Mixed models for hierarchical & longitudinal data
Michael Friendly
Psychology 6140

Why Mixed models?
- More flexible for repeated measures 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)

Classical GLM
- The classical GLM, $y = X \beta + \epsilon$, assumes:
  - All observations are conditionally independent
    - $\rightarrow$ residuals, $\epsilon_i$, are uncorrelated
  - The model parameters, $\beta$, are fixed (non-random)
    - $\rightarrow$ 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

Fixed vs. random factors

<table>
<thead>
<tr>
<th></th>
<th>Fixed</th>
<th>Random</th>
</tr>
</thead>
<tbody>
<tr>
<td>Levels</td>
<td>Given # of possibilities</td>
<td>Selected at random from a population</td>
</tr>
<tr>
<td>New experiment</td>
<td>Use same levels</td>
<td>Use different levels from same population</td>
</tr>
<tr>
<td>Goal</td>
<td>Estimate means of fixed levels</td>
<td>Estimate variance of population of means, $\sigma^2_\mu$</td>
</tr>
<tr>
<td>Inference</td>
<td>Only for levels used $H_0: \mu_1 = \mu_2 = \ldots = \mu_k$</td>
<td>For all levels in the population $H_0: \sigma^2_\mu = 0$ $H_a: \sigma^2_\mu &gt; 0$</td>
</tr>
</tbody>
</table>
Terminology

- Different names for this modeling approach:
  - Hierarchical linear models (HLM)
  - Multilevel models (MLM)
  - Mixed models
  - 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

- 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?
HSB data: sector effects?

slight improvement: separate lines for each sector

problem: school dependency still ignored

- 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

Fixed effect approach

---

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;

NB: crucial to control for school here

One line for each school
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—infusion to these schools only, not a pop of schools.

Still assumes conditional independence and constant within-school $\sigma^2$.

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

$$y_{ij} = \beta_{0j} + \beta_{1j} \text{CSES}_{ij} + e_{ij}$$

Capture the coefficients, $(\beta_{0j}, \beta_{1j})$ and analyze these in relation to:

- Sector, school size, ...

Results are interpretable

- Public: lower math at mean CSES
- Public: greater slope for CSES

But:

- doesn’t take $n_i$ into account
- std errors still too small
- inferences maybe wrong
- hard to handle other nestings

A better (joint) plot shows individual slopes and intercepts in $\beta$ space

Data ellipses show the covariation within groups
Analyzing school-level variation

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: \( \sigma^2_{\text{schools}} > 0 \)?
  - 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: \( y_{ij} = \beta_{0j} + e_{ij} \)
  - \( \beta_{0j} \) 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: \( \gamma_{00} \) = grand mean of \( y \)
  \( u_{0j} \) = deviation of group \( j \) from GM

(Notation: I’m using \( \Gamma_{\text{fixed}} \) for fixed parameters, \( \text{Roman} \) for random parameters)
Basic multilevel model: ICC

- ICC: express variance of group means, $\beta_{0j}$, as proportion of total variance of $y_{ij}$
  
  \[ ICC = \frac{\tau_{00}}{\tau_{00} + \sigma^2} \]

- $0 \leq ICC \leq 1$: proportion of variance accounted for by school mean variation
  - ICC $= 0$: little variation among school means
  - ICC $= 1$: most variation of $y_{ij}$ ac'td 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>;
  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=schooltype=un;
  run;
  ```

- 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=schooltype=un;
  run;
  ```

- Other useful output:
  - Covariance Parameter Estimates
  - Solution for Fixed Effects
  - 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)

\[
y_{ij} = \beta_{0j} + \beta_{1j} \text{CSES}_{ij} + e_{ij} \]
\[
\beta_{0j} = \gamma_{00} + u_{0j} \quad \text{(intercept)}
\]
\[
\beta_{1j} = \gamma_{10} \quad \text{(slope)}
\]

Combined model:

\[
y_{ij} = [\gamma_{00} + \gamma_{10} \text{CSES}_{ij}] + [u_{0j} + u_{1j} \text{CSES}_{ij}] + e_{ij}
\]

Covariance Parameter Estimates

| Cov Parm | Subject | Estimate | Standard Error | Z Value | Pr > |t| |
|----------|---------|----------|----------------|---------|------|---|
| UN(1,1)  | school  | 8.6677   | 1.0784         | 8.04    | <.0001 | |
| Residual |         | 37.0108  | 0.6246         | 59.25   | <.0001 | |

