Scientific Software International

HLMHCM model for the growth data

1. Introduction to HLMHCM models 1
2. Description of the model 1
3. Creating the MDM file 3
4. Creating the command file...9
5. Interpreting the output 10

1. Introduction to HLMHCM models

In HCM2, level-1 units are nested within cells and cross-classified by two higher-level factors. HLMHCM adds a level within the cells. For example, we may have a growth model for each of a set of students, all of whom live in the same neighborhood and attend the same school. We would say that level-1 units (repeated measures) are nested within level-2 units (children); level-2 units are crossed by rows (neighborhoods) and columns (schools). Another example might involve repeated item responses at a given time for a student encountering a given teacher. The level-1 units are the item responses, nested within occasions (level-2) crossed by rows (students) and columns (teachers).

2. Description of the model

A general hierarchical HLMHCM has three sub-models: a level-1 model and a level-2 model within each cell; and a level-3 model or between-cell model that incorporates row and column effects.

Formally, there are $m = 1,2,\ldots, n_{ijk}$ level-1 units (*e.g.*, repeated measurement of student achievement) nested within level-2 (*e.g.*, students) $I = 1,\ldots, n_{jk}$ nested within cells cross-classified by $j = 1,..., J$ rows (*e.g.*, neighborhoods) and $k = 1,..., K$ columns (*e.g.*, schools).

Here is an example of a data layout for three waves of developmental data ($n_{ijk} = 3$) nested within $J = 10$ students nested within cells cross-classified by $J = 3$ neighborhoods (rows) and $K = 3$ schools (columns):

Organization of data of the HLMHCM example

	$School_1$	$School_2$	$School_3$
$Neighborhood_1$	Y_{1111}, Y_{2111}, Y_{3111} of Stud 1 Y_{1211}, Y_{2211}, Y_{3211} of Stud 2	Y_{1311}, Y_{2311}, Y_{3311} of Stud 3	
$Neighborhood_3$	Y_{1411}, Y_{2411}, Y_{3411} of Stud 4	Y_{1511}, Y_{2511}, Y_{3511} of Stud 5 Y_{1611}, Y_{2611}, Y_{3611} of Stud 6	Y_{1711}, Y_{2711}, Y_{3711} of Stud 7
$Neighborhood_3$	Y_{1811}, Y_{2811}, Y_{3811} of Stud 8 Y_{1911}, Y_{2911}, Y_{3911} of Stud 9		Y_{11011}, Y_{21011}, Y_{31011} of Stud 10

Note that the repeated developmental data are nested within individual students nested within cells cross-classified by neighborhoods and schools. Note that unlike in HCM3, the students *never* leave the neighborhood or school of origin.

Level-1 or "within-unit" model

We represent in the level-1 model the outcome *Y* for response *m* of the level-2 unit *i* cross-classified by row *j* and column *k*.

$$\begin{aligned} Y_{mijk} &= \psi_{0ijk} + \psi_{1ijk} a_{1ijk} + \psi_{2ijk} a_{2ijk} + \cdots + \psi_{pijk} a_{pijk} + \xi_{mijk} \\ &= \psi_{0ijk} + \sum_{p=1}^{P} \psi_{pijk} a_{pijk} + \xi_{mijk} \end{aligned}$$

where

ψ_{0ijk} is the intercept, the expected value of Y_{mijk} when all explanatory variables are set to zero;

ψ_{pijk} are level-1 coefficients of predictors a_{pijk} (p=1,2,…,P) ;

ξ_{mijk} is the level-1 random effect; and

σ^2 is the variance of ξ_{mijk}, that is the level-1 variance. Here we assume that the random term $\xi_{mijk} \sim N(0,\sigma^2)$.

Level-2 or "between-unit" or "within-cell" model

Each of the ψ_{pijk} (p=0,1,…,P) coefficients in the level-1 model becomes an outcome variable in the level-2 or within-cell model:

$$\begin{aligned} \psi_{pijk} &= \pi_{p0jk} + \pi_{p1jk}\alpha_{p01k} + \pi_{p2jk}\alpha_{p02k} + \cdots + \pi_{pQ_p jk}\alpha_{Q_p jk} + e_{pijk} \\ &= \pi_{p0jk} + \sum_{q=1}^{Q_p} \pi_{pqjk}\alpha_{p0jk} + e_{pijk} \end{aligned}$$

π_{p0jk} is the intercept, the expected value of ψ_{pijk} when all explanatory variables are set to zero;

π_{pqjk} are level-1 coefficients of predictors α_{p0jk} (p=1,2,…,P);

e_{pijk} is the level-2 or within-cell random effect, and

τ is the variance-covariance matrix of e_{pijk}, that is the level-2 variance. Here we assume that the random term $e_{pijk} \sim N(0,\tau)$. The vector containing elements e_{pijk} is assumed multivariate normal with a mean zero and a full covariance matrix, τ.

