Methods: Mind the Gap
Webinar Series

An Introduction to Cross-classified, Multiple Membership, and Dynamic Group Multilevel Models

Presented by:

Donald Hedeker, Ph.D.

University of Chicago

National Institutes of Health
Office of Disease Prevention

Outline

- Cross-sectional clustered data (at 1 timepoint)
  - Multilevel introduction - slides 3-4
  - Cross-classified multilevel models - slides 5-19
  - Multiple membership multilevel models - slides 20-31
- Longitudinal & clustered data
  - Introduction - slide 32
  - Acute-effects cross-classified multilevel models - slides 33-39
  - Dynamic group multilevel models - slides 40-46
- Summary - slide 47
- References - slides 48-49
- Appendix with SPSS & Stata examples - slides 50-73

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)

data collected on	*who are clustered within*
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

Why do Multilevel Data Analysis?

- assess amount of variability due to each level (*e.g.*, classroom variance and student variance)
- model level 1 outcome in terms of effects at both levels

$$student\ outcome = fn(student\ variables + class\ variables)$$

- assess interaction between level effects (*e.g.*, student outcome influenced by class size for males, not females)
- Responses are not independent - students within classrooms share influencing factors

$\Rightarrow$ Multilevel analysis - another example of *Golden Rule of Statistics*: "one person's error term is another person's (or many persons') career"

What are Cross-Classified Multilevel Data?

- Subjects are classified by two or more types of clusters, but clusters are not hierachicial or nested within one another

data collected on	*who are clustered within*
students	schools and neighborhoods
patients	providers and hospitals

Here,

- students from the same neighborhoods go to different schools
- patients can be seen by the same provider at different hospitals

Nested or Hierarchical Structure

	School 1	School 2	School 3	School 4
Neighborhood 1	x x x	x x		
Neighborhood 2			x x x x	x x x x x

students nested within schools within neighborhoods

Crossed Structure

	School 1	School 2	School 3	School 4
Neighborhood 1	x x x x	x x	x x x	x x
Neighborhood 2	x x x		x x	x x x x

students nested within crossing of schools and neighborhoods

$\Rightarrow$ interest is on assessing effects of schools and neighborhoods on student outcomes

Example: Cross-Classified Multilevel Models. LEMMA Module 12. (http://www.bristol.ac.uk/cmm/learning/online-course/course-topics.html)

Data from 3,435 children who attended 148 primary schools and 19 secondary schools in Scotland.

- VRQ: A verbal reasoning score from tests pupils took when they entered secondary school
- ATTAIN: Attainment score of pupils at age 16
- PID: Primary school identifying code
- SEXF: Pupil's gender (0 = boy and 1 = girl)
- SC: Pupil's social class scale (continuous score from low to high social class)
- SID: Secondary school identifying code
- FED: Father's education (0,1)
- CHOICE: Choice of secondary school that they attend (1=first choice, ... 4=fourth choice)
- MED: Mother's education (0,1)

Excel file: xwdata_9var.xlsx

	A	B	C	D	E	F	G	H	I
1	VRQ	ATTAIN	PID	SEXF	SC	SID	FED	CHOICE	MED
2	111	10	1	0	0	9	0	1	0
3	74	2	1	1	0	9	0	1	0
4	92	4	1	1	1	9	1	1	0
5	98	4	1	0	0	9	0	1	0
6	101	8	1	0	31	1	0	4	0
7	92	5	1	1	0	9	0	1	0
8	121	10	1	0	0	9	0	1	0
9	90	5	1	1	0	9	1	1	1
10	81	2	1	1	0	9	0	1	0
11	96	4	1	1	0	9	0	1	0
12	70	1	1	0	0	18	0	1	0
13	76	2	1	1	0	1	0	4	0
14	101	4	1	1	31	9	1	1	0
15	112	9	1	0	0	9	0	1	0
16	103	6	1	1	0	1	0	2	1
17	96	3	1	1	31	9	0	1	0
18	81	3	1	0	0	9	0	1	0
19	100	6	1	0	0	9	0	1	1
20	109	8	1	1	20	9	0	1	0
21	107	9	1	1	0	9	1	1	0
22	102	10	1	0	20	9	0	1	0
23	83	6	1	1	0	1	0	1	0
24	71	2	1	1	0	1	0	1	0
25	105	4	1	1	0	9	0	1	0
26	96	4	1	1	31	9	1	1	0
27	75	2	1	1	0	9	0	1	0
28	81	2	1	0	0	9	0	1	0
29	91	4	1	0	0	9	1	1	0
30	89	6	1	0	0	1	0	1	1
31	83	4	1	1	0	1	0	2	0

data are sorted by Primary School ID (PID), but appear unsorted in terms of Secondary School ID (SID)

SAS PROC MIXED code: `Fife_CrossClassified.sas`

Null model

```
PROC MIXED  DATA=xw COVTEST;
  CLASS pid sid ;
  MODEL attain =  / S;
  RANDOM pid sid;
RUN;
```

Model including verbal reasoning score as a covariate

```
PROC MIXED  DATA=xw COVTEST;
  CLASS pid sid;
  MODEL attain = vrq / S;
  RANDOM pid sid;
RUN;
```

`SAS abbreviation: S=SOLUTION`

The Mixed Procedure

Model Information

Dependent Variable	ATTAIN
Estimation Method	REML

Class Level Information

Class	Levels	Values
PID	148	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148
SID	19	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19

Number of Observations

Number of Observations Read	3435
Number of Observations Used	3435

Covariance Parameter Estimates

Cov Parm	Estimate	Standard Error	Z Value	Pr > Z
PID	1.1300	0.2074	5.45	<.0001
SID	0.3722	0.1743	2.14	0.0164
Residual	8.1107	0.2004	40.46	<.0001

Fit Statistics

-2 Res Log Likelihood	17150.8
AIC (Smaller is Better)	17156.8
AICC (Smaller is Better)	17156.8
BIC (Smaller is Better)	17165.8

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	5.5017	0.1787	18	30.79	<.0001

Intraclass correlations

For the null model (no covariates), Residual variance = 8.1107, Primary school variance = 1.13, Secondary school variance = 0.3722

Students from same primary school, but different secondary schools

$$ICC = \frac{1.13}{8.1107 + 1.13 + 0.3722} = 0.118$$

Students from same secondary school, but different primary schools

$$ICC = \frac{0.3722}{8.1107 + 1.13 + 0.3722} = 0.039$$

Students from same primary and secondary schools

$$ICC = \frac{1.13 + 0.3722}{8.1107 + 1.13 + 0.3722} = 0.156$$

Crossed random-effects model with random interaction

- For this, need combinations of primary and secondary schools where there are several observations
- Can relax assumption of additive random effects
- Some secondary schools might benefit students from particular primary schools more than other primary schools

Crossed structure with singletons only - CANNOT estimate interaction

	Secondary 1	Secondary 2	Secondary 3	Secondary 4
Primary 1	x	x	x	
Primary 2	x	x		x
Primary 3	x	x	x	x
Primary 4	x		x	x

Crossed structure with duplications - CAN estimate interaction

	Secondary 1	Secondary 2	Secondary 3	Secondary 4
Primary 1	xxx	xx	xxx	
Primary 2	xx	x		xxx
Primary 3	x	xxx	xx	xx
Primary 4	xxxx		x	xxxx

Crossed random effects with interaction

SAS

```
PROC MIXED COVTEST;
  CLASS pid sid ;
  MODEL attain =  / S;
  RANDOM pid sid pid*sid;
```

SAS output: model with cluster interactions

Covariance Parameter Estimates

Cov Parm	Estimate	Standard Error	Z Value	Pr > Z
PID	0.9144	0.2954	3.10	0.0010
SID	0.3376	0.1721	1.96	0.0249
PID*SID	0.2335	0.2479	0.94	0.1731
Residual	8.0873	0.2012	40.20	<.0001

Fit Statistics

-2 Res Log Likelihood	17149.7
AIC (Smaller is Better)	17157.7
AICC (Smaller is Better)	17157.7
BIC (Smaller is Better)	17169.6

