Overview and resources

- Overview
- Listserv:
  http://www.jiscmail.ac.uk/lists/multilevel.html
- Web site and links:
  www.uoregon.edu/~stevensj/HLM-II
- Software:
  HLM    MLWinN    Mplus    SAS
  SPSS   R and S-Plus   WinBugs
Workshop Overview

- Rationale for multilevel modeling
- Four examples as demonstrations of the power and flexibility of multilevel models
  - Achievement gap
  - Meta analysis
  - Longitudinal models of school effects
  - Interrupted time series
- Introduction to several technical issues as we discuss examples
- Lots of “how-to” information in last year’s workshop
Grouping and membership in particular units and clusters are important.
Hierarchical Data Structures

Many social and natural phenomena have a nested or clustered organization:

- Children within classrooms within schools
- Patients in a medical study grouped within doctors within different clinics
- Children within families within communities
- Employees within departments within business locations
Hierarchical Data Structures

More examples of nested or clustered organization:

- Children within peer groups within neighborhoods
- Respondents within interviewers or raters
- Effect sizes within studies within methods (meta-analysis)
- Multistage sampling
- Time of measurement within persons within organizations
Simpson’s Paradox: Clustering Is Important

Well known paradox in which performance of individual groups is reversed when the groups are combined.

<table>
<thead>
<tr>
<th></th>
<th>Quiz 1</th>
<th>Quiz 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gina</td>
<td>60.0%</td>
<td>10.0%</td>
</tr>
<tr>
<td>Sam</td>
<td>90.0%</td>
<td>30.0%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Quiz 1</th>
<th>Quiz 2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gina</td>
<td>60 / 100</td>
<td>1 / 10</td>
<td>61 / 110</td>
</tr>
<tr>
<td>Sam</td>
<td>9 / 10</td>
<td>30 / 100</td>
<td>39 / 110</td>
</tr>
</tbody>
</table>
Simpson’s Paradox: Other Examples

2006 US School study:
• In past research, private schools achieve higher than public schools
• Study was expected to provide additional support to the idea that private and charter schools perform better
• USED study (using multilevel modeling):
  • Unanalyzed math and reading higher for private schools
  • After taking demographic grouping into account, there was little difference between public and private and differences were almost equally split in favor of each school type

1975 Berkeley sex bias case:
• UCB sued for bias by women applying to grad school
  • Admission figures showed then men were more likely to be admitted
  • When analyzed by individual department, it turned out that men applied more to high admission departments and women to low admission departments
  • “When the Oakies left Oklahoma and moved to California, it raised the IQ of both states.”
    – Will Rogers
First Example: Does Multilevel Modeling Matter?

- The Analysis of School Effects
  - Individual Level Analysis
  - Analysis of School Level Aggregates
  - Multilevel Analysis
- The Intraclass Correlation Coefficient (ICC)
- Fixed and Random Effects
Why Is Multilevel Analysis Needed?

- Nesting creates dependencies in the data
  - Dependencies violate the assumptions of traditional statistical models (“independence of error”, “homogeneity of regression slopes”)
  - Dependencies result in inaccurate statistical estimates
- Important to understand variation at different levels
Decisions About Multilevel Analysis

- Properly modeling multilevel structure often matters (and sometimes a lot)
- Partitioning variance at different levels is useful
  - tau and sigma ($\sigma^2_Y = \tau^2 + \sigma^2$)
  - policy & practice implications
- Correct coefficients and unbiased standard errors
- Cross-level interaction
- Understanding and modeling site or cluster variability
Example 1: Achievement Gap

Data Example from New Mexico State accountability system, 2001 reading data for grade 6 children, N = 5,544, j=36

Example used here examines relationship between ethnicity (Hispanic, Native American, Other, or White) and reading achievement as measured on the TerraNova standardized test.

First analysis considers all 5,544 students without taking school membership into account.

Second analysis considers the 36 schools without taking students into account.

Third analysis considers both the 5,544 students and the 36 schools using a multilevel modeling approach.
Disaggregated analysis (N = 5,544 students)

### Model Summary

<table>
<thead>
<tr>
<th>Model</th>
<th>R</th>
<th>R Square</th>
<th>Adjusted R Square</th>
<th>Std. Error of the Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.389</td>
<td>.151</td>
<td>.151</td>
<td>36.128</td>
</tr>
</tbody>
</table>

### ANOVA

<table>
<thead>
<tr>
<th>Model</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Regression</td>
<td>3</td>
<td>429272.082</td>
<td>328.890</td>
<td>.000&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Residual</td>
<td>5540</td>
<td>1305.214</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>5543</td>
<td>8518703</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> Predictors: (Constant), OTHER, AMIND, HISP

<sup>b</sup> Dependent Variable: READ01

### Coefficients

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Std. Error</td>
<td>Beta</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>(Constant)</td>
<td>701.164</td>
<td>.795</td>
<td>881.726</td>
</tr>
<tr>
<td></td>
<td>HISP</td>
<td>-31.449</td>
<td>1.046</td>
<td>-30.078</td>
</tr>
<tr>
<td></td>
<td>AMIND</td>
<td>-38.740</td>
<td>2.390</td>
<td>-16.211</td>
</tr>
<tr>
<td></td>
<td>OTHER</td>
<td>-22.486</td>
<td>1.993</td>
<td>-11.285</td>
</tr>
</tbody>
</table>

<sup>a</sup> Dependent Variable: READ01
Disaggregated analysis (N = 5,544 students)

\[ Y = 701.164 -31.449(X_1) -38.740(X_2) -22.486(X_3) + r \]

Interpretation: White students average 6\textsuperscript{th} grade reading performance is about 701 points; Hispanic students score on average 31 points less, American Indian students score on average 39 points less, and other ethnic categories of students score on average about 23 points less.
Another alternative is to analyze data at the aggregated group level.

