Module 11: Three-Level Multilevel Models

MLwiN Practical[1]

George Leckie and Rob French
Centre for Multilevel Modelling

Pre-requisites

- Modules 1-5

Contents

[1] This MLwiN practical is adapted from the corresponding Stata practical: Leckie, G. (2013). Three-Level Multilevel Models – Stata Practical. LEMMA VLE Module 12, 1-52. Accessed at http://www.bristol.ac.uk/cmm/learning/course.html.

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

Leckie, G. and French, R. (2013). Three-Level Multilevel Models – MLwiN Practical. LEMMA VLE Module 11, 1-56. 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

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 × 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**
schoolid	School ID
classid	Class ID
studentid	Student ID
postthks	Postintervention THKS score. Scores range from 0 to 7, with a higher score indicating a higher tobacco and health knowledge
prethks	Baseline THKS score. Scores are measured on the same scale as **postthks**.
cc	Classroom curriculum (CC) (0 = no CC, 1 = CC)
tv	Television (TV) (0 = no TV, 1 = TV)
ccXtv	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.
cons	A column of ones. This variable will be included as an explanatory variable in all models and its coefficient will be the intercept.

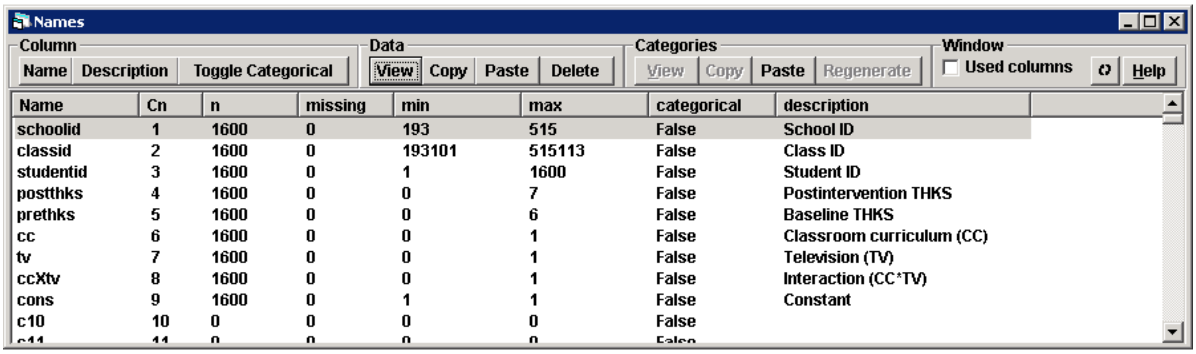
P11.1 Examining and Describing the Data

Open the worksheet '11.1.wsz'

From within the LEMMA Learning Environment

- Go to **Module 11: Three-Level Multilevel Models**, and scroll down to **MLwiN Datafiles**
- Click **'11.1.wsz'** to open the worksheet

The **Names** window will appear.



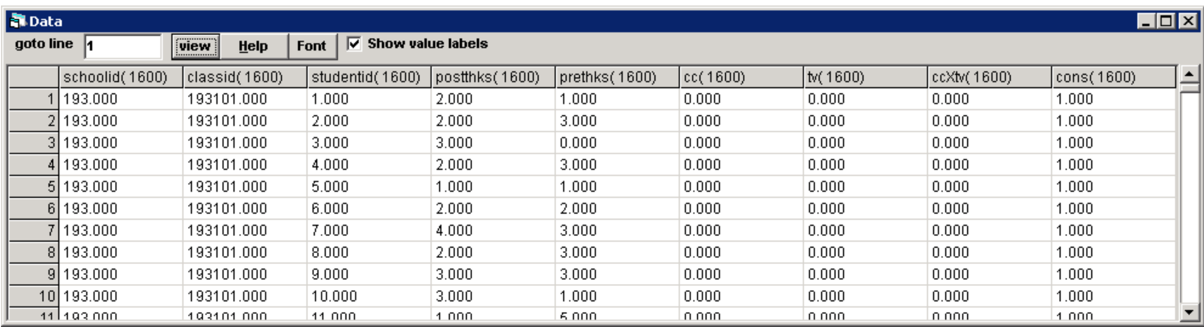
Name	Cn	n	missing	min	max	categorical	description
schoolid	1	1600	0	193	515	False	School ID
classid	2	1600	0	193101	515113	False	Class ID
studentid	3	1600	0	1	1600	False	Student ID
postthks	4	1600	0	0	7	False	Postintervention THKS
prethks	5	1600	0	0	6	False	Baseline THKS
cc	6	1600	0	0	1	False	Classroom curriculum (CC)
tv	7	1600	0	0	1	False	Television (TV)
ccXtv	8	1600	0	0	1	False	Interaction (CC*TV)
cons	9	1600	0	1	1	False	Constant
c10	10	0	0	0	0	False	

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**



	schoolid(1600)	classid(1600)	studentid(1600)	postthks(1600)	prethks(1600)	cc(1600)	tv(1600)	ccXtv(1600)	cons(1600)
1	193.000	193101.000	1.000	2.000	1.000	0.000	0.000	0.000	1.000
2	193.000	193101.000	2.000	2.000	3.000	0.000	0.000	0.000	1.000
3	193.000	193101.000	3.000	3.000	0.000	0.000	0.000	0.000	1.000
4	193.000	193101.000	4.000	2.000	3.000	0.000	0.000	0.000	1.000
5	193.000	193101.000	5.000	1.000	1.000	0.000	0.000	0.000	1.000
6	193.000	193101.000	6.000	2.000	2.000	0.000	0.000	0.000	1.000
7	193.000	193101.000	7.000	4.000	3.000	0.000	0.000	0.000	1.000
8	193.000	193101.000	8.000	2.000	3.000	0.000	0.000	0.000	1.000
9	193.000	193101.000	9.000	3.000	3.000	0.000	0.000	0.000	1.000
10	193.000	193101.000	10.000	3.000	1.000	0.000	0.000	0.000	1.000

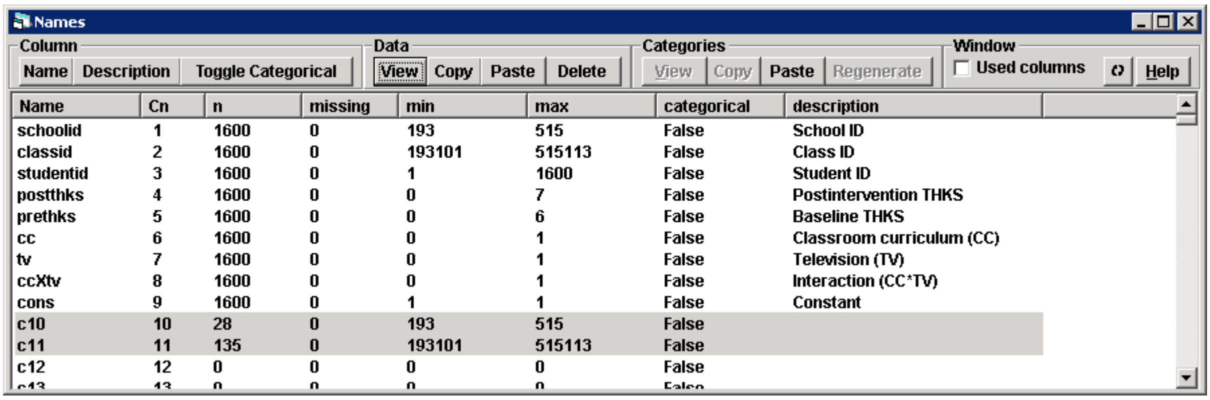
We see, for example, that student 1 was taught in class 193101 within school 193. The student scored 1 out of 7 on the THKS at baseline (**prethks**) and 2 out of 7 at

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.



Name	Cn	n	missing	min	max	categorical	description
schoolid	1	1600	0	193	515	False	School ID
classid	2	1600	0	193101	515113	False	Class ID
studentid	3	1600	0	1	1600	False	Student ID
postthks	4	1600	0	0	7	False	Postintervention THKS
prethks	5	1600	0	0	6	False	Baseline THKS
cc	6	1600	0	0	1	False	Classroom curriculum (CC)
tv	7	1600	0	0	1	False	Television (TV)
ccXtv	8	1600	0	0	1	False	Interaction (CC*TV)
cons	9	1600	0	1	1	False	Constant
c10	10	28	0	193	515	False	
c11	11	135	0	193101	515113	False	
c12	12	0	0	0	0	False	
c13	13	0	0	0	0	False	

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.

		Television-based programme (TV)		
		No	Yes	Total
Classroom Curriculum (CC)	No	7	7	14
	Yes	7	7	14
	Total	14	14	28

Tabulating **cc** by **tv** at the classroom level shows that the number of classes assigned to each condition ranges from 31 to 36. The number of classes varies across conditions due to schools varying in size: some schools have as few as 1 class involved in the study, other schools have as many as 13 classes involved in the study.

		Television-based programme (TV)		
		No	Yes	Total
Classroom Curriculum (CC)	No	34	36	70
	Yes	34	31	65
	Total	68	67	135

Finally, tabulating **cc** by **tv** at the student level shows that the number of students assigned to each condition ranges from 380 and 421. The number of students varies across conditions due to the number of classrooms varying across schools and the number of students varying across classrooms.

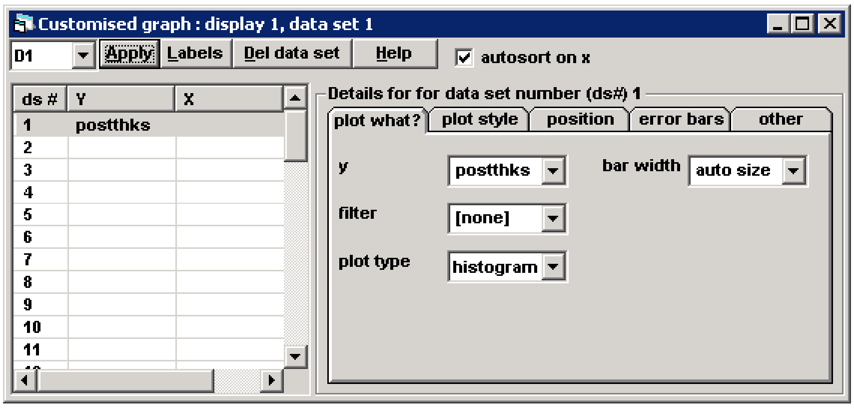
		Television-based programme (TV)		
		No	Yes	Total
Classroom Curriculum (CC)	No	421	416	837
	Yes	380	383	763
	Total	801	799	1600

[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.

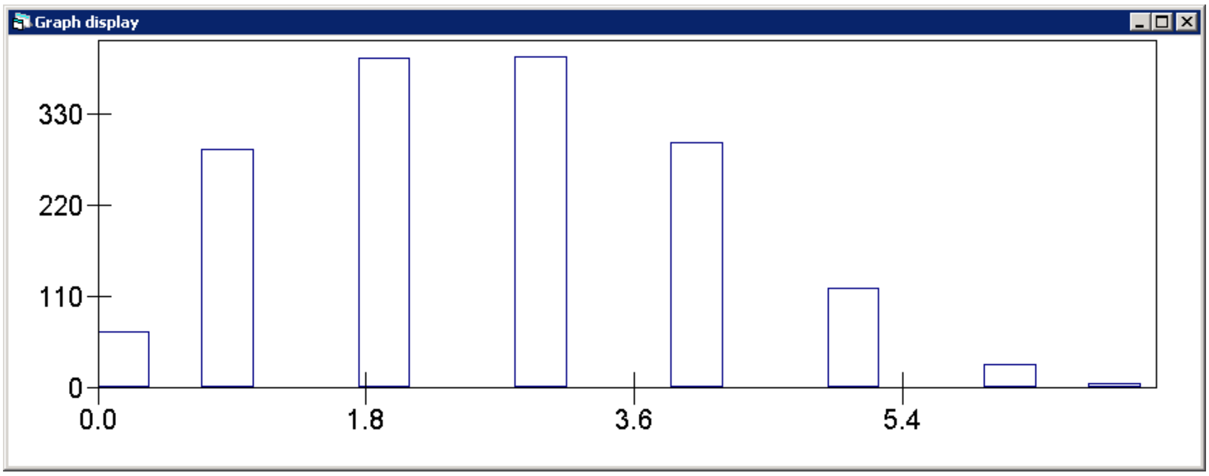
P11.1.2 Summarising the response and predictor variables

We start by plotting the distribution of our response variable, postintervention THKS scores (**postthks**).

- From the **Graphs** menu, select **Customised Graph(s)**
- On the **Plot what?** tab, select **histogram** from the **plot type** drop-down list
- In the **y** drop-down list, select **postthks**
- Check that the window matches that shown below and then click **Apply**

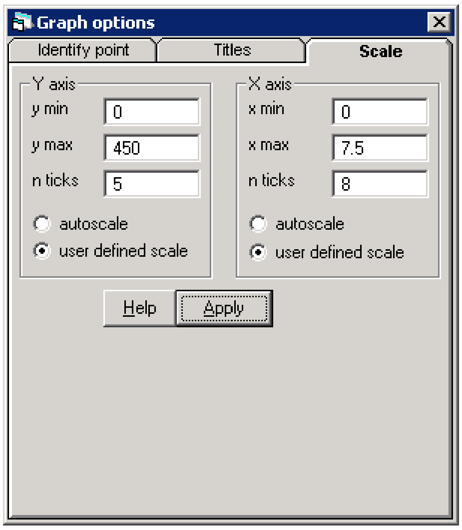


You should see the following graph

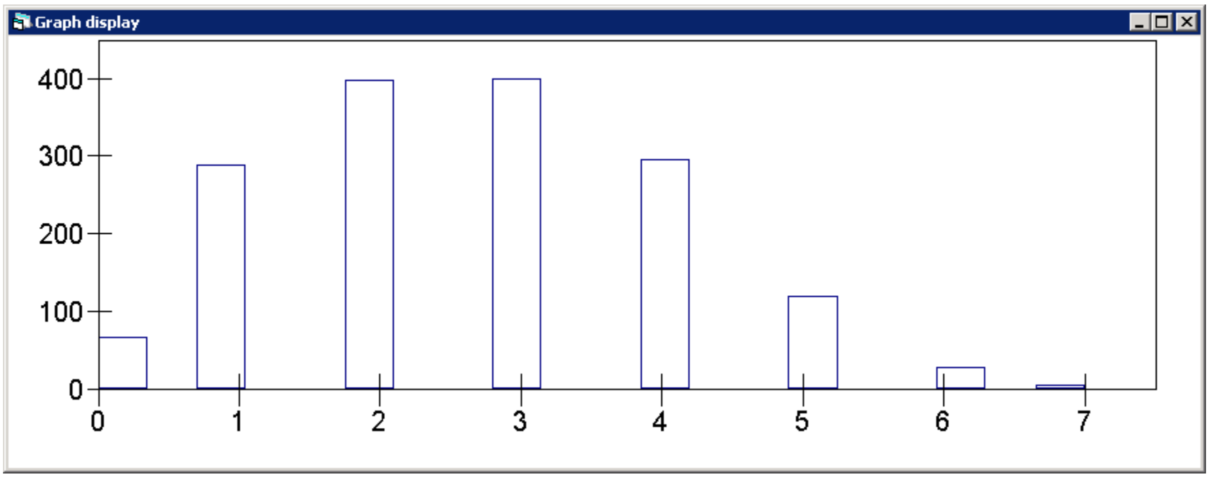


We can make this graph more readable by altering the labelling of the y-axis and x-axis scales.

- Left click anywhere on the graph to open the **Graph options** window
- On the **Scale** tab, in the **Y axis** options, select the **user defined scale** radio button, change **y max** to **450** and **n ticks** to **5**
- In the **X axis** options, select the **user defined scale** radio button, change the **x max** to **7.5** and **n ticks** to **8**
- Check that the window matches that shown below and then click **Apply**



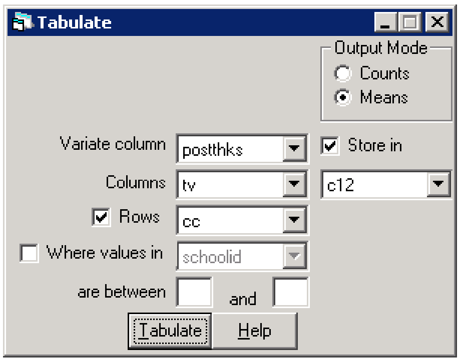
You should see the following graph



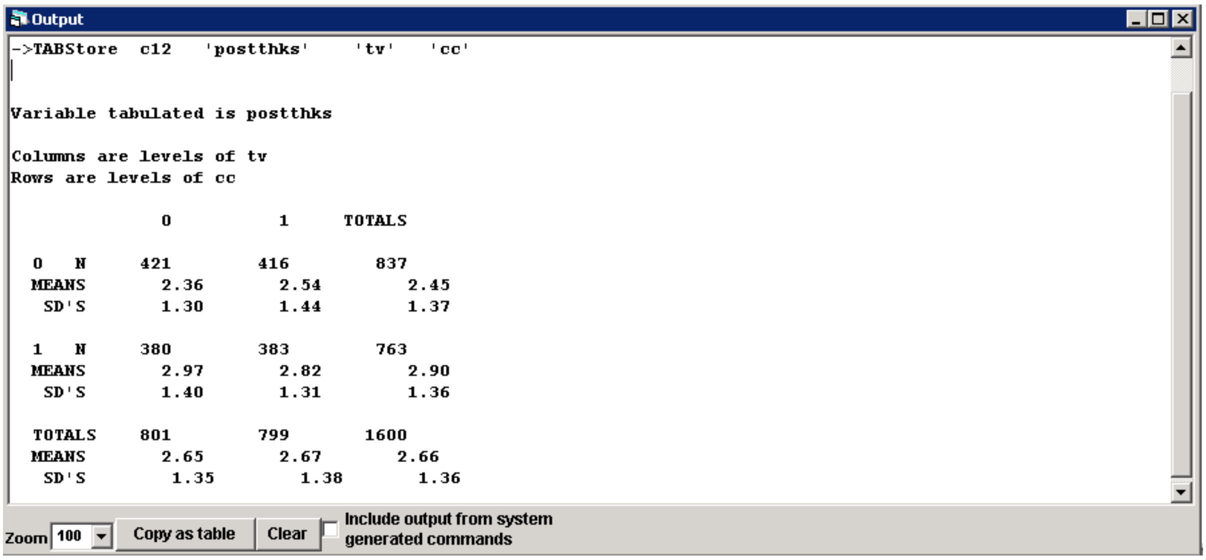
The graph shows that **postthks** is approximately normally distributed. Remember though, that in single-level and multilevel models it is the residuals that are assumed to be normal, not the response; we will check this assumption in P11.2.4.

