



Two-level logistic model

Contents

3.1	INTRODUCTION	1
3.1.1	<i>Link functions</i>	1
3.1.2	<i>Methods of estimation</i>	4
3.2	MODELS BASED ON A SUBSET OF THE TVSFP DATA	6
3.2.1	<i>The data</i>	6
3.2.2	<i>A 2-level random intercept logistic model with 2 predictors</i>	11
3.2.3	<i>A 2-level random intercept logistic regression model</i>	28
3.2.4	<i>A 3-level random intercept logistic regression model</i>	30

3.1 Introduction

The nominal and ordinal outcome models can be seen as generalizations of the binary outcome model. In order to understand these models, an understanding of the binary outcome model is required.

A binary random variable is a discrete random variable that has only two possible values, such as whether a subject dies (event) or lives (non-event). Such events are often described as success versus failure, and coded using the values 0 or 1. Consequently, the assumption that this type of outcome variable has a normal distribution does not hold anymore.

The most common distribution used for a binary outcome is the Bernoulli distribution, which takes a value 1 with probability of success p and a value 0 with probability of failure $q = 1 - p$. The selection of the distribution for the outcome variable is not fixed. For example, if the occurrence is very rare, the Poisson distribution can be used.

3.1.1 Link functions

In the case of a binary variable, observed values are usually assigned as either 0 or 1. When such a variable is treated as if it were continuous, predicted values, indicating the probability of the event occurring, can fall outside the (0,1) interval. Moreover, the assumption of

normality at level 1 is not realistic as the random effects can no longer be assumed to have a normal distribution or to have homogeneous variance.

The multilevel generalized linear model (MGLM) generalizes the multilevel model for continuous outcomes by additionally allowing for error distributions from the exponential family (see, for example, McCullagh & Nelder, 1989). Let y denote the outcome variable, and $E(y)$ the expected value of y . The key to MGLM models is that a nonlinear relationship between $E(y)$ and β is allowed, with the aid of a link function.

Suppose that $\mathbf{x} = (x_1 \dots x_n)$ is the vector of all the predictors and that $\boldsymbol{\beta} = (\beta_1 \dots \beta_n)$ is the vector of unknown regression parameters. In the models discussed up to now, it was assumed that the outcomes were normally distributed variables and that a model of the form

$$y_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{v}_i + e_{ij}, \quad j = 1, 2, \dots, n_i$$

could be used to describe the relationship between the outcome and predictor variables. The vector \mathbf{z}'_{ij} denotes a design vector for the random effects contained in the vector \mathbf{v}_i , and \mathbf{x}'_{ij} is the design vector for the predictors in the fixed part of the model with corresponding vector $\boldsymbol{\beta}$ of regression parameters. The covariance matrix of \mathbf{v}_i is denoted by $\boldsymbol{\Phi}_{(2)}$ and the variance of e_{ij} by σ_e^2 .

The link function specifies a nonlinear transformation between the linear predictor η and the assumed distribution function. These link functions transform the observed outcome value to a function $\eta = \mathbf{x}'\boldsymbol{\beta}$ and ensure that the predicted probability lies within the (0,1) interval. Instead of y , η is being analyzed. For the binary outcome, the probability of success η is the predictor of interest.

The most commonly used link functions are the log, logit, probit and complementary log-log link functions. The log link generally is used for the count variable with Poisson distribution, which will be discussed in the next chapter. The link functions available in SuperMix include the logit, probit and complementary log-log functions for models with an ordinal dependent variable, and the logit link function for models with a nominal dependent variable. Table 4.1 shows these link functions, along with their distribution functions (CDF), means and variances.

Table 4.1: Link functions for the Bernoulli distribution

Link name	Link function $F^{-1}(p), 0 < p < 1$	CDF $-\infty < w < \infty$	Mean	Variance
logit (logistic)	$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right)$	$\frac{e^w}{1+e^w}$	0	$\frac{\pi^2}{3}$
probit	$\Phi^{-1}(p)$, where Φ^{-1} is the inverse of the standard normal cumulative distribution	$\Phi(w)$	0	1
complementary log-log	$\log(-\log(1-p))$	$1 - \exp(-\exp(w))$	-0.577	$\frac{\pi^2}{6}$

These link functions map the probability η with an open interval (0,1) to the entire set of real numbers \mathbb{R} . Figure 4.1 illustrates how a real number w is transformed to the probability η .

As shown below, the logit and probit link functions are both symmetric around a value of 0. The logit function has a larger variance. The complementary log-log link function is asymmetric. When the probability of a successful outcome (p) is extremely small or large, the linear relationship does not hold. Understanding the nature of the link function used in an analysis is essential to the correct interpretation of the results.

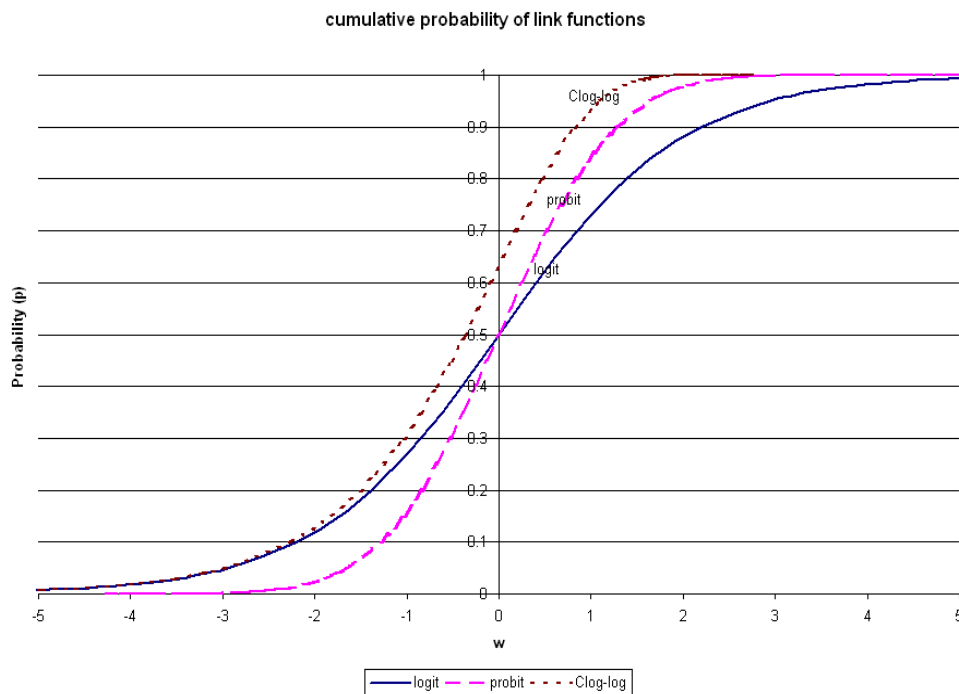


Figure 4.1: Cumulative density of link functions

3.1.2 Methods of estimation

For models with binary, ordinal, count, and nominal outcomes, SuperMix offers two methods of estimation: maximization of the posterior distribution (MAP) and numerical integration (adaptive and non-adaptive quadrature) to obtain parameter and standard error estimates.

The MAP method of estimation can be used to obtain a point estimate of an unobserved quantity on the basis of empirical data. It is closely related to Fisher's method of maximum likelihood (ML), but employs an augmented optimization objective which incorporates a prior distribution over the quantity one wants to estimate.

Quadrature is a numeric method for evaluating multi-dimensional integrals. For mixed effect models with count and categorical outcomes, the log-likelihood function is expressed as the sum of the logarithm of integrals, where the summation is over higher-level units, and the dimensionality of the integrals equals the number of random effects.

Typically, MAP estimates are used as starting values for the quadrature procedure. When the number of random effects is large, the quadrature procedures can become computationally intensive. In such cases, MAP estimation is usually selected as the final method of estimation. Numerical quadrature, as implemented in SuperMix, offers users a choice between adaptive and non-adaptive quadrature. Quadrature uses a quadrature rule, *i.e.*, an approximation of the definite integral of a function, usually stated as a weighted sum of function values at specified points within the domain of integration.

Adaptive quadrature generally requires fewer points and weights to yield estimates of the model parameters and standard errors that are as accurate as would be obtained with more points and weights in non-adaptive quadrature. The reason for that is that the adaptive quadrature procedure uses the empirical Bayes means and covariances, updated at each iteration to essentially shift and scale the quadrature locations of each higher-level unit in order to place them under the peak of the corresponding integral.

A brief description of MAP estimation and quadrature follows below.

MAP estimation

For level-2 unit i , let $v_{i1}, v_{i2}, \dots, v_{ir}$ denote the random effects and $y_{i1}, y_{i2}, \dots, y_{in_i}$ the outcomes. Let $f(\mathbf{v}_i, \mathbf{y}_i)$ denote the joint distribution of $\mathbf{v}_i = (v_{i1}, v_{i2}, \dots, v_{ir})$ and $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{in_i})$.

Using standard results for conditional distributions, it follows that

$$f(\mathbf{v}_i | \mathbf{y}_i) = f(\mathbf{y}_i | \mathbf{v}_i) f(\mathbf{v}_i) / f(\mathbf{y}_i).$$

By taking logarithms on both sides of the equation, the following density function is obtained:

$$\ln f(\mathbf{v}_i | \mathbf{y}_i) = \ln f(\mathbf{y}_i | \mathbf{v}_i) + \ln f(\mathbf{v}_i) - K$$

where K is a constant. Mode estimates \hat{v}_i of the random effects and estimates $\hat{\boldsymbol{\beta}}$ of the fixed parameters are obtained by iteratively solving the equations

$$\frac{\partial}{\partial \gamma_k} \ln f(\mathbf{v}_i | \mathbf{y}_i) = 0,$$

where γ_k is a typical element of the unknown parameters $v_{i1}, v_{i2}, \dots, v_{ir}$ and $\beta_1, \beta_2, \dots, \beta_p$.

As a by-product of the iterative procedure, estimates of $\text{cov}\left(\hat{v}_i\right), i = 1, 2, \dots, N$ are obtained and these, in turn, are used to estimate $\Phi_{(2)} = \text{cov}(\mathbf{v}_i)$.

