Two-Level Clustered Data Example: Students within Schools
94 schools; 13,802 students; between 31-515 students in each school (*M* = 275)

SAS Code for Data Manipulation:

```
* Importing data into work library;
%LET example = F:\Example Data\School Data;
LIBNAME example "&example.";
DATA work.grade10; SET example.grade10;
      * Selecting cases that are complete for analysis variables;
      WHERE NMISS(studentID, schoolID, boyvsgirl, frlunch, math)=0;
      LABEL  studentID=    "studentID: Student ID number"
             schoolID=     "schoolID: School ID number"
             boyvsgirl=    "boyvsgirl: Boy=0, Girl=1"
             frlunch=      "frlunch: 0=No, 1=Free/Reduced Lunch"
             math=         "math: Math Test Score Outcome"; RUN;

* Getting school means to use as predictors;
PROC SORT DATA=work.grade10; BY schoolID studentID; RUN;
PROC MEANS NOPRINT N DATA= work.grade10;
      BY schoolID;
      VAR boyvsgirl frlunch math;
      OUTPUT OUT=SchoolMeans
             MEAN(boyvsgirl frlunch math)= SMboyvsgirl SMfrlunch SMmath; RUN;

* Labeling new school mean variables;
DATA work.SchoolMeans; SET work.SchoolMeans;
      SchoolN = _FREQ_;    * Saving N per school;
      DROP _TYPE_ _FREQ_; * Dropping unneeded SAS-created variables;
      LABEL  SMboyvsgirl= "SMboyvsgirl: School Mean Boy=0, Girl=1"
             SMfrlunch=    "SMfrlunch: School Mean 0=No, 1=Free/Reduced Lunch"
             SMmath=       "SMmath: School Mean Math Outcome"
             SchoolN=      "SchoolN: # Students Contributing Data"; RUN;

* Merging school means back with individual data;
DATA work.grade10; MERGE work.grade10 work.SchoolMeans; BY schoolID;
      * Selecting only schools with data from at least 30 students;
      IF SchoolN < 31 THEN DELETE; RUN;

* Outputting table of descriptives to rtf document;
ODS RTF FILE="&example.\Descriptive Stats.rtf";
TITLE "Getting means to center predictors with";
PROC MEANS MEAN STDDEV MIN MAX DATA=work.grade10;
      VAR math boyvsgirl frlunch SMmath SMboyvsgirl SMfrlunch SchoolN;
RUN; TITLE; ODS RTF CLOSE;
```

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
math	math: Math Test Score Outcome	13082	48.1185599	17.2590473	0	83.0000000
boyvsgirl	boyvsgirl: Boy=0, Girl=1	13082	0.4981654	0.5000157	0	1.0000000
frlunch	frlunch: 0=No, 1=Free/Reduced Lunch	13082	0.3075218	0.4614850	0	1.0000000
SMmath	SMmath: School Mean Math Outcome	13082	48.1185599	6.8181301	29.4509804	61.6136364
SMboyvsgirl	SMboyvsgirl: School Mean Boy=0, Girl=1	13082	0.4981654	0.0422383	0.3333333	0.6842105
SMfrlunch	SMfrlunch: School Mean 0=No, 1=Free/Reduced Lunch	13082	0.3075218	0.2220852	0	0.8032787
SchoolN	SchoolN: # Students Contributing Data	13082	274.9501605	155.3319041	31.0000000	515.0000000

```
* Centering school mean predictors;
DATA work.grade10; SET work.grade10;
      SMboyvsgirl50 = SMboyvsgirl - .50; LABEL SMboyvsgirl50= "SMboyvsgirl50: 0=.50";
      SMfrlunch30 = SMfrlunch - .30;     LABEL SMfrlunch30= "SMfrlunch30: 0=.30";
RUN;
```

SPSS Code for Data Manipulation:

```
* SPSS code to import data and create/center predictors.
GET FILE = "example/grade10.sav".
DATASET NAME grade10 WINDOW=FRONT.
VARIABLE LABELS
       studentID     "studentID: Student ID number"
       schoolID      "schoolID: School ID number"
       districtID    "districtID: District ID number"
       boyvsgirl     "boyvsgirl: Boy=0, Girl=1"
       frlunch       "frlunch: 0=No, 1=Free/Reduced Lunch"
       math          "math: Math Test Score".
* Selecting complete cases for analysis.
SELECT IF (NMISS(studentID, schoolID, boyvsgirl, frlunch, math)=0).
EXECUTE.
* Getting school means to use as level-2 predictors - SPSS 14+ can merge them back automatically.
SORT CASES BY schoolID studentID.
AGGREGATE
       /OUTFILE=* MODE=ADDVARIABLES
       /PRESORTED
       /BREAK = schoolID
       /SMboyvsgirl = MEAN(boyvsgirl)
       /SMfrlunch = MEAN(frlunch)
       /SMmath = MEAN(math)
       /SchoolN = N.
* Labeling new school mean variables.
VARIABLE LABELS
       SMboyvsgirl   "SMboyvsgirl: School Mean Boy=0, Girl=1"
       SMfrlunch     "SMfrlunch: School Mean 0=No, 1=Free/Reduced Lunch"
       SMmath        "SMmath: School Mean Math Outcome"
       SchoolN       "SchoolN: # Students Contributing Data".
* Selecting schools with data from at least 30 students.
SELECT IF (SchoolN GT 30).
* Descriptive statistics.
DESCRIPTIVES VARIABLES=math boyvsgirl frlunch SMmath SMboyvsgirl SMfrlunch SchoolN
  /STATISTICS=MEAN STDDEV MIN MAX.
* Centering school mean predictors.
COMPUTE SMboyvsgirl50 = SMboyvsgirl - .50.
COMPUTE SMfrlunch30 = SMfrlunch - .30.
VARIABLE LABELS
       SMboyvsgirl50 "SMboyvsgirl50: 0=.50"
       SMfrlunch30   "SMfrlunch30: 0=.30".
EXECUTE.
```