The aggregated analysis considers the 36 middle schools without taking students into account.
Aggregated analysis ($J = 36$ schools)

### Model Summary

<table>
<thead>
<tr>
<th>Model</th>
<th>R</th>
<th>R Square</th>
<th>Adjusted R Square</th>
<th>Std. Error of the Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.895</td>
<td>.801</td>
<td>.782</td>
<td>7.17838</td>
</tr>
</tbody>
</table>

### ANOVA

<table>
<thead>
<tr>
<th>Model</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>6630.867</td>
<td>3</td>
<td>2210.289</td>
<td>42.894</td>
<td>.000a</td>
</tr>
<tr>
<td>Residual</td>
<td>1648.933</td>
<td>32</td>
<td>51.529</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>8279.800</td>
<td>35</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. Predictors: (Constant), OTHER, HISP, AMIND
b. Dependent Variable: READING

### Coefficients

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>(Constant)</td>
<td>715.355</td>
<td>4.203</td>
<td></td>
</tr>
<tr>
<td></td>
<td>HISP</td>
<td>-50.789</td>
<td>5.095</td>
<td>-1.031</td>
</tr>
<tr>
<td></td>
<td>AMIND</td>
<td>-60.006</td>
<td>6.155</td>
<td>-1.027</td>
</tr>
</tbody>
</table>

a. Dependent Variable: READING
Aggregated analysis ($J = 36$ schools)

\[ Y = 715.355 - 50.789(X_1) - 60.006(X_2) - 70.699(X_3) + r \]

Interpretation: White students average 6th grade reading performance is about 715 points; Hispanic students score on average 51 points less, American Indian students score on average 60 points less, and other ethnic categories of students score on average about 71 points less.
Multilevel Models

Unlike the two previous single-level regression models, multilevel modeling takes both levels (students and schools) into account simultaneously:

\[ Y_{ij} = \beta_{0j} + \beta_1(X_1) + r_{ij} \quad \text{Level 1} \]

\[ \beta_{0j} = \gamma_{00} + u_{0j} \quad \text{Level 2} \]

\[ \beta_1 = \gamma_{10} + u_{1j} \quad \text{Level 2} \]

Note that level 1 regression model parameters become outcomes at level 2.
Multilevel Models

- Variance associated with the level 1 units (students) is partitioned from variance associated with level 2 units (schools)
- In essence, a different regression model is fit within each school
- Differences in model parameters (slopes and intercepts) can then be analyzed from one school to another
- A fundamental question in multilevel analysis is how much the outcome differs in relation to the level 2 grouping variable (e.g., schools); this relationship is estimated by the intraclass correlation coefficient (ICC)
The Intraclass Correlation Coefficient (ICC) measures the correlation between a grouping factor and an outcome measure.

In common notation there are 1 to J groups.

If participants do not differ from one group to another, then the ICC = 0.

As participants’ outcome scores differ due to membership in a particular group, the ICC grows large.
Intraclass Correlation Coefficient (\(\rho\))

Total \(\sigma^2_Y = \tau^2 + \sigma^2\)

\[
\text{ICC} = \frac{\text{between unit variance}}{\text{total variance}}
\]

\[= \frac{\tau^2}{\tau^2 + \sigma^2}\]

When the ICC is 0, multilevel modeling is not needed and power is the same as a non-nested design.
Multilevel Analysis (N = 5,544 students nested in J = 36 schools)

Third analysis considers both the 5,544 students and the 36 schools using a multilevel modeling approach.

Within school variance = 226.091
Between school Variance = 1304.875
ICC = .148

Estimation of ICC an important result with policy implications, in and of itself. Over a large number of SER studies, ICC ranges from about 10-20%. 
### Comparing the Three Analyses

<table>
<thead>
<tr>
<th>Model</th>
<th>$R^2$</th>
<th>$F$</th>
<th>$b$</th>
<th>$SE$</th>
<th>$B$</th>
<th>$t$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Disaggregated</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>.389</td>
<td>328.890</td>
<td>701.164</td>
<td>.795</td>
<td>-.401</td>
<td>-30.078</td>
</tr>
<tr>
<td>Hispanic</td>
<td></td>
<td></td>
<td>-31.449</td>
<td>1.046</td>
<td>-.208</td>
<td>-16.211</td>
</tr>
<tr>
<td>Amer. Indian</td>
<td></td>
<td></td>
<td>-38.740</td>
<td>2.390</td>
<td>-.147</td>
<td>-11.285</td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td>-22.486</td>
<td>1.993</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Aggregated</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>.895</td>
<td>42.894</td>
<td>715.355</td>
<td>4.203</td>
<td>-1.031</td>
<td>-9.969</td>
</tr>
<tr>
<td>Hispanic</td>
<td></td>
<td></td>
<td>-50.789</td>
<td>5.095</td>
<td>-1.027</td>
<td>-9.750</td>
</tr>
<tr>
<td>Amer. Indian</td>
<td></td>
<td></td>
<td>-60.006</td>
<td>6.155</td>
<td>-.305</td>
<td>-3.282</td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td>-70.699</td>
<td>21.540</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Multilevel</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>Level 1</td>
<td>.156</td>
<td>695.412</td>
<td>1.722</td>
<td>-.308</td>
<td>-16.102</td>
</tr>
<tr>
<td>Amer. Indian</td>
<td>Level 2</td>
<td>.697</td>
<td>-24.109</td>
<td>1.497</td>
<td>-.154</td>
<td>-10.504</td>
</tr>
<tr>
<td>Other</td>
<td>Level 2</td>
<td></td>
<td>-28.703</td>
<td>2.733</td>
<td>-.131</td>
<td>-8.541</td>
</tr>
<tr>
<td>Level 1</td>
<td></td>
<td></td>
<td>-19.703</td>
<td>2.307</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Multilevel Model Specification

- Another important difference in the approaches is the greater flexibility of model specification in HLM

- Multilevel models preserve information about individual differences (level 1 variance)

- Multilevel models take groups into account and explicitly model group effects (level 2 variance)

- Multilevel models allow for the examination of interactions between the two levels
Multilevel Model Specification

- In single level regression models, only fixed effects are possible for many parameters (all groups the same on many model parameters; i.e., homogeneity of regression slopes assumption)

- How to conceptualize and model group level variation?

