Introduction to Hierarchical Linear Model

Hsueh-Sheng Wu
CFDR Workshop Series
May 20, 2013

BGSU
Center for Family and Demographic Research

Outline

- What is Hierarchical Linear Model?
- Why do nested data create analytic problems?
- Graphic presentation
- Building HLM models
- Basic HLM models
  - Unconditional Random Intercept Model
  - Random Intercept Model with a level 2 predictor
  - Random-coefficient Model with a Level 1 predictor
  - Random-coefficient Model with predictors from two different levels
  - Unconditional Growth Curve Model without predictors
  - Growth Curve Model with a level 1 predictor
- SAS codes for basic HLM models
- Stata codes for basic HLM models
- Conclusions

What Is Hierarchical Linear Model?

- A statistical technique that takes account the nested structure of the data when modeling the linear relations among parameters.
- Social scientists often deal with nested data because individuals are embedded within their environments.
- Two types of nested structures:
    - In cross-sectional data, individuals nested within groups.
    - In panel data, individual data collected at different time points are viewed as nested within individuals.
- Applications
    - Multilevel analysis
    - Growth curve analysis
    - Meta-Analysis
- Software: SAS, Stata, HLM, SPSS, R, LISREL, and Mplus

Why Do Nested Data Create Analytic Problems?

- Statistically, nested data structure violates thee assumptions of regression:
  - Independent observations
  - Independent error terms
  - Equal variances of errors for all observations

- Empirically, when one of these assumptions is violated, the estimates are biased.

Why Do Nested Data Create Analytic Problems? (Cont.)

- Hypothetical Data Example from Snijders and Bosker (1999)

Table 1. Sample data from 10 respondents

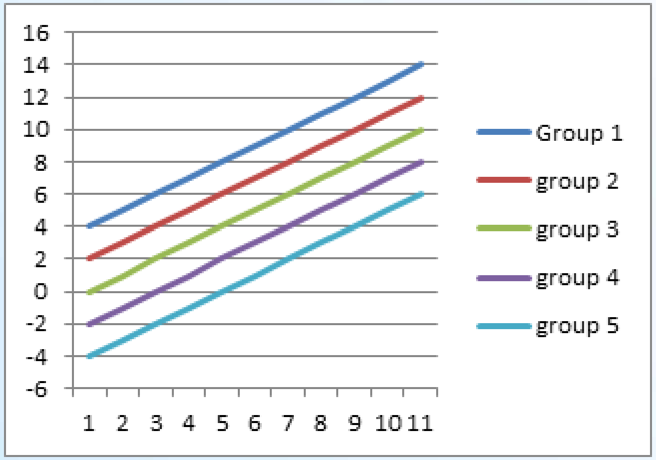
group	id	x	y
1	1	1	5
1	2	3	7
2	3	2	4
2	4	4	6
3	5	3	3
3	6	5	5
4	7	4	2
4	8	6	4
5	9	5	1
5	10	7	3

BGSU Center for Family and Demographic Research

Why Do Nested Data Create Analytic Problems? (Cont.)

- Use information from both individual and group levels, that is, we regress Y on X within each group.

Y = 4 + X	(group 1)
Y = 2 + X	(group 2)
Y = 0 + X	(group 3)
Y = -2 + X	(group 4)
Y = -4 + X	(group 5)



- The finding shows a positive association between X and Y for all five group, suggesting that X should be positively associated with Y when all of the five groups were aggregated.

Why Do Nested Data Create analytic Problems? (Cont.)

- Disaggregation analysis (analyzing the data at the individual level and ignoring the fact that respondents are from different groups)

```
. reg y x

      Source |       SS       df       MS              Number of obs =      10
-------------+------------------------------           F(  1,     8) =    1.00
       Model |  3.33333333     1  3.33333333           Prob > F      =  0.3466
    Residual |  26.6666667     8  3.33333333           R-squared     =  0.1111
-------------+------------------------------           Adj R-squared =  0.0000
       Total |          30     9  3.33333333           Root MSE      =  1.8257

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |  -.3333333   .3333333    -1.00   0.347    -1.102001    .4353347
       _cons |   5.333333   1.452966     3.67   0.006     1.982787     8.68388
------------------------------------------------------------------------------
```

- Y= 5.33 -.33X + e
- One unit increases in X leads to 0.33 unit decrease in Y.
- The result is different from what we would have expected from the previous analysis.

BGSU Center for Family and Demographic Research

Why Do Nested Data Create Analytic Problems? (Cont.)

