Lab 8: Three level Normal, Math Achievement data
(From pages 463-4 (241-2, 1st ed.) of *Multilevel and Longitudinal Modeling Using Stata*)

Data: The math-achievement dataset in *Multilevel and Longitudinal Modeling Using Stata* contains information from the U.S. Sustaining Effects Study, which is a longitudinal study of children's academic progress during the six years of elementary school (kindergarten and 1^{st} through 5^{th} grade). We have repeated observations on 1,721 students from 60 public elementary schools in urban areas. Hence we have a three-level data structure: repeated observation within child within school.

Variables

- Level 1 (repeated observations within a child)
  - math: math-test score from item response model (treat as though normal)
  - year: 'centered' year of study (1 through 6 minus 3.5)
  - grade: grade level of child at time of observation - sometimes repeats
  - retained: indicator for child being held back a grade (1 = retained, 0 = not retained)

- Level 2 (child)
  - child: child id
  - female: dummy variable for gender (1 = female, 0 = male)
  - black: dummy variable for being African American
  - hispanic: dummy variable for being Hispanic

- Level 3 (school)
  - school: school id
  - size: number of students enrolled in school
  - lowinc: percentage of students from low income families
  - mobility: percentage of students moving during the course of a school year

Goals:

(1) Describe and explore data structure with three levels.
(2) Fit 3-level models with a Normal outcome using xtmixed.
(3) Interpret model parameters (effect coefficients and variance components).

I. Exploratory Data Analysis

Let's first make sure we understand the data structure. We can use the xtdes command to examine the different patterns of observations taken on children in the dataset, but which time variable do we use -- grade, year, or something else?

- Grade doesn't necessarily represent time because some children repeat grades.

- Year is not an integer variable, and xtdes only accepts integer time variables, so we will modify it to be an integer.

- We could also create an observation number variable, but since the year variable is already simple and useful, we will continue with that.

```
. gen yr=year+3.5
. xtdes, i(child) t(yr) patterns(30)
   child:  1, 2, ..., 1721                                  n =       1721
      yr:  1, 2, ..., 6                                     T =          6
           Delta(yr) = 1 unit
           Span(yr)  = 6 periods
           (child*yr uniquely identifies each observation)
Distribution of T_i:   min      5%     25%       50%       75%     95%     max
                         2       3       3         4         5       5       6

      Freq.  Percent    Cum. |  Pattern
 ---------------------------+---------
      783     45.50   45.50 |  .11111
      259     15.05   60.55 |  .111..
      185     10.75   71.30 |  ...111
      158      9.18   80.48 |  .1111.
      142      8.25   88.73 |  ..1111
       52      3.02   91.75 |  111111
       49      2.85   94.60 |  111...
       46      2.67   97.27 |  ..111.
       19      1.10   98.37 |  1111..
       11      0.64   99.01 |  11111.
        8      0.46   99.48 |  .1.111
        3      0.17   99.65 |  .1.1..
        2      0.12   99.77 |  ...1.1
        2      0.12   99.88 |  .1.11.
        1      0.06   99.94 |  .1.1.1
        1      0.06  100.00 |  .111.1
 ---------------------------+---------
     1721    100.00         |  XXXXXX
```

From this description, we see that the study lasted for six years, and there were only 52 children measured in all six study years. Most children (all but 17) were measured consecutively. All children were measured at least twice. We can use the xtsum command to give estimates of the mean math score, and its variability among schools and among children.

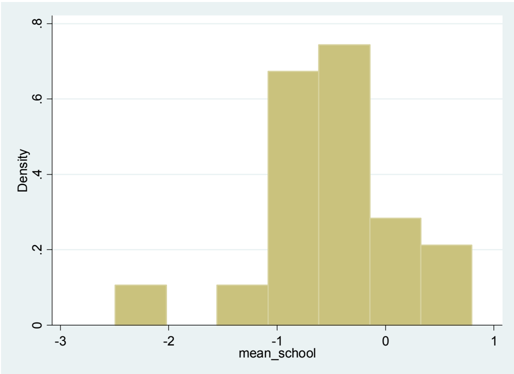
Let's ignore clustering due to subject for now:

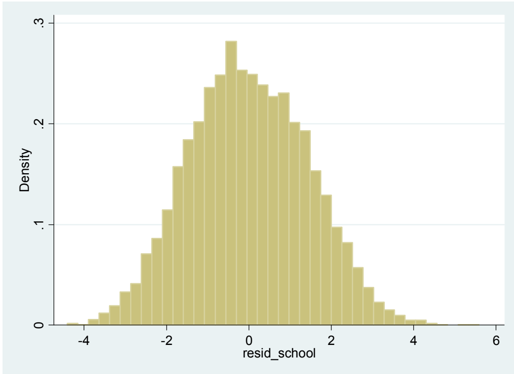
```
. xtsum math, i(school)

Variable         |      Mean   Std. Dev.       Min        Max |    Observations
-----------------+--------------------------------------------+----------------
math     overall | -.5369243   1.534696     -5.219      5.766 |     N =    7230
         between |             .6380456  -2.493857   .7969333 |     n =      60
         within  |             1.433215   -4.93981   4.795438 | T-bar =   120.5
```

Using `school` as the grouping variable, we note that the within school standard deviation (1.433) is much larger than the between school standard deviation (0.638). The within school variance is capturing both the variability among students at the same school and the variability among repeated observations on each student.

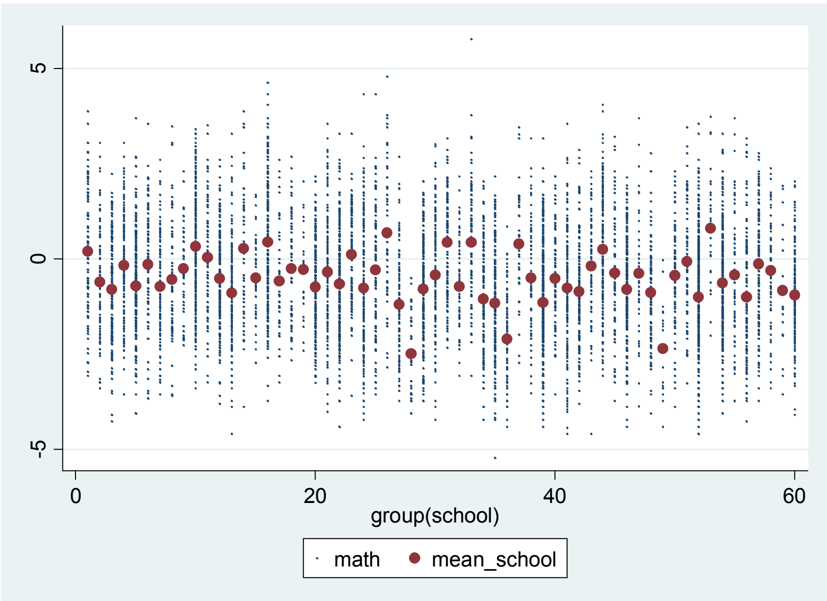
How are the above statistics calculated? Can we describe the above graphically?

```
. sort(school)
. by school: egen mean_school = mean(math)
. gen resid_school = math - mean_school
. by school: replace mean_school = . if _n > 1
```






```
. egen school_id = group(school)
. twoway (scatter math school_id, msymbol(p) )(scatter mean_school school_id)
```

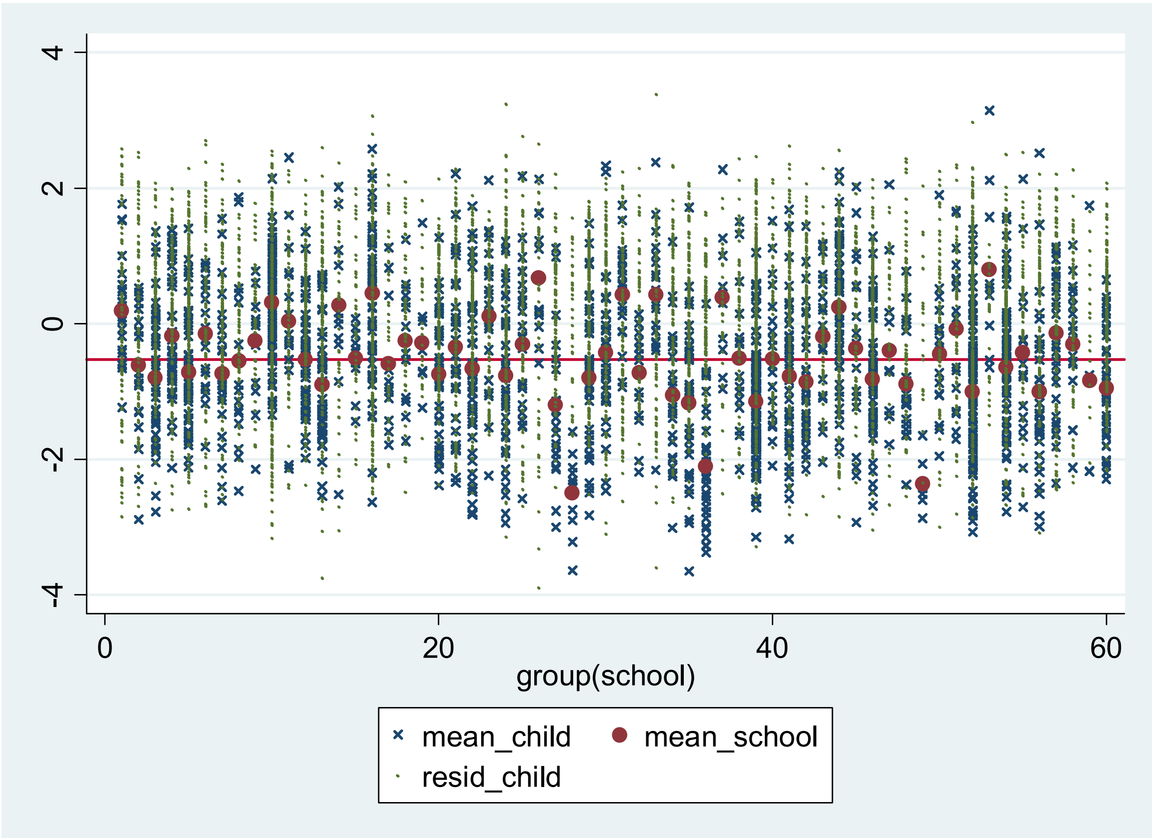


We can do the same by treating each child as a cluster.

```
. xtsum math, i(child)

Variable         |      Mean   Std. Dev.       Min        Max |    Observations
-----------------+--------------------------------------------+----------------
math     overall | -.5369243   1.534696     -5.219      5.766 |     N =    7230
         between |             1.121831    -3.6545      3.141 |     n =    1721
         within  |             1.076138  -4.435124   2.851075 | T-bar = 4.20105
```

Using `child` as the grouping variable, we can get a sense of what the within student variability looks like, but the between student variability doesn't take into account the fact that children are nested within schools. The between child standard deviation (1.122) captures both the variability between schools and the variability between students in the same school.

```
. sort child
. by child: egen mean_child = mean(math)
. gen resid_child = math - mean_child
. by child: replace mean_child = . if _n > 1
. egen child_id = group(child)
. twoway (scatter math child_id, msymbol(p) )(scatter mean_child child_id)
. twoway (scatter mean_child school_id, msymbol(x)) (scatter mean_school
school_id) (scatter resid_child school_id, msymbol(p)),  yline(-.53)
```



What are some of the "variations" (variance components) due to clustering shown in the above scatter plot?

Can we formulate a multi-level model that describes variation at different levels?

The ultimate goal is to examine covariate effects after accounting for variations due to clustering.

II. Two-level variance component with a random intercept for school

$$math_{ijk} = \beta_0 + U_i + \beta_1 year_{ijk} + \varepsilon_{ijk}$$

- *i* indexes school,
- *j* indexes child,
- *k* indexes observation.
- $U_i \sim N(0, \psi^{(3)})$ is a random intercept deviation for school *i*. The variance parameter $\psi^{(3)}$ has a superscript 3 to denote that it is the variance of a random effect at level three (school).
- $\varepsilon_{ijk} \sim N(0, \theta)$.

Interpretation for the coefficient on *year*?

```
. xtmixed math year || school:, nolog mle
Mixed-effects ML regression                     Number of obs      =      7230
Group variable: school                          Number of groups   =        60
                                                Obs per group: min =        18
                                                               avg =     120.5
                                                               max =       387
                                                Wald chi2(1)       =   7756.87
Log likelihood = -10343.209                     Prob > chi2        =    0.0000
------------------------------------------------------------------------------
        math |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        year |    .751992   .0085383    88.07   0.000     .7352573    .7687267
       _cons |  -.7699016   .0606114   -12.70   0.000    -.8886977   -.6511055
------------------------------------------------------------------------------
------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
school: Identity             |
                   sd(_cons) |   .4552395   .0441276      .3764702    .5504898
-----------------------------+------------------------------------------------
                sd(Residual) |   .9989248   .0083418      .9827082    1.015409
------------------------------------------------------------------------------
LR test vs. linear regression: chibar2(01) =  1235.39 Prob >= chibar2 = 0.0000
-----------------------------+------------------------------------------------

. estimates store models
```

1. Should _cons be -.53 (the overall math average) on our previous graph?
2. Try running the model without *year*. What estimates describe the between and within school variation?

We have two other ways to estimate the parameters of this model:

```
. xtreg math year, i(school) nolog mle

Random-effects ML regression                    Number of obs      =      7230
Group variable: school                          Number of groups   =        60
Random effects u_i ~ Gaussian                   Obs per group: min =        18
                                                               avg =     120.5
                                                               max =       387
                                                LR chi2(1)         =   5269.11
Log likelihood  = -10343.209                    Prob > chi2        =    0.0000
------------------------------------------------------------------------------
        math |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        year |    .751992   .0085409    88.05   0.000     .7352521     .768732
       _cons |  -.7699016   .0606118   -12.70   0.000    -.8886985   -.6511047
-------------+----------------------------------------------------------------
    /sigma_u |   .4552394   .0441276                      .3764702    .5504897
    /sigma_e |   .9989248   .0083418                      .9827082    1.015409
         rho |   .1719725   .0277211                      .1231205    .2316966
------------------------------------------------------------------------------
Likelihood-ratio test of sigma_u=0: chibar2(01)= 1235.39 Prob>=chibar2 = 0.000
```

```
. gllamm math year, i(school) nolog

number of level 1 units = 7230
number of level 2 units = 60
Condition Number = 1.8706763

log likelihood = -10355.384

------------------------------------------------------------------------------
        math |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        year |   .7523247   .0085099    88.41   0.000     .7356455    .7690038
       _cons |  -.7938479   .0151558   -52.38   0.000    -.8235527    -.764143
------------------------------------------------------------------------------

Variance at level 1
------------------------------------------------------------------------------
  1.0066166 (.01677245)
Variances and covariances of random effects
------------------------------------------------------------------------------
***level 2 (school)
    var(1): .20660673 (.0130242)
```

Note that xtreg and xtmixed used identical fitting procedures, and, accordingly, give identical results. Also note that from gllamm, the square root of the variance at level 1 sqrt(1.0066) = 1.0033 is equivalent to sigma_e which was estimated by xtmixed and xtreg to be sigma_e = 0.9989. If we compare the estimates of the sd of the random intercept for schools, we will see that gllamm estimated sqrt(.2066) = 0. 4545 while xtmixed and xtreg estimated 0.4552. These results are pretty close, but if we want gllamm to get a more precise estimate, we can specify nip() and adapt for more precise estimation (but at the expense of taking longer to run!)

```
. gllamm math year, i(school) nip(15) adapt

number of level 1 units = 7230
number of level 2 units = 60
Condition Number = 7.2589745

log likelihood = -10343.209

------------------------------------------------------------------------------
        math |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        year |    .751992   .0085409    88.05   0.000     .7352521    .7687319
       _cons |  -.7699022    .060612   -12.70   0.000    -.8886995   -.6511049
------------------------------------------------------------------------------
Variance at level 1
------------------------------------------------------------------------------
  .9978509 (.01666573)
Variances and covariances of random effects
------------------------------------------------------------------------------
***level 2 (school)
    var(1): .20724444 (.04017748)
------------------------------------------------------------------------------
```

Comparing our 'improved' gllamm estimates to the results from xtmixed and xtreg, we see that they are very similar. The sd of the random intercept is now estimated to be sqrt(.2072) = .4552, and the within school sd is estimated to be sqrt(.9979) = .9989.

This model assumes that math scores are a linear function of time and, ***conditional on a school and time***, the math scores within this school are independent. Perhaps this is not very reasonable because we know that there are students with repeated measures in each school!

III. Two-level variance component with a random intercept for child

$$math_{ijk} = \beta_0 + W_{ij} + \beta_1 year_{ijk} + \varepsilon_{ijk}$$

$W_{ij} \sim \mathrm{N}(0, \psi^{(2)})$ is a random intercept deviation for child *j* in school *i*. The variance parameter $\psi^{(2)}$ has a superscript 2 to denote that it is the variance of a random effect at level two (child).

```
. xtmixed math year || child:, nolog mle
Mixed-effects ML regression                     Number of obs      =      7230
Group variable: child                           Number of groups   =      1721
                                                Obs per group: min =         2
                                                               avg =       4.2
                                                               max =         6
                                                Wald chi2(1)       =  19156.93
Log likelihood = -8515.4377                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
        math |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        year |   .7474525   .0054003   138.41   0.000      .736868    .7580369
       _cons |  -.8386747     .02363   -35.49   0.000    -.8849885   -.7923608
------------------------------------------------------------------------------
```

```
------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
child: Identity              |
                   sd(_cons) |   .9315118   .0174854      .8978639    .9664207
-----------------------------+------------------------------------------------
                sd(Residual) |   .5890149   .0056113       .578119     .600116
------------------------------------------------------------------------------
LR test vs. linear regression: chibar2(01) =  4890.93 Prob >= chibar2 = 0.0000

. estimates store modelC
```

This model assumes that math scores are a linear function of time and, ***conditional on a child and time***, the repeated math scores are independent. This might be an okay model, but it doesn't take into account clustering of children by school.

IV. Three-level variance component, accounting for clustering of children within schools, including a random intercept for child and a random intercept for school

$$math_{ijk} = \beta_0 + U_i + W_{ij} + \beta_1 year_{ijk} + \varepsilon_{ijk}$$

$U_i \sim N(0, \psi^{(3)})$: random intercept deviation for school *i* from a typical (average) school.
$W_{ij} \sim N(0, \psi^{(2)})$: random intercept deviation for child *j* within school *i* from a typical child within school *i*. (i.e. $\beta_0 + U_i$)

```
. xtmixed math year|| school: || child:, nolog mle

Mixed-effects ML regression                     Number of obs      =      7230
-----------------------------------------------------------
                |   No. of       Observations per Group
 Group Variable |   Groups    Minimum    Average    Maximum
----------------+------------------------------------------
         school |       60         18      120.5        387
          child |     1721          2        4.2          6
-----------------------------------------------------------

                                                Wald chi2(1)       =  19120.98
Log likelihood = -8373.5216                     Prob > chi2        =    0.0000
------------------------------------------------------------------------------
        math |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        year |   .7461302   .0053958   138.28   0.000     .7355545    .7567058
       _cons |  -.7806069    .060579   -12.89   0.000    -.8993395   -.6618743
------------------------------------------------------------------------------
------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
school: Identity             |
                   sd(_cons) |   .4280823   .0462896      .3463257    .5291391
-----------------------------+------------------------------------------------
child: Identity              |
                   sd(_cons) |   .8184857   .0160566      .7876127    .8505689
-----------------------------+------------------------------------------------
                sd(Residual) |   .5890159   .0056111      .5781204    .6001168
------------------------------------------------------------------------------
LR test vs. linear regression:       chi2(2) =  5174.77   Prob > chi2 = 0.0000

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

```
. estimates store modelSC
```

Use a likelihood ratio test to test the null hypothesis that variance component for child is zero

```
. lrtest modelS modelSC

Likelihood-ratio test                              LR chibar2(01) =   3939.37
(Assumption: modelS nested in modelSC)             Prob > chibar2 =    0.0000
```

So we need to include the random intercept for child since with a p-value of <0.001 we reject the null hypothesis that the variance of the random intercept for child is zero.

We can also test the null hypothesis that the variance component for school is zero:

```
. lrtest modelC modelSC

Likelihood-ratio test                              LR chibar2(01) =    283.83
(Assumption: modelC nested in modelSC)             Prob > chibar2 =    0.0000
```

It follows that we also need to include the random intercept for school.

Intraclass Correlations:

1. $\rho(school) = \dfrac{\psi^{(3)}}{\psi^{(2)} + \psi^{(3)} + \theta}$ where θ is the variance of ε_{ijk} is defined to be the ICC between measurements from same school different child

```
. display .43^2 /(.43^2 + .82^2 + .59^2)
.15339306
```

2. $\rho(child, school) = \dfrac{\psi^{(2)} + \psi^{(3)}}{\psi^{(2)} + \psi^{(3)} + \theta}$ is the ICC between measurements from same school same child

```
. display (.43^2 + .82^2 )/(.43^2 + .82^2 + .59^2)
.71121619
```

Note that $\rho(child, school)$ is always greater than $\rho(school)$!!!

V. Incorporating covariates as fixed effects:

First we add child-level covariates:

```
. xtmixed math time female hispanic black || school: || child:, nolog mle

Mixed-effects ML regression                     Number of obs      =      7230

-----------------------------------------------------------
                |   No. of       Observations per Group
 Group Variable |   Groups    Minimum    Average    Maximum
----------------+------------------------------------------
         school |       60         18      120.5        387
          child |     1721          2        4.2          6
-----------------------------------------------------------
                                                Wald chi2(4)       =  19209.40
Log likelihood = -8343.9671                     Prob > chi2        =    0.0000
------------------------------------------------------------------------------
        math |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        year |   .7464291   .0053939   138.38   0.000     .7358573    .7570009
      female |  -.0029297   .0419391    -0.07   0.944    -.0851288    .0792695
    hispanic |  -.3624078   .0873684    -4.15   0.000    -.5336468   -.1911689
       black |   -.619737     .07792    -7.95   0.000    -.7724573   -.4670167
       _cons |  -.3400036   .0798143    -4.26   0.000    -.4964368   -.1835703
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
school: Identity             |
                   sd(_cons) |   .3508068   .0405635      .2796689    .4400397
-----------------------------+------------------------------------------------
child: Identity              |
                   sd(_cons) |   .8075023   .0158912      .7769492    .8392569
-----------------------------+------------------------------------------------
                sd(Residual) |   .5890221   .0056112      .5781264    .6001232
------------------------------------------------------------------------------
LR test vs. linear regression:       chi2(2) =  4692.66   Prob > chi2 = 0.0000

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

Note that the standard deviation of the random intercept for school decreases after controlling for these student-level characteristics. The student body of each of these schools must differ in terms of these student-level characteristics. Controlling for these student level characteristics removes some of the unexplained variability at the school-level that used to be explained by a larger variance of the random intercepts for school. Therefore we in this model the variance of the random intercepts for schools is smaller.

We will drop the child-level covariate female and add in some school-level covariates.

Incorporating school-level covariates as fixed effects:

```
. xtmixed math year hispanic black lowinc size mobility|| school: || child:,
mle

Performing EM optimization:
```

```
Performing gradient-based optimization:

Iteration 0:   log likelihood = -8328.2506
Iteration 1:   log likelihood = -8328.2506

Computing standard errors:

Mixed-effects ML regression                     Number of obs      =      7230

-----------------------------------------------------------
                |   No. of       Observations per Group
 Group Variable |   Groups    Minimum    Average    Maximum
----------------+------------------------------------------
         school |       60         18      120.5        387
          child |     1721          2        4.2          6
-----------------------------------------------------------

                                                Wald chi2(6)       =  19297.11
Log likelihood = -8328.2506                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
        math |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        year |   .7463203   .0053926   138.40   0.000     .7357511    .7568895
    hispanic |  -.2965714   .0877363    -3.38   0.001    -.4685315   -.1246113
       black |  -.5250947   .0786526    -6.68   0.000    -.6792509   -.3709385
      lowinc |  -.0052022   .0018149    -2.87   0.004    -.0087594   -.0016451
        size |  -.0000372    .000133    -0.28   0.780    -.0002978    .0002235
    mobility |  -.0120827   .0034534    -3.50   0.000    -.0188513   -.0053142
       _cons |   .4202531   .1428266     2.94   0.003     .1403182     .700188
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
school: Identity             |
                   sd(_cons) |   .2490489    .033076      .1919716    .3230964
-----------------------------+------------------------------------------------
child: Identity              |
                   sd(_cons) |   .8069863   .0158626      .7764875     .838683
-----------------------------+------------------------------------------------
                sd(Residual) |   .5890204   .0056112      .5781248    .6001214
------------------------------------------------------------------------------
LR test vs. linear regression:       chi2(2) =  4325.16   Prob > chi2 = 0.0000

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

The standard deviation of the random intercept for school decreases even more, so we have removed more of the unexplained variability between schools by controlling for these school-level covariates. If we add in enough school-level covariates so that, having controlled for all these school-level covariates, (i.e. controlling for all the school-level confounders) the sd of the random intercept for school is zero, we wouldn't need to include the random intercept for school. The random intercept for school in effect "mops up" unexplained variability between schools. When all the variability between schools is explained, we no longer need a random intercept for schools.

Recall that we can write the above model in separate levels where the cluster-level covariates directly model the random intercept components.

We can test if we should be including any variables that control for SES, at either the school or child level.

```
. test hispanic black lowinc mobility

 ( 1)  [math]hispanic = 0
 ( 2)  [math]black = 0
 ( 3)  [math]lowinc = 0
 ( 4)  [math]mobility = 0

           chi2(  4) =  113.60
         Prob > chi2 =    0.0000
```

So, we should be including at least one of the above variables. You could then test individually whether you need each variable by looking at the p-value in the regression output for the coefficient on each variable.

VI. Add in a random slope on year at the child level:

The corresponding equation is:

$$math_{ijk} = \beta_0 + U_i + W_{ij} + (\beta_1 + A_{ij})year_{ijk} + \beta_2 H_{ij} + \beta_3 B_{ij} + \beta_4 LI_i + \beta_5 M_i + \varepsilon_{ijk} \,,$$

where:

- A_{jk} is a random slope on time at the child-level,
- H_{ij} is the indicator for child j in school i being hispanic,
- B_{ij} is the indicator for child j in school i being black,
- LI_i is the low income percentage for school i,
- M_i is the proportion of children moving in school i,

The two child-level random effects are distributed multivariate normal,

$$(W_{ij}, A_{ij}) \sim \text{MVN}(0,\text{V}).$$

U_i is distributed as in earlier models. V describes the covariance between the W and A for each child.

We will first use the default for the covariance structure between random effects at the child level, (i.e., the child's random intercept and random slope on year are **independent** and V is the identity matrix.)

```
. xtmixed math year hispanic black lowinc mobility|| school: || child: year,
nolog mle

Mixed-effects ML regression                     Number of obs      =      7230

-----------------------------------------------------------
                |   No. of       Observations per Group
 Group Variable |   Groups    Minimum    Average    Maximum
----------------+------------------------------------------
```

```
      school |         60         18      120.5        387
       child |       1721          2        4.2          6
-----------------------------------------------------------

                                                Wald chi2(5)       =  13550.34
Log likelihood = -8250.3809                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
        math |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        year |   .7474726   .0064516   115.86   0.000     .7348277    .7601175
    hispanic |  -.3009228   .0869973    -3.46   0.001    -.4714343   -.1304113
       black |  -.5159753   .0782742    -6.59   0.000    -.6693899   -.3625607
      lowinc |  -.0050929   .0017953    -2.84   0.005    -.0086117   -.0015742
    mobility |  -.0120803   .0033991    -3.55   0.000    -.0187424   -.0054181
       _cons |   .3855081   .1320357     2.92   0.004      .126723    .6442933
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
school: Identity             |
                   sd(_cons) |   .2483921   .0328047      .1917437    .3217767
-----------------------------+------------------------------------------------
child: Independent           |
                    sd(year) |   .1526224   .0077812      .1381088    .1686613
                   sd(_cons) |   .8045489   .0157675      .7742311    .8360539
-----------------------------+------------------------------------------------
                sd(Residual) |   .5460603   .0060043       .534418    .5579562
------------------------------------------------------------------------------
LR test vs. linear regression:       chi2(3) =  4484.28   Prob > chi2 = 0.0000

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

Assess the goodness of fit of this model.

```
. estat ic

------------------------------------------------------------------------------
       Model |    Obs    ll(null)   ll(model)     df          AIC         BIC
-------------+----------------------------------------------------------------
           . |   7230           .   -8250.381     10     16520.76    16589.62
------------------------------------------------------------------------------
```

Second, allow for correlation between random effects at the child level, (i.e., the child's random intercept and random slope on year are now allowed to be **correlated** and V is unstructured.)

```
. xtmixed math time hispanic black grade lowinc mobility|| school: || child:
year, cov(unstructured) nolog mle

Mixed-effects ML regression                     Number of obs      =      7230

-----------------------------------------------------------
                |   No. of       Observations per Group
 Group Variable |   Groups    Minimum    Average    Maximum
----------------+------------------------------------------
         school |       60         18      120.5        387
          child |     1721          2        4.2          6
-----------------------------------------------------------

                                                Wald chi2(5)       =  13949.71
Log likelihood = -8212.9152                     Prob > chi2        =    0.0000
```

```
------------------------------------------------------------------------------
        math |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        year |   .7477856   .0063515   117.73   0.000     .7353369    .7602343
    hispanic |  -.3253004   .0850026    -3.83   0.000    -.4919024   -.1586983
       black |  -.4850724   .0765025    -6.34   0.000    -.6350146   -.3351302
      lowinc |  -.0040969   .0018017    -2.27   0.023    -.0076281   -.0005657
    mobility |  -.0114279   .0034265    -3.34   0.001    -.0181438   -.0047121
       _cons |   .2618913   .1326609     1.97   0.048     .0018807    .5219019
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
school: Identity             |
                   sd(_cons) |   .2534791   .0326095      .1969867    .3261725
-----------------------------+------------------------------------------------
child: Unstructured          |
                   sd(year)  |    .146334   .0079666      .1315239    .1628117
                   sd(_cons) |   .7914701   .0156571        .76137    .8227602
            corr(year,_cons) |   .4371882   .0503603      .3334645    .5304544
-----------------------------+------------------------------------------------
                sd(Residual) |   .5488879   .0060859      .5370885    .5609465
------------------------------------------------------------------------------
LR test vs. linear regression:       chi2(4) =  4559.21   Prob > chi2 = 0.0000

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

Assess the goodness of fit of this model.

```
. estat ic

------------------------------------------------------------------------------
       Model |    Obs    ll(null)   ll(model)     df          AIC         BIC
-------------+----------------------------------------------------------------
           . |   7230           .   -8212.915     11     16447.83    16523.58
------------------------------------------------------------------------------
```

Using AIC and BIC as model selection criteria, we would choose to stick with the unstructured correlation for the random effects on the child-level since this model has smaller AIC and BIC. (NOTE: In this case, we could have also looked at the regression output for corr(year, cons) and noted that the 95% CI doesn't contain zero, so we probably do want to allow for correlation between the two random effects.)

The correlation of 0.44 between the random effects on the child-level means that for children who tend to have a higher value of the random intercept (a higher baseline math score), they also tend to have a higher random slope on year (they improve math scores at a greater rate).

Another look at our most complicated model yet!

We have been working with the "one big model" form:

$$math_{ijk} = \beta_0 + U_i + W_{ij} + (\beta_1 + A_{ij})\, year_{ijk} + \beta_2 H_{ij} + \beta_3 B_{ij} + \beta_4 LI_i + \beta_5 M_i + \varepsilon_{ijk}$$

$U_i \sim N(0, \psi^{(3)})$
$W_{ij} \sim N(0, \psi^{(2)}{}_1)$ $\quad A_{ij} \sim N(0, \psi^{(2)}{}_2)$ $\quad cor(W_{ij}, A_{ij}) = \rho$
$\varepsilon_{ijk} \sim N(0, \psi^{(1)})$

The above model is equivalent to the random "intercept" and "slope" form:

$$math_{ijk} = \beta_{0,ij} + \beta_{1,ij}\, year_{ijk} + \beta_2 H_{ij} + \beta_3 B_{ij} + \beta_4 LI_i + \beta_5 M_i + \varepsilon_{ijk}$$

$\beta_{0,ij} \sim N(\beta_0, \psi^{(2)}{}_1 + \psi^{(3)})$ $\quad \beta_{1,ij} \sim N(\beta_1, \psi^{(2)}{}_2)$ $\quad cor(\beta_{0,ij}, \beta_{1,ij}) = \rho$
$\varepsilon_{ijk} \sim N(0, \psi^{(1)})$

The above model is also equivalent to the "multi-level" form:
(Note how the subscripts for the covariatline up nicely within each level)

School Level

$\beta_{0,i} = \beta_0 + U_i + \beta_4 LI_i + \beta_5 M_i,$ $\qquad U_i \sim N(0, \psi^{(3)})$

Child Level

$\beta_{0,ij} = \beta_{0,i} + W_{ij} + \beta_2 H_{ij} + \beta_3 B_{ij},$ $\qquad W_{ij} \sim N(0, \psi^{(2)}{}_1)$

$\beta_{1,ij} = \beta_1 + A_{ij},$ $\qquad A_{ij} \sim N(0, \psi^{(2)}{}_2)$

$cor(\beta_{0,ij}, \beta_{1,ij}) = \rho$

Observation Level

$math_{ijk} = \beta_{0,ij} + \beta_{1,ij}\, year_{ijk} + \varepsilon_{ijk},$ $\qquad \varepsilon_{ijk} \sim N(0, \psi^{(1)})$

We will see more models similar to this when we look at "cross-level interaction!"