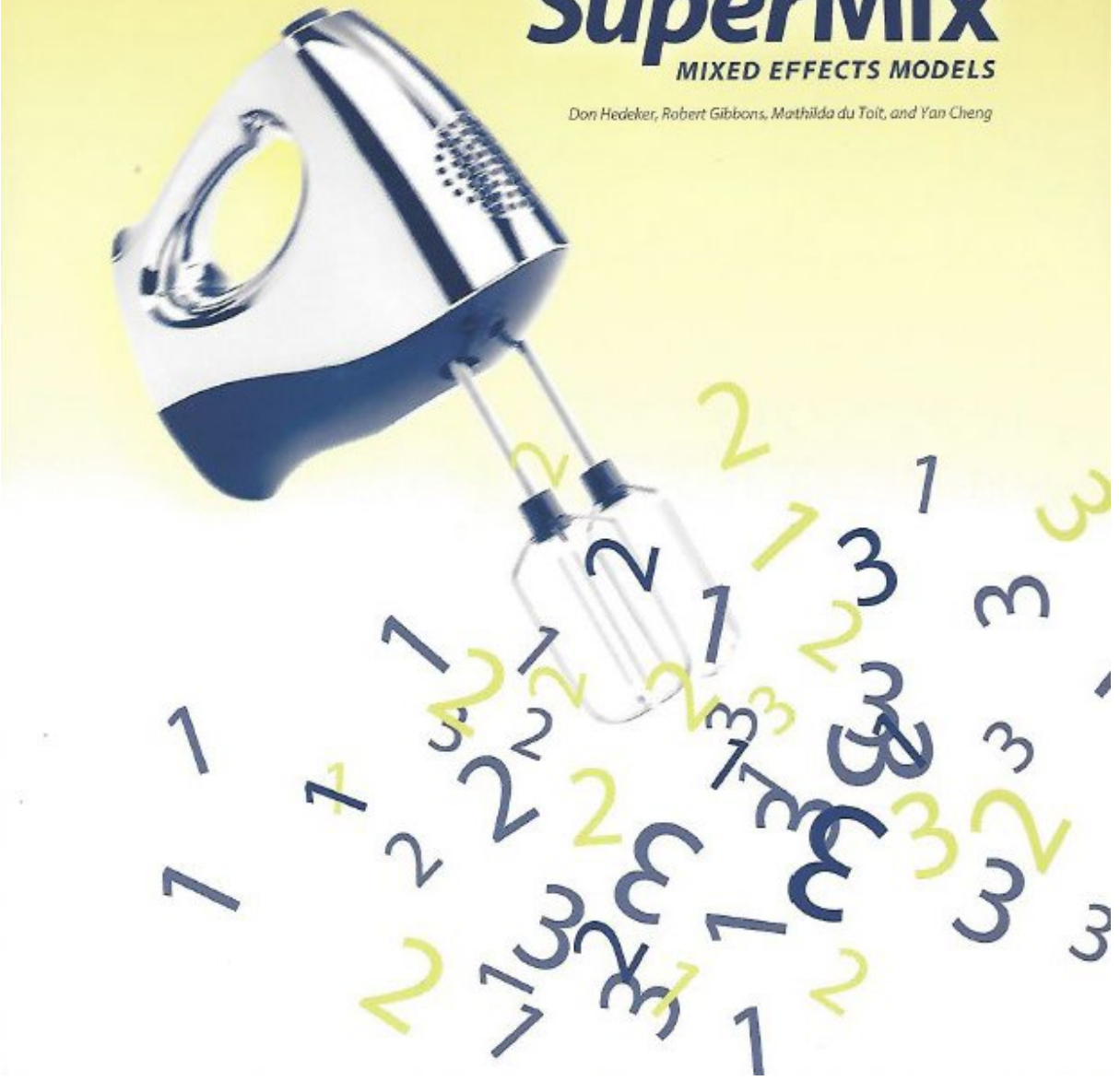
SSI SCIENTIFIC SOFTWARE INTERNATIONAL

SuperMix

MIXED EFFECTS MODELS

Don Hedeker, Robert Gibbons, Mathilda du Toit, and Yan Cheng



Cover by Clint Smith.

1 2 3 4 5 6 7 8 9 0 11 10 09 08

ISBN: 978-0-89498-056-5 (paper)

Please download the examples from our website and unzip them into a convenient folder location.

Table of contents

List of tables and figures

1 Introduction to mixed-effects models

Hierarchical structures are often encountered in numerous research areas. Consider, for example, the study of the effect of administering medication, such as an antidepressant, over time to a patient diagnosed with depression. For each patient, the effect of the drug over time can be modeled in terms of the time since the start of treatment, and also in terms of any other information obtained at the time of each measurement during the study. Measures of family support at the time of measurement can also be incorporated into such a model. The outcome would be described as a function of the information collected at the measurement level, and could be viewed as a measurement-level model for each individual patient. However, the gender of the patient, and other characteristics that may influence the outcome but that do not change over time, cannot easily be accommodated in the model proposed, as the model is at a measurement, rather than a patient, level. It may also be of interest to compare patients in terms of their improvement trajectories, which is easier when outcomes are described in terms of patients rather than measurements.

To allow us to study all of these areas of interest simultaneously, a model that acknowledges the data's inherent hierarchical structure (measurements nested within individual patients), and allows the study of both measurement- and patient-level models along with the way these models are related to each other, is needed. As patients may drop out during the study period, the model should also be suitable for the analysis of unbalanced longitudinal data where each individual may be measured at a different number of occasions, or even at different time points.

In this chapter, data from a study described in Vonesh & Carter (1992) that focused on the assessment of high-flux hemodialyzers' *in vivo* ultrafiltration are used to illustrate the need for and basic characteristics of a mixed-effects regression model. While the eventual application of these findings will be in a medical field, the testing of the dialyzers discussed here may be of interest to any researcher who intends modeling repeated measures data.

The ultrafiltration rates of 20 high-flux dialyzers were measured at seven different transmembrane pressures. The unit of measurement for transmembrane pressure was dmHg, and the filtration rate was recorded in mL/hr. These data, also analyzed in Littell, Milliken, Stroup & Wolfinger (1996), are perfectly balanced in that all seven measurements are available for each of the hemodialyzers. The hemodialyzers, machines for filtering impurities from the blood, are the units within which the actual measurements are nested. Data for 10 of the dialyzers are shown in Table 1.1.

Table 1.1: Data for 10 hemodialyzers from Vonesh & Carter data

Device ID	Supply	Pressure	Rate	Device ID	Supply	Pressure	Rate
11.000	1.000	28.500	1.500	16.000	1.000	23.500	3.600
11.000	1.000	52.000	15.400	16.000	1.000	48.000	20.490
11.000	1.000	100.500	32.520	16.000	1.000	101.000	41.880
11.000	1.000	150.000	42.440	16.000	1.000	149.000	49.990
11.000	1.000	198.500	48.570	16.000	1.000	199.000	57.670
11.000	1.000	249.000	53.690	16.000	1.000	248.000	62.480
11.000	1.000	299.500	53.660	16.000	1.000	300.500	62.150
12.000	1.000	29.500	6.420	17.000	1.000	23.500	1.170
12.000	1.000	51.500	20.250	17.000	1.000	48.500	17.680
12.000	1.000	101.000	43.050	17.000	1.000	102.500	39.700
12.000	1.000	148.000	58.110	17.000	1.000	151.500	52.680
12.000	1.000	200.000	61.990	17.000	1.000	199.000	61.800
12.000	1.000	248.000	60.910	17.000	1.000	251.000	61.480
12.000	1.000	300.500	63.600	17.000	1.000	302.000	61.420
13.000	1.000	25.500	3.880	18.000	1.000	26.000	1.890
13.000	1.000	50.000	19.160	18.000	1.000	51.500	18.510
13.000	1.000	98.000	37.650	18.000	1.000	97.000	37.220
13.000	1.000	149.000	47.900	18.000	1.000	150.500	52.350

Table 1.1: Data for 10 hemodialyzers from Vonesh & Carter data (continued)

13.000	1.000	201.500	54.490	18.000	1.000	199.000	60.910
13.000	1.000	251.000	53.170	18.000	1.000	250.000	62.980
13.000	1.000	298.000	59.350	18.000	1.000	299.500	64.770
14.000	1.000	40.000	10.940	19.000	1.000	35.500	10.410
14.000	1.000	47.000	13.470	19.000	1.000	48.000	19.320
14.000	1.000	101.000	35.350	19.000	1.000	102.500	43.770
14.000	1.000	151.500	45.340	19.000	1.000	150.000	51.230
14.000	1.000	198.000	49.440	19.000	1.000	199.000	58.090
14.000	1.000	251.000	53.630	19.000	1.000	250.000	54.090
14.000	1.000	300.000	56.430	19.000	1.000	300.500	62.010
15.000	1.000	29.000	4.050	20.000	1.000	28.000	5.710
15.000	1.000	49.500	16.590	20.000	1.000	50.500	20.500
15.000	1.000	101.500	40.520	20.000	1.000	100.000	39.410
15.000	1.000	152.000	52.840	20.000	1.000	149.000	50.100
15.000	1.000	202.000	60.440	20.000	1.000	200.000	55.160
15.000	1.000	250.000	64.830	20.000	1.000	250.500	61.190
15.000	1.000	297.500	63.830	20.000	1.000	302.000	50.720

Of interest here is the relationship between the ultrafiltration rate, denoted as Rate in Table 1.1, and the associated transmembrane pressure, indicated as Pressure in the table. The blood flow rate, as represented by the column with header Supply, is also of potential interest.

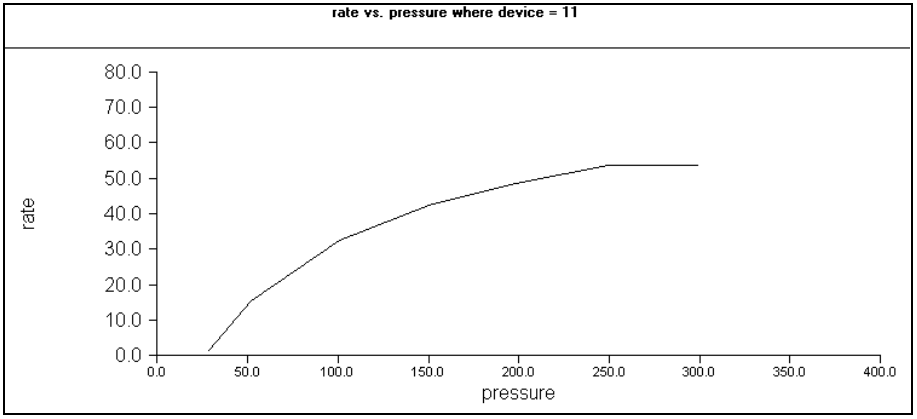
The data as a whole can be viewed as having a hierarchical structure, with measurement-related characteristics of the hemodialyzers at seven measurement occasions; all measurements for each dialyzer are therefore *nested* within that dialyzer. The dialyzers, in turn, form the next level of the hierarchy, and any machine-specific characteristics may be used as potential predictors at this level.

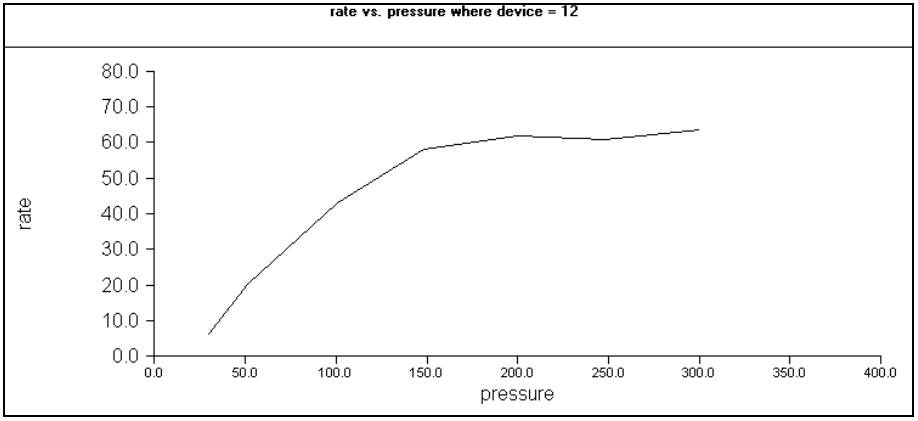
Fixed-effects regression ignoring data clustering

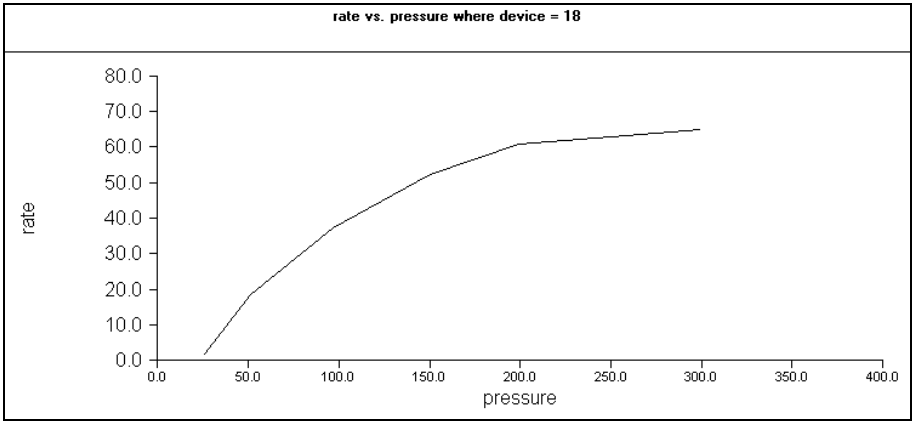
Before proceeding with a mixed-effects analysis of these data, we first look at a fixed-effects analysis that ignores the clustering of measurements within dialyzers. Note that SuperMix can be used for this purpose, and that the analysis is essentially equivalent to performing a traditional multiple linear regression analysis using maximum likelihood, and not least squares, estimation.

Using the information for the second set of 10 dialyzers, for which 70 measurements were available, we now explore the relationship between the Rate of filtration, which serves as our outcome variable, and the transmembrane Pressure at which the measurement was made. Line plots of this relationship for some of the dialyzers are shown in Figure 1.1. These graphs were obtained using SuperMix's exploratory graphs option. Detailed information on how to create such graphs are given elsewhere in the manual.

It is clear from these graphs that the relationship between the observed Rate and Pressure at which the measurement was made will be inadequately described by a first-order polynomial. For dialyzer 12 the slope of the line is steep initially, but the curve flattens out at a pressure of about 100 dmHg. This trend is not as clearly observed for the other dialyzers. Also, there seems to be evidence of differences in the rates of dialyzers 18, 19, and 20 towards the higher end of the pressure scale. We conclude that a higher-order polynomial will probably offer a better description of the relationship, and that it may also be wise to make provision for differences between devices (dialyzers).


rate vs. pressure where device = 11
rate
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
0.0
50.0
100.0
150.0
200.0
250.0
300.0
350.0
400.0
pressure


rate vs. pressure where device = 12
rate
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
0.0
50.0
100.0
150.0
200.0
250.0
300.0
350.0
400.0
pressure


rate vs. pressure where device = 18
rate
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
0.0
50.0
100.0
150.0
200.0
250.0
300.0
350.0
400.0
pressure

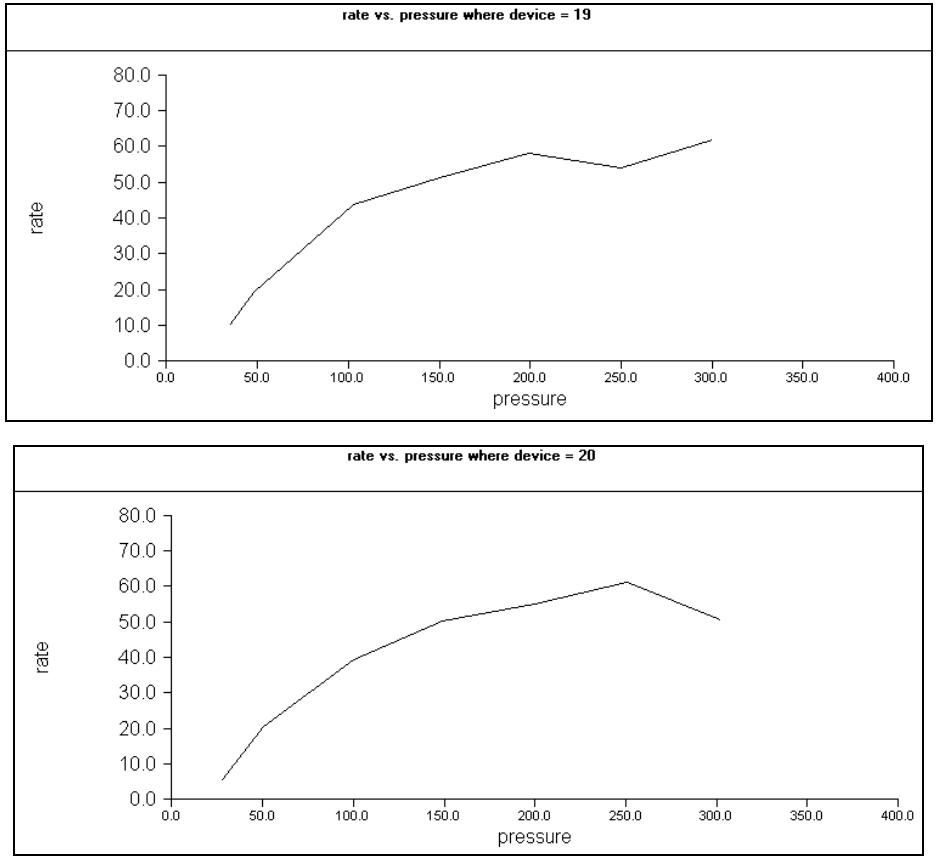


Figure 1.1: Exploratory graphs of rate versus pressure for hemodialyzers

Figure 1.2 represents the same lines for all ten dialyzers simultaneously. While there seems to be little difference in their behavior at the lower level of the pressure scale, the divergence in the plotted lines at higher pressure levels can be seen clearly.

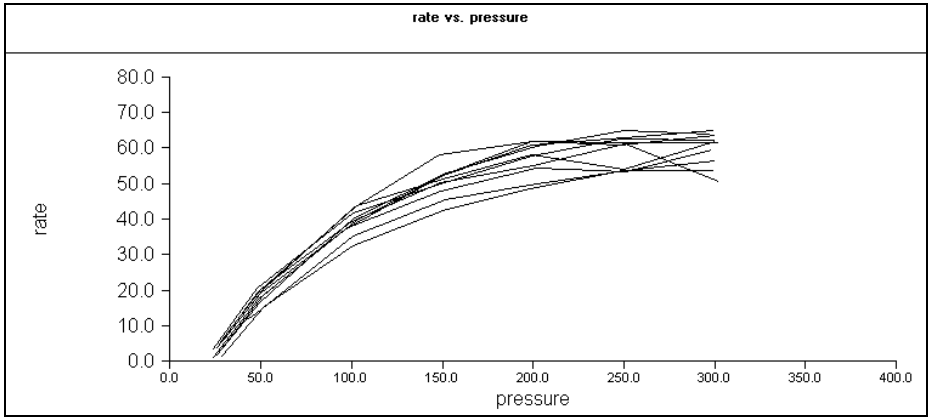


Figure 1.2: Exploratory graphs of rate versus pressure for 10 hemodialyzers

In terms of the variables shown in Table 1.1, we now fit a model of the form

$$y_{ij} = \beta_0 + \beta_1 \left(\text{PRESSURE}\right)_{ij} + \beta_2 \left(\text{PRESSURE}\right)_{ij}^2 + e_{ij} \qquad (1.1)$$

where y_{ij} denotes the Rate measurement at time j (j = 1, 2, 3, …7) for hemodialyzer i. $(\text{PRESSURE})_{ij}$ indicates the associated transmembrane pressure, $(\text{PRESSURE})_{ij}^2$ the squared value of the pressure, and e_{ij} measurement error. The coefficients β_0, β_1, and β_2 are the fixed, but unknown, parameters to be estimated. The e_{ij} are assumed to have a normal distribution, with mean 0 and variance σ^2.

For this analysis, we obtain estimates of β_0, β_1, and β_2 of –6.5847, 0.5281 and –0.0011 respectively. The estimated Rate is plotted against pressure in Figure 1.3. In addition, an estimate of σ^2 of 41.34095 was obtained. The results show that the average predicted Rate, $\hat{\beta}_0$, at a pressure of zero is –6.5847. However, a value of 0 is outside the range of 23.50 to 303.00 of observed pressure values. As such, the interpretation of the estimate of β_0 in this context is difficult, and we would rather look at the predicted rate for the lowest observed pressure. Another alternative is to transform the values of the variables Pressure and Pressure^2 in such a way that

interpretation of the intercept estimate is meaningful. Examples of such transformations are given in the chapters to follow.

The coefficient representing the effect of the predictor Pressure, $\hat{\beta}_1$, indicates a predicted increase in Rate with increased pressure: an increase of 0.52807 mL/hr in the Rate is expected for each increase of 1 dmHg in transmembrane pressure. The coefficient β_1 is commonly referred to as a "slope" coefficient, as it indicates both the direction of the relationship between the predictor and the outcome, and the magnitude of the expected change in outcome associated with changes in the predictor.

Similarly, the relationship between the squared values of transmembrane pressure (Pressure^2) and the ultrafiltration rate is estimated to be negative: higher values of pressure are predicted to lead to lower predicted rates. The statistical significance of this estimated coefficient indicates that the relationship between pressure and filtration rate is not truly linear, and that the use of a higher-order polynomial may provide a better description of the data. However, while the estimates of β_1 and β_2 are of interest individually, when evaluating the relationship between the transmembrane pressure and the ultrafiltration rate, both estimates should be taken into account. A increase of 1 dmHg in pressure will lead to a change in expected filtration rate of 0.52807(1) –0.0011(1) = 0.52697. From this result, we conclude that while the filtration rate and pressure generally shows a positive relationship, this relationship is bound to change with increased pressure. The higher the pressure, the bigger the impact of the estimate of β_2 in the prediction of the rate through use of the formula

$$\hat{y}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 \left(\text{PRESSURE}\right)_{ij} + \hat{\beta}_2 \left(\text{PRESSURE}\right)_{ij}^2$$

The lowest observed pressure is 23.50, and the predicted rate of filtration is thus

$$\begin{aligned}\hat{y}_{ij} &= \hat{\beta}_0 + 0.52807(\text{PRESSURE})_{ij} - 0.0011(\text{PRESSURE})_{ij}^2 \\ &= \text{-}6.5847+0.52807(23.5)\text{-}0.0011(23.5)^2 \\ &= -6.5847 + 12.4096 - 0.6075 \\ &= 5.2174.\end{aligned}$$

For the highest observed pressure of 303, the predicted filtration rate follows as

$$\begin{aligned}\hat{y}_{ij} &= \text{-}6.5847+0.52807(303)\text{-}0.0011(303)^2 \\ &= -6.5847 + 160.0052 - 100.9899 \\ &= 52.4306.\end{aligned}$$

The fixed-effects regression line over all measurements is shown in Figure 1.3 below.

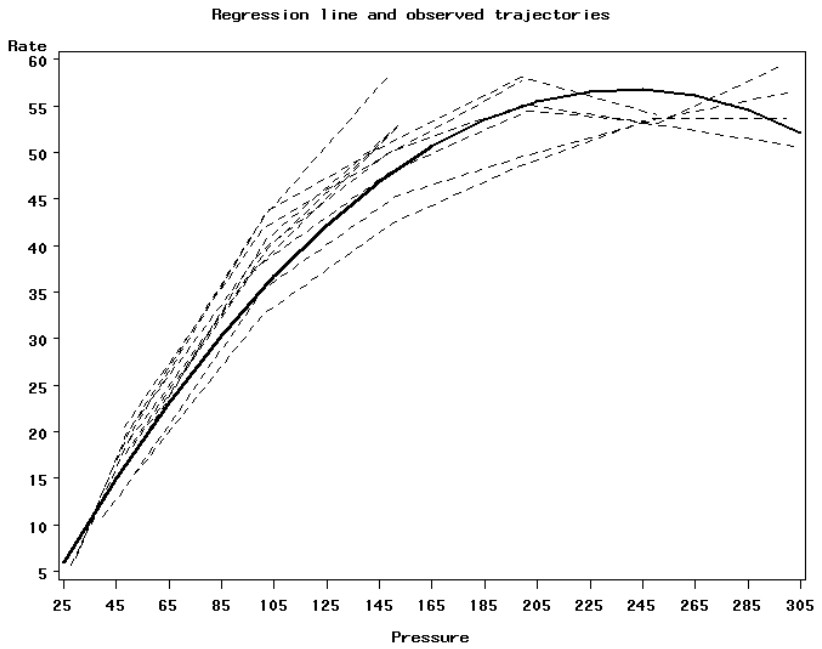


Figure 1.3: Fixed-effects regression line for 10 dialyzers

Fixed-effects regression including data clustering

As noted by Hedeker, Gibbons & Flay (1994) and others, ignoring the data clustering often results in statistical tests which are too liberal, resulting in falsely rejecting the null hypothesis too often. In terms of our data, where multiple measurements "belong" to each dialyzer, it is reasonable to expect that measurements for a given dialyzer may be more similar to each other than to any other measurement, regardless of the dialyzer it was obtained for. Thus, it may be reasonable to assume that the measurements for a given dialyzer may be correlated. In addition, if it is indeed true that the transmembrane pressure applied impacts on the transfer rate, ignoring the clustering effect may lead to erroneous conclusions concerning the relationship between pressure and transfer rate.

To start addressing these concerns, we modify the previous model to take the clustering of measurements within dialyzers into account. We do so by fitting a line similar to that given in Equation (1.1) for each individual dialyzer. Table 1.2 shows the estimates of β_0 and β_1 for individual dialyzers, and Figure 1.4 a graphical representation of the results.

The estimated coefficients for the intercepts and time slopes of the dialyzers ($\hat{\beta}_0$ and $\hat{\beta}_1$ respectively) in Table 1.2 show that the predicted intercepts of dialyzers differ considerably. Device/dialyzer number 15 has a predicted initial transfer rate of –10.885, which is considerably lower than the predicted initial rate of –3.645 for dialyzer 19. Recall that in the previous analysis, we obtained a value of –6.585 for $\hat{\beta}_0$, which does not provide an adequate description of the initial status of any of the dialyzers except perhaps dialyzers 13, 14, 16, and 20. A "one size fits all" policy for obtaining an estimate of the initial status of patients is clearly inadequate, and does not describe the initial status for individual dialyzers satisfactorily.

Table 1.2: Regression results for 10 dialyzers: taking clustering into account

Device	Intercept	Pressure	(Pressure)^2
11	–9.206	0.486	–0.001
12	–9.024	0.629	–0.001
13	–5.115	0.500	–0.001
14	–5.008	0.454	–0.001
15	–10.885	0.602	–0.001
16	–5.255	0.537	–0.001
17	–10.614	0.608	–0.001
18	–10.582	0.590	–0.001
19	–3.645	0.520	–0.001
20	–7.911	0.589	–0.001
overall	–6.585	0.528	–0.001

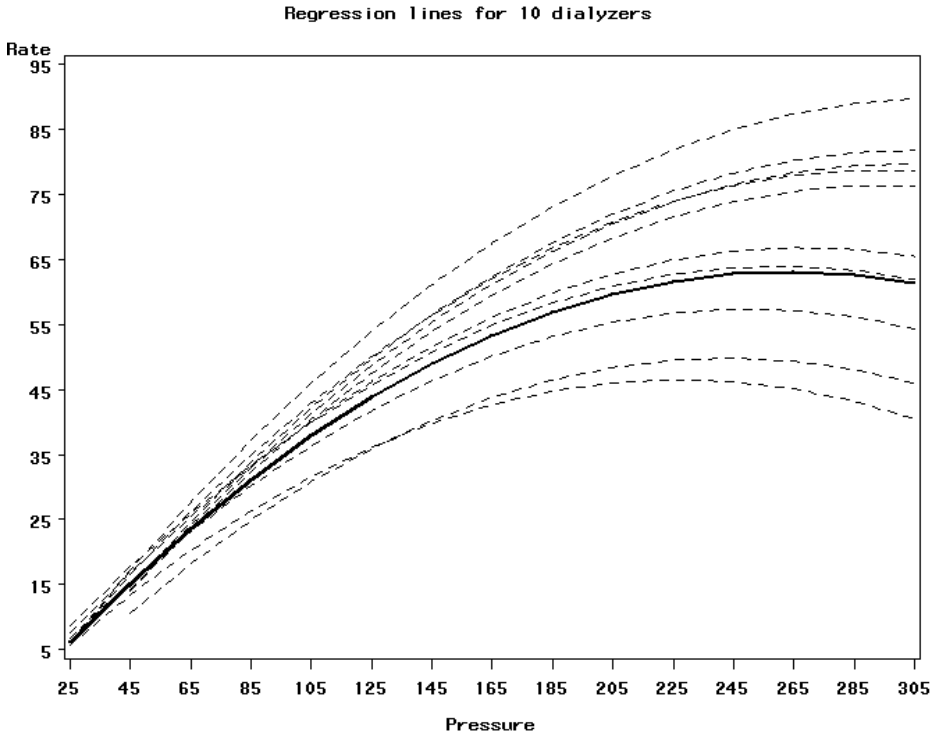


Figure 1.4: Individual fixed-effects regression lines for 10 dialyzers

This conclusion is also apparent from Figure 1.4. While the differences in transfer rates at the lower end of the pressure range are not as clear from the graph as they are in Table 1.2, the graph indicates even larger differences between the dialyzers at high transmembrane pressure. Not only will individual differences in initial transfer rate between devices have to be addressed, but differences in their rates of transfer over the range of applied transmembrane pressure will have to be accommodated in the model.

Fixed-effects regression with dummy variables

Up to this point we have considered two approaches for the modeling of the transfer rates. In the first, all the data were pooled and a common regression model was fitted to the data. In the second approach, a regression line was fitted to each

dialyzer's measurements. A summary of the estimated intercepts and slopes showed substantial between-dialyzer variation. The disadvantage of the second approach is that ten separate regression models are fitted. Ideally, a researcher would want to fit a single model that conveys information about between-subject variability.

One approach would be to do a regression analysis with dummy variables. Table 1.3 below shows the data for the first and last dialyzers. We use a dummy variable to represent each dialyzer, coded as follows:

$$\begin{aligned} D_j &= 1 \text{ for dialyzer } j = 1, 2, \ldots, 10 \\ &= 0 \text{ otherwise.} \end{aligned}$$

Table 1.3: Results of dummy variable model

device	rate	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	Pressure
11	1.50	1	0	0	0	0	0	0	0	0	0	28.5
11	15.40	1	0	0	0	0	0	0	0	0	0	52.0
11	32.52	1	0	0	0	0	0	0	0	0	0	100.5
11	42.44	1	0	0	0	0	0	0	0	0	0	150.0
11	48.57	1	0	0	0	0	0	0	0	0	0	198.5
11	53.69	1	0	0	0	0	0	0	0	0	0	249.0
11	53.66	1	0	0	0	0	0	0	0	0	0	299.5
20	5.71	0	0	0	0	0	0	0	0	0	1	28.0
20	20.50	0	0	0	0	0	0	0	0	0	1	50.5
20	39.41	0	0	0	0	0	0	0	0	0	1	100.0
20	50.10	0	0	0	0	0	0	0	0	0	1	149.0
20	55.16	0	0	0	0	0	0	0	0	0	1	200.0
20	61.19	0	0	0	0	0	0	0	0	0	1	250.5
20	50.72	0	0	0	0	0	0	0	0	0	1	302.0

The following regression model is fitted to the data:

$$\text{RATE}_{ij} = \alpha_0 \left(\text{D}_1\right)_{ij} + \alpha_1 \left(\text{D}_2\right)_{ij} + \text{K} + \alpha_9 \left(\text{D}_{10}\right)_{ij} + \alpha_{10} \left(\text{PRESSURE}\right)_{ij} + e_{ij}.$$

This model allows for the estimation of individual intercept coefficients, but a common slope parameter α_{10}.

Table 1.4 contains a summary of the results of this analysis. With the exception of the first and fourth dialyzers (represented by the dummy variables D1 and D4), the estimated coefficients associated with the individual dialyzers are all significantly different from zero at a 5% level of significance. From these results, we expect transfer rates for the second device to be much higher than for the first device, as reflected by the parameter estimates of 14.7153 and 5.2222 respectively. The transmembrane pressure also has a significant and positive relationship to the rate of transfer: for each increase of 1 dmHg in pressure, the rate of transfer is expected to be 0.1959 ml/hr higher.

Table 1.4: Results of regression model with dummy variables

Variable	Parameter estimate	Standard error	t-Value	Pr > |t|
D1	5.2222	3.9381	1.33	0.1899
D2	14.7153	3.9385	3.74	0.0004
D3	9.3364	3.9343	2.37	0.0209
D4	7.3311	3.9463	1.86	0.0682
D5	13.0271	3.9408	3.31	0.0016
D6	12.6855	3.9312	3.23	0.0020
D7	12.1007	3.9381	3.07	0.0032
D8	12.6124	3.9347	3.21	0.0022
D9	12.3179	3.9439	3.12	0.0028
D10	10.1676	3.9397	2.58	0.0124
Pressure	0.1959	0.0117	16.68	<.0001

Although this model is a compromise between the models for pooled data and separate models for dialyzers' data, the number of parameters to be estimated is proportional to the number of dialyzers and does not allow for the estimation of individual slopes. These issues have led researchers over time to develop mixed-effects models.

Random-intercept model

From the results of the previous models, we concluded that it is not reasonable to assume that the initial transfer rates of dialyzers, or their change in transfer rate with increased transmembrane pressure, can be described adequately by average intercept and slope estimates while the clustering of measurements within individual dialyzers was ignored. While the second of these analyses, where fixed-effects regression lines were fitted for each dialyzer and thus the clustering of measurements was acknowledged, provided better information per dialyzer, neither of these models allows us to obtain average intercept or slope coefficients while simultaneously incorporating the effect of measurements nested within individual devices.

To study differences in the behavior of dialyzers with pressure changes, while acknowledging the clustering of measurements and allowing for differences between devices in initial transfer rate, a random-effects model is needed. From the results obtained thus far, we will have to accommodate not only differences in initial status between dialyzers, but also differences in the slopes of the rates over the range of applied transmembrane pressure.

We start by specifying a model which takes clustering of measurements within dialyzers into account, while allowing the initial transfer rate to vary from device to device. This model, a so-called random-intercept model, contains both fixed and random effects, and can be expressed as

$$y_{ij} = \beta_0 + \beta_1(\text{PRESSURE})_{ij} + \beta_2(\text{PRESSURE})_{ij}^2 + v_{i0} + e_{ij} \qquad (1.2)$$

where y_{ij} denotes the Rate measurement at measurement j (j = 0, 1, 2, 3, 4, 5, 6, or 7) for dialyzer i, $(\text{PRESSURE})_{ij}$ the associated transmembrane pressure, $(\text{PRESSURE})_{ij}^2$ the squared value of $(\text{PRESSURE})_{ij}$, and e_{ij} measurement error. The coefficients β_0, β_1 and β_2 are the fixed, but unknown, parameters to be estimated. The coefficient v_{i0}, in contrast, denotes a random parameter, and represents the amount by which the intercept of dialyzer i differs from the average (fixed) intercept for all devices, as represented by β_0. By including v_{i0}, we allow

intercepts to vary randomly over the dialyzers. We assume that v_{i0} is normally distributed with mean 0 and variance $\phi_{(2)}$ and that the e_{ij}, too, as in the first model, have a normal distribution with mean 0 and variance σ^2 for all dialyzers.

In contrast to the model in (1.1), where all unexplained variations in transfer rates were captured by e_{ij}, the current model assumes that there are two potential sources of unexplained variation: variation between measurements as represented by e_{ij}, and variation between dialyzers in terms of their intercepts, as represented by v_{i0}. Viewing the measurements as the lowest level of a nested structure in our data, with measurements nested within devices, we refer to σ^2 as the level-1 (measurement-level) variance and to $\phi_{(2)}$ as the level-2 (dialyzer-level) variance.

In fitting this model, data from all 20 devices are used. The results of the analysis are reported in Table 1.5. All of the estimated coefficients are statistically significant at a 5% level of significance. We see that the rate of transfer is expected to increase with an increase in pressure. However, as pressure increases, the squared value of pressure increases quickly, and the small negative coefficient for this will lead to larger decreases in transfer rate at high pressures. At first glance, these estimates indicate a somewhat nonlinear curve.

Table 1.5: Results of random-intercept model

Parameter	Estimate	Standard error
Intercept	–6.56547	1.56214
Pressure	0.52792	0.01840
Pressure^2	–0.00114	0.00006
var(v_{i0})	16.28786	6.29943
var(e_{ij})	25.05420	3.23435

What is really interesting, and something we have not been able to look at previously, is the amount of variation within and between devices. While most of the variation is at measurement level, *i.e.* within devices, as indicated by $\text{var}(e_{ij}) = 25.0542$, there is a sizable amount of variation in the intercepts of the devices themselves. As this estimated coefficient is statistically significant, it indicates that it is not adequate to try and describe the intercepts of the devices using a single, common fixed effect as we have done previously. If we had more characteristics of the individual devices, these could have been added to our current model in an attempt to explain away the variation in device intercepts. Likewise, we could have used any other type of measurement made at the measurement occasions to explain more of the residual variation. In Chapter 3, models with a continuous outcome are described in which the use of additional characteristics at both levels is illustrated.

In addition to these estimates, which describe the average estimated intercept and slope over all devices, we also obtain estimates for the unique deviations from the intercept associated with each of the individual devices. The estimates of the deviations of the predicted from the observed values are depicted graphically in Figure 1.5. The residuals associated with devices 11 and 20 are highlighted: residuals for device 11 are shown as square black boxes, and those for device 20 as asterisks. We see that almost all the residuals are within a (–10,10) interval. For device 11, the residuals are closer to zero in value at lower transfer rates, but vary quite a bit more above a transfer rate of 40. The residuals for device 20, however, vary more over the entire rate of transfer range.

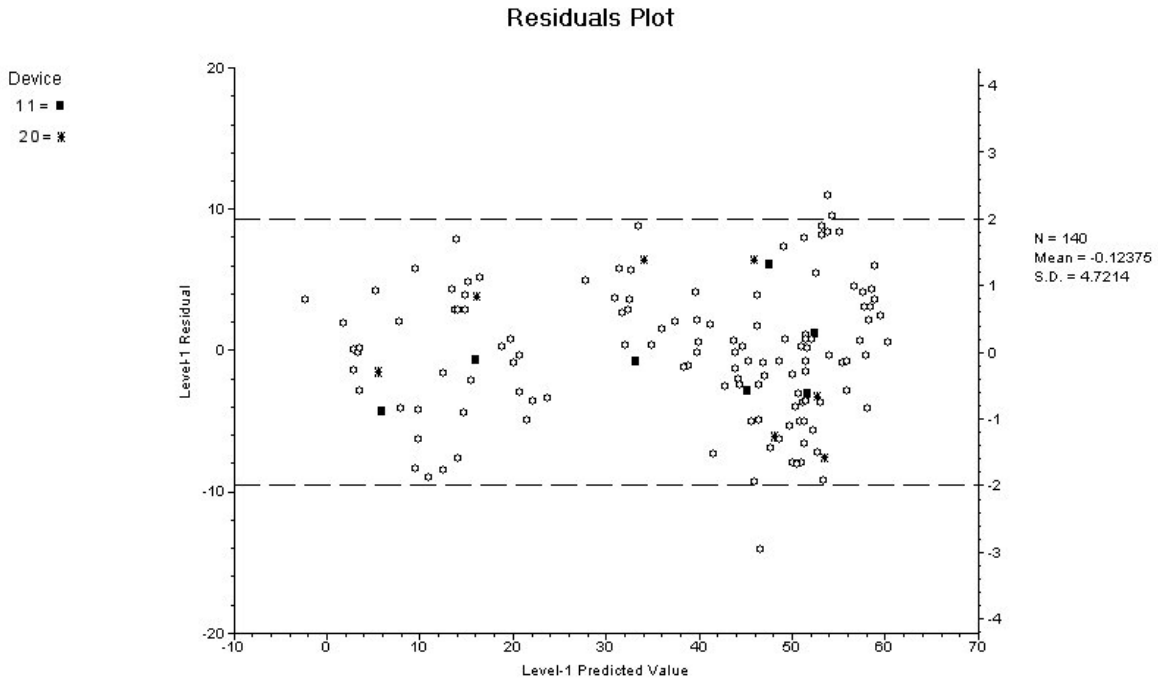


Figure 1.5: Level-1 residuals plotted against level-1 predicted values

Another way to look at these results is to inspect confidence intervals for the deviations of the device intercepts from the estimated value of –6.565. These are shown in Figure 1.6. The units appear in numerical order, and we can see that the 95% confidence interval for the intercept of device 20 is approximately centered above 0, while that of device 11 is centered below zero. Looking at the confidence intervals for devices 1 and 12, our result that there is significant variation in the device intercepts makes sense.

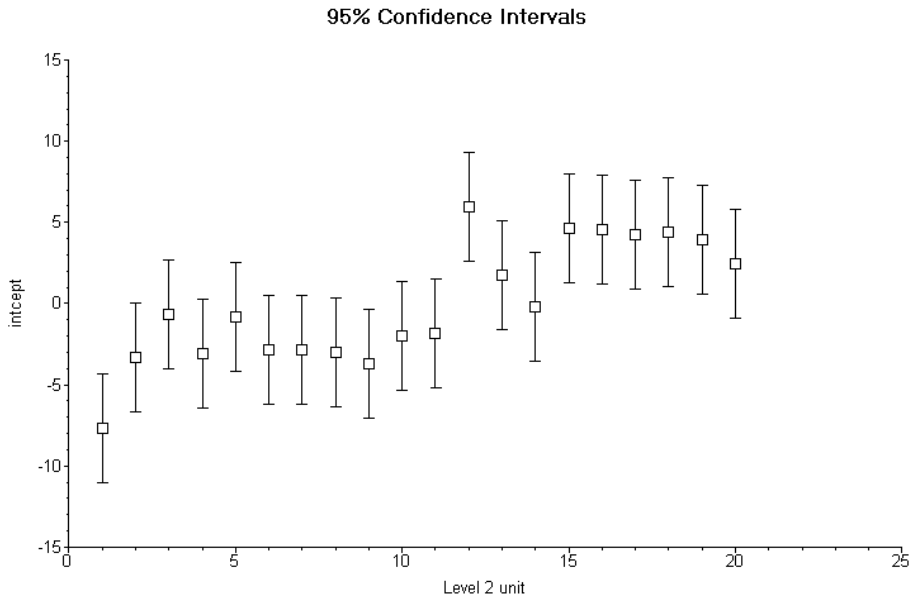


Figure 1.6: 95% confidence intervals for 20 devices

Intraclass correlation

The intraclass correlation is a measure of the degree of dependence of the higher-level units, in this case the devices. It is realistic to assume that measurements from the same device are more alike with respect to certain traits than measurements from different devices.

For data having a two-level hierarchical structure, the intraclass correlation ρ is defined as the proportion of the variance in the outcome variable that is between the second-level units:

$$\rho = \frac{\text{between group variability}}{\text{between group variability} + \text{within group variability}}$$

In the current example, we obtain $\hat{\rho}$ as

$$\hat{\rho} = \frac{16.28786}{16.28786 + 25.05420} = 0.39398.$$

As pointed out by Kreft and de Leeuw (1998), if intraclass correlation is present, as is usually the case when we are dealing with clustered data, the assumption of independent observations in the traditional linear model is violated. They also pointed out that tests of significance lean heavily on the number of independent observations and that the existence of intraclass correlation makes the test of significance in traditional linear models too liberal. Barcikowski (1981) shows that in most applications of analysis of variance, the standard errors of the parameter estimates will be underestimated and that even a small intraclass correlation can inflate the alpha level substantially.

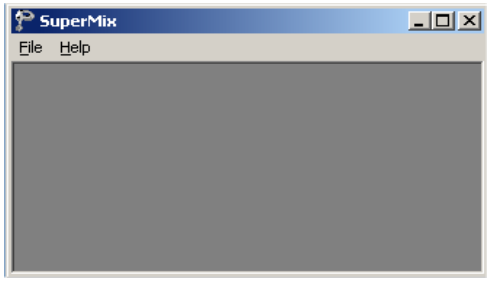
While the random-intercept model has allowed us to accommodate some of our modeling concerns for an unbalanced data set such as the nesting of measurements within devices and allowing intercepts to vary over devices, other concerns remain. From the results shown in Table 1.5, we know that there is a sizable amount of variation between devices, variation that may be explained by the inclusion of additional device characteristics in the model. To address these concerns, extended models are required. Examples of such models, based on the Reisby data, are shown in detail in Section 3.2.

2 Graphical User Interface

The SuperMix graphical user interface (GUI) consists of a main window, a spreadsheet window, and a graph window. The main window is used to create or open SuperMix data files, whereas the spreadsheet window is used to display SuperMix data files and to allows access to the **Model Setup** window. The graph window is used to display SuperMix graph files. SuperMix data files have the default extension **.ss3** and are known as **ss3** or spreadsheet files, while SuperMix model files have the default extension **.mum**. SuperMix graph files have the default extension **.mug**. The main window and its menus and dialog boxes are reviewed in the next section, and the menus and dialogs of the spreadsheet and graph windows are reviewed in the sections to follow.

2.1 The main window

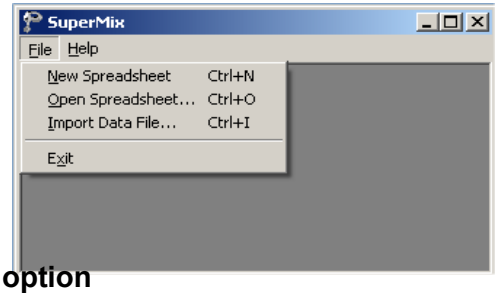
The SuperMix main window is accessed when you start the program. SuperMix can be opened from the **Programs** option on the Windows **Start** menu, by double-clicking on the SuperMix application or by clicking on a shortcut for SuperMix. Any of these actions opens the following main window.



The SuperMix main window consists of a **File** menu and a **Help** menu. These menus are reviewed separately in the following two sections.

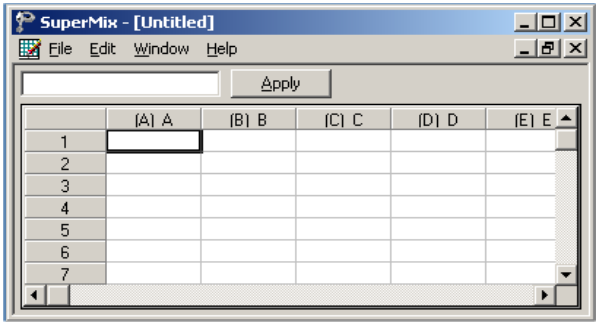
2.1.1 The File menu

The options on the **File** menu of the SuperMix main window provide access to a sequence of three dialog boxes that can be used to create or to open a SuperMix data file in a spreadsheet format.



The New Spreadsheet option

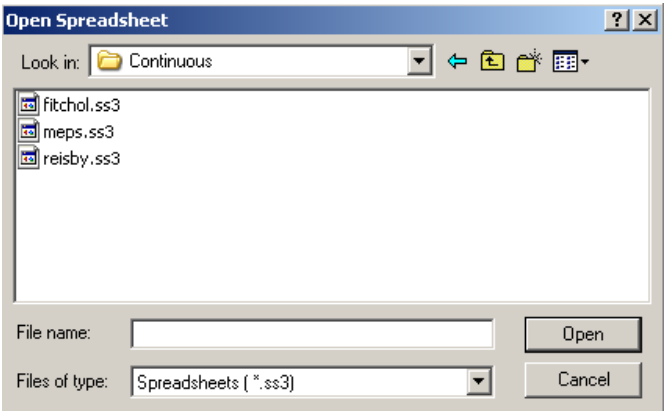
Click on the **New Spreadsheet** option to open an empty SuperMix spreadsheet window.



You can use the window above to enter data manually. Use the **Save As** option on the **File** menu to save the data to an **ss3** file. Alternatively, data can be imported into the empty spreadsheet via the **File**, **Import Data File** option.

The Open Spreadsheet option

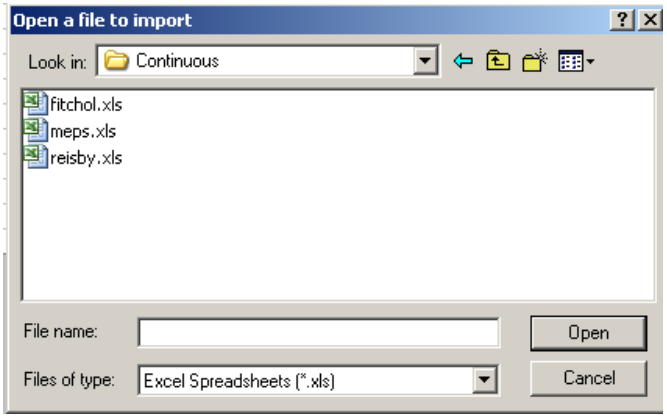
The **Open** option on the **File** menu is used to open an existing SuperMix data file. Click on the **Open Spreadsheet** option to load the following **Open Spreadsheet** dialog box.



Next, browse for the **ss3** file, select it, and click on the **Open** button to open the SuperMix spreadsheet window.

The Import Data File option

Use the **Import Data File** option on the **File** menu to convert the data in a Microsoft Excel workbook (***.xls**), statistical files and databases (SAS, SPSS, etc.) or a comma delimited text file to a SuperMix data file. To import an Excel data file, click on the **Import Data File** option to load the following **Open a file to import** dialog box.



Next, browse for the Microsoft Excel workbook or the text file and select it. Click on the **Open** button to load the **Save As** dialog box. Enter a name for the **ss3** file and click on the **Save** button to open the SuperMix data file in a spreadsheet window.

The Exit option

Close the SuperMix main window by clicking on the **Exit** option on the **File** menu.

2.1.2 The Help menu

The options on the **Help** menu on the SuperMix main window provide access to the contents of the SuperMix online help file, the SuperMix user's guide, the SuperMix website, technical support and other information.

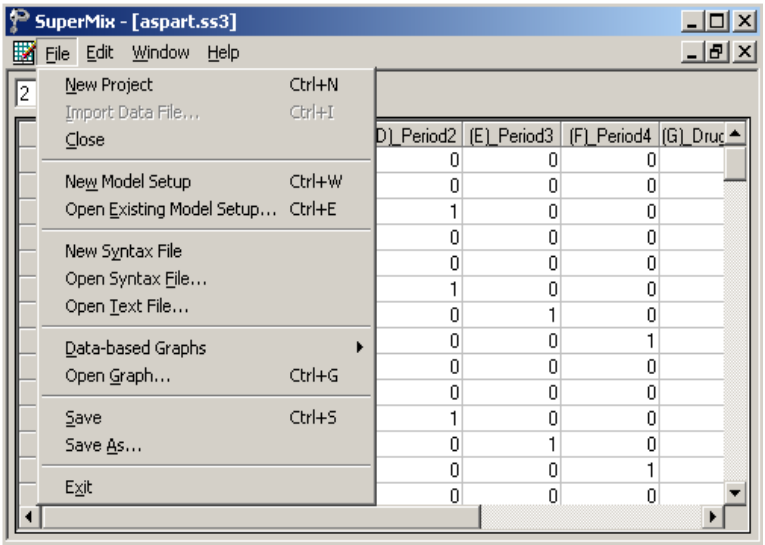
2.2 The spreadsheet window

The SuperMix spreadsheet window is used to display a new or existing SuperMix data file. The menus on the spreadsheet window can be used to manipulate the data entries in an existing SuperMix data file. It is also used to access the **Model Setup** window, which is used to specify a mixed-effects model and to edit existing SuperMix model files. These menus can also be used to create new or open existing SuperMix graph files. In Section 2.5 some basic spreadsheet operations are

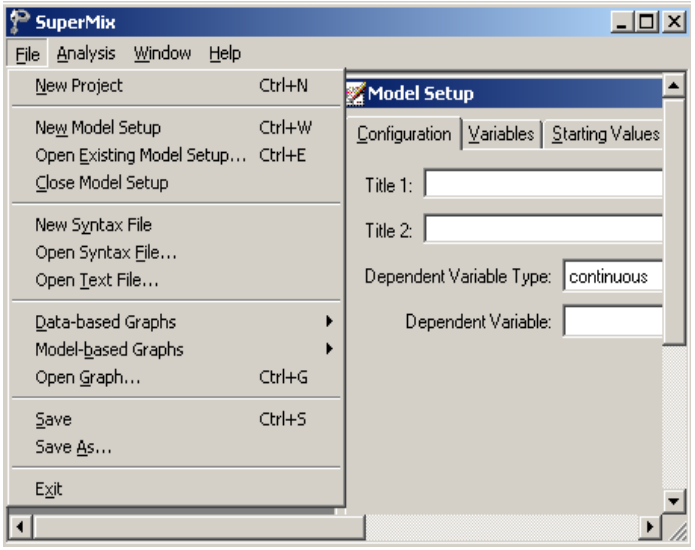
illustrated. In the sections to follow, we review the four menus of the SuperMix spreadsheet window.

2.2.1 The File menu

The options on the **File** menu of the spreadsheet window are used to open a new SuperMix project, open an existing **ss3** file, create a new SuperMix model (.**mum**) file, edit an existing model file, or convert an existing MIX definition file to a SuperMix model file. It is also used to create or edit a SuperMix graph file. An example of the **File** menu is shown below.



When an **ss3** file as well as a SuperMix model file are opened, the **File** menu changes as shown in the following window.



The New Project option

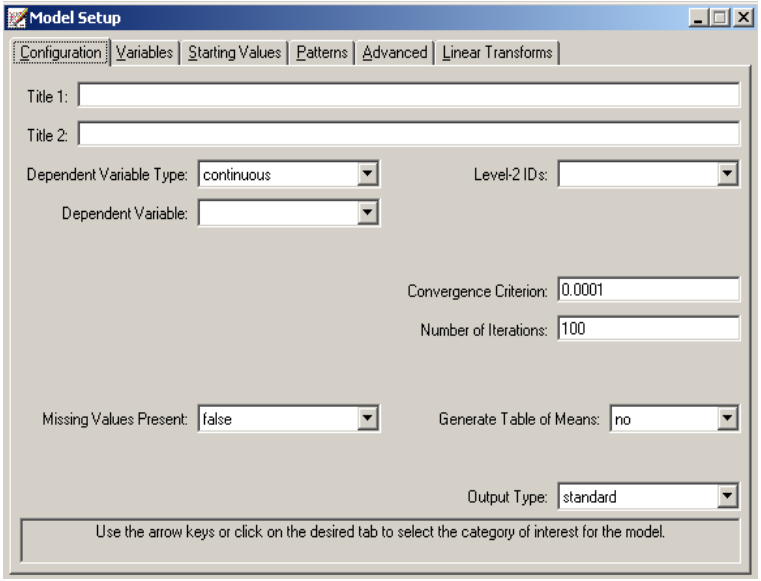
The **New Project** option is used to open an independent SuperMix main window.

The Exit option

The **Exit** option is used to close the current open SuperMix main window.

The New Model Setup option

The **New Model Setup** option of the spreadsheet window provides access to the **Configuration**, **Variables**, **Starting Values**, **Patterns**, **Advanced** and **Linear Transforms** screens of the **Model Setup** window shown below. Each screen is opened by clicking on the corresponding tab.