- Do groups vary on the model parameters (fixed versus random effects)?

- Can group level information predict outcomes?
The Single-Level, Fixed Effects Regression Model

\[ Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + r_i \]

- The parameters \( \beta_{kj} \) are considered fixed
  - One for all and all for one
  - Same values for all \( i \) and \( j \); the single level model
- The \( r_i \)'s are random: \( r_i \sim N(0, \sigma) \) and independent
- But what if the \( \beta_{kj} \) were random and variable?
Modeling variation at Level 2: Intercepts as Outcomes

\[
Y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + r_{ij}
\]

\[
\beta_{0j} = \gamma_{00} + \gamma_{0j}W_j + u_{0j}
\]

\[
\beta_{1j} = \gamma_{10} + u_{1j}
\]

- Predictors (W’s) at level 2 are used to model variation in intercepts between the j units
Modeling Variation at Level 2: Slopes as Outcomes

\[ Y_{ij} = \beta_{0j} + \beta_{1j} X_{1ij} + r_{ij} \]

\[ \beta_{0j} = \gamma_{00} + \gamma_{0j} W_j + u_{0j} \]

\[ \beta_{1j} = \gamma_{10} + \gamma_{1j} W_j + u_{1j} \]

- Do slopes vary from one j unit to another?
- \( W \)'s can be used to predict variation in slopes as well
Fixed vs. Random Effects

- Fixed Effects represent discrete, purposefully selected or existing values of a variable or factor
  - Fixed effects exert constant impact on DV
  - Random variability only occurs as a within subjects effect (level 1)
  - Should only generalize to particular fixed values used

- Random Effects represent more continuous or randomly sampled values of a variable or factor
  - Random effects exert variable impact on DV
  - Variability occurs at level 1 and level 2
  - Can study and model variability
  - Can generalize to population of values
Fixed vs. Random Effects?

- Use fixed effects if
  - The groups are regarded as unique entities
  - If group values are determined by researcher through design or manipulation
  - Small j (< 10); improves power
- Use random effects if
  - Groups regarded as a sample from a larger population
  - Researcher wishes to test effects of group level variables
  - Researcher wishes to understand group level differences
  - Small j (< 10); improves estimation
Variance Components Analysis

- VCA allows estimation of the size of random variance components
  - Important issue when unbalanced designs are used
  - Iterative procedures must be used (usually ML estimation)
- Allows significance testing of whether there is variation in the components (parameters) across units
Achievement Gap Example Again

- Random effects allows parameters to vary across schools
- Introduces an entirely different set of research questions, for example:
  - Does the relationship between reading achievement and ethnic group differ from one school to another?
  - Can the differences in the ethnicity-reading achievement relationship be explained by characteristics of the schools?
Multilevel Analysis with Level 2 Predictors
(N = 5, 544 students nested within J = 36 schools)

Final estimation of fixed effects
(with robust standard errors)

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>T-ratio</th>
<th>d.f.</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRCPT1, B0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTRCPT2, G00</td>
<td>694.377894</td>
<td>1.590546</td>
<td>436.566</td>
<td>34</td>
<td>0.000</td>
</tr>
<tr>
<td>ANGLO, G01</td>
<td>24.756614</td>
<td>5.684485</td>
<td>4.355</td>
<td>34</td>
<td>0.000</td>
</tr>
<tr>
<td>HISP slope, B1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTRCPT2, G10</td>
<td>-22.994825</td>
<td>1.654343</td>
<td>-13.900</td>
<td>34</td>
<td>0.000</td>
</tr>
<tr>
<td>ANGLO, G11</td>
<td>3.049102</td>
<td>5.680311</td>
<td>0.537</td>
<td>34</td>
<td>0.594</td>
</tr>
<tr>
<td>AMIND slope, B2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTRCPT2, G20</td>
<td>-27.142110</td>
<td>2.525960</td>
<td>-10.745</td>
<td>34</td>
<td>0.000</td>
</tr>
<tr>
<td>ANGLO, G21</td>
<td>23.113476</td>
<td>9.189661</td>
<td>2.515</td>
<td>34</td>
<td>0.017</td>
</tr>
<tr>
<td>OTHER slope, B3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTRCPT2, G30</td>
<td>-20.440687</td>
<td>2.350818</td>
<td>-8.695</td>
<td>34</td>
<td>0.000</td>
</tr>
<tr>
<td>ANGLO, G31</td>
<td>18.360940</td>
<td>9.255701</td>
<td>1.984</td>
<td>34</td>
<td>0.055</td>
</tr>
</tbody>
</table>
Summary of Example 1 – Structure Matters!

- Correct statistical estimates
- ICC, separating parts from whole
- Understanding relations within and across levels
Example 2: Meta-Analysis

- Can estimation techniques used in HLM provide a more sophisticated way to synthesize quantitative results across studies?
- Example from Raudenbush & Bryk (2002)
  - Teacher expectancy ("the Pygmalion effect")
  - Contentious literature (see Wineburg, 1987; Rosenthal, 1987)
- Parameter Reliability
- Empirical Bayes Estimation
Statistical Estimation in HLM Models

- Estimation Methods
  - FML
  - RML
  - Empirical Bayes estimation

- Parameter estimation
  - Coefficients and standard errors
  - Variance Components

- Parameter reliability
Estimation Methods: Maximum Likelihood (ML)

- ML estimates model parameters by estimating a set of population parameters that maximize a likelihood function.
- The likelihood function provides the probabilities of observing the sample data given particular parameter estimates.
- ML methods produce parameters that maximize the probability of finding the observed sample data.
### Estimation Methods

<table>
<thead>
<tr>
<th>RML – Restricted Maximum Likelihood</th>
<th>FML – Full Maximum Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>only the variance components</td>
<td>both the regression coefficients and the variance components are included in the likelihood function</td>
</tr>
</tbody>
</table>

