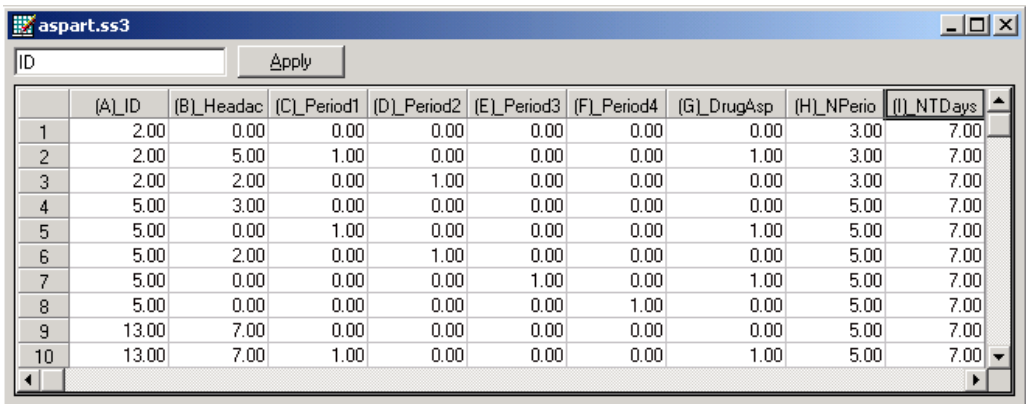
Contents

Two-level models for count outcomes *1*
The data 1
The model 2
Example: Poisson regression with a random intercept 3
Example: Mixed-effects analysis with an offset variable 8

Two-level models for count outcomes

The data

The data for this example are taken from a paper by McKnight and Van Den Eeden (1993), who reported on the number of headaches in a two treatment, multiple period crossover trial. Specifically, the number of headaches per week was repeatedly measured for 27 patients. Following a seven-day placebo run-in period, subjects received either aspartame or placebo in four seven-day treatment periods according to the double-blind crossover treatment design. Each treatment period was separated by a washout day. The sample size is 122. Data for the first 10 observations of all the variables used in this section are shown below in the form of a SuperMix spreadsheet window for **aspart.ss3**.


aspart.ss3

ID | Apply

	(A)_ID	(B)_Headac	(C)_Period1	(D)_Period2	(E)_Period3	(F)_Period4	(G)_DrugAsp	(H)_NPerio	(I)_NTDays
1	2.00	0.00	0.00	0.00	0.00	0.00	0.00	3.00	7.00
2	2.00	5.00	1.00	0.00	0.00	0.00	1.00	3.00	7.00
3	2.00	2.00	0.00	1.00	0.00	0.00	0.00	3.00	7.00
4	5.00	3.00	0.00	0.00	0.00	0.00	0.00	5.00	7.00
5	5.00	0.00	1.00	0.00	0.00	0.00	1.00	5.00	7.00
6	5.00	2.00	0.00	1.00	0.00	0.00	0.00	5.00	7.00
7	5.00	0.00	0.00	0.00	1.00	0.00	1.00	5.00	7.00
8	5.00	0.00	0.00	0.00	0.00	1.00	0.00	5.00	7.00
9	13.00	7.00	0.00	0.00	0.00	0.00	0.00	5.00	7.00
10	13.00	7.00	1.00	0.00	0.00	0.00	1.00	5.00	7.00

The variables of interest are:

- ID is the patient ID (27 patients in total).
- HeadAche is the number of headaches during the week (from 0 to 7).
- Period1 is a period 1 treatment indicator (1 for the first treatment period and 0 otherwise).
- Period2 is a period 2 treatment indicator (1 for the second treatment period and 0 otherwise).
- Period3 is a period 3 treatment indicator (1 for the third treatment period and 0 otherwise).
- Period4 is a period 4 treatment indicator (1 for the fourth treatment period and 0 otherwise).
- DrugAsp indicates the type of drug being used for the treatment, (0 = placebo and 1 = aspartame). 75 observations used placebo and 47 used aspartame.
- Nperiods is the number of periods the individual was observed (from 2 to 5).
- NTDays is the number of treatment days in the period (from 1 to 7).

The model

A general two-level Poisson regression model for a count response variable y depending on a set of r predictors $x_1, x_1, \ldots, x_r$ may be expressed as

$$\ln\left(\mu_{ij}\right) = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{u}_i$$

where y_{ij} denotes the value of y for the j-th level-1 unit nested within the i-th level-2 unit for $i = 1, 2, \cdots, N$ and $j = 1, 2, \ldots, n_i$, the scalar product $\mathbf{x}'_{ij}\boldsymbol{\beta}$ is the fixed part of the model, and $\mathbf{z}'_{ij}\mathbf{u}_i$ denotes the random part of the model at level 2. For the fixed part of the model, $\mathbf{x}'_{ij}$ is a typical row of the design matrix while the vector $\boldsymbol{\beta}$ contains the fixed, but unknown parameters to be estimated. In the case of the random part of the model at level 2, $\mathbf{z}'_{ij}$ represents a typical row of the design matrix, and $\mathbf{u}_i$ the vector of random level-2 effects to be estimated.

