2 Models for binary outcomes

2.1 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.

2.1.1 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 $\mathbf{x} = \left(x_1 \cdots x_n\right)$ is the vector of all the predictors and that $\boldsymbol{\beta} = \left(\beta_1 \cdots \beta_n\right)$ 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

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

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 $\Phi_{(2)}$ and the variance of ε_{ij} by σ_{ε}^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 XXX.1 shows these link functions, along with their distribution functions (CDF), means and variances.

Table XXX.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	$\log(-\log(1-p))$	$1\text{-}\exp(-\exp(w))$	-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 XXX.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.

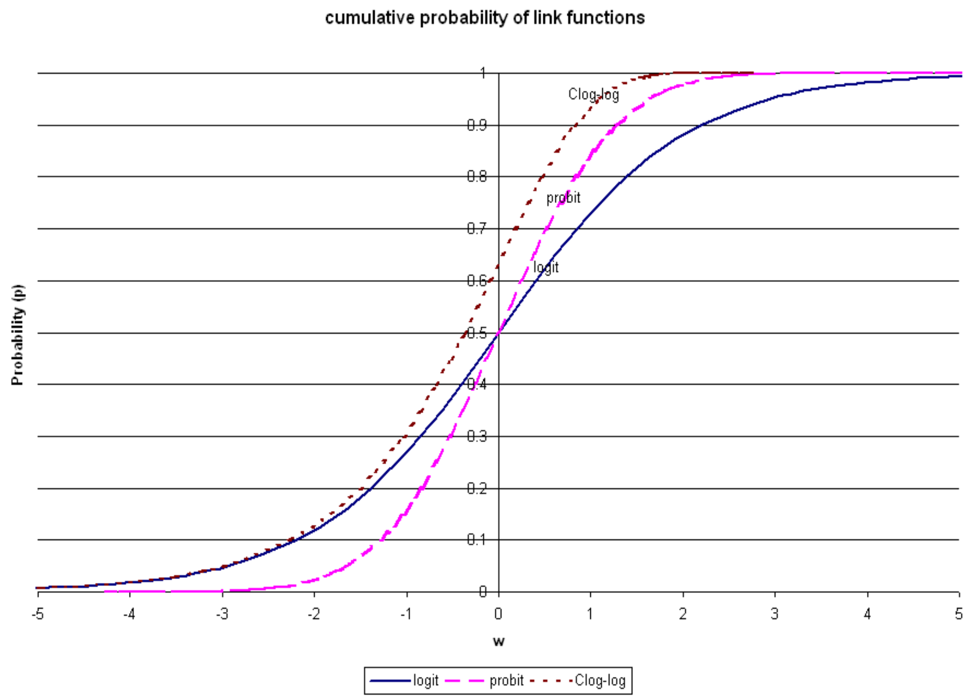


Figure XXX.1 Cumulative density of link functions

In this chapter, we will consider examples of two- and three-level models based on two data sets, both with binary outcome variables. These data will also be used to illustrate the ordinal outcome model in Chapter XXX.

2.1.2 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 full description of these methods is given in Chapter XXX. A brief description of MAP estimation and quadrature follows below.

2.1.2.1 MAP estimation

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

Using standard results for conditional distributions, it follows that

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

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

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

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

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

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

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

2.1.2.2 Numerical quadrature

Since

$$f\left(\mathbf{y}_i, \mathbf{v}_i\right) = f\left(\mathbf{y}_i \mid \mathbf{v}_i\right) f\left(\mathbf{v}_i\right)$$

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

$$f\left(\mathbf{y}_i\right) = \int_{v_1} \ldots \int_{v_r} f\left(\mathbf{y}_i \mid \mathbf{v}_i\right) f\left(\mathbf{v}\right) dv_1 \ldots . dv_r .$$

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

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

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

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

is approximated by the sum

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

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.

2.2 Models based on a subset of the TVSFP data

2.2.1 The data

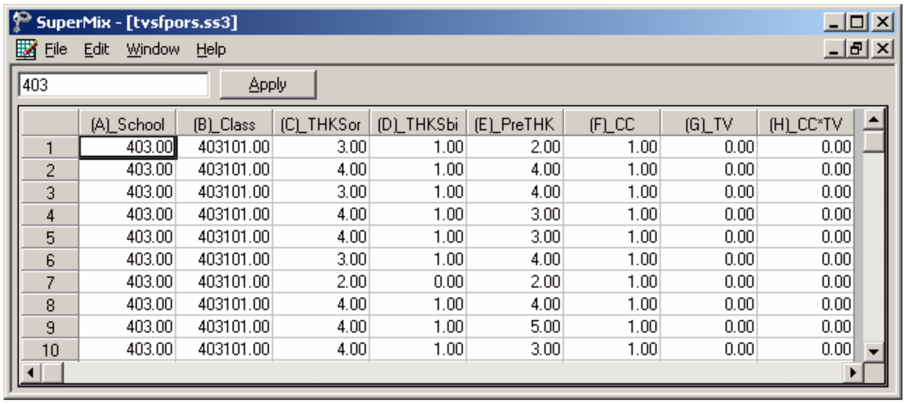
The data are from the Television School and Family Smoking Prevention and Cessation Project (TVSFP) study (Flay, *et. al.*, 1988). The study was designed to test the independent and combined effects of a school-based social-resistance curriculum and a television-based program in terms of tobacco use and cessation. The data from the study included a total of 1,600 students with both pre- and post-intervention scores from 135 classrooms drawn from 28 schools. Schools were randomized to one of four study conditions:

- A social-resistance classroom curriculum,
- A media (television) intervention,
- A social-resistance classroom curriculum combined with a mass-media intervention, and
- A no-treatment control group

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 **tvsfpors.ss3**, located in the **Examples\Binary** subfolder.



	(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

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 an 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.

2.2.1.1 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 XXX.2 below.

Table XXX.2: 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 XXX.2. 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 XXX.3 below.

Table XXX.3: Observed proportion of high post–intervention scores

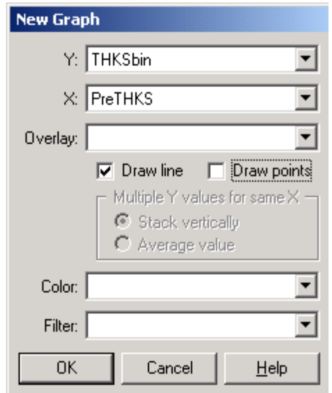
Study condition	Proportion	odds	logits
CC = 0, TV = 0	0.4157	0.711	-0.340
CC = 0, TV = 1	0.4832	0.935	-0.067
CC = 1, TV = 0	0.6316	1.714	0.539
CC = 1, TV = 1	0.6031	1.520	0.418

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 XXX.2. 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.

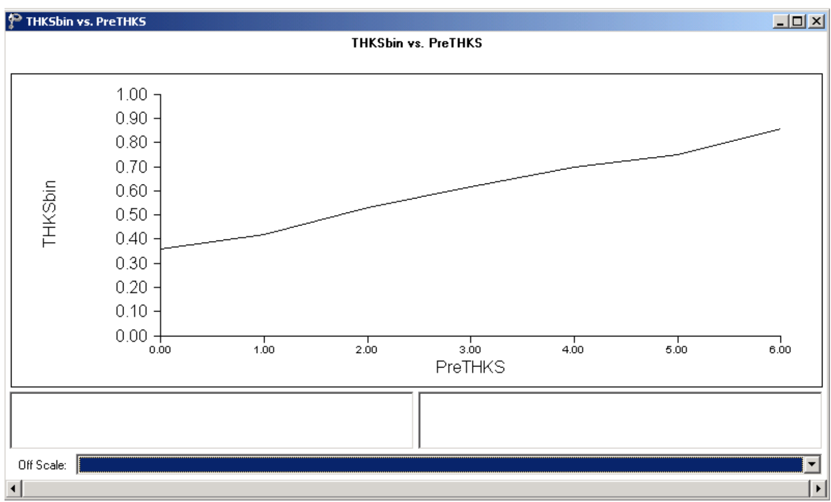
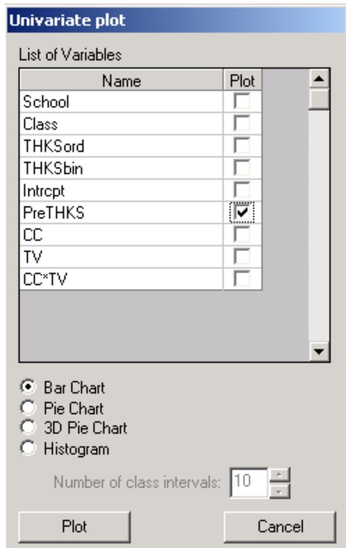


Figure XXX.2: Exploratory graph of THKSbin vs. PreTHKS

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.

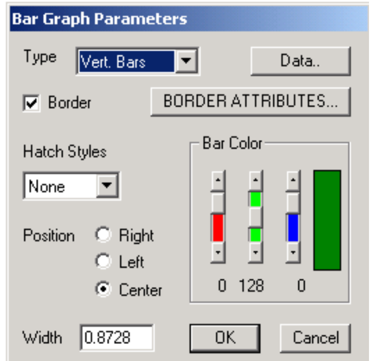


Figure XXX.3 below shows both the graphing window with bar chart and the data in spreadsheet format. Note that only 55 of the 1600 observations showed a score of 5 or higher, and that no student obtained a post-intervention score of 7 out of 7.

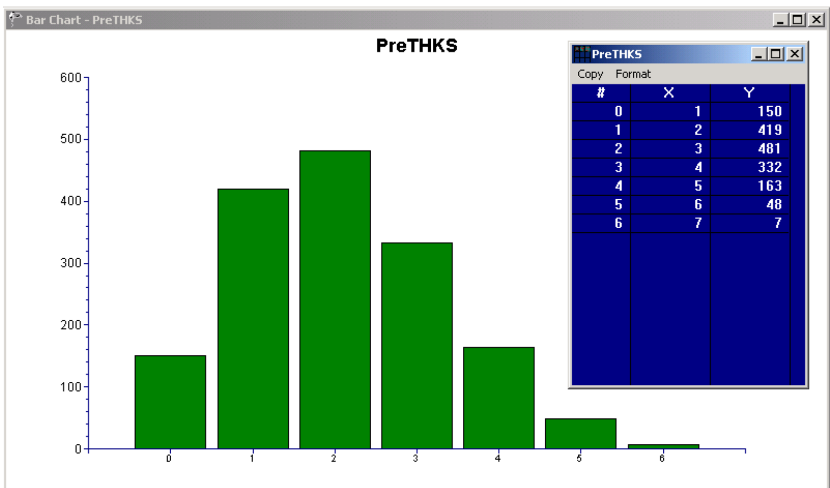


Figure XXX.3: Bar chart of PreTHKS values

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 XXX.4 below, and show that the mean pre-intervention scores do not differ much.

