

# Module 11: Three-Level Multilevel Models

## Stata Practical

*George Leckie*  
Centre for Multilevel Modelling

### Pre-requisites

- Modules 1-5

## Contents

Introduction to the Television School and Family Smoking Prevention and Cessation Project .....	3
<b>P11.1 Examining and Describing the Data.....</b>	<b>5</b>
P11.1.1 Exploring the three-level data structure .....	6
P11.1.2 Summarising the response and predictor variables .....	10
<b>P11.2 A Three-Level Model of THKS.....</b>	<b>16</b>
P11.2.1 Specifying and fitting the three-level model .....	16
P11.2.2 Interpretation of the model output .....	19
P11.2.3 Calculating coverage intervals, variance partition coefficients (VPCs) and intraclass correlation coefficients (ICCs) .....	21
P11.2.4 Predicting and examining school and classroom effects .....	24
<b>P11.3 Adding Predictor Variables.....</b>	<b>31</b>
P11.3.1 Adding student level predictor variables .....	31
P11.3.2 Adding school level predictor variables .....	34
<b>P11.4 Adding Random Coefficients .....</b>	<b>42</b>
P11.4.1 Adding classroom level random coefficients.....	42
P11.4.2 Adding cross-level interactions .....	45
<b>Further Reading.....</b>	<b>51</b>
<b>References .....</b>	<b>52</b>

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

Leckie, G. (2013). Three-Level Multilevel Models - Stata Practical. LEMMA VLE  
Module 11, 1-52.  
<http://www.bristol.ac.uk/cmm/learning/course.html>

Address for correspondence:

George Leckie  
Centre for Multilevel Modelling  
University of Bristol  
2 Priory Road  
Bristol, BS8 1TX  
UK

[g.leckie@bristol.ac.uk](mailto: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.

		Television-based programme (TV)	
		No	Yes
Classroom Curriculum (CC)	No	Neither intervention	TV only
	Yes	CC only	CC and TV

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, 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

Variable name	Description and codes
<b>schoolid</b>	School ID
<b>classid</b>	Class ID
<b>studentid</b>	Student ID
<b>postthks</b>	Postintervention THKS score. Scores range from 0 to 7, with a higher score indicating a higher tobacco and health knowledge
<b>prethks</b>	Baseline THKS score. Scores are measured on the same scale as <b>postthks</b> .
<b>cc</b>	Classroom curriculum (CC) (0 = no CC, 1 = CC)
<b>tv</b>	Television (TV) (0 = no TV, 1 = TV)
<b>ccXtv</b>	CC × TV, the interaction between CC and TV. The variable is constructed by multiplying the variables <b>cc</b> and <b>tv</b> . Note that <b>ccXtv</b> is also binary and 1 = both CC and TV and 0 otherwise.

## P11.1 Examining and Describing the Data

Load '11.1.dta' into memory and open the do-file for this lesson

From within the LEMMA learning environment

- Go to **Module 11: Three-Level Multilevel Models**, and scroll down to **Stata files**
- Click '**11.1.dta**' to open the dataset

and use the `describe` command to produce a summary of the dataset

```
. describe

Contains data from 11.1.dta
  obs:          1,600
  vars:           8                      25 Jul 2012 14:43
  size:         20,800
-----
storage  display      value
variable name  type    format      label      variable label
-----
schoolid      int      %8.0g              School ID
classid       float    %12.0g             Class ID
studentid     int      %9.0g              Student ID
postthks      byte     %8.0g             Postintervention THKS
prethks       byte     %8.0g             Baseline THKS
cc            byte     %8.0g             Classroom curriculum (CC)
tv            byte     %8.0g             Television (TV)
ccXtv        byte     %8.0g             Interaction (CC*TV)
-----
Sorted by:  schoolid  classid
```

The data consist of 1,600 observations on 8 variables and each variable has been given a variable label.

Standard univariate summary statistics can be requested using the `summarize` command

```
. summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
schoolid	1600	421.9388	112.6662	193	515
classid	1600	422042.3	112667.2	193101	515113
studentid	1600	800.5	462.0245	1	1600
postthks	1600	2.661875	1.382929	0	7
prethks	1600	2.069375	1.26018	0	6
cc	1600	.476875	.4996211	0	1
tv	1600	.499375	.5001559	0	1
ccXtv	1600	.239375	.4268354	0	1

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 using the `list` command to list the data for the first 10 students in the data. We use the `compress` option to abbreviate the variable names and to therefore compress the width of the columns in the output. Doing this prevents each line of the output from being wrapped on to a second line and so ensures readability.

```
. list in 1/10, compress
```

	sch~d	clas~d	stu~d	pos~s	pre~s	cc	tv	ccXtv
1.	193	193101	1	2	1	0	0	0
2.	193	193101	2	2	3	0	0	0
3.	193	193101	3	3	0	0	0	0
4.	193	193101	4	2	3	0	0	0
5.	193	193101	5	1	1	0	0	0
6.	193	193101	6	2	2	0	0	0
7.	193	193101	7	4	3	0	0	0
8.	193	193101	8	2	3	0	0	0
9.	193	193101	9	3	3	0	0	0
10.	193	193101	10	3	1	0	0	0

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 (`pre~s`) and 2 out of 7 at postintervention (`pos~s`). The variables `cc` and `tv` (and therefore `ccXtv`) are both zero and so school 193 received neither intervention.

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.