**Restricted:** Sequentially estimates the fixed effects and then the variance components.

**Goodness of fit statistics (deviance tests) apply only to the random effects.**

**RML only tests hypotheses about the VCs (and the models being compared must have identical fixed effects).**

**Goodness of fit statistics apply to the entire model (both fixed and random effects).**

Check on software default.
Estimation Methods

- RML expected to lead to better estimates especially when \( j \) is small
- FML has two advantages:
  - Computationally easier
  - With FML, overall chi-square statistic tests both regression coefficients and variance components, with RML only variance components are tested
  - Therefore if fixed portion of two models differ, must use FML for nested deviance tests
Several algorithms exist for existing HLM models:
- Expectation-Maximization (EM)
- Fisher scoring
- Iterative Generalized Least Squares (IGLS)
- Restricted IGLS (RIGLS)

All are iterative search and evaluation procedures
Model Estimation

- Iterative estimation methods usually begin with a set of start values
- Start values are tentative values for the parameters in the model
  - Program begins with starting values (usually based on OLS regression at level 1)
  - Resulting parameter estimates are used as initial values for estimating the HLM model
Model Estimation

- Start values are used to solve model equations on first iteration
- This solution is used to compute initial model fit
- Next iteration involves search for better parameter values
- New values evaluated for fit, then a new set of parameter values tried
- When additional changes produce no appreciable improvement, iteration process terminates (convergence)
- Note that convergence and model fit are very different issues
Parameter estimation

- Coefficients and standard errors estimated through maximum likelihood procedures (usually)
  - The ratio of the parameter to its standard error produces a Wald test evaluated through comparison to the normal distribution ($z$)
  - In HLM software, a more conservative approach is used:
    - t-tests are used for significance testing
    - t-tests more accurate for fixed effects, small n, and nonnormal distributions

- Standard errors
- Variance components
"I think you should be more explicit here in step two."
Parameter reliability

- Analogous to score reliability: ratio of true score variance to total variance (true score + error)
- In HLM, ratio of true parameter variance to total variability
- For example, in terms of intercepts, parameter reliability, $\lambda$, is:

$$\lambda_j = \frac{\text{Var}(\beta_{0j})}{\text{Var}(\bar{Y}_j)} = \frac{\tau_{00}^2}{\tau_{00}^2 + \sigma^2/n_j}$$
Parameter reliability

\[ \lambda_j = \frac{n_j \rho_I}{1 + (n_j - 1) \rho_I} \]

<table>
<thead>
<tr>
<th>( n_j )</th>
<th>.05</th>
<th>.10</th>
<th>.20</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>.21</td>
<td>.36</td>
<td>.56</td>
</tr>
<tr>
<td>10</td>
<td>.34</td>
<td>.53</td>
<td>.71</td>
</tr>
<tr>
<td>20</td>
<td>.51</td>
<td>.69</td>
<td>.83</td>
</tr>
<tr>
<td>30</td>
<td>.61</td>
<td>.77</td>
<td>.88</td>
</tr>
<tr>
<td>50</td>
<td>.72</td>
<td>.85</td>
<td>.93</td>
</tr>
<tr>
<td>100</td>
<td>.84</td>
<td>.92</td>
<td>.96</td>
</tr>
</tbody>
</table>
Parameter reliability

\[ \lambda_j = \frac{n_j \rho_I}{1 + (n_j - 1) \rho_I} \]

Parameter Reliability

- ICC = 0.05
- ICC = 0.10
- ICC = 0.20

n per cluster
It is often of interest to estimate the random group effects \((\beta_{0j}, \beta_{1j})\).

This is accomplished using Empirical Bayes (EB) estimation.

The basic idea of EB estimation is to predict group values using two kinds of information:

- Group \(j\) data
- Population data obtained from the estimation of the regression model
Empirical Bayes

- If information from only group $j$ is used to estimate then we have the OLS estimate:

$$\beta_{0j} = \bar{Y}_j$$

- If information from only the population is used to estimate then the group is estimated from the grand mean:

$$\gamma_{00} = \bar{Y}_{..} = \sum_{j=1}^{N} \frac{n_j}{N} \bar{Y}_j$$
Empirical Bayes

A third possibility is to combine group level and population information. The optimal combination is an average weighted by parameter reliability:

\[ \beta^{EB}_{0j} = \lambda_j \beta_{0j} + (1 - \lambda_j) \gamma_{00} \]

The larger the reliability, the greater the weight of the group mean. The smaller the reliability, the greater the weight of the grand mean.

This results in the “posterior means” or EB estimates.
Bayesian Estimation

- Use of prior and posterior information improves estimation (depending on purpose)
- Estimates “shrink” toward the grand mean as shown in formula
- Amount of shrinkage depends on the “badness” of the unit estimate
  - Low reliability results in greater shrinkage (if $\lambda = 1$, there is no shrinkage; if $\lambda = 0$, shrinkage is complete, $\gamma_{00}$)
  - Small n-size within a j unit results in greater shrinkage, “borrowing” from larger units

$$\beta_{0,j}^{EB} = \lambda_j \beta_{0,j} + (1 - \lambda_j) \gamma_{00}$$
“Frankly, Harold, you’re beginning to bore everyone with your statistics.”
Example 2: Meta-Analysis

- Can estimation techniques used in HLM provide a more sophisticated way to synthesize quantitative results across studies?

- Example from Raudenbush & Bryk (2002)
  - Teacher expectancy ("the Pygmalion effect")
  - Contention

- Approach takes the standard error of effect size into account:

\[
SE(d_j) = (\tau + V_j)^{1/2}, \text{ where } V_j = 1/(n_j - 3)
\]

Note the effect of sample size on the standard error of the effect size.
Example 2: Meta-Analysis

- Term coined by Gene Glass in his 1976 AERA Presidential address
- An alternative to the traditional literature review
- Allows the reviewer to quantitatively combine and analyze the results from multiple studies
- Traditional literature review is based on the reviewer’s analysis and synthesis of study themes or conclusions
What is Meta-Analysis (MA)?

