



Two-level models for ordinal outcomes

The data

To illustrate the application of the mixed-effects ordinal probit regression model to longitudinal data, we examined data collected in the NIMH Schizophrenia Collaborative Study on treatment related changes in overall severity. For the present illustration of the ordinal mixed-effects model, we recoded the seven ordered categories of Imps79 into four: 1) normal or borderline mentally ill, 2) mildly or moderately ill, 3) markedly ill, and 4) severely or among the most extremely ill.

For both models discussed in this section and the next example, the repeated ordinal IMPS score (Imps79O) is modeled in terms of a dummy-coded drug effect (TxDrug: placebo = 0 and drug = 1), a time effect (SqrtWeek: square root of week) and a drug by time interaction (Tx*SWeek). In terms of the random effects, this example specifies a random intercepts model, while the next example will show a model that makes allowance for patients to vary in terms of both their intercept and their trend across time (random intercepts and slopes).

The models

To motivate the ordinal regression model, it is often assumed that there is an unobservable latent variable (y) which is related to the actual response through the "threshold concept". For the dichotomous model, one threshold value is assumed, while for the ordinal model, a series of threshold values $\gamma_1, \gamma_2, \dots, \gamma_{J-1}$, where J equals the number of ordered categories, $\gamma_0 = -\infty$, and $\gamma_J = \infty$, is assumed. Here, a response occurs in category j ($Y = j$) if the latent response process y exceeds the threshold value γ_{j-1} , but does not exceed the threshold value γ_j .

Assume that there are $i = 1, \dots, N$ level-2 units and $k = 1, \dots, n_i$ level-1 units nested within each level-2 unit. The mixed-effects regression model for the latent response variable y_{ik} can be written as follows:

$$y_{ik} = \mathbf{x}'_{ik} \boldsymbol{\beta}_i + \mathbf{w}'_{ik} \boldsymbol{\alpha} + e_{ik}$$

where \mathbf{w}_{ik} is the $p \times 1$ covariate vector and \mathbf{x}_{ik} is the design vector for the r random effects, both vectors being for the k -th level-1 unit nested within level-2 unit i . Also, $\boldsymbol{\alpha}$ is the $p \times 1$ vector of unknown fixed regression parameters, $\boldsymbol{\beta}_i$ is the $r \times 1$ vector of unknown random effects for the level-2 unit i , and e_{ik} are the model residuals. The distribution of the random effects is assumed to be multivariate normal with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$, and the residuals are assumed to be independently normally distributed with mean 0 and variance σ^2 .

A model with probit link function and random intercept

In the first example of this section we consider the model

$$y_{ik} = \beta_{i0} + \alpha_0 + \alpha_1 \text{TxDrug}_{ik} + \alpha_2 \text{SqrtWeek}_{ik} + \alpha_3 \text{Tx} * \text{SWeek}_{ik} + e_{ik}$$

With the above mixed-effects regression model for the underlying and unobservable variable y_{ik} , the probability, for a given level-2 unit i , that $Y_k = j$ (a response occurs in category j), conditional on $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, is given by the following equation:

$$P(Y_k = j | \boldsymbol{\beta}_i, \boldsymbol{\alpha}) = \Phi[(\gamma_j - z_{ik}) / \sigma] - \Phi[(\gamma_{j-1} - z_{ik}) / \sigma]$$

where $z_{ik} = \mathbf{x}'_{ik} \boldsymbol{\beta}_i + \mathbf{w}'_{ik} \boldsymbol{\alpha}$ and $\Phi(\cdot)$ represents the cumulative standard normal density function. Without loss of generality, the origin and unit of z may be chosen arbitrarily. For convenience, let $\gamma_1 = 0$ and $\sigma = 1$, then

$$P(\text{Imps79O}_{ik} = j | \boldsymbol{\beta}_i, \boldsymbol{\alpha}) = \Phi(\gamma_j - z_{ik}) - \Phi(\gamma_{j-1} - z_{ik}).$$

A model with probit link function with random intercept and slope

In the second example we again consider the model

$$y_{ik} = \beta_{i0} + \beta_{i0} \text{SqrtWeek}_{ik} + \alpha_0 + \alpha_1 \text{TxDrug}_{ik} + \alpha_2 \text{SqrtWeek}_{ik} + \alpha_3 \text{Tx} * \text{SWeek}_{ik} + e_{ik}$$

where the vector of unknown parameters β_i in the mixed-effects regression model

$$y_{ik} = \mathbf{x}'_{ik} \beta_i + \mathbf{w}'_{ik} \alpha + e_{ik}$$

is extended to include both intercept and slope (SqrtWeek) effects.

Example: Probit link function with random intercept

Preparing the data

The model is fitted to the data in **schizx1.ss3**. Use the **File, Open** option to activate an **Open** dialog box, and browse for the file **schizx1.xls** in the **Examples** folder.

	(A) Patient	(B) Imps79	(C) Imps79	(D) Imps79	(E) TxDrug	(F) Week	(G) SqrtWeek	(H) Tx*SWeek
1	1103	5.50	1	4	1	0	0.00	0.00
2	1103	3.00	0	2	1	1	1.00	1.00
3	1103	-9.00	-9	-9	1	2	1.41	1.41
4	1103	2.50	0	2	1	3	1.73	1.73
5	1103	-9.00	-9	-9	1	4	2.00	2.00
6	1103	-9.00	-9	-9	1	5	2.24	2.24
7	1103	4.00	1	2	1	6	2.45	2.45
8	1104	6.00	1	4	1	0	0.00	0.00
9	1104	3.00	0	2	1	1	1.00	1.00
10	1104	-9.00	-9	-9	1	2	1.41	1.41

Setting up the analysis

The next step is to describe the model to be fitted. We use the SuperMix interface to provide the model specifications. From the main menu bar, select the **File, New Model Setup** option.