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	5.4988	0.1730	18	31.79	<.0001

Test of interaction effect of clustering

- model assuming no interaction: -2 Res Log L = 17150.8
- model allowing for interaction: -2 Res Log L =17149.7
- LR $X_1^2 = 17150.8 - 17149.7 = 1.1$; not significant

$\Rightarrow$ No interaction effect on clustering; additive model is reasonable

Intraclass correlations - model with RE interaction

Null model (no covariates), Residual var = 8.0873, Primary var = 0.9144, Secondary var = 0.3376, Interaction var = 0.2335

Students from same primary school, but different secondary schools

$$ICC = \frac{0.9144}{8.0873 + 0.9144 + 0.3376 + 0.2335} = 0.096$$

Students from same secondary school, but different primary schools

$$ICC = \frac{0.3376}{8.0873 + 0.9144 + 0.3376 + 0.2335} = 0.035$$

Students from same primary and secondary schools

$$ICC = \frac{0.9144 + 0.3376 + 0.2335}{8.0873 + 0.9144 + 0.3376 + 0.2335} = 0.155$$

Crossed random effect models - show me the equations!

Subject k nested within crossing of primary schools i and secondary schools j

model with additive random effects

$$y_{ijk} = \boldsymbol{x}'_{ijk}\boldsymbol{\beta} + \upsilon_i + \upsilon_j + \epsilon_{ijk}$$

$$\upsilon_i \sim N(0, \sigma_P^2) \quad \upsilon_j \sim N(0, \sigma_S^2) \quad \epsilon_{ijk} \sim N(0, \sigma_\epsilon^2)$$

model with random effects interaction

$$y_{ijk} = \boldsymbol{x}'_{ijk}\boldsymbol{\beta} + \upsilon_i + \upsilon_j + \upsilon_{ij} + \epsilon_{ijk}$$

$$\upsilon_i \sim N(0, \sigma_P^2) \quad \upsilon_j \sim N(0, \sigma_S^2) \quad \upsilon_{ij} \sim N(0, \sigma_{PS}^2) \quad \epsilon_{ijk} \sim N(0, \sigma_\epsilon^2)$$

Multiple Membership Models

- Subjects are nested within more than one cluster

data collected on	*who are clustered within*
students	more than one teacher
patients	more than one provider

Here,

- assume there are known weights that represent the degree of membership for a subject to the different clusters
- sum of weights equals one
- possibly do sensitivity analysis to examine how different choices for weights affect results

Nested or Hierarchical Structure weights

	Teacher 1	Teacher 2	Teacher 3	Teacher 4
Student 1	1	0	0	0
Student 2	0	1	0	0
Student 3	0	0	1	0
Student 4	0	0	0	1

each student nested within one teacher only

Multiple Membership Structure weights

	Teacher 1	Teacher 2	Teacher 3	Teacher 4
Student 1	.25	.25	.25	.25
Student 2	0	1	0	0
Student 3	.33	.33	0	.33
Student 4	0	0	.5	.5

students (possibly) nested within multiple teachers

Example: Multiple Membership Multilevel Models. LEMMA Module 13. (http://www.bristol.ac.uk/cmm/learning/online-course/course-topics.html)

Simulated data from 1,000 patients who were treated in all by 25 nurses: 400 treated by only one nurse, 300 treated by two nurses, 200 by three nurses, and 100 by four nurses.

- `patient`: Patient ID
- `satis`: Patient post-op satisfaction (mean=0, std=1)
- `assess`: Patient pre-op assessment (mean=0, std=1); higher scores are better
- `nurses`: Number of nurses seen by the patient (1 to 4)
- `n1st`: Nurse ID for patient's 1st nurse
- `n2nd`: Nurse ID for patient's 2nd nurse
- `n3rd`: Nurse ID for patient's 3rd nurse
- `n4th`: Nurse ID for patient's 4th nurse
- `p1`: Proportion of time with nurse 1
- `p2`: Proportion of time with nurse 2
- $\vdots \quad \vdots$
- `p25`: Proportion of time with nurse 25
- `h1`: Job Happiness score for nurse 1
- `h2`: Job Happiness score for nurse 2
- $\vdots \quad \vdots$
- `h25`: Job Happiness score for nurse 25

Excel file: nursedat2.xlsx - some subjects seen by only 1 nurse

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	patient	satis	assess	nurses	n1st	n2nd	n3rd	n4th	p1	p2	p3	p4	p5	p6
2	1	2.08535	0.779168	1	24	0	0	0	0	0	0	0	0	0
3	2	-0.12707	0.675926	1	10	0	0	0	0	0	0	0	0	0
4	3	-0.33507	-2.2283	1	15	0	0	0	0	0	0	0	0	0
5	4	-0.89718	0.627285	1	16	0	0	0	0	0	0	0	0	0
6	5	1.160396	-0.39504	1	5	0	0	0	0	0	0	0	1	0
7	6	-0.19741	0.219889	1	19	0	0	0	0	0	0	0	0	0
8	7	1.679402	0.93098	1	4	0	0	0	0	0	0	1	0	0
9	8	1.333309	1.040585	1	25	0	0	0	0	0	0	0	0	0
10	9	-1.18201	-0.63619	1	23	0	0	0	0	0	0	0	0	0
11	10	-0.76671	-0.83759	1	5	0	0	0	0	0	0	0	1	0
12	11	0.606306	-0.49204	1	1	0	0	0	1	0	0	0	0	0
13	12	-1.12415	-1.09006	1	12	0	0	0	0	0	0	0	0	0
14	13	1.334182	-1.77805	1	18	0	0	0	0	0	0	0	0	0
15	14	0.292212	0.05709	1	15	0	0	0	0	0	0	0	0	0
16	15	-0.83327	0.118375	1	10	0	0	0	0	0	0	0	0	0
17	16	0.238606	0.787561	1	22	0	0	0	0	0	0	0	0	0
18	17	-0.04273	-0.34001	1	9	0	0	0	0	0	0	0	0	0
19	18	-0.42192	-0.49124	1	25	0	0	0	0	0	0	0	0	0
20	19	-0.41706	0.5773	1	19	0	0	0	0	0	0	0	0	0
21	20	0.96667	0.53848	1	14	0	0	0	0	0	0	0	0	0
22	21	-2.4744	-0.92708	1	8	0	0	0	0	0	0	0	0	0
23	22	-1.06177	-0.42448	1	2	0	0	0	0	1	0	0	0	0
24	23	-1.1788	0.163695	1	21	0	0	0	0	0	0	0	0	0
25	24	0.048997	-0.19239	1	3	0	0	0	0	0	1	0	0	0

value of 1 for only one of the variables `p1` to `p25` (all others equal 0)