- Meta-analysis
  - Collects empirical results from multiple studies
  - Expresses all results on a common scale, effect size
  - Can analyze covariates of effect size
  - Draws conclusions about the “overall” effect across studies no matter what the original study conclusions were
- Thus a MA becomes a research study on research studies, hence the term "meta"
Example 2: Meta-Analysis through HLM

- Implemented through interactive DOS-based HLM programs rather than the Windows interface
- Involves estimation based on the observed variance-covariance matrix
- In this example, the v-c matrix is simply the study effect sizes and their standard errors
- Data file prepared with relevant variables (effect size, variance of effect size, predictors)
- Then an HLM “.mdm” file is created
<table>
<thead>
<tr>
<th>studyid</th>
<th>effsize</th>
<th>variance</th>
<th>weeks</th>
<th>var</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.030</td>
<td>0.016</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.120</td>
<td>0.022</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-0.140</td>
<td>0.028</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1.180</td>
<td>0.139</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.260</td>
<td>0.136</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>-0.060</td>
<td>0.011</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>-0.020</td>
<td>0.011</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>-0.320</td>
<td>0.048</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.270</td>
<td>0.027</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.800</td>
<td>0.063</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>0.540</td>
<td>0.091</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>0.180</td>
<td>0.050</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>-0.020</td>
<td>0.084</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>0.230</td>
<td>0.084</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>-0.180</td>
<td>0.025</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>-0.060</td>
<td>0.028</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>0.300</td>
<td>0.019</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>0.070</td>
<td>0.009</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>-0.070</td>
<td>0.030</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Will you be starting with raw data? y
Is the input file a v-known file? y
How many level-1 statistics are there? 1
How many level-2 predictors are there? 1
   Enter 8 character name for level-1 variable number 1: effsize

   Enter 8 character name for level-2 variable number 1: weeks
Input format of raw data file (the first field must be the character ID)
   format: <a2.3f12.3>
What file contains the data? c:\expect.dat

Enter name of MDM file: c:\expect.mdm
Will you be starting with raw data?  n
Enter name of MDM file: c:\expect.mdm

SPECIFYING AN HLM2 MODEL

Level-1 predictor variable specification
Which level-1 predictors do you wish to use?
  The choices are:
  For EFFSIZE enter 1
  level-1 predictor? (Enter 0 to end) 1

Level-2 predictor variable specification
Which level-2 variables do you wish to use?
  The choices are:
  For WEEKS enter 1
  Which level-2 predictors to model EFFSIZE?
  Level-2 predictor? (Enter 0 to end) 1

ADDITIONAL PROGRAM FEATURES

Select the level-2 variables that you might consider for inclusion as predictors in subsequent models.
  The choices are:
  For WEEKS enter 1
  Which level-2 variables to model EFFSIZE?
  Level-2 variable? (Enter 0 to end)
  Do you wish to use any of the testing procedures? _
The HLM analysis allows the use of Bayesian estimation methods to temper the estimates of study effect sizes.
Use of a covariate to account for variation in study effect size: Teacher expectancy as a function of unfamiliarity
Summary of Example 2 – Estimation Methods

- Advanced estimation methods (ML and Bayesian)
- More realistic estimates of model parameters tempered by available information (e.g., n, reliability)
Example 3: Longitudinal Models

- Growth models as an Alternative to NCLB Adequate Yearly Progress (AYP)
- HLM as a more flexible means to model repeated measures
  - Individual growth curves
  - Ability to model growth parameters

No Child Left Behind

- Purpose of legislation is to ensure the learning of all children
- Schools (and districts and states) judged on whether a sufficient proportion of students are learning each year
- Measure and report “Adequate Yearly Progress” (AYP) in each content area
- Disaggregation of results by ethnicity, economic advantage, disability, and ELL
- But does NCLB AYP validly reflect student learning?
No Child Left Behind

- NCLB and other recent federal mandates and programs place strong emphasis on “evidence based” or “scientifically based” research.

- Scientifically based research “...means research that involves the application of rigorous, systematic, and objective procedures to obtain reliable and valid knowledge relevant to education activities and programs” (NCLB, 2001)
No Child Left Behind

- However, NCLB methods appear to contradict the federal push for more rigorous, scientifically based evidence.

- Collectively, NCLB regulations prescribe an unusual form of case study design that must be used to evaluate school effectiveness for AYP.
NCLB accountability requirements impose a nonequivalent-groups, case study design for the evaluation of school effectiveness:

<table>
<thead>
<tr>
<th>Group</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A (4th grade)</td>
<td>X? O₁</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group B (4th grade)</td>
<td></td>
<td>X? O₂</td>
<td></td>
</tr>
<tr>
<td>Group C (4th grade)</td>
<td></td>
<td></td>
<td>X? O₃</td>
</tr>
</tbody>
</table>

- X? is used to indicate unknown treatment implementation
- AYP in NCLB is a simple comparison of one $O_t$ to a calculated target for improvement
How to Measure School Effectiveness?

- Estimating the impact a school has on students is a complex task; a problem in research or program evaluation design

- One of the most important challenges is separating “intake” to the school from “value added” by the school

- Raudenbush and Willms (1995) Type A and Type B effects or total causal effects vs. school effects

- Intake represents confounding pre-existing student differences as well as previous learning

- Intake also represents differences in group composition from school to school
The Analysis of Change

- Cross sectional comparisons do not likely measure change effectively/accurately
- Individual growth curve analysis an important tool for analyzing change
- HLM models are one mechanism for estimating growth curves
- Height analogy
Analogy: Measuring Physical Development

Measure Height

2004
AYP defined by requiring 100% of children to be at least 6’0” by 2014 and projecting backwards to year in which height is first measured

All children must grow enough in each year to show AYP; all children must be tall by 2014
Get your facts first, and then you can distort them as much as you please.

