Example of Crossed Random Effects Models: Kids Nested within Primary and Secondary Schools

Crossed random effects models (also known as cross-classified models) are useful in situations in which people belong to more than one cluster or move between clusters, and thus are not strictly nested within clusters. A simulated data example is shown below from Hox chapter 7, in which kids are nested within primary schools AND within secondary schools, but primary and secondary schools are crossed with each other at level 2 (1000 kids, 30 secondary schools, 50 primary schools).

```
LIBNAME hox  'F:\Example Data\Hox Crossed Schools';
DATA pupcross; SET hox.pupcross; run;

* Getting means per primary school and secondary school of kid predictor;
PROC SORT DATA=pupcross; BY pschool; RUN;
PROC MEANS NOPRINT DATA=pupcross; BY pschool; VAR pupSES;
       OUTPUT OUT=Pses MEAN(pupSES)=pmSES; RUN;
PROC SORT DATA=pupcross; BY sschool; RUN;
PROC MEANS NOPRINT DATA=pupcross; BY sschool; VAR pupSES;
       OUTPUT OUT=Sses MEAN(pupSES)=smSES; RUN;

* Merge back into individual data;
PROC SORT DATA=pupcross; BY pschool; RUN;
DATA pupcross; MERGE pupcross Pses; BY pschool; DROP _TYPE_ _FREQ_;
       LABEL pmSES= "Primary School Mean SES"; RUN;
PROC SORT DATA=pupcross; BY sschool; RUN;
DATA pupcross; MERGE pupcross Sses; BY sschool; DROP _TYPE_ _FREQ_;
       LABEL smSES= "Secondary School Mean SES"; RUN;

* Get means to center with;
PROC MEANS DATA=pupcross; VAR achiev pdenom sdenom pupSES pmSES smSES; RUN;
DATA pupcross; SET pupcross;
       pupSES4 = pupSES - 4; LABEL pupSES4= "Student SES (0=4)";
       pmSES4  = pmSES - 4;  LABEL pmSES4=  "Primary School Mean SES (0=4)";
       smSES4  = smSES - 4;  LABEL smSES4=  "Secondary School Mean SES (0=4)"; RUN;
```

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
ACHIEV	achievement score in secondary school	1000	6.3435000	0.8676812	3.9000000	9.9000000
PDENOM	primary school denominational?	1000	0.5770000	0.4942826	0	1.0000000
SDENOM	secondary school denominational?	1000	0.6660000	0.4718757	0	1.0000000
PUPSES	pupil ses	1000	4.0980000	1.3979808	1.0000000	6.0000000
pmSES	Primary School Mean SES	1000	4.0980000	0.2748875	3.4705882	4.7333333
smSES	Secondary School Mean SES	1000	4.0980000	0.2351805	3.4705882	4.5238095

We can start with a basic model in which we assume that academic achievement for child i who went to primary school p and secondary school s can be modeled by this equation: $Y_{kps} = \gamma_{000} + U_{00s} + e_{kps}$ in which achievement in 9th grade (i.e., in secondary school) is expected to be correlated among kids from the same secondary school (i.e., a random intercept at level 2 for secondary school).

```
TITLE1 'SAS Empty Model: Secondary Schools';
TITLE2 'Random Intercept Only for Secondary Schools Only';
PROC MIXED DATA=pupcross COVTEST NOCLPRINT NOITPRINT MAXIT=2000 METHOD=REML;
       CLASS pupil pschool sschool;
       MODEL achiev =   / SOLUTION DDFM=Satterthwaite;
       RANDOM INTERCEPT / SUBJECT=sschool TYPE=UN; *Level 2 variance for secondary; run;
```

Dimensions	
Covariance Parameters	2
Columns in X	1
Columns in Z Per Subject	1
Subjects	30 # of secondary schools
Max Obs Per Subject	48 # kids per secondary school

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z	
UN(1,1)	SSCHOOL	0.07206	0.02449	2.94	0.0016	Secondary Random Intercept Variance
Residual		0.6833	0.03102	22.03	<.0001	Residual Pupil Variance

Fit Statistics

-2 Res Log Likelihood	2504.7
AIC (smaller is better)	2508.7
AICC (smaller is better)	2508.7
BIC (smaller is better)	2511.5

