

Lab 2: TWO-LEVEL NORMAL MODELS with school children popularity data

Purpose: Introduce basic two-level models for normally distributed responses using STATA. In particular, we discuss

- Random intercept models without covariates
- Random intercept models with covariates
- Random intercept models with covariates and random slopes
- Cross-level interactions

We will use the following 5 variables in the **popular.dta** dataset:

pupil: pupil identification number

school: school identification number

popular: the outcome variable ‘popularity’ (Y), measured by a self-rating scale that range from 0 (very unpopular) to 10 (very popular).

sex: the pupil sex, 0 – boy 1—girl

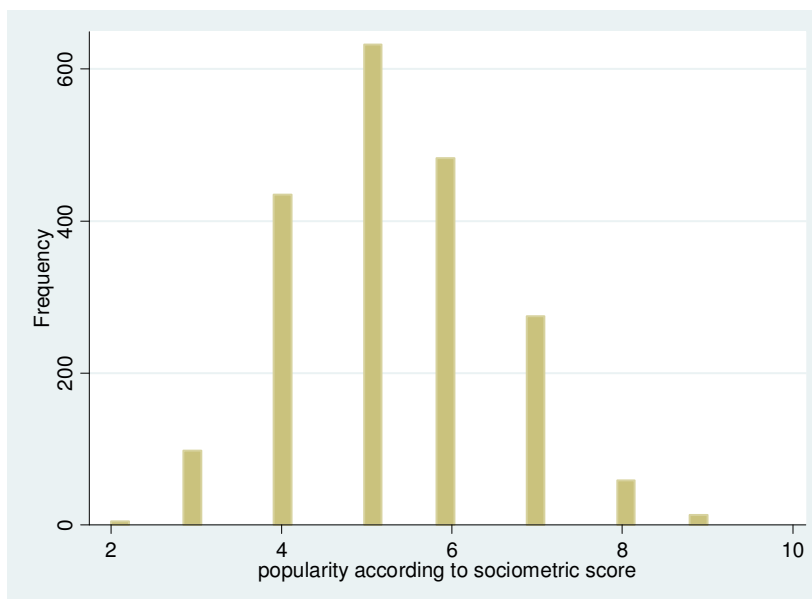
texp: teacher experience in years

The data are from 2000 pupils from 100 classes (each at a different school), the average number of students per school (class) is 20. Therefore, we have pupils nested within schools, and we need to account for the possible correlation between pupils in the same school in our model.

Initial EDA:

Our outcome variable of interest is self-reported popularity score:

```
. hist popular, freq
(bin=33, start=2, width=.21212121)
```



Q1. What is the average self-rating score?

Two-stage model of the popularity score of pupil (j) in school (i)

MODEL 1:

The **intercept-only (variance components)** model:

$$\begin{aligned} popular_{ij} | U_i &= \beta_0 + U_i + \varepsilon_{ij} \\ \varepsilon_{ij} &\sim N(0, \sigma^2) \\ U_i &\sim N(0, \tau^2) \end{aligned}$$

Interpretations of model parameters:

$\beta_0 + U_i$: average score for pupils in school i

β_0 : average score for pupils in a typical school ($U_i=0$)

U_i : school-level random intercept (random effect). Represents the difference between the average popularity score for a specific school(i) and the average popularity score of a typical ($U_i=0$) school.

τ^2 : variance of the random intercept. Represents the variability (or amount of dispersion) in specific schools' average scores around the average score of the typical school.

ε_{ij} : difference between the popularity score for child(ij) and the average popularity score in school(i)

σ^2 : variance of the error. Represents the variability of individual children's scores in school(i) around the average score for school(i).

