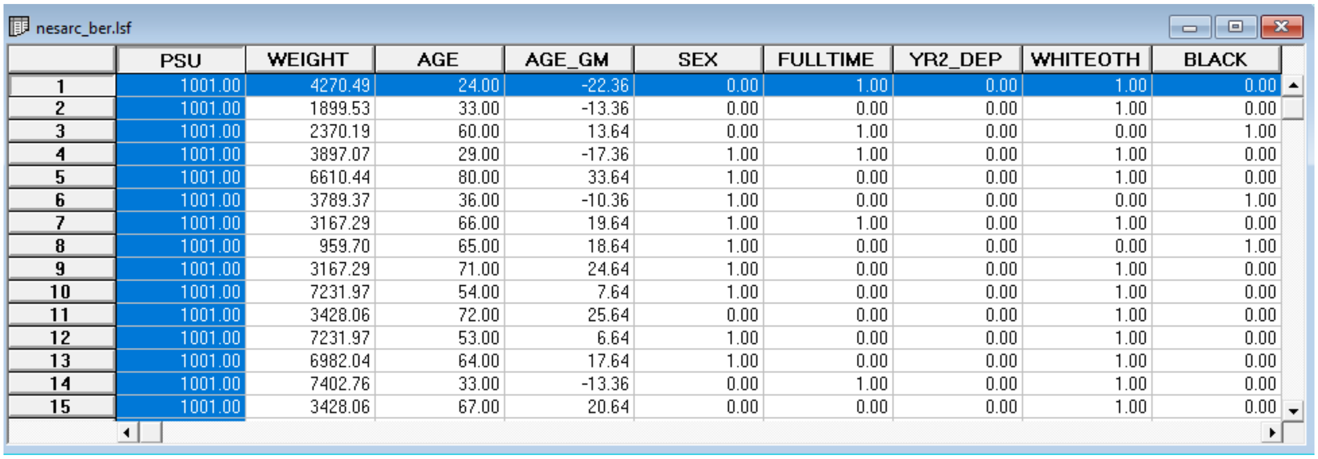
Bernoulli model for NESARC data

Contents

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 contains information on the occurrences of major depression, family history of major depression and dysthymia of dysthymia respondents.



nesarc_ber.lsf

	PSU	WEIGHT	AGE	AGE_GM	SEX	FULLTIME	YR2_DEP	WHITEOTH	BLACK
1	1001.00	4270.49	24.00	-22.36	0.00	1.00	0.00	1.00	0.00
2	1001.00	1899.53	33.00	-13.36	0.00	0.00	0.00	1.00	0.00
3	1001.00	2370.19	60.00	13.64	0.00	1.00	0.00	0.00	1.00
4	1001.00	3897.07	29.00	-17.36	1.00	1.00	0.00	1.00	0.00
5	1001.00	6610.44	80.00	33.64	1.00	0.00	0.00	1.00	0.00
6	1001.00	3789.37	36.00	-10.36	1.00	0.00	0.00	0.00	1.00
7	1001.00	3167.29	66.00	19.64	1.00	1.00	0.00	1.00	0.00
8	1001.00	959.70	65.00	18.64	1.00	0.00	0.00	0.00	1.00
9	1001.00	3167.29	71.00	24.64	1.00	0.00	0.00	1.00	0.00
10	1001.00	7231.97	54.00	7.64	1.00	0.00	0.00	1.00	0.00
11	1001.00	3428.06	72.00	25.64	0.00	0.00	0.00	1.00	0.00
12	1001.00	7231.97	53.00	6.64	1.00	0.00	0.00	1.00	0.00
13	1001.00	6982.04	64.00	17.64	1.00	0.00	0.00	1.00	0.00
14	1001.00	7402.76	33.00	-13.36	0.00	1.00	0.00	1.00	0.00
15	1001.00	3428.06	67.00	20.64	0.00	0.00	0.00	1.00	0.00

The variables of interest are:

- PSU denotes the Census 2000/2001 Supplementary Survey (C2SS) primary sampling unit.

- WEIGHT represents the NESARC weights sample results used to **estimate the national level statistics.**
- AGE represents the age at onset of first major depression episode, while AGE_GM is AGE, grand mean centered. That is AGE- 46.357, where 46.357 is the mean value of AGE.
- SEX is the gender of the respondent, where 0 denotes males and 1 females.
- FULLTIME =1 if a respondent is employed fulltime and 0 otherwise
- YR2_DEP is the outcome variable of interest, and indicates whether a patient had an episode or not.
- WHITEOTH is an ethnicity dummy variable coded 1 if white and 0 otherwise
- BLACK is an ethnicity dummy variable coded 1 if black and 0 otherwise
- HISP is an ethnicity dummy variable coded 1 if Hispanic and 0 otherwise.
- YOUNG, MIDDLE and OLD are 0,1 dummy variables representing AGE

2. Exploring the data

Inspecting the distribution of the variables before starting with the model is important. Using the **Data Screening** option, we obtain the univariate distributions shown below. We note that the proposed outcome variable indicates observed occurrences for only 5.6% of the observed data. There are considerably more females than females among respondents.

```
Univariate Distributions for Ordinal Variables

 SEX              Frequency Percentage Bar Chart
      0   17942        42.9    □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
      1   23907        57.1    □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

 FULLTIME         Frequency Percentage Bar Chart
      0   20203        48.3    □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
      1   21646        51.7    □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

 YR2_DEP          Frequency Percentage Bar Chart
      0   39510        94.4    □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
      1    2339         5.6    □□□

 WHITEOTH         Frequency Percentage Bar Chart
      0   16028        38.3    □□□□□□□□□□□□□□□□□□□□□□□□□□□□□
      1   25821        61.7    □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

 BLACK            Frequency Percentage Bar Chart
      0   33570        80.2    □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
      1    8279        19.8    □□□□□□□□□□□□

 HISP             Frequency Percentage Bar Chart
      0   34100        81.5    □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
      1    7749        18.5    □□□□□□□□□□□

 YOUNG            Frequency Percentage Bar Chart
      0   29211        69.8    □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
      1   12638        30.2    □□□□□□□□□□□□□□□□□□□□

 MIDDLE           Frequency Percentage Bar Chart
      0   29145        69.6    □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
      1   12704        30.4    □□□□□□□□□□□□□□□□□□□□
```

```
OLD              Frequency Percentage Bar Chart
     0   25342        60.6    □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
     1    16507        39.4    □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
```

3. The model

The first model fitted to the data explores the relationship between YR2_DEP and the variables indicating age, employment, gender and ethnicity.