Level-3 model or "between-cell" model

Each of the π_{pqjk} (q = 0, 1, …, Q_p) coefficients in the level-2 or within-cell model becomes an outcome variable in the level-3 or between-cell model:

$$\pi_{pqjk} = \theta_{pq0} + (\beta_{pq1} + b_{pq1j})X_{1k} + (\beta_{pq2} + b_{pq2j})X_{2k} + \cdots + (\beta_{pqR_p} + b_{pqR_pj})X_{R_qj} +$$
$$(\gamma_{pq1} + c_{pq1k})W_{1j} + (\gamma_{pq2} + c_{ps2k})W_{2j} + \cdots + (\gamma_{pqS_p} + c_{pqS_pk})W_{S_pj} +$$
$$b_{pq0j} + c_{pq0k}$$

where

θ_{pq}, β_{pqr} are the fixed effects of column-specific predictors $X_{qk}, r = 1,\ldots, R_p$,
b_{pqrj} are the random effects associated with column-specific predictors X_{rk}. They vary randomly over rows j = 1,..., J;
γ_{pqs} are the fixed coefficients of row-specific predictors W_{sj}, $s = 1,\ldots,S_p$;
c_{pqsk} are the random effects associated with row-specific predictors W_{sj}. They vary randomly over columns k = 1,…, K; and
b_{pqrj}, and c_{pqsk} are residual row- and column-specific random effects, respectively, on π_{pqjk}, after taking into account X_{rk} and W_{sj}.

The vector containing elements b_{pqrj} is assumed multivariate normal with a mean zero and a full covariance matrix Ω. Similarly, the vector with elements c_{pqsk} is assumed multivariate normal with mean vector zero and full covariance matrix Δ.

3. Creating the MDM file

In constructing the MDM file, there is the same range of options for data input as for HLM2. HLMHCM requires three IDs, one for the level-2 (students in our illustration) units, and one for the units of each of the higher-level factors (school and neighborhood), and **the level-2 IDs have to be sorted**. As there are more schools than neighborhoods in our example, we follow the convention adopted for HCM2 and designate school as the row factor and neighborhood as the column factor.

Data input requires a level-1 within-unit file (a time-series student achievement data file in our example), a level-2 or between unit (student-level) file, a level-3 row-factor (school-level) file, and a level-3 column-factor (neighborhood-level) file.

Level-1 file. The level-1 or within-cell file, GROWTH.SAV has 2008 observations collected on 567 students beginning at grade one and followed up annually thereafter for six years. The image below shows the time series data for the first three students. All of them have complete data; typically there are three or four observations per child. Following the student ID field are that student's values on two variables:

- AGE8
  The age of the child minus 8 at each testing occasion. Therefore, it is 0 at age 8, 1 at age 9, etc.
- MATH
  A math test score in an IRT metric.

	studid	age8	math
1	1	-0.420	2.100
2	1	0.580	3.000
3	1	1.580	4.300
4	1	2.580	6.200
5	1	3.580	7.300
6	1	4.580	8.100
7	2	-0.053	2.600
8	2	0.947	2.933
9	2	1.947	3.500
10	2	2.947	4.000
11	2	3.947	5.400
12	2	4.947	7.200
13	3	-0.299	2.700
14	3	0.701	3.400
15	3	1.701	4.400
16	3	2.701	5.200
17	3	3.701	7.100
18	3	4.701	8.500

We see that the first student was about seven and a half years old (AGE8 = –0.420) during the first data collection wave with a math score of 2.1.

Level-2 file. The level-2 units in the illustration are 567 students. The data are stored in the file STUDENT.SAV. The level-2 data for the first eight children are listed below. The first ID is the level-3 row-factor (*i.e.*, school) ID, the second ID is the level-3 column factor (*i.e.*, neighbor) ID, and the third ID is the level-2 (*i.e.*, student) ID. **Note that the level-2 files must be sorted in the same order of level-2 ID.**

There are three variables:

- FEMALE (1 = female, 0 = male)
- BLACK (1 = African-American, 0 = other)
- HISPANIC (1 = Hispanic, 0 = other)

We see, for example, that student 1 who attended school 175 and resided in neighborhood 68 is a African-American male (FEMALE = 0, BLACK = 1, HISPANIC = 0).

	schid	neighid	studid	female	black	hispanic
1	175	68	1	0.000	1.000	0.00
2	99	68	2	0.000	1.000	0.00
3	99	68	3	0.000	1.000	0.00
4	109	23	4	1.000	0.000	0.00
5	165	72	5	0.000	1.000	0.00
6	207	72	6	1.000	1.000	0.00
7	6	72	7	1.000	1.000	0.00
8	148	72	8	0.000	0.000	0.00

Level-3 row-factor file. The level-3 row-factor (school) level file, SCHOOL.SAV, consists of data on 1 variable for 224 schools. The variable is SCHPOV, which is an indicator of school poverty, as measured by the percentage of the total number of students enrolled in free or subsidized lunch programs.

