



Two-level models for nominal outcomes

Contents

3.1	MODELS FOR THE NHIS DATA	1
3.1.1	<i>The data</i>	2
3.1.2	<i>The model</i>	4
3.1.3	<i>A random intercept model with fourteen predictors</i>	5
3.1.4	<i>A random intercept model with ten predictors</i>	24

3.1 Models for the NHIS data

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

3.1.1 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 SuperMix spreadsheet file, named **nih_recoded.ss3**.

	(A) CSTRATM	(B) CPSUM	(C) PATWT	(D) PASTVIS	(E) NUMMED	(F) GENDER	(G) USETOBA
1	20102101	100013	50245	3	2	1	1
2	20102101	100013	50245	3	2	1	0
3	20102101	100013	50245	3	4	1	0
4	20102101	100013	50245	3	2	1	0
5	20102101	100013	50245	2	1	0	0
6	20102101	100015	72581	4	0	1	0
7	20102101	100015	72581	3	2	1	0
8	20102101	100015	72581	3	2	0	0
9	20102101	100015	72581	1	1	0	0
10	20102101	100015	72581	3	0	1	0

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 7.1: 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

3.1.2 The model

3.1.2.1 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 $1, 2, 3, \dots, C$.

Let $P_{ijk} = P(y_{ijk} = c | \boldsymbol{\beta}_c, \mathbf{v}_{ic}, \mathbf{v}_{ijc})$ 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 \mathbf{v}_{ijc} and the $(r \times 1)$ vector of level-3 random effects \mathbf{v}_{ic} . It is further assumed that the level-2 random effects \mathbf{v}_{ijc} are independent and identically distributed (i.i.d.) as a $N(\mathbf{0}, \boldsymbol{\Phi}_{(2)})$ random variable. Uncorrelated with \mathbf{v}_{ijc} , the level-3 random effects are i.i.d. $N(\mathbf{0}, \boldsymbol{\Phi}_{(3)})$. The scalar y_{ijk} denotes the value of the nominal variable associated with level-1 unit k , $k = 1, 2, \dots, n_{ij}$, nested within level-2 unit j , $j = 1, 2, \dots, n_i$, which in turn is nested within the i -th level-3 unit, where $i = 1, 2, \dots, N$. The probabilities P_{ijk} are computed as

$$\begin{aligned} P_{ijk} &= P(y_{ijk} = c | \boldsymbol{\beta}_c, \mathbf{v}_{ic}, \mathbf{v}_{ijc}) \\ &= \frac{\exp(\eta_{ijk})}{1 + \sum_{h=1}^{C-1} \exp(\eta_{ijk})}, \quad c = 1, 2, \dots, C-1 \end{aligned}$$

where

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

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.

3.1.2.2 Random intercept model with two explanatory variables

For the `nihs_recoded.ss3` 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

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

where

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

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

SuperMix 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

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

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

3.1.3 A random intercept model with fourteen predictors

3.1.3.1 Preparing the data

The model to be fitted to the data is contained in **nihs_recoded.ss3**. This file was created from the SPSS data file **nihs_recoded.sav** as follows.

Use the **File, Import Data File** option to activate the display of an **Open** dialog box. Browse for the file **nihs_recoded.sav** in the **Examples\Nominal** folder. Select the file and click the **Open** button to display **nihs_recoded.ss3**.

3.1.3.2 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 **Data-based Graphs** option. Prior to making visual presentations, it is a good idea to assign labels to the categories of the nominal and ordinal variables. To illustrate, right click on the PASTVIS header and

select the **Column Properties** option to display the **Column Properties** dialog. Select the **Nominal** or **Ordinal** option to obtain the list of values assigned to the categories of PASTVIS.

(C) PATWT	(D) PASTVIS	(E) GENDER	(F) AGER
50245	3		
50245	3		
50245	3		
50245	3		
50245	2		
72581	4		
72581	3		

Enter the labels None, 1 to 2, 3 to 5 and 6 and more as shown below and click **OK**.

Column Properties

Header: PASTVIS

Number of Decimal Places: 0

Categorical Continuous

Create dummies when used as a predictor

	Value	Label
1	1	None
2	2	1 to 2
3	3	3 to 5
4	4	6 and more

OK Cancel

Category labels can be assigned to the variables GENDER and AGER in a similar way. For example, right-click on the AGER header and enter labels as shown below.

Column Properties

Header: AGER

Number of Decimal Places: 0

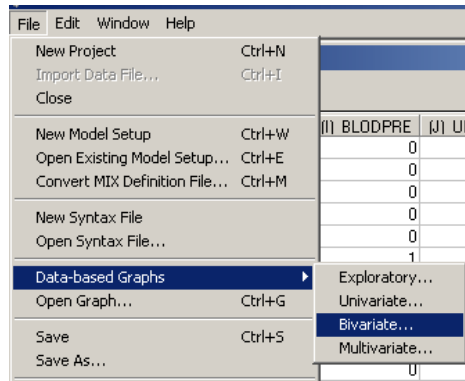
Categorical Continuous

Create dummies when used as a predictor

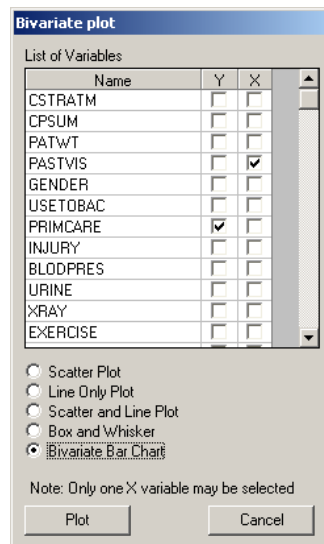
	Value	Label
1	1	Under 15
2	2	15 to 24
3	3	25 to 44
4	4	45 to 64

OK Cancel

Remember to use the **File, Save** option to permanently store the labeling information. From the main menu bar, select the **File, Data-based Graphs, Bivariate** option.



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



Next, click the **Bivariate Bar Chart** radio button and then the **Plot** button to obtain the bivariate bar chart of PRIMCARE vs 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.