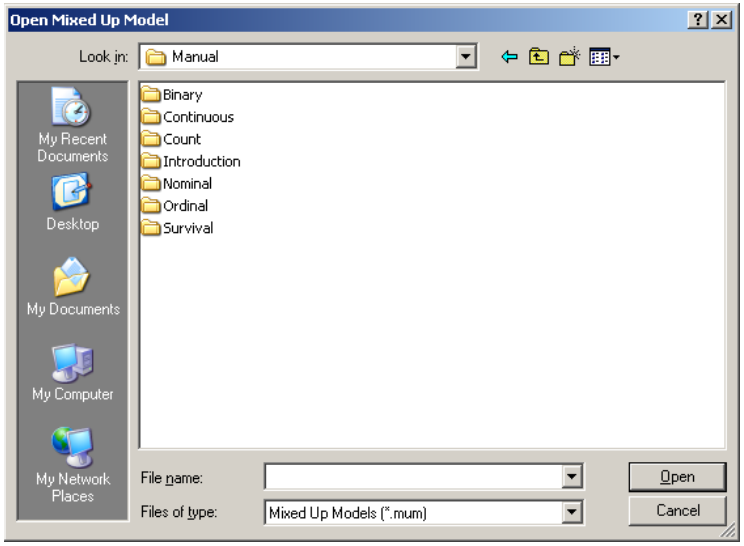


These screens are used to specify a mixed-effects model to be fitted to the data in the open spreadsheet window. The appearance of the screens depends on the type of outcome (dependent) variable (continuous, count, ordered, or nominal) that is selected on the **Configuration** screen shown above. A detailed description of each of these screens is given in Section 2.4. Once a model is defined, it can be saved as a **.mum** file.

The Open Existing Model Setup option

The **Open Existing Model Setup** option is used to open the **Model Setup** window of an existing SuperMix model file. This is accomplished by clicking on the **Open Existing Model Setup** option, which loads the following **Open Mixed Up Model** dialog box.

Browse for the desired SuperMix model file, select it, and click on the **Open** button to load the **Model Setup** window for the selected SuperMix model file.



The Close Model Setup option

The **Close Model Setup** option is used to close any SuperMix Model Setup dialog box that is currently open.

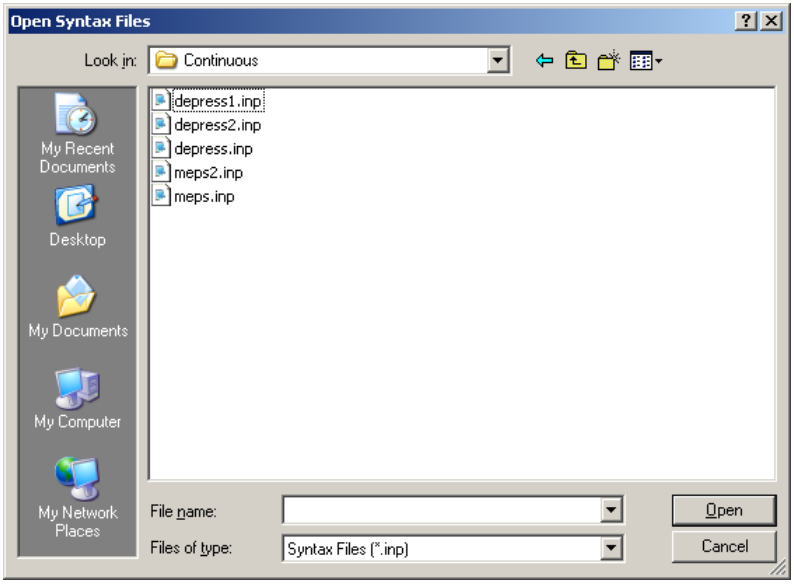
The New Syntax File option

The **New Syntax File** option is used to open a blank syntax window.

The Open Syntax File option

The **Open Syntax File** option is used to open an existing syntax file (**.inp**). This is accomplished by clicking on the **Open Syntax File** option, which leads to the display of the **Open Syntax File** dialog box.

Browse for the desired SuperMix syntax file, select it, and click on the **Open** button to open the syntax window for the selected SuperMix model file.



The Open Text File option

The **Open Text File** option is used to open any existing text file.

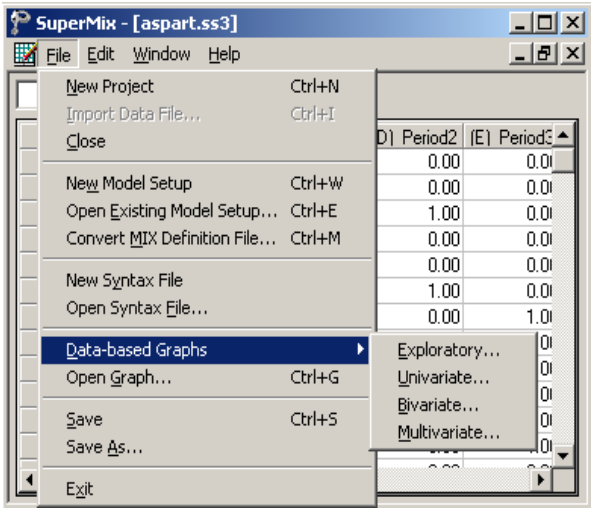
The Data-based Graphs pop-up menu

The **Data-based Graphs** pop-up menu is used to create a new SuperMix graph from the data displayed in the open **ss3** file in a SuperMix graph window. The menus and dialogs of the SuperMix graph window for new and existing SuperMix graphs are reviewed in Chapter 4 of the SuperMix primer. Various illustrations are given throughout Chapters 3 to 8 of this manual.

- The **Exploratory** option is used to produce single or overlay color-coded *Y* against *X* plots. Groups of plots are obtained by using a filter variable.
- The **Univariate** option on the **Data-based Graphs** pop-up menu is used to create a bar chart, a pie chart or a histogram for the data displayed in the spreadsheet window.
- The **Bivariate** option on the **Data-based Graphs** pop-up menu is used to create a scatter plot, a line plot, a combination line and scatter plot, a box-

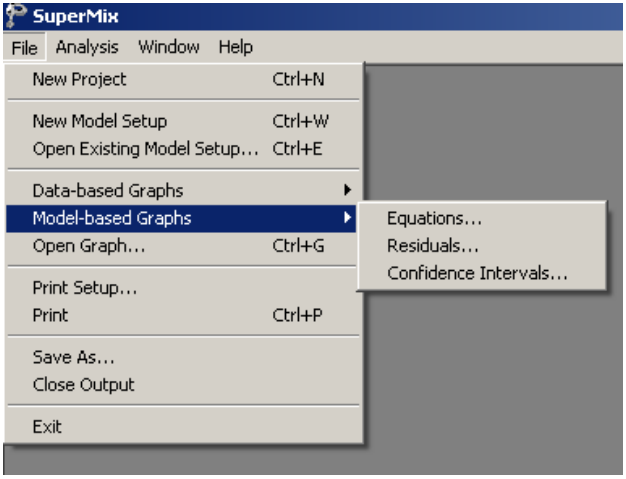
and-whisker plot, or a 3-dimensional bar chart for the data in the open SuperMix data file.

- The **Multivariate** option on the **Data-based Graphs** pop-up menu is used to make a matrix scatter plot based on the data in the open **ss3** file. This provides an organized way of simultaneously looking at a set of bivariate plots.



The Model-based Graphs pop-up menu

The options on the **Model-based Graphs** pop-up menu are activated when a model setup file is opened. These options are used to create a new SuperMix graph from the data displayed in the open spreadsheet window. The menus and dialogs of the SuperMix graph window for new and existing SuperMix graphs are reviewed in Chapter 4 of the SuperMix primer. Various illustrations are given throughout Chapters 3 to 8 of this manual.

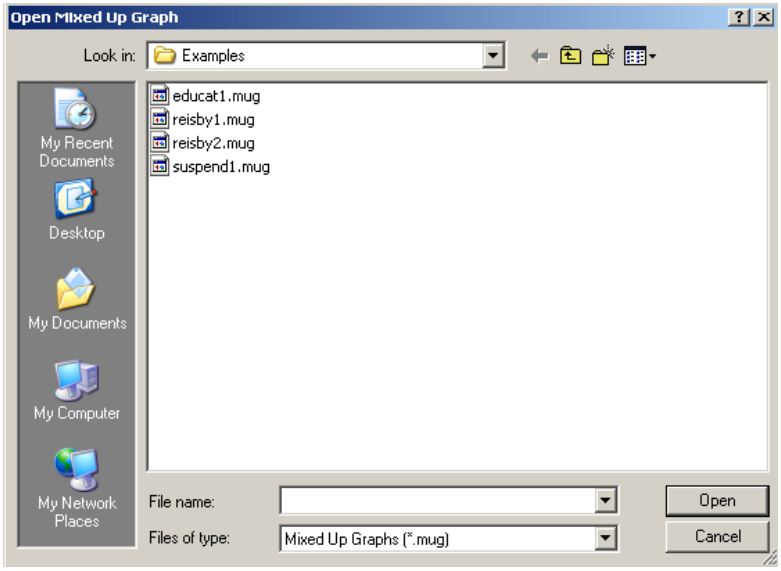


Available options are:

- The **Equations** option on the **Model-based Graphs** pop-up menu loads the **Plot Equations for** dialog box which can be used to plot model equations of an outcome variable for given values of the predictors in the model.
- The **Residuals** option on the **Model-based Graphs** pop-up menu provides access to the **Plot of Residuals** dialog box, which is used to create a residual plot for the residuals based on the current SuperMix analysis.
- The **Confidence Intervals** option is used to open the **95% C.I. for Level-1 Variables** dialog box, which is used to create confidence interval plots.

The Open Graph option

The **Open Graph** option is used to open an existing SuperMix graph file with a default extension **.mug**. You first click on the **Open Graph** option to load the following **Open Mixed Up Graph** dialog box.



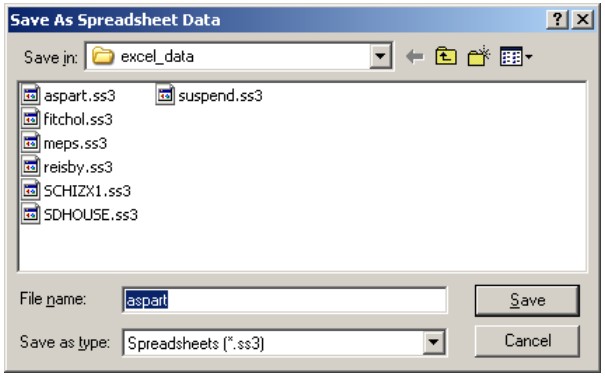
The next steps are to browse for the desired SuperMix graph file, select it, and click on the **Open** button to open the graph window for the selected SuperMix graph file.

The Save option

The **Save** option on the **File** menu is used to save any changes made to the data or the model setup file (mum). Please note that any change to the data will not be saved to file unless you use this option or the **Save As** option.

The Save As option

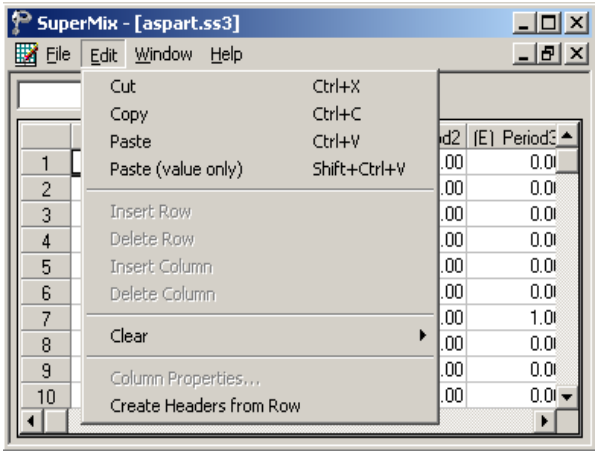
The **Save As** option on the **File** menu is used to save the opened **ss3** file or **mum** file as another SuperMix data file or mum file. To save the spreadsheet data as another file, select the **Save As** option to load the following **Save As Spreadsheet Data** dialog box.



Enter the file name in the **File name** string field and click on the **Save** button to save the SuperMix data file.

2.2.2 The Edit menu

The options on the **Edit** menu of the SuperMix spreadsheet window are used to edit the data entries of the open SuperMix data file. To use these options, select the data to be edited (cell(s), row(s) or columns(s)). Then click on the **Edit** menu to produce the following window.



The next step is to select one of the options available, which have the following effects on the selected data:

- The **Cut** option cuts the data selection from the spreadsheet window and places it into the Windows clipboard.
- The **Copy** option places the data selection in the Windows clipboard.
- The **Paste** option pastes data from the Windows clipboard into the selected area of the spreadsheet window.
- The **Paste (value only)** option pastes only the actual values (ignoring the formats) of the data from the Windows clipboard into the selected area of the spreadsheet window.
- The **Clear** option replaces the selected data with empty cell(s). Choosing this option activates the following drop-down menu.

  - The **Clear All** option deletes the values and the formulas of the selected data.
  - The **Clear Data** option deletes the values of the data selection, but leaves the corresponding formulas intact.
  - The **Clear Formula** option deletes the formulas of the selected data, but not the corresponding values.
- The **Create Header from Row** creates spreadsheet headers that correspond to the labels in the selected row.

2.2.3 The Window menu

The **Window** menu is used to toggle between open spreadsheet windows.

2.2.4 The Help menu

The **Help** menu of the spreadsheet window is identical to that of the main window and is reviewed in Section 2.1.2.

2.3 The graph window

The SuperMix graph window is opened by creating a new SuperMix graph or by opening an existing SuperMix graph file. We accomplish this by using one of the options on the **Data-based Graphs** pop-up menu or the **Open Graph** option or one of the options on the **Model-based Graphs** pop-up menu (if a SuperMix model file is also open) on the **File** menu of the spreadsheet window reviewed in Section 2.2.1. The menus and dialogs of the SuperMix graph window for new and existing SuperMix graphs are reviewed in Chapter 4 of the SuperMix primer. Various illustrations are given throughout Chapters 3 to 7 of this manual.

2.4 The Model Setup window

A SuperMix **mum** file (model setup file) is always associated with an **ss3** file (data spreadsheet file). This ensures that variable selections are maintained correctly in the **mum** file, regardless of changes to the header text and cut/paste/move operations on the columns of the **ss3** file. For this reason, the **Model Setup** window is accessed via the **File** menu of the spreadsheet window. This is done by selecting the **New Model Setup** or **Open Existing Model Setup** options. The **Model Setup** window has six tabs. By clicking on a tab, the corresponding **Configuration**, **Variables**, **Starting Values**, **Patterns**, **Advanced**, or **Linear Transforms** screen is accessed. The appearance of a screen depends on the type of outcome variable selected. The purpose of a field is displayed at the bottom of the screen when the field is clicked. Tables 2.1 to 2.13 are summaries of these descriptions.

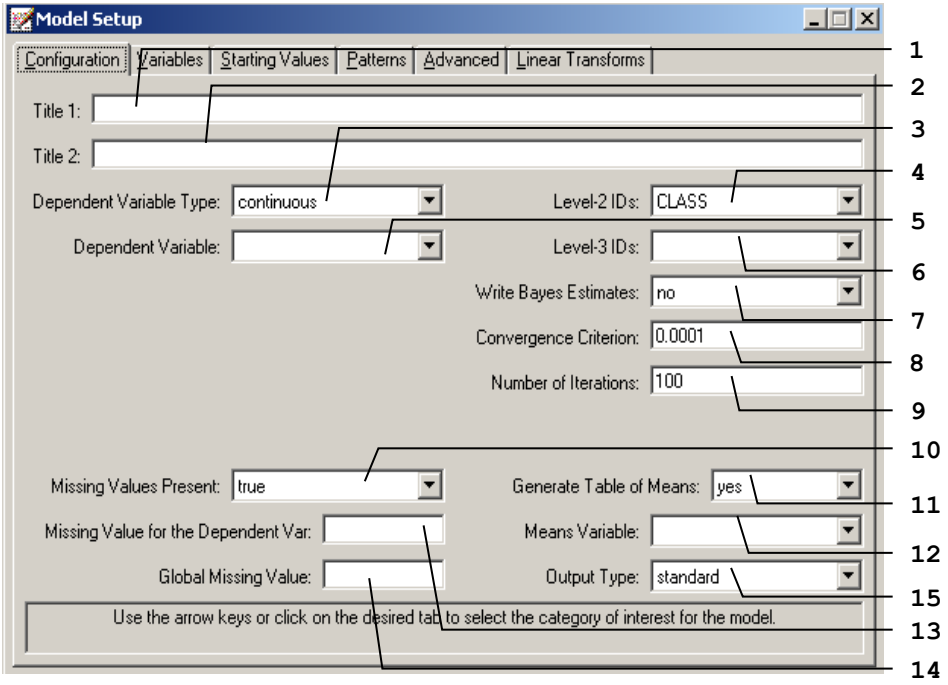
2.4.1 The Configuration screen

The **Configuration** screen is used to provide a title for the analysis, to select the type and name of the outcome (dependent) variable, and to indicate identifiers of the level-2 and level-3 units. Additionally, it contains options that control the amount of information to be saved to file and the parameters of the optimization procedure. When the **New Model Setup** or **Open Existing Model Setup** options on the **File** menu are used, the **Configuration** screen is, by default, the first screen displayed.

The same **Configuration** screen is used for continuous and count outcomes, but its contents change when the dependent variable type is ordinal or nominal. The screen is the same for ordinal and nominal outcome types. The two cases are discussed separately below.

Configuration screen for continuous and count outcomes

An example of the **Configuration** screen of the **Model Setup** window for a continuous response variable is shown below. The layout is identical when the dependent variable type is changed from **continuous** to **count** (see Chapter 5 for examples based on a count outcome variable).



The 15 possible entries on the **Configuration** screen of the **Model Setup** window for continuous or count response variables are summarized in Table 2.1.

Table 2.1: Entries on the Configuration screen of the Model Setup window for continuous and count outcomes

Number	Caption	Purpose	Type	Action	Options
1	Title 1	To specify the first line of the title to be listed in the output file.	Text box	Enter a string of not more than 60 characters.	
2	Title 2	To specify the second line of the title to be listed in the output file.	Text box	Enter a string of not more than 60 characters.	
3	Dependent Variable Type	To specify the variable type for the response variable.	Drop-down list box	Select an option from the drop-down list box.	**continuous** (default)
					ordered
					nominal
					count
4	Level-2 IDs	To specify the variable that defines the second level of the hierarchy in the data.	Drop-down list box	Select a variable from the drop-down list box.	
5	Dependent Variable	To specify the response variable of the model.	Drop-down list box	Select a variable from the drop-down list box.	
6	Level-3 IDs	To specify the variable that defines the 3rd level of the hierarchy in the data.	Drop-down list box	Select a variable from the drop-down list box.	
7	Write Bayes Estimates	To request a text file for the Bayes estimates.	Drop-down list box	Select an option from the drop-down list box.	**no** (default)
					means only
					means & (co)variances
8	Convergence Criterion	To specify the convergence criterion for the iterative algorithm.	Text box	Enter a non-zero positive real number if the default of 0.0001 is not desired.	

Table 2.1: Entries on the Configuration screen of the Model Setup window for continuous and count outcomes (continued)

9	Number of Iterations	To specify the maximum number of iterations for the iterative algorithm.	Text box	Enter a positive integer if the default of 100 is not desired.	
10	Missing Values Present	To specify the missing value status of the data.	Drop-down list box	Select an option from the drop-down list box.	false (default) **true**
11	Generate Table of Means	To request the printing of a table as part of the output.	Drop-down list box	Select an option from the drop-down list box.	no (default) **yes**
12	Means Variable	To specify the variable for which the tables should be created (see 8).	Drop-down list box	Select a variable from the drop-down list box.	
13	Missing Value for the Dependent Var	To specify the missing value code for the response variable.	Text box	Enter a real number.	
14	Global Missing Value	To specify the global missing value code.	Text box	Enter a real number.	
15	Output Type	To request different type of output.	Drop-down list box	Select an option from drop-down list box.	**standard** (default) iterative details simulation information

Configuration screen for ordered, nominal and binary outcomes

The following screen is an example of the **Configuration** screen of the **Model Setup** window in the case of an ordered response variable. An example of this screen for a nominal outcome variable is given in Chapter 7.

As shown in the image below, the 5 entries shown in bold typeface are either new or different compared with those on the **Configuration** screen of the **Model Setup** window for continuous or count outcome variables. These 5 entries are summarized in Table 2.2. Please refer to Table 2.1 for the information about all the other entries.

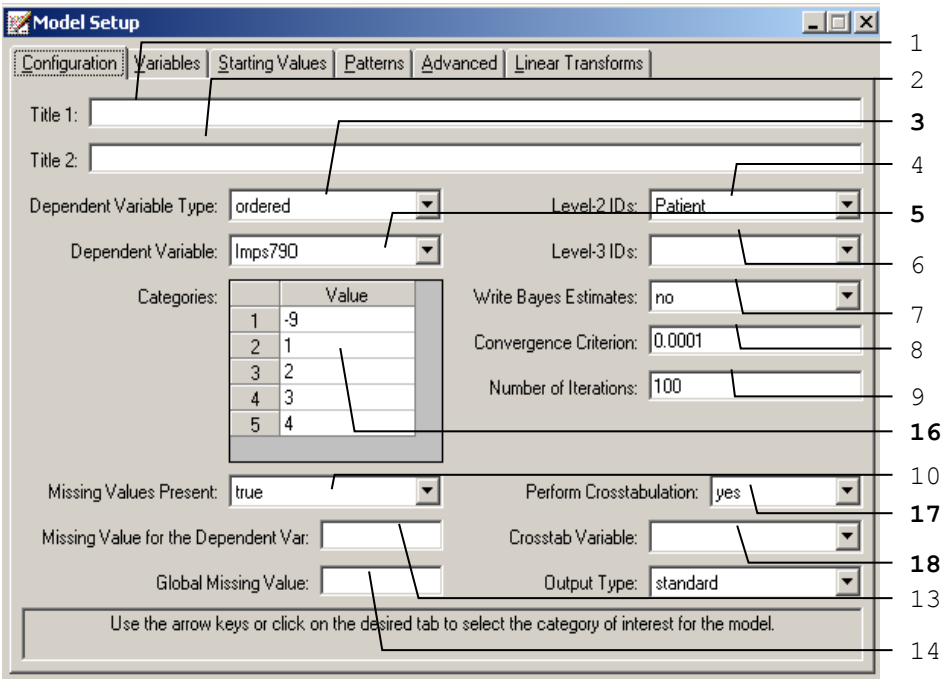


Table 2.2: Entries of the configuration screen for ordered and nominal outcomes

Number	Caption	Purpose	Type	Action	Options
3	Dependent Variable Type	To specify the variable type for the response variable.	Drop-down list box	Select an option from the drop-down list box.	continuous (default) **ordered** nominal count
5	Dependent Variable	To specify the response variable of the model.	Drop-down list box	Select a variable from the drop-down list box.	
16	Categories	To show the value of each category of the ordered dependent variable selected in 5.	Grid box		

Table 2.2: Entries of the configuration screen for ordered and nominal outcomes (continued)

17	Perform Crosstabulation	To specify a cross tabulation of selected variable by the outcome variable.	Drop-down list box	Select an option from the drop-down list box.	no (default) **yes**
18	Crosstab Variable	To specify the variable to be crosstabulated with the outcome variable (see 8).	Drop-down list box	Select a variable from the drop-down list box.	

2.4.2 The Variables screen

Besides the variables screen for the ordered outcome, which doesn't include the option to select an intercept as an explanatory variable, this screen has the same appearance for all outcome types and is used to select explanatory variables and random effects. The unknown model parameters are the coefficients of the explanatory variables and the variances and covariances of the random effects. The appearance of the **Variables** screen depends on the number of levels of the model. For a two-level model, the **3** columns in the **Available** grid and the **L-3** grid will be hidden. By default, an intercept term is included in the fixed part (explanatory variables) and in the random part (random effects) of the model.

The following screen is an example of the **Variables** screen of the **Model Setup** window that is used for variable selection for continuous, count, or nominal response variables. The 9 possible entries of the **Variables** screen of the **Model Setup** window for continuous, count, or nominal response variables are summarized in Table 2.3.

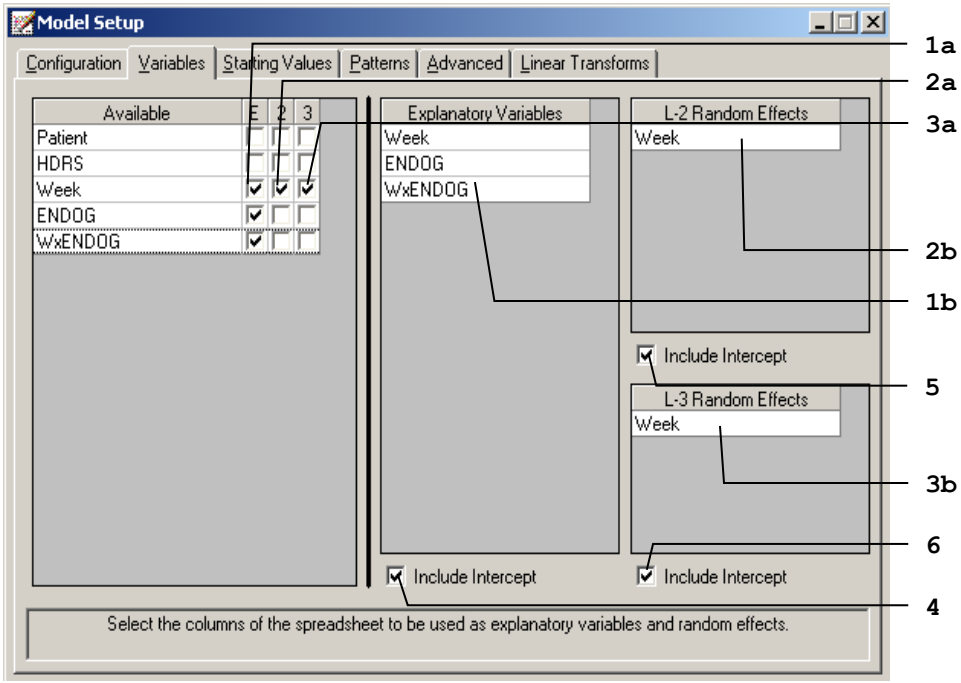


Table 2.3: Entries of the Variables screen

Number		Caption	Purpose	Type	Action	Options
1	a	E	To specify the explanatory variable(s) of the model.	Column of check box(es)	Check the E column(s) of the variable(s).	
	b	Explanatory Variables	Displays the variable(s) selected in 1a.	Grid box		
2	a	2	To specify the level-2 random effects of the model.	Column of check box(es)	Check the 2 column(s) of the variable(s).	
	b	L-2 Random Effects	Displays the variable(s) selected in 2a.	Grid box		

Table 2.3: Entries of the Variables screen (continued)

3	a	3	To specify the level-3 random effects of the model.	Column of check box(es)	Check the 3 column(s) of the variable(s).	
	b	L-3 Random Effects	Displays the variable(s) selected in 3a.	Grid box		
4		Include Intercept	To specify an intercept term for the fixed part of the model.	Check box	Uncheck the check box if an intercept is not desired.	**Check (default)**
						Uncheck
5		Include Intercept	To specify a random intercept at level-2 of the model.	Check box	Uncheck the check box if a level-2 random intercept is not desired.	**Check (default)**
						Uncheck
6		Include Intercept	To specify a random intercept at level-3 of the model.	Check box	Uncheck the check box if a level-3 random intercept is not desired.	**Check (default)**
						Uncheck

2.4.3 The Starting Values screen

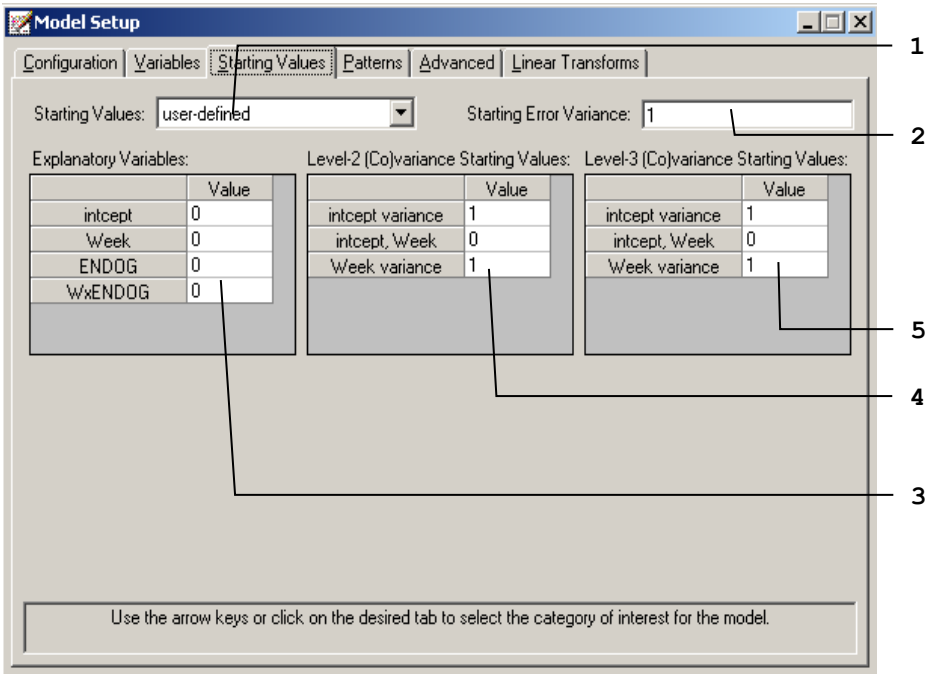
The unknown parameters in a mixed-effects model cannot, in general, be obtained as a closed-form expression. To estimate these parameters, use is made of an iterative procedure based on the method of maximum likelihood. For count, ordinal, and nominal outcomes, the likelihood function is approximated by numerical integration. For more than one random effect, this procedure is computationally intensive. All iterative procedures start with initial estimates of the values of the unknown parameters and, at each iteration, the algorithm attempts to improve this estimate until convergence is obtained. The closer these initial estimates (the starting values) are to the maximum likelihood solution, the fewer iterations are needed to obtain convergence and reach the final solution.

SuperMix automatically generates starting values for the model parameters and typically these values are sufficient to ensure convergence. There may, however, be cases where a model with many parameters takes a long time to run, and if small modifications are made to such a model, one can use the parameter estimates from

the previous analysis as starting values for the next analysis. Alternatively, one may want to fix some of the parameter values at specific values, for example, the slope coefficient of variable *X* at 0.1. This can be accomplished by selecting the user-defined option and entering this value for *X*. Note that the value of 0.1 will remain fixed during the optimization procedure if it is specified as fixed for *X* on the **Patterns** screen discussed in Section 2.4.4.

Starting Values screen for continuous or count outcomes

An example of the **Starting Values** screen of the **Model Setup** window for a continuous or count response variable is shown below.



An example of the **Starting Values** screen of the **Model Setup** window for continuous or count outcomes is shown above. The 5 possible entries of the **Starting Values** screen of the **Model Setup** window for count or nominal response variables are summarized in Table 2.4.

Table 2.4: Entries of the Starting Values screen for continuous and count outcomes

Number	Caption	Purpose	Type	Action	Options
1	Starting Values	To specify the type of starting values to be used.	Drop-down list box	Select an option from the drop-down list box.	automatic (default) **user-defined**
2	Starting Error Variance	To specify the starting error variances.	Text box	Enter a integer if the default of 1 is not desired.	
3	Explanatory Variables	To specify the starting value(s) for the coefficients of explanatory variable(s) of the fixed part of the model.	Grid box	Enter a real number in the corresponding Value box(es) of the variable(s) of interest.	
4	Level 2 (Co)variances Starting Values	To specify the starting value(s) for the variance(s) and/or covariance(s) of the level-2 random effects.	Grid box	Enter a real number (positive for variances) in the corresponding Value box(es) of the variable(s) of interest.	
5	Level 3 (Co)variances Starting Values	To specify the starting value(s) for the variance(s) or covariance(s) of the level-3 random effects.	Grid box	Enter a real number (positive for variances) in the corresponding Value box(es) of the variable(s) of interest.	

Starting Values screen for ordered outcomes

For ordinal outcomes, additional grid boxes appear to allow for user-specified starting values of thresholds and threshold-explanatory variable(s) interaction(s). The following screen is an example of the **Starting Values** screen of the **Model Setup** window.

As shown in the image below, the single entry shown in bold typeface is different from those of the **Starting Values** screen of the **Model Setup** window for continuous or count outcome variables. This entry is described in Table 2.5. Please refer to Table 2.4 for the information about all the other entries.

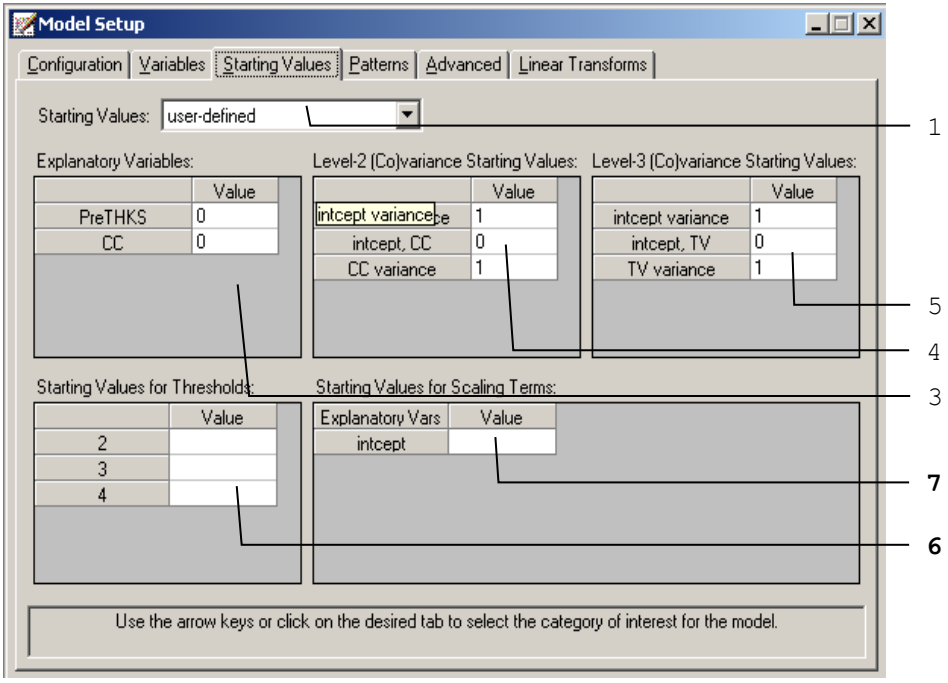


Table 2.5: Entry of the Starting Values screen for ordered outcomes

Number	Caption	Purpose	Type	Action
6	Starting Values for Thresholds	Enter the starting values for the thresholds.	Grid box	Enter real numbers. The values must be monotonically increasing.
7	Starting Values for threshold interactions	Enter the starting values for the threshold interaction terms.	Grid box	Enter a real number in each of the corresponding Value box(es) of the variable(s) of interest.

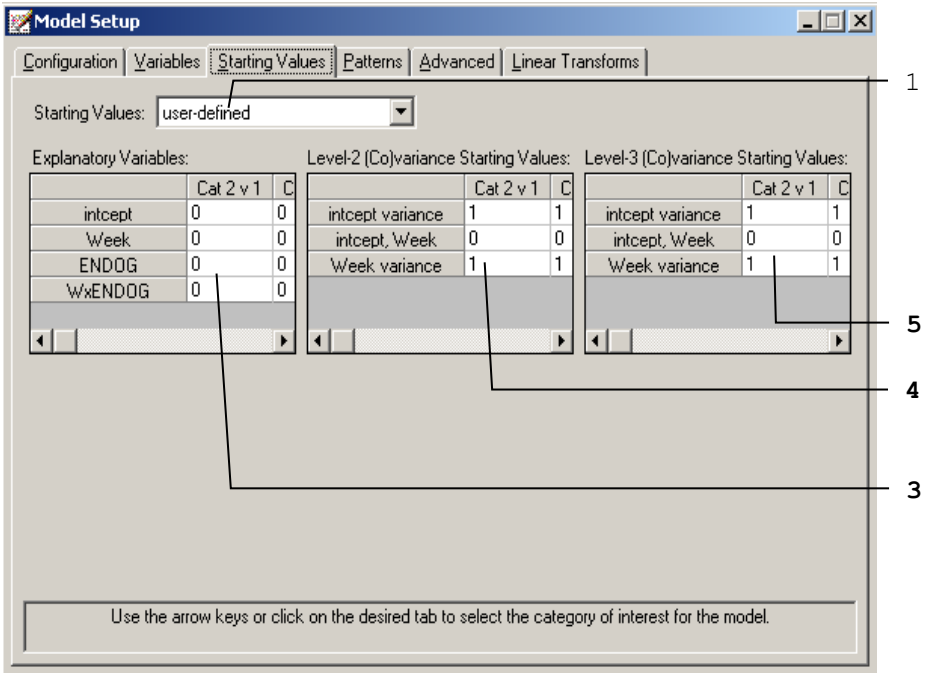
Starting Values screen for nominal or binary outcomes

When the nominal outcome is selected, the grid boxes appear differently with a slide bar as shown below.

Table 2.6: Entries of the Starting Values screen for nominal outcomes

Number	Caption	Purpose	Type	Action
3	Explanatory Variables	To specify the starting value(s) for the coefficients of explanatory variable(s) of the fixed part of the model.	Grid box with slide bar	Enter a real number in the corresponding Value box(es) of the variable(s) of interest.
4	Level 2 (Co)variances Starting Values	To specify the starting value(s) for the variance(s) and/or covariance(s) of the level-2 random effects.	Grid box with slide bar	Enter a real number (positive for variances) in the corresponding Value box(es) of the variable(s) of interest.
5	Level 3 (Co)variances Starting Values	To specify the starting value(s) for the variance(s) or covariance(s) of the level-3 random effects.	Grid box with slide bar	Enter a real number (positive for variances) in the corresponding Value box(es) of the variable(s) of interest.

The 3 different entries shown in bold typeface are either new or different compared with those on the **Starting Values** screen of the **Model Setup** window for continuous or count outcome variables. These 4 entries are summarized in Table 2.6. Please refer to Table 2.4 for the information about all the other entries.



2.4.4 The Patterns screen

This screen is used to specify patterns or structures for the coefficients of the explanatory variables and variances and covariances of the random effects. A typical **Patterns** screen is shown below. Note that the default numbers for these patterns (1, 2, 3, …) are dependent on the number of parameters listed in a grid. The default numbers indicate that all parameters are set free. On the other hand, if a number is replaced by a '0', the corresponding parameter is fixed to the default or user-specified value on the **Starting Values** screen.

Patterns screen for continuous, ordered, nominal and count outcomes

The 6 different entries of the **Patterns** screen of the **Model Setup** window for continuous, ordered, nominal, or count response variables are summarized in Table 2.7. For an ordinal outcome variable, provision is also made for entering user-defined values for threshold parameters.

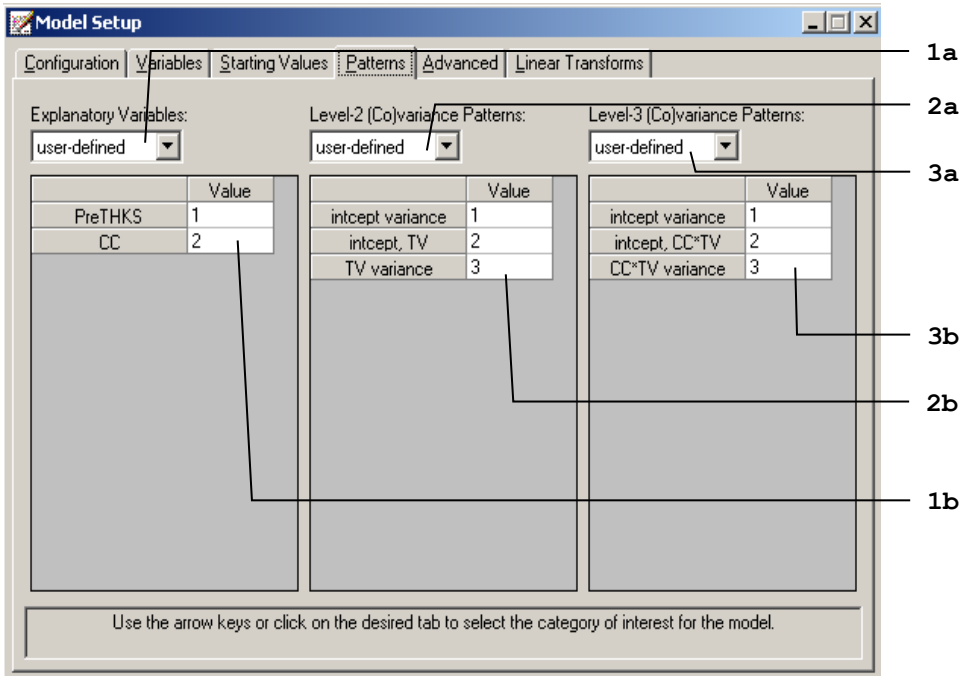


Table 2.7: Entries of the Patterns screen for continuous, count and nominal outcomes

Number		Caption	Purpose	Type	Action	Options
1	a	Explanatory Variables	To specify the pattern type for the coefficients in the fixed part of the model.	Drop-down list box	Select an option from the drop-down list box.	free (default) **user-defined**
	b		To specify the pattern for the covariance matrix of the fixed part of the model.	Text box	Enter integer values >= 0	
2	a	Level-2 (Co)variance Patterns	To specify the pattern type for the covariance matrix of the level-2 random effects.	Drop-down list box	Select an option from the drop-down list box.	correlated (default) independent unidimensional **user-defined**
	b		To specify the pattern for the covariance matrix of the level-2 random effects.	Text box	Enter integer values >= 0	
3	a	Level-3 (Co)variance Patterns	To specify the pattern type for the covariance matrix of the level-3 random effects.	Drop-down list box	Select an option from the drop-down list box.	correlated (default) independent unidimensional **user-defined**
	b		To specify the pattern for the covariance matrix of the level-3 random effects.	Text box	Enter integer values >= 0	

Examples of Patterns:

- The pattern below is used to constrain the coefficients of Treatment 1 and Treatment 2 to be equal. Likewise, the coefficients of Treatment 3 and Treatment 4 are constrained to be equal.

Explanatory Variables	Pattern
Treatment 1	1
Treatment 2	1
Treatment 3	3
Treatment 4	3

Note that a number cannot be larger than the row number on the grid. For example, the following pattern is not recognized by SuperMix:

Explanatory Variables	Pattern
Treatment 1	2
Treatment 2	2
Treatment 3	4
Treatment 4	4

- The table below shows three possible patterns for the level-2 variances and covariances of the random effects Time1, Time2, Time3, and Time4.

Level-3 co(variance)	Pattern 1	Pattern 2	Pattern 3
Variance, Time1	1	1	1
Time1, Time2	2	2	2
Variance, Time2	1	3	3
Time1, Time3	2	0	0
Time2, Time3	2	2	0
Variance, Time3	1	6	6
Time1, Time4	2	0	0
Time2, Time4	2	0	0
Time3, Time4	2	2	9
Variance, Time4	1	10	10

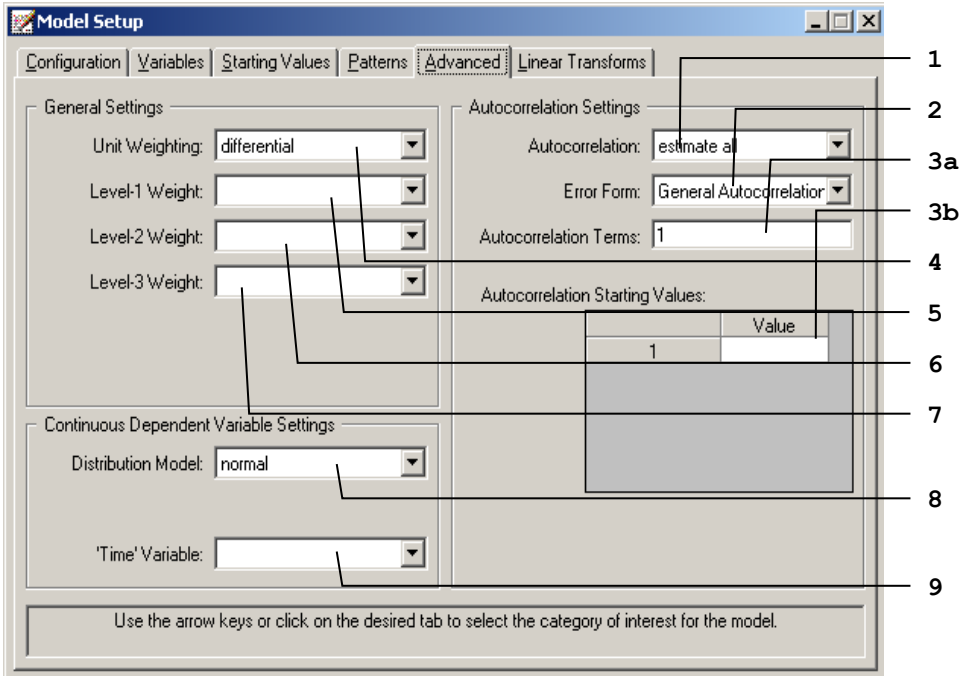
Pattern 1 restricts all the variances to be equal and, likewise, all the covariances to be equal. Pattern 2 specifies that all variances should be estimated freely, all covariances one time unit apart are set equal, and all covariances more than one time unit apart are fixed at the values specified on the **Starting Values** screen, the default for covariances being zero. Pattern 3 specifies that Time1 and Time2 are correlated, but uncorrelated with Time3 and Time4, which are correlated with each other.

2.4.5 The Advanced screen

The appearance of the **Advanced** screen depends on the type of outcome variable selected on the **Configuration** screen, and is used to change default settings used in SuperMix. Specific examples of the use of this screen are given in Chapters 3 to 8. Screens for the various outcome types are given next.

Advanced screen for continuous outcomes – normal distribution

In repeated measurement studies, the assumption of uncorrelated identically distributed level-1 error terms is often unrealistic. The options on the **Advanced** screen shown below allow for correlated level-1 error terms that follow a time series process.



The 9 different entries of the **Advanced** screen of the **Model Setup** window for continuous response variables are summarized in Table 2.8(a).

Table 2.8(a): Entries of the Advanced screen for continuous outcomes with normal distribution

Number		Caption	Purpose	Type	Action	Options
1		Autocorrelation	To specify the type of autocorrelation terms.	Drop-down list box	Select an option from the drop-down list box.	no AC terms (default)
						fixed AC terms
						estimate all
2		Error Form	To specify a time series model for the auto-correlated errors.	Drop-down list box	Select an option from the drop-down list box.	stationary AR1 (default)
						non-stationary AR1
						stationary MA1
						stationary ARMA(1,1)
						general Auto-correlation
3	a	Autocorrelation Terms	To specify the number of autocorrelation terms.	Text box	Enter an integer if the default 1 is not desired.	
	b	Autocorrelation Starting Values	To specify the starting value(s) for the autocorrelation(s).	Grid box	Enter a real number in the region of [-0.99, 0.99].	
4		Unit Weighting	To select equal or differential weighting for the unites for continuous dependent variable.	Drop-down list box	Select an option from the drop-down list box.	equal (default)
						differential
5		Level-1 Weight	To specify the weight variable that defines the first level of the hierarchy in the data.	Drop-down list box	Select a variable from the drop-down list box.	

Table 2.8(a): Entries of the Advanced screen for continuous outcomes with normal distribution (continued)

6	Level-2 Weight	To specify the weight variable that defines the second level of the hierarchy in the data.	Drop-down list box	Select a variable from the drop-down list box.	
7	Level-3 Weight	To specify the weight variable that defines level-3 of the hierarchy in the data.	Drop-down list box	Select a variable from the drop-down list box.	
8	Distribution Model	To select an appropriate distribution model.	Drop-down list box	Select a distribution from the drop-down list.	**normal** (default) gamma inverse Gaussian
9	'Time' Variable	To specify the time variable.	Drop-down list box	Select a variable from the drop-down list box.	

Advanced screen for continuous outcomes – gamma/inverse Gaussian distribution

When the gamma or inverse Gaussian distribution is selected, the **Advanced** screen is a little different from when the normal distribution is selected as shown below. The 4 different entries of the **Advanced** screen of the **Model Setup** window for continuous response variables are summarized in Table 2.8(b).

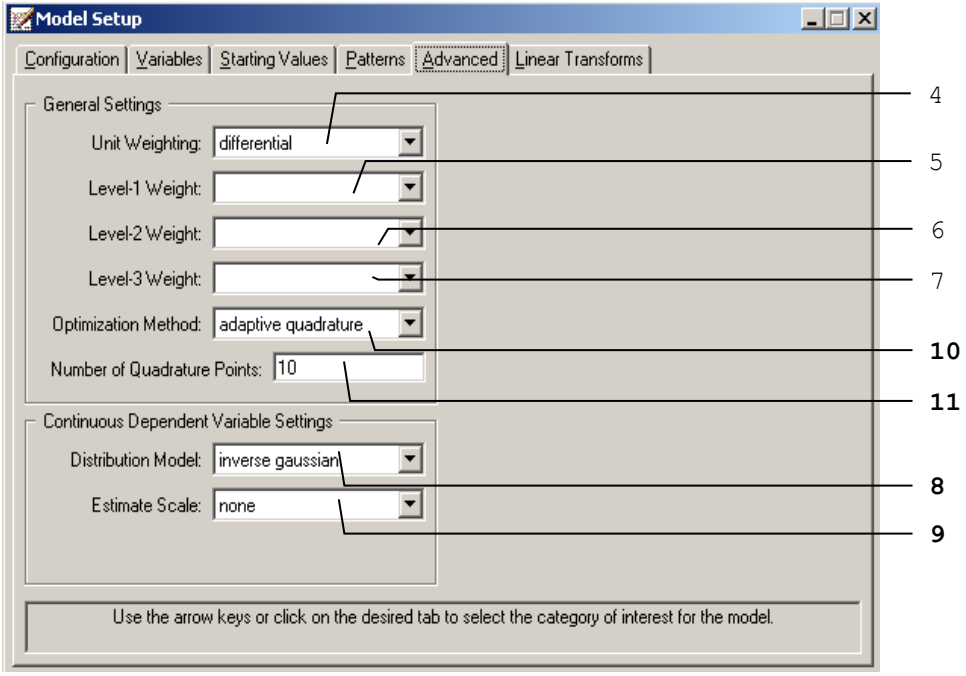


Table 2.8(b): Entries of the Advanced screen for continuous outcomes with gamma/inverse Gaussian distribution

Number	Caption	Purpose	Type	Action	Options
8	Distribution Model	To select an appropriate distribution model.	Drop-down list box	Select a distribution from the drop-down list.	normal (default) gamma **inverse Gaussian**
9	Estimate Scale	To specify the method for estimating the scale.	Drop-down list box	Select an estimated scale from the drop-down list box.	**none** (default) deviance Pearson

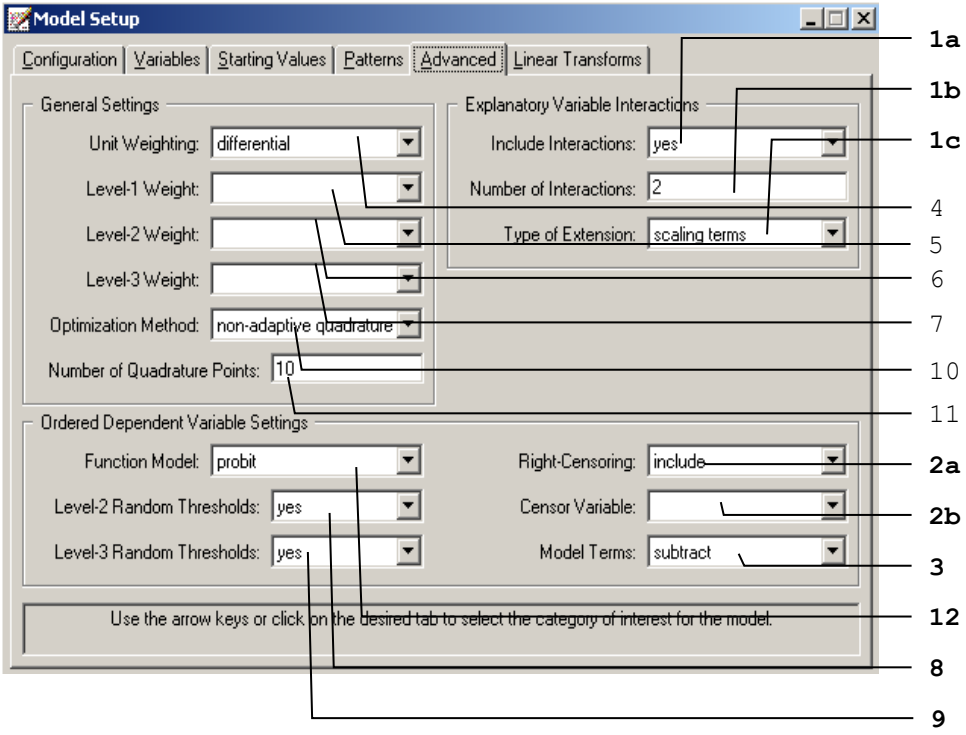
Table 2.8(b): Entries of the Advanced screen for continuous outcomes with gamma/inverse Gaussian distribution (continued)

10	Optimization Method	To select the optimization method.	Drop-down list box	Select an optimization method from the drop-down list box.	maximum posterior **adaptive quadrature** non-adaptive quadrature (default)
11	Number of Quadrature Points	To enter the quadrature points (per random-effect dimension) to use in the numerical integration.	Text box	Enter an integer if the default 10 is not desired. It is usually set to 10 for 1 effect and 5 to 10 for 2 or 3 effects.	

Advanced screen for ordered outcomes

An important feature of mixed-effects models with ordered outcomes is the inclusion of threshold parameters in the model. As illustrated in Chapter 6, the number of threshold parameters equals $C-2$, where C is the number of distinct categories of the outcome variable. If right-censoring (see Chapter 8) is included in the model specification, the number of thresholds becomes $C-1$ and a **Censor Variable** is selected. The mixed-effects model for ordinal outcomes additionally allows for the inclusion of **Explanatory Variable**-threshold interaction terms. If entry number 10 in the screen below is set equal to 2, for example, then interaction terms of the first two explanatory variables with each of the thresholds are included in the model. One can also select a weight variable, link function (**Function model**) and the number of quadrature points to be used for the approximation of the likelihood function by numerical integration.

Table 2.9 gives a summary of the 12 possible entries of the **Advanced** screen of the **Model Setup** window for an ordered response variable.



As shown in the above image, the 6 entries shown in bold typeface are either new or different compared with those on the **Advanced** screen of the **Model Setup** window for continuous outcome variables. These 6 entries are summarized in Table 2.9. Please refer to Table 2.8(a) and (b) for the information about all the other entries.