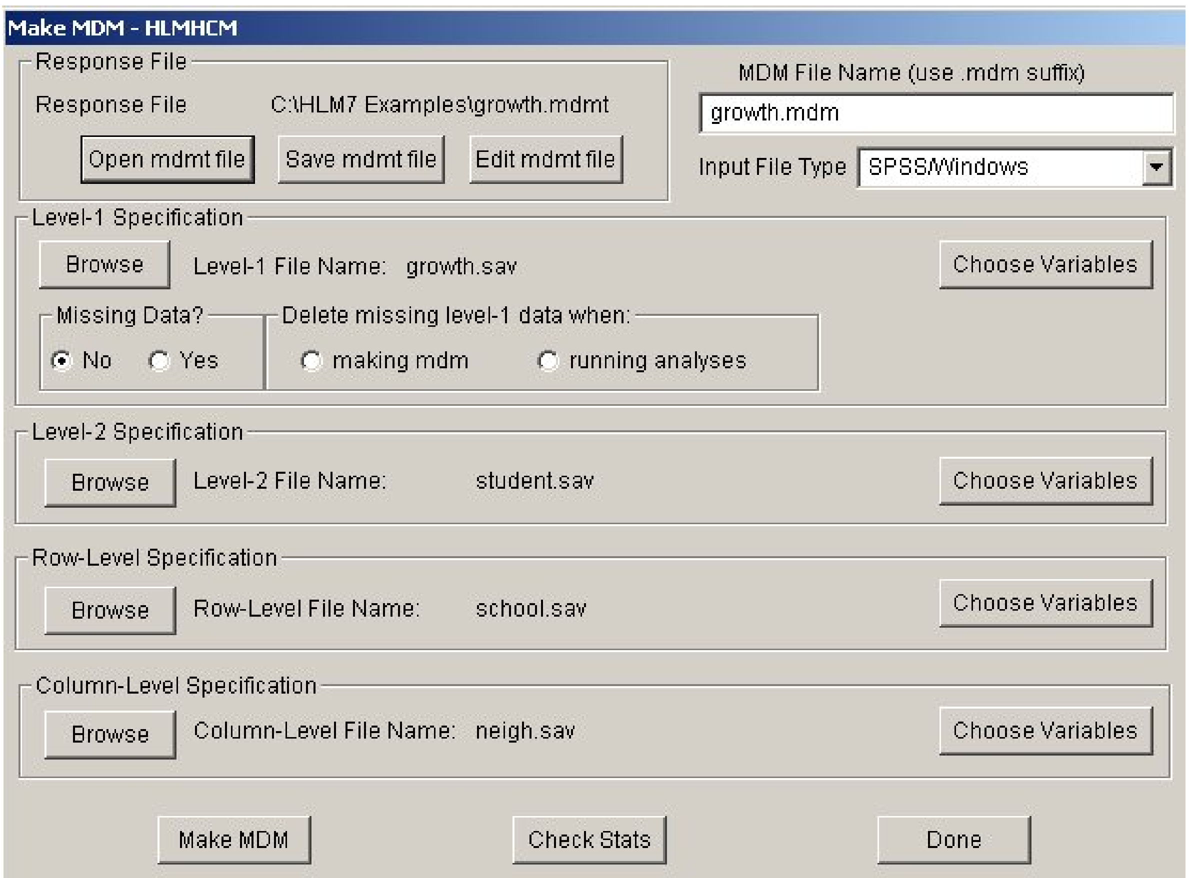
We see that the first school, school 1, has 91% of its students enrolled in free or subsidized lunch programs.

	schid	schpov
1	1	91.400
2	2	78.900
3	3	63.200
4	4	93.000
5	5	71.800
6	6	92.600
7	7	96.200
8	8	83.400