STATA Code for Data Manipulation:

```
 * label existing variables
label variable boyvsgirl "boyvsgirl: Student Gender 0=Boy, 1= Girl"
label variable frlunch "frlunch: Student Free/Reduced Lunch 0=No 1=Yes"
label variable math "math: Student Free/Reduced Lunch 0=No 1=Yes"
 * get school means of variables and label them
egen SMboyvsgirl = mean(boyvsgirl), by (schoolID)
egen SMfrlunch   = mean(frlunch),   by (schoolID)
egen SMmath      = mean(math),      by (schoolID)
label variable SMboyvsgirl "SMboyvsgirl: School Mean Boy=0, Girl=1"
label variable SMfrlunch   "SMfrlunch: School Mean 0=No, 1=Free/Reduced Lunch"
label variable SMmath      "SMmath: School Mean Math Outcome"
 * get number of students per school
egen SchoolN = count(studentID), by (schoolID)
label variable SchoolN= "SchoolN: # Students Contributing Data"
 * then drop schools with <= 30 students
drop if SchoolN < 31
 * centering school mean predictors
gen SMboyvsgirl50 = SMboyvsgirl - .50
label variable SMboyvsgirl50 "SMboyvsgirl50: Percentage Students who are Girls (0=50%)"
gen SMfrlunch30 = SMfrlunch - .30
label variable SMfrlunch30 "SMfrlunch30: Percentage Students with Free Lunch (0=30%)"
```

Model 1a: Two-Level Empty Means, Random Intercept for Math Test Outcome

```
TITLE1 "SAS Model 1a: 2-Level Empty Means, Random Intercept for Math Outcome";
PROC MIXED DATA=work.grade10 NOCLPRINT NOITPRINT COVTEST NAMELEN=100 IC METHOD=ML;
       CLASS schoolID studentID;
       MODEL math = / SOLUTION DDFM=Satterthwaite;
       RANDOM INTERCEPT / TYPE=UN SUBJECT=schoolID; RUN;
```

```
TITLE "SPSS Model 1a: 2-Level Empty Means, Random Intercept for Math Outcome".
MIXED math BY schoolID studentID
       /METHOD = ML
       /PRINT  = SOLUTION TESTCOV
       /FIXED  =
       /RANDOM = INTERCEPT | SUBJECT(schoolID) COVTYPE(UN).
```

```
 * STATA Model 1a: 2-Level Empty Means, Random Intercept for Math Outcome
xtmixed math  , || schoolID: , ///
       variance ml covariance(un) residuals(independent),
       estat ic, n(94)
```

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
UN(1,1)	schoolID	44.9335	7.0391	6.38	<.0001
Residual		253.18	3.1415	80.59	<.0001

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
1	1857.08	<.0001

This is the −2ΔLL test of whether we need any random effects in the model. Right now all we have is a random intercept (so df=1)

Information Criteria

Neg2LogLike	Parms	AIC	AICC	HQIC	BIC
109791	3	109797	109797	109800	109805

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	47.7561	0.7192	94.9	66.40	<.0001

Calculate the ICC (correlation of students in same school in math):

Calculate the design effect using mean #students per school: Design Effect = 1 + ((*n* – 1) * ICC)

Calculate the Effective N: $N_{effective}$ = (#Total Obs) / Design Effect

Calculate 95% random effect CI for the intercept across schools: Fixed effect ± 1.96*SQRT(variance)

Model 1b: Two-Level Empty Means, Random Intercept for Student Gender Predictors (no SPSS)

```
TITLE1 "SAS Model 1b: 2-Level Empty Means, Random Intercept for Binary Gender Predictor";
PROC GLIMMIX DATA=work.grade10 NOCLPRINT NOITPRINT METHOD=QUAD (QPOINTS=7);
       CLASS schoolID studentID;
       MODEL boyvsgirl (DESCENDING) =  / SOLUTION LINK=LOGIT DIST=BINARY DDFM=BW;
       RANDOM INTERCEPT / TYPE=UN SUBJECT=schoolID;
       COVTEST "Need Random Intercept?" 0; RUN;

 * STATA Model 1b: 2-Level Empty Means, Random Intercept for Binary Gender Predictor
xtmelogit boyvsgirl,  ||  schoolID:  , variance covariance(unstructured) intpoints(7),
       estat ic, n(94)
```

```
           Fit Statistics
-2 Log Likelihood           18135.09
AIC  (smaller is better)    18139.09
AICC (smaller is better)    18139.09
BIC  (smaller is better)    18144.18
CAIC (smaller is better)    18146.18
HQIC (smaller is better)    18141.15

      Covariance Parameter Estimates
Cov                                 Standard
Parm       Subject     Estimate        Error
UN(1,1)    schoolID    0.001441     0.003211
```

For logit models for binary outcomes, the ICC is calculated as random intercept variance / (random intercept variance + 3.29)
ICC = 0.001441 / (0.001441 + 3.29) = .0004

```
                   Solutions for Fixed Effects
                         Standard
Effect         Estimate     Error       DF    t Value    Pr > |t|
Intercept      -0.00614   0.01849       93      -0.33      0.7405
```

The intercept is log odds of being a girl in a school with U_{0j} = 0 → prob = .4985

```
                        Tests of Covariance Parameters
                           Based on the Likelihood
Label                      DF    -2 Log Like      ChiSq    Pr > ChiSq    Note
Need Random Intercept?      1          18135       0.24        0.3134    MI
MI: P-value based on a mixture of chi-squares.
```

This is the −2ΔLL test of whether we need a random intercept for gender.