Table 2.9: Entries of the Advanced screen for ordered outcomes

Number		Caption	Purpose	Type	Action	Options
1	a	Include Interactions	To indicate if explanatory variable interactions should be include in the model.	Drop-down list box	Select an option from the drop-down list box.	no (default) **yes**
	b	Number of Interactions	To specify the number of interactions.	Text box	Enter an integer if the default maximum allowable value is not desired.	
	c	Type of Extension	To select whether to treat the explanatory variables as having scaling effects or threshold interactions.	Drop-down list box	Select an option from the drop-down list box.	scaling terms **threshold interactions (default)**
2	a	Right-Censoring	To specify is right-censoring is included.	Drop-down list box	Select an option from the drop-down list box.	none (default) **include**
	b	Censor Variable	To specify the censor variable.	Drop-down list box	Select a variable from the drop-down list box.	
3		Model Terms	To select subtracting or adding the model terms to the threshold.	Drop-down list box	Select an option from the drop-down list box.	**subtract means** $(\gamma - X'\beta)$ **(default)** add, means $(\gamma + X'\beta)$
8		Level-2 Random Thresholds	To specify if there are thresholds for the level-2 random effects.	Drop-down list box	Select an option from the drop-down list box.	no (default) **yes**
9		Level-3 Random Thresholds	To specify if there are thresholds for the level-3 random effects.	Drop-down list box	Select an option from the drop-down list box.	no (default) **yes**
12		Function Model	To specify the link function.	Drop-down list box	Select an option from the drop-down list box.	**probit** (default), logistic, complementary log-log, log-log

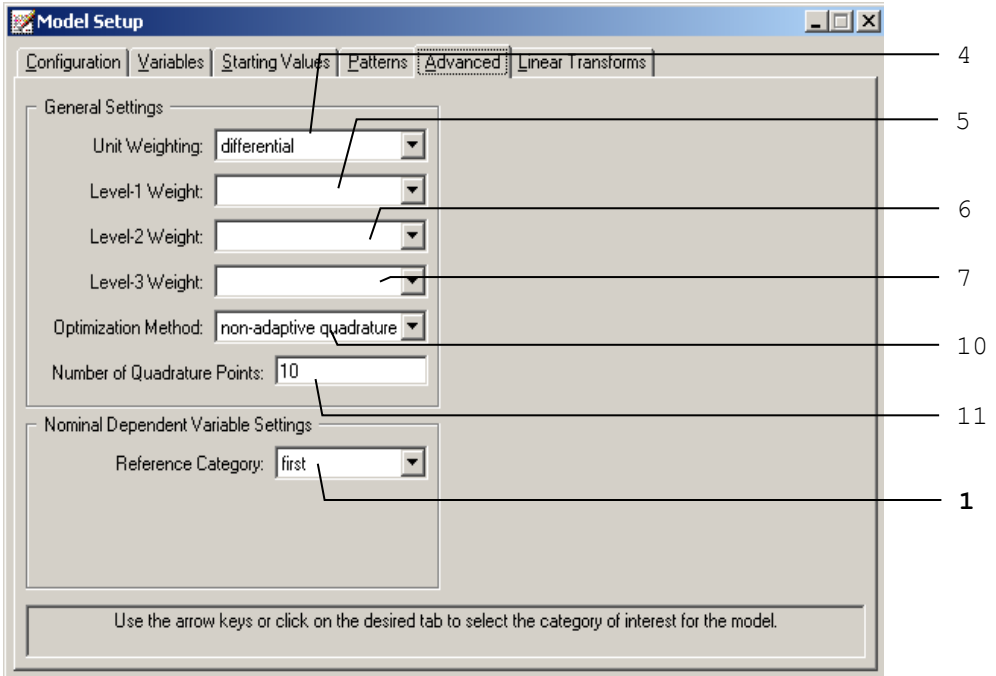
Advanced screen for nominal outcomes

The general settings of this screen is similar to the one used for the continuous and ordinal variables, but nominal dependent variable settings is different as shown in the bold entry number 1 below.

Table 2.10: Entries of the Advanced screen for nominal outcomes

Number	Caption	Purpose	Type	Action	Options
1	Reference Category	To select whether the first or last category of the outcome should be used as the reference category.	Drop-down list box	Select a reference category from the drop-down list.	**first** (default)
					last

Note that all the information for the other 6 entries are given in Table 2.8(a) and (b).



Advanced screen for count outcomes with Poisson distribution

The screen below has the same functionality as the screen used for nominal outcomes, except that provision is made here for the specification of an **Offset Variable**. In practice, it can occur that the coefficient of some covariate is assumed to be unity. This covariate is commonly known as an offset variable. Offsets are typically used when the response variable is a rate rather than a number or count.

The 2 entries pertaining to the offset variable on the **Advanced** screen of the **Model Setup** window for a count outcome are summarized in Table 2.11(a). Note that entries in the **Advanced** screen for the count outcome variable are similar to those for the ordered outcome. The information for the other 2 entries are given in Table 2.9.

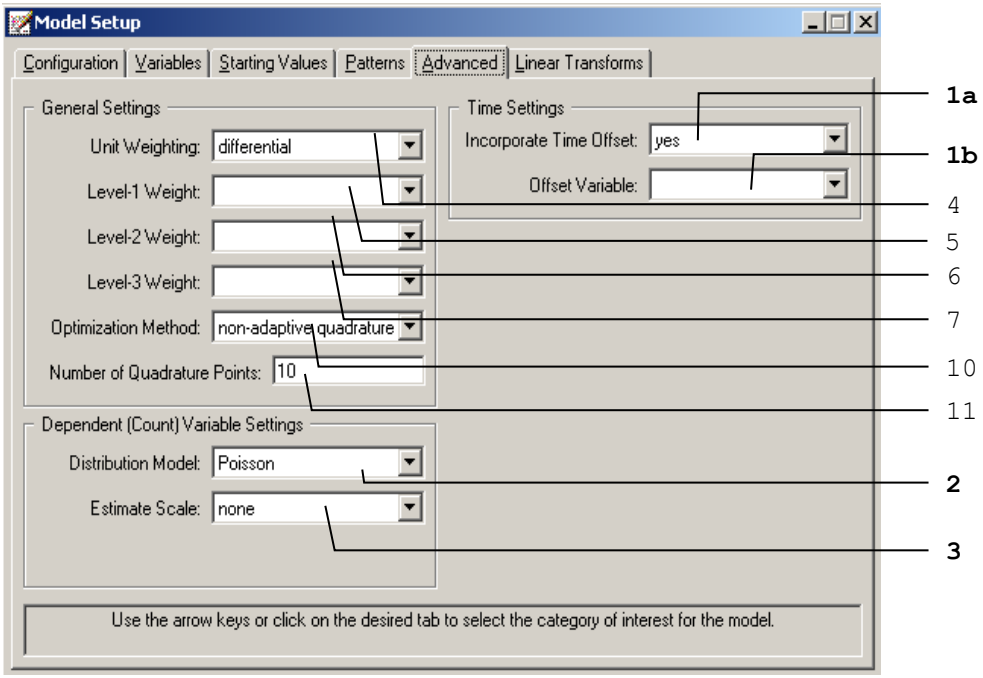
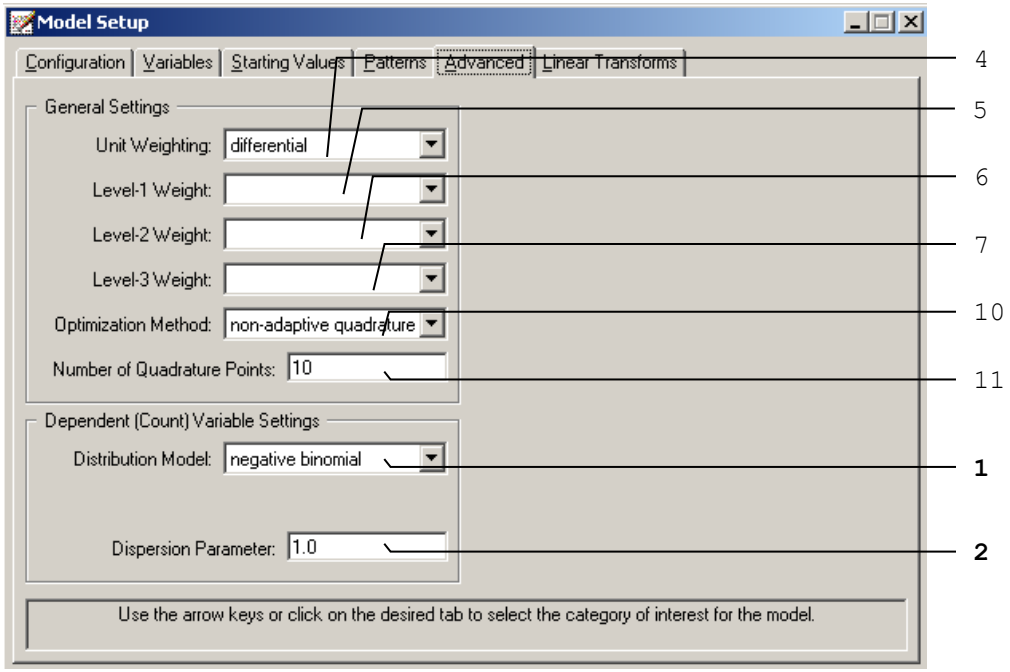


Table 2.11(a): Entries of the Advanced screen for counts with Poisson distribution

Number		Caption	Purpose	Type	Action	Options
1	a	Incorporate Time Offset	To select whether or not to include an offset variable.	Drop-down list box	Select an option from the drop-down list box.	no (default) **yes**
	b	Offset Variable	To specify the offset variable.	Drop-down list box	Select a variable from the drop-down list box.	
2		Distribution Model	To select an appropriate distribution model.	Drop-down list box.	Select a distribution from the drop-down list box.	**Poisson** (default) negative binomial
3		Estimate Scale	To specify the method for estimating the scale.	Drop-down list box	Select an estimated scale from the drop-down list box.	**none** (default) deviance Pearson

Advanced screen for count outcomes with negative binomial distribution

When the negative binomial distribution is selected, the **Advanced** screen of the count variable is slightly different as shown below.



As shown above, the bold font entries 1 and 2 are different from the previous screen.

Table 2.11(b): Entries of the Advanced screen for count outcomes with negative binomial distribution

Number	Caption	Purpose	Type	Action	Options
1	Distribution Model	To select an appropriate distribution model.	Drop-down list box.	Select a distribution from the drop-down list box.	Poisson (default) **negative binomial**
2	Dispersion Parameter	To enter the dispersion parameter for the negative binomial model.	Text box	Enter any numeric value greater than 0.0. The default value is 1.0.	

Advanced screen for binary outcomes

When the negative binomial distribution is selected, the **Advanced** screen of the count variable is slightly different: the bold font entries 1 and 2 are different from the previous screen.

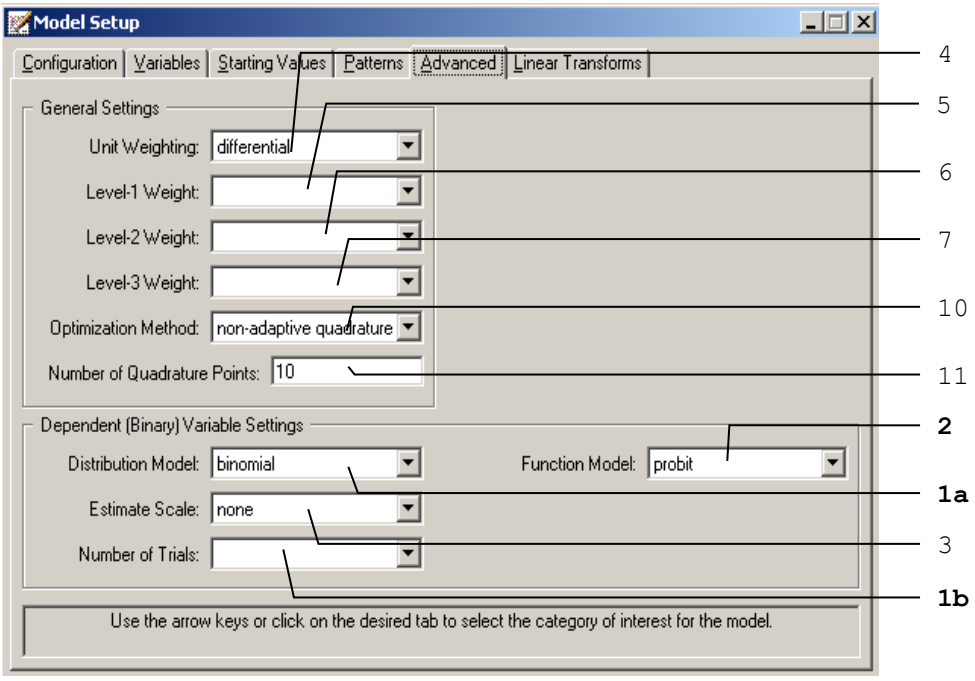


Table 2.12: Entries of the Advanced screen for binary outcomes

Number		Caption	Purpose	Type	Action	Options
1	a	Distribution Model	To select an appropriate distribution model.	Drop-down list box.	Select a distribution from the drop-down list box.	Bernoulli (default) **binomial**
	b	Number of Trials	To select the column of the spreadsheet, which contains the no. of trials.	Drop-down list box	Select a variable from the drop-down list box.	

Table 2.12: Entries of the Advanced screen for binary outcomes (continued)

2	Function Model	To select an appropriate link function.	Drop-down list box.	Select a link function from the drop-down list box.	**probit** (default)
					logistic
					complementary log-log
					log-log
3	Estimate Scale	To specify the method for estimating the scale.	Drop-down list box	Select an estimated scale from the drop-down list box.	**none** (default)
					deviance
					Pearson

2.4.6 The Linear Transforms screen

Linear transforms are used to test hypotheses of the type

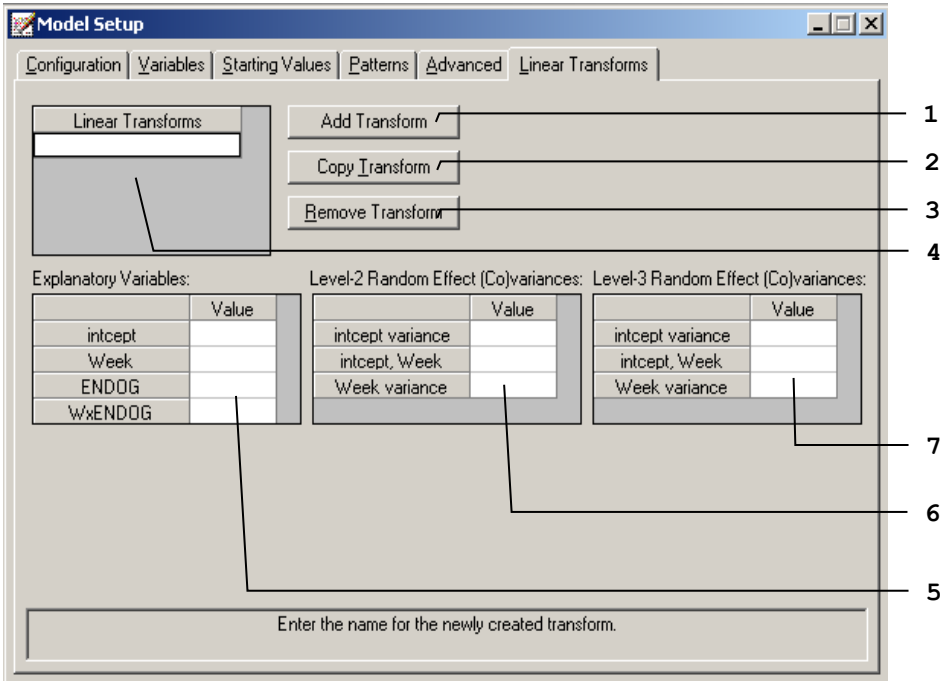
$$H_0 : c_1\beta_1 + c_2\beta_2 + c_3\beta_3 + ... + c_k\beta_k = 0$$

where $\beta_1, \beta_2, \beta_3, ..., \beta_k$ are model parameters and $c_1, c_2, c_3, ..., c_k$ user-specified real-valued coefficients. For each linear transform, a Z-statistic and an associated two-tailed p-value are saved to the output file. The Z-statistic is a function of the estimated parameters and the large sample covariance matrix of the estimates. The value of the linear transform when the parameters are replaced with their estimates is also of interest. This value also appears in the output.

An example of the **Linear Transform** screen in the case of an ordinal outcome variable is given below (see Chapter 6 for an additional example.) For continuous, count and nominal variables the **Linear Transform** screens are identical, but it differ from that for an ordinal outcome in that the screen for an ordinal outcome contains threshold parameter information.

Linear Transforms screen for continuous and count outcomes

The buttons and options on the **Linear Transforms** screen are shown below.



A summary of the 3 buttons and 4 different entries of the **Linear Transforms** screen of the **Model Setup** window for an ordered outcome is given in Table 2.12.

Table 2.13: Entries of the Linear Transforms screen for continuous and count outcomes

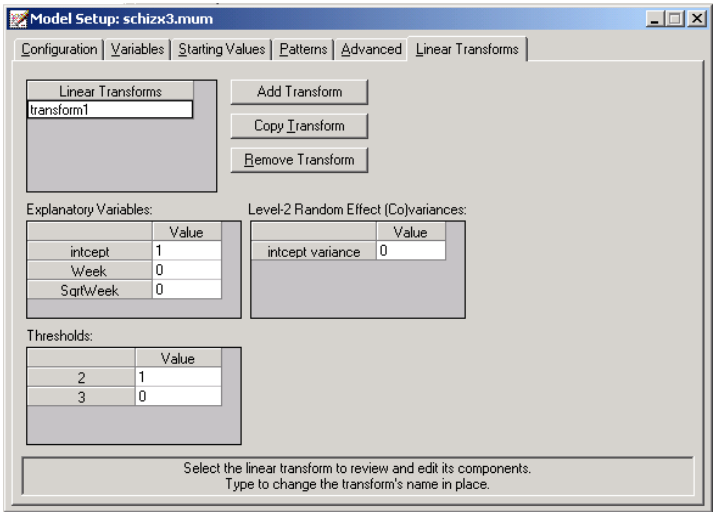
Number	Caption	Purpose	Type	Action
1	Add Transform	To create a new linear transform in 4.	Click Button	Click to add a blank transform.
2	Copy Transform	To create a copy the selected transform in 4 with a different name.	Click Button	Click to copy and paste the select transform.

Table 2.13: Entries of the Linear Transforms screen for continuous and count outcomes (continued)

3	Remove Transform	To delete the selected transform in 4.	Click Button	Click on the button to delete the selected transform.
4	Linear Transforms	To select the linear transform and edit it's components.	Grid box	Enter string(s) as names for transforms.
5	Explanatory Variables	To specify the coefficient(s) for the linear transformation(s) of the fixed part of the model.	Grid box	Enter real number(s).
6	Level-2 Random Effect (Co)variances	To specify the coefficient(s) for the linear transformation(s) of the covariance matrix of the level-2 random effects.	Grid box	Enter real number(s).
7	Level-3 Random Effect (Co)variances	To specify the coefficient(s) for the linear transformation(s) of the covariance matrix of the level-3 random effects.	Grid box	Enter real number(s).

For example, in the **Linear Transforms** screen shown below we wish to test the hypothesis that

$$\beta_{intercept} + \beta_{Threshold1} = 0.$$



Linear Transforms screen for ordered outcomes

Additional grid boxes appear to allow for user-specified starting values of thresholds and threshold-explanatory variable(s) interaction (s) for ordinal outcomes. The following screen is an example of the **Linear Transforms** screen of the **Model Setup** window.

Only the 2 entries shown in bold typeface are either new or different compared with those on the **Linear Transforms** screen of the **Model Setup** window for continuous or count outcome variables. These 2 entries are summarized in Table 2.13. Please refer to Table 2.12 for the information about all the other entries.

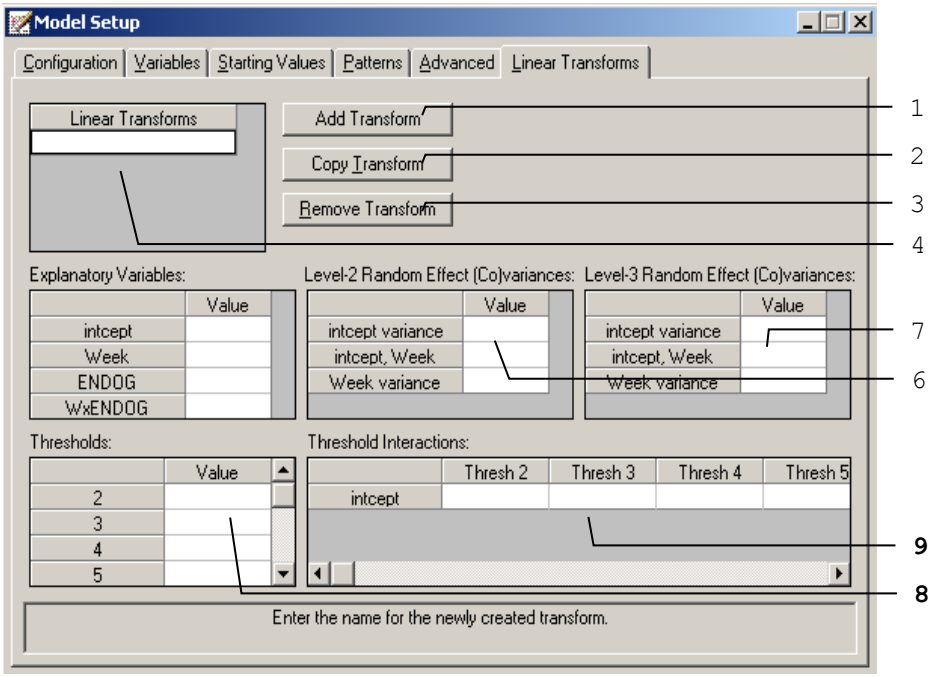
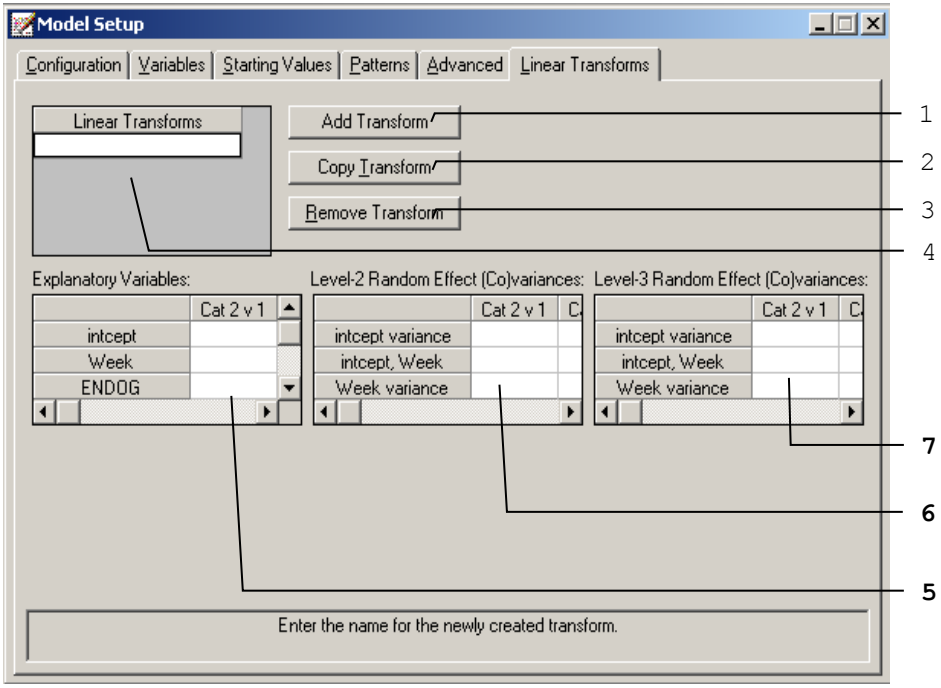


Table 2.14: Entries of the Linear Transforms screen for ordered outcomes

Number	Caption	Purpose	Type	Action
8	Thresholds	To specify the coefficient(s) for the linear transformation(s) of the thresholds.	Grid box	Enter real number(s).
9	Thresholds Interactions	To enter thresholds for the selected transform	Grid box	Enter real number(s).

Linear Transforms screen for nominal and binary outcomes

Additional slide bar appears for as shown in the following screen is an example of the **Linear Transforms** screen of the **Model Setup** window.



Only the 2 entries shown in bold typeface are either new or different compared with those on the **Linear Transforms** screen of the **Model Setup** window for continuous or count outcome variables. These 2 entries are summarized in Table 2.14. Please refer to Table 2.12 for the information about all the other entries.

Table 2.15: Entries of the Linear Transforms screen for nominal outcomes

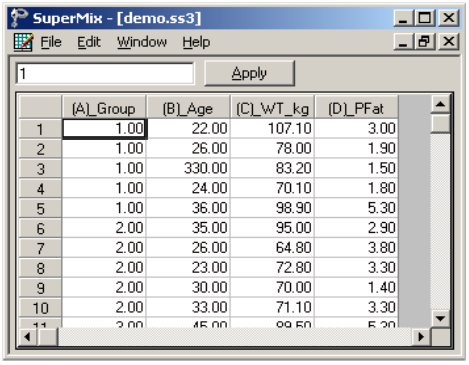
Number	Caption	Purpose	Type	Action
5	Explanatory Variables	To specify the coefficient(s) for the linear transformation(s) of the fixed part of the model.	Grid box with slide bar	Enter real number(s).
6	Level-2 Random Effect (Co)variances	To specify the coefficient(s) for the linear transformation(s) of the covariance matrix of the level-2 random effects.	Grid box with slide bar	Enter real number(s).
7	Level-3 Random Effect (Co)variances	To specify the coefficient(s) for the linear transformation(s) of the covariance matrix of the level-3 random effects.	Grid box with slide bar	Enter real number(s).

2.5 Data manipulation

The SuperMix spreadsheet can be manipulated in various ways. Rows and columns can be changed directly, and simple computations or more complex built-in functions can be used in individual cells. Some of these manipulations are discussed and illustrated in the following sections, using **demo.ss3**. There are 15 cases (patients) in the data set.

The spreadsheet window for **demo.ss3** is opened as follows:

- Use the **Open** option on the **File** menu of the main window to load the **Open Spreadsheet** dialog box.
- Browse for the file **demo.ss3** in the **Examples** folder.
- Select the file and click on the **Open** button to open the following SuperMix spreadsheet window.



The variables include:

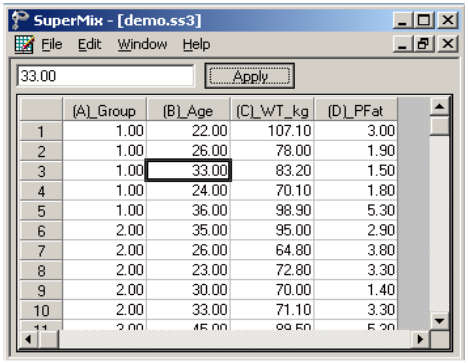
- Group is a variable with 3 categories, indicating the group number of the patient (5 patients in each group).
- Age is the age of the patient.
- WT_kg denotes the weight of the patient in kg.
- PFat is a measure of percentage body fat.

2.5.1 Basic data manipulations

It is important to note that any change of the data file will not be saved unless you use the **Save** option on the **File** menu.

Cells

A careful examination of the data shows that the Age entry of the 3rd observation is 330.00. This is obviously a typing error. Upon further investigation, it turns out that the correct age value is 33. To correct this error, select the cell, change the value of the formula box from 330.00 to 33.00, and then click on the **Apply** button.



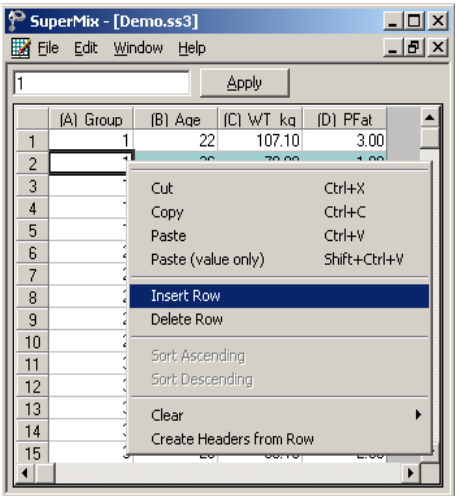
	(A)_Group	(B)_Age	(C)_WT_kg	(D)_PFat
1	1.00	22.00	107.10	3.00
2	1.00	26.00	78.00	1.90
3	1.00	33.00	83.20	1.50
4	1.00	24.00	70.10	1.80
5	1.00	36.00	98.90	5.30
6	2.00	35.00	95.00	2.90
7	2.00	26.00	64.80	3.80
8	2.00	23.00	72.80	3.30
9	2.00	30.00	70.00	1.40
10	2.00	33.00	71.10	3.30

Rows

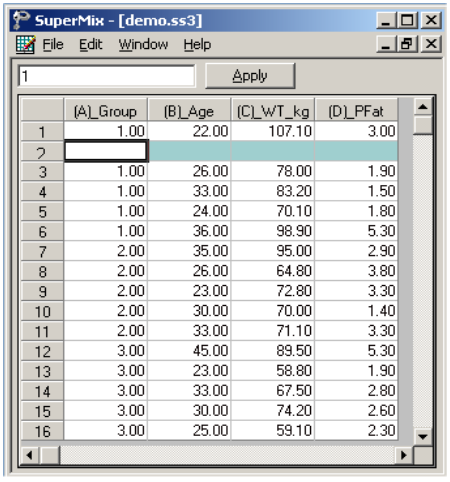
To work with a row (or rows) of the data file, click on the row tab(s) to select the complete row(s) and then right-click on the selection to display all the options from the pop-up menu.

Insert a row

For example, to insert another row (observation) between the first and the second rows, select the second row by clicking on the row 2 tab, right-click on the selected row to activate the menu and select the **Insert Row** option to create the window as shown below.



An empty row is added to the spreadsheet above the previous second row and the total sample size is changed to 16 as shown below.



Delete a row

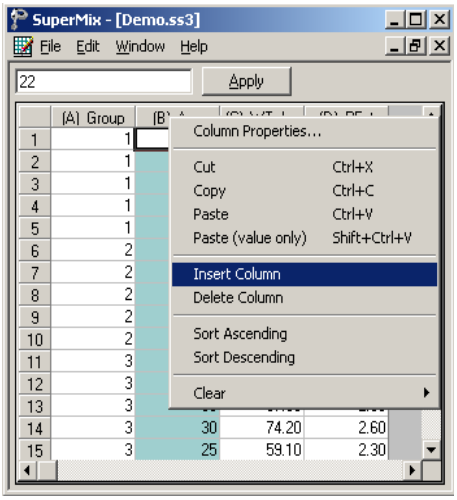
To delete the empty row that was inserted, select the second row by clicking on the row 2 tab. Right click on the selected row and select the **Delete Row** option to delete the second row.

Columns

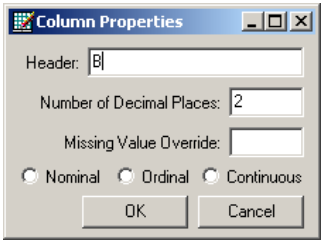
To work with a column (or columns) of the data file, first click on the column header(s) to select the column(s) and then right-click on one of the selected columns to see all the options listed on the pop-up menu.

Insert a column

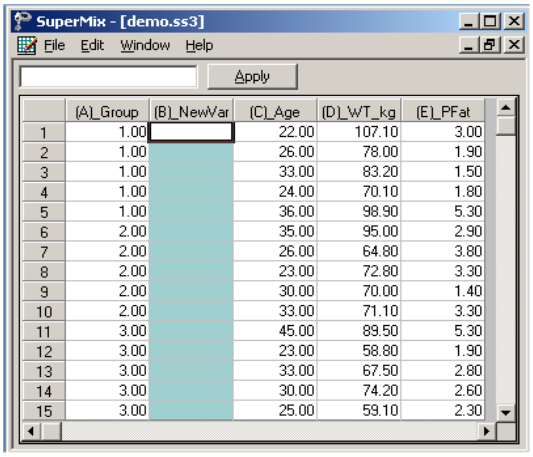
To insert another variable between (A)_Group and (B)_Age, first click on the header of (B)_Age, right-click on the column to activate the menu and select the **Insert Column** option to create the window as shown below.



A new column (variable) is added to the spreadsheet. Change the variable name by selecting the column header, right-click and select **Column Properties** to load the dialog box as shown below.



Input a variable name, such as NewVar, and then click on the **OK** button to return to the spreadsheet window as shown below.



	(A)_Group	(B)_NewVar	(C)_Age	(D)_WT_kg	(E)_PFat
1	1.00		22.00	107.10	3.00
2	1.00		26.00	78.00	1.90
3	1.00		33.00	83.20	1.50
4	1.00		24.00	70.10	1.80
5	1.00		36.00	98.90	5.30
6	2.00		35.00	95.00	2.90
7	2.00		26.00	64.80	3.80
8	2.00		23.00	72.80	3.30
9	2.00		30.00	70.00	1.40
10	2.00		33.00	71.10	3.30
11	3.00		45.00	89.50	5.30
12	3.00		23.00	58.80	1.90
13	3.00		33.00	67.50	2.80
14	3.00		30.00	74.20	2.60
15	3.00		25.00	59.10	2.30

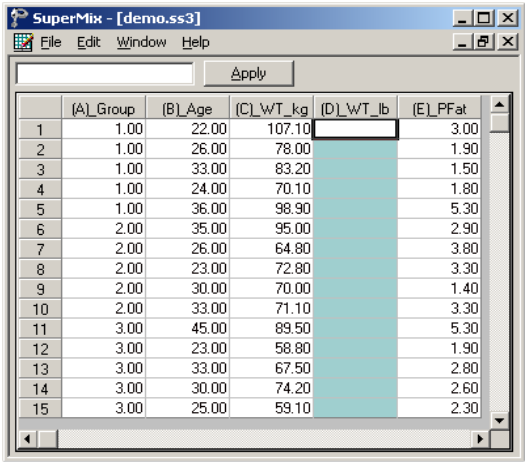
Delete a column

To delete the NewVar column that was inserted, select the variable NewVar by clicking on the column header. Right-click and select the **Delete Column** option to delete column B.

2.5.2 Simple computations

Assigning values to a new variable

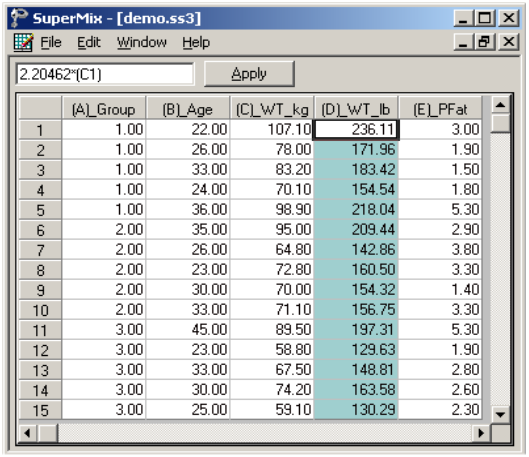
In **demo.ss3**, the variable WT_kg is a variable denoting weight in kilograms. We would like to use the variable WT_lb, that is, the corresponding weight in pounds. To create this variable, first insert a column and change the column header to WT_lb as illustrated earlier in Section 2.5.1 to generate the following spreadsheet window.



	(A)_Group	(B)_Age	(C)_WT_kg	(D)_WT_lb	(E)_PFat
1	1.00	22.00	107.10		3.00
2	1.00	26.00	78.00		1.90
3	1.00	33.00	83.20		1.50
4	1.00	24.00	70.10		1.80
5	1.00	36.00	98.90		5.30
6	2.00	35.00	95.00		2.90
7	2.00	26.00	64.80		3.80
8	2.00	23.00	72.80		3.30
9	2.00	30.00	70.00		1.40
10	2.00	33.00	71.10		3.30
11	3.00	45.00	89.50		5.30
12	3.00	23.00	58.80		1.90
13	3.00	33.00	67.50		2.80
14	3.00	30.00	74.20		2.60
15	3.00	25.00	59.10		2.30

Select the column containing the variable WT_lb, input the function **2.20462*(C1)** in the formula box and click on the **Apply** button to get the new variable WT_lb as shown below. The formula applies to each row of (D)_WT_lb provided that

- this column is selected (highlighted)
- the first cell of the variable(s) in the formula, in this case C1, is referenced.



	(A)_Group	(B)_Age	(C)_WT_kg	(D)_WT_lb	(E)_PFat
1	1.00	22.00	107.10	236.11	3.00
2	1.00	26.00	78.00	171.96	1.90
3	1.00	33.00	83.20	183.42	1.50
4	1.00	24.00	70.10	154.54	1.80
5	1.00	36.00	98.90	218.04	5.30
6	2.00	35.00	95.00	209.44	2.90
7	2.00	26.00	64.80	142.86	3.80
8	2.00	23.00	72.80	160.50	3.30
9	2.00	30.00	70.00	154.32	1.40
10	2.00	33.00	71.10	156.75	3.30
11	3.00	45.00	89.50	197.31	5.30
12	3.00	23.00	58.80	129.63	1.90
13	3.00	33.00	67.50	148.81	2.80
14	3.00	30.00	74.20	163.58	2.60
15	3.00	25.00	59.10	130.29	2.30

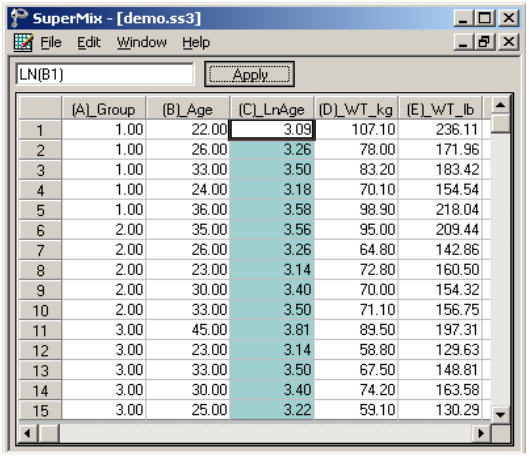
Save the changes to **demo.ss3** by clicking on the **Save** option on the **File** menu.

2.5.3 Built-in functions

LN function

In mixed-effects models, we often consider the natural log of a variable. For example, the natural log of Age in **demo.ss3** may be a more appropriate explanatory variable than the original age given in years. The variable LnAge can easily be created by using the options available in the SuperMix spreadsheet.

First, create a new column with the header LnAge. Next, select the column containing the variable LnAge, input the function **LN(B1)** in the formula box, and click on the **Apply** button. Each value of the new variable LnAge is the natural log of the corresponding values of the variable Age as shown below.

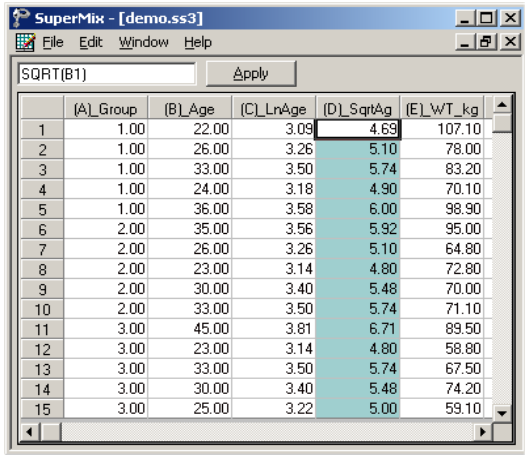


	(A)_Group	(B)_Age	(C)_LnAge	(D)_WT_kg	(E)_WT_lb
1	1.00	22.00	3.09	107.10	236.11
2	1.00	26.00	3.26	78.00	171.96
3	1.00	33.00	3.50	83.20	183.42
4	1.00	24.00	3.18	70.10	154.54
5	1.00	36.00	3.58	98.90	218.04
6	2.00	35.00	3.56	95.00	209.44
7	2.00	26.00	3.26	64.80	142.86
8	2.00	23.00	3.14	72.80	160.50
9	2.00	30.00	3.40	70.00	154.32
10	2.00	33.00	3.50	71.10	156.75
11	3.00	45.00	3.81	89.50	197.31
12	3.00	23.00	3.14	58.80	129.63
13	3.00	33.00	3.50	67.50	148.81
14	3.00	30.00	3.40	74.20	163.58
15	3.00	25.00	3.22	59.10	130.29

SQRT function

To add another variable, SqrtAge, which is the square root of Age, we proceed as follows. As above, first create a new column with the header SqrtAge.

Select the column containing the variable SqrtAge, input the function **SQRT(B1)** in the formula box, and click on the **Apply** button. Each value of the new variable SqrtAge is the square root value of the corresponding value of the variable Age as shown below.



	(A)_Group	(B)_Age	(C)_LnAge	(D)_SqrtAg	(E)_WT_kg
1	1.00	22.00	3.09	4.69	107.10
2	1.00	26.00	3.26	5.10	78.00
3	1.00	33.00	3.50	5.74	83.20
4	1.00	24.00	3.18	4.90	70.10
5	1.00	36.00	3.58	6.00	98.90
6	2.00	35.00	3.56	5.92	95.00
7	2.00	26.00	3.26	5.10	64.80
8	2.00	23.00	3.14	4.80	72.80
9	2.00	30.00	3.40	5.48	70.00
10	2.00	33.00	3.50	5.74	71.10
11	3.00	45.00	3.81	6.71	89.50
12	3.00	23.00	3.14	4.80	58.80
13	3.00	33.00	3.50	5.74	67.50
14	3.00	30.00	3.40	5.48	74.20
15	3.00	25.00	3.22	5.00	59.10

Save the changes to **demo.ss3** by clicking on the **Save** option on the **File** menu.

Table 2.16 contains a selection of the built-in functions in SuperMix. A list of values can be any of the following types:

- (B1, B2, B3) is the selection of the first three values of the variable in column B.
- (A1:A15) selects all the values of the variable in column A. Starting with the first and ending with the one in row 15.
- (A1:A5, A7, A11:A15) selects the values of row 1 to row 5, row 7, and row 11 to row 15 of column A.
- (A3:A6, B7, D12:D15) contains a list of values from more than one column. This selection includes the values of row 3 to row 6 of column A, row 7 of column B and row 12 to row 15 of column D.

Table 2.16: Selection of SuperMix functions

Function	Definition
ABS(value)	Absolute value
AVERAGE (list of values)	Average
EXP(value)	Exponent base e
LN(value)	Natural log
LOG(value)	Logarithm
MAX(list of values)	Maximum value
MEDIAN(list of values)	Median
MIN(list of values)	Minimum value
MODE(list of values)	Mode
SQRT(value)	Square root
SQUARE(value)	Square

2.5.4 Other useful data manipulations

The data manipulation capabilities of the SuperMix spreadsheet window can be used to create interaction terms, essentially a product of variables, for use in modeling. It can also be used to perform grand mean centering of variables.

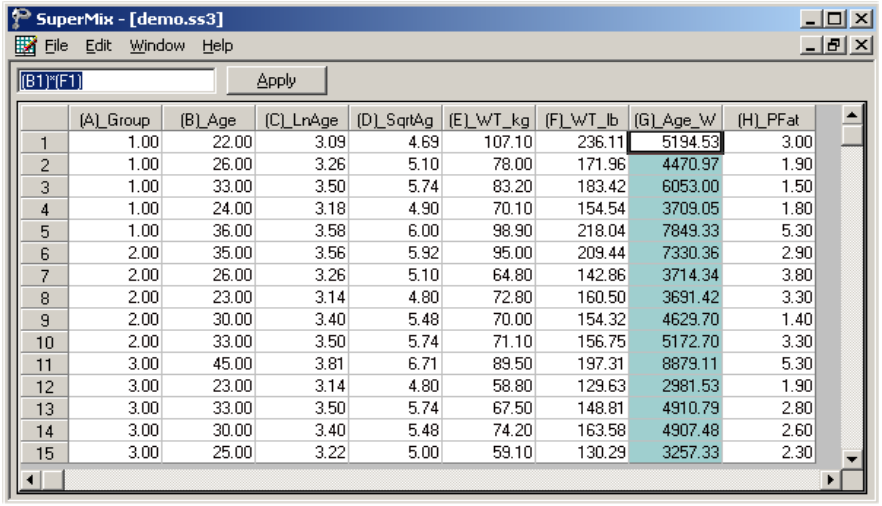
Absolute references

If you do not want SuperMix to adjust references when you copy a formula to a different cell, use an **absolute** reference. For example, if your formula multiplies cell A5 with cell C1 (=A5*C1) and you copy the formula to another cell, SuperMix will adjust both references. You can create an absolute reference to cell C1 by placing a **dollar sign** ($) before the parts of the reference that do not change. To create an absolute reference to cell C1, for example, add dollar signs to the formula as follows: = A5*C1.

Creating an interaction term

Suppose that we want to study the possible interaction between a subject's age and weight (in pounds). This product of Age and WT_lb, is created in the SuperMix spreadsheet window as follows.

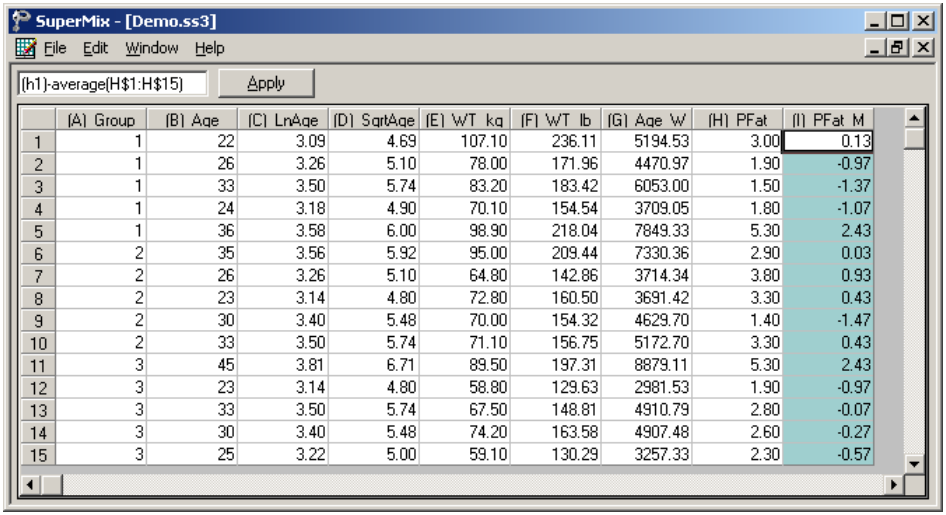
First, create a new column with header Age_WT. Then, select this column, and input the function **(B1)*(E1)** in the formula box. Click on the **Apply** button. Each value of the new variable Age_WT is equal to the product of the corresponding values of Age and WT_lb as shown below.



	(A)_Group	(B)_Age	(C)_LnAge	(D)_SqrtAg	(E)_WT_kg	(F)_WT_lb	(G)_Age_W	(H)_PFat
1	1.00	22.00	3.09	4.69	107.10	236.11	5194.53	3.00
2	1.00	26.00	3.26	5.10	78.00	171.96	4470.97	1.90
3	1.00	33.00	3.50	5.74	83.20	183.42	6053.00	1.50
4	1.00	24.00	3.18	4.90	70.10	154.54	3709.05	1.80
5	1.00	36.00	3.58	6.00	98.90	218.04	7849.33	5.30
6	2.00	35.00	3.56	5.92	95.00	209.44	7330.36	2.90
7	2.00	26.00	3.26	5.10	64.80	142.86	3714.34	3.80
8	2.00	23.00	3.14	4.80	72.80	160.50	3691.42	3.30
9	2.00	30.00	3.40	5.48	70.00	154.32	4629.70	1.40
10	2.00	33.00	3.50	5.74	71.10	156.75	5172.70	3.30
11	3.00	45.00	3.81	6.71	89.50	197.31	8879.11	5.30
12	3.00	23.00	3.14	4.80	58.80	129.63	2981.53	1.90
13	3.00	33.00	3.50	5.74	67.50	148.81	4910.79	2.80
14	3.00	30.00	3.40	5.48	74.20	163.58	4907.48	2.60
15	3.00	25.00	3.22	5.00	59.10	130.29	3257.33	2.30

Grand mean centering

It is often useful to center a predictor variable around its grand mean. To illustrate, we grand mean center the variable PFat.



	(A) Group	(B) Age	(C) LnAge	(D) SqrtAge	(E) WT kg	(F) WT lb	(G) Age W	(H) PFat	(I) PFat M
1	1	22	3.09	4.69	107.10	236.11	5194.53	3.00	0.13
2	1	26	3.26	5.10	78.00	171.96	4470.97	1.90	-0.97
3	1	33	3.50	5.74	83.20	183.42	6053.00	1.50	-1.37
4	1	24	3.18	4.90	70.10	154.54	3709.05	1.80	-1.07
5	1	36	3.58	6.00	98.90	218.04	7849.33	5.30	2.43
6	2	35	3.56	5.92	95.00	209.44	7330.36	2.90	0.03
7	2	26	3.26	5.10	64.80	142.86	3714.34	3.80	0.93
8	2	23	3.14	4.80	72.80	160.50	3691.42	3.30	0.43
9	2	30	3.40	5.48	70.00	154.32	4629.70	1.40	-1.47
10	2	33	3.50	5.74	71.10	156.75	5172.70	3.30	0.43
11	3	45	3.81	6.71	89.50	197.31	8879.11	5.30	2.43
12	3	23	3.14	4.80	58.80	129.63	2981.53	1.90	-0.97
13	3	33	3.50	5.74	67.50	148.81	4910.79	2.80	-0.07
14	3	30	3.40	5.48	74.20	163.58	4907.48	2.60	-0.27
15	3	25	3.22	5.00	59.10	130.29	3257.33	2.30	-0.57

To do so, first create a new column with the header of PFat_Mea. Then, select the PFat_Mea column, input the function **(H1)–AVERAGE(H$1:H$15)** in the formula box and click on the **Apply** button. Each value of the new variable of PFat_Mea now contains the difference between the corresponding original PFat value and the grand mean of all the PFat values. As illustrated above, the spreadsheet functions are not case sensitive.

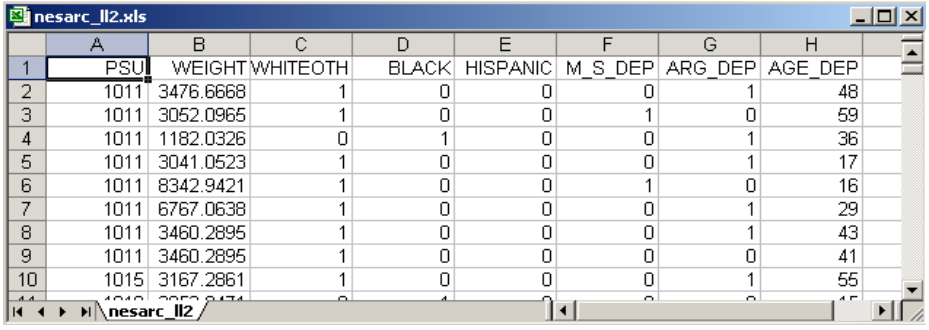
3 Models for continuous outcomes

3.1 Models based on a subset of the NESARC data

3.1.1 The data

The data set is from the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC), a longitudinal survey with its first wave fielded in 2001–2002. The NESARC is a representative sample of the United States population, and 43,093 Americans participated in the first wave of the survey. The NESARC survey was conducted and sponsored by the National Institute on Alcohol Abuse and Alcoholism (NIAAA). Detailed information is available at http://niaaa.census.gov/index.html.

Section 4 of the NESARC data documentation describes data regarding major depression, family history of major depression and dysthymia. Together with the demographic information in Section 1, we produced the **nesarc_ll2.xls** data set as shown below. There are 2,339 dysthymia respondents in the survey. After listwise deletion, the sample size is 1,698.



nesarc_ll2.xls

	A	B	C	D	E	F	G	H
1	PSU	WEIGHT	WHITEOTH	BLACK	HISPANIC	M_S_DEP	ARG_DEP	AGE_DEP
2	1011	3476.6668	1	0	0	0	1	48
3	1011	3052.0965	1	0	0	1	0	59
4	1011	1182.0326	0	1	0	0	1	36
5	1011	3041.0523	1	0	0	0	1	17
6	1011	8342.9421	1	0	0	1	0	16
7	1011	6767.0638	1	0	0	0	1	29
8	1011	3460.2895	1	0	0	0	1	43
9	1011	3460.2895	1	0	0	0	0	41
10	1015	3167.2861	1	0	0	0	1	55

nesarc_ll2

The variables of interest are:

- PSU is the Census 2000/2001 Supplementary Survey (C2SS) primary sampling unit (PSU).
- WEIGHT is the final weight, calculated as the product of the NESARC base weight and other individual weighting factors.
- WHITEOTH represents the white and other ethnicities, excluding African American and Hispanic. It is recoded from items S1Q1C, S1Q1D2, S1Q1D3 and S1Q1D5 in the NESARC source code (1 for white and other, 0 for African American and Hispanic).
- BLACK represents African Americans. It is recoded from items S1Q1C and S1Q1D3 in the NESARC source code (1 for African American, 0 for others).
- HISPANIC is an indicator for Hispanic. It is recoded from items S1Q1C, S1Q1D3 and S1Q1D5 (1 for Hispanic, 0 for others).
- M_S_DEP is recoded from item S4BQ10C. It is the response to the statement "Any of natural mother's full sisters ever depressed," with 1 for "Yes," and 0 for "No."
- ARG_DEP is recoded from item S4CQ43. It represents the response to the statement "Had arguments/friction with family, friends, people at work, or anyone else," with 1 for "Yes," 0 for "No."
- AGE_DEP is a renamed version of item S4CQ7AR. It represents the age at onset of first episode of dysthymia.

Inspection of the data shows that only about 2% of 43,093 respondents are of Asian and Pacific origin. Due to the skewness of the distribution of ethnicity, we recoded the variables representing ethnic origin. The resulting variable WHITEOTH represents this recoding of respondents as being either white or from other ethnic groups (blacks and Hispanics excluded).

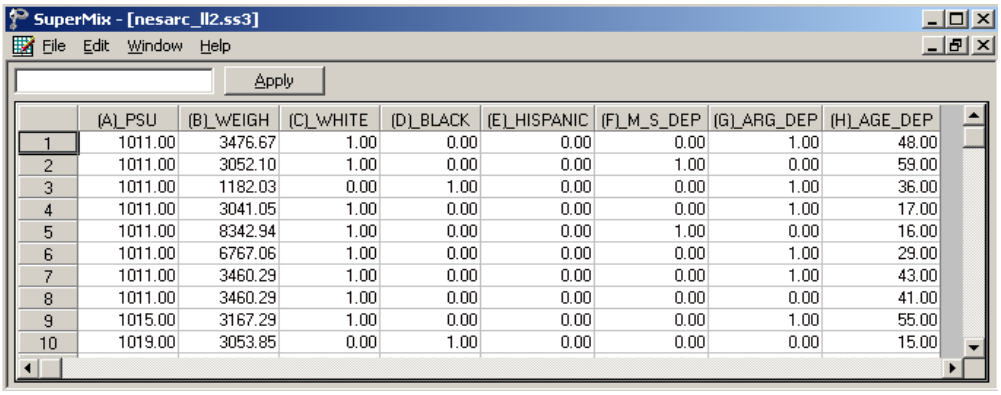
3.1.1.1 Importing the data and defining variable types

The data set shown previously is available in the form of a spreadsheet file, named **nesarc_II2.xls**. This file contains a subset of the original NESARC data, *i.e.* data for

the 1,698 respondents who reported some form of depression and for whom complete information on variables of interest was available.

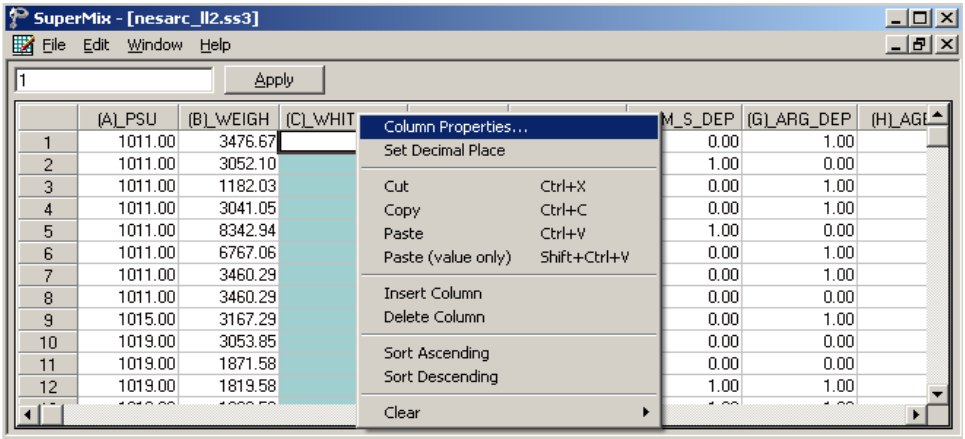
The first step is to create the SuperMix spreadsheet file (*.**ss3**) from the Excel file:

- Use the **Import Data File** option on the **File** menu to load the **Open** dialog box.
- Browse for the file **nesarc_ll2.xls** in the **examples** folder of the SuperMix installation folder.
- Select the file and click on the **Open** button to open the following SuperMix spreadsheet window **nesarc_ll2.ss3**.

