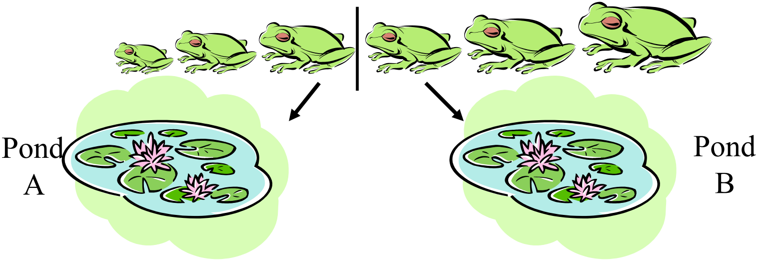
Introduction and Background to Multilevel Analysis

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



History, cont.

- 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

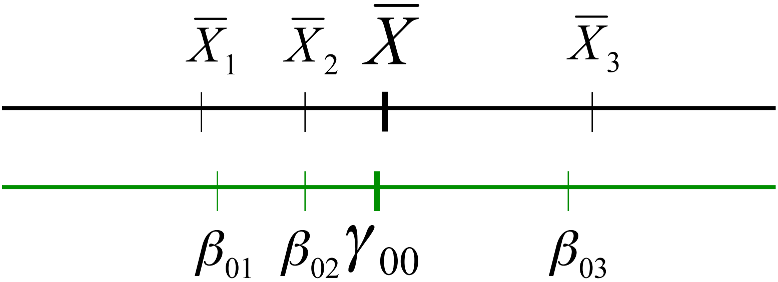
- Students nested in classrooms
- Students nested in schools
- Students nested in classrooms nested in schools
- Measurement occasions nested inside individuals
- Students cross-classified by middle and high school

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
- $1 + 1 = 1.5$



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.)

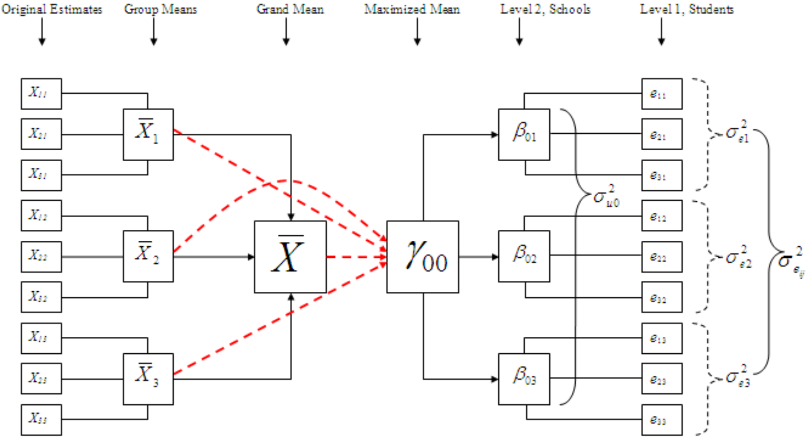
- This means that any individual score within each group may be thought of as

$$Y_{ij} = \beta_j + e_{ij} \begin{cases} Y_{11} = \beta_1 + e_{11} \\ Y_{21} = \beta_1 + e_{21} \\ Y_{31} = \beta_1 + e_{31} \\ \dots \\ Y_{ij} = \beta_j + e_{ij} \end{cases}$$

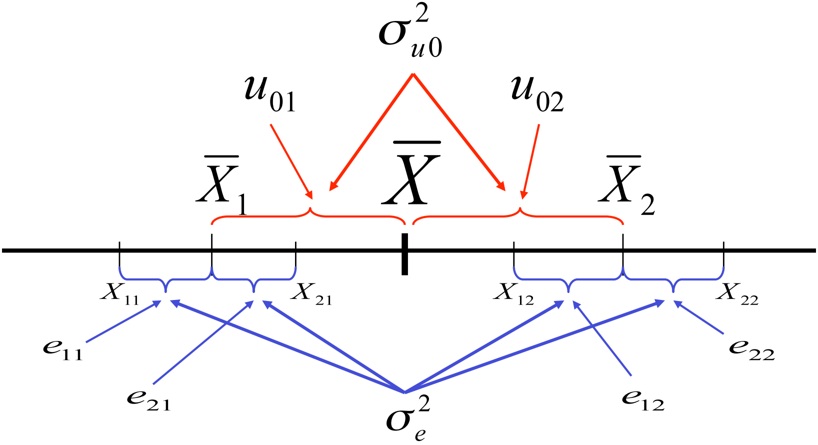
- And any group score may be modeled as:

$$\beta_j = \gamma_{00} + u_{0j} \begin{cases} \beta_1 = \gamma_{00} + u_{01} \\ \beta_2 = \gamma_{00} + u_{02} \\ \beta_3 = \gamma_{00} + u_{03} \\ \dots \\ \beta_j = \gamma_{00} + u_{0j} \end{cases}$$

Graphical Example of Multilevel ANOVA



Understanding Errors



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

```
> 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
```

```
> 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.

```
> 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
           Df Sum Sq Mean Sq F value    Pr(>F)
GROUP      15 3859.4 257.292     130 < 2.2e-16
Residuals 144  285.0   1.979
```

Running the Same Model as a Multilevel ANOVA

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

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

```
> 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:
```

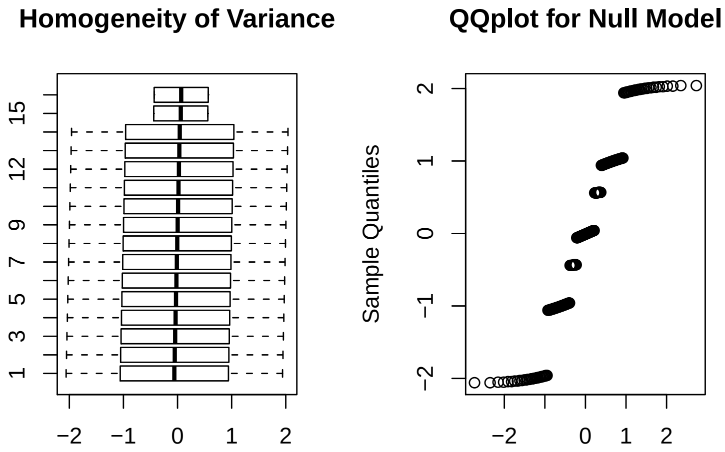
Comparing OLS and MLA Estimates

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

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

Checking Assumptions

```
> 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 χ^2

$$\chi^2 = -2 * \ell$$

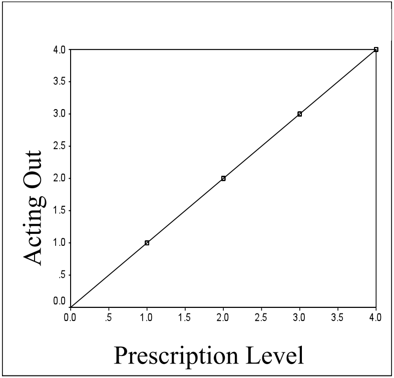
- Akaike Information Criteria (AIC)

$$AIC = -2 * \ell + 2K$$

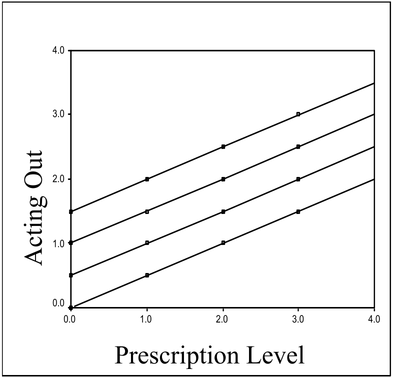
- Bayesian Information Criteria (BIC)

