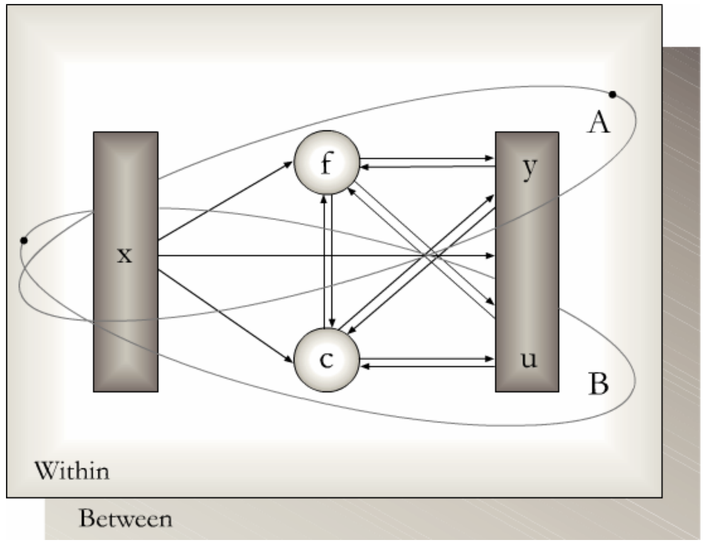
Multilevel Modeling With Latent Variables Using Mplus

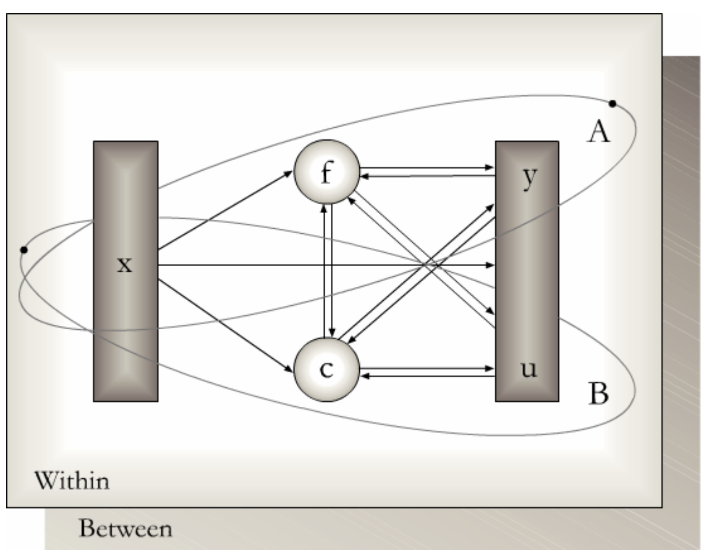
Linda K. Muthén
Bengt Muthén

Table Of Contents

General Latent Variable Modeling Framework



General Latent Variable Modeling Framework



Analysis With Multilevel Data

Used when the 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
      - Stratification, sampling weights, clustering (Asparouhov, 2005)

Analysis With Multilevel Data (Continued)

- Multilevel perspective
  - Disaggregated modeling – multilevel modeling
    - TYPE = TWOLEVEL
  - Multivariate modeling
    - TYPE = GENERAL

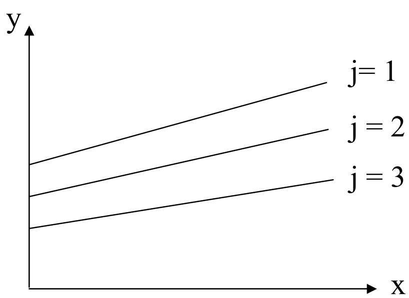
Analysis Areas

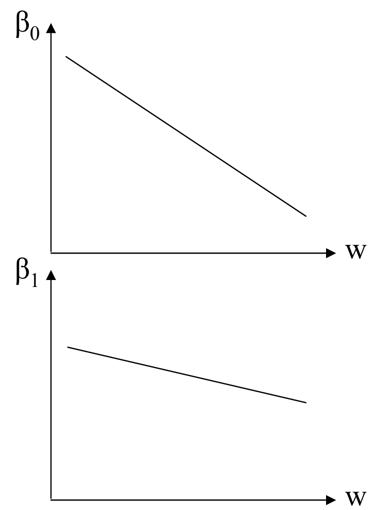
- Multilevel regression analysis
- Multilevel path analysis
- Multilevel factor analysis
- Multilevel SEM
- Multilevel latent class analysis
- Multilevel growth modeling
- Multilevel 2-part growth modeling
- Multilevel growth mixture modeling

Cluster-Specific Regressions

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

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\, w_j + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}\, w_j + u_{1j}$$





Multilevel Regression Analysis With Random Intercepts And Random Slopes In Multilevel Terms

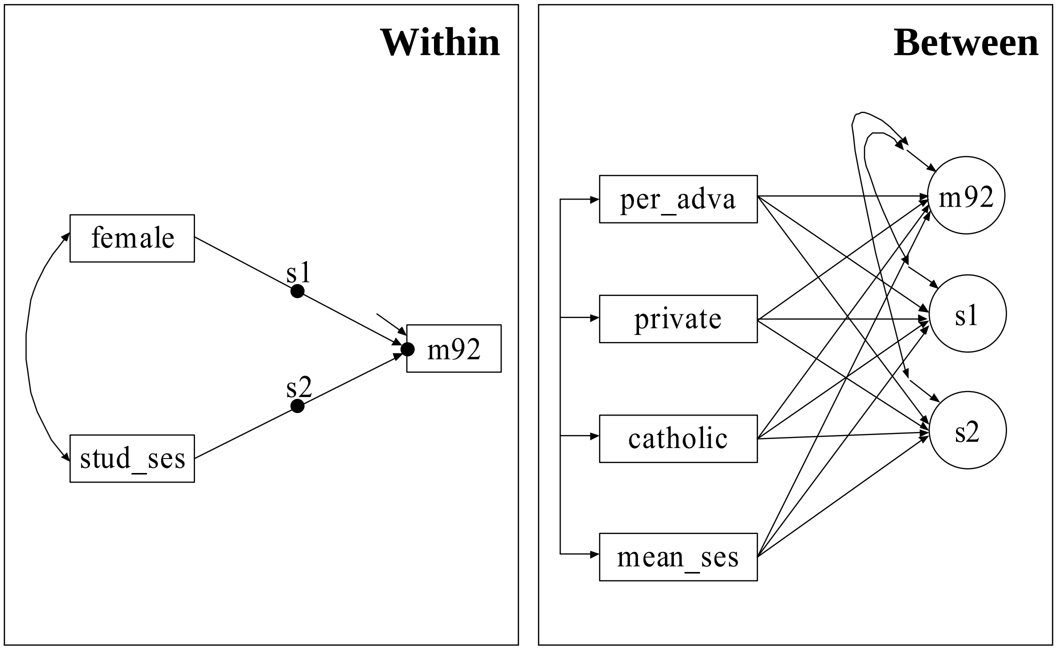
Two-level analysis (individual i in cluster j):

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

Random intercepts, random slopes:

$$\begin{aligned} &Level\ 1\ (Within) : y_{ij} = \beta_{0j} + \beta_{1j}\, x_{ij} + r_{ij}, && (8)\\ &Level\ 2\ (Between) : \beta_{0j} = \gamma_{00} + \gamma_{01}\, w_{\mathrm{j}} + u_{0\mathrm{j}}, && (9)\\ &Level\ 2\ (Between) : \beta_{1j} = \gamma_{10} + \gamma_{11}\, w_j + u_{1j}. && (10)\end{aligned}$$

- Mplus gives the same estimates as HLM/MLwiN ML (not REML): $V(r)$ (residual variance for level 1), γ_{00}, γ_{01}, γ_{10}, γ_{11}, $V(u_0)$, $V(u_1)$, $Cov(u_0, u_1)$
- Centering of x: subtracting grand mean or group (cluster) mean
- Model testing with varying covariance structure, marginal covariance matrix for y



Input For Multilevel Regression Model

```
TITLE:     multilevel 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);
```

Input For Multilevel Regression Model

```
ANALYSIS: TYPE = TWOLEVEL RANDOM MISSING;

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

          %BETWEEN%
          s1 WITH m92; s2 WITH m92;
          m92 s1 s2 ON per_adva private catholic mean_ses;

OUTPUT:   TECH8 SAMPSTAT;
```


Output Excerpts For Multilevel Regression Model (Continued)

N = 10,933

Summary of Data

```
Number of clusters     902

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   7400  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 For Multilevel Regression Model (Continued)

```
                3459      917   58687   81919   37741   63302   63143
    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 For Multilevel Regression Model (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 For Multilevel Regression Model (Continued)

```
S2         ON
   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                     54.886        0.428      128.231
   S1                      -0.856        0.507       -1.688
   S2                       4.075        0.309       13.208
 Residual Variances
   M92                      8.679        1.003        8.649
   S1                       5.740        1.411        4.066
   S2                       0.307        0.527        0.583
```


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 (Version 3)
- Continuous latent variables (Version 3)

Numerical Integration

Numerical integration is needed with maximum likelihood estimation when the posterior distribution for the latent variables does not have a closed form expression. This occurs for models with categorical outcomes that are influenced by continuous latent variables, for models with interactions involving continuous latent variables, and for certain models with random slopes such as multilevel mixture models.

When the posterior distribution does not have a closed form, it is necessary to integrate over the density of the latent variables multiplied by the conditional distribution of the outcomes given the latent variables. Numerical integration approximates this integration by using a weighted sum over a set of integration points (quadrature nodes) representing values of the latent variable.

Numerical Integration (Continued)

Numerical integration is computationally heavy and thereby time-consuming because the integration must be done at each iteration, both when computing the function value and when computing the derivative values. The computational burden increases as a function of the number of integration points, increases linearly as a function of the number of observations, and increases exponentially as a function of the dimension of integration, that is, the number of latent variables for which numerical integration is needed.

