



Two-level ordinal analysis

Contents

3.1.1	<i>The data</i>	1
3.1.2	<i>An ordinal regression model with random intercept</i>	6

3.1.1 The data

To illustrate the application of the mixed-effects ordinal logistic regression model to longitudinal data, we examined data collected in the NIMH Schizophrenia Collaborative Study on treatment-related changes in overall severity. Specifically, Item 79 of the Inpatient Multidimensional Psychiatric Scale (IMPS; Lorr & Klett, 1966) was used. In this study, patients were randomly assigned to receive one of four medications: placebo, chlorpromazine, fluphenazine, or thioridazine. Since previous analyses (Longford, 1993, and Gibbons & Hedeker, 1994) revealed similar effects for the three anti-psychotic drug groups, they were combined in the present analysis. Finally, again based on previous analysis, a square root transformation of time was chosen to linearize the relationship of the IMPS79 scores over time.

Data for the first 10 observations are shown below in the form of a SuperMix spreadsheet file, named **schizx.ss3**.

	(A) ID	(B) IMPS79	(C) IMPS79D	(D) IMPS79O	(E) DRUG	(F) WEEK	(G) SQRTWEEK	(H) WSQRTxDRUG
1	1103.00	5.50	1.00	4.00	1.00	0.00	0.00	
2	1103.00	3.00	0.00	2.00	1.00	1.00	1.00	
3	1103.00	-9.00	-9.00	-9.00	1.00	2.00	1.41	
4	1103.00	2.50	0.00	2.00	1.00	3.00	1.73	
5	1103.00	-9.00	-9.00	-9.00	1.00	4.00	2.00	
6	1103.00	-9.00	-9.00	-9.00	1.00	5.00	2.24	
7	1103.00	4.00	1.00	2.00	1.00	6.00	2.45	
8	1104.00	6.00	1.00	4.00	1.00	0.00	0.00	
9	1104.00	3.00	0.00	2.00	1.00	1.00	1.00	
10	1104.00	-9.00	-9.00	-9.00	1.00	2.00	1.41	

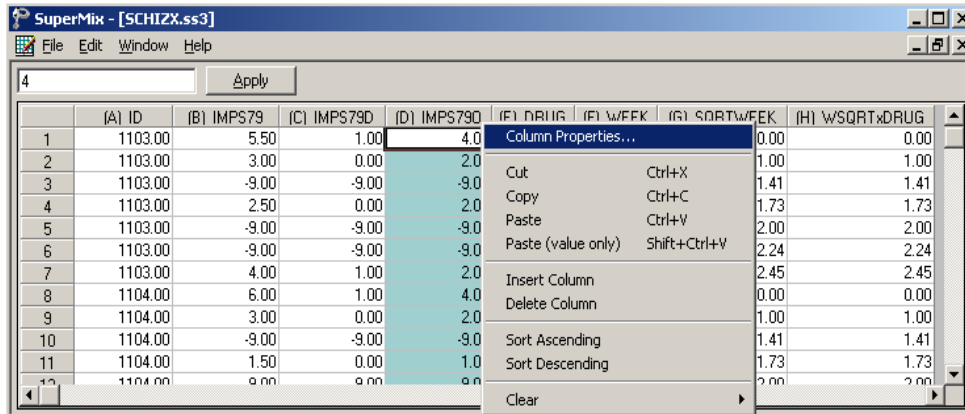
The variables of interest are:

- ID indicates the subject (437 patients in total).
- IMPS79 represents the original score on Item 79 of the Inpatient Multidimensional Psychiatric Scale. It was scored as: 1 = normal, or not at all ill; 2 = borderline mentally ill; 3 = mildly ill; 4 = moderately ill; 5 = markedly ill; 6 = severely ill; and 7 = among the most extremely ill.
- IMPS79D is a recoded version of the same scale, but in binary form, where scores up to, but excluding 3.5 were coded 0, and scores of 3.5 or higher were coded 1. The value "0" is associated with measurements classified as normal, borderline, mildly, or moderately mentally ill, while the value "1" was assigned to measurements corresponding to "markedly ill" through "most extremely ill."
- IMPS79O is also a recoded version of the same scale, but with the 7 original categories reduced to four: 1 = normal or borderline mentally ill, 2 = mildly or moderately ill, 3 = markedly ill, and 4 = severely or among the most extremely ill.
- DRUG indicates the treatment group, where 0 indicates the placebo patients, and 1 refers to the drug patients.
- WEEK represents the time during the course of the study when a specific measurement was made, and ranges between 0 and 6.
- SQRTWEEK is the square root of WEEK. This variable is generated within the SuperMix spreadsheet.
- WSQRTxDRUG is the product of the treatment group and the square root of WEEK.

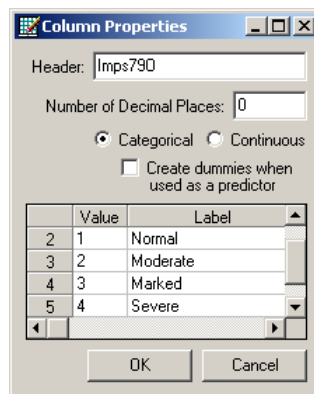
In this data file, each subject's data consist of seven lines, these being the repeated measurements on seven occasions. Notice that there are missing value codes (-9) for some subjects at specific time points. The data from these time points will not be used in the analysis, but data from these subjects at other time points where there are no missing data will be used in the analysis. Thus, for inclusion into the analysis, a subject's data (both the dependent variable and all model covariates being used in a particular analysis) at a specific time point must be complete. The number of repeated observations per subject then depends on the number of time points for which there are non-missing data for that subject. The specification of missing data codes will be illustrated in the model specification section to follow.

