Two-level probit model

Contents

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

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

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

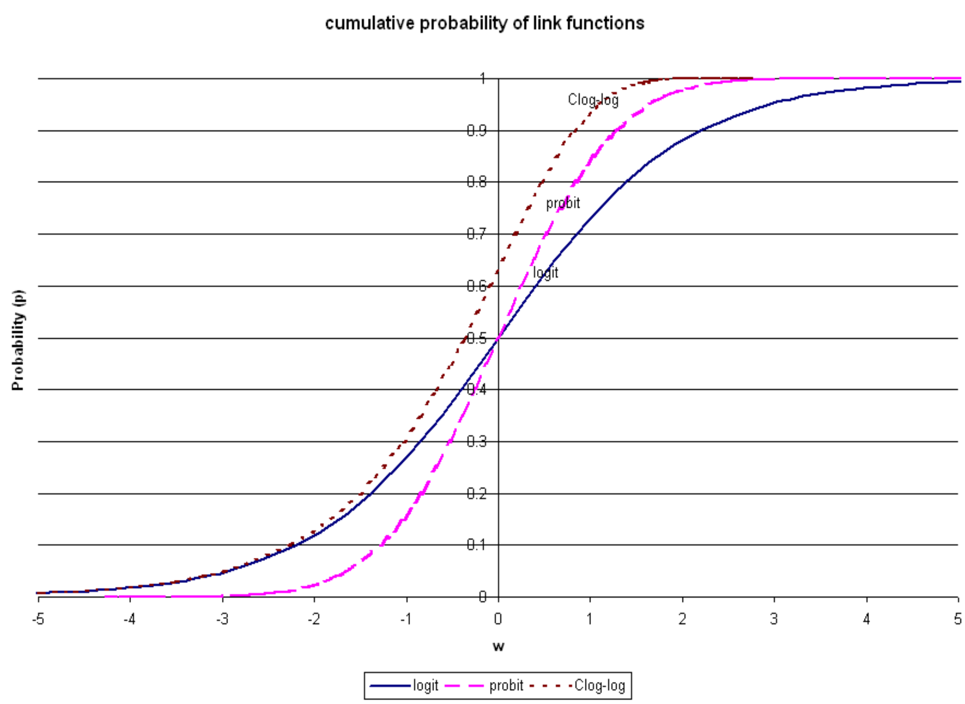


Figure 4.1: Cumulative density of link functions

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 $v_{i1}, v_{i2}, ..., v_{ir}$ 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 $\hat{v}_i$ 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 $\mathrm{cov}\left(\hat{v}_i\right), i = 1, 2, ..., N$ are obtained and these, in turn, are used to estimate $\Phi_{(2)} = \mathrm{cov}(\mathbf{v}_i)$.

Numerical quadrature

Since

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

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

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

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

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

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 the subset of NESARC data

3.2.1 The data

The data set is from the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC). This data file has been used in some of the examples in Section 3.1. Detailed information about the survey is available at the NIAAA website at http://niaaa.census.gov/index.html. We focus on information regarding occurrences of major depression, family history of major depression and dysthymia. 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 $35 \leq \text{AGE} < 50$ have the value 1. Otherwise, MIDDLE = 0.
- OLD is recoded from AGE. Respondents with $\text{AGE} \geq 50$ 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.