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|	
Intercept	6.3386	0.05583	29	113.53	<.0001	Grand Mean Achievement Score

---

However, because primary schools may have lasting effects, it might make sense to also allow a random intercept for primary school that is <u>crossed</u> at level 2 with the random intercept for secondary school:

$$Y_{kps} = \gamma_{000} + U_{0p0} + U_{00s} + e_{kps}$$

```
TITLE1 'SAS Empty Model: Primary by Secondary School Crossed';
TITLE2 'Random Intercept Only for Primary and Secondary Schools';
PROC MIXED DATA=pupcross COVTEST NOCLPRINT NOITPRINT MAXIT=2000 METHOD=REML;
      CLASS pupil pschool sschool;
      MODEL achiev =   / SOLUTION DDFM=Satterthwaite;
      RANDOM INTERCEPT / SUBJECT=sschool TYPE=UN; *Level 2 variance for secondary;
      RANDOM INTERCEPT / SUBJECT=pschool TYPE=UN; *Level 2 variance for primary; run;
```

Dimensions

Covariance Parameters	3
Columns in X	1
Columns in Z Per Subject	80
Subjects	1
Max Obs Per Subject	1000

Notice that SAS thinks we have 1 subject with 1000 observations—that's ok. What it refers to is how many cases share the exact same **V** matrix, which never occurs here.

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z	
UN(1,1)	SSCHOOL	0.06666	0.02190	3.04	0.0012	Secondary Random Intercept Variance
UN(1,1)	PSCHOOL	0.1719	0.04018	4.28	<.0001	Primary Random Intercept Variance
Residual		0.5131	0.02390	21.47	<.0001	Residual Pupil Variance

Fit Statistics

-2 Res Log Likelihood	2321.1
AIC (smaller is better)	2327.1
AICC (smaller is better)	2327.1
BIC (smaller is better)	2321.1

Do we need both random intercepts?

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	6.3486	0.07890	66	80.46	<.0001

<u>Of the total variation of 0.75166 (from summing all three orthogonal variances):</u>

0.06666 / 0.75166 = .089 reflects mean achievement differences between secondary schools

0.1719 / 0.75166 = .229 reflects mean achievement differences between primary schools

0.5131 / 0.75166 = .683 reflects achievement differences between kids with same schooling

Adding primary school and secondary school denomination as control predictors…

$$Y_{kps} = \gamma_{000} + \gamma_{010}(\text{PrimDenom}_p) + \gamma_{001}(\text{SecDenom}_s) + U_{0p0} + U_{00s} + e_{kps}$$

```
TITLE1 'SAS Predictive Model: Primary by Secondary School Crossed';
TITLE2 'With School Denomination Variables';
PROC MIXED DATA=pupcross COVTEST NOCLPRINT NOITPRINT MAXIT=2000 METHOD=REML;
       CLASS pupil pschool sschool;
       MODEL achiev = pdenom sdenom / SOLUTION DDFM=Satterthwaite;
       RANDOM INTERCEPT / SUBJECT=sschool TYPE=UN; *Level 2 variance for secondary;
       RANDOM INTERCEPT / SUBJECT=pschool TYPE=UN; *Level 2 variance for primary; run;
```

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z	
UN(1,1)	SSCHOOL	0.06017	0.02044	2.94	0.0016	leftover secondary intercept variance
UN(1,1)	PSCHOOL	0.1679	0.03976	4.22	<.0001	leftover primary intercept variance
Residual		0.5129	0.02388	21.47	<.0001	residual pupil variance

Fit Statistics

-2 Res Log Likelihood	2320.3
AIC (smaller is better)	2326.3
AICC (smaller is better)	2326.4
BIC (smaller is better)	2320.3

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|	
Intercept	6.1115	0.1259	79.1	48.55	<.0001	
PDENOM	0.1868	0.1276	48.2	1.46	0.1495	Effect of Denomination of Primary
SDENOM	0.1899	0.09853	46.5	1.93	0.0601	Effect of Denomination of Secondary

