



HCM2 model for the Scotland data

1. Introduction to HCM2 models	1
2. Description of the model	1
3. Creating the MDM file	3
4. Creating the command file.....	7
5. Interpreting the output	8

1. Introduction to HCM2 models

All of the applications discussed thus far have involved a strictly hierarchical data structure. Such nesting structures would occur, for example, in a study of neighborhood and school effects on child development in which all children living in the same neighborhood attended the same school, with multiple neighborhoods per school. In this case we would have children at level 1 nested within neighborhoods at level 2 and neighborhoods nested within schools at level 3. Alternatively, we might have a nested structure in which every child attending a given school lived in the same neighborhood, with multiple schools per neighborhood. In this case, we would have children nested within schools nested within neighborhoods. HLM3 can be used to accommodate such three-level nested data structures. However, we typically find, in fact, that children who reside in a specific neighborhood can enroll in one of several schools, and each school might draw students from several neighborhoods. In this case, the data gathered will no longer have a purely nested structure. Instead, a cross-classification of students by two higher-level factors, neighborhoods and schools, arises. To handle this more complex data structure while modeling the developmental influences of neighborhoods and schools requires the use of cross-classified random effects models (HCM2).

2. Description of the model

A general random cross-classified model consists of two sub-models: the level-1 or within-cell model and level-2 or between-cell model. The cells refer to the cross-classifications by the two higher-level factors. For example, if the research problem consists of data on students cross-classified by neighborhoods and schools, a cell consists of a set of students who live in the same neighborhood and attend the same school. The level-1 or within-cell model will represent the

relationships among the student-level variables for those students while the level-2 or between-cell model will capture the influences of neighborhood- and school-level factors. Formally, there are $i = 1, 2, \dots, n_{jk}$ level-1 units (*e.g.*, students) nested within each cell cross-classified by $j = 1, \dots, J$ units of the first higher-level factor (*e.g.*, neighborhoods), designated as rows, and $k = 1, \dots, K$ units of the second higher-level factor (*e.g.*, schools), designated as columns.

Level-1 or “within-cell” model

We represent in the level-1 or within-cell model the outcome for case i nested within row j and column k of the two-way classification:

$$Y_{ijk} = \pi_{0jk} + \pi_{1jk}a_{1ijk} + \pi_{2jk}a_{2ijk} + \mathbf{L} + \pi_{pjk}a_{pijk} + e_{ijk}$$

where

π_{0jk} is the intercept, the expected value of Y_{ijk} within cell jk when all explanatory variables are set to zero;
 π_{pjk} are the level-1 coefficients of predictors a_{pijk} , for $p = 1, \dots, P$;
 e_{ijk} is the level-1 or within-cell random effect; and
 σ^2 is the variance of e_{ijk} , that is the level-1 or within-cell variance. Here we assume that the random term $e_{ijk} \sim N(0, \sigma^2)$.

Level-2 or “between-cell” model

Each of the π_{pjk} coefficients in the level-1 or within-cell model becomes an outcome variable in the level-2 or between-cell model:

$$\begin{aligned} \pi_{pjk} = & \theta_{p0} + (\beta_{p1} + b_{p1j})X_{1k} + (\beta_{p2} + b_{p2j})X_{2k} + \mathbf{L} + (\beta_{pQ_p} + b_{pQ_pj})X_{Q_pk} + \\ & (\gamma_{p1} + c_{p1k})W_{1j} + (\gamma_{p2} + c_{p2k})W_{2j} + \mathbf{L} + (\gamma_{pR_p} + c_{pR_pk})W_{R_pj} + \\ & b_{p0j} + c_{p0k} \end{aligned}$$

where

θ_{p0} is the model intercept, the expected value of π_{pjk} when all explanatory variables are set to zero;
 β_{pq} are the fixed effects of column-specific predictors X_{qk} , $q = 1, \dots, Q_p$;
 b_{pqj} are the random effects associated with column-specific predictors X_{qk} . They vary randomly over rows $j = 1, \dots, J$;
 γ_{pr} are the fixed effects of row-specific predictors W_{rj} , $r = 1, \dots, R_p$;

c_{prk} are the random effects associated with row-specific predictors W_{rj} . They vary randomly over columns $k = 1, \dots, K$, and; b_{p0j} and c_{p0k} are residual row and column random effects, respectively, on π_{pjk} , after taking into account X_{qk} and W_{rj} . We assume that $b_{p0j} \sim N(0, \tau_{pb00})$, $c_{p0k} \sim N(0, \tau_{pc00})$, and that the effects are independent of each other.

