Models for nominal outcomes using NHIS data

Contents

1. Introduction

In statistics, the kinds of significance tests and model fitting procedures that are appropriate depend on the level of measurement of the variables concerned. A widely accepted classification scheme, proposed by Stevens (1946), is listed below and consists of four levels of measurement:

- nominal (also categorical or discrete)
- ordinal
- interval
- ratio

Interval and ratio variables are usually grouped together as continuous variables.

In the case of nominal variables there are no "less than" or "greater than" relations among the categories of the variable and operations such as addition or multiplication do not exist.

Examples of nominal variables are

- Cancer Type (1 = breast, 2 = lung, 3 = brain, 4 = leukemia, 5 = liver, 6 = colon, 7 = other),

- Smoking Status (1 = never smoked, 2 = former smoker, 3 = current smoker),
- Preference for U.S. President (1 = Democrat, 2 = Republican, 3 = Independent),
- Type of Sweetener (1 = sugar, 2 = saccharin, 3 = aspartame, 4 = other),
- Pain Reliever (1 = Acetaminophen, 2 = Aspirin, 3 = Ibuprofen, 4 = Ketoprofen, 5 = Naproxen, 6 = other).

In many research situations, the underlying variable type is continuous. However, to ensure anonymity of respondents, information is obtained by categorizing variables. For example:

- Annual Income (1 = not employed, 2 = less or equal to $20,000, 3 = more than $20,000 but less than or equal to $50,000, 4 = more than $50,000 but less than or equal to $100,000, 5 = more than $100,000)
- Age when diagnosed (1 = not applicable, 2 = younger than 25 years, 3 = 25 years or older but less than 50 years, 4 = 50 years or older but less than 70 years, 5 = 70 years and older).

In both the cases above, the available data values are coded 1, 2, 3, 4 and 5. Arithmetic operation with these codes will not provide accurate estimates of the actual age and income characteristics and in both cases the first category makes "less than" and "more than" comparisons less feasible.

In this guide we illustrate the analysis of a nominal outcome variable by fitting a three-level model to health related data.

2. The data

The data set comes from the data library of the National Health Interview Survey (NHIS). The NHIS is a national longitudinal health survey. During 2002, background data and data on the health conditions of a sample of 28,737 participants were obtained. The 2002 sample was stratified into 64 strata and into 601 PSUs. Using this data, we created a subset consisting of 57 strata (the level-3 units), 399 PSUs (the level-2 units) and 6445 participants. A partial list of the data is given below in the form of a LISREL spreadsheet file named **nihs_subset.lsf** in the **Multilevel Generalized Linear Model Example** folder.

nihs_subset.LSF

	CSTRATM	CPSUM	PATWT	PASTVIS	NUMMED	GENDER	USETOBAC
1	20102101.00	100013.00	50245.00	3.00	2.00	1.00	1.00
2	20102101.00	100013.00	50245.00	3.00	2.00	1.00	0.00
3	20102101.00	100013.00	50245.00	3.00	4.00	1.00	0.00
4	20102101.00	100013.00	50245.00	3.00	2.00	1.00	0.00
5	20102101.00	100013.00	50245.00	2.00	1.00	0.00	0.00
6	20102101.00	100015.00	72581.00	4.00	0.00	1.00	0.00
7	20102101.00	100015.00	72581.00	3.00	2.00	1.00	0.00
8	20102101.00	100015.00	72581.00	3.00	2.00	0.00	0.00
9	20102101.00	100015.00	72581.00	1.00	1.00	0.00	0.00
10	20102101.00	100015.00	72581.00	3.00	0.00	1.00	0.00

A description of the variables is as follows:

- CSTRATM is the stratum used as level-3 ID (57 strata).
- CPSUM is the primary sampling unit (PSU) and is used as level-2 ID (399 clusters).
- PATWT is the participant design weight.
- PASTVIS is the value of the nominal variable for the number of visits to a medical doctor during the past 12 months (1 = none or unknown, 2 = 1 to 2, 3 = 3 to 5, 4 = 6 medications and more).
- NUMMED is the number of medications.
- GENDER, where 0 = Female and 1 = Male.
- USETOBAC indicates whether a participant smoked cigarettes or not, where 0 = no and 1 = yes.
- PRIMCARE, where 0 = none and 1 = participant has primary care.
- INJURY indicates whether a participant suffered from an injury or not (0 = no, 1 = yes).
- BLODPRES, where 0 = blood pressure not measured and 1 = blood pressure measured.
- URINE, where 0 = no urine tested, 1 = tested.
- XRAY, where 0 = no X rays taken and 1 = X ray taken.
- EXERCISE, where 0 = no exercise and 1 = participant does exercise.
- RACER indicates the ethnicity of a participant where 1 = White, 2 = Black and 3 = Other.
- AGER indicates in which age category a participant belongs. Coded as follows: 1 = Under 15, 2 = 15 to 24, 3 = 25 to 44, 4 = 45 to 64, 5 = 65 to 74, 6 = 75 and older.
- AGE1 to AGE5 are five dummy variables coded as follows:

Table 8: Dummy variables

Age	AGE1	AGE2	AGE3	AGE4	AGE5
Under 15	1	0	0	0	0
15 to 24	0	1	0	0	0
25 to 44	0	0	1	0	0
45 to 64	0	0	0	1	0
65 to 74	0	0	0	0	1
75 and older	0	0	0	0	0

3. The model

A general multilevel nominal model

In the nominal case we need to consider the values corresponding to the unordered multiple categories of the response variable. We thus assume that the C response categories are coded as $1,2,3,...,C$.

Let $P_{ijkc} = P\left(y_{ijk} = c \mid \boldsymbol{\beta}_c, \boldsymbol{\upsilon}_{ic}, \boldsymbol{\upsilon}_{ijc}\right)$ denote the probability that a response occurs in category c, conditional on a $(p \times 1)$ vector of fixed regression parameters $\boldsymbol{\beta}_c$, the $(m \times 1)$ vector of level-2 random effects $\boldsymbol{\upsilon}_{ijc}$ and the $(r \times 1)$ vector of level-3 random effects $\boldsymbol{\upsilon}_{ic}$. It is further assumed that the level-2 random effects $\boldsymbol{\upsilon}_{ijc}$ are independent and identically distributed (i.i.d.) as a $N\left(\mathbf{0}, \boldsymbol{\Phi}_{(2)}\right)$

random variable. Uncorrelated with $\boldsymbol{\upsilon}_{ijc}$, the level-3 random effects are i.i.d. $N\left(\mathbf{0}, \mathbf{\Phi}_{(3)}\right)$. The scalar y_{ijk} denotes the value of the nominal variable associated with level-1 unit *k*, $k = 1, 2, \ldots, n_{ij}$, nested within level-2 unit *j*, $j = 1, 2, \ldots n_i$, which in turn is nested within the *i*-th level-3 unit, where $i = 1, 2, \ldots, N$. The probabilities P_{ijkc} are computed as

$$\begin{aligned} P_{ijkc} &= P\left(y_{ijk} = c \mid \boldsymbol{\beta}_c, \boldsymbol{\upsilon}_{ic}, \boldsymbol{\upsilon}_{ijc}\right) \\ &= \frac{\exp\left(\eta_{ijkc}\right)}{1 + \sum_{h=1}^{C-1} \exp\left(\eta_{ijkc}\right)}, \qquad c = 1, 2, \ldots, C-1 \end{aligned}$$

where

$$\eta_{ijkc} = \mathbf{x}'_{ijk} \boldsymbol{\beta}_c + \mathbf{z}'_{ijk(2)} \boldsymbol{\upsilon}_{ijc} + \mathbf{z}'_{ijk(3)} \boldsymbol{\upsilon}_{ic}$$

Note that $\mathbf{x}'_{ijk}$, $\mathbf{z}'_{ijk(2)}$ and $\mathbf{z}'_{ijk(3)}$ are design vectors for the explanatory variables and the level-2 and level-3 random effects respectively.