For the binary case with logit link considered here

$$\text{Prob}(YR2\_DEP_{ij} = 1) = \frac{e^{\eta_{ij}}}{1+e^{\eta_{ij}}}$$

where η_{ij} represents the log of the odds of success, in this case the log odds of an episode during the time period. With the logit link function, the probability $\text{Prob}(y_{ij} = 1 | \boldsymbol{\beta})$ is transformed to lie in the interval (0,1). The two-level model, taking PSU into account as the level-2 ID can be expressed as

The level-1 model is

$$\begin{gathered}\log\left(\eta_{ij}\right) = \beta_0 + \beta_1 \times AGE\_GM_{ij} + \beta_2 \times FULLTIME_{ij} + \beta_3 \times SEX_{ij} + \\ \beta_4 \times BLACK_{ij} + \beta_5 \times YOUNG_{ij} + \beta_6 \times MIDDLE_{ij} + e_{ij}\end{gathered}$$

The level-2 model is

$$\begin{aligned} b_{0i} &= \beta_0 + u_{0i} \\ b_{1i} &= \beta_1 + u_{1i} \\ b_{2i} &= \beta_2 \\ b_{3i} &= \beta_3 \\ b_{4i} &= \beta_4 \\ b_{5i} &= \beta_5 \\ b_{6i} &= \beta_6 \end{aligned}$$

where

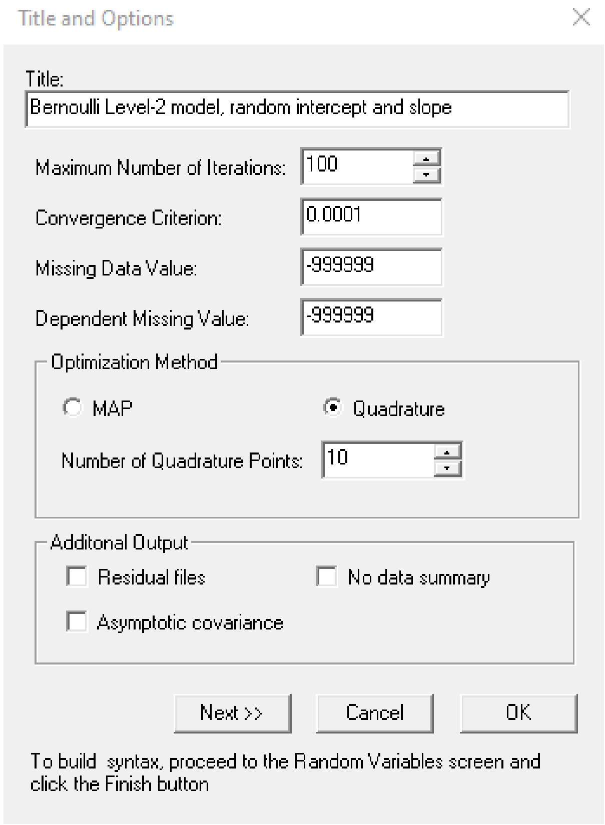
$$\begin{gathered} e_i \sim N\left(0, \sigma^2 \mathbf{I}_i\right) \\ \mathbf{u}_i \sim N\left(0, \boldsymbol{\Sigma}_i\right) \end{gathered}$$

β_0 denotes the average expected η_{ij}, which can be converted to the expected probability of an episode occurring in the time period studied. β_3 denotes the coefficient of the predictor variable SEX (slope) in the fixed part of the model. The random coefficients u_{i0} and e_{ij} denote the variation in the average expected YR2_DEP value between PSUs and between patients respectively.

4. Setting up the analysis

Open the LISREL spreadsheet **nesarc_ber.lsf** used during the exploratory analysis discussed previously. The next step is to describe the model to be fitted. We use the LISREL interface to provide the model specifications. From the main menu bar, select the **Multilevel, Generalized Linear Model, Title and Options** option.

The multilevel generalized linear model contains five consecutive dialogs boxes. The **Titles and Options** dialog box as shown below enables the user to input the title, maximum number of iteration, convergence criterion, missing values, and method and request additional output. Enter a title for the analysis in the **Title** text boxes (optional) and keep all the other settings as default.



Title and Options

Title: Bernoulli Level-2 model, random intercept and slope

Maximum Number of Iterations: 100

Convergence Criterion: 0.0001

Missing Data Value: -999999

Dependent Missing Value: -999999

Optimization Method

○ MAP ● Quadrature

Number of Quadrature Points: 10

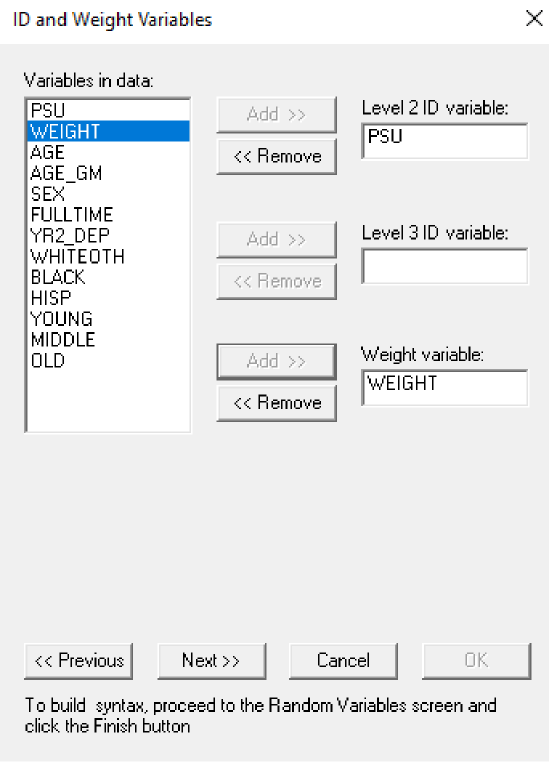
Additonal Output

□ Residual files □ No data summary

□ Asymptotic covariance

To build syntax, proceed to the Random Variables screen and click the Finish button

Proceed to the **ID and Weights** screen by clicking on the **Next** button. Highlight PSU from the **Variables in data** list and click on the upper **Add** button to select is as the **Level-2 ID variable**. Similarly, highlight the variable WEIGHT and click on the lower **Add** button to select it as the **Weight variable** and obtain the screen shown below.