Solution for Fixed Effects

| Effect   | Estimate  | Standard Error | DF  | t Value | Pr > |t| |
|----------|-----------|----------------|-----|---------|------|---|
| Intercept| 12.6493   | 0.2444         | 159 | 51.75   | <.0001 | |
| cses     | 2.1912    | 0.1087         | 7024| 20.17   | <.0001 | |

Random intercepts, random slopes

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

Reduced:

\[
y_{ij} = [\gamma_{00} + \gamma_{10} \text{CSES}_{ij}] + [u_{0j} + u_{1j} \text{CSES}_{ij}] + e_{ij}
\]
Random intercepts, random slopes

- Fixed effects estimates are similar to OLS
  - Mixed: \( E(y|\text{CSES}) = 12.65 + 2.19 \text{CSES} \)
  - OLS: \( E(y|\text{CSES}) = 12.76 + 2.19 \text{CSES} \)
- Standard errors more realistic with mixed model
  - (often larger—non-independence)
- Variance components: relative size of random effects
  - School means (overall achievement) \( \hat{\sigma}_{00}^2 = 8.68 \)
  - School slopes (~1/equity) \( \hat{\sigma}_{11}^2 = 0.69 \)
  - 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:
  \[
  y_{ij} = [\gamma_{00} + \gamma_{01} \text{SECT}_j + \gamma_{10} \text{CSES}_{ij} + \gamma_{11} \text{SECT}_j \text{CSES}_{ij}] + [u_{0j} + u_{1j} \text{CSES}_{ij}] + e_{ij}
  \]
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|\text{CSES}) = 11.41 + 2.80 \times \text{CSES}$$
  - Sector 1 (Catholic): $E(y|\text{CSES}) = 14.21 + 1.46 \times \text{CSES}$

  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

- 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 differences 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, $\hat{\sigma}^2$ still large: other lev 1?

Comparing models

<table>
<thead>
<tr>
<th>Parameter</th>
<th>0: Random ANOVA</th>
<th>1: Random Intercepts</th>
<th>2: Random Int &amp; Slope</th>
<th>3: Including SECTOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Int var</td>
<td>$\hat{\tau}_{00}$</td>
<td>8.610</td>
<td>8.668</td>
<td>8.677</td>
</tr>
<tr>
<td>Slope var</td>
<td>$\hat{\tau}_{11}$</td>
<td></td>
<td>0.692</td>
<td>0.266</td>
</tr>
<tr>
<td>Resid var</td>
<td>$\hat{\sigma}^2$</td>
<td>39.149</td>
<td>37.011</td>
<td>36.700</td>
</tr>
<tr>
<td>Intercept</td>
<td>$\hat{\gamma}_{00}$</td>
<td>12.637</td>
<td>12.649</td>
<td>12.649</td>
</tr>
<tr>
<td>CSES</td>
<td>$\hat{\gamma}_{10}$</td>
<td>2.191</td>
<td>2.193</td>
<td>2.803</td>
</tr>
<tr>
<td>sector</td>
<td>$\hat{\gamma}_{01}$</td>
<td></td>
<td>2.799</td>
<td></td>
</tr>
<tr>
<td>CSES*sector</td>
<td>$\hat{\gamma}_{11}$</td>
<td></td>
<td>-1.341</td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>47120.8</td>
<td>46728.0</td>
<td>46722.2</td>
<td>46511.7</td>
</tr>
</tbody>
</table>

Estimating random effects: BLUEs & BLUPS

- OLS regressions (within School) give Best Linear Unbiased Estimates (BLUEs) of
  - $\hat{\beta}_j = \left(\hat{\beta}_{0j}\right)$ with $\text{Var}(\hat{\beta}_j) = \hat{\sigma}^2(X_j^T X_j)^{-1}$

- Another estimate comes from random intercepts and slopes
  - $\hat{\mu}_j = \left(\hat{\mu}_{0j}\right)$ with $\text{Var}(\hat{\mu}_j) = \left(\hat{\tau}_{00} \hat{\tau}_{01} \hat{\tau}_{11}\right) = \hat{T}$

- A better estimate --- the BLUP (Best Linear Unbiased Predictor) is a weighted average of these, using $1/\text{Var}$ as weights
  - $\hat{\beta}_j = \left[\hat{\beta}_j \left[\text{Var}(\hat{\beta}_j)\right]^{-1} + \hat{\mu}_j \hat{T}^{-1}\right] \times \left[\text{Var}(\hat{\beta}_j)\right]^{-1} + \hat{T}^{-1}$

- 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 “smoothes” estimates toward the pooled estimate

Results for Model 3: Random intercepts and slopes

The BLUP estimates of $\hat{\beta}_{0j}$ are shrunk towards the OLS estimate but only slightly, because there is a large variance component for intercepts $\tau_{00}$.
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, $\tau_{11}$.

