Bernoulli distribution with complementary log-log link function

Contents

1. The data

The data set forms part of the data library of the Alcohol and Drug Services Study (ADSS). The ADSS is a national study of substance abuse treatment facilities and clients. Background data and data on the substance abuse of a sample of 1752 clients were obtained. The sample was stratified by census region (CENREG) and within each stratum a sample was obtained for each of three facility treatment types (FACTYPE) within each of the four census regions. More information on the ADSS and the data are available at http://webapp.icpsr.umich.edu/cocoon/ICPSR-STUDY/03088.xml.

The specific data set is provided in the **Multilevel Generalized Linear Model Examples** folder as the LSF file **Depress.LSF**. The first portion of this file is shown in the following LSF window.

Depress.lsf

	depr	sex	race_d	SEXxRace	A2TWA0	LEV2ID
1	1.00	0.00	1.00	0.00	316.80	2.00
2	1.00	0.00	0.00	0.00	276.60	2.00
3	1.00	1.00	0.00	0.00	276.60	2.00
4	1.00	0.00	0.00	0.00	276.60	2.00
5	1.00	1.00	0.00	0.00	276.60	2.00
6	0.00	0.00	0.00	0.00	276.60	2.00
7	1.00	1.00	0.00	0.00	276.60	2.00
8	1.00	1.00	0.00	0.00	276.60	2.00
9	1.00	0.00	0.00	0.00	276.60	2.00
10	1.00	0.00	0.00	0.00	276.60	2.00

The variables of interest are:

- DEPR addresses the question of whether the patient has depression. (1 = Yes; 0 = No)
- A2TWA0 is the sampling weight
- SEX is a dummy variable indicating the gender (0 for male and 1 for female) of the client
- RACE_D is a dummy variable representing the ethnicity (0 for black and 1 for white) of the client
- SEXxRACE is the interaction term of gender and race.
- LEV2ID is the variable used to identify the level-2 ID or grouping variable.

1. The model

In a previous model, the probit link function is used. We now fit the same model by using a complementary log-log link function.

The complementary log-log link function is defined as

$$1\text{-exp}\left(-\exp\left(\eta_{ij}\right)\right)$$

The level-1 and level-2 models are unchanged.

Level-1 model:

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

Level-2 model:

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

2. Edit the existing syntax file

To obtain the model discussed above, we can either go through the multilevel generalized linear model dialog boxes as we did in the previous models or modify the existing syntax file directly. Here, we will illustrate how to modify the syntax file generated in the previous example.

First, open the syntax file for the previous model and save it under a different name so as not to overwrite the syntax file associated with the previous analysis. To do so, click on the **File, Open** option on the LISREL main window. Keep the **Syntax Only (*.spl, *.lis, *.prl)** dropdown list unchanged. Browse for the **Multilevel Generalized Linear Model Examples** folder, in which we saved the syntax file (**depress2.prl**) of the previous model. Double click on the **depress2.prl** to open it. Select the **File**, **Save As** option, and provide a new name (**depress3.prl**) for the model specification file.

As a first step, give an appropriate title by changing the Title line. Next, change the probit link function to a complementary log-log link function by modifying the syntax line Link=PROBIT to Link = CLL to produce the following syntax file. The CLL keyword refers to the complementary log-log link.

Run the analysis by selecting the **Run PRELIS** button to generate the output file **depress3.out**, which is saved to the same folder as the syntax file.

3. Discussion of results

Portions of the output file **depress3.out** are shown below.

Model and data description

The model and data descriptions clearly show that complementary log-log (CLL) link function is used in the model.

```
Depress3.OUT
              Model and Data Descriptions

     Sampling Distribution                     = Bernoulli
     Link Function                             = Complementary Log-Log (CLL)
     PROB(Success)= 1.0-EXP[-EXP(ETA)]

     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
```

Results for the model with fixed and random effects

Number of iterations and fit statistics

Six iterations were needed to obtain convergence. The likelihood ratio test, Pearson chi-square, and Akaike's and Schwarz's criteria are given after to the iteration number.

```
Depress3.OUT

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

   Number of quadrature points =             10
   Number of free parameters =                5
   Number of iterations used =                2

   -2lnL (deviance statistic) =      2891.16332
   Akaike Information Criterion      2901.16332
   Schwarz Criterion                 2929.67610
```

Estimated regression weights

The output describing the estimated regression weights is shown next. The estimates are shown in the column with heading Estimate and correspond to the coefficients β_0, β_1, β_2 and β_3 in the model specification. From the z-values and associated exceedance probabilities, we see that the intercept and the regression weight for SEXxRace is not significant at a 10% level of significance.

```
Depress3.OUT

                        Estimated regression weights

                                    Standard
     Parameter      Estimate          Error      z Value      P Value
     ---------      --------       --------      -------      -------
     intcept         -0.4403         0.1397      -3.1525       0.0016
     sex              0.5005         0.0771       6.4902       0.0000
     race_d          -0.5452         0.1048      -5.2029       0.0000
     SEXxRace         0.1646         0.1678       0.9808       0.3267


                Estimated level 2 variances and covariances

                                             Standard
     Parameter               Estimate           Error      z Value      P Value
     ---------               --------        --------      -------      -------
     intcept/intcept           0.1714          0.0852       2.0124       0.0442


   | Calculation of the intracluster correlation
   -------------------------------------------
   residual variance = pi*pi / 6 (assumed)
   cluster  variance =     0.1714

   intracluster correlation =     0.1714 / (     0.1714 + (pi*pi/6)) = 0.094
```

The estimated coefficient for the intercept is negative in value in this model. The estimated coefficient associated with gender (SEX) is – 0.5508. The estimate for the ethnicity indicator (RACE_D) also shows that white clients are likely to have a higher $\hat{\eta}$ value. The coefficient representing the iinteraction of gender and ethnicity is also negative and implies a decrease in $\hat{\eta}$. To transform these results into probabilities, we use the complementary log-log link function.

Interpreting estimated regression weights by using the link function

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

$$\begin{aligned}\hat{\eta}_{ij} &= \hat{b}_{0i} + \hat{b}_{1i} \times (\text{SEX})_{ij} + \hat{b}_{2i} \times (\text{RACE\_d})_{ij} + \hat{b}_{3i} \times (\text{SEXxRACE})_{ij} \\ &= -0.4403 + 0.5005 \times (\text{SEX})_{ij} - 0.5452 \times (\text{RACE\_d})_{ij} + 0.1646 \times (\text{SEXxRACE})_{ij}\end{aligned}.$$

For a black males, we have SEX = 0, RACE_d = 0 and SEXxRace = 0 thus

$$\hat{\eta}_{ij} = -0.4403.$$

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

$$\begin{aligned}\hat{\eta}_{ij} &= -0.4403 + 0.5005 \\ &= 0.0602\end{aligned}.$$

Next, we transform the $\hat{\eta}_{ij}$'s into the corresponding probabilities by using the complementary log-log link function. Taking the black males as an example, we calculate their probability of being depressed as

$$\text{Prob}(\widehat{DEPR}_{ij} = 1) = 1\text{-exp}\left(-\exp\left(\eta_{ij}\right)\right) = 61.18\%$$

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	-0.4403	47.47%
Black, female	sex = 1, race_d = 0	0.0602	65.43%
White, male	sex = 0, race_d = 1	-0.9855	31.15%
White, female	sex = 1, race_d = 1	-0.3204	51.61%

In all three binary models, even though the estimated $\hat{\eta}$'s are different, the estimated probabilities are close to each other in all groups.