Negative binomial model for the 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 2339 dysthymia respondents. After list-wise deletion, the sample size is 1981.

nesarc_poi.LSF

	PSU	FINWT	CONCENTR	AGE_ONS	N_DEP
1	1011.00	7256.15	0.00	51.00	1.00
2	1011.00	3476.67	1.00	48.00	1.00
3	1011.00	3052.10	1.00	59.00	1.00
4	1011.00	1182.03	1.00	36.00	2.00
5	1011.00	3041.05	1.00	17.00	1.00
6	1011.00	8342.94	0.00	16.00	1.00
7	1011.00	6767.06	1.00	29.00	1.00
8	1011.00	3460.29	1.00	43.00	1.00
9	1015.00	3167.29	1.00	55.00	1.00
10	1018.00	1014.56	0.00	37.00	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.
- CONCENTR contains the information captured in field S4CQ3A6 of the NESARC data. It represents the response to the statement "Often had trouble concentrating/keeping mind on things," with 1 indicating "Yes," and 0 indicating "No."
- AGE_ONS is based on field S4CQ7AR of the NESARC data. It represents the age at onset of first episode.
- N_DEP is recoded from field S4CQ6A of the NESARC data and gives the number of depression/dysthymia episodes. This is the count variable we would like to use as outcome variable in the examples to follow.

1. The model

In the previous section a Poisson model was fitted to the data. It was also noted that a Poisson distribution has an important property: the mean number of occurrences is equal to the variance. The negative binomial distribution can be used as an alternative to the Poisson distribution. It is especially useful for discrete data that assumes values 0, 1, 2, 3… whose sample variance exceeds the sample mean. In such cases, the observations are over-dispersed with respect to a Poisson distribution, for which the mean is equal to the variance. Since the negative binomial distribution has one more parameter than the Poisson, the second parameter can be used to adjust the variance independently of the mean. It can be shown that a model based on the negative binomial distribution with a dispersion parameter close to zero will produce results that correspond closely to those obtained for the Poisson model. In this section, we fit a negative binomial model, utilizing the same predictors and a small dispersion parameter, to the NESARC data. Subsequently larger values of the dispersion parameter will be used to study the impact on parameter estimates and the deviance statistic. Again, adaptive quadrature is used as the method of optimization.

The negative binomial distribution can be expressed as

$$f\left(y_i\right)=\frac{\Gamma\left(y_i+1/\alpha\right)}{\Gamma\left(y_i+1\right)\Gamma\left(1/\alpha\right)}\times\frac{\left(\alpha\mu_i\right)^{y_i}}{\left(1+\alpha\mu_i\right)^{y_i+1/\alpha}}$$

with $\sigma^2\left(y_i\right)=\mu_i+\alpha\mu_i^2$ where α denotes an additional parameter and it can no longer be assumed that the variance is a known function of the mean. We assume α to be a fixed parameter.

The model fitted to the data explores the relationship between N_DEP and the variables indicating concentration (or lack thereof) and age, as represented by the variables CONCENTR and AGE_ONS.

The level-1 model is

$$\log\left[\mathrm{E}\left(\mathrm{N\_DEP}_{ij}\right)\right]=\beta_0+\beta_1\times\mathrm{CONC\_DEP}_{ij}+\beta_2\times\mathrm{AGE\_DEP}_{ij}$$