Variables in data:

PSU
WEIGHT
AGE
AGE_GM
SEX
FULLTIME
YR2_DEP
WHITEOTH
BLACK
HISP
YOUNG
MIDDLE
OLD

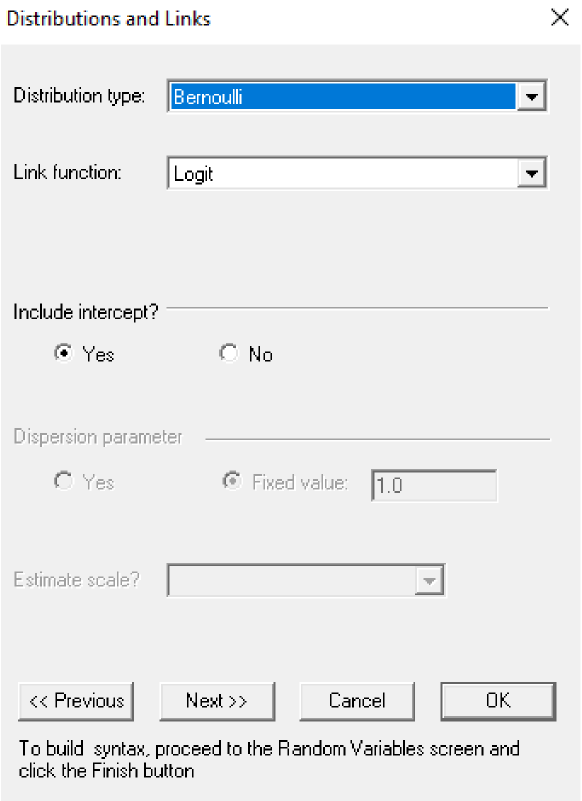
Level 2 ID variable: PSU

Level 3 ID variable:

Weight variable: WEIGHT

To build syntax, proceed to the Random Variables screen and click the Finish button

Click on the **Next** button to load the **Distribution and Links** dialog box. Select **Binomial** from the **Distribution type** dropdown list box. By default, the logit link function is selected. Keep the other default settings unchanged as shown below, and click on the **Next** button.



Distribution type: Bernoulli

Link function: Logit

Include intercept?

● Yes ○ No

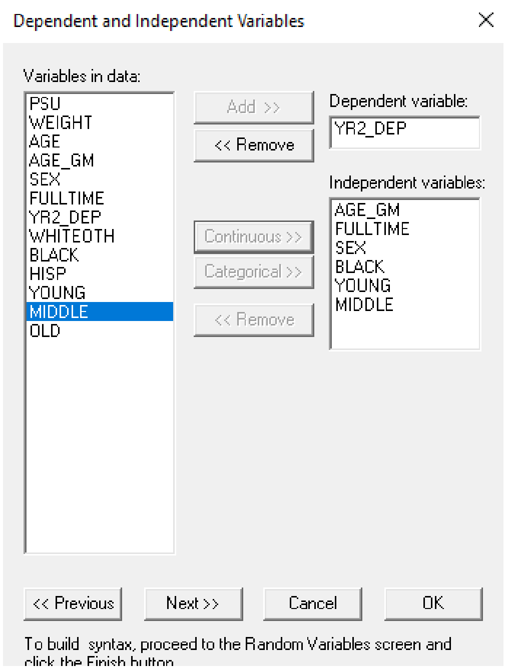
Dispersion parameter

○ Yes ● Fixed value: 1.0

Estimate scale?

To build syntax, proceed to the Random Variables screen and click the Finish button

On the **Dependent and Independent Variables** dialog box screen, first select YR2_DEP and click on the upper **Add** button to define it as the **Dependent variable**. Then, select SEX and the other predictors and click on the **Continuous** button to add these variables in the **Independent variables** list box as shown below.



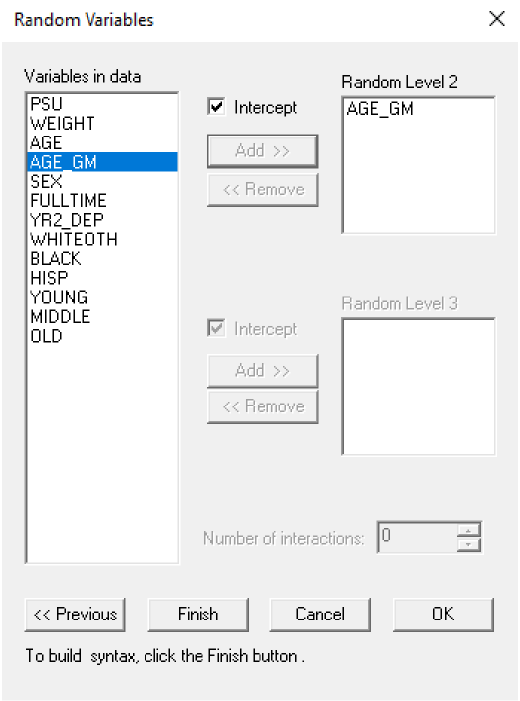
Dependent and Independent Variables

Variables in data: PSU, WEIGHT, AGE, AGE_GM, SEX, FULLTIME, YR2_DEP, WHITEOTH, BLACK, HISP, YOUNG, MIDDLE, OLD

Dependent variable: YR2_DEP

Independent variables: AGE_GM, FULLTIME, SEX, BLACK, YOUNG, MIDDLE

To build syntax, proceed to the Random Variables screen and click the Finish button



Random Variables

Variables in data: PSU, WEIGHT, AGE, AGE_GM, SEX, FULLTIME, YR2_DEP, WHITEOTH, BLACK, HISP, YOUNG, MIDDLE, OLD

☑ Intercept

Random Level 2: AGE_GM

☑ Intercept

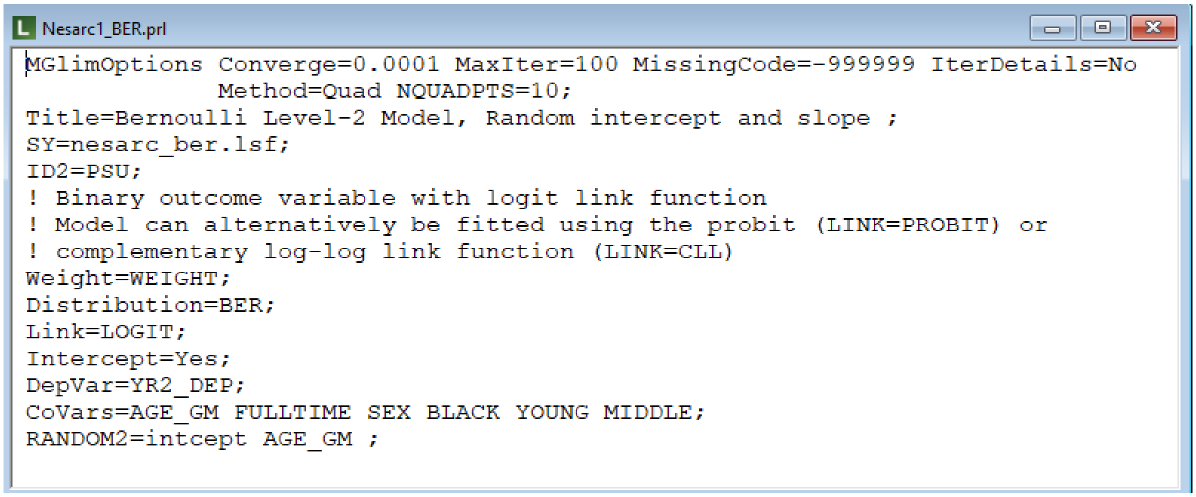
Random Level 3:

Number of interactions: 0

To build syntax, click the Finish button .