The specific Poisson regression model with a random intercept for the number of headaches may be expressed as

$$\hat{\mu}_{ij} = \exp\left(\beta_0 + \beta_1 * \text{Period1}_{ij} + \beta_2 * \text{Period2}_{ij} + \beta_3 * \text{Period3}_{ij} + \beta_4 * \text{Period4}_{ij} + \beta_5 * \text{DrugAsp}_{ij} + u_{i0}\right)$$

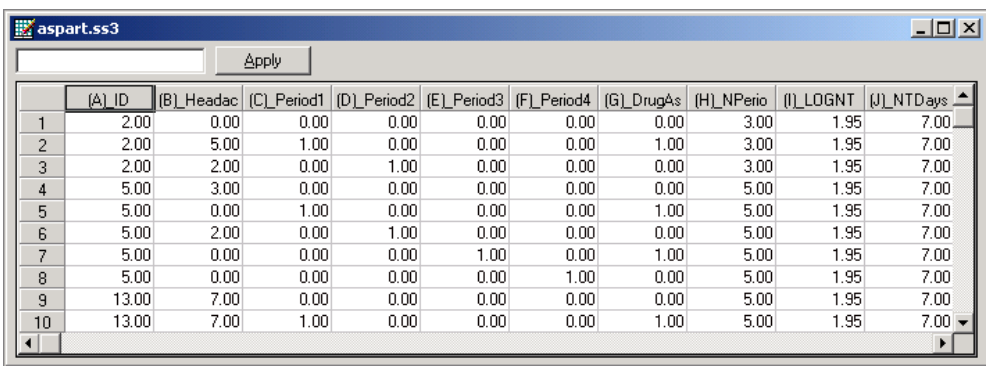
where $\hat{\mu}_{ij}$ denotes the mean number of headaches of patient i for treatment period j; Period2_{ij}, Period2_{ij}, Period3_{ij} and Period4_{ij} denote the values of the dummy variables Period1, Period2, Period3 and Period4 for patient i for treatment period j respectively; DrugAsp_{ij} denotes the value of the DrugAsp for patient i for treatment period j; β_0, β_1, β_2, β_3, β_4 and β_5 denote unknown parameters; and u_{i0} denotes the random intercept for patient i for $i = 1, 2, \ldots, 27$ and $j = 1, 2, 3, 4$. This model is fitted to the data in **aspart.ss3** as follows.

Example: Poisson regression with a random intercept

Importing the data

The first step is to create the **ss3** file, **aspart.ss3**, from the Excel workbook **aspart.xls**. This is accomplished as follows:

- Use the **Import Data File** option on the **File** menu to open the **Open** dialog box.
- Browse for the file **aspart.xls** in the **Examples\Primer\Count** folder.
- Select the file and click on the **Open** button to open the following SuperMix spreadsheet window for **aspart.ss3**.



	(A)_ID	(B)_Headac	(C)_Period1	(D)_Period2	(E)_Period3	(F)_Period4	(G)_DrugAs	(H)_NPerio	(I)_LOGNT	(J)_NTDays
1	2.00	0.00	0.00	0.00	0.00	0.00	0.00	3.00	1.95	7.00
2	2.00	5.00	1.00	0.00	0.00	0.00	1.00	3.00	1.95	7.00
3	2.00	2.00	0.00	1.00	0.00	0.00	0.00	3.00	1.95	7.00
4	5.00	3.00	0.00	0.00	0.00	0.00	0.00	5.00	1.95	7.00
5	5.00	0.00	1.00	0.00	0.00	0.00	1.00	5.00	1.95	7.00
6	5.00	2.00	0.00	1.00	0.00	0.00	0.00	5.00	1.95	7.00
7	5.00	0.00	0.00	0.00	1.00	0.00	1.00	5.00	1.95	7.00
8	5.00	0.00	0.00	0.00	0.00	1.00	0.00	5.00	1.95	7.00
9	13.00	7.00	0.00	0.00	0.00	0.00	0.00	5.00	1.95	7.00
10	13.00	7.00	1.00	0.00	0.00	0.00	1.00	5.00	1.95	7.00

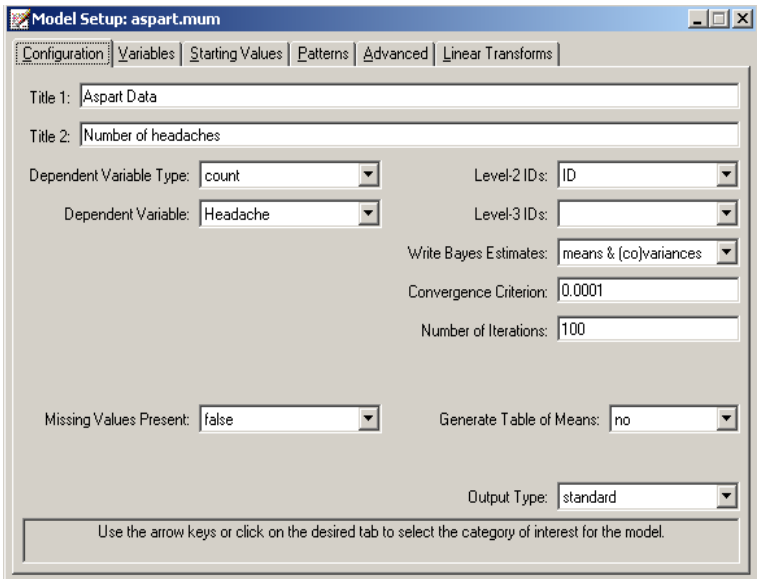
After selecting the **File**, **Save** option from the main menu bar, we are ready to fit the Poisson regression model with a random intercept for the number of headaches to the data in **aspart.ss3**.