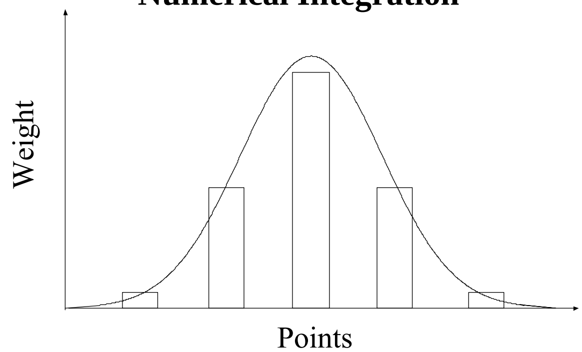
Practical Aspects Of Numerical Integration

- Types of numerical integration available in Mplus with or without adaptive quadrature
  - Standard (rectangular, trapezoid) – default with 15 integration points per dimension
  - Gauss-Hermite
  - Monte Carlo
- Computational burden for latent variables that need numerical integration
  - One or two latent variables Light
  - Three to five latent variables Heavy
  - Over five latent variables Very heavy

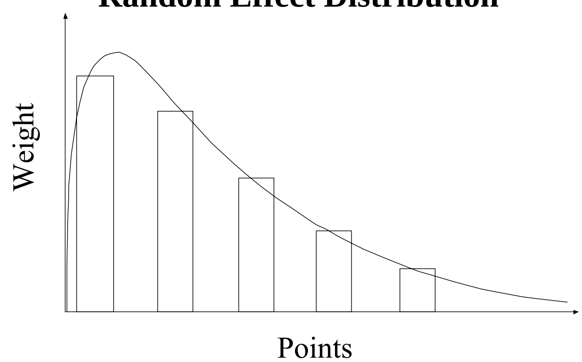
Practical Aspects Of Numerical Integration (Continued)

- Suggestions for using numerical integration
  - Start with a model with a small number of random effects and add more one at a time
  - Start with an analysis with TECH8 and MITERATIONS=1 to obtain information from the screen printing on the dimensions of integration and the time required for one iteration and with TECH1 to check model specifications
  - With more than 3 dimensions, reduce the number of integration points to 5 or 10 or use Monte Carlo integration with the default of 500 integration points
  - If the TECH8 output shows large negative values in the column labeled ABS CHANGE, increase the number of integration points to improve the precision of the numerical integration and resolve convergence problems

Numerical Integration

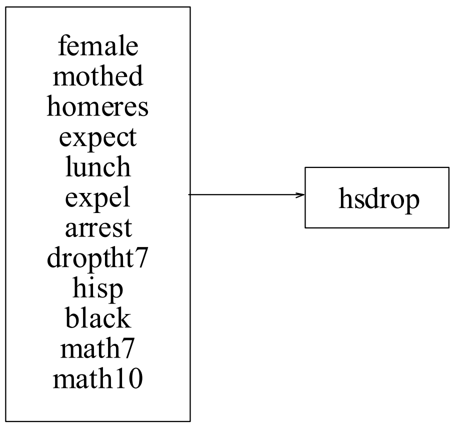


Nonparametric Estimation Of The Random Effect Distribution

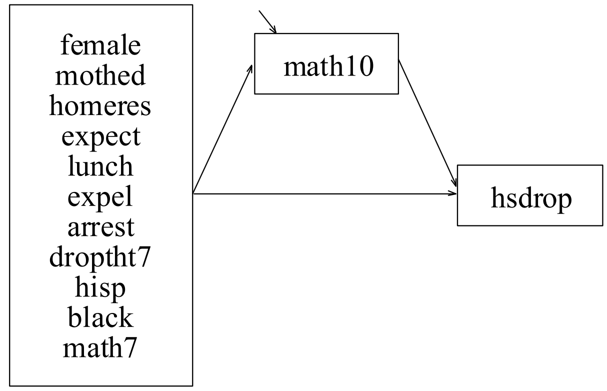


Twolevel Path Analysis With Categorical Outcomes

Logistic Regression

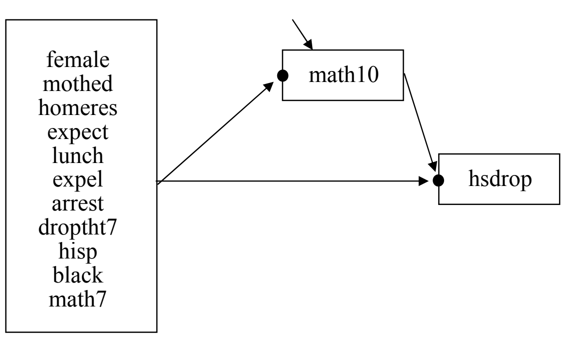


Path Model

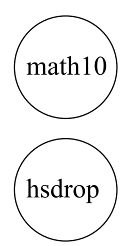


Two-Level Path Analysis

Within



Between



Input For A Twolevel Path Analysis Model With A Categorical Outcome And Missing Data On The Mediating Variable

```
TITLE:      a twolevel path analysis with a categorical outcome
            and missing data on the mediating variable
DATA:       FILE = lsayfull_dropout.dat;
VARIABLE:   NAMES = female mothed homeres math7 math10 expel
            arrest hisp black hsdrop expect lunch droptht7
            schcode;
            MISSING = ALL (999);
            CATEGORICAL = hsdrop;
            CLUSTER = schcode;
            WITHIN = female mothed homeres expect math7 lunch
            expel arrest droptht7 hisp black;
ANALYSIS:   TYPE = TWOLEVEL MISSING;
            ESTIMATOR = ML;
            ALGORITHM = INTEGRATION;
            INTEGRATION = MONTECARLO (500);
```


Input For A Twolevel Path Analysis Model With A Categorical Outcome And Missing Data On The Mediating Variable (Continued)

```
MODEL:
            %WITHIN%
            hsdrop ON female mothed homeres expect math7 math10
            lunch expel arrest droptht7 hisp black;
            math10 ON female mothed homeres expect math7 lunch
            expel arrest droptht7 hisp black;

            %BETWEEN%
            hsdrop*1; math10*1;

OUTPUT:     PATTERNS SAMPSTAT STANDARDIZED TECH1 TECH8;
```

Output Excerpts A Twolevel Path Analysis Model With A Categorical Outcome And Missing Data On the Mediating Variable

Summary Of Data

```
Number of patterns          2
Number of clusters         44

Size (s)        Cluster ID with Size s
   12              304
   13              305
   36              307     122
   38              106     112
   39              138     109
   40              103
   41              308
   42              146     120
   43              102     101
   44              303     143
   45              141
```


Output Excerpts A Twolevel Path Analysis Model With A Categorical Outcome And Missing Data On the Mediating Variable (Continued)

```
Size (s)        Cluster ID with Size s
   46              144
   47              140
   49              108
   50              126     111     110
   51              127     124
   52              137     117     147     118     301     136
   53              142     131
   55              145     123
   57              135     105
   58              121
   59              119
   73              104
   89              302
   93              309
  118              115
```

Output Excerpts A Twolevel Path Analysis Model With A Categorical Outcome And Missing Data On the Mediating Variable (Continued)

Model Results

	Estimates	S.E.	Est./S.E.	Std	StdYX
Within Level					
HSDROP ON					
FEMALE	0.323	0.171	1.887	0.323	0.077
MOTHED	-0.253	0.103	-2.457	-0.253	-0.121
HOMERES	-0.077	0.055	-1.401	-0.077	-0.061
EXPECT	-0.244	0.065	-3.756	-0.244	-0.159
MATH7	-0.011	0.015	-0.754	-0.011	-0.055
MATH10	-0.031	0.011	-2.706	-0.031	-0.197
LUNCH	0.008	0.006	1.324	0.008	0.074
EXPEL	0.947	0.225	4.201	0.947	0.121
ARREST	0.068	0.321	0.212	0.068	0.007
DROPTHT7	0.757	0.284	2.665	0.757	0.074
HISP	-0.118	0.274	-0.431	-0.118	-0.016
BLACK	-0.086	0.253	-0.340	-0.086	-0.013

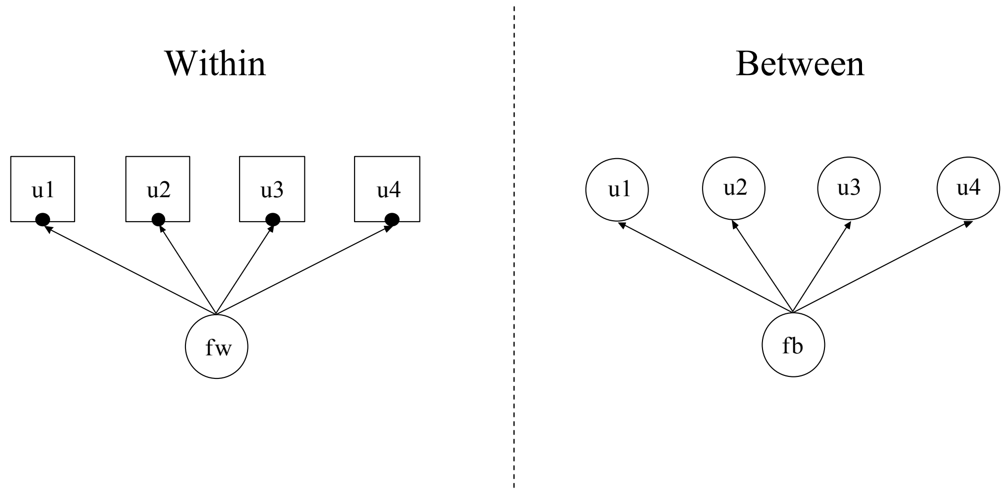
Output Excerpts A Twolevel Path Analysis Model With A Categorical Outcome And Missing Data On the Mediating Variable (Continued)

	Estimates	S.E.	Est./S.E.	Std	StdYX
MATH10 ON					
FEMALE	-0.841	0.398	-2.110	-0.841	-0.031
MOTHED	0.263	0.215	1.222	0.263	0.020
HOMERES	0.568	0.136	4.169	0.568	0.070
EXPECT	0.985	0.162	6.091	0.985	0.100
MATH7	0.940	0.023	40.123	0.940	0.697
LUNCH	-0.039	0.017	-2.308	-0.039	-0.059
EXPEL	-1.293	0.825	-1.567	-1.293	-0.026
ARREST	-3.426	1.022	-3.353	-3.426	-0.054
DROPTHT7	-1.424	1.049	-1.358	-1.424	-0.022
HISP	-0.501	0.728	-0.689	-0.501	-0.010
BLACK	-0.369	0.733	-0.503	-0.369	-0.009

Output Excerpts A Twolevel Path Analysis Model With A Categorical Outcome And Missing Data On the Mediating Variable (Continued)

```
                    Estimates       S.E.  Est./S.E.     Std     StdYX

Residual Variances
   MATH10             62.010       2.162    28.683   62.010    0.341

Between Level
 Means
   MATH10             10.226       1.340     7.632   10.226    5.276
Thresholds
   HSDROP$1           -1.076       0.560    -1.920
Variances
   HSDROP              0.286       0.133     2.150    0.286    1.000
   MATH10              3.757       1.248     3.011    3.757    1.000
```


Twolevel Factor Analysis With Categorical Outcomes



Input For A Two-Level Factor Analysis Model With Categorical Outcomes

```
TITLE:      this is an example of a two-level factor analysis
            model with categorical outcomes
DATA:       FILE = catrep1.dat;
VARIABLE:   NAMES ARE u1-u6 clus;
            CATEGORICAL = u1-u6;
            CLUSTER = clus;
ANALYSIS:   TYPE = TWOLEVEL;
            ESTIMATION = ML;
            ALGORITHM = INTEGRATION;
MODEL:
            %WITHIN%
            fw BY u1@1
            u2 (1)
            u3 (2)
            u4 (3)
            u5 (4)
            u6 (5);
```