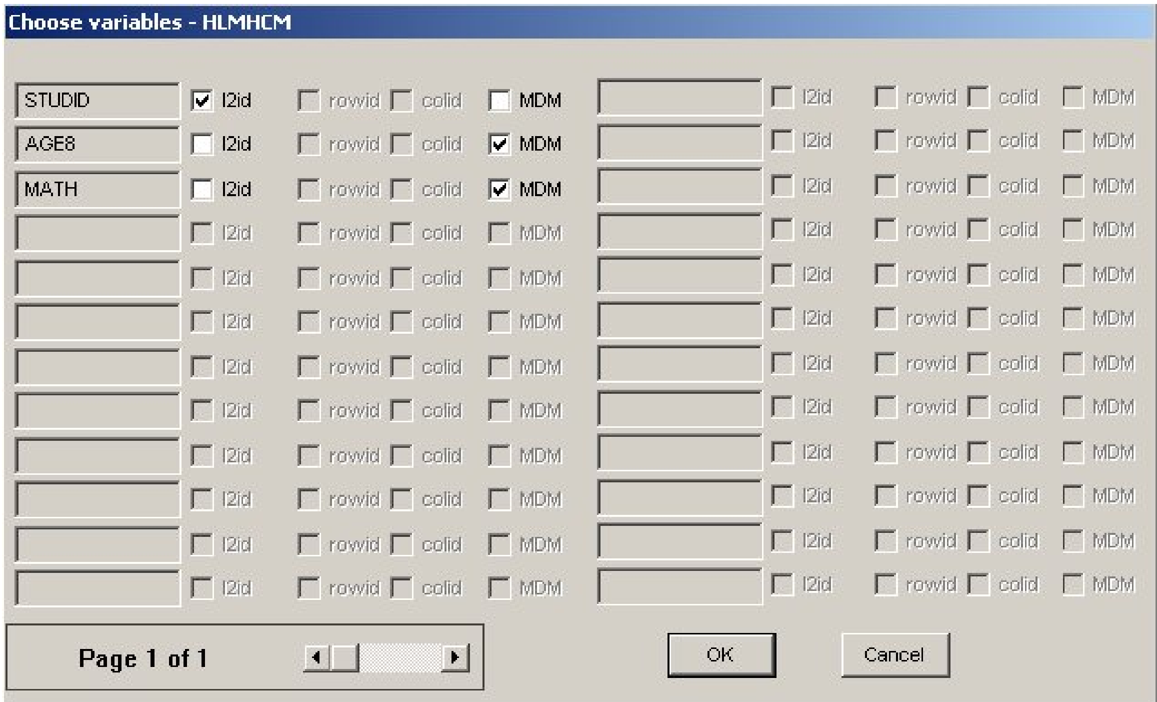
Level-3 column-factor file. The level-3 row-factor (neighborhood) level file, NEIGH.SAV, consists of data on 1 variable for 74 neighborhoods. The variable is DISADV (a scale measuring social deprivation, which incorporates information on the poverty concentration, health, and housing stock of a local community). A measure of neighborhood disadvantage, constructed through an oblique factor analysis from the 1990 decennial census data, tapped the level of poverty and unemployment, and the percentage of families that were headed by females and percentage on welfare (Sampson & Raudenbush, 1999; Sampson, Raudenbush, & Earls, 1997).

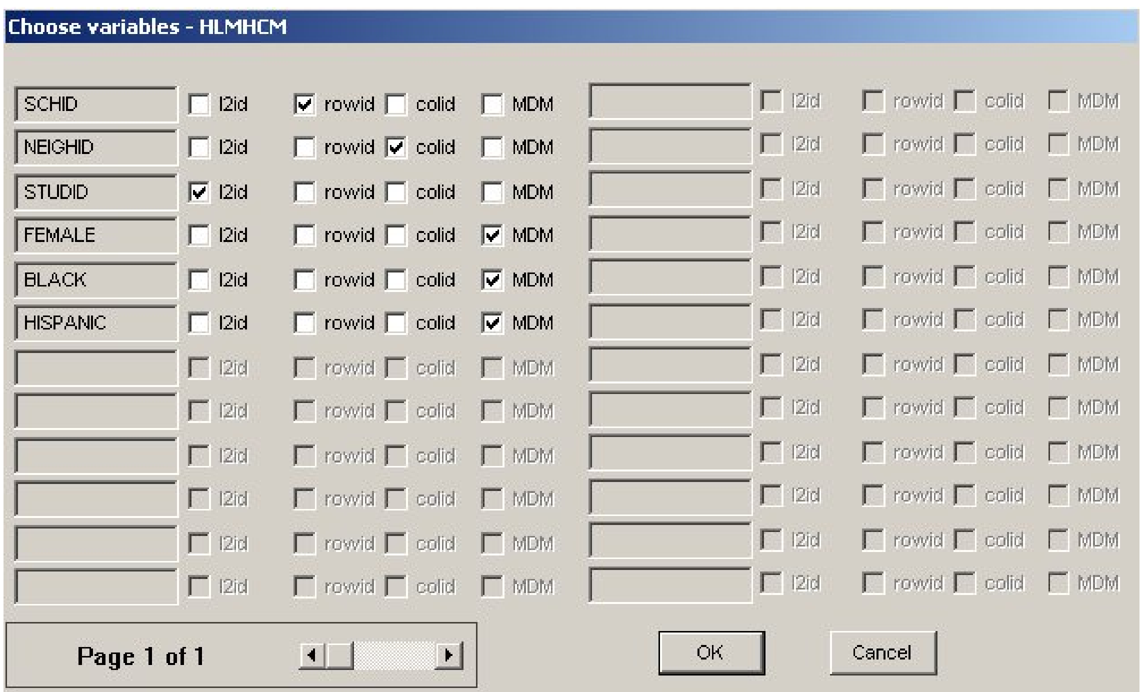
	neighid	disadv
1	1	0.449
2	2	-0.553
3	3	-0.382
4	4	-0.563
5	5	-0.313
6	6	-0.566
7	7	0.160
8	8	1.222

