



## Binary models with probit link function

### Contents

1.	The model.....	1
2.	Setting up the analysis .....	2
3.	The syntax file .....	5
4.	Discussion of results.....	5

### 1. The model

In the model just discussed the logistic link function was used. We now will fit another model by using the probit link function. We will only include a random intercept at level-2. On the other hand, the interaction between gender and race will be considered at level-1. The model can be expressed as follows.

The probit link function is

$$\text{Prob}(\text{DEPR}_{ij} = \gamma) = \Phi^{-1}(\eta)$$

where  $\eta_{ij}$  represents the log of the odds of success. With the probit link function, the probability  $\text{Prob}(y_{ij} = 1 | \beta)$  is transformed to lie in the interval (0,1). And (for the current model) the two-level model can be expressed as

Level-1 model:

$$\eta_{ij} = b_{0i} + b_{1i} \times (\text{SEX})_{ij} + b_{2i} \times (\text{RACE\_d})_{ij} + b_{3i} \times (\text{SEXxRACE})_{ij} + e_{ij}$$

Level-2 model:

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

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

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

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

The current model only assumes the intercepts differ between PSUs, but the slopes are the same for all the level-2 units.

## 2. Setting up the analysis

Open the LISREL spreadsheet **depress.isf** used previously, Select the **Multilevel, Generalized Linear Model, Title and Options** option. Input the new analysis title in the **Titles and Options** dialog box and keep all the other settings as default.

The screenshot shows the 'Title and Options' dialog box with the following settings:

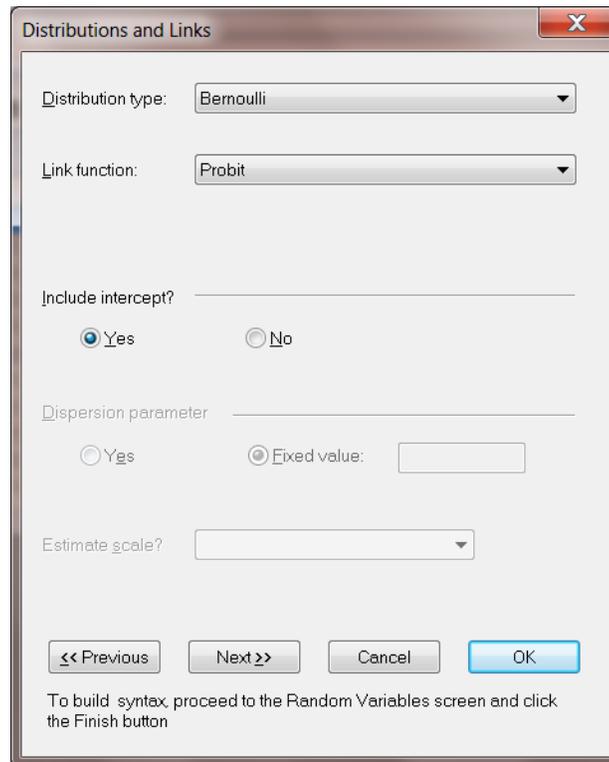
- Title: Bernoulli-Probit model based on the Depression Data
- Maximum Number of Iterations: 100
- Convergence Criterion: 0.0001
- Missing Data Value: -999999
- Dependent Missing Value: -999999
- Optimization Method:  Quadrature,  MAP
- Number of Quadrature Points: 10
- Additional Output:  Residual files,  No data summary,  Asymptotic covariance

Buttons: Next >>, Cancel, OK

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. The settings of this screen are exactly the same as the previous model. Highlight LEV2ID from the **Variables in data** list and click on the upper **Add** button to select it as the **Level-2 ID variable**. Similarly, highlight the A2TWA0 and click on the lower **Add** button to select it as the **Weight variable** to obtain the screen shown below.

Click on the **Next** button to load the **Distribution and Links** dialog box. Select **Binomial** from the **Distribution type** dropdown list box. Select **Probit** from the **Link function** drop-down list. Keep the other default settings unchanged as shown below, and click on the **Next** button.