Next, we calculate student level means for the postintervention scores, separately for each condition. We shall also store these statistics in the column **c12** so that we can graph them later. These statistics replicate those presented in Table 1 of Hedeker *et al.* (1994).

- From the **Basic statistics** menu, select **Tabulate**
- Select **Means** as the **Output Mode**
- Select **postthks** in the **Variate column** drop-down list
- Select **tv** in the **Columns** drop-down list
- Tick the **Rows** checkbox
- Select **cc** in the **Rows** drop-down list
- Tick the **Store in** checkbox
- Select **c12** from the **Store in** drop-down list
- Check that the window matches that shown below and then click **Tabulate**



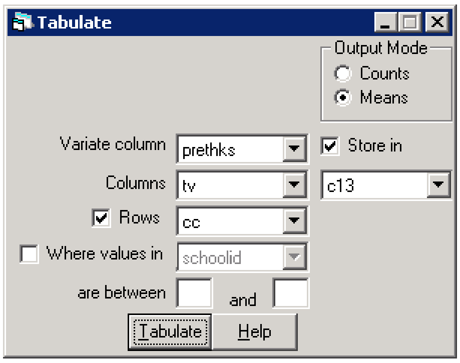
You should see the following output



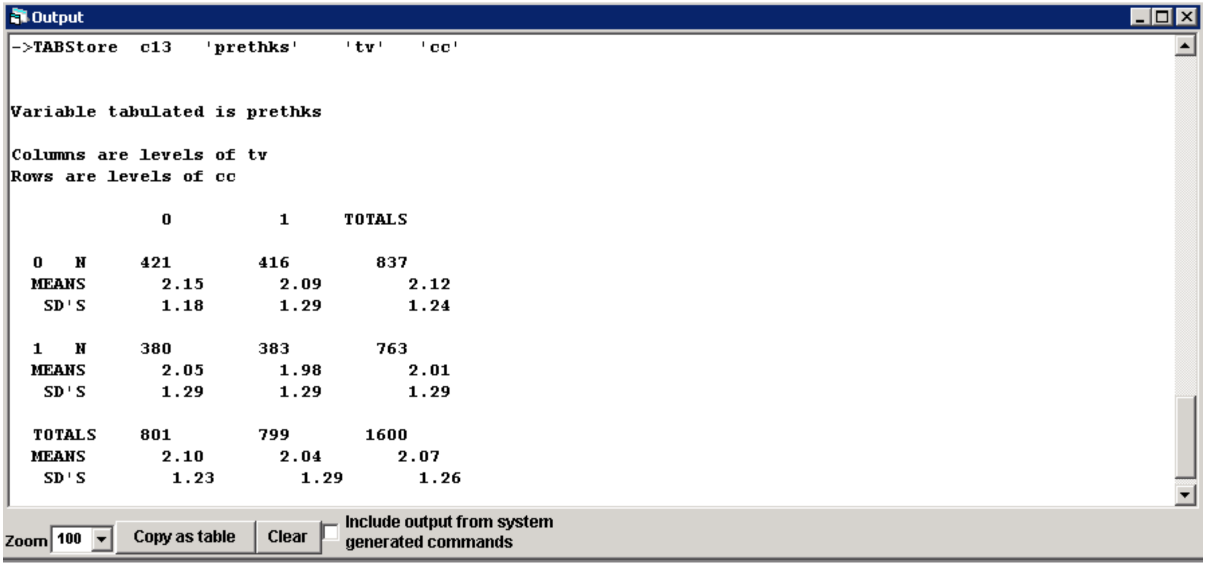
As schools were randomised to the four conditions, it seems reasonable to expect that there should be no baseline differences in THKS scores across conditions and that we should therefore be able to interpret the mean differences seen here as the causal effects of CC and TV. However, we should and can check this assumption by additionally calculating the baseline means separately for each condition.

The baseline means for each condition can be calculated (and stored in column **c13**) as follows.

- Select **prethks** in the **Variate column** drop-down list
- Select **c13** from the **Store in** drop-down list
- Check that the window matches that shown below and then click **Tabulate**



You should see the following output.

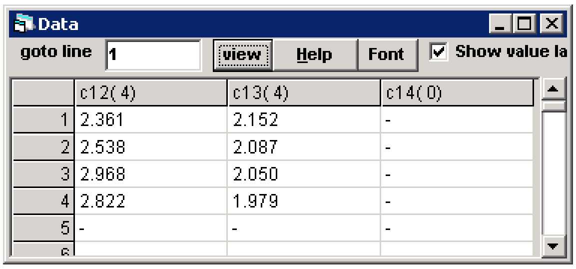


```
->TABStore  c13   'prethks'    'tv'   'cc'


Variable tabulated is prethks

Columns are levels of tv
Rows are levels of cc

              0           1      TOTALS

  0   N     421         416         837
 MEANS        2.15        2.09        2.12
  SD'S        1.18        1.29        1.24

  1   N     380         383         763
 MEANS        2.05        1.98        2.01
  SD'S        1.29        1.29        1.29

 TOTALS     801         799        1600
 MEANS        2.10        2.04        2.07
  SD'S         1.23        1.29        1.26
```


Interestingly, we see that the baseline means do vary somewhat across the four conditions. Perhaps this is not surprising given that there are only 7 schools in each condition. Had 70 schools instead been randomly assigned to each condition, we would expect the baseline means to be considerably more similar than they are here.

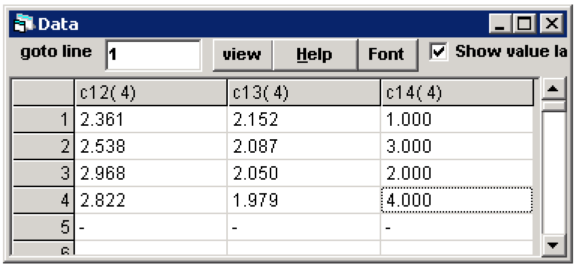
It is often easier to examine descriptive statistics such as those calculated above visually. We shall do this here by plotting a bar chart of the baseline and postintervention means across conditions. First, we examine the columns where we have stored these means.

- In the **Names** window, select the columns **c12**, **c13** and **c14**
- In the **Names** window toolbar, under **Data**, click **View** to open the **Data** window shown below



The mean preintervention and postintervention scores are stored in columns **c13** and **c12**, respectively. Column **c14** is currently empty. The four rows index the four study conditions: (1) Neither intervention; (2) CC only; (3) TV only; and (4) CC and TV. However, they are not in the expected order: the means for the CC only and TV only conditions are presented in rows 3 and 2, rather than the other way around. To clarify matters, we generate a new variable in the empty column **c14** to index the four conditions.

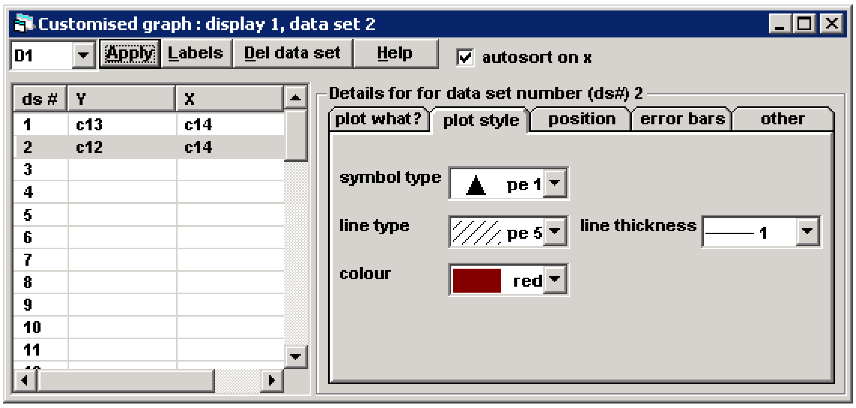
- Type the values **1**, **3**, **2** and **4** into the first four rows of **c14** so that the **Data** window matches that shown below.



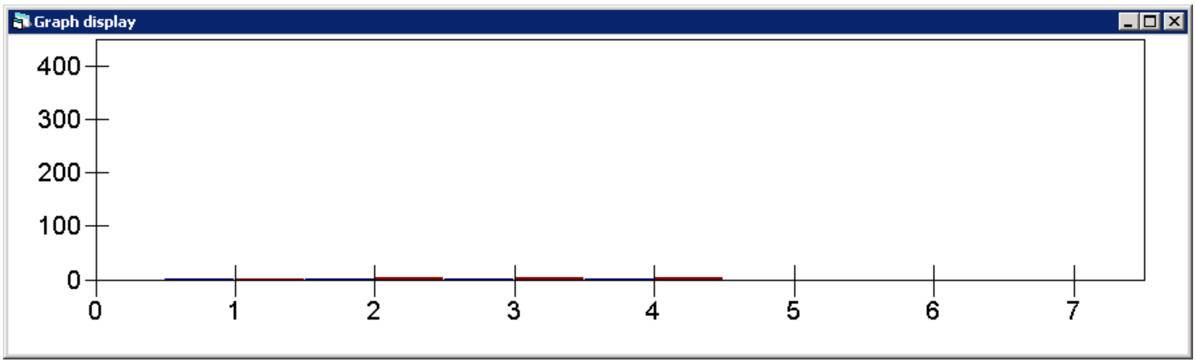
Column **c14** now indexes the four study conditions: 1 = Neither intervention; 2 = CC only; 3 = TV only; 4 = CC and TV.

We are now ready to plot the bar chart.

- From the **Graphs** menu, select **Customised Graph(s)**
- On the **plot what?** tab, select **bar** from the **plot type** drop-down list
- For the **y** drop-down list select **c13**
- For the **x** drop-down list select **c14**
- Select the **plot style** tab, change the **line type** drop-down list to **type 4**
- Click on the **second row** of the table
- Select the **plot what?** tab, select **bar** from the **plot type** drop-down list
- For the **y** drop-down list select **c12**
- For the **x** drop-down list select **c14**
- Select the **plot style** tab, change the **line type** drop-down list to **type 5**
- Change the **colour** drop-down list to **red**
- Check that the window matches that shown below and then click **Apply**



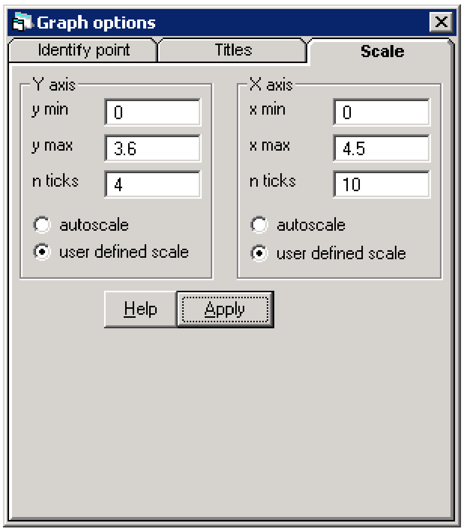
You should see the following graph



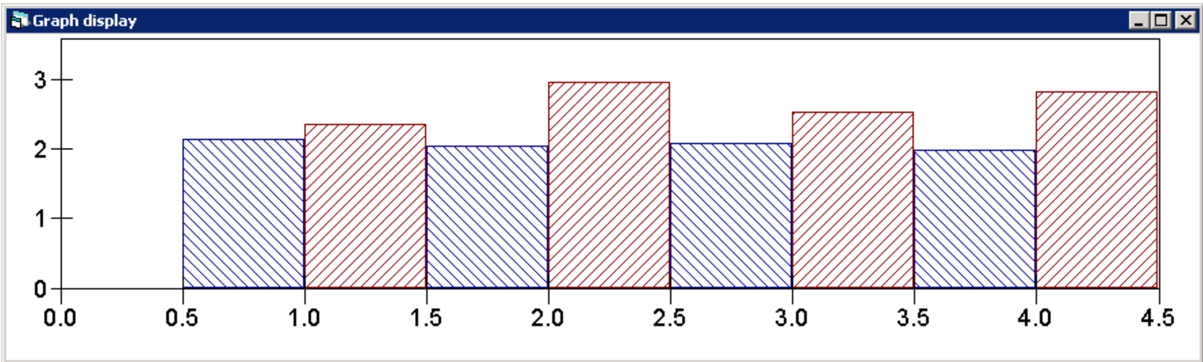
The graph is not very easy to read as the y-axis ranges from 0 to 400 while the mean preintervention and postintervention scores for each condition will lie between the minimum and maximum scores for the test, 0 and 7 respectively. The reason for the strange scaling of the y-axis is that the graph has remembered the y-axis scaling which we specified when plotting the histogram of postintervention scores.

We can make the current graph easier to read by rescaling the y-axis.

- Left click anywhere on the graph to open the **Graph options** window
- On the **Scale** tab, in the **Y axis** options, select the **user defined scale** radio button, change **y max** to **3.6** and **n ticks** to **4**
- In the **X axis** options, select the **user defined scale** radio button, change the **x max** to **4.5** and **n ticks** to **10**
- Check that the window matches that shown below and then click **Apply**



You should see the following graph



The blue and red bars present the mean baseline and postintervention scores across the four conditions, respectively. Moving from left to right along the x-axis, the first two bars give the mean scores for the neither condition, the next two bars present the mean scores for the CC only condition, the fifth and sixth bars correspond to the TV only condition, while the final two bars relate to the CC and TV condition.

Looking at the baseline means (plotted in blue), we see that while they vary little across the four conditions, the baseline mean is highest for the neither condition i.e. the no-treatment control group (the first blue bar). This is important as it means that if we fail to account for this condition having a higher mean score at baseline, we will understate any possible positive causal effects of CC and TV. We must therefore include **prethks** as a predictor variable in our multilevel models.

If we now turn our attention to the postintervention means (plotted in red), we see that they are higher than the baseline means for all four conditions. The fact that the postintervention mean is higher than the baseline mean even for the neither condition suggests that there is a maturation effect whereby students' tobacco and health knowledge improves as they age even in absence of learning interventions such as CC and TV.

The postintervention means also reveal a strong positive effect of the CC intervention that applies both to those not receiving the TV intervention (2.97 vs. 2.36) and for those receiving the TV intervention (2.82 vs. 2.54). The TV effect, however, is less clear. The TV intervention appears to have a positive effect for those not receiving the CC intervention (2.54 vs. 2.36), but a negative effect for those receiving the CC intervention (2.82 vs. 2.97).

To summarise, the CC intervention appears to be effective in increasing students' THKS scores, irrespective of whether they receive the TV intervention. The TV intervention, however, appears only effective for those students who do not also receive the CC intervention.

In the following section we shall fit a series of three-level multilevel models to examine whether these effects remain after adjusting for baseline variation in tobacco and health knowledge and, importantly, to also examine whether any such effects are statistically significant.

P11.2 A Three-Level Model of THKS

Open the worksheet '11.2.wsz'

From within the LEMMA Learning Environment
- Go to **Module 11: Three-Level Multilevel Models**, and scroll down to **MLwiN Datafiles**
- Click '**11.2.wsz**' to open the worksheet

P11.2.1 Specifying and fitting the three-level model

We start by specifying and fitting a three-level variance components model to students' postintervention THKS scores. This model includes only an intercept, school and classroom random effects, and a student level residual error term; the model makes no adjustments for predictor variables. The model simply decomposes the total variance in students' postintervention THKS scores into separate school, classroom and student variance components.

