PRINCETON UNIVERSITY

Multilevel Analysis

(ver. 1.0)

Oscar Torres-Reyna
Data Consultant
otorres@princeton.edu

http://dss.princeton.edu/training/

Use multilevel model whenever your data is grouped (or nested) in more than one category (for example, states, countries, etc).

Multilevel models allow:

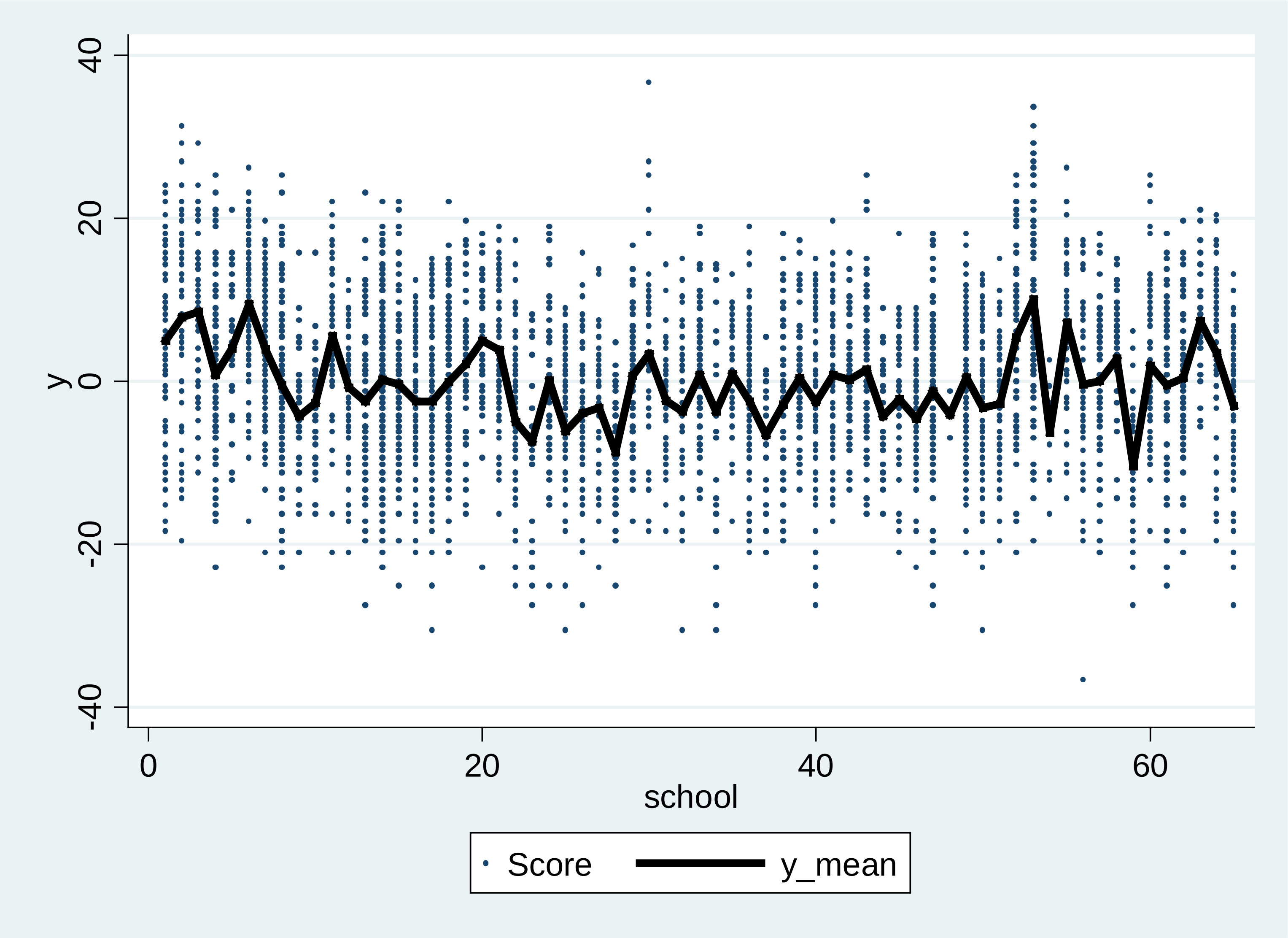
- Study effects that vary by entity (or groups)
- Estimate group level averages