Numerical quadrature

Since

$$f(\mathbf{y}_i, \mathbf{v}_i) = f(\mathbf{y}_i | \mathbf{v}_i) f(\mathbf{v}_i)$$

it follows that the marginal distribution of \mathbf{y}_i can be obtained as the solution to the multi-dimensional integral

$$f(\mathbf{y}_i) = \int_{v_1} \dots \int_{v_r} f(\mathbf{y}_i | \mathbf{v}_i) f(\mathbf{v}_i) dv_1 \dots dv_r.$$

Since it is assumed that $\mathbf{v}_i \sim N(\mathbf{0}, \Phi_{(2)})$ it follows, for example, that

$$f(\mathbf{v}_i) = (2\pi)^{-r/2} |\mathbf{\Phi}_{(2)}|^{-1/2} \exp\left[-\frac{1}{2} \mathbf{v}_i' \mathbf{\Phi}_{(2)}^{-1} \mathbf{v}_i\right].$$

In general, a closed-form solution to this integral does not exist. To evaluate integrals of the type described above, we use a direct implementation of Gauss-Hermite quadrature (see, *e.g.*, Krommer & Ueberhuber, 1994, Section 4.2.6 and Stroud & Sechrest, 1966, Section 1).

With this rule, an integral of the form

$$I(t) = \int f(t) \exp[-t^2] dt$$

is approximated by the sum

$$I(t) \approx \sum_{u=1}^Q w_u f(z_u),$$

where w_u and z_u are weights and nodes of the Hermite polynomial of degree Q . A Q -point adaptive quadrature rule is a quadrature rule constructed to yield an exact result for polynomials of degree $2Q-1$, by a suitable choice of the n points x_i and n weights w_i .

3.2 Models based on a subset of the TVSFP data

3.2.1 The data

The data are from the Television School and Family Smoking Prevention and Cessation Project (TVSFP) study (Flay, *et. al.*, 1988) described in Section 3.3. The study was designed to test the independent and combined effects of a school-based social-resistance curriculum and a television-based program in terms of tobacco use and cessation.

A tobacco and health knowledge scale was used in classifying subjects as knowledgeable or not. In its original form, the student's score was defined as the number of correct answers to seven items on tobacco and health knowledge. The structure of this study indicates a three-level hierarchical structure. However, we will first consider two two-level structures. In the first, students are nested within schools; in the second, students are nested within classrooms. Finally, a three-level model recognizing the role of both classroom and school in the hierarchical structure of the data will be considered.

Data for the first 10 participants on most of the variables used in this section are shown below in the form of a SuperMix spreadsheet file, named **tvsfors.ss3**, located in the **Examples\Binary** subfolder.

SuperMix - [tvspors.ss3]

File Edit Window Help

403 Apply

	(A)_School	(B)_Class	(C)_THKSord	(D)_THKSbi	(E)_PreTHK	(F)_CC	(G)_TV	(H)_CC*TV
1	403.00	403101.00	3.00	1.00	2.00	1.00	0.00	0.00
2	403.00	403101.00	4.00	1.00	4.00	1.00	0.00	0.00
3	403.00	403101.00	3.00	1.00	4.00	1.00	0.00	0.00
4	403.00	403101.00	4.00	1.00	3.00	1.00	0.00	0.00
5	403.00	403101.00	4.00	1.00	3.00	1.00	0.00	0.00
6	403.00	403101.00	3.00	1.00	4.00	1.00	0.00	0.00
7	403.00	403101.00	2.00	0.00	2.00	1.00	0.00	0.00
8	403.00	403101.00	4.00	1.00	4.00	1.00	0.00	0.00
9	403.00	403101.00	4.00	1.00	5.00	1.00	0.00	0.00
10	403.00	403101.00	4.00	1.00	3.00	1.00	0.00	0.00

The variables of interest are:

- School indicates the school a student is from (28 schools in total).
- Class identifies the classroom (135 classrooms in total).
- THKSord represents the tobacco and health knowledge scale, with 4 categories ranging between 1 and 4. The frequency distribution of the post-intervention THKS scores indicated that approximately half the students had scores of 2 or less, and half of 3 or greater. In terms of quartiles, four ordinal classifications were suggested corresponding to 0 – 1, 2, 3, and 4 – 7 correct responses.
- THKSbin is a recoded version of the ordinal variable THKSord, but in binary form: a value of "0" indicates an original scale score of 1 or 2, while a value of "1" indicates a scale score of 3 or 4. This variable will serve as our outcome variable in the current chapter.
- PreTHKS indicates a student's score prior to intervention, *i.e.*, the number correct of 7 items.
- CC is a binary variable indicating whether a social-resistance classroom curriculum was introduced, with 0 indicating "no" and 1 "yes."
- TV is an indicator variable for the use of media (television) intervention, with a "1" indicating the use of media intervention, and "0" the absence thereof.
- CC*TV is the product of the variables TV and CC, and represents the CC by TV interaction.

In this chapter, we consider models for binary outcomes, using quadrature as method of estimation.

3.2.1.1 Exploring the data

Crosstabulation

The focus is on the influence of the intervention on the tobacco health knowledge scores of the students, as represented by the binary outcome variable THKSbin. A cross-tabulation of the variables CC and TV for the two categories of the binary variable THKSbin is given in Table 4.1 below.

Table 4.1: Crosstabulation of CC, TV and THKSbin

THKSbin			CC		Total
			0	1	
0	TV	0	246	140	386
		1	215	152	367
Total			461	292	753
1	TV	0	175	240	415
		1	201	231	432
Total			376	471	847

The proportion of students with high scale scores (THKSbin = 1) in each of the four cells formed by the categories of CC and TV can be derived from Table 4.1. For example, 246 students in the category CC = 0, TV = 0 had a low score (THKSbin = 0), while 175 students had a high score (THKSbin = 1). The proportion of students in this cell with a high score is thus $\frac{175}{175 + 246} = 0.4157$. The observed proportions of high scores are summarized in Table 4.2 below.

Table 4.2: Observed proportion of high post-intervention scores

Study condition	Proportion	odds	logits
CC = 0, TV = 0	0.4157	0.711	-0.340
CC = 0, TV = 1	0.4832	0.935	-0.067
CC = 1, TV = 0	0.6316	1.714	0.539
CC = 1, TV = 1	0.6031	1.520	0.418

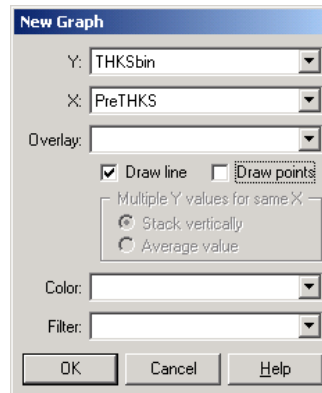
Proportions less than 0.5 indicate odds less than one and negative logits, while proportions above 0.5 yield odds greater than one and positive logit values. We note that, based on the observed proportion of high post-intervention scores, the use of only the social-resistance classroom curriculum (as represented by the variable CC) seemed the most successful, followed by the use of both curriculum and media intervention (CC = 1, TV = 1).

Exploratory graphs

The pre-intervention scores of the students may be useful as a covariate in the analysis. To get an idea of the relationship between the scale score PreTHKS and the post-intervention

score THKSbin, an exploratory graph is created. Select the **Data-based Graphs, Exploratory** option from the **File** menu.

The **New Graph** dialog box is activated. Select the binary outcome variable THKSbin as the **Y** variable and the pre-intervention score PreTHKS as the **X** variable. Uncheck the **Draw points** check box, which is checked by default, to obtain the settings as shown.



Click **OK** to obtain Figure 4.1. The value associated with the tick marks on the X-axis represents the proportion of students with that PreTHKS score that had a value of 1 on the THKSbin variable, in other words the proportion of students with a post-intervention score of 3 or 4. We note that the relationship is reasonably linear, and that higher post-intervention scores are more often observed for students with high pre-intervention scores, which is what one intuitively would expect.

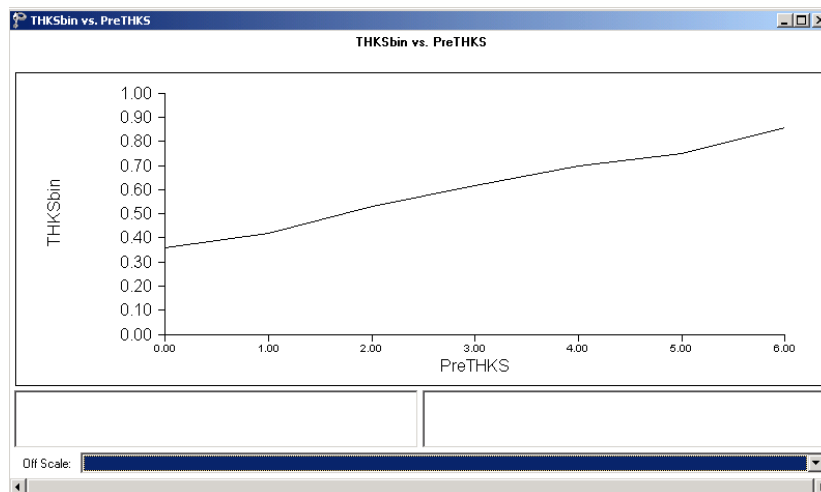
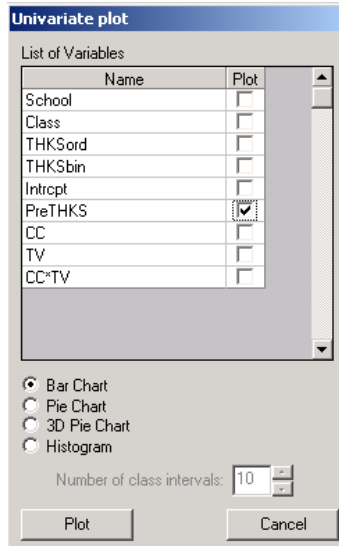


Figure 4.1: Exploratory graph of THKSbin vs. PreTHKS

Univariate graphs

We now take a closer look at the distribution of the pre-intervention scores by utilizing the **Data-based Graphs, Univariate** option on the **File** menu. By default, a bar chart will be produced. Select the variable PreTHKS in the **Plot** column, and click **Plot**.



By clicking anywhere in the bars, the **Bar Graph Parameters** dialog box is activated. Click the **Data** button and then **OK** to display the data used to construct the bar chart.

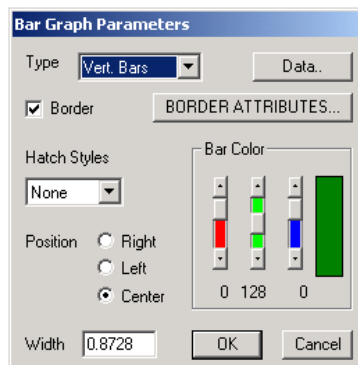


Figure 4.2 below shows both the graphing window with bar chart and the data in spreadsheet format. Note that only 55 of the 1600 observations showed a score of 5 or higher, and that no student obtained a post-intervention score of 7 out of 7.