Excel file: nursedat2.xlsx - some subjects seen by 2 nurses

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
402	401	-0.58681	0.191419	2	22	16	0	0	0	0	0	0	0	0	0	0
403	402	1.168098	0.177525	2	11	6	0	0	0	0	0	0	0	0.45	0	0
404	403	0.797301	0.335763	2	8	2	0	0	0	0.55	0	0	0	0	0	0.45
405	404	-0.61233	1.781641	2	6	12	0	0	0	0	0	0	0	0.61	0	0
406	405	-0.08651	0.250836	2	21	13	0	0	0	0	0	0	0	0	0	0
407	406	1.249969	-0.70651	2	1	11	0	0	0.74	0	0	0	0	0	0	0
408	407	-1.90045	-0.65104	2	4	22	0	0	0	0	0	0.47	0	0	0	0
409	408	-1.13544	-0.03925	2	12	20	0	0	0	0	0	0	0	0	0	0
410	409	0.402955	0.168182	2	13	9	0	0	0	0	0	0	0	0	0	0
411	410	0.72849	0.266455	2	20	12	0	0	0	0	0	0	0	0	0	0
412	411	1.053603	1.229129	2	10	15	0	0	0	0	0	0	0	0	0	0
413	412	-0.11927	-1.09804	2	13	11	0	0	0	0	0	0	0	0	0	0
414	413	1.613424	1.555685	2	22	4	0	0	0	0	0	0.29	0	0	0	0
415	414	1.050873	1.531523	2	4	21	0	0	0	0	0	0.4	0	0	0	0
416	415	1.30818	0.140375	2	22	5	0	0	0	0	0	0	0.65	0	0	0
417	416	0.117984	1.92247	2	24	25	0	0	0	0	0	0	0	0	0	0
418	417	-1.72695	-1.71542	2	17	1	0	0	0.39	0	0	0	0	0	0	0
419	418	0.460634	1.503006	2	20	22	0	0	0	0	0	0	0	0	0	0
420	419	-1.30291	-1.80717	2	15	23	0	0	0	0	0	0	0	0	0	0
421	420	0.834141	-0.59873	2	15	4	0	0	0	0	0	0.63	0	0	0	0
422	421	-1.75112	1.245672	2	13	6	0	0	0	0	0	0	0	0.56	0	0
423	422	1.034211	1.46297	2	3	17	0	0	0	0	0.74	0	0	0	0	0
424	423	1.118368	-0.40199	2	18	6	0	0	0	0	0	0	0	0.56	0	0
425	424	0.573003	-0.20438	2	13	10	0	0	0	0	0	0	0	0	0	0
426	425	1.315172	0.107181	2	5	14	0	0	0	0	0	0	0.27	0	0	0

Two of the variables `p1` to `p25` are non-zero and their sum equals 1

Similarly, for patients seen by 3 (or 4) nurses: three (or four) of the variables `p1` to `p25` are non-zero and their sum equals 1

Excel file: nursedat2.xlsx - Job Happiness values (mean=0, sd=1) of nurses

AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ
h1	h2	h3	h4	h5	h6	h7	h8	h9	h10
1.767854	-0.00445	-0.03987	-0.3772	-1.3821	0.922799	-0.11437	-1.30807	0.786372	1.034006
1.767854	-0.00445	-0.03987	-0.3772	-1.3821	0.922799	-0.11437	-1.30807	0.786372	1.034006
1.767854	-0.00445	-0.03987	-0.3772	-1.3821	0.922799	-0.11437	-1.30807	0.786372	1.034006
1.767854	-0.00445	-0.03987	-0.3772	-1.3821	0.922799	-0.11437	-1.30807	0.786372	1.034006
1.767854	-0.00445	-0.03987	-0.3772	-1.3821	0.922799	-0.11437	-1.30807	0.786372	1.034006
1.767854	-0.00445	-0.03987	-0.3772	-1.3821	0.922799	-0.11437	-1.30807	0.786372	1.034006
1.767854	-0.00445	-0.03987	-0.3772	-1.3821	0.922799	-0.11437	-1.30807	0.786372	1.034006
1.767854	-0.00445	-0.03987	-0.3772	-1.3821	0.922799	-0.11437	-1.30807	0.786372	1.034006
1.767854	-0.00445	-0.03987	-0.3772	-1.3821	0.922799	-0.11437	-1.30807	0.786372	1.034006
1.767854	-0.00445	-0.03987	-0.3772	-1.3821	0.922799	-0.11437	-1.30807	0.786372	1.034006
1.767854	-0.00445	-0.03987	-0.3772	-1.3821	0.922799	-0.11437	-1.30807	0.786372	1.034006
1.767854	-0.00445	-0.03987	-0.3772	-1.3821	0.922799	-0.11437	-1.30807	0.786372	1.034006
1.767854	-0.00445	-0.03987	-0.3772	-1.3821	0.922799	-0.11437	-1.30807	0.786372	1.034006
1.767854	-0.00445	-0.03987	-0.3772	-1.3821	0.922799	-0.11437	-1.30807	0.786372	1.034006
1.767854	-0.00445	-0.03987	-0.3772	-1.3821	0.922799	-0.11437	-1.30807	0.786372	1.034006
1.767854	-0.00445	-0.03987	-0.3772	-1.3821	0.922799	-0.11437	-1.30807	0.786372	1.034006
1.767854	-0.00445	-0.03987	-0.3772	-1.3821	0.922799	-0.11437	-1.30807	0.786372	1.034006
1.767854	-0.00445	-0.03987	-0.3772	-1.3821	0.922799	-0.11437	-1.30807	0.786372	1.034006
1.767854	-0.00445	-0.03987	-0.3772	-1.3821	0.922799	-0.11437	-1.30807	0.786372	1.034006
1.767854	-0.00445	-0.03987	-0.3772	-1.3821	0.922799	-0.11437	-1.30807	0.786372	1.034006
1.767854	-0.00445	-0.03987	-0.3772	-1.3821	0.922799	-0.11437	-1.30807	0.786372	1.034006
1.767854	-0.00445	-0.03987	-0.3772	-1.3821	0.922799	-0.11437	-1.30807	0.786372	1.034006

Nurse-level covariate doesn't change value within a column (same nurse), only across columns `h1` to `h25`

SAS syntax: nursedat2.sas

```
PROC IMPORT OUT= nursedat2 DATAFILE="C:\temp\nursedat2.xlsx"
            DBMS=xlsx;
            GETNAMES=YES;
RUN;

PROC MEANS DATA=nursedat2;
RUN;

DATA ndats;
SET nursedat2;
cons = 1;
ARRAY ps(25)  p1-p25;
ARRAY hs(25)  h1-h25;
     happiness = 0;
     DO i = 1 TO 25;
        happiness = happiness + ps(i)*hs(i);
     END;
```

```
PROC MIXED  DATA=ndats COVTEST;
  CLASS cons;
  MODEL satis =  / SOLUTION;
  RANDOM p1-p25 / SUBJECT=cons TYPE=TOEPLITZ(1);

PROC MIXED  DATA=ndats COVTEST;
  CLASS cons;
  MODEL satis = assess / SOLUTION;
  RANDOM p1-p25 / SUBJECT=cons TYPE=TOEPLITZ(1);

PROC MIXED  DATA=ndats COVTEST;
  CLASS cons;
  MODEL satis = assess happiness / SOLUTION;
  RANDOM p1-p25 / SUBJECT=cons TYPE=TOEPLITZ(1);
RUN;
```

- treats all observations as being nested within one cluster (`SUBJECT= cons`)
- indicates the 25 proportion variables as random effects (`RANDOM p1-p25`)
- 25 random effects are uncorrelated and have the same variance (`TYPE=TOEPLITZ(1)`)

The Mixed Procedure

Model Information

Dependent Variable	satis
Covariance Structure	Banded Toeplitz
Subject Effect	cons
Estimation Method	REML

Class Level Information

Class	Levels	Values
cons	1	1

Dimensions

Covariance Parameters	2
Columns in X	3
Columns in Z per Subject	25
Subjects	1
Max Obs per Subject	1000

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
Variance	cons	0.1858	0.06142	3.03	0.0012
Residual		0.5880	0.02664	22.07	<.0001

Fit Statistics

-2 Res Log Likelihood	2371.0
AIC (Smaller is Better)	2375.0
AICC (Smaller is Better)	2375.0
BIC (Smaller is Better)	2371.0

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	-0.03124	0.08965	974	-0.35	0.7276
assess	0.4913	0.02531	974	19.41	<.0001
happiness	0.2938	0.09084	974	3.23	0.0013

Multi-Membership Multilevel Model

Patient j nested within nurse(s) i $(i = 1, \ldots, N)$

Model with only patient covariates

$$y_j = \boldsymbol{x}_j'\boldsymbol{\beta} + \sum_{i=1}^{N} p_{ij}\upsilon_i + \epsilon_j$$

$$\upsilon_i \sim N(0, \sigma_\upsilon^2) \qquad \epsilon_{ij} \sim N(0, \sigma_\epsilon^2) \qquad \sum_{i=1}^{N} p_{ij} = 1 \; \forall \; j$$

