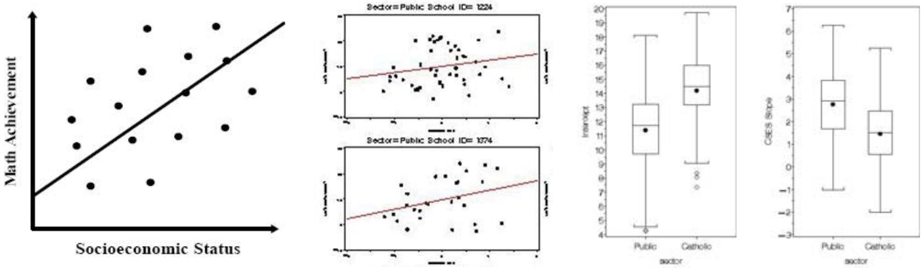


Mixed models for hierarchical & longitudinal data

Michael Friendly
Psychology 6140

Classical GLM

- The classical GLM, $\mathbf{y} = \mathbf{X}\,\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, assumes:
  - All observations are conditionally independent
    - → residuals, ε_i, are uncorrelated
  - The model parameters, $\boldsymbol{\beta}$, are fixed (non-random)
    - → only the residuals are random effects
- These assumptions are commonly violated:
  - Repeated measures & split-plot designs
  - Longitudinal and growth models
    - E.g., subjects $\subset$ groups $\subset$ time (age)
  - Hierarchical & multi-level designs
    - E.g., children $\subset$ classes $\subset$ schools $\subset$ counties …
    - patients $\subset$ therapists $\subset$ treatment type

Why Mixed models?

- More flexible for repeated measure or longitudinal data than univariate or multivariate approaches based on PROC GLM
  - Allows *missing data* (GLM w/ REPEATED discards missing)
  - Does *not* require measurements at the *same* time points
  - Provides a wide class of var-cov structures for dependent data (sometimes interest in modeling this)
    - E.g., unstructured (MANOVA), compound symmetry, AR(1), …
- Provides GLS, ML and REML estimates
  - More efficient than OLS
  - AIC & BIC for model selection
  - Better estimates of variance components than traditional ANOVA based on E(MS)

Fixed vs. random factors

Fixed and random factors differ in the scope of inference

	Fixed	Random
Levels	Given # of possibilities	Selected at random from a population
New experiment	Use *same* levels	Use *different* levels from same population
Goal	Estimate *means* of fixed levels	Estimate *variance* of population of means, σ^2_μ
Inference	Only for levels used $H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$	For all levels in the population $H_0 : \sigma^2_\mu = 0 \quad H_a : \sigma^2_\mu > 0$

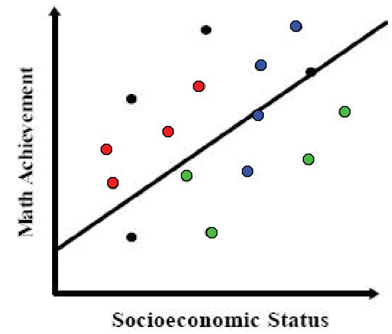
Terminology

- Different names for this modeling approach:
  - Hierarchical linear models (HLM)
  - Multilevel models (MLM)
  - Mixed-effects models (fixed & random)
  - Variance component models
  - Random-effects models
  - Random-coefficients regression models
  - …

Different names arose partially because these methods were re-invented in a variety of fields (psychology, education, agronomy, economics, ...), each with different slants and emphasis.

Main example: math achievement & SES

- Predicting math achievement
- Model: $y_i = \beta_0 + \beta_1\ \text{SES} + \varepsilon_i$
- OLS:
  - Best, unbiased estimates *iff* assumptions are met
- But:
  - Kids in same class not independent
  - Classes in same school, ditto
- Effect:
  - Have <N independent obs.
  - Std. errors overly optimistic– p-values too small
  - Other effects not controlled
  - Worse with unbalanced data

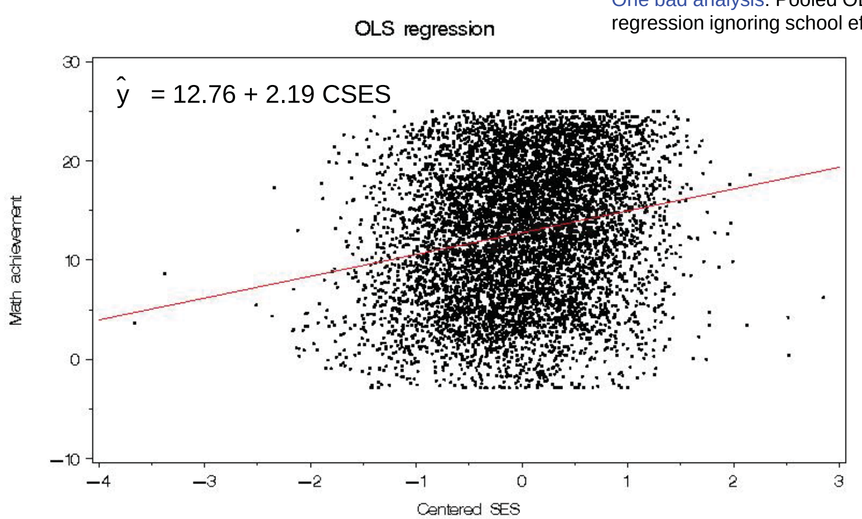


High School & Beyond data

7185 students, 160 schools

Predictors: CSES, school size, female, minority, …

One bad analysis: Pooled OLS regression ignoring school effects



High School & Beyond data

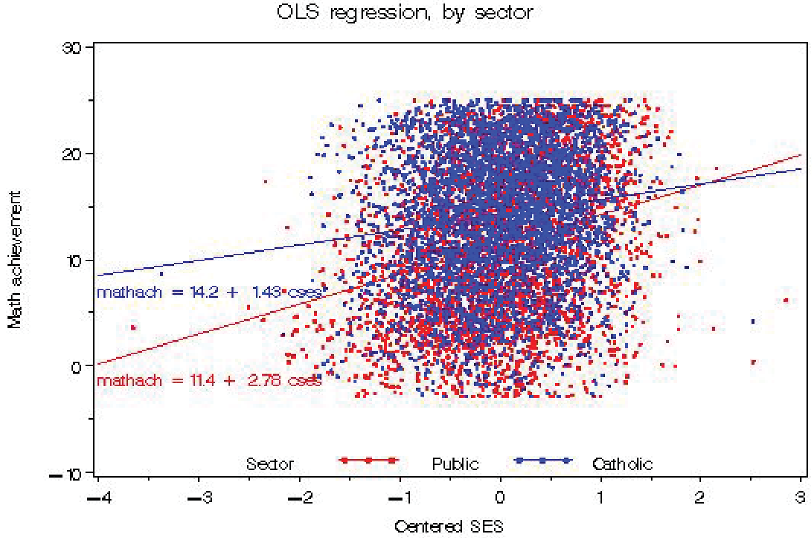
- Response variable: math achievement
- Level 1 predictors (students)
  - Minority? (0/1)
  - Female? (0/1)
  - SES: student SES (parent education, occupation, income)
  - CSES: mean-centered SES = (SES – meanSES)
- Level 2 predictors (schools)
  - Size – school enrollment
  - Sector – public or Catholic (private)? – (0/1)
  - meanSES – school mean of SES
  - PrAcad – proportion of students in academic track
  - DisClim – scale for disciplinary climate in school
  - HIminty – more than 40% minority students?

constant for all students in a given school

HSB data: sector effects?

slight improvement: separate lines for each sector

problem: school dependency still ignored



Fixed effect approach

- If predictors in model can account for correlations of residuals, then conditional independence will be satisfied
- E.g., add school effect to adjust for mean differences among schools

NB: crucial to control for school here

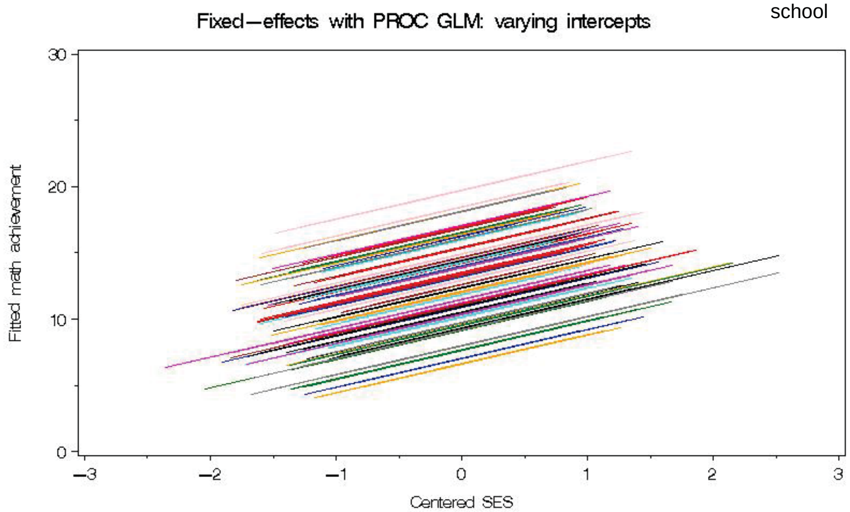
```
*-- fixed-effect approach via PROC GLM;
title 'Fixed-effects with PROC GLM: varying intercepts';
proc glm data=hsbmix;
      class school;
      model mathach = cses school ;
      output out=glm1 p=predict r=residual;
run;
```