Click on the **Next** button to proceed to the **Random Variables** dialog box once these settings have been defined. Keep the **Intercept** check box checked so as to include a level-2 intercept and add AGE_GM to the **Random Level-2** field to allow this slope to vary randomly over PSUs.

Click on the **Finish** button to generate the PRELIS syntax file (**.prl**) that corresponds to the above settings. Select the **File**, **Save As** option, and provide a name (**nesarc1_BER.prl**) for the model specification file. The default folder for the syntax to be saved in is the same folder used for the data file.



```
MGlimOptions Converge=0.0001 MaxIter=100 MissingCode=-999999 IterDetails=No
             Method=Quad NQUADPTS=10;
Title=Bernoulli Level-2 Model, Random intercept and slope ;
SY=nesarc_ber.lsf;
ID2=PSU;
! Binary outcome variable with logit link function
! Model can alternatively be fitted using the probit (LINK=PROBIT) or
! complementary log-log link function (LINK=CLL)
Weight=WEIGHT;
Distribution=BER;
Link=LOGIT;
Intercept=Yes;
DepVar=YR2_DEP;
CoVars=AGE_GM FULLTIME SEX BLACK YOUNG MIDDLE;
RANDOM2=intcept AGE_GM ;
```

The syntax file

The syntax file contains the following information:

- The MGlimOptions keyword requests the MGLIM module to run. The first two lines, together with the Title line, correspond to the settings entered in the **Title and Options** dialog box.
- The SY line indicates the location and name of the ***.lsf** data file.
- PSU is the level-2 ID variable, while Weight corresponds with the weight variable. These are defined in the **ID and Weights** dialog box.
- The syntax lines for Distribution, Link and level-1 Intercept are set up in the **Distribution and Links** dialog box.
- The DepVar line, which represents the dependent variable and the CoVars line, which represents the covariate variable, are defined in the **Dependent and Independent Variables** dialog box.
- Finally, the RANDOM2 syntax line corresponds to the **Random Variables** dialog box.

Understanding how the syntax works enables the user to make changes directly to the syntax file. Run the analysis by selecting the **Run PRELIS** button to generate the output file. The output file has the same file name as the syntax file with a different extension **.out**. It is saved in the same folder as the syntax file.

5. Discussion of results

Portions of the output file are shown below.

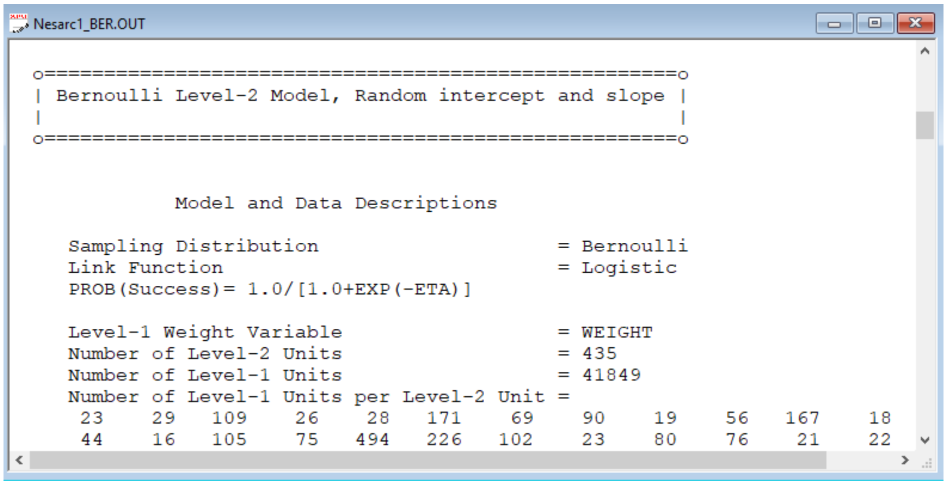
Program information and syntax

At the top of the output file, program information is given. It states the version number, corporate and technical support information, the date and time of analysis, and the locations of data file and syntax file.

Program information is followed by the Multilevel GLIM syntax. This section echoes the contents of the syntax.

Model and data description

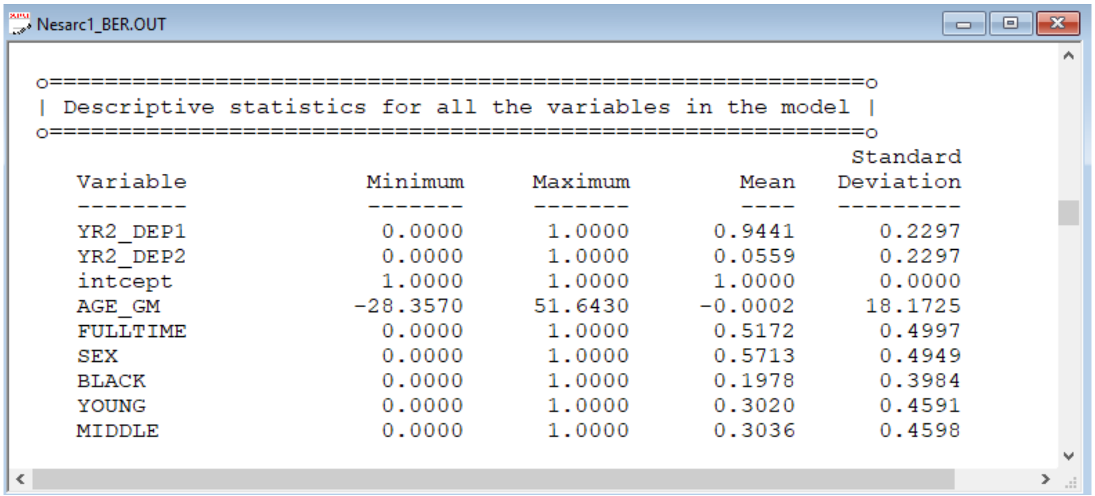
In the next section of the output file as shown above, descriptions of the distribution, the link function, the weight variable and the hierarchical structure of the data are provided. Data from a total of 435 level-2 units and 41,849 respondents were included at levels 2 and 1 of the model. In addition, a summary of the number of respondents nested within each level-2 unit is provided.



```
o=====================================================o
| Bernoulli Level-2 Model, Random intercept and slope |
|                                                     |
o=====================================================o


            Model and Data Descriptions

   Sampling Distribution                     = Bernoulli
   Link Function                             = Logistic
   PROB(Success)= 1.0/[1.0+EXP(-ETA)]

   Level-1 Weight Variable                   = WEIGHT
   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
```

