Residual)	39.46106	1.024013	37.50421	41.52
-----+-----					

LR test vs. linear model: chibar2(01) = 166.31 Prob >= chibar2 = 0.0000

2.1.5.2 ICC

```
estat icc
```

Intraclass correlation

Level	ICC	Std. err.	[95% conf. interval]	
-----+-----				
country	.0745469	.0201254	.0434963	.1248696
-----+-----				

2.1.6 Conditional Model

```
mixed outcome warmth physical_punishment identity i.intervention HDI || country: warmth // ml
est store crosssectional // store estimates
```

Performing EM optimization ...

Performing gradient-based optimization:

Iteration 0: Log likelihood = -9626.6279

Iteration 1: Log likelihood = -9626.607

Iteration 2: Log likelihood = -9626.607

Computing standard errors ...

Mixed-effects ML regression

Group variable: country

Number of obs = 3,000

Number of groups = 30

Obs per group:

min = 100

avg = 100.0

max = 100

Wald chi2(5) = 334.14

Prob > chi2 = 0.0000

Log likelihood = -9626.607

outcome	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
warmth	.8345368	.0637213	13.10	0.000	.7096453	.9594282
physical_punishment	-.9916657	.0797906	-12.43	0.000	-1.148052	-.8352791
identity	-.3004767	.2170295	-1.38	0.166	-.7258466	.1248933
1.intervention	.6396427	.2174519	2.94	0.003	.2134448	1.065841
HDI	-.003228	.0199257	-0.16	0.871	-.0422817	.0358256
_cons	51.99991	1.371257	37.92	0.000	49.3123	54.68753

Random-effects parameters	Estimate	Std. err.	[95% conf. interval]	
country: Independent				
var(warmth)	.0227504	.0257784	.0024689	.2096436
var(_cons)	2.963975	.9737647	1.556777	5.643163

<hr/>				
var(Residual)		34.97499	.9097109	33.23668 36.80422

LR test vs. linear model: chi2(2) = 205.74 Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.

2.2 Longitudinal Model

2.2.1 Get Data

```
use "simulated_multilevel_longitudinal_data.dta", clear
```

2.2.2 The Equation

$$\text{outcome}_{ij} = \beta_0 + \beta_1 \text{parental warmth} + \beta_2 \text{physical punishment} + \beta_3 \text{time} +$$

$$\beta_4 \text{identity}_2 + \beta_5 \text{intervention} + \beta_5 \text{HDI} +$$

$$u_{0j} + u_{1j} \times \text{parental warmth} +$$

$$v_{0i} + v_{1i} \times t + e_{ij}$$

2.2.3 Descriptive Statistics

```
summarize // descriptive statistics
```

Variable	Obs	Mean	Std. dev.	Min	Max
country	9,000	15.5	8.655922	1	30
HDI	9,000	64.76667	17.2437	33	87
family	9,000	50.5	28.86767	1	100
id	0				

identity	9,000	.4976667	.5000223	0	1
-----+-----					
intervention	9,000	.4843333	.4997823	0	1
t	9,000	2	.8165419	1	3
physical_p~t	9,000	2.485333	1.373639	0	5
warmth	9,000	3.514222	1.8839	0	7
outcome	9,000	53.37768	6.572285	29.60798	79.02199

2.2.4 Alternate Plot

```

encode id, generate(idNUMERIC) // numeric version of id

* spagplot outcome t if idNUMERIC <= 10, id(idNUMERIC) scheme(stcolor)

twoway (lfit outcome t) (scatter outcome t) if idNUMERIC <= 10, by(idNUMERIC) scheme(stcolor)

graph export spagplot2.png, width(1000) replace

```

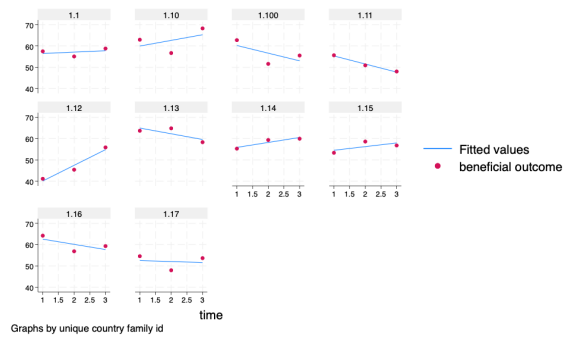


Figure 2.2: Alternate Plot of Outcome by Time by Individual; First 10 Observations

2.2.5 Unconditional Model

2.2.5.1 Model

```

mixed outcome || country: || id: // unconditional model

```

2.2.5.2 ICC

```
estat icc
```

Intraclass correlation

Level	ICC	Std. err.	[95% conf. interval]	
country	.0748336	.0190847	.0450028	.1219141
id country	.3462837	.0171461	.3134867	.3806097

2.2.6 Conditional Model

```
mixed outcome t warmth physical_punishment i.identity i.intervention HDI || country: warmth
est store longitudinal // store estimates
```

Performing EM optimization ...