Model 1c: Two-Level Empty Means, Random Intercept for Student Free/Reduced Lunch Predictor (no SPSS)

```
TITLE1 "SAS Model 1c: 2-Level Empty Means, Random Intercept for Binary Lunch Predictor";
PROC GLIMMIX DATA=work.grade10 NOCLPRINT NOITPRINT METHOD=QUAD (QPOINTS=7);
       CLASS schoolID studentID;
       MODEL frlunch (DESCENDING) =  / SOLUTION LINK=LOGIT DIST=BINARY DDFM=BW;
       RANDOM INTERCEPT / TYPE=UN SUBJECT=schoolID;
       COVTEST "Need Random Intercept?" 0; RUN;

* STATA Model 1c: 2-Level Empty Means, Random Intercept for Binary Lunch Predictor
xtmelogit frlunch,  ||  schoolID:  , variance covariance(unstructured) intpoints(7),
       estat ic, n(94)
```

```
           Fit Statistics
-2 Log Likelihood           13172.43
AIC  (smaller is better)    13176.43
AICC (smaller is better)    13176.43
BIC  (smaller is better)    13181.52
CAIC (smaller is better)    13183.52
HQIC (smaller is better)    13178.48
      Covariance Parameter Estimates
Cov                                 Standard
Parm       Subject     Estimate        Error
UN(1,1)    schoolID      1.9545       0.3315
```

For logit models for binary outcomes, the ICC is calculated as random intercept variance / (random intercept variance + 3.29)
ICC = 1.9545 / (1.9545 + 3.29) = .327

Solutions for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	-1.1721	0.1494	93	-7.85	<.0001

The intercept is log odds of a student getting free lunch in a school with $U_{0j} = 0$ → probability = .2365

Tests of Covariance Parameters Based on the Likelihood

Label	DF	-2 Log Like	ChiSq	Pr > ChiSq	Note
Need Random Intercept?	1	16146	2973.46	<.0001	MI

MI: P-value based on a mixture of chi-squares.

This is the −2ΔLL test of whether we need a random intercept for frlunch.

Model 2a: Predicting Math with a Fixed Effect of Student Gender (Level 1)

```
TITLE1 "SAS Model 2a: Predicting Math with a Fixed Effect of Student Gender";
PROC MIXED DATA=work.grade10 NOCLPRINT NOITPRINT COVTEST NAMELEN=100 IC METHOD=ML;
      CLASS schoolID studentID;
      MODEL math = boyvsgirl / SOLUTION DDFM=Satterthwaite;
      RANDOM INTERCEPT / TYPE=UN SUBJECT=schoolID; RUN;
```

```
TITLE "SPSS Model 2a: Predicting Math with a Fixed Effect of Student Gender".
MIXED math BY schoolID studentID WITH boyvsgirl
      /METHOD = ML
      /PRINT  = SOLUTION TESTCOV
      /FIXED  = boyvsgirl
      /RANDOM = INTERCEPT | SUBJECT(schoolID) COVTYPE(UN).
```

```
 * STATA Model 2a: Predicting Math with a Fixed Effect of Student Gender
xtmixed math c.boyvsgirl , || schoolID: ,  variance ml covariance(un) residuals(independent),
      estat ic, n(94)
```

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
UN(1,1)	schoolID	44.8203	7.0210	6.38	<.0001 → random intercept variance down 0.25%
Residual		253.00	3.1394	80.59	<.0001 → residual variance down 0.07%

Information Criteria

Neg2LogLike	Parms	AIC	AICC	HQIC	BIC	CAIC
109782	4	109790	109790	109794	109800	109804

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	47.3300	0.7319	102	64.66	<.0001
boyvsgirl	0.8462	0.2791	13E3	3.03	0.0024

What does the effect of student gender represent in this model?

What are we assuming about the effect of student gender?