The vector of random row effects b_{pqj} ($p = 0, \dots, P; q = 0, \dots, Q_p$) is assumed multivariate normal with a mean zero and a full covariance matrix τ . Similarly the vector of random column effects c_{prk} ($p = 0, \dots, P; r = 0, \dots, R_p$) is assumed multivariate normal with mean vector zero and full covariance matrix Δ .

This is the first of a set of three HCM2 models fitted to the Scotland data.

3. Creating the MDM file

Chapter 12 of *Hierarchical Linear Models* discusses two applications of cross-classified random effects models, one with cross-sectional, and the other with longitudinal data. The first application is from a study of neighborhood and school effects on educational attainment in Scotland (Garner & Raudenbush, 1991). Some of the children in this study enrolled in schools located in neighborhoods that were different from the ones they resided in. These students were thus cross-classified by neighborhoods and schools. The range of options for data input are the same as for HLM2 and HLM3. Similar to HLM3, HCM2 requires two IDs, one for each higher-level unit, and the IDs have to be sorted. The two higher-level units in our example are neighborhoods and schools. Whereas the user can choose either higher-level unit as the row or column factor, we adopt the convention that the data are arranged such that the level with more units becomes the row factor and the level with fewer units becomes the column factor. Thus, we will designate the neighborhood ($N = 542$) as the row factor and school ($N = 17$) as the column factor.

Data input requires a level-1 file (student-level file), a level-2 row-factor (neighborhood-level) file, and a level-2 column-factor (school-level) file.

Level-1 file. The level-1 or within-cell file, ATTAIWN.SAV has 2,310 students and 8 variables. The two IDs are NEIGHID for neighborhoods and SCHID for schools. The variables are:

- ATTAI (a measure of educational attainment)
- P7VRQ (Primary 7 verbal reasoning quotient)
- P7READ (Primary 7 reading test scores)
- DAD OCC (father's occupation scaled on the Hope-Goldthorpe scale in conjunction with the Registrar General's social-class index (Willms, 1986))
- DAD UNEMP, an indicator for father's unemployment status (1 if unemployed, 0 otherwise)
- DAD ED, an indicator for father's educational level (1 if schooling past the age of 15, 0 otherwise)
- MOM ED, an indicator for mother's educational level (1 if schooling past the age of 15, 0 otherwise)

otherwise)

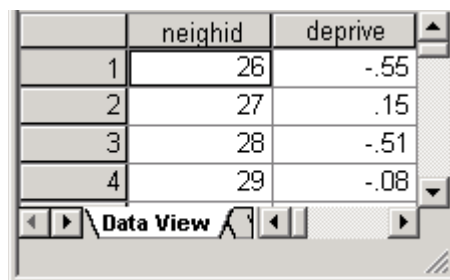
- MALE, an indicator for student gender (1 if male, 0 if female)

Data for the first 15 observations are shown below. Note that five students from Neighborhood 26 and one from Neighborhood 27 attended School 20. These first six observations provided information about two neighborhood-by-school combinations or cells. One of the next nine students living in Neighborhood 29 attended School 18 and the other eight went to School 20. They provided data for two cross-classified neighborhood-by-school cells (see Table 12.1 in *Hierarchical Linear Models*, p. 374, for a display of the organization of the data by counts in each neighborhood-by-school cell).