Input For A Two-Level Factor Analysis Model With Categorical Outcomes

```
          %BETWEEN%
          fb BY u1@1
          u2 (1)
          u3 (2)
          u4 (3)
          u5 (4)
          u6 (5);
OUTPUT:   TECH1 TECH8;
```

$$\lambda f_{ij} = \lambda (f_j^B + f_{ij}^W)$$

Output Excerpts Two-Level Factor Analysis Model With Categorical Outcomes

Tests Of Model Fit

```
Loglikelihood

          HO Value                              -3696.117
Information Criteria

          Number of Free Parameters                    13
          Akaike (AIC)                           7418.235
          Bayesian (BIC)                         7481.505
          Sample-Size Adjusted BIC               7440.217
             (n* = (n + 2) / 24)
```

Output Excerpts Two-Level Factor Analysis Model With Categorical Outcomes (Continued)

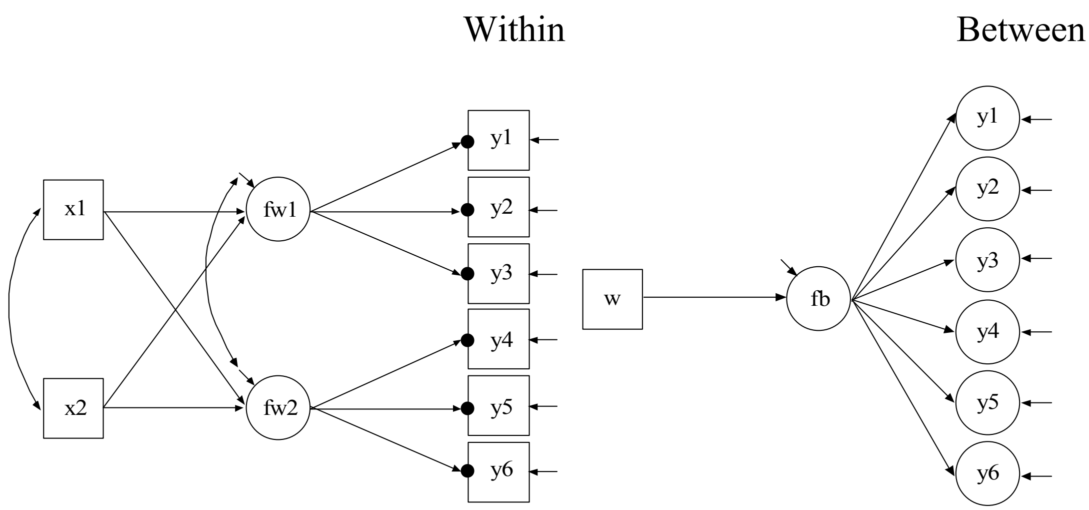
Model Results

```
                     Estimates       S.E.  Est./S.E.
Within Level
  FW        BY
     U1                  1.000      0.000      0.000
     U2                  0.915      0.146      6.264
     U3                  1.087      0.169      6.437
     U4                  1.058      0.164      6.441
     U5                  1.191      0.185      6.449
     U6                  1.143      0.178      6.439

  Variances
     FW                  0.834      0.191      4.360
```


Output Excerpts Two-Level Factor Analysis Model With Categorical Outcomes (Continued)

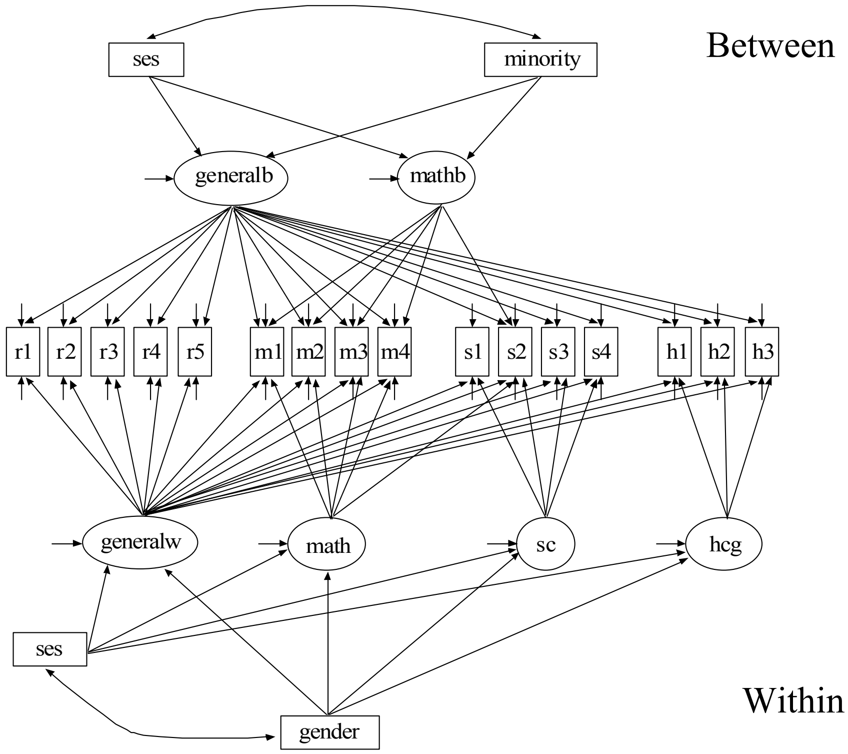
```
                     Estimates       S.E.  Est./S.E.
Between Level
  FB        BY
     U1                  1.000      0.000      0.000
     U2                  0.915      0.146      6.264
     U3                  1.087      0.169      6.437
     U4                  1.058      0.164      6.441
     U5                  1.191      0.185      6.449
     U6                  1.143      0.178      6.439

  Thresholds
     U1$1               -0.206      0.096     -2.150
     U2$1                0.001      0.091      0.007
     U3$1               -0.016      0.100     -0.156
     U4$1               -0.064      0.098     -0.652
     U5$1               -0.033      0.105     -0.315
     U6$1               -0.021      0.102     -0.209

  Variances
     FB                  0.496      0.139      3.562
```

Two-Level Factor Analysis with Covariates

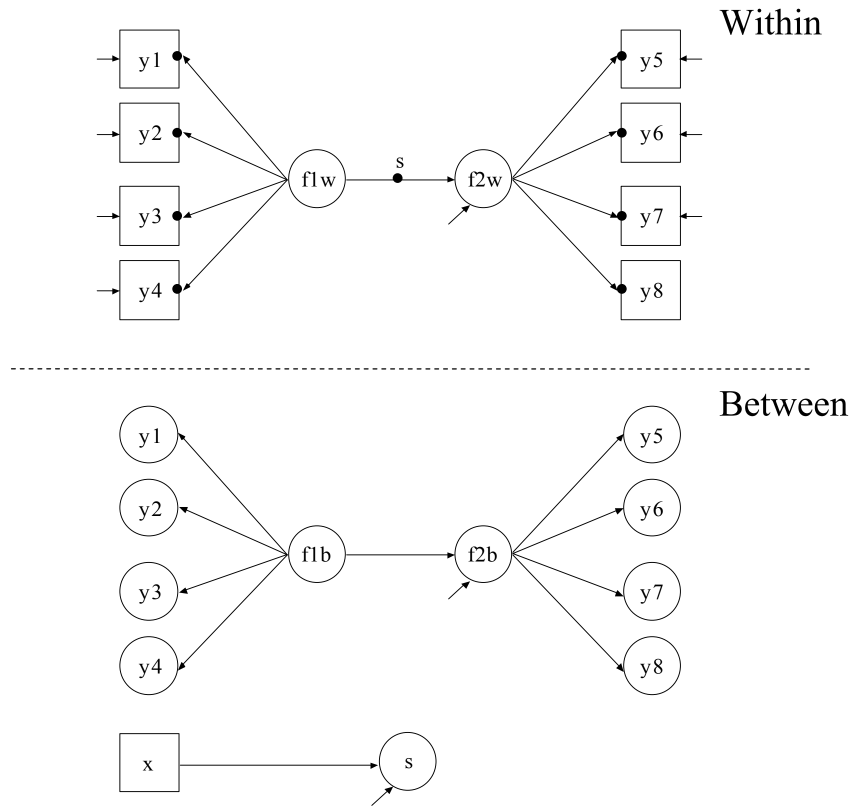


NELS Data

- The Data—National Education Longitudinal Study (NELS:88)
  - Base year Grade 8—followed up in Grades 10 and 12
  - Student sampled within 1,035 schools—approximately 26 students per school
  - Variables—reading, math, science, history-citizenship-geography, and background variables
- Data for the analysis—reading, math, science, history-citizenship-geography, gender, individual SES, school SES, and minority status



Twolevel SEM: Random Slopes For Regressions Among Factors



Input For A Twolevel SEM With A Random Slope

```
TITLE:      a twolevel SEM with a random slope

DATA:       FILE = etaeta3.dat;

VARIABLE:   NAMES ARE y1-y8 x clus;
            CLUSTER = clus;
            BETWEEN = x;

ANALYSIS:   TYPE = TWOLEVEL RANDOM MISSING;
            ALGORITHM = INTEGRATION;
```

Input For A Twolevel SEM With A Random Slope (Continued)

```
MODEL:
              %WITHIN%
              f1w BY y1@1
              y2 (1)
              y3 (2)
              y4 (3);
              f2w BY y5@1
              y6 (4)
              y7 (5)
              y8 (6);
              s | f2w ON f1w;

              %BETWEEN%
              f1b BY y1@1
              y2 (1)
              y3 (2)
              y4 (3);
              f2b BY y5@1
              y6 (4)
              y7 (5)
              y8 (6);
              f2b ON f1b;
              s ON x;
OUTPUT:       TECH1 TECH8;
```


Output Excerpts Twolevel SEM With A Random Slope

Tests Of Model Fit

```
Loglikelihood

          HO Value                             -12689.557
Information Criteria

          Number of Free Parameters                    30
          Akaike (AIC)                          25439.114
          Bayesian (BIC)                        25585.122
          Sample-Size Adjusted BIC              25489.843
            (n* = (n + 2) / 24)
```

Output Excerpts Twolevel SEM With A Random Slope (Continued)

Model Results

		Estimates	S.E.	Est./S.E.
Within Level				
F1W	BY			
Y1		1.000	0.000	0.000
Y2		0.992	0.035	28.597
Y3		0.978	0.041	23.593
Y4		1.001	0.037	26.884
F2W	BY			
Y5		1.000	0.000	0.000
Y6		0.978	0.028	34.417
Y7		1.049	0.030	35.174
Y8		1.008	0.026	38.090
F1W	WITH			
F2W		0.000	0.000	0.000

Output Excerpts Twolevel SEM With A Random Slope (Continued)

	(Estimates	S.E.	Est./S.E.)
Variances			
F1W	1.016	0.082	12.325
F2W	0.580	0.063	9.144
Residual Variances			
Y1	0.979	0.063	15.517
Y2	0.949	0.056	16.854
Y3	1.052	0.060	17.406
Y4	0.971	0.053	18.174
Y5	1.039	0.057	18.187
Y6	1.062	0.058	18.292
Y7	0.941	0.058	16.191
Y8	1.076	0.060	17.835