The level-2 model is

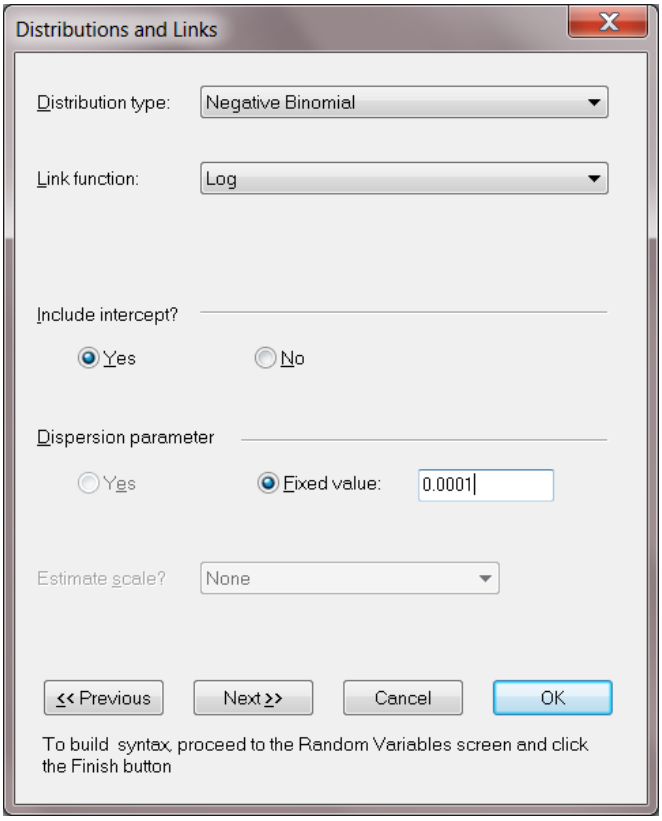
$$\beta_0=b_{00}+v_{i0},\ \beta_1=b_{10}$$

and

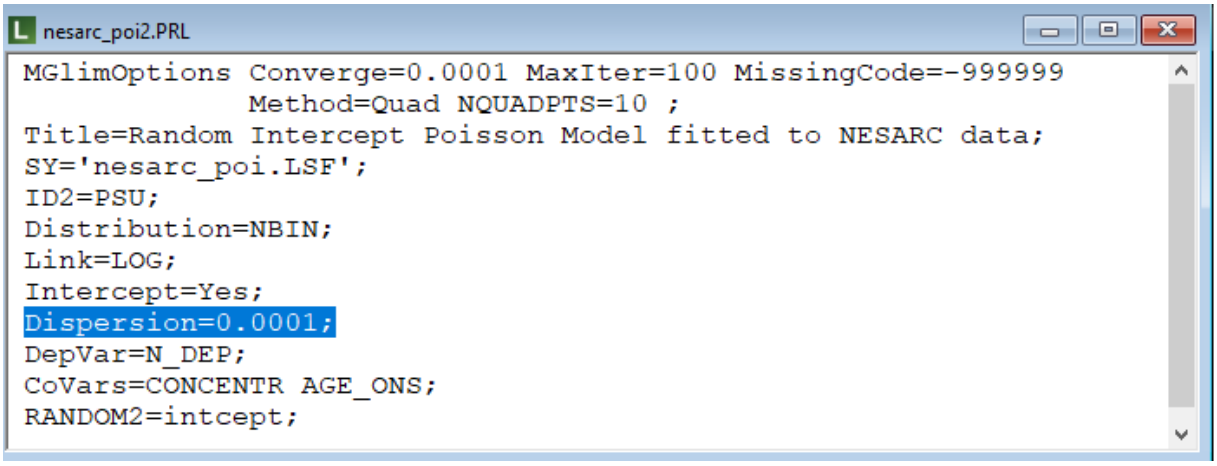
$$\beta_2 = b_{20}$$

2. Setting up the analysis

Make sure that the spreadsheet **nesarc_poi.lsf** is the active window and keeping all the other settings for a **Multilevel**, **Generalized Linear Model** unchanged, set the **Distribution Model** to **negative binomial**, and the **Dispersion Parameter** to 0.0001 as shown below.



Save the revised syntax file as **nesarc_poi2.prl**, and click the **PRELIS**, **Run** icon button to start the iterative process.