The model is written as

$$\textbf{postthks}_{ijk} = \beta_0 + v_k + u_{jk} + e_{ijk}$$

$$v_k \sim \mathrm{N}(0, \sigma_v^2)$$

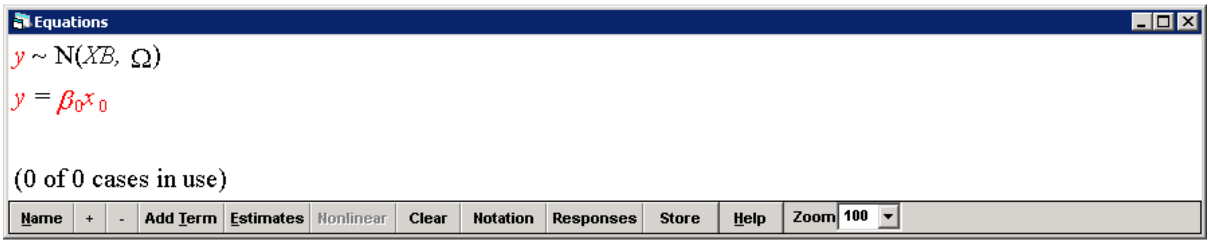
$$u_{jk} \sim \mathrm{N}(0, \sigma_u^2)$$

$$e_{ijk} \sim \mathrm{N}(0, \sigma_e^2)$$

where $\textbf{postthks}_{ijk}$ is the observed postintervention THKS score for student i ($i = 1, \ldots, 1600$) in classroom j ($j = 1, \ldots, 135$) in school k ($k = 1, \ldots, 28$), β_0 is the mean score across all schools, v_k is the effect of school k, u_{jk} is the effect of classroom j, and e_{ijk} is the student level residual error term. The school, classroom effects and the student level residual errors are assumed independent and normal distributed with zero means and constant variances.

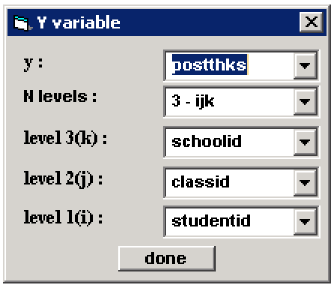
We will now specify the above three-level model in MLwiN.

- From the **Model** menu, select **Equations**

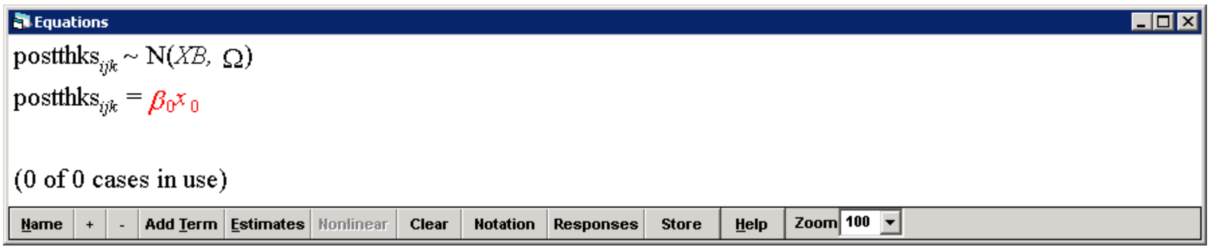


First, specify the response variable, and the three levels of the model.

- Click on the red y to open the **Y variable** window
- In the **y** drop-down list, select **postthks**
- In the **N levels** drop-down list, select **3-ijk**
- In the **level 3(k)** drop-down list, select **schoolid**
- In the **level 2(j)** drop-down list, select **classid**
- In the **level 1(i)** drop-down list, select **studentid**
- Check that the window matches that shown below and then click **done**

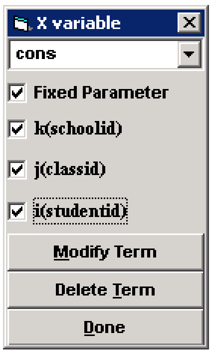


The **Equations** window should update to show **postthks** as the response variable.

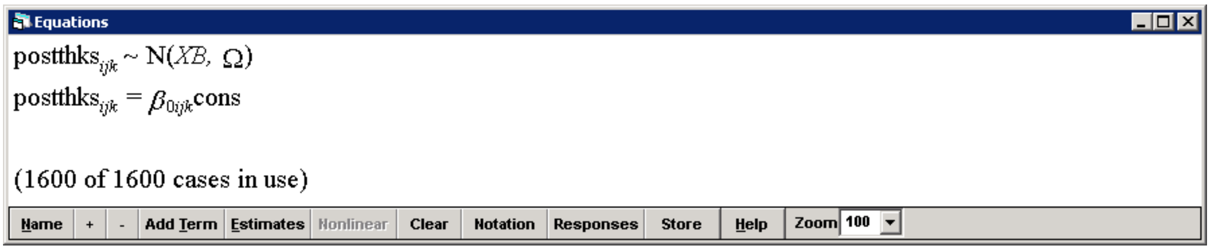


Next, we specify the constant: the variable associated with the intercept coefficient.

- Click on the red $\beta_0 x_0$ term to open the **X variable** window
- From the drop-down list, select **cons**
- Check **k(schoolid)**, **j(classid)** and **i(studentid)**
- Check that the window matches that shown below and then click **Done**

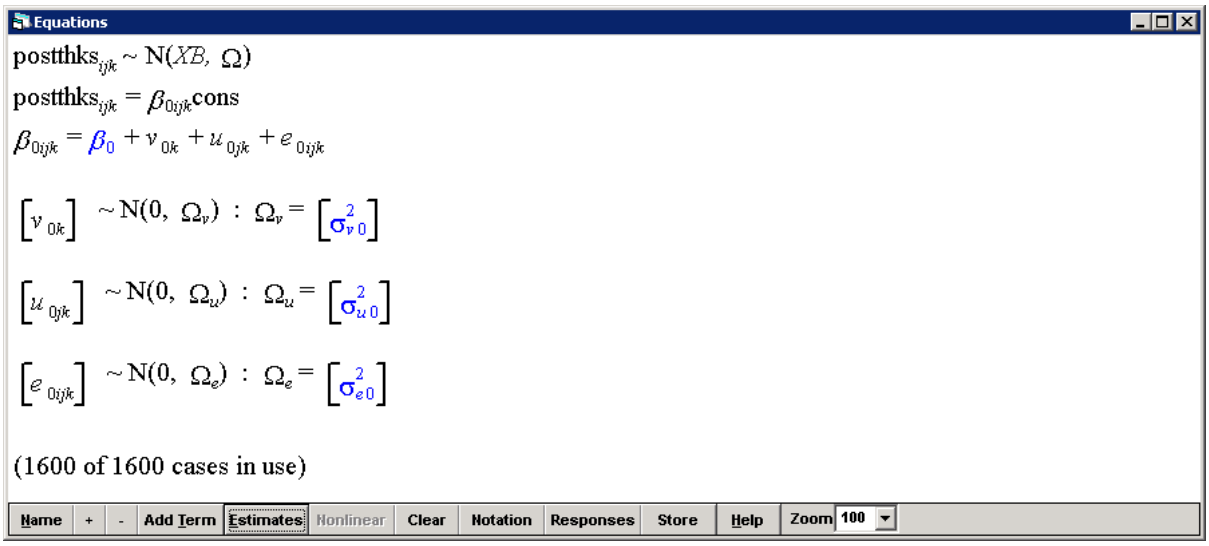


The **Equations** window should update to show that the variable **cons** is associated with the coefficient β_0.



The random effects and distributional assumptions are currently hidden and so next we reveal the full model specification.

- In the **Equations** window toolbar, click **Estimates** once

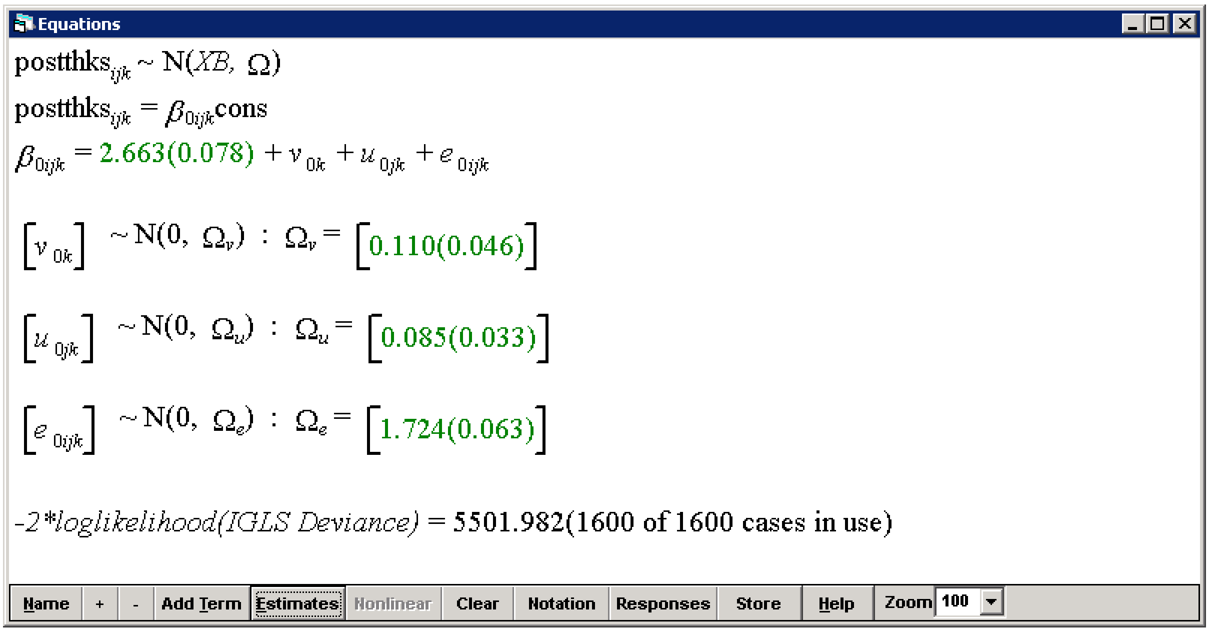


The four parameters to be estimated are highlighted in blue. The first parameter is the intercept, the second parameter is the between-school variance, the third parameter is the within-school-between-classroom variance, and the fourth parameter is the within-classroom-between-student variance.

We now fit the three-level model.

- Click **Start**
- In the **Equations** window toolbar, click **Estimates** once

You should see the following model results. The parameter estimates and standard errors (reported in parentheses) are presented in green (as opposed to the default colouring of blue used when specifying the model) to indicate that the model has successfully converged.

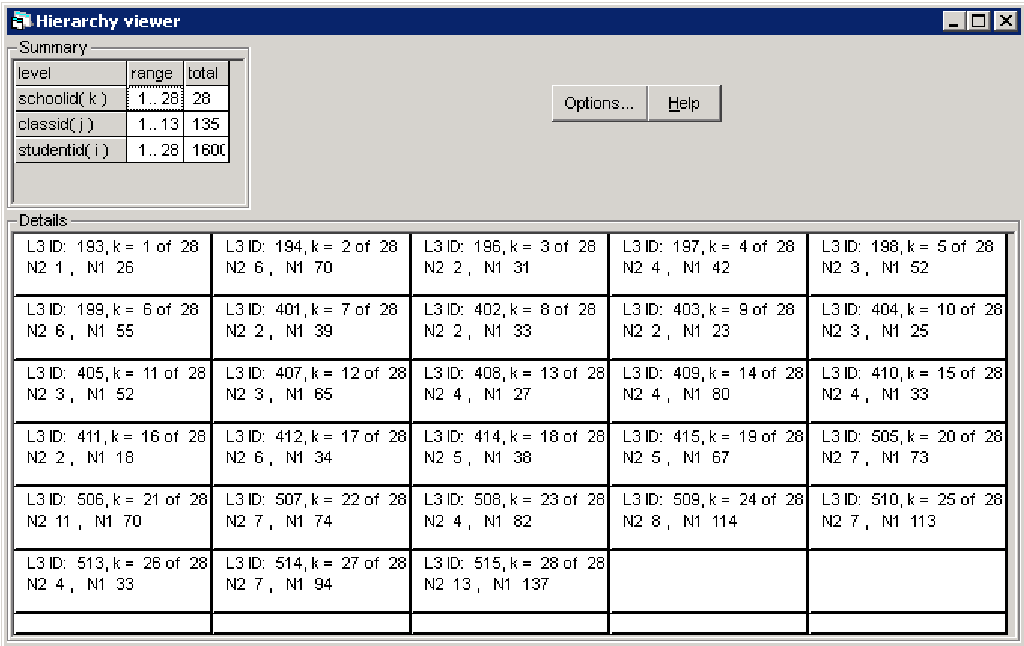


Store the estimation results.

- In the **Equations** window toolbar, click **Store**
- Type ‘**model1**’ into the box
- Click **OK**

Before we interpret the model results, we shall check that the three-level hierarchy assumed by the model matches that found in the data. We can do this by examining the output of the **Hierarchy viewer** window.

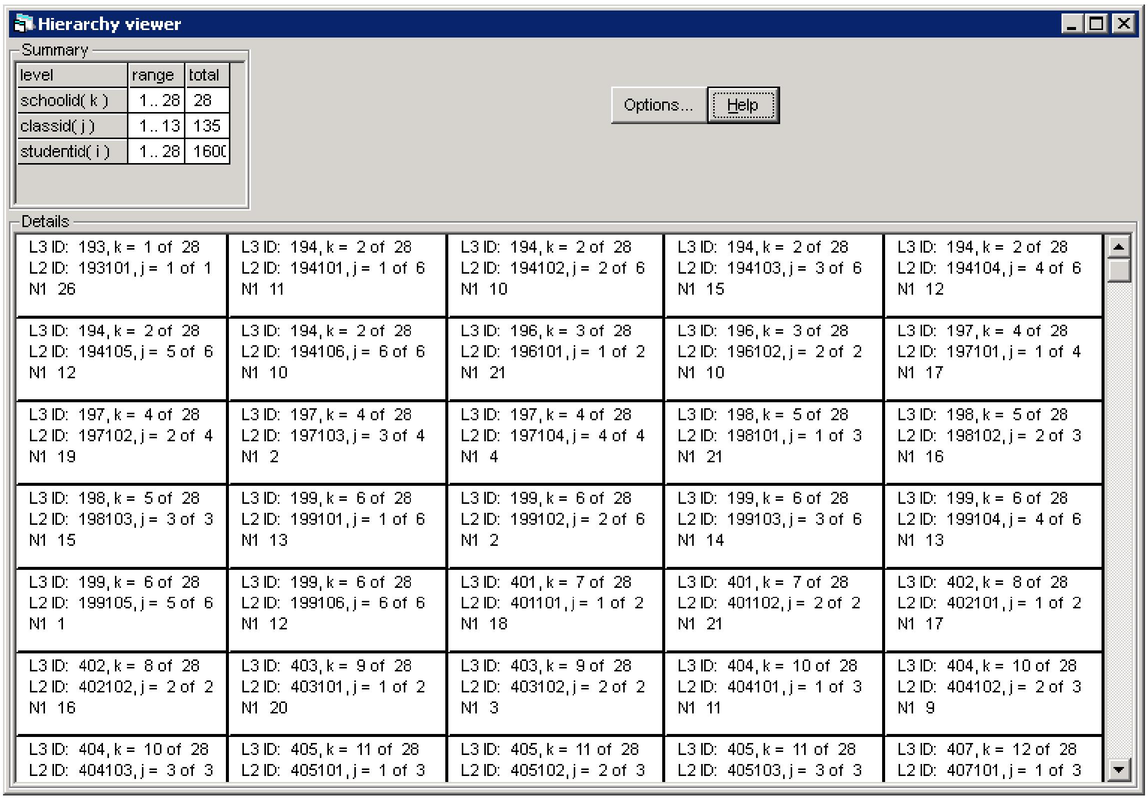
- From the **Model** menu, select **Hierarchy viewer**



The **Summary** section of the window shows that the model correctly identifies that there are 28 schools, 135 classes and 1600 students. The **Details** section of the window reports the number of classes and the number of students per school. We see, for example, that the school 193 (`L3 ID: 193`) has only one class (`N2 1`) and that 26 students (`N1 26`) attend that class.

Where schools have multiple classrooms, we can also check the number of students per class.

- In the **Hierarchy viewer** window, click the **Options…** button
- Select **2:classid** from the **display at level** drop-down list
- Click **Done**



We see, for example, that in school 194 (`L3 ID: 194`) there are 11 students in the first class (`L2 ID: 194101` and `N1 11`) and 10 students in the second class (`L2 ID: 194102` and `N1 10`).

P11.2.2 Interpretation of the model output

The only coefficient in the fixed part of the model is the intercept and this is estimated to be 2.663, with a standard error or 0.078. Thus, at postintervention, the mean student is predicted to score 2.663 out of 7 on the THKS scale. The *z*-ratio for this parameter estimate could be calculated (by dividing the parameter estimate by the standard error), however, as the THKS scale ranges from 0 to 7, it is of no interest to test whether the intercept is significantly different from zero.

Below the fixed part of the model are the estimates for the variance components. The between-school variance is estimated as 0.110, the within-school-between-classroom variance is estimated as 0.085, while and the within-classroom-between-student variance is estimate as 1.724. We shall interpret the relative magnitude of these variances in P11.2.3.

Note that although standard errors are reported for these variances, they should not be used to assess the significance of these parameters (for example, by calculating *z*-ratios and *p*-values). The reason for this is that Wald tests on variance parameters are approximate as they assume that the sampling distributions of these parameters are asymptotically normal when in fact they are positively skewed (i.e. they have long right-hand tails). Likelihood ratio (LR) tests, which do not rely on the assumption of asymptotic normal sampling distributions, should therefore be used to test the significance of variance parameters, not Wald tests.