Table XXX.4: 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

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

2.2.2.1 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 2.1.1, 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:

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

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:

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

Level 2 model:

$$\begin{aligned} b_{0i} &= \beta_0 + v_{0i} \\ b_{1i} &= \beta_1 \\ b_{2i} &= \beta_2 \\ b_{3i} &= \beta_3 \\ b_{4i} &= \beta_4 \end{aligned}$$

An equivalent expression for the model is

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

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 $\beta_1, \beta_2, \ldots, \beta_4$ can be interpreted as follows:

- β_0 is the THKS logit for CC = 0, 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 CC = 0, TV = 0 subgroup.
- β_1 = the logit difference between (CC = 1, TV = 0) and (CC = 0, TV = 0) for the case where 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:

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

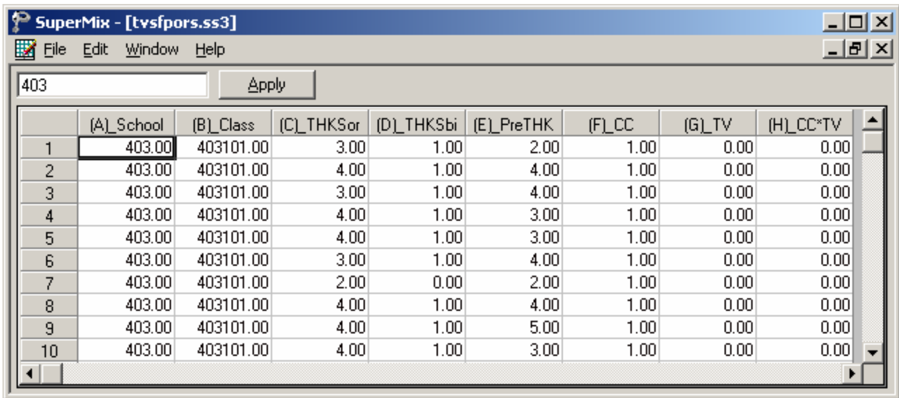
- β_3 is the difference in logit attributable to the interaction between the two intervention methods.

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

2.2.2.2 Setting up the analysis

Using the data in **tvsfpors.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 **tvsfpors.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.

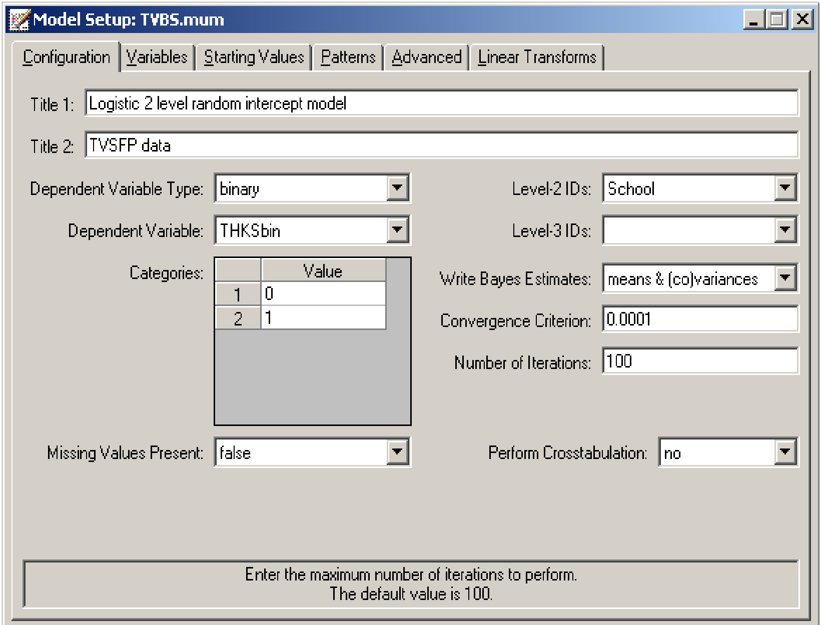


	(A)_School	(B)_Class	(C)_THKSor	(D)_THKSbi	(E)_PreTHK	(F)_CC	(G)_TV	(H)_CC*TV
1	403.00	403101.00	3.00	1.00	2.00	1.00	0.00	0.00
2	403.00	403101.00	4.00	1.00	4.00	1.00	0.00	0.00
3	403.00	403101.00	3.00	1.00	4.00	1.00	0.00	0.00
4	403.00	403101.00	4.00	1.00	3.00	1.00	0.00	0.00
5	403.00	403101.00	4.00	1.00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

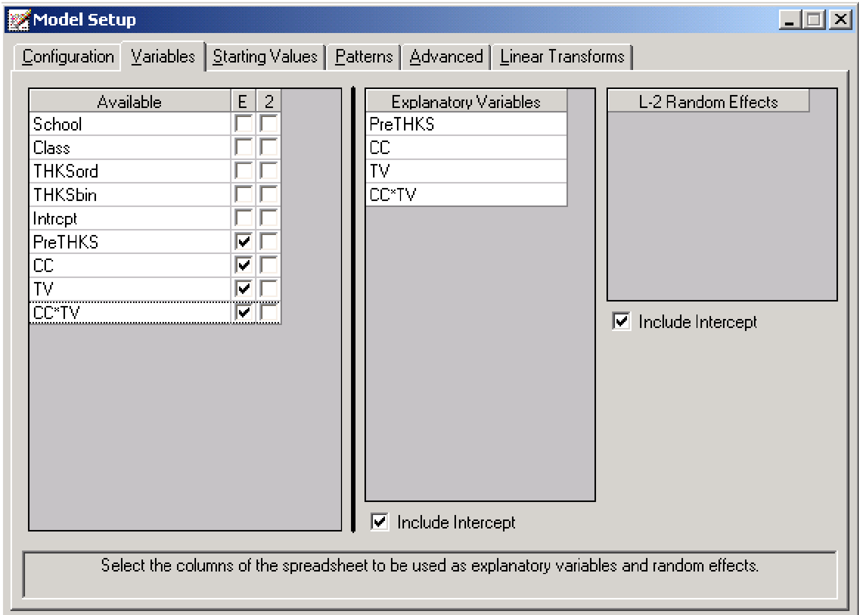
Next, we use the SuperMix interface to provide the model specifications. From the main menu bar, select the **File, New Model Setup** option.

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.

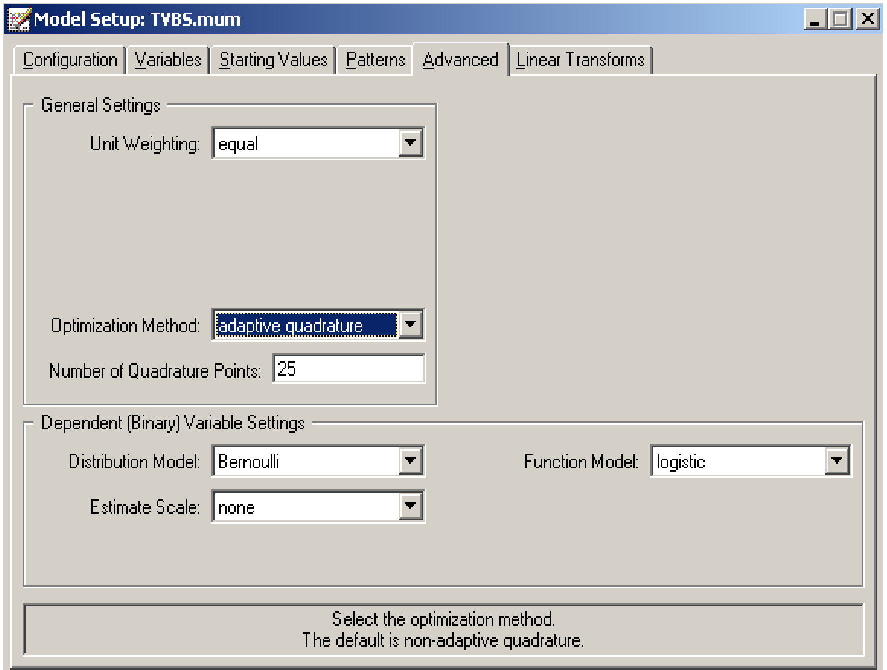
As a first step, the binary outcome variable THKSbin is selected from the **Dependent Variable** drop-down list box. The type of outcome is specified as binary using the drop-down list box in the **Dependent Variable Type** field. Once this selection is made, the **Categories** field is displayed. The 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 **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.



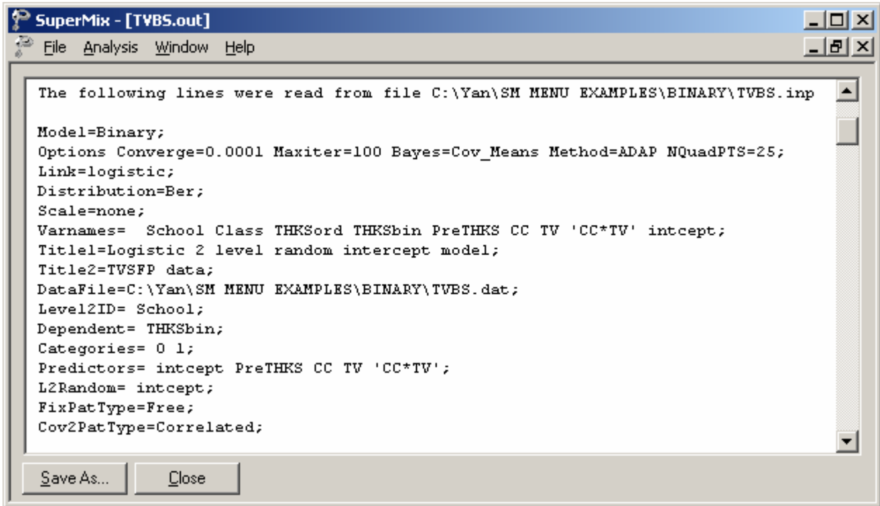
Before running the analysis, the model specifications have to be saved. Select the **File**, **Save** option, and provide a name for the model specification file, for example **TVBS.mum**. Run the analysis by selecting the **Run** option from the **Analysis** menu.

2.2.2.3 Discussion of results

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

Syntax

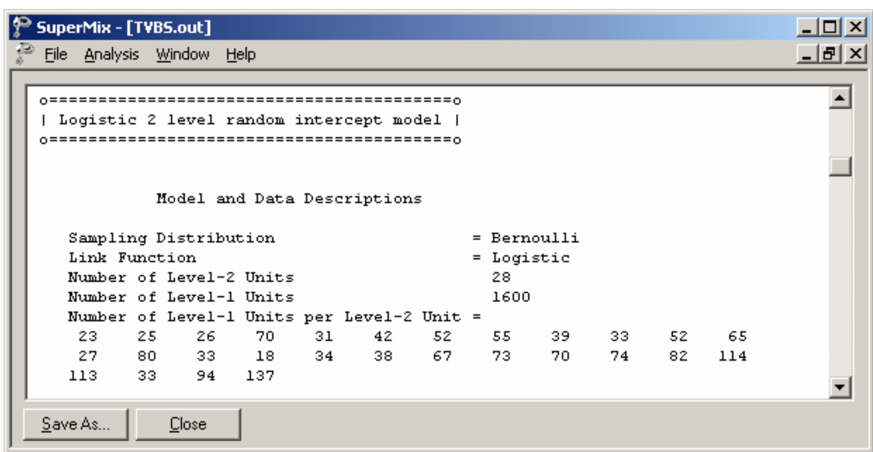
At the top of the file, the syntax saved to the **TVBS.mum** file is shown. The first part states the selection of iteration control options, requests for Bayes residuals, and the specifications necessary to define the model fitted as an binary model with a logistic link function. The second part of 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:\Yan\SM MENU EXAMPLES\BINARY\TVBS.inp

Model=Binary;
Options Converge=0.0001 Maxiter=100 Bayes=Cov_Means Method=ADAP NQuadPTS=25;
Link=logistic;
Distribution=Ber;
Scale=none;
Varnames=  School Class THKSord THKSbin PreTHKS CC TV 'CC*TV' intcept;
Title1=Logistic 2 level random intercept model;
Title2=TVSFP data;
DataFile=C:\Yan\SM MENU EXAMPLES\BINARY\TVBS.dat;
Level2ID= School;
Dependent= THKSbin;
Categories= 0 1;
Predictors= intcept PreTHKS CC TV 'CC*TV';
L2Random= intcept;
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.