Some advantages:

- Regular regression ignores the average variation between entities.
- Individual regression may face sample problems and lack of generalization

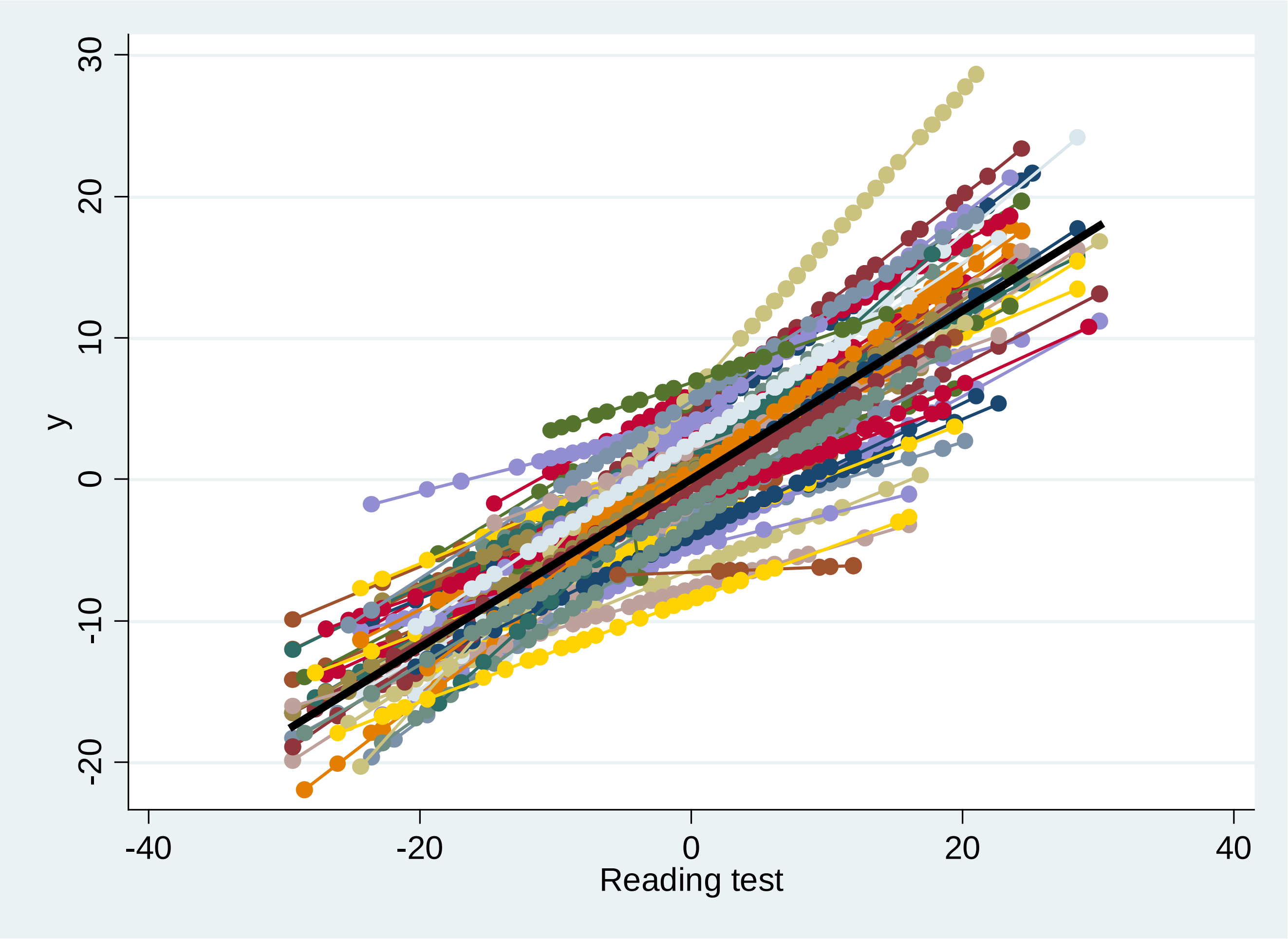
Variation between entities

```
use http://dss.princeton.edu/training/schools.dta
bysort school: egen y_mean=mean(y)
twoway scatter y school, msize(tiny)  || connected y_mean school, connect(L)
clwidth(thick) clcolor(black) mcolor(black) msymbol(none) || , ytitle(y)
```



```
statsby inter=_b[_cons] slope=_b[x1], by(school) saving(ols, replace): regress y x1
sort school
merge school using ols
drop _merge
gen yhat_ols = inter + slope*x1
sort school x1
separate y, by(school)
separate yhat_ols, by(school)
twoway connected yhat_ols1-yhat_ols65 x1 || lfit y x1, clwidth(thick) clcolor(black)
legend(off) ytitle(y)
```