Random intercept model with two explanatory variables

For the **nihs_subset.lsf** data set considered earlier, let PASTVIS denote the outcome variable. Assume further that GENDER and EXERCISE are the only predictors and that only intercepts are allowed to vary randomly across level-3 and level-2 units. The corresponding estimated probability model is given by

$$P\left(\text{PASTVIS}_k = c\right) = \frac{\exp\left(\eta_{ijkc}\right)}{1 + \sum_{h=1}^{3} \exp\left(\eta_{ijkh}\right)}, \qquad c = 1, 2, 3$$

where

$$\eta_{ijkh} = \beta_{0h} + \beta_{1h} \times \text{GENDER}_k + \beta_{1h} \times \text{EXERCISE}_k + \upsilon_{ijh} + \upsilon_{ih}$$

and where PASTVIS_k, GENDER_k and EXERCISE_k denote values of the variables for client *k* nested within unit (i, j). Note that for PASTVIS the number of categories is $C = 4$.

Remarks:

The probability $P\left(\text{PASTVIS}_k = 4\right)$ is obtained as $1 - \sum_{c=1}^{3} P\left(\text{PASTVIS}_k = c\right)$. In the formulation above, we used the last category as the so-called reference category.

MGLIM allows the user to select the first or the last category as the reference category. If the first category is selected as reference category, then

$$P(\text{PASTVIS}_k = c) = \frac{\exp(\eta_{ijkc})}{1 + \sum_{h=2}^{4} \exp(\eta_{ijkh})}, \qquad c = 2,3,4.$$

$$P(\text{PASTVIS}_k = 1) = 1 - \sum_{c=2}^{4} P(\text{PASTVIS}_k = c)$$
.

4. A random intercept model with fourteen predictors

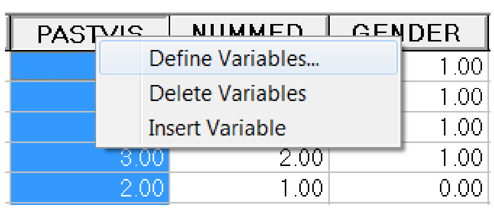
Preparing the data

The model to be fitted to the data is contained in **nihs_subset1.lsf**. This file was created from the SPSS data file **nihs_subset1.sav** as follows.

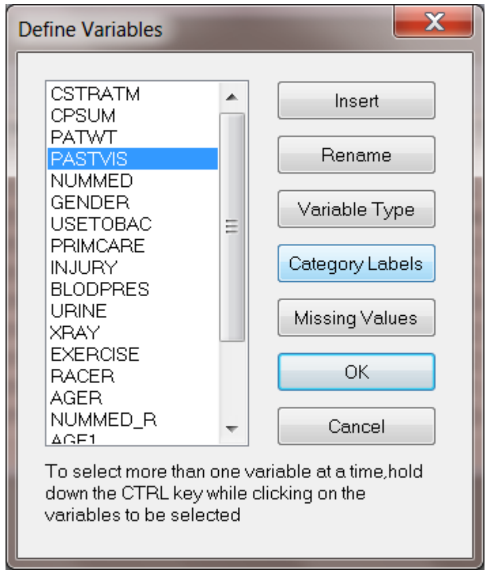
Use the **File**, **Import Data File** option to activate the display of an **Open** dialog box. From the **Files of type** drop-down list, select **SPSS Data File (*.sav)**. Browse for the file **nihs_subset.lsf**. Select the file and click the **Open** button to activate **Save As** dialog box. Enter the file name **nihs_subset.lsf** and click on the Save button to display **nihs_subset.lsf** in the LISREL spreadsheet.

Exploring the data

To obtain some insight into the distributional properties and possible relationships between variables, it is useful to present these properties graphically using the **Graphs** option. Prior to making visual presentations, it is a good idea to assign labels to the categories of the nominal and ordinal variables. First, highlight the column of PASTVIS by clicking on its header. Then right click and select the **Define Variables** option as shown below to open the **Define Variables** dialog box.



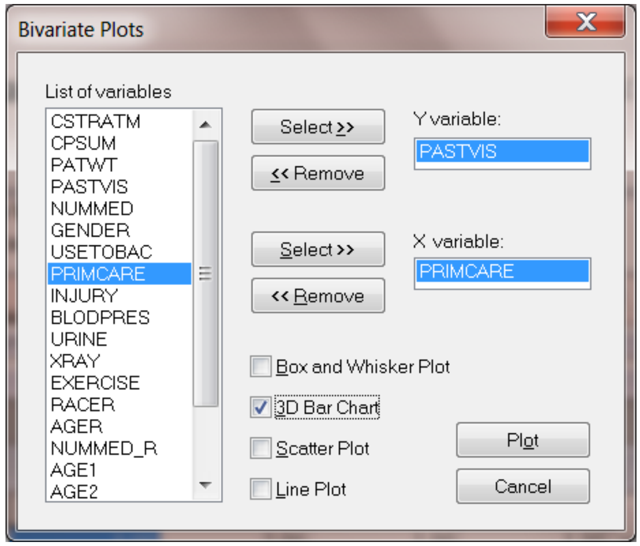
Select variable PASTVIS as shown below to activate all the options on the **Define Variables** dialog box.