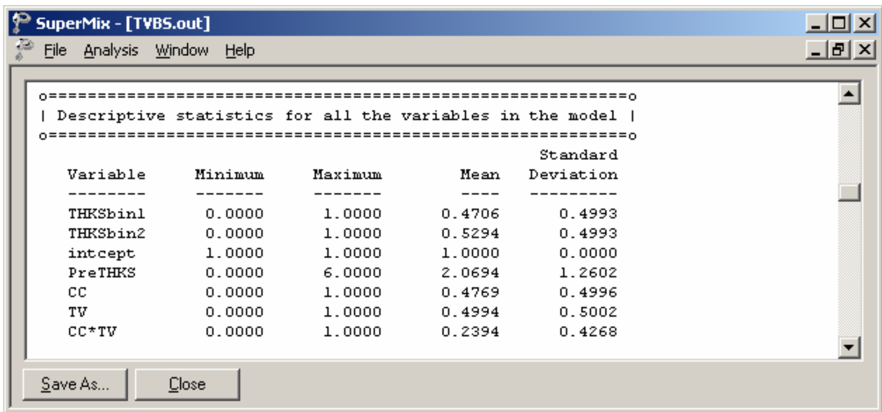



```
o========================================o
| Logistic 2 level random intercept model |
o========================================o


            Model and Data Descriptions

   Sampling Distribution                     = Bernoulli
   Link Function                             = Logistic
   Number of Level-2 Units                     28
   Number of Level-1 Units                     1600
   Number of Level-1 Units per Level-2 Unit =
    23     25     26     70     31     42     52     55     39     33     52     65
    27     80     33     18     34     38     67     73     70     74     82    114
   113     33     94    137
```

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.

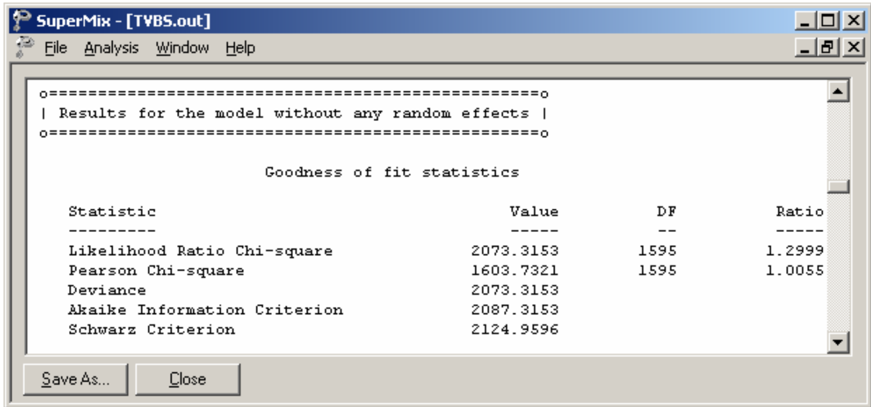


| Descriptive statistics for all the variables in the model |

Variable	Minimum	Maximum	Mean	Standard Deviation
THKSbin1	0.0000	1.0000	0.4706	0.4993
THKSbin2	0.0000	1.0000	0.5294	0.4993
intcept	1.0000	1.0000	1.0000	0.0000
PreTHKS	0.0000	6.0000	2.0694	1.2602
CC	0.0000	1.0000	0.4769	0.4996
TV	0.0000	1.0000	0.4994	0.5002
CC*TV	0.0000	1.0000	0.2394	0.4268

Results for the model without any random effects

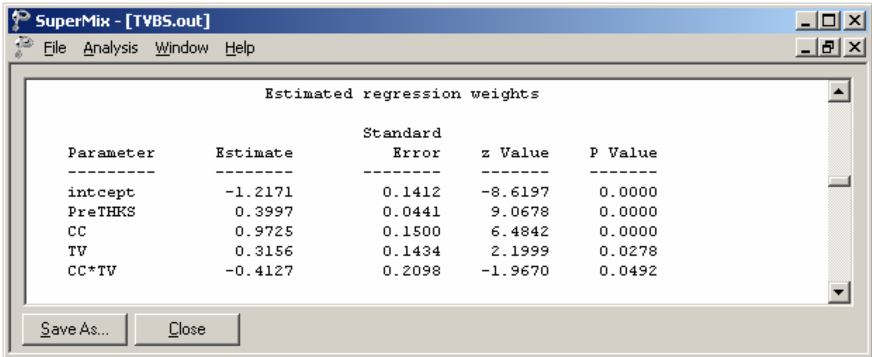
Descriptive statistics are followed by parameter estimates obtained under the assumption that all random effects are zero. The parameter values for the predictors CC, TV, CC*TV and PreTHKS are given in the first column, followed by the standard errors and *z*- and *p*-values.



| Results for the model without any random effects |

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
Deviance	2073.3153		
Akaike Information Criterion	2087.3153		
Schwarz Criterion	2124.9596		

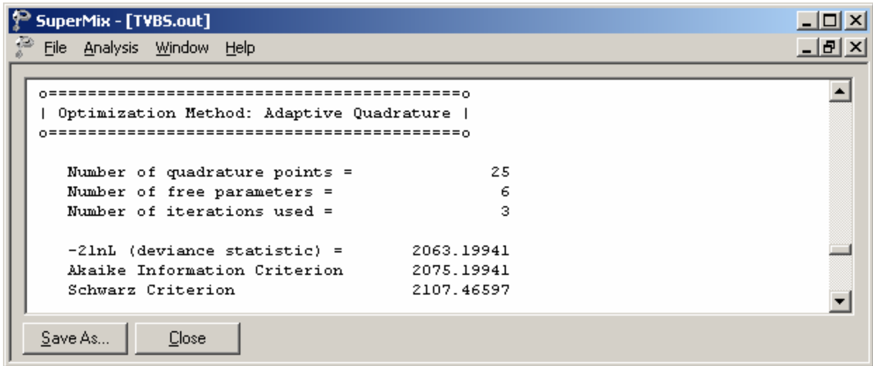


Estimated regression weights

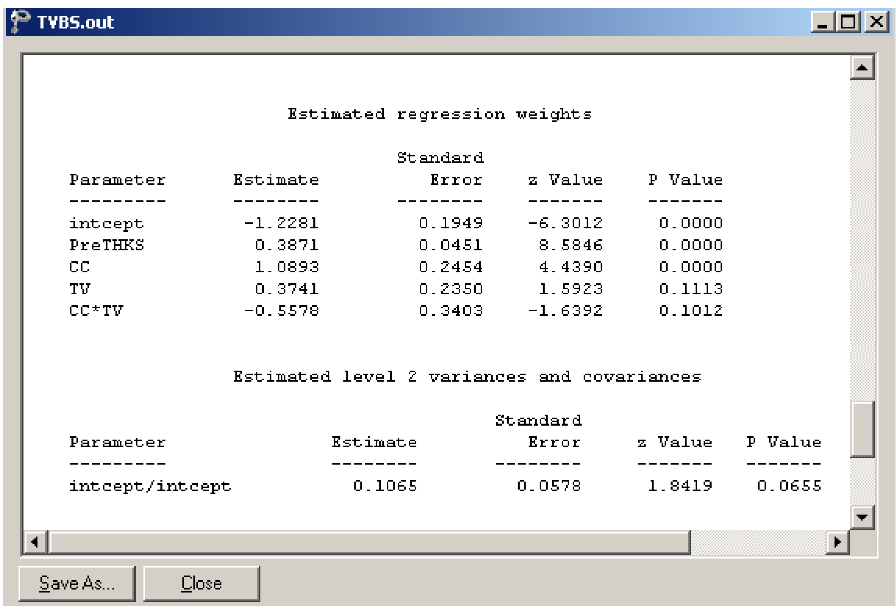
Parameter	Estimate	Standard Error	z Value	P Value
intcept	-1.2171	0.1412	-8.6197	0.0000
PreTHKS	0.3997	0.0441	9.0678	0.0000
CC	0.9725	0.1500	6.4842	0.0000
TV	0.3156	0.1434	2.1999	0.0278
CC*TV	-0.4127	0.2098	-1.9670	0.0492

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 quad 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 $\beta_0, \beta_1, \ldots, \beta_4$ 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.


TVBS.out

Estimated regression weights

Parameter	Estimate	Standard Error	z Value	P Value
intcept	-1.2281	0.1949	-6.3012	0.0000
PreTHKS	0.3871	0.0451	8.5846	0.0000
CC	1.0893	0.2454	4.4390	0.0000
TV	0.3741	0.2350	1.5923	0.1113
CC*TV	-0.5578	0.3403	-1.6392	0.1012

Estimated level 2 variances and covariances

Parameter	Estimate	Standard Error	z Value	P Value
intcept/intcept	0.1065	0.0578	1.8419	0.0655

Save As... Close

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.

2.2.2.4 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

$$\hat{\eta}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 \times \mathrm{CC}_i + \hat{\beta}_2 \times \mathrm{TV}_i + \hat{\beta}_3 \times \mathrm{CC}_i \times \mathrm{TV}_i + \hat{\beta}_4 \times \mathrm{PreTHKS}_{ij}$$

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),

$$\hat{\eta}_{ij} = \hat{\beta}_0 + \hat{\beta}_2 \times \mathrm{TV}_i.$$

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

$$\hat{\eta}_{ij} = \widehat{\beta}_0 + \widehat{\beta}_1 \times \mathrm{CC}_i + \widehat{\beta}_4 \times \mathrm{PreTHKS}_{ij}$$

and

$$\hat{\eta}_{ij} = \widehat{\beta}_0 + \widehat{\beta}_1 \times \mathrm{CC}_i + \widehat{\beta}_2 \times \mathrm{TV}_i + \widehat{\beta}_3 \times \mathrm{CC}_i \times \mathrm{TV}_i + \widehat{\beta}_4 \times \mathrm{PreTHKS}_{ij}$$

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

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

Using the $\widehat{\beta}_0$ and $\widehat{\beta}_4$ estimates of -1.2281 and 0.3871 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

$$\begin{aligned}\mathrm{Prob}(\mathrm{THKSbin}_{ij} = 1 \mid \mathrm{CC} = \mathrm{TV} = 0; \mathrm{PreTHKS} = 2.152) &= \frac{e^{-1.2281+0.3871(2.152)}}{1+e^{-1.2281+0.3871(2.152)}} \\ &= \frac{e^{-0.39506}}{1+e^{0.39506}} \\ &= 0.40250\end{aligned}$$

and

$$\begin{aligned}\mathrm{Prob}(\mathrm{THKSbin}_{ij} = 1 \mid \mathrm{CC} = \mathrm{TV} = \mathrm{PreTHKS} = 0) &= \frac{e^{-1.2281}}{1+e^{-1.2281}} \\ &= 0.22651.\end{aligned}$$

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 XXX.5, 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 XXX.5: 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.

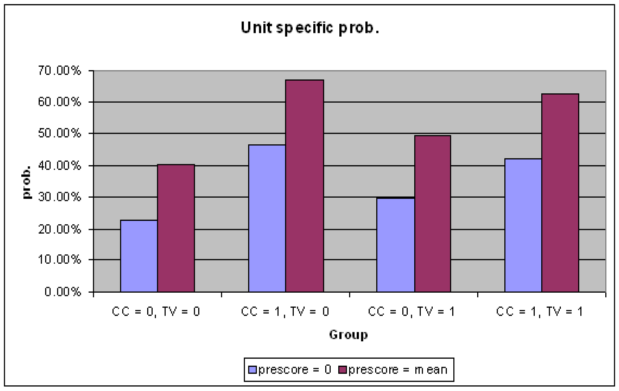


Figure XXX.4: Bar chart of estimated unit-specific probabilities

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. 65.81%).