Performing gradient-based optimization:

Iteration 0: Log likelihood = -28523.49

Iteration 1: Log likelihood = -28499.987

Iteration 2: Log likelihood = -28499.739

Iteration 3: Log likelihood = -28499.604

Iteration 4: Log likelihood = -28499.603

Computing standard errors ...

Mixed-effects ML regression

Number of obs = 9,000

Grouping information

Group variable	No. of groups	Observations per group		
		Minimum	Average	Maximum
country	30	300	300.0	300

id		3,000	3	3.0	3		

Log likelihood = -28499.603				Wald chi2(6) = 1096.15		Prob > chi2 = 0.0000	

outcome	Coefficient	Std. err.	z	P> z	[95% conf. interval]		
-----+-----							
t	.943864	.0658716	14.33	0.000	.814758	1.07297	
warmth	.9134959	.0423732	21.56	0.000	.830446	.9965459	
physical_punishment	-1.007897	.0497622	-20.25	0.000	-1.105429	-.9103647	
1.identity	-.1276926	.1515835	-0.84	0.400	-.4247908	.1694057	
1.intervention	.8589966	.1519095	5.65	0.000	.5612596	1.156734	
HDI	-.0005657	.0196437	-0.03	0.977	-.0390666	.0379352	
_cons	50.46724	1.338318	37.71	0.000	47.84418	53.09029	

Random-effects parameters		Estimate	Std. err.	[95% conf. interval]			
-----+-----							
country: Independent							
var(warmth)	.0107586	.0127845		.0010478	.1104703		
var(_cons)	3.167085	.9146761		1.798154	5.578181		
-----+-----							
id: Independent							
var(t)	3.58e-09	7.06e-07		3.5e-177	3.7e+159		
var(_cons)	8.387275	.4724188		7.510631	9.366242		
-----+-----							
var(Residual)	26.02733	.4753701		25.11211	26.97592		

LR test vs. linear model: chi2(4) = 1247.03				Prob > chi2 = 0.0000			

Note: LR test is conservative and provided only for reference.

2.3 Nice Table of Results

```
etable, estimates(crosssectional longitudinal) ///
showstars showstarsnote /// show stars and note
column(estimate) // column is modelname
```

	crosssectional	longitudinal
parental warmth in past week	0.835 ** (0.064)	0.913 ** (0.042)
physical punishment in past week	-0.992 ** (0.080)	-1.008 ** (0.050)
hypothetical identity group variable	-0.300 (0.217)	
recieved intervention		
1	0.640 ** (0.217)	0.859 ** (0.152)
Human Development Index	-0.003 (0.020)	-0.001 (0.020)
time		0.944 ** (0.066)
hypothetical identity group variable		
1		-0.128 (0.152)
Intercept	52.000 ** (1.371)	50.467 ** (1.338)
var(warmth)	0.023 (0.026)	0.011 (0.013)
var(_cons)	2.964 (0.974)	3.167 (0.915)
var(e)	34.975 (0.910)	26.027 (0.475)
var(_cons)		8.387 (0.472)
var(t)		0.000 (0.000)
Number of observations	3000	9000
** p<.01, * p<.05		

2.4 QUESTIONS???

3 Cross-Classified Models

3.1 Introduction

A two level multilevel model imagines that *Level 1* units are nested in *Level 2* units. A three level multilevel model imagines that *Level 1* units are nested in *Level 2* units, which are in turn nested in *Level 3*.

A cross-classified model imagines that the nesting is not hierarchical, but rather that there are two sets of clusters or nestings in which individuals may be nested.

3.2 Get Data

```
use "simulated_multilevel_longitudinal_data.dta", clear
```

3.3 Cross Classified Model

We can treat these random effects as being *cross classified*.

This might be useful if we had data where individuals lived in different countries at different times.

However, because `id` is in fact nested inside `country`, in this case, estimating the random effects as cross classified will be more time consuming, but will give us equivalent results to a three level model.

3.3.1 Standard (Less Computationally Efficient) Syntax

The below syntax will take a very long time to run with the full sample, and thus we have commented it out.


```
* mixed outcome t warmth physical_punishment || _all: R.country || _all: R.id
* est store crossed1
```

The documentation notes that we can use a *much* more computationally efficient version of the above command, which is what we do in these notes. The user can verify that both versions of the command will produce equivalent results.