Model 2b: Adding a Fixed Effect of School Proportion of Girls (Level 2)

```
TITLE1 "SAS Model 2b: Adding Fixed Effect of School Proportion Girls";
PROC MIXED DATA=work.grade10 NOCLPRINT NOITPRINT COVTEST NAMELEN=100 IC METHOD=ML;
      CLASS schoolID studentID;
      MODEL math = boyvsgirl SMboyvsgirl50 / SOLUTION DDFM=Satterthwaite OUTPM=work.BoySave;
      RANDOM INTERCEPT  / TYPE=UN SUBJECT=schoolID;
      ESTIMATE "Gender Between-School Effect"  boyvsgirl 1 SMboyvsgirl50 1;
RUN; PROC CORR NOSIMPLE DATA=work.BoySave; VAR math pred; RUN;
```

```
TITLE "SPSS Model 2b: Adding Fixed Effect of School Proportion Girls".
MIXED math BY schoolID studentID WITH boyvsgirl SMboyvsgirl50
      /METHOD = ML
      /PRINT  = SOLUTION TESTCOV
      /FIXED  = boyvsgirl SMboyvsgirl50
      /RANDOM = INTERCEPT | SUBJECT(schoolID) COVTYPE(UN)
       /SAVE = FIXPRED(boypred)
       /TEST = "Gender Between-School Effect" boyvsgirl 1 SMboyvsgirl50 1.
CORRELATIONS /VARIABLES = math boypred.
```

```
 * STATA Model 2b: Adding Fixed Effect of School Proportion Girls
xtmixed math c.boyvsgirl c.SMboyvsgirl50, || schoolID: , ///
      variance ml covariance(un) residuals(independent),
      estat ic, n(94),
      predict boypred,
      estimates store FixGender,
      lincom 1*boyvsgirl + 1*SMboyvsgirl50  // gender between-school effect
corr math boypred
```

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
UN(1,1)	schoolID	43.4390	6.7859	6.40	<.0001 → random intercept variance down 3.08%
Residual		253.00	3.1392	80.59	<.0001 → residual variance not reduced further

Information Criteria

Neg2LogLike	Parms	AIC	AICC	HQIC	BIC	CAIC
109779	5	109789	109789	109794	109801	109806

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	47.2605	0.7229	103	65.37	<.0001
boyvsgirl	0.8352	0.2791	13E3	2.99	0.0028
SMboyvsgirl50	20.8313	11.9611	103	1.74	0.0846

Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > |t|
Gender Between-School Effect	21.6665	11.9578	103	1.81	0.0729

Pearson Correlation Coefficients, N = 13082
Prob > |r| under H0: Rho=0

	math	Pred
math	1.00000	0.03016
math: Math Test Score Outcome		0.0006

R = .03016, so total R^2 ~ .001

What does the effect of school mean gender represent in this model?

Model 2c: Adding a Random Effect of Student Gender (over Schools)

```
TITLE1 "SAS Model 2c: Adding Random Effect of Student Gender";
PROC MIXED DATA=work.grade10 NOCLPRINT NOITPRINT COVTEST NAMELEN=100 IC METHOD=ML;
      CLASS schoolID studentID;
      MODEL math = boyvsgirl SMboyvsgirl50 / SOLUTION DDFM=Satterthwaite;
      RANDOM INTERCEPT boyvsgirl / G TYPE=UN SUBJECT=schoolID; RUN;
```

```
TITLE "SPSS Model 2c: Adding Random Effect of Student Gender".
MIXED math BY schoolID studentID WITH boyvsgirl SMboyvsgirl50
      /METHOD = ML
      /PRINT  = SOLUTION TESTCOV G
      /FIXED  = boyvsgirl SMboyvsgirl50
      /RANDOM = INTERCEPT boyvsgirl | SUBJECT(schoolID) COVTYPE(UN).
```

```
* STATA Model 2c: Adding Random Effect of Student Gender
xtmixed math c.boyvsgirl c.SMboyvsgirl50, || schoolID: boyvsgirl,  ///
      variance ml covariance(un) residuals(independent),
      estat recovariance, level(schoolID),
      estat ic, n(94),
      estimates store RandGender,
      lrtest RandGender FixGender
```

Estimated G Matrix

Row	Effect	School ID number	Col1	Col2
1	Intercept	125	44.0992	-0.7533
2	boyvsgirl	125	-0.7533	0.6203

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	schoolID	44.0992	7.1632	6.16	<.0001
UN(2,1)	schoolID	-0.7533	2.1120	-0.36	0.7213
UN(2,2)	schoolID	0.6203	0.8957	0.69	0.2443
Residual		252.84	3.1432	80.44	<.0001

Information Criteria

Neg2LogLike	Parms	AIC	AICC	HQIC	BIC	CAIC
109778	7	109792	109792	109799	109810	109817

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	47.2521	0.7289	94.9	64.82	<.0001
boyvsgirl	0.8553	0.3005	55.1	2.85	0.0062
SMboyvsgirl50	20.5425	11.9666	103	1.72	0.0890

Is model 2c better than model 2b? How do we know?

So what does this mean about the effect of student gender?

Model 3a: Adding a Fixed Effect of Student Free/Reduced Lunch (Level 1)

```
TITLE1 "SAS Model 3a: Adding Fixed Effect of Student Free/Reduced Lunch";
PROC MIXED DATA=work.grade10 NOCLPRINT NOITPRINT COVTEST NAMELEN=100 IC METHOD=ML;
      CLASS schoolID studentID;
      MODEL math = boyvsgirl SMboyvsgirl50 frlunch / SOLUTION DDFM=Satterthwaite;
      RANDOM INTERCEPT / TYPE=UN SUBJECT=schoolID; RUN;


TITLE "SPSS 3a: Adding Fixed Effect of Student Free/Reduced Lunch".
MIXED math BY schoolID studentID WITH boyvsgirl SMboyvsgirl50 frlunch
      /METHOD = ML
      /PRINT  = SOLUTION TESTCOV
      /FIXED  = boyvsgirl SMboyvsgirl50 frlunch
      /RANDOM = INTERCEPT | SUBJECT(schoolID) COVTYPE(UN).


 * STATA Model 3a: Adding Fixed Effect of Student Free/Reduced Lunch
xtmixed math c.boyvsgirl c.SMboyvsgirl50 c.frlunch, || schoolID: , ///
      variance ml covariance(un) residuals(independent),
      estat ic, n(94)
```

```
                 Covariance Parameter Estimates
                                   Standard         Z
Cov Parm     Subject     Estimate     Error     Value      Pr > Z
UN(1,1)      schoolID     26.1636    4.3105      6.07      <.0001 → random intercept var down by 39.77%
Residual                   239.14    2.9675     80.59      <.0001 → residual variance down by 5.48%

                           Information Criteria
Neg2LogLike     Parms        AIC        AICC        HQIC        BIC        CAIC
     109003         6     109015      109015      109022     109031      109037

                    Solution for Fixed Effects
                                Standard
Effect             Estimate        Error       DF     t Value     Pr > |t|
Intercept           50.1325       0.5868      110       85.44       <.0001
boyvsgirl            0.8555       0.2714     13E3        3.15       0.0016
SMboyvsgirl50       15.0796       9.5669      104        1.58       0.1180
frlunch             -9.4298       0.3316     13E3      -28.43       <.0001
```