Model with a nurse covariate x_i

$$y_j = \boldsymbol{x}_j'\boldsymbol{\beta} + \left(\sum_{i=1}^{N} p_{ij}x_i\right)\beta_{x_i} + \sum_{i=1}^{N} p_{ij}\upsilon_i + \epsilon_j$$

Model with $N \times p$ nurse covariate matrix $\boldsymbol{X}$ and $N \times 1$ nurse weight vector $\boldsymbol{p}_j$ for patient j

$$y_j = \boldsymbol{x}_j'\boldsymbol{\beta} + \boldsymbol{p}_j'\boldsymbol{X}\boldsymbol{\beta}_X + \boldsymbol{p}_j'\upsilon_i + \epsilon_j$$

Intraclass correlation - Leckie (2013)

- pairwise correlation of patients DV within a given cluster (conditional on covariates)
- ICCs vary depending on patient weights

Correlation for two patients cared for by the same nurse for the entire hospital stay (null model: nurse var = .2439, error var = .8144)

$$ICC = \frac{\sigma_\upsilon^2}{\sigma_\upsilon^2 + \sigma_\epsilon^2} = \frac{.2439}{.2439 + .8144} = 0.230$$

Correlation for two patients j and j' cared for by the same nurse, say nurse 2, for half their stays ($p_{2j} = p_{2j'} = .5$), but different nurses for the rest of their stays, say $p_{5j} = .5$ and $p_{8j'} = .5$

$$ICC = \frac{p_{2j}p_{2j'}\sigma_\upsilon^2}{\sqrt{\left(p_{2j}^2 + p_{5j}^2\right)\sigma_\upsilon^2 + \sigma_\epsilon^2}\sqrt{\left(p_{2j'}^2 + p_{8j'}^2\right)\sigma_\upsilon^2 + \sigma_\epsilon^2}} = \frac{.25 \times .2439}{.5 \times .2439 + .8144} = 0.065$$

Longitudinal and Clustered - Cafri, Hedeker, & Aarons (2015)

Hierarchical Structure

	Class 1	Class 2	Class 3	Class 4
Subject 1	$y_{111}, y_{211}, y_{311}$			
Subject 2			$y_{123}, y_{223}, y_{323}$	

y_{ijk} = repeated observation i nested within subject j within class k

Crossed Structure

	Class 1	Class 2	Class 3	Class 4
Subject 1	y_{111}	y_{212}		y_{314}
Subject 2	y_{321}		y_{123}, y_{223}	

y_{ijk} = repeated obs i nested within subject j, crossed with class k

$\Rightarrow$ subjects can be in different classes at different timepoints

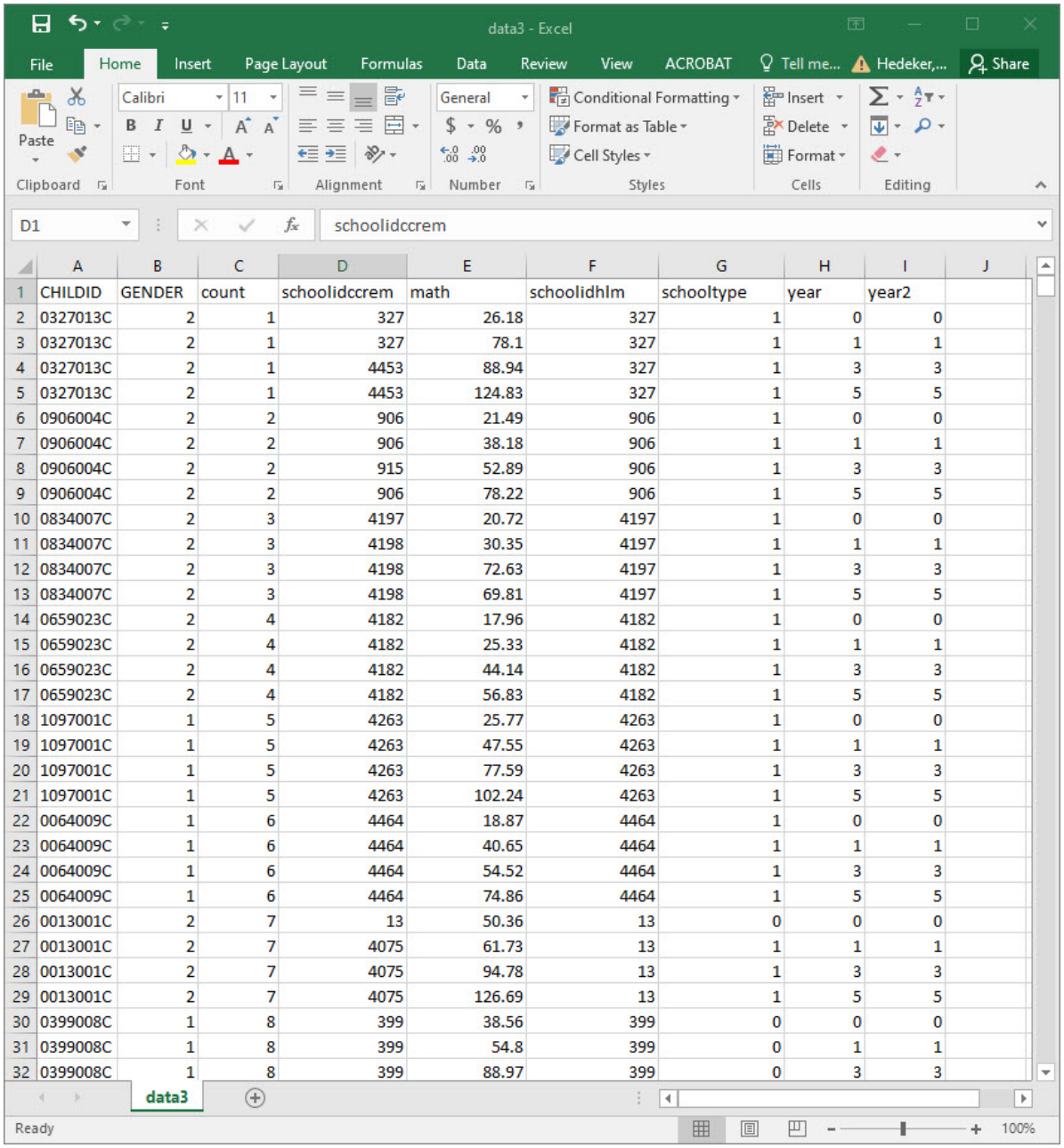
Acute-Effects Cross-Classified Multilevel Models

- Repeated observations within subjects, with subjects in (potentially) different clusters at different timepoints
  - Subjects and clusters are crossed random effects
- Add random cluster effects to longitudinal multilevel model
  - Multiple random subject effects usually (e.g., random intercept and time effects)
  - Random intercept for clusters
- Cluster effect for a given subject and timepoint, depends on which cluster a subject is in at that timepoint

Example: Cafri, Hedeker, & Aarons (2015)

- 500 children (a subset) from Early Childhood Longitudinal Study–Kindergarten Cohort
- Math achievement is the outcome; student gender and school type (public vs. private) are predictors
- Outcome is measured at four timepoints: kindergarten, first grade, third grade, fifth grade
- A total of 245 schools
  - between kindergarten and first grade, 4.8% of the children changed schools
  - between first and third grade, 8.6% changed schools
  - between third and fifth grade, 8.0% changed schools

