Mplus Short Courses
Day 5A

Multilevel Modeling With Latent Variables Using Mplus

Linda K. Muthén
Bengt Muthén

1

Table Of Contents

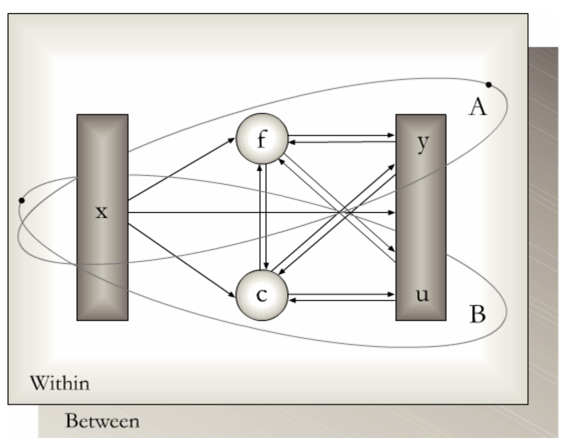
2

Mplus Background

- Inefficient dissemination of statistical methods:
  - Many good methods contributions from biostatistics, psychometrics, etc are underutilized in practice
- Fragmented presentation of methods:
  - Technical descriptions in many different journals
  - Many different pieces of limited software
- Mplus: Integration of methods in one framework
  - Easy to use: Simple, non-technical language, graphics
  - Powerful: General modeling capabilities
- Mplus versions
  - V1: November 1998
  - V2: February 2001
  - V3: March 2004
  - V4: February 2006
- Mplus team: Linda & Bengt Muthén, Thuy Nguyen, Tihomir Asparouhov, Michelle Conn

3

General Latent Variable Modeling Framework



4

Mplus

Several programs in one

- Structural equation modeling
- Item response theory analysis
- Latent class analysis
- Latent transition analysis
- Survival analysis
- Multilevel analysis
- Complex survey data analysis
- Monte Carlo simulation

Fully integrated in the general latent variable framework

5

Overview

Single-Level Analysis

	Cross-Sectional	Longitudinal
Continuous Observed And Latent Variables	***Day 1*** Regression Analysis Path Analysis Exploratory Factor Analysis Confirmatory Factor Analysis Structural Equation Modeling	***Day 2*** Growth Analysis
Adding Categorical Observed And Latent Variables	***Day 3*** Regression Analysis Path Analysis Exploratory Factor Analysis Confirmatory Factor Analysis Structural Equation Modeling Latent Class Analysis Factor Mixture Analysis Structural Equation Mixture Modeling	***Day 4*** Latent Transition Analysis Latent Class Growth Analysis Growth Analysis Growth Mixture Modeling Discrete-Time Survival Mixture Analysis Missing Data Analysis

6

Overview (Continued)

Multilevel Analysis

	Cross-Sectional	Longitudinal
Continuous Observed And Latent Variables	***Day 5*** Regression Analysis Path Analysis Exploratory Factor Analysis Confirmatory Factor Analysis Structural Equation Modeling	***Day 5*** Growth Analysis
Adding Categorical Observed And Latent Variables	***Day 5*** Latent Class Analysis Factor Mixture Analysis	***Day 5*** Growth Mixture Modeling

Analysis With Multilevel Data

Used when data have been obtained by cluster sampling and/or unequal probability sampling to avoid biases in parameter estimates, standard errors, and tests of model fit and to learn about both within- and between-cluster relationships.

Analysis Considerations

- Sampling perspective
  - Aggregated modeling – SUDAAN
    - TYPE = COMPLEX
      - Clustering, sampling weights, stratification (Asparouhov, 2005)

Analysis With Multilevel Data (Continued)

- Multilevel perspective
  - Disaggregated modeling – multilevel modeling
    - TYPE = TWOLEVEL
      - Clustering, sampling weights, stratification
  - Multivariate modeling
    - TYPE = GENERAL
      - Clustering, sampling weights, stratification
- Combined sampling and multilevel perspective
  - TYPE = COMPLEX TWOLEVEL
    - Clustering, sampling weights, stratification

9

Analysis With Multilevel Data (Continued)

Analysis Areas

- Multilevel regression analysis
- Multilevel path analysis
- Multilevel factor analysis
- Multilevel SEM
- Multilevel growth modeling
- Multilevel latent class analysis
- Multilevel latent transition analysis
- Multilevel growth mixture modeling

10

Complex Survey Data Analysis

11

Intraclass Correlation

Consider nested, random-effects ANOVA for unit i in cluster j,

$$y_{ij} = v + \eta_j + \varepsilon_{ij} \; ; \; i = 1, 2, \ldots, n_j \; ; \; j = 1, 2, \ldots, J. \qquad (44)$$

Random sample of J clusters (e.g. schools).