The final line of output in the **Equations** window reports the deviance statistic (D = 5501.982, calculated as minus two times the log-likelihood). The difference in deviances between two nested models gives the likelihood ratio test statistic for comparing the fit of the two models. For example, the deviance of a simpler single-level model with no school effects and no classroom effects (output not shown) is D = 5577.054. The LR test statistic and associated p-value for testing whether the three-level model is preferred to the single-level model are then: χ_2^2 = 75.07, p < 0.001. The p-value is effectively zero and so the three-level model offers a significantly better fit to the data than the single-level model. We can therefore conclude that the 1,600 students do not act as 1,600 independent observations; rather, students are clustered by classrooms and schools. LR tests (output not reported), which compare this three-level model to the simpler two-level students-within-schools model (χ_1^2 = 11.67, p < 0.001) and the two-level students-within-classrooms model (χ_1^2 = 19.89, p < 0.001), confirm that both the school variance and the classroom variance are separately significant. Students from the same school are therefore significantly more alike than students from different schools. Similarly, students taught in the same classroom are significantly more homogenous than schoolmates taught in two different classrooms. Put differently, the postintervention THKS scores vary significantly across schools and across classrooms. A multilevel approach to analyse the data is clearly favoured over a single-level approach and also over carrying out either of the potential two-level analyses of these data.

P11.2.3 Calculating coverage intervals, variance partition coefficients (VPCs) and intraclass correlation coefficients (ICCs)

There are several approaches to interpreting variance components in multilevel models and we shall consider three of these here: (1) coverage intervals; (2) variance partition coefficients (VPCs); and (3) intraclass correlation coefficients (ICCs). A complete introduction to these approaches is given in C11.2.4.

Coverage intervals

Coverage intervals enable us to interpret the absolute magnitude of variance components in the metric of the response variable. For example, the model implied 95% range in school effects is calculated as

$$(-1.96\sigma_v, +1.96\sigma_v) = \left(-1.96\sqrt{0.110}, +1.96\sqrt{0.110}\right) = (-0.650, +0.650)$$

where the use of ± 1.96 reflects the fact that 95 per cent of the probability mass of a normal distribution lies within approximately ± 1.96 standard deviations of the mean.[3] Thus, the derivation of coverage intervals is based on the model assumption that the random effects are normally distributed. We see that schools at the 97.5$^{\text{th}}$ percentile are estimated to score 1.300 $(= 2 \times 0.650)$ of a standard deviation higher than schools at the 2.5$^{\text{th}}$ percentile.

Variance partition coefficients (VPCs)

Variance partition coefficients (VPCs) report the proportion of the observed response variation that lies at each level of the model hierarchy.[4] They therefore allow us to establish the relative importance of schools, classrooms and students as sources of variation of students' postintervention THKS scores.

The school level VPC is calculated as

$$\text{VPC}_v = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_u^2 + \sigma_e^2} = \frac{0.110}{0.110 + 0.085 + 1.724} = 0.057$$

The classroom level VPC is calculated as

$$\text{VPC}_u = \frac{\sigma_u^2}{\sigma_v^2 + \sigma_u^2 + \sigma_e^2} = \frac{0.085}{0.110 + 0.085 + 1.724} = 0.044$$

[3] In conditional models, coverage intervals are based on the residual rather than the observed responses. Coverage intervals based on conditional models therefore measure the expected range in adjusted outcomes.

[4] In conditional models, VPCs are based on the residual rather than the observed responses. VPCs based on conditional models therefore measure the proportion of outcome variation *unexplained by the predictor variables* that lies at each level of the model hierarchy.

The student level VPC is calculated as

$$\mathrm{VPC}_e = \frac{\sigma_e^2}{\sigma_v^2 + \sigma_u^2 + \sigma_e^2} = \frac{1.724}{0.110 + 0.085 + 1.724} = 0.898$$

We see that 5.7% of the variation in postintervention THKS scores lies between schools, 4.4% lies within schools between classrooms and 89.8% lies within classrooms between students. Thus, there is only modest variation in students' mean tobacco and health knowledge across schools and classrooms; most of the variation in students' knowledge is seen within their classrooms.

Intraclass correlation coefficients (ICCs)

Intraclass correlation coefficients (ICCs) measure the model implied correlation (i.e. similarity or homogeneity) of the observed responses within a given cluster.[5]

The school level ICC is calculated as the correlation between two students i and i' within the same school k, but different classrooms j and j'

$$\begin{aligned}\rho_v &= \mathrm{corr}\left(\mathbf{postthks}_{ijk}, \mathbf{postthks}_{i'j'k}\right) = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_u^2 + \sigma_e^2} \\ &= \frac{0.110}{0.110 + 0.085 + 1.724} = 0.057\end{aligned}$$

Thus, for this model, the school level ICC coincides with the school level VPC. However, this equivalence will not hold in more complex models, such as those including random coefficients.

The classroom level ICC is calculated as the correlation between two students i and i' within the same classroom j and therefore the same school k

$$\begin{aligned}\rho_{vu} &= \mathrm{corr}\left(\mathbf{postthks}_{ijk}, \mathbf{postthks}_{i'jk}\right) = \frac{\sigma_v^2 + \sigma_u^2}{\sigma_v^2 + \sigma_u^2 + \sigma_e^2} \\ &= \frac{0.110 + 0.085}{0.110 + 0.085 + 1.724} = 0.102\end{aligned}$$

Here we see that the classroom level ICC does not coincide with the classroom level VPC. The between-school variance σ_v^2 appears on the numerator of the expression for the classroom level ICC, but does not appear on the numerator of the expression for classroom level VPC.

[5] In conditional models, ICCs are based on the residual rather than the observed responses. ICCs in conditional models therefore measure the similarity in outcomes having adjusted for the predictor variables; that is, the *similarity in unexplained outcomes*. ICCs based on conditional models are sometimes referred to as adjusted ICCs.

The correlation between two students i and i' from different schools k and k' and therefore from different classrooms is assumed zero

$$\rho_{vue} = \mathrm{corr}\left(\mathbf{postthks}_{ijk}, \mathbf{postthks}_{i'j'k'}\right) = 0$$

We see that the school ICC is 0.057, while the classroom ICC is 0.102. Thus, scores on students in the same school are slightly correlated, while scores on students within the same classroom have a somewhat higher correlation. Put differently, students from the same school, or even the same classroom, are not especially similar in their postintervention THKS scores.

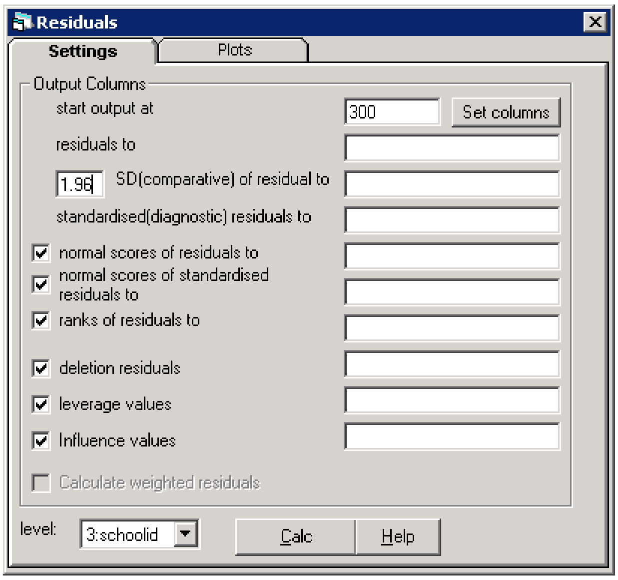
In summary, the VPCs and ICCs show that there is a relatively low degree of clustering in the data. We only see relatively small school level differences in postintervention THKS scores. As we are yet to account for the interventions in our model, this interestingly suggests that the CC and TV interventions do not have particularly strong effects. It is also interesting to note that even with this low degree of clustering the three-level model was significantly preferred to the single-level model and each of its two-level counterparts (See P11.2.2).

P11.2.4 Predicting and examining school and classroom effects

Having fitted the model, we can predict posterior estimates of the school and classroom effects together with their associated standard errors. We can examine these predictions to check whether the random effects at each level are normally distributed. We can also examine them in order to make inferences about specific schools or classrooms.

We can predict the random effects and associated standard errors at both the school- and classroom-level using the **Residuals** window. We start by predicting the school-level residuals.

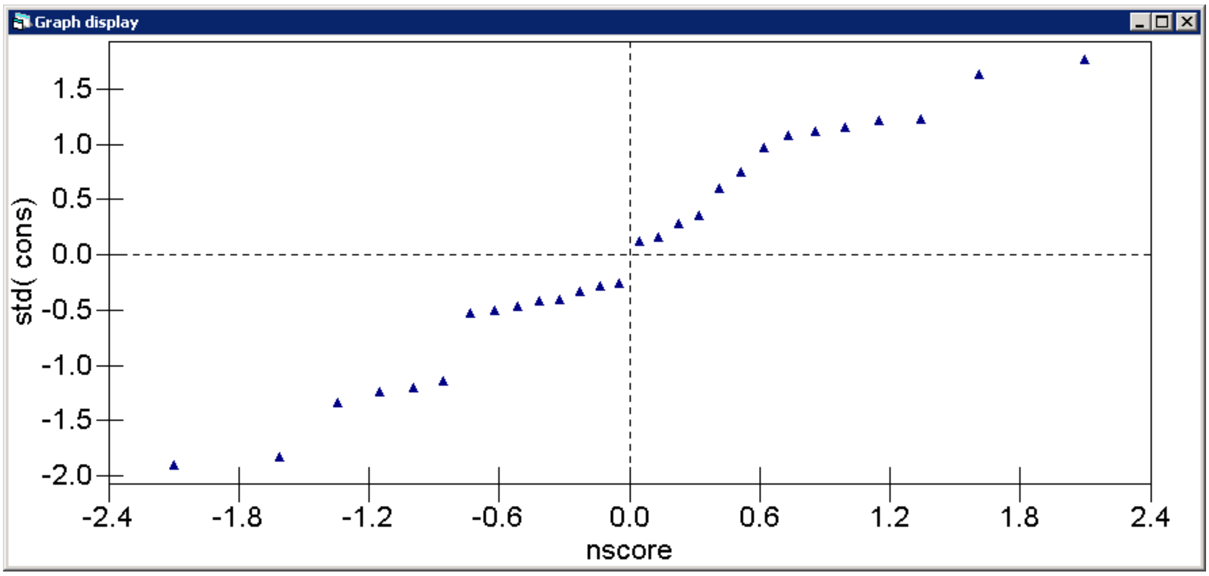
- From the Model menu, select **Residuals**
- In the **level** drop-down list, select **3:schoolid**
- In the box to the left of **SD(comparative) of residual to**, change the number from **1.0** to **1.96**, so that we obtain the 95% confidence intervals
- Check that the window matches that shown below and then click **Calc**



If you look in the **Names** window at this point, you will see that nine new variables have been created in columns **c300** through **c308**.

We first plot a quantile-quantile plot to check whether the school-level random effects are normally distributed.

- Click on the **Plots** tab in the **Residuals** window
- Select **standardised residual x normal scores**
- Click **Apply**

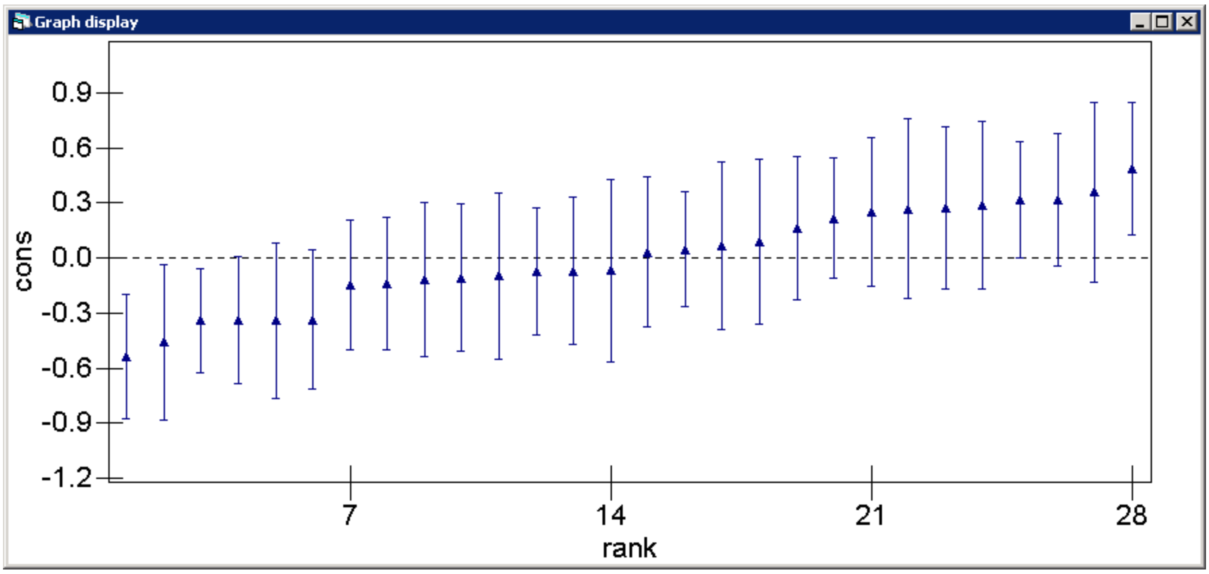


If the random effects are normally distributed, all the data will be plotted along the 45 degree line. While the schools do not lie on the line, they all lie close to the line suggesting that the predicted effects are approximately normally distributed.

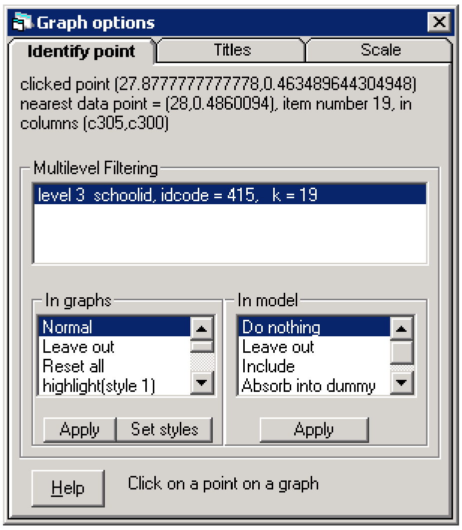
Next we examine the magnitudes of the school effects and we will count how many schools differ significantly from the average school. We will do this by plotting a 'caterpillar plot' of the school effects.

- Return to the **Plots** tab of the **Residuals** window
- Select **residual +/-1.96 sd x rank**
- Click **Apply**

You should see the following plot.



This graph is interactive. By clicking on the points we can identify which schools they refer to. So, for example, clicking on the highest ranked school residual in the plot produces the following window



which shows that the highest scoring school is school 415.

Notice that the confidence intervals around the predicted effects vary greatly in their length; smaller schools (e.g. school 411, ranked 22, which has 18 students) will have longer intervals than larger schools (e.g. school 515, ranked 3, which has 137 students).

The plot shows that only six out of the 28 schools differ significantly from the average school. Four schools (506, 513, 515 and 507) score significantly lower than average, two schools (510 and 415) score significantly higher than average.

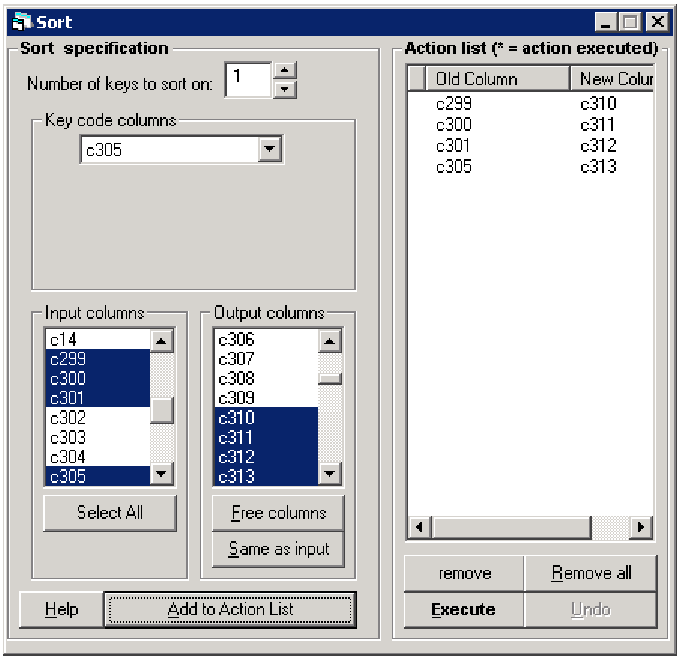
Note that because we have not yet accounted for baseline THKS scores, we cannot interpret these predicted effects in any sense as the effects of schools on students' tobacco and health knowledge. The effects plotted here are very likely to reflect not just school effects, but to also reflect school differences in students' tobacco and health knowledge that were present at baseline and persist through to postintervention (i.e. selection effects).

Next, we rank schools by the magnitude of their predicted effects. First, we use the **UNIQ** command to generate a new 'short' version of the school identifier variable which takes one record per school. The new variable will then appear in the **Names** window.