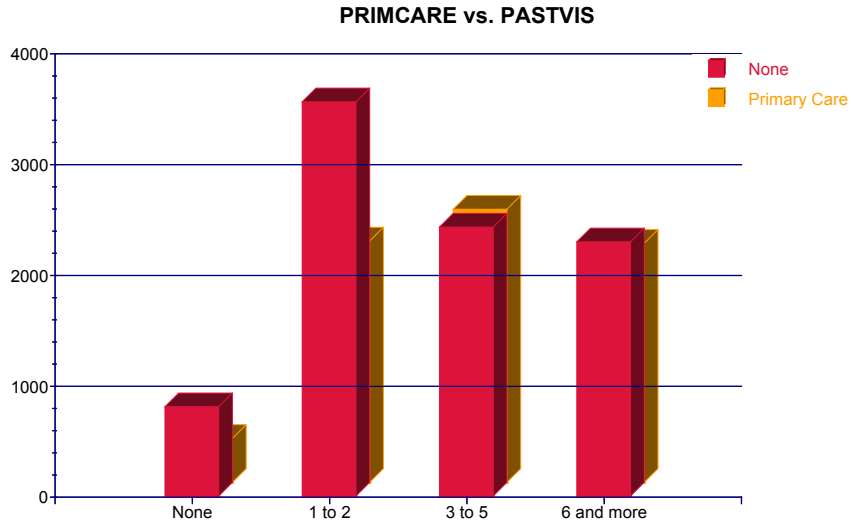


Figure 7.1: Bivariate bar chart of PRIMCARE vs PASTVIS

To investigate the relationship between the variables BLODPRES and AGER, select BLODPRES as the Y variable and AGER as the X variable.

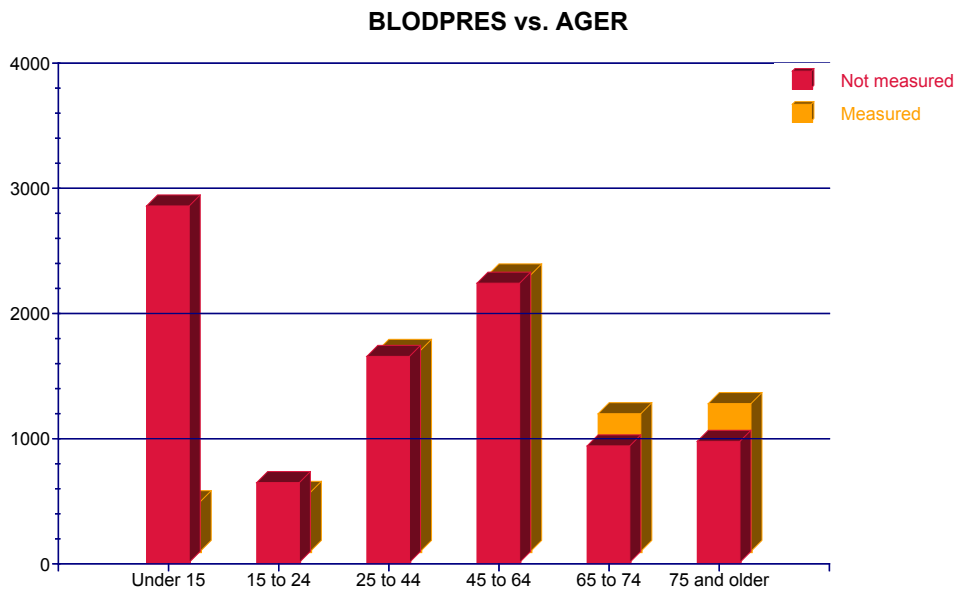


Figure 7.2: Bivariate bar chart of BLODPRES vs AGER

From Figure 7.2 it is evident that it is much more likely that blood pressure will be measured for patients older than 45 than for younger patients.

3.1.3.3 Setting up the analysis

From the main menu bar, select the **File, New Model Setup** option. In this example, only the **Configuration, Variables** and **Advanced** tabs of the **Model Setup** window are used. By default, the **Configuration** tab is displayed first.

Start by providing titles for the analysis in the **Title 1** and **Title 2** text boxes. Next, select the outcome variable PASTVIS from the **Dependent Variable** drop-down list box and indicate the type of outcome as nominal using the **Dependent Variable Type** drop-down list box. Select CPSUM and CSTRATM as the **level-2** and **level-3 ID** variables and choose **means & (co)variances** as the **Write Bayes Estimates** option. Note that the current data set contains no missing values.

Model Setup: NIH51.mum

Configuration | Variables | Starting Values | Patterns | Advanced | Linear Transforms

Title 1: National Health Interview Survey Data

Title 2: Level-3 and Level-2 Random Intercepts

Dependent Variable Type: nominal

Level-2 ID: CPSUM

Dependent Variable: PASTVIS

Level-3 ID: CSTRATM

Categories:

	Value
1	1
2	2
3	3
4	4

Write Bayes Estimates: means & (co)variances

Convergence Criterion: 0.0001

Number of Iterations: 100

Missing Values Present: false

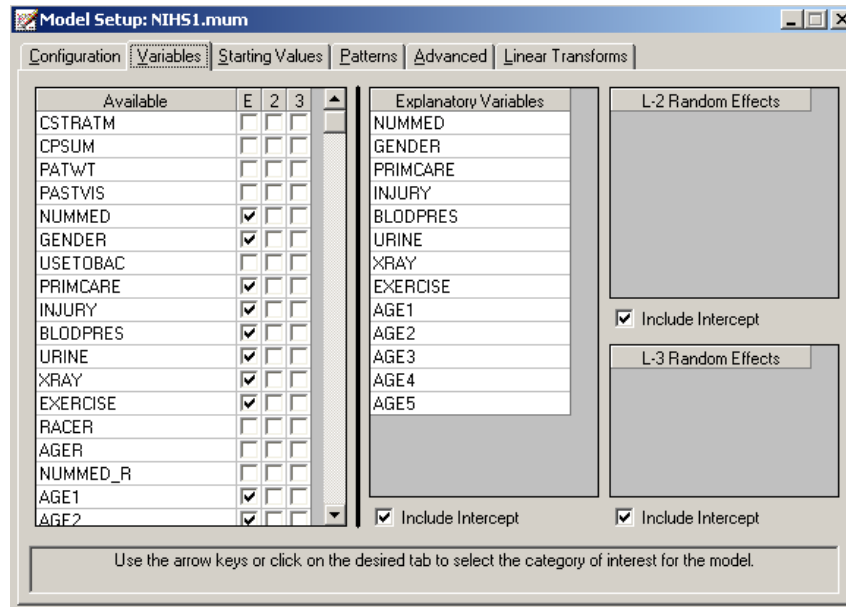
Perform Crosstabulation: no

Output Type: standard

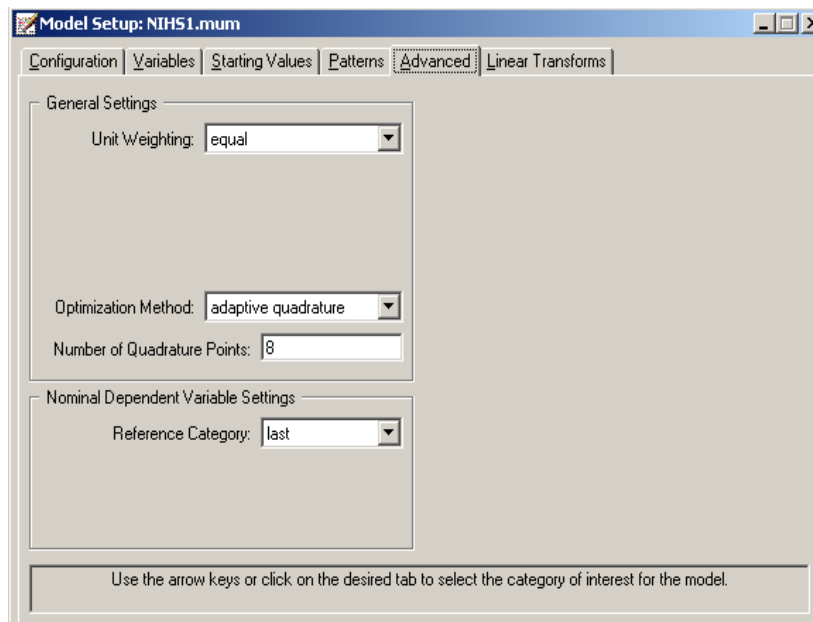
Use the arrow keys or click on the desired tab to select the category of interest for the model.

The **Variables** tab is used to specify the fixed and random effects to be included in the model. Start by selecting the explanatory (fixed) variables, using the check boxes in the **Available** grid. The image below shows the completed selection of all predictors.

By default, the inclusion of a fixed intercept coefficient and random intercepts at level 2 and level 3 is assumed. As these selections correspond to the model we intend fitting to the data, no further changes are made on this tab. Click on the **Advanced** tab and request adaptive quadrature as the **Optimization Method** and enter the number 8 as the **Number of Quadrature Points**.



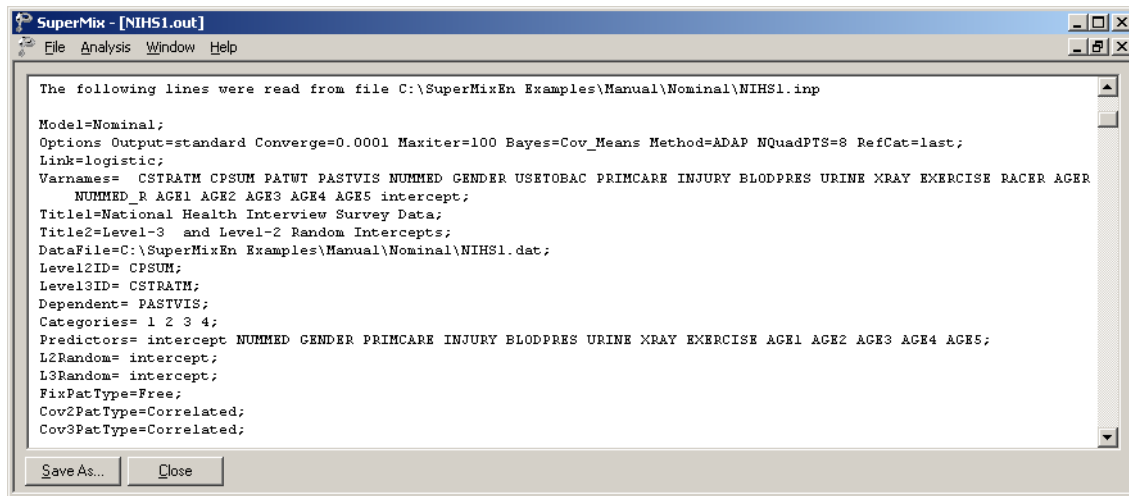
Before running the analysis, save the model specification using the **File, Save** option from the main menu bar. Provide a name for the model specification file, for example **NIHS1.mum**, and then run the analysis using the **Analysis, Run** option.