With timepoint as i and individual as j, this is a repeated measures model with random intercepts.

Consider the covariance and variances for cluster members $i = k$ and $i = l$,

$$Cov(y_{kj}, y_{lj}) = V(\eta), \qquad (45)$$

$$V(y_{kj}) = V(y_{lj}) = V(\eta) + V(\varepsilon), \qquad (46)$$

resulting in the intraclass correlation

$$\rho(y_{kj}, y_{lj}) = V(\eta)/[V(\eta) + V(\varepsilon)]. \qquad (47)$$

Interpretation: Between-cluster variability relative to total variation, intra-cluster homogeneity.

12

NLSY Household Clusters

Household Type (# of respondents)	# of Households*	Intraclass Correlations for Siblings: Year	Heavy Drinking
Single	5,944	1982	0.19
Two	1,985	1983	0.18
Three	634	1984	0.12
Four	170	1985	0.09
Five	32	1988	0.04
Six	5	1989	0.06

Total number of households: 8,770

Total number of respondents: 12,686

Average number of respondents per household: 1.4

*Source: NLS User's Guide, 1994, p.247

13

Design Effects

Consider cluster sampling with equal cluster sizes and the sampling variance of the mean.

V_C: correct variance under cluster sampling
V_{SRS}: variance assuming simple random sampling

$V_C \geq V_{SRS}$ but cluster sampling more convenient, less expensive.

$$DEFF = V_C / V_{SRS} = 1 + (s - 1)\,\rho, \qquad (47)$$

where s is the common cluster size and ρ is the intraclass correlation (common range: 0.00 – 0.50).

14

Random Effects ANOVA Example

200 clusters of size 10 with intraclass correlation 0.2 analyzed as:

- TYPE = TWOLEVEL
- TYPE = COMPLEX
- Regular analysis, ignoring clustering

  DEFF = 1 + 9 * 0.2 = 2.8

Input For Two-Level Random Effects ANOVA Analysis

```
TITLE:      Random effects ANOVA data
            Two-level analysis with balanced data

DATA:       FILE = anova.dat;

VARIABLE:   NAMES = y cluster;
            USEV = y;
            CLUSTER = cluster;

ANALYSIS:   TYPE = TWOLEVEL;

MODEL:
            %WITHIN%
            y;
            %BETWEEN%
            y;
```

Output Excerpts Two-Level Random Effects ANOVA Analysis

Model Results

	Estimates	S.E.	Est./S.E.
Within Level			
Variances			
Y	0.779	0.025	31.293
Between Level			
Means			
Y	0.003	0.038	0.076
Variances			
Y	0.212	0.028	7.496

17

Input For Complex Random Effects ANOVA Analysis

```
TITLE:      Random effects ANOVA data
            Complex analysis with balanced data

DATA:       FILE = anova.dat;

VARIABLE:   NAMES = y cluster;
            USEV = y;
            CLUSTER = cluster;

ANALYSIS:   TYPE = COMPLEX;
```

18

Output Excerpts Complex Random Effects ANOVA Analysis

Model Results

	Estimates	S.E.	Est./S.E.
Means			
Y	0.003	0.038	0.076
Variances			
Y	0.990	0.036	27.538

19

Input For Random Effects ANOVA Analysis Ignoring Clustering

```
TITLE:      Random effects ANOVA data
            Ignoring clustering

DATA:       FILE = anova.dat;

VARIABLE:   NAMES = y cluster;
            USEV = y;
!           CLUSTER = cluster;

ANALYSIS:   TYPE = MEANSTRUCTURE;
```

20

Output Excerpts Random Effects ANOVA Analysis Ignoring Clustering

Model Results

	Estimates	S.E.	Est./S.E.
Means			
Y	0.003	0.022	0.131
Variances			
Y	0.990	0.031	31.623

Note: The estimated mean has SE = 0.022 instead of the correct 0.038

Further Readings On Complex Survey Data

Asparouhov, T. (2005). Sampling weights in latent variable modeling. Structural Equation Modeling, 12, 411-434.

