Introduction and Background to Multilevel Analysis

Dr. J. Kyle Roberts

Southern Methodist University
Simmons School of Education and Human Development
Department of Teaching and Learning
Background and History of Multilevel Analysis

- Robinson (1950) and the problem of contextual effects
- The “Frog-Pond” Theory
• Statistical models that are not hierarchical sometimes ignore nesting structure and therefore report underestimated standard errors.

• Multilevel techniques are more efficient than other typical OLS techniques.

• Multilevel techniques assume a General Linear Model framework and can thus perform all types of analyses.

• Multilevel techniques can go beyond questions of “Do schools differ?” to ask questions of “Why do schools differ?”
Types of Multilevel Structures

- Students nested in classrooms
- Students nested in schools
- Students nested in classrooms nested in schools
- Measurement occasions nested inside individuals
- Students cross-classified by middle and high school
Do We Really Need Multilevel Analysis?

- “All data are multilevel!”
- The problem of independence of observations
  - GPA and the SAT in different high schools
  - SES as it relates to school achievement
- The inefficiency of OLS techniques
Differences Between Multilevel and OLS Methods

- MLA is based on maximum likelihood and empirical Bayesian techniques
- $1 + 1 = 1.5$

\[
\bar{X}_1 \quad \bar{X}_2 \quad \bar{X} \quad \bar{X}_3
\]

\[
\beta_{01} \quad \beta_{02} \quad \gamma_{00} \quad \beta_{03}
\]
Notating the Multilevel ANOVA

- Recall from Analysis of Variance (ANOVA) that we may notate a model testing for differences between groups as:

\[ Y_{ij} = \mu + \alpha_j + \epsilon_{ij}, \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2), j = 1, \ldots, k, i = 1, \ldots, n_j \]

- We can further notate the above model as:

\[ Y_{ij} = \gamma_{00} + u_{0j} + e_{ij} \]

- which may be decomposed as a Level-1 model

\[ Y_{ij} = \beta_{0j} + e_{ij} \]

- and Level-2 model

\[ \beta_{0j} = \gamma_{00} + u_{0j} \]
Notating the Multilevel ANOVA (cont.)

- This means that any individual score within each group may be thought of as

\[
Y_{ij} = \beta_j + e_{ij}
\]

\[
\begin{align*}
Y_{11} &= \beta_1 + e_{11} \\
Y_{21} &= \beta_1 + e_{21} \\
Y_{31} &= \beta_1 + e_{31} \\
&\vdots \\
Y_{ij} &= \beta_j + e_{ij}
\end{align*}
\]

- And any group score may be modeled as:

\[
\begin{align*}
\beta_1 &= \gamma_{00} + u_{01} \\
\beta_2 &= \gamma_{00} + u_{02} \\
\beta_3 &= \gamma_{00} + u_{03} \\
&\vdots \\
\beta_j &= \gamma_{00} + u_{0j}
\end{align*}
\]
Graphical Example of Multilevel ANOVA

Original Estimates → Group Means → Grand Mean → Maximized Mean → Level 2, Schools → Level 1, Students

$X_{11}$ → $X_{12}$ → $X_{13}$ → $X_{14}$ → $\bar{X}_1$ → $\bar{X}$ → $\gamma_{00}$ → $\beta_{01}$ → $a_{11}$ → $\sigma^2_{e1}$

$X_{21}$ → $X_{22}$ → $X_{23}$ → $X_{24}$ → $\bar{X}_2$ → $\bar{X}$ → $\gamma_{00}$ → $\beta_{02}$ → $a_{21}$ → $\sigma^2_{e2}$

$X_{31}$ → $X_{32}$ → $X_{33}$ → $X_{34}$ → $\bar{X}_3$ → $\bar{X}$ → $\gamma_{00}$ → $\beta_{03}$ → $a_{31}$ → $\sigma^2_{e3}$

$X_{41}$ → $X_{42}$ → $X_{43}$ → $X_{44}$
Understanding Errors

\[ \sigma^2_u \]

\[ u_{01} \]

\[ u_{02} \]

\[ \bar{X}_1 \]

\[ \bar{X} \]

\[ \bar{X}_2 \]

\[ X_{11} \]

\[ X_{21} \]

\[ X_{12} \]

\[ X_{22} \]

\[ e_{11} \]

\[ e_{21} \]

\[ e_{12} \]

\[ e_{22} \]

\[ \sigma^2_e \]
Heuristic Example of Multilevel ANOVA

- Suppose that we have a dataset in which we have scores from 160 students nested inside 16 different schools.
- The dataset may be found at http://faculty.smu.edu/kyler/courses/7309/sciach.txt

```r
> sciach <- read.table("sciach.txt", header = T)
> str(sciach)

'data.frame': 160 obs. of 10 variables:
$ ID : int 1 2 3 4 5 6 7 8 9 10 ...
$ GROUP : int 1 1 1 1 1 1 1 1 1 1 ...
$ SCIENCE : int 1 1 2 2 3 3 4 4 5 5 ...
$ URBAN : int 8 7 7 6 6 5 5 5 3 2 ...
$ GENDER : int 1 1 1 1 1 2 2 2 2 2 ...
$ CONS : int 1 1 1 1 1 1 1 1 1 1 ...
$ URB.MEAN : num -6.43 -7.43 -7.43 -8.43 -8.43 ...
$ SCH.RES : num 5 5 5 5 5 5 5 5 5 5 ...
$ SCH.RES.MEAN: num -4.97 -4.97 -4.97 -4.97 -4.97 ...
$ GEND.FAC : Factor w/ 2 levels "Female","Male": 2 2 2 2 2 1 1 1 1 1

> sciach$GROUP <- factor(sciach$GROUP)
```
ANOVA of Science Achievement Data

