



Two-level weighted Poisson model

3.1.1 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 file has been used in some of the examples in Chapter 3, and contains information on the occurrences of major depression, family history of major depression and dysthymia of 2339 dysthymia respondents. After listwise deletion, the sample size is 1981.

	(A) PSU	(B) FINWT	(C) CONC	(D) AGE	(E) N	DEP
1	1011	7256.15	0	51		1
2	1011	3476.67	1	48		1
3	1011	3052.10	1	59		1
4	1011	1182.03	1	36		2
5	1011	3041.05	1	17		1
6	1011	8342.94	0	16		1
7	1011	6767.06	1	29		1
8	1011	3460.29	1	43		1
9	1015	3167.29	1	55		1
10	1018	1014.56	0	37		1

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.
- CONC_DEP 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_DEP 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.

3.1.1.1 Exploring the data

Inspecting the distribution of the intended outcome variable, N_DEP, before starting with the model is important. In the case of a count variable, this can easily be done by producing a bar chart of the observed frequencies of occurrence captured by the count variable. Select the **File, Data-based Graph, Univariate** option from the main SuperMix window and request a bar chart before clicking the **Plot** button.

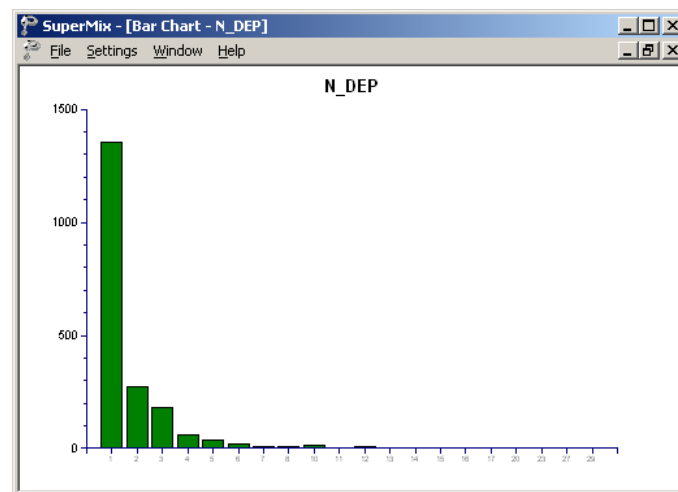


Figure 5.3: Bar chart for count variable N_DEP

The frequency bar chart for the count variable N_DEP shown in Figure 5.3 is obtained. We note that the number of depression episode ranges from 1 to 29, with most respondents having a small number of reported episodes of depression.

3.1.2 Weighted 2-level models

3.1.2.1 The data

The sampling frame of many multistage surveys frequently entails selection of units with known, but unequal, selection probabilities. This situation is the result of a number of design factors, of which the cost of doing the survey is an important consideration. When this is the case, it is appropriate to weight observations in order to produce unbiased estimates of population parameters.

The data also included a weight variable. The variable 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. In this section, we explore the effect of inclusion of the weights on the results obtained in other sections.

3.1.2.2 Setting up the analysis

The models remain the same, with only the selection of the weight variable on the **Advanced** tab of the **Model Specification** screen to be added. Below, we show how this is done in the case of the Poisson distribution model.

Open the model specification file for the Poisson distribution model (**nesarc_poi1w.mum**) and click on the **Advanced** tab. Change the **Unit Weighting** field from its default value of equal to differential. Next, select the variable FINWT from the **Assigned Weight** drop-down list box that appears when the selection has been made in the **Unit Weighting** field. The completed **Advanced** tab is shown below.

The screenshot shows the 'Model Setup: nesarc_poi1w.mum' dialog box with the 'Advanced' tab selected. The dialog has several tabs: Configuration, Variables, Starting Values, Patterns, Advanced, and Linear Transforms. The 'Advanced' tab contains the following settings:

- General Settings:**
 - Unit Weighting: differential (dropdown)
 - Level-1 Weight: FINWT (dropdown)
 - Level-2 Weight: (empty dropdown)
 - Optimization Method: adaptive quadrature (dropdown)
 - Number of Quadrature Points: 10 (text input)
- Time Settings:**
 - Incorporate Time Offset: no (dropdown)
- Dependent (Count) Variable Settings:**
 - Distribution Model: Poisson (dropdown)
 - Estimate Scale: none (dropdown)

At the bottom of the dialog, there is a note: 'Use the arrow keys or click on the desired tab to select the category of interest for the model.'

Save the specification file as **nesarc_poi1w.mum**, and run the analysis.

3.1.2.3 Discussion of results

Results for this analysis are reported in Table 5.2 below. The results from the unweighted Poisson distribution model are included in order to facilitate evaluation of the impact of the weights on the results.

Table 5.2: Comparison of results for weighted and unweighted Poisson models

Parameter	Unweighted model		Weighted model	
	Estimate	Standard error	Estimate	Standard error
intcept	0.7982	0.0641	0.7229	0.0659
CONC_DEP	0.2922	0.0510	0.3055	0.0532
AGE_DEP	−0.0165	0.0012	−0.0156	0.0013
Level-2 variance	0.1347	0.0184	0.1373	0.0189

Results for the two models are very similar, and interpretation of the results of both models would lead to the same conclusions, both in terms of significance and in terms of the expected number of depression episodes. However, this is more the exception than the rule – users are cautioned to use weight variables whenever they are available in order to prevent skewed or biased results that may occur when weights are excluded in the analysis of a disproportionally drawn sample.