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 65.81% 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

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

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 $\boldsymbol{\beta}$ of regression parameters. The covariance matrix of $\mathbf{v}_i$ is denoted by $\Phi_{(v)}$ and the variance of ε_{ij} by σ_{ε}^2.

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

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

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

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

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 XXX.6. To illustrate, we calculate the estimated population-average probabilities for a few of the subgroups.

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

$$P(\mathrm{THKSbin}_{ij} = 1 \mid \mathrm{CC} = \mathrm{TV} = 0, \mathrm{PreTHKS} = 2.152) = \frac{e^{\eta_{ij}^*}}{1+e^{\eta_{ij}^*}}$$
$$= \frac{0.67786}{1.67786} = 40.40\%.$$

Similarly, for the group where TV = CC = 0 and PreTHKS = 0, we find that

$$\hat{\eta}_{ij} = -1.2281$$

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

Table XXX.6: Estimated population-average probabilities

Group	prescore	prob.	prescore	prob.
CC = 0, TV = 0	0	22.99%	2.152	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%	1.979	60.74%

A comparison of these probabilities with the observed ratios given in Table XXX.7 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 XXX.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%	40.25%	40.40%
CC = 1, TV = 0	63.16%	65.81%	65.57%
CC = 0, TV = 1	48.32%	48.85%	48.87%
CC = 1, TV = 1	60.31%	60.91%	60.74%

2.2.2.5 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.

TVBS.ba2 - Notepad

File Edit Format View Help

1.00	403.00	1	0.2890865	0.0746508	intcept
1.00	404.00	1	0.1160136	0.0660369	intcept
1.00	193.00	1	0.0893923	0.0651102	intcept
1.00	194.00	1	0.0983555	0.0392957	intcept
1.00	196.00	1	0.2469961	0.0661559	intcept
1.00	197.00	1	0.0137773	0.0528812	intcept
1.00	198.00	1	-0.2769779	0.0466032	intcept
1.00	199.00	1	-0.1872501	0.0453496	intcept
1.00	401.00	1	0.3178720	0.0580872	intcept
1.00	402.00	1	-0.0834452	0.0591895	intcept
1.00	405.00	1	-0.0580730	0.0463621	intcept
1.00	407.00	1	0.4107440	0.0408293	intcept
1.00	408.00	1	-0.1502285	0.0664276	intcept
1.00	409.00	1	0.2301998	0.0354459	intcept
1.00	410.00	1	-0.4219231	0.0585430	intcept
1.00	411.00	1	0.2728001	0.0755663	intcept
1.00	412.00	1	0.0151303	0.0586932	intcept
1.00	414.00	1	-0.1034214	0.0556330	intcept
1.00	415.00	1	0.3657644	0.0440240	intcept
1.00	505.00	1	-0.2049495	0.0377749	intcept
1.00	506.00	1	-0.4732580	0.0411038	intcept
1.00	507.00	1	-0.2183398	0.0381740	intcept
1.00	508.00	1	0.2533768	0.0358708	intcept
1.00	509.00	1	-0.1449054	0.0281281	intcept
1.00	510.00	1	-0.1526582	0.0291689	intcept
1.00	513.00	1	-0.3504516	0.0621495	intcept
1.00	514.00	1	0.1042001	0.0328937	intcept
1.00	515.00	1	0.0021722	0.0244088	intcept

Ln 1, Col 1

The mean of the empirical Bayes estimates is – 0.0002. The estimates ranged from – 0.473258 for school 506 to 0.410744 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

$$\text{Prob}(\text{THKSbin}_{ij} = 1 \mid \text{CC} = 0, \text{PreTHKS} = 3, \text{ID} = 506) = \frac{e^{-0.473258+0.3741+0.3871(3)}}{1+e^{-0.473258+0.3741+0.3871(3)}}$$

$$= \frac{e^{1.062142}}{1+e^{1.062142}} = 0.7431$$

and

$$\text{Prob}(\text{THKSbin}_{ij} = 1 \,|\, \text{CC} = 0, \text{PreTHKS} = 3, \text{ID} = 407) = \frac{e^{0.410744+0.3741+0.3871(3)}}{1+e^{0.410744+0.3741+0.3871(3)}}$$

$$= \frac{e^{1.946144}}{1+e^{1.946144}} = 0.8750$$

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 74.31% chance of obtaining a high post-intervention score.

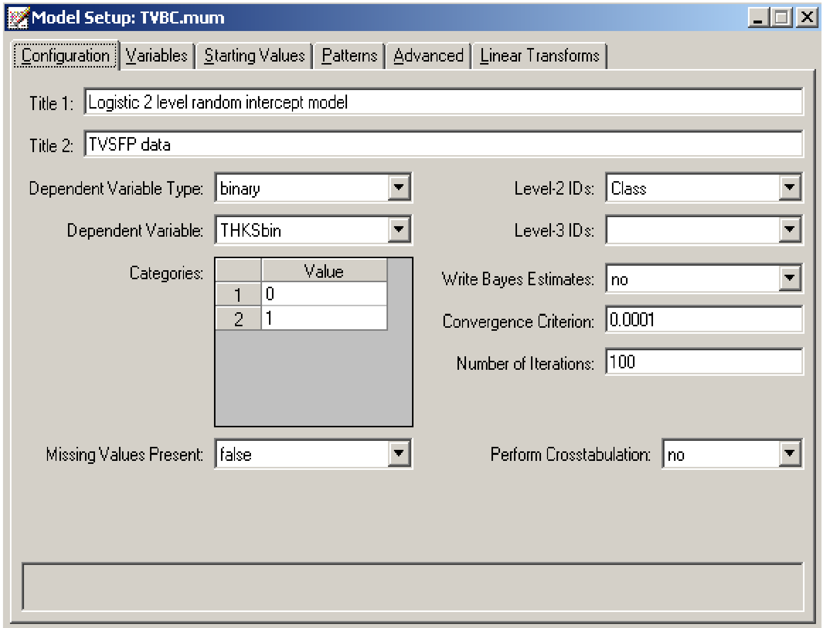
2.2.3 A 2-level random intercept logistic regression model

Using the same data (**tvsfpors.ss3)** and syntax 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.

2.2.3.1 Setting up the analysis

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

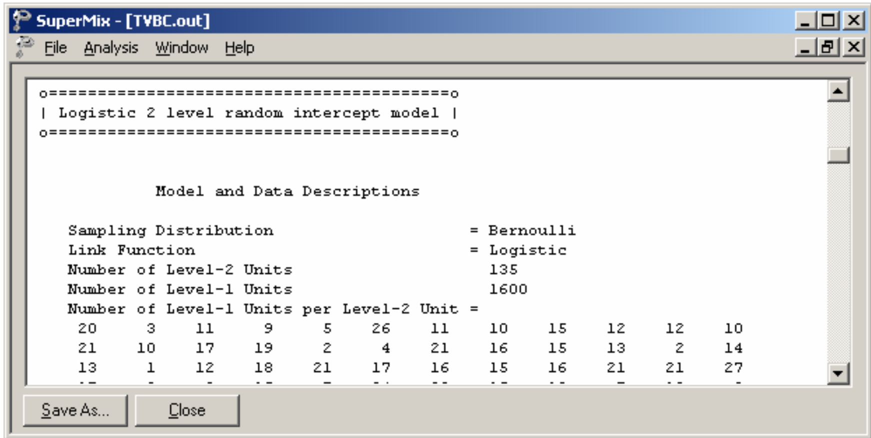
The biggest change to be made to the syntax file is in terms of the ID variable. Change the **Level-2 IDs** field on the **Configuration** tab of the **Model Setup** dialog box from School to Class, as shown below. Also, turn of the writing out of Bayes estimates by setting the **Write Bayes Estimates**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



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

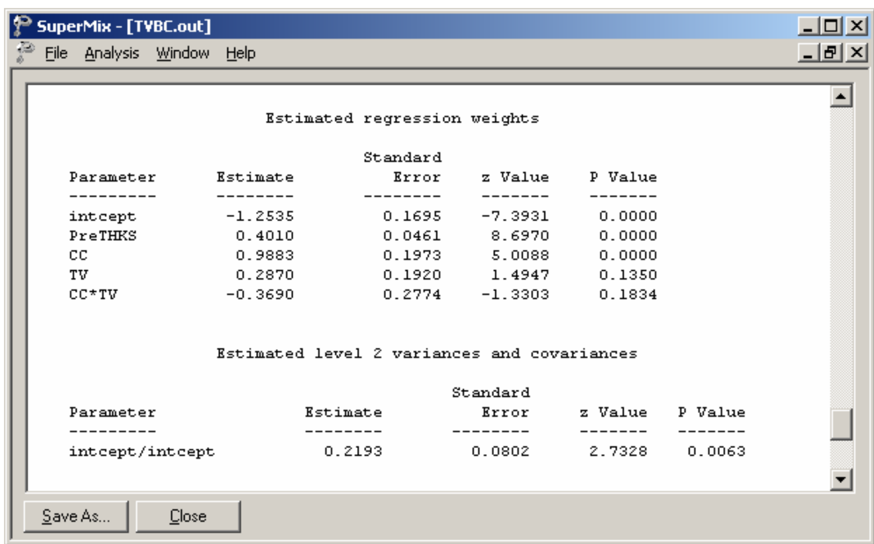
2.2.3.2 Discussion of results

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.




```
o========================================o
| Logistic 2 level random intercept model |
o========================================o


          Model and Data Descriptions

   Sampling Distribution                     = Bernoulli
   Link Function                             = Logistic
   Number of Level-2 Units                     135
   Number of Level-1 Units                     1600
   Number of Level-1 Units per Level-2 Unit =
     20     3    11     9     5    26    11    10    15    12    12    10
     21    10    17    19     2     4    21    16    15    13     2    14
     13     1    12    18    21    17    16    15    16    21    21    27
```

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




```
                    Estimated regression weights

                                    Standard
   Parameter        Estimate          Error     z Value     P Value
   ---------        --------       --------     -------     -------
   intcept           -1.2535         0.1695     -7.3931      0.0000
   PreTHKS            0.4010         0.0461      8.6970      0.0000
   CC                 0.9883         0.1973      5.0088      0.0000
   TV                 0.2870         0.1920      1.4947      0.1350
   CC*TV             -0.3690         0.2774     -1.3303      0.1834


               Estimated level 2 variances and covariances

                                             Standard
   Parameter                 Estimate          Error     z Value     P Value
   ---------                 --------       --------     -------     -------
   intcept/intcept             0.2193         0.0802      2.7328      0.0063
```

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

$$\hat{\eta}_{ij} = -1.2535 + 0.9883 \times \mathrm{CC}_i + 0.2870 \times \mathrm{TV}_i - 0.369 \times \left(\mathrm{CC} \times \mathrm{TV}\right)_i$$
$$+0.401 \times \mathrm{PreTHKS}_{ij}$$

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

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

A comparison of unit-specific and population-average predicted probabilities for the current model are given in Table XXX.8. For comparison purposes, similar results for the previous model can be found in Table XXX.7.

Table XXX.8: 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%

2.2.4 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.