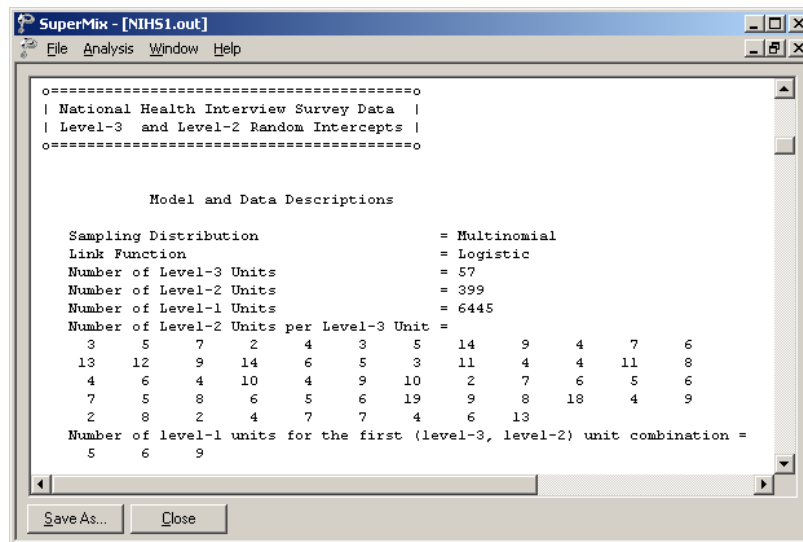
3.1.3.4 Discussion of results

The command syntax generated by the graphical user interface is saved to the file **NIHS1.inp**. This file can be edited by using the **File, Open Syntax File** option. For example, the predictors

GENDER, BLODPRES, URINE and XRAY can be removed, after which the syntax file can be saved as **NIHS2.inp**. To run **NIHS2.inp**, use the **Analysis, Run** option as before.



Model and data description



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.

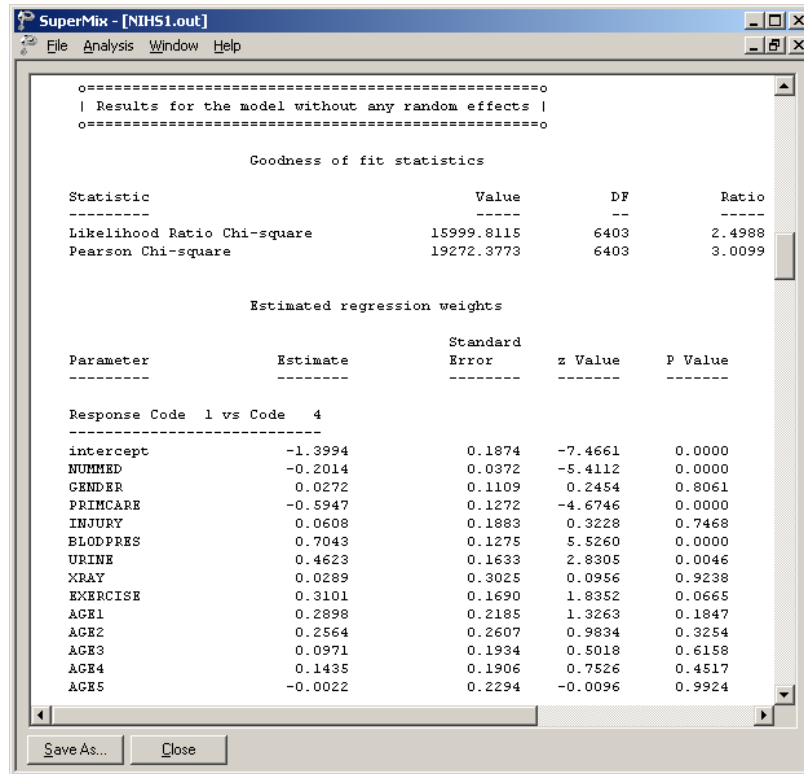
Variable	Minimum	Maximum	Mean	Standard Deviation
PASTVIS1	0.0000	1.0000	0.0692	0.2538
PASTVIS2	0.0000	1.0000	0.3677	0.4822
PASTVIS3	0.0000	1.0000	0.2858	0.4518
PASTVIS4	0.0000	1.0000	0.2773	0.4477
intercept	1.0000	1.0000	1.0000	0.0000
NUMMED	0.0000	6.0000	1.5286	1.6863
GENDER	0.0000	1.0000	0.4196	0.4935
PRIMCARE	0.0000	1.0000	0.3992	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	0.3054
XRAY	0.0000	1.0000	0.0380	0.1912
EXERCISE	0.0000	1.0000	0.1047	0.3062
AGE1	0.0000	1.0000	0.1966	0.3974
AGE2	0.0000	1.0000	0.0644	0.2455
AGE3	0.0000	1.0000	0.2284	0.4198
AGE4	0.0000	1.0000	0.2667	0.4423
AGE5	0.0000	1.0000	0.1156	0.3198

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 7.2.

Table 7.2: Age distribution of respondents

Age	Percentage
Younger than 15	19.7
15 to 24	6.4
25 to 44	22.8
45 to 64	26.7
65 to 74	11.6
75 and older	12.8*

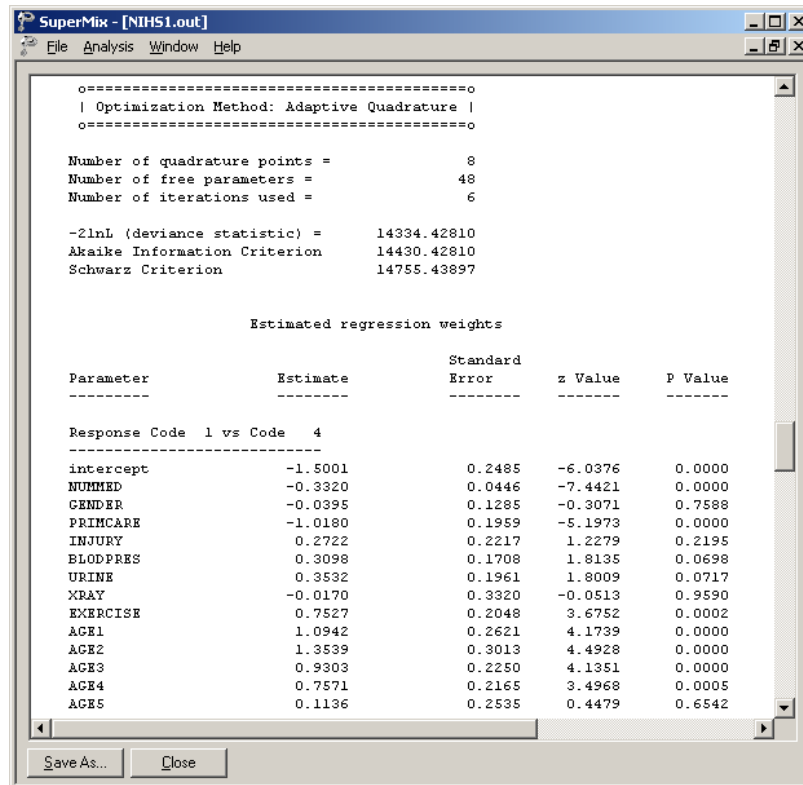
*: calculated as $100 - (19.7 + 6.4 + 22.8 + 26.7 + 11.6)$



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, 5 iterations were required to reach convergence. The deviance statistic ($-2\ln L$) allows the user to compare the current model with other nested models.