What does the effect of student free/reduced lunch represent in this model?

What are we assuming about the effect of student free/reduced lunch?

Model 3b: Adding a Fixed Effect of School Proportion Free/Reduced Lunch (Level 2)

```
TITLE1 "SAS Model 3b: Adding Fixed Effect of School Proportion Free/Reduced Lunch";
PROC MIXED DATA=work.grade10 NOCLPRINT NOITPRINT COVTEST NAMELEN=100 IC METHOD=ML;
      CLASS schoolID studentID;
      MODEL math = boyvsgirl SMboyvsgirl50 frlunch SMfrlunch30
                   / SOLUTION DDFM=Satterthwaite OUTPM=work.LunchSave;
      RANDOM INTERCEPT / TYPE=UN SUBJECT=schoolID;
      ESTIMATE "FR Lunch Between-School Effect"  frlunch 1 SMfrlunch30 1;
RUN; PROC CORR NOSIMPLE DATA=work.LunchSave; VAR math pred; RUN;
```

```
TITLE "SPSS Model 3b: Adding Fixed Effect of School Proportion Free/Reduced Lunch".
MIXED math BY schoolID studentID WITH boyvsgirl SMboyvsgirl50 frlunch SMfrlunch30
      /METHOD = ML
      /PRINT  = SOLUTION TESTCOV
      /FIXED  = boyvsgirl SMboyvsgirl50 frlunch SMfrlunch30
      /RANDOM = INTERCEPT | SUBJECT(schoolID) COVTYPE(UN)
       /SAVE = FIXPRED(lunchpred)
       /TEST = "FR Lunch Between-School Effect" frlunch 1 SMfrlunch30 1.
CORRELATIONS /VARIABLES = math lunchpred.
```

```
 * STATA Model 3b: Adding Fixed Effect of School Proportion Free/Reduced Lunch
xtmixed math c.boyvsgirl c.SMboyvsgirl50 c.frlunch c. SMfrlunch30, || schoolID: , ///
      variance ml covariance(un) residuals(independent),
      estat ic, n(94),
      predict lunchpred,
      estimates store FixFRLunch,
      lincom 1*frlunch + 1*SMfrlunch30  // FR lunch between-school effect
corr math lunchpred
```

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z	
UN(1,1)	schoolID	13.3767	2.5214	5.31	<.0001	→ random intercept var down by 48.87%
Residual		239.21	2.9693	80.56	<.0001	→ residual variance up by 0.03%

Information Criteria

Neg2LogLike	Parms	AIC	AICC	HQIC	BIC	CAIC
108954	7	108968	108968	108975	108986	108993

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	50.1574	0.4542	113	110.44	<.0001
boyvsgirl	0.8549	0.2714	13E3	3.15	0.0016
SMboyvsgirl50	5.6834	7.4083	101	0.77	0.4448
frlunch	-9.1757	0.3343	13E3	-27.45	<.0001
SMfrlunch30	-16.5983	2.0143	83.8	-8.24	<.0001

Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > |t|
FR Lunch Between-School Effect	-25.7740	1.9863	79.2	-12.98	<.0001

Pearson Correlation Coefficients, N = 13082
Prob > |r| under H0: Rho=0

	math	Pred
math	1.00000	0.40455
math: Math Test Score Outcome		<.0001

$\boxed{R = .40455\text{, so total } R^2 \sim .164}$

What does the effect of school proportion free/reduced lunch represent in this model?

What does the effect of student free/reduced lunch NOW represent in this model?

Model 3c: Adding a Random Effect of Student Free/Reduced Lunch (over Schools)

```
TITLE1 "SAS Model 3c: Adding Random Effect of Student Free/Reduced Lunch";
PROC MIXED DATA=work.grade10 NOCLPRINT NOITPRINT COVTEST NAMELEN=100 IC METHOD=ML;
      CLASS schoolID studentID;
      MODEL math = boyvsgirl SMboyvsgirl50 frlunch SMfrlunch30 / SOLUTION DDFM=Satterthwaite;
      RANDOM INTERCEPT frlunch / G TYPE=UN SUBJECT=schoolID; RUN;


TITLE "SPSS Model 3c: Adding Random Effect of Student Free/Reduced Lunch".
MIXED math BY schoolID studentID WITH boyvsgirl SMboyvsgirl50 frlunch SMfrlunch30
      /METHOD = ML
      /PRINT  = SOLUTION TESTCOV G
      /FIXED  = boyvsgirl SMboyvsgirl50 frlunch SMfrlunch30
      /RANDOM = INTERCEPT frlunch | SUBJECT(schoolID) COVTYPE(UN).


 * STATA Model 3c: Adding Random Effect of Student Free/Reduced Lunch
xtmixed math c.boyvsgirl c. SMboyvsgirl50 c.frlunch c. SMfrlunch30, || schoolID: frlunch,  ///
      variance ml covariance(un) residuals(independent),
      estat recovariance, level(schoolID),
      estat ic, n(94),
      estimates store RandFRLunch
      lrtest RandFRLunch FixFRLunch
```