CSV file: data3.csv - Math achievement in 500 children over time



CHILDID	GENDER	count	schoolidccrem	math	schoolidhlm	schooltype	year	year2
0327013C	2	1	327	26.18	327	1	0	0
0327013C	2	1	327	78.1	327	1	1	1
0327013C	2	1	4453	88.94	327	1	3	3
0327013C	2	1	4453	124.83	327	1	5	5
0906004C	2	2	906	21.49	906	1	0	0
0906004C	2	2	906	38.18	906	1	1	1
0906004C	2	2	915	52.89	906	1	3	3
0906004C	2	2	906	78.22	906	1	5	5
0834007C	2	3	4197	20.72	4197	1	0	0
0834007C	2	3	4198	30.35	4197	1	1	1
0834007C	2	3	4198	72.63	4197	1	3	3
0834007C	2	3	4198	69.81	4197	1	5	5
0659023C	2	4	4182	17.96	4182	1	0	0
0659023C	2	4	4182	25.33	4182	1	1	1
0659023C	2	4	4182	44.14	4182	1	3	3
0659023C	2	4	4182	56.83	4182	1	5	5
1097001C	1	5	4263	25.77	4263	1	0	0
1097001C	1	5	4263	47.55	4263	1	1	1
1097001C	1	5	4263	77.59	4263	1	3	3
1097001C	1	5	4263	102.24	4263	1	5	5
0064009C	1	6	4464	18.87	4464	1	0	0
0064009C	1	6	4464	40.65	4464	1	1	1
0064009C	1	6	4464	54.52	4464	1	3	3
0064009C	1	6	4464	74.86	4464	1	5	5
0013001C	2	7	13	50.36	13	0	0	0
0013001C	2	7	4075	61.73	13	1	1	1
0013001C	2	7	4075	94.78	13	1	3	3
0013001C	2	7	4075	126.69	13	1	5	5
0399008C	1	8	399	38.56	399	0	0	0
0399008C	1	8	399	54.8	399	0	1	1
0399008C	1	8	399	88.97	399	0	3	3

- CHILDID: child id variable
- GENDER: child's gender (1=male, 2=female); after reading the CSV file this is recoded to genderf (0=male, 1=female)
- count: sequential count of 500 children
- schoolidccrem: school id for child at each timepoint
- math: math achievement (outcome)
- schoolidhlm: school id for child's school at kindergarten
- schooltype: 0=private, 1=public
- year: year of school (0=kindergarten, 1=first grade, 3=third grade, 5=fifth grade)
- year2: duplicate of year variable

SAS syntax: `Example1_mixed.sas`
Acute-Effects Cross-Classified Multilevel Model

```
PROC MIXED;
CLASS childid schoolidccrem ;
MODEL math = year genderf schooltype genderf*year/S;
RANDOM INTERCEPT year/SUB=childid TYPE=UN;
RANDOM INTERCEPT /SUB=schoolidccrem;
RUN;
```

- predictors are year, genderf, schooltype, and genderf by year interaction
- random subject intercept and year effects (which are allowed to be correlated)
- random school effect (i.e., id variable **`schoolidccrem`** indicates which school a child is in at each timepoint); a hierarchical three-level model using **`schoolidhlm`** would only consider clustering of first school at kindergarten

```
-2 Res Log Likelihood         15305.8
```

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	37.0368	1.3843	88	26.75	<.0001
year	15.7188	0.2182	498	72.04	<.0001
genderf	-0.8006	0.9273	911	-0.86	0.3881
schooltype	-5.5011	1.4327	911	-3.84	0.0001
year*genderf	-0.4627	0.3055	911	-1.51	0.1302

- Mean math score is approximately 37 (in kindergarten for males in private schools)
- Math scores increasing 15.72 points per year ($p < .0001$)
- Public schools are lower by approximately 5.5 points ($p < .0001$)
- No significant gender or gender by time interaction

```
    Covariance Parameter Estimates

Cov Parm      Subject          Estimate

UN(1,1)       CHILDID           58.4328
UN(2,1)       CHILDID           14.6621
UN(2,2)       CHILDID            7.7330
Intercept     schoolidccrem     27.0622
Residual                        55.0195
```

- Comparing this model to simpler 2-level model without the clustering of schools (not shown) yields:

  $\chi^2_1 = 15343.7 - 15305.8 = 37.9$; highly significant
- Appreciable clustering of math scores attributable to schools
- Schools vary in terms of higher/lower average math scores

Dynamic Group Multilevel Models

- So far, subjects have been considered either in the same group (cluster) across time (hierarchical), or in different groups (clusters) at different timepoints (crossed)
- An implicit assumption of both is that the cluster effects are equal across time and perfectly correlated (a dubious assumption)
- Dynamic group (cluster) multilevel models (Bauer et al., 2013) relax this assumption, allowing different variance and correlation structures for the group (cluster) effects over time
- Many structures are possible including unstructured, toeplitz, first-order autoregressive; most general is unstructured, which allows variances and correlations across timepoints to all be different

SAS syntax: `Example1_mixed.sas`

Acute-Effects Cross-Classified and Dynamic Group Models

Acute-Effects Cross-Classified Multilevel Model

```
PROC MIXED;
CLASS childid schoolidccrem ;
MODEL math = year genderf schooltype genderf*year/S;
RANDOM INTERCEPT year/SUB=childid TYPE=UN;
RANDOM INTERCEPT /SUB=schoolidccrem;
```

Dynamic Group Multilevel Model

```
PROC MIXED;
CLASS childid schoolidccrem year2;
MODEL math = year genderf schooltype genderf*year/S;
RANDOM INTERCEPT year/SUB=childid TYPE=UN;
RANDOM year2 /SUB=schoolidccrem TYPE=UN;
```

- `year2` is a duplicate of `year`, but former needs to be on the `CLASS` and `RANDOM` statements
- `TYPE=UN` specifies the unstructured variance-covariance matrix (for the school effects across the four timepoints)

SAS: Dynamic Group Multilevel Model Output

```
          Fit Statistics

-2 Res Log Likelihood         15129.4

                    Solution for Fixed Effects

                              Standard
Effect          Estimate       Error      DF    t Value    Pr > |t|

Intercept        34.9419      1.3074     385      26.73      <.0001
year             15.7547      0.2307     498      68.29      <.0001
genderf          -0.8928      0.9102     614      -0.98      0.3270
schooltype       -6.0861      1.3517     614      -4.50      <.0001
year*genderf     -0.3667      0.2923     614      -1.25      0.2101
```

- Mean math score is approximately 35 (in kindergarten for males in private schools)
- Math scores increasing 15.75 points per year ($p < .0001$)
- Public schools are lower by approximately 6 points ($p < .0001$)
- No significant gender or gender by time interaction

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
UN(1,1)	CHILDID	67.2172
UN(2,1)	CHILDID	12.6890
UN(2,2)	CHILDID	6.7432
UN(1,1)	schoolidccrem	17.3132
UN(2,1)	schoolidccrem	23.7342
UN(2,2)	schoolidccrem	81.4858
UN(3,1)	schoolidccrem	12.0855
UN(3,2)	schoolidccrem	53.3208
UN(3,3)	schoolidccrem	56.7376
UN(4,1)	schoolidccrem	8.9422
UN(4,2)	schoolidccrem	26.3796
UN(4,3)	schoolidccrem	40.8660
UN(4,4)	schoolidccrem	45.7929
Residual		37.6731

- School clustering effect is largest at first grade (estimate = 81.49), smallest at kindergarten (estimate = 17.31), and intermediate at third and fifth grades (estimates = 56.74 and 45.79, respectively)

Comparing Models

- Acute-effects model assumes school effects are constant over time, dynamic-effects model relaxes this assumption
- Can use likelihood-ratio test to compare models and test this assumption
- Here, acute-effects model includes 5 variance-covariance parameters, whereas dynamic-effects UN model has 14 parameters; degrees of freedom for this test is 9
- $\chi^2_9 = 15305.8 - 15129.4 = 176.4$; which is highly significant
- Acute-effects model is rejected in favor of the more general dynamic-effects UN model
- Variance of school effects varies over time, and school effects are differentially correlated over time