Individual regressions (no-pooling approach)



NOTE: Stata 13, the command changed to "mixed"

Varying-intercept model (null)

$$y_i = \alpha_{j[i]} + \varepsilon_i$$

```
. xtmixed y || school: , mle nolog

Mixed-effects ML regression                     Number of obs      =      4059
Group variable: school                          Number of groups   =        65

                                                Obs per group: min =         2
                                                               avg =      62.4
                                                               max =       198

                                                Wald chi2(0)       =         .
Log likelihood = -14851.502                     Prob > chi2        =         .

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |  -.1317104    .536272    -0.25   0.806    -1.182784    .9193634
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
school: Identity             |
                   sd(_cons) |   4.106553   .3999163      3.392995    4.970174
-----------------------------+------------------------------------------------
                sd(Residual) |   9.207357   .1030214      9.007636    9.411505
------------------------------------------------------------------------------
LR test vs. linear regression: chibar2(01) =   498.72 Prob >= chibar2 = 0.0000
```

Mean of state level intercepts

Standard deviation at the school level (level 2)

Standard deviation at the individual level (level 2)

Ho: Random-effects = 0

$$Intraclass\_correlation = \frac{(sigma\_u)^2}{(sigma\_u)^2 + (sigma\_e)^2} = \frac{sd(\_cons)^2}{sd(\_cons)^2 + sd(residual)^2} = \frac{4.11^2}{4.11^2 + 9.21^2} = 0.17$$

If the interclass correlation (IC) approaches 0 then the grouping by counties (or entities) are of no use (you may as well run a simple regression). If the IC approaches 1 then there is no variance to explain at the individual level, everybody is the same.

“An intraclass correlation tells you about the correlation of the observations (cases) within a cluster” (http://www.ats.ucla.edu/stat/Stata/Library/cpsu.htm)

Varying-intercept model (one level-1 predictor) $y_i = \alpha_{j[i]} + \beta x_i + \varepsilon_i$

```
. xtmixed y x1 || school: , mle nolog

Mixed-effects ML regression                     Number of obs      =      4059
Group variable: school                          Number of groups   =        65

                                                Obs per group: min =         2
                                                               avg =      62.4
                                                               max =       198

                                                Wald chi2(1)       =   2042.57
Log likelihood = -14024.799                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .5633697   .0124654    45.19   0.000     .5389381    .5878014
       _cons |   .0238706   .4002258     0.06   0.952    -.7605576    .8082987
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
school: Identity             |
                   sd(_cons) |   3.035271   .3052516      2.492262     3.69659
-----------------------------+------------------------------------------------
                sd(Residual) |   7.521481   .0841759      7.358295    7.688285
------------------------------------------------------------------------------
LR test vs. linear regression: chibar2(01) =   403.27 Prob >= chibar2 = 0.0000
```

Mean of state level intercepts

Standard deviation at the school level (level 2)

Standard deviation at the individual level (level 2)

Ho: Random-effects = 0

$$Intraclass\_correlation = \frac{(sigma\_u)^2}{(sigma\_u)^2 + (sigma\_e)^2} = \frac{sd(\_cons)^2}{sd(\_cons)^2 + sd(residual)^2} = \frac{3.03^2}{3.03^2 + 7.52^2} = 0.14$$

If the interclass correlation (IC) approaches 0 then the grouping by counties (or entities) are of no use (you may as well run a simple regression). If the IC approaches 1 then there is no variance to explain at the individual level, everybody is the same.