2.2.4.1 The model

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

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

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

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

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

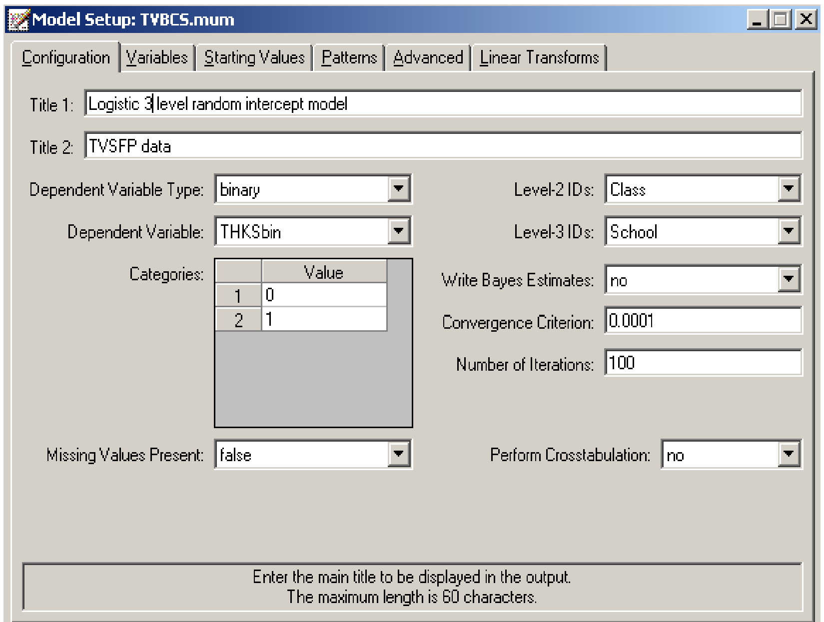
With classrooms nested within schools, however, a third level of the hierarchy is defined. At this level, the level-2 coefficients become outcomes again, and can potentially vary over the schools (level-3 units). In the current model, we allow only the intercept to vary randomly over the schools.

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

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

2.2.4.2 Setting up the analysis

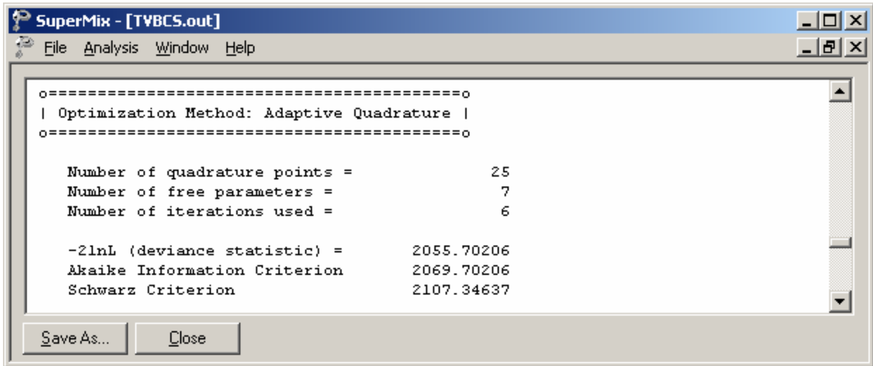
We modify our model setup saved to the syntax file **TVBS.mum** by first using the **Open Existing Model Setup** option on the **File** menu to retrieve the syntax file. Then click on **File, Save** as to save the model setup in a new file, such as **TVBSC.mum**. Next, select CLASS as the **Level-2 ID**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



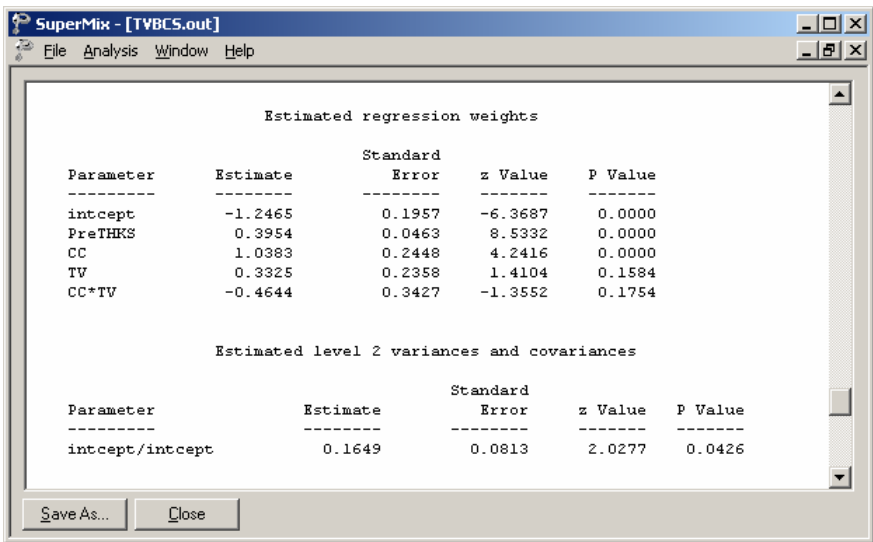
Keep all the other settings unchanged. Save the changes to the file **TVBCS.mum** and select the **Run** option on the **Analysis** menu to run the analysis.

2.2.4.3 Discussion of results

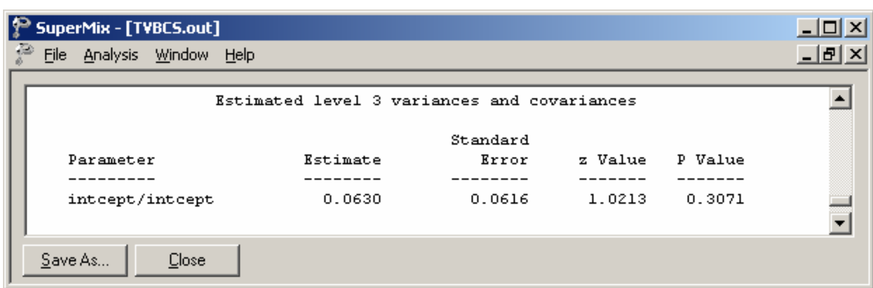
The portions of the output file **TBVSC.out** containing the estimates of the fixed and random coefficients in the current model are shown below.



```
SuperMix - [TVBCS.out]
File  Analysis  Window  Help

o==========================================o
| Optimization Method: Adaptive Quadrature |
o==========================================o

   Number of quadrature points =            25
   Number of free parameters =               7
   Number of iterations used =               6

   -2lnL (deviance statistic) =     2055.70206
   Akaike Information Criterion     2069.70206
   Schwarz Criterion                2107.34637

Save As...   Close
```



```
SuperMix - [TVBCS.out]
File  Analysis  Window  Help

                        Estimated regression weights

                                   Standard
   Parameter      Estimate           Error      z Value     P Value
   ---------      --------        --------      -------     -------
   intcept         -1.2465          0.1957      -6.3687      0.0000
   PreTHKS          0.3954          0.0463       8.5332      0.0000
   CC               1.0383          0.2448       4.2416      0.0000
   TV               0.3325          0.2358       1.4104      0.1584
   CC*TV           -0.4644          0.3427      -1.3552      0.1754


                   Estimated level 2 variances and covariances

                                            Standard
   Parameter                Estimate           Error     z Value     P Value
   ---------                --------        --------     -------     -------
   intcept/intcept            0.1649          0.0813      2.0277      0.0426

Save As...   Close
```



```
SuperMix - [TVBCS.out]
File  Analysis  Window  Help
                   Estimated level 3 variances and covariances

                                            Standard
   Parameter                Estimate           Error     z Value     P Value
   ---------                --------        --------     -------     -------
   intcept/intcept            0.0630          0.0616      1.0213      0.3071

Save As...   Close
```

Results for this model are compared to those obtained using the two 2-level models in Table XXX.9. 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.

Table XXX.9: Comparison of results for three models with binary variable THKSbin as outcome

Coefficient		**2-level: CLASS as ID**	**2-level: SCHOOL as ID**	**3-level**
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.

2.2.4.4 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

$$\text{Level-1: } \widehat{\text{error var}} = \pi^2/3 = 3.2899$$
$$\text{Level-2: } \widehat{\text{class var}} = 0.1649$$
$$\text{Level-3: } \widehat{\text{school var}} = 0.0630.$$

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

Similarity of students within the same school:

$$ICC = \frac{\sigma^2_{v(3)}}{\sigma^2_{v(3)} + \sigma^2_{v(2)} + \sigma^2} = \frac{0.063}{0.063 + 0.1649 + 3.28986}$$

$$= 0.0179.$$

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

$$ICC = \frac{\sigma^2_{v(2)}}{\sigma^2_{v(3)} + \sigma^2_{v(2)} + \sigma^2} = \frac{0.1649}{0.063 + 0.1649 + 3.28986}$$

$$= 0.04688.$$

Similarity of classes within the same school:

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

$$= 0.7236.$$

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

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

The estimated unit-specific probabilities are calculated using

$$\hat{\eta}_{ijk} = -1.2465 + 1.0383 \times \mathrm{CC}_i + 0.3325 \times \mathrm{TV}_i - 0.4.644 \times \mathrm{CC}_i \times \mathrm{TV}_i$$
$$+0.3954 \times \mathrm{PreTHKS}_{ijk}$$

and

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

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.

2.3 Models based on the subset of NESARC data

2.3.1 The data

The data set is from the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC), which was designed to be a longitudinal survey with its first wave fielded in 2001-2002. This data file has been used in some of the examples in Chapter XXX. Detailed information about the survey is available at http://niaaa.census.gov/index.html.

In Section 4 of the description of the NESARC study, information on the data regarding occurrences of major depression, family history of major depression and dysthymia are collected. This information was used, in combination with the demographic information provided in Section 1 of the study description, to produce the **nesarc_berc.ss3** data set used in this section. The image below shows the first ten records of this data set. There are 2339 dysthymia respondents in the survey; after listwise deletion, the sample size is 1981.

nesarc_berc.ss3

Apply

	(A) PSU	(B) FINWT	(C) AGE	(D) SEX	(E) FULLTI	(F) YR2 D	(G) WHITE
1	1001.00	4270.49	24.00	0.00	1.00	0.00	1.00
2	1001.00	1899.53	33.00	0.00	0.00	0.00	1.00
3	1001.00	2370.19	60.00	0.00	1.00	0.00	0.00
4	1001.00	3897.07	29.00	1.00	1.00	0.00	1.00
5	1001.00	6610.44	80.00	1.00	0.00	0.00	1.00
6	1001.00	3789.37	36.00	1.00	0.00	0.00	0.00
7	1001.00	3167.29	66.00	1.00	1.00	0.00	1.00
8	1001.00	959.70	65.00	1.00	0.00	0.00	0.00
9	1001.00	3167.29	71.00	1.00	0.00	0.00	1.00
10	1001.00	7231.97	54.00	1.00	0.00	0.00	1.00

The variables of interest are:

- PSU denotes the Census 2000/2001 Supplementary Survey (C2SS) primary sampling unit.
- FINWT represents the NESARC weights sample results used to form national level estimates. The final weight is the product of the NESARC base weight and other individual weighting factors.
- AGE represents the age of the respondent.
- SEX is the gender of the respondent (1 for male, 0 for female).
- FULLTIME is recoded from question S1Q7A1. It is the response to the statement "present situation includes working full time (35+ hours a week)" with 1 indicating yes and 0 indicating no.
- YR2_DEP is the observed response to the statement that the respondent had a period of at least 2 years with low mood, and being sad or depressed most of day (1 = yes, 0 = no.) It is recoded from S4CQ1 in the source data.
- WHITEOTH represents the origin of white and other ethnicities, excluding Black and Hispanic. It is recoded from items S1Q1C, S1Q1D2, S1Q1D3 and S1Q1D5 in the NESARC source code (1 for white and other, 0 for black and Hispanic).
- BLACK represents African American respondents in the sample. It is recoded from S1Q1C and S1Q1D3 (1 for African American, 0 for others).
- HISPANIC is an indicator for Hispanic respondents in the sample data. It is recoded from S1Q1C, S1Q1D3 and S1Q1D5 (1 for Hispanic, 0 for others).
- YOUNG is recoded from AGE. Respondents younger than 35 have the value 1; otherwise, YOUNG = 0.
- MIDDLE is recoded from AGE. Respondents with $35 \leq AGE < 50$ have the value 1. Otherwise, MIDDLE = 0.
- OLD is recoded from AGE. Respondents with $AGE \geq 50$ have the value 1. Otherwise, OLD = 0.