In fact, at the end of handout we verify the similarity of both sets of syntax using a random sample.

3.3.2 Cross Classified With Computationally Efficient Syntax

```
mixed outcome t warmth physical_punishment || _all: R.country || id:
est store crossed2 // store crossed effects result
```

Performing EM optimization ...

Performing gradient-based optimization:
 Iteration 0: Log likelihood = -28516.314
 Iteration 1: Log likelihood = -28516.277
 Iteration 2: Log likelihood = -28516.277

Computing standard errors ...

Mixed-effects ML regression Number of obs = 9,000

Grouping information

Group variable		No. of groups	Observations per group		
			Minimum	Average	Maximum
-----+-----					
_all		1	9,000	9,000.0	9,000
id		3,000	3	3.0	3

Log likelihood = -28516.277 Wald chi2(3) = 1168.69
Prob > chi2 = 0.0000

	outcome	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
	t	.9434605	.065866	14.32	0.000	.8143654	1.072556
	warmth	.9053924	.0380439	23.80	0.000	.8308277	.9799572
physical_punishment		-1.014385	.0499354	-20.31	0.000	-1.112257	-.916514
	_cons	50.8301	.4123007	123.28	0.000	50.022	51.63819

Random-effects parameters		Estimate	Std. err.	[95% conf. interval]	
_all: Identity					
	var(R.country)	3.429974	.930313	2.015668	5.836634
id: Identity					
	var(_cons)	8.608872	.4757699	7.725107	9.59374
	var(Residual)	26.02862	.4752444	25.11363	26.97695

LR test vs. linear model: $\chi^2(2) = 1260.84$ Prob > $\chi^2 = 0.0000$

Note: LR test is conservative and provided only for reference.

3.4 Three Level Model

```
mixed outcome t warmth physical_punishment || country: || id: // 3 level w/ random intercept
est store threelevel // store random intercept model
```

Performing EM optimization ...

Performing gradient-based optimization:

Iteration 0: Log likelihood = -28516.314

Iteration 1: Log likelihood = -28516.277

Iteration 2: Log likelihood = -28516.277

Computing standard errors ...

Mixed-effects ML regression

Number of obs = 9,000

Grouping information

Group variable		No. of groups	Observations per group		
			Minimum	Average	Maximum
country		30	300	300.0	300
id		3,000	3	3.0	3

Log likelihood = -28516.277

Wald chi2(3) = 1168.69

Prob > chi2 = 0.0000

outcome	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
t	.9434605	.065866	14.32	0.000	.8143654	1.072556
warmth	.9053924	.0380439	23.80	0.000	.8308277	.9799572
physical_punishment	-1.014385	.0499354	-20.31	0.000	-1.112257	-.916514
_cons	50.8301	.4123007	123.28	0.000	50.022	51.63819

Random-effects parameters		Estimate	Std. err.	[95% conf. interval]	
country: Identity					
	var(_cons)	3.429974	.930313	2.015668	5.836634
id: Identity					
	var(_cons)	8.608872	.4757699	7.725107	9.59374
	var(Residual)	26.02862	.4752444	25.11363	26.97695

LR test vs. linear model: chi2(2) = 1260.84

Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.

3.5 Nice Table of Results of Three Level and Cross Classified Model

```
etable, estimates(threelevel crossed2), ///
showstars showstarsnote /// show stars and note
column(estimate) // column is modelname
```

```
invalid 'showstars'  
r(198);
```

```
r(198);
```

3.6 Verification of Syntax Equivalence for Cross Classified Model

```
keep if family <= 5 // random sample of families

quietly mixed outcome t warmth physical_punishment || _all: R.country || _all: R.id

est store crossed1A // less efficient syntax

quietly mixed outcome t warmth physical_punishment || _all: R.country || id:

est store crossed2A // more efficient syntax

etable, estimates(crossed1A crossed2A) ///
showstars showstarsnote /// show stars and note
column(estimate) // column is modelname
```

(8,550 observations deleted)

crossed1A crossed2A

```

-----
time                0.745 **    0.745 **
                   (0.281)    (0.281)
parental warmth in past week  0.871 **    0.871 **
                   (0.160)    (0.160)
physical punishment in past week -1.262 ** -1.262 **
                   (0.206)    (0.206)
Intercept           51.755 **    51.755 **
                   (1.009)    (1.009)
var(R_country)       2.245
                   (1.319)    (1.319)
var(R_id)            5.425
                   (1.843)
var(e)              23.638    23.638
                   (1.933)    (1.933)
var(_cons)          5.425
                   (1.843)
Number of observations      450    450
-----
** p<.01, * p<.05

```

3.7 QUESTIONS???