- From the **Data Manipulation** menu, select **Command interface**
- Type the following command into the bottom pane of the window and then press **Enter**
  `UNIQ 'schoolid' c299`

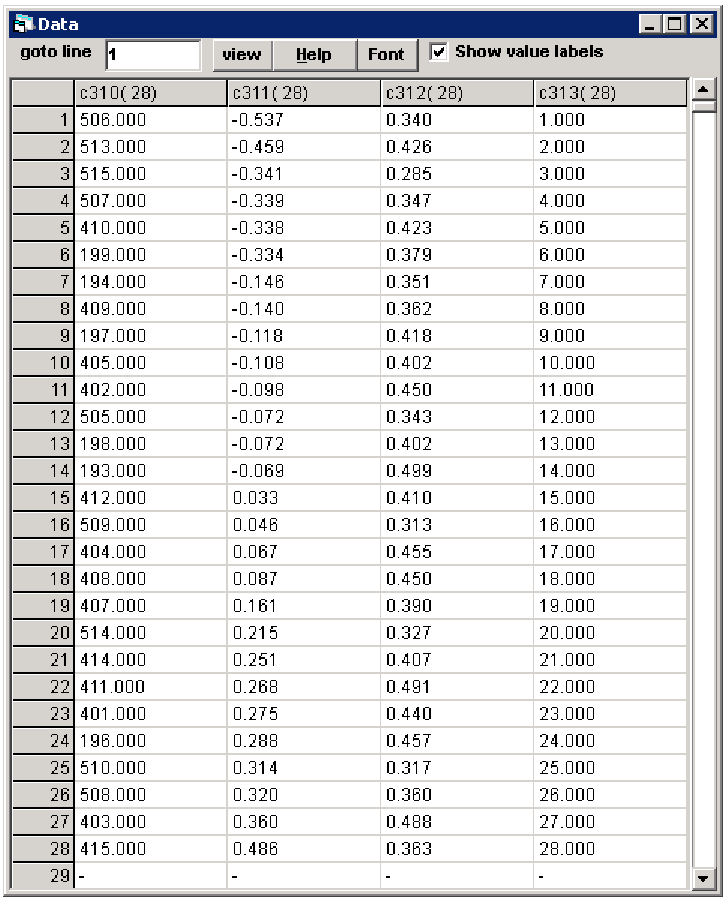
Next, we use the **Sort** window to sort the data by the residual rank variable (which was created with the residuals and is found in **c305**), storing the newly ordered variables in columns **c310-c313**.

- From the **Data Manipulation** menu, select **Sort**
- From the **Key code columns** drop down list, select **c305**
- From the **input columns** variable list, select **c299**, **c300**, **c301** and **c305**
- From the **output columns** variable list, select columns **c310-c313**
- Click **Add to action list**
- Check that the window matches that shown below and then click **Execute**



Finally, view the ranked school effects

- In the **Names** window, tick the **Used columns** checkbox
- Select columns **c310** to **c313**
- From the **Data** toolbar in the **Names** window, click **View**



	c310(28)	c311(28)	c312(28)	c313(28)
1	506.000	-0.537	0.340	1.000
2	513.000	-0.459	0.426	2.000
3	515.000	-0.341	0.285	3.000
4	507.000	-0.339	0.347	4.000
5	410.000	-0.338	0.423	5.000
6	199.000	-0.334	0.379	6.000
7	194.000	-0.146	0.351	7.000
8	409.000	-0.140	0.362	8.000
9	197.000	-0.118	0.418	9.000
10	405.000	-0.108	0.402	10.000
11	402.000	-0.098	0.450	11.000
12	505.000	-0.072	0.343	12.000
13	198.000	-0.072	0.402	13.000
14	193.000	-0.069	0.499	14.000
15	412.000	0.033	0.410	15.000
16	509.000	0.046	0.313	16.000
17	404.000	0.067	0.455	17.000
18	408.000	0.087	0.450	18.000
19	407.000	0.161	0.390	19.000
20	514.000	0.215	0.327	20.000
21	414.000	0.251	0.407	21.000
22	411.000	0.268	0.491	22.000
23	401.000	0.275	0.440	23.000
24	196.000	0.288	0.457	24.000
25	510.000	0.314	0.317	25.000
26	508.000	0.320	0.360	26.000
27	403.000	0.360	0.488	27.000
28	415.000	0.486	0.363	28.000
29	-	-	-	-

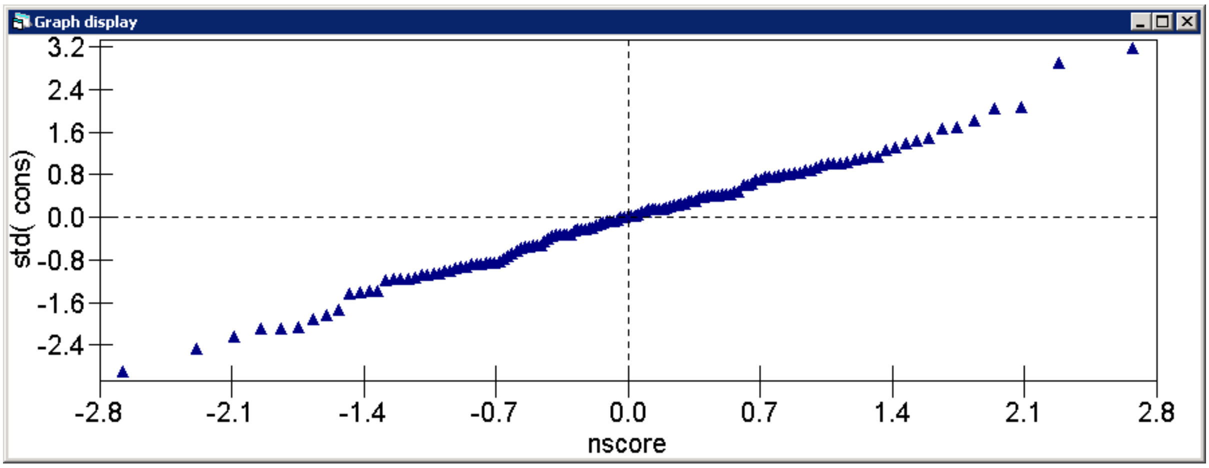
From these values we see that school 506, with a score of -0.537, is predicted to be the lowest scoring school while school 415, with a score of 0.486, is predicted to be the highest scoring school. The difference between the highest and the lowest scoring schools is just over 1 point, which is fairly sizeable given that the THKS scale ranges from 0 to 7.

We can also view the classroom level random effects using a similar approach to that used for the school level random effects. We start by predicting the classroom level residuals.

- From the **Model** menu, select **Residuals**
- In the **Residuals** window, select the **Settings** tab if not already selected
- Change the **level** drop-down list from **3:schoolid** to **2:classid**
- In the **start output at** box, type **'400'**, so as not to overwrite the level 3 residuals and associated terms
- In the box to the left of **SD(comparative) of residual to**, change the number from **1.0** to **1.96**
- Click **Calc**

We generate a quantile-quantile plot to check whether the classroom level residuals are normally distributed.

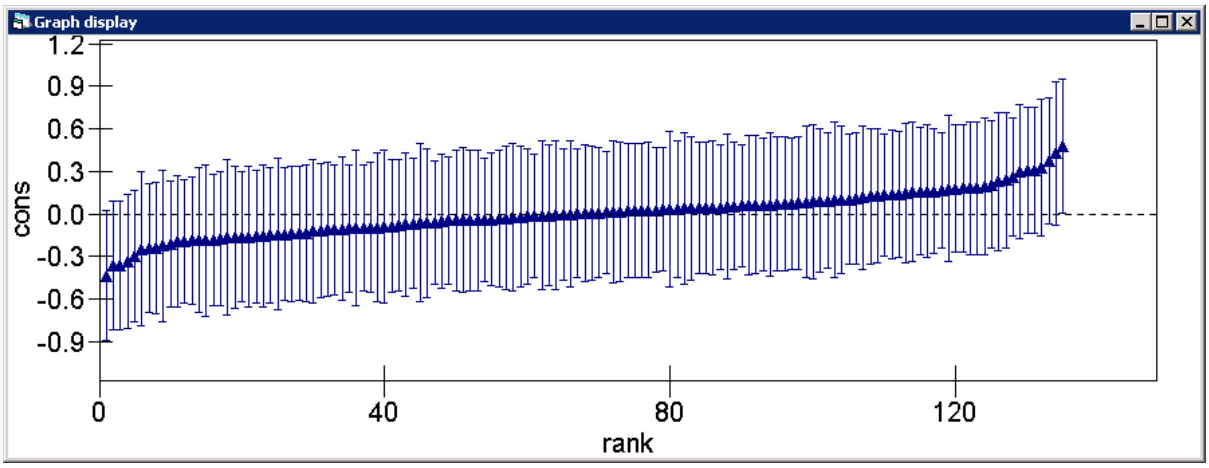
- Click on the **Plots** tab in the **Residuals** window
- Select **standardised residual x normal scores**
- Click **Apply**



The quantile-quantile plot for the classroom effects shows that these effects are also approximately normal, although there is some indication that the distribution of effects has slightly heavier tails than would be expected from a normal distribution.

Finally, we generate the caterpillar plot of the classroom level residuals

- Return to the **Plots** tab of the **Residuals** window
- Select **residual +/-1.96 sd x rank**
- Click **Apply**



We see that the vast majority of classrooms cannot be distinguished from the overall average.

Note that the classroom effects calculated and examined above are net of the effects of the schools in which classrooms are located. In many ways it may be more interesting to calculate and examine the combined school and classroom effect and to consider how many of these differ from the overall average. We leave this as an exercise for the reader.

P11.3 Adding Predictor Variables

Open the worksheet '11.3.wsz'

From within the LEMMA Learning Environment
- Go to **Module 11: Three-Level Multilevel Models**, and scroll down to **MLwiN Datafiles**
- Click '**11.3.wsz**' to open the worksheet

In this lesson, we shall introduce student and school level predictor variables into the three-level model.

P11.3.1 Adding student level predictor variables

We begin by including students' baseline THKS scores (**prethks**). Our exploratory analyses (P11.1.2) revealed that baseline THKS scores vary across the four conditions and so it is essential to adjust for this variable when we come to examine the effects of CC and TV on students' tobacco and health knowledge.

The model is written as

$$\mathbf{postthks}_{ijk} = \beta_0 + \beta_1 \mathbf{prethks}_{ijk} + v_k + u_{jk} + e_{ijk}$$

$$v_k \sim \mathrm{N}(0, \sigma_v^2)$$

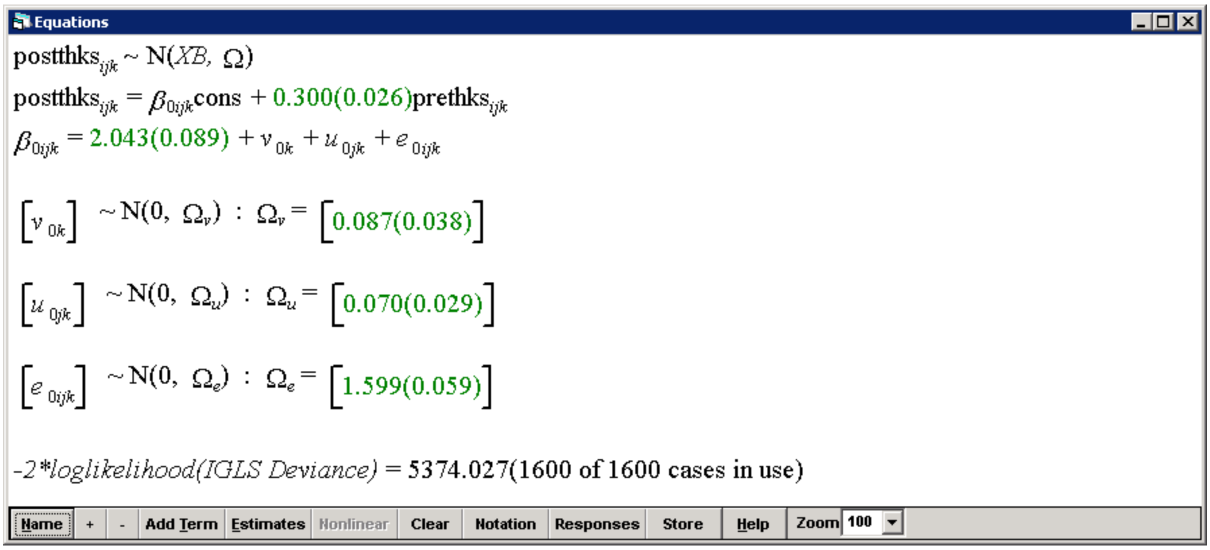
$$u_{jk} \sim \mathrm{N}(0, \sigma_u^2)$$

$$e_{ijk} \sim \mathrm{N}(0, \sigma_e^2)$$

Specify and fit the model.

- From the **Model** menu, select **Equations**
- In the **Equations** window, click **Add Term**
- Select **prethks** from the **variable** drop-down list and click **Done**
- Click **Start**

You should obtain the following results



Store the estimation results.

- In the **Equations** window toolbar, click **Store**
- Type **'model2'** into the box
- Click **OK**

The coefficient on **prethks** is 0.300 and so students who score one point higher at baseline are predicted to score 0.300 points higher postintervention. This effect is highly statistically significant with a z-ratio of 11.54, calculated as the estimate divided by the standard error (11.54 = 0.300/0.026).

Adjusting for baseline THKS scores reduces the three variance parameters. The school level variance drops from 0.110 in the unconditional model to 0.087 in this model, a drop of 20%. The classroom level variance drops from 0.085 to 0.070, a drop of 18%. The student level variance drops from 1.724 to 1.599, a drop of 7%. The large decline in the classroom level variance and in particular the school level variance shows that there are large baseline differences in students' tobacco and health knowledge between classrooms and between schools.

The deviance statistic for this model is D = 5374.027. LR tests (output not shown) confirm that the three-level model is still preferred to its single-level counterpart (χ^2_2 = 63.77, p < 0.001), its two-level students-within-schools counterpart (χ^2_1 = 10.02, p = 0.002), and its students-within-classes counterpart (χ^2_1 = 18.75, p = 0.001). Thus, it is important to retain the school and classroom random effects in the model, even after adjusting for students' baseline scores.

P11.3.2 Adding school level predictor variables

Next we model the effects of the CC and TV interventions. Recall that the two interventions led to four study conditions.

(1) Neither intervention (a no-treatment control group);
(2) CC only;
(3) TV only;
(4) CC and TV.

We will model the effects of these two interventions by including binary indicator variables for CC and TV (**cc** and **tv**) along with their interaction (**ccXtv**). The model is written as

$$\mathbf{postthks}_{ijk} = \beta_0 + \beta_1\mathbf{prethks}_{ijk} + \beta_2\mathbf{cc}_k + \beta_3\mathbf{tv}_k + \beta_4\mathbf{ccXtv}_k + v_k + u_{jk} + e_{ijk}$$

$$v_k \sim \mathrm{N}(0, \sigma_v^2)$$

$$u_{jk} \sim \mathrm{N}(0, \sigma_u^2)$$

$$e_{ijk} \sim \mathrm{N}(0, \sigma_e^2)$$

It is helpful to write out the fixed part of the model separately for the four conditions

Neither intervention: $\beta_0 + \beta_1\mathbf{prethks}_{ijk}$

CC only: $\beta_0 + \beta_1\mathbf{prethks}_{ijk} + \beta_2\mathbf{cc}_k$

TV only: $\beta_0 + \beta_1\mathbf{prethks}_{ijk} \qquad + \beta_3\mathbf{tv}_k$

CC and TV: $\beta_0 + \beta_1\mathbf{prethks}_{ijk} + \beta_2\mathbf{cc}_k + \beta_3\mathbf{tv}_k + \beta_4\mathbf{ccXtv}_k$

The following table summarises four effects of interest and how they are obtained from the model parameters.

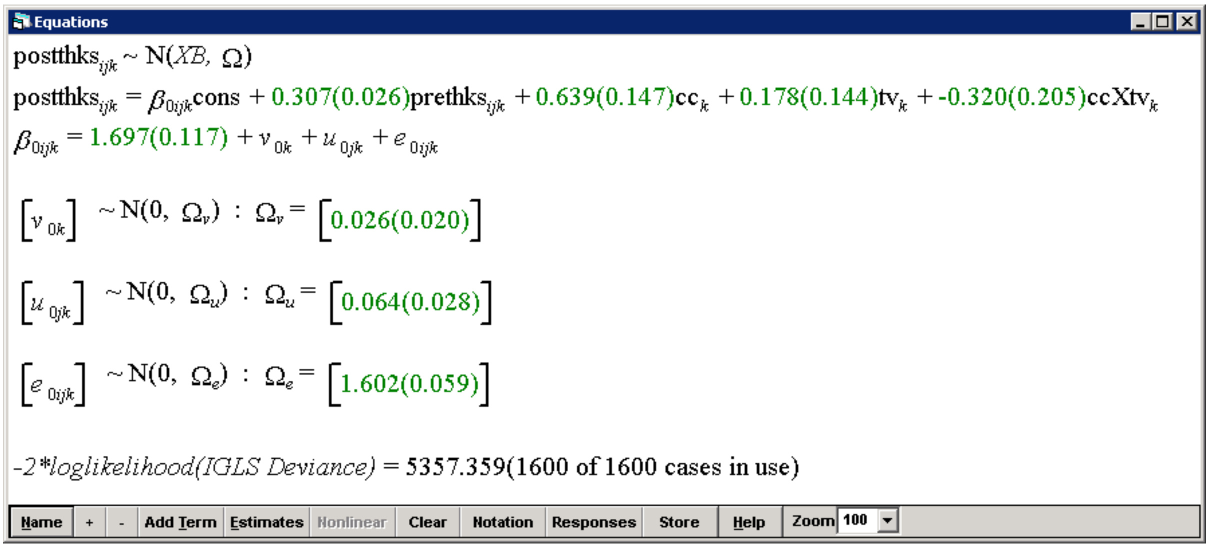
Parameter	Interpretation
β_2	effect of CC on non-TV students
$\beta_2 + \beta_4$	effect of CC on TV students
β_3	effect of TV on non-CC students
$\beta_3 + \beta_4$	effect of TV on CC students

The interaction coefficient β_4 can be interpreted in two ways. First, it can be interpreted as the difference between the effect of CC on TV students and the effect of CC on non-TV students. Second, it can be interpreted as the difference between the effect of TV on CC students and the effect of TV on non-CC students.