Expressing Covariance Matrix as a Correlation Matrix

```
/* get correlation matrix of school random effects over time */
data covs;
input t1 t2 t3 t4;
datalines;
17.3132 23.7342 12.0855 8.9422
23.7342 81.4858 53.3208 26.3796
12.0855 53.3208 56.7376 40.8660
8.9422  26.3796 40.8660 45.7929
;
run;

proc iml;
use covs;
read all var{t1 t2 t3 t4} into cov ;
print cov;
corr = inv(sqrt(diag(cov)))*cov*inv(sqrt(diag(cov)));
print corr;
run;
quit;
```

```
cov

  17.3132   23.7342   12.0855    8.9422
  23.7342   81.4858   53.3208   26.3796
  12.0855   53.3208   56.7376    40.866
   8.9422   26.3796    40.866   45.7929

corr

        1 0.6318951 0.3856031 0.3175826
0.6318951         1  0.784188 0.4318453
0.3856031  0.784188         1  0.801729
0.3175826 0.4318453  0.801729         1
```

- School effects are moderately to highly correlated (but not really equal)
- Correlation of school effects diminishes as time interval increases, and increases within a lag

Summary: Focus on non-hierarchical multilevel models

- Cross-classified models: when subjects are classified in two or more cluster types, but cluster types are not nested within each other
  - Students within primary and secondary schools

- Multi-membership models: when subjects are clustered within potentially multiple clusters (at the same level)
  - Patients seen by multiple nurses

- Cross-classified longitudinal models: when subjects are measured over time and (potentially) in different clusters across time
  - Acute effects: cluster effect does not vary over time: assumes cluster effects are perfectly correlated over time
  - Dynamic effects: cluster effect can vary over time; allows for different variance and correlation structures

- Can be computationally demanding because of specification of many random effects; SAS HPMIXED is potentially useful for this

References

- Bauer, D. J., Gottfredson, N. C., Dean, D., & Zucker, R. A. (2013). Analyzing repeated measures data on individuals nested within groups: Accounting for dynamic group effects. Psychological Methods, 18, 1–14.
- Beretvas, S. N. (2010). Cross-classified and multiple membership models. In J. J. Hox & J. K. Roberts (Eds.), Handbook of advanced multilevel analysis. Texas: Psychology Press.
- Beretvas, S. N., & Murphy, D. L. (2013). An evaluation of information criteria use for correct cross-classified random effects model selection. Journal of Experimental Education, 81, 429–463.
- Cafri G, Hedeker D, & Aarons GA (2015). An introduction and integration of cross-classified, multiple membership, and dynamic group random-effects models. Psychological Methods, 20(4), 407-421.
- Chung, H., & Beretvas, S. N. (2012). The impact of ignoring multiple membership data structures in multilevel models. British Journal of Mathematical and Statistical Psychology, 65, 185–200
- Grady, M. W., & Beretvas, S. N. (2010). Incorporating student mobility in achievement growth modeling: A cross-classified multiple membership growth curve model. Multivariate Behavioral Research, 45, 393–419.
- Hill, P. W., & Goldstein, H. (1998). Multilevel modeling of educational data with cross-classification and missing identification units. Journal of Educational and Behavioral Statistics, 23, 117–128.
- Leckie, G. (2013). Cross-classified multilevel models. LEMMA VLE Module 12, 1–60
- Luo, W., & Kwok, O. (2009). The impacts of misspecifying crossclassified random effects models. Multivariate Behavioral Research, 44, 182–212.
- Meyers, J. L., & Beretvas, S. N. (2006). The impact of inappropriate modeling of cross-classified data structures. Multivariate Behavioral Research, 41, 473–497.

- Palardy, G. J. (2010). The multilevel crossed random effects growth model for estimating teacher and school effects: Issues and extensions. Educational and Psychological Measurement, 70, 401–419.
- Rasbash, J., & Browne, W. (2008). Non-hierarchical multilevel models. In J. De Leeuw & E. Meijer (Eds.), Handbook of quantitative multilevel analysis (pp. 301–334). New York, NY: Springer.
- Rasbash, J., & Goldstein, H. (1994). Efficient analysis of mixed hierarchical and cross-classified random structures using a multilevel model. Journal of Educational and Behavioral Statistics, 19(4), 337–350.
- Wolff Smith, L. J., & Beretvas, S. N. (2014). The impact of using incorrect weights with the multiple membership random effects model. Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 10, 31–42.
- Raudenbush, S. W. (1993). A crossed random effects model for unbalanced data with applications in cross-sectional and longitudinal research. Journal of Educational Statistics, 18, 321–349.
- Shi, Y., Leite, W., & Algina, J. (2010). The impact of omitting the interaction between crossed factors in cross-classified random effects modelling. British Journal of Mathematical and Statistical Psychology, 63(1), 1–15.

Appendix: SPSS examples

SPSS syntax after reading in Excel file:
`Fife_CrossClassified.sps`

Null Model

```
MIXED
  attain BY pid sid
  /PRINT = SOLUTION TESTCOV
  /RANDOM pid sid .
```

Model with `vrq` as covariate

```
MIXED
  attain WITH vrq BY pid sid
  /FIXED = vrq
  /PRINT = SOLUTION TESTCOV
  /RANDOM pid sid .
```

Crossed random effects with interaction

SPSS

```
MIXED
  attain BY pid sid
  /PRINT = SOLUTION TESTCOV
  /RANDOM pid sid pid*sid .
```

SPSS syntax: nursedat2.sps

```
GET FILE='C:\MixDemo\nursedat2.sav'.

COMPUTE cons=1.

COMPUTE happy = 0.
VECTOR ps = p1 TO p25 / hs = h1 TO h25.
LOOP i = 1 TO 25.
+ COMPUTE happy = happy + ps(i) * hs(i).
END LOOP.

MIXED
  satis WITH assess happy p1 TO p25
  /FIXED = assess happy
  /PRINT = SOLUTION TESTCOV
  /RANDOM  p1 p2 p3 p4 p5 p6 p7 p8 p9 p10 p11 p12
           p13 p14 p15 p16 p17 p18 p19 p20 p21 p22
           p23 p24 p25  | SUBJECT(cons) COVTYPE(ID).
```

SPSS syntax: `Example1_mixed.sps`

Acute-Effects Cross-Classified Multilevel Model

```
MIXED
  math WITH year genderf schooltype
  /FIXED = year genderf schooltype genderf*year
  /PRINT = SOLUTION
  /RANDOM INTERCEPT year  | SUBJECT(childid) COVTYPE(UN)
  /RANDOM INTERCEPT | SUBJECT(schoolidccrem).
```

- predictors are year, genderf, schooltype, and genderf by year interaction
- random subject intercept and year effects (which are allowed to be correlated)
- random school effect (i.e., id variable **`schoolidccrem`** indicates which school a child is in at each timepoint)

Appendix: Stata examples

- SAS and SPSS perform cross-classified analyses seamlessly, but not all software can

- Some hierarchical multilevel software programs (e.g., Stata) can be "tricked" into running cross-classified models if they allow
  - 3-level models
  - Equality constraints on variances of random effects
  - Zero covariances of random effects
  - described in: Rasbash & Goldstein (1994). Efficient Analysis of Mixed Hierarchical and Cross-Classified Random Structures Using a Multilevel Model. *Journal of Educational and Behavioral Statistics, 19(4)*:337-350.