"An intraclass correlation tells you about the correlation of the observations (cases) within a cluster" (http://www.ats.ucla.edu/stat/Stata/Library/cpsu.htm)

Varying-intercept, varying-coefficient model $y_i = \alpha_{j[i]} + \beta_{j[i]} x_i + \varepsilon_i$

```
. xtmixed y x1 || school: x1, mle nolog covariance(unstructure)

Mixed-effects ML regression                     Number of obs      =      4059
Group variable: school                          Number of groups   =        65

                                                Obs per group: min =         2
                                                               avg =      62.4
                                                               max =       198

                                                Wald chi2(1)       =    779.80
Log likelihood = -14004.613                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .5567291   .0199367    27.92   0.000     .5176539    .5958043
       _cons |  -.1150841   .3978336    -0.29   0.772    -.8948236    .6646554
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
school: Unstructured         |
                      sd(x1) |   .1205631   .0189827      .0885508    .1641483
                   sd(_cons) |   3.007436   .3044138      2.466252    3.667375
              corr(x1,_cons) |   .4975474   .1487416      .1572843    .7322131
-----------------------------+------------------------------------------------
                sd(Residual) |   7.440788   .0839482      7.278059    7.607157
------------------------------------------------------------------------------
LR test vs. linear regression:       chi2(3) =   443.64   Prob > chi2 = 0.0000

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

Mean of state level intercepts

Standard deviation at the school level (level 2)

Standard deviation at the individual level (level 2)

Ho: Random-effects = 0

$$Intraclass\_correlation = \frac{(sigma\_u)^2}{(sigma\_u)^2 + (sigma\_e)^2} = \frac{sd(\_cons)^2 + sd(x1)}{sd(\_cons)^2 + sd(x1) + sd(residual)^2} = \frac{0.12^2 + 3.01^2}{0.12^2 + 3.01^2 + 7.44^2} = 0.14$$

Varying-slope model

$$y_i = \alpha + \beta_{j[i]} x_i + \varepsilon_i$$

```
. xtmixed y x1 || _all: R.x1, mle nolog

Mixed-effects ML regression                     Number of obs      =      4059
Group variable: _all                            Number of groups   =         1

                                                Obs per group: min =      4059
                                                               avg =    4059.0
                                                               max =      4059


                                                Wald chi2(1)       =   2186.09
Log likelihood = -14226.433                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .5950551   .0127269    46.76   0.000     .5701108    .6199995
       _cons |  -.011948    .1263914    -0.09   0.925    -.2596706    .2357746
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
_all: Identity               |
                   sd(R.x1)  |   .0003388   .1806391             0           .
-----------------------------+------------------------------------------------
               sd(Residual)  |   8.052417    .089372      7.879142    8.229502
------------------------------------------------------------------------------
LR test vs. linear regression: chibar2(01) =     0.00 Prob >= chibar2 = 1.0000
```

Mean of state level intercepts

Standard deviation at the school level (level 2)

Standard deviation at the individual level (level 2)

Postestimation

Comparing models using likelihood-ration test

Use the likelihood-ratio test (`lrtest`) to compare models fitted by maximum likelihood. This test compares the “log likelihood” (shown in the output) of two models and tests whether they are significantly different.