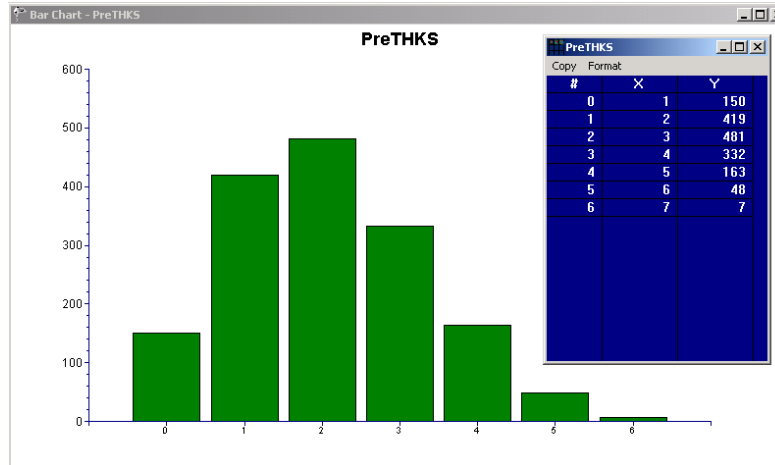


Figure 4.2: Bar chart of PreTHKS values

Finally, we also take a look at the mean pre-intervention scores of the students for each of the four subgroups. These are summarized in Table 4.3 below, and show that the mean pre-intervention scores do not differ much.

Table 4.3: Mean pre-intervention scores

Study condition	Mean
CC = 0, TV = 0	2.152
CC = 0, TV = 1	2.087
CC = 1, TV = 0	2.050
CC = 1, TV = 1	1.979

3.2.2 A 2-level random intercept logistic model with 2 predictors

3.2.2.1 The model

The outcome variable THKSbin used here is binary. It assumes a value of "0" when the original scale score was either 1 or 2, and a value of "1" for an original scale score of 3 or 4. The predicted value of the outcome can be viewed as the predicted probability that THKSbin is 1. As explained in Section 4.1.1, predicted values outside the interval [0,1] would not be meaningful and a model constraining predicted values to lie in this interval would be appropriate, in contrast with the model for a continuous outcome (see above) where effect sizes outside this interval would be interpretable. In addition, the assumption of normality at level 1 is not realistic, as the level-1 random effect can only assume one of two values: 0 or 1. This random effect can thus not have homogeneous variance.

In order to insure that the predicted values lie within the (0,1) interval, a transformation of the level-1 predicted probability can be used. For the binary case considered here, the following link function is used:

$$\text{Prob}(\text{THKS}_{ij} = 1 | \boldsymbol{\beta}, \mathbf{v}) = \frac{e^{\eta_{ij}}}{1 + e^{\eta_{ij}}}$$

where η_{ij} represents the log of the odds of success.

For the current model, we want to explore the relationship between the post-intervention scores and the type of intervention applied. This relationship can be expressed as

Level 1 model:

$$\eta_{ij} = b_{0i} + b_{1i} \times \text{CC}_i + b_{2i} \times \text{TV}_i + b_{3i} \times \text{CC}_i * \text{TV}_{ij} + b_{4i} \times \text{PreTHKS}_{ij} + e_{ij}$$

Level 2 model:

$$b_{0i} = \beta_0 + v_{0i}$$

$$b_{1i} = \beta_1$$

$$b_{2i} = \beta_2$$

$$b_{3i} = \beta_3$$

$$b_{4i} = \beta_4$$

An equivalent expression for the model is

$$\eta_{ij} = \beta_0 + \beta_1 \times \text{CC}_i + \beta_2 \times \text{TV}_i + \beta_3 \times \text{CC}_i * \text{TV}_{ij} + \beta_4 \times \text{PreTHKS}_{ij} + v_{0i} + e_{ij}.$$

The interpretation of the logistic regression model is made in terms of the logits, as the model is linear in terms of the logits. Thus the coefficients $\beta_1, \beta_2, \dots, \beta_4$ can be interpreted as follows:

- β_0 is the THKS logit for $\text{CC} = 0, \text{TV} = 0$, that is the log odds of a positive outcome for an individual from the control group where no intervention was made and with a pre-intervention score of 0. One could also refer to β_0 as the PreTHKS adjusted logit for the $\text{CC} = 0, \text{TV} = 0$ subgroup.
- β_1 = the logit difference between $(\text{CC} = 1, \text{TV} = 0)$ and $(\text{CC} = 0, \text{TV} = 0)$ for the case where $\text{PreTHKS} = 0$:

$$\eta_{ij} = \beta_0 + (\beta_1 + \beta_3 \text{TV}_i) \text{CC}_i + \beta_2 \text{TV}_i + \beta_4 \text{PreTHKS}_{ij} + v_{0i},$$

in other words, the PreTHKS adjusted logit difference between these two subgroups.

- β_2 = the logit difference between (TV = 1, CC = 0) and (TV = 0, CC = 0) with PreTHKS = 0:

$$\eta_{ij} = \beta_0 + (\beta_2 + \beta_3 CC_i) TV_i + \beta_1 CC_i + \beta_4 \text{PreTHKS}_{ij} + v_{0i}.$$

- β_3 is the difference in logit attributable to the interaction between the two intervention methods.

The interpretation of the coefficients is dependent on the coding of the variables used in the model.

3.2.2.2 Setting up the analysis

Using the data in **tvspors.ss3**, we consider the situation where students are nested within schools, and fit a two-level model with the binary variable THKSbin as outcome. We wish to examine the relationships between the outcome and the two intervention methods employed, simultaneously taking students' pre-intervention scores into account. To do so, we use the model described above with schools as the level-2 units.

Use the **File, Open Spreadsheet** option to activate the **Open** dialog box. Browse for the file **tvspors.ss3** in the **Examples\Binary** folder. Select the file and click the **Open** button to return to the main SuperMix window, where the contents of the SuperMix system file are displayed.

	(A)_School	(B)_Class	(C)_THKSor	(D)_THKSbi	(E)_PreTHK	(F)_CC	(G)_TV	(H)_CC*TV
1	403.00	403101.00	3.00	1.00	2.00	1.00	0.00	0.00
2	403.00	403101.00	4.00	1.00	4.00	1.00	0.00	0.00
3	403.00	403101.00	3.00	1.00	4.00	1.00	0.00	0.00
4	403.00	403101.00	4.00	1.00	3.00	1.00	0.00	0.00
5	403.00	403101.00	4.00	1.00	3.00	1.00	0.00	0.00
6	403.00	403101.00	3.00	1.00	4.00	1.00	0.00	0.00
7	403.00	403101.00	2.00	0.00	2.00	1.00	0.00	0.00
8	403.00	403101.00	4.00	1.00	4.00	1.00	0.00	0.00
9	403.00	403101.00	4.00	1.00	5.00	1.00	0.00	0.00
10	403.00	403101.00	4.00	1.00	3.00	1.00	0.00	0.00

Next, we use the SuperMix interface to provide the model specifications. From the main menu bar, select the **File, New Model Setup** option.

The **Model Setup** dialog box that appears has six tabs: **Configuration, Variables, Starting Values, Patterns, Advanced, and Linear Transforms**. In this example, only three of the tabs are used.

As a first step, the binary outcome variable THKSbin is selected from the **Dependent Variable** drop-down list box. The type of outcome is specified as binary using the drop-down list box in the **Dependent Variable Type** field. Once this selection is made, the **Categories** field is displayed. The school identification variable is used to define the hierarchical structure of the data, and is selected as the **Level-2 ID** from the **Level-2 IDs** drop-down list box. A title for the analysis (optional) is entered in the **Title** fields. A convergence criterion of 0.0001 is requested. By default, the maximum number of iterations performed is set to 100. Empirical Bayes residuals, written to additional output files, are requested by setting the **Write Bayes Estimates** option to **means and (co)variances**. Default settings for all other options associated with this tab are used. Proceed to the **Variables** tab by clicking on this tab.

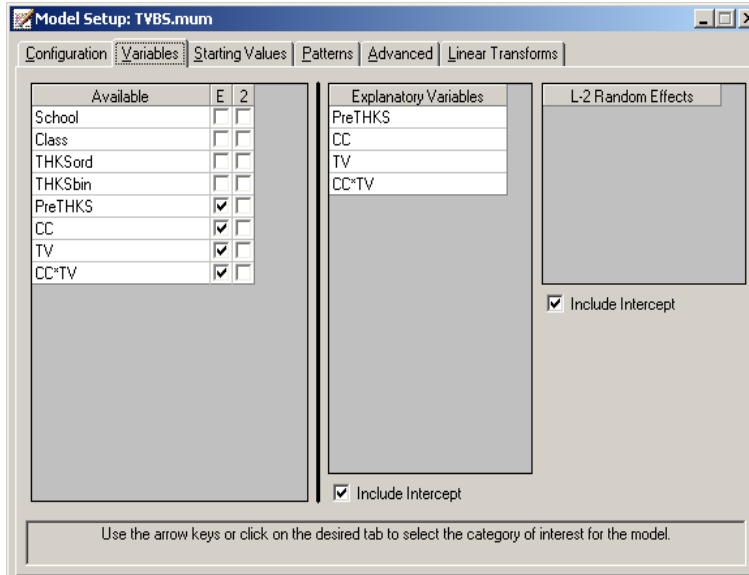
The screenshot shows the 'Model Setup: TVBS.mum' dialog box with the 'Configuration' tab selected. The fields are as follows:

- Title 1: Logistic 2 level random intercept model
- Title 2: TVSF data
- Dependent Variable Type: binary
- Dependent Variable: THKSbin
- Level-2 IDs: School
- Level-3 IDs: (empty)
- Write Bayes Estimates: means & (co)variances
- Convergence Criterion: 0.0001
- Number of Iterations: 100
- Missing Values Present: false
- Perform Crosstabulation: no
- Output Type: standard