	neighid	schid	attain	p7vrq	p7read	dadocc	dadunemp	daded	momed	male
1	26.00	20.00	-1.33	-1.03	-.87	-3.45	.00	.00	.00	1.00
2	26.00	20.00	-1.33	-10.03	-27.87	-3.45	.00	.00	.00	.00
3	26.00	20.00	1.52	1.97	11.13	-9.09	.00	.00	.00	.00
4	26.00	20.00	.56	2.97	6.13	2.32	.00	.00	.00	.00
5	26.00	20.00	1.52	17.97	17.13	16.20	.00	.00	.00	1.00
6	27.00	20.00	-.13	3.97	-.87	-3.45	.00	.00	1.00	.00
7	29.00	18.00	.03	8.97	6.13	16.20	.00	1.00	1.00	1.00
8	29.00	20.00	-1.33	-17.03	-23.87	-3.45	.00	.00	.00	1.00
9	29.00	20.00	-1.33	-8.03	-4.87	-3.45	.00	.00	.00	.00
10	29.00	20.00	.16	1.97	-4.87	-11.49	.00	.00	.00	.00
11	29.00	20.00	.56	-.03	-5.87	-3.45	.00	.00	.00	.00
12	29.00	20.00	.16	4.97	11.13	-11.49	.00	.00	.00	1.00
13	29.00	20.00	.74	-4.03	.13	-3.45	.00	.00	.00	1.00
14	29.00	20.00	-1.33	-15.03	-25.87	-3.45	.00	.00	.00	.00
15	29.00	20.00	-.36	-8.03	-13.87	-3.45	.00	.00	.00	.00

Level-2 row-factor file

For our neighborhood example, the level-2 row-factor (neighborhood) level file, ATTAINR.SAV, consists data on 1 variable for 542 neighborhoods. The variable is DEPRIVE (a scale measuring social deprivation, which incorporates information on the poverty concentration, health, and housing stock of a local community). Data from the first 4 neighborhoods are shown below.



	neighid	deprive
1	26	-.55
2	27	.15
3	28	-.51
4	29	-.08

Level-2 column-factor file

The level-2 column-factor (neighborhood) file, ATTAINCO.SAV, has 17 schools and 1 variable. The variable is DUMMY, a dummy variable. Data for the first 4 schools are shown below.

	schid	dummy
1	1.00	1.00
2	2.00	1.00
3	3.00	1.00
4	4.00	1.00

Data View

The steps for the construction of the MDM for HCM2 are similar to the ones described earlier. Select **HCM2** 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 ATTAIN.MDMT contains a log of the input responses used to create the MDM file, ATTAIN.MDM, using ATTAINW.SAV, ATTAINR.SAV, and ATTAINCO.SAV. Figure 13.4 displays the dialog box used to create the MDM file. The dialog boxes for the within-cell file, ATTAINW.SAV, the row-factor file, ATTAINR.SAV, and the column-factor file, ATTAINCO.SAV are given below.

Make MDM - HCM2

MDM template file

File Name: C:\HLM\Examples\ATTAIN.MDMT

MDM File Name (use .mdm suffix): ATTAIN.MDM

Input File Type: SPSSWindows

Open mdmt file Save mdmt file Edit mdmt file

Level-1 Specification

Browse Level-1 File Name: C:\HLM\Examples\attainw.sav Choose Variables

Missing Data?