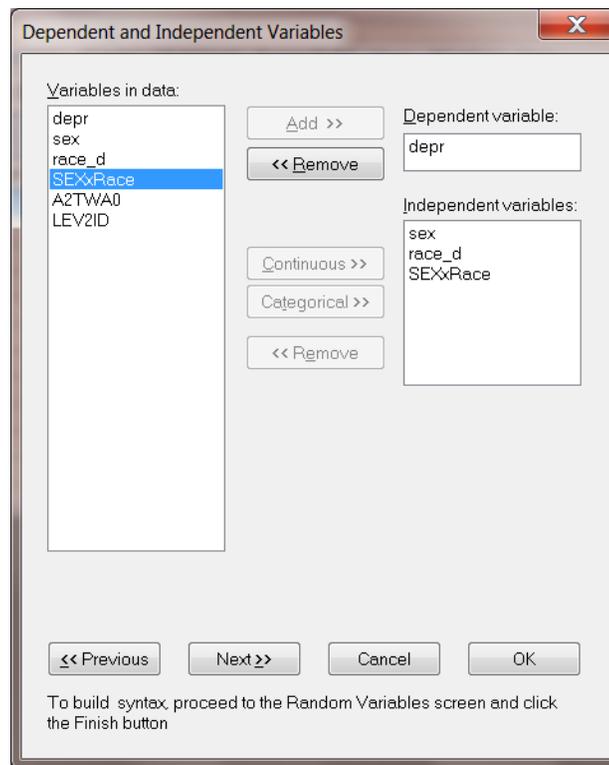


The screenshot shows the 'Distributions and Links' dialog box with the following settings:

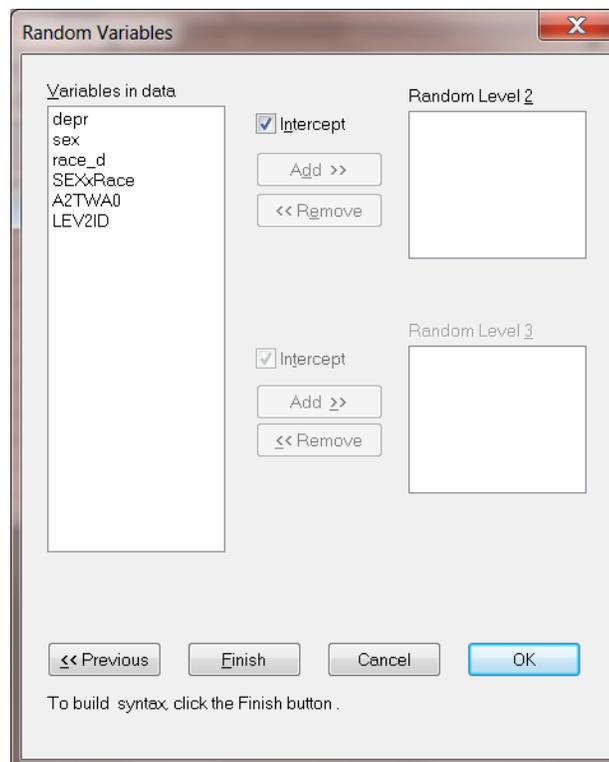
- Distribution type:** Bernoulli
- Link function:** Probit
- Include intercept?:** Yes (selected)
- Dispersion parameter:** Fixed value (selected) with an empty input field
- Estimate scale?:** No

At the bottom, there are four buttons: '<< Previous', 'Next >>', 'Cancel', and 'OK'. The 'Next >>' button is highlighted in blue. Below the buttons, there is a note: 'To build syntax, proceed to the Random Variables screen and click the Finish button'.