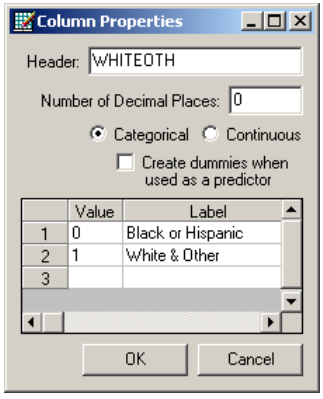


	(A)_PSU	(B)_WEIGH	(C)_WHITE	(D)_BLACK	(E)_HISPANIC	(F)_M_S_DEP	(G)_ARG_DEP	(H)_AGE_DEP
1	1011.00	3476.67	1.00	0.00	0.00	0.00	1.00	48.00
2	1011.00	3052.10	1.00	0.00	0.00	1.00	0.00	59.00
3	1011.00	1182.03	0.00	1.00	0.00	0.00	1.00	36.00
4	1011.00	3041.05	1.00	0.00	0.00	0.00	1.00	17.00
5	1011.00	8342.94	1.00	0.00	0.00	1.00	0.00	16.00
6	1011.00	6767.06	1.00	0.00	0.00	0.00	1.00	29.00
7	1011.00	3460.29	1.00	0.00	0.00	0.00	1.00	43.00
8	1011.00	3460.29	1.00	0.00	0.00	0.00	0.00	41.00
9	1015.00	3167.29	1.00	0.00	0.00	0.00	1.00	55.00
10	1019.00	3053.85	0.00	1.00	0.00	0.00	0.00	15.00

Next, we define the variable types. Highlight WHITEOTH by clicking on the variable name, and then right click to open the following pop-up menu. Select the **Column Properties** option



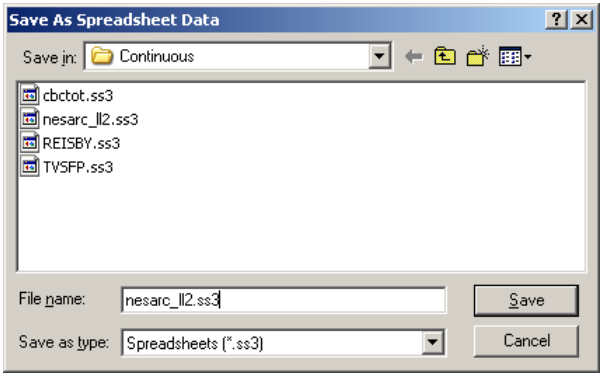
to open the **Column Properties** dialog box. Checking the **Nominal** radio button enables the user to define the labels. Input correct labels for the different categories as shown below.



Similarly define BLACK, HISPANIC, M_S_DEP and ARG_DEP as nominal variables and define AGE_DEP as continuous.

To save the **nesarc_ll2.ss3** spreadsheet, select the **Save As** option from the **File** menu to load the **Save As Spreadsheet Data** dialog box, and then enter the desired file

name in the **File name** string field as shown below. Click on the **Save** button when done.



3.1.1.2 Exploring the data

Graphics are often a useful data-exploring technique through which the researcher may familiarize her- or himself with the data. Relationships and trends may be conveyed in an informal and simplified visual form via graphical displays. SuperMix offers both data-based and model-based graphs. Data-based graphing options are accessed via the **File, Data-based Graphs** option once a SuperMix data file (**.ss3**) is opened, and include **Exploratory**, **Univariate**, **Bivariate** and **Multivariate** graphs as shown on the pop-up menu below. Model-based graphs are available after the analysis has been performed, and will be discussed later in this section.

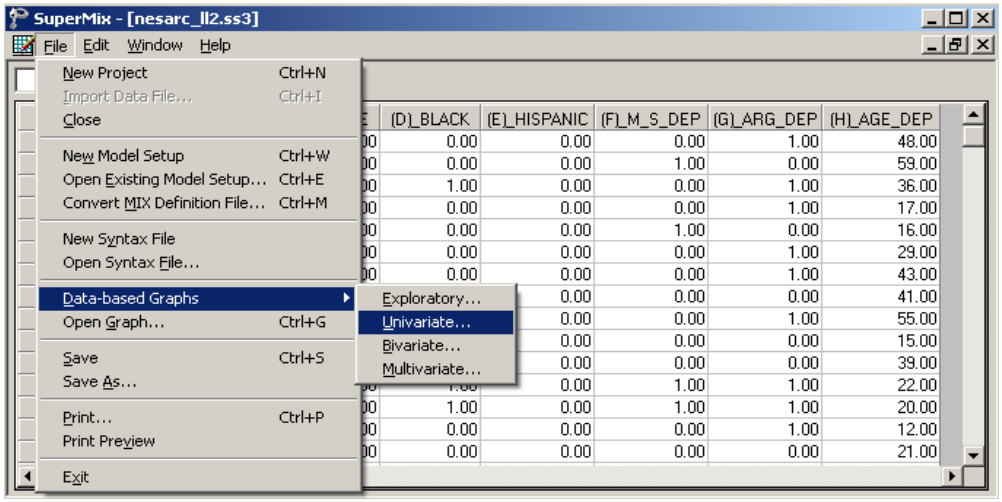
In the case of data-based graphs, we distinguish between three categories: univariate, bivariate, and multivariate graphs. Univariate graphs are particularly useful to obtain an overview of the characteristics of a variable. However, they do not necessarily offer the tools needed to explore longitudinal data as completely as one would wish. For that purpose, bivariate and multivariate data-based graphs are more appropriate.

Univariate graphs

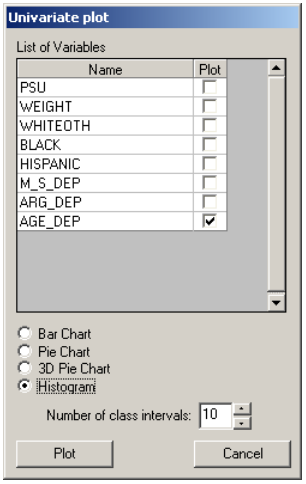
The pop-up menu below shows the data-based graphing options currently available in SuperMix. As a first step, we take a look at the distribution of age at onset of first depression episode (AGE_DEP), which is the potential dependent variable in this study.

Histograms

A histogram represents the frequency of cases per unit interval. It gives a good picture of the distribution of a variable. To create a histogram for AGE_DEP, select the **Univariate** option from the **Data-based Graphs** menu as shown below.



The **Univariate plot** dialog box appears. Select the variable AGE_DEP and indicate that a **Histogram** is to be graphed. The desired number of intervals shown on the histogram is controlled by the **Number of class intervals** field. It is specified as 18 in this case. Click the **Plot** button to display the histogram.



The histogram, as seen below, shows that the distribution of AGE_DEP is nearly symmetrical, and should satisfy the normality assumptions implicit in a multilevel model.

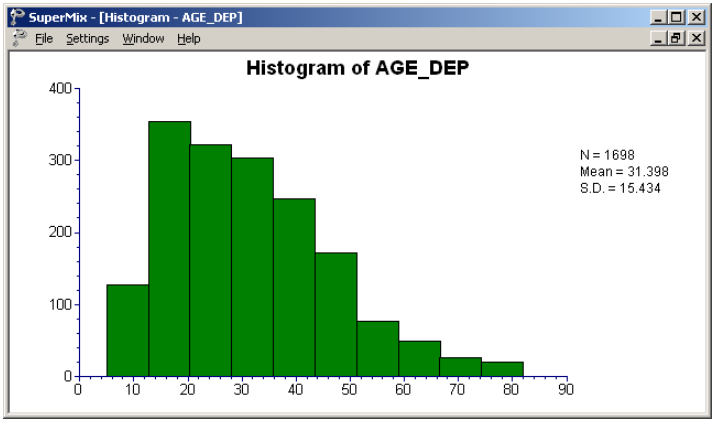


Figure 3.1: Histogram of the variable AGE_DEP

3.1.2 2-level random intercept model with 2 predictors

3.1.2.1 The model

A two-level multilevel model consists of two submodels, one at each level of the hierarchy. A general two-level model for a continuous response variable y depending on a set of p predictors $x_1, x_2, \ldots, x_r$ can be written in the form

$$y_{ij} = \mathbf{x}_{ij}^{'}\boldsymbol{\beta} + \mathbf{z}_{ij}^{'}\mathbf{v}_i + e_{ij}$$

where $i = 1, 2, \ldots, N$ denotes the level-2 units, and $j = 1, 2, \ldots, n_i$ the level-1 units. In this context, y_{ij} represents the response of individual j, nested within level-2 unit i. The model shown here consists of a fixed and a random part. The fixed part of the model is represented by the vector product $\mathbf{x}_{ij}^{'}\boldsymbol{\beta}$, where $\mathbf{x}_{ij}^{'}$ is a typical row of the design matrix of the fixed part of the model with, as elements, a subset of the p predictors. The vector $\boldsymbol{\beta}$ contains the fixed, but unknown parameters to be estimated. $\mathbf{z}_{ij}^{'}\mathbf{v}_i$ and e_{ij} denote the random part of the model at levels 2 and 1 respectively. For example, $\mathbf{z}_{ij}^{'}$ represents a typical row of the design matrix of the random part at level 2, and $\mathbf{v}_i$ the vector of random level-2 effects to be estimated. It is assumed that $v_{01}, v_{02}, \ldots, v_{0N}$ are independently and identically distributed (i.i.d.) with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Phi}_{(v)}$. Similarly, the e_{ij} are assumed i.i.d., with mean 0 and variance σ^2.

The first model fitted to the NESARC data explores the relationship between AGE_DEP and the maternal-side depression and argument involvement, as represented by the variables M_S_DEP and ARG_DEP. The level-1 model is at a patient level, while the level-2 model is at a PSU level. The model can be expressed as

Level-1 model:

$$\text{AGE\_DEP}_{ij} = b_{0i} + b_{1i} \times \left(\text{MS\_DEP}\right)_{ij} + b_{2i} \times \left(\text{ARG\_DEP}\right)_{ij} + e_{ij}$$

Level-2 model:

$$b_{0i} = \beta_0 + v_{0i}$$
$$b_{1i} = \beta_1$$
$$b_{2i} = \beta_2$$

where

$$e_i : N\left(0, \sigma^2 \mathbf{I}_i\right)$$
$$\mathbf{v}_i : N\left(0, \mathbf{\Sigma}_i\right)$$

β_0 denotes the average expected age at onset of the first episode and β_1 denotes the coefficient of the predictor variable M_S_DEP (slope) in the fixed part of the model. Given that the variable M_S_DEP is an indicator variable, β_1 is in effect the expected change in age at onset for patients who reported maternal-side depression. Likewise, β_2 is in effect the expected change in age at onset for patients who reported arguments and stress. The random coefficients v_{i0} and e_{ij} denote the variation in the average expected AGE_DEP value between PSUs and between patients respectively.

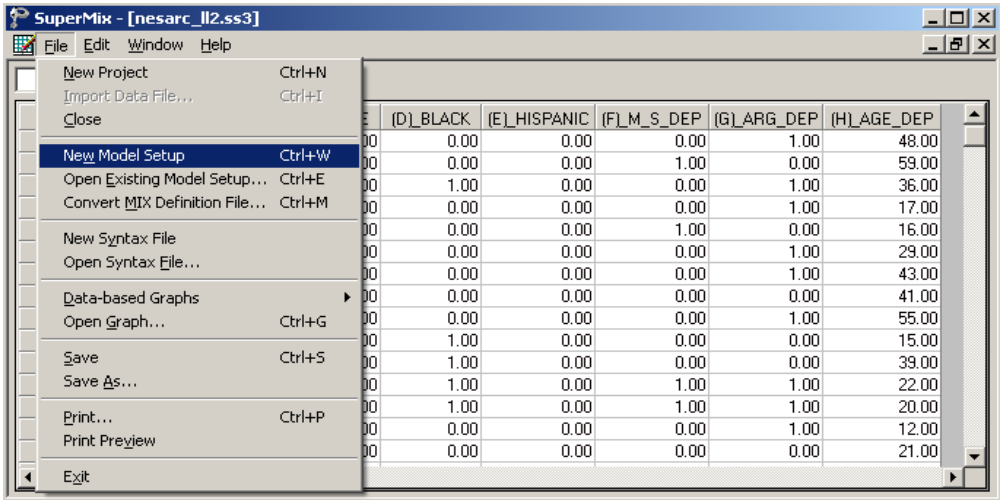
The model can also be written in so-called mixed model notation, as shown below.

$$\text{AGE\_DEP}_{ij} = \beta_0 + \beta_1 * \text{M\_S\_DEP}_{ij} + \beta_2 * \text{ARG\_DEP}_{ij} + v_{i0} + e_{ij}$$

3.1.2.2 Setting up the analysis

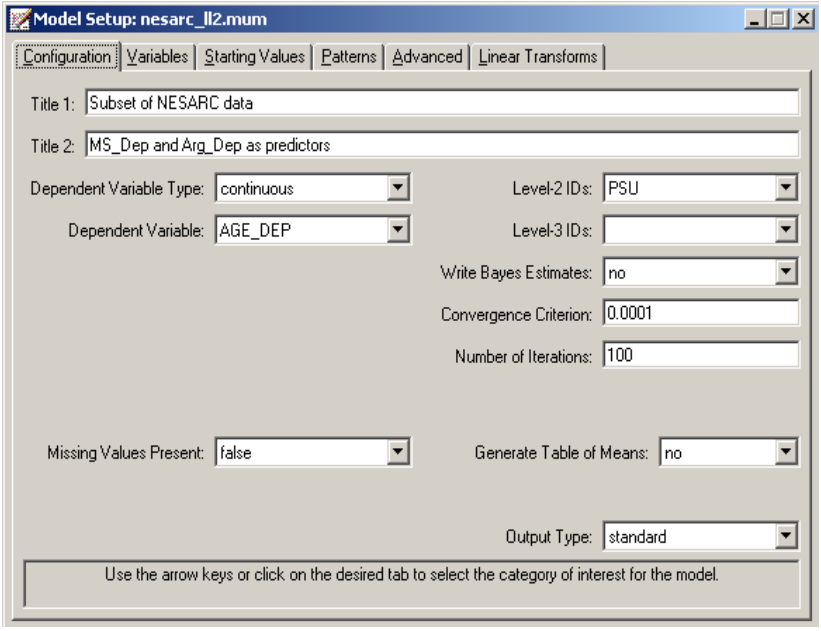
Open the SuperMix spreadsheet **nesarc_ll2.ss3** used during the exploratory analysis discussed previously. The next step is to describe the model to be fitted. We use the SuperMix interface to provide the model specifications. From the main menu bar, select the **File**, **New Model Setup** option.

The **Model Setup** window that appears has six tabs. In this example, only the screens associated with the first two tabs are used. Information entered on these tabs are subsequently saved to a syntax file (*.**mum**) that can be retrieved later as needed.



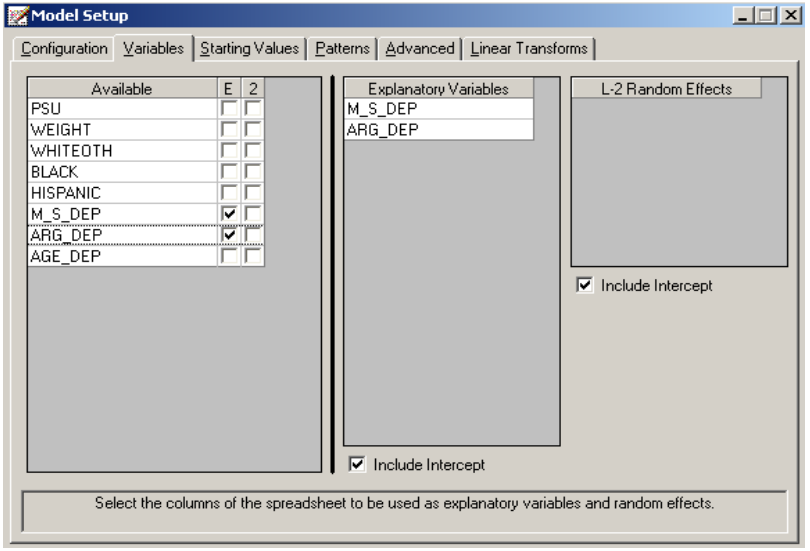
The **Configuration** screen is the first tab on the **Model Setup** window. It enables the user to define the outcome variable, level-2 and level-3 IDs. Some other settings such as missing values, the convergence criterion, the number of iterations, etc. can be specified here. For all the available settings, please refer to Section 2.4. To obtain the model we discussed, proceed as follows:

- Select the continuous outcome variable AGE_DEP from the **Dependent Variable** drop-down list box.
- Select PSU from **Level-2 ID** drop-down list box.
- Enter a title for the analysis in the **Title** text boxes (optional).
- Keep all the other settings on the **Configuration** screen at their default values. Proceed to the **Variables** screen by clicking on that tab.

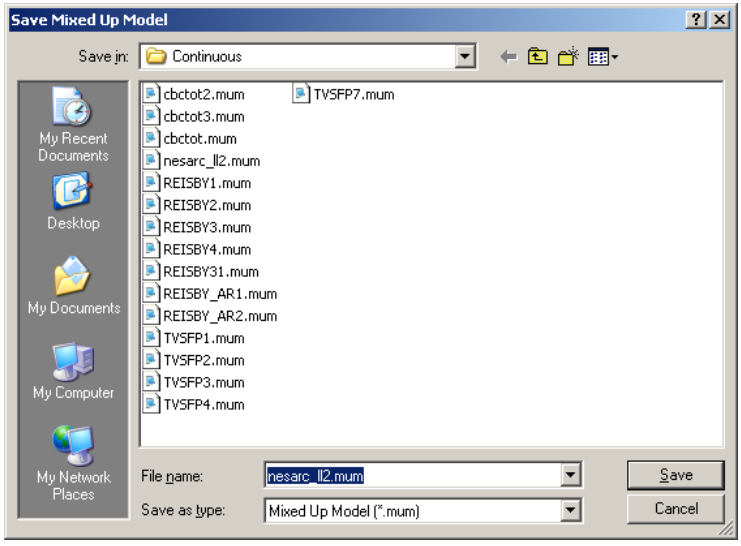


The **Variables** screen is used to specify the fixed and random effects to be included in the model. This screen shows the list of variables available for analysis and next to it two columns, with headings **E** (for explanatory variables) and **2** (for level-2 random effects). Select the explanatory (fixed) variables by checking the **E** check boxes next to the variables M_S_DEP and ARG_DEP in the **Available** grid at the left of the screen. Note that, as the variables are selected, they are listed in the **Explanatory Variables** grid. After selecting all the explanatory variables, the screen shown below is obtained.

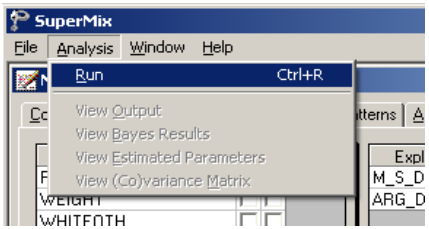
Note that the **Include Intercept** check boxes in the **Explanatory Variables** grid and **L-2 Random Effects** grid are checked by default, indicating that an intercept term will automatically be included in the fixed and random parts of the model.



Before running the analysis, the model specifications have to be saved. Select the **File**, **Save As** option, provide a name (**nesarc_ll2.mum**) for the model specification file, and save.



Run the analysis by selecting the **Run** option from the **Analysis** menu. The standard output file opens. It can also be viewed by selecting the **View Output** option from the same menu.



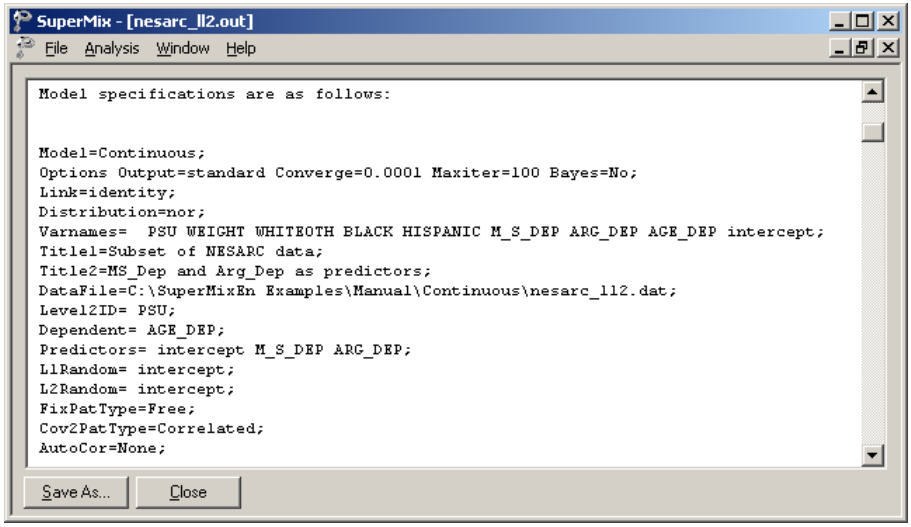
3.1.2.3 Discussion of results

Portions of the output file **nesarc_II2.out** are shown below.

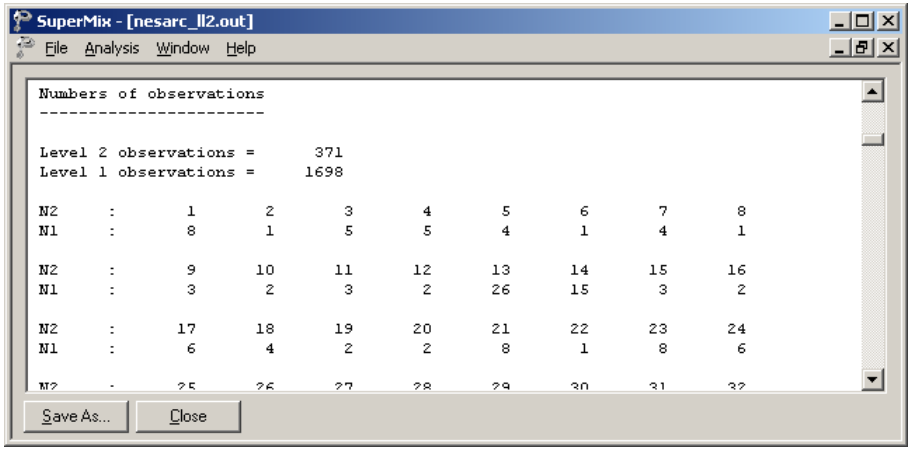
Program information and syntax

At the top of the output file, program information is given. It states the type, date and time of analysis, and provides contact information for technical support.

Program information is followed by model specifications. This section echoes the contents of the syntax file **nesarc_II2.mum**. For more information on syntax and keywords, please see Section 9.7.



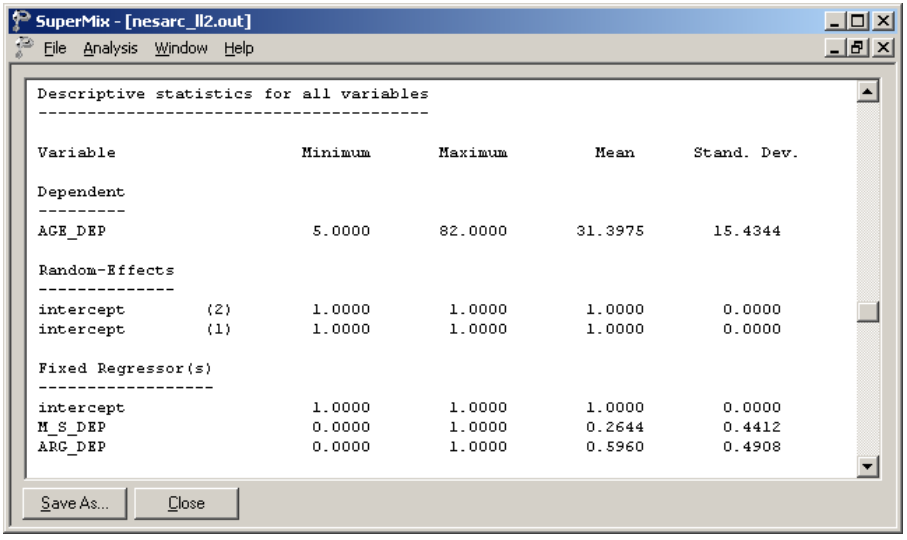
Model and data description



In the next section of the output file as shown above, a description of the hierarchical structure of the data is provided. Data from a total of 371 PSUs and 1,698 respondents were included at levels 2 and 1 of the model. In addition, a summary of the number of respondents nested within each PSU is provided. For example, the PSU with N2:14 had 15 respondents. Note that N2:2 had only 1 observation, which means that the estimation for this PSU might not be reliable.

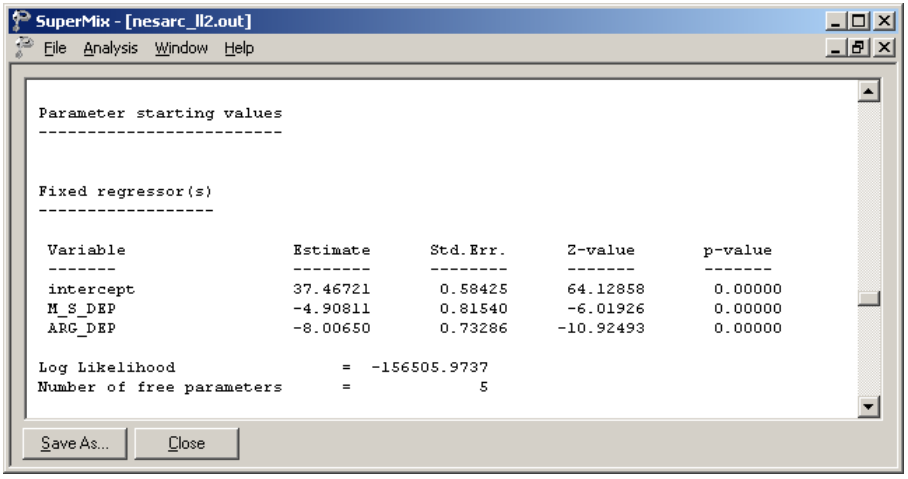
Descriptive statistics and starting values

The data summary is followed by descriptive statistics for all the variables included in the model. We note that the observed average age at the onset of depression is approximately 31 years.

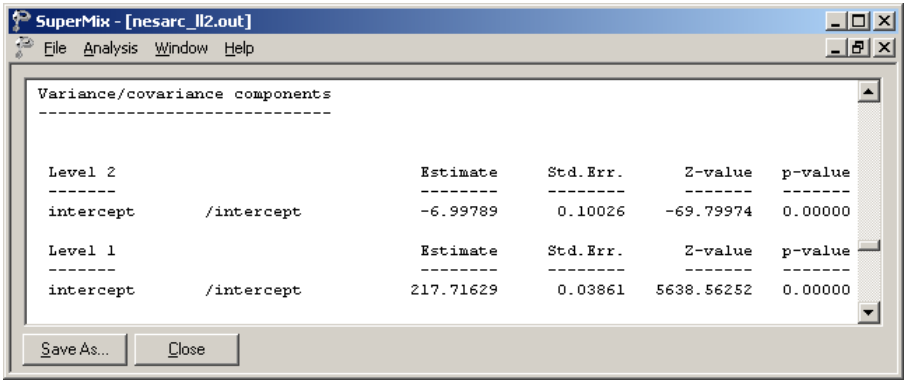


Descriptive statistics are followed by the starting values of the parameters that were used in the initial step of the iterative algorithm. These starting values are obtained by ordinary least squares (OLS) regression, which calculates the estimates by minimizing the sum of the squares of the residuals.

The starting values for the **fixed regressor(s)** are shown below. The **log likelihood** value and **number of free parameters** of the OLS regression are given in this part of the output.

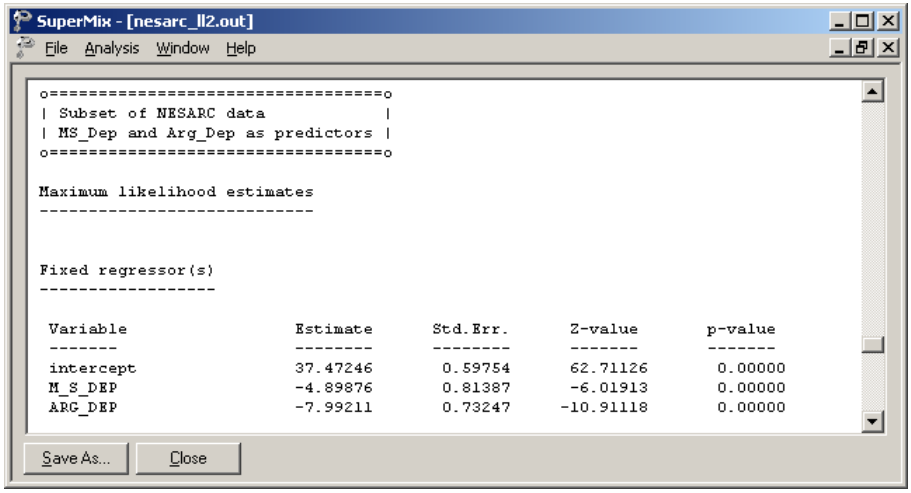


The starting values for the random effects are given next.



Fixed effects results

The output describing the estimated **fixed effects** after convergence is shown next. The estimates are shown in the column with heading Estimate, and correspond to the coefficients β_0, β_1 and β_2 in the model specification. From the z-values and associated exceedance probabilities, we see that all three estimates are highly significant.

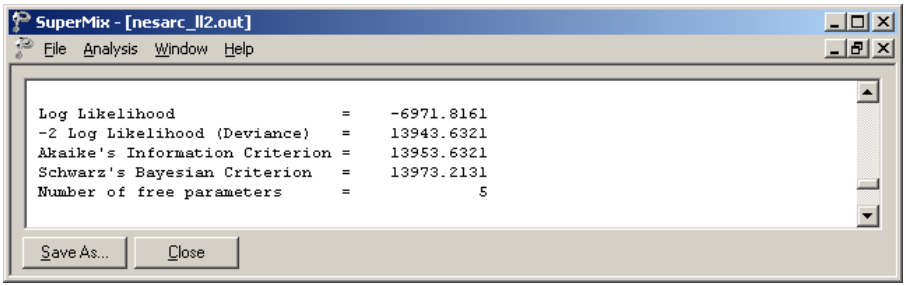


The estimated intercept is 37.472, which means that the average age of the first episode onset of the dysthymia respondents who do not have mother-side depression history and don't argue with others is around 37.4. The estimated coefficients associated with the mother-side history of depression (M_S_DEP) is – 4.898, which indicates that the respondents who have maternal-side depression history tend to get the first episode about five years earlier than those who do not (given the same response on ARG_DEP). The estimate for the indicator of argument involvement (ARG_DEP) shows that a respondent who has argument(s) with others is likely to have a first episode of depression about eight years earlier than a respondent who did not report arguing.

Fit statistics

In addition to the likelihood function value at convergence, a number of related statistical measures for assessing model adequacy are available. The most common of these are the likelihood ratio test and Akaike's and Schwarz's criteria. Both the Akaike information criterion (AIC) and the Schwarz Bayesian criterion (SBC) are functions of the number of estimated parameters, and therefore "penalize" models with large numbers of parameters. In the SuperMix output file, all three of these are

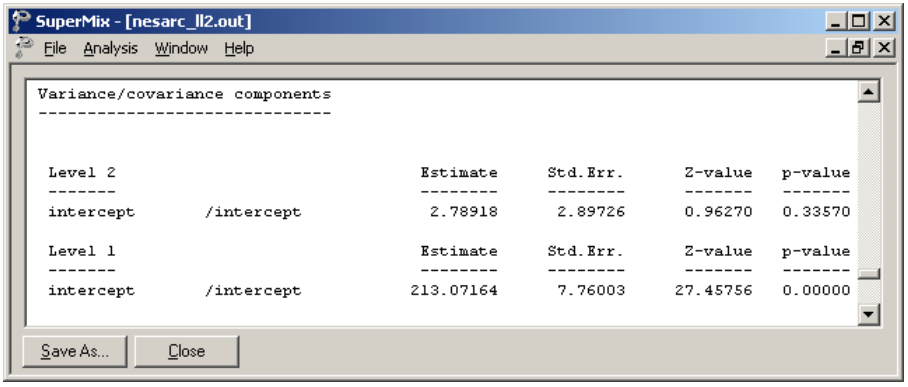
reported. A χ^2 scale factor, with which a χ^2-value obtained from the difference between two deviance statistics should be multiplied to yield a corrected χ^2 statistic in the case of a weighted analysis, may also be found in this section.



- The deviance is defined as $-2\ln L$. For a pair of nested models, the difference in $-2\ln L$ values has a χ^2 distribution, with degrees of freedom equal to the difference in number of parameters estimated in the models compared.
- The AIC was originally proposed for time-series models, but is also used in regression. It is defined as $-2\ln L + 2r$, where r denotes the number of parameters estimated in the model. The model with minimum AIC, in a set of nested models, will be the most parsimonious according to this criterion.
- The SBC is defined as $-2\ln L + r\log n$, where n denotes the number of units at the highest level of the hierarchy. A smaller value of this criterion would indicate the most parsimonious of the models being compared.

Random effects results

The output for the random part of the model follows, and is shown in the image below. In the case of a model with only a random intercept, there are two variances of interest: the variation in the random intercept over the patients, and the residual variation at level 1 over the measurements. There is no significant variation in the average estimated AGE_DEP at level 2 (p = 0.33). This indicates that the expected average age at onset of depression does not differ significantly from PSU to PSU (the level-2 units). Significant differences between the patients (the level-1 units) are reported (p = 0.00).



3.1.2.4 Interpreting the results

Model-based graphs

Activate the **Model Setup** window by clicking on it. Using the **Plot Equations for: AGE_DEP** dialog box that appears when the **File**, **Model-based Graphs**, **Equations** option is selected, we can graphically depict the trend in expected age at onset of depression, taking the values of the predictors M_S_DEP and ARG_DEP into account. The dialog box below shows the selection of the predictor M_S_DEP. Marking of the plots by ARG_DEP is also requested. Two graphs will thus be displayed on the same set of axes: one for each value of the indicator variable ARG_DEP. By default, all

variables present in the model, but not selected for inclusion in the graph, will be assumed to have a value of 0.

The graph below shows the result obtained when the **Plot** button is clicked after completion of the **Plot Equations for: AGE_DEP** dialog box as shown above. We note that patients who did not report arguing are expected to experience onset approximately 8 years later than patients reporting involvement in arguments.

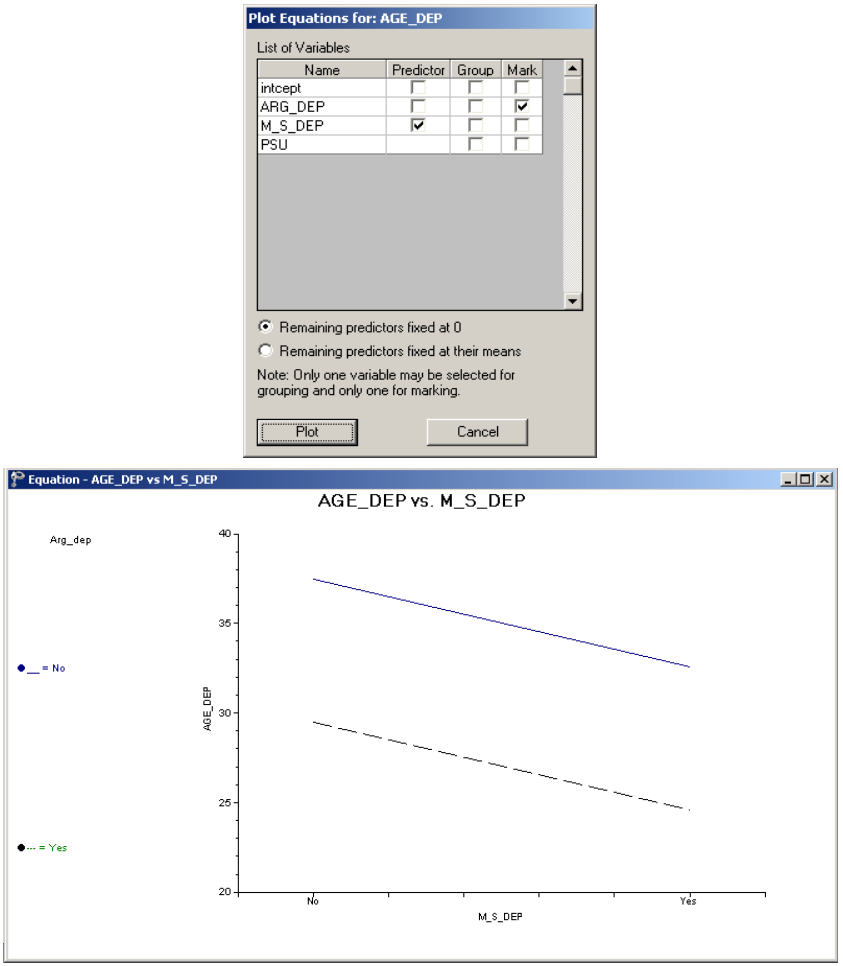


Figure 3.2: Plot of AGE_DEP versus M_S_DEP for 2 groups

A similar plot for the predictor ARG_DEP is given next. This graph was obtained by swapping the positions of the M_S_DEP and ARG_DEP variables on the **Plot Equations for: AGE_DEP** dialog box. Note that patients with maternal-side depression had their first episode approximately 5 years earlier than patients with no history of maternal-side depression. The two graphs shown represent the graphic interpretation of the fixed effect estimates shown previously.

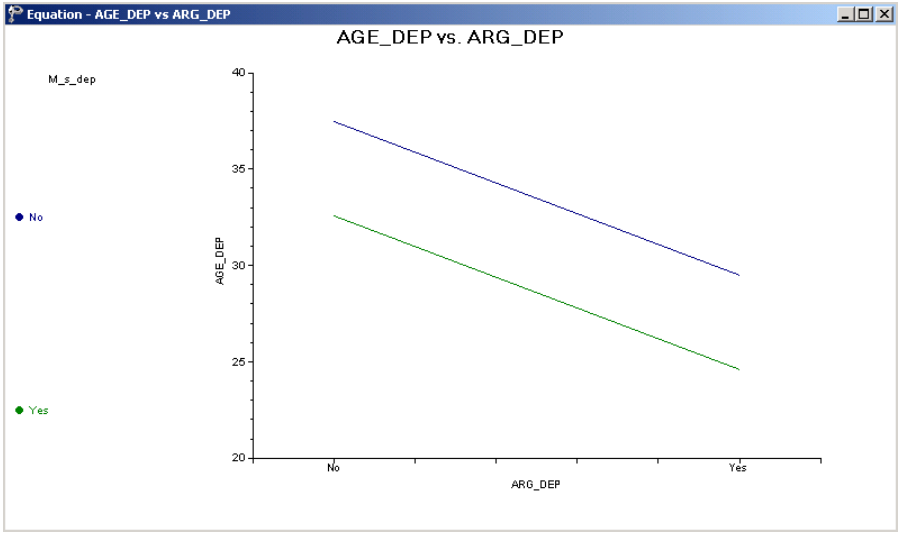


Figure 3.3: Plot of AGE_DEP versus M_S_DEP for 2 groups

ICCs and % variance explained

By calculating the total variation in the age at onset as explained by the current model, we can obtain an estimate of the intracluster correlation coefficient. We first need to calculate the total variation in the outcome variable, which for this model is defined as $\hat{\text{var}}(e_{ij}) + \hat{\text{var}}(v_{i0})$.

The intracluster coefficient is then defined as

$$ICC = \frac{\hat{\text{var}}(v_{i0})}{\hat{\text{var}}(e_{ij}) + \hat{\text{var}}(v_{i0})}$$

and represents the proportion of variation in age at onset that is between the groups (PSUs). An estimate of the percentage of variation in the outcome at a PSU level is obtained as

$$\frac{2.78918}{2.78918 + 213.07164} \times 100\% = 1.29\%$$

indicating that only 1.29% of the total variance is explained at PSU level; the rest of the variance remains at the respondent level.

3.1.3 A 2-level random intercept model with 4 predictors

3.1.3.1 The model

In the previous section, we modeled the outcome variable AGE_DEP as a function of M_S_DEP and ARG_DEP. The extended model discussed in this section takes the ethnicity of a respondent into consideration. The model fitted is expressed as follows:

$$\begin{aligned}\text{AGE\_DEP}_{ij} = {} & \beta_0 + \beta_1 * \text{BLACK}_{ij} + \beta_2 * \text{HISPANIC}_{ij} \\ & + \beta_3 * \text{M\_S\_DEP}_{ij} + \beta_4 * \text{ARG\_DEP}_{ij} + v_{i0} + e_{ij}.\end{aligned}$$

As before, β_0 denotes the average expected age at the onset of first episode, $\beta_1, \beta_2, \ldots, \beta_4$ indicate the estimated coefficients associated with the fixed part of the model, and v_{i0} and e_{ij} represent the random part of the model.

Recall from Section 3.1 that ethnicity was represented by 3 indicator variables, namely WHITEOTH, BLACK and HISPANIC. In the model formulated above, only two of these variables have been included. This was done since the inclusion of all three

indicators and the intercept term in the model would cause collinearity between the fixed effects. Any of the respondents will have a value of "1" on one of the three ethnicity indicators. If the values of the indicators are added together in a column-wise fashion, a column of 1s will result. The intercept variable is represented by just such a column of 1s in the program. If a linear combination of a subset of the columns of the design matrix is a constant multiple of another column, a condition referred to as multicollinearity is present and the model cannot be estimated properly.

Consider an example where three respondents, one from each of the three ethnic groups, are considered:

Patient	WHITEOTH	BLACK	HISPANIC	Sum of Ethnicity var.	Intercept
1	1	0	0	1	1
2	0	1	0	1	1
3	0	0	1	1	1

There are two ways in which the model can be formulated to avoid running into this problem. The first is to exclude the intercept and use only the three ethnicity indicators. Such a model, as shown below,

$$\begin{aligned}\text{AGE\_DEP}_{ij} = \beta_0 * \text{WHITEOTH}_{ij} + \beta_1 * \text{BLACK}_{ij} + \beta_2 * \text{HISPANIC}_{ij} \\ + \beta_3 * \text{M\_S\_DEP}_{ij} + \beta_4 * \text{ARG\_DEP}_{ij} + v_{i0} + e_{ij}\end{aligned}$$

would not offer an estimated coefficient of the average age at onset. Instead, the expected average age at onset for each of the three ethnic groups may be deduced from the estimated coefficients for WHITEOTH, BLACK and HISPANIC.

Alternatively, one can drop one of the ethnicity indicators from the model while retaining the intercept coefficient. This is what we have opted to do in the current example:

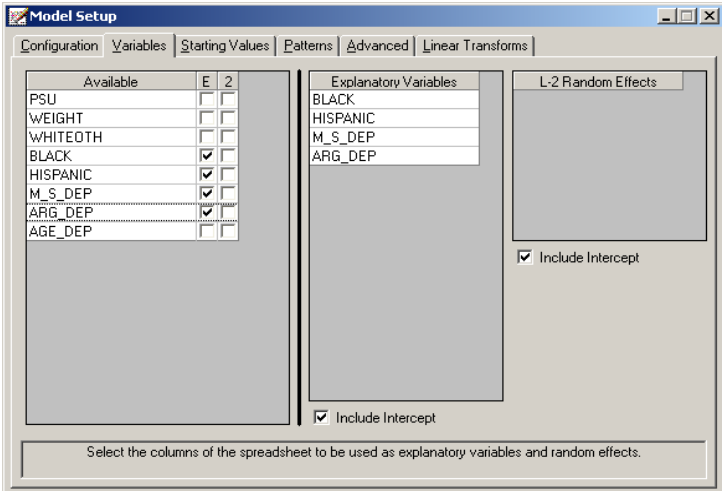
$$\begin{aligned}\text{AGE\_DEP}_{ij} = \beta_0 + \beta_1 * \text{BLACK}_{ij} + \beta_2 * \text{HISPANIC}_{ij} \\ + \beta_3 * \text{M\_S\_DEP}_{ij} + \beta_4 * \text{ARG\_DEP}_{ij} + v_{i0} + e_{ij}\end{aligned}$$

In the case of this formulation, the intercept coefficient represents the expected average age at onset for a patient with a value of zero on all the predictors. But if the indicators BLACK and HISPANIC assume a value of 0, it implies that the remaining ethnicity variable WHITEOTH must have a value of 1. As a result, the interpretation of the intercept coefficient would be the expected average onset age for a patient who is white or from some other ethnic origin (excluding African American and Hispanic). This ethnic group thus becomes the reference group in the current analysis. Any of the ethnic groups can be used as the reference group by simply adjusting the coding of the indicator variables; the only proviso being that the group of interest have sufficient data to serve as stable reference group.

3.1.3.2 Setting up the analysis

The SuperMix spreadsheet **nesarc_ll2.ss3** and the model specification file **nesarc_ll2.mum** discussed in the previous example are used a point of departure.

With the model specification file open, click on the **Variables** tab of the **Model Setup** window. Add the predictors BLACK and HISPANIC to the model by checking the boxes next to these variables in the **E** column, as shown below.

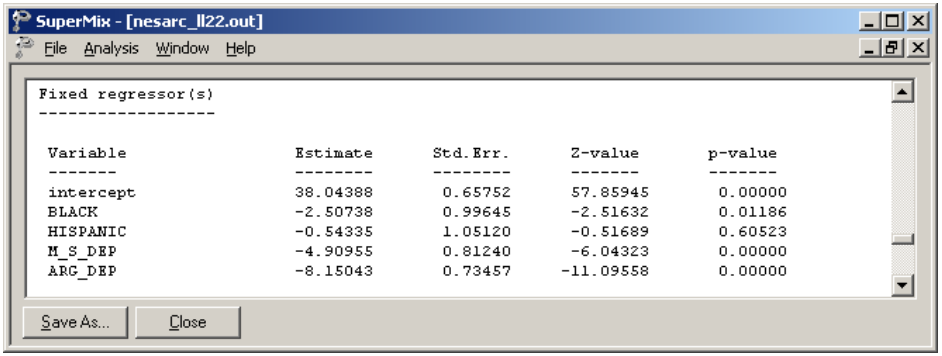


Save the modified model as **nesarc_II22.mum** specification file, and select the **Run** option from the **Analysis** menu to perform the analysis.

3.1.3.3 Discussion of results

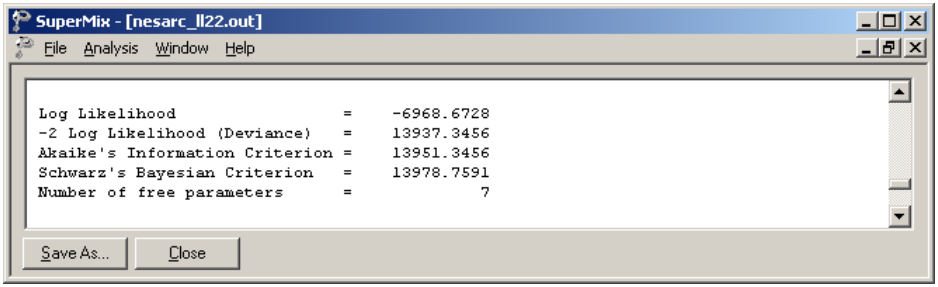
Fixed effects results

The maximum likelihood estimates of the coefficients in the fixed part of the model are shown below. Statistically the estimate for HISPANIC is not significant (p =0.61). Both estimates for BLACK and HISPANIC are negative, which indicates that African American and Hispanic respondents tend to have an earlier onset of the first episode when compare with patients from white and other ethnic groups.



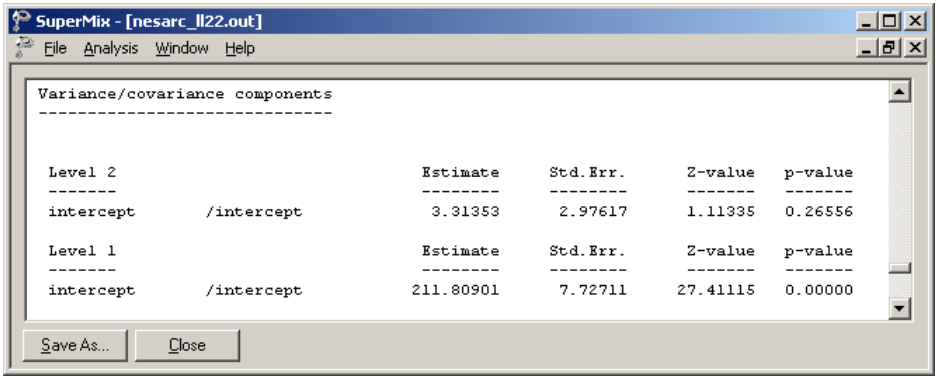
Fit statistics

Fit statistics for the current model are reported as shown below.



Random effects results

The output for the **random part** of the model is given next.



The random intercept effect at level 2 is not significant. As before, most of the variation in scores is found at a respondent level, with only about 2% of the variation remaining at the PSU level.

3.1.3.4 Interpreting the results

Estimated outcomes for different groups

The estimated outcome for any patient can be obtained using the formula

$$\widehat{\text{AGE\_DEP}}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 * \text{BLACK}_{ij} + \hat{\beta}_2 * \text{HISPANIC}_{ij} + \hat{\beta}_3 * \text{M\_S\_DEP}_{ij} + \hat{\beta}_4 * \text{ARG\_DEP}_{ij}$$

For a white respondent, the expected AGE_DEP can be calculated as

$$\begin{aligned}\widehat{\text{AGE\_DEP}}_{ij} &= \hat{\beta}_0 + \hat{\beta}_3 * \text{M\_S\_DEP}_{ij} + \hat{\beta}_4 * \text{ARG\_DEP}_{ij} \\ &= 38.04388 - 4.90955 \times \text{M\_S\_DEP}_{ij} - 8.15043 \times \text{ARG\_DEP}_{ij}.\end{aligned}$$

For African American respondents BLACK = 1, and thus the formula used to predict their AGE_DEP scores reduces to

$$\begin{aligned}\widehat{\text{AGE\_DEP}}_{ij} &= \hat{\beta}_0 + \hat{\beta}_1 * \text{BLACK}_{ij} + \hat{\beta}_3 * \text{M\_S\_DEP}_{ij} + \hat{\beta}_4 * \text{ARG\_DEP}_{ij} \\ &= 38.04388 - 2.50738 \times 1 - 4.90955 \times \text{M\_S\_DEP}_{ij} - 8.15043 \times \text{ARG\_DEP}_{ij}.\end{aligned}$$

The formula for a patient of Hispanic origin can be derived in a similar way. In 3.1, the same expected ages of the first episode onset for different groups are calculated based on the formulas above.

Table 3.1: Expected AGE_DEP for various groups of patients

Origin	M_S_DEP = No ARG_DEP = No	M_S_DEP = Yes ARG_DEP = No	M_S_DEP = No ARG_DEP = Yes	M_S_DEP = Yes ARG_DEP = Yes
White & Other	38.04	33.13	29.89	24.98
African American	35.54	30.63	27.39	22.48
Hispanic	37.50	32.59	29.35	24.44

The results show that the respondent who has a history of maternal-side depression or gets involved into arguments generally has an earlier onset age for the first episode. For the respondents with the same M_S_DEP and ARG_DEP values, the average first episode onset ages of African American respondents are the lowest. We also conclude that a patient involved in arguments (ARG_DEP = 1) is likely to

have an earlier onset age of depression than a patient with maternal-side depression only (M_S_DEP = 1).

Fit statistics and % variation explained

Table 3.2 shows the fit indices for the previous and current models.

TABLE 3.2: Comparison of random intercept models for NESARC data

Fit indices	Model with 2 indicators	Model with 4 indicators	Difference
Log Likelihood	–6971.8161	–6968.6728	
–2 Log Likelihood (Deviance)	13943.6321	13937.3456	6.2865
Akaike's Information Criterion	13953.6321	13951.3456	2.2865
Schwarz's Bayesian Criterion	13973.2131	13978.7591	–5.5460
Number of free parameters	5	7	

The difference in deviances can be used to assess the model fit. This method is valid for nested models. A nested model may be defined as any submodel of a given model that is based on the same number of observations. Given the difference in structure between the 2-level models these models cannot, however, be compared to each other.

The difference in the deviances follows a χ^2 distribution, where the degree of freedom is the difference of numbers of free parameters.

$$\left(-2\ln_{model1}\right)-\left(-2\ln_{model2}\right)\sim\chi^2\left(d.f.\left(-2\ln_{model2}\right)-\left(-2\ln_{model1}\right)\right)$$

When the deviances of the two models are compared, a χ^2-statistic of 13943.6321 – 13937.3456 = 6.2865 with 7 – 5 = 2 degrees of freedom is obtained. This indicates that the current model fits the data better. The AIC decreased from 13953.6321 to 13951.3456, and also favors the use of the 4-predictor model. The SBC, however, increased slightly, from 13973.2131 to 13978.7591, and thus favors the model previously fitted as the more parsimonious. The definitions of these indices are

given in the discussion of the output of the previous model. Note, however, that the changes in all three criteria are rather small.

The estimated percentages of variation in outcome at respondent level can be calculated using the variance components reported in the random effects part of the output file:

$$\frac{211.80901}{211.80901 + 3.31353} \times 100\% = 98.46\%.$$

Once the additional level-1 predictors are taken into account, there does not seem to be significant random variation in the outcome over the intercepts of the level-2 units. The estimated average onset age of the first episode does not vary significantly from PSU to PSU.

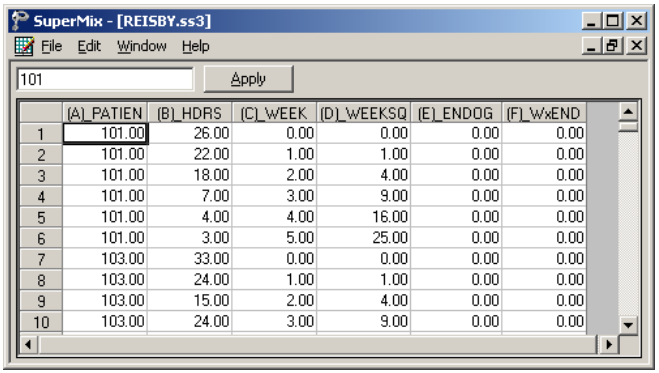
3.2 Models based on the Reisby data

3.2.1 The data

The data set is from a study described in Reisby *et. al.* (1977) that focused on the longitudinal relationship between imipramine (IMI) and desipramine (DMI) plasma levels and clinical response in 66 depressed inpatients (37 endogenous and 29 non-endogenous). Following a placebo period of 1 week, patients received 225 mg/day doses of imipramine for four weeks. In this study, subjects were rated with the Hamilton depression rating scale (HDRS) twice during the baseline placebo week (at the start and end of this week) as well as at the end of each of the four treatment weeks of the study. Plasma level measurements of both IMI and its metabolite DMI were made at the end of each week. The sex and age of each patient were recorded and a diagnosis of endogenous or non-endogenous depression was made for each patient.

Although the total number of subjects in this study was 66, the number of subjects with all measures at each of the weeks fluctuated: 61 at week 0 (start of placebo week), 63 at week 1 (end of placebo week), 65 at week 2 (end of first drug treatment week), 65 at week 3 (end of second drug treatment week), 63 at week 4 (end of third

drug treatment week), and 58 at week 5 (end of fourth drug treatment week). The sample size is 375. Data for the first 10 observations of all the variables used in this section are shown below in the form of a SuperMix spreadsheet file, named **reisby.ss3**.