Alternative way to run cross-classified multilevel models

- Identify cluster level with fewest number of clusters; here, 148 primary schools and 19 secondary schools
- Create indicator variables for the secondary schools, `secs1-secs19` (0/1), which indicate the secondary school that a student belongs to (each student belongs to only one)
- Create a variable `cons` that equals 1 for all observations in the dataset
- At the third level, specify `cons` as the level-3 ID variable, and the 19 indicator variables `secs1-secs19` as random effects with EQUAL variance and zero covariances
- At the second level, specify the primary school ID nested within the level-3 ID (`cons`) and specify a random intercept

Secondary School Indicator Variables

ID	secs1	secs2	secs3	secs4	secs5	secs6	secs7	...	secs19
sid = 1	1	0	0	0	0	0	0	...	0
sid = 2	0	1	0	0	0	0	0	...	0
sid = 3	0	0	1	0	0	0	0	...	0
sid = 4	0	0	0	1	0	0	0	...	0
sid = 5	0	0	0	0	1	0	0	...	0
sid = 6	0	0	0	0	0	1	0	...	0
sid = 7	0	0	0	0	0	0	1	...	0
...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...
sid = 19	0	0	0	0	0	0	0	...	1

Stata for cross-classified multilevel

- Stata uses this 3-level approach for cross-classified multilevel models
- It creates the school indicator variables, so you don't have to
- Uses ML estimation by default
- remember that Stata is case-sensitive

Stata syntax for cross-classified: `fife_crossclassified.do`

```
log using u:\Stata_Examples\fife_crossclassified.log
import excel "u:\Stata_Examples\xwdata_9var.xlsx",  ///
sheet("Sheet1") firstrow
summarize
mixed ATTAIN || _all: R.SID || PID:, reml
```

From me.pdf:

- `_all` identifes all the observations as one big group and `R.SID` tells `mixed` to treat `SID` as a factor variable (or equivalently, as a set of overparameterized indicator variables identifying `SID` groups)
- The `R.varname` notation is equivalent to giving a list of overparameterized (none dropped) indicator variables for use in a random-effects specification. When you use `R.varname`, mixed-effects commands handle the calculations internally rather than creating the indicators in the data. Because the set of indicators is overparameterized, `R.varname` implies `noconstant`

```
. import excel "u:\Stata_Examples\xwdata_9var.xlsx", ///
> sheet("Sheet1") firstrow
(9 vars, 3,435 obs)

. summarize

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
         VRQ |      3,435    97.80437    13.29291         70        140
      ATTAIN |      3,435    5.678603    3.058504          1         10
         PID |      3,435     70.7377    45.02572          1        148
        SEXF |      3,435    .4937409    .5000336          0          1
          SC |      3,435    6.844833    10.88761          0         31
-------------+---------------------------------------------------------
         SID |      3,435    10.21951     5.55694          1         19
         FED |      3,435    .2754003    .4467808          0          1
      CHOICE |      3,435    1.195342    .6502375          1          4
         MED |      3,435     .342067     .474471          0          1
```

```
. mixed ATTAIN || _all: R.SID || PID:, reml


Mixed-effects REML regression                   Number of obs     =      3,435

        Grouping information
        -------------------------------------------------------------
                        |     No. of       Observations per group
         Group variable |     groups    Minimum    Average    Maximum
        ----------------+--------------------------------------------
                   _all |          1      3,435    3,435.0      3,435
                    PID |        148          1       23.2         72
        -------------------------------------------------------------

                                                Wald chi2(0)      =          .
Log restricted-likelihood = -8575.3795          Prob > chi2       =          .

------------------------------------------------------------------------------
      ATTAIN | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
       _cons |   5.501727   .1786809    30.79   0.000     5.151519    5.851935
------------------------------------------------------------------------------
```

```
------------------------------------------------------------------------------
  Random-effects parameters  |   Estimate   Std. err.     [95% conf. interval]
-----------------------------+------------------------------------------------
_all: Identity               |
                  var(R.SID) |   .3722272   .1743284      .1486475    .9320919
-----------------------------+------------------------------------------------
PID: Identity                |
                  var(_cons) |   1.130027   .2073717      .7886506    1.619173
-----------------------------+------------------------------------------------
               var(Residual) |   8.110684   .2004489      7.727175    8.513228
------------------------------------------------------------------------------
LR test vs. linear model: chi2(2) = 280.57                Prob > chi2 = 0.0000

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

Crossed random effects with interaction

```
. mixed ATTAIN || _all: R.SID || PID: || SID:, reml

Mixed-effects REML regression                   Number of obs     =      3,435

        Grouping information
        -------------------------------------------------------------
                        |     No. of       Observations per group
         Group variable |     groups    Minimum    Average    Maximum
        ----------------+--------------------------------------------
                   _all |          1      3,435    3,435.0      3,435
                    PID |        148          1       23.2         72
                    SID |        303          1       11.3         72
        -------------------------------------------------------------

                                                Wald chi2(0)      =          .
Log restricted-likelihood = -8574.8294          Prob > chi2       =          .

------------------------------------------------------------------------------
      ATTAIN | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
       _cons |    5.49875   .1731218    31.76   0.000     5.159438    5.838063
```

```
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects parameters  |   Estimate   Std. err.     [95% conf. interval]
-----------------------------+------------------------------------------------
_all: Identity               |
                  var(R.SID) |   .3384038   .1726538       .124495    .9198535
-----------------------------+------------------------------------------------
PID: Identity                |
                  var(_cons) |   .9146271   .2954226      .4856325    1.722584
-----------------------------+------------------------------------------------
SID: Identity                |
                  var(_cons) |   .2334204   .2478765       .029122    1.870926
-----------------------------+------------------------------------------------
               var(Residual) |   8.087217   .2011691       7.70239     8.49127
------------------------------------------------------------------------------
LR test vs. linear model: chi2(3) = 281.67                Prob > chi2 = 0.0000

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

Stata syntax: `nursedat2.do`

```
log using u:\Stata_Examples\nursedat2.log
import excel "u:\Stata_Examples\nursedat2.xlsx", sheet("Sheet1") firstrow
summarize

forvalues j = 1/25 {
          generate p`j'Xh`j' = p`j'*h`j'
  }
egen happiness = rsum(p1Xh1-p25Xh25)

mixed satis  || _all: p1-p25, nocons covariance(identity) reml
mixed satis assess || _all: p1-p25, nocons covariance(identity) reml
mixed satis assess happiness || _all: p1-p25, nocons covariance(identity) reml
log close
```

- must use forward and backward quotation marks in the `generate` statement
- `rsum` = row sum of the variables `p1Xh1-p25Xh25`
- `covariance(identity)` = equal variances for the random effects, all covariances are zero

```
-------------------------------------------------------------------------------------------
      name:  <unnamed>
       log:  u:\Stata_Examples\nursedat2.log
  log type:  text

. import excel "u:\Stata_Examples\nursedat2.xlsx", sheet("Sheet1") firstrow

. summarize

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
     patient |      1000       500.5    288.8194          1       1000
       satis |      1000   -.0387974    .9809753  -2.981966   3.040571
      assess |      1000    .0204101    .9700845    -2.9099    3.22795
      nurses |      1000           2      1.0005          1          4
        n1st |      1000      12.814    7.052314          1         25
-------------+--------------------------------------------------------
        n2nd |      1000       7.811    8.542614          0         25
        n3rd |      1000       3.751    6.937227          0         25
        n4th |      1000       1.264    4.532863          0         25
          p1 |      1000      .04428    .1633541          0          1
          p2 |      1000      .04878     .178631          0          1
-------------+--------------------------------------------------------
          p3 |      1000      .03483    .1418933          0          1
          p4 |      1000      .03641    .1521412          0          1
          p5 |      1000      .04105    .1640424          0          1
          p6 |      1000      .03865     .146819          0          1
          p7 |      1000      .03965    .1549309          0          1
```

p8	1000	.03734	.1517833	0	1
p9	1000	.0429	.1720581	0	1
p10	1000	.04406	.1767704	0	1
p11	1000	.03236	.140788	0	1
p12	1000	.04978	.1763305	0	1
p13	1000	.04874	.1728985	0	1
p14	1000	.04426	.1630725	0	1
p15	1000	.04346	.1710687	0	1
p16	1000	.03532	.1456603	0	1
p17	1000	.02987	.1288965	0	1
p18	1000	.04052	.162613	0	1
p19	1000	.03883	.154426	0	1
p20	1000	.03279	.1439258	0	1
p21	1000	.04696	.170115	0	1
p22	1000	.04011	.1623559	0	1
p23	1000	.03768	.1503983	0	1
p24	1000	.03576	.1463942	0	1
p25	1000	.03561	.1463354	0	1
h1	1000	1.767854	0	1.767854	1.767854
h2	1000	-.0044529	0	-.0044529	-.0044529
h3	1000	-.0398703	0	-.0398703	-.0398703
h4	1000	-.3772039	0	-.3772039	-.3772039
h5	1000	-1.382097	0	-1.382097	-1.382097