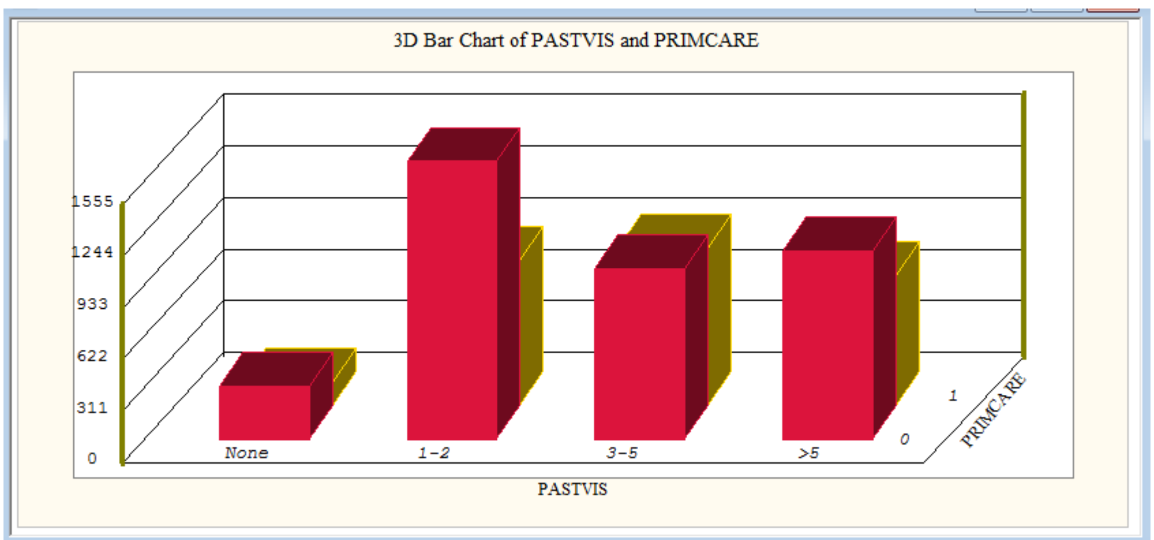
Click on the **Category Labels** option to activate the **Category Labels for PASTVIS** dialog box. Enter the labels None, 1 to 2, 3 to 5 and >5 as shown below and click **OK**.

Click on the **OK** button to return to the **Define Variables** dialog box. Click on **OK** button to return to the LSF window. Save the change to the data set by clicking on the **File**, **Save** option.

From the main menu bar, select the **Graphs**, **Bivariate** option. By clicking on the **Bivariate** tab of the pop-up menu, the **Bivariate plot** dialog box is invoked. Select PASTVIS as the Y variable and PRIMCARE as the X variable.

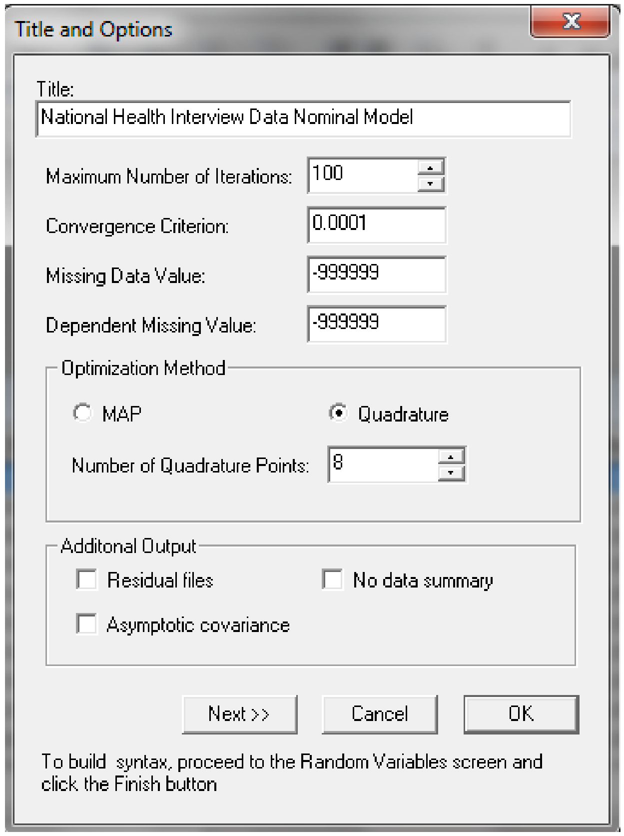


Next, check the **3D Bar Chart** check box and then the **Plot** button to obtain the bivariate bar chart of PRIMCARE versus PASTVIS. The graph below shows that there is an increase in the use of primary care with the number of visits to a medical doctor.

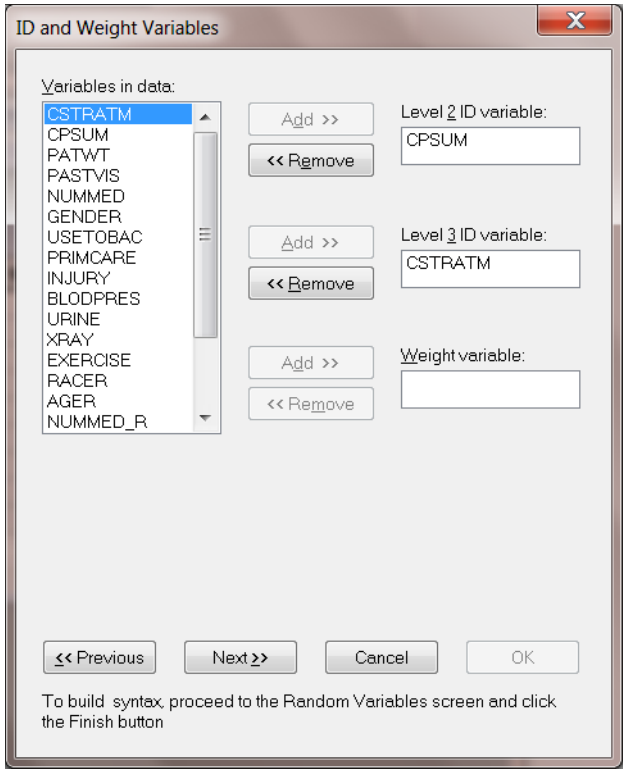


5. Setting up the analysis

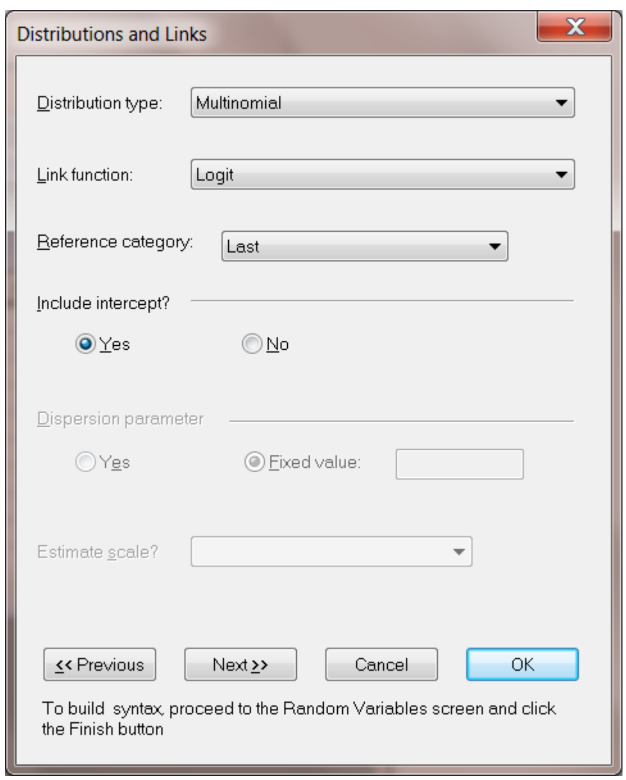
From the main menu bar of the LSF window, select the **Multilevel**, **Generalized Linear Model**, **Title and Options** option. Enter a title for the analysis in the **Title** text boxes. Keep the default settings for the **Maximum Number of Iterations**, **Convergence Criterion** and the **Missing Data Value**. Activate **Quadrature** radio button in the **Optimization Method** section and change the **Number of Quadrature Points** to 8 to obtain the above screen. Proceed to the **ID and Weight Variables** dialog box by clicking on the **Next** button.