Descriptive statistics

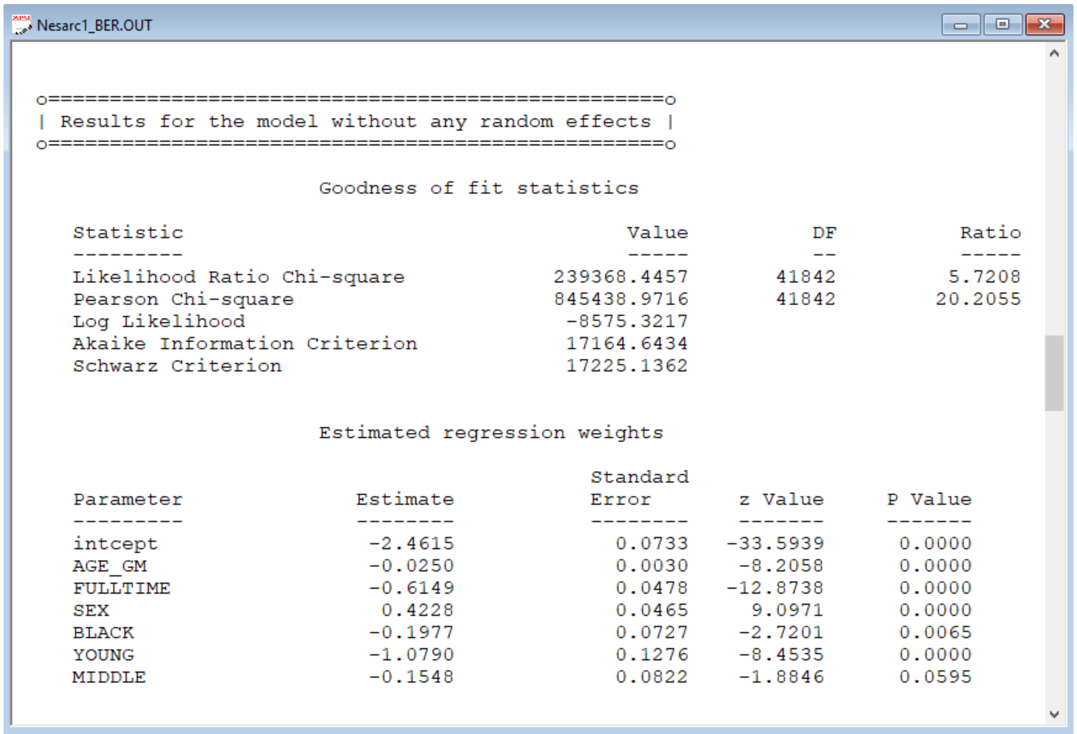
The data summary is followed by descriptive statistics for all the variables included in the model. Since YR2_DEP is defined as a binary variable, it is presented by two dummy variables YR2_DEP1 and YR2_SEP2.



```
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
  intcept                1.0000       1.0000      1.0000     0.0000
  AGE_GM               -28.3570      51.6430     -0.0002    18.1725
  FULLTIME               0.0000       1.0000      0.5172     0.4997
  SEX                    0.0000       1.0000      0.5713     0.4949
  BLACK                  0.0000       1.0000      0.1978     0.3984
  YOUNG                  0.0000       1.0000      0.3020     0.4591
  MIDDLE                 0.0000       1.0000      0.3036     0.4598
```

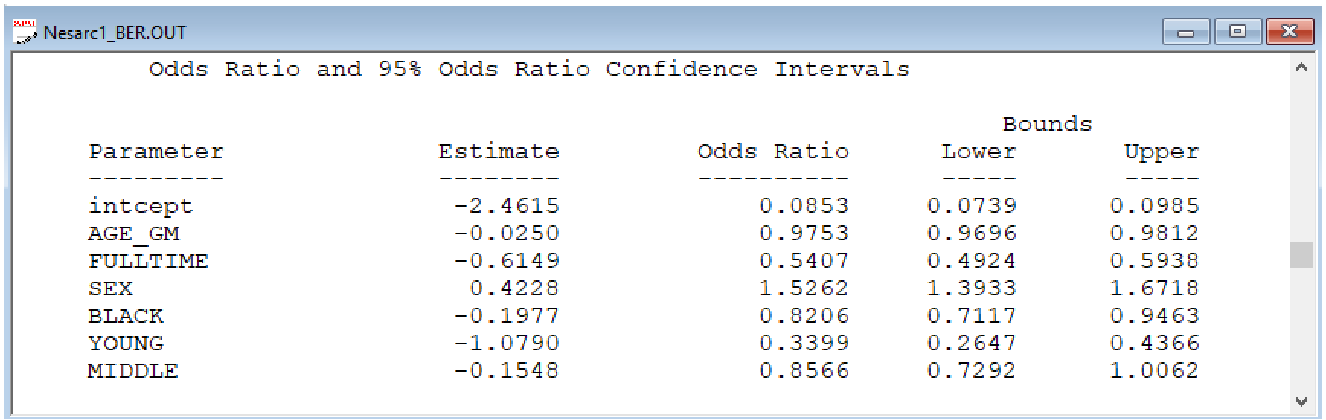
Results for the model without any random effects

Descriptive statistics are followed by the results for the model without any random effects. These parameters are used in the initial step of the iterative algorithm. They are obtained by ordinary weighted least squares (WLS) regression. The goodness of WLS fit statistics are also given as shown below.



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

                    Goodness of fit statistics

  Statistic                                Value          DF          Ratio
  ---------                                -----          --          -----
  Likelihood Ratio Chi-square        239368.4457       41842         5.7208
  Pearson Chi-square                 845438.9716       41842        20.2055
  Log Likelihood                      -8575.3217
  Akaike Information Criterion        17164.6434
  Schwarz Criterion                   17225.1362


                    Estimated regression weights

                                          Standard
  Parameter               Estimate           Error     z Value     P Value
  ---------               --------        --------    --------     -------
  intcept                  -2.4615          0.0733    -33.5939      0.0000
  AGE_GM                   -0.0250          0.0030     -8.2058      0.0000
  FULLTIME                 -0.6149          0.0478    -12.8738      0.0000
  SEX                       0.4228          0.0465      9.0971      0.0000
  BLACK                    -0.1977          0.0727     -2.7201      0.0065
  YOUNG                    -1.0790          0.1276     -8.4535      0.0000
  MIDDLE                   -0.1548          0.0822     -1.8846      0.0595
```

This is followed by odds ratios and 95% confidence intervals for the odds ratios.



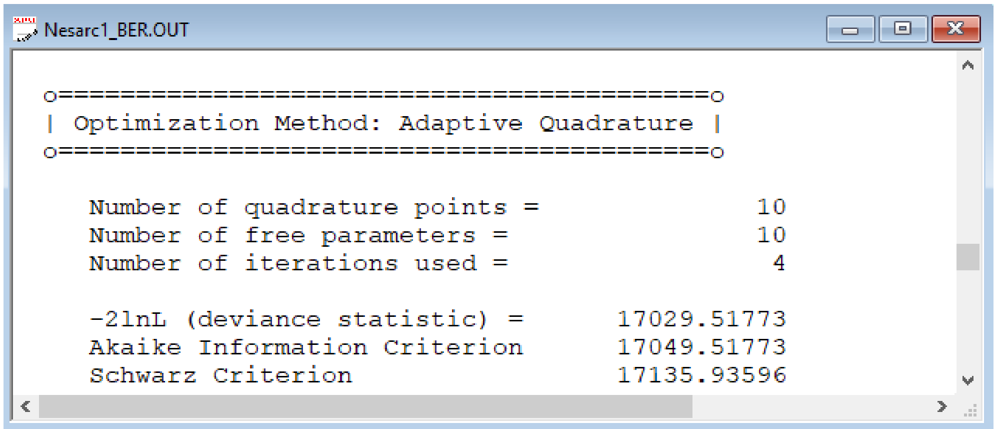
```
         Odds Ratio and 95% Odds Ratio Confidence Intervals

                                                                  Bounds
     Parameter                Estimate         Odds Ratio      Lower        Upper
     ---------                --------         ----------      -----        -----
     intcept                   -2.4615             0.0853     0.0739       0.0985
     AGE_GM                    -0.0250             0.9753     0.9696       0.9812
     FULLTIME                  -0.6149             0.5407     0.4924       0.5938
     SEX                        0.4228             1.5262     1.3933       1.6718
     BLACK                     -0.1977             0.8206     0.7117       0.9463
     YOUNG                     -1.0790             0.3399     0.2647       0.4366
     MIDDLE                    -0.1548             0.8566     0.7292       1.0062
```

Results for the model with fixed and random effects

Number of iterations and fit statistics

The total number of (macro) iterations is reported. As shown below, there are 58 iterations to get the converged results.

In addition to the likelihood function value at convergence, a number of related statistical measures for assessing model adequacy are available. The most common of these are the likelihood ratio test, Pearson chi-square, and Akaike's and Schwarz's criteria. Both the Akaike information criterion (AIC) and the Schwarz Bayesian criterion (SBC) are functions of the number of estimated parameters, and therefore "penalize" models with large numbers of parameters. In the LISREL output file, all three of these are reported. A chi-square scale factor, with which a chi-square value obtained from the difference between two deviance statistics should be multiplied to yield a corrected chi-square statistic in the case of a weighted analysis, may also be found in this section.



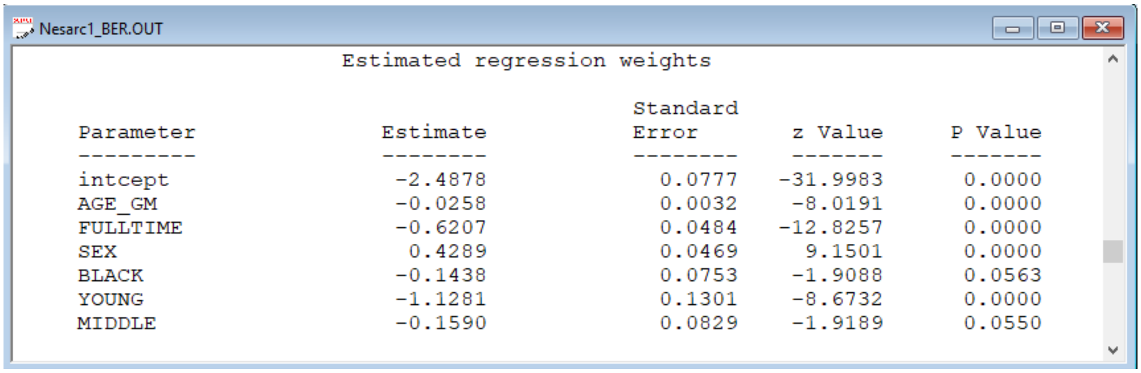
```
 o============================================o
 | Optimization Method: Adaptive Quadrature |
 o============================================o

    Number of quadrature points =              10
    Number of free parameters =                10
    Number of iterations used =                 4

    -2lnL (deviance statistic) =      17029.51773
    Akaike Information Criterion      17049.51773
    Schwarz Criterion                 17135.93596
```

- The Pearson Chi-square is defined as $\chi_P^2 = \sum_{i=1}^{N}\sum_{j=1}^{n_i}\sum_{k=1}^{n_{ij}} \frac{w_{ijk}\left(y_{ijk}-\hat{\mu}_{ijk}\right)^2}{\hat{\sigma}^2\left(y_{ijk}\right)}$.
- The deviance is defined as $-2\ln L$. For a pair of nested models, the difference in $-2\ln L$ values has a χ^2 distribution, with degrees of freedom equal to the difference in number of parameters estimated in the models compared.

- The AIC was originally proposed for time-series models, but is also used in regression. It is defined as $-2\ln L + 2r$, where r denotes the number of parameters estimated in the model. The model with minimum AIC, in a set of nested models, will be the most parsimonious according to this criterion.
- The SBC is defined as $-2\ln L + r\log n$, where n denotes the number of units at the highest level of the hierarchy. A smaller value of this criterion would indicate the most parsimonious of the models being compared.

Estimated regression weights

The output describing the estimated regression weights after fit statistics is shown next. The estimates are shown in the column with heading Estimate and correspond to the coefficients β_0, β_1 etc. in the model specification. From the z-values and associated exceedance probabilities, we see that most of the estimates are highly significant at 10% level.



```
                    Estimated regression weights

                                                  Standard
     Parameter                  Estimate          Error          z Value      P Value
     ---------                  ---------         ---------      -------      -------
     intcept                      -2.4878           0.0777      -31.9983       0.0000
     AGE_GM                       -0.0258           0.0032       -8.0191       0.0000
     FULLTIME                     -0.6207           0.0484      -12.8257       0.0000
     SEX                           0.4289           0.0469        9.1501       0.0000
     BLACK                        -0.1438           0.0753       -1.9088       0.0563
     YOUNG                        -1.1281           0.1301       -8.6732       0.0000
     MIDDLE                       -0.1590           0.0829       -1.9189       0.0550
```

The estimated intercept is -2.4878, which is the average logit. The estimated coefficients associated with gender (SEX) is 0.4289, which indicates that the female respondents (SEX = 1) have a larger $\hat{\eta}$. The estimate for the indicator of race (BLACK) shows that white clients have higher $\hat{\eta}$ values. Younger respondents have lower $\hat{\eta}$ values than older respondents judging by the size of the estimates for the two variables YOUNG and MIDDLE. To describe the η's in a more accessible way to readers of reports, we need the link functions to transform them into probabilities.

Interpreting estimated regression weights by using link function

We consider four respondents: all white (BLACK=0) of AGE equal to the grand mean. As the grand mean age is 46.357, both respondents would thus have a value of 1 on the variable MIDDLE and a value of 0 on the other age-related variable YOUNG. This implies that the 4 respondents only differ on gender and employment. We substitute the regression weights and obtain the function for $\hat{\eta}_{ij}$

$$\begin{aligned}\hat{\eta}_{ij} &= \hat{b}_{0i} + \hat{b}_{3i} \times (\text{SEX})_{ij} + \hat{b}_{2i} \times (\text{FULLTIME})_{ij} \\ &= -2.4878 + 0.4289 \times (\text{SEX})_{ij} - 0.6207 \times (\text{FULLTIME})_{ij}\end{aligned}.$$

By substituting the values of gender and employment status, we obtain four estimates of η_{ij}. Next, we transform the $\hat{\eta}_{ij}$'s into corresponding probabilities by using the logit link function

$$\text{Prob}(\text{DEPR}_{ij} = 1) = \frac{e^{\hat{\eta}_{ij}}}{1+e^{\hat{\eta}_{ij}}} = \frac{1}{1+e^{-\hat{\eta}_{ij}}} =$$

to obtain the results in the table below.

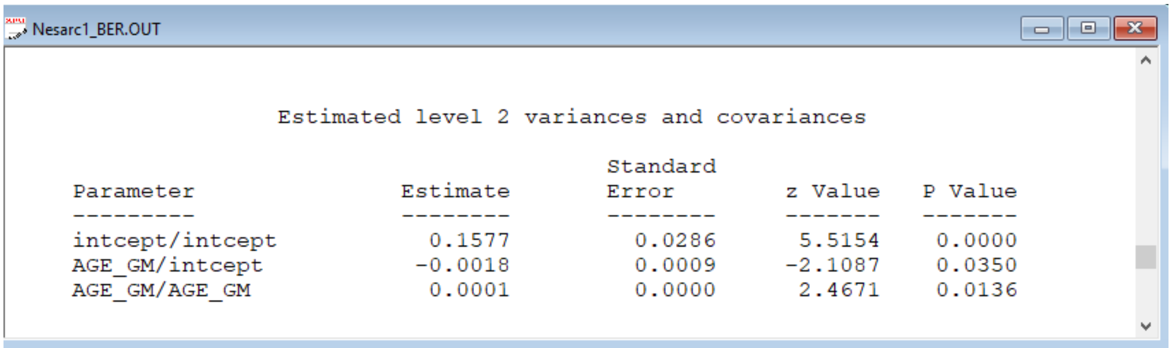
Respondent	**Code**	$\hat{\eta}$	**Prob (DEPR = 1)**
Male, employed	sex = 0, fulltime = 1	-3.1085	4.28%
Male, not employed	sex = 0, fulltime = 0	-2.4878	7.67%
Female, employed	sex = 1, fulltime = 1	-2.6796	6.42%
Female, not employed	sex = 1, fulltime = 0	-2.0589	11.32%

We can conclude that the estimated probability for a unemployed female is the highest at 11.32%. From the results, we conclude that females are more likely to have an episode, and this risk increases if they are not employed full time.

Group	**Code**	$\hat{\eta}$	**Prob (DEPR = 1)**
Black, male	sex = 0, race_d = 0	0.1018	47.53%
Black, female	sex = 1, race_d = 0	0.6848	66.48%
White, male	sex = 0, race_d = 1	-0.7450	32.19%
White, female	sex = 1, race_d = 1	0.0388	50.97%

Estimated level-2 variance

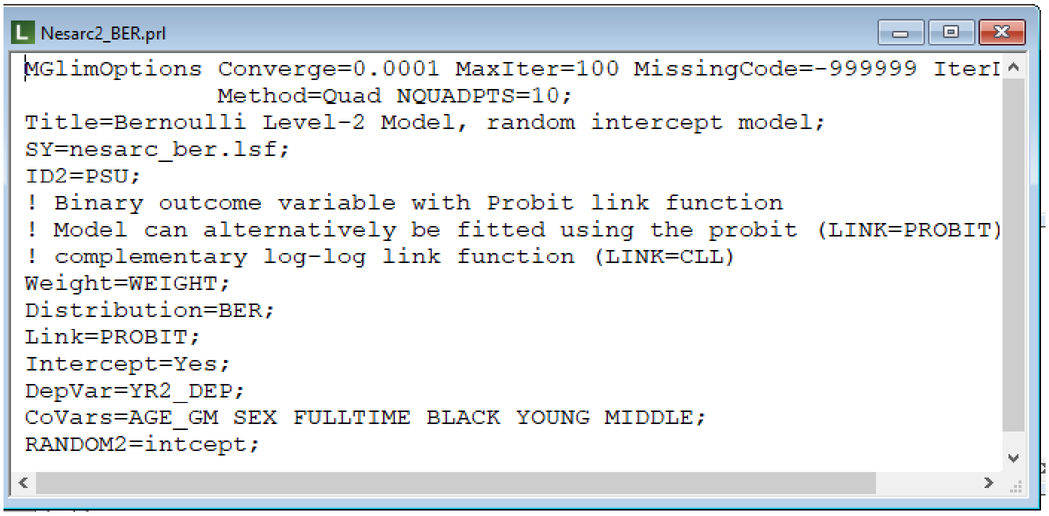
The output for the estimated level-2 variance is shown in the image below. The *p* value of intercept shows the probability of having an episode within the time period of study differs significantly from PSU to PSU (the level-2 units).



```
                    Estimated level 2 variances and covariances

                                                Standard
     Parameter                Estimate          Error          z Value     P Value
     ---------                --------          --------       -------     -------
     intcept/intcept            0.1577            0.0286         5.5154      0.0000
     AGE_GM/intcept            -0.0018            0.0009        -2.1087      0.0350
     AGE_GM/AGE_GM              0.0001            0.0000         2.4671      0.0136
```

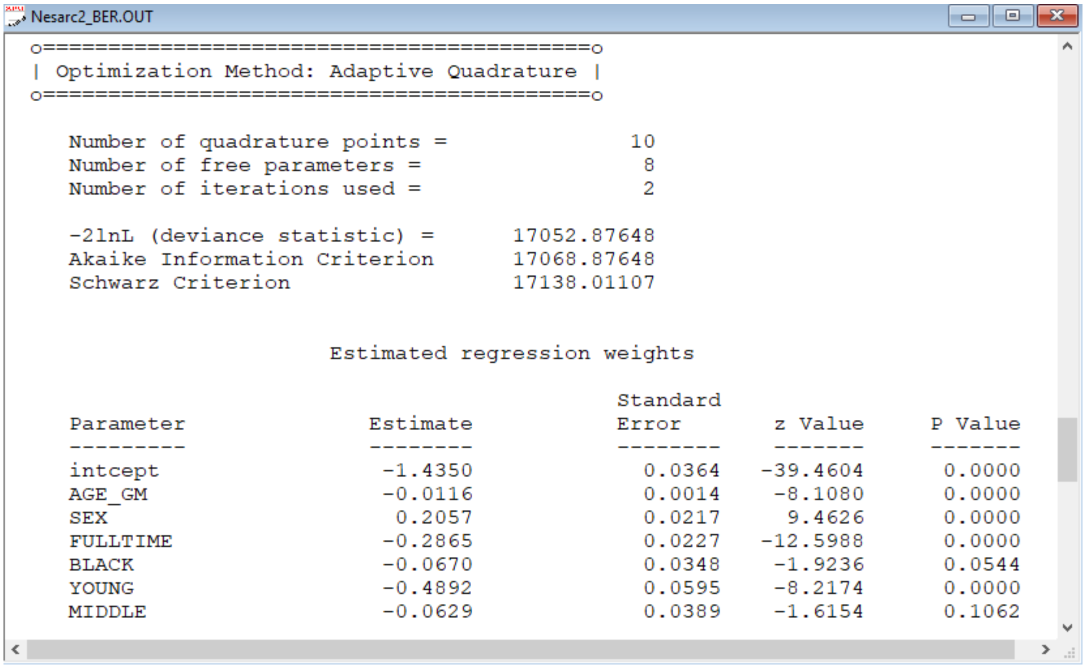
6. Alternative model

Alternatively, we can fit a model with probit link function to these data. The syntax for this model is given in **nesarc2_ber.prl**.



```
MGlimOptions Converge=0.0001 MaxIter=100 MissingCode=-999999 IterI
             Method=Quad NQUADPTS=10;
Title=Bernoulli Level-2 Model, random intercept model;
SY=nesarc_ber.lsf;
ID2=PSU;
! Binary outcome variable with Probit link function
! Model can alternatively be fitted using the probit (LINK=PROBIT)
! complementary log-log link function (LINK=CLL)
Weight=WEIGHT;
Distribution=BER;
Link=PROBIT;
Intercept=Yes;
DepVar=YR2_DEP;
CoVars=AGE_GM SEX FULLTIME BLACK YOUNG MIDDLE;
RANDOM2=intcept;
```

From the results of this analysis, we see higher fit statistics, leading us to conclude that the previous model fitted the data better. Though the estimates are obviously different, we note tht gender is again the only predictor with a positive estimated regression weight.



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

   Number of quadrature points =             10
   Number of free parameters =                8
   Number of iterations used =                2

   -2lnL (deviance statistic) =      17052.87648
   Akaike Information Criterion      17068.87648
   Schwarz Criterion                 17138.01107


                        Estimated regression weights

                                           Standard
   Parameter              Estimate          Error       z Value      P Value
   ---------              --------         --------     -------      -------
   intcept                 -1.4350           0.0364    -39.4604       0.0000
   AGE_GM                  -0.0116           0.0014     -8.1080       0.0000
   SEX                      0.2057           0.0217      9.4626       0.0000
   FULLTIME                -0.2865           0.0227    -12.5988       0.0000
   BLACK                   -0.0670           0.0348     -1.9236       0.0544
   YOUNG                   -0.4892           0.0595     -8.2174       0.0000
   MIDDLE                  -0.0629           0.0389     -1.6154       0.1062
```

The final output is the population-average results. The regression parameters in multilevel generalized linear models have the “unit-specific” or conditional interpretation, in contrast to the “population-average” or marginal estimates that represent the unconditional covariate effects. LISREL uses numerical quadrature to obtain population-average estimates from their unit-specific counterparts in models with multiple random effects. Standard errors for the population-average estimates are derived using the delta method.

Under the model we fitted, the predicted probability for case *ij*, given u_{0j}, would be

$$E\left(Y_{ij} \mid u_{0j}\right)=\frac{1}{1+\exp\left\{-\left(\begin{array}{l}\beta_0+\beta_1\times AGE\_GM_{ij}+\beta_2\times FULLTIME_{ij}+\beta_3\times SEX_{ij}+\\ \beta_4\times BLACK_{ij}+\beta_5\times YOUNG_{ij}+\beta_6\times MIDDLE_{ij}+u_{0j}\end{array}\right)\right\}}.$$

while the population average model would be

$$E\left(Y_{ij} \mid\right)=\frac{1}{1+\exp\left\{-\left(\begin{array}{l}\beta_0+\beta_1\times AGE\_GM_{ij}+\beta_2\times FULLTIME_{ij}+\beta_3\times SEX_{ij}+\\ \beta_4\times BLACK_{ij}+\beta_5\times YOUNG_{ij}+\beta_6\times MIDDLE_{ij}\end{array}\right)\right\}}.$$

Users will need to take care in choosing unit-specific versus population-average results for their research. The choice will depend on the specific research questions that are of interest. If one were primarily interested in how a change in one of the predictors, for example FULLTIME, can be expected to affect a particular individual PSU's mean, one would use the unit-specific model. If one were interested in how a change in FULLTIME can be expected to affect the overall population mean, one would use the population-average model.