File	Edit	Window	Help
New Project			Ctrl+N
Import Data File...			Ctrl+I
Close			
New Model Setup			Ctrl+W
Open Existing Model Setup...			Ctrl+E
Convert MIX Definition File...			Ctrl+M

The **Model Setup** window that appears has six tabs. In this example, the **Configuration**, **Variables**, and **Advanced** tabs are used in model specification. By default, the **Configuration** screen is displayed on top.

As a first step, enter a title for the analysis in the **Title 1** and **Title 2** text boxes. The patient identification variable (Patient) is used to define the hierarchical structure of the data, and is selected as the Level-2 ID from the **Level-2 IDs** drop-down list box. Select the ordinal outcome variable Imps790 from the **Dependent Variable** drop-down list box. Specify the type of outcome as ordered using the **Dependent Variable Type** drop-down list box. Once this selection is made, the **Categories** grid is displayed, with the distinct values of the categories shown.

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. In terms of the missing value specification, notice that even though missing values are coded only for the dependent variable in the input data file, numeric missing value codes must be specified in the model specification file for all model terms (if **Missing Values Present = True**). In this case, the value -9 was specified for all variables since for the dependent variable this value is the correct missing value code, while for all other model terms (Intcept, TxDrug, SqrtWeek, and Tx*SWeek) this value was never observed. Once -9 is entered in the **Missing Value for Dependent Var** text box, the categories grid will automatically be updated to display 1, 2, 3, and 4 only.

The screenshot shows a software configuration window with the following settings:

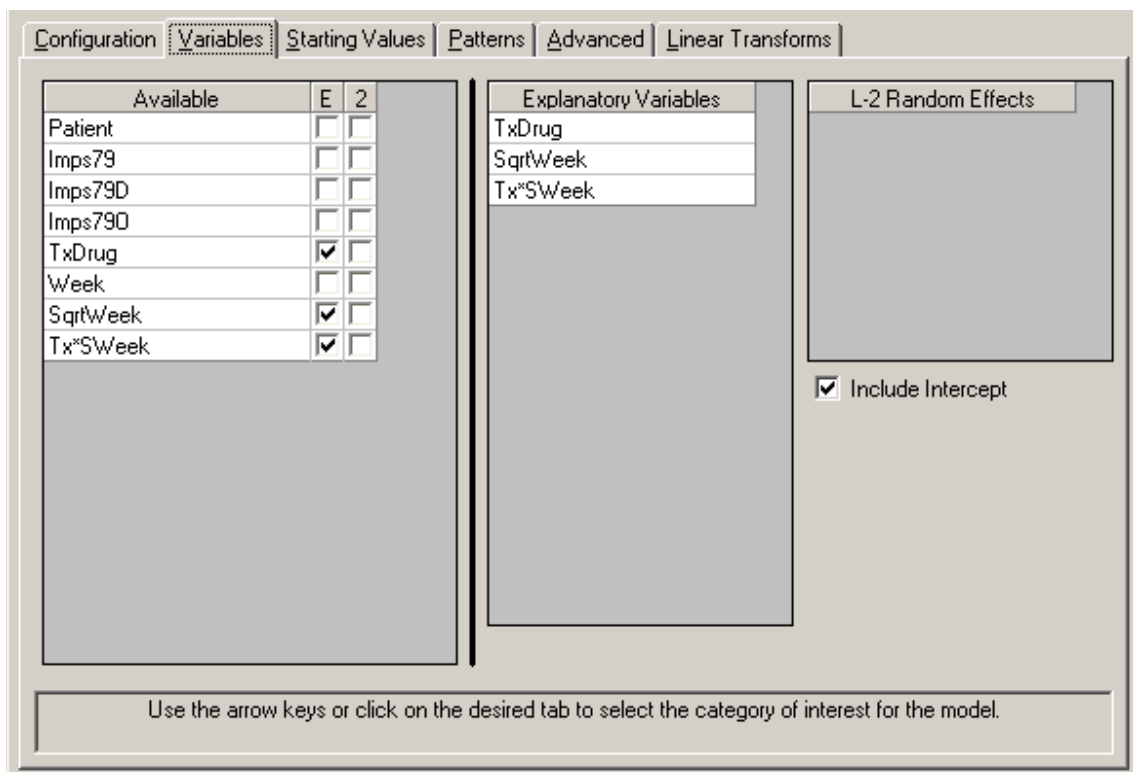
- Title 1:** NIMH Schiz data - two groups-seven timepoints
- Title 2:** IMPS79 (binary) across SQRT week - one random effect
- Dependent Variable Type:** ordered
- Dependent Variable:** Imps790
- Level-2 IDs:** Patient
- Level-3 IDs:** (empty)
- Categories:**

	Value
1	1
2	2
3	3
4	4
- Write Bayes Estimates:** means only
- Convergence Criterion:** 0.0001
- Number of Iterations:** 100
- Missing Values Present:** true
- Perform Crosstabulation:** yes
- Missing Value for the Dependent Var:** -9
- Crosstab Variable:** SqrtWeek
- Global Missing Value:** -9

Select between writing the Bayes estimates to an optional results file or supressing them.

Request a crosstabulation of the variable SqrtWeek by the response variable Imps790 by selecting the **yes** option from the **Perform Crosstabulation** drop-down list box, followed by the selection of SqrtWeek as the **Crosstab Variable**. Finally, request the writing out of Bayes estimates using the **Write Bayes Estimates** drop-down list box. The default setting for the **Number of Iterations** text box is used. Proceed to the **Variables** screen by clicking on this tab.