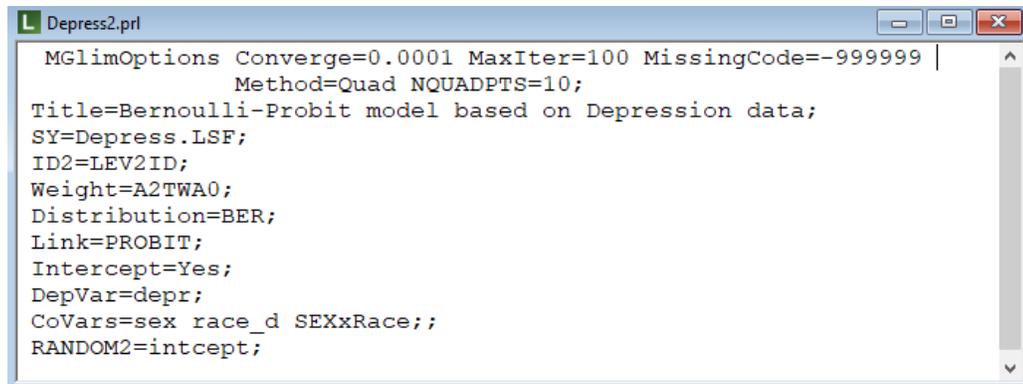
On the **Dependent and Independent Variables** dialog box screen, first select DEPR and click on the upper **Add** button to define it as the **Dependent variable**. Then select the variables SEX, RACE\_D and SEXxRace and click on the **Continuous** button to add them to the **Independent variables** list box as shown below.



Click on the **Next** button to proceed to the **Random Variables** dialog box. Keep the **Intercept** check box checked to include the random effect for the level-2 intercept.



Click 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 (**depress2.prl**) for the model specification file. The default folder to which the the syntax file will be save is the folder in which the data file is located.



```
Depress2.prl
MGLimOptions Converge=0.0001 MaxIter=100 MissingCode=-999999 |
Method=Quad NQUADPTS=10;
Title=Bernoulli-Probit model based on Depression data;
SY=Depress.LSF;
ID2=LEV2ID;
Weight=A2TWA0;
Distribution=BER;
Link=PROBIT;
Intercept=Yes;
DepVar=depr;
CoVars=sex race_d SEXxRace;;
RANDOM2=intcept;
```

### 3. The syntax file

Note that the following syntax lines are different from the previous model:

- Link = PROBIT defines the probit link function, which is defined in the **Distribution and Links** dialog box.
- CoVars = sex race\_d SEXxRace, now includes one more independent variable than was the case in the previous model.

Run the analysis by selecting the **Run PRELIS** button to generate the output file **depress2.out**, which is saved to the same folder as the syntax file.

### 4. Discussion of results

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

#### Model and data descriptions

The program info and the syntax are printed on the top of the output file. In the next section of the output file as shown below, descriptions of the distribution, the link function, the weight variable and the hierarchical structure of the data is provided. Note that probit link function is used as the link function.

```

Depress2.OUT
=====0
| Bernoulli-Probit model based on Depression data |
|
|-----0

                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        = A2TWA0
Number of Level-2 Units        = 10
Number of Level-1 Units        = 2214
Number of Level-1 Units per Level-2 Unit =
    62   598   34   126   416   148   363   141   246   80

```

**Descriptive statistics**

The data summary is followed by descriptive statistics for all the variables included in the model.

```

Depress2.OUT
=====0
| Descriptive statistics for all the variables in the model |
|-----0

Variable      Minimum      Maximum      Mean      Standard
-----      -
depr1         0.0000      1.0000      0.5890      0.4921
depr2         0.0000      1.0000      0.4110      0.4921
intcept       1.0000      1.0000      1.0000      0.0000
sex           0.0000      1.0000      0.2882      0.4530
race_d        0.0000      1.0000      0.3071      0.4614
SEXxRace      0.0000      1.0000      0.0980      0.2974

```

**Results for the model without any random effects**

The descriptive statistics is followed by the results for the model without any random effects. The deviance can be used for comparison tests of nested models. The estimated regression weights without any random effect are given below.

0=====0  
 | Results for the model without any random effects |  
 0=====0

Goodness of fit statistics

Statistic	Value	DF
Likelihood Ratio Chi-square	3325.8392	2210
Pearson Chi-square	2578.1501	2210

Estimated regression weights

Parameter	Estimate	Standard Error	z Value	P Value
intcept	-0.0857	0.0357	-2.4036	0.0162
sex	0.4175	0.0703	5.9361	0.0000
race_d	-0.3712	0.0762	-4.8706	0.0000
SEXxRace	0.0610	0.1396	0.4368	0.6623

### Results for the model with fixed and random effects

#### Number of iterations and fit statistics

As shown below, iteration was needed to obtain convergence. The likelihood ratio test, Akaike's and Schwarz's criteria are given after the iteration number. For detailed information about these statistics, please refer to the previous section.