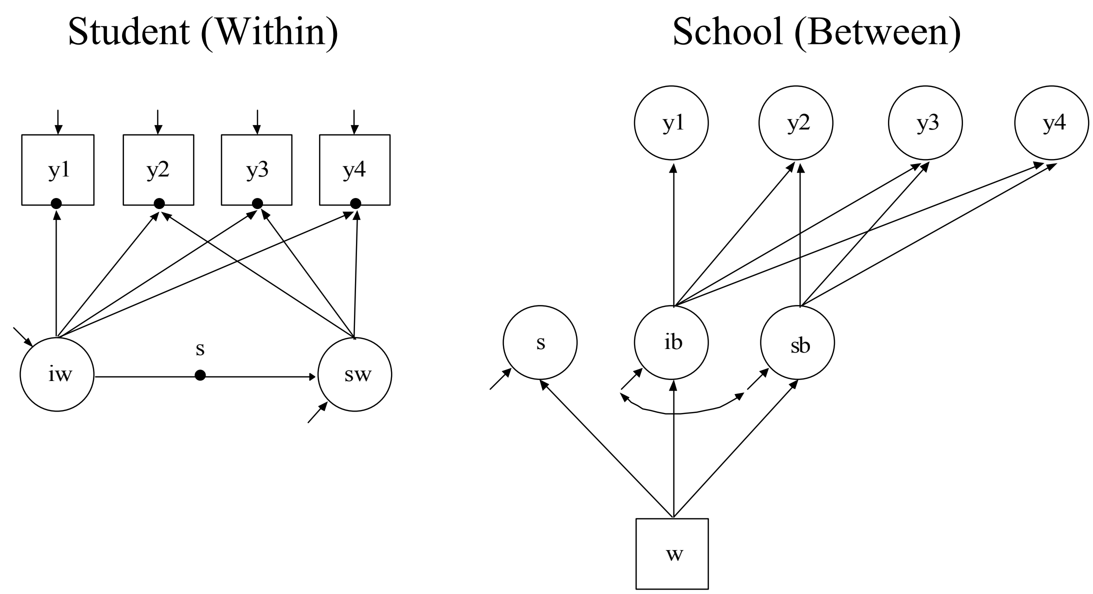
Output Excerpts Twolevel SEM With A Random Slope (Continued)

```
                  (Estimates       S.E.  Est./S.E.)
Between Level
 F1B         BY
    Y1               1.000       0.000     0.000
    Y2               0.992       0.035    28.597
    Y3               0.978       0.041    23.593
    Y4               1.001       0.037    26.884
 F2B         BY
    Y5               1.000       0.000     0.000
    Y6               0.978       0.028    34.417
    Y7               1.049       0.030    35.174
    Y8               1.008       0.026    38.090
 F2B         ON
    F1B              0.180       0.080     2.248
```


Output Excerpts Twolevel SEM With A Random Slope (Continued)

```
                  (Estimates       S.E.  Est./S.E.)
 S           ON
    X                0.999       0.082    12.150
Intercepts
    Y1              -0.099       0.063    -1.560
    Y2              -0.011       0.064    -0.175
    Y3              -0.069       0.067    -1.034
    Y4              -0.001       0.065    -0.017
    Y5               0.030       0.062     0.475
    Y6              -0.008       0.064    -0.129
    Y7               0.041       0.064     0.635
    Y8               0.002       0.071     0.035
    S                0.777       0.073    10.604
Variances
    F1B              0.568       0.096     5.900
Residual Variances
    F2B              0.237       0.056     4.211
    S                0.420       0.088     4.756
```

Multilevel Modeling With A Random Slope For Latent Variables



Multilevel Estimation, Testing, Modification, And Identification

Estimators

- Muthén's limited information estimator (MUML) – random intercepts
  - ESTIMATOR = MUML
  - Muthén's limited information estimator for unbalanced data
  - Maximum likelihood for balanced data
- Full-information maximum likelihood (FIML) – random intercepts and random slopes
  - ESTIMATOR = ML, **MLR**, MLF
  - Full-information maximum likelihood for balanced and unbalanced data
  - Robust maximum likelihood estimator
  - MAR missing data
  - Asparouhov and Muthén

Multilevel Estimation, Testing, Modification, And Identification (Continued)

Tests of Model Fit

- MUML – chi-square, robust chi-square, CFI, TLI, RMSEA, and SRMR
- FIML – chi-square, robust chi-square, CFI, TLI, RMSEA, and SRMR
- FIML with random slopes – no tests of model fit

Model Modification

- MUML – modification indices not available
- FIML – modification indices available

Model identification is the same as for CFA for both the between and within parts of the model.

Practical Issues Related To The Analysis Of Multilevel Data

Size Of The Intraclass Correlation

- Small intraclass correlations can be ignored but important information about between-level variability may be missed by conventional analysis
- The importance of the size of an intraclass correlation depends on the size of the clusters
- Intraclass correlations are attenuated by individual-level measurement error
- Effects of clustering not always seen in intraclass correlations

Practical Issues Related To The Analysis Of Multilevel Data (Continued)

Within-Level And Between-Level Variables

- Variables measured on the within-level can be used in both the between-level and within-level parts of the model or only in the within-level part of the model (WITHIN=)
- Variables measured on the between-level can be used only in the between-level part of the model (BETWEEN=)

Sample Size

- There should be at least 30-50 between-level units (clusters)
- Clusters with only one observation are allowed

Steps In SEM Multilevel Analysis For Continuous Outcomes

- Explore SEM model using the sample covariance matrix from the total sample

- Estimate the SEM model using the pooled-within sample covariance matrix

- Investigate the size of the intraclass correlations and DEFF's

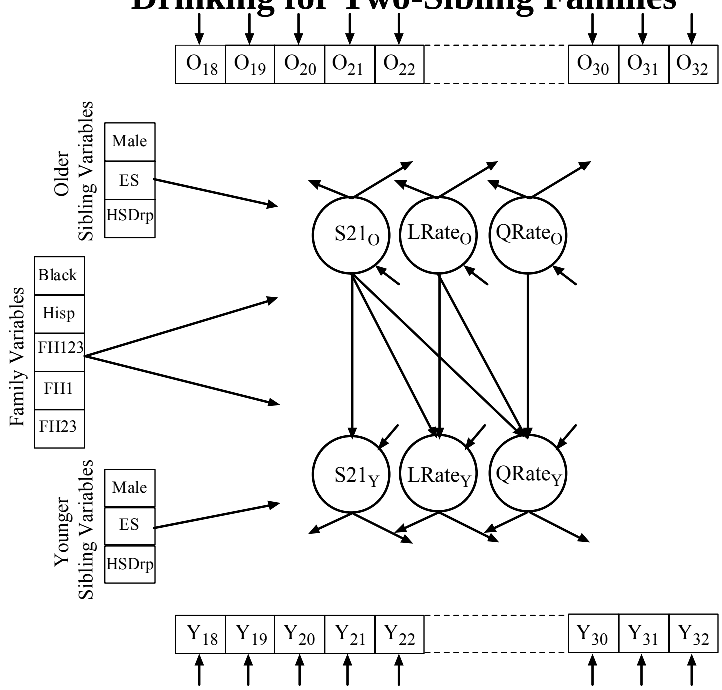
- Explore the between structure using the estimated between covariance matrix

- Estimate and modify the two-level model suggested by the previous steps

Multivariate Modeling of Family Members

- Multilevel modeling: clusters independent, model for between- and within-cluster variation, units within a cluster statistically equivalent
- Multivariate approach: clusters independent, model for all variables for each cluster unit, different parameters for different cluster units.
  - used in the latent variable growth modeling, where the cluster units are the repeated measures over time
  - allows for different cluster sizes by missing data techniques
  - more flexible than the multilevel approach, but computationally convenient only for applications with small cluster sizes (e.g. twins, spouses)

Figure 1. A Longitudinal Growth Model of Heavy Drinking for Two-Sibling Families



Source: Khoo, S.T. & Muthen, B. (2000). Longitudinal data on families: Growth modeling alternatives. Multivariate Applications in Substance Use Research, J. Rose, L. Chassin, C. Presson & J. Sherman (eds.), Hillsdale, N.J.: Erlbaum, pp. 43-78.

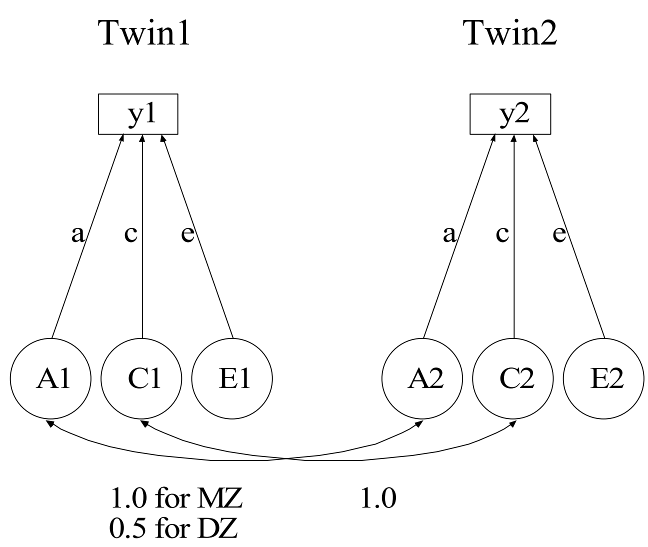
Input For Multivariate Modeling Of Family Data

```
TITLE:     Multivariate Modeling Of Family Data
           One Observation Per Family
DATA:      FILE IS multi.dat;
VARIABLE:  NAMES ARE o18-o32 y18-y32 omale oes ohsdrop ymale yoes
           yhsdrop black hisp fh123 fh1 hf123;
MODEL:     s21o BY o18-o32@1;
           lrateo BY o18@0 o19@1 o20@2 o21@3 o22@4 o23@5 o24@6
           o25@7 o26@8 o27@9 o28@10 o29@11 o30@12 o31@13 o32@14;
           qrateo BY o18@0 o19@1 o20@4 o21@9 o22@16 o23@25 o24@36
           o25@49 o26@64 o27@81 o28@100 o29@121 o30@144 o31@169
           o32@196;
           s21y BY y18-y32@1;
           lratey BY y18@0 y19@1 y20@2 y21@3 y22@4 y23@5 y24@6
           y25@7 y26@8 y27@9 y28@10 y29@11 y30212 y31@13 y32@14;
           qratey BY y18@0 y19@1 y20@4 y21@9 y22@16 y23@25 y24@36
           y25@49 y26@64 y27@81 y28@100 y29@121 y30@144 y31@169
           y32@196;
           s21o ON omale oes ohsdrop black hisp fh123 fh1 fh23;
           221y ON ymale yes yhsdrop black hisp fh123 fh1 fh23;
           s21y ON s21o;
           lratey ON s21o lrateo;
           qratey ON s21o lrateo qrateo;
           [o18-y32@0 s21o-qratey];
```


Input For Multivariate Modeling Of Family Data (Continued)