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

3.2.2 A 2-level random intercept probit model

3.2.2.1 The model

In previous models 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

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

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

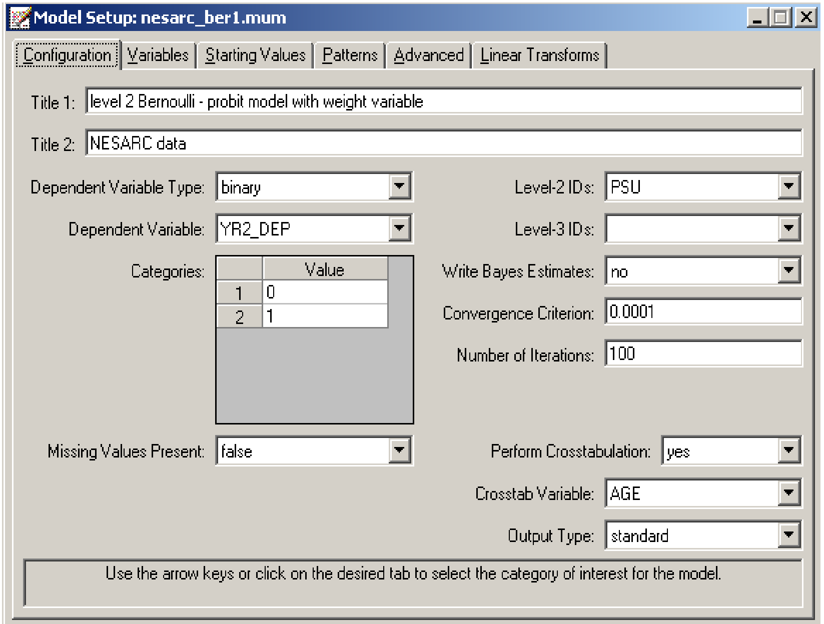
$$\eta_{ij} = b_0 + b_1 \times \text{AGE}_{ij} + b_2 \times \text{SEX}_{ij} + b_3 \times \text{FULLTIME}_{ij} + b_{i0} + v_{i0} + e_{ij}$$

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).

3.2.2.2 Setting up the analysis

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

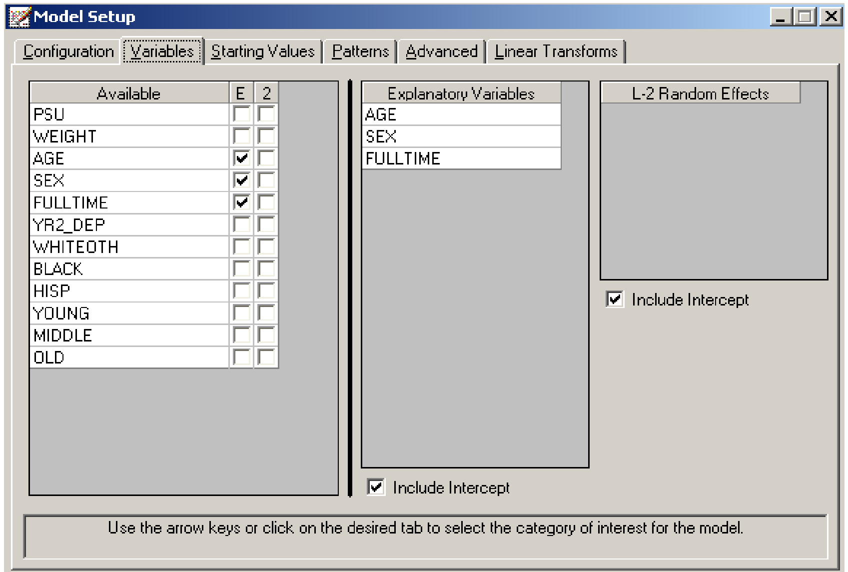
The **Configuration** screen is the first tab on the **Model Setup** dialog box. It is used to define the outcome variable and level-2 and level-3 IDs. Some other settings such as missing values, convergence criterion, number of iterations, etc. can also be specified here.



To obtain the model shown above, proceed as follows.

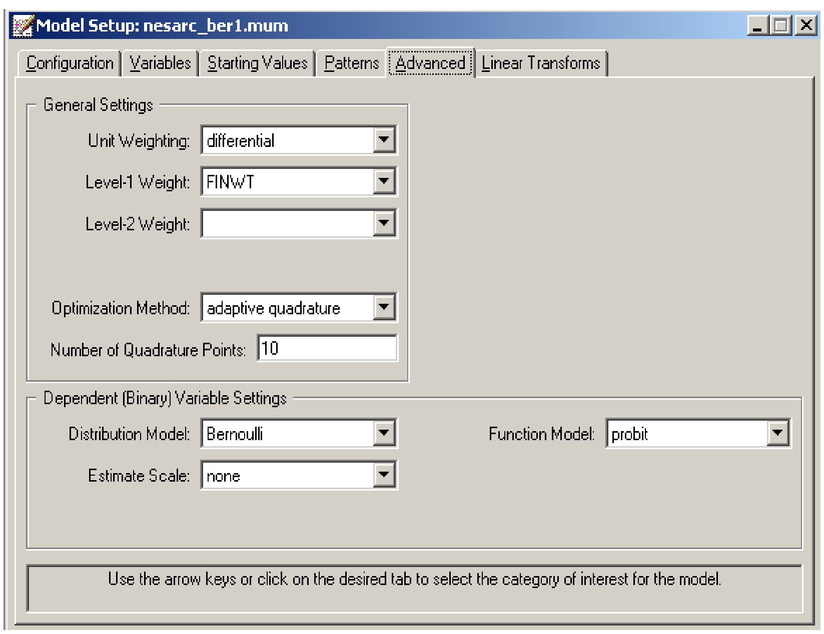
- Select the **binary** option from the **Dependent Variable Type** drop-down list.
- Select the outcome variable YR2_DEP from the **Dependent Variable type** drop-down list box.
- Select PSU from the **Level-2 ID** drop-down list box.
- Enter a title for the analysis in the **Title** text boxes if needed (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**.