---

Adding a fixed effect of pupil SES…

$$Y_{kps} = \gamma_{000} + \gamma_{010}(\text{PrimDenom}_p) + \gamma_{001}(\text{SecDenom}_s) + \gamma_{100}(\text{pupSES}_{kps} - 4) + U_{0p0} + U_{00s} + e_{kps}$$

```
TITLE1 'SAS Predictive Model: Primary by Secondary School Crossed';
TITLE2 'With a Fixed Effect of Pupil SES';
PROC MIXED DATA=pupcross COVTEST NOCLPRINT NOITPRINT MAXIT=2000 METHOD=REML;
       CLASS pupil pschool sschool;
       MODEL achiev = pdenom sdenom pupSES4 / SOLUTION DDFM=Satterthwaite;
       RANDOM INTERCEPT / SUBJECT=sschool TYPE=UN; *Level 2 variance for secondary;
       RANDOM INTERCEPT / SUBJECT=pschool TYPE=UN; *Level 2 variance for primary; run;
```

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z	
UN(1,1)	SSCHOOL	0.05710	0.01951	2.93	0.0017	leftover secondary intercept variance
UN(1,1)	PSCHOOL	0.1686	0.03966	4.25	<.0001	leftover primary intercept variance
Residual		0.4915	0.02290	21.46	<.0001	leftover residual pupil variance

Fit Statistics

-2 Res Log Likelihood	2285.1
AIC (smaller is better)	2291.1
AICC (smaller is better)	2291.1
BIC (smaller is better)	2285.1

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|	
Intercept	6.1095	0.1246	78.9	49.02	<.0001	
PDENOM	0.1890	0.1274	48.2	1.48	0.1446	
SDENOM	0.1745	0.09620	46	1.81	0.0763	
pupSES4	0.1066	0.01634	943	6.52	<.0001	But what does this effect mean??

Adding pupil SES contextual effects…

$$Y_{kps} = \gamma_{000} + \gamma_{010}(\text{PrimDenom}_p) + \gamma_{020}(\text{pmSES}_p - 4) + \gamma_{001}(\text{SecDenom}_s) + \gamma_{002}(\text{smSES}_s - 4) + \gamma_{100}(\text{pupSES}_{kps} - 4) + U_{0p0} + U_{00s} + e_{kps}$$

```
TITLE2 'With Pupil SES Contextual Effects';
PROC MIXED DATA=pupcross COVTEST NOCLPRINT NOITPRINT MAXIT=2000 METHOD=REML;
       CLASS pupil pschool sschool;
       MODEL achiev = pdenom sdenom pupSES4 pmSES4 smSES4 / SOLUTION DDFM=Satterthwaite;
       RANDOM INTERCEPT / SUBJECT=sschool TYPE=UN; *Level 2 variance for secondary;
       RANDOM INTERCEPT / SUBJECT=pschool TYPE=UN; *Level 2 variance for primary;
 run;
```

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
UN(1,1)	SSCHOOL	0.05773	0.02010	2.87	0.0020
UN(1,1)	PSCHOOL	0.1726	0.04092	4.22	<.0001
Residual		0.4916	0.02291	21.46	<.0001

Fit Statistics

-2 Res Log Likelihood	2286.9
AIC (smaller is better)	2292.9
AICC (smaller is better)	2292.9
BIC (smaller is better)	2286.9

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|	
Intercept	6.0985	0.1296	76.3	47.06	<.0001	
PDENOM	0.1884	0.1288	47.2	1.46	0.1502	
SDENOM	0.1733	0.09655	44.3	1.79	0.0796	
pupSES4	0.1057	0.01645	923	6.43	<.0001	
pmSES4	-0.03991	0.2278	48.3	-0.18	0.8617	No incremental effect of primary mean SES
smSES4	0.1607	0.2033	28.7	0.79	0.4357	No incremental effect of secondary mean SES

Adding random effect of pupil SES across secondary schools:

$$Y_{kps} = \gamma_{000} + \gamma_{010}(\text{PrimDenom}_p) + \gamma_{020}(\text{pmSES}_p - 4) + \gamma_{001}(\text{SecDenom}_s) + \gamma_{002}(\text{smSES}_s - 4) + \gamma_{100}(\text{pupSES}_{kps} - 4) + U_{0p0} + U_{00s} + U_{10s}(\text{pupSES}_{kps} - 4) + e_{kps}$$

```
TITLE2 'With Random Pupil SES Effect for Secondary';
PROC MIXED DATA=pupcross COVTEST NOCLPRINT NOITPRINT MAXIT=2000 METHOD=REML;
       CLASS pupil pschool sschool;
       MODEL achiev = pdenom sdenom pupSES4 pmSES4 smSES4 / SOLUTION DDFM=Satterthwaite;
       RANDOM INTERCEPT pupSES4 / SUBJECT=sschool TYPE=UN; *Level 2 variance for secondary;
       RANDOM INTERCEPT         / SUBJECT=pschool TYPE=UN; *Level 2 variance for primary;  run;
```