The **Variables** screen is used to specify the fixed and random effects to be included in the model. Select the explanatory (fixed) variables TxDrug, SqrtWeek, and Tx*SWeek using the check boxes next to the variable names in the **E** column of the **Available** grid. After selecting all the explanatory variables, the random effect(s) at level 2 must be selected. By default, an intercept will be included in the fixed part of the model, as indicated by the check box next to **Include Intercept** in the **Explanatory Variables** grid. Similarly, a random intercept at level 2 is assumed. The intercept is assumed to vary randomly over higher levels of the hierarchy, while the slopes of the predictors TxDrug, SqrtWeek, and Tx*SWeek are assumed to be adequately described by common, fixed coefficients. The completed **Variables** screen is shown below. To change estimation settings such as the number of quadrature points used in estimation, the **Advanced** screen is used. Click **Advanced** to access this screen and complete the model specification process.



Available	E	2
Patient	<input type="checkbox"/>	<input type="checkbox"/>
Imps79	<input type="checkbox"/>	<input type="checkbox"/>
Imps79D	<input type="checkbox"/>	<input type="checkbox"/>
Imps79O	<input type="checkbox"/>	<input type="checkbox"/>
TxDrug	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Week	<input type="checkbox"/>	<input type="checkbox"/>
SqrtWeek	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Tx*SWeek	<input checked="" type="checkbox"/>	<input type="checkbox"/>

Explanatory Variables

TxDrug

SqrtWeek

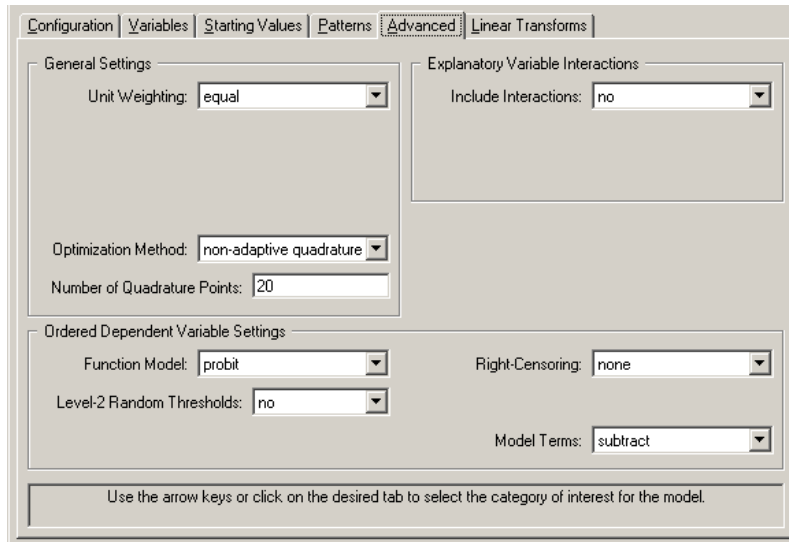
Tx*SWeek

L-2 Random Effects

Include Intercept

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

Set the **Function Model** option to probit. All other settings on this screen are left at their previous values, and we can save our model specification to file prior to running the analysis. Select the **File, Save** option, provide a name for the model specification file (such as **schizx2.mum**), and run the analysis by selecting the **Run** option from the **Analysis** menu.



Discussion of results

Portions of the output file **schizx2.out** are shown below.

```
SuperMix - [schizx2.out]
File Analysis Window Help

NIMH Schiz Data - 2grps - 7 timepoints
IMPS79 (ordinal) across Sqrt week - 1 random effect

Response function: normal

Random-effects distribution: normal

Covariate(s) and random-effect(s) mean subtracted from thresholds
==> positive coefficient = positive association between regressor
and ordinal outcome
```

In the first section of the output file, a description of the model specifications is provided. The use of a normal response function (probit link function), and the assumption of a normal distribution of random effects are reported. This is followed by a summary of the number of measurements nested within each subject. Note that for some of the 437 subjects, only 2 measurements are available for analysis.

```

SuperMix - [schizx2.out]
File Analysis Window Help

Numbers of observations
-----
Level 1 observations = 1603
Level 2 observations = 437

The number of level 1 observations per level 2 unit are:

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 3 4 3 4 4 4 3 4 4 2 2 4 5 4 2 4
4 3 4 4 3 2 3 4 4 4 4 4 4 2 4 4 4 5 4
4 2 2 4 2 4 4 3 3 4 4 4 4 4 4 4 4 3 3
4 2 3 4 4 4 2 5 3 4 4 2 4 4 4 2 4 4 4
4 4 4 4 4 4 5 2 4 3 4 4 2 2 4 4 4 4 4
2 4 4 4 4 4 4 4 4 4 4 4 4 2 4 4 2 4 4
4 4 3 4 2 4 4 3 2 3 4 4 3 3 4 3 4 4 4
4 4 4 4 4 4 4 4 4 4 4 2 3 3 5 4 3 4 4
3 2 4 4 4 4 4 3 3 4 4 4 4 4 4 4 4 4 4

```

```

SuperMix - [schizx2.out]
File Analysis Window Help

Descriptive statistics for all variables
-----
Variable           Minimum           Maximum           Mean           Stand. Dev.
IMPS790             1.00000           4.00000           2.79601         1.02840
intcept             1.00000           1.00000           1.00000         0.00000
intcept             1.00000           1.00000           1.00000         0.00000
TxDRUG              0.00000           1.00000           0.76419         0.42464
SQRTWEEK            0.00000           2.44950           1.22041         0.89651
TxSQRT              0.00000           2.44950           0.94424         0.94541

Categories of the response variable IMPS790
-----
Category           Frequency           Proportion
1.00                190.00              0.11853
2.00                474.00              0.29570
3.00                412.00              0.25702
4.00                527.00              0.32876

```

The data summary is followed by descriptive statistics for all the variables included in the model. We note that 33% of the measurements were in the highest category of the outcome variable and correspond to the "severely or among the most extremely ill" group. Only 12% of measurements are in the first category ("normal, not at all ill").