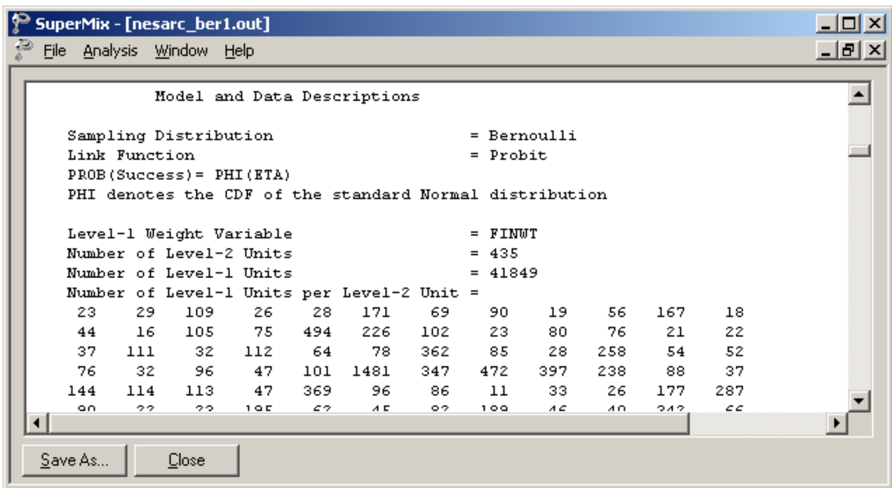
Save the model specifications to the file **nesarc_ber1.mum** and run the analysis.

3.2.2.3 Discussion of results

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

Model and data description

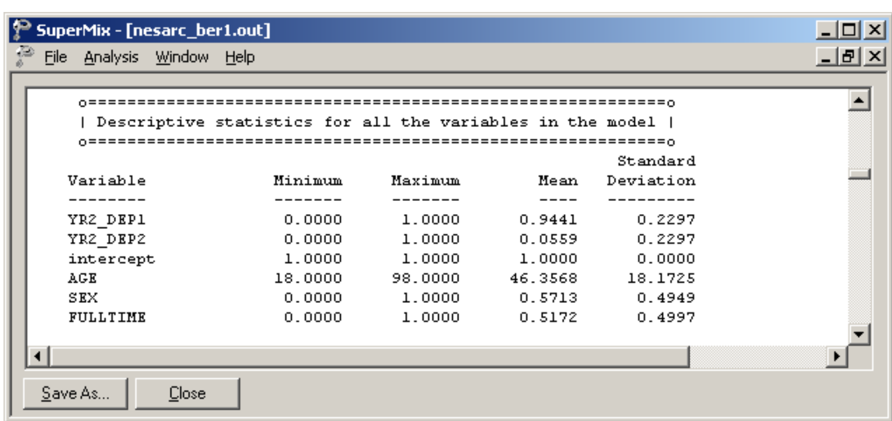
As shown in the model and data description section, the Bernoulli sampling distribution and probit link function are specified. The weight variable FINWT is used to include sampling weight. There are 41,849 observations from 435 PSUs included in the data we are analyzing.




```
           Model and Data Descriptions

    Sampling Distribution                     = Bernoulli
    Link Function                             = Probit
    PROB(Success)= PHI(ETA)
    PHI denotes the CDF of the standard Normal distribution

    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
```