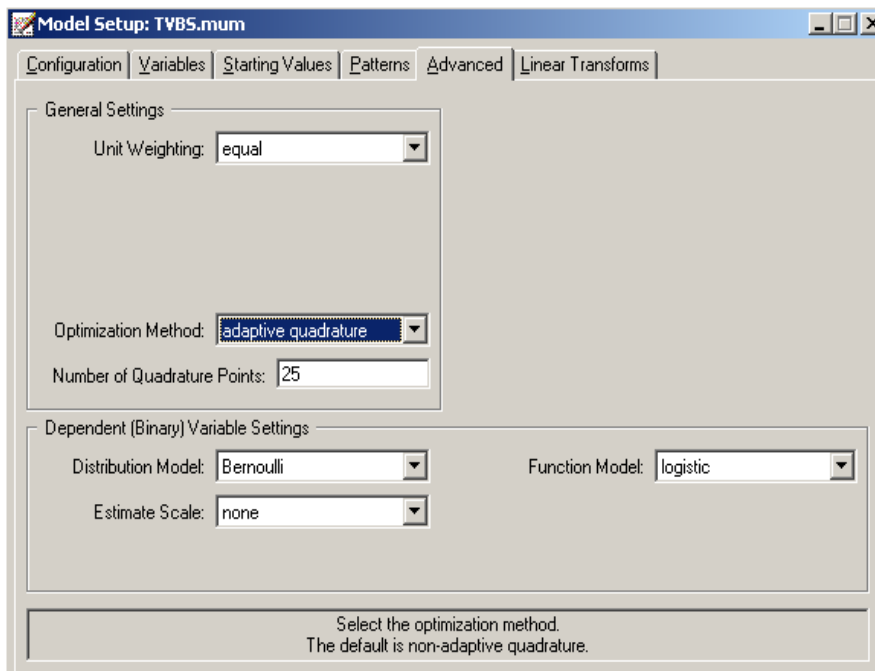
The 'Categories' table is shown below:

	Value
1	0
2	1

The **Variables** tab is used to specify the fixed and random effects to be included in the model. Start by selecting the explanatory (fixed) variables using the first column of boxes in the **Available** group field. The first variable selected is PreTHKS, followed by CC, TV, and the interaction term CC*TV. After selecting these explanatory variables, the random effect(s) at level 2 must be selected. In this case, we wish to allow only the intercept to vary randomly over the schools. By default, the intercept is assumed to vary randomly over higher levels of the hierarchy as indicated by the checked box for the **Include Intercept** option in the **L-2 Random Effects** group field. A common fixed intercept coefficient is also included, as shown in the **Explanatory Variables** group field.



We opt to increase the number of quadrature points to be used during estimation. To do so, click the **Advanced** tab. First select **adaptive quadrature** from the **Optimization Method** drop-down list box, then change the **Number of Quadrature Points** field to 25. The default distribution for a binary outcome variable is **Bernoulli** and the default link function is **probit**. Change **probit** to **logistic** by using the drop-down list box in the **Function Model** field.



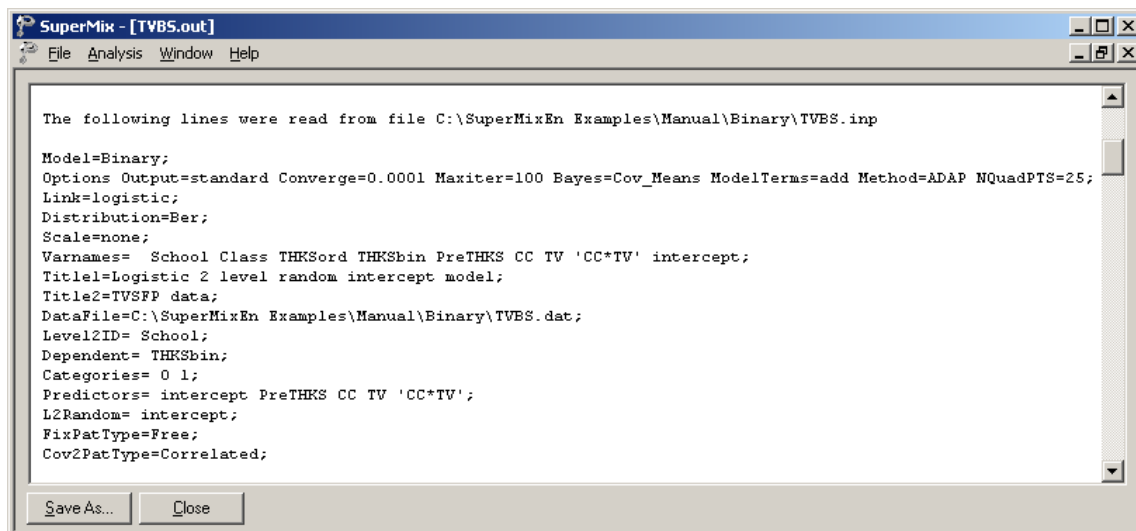
Before running the analysis, the model specifications have to be saved. Select the **File, Save** option, and provide a name for the model specification file, for example **TVBS.mum**. Run the analysis by selecting the **Run** option from the **Analysis** menu.

3.2.2.3 Discussion of results

Portions of the output file **tvbs.out** are shown below.

Syntax

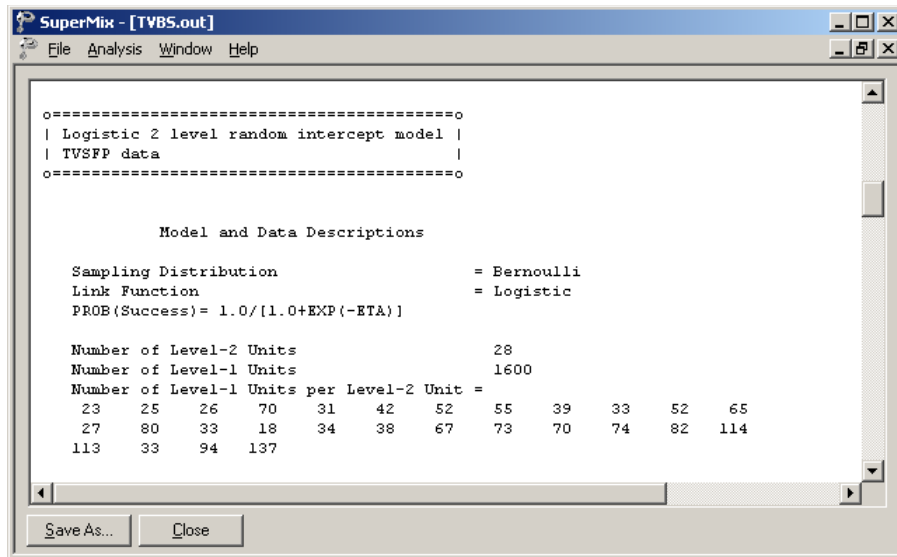
At the top of the file, the syntax saved to the **TVBS.mum** file is shown. The first part states the selection of iteration control options, requests for Bayes residuals, and the specifications necessary to define the model fitted as an binary model with a logistic link function. The second part of the syntax provides information on the structure of the data, the name and structure of the outcome variable, and the predictors included in the model. Text to the left of the equal sign in each line denote keywords recognized by the program; text to the right are either keywords (for example, in the case of `Cov2PatType = Correlated`) or variable names as given in the **ss3** file (for example, `Level2ID = School`).



```
The following lines were read from file C:\SuperMixEn Examples\Manual\Binary\TVBS.inp
Model=Binary;
Options Output=standard Converge=0.0001 Maxiter=100 Bayes=Cov_Means ModelTerms=add Method=ADAP NQuadPTS=25;
Link=logistic;
Distribution=Ber;
Scale=none;
Varnames= School Class THKSord THKSbin PreTHKS CC TV 'CC*TV' intercept;
Title1=Logistic 2 level random intercept model;
Title2=TVSFP data;
DataFile=C:\SuperMixEn Examples\Manual\Binary\TVBS.dat;
Level2ID= School;
Dependent= THKSbin;
Categories= 0 1;
Predictors= intercept PreTHKS CC TV 'CC*TV';
L2Random= intercept;
FixPatType=Free;
Cov2PatType=Correlated;
```

Model and data description

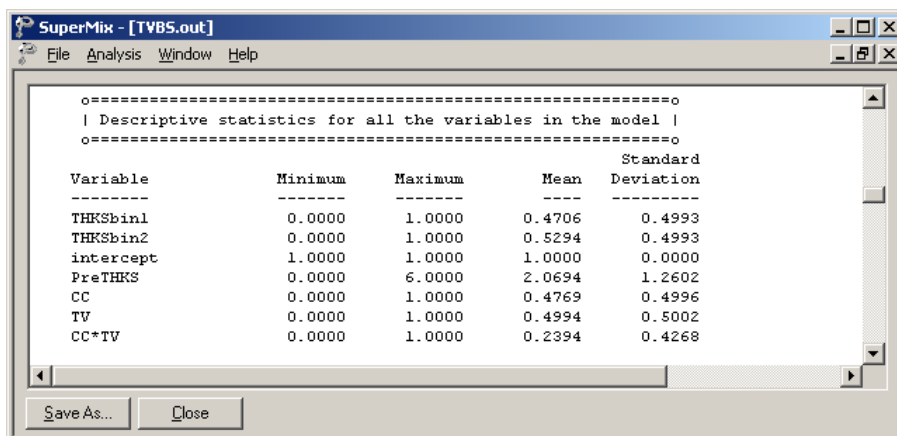
The next section of the output file contains a description of the hierarchical structure and model specifications.



The use of a logistic response function (logit link function) with the assumption of a Bernoulli distribution is indicated. This is followed by a summary of the number of students nested within each school. The number of students per school (level-2 unit) ranges between 23 and 137.

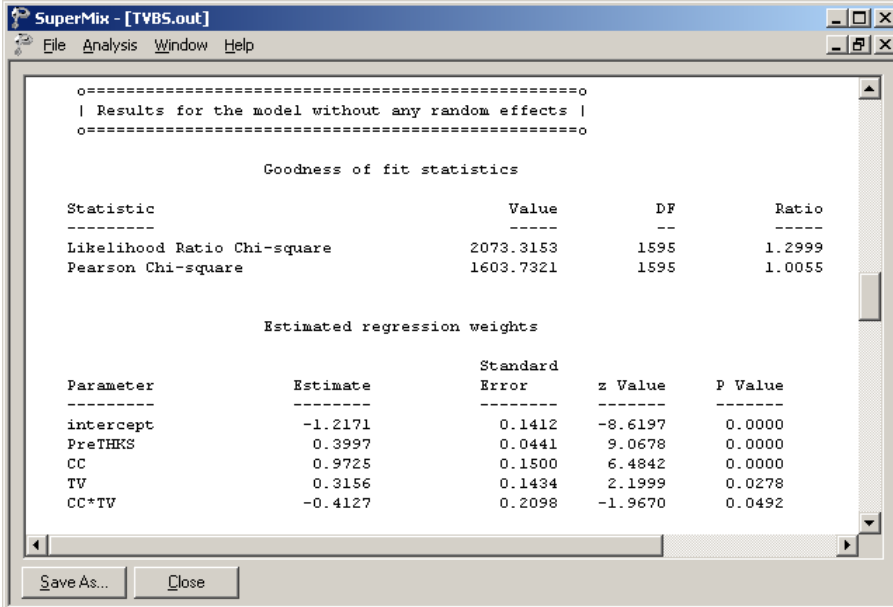
Descriptive statistics

The data summary is followed by descriptive statistics for all variables included in the model. We note that 47% of the students had a value of 0 on the binary knowledge score outcome variable THKSbin, and 53% a value of 1.