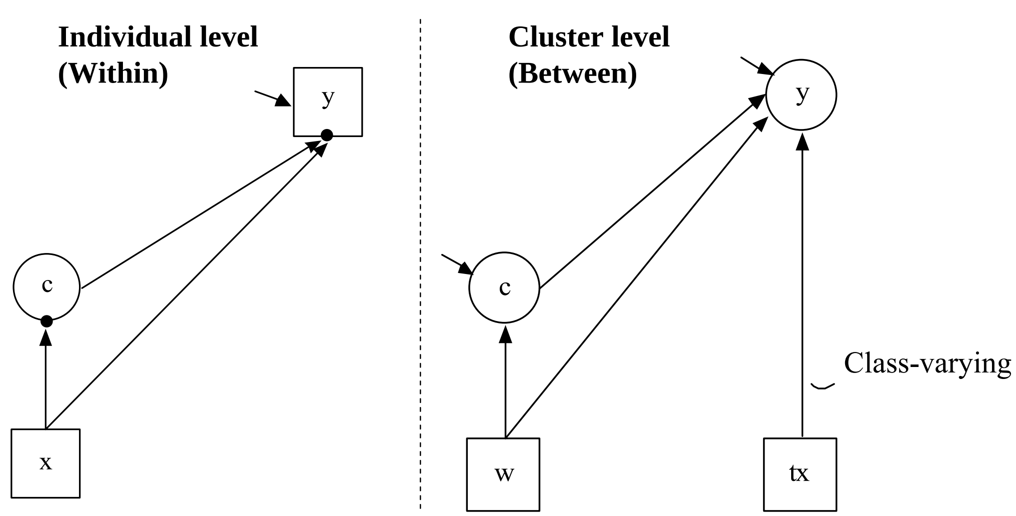
```
!New Version 3 Language For Growth Models

!MODEL     s21o lrateo qrateo | o18@0 o19@1 o20@2 o21@3 o22@4
           o23@5 o24@6 o25@7 o26@8 o27@9 o28@10 o29@11 o30@12
           o31@13 o32@14;
           s21y lratey qratey | y18@0 y19@1 y20@2 y21@3 y22@4 y23@5
           y24@6 y25@7 y26@8 y27@9 y28@10 y29@11 y30@12
           y31@13 y32@14;
           s12o ON omale oes ohsdrop black hisp fh123 fh1 fh23;
           221y ON ymale yes yhsdrop black hisp fh123 fh1 fh23;
           s21y ON s21o;
           lratey ON s21o lrateo;
           qratey ON s21o lrateo qrateo;
```

Twin Modeling

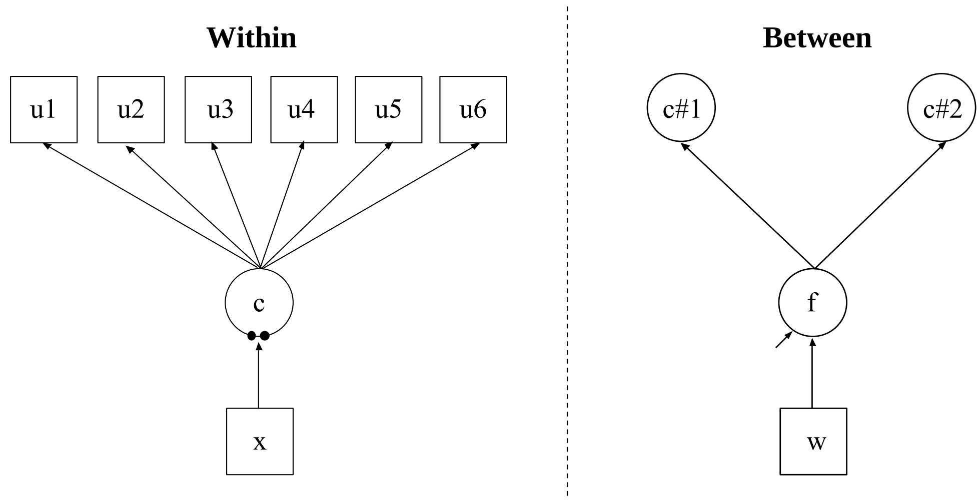


Multilevel Mixture Modeling

Two-Level Regression Mixture Modeling: Group-Randomized CACE



Two-Level Latent Class Analysis



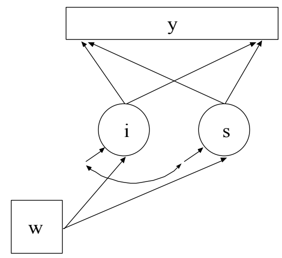
Multilevel Growth Models

Growth Modeling Approached in Two Ways: Data Arranged As Wide Versus Long

- Wide: Multivariate, Single-Level Approach

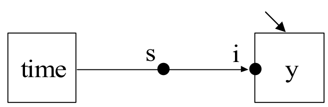
$y_{ti} = i_i + s_i \times time_{ti} + \varepsilon_{ti}$

i_i regressed on w_i
s_i regressed on w_i

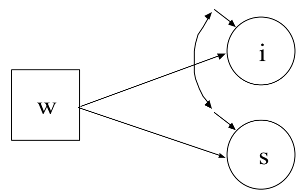


- Long: Univariate, 2-Level Approach (cluster = id)

Within



Between



Growth Modeling Approached in Two Ways: Data Arranged As Wide Versus Long (Continued)

- Wide (one person):

		t1	t2	t3	t1	t2	t3	
Person i:	id	y1	y2	y3	x1	x2	x3	w

- Long (one cluster):

Person i:	t1	id	y1	x1	w
	t2	id	y2	x2	w
	t3	id	y3	x3	w

Three-Level Modeling In Multilevel Terms

Time point t, individual i, cluster j.

y_{tij} : individual-level, outcome variable
a_{1tij} : individual-level, time-related variable (age, grade)
a_{2tij} : individual-level, time-varying covariate
x_{ij} : individual-level, time-invariant covariate
w_j : cluster-level covariate

Three-level analysis (Mplus considers Within and Between)

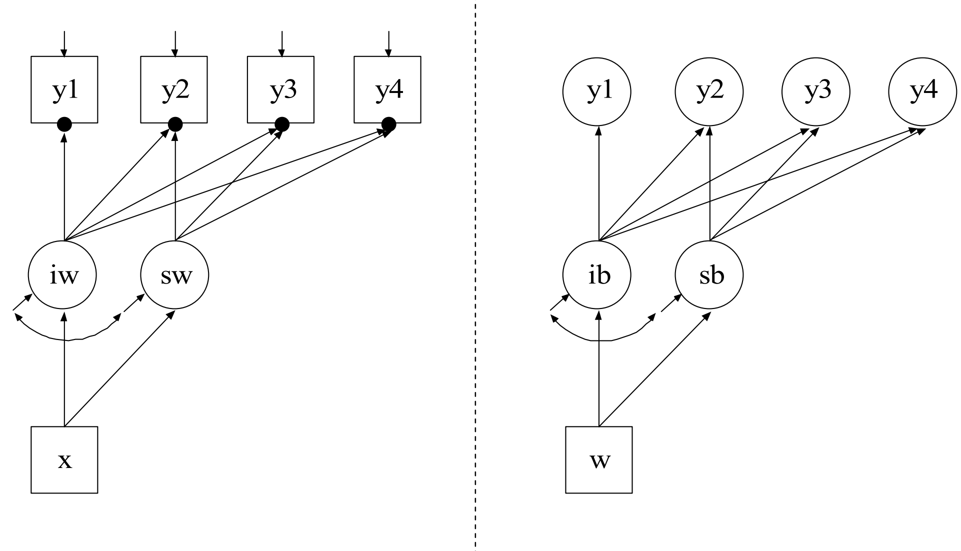
$$\text{Level 1 (Within)} : \; y_{tij} = \pi_{0ij} + \pi_{1ij}\, a_{1tij} + \pi_{2tij}\, a_{2tij} + e_{tij}, \tag{1}$$

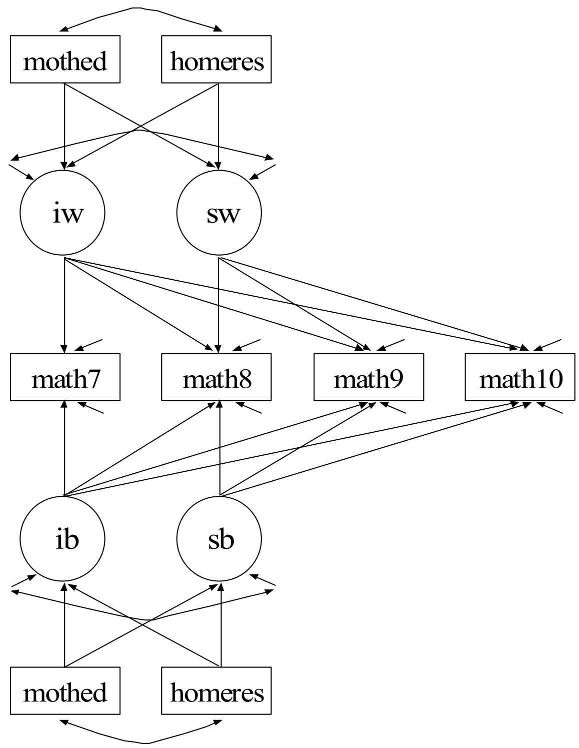
$$\text{Level 2 (Within)} : \; \begin{cases} \pi_{0ij} = \beta_{00j} + \beta_{01j}\, x_{ij} + r_{0ij}, \\ \pi_{1ij} = \beta_{10j} + \beta_{11j}\, x_{ij} + r_{1ij}, \\ \pi_{2tij} = \beta_{20tj} + \beta_{21tj}\, x_{ij} + r_{2tij}. \end{cases} \tag{2}$$

$$\text{Level 3 (Between)} : \; \begin{cases} \beta_{00j} = \gamma_{000} + \gamma_{001}\, w_j + u_{00j}, \\ \beta_{10j} = \gamma_{100} + \gamma_{101}\, w_j + u_{10j}, \\ \beta_{20tj} = \gamma_{200t} + \gamma_{201t}\, w_j + u_{20tj}, \\ \beta_{01j} = \gamma_{010} + \gamma_{011}\, w_j + u_{01j}, \\ \beta_{11j} = \gamma_{110} + \gamma_{111}\, w_j + u_{11j}, \\ \beta_{21tj} = \gamma_{2t0} + \gamma_{2t1}\, w_j + u_{2tj}. \end{cases} \tag{3}$$

Two-Level Growth Modeling (3-Level Modeling)

Within **Between**





Input For LSAY Two-Level Growth Model With Free Time Scores And Covariates

```
TITLE:     LSAY two-level growth model with free time scores
           and covariates
DATA:      FILE IS lsay98.dat;
           FORMAT IS 3f8 f8.4 8f8.2 3f8 2f8.2;
VARIABLE:  NAMES ARE cohort id school weight math7 math8 math9
           math10 att7 att8 att9 att10 gender mothed homeres;
           USEOBS = (gender EQ 1 AND cohort EQ 2);
           MISSING = ALL (999);
           USEVAR = math7-math10 mothed homeres;
           CLUSTER = school;
ANALYSIS:  TYPE = TWOLEVEL;
           ESTIMATOR = MUML;
```