(At least) three ways to fit this multilevel model for a continuous outcome using STATA, in order of generality:

1. xtreg, re mle

- estimates model parameters by finding the maximum of a closed form solution for the likelihood (fast)
- can only do two level models
- can only do a random intercept, not random slopes
- only for continuous outcomes
- using the pa option for xtreg is equivalent to using xtgee, so with xtreg you can easily switch from doing a cluster-specific random intercept model to doing a marginal GEE model
- automatically gives you the ICC (ρ) in the output
- by default, returns the **sd** of the random intercept

2. xtmixed, mle

- estimates model parameters by finding the maximum of a closed form solution for the likelihood (fast)
- very close to being equivalent to xtreg, re mle
- as many levels as you want (within reason!!)
- can do random intercepts and slopes
- only for continuous outcomes
- by default returns the **sd** of the random effects

3. gllamm

- estimates model parameters using quadrature methods to approximate the likelihood function, and then finds the maximum of the approximated likelihood (slow!)
- results won't be exactly the same as xtmixed, but they should be close. You can increase nip() and use the adapt option to get better estimates in gllamm (even slower!)
- can do many levels
- can do random intercepts and slopes
- for many types of outcomes (continuous, binary, poisson, etc...)
- by default returns the **variance** of the random effects

For continuous outcomes, xtmixed is preferred because it uses a closed form solution for the likelihood and hence produces accurate results relatively quickly.

Syntax for fitting the intercept-only model

xtreg

```
. xtreg popular, re i(school) mle
```

```
Random-effects ML regression          Number of obs    =    2000
Group variable: school                Number of groups =    100

Random effects u_i ~ Gaussian        Obs per group: min =    16
                                      avg =    20.0
                                      max =    26

Log likelihood = -2556.3612          Wald chi2(0)     =    0.00
                                      Prob > chi2      =    .
```

popular	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
_cons	5.307603	.0950194	55.86	0.000	5.121369 5.493838
/sigma_u	.9331052	.0684368			.8081665 1.077359
/sigma_e	.7991726	.0129644			.7741625 .8249907
rho	.5768565	.0367346			.5039739 .6471936

Likelihood-ratio test of sigma_u=0: chibar2(01) = 1376.81 Prob>=chibar2 = 0.000

xtmixed

```
. xtmixed popular || school:, mle
```

```
Mixed-effects ML regression          Number of obs    =    2000
Group variable: school                Number of groups =    100

Obs per group: min =    16
                  avg =    20.0
                  max =    26

Log likelihood = -2556.3612          Wald chi2(0)     =    .
                                      Prob > chi2      =    .
```

popular	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
_cons	5.307603	.0950231	55.86	0.000	5.121361 5.493845

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]
school: Identity			
sd(_cons)	.9331053	.0684433	.8081556 1.077374
sd(Residual)	.7991726	.0129645	.7741624 .8249907

LR test vs. linear regression: chibar2(01) = 1376.81 Prob >= chibar2 = 0.0000

gllamm

. gllamm popular, i(school) adapt

Running adaptive quadrature

```
Iteration 0: log likelihood = -2860.4937
Iteration 1: log likelihood = -2772.8365
Iteration 2: log likelihood = -2556.54
Iteration 3: log likelihood = -2556.3615
Iteration 4: log likelihood = -2556.3613
```

Adaptive quadrature has converged, running Newton-Raphson

```
Iteration 0: log likelihood = -2556.3613
Iteration 1: log likelihood = -2556.3613 (backed up)
Iteration 2: log likelihood = -2556.3612
```

```
number of level 1 units = 2000
number of level 2 units = 100
```

```
Condition Number = 5.8577667
gllamm model
log likelihood = -2556.3612
```

popular	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
_cons	5.307603	.0950232	55.86	0.000	5.121361 5.493845

Variance at level 1

```
.63867684 (.02072167)
```

Variances and covariances of random effects

```
***level 2 (school)
var(1): .87068733 (.12772979)
```

xtmixed can give you the variance of the error and the random intercept, not the sd

. xtmixed popular || school:, mle var

popular	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
_cons	5.307603	.0950231	55.86	0.000	5.121361 5.493845

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]
school: Identity			
var(_cons)	.8706855	.1277297	.6531154 1.160734
var(Residual)	.6386768	.0207217	.5993275 .6806097

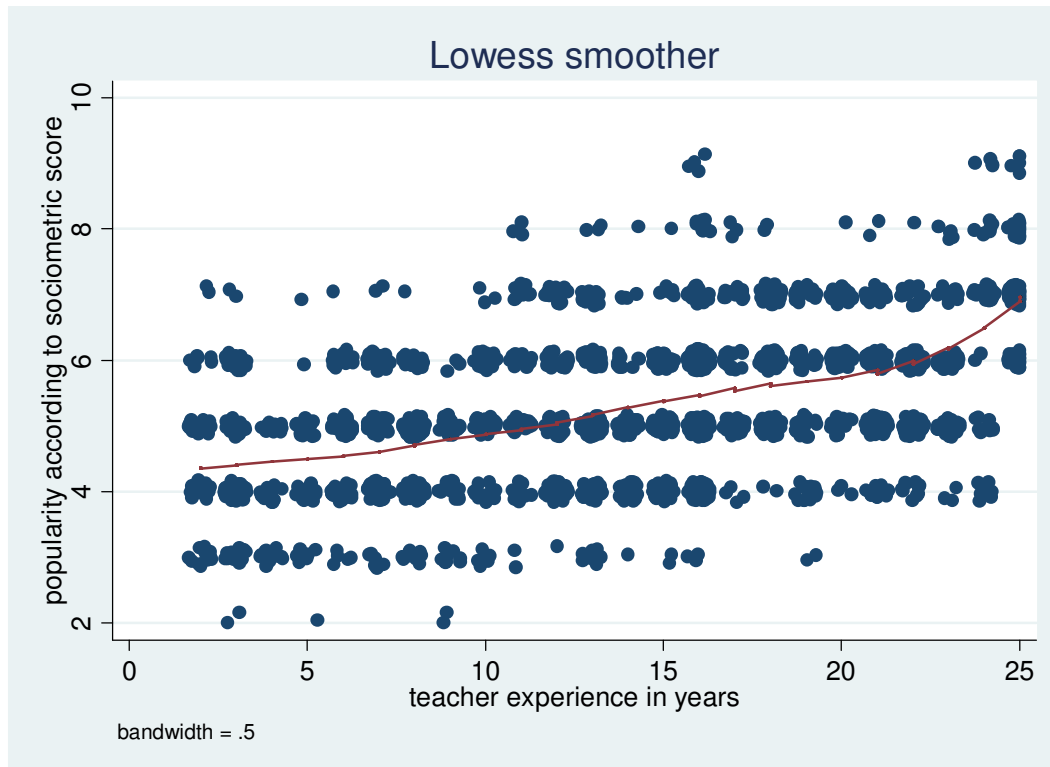
```
LR test vs. linear regression: chibar2(01) = 1376.81 Prob >= chibar2 = 0.0000
```

Q2. Do gender and teaching experience affect the self-rating score?

Gender is a level-1 covariate (it varies for each child (j) nested in school (i)). Teaching experience is a level-2 covariate since it is fixed for all students in a given class (school).

Exploratory analysis of the relationship between `texp` and popularity *ignoring clustering*

```
. lowess popular texp, bw(0.5) jitter(3)
```



But what we'll really be modeling is how the average popularity scores of children in a given school vary according to the covariate values of student gender and teacher's experience since we have the model:

MODEL 2:

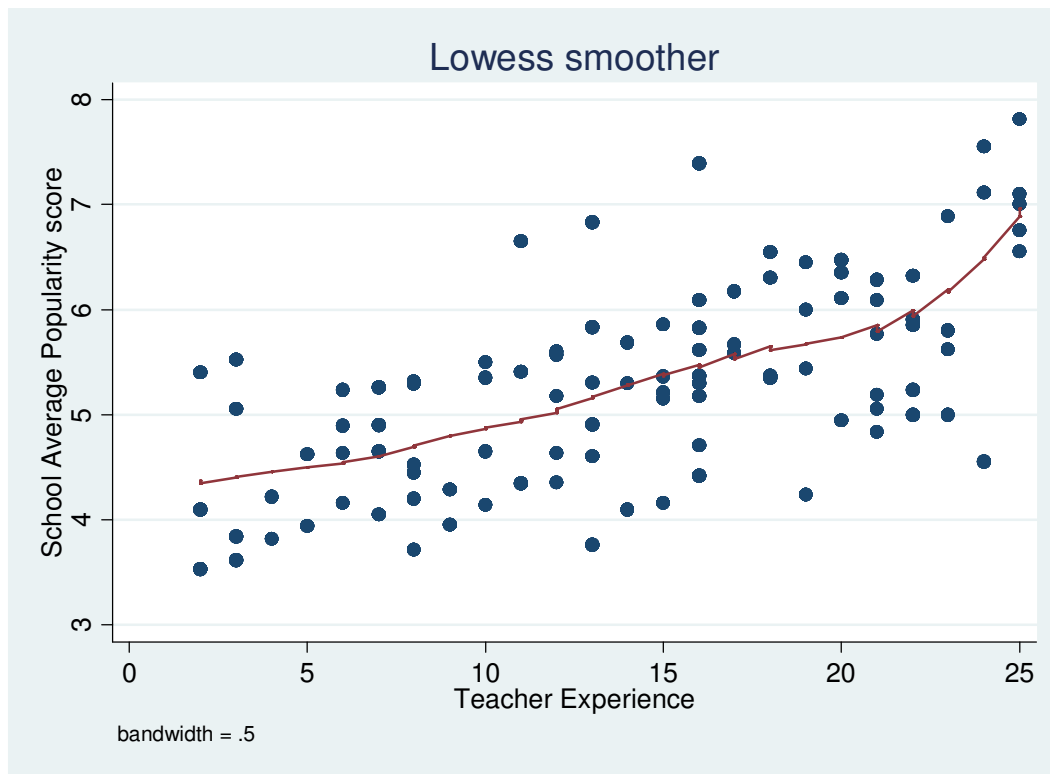
$$popular_{ij} | U_i, girl_{ij}, texp_i = (\beta_0 + U_i) + \beta_1 girl_{ij} + \beta_2 texp_i + \varepsilon_{ij}$$

Or, in other words,

$$E(popular_{ij} | U_i, girl_{ij}, texp_i) = (\beta_0 + U_i) + \beta_1 girl_{ij} + \beta_2 texp_i$$

A more informative plot might look at the relationship between the **mean** popularity score at each school and teacher experience.

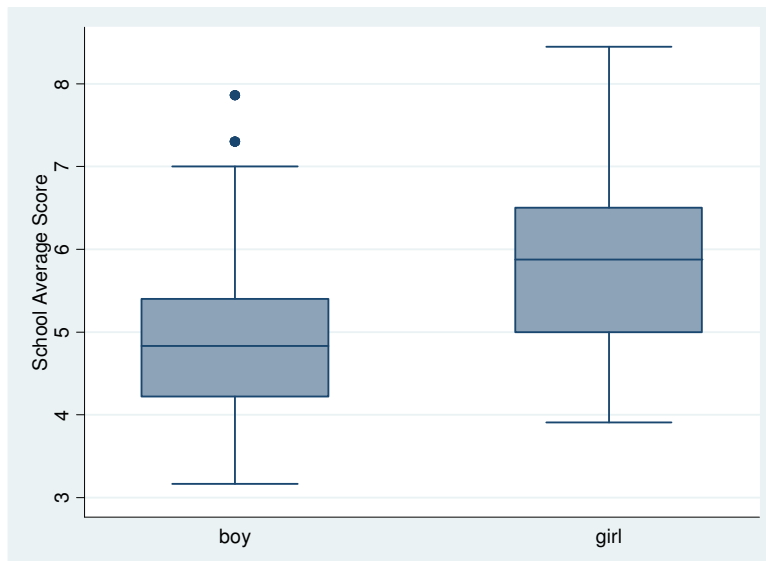
```
. sort school
. by school: egen mscore_sch=mean(popular)
. lowess mscore_sch texp, ytitle("School Average Popularity score")
xtitle("Teacher Experience") bw(0> .5)
```



We find a positive association between the teachers' experience and the average popularity score for each school.

Next we explore the relationship between the **mean** popularity score at each school for girls versus boys.

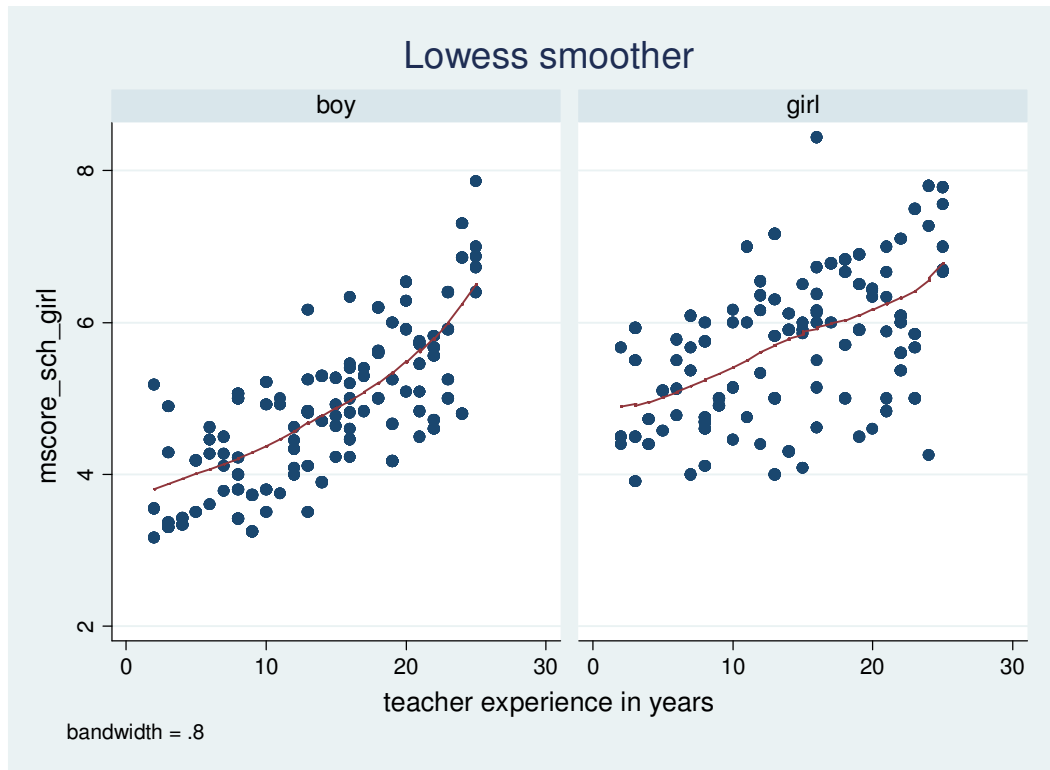
```
. sort school girl
. by school girl: egen mscore_sch_girl = mean(popular)
. sort girl
. graph box mscore_sch_girl, over(girl) ytitle("School Average Score")
```



We find the average reported scores for girls are higher than for boys.

Do we see evidence of an interaction?

```
. lowess mscore_sch_girl texp, by(girl)
```



The slopes are relatively similar, but the two curves definitely start at different points, so for this example we'll continue by just including a linear term for teacher's experience and an indicator variable for pupil gender, but not an interaction term (an interaction term would allow the slope on the relation between average popularity score and texp to vary according gender).

Random intercept model with fixed effects for texp and girl

```
. xtmixed popular texp girl || school:, mle
```

```
Mixed-effects ML regression
Group variable: school
Number of obs      =      2000
Number of groups   =       100

Obs per group: min =       16
                avg  =      20.0
                max  =       26

Wald chi2(2)      =      820.79
Prob > chi2       =      0.0000

Log likelihood = -2214.2871
```

popular	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
texp	.0934455	.0107449	8.70	0.000	.0723858 .1145052
girl	.8447125	.0309418	27.30	0.000	.7840676 .9053574
_cons	3.560661	.1697652	20.97	0.000	3.227927 3.893394

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
school: Identity				
sd(_cons)	.6897604	.0511783	.5964053	.7977284
sd(Residual)	.6779987	.0109988	.6567804	.6999024

LR test vs. linear regression: chibar2(01) = 1107.01 Prob >= chibar2 = 0.0000

Interpretation of the coefficient on 'texp':

For a given school, we estimate that increasing the teacher's experience by one year would result in average popularity scores that are .09 points higher for both male and female students.

Interpretation of the coefficient on 'girl':

Within a school, we estimate that the average popularity scores are .845 points higher for girls than for boys, controlling for teacher's experience.

Q3. Does the difference in average self-rating score between female and male students vary across schools?

To address this question using EDA, we'll first create a variable that contains the mean scores for boys for each school and then another variable that contains the mean score for girls for each school.

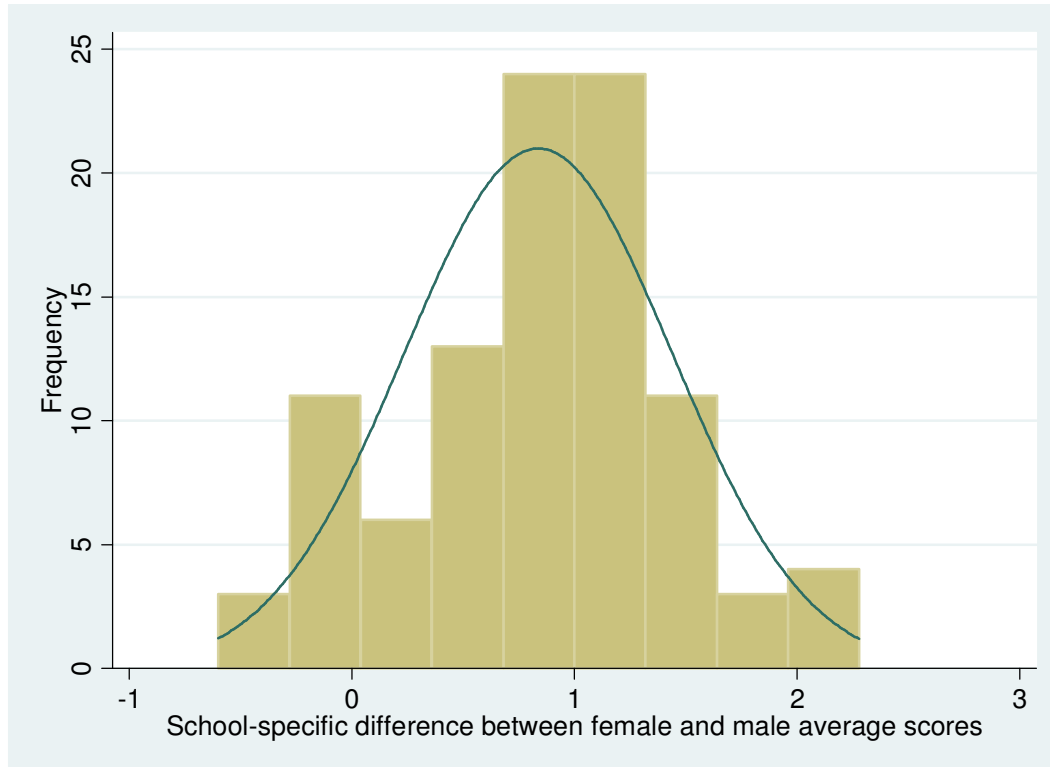
```
. sort school girl
. by school: gen meanpop_boys = mscore_sch_girl[1]
. gen boy = 1-girl
. sort school boy
. by school: gen meanpop_girls = mscore_sch_girl[1]
```

