Multilevel Models I

Introduction, implementation, interpretation

Michele Claibourn, StatLab

June 20/21, 2013

Research Data Services in the Library

- Research Data Services: `www.library.virginia.edu/services/`
  - Data management plans
  - GIS training and consultations
  - Locating data, archiving data
- StatLab Services: `statlab.library.virginia.edu`
  - Individual consulting: advice, training or feedback on quantitative research
  - Workshops: Tentative fall schedule is available
- Upcoming Events

Multilevel Data

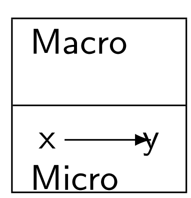
Multilevel data are meaningfully grouped into relevant larger groups or contexts, e.g., lower level units are nested within higher level units:

- Students within schools, individuals within familes
- Voters nested precincts/districts/states, or within elections
- Survey respondents within sampling strata or geographic units
- Suspects on trial within particular judges/jurisdictions
- Legislators within different legislatures
- Repeated observations, time-series cross-section data, missing data

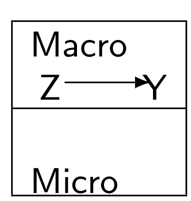
For example (multilevel1.do)

Multilevel Propositions

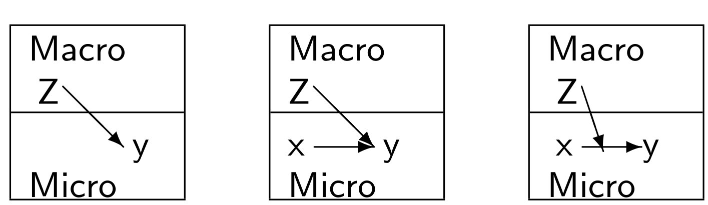
- Micro-level propositions



- Macro-level propositions



- Macro-micro propositions



But why do I need a special model?

Or less appropriate ways of dealing with multilevel data...

- Aggregating (least common)

  E.g., a regression of the means.
- Disaggregating (more common)

  E.g., estimate j separate models for j separate groups.
- Pooling (very common)

  E.g., pool all observations, treat data as independent observations all from common population.

What could possibly go wrong? (multilevel1.do)

The Intra-Class Correlation

The intra-class correlation, ρ, or degree of dependence among observations

$$\rho = \frac{\tau^2}{\tau^2 + \sigma^2}$$

where τ^2 =between-group variance and σ^2 =within-group variance.
ρ increases as variation between groups increases or variation within groups decreases.

$$\hat{\sigma}^2 = \frac{1}{N - J} \sum_{j=1}^{J} \sum_{i=1}^{N_j} (Y_{ij} - \bar{Y}_{.j})^2$$

$$\hat{\tau^2} = \frac{1}{J - 1} \sum_{j=1}^{J} (\bar{Y}_{.j} - \bar{Y}_{..})^2$$

Context, context, context

- Contextual variation can have consequences
- Multilevel models provide a way to incorporate such information, can provide new insight
- Greater Efficiency (relative to disaggregation)! Less Bias (relative to "naive pooling")!
- Are multilevel models always better? No! Requires more data, more assumptions (more later)

Level 1 Model

The multilevel model will have a level-1 model and a level-2 model (at least), which will be combined to produce the estimated model.

The level-1 model is:

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

where i indicates level-1 observations and j indicates level-2 observtions.

Let:

$$E(\beta_{0j}) = \gamma_0 \qquad Var(\beta_{0j}) = \tau_{00} = \tau_0^2$$
$$E(\beta_{1j}) = \gamma_1 \qquad Var(\beta_{1j}) = \tau_{11} = \tau_1^2$$
$$Cov(\beta_{0j}, \beta_{1j}) = \tau_{01}$$

Level 2 Model

To *explain* the variation in intercepts and slopes, we specify the level-2 model:

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01} z_j + u_{0j} \\ \beta_{1j} &= \gamma_{10} + \gamma_{11} z_j + u_{1j} \end{aligned} \tag{2}$$

where z_j represents a macro-level independent variable.

u_{0j} and u_{1j} are disturbance terms; the variance of these represents the deviation of unit j from the average intercept and slope.

Some Simplifying Assumptions

Some simplifying assumptions

$$E(\varepsilon_{ij}) = 0$$
$$Var(\varepsilon_{ij}) = \sigma^2$$
$$\varepsilon_{ij} \sim N(0, \sigma^2)$$
$$E \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$
$$Var \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} = \begin{bmatrix} \tau_0^2 & \tau_{01} \\ \tau_{10} & \tau_1^2 \end{bmatrix} = \mathbf{T}$$
$$\mathbf{U}_j \sim \mathcal{N}(\mathbf{0}, \mathbf{T})$$
$$Cov(u_{0j}, \varepsilon_{ij}) = Cov(u_{1j}, \varepsilon_{ij}) = 0$$

The Estimated Equation

The model we actually estimate is the reduced-form equation obtained by combining the level-1 and level-2 models:

$$y_{ij} = \gamma_{00} + \gamma_{01} z_j + \gamma_{10} x_{ij} + \gamma_{11} z_j x_{ij} + u_{0j} + u_{1j} x_{ij} + \varepsilon_{ij} \tag{3}$$

Things to note:

- The cross-level interaction, $z_j x_{ij}$, is more obvious in this form.
- The error term is complex: non-independent and heteroskedastic.

