Multilevel Modeling

Using R and Stata

Desmond D. Wallace and Scott J. LaCombe

Department of Political Science
The University of Iowa
Iowa City, IA

March 27, 2019

Multi-Stage Sampling

- OLS assumptions imply utilization of Simple Random Sampling (SRS)
- However, due to cost-efficiency, multi-stage sampling approaches may be utilized instead.
- Researcher may randomly sample grouping units instead of individuals (cluster sampling)
- Examples
  - Students nested in schools
  - Respondents nested in states (countries)
  - Patients nested in hospitals

Applying OLS to Multilevel Data

- Biased standard errors
- Model Misspecification
  - Not accounting for everything one should in model
  - Biased coefficient estimates

New Approach

- Best approach to analyzing nested data is a statistical approach that accounts for both within-group and between-group variation *simultaneously*
- One approach is to conceive within-group and between-group variation as random variability
- One can achieve this by including *random coefficient(s)* in the statistical model

Multilevel Model

- Multilevel Model (MLM) is a model where the parameters vary at more than one level
- Features more than one error term
- Variation can occur with respect to the intercept (*random intercept*) and/or the slope (*random slope*)
- This approach leads to corrected standard errors and correct model specification

Multilevel Model

- Model coefficients are now a combination of both fixed and random components
  - Fixed Coefficient – An unknown constant of nature
  - Random Coefficient – One which varies from sample of groups to sample of groups
- Random coefficients are not estimated, and are instead predicted

Null (Variance Components) Model

- Predicting the outcome from only an intercept that varies between groups
- The 2-level null model takes the following form:
  - Level-1 Model: $y_{ij} = \beta_{0j} + \varepsilon_{ij}$
  - Level-2 Model: $\beta_{0j} = \gamma_{00} + U_{0j}$ where
    - γ_{00} – Average (general) intercept holding across all groups (fixed effect)
    - U_{0j} – Group-specific effect on the intercept (random effect)
  - Full Specification: $y_{ij} = \gamma_{00} + U_{0j} + \varepsilon_{ij}$
- Interested in general mean value for y_{ij} (γ_{00}) and deviation between overall mean and group-specific effects for the intercept (U_{0j})

Null Model Assumptions

1. Groups are a random sample from the population of all possible groups
2. U_{0j} is randomly drawn from a population distribution with mean 0 and variance τ_0^2
3. τ_0^2 (Variance of U_{0j}) and σ^2 (variance of ε_{ij}) are uncorrelated

Example: Math Achievement

- Goal: Partition the value of students' math achievement scores into an overall mean and group-specific random effects.
- Model Specification
  - Level-1 Model (student-level): $\text{mathach}_{ij} = \beta_{0j} + \varepsilon_{ij}$
  - Level-2 Model (school-level): $\beta_{0j} = \gamma_{00} + U_{0j}$
  - Full Model: $\text{mathach}_{ij} = \gamma_{00} + U_{0j} + \varepsilon_{ij}$

Example: Math Achievement

- γ_{00}: Overall mean of student's math achievement scores
- U_{0j}: Unique effect of school j on mean math achievement score

2-Level Model Including Level-1 Covariates

- Level-1 Model: $y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + \varepsilon_{ij}$
- Level-2 Models
  - Intercept: $\beta_{0j} = \gamma_{00} + U_{0j}$ where

    γ_{00} – Average (general) intercept holding across all groups (fixed effect)

    U_{0j} – Group-specific effect on the intercept (random effect)
  - Slope: $\beta_{1j} = \gamma_{10}$ where

    γ_{10} – Amount of increase (decrease) in dependent variable for a one-unit change in x_{ij} (fixed effect)
- Full Specification: $y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + U_{0j} + \varepsilon_{ij}$

Assumptions

1. U_{0j} and ε_{ij} are mutually independent with mean 0, given the values of x_{ij}
2. U_{0j} is randomly drawn from a population distribution with mean 0 and variance τ_0^2
3. Population variance of level-1 residuals, σ^2, is constant across groups
4. U_{0j} are interpreted as group-level residuals, or group effects left unexplained by x_{ij}
5. **Unexplained variability at multiple levels is essence of multilevel modeling**

Example: Math Achievement

- Goal: Examine the influence students' socioeconomic status (SES) has on math achievement scores while controlling for students' minority and gender identification.
- Model Specification
  - Level-1 Model (student-level):
    $\text{mathach}_{ij} = \beta_{0j} + \beta_{1j}\text{SES}_{ij} + \beta_{2j}\text{minority}_{ij} + \beta_{3j}\text{female}_{ij} + \varepsilon_{ij}$
  - Level-2 Models (school-level):
    - $\beta_{0j} = \gamma_{00} + U_{0j}$
    - $\beta_{1j} = \gamma_{10}$
    - $\beta_{2j} = \gamma_{20}$
    - $\beta_{3j} = \gamma_{30}$
  - Full Model:
    $\text{mathach}_{ij} = \gamma_{00} + \gamma_{10}\text{SES}_{ij} + \gamma_{20}\text{minority}_{ij} + \gamma_{30}\text{female}_{ij} + U_{0j} + \varepsilon_{ij}$

Example: Math Achievement

- γ_{00}: Overall mean of student's math achievement scores
- γ_{10}: Effect SES has on math achievement
- γ_{20}: Difference in math achievement between minorities and non-minorities
- γ_{30}: Difference in math achievement between females and males
- U_{0j}: Unique effect of school j on mean math achievement score