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.

SuperMix - [NIHS1.out]

File Analysis Window Help

Response Code 2 vs Code 4

Parameter	Estimate	Standard Error	z Value	P Value
intercept	0.3740	0.1731	2.1604	0.0307
NUMMED	-0.2360	0.0285	-8.2797	0.0000
GENDER	0.0900	0.0842	1.0680	0.2855
PRIMCARE	-0.9170	0.1456	-6.2977	0.0000
INJURY	0.2307	0.1503	1.5347	0.1249
BLODPRES	0.3190	0.1254	2.5427	0.0110
URINE	-0.0137	0.1448	-0.0943	0.9249
XRAY	0.2636	0.2075	1.2703	0.2040
EXERCISE	0.3829	0.1472	2.6020	0.0093
AGE1	1.0191	0.1737	5.8665	0.0000
AGE2	1.2608	0.2030	6.2122	0.0000
AGE3	0.6969	0.1500	4.6465	0.0000
AGE4	0.6645	0.1409	4.7150	0.0000
AGE5	0.1072	0.1595	0.6722	0.5014

Response Code 3 vs Code 4

Parameter	Estimate	Standard Error	z Value	P Value
intercept	0.3444	0.1570	2.1928	0.0283
NUMMED	-0.0718	0.0249	-2.8840	0.0039
GENDER	0.0083	0.0801	0.1040	0.9172
PRIMCARE	-0.3006	0.1197	-2.5111	0.0120
INJURY	-0.0516	0.1449	-0.3560	0.7218
BLODPRES	0.2112	0.1126	1.8757	0.0607
URINE	-0.1531	0.1416	-1.0814	0.2795
XRAY	-0.0242	0.2025	-0.1196	0.9048
EXERCISE	0.3251	0.1345	2.4175	0.0156
AGE1	0.2837	0.1552	1.8274	0.0676
AGE2	0.4731	0.1969	2.4024	0.0163
AGE3	0.2068	0.1376	1.5029	0.1329
AGE4	0.1961	0.1287	1.5240	0.1275
AGE5	-0.1325	0.1458	-0.9093	0.3632

Save As... Close

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.

SuperMix - [NIHS1.out]

File Analysis Window Help

Estimated level 2 variances and covariances

Parameter	Estimate	Standard Error	z Value	P Value
intercept1/intercept1	4.7074	0.4313	10.9144	0.0000
intercept2/intercept2	3.2371	0.2784	11.6272	0.0000
intercept3/intercept3	1.0766	0.1381	7.7937	0.0000

Save As... Close

SuperMix - [NIHS1.out]

File Analysis Window Help

Estimated level 3 variances and covariances

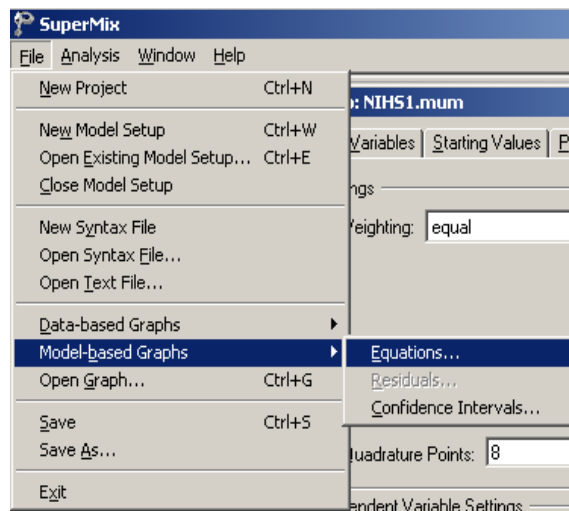
Parameter	Estimate	Standard Error	z Value	P Value
intercept1/intercept1	1.0104	0.2522	4.0058	0.0001
intercept2/intercept2	0.9219	0.1878	4.9088	0.0000
intercept3/intercept3	0.8488	0.1525	5.5663	0.0000

Save As... Close

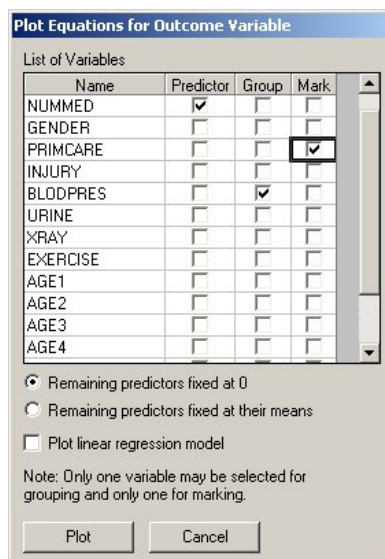
3.1.3.5 Interpreting the output

Model-based graphs

The **Model-based Graphs** option is available via the **File** menu if either the output window or the **Model Setup** window is active.



Select the **Equations** option to open the **Plot Equations** dialog box. Select NUMMED as the predictor, BLODPRES as the **Group** variable, and **Mark** each plot by the categories of PRIMCARE.



Click **Plot** to obtain the plot of the logit of PASTVIS against NUMMED for category 1 versus category 4.