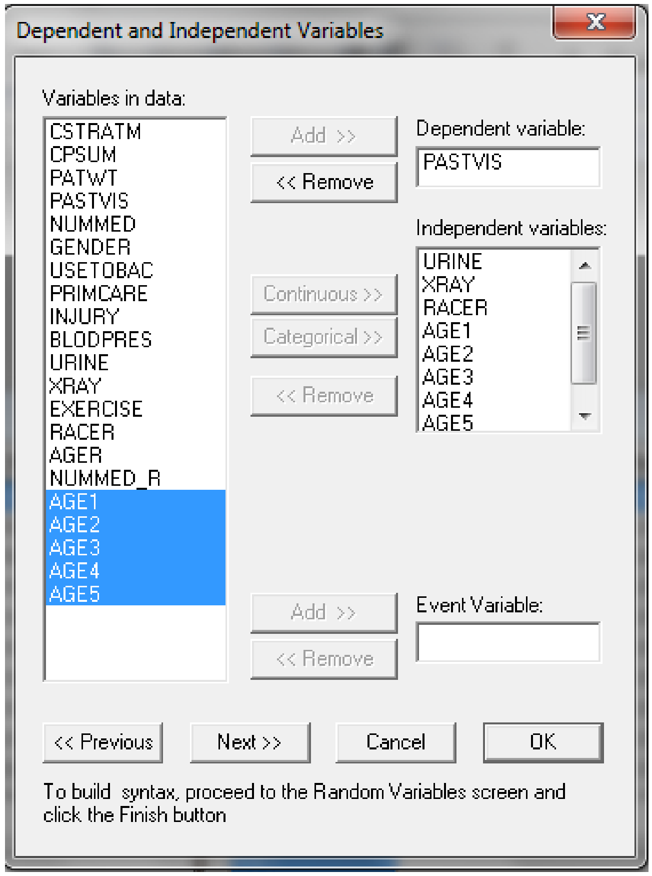
Select CPSUM from the **Variables in data** list box. Click on the upper **Add** button of the **Level-2 ID variable** section to define the level-2 ID. Similarly define CSTRATM and click on the middle **Add** button to define it as **Level-3 ID variable**.



Proceed to the **Distribution and Links** dialog box by clicking on the **Next** button. Keep all the default settings on this dialog box as shown below.

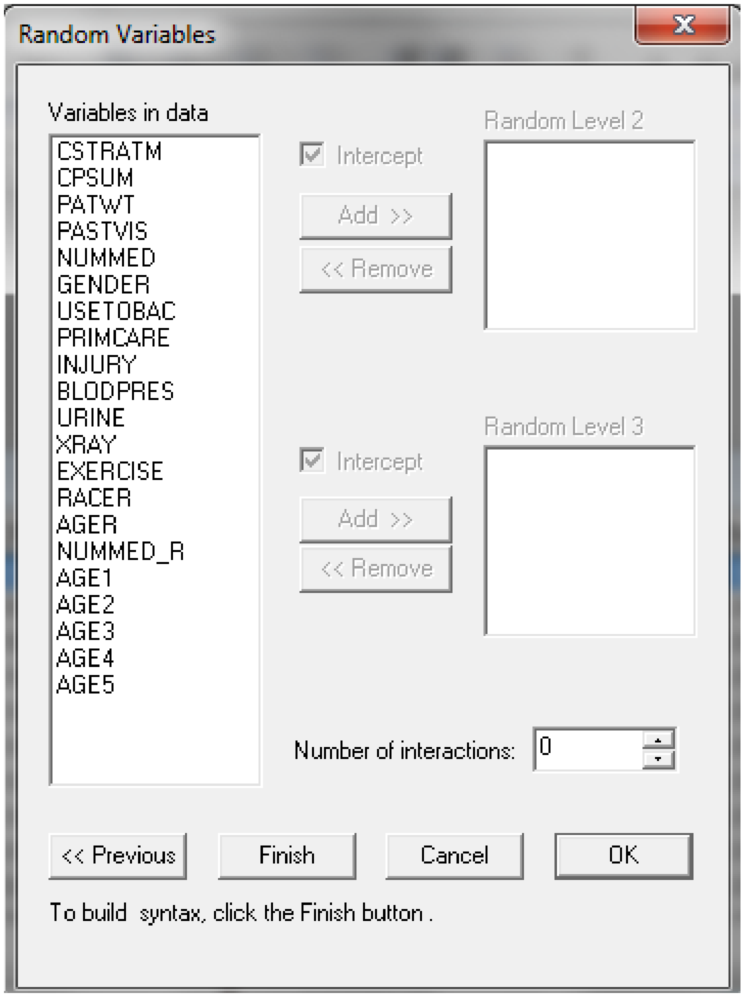


Click on the **Next** button to proceed to the **Dependent and Independent Variables** dialog box.

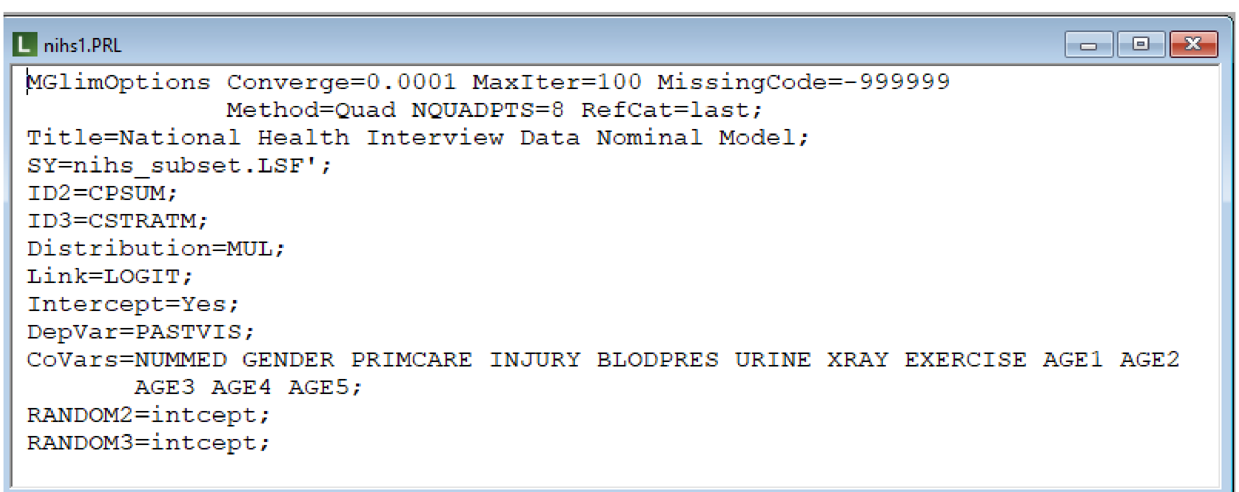


The **Dependent and Independent Variables** dialog box is used to specify the dependent and independent variables. First, select the dependent variable PASTVIS from the **Variables in data** list box and then click on the **Add** button to define it as the **Dependent variable**. Next, select NUMMED, GENDER, PRIMCARE, INJURY, BLODPRES, URINE, XRAY, EXERCISE, AGE1, AGE2, AGE3, AGE4 and AGE5 and click on the **Continuous** button to add them as **Independent variables** as shown below.