Chambers, R.L. & Skinner, C.J. (2003). Analysis of survey data. Chichester: John Wiley & Sons.

Kaplan, D. & Ferguson, A.J (1999). On the utilization of sample weights in latent variable models. Structural Equation Modeling, 6, 305-321.

Korn, E.L. & Graubard, B.I (1999). Analysis of health surveys. New York: John Wiley & Sons.

Patterson, B.H., Dayton, C.M. & Graubard, B.I. (2002). Latent class analysis of complex sample survey data: application to dietary data. Journal of the American Statistical Association, 97, 721-741.

Skinner, C.J., Holt, D. & Smith, T.M.F. (1989). Analysis of complex surveys. West Sussex, England: Wiley.

Stapleton, L. (2002). The incorporation of sample weights into multilevel structural equation models. Structural Equation Modeling, 9, 475-502.

Two-Level Regression Analysis

23

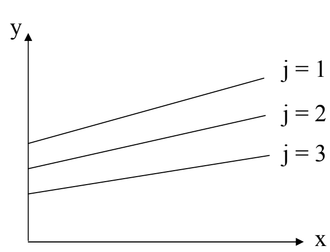
Cluster-Specific Regressions

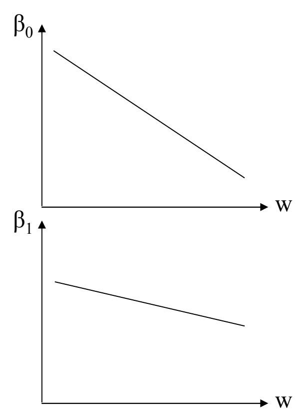
Individual *i* in cluster *j*

(1) $y_{ij} = \beta_{0j} + \beta_{1j}\, x_{ij} + r_{ij}$

(2a) $\beta_{0j} = \gamma_{00} + \gamma_{01}\, w_j + u_{0j}$

(2b) $\beta_{1j} = \gamma_{10} + \gamma_{11}\, w_j + u_{1j}$





24

Two-Level Regression Analysis With Random Intercepts And Random Slopes In Multilevel Terms

Two-level analysis (individual i in cluster j):

y_{ij} : individual-level outcome variable
x_{ij} : individual-level covariate
w_j : cluster-level covariate

Random intercepts, random slopes:

$$Level\ 1\ (Within)\ :\ y_{ij} = \beta_{0j} + \beta_{1j}\, x_{ij} + r_{ij}, \qquad (1)$$

$$Level\ 2\ (Between)\ :\ \beta_{0j} = \gamma_{00} + \gamma_{01}\, w_j + u_{0j}, \qquad (2a)$$

$$Level\ 2\ (Between)\ :\ \beta_{1j} = \gamma_{10} + \gamma_{11}\, w_j + u_{1j}. \qquad (2b)$$

- Mplus gives the same estimates as HLM/MLwiN ML (not REML):
  - $V(r)$ (residual variance for level 1)
  - $\gamma_{00}, \gamma_{01}, \gamma_{10}, \gamma_{11}, V(u_0), V(u_1), Cov(u_0, u_1)$
- Centering of x: subtracting grand mean or group (cluster) mean