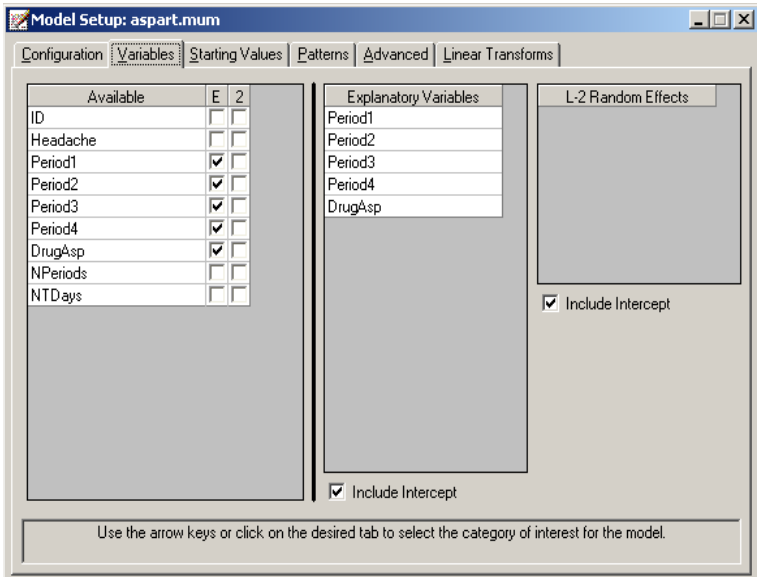
Setting up the analysis

Start by selecting the **New Model Setup** option on the **File** menu to open the **Model Setup** window.

On the **Configuration** screen, we first enter the titles Aspartame Data – Repeated Headaches across Time and random intercept and 5 covariates for the analysis in the **Title 1** and **Title 2** text

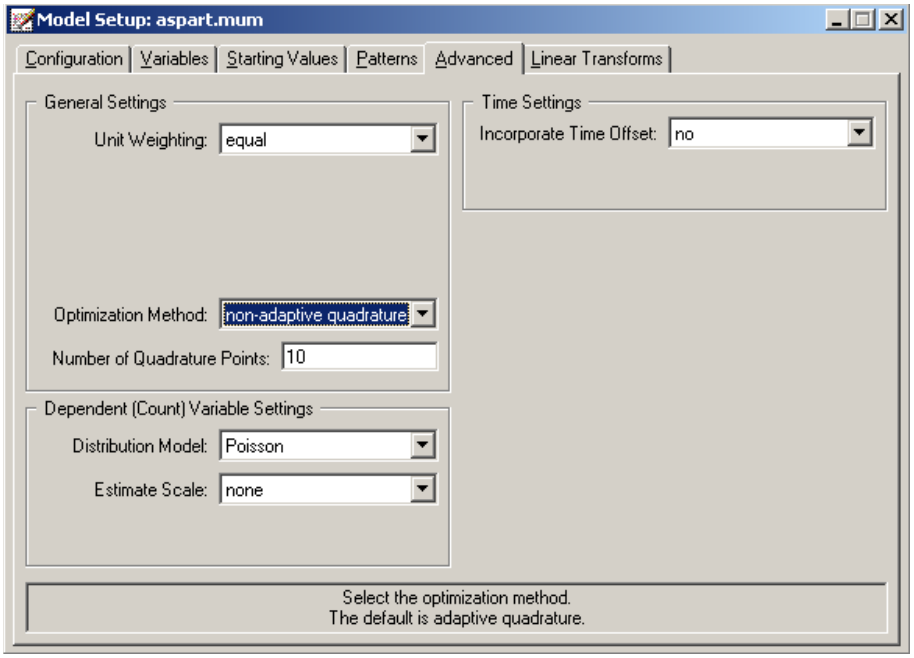
boxes respectively. The count outcome variable HeadAche is selected from the **Dependent Variable** drop-down list box. The **Dependent Variable Type** drop-down list box is used to indicate that the outcome variable is a **count**. The variable ID, which defines the levels of the hierarchy, is selected as the Level-2 ID from the **Level-2 IDs** drop-down list box.





Next, click on the **Variables** tab to proceed with variable selection. The variables Period1, Period2, Period3, Period4, and DrugAsp are specified as the fixed effects of the model by checking the **E** check boxes for Period1, Period2, Period3, Period4, and DrugAsp in the **Available** grid. These actions produce the **Variables** screen as shown above.

Finally, we enter the number of quadrature points, in this case 20, in the **Number of Quadrature Points** text box on the **Advanced** screen as shown below. Also, change the Optimization Method to non-adaptive quadrature.