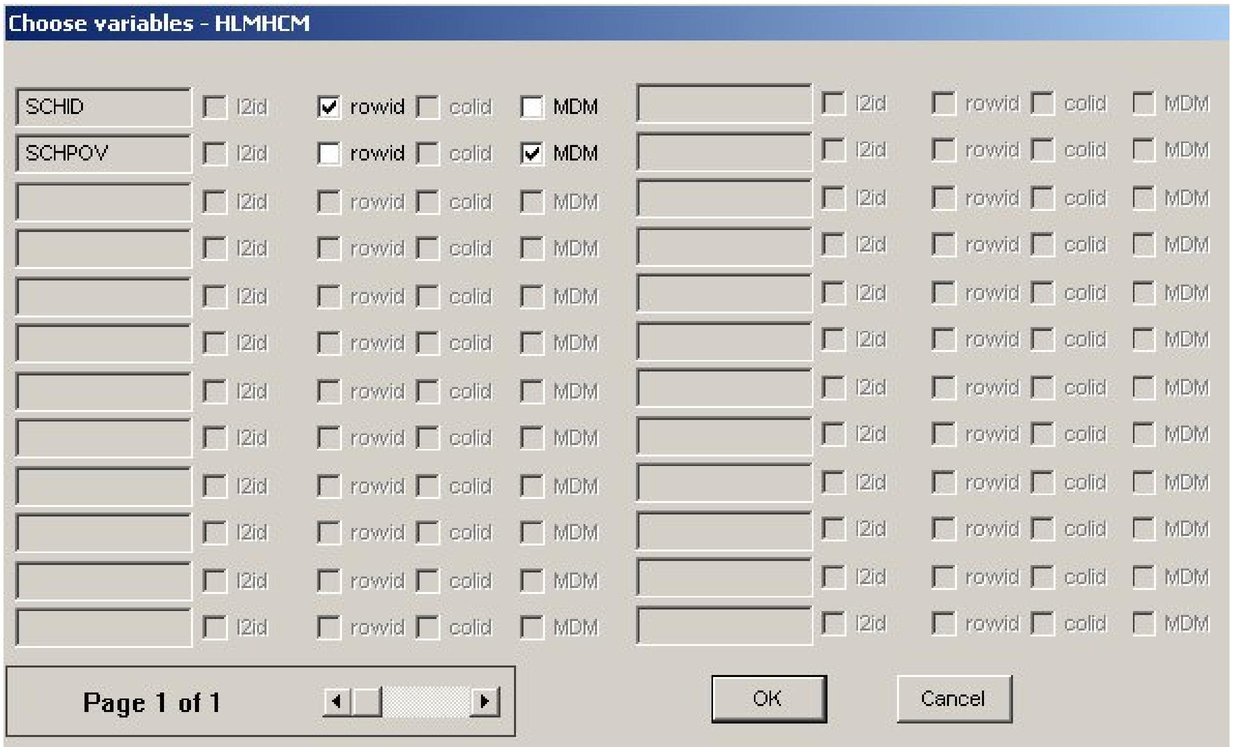
In sum, there are two variables at level 1, three at level 2, and one for each of the level-3 factors.