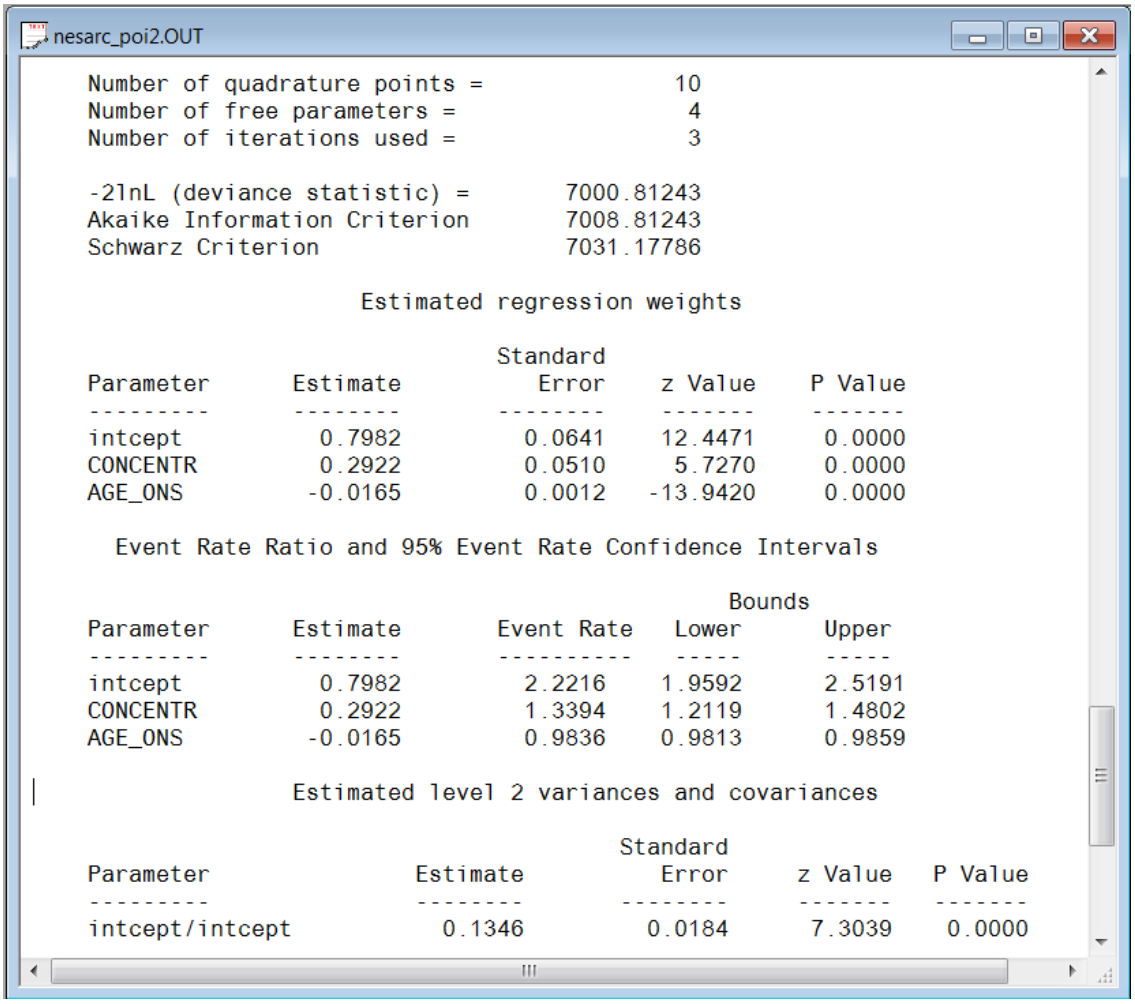


3. Discussion of results

Portions of the output file **nesarc_poi2.out** are shown below.

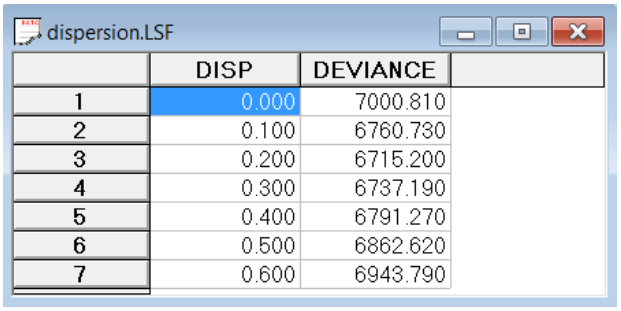
Fixed and random effect results

The estimated regression coefficients for fixed effects in the model are shown below. Recall that the estimated coefficients of the intercept, CONCENTR, and AGE_ONS under the Poisson model reported in the previous section were 0.7982, 0.2922, and -0.0165 respectively. The estimated variation in the average estimated N_DEP at level-2 was 0.1347, and highly significant. Note that with the dispersion parameter set at 0.0001, results of the two models are almost identical.