R:
```
mod1 <- lm(mathach ~ cses + school, data = hsbmix)
```

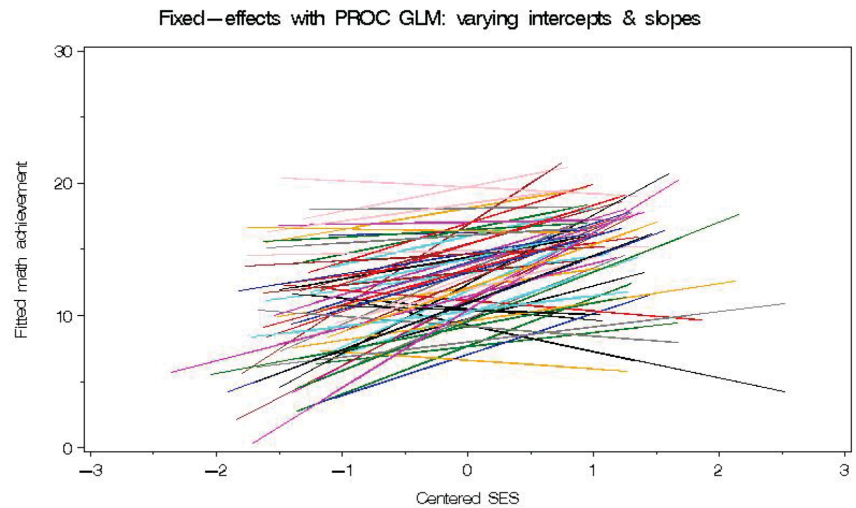


```
title 'Fixed-effects with PROC GLM: varying intercepts';
proc glm data=hsbmix;
      class school;
      model mathach = cses  school ;
      output out=glm1 p=predict r=residual;
run;
```

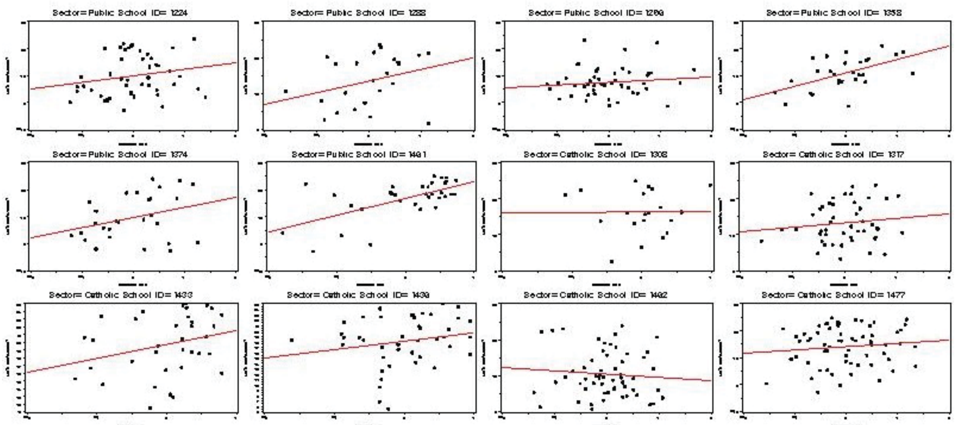
One line for each school




```
title 'Fixed-effects with PROC GLM: varying intercepts & slopes';
proc glm data=hsbmix;
      class school;
      model mathach = cses school cses*school ;
      output out=glm2 p=predict r=residual;
run;
```



- Separate plots for a subset of schools shows considerable variability in intercepts and slopes– how do these relate to school-level variables?
- But this treats the school parameters as fixed – inference to these schools only, not a popⁿ of schools.
- Still assumes conditional independence and constant within-school σ^2



Analyzing school-level variation

- We could just fit a separate regression model for each of the 160 schools

$$y_{ij} = \beta_{0j} + \beta_{1j}\, CSES_{ij} + e_{ij}$$

- Capture the coefficients, (β_{0j} , β_{1j}) and analyze these in relation to:
  - Sector, school size, …

```
proc reg data=hsbmix outest=parms;
   by sector school;
   model mathach = cses;
proc glm data=parms;
   model Int slope = sector;
```

```
sector  school   Int  slope
  Pub     1224   9.73  2.51
  Pub     1288  13.53  3.26
  Pub     1296   7.64  1.08
  ...
  Cat     1308  16.26  0.12
  Cat     1317  13.18  1.27
  Cat     1433  19.73  1.60
  ...
```


Analyzing school-level variation

Results are interpretable

• Public: lower math at mean CSES

• Public: greater slope for CSES

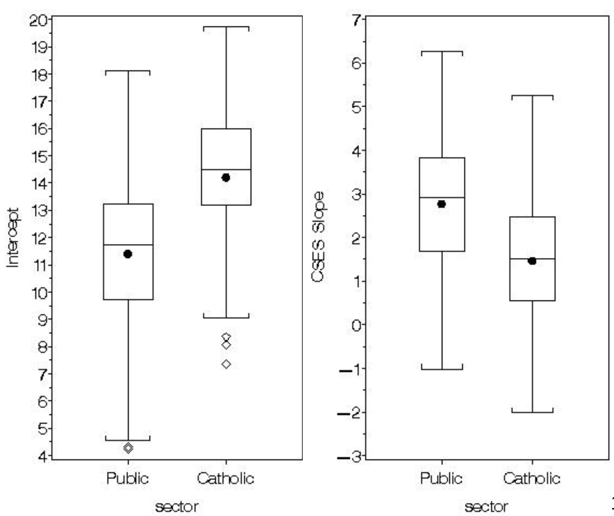
But:

• doesn't take n_j into account

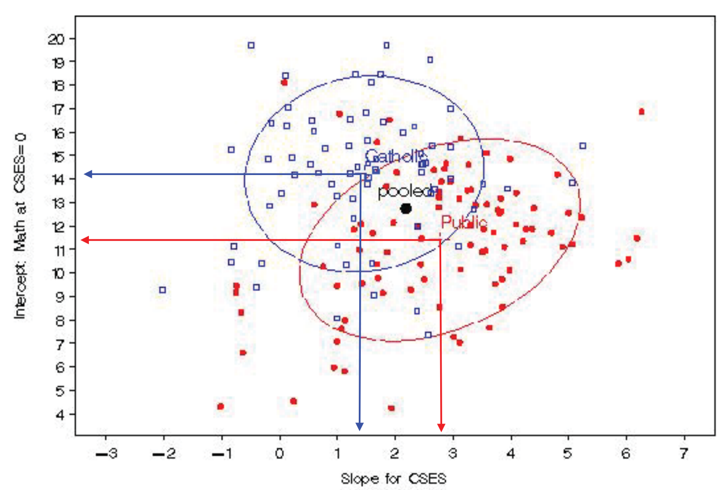
• std errors still too small

• inferences maybe wrong

• hard to handle other nestings



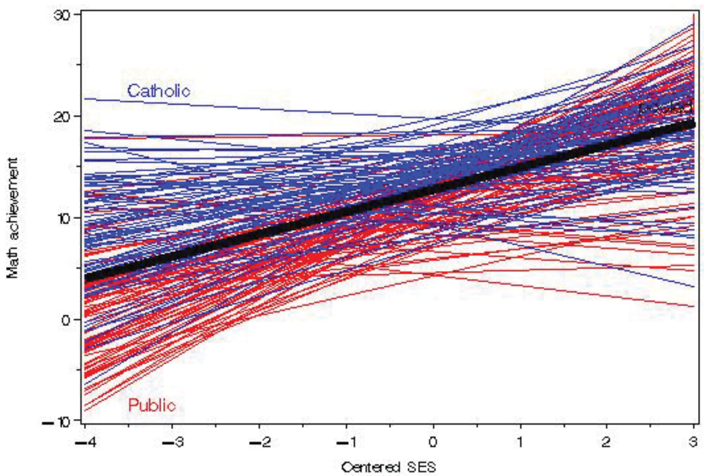
Analyzing school-level variation



A better (joint) plot shows individual slopes and intercepts in β space

Data ellipses show the covariation within groups

Analyzing school-level variation



The same plot in data space

SES has a larger impact (slope) in public schools

Math ach. higher in Catholic schools

Multilevel (mixed) model approach

- Multilevel model treats both students *and* schools as sampling units from some populations
- In particular, schools are considered another random effect, with some distribution
  - → we can estimate the variance due to schools
    - Allows inference about pop^n of schools: H0: $\sigma^2_{schools} > 0$?
  - → we can model relations at different levels:
    - L1: student variables (IQ, sex, minority),
    - L2: class-level variables (teacher experience, class size),
    - L3: school-level variables (public vs. private, school size)