- We can run a simple ANOVA on the science achievement data comparing the means of the 16 schools.

```R
> mean(sciach$SCIENCE)
[1] 10.6875

> with(sciach, tapply(SCIENCE, GROUP, mean))
   1   2   3   4   5   6   7   8   9  10  11  12
 3.0 4.0 5.0 6.0 7.0 8.0 9.0 10.0 11.0 12.0 13.0 14.0
   13  14  15  16
15.0 16.0 18.5 19.5

> anova(aov(SCIENCE ~ GROUP, sciach))

Analysis of Variance Table

Response: SCIENCE

<table>
<thead>
<tr>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GROUP</td>
<td>15</td>
<td>3859.4</td>
<td>257.292</td>
<td>130</td>
</tr>
<tr>
<td>Residuals</td>
<td>144</td>
<td>285.0</td>
<td>1.979</td>
<td></td>
</tr>
</tbody>
</table>
```
Running the Same Model as a Multilevel ANOVA

- Recall from the multilevel ANOVA notation that we want to test:

\[ Y_{ij} = \gamma_{00} + u_{0j} + e_{ij} \]

```r
> m0 <- lme(SCIENCE ~ 1, random = ~1 | GROUP, sciach)
> summary(m0)

Linear mixed-effects model fit by REML
  Data: sciach
    AIC   BIC   logLik
  643.8561 653.0628 -318.9281

Random effects:
  Formula: ~1 | GROUP
            (Intercept) Residual
            StdDev: 5.052846 1.406829

Fixed effects: SCIENCE ~ 1
  Value Std.Error   DF  t-value p-value
  (Intercept) 10.6875 1.268098 144   8.427975 0

Standardized Within-Group Residuals:

  Min      Q1     Med      Q3     Max
-1.4637  -0.7214  -0.0065   0.7084  1.4507
```
## Comparing OLS and MLA Estimates

```r
> cbind(means = with(sciach, tapply(SCIENCE, GROUP, + mean)), coef(m0))

<table>
<thead>
<tr>
<th>means</th>
<th>(Intercept)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.0</td>
</tr>
<tr>
<td>2</td>
<td>4.0</td>
</tr>
<tr>
<td>3</td>
<td>5.0</td>
</tr>
<tr>
<td>4</td>
<td>6.0</td>
</tr>
<tr>
<td>5</td>
<td>7.0</td>
</tr>
<tr>
<td>6</td>
<td>8.0</td>
</tr>
<tr>
<td>7</td>
<td>9.0</td>
</tr>
<tr>
<td>8</td>
<td>10.0</td>
</tr>
<tr>
<td>9</td>
<td>11.0</td>
</tr>
<tr>
<td>10</td>
<td>12.0</td>
</tr>
<tr>
<td>11</td>
<td>13.0</td>
</tr>
<tr>
<td>12</td>
<td>14.0</td>
</tr>
<tr>
<td>13</td>
<td>15.0</td>
</tr>
<tr>
<td>14</td>
<td>16.0</td>
</tr>
<tr>
<td>15</td>
<td>18.5</td>
</tr>
<tr>
<td>16</td>
<td>19.5</td>
</tr>
<tr>
<td></td>
<td>3.059135</td>
</tr>
<tr>
<td></td>
<td>4.051442</td>
</tr>
<tr>
<td></td>
<td>5.043750</td>
</tr>
<tr>
<td></td>
<td>6.036058</td>
</tr>
<tr>
<td></td>
<td>7.028365</td>
</tr>
<tr>
<td></td>
<td>8.020673</td>
</tr>
<tr>
<td></td>
<td>9.012981</td>
</tr>
<tr>
<td></td>
<td>10.005288</td>
</tr>
<tr>
<td></td>
<td>10.997596</td>
</tr>
<tr>
<td></td>
<td>11.989904</td>
</tr>
<tr>
<td></td>
<td>12.982212</td>
</tr>
<tr>
<td></td>
<td>13.974519</td>
</tr>
<tr>
<td></td>
<td>14.966827</td>
</tr>
<tr>
<td></td>
<td>15.959135</td>
</tr>
<tr>
<td></td>
<td>18.439904</td>
</tr>
<tr>
<td></td>
<td>19.432212</td>
</tr>
</tbody>
</table>
```
Checking Assumptions

```r
> par(mfrow = c(1, 2))
> boxplot(resid(m0) ~ GROUP, sciach, horizontal = T, 
>       main = "Homogeneity of Variance")
> qqnorm(resid(m0), main = "QQplot for Null Model")
```
Model Fit Indices