- Aggregate analysis (focusing on group-level variables and ignore individual-level variables; that is, using the average of X to predict the average of Y at the group level):

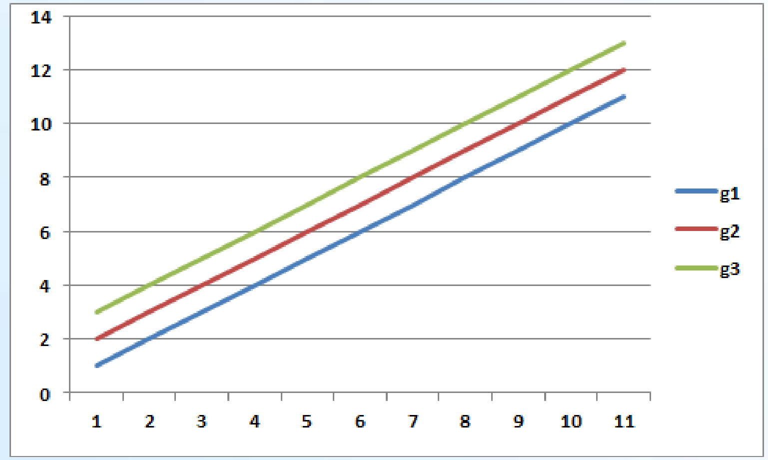
```
reg mean_y mean_x

      Source |       SS       df       MS              Number of obs =       5
-------------+------------------------------           F(  1,     3) =       .
       Model |         10     1          10            Prob > F      =       .
    Residual |          0     3           0            R-squared     =  1.0000
-------------+------------------------------           Adj R-squared =  1.0000
       Total |         10     4         2.5            Root MSE      =       0

------------------------------------------------------------------------------
      mean_y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      mean_x |         -1          .        .       .            .           .
       _cons |          8          .        .       .            .           .
------------------------------------------------------------------------------
```

- Y = 8 – X + e
- One unit increases in X leads to 1 unit decrease in Y.
- This method also reduce the power, because groups are the unit of analysis.
- In addition, the result is different from what we would have expected from the previous analyses.

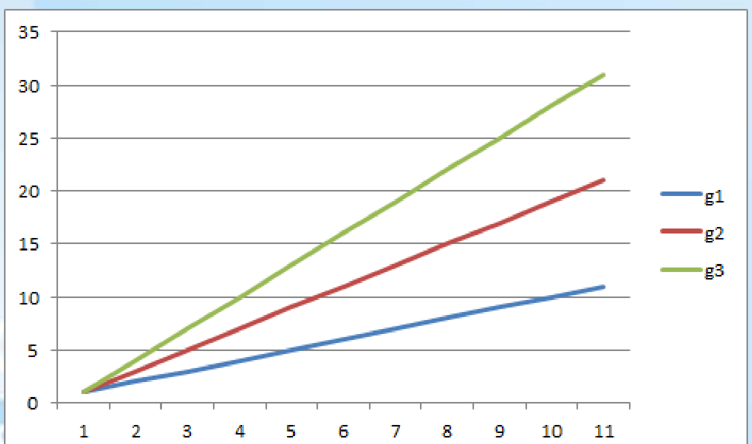
Graphic Presentation of Fixed or Random Intercepts or Slopes

A. Groups with same intercept and slope

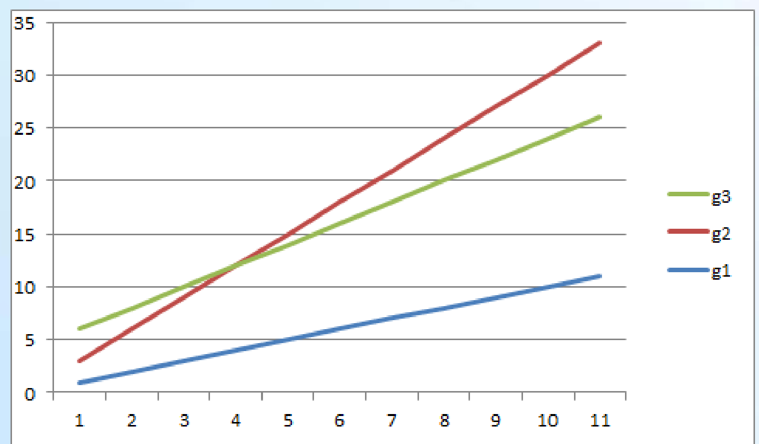
B. Groups with different intercepts and same slope