We recoded the ethnicity variables because of the unbalanced numbers of respondents from different ethnicities in the original NESARC data. While weights are supplied with the data and should be used to adjust for the disproportionality of the sample, the use of indicator variables offers the opportunity to obtain estimated coefficients for individual groups while using one of the other ethnic groups as a reference group. The recoding of ethnicity is discussed in detail in Section XXXXX . Similarly, an explanation of the motivation for the replacement of AGE with 3 indicator variables representing different age groups is given in Section XXXXX.

In this section, we discuss the fitting of three Bernoulli models to these data.

2.3.2 A 2-level random intercept probit model

2.3.2.1 The model

In the previous models (see Section 2.2) the logistic link function was used. We now fit a model by using the probit link function.

The outcome variable of interest is YR2_DEP has the values 0 or 1. For this binary outcome variable

$$\text{Prob}(\text{YR2\_DEP}_{ij} = 1 \mid \boldsymbol{\beta}_i) = \Phi^{-1}\left(\eta_{ij}\right)$$

where η_{ij} represents the log of the odds of success, and can be expressed as

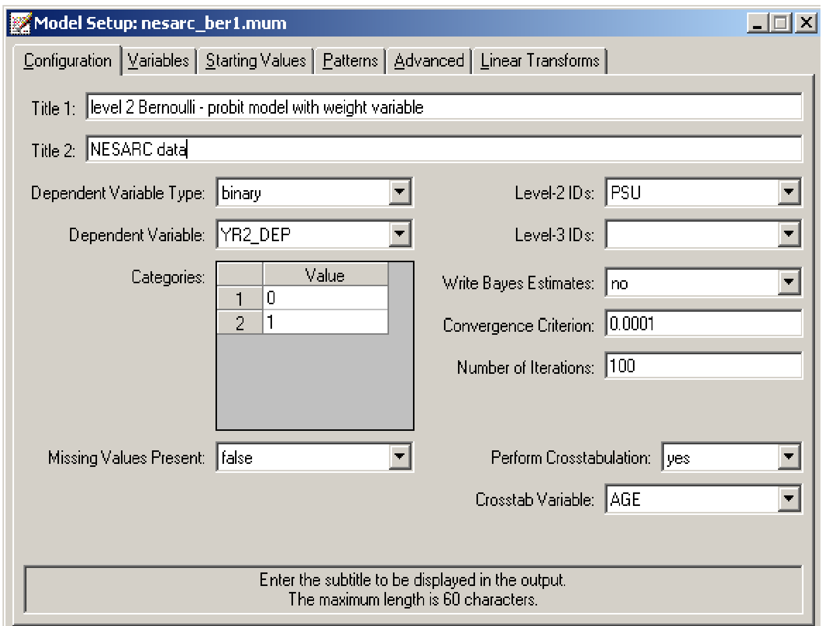
$$\eta_{ij} = b_0 + b_1 \times \text{AGE}_{ij} + b_2 \times \text{SEX}_{ij} + b_3 \times \text{FULLTIME}_{ij} + b_{i0} + v_{i0} + e_{ij}$$

for the intended model. This transformation, commonly referred to as the probit link function, constrains $\text{Prob}(y_{ij} = 1 \mid \boldsymbol{\beta})$ to lie in the interval (0,1).

2.3.2.2 Setting up the analysis

Open the SuperMix spreadsheet **nesarc_berc.ss3**. From the main menu bar, select the **File, New Model Setup** option.

- The **Configuration** screen is the first tab on the **Model Setup** dialog box. It is used to define the outcome variable and level-2 and level-3 IDs. Some other settings such as missing values, convergence criterion, number of iterations, etc. can also be specified here. For all the available settings, please refer to Chapter XXXX.

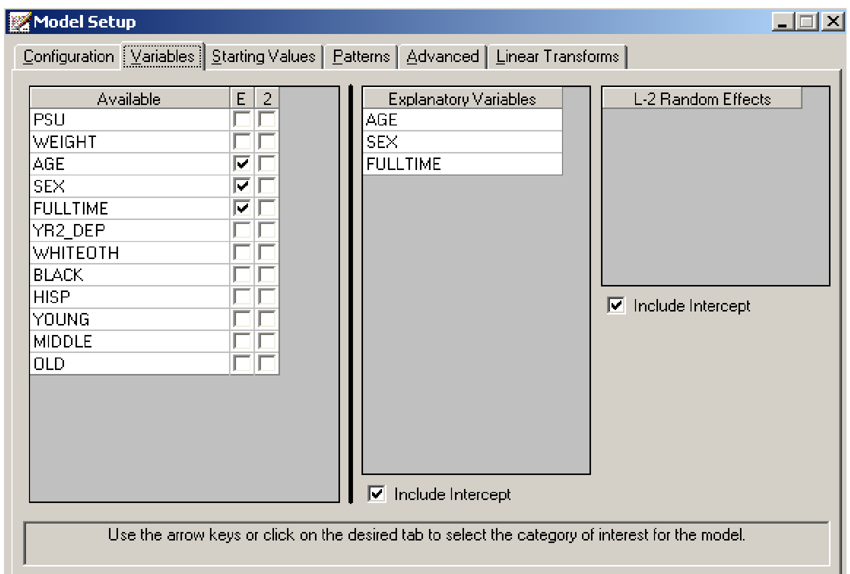


To obtain the model shown above, proceed as follows.

- Select the **binary** option from the **Dependent Variable Type** drop-down list.
- Select the outcome variable YR2_DEP from the **Dependent Variable type** drop-down list box.

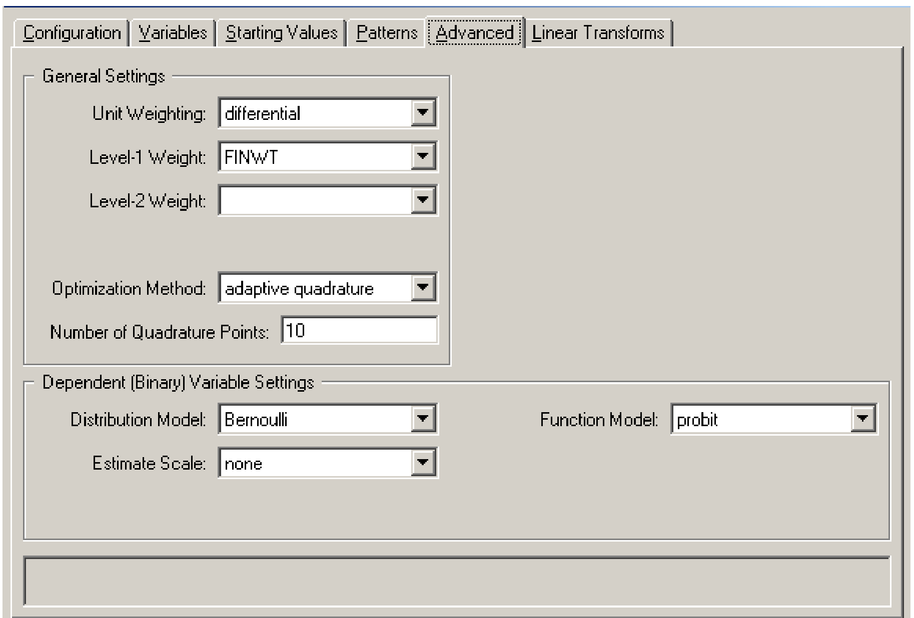
- Select PSU from the **Level-2 ID** drop-down list box.
- Enter a title for the analysis in the **Title** text boxes if needed (optional).
- Request a crosstabulation of the outcome variable against AGE by selecting Yes from the **Perform Crosstabulation** drop-down list box, and select AGE as **Crosstab Variable**.
- Keep all the other settings on the **Configuration** screen at their default values. Proceed to the **Variables** screen by clicking on this tab.

The **Variables** screen is used to specify the fixed and random effects to be included in the model. Select the explanatory (fixed) variables using the **E** check boxes next to the variables AGE, SEX and FULLTIME in the **Available** grid at the left of the screen. After selecting all the explanatory variables, the screen shown below is obtained. The **Include Intercept** check box in the **Explanatory Variables** grid is checked by default, indicating that an intercept term will automatically be included in the fixed part of the model.



The **Advanced** tab enables the user to define the weight variable. Weights are often used in complex sampling to adjust the existing sample for known biases. In SuperMix, the weight is normalized by default. To include a weight variable, proceed as follows.

- Select differential from the **Unit Weight** drop-down list to activate the **Assigned Weight.**
- Select FINWT from the drop-down list of the **Level-1 Weight field**.
- Select adaptive quadrature as method of estimation in the **Optimization Method** field.



Save the model specifications to the file **nesarc_ber1.mum** and run the analysis.

2.3.2.3 Discussion of results

Portions of the output file **nesarc_ber1.out** are shown below.

Model and data description

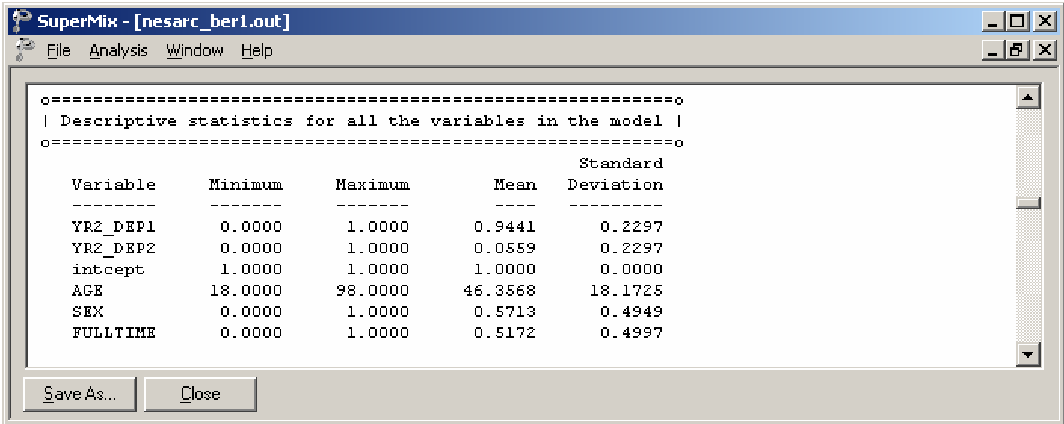
As shown in the model and data description section, the Bernoulli sampling distribution and probit link function are specified. The weight variable FINWT is used to include sampling weight. There are 419849 observations from 435 PSUs included in the data we are analyzing.

```
             Model and Data Descriptions

    Sampling Distribution                    = Bernoulli
    Link Function                            = Probit
    Level-1 Weight Variable                  = FINWT
    Number of Level-2 Units                  = 435
    Number of Level-1 Units                  = 41849
    Number of Level-1 Units per Level-2 Unit =
      23    29   109    26    28   171    69    90    19    56   167    18
      44    16   105    75   494   226   102    23    80    76    21    22
      37   111    32   112    64    78   362    85    28   258    54    52
      76    32    96    47   101  1481   347   472   397   238    88    37
     144   114   113    47   369    96    86    11    33    26   177   287
      90    22    23   195    62    45    82   189    46    40   342    66
      64    53    43   162    67    49   374   593    80    85   256    80
     100    49    93   130   144    61    44   590   125    15    33    30
```

Save As... Close

Descriptive statistics

The data summary is followed by descriptive statistics for all the variables included in the model. As shown below, about 94.41% of the respondents did not have a 2+ year period of low moods or being sad or depressed most of day.