Input For LSAY Two-Level Growth Model With Free Time Scores And Covariates (Continued)

```
MODEL:
            %WITHIN%
            iw BY math7-math10@1;
            sw BY math7@0 math8@1
            math9*2  (1)
            math10*3 (2);
            iw sw ON mothed homeres;
            %BETWEEN%
            ib BY math7-math10@1;
            sb BY math7@0 math8@1
            math9*2  (1)
            math10*3 (2);
            [math7-math10@0 ib sb];
            ib sb ON mothed homeres;
OUTPUT      SAMPSTAT STANDARDIZED RESIDUAL;
```


Input For LSAY Two-Level Growth Model With Free Time Scores And Covariates (Continued)

Version 3

```
!           %WITHIN%
!           iw sw | math7@0 math8@1
!           math9*2  (1)
!           math10*3 (2);
!           iw sw ON mothed homeres;
!           %BETWEEN%
!           ib sb | math7@0 math8@1
!           math9*2  (1)
!           math10*3 (2);
!           ib sb ON mothed homeres;
```

Output Excerpts LSAY Two-Level Growth Model With free Time Scores And Covariates

Summary of Data

```
Number of clusters        50

 Size (s)  Cluster ID with Size s

   1         114
   2         136
   6         132     304
   7         103
   8         102     109
   9         111     305
  14         134     118
  15         106     138     110     116
  16         105     122
  17         131     101     146     133     128
  18         141
  19         124     147     303     137     143
  20         112     129     142     307
```


Output Excerpts LSAY Two-Level Growth Model With free Time Scores And Covariates (Continued)

```
 Size (s)  Cluster ID with Size s

  21         120     145
  22         144     127
  23         140     121     139     308
  24         119
  25         123
  26         301     117
  27         108
  29         135
  33         115
  34         104
  39         309
  40         302
Average cluster size 18.627
Estimated Intraclass Correlations for the Y Variables
```

Variable	Intraclass Correlation	Variable	Intraclass Correlation	Variable	Intraclass Correlation
MATH7	0.199	MATH8	0.149	MATH9	0.168
MATH10	0.165				

Output Excerpts LSAY Two-Level Growth Model With free Time Scores And Covariates (Continued)

Tests Of Model Fit

```
Chi-square Test of Model Fit
            Value                       24.058*
            Degrees of Freedom              14
            P-Value                     0.0451
CFI / TLI
            CFI                          0.997
            TLI                          0.995
RMSEA (Root Mean Square Error Of Approximation)
            Estimate                     0.028
SRMR (Standardized Root Mean Square Residual)
            Value for Between            0.048
            Value for Within             0.007
```


Output Excerpts LSAY Two-Level Growth Model With free Time Scores And Covariates (Continued)

Model Results

Within Level

SW BY					
MATH8	1.000	0.000	0.000	1.073	0.128
MATH9	2.487	0.163	15.220	2.670	0.288
MATH10	3.589	0.223	16.076	3.853	0.368
IW ON					
MOTHED	**1.780**	**0.232**	**7.665**	**0.246**	**0.226**
HOMERES	**0.892**	**0.221**	**4.031**	**0.124**	**0.173**
SW ON					
MOTHED	**0.053**	**0.063**	**0.836**	**0.049**	**0.045**
HOMERES	**0.135**	**0.044**	**3.047**	**0.125**	**0.176**
SW WITH					
IW	2.112	0.522	4.044	0.273	0.273

Output Excerpts LSAY Two-Level Growth Model With free Time Scores And Covariates (Continued)

```
HOMERES  WITH
   MOTHED               0.261    0.039    6.709    0.261    0.203
Residual Variances
   MATH7               12.748    1.434    8.888   12.748    0.197
   MATH8               12.298    0.893   13.771   12.298    0.174
   MATH9               14.237    1.132   12.578   14.237    0.166
   MATH10              24.829    2.230   11.133   24.829    0.226
   IW                  47.060    3.069   15.333    0.903    0.903
   SW                   1.110    0.286    3.879    0.964    0.964
Variances
   MOTHED               0.841    0.049   17.217    0.841    1.000
   HOMERES              1.970    0.069   28.643    1.970    1.000
```


Output Excerpts LSAY Two-Level Growth Model With free Time Scores And Covariates (Continued)

```
                  Estimates    S.E.    Est./S.E. Std      StdYX
```

Between Level

```
SB        BY
   MATH8                1.000    0.000    0.000    0.196    0.052
   MATH9                2.487    0.163   15.220    0.488    0.119
   MATH10               3.589    0.223   16.076    0.704    0.115
IB        ON
   MOTHED              -1.225    2.587   -0.474   -0.362   -0.107
   HOMERES              7.160    1.847    3.876    2.117    1.011
SB        ON
   MOTHED               0.995    0.647    1.538    5.073    1.493
   HOMERES              0.017    0.373    0.045    0.086    0.041
SB        WITH
   IB                   0.382    0.248    1.538    0.575    0.575
```

Output Excerpts LSAY Two-Level Growth Model With free Time Scores And Covariates (Continued)

HOMERES WITH					
MOTHED	0.103	0.019	5.488	0.103	0.733
Residual Variances					
MATH7	2.059	0.552	3.732	2.059	0.153
MATH8	0.544	0.268	2.033	0.544	0.039
MATH9	0.105	0.213	0.493	0.105	0.006
MATH10	1.395	0.504	2.767	1.395	0.067
IB	1.428	1.690	0.845	0.125	0.125
SB	-0.051	0.071	-0.713	-1.321	-1.321
Variances					
MOTHED	0.087	0.023	3.801	0.087	1.000
HOMERES	0.228	0.056	4.066	0.228	1.000
Means					
MOTHED	2.307	0.043	53.277	2.307	7.838
HOMERES	3.108	0.062	50.375	3.108	6.509
Intercepts					
IB	33.510	2.678	12.512	9.909	9.909
SB	0.163	0.776	0.210	0.830	0.830

Output Excerpts LSAY Two-Level Growth Model With free Time Scores And Covariates (Continued)

R-Square

Within Level

Observed Variable	R-Square
MATH7	0.803
MATH8	0.826
MATH9	0.834
MATH10	0.774

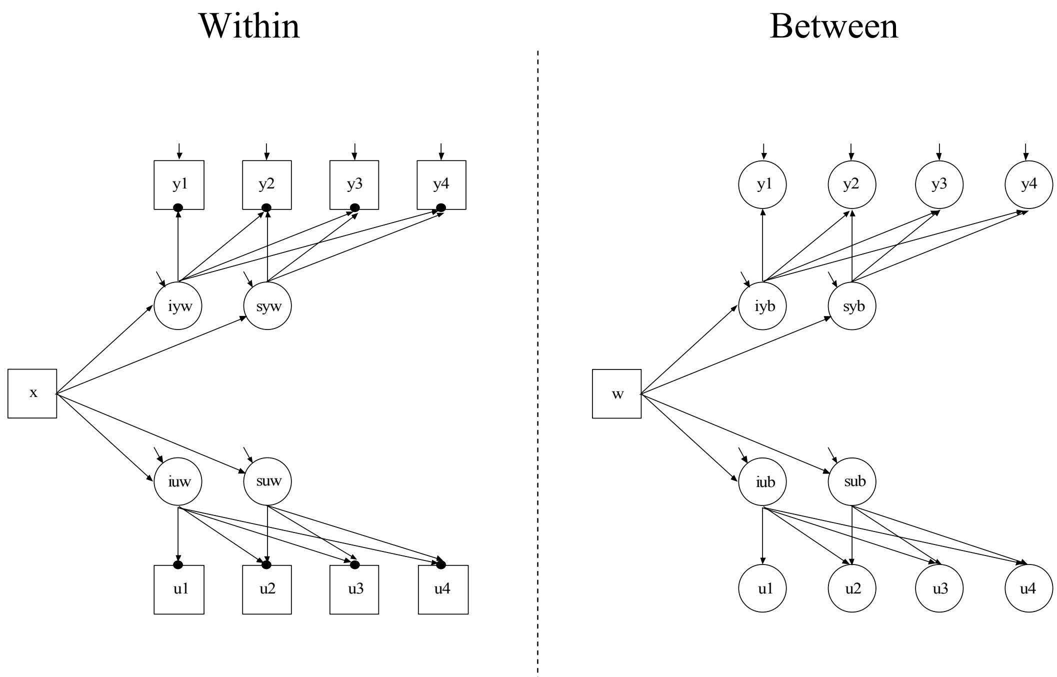
Latent Variable	R-Square
IW	0.097
SW	0.036

Output Excerpts LSAY Two-Level Growth Model With free Time Scores And Covariates (Continued)

R-Square

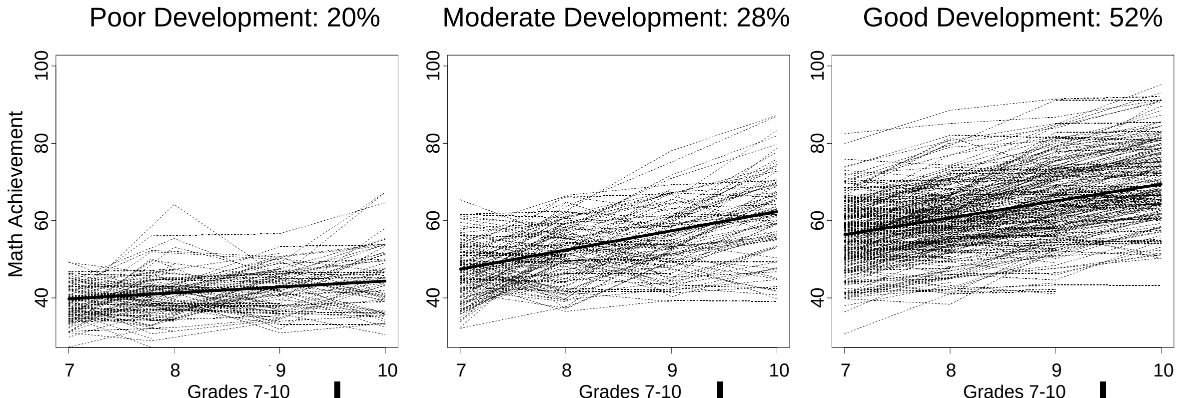
```
Between Level

    Observed
    Variable       R-Square

    MATH7          0.847
    MATH8          0.961
    MATH9          0.994
    MATH10         0.933

    Latent
    Variable       R-Square

    IW             0.875
    SW             Undefined    0.23207E+01
```