Before we can run the analysis, we have to save the model specifications to file. This is accomplished by using the **Save** option on the **File** menu to open a **Save Mixed Up Model** dialog box. First enter the name **aspart.mum** in the **File name** text box and then click on the **Save** button to save the file. The analysis is run by selecting the **Run** option from the **Analysis** menu. This produces the corresponding output file **aspart.out**.

Discussion of results

Portions of this output file are shown below.

```
o=====================o
| Aspart Data         |
| Number of headaches |
o=====================o


            Model and Data Descriptions

   Sampling Distribution                    = Poisson
   Link Function                            = Log
   Number of Level-2 Units                    27
   Number of Level-1 Units                    122
   Number of Level-1 Units per Level-2 Unit =
     3     5     5     3     5     5     5     5     5     5     4     2
     5     5     5     5     5     5     5     5     5     5     5     5
     5     3     2
```

The above output file indicates that there are 27 subjects with 122 observations nested within them. The number of observations per subject varies between 2 and 5.

The descriptive statistics for all the variables in the model are shown next.

```
o==========================================================o
| Descriptive statistics for all the variables in the model |
o==========================================================o
                                                  Standard
   Variable      Minimum      Maximum        Mean   Deviation
   --------      -------      -------        ----   ---------
   Headache       0.0000       7.0000      1.6803      1.8863
   intcept        1.0000       1.0000      1.0000      0.0000
   Period1        0.0000       1.0000      0.2213      0.4168
   Period2        0.0000       1.0000      0.2049      0.4053
   Period3        0.0000       1.0000      0.1803      0.3860
   Period4        0.0000       1.0000      0.1721      0.3791
   DrugAsp        0.0000       1.0000      0.3852      0.4887
```

Following the descriptive statistics, the results for the model without any random effect is given, as shown below.

SuperMix - [aspart.out]

File Analysis Window Help

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

                    Goodness of fit statistics

  Statistic                                  Value          DF          Ratio
  ---------                                  -----          --          -----
  Likelihood Ratio Chi-square              243.8257         116         2.1019
  Pearson Chi-square                       253.8934         116         2.1887


                    Estimated regression weights

                                  Standard
  Parameter       Estimate          Error     z Value    P Value
  ---------       --------        --------    -------    -------
  intcept           0.4654         0.1525     3.0516     0.0023
  Period1           0.0916         0.2265     0.4043     0.6860
  Period2           0.0131         0.2276     0.0575     0.9542
  Period3          -0.2245         0.2471    -0.9084     0.3637
  Period4          -0.1840         0.2540    -0.7242     0.4689
  DrugAsp           0.2332         0.1596     1.4612     0.1440


    Event Rate Ratio and 95% Event Rate Confidence Intervals

                                                Bounds
  Parameter       Estimate        Event Rate   Lower      Upper
  ---------       --------        ----------   -----      -----
  intcept           0.4654          1.5926    1.1811     2.1474
  Period1           0.0916          1.0959    0.7030     1.7085
  Period2           0.0131          1.0132    0.6486     1.5827
  Period3          -0.2245          0.7989    0.4923     1.2967
  Period4          -0.1840          0.8320    0.5057     1.3688
  DrugAsp           0.2332          1.2626    0.9235     1.7263
```

Save As... Close

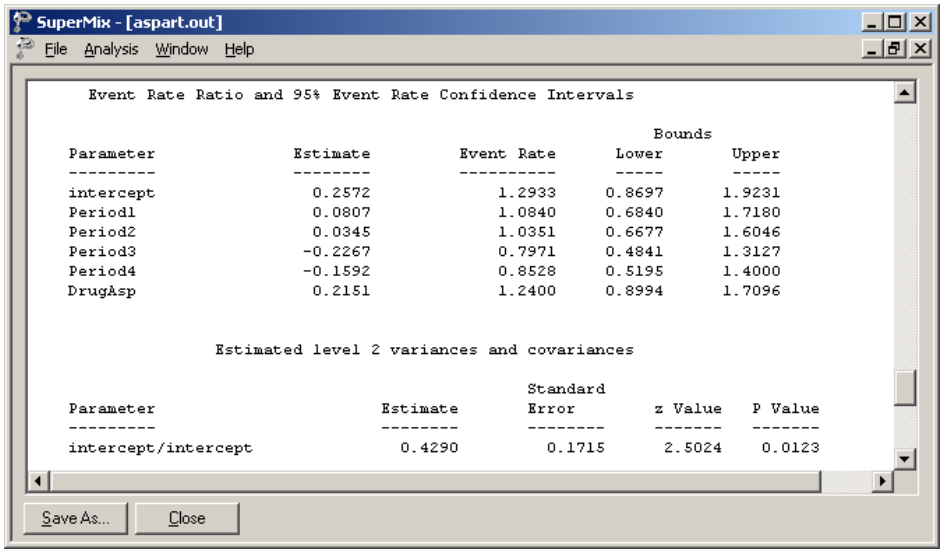
Next, the final results using non-adaptive quadrature optimization method is given. As shown below, the model converged after 4 iterations. The fixed part of the estimates is given. The *p*-values for all these estimates are not significant at 10% level.

SuperMix - [aspart.out]

File Analysis Window Help

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

Number of quadrature points =              20
Number of free parameters =                 7
Number of iterations used =                 4

-2lnL (deviance statistic) =        406.34905
Akaike Information Criterion        420.34905
Schwarz Criterion                   439.97720


                    Estimated regression weights

                                        Standard
Parameter              Estimate         Error       z Value     P Value
---------              --------         --------    -------     -------
intercept                0.2572           0.2024     1.2705      0.2039
Period1                  0.0807           0.2349     0.3434      0.7313
Period2                  0.0345           0.2237     0.1542      0.8775
Period3                 -0.2267           0.2545    -0.8909      0.3730
Period4                 -0.1592           0.2529    -0.6295      0.5290
DrugAsp                  0.2151           0.1638     1.3129      0.1892
```