Descriptive statistics

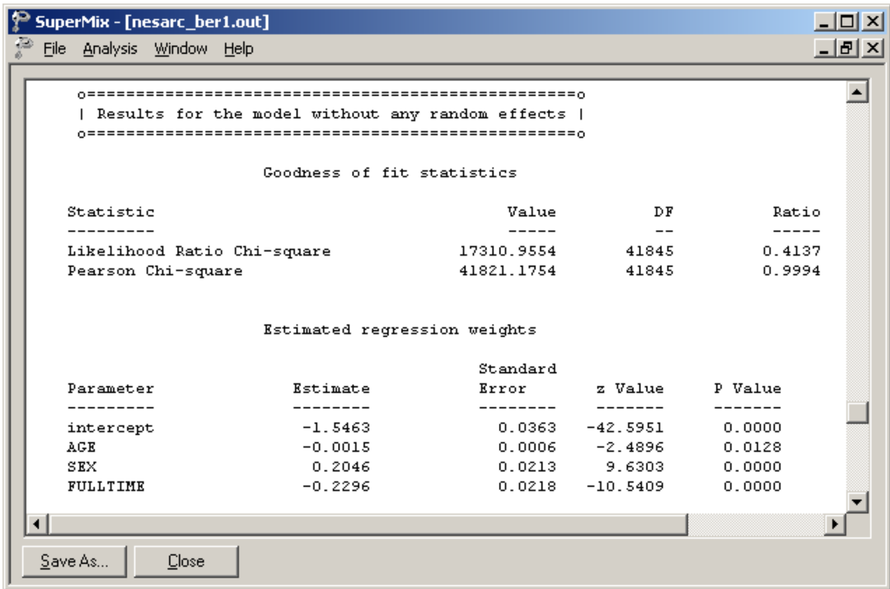
The data summary is followed by descriptive statistics for all the variables included in the model. As shown below, about 94.41% of the respondents did not have a 2+ year period of low moods or being sad or depressed most of day.




```
     o==========================================================o
     | Descriptive statistics for all the variables in the model |
     o==========================================================o
                                                          Standard
    Variable           Minimum     Maximum        Mean   Deviation
    --------           -------     -------        ----   ---------
    YR2_DEP1            0.0000      1.0000      0.9441      0.2297
    YR2_DEP2            0.0000      1.0000      0.0559      0.2297
    intercept           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
```

Results for the model without any random effects

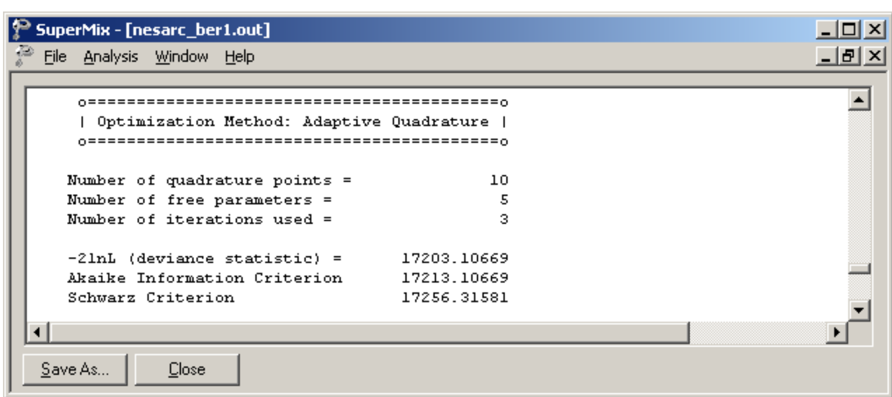
Descriptive statistics are followed by the results for the model without any random effects. These results are used as the starting values for the model with random effects.