Results for the model without any random effects

Descriptive statistics are followed by parameter estimates obtained under the assumption that all random effects are zero. The parameter values for the predictors CC, TV, CC*TV and PreTHKS are given in the first column, followed by the standard errors and z- and p -values.



o-----o
| Results for the model without any random effects |
o-----o

Goodness of fit statistics

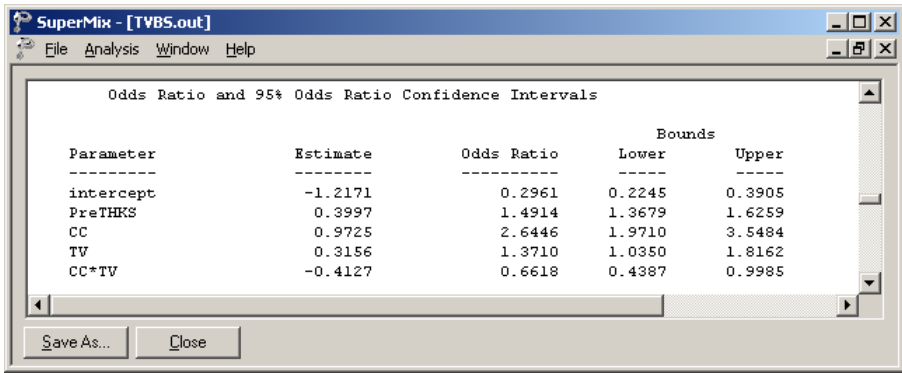
Statistic	Value	DF	Ratio
Likelihood Ratio Chi-square	2073.3153	1595	1.2999
Pearson Chi-square	1603.7321	1595	1.0055

Estimated regression weights

Parameter	Estimate	Standard Error	z Value	P Value
intercept	-1.2171	0.1412	-8.6197	0.0000
PreTHKS	0.3997	0.0441	9.0678	0.0000
CC	0.9725	0.1500	6.4842	0.0000
TV	0.3156	0.1434	2.1999	0.0278
CC*TV	-0.4127	0.2098	-1.9670	0.0492

Save As... Close

This is followed by the odds ratios and associated confidence intervals. The odds ratios are the exponents ($e^{\hat{\beta}}$) of the estimated regression coefficients.



Odds Ratio and 95% Odds Ratio Confidence Intervals

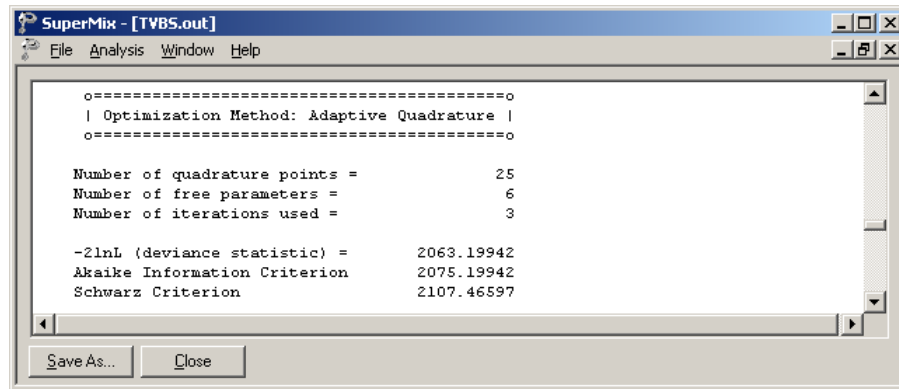
Parameter	Estimate	Odds Ratio	Bounds	
			Lower	Upper
intercept	-1.2171	0.2961	0.2245	0.3905
PreTHKS	0.3997	1.4914	1.3679	1.6259
CC	0.9725	2.6446	1.9710	3.5484
TV	0.3156	1.3710	1.0350	1.8162
CC*TV	-0.4127	0.6618	0.4387	0.9985

Save As... Close

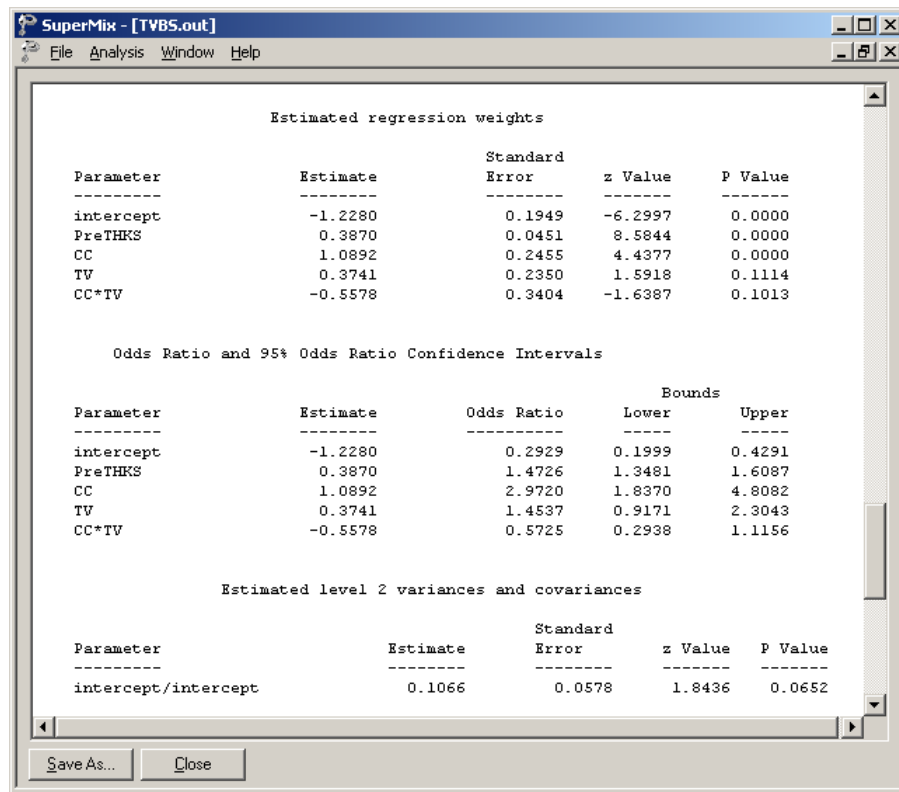
Results for the model fitted with adaptive quadrature

The output describing the estimated parameters after convergence is shown next. Three iterations were required to obtain convergence. The number of quadrature points used per

dimension was 25. The likelihood function value at convergence as well as the deviance are also given, and may be used to compare a set of nested models.



The estimates are shown in the column with heading Estimate, and correspond to the coefficients $\beta_0, \beta_1, \dots, \beta_4$ in the model specification. Significant effects of PreTHKS and CC are observed. The variation in the intercept over the schools is estimated as 0.1065, and from the associated p -value we conclude that there is significant variation, at a 10% level of significance, in the intercept between the schools included in this analysis.

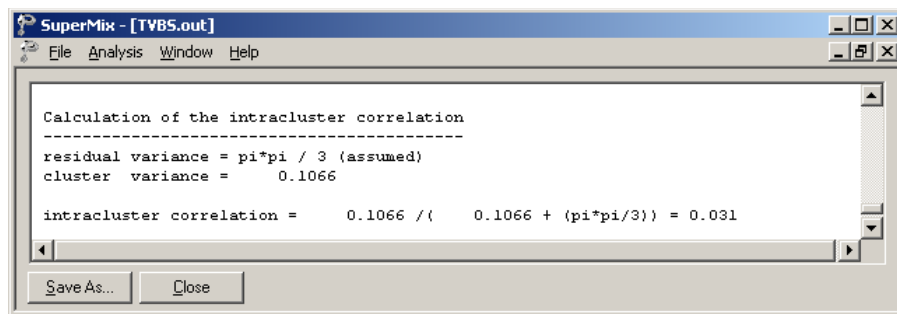


In the case of the fixed effects, a 2-tailed p -value is used, as the alternative hypothesis considered here is of the form $H_1 : \beta \neq 0$. As variances are constrained to be elements of the interval $[0, +\infty)$, the p -values used for these effects are 1-tailed.

If the model is true, it is assumed that the level-1 error variance is equal to $\pi^2 / 3 = 3.29895$ for the logistic link function (see, e.g., Hedeker & Gibbons (2006), p. 157), where π represents the constant 3.141592654.

Thus the estimated ratio between level-2 variation and the total variation is calculated as

$$ICC = \frac{0.1065}{0.1065 + \pi^2 / 3} = 0.031$$



This indicates that almost all variation is attributable to students, rather than to the schools.

3.2.2.4 Interpreting the adaptive quadrature results

The expected log-odds of having a high post-intervention knowledge score (THKSbin score of 1) for a student with a zero value on all the predictors (that is, no social-resistance curriculum, no media intervention, and a pre-intervention knowledge score of 0) is represented by the estimated intercept of -1.2281 . When a social-resistance curriculum was in place ($CC = 1$), or a mass-media intervention was performed ($TV = 1$), the log-odds of a typical student is expected to increase, as indicated by the positive estimated coefficients for CC and TV . Similarly, a higher score on the pre-intervention knowledge test is associated with higher log-odds of a higher post-intervention knowledge score. It can be concluded from the results that the implementation of a classroom curriculum was more likely to lead to a higher post-intervention knowledge score than was the case when mass-media intervention was used. In contrast, the log-odds of a high post-intervention knowledge score was expected to be lower for a typical student from a school where both social resistance classroom curriculum and mass-media intervention defined the study condition for that school, as the estimated coefficient for the interaction term $CC*TV$ was negative.

Estimated outcomes for different groups: unit-specific results

To evaluate the expected effect of CC, TV, CC*TV, and PreTHKS on the predicted probability that the post-intervention score is equal to 1, we use the following expression for the predicted log odds of success

$$\hat{\eta}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 \times CC_i + \hat{\beta}_2 \times TV_i + \hat{\beta}_3 \times CC_i \times TV_i + \hat{\beta}_4 \times \text{PreTHKS}_{ij}$$

for the four groups defined by the categories of CC and TV. Note the similarity of this equation with that given for η_{ij} earlier: random coefficients are not included, as their expected value is 0.