Specify and fit the model.

- In the **Equations** window, click **Add Term** and select **cc** from the **variable** drop-down list, then click **Done**
- Repeat this process to add **tv** and then **ccXtv** to the model
- Click **Start**

You should obtain the following results.

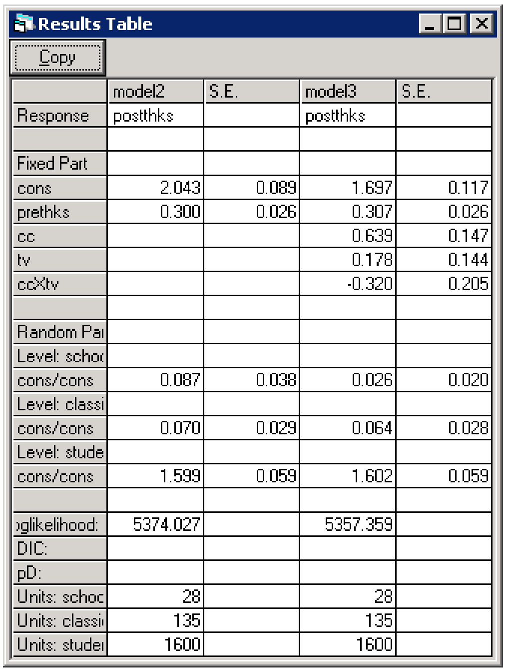


Store the estimation results.

- In the **Equations** window toolbar, click **Store**
- Type **'model3'** into the box
- Click **OK**

We can perform an LR test to confirm that the additional predictors significantly improve the fit of the model.

- From the **Model** menu, select **Manage stored models**
- Select both **model2** and **model3**
- Click **Compare**

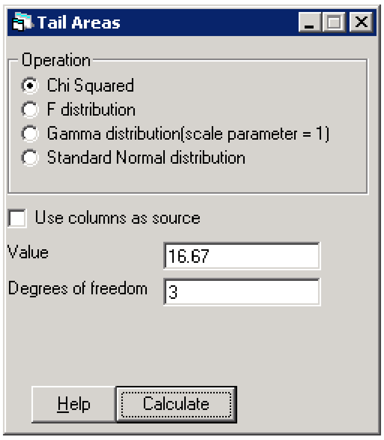


	model2	S.E.	model3	S.E.
Response	postthks		postthks	
Fixed Part				
cons	2.043	0.089	1.697	0.117
prethks	0.300	0.026	0.307	0.026
cc			0.639	0.147
tv			0.178	0.144
ccXtv			-0.320	0.205
Random Pai				
Level: schoo				
cons/cons	0.087	0.038	0.026	0.020
Level: classi				
cons/cons	0.070	0.029	0.064	0.028
Level: stude				
cons/cons	1.599	0.059	1.602	0.059
›glikelihood:	5374.027		5357.359	
DIC:				
pD:				
Units: schoo	28		28	
Units: classi	135		135	
Units: studer	1600		1600	

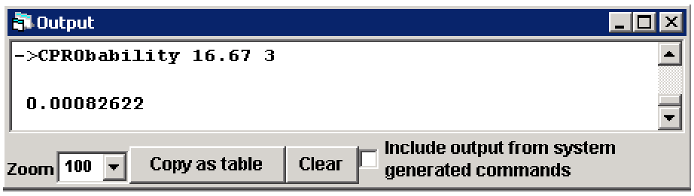
The **Results Table** reports a deviance statistic of 5374.027 for Model 2 and a deviance statistic of 5357.359 for Model 3. The LR test statistic (difference in deviances) is 16.67. We can calculate the p-value associated with this test statistic using the **Tail Areas** window.

- From the **Basic Statistics** menu, select **Tail Areas**
- Select **Chi Squared**
- Type '**16.67**' next to **Value**

- Type '**3**' next to **Degrees of freedom**
- Check that the window matches that shown below and then click **Calculate**



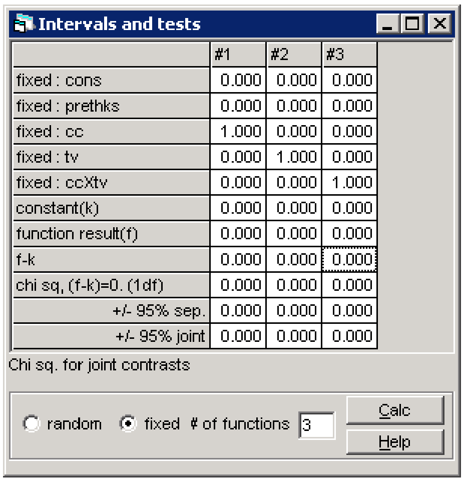
You should see the following output



The p-value is effectively zero (χ^2_3 = 16.67, p < 0.001). The LR test therefore confirms that that the additional predictors significantly improve the fit of the model.

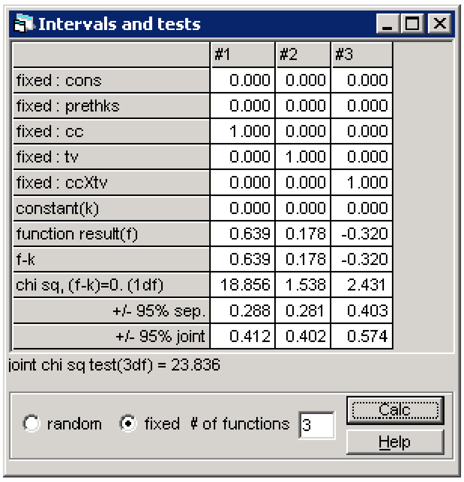
Note that we could have instead used a Wald test to confirm the joint significance of the three school level variables.[6]

- From the **Model** menu, select **Intervals and tests**
- Select the **fixed** radio button
- Type **3** next to **# of functions**
- Replace the **0** with **1** in the **fixed: cc** row of the **#1** column
- Replace the **0** with **1** in the **fixed: tv** row of the **#2** column
- Replace the **0** with **1** in the **fixed: cc*tv** row of the **#3** column
- Check that the window matches that shown below and then click **Calc**



[6] Whether one uses LR or Wald tests to test fixed part parameters is largely a matter of personal preference. The two tests are asymptotically equivalent to one another, although they may produce different conclusions in small samples. Note though that LR tests of fixed part parameters cannot be based on RIGLS estimates.

You should obtain the following results.


Intervals and tests

	#1	#2	#3
fixed : cons	0.000	0.000	0.000
fixed : prethks	0.000	0.000	0.000
fixed : cc	1.000	0.000	0.000
fixed : tv	0.000	1.000	0.000
fixed : ccXtv	0.000	0.000	1.000
constant(k)	0.000	0.000	0.000
function result(f)	0.639	0.178	-0.320
f-k	0.639	0.178	-0.320
chi sq, (f-k)=0. (1df)	18.856	1.538	2.431
+/- 95% sep.	0.288	0.281	0.403
+/- 95% joint	0.412	0.402	0.574

joint chi sq test(3df) = 23.836

random fixed # of functions 3 Calc Help

The parameters **cc**, **tv** and **ccXtv** are jointly significant (χ^2_3 = 23.84, p < 0.001).

Adjusting for the CC and TV main effects and their interaction reduces the between-school variance from 0.087 to 0.026, a drop of 70%. Thus, 70% of the variation in tobacco and health knowledge between schools, having accounted for baseline differences in students' knowledge, is attributable to the implementation of the CC and TV interventions. Interestingly, the fact that 30% of the between-school variation is not explained by the implementation of the interventions suggests that there are other unaccounted for school level factors which are leading to differences between schools in students' knowledge.

Adjusting for the CC and TV main effects and their interaction has minimal impact on the magnitude of the classroom level and student level variances and this is to be expected as **cc**, **tv** and **ccXtv** are school level predictors.

We now turn our attention to the four effects of interest.

Parameter	**Interpretation**	**Estimate**
β_2	effect of CC on non-TV students	0.639
$\beta_2 + \beta_4$	effect of CC on TV students	0.319 = 0.639 -0.320
β_3	effect of TV on non-CC students	0.178
$\beta_3 + \beta_4$	effect of TV on CC students	-0.142 = 0.178 -0.320

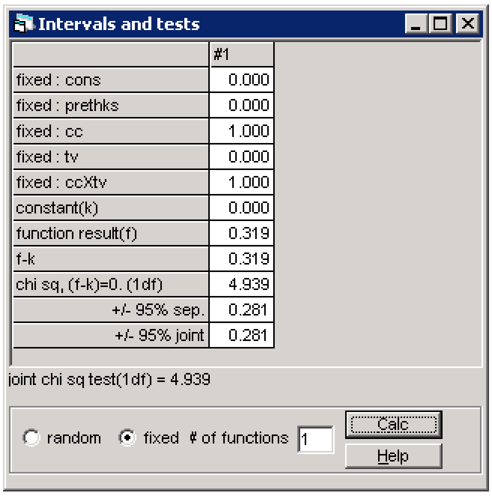
The results suggest that the CC intervention has a positive effect on students, irrespective of whether they receive the TV intervention (0.319) or not (0.639). The effect of the TV intervention, however, is less clear. The results suggest that the TV intervention has a positive effect on students who do not receive the CC intervention (0.178), but a negative effect on students who received the CC intervention (-0.142). These results agree with those seen in P11.1.2.

Importantly, before we draw any firm conclusions, we must check the separate significance of these four inferences. That is, while the above joint significance test reveals that the four conditions have significantly different effects from one another, we must additionally employ separate significance tests to establish whether each of the four effects of interest are individually significant. We can use the **Intervals and Tests** window to formally test the individual significance of these effects. For the first and third effects which each involve only one parameter, we have already calculated individual Wald test statistics as part of the previous joint Wald test reported above. The effect of CC on non-TV students was calculated as (β_2 = 0.639, χ_1^2 = 18.856, p < 0.001), while the effect of TV on non-CC students was calculated as (β_3 = 0.178, χ_1^2 = 1.538, p < 0.215). However, in order to calculate the effect of CC on TV students ($\beta_2 + \beta_4$) and the effect of TV on CC students ($\beta_3 + \beta_4$) we must conduct two further tests.

First, consider the effect of CC on TV students.

- From the **Model** menu, select **Intervals and tests**
- Select the **fixed** radio button
- Type **1** next to **# of functions**
- Replace the **0** with **1** in the **fixed: cc** row of the **#1** column
- Replace the **0** with **1** in the **fixed: ccXtv** row of the **#1** column
- Click **Calc**

You should obtain the results shown below.



	#1
fixed : cons	0.000
fixed : prethks	0.000
fixed : cc	1.000
fixed : tv	0.000
fixed : ccXtv	1.000
constant(k)	0.000
function result(f)	0.319
f-k	0.319
chi sq, (f-k)=0. (1df)	4.939
+/- 95% sep.	0.281
+/- 95% joint	0.281

joint chi sq test(1df) = 4.939

random fixed # of functions 1 Calc Help

As expected, the combination of **cc** and **ccXtv** ($\beta_2 + \beta_4$ = 0.319) is significant (χ_2^2 = 4.939, p = 0.026). We can repeat the above steps to test the effect of TV on CC students and doing this shows this effect ($\beta_3 + \beta_4$ = -0.142) to be not significant (χ_2^2 = 0.937, p = 0.333).

In sum, the results reveal that both CC effects are significant and that both TV effects are not significant. Thus, the CC intervention appears to be an effective means of improving students' tobacco and health knowledge. The TV intervention, on the other hand, has no discernible effect.

What happens if we fail to account for baseline THKS scores?
Earlier we stressed the importance of adjusting for baseline differences in THKS scores. It is interesting to consider what we would have concluded had we failed to do so. We therefore refit the previous model excluding **prethks**.

$$\mathbf{postthks}_{ijk} = \beta_0 + \beta_1\mathbf{cc}_k + \beta_2\mathbf{tv}_k + \beta_3\mathbf{ccXtv}_k + v_k + u_{jk} + e_{ijk}$$

$$v_k \sim \mathrm{N}(0, \sigma_v^2)$$

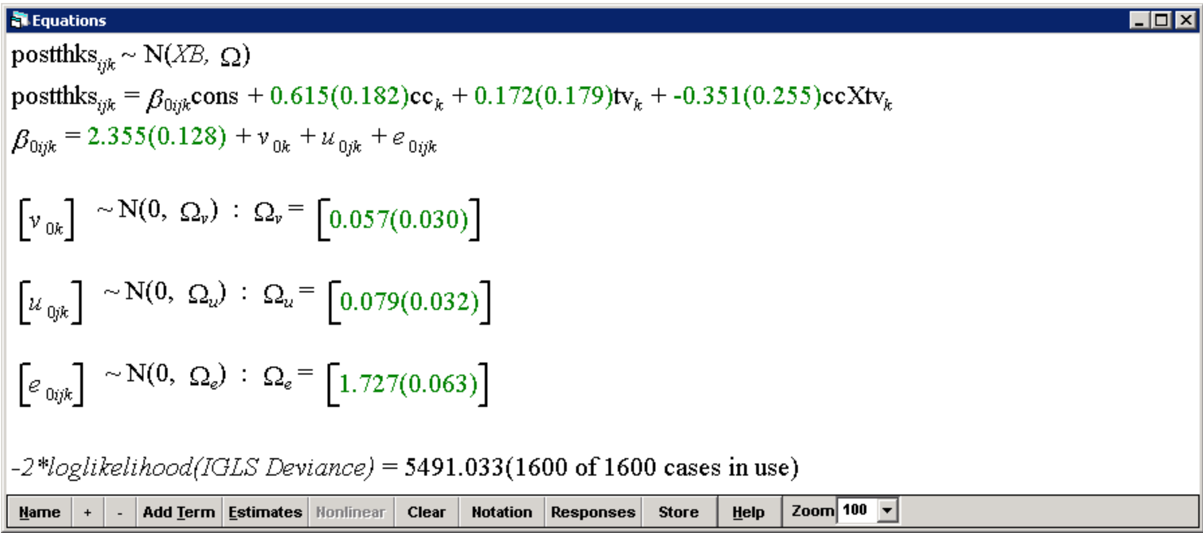
$$u_{jk} \sim \mathrm{N}(0, \sigma_u^2)$$

$$e_{ijk} \sim \mathrm{N}(0, \sigma_e^2)$$

Fit the model.

- In the **Equations** window, click on the **prethks** term to open the **X variable** window
- Click **Delete Term**
- Click **Start**

You should obtain the following results.



Store the estimation results.

- In the **Equations** window toolbar, click **Store**
- Type '**model4**' into the box
- Click **OK**

We could use the **Intervals and tests** window to again test the significance of the four effects of interest: CC on non-TV students (β_1 = 0.615, χ_1^2 = 11.365, p < 0.001); CC on TV students ($\beta_1 + \beta_3$ = 0.264, χ_2^2 = 2.211, p < 0.137); TV on non-CC students (β_2 = 0.172, χ_1^2 = 0.923 , p < 0.337); and TV on CC students ($\beta_2 + \beta_3$ = -0.179, χ_2^2 = 0.975, p < 0.324). The magnitudes of the effects are broadly similar to before. This is expected since the randomisation of schools to the four study conditions should mean that there is no meaningful association between the conditions and baseline THKS scores. However, we no longer find the effect of the CC intervention on students who also receive the TV intervention to be significant. What we can conclude from this is that failing to adjust for students' baseline scores will affect our conclusions about the effectiveness of the interventions and this is despite the fact that students were effectively randomly assigned to the four different conditions. One reason why the effects in the previous model are less precise than the effects in the current model is because adjusting for students' baseline scores reduces the residual school variation and therefore lead to more precise estimates of any school level variables included in the model.

P11.4 Adding Random Coefficients

Open the worksheet '11.4.wsz'

From within the LEMMA Learning Environment

- Go to **Module 11: Three-Level Multilevel Models**, and scroll down to **MLwiN Datafiles**
- Click **'11.4.wsz'** to open the worksheet

In this lesson, we shall introduce classroom level random coefficients into the three-level model.