	(A)_PATIEN	(B)_HDRS	(C)_WEEK	(D)_WEEKSQ	(E)_ENDOG	(F)_WxEND
1	101.00	26.00	0.00	0.00	0.00	0.00
2	101.00	22.00	1.00	1.00	0.00	0.00
3	101.00	18.00	2.00	4.00	0.00	0.00
4	101.00	7.00	3.00	9.00	0.00	0.00
5	101.00	4.00	4.00	16.00	0.00	0.00
6	101.00	3.00	5.00	25.00	0.00	0.00
7	103.00	33.00	0.00	0.00	0.00	0.00
8	103.00	24.00	1.00	1.00	0.00	0.00
9	103.00	15.00	2.00	4.00	0.00	0.00
10	103.00	24.00	3.00	9.00	0.00	0.00

The variables of interest are:

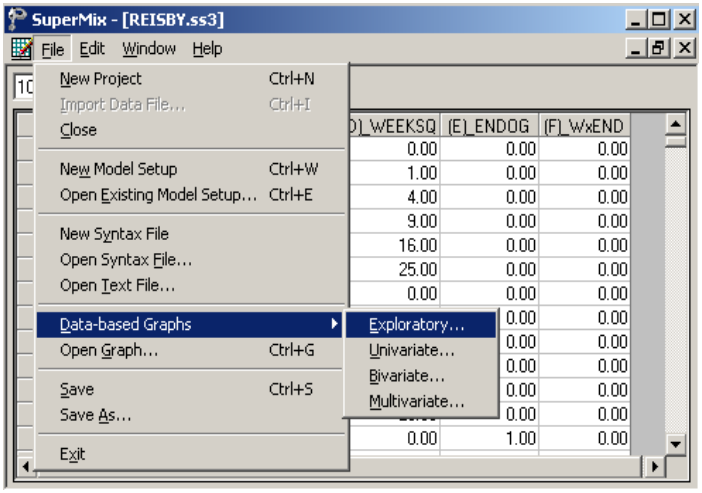
- Patient is the patient ID (66 patients in total).
- HDRS is the Hamilton depression rating scale.
- WEEK represents the week (0, 1, 2, 3, 4 or 5) at which a measurement was made.
- WEEKSQ represents the squared values of WEEK. The creation of this variable is illustrated in Section 3.2.1.1.
- ENDOG is a dummy variable for the type of depression a patient was diagnosed with (1 for endogenous depression and 0 for non-endogenous depression).
- WxENDOG represents the interaction between WEEK and ENDOG, and is the product of WEEK and ENDOG.

3.2.1.1 Exploring the data

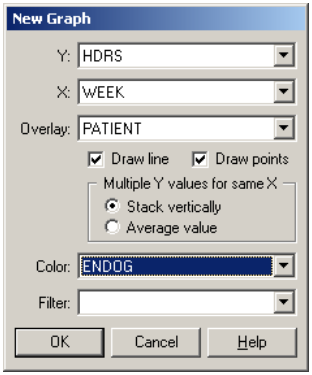
Graphing the observed data

In the previous example, we have shown a number of data-based graphs. Here, we use the **Exploratory** option of the **Data-Based Graphs** menu to explore the data in the **reisby.ss3** spreadsheet, stored in the **Continuous** subfolder.

Start by opening the data file in the SuperMix spreadsheet. Then select the **Data-based Graphs**, **Exploratory** option on the **File** menu as shown below to activate the **New Graph** dialog box.



Specify HDRS as the dependent (vertical axis) variable by selecting it from the **Y** drop-down list box and WEEK as the independent (horizontal axis) variable by selecting it from the **X** drop-down list box. A graph on the same axis system is created for each patient by selecting the variable Patient from the **Overlay** drop-down list box. Furthermore, each graph is assigned a color by selecting ENDOG from the **Color** drop-down list box to produce the following **New Graph** dialog box.



Click on the **OK** button to produce the following graph of the reaction trajectories over time for the 66 inpatients.

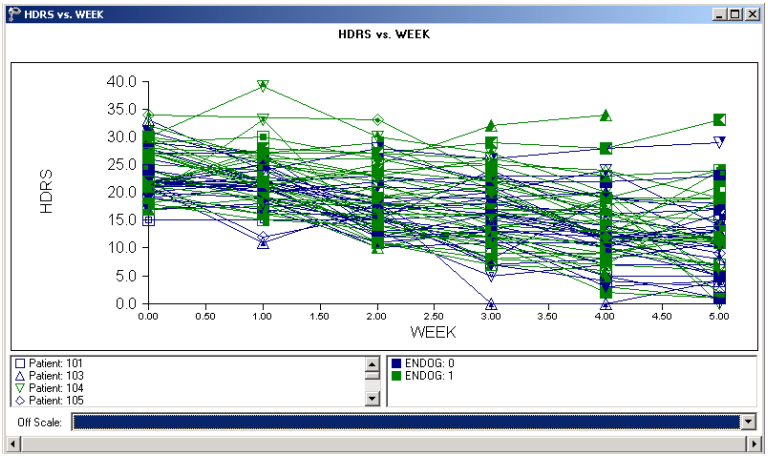
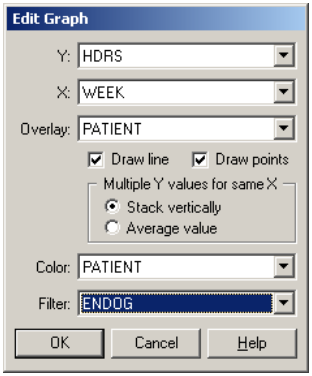


Figure 3.4: Reaction trajectories over time for 66 patients

To modify the existing graphic display, select the **Edit Graph** option from the **Settings** menu to load the **Edit Graph** dialog box. To obtain different graphs for the two categories of the covariate ENDOG, select it from the **Filter** drop-down list box to produce the following **Edit Graph** dialog box.



Click on the **OK** button to open the following graphics window.

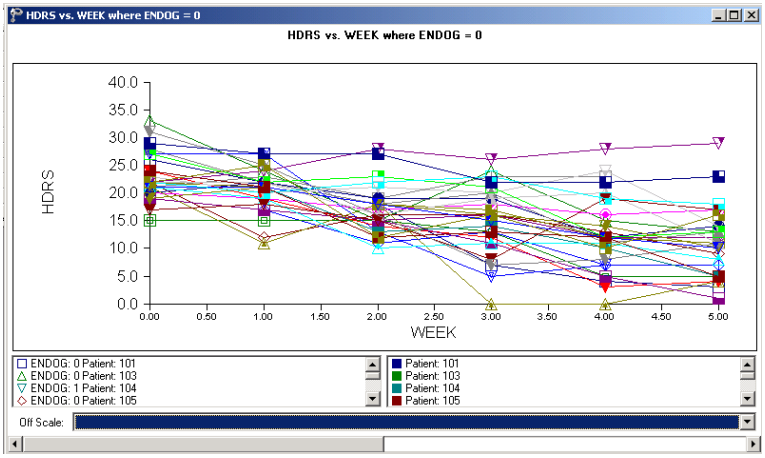


Figure 3.5: Reaction trajectories over time for patients with ENDOG=0

At the bottom of the graphics window is a "slider" with left and right arrows. By clicking on the right arrow, one can obtain the next graphic shown below and by clicking on the left arrow, the graphic above.

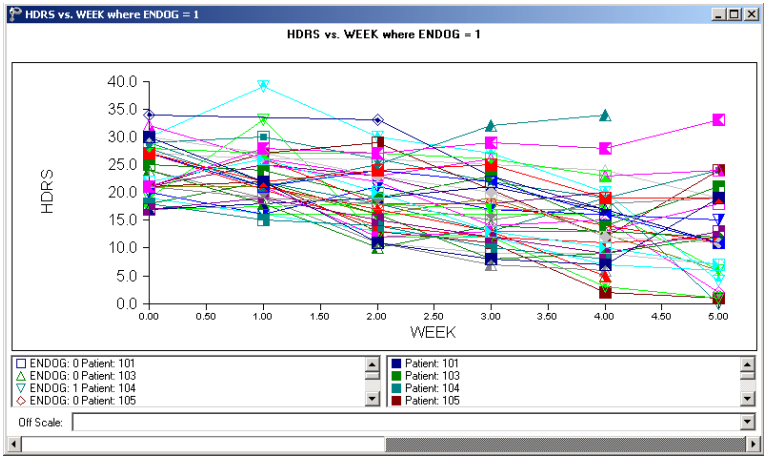


Figure 3.6: Reaction trajectories over time for patients with ENDOG=1

The above graphs show a general, approximately linear decline over time and an increase in the variability of the HDRS scores across time for both types of depression.

3.2.2 A 2-level random intercept-and-slope model

From the graphical display obtained in the previous section, it seems as if the HDRS scores follow an approximately linear trend over time, decreasing over the course of the study. It is also apparent, however, that patients not only start out at different levels but also have differences in the slopes of the HDRS against WEEK lines. In this section, we explore a model that allows patients not only to have unique intercepts, but also unique slopes across time. In other words, we allow both intercept and WEEK (slope) to vary randomly over patients. The image below demonstrates the meaning of the random slope and random intercept in a hypothetical 2-level model.

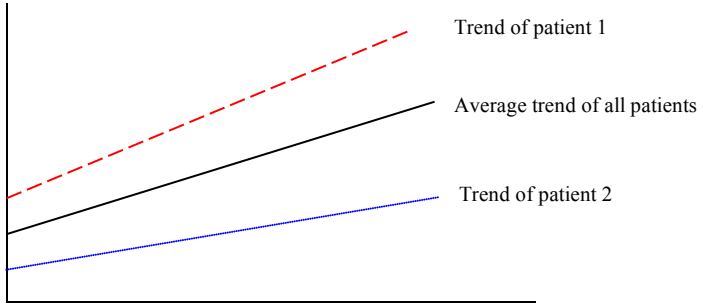


Figure 3.7: Score trends for individual patients

3.2.2.1 The model

The random intercept-and-slope model for the response variable HDRS may be expressed as

$$\text{HDRS}_{ij} = \beta_0 + \beta_1 \times (\text{WEEK})_{ij} + v_{i0} + v_{i1}(\text{WEEK})_{ij} + e_{ij}$$

We can rewrite the model in the following way.

Level-1 model:

$$\text{HDRS}_{ij} = b_{0i} + b_{1i} \times (\text{WEEK})_{ij} + e_{ij}$$

Level-2 model:

$$b_{0i} = \beta_0 + v_{0i}$$
$$b_{1i} = \beta_1 + v_{1i}$$

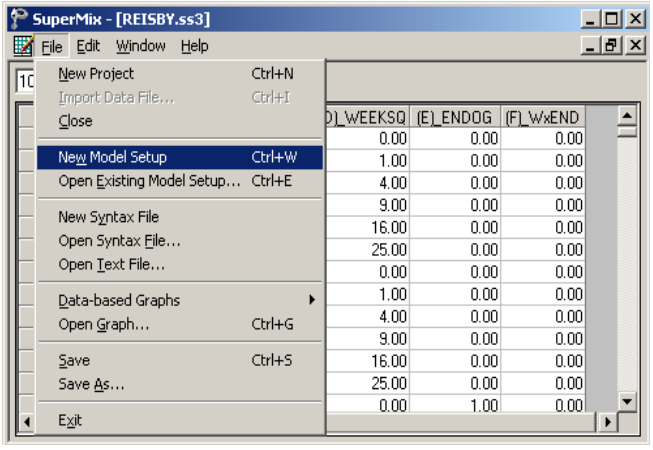
where

$$e_i : N(0, \sigma^2 \mathbf{I}_i)$$
$$\mathbf{v}_i : N(\mathbf{0}, \Phi_{(v)})$$

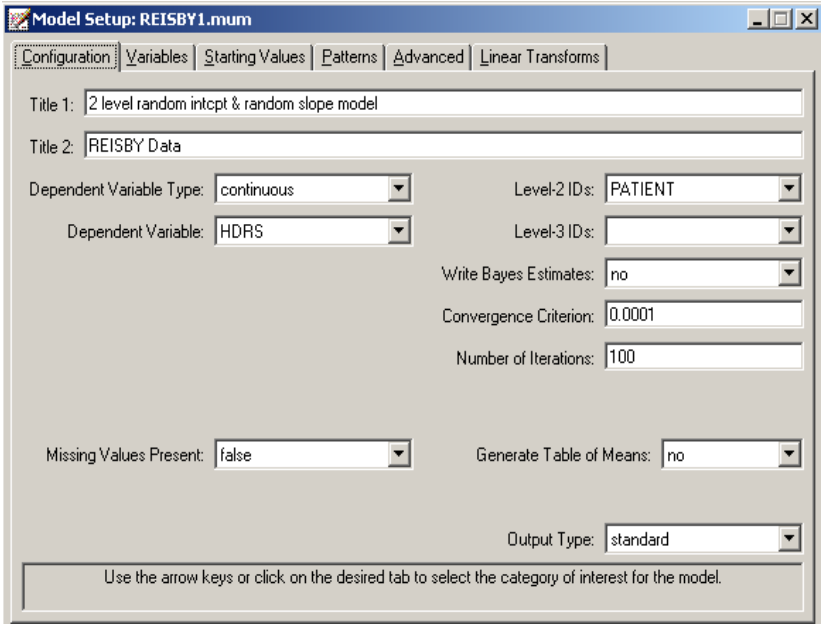
β_0 denotes the average expected depression rating scale value, β_1 denotes the coefficient of the predictor variable WEEK (slope) in the fixed part of the model, v_{1i} denotes the variation in the slopes over patients, and v_{0i} and e_{ij} denote the variation in the average expected HDRS value over patients and between patients respectively. Furthermore, i = 1,2,…,66 refers to the 66 patients; j = 1, 2, …, n_i refers to the j^{th} observation for patient i. The maximum value for n_i is 6.

3.2.2.2 Setting up the analysis

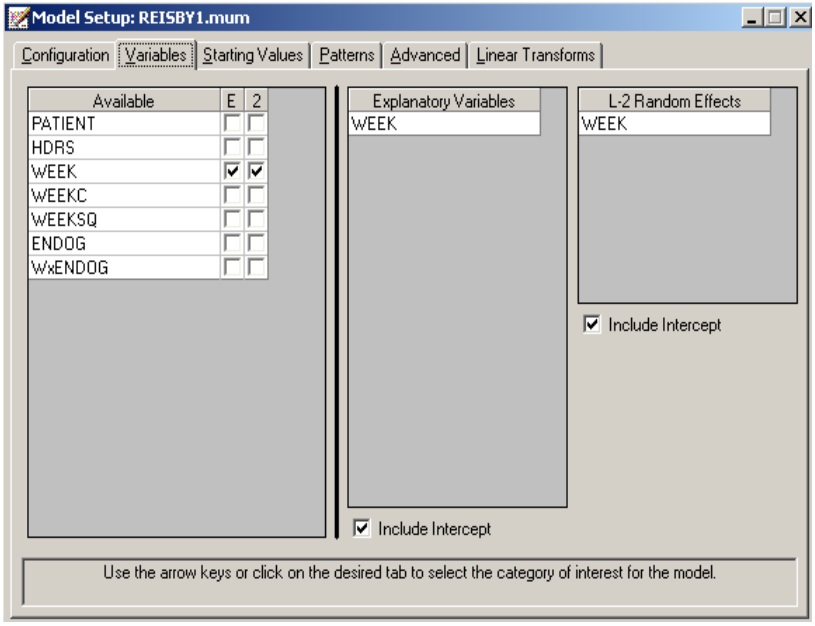
Start by opening the **reisby.ss3** file as a SuperMix spreadsheet. Next, select the **New Model Setup** option on the **File** menu as shown below to load the **Model Setup** window.



Starting with the **Configuration** screen, enter the (optional) title in the **Title 1** and **Title 2** text boxes respectively. The continuous outcome variable HDRS is selected from the **Dependent Variable** drop-down list box. The variable Patient, which defines the levels of the hierarchy, is selected as the Level-2 ID from the **Level-2 IDs** drop-down list box to produce the following **Configuration** screen.



Click the **Variables** tab to proceed to the **Variables** screen of the **Model Setup** window. The variable Week is specified as the covariate of the fixed part of the model by checking the **E** check box for WEEK in the **Available** grid. Mark the **2** check box for Week in the **Available** grid to specify the random slope at level 2 of the model. After completion, the **Variables** screen should look as shown below.

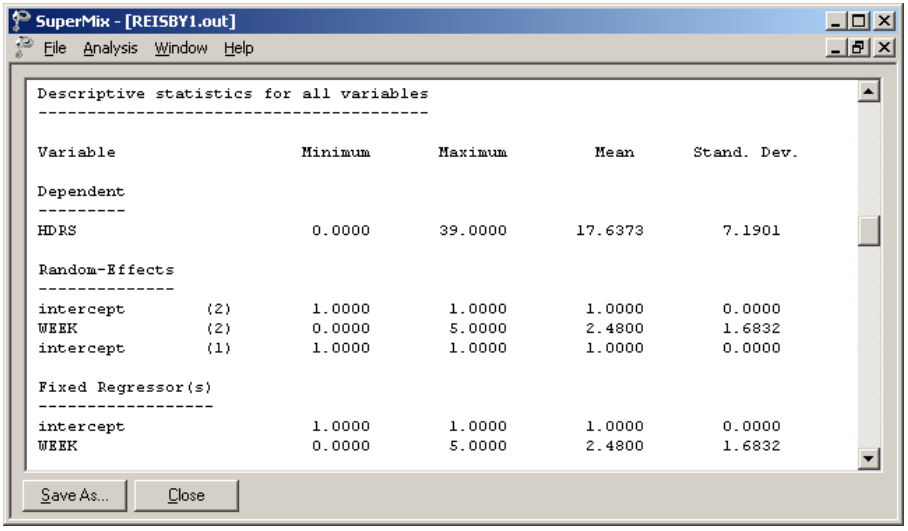


Before the analysis can be run, save the model specifications to **reisby1.mum**. Run the model to produce the output file **reisby1.out**.

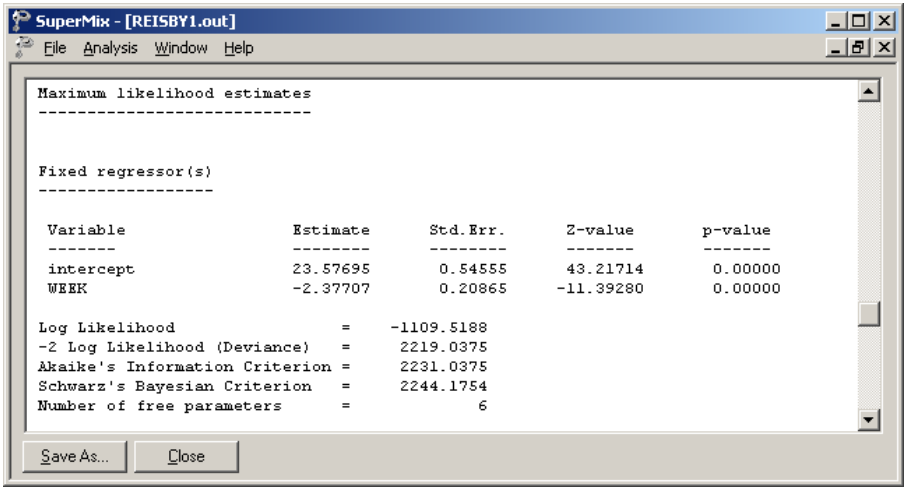
3.2.2.3 Discussion of results

Descriptive statistics

The section of the output file shown below contains the descriptive statistics for all variables in the current model specification. If all patients' data were complete, the average for the time variable WEEK would have been exactly 2.5; the value of 2.48 indicates that the number of patients with information at each time point fluctuates somewhat.

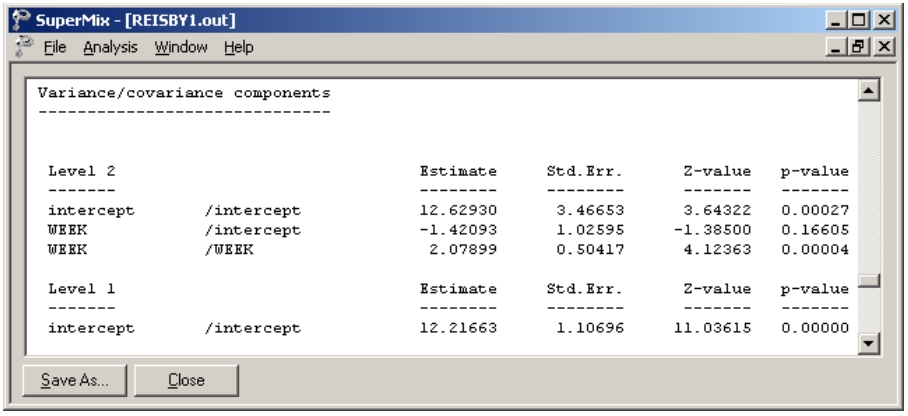


3.2.2.4 Interpreting the results



The summary of the hierarchical structure of the data shows how the 375 measurements are nested within the 66 patients. It also indicates that the number of repeated measurements per patient varies from 4 to 6 observations. The convergence

is attained in 5 iterations. The output file contains the final estimates of the fixed and random coefficients included in the model, along with some goodness of fit measures as shown .



Fixed effects results

The results show a highly significant coefficient ($p < 0.00001$) for the time effect, as represented by the variable WEEK. At the beginning of the study, when WEEK = 0, the average expected HDRS score is 23.57695. For each subsequent week, a decrease of 2.37707 in average HDRS score is expected. At the end of the study period, the average expected HDRS score is 23.57695 – 5(2.37707) = 11.6916.

Random effects results

With the exception of the WEEK-intcept covariance, all variance components are highly significant, as shown in the *p*-value column. From the output above we have $\hat{\text{var}}(v_{i0}) = 12.62930$, $\hat{\text{var}}(v_{i1}) = 2.07899$, $\hat{\text{cov}}(v_{i0}, v_{i1}) = -1.42093$, and $\hat{\text{var}}(e_{ij}) = 12.21663$. Typically, one would expect most of the variation in HDRS scores at the measurement level, and thus would expect $\hat{\text{var}}(e_{ij})$ to be larger than any of the other variances/covariances. With these data, however, there is more variation in the random intercepts over patients than in the measurements nested within patients. Due to this, it may be of interest to take a closer look at the variation in HDRS scores at the two levels of the hierarchy.

Fit statistics and ICC

In the case of a model with only a random intercept, there are two variances of interest: the variation in the random intercept over the patients (the level-2 units), and the residual variation at level 1, over the measurements. By calculating the total variation in the HDRS score explained by such a model, obtained as $\hat{\text{var}}(e_{ij}) + \hat{\text{var}}(v_{i0})$, we can obtain an estimate of the intracluster correlation coefficient.

The intracluster coefficient is defined as

$$ICC = \frac{\hat{\text{var}}(v_{i0})}{\hat{\text{var}}(e_{ij}) + \hat{\text{var}}(v_{i0})}$$

and would, for a random intercept model for this data, represent the proportion of variation in HDRS scores between patients. The term intracluster correlation coefficient applies to random intercept models only; in more complicated models the focus is on explanation of variation in various coefficients.

In the current model, the situation is somewhat more complicated due to the inclusion of both random intercept and random slope. This implies a possible correlation between the level-2 random effects. When calculating an estimate of the total variation, the covariance(s) between random effects have to be taken into account in any attempt to estimate the proportion of variation in outcome at any level or for any random coefficient. In addition, the inclusion of a covariate such as ENDOG can affect the variance estimates.

The total variation in HDRS scores over patients is defined as

$$\text{Var(level 2)} = \text{var}(v_{i0}) + \text{var}(v_{i1})(\text{WEEK})_{ij}^2 + 2\left[\text{cov}(v_{i0}, v_{i1})\right](\text{WEEK})_{ij}$$

The total variation is a function of the value assumed by the predictor WEEK, which has a random slope. As such, the total variation at the beginning of the study is

$$\begin{aligned}\text{Var(level 2)} &= \text{var}(v_{i0}) + \text{var}(v_{i1})(0)^2 + 2\left[\text{cov}(v_{i0}, v_{i1})\right](0) \\ &= \text{var}(v_{i0})\end{aligned}$$

while at the end of the study we have

$$\begin{aligned}\text{Var(level 2)} &= \text{var}(v_{i0}) + \text{var}(v_{i1})(5)^2 + 2\left[\text{cov}(v_{i0}, v_{i1})\right](5) \\ &= \text{var}(v_{i0}) + 25\,\text{var}(v_{i1}) + 10\,\text{cov}(v_{i0}, v_{i1})\end{aligned}$$

An estimate of the total variation at this level can be obtained by using the estimates of the variances and covariance obtained under this model. By substituting $\hat{\text{var}}(v_{i0})$, $\hat{\text{var}}(v_{i1})$, and $\hat{\text{cov}}(v_{i0}, v_{i1})$ into the equations above, we obtain the estimated variation in HDRS scores over patients at different points during the study period.

At the beginning of the study, the estimated total variation in HDRS scores over patients is simply the estimated variation in the random intercept, *i.e.*, $\hat{\text{var}}(v_{i0})$ = 12.62930. At the end of the study, the total variation at level-2 is estimated as

$$\begin{aligned}\hat{\text{var}}(\textit{level } 2) &= \hat{\text{var}}(v_{i0}) + 25\,\hat{\text{var}}(v_{i1}) + 10\,\hat{\text{cov}}(v_{i0}, v_{i1}) \\ &= 12.62930 + 25(2.07899) + 10(-1.42093) \\ &= 50.39475.\end{aligned}$$

At the beginning of the study we obtain

$$\begin{aligned}\frac{\hat{\text{var}}(\text{level } 2)}{\hat{\text{var}}(\text{level } 2) + \hat{\text{var}}(\text{level } 1)} &= \frac{12.62930}{12.62930 + 12.21663} \\ &= 0.5083\end{aligned}$$

and thus conclude that 50.8% of the variation in HDRS scores at this time is over patients. At the end of the study, we find that

$$\begin{aligned}\frac{\hat{\text{var}}(\text{level } 2)}{\hat{\text{var}}(\text{level } 2) + \hat{\text{var}}(\text{level } 1)} &= \frac{50.39475}{50.39475 + 12.21663} \\ &= 0.8049,\end{aligned}$$

so that only 20% of the variation in HDRS scores are estimated to be at the measurement level, with 80% at the patient level. As mentioned before, the total variation in HDRS scores is a function of the time of measurement, as represented by the variable WEEK. The very different estimates of variation at a patient level show how the introduction of an important predictor, in this case at the measurement level, can have an impact on variance estimates at a different level of the hierarchy. By the end of the study period, the residual variation over measurements has been dramatically reduced, this being explained to a large extent by the inclusion of the time effect. Most of the remaining unexplained variation is at the patient level.

As a result of this finding and in the light of our original research question, whether the initial depression classification of a patient is also related to the HDRS scores over the time in which medication is administered, the model will be extended to include the covariate ENDOG. This dichotomous variable assumes a value of 1 when endogenous depression was diagnosed, and 0 if not. In addition, we will provide for a possible interaction between depression classification and measurement occasion by including the interaction term WxENDOG in the model. While WxENDOG can be

viewed as a cross-level interaction, as WEEK is a measurement-level variable and ENDOG a patient-level variable, the inclusion of the patient-level variable ENDOG may enable us to explain more of the remaining variation in the random intercepts and slopes at the patient level.

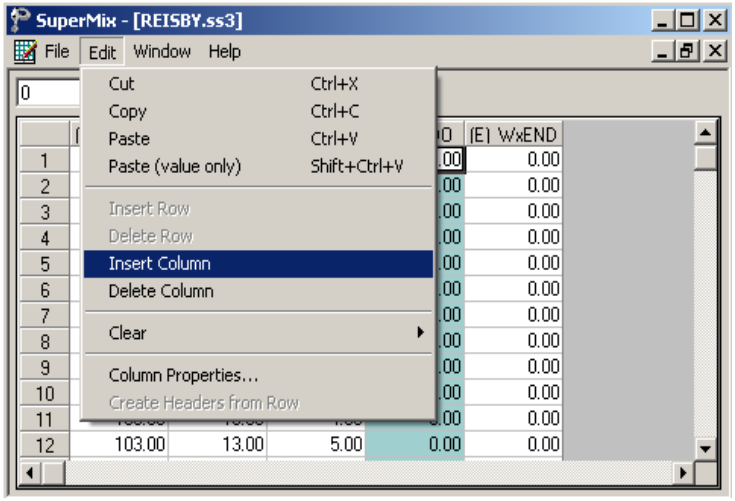
3.2.3 A 2-level random intercept-and-slope model with centered predictor

In the previous example, the time variable WEEK is coded from 0 to 5 and indicates the number of weekly follow-ups. The estimated average intercept of 23.577 obtained for this model represented the expected average HDRS score at the beginning of the study, *i.e.* WEEK = 0. An alternative formulation of the model that can be considered is one in which the estimated average intercept represents the expected average HDRS score midway through the study period. This linear transformation of the predictor variable WEEK, in which the grand mean of the variable is subtracted from each observed WEEK value, is referred to as grand mean centering. While the model based on the "raw" data and the model utilizing grand mean centered variables can be shown to be mathematically equivalent, the coefficients in these models have very different meanings.

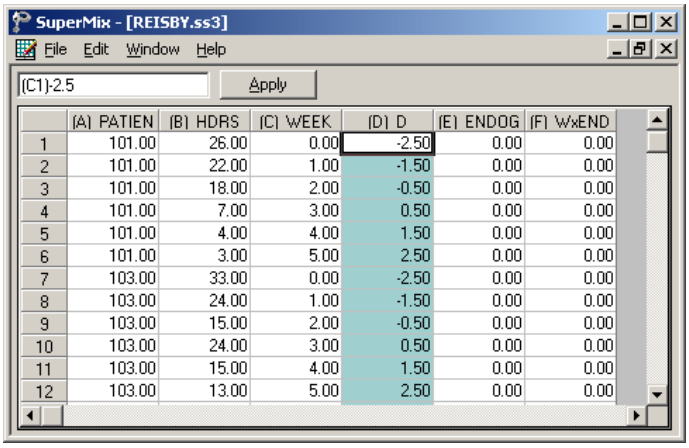
3.2.3.1 Preparing the data

Recall that the descriptive statistics in the previous model indicated a mean value over all level-1 observations of WEEK equal to 2.48. This is the true observed mean, compared to the value of 2.5 that would have been obtained if all patients had complete data over the course of the study. Here, we opt to use the value of 2.5 to center the WEEK variable.

To grand mean center the predictor WEEK, proceed as follows. Open the **reisby.ss3** in the SuperMix spreadsheet, then highlight the column WEEK. Select the **Insert Column** option on the **Edit** menu as shown below to insert a blank column named D after WEEK.

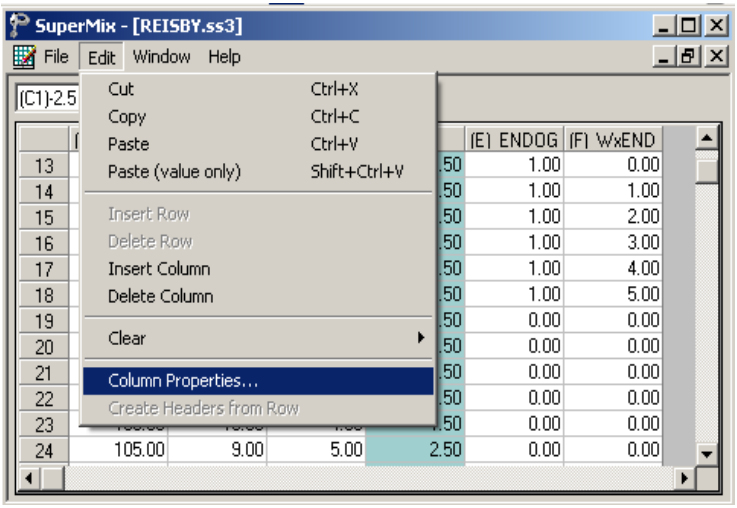


Keep the column D highlighted, type the formula (C1)–2.5 in the string field of the top-left corner and click on the **Apply** button to produce the following screen.

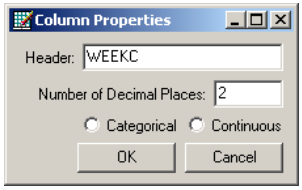


	(A) PATIEN	(B) HDRS	(C) WEEK	(D) D	(E) ENDOG	(F) WxEND
1	101.00	26.00	0.00	-2.50	0.00	0.00
2	101.00	22.00	1.00	-1.50	0.00	0.00
3	101.00	18.00	2.00	-0.50	0.00	0.00
4	101.00	7.00	3.00	0.50	0.00	0.00
5	101.00	4.00	4.00	1.50	0.00	0.00
6	101.00	3.00	5.00	2.50	0.00	0.00
7	103.00	33.00	0.00	-2.50	0.00	0.00
8	103.00	24.00	1.00	-1.50	0.00	0.00
9	103.00	15.00	2.00	-0.50	0.00	0.00
10	103.00	24.00	3.00	0.50	0.00	0.00
11	103.00	15.00	4.00	1.50	0.00	0.00
12	103.00	13.00	5.00	2.50	0.00	0.00

Rename the newly created variable to WEEKC by first highlighting the column, then selecting the **Column Properties** option on the **Edit** menu as shown below.



Input the desired variable name, *e.g.* WEEKC, in the **Header** string field as shown below and click on the **OK** button. By default, all variables are assumed to be continuous.



Save the changes to **reisby.ss3** by selecting the **Save** option on the **File** menu.

3.2.3.2 The model

The revised random intercept-and-slope model for the response variable HDRS may be expressed as

$$\text{HDRS}_{ij} = \beta_0 + \beta_1 \times \left(\text{WEEKC}\right)_{ij} + v_{i0} + v_{i1}\left(\text{WEEKC}\right)_{ij} + e_{ij}$$

or, alternatively, as

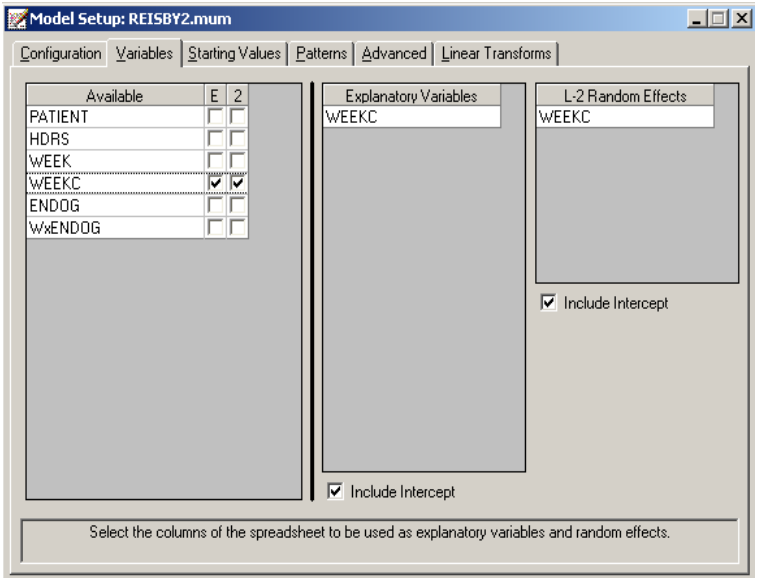
$$\mathrm{HDRS}_{ij} = \beta_0 + \beta_1 \times \left[\left(\mathrm{WEEK} \right)_{ij} - \overline{\mathrm{WEEK}} \right] + v_{i0} + v_{i1} \left[\left(\mathrm{WEEK} \right)_{ij} - \overline{\mathrm{WEEK}} \right] + e_{ij}$$

where $\overline{\mathrm{WEEK}} = 2.5$.

3.2.3.3 Setting up the analysis

Open the previous model setup for **reisby1.mum**. Save the file as **reisby2.mum** by using the **Save As** option on the **File** menu. Change the title on the **Configuration** tab if desired.

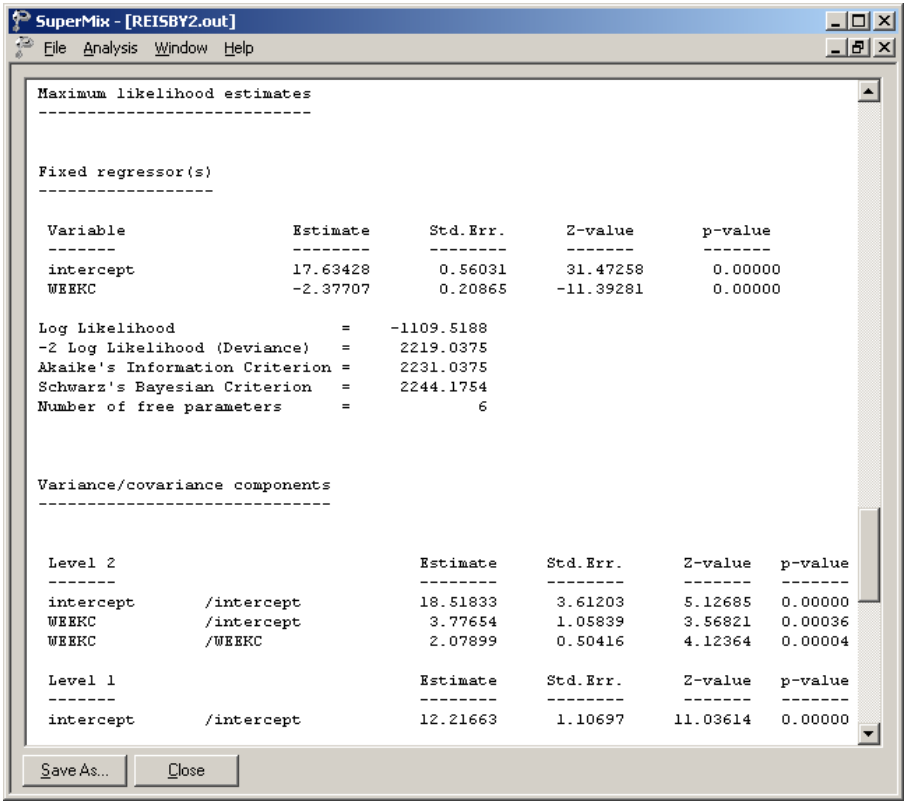
Click on the **Variables** tab and select WEEKC both as **Explanatory Variable** and **L-2 Random Effects** instead of WEEK as shown below.



Save the changes to the file **reisby2.mum**. Select the **Run** option on the **Analysis** menu to produce the output file **reisby2.out**. Use the **Analysis**, **View Output** option to open the output file.

3.2.3.4 Discussion of results

The output file contains the final estimates of the fixed and random coefficients included in the model, along with some goodness of fit measures as given below. Note that the use of grand mean centering of the time variable has no effect on the fit statistics.



3.2.3.5 Interpreting the results

Comparison of models

Table 3.3 contains the estimates and standard errors of the above two analyses. The coefficient for WEEKC is the same as for the uncentered variable WEEK. However, the variance of the random intercept ($\sigma^2_{v_0}$) and the covariance term $\sigma_{v_0 v_1}$ have changed. The covariance between the intercept and the WEEKC slope is now significant.

Table 3.3: Estimates and standard errors for two models

Coefficient	Level-2 model	
	WEEK = 0 ~ 5	WEEKC = –2.5 ~ 2.5
β_0	23.57695	17.63428
	(0.54555)	(0.56031)
β_1	–2.37707	–2.37707
	(0.20865)	(0.20865)
$\sigma^2_{v_0}$	12.6293	18.51833
	(3.46653)	(3.61203)
$\sigma_{v_0 v_1}$	–1.42093	3.77654
	(1.02595)	(1.05839)
$\sigma^2_{v_1}$	2.07899	2.07899
	(0.50417)	(0.50416)
σ^2_e	12.21663	12.21663
	(1.10697)	(1.10697)

Table 3.3: Estimates and standard errors for two models (continued)

Deviance	2219.0375	2219.0375
AIC	2231.0375	2231.0375
SBC	2244.1754	2244.1754
Number of free parameters	6	6

As shown above, the estimates of the slope and its variance are the same. This is because the scale of WEEK was not changed; only its location changed. The estimated intercept decreased from 23.58 to 17.63, which corresponds to the average HDRS score at week 2.5 instead of week 0. Similarly, the $\sigma^2_{v_0}$ of intercept increased to 18.52, which shows the increase of the individual variance at week 2.5. The change of $\sigma_{v_0 v_1}$ is interesting: not only the value changed, but also the sign. The covariance of the first analysis tells us that the higher the variance of intercept, the lower the variance of slope. Or say, at week 1, the HDRS score decreases at a faster rate for those patients who started with higher HDRS. However, at week 2.5, the patients with higher HDRS tend to improve less.

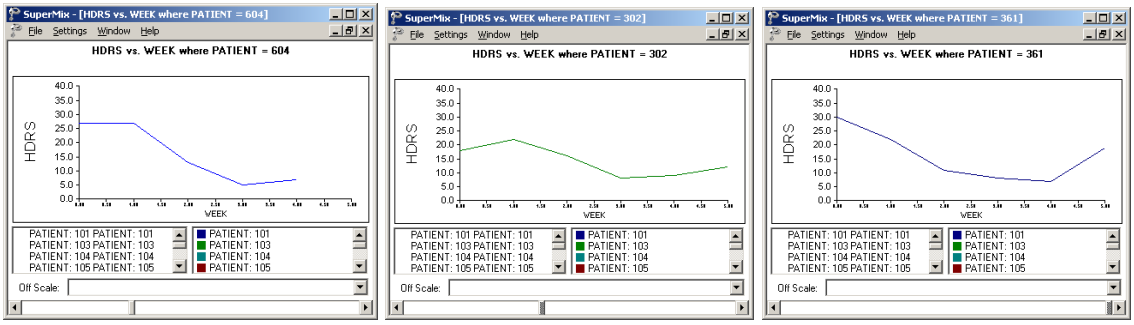


Figure 3.8: Changes in covariance over time

When looking at the three HDRS versus WEEK plots for patient 604, 302 and 361, we can see why this could happen. The graphs show the change of $\sigma_{v_0 v_1}$ from week 0 to week 2.5.

3.2.4 A random intercept-and-slope with a covariate and an interaction term

The type of depression a patient was diagnosed with was recorded as part of the study and information on this patient characteristic is represented by the variable ENDOG, which assumes a value of 1 for patients with endogeneous depression and 0 otherwise. Including this variable in the model allows us to explore the potential relationship between a patient's HDRS score and the type of depression the patient was diagnosed with. Moreover, it is possible that the trend in HDRS scores over the study period may differ for the two ENDOG groups. Including an interaction term between the time of measurement and the type of depression in the model will allow us to evaluate this potential relationship as well.

3.2.4.1 The model

We now include ENDOG and WxENDOG in the level-1 model. ENDOG is a dummy variable representing the type of depression a patient was diagnosed with, and WxENDOG represents the interaction between WEEK and ENDOG. The model shows changes at both levels: at level 2, the covariate ENDOG is now included, while at level 1 the interaction between WEEK and ENDOG, which can potentially change from week to week, is added. The revised model for the response variable HDRS may be expressed as

Level-1 model:

$$\mathrm{HDRS}_{ij} = b_{0i} + b_{1i} \times \left(\mathrm{WEEK}\right)_{ij} + b_{2i} \times \left(\mathrm{WxENDOG}\right)_{ij} + e_{ij}$$

Level-2 model:

$$\begin{aligned} b_{0i} &= \beta_0 + \beta_3 \times \left(\mathrm{ENDOG}\right)_i + v_{0i} \\ b_{1i} &= \beta_1 + v_{1i} \\ b_{2i} &= \beta_2 \end{aligned}$$

or, in mixed model formulation, as

$$\text{HDRS}_{ij} = \beta_0 + \beta_1 \times (\text{WEEK})_{ij} + \beta_2 \times (\text{WxENDOG})_{ij} + \beta_3 \times (\text{ENDOG})_i + v_{0i} + v_{1i} \times (\text{WEEK})_{ij} + e_{ij}$$

where β_0 denotes the average HDRS level at week 0 for the non-endogenous depression patients (ENDOG=0), β_1 refers to the weekly improvement for the non-endogenous group, β_2 indicates the expected change in HDRS score for a unit change in the value of the interaction term WxENDOG, and β_3 refers to the average expected change in HDRS level for endogenous patients. v_{0i} is the individual deviation from the average intercept. v_{1i} denotes the average deviation from the slope, or say, average improvement of the HDRS.

We can also write the model in terms of our original variables (WEEK and ENDOG) as:

Level-1 model:

$$\text{HDRS}_{ij} = b_{0i} + b_{1i} \times (\text{WEEK})_{ij} + b_{2i} \times (\text{WxENDOG})_{ij} + e_{ij}$$

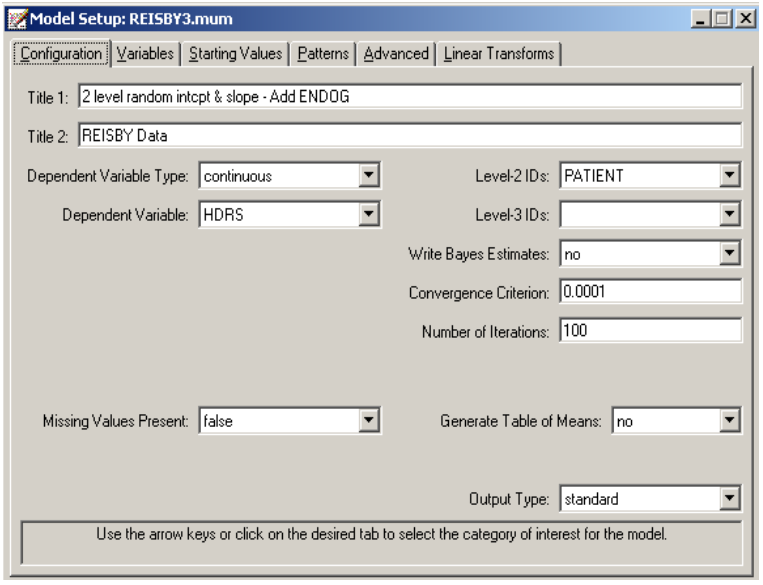
Level-2 model:

$$b_{0i} = \beta_0 + \beta_2 \times (\text{ENDOG})_i + v_{0i}$$
$$b_{1i} = \beta_1 + \beta_4 \times (\text{ENDOG})_i + v_{1i}$$

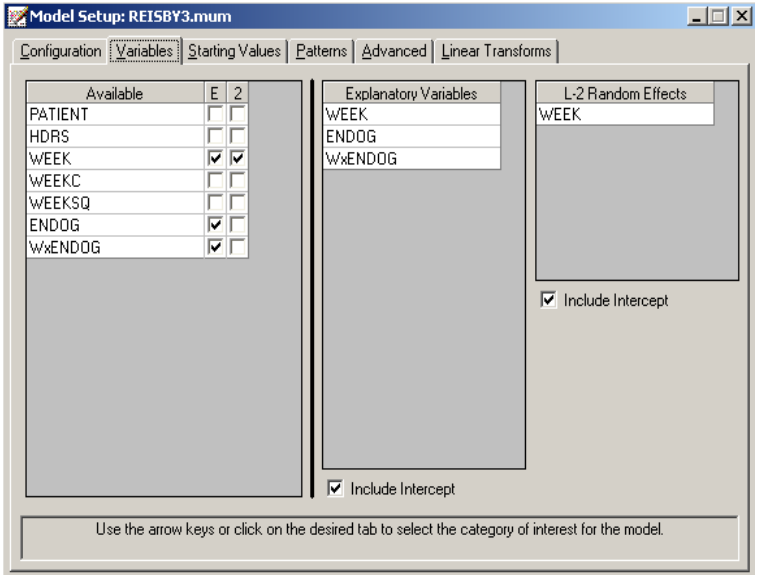
3.2.4.2 Setting up the analysis

To create the model specifications for this model, we start by opening **reisby.ss3** in a SuperMix spreadsheet window. Then we use the **Open Existing Model Setup** option on the **File** menu to load the **Model Setup** window for **reisby1.mum**. Save the file as **reisby3.mum** by using the **Save As** option on the **File** menu. Change the string in the

Title 1 text box on the **Configuration** screen to reflect the new model, thereby producing the following dialog box.



Next, click on the **Variables** tab to proceed to the **Variables** screen of the **Model Setup** window.



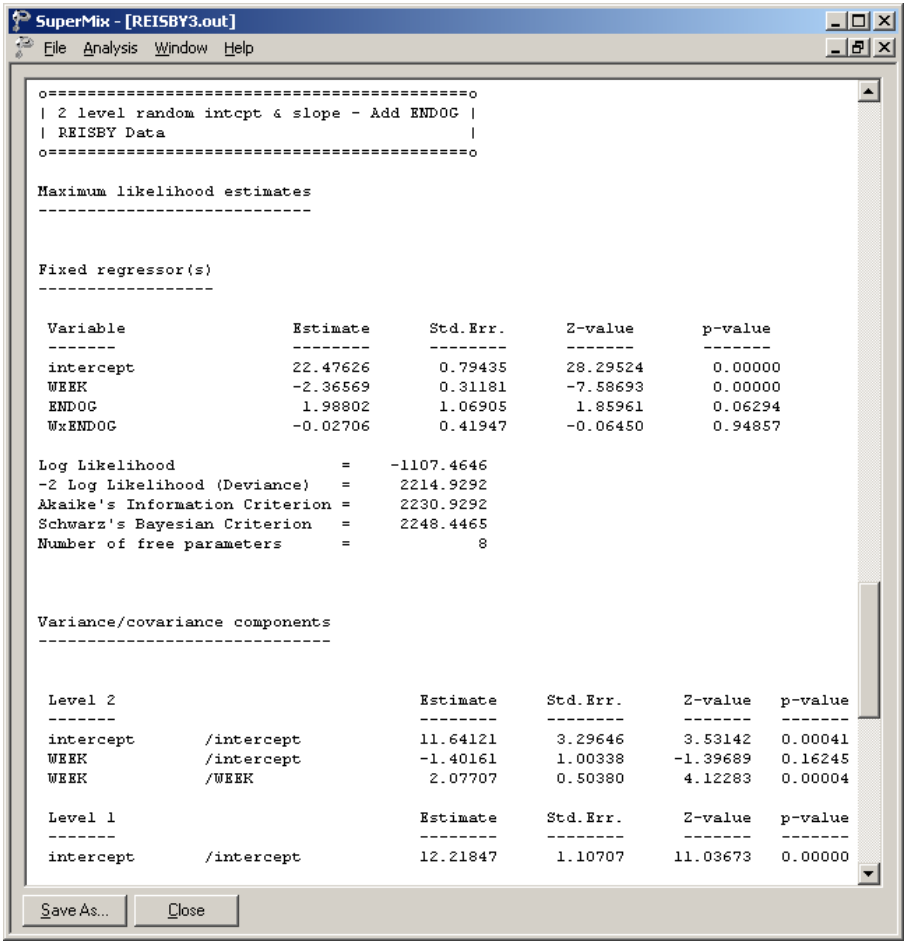
The two covariates are specified by checking the **E** check boxes for ENDOG and WxENDOG respectively in the **Available** grid respectively to produce the following **Variables** tab.

Save the changes to the file **reisby3.mum**. To fit the revised model to the data, select the **Run** option on the **Analysis** menu to produce the output file **reisby3.out**.

3.2.4.3 Interpreting the results

Fixed effects results

A portion of the output file **reisby3.out** is shown below.



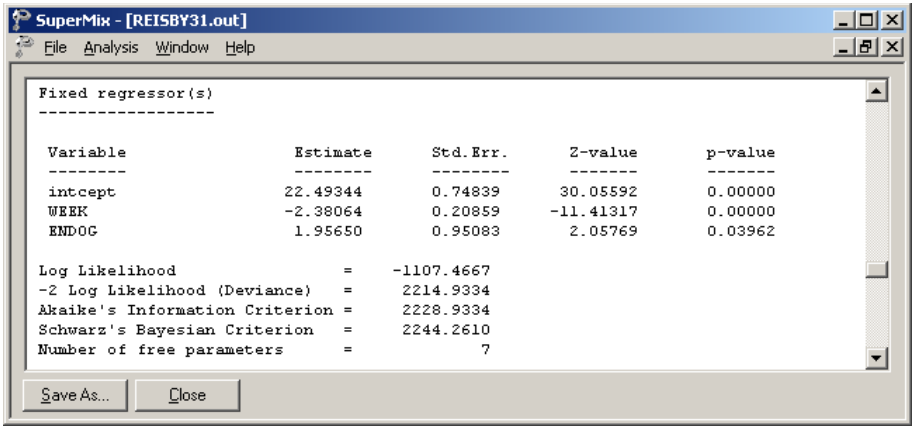
The interaction WxENDOG between the time variable WEEK and the depression classification variable ENDOG, is not significant. Given this, we can take a closer look at the estimated coefficients for the main effects WEEK and ENDOG respectively. Note, however, that the *p*-value for the ENDOG coefficient is larger than 0.05, and thus can only be considered significant at a 10% level of significance. The effect of time, on the other hand, is found to be highly significant. While the average HDRS score is predicted to decrease by –2.37 score scale units each week, patients classified as having endogenous depression (*i.e.*, ENDOG = 1) are predicted to have a HDRS score of 2 units higher at all occasions.

To obtain the predicted average HDRS scores, the estimates obtained from the output are used:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1(\text{WEEK}) + \hat{\beta}_2(\text{ENDOG}) + \hat{\beta}_3(\text{WxENDOG})$$
$$= 22.47626 - 2.36569(\text{WEEK}) + 1.98802(\text{ENDOG}) - 0.02706(\text{WxENDOG})$$

Model comparison

A question that arises from inspection of the results obtained thus far is whether the interaction term contributes overall to the explanation of the variation in the HDRS scores. To test this, we can fit a model without the interaction term and use the deviance reported in the output to compare results for the model with interaction and the model without this term. The relevant output from an analysis without the interaction term is shown below. We note that the deviance obtained for the simpler model is almost identical to that of the model considered in this section. Based on this, we conclude that a model without the interaction WxENDOG would fit the data as well as the one with the interaction term included.



In addition, we can test the hypothesis that the model with covariate (ENDOG) fits the data better than the random intercept and slope model considered previously. To test this hypothesis, we calculate the difference between the –2 log likelihood value

obtained for the previous model and the –2 log likelihood value for the current model. It can be shown that this difference of 2219.04 – 2214.93 = 4.11 has a χ^2 distribution with associated degrees of freedom equal to the difference in the number of parameters estimated in the two examples, *i.e.*, 8 – 7 = 1 degrees of freedom. Since the *p*-value for this test statistic is less than 0.05, it is concluded that the random intercept-and-slope model with ENDOG as a covariate provides a better description of the data than the original random intercept-and-slope model. This finding is supported by the fact that the *p*-value for ENDOG when the interaction effect between WEEK and ENDOG is excluded equals 0.04.

3.2.5 A random intercept-and-slope quadratic model

3.2.5.1 The model

In this section we include an additional predictor and a random term to examine a possible quadratic response trend in HDRS scores over time. Keeping the level-2 model the same as before, the corresponding model for the response variable HDRS may be expressed as

Level-1 model:

$$\text{HDRS}_{ij} = b_{0i} + b_{1i} \times \left(\text{WEEK}\right)_{ij} + b_{2i} \times \left(\text{WEEK}^2\right)_{ij} + e_{ij}$$

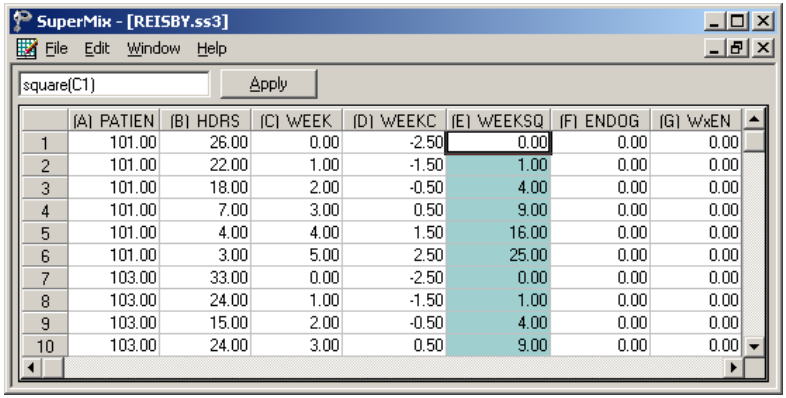
Level-2 model:

$$b_{0i} = \beta_0 + v_{0i}$$
$$b_{1i} = \beta_1 + v_{1i}$$
$$b_{2i} = \beta_2 + v_{2i}$$

3.2.5.2 Preparing the data

Create a new blank variable named WEEKSQ as shown in section 2.5.1. Highlight the column WEEKSQ, type the formula SQUARE(C1) where C = WEEK in the string

field and click on the **Apply** button to produce the following screen. Save the change to **reisby.ss3**.