For a typical student with a PreTHKS score of 0 from any school where no media television intervention and no social-resistance classroom curriculum was implemented, $CC = TV = 0$, and thus

$$\hat{\eta}_{ij} = \hat{\beta}_0$$

In the case of a typical student with a PreTHKS score of 0 from any school where only media television intervention was implemented ($TV = 1$),

$$\hat{\eta}_{ij} = \hat{\beta}_0 + \hat{\beta}_2 \times TV_i.$$

The equations for similar students from a school with only a social-resistance classroom curriculum implemented ($CC = 1, TV = 0$), and from a school with both interventions implemented ($TV = 1, CC = 1$) are

$$\hat{\eta}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 \times CC_i + \hat{\beta}_4 \times \text{PreTHKS}_{ij}$$

and

$$\hat{\eta}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 \times CC_i + \hat{\beta}_2 \times TV_i + \hat{\beta}_3 \times CC_i \times TV_i + \hat{\beta}_4 \times \text{PreTHKS}_{ij}$$

respectively.

For a student with an average PreTHKS score (2.152, see exploratory analysis) from any school with similar values of CC and TV we find that

$$\begin{aligned}\hat{\eta}_{ij} &= \hat{\beta}_0 + \hat{\beta}_4 * \text{PreTHKS}_{ij} \\ &= \hat{\beta}_0 + \hat{\beta}_4 * 2.152.\end{aligned}$$

Using the $\hat{\beta}_0$ and $\hat{\beta}_4$ estimates of -1.2280 and 0.3870 respectively as obtained for the current analysis, we can calculate the estimated probability of a THKSbin score of 1 for typical students with PreTHKS scores of 2.152 and 0 respectively as

$$\begin{aligned} \text{Prob}(\text{THKSbin}_{ij} = 1 | \text{CC} = \text{TV} = 0; \text{PreTHKS} = 2.152) &= \frac{e^{-1.2280+0.3870(2.152)}}{1 + e^{-1.2280+0.3870(2.152)}} \\ &= \frac{e^{-0.39518}}{1 + e^{0.39518}} \\ &= 0.40247 \end{aligned}$$

and

$$\begin{aligned} \text{Prob}(\text{THKSbin}_{ij} = 1 | \text{CC} = \text{TV} = \text{PreTHKS} = 0) &= \frac{e^{-1.2280}}{1 + e^{-1.2280}} \\ &= 0.22653. \end{aligned}$$

A student with an average observed score of PreTHKS is almost twice as likely to have a THKSbin score of 1 as a student with the lowest observed score on the same variable. Note that we opted to use the mean pre-intervention score for this specific subgroup.

On the other end of the scale in terms of intervention, we have schools where both a social-resistance classroom curriculum and a mass-media intervention were implemented (CC = TV = 1). For two typical students from these schools, an observed PreTHKS score of 0 or the mean score of 1.979 will imply a predicted probability of a THKSbin score of 1 of 0.4201 for the first and 0.6091 for the second. Again, the higher the pre-intervention score, the higher the predicted probability of a high post-intervention score.

In Table 4.4, the estimated probabilities of high post-intervention scores on the tobacco and health questionnaire are given for typical students with high or low pre-intervention scores for each of the subpopulations formed by mass-media intervention and implementation of social-resistance classroom curriculum.

Table 4.4: Estimated unit-specific probability of a high post-intervention knowledge score

Group	prescore	prob.	prescore	prob.
CC = 0, TV = 0	0	22.65%	2.152	40.25%
CC = 1, TV = 0	0	46.54%	2.05	65.81%
CC = 0, TV = 1	0	29.86%	2.87	48.85%
CC = 1, TV = 1	0	42.01%	1.979	60.91%

These estimated probabilities can also be presented graphically, as shown in the bar chart below.

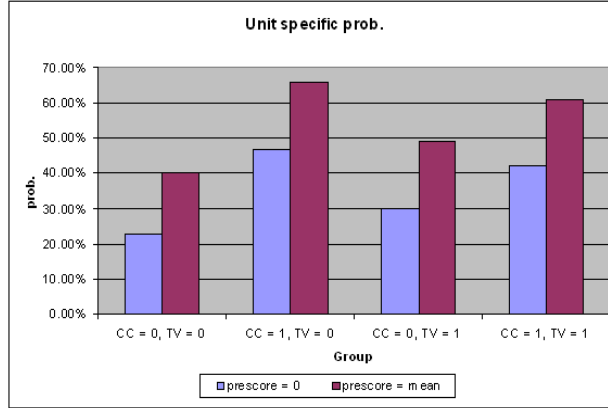


Figure 4.3: Bar chart of estimated unit-specific probabilities

Students with a high pre-intervention score were predicted to have a high post-intervention score too, regardless of the study conditions. Similarly, students with a low pre-intervention score were generally likely to have a low post-intervention score too. If only curriculum intervention (CC = 1) was used, scores for students were likely to be higher regardless of their pre-intervention scores. On both ends of the pre-intervention knowledge score scale, in groups where mass-media intervention was used (TV = 1), scores were predicted to be higher than where media intervention was not used, except when both mass-media and curriculum intervention were used. For these groups, with CC = TV = 1, the estimated probabilities of a high post-intervention score were actually lower than for the group where only a classroom curriculum was used (42.01% vs. 46.54%, and 60.91% vs. 65.80%).

We conclude that for most students, the implementation of a social-resistance classroom curriculum is more likely to be effective in increasing their knowledge (predicted probabilities of a high score being 46.54% and 65.80% respectively) than mass-media intervention (predicted probabilities of a high score being 29.86% and 48.85% respectively). The control group, where neither method was implemented, had the lowest predicted knowledge scores (22.65% and 40.25% respectively). While the implementation of both procedures was associated with higher probabilities than either the control group or the group where only mass-media intervention was used, its predicted gain was disappointing when compared to the use of only social-resistance curriculum implementation. Generally speaking, the implementation of a curriculum only seems to be most effective in increasing the predicted knowledge of students on the tobacco and health questionnaire.

Estimated outcomes for different groups: population-average results

In the introduction to this section, we defined the latent response variable model as

$$y_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{v}_i + e_{ij}, \quad j = 1, 2, \dots, n_i$$

where \mathbf{z}'_{ij} denotes a design vector for the random effects contained in the vector \mathbf{v}_i , and \mathbf{x}'_{ij} the design vector for the predictors in the fixed part of the model with corresponding vector

β of regression parameters. The covariance matrix of \mathbf{v}_i is denoted by $\Phi_{(v)}$ and the variance of e_{ij} by σ_e^2 .

For a probit link function $\sigma_e^2=1$, and for a logistic link function it is assumed to be $\sigma_e^2 = \pi^2/3$. Under the assumption that \mathbf{v}_i and e_{ij} are independently distributed, it follows that

$$\sigma_{y_{ij}}^2 = \mathbf{z}'_{ij} \Phi_{v_i} \mathbf{z}_{ij} + \sigma_e^2.$$

The design effect d_{ij} is defined in terms of σ_e^2 and $\sigma_{y_{ij}}^2$:

$$d_{ij} = \frac{\sigma_{y_{ij}}^2}{\sigma_e^2}.$$

This design effect may be used to obtain the estimated population-average probabilities in a similar fashion as the unit-specific probabilities, but with replacing $\hat{\eta}_{ij}$ with $\hat{\eta}_{ij}^* = \hat{\eta}_{ij} / \sqrt{d_{ij}}$ (Hedeker & Gibbons, 2006).

We can compare these estimated population-average probabilities with the observed data for the four groups formed by the categories of TV and CC as shown in Table 4.5. To illustrate, we calculate the estimated population-average probabilities for a few of the subgroups.

From the output, we have $\text{var}(v_{i0}) = 0.1065$, where v_{i0} denotes the random intercept coefficient. In this case, $\mathbf{z}'_{ik} = \mathbf{1}$ and hence, with $\sigma_e^2 = \pi^2/3$ for the logistic link,

$$\sigma_{y_{ij}}^2 = 1 \times 0.1065 \times 1 + 3.2899 = 3.3964.$$

Therefore

$$d_{ij} = \frac{3.3964}{3.2899} = 1.0324.$$

To obtain the population-average probability estimates, we now replace the $\hat{\eta}_{ij}$ values calculated for the unit-specific case with $\hat{\eta}_{ij}^* = \hat{\eta}_{ij} / \sqrt{d_{ij}}$.

For the subgroup where TV = CC = 0 and the mean PreTHKS value is equal to 2.152, for example, we find that

$$\begin{aligned}\hat{\eta}_{ij} &= -1.2281 + 0.3871(2.152) \\ &= -0.39506\end{aligned}$$

so that

$$\begin{aligned}\hat{\eta}_{ij}^* &= -0.39506 / \sqrt{1.0324} \\ &= -0.38881\end{aligned}$$

and

$$\begin{aligned}P(\text{THKSbin}_{ij} = 1 \mid \text{CC} = \text{TV} = 0, \text{PreTHKS} = 2.152) &= \frac{e^{\hat{\eta}_{ij}^*}}{1 + e^{\hat{\eta}_{ij}^*}} \\ &= \frac{0.67786}{1.67786} = 40.40\%.\end{aligned}$$

Similarly, for the group where $\text{TV} = \text{CC} = 0$ and $\text{PreTHKS} = 0$, we find that

$$\begin{aligned}\hat{\eta}_{ij} &= -1.2281 \\ \hat{\eta}_{ij}^* &= -1.2281 / 1.01606 \\ &= -1.2087.\end{aligned}$$

Table 4.5: Estimated population-average probabilities

Group	prescore	prob.	prescore	prob.
CC = 0, TV = 0	0	22.99%	2.15	40.40%
CC = 1, TV = 0	0	46.59%	2.05	65.57%
CC = 0, TV = 1	0	30.14%	2.87	48.87%
CC = 1, TV = 1	0	42.13%	1.98	60.74%

A comparison of these probabilities with the observed ratios given in Table 4.6 for the control group at the end of the study indicates that the population-average results are slightly closer to the observed ratios than is the case for the unit-specific results. Recall that $\sqrt{d_{ij}} = 1.0161$. The extent of differences between unit-specific and population-average results is highly dependent on the "scaling" induced by dividing the $\hat{\eta}_{ij}$'s by $\sqrt{d_{ij}}$.

Table 4.6: Observed and predicted proportions of high post-intervention scores

Group	Proportion observed	Unit-specific predicted prob.	Population-average predicted prob.
CC = 0, TV = 0	41.57%	40.25%	40.40%
CC = 1, TV = 0	63.16%	65.80%	65.57%
CC = 0, TV = 1	48.32%	48.84%	48.86%
CC = 1, TV = 1	60.31%	60.91%	60.74%

3.2.2.5 Interpreting the contents of the level-2 residual file

In addition to the standard output file, the **Write Bayes Estimates** field on the **Configuration** tab of the **Model Setup** dialog was used to request Bayes estimates for the individual random terms. These estimates are written to the file **TVBS.ba2**. The first few lines of this file are shown below.