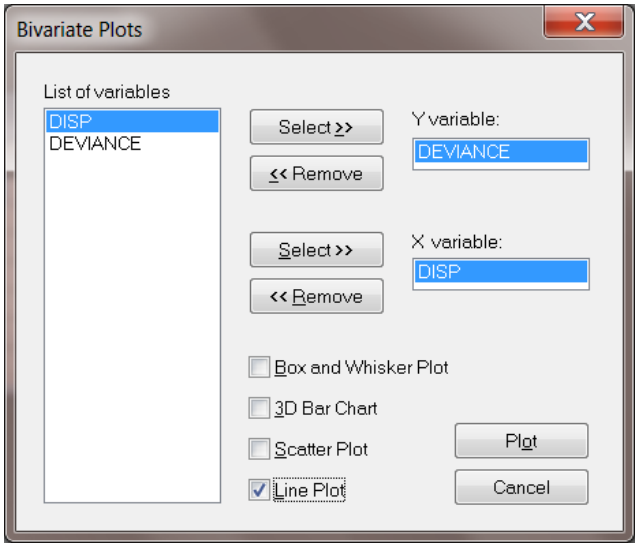
4. Changing the value of the dispersion parameter

Save the syntax file for the negative binomial model as **nesarc_poi3.prl** and change the dispersion value to 0.1 before running the analysis and making a note of the deviance value (6760.73 in this case). Repeat the procedure for dispersion values of 0.2, 0.3, 0.4, 0.5 and 0.6 respectively. The list of (dispersion; deviance)-values are shown below in the LISREL data file **dispersion.lsf**.

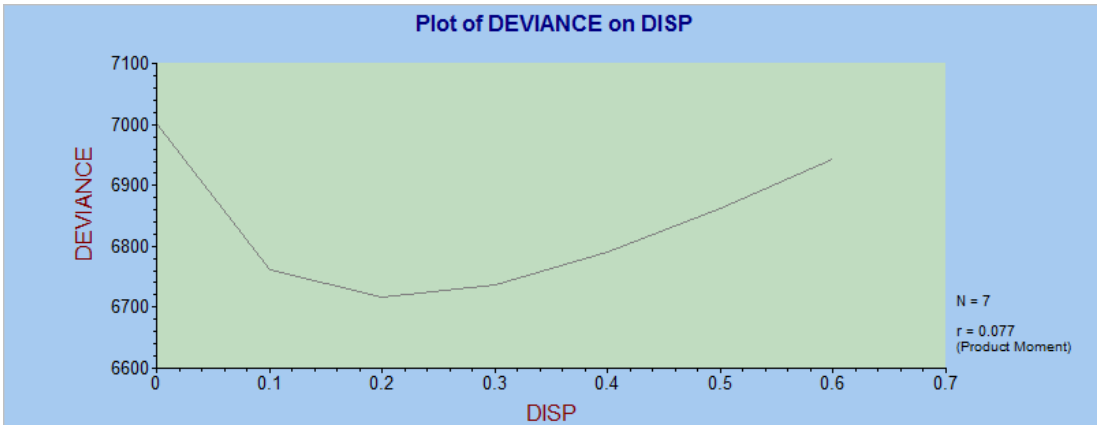


	DISP	DEVIANCE
1	0.000	7000.810
2	0.100	6760.730
3	0.200	6715.200
4	0.300	6737.190
5	0.400	6791.270
6	0.500	6862.620
7	0.600	6943.790