Click on the **Next** button to activate the **Random Variables** dialog box. By default, the **Intercept** check box in the **Random Level-2** and **Random Level-3** are checked, indicating the inclusion of a random intercept at level-2 and 3 in the model. Keep the default settings as shown below and click on the **Finish** button to generate the PRELIS syntax (**prl**) file.



Before running the analysis, the PRELIS syntax file should be saved. Select the **File**, **Save As** option, and provide a name (**nihs1.prl**) for the syntax file. Run the analysis by selecting the **Run PRELIS** icon as shown below.




```
MGlimOptions Converge=0.0001 MaxIter=100 MissingCode=-999999
             Method=Quad NQUADPTS=8 RefCat=last;
Title=National Health Interview Data Nominal Model;
SY=nihs_subset.LSF';
ID2=CPSUM;
ID3=CSTRATM;
Distribution=MUL;
Link=LOGIT;
Intercept=Yes;
DepVar=PASTVIS;
CoVars=NUMMED GENDER PRIMCARE INJURY BLODPRES URINE XRAY EXERCISE AGE1 AGE2
       AGE3 AGE4 AGE5;
RANDOM2=intcept;
RANDOM3=intcept;
```

6. Discussion of results

Model and data description

```
nihs1.OUT

o============================================o
| National Health Interview Data Nominal Model |
o============================================o

            Model and Data Descriptions

   Sampling Distribution                    = Multinomial
   Link Function                            = Logistic
   Number of Level-3 Units                  = 57
   Number of Level-2 Units                  = 399
   Number of Level-1 Units                  = 6444
   Number of Level-2 Units per Level-3 Unit =
     3     5     7     2     4     3     5    14     9     4     7     6
    13    12     9    14     6     5     3    11     4     4    11     8
     4     6     4    10     4     9    10     2     7     6     5     6
     7     5     8     6     5     6    19     9     8    18     4     9
     2     8     2     4     7     7     4     6    13
   Number of level-1 units for the first (level-3, level-2) unit combination =
     5     6     9
```

The first part of the output file gives a description of the model specifications. This is followed by a data summary of the number of observations nested within each subject.

Descriptive statistics and starting values

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

```
nihs1.OUT
o========================================================o
| Descriptive statistics for all the variables in the model |
o========================================================o
                                                     Standard
   Variable      Minimum      Maximum       Mean    Deviation
   --------      -------      -------       ----    ---------
   PASTVIS1       0.0000       1.0000     0.0691       0.2536
   PASTVIS2       0.0000       1.0000     0.3678       0.4822
   PASTVIS3       0.0000       1.0000     0.2858       0.4519
   PASTVIS4       0.0000       1.0000     0.2773       0.4477
   intcept        1.0000       1.0000     1.0000       0.0000
   NUMMED         0.0000       6.0000     1.5286       1.6864
   GENDER         0.0000       1.0000     0.4196       0.4935
   PRIMCARE       0.0000       1.0000     0.3993       0.4898
   INJURY         0.0000       1.0000     0.0906       0.2871
   BLODPRES       0.0000       1.0000     0.4008       0.4901
   URINE          0.0000       1.0000     0.1041       0.3055
   XRAY           0.0000       1.0000     0.0380       0.1913
   EXERCISE       0.0000       1.0000     0.1047       0.3063
   AGE1           0.0000       1.0000     0.1966       0.3975
   AGE2           0.0000       1.0000     0.0644       0.2455
   AGE3           0.0000       1.0000     0.2284       0.4199
   AGE4           0.0000       1.0000     0.2668       0.4423
   AGE5           0.0000       1.0000     0.1155       0.3196
```

Each category of the nominal outcome variable is denoted as PASTVIS_i, $i = 1, 2, 3, 4$. From the output it can be seen that the distribution of respondents over these categories are 6.9%, 36.8%, 28.6%, and 27.7% respectively. The age distribution is given in Table 9.

Table 9: Age distribution of respondents

Age	**Percentage**
Younger than 15 (AGE1)	19.7
15 to 24 (AGE2)	6.4
25 to 44 (AGE3)	22.8
45 to 64 (AGE4)	26.7
65 to 74 (AGE5)	11.6
75 and older	12.8*

*: calculated as 100 – (19.7+6.4+22.8+26.7+11.6)

```
nihs1.OUT

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

                    Goodness of fit statistics

   Statistic                               Value         DF
   ---------                               -----         --
   Likelihood Ratio Chi-square        15994.0450       6402
   Pearson Chi-square                 19269.4871       6402

                    Estimated regression weights

                                   Standard
   Parameter      Estimate           Error     z Value     P Value
   ---------      --------        --------     -------     -------

   Response Code  1 vs Code    4
   ----------------------------
   intcept         -1.4038          0.1876     -7.4847      0.0000
   NUMMED          -0.2022          0.0373     -5.4238      0.0000
   GENDER           0.0325          0.1110      0.2926      0.7698
   PRIMCARE        -0.5933          0.1273     -4.6618      0.0000
   INJURY           0.0623          0.1883      0.3311      0.7406
   BLODPRES         0.7097          0.1276      5.5633      0.0000
   URINE            0.4633          0.1634      2.8359      0.0046
   XRAY             0.0299          0.3026      0.0988      0.9213
   EXERCISE         0.3110          0.1690      1.8399      0.0658
   AGE1             0.2907          0.2186      1.3298      0.1836
   AGE2             0.2566          0.2607      0.9842      0.3250
   AGE3             0.0970          0.1934      0.5016      0.6160
   AGE4             0.1434          0.1906      0.7523      0.4519
   AGE5            -0.0250          0.2305     -0.1086      0.9135
```

The estimated parameters for the model, assuming no random effects, are reported next. For each response code i versus code 4, $i = 1, 2, 3,$ there are 14 parameter estimates. Only the estimates for response code 1 versus response code 4 are displayed. Comparing these estimates with those obtained when allowance is made for the hierarchical structure of the data, a considerable difference is apparent.