Save As... Close



SuperMix - [aspart.out]

File Analysis Window Help

Event Rate Ratio and 95% Event Rate Confidence Intervals

Parameter	Estimate	Event Rate	Bounds Lower	Bounds Upper
intercept	0.2572	1.2933	0.8697	1.9231
Period1	0.0807	1.0840	0.6840	1.7180
Period2	0.0345	1.0351	0.6677	1.6046
Period3	-0.2267	0.7971	0.4841	1.3127
Period4	-0.1592	0.8528	0.5195	1.4000
DrugAsp	0.2151	1.2400	0.8994	1.7096

Estimated level 2 variances and covariances

Parameter	Estimate	Standard Error	z Value	P Value
intercept/intercept	0.4290	0.1715	2.5024	0.0123

Save As... Close

The random-effect variance is estimated as 0.429, with a *p*-value of 0.0123. All the regression coefficients in the fixed part of the model are non-significant.

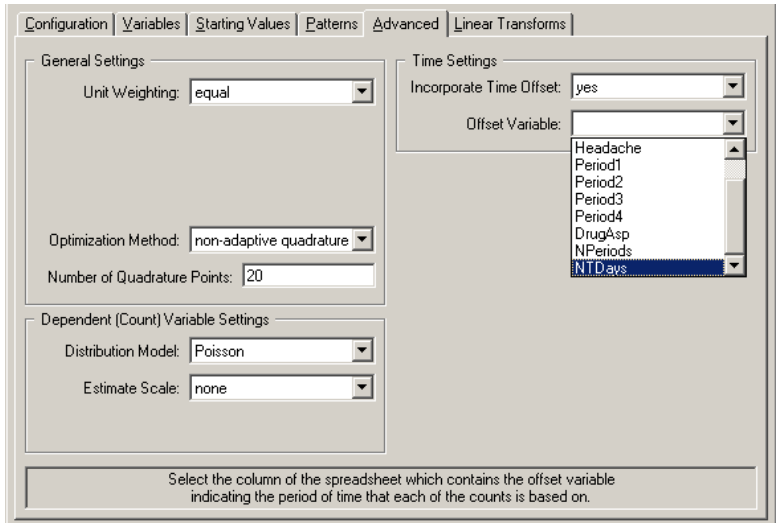
Example: Mixed-effects analysis with an offset variable

The previous analysis has assumed that the counts were all observed for the same number of days. However, this was not the case since the number of treatment days in the period (the 9th field of the input data file) did vary to some degree. Most of the counts were based on the full seven days in the week; however, some observations were made only for 1 day in the given week. To take this into account, we need to specify a so-called OFFSET variable. The offset variable indicates the amount of time that each count is based on. If OFFSET = no is specified, then SuperMix assumes that all counts are based on the same amount of time.

Setting up the analysis

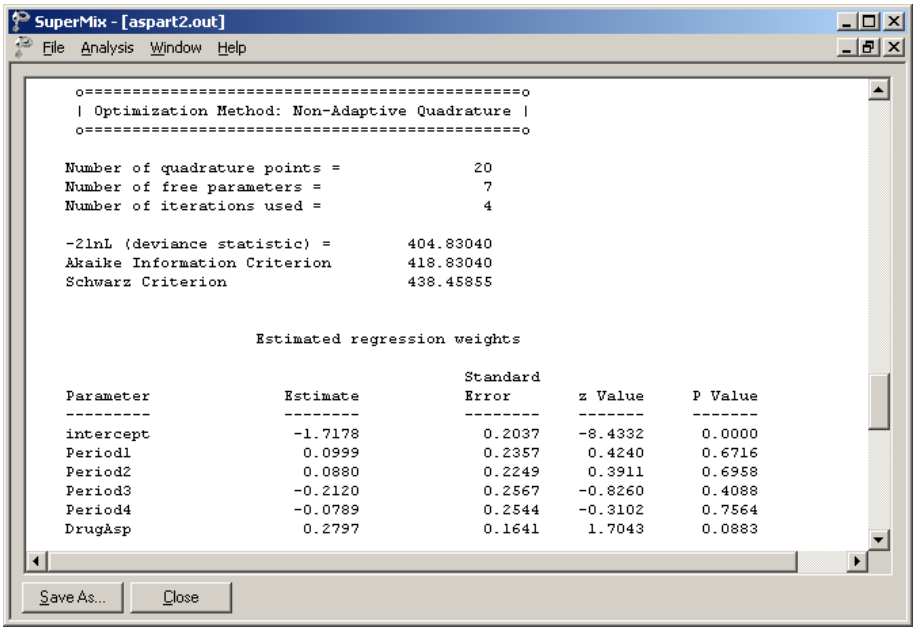
To create the model specifications for this model, start by opening **aspart.ss3** in a SuperMix spreadsheet window. Then, use the **Open Existing Model Setup** option on the **File** menu to open the **Model Setup** window for **aspart.mum**. On the **Configuration** screen, extend the title in the **Title 1** text box by adding the string "with Offset Variable". Next, click on the **Advanced** tab of the **Model Setup** window. Select yes from the **Incorporate Time Offset** drop-down list to activate the **Offset Variable** drop-down list box. Select the variable NTDays from the drop-down list of **Offset Variable** to produce the following **Advanced** screen.