- Mark Twain quoted by Rudyard Kipling in *From Sea to Shining Sea*
Measuring Height Using Longitudinal Methods

2004

Growth

2005
Longitudinal models using HLM

- Level 1 defined as repeated measurement occasions
- Levels 2 and 3 defined as higher levels in the nested structure
- For example, longitudinal analysis of student achievement:
  - Level 1 = achievement scores at times 1 – t
  - Level 2 = student characteristics
  - Level 3 = school characteristics
Three important advantages of the HLM approach to repeated measures:

- Times of measurement can vary from one person to another
- Data do not need to be complete on all measurement occasions
- Growth parameters can be modeled at higher levels
HLM Longitudinal models

Level-1

\[ Y_{tij} = \pi_{0ij} + \pi_{1ij}(\text{time}) + e_{tij} \]

Level-2

\[ \pi_{0ij} = \beta_{00j} + \beta_{01j}(X_{ij}) + r_{0ij} \]

\[ \pi_{1ij} = \beta_{10j} + \beta_{11j}(X_{ij}) + r_{1ij} \]

Level-3

\[ \beta_{00j} = \gamma_{000} + \gamma_{001}(W_{1j}) + u_{00j} \]

\[ \beta_{10j} = \gamma_{100} + \gamma_{101}(W_{1j}) + u_{10j} \]
Curvilinear Longitudinal models

Level-1

\[ Y_{tij} = \pi_{0ij} + \pi_{1ij}(\text{time}) + \pi_{2ij}(\text{time}^2) + e_{tij} \]

Level-2

\[ \pi_{0ij} = \beta_{00j} + \beta_{01j}(X_{ij}) + r_{0ij} \]
\[ \pi_{1ij} = \beta_{10j} + \beta_{11j}(X_{ij}) + r_{1ij} \]
\[ \pi_{2ij} = \beta_{20j} + \beta_{21j}(X_{ij}) + r_{2ij} \]

Level-3

\[ \beta_{00j} = \gamma_{000} + \gamma_{001}(W_{1j}) + u_{00j} \]
\[ \beta_{10j} = \gamma_{100} + \gamma_{101}(W_{1j}) + u_{10j} \]
\[ \beta_{20j} = \gamma_{200} + \gamma_{201}(W_{2j}) + u_{20j} \]
Mathematics Achievement Predicted by Individual Characteristics

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Coefficient</th>
<th>SE</th>
<th>t</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>School Mean Achievement, $\gamma_{000}$</td>
<td>663.54</td>
<td>1.28</td>
<td>513.86</td>
<td>241</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>White Student, $\gamma_{010}$</td>
<td>14.62</td>
<td>0.77</td>
<td>18.88</td>
<td>241</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>LEP, $\gamma_{020}$</td>
<td>-16.00</td>
<td>1.19</td>
<td>-13.50</td>
<td>241</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Title 1 Student, $\gamma_{030}$</td>
<td>-11.10</td>
<td>1.44</td>
<td>-7.71</td>
<td>241</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Special Education, $\gamma_{040}$</td>
<td>-33.09</td>
<td>1.88</td>
<td>-17.62</td>
<td>241</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Modified Test, $\gamma_{050}$</td>
<td>-16.83</td>
<td>2.63</td>
<td>-6.40</td>
<td>241</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Free Lunch Student, $\gamma_{060}$</td>
<td>-7.75</td>
<td>1.13</td>
<td>-6.85</td>
<td>241</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Gender, $\gamma_{070}$</td>
<td>-1.21</td>
<td>0.59</td>
<td>-2.03</td>
<td>241</td>
<td>.042</td>
</tr>
<tr>
<td>School Linear Growth, $\gamma_{100}$</td>
<td>19.40</td>
<td>0.70</td>
<td>27.88</td>
<td>241</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>White Student, $\gamma_{110}$</td>
<td>-1.20</td>
<td>0.64</td>
<td>-1.86</td>
<td>241</td>
<td>.062</td>
</tr>
<tr>
<td>LEP, $\gamma_{120}$</td>
<td>0.70</td>
<td>1.13</td>
<td>0.60</td>
<td>241</td>
<td>.547</td>
</tr>
<tr>
<td>Title 1 Student, $\gamma_{130}$</td>
<td>-2.58</td>
<td>0.95</td>
<td>-2.72</td>
<td>241</td>
<td>.007</td>
</tr>
<tr>
<td>Special Education, $\gamma_{140}$</td>
<td>-2.16</td>
<td>1.67</td>
<td>-1.29</td>
<td>241</td>
<td>.196</td>
</tr>
<tr>
<td>Modified Test, $\gamma_{150}$</td>
<td>-2.43</td>
<td>2.47</td>
<td>-0.99</td>
<td>241</td>
<td>.325</td>
</tr>
<tr>
<td>Free Lunch Student, $\gamma_{160}$</td>
<td>-0.75</td>
<td>1.03</td>
<td>-0.73</td>
<td>241</td>
<td>.466</td>
</tr>
<tr>
<td>Gender, $\gamma_{170}$</td>
<td>-4.68</td>
<td>0.59</td>
<td>-7.98</td>
<td>241</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Fixed Effect</td>
<td>Coefficient</td>
<td>SE</td>
<td>t</td>
<td>df</td>
<td>p</td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>-------------</td>
<td>-----</td>
<td>---------</td>
<td>----</td>
<td>-------</td>
</tr>
<tr>
<td>School Curvilinear Growth, $\gamma_{200}$</td>
<td>-2.09</td>
<td>0.21</td>
<td>-9.78</td>
<td>241</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>White Student, $\gamma_{210}$</td>
<td>0.48</td>
<td>0.20</td>
<td>2.35</td>
<td>241</td>
<td>.019</td>
</tr>
<tr>
<td>LEP, $\gamma_{220}$</td>
<td>-0.10</td>
<td>0.36</td>
<td>-0.27</td>
<td>241</td>
<td>.790</td>
</tr>
<tr>
<td>Title 1 Student, $\gamma_{230}$</td>
<td>0.61</td>
<td>0.28</td>
<td>2.17</td>
<td>241</td>
<td>.030</td>
</tr>
<tr>
<td>Special Education, $\gamma_{240}$</td>
<td>0.61</td>
<td>0.50</td>
<td>1.22</td>
<td>241</td>
<td>.224</td>
</tr>
<tr>
<td>Modified Test, $\gamma_{250}$</td>
<td>-0.10</td>
<td>0.75</td>
<td>-0.14</td>
<td>241</td>
<td>.890</td>
</tr>
<tr>
<td>Free Lunch Student, $\gamma_{260}$</td>
<td>0.26</td>
<td>0.33</td>
<td>0.79</td>
<td>241</td>
<td>.427</td>
</tr>
<tr>
<td>Gender, $\gamma_{270}$</td>
<td>1.05</td>
<td>0.19</td>
<td>5.64</td>
<td>241</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>School Level</th>
<th>Level-1</th>
<th>Level-2</th>
<th>Variance Explained</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Achievement, $u_{00}$</td>
<td>242.78</td>
<td>184.89</td>
<td>23.8%</td>
</tr>
<tr>
<td>Linear Growth, $u_{10}$</td>
<td>41.46</td>
<td>30.68</td>
<td>26.0%</td>
</tr>
<tr>
<td>Curvilinear Growth, $u_{10}$</td>
<td>2.94</td>
<td>2.60</td>
<td>11.6%</td>
</tr>
</tbody>
</table>
Mathematics Achievement Predicted by School Characteristics