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z	Your turn to label these!
UN(1,1)	SSCHOOL	0.05583	0.01951	2.86	0.0021	
UN(2,1)	SSCHOOL	0.009256	0.005898	1.57	0.1166	
UN(2,2)	SSCHOOL	0.004633	0.003324	1.39	0.0817	
UN(1,1)	PSCHOOL	0.1710	0.04058	4.22	<.0001	
Residual		0.4833	0.02283	21.16	<.0001	

Fit Statistics

-2 Res Log Likelihood	2280.8
AIC (smaller is better)	2290.8
AICC (smaller is better)	2290.9
BIC (smaller is better)	2280.8

Do we need the random pupil SES slope over secondary schools? What kind of effects would explain that variance?

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	6.1049	0.1269	74.8	48.10	<.0001
PDENOM	0.1960	0.1283	47.3	1.53	0.1333
SDENOM	0.1597	0.09182	42	1.74	0.0893
pupSES4	0.1055	0.02073	27.7	5.09	<.0001
pmSES4	-0.03258	0.2267	48.2	-0.14	0.8863
smSES4	0.1213	0.1916	28.4	0.63	0.5317

Adding random effect of pupil SES across primary schools, too:

$$Y_{kps} = \gamma_{000} + \gamma_{010}(\text{PrimDenom}_p) + \gamma_{020}(\text{pmSES}_p - 4) + \gamma_{001}(\text{SecDenom}_s) + \gamma_{002}(\text{smSES}_s - 4) + \gamma_{100}(\text{pupSES}_{kps} - 4) + U_{0p0} + U_{1p0}(\text{pupSES}_{kps} - 4) + U_{00s} + U_{10s}(\text{pupSES}_{kps} - 4) + e_{kps}$$

```
TITLE2 'With Random Pupil SES Effect for Primary';
PROC MIXED DATA=pupcross COVTEST NOCLPRINT NOITPRINT MAXIT=2000 METHOD=REML;
       CLASS pupil pschool sschool;
       MODEL achiev = pdenom sdenom pupSES4 pmSES4 smSES4 / SOLUTION DDFM=Satterthwaite;
       RANDOM INTERCEPT pupSES4 / SUBJECT=sschool TYPE=UN; *Level 2 variance for secondary;
       RANDOM INTERCEPT pupSES4 / SUBJECT=pschool TYPE=UN; *Level 2 variance for primary;
 run;
```

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z	Your turn to label these!
UN(1,1)	SSCHOOL	0.05355	0.01884	2.84	0.0022	
UN(2,1)	SSCHOOL	0.008731	0.005637	1.55	0.1214	
UN(2,2)	SSCHOOL	0.004225	0.003229	1.31	0.0954	
UN(1,1)	PSCHOOL	0.1615	0.03846	4.20	<.0001	
UN(2,1)	PSCHOOL	0.01810	0.009498	1.91	0.0567	
UN(2,2)	PSCHOOL	0.009637	0.004384	2.20	0.0140	
Residual		0.4656	0.02247	20.72	<.0001	

Fit Statistics

-2 Res Log Likelihood	2266.1
AIC (smaller is better)	2280.1
AICC (smaller is better)	2280.2
BIC (smaller is better)	2266.1

Do we need the random pupil SES slope over primary schools? What kind of effects would explain that variance?

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	6.1126	0.1219	74.2	50.15	<.0001
PDENOM	0.1973	0.1209	46.5	1.63	0.1094
SDENOM	0.1414	0.09045	42	1.56	0.1255
pupSES4	0.1069	0.02469	35.7	4.33	0.0001
pmSES4	-0.01314	0.2148	48.3	-0.06	0.9515
smSES4	0.1562	0.1889	28.6	0.83	0.4153