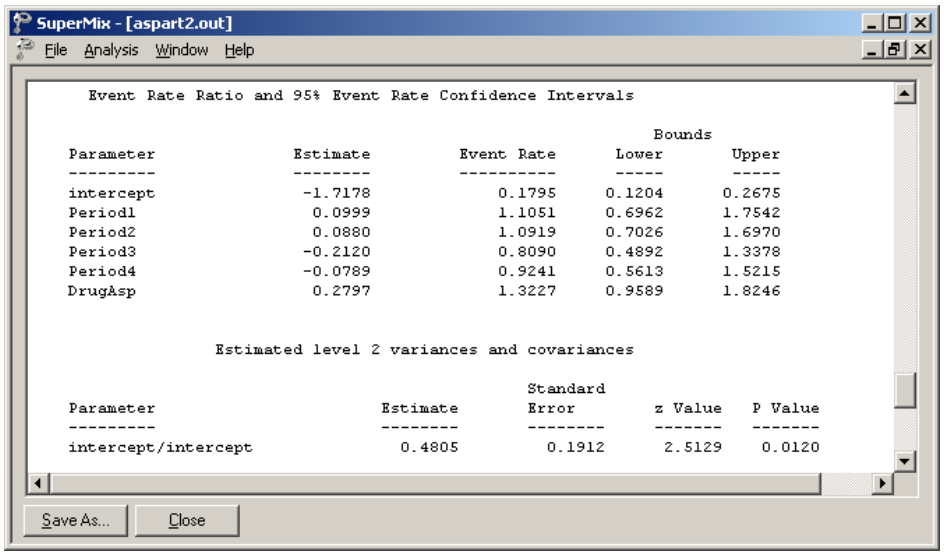
Save the changes to the file **aspart2.mum** by using the **Save As** option on the **File** menu. To fit the revised model to the data, select the **Run** option on the **Analysis** menu to produce the output file **aspart2.out**.



Discussion of results

A portion of this output file is shown below.





SuperMix - [aspart2.out]

File Analysis Window Help

Event Rate Ratio and 95% Event Rate Confidence Intervals

Parameter	Estimate	Event Rate	Bounds Lower	Bounds Upper
intercept	-1.7178	0.1795	0.1204	0.2675
Period1	0.0999	1.1051	0.6962	1.7542
Period2	0.0880	1.0919	0.7026	1.6970
Period3	-0.2120	0.8090	0.4892	1.3378
Period4	-0.0789	0.9241	0.5613	1.5215
DrugAsp	0.2797	1.3227	0.9589	1.8246

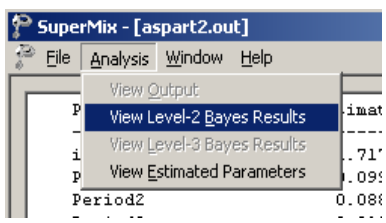
Estimated level 2 variances and covariances

Parameter	Estimate	Standard Error	z Value	P Value
intercept/intercept	0.4805	0.1912	2.5129	0.0120

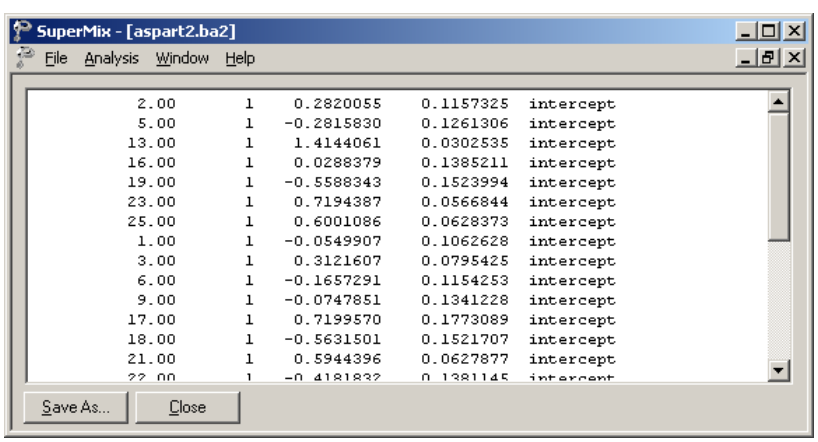
Save As... Close

Here, we see a marginally significant positive relationship between drug treatment and number of headaches. All time effects are again non-significant.

As mentioned, the empirical Bayes estimates of the random effects are written to the file **aspart2.ba2** as shown below at the conclusion of the SuperMix run. To view the file, click the **Analysis**, **View Level-2 Bayes Results** option on the output window on as shown below.