Basic multilevel model: random-effects ANOVA

- Ignore CSES for now: examine mean differences in the pop^n of schools
- Level 1 model: $y_{ij} = \beta_{0j} + e_{ij}$
  - β_{0j} is the mean for school *j*, with some distribution in the pop^n of schools
  - e_{ij} is the residual for person *i* in school *j*
- Level 2 model: $\beta_{0j} = \gamma_{00} + u_{0j}$

where: γ_{00} = grand mean of y

u_{0j} = deviation of group j from GM

(Notation: I'm using Γρεεκ for fixed parameters, Roman for random parameters)

Basic multilevel model: random-effects ANOVA

- Substitute Level 2 into Level 1: "reduced-form model":

$$y_{ij} = \overbrace{\gamma_{00} + u_{0j}}^{\beta_{0j}} + e_{ij}$$

- Now, assume u_{0j} & e_{ij} are independent, and

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

$$u_{0j} \sim N(0, \tau_{00})$$

- Now have two variance components

$$var(y_{ij}) = \tau_{00} + \sigma^2$$

school residual

Basic multilevel model: ICC

- ICC: express variance of group means, β_{0j}, as proportion of total variance of y_{ij}

$$ICC = \frac{\tau_{00}}{\tau_{00} + \sigma^2} \qquad \frac{\text{school variance}}{\text{total}}$$

- $0 \le ICC \le 1$: proportion of variance accounted for by school mean variation
  - ICC ≈ 0 : little variation among school means
  - ICC ≈ 1 : most variation of y_{ij} acct'd for by school means
  - Fixed effects model: assumes $\tau_{00} = 0 \rightarrow$ ICC=0

Estimating multilevel models: PROC MIXED

- Basic syntax:

```
proc mixed data=<dataset> <options>;
      class <class variables>;
      model <dependent> = <fixed-effects> < / options>;
      random <random-effects> </options>;
```

NB: fixed & random effects are specified separately

- Example:

```
title 'Mixed model 0: random-effects ANOVA';
proc mixed data=hsbmix noclprint covtest method=reml;
      class school;
      model mathach = / solution ddfm=bw outp=mix0;
      random intercept / sub=school type=un;
run;
```


Output:

The Mixed Procedure

```
                Covariance Parameter Estimates

                                       Standard         Z
Cov Parm      Subject     Estimate        Error     Value        Pr Z

UN(1,1)       school        8.6097       1.0778      7.99      <.0001
Residual                   39.1487       0.6607     59.26      <.0001

                  Solution for Fixed Effects

                            Standard
Effect        Estimate         Error       DF    t Value    Pr > |t|

Intercept      12.6370        0.2443      159      51.72      <.0001
```

$\hat{\tau}_{00}$ $\hat{\sigma}^2$ $\hat{\gamma}_{00}$

- Variance component for schools: τ_{00} = 8.6 → signif between school variation
- τ_{00} is also the implied covariance of students within the same school
- Level 1 residual variance: σ^2 = 39.1 → signif within school variation
- ICC = 8.6/(8.6 + 39.1) = .18 → 18% of total variance due to school means
- Fixed effect analysis inappropriate: dependency within schools

Other useful output:

```
                            Iteration History
 Iteration    Evaluations    -2 Res Log Like       Criterion
         0              1    48102.91726234
         1              2    47116.81230623      0.00000109
         2              1    47116.79350024      0.00000000

                      Convergence criteria met.

                            Fit Statistics

-2 Res Log Likelihood          47116.8
AIC (smaller is better)        47120.8
AICC (smaller is better)       47120.8
BIC (smaller is better)        47126.9

                    Null Model Likelihood Ratio Test

      DF    Chi-Square      Pr > ChiSq

       1        986.12          <.0001
```

Don't interpret output if model fails to converge

Useful for comparing the fit of different models for the same data

Test of present model vs. one that assumes independence and homoscedasticity (std. OLS assumptions)

Random effects regression: random intercept

- Use CSES as a student, level 1 predictor
- For now, allow only intercept to be random

Level 1 model (student)	Level 2 model (school)
$y_{ij} = \beta_{0j} + \beta_{1j}\,\text{CSES}_{ij} + \color{red}{e_{ij}}$	$\beta_{0j} = \gamma_{00} + \color{red}{u_{0j}}$ (intercept) $\beta_{1j} = \gamma_{10}$ (slope)

Combined model:

$$y_{ij} = \underbrace{[\gamma_{00} + \gamma_{10}\,\text{CSES}_{ij}]}_{\substack{\text{fixed} \\ \text{(MODEL stmt)}}} + \underbrace{[\overset{\tau_{00}}{\color{red}{u_{0j}}}] + \overset{\sigma^2}{\color{red}{e_{ij}}}}_{\substack{\text{random} \\ \text{(RANDOM stmt)}}}$$

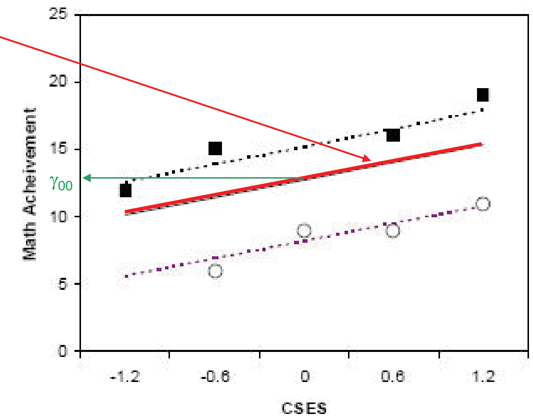
Random effects regression: random intercept

Fixed effects: avg. regression

$\gamma_{00} + \gamma_{10}\,\text{CSES}_{ij}$

τ_{00} : variance of intercepts (school means)

σ^2 : residual var. within schools (around indiv. lines)




```
title 'Mixed model 1: random intercepts';
proc mixed data=hsbmix noclprint covtest method=reml;
       class school;
       model mathach = cses / solution ddfm=bw;
       random intercept / sub=school type=un;
run;
```

```
                    Covariance Parameter Estimates

                                          Standard         Z
Cov Parm     Subject      Estimate          Error      Value          Pr Z

UN(1,1)      school         8.6677         1.0784       8.04        <.0001
Residual                   37.0108         0.6246      59.25        <.0001

                      Solution for Fixed Effects

                              Standard
Effect        Estimate           Error       DF    t Value    Pr > |t|

Intercept      12.6493          0.2444      159      51.75      <.0001
cses            2.1912          0.1087     7024      20.17      <.0001
```

$\hat{\tau}_{00}$ (UN(1,1)), $\hat{\sigma}^2$ (Residual), $\hat{\gamma}_{00}$ (Intercept), $\hat{\gamma}_{10}$ (cses)

Random intercepts, random slopes

- Test whether slope of CSES relation varies over schools– allow it to vary!
- Level 1: $y_{ij} = \beta_{0j} + \beta_{1j}\,\text{CSES}_{ij} + \color{red}{e_{ij}}$
- Level 2: $\beta_{0j} = \gamma_{00} + \color{red}{u_{0j}}$ (intercept)
  $\beta_{1j} = \gamma_{10} + \color{red}{u_{1j}}$ (slope)
- Reduced:

$$y_{ij} = \underbrace{[\gamma_{00} + \gamma_{10}\,\text{CSES}_{ij}]}_{\text{fixed}} + \underbrace{[\color{red}{u_{0j}} + \color{red}{u_{1j}}\,\text{CSES}_{ij}] + \color{red}{e_{ij}}}_{\text{random}}$$

Variances: τ_{00} τ_{11} σ^2

Covariance: τ_{01}

```
title 'Mixed model 2: random intercepts & slopes';
proc mixed data=hsbmix noclprint covtest method=reml;
      class school;
      model mathach = cses / solution ddfm=bw outp=mix2;
      random intercept cses / sub=school type=un;
```