The following portion of the output is a crosstabulation of the seven distinct values of the variable SqrtWeek by the four categories of the outcome variable IMPS790. We note that there are relatively few observations for the third, fifth and sixth weeks. For example, for week 5 (SqrtWeek = 2.24), measurements on only 9 of the 437 patients are available.

SuperMix - [schizx2.out]

File Analysis Window Help

Crosstabulation of variable SQRWEEK by the response variable IMPS790

SQRWEEK	IMPS790				Total
	1.00	2.00	3.00	4.00	
0.00	1.0 (0.00)	54.0 (0.12)	122.0 (0.28)	257.0 (0.59)	434.0
1.00	23.0 (0.05)	135.0 (0.32)	124.0 (0.29)	144.0 (0.34)	426.0
1.41	3.0 (0.21)	4.0 (0.29)	2.0 (0.14)	5.0 (0.36)	14.0
1.73	54.0 (0.14)	132.0 (0.35)	113.0 (0.30)	75.0 (0.20)	374.0
2.00	5.0 (0.45)	3.0 (0.27)	2.0 (0.18)	1.0 (0.09)	11.0
2.24	3.0 (0.33)	4.0 (0.44)	0.0 (0.00)	2.0 (0.22)	9.0
2.45	101.0 (0.30)	142.0 (0.42)	49.0 (0.15)	43.0 (0.13)	335.0
Total	190.0	474.0	412.0	527.0	1603.0

The starting values for all parameters are given next. Starting values for the predictors TxDrug, SqrtWeek and TX*SWeek are given in the first line (covariates), while the starting value for the variance component associated with the random level-2 intercept is given in the second line (var. terms). In 18% of the subjects, no change in the category assigned for measurements was observed, as indicated by the last two lines shown below. Since the first threshold is fixed at 0 for identification purposes, starting values for the second and third thresholds only are listed.

SuperMix - [schizx2.out]

File Analysis Window Help

Starting values

covariates	0.000	0.000	0.000	0.000
var. terms	0.316			
thresholds	1.037	1.700		

==> The number of level 2 observations with non-varying responses
= 79 (18.08 percent)

The output summarizing the estimated parameters after convergence is shown next. Eighteen iterations were required to obtain convergence. The estimates are shown in the column with heading Estimate, and correspond to the coefficients α_0 , α_1 , α_2 , and α_3 in the model specification. The effect of TxDrug was the only nonsignificant effect.


```

-----
* Final Results - Maximum Marginal Likelihood Estimates *
-----
Total Iterations = 16
Quad Pts per Dim = 20
Log Likelihood = -1699.739
Deviance (-2logL) = 3399.479
Ridge = 0.000

Variable      Estimate      Stand. Error      Z      p-value
-----
TxDrug        -0.05128      0.17780          -0.28840  0.77304 (2)
SqrtWeek      -0.45912      0.06796          -6.75582  0.00000 (2)
Tx*SWeek      -0.67222      0.07481          -8.98573  0.00000 (2)
intcept       3.36602      0.18668          18.03054  0.00000 (2)

Random effect variance term (standard deviation)
intcept       1.10776      0.07010          15.80332  0.00000 (1)

Thresholds (for identification: threshold 1 = 0)
2             1.72926      0.07030          24.59741  0.00000 (1)
3             2.93962      0.08989          32.70065  0.00000 (1)

note: (1) = 1-tailed p-value
      (2) = 2-tailed p-value
-----
Save As...  Close

```

The variation in the intercept over the subjects is estimated as $1.10776^2 = 1.2271$, 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 $[0, +\infty)$, the p -values used for these effects are 1-tailed. No estimate of the level-1 variance is given, as there is no single level-1 variance for this model.

The results indicate that the treatment groups do not differ significantly at baseline (TxDrug 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 significance of, the Tx*SWeek interaction which forms part of the model.

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

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

where it is assumed that $\gamma_0 = -\infty$ and $\gamma_{NCAT} = +\infty$. In addition, γ_1 is usually set to 0 to avoid identification problems. For the present example, $NCAT = 4$, and from the output we see that $\hat{\gamma}_2 = 1.72926$ and $\hat{\gamma}_3 = 2.93962$. These values are used in combination with the coefficients of TxDrug, SqrtWeek, and Tx*SWeek to calculate estimated outcomes for different groups of patients.

Estimated unit-specific probabilities

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

$$\hat{z}_{ik} = 3.36602 - 0.05128 \text{TxDrug}_{ik} - 0.45912 \text{SqrtWeek}_{ik} - 0.67222 \text{Tx} \times \text{SWeek}_{ik}.$$

For a typical patient from the control group (TxDrug = 0, SqrtWeek = Tx*SWeek = 0),

$$\hat{z}_{ik} = 3.36602.$$

The probabilities that a typical patient from the control group responded in a specific category at the start of the study are obtained by noting that $\gamma_0 = -\infty$, $\gamma_1 = 0$, $\hat{\gamma}_2 = 1.72926$, and $\hat{\gamma}_3 = 2.93962$.

Let $p = \Phi(a) = P(Z \leq a)$ where Z has a standard normal distribution. To compute p for a given value of a , use is made of a probability calculator. These calculators can, for example, be found on the internet.