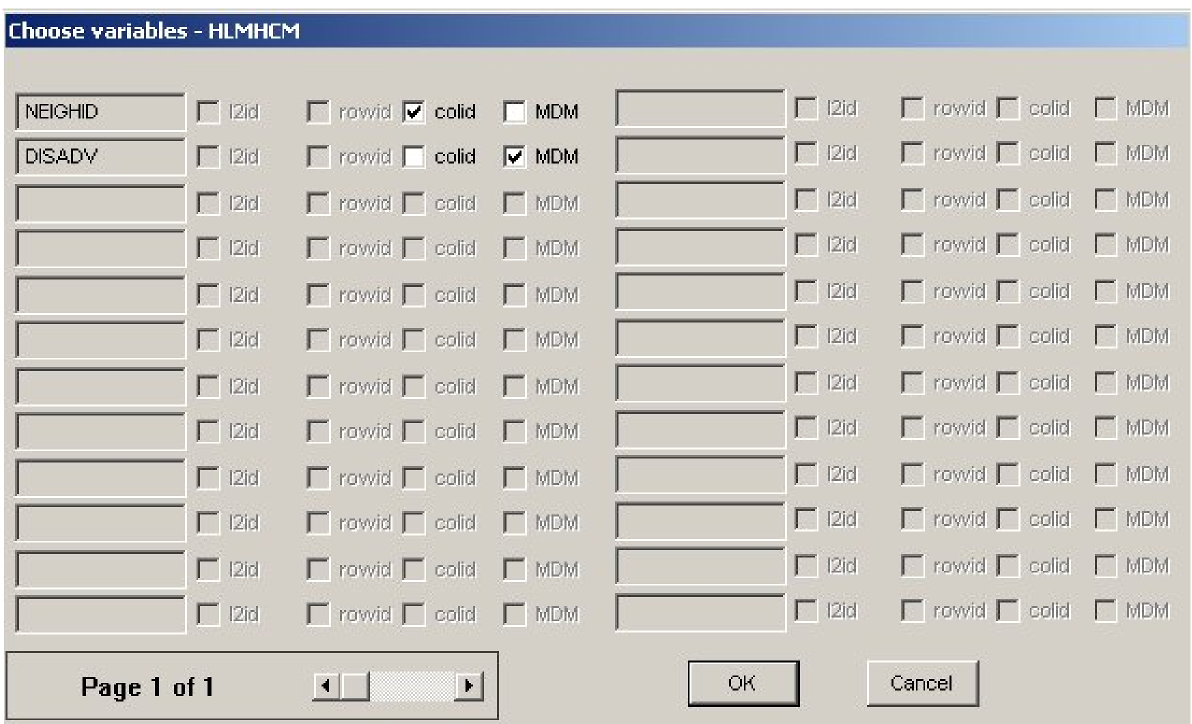


The steps for the construction of the MDM for HLMHCM2 are similar to that for other modules. The user will select **HLMHCM** in the **Select MDM type** dialog box. Note that the program can handle missing data at level 1 or within cell only. The MDM template file, GROWTH.MDMT, contains a log of the input responses used to create the MDM file, GROWTH.MDM, using GROWTH.SAV, STUDNET.SAV, SCHOOL.SAV, and NEIGH.SAV. This information is displayed in the dialog box shown above.. Below we show the dialog boxes for the level-1 file, GROWTH.SAV, the level-2 file, STUDENT.SAV, the level-3 row file, SCHOOL.SAV, and the level-3 column file, SCHOOL.SAV.


Choose variables - HLMHCM
STUDID
AGE8
MATH
l2id
rowid
colid
MDM
Page 1 of 1
OK
Cancel


Choose variables - HLMHCM
SCHID
NEIGHID
STUDID
FEMALE
BLACK
HISPANIC
l2id
rowid
colid
MDM
Page 1 of 1
OK
Cancel


Choose variables - HLMHCM
SCHID
l2id
rowid
colid
MDM
SCHPOV
Page 1 of 1
OK
Cancel

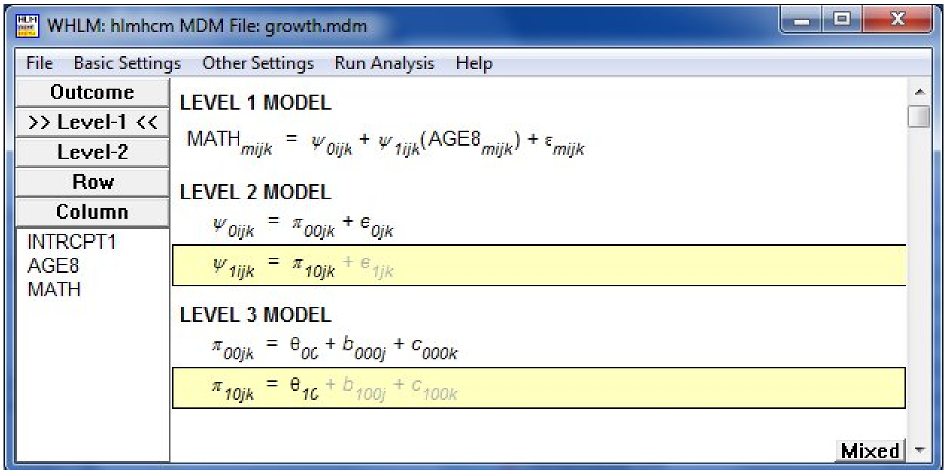

Choose variables - HLMHCM
NEIGHID
l2id
rowid
colid
MDM
DISADV
Page 1 of 1
OK
Cancel

4. Creating the command file

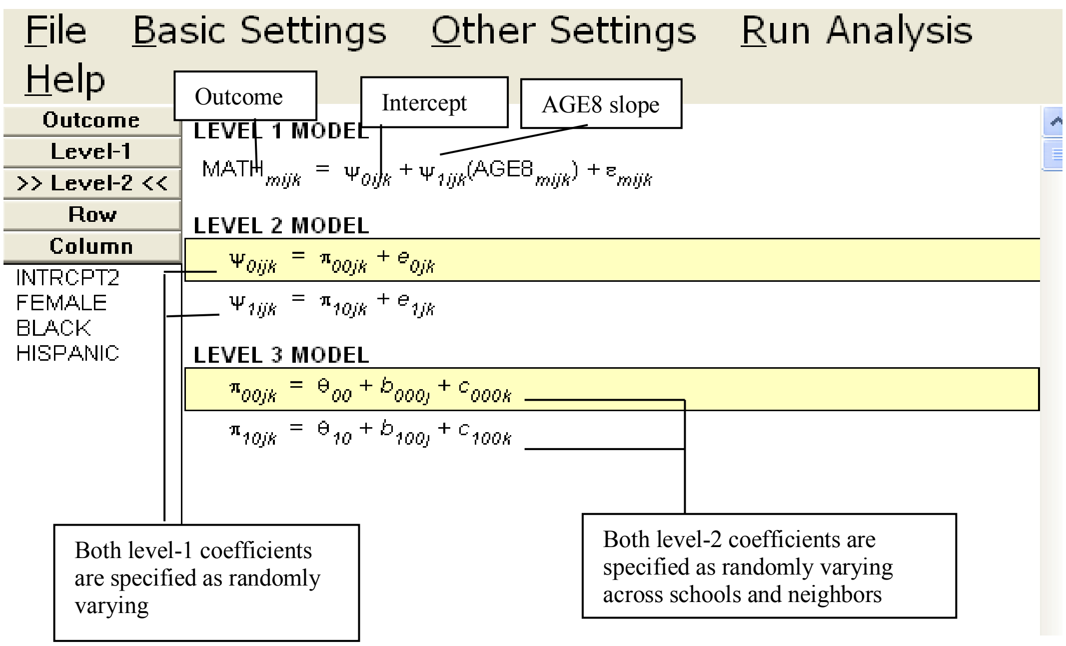
Once the MDM file is constructed, it can be used as input for the analysis. Model specification has three steps:

1. Specification of the level-1 model. In our case we shall model mathematics achievement (MATH) as the outcome, to be predicted by AGE8. Hence, the level-1 model will have two coefficients for each student: the intercept and the AGE slope.
2. Specification of the level-2 prediction model. Here each level-1 coefficient – the intercept and the AGE8 slope in our example – becomes an outcome variable. We may select certain student characteristics to predict each of these level-1 coefficients. In principle, the level-2 parameters then describe the distribution of growth curves cross-classified by schools and neighborhoods.
3. Specification of level-1 coefficients as random or non-random across level-two units. We shall model the intercept and the AGE8 slope as varying randomly across the students cross-classified by schools and neighborhoods.
4. Specification of the level-3 row- and/or column-factor prediction model. Here each level-2 coefficient becomes an outcome, and we can select row- and/or column-factor variables to predict school-to-school and neighbor-to-neighbor variation in these level-2 coefficients. In principle, this model specifies how schools and neighborhoods differ with respect to the distribution of growth curves within them.
5. Specification of the residual row and column as random or non-random, the effects associated with row-specific predictors as varying randomly or fixed over columns, and the effects associated with column-specific predictors as varying randomly or fixed over rows. We shall test whether the associations between neighborhood disadvantage (a column-specific predictor) and growth parameters vary over schools.

In this example, the first in a set of two based on these data, we set up an unconditional linear growth model for the Growth data. From the WHLM window, open the **File** menu. Choose **Create a new model using an existing MDM file** to open an **Open MDM File** dialog box. Open the existing MDM file (GROWTH.MDM in our example). Click on the name of the outcome variable (MATH in our example). Click **Outcome variable**. Next, select AGE8 as level-1 predictor. The specified model will appear in equation format.



Model the intercept and the AGE8 slope as varying randomly across the students cross-classified by schools and neighborhoods by clicking on the equations for ψ_{1ijk} and π_{10jk}, and then activating the random effects e_{1jk} and c_{100k} respectively. The final model is shown below.



Save the model before running the analysis.

5. Interpreting the output

The results of the analysis are given below.