P11.4.1 Adding classroom level random coefficients

While it is schools which were randomly assigned to the four conditions, the implementation of the CC and TV interventions was carried out at the classroom level. Our inclusion of classroom level effects in all our models recognises that some classrooms may be more successful at implementing the interventions than others. The classroom level variance provides a measure of the extent to which classrooms vary in this respect. However, what all our previous models have implicitly assumed is that the extent to which classrooms vary is the same across all four conditions. There are, however, good reasons to expect this not to be the case. For example, one might expect it to be easier for teachers to implement the TV intervention in their classrooms than it is to implement the CC intervention. If this is the case then we might expect TV only teachers to vary less in their ability to implement their intervention than CC only teachers vary in their ability to implement their intervention. Such a scenario would likely be reflected in students' scores. We would expect students in TV only classrooms to have less heterogeneous postintervention scores than the students in CC only classrooms. We can explore this hypothesis by estimating separate classroom level variances for each of the four conditions.

First we need to generate a series of binary indicator variables for the four study conditions.

- 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

```
CALC c10 = ('cc'==0 & 'tv'==0)
CALC c11 = ('cc'==1 & 'tv'==0)
CALC c12 = ('cc'==0 & 'tv'==1)
CALC c13 = ('cc'==1 & 'tv'==1)
NAME c10 'neither'
NAME c11 'cc_only'
NAME c12 'tv_only'
NAME c13 'cc_and_tv'
```

The model is written as

$$\mathbf{postthks}_{ijk} = \beta_0 + \beta_1\mathbf{prethks}_{ijk} + \beta_2\mathbf{cc}_k + \beta_3\mathbf{tv}_k + \beta_4\mathbf{ccXtv}_k$$

$$+v_k + u_{5jk}\mathbf{neither}_k + u_{6jk}\mathbf{cc\_only}_k + u_{7jk}\mathbf{tv\_only}_k$$

$$+u_{8jk}\mathbf{cc\_and\_tv}_k + e_{ijk}$$

$$v_k \sim \mathrm{N}(0, \sigma_v^2)$$

$$\begin{pmatrix} u_{5jk} \\ u_{6jk} \\ u_{7jk} \\ u_{8jk} \end{pmatrix} \sim \mathrm{N}\left\{\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{u5}^2 & & & \\ 0 & \sigma_{u6}^2 & & \\ 0 & 0 & \sigma_{u7}^2 & \\ 0 & 0 & 0 & \sigma_{u8}^2 \end{pmatrix}\right\}$$

$$e_{ijk} \sim \mathrm{N}(0, \sigma_e^2)$$

The four sets of classroom effects u_{5jk}, u_{6jk}, u_{7jk} and u_{8jk} are modelled as independent (the different sets of effects are not allowed to covary) as each classroom can only belong to one of the four study conditions.

Note that we have only entered the variables **neither**, **cc_only**, **tv_only** and **cc_and_tv** into the level 2 random part of the model. They do not appear in the fixed part of the model. This is because we have already accounted for the mean differences between the four conditions through the inclusion of the constant and the variables **cc**, **tv** and **ccXtv**. The resulting level 2 variance function is

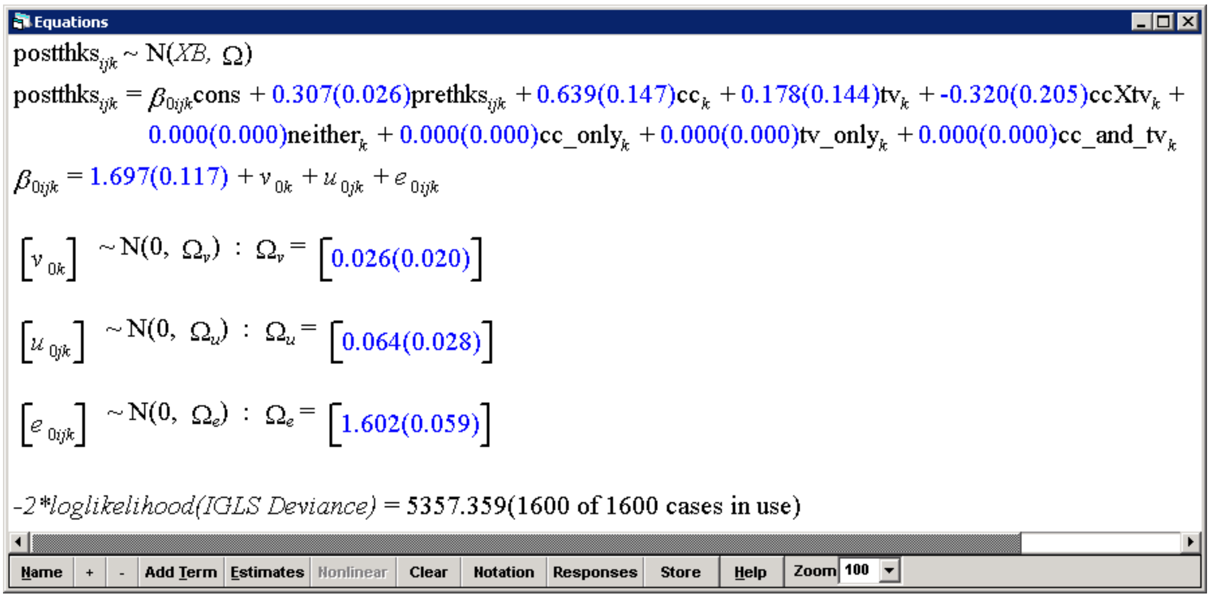
$$\mathrm{var}\left(u_{5jk}\mathbf{neither}_k + u_{6jk}\mathbf{cc\_only}_k + u_{7jk}\mathbf{tv\_only}_k + u_{8jk}\mathbf{cc\_and\_tv}_k\right)$$

$$= \sigma_{u5}^2\mathbf{neither}_k + \sigma_{u6}^2\mathbf{cc\_only}_k + \sigma_{u7}^2\mathbf{tv\_only}_k + \sigma_{u8}^2\mathbf{cc\_and\_tv}_k$$

Add the four new variables to the model

- From the **Model** menu, select **Equations**
- Click on **Add Term**, and select **neither** from the **variable** drop-down list and then click **Done**
- Repeat this process to add **cc_only**, **tv_only** and **cc_and_tv** to the model

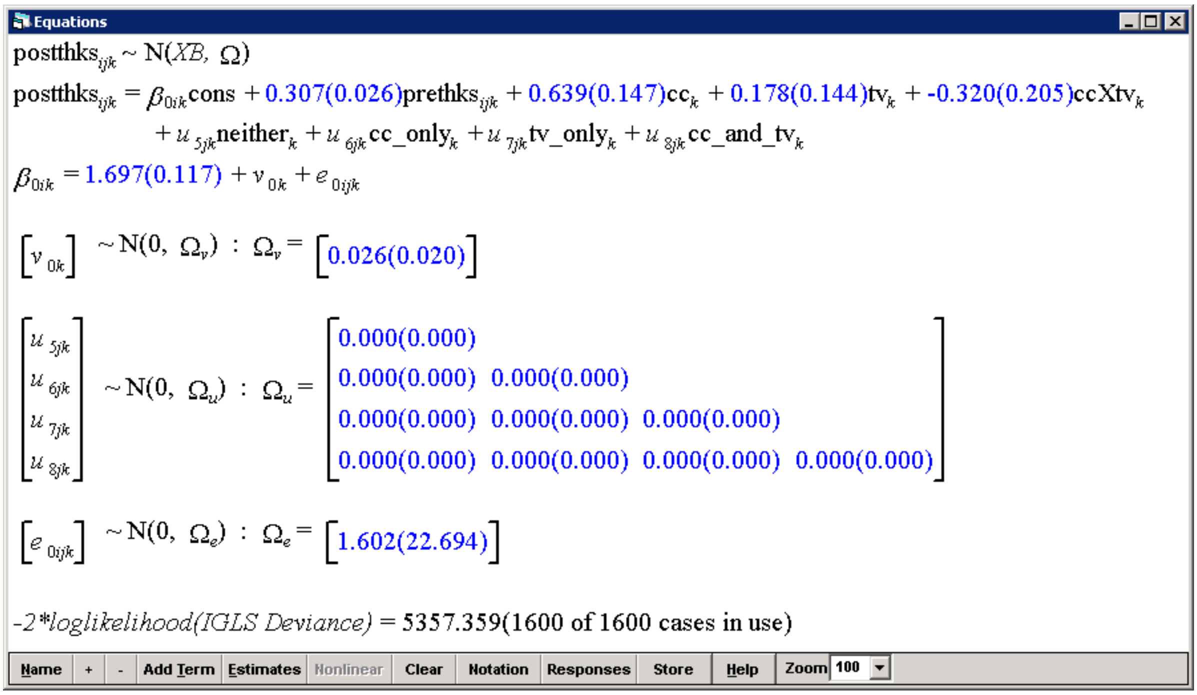
The **Equations** window should now match that shown below.



Remove the four new variables from the fixed part of the model and enter them instead into the random part of the model at the classroom level.

- Click on **cons**, uncheck the **j(classid)** checkbox and then click **Done**
- Click on **neither**, uncheck the **Fixed Parameter** checkbox, tick the **j(classid)** checkbox and then click **Done**
- Click on **cc_only**, uncheck the **Fixed Parameter** checkbox, tick the **j(classid)** checkbox and then click **Done**
- Click on **tv_only**, uncheck the **Fixed Parameter** checkbox, tick the **j(classid)** checkbox and then click **Done**
- Click on **cc_and_tv**, uncheck the **Fixed Parameter** checkbox, tick the **j(classid)** checkbox and then click **Done**

The **Equations** window should now match that shown below.



Make the classroom level variance-covariance matrix diagonal so that the four classroom effects are modelled as independent. Then fit the model.

- Click $\boldsymbol{\Omega}_u$ and then select **set diagonal matrix**
- Click **Start**

You should obtain the following results.

$$\text{postthks}_{ijk} \sim \text{N}(XB, \Omega)$$

$$\text{postthks}_{ijk} = \beta_{0ik}\text{cons} + 0.308(0.026)\text{prethks}_{ijk} + 0.643(0.145)\text{cc}_k + 0.162(0.149)\text{tv}_k + -0.306(0.204)\text{ccXtv}_k + u_{5jk}\text{neither}_k + u_{6jk}\text{cc\_only}_k + u_{7jk}\text{tv\_only}_k + u_{8jk}\text{cc\_and\_tv}_k$$

$$\beta_{0ik} = 1.695(0.116) + v_{0k} + e_{0ijk}$$

$$\begin{bmatrix} v_{0k} \end{bmatrix} \sim \text{N}(0,\ \Omega_v) \ : \ \Omega_v = \begin{bmatrix} 0.025(0.019) \end{bmatrix}$$

$$\begin{bmatrix} u_{5jk} \\ u_{6jk} \\ u_{7jk} \\ u_{8jk} \end{bmatrix} \sim \text{N}(0,\ \Omega_u) \ : \ \Omega_u = \begin{bmatrix} 0.065(0.052) & & & \\ 0 & 0.050(0.050) & & \\ 0 & 0 & 0.111(0.066) & \\ 0 & 0 & 0 & 0.027(0.044) \end{bmatrix}$$

$$\begin{bmatrix} e_{0ijk} \end{bmatrix} \sim \text{N}(0,\ \Omega_e) \ : \ \Omega_e = \begin{bmatrix} 1.602(0.059) \end{bmatrix}$$

$$-2{*}loglikelihood(IGLS\ Deviance) = 5356.220(1600 \text{ of } 1600 \text{ cases in use})$$

Store the estimation results.

- In the **Equations** window toolbar, click **Store**
- Type '**model5**' into the box
- Click **OK**

An LR test (output not shown) comparing this model (Model 5) to one which assumes a constant classroom level variance across the four conditions (Model 3) strongly rejects the current model (χ^2_3 = 1.14, p = 0.7678). Thus, we find no evidence that classroom level heterogeneity varies across the study conditions.

P11.4.2 Adding cross-level interactions

The previous models all assumed that the CC and TV interventions have the same effect on all students, irrespective of the level of their prior tobacco and health knowledge. It may well be the case, however, that the effectiveness of each of these interventions is a function of students' baseline THKS scores. That is, perhaps the CC intervention is relatively more effective for students with high prior knowledge, while the TV intervention might be relatively more effective for students with low prior knowledge. We can explore such hypotheses by introducing into our model cross-level interaction variables between the three school level variables **cc**, **tv** and **ccXtv** and the single student level variable **prethks**.

The model, including these interactions, is written as

$$
\begin{aligned}
\mathbf{postthks}_{ijk} = {} & \beta_0 + \beta_1 \mathbf{prethks}_{ijk} \\
& + \beta_2 \mathbf{cc}_k + \beta_3 \mathbf{tv}_k + \beta_4 \mathbf{ccXtv}_k \\
& + \beta_5 \mathbf{ccXprethks}_{ijk} + \beta_6 \mathbf{tvXprethks}_{ijk} + \beta_7 \mathbf{ccXtvXprethks}_{ijk} \\
& + v_k + u_{jk} + e_{ijk}
\end{aligned}
$$

$$v_k \sim \mathrm{N}(0, \sigma_v^2)$$

$$u_{jk} \sim \mathrm{N}(0, \sigma_u^2)$$

$$e_{ijk} \sim \mathrm{N}(0, \sigma_e^2)$$

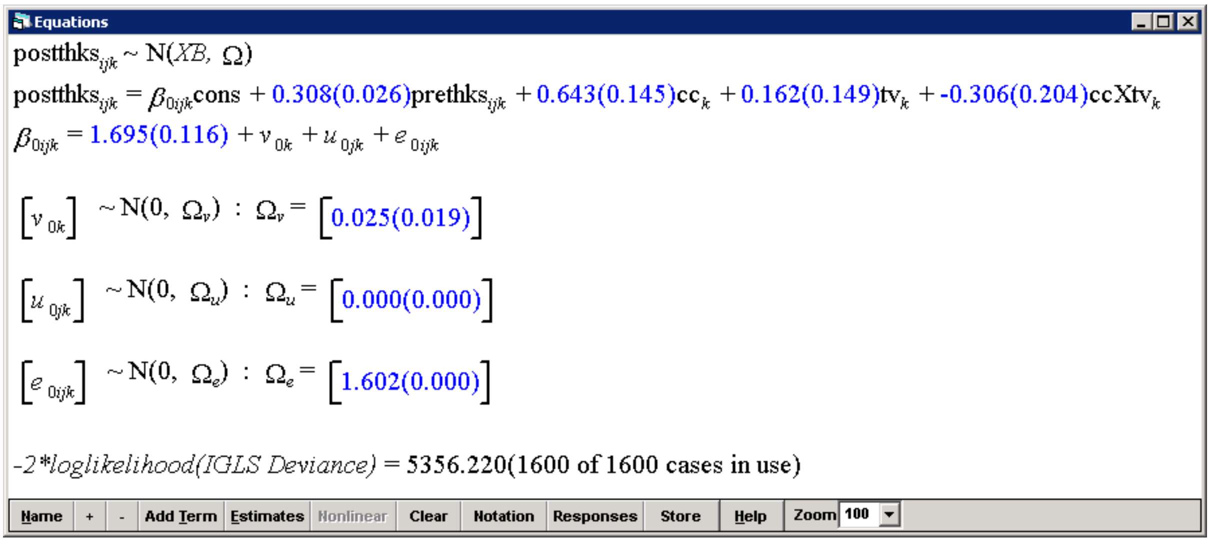
The following table summarises the four effects of interest and how they are obtained from the model parameters.

Parameter	Interpretation
$\beta_2 + \beta_5 \mathbf{ccXprethks}_{ijk}$	effect of CC on non-TV students
$\beta_2 + \beta_4 + \beta_5 \mathbf{ccXprethks}_{ijk} + \beta_7 \mathbf{ccXtvXprethks}_{ijk}$	effect of CC on TV students
$\beta_3 + \beta_6 \mathbf{tvXprethks}_{ijk}$	effect of TV on non-CC students
$\beta_3 + \beta_4 + \beta_6 \mathbf{tvXprethks}_{ijk} + \beta_7 \mathbf{ccXtvXprethks}_{ijk}$	effect of TV on CC students

Notice that the magnitude of each of these effects is now a function of **prethks**.

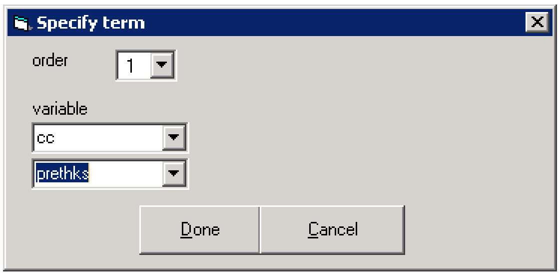
Begin to specify the model by first reverting back to a simple random intercept at the classroom level.

- In the **Equations** window, click **neither** and then click **Delete term**
- Delete the terms **cc_only**, **tv_only** and **cc_and_tv** in the same way
- Click on the **cons** and ensure all the checkboxes are ticked then click **Done**



Next add in the cross-level interaction between **cc** and **prethks**.