Since $\hat{z}_{ik} = 3.36602$, it follows that

$$\begin{aligned}\Phi(\gamma_0 - z_{ik}) &= \Phi(-\infty) = 0 \\ \Phi(\gamma_1 - z_{ik}) &= \Phi(-3.36602) = 0.0004 \\ \Phi(\gamma_2 - z_{ik}) &= \Phi(-1.63676) = 0.0508 \\ \Phi(\gamma_3 - z_{ik}) &= \Phi(-0.42640) = 0.3349.\end{aligned}$$

Therefore

$$\begin{aligned}P(\text{Imps79O} = 1) &= \Phi(\gamma_1 - z_{ik}) - \Phi(\gamma_0 - z_{ik}) \\ &= 0.0004 \\ P(\text{Imps79O} = 2) &= \Phi(\gamma_2 - z_{ik}) - \Phi(\gamma_1 - z_{ik}) \\ &= 0.0508 - 0.0004 \\ &= 0.0504\end{aligned}$$

$$\begin{aligned}
P(\text{Imps79O} = 3) &= \Phi(\gamma_3 - z_{ik}) - \Phi(\gamma_2 - z_{ik}) \\
&= 0.3349 - 0.0508 \\
&= 0.2841
\end{aligned}$$

and

$$\begin{aligned}
P(\text{Imps79O} = 4) &= 1 - P(\text{Imps79O} = 3) - P(\text{Imps79O} = 2) - P(\text{Imps79O} = 1) \\
&= 1 - 0.2841 - 0.0504 - 0.0004 \\
&= 0.6651.
\end{aligned}$$

for the four categories of Imps79O. We can compare these predicted probabilities with the observed data for the control group as shown in Table 3.3 below. The ratio of responses in the categories of Imps79O when SqrtWeek = 0 are 0.000, 0.102, 0.336 and 0.561 respectively. For example, if Imps79O = 4, the ratio is $\frac{60}{107}$.

Table 3.3: Crosstabulation of Imps79O and SqrtWeek for control group

		SqrtWeek						Total	
		.0000	1.0000	1.4142	1.7321	2.0000	2.2361		2.4495
Imps79O	1	0	2	0	3	0	0	8	13
	2	11	25	0	24	0	1	24	85
	3	36	26	1	26	1	0	16	106
	4	60	52	4	34	1	1	22	174
Total		107	105	5	87	2	2	70	378

At the end of the study, the corresponding ratios for the observed data were 0.00, 0.10, 0.34, and 0.56 respectively. Here, the variable SqrtWeek = 2.4495 and thus

$$\begin{aligned}
\hat{z}_{ik} &= 3.36602 - 0.45912 \text{ SqrtWeek}_{ik} \\
&= 3.36602 - (0.45912)(2.4495) \\
&= 2.24141.
\end{aligned}$$

The probabilities that a typical patient from the control group responded in a specific category at the conclusion of the study period are obtained as

$$\Phi(\gamma_1 - z_{ik}) = \Phi(-2.24141) = 0.01250$$

$$\Phi(\gamma_2 - z_{ik}) = \Phi(-0.51214) = 0.30428$$

$$\Phi(\gamma_3 - z_{ik}) = \Phi(0.69821) = 0.75748.$$

Therefore

$$\begin{aligned} P(\text{Imps79O} = 1) &= \Phi(\gamma_1 - z_{ik}) - \Phi(\gamma_0 - z_{ik}) \\ &= 0.01250 \end{aligned}$$

$$\begin{aligned} P(\text{Imps79O} = 2) &= \Phi(\gamma_2 - z_{ik}) - \Phi(\gamma_1 - z_{ik}) \\ &= 0.30428 - 0.01250 \\ &= 0.29178 \end{aligned}$$

$$\begin{aligned} P(\text{Imps79O} = 3) &= \Phi(\gamma_3 - z_{ik}) - \Phi(\gamma_2 - z_{ik}) \\ &= 0.75748 - 0.30428 \\ &= 0.45320 \end{aligned}$$

and

$$\begin{aligned} P(\text{Imps79O} = 4) &= 1 - P(\text{Imps79O} = 3) - P(\text{Imps79O} = 2) - P(\text{Imps79O} = 1) \\ &= 1 - (0.45320 + 0.29178 + 0.01250) \\ &= 0.24252. \end{aligned}$$

The estimated probabilities for a typical patient from the treatment group at the start of the study (TxDrug = 1, SqrtWeek = Tx*SWeek = 0) can be calculated in a similar fashion, using

$$\begin{aligned} \hat{z}_{ik} &= 3.36602 - 0.05128 \\ &= 3.31474. \end{aligned}$$

The probabilities that a typical patient from the treatment group responded in a specific category at the start of the study are obtained as

$$\Phi(\gamma_1 - z_{ik}) = \Phi(-3.31474) = 0.00046$$

$$\Phi(\gamma_2 - z_{ik}) = \Phi(-1.58548) = 0.05643$$

$$\Phi(\gamma_3 - z_{ik}) = \Phi(-0.37512) = 0.35379.$$

Therefore

$$\begin{aligned} P(\text{Imps79O} = 1) &= \Phi(\gamma_1 - z_{ik}) - \Phi(\gamma_0 - z_{ik}) \\ &= 0.00046 \end{aligned}$$

$$\begin{aligned}
P(\text{Imps79O} = 2) &= \Phi(\gamma_2 - z_{ik}) - \Phi(\gamma_1 - z_{ik}) \\
&= 0.05643 - 0.00046 \\
&= 0.05597
\end{aligned}$$

$$\begin{aligned}
P(\text{Imps79O} = 3) &= \Phi(\gamma_3 - z_{ik}) - \Phi(\gamma_2 - z_{ik}) \\
&= 0.35379 - 0.05643 \\
&= 0.29736
\end{aligned}$$

and

$$\begin{aligned}
P(\text{Imps79O} = 4) &= 1 - P(\text{Imps79O} = 3) - P(\text{Imps79O} = 2) - P(\text{Imps79O} = 1) \\
&= 1 - 0.29736 - 0.05597 - 0.00046 \\
&= 0.64621
\end{aligned}$$

for the four groups defined by the categories of Imps79O.