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Coefficient</th>
<th>SE</th>
<th>t</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>School Mean Achievement, $\gamma_{000}$</td>
<td>662.53</td>
<td>1.07</td>
<td>620.80</td>
<td>237</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Percent Bilingual Students, $\gamma_{001}$</td>
<td>4.19</td>
<td>4.00</td>
<td>1.05</td>
<td>237</td>
<td>.295</td>
</tr>
<tr>
<td>Percent LEP Students, $\gamma_{002}$</td>
<td>-0.99</td>
<td>4.56</td>
<td>-0.22</td>
<td>237</td>
<td>.828</td>
</tr>
<tr>
<td>Percent White Students, $\gamma_{003}$</td>
<td>19.55</td>
<td>3.72</td>
<td>5.25</td>
<td>237</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Percent Free Lunch, $\gamma_{004}$</td>
<td>-5.29</td>
<td>3.18</td>
<td>-1.67</td>
<td>237</td>
<td>.096</td>
</tr>
<tr>
<td>School Mean Linear Growth, $\gamma_{100}$</td>
<td>19.18</td>
<td>0.71</td>
<td>26.87</td>
<td>237</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Percent Bilingual Students, $\gamma_{101}$</td>
<td>-0.17</td>
<td>1.98</td>
<td>-0.09</td>
<td>237</td>
<td>.932</td>
</tr>
<tr>
<td>Percent LEP Students, $\gamma_{102}$</td>
<td>2.90</td>
<td>2.85</td>
<td>1.02</td>
<td>237</td>
<td>.309</td>
</tr>
<tr>
<td>Percent White Students, $\gamma_{003}$</td>
<td>3.51</td>
<td>2.74</td>
<td>1.28</td>
<td>237</td>
<td>.201</td>
</tr>
<tr>
<td>Percent Free Lunch, $\gamma_{004}$</td>
<td>-3.67</td>
<td>2.23</td>
<td>-1.65</td>
<td>237</td>
<td>.099</td>
</tr>
</tbody>
</table>
Mathematics Achievement Predicted by School Characteristics

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Coefficient</th>
<th>SE</th>
<th>t</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>School Curvilinear Growth, $\gamma_{200}$</td>
<td>-1.99</td>
<td>0.22</td>
<td>-9.10</td>
<td>237</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Percent Bilingual Students, $\gamma_{201}$</td>
<td>-0.12</td>
<td>0.57</td>
<td>-0.21</td>
<td>237</td>
<td>.834</td>
</tr>
<tr>
<td>Percent LEP Students, $\gamma_{202}$</td>
<td>0.39</td>
<td>0.84</td>
<td>0.46</td>
<td>237</td>
<td>.643</td>
</tr>
<tr>
<td>Percent White Students, $\gamma_{203}$</td>
<td>-1.11</td>
<td>0.75</td>
<td>-1.48</td>
<td>237</td>
<td>.138</td>
</tr>
<tr>
<td>Percent Free Lunch, $\gamma_{204}$</td>
<td>-1.17</td>
<td>0.64</td>
<td>1.84</td>
<td>237</td>
<td>.065</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>School Level Variance Component</th>
<th>Level-1</th>
<th>Level-2</th>
<th>Level-3</th>
<th>Variance Explained*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Achievement, $u_{00}$</td>
<td>242.78</td>
<td>184.89</td>
<td>123.96</td>
<td>33.0%</td>
</tr>
<tr>
<td>Linear Growth, $u_{10}$</td>
<td>41.46</td>
<td>30.68</td>
<td>29.54</td>
<td>3.7%</td>
</tr>
<tr>
<td>Curvilinear Growth, $u_{10}$</td>
<td>2.94</td>
<td>2.60</td>
<td>2.49</td>
<td>4.2%</td>
</tr>
</tbody>
</table>

* Percent level 2 residual variance explained by level 3 model.
Hispanic Students
Native American Students
White Students
Achievement Status versus Achievement Growth

- Also important to note that the two research design approaches and the two parameters represent very different things.
- In this example, the correlation between status and growth parameters was -.378.
  - Has important policy implications.
  - Varies substantially across content, assessment, and state system.
Inferences about School Performance

Conclusions depend on which outcome is chosen.

Validity of conclusions depends on methods used.