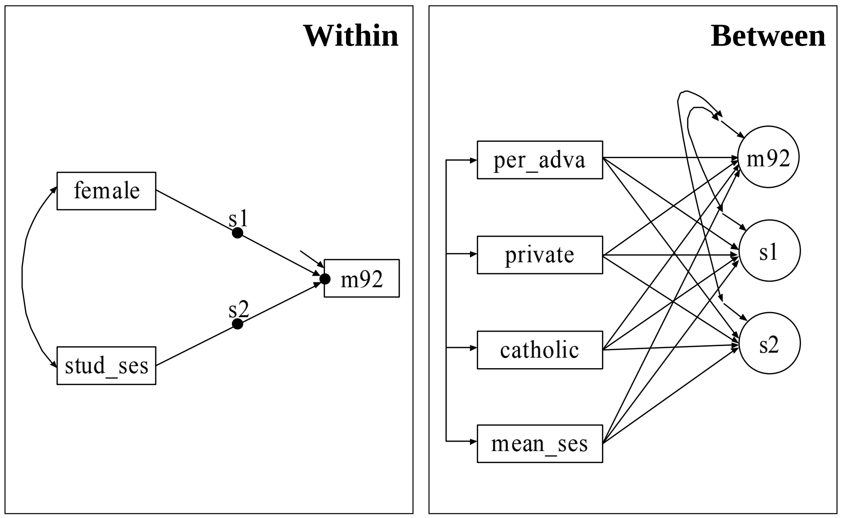
25

NELS Data

- The data—National Education Longitudinal Study (NELS:88)
  - Base year Grade 8—followed up in Grades 10 and 12
  - Students sampled within 1,035 schools—approximately 26 students per school, n = 14,217
  - Variables—reading, math, science, history-citizenship-geography, and background variables

26

NELS Math Achievement Regression



27

Input For NELS Math Achievement Regression

```
TITLE:     NELS math achievement regression

DATA:      FILE IS completev2.dat;
           ! National Education Longitudinal Study (NELS)
           FORMAT IS f8.0 12f5.2 f6.3 f11.4 23f8.2
           f18.2 f8.0 4f8.2;

VARIABLE:  NAMES ARE school r88 m88 s88 h88 r90 m90 s90 h90 r92
           m92 s92 h92 stud_ses f2pnlwt transfer minor coll_asp
           algebra retain aca_back female per_mino hw_time
           salary dis_fair clas_dis mean_col per_high unsafe
           num_frie teaqual par_invo ac_track urban size rural
           private mean_ses catholic stu_teac per_adva tea_exce
           tea_res;

           USEV = m92 female stud_ses per_adva private catholic
           mean_ses;

           !per_adva = percent teachers with an MA or higher

           WITHIN = female stud_ses;
           BETWEEN = per_adva private catholic mean_ses;
           MISSING = blank;
           CLUSTER = school;
           CENTERING = GRANDMEAN (stud_ses per_adva mean_ses);
```

28

Input For NELS Math Achievement Regression (Continued)

```
ANALYSIS: TYPE = TWOLEVEL RANDOM MISSING;

MODEL:
          %WITHIN%
          s1 | m92 ON female;
          s2 | m92 ON stud_ses;

          %BETWEEN%
          m92 s1 s2 ON per_adva private catholic mean_ses;
          m92 WITH s1 s2;
OUTPUT:   TECH8 SAMPSTAT;
```


Output Excerpts NELS Math Achievement Regression

N = 10,933

Summary of Data

Number of clusters 902

Size (s) Cluster ID with Size s

Size (s)	Cluster ID with Size s							
1	89863	75862	52654	1995	32661	89239	56214	
2	41743	81263	45025	26790	60281	82860	56241	21474
	4570	27159	11662	87842	38454			
3	65407	61407	83048	42640	41412	67708	83085	39685
	40402	93469	98582	68595	11517	17543	75498	81069
	66512							
4	31646	68153	85508	26234	83390	60835	74400	20770
	5095	10904	93569	38063	86733	66125	51670	10910
	98461	44395	95317	64112	50880	77381	12835	47555
	9208	93859	35719	67574	20048	34139	25784	80675
5	14464	74791	18219	10468	72193	97616	15773	877
	9471	83234	68254	68028	70718	3496	6842	45854

Output Excerpts NELS Math Achievement Regression (Continued)

```
     22         79570   15426   97947   93599   85125   10926    4603
     23          6411   60328   70024   67835
     24         36988   22874   50626   19091
     25         56619   59710   34292   18826   62209
     26         44586   67832   16515
     27         82887
     28           847   76909
     30         36177
     31         12786   53660   47120   94802
     32         80553
     34         53272
     36         89842   31572
     42         99516
     43         75115

Average cluster size  12.187
Estimated Intraclass Correlations for the Y Variables
          Intraclass
Variable Correlation
M92          0.107
```


Output Excerpts NELS Math Achievement Regression (Continued)