SuperMix - [nesarc_ber1.out]

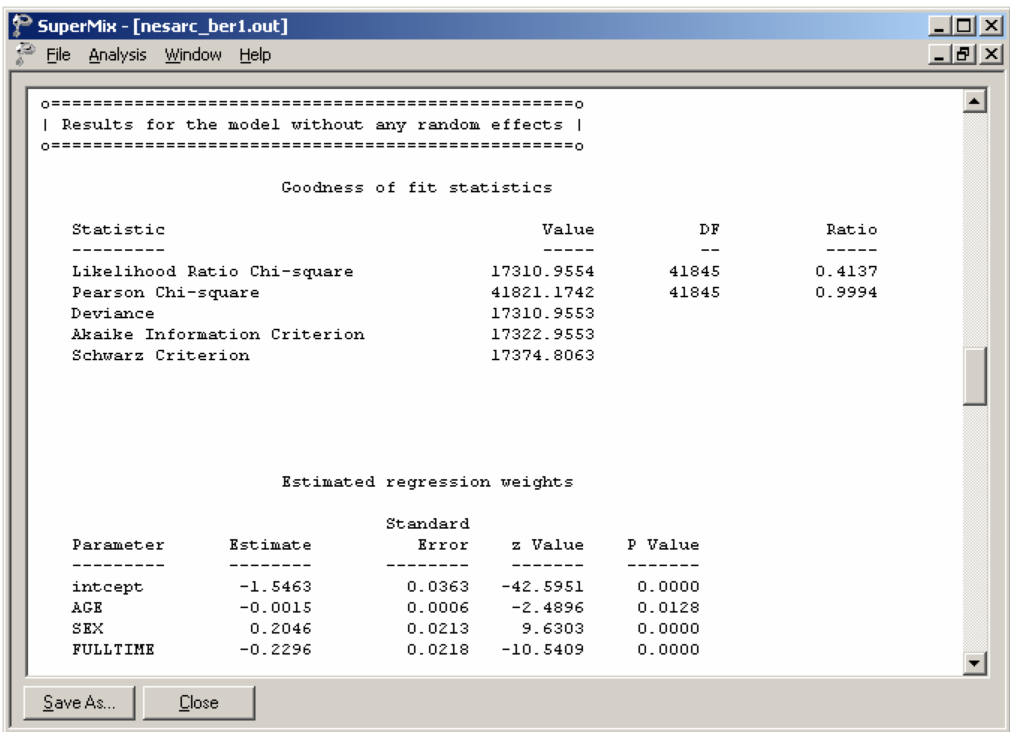
File Analysis Window Help

| Descriptive statistics for all the variables in the model |

Variable	Minimum	Maximum	Mean	Standard Deviation
YR2_DEP1	0.0000	1.0000	0.9441	0.2297
YR2_DEP2	0.0000	1.0000	0.0559	0.2297
intcept	1.0000	1.0000	1.0000	0.0000
AGE	18.0000	98.0000	46.3568	18.1725
SEX	0.0000	1.0000	0.5713	0.4949
FULLTIME	0.0000	1.0000	0.5172	0.4997

Save As... Close

Results for the model without any random effects



SuperMix - [nesarc_ber1.out]

File Analysis Window Help

| Results for the model without any random effects |

Goodness of fit statistics

Statistic	Value	DF	Ratio
Likelihood Ratio Chi-square	17310.9554	41845	0.4137
Pearson Chi-square	41821.1742	41845	0.9994
Deviance	17310.9553		
Akaike Information Criterion	17322.9553		
Schwarz Criterion	17374.8063		

Estimated regression weights

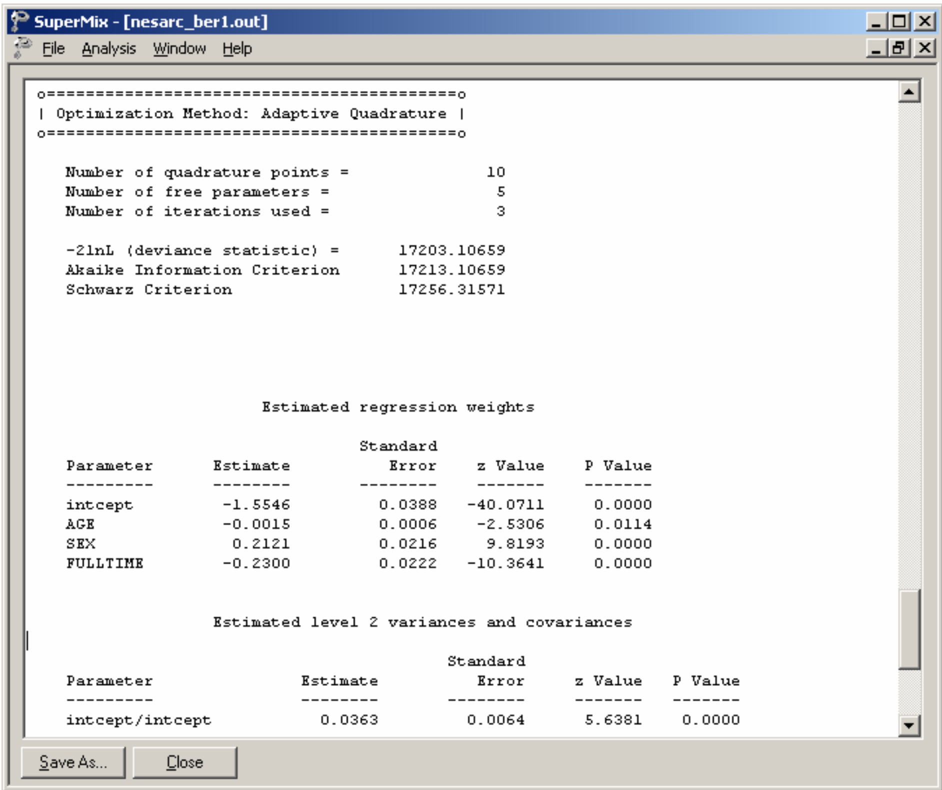
Parameter	Estimate	Standard Error	z Value	P Value
intcept	-1.5463	0.0363	-42.5951	0.0000
AGE	-0.0015	0.0006	-2.4896	0.0128
SEX	0.2046	0.0213	9.6303	0.0000
FULLTIME	-0.2296	0.0218	-10.5409	0.0000

Save As... Close

Descriptive statistics are followed by the results for the model without any random effects. These results are used as the starting values for the model with random effects.

Results for the model with random effects

The total number of iterations, the goodness of fit statistics and the estimated regression weights are shown below.




```
o========================================o
| Optimization Method: Adaptive Quadrature |
o========================================o

   Number of quadrature points =             10
   Number of free parameters =                5
   Number of iterations used =                3

   -2lnL (deviance statistic) =      17203.10659
   Akaike Information Criterion      17213.10659
   Schwarz Criterion                 17256.31571



                    Estimated regression weights

                              Standard
   Parameter      Estimate       Error     z Value     P Value
   ---------      --------    --------     -------     -------
   intcept         -1.5546      0.0388    -40.0711      0.0000
   AGE             -0.0015      0.0006     -2.5306      0.0114
   SEX              0.2121      0.0216      9.8193      0.0000
   FULLTIME        -0.2300      0.0222    -10.3641      0.0000


               Estimated level 2 variances and covariances

                                       Standard
   Parameter              Estimate        Error     z Value     P Value
   ---------              --------     --------     -------     -------
   intcept/intcept          0.0363       0.0064      5.6381      0.0000
```

The estimated intercept coefficient is -1.5546. The estimated coefficient associated with AGE is – 0.0015, which implies that for every year increase in age of a typical respondent, the estimated probit $\hat{\eta}_{ij}$ is expected to decrease by 0.0015. The coefficient seems small, but keep in mind that age has a wide range, and consequently this estimate may have a big effect on the overall probability. The estimated coefficient associated with gender is 0.2121, which indicates that the male respondents (SEX = 1) have a larger $\hat{\eta}_{ij}$. The estimate for the indicator of FULLTIME shows that respondents with full-time jobs were expected to have a lower $\hat{\eta}_{ij}$ value than respondents with a similar profile in terms of age and gender but without full-time employment.

2.3.2.4 Interpreting the adaptive quadrature results

The probit link function is now used to transform these estimates into probabilities. First, we substitute the regression weights and obtain an expression for $\hat{\eta}_{ij}$:

$$\begin{aligned}\hat{\eta}_{ij} &= \hat{b}_{0i} + \hat{b}_{1i} \times (\text{AGE})_{ij} + \hat{b}_{2i} \times (\text{SEX})_{ij} + \hat{b}_{3i} \times (\text{FULLTIME})_{ij} \\ &= -1.5546 - 0.0015 \times (\text{AGE})_{ij} + 0.2121 \times (\text{SEX})_{ij} - 0.23 \times (\text{FULLTIME})_{ij}.\end{aligned}$$

For a typical 30-year-old male with a full-time job, SEX = 1, FULLTIME = 1 and AGE = 30, and thus

$$\begin{aligned}\hat{\eta}_{ij} &= -1.5546 - 0.0015 \times 30 + 0.2121 - 0.23 \\ &= -1.6025.\end{aligned}.$$

Transform the $\hat{\eta}_{ij}$ into the corresponding probability by using the probit link function:

$$\text{Prob}(\widehat{\text{YR2\_DEP}}_{ij} = 1) = \Phi^{-1}(-1.60937) = 0.0545.$$

In terms of percentages, 5.45% of males with this profile would be expected to suffer from long-term depression episodes. Similarly, the probability of having a depression episode of 2+ years' duration for different gender and age combinations can be calculated. These probabilities, expressed as percentages, are reported in Table XXX.10 below.

Table XXX.10: % probabilities of having a depression episode

Age	20	30	40	50	60	70
not fulltime, female	5.65%	5.48%	5.32%	5.16%	5.00%	4.85%
not fulltime, male	8.50%	8.26%	8.04%	7.82%	7.60%	7.39%
fulltime, female	3.48%	3.37%	3.25%	3.15%	3.04%	2.94%
fulltime, male	5.45%	5.29%	5.13%	4.97%	4.82%	4.67%

In general, males without full-time employment were more likely to have depression episodes than their female counterparts. Surprisingly, this is also true of males with full-time employment.

These probabilities can also be depicted in a line plot as shown below. The line associated with males without full-time jobs is considerably higher than for any other groups, again illustrating that this group has the highest probability of having 2+ years' period with low mood regardless of their age. For all the correspondents, as they grow older, the probability of having lengthy depression episodes decreased.