The first few lines of this file are shown below.



SuperMix - [aspart2.ba2]

File Analysis Window Help

2.00	1	0.2820055	0.1157325	intercept
5.00	1	-0.2815830	0.1261306	intercept
13.00	1	1.4144061	0.0302535	intercept
16.00	1	0.0288379	0.1385211	intercept
19.00	1	-0.5588343	0.1523994	intercept
23.00	1	0.7194387	0.0566844	intercept
25.00	1	0.6001086	0.0628373	intercept
1.00	1	-0.0549907	0.1062628	intercept
3.00	1	0.3121607	0.0795425	intercept
6.00	1	-0.1657291	0.1154253	intercept
9.00	1	-0.0747851	0.1341228	intercept
17.00	1	0.7199570	0.1773089	intercept
18.00	1	-0.5631501	0.1521707	intercept
21.00	1	0.5944396	0.0627877	intercept
22.00	1	-0.4181832	0.1381145	intercept

Save As... Close

The file **aspart2.ba2** contains four pieces of information per individual:

- the individual's ID,
- the number of the random effect,
- the empirical Bayes estimate for that individual (which is the mean of the posterior distribution), and
- the associated posterior standard deviation, and
- the name of the relevant random coefficient.

Since they are estimates of u_{i0} for each individual, the empirical Bayes estimates are expressed on the standard normal scale. Inspection of these estimates indicates that subject 13 has a very high score. This person's estimate of 1.043 (with standard deviation .016) suggests a very high level of headaches. This agrees well with the raw data, which reveals that this person recorded 7 headaches on four occasions and 6 on the only other occasion.

Graphical displays

Figure 3.9 is a comparison (represented by a dotted line) of the predicted average number of headaches reported by each patient when taking a placebo (left axis) as opposed to the predicted average number when the treatment is aspartame (right axis). From the graphical display, it appears as if all of the lines (each representing a patient) have a positive slope. The slopes become steeper as the number of headaches increases. This suggests an increase, albeit small, in the expected average number of headaches when aspartame is used. Note that patient 13, who reported a consistently high number of headaches at all occasions, was excluded from this graph.

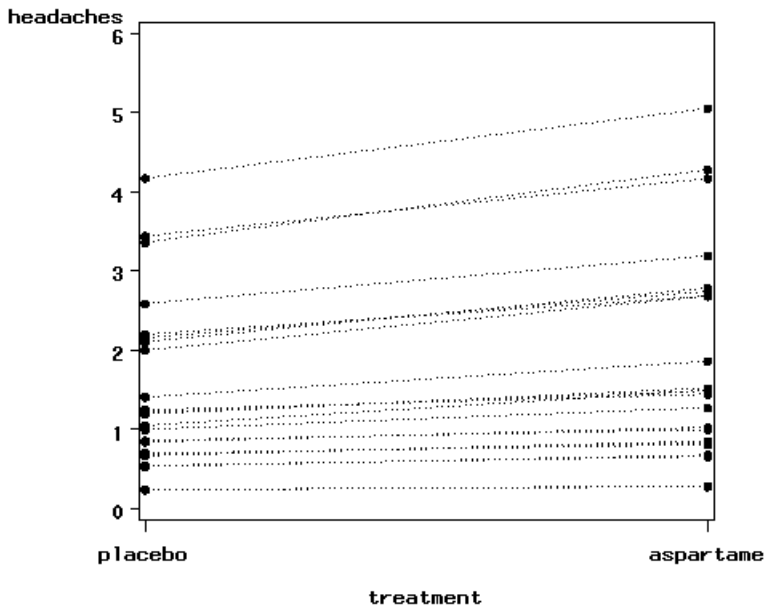
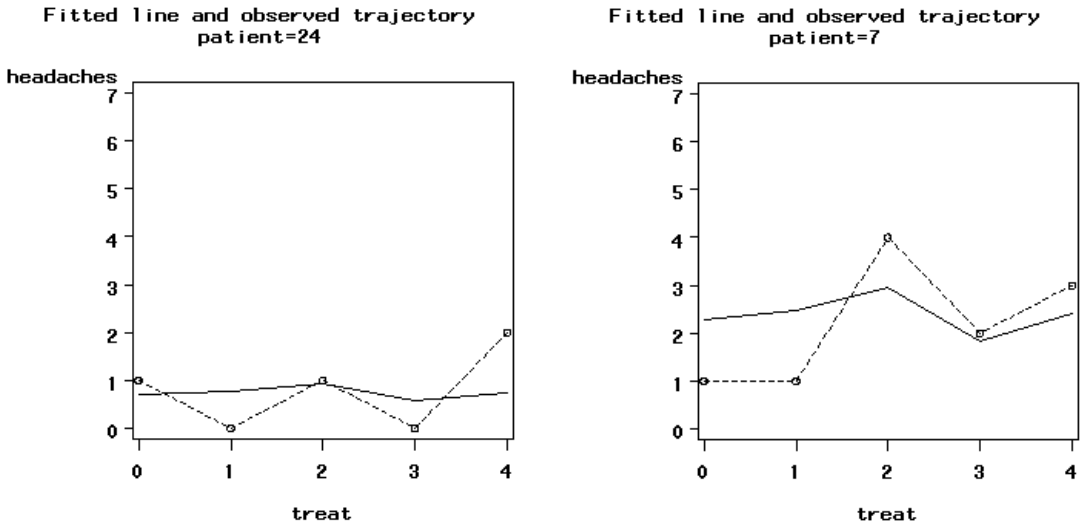