Two-Level, Two-Part Growth Modeling



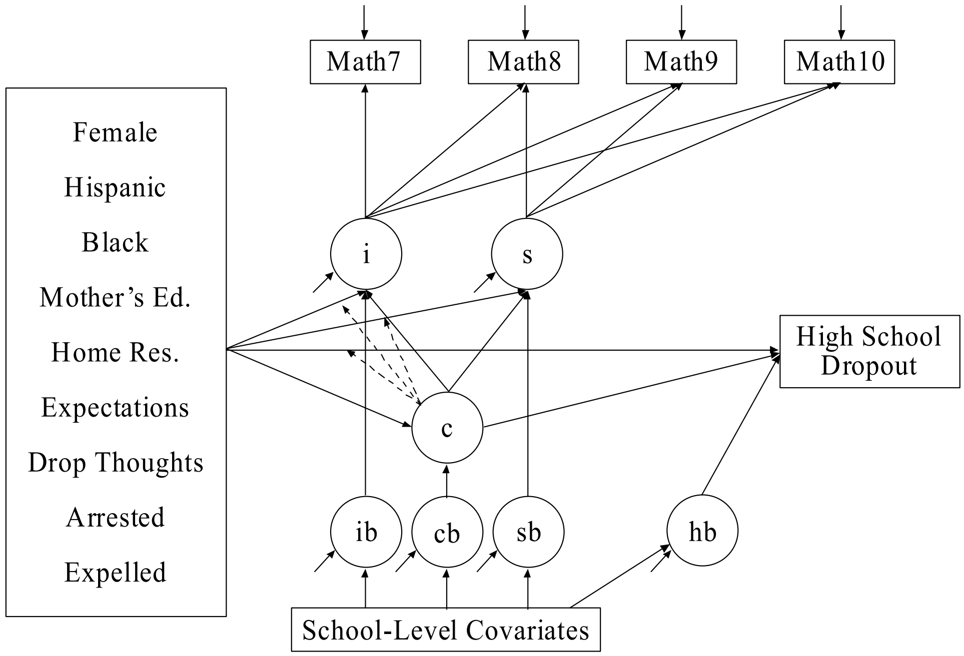
Multilevel Growth Mixture Modeling

Growth Mixture Modeling: LSAY Math Achievement Trajectory Classes And The Prediction Of High School Dropout



Dropout: 69% 8% 1%

Multilevel Growth Mixture Modeling



Input For A Multilevel Growth Mixture Model For LSAY Math Achievement

```
TITLE:      multilevel growth mixture model for LSAY math
            achievement

DATA:       FILE = lsayfull_Dropout.dat;

VARIABLE:   NAMES = female mothed homeres math7 math8 math9 math10
            expel arrest hisp black hsdrop expect lunch mstrat
            droptht7;
            !lunch = % of students eligible for full lunch
            !assistance (9th)
            !mstrat = ratio of students to full time math
            !teachers (9th)
            MISSING = ALL (9999);
            CATEGORICAL = hsdrop;
            CLASSES = c (3);
            CLUSTER = schcode;
            WITHIN = female mothed homeres expect droptht7 expel
            arrest hisp black;
            BETWEEN = lunch mstrat;
```

Input For A Multilevel Growth Mixture Model For LSAY Math Achievement (Continued)

```
DEFINE:     lunch = lunch/100;
            mstrat = mstrat/1000;

ANALYSIS:   TYPE = MIXTURE TWOLEVEL MISSING;
            ALGORITHM = INTEGRATION;

OUTPUT:     SAMPSTAT STANDARDIZED TECH1 TECH8;

PLOT:       TYPE = PLOT3;
            SERIES = math7-math10 (s);
```


Input For A Multilevel Growth Mixture Model For LSAY Math Achievement (Continued)

```
MODEL:
            %WITHIN%
            %OVERALL%
            i s | math7@0 math8@1 math9@2 math10@3;
            i s c#1 c#2 hsdrop ON female hisp black mothed homeres
            expect droptht7 expel arrest;

            %c#1%
            [i*40 s*1];
            math7-math10*20;
            i*13 s*3;

            %c#2%
            [i*40 s*5];
            math7-math10*30;
            i*8 s*3;
            i s ON female hisp black mothed homeres expect
            droptht7 expel arrest;
```

Input For A Multilevel Growth Mixture Model For LSAY Math Achievement (Continued)

```
%c#3%
[i*45 s*3];
math7-math10*10;
i*34 s*2;
i s ON female hisp black mothed homeres expect
dropththt7 expel arrest;

%BETWEEN%
%OVERALL%
ib | math7-math10@1; [ib@0];
ib*1; hsdrop*1; ib WITH hsdrop;
math7-math10@0;
ib c#1 c#2 hsdrop ON lunch mstrat;

%c#1%
[hsdrop$1*-.3];

%c#2%
[hsdrop$1*.9];

%c#3%
[hsdrop$1*1.2];
```


Output Excerpts A Multilevel Growth Mixture Model For LSAY Math Achievement

Summary of Data

```
Number of patterns          13
Number of y patterns        13
Number of u patterns         1
Number of clusters          44

Size (s)     Cluster ID with Size s

   12           304
   13           305
   38           112
   39           109
   40           138
   42           120
   43           307
   44           303
   45           143        146
```

Output Excerpts A Multilevel Growth Mixture Model For LSAY Math Achievement (Continued)

46	101					
48	144	106				
51	102	308				
52	136	118	133	111		
53	140	142	108	131	122	124
54	301	117	127	137	126	
55	103	141	123			
56	110					
57	147					
58	121	105	145	135		
59	119					
73	104					
89	302					
94	309					
118	115					

Output Excerpts A Multilevel Growth Mixture Model For LSAY Math Achievement (Continued)

```
MAXIMUM LOG-LIKELIHOOD VALUE FOR THE UNRESTRICTED (H1) MODEL IS
-36393.088

THE STANDARD ERRORS OF THE MODEL PARAMETER ESTIMATES MAY NOT BE
TRUSTWORTHY FOR SOME PARAMETERS DUE TO A NON-POSITIVE DEFINITE
FIRST-ORDER DERIVATIVE PRODUCT MATRIX. THIS MAY BE DUE TO THE
STARTING VALUES BUT MAY ALSO BE AN INDICATION OF MODEL
NONIDENTIFICATION. THE CONDITION NUMBER IS -0.758D-16. PROBLEM
INVOLVING PARAMETER 54.

THE NONIDENTIFICATION IS MOST LIKELY DUE TO HAVING MORE
PARAMETERS THAN THE NUMBER OF CLUSTERS. REDUCE THE NUMBER OF
PARAMETERS.

THE MODEL ESTIMATION TERMINATED NORMALLY
```

Output Excerpts A Multilevel Growth Mixture Model For LSAY Math Achievement (Continued)

Tests Of Model Fit

```
Loglikelihood

               HO Value                           -26247.205
Information Criteria

               Number of Free Parameters                 122
               Akaike (AIC)                        52738.409
               Bayesian (BIC)                      53441.082
               Sample-Size Adjusted BIC            53053.464
                  (n* = (n + 2) / 24)
               Entropy                                 0.632

FINAL  CLASS COUNTS AND PROPORTIONS OF TOTAL SAMPLE SIZE BASED
ON ESTIMATED POSTERIOR PROBABILITIES

 Class 1        686.43905          0.29285
 Class 2        430.83877          0.18380
 Class 3       1226.72218          0.52335
```


Output Excerpts A Multilevel Growth Mixture Model For LSAY Math Achievement (Continued)

Model Results

```
                         Estimates     S.E.  Est./S.E.    Std      StdYX


 Between Level
 CLASS 1
  IB         ON
     LUNCH              -1.805      1.310    -1.378   -0.851   -0.176
     MSTRAT            -13.365      3.086    -4.331   -6.299   -0.448

  HSDROP     ON
     LUNCH               1.087      0.543     2.004    1.087    0.290
     MSTRAT             -0.178      1.478    -0.120   -0.178   -0.016

  IB         WITH
     HSDROP             -0.416      0.328    -1.267   -0.196   -0.253
```

Output Excerpts A Multilevel Growth Mixture Model For LSAY Math Achievement (Continued)

```
Intercepts
    MATH7           0.000     0.000     0.000     0.000     0.000
    MATH8           0.000     0.000     0.000     0.000     0.000
    MATH9           0.000     0.000     0.000     0.000     0.000
    MATH10          0.000     0.000     0.000     0.000     0.000
    IB              0.000     0.000     0.000     0.000     0.000

Residual Variances
    HSDROP          0.550     0.216     2.542     0.550     0.915
    MATH7           0.000     0.000     0.000     0.000     0.000
    MATH8           0.000     0.000     0.000     0.000     0.000
    MATH9           0.000     0.000     0.000     0.000     0.000
    MATH10          0.000     0.000     0.000     0.000     0.000
    IB              3.456     1.010     3.422     0.768     0.768
```


Output Excerpts A Multilevel Growth Mixture Model For LSAY Math Achievement (Continued)

Model Results

```
LATENT CLASS REGRESSION MODEL PART
                  Estimates    S.E.   Est./S.E.

Within Level

 C#1       ON
    FEMALE         -0.751     0.188    -3.998
    HISP            0.094     0.705     0.133
    BLACK           0.900     0.385     2.339
    MOTHED         -0.003     0.106    -0.028
    HOMERES        -0.060     0.069     0.864
    EXPECT         -0.251     0.074    -3.406
    DROPTHT7        1.616     0.451     3.583
    EXPEL           0.698     0.337     2.068
    ARREST          1.093     0.384     2.842
```

Output Excerpts A Multilevel Growth Mixture Model For LSAY Math Achievement (Continued)

	(Estimates	S.E.	Est./S.E.)
C#2 ON			
FEMALE	-1.610	0.450	-3.577
HISP	1.144	0.466	2.458
BLACK	-0.961	0.656	-1.465
MOTHED	-0.234	0.139	-1.684
HOMERES	0.102	0.094	1.085
EXPECT	0.056	0.089	0.628
DROPTHT7	0.570	0.657	0.869
EXPEL	1.217	0.397	3.068
ARREST	1.133	0.580	1.951

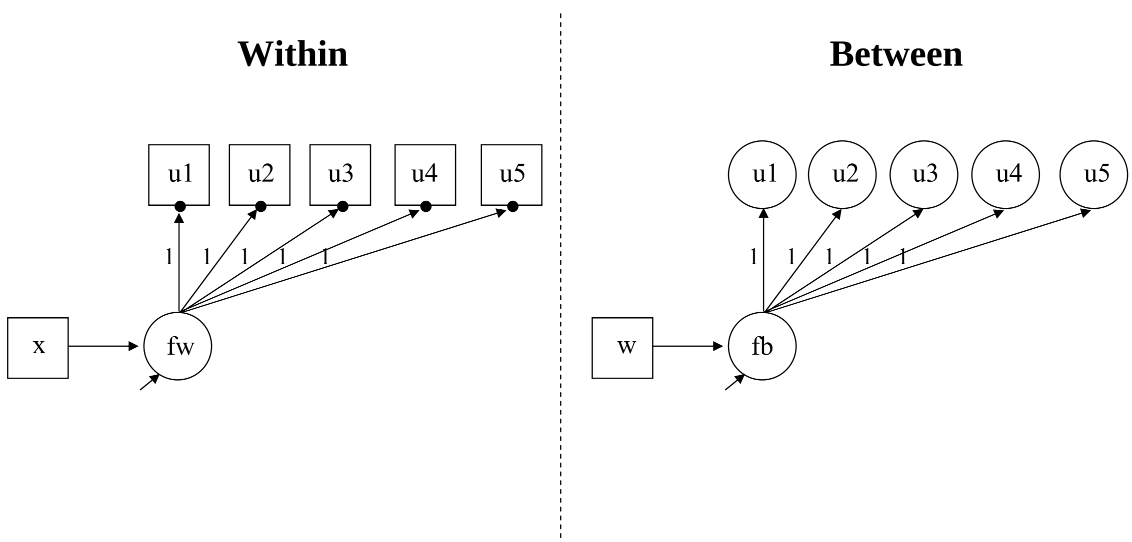
Output Excerpts A Multilevel Growth Mixture Model For LSAY Math Achievement (Continued)

	(Estimates	S.E.	Est./S.E.)
Intercepts			
C#1	0.495	0.535	0.921
C#2	-0.533	0.627	-0.849
Between Level			
C#1 ON			
LUNCH	2.265	0.706	3.208
MSTRAT	-2.876	2.909	-0.988
C#2 ON			
LUNCH	-0.088	1.343	-0.065
MSTRAT	-0.608	2.324	-0.262

Multilevel Discrete-Time Survival Analysis

- Muthén and Masyn (2005) in Journal of Educational and Behavioral Statistics
- Masyn dissertation
- Asparouhov and Muthén

Discrete-Time Survival Frailty Modeling



References

(To request a Muthén paper, please email bmuthen@ucla.edu.)