At the end of the study, the corresponding ratios for the observed data were 0.35, 0.45, 0.12, and 0.08 respectively. Here, the variable SqrtWeek = 2.4495 and thus

$$\begin{aligned}
\hat{z}_{ik} &= 3.36602 - 0.05128 \text{ TxDrug}_{ik} - 0.45912 \text{ SqrtWeek}_{ik} - 0.67222 \text{ Tx} * \text{SWeek}_{ik} \\
&= 3.36602 - 0.05128 - (0.45912)(2.4495) - (0.67222)(2.4495) \\
&= 0.54354.
\end{aligned}$$

The probabilities that a typical patient from the treatment group responded in a specific category at the conclusion of the study period are obtained as

$$\Phi(\gamma_1 - z_{ik}) = \Phi(-0.54354) = 0.29338$$

$$\Phi(\gamma_2 - z_{ik}) = \Phi(1.18572) = 0.88213$$

$$\Phi(\gamma_3 - z_{ik}) = \Phi(2.39608) = 0.99171.$$

Therefore

$$\begin{aligned}
P(\text{Imps79O} = 1) &= \Phi(\gamma_1 - z_{ik}) - \Phi(\gamma_0 - z_{ik}) \\
&= 0.29338
\end{aligned}$$

$$\begin{aligned}
P(\text{Imps79O} = 2) &= \Phi(\gamma_2 - z_{ik}) - \Phi(\gamma_1 - z_{ik}) \\
&= 0.88213 - 0.29338 \\
&= 0.58875
\end{aligned}$$

$$\begin{aligned}
P(\text{Imps79O} = 3) &= \Phi(\gamma_3 - z_{ik}) - \Phi(\gamma_2 - z_{ik}) \\
&= 0.99171 - 0.88213 \\
&= 0.10957
\end{aligned}$$

and

$$\begin{aligned}
P(\text{Imps79O} = 4) &= 1 - P(\text{Imps79O} = 3) - P(\text{Imps79O} = 2) - P(\text{Imps79O} = 1) \\
&= 1 - 0.10957 - 0.58875 - 0.29339 \\
&= 0.008.
\end{aligned}$$

Estimated population-average probabilities

In this section, we defined the latent response variable model as

$$y_{ik} = \mathbf{x}'_{ik} \boldsymbol{\beta}_i + \mathbf{w}'_{ik} \boldsymbol{\alpha} + e_{ik}, \quad k = 1, 2, \dots, n_i,$$

where \mathbf{x}_{ik} denotes a design vector for the random effects $\boldsymbol{\beta}_i$ and \mathbf{w}_{ik} the design vector for the predictors in the fixed part of the model with corresponding vector $\boldsymbol{\alpha}$ of regression parameters. Denote the covariance matrix of $\boldsymbol{\beta}_i$ by $\boldsymbol{\Sigma}_{\beta_i}$ and the variance of e_{ik} by σ^2 .

For a probit link function $\sigma^2 = 1$, and for a logistic link function $\sigma^2 = \pi^2/3$. Under the assumption that $\boldsymbol{\beta}_i$ and e_{ik} are independently distributed, it follows that

$$\sigma_{y_{ik}}^2 = \mathbf{x}'_{ik} \boldsymbol{\Sigma}_{\beta_i} \mathbf{x}_{ik} + \sigma^2.$$

Let

$$d_{ik} = \frac{\sigma_{y_{ik}}^2}{\sigma^2}.$$

The quantity d_{ik} is called the design effect. The estimated population-average probabilities (Hedeker & Gibbons, 2006) are obtained in a similar fashion as the unit-specific probabilities, but with replacing $\hat{\gamma}_j - \hat{z}_{ik}$ by $\left(\hat{\gamma}_j - \hat{z}_{ik} \right) / \sqrt{d_{ik}}$.

To illustrate, we calculate the estimated population-average probabilities for the control group associated at the end of the study.

From the output, we have $\text{var}(\beta_{i1}) = (1.10776)^2$, where β_{i1} denotes the intercept coefficient. In this case, $\mathbf{x}'_{ik} = \mathbf{1}$ and hence, with $\sigma^2 = 1$ for the probit link,

$$\sigma^2_{y_{ik}} = 1 \times 1.227 \times 1 + 1 = 2.227.$$

Therefore

$$d_{ik} = \frac{2.227}{1} = 2.227.$$

Using the unit-specific values for $\hat{\gamma}_j - \hat{z}_{ik}$ obtained previously, it follows that

$$\begin{aligned} P(\text{Imps79O} = 1) &= \Phi(-2.2414 / \sqrt{2.227}) \\ &= 0.06655 \end{aligned}$$

$$\begin{aligned} P(\text{Imps79O} = 2) &= \Phi(-0.51214 / \sqrt{2.227}) - \Phi(-1.50197) \\ &= 0.36573 - 0.06655 \\ &= 0.29918 \end{aligned}$$

$$\begin{aligned} P(\text{Imps79O} = 3) &= \Phi(0.69821 / \sqrt{2.227}) - \Phi(-0.34319) \\ &= 0.68006 - 0.36573 \\ &= 0.31433 \end{aligned}$$

and

$$\begin{aligned} P(\text{Imps79O} = 4) &= 1 - P(0.0665 + 0.29918 + 0.31433) \\ &= 0.31994. \end{aligned}$$

The estimated population-average probabilities for the control group are summarized in Table 3.4 below.