Tests of Model Fit

```
Loglikelihood
          H0 Value                     -39390.404
Information Criteria
          Number of Free parameters         21
          Akaike (AIC)                  78822.808
          Bayesian (BIC)                78976.213
          Sample-Size Adjusted BIC      78909.478
             (n* = (n + 2) / 24)
```

Model Results

```
                    Estimates     S.E.    Est./S.E.
Within Level
Residual
Variances
   M92                70.577      1.149     61.442
Between Level
S1          ON
   PER_ADVA            0.084      0.841      0.100
   PRIVATE            -0.134      0.844     -0.159
   CATHOLIC           -0.736      0.780     -0.944
   MEAN_SES           -0.232      0.428     -0.542
```

Output Excerpts NELS Math Achievement Regression (Continued)

```
S2          ON         Estimates       S.E.   Est./S.E.
   PER_ADVA                1.348      0.521       2.587
   PRIVATE                -1.890      0.706      -2.677
   CATHOLIC               -1.467      0.562      -2.612
   MEAN_SES                1.031      0.283       3.640
M92         ON
   PER_ADVA                0.195      0.727       0.268
   PRIVATE                 1.505      1.108       1.358
   CATHOLIC                0.765      0.650       1.178
   MEAN_SES                3.912      0.399       9.814
S1           WITH
   M92                    -4.456      1.007      -4.427
 S2          WITH
   M92                     0.128      0.399       0.322
Intercepts
   M92                    55.136      0.185     297.248
   S1                     -0.819      0.211      -3.876
   S2                      4.841      0.152      31.900
 Residual Variances
   M92                     8.679      1.003       8.649
   S1                      5.740      1.411       4.066
   S2                      0.307      0.527       0.583
```

33

Cross-Level Influence

Between-level (level 2) variable *w* influencing within-level (level 1) *y* variable:

Random intercept

$$y_{ij} = \beta_{0j} + \beta_1 x_{ij} + r_{ij}$$

$$\beta_{0j} = \overbrace{\gamma_{00} + \gamma_{01} w_j + u_{0j}}$$

Mplus:

```
MODEL:
        %WITHIN%;
        y ON x;    ! estimates beta1
        %BETWEEN%;
        y ON w;    ! y is the same as beta0
                   ! estimates gamma01
```

34

Cross-Level Influence (Continued)

Cross-level interaction, or between-level (level 2) variable moderating a within level (level 1) relationship:

Random slope

$$y_{ij} = \beta_0 + \beta_{1j}\, x_{ij} + r_{ij}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}\, w_j + u_{1j}$$

Mplus:

```
MODEL:

        %WITHIN%;
        beta1 | y ON x;
        %BETWEEN%;
        beta1 ON w;      ! estimates gamma11
```

35

Random Slopes

- In single-level modeling random slopes β_i describe variation across individuals i,

$$y_i = \alpha_i + \beta_i x_i + \varepsilon_i, \quad (100)$$

$$\alpha_i = \alpha + \zeta_{0i}, \quad (101)$$

$$\beta_i = \beta + \zeta_{1i}, \quad (102)$$

resulting in heteroscedastic residual variances

$$V(y_i \mid x_i) = V(\beta_i)\; x_i^2 + \theta. \quad (103)$$

- In two-level modeling random slopes β_j describe variation across clusters j

$$y_{ij} = a_j + \beta_j\, x_{ij} + \varepsilon_{ij}, \quad (104)$$

$$a_j = a + \zeta_{0j}, \quad (105)$$

$$\beta_j = \beta + \zeta_{1j}, \quad (106)$$

A small variance for a random slope typically leads to slow convergence of the ML-EM iterations. This suggests respecifying the slope as fixed.

Mplus allows random slopes for predictors that are

- Observed covariates
- Observed dependent variables
- Continuous latent variables

36

Further Readings On Multilevel Regression Analysis

Ludtke Marsh, Robitzsch, Trautwein, Asparouhov, Muthen (2007). Analysis of group level effects using multilevel modeling: Probing a latent covariate approach. Submitted for publication.

Raudenbush, S.W. & Bryk, A.S. (2002). Hierarchical linear models: Applications and data analysis methods. Second edition. Newbury Park, CA: Sage Publications.

Snijders, T. & Bosker, R. (1999). Multilevel analysis. An introduction to basic and advanced multilevel modeling. Thousand Oakes, CA: Sage Publications.

Logistic And Probit Regression