h6	1000	.9227986	0	.9227986	.9227986
h7	1000	-.1143707	0	-.1143707	-.1143707
h8	1000	-1.308072	0	-1.308072	-1.308072
h9	1000	.7863722	0	.7863722	.7863722
h10	1000	1.034006	0	1.034006	1.034006
h11	1000	1.563056	0	1.563056	1.563056
h12	1000	-1.421343	0	-1.421343	-1.421343
h13	1000	.5874567	0	.5874567	.5874567
h14	1000	-.3803294	0	-.3803294	-.3803294
h15	1000	1.291649	0	1.291649	1.291649
h16	1000	-1.985823	0	-1.985823	-1.985823
h17	1000	1.049627	0	1.049627	1.049627
h18	1000	-1.010235	0	-1.010235	-1.010235
h19	1000	-.7836885	0	-.7836885	-.7836885
h20	1000	1.097389	0	1.097389	1.097389
h21	1000	-.9833028	0	-.9833028	-.9833028
h22	1000	-.2067423	0	-.2067423	-.2067423
h23	1000	-.6376706	0	-.6376706	-.6376706
h24	1000	.1376409	0	.1376409	.1376409
h25	1000	.253908	0	.253908	.253908

```
. mixed satis  || _all: p1-p25, nocons covariance(identity) reml

Mixed-effects REML regression                   Number of obs     =      1,000
Group variable: _all                            Number of groups  =          1
                                                Obs per group:
                                                              min =      1,000
                                                              avg =    1,000.0
                                                              max =      1,000
                                                Wald chi2(0)      =          .
Log restricted-likelihood =  -1344.359          Prob > chi2       =          .

------------------------------------------------------------------------------
       satis | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
       _cons |  -.0264564   .1029095    -0.26   0.797    -.2281553    .1752425
------------------------------------------------------------------------------


------------------------------------------------------------------------------
  Random-effects parameters  |   Estimate   Std. err.     [95% conf. interval]
-----------------------------+------------------------------------------------
_all: Identity               |
             var(p1..p25)(1) |   .2438996   .0799165      .1283232    .4635718
-----------------------------+------------------------------------------------
               var(Residual) |   .8144145   .0368878      .7452319    .8900196
------------------------------------------------------------------------------
LR test vs. linear model: chibar2(01) = 114.85        Prob >= chibar2 = 0.0000
(1) p1 p2 p3 p4 p5 p6 p7 p8 p9 p10 p11 p12 p13 p14 p15 p16 p17 p18 p19 p20 p21 p22 p23 p24 p25
```

```
. mixed satis assess || _all: p1-p25, nocons covariance(identity) reml

Mixed-effects REML regression                   Number of obs     =      1,000
Group variable: _all                            Number of groups  =          1
                                                Obs per group:
                                                              min =      1,000
                                                              avg =    1,000.0
                                                              max =      1,000
                                                Wald chi2(1)      =     375.00
Log restricted-likelihood = -1188.5094          Prob > chi2       =     0.0000

------------------------------------------------------------------------------
       satis | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
      assess |   .4904096   .0253247    19.36   0.000      .440774    .5400451
       _cons |  -.0331122   .1062326    -0.31   0.755    -.2413243    .1750999
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects parameters  |   Estimate   Std. err.     [95% conf. interval]
-----------------------------+------------------------------------------------
_all: Identity               |
             var(p1..p25)(1) |   .2670523   .0840303      .1441314     .494805
-----------------------------+------------------------------------------------
               var(Residual) |   .5881106   .0266514      .5381275    .6427363
------------------------------------------------------------------------------
LR test vs. linear model: chibar2(01) = 180.45        Prob >= chibar2 = 0.0000
(1) p1 p2 p3 p4 p5 p6 p7 p8 p9 p10 p11 p12 p13 p14 p15 p16 p17 p18 p19 p20 p21 p22 p23 p24 p25
```

```
. mixed satis assess happiness || _all: p1-p25, nocons covariance(identity) reml

Mixed-effects REML regression                   Number of obs     =      1,000
Group variable: _all                            Number of groups  =          1
                                                Obs per group:
                                                              min =      1,000
                                                              avg =    1,000.0
                                                              max =      1,000
                                                Wald chi2(2)      =     385.08
Log restricted-likelihood = -1185.4774          Prob > chi2       =     0.0000

------------------------------------------------------------------------------
       satis | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
      assess |   .4912597   .0253147    19.41   0.000     .4416437    .5408757
   happiness |   .2937832   .0908374     3.23   0.001     .1157451    .4718213
       _cons |  -.0312412    .089651    -0.35   0.727    -.2069539    .1444715
------------------------------------------------------------------------------


------------------------------------------------------------------------------
  Random-effects parameters  |   Estimate   Std. err.     [95% conf. interval]
-----------------------------+------------------------------------------------
_all: Identity               |
             var(p1..p25)(1) |   .1858517   .0614233      .0972406    .3552102
-----------------------------+------------------------------------------------
               var(Residual) |   .5880194   .0266431      .5380514    .6426278
------------------------------------------------------------------------------
LR test vs. linear model: chibar2(01) = 125.61        Prob >= chibar2 = 0.0000
(1) p1 p2 p3 p4 p5 p6 p7 p8 p9 p10 p11 p12 p13 p14 p15 p16 p17 p18 p19 p20 p21 p22 p23 p24 p25
```

Stata syntax: `Example1_mixed.do`

Acute-Effects Cross-Classified Multilevel Model

```
mixed math year genderf schooltype c.genderf#c.year ///
    || _all: R.schoolidccrem || childid: year, covariance(un) reml
```

- predictors are year, genderf, schooltype, and genderf by year interaction
- random subject intercept and year effects (which are allowed to be correlated)
- random school effect (i.e., id variable **`schoolidccrem`** indicates which school a child is in at each timepoint)
  - uses 3-level trick for the cross-classified random school effect; **`_all`** identifes all the observations as one big group and **`R.schoolidccrem`** tells **`mixed`** to treat **`schoolidccrem`** as a factor variable (i.e., as a set of indicator variables identifying **`schoolidccrem`** groups)

```
Mixed-effects REML regression                   Number of obs     =      2,000

        Grouping information
        -------------------------------------------------------------
                        |     No. of       Observations per group
         Group variable |     groups    Minimum    Average    Maximum
        ----------------+--------------------------------------------
                   _all |          1      2,000    2,000.0      2,000
                childid |        500          4        4.0          4
        -------------------------------------------------------------

                                                Wald chi2(4)      =   10248.49
Log restricted-likelihood = -7652.9171          Prob > chi2       =     0.0000

------------------------------------------------------------------------------
            math | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-----------------+------------------------------------------------------------
            year |   15.71882   .2181831    72.04   0.000     15.29119    16.14646
         genderf |  -.8005805    .927255    -0.86   0.388    -2.617967    1.016806
      schooltype |  -5.501093   1.432701    -3.84   0.000    -8.309136    -2.69305
                 |
c.genderf#c.year |   -.462735   .3055153    -1.51   0.130    -1.061534     .136064
                 |
           _cons |   37.03682   1.384354    26.75   0.000     34.32354     39.7501
------------------------------------------------------------------------------
```

```
------------------------------------------------------------------------------
  Random-effects parameters  |   Estimate   Std. err.     [95% conf. interval]
-----------------------------+------------------------------------------------
_all: Identity               |
        var(R.schoolidccrem) |   27.06586   6.164576       17.3201    42.29541
-----------------------------+------------------------------------------------
childid: Unstructured        |
                   var(year) |   7.733038     .75443      6.387155    9.362521
                  var(_cons) |   58.43265   6.991829      46.21724    73.87664
             cov(year,_cons) |     14.662   1.708738      11.31293    18.01106
-----------------------------+------------------------------------------------
               var(Residual) |   55.01896   2.509066      50.31465    60.16311
------------------------------------------------------------------------------
LR test vs. linear model: chi2(4) = 1565.41               Prob > chi2 = 0.0000

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