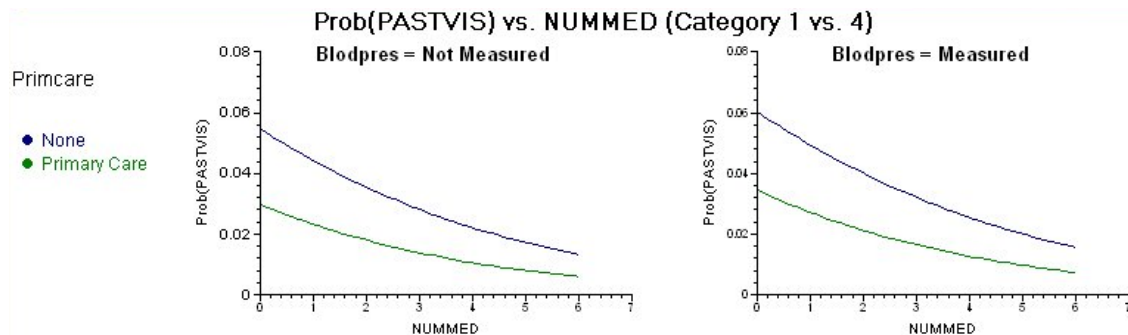


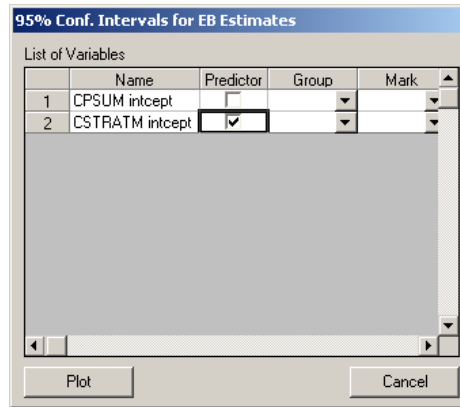
Figure 7.3: Plot of Prob(PASTVIS) vs. NUMMED

Similar plots are obtained for category 2 vs. 3 and category 3 vs. 4 by clicking the slider (not shown) beneath the graphical displays in Figure 7.3. Figure 7.3 shows smaller probabilities for the category Primary Care (lower of the lines) compared to the category none (uppermost lines) of the variable PRIMCARE.

Confidence intervals for EB estimates

In addition to the estimate of the regression coefficients in the fixed part of the model and the estimates of the variances of the random intercept coefficients, we also obtain estimates for the unique deviations from the estimated population intercept coefficients at levels 2 and 3. These estimates, also known as empirical Bayes residuals, are written to **NIHS1_res.bay2** and **NIHS1_res.bay3**. these files can be imported into SuperMix as ***.ss3** spreadsheet files. Using the **File, Model-Based Graphs, Confidence Intervals** option, a graphical display of 95% confidence intervals can be obtained for both level-2 and level-3 units.

In the **95% Confidence Intervals for EB Estimates** dialog box shown below, the variable CSTRATM (level-3 ID) is selected. Click **Plot** to obtain a visual presentation of the confidence intervals.



Each confidence interval is calculated as $\hat{v}_{0i} \pm 1.96\hat{\sigma}_{v_{0i}}$, where $\hat{\sigma}_{v_{0i}}$ is the square root of the variance of \hat{v}_{0i} . Figure 7.4 shows that the estimated deviances from $\hat{\beta}_0$ exhibit a reasonable amount of variation across the level-3 units. It also indicates substantial variation in the estimated EB variances across the level-3 units.

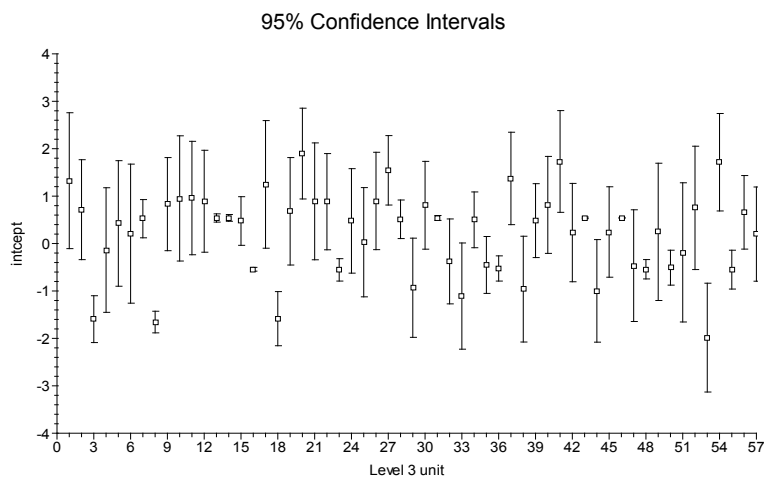


Figure 7.4: 95% Confidence intervals for level-3 EB estimates

To obtain the list of lower and upper confidence values and of the estimated EB residuals, double-click with the left mouse button anywhere in the graph area.

Unit	Lower Conf	Estimate	Upper Conf
1	-0.10679	1.3257311	2.7583
2	-0.33892	0.7145276	1.768
3	-2.0872	-1.5939368	-1.1006
4	-1.45	-0.1362415	1.1775
5	-0.89933	0.4244613	1.7482
6	-1.2565	0.2092281	1.6749
7	0.12184	0.5242952	0.92675
8	-1.8828	-1.6546821	-1.4265
9	-0.14985	0.8318654	1.8136
10	-0.36842	0.9517006	2.2718
11	-0.23608	0.9601461	2.1564
12	-0.18239	0.8925091	1.9674
13	0.45277	0.5397403	0.62671
14	0.47023	0.5411473	0.61207
15	-0.036827	0.4755487	0.98792

From the display above, it follows that the 95% confidence interval for unit 3, for example, is given by $(-2.087, -1.101)$.