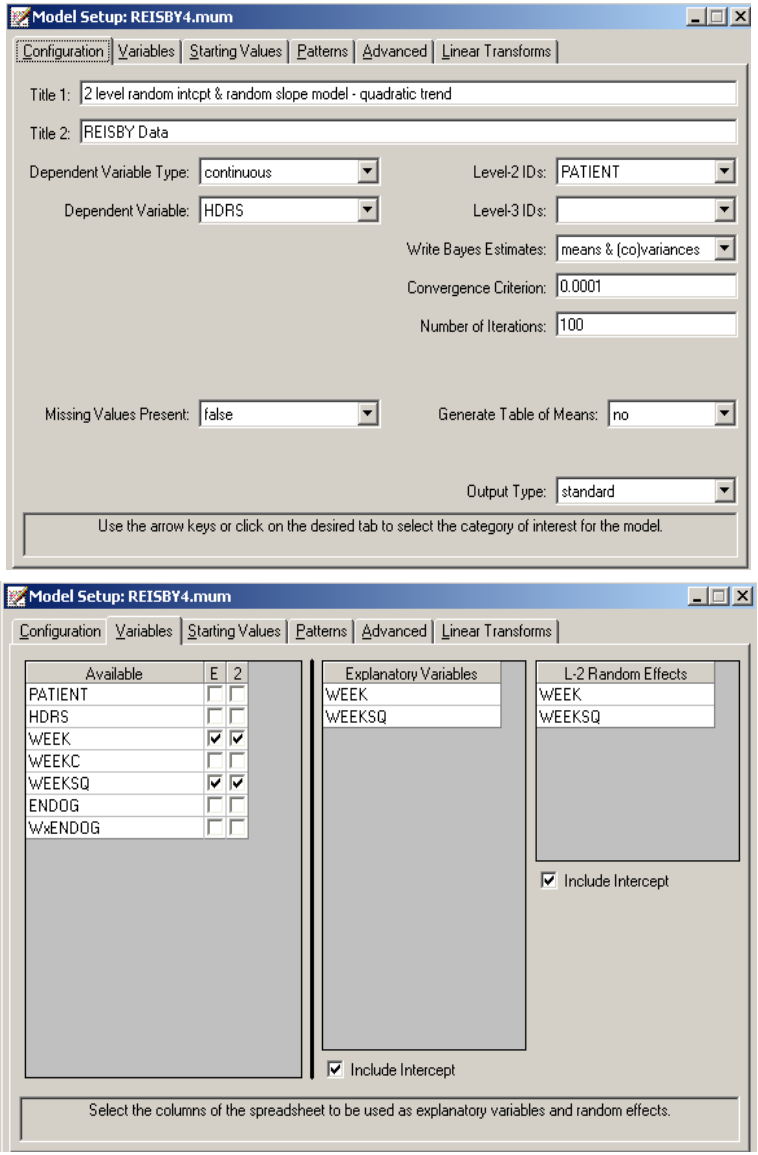


	(A) PATIEN	(B) HDRS	(C) WEEK	(D) WEEKC	(E) WEEKSQ	(F) ENDOG	(G) WxEN
1	101.00	26.00	0.00	-2.50	0.00	0.00	0.00
2	101.00	22.00	1.00	-1.50	1.00	0.00	0.00
3	101.00	18.00	2.00	-0.50	4.00	0.00	0.00
4	101.00	7.00	3.00	0.50	9.00	0.00	0.00
5	101.00	4.00	4.00	1.50	16.00	0.00	0.00
6	101.00	3.00	5.00	2.50	25.00	0.00	0.00
7	103.00	33.00	0.00	-2.50	0.00	0.00	0.00
8	103.00	24.00	1.00	-1.50	1.00	0.00	0.00
9	103.00	15.00	2.00	-0.50	4.00	0.00	0.00
10	103.00	24.00	3.00	0.50	9.00	0.00	0.00

3.2.5.3 Setting up the analysis

Again, we can modify the model setup file of **reisby1.mum** by first opening it, then saving the file as **reisby4.mum**. Change the title on the **Configuration** tab and request Bayes estimates by selecting the **means & (co)variances** option from the **Write Bayes Estimates** drop-down list.

Next, click on the **Variables** tab to proceed to the **Variables** screen of the **Model Setup** window. The two covariates are specified by checking the **E** and **2** check boxes for WEEKSQ in the **Available** grid to produce the **Variables** screen shown below.

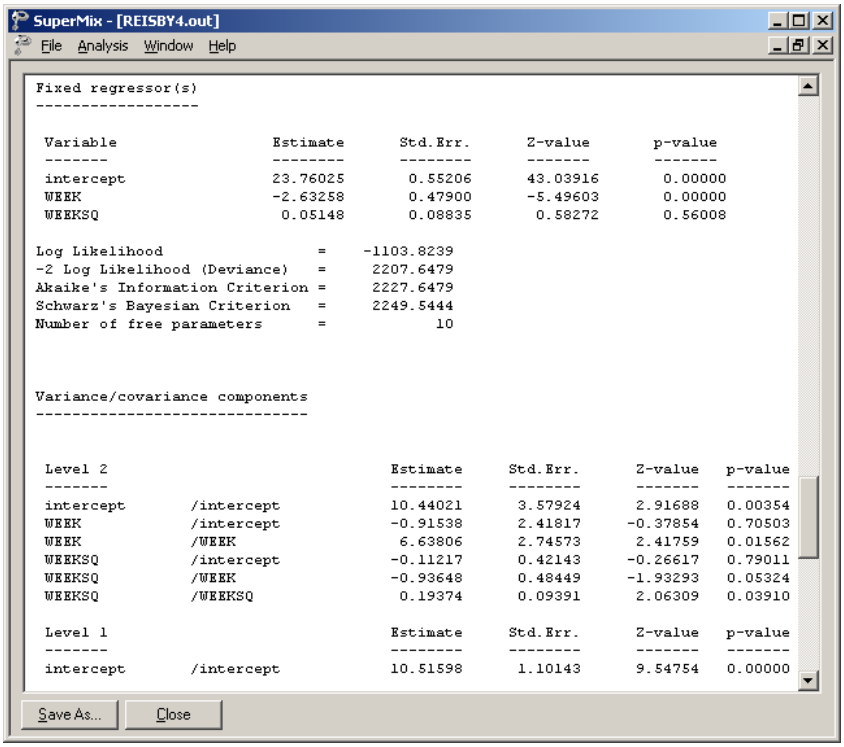


Save the changes to the file **reisby4.mum** and run the model.

3.2.5.4 Interpreting the results

A portion of the output file **reisby4.out** is shown below.

Fixed effects results



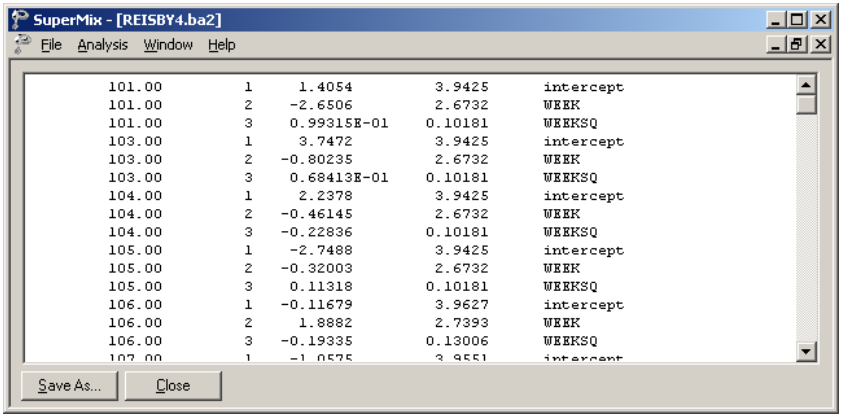
The level-1 estimate of the WEEKSQ coefficient is 0.05, which turns out not to be significant (p = 0.56). On the other hand, the WEEKSQ random effect is significant at a 5% level (p = 0.04). Comparing the present results with those reported in **reisby1.out**, we see that the deviance difference of 2219.04 – 2207.65 = 11.19 with 10 – 7 = 3 degrees of freedom, indicating an improved overall model fit at a 5% significance level. These results imply that, although the mean trend of HDRS scores over time is linear, some of the individuals' trajectories are quadratic.

3.2.5.5 Residuals

Level 2 Bayes results

Up to this point, we have considered results averaged over all patients. We now turn our attention to the residual file **reisby4.ba2**, which offers the opportunity to take a closer look at the results by individual patient. After running the above model, select the **Analysis**, **View L-2 Bayes Results** option to open the image below. The contents of this file are displayed for the first 5 patients. Three lines of information are given for each patient, containing, in order of appearance,

- the number of the patient in the data set,
- the number of the empirical Bayes coefficient,
- the empirical Bayes estimate,
- the estimated variance of the Bayes coefficient, and
- the name of the associated coefficient as used in the model.



To obtain patient-specific predicted HDRS scores, the empirical Bayes estimate for each patient have to be taken into account, as these estimates indicate the extent to which the random intercept or slope for that patient deviates from the intercept and slope over all patients. Patient-specific predicted HDRS scores are calculated as

$$\hat{y}_{ij} \mid \hat{\beta} = 23.76025\text{-}2.63258 \times \text{WEEK}_{ij} + 0.05148 \times \text{WEEK}^2_{ij}$$
$$+ \hat{v}_{0i} + \hat{v}_{1i} \times \text{WEEK}_{ij} + \hat{v}_{2i} \times \text{WEEK}^2_{ij}$$

For the first patient shown in the residual file above, we have $\hat{v}_{i0}$ = 1.4054, $\hat{v}_{i1}$ = –2.6506 and $\hat{v}_{i2}$ = 0.099315. From this information, we can already tell that the intercept for the patient is higher than average, but that the WEEK slope for this patient is lower than average. The positive value of the quadratic term indicates that the decreasing rate slows down more quickly than average with an increase in time. The predicted HDRS score for this patient (PATIENT = 101) is found to be

$$\hat{y}_{ij} \mid \hat{\beta} = 23.76025\text{-}2.63258 \times \text{WEEK}_{ij} + 0.05148 \times \text{WEEK}^2_{ij}$$
$$+ 1.4054\text{-}2.6506 \times \text{WEEK}_{ij} + 0.099315 \times \text{WEEK}^2_{ij} .$$

Substituting the WEEK with values 0, 1, …, 5, we get the predicted HDRS scores for Patient 101, and similarly, for all the other patients. Table 3.4 and the graphical display below give the predicted HDRS for the first 5 patients.

Table 3.4: Predicted HDRS values for selected patients

	Patient 101	Patient 103	Patient 104	Patient 105	Patient 106	Population Avg.
Week 0	25.166	27.507	25.998	21.011	23.643	23.760
Week 1	20.033	24.192	22.727	18.224	22.757	21.179
Week 2	15.202	21.117	19.102	15.765	21.587	18.701
Week 3	10.673	18.282	15.124	13.636	20.133	16.326
Week 4	6.446	15.686	10.792	11.836	18.396	14.054
Week 5	2.520	13.330	6.106	10.365	16.375	11.884

We find that Patient 101 had a higher initial HDRS score, but over time obtained a lower than average score. For Patient 103, a higher than average predicted HDRS score is obtained at each time point. In contrast, Patient 105 scored lower at each

time point. The quadratic term doesn't affect much of the population average; however the effect is obvious for Patients 105 and 106.

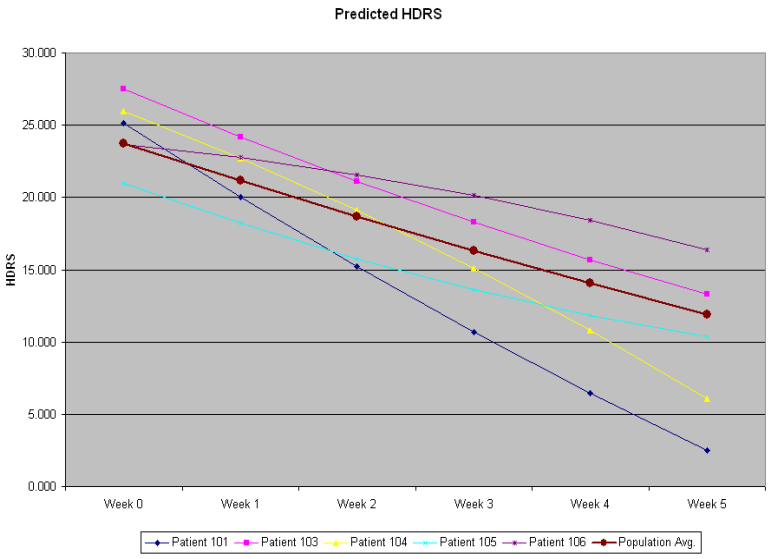


Figure 3.9: Predicted HDRS for selected patients

Model-based graphs

Residual plot

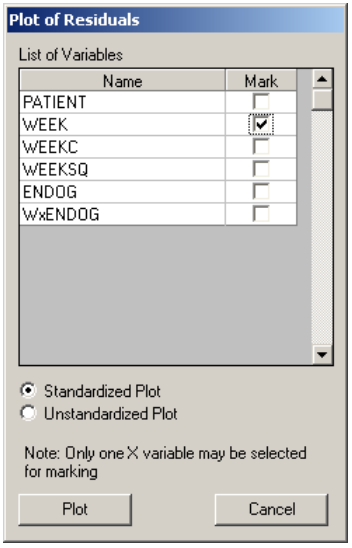
Level-1 residuals can also be obtained, either for a typical or specific patient, by using the empirical Bayes estimates. The residuals for a typical patient are obtained as

$$\begin{aligned}\text{Patient residual} &= \text{Observed HDRS score} - \hat{y} \mid \beta \\ &= \text{Observed HDRS score} - [23.76025 - 2.63258 \times \text{WEEK}_{ij} \\ &\quad + 0.05148 \times \text{WEEK}^2_{ij}\end{aligned}$$

The residuals for a specific patient use the additional information given by the empirical Bayes residuals and have the form

$$\begin{aligned}\text{Patient-specific residual} &= \text{Observed HDRS score} - \hat{y} \mid \beta \\ &= \text{Observed HDRS score} - [23.76025\text{-}2.63258 \times \text{WEEK}_{ij} \\ &\quad + 0.05148 \times \text{WEEK}^2_{ij} + \hat{v}_{0i} + \hat{v}_{10} \times \text{WEEK}_{ij} + \hat{v}_{20} \times \text{WEEK}^2_{ij}]\end{aligned}$$

Select the **Residuals** option on the **File**, **Model-based Graphs** menu to activate the **Plot of Residuals** dialog box. Check the **Mark** check box for WEEK as shown below, then click on the **Plot** button.



The graph obtained, as shown below, shows that, in general, the range of the level-1 residuals is $(-5; 5)$.

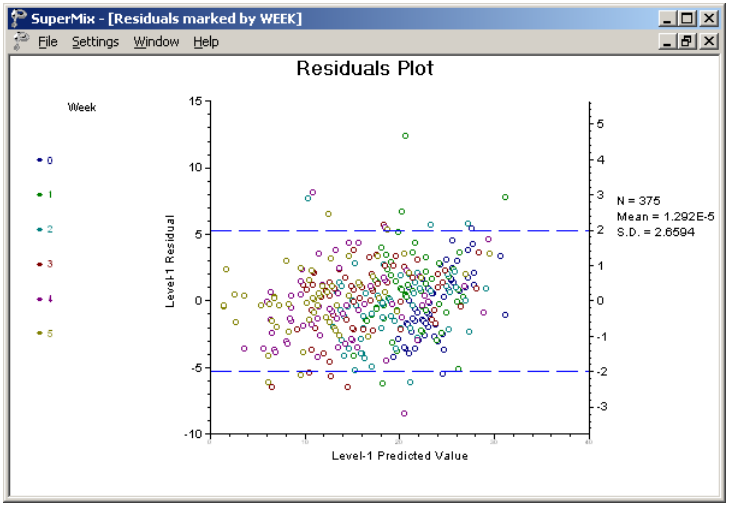
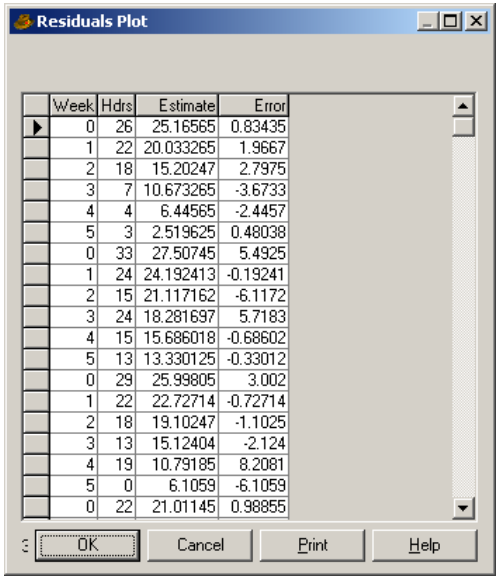


Figure 3.10: Plot of level-1 residuals vs. predicted values

Inspection of these residuals can be useful in examining the distributional assumptions for the level-1 data, in this case at the measurement level. For the current example, residuals for a typical patient have a mean of 0.000 with standard error of 2.66. Double-click on the middle of the graph to open an additional window that shows the detailed residual data for each observation.

We note that the estimate for Patient 101 at the beginning of the study was 25.166, and 2.520 at the end of the study. On both occasions, the residuals associated with these estimates were positive, indicating that the estimates are above estimated average.



3.2.6 A 2-level random intercept-and-slope model with autocorrelated errors

In the mixed models discussed so far, it was assumed that the level-1 errors are conditionally independent from each other. However, the errors could be correlated over time. Different types of correlated error structures are available in SuperMix: the first-order stationary autoregressive process, stationary AR(1), the first-order non-stationary autoregressive process, non-stationary AR(1), the first-order stationary moving average process, MA(1), the first-order stationary autoregressive moving average process, ARMA(1), and a general Toeplitz autocorrelation structure.

The stationary AR(1) and ARMA(1) use the stationary assumption, that is that the variance of errors is constant over time and that the covariance of errors from differing times depends only on the time interval between these time points and not on the starting time point. The assumption of stationarity is relaxed in the other two types of models. In SuperMix, the maximum marginal likelihood (MML) solution at convergence is obtained by first using the EM algorithm and then Fisher scoring iterations.

3.2.6.1 The non-stationary AR(1) model

The model here is essentially the same as the one we had in section 3.2.2, apart from the autocorrelated error term.

Level-1 model:

$$\text{HDRS}_{ij} = b_{0i} + b_{1i} \times \text{WEEK}_{ij} + b_{2i} \times \text{WxENDOG}_{ij} + e_{ij},$$

where

$$e_{ij} = \rho\, e_{i,j-1} + \varepsilon_{ij}$$

with ρ denoting the AR coefficient.

Level-2 model:

$$\begin{aligned} b_{0i} &= \beta_0 + \beta_3 \times \text{ENDOG}_i + v_{0i} \\ b_{1i} &= \beta_1 + v_{1i} \\ b_{2i} &= \beta_2 \end{aligned}$$

We can rewrite the model as follows:

$$\begin{aligned} \text{HDRS}_{ij} &= \beta_0 + \beta_1 \times \text{WEEK}_{ij} + \beta_2 \times \text{ENDOG}_{ij} + \beta_3 \times \text{WEEK} \times \text{ENDOG}_{ij} \\ &\quad + u_{0i} + u_{1i} \times \text{WEEK}_{ij} + e_{ij} \\ &= \beta_0 + \left(\beta_1 + u_{1i}\right) \times \text{WEEK}_{ij} + \beta_2 \times \text{ENDOG}_{ij} + \beta_3 \times \text{WEEK} \times \text{ENDOG}_{ij} \\ &\quad + e_{ij} + u_{0i} \end{aligned}$$

The difference between the present and previous models lies in the assumption concerning the error term. Previously, we assumed that $\mathbf{e}_i = (e_{i1}, e_{i2}, ..., e_{in_i})'$: $N\left(\mathbf{0}, \sigma^2 \mathbf{I}_i\right)$, where $\mathbf{I}_i$ is an identity matrix of order $n_i \times n_i$. Now we assume that the errors are autocorrelated, and that $\mathbf{e}_i$: $N\left(\mathbf{0}, \sigma^2 \mathbf{\Omega}_i\right)$, where $\mathbf{\Omega}_i$ is the autocorrelation matrix.

The analysis – step 1: starting values from a non-AR model

Two types of iteration algorithms, EM and Fisher scoring, are used for fitting an autoregressive model:

- The EM solution proceeds by assigning starting values for the structural and population parameters.
- The Fisher scoring procedure utilizes the first derivatives and expected values of the second derivatives to obtain improved parameter estimates.

Although the Fisher scoring solution is a significant improvement in terms of speed of convergence over the EM solution, it can fail in the estimation of the covariance matrix of the random effects when these terms become very small. The most reliable way to minimize the chance of encountering convergence problems is first obtaining the starting values by running the model without autocorrelated errors, then substituting the starting values obtained prior to fitting the AR model.

Recall that in Section 3.2.4 we fitted the model

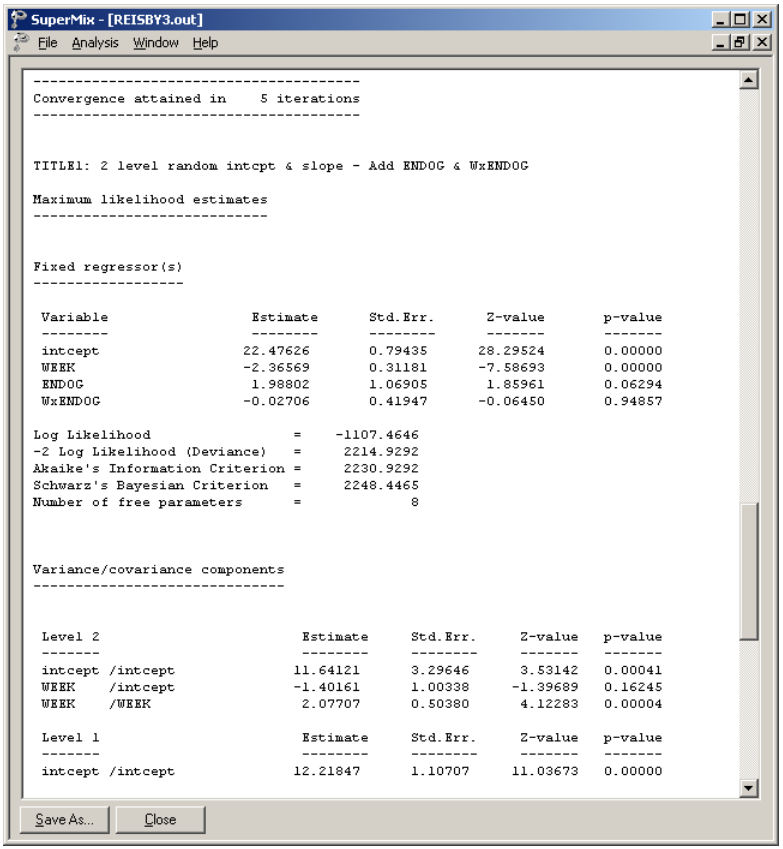
Level-1 model:

$$\text{HDRS}_{ij} = b_{0i} + b_{1i} \times \text{WEEK}_{ij} + b_{2i} \times \text{WxENDOG}_{ij} + e_{ij}$$

Level-2 model:

$$b_{0i} = \beta_0 + \beta_3 \times \text{ENDOG}_i + v_{0i}$$
$$b_{1i} = \beta_1 + v_{1i}$$
$$b_{2i} = \beta_2$$

The estimates obtained for that model are repeated below. The level-2 estimated variance of intercept and WEEK are 11.64121 and 2.07707 respectively. The estimated level-2 covariance is –1.40161. The estimated level-1 variance is 12.21847. These numbers will be used as the starting values in the non-stationary AR model to be fitted next.

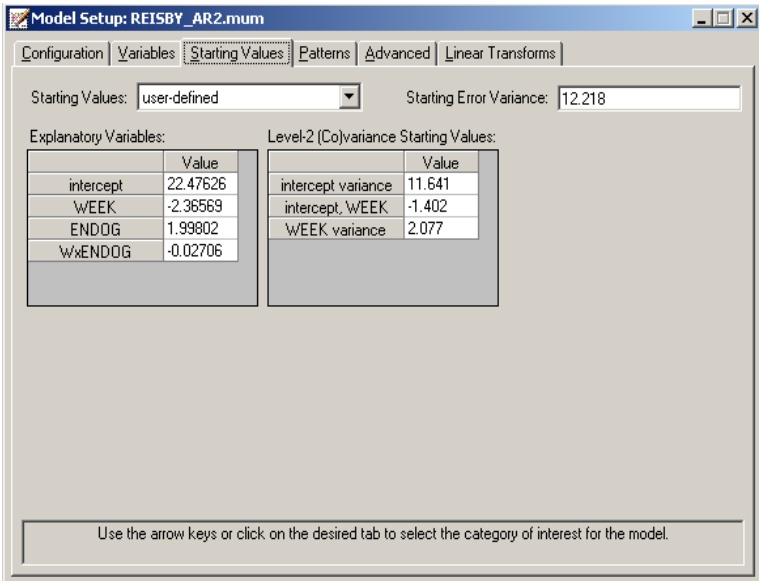



```
----------------------------------------
Convergence attained in    5 iterations
----------------------------------------


TITLE1: 2 level random intcpt & slope - Add ENDOG & WxENDOG

Maximum likelihood estimates
----------------------------


Fixed regressor(s)
------------------

 Variable                Estimate      Std.Err.      Z-value       p-value
 --------                --------      --------      -------       -------
 intcept                22.47626       0.79435      28.29524       0.00000
 WEEK                   -2.36569       0.31181      -7.58693       0.00000
 ENDOG                   1.98802       1.06905       1.85961       0.06294
 WxENDOG                -0.02706       0.41947      -0.06450       0.94857

Log Likelihood               =    -1107.4646
-2 Log Likelihood (Deviance) =     2214.9292
Akaike's Information Criterion =     2230.9292
Schwarz's Bayesian Criterion =     2248.4465
Number of free parameters    =             8



Variance/covariance components
------------------------------


 Level 2                        Estimate      Std.Err.      Z-value   p-value
 -------                        --------      --------      -------   -------
 intcept /intcept              11.64121       3.29646       3.53142   0.00041
 WEEK    /intcept              -1.40161       1.00338      -1.39689   0.16245
 WEEK    /WEEK                  2.07707       0.50380       4.12283   0.00004

 Level 1                        Estimate      Std.Err.      Z-value   p-value
 -------                        --------      --------      -------   -------
 intcept /intcept              12.21847       1.10707      11.03673   0.00000
```

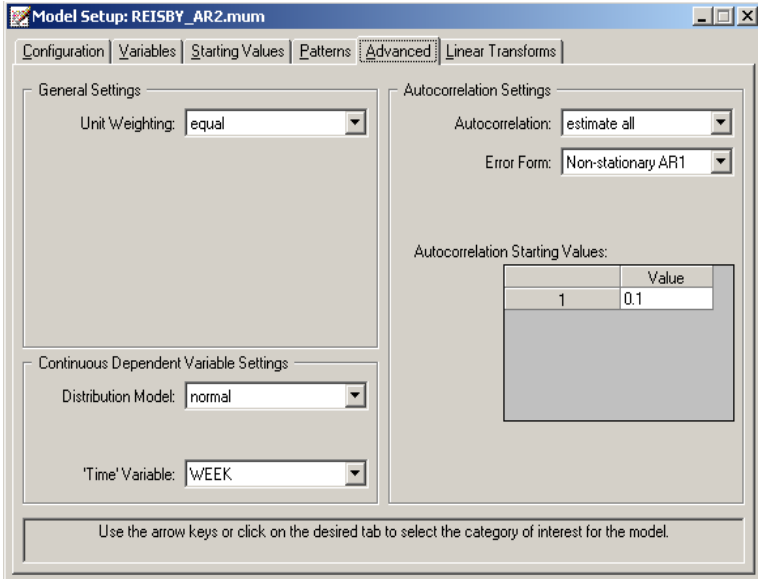
The analysis – step 2: non-stationary AR model

We modify the model setup file, **reisby3.mum,** by first saving the file as **reisby_ar2.mum**. Change the title on the **Configuration** screen. Keep the settings of the **Variables** tab the same as before.

Click on the **Starting Values** tab. Select the **user-defined** option from the **Starting Values** drop down list to activate the grid fields for the starting values. Input the starting values we obtained from **reisby.out** to generate the following screen.



Click on the **Advanced** tab to proceed to the **Advanced** screen. First, select the **estimate all** option from the **Autocorrelation** drop down list; then select **Non-stationary AR1** as the **Error Form** and specify WEEK as the **'Time' Variable**. Input 0.1 in the **Autocorrelation Starting Values** grid field to get the **Advanced** screen as shown below.



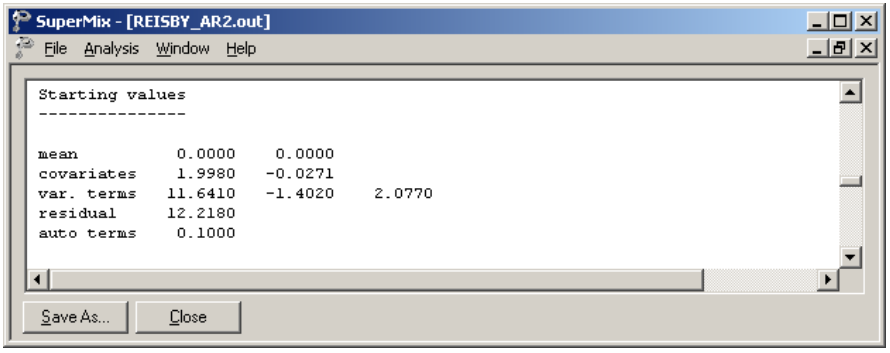
Save the changes to **reisby_ar2.mum** and run the model to produce the output file **reisby_ar2.out**.

3.2.6.2 Interpreting the output

The output for the AR model first shows the syntax information of the model setup. The number of observations, hierarchical structure of the 2-level model and descriptive statistics follow next.

The starting values

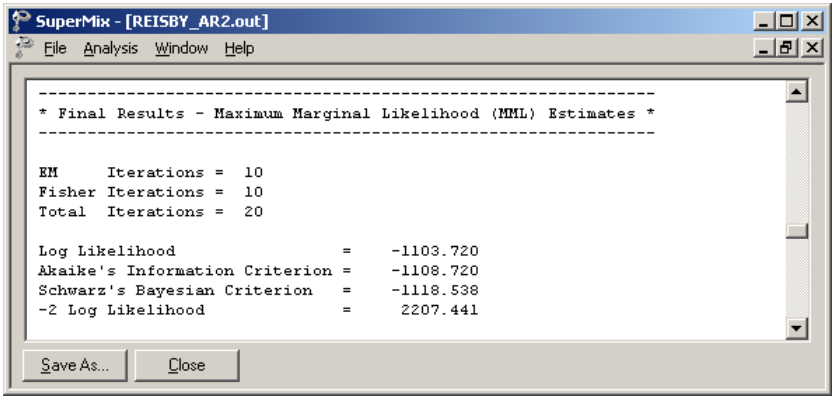
The starting values could either be user-defined or program generated. In our case the user-defined starting values are listed below.



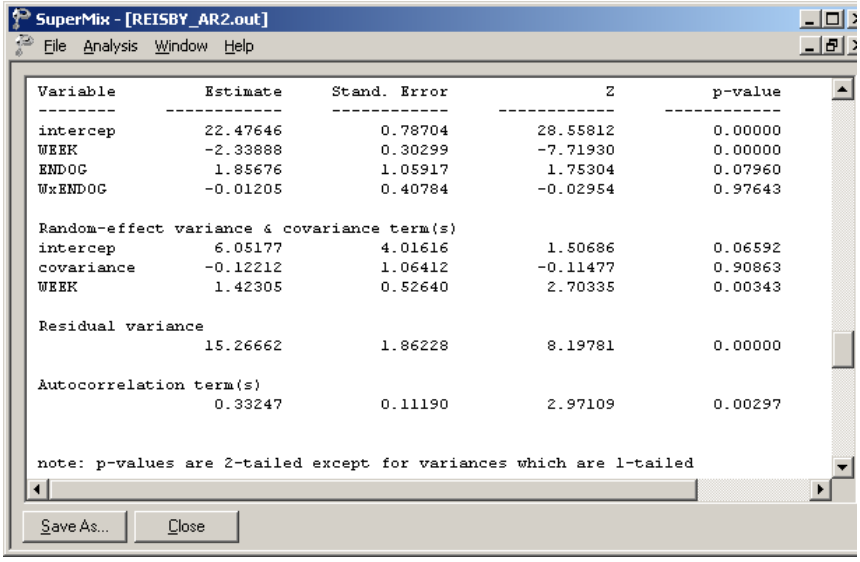
The **Starting values** section in the output file corresponds with the starting values we specified in the **Starting Values** and **Advanced** screens. The **mean** row refers to the starting values for the fixed regressors, which are intercept and WEEK in this example. The **covariates** row contains the starting values for ENDOG and WxENDOG. The elements of the **var. terms** row are the starting values for the level-2 variance/covariance matrix. The **residual** value is the starting error variance. The **auto term(s)** is the autocorrelation starting value(s).

The maximum marginal likelihood (MML) estimates

The starting values section is followed by the **Final Results**. The maximum marginal likelihood (MML) solution at convergence is obtained by first using the EM algorithm and then Fisher scoring iterations. The AIC, SBC and –2 log likelihood (deviance) are given right below the iteration information.



As shown above, the convergence is obtained after 10 EM and 10 Fisher iterations. The log likelihood value can be used to perform likelihood ratio tests.



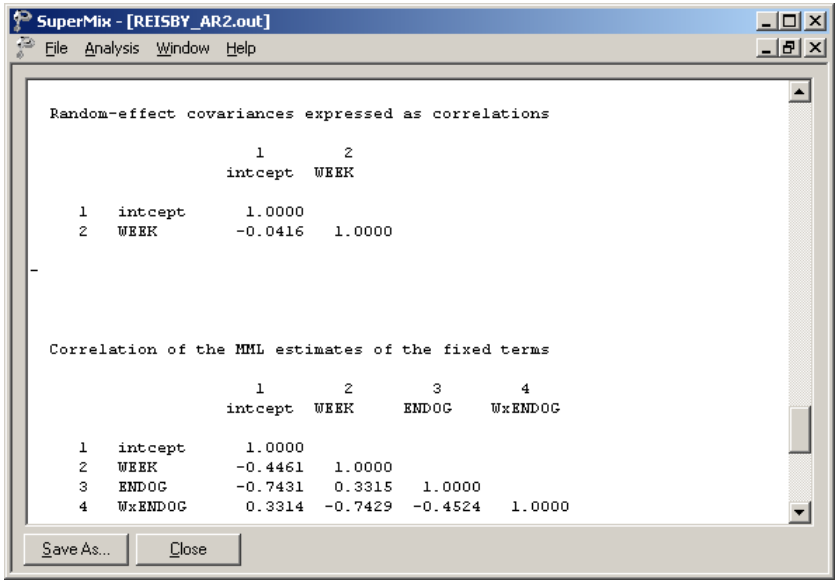
For each model parameter, maximum marginal likelihood estimates, standard errors, *z*-values, and *p*-values are provided. These *p*-values are two-tailed, except for the variance parameters where one-tailed *p*-values are given.

Considering the estimated fixed effects, the initial level of severity for non-endogenous patients is approximately 22.5 on the HDRS, while the endogenous patients start about 1.9 units higher. The difference in initial severity is almost significant ($p < 0.0790$). The reason that the intercept and endogenous effect reflect HDRS levels at week 0 is due to the coding of WEEK that was used, namely, 0 to 6. Using other codings of WEEK would change the meaning of these regression coefficients.

Both groups exhibit an overall weekly rate of improvement of roughly 2.3 units which is highly significant. In terms of the random-effect variance and covariance terms, there is a significant rate of improvement ($p < 0.00343$). The variation in patients' initial severity is marginally significant at 0.066. However, the overall covariation between those two terms are significant at a 10% level ($p > 0.90864$).

Correlation of the MML estimates

Finally, correlation matrices are also provided for the estimates of all model parameters. It is important to realize that these correlation matrices are not correlations of the variables themselves, but correlations of the estimated model parameters. These matrices may be helpful in determining the degree to which collinearity is present in terms of the model parameters.



It is interesting to note that, when the correlations are rounded to two decimal places, equalities exist between the correlations:

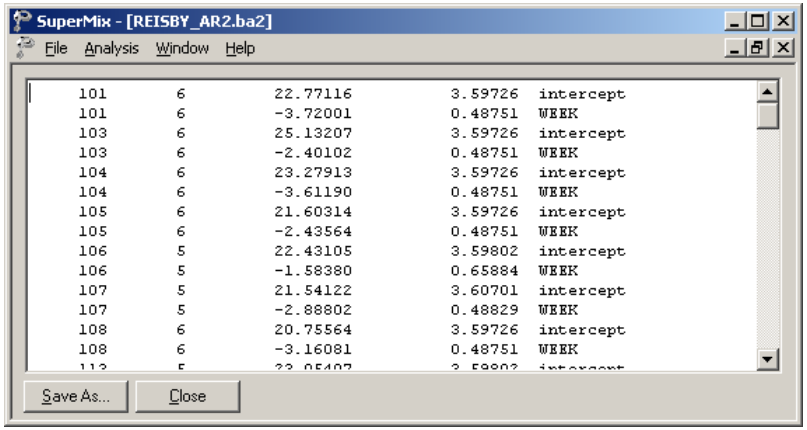
$$r(\text{INTCEPT}, \text{ENDOG}) = r(\text{WEEK}, \text{WxENDOG}) = -0.74$$
$$r(\text{INTCEPT}, \text{WEEK}) = r(\text{ENDOG}, \text{WxENDOG}) = -0.45$$
$$r(\text{ENDOG}, \text{WEEK}) = r(\text{INTCEPT}, \text{WxENDOG}) = 0.33$$

From the fixed effect results we see that the WxENDOG effect was not significant ($p = 0.976$). It is reasonable to assume that, with the interaction term omitted from the model, the correlations between the intercept, ENDOG and WEEK coefficients will be close to those reported above.

Level-2 Bayes results

The residual file **reisby_ar2.ba2** offers the opportunity to take a closer look at the results by individual patient. Select the **Analysis**, **View L-2 Bayes Results** option to open the Bayes results as shown below. The contents of this file are displayed for the first 7 patients. Two lines of information are given for each patient, containing, in order of appearance,

- the number of the patient in the data set,
- the number of the empirical Bayes coefficient,
- the empirical Bayes estimate,
- the estimated variance of the Bayes coefficient, and
- the name of the associated coefficient as used in the model.



The user can obtain patient-specific predicted HDRS scores using the empirical Bayes estimates for each patient by using the method discussed earlier in Section 3.2.5.5.

Model comparison

In Table 3.5, the estimates of the regular model without an autoregressive term and the non-stationary AR(1) are summarized. Note that the AIC and BIC values obtained from the AR(1) model were multiplied by – 2.0 in order to facilitate comparison over the models.

TABLE 3.5: Comparison of models with and without AR(1) term

	no AR term	Non-stationary AR(1)
intcept	22.47626	22.47646
	(0.79435)	(0.78704)
WEEK	–2.36569	–2.33888
	(0.31181)	(0.30299)
ENDOG	1.98802	1.85677
	(1.06905)	(1.05917)
WxENDOG	–0.02706	–0.01205
	(0.41947)	(0.40784)
Log Likelihood	–1107.4646	–1103.72
Akaike's Information Criterion	2214.9292	2217.44
Schwarz's Bayesian Criterion	2230.9292	2237.076
–2 Log Likelihood	2248.4465	2207.441
Number of free parameters	8	9

We notice that the estimates of both models are close to each other. The estimated variances of the non-stationary AR(1) model are smaller for all the parameters. The deviance is 2248.4465 – 2207.441 = 41.0055 with 1 degree of freedom, which is highly significant. Thus, we conclude that in this example, the non-stationary AR(1) model fits the data better.

3.3 Models based on the TVSFP data

3.3.1 The data

The data set used here is from the Television School and Family Smoking Prevention and Cessation Project (TVSFP) (Flay *et. al.*, 1988). The study was designed to test independent and combined effects of a school-based social-resistance curriculum and a television-based program in terms of tobacco use and cessation. The data from the study included a total of 1,600 students from 135 classrooms drawn from 28 schools.

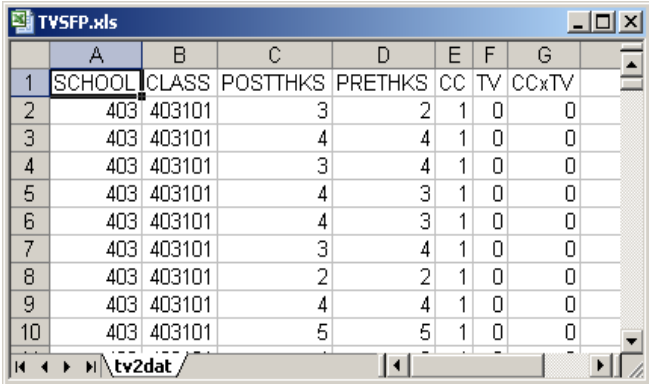
Schools were randomized to one of four study conditions:

- a social-resistance classroom curriculum
- a media (television) intervention
- a social-resistance classroom curriculum combined with a mass-media intervention, and
- a no-treatment control group

A tobacco and health knowledge scale (THKS) was used in classifying subjects as knowledgeable or not. In its original form, the student's score was defined as the number of correct answers to seven items on tobacco and health knowledge.

While the structure of this study indicates a three-level hierarchical structure, the present application uses these data to fit a two-level model, with students nested within either classes or schools, in order to present an introduction to the analysis of ordinal outcomes. A 3-level model is presented in Sections 3.3.4 and 3.3.5.

Data for the first 10 students on most of the variables used in this section are shown below in the form of an SuperMix spreadsheet file, named **TVSFP.ss3**.



TVSFP.xls

	A	B	C	D	E	F	G
1	SCHOOL	CLASS	POSTTHKS	PRETHKS	CC	TV	CCxTV
2	403	403101	3	2	1	0	0
3	403	403101	4	4	1	0	0
4	403	403101	3	4	1	0	0
5	403	403101	4	3	1	0	0
6	403	403101	4	3	1	0	0
7	403	403101	3	4	1	0	0
8	403	403101	2	2	1	0	0
9	403	403101	4	4	1	0	0
10	403	403101	5	5	1	0	0

tv2dat

The variables of interest are:

- SCHOOL indicates the school a student is from (28 schools in total).
- CLASS identifies the classroom (135 classrooms in total).
- POSTTHKS represents the post-intervention tobacco and health knowledge scale. It is treated as a continuous variable in the examples in this chapter. See Sections 4.2 and 6.2 for examples where POSTTHKS is treated as a binary or ordinal outcome.
- PRETHKS indicates the pre-intervention THKS score.
- CC is a binary variable indicating whether a social-resistance classroom curriculum was introduced, where 0 indicates "no" and 1 "yes."
- TV is an indicator variable for the use of media (television) intervention, with a "1" indicating the use of media intervention, and "0" the absence thereof.
- CCxTV was constructed by multiplying the variables TV and CC, and represents the CC by TV interaction.

3.3.1.1 Exploring the data

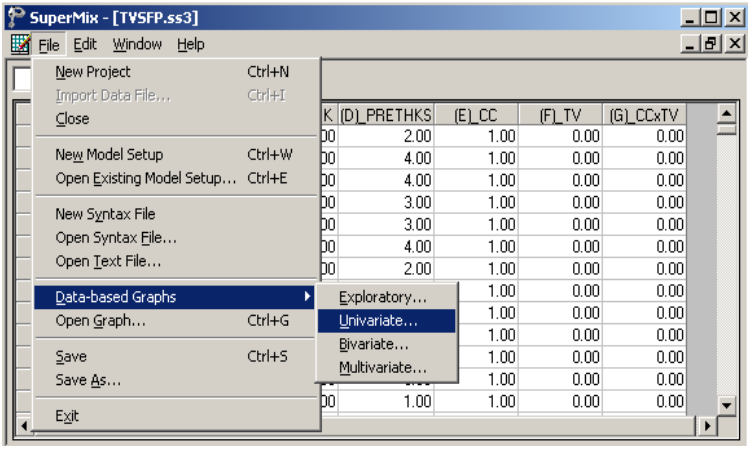
In this section, a univariate bar chart and a bivariate box-and-whisker plot are given. More information on other types of plots available are given in Chapter 4 of the SuperMix primer.

Univariate graphs

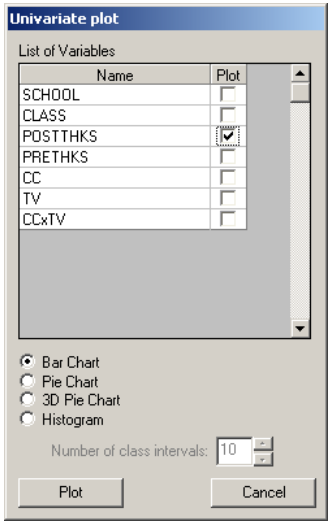
The pop-up menu below shows the data-based graphing options currently available in SuperMix. As a first step, we will take a closer look at the distribution of the total post-intervention scores (POSTTHKS), which is the potential dependent variable in this study. While scores such as these are not truly continuous variables, they are often treated as if they were.

Bar chart

To do so, select the **Univariate** option from the **Data-based Graphs** menu as shown below.



The **Univariate plot** dialog box appears. Select the variable POSTTHKS and indicate that a **Bar Chart** is to be graphed. Click the **Plot** button to display the bar chart.



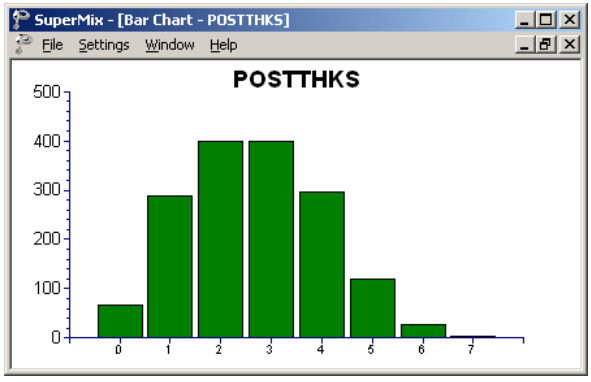


Figure 3.11: Bar chart of POSTTHKS scores

The bell-shaped bar chart above shows that the variable POSTTHKS is approximately normally distributed. Note that histograms are usually used for the depiction of the distribution of a continuous variable.

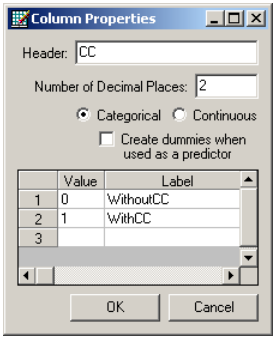
Bivariate graphs

It is hoped that the social-resistance classroom curriculum (CC), the television intervention (TV) and the CC and TV interaction combination (CCxTV) would affect the tobacco and health knowledge (POSTTHKS). Before we start with the model, we would like to show a box-and-whisker plot of POSTTHKS for each category of CC.

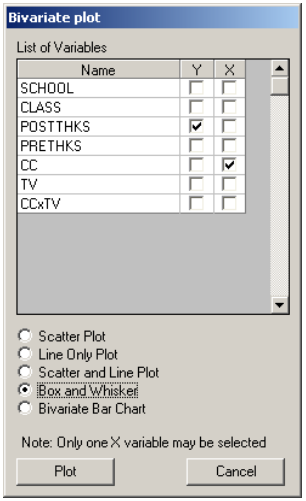
Box-and-whisker plots

A box-and-whisker plot is useful for depicting the locality, spread and skewness of variables in a data set and may be used to examine the distributions of continuous variables, such as for the different values of discrete valued predictors. This option is accessed via the **Data-based Graphs**, **Bivariate** option on the **File** menu.

To assign labels to the categories of CC, right-click on the CC column in the spreadsheet and select **Column Properties**. On the **Column Properties** dialog box, select the **Nominal** option and assign the appropriate labels and save the data file.



The **Bivariate plot** dialog box is completed as shown below: select the outcome variable POSTTHKS as the **Y**-variable of interest, and the predictor CC to be plotted on the **X**-axis. Check **the Box and Whisker** option, and click **Plot**.



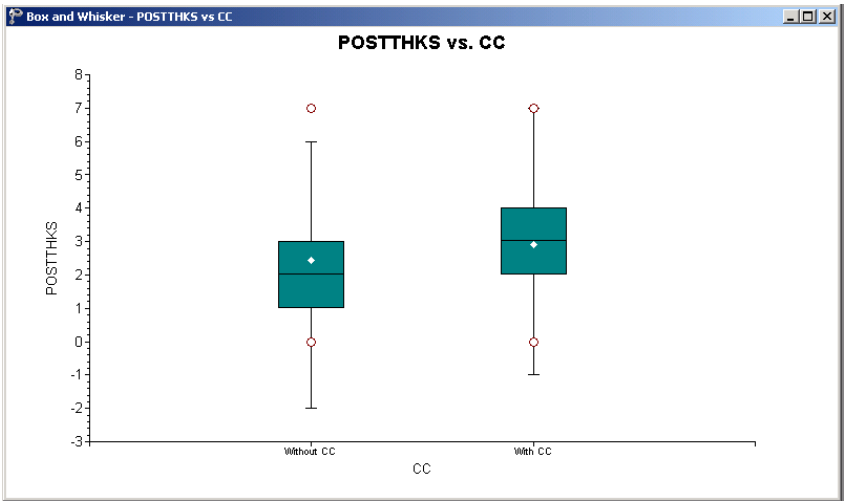


Figure 3.12: Box-and-whisker plots of POSTTHKS scores for different CC values

The bottom line of a box represents the first quartile (q_1), the top line the third quartile (q_3), and the in-between line the median (me). The arithmetic mean is represented by a diamond. Here, the mean of POSTTHKS is lower in the group without the social-resistance classroom curriculum (CC). The box-and-whisker plot indicates a positive relationship between CC and POSTTHKS.

3.3.2 A 2-level random intercept model using classroom as level-2 ID

3.3.2.1 The model

The first model fitted to the data explores the cluster effects of each classroom on the outcome. The mixed model can be expressed as

$$\text{POSTTHKS}_{ij} = \beta_0 + \beta_1 \text{CC}_i + \beta_2 \text{TV}_i + \beta_3 (\text{CC}_i \times \text{TV}_i) + v_{0i} + e_{ij},$$

where v_{0i} represents the classroom influence on POSTTHKS. To understand the model better, we can rewrite the model in the following way. The level-1 or within-cluster model is shown below.

Level-1 model: $(j = 1, ..., n_i)$

$$\text{POSTTHKS}_{ij} = b_{0i} + e_{ij},$$

$$e_{ij} : \ NID(0, \sigma^2)$$

The level-1 model estimates POSTTHKS as a function of the intercept b_{0i} and error term e_{ij}. Subscript i denotes the subscript for classroom, while subscript j refers to the student j. n_i is used to denote the number of students in each classroom. Because we have different numbers of students in different classrooms, n_i also varies. In this data set, $1 \le n_i \le 28$.

The level-2, or between-cluster, model describes the intercept b_{0i} as a function of cluster characteristics.

Level-2 model: $(i = 1, \text{L} \ , N)$

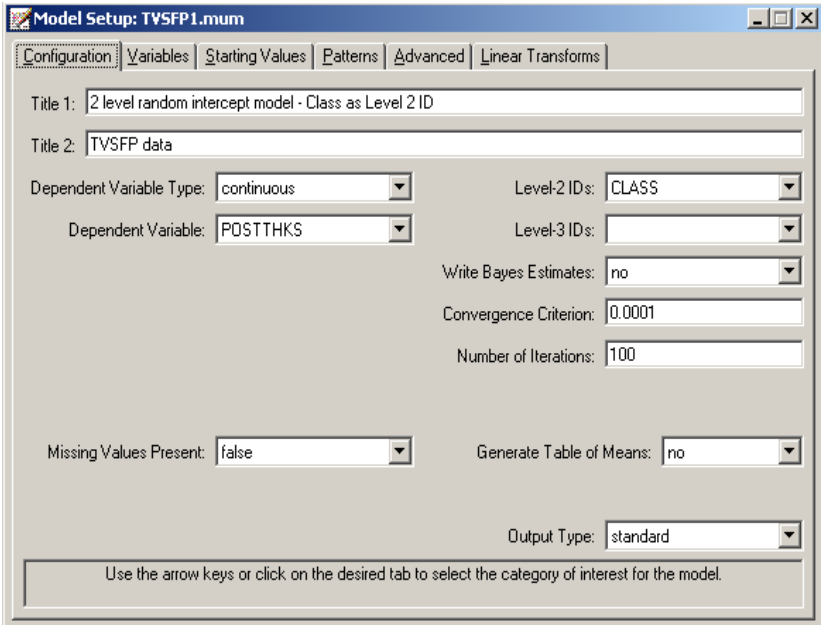
$$b_{0i} = \beta_0 + \beta_1 \text{CC}_i + \beta_2 \text{TV}_i + \beta_3 (\text{CC}_i \times \text{TV}_i) + v_{0i}$$

$$v_{0i} : NID(0, \sigma_v^2)$$

As shown above, the intercept b_{0i} is estimated as a function of the population average β_0, the covariates CC_i, TV_i, and $\mathrm{CC}_i \times \mathrm{TV}_i$, and the classroom difference v_{0i}. The coefficient v_{0i} represents the amount that unit *i* deviates from the average β_0, after controlling for the effects of the covariates included. The level-2 residual v_{0i} is assumed to follow $NID(0, \sigma_v^2)$ for all the *i*s. If $v_{0i} = 0$ for all i, which implies $\sigma_v^2 = 0$, the model is the same as the ordinary regression model.

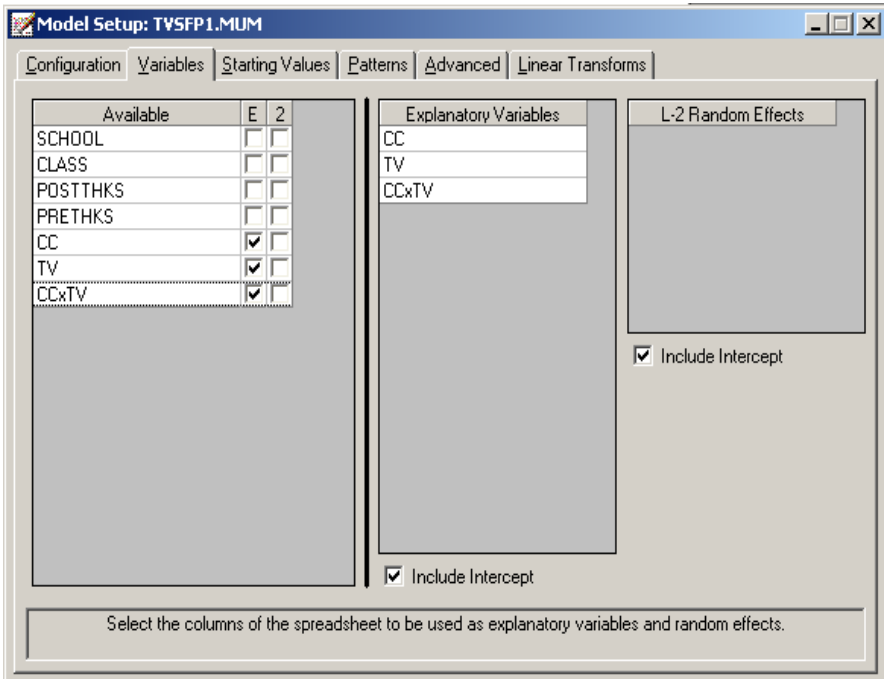
3.3.2.2 Setting up the analysis

Open the SuperMix spreadsheet **TVSFP.ss3** used during the exploratory analysis discussed previously in this chapter. 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.