Fixed effects estimates and fit statistics

The final results obtained using adaptive quadrature are given next. Using 8 quadrature points, 6 iterations were required to reach convergence. The deviance statistic ($-2\ln L$) allows the user to compare the current model with other nested models.

nihs1.OUT

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

   Number of quadrature points =                8
   Number of free parameters =                 48
   Number of iterations used =                  5

   -2lnL (deviance statistic) =       14329.74536
   Akaike Information Criterion       14425.74536
   Schwarz Criterion                  14750.74879


                       Estimated regression weights

                                     Standard
   Parameter       Estimate           Error      z Value      P Value
   ---------       --------         --------     -------      -------

   Response Code  1 vs Code   4
   ----------------------------
   intcept          -1.5012           0.2483     -6.0447       0.0000
   NUMMED           -0.3325           0.0446     -7.4481       0.0000
   GENDER           -0.0355           0.1285     -0.2764       0.7822
   PRIMCARE         -1.0197           0.1954     -5.2190       0.0000
   INJURY            0.2728           0.2221      1.2282       0.2194
   BLODPRES          0.3133           0.1707      1.8351       0.0665
   URINE             0.3527           0.1962      1.7979       0.0722
   XRAY             -0.0141           0.3320     -0.0423       0.9662
   EXERCISE          0.7546           0.2047      3.6868       0.0002
   AGE1              1.0955           0.2621      4.1803       0.0000
   AGE2              1.3540           0.3013      4.4941       0.0000
   AGE3              0.9306           0.2249      4.1374       0.0000
   AGE4              0.7571           0.2165      3.4969       0.0005
   AGE5              0.0916           0.2545      0.3598       0.7190
```

A study of the p-values associated with the parameter estimates indicates that the estimated GENDER, INJURY, URINE, and XRAY coefficients are not significant, regardless of the values of the category of the outcome variable.

Random effect estimates

The last part of the output file shows the variance estimates for the level-2 and level-3 random effects. Both effects are highly significant.

nihs1.OUT

Estimated level 2 variances and covariances

Parameter	Estimate	Standard Error	z Value	P Value
intcept1/intcept1	4.7004	0.4312	10.8997	0.0000
intcept2/intcept2	3.2406	0.2791	11.6099	0.0000
intcept3/intcept3	1.0800	0.1386	7.7938	0.0000

Estimated level 3 variances and covariances

Parameter	Estimate	Standard Error	z Value	P Value
intcept1/intcept1	1.0122	0.2513	4.0287	0.0001
intcept2/intcept2	0.9227	0.1875	4.9221	0.0000
intcept3/intcept3	0.8511	0.1523	5.5881	0.0000

7. Interpreting the output

Estimated unit-specific probabilities

The estimated regression coefficients given in the adaptive quadrature portion of the output provide the information necessary to compute unit-specific probabilities for a typical participant that is associated with each possible combination of the predictor variables. For example, consider a typical female patient (GENDER = 0) that received 3 medications (NUMMED = 3), has primary care (PRIMCARE = 1), had no injuries (INJURY = 0), did not have a blood pressure or urine test (BLODPRES = URINE = 0), does not exercise (EXERCISE = 0), and is in the age group 25 to 44 (AGE3 = 1).

For response code 1 vs. code 4:

$$\begin{aligned}\hat{\eta}_{ijk1} &= -1.5004 - 0.3320(\text{NUMMED}_{ijk}) - 0.0395(\text{GENDER}_{ijk}) - 1.0176(\text{PRIMCARE}_{ijk}) \\ &\quad +....1.0945(\text{AGE1}_{ijk}) + 1.3539(\text{AGE2}_{ijk}) + 0.9306(\text{AGE3}_{ijk}) \\ &\quad +0.7572(\text{AGE4}_{ijk}) + 0.1136(\text{AGE5}_{ijk}) \\ &= -1.5004 - 3(0.3320) - 1(1.0176) + 1(0.9306) \\ &= -2.5834\end{aligned}$$

so that $\exp\left(\hat{\eta}_{ijk1}\right) = 0.0755.$

For response code 2 vs. 4, we find that

$$\begin{aligned}\hat{\eta}_{ijk2} &= 0.3737 - 3(0.2360) - 1(0.9167) + 1(0.6972) \\ &= -0.5538\end{aligned}$$

and thus

$$\exp\left(\hat{\eta}_{ijk2}\right) = 0.5748.$$

For response code 3 vs. code 4

$$\begin{aligned}\hat{\eta}_{ijk3} &= 0.3440 - 3(0.0718) - 1(0.3004) + 1(0.2070) \\ &= 0.0352\end{aligned}$$

and thus

$$\exp\left(\hat{\eta}_{ijk3}\right) = 1.0358.$$

Using these values, it follows that

$$\begin{aligned}&Prob(respondent\ not\ seen\ doctor\ previously) \\ &= \frac{0.0755}{1 + 0.0755 + 0.5748 + 1.0358} \\ &= 0.0281.\end{aligned}$$

The next two tables contain a selection of unit-specific probabilities for the four categories of PASTVIS for females (GENDER = 0).

Unit-specific probabilities for females with XRAY = no, INJURY = no, URINE = no, and BLODPRES = no

NUMMED	PRIM CARE	EXER CISE	AGE	ETA1	ETA2	ETA3	PROB1	PROB2	PROB3	PROB4
none	no	no	< 15	−0.406	1.393	0.628	0.088	0.532	0.248	0.132
none	no	no	25:44	−0.570	1.071	0.551	0.091	0.469	0.279	0.161
none	no	no	>= 75	−1.500	0.374	0.344	0.055	0.356	0.345	0.245
none	no	yes	< 15	0.347	1.776	0.953	0.130	0.541	0.238	0.092
none	no	yes	25:44	0.183	1.454	0.876	0.135	0.482	0.270	0.113
none	no	yes	>= 75	−0.748	0.757	0.669	0.085	0.384	0.351	0.180
none	yes	no	< 15	−1.424	0.476	0.328	0.057	0.380	0.327	0.236
none	yes	no	25:44	−1.587	0.154	0.251	0.056	0.319	0.351	0.274
none	yes	no	>= 75	−2.518	−0.543	0.044	0.030	0.215	0.386	0.370
none	yes	yes	< 15	−0.671	0.859	0.653	0.088	0.408	0.332	0.173
none	yes	yes	25:44	−0.835	0.537	0.576	0.088	0.348	0.361	0.203
none	yes	yes	>= 75	−1.765	−0.160	0.369	0.049	0.246	0.417	0.288
three	no	no	< 15	−1.402	0.685	0.413	0.052	0.418	0.319	0.211
three	no	no	25:44	−1.566	0.363	0.336	0.052	0.355	0.346	0.247
three	no	no	>= 75	−2.496	−0.334	0.129	0.028	0.244	0.387	0.341
three	no	yes	< 15	−0.649	1.068	0.738	0.080	0.446	0.321	0.153
three	no	yes	25:44	−0.813	0.746	0.661	0.081	0.384	0.353	0.182
three	no	yes	>= 75	−1.744	0.049	0.454	0.046	0.276	0.414	0.263
three	yes	no	< 15	−2.420	−0.232	0.112	0.030	0.264	0.373	0.333
three	yes	no	25:44	−2.584	−0.554	0.035	0.028	0.214	0.386	0.372
three	yes	no	>= 75	−3.514	−1.251	−0.172	0.014	0.133	0.390	0.463
three	yes	yes	< 15	−1.667	0.151	0.437	0.048	0.298	0.397	0.256
three	yes	yes	25:44	−1.831	−0.171	0.360	0.047	0.245	0.417	0.291
three	yes	yes	>= 75	−2.761	−0.868	0.153	0.024	0.158	0.440	0.378

From these tables we conclude that the proportion of female patients, regardless of age group, that indicated no prior visits to a medical practitioner (PASTVIS = 1) is generally low. Females who exercise have a lower probability of having several past visits when compared to those who do not exercise.

Unit-specific probabilities for females with XRAY = no, INJURY = no, URINE = no, and BLODPRES = yes

NUMMED	PRIM CARE	EXER CISE	AGE	ETA1	ETA2	ETA3	PROB1	PROB2	PROB3	PROB4
none	No	no	< 15	-0.096	1.712	0.839	0.093	0.567	0.237	0.102
none	No	no	25:44	-0.260	1.390	0.762	0.097	0.506	0.270	0.126
none	No	no	>=75	-1.191	0.693	0.555	0.060	0.396	0.345	0.198
none	No	yes	< 15	0.657	2.095	1.164	0.135	0.570	0.225	0.070
none	No	yes	25:44	0.493	1.773	1.088	0.142	0.512	0.258	0.087
none	No	yes	>=75	-0.438	1.076	0.880	0.092	0.420	0.345	0.143
none	Yes	no	< 15	-1.114	0.795	0.539	0.062	0.421	0.326	0.190
none	Yes	no	25:44	-1.278	0.473	0.462	0.062	0.359	0.355	0.224
none	Yes	no	>=75	-2.208	-0.224	0.255	0.034	0.250	0.403	0.313
none	Yes	yes	< 15	-0.361	1.178	0.864	0.095	0.444	0.324	0.137
none	Yes	yes	25:44	-0.525	0.856	0.787	0.096	0.383	0.358	0.163
none	Yes	yes	>=75	-1.456	0.159	0.580	0.056	0.280	0.426	0.239
three	No	no	< 15	-1.092	1.004	0.624	0.057	0.460	0.315	0.169
three	No	no	25:44	-1.256	0.682	0.547	0.057	0.396	0.346	0.200