Specifications for this HLM-HCM run

Problem Title: UNCONDITIONAL LINEAR GROWTH MODEL
The data source for this run = growth.mdm

The command file for this run = growth1.hlm
Output file name = growth1.html
The maximum number of level-1 units = 2008
The maximum number of level-2 units = 567
The maximum number of row units = 224
The maximum number of column units = 74
The maximum number of iterations = 100
Method of estimation: full maximum likelihood

The outcome variable is MATH

Summary of the model specified

Level-1 Model

$$\text{MATH}_{mijk} = \psi_{0ijk} + \psi_{1ijk}*(\text{AGE8}_{mijk}) + \varepsilon_{mijk}$$

Level-2 Model

$$\psi_{0ijk} = \pi_{00jk} + e_{0jk}$$
$$\psi_{1ijk} = \pi_{10jk} + e_{1jk}$$

Row/Column Model

$$\pi_{00jk} = \theta_{00} + b_{000j} + c_{000k}$$
$$\pi_{10jk} = \theta_{10} + b_{100j} + c_{100k}$$

For starting values, data from 1967 level-1, 526 level-2, 219 rows, and 74 column records were used

Final Results - iteration 814

Iterations stopped due to small change in likelihood function

σ^2 = 0.16452

τ

INTRCPT1	AGE8
INTRCPT2,e_0	INTRCPT2,e_{1jk}
0.27574	0.07972
0.07972	0.03283

τ (as correlations)

1.000	0.838
0.838	1.000

Note that the estimated correlation between true status at AGE = 8 and true rate of change is estimated to be 0.838 for students in the same cell cross-classified by schools and neighborhoods.

Ω

INTRCPT1 INTRCPT2 ICPTROW,b_{000}	AGE8 INTRCPT2 ICPTROW,b_{100}
0.10927	-0.00606
-0.00606	0.00580

Ω (as correlations)

1.000	-0.241
-0.241	1.000

Note that the estimated correlation between true school mean status at AGE = 8 and true school-mean rate of change is estimated to be -0.241.

Δ

INTRCPT1 INTRCPT2 ICPTCOL,c_{000}	AGE8 INTRCPT2 ICPTCOL,c_{100}
0.02840	0.01363
0.01363	0.00720

Δ (as correlations)

1.000	0.954
0.954	1.000

Note that the estimated correlation between true neighborhood mean status at AGE = 8 and true neighborhood-mean rate of change is estimated to be 0.954.

The value of the log-likelihood function at iteration 814 = -1.917348E+003

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. *d.f.*	*p*-value
For INTRCPT1, π_0					
INTRCPT2, INTERCEPT, θ_{00}	2.257403	0.042925	52.589	274	<0.001
For AGE8, π_1					
INTRCPT2, INTERCEPT, θ_{10}	0.880177	0.016734	52.598	274	<0.001

The above table indicates that the average growth rate is significantly positive at 0.880 logits per year, $t = 52.598$.

Final estimation of level-1 and level-2 variance components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	*p*-value
INTRCPT1, e_0	0.52510	0.27574	268	4818.18751	<0.001
AGE8, e_{1jk}	0.18119	0.03283	268	1465.94774	<0.001
σ^2,ε	0.40561	0.16452			

Note: The chi-square statistics reported above are based on only 526 of 567 units that had sufficient data for computation. Fixed effects and variance components are based on all the data.

The results above indicate significant variability among children cross-classified by schools and neighborhoods in terms of mean status at AGE = 8 (χ^2 = 4818.18751, *df* = 268) and in terms of yearly rate of change (χ^2 = 1465.94774, *df* = 268).

Final estimation of row level variance components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	*p*-value
INTRCPT1/ INTRCPT2/ ICPTROW,b_{000}	0.33055	0.10927	224	87.39230	>0.500
AGE8/ INTRCPT2/ ICPTROW,b_{100}	0.07616	0.00580	224	201.21512	>0.500

The results above indicate there is no significant variability among schools in terms of mean status at AGE = 8 (χ^2 = 87.39230, *df* = 224) and in terms of yearly rates of change (χ^2 = 201.21512, *df* = 224).

Final estimation of column level variance components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	*p*-value
INTRCPT1/INTRCPT2/ ICPTCOL,c_{000}	0.16851	0.02840	73	1316.77855	<0.001
AGE8/INTRCPT2/ ICPTCOL,c_{100}	0.08484	0.00720	73	831.88840	<0.001

The results above indicate significant variability among neighbors in terms of mean status at AGE = 8 (χ^2 = 1316.77855, *df* = 73) and in terms of yearly rates of change (χ^2 = 831.88840, *df* = 73).

Statistics for the current model

Deviance = 3834.695088
Number of estimated parameters = 12