We'll subtract the average boys score from the average females score to get the difference in scores across genders for each school.

```
. gen genddiff = meanpop_girls - meanpop_boys
. replace genddiff =. if pupil!=1
```

We then make a histogram of the gender differences for each school to see how much heterogeneity (or variability) we have between schools in our gender difference.

```
. hist genddiff, norm freq xtitle(School-specific difference between
female and male average scores)
```



We do see quite a bit of heterogeneity in the gender differences across the different school. We can build the heterogeneity of the gender effect into our model:

Random intercept model with fixed effects for *texp* and *girl* and a random coefficient on pupil gender

MODEL 3:

$$popular_{ij} | U_{0i}, U_{1i}, girl_{ij}, texp_i = (\beta_0 + U_{0i}) + (\beta_1 + U_{1i})girl_{ij} + \beta_2 texp_i + \varepsilon_{ij}$$

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

$$\begin{pmatrix} U_{0i} \\ U_{1i} \end{pmatrix} \sim MVN(0, \Sigma)$$

In the above model, popularity score is function of pupil gender and teaching experience. We allow for different baseline scores for different schools by using a *random intercept*, and we allow different gender effects for different schools by using a *random slope* for gender.

Note: We could NOT fit a model that includes a random slope on teaching experience because teaching experience does not vary within school.


```
. gen girlXtexp = girl*texp
```

```
. xtmixed popular texp girl girlXtexp || school: girl, cov(unstr) mle
```

```
Mixed-effects ML regression      Number of obs      =      2000
Group variable: school           Number of groups   =       100

                                Obs per group: min =       16
                                avg =       20.0
                                max =       26

                                Wald chi2(3)      =     365.74
Log likelihood = -2122.925        Prob > chi2        =     0.0000
```

popular	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
texp	.1102293	.0101287	10.88	0.000	.0903774 .1300811
girl	1.329479	.1317029	10.09	0.000	1.071346 1.587612
girlXtexp	-.0340251	.0083716	-4.06	0.000	-.0504331 -.0176172
_cons	3.313651	.1593869	20.79	0.000	3.001258 3.626044

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]
school: Unstructured			
sd(girl)	.4692521	.0458652	.3874439 .5683341
sd(_cons)	.6347378	.0495438	.5446967 .7396631
corr(girl,_cons)	.0798403	.1247735	-.1645989 .3150401
sd(Residual)	.626432	.0104426	.6062956 .6472371

```
LR test vs. linear regression:      chi2(3) = 1269.28  Prob > chi2 = 0.0000
```

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

The estimate of fixed coefficient on teacher's experience is similar for both model 3 and model 4. However, the regression slope for pupil gender is considerable larger in model 4 (the model with the cross-level interaction between gender and teacher's experience) and this coefficient now has the interpretation of being the school-specific difference in average scores between genders when the teacher's experience is zero for a typical school.

The coefficient on the interaction between gender and teacher experience is estimated as -0.03, which is statistically significant. The negative value means the difference between girls and boys is smaller with more experienced teachers. The standard deviation of the random slope on gender decreases from 0.52 to 0.47, which means that the variation in teacher experience between schools explains some of the variability between schools in the estimated gender difference in self-reported popularity scores.