No Yes

Row-Level Specification

Browse Row-Level File Name: C:\HLM\Examples\attainr.sav Choose Variables

Column-Level Specification

Browse Column-Level File Name: C:\HLM\Examples\attainco.sav Choose Variables

Make MDM Check Stats Done

Choose variables - HCM2

NEIGHID	<input checked="" type="checkbox"/> rowid	<input type="checkbox"/> colid	<input type="checkbox"/> in MDM		<input type="checkbox"/> rowid	<input type="checkbox"/> colid	<input type="checkbox"/> in MDM
SCHID	<input type="checkbox"/> rowid	<input checked="" type="checkbox"/> colid	<input type="checkbox"/> in MDM		<input type="checkbox"/> rowid	<input type="checkbox"/> colid	<input type="checkbox"/> in MDM
ATTAIN	<input type="checkbox"/> rowid	<input type="checkbox"/> colid	<input checked="" type="checkbox"/> in MDM		<input type="checkbox"/> rowid	<input type="checkbox"/> colid	<input type="checkbox"/> in MDM
P7VRQ	<input type="checkbox"/> rowid	<input type="checkbox"/> colid	<input checked="" type="checkbox"/> in MDM		<input type="checkbox"/> rowid	<input type="checkbox"/> colid	<input type="checkbox"/> in MDM
P7READ	<input type="checkbox"/> rowid	<input type="checkbox"/> colid	<input checked="" type="checkbox"/> in MDM		<input type="checkbox"/> rowid	<input type="checkbox"/> colid	<input type="checkbox"/> in MDM
DADOCC	<input type="checkbox"/> rowid	<input type="checkbox"/> colid	<input checked="" type="checkbox"/> in MDM		<input type="checkbox"/> rowid	<input type="checkbox"/> colid	<input type="checkbox"/> in MDM
DADUNEMP	<input type="checkbox"/> rowid	<input type="checkbox"/> colid	<input checked="" type="checkbox"/> in MDM		<input type="checkbox"/> rowid	<input type="checkbox"/> colid	<input type="checkbox"/> in MDM
DADED	<input type="checkbox"/> rowid	<input type="checkbox"/> colid	<input checked="" type="checkbox"/> in MDM		<input type="checkbox"/> rowid	<input type="checkbox"/> colid	<input type="checkbox"/> in MDM
MOMED	<input type="checkbox"/> rowid	<input type="checkbox"/> colid	<input checked="" type="checkbox"/> in MDM		<input type="checkbox"/> rowid	<input type="checkbox"/> colid	<input type="checkbox"/> in MDM
MALE	<input type="checkbox"/> rowid	<input type="checkbox"/> colid	<input checked="" type="checkbox"/> in MDM		<input type="checkbox"/> rowid	<input type="checkbox"/> colid	<input type="checkbox"/> in MDM
	<input type="checkbox"/> rowid	<input type="checkbox"/> colid	<input type="checkbox"/> in MDM		<input type="checkbox"/> rowid	<input type="checkbox"/> colid	<input type="checkbox"/> in MDM
	<input type="checkbox"/> rowid	<input type="checkbox"/> colid	<input type="checkbox"/> in MDM		<input type="checkbox"/> rowid	<input type="checkbox"/> colid	<input type="checkbox"/> in MDM