Table 3.4: Estimated unit-specific probabilities for control group

		SqrtWeek						Total
		Start of study			End of study			
		Obs. Freq	Ratio	Pred. Prob	Obs. Freq	Ratio	Pred. Prob	
Imps79O	1	0	0	0.01	8	0.11	0.07	8
	2	11	0.10	0.12	24	0.34	0.30	35
	3	36	0.34	0.25	16	0.23	0.31	52
	4	60	0.56	0.61	22	0.31	0.32	82
Total		107			70			177

A comparison of these probabilities with the observed ratios given in Table 3.4 for the control group at the end of the study indicate that the population-average results are closer to the observed ratios than is the case for the unit-specific results.

Results for the treatment group are obtained in a similar way and are summarized in Table 3.5 below.

Table 3.5: Estimated population-average probabilities for treatment group

		SqrtWeek						Total
		Start of study			End of study			
		Obs. Freq	Ratio	Pred. Prob	Obs. Freq	Ratio	Pred. Prob	
Imps790	1	1	0.003	0.01	93	0.35	0.36	94
	2	43	0.13	0.13	118	0.45	0.43	161
	3	86	0.26	0.26	33	0.12	0.16	119
	4	197	0.60	0.60	21	0.08	0.05	218
Total		327			287			592

It is interesting to note that the observed frequencies at the start and end of the study are 107 and 70 for the control group, and 327 and 287 for the experimental group. The larger number of observations available in the latter case may partially explain why the predicted probabilities are closer to the observed probabilities for the experimental group when compared to those for the control group. The results clearly demonstrate the effectiveness of a seven week treatment, and the significant drug by time interaction effect: at the end of the study, the estimated population-average proportion of patients in categories 1 or 2 (normal or borderline) are 0.37 for the control group and 0.79 for the treatment group. The corresponding estimated population-average probabilities at the start of the study are 0.13 and 0.14 for the control and experimental groups respectively.

Finally, the estimated intraclass correlation (which in this case is the *intrapatient* correlation) for this analysis is given, along with the variation in intercepts over patients.

```

Calculation of the intraclass correlation
-----
residual variance = 1 (assumed)
cluster variance = (1.108 * 1.108) = 1.227

intraclass correlation = 1.227 / ( 1.227 + 1.000) = 0.551

```

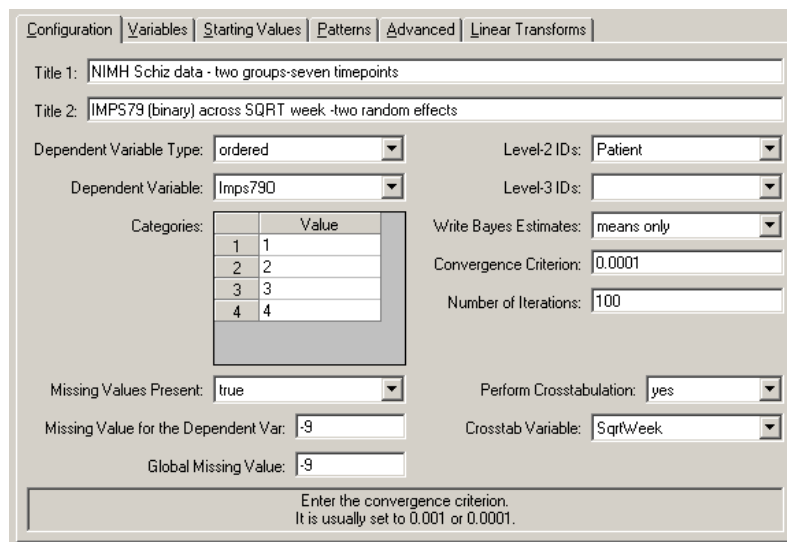
Save As... Close

Example: Probit link function with random intercept and slope

To fit a model allowing patients to vary in terms of both their intercept and their trend across time (random intercepts and slopes), the model in the previous example is extended to include both intercept and slope (SqrtWeek) effects.

Setting up the analysis

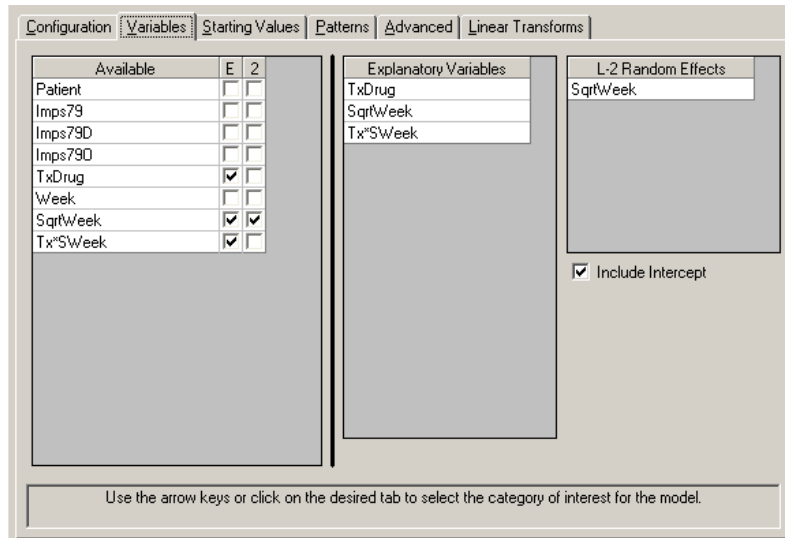
The model is fitted to the data in **schizx1.ss3**, and the model specification file **schizx2.mum** used previously is modified to include a random slope. The first step is to open the **ss3** file, and then the model specification file **schizx3.mum**. This is accomplished by using the **File, Open** option to activate the display of an **Open** dialog box, and browsing for the file **schizx1.ss3** in the **Examples** folder. Next, use the **File, Open Existing Model Setup** option to activate an **Open** dialog box and browse for the file **schizx2.mum** in the **Examples** folder.