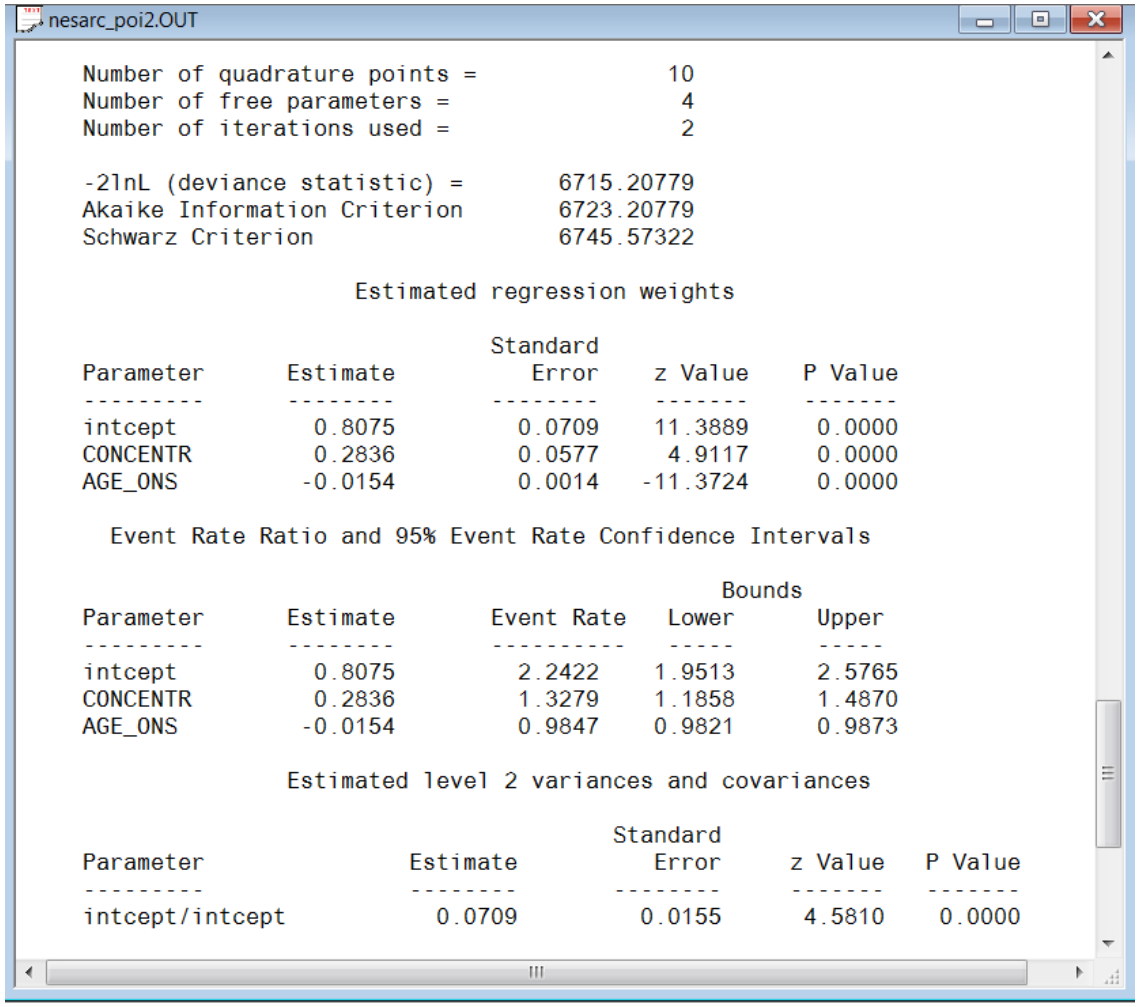
From the main menu-bar select the **Graphs, Bivariate Plots** option to request a line plot of the deviance-statistic against dispersion.



The plot shown below, shows that the best model fit is obtained for a dispersion-value of 0.2 (best fit corresponds to smallest deviance value).



The output for the fixed and random effects is listed below. All the parameter estimates are significant, but different from those reported for the Poisson model.


nesarc_poi2.OUT
Number of quadrature points = 10
Number of free parameters = 4
Number of iterations used = 2
-2lnL (deviance statistic) = 6715.20779
Akaike Information Criterion 6723.20779
Schwarz Criterion 6745.57322
Estimated regression weights
Parameter Estimate Standard Error z Value P Value
intcept 0.8075 0.0709 11.3889 0.0000
CONCENTR 0.2836 0.0577 4.9117 0.0000
AGE_ONS -0.0154 0.0014 -11.3724 0.0000
Event Rate Ratio and 95% Event Rate Confidence Intervals
Parameter Estimate Event Rate Bounds Lower Upper
intcept 0.8075 2.2422 1.9513 2.5765
CONCENTR 0.2836 1.3279 1.1858 1.4870
AGE_ONS -0.0154 0.9847 0.9821 0.9873
Estimated level 2 variances and covariances
Parameter Estimate Standard Error z Value P Value
intcept/intcept 0.0709 0.0155 4.5810 0.0000