- Click **Add term** and select **1** from the **order** drop-down list
- Select **cc** from the first **variable** drop-down list
- Select **prethks** from the second **variable** drop-down list
- Check that the window matches that shown below and then click **Done**

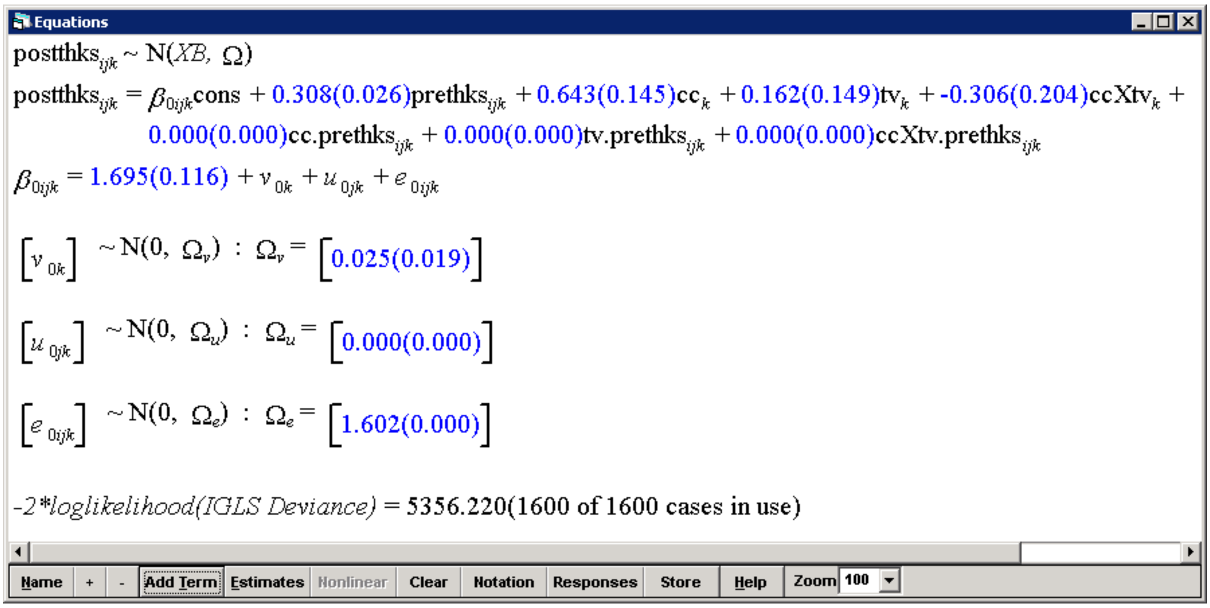


Repeat this process in order to add the remaining two cross-level interactions

- Click **Add term** and select **1** from the **order** drop-down list
- Select **tv** from the first **variable** drop-down list
- Select **prethks** from the second **variable** drop-down list

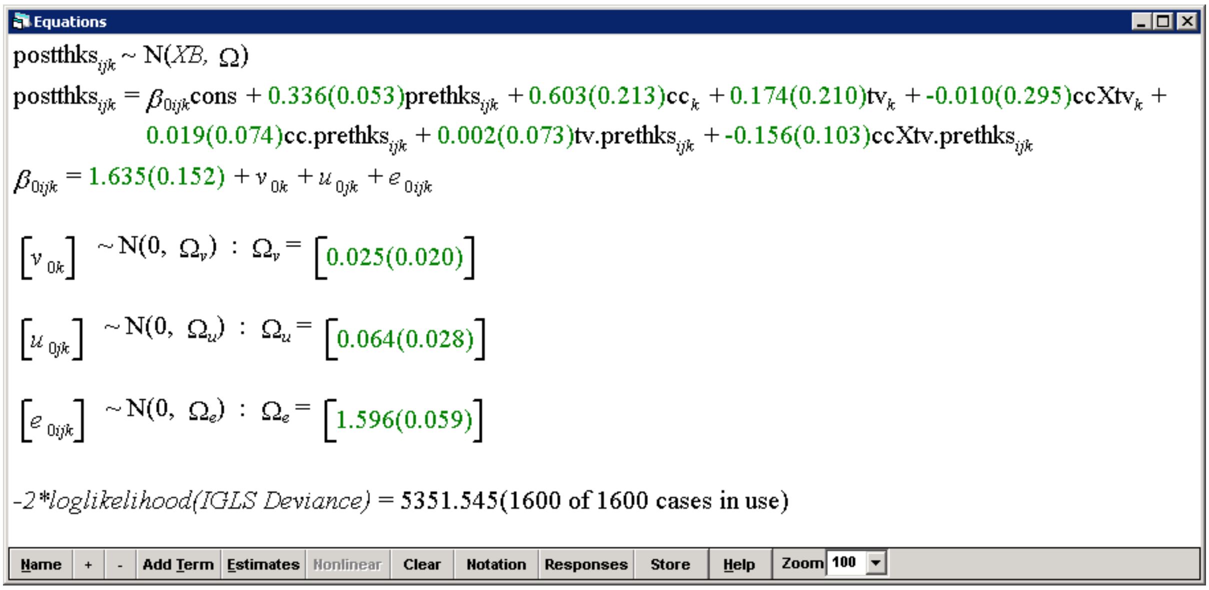
- Click **Done**
- Click **Add term** and select **1** from the **order** drop-down list
- Select **ccXtv** from the first **variable** drop-down list
- Select **prethks** from the second **variable** drop-down list
- Click **Done**

The **Equations** window should now look as follows.



Fit the model.

- Click **Start**



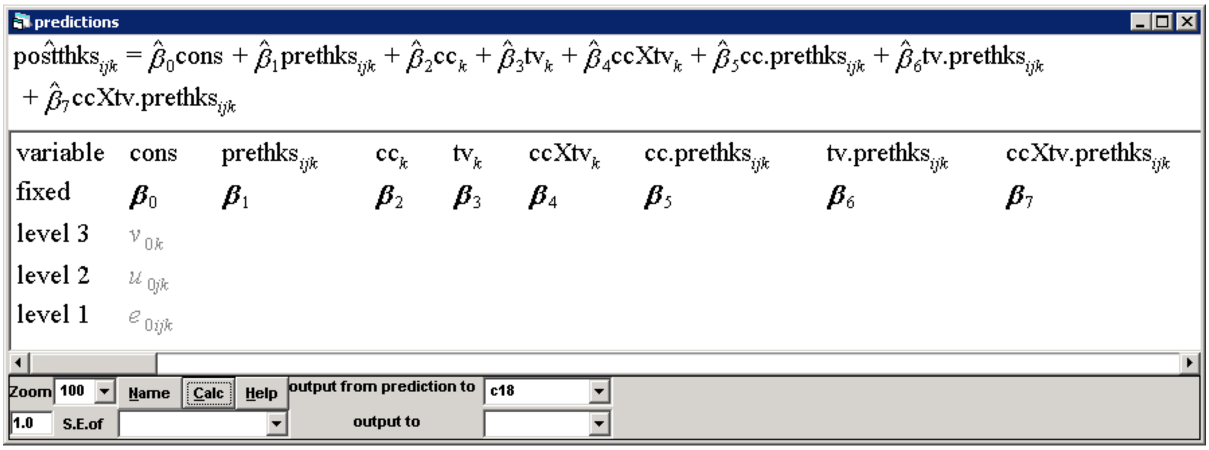
Store the estimation results.

- In the **Equations** window toolbar, click **Store**
- Type ‘**model6**’ into the box
- Click **OK**

An LR test (χ^3_2 = 5.81, *p* = 0.1211) comparing this model (Model 6) to the model with no cross-level interactions (Model 3) shows that including the three interaction terms does not significantly improve the fit of the model. This suggests that the effects of the CC and TV interventions do not actually vary by students' levels of prior knowledge. However, looking at the coefficients of the cross-level interactions, we see that the first two coefficients (**ccXprethks** and **tvXprethks**) are effectively zero while the third coefficient (**ccXtvXprethks**), although not significant is fairly large. Indeed the magnitude of the third coefficient suggests that the relationship between baseline and postintervention scores for students receiving both interventions is approximately half as strong as it is for students in the other three conditions.

We can visualise these results by calculating model predictions and then plotting them against students' baseline scores. First we predict students' scores.

- From the **Model** menu, select **Predictions**
- Click on **fixed** and select **Include all fixed coefficients**
- In the **output from prediction to** drop-down list, select **c18**
- Check that the window matches that shown below and then click **Calc**

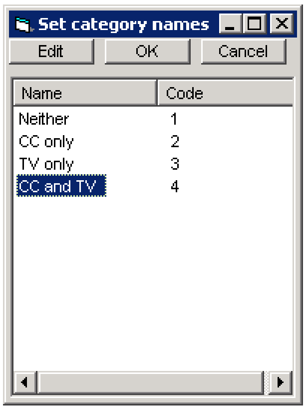


We then plot the predicted postintervention scores against students' baseline scores separately for the four conditions. However, first we need to generate a category indicator variable for the four conditions.

- 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

```
CALC c19 = 1 + ‘cc’ + 2*‘tv’
NAME c19 ‘condition’
```

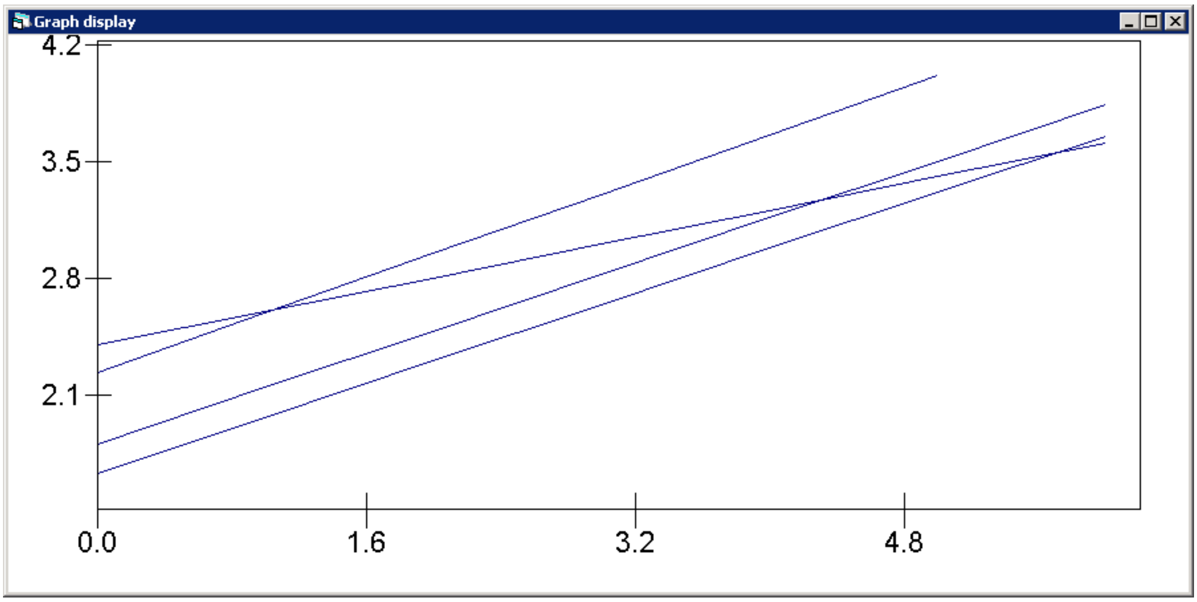
- In the **Names** window, select the variable **condition** and then, under **Column**, click the **Toggle Categorical** button to declare the variable to be a categorical variable
- Under **Categories**, click on **View** to open the **Set category names** window
- Click **condition_1** to highlight the value label associated with **Code 1** and then click the **Edit** button and rename the value label to '**Neither**'
- Repeat this process to rename the value labels associated with **Code 2**, **3** and **4** to '**CC only**', '**TV only**' and '**CC and TV**', respectively
- Check that the window matches that shown below and then click **OK**



We can now proceed to plot the graph.

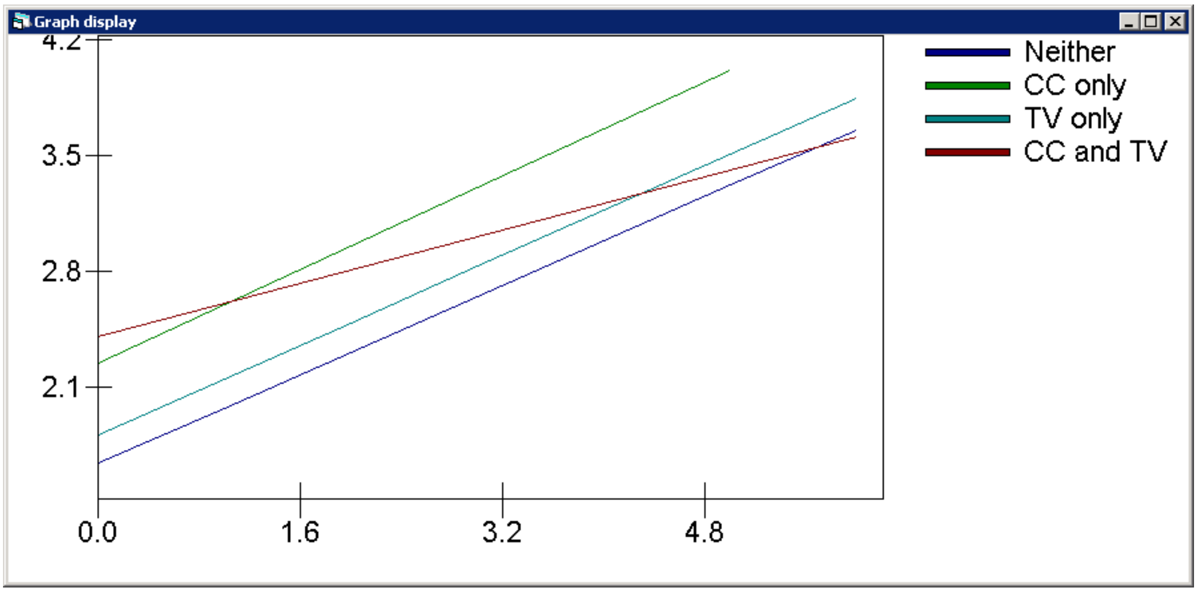
- From the **Graphs** menu, select **Customised Graph(s)**
- On the **plot what?** tab, select **line** from the **plot type** drop-down list
- For the **y** drop-down list select **c18**
- For the **x** drop-down list select **prethks**
- For the **group** drop-down list select **condition**
- Click **Apply**

You should see the following plot.



We can make this plot more informative by colour coding the lines and by adding a legend to the graph.

- In the **Customised graph** window, select the **plot style** tab
- Next to the **colour** drop-down list select **16 rotate**
- On the **other** tab, tick the **group code** checkbox
- Click **Apply**



The plot clearly shows that the relationship between baseline and postintervention scores is much shallower for students receiving both interventions (CC & TV) than it is for students in the other three groups. The plot also suggests that receiving both interventions relative to receiving neither intervention is effective for students with low prior knowledge, but not so for students with high prior knowledge. For example students with a baseline score of 0 who receive both interventions are predicted to score approximately two-thirds of a point higher than equivalent students who receive neither intervention. However, students with a baseline score of 6 who receive both interventions are not predicted to score any higher than equivalent students who receive neither intervention.

Graphing the model predictions has shown us that the predicted lines for the Neither, CC only and TV only groups are effectively parallel, while the line for the CC & TV group is substantially shallower. Given this, we might choose to simplify the current model by forcing the slopes for the first three groups to be the same. This can be achieved by simply removing the variables **ccXprethks** and **tvXprethks** from the model. Interestingly, if we do this and refit the model we find that this new model with the single cross-level interaction **ccXtvXprethks** is preferred over the model with no cross-level interactions (χ_1^2 = 5.73, *p* = 0.017). This suggests that the relationship seen for students in the CC & TV group is in fact significantly different from that of the other groups combined.

Further Reading

Readers interested in the tobacco and health application analysed in this practical are referred to the original study by Flay *et al.* (1989) and the subsequent three-level multilevel analysis by Hedeker *et al.* (1994) for further information.

Researchers familiar with the R or Stata software packages may wish to fit three-level and other multilevel models available in MLwiN by calling MLwiN from within R or Stata using the R2MLwiN (Zhang et al., 2012) and runmlwin (Leckie and Charlton, 2013) commands, respectively.

References

Flay, B. R., Brannon, B. R., Johnson, C. A., Hansen, W. B., Ulene, A. L., Whitney-Saltiel, D. A., Gleason, L. R., Sussman, S., Gavin, M., Glowacz, K. M., Sobol, D. F., and Spiegel, D. C. (1989). The Television, School and Family Smoking Cessation and Prevention Project: I. Theoretical basis and program development. *Preventive Medicine*, 17, 585-607.

Hedeker, D., Gibbons, R. D., and Flay, B. R. (1994). Random-Effects Regression Models for Clustered Data: With an Example from Smoking Prevention Research. *Journal of Consulting and Clinical Psychology*, 62, 757-765.

Leckie, G. and Charlton, C. (2013). runmlwin - A Program to Run the MLwiN Multilevel Modelling Software from within Stata. *Journal of Statistical Software*. *Forthcoming*. http://www.bristol.ac.uk/cmm/media/runmlwin/jss.pdf

Zhang, Z., Charlton, C., Parker, R., Leckie, G. and Browne, W. J. (2012). Package 'R2MLwiN'. Centre for Multilevel Modelling, University of Bristol. http://cran.r-project.org/web/packages/R2MLwiN/R2MLwiN.pdf