Covariance Parameter Estimates

	Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
$\hat{\tau}_{00}$	UN(1,1)	school	8.6769	1.0786	8.04	<.0001
	UN(2,1)	school	0.05075	0.4062	0.12	0.9006
$\hat{\tau}_{11}$	UN(2,2)	school	0.6940	0.2808	2.47	0.0067
$\hat{\sigma}^2$	Residual		36.7006	0.6258	58.65	<.0001

Solution for Fixed Effects

	Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
$\hat{\gamma}_{00}$	Intercept	12.6493	0.2445	159	51.75	<.0001
$\hat{\gamma}_{10}$	cses	2.1932	0.1283	7024	17.10	<.0001

Random intercepts, random slopes

- Fixed effects estimates are similar to OLS
  - Mixed: E(y|CSES) = 12.65 + 2.19 CSES
  - OLS: E(y|CSES) = 12.76 + 2.19 CSES
- Standard errors more realistic with mixed model
  - (often larger– non-independence)
- Variance components: relative size of random effects
  - School means (overall achievement) $\hat{\tau}_{00} = 8.68$
  - School slopes (~ 1/equity) $\hat{\tau}_{11} = 0.69$
  - Residual variance $\hat{\sigma}^2 = 36.7$

Intercepts & slopes as outcomes

- The level 2 (school) models can now consider other school-level predictors
- E.g., how do intercepts and slopes on CSES differ by Sector (Public/Catholic)?
- Add Sector to the level 2 model:
- Level 2: $\beta_{0j} = \gamma_{00} + \gamma_{01}\,\text{SECT}_j + u_{0j}$ (intercept)
  $\beta_{1j} = \gamma_{10} + \gamma_{11}\,\text{SECT}_j + u_{1j}$ (slope)
- Reduced:

$$y_{ij} = [\gamma_{00} + \gamma_{01}\,\text{SECT}_j + \gamma_{10}\,\text{CSES}_{ij} + \gamma_{11}\,\text{SECT}_j\,\text{CSES}_{ij}] \quad \text{fixed}$$
$$+ [u_{0j} + u_{1j}\,\text{CSES}_{ij}] + e_{ij} \quad \text{random}$$


```
title 'Mixed model 3: intercepts & slopes as outcomes';
proc mixed data=hsbmix noclprint covtest method=reml;
      class school;
      model mathach = cses | sector / solution ddfm=bw outp=mix3;
      random intercept cses / sub=school type=un;
```

Covariance Parameter Estimates

	Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
$\hat{\tau}_{00}$	UN(1,1)	school	6.7414	0.8649	7.79	<.0001
	UN(2,1)	school	1.0503	0.3421	3.07	0.0021
$\hat{\tau}_{11}$	UN(2,2)	school	0.2656	0.2288	1.16	0.1228
$\hat{\sigma}^2$	Residual		36.7066	0.6258	58.66	<.0001

Solution for Fixed Effects

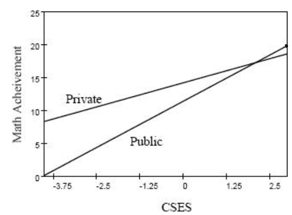
	Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
$\hat{\gamma}_{00}$	Intercept	11.4106	0.2928	158	38.97	<.0001
$\hat{\gamma}_{10}$	cses	2.8028	0.1550	7023	18.09	<.0001
$\hat{\gamma}_{01}$	sector	2.7995	0.4393	158	6.37	<.0001
$\hat{\gamma}_{11}$	cses*sector	-1.3411	0.2338	7023	-5.74	<.0001

Interpreting the output

- Fixed effects:
  - $\hat{\gamma}_{00}$ = 11.41 = avg math achievement in public schools
  - $\hat{\gamma}_{10}$ = 2.80 = slope for CSES effect in public schools
  - $\hat{\gamma}_{01}$ = 2.80 = increment in avg mathach in Catholic schools
  - $\hat{\gamma}_{11}$ = -1.34 = change in CSES slope for Catholic schools
- Thus, predicted effects are
  - Sector 0 (public): E(y|CSES) = 11.41 + 2.80 CSES
  - Sector 1 (Catholic): E(y|CSES) = 14.21 + 1.46 CSES

Children of avg SES do better in Catholic schools

Performance in Catholic schools depends less on SES



Interpreting the output

- Random effects
  - $\hat{\tau}_{00}$ = 6.74 : residual intercept variance, controlling for sector
  - $\hat{\tau}_{11}$ = .266 : residual CSES slope variance, " " "
  - $\hat{\sigma}^2$ = 36.7 : residual variance w/in schools, " " CSES
- Evaluating the impact of the level 2 predictor:
  - $\hat{\tau}_{00}$ decreased from 8.68 to 6.74: decrease of 22%
  - $\hat{\tau}_{11}$ decreased from .692 to .266: decrease of 62%
  - But, $\hat{\tau}_{11}$ is no longer signif > 0 : residual diff^{ces} in slope are minimal after sector is accounted for
  - Intercept variance $\hat{\tau}_{00}$ still large: perhaps other lev 2 predictors?
  - Residual (within-school) variance, σ^2 still large: other lev 1?

Mixed models in R

`lme()` in the nlme package uses a similar syntax for fixed and random effects, from the reduced-form equation.

```
library(nlme)
# random intercepts
lme.1 <- lme(mathach ~ cses,
        random = ~ 1 | school, data=hsbmix)
# random intercept and slope
lme.2 <- lme(mathach ~ cses,
        random = ~ 1 + cses | school, data=hsbmix)
# intercepts and slopes vary with sector
lme.3 <- lme(mathach ~ sector*cses,
        random = ~ 1 + cses | school, data=hsbmix)
```

NB: `lme()` assumes an intercept for the fixed and random terms

Comparing models

	Parameter		0: Random ANOVA	1: Random Intercepts	2: Random Int & Slope	3: Including SECTOR
Variances	Int var	$\hat{\tau}_{00}$	8.610	8.668	8.677	6.741
	Slope var	$\hat{\tau}_{11}$			0.692	0.266
	Resid var	$\hat{\sigma}^2$	39.149	37.011	36.700	36.707
Fixed effects	Intercept	$\hat{\gamma}_{00}$	12.637	12.649	12.649	11.411
	CSES	$\hat{\gamma}_{10}$		2.191	2.193	2.803
	sector	$\hat{\gamma}_{01}$				2.799
	CSES*sector	$\hat{\gamma}_{11}$				-1.341
	AIC		47120.8	46728.0	46722.2	46511.7

Estimating random effects: BLUEs & BLUPS

- OLS regressions (within School) give Best Linear Unbiased Estimates (BLUEs) of

$$\hat{\boldsymbol{\beta}}_j = \begin{pmatrix} \hat{\beta}_{0j} \\ \hat{\beta}_{1j} \end{pmatrix} \quad \text{with} \quad Var(\hat{\boldsymbol{\beta}}_j) = \hat{\sigma}^2 (X_j^T X_j)^{-1}$$

- Another estimate comes from random intercepts and slopes

$$\hat{\mathbf{u}}_j = \begin{pmatrix} \hat{u}_{0j} \\ \hat{u}_{1j} \end{pmatrix} \quad \text{with} \quad Var(\hat{\mathbf{u}}_j) = \begin{pmatrix} \hat{\tau}_{00} & \\ \hat{\tau}_{01} & \hat{\tau}_{11} \end{pmatrix} = \hat{\mathbf{T}}$$

- A better estimate --- the BLUP (Best Linear Unbiased Predictor) is a weighted average of these, using 1/Var as weights

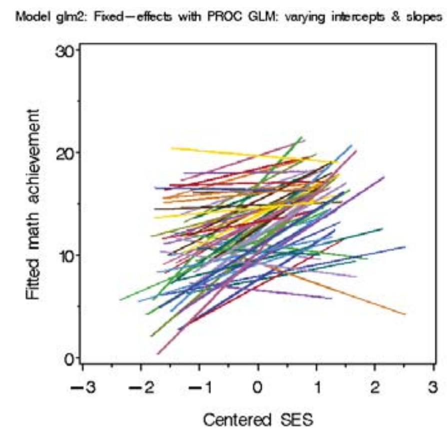
$$\tilde{\boldsymbol{\beta}}_j = \left\{ \underbrace{\hat{\boldsymbol{\beta}}_j \left[Var(\hat{\boldsymbol{\beta}}_j) \right]^{-1}}_{\text{OLS}} + \underbrace{\hat{\mathbf{u}}_j \hat{\mathbf{T}}^{-1}}_{\text{Random}} \right\} \left[\left[Var(\hat{\boldsymbol{\beta}}_j) \right]^{-1} + \hat{\mathbf{T}}^{-1} \right]^{-1}$$