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:

$$\begin{aligned}
 \hat{\eta}_{ijk1} &= -1.5004 - 0.3320(NUMMED_{ijk}) - 0.0395(GENDER_{ijk}) - 1.0176(PRIMCARE_{ijk}) \\
 &\quad + \dots + 1.0945(AGE1_{ijk}) + 1.3539(AGE2_{ijk}) + 0.9306(AGE3_{ijk}) \\
 &\quad + 0.7572(AGE4_{ijk}) + 0.1136(AGE5_{ijk}) \\
 &= -1.5004 - 3(0.3320) - 1(1.0176) + 1(0.9306) \\
 &= -2.5834
 \end{aligned}$$

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

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

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

Table 7.3: Unit-specific probabilities for females with XRAY = no, INJURY = no, URINE = no, and BLODPRES = no

NUMMED	PRIM CARE	EXERCISE	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.

Table 7.4: Unit-specific probabilities for females with XRAY = no, INJURY = no, URINE = no, and BLODPRES = yes

NUMMED	PRIM CARE	EXERCISE	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 Tables 7.3 and 7.4 by the square root of the corresponding design effects. For the intercepts-only model, this quantity is obtained as

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

For the logistic model it is assumed that

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

Therefore

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

Similarly,

$$\begin{aligned}\sqrt{d_2} &= \sqrt{(3.237 + 0.921 + 3.290) / 3.290} \\ &= \sqrt{2.264} \\ &= 1.5046\end{aligned}$$

and

$$\begin{aligned}\sqrt{d_3} &= \sqrt{(1.077 + 0.848 + 3.290) / 3.290} \\ &= \sqrt{1.585} \\ &= 1.2590.\end{aligned}$$

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 Tables 7.5 and 7.6 below.

Table 7.5: Population-average probabilities for females with XRAY = no, INJURY = no, URINE = no, and BLODPRES = no

NUMMED	PRIM CARE	EXERCISE	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

Table 7.6: Population-average probabilities for females with XRAY = no, INJURY = no, URINE = no, and BLODPRES = no

NUMMED	PRIM CARE	EXERCISE	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

3.1.4 A random intercept model with ten predictors

3.1.4.1 Setting up the analysis

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.

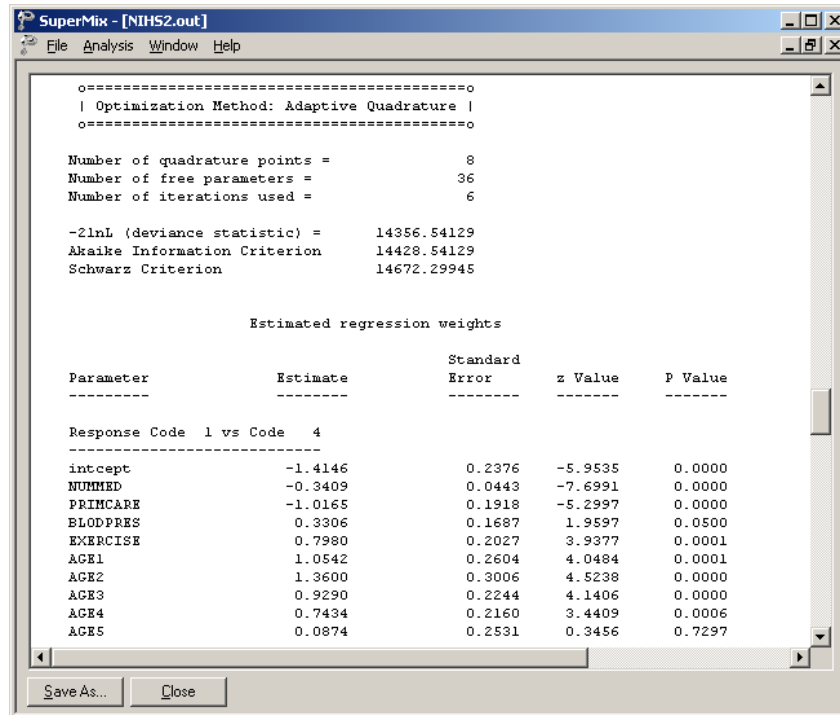
To run the model without these fixed effects, use the **File, Open Syntax File** option and select the command syntax previously saved to the file **nihs1.inp**. Delete the variables GENDER, INJURY, URINE and XRAY from the Predictors paragraph and save the modified syntax file as **nihs2.inp**. To run this syntax file, select the **Run** option from the **Analysis** menu.


```
Model=Nominal;
Options Output=standard Converge=0.0001 Maxiter=100 Bayes=Cov_Means Method=ADAP NQuadPTS=8 RefCat=last;
Link=logistic;
Varnames= CSTRATM CPSUM PATWT PASTVIS NUMMED GENDER USETOBAC PRIMCARE INJURY BLODPRES URINE XRAY EXERCISE RACER AGER
          NUMMED_R AGE1 AGE2 AGE3 AGE4 AGE5 intercept;
Title1=National Health Interview Survey Data;
Title2=Level-3 and Level-2 Random Intercepts;
DataFile=C:\SuperMixEx Examples\Manual\Nominal\NIHS1.dat;
Level2ID= CPSUM;
Level3ID= CSTRATM;
Dependent= PASTVIS;
Categories= 1 2 3 4;
Predictors= intercept NUMMED PRIMCARE BLODPRES EXERCISE AGE1 AGE2 AGE3 AGE4 AGE5;
L2Random= intercept;
L3Random= intercept;
FixPatType=Free;
Cov2PatType=Correlated;
Cov3PatType=Correlated;
```

3.1.4.2 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 14334.43, with 48 free parameters. For the current model, the deviance statistic is equal to 14356.54 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 $14356.34 - 14334.43 = 21.91$, with 12 degrees of freedom. Since the associated p -value equals 0.0385, 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 7.7.

Table 7.7: Akaike and Schwarz fit criteria for two nested models

Fit statistic	14 predictors	10 predictors
Akaike	14430.43	14428.54
Schwarz	14755.44	14672.30

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.

