Module 11: Three-Level Multilevel Models

MLwiN Practical

George Leckie and Rob French
Centre for Multilevel Modelling

Pre-requisites

- Modules 1-5

Contents

Introduction to the Television School and Family Smoking Prevention and Cessation Project ................................................................. 2

P11.1 Examining and Describing the Data ........................................ 4
  P11.1.1 Exploring the three-level data structure ................................ 4
  P11.1.2 Summarising the response and predictor variables .............. 6

P11.2 A Three-Level Model of THKS.............................................. 14
  P11.2.1 Specifying and fitting the three-level model ...................... 14
  P11.2.2 Interpretation of the model output ................................... 19
  P11.2.3 Calculating coverage intervals, variance partition coefficients (VPCs) and intraclass correlation coefficients (ICCs) .................. 20
  P11.2.4 Predicting and examining school and classroom effects ......... 22

P11.3 Adding Predictor Variables ................................................. 29
  P11.3.1 Adding student level predictor variables .......................... 29
  P11.3.2 Adding school level predictor variables ............................ 31

P11.4 Adding Random Coefficients ............................................. 38
  P11.4.1 Adding classroom level random coefficients ..................... 38
  P11.4.2 Adding cross-level interactions .................................... 42

Further Reading ........................................................................... 48

References ................................................................................. 49

Module 11 (MLwiN Practical): Three-Level Multilevel Models

If you find this module helpful and wish to cite it in your research, please use the following citation:

http://www.bristol.ac.uk/cmm/learning/course.html

Address for correspondence:

George Leckie
Centre for Multilevel Modelling
University of Bristol
35 Berkeley Square
Bristol, BS8 1JA
UK

g.leckie@bristol.ac.uk
Introduction to the Television School and Family Smoking Prevention and Cessation Project

We will analyse data from the Television School and Family Smoking Prevention and Cessation Project (TVSFP) (Flay et al., 1989). The project was designed to test the effect of two different school-based interventions on student tobacco and health knowledge: (1) A social-resistance classroom curriculum (CC); and (2) A television-based programme.

The study sample involved schools with seventh-grade students (age 12 to 13 years) in Los Angeles and San Diego, California. Schools were randomized to one of the four study conditions formed by crossing the two interventions in a $2 \times 2$ design.

<table>
<thead>
<tr>
<th>Classroom Curriculum (CC)</th>
<th>Television-based programme (TV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Neither intervention</td>
<td>TV only</td>
</tr>
<tr>
<td>CC only</td>
<td>CC and TV</td>
</tr>
</tbody>
</table>

The two interventions were delivered to the seventh-grade students in these schools in spring 1986. Students were baselined in January 1986, completed an immediate postintervention questionnaire in April 1986, a one-year follow-up questionnaire in April 1987, and a two-year follow-up questionnaire in April 1988. At each time point, students’ knowledge was assessed using a tobacco and health knowledge scale (THKS), constructed as the number of correct answers to seven binary questionnaire items.

The data were restudied by Hedeker et al. (1994) who used them to illustrate the importance of clustering in clinical and public health research and how multilevel models could be used to account for two-level and three-level hierarchical clustering structures. They concentrated on the sub sample of students who studied at 28 Los Angeles schools and only analysed data from the baseline and postintervention time points. Students who missed data at either time point were listwise deleted.

In this Module, we will explore the three-level hierarchical structure of the data: students (level 1) in classrooms (level 2) in schools (level 3). We will fit three-level multilevel models to examine the relative importance of schools and classrooms as influences on student tobacco and health knowledge and we will pay particular attention to assessing the possible causal effects of the CC and TV interventions.

There is good reason to expect both school and classroom effects on students’ THKS scores. While schools were randomly assigned to the four study conditions, implementation of the CC and TV interventions were carried out at the classroom level. It seems very likely that some schools and teachers would have been more enthused about the interventions than others and this is likely to have had a direct effect on the success of the interventions. We therefore expect to see both between-school and within-school-between-classroom variation in students’ THKS scores.
scores, even after accounting for baseline differences in their tobacco and health knowledge.

We use the Hedeker et al. sub sample of the original data. The data consist of 1,600 students (level 1) nested within 135 classrooms (level 2) nested within 28 schools (level 3).

The response variable is students’ postintervention THKS. We shall treat this score as a continuous response variable in our multilevel models, though we note that we could equally treat this response as ordinal and therefore fit ordinal response multilevel models (see Module 9). The predictor variables of key interest are the school level binary indicators of whether each school was randomly assigned to the CC or TV interventions. The predictor variables also include students’ baseline THKS scores. We will include this predictor variable in our models to adjust for baseline variation in students’ tobacco and health knowledge.