$$BIC = -2 * \ell + K * Ln(N)$$

Fixed and Random Effects

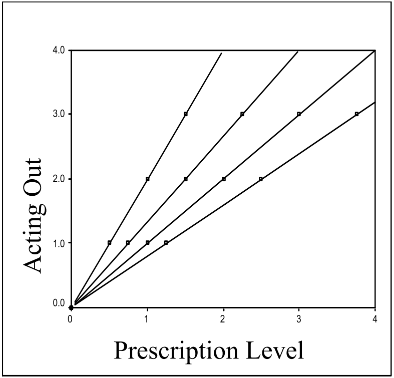


(a) No Random Effects

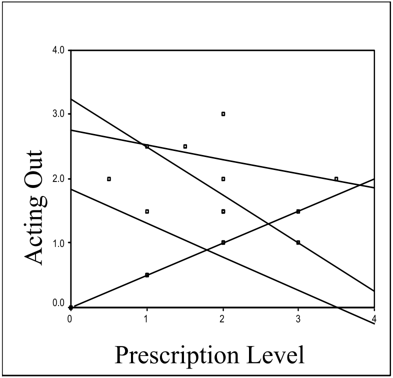


(b) Random Intercepts

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
  - ϵ - 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
  - γ_{00} - the y-intercept or the expected value when the covariate is 0
  - γ_{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 γ_{00}
  - ϵ_{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:

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

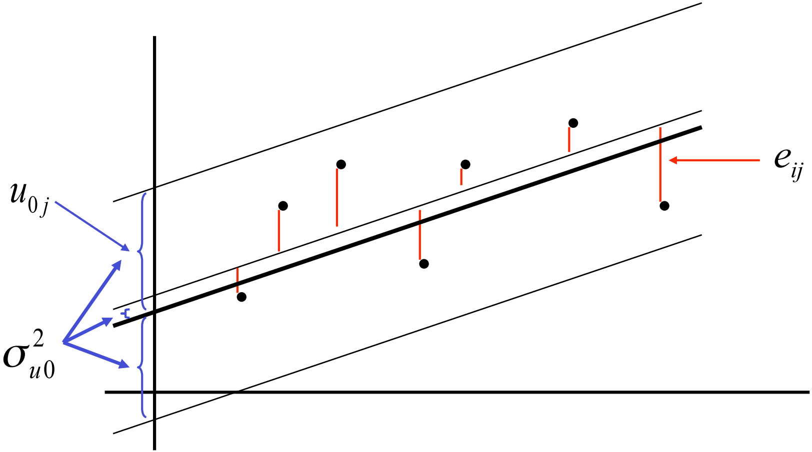
  - Level 2:

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

  - where:

$$\begin{aligned} u_{0j} &\sim \mathcal{N}(0, \sigma^2_{u_{0j}}) \\ \epsilon_{ij} &\sim \mathcal{N}(0, \sigma^2_{\epsilon_{ij}}) \end{aligned}$$

Understanding Errors Again

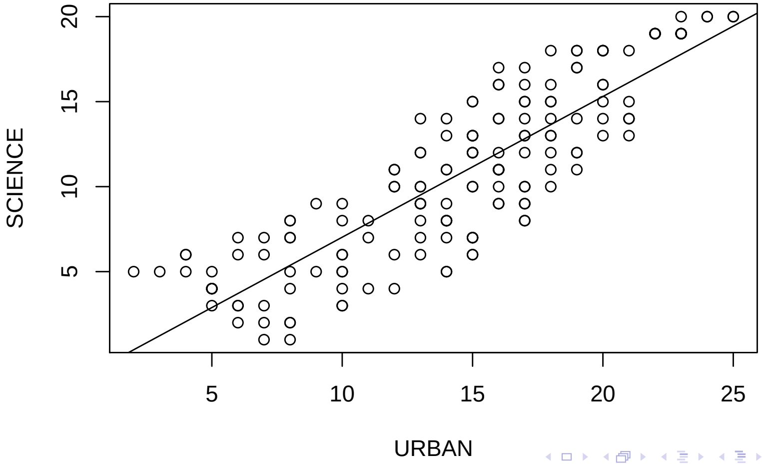


Running the Linear Model in R

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

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

Residuals:
    Min      1Q  Median      3Q     Max
-5.3358 -2.1292  0.4919  2.0432  5.0090

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.25108    0.59371  -2.107   0.0367
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 on 1 and 158 DF,  p-value: < 2.2e-16
```

Running the Linear Model in R (cont.)

```
> plot(SCIENCE ~ URBAN, sciach)
> abline(lm(SCIENCE ~ URBAN, sciach))
```



Running the Multilevel Model in 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

```
> anova(m0, m1)
   Model df      AIC      BIC    logLik   Test  L.Ratio
m0     1  3 643.8561 653.0628 -318.9281
m1     2  4 508.0940 520.3444 -250.0470 1 vs 2 137.7621
   p-value
m0
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.