three	No	no	>=75	-2.187	-0.015	0.340	0.032	0.281	0.401	0.286
three	No	yes	< 15	-0.339	1.387	0.949	0.086	0.482	0.311	0.120
three	No	yes	25:44	-0.503	1.065	0.872	0.088	0.421	0.347	0.145
three	No	yes	>=75	-1.434	0.368	0.665	0.052	0.312	0.420	0.216
three	Yes	no	< 15	-2.110	0.087	0.323	0.034	0.304	0.384	0.278
three	Yes	no	25:44	-2.274	-0.235	0.246	0.032	0.249	0.403	0.315
three	Yes	no	>=75	-3.204	-0.932	0.039	0.016	0.159	0.420	0.404
three	Yes	yes	< 15	-1.357	0.470	0.649	0.054	0.335	0.401	0.210
three	Yes	yes	25:44	-1.521	0.148	0.572	0.053	0.279	0.427	0.241
three	Yes	yes	>=75	-2.452	-0.549	0.365	0.028	0.186	0.464	0.322

Estimated population-average probabilities

The population-average probabilities are obtained by dividing the ETA1, ETA2 and ETA3 values given in the previous two tables by the square root of the corresponding design effects. For the intercepts-only model, this quantity is obtained as

$$\hat{d}_c = \left[\operatorname{var}\left(v_{ij00}\right) + \operatorname{var}\left(v_{ic0}\right) + \operatorname{var}\left(e_{ijk}\right)\right] / \operatorname{var}\left(e_{ijk}\right), \quad c = 1,2,3.$$

For the logistic model it is assumed that

$$\operatorname{var}\left(e_{ijk}\right) = \frac{\pi^2}{3} = 3.290.$$

Therefore

$$\begin{aligned}\sqrt{d_1} &= \sqrt{\left(4.707 + 1.009 + 3.290\right) / 3.290} \\ &= \sqrt{2.737} \\ &= 1.6545.\end{aligned}$$

Similarly,

$$\sqrt{d_2} = \sqrt{(3.237 + 0.921 + 3.290)/3.290}$$

$$= \sqrt{2.264}$$

$$= 1.5046$$

and

$$\sqrt{d_3} = \sqrt{(1.077 + 0.848 + 3.290)/3.290}$$

$$= \sqrt{1.585}$$

$$= 1.2590.$$

Using these values, we obtain the population-average probabilities for the four categories of PASTVIS for a female respondent. Summaries of a selected number of population-average probabilities are given in the tables below.

Population-average probabilities for females with XRAY = no, INJURY = no, URINE = no, and BLODPRES = no

NUMMED	PRIM CARE	EXER CISE	AGE	ETA1	ETA2	ETA3	PROB1	PROB2	PROB3	PROB4
none	No	no	< 15	-0.245	0.926	0.499	0.131	0.424	0.277	0.168
none	No	no	25:44	-0.344	0.712	0.438	0.134	0.385	0.293	0.189
none	No	no	>= 75	-0.907	0.248	0.273	0.101	0.320	0.329	0.250
none	No	yes	< 15	0.210	1.180	0.757	0.162	0.427	0.280	0.131
none	No	yes	25:44	0.111	0.966	0.696	0.165	0.389	0.297	0.148
none	No	yes	>= 75	-0.452	0.503	0.532	0.128	0.331	0.341	0.200
none	Yes	no	< 15	-0.860	0.317	0.260	0.103	0.335	0.317	0.244
none	Yes	no	25:44	-0.960	0.102	0.199	0.103	0.299	0.329	0.269
none	Yes	no	>= 75	-1.522	-0.361	0.035	0.074	0.236	0.351	0.339
none	Yes	yes	< 15	-0.405	0.571	0.519	0.130	0.346	0.328	0.195
none	Yes	yes	25:44	-0.504	0.357	0.457	0.131	0.310	0.343	0.217
none	Yes	yes	>= 75	-1.067	-0.106	0.293	0.096	0.251	0.374	0.279
three	No	no	< 15	-0.847	0.455	0.328	0.098	0.359	0.316	0.228
three	No	no	25:44	-0.946	0.241	0.267	0.098	0.321	0.329	0.252
three	No	no	>= 75	-1.509	-0.222	0.102	0.071	0.256	0.354	0.320
three	No	yes	< 15	-0.392	0.710	0.586	0.123	0.369	0.326	0.182
three	No	yes	25:44	-0.491	0.496	0.525	0.124	0.332	0.342	0.202
three	No	yes	>= 75	-1.054	0.032	0.360	0.091	0.271	0.376	0.262
three	Yes	no	< 15	-1.462	-0.154	0.089	0.073	0.269	0.344	0.314
three	Yes	no	25:44	-1.562	-0.368	0.028	0.072	0.236	0.351	0.341
three	Yes	no	>= 75	-2.124	-0.831	-0.136	0.049	0.179	0.359	0.412
three	Yes	yes	< 15	-1.007	0.101	0.347	0.094	0.285	0.364	0.257
three	Yes	yes	25:44	-1.107	-0.114	0.286	0.093	0.251	0.375	0.281
three	Yes	yes	>= 75	-1.669	-0.577	0.122	0.065	0.195	0.392	0.347

Population-average probabilities for females with XRAY = no, INJURY = no, URINE = no, and BLODPRES = no

NUMMED	PRIM CARE	EXER CISE	AGE	ETA1	ETA2	ETA3	PROB1	PROB2	PROB3	PROB4
none	No	no	< 15	-0.058	1.138	0.667	0.135	0.445	0.278	0.143
none	No	no	25:44	-0.157	0.924	0.605	0.138	0.406	0.295	0.161
none	No	no	>= 75	-0.720	0.460	0.441	0.105	0.343	0.336	0.216
none	No	yes	< 15	0.397	1.392	0.925	0.165	0.446	0.279	0.111
none	No	yes	25:44	0.298	1.178	0.864	0.169	0.408	0.298	0.126
none	No	yes	>= 75	-0.265	0.715	0.699	0.132	0.351	0.346	0.172
none	Yes	no	< 15	-0.673	0.529	0.428	0.108	0.358	0.324	0.211
none	Yes	no	25:44	-0.772	0.314	0.367	0.108	0.320	0.338	0.234
none	Yes	no	>= 75	-1.335	-0.149	0.202	0.079	0.257	0.366	0.299
none	Yes	yes	< 15	-0.218	0.783	0.686	0.134	0.366	0.332	0.167
none	Yes	yes	25:44	-0.317	0.569	0.625	0.136	0.329	0.348	0.186
none	Yes	yes	>= 75	-0.880	0.106	0.461	0.101	0.270	0.386	0.243
Three	No	no	< 15	-0.660	0.667	0.496	0.101	0.382	0.321	0.196
Three	No	no	25:44	-0.759	0.453	0.434	0.102	0.343	0.337	0.218
Three	No	no	>= 75	-1.322	-0.010	0.270	0.075	0.278	0.367	0.280
Three	No	yes	< 15	-0.205	0.922	0.754	0.126	0.390	0.329	0.155
Three	No	yes	25:44	-0.304	0.708	0.693	0.128	0.352	0.347	0.173
Three	No	yes	>= 75	-0.867	0.244	0.528	0.096	0.291	0.386	0.228
Three	Yes	no	< 15	-1.275	0.058	0.257	0.077	0.292	0.356	0.275
Three	Yes	no	25:44	-1.374	-0.156	0.196	0.076	0.257	0.366	0.301
Three	Yes	no	>= 75	-1.937	-0.619	0.031	0.053	0.198	0.380	0.368
Three	Yes	yes	< 15	-0.820	0.313	0.515	0.098	0.305	0.374	0.223
Three	Yes	yes	25:44	-0.919	0.098	0.454	0.098	0.271	0.386	0.245
Three	Yes	yes	>= 75	-1.482	-0.365	0.290	0.070	0.213	0.410	0.307

8. A random intercept model with ten predictors

In the previous example, we included 14 possible predictors of PASTVIS in the fixed part of the model. The output indicated that the variables GENDER, INJURY, URINE and XRAY did not contribute significantly to explaining the variation in PASTVIS outcomes.

To run the model without these fixed effects, use the **File**, **Open Syntax File** option and select the command syntax previously saved to the file **NHIS1.prl**. Delete the variables GENDER, INJURY, URINE and XRAY from the Predictors paragraph and save the modified syntax file as **NHIS2.prl**. To run this syntax file, select the **Run** option from the **Analysis** menu.

9. Interpreting the output

Fit statistics

Only a portion of the output file **NIHS2.out** is shown below. Recall that the deviance statistic for the previous model was 14329.75, with 48 free parameters. For the current model, the deviance statistic is equal to 14351.80 and the number of free parameters is equal to 36. To test whether the removal of GENDER, INJURY, URINE and XRAY made a significant difference to the model fit, we use the fact that the difference in deviance statistics for two nested models follows a χ^2-distribution with degrees of freedom equal to the difference in the number of parameters estimated.

The χ^2-value obtained for this test is 14351.80 – 14329.75 = 22.05, with 12 degrees of freedom. Since the associated p-value equals 0.04, the χ^2-value is significant at the 5% level, but not at the 1% level of significance. We therefore conclude that, based on the χ^2-difference test, we do not have a definitive answer to the question of whether the 4 predictors should remain in the model or not. A summary of the Akaike and Schwarz criteria is shown in Table 14.

Table 14: Akaike and Schwarz fit criteria for two nested models

Fit statistic	14 predictors	10 predictors
Akaike	14425.75	14423.80
Schwarz	14750.75	14667.55

Each of these criteria states that the model with the smallest value is the model to be selected. Based on this decision rule, we conclude that the model without the four predictors should be used, since it is more parsimonious and very little information regarding the explanation of variation in PASTVIS is lost.

Odds ratios and 95% confidence intervals for the odds ratios

An odds ratio of 1 indicates the event under study is equally likely in both the outcome category of interest and in the reference category. An odds ratio greater than 1 indicates that the event is more likely to occur in the category of interest.