Select the continuous outcome variable POSTTHKS from the **Dependent Variable** drop-down list box. Select the classroom number CLASS from the **Level-2 IDs** drop-down list box. Enter a title for the analysis in the **Title** text boxes. In this example, default settings for all other options associated with the **Configuration** screen are used.

Proceed to the **Variables** screen by clicking on that tab. The **Variables** screen is used to specify the fixed and random effects to be included in the model. Select the explanatory (fixed) variables using the **E** check boxes next to the variables names in the **Available** grid at the left of the screen. Note that, as the variables are selected, the selected variables are listed in the **Explanatory Variables** grid. After selecting all the explanatory variables, the screen shown below is obtained. 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.



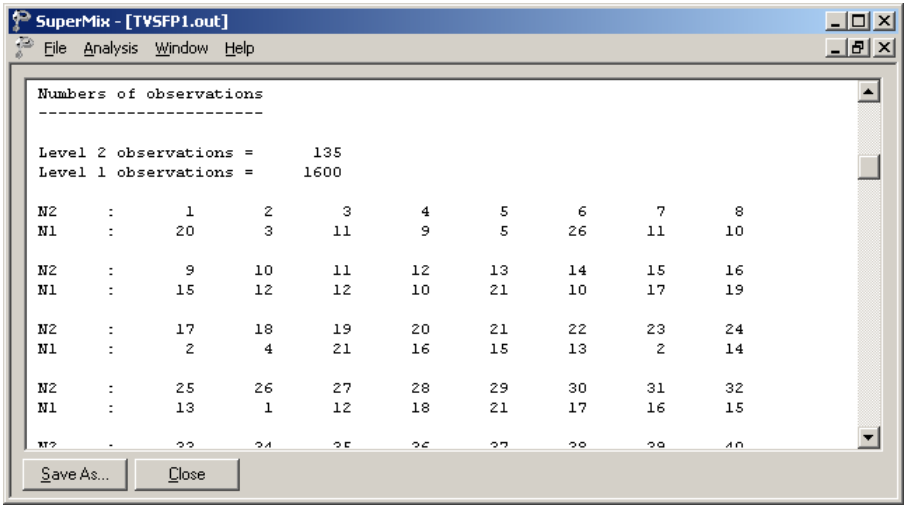
Next, specify the random effects at level 2 the hierarchy. In this example, we want to fit a model with random intercepts at level 2. By default, the **Include Intercept** check box in the **L-2 Random Effects** grid is checked. If this box is left checked, and no additional random effects are indicated using the **2** column in the **Available** grid to the left, the model fitted will be the random-intercepts-only model we intend to use. No further changes on this screen are necessary.

Before running the analysis, the model specifications have to be saved. Select the **File**, **Save As** option, and provide a name (**TVSFP1.mum**) for the model specification file. Run the analysis by selecting the **Run** option from the **Analysis** menu.

3.3.2.3 Discussion of results

Model and data description

In the **numbers of observations** section, a summary of the hierarchical structure is provided.

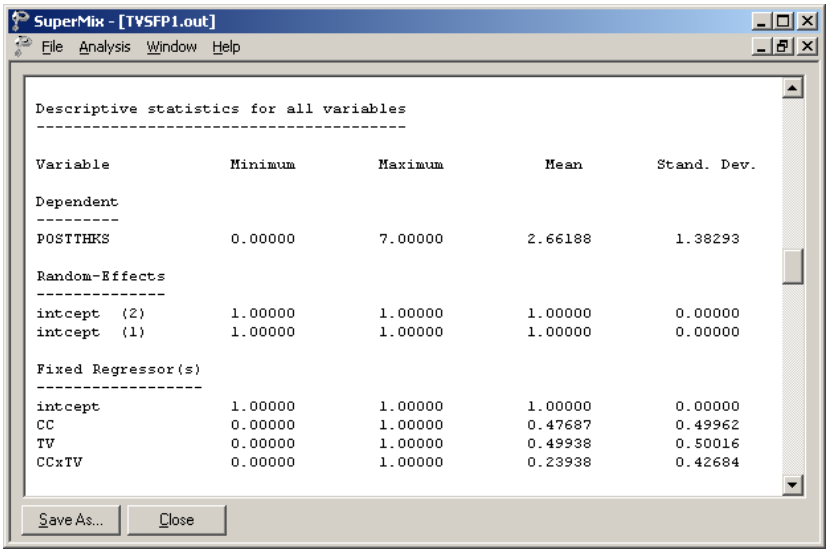


As shown above, data from a total of 1600 students within 135 classrooms were included at levels 2 and 1 of the model. This corresponds to the study design

described earlier. In addition, a summary of the number of students nested within each classroom is provided. The classroom with N2 = 6, for example, had 26 students (N1: 26). By contrast, classroom 26 had only 1 student.

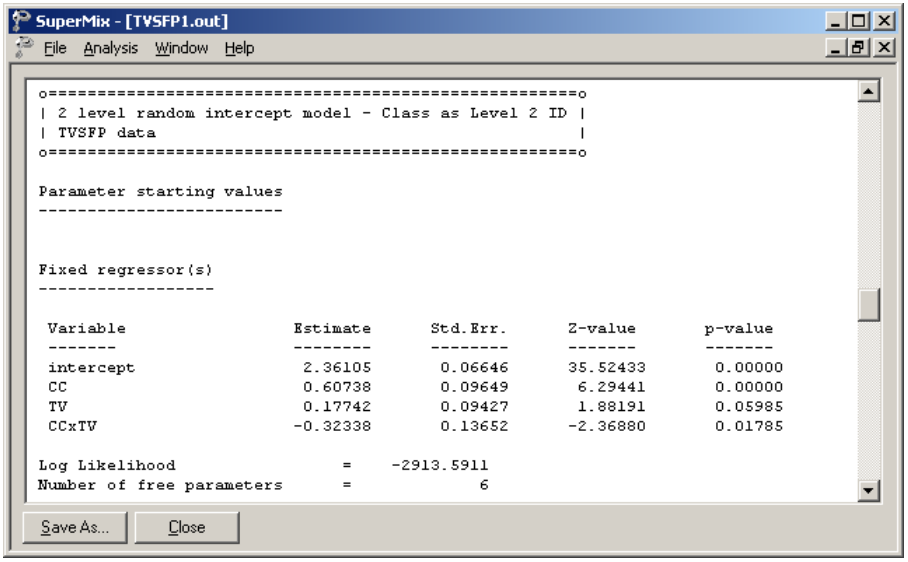
Descriptive statistics and starting values

Next, the **descriptive statistics for all variables** are given. The minimum value, maximum value, mean and standard deviation are given for all the variables included in the model. For example, the mean POSTTHKS is 2.6618 with a standard deviation of 1.38293.

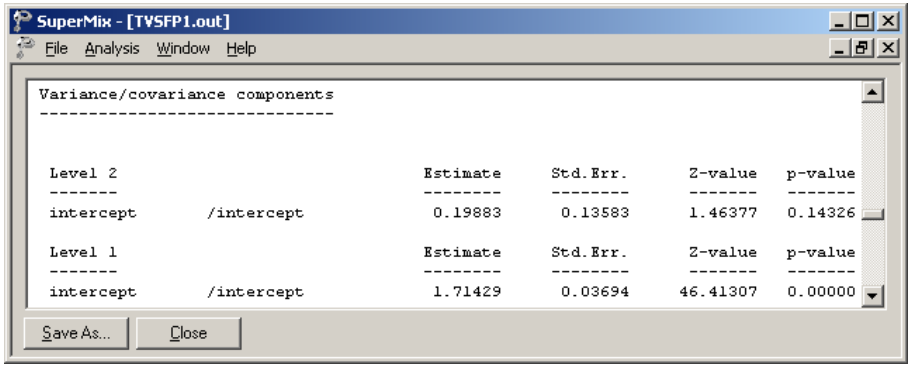


Starting values – OLS estimates

The starting values for the **fixed regressor(s)** are shown below. The **log likelihood** value and **number of free parameters** of the OLS regression are given in this part of the output.



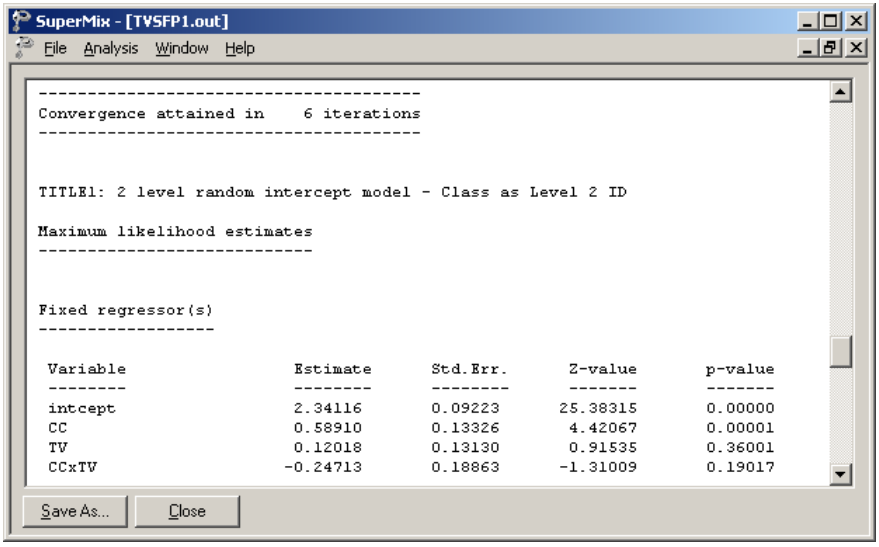
After the **number of free parameters**, the starting values of **variance/covariance components** are reported as shown.



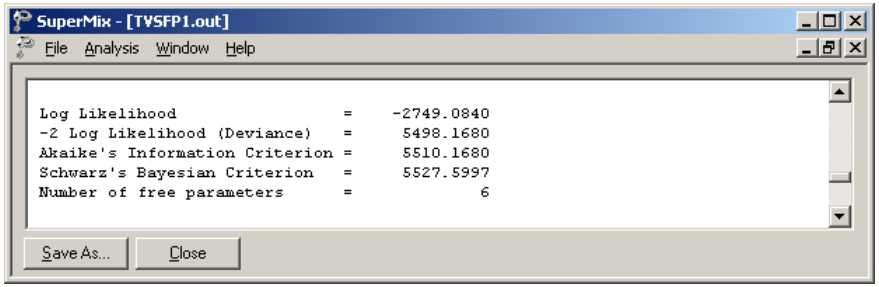
Fixed effects estimates

The number of iterations needed to obtain convergence is given after the starting values. The output describing the estimated **fixed regressor(s)** after convergence is shown next.

As shown below, the estimates for CC and TV are both positive. On average, a social-resistance classroom curriculum can improve the tobacco and health knowledge by 0.58910, and television intervention can increase the POSTTHKS score by 0.12018. However the estimate of CCxTV is negative, which implies that the students who had both CC and TV are expected to show a decrease of 0.24713 in their POSTTHKS score. The estimates associated with intercept and TV are highly significant, but estimates of the other two coefficients are not statistically significantly different from zero.

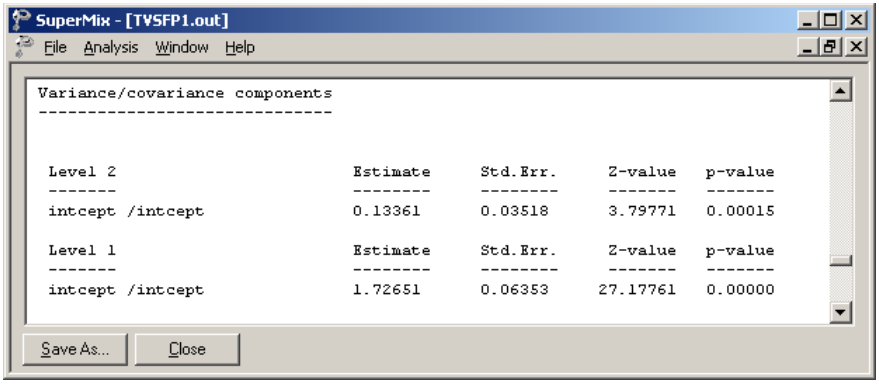


The estimates for the fixed regressors and model fit statistics are given next. For more information on these statistics, see Section 3.1.2.3.



Random effect estimates

The estimates for the random part of the model are reported next. The variation in the average estimated intercept at level 2 is highly significant, which indicates that the classroom difference in intercepts does help to explain the variation in POSTTHKS scores.



The covariance and correlation matrix of level-2 and level-1 random effects are given in matrix format at the end of the output file. These values are the same as the estimates of variance/covariance components as shown above.

3.3.2.4 Interpreting the results

Estimated outcomes for different groups

For a student who participated in neither social-resistance classroom curriculum nor television intervention (CC = 0; TV = 0), the expected POSTTHKS is equal to just the intercept 2.36105. For a student who participated in both programs (CC = 1; TV = 1; CCxTV = 1), the predicted POSTTHKS is calculated as follows:

$$\begin{aligned}\widehat{\text{POSTTHKS}}_{ij} &= \hat{\beta}_0 + \hat{\beta}_1 \text{CC}_i + \hat{\beta}_2 \text{TV}_i + \hat{\beta}_3(\text{CC}_i \times \text{TV}_i) \\ &= 2.34116 + 0.5891 + 0.12018\text{-}0.24713 \\ &= 2.80331\end{aligned} \quad .$$

Fit statistics and % variation explained

An estimate of the percentage of variation in the outcome at classroom level is obtained as

$$\frac{0.13361}{0.13361 + 1.72651} \times 100\% = 7.18\%$$

indicating that about 7.18% of the total variance lies between the clusters/classrooms and that 92.82% of the variance remains at the student level.

3.3.3 2-level random intercept model by using school as level-2 ID

The model in the previous section shows that only about 7% of the total variation in outcome is at the classroom level. The question that arises is whether clustering within schools may provide a better explanation of the way in which post-intervention scores vary. In this section, the model is fitted using SCHOOL, rather than classroom, as the level-2 ID.

3.3.3.1 The model

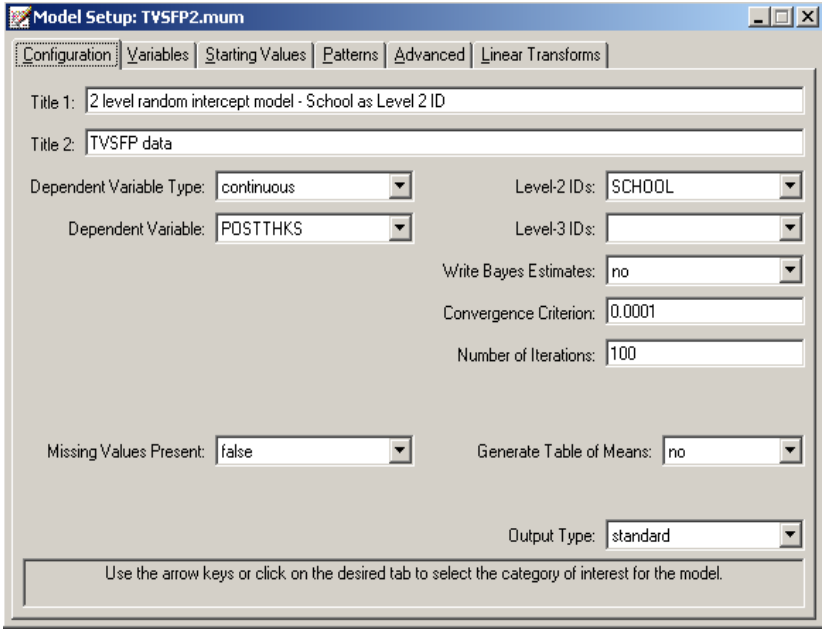
The mathematical equation of the model to be fitted is exactly the same as for the previous model.

$$\text{POSTTHKS}_{ij} = \beta_0 + \beta_1 \text{CC}_i + \beta_2 \text{TV}_i + \beta_3 (\text{CC}_i \times \text{TV}_i) + v_{0i} + e_{ij},$$

The difference here is in the meaning of the subscript *i*. In the previous model, we used *i* to refer the classroom. However, the *i*s here refer to the schools.

3.3.3.2 Setting up the analysis

To create the model specifications for this model, we start by opening **TVSFP.ss3** in a SuperMix spreadsheet window.



We use the **Open Existing Model Setup** option on the **File** menu to load the **Model Setup** window for **TVSFP1.mum**. Click on **File**, **Save as** to save the model setup in a

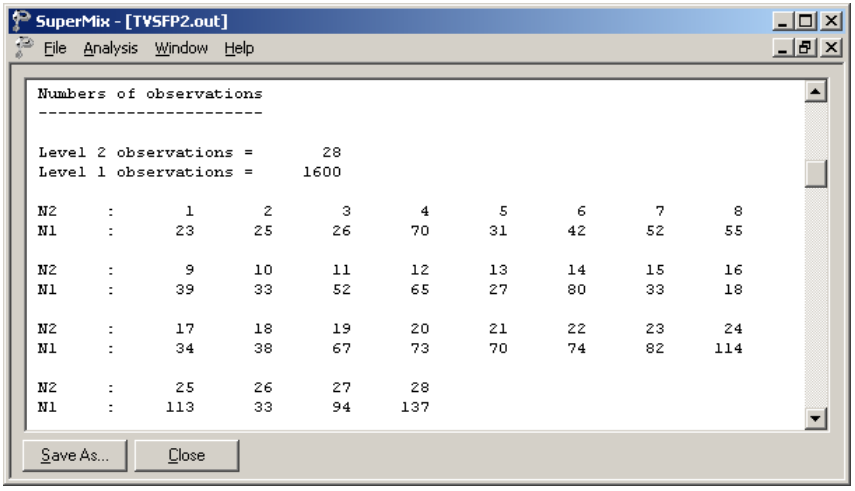
new file, such as **TVSFP2.mum**. Next, change the string in the **Title 1** text box on the **Configuration** screen, and select SCHOOL as the **Level-2 ID** as shown above.

Keep all the other settings unchanged. Save the changes to the file **TVSFP2.mum** and select the **Run** option on the **Analysis** menu to produce the output file **TVSFP2.out**.

3.3.3.3 Discussion of results

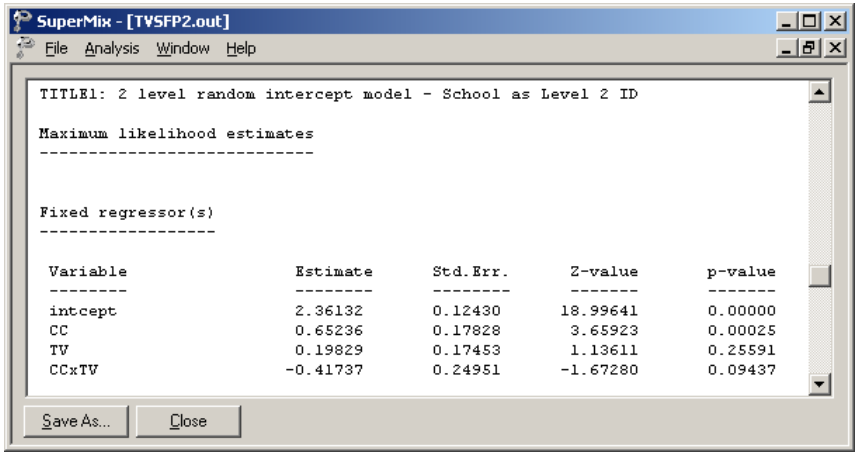
Model and data description

The **number of observations** section clearly shows that the data set contains 28 schools and each school has between 18 and 137 students as shown below.

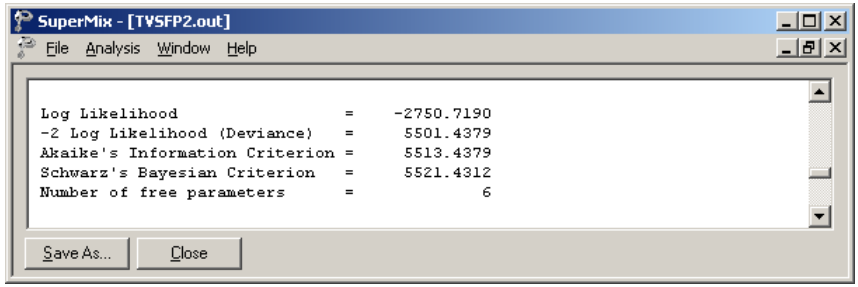


Fixed effects estimates and descriptive statistics

The estimates for the fixed estimates as shown below are close to the estimates in the previous example, but not exactly the same. For example, the estimate for CC increased by 0.06326 (0.65236 – 0.58910 = 0.06326), and the estimate for the effect of television intervention is about 0.07811 higher when using school as the level-2 ID. However, the estimate of the interaction of CC and TV is about 0.17 lower.

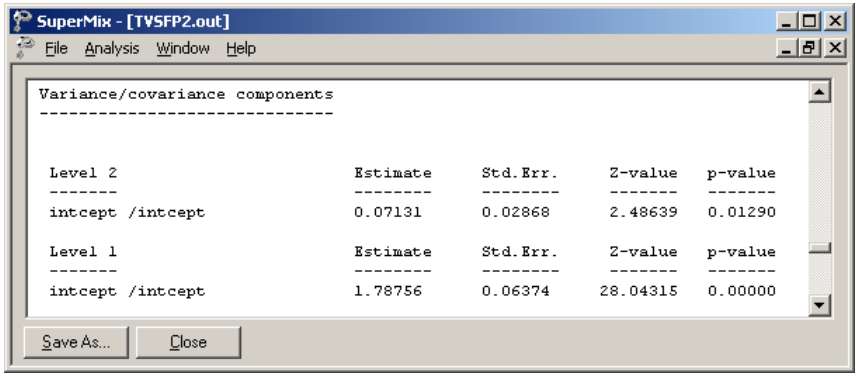


Both the deviance and Akaike information criterion (AIC) are slightly higher than the previous model. The SBC is smaller.



Random effect estimates and covariance/correlation matrices

The estimates for the random part of the model are reported next.



The variation in the average estimated intercept at level 2 is highly significant, which indicates that the difference in school intercepts also explains the variation of POSTTHKS scores. Similarly, we can calculate that about 3.84% of the total variance can be explained by the school difference:

$$\frac{0.07131}{0.07131+1.78756}\times 100\% = 3.84\%.$$

3.3.4 A 3-level random intercept model using class and school as IDs

The previous two models show that both school and classroom contribute to the explanation of the total variation of the POSTTHKS scores. We now construct a three-level model that uses both CLASS and SCHOOL as level-2 and level-3 IDs.

3.3.4.1 The model

The level-1 and level-2 models are the same as the previous two models, as shown below.

Level-1 model $(k = 1,\ldots,n_{ij})$

$$\text{POSTTHKS}_{ijk} = b_{0ij} + e_{ijk},$$

$$e_{ijk} : NID(0,\sigma^2)$$

Level-2 model $(j = 1,...,n_i)$

$$b_{0ij} = b_{0i} + b_{1i}\mathrm{CC}_{ij} + b_{2i}\mathrm{TV}_{ij} + b_{3i}(\mathrm{CC}_{ij} \times \mathrm{TV}_{ij}) + v_{0ij}$$

$$v_{0ij} : \ NID(0, \sigma^2_{v(2)})$$

Level-3 model $(i = 1,...,N)$

$$\begin{aligned} b_{0i} &= \beta_0 + v_{0i} \\ b_{1i} &= \beta_1 \\ b_{2i} &= \beta_2 \\ b_{3i} &= \beta_3 \end{aligned}$$

$$v_{0i} : \ NID(0, \sigma^2_{v(3)})$$

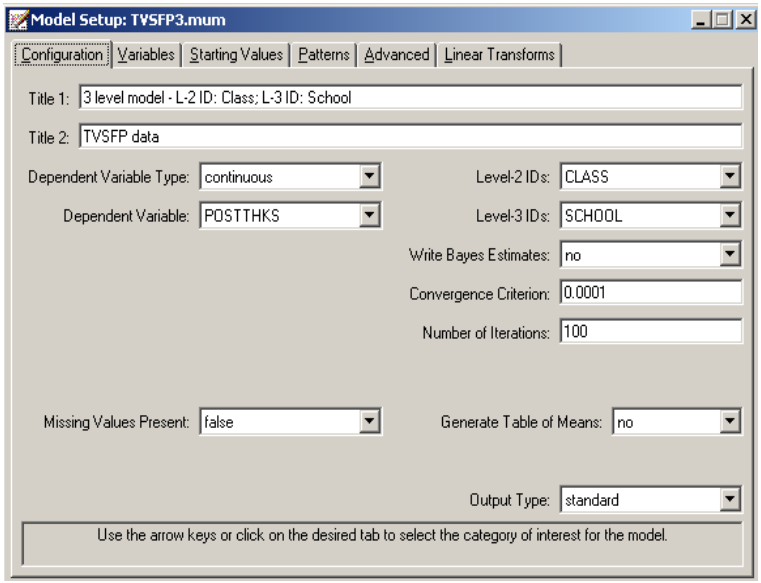
In this mixed model the intercept b_{0ij} is estimated by a level-2 equation. It indicates that classroom *j*'s initial value is not only determined by the population average b_{0i}, but also by the classroom difference v_{0ij}. The level-2-intercept b_{0ij} is estimated by a level-3 equation which takes the school difference v_{0i} into consideration, where *i* denotes the school ID.

The above model can also be written in the following format.

$$\mathrm{POSTTHKS}_{ijk} = \beta_0 + \beta_1\mathrm{CC}_{ij} + \beta_2\mathrm{TV}_{ij} + \beta_3(\mathrm{CC}_{ij} \times \mathrm{TV}_{ij}) + v_{0ij} + v_{0i} + e_{ijk}\,.$$

3.3.4.2 Setting up the analysis

We modify our model setup saved to the syntax file **TVSFP1.mum** by first using the **Open Existing Model Setup** option on the **File** menu of the **TVSFP.ss3** window to retrieve the syntax file. Then click on **File**, **Save as** to save the model setup in a new file, such as **TVSFP3.mum**.

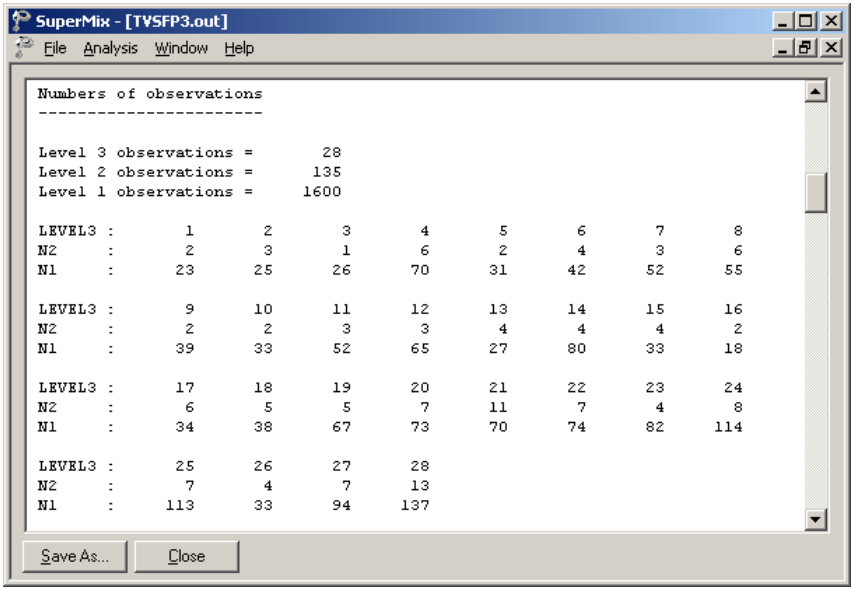


Next, we change the string in the **Title 1** text box on the **Configuration** screen, and select SCHOOL as the **Level-3 ID** as shown below. We now have both level-2 and level-3 IDs selected. Keep all the other settings unchanged. Save the changes to the file **TVSFP3.mum** and select the **Run** option on the **Analysis** menu to produce the output file **TVSFP3.out**.

3.3.4.3 Discussion of results

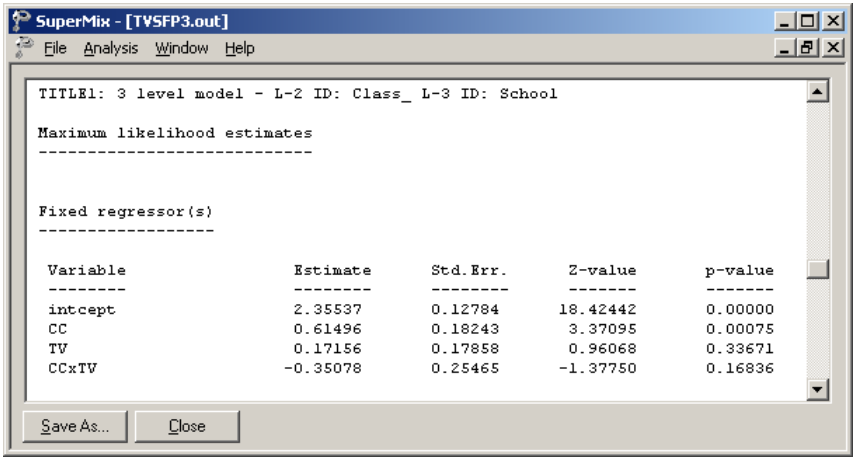
Model and data description

The **number of observations** section clearly shows the hierarchical structure of the data. The data contains 1600 students from 135 classes nested in 28 schools. In school number 20 (LEVEL 3: 20), the data of 73 students (N1: 73) from 7 (N2: 7) classes are present in this data set.

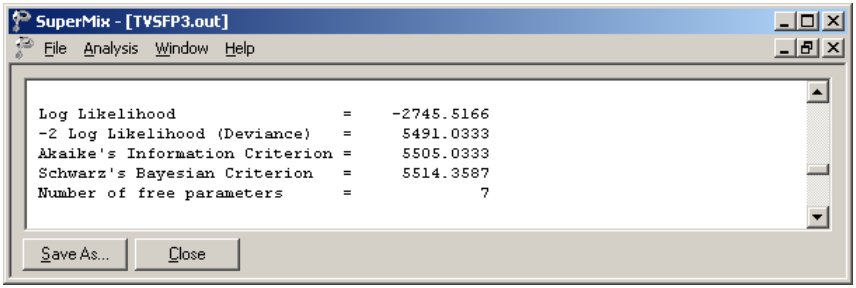


Fixed effects estimates

As shown below, the estimates are not markedly different from the estimates of the previous two models.

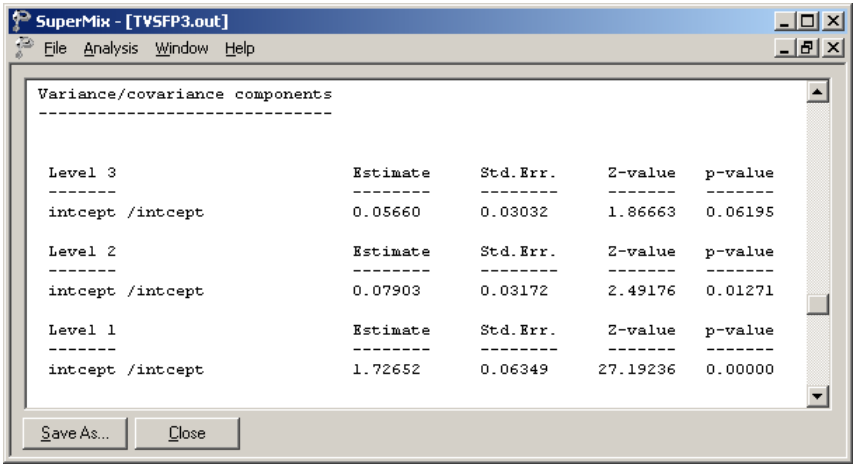


Both the deviance and Akaike information criterion (AIC) are slightly higher than the previous model. The SBC is smaller.



Random effect estimates

The estimates for the random part of the model are reported next.



The estimated level-2 random effect is highly significant (p = 0.08), but the level-3 is not (p = 0.06).

3.3.4.4 Interpreting the results

Fit statistics and % variation explained

The variation of POSTTHKS scores can be explained by individual differences, classroom differences and school differences.

For schools,

$$\frac{0.05660}{0.05660+0.07903+1.72652}\times 100\% = 3.04\%,$$

while for classrooms

$$\frac{0.07903}{0.05660+0.07903+1.72652}\times 100\% = 4.24\%.$$

As calculated above, the school difference contributes 3.04% to the explanation of the total variance in the outcome, and classroom difference contributes 4.24%. The rest, 92.72% of the variation, is explained by the student differences.

Comparison of models

In Table 3.6 the estimates of the previous three models and OLS in this chapter are summarized. The three-level estimates all lie between the corresponding two level-2 estimates.

Table 3.6: Comparison of OLS and mixed model results

	OLS Estimates	Mixed Model		
		L-2 model		L-3 model
		L-2 ID: CLASS	L-2 ID: SCHOOL	L-2 ID: CLASS L-3 ID: SCHOOL
intcept	2.361	2.341	2.361	2.355
	(0.066)	(0.092)	(0.124)	(0.128)
CC	0.607	0.589	0.652	0.615
	(0.096)	(0.133)	(0.178)	(0.182)
TV	0.177	0.120	0.198	0.172
	(0.094)	(0.131)	(0.175)	(0.179)
CCxTV	–0.323	–0.247	–0.417	–0.351
	(0.137)	(0.189)	(0.250)	(0.255)
Deviance		5498.168	5501.438	5491.033
AIC		5510.168	5513.438	5505.033
SBC		5527.600	5521.431	5514.359
Number of free parameters		6	6	7

3.3.5 A 3-level random intercept model including pre-THKS score

The PRETHKS variable indicates the observed score before implementation of intervention. It might have an impact on the POSTTHKS scores. In this section, a three-level model including the PRETHKS as predictor is fitted.

3.3.5.1 The model

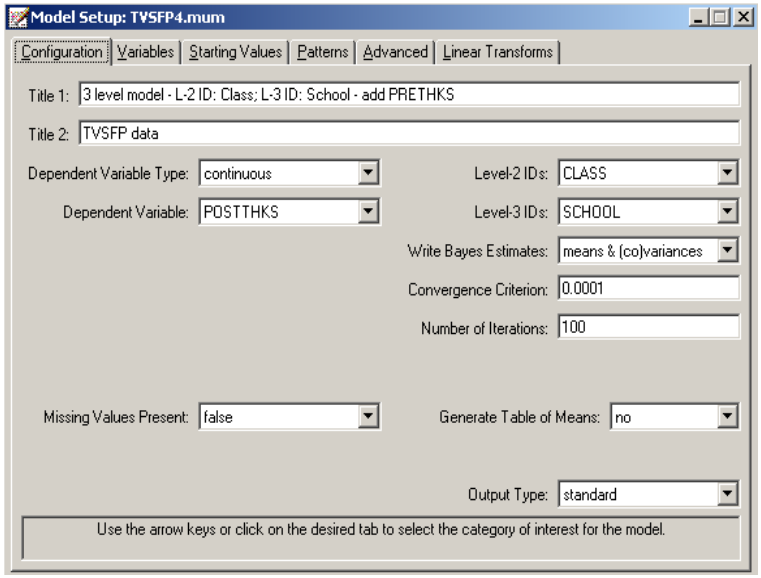
The only difference between this model and the previous one is the addition of the variable PRETHKS as a covariate:

$$\text{POSTHKS}_{ijk} = \beta_0 + \beta_1 \text{CC}_{ij} + \beta_2 \text{TV}_{ij} + \beta_3 (\text{CC}_{ij} \times \text{TV}_{ij}) + \beta_4 \left(\text{PRETHKS}_{ijk} \right) + v_{0ij} + v_{0i} + e_{ijk}.$$

From the subscripts associated with the coefficients, we note that while CC, TV and CCxTV were measured at a classroom level, the pre-intervention score PRETHKS is measured on the individual level. Such a variable may also be referred to as a level-1 predictor, while CC, TV and CCxTV may be called level-2 predictors, covariates, or mediating effects.

3.3.5.2 Setting up the analysis

The easiest way to set up this model is to modify the model setup in the syntax file **TVSFP3.mum** by first using the **Open Existing Model Setup** option on the **File** menu. Then click on **File**, **Save as** to save the model setup in a new file, such as **TVSFP4.mum**.



Next, we change the string in the **Title 1** text box on the **Configuration** screen. Notice that we would like to request Bayes estimates as part of the program output. To do

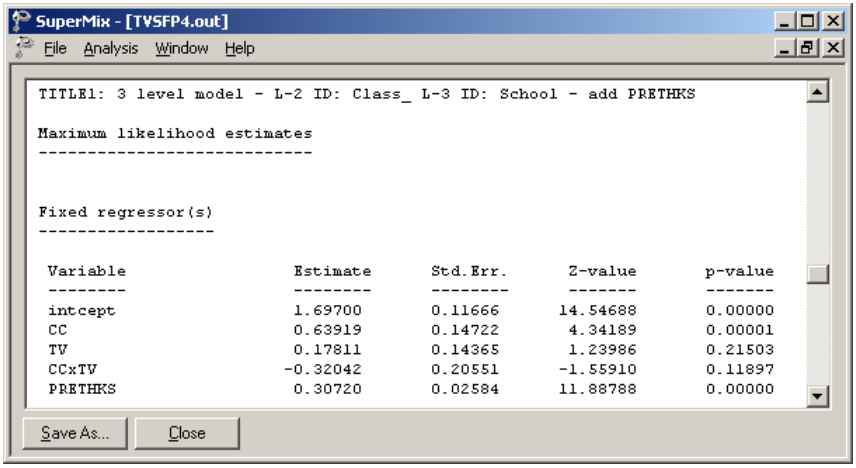
so, select **means & (co)variances** option from the **Write Bayes Estimates** drop down list as shown above.

Click on the **Variables** tab and select PRETHKS as an additional **Explanatory Variable** by checking the corresponding **E** check box. Save the changes to the file **TVSFP4.mum** and select the **Run** option on the **Analysis** menu to produce the output file **TVSFP4.out**.

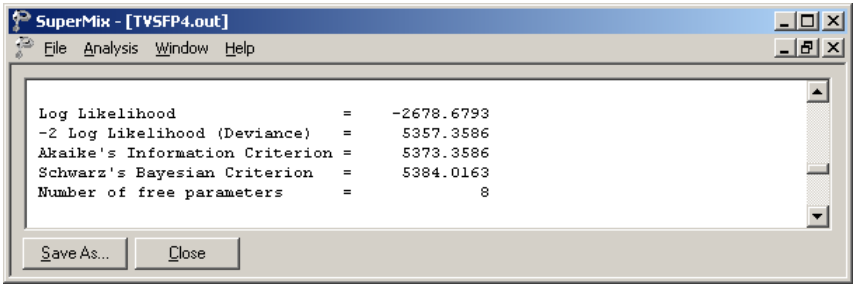
3.3.5.3 Discussion of results

Fixed effects estimates and descriptive statistics

As shown below, the estimated coefficient of PRETHKS is highly significant. The estimate of the intercept coefficient decreased because part of the variation in the intercept can now be explained by PRETHKS.

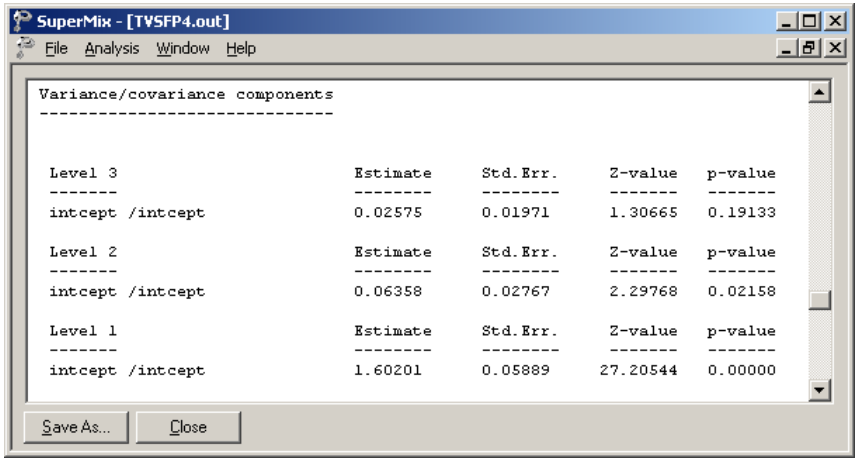


The fit statistics are given below. A comparison of these two three-level examples will be given in the next section.



Random effect estimates

The third-level random intercept estimate is not significant at a 5% level of significance, which implies that after taking PRETHKS into account, the school differences are not significant.



3.3.5.4 Interpreting the results

Estimated outcomes for different groups

For example, if a typical student who only participated in television intervention had a PRETHKS score of 2 (CC = 0; TV = 1; CCxTV = 0), the expected POSTTHKS score is calculated as follows:

$$\begin{aligned}\widehat{\text{POSTTHKS}}_{ijk} &= \hat{\beta}_{00} + \hat{\beta}_{02}\,\text{TV}_{ij} + \hat{\beta}_{04}\left(\text{PRETHKS}_{ijk}\right) \\ &= 1.697 + 0.17811 + 2 \times 0.3072 \\ &= 2.48951.\end{aligned}$$

ICCs and R square

ICCs

The so-called ICC (interclass correlation) measures the proportion of variation in the outcome variable between units at the different levels. It is occasionally referred to as the cluster effect, and is defined as the ratio of the between-cluster variance to the total variance. From the output for the random effects, we have

$$\begin{aligned}&\text{Level-1: estimated}\left(\text{error var}\right) = 1.6020 \\ &\text{Level-2: estimated}\left(\text{class var}\right) = 0.0636 \\ &\text{Level-3: estimated}\left(\text{school var}\right) = 0.0258.\end{aligned}$$

Based on this information, we can calculate the ICCs as shown below.

Similarity of students within the same school:

$$ICC = \frac{0.0258}{1.6020 + 0.0636 + 0.0258} = 0.0153$$

Similarity of students within the same classrooms (and schools):

$$ICC = \frac{0.0636 + 0.0258}{1.6020 + 0.0636 + 0.0258} = 0.0529$$

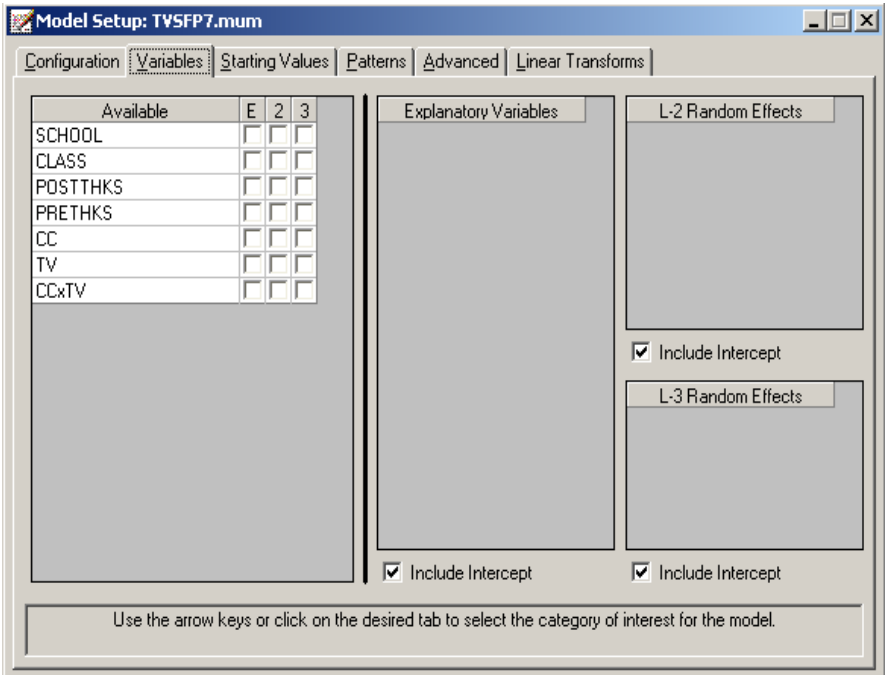
Similarity of classes within the same school:

$$ICC = \frac{0.0258}{0.0636 + 0.0258} = 0.289$$

R-square

Another way to evaluate the explanation of variation in the outcome is to compute a statistic analogous to the familiar R^2 used in multiple linear regression. In a multilevel model, however, there is an R^2 for each variance component. Use of these statistics is not without problems, however, because the R^2 may at times have negative values, and in other cases the addition of explanatory variables can lead to an increase rather than a decrease in variance components. The more complex a hierarchical model is, the more likely is the occurrence of anomalies in variance-explained statistics.

To calculate the R^2s for different levels of the level-3 model, we first need to get the variances for the null model, which is a 3-level model with no covariates. Open the file **TVSFP4.mum**, click on the **Variables** tab, and uncheck the check boxes of the selected **Explanatory Variables** as shown below.



Save the setup as **TVSFP7.mum** and run the model to get the following output of the variance/covariance component.

The R^2 s are calculated as

$$R_1^2 = 1 - \frac{\hat{\sigma}_p^2}{\hat{\sigma}_0^2} \qquad R_2^2 = 1 - \frac{\hat{\sigma}_{v_{(2)p}}^2}{\hat{\sigma}_{v_{(2)0}}^2} \qquad R_3^2 = 1 - \frac{\hat{\sigma}_{v_{(3)p}}^2}{\hat{\sigma}_{v_{(3)0}}^2}$$

where subscript 0 refers to a model with no covariates (*i.e.*, the null model, **TVSFP7.out**) and subscript p refers to a model with p covariates (*i.e.*, the full model, **TVSFP4.out**). The R^2 s for different levels are given in Table 3.7.

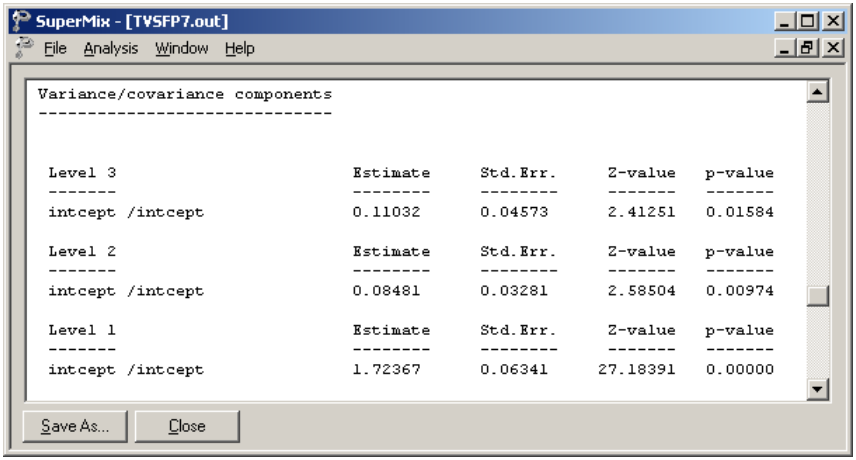


Table 3.7: R^2 values for a set of nested models

level	variance	null	full	R^2
1 (students)	$\hat{\sigma}^2$	1.724	1.602	.071
2 (classrooms)	$\hat{\sigma}^2_{\upsilon_{(2)}}$	.085	.064	.247
3 (schools)	$\hat{\sigma}^2_{\upsilon_{(3)}}$	.110	.026	.764

In the current example, only the intercept coefficient is allowed to vary randomly over classrooms and schools, thus making the calculation of the R^2 relatively straightforward. In the case of models with random slopes, the calculation of R^2 statistics becomes more difficult. For an extensive discussion of the rationale and calculation of R^2 statistics, the user is referred to Snijders & Bosker (2000).

Model fit statistics and comparison of models

Now, we consider two level-2 models using the same covariates but different level-2 IDs: one uses CLASS as level-2 ID, the other uses SCHOOL. The models' setups are

given in **TVSFP5.mum** and **TVSFP6.mum**. The comparison of estimates is summarized in the Table 3.8.

Table 3.8: Comparison of OLS and mixed model results

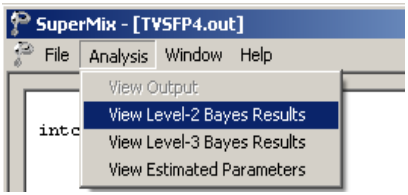
	OLS Estimates	Mixed Model		
		L-2 model		L-3 model
		L-2 ID: CLASS	L-2 ID: SCHOOL	L-2 ID: CLASS L-3 ID: SCHOOL
intcept	1.6613	1.6776	1.6952	1.6970
	(0.0844)	(0.0988)	(0.1145)	(0.1167)
CC	0.6406	0.6330	0.6601	0.6392
	(0.0921)	(0.1186)	(0.1440)	(0.1472)
TV	0.1987	0.1597	0.2024	0.1781
	(0.0900)	(0.1167)	(0.1401)	(0.1437)
CCxTV	–0.3216	–0.2747	–0.3697	–0.3204
	(0.1303)	(0.1678)	(0.2011)	(0.2055)
PRETHKS	0.3252	0.3116	0.3103	0.3072
	(0.0259)	(0.0258)	(0.0259)	(0.0258)
error variance		1.6030	1.6523	1.6020
		(0.0589)	(0.0589)	(0.0589)
class variance		0.0870		0.0636
		(0.0277)		(0.0277)
school variance			0.0372	0.0258
			(0.0184)	(0.0197)
Deviance		5359.9641	5366.0133	5357.3586
AIC		5373.9641	5380.0133	5373.3586
SBC		5394.3010	5389.3387	5384.0163
Number of free parameters		7	7	8

When comparing the deviances, AIC and SBC of the level-3 model with the level-2 models, we conclude that the three-level model has a better fit to the data.

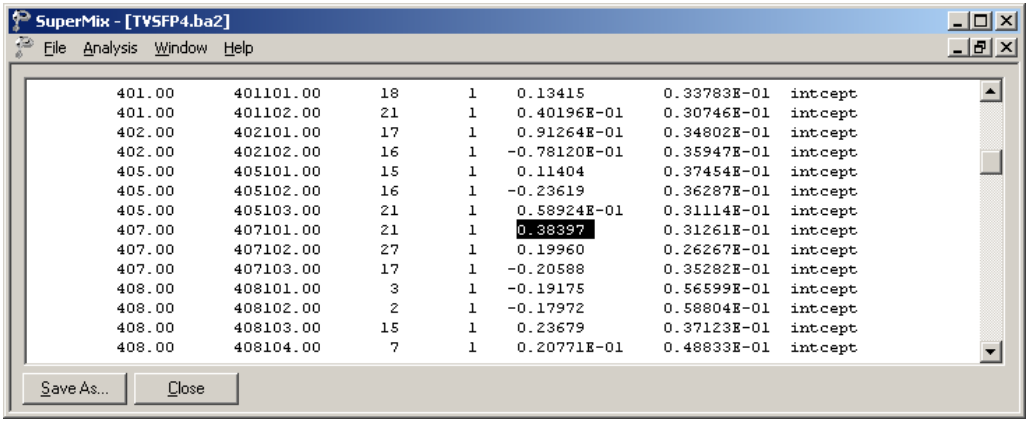
3.3.5.5 Residuals

Level-2 Bayes results

Returning to the **TVSFP4.mum** output, click on the **Analysis** menu of the output window or the model set up window, and note that **View Level-2 Bayes Results** is now activated. Select the option to open the level-2 Bayes results.



Note that the default extension for the level-2 Bayes estimates is **.ba2**. Part of the file is shown below.



The representations of these seven columns are given in order below:

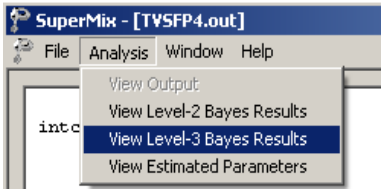
- Column 1: level-3 ID, which is school in our example.
- Column 2: level-2 ID, which refers to classroom.
- Column 3: number of the observations within level-2 ID, number of students within each classroom.
- Column 4: the number of the empirical Bayes coefficients.
- Column 5: the empirical Bayes estimate.
- Column 6: the estimated variance of the Bayes coefficient.
- Column 7: the name of the associated coefficient as used in the model.

Classroom 407102 has the largest Bayes estimate with a value of 0.38397. When considering the class difference, the predicted POSTTHKS score for a student in this specific class who only participated in television intervention with a PRETHKS score of 2 (CC = 0; TV = 1; CCxTV = 0) is calculated as follows.

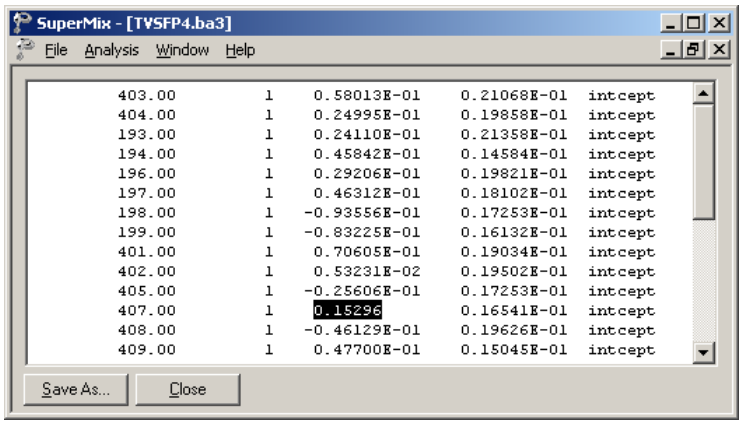
$$\begin{aligned}\widehat{\text{POSTTHKS}}_{ijk} &= \hat{\beta}_0 + \hat{\beta}_2\,\text{TV}_{ij} + \hat{\beta}_4\left(\text{PRETHKS}_{ijk}\right) + \hat{u}_{0i} \\ &= 1.697 + 0.17811 + 2 \times 0.3072 + 0.38397 \\ &= 2.87348.\end{aligned}$$

Level-3 Bayes results

Similarly, the level-3 Bayes results can be viewed by clicking on the **Analysis**, **View Level-3 Bayes Results**.



Part of the **TVSFP.ba3** is shown below.



The same classroom (ID = 407102) discussed above is nested in school number 407. Now, considering both the class and school differences, the estimated POSTTHKS for a student from this classroom who only participated in television intervention with a pre-intervention score of 2 (CC = 0; TV = 1; CCxTV = 0) is calculated as follows.

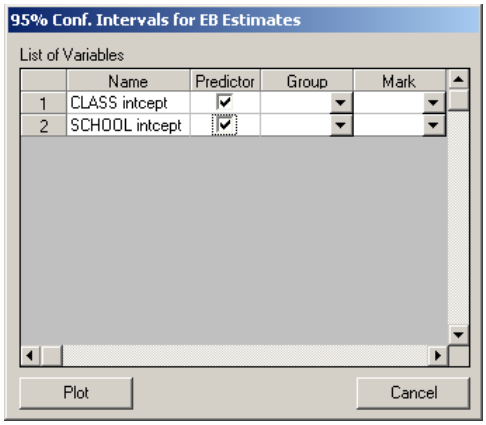
$$\begin{aligned}\widehat{\text{POSTTHKS}}_{ijk} &= \hat{\beta}_0 + \hat{\beta}_2\,\text{TV}_{ij} + \hat{\beta}_4\left(\text{PRETHKS}_{ijk}\right) + \hat{v}_{0ij} + \hat{v}_{0i} \\ &= 1.697 + 0.17811 + 2 \times 0.3072 + 0.38397 + 0.15296 \\ &= 3.02644.\end{aligned}$$

Confidence intervals for random coefficients

The **Confidence Intervals** option on the **File**, **Model-based Graphs** menu provides the option to display confidence intervals for the empirical Bayes estimates of the random effects specified in a given model. This option is now used to examine the confidence intervals of the random intercepts for the schools, which represent the highest level of the hierarchy in the current example.

Select the **Confidence Intervals** option on the **File**, **Model-based Graphs** menu to activate the **95% Conf. Intervals for EB estimates** dialog box. Two graphs of the

confidence intervals for the empirical Bayes estimates of the intercepts at the classroom level and school level are obtained by selecting CLASS intcept and SCHOOL intcept in the **Predictor** column before clicking **Plot**.