High Status, Low Growth

Low Status, Low Growth

Low Status, High Growth

High Status, High Growth
Relationships Between the Proportion of Free-Reduced Price Lunch (FRPL) in the School and Status \( (r = -0.56) \) or Growth \( (r = -0.17) \) in New Mexico Middle School Mathematics Achievement.
Relationships Between the Proportion of Limited English Proficient (LEP) Students in the School and Status ($r = -0.51$) or Growth ($r = -0.06$) in New Mexico Middle School Mathematics Achievement.
Relationships Between the Proportion of Hispanic Students in the School and 
Status ($r = -.30$) or Growth ($r = -.05$) in New Mexico Middle School 
Mathematics Achievement.
Relationships Between the Proportion of Native American Students in the School and Status ($r = -.35$) or Growth ($r = -.02$) in New Mexico Middle School Mathematics Achievement.
Relationship Between Schools Ranked on Status and Schools Ranked on Growth ($r = -.75$) in New Mexico Elementary School Reading Achievement.
Classifying Schools using Status or Growth: Rankings can Differ Substantially

Relationship Between Schools Ranked on Status and Schools Ranked on Growth in New Mexico Middle School Mathematics Achievement \( (r = -.12) \).
Example 4: Interrupted Time Series

- Variation on longitudinal growth models presented earlier
- Flexible modeling of intervention effects over time
- In progress study of reading intervention in Bethel School District
  - Examine effects of time of intervention on reading performance
  - Examine effects of “dosage” of intervention on reading performance
## Interrupted Time Series Designs: Change in Intercept

\[ Y_{ij} = \pi_{0i} + \pi_{1i} Time_{ij} + \pi_{2i} Treatment_{ij} + \varepsilon_{ij} \]

**When Treatment = 0:**

\[ Y_{ij} = \pi_{0i} + \pi_{1i} Time_{ij} + \varepsilon_{ij} \]

**When Treatment = 1:**

\[ Y_{ij} = (\pi_{0i} + \pi_{2i}) + \pi_{1i} Time_{ij} + \varepsilon_{ij} \]
Interrupted Time Series Designs

Treatment effect on level: \((\pi_0 + \pi_2)\)
Interrupted Time Series Designs: Change in Slope

When Treatment = 1:

\[ Y_{ij} = \pi_0 + \pi_1 Time_{ij} + \pi_3 Treatment Time_{ij} + \epsilon_{ij} \]

When Treatment = 0:

\[ Y_{ij} = \pi_0 \]

Treatment time expressed as 0’s before treatment and time intervals post-treatment (i.e., 0, 0, 0, 1, 2, 3)
Interrupted Time Series Designs

Treatment effect on slope: \((+\pi_3)\)
Change in Intercept \textbf{and} Slope

\[
Y_{ij} = \pi_{0i} + \pi_{1i} Time_{ij} + \pi_{2i} Treatment + \pi_{3i} Treatment Time_{ij} + \epsilon_{ij}
\]

Effect of treatment on intercept

Effect of treatment on slope

When \text{Treatment} = 0:

\[
Y_{ij} = \pi_{0i} + \pi_{1i} Time_{ij} + \epsilon_{ij}
\]
Interrupted Time Series Designs

Effect of treatment on intercept

Effect of treatment on slope
Preliminary example modeling
“What I did last summer”

- Prior to evaluating our treatment effects:
  - Examine nature of growth function
  - Explore the effect of summer drop in performance
Effect of summer break on intercept?

Effect of summer break on slope?
Testing change in intercept and slope after summer break:

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Coefficient</th>
<th>Error</th>
<th>T-ratio</th>
<th>d.f.</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRCPT1, P0</td>
<td>11.934266</td>
<td>1.496115</td>
<td>7.977</td>
<td>6</td>
<td>0.000</td>
</tr>
<tr>
<td>TIME slope, P1</td>
<td>24.969961</td>
<td>0.588547</td>
<td>42.426</td>
<td>6</td>
<td>0.000</td>
</tr>
<tr>
<td>INTERCHA slope, P2</td>
<td>-25.886013</td>
<td>0.844983</td>
<td>-30.635</td>
<td>6</td>
<td>0.000</td>
</tr>
<tr>
<td>SLOPECHA slope, P3</td>
<td>-0.777770</td>
<td>0.899231</td>
<td>-0.865</td>
<td>6</td>
<td>0.421</td>
</tr>
</tbody>
</table>
### Final estimation of level-1 and level-2 variance components:

<table>
<thead>
<tr>
<th>Random Effect</th>
<th>Standard Deviation</th>
<th>Variance Component</th>
<th>df</th>
<th>Chi-square</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRCPT1, R0</td>
<td>27.84145</td>
<td>775.14633</td>
<td>1362</td>
<td>12718.00625</td>
<td>0.000</td>
</tr>
<tr>
<td>TIME slope, R1</td>
<td>10.65064</td>
<td>113.43621</td>
<td>1362</td>
<td>3795.84626</td>
<td>0.000</td>
</tr>
<tr>
<td>INTERCHA slope, R2</td>
<td>4.62646</td>
<td>21.40417</td>
<td>1362</td>
<td>1508.33545</td>
<td>0.003</td>
</tr>
<tr>
<td>SLOPECHA slope, R3</td>
<td>10.03375</td>
<td>100.67614</td>
<td>1362</td>
<td>2352.66736</td>
<td>0.000</td>
</tr>
<tr>
<td>Level-1, E</td>
<td>9.33343</td>
<td>87.11298</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Final estimation of level-3 variance components:

<table>
<thead>
<tr>
<th>Random Effect</th>
<th>Standard Deviation</th>
<th>Variance Component</th>
<th>df</th>
<th>Chi-square</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRCPT1/INTRCPT2, U00</td>
<td>3.49659</td>
<td>12.22617</td>
<td>6</td>
<td>31.18004</td>
<td>0.000</td>
</tr>
<tr>
<td>TIME/INTRCPT2, U10</td>
<td>1.28195</td>
<td>1.64341</td>
<td>6</td>
<td>19.29324</td>
<td>0.004</td>
</tr>
<tr>
<td>INTERCHA/INTRCPT2, U20</td>
<td>1.86312</td>
<td>3.47123</td>
<td>6</td>
<td>20.65449</td>
<td>0.002</td>
</tr>
<tr>
<td>SLOPECHA/INTRCPT2, U30</td>
<td>2.15169</td>
<td>4.62977</td>
<td>6</td>
<td>39.43444</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Variation across students and schools?
A last thought or two:

- Better modeling tools can expand the richness of research questions
- Better models allow more nuanced understanding of educational and social phenomena

Supposing is good, but finding out is better.

- Mark Twain's Autobiography
Bibliography


Bibliography


Bibliography