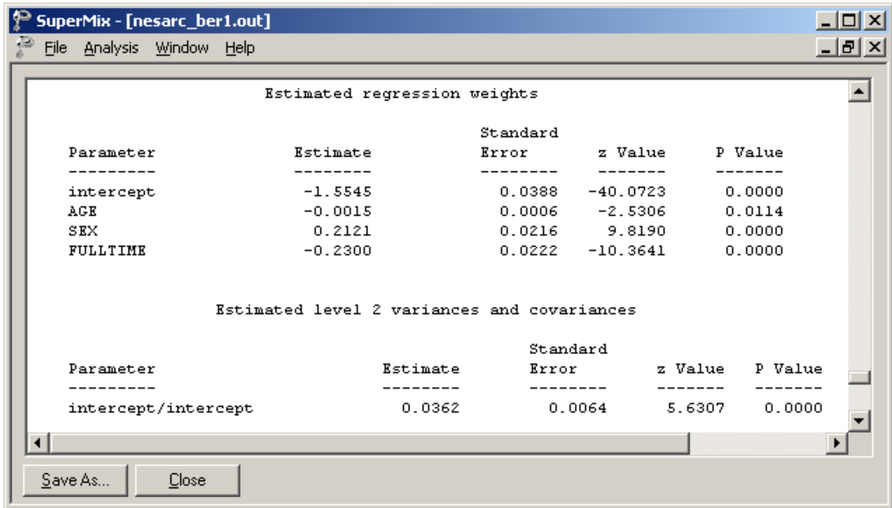
```
SuperMix - [nesarc_ber1.out]
File  Analysis  Window  Help

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

                 Goodness of fit statistics

Statistic                            Value        DF        Ratio
---------                            -----        --        -----
Likelihood Ratio Chi-square     17310.9554     41845       0.4137
Pearson Chi-square              41821.1754     41845       0.9994


                 Estimated regression weights

                                      Standard
Parameter          Estimate           Error       z Value    P Value
---------          --------           --------    -------    -------
intercept           -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
```

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.



```
SuperMix - [nesarc_ber1.out]
File  Analysis  Window  Help

 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) =      17203.10669
Akaike Information Criterion      17213.10669
Schwarz Criterion                 17256.31581

Save As...   Close
```


SuperMix - [nesarc_ber1.out]

File Analysis Window Help

Estimated regression weights

Parameter	Estimate	Standard Error	z Value	P Value
intercept	-1.5545	0.0388	-40.0723	0.0000
AGE	-0.0015	0.0006	-2.5306	0.0114
SEX	0.2121	0.0216	9.8190	0.0000
FULLTIME	-0.2300	0.0222	-10.3641	0.0000

Estimated level 2 variances and covariances

Parameter	Estimate	Standard Error	z Value	P Value
intercept/intercept	0.0362	0.0064	5.6307	0.0000

Save As... Close

The estimated intercept coefficient is – 1.5544. 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.

3.2.2.4 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:

$$\text{Prob}\left(\text{YR2\_}\hat{\text{DEP}}_{ij}=1\right)=\Phi^{-1}\left(-1.60937\right)=0.0545.$$

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 4.9 below.

Table 4.9: % probabilities of having a depression episode

Age	20	30	40	50	60	70
not fulltime, female	5.66%	5.49%	5.32%	5.16%	5.00%	4.85%
not fulltime, male	8.50%	8.27%	8.04%	7.82%	7.60%	7.39%
fulltime, female	3.48%	3.37%	3.26%	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.

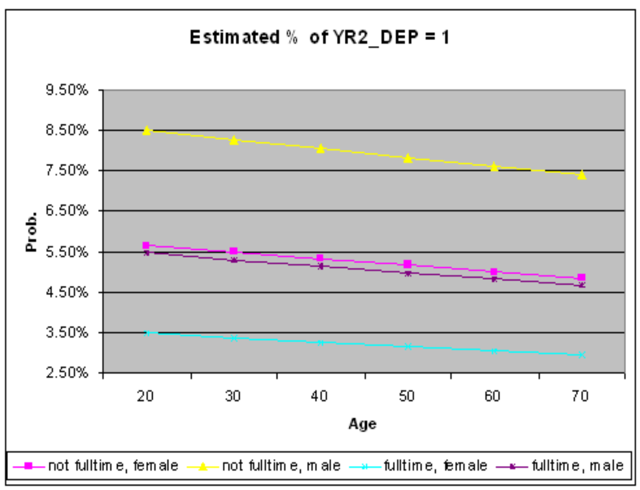


Figure 4.4: Expected probabilities for subgroups

These probabilities can also be depicted in Figure 4.4. 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.