The dataset contains the following variables

<table>
<thead>
<tr>
<th>Variable name</th>
<th>Description and codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>schoolid</td>
<td>School ID</td>
</tr>
<tr>
<td>classid</td>
<td>Class ID</td>
</tr>
<tr>
<td>studentid</td>
<td>Student ID</td>
</tr>
<tr>
<td>postthks</td>
<td>Postintervention THKS score. Scores range from 0 to 7, with a higher score indicating a higher tobacco and health knowledge</td>
</tr>
<tr>
<td>prethks</td>
<td>Baseline THKS score. Scores are measured on the same scale as postthks.</td>
</tr>
<tr>
<td>cc</td>
<td>Classroom curriculum (CC) (0 = no CC, 1 = CC)</td>
</tr>
<tr>
<td>tv</td>
<td>Television (TV) (0 = no TV, 1 = TV)</td>
</tr>
<tr>
<td>ccXtv</td>
<td>CC × TV, the interaction between CC and TV. The variable is constructed by multiplying the variables cc and tv. Note that ccXtv is also binary and 1 = both CC and TV and 0 otherwise.</td>
</tr>
<tr>
<td>cons</td>
<td>A column of ones. This variable will be included as an explanatory variable in all models and its coefficient will be the intercept.</td>
</tr>
</tbody>
</table>
P11.1 Examining and Describing the Data

Open the worksheet ‘11.1.ws’

From within the LEMMA Learning Environment

▪ Go to Module 11: Three-Level Multilevel Models, and scroll down to MLwiN Datafiles
▪ Click ‘11.1.ws’ to open the worksheet

The Names window will appear.

The data consist of 1,600 observations on 9 variables and each variable has been given a variable label. We see, for example, that the response variable postthks ranges from 0 to 7. We shall describe a range of summary statistics for the response and predictor variables in P11.1.2.

P11.1.1 Exploring the three-level data structure

We start by looking in more detail at the structure of the data for the first 10 students.

▪ In the Names window, select all nine variables schoolid through to cons (use the Shift button on the keyboard to select multiple variables)
▪ Under the Data toolbar of the Names window, click View

We see, for example, that student 1 was taught in class 193101 within school 193. The student scored 1 out of 7 on the THKS at baseline (prethks) and 2 out of 7 at postintervention (postthks). The variables cc and tv (and therefore ccXtv) are both zero and so school 193 received neither intervention.
Next, we use the **Command interface** window to confirm that the number of schools and classrooms in the data are 28 and 135, respectively. Specifically, we use the `UNIQ` command to generate new ‘short’ versions of the school and classroom identifier variables which take one record per group.

- From the **Data Manipulation** menu, select **Command interface**
- Type the following into the bottom pane of the window and press **Enter** after typing each command

  ```
  UNIQ ‘schoolid’ c10
  UNIQ ‘classid’ c11
  ```

The **Names** window should update and show the following.

![Names window](image)

The new variable `c10` now contains a single record for each unique school, while the new variable `c11` contains a single record for each unique classroom. The number of records for each of these new variables, 28 and 135, confirms that there are indeed 28 schools and 135 classrooms in the data.

Next, we will explore the distribution of schools, classrooms and students across the four study conditions outlined in our introduction to the data: (1) Neither intervention; (2) CC only; (3) TV only; and (4) CC and TV.

Tabulating `cc` by `tv` at the school level shows seven schools were assigned to each condition.$^2$ The data are therefore balanced, at the school level, across conditions. Note, however, that balance at level 3, or any other level, is by no means a requirement when fitting three-level, or any other, multilevel models.

<table>
<thead>
<tr>
<th>Classroom Curriculum (CC)</th>
<th>Television-based programme (TV)</th>
<th>No</th>
<th>Yes</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td></td>
<td>7</td>
<td>7</td>
<td>14</td>
</tr>
<tr>
<td>Yes</td>
<td></td>
<td>7</td>
<td>7</td>
<td>14</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>14</td>
<td>14</td>
<td>28</td>
</tr>
</tbody>
</table>

$^2$ We do not present the step-by-step instructions to replicate this or subsequent cross-tabulations shown in this section as they are somewhat involved. Such higher-level cross-tabulations are more easily carried out in standard statistical software packages such as R, SPSS or Stata.
This document is only the first few pages of the full version. To see the complete document please go to learning materials and register: 
http://www.cmm.bris.ac.uk/lemma

The course is completely free. We ask for a few details about yourself for our research purposes only. We will not give any details to any other organisation unless it is with your express permission.