3.1.1.1 Defining column properties

Defining column properties for the ordinal data is recommended. We use the column of IMPS79O as an example. First, highlight the column of IMPS79O by clicking on its header. Then right click and select the **Column Properties** option as shown below to open the **Column Properties** dialog box.



The header of the **Column Properties** dialog box indicates the current variable name. Keep the default number of decimal places unchanged. Select the **Categorical** radio button to activate the grid field to enter the labels for each category as shown below.



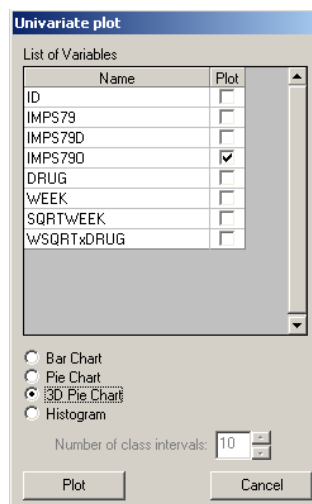
Click on the **OK** button and save the change to the data set by clicking on the **File, Save** option.

The outcome variable: univariate graphs

As a first step, we take a look at the ordinal variable IMPS790 which is the potential dependent variable in this study.

Pie chart

To generate a pie chart for IMPS790, first open the **schizx.ss3** SuperMix spreadsheet. Next, select the **File, Data-based Graphs, Univariate** option to load the **Univariate plot** dialog box. Select the variable IMPS790 and indicate that a **3D Pie Chart** is to be graphed as shown below.



Click the **Plot** button to display the following pie chart.

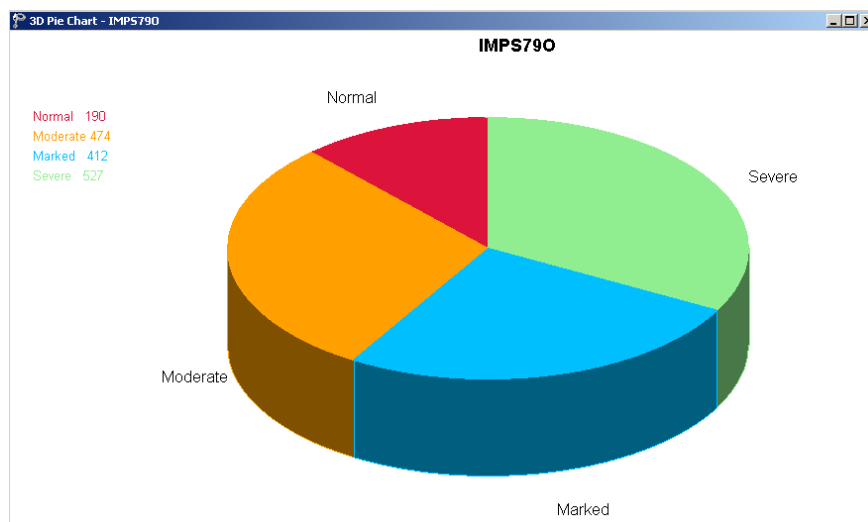


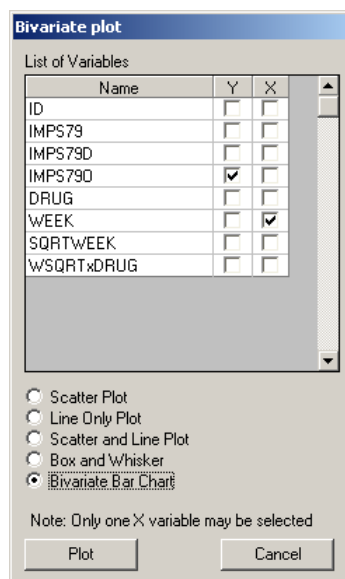
Figure 6.6: Pie chart of IMPS790 values

Note that most of the observations fall into the Severe illness category. Keep in mind that the pie chart takes all observations, regardless of the time of measurement, into account. As such, it is informative about the distribution of all observed values of the potential outcome, but does not provide any information on possible trends in illness level over time.

Relationships between variables: bivariate bar chart

It is hoped that the severity of the illness (IMPS790) will decrease over the treatment period. Before considering fitting a model to these data, we would like to explore the relationship between IMPS790 and WEEK using a bivariate bar chart.

Bivariate bar chart



A bivariate bar chart is accessed via the **Data-based Graphs, Bivariate** option on the **File** menu. The **Bivariate plot** dialog box is completed as below: select the outcome variable IMPS79O as the **Y**-variable of interest, and the predictor WEEK to be plotted on the **X**-axis. Check the **Bivariate Bar Chart** option, and click **Plot**.

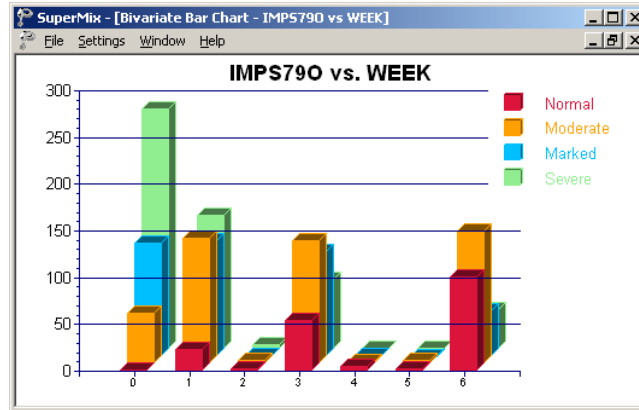


Figure 6.7: Bar chart of IMPS790 vs. WEEK

As shown above, most patients did not participate in the study at weeks 2, 4 and 5. At the beginning of the study (week 0), a large percentage of patients are markedly or severely ill. By the end of the study (week 6), most patients are reported as normal or moderate.