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.
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

```
ods 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;
random intercept cses / sub=school type=un;
run;
ods graphics off;
ods pdf close;
```

Influence plots

With many level 2 clusters, influential ones are less likely

Taxonomy of models

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

<table>
<thead>
<tr>
<th></th>
<th>Fixed (MODEL stmt)</th>
<th>Random</th>
<th>Combined formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random effects ANOVA</td>
<td>Intercept</td>
<td>Int</td>
<td>$y_i = \gamma_0 + u_{0j} + e_i$</td>
</tr>
<tr>
<td>Means as outcomes</td>
<td>Int G</td>
<td>Int</td>
<td>$y_i = \gamma_0 + \gamma_1 G_j + u_{0j} + e_i$</td>
</tr>
<tr>
<td>Random intercepts</td>
<td>Int X</td>
<td>Int</td>
<td>$y_i = \gamma_0 + \gamma_1 X_j + u_{0j} + e_i$</td>
</tr>
<tr>
<td>Random coefficients</td>
<td>Int X</td>
<td>Int X</td>
<td>$y_i = \gamma_0 + \gamma_1 X_j + u_{0j} + u_{1j} X_j + e_i$</td>
</tr>
<tr>
<td>Intercepts, slopes as outcomes</td>
<td>Int X G G*X</td>
<td>Int X</td>
<td>$y_i = \gamma_0 + \gamma_1 X_j + \gamma_0 G_j + \gamma_1 G_j X_j + u_{0j} + u_{1j} X_j + e_i$</td>
</tr>
<tr>
<td>Non-random slopes</td>
<td>Int X G G*X</td>
<td>Int</td>
<td>$y_i = \gamma_0 + \gamma_1 X_j + \gamma_0 G_j + \gamma_1 G_j X_j + u_{0j} + e_i$</td>
</tr>
</tbody>
</table>
The general linear mixed model

- Consider the outcomes, \( y_{ij}, i=1,\ldots,n_j \) within level 1 units \( j=1,\ldots,J \). \( y_j \) is the response vector for group \( j \).
- For group \( j \), the GLMM is

\[
y_j = \begin{bmatrix} X_j & Z_j \end{bmatrix} \gamma_j + Z_j u_j + \epsilon_j
\]

\[
\text{var}(\mathbf{u}) = \mathbf{T} = \begin{bmatrix} r_{00} & \cdots & \cdots & r_{0q} \\
\vdots & \ddots & \cdots & \ddots \\
r_{0q} & \cdots & \cdots & r_{qq} \end{bmatrix}
\]

\[ y_j = \gamma_j + \mathbf{X}_j \beta_j + \mathbf{Z}_j u_j + \epsilon_j \]

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, \( u_j \) and residuals, \( \epsilon_j \), are normally distributed, and mutually independent

\[
\begin{bmatrix} \mathbf{u}_j \\ \mathbf{e}_j \end{bmatrix} \sim \mathcal{N}\left( \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{T} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_j \end{bmatrix} \right)
\]

- The variance of \( \mathbf{y}_j \) is therefore \( \mathbf{Z}_j \mathbf{T} \mathbf{Z}_j^T + \mathbf{S}_j \)
- In most cases, \( \mathbf{T} \) is unstructured— all var/cov freely estimated & \( \mathbf{S}_j = \sigma^2 \mathbf{I} \)

- But mixed model allows more restricted & specialized structures
  - E.g., could estimate separate \( \mathbf{T} \) matrices for public/Catholic
  - Longitudinal data: \( \mathbf{S}_j \) = autoregressive (\( \rho_{ij} = \rho^{k-j} \))
  - If \( \mathbf{S}_j = \sigma^2 \mathbf{I} \) and no random effects, this reduces to std model

Covariance structures for \( \mathbf{T} \) & \( \mathbf{S} \)

<table>
<thead>
<tr>
<th>Structure (TYPE= option)</th>
<th>Parameters</th>
<th>(i,j)th element</th>
<th>Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unstructured UN</td>
<td>( t(t+1)/2 )</td>
<td>( \sigma_{ij} )</td>
<td>( \sigma^2 \begin{bmatrix} \sigma_1 &amp; \cdots &amp; \sigma_{ij} \ \cdots &amp; \cdots &amp; \cdots \ \sigma_1 &amp; \cdots &amp; \sigma_{ij} \end{bmatrix} )</td>
</tr>
<tr>
<td>Compound Symmetry CS</td>
<td>2</td>
<td>( \sigma_1 + \sigma_{ij} )</td>
<td>( \sigma^2 \begin{bmatrix} \sigma_1 &amp; \cdots &amp; \sigma_{ij} \ \cdots &amp; \cdots &amp; \cdots \ \sigma_1 &amp; \cdots &amp; \sigma_{ij} \end{bmatrix} )</td>
</tr>
<tr>
<td>First-order autoregressive AR(1)</td>
<td>2</td>
<td>( \sigma^2 \rho^{j-i} )</td>
<td>( \sigma^2 \begin{bmatrix} 1 &amp; \rho &amp; \rho^2 &amp; \rho^3 \ \rho &amp; 1 &amp; \rho &amp; \rho^2 \ \rho^2 &amp; \rho &amp; 1 &amp; \rho \ \rho^3 &amp; \rho^2 &amp; \rho &amp; 1 \end{bmatrix} )</td>
</tr>
</tbody>
</table>

There are many more possibilities for special forms of dependence.
Mostly, these are used in special situations; the GLMM provides them.

<table>
<thead>
<tr>
<th>Structure</th>
<th>Description</th>
<th>Parameters</th>
<th>( (i,j) \text{th element} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANTE(1)</td>
<td>Ante-dependence ( 2r - 1 )</td>
<td>( \sigma \sigma \Gamma_{i,j}^{r-1} \rho_k )</td>
<td>( \sigma )</td>
</tr>
<tr>
<td>AR(1)</td>
<td>Autoregressive ( 1 ) ( r )</td>
<td>( \sigma^2 \rho_{i-j} )</td>
<td>( \sigma )</td>
</tr>
<tr>
<td>ARH(1)</td>
<td>Heterogeneous AR(1) ( t + 1 )</td>
<td>( \sigma \gamma_r )</td>
<td>( \gamma_r )</td>
</tr>
<tr>
<td>ARMA(1,1)</td>
<td>ARMA(1,1) ( 3 )</td>
<td>( \sigma^2 \gamma_r )</td>
<td>( \gamma_r )</td>
</tr>
<tr>
<td>CS</td>
<td>Compound Symmetry ( 2 )</td>
<td>( \sigma_1 + \sigma^2_1(i = j) )</td>
<td>( \sigma_1 )</td>
</tr>
<tr>
<td>CSH</td>
<td>Heterogeneous CS ( t + 1 )</td>
<td>( \sigma \sigma \gamma )</td>
<td>( \gamma )</td>
</tr>
<tr>
<td>FA(q)</td>
<td>Factor Analytic ( 2(2q + 1) )</td>
<td>( \Sigma_{i=q}^{m} \lambda_{i,j} )</td>
<td>( \lambda_{i,j} )</td>
</tr>
<tr>
<td>FAD(q)</td>
<td>No Diagonal FA ( 2(2q + 1) )</td>
<td>( \Sigma_{i=q}^{m} \lambda_{i,j} )</td>
<td>( \lambda_{i,j} )</td>
</tr>
<tr>
<td>FAQR(q)</td>
<td>Equal Diagonal FA ( 2(2q + 1) )</td>
<td>( \Sigma_{i=q}^{m} \lambda_{i,j} )</td>
<td>( \lambda_{i,j} )</td>
</tr>
<tr>
<td>HP</td>
<td>Huyhn-Field ( t + 1 )</td>
<td>( \sigma^2_i + \sigma^2_j \lambda )</td>
<td>( \sigma^2_i, \sigma^2_j, \lambda )</td>
</tr>
<tr>
<td>LIN(q)</td>
<td>General Linear ( q )</td>
<td>( \Sigma_{i=1}^{q} A_{i,j} )</td>
<td>( A_{i,j} )</td>
</tr>
<tr>
<td>TOEP</td>
<td>Toeplitz ( t )</td>
<td>( \sigma_i )</td>
<td>( \sigma_i )</td>
</tr>
<tr>
<td>TOEP(q)</td>
<td>Banded Toeplitz ( q )</td>
<td>( \sigma_i \gamma_i )</td>
<td>( \gamma_i )</td>
</tr>
<tr>
<td>TOEPH</td>
<td>Heterogeneous TOEP ( 2r - 1 )</td>
<td>( \sigma \sigma \beta )</td>
<td>( \beta )</td>
</tr>
<tr>
<td>TOEPH(q)</td>
<td>Banded Hetero TOEP ( t + q - 1 )</td>
<td>( \sigma \sigma \beta \gamma )</td>
<td>( \beta \gamma )</td>
</tr>
<tr>
<td>UN</td>
<td>Unstructured ( t(t+1)/2 )</td>
<td>( \sigma )</td>
<td>( \sigma )</td>
</tr>
</tbody>
</table>

These require fewer parameters than the UNstructured (MANOVA) model.

Other cov. structures handle spatial dependence.

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 allow:
- Different number of time points over subjects
- Different time locations over subjects
- Time-varying predictors
- Several levels: individual \( \subset \) treatment \( \subset \) 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 \( \text{TIME}=0 \) \( \rightarrow \) initial status, or center (average status, etc.)

\[ y_i = \beta_{0j} + \beta_{1j}(\text{TIME}_i) + e_{ij} \text{ where } e_{ij} \sim \mathcal{N}(0, \sigma^2) \]

\[ \beta_{0j} = \gamma_{00} + u_{0j}, \beta_{1j} = \gamma_{10} + u_{1j}, \text{ where } \begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim \mathcal{N} \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix} \right] \]
Unconditional linear growth model

- Reduced form (combined model):
  \[ y_{ij} = [\gamma_{00} + \gamma_{10}TIME_{ij}] + [u_{0j} + u_{ij}TIME_{ij} + e_{ij}] \]

- 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

  \[ y_{ij} = \beta_{0j} + \beta_{ij}(TIME_{ij}) + \epsilon_{ij} \]

  where \( \epsilon_{ij} \sim N(0, \sigma^2) \)

  Combined model:

  \[ y_{ij} = [\gamma_{00} + \gamma_{10}(TIME_{ij}) + \gamma_{01}Treat_j + \gamma_{11}TIME_{ij}Treat_j] \]

  \[ + [u_{0j} + u_{ij}(TIME_{ij}) + e_{ij}] \]

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;
  ```

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

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

Estimated mean math achievement in grade 7
Estimated yearly change in math achievement

Covariance Parameter Estimates

<table>
<thead>
<tr>
<th>Cov Parm</th>
<th>Subject</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Value</th>
<th>Pr Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>UN(1,1)</td>
<td>LSAYID</td>
<td>62.4944</td>
<td>3.3638</td>
<td>18.58</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>UN(2,1)</td>
<td>LSAYID</td>
<td>6.4550</td>
<td>0.7011</td>
<td>9.21</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>UN(2,2)</td>
<td>LSAYID</td>
<td>3.2164</td>
<td>0.2906</td>
<td>11.07</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Residual</td>
<td></td>
<td>37.1645</td>
<td>0.8552</td>
<td>43.46</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

Random effects:

Variance in initial status
Variance of level 1 residuals
Variance in rate of change

Hypothesis tests for variance components

Adding level 2 predictors: Race

Level 1: Within person

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

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

Level 2: Between person

\[ \beta_{0j} = \gamma_{00} + \gamma_{01}Black_j + u_{0j}, \]
\[ \beta_{1j} = \gamma_{10} + \gamma_{11}(Grade_{ij} - 7)Black_j + u_{1j}, \]

where \( \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{bmatrix} \right) \)

Combined model:

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

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

Plotting means: %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=r);

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

```
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

```
       Estimate      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=ri);
```

Some extensions

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

  \[
  \begin{align*}
  y_j &| \mu_j = \text{Binomial}(\mu_j) \\
  \eta_j & = g(\mu_j) \quad \text{e.g.,} \ \eta_j = \log(\mu_j / 1 - \mu_j) \\
  \eta_j & = \left[ X_j \ Z_j \right] \gamma + Z_j u_j
  \end{align*}
  \]

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

### Curve type | Level 1 model
---|---
Hyperbolic | $y_{ij} = \beta_0 + \frac{1}{\beta_1 T_{ij}} + e_{ij}$
Exponential | $y_{ij} = \beta_0 e^{\beta_1 T_{ij}} + e_{ij}$

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