3.2.3 A 2-level random intercept model with additional predictors

3.2.3.1 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.

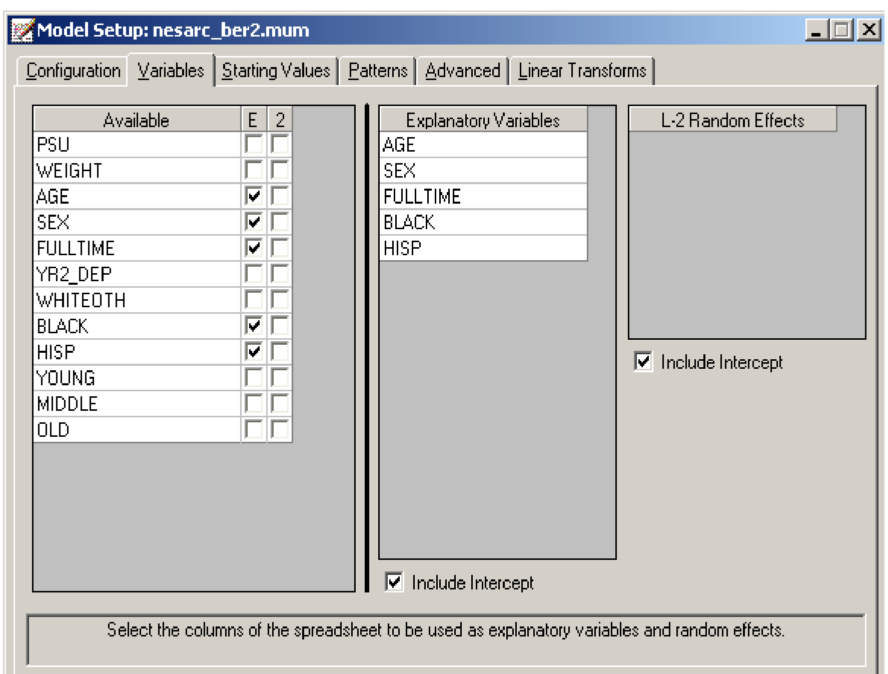
For the current model, the log of the odds of success (η_{ij}) can be expressed as

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

3.2.3.2 Setting up the analysis

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

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.



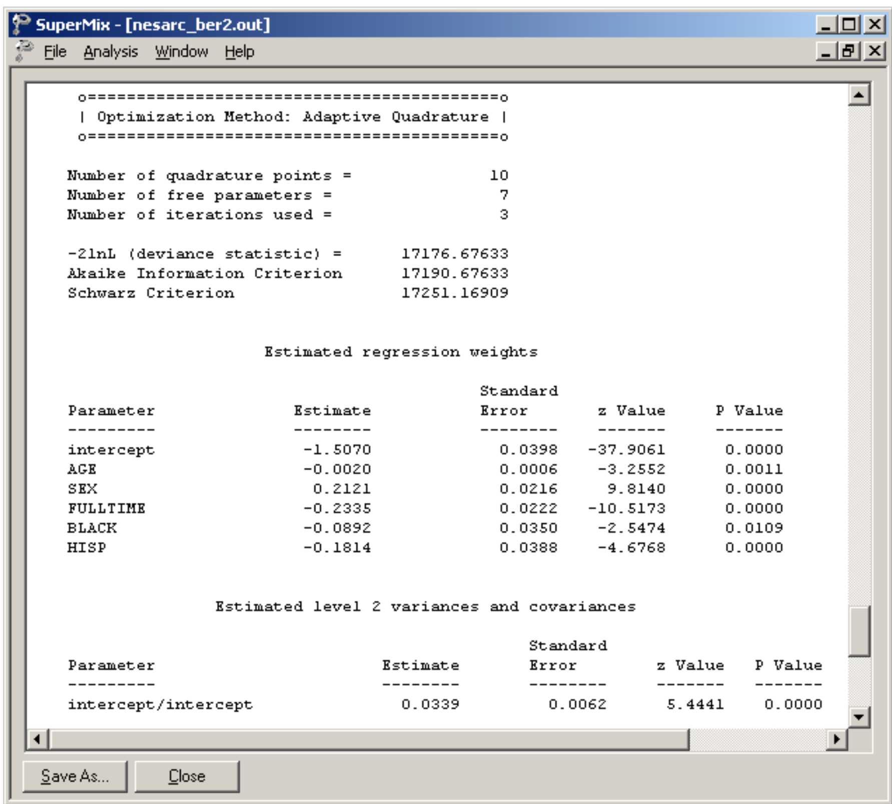
Save the modified model specification file, and select the **Run** option from the **Analysis** menu to perform the analysis.

3.2.3.3 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: Adaptive Quadrature |
o=========================================o

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

-2lnL (deviance statistic) =      17176.67633
Akaike Information Criterion      17190.67633
Schwarz Criterion                 17251.16909


                    Estimated regression weights

                                        Standard
Parameter             Estimate          Error      z Value      P Value
---------             --------          --------   -------      -------
intercept              -1.5070          0.0398    -37.9061      0.0000
AGE                    -0.0020          0.0006     -3.2552      0.0011
SEX                     0.2121          0.0216      9.8140      0.0000
FULLTIME               -0.2335          0.0222    -10.5173      0.0000
BLACK                  -0.0892          0.0350     -2.5474      0.0109
HISP                   -0.1814          0.0388     -4.6768      0.0000


               Estimated level 2 variances and covariances

                                             Standard
Parameter                     Estimate       Error        z Value   P Value
---------                     --------       --------     -------   -------
intercept/intercept             0.0339       0.0062       5.4441    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 –1.5069, 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.

3.2.3.4 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:

$$\begin{aligned}\hat{\eta}_{ij} &= -1.5069 - 0.0020 \times \text{AGE}_{ij} + 0.2121 \times \text{SEX}_{ij} - 0.2335 \times \text{FULLTIME}_{ij} \\ &\quad -0.0891 \times \text{BLACK}_{ij} - 0.1814 \times \text{HISP}_{ij}.\end{aligned}$$

Table 4.10 contains a subset of these estimated probabilities. Only typical respondents 30 or 50 years old are considered here, and probabilities are expressed as percentages.

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.

Table 4.10: % probabilities of having depression episodes for selected age groups

Age	30			50		
Ethnicity	White	Black	Hispanic	White	Black	Hispanic
not fulltime, female	5.86%	4.89%	4.02%	5.40%	4.49%	3.69%
not fulltime, male	8.77%	7.44%	6.22%	8.15%	6.89%	5.75%
fulltime, female	3.59%	2.94%	2.38%	3.29%	2.68%	2.16%
fulltime, male	5.61%	4.67%	3.84%	5.17%	4.30%	3.52%

The results in Table 4.10 can also be depicted as a bar chart. Figure 4.5 shows that white respondents are more likely to get depressed for a long period than African American or Hispanic respondents.

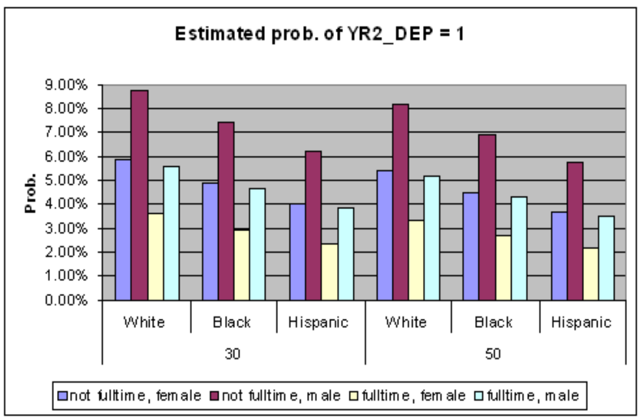


Figure 4.5: Estimated probabilities for subgroups

Model comparison

Since the two models in this section are nested models, the χ^2 difference test can be used. The deviances, AIC, and SBC statistics for these models are summarized in Table 4.11. These statistics suggest that the second model fits the data better.

Table 4.11: Model comparison

Statistic	Model 1	Model 2	difference	Difference in d.f.
$-2\ln L$ (deviance statistic)	17203.107	17176.677	26.430	2
Akaike Information Criterion	17213.107	17190.677	22.430	2
Schwarz Criterion	17256.316	17251.169	5.147	2