Categories:	Value
1	1
2	2
3	3
4	4

The **Model Setup** window, with entries corresponding to the contents of the model specification file **schizx2.mum** is displayed. After adjusting the title in the **Title 2** text box of the **Configuration** screen to reflect the intended changes to the model, proceed to the **Variables** screen by clicking on this tab.

The only change needed on the **Variables** screen is to add the variable SqrtWeek to the **L-2 Random Effects** grid. This is done by checking the check box next to this variable's name under the **2** column of the **Available** grid. Doing so leads to the display of the variable name SqrtWeek in the **L-2 Random Effects** grid as shown below.



Before running the analysis, the model specifications have to be saved. Select the **File, Save** option, and provide a new name, for example **schizx3.mum**, for the model specification file. Run the analysis by selecting the **Run** option from the **Analysis** menu.

Discussion of results

Portions of the output file **schizx3.out** are shown below. The estimates are shown in the column with heading Estimate, and correspond to the coefficients α_0 , α_1 , α_2 , and α_3 in the model specification. Results for the fixed effects in this model are similar to those obtained for the random intercepts model. There is significant random variation in the random intercepts and slopes of the patients. For this model, we have $\hat{\gamma}_2 = 2.17905$ and $\hat{\gamma}_3 = 3.64340$, both somewhat higher than those for the previous model ($\hat{\gamma}_2 = 1.72926$ and $\hat{\gamma}_3 = 2.93962$.)

```

-----
* Final Results - Maximum Marginal Likelihood Estimates *
-----
Total Iterations = 27
Quad Pts per Dim = 20
Log Likelihood = -1663.524
Deviance (-2logL) = 3327.049
Ridge = 0.100

Variable      Estimate      Stand. Error      Z      p-value
-----
TxDrug        0.03242      0.22286          0.14547  0.88434 (2)
SqrtWeek     -0.50590      0.12934         -3.91143  0.00009 (2)
Tx*SWeek     -0.94264      0.14762         -6.38557  0.00000 (2)
intcept      4.10110      0.25090         16.34523  0.00000 (2)

Random effect variance & covariance terms (Cholesky of var-covariance matrix)
intcept      1.47374      0.14084         10.46387  0.00000 (1)
covariance   -0.30708      0.09008         -3.40888  0.00065 (2)
SqrtWeek     0.72456      0.06985         10.37301  0.00000 (1)

Thresholds (for identification: threshold 1 = 0)
2            2.17905      0.10995         19.81824  0.00000 (1)
3            3.64340      0.14410         25.28399  0.00000 (1)

note: (1) = 1-tailed p-value
      (2) = 2-tailed p-value

```

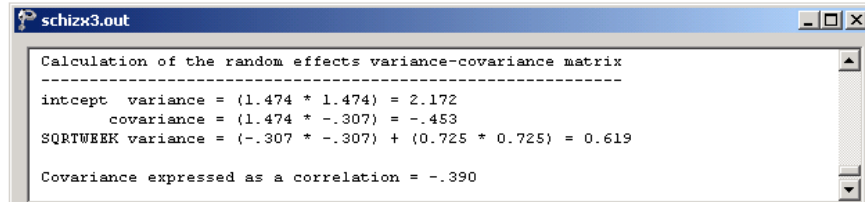
To compare this model to the random intercepts model previously fitted, the likelihood-ratio χ^2 test can be used. Based on this test, there is clear evidence of significant variation in the linear time trends:

$$\chi^2 = -2[-1699.739 - (-1663.354)] = 72.77,$$

(2 degrees of freedom and $p < 0.001$) over and above the individual intercept variation. Significant negative association between the intercept and linear time terms is indicated, suggesting that those patients with the highest initial severity show the greatest improvement across time (e.g., largest negative time trends).

Note that variances are, by definition, non-negative. It was shown by Miller (1977) and Self & Liang (1987) that, when testing the hypothesis that the variance of a random intercept or slope is zero, the alternative hypothesis is one-sided. This implies that an approximate p -value is obtained as half the p -value obtained from a χ^2 statistic based on the difference in deviances. Snijders & Bosker (1999), compares a random intercept and slope model with a random intercept model. For their example, $\chi^2 = 8.5$ at 2 degrees of freedom with tail probability of $p < 0.02$, so halving the p -value yields a significant outcome with $p < 0.01$. Generally speaking, the likelihood-ratio test applied to random coefficients without the adjustment of p -value is conservative.

Finally, the estimated intracluster correlation (which in this case is the *intrapatient* correlation) for this analysis is given. The calculation of variance and covariance components is shown, and we note that most of the variation in *lmps790* is in intercept, rather than in the slope of *SqrtWeek* (2.172 vs. 0.619).



```
schizx3.out
-----
Calculation of the random effects variance-covariance matrix
-----
intcept variance = (1.474 * 1.474) = 2.172
covariance = (1.474 * -.307) = -.453
SQRWEEK variance = (-.307 * -.307) + (0.725 * 0.725) = 0.619
Covariance expressed as a correlation = -.390
```