General

Asparouhov, T. (2005). Sampling weights in latent variable modeling. Mplus Web Notes: No. 7. Forthcoming in Structural Equation Modeling.

Analysis With Multilevel Data

Cross-sectional Data

Harnqvist, K., Gustafsson, J.E., Muthén, B. & Nelson, G. (1994). Hierarchical models of ability at class and individual levels. Intelligence, 18, 165-187. (#53)

Heck, R.H. (2001). Multilevel modeling with SEM. In G.A. Marcoulides & R.E. Schumacker (eds.), New Developments and Techniques in Structural Equation Modeling (pp. 89-127). Lawrence Erlbaum Associates.

Hox, J. (2002). Multilevel analysis. Techniques and applications. Mahwah, NJ: Lawrence Erlbaum.

Kaplan, D. & Elliott, P.R. (1997). A didactic example of multilevel structural equation modeling applicable to the study of organizations. Structural Equation Modeling: A Multidisciplinary Journal, 4, 1-24.

References (Continued)

Kaplan, D. & Kresiman, M.B. (2000). On the validation of indicators of mathematics education using TIMSS: An application of multilevel covariance structure modeling. International Journal of Educational Policy, Research, and Practice, 1, 217-242.

Kreft, I. & de Leeuw, J. (1998). Introducing multilevel modeling. Thousand Oakes, CA: Sage Publications.

Longford, N.T., & Muthén, B. (1992). Factor analysis for clustered observations. Psychometrika, 57, 581-597. (#41)

Muthén, B. (1989). Latent variable modeling in heterogeneous populations. Psychometrika, 54, 557-585. (#24)

Muthén, B. (1990). Mean and covariance structure analysis of hierarchical data. Paper presented at the Psychometric Society meeting in Princeton, N.J., June 1990. UCLA Statistics Series 62. (#32)

Muthén, B. (1991). Multilevel factor analysis of class and student achievement components. Journal of Educational Measurement, 28, 338-354. (#37)

Muthén, B. (1994). Multilevel covariance structure analysis. In J. Hox & I. Kreft (eds.), Multilevel Modeling, a special issue of Sociological Methods & Research, 22, 376-398. (#55)

Muthén, B., Khoo, S.T. & Gustafsson, J.E. (1997). Multilevel latent variable modeling in multiple populations. (#74)

References (Continued)

Muthén, B. & Satorra, A. (1995). Complex sample data in structural equation modeling. In P. Marsden (ed.), Sociological Methodology 1995, 216-316. (#59)

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

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

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

Longitudinal Data

Khoo, S.T. & Muthén, B. (2000). Longitudinal data on families: Growth modeling alternatives. Multivariate Applications in Substance use Research, J. Rose, L. Chassin, C. Presson & J. Sherman (eds.), Hillsdale, N.J.: Erlbaum, pp. 43-78. (#79)

Masyn, K. E. (2003). Discrete-time survival mixture analysis for single and recurrent events using latent variables. Doctoral dissertation, University of California, Los Angeles.

References (Continued)

Muthén, B. (1997). Latent variable modeling with longitudinal and multilevel data. In A. Raftery (ed.) Sociological Methodology (pp. 453-480). Boston: Blackwell Publishers.

Muthén, B. (1997). Latent variable growth modeling with multilevel data. In M. Berkane (ed.), Latent Variable Modeling with Application to Causality (149-161), New York: Springer Verlag.

Muthén, B. & Masyn, K. (in press). Discrete-time survival mixture analysis. Journal of Educational and Behavioral Statistics, Spring 2005.

General

Mplus Analysis

Asparouhov, T. & Muthén, B. (2003a). Full-information maximum-likelihood estimation of general two-level latent variable models. In preparation.

Asparouhov, T. & Muthén, B. (2003b). Maximum-likelihood estimation in general latent variable modeling. In preparation.

Muthén, B. (2002). Beyond SEM: General latent variable modeling. Behaviormetrika, 29, 81-117.

Muthén, B. & Asparouhov, T. (2003b). Advances in latent variable modeling, part II: Integrating continuous and categorical latent variable modeling using Mplus. In preparation.

Numerical integration

Aitkin, M. A general maximum likelihood analysis of variance components in generalized linear models. Biometrics, 1999, 55, 117-128.

Bock, R.D. & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. Psychometrika, 46, 443-459.

General

Muthén, L. & Muthén, B. (1998-2003). Mplus User's Guide. Los Angeles, CA: Muthén & Muthén.

Muthén, L.K. and Muthén, B. (2002). How to use a Monte Carlo study to decide on sample size and determine power. Structural Equation Modeling, 4, 599-620.

Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics, 34, 1-13.

SEM

Latent variable interactions

Klein, A. & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. Psychometrika, 65, 457-474.

Indirect effects

MacKinnon, D.P., Lockwood, C.M., Hoffman, J.M., West, S.G. & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. Psychological Methods, 7, 83-104.

Shrout, P.E. & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. Psychological Methods, 7, 422-445.

Growth Modeling

Carlin, J.B., Wolfe, R., Brown, C.H., Gelman, A. (2001). A case study on the choice, interpretation and checking of multilevel models for longitudinal binary outcomes. Biostatistics, 2, 397-416.

References (Continued)

Duan, N., Manning, W.G., Morris, C.N. & Newhouse, J.P. (1983). A comparison of alternative models for the demand for medical care. Journal of Business and Economic Statistics, 1, 115-126.

Farrington, D.P. & West, D.J. (1990). The Cambridge study in delinquent development: A prospective longitudinal study of 411 males. In Criminality: Personality, Behavior, and Life History, edited by Hans-Jurgen Kernere and G. Kaiser. New York: Springer – Verlag.

Hedeker, D. (2000). A fully semi-parametric mixed-effects regression model for categorical outcomes. Presented at the Joint Statisticacl Meetings, Indianapolis, IN, 2000.

Hedeker, D. & Gibbons, R.D. (1994). A random-effects ordinal regression model for multilevel analysis. Biometrics, 50, 933-944.

Olsen, M.K. & Schafer, J.L. (2001). A two-part random effects model for semicontinuous longitudinal data. Journal of the American Statistical Association, 96, 730-745.

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

Tubin, J. (1958). Estimation of relationships for limited dependent variables. Econometrica, 26, 24-36.

References (Continued)

Multilevel Modeling

Choi, K.C. (2002). Latent variable regression in a three-level hierarchical modeling framework: A fully Bayesian approach. Doctoral dissertation, University of California, Los Angeles.

Seltzer, M., Choi, K., Thum, Y.M. (2002). Examining relationships between where students start and how rapidly they progress: Implications for conducting analyses that help illuminate the distribution of achievement within schools. CSE Technical Report 560. CRESST, University of California, Los Angeles.

Mixture Modeling

Albert, P.A., McShane, L.M., Shih, J.H. & The US SCI Bladder Tumor Marker Network (2001). Latent class modeling approaches for assessing diagnostic error without a gold standard: With applications to p.53 immunohistochemical assays in bladder tumors. Biometrics, 57, 610-619.

Heinen, D. (1996). Latent class and discrete latent trait models: similarities and differences. Thousand Oakes, CA: Sage Publications.

References (Continued)

Land, K.C. (2001). Introduction to the special issue on finite mixture models. Sociological Methods & Research, 29, 275-281.

Lo, Mendell, & Rubin (2001). Testing the number of components in a normal mixture. Biometrika, 88, 767-778.

McLachlan, G.J. & Peel, D. (2000). Finite mixture models. New York: Wiley & Sons.

Muthén, B. (2001b). Two-part growth mixture modeling. Draft.

Muthén, B. (2003a). Statistical and substantive checking in growth mixture modeling. Psychological Methods, 8, 369-377.

Muthén, B. (2003b). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. Forthcoming in D. Kaplan (ed.), Handbook of Quantitative Methodology for the Social Sciences, Sage Publications.

Muthén, B. (2003c). Advances in latent variable modeling of heterogeneous phenotypes. Presentation at NICA, May 2003.

Muthén, B. (2003d). Multilevel growth mixture modeling: Math achievement trajectory classes and high school dropout. Presentation at AERA, March 2003.

References (Continued)

Muthén, B., Kreuter, F. & Asparouhov, T. (2003). Applications of growth mixture modeling to non-normal outcomes. In preparation.

Muthén, B. & Masyn, K. (2001). Mixture discrete-time survival analysis. Submitted to Journal of Educational and Behavioral Statistics.

Muthén, B. & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. Biometrics, 55, 463-469.

Nagin, D.S. & Land, K.C. (1993). Age, criminal careers, and population heterogeneity: Specification and estimation of a nonparametric, mixed Poisson model. Criminology, 31, 327-362.

Qu, Y., Tan, M., & Kutner, M.H. (1996). Random effects models in latent class analysis for evaluating accuracy of diagnostic tests. Biometrics, 52, 797-810.

Nagin, D.S. (1999). Analyzing developmental trajectories: A semi-parametric, group-based approach. Psychological Methods, 4, 139-157.

Roeder, K., Lynch, K.G., & Nagin, D.S. (1999). Modeling uncertainty in latent class membership: A case study in criminology. Journal of the American Statistical Association, 94, 766-776.

References (Continued)

Rumberger, R.W. & Larson, K.A. (1998). Student mobility and the increased risk of high school dropout. American Journal of Education, 107, 1-35.

Singer, J.D. & Willett, J.B. (1993). It's about time: using discrete-time survival analysis to study duration and the timing of events. Journal of Educational Statistics, 18, 155-195.

Wang, C.P., Brown, C.H. & Bandeen-Roche, K. (2002). Residual diagnostics for growth mixture models: Examining the impact of a preventive intervention on multiple trajectories of aggressive behavior. Manuscript submitted for publication.

Yamamoto, K. & Gitomer, D. (1993). Application of a HYBRID model to a test of cognitive skill representation. In N. Frederiksen, R. Mislevy, and I Beijar (eds.), Test theory for new generation of tests. Hillsdale, N.J.: LEA.