Random Slopes

- Belief the relationship between independent and dependent variables differs across groups
- The 2-level model takes the following form:
  - Level-1 Model: $y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + \varepsilon_{ij}$
  - Level-2 Models
    - Intercept: $\beta_{0j} = \gamma_{00} + U_{0j}$ where

      γ_{00} – Average (general) intercept holding across all groups (fixed effect)

      U_{0j} – Group-specific effect on the intercept (random effect)
    - Slope: $\beta_{1j} = \gamma_{10} + U_{1j}$ where

      γ_{10} – Average relationship of x_{ij} and y_{ij} across groups (fixed effect)

      U_{1j} – Group-specific variation of the relationship between x_{ij} and y_{ij} (random effect)
  - Full Specification: $y_{ij} = \gamma_{00} + \gamma_{10} x_{ij} + U_{0j} + U_{1j} x_{ij} + \varepsilon_{ij}$

Assumptions

- All residuals (U_{0j}, U_{1j}, and ε_{ij}) have mean 0, given the values of the independent variable(s)
- The pair of random effects (U_{0j}, U_{1j}) are independent and identically distributed ($i.i.d$)
- (U_{0j}, U_{1j}) are independent of (ε_{ij})
- ε_{ij} is $i.i.d$

Example: Math Achievement

- Goal: Examine the influence students' socioeconomic status (SES) has on math achievement scores while controlling for students' minority and gender identification, while accounting for the effect of SES varying across schools.
- Model Specification
  - Level-1 Model (student-level):
    $\text{mathach}_{ij} = \beta_{0j} + \beta_{1j}\text{SES}_{ij} + \beta_{2j}\text{minority}_{ij} + \beta_{3j}\text{female}_{ij} + \varepsilon_{ij}$
  - Level-2 Models (school-level):
    - $\beta_{0j} = \gamma_{00} + U_{0j}$
    - $\beta_{1j} = \gamma_{10} + U_{1j}$
    - $\beta_{2j} = \gamma_{20}$
    - $\beta_{3j} = \gamma_{30}$
  - Full Model:
    $\text{mathach}_{ij} = \gamma_{00} + \gamma_{10}\text{SES}_{ij} + \gamma_{20}\text{minority}_{ij} + \gamma_{30}\text{female}_{ij} + U_{0j} + U_{1j} + \varepsilon_{ij}$

Example: Math Achievement

- γ_{00}: Overall mean of student's math achievement scores
- γ_{10}: Average effect SES has on math achievement
- γ_{20}: Difference in math achievement between minorities and non-minorities
- γ_{30}: Difference in math achievement between females and males
- U_{0j}: Unique effect of school j on mean math achievement score
- U_{1j}: Unique effect of school j on SES effect on math achievement score

Explaining Random Intercept and Random Slope Variation

- So far, coefficients have been the sum of an average and random effect.
- One could further explain this random variability via inclusion of group-level variables (Z)
- Example (Single group-level variable):
  - Random Intercept: $\beta_{0j} = \gamma_{00} + \gamma_{01} z_j + U_{0j}$
  - Random Slope: $\beta_{1j} = \gamma_{10} + \gamma_{11} z_j + U_{1j}$
- Including a group-level variable in the random intercept equation leads to a main effect of z_j
- Including a group-level variable in the random slope equation leads to an interaction effect of $z_j x_{ij}$ (Cross-level Interaction)
- Just as with level-1 variables, can feature multiple level-2 variables

Example: Math Achievement

- Goal: Examine the influence students' socioeconomic status (SES) has on math achievement scores.
- Model Specification
  - Level-1 Model (student-level):
    $\text{mathach}_{ij} = \beta_{0j} + \beta_{1j}\text{SES}_{ij} + \beta_{2j}\text{minority}_{ij} + \beta_{3j}\text{female}_{ij} + \varepsilon_{ij}$
  - Level-2 Models (school-level):
    - $\beta_{0j} = \gamma_{00} + \gamma_{01}\text{size}_j + \gamma_{02}\text{sector}_j + U_{0j}$
    - $\beta_{1j} = \gamma_{10} + \gamma_{11}\text{size}_j + U_{1j}$
    - $\beta_{2j} = \gamma_{20}$
    - $\beta_{3j} = \gamma_{30}$
  - Full Model: $\text{mathach}_{ij} = \gamma_{00} + \gamma_{01}\text{size}_j + \gamma_{02}\text{sector}_j + \gamma_{10}\text{SES}_{ij} + \gamma_{11}\text{size}_j * \text{SES}_{ij} + \gamma_{20}\text{minority}_{ij} + \gamma_{30}\text{female}_{ij} + U_{0j} + U_{1j} + \varepsilon_{ij}$

Example: Math Achievement

- γ_{00}: Overall mean of student's math achievement scores
- γ_{01}: Effect school size has on overall mean of student's math achievement scores when SES $= 0$
- γ_{02}: Difference in overall mean of student's math achievement scores for schools in sectors coded as 1 compared to schools in sectors coded as 0.
- γ_{10}: Average effect SES has on math achievement when size $= 0$
- γ_{11}: Average effect SES has on math achievement depends on school size
- γ_{20}: Difference in math achievement between minorities and non-minorities
- γ_{30}: Difference in math achievement between females and males
- U_{0j}: Unique effect of school j on mean math achievement score
- U_{1j}: Unique effect of school j on SES effect on math achievement score