C. Groups with same intercept and different slopes



D. Groups with different intercepts and slopes



Building HLM Models

- Find a research topic
- Find a data set with variables from different levels
- Contemplate whether the intercept or slope(s) can vary at the higher level
- Always write the equations for the full model. You start with the equation for the lower level, then with the ones for the higher level, and finally combines these two sets of equations.
- Choose software to run the models

Basic HLM Models

Multi-level models:

- Unconditional Random Intercept Model
- Random Intercept Model with a level 2 predictor
- Random-coefficient Model with a Level 1 predictor
- Random-coefficient Model with predictors from two different levels

Growth Curve models:

- Unconditional Growth Curve Model without predictors
- Growth Curve Model with a level 1 predictor

Unconditional Random Intercept Model

Research Question: We would like to know whether students from different schools have different levels of mathematics achievement.

Level 1 Model:

$$MATHACH_{ij} = \beta_{0j} + r_{ij}$$

Level 2 Model:

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

Full Model:

$$MATHACH_{ij} = \gamma_{00} + u_{0j} + r_{ij}$$

Intra-class association

- Intra-class tells you whether the variance attributable to the higher-level units, relative to that from the lower-level units, is significant
- If it is significant, it justifies the use of HLM models.
- Intra-class association = $u_0 / (u_0 + r_{ij})$ $= 8.6097/(8.6097 + 39.1487) = 0.18$

BGSU Center for Family and Demographic Research

Random Intercept Model with a Level 2 Predictor

Research question is whether the SES level of schools are associated with the mathematic achievements of students.

Level 1 Model:

$$MATHACH_{ij} = \beta_{0j} + r_{ij}$$

Level 2 Model:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(MEANSES) + u_{0j}$$

Full model:

$$MATHACH_{ij} = \gamma_{00} + \gamma_{01}(MEANSES) + u_{0j} + r_{ij}$$

BGSU Center for Family and Demographic Research

Random-coefficient Model with a Level 1 Predictor

Research Questions is whether student's SES level, relative to that of other students in their schools, is associated with the mathematics achievement levels, and if so, whether such associations differ for schools.

Level 1 Model:

$$MATHACH_{ij} = \beta_{0j} + \beta_{1j}\,(SES - MEANSES) + r_{ij}$$

Level 2 Model:

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

Full Model:

$$MATHACH_{ij} = \gamma_{00} + \gamma_{10}(SES - MEANSES) + u_{0j} + u_{1j}(SES - MEANSES) + r_{ij}$$

BGSU Center for Family and Demographic Research

Random-coefficient Model with Predictors from Two Different Levels

Research Questions is whether student's SES level, relative to that of other students in their schools, is associated with the mathematics achievement levels, and if so, whether the associations differ for public schools than for private schools.

Level 1 model:

$$MATHACH_{ij} = \beta_{0j} + \beta_{1j}\,(SES - MEANSES) + r_{ij}$$

Level 2 model:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(MEANSES) + \gamma_{02}(SECTOR) + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}(MEANSES) + \gamma_{12}(SECTOR) + u_{1j}$$

Full Model:

$$MATHACH_{ij} = \gamma_{00} + \gamma_{01}(MEANSES) + \gamma_{02}(SECTOR) + \gamma_{10}\,(SES - MEANSES) + \gamma_{11}(MEANSES)^*\,(SES - MEANSES) + \gamma_{12}(SECTOR)^*\,(SES - MEANSES) + u_{0j} + u_{1j}(SES - MEANSES) + r_{ij}$$

Unconditional Growth Curve Models

Research Question is whether the change of dependent variable (Y) has a linear association with time.

Level 1 model:

$$Y_{ij}=p_{0j} + p_{1j}(\mathrm{TIME})_{ij} + r_{ij}$$

Level 2 model:

$$p_{0j}=b_{00} + u_{0j}$$
$$p_{1j}=b_{10} + u_{1j}$$

Full Model:

$$Y_{ij}=[b_{00} + b_{10}\mathrm{TIME}_{ij}] + [u_{0j} + u_{1j}\mathrm{TIME}_{ij} + r_{ij}]$$

Center for Family and Demographic Research

Growth Curve Model with a Level 2 Predictor

Research Question is whether the change of dependent variable (i.e., the change trajectory of Y) has a linear association with time and whether the individuals' characteristics will influence this trajectory.

Level 1 Model:

$$Y_{ij}=p_{0j} + p_{1j}(\text{TIME})_{ij} + r_{ij}$$

Level 2 Model:

$$p_{0j}=b_{00} + b_{01}\text{COVAR}_{j} + u_{0j}$$
$$p_{1j}=b_{10} + b_{11}\text{COVAR}_{j} + u_{1j}$$

Full Model:

$$Y_{ij}=b_{00} + b_{10}(\text{TIME})_{ij} + b_{01}(\text{COVAR})_{ij} + b_{11}(\text{COVAR})(\text{TIME})_{ij} + u_{0j} + u_{1j}(\text{TIME})_{ij} + r_{ij}$$

BGSU Center for Family and Demographic Research

SAS Codes for Multi-Level Models

- Unconditional Random Intercept Model:

```
proc mixed data = in.hsb12 covtest noclprint;
class school;
model mathach = / solution;
random intercept / subject = school;
run;
```

- Random Intercept Model with a level 2 predictor:

```
proc mixed data = in.hsb12 covtest noclprint;
class school;
model mathach = meanses / solution ddfm = bw;
random intercept / subject = school;
run;
```

SAS Codes for Multi-Level Models (cont.)

- Random-coefficient Model with a Level 1 predictor:

```
data hsbc; set in.hsb12;
cses = ses - meanses;
run;
proc mixed data = hsbc noclprint covtest noitprint;
class school;
model mathach = cses / solution ddfm = bw notest;
random intercept cses / subject = school type = un gcorr;
run;
```

- Random-coefficient Model with predictors from two different levels:

```
proc mixed data = hsbc noclprint covtest noitprint;
class school;
model mathach = meanses sector cses meanses*cses sector*cses /
   solution ddfm = bw notest;
random intercept cses / subject = school type = un;
run;
```

SAS Codes for Growth Curve Models

- Unconditional Growth Curve Model without predictors:

```
proc mixed data = in.willett noclprint covtest;
class id;
model y = time /solution ddfm = bw notest;
random intercept time / subject = id type = un;
run;
```

- Growth Curve Model with a level 1 predictor:

```
data in.willett2;
set in.willett;
wave = time; ccovar = covar - 113.4571429;
run;
proc mixed data = in.willett2 noclprint covtest;
class id;
model y = time ccovar time*ccovar /solution ddfm = bw notest;
random intercept time / subject = id type = un gcorr;
run;
```

BGSU Center for Family and Demographic Research

Stata Codes for Multi-Level Models

- Unconditional Random Intercept Model

  xtmixed mathach || school: , variance

- Random Intercept Model with a level 2 predictor

  xtmixed mathach meanses || school: , variance

- Random-coefficient Model with a Level 1 predictor

  xtmixed mathach cses || school: cses, variance cov(un)

- Random-coefficient Model with predictors from two different levels

  generate msesXcses = meanses*cses

  generate secXcses = sector*cses

  xtmixed mathach meanses sector cses msesXcses secXcses || school: cses, variance cov(un)

Stata Codes for Growth Curve Models

- Unconditional Growth Curve Model without predictors

  xtmixed y time || id: time, variance cov(un)

- Growth Curve Model with a level 1 predictor

  xtmixed y time ccovar timeBYccovar || id: time, variance cov(un)

Conclusions

- Social scientists often deal with nested data because individuals are embedded within their environments.
- Hierarchical modeling allows researchers to take into account the associations among variables from different levels.
- You can do simple HLM models with SAS, Stata, HLM, SPSS, R, LISREL, or Mplus. However, some complex models may only be analyzed with a certain software.
- If you have any questions, please send me a note.

BGSU Center for Family and Demographic Research

References

Books:

Judith Singer (1998) Using SAS PROC MIXED to Fit Multilevel Models, Hierarchical Models, and Individual Growth Models. *Journal of Educational and Behavioral Statistics,23*, 323-355.

Maimon, David and Danielle Kuhl (2008) Social Control and Youth Suicidality: Situating Durkheim's Ideas in A Multilevel Framework. *American Sociological Review,* 73, 921-943.

Multilevel Models in Family Research (2002) *Some Conceptual and Methodological Issues. Journal of Marriage and Family*, 64,280-294.

Webpages for Multilevel models:

Princeton University
http://data.princeton.edu/pop510/default.html

UCLA
http://www.ats.ucla.edu/stat/sas/topics/MLM.htmhttp://www.ats.ucla.edu/stat/stata/topics/MLM.htm

BGSU

Center for Family and Demographic Research