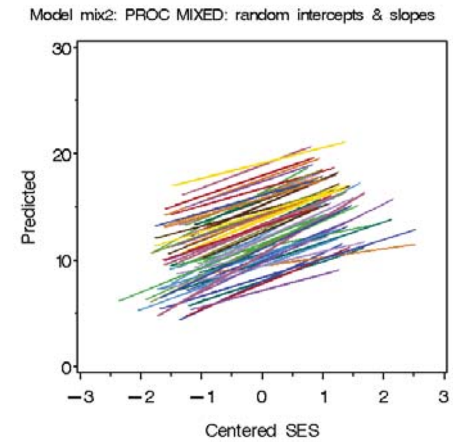
- This “borrows strength” --- **optimally** combines the information from school j with information from all schools

Estimating random effects: BLUEs & BLUPS

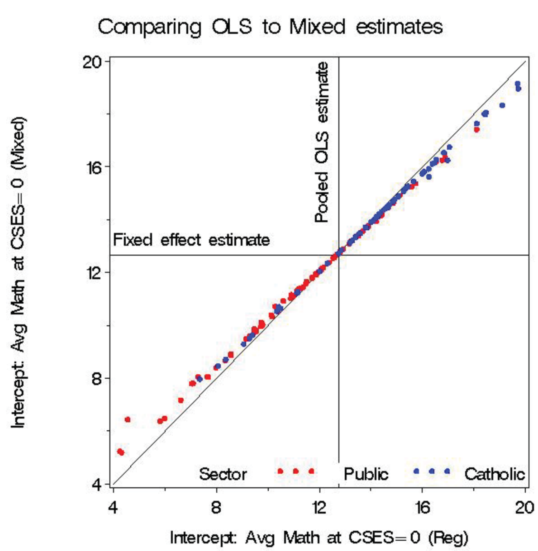
Comparing OLS to Mixed estimates

- OLS treats each school separately
- Mixed model “smooths” estimates toward the pooled estimate





Estimating random effects: BLUEs & BLUPS



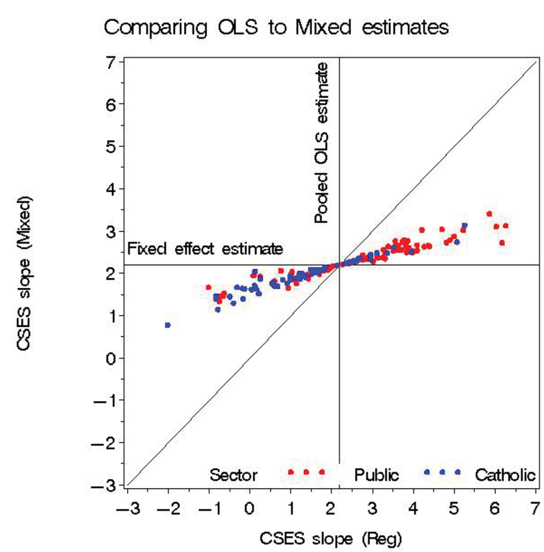
Results for Model 3: Random intercepts and slopes

The BLUP estimates of $\tilde{\beta}_{0j}$ are shrunk towards the OLS estimate

But only slightly, because there is a large variance component for intercepts $\hat{\tau}_{00}$

Thus, the mixed estimates of u_{0j} have a small weight

Estimating random effects: BLUEs & BLUPS

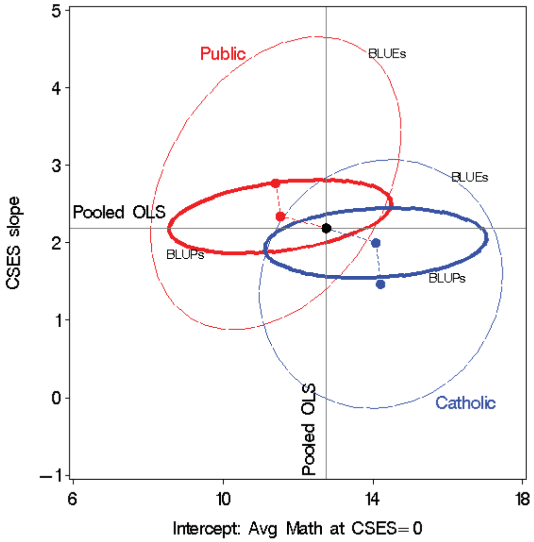


The mixed model estimates of slopes for CSES are shrunk much more because there is a smaller variance component for slopes, $\hat{\tau}_{11}$

Typically, we are not interested directly in the random effects for individual schools;

However, the same idea applies to other estimates based on the random effects, e.g., estimating the mean difference between Public & Catholic schools at given values of CSES or other predictors

Estimating random effects: BLUEs & BLUPS



The mixed model estimates of slopes for CSES are shrunk much more because there is a smaller variance component for slopes, $\hat{\tau}_{11}$

Diagnostics and influence measures

- As in the GLM, regression diagnostics are available for mixed models in SAS [Uses ODS Graphics]
  - Influence of deleting observations at Level 1 (individual) or Level 2 (cluster)
  - Plots of Cook's D and other influence measures

Boxplots of residuals

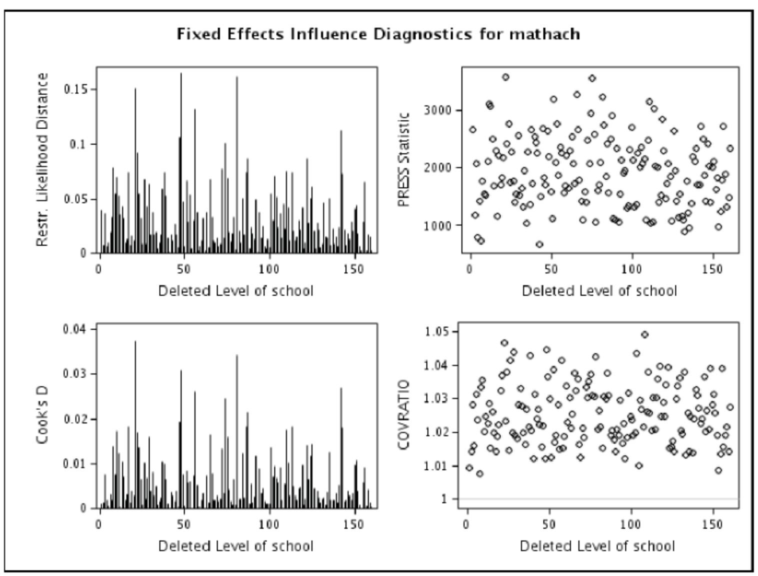
```
ods pdf file='hsbmix3.pdf';
ods graphics on;
title 'Mixed model: intercepts & slopes as outcomes';
proc mixed data=hsbmix noclprint covtest method=reml boxplot ;
  class school;
  model mathach = cses | sector / solution ddfm=bw
            influence (effect=school estimates);
  random intercept cses / sub=school type=un;
run;
ods graphics off;
ods pdf close;
```

Influence of schools

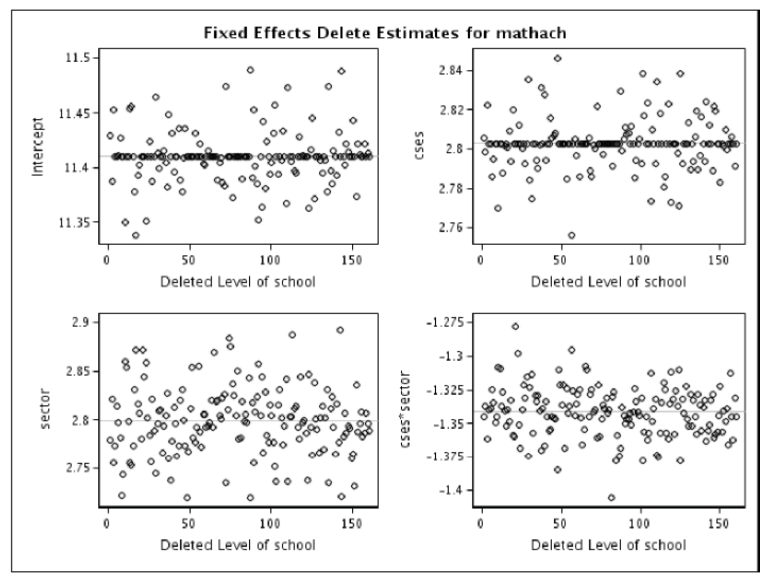
Influence on estimates

Influence plots

With many level 2 clusters, influential ones are less likely



Influence plots



Taxonomy of models

Consider: X as a Level 1 (individual) predictor; G as a Level 2 (group) predictor