Four pieces of information per school are given:

- all 1s for the level-2 model,
- the school's ID,
- the value of random intercept,
- the empirical Bayes estimate,
- the associated posterior variance for the school estimate, and
- the name of the associated random coefficient.

School ID	TV	CC	Intercept	Label
403.00	1	0.2892868	0.0747025	intercept
404.00	1	0.1160913	0.0660764	intercept
193.00	1	0.0894222	0.0651482	intercept
194.00	1	0.0983612	0.0393094	intercept
196.00	1	0.2471524	0.0661965	intercept
197.00	1	0.0137577	0.0529062	intercept
198.00	1	-0.2770740	0.0466224	intercept
199.00	1	-0.1873500	0.0453684	intercept
401.00	1	0.3180534	0.0581188	intercept
402.00	1	-0.0835122	0.0592211	intercept
405.00	1	-0.0581176	0.0463815	intercept
407.00	1	0.4108713	0.0408447	intercept
408.00	1	-0.1503098	0.0664666	intercept
409.00	1	0.2302431	0.0354570	intercept
410.00	1	-0.4221328	0.0585741	intercept
411.00	1	0.2729730	0.0756187	intercept
412.00	1	0.0151462	0.0587242	intercept
414.00	1	-0.1034631	0.0556605	intercept
415.00	1	0.3659302	0.0440425	intercept
505.00	1	-0.2049965	0.0377875	intercept
506.00	1	-0.4734572	0.0411202	intercept
507.00	1	-0.2184398	0.0381874	intercept
508.00	1	0.2534393	0.0358827	intercept
509.00	1	-0.1449230	0.0281350	intercept
510.00	1	-0.1526762	0.0291762	intercept
513.00	1	-0.3506720	0.0621859	intercept
514.00	1	0.1042555	0.0329034	intercept
515.00	1	0.0021403	0.0244141	intercept

The mean of the empirical Bayes estimates is -0.0002 . The estimates ranged from -0.473614 for school 506 to 0.4110043 for school 407. In both cases a mass-media intervention procedure was applied, and thus $TV = 1$, but $CC = CC*TV = 0$. For students with a PreTHKS score of 3 from each of these schools, this implies

$$\begin{aligned} \text{Prob}(\text{THKSbin}_{ij} = 1 \mid CC = 0, \text{PreTHKS} = 3, \text{ID} = 506) &= \frac{e^{-0.473614+0.3741+0.3870(3)}}{1 + e^{-0.473614+0.3741+0.3870(3)}} \\ &= \frac{e^{1.061486}}{1 + e^{1.061486}} = 0.7430 \end{aligned}$$

and

$$\begin{aligned} \text{Prob}(\text{THKSbin}_{ij} = 1 \mid CC = 0, \text{PreTHKS} = 3, \text{ID} = 407) &= \frac{e^{0.4110043+0.3741+0.3870(3)}}{1 + e^{0.4110043+0.3741+0.3870(3)}} \\ &= \frac{e^{1.9461043}}{1 + e^{1.9461043}} = 0.8750 \end{aligned}$$

respectively. The fact that the intercept for school 407 lies higher than the average is reflected in the higher probability (87.5%) that a student with average pre-intervention knowledge score will obtain a high post-intervention score. School 506, on the other hand, has an intercept far below the average, and a student from this school has, in effect, a 74.30% chance of obtaining a high post-intervention score.

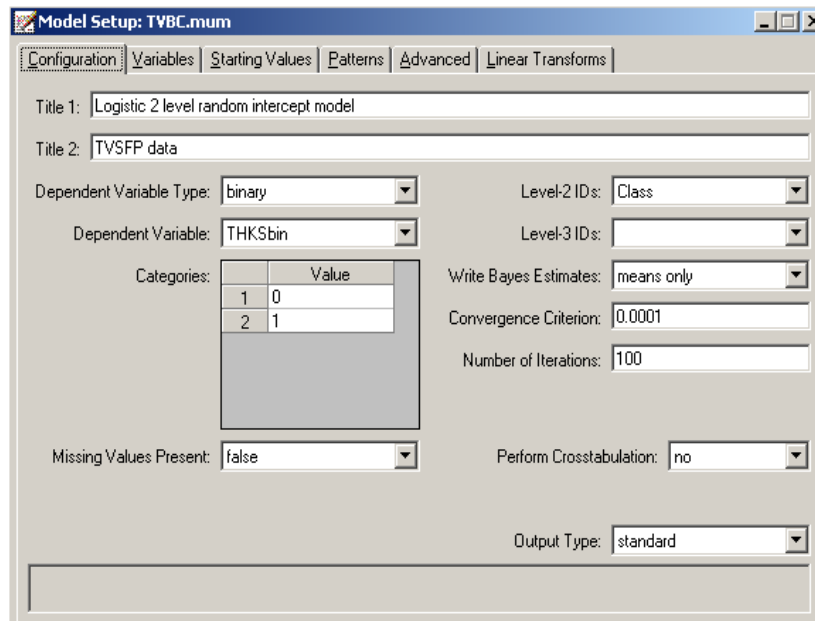
3.2.3 A 2-level random intercept logistic regression model

Using the same data (**tvspors.ss3**) and model setup file **TVBS.mum** from the previous example, we now consider the situation where students are nested within classrooms and fit a two-level model of the form described earlier, again with the binary variable THKSbin as outcome.

3.2.3.1 Setting up the analysis

Use the **File, Open Spreadsheet** option to re-open the previously used spreadsheet **tvspors.ss3** from the **Examples\Binary** folder. Next, use the **File, Open Existing Model Setup** option to browse and open the syntax file **TVBS.mum**.

The biggest change to be made to the syntax file is in terms of the ID variable. Change the **Level-2 IDs** field on the **Configuration** tab of the **Model Setup** dialog box from **School** to **Class**, as shown below. Also, turn off the writing out of Bayes estimates by setting the **Write Bayes Estimates** field to **no**.

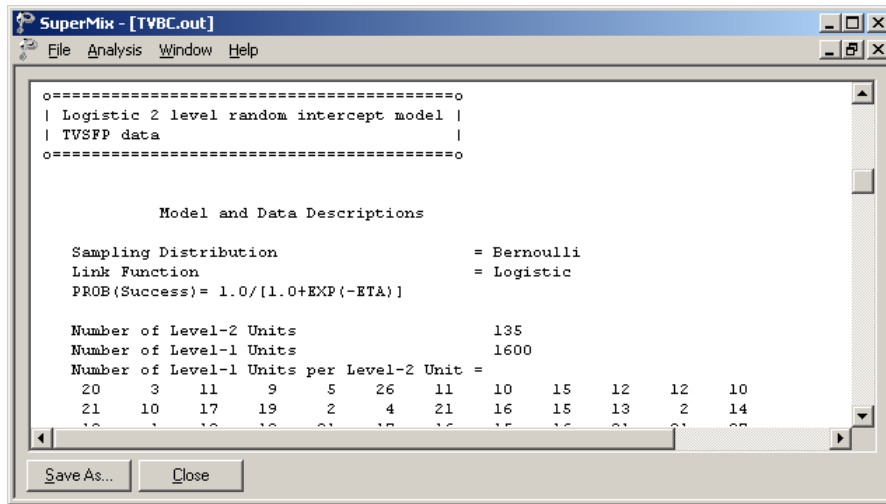


Save the revised syntax file under a new name such as **TVBC.mum** and run the analysis.

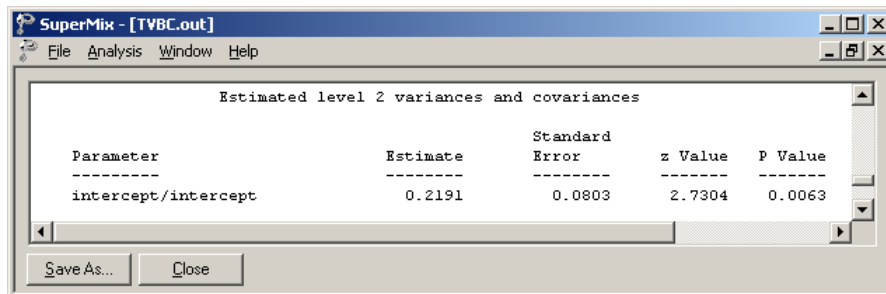
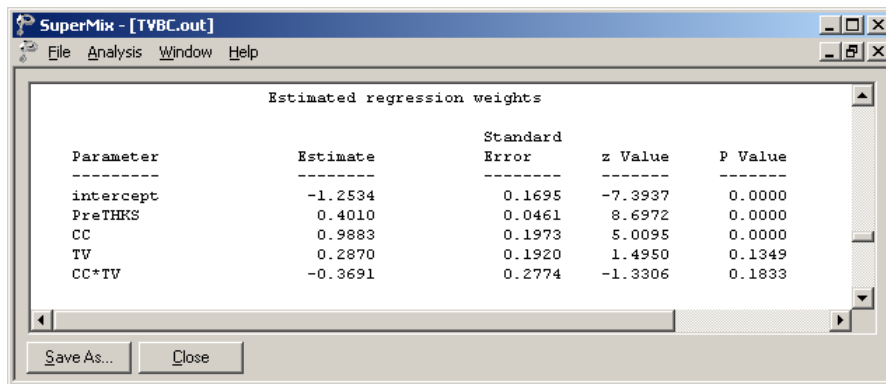
3.2.3.2 Discussion of results

Partial output for this run is provided below. The summary of units now reflects the number of students nested within each classroom. The number of students per class (level-2 unit)

ranges between 2 and 28. In this analysis, there were 135 level-2 units, compared to 28 in the previous analysis.



Estimated coefficients with adaptive quadrature and the estimated level-2 variances are given below.



The estimates for the classroom analysis are very similar to those of the school analysis. All estimated fixed coefficients are slightly lower than was the case in the previous analysis.

There seems to be more variation between classrooms than between schools, as indicated by the estimated variation in the random intercept of 0.2193, compared to the similar estimate of 0.1065 in the school analysis.

The estimates can again be used to obtain predicted probabilities by first calculating the $\hat{\eta}_{ij}^*$ s, using the formulae

$$\hat{\eta}_{ij} = -1.2535 + 0.9883 \times CC_i + 0.2870 \times TV_i - 0.369 \times (CC \times TV)_i + 0.401 \times PreTHKS_{ij}$$

and $\hat{\eta}_{ij}^* = \hat{\eta}_{ij} / \sqrt{d_{ij}}$ where

$$d_{ij} = \frac{\sigma_{y_{ij}}^2}{\sigma_e^2} = \frac{0.2193 + \pi^2 / 3}{\pi^2 / 3} = \frac{0.2193 + 3.289865}{3.289865} = 1.0666.$$

A comparison of unit-specific and population-average predicted probabilities for the current model are given in Table 4.7. For comparison purposes, similar results for the previous model can be found in Table 4.7.