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	schoolID	19.7443	3.7094	5.32	<.0001
UN(2,1)	schoolID	-11.7978	3.1478	-3.75	0.0002
UN(2,2)	schoolID	12.7448	3.3242	3.83	<.0001
Residual		236.63	2.9443	80.37	<.0001

Information Criteria

Neg2LogLike	Parms	AIC	AICC	HQIC	BIC	CAIC
108866	9	108884	108884	108893	108907	108916

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	49.7910	0.5309	97.4	93.79	<.0001
boyvsgirl	0.8961	0.2703	13E3	3.31	0.0009
SMboyvsgirl50	4.3223	7.2337	104	0.60	0.5515
frlunch	-8.4552	0.5621	98.2	-15.04	<.0001
SMfrlunch30	-16.8499	1.9441	77.7	-8.67	<.0001

Is model 3c better than model 3b? How do we know?

So what does this mean about the effect of student free/reduced lunch?

Calculate a 95% random effects confidence interval for the random FRlunch slope:

Model 3d: Adding a Cross-Level Interaction of Student by School Free/Reduced Lunch

```
TITLE1 "SAS Model 3d: Adding Cross-Level Interaction of Student by School Free/Reduced Lunch";
PROC MIXED DATA=work.grade10 NOCLPRINT NOITPRINT COVTEST NAMELEN=100 IC METHOD=ML;
      CLASS schoolID studentID;
      MODEL math = boyvsgirl SMboyvsgirl50 frlunch SMfrlunch30 frlunch*SMfrlunch30
                   / SOLUTION DDFM=Satterthwaite;
      RANDOM INTERCEPT frlunch / TYPE=UN SUBJECT=schoolID; RUN;

TITLE "SPSS Model 3d: Adding Cross-Level Interaction of Student by School Free/Reduced Lunch".
MIXED math BY schoolID studentID WITH boyvsgirl SMboyvsgirl50 frlunch SMfrlunch30
      /METHOD = ML
      /PRINT  = SOLUTION TESTCOV
      /FIXED  = boyvsgirl SMboyvsgirl50 frlunch SMfrlunch30 frlunch*SMfrlunch30
      /RANDOM = INTERCEPT frlunch | SUBJECT(schoolID) COVTYPE(UN).

* STATA Model 3d: Adding Cross-Level Interaction of Student by School Free/Reduced Lunch
xtmixed math c.boyvsgirl c.SMboyvsgirl50 c.frlunch c.smfrlunch30 c.frlunch#c.smfrlunch30, ///
      || schoolID: frlunch, variance ml covariance(un) residuals(independent),
      estat ic, n(94)
```

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	schoolID	19.5909	3.6558	5.36	<.0001
UN(2,1)	schoolID	-11.1933	3.0624	-3.66	0.0003
UN(2,2)	schoolID	11.8021	3.1679	3.73	<.0001
Residual		236.62	2.9439	80.37	<.0001

Information Criteria

Neg2LogLike	Parms	AIC	AICC	HQIC	BIC	CAIC
108863	10	108883	108883	108894	108909	108919

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	49.7497	0.5300	99.7	93.87	<.0001
boyvsgirl	0.8996	0.2703	13E3	3.33	0.0009
SMboyvsgirl50	4.8849	7.2788	103	0.67	0.5037
frlunch	-8.7042	0.5679	112	-15.33	<.0001
SMfrlunch30	-19.2722	2.4760	89.2	-7.78	<.0001
frlunch*SMfrlunch30	4.3141	2.6386	103	1.64	0.1051

What does the effect of student free/reduced lunch NOW represent in this model?

What does the effect of school proportion free/reduced lunch NOW represent in this model?

What does the cross-level interaction of student by school free/reduced lunch represent?

Which variance component should be reduced (non-significantly)?

Calculate its proportion reduction relative to model 3c:

Model 3e: Adding a Level-2 Interaction of Quadratic School Free/Reduced Lunch

```
TITLE1 "SAS Model 3e: Adding Level-2 Interaction of Quadratic School Free/Reduced Lunch";
PROC MIXED DATA=work.grade10 NOCLPRINT NOITPRINT COVTEST NAMELEN=100 IC METHOD=ML;
      CLASS schoolID studentID;
      MODEL math = boyvsgirl SMboyvsgirl50 frlunch SMfrlunch30 frlunch*SMfrlunch30
                   SMfrlunch30*SMfrlunch30 / SOLUTION DDFM=Satterthwaite OUTPM=work.TotalSave;
      RANDOM INTERCEPT frlunch / TYPE=UN SUBJECT=schoolID;
ESTIMATE "FR Lunch Between-School Main Effect" frlunch 1 SMfrlunch30 1;
ESTIMATE "FR Lunch Between-School Interaction" frlunch*SMfrlunch30 1 SMfrlunch30*SMfrlunch30 1;
RUN; PROC CORR NOSIMPLE DATA=work.TotalSave; VAR math pred; RUN;

TITLE "SPSS Model 3e: Adding Level-2 Interaction of Quadratic School Free/Reduced Lunch".
MIXED math BY schoolID studentID WITH boyvsgirl SMboyvsgirl50 frlunch SMfrlunch30
      /METHOD = ML
      /PRINT  = SOLUTION TESTCOV
      /FIXED  = boyvsgirl SMboyvsgirl50 frlunch SMfrlunch30 frlunch*SMfrlunch30
                SMfrlunch30*SMfrlunch30
      /RANDOM = INTERCEPT frlunch | SUBJECT(schoolID) COVTYPE(UN)
      /SAVE = FIXPRED(totalpred)
 /TEST = "FR Lunch Between-School Main Effect" frlunch 1 SMfrlunch30 1
 /TEST = "FR Lunch Between-School Interaction" frlunch*SMfrlunch30 1 SMfrlunch30*SMfrlunch30 1.
CORRELATIONS /VARIABLES = math totalpred.

 * interaction terms for for lincom statements
gen lunchcross = SMfrlunch30*frlunch
gen SMfrlunch30sq = SMfrlunch30*SMfrlunch30
 * STATA Model 3e: Adding Cross-Level Interaction of Student by School Free/Reduced Lunch
xtmixed math c.boyvsgirl c.SMboyvsgirl50 c.frlunch c.SMfrlunch30 c.lunchcross ///
      c.SMfrlunch30sq, || schoolID: frlunch, variance ml covariance(un) residuals(independent),
      estat ic, n(94),
      predict totalpred,
      lincom 1*frlunch + 1*SMfrlunch30          // FR lunch between-school main effect
      lincom 1*lunchcross + 1*SMfrlunch30sq     // FR lunch between-school interaction
corr math totalpred
```

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	schoolID	18.8449	3.5480	5.31	<.0001
UN(2,1)	schoolID	-10.8805	3.0207	-3.60	0.0003
UN(2,2)	schoolID	11.8613	3.1882	3.72	<.0001
Residual		236.61	2.9439	80.37	<.0001

Information Criteria

Neg2LogLike	Parms	AIC	AICC	HQIC	BIC	CAIC
108861	11	108883	108883	108894	108911	108922

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	50.3622	0.6594	114	76.37	<.0001
boyvsgirl	0.9016	0.2703	13E3	3.34	0.0009
SMboyvsgirl50	3.0898	7.2988	100	0.42	0.6730
frlunch	-8.8417	0.5775	114	-15.31	<.0001
SMfrlunch30	-17.9370	2.5918	88.1	-6.92	<.0001
frlunch*SMfrlunch30	5.5170	2.7679	108	1.99	0.0488
SMfrlunch30*SMfrlunch30	-13.5458	8.9313	86.4	-1.52	0.1330

Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > |t|
FR Lunch Between-School Main Effect	-26.7787	2.6001	90.5	-10.30	<.0001
FR Lunch Between-School Interaction	-8.0288	8.5365	74.4	-0.94	0.3500

```
      Pearson Correlation Coefficients, N = 13082
               Prob > |r| under H0: Rho=0
                                          math         Pred
math                                   1.00000      0.40589
math: Math Test Score Outcome                        <.0001
```

R = .40589, so total R^2 = .165

What does the cross-level interaction of student by school free/reduced lunch NOW represent?

What does the level-2 interaction of quadratic school free/reduced lunch represent?

Which variance component should be reduced (non-significantly)?

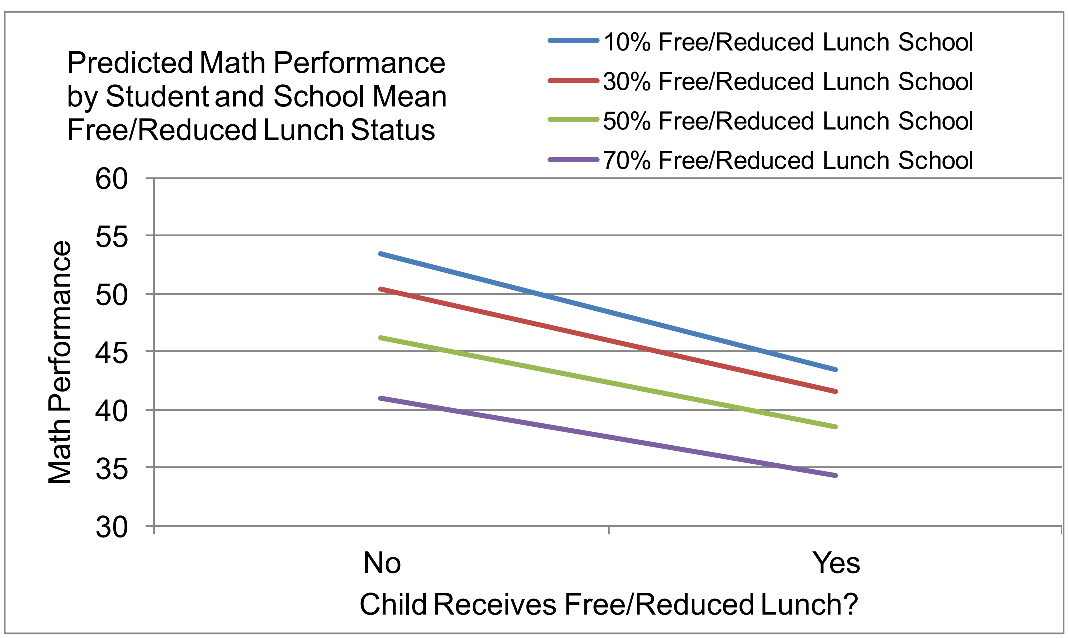
Calculate its proportion reduction relative to model 3d:

Equation for this final model:

Level 1:

Level 2:

Figure 1: Plot of model-predicted math by free/reduced lunch status



Sample Results Section (note that "smushed" models are not reported)…

The extent to which student gender and free/reduced lunch status could predict student math outcomes was examined in a series of multilevel models in which the 13,802 students were modeled as nested within their 94 schools. Maximum likelihood (ML) was used in estimating and reporting all model parameters; denominator degrees of freedom were estimated using the Satterthwaite method. The significance of fixed effects was evaluated with individual Wald tests (i.e., of estimate / SE), whereas random effects were evaluated via likelihood ratio tests (i.e., −2ΔLL with degrees of freedom equal to the number of new random effects variances and covariances). Effect size was evaluated via pseduo-R^2 values for the proportion reduction in each variance component, as well as with total R^2, the squared correlation between the actual math outcome values and the math outcomes predicted by the model fixed effects.

As derived from an empty means, random intercept model, student math scores had an intraclass correlation of .15, indicating that 15% of the variance in math scores was between schools. A 95% random effects confidence interval, calculated as fixed intercept ± 1.96*SQRT(random intercept variance), revealed that 95% of the sample schools were predicted to have intercepts for school mean math scores between 35 to 61. With respect to the student-level binary predictors, boys who did not receive free/reduced lunch were treated as the reference group. The intraclass correlations for student gender and free/reduced lunch status, as derived from empty means, random intercept logistic models, were ~0 and .33, respectively. Thus, there was little to no variability across schools in the proportion of student who were girls, but significant variability across schools in the proportion of students who received free/reduced lunch. Their contextual effects at level 2 were represented by the school proportions of girls and of students who receive free/reduced lunch; these predictors were centered near their sample means (i.e., at .50 and .30, respectively).

We first examined the effect of student gender at each level. Although the within-school gender effect was significant, such that girls had higher math scores by 0.84, it only accounted for 0.07% of the residual variance (i.e., individual differences in math within the same school). The between-school and contextual gender effects were not significant, indicating no additional impact on math scores related to gender composition of the school (i.e., neither before nor after controlling for student gender). Gender composition accounted for 3.33% of the random intercept variance (i.e., between-school differences in mean math scores), resulting in a total $R^2 < .001$. Finally, a random slope for the effect of gender did not result in a significant improvement in model fit, −2ΔLL(2) = 0.6, *p* = .74, indicating that the small advantage for girls was equivalent across schools. The random slope for gender was thus removed before continuing.

The effects of free/reduced lunch status at each level were then added to the model. The within-school effect was significant and accounted for 5.45% of the remaining residual variance, and indicated that students who receive free/reduced lunch are expected to have lower math scores than other students in their school by 9.18. The between-school effect was also significant and accounted for 69.21% of the remaining random intercept variance, and indicated that for every additional 10% of students who receive free/reduced lunch, that school's mean math score is expected to be lower by 2.58. After controlling for student free/reduced lunch, the contextual free/reduced lunch effect of −1.66 per additional 10% of students was still significant. Finally, a random slope for the effect of free/reduced lunch resulted in a significant improvement in model fit, −2ΔLL(2) = 88.2, *p* < .001, indicating that the size of the disadvantage related to free/reduced lunch differed significantly across schools. A 95% random effects confidence interval for the student free/reduced lunch effect, calculated as fixed slope ± 1.96*SQRT(random slope variance), revealed that 95% of the schools were predicted to have lunch-related gaps between students ranging from −15.45 to −1.46.

The extent to which school differences in the lunch-related disadvantage in math could be predicted from school lunch composition was then examined by adding a cross-level intra-variable interaction between the student and school lunch predictors, as well as the quadratic effect of school lunch composition to control for a contextual interaction effect. The within-school lunch effect was significantly moderated by school lunch composition (which reduced its random slope variance by 7.49%), although the between-school and contextual effects were not significant, reducing the random intercept variance by another 4.45%, for a total $R^2 = .165$. The significant intra-variable cross-level interaction, as shown by the nonparallel slopes of the lines in Figure 1, indicated that the lunch-related disadvantage in math scores of 8.84, as found for students receiving free/reduced lunch in schools in which 30% of students received free/reduced lunch, became significantly less negative by 0.55 for every additional 10% of students who received free/reduced lunch. Alternatively, the contextual school effect of −1.79 per 10% free/reduced lunch students (in baseline students in schools with 30% free/reduced lunch students) was reduced by 0.55 in free/reduced lunch students. The level-2 quadratic effect, seen by the widening distance between the lines in Figure 1, indicated that the same contextual school effect became nonsignificantly more negative by 1.35 for every additional 10% free/reduced lunch students (i.e., controlling for student lunch status), or that the between-school effect of −2.6 per 10% students became nonsignificantly more negative by 0.80 per 10% students (i.e., not controlling for student lunch status).