```
/*Fitting random intercepts and storing results*/
quietly xtmixed y x1 || school: , mle nolog
estimates store ri

/*Fitting random coefficients and storing results*/
quietly xtmixed y x1 || school: x1, mle nolog covariance(unstructure)
estimates store rc

/*Running the likelihood-ratio test to compare*/
lrtest ri rc
```

```
. lrtest ri rc

Likelihood-ratio test                                 LR chi2(2)  =     40.37
(Assumption: ri nested in rc)                         Prob > chi2 =    0.0000

Note: LR test is conservative
```

The null hypothesis is that there is no significant difference between the two models. If `Prob>chi2<0.05`, then you may reject the null and conclude that there is a statistically significant difference between the models. In the example above we reject the null and conclude that the random coefficients model provides a better fit (it has the lowest log likelihood)

Varying-intercept, varying-coefficient model: postestimation

```
. xtmixed y x1 || school: x1, mle nolog covariance(unstructured) variance

Mixed-effects ML regression                     Number of obs      =      4059
Group variable: school                          Number of groups   =        65

                                                Obs per group: min =         2
                                                               avg =      62.4
                                                               max =       198

                                                Wald chi2(1)       =    779.80
Log likelihood = -14004.613                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .5567291   .0199367    27.92   0.000     .5176539    .5958043
       _cons |  -.1150841   .3978336    -0.29   0.772    -.8948236    .6646554
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
school: Unstructured         |
                     var(x1) |   .0145355   .0045772      .0078412    .0269446
                  var(_cons) |    9.04467    1.83101      6.082398    13.44964
                cov(x1,_cons)|   .1804036   .0691515      .0448692     .315938
-----------------------------+------------------------------------------------
               var(Residual) |   55.36533   1.249282      52.97014    57.86883
------------------------------------------------------------------------------
LR test vs. linear regression:       chi2(3) =   443.64   Prob > chi2 = 0.0000

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

Mean of state level intercepts

Standard deviation at the school level (level 2)

Standard deviation at the individual level (level 2)

$$Intraclass\_correlation = \frac{(sigma\_u)}{(sigma\_u)+(sigma\_e)} = \frac{\text{var}(\_cons)+\text{var}(x1)}{\text{var}(\_cons)^2+\text{var}(x1)+\text{var}(residual)} = \frac{0.014+9.045}{0.014+9.045+55.365} = 0.14$$

Postestimation: variance-covariance matrix

```
. xtmixed y x1 || school: x1, mle nolog covariance(unstructured) variance

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
school: Unstructured         |
                     var(x1) |   .0145355   .0045772      .0078412    .0269446
                  var(_cons) |    9.04467    1.83101      6.082398    13.44964
                cov(x1,_cons)|   .1804036   .0691515      .0448692     .315938
-----------------------------+------------------------------------------------
               var(Residual) |   55.36533   1.249282      52.97014    57.86883
------------------------------------------------------------------------------
LR test vs. linear regression:       chi2(3) =   443.64   Prob > chi2 = 0.0000

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

```
. estat recovariance

Random-effects covariance matrix for level school

             |        x1      _cons
-------------+----------------------
          x1 |  .0145355
       _cons |  .1804036    9.04467
```

Variance-covariance matrix

```
. estat recovariance, correlation

Random-effects correlation matrix for level school

             |        x1      _cons
-------------+----------------------
          x1 |         1
       _cons |  .4975474          1
```

The correlation between the intercept and x1 shows a close relationship between the average of y and x1.

Postestimation: estimating random effects (group-level errors)

$$y_i = \alpha_{j[i]} + \beta_{j[i]} x_i + \varepsilon_i \quad \Rightarrow \quad y_i = \underbrace{\alpha_{j[i]} + \beta_{j[i]} x_i}_{\text{Fixed-effects}} + \underbrace{u_{\alpha_i} + u_{\beta_{j[i]}}}_{\text{Random-effects}} + \varepsilon_i$$

To estimate the random effects *u*, use the command `predict` with the option `reffects`, this will give you the best linear unbiased predictions (BLUPs) of the random effects which basically show the amount of variation for both the intercept and the estimated beta coefficient(s). After running `xtmixed`, type

```
predict u*, reffects
```

Two new variables are created

u1 "BLUP r.e. for school: x1" ------- /* u_β */
u2 "BLUP r.e. for school: _cons" --- /* u_α */

Postestimation: estimating random effects (group-level errors)

$$y_i = -0.12 + 0.56x1 \rightarrow y_i = \underbrace{-0.12 + 0.56x1}_{\text{Fixed-effects}} + \underbrace{u_\alpha + u_\beta}_{\text{Random-effects}}$$

To explore some results type:

```
bysort school: generate groups=(_n==1)  /*_n==1 selects the first
                                         case of each group */
list school u2 u1 if school<=10 & groups
```

```
. list school u2 u1 if school<=10 & groups
```

	school	u2	u1
1.	1	3.749336	.1249755
74.	2	4.702129	.1647261
129.	3	4.79768	.0808666
181.	4	.3502505	.1271821
260.	5	2.462805	.0720576
295.	6	5.183809	.0586242
375.	7	3.640942	-.1488697
463.	8	-.121886	.0068855
565.	9	-1.767982	-.0886194
599.	10	-3.139076	-.1360763

Here *u2* and *u1* are the group level errors for the intercept and the slope respectively. For the first school the equation would be:

$$y_1 = -0.12 + 0.56x1 + 3.75 + 0.12 = (-0.12 + 3.75) + (0.56 + 0.12)x1 = 3.63 + 0.68x1$$

$$y_1 = -0.12 + 0.56x1 + 3.75 + 0.12 = (-0.12 + 3.75) + (0.56 + 0.12)x1 = 3.63 + 0.68x1$$

To estimate intercepts and slopes per school type :

```
gen intercept = _b[_cons] + u2
gen slope = _b[x1] + u1
list school intercept slope if school<=10 & groups
```

Compare the coefficients for school 1 above

```
. list school intercept slope if school<=10 & groups
```

	school	intercept	slope
1.	1	3.634251	.6817045
74.	2	4.587045	.7214552
129.	3	4.682596	.6375957
181.	4	.2351664	.6839111
260.	5	2.347721	.6287867
295.	6	5.068725	.6153533
375.	7	3.525858	.4078594
463.	8	-.2369701	.5636145
565.	9	-1.883067	.4681097
599.	10	-3.254161	.4206528

Using intercept and slope you can estimate `yhat`, type

```
gen yhat= intercept + (slope*x1)
```

Or, after `xtmixed` type:

```
predict yhat_fit, fitted
```

```
list school yhat yhat_fit if school<=10 & groups
```

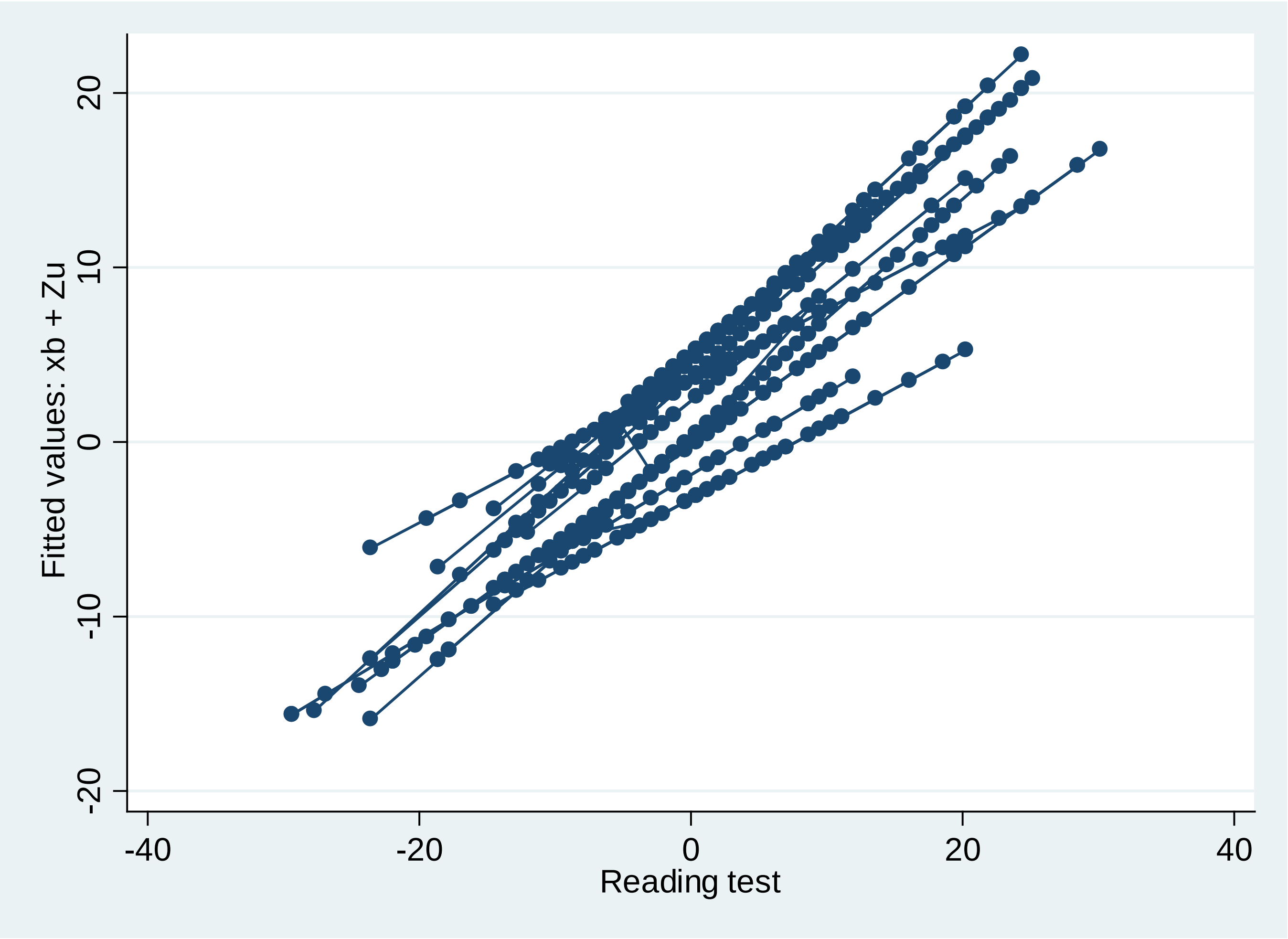
```
. list school yhat yhat_fit if school<=10 & groups
```

	school	yhat	yhat_fit
1.	1	-12.42943	-12.42943
74.	2	-15.3951	-15.3951
129.	3	-7.179871	-7.179871
181.	4	-15.88052	-15.88052
260.	5	-5.193317	-5.193318
295.	6	-3.836668	-3.836667
375.	7	-6.084939	-6.084939
463.	8	-13.98353	-13.98353
565.	9	-15.62209	-15.62209
599.	10	-9.341847	-9.341847

Postestimation: fittied values (graph)

You can plot individual regressions, type