```
nihs2.OUT
     Response Code  1 vs Code   4
     ----------------------------
     intcept          -1.4149        0.2429     0.1525     0.3871
     NUMMED           -0.3414        0.7108     0.6517     0.7753
     PRIMCARE         -1.0147        0.3625     0.2490     0.5278
     BLODPRES          0.3338        1.3963     1.0033     1.9431
     EXERCISE          0.7986        2.2225     1.4941     3.3060
     AGE1              1.0546        2.8707     1.7234     4.7817
     AGE2              1.3594        3.8940     2.1604     7.0186
     AGE3              0.9288        2.5316     1.6309     3.9296
     AGE4              0.7429        2.1021     1.3765     3.2101
     AGE5              0.0656        1.0678     0.6491     1.7566

     Response Code  2 vs Code   4
     ----------------------------
     intcept           0.4636        1.5898     1.1447     2.2080
     NUMMED           -0.2382        0.7881     0.7456     0.8329
     PRIMCARE         -0.9058        0.4042     0.3057     0.5345
     BLODPRES          0.2925        1.3398     1.0492     1.7108
     EXERCISE          0.3891        1.4757     1.1091     1.9635
     AGE1              0.9978        2.7122     1.9344     3.8027
     AGE2              1.2602        3.5261     2.3691     5.2483
     AGE3              0.6881        1.9899     1.4823     2.6713
     AGE4              0.6576        1.9302     1.4647     2.5435
     AGE5              0.0956        1.1003     0.8053     1.5033
```

The intercept coefficient is the expected log-odds that a participant in the present study indicated no past visits (PASTVIS = 1) relative to the category PASTVIS = 4 (6 or more visits), given that the remaining predictors are held constant at zero. The estimated conditional expected log-odds is –1.4156, corresponding to an odds ratio of exp(–1.4156)=0.2428. This implies that a qualifying participant (a participant with NUMMED = 0, GENDER = 0, …, AGE5 = 0) has 0.2427 times the odds of having had no previous visits, as opposed to 6 or more visits.

The 95% confidence interval for the odds ratio is obtained by first computing a 95% confidence interval for the intercept coefficient. This confidence interval is given by

$$\hat{\beta}_0 \pm 1.96\, std.error\left(\hat{\beta}_0\right).$$

From the output, it follows that this interval is

$$\begin{aligned}&\left(-1.4161-1.96\times 0.2377; -1.4161+1.96\times 0.2377\right)\\&=\left(-1.8822; -0.9500\right).\end{aligned}$$

Using these values, we obtain the 95% confidence interval for the odds ratio as

$$\begin{aligned}&\left(\exp\left(-1.8822\right); \exp(-0.9500)\right)\\&=\left(0.1523; 0.3867\right).\end{aligned}$$