Page 1 of 1

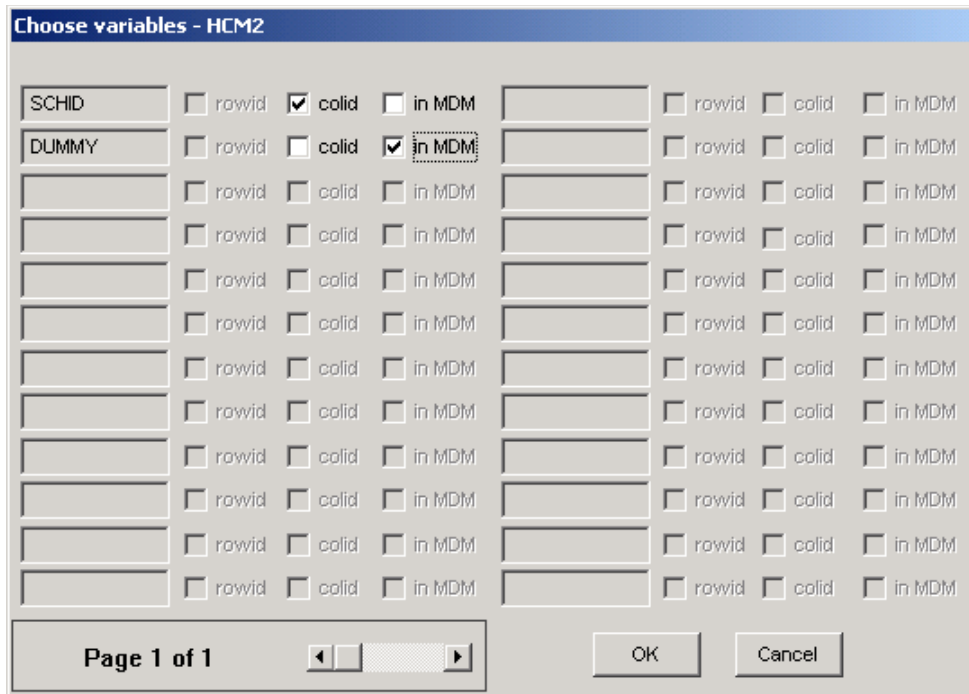
OK Cancel

Choose variables - HCM2

NEIGHID	<input checked="" type="checkbox"/> rowid	<input type="checkbox"/> colid	<input type="checkbox"/> in MDM		<input type="checkbox"/> rowid	<input type="checkbox"/> colid	<input type="checkbox"/> in MDM
DEPRIVE	<input type="checkbox"/> rowid	<input type="checkbox"/> colid	<input checked="" type="checkbox"/> in MDM		<input type="checkbox"/> rowid	<input type="checkbox"/> colid	<input type="checkbox"/> in MDM
	<input type="checkbox"/> rowid	<input type="checkbox"/> colid	<input type="checkbox"/> in MDM		<input type="checkbox"/> rowid	<input type="checkbox"/> colid	<input type="checkbox"/> in MDM
	<input type="checkbox"/> rowid	<input type="checkbox"/> colid	<input type="checkbox"/> in MDM		<input type="checkbox"/> rowid	<input type="checkbox"/> colid	<input type="checkbox"/> in MDM
	<input type="checkbox"/> rowid	<input type="checkbox"/> colid	<input type="checkbox"/> in MDM		<input type="checkbox"/> rowid	<input type="checkbox"/> colid	<input type="checkbox"/> in MDM
	<input type="checkbox"/> rowid	<input type="checkbox"/> colid	<input type="checkbox"/> in MDM		<input type="checkbox"/> rowid	<input type="checkbox"/> colid	<input type="checkbox"/> in MDM
	<input type="checkbox"/> rowid	<input type="checkbox"/> colid	<input type="checkbox"/> in MDM		<input type="checkbox"/> rowid	<input type="checkbox"/> colid	<input type="checkbox"/> in MDM
	<input type="checkbox"/> rowid	<input type="checkbox"/> colid	<input type="checkbox"/> in MDM		<input type="checkbox"/> rowid	<input type="checkbox"/> colid	<input type="checkbox"/> in MDM
	<input type="checkbox"/> rowid	<input type="checkbox"/> colid	<input type="checkbox"/> in MDM		<input type="checkbox"/> rowid	<input type="checkbox"/> colid	<input type="checkbox"/> in MDM
	<input type="checkbox"/> rowid	<input type="checkbox"/> colid	<input type="checkbox"/> in MDM		<input type="checkbox"/> rowid	<input type="checkbox"/> colid	<input type="checkbox"/> in MDM
	<input type="checkbox"/> rowid	<input type="checkbox"/> colid	<input type="checkbox"/> in MDM		<input type="checkbox"/> rowid	<input type="checkbox"/> colid	<input type="checkbox"/> in MDM
	<input type="checkbox"/> rowid	<input type="checkbox"/> colid	<input type="checkbox"/> in MDM		<input type="checkbox"/> rowid	<input type="checkbox"/> colid	<input type="checkbox"/> in MDM

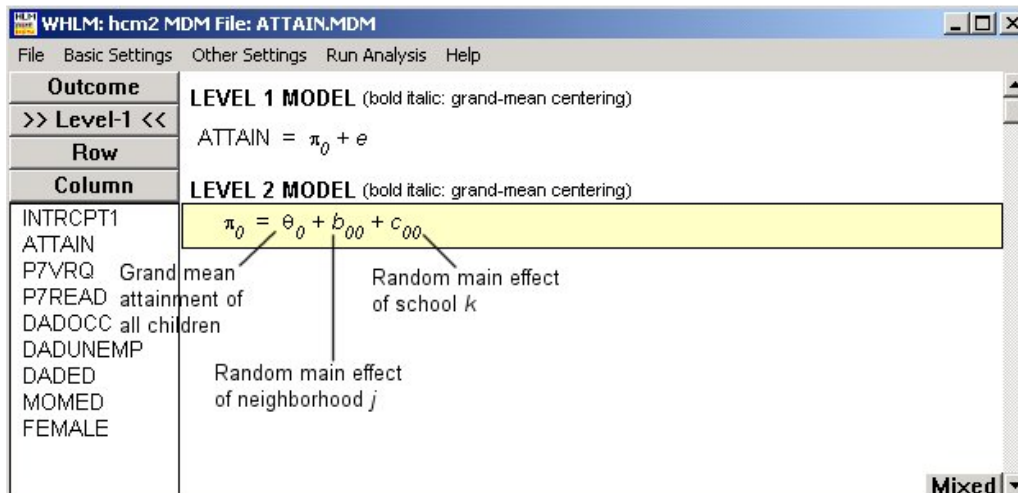
Page 1 of 1

OK Cancel



4. Creating the command file

Once the MDM file is constructed, it can be used as input for the analysis. We first specify a model with no student-, neighborhood-, or school-level predictors. The purpose is to estimate the components of variation that lie between neighborhoods, between schools, and within cells. This is the first in a set of three models fitted to these 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 (ATTAIN.MDM in our example). Click on the name of the outcome variable (ATTAIN in our example). Click **Outcome**

variable. The specified model will appear in equation format.

Save the model before running the analysis.

5. Interpreting the output

The results of the analysis are given below.

Problem Title: Unconditional model

The data source for this run = ATTAIN.MDM

The command file for this run = attain1.hlm

Output file name = hcm2.html

The maximum number of level-1 units = 2310

The maximum number of row-level units = 524

The maximum number of column-level units = 17

The maximum number of iterations = 100

Method of estimation: full maximum likelihood

The maximum number of iterations = 100

Z-structure: independent

The outcome variable is ATTAIN

Summary of the model specified

Level-1 Model

$$\text{ATTAIN}_{ijk} = \pi_{0jk} + e_{ijk}$$

Level-2 Model

$$\pi_{0jk} = \theta_0 + b_{00j} + c_{00k}$$

For starting values, data from 2310 level-1, 524 row-level and 17 column-level records were used

Final Results - iteration 21

Iterations stopped due to small change in likelihood function

$$\sigma^2 = 0.79909$$

Trows

INTRCPT1

ICPTROW, b_{00j}

0.14105

Tcolumns

INTRCPT1

ICPTCOL, c_{00k}

0.07546

The intra-neighborhood correlation, the correlation between outcomes of two students who live in the same neighborhood but attend different schools, is estimated to be:

$$\begin{aligned} \hat{C}orr(Y_{ijk}, Y_{ijk'\varepsilon}) &= \frac{\hat{\tau}_{b00}}{\hat{\tau}_{b00} + \hat{\tau}_{c00} + \hat{\sigma}^2} \\ &= \frac{0.141}{0.141 + 0.075 + 0.799} \\ &= 0.139. \end{aligned}$$

Thus, about 13.9% of the total variance lies between neighborhoods.

The intra-school correlation is the correlation between outcomes of two students who attend the same school but live in different neighborhoods:

$$\begin{aligned} \hat{C}orr(Y_{ijk}, Y_{i'jk}) &= \frac{\hat{\tau}_{c00}}{\hat{\tau}_{b00} + \hat{\tau}_{c00} + \hat{\sigma}^2} \\ &= \frac{0.075}{0.141 + 0.075 + 0.799} \\ &= 0.074. \end{aligned}$$

That is, about 7.4% of the variation lies within schools.

The intra-cell correlation is the correlation between outcomes of two students who live in the same neighborhood and attend the same school:

$$\begin{aligned} \hat{C}orr(Y_{ijk}, Y_{ijk'\varepsilon}) &= \frac{\hat{\tau}_{b00} + \hat{\tau}_{c00}}{\hat{\tau}_{b00} + \hat{\tau}_{c00} + \hat{\sigma}^2} \\ &= \frac{0.141 + 0.075}{0.141 + 0.075 + 0.799} \\ &= 0.212. \end{aligned}$$

Thus, according to the fitted model, about 21% of the variance lies between cells.

The value of the log-likelihood function at iteration 21 = -3.178356E+003

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, π_0					
INTERCEPT, θ_0	0.075357	0.072226	1.043	1769	0.297

Final estimation of row and level-1 variance components:

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1/ ICPTROW, b_{00j}	0.37556	0.14105	523	904.83225	<0.001
level-1, e	0.89392	0.79909			

Final estimation of column level variance components:

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1/ ICPTCOL, c_{00k}	0.27470	0.07546	16	120.45262	<0.001

Statistics for the current model

Deviance = 6356.711470

Number of estimated parameters = 4