3.1.2 An ordinal regression model with random intercept

3.1.2.1 Introduction

An ordinal variable is a categorical variable where there is a logical ordering to the categories. In most cases, treating an ordinal outcome as a continuous variable is inadvisable. As in the case of a binary outcome variable, a link function is used in order to take the ceiling and floor effects of the ordinal outcome into account. The available link functions in SuperMix include probit, logistic, complementary log-log and log-log.

3.1.2.2 The model

Let the outcome variable be coded into c categories, where $c = 1, 2, \dots, C$. In this example, the ordinal variable IMPS790 defines the severity of the illness in terms of four categories, and thus $C = 4$. As ordinal models utilize cumulative comparisons of the categories, define the cumulative probabilities for the C categories of the outcome Y as $P_{ijc} = \Pr(Y_{ij} \leq c) = \sum_{k=1}^c p_{ijk}$, where p_{ijk} represents the probability that the response of the j th measurement on patient i occurs in category k .

The type of drug, time elapsed since start of treatment, and the interaction between drug taken and time elapsed are of interest as predictors. The logistic regression model with IMPS790 as outcome can then be written as

Level 1 model:

$$y_{ij} = \log\left(\frac{P_{ijc}}{1 - P_{ijc}}\right) = \gamma_c - [b_{0i} + b_{1i}\text{DRUG}_i + b_{2i}\text{SQRTWEEK}_i + b_{3i}(\text{WSQRT} \times \text{DRUG})_i],$$

$$j = 1, \dots, n_i; c = 1, 2, \dots, C - 1$$

Level 2 model:

$$b_{0i} = \beta_0 + v_{0i}, \quad i = 1, \dots, N$$

$$b_{1i} = \beta_1$$

$$b_{2i} = \beta_2$$

$$b_{3i} = \beta_3$$

The cumulative probability can be expressed by

$$P_{ijc} = \frac{e^{\gamma_c - [b_{0i} + b_{1i}\text{DRUG}_i + b_{2i}\text{SQRTWEEK}_i + b_{3i}(\text{WSQRT} \times \text{DRUG})_i]}}{1 + e^{\gamma_c - [b_{0i} + b_{1i}\text{DRUG}_i + b_{2i}\text{SQRTWEEK}_i + b_{3i}(\text{WSQRT} \times \text{DRUG})_i]}}$$

To obtain the probability for category c ,

$$p_{ij,c} = P_{ij,c+1} - P_{ij,c}$$

As shown above, the intercept b_{0i} is estimated by a level-2 equation. It indicates that patient i 's initial IMPS790 value is not only determined by the population average β_0 , but also by the patient difference v_{0i} . In other words, patients may have different average intercepts, and the model makes provision for this eventuality. The slopes are assumed to be the same for all the patients, which implies that each patient's trend line is parallel to the population trend.

The connection between an ordinal outcome variable y with C categories and an underlying continuous variable y^* is

$$y = c \leftrightarrow \gamma_{j-1} \leq y^* \leq \gamma_j, \quad c = 1, 2, \dots, C$$

where it is assumed that $\gamma_0 = -\infty$ and $\gamma_C = +\infty$. In addition, γ_1 is usually set to 0 to avoid identification problems.

3.1.2.3 Setting up the analysis

Open the SuperMix spreadsheet **schizx.ss3** and select the **File, New Model Setup** option.

Model Setup: SCHIZX1.mum

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

Title 1: Random intercept ordinal logistic regression model

Title 2: NIMH SCHIZ data

Dependent Variable Type: ordered

Dependent Variable: IMPS790

	Value
1	1
2	2
3	3
4	4

Level-2 IDs: ID

Level-3 IDs:

Write Bayes Estimates: no

Convergence Criterion: 0.0001

Number of Iterations: 100

Missing Values Present: true

Missing Value for the Dependent Var: -9

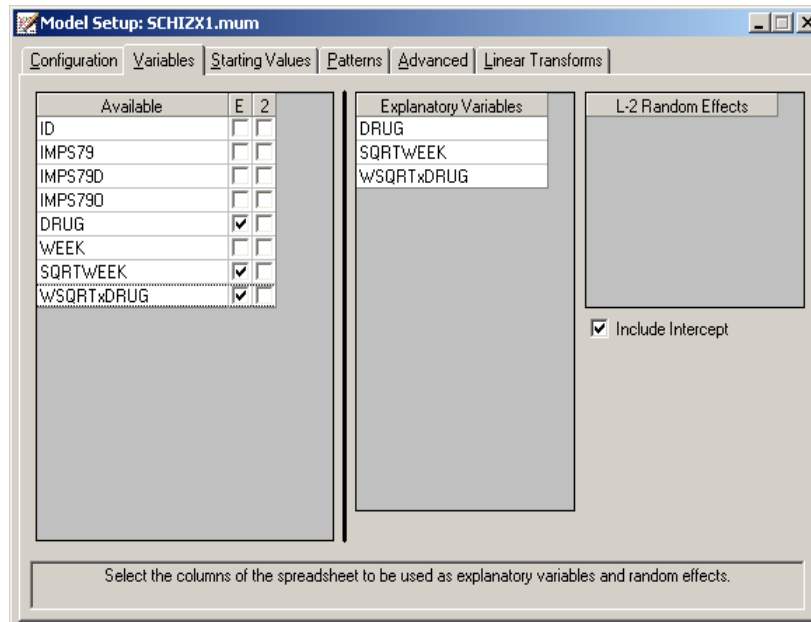
Global Missing Value: -9

Perform Crosstabulation: no

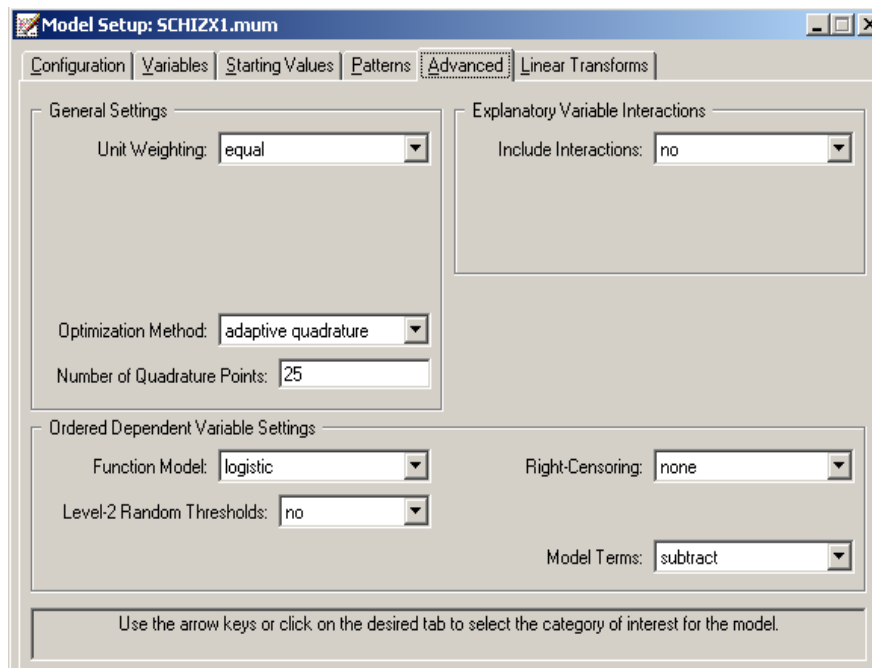
Output Type: standard

In the **Configuration** screen of the **Model Setup** window, enter a title for the analysis in the **Title** text boxes. Select **ordered** from the **Dependent Variable Type** drop-down list box. Select the outcome variable **IMPS790** from the **Dependent Variable** drop-down list box. Once this selection has been made, the **Categories** grid is displayed, with the distinct values of the categories shown.

We notice that the missing value **-9** is also included as a category. The **Missing Values Present** drop-down list box is used to specify the values of missing data for both outcome and predictors. As a first step, set the value of the **Missing Values Present** drop-down list box to **True**. The appearance of the screen will change when this is done, and text boxes for the specification of the missing data codes are displayed. Start by entering the value **-9** in the **Missing Value for the Dependent Var** text box. Do the same for all the predictors included in the model by entering **-9** in the **Global Missing Value** text box. Finally, select the patient ID from **Level-2 IDs** drop-down list box to produce the **Configuration** screen seen above.



Proceed to the **Variables** screen by clicking on this tab. The **Variables** tab is used to specify the fixed and random effects to be included in the model. Select DRUG, SQRTWEEK and WSQRTxDRUG as explanatory (fixed) variables using the **E** check boxes next to the variables names in the **Available** grid at the left of the screen. The **Include Intercept** check box in the **Explanatory Variables** grid is checked by default, indicating that an intercept term will automatically be included in the fixed part of the model. Next, specify the random effects at level 2 of the hierarchy. In this example, we want to fit a model with random intercepts at level 2. By default, the **Include Intercept** check box in the **L-2 Random Effects** is checked, indicating the inclusion of a random intercept at this level in the model.



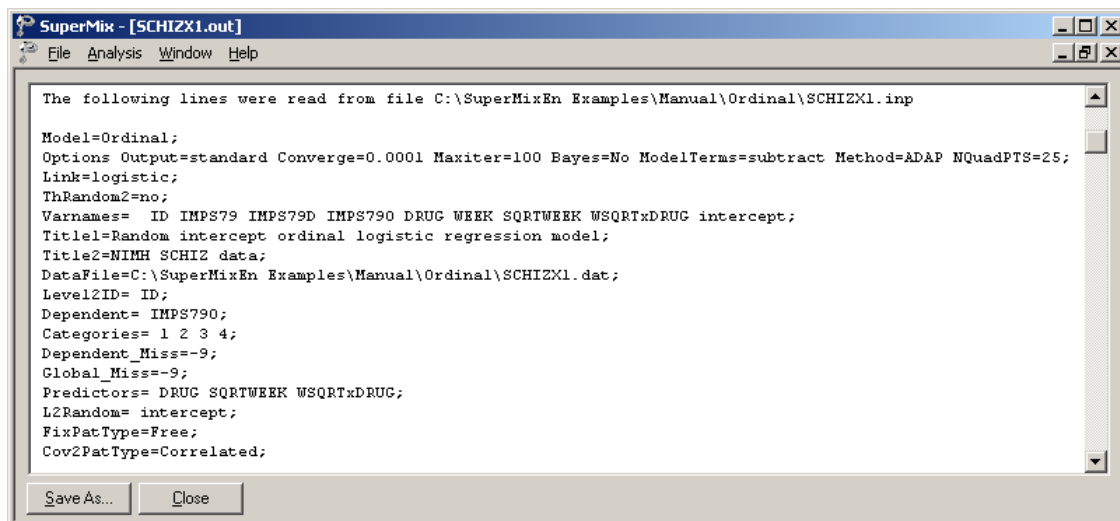
The default link function for the ordinal outcome variable is the probit link function. To change it to the logistic link function corresponding to the model formulation above, click on the **Advanced** tab and select the **logistic** link function from the **Function Model** drop-down list box as shown above. Use 25 quadrature points.