- Chi-square or $\chi^2$
  \[ \chi^2 = -2 \times \ell \]
- Akaike Information Criteria (AIC)
  \[ AIC = -2 \times \ell + 2K \]
- Bayesian Information Criteria (BIC)
  \[ BIC = -2 \times \ell + K \times Ln(N) \]
Fixed and Random Effects

(a) No Random Effects

(b) Random Intercepts
Fixed and Random Effects

(c) Random Slopes

(d) Random Intercepts and Slopes
Recalling the OLS Linear Model

Consider the following 1-level regression equation:

\[ y = a + bx + \epsilon \]

where:
- \( y \) - the response variable
- \( a \) - the \( y \)-intercept or the expected value when the covariate is 0
- \( b \) - the expected change in the response variable (\( y \)) for every one unit change in the covariate
- \( x \) - the covariate
- \( \epsilon \) - the residual term

This model may also be written as:

\[ y = \beta_0 + \beta_1 x + \epsilon \]
Adding a Random Effect for the Intercept

- We may further modify this equation to allow for variation among the intercepts for each pre-identified group such that:

\[ y = \beta_0 + \beta_1 x + \epsilon \]

now becomes:

\[ y_{ij} = \gamma_{00} + \gamma_{10} x_{ij} + u_{0j} + \epsilon_{ij} \]

where:

- \( y_{ij} \) - is the response variable for individual \( i \) in group \( j \)
- \( \gamma_{00} \) - the \( y \)-intercept or the expected value when the covariate is 0
- \( \gamma_{10} \) - the expected change in the response variable (\( y \)) for every one unit change in the covariate
- \( x_{ij} \) - the covariate term for each individual; the subscripts \( i \) and \( j \) mean that this variable is measured at the first level
- \( u_{0j} \) - the residual term defining the random variation of each of the group intercepts around the grand intercept \( \gamma_{00} \)
- \( \epsilon_{ij} \) - the residual term defining the random variation of each person around their predicted group regression equation.
Breaking Down the Mixed Effects Model into Levels

- From the previous slide, our mixed effects model:

\[ y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + u_{0j} + \epsilon_{ij} \]

may be thought of as a 2-level model where:

- Level 1:

\[ y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + \epsilon_{ij} \]

- Level 2:

\[ \beta_{0j} = \gamma_{00} + u_{0j} \]
\[ \beta_{1j} = \gamma_{10} \]

- where:

\[ u_{0j} \sim \mathcal{N}(0, \sigma_{u_{0j}}^2) \]
\[ \epsilon_{ij} \sim \mathcal{N}(0, \sigma_{\epsilon_{ij}}^2) \]
Understanding Errors Again

\[ u_{0j} \]

\[ \sigma^2_{u0} \]

\[ e_{ij} \]
Running the Linear Model in R

```r
> m.lm <- lm(SCIENCE ~ URBAN, sciach)
> summary(m.lm)

Call:
  lm(formula = SCIENCE ~ URBAN, data = sciach)

Residuals:
            Min          1Q   Median          3Q          Max
  -5.335800  -2.129210   0.491930   2.043180   5.009010

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   -1.25108   0.59371  -2.107  0.03670 *
URBAN          0.82763   0.03863  21.425  <2e-16 ***

Residual standard error: 2.592 on 158 degrees of freedom
Multiple R-squared:  0.7439, Adjusted R-squared:  0.7423
F-statistic: 459.1 on 1 and 158 DF,  p-value: < 2.2e-16
```
Running the Linear Model in R (cont.)

```r
> plot(SCIENCE ~ URBAN, sciach)
> abline(lm(SCIENCE ~ URBAN, sciach))
```
Running the Multilevel Model in R

```r
> m1 <- lme(SCIENCE ~ URBAN, random = ~1 | GROUP,
+     sciach)
> summary(m1)

Linear mixed-effects model fit by REML
  Data: sciach
    AIC    BIC   logLik
508.094 520.3444 -250.047

Random effects:
  Formula: ~1 | GROUP
        (Intercept) Residual
    StdDev:  9.29817  0.809449

Fixed effects: SCIENCE ~ URBAN
       Value Std.Error  DF  t-value p-value
(Intercept)  22.302911  2.4263101 143   9.192111    0
   URBAN     -0.805228  0.0479985 143  -16.776087    0

Correlation:
       (Intr)
URBAN -0.285
```
### Comparing Models

```r
> anova(m0, m1)

<table>
<thead>
<tr>
<th>Model</th>
<th>df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>Test</th>
<th>L.Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>m0</td>
<td>1</td>
<td>643.8561</td>
<td>653.0628</td>
<td>-318.9281</td>
<td></td>
<td></td>
</tr>
<tr>
<td>m1</td>
<td>2</td>
<td>508.0940</td>
<td>520.3444</td>
<td>-250.0470</td>
<td>1 vs 2</td>
<td>137.7621</td>
</tr>
</tbody>
</table>

**p-value**

- m0: <.0001
- m1: <.0001

- So we can see that by the addition of a single fixed effect to our model, we reduced the AIC by $\sim 135$ and the BIC by $\sim 133$. 