One-way ANOVA with Random Effects

If $\beta_{1j} = 0 \, \forall \, j$ and $\gamma_{01} = 0$, Equations (1-3) reduce to an ANOVA with random effects.

The level-1 model::

$$y_{ij} = \beta_{0j} + \varepsilon_{ij}$$

The level-2 model:

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

Reduced-form model:

$$y_{ij} = \gamma_{00} + u_{0j} + \varepsilon_{ij}$$

Let's see it in action!

- The ICC

Random Intercept Model

Adding a level-1 variable produces the level-1 model:

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

The level-2 models:

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

Reduced-form model:

$$y_{ij} = \gamma_{00} + \gamma_{10} x_{ij} + u_{0j} + \varepsilon_{ij}$$

In action!

- Inference for fixed coefficients: Hypotheses of the form

$$\begin{aligned} H_0 : &\quad \gamma_{pq} = 0 \\ H_a : &\quad \gamma_{pq} \neq 0 \end{aligned}$$

  can be tested with a standard z-test.

Random Coefficient Model

Allowing the slope to vary produces the level-1 model:

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

The level-2 models:

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

Reduced-form model:

$$y_{ij} = \gamma_{00} + \gamma_{10} x_{ij} + u_{0j} + u_{1j} x_{ij} + \varepsilon_{ij}$$

Action!

- Inference for variance components: The hypothesis is one-sided

$$\begin{aligned} H_0 &: \tau_p^2 = 0 \\ H_a &: \tau_p^2 > 0 \end{aligned}$$

  and the L-R test is preferred.

Mixed Effects Model

Adding level-2 covariates to account for the variation in intercepts and slopes produces the "full" mixed-effects model outlined in Equations (1)-(3):

The level-1 model:

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

The level-2 model:

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

The combined model:

$$y_{ij} = \gamma_{00} + \gamma_{01}z_j + \gamma_{10}x_{ij} + \gamma_{11}z_jx_{ij} + u_{0j} + u_{1j}x_{ij} + \varepsilon_{ij}$$

Action!

- Level-2 PRE: $R^2_{pj} = 1 - \frac{\hat{\tau}^2_{p(lme)}}{\hat{\tau}^2_{p(rcm)}}$

The Linear Mixed-Effects Model, Level 1

The level-1 model in scalar form, with P level-1 predictors:

$$\begin{aligned}y_{ij} &= \beta_{0j} + \beta_{1j}x_{1ij} + \beta_{2j}x_{2ij} + \cdots + \beta_{pj}x_{pij} + \varepsilon_{ij}\\ &= \beta_{0j} + \sum_{p=1}^{P}\beta_{pj}x_{pij} + \varepsilon_{ij}\end{aligned}$$

The usual assumptions that $\varepsilon_{ij} \sim N(0, \sigma^2)$.
In matrix form:

$$Y_j = X_j\boldsymbol{\beta}_j + \boldsymbol{\varepsilon}_j$$

Assume: $\varepsilon_j \sim N(\mathbf{0}, \sigma^2\mathbf{I})$.

The Linear Mixed-Effects Model, Level 2

The level-2 model in scalar form, with Q level-2 predictors:

$$\begin{aligned}
\beta_{0j} &= \gamma_{00} + \gamma_{01} z_{1j} + \gamma_{02} z_{2j} + \cdots + \gamma_{0q} x_{qj} + u_{0j} \\
&= \gamma_{00} + \sum_{q=1}^{Q} \gamma_{0q} z_{qj} + u_{0j} \\
\beta_{pj} &= \gamma_{p0} + \gamma_{p1} z_{1j} + \gamma_{p2} z_{2j} + \cdots + \gamma_{pq} x_{qj} + u_{pj} \\
&= \gamma_{p0} + \sum_{q=1}^{Q} \gamma_{pq} z_{qj} + u_{p}
\end{aligned}$$

Assumptions: $E(u_{pj}) = 0, V(u_{pj}) = \tau_{pp} = \tau_p^2, Cov(u_{pj}, u_{p'j}) = \tau_{pp'}$
In matrix form:

$$\boldsymbol{\beta}_j = \boldsymbol{Z}_j \boldsymbol{\gamma} + \boldsymbol{u}_j$$

Assume: $\mathbf{u}_j \sim N_p(0, \mathbf{T})$

The Linear Mixed-Effects Model

By substitution, the linear mixed-effects model (in scalar) is:

$$\begin{aligned}
y_{ij} &= \gamma_{00} + \sum_{q=1}^{Q} \gamma_{0q} z_{qj} + \sum_{p=1}^{P} \gamma_{p0} x_{pij} + \sum_{q=1}^{Q} \sum_{p=1}^{P} \gamma_{pq} z_{qj} x_{pij} + \\
& u_{0j} + \sum_{p=1}^{P} u_{pj} x_{pij} + \varepsilon_{ij}
\end{aligned}$$

The same model in matrix:

$$\boldsymbol{Y}_j = \boldsymbol{X}_j \boldsymbol{Z}_j \boldsymbol{\gamma} + \boldsymbol{X}_j \boldsymbol{u}_j + \boldsymbol{\varepsilon}_j$$

Multilevel Models II

Next week

- More on estimation and inference: inference for multiple parameters, variance paramaters; restricted maximum likelihood; sample sizes
- More on model assessment: exploratory analysis and model building, model assumptions, the role of centering
- Generalizing the model: three-level models, generalized linear models, longitudinal applications