	Fixed (MODEL stmt)	Random	Combined formula
Random effects ANOVA	Intercept	Int	$y_{ij} = \gamma_{00} + u_{0j} + e_{ij}$
Means as outcomes	Int G	Int	$y_{ij} = \gamma_{00} + \gamma_{01} G_j + u_{0j} + e_{ij}$
Random intercepts	Int X	Int	$y_{ij} = \gamma_{00} + \gamma_{10} X_{ij} + u_{0j} + e_{ij}$
Random coefficients	Int X	Int X	$y_{ij} = \gamma_{00} + \gamma_{10} X_{ij} + u_{0j} + u_{1j} X_{ij} + e_{ij}$
Intercepts, slopes as outcomes	Int X G G*X	Int X	$y_{ij} = \gamma_{00} + \gamma_{10} X_{ij} + \gamma_{01} G_j + \gamma_{11} G_j X_{ij} + u_{0j} + u_{1j} X_{ij} + e_{ij}$
Non-random slopes	Int X G G*X	Int	$y_{ij} = \gamma_{00} + \gamma_{10} X_{ij} + \gamma_{01} G_j + \gamma_{11} G_j X_{ij} + u_{0j} + e_{ij}$

The general linear mixed model

- Consider the outcomes, y_{ij}, $i=1,\dots,n_j$ within level 1 units $j=1,\dots,J$. $\mathbf{y}_j$ is the response vector for group j.
- For group j, the GLMM is

$$\mathbf{y}_j = [\mathbf{X}_j \quad \mathbf{Z}_j] \begin{pmatrix} \gamma \\ \mathbf{u}_j \end{pmatrix} + \epsilon_j = \mathbf{X}_j \gamma + \mathbf{Z}_j \mathbf{u}_j + \epsilon_j$$

fixed predictors (level 1,2)

random predictors (level 2)

fixed parameters

random parameters

$$\text{var}(\mathbf{u}) = \mathrm{T} = \begin{bmatrix} \tau_{00} & & \\ \vdots & \ddots & \\ \tau_{0q} & \cdots & \tau_{qq} \end{bmatrix}$$

The general linear mixed model

- For example, the model with sector:

$y_{ij} = [\gamma_{00} + \gamma_{01}\ \text{SECT}_j + \gamma_{10}\ \text{CSES}_{ij} + \gamma_{11}\ \text{SECT}_j\ \text{CSES}_{ij}]$
$\quad + [u_{0j} + u_{1j}\ \text{CSES}_{ij}] + e_{ij}$

$$\begin{bmatrix} y_{1j} \\ y_{2j} \\ \vdots \\ y_{n_j j} \end{bmatrix} = \begin{bmatrix} 1 & \text{SECTOR}_j & \text{CSES}_{1j} & (\text{SECTOR}_j)(\text{CSES}_{1j}) \\ 1 & \text{SECTOR}_j & \text{CSES}_{2j} & (\text{SECTOR}_j)(\text{CSES}_{2j}) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \text{SECTOR}_j & \text{CSES}_{n_j j} & (\text{SECTOR}_j)(\text{CSES}_{n_j j}) \end{bmatrix} \begin{bmatrix} \gamma_{00} \\ \gamma_{01} \\ \gamma_{10} \\ \gamma_{11} \end{bmatrix} + \begin{bmatrix} 1 & \text{CSES}_{1j} \\ 1 & \text{CSES}_{2j} \\ \vdots & \vdots \\ 1 & \text{CSES}_{n_j j} \end{bmatrix} \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} + \begin{bmatrix} e_{1j} \\ e_{2j} \\ \vdots \\ e_{n_j j} \end{bmatrix}$$

$$\mathbf{y}_j = \mathbf{X}_j \gamma + \mathbf{Z}_j \mathbf{u}_j + \mathbf{e}_j$$

Note that level 1 predictors (CSES) *vary* over cases w/in schools;

Level 2 predictors (SECTOR) are *constant* w/in schools

The general linear mixed model

- Specifying distributions & covariance structure
  - Typically assume that both the random effects, $\mathbf{u}_j$ and residuals, $\mathbf{e}_j$, are normally distributed, and mutually independent

$$\begin{pmatrix} \mathbf{u}_j \\ \mathbf{e}_j \end{pmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{T} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_j \end{bmatrix} \right)$$

var-cov matrix of random effects

var-cov matrix of level 1 resids: typically $\sigma^2 \mathbf{I}_n$

- The variance of $\mathbf{y}_j$ is therefore $\mathbf{Z}\ \mathbf{T}\ \mathbf{Z}' + \mathbf{\Sigma}_j$
- In most cases, $\mathbf{T}$ is unstructured– all var/cov freely estimated & $\mathbf{\Sigma}_j = \sigma^2\ \mathbf{I}$
  - But mixed model allows more restricted & specialized structures
  - E.g., could estimate separate $\mathbf{T}$ matrices for public/Catholic
  - Longitudinal data: $\mathbf{\Sigma}_j$ = autoregressive ($\rho_{kl} = \rho^{|k-l|}$)
- If $\mathbf{\Sigma}_j = \sigma^2\ \mathbf{I}$ and no random effects, this reduces to std model

Covariance structures for T & Σ

Structure (TYPE= option)	Parameters	(i,j)th element	Form
Unstructured **UN**	t(t+1)/2	σ_{ij}	$\sigma^2 \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{bmatrix}$
Compound Symmetry **CS**	2	$\sigma_1 + \sigma^2 1(i=j)$	$\sigma^2 \begin{bmatrix} \sigma^2+\sigma_1 & \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma^2+\sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma^2+\sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma_1 & \sigma^2+\sigma_1 \end{bmatrix}$
First-order autoregressive **AR(1)**	2	$\sigma^2 \rho^{\lvert i-j \rvert}$	$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$

There are many more possibilities for special forms of dependence

Mostly, these are used in special situations; the GLMM provides them.

Table 56.13 Covariance Structures

Structure	Description	Parms	(i,j)th element
ANTE(1)	Ante-dependence	$2t-1$	$\sigma_i \sigma_j \prod_{k=i}^{j-1} \rho_k$
AR(1)	Autoregressive(1)	2	$\sigma^2 \rho^{\lvert i-j \rvert}$
ARH(1)	Heterogeneous AR(1)	$t+1$	$\sigma_i \sigma_j \rho^{\lvert i-j \rvert}$
ARMA(1,1)	ARMA(1,1)	3	$\sigma^2 [\gamma \rho^{\lvert i-j \rvert - 1} 1(i \neq j) + 1(i=j)]$
CS	Compound Symmetry	2	$\sigma_1 + \sigma^2 1(i=j)$
CSH	Heterogeneous CS	$t+1$	$\sigma_i \sigma_j [\rho 1(i \neq j) + 1(i=j)]$
FA(q)	Factor Analytic	$\frac{q}{2}(2t-q+1)+t$	$\Sigma_{k=1}^{\min(i,j,q)} \lambda_{ik} \lambda_{jk} + \sigma_i^2 1(i=j)$
FA0(q)	No Diagonal FA	$\frac{q}{2}(2t-q+1)$	$\Sigma_{k=1}^{\min(i,j,q)} \lambda_{ik} \lambda_{jk}$
FA1(q)	Equal Diagonal FA	$\frac{q}{2}(2t-q+1)+1$	$\Sigma_{k=1}^{\min(i,j,q)} \lambda_{ik} \lambda_{jk} + \sigma^2 1(i=j)$
HF	Huynh-Feldt	$t+1$	$(\sigma_i^2 + \sigma_j^2)/2 + \lambda 1(i \neq j)$
LIN(q)	General Linear	q	$\Sigma_{k=1}^{q} \theta_k \mathbf{A}_{ij}$
TOEP	Toeplitz	t	$\sigma_{\lvert i-j \rvert+1}$
TOEP(q)	Banded Toeplitz	q	$\sigma_{\lvert i-j \rvert+1} 1(\lvert i-j \rvert < q)$
TOEPH	Heterogeneous TOEP	$2t-1$	$\sigma_i \sigma_j \rho_{\lvert i-j \rvert}$
TOEPH(q)	Banded Hetero TOEP	$t+q-1$	$\sigma_i \sigma_j \rho_{\lvert i-j \rvert} 1(\lvert i-j \rvert < q)$
UN	Unstructured	$t(t+1)/2$	σ_{ij}

These require fewer parameters than the UNstructured (MANOVA) model

Other cov. structures handle spatial dependence

... even more

Multilevel models for longitudinal data

- Longitudinal data traditionally modeled as a repeated measure design--- simple!

  e.g.
  ```
  proc glm data=weightloss;
     class treat;
     model week1-week4 = treat;
     repeated week 4 (polynomial);
  ```
- But:
  - Requires: *complete* data, *same* time points for all
  - Does not allow *time-varying predictors* (e.g., exercise)
  - Restrictive assumptions: *compound symmetry*

Multilevel models for longitudinal data

- Multilevel models allow:
  - Different number of time points over subjects
  - Different time locations over subjects
  - Time-varying predictors
  - Several levels: individual ⊂ treatment ⊂ center
- Can model interactions with time
  - Do effects get larger? Smaller?
- Can allow for covariance structures appropriate to longitudinal data

Unconditional linear growth model

- Simplest model: scores change linearly over time, with random slopes and intercepts
  - NB: Define TIME so TIME=0 → **initial** status, or center (**average** status, etc.)