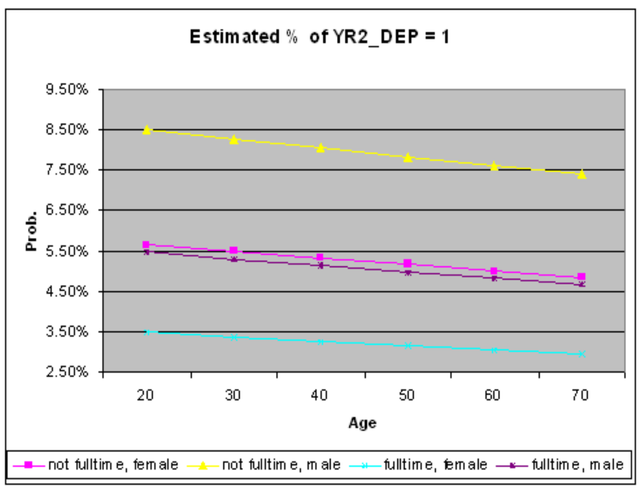


Figure XXX.5: Expected probabilities for subgroups

2.3.3 A 2-level random intercept model with additional predictors

2.3.3.1 The model

In the previous section, we modeled the outcome variable YR2_DEP in terms of its relationship with the predictors AGE, SEX and FULLTIME. The model discussed in this section takes the ethnicity of patients into consideration by including two dummy variables, BLACK and HISP. Since the group of WHITEOTH is not included, it is automatically regarded as the reference category.

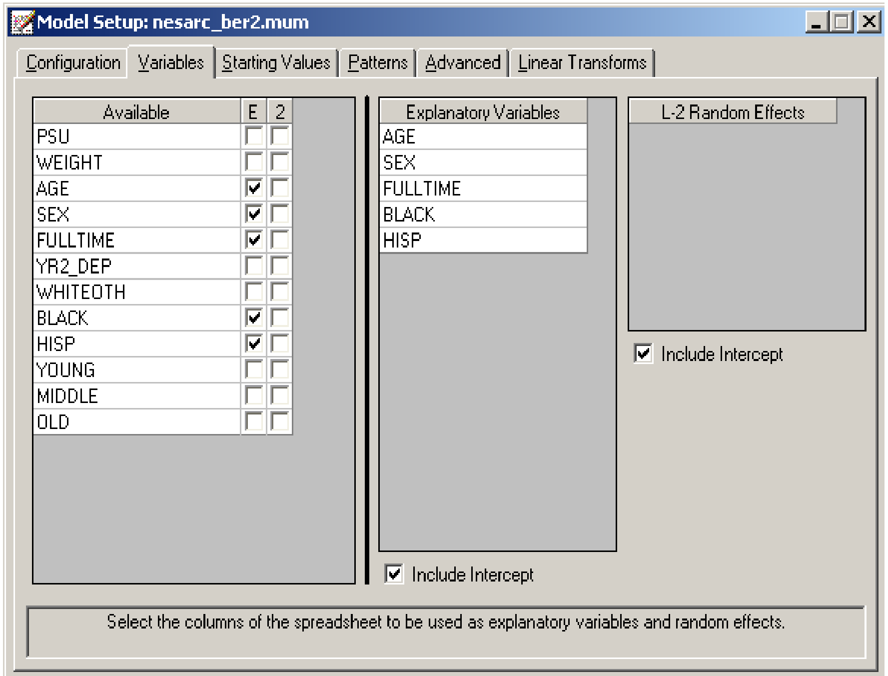
For the current model, the log of the odds of success (η_{ij}) can be expressed as

$$\eta_{ij} = b_0 + b_1 \times \text{AGE}_{ij} + b_2 \times \text{SEX}_{ij} + b_3 \times \text{FULLTIME}_{ij} + b_4 \times \text{BLACK}_{ij} + b_5 \times \text{HISP}_{ij}$$
$$+ v_{0i} + e_{ij}.$$

2.3.3.2 Setting up the analysis

We can modify the model setup file **nesarc_ber1.ss3** by opening it and then saving it under a different name, such as **nesarc_ber2.ss3**.

Click on the **Variables** tab of the **Model Setup** window. Add the predictors BLACK and HISP to the model by checking the boxes next to these variables in the **E** column, as shown below.



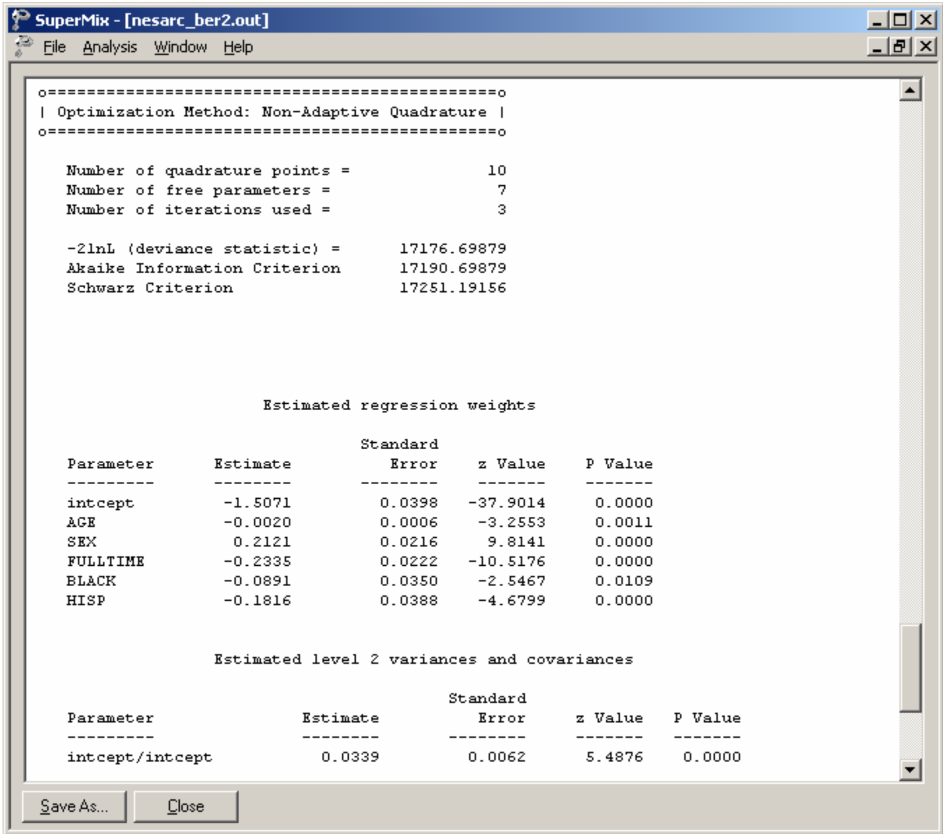
Save the modified model specification file, and select the **Run** option from the **Analysis** menu to perform the analysis.

2.3.3.3 Discussion of results

Portions of the output file **nesarc_berc2.out** are shown below.

Results for the model fitted with adaptive quadrature

The goodness of fit statistics are shown below. Since the previous model can be considered as a submodel of the current model, the deviances of these two models can be used to perform a χ^2 test to evaluate possible improvement in model fit.



SuperMix - [nesarc_ber2.out]

File Analysis Window Help

```
o============================================o
| Optimization Method: Non-Adaptive Quadrature |
o============================================o

   Number of quadrature points =             10
   Number of free parameters =                7
   Number of iterations used =                3

   -2lnL (deviance statistic) =      17176.69879
   Akaike Information Criterion      17190.69879
   Schwarz Criterion                 17251.19156
```

Estimated regression weights

Parameter	Estimate	Standard Error	z Value	P Value
intcept	-1.5071	0.0398	-37.9014	0.0000
AGE	-0.0020	0.0006	-3.2553	0.0011
SEX	0.2121	0.0216	9.8141	0.0000
FULLTIME	-0.2335	0.0222	-10.5176	0.0000
BLACK	-0.0891	0.0350	-2.5467	0.0109
HISP	-0.1816	0.0388	-4.6799	0.0000

Estimated level 2 variances and covariances

Parameter	Estimate	Standard Error	z Value	P Value
intcept/intcept	0.0339	0.0062	5.4876	0.0000

Save As... Close

The output describing the estimated fixed effects after convergence is shown next. As shown above the estimated logit for the intercept is -1.5071, the estimated logit associated with AGE is -0.002, etc. It is interesting to note that the only positive estimate is for gender. Males are thus more likely to show long-term depression, while it will be less likely in those who are older or fully employed. The ethnicity indicators' coefficients also indicate that white respondents are most likely to have depression, with the Hispanic population the least likely.

2.3.3.4 Interpreting the adaptive quadrature results

Estimated outcomes for different groups: unit-specific results

To evaluate the simultaneous impact of these estimates on the expected probabilities for respondents from the subgroups formed by the categories of age, gender, and ethnicity, we may use the estimated regression weights and the link function to calculate probabilities of having depression in the same way as for the previous model.

For the current model, $\hat{\eta}_{ij}$ can be expressed as:

$$\begin{aligned}\hat{\eta}_{ij} = &-1.5071 - 0.0020\times \mathrm{AGE}_{ij} + 0.2121\times \mathrm{SEX}_{ij} - 0.2335\times \mathrm{FULLTIME}_{ij} \\ &-0.0891\times \mathrm{BLACK}_{ij} - 0.1816\times \mathrm{HISP}_{ij}.\end{aligned}$$

Table XXX.11 contains a subset of these estimated probabilities. Only typical respondents 30 or 50 years old are considered here, and probabilities are expressed as percentages.

Table XXX.11: % Probabilities of having depression episodes for selected age groups

Age	30			50		
Ethnicity	**White**	**Black**	**Hispanic**	**White**	**Black**	**Hispanic**
not fulltime, female	5.85%	4.88%	4.02%	5.40%	4.49%	3.68%
not fulltime, male	8.77%	7.44%	6.22%	8.15%	6.89%	5.74%
fulltime, female	3.59%	2.94%	2.37%	3.28%	2.68%	2.16%
fulltime, male	5.61%	4.67%	3.84%	5.17%	4.29%	3.51%

Younger white males without full-time employment have the highest risk of having long-term depression, while female Hispanic respondents with full-time employment were least at risk.

The results in Table XXX.11 can also be depicted as a bar chart. Figure xxx.6 shows that white respondents are more likely to get depressed for a long period than African American or Hispanic respondents.

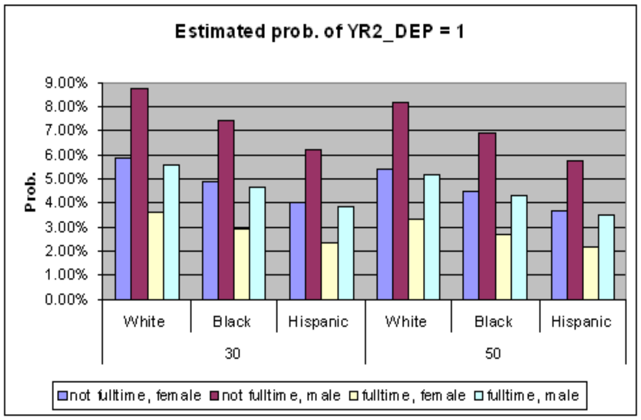


Figure XXX.6: Estimated probabilities for subgroups

Model comparison

Since the two models in this section are nested models, the chi-square difference test can be used. The deviances, AIC, and SBC statistics for these models are summarized in Table XXX.12. These statistics suggest that the second model fits the data better.

Table XXX12: Model comparison

Statistic	Model 1	Model 2	difference	Difference in d.f.
$-2\ln L$ (deviance statistic)	17203.107	17176.699	26.4078	2
Akaike Information Criterion	17213.107	17190.699	22.4078	2
Schwarz Criterion	17256.316	17251.192	5.12415	2

References

Krommer, A.R. & Ueberhuber, C.W. (1994). *Numerical Integration on advanced computer systems.* New York: Springer-Verlag.

Stroud, A.H. & Secrest, D. (1966). *Gaussian quadrature formulas.* New Jersey: Prentice-Hall.

McCullagh & Nelder, 1989

Flay, *et. al.*, 1988

Hedeker & Gibbons, 2006

National Epidemiologic Survey on Alcohol and Related Conditions (NESARC)