Before running the analysis, the model specifications have to be saved. Select the **File, Save As** option, and provide a name (**SCHIZX1.mum**) for the model specification file. Run the analysis by selecting the **Run** option from the **Analysis** menu.

3.1.2.4 Discussion of results

Syntax

The syntax corresponding to the model setup is given in the **model specifications**. These lines of SuperMix syntax are saved as a ***.inp** file with the same name as the model setup file (***.mum**). At the top of the output file, the syntax lines are printed as shown below.



The screenshot shows a window titled "SuperMix - [SCHIZX1.out]" with a menu bar containing "File", "Analysis", "Window", and "Help". The main text area displays the following syntax:

```
The following lines were read from file C:\SuperMixEn Examples\Manual\Ordinal\SCHIZX1.inp

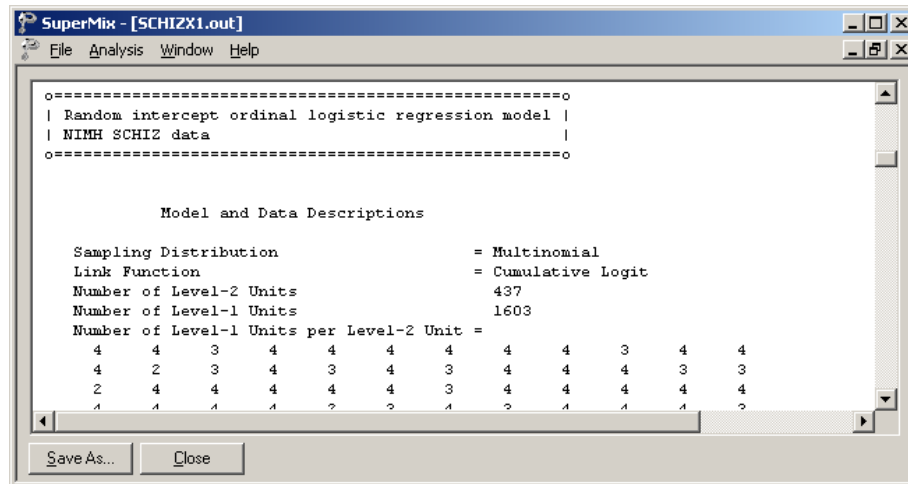
Model=Ordinal;
Options Output=standard Converge=0.0001 Maxiter=100 Bayes=No ModelTerms=subtract Method=ADAP NQuadPTS=25;
Link=logistic;
ThRandom2=no;
VarNames= ID IMPS79 IMPS79D IMPS790 DRUG WEEK SQRTWEEK WSQRTxDRUG intercept;
Title1=Random intercept ordinal logistic regression model;
Title2=NIMH SCHIZ data;
DataFile=C:\SuperMixEn Examples\Manual\Ordinal\SCHIZX1.dat;
Level2ID= ID;
Dependent= IMPS790;
Categories= 1 2 3 4;
Dependent_Miss=-9;
Global_Miss=-9;
Predictors= DRUG SQRTWEEK WSQRTxDRUG;
L2Random= intercept;
FixPatType=Free;
Cov2PatType=Correlated;
```

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

The first part indicates that an ordinal outcome is analyzed, states the selection of iteration control options, does not request Bayes residuals, and contains all the specifications necessary to define the model fitted as an ordinal model with logistic link function. The second part of the syntax provides information on the structure of the data, the name and structure of the outcome variable, the missing values and the predictors included in the model.

Model and data description

The next section of the output file contains a description of the hierarchical structure and model specifications. The use of a logistic response function (logit link function) with the assumption of a normal distribution of random effects is indicated. This is followed by a summary of the number of observations nested within each patient. As shown below, 437 patients with a total of 1603 observations are included in this study after listwise deletion. The number of observations per patient (level 2 unit) varies between 2 and 5.



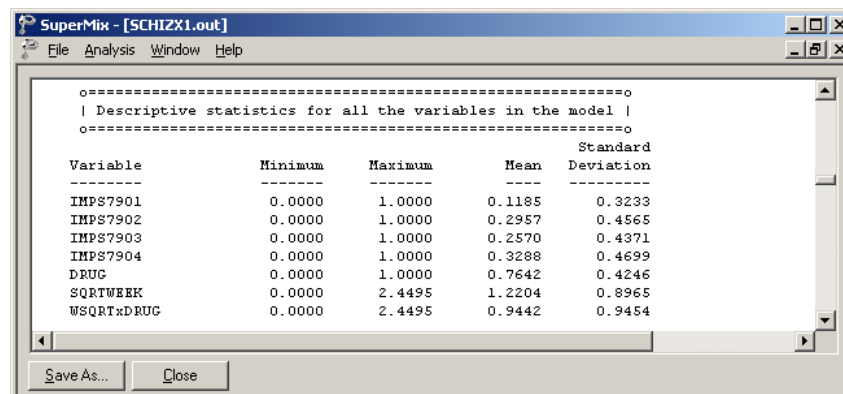
```
o=====o
| Random intercept ordinal logistic regression model |
| NIMH SCHIZ data |
o=====o

Model and Data Descriptions

Sampling Distribution          = Multinomial
Link Function                 = Cumulative Logit
Number of Level-2 Units      = 437
Number of Level-1 Units      = 1603
Number of Level-1 Units per Level-2 Unit =
4 4 3 4 4 4 4 4 4 3 4 4
4 2 3 4 3 4 3 4 4 4 3 3
2 4 4 4 4 4 3 4 4 4 4 4
4 4 4 4 2 2 4 2 4 4 4 2
```

Descriptive statistics and starting values

Next, the descriptive statistics for all the variables are given. Notice that the variable name WSQRTxDRUG is truncated to WSQRTxDR. This is because SuperMix only recognizes the first 8 characters of a variable name.



```
o=====o
| Descriptive statistics for all the variables in the model |
o=====o

Variable          Minimum    Maximum    Mean    Standard
-----          -
IMPS7901          0.0000    1.0000    0.1185    0.3233
IMPS7902          0.0000    1.0000    0.2957    0.4565
IMPS7903          0.0000    1.0000    0.2570    0.4371
IMPS7904          0.0000    1.0000    0.3288    0.4699
DRUG              0.0000    1.0000    0.7642    0.4246
SQRTWEEK          0.0000    2.4495    1.2204    0.8965
WSQRTxDRUG       0.0000    2.4495    0.9442    0.9454
```

Descriptive statistics are followed by the starting values of parameters.

SuperMix - [SCHIZX1.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    3756.1953            1597            2.3520
Pearson Chi-square            4426.5410            1597            2.7718

```

Save As... Close

SuperMix - [SCHIZX1.out]

File Analysis Window Help

```

                Estimated regression weights

Parameter                Estimate                Standard
-----                -                Error                z Value                P Value
Threshold1                -3.8073                0.1899                -20.0532            0.0000
Threshold2                -1.7602                0.1703                -10.3375            0.0000
Threshold3                -0.4221                0.1636                -2.5796            0.0099
DRUG                -0.0006                0.1883                -0.0032            0.9974
SQRWEEK                -0.5366                0.1108                -4.8427            0.0000
WSQRTxDRUG                -0.7510                0.1277                -5.8817            0.0000

                Odds Ratio and 95% Odds Ratio Confidence Intervals

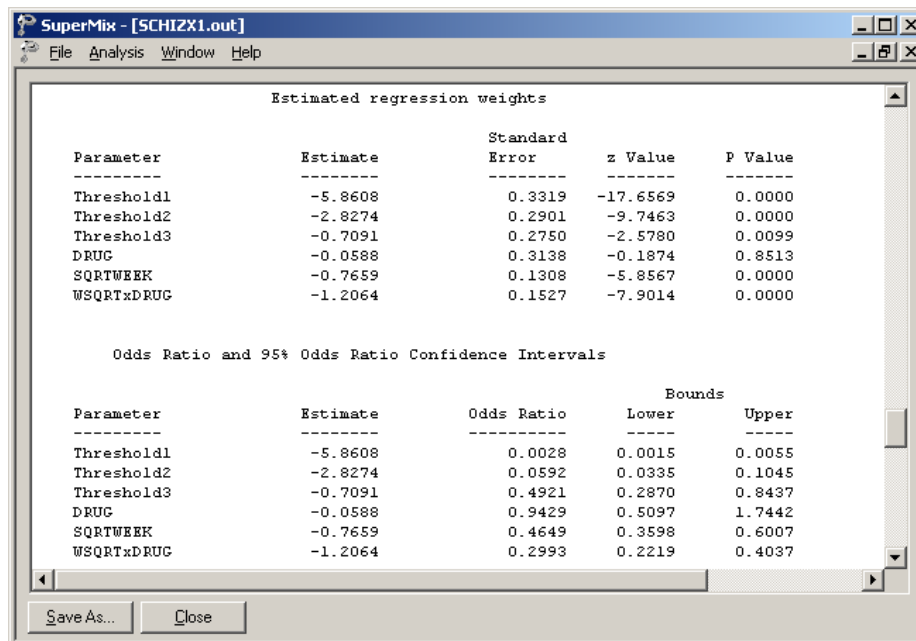
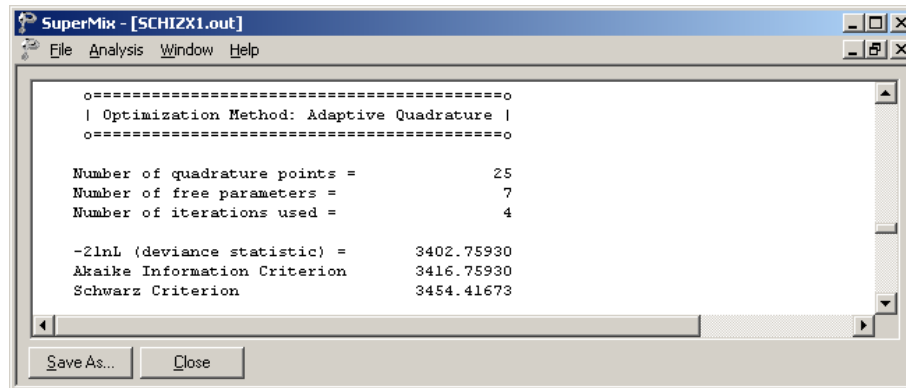
Parameter                Estimate                Odds Ratio                Bounds
-----                -                -                Lower                Upper
Threshold1                -3.8073                0.0222                0.0153            0.0322
Threshold2                -1.7602                0.1720                0.1232            0.2402
Threshold3                -0.4221                0.6557                0.4758            0.9036
DRUG                -0.0006                0.9994                0.6909            1.4456
SQRWEEK                -0.5366                0.5847                0.4706            0.7266
WSQRTxDRUG                -0.7510                0.4719                0.3674            0.6061

```

Save As... Close

Fixed effects estimates

The final results after 16 iterations are shown next. The estimates are shown in the column with heading Estimate, and correspond to the coefficients $\beta_0, \beta_1, \dots, \beta_3$ in the model specification. The standard error, z-value and p-value are also printed.



The variation in the intercept over the subjects is estimated as 3.7739, and from the associated p -value we conclude that there is significant variation in the (random) intercept between the patients included in this analysis. In the case of the fixed effects, a 2-tailed p -value is used, as the alternative hypothesis considered here is of the form $H_1: \beta \neq 0$. As variances are constrained to be elements of the interval $[0, +\infty)$ and thresholds are constrained so that $\gamma_1 \leq \gamma_2 \leq \gamma_3$, the p -values used for these effects are 1-tailed. The results indicate that the treatment groups do not differ significantly at baseline (the estimated DRUG coefficient is not significant). The placebo group seems to improve over time, as the SQRTWEEK coefficient is both significant and negative. Note that the interpretation of the main effects depends on the coding of the variable, and on the significance of the WSQRTxDR interaction which forms part of the model.

Estimated regression weights

Parameter	Estimate	Standard Error	z Value	P Value
intcept	5.8608	0.3319	17.6569	0.0000
Threshold2	3.0333	0.1375	22.0644	0.0000
Threshold3	5.1517	0.1822	28.2714	0.0000

Odds Ratio and 95% Odds Ratio Confidence Intervals

Parameter	Estimate	Odds Ratio	Bounds	
			Lower	Upper
intcept	5.8608	350.9925	183.1293	672.7253
Threshold2	3.0333	20.7661	15.8611	27.1880
Threshold3	5.1517	172.7219	120.8472	246.8642

Estimated level 2 variances and covariances

Parameter	Estimate	Standard Error	z Value	P Value
intercept/intercept	3.7776	0.4653	8.1180	0.0000

As noted before, it is assumed that $\gamma_0 = -\infty$ and $\gamma_C = +\infty$. For the present example, $C = 4$, and from the output we see that $\hat{\gamma}_1 = -5.8593$, $\hat{\gamma}_2 = -2.8264$ and $\hat{\gamma}_3 = -0.7085$. These values are used in combination with the coefficients of DRUG, SQRTWEEK, and WSQRTxDR to calculate estimated outcomes for different groups of patients.

Intraclass correlation (ICC)

Below the estimate the intraclass correlation (ICC) is given. The residual variance for the logistic link function is assumed to be $\pi^2/3$.

Calculation of the intraclass correlation

```

-----
residual variance = pi*pi / 3 (assumed)
cluster variance = 3.7776

intraclass correlation = 3.7776 / ( 3.7776 + (pi*pi/3) ) = 0.535

```

The ICC in this model refers to the intra-person correlation. It is reported as 0.534, which is fairly high. Generally, the shorter the interval between the repeated measurements, the higher the ICCs will be.

3.1.2.5 Interpreting the output

Estimated outcomes for groups: unit-specific probabilities

To evaluate the expected effect of the treatment group and the square root of time of treatment, while allowing for the interaction between treatment and the square of time, we use the expression below:

$$\log\left(\frac{\hat{P}_{ijc}}{1-\hat{P}_{ijc}}\right) = \hat{\gamma}_c - \left[\hat{b}_{1i} \text{DRUG}_i + \hat{b}_{2i} \text{SQRTWEEK}_i + \hat{b}_{3i} (\text{WSQRT} \times \text{DRUG})_i \right]$$

or, in the notation introduced in Section 6.3.2.2,

$$\begin{aligned} \log\left(\frac{\hat{P}_{ijc}}{1-\hat{P}_{ijc}}\right) &= \hat{\eta}_{ijc} \\ &= \hat{\gamma}_c - 0.0585 \times \text{DRUG}_i + 0.7658 \times \text{SQRTWEEK}_i \\ &\quad + 1.2061 \times (\text{WSQRT} \times \text{DRUG})_i. \end{aligned}$$

When $c = 1$, we find that, for a patient from the control group ($\text{DRUG} = 0$, $\text{SQRTWEEK} = \text{WSQRT} \times \text{DR} = 0$),

$$\begin{aligned} \log\left(\frac{\hat{P}_{ij1}}{1-\hat{P}_{ij1}}\right) &= \hat{\eta}_{ij1} = -5.8593 \\ \hat{P}_{ij1} &= \frac{e^{\hat{\eta}_{ij1}}}{1+e^{\hat{\eta}_{ij1}}} = 0.0028 \end{aligned}$$

Similarly, the probabilities that a typical patient from the control group responded in a specific category at the start of the study are obtained by using $\hat{\gamma}_2 = -2.8264$, and $\hat{\gamma}_3 = -0.7085$.

The cumulative probabilities we calculated are

$$\hat{P}_{ij2} = \frac{e^{\hat{\eta}_{ij2}}}{1 + e^{\hat{\eta}_{ij2}}} = \frac{e^{-2.8264}}{1 + e^{-2.8264}} = 0.0559$$

$$\hat{P}_{ij3} = \frac{e^{\hat{\eta}_{ij3}}}{1 + e^{\hat{\eta}_{ij3}}} = \frac{e^{-0.7085}}{1 + e^{-0.7085}} = 0.3299.$$

Thus, the estimated category probabilities we have for such a group (category 1 to 4) are obtained as

$$\hat{p}_{ij1} = 0.0028 - 0 = 0.0028$$

$$\hat{p}_{ij2} = 0.0559 - 0.0028 = 0.0531$$

$$\hat{p}_{ij3} = 0.3299 - 0.059 = 0.2740$$

$$\hat{p}_{ij4} = 1 - 0.3299 = 0.6701.$$

For this group of patients (DRUG = 0) at the starting week, the expected percentages of patients in each of the categories are as follows: 0.3% of the patients are normal or borderline mentally ill; 5.3% of the patients are mildly or moderately ill; 27.4% are markedly ill and 67% are severely or extremely ill. Similarly, we can calculate the estimated percentages for both groups at all the time points as shown in Table 6.8.

The contents of Table 6.8 can be graphically represented as shown in Figures 6.8 and 6.9. It clearly shows that the numbers of markedly and severely ill patients decrease dramatically over time. The improvement for the drug patients is larger than the placebo patients.

Table 6.8: Estimated % for both groups at 7 time points

severity	Drug patients (drug = 1)				Placebo patients (drug = 0)			
	normal	moderate	marked	severe	normal	moderate	marked	severe
week 0	0.30%	5.61%	28.39%	65.70%	0.28%	5.31%	27.40%	67.01%
week 1	0.65%	11.25%	40.99%	47.11%	2.01%	27.84%	48.11%	22.04%
week 2	0.89%	14.76%	45.02%	39.34%	4.43%	44.62%	39.84%	11.10%
week 3	1.13%	18.00%	47.16%	33.71%	7.99%	56.32%	29.43%	6.26%
week 4	1.38%	21.13%	48.21%	29.28%	12.84%	62.51%	20.87%	3.79%
week 5	1.65%	24.17%	48.50%	25.69%	19.00%	63.96%	14.63%	2.41%
week 6	1.94%	27.13%	48.24%	22.69%	26.32%	61.79%	10.29%	1.60%

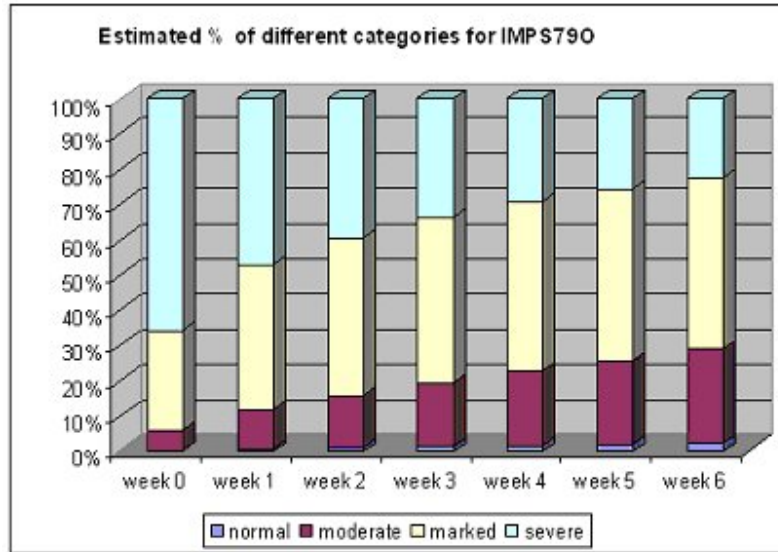


Figure 6.8: Estimated percentage of patients over time (treatment group)

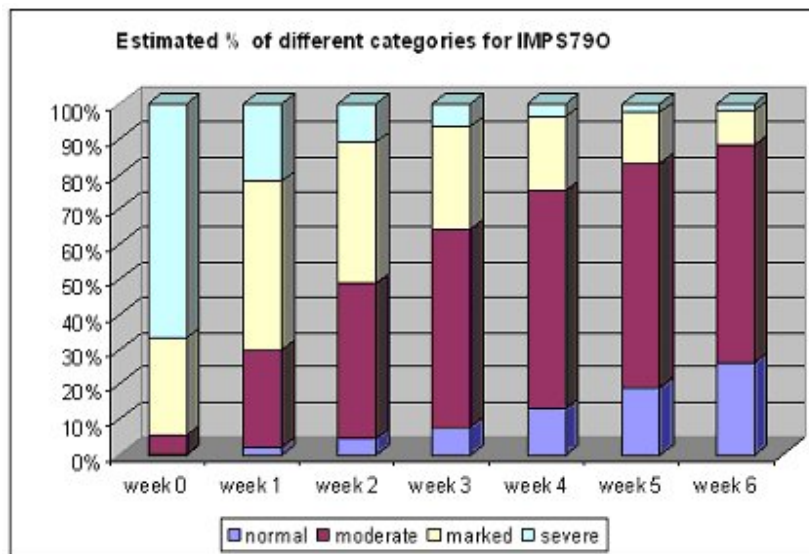


Figure 6.9: Estimated percentage of patients over time (control group)