Figure 3.9: Predicted average number of headaches for placebo and aspartame



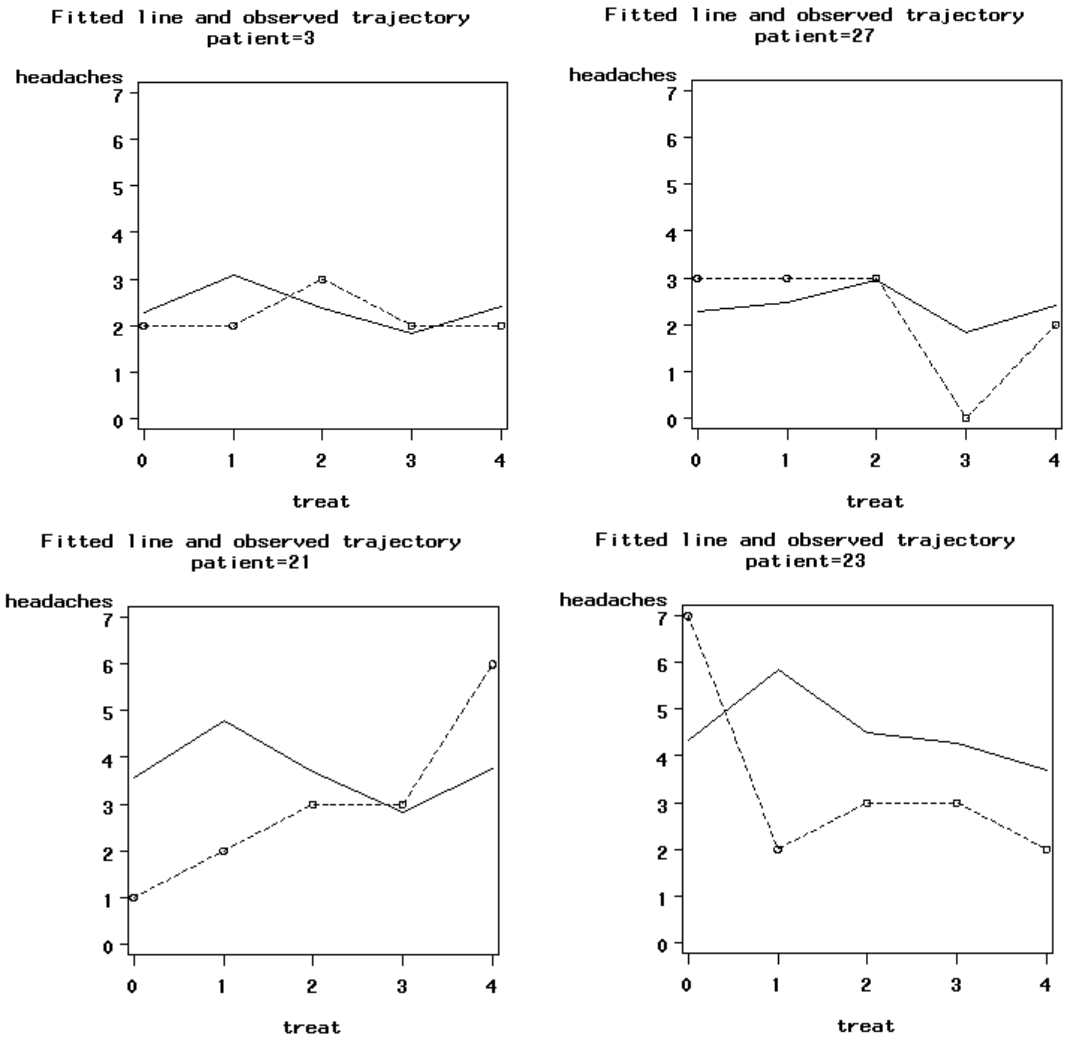


Figure 3.10: Fitted and observed trajectories

Figure 3.10 is a graphical display of the fitted trajectory (solid line) and observed trajectory (dotted line) for a sample of 6 patients. These displays are ordered from a patient who reported a relatively small number of headaches at the different treatment occasions to one who reported a relatively high number of headaches at the treatment occasions. A study of the fitted and observed trajectories reveals that, in general, the model fit is best when the number of headaches is smaller and becomes less accurate as the number of headaches increases. For patient 13, who is not represented in the graphical display, the number of predicted headaches is almost twice the number observed.

The fitted lines were obtained as

$$\hat{y} = \exp(-1.7178 + 0.0999(\text{Period1})_{ij} + 0.0880(\text{Period2})_{ij} - 0.2120(\text{Period3})_{ij}$$
$$-0.0789(\text{Period4})_{ij} + 0.2797(\text{DrugAsp})_{ij} + \hat{u}_{i0}$$

where $\hat{u}_{i0}$ is obtained from the **aspart2.ba2** file, shown in the **discussion of results** section.