Table 4.7: Observed and predicted proportions of high post-intervention scores

Group	Proportion observed	Unit-specific predicted prob.	Population-average predicted prob.
CC = 0, TV = 0	41.57%	40.36%	40.66%
CC = 1, TV = 0	63.16%	63.57%	63.16%
CC = 0, TV = 1	48.32%	46.76%	46.87%
CC = 1, TV = 1	60.31%	60.98%	60.64%

3.2.4 A 3-level random intercept logistic regression model

Having fitted 2-level models where students were nested within either classrooms or schools thus far, we now consider a 3-level model with both classroom and school defining levels of the hierarchy.

3.2.4.1 The model

The level-1 and level-2 models are the same as for the previous two models, as shown below.

Level 1 model ($k = 1, \dots, n_{ij}$):

$$\text{THKSbin}_{ijk} = b_{0ij} + b_{1ij} \text{PRETHKS}_{ijk} + e_{ijk}$$

Level-2 model ($j = 1, \dots, n_i$):

$$b_{0ij} = b_{00i} + b_{01i} \text{CC}_{ij} + b_{02i} \text{TV}_{ij} + b_{03i} (\text{CC}_{ij} \times \text{TV}_{ij}) + v_{0ij}$$
$$b_{1ij} = b_{10i}$$

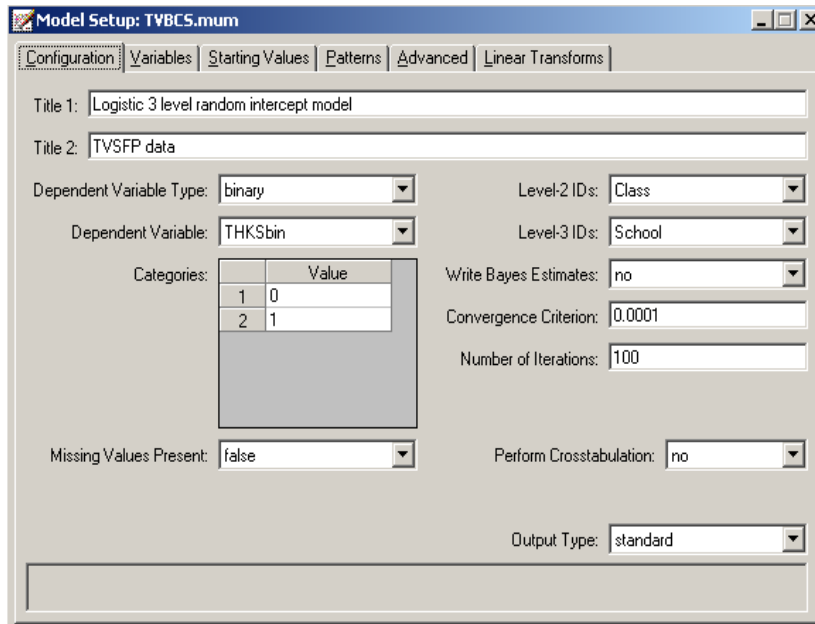
With classrooms nested within schools, however, a third level of the hierarchy is defined. At this level, the level-2 coefficients become outcomes again, and can potentially vary over the schools (level-3 units). In the current model, we allow only the intercept to vary randomly over the schools.

Level-3 model ($i = 1, \dots, N$)

$$b_{00i} = \beta_0 + v_{0i}$$
$$b_{01i} = \beta_1$$
$$b_{02i} = \beta_2$$
$$b_{03i} = \beta_3$$
$$b_{10i} = \beta_4$$

3.2.4.2 Setting up the analysis

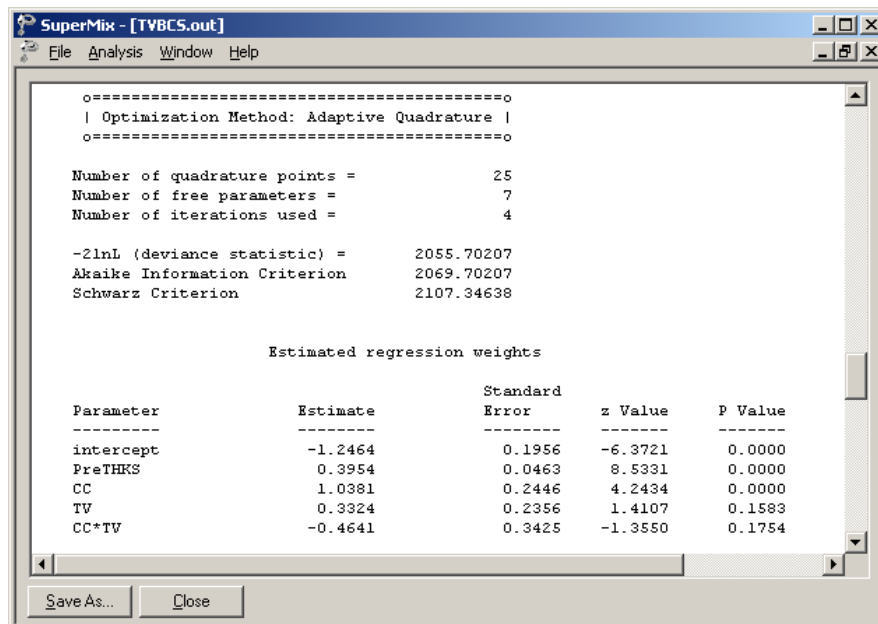
We modify our model setup saved to the syntax file **TVBS.mum** by first using the **Open Existing Model Setup** option on the **File** menu to retrieve the syntax file. Then click on **File, Save** as to save the model setup in a new file, such as **TVBCS.mum**. Next, select CLASS as the **Level-2 ID** and SCHOOL as the **Level-3 IDs** as shown below. We now have both level-2 and level-3 IDs selected.



Keep all the other settings unchanged. Save the changes to the file **TVBCS.mum** and select the **Run** option on the **Analysis** menu to run the analysis.

3.2.4.3 Discussion of results

The portions of the output file **TVBCS.out** containing the estimates of the fixed and random coefficients in the current model are shown below.



Parameter	Estimate	Standard Error	z Value	P Value
intercept/intercept	0.1650	0.0814	2.0271	0.0426

Parameter	Estimate	Standard Error	z Value	P Value
intercept/intercept	0.0627	0.0618	1.0145	0.3103

Table 4.8: Comparison of results for three models with binary variable THKSbin as outcome

Coefficient		2-level:	2-level:	3-level
		CLASS as ID	SCHOOL as ID	
Fixed effects:				
Intercept	estimate	-1.2535	-1.228	-1.2465
	standard error	0.1695	0.1949	0.1957
PRETHKS	estimate	0.401	0.3871	0.3954
	standard error	0.0461	0.0451	0.0463
CC	estimate	0.9883	1.0893	1.0383
	standard error	0.1973	0.2454	0.2448
TV	estimate	0.287*	0.3741*	0.3325*
	standard error	0.192	0.235	0.2358
CCxTV	estimate	-0.369*	-0.5578*	-0.4644*
	standard error	0.2774	0.3403	0.3427
Random effects:				
Var(between classrooms)	estimate	0.2193		0.1649
	standard error	0.0802		0.0813
Var(between schools)	estimate		0.1065	0.063*
	standard error		0.0578	0.0616

*: Not significant at 5% level of significance.

Results for this model are compared to those obtained using the two 2-level models in Table 4.8. Generally, there is close agreement between the models in terms of both the sign and size of the effects. Note that the only intervention method that consistently has an estimated coefficient significantly different from zero is CC. While use of the media intervention (TV) can positively influence the post-intervention score, it seems clear that using both methods simultaneously does not have any real benefits.

3.2.4.4 Interpreting the adaptive quadrature results

3-level ICCs

Intraclass correlation coefficients can be obtained for the three-level dichotomous outcome model. As mentioned earlier, it is assumed that the level-1 error variance is equal to $\pi^2/3$ for the logistic link function if the model is true (see, *e.g.*, Hedeker & Gibbons (2006), p. 157). Using this approximation, the formulae for the standard ICCs can be adjusted.

From the output for the random effects, we have

$$\text{Level-1: estimated (error var)} = \pi^2/3 = 3.2899$$

$$\text{Level-2: estimated (class var)} = 0.1649$$

$$\text{Level-3: estimated (school var)} = 0.0630.$$

Based on this information, we can calculate the ICC as shown below.

Similarity of students within the same school:

$$\begin{aligned} ICC &= \frac{\sigma_{v(3)}^2}{\sigma_{v(3)}^2 + \sigma_{v(2)}^2 + \sigma_e^2} = \frac{0.063}{0.063 + 0.1649 + 3.28986} \\ &= 0.0179. \end{aligned}$$

Similarity of students within the same classrooms (and schools):

$$\begin{aligned} ICC &= \frac{\sigma_{v(2)}^2}{\sigma_{v(3)}^2 + \sigma_{v(2)}^2 + \sigma_e^2} = \frac{0.1649}{0.063 + 0.1649 + 3.28986} \\ &= 0.04688. \end{aligned}$$

Similarity of classes within the same school:

$$ICC = \frac{\sigma_{v(2)}^2}{\sigma_{v(3)}^2 + \sigma_{v(2)}^2} = \frac{0.1649}{0.063 + 0.1649}$$

$$= 0.7236.$$

Estimated unit-specific and population-average probabilities

Under the assumption that \mathbf{v}_i , \mathbf{v}_{ij} and ε_{ijk} are independently distributed, it follows that for the three-level model the design effect is defined as

$$d_{ijk} = \frac{(\sigma_{v(3)}^2 + \sigma_{v(2)}^2 + \sigma_e^2)}{\sigma_e^2} = 1.0692.$$

The estimated unit-specific probabilities are calculated using

$$\hat{\eta}_{ijk} = -1.2465 + 1.0383 \times CC_i + 0.3325 \times TV_i - 0.4.644 \times CC_i \times TV_i$$

$$+ 0.3954 \times \text{PreTHKS}_{ijk}$$

and

$$\text{Prob}(\text{THKSbin} = 1 | \boldsymbol{\beta}) = \frac{1}{1 + e^{-\eta_{ijk}}}$$

The estimated population-average probabilities (Hedeker & Gibbons, 2006) are obtained in a similar fashion as the unit-specific probabilities after replacing $\hat{\eta}_{ijk}$ with $\hat{\eta}_{ijk}^* = \hat{\eta}_{ijk} / \sqrt{d_{ijk}}$ in the second of the equations shown above.