Level 1: $y_{ij} = \beta_{0j} + \beta_{1j}(TIME_{ij}) + e_{ij}$ where $e_{ij} \sim \mathcal{N}(0, \sigma^2)$

Level 2: $\begin{aligned}\beta_{0j} &= \gamma_{00} + u_{0j}, \\ \beta_{1j} &= \gamma_{10} + u_{1j},\end{aligned}$ where $\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix}\right]$

Unconditional linear growth model

- Reduced form (combined model):

$$y_{ij} = \underbrace{[\gamma_{00} + \gamma_{10} TIME_{ij}]}_{\text{fixed}} + \underbrace{[u_{0j} + u_{1j} TIME_{ij} + e_{ij}]}_{\text{random}}$$

- Fitting:

```
proc mixed covtest;
    class id;
    model y = time / solution;
    random intercept time/ subject=id type=un;
```

- Can easily include non-linear terms, eg, $TIME^2$

Linear growth, person-level predictor

- Now, begin to predict person-level intercepts and slopes

Level 1: Within person

$$y_{ij} = \beta_{0j} + \beta_{1j}(TIME_{ij}) + e_{ij}$$

where $e_{ij} \sim \mathcal{N}(0, \sigma^2)$

Level 2: Between person

$$\begin{aligned}\beta_{0j} &= \gamma_{00} + \gamma_{01} Treat_j + u_{0j}, \\ \beta_{1j} &= \gamma_{10} + \gamma_{11} Treat_j + u_{1j},\end{aligned}$$

where $\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix}\right]$

Combined model:

$$\begin{aligned} y_{ij} = &[\gamma_{00} + \gamma_{10}(TIME_{ij}) + \gamma_{01} Treat_j + \gamma_{11} TIME_{ij} Treat_j] \\ &+ [u_{01} + u_{11}(TIME_{ij}) + e_{ij}] \end{aligned}$$

Linear growth, person-level predictor

- Fitting:

```
proc mixed covtest;
    class id treat;
    model y = time treat time*treat / solution;
    random intercept time/ subject=id type=un;
```

$$\begin{aligned} y_{ij} = &[\gamma_{00} + \gamma_{10}(TIME_{ij}) + \gamma_{01} Treat_j + \gamma_{11} TIME_{ij} Treat_j] \\ &+ [u_{01} + u_{11}(TIME_{ij}) + e_{ij}] \end{aligned}$$

Example: Math achievement, grade 7-11

- Research Qs:
  - At what rate does math achievement increase?
  - Is rate of increase related to race, controlling for SES and gender?
- Sample: Longitudinal Study of American Youth, N=1322
- Variables:
  - LSAYid: person ID variable
  - Female (male=0; female=1)
  - Black
  - Grade (7—11): center on initial level– Grade7 = Grade-7
  - MathIRT (math achievement, IRT scaled) --- Outcome variable!
  - MathATT (attitude about mathematics, centered at grand mean) – a time-varying covariate
  - SES (continuous)

Data: Math achievement, grade 7-11

LSAYID	grade	grade7	female	black	mathirt	mathatt	ses
101101	7	0	0	0	67.89	-2.83	0.37
101101	8	1	0	0	63.44	-0.33	0.37
101101	9	2	0	0	67.05	-0.91	0.37
101101	10	3	0	0	73.60	-0.08	0.37
101101	11	4	0	0	76.24	-0.99	0.37
101102	7	0	0	0	58.04	1.67	0.22
101102	8	1	0	0	64.60	2.17	0.22
101102	9	2	0	0	66.31	0.34	0.22
101102	10	3	0	0	68.63	0.67	0.22
101102	11	4	0	0	67.69	0.17	0.22
101106	7	0	1	0	65.25	0.09	-0.78
101106	8	1	1	0	60.69	0.67	-0.78
101106	9	2	1	0	58.06	1.17	-0.78
101106	10	3	1	0	60.48	-0.58	-0.78
101106	11	4	1	0	76.12	-0.99	-0.78
101111	7	0	1	0	59.40	1.34	0.03
101111	8	1	1	0	54.78	0.92	0.03
101111	9	2	1	0	59.35	-1.08	0.03
101111	10	3	1	0	63.01	-0.49	0.03
101111	11	4	1	0	64.88	-1.41	0.03
...	..	..	..	..	...	...	...

Data is in long format!

Unconditional linear growth model

```
proc mixed data=mathach noclprint covtest method=ml;
   title 'Model A:  Unconditional linear growth model';
   class lsayid;
   model mathirt = grade7 / solution ddfm=bw notest;
   random intercept grade7 /subject=lsayid type=un;
   run;
```

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	52.3660	0.2541	1321	206.10	<.0001
grade7	2.8158	0.07322	5102	38.46	<.0001

$\hat{\gamma}_{00}$ Estimated mean math achievement in grade 7

$\hat{\gamma}_{10}$ Estimated yearly change in math ach.

Unconditional linear growth model

Random effects:

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	LSAYID	62.4944	3.3638	18.58	<.0001
UN(2,1)	LSAYID	6.4550	0.7011	9.21	<.0001
UN(2,2)	LSAYID	3.2164	0.2906	11.07	<.0001
Residual		37.1645	0.8552	43.46	<.0001

Variance in initial status $\hat{\tau}_{00}$

Variance of level 1 residuals $\hat{\sigma}^2$

Hypothesis tests for variance components

Variance in rate of change $\hat{\tau}_{11}$

Adding level 2 predictors: Race

Level 1: Within person

$$y_{ij} = \beta_{0j} + \beta_{1j}(Grade_{ij} - 7) + e_{ij}$$

where $e_{ij} \sim \mathcal{N}(0, \sigma^2)$

Level 2: Between person

$$\beta_{0j} = \gamma_{00} + \gamma_{01} Black_j + u_{0j},$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11} Black_j + u_{1j},$$

where $\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix}\right]$

Combined model:

$$y_{ij} = [\gamma_{00} + \gamma_{10}(Grade_{ij} - 7) + \gamma_{01} Black_j + \gamma_{11}(Grade_{ij} - 7) Black_j] + [u_{01} + u_{11}(Grade_{ij} - 7) + \epsilon_{ij}]$$

Adding level 2 predictors: Race

```
proc mixed data=mathach noclprint covtest method=ml;
  title2 'Model B:  Adding the effect of race';
  class lsayid;
  model mathirt = grade7 black black*grade7 / solution ddfm=bw outpm=modelb;
  random intercept grade7 /subject=lsayid type=un;
  run;
```

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	53.0170	0.2638	1320	201.00	<.0001
grade7	2.8688	0.07747	5101	37.03	<.0001
black	-5.9336	0.7969	1320	-7.45	<.0001
grade7*black	-0.4822	0.2341	5101	-2.06	0.0395

Plotting means

Using Level 2 coefficients:

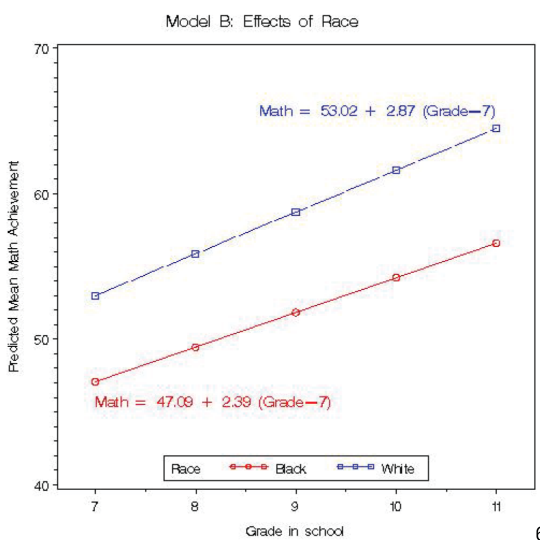
$$\hat{\beta}_{0j} = 53.02 - 5.93 Black$$
$$\hat{\beta}_{1j} = 2.87 - .48 Black$$

Substituting Black=0:

Math = 53.02 + 2.87(Grade-7)

Substituting Black=1

Math = 47.09 + 2.39(Grade-7)



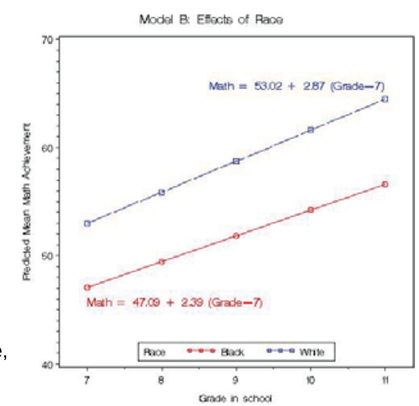
Plotting means: %meanplot

Get predicted means with **outpm** option

Plot with %meanplot

```
proc mixed data=mathach noclprint covtest method=ml;
  title2 'Model B:  Adding the effect of race';
  class lsayid;
  model mathirt = grade7 black black*grade7 /
                  solution ddfm=bw outpm=modelb;
  random intercept grade7/subject=lsayid type=un;
  run;

axis1 label=(a=90 'Predicted Mean Math Achievement')
%meanplot(data=modelb, response=pred, class=Grade Race,
          colors=red blue, lines=1 5, interp=rl);
```



In general, it is easier to interpret model results from a plot of means than a table of coefficients. Error bars or CIs help to show precision.

Adding more predictors

- Add SES as a Level 2 predictor of both initial level and rate of change
- Remove Black as Level 2 predictor of rate of change
- Add FEMALE as a level 2 predictor of initial level

Level 1: Within person	Level 2: Between person
$y_{ij} = \beta_{0j} + \beta_{1j}(Grade_{ij} - 7) + e_{ij}$ where $e_{ij} \sim \mathcal{N}(0, \sigma^2)$	$\beta_{0j} = \gamma_{00} + \gamma_{01} Black_j + \gamma_{02} SES + \gamma_{03} Female + u_{0j}$, $\beta_{1j} = \gamma_{10} + \gamma_{11} SES_j + u_{1j}$, where $\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix}\right]$

$$y_{ij} = [\gamma_{00} + \gamma_{10}(Grade_{ij} - 7) + \gamma_{01} Black_j + \gamma_{02} SES + \gamma_{03} Female + \gamma_{11}(Grade_{ij} - 7) SES_j] + [u_{01} + u_{11}(Grade_{ij} - 7) + \epsilon_{ij}]$$

Adding more predictors

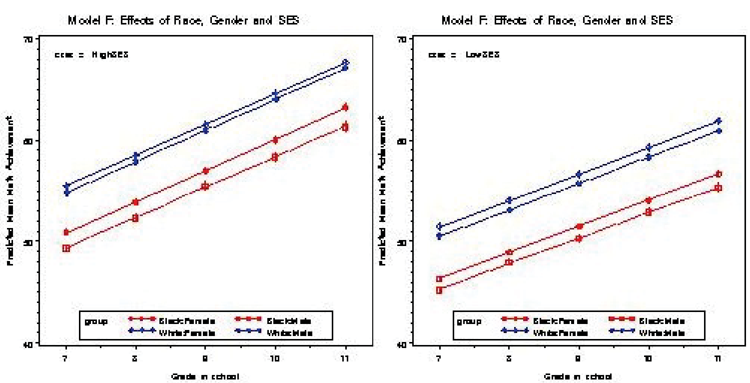
```
proc mixed data=mathach noclprint noinfo covtest method=ml;
  title2 'Model F:  Effect of SES only on rate of change';
  class lsayid;
  model mathirt = grade7 black ses female ses*grade7
              / solution ddfm=bw notest outpm=modelf;
  random intercept grade7 /subject=lsayid type=un;
  run;
```

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	52.4013	0.3504	1318	149.55	<.0001
grade7	2.8077	0.07286	5101	38.53	<.0001
black	-4.7982	0.7693	1318	-6.24	<.0001
ses	3.6159	0.3375	1318	10.71	<.0001
female	0.8183	0.4751	1318	1.72	0.0852
grade7*ses	0.3953	0.1017	5101	3.89	0.0001

Plotting means

```
data modelf;
  set modelf;
  group = put(black, race.) || ':' || put(female, sex.);
  cses = put(ses, ses.);
%meanplot(data=modelf, response=pred, class=grade group cses,
          colors=red red blue blue, lines=1 5 1 5, interp=rl);
```



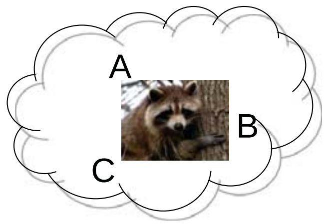
Some extensions

- **Generalized** linear mixed models
  - Analogous to extension of classical GLM to non-normal response distributions (PROC GENMOD; `glm()` in R)
  - E.g., binary outcomes (logistic), frequencies (Poisson), etc.
  - SAS: PROC GLIMMIX; R: `glmer()` in lme4 package
  - Model

Response distribution	$y_j \mid \mu_j = Binomial(\mu_j)$
Link function	$\boldsymbol{\eta}_j = g(\mu_j)$ e.g., $\boldsymbol{\eta}_j = \log(\mu_j / 1 - \mu_j)$
Linear predictor	$\boldsymbol{\eta}_j = \left[\underbrace{\mathbf{X}_j}_{\text{Fixed}} \quad \underbrace{\mathbf{Z}_j}_{\text{Random}}\right] \begin{pmatrix} \gamma \\ \mathbf{u}_j \end{pmatrix} = \mathbf{X}_j \gamma + \mathbf{Z}_j \mathbf{u}_j$

Example: Where the raccoons are?

- Raccoons photo'd in a park
- 3 sites: A, B, C
  - Spatial characteristics?
- Longitudinal:
  - L3: Year (1-5)
  - L2: Season (Fall, Spring)
  - L1: Week (1-4)
- Response: raccoon? (0/1)
  - Model: logistic
  - Fixed: Site Year Season Week
  - Random: Int Site? Week?



Standard logistic model could be used, but doesn't take dependencies into acct.

Mixed model can estimate Site variance, etc.

Some extensions

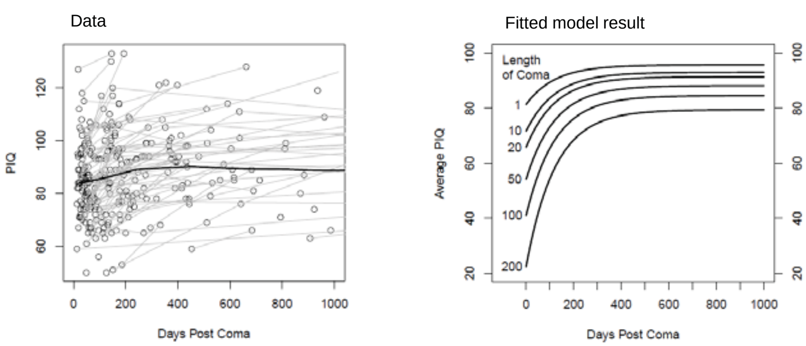
- **Non-linear** mixed models
  - Analogous to non-linear models with classical assumptions (independence, homoscedasticity)
  - Includes most generalized linear mixed models
  - Plus others, e.g., exponential growth/decay
  - SAS: PROC NLMIXED; R: `nlme()`

Curve type	Level 1 model
Hyperbolic	$y_{ij} = \beta_{0j} - \dfrac{1}{\beta_{1j} Time_{ij}} + e_{ij}$
Exponential	$y_{ij} = \beta_{0j} e^{\beta_{1j} Time_{ij}} + e_{ij}$

Example: Recovery from coma

Data from Wong etal. (2011) on recovery of performance IQ following a traumatic brain injury for patients in coma for varying length of time.

- Only 1.7 time points per patient on average!
- Use model of exponential growth



Summary

- Mixed models
  - Powerful methods for handling non-independence
  - Optimal compromise between pooling (ignoring nested structure) and by-group analysis
  - Highly flexible: incomplete data, various covariance structure, …
- Hierarchical data
  - Clear separation between effects at Level 1, Level 2, …
- Longitudinal data
  - Allows unequal time points, time-varying predictors
- Downside:
  - Classical GLM w/ fixed effects: familiar *F*, *t* tests (maybe wrong)
  - Need to understand the mixed model to interpret random effects & variance components