0=====0  
 | Optimization Method: Adaptive Quadrature |  
 0=====0

Number of quadrature points =	10
Number of free parameters =	5
Number of iterations used =	1
-2lnL (deviance statistic) =	2892.01070
Akaike Information Criterion	2902.01070
Schwarz Criterion	2930.52348

#### Estimated regression weights

The output describing the estimated regression weights for the multilevel model is shown next. The estimates are shown in the column with heading Estimate and correspond to the coefficients  $\beta_0$ ,  $\beta_1$ ,  $\beta_2$  and  $\beta_3$  in the model specification. From the z-values and associated exceedance probabilities, we see that the intercept and the regression weight for SEXxRace are not significant at a 10% level of significance.

Parameter	Estimate	Standard Error	z Value	P Value
intcept	-0.0544	0.1186	-0.4582	0.6468
sex	0.4612	0.0723	6.3824	0.0000
race_d	-0.4207	0.0805	-5.2282	0.0000
SEXxRace	0.0805	0.1430	0.5628	0.5736

The estimated intercept has changed to -0.0544 in this second model. The estimated coefficient associated with gender (sex) is now 0.4612, which indicates that the female respondents (sex = 1) have a smaller  $\hat{\eta}$ . The estimate for the indicator of race (race\_d) shows that white clients have a higher  $\hat{\eta}$  value in the current model. The interaction term of gender and race is positive, but not significant. The probit link function is needed to transform these results into probabilities.

### Interpreting estimated regression weights by using link function

First, we substitute the regression weights and obtain the function for  $\hat{\eta}_{ij}$

$$\begin{aligned} \hat{\eta}_{ij} &= \hat{b}_{0i} + \hat{b}_{1i} \times (\text{sex})_{ij} + \hat{b}_{2i} \times (\text{race\_d})_{ij} + \hat{b}_{3i} \times (\text{SEXxRace})_{ij} \\ &= -0.0544 + 0.4612 \times (\text{sex})_{ij} - 0.4207 \times (\text{race\_d})_{ij} + 0.0805 \times (\text{SEXxRace})_{ij} \end{aligned}$$

For a black male, we have SEX = 0, RACE\_d = 0 and SEXxRace = 0 thus

$$\hat{\eta}_{ij} = -0.0544$$

Similarly, the calculation of  $\hat{\eta}_{ij}$  for a black female (SEX = 1, RACE\_d = 0 and SEXxRace = 0) is

$$\begin{aligned} \hat{\eta}_{ij} &= -0.0544 + 0.4612 \times 1 \\ &= 0.4068 \end{aligned}$$

Next, we transform the  $\hat{\eta}_{ij}$ 's into corresponding probabilities by using the probit link function. Taking black males as the example, the probability is calculated as shown below.

$$\text{Prob}(\text{DEPR}_{ij} = 1) = \Phi(-0.0544) = 47.83\%$$

Similarly, the probabilities of having depression for different gender and ethnicity groups are reported in the following table.

Group	Code	$\hat{\eta}$	Prob (DEPR = 1)
Black, male	sex = 0, race_d = 0	-0.0544	47.83%
Black, female	sex = 1, race_d = 0	0.4068	65.79%
White, male	sex = 0, race_d = 1	-0.4751	31.74%
White, female	sex = 1, race_d = 1	0.0666	52.65%

### Estimated level-2 variance and covariance

The output for the estimated level-2 variance and covariance is shown in the image below. In the previous model, we have discussed the interpretation of these in detail. In this model, the random part of the current model is simpler than in the previous model.

The screenshot shows a window titled "Depress2.OUT" with the following text:

```

Estimated level 2 variances and covariances

Parameter          Estimate      Standard      z Value      P Value
-----          -----          -----          -----          -----
intcept/intcept    0.1231         0.0625         1.9685       0.0490

Level 2 covariance matrix

intcept      intcept
intcept      0.123079

Calculation of the intraclass correlation
-----
residual variance = 1 (assumed)
cluster variance = 0.1231

intraclass correlation = 0.1231 / ( 0.1231 + 1.000 ) = 0.110

```