```
twoway connected yhat_fit x1 if school<=10, connect(L)
```

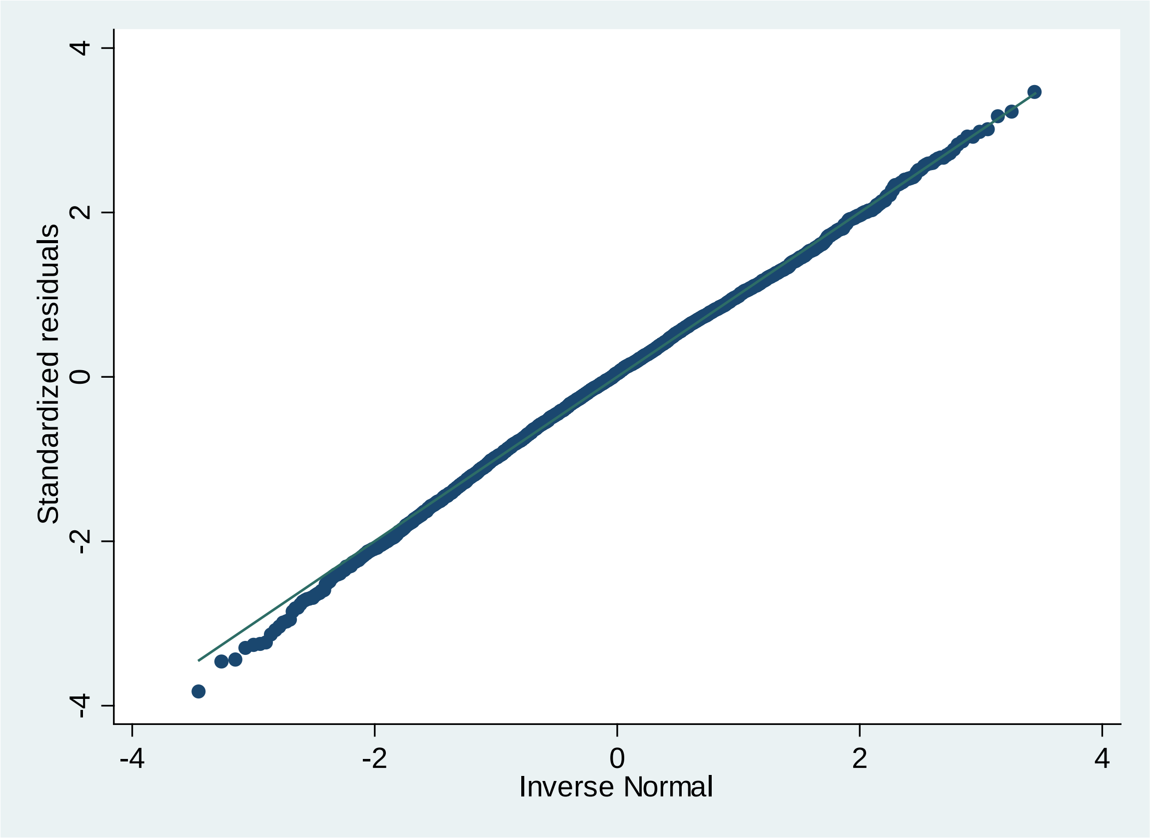


After `xtmixed` you can get the residuals by typing:

```
predict resid, residuals
predict resid_std, rstandard /* residuals/sd(Residual) */
```

A quick check for normality in the residuals

```
qnorm resid_std
```



Useful links / Recommended books / References

- DSS Online Training Section http://dss.princeton.edu/training/
- UCLA Resources http://www.ats.ucla.edu/stat/
- Princeton DSS Libguides http://libguides.princeton.edu/dss

Books/References

- "Beyond "Fixed Versus Random Effects": A framework for improving substantive and statistical analysis of panel, time-series cross-sectional, and multilevel data" / Brandom Bartels http://polmeth.wustl.edu/retrieve.php?id=838
- "Robust Standard Errors for Panel Regressions with Cross-Sectional Dependence" / Daniel Hoechle, http://fmwww.bc.edu/repec/bocode/x/xtscc_paper.pdf
- *An Introduction to Modern Econometrics Using Stata*/ Christopher F. Baum, Stata Press, 2006.
- *Data analysis using regression and multilevel/hierarchical models* / Andrew Gelman, Jennifer Hill. Cambridge ; New York : Cambridge University Press, 2007.
- *Data Analysis Using Stata*/ Ulrich Kohler, Frauke Kreuter, 2nd ed., Stata Press, 2009.
- *Designing Social Inquiry: Scientific Inference in Qualitative Research* / *Gary King, Robert O. Keohane, Sidney Verba, Princeton University Press*, 1994.
- *Econometric analysis* / William H. Greene. 6th ed., Upper Saddle River, N.J. : Prentice Hall, 2008.
- *Introduction to econometrics* / James H. Stock, Mark W. Watson. 2nd ed., Boston: Pearson Addison Wesley, 2007.
- *Statistical Analysis: an interdisciplinary introduction to univariate & multivariate methods /* Sam Kachigan, New York : Radius Press, c1986
- *Statistics with Stata (updated for version 9) /* Lawrence Hamilton, Thomson Books/Cole, 2006
- *Unifying Political Methodology: The Likelihood Theory of Statistical Inference* / Gary King, Cambridge University Press, 1989