SuperMix - [NIHS2.out]

File Analysis Window Help

Response Code 2 vs Code 4

Parameter	Estimate	Standard Error	z Value	P Value
intcept	0.4640	0.1675	2.7706	0.0056
NUMMED	-0.2383	0.0283	-8.4299	0.0000
PRIMCARE	-0.9067	0.1424	-6.3668	0.0000
BLODPRES	0.2945	0.1250	2.3568	0.0184
EXERCISE	0.3895	0.1457	2.6728	0.0075
AGE1	0.9974	0.1725	5.7835	0.0000
AGE2	1.2599	0.2029	6.2098	0.0000
AGE3	0.6874	0.1502	4.5759	0.0000
AGE4	0.6574	0.1408	4.6697	0.0000
AGE5	0.0952	0.1592	0.5977	0.5500

Response Code 3 vs Code 4

Parameter	Estimate	Standard Error	z Value	P Value
intcept	0.3374	0.1514	2.2284	0.0259
NUMMED	-0.0693	0.0246	-2.8155	0.0049
PRIMCARE	-0.3062	0.1174	-2.6084	0.0091
BLODPRES	0.1940	0.1124	1.7253	0.0845
EXERCISE	0.3043	0.1334	2.2809	0.0226
AGE1	0.2898	0.1537	1.8855	0.0594
AGE2	0.4733	0.1966	2.4071	0.0161
AGE3	0.2069	0.1376	1.5039	0.1326
AGE4	0.1978	0.1286	1.5374	0.1242
AGE5	-0.1277	0.1456	-0.8767	0.3806

Save As... Close

The next portion of the output contains estimates of the variances of the level-2 and level-3 random effects for each response code relative to the reference code. From the output it is evident that there are highly significant deviances from the estimated population intercepts at both level-2 and level-3.

SuperMix - [NIHS2.out]

File Analysis Window Help

Estimated level 2 variances and covariances

Parameter	Estimate	Standard Error	z Value	P Value
intercept1/intercept1	4.6577	0.4312	10.8016	0.0000
intercept2/intercept2	3.2423	0.2826	11.4727	0.0000
intercept3/intercept3	1.0897	0.1408	7.7406	0.0000

Save As... Close

SuperMix - [NIHS2.out]

File Analysis Window Help

Estimated level 3 variances and covariances

Parameter	Estimate	Standard Error	z Value	P Value
intcept1/intcept1	1.0109	0.2424	4.1708	0.0000
intcept2/intcept2	0.9627	0.1840	5.2330	0.0000
intcept3/intcept3	0.8552	0.1472	5.8092	0.0000

Save As... Close

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.

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.

Parameter	Estimate	Odds Ratio	Bounds	
			Lower	Upper
Response Code 1 vs Code 4				
intcept	-1.4146	0.2430	0.1525	0.3872
NUMMED	-0.3409	0.7111	0.6520	0.7756
PRINCARE	-1.0165	0.3619	0.2485	0.5270
BLODPRES	0.3306	1.3918	0.9999	1.9372
EXERCISE	0.7980	2.2212	1.4930	3.3045
AGE1	1.0542	2.8697	1.7226	4.7808
AGE2	1.3600	3.8961	2.1614	7.0231
AGE3	0.9290	2.5319	1.6311	3.9303
AGE4	0.7434	2.1030	1.3770	3.2117
AGE5	0.0874	1.0914	0.6646	1.7922
Response Code 2 vs Code 4				
intcept	0.4640	1.5904	1.1454	2.2084
NUMMED	-0.2383	0.7879	0.7455	0.8328
PRINCARE	-0.9067	0.4038	0.3055	0.5339
BLODPRES	0.2945	1.3425	1.0508	1.7151
EXERCISE	0.3895	1.4762	1.1094	1.9641
AGE1	0.9974	2.7112	1.9336	3.8016
AGE2	1.2599	3.5249	2.3684	5.2462
AGE3	0.6874	1.9886	1.4814	2.6695
AGE4	0.6574	1.9297	1.4644	2.5428
AGE5	0.0952	1.0999	0.8050	1.5027

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

$$\hat{\beta}_0 \pm 1.96 \text{ std.error}(\hat{\beta}_0).$$

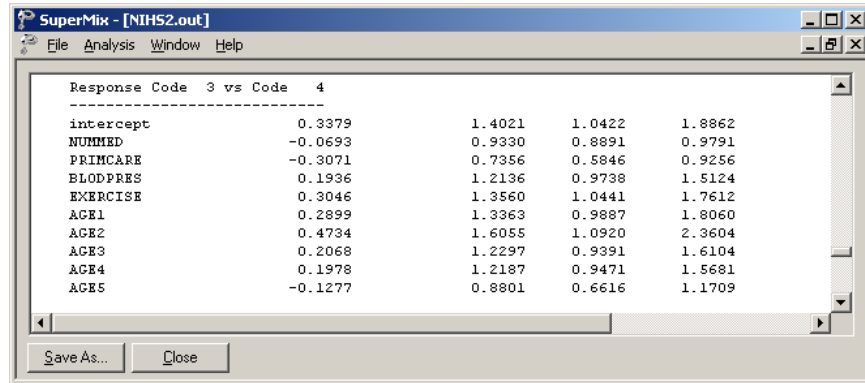
From the output, it follows that this interval is

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

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

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

The odds ratios and confidence intervals for response codes 1, 2 and 3 versus 4 are given as part of the output for this model, as shown below.



The screenshot shows a window titled "SuperMix - [NIHS2.out]" with a menu bar (File, Analysis, Window, Help). The main content area displays the following table:

Response Code	3 vs Code	4		
intercept	0.3379	1.4021	1.0422	1.8862
NUMMED	-0.0693	0.9330	0.8891	0.9791
PRIMCARE	-0.3071	0.7356	0.5846	0.9256
BLODPRES	0.1936	1.2136	0.9738	1.5124
EXERCISE	0.3046	1.3560	1.0441	1.7612
AGE1	0.2899	1.3363	0.9887	1.8060
AGE2	0.4734	1.6055	1.0920	2.3604
AGE3	0.2068	1.2297	0.9391	1.6104
AGE4	0.1978	1.2187	0.9471	1.5681
AGE5	-0.1277	0.8801	0.6616	1.1709

At the bottom of the window, there are two buttons: "Save As..." and "Close".