The graph obtained, as shown below, shows that, in general, the range of the confidence intervals for the level-3 empirical Bayes estimates of the intercepts is (–0.2; 0.2), and the range for level-2 is about (–0.4; 0.4).

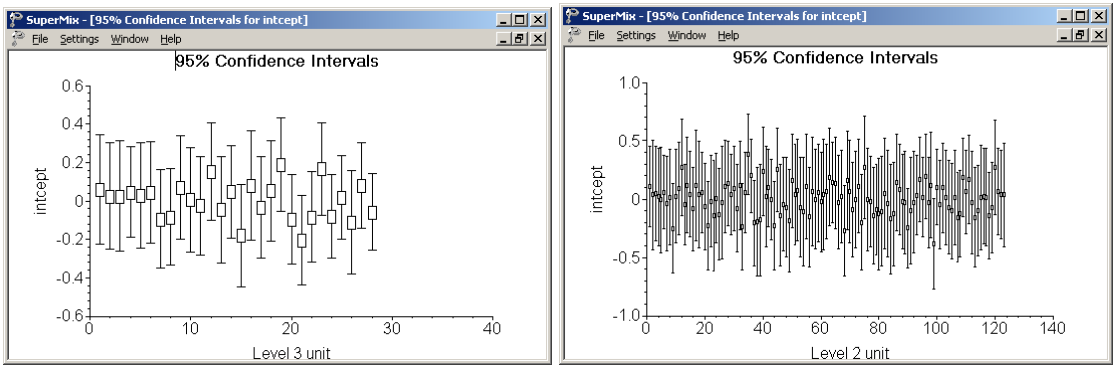


Figure 3.13: 95% confidence intervals for level-2 Bayes estimates

The deviations from the estimated population intercept over schools are also apparent. Each confidence interval is obtained using the formula

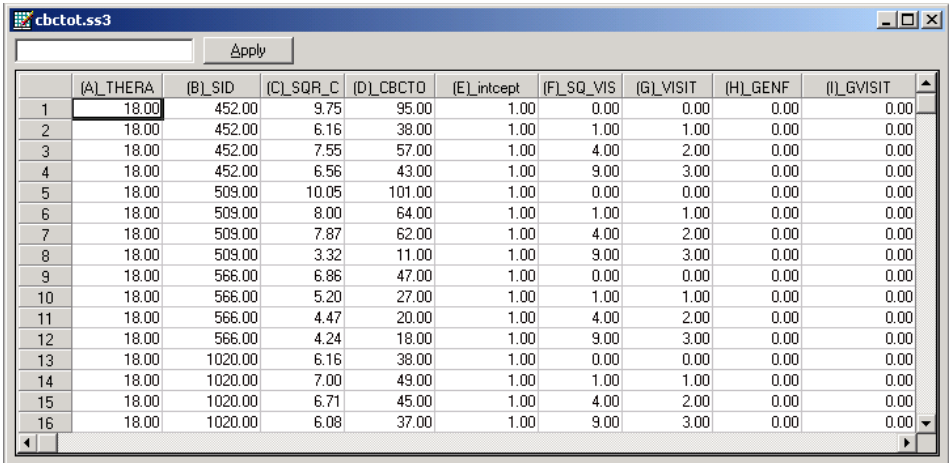
$$Empirical\ Bayes\ residual \pm 1.96\sqrt{var\left(Empirical\ Bayes\ residual\right)}.$$

3.4 3-level continuous example using a subset of Schoenwald data

3.4.1 The data

The data set for this example is taken from a study described in Schoenwald & Henggeler (2005). Children in the study were assigned to therapists and followed across time. In this study, respondents were rated with the Child Behavioral Checklist (Achenbach, 1991) at four occasions. The gender of each respondent was also recorded.

Although the total number of patients in this study was 1,951, the number of patients treated by any single therapist ranged between 1 and 19. A total of 7,127 measurements were made for all patients over the course of the study. Data for the observations of all the variables for the first four patients treated by therapist number 18 are shown below in the form of a SuperMix spreadsheet file, named **cbtot.ss3**.



	(A)_THERA	(B)_SID	(C)_SQR_C	(D)_CBCTO	(E)_intcept	(F)_SQ_VIS	(G)_VISIT	(H)_GENF	(I)_GVISIT
1	18.00	452.00	9.75	95.00	1.00	0.00	0.00	0.00	0.00
2	18.00	452.00	6.16	38.00	1.00	1.00	1.00	0.00	0.00
3	18.00	452.00	7.55	57.00	1.00	4.00	2.00	0.00	0.00
4	18.00	452.00	6.56	43.00	1.00	9.00	3.00	0.00	0.00
5	18.00	509.00	10.05	101.00	1.00	0.00	0.00	0.00	0.00
6	18.00	509.00	8.00	64.00	1.00	1.00	1.00	0.00	0.00
7	18.00	509.00	7.87	62.00	1.00	4.00	2.00	0.00	0.00
8	18.00	509.00	3.32	11.00	1.00	9.00	3.00	0.00	0.00
9	18.00	566.00	6.86	47.00	1.00	0.00	0.00	0.00	0.00
10	18.00	566.00	5.20	27.00	1.00	1.00	1.00	0.00	0.00
11	18.00	566.00	4.47	20.00	1.00	4.00	2.00	0.00	0.00
12	18.00	566.00	4.24	18.00	1.00	9.00	3.00	0.00	0.00
13	18.00	1020.00	6.16	38.00	1.00	0.00	0.00	0.00	0.00
14	18.00	1020.00	7.00	49.00	1.00	1.00	1.00	0.00	0.00
15	18.00	1020.00	6.71	45.00	1.00	4.00	2.00	0.00	0.00
16	18.00	1020.00	6.08	37.00	1.00	9.00	3.00	0.00	0.00

The variables of interest are:

- THERAPIS is the therapist ID (446 in total).
- SID is the patient ID (1951 in total).
- CBTOT is the total score of the Child Behavior Checklist.
- INT is a column of ones, representing an (optional) intercept.
- VISIT represents the visit number (0, 1, 2, or 3) at which a measurement was made.
- GENF is an indicator variable for gender, and assumes the value 0 for males and 1 for females.
- GVISIT represents the interaction between GENF and VISIT, and is the product of GENF and VISIT.

3.4.2 Exploring the data

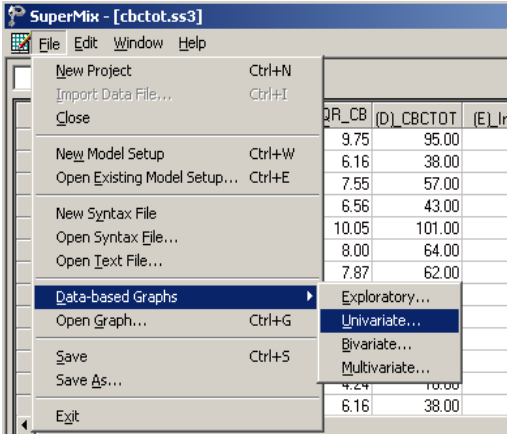
Relationships between variables, and trends over time in repeated measurement data, may be conveyed in an informal and simplified visual form via graphical displays. SuperMix offers both data-based and model-based graphs. Data-based graphing options are accessed via the **File**, **Data-based Graphs** option once a SuperMix data file (**.ss3**) is opened, while model-based graphs are available after the analysis has been performed, and will be discussed later in this section.

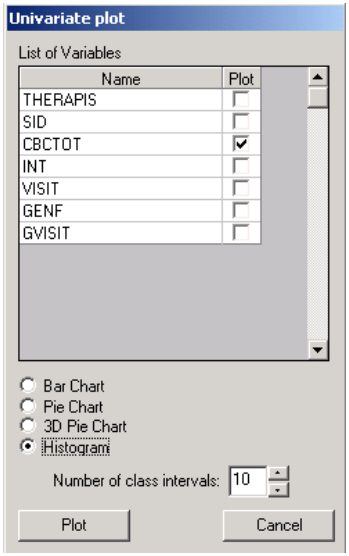
In the case of data-based graphs, we distinguish between three categories: univariate, bivariate, and multivariate graphs. Univariate graphs are particularly useful to obtain an overview of the characteristics of a single variable. In the sections to follow, we use data-based graphs to take a closer look at some of the variables in these data.

3.4.2.1 Univariate graphs

Histograms

As a first step, we take a look at the distribution of the total score on the Child Behavior Checklist (CBTOT) which is the potential dependent variable in this study. While scores such as these are not truly continuous variables, they are often treated as if they were. However, like personal income, the distribution of a score often is skewed. As a first step, we will take a closer look at the distribution of the intended outcome variable CBTOT. To do so, select the **Univariate** option from the **Data-based Graphs** menu as shown below.





The **Univariate plot** dialog box appears. Select the variable CBTOT and indicate that a **Histogram** is to be graphed. Note that the number of class intervals shown on the histogram is controlled by the **Number of class intervals** field, which is left at the default value of 10 in this case. Click the **Plot** button to display the histogram.

The histogram below shows that the distribution of total scores (CBCTOT) on the Child Behavior Checklist (CBC) is markedly asymmetrical. Given the normality assumptions used in fitting a 3-level linear multilevel model, it would be inappropriate to use CBCTOT in its current state. A transformation of this variable is required before it would make a suitable outcome variable for the intended analysis.

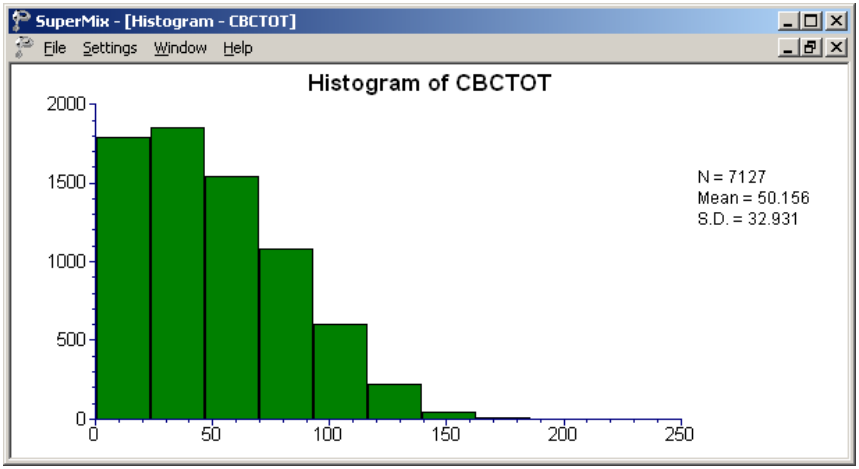
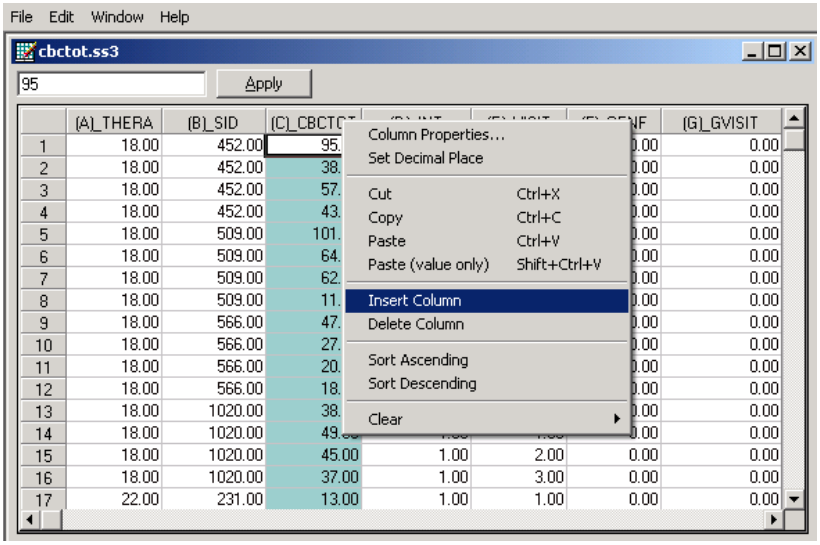


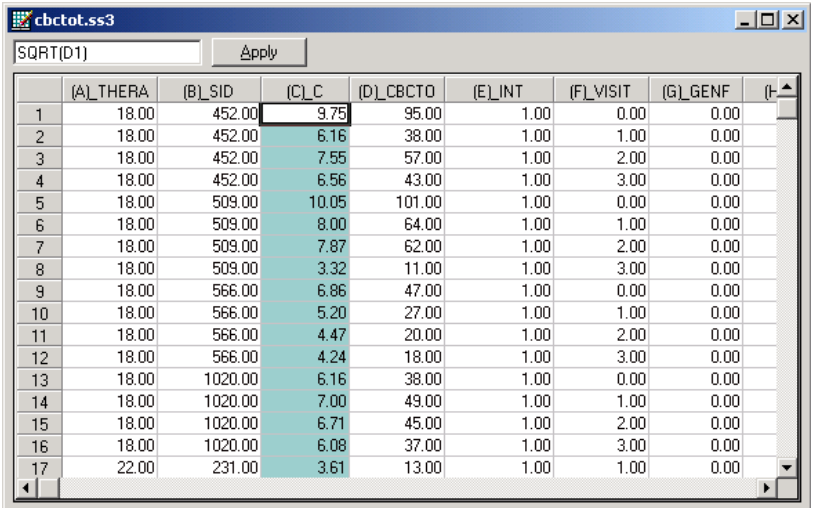
Figure 3.14: Histogram of the variable CBCTOT

Transformation of variables

Common transformations used in the case of skewed variables include the natural logarithm of the variable in question, or the square root of the variable. We opt to explore the possibility of using the square root of the total score as outcome. To do so, a new variable containing the square root of the current total scores has to be created in the SuperMix spreadsheet. Right-click on the column with CBCTOT as heading, and select the **Insert Column** option from the pop-up menu that appears.

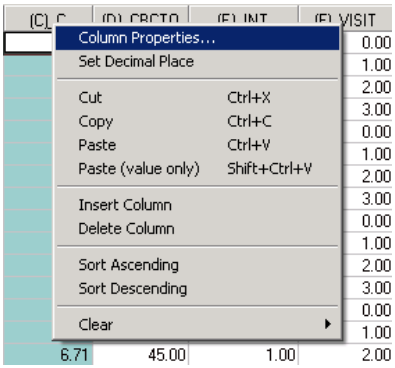


Select the new column and input the function **SQRT(D1)** in the formula box as shown below. Click the **Apply** button. Each value of the new variable is the square root value of the corresponding value of the variable CBCTOT as shown below.

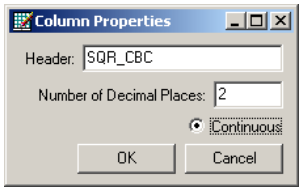


	(A)_THERA	(B)_SID	(C)_C	(D)_CBCTO	(E)_INT	(F)_VISIT	(G)_GENF
1	18.00	452.00	9.75	95.00	1.00	0.00	0.00
2	18.00	452.00	6.16	38.00	1.00	1.00	0.00
3	18.00	452.00	7.55	57.00	1.00	2.00	0.00
4	18.00	452.00	6.56	43.00	1.00	3.00	0.00
5	18.00	509.00	10.05	101.00	1.00	0.00	0.00
6	18.00	509.00	8.00	64.00	1.00	1.00	0.00
7	18.00	509.00	7.87	62.00	1.00	2.00	0.00
8	18.00	509.00	3.32	11.00	1.00	3.00	0.00
9	18.00	566.00	6.86	47.00	1.00	0.00	0.00
10	18.00	566.00	5.20	27.00	1.00	1.00	0.00
11	18.00	566.00	4.47	20.00	1.00	2.00	0.00
12	18.00	566.00	4.24	18.00	1.00	3.00	0.00
13	18.00	1020.00	6.16	38.00	1.00	0.00	0.00
14	18.00	1020.00	7.00	49.00	1.00	1.00	0.00
15	18.00	1020.00	6.71	45.00	1.00	2.00	0.00
16	18.00	1020.00	6.08	37.00	1.00	3.00	0.00
17	22.00	231.00	3.61	13.00	1.00	1.00	0.00

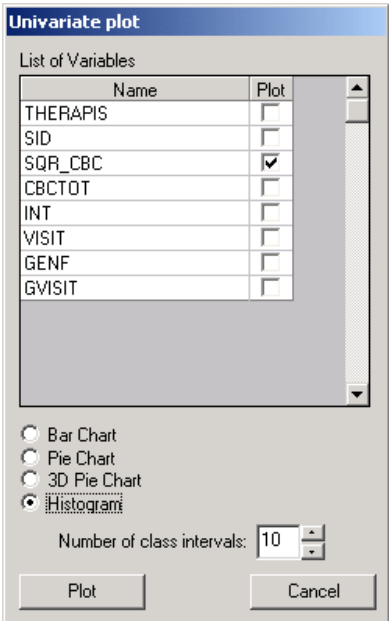
To rename the new variable, right-click again on the column header and select the **Column Properties** option.



Complete the **Header** field in the **Column Properties** dialog box as shown below. Also indicate that this is a continuous variable by selecting the **Continuous** option before clicking the **OK** button.



Check the distribution of the square root of the total score on the Child Behavior Checklist (SQR_CBC) by selecting the **Univariate** option from the **Data-based Graphs** menu to activate the **Univariate plot** dialog box. After selecting SQR_CBC by checking the appropriate box in the **Plot** column, select the **Histogram** option as before, and click **Plot**.



The histogram for the variable SQR_CBC is appreciably more symmetric than was the case for the original variable CBCTOT, as evident from the histogram shown below.

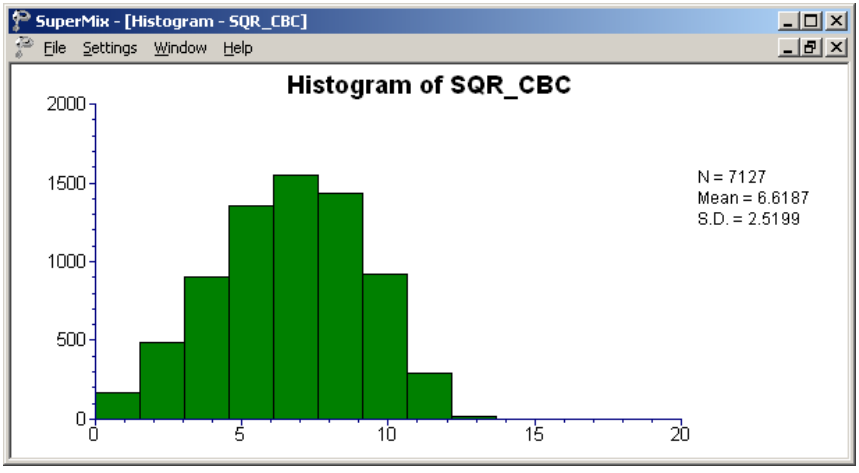


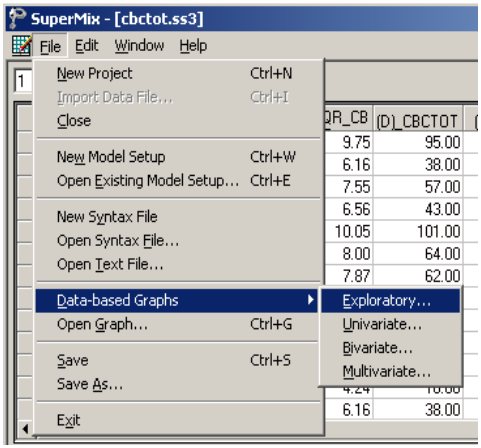
Figure 3.15: Histogram of the variable SQR_CBC

3.4.2.2 Bivariate graphs

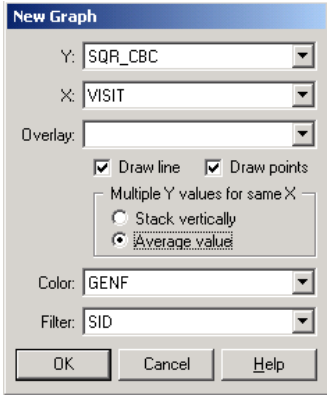
It is hoped that the total scores of patients would change over time, *i.e.*, with successive visits to their therapists. In addition, it is hypothesized that the gender of a patient may also have some relationship to the total score of a patient. Bivariate plots of possible relationships are a handy tool for the exploration of possible relationships.

Exploratory graphs

To explore the relationship between the time since the start of therapy and the square root of the total score, select the **Data-based Graphs**, **Exploratory** option from the **File** menu to activate the **New Graph** dialog box.

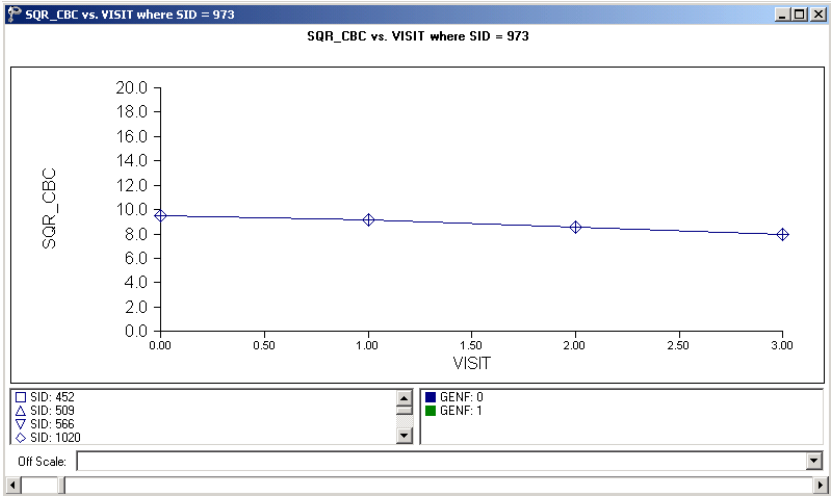


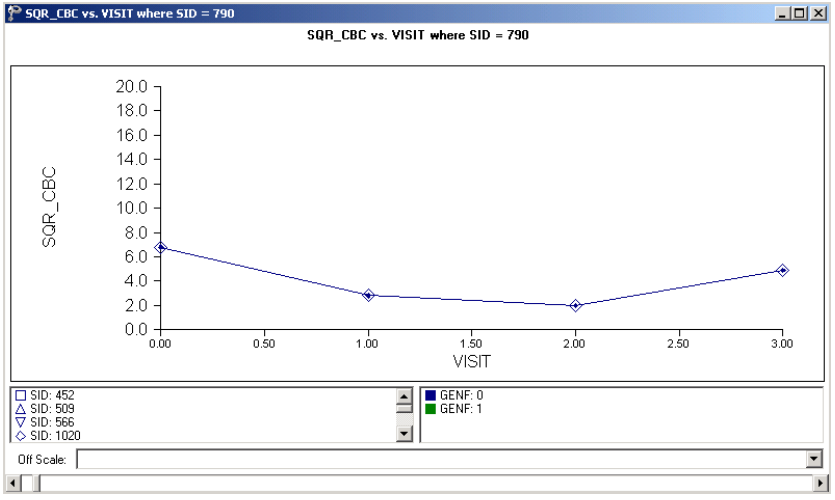
Select the outcome variable SQR_CBC as the **Y**-variable and VISIT as the **X** variable. Add the variable representing gender, GENF, from the **Color** field. Doing so will lead to the graphs of the gender groups to be displayed in different colors (blue and green being the default colors for two groups). Select the patient ID, as denoted by the variable SID, as the **Filter** variable to obtain individual graphs for patients. Click **OK** after completing the fields on this dialog box.



Graphs for patients with SIDs equal to 973, 790, and 2233 are shown below. These are but three of the 1951 graphs created via the graphing specification described above. Plotting symbols for each patient are shown at the bottom left of the graphing window, and the legend for gender groups to the right. The slider at the bottom of the window is used to move from one graph to another.

For the first patient, with SID equal to 973, a roughly linear decrease in the outcome is observed as the visit number increases. This is not the case for patient 790, where an almost parabolic curve is observed, or for patient 2233 where an inverted parabola seems to be the most obvious line to fit. It can be concluded from these graphs that the relationship between SQR_CBC and VISIT differs from patient to patient, and moreover that it may not be strictly linear. The possible inclusion of a quadratic function of the time of measurement, *i.e.* VISIT, should be explored. No definite trend is immediately apparent for gender groups within the wide variety of curves plotted, but the possibility of an interaction between the gender and the number of the visit cannot be excluded.


SQR_CBC vs. VISIT where SID = 973
SQR_CBC vs. VISIT where SID = 973
SQR_CBC
20.0
18.0
16.0
14.0
12.0
10.0
8.0
6.0
4.0
2.0
0.0
0.00
0.50
1.00
1.50
2.00
2.50
3.00
VISIT
SID: 452
SID: 509
SID: 566
SID: 1020
GENF: 0
GENF: 1
Off Scale:


SQR_CBC vs. VISIT where SID = 790
SQR_CBC vs. VISIT where SID = 790
SQR_CBC
20.0
18.0
16.0
14.0
12.0
10.0
8.0
6.0
4.0
2.0
0.0
0.00
0.50
1.00
1.50
2.00
2.50
3.00
VISIT
SID: 452
SID: 509
SID: 566
SID: 1020
GENF: 0
GENF: 1
Off Scale:

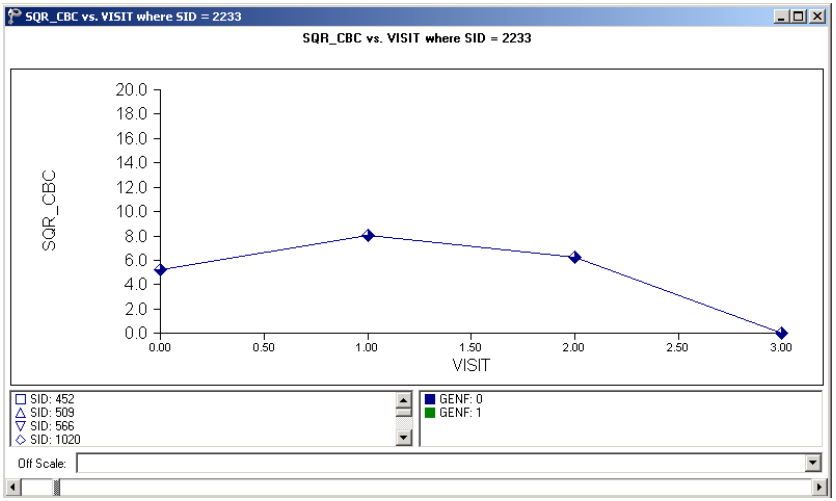
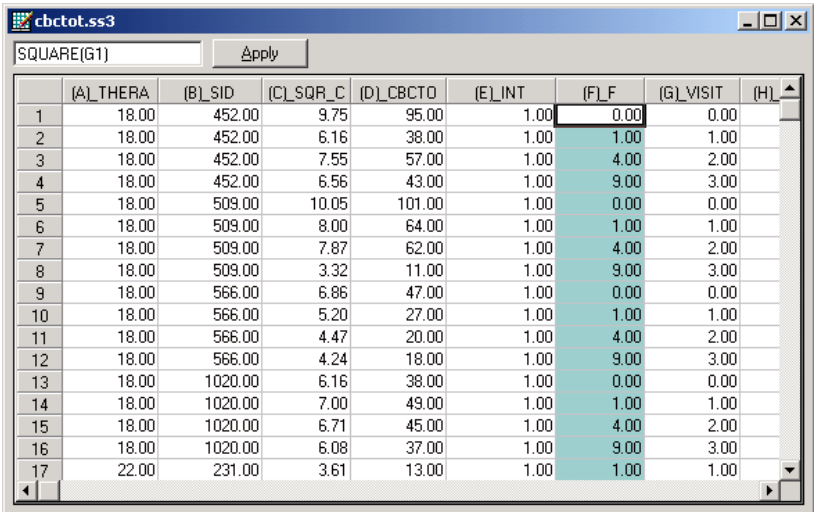


Figure 3.16: Relationship between SQR_CBC and VISIT for selected patients

Transforming a variable

To examine the relationship between the outcome and the quadratic value of VISIT, a new variable has to be created. This is done in a similar way to adding the square root of the total score. First insert a column, then type the appropriate function into the formula box as shown below. Click the **Apply** button. Each value of the new variable is the squared value of the corresponding value of the variable VISIT as shown below.

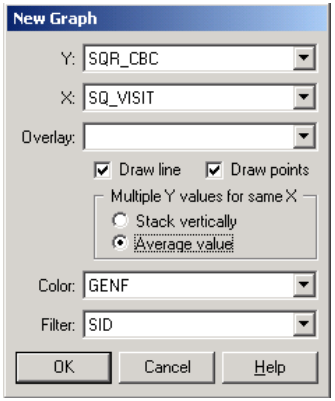


	(A)_THERA	(B)_SID	(C)_SQR_C	(D)_CBCTO	(E)_INT	(F)_F	(G)_VISIT	(H)_
1	18.00	452.00	9.75	95.00	1.00	0.00	0.00	
2	18.00	452.00	6.16	38.00	1.00	1.00	1.00	
3	18.00	452.00	7.55	57.00	1.00	4.00	2.00	
4	18.00	452.00	6.56	43.00	1.00	9.00	3.00	
5	18.00	509.00	10.05	101.00	1.00	0.00	0.00	
6	18.00	509.00	8.00	64.00	1.00	1.00	1.00	
7	18.00	509.00	7.87	62.00	1.00	4.00	2.00	
8	18.00	509.00	3.32	11.00	1.00	9.00	3.00	
9	18.00	566.00	6.86	47.00	1.00	0.00	0.00	
10	18.00	566.00	5.20	27.00	1.00	1.00	1.00	
11	18.00	566.00	4.47	20.00	1.00	4.00	2.00	
12	18.00	566.00	4.24	18.00	1.00	9.00	3.00	
13	18.00	1020.00	6.16	38.00	1.00	0.00	0.00	
14	18.00	1020.00	7.00	49.00	1.00	1.00	1.00	
15	18.00	1020.00	6.71	45.00	1.00	4.00	2.00	
16	18.00	1020.00	6.08	37.00	1.00	9.00	3.00	
17	22.00	231.00	3.61	13.00	1.00	1.00	1.00	

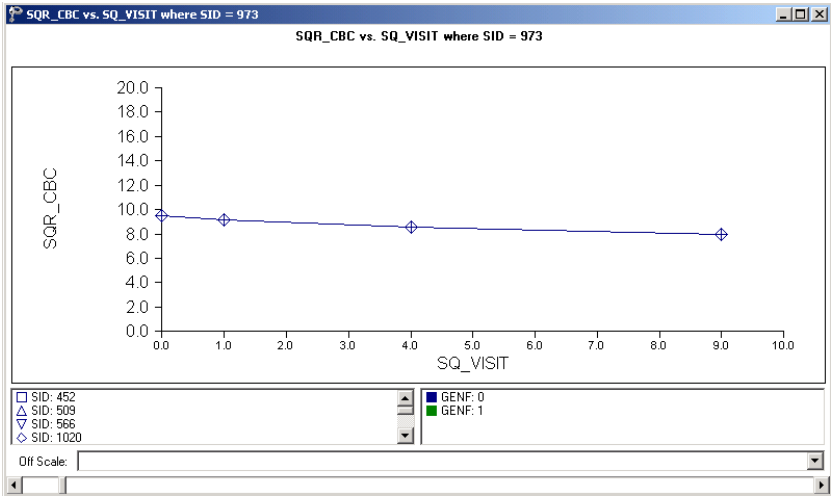
Right-click on the header of the newly inserted column to activate the **Column Properties** dialog box and enter a variable name such as SQ_VISIT into the **Header** field. Click **OK** to return to the spreadsheet.

3.4.2.3 Exploratory graphs

Remake the bivariate graphs shown previously for SQR_CBC and VISIT, using the squared value of VISIT (SQ_VISIT) instead. The completed **New Graph** dialog box, accessed via the **Data-based Graphs**, **Exploratory** option, is shown below. Click **OK** to display the graphs for individual patients.



Very little change in the shape of the plots is observed in the graphs obtained for the three patients considered earlier. To follow up on the possibility of a nonlinear relationship between the outcome and the visit number, both of the variables VISIT and SQ_VISIT will be included in the first model fitted, where the relationship of each with the outcome can be evaluated in the presence of the other.



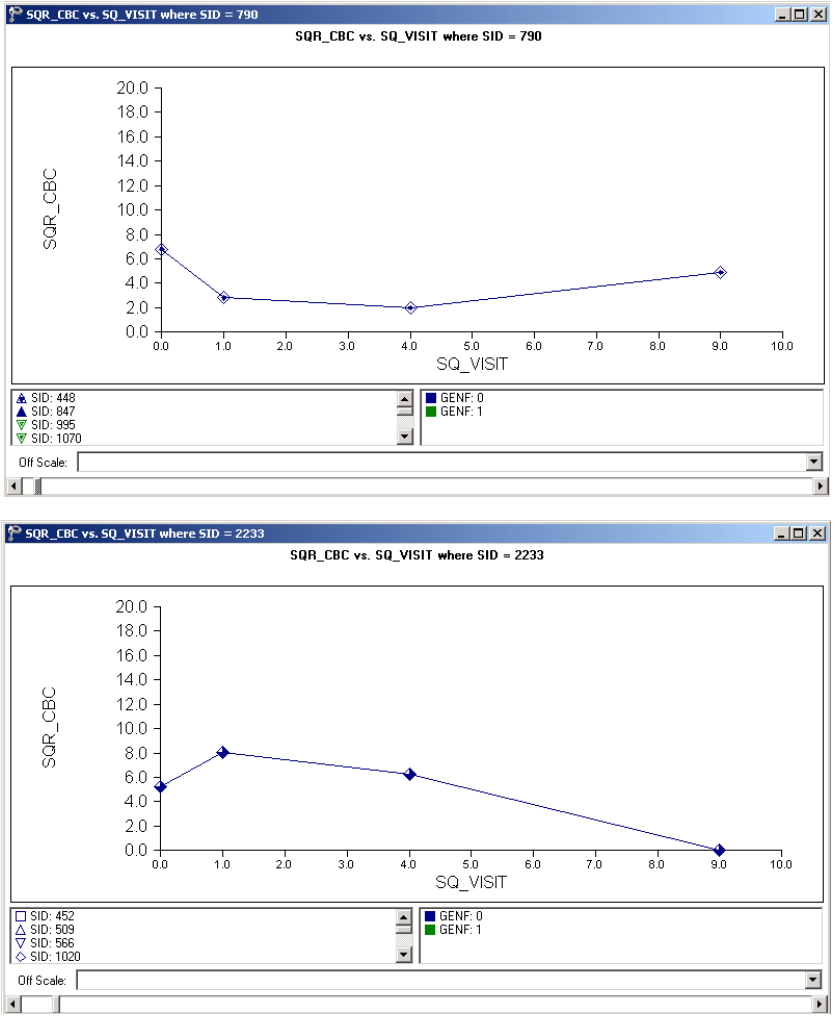
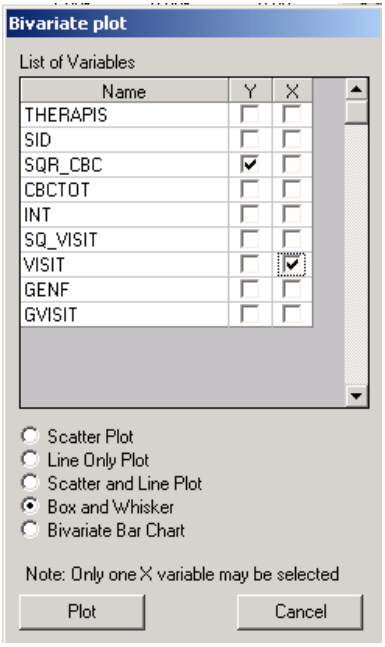


Figure 3.17: relationship between SQR_CBC and SQ_VISIT for selected patients

Box-and-whisker plots

Another bivariate plot of interest is a box-and-whisker plot, which may be used to examine the distributions of continuous variables such as for the different values of discrete valued predictors. This option, accessed via the **Data-based Graphs**, **Bivariate** option on the **File** menu, is now used to take a closer look at the distribution of the transformed outcome variable at different visits, and for the two gender groups.

The **Bivariate plot** dialog box is completed as follows: select the outcome variable SQR_CBC as the **Y**-variable of interest, and the predictor VISIT to be plotted on the **X**-axis. Check the **Box and Whisker** option, and click **Plot**.



In the plot shown below, the box-and-whisker plots for the square root of the **CBCTOT** scores are shown at each of the measurement occasions. Recall that the bottom line of a box represents the first quartile (q_1), the top line the third quartile

(q_3), and the in-between line the median (me). Here, the arithmetic mean is represented by a diamond. A decrease in the mean **HDRS** rating is observed over the course of the study. In addition, the larger distances between the extremes of the boxes at the later measurement occasions indicate more variability in the transformed **CBCTOT** scores towards the end of the study.

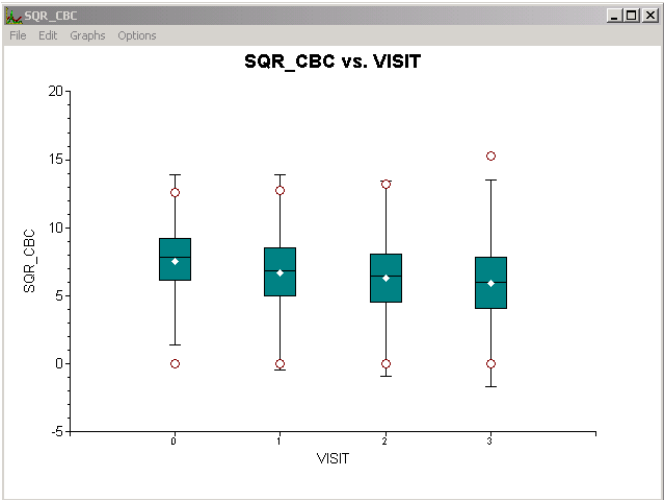


Figure 3.18: Box-and-whisker plot of SQR_CBC vs. VISIT

When a similar plot is made for the original total score as represented by the variable CBCTOT, it is clear that the distributions of the transformed scores, though still exhibiting more variability at later visits, are closer to normal for the transformed variable (figure below).

Figure 3.19: Box-and-whisker plot of CBCTOT vs. VISIT

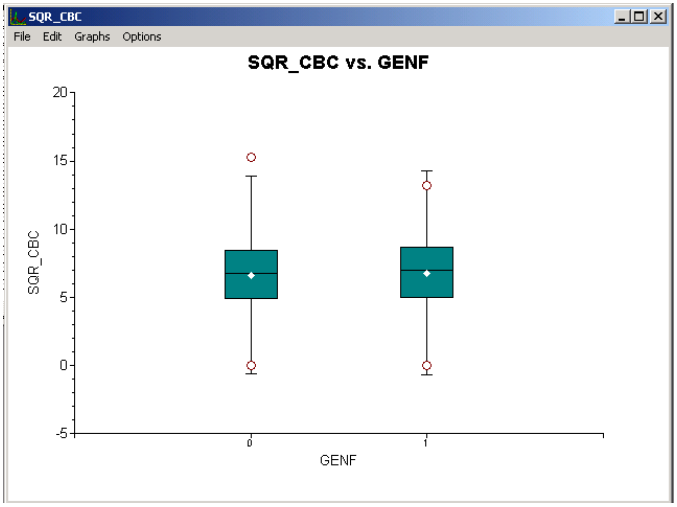


Figure 3.20: Box-and-whisker plot of SQR_CBC vs. GENF

A box-and-whisker plot of the transformed scores for the two gender groups can easily be obtained. Simply close the graph window shown above, deselect VISIT as the **X**-variable and select the indicator of gender GENF instead. Click **Plot** to obtain

the box-and-whisker plot shown below. A slightly larger range of scores is observed for males (GENF = 0) than for females (GENF = 1).

When Figure 3.20 is compared to a similar one for the untransformed outcome variable CBCTOT, the same tendency towards a less normal distribution is observed, particularly with respect to the total scores of male patients.

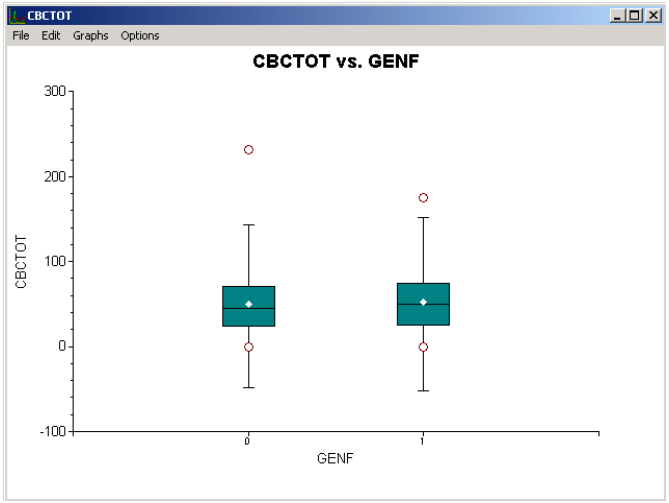


Figure 3.21: Box-and-whisker plot of CBCTOT vs. GENF

Bivariate bar charts

Another bivariate plot that may provide insight is a plot of gender by the number of visits. The **Bivariate** option on the **File**, **Data-based Graphs** menu is again used to access the **Bivariate plot** dialog box. Select VISIT as the **Y**-variable and GENF as the **X**-variable, and request a bivariate bar chart prior to clicking the **Plot** button.

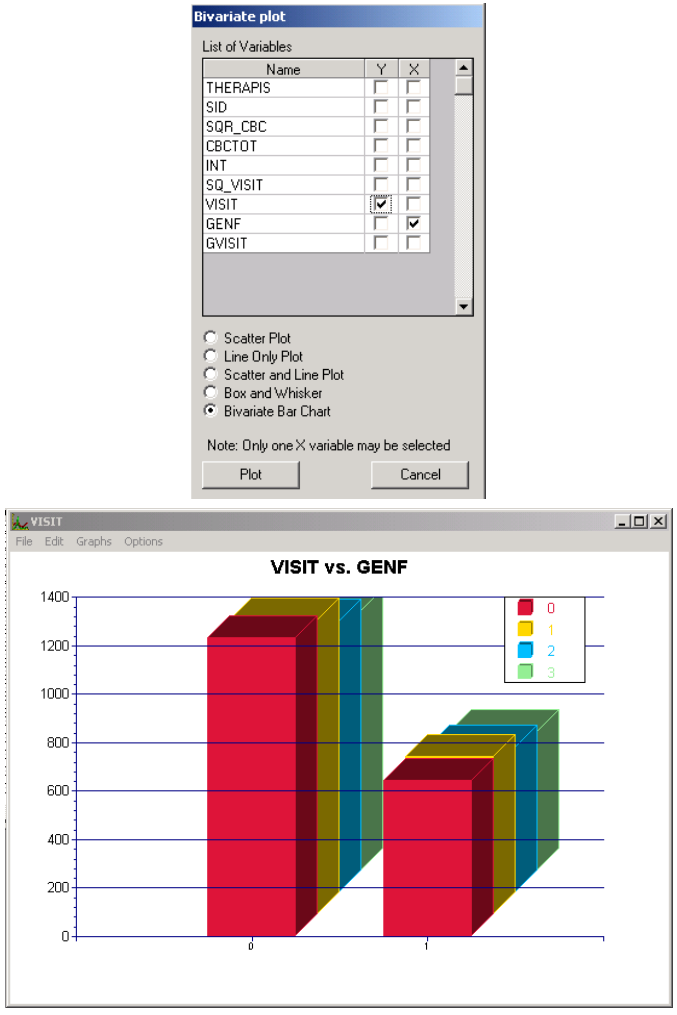


Figure 3.22: Bivariate chart of VISIT vs GENF

The bar chart for VISIT vs. GENF shows not only that more males than females are present in the data, but also that roughly equal numbers of observations/scores are available for the two groups at each of the visits. The pattern in terms of the number of observations available at each visit is the same for the two gender groups.

3.4.3 Fitting a growth curve model to the data

3.4.3.1 The model

The first model fitted to the data explores the relationship between SQR_CBC and the visit number, as represented by the variables VISIT and SQ_VISIT:

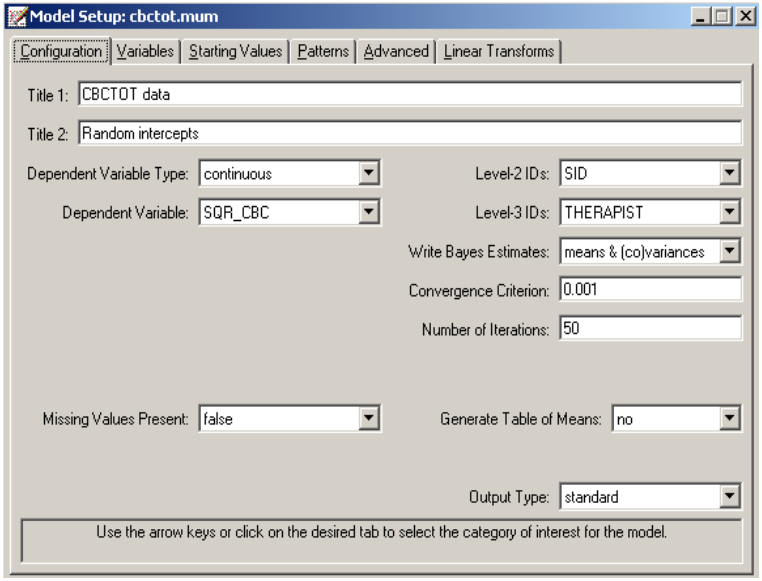
$$\text{SQR\_CBC}_{ijk} = \beta_0 + \beta_1 * \text{VISIT}_{ijk} + \beta_2 * \text{SQ\_VISIT}_{ijk} + v_{i0} + v_{ij0} + e_{ijk}$$

In this model, β_0 denotes the average expected total score, and β_1 and β_2 indicate the estimated coefficients associated with the fixed part of the model which contains the predictor variables VISIT and SQ_VISIT. The random part of the model is represented by v_{i0}, v_{ij0} and e_{ijk}, which denote the variation in average total score over therapists, between patients (or, in other words, over patients nested within therapists) and between measurements at the lowest level of the hierarchy.

3.4.3.2 Setting up the analysis

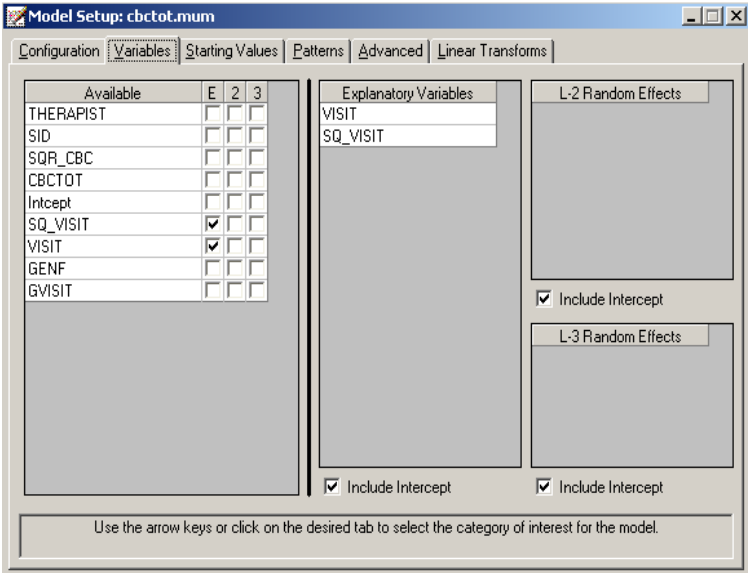
Open the SuperMix spreadsheet **cbtot.ss3**. The next step is to describe the model to be fitted. We use the SuperMix interface to provide the model specifications. From the main menu bar, select the **File, New Model Setup** option.

Select the continuous outcome variable SQR_CBC from the **Dependent Variable** drop-down list box on the **Configuration** tab. The therapist number THERAPIS and respondent identification code SID used to define the levels of the hierarchy are specified as **Level-3 ID** and **Level-2 ID** respectively by selecting them from the **Level-3 IDs** and **Level-2 IDs** drop-down list boxes. Enter a title for the analysis in the **Title** text boxes. Select the **means & (co)variances** option from the **Write Bayes estimates** drop-down list box to request the writing of residuals to an external file. In this example, default settings for all other options associated with the **Configuration** screen are used. Proceed to the **Variables** screen by clicking on that tab.



The **Variables** screen is used to specify the fixed and random effects to be included in the model. Select the explanatory (fixed) variables using the **E** check boxes next to the variables names in the **Available** grid at the left of the screen. Note that, as the variables are selected, the selected variables are listed in the **Explanatory Variables** grid. After selecting all the explanatory variables, the screen shown below is obtained. 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 levels 2 and 3 of the hierarchy. In this example, we want to fit a model with random intercepts at levels 2 and 3. By default, the **Include Intercept** check boxes in both the **L-2 Random Effects** and **L-3 Random effects** grids are checked. If these boxes are left checked, and no additional random effects are indicated using the **2** column in the **Available** grid to the left, the model fitted will be the random-intercepts-only model we intend to use. No further changes on this screen are necessary.

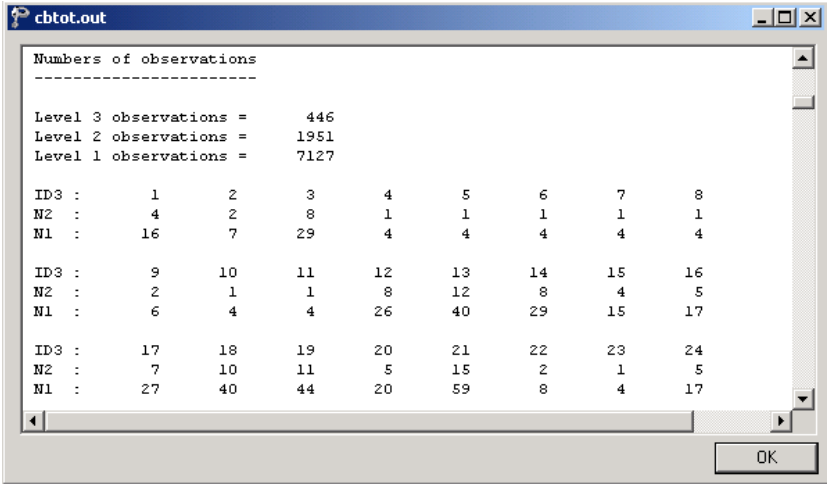


Before running the analysis, the model specifications have to be saved. Select the **File**, **Save As** option, and provide a name (**cbctot.mum**) for the model specification file. Run the analysis by selecting the **Run** option from the **Analysis** menu.

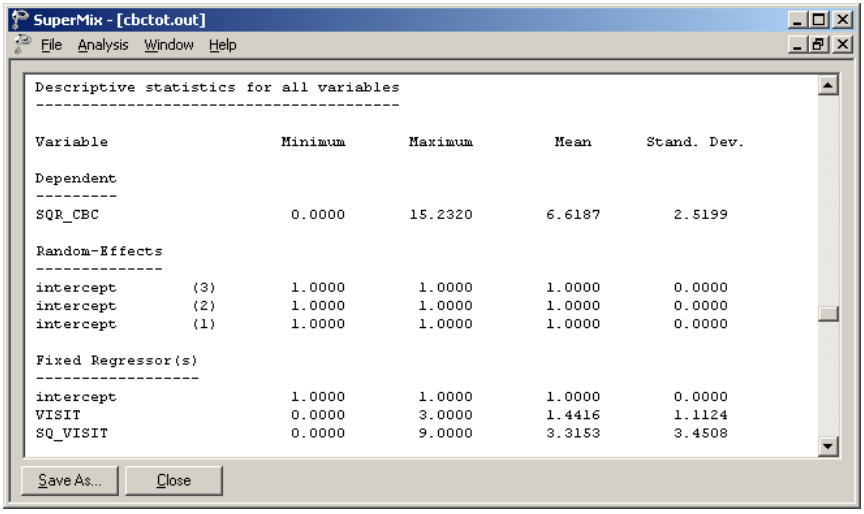
3.4.3.3 Discussion of results

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

In the first section of the output file, a description of the hierarchical structure is provided. Data from a total of 446 therapists and 1951 patients at 7127 measurement occasions were included at levels 3, 2 and 1 of the model. This corresponds to the study design described earlier. In addition, a summary of the number of patients and measurements nested within each therapist is provided. The therapist with ID3 = 21, for example, had 15 patients (N2: 15). These patients were measured at 59 occasions. By contrast, therapist 23 had only 1 patient, for whom 4 measurements were made.

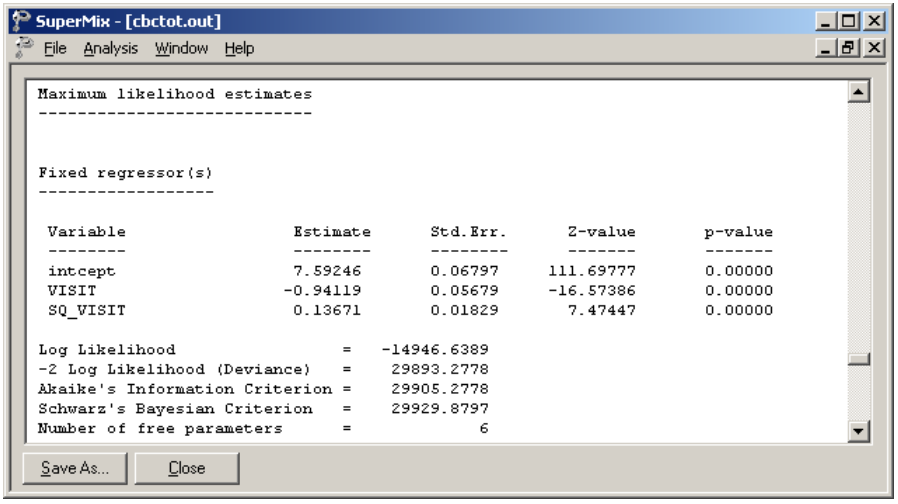


The data summary is followed by descriptive statistics for all the variables included in the model. The mean of 6.61867 reported for the outcome SQR_CBC translates to a total score of 43.806 on the Child Behavior Checklist.



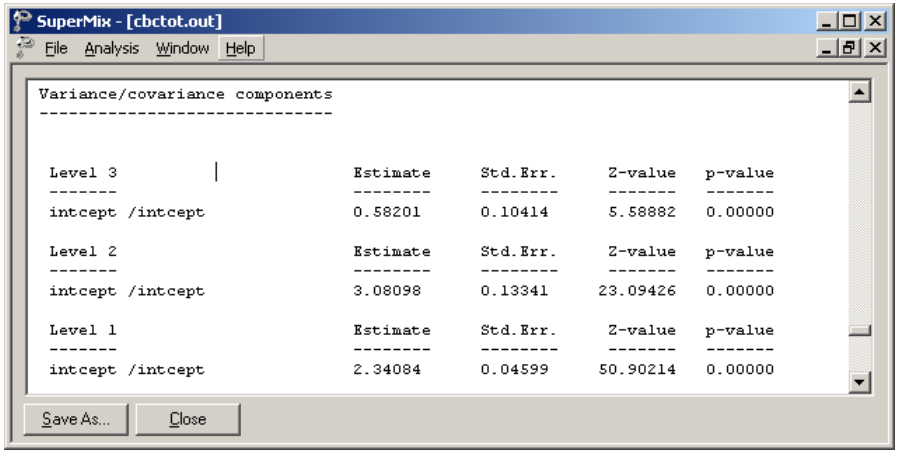
Fixed effects results

The output describing the estimated **fixed effects** after convergence is shown next. The estimates are shown in the column with heading Estimate, and correspond to the coefficients $\beta_0, \beta_1, \ldots, \beta_3$ in the model specification. From the z-values and associated exceedance probabilities, we see that the coefficients associated with both the time of measurement (VISIT) and squared value of the time of measurement (SQ_VISIT) are highly significant. The significance of the estimate associated with SQ_VISIT supports the tentative conclusion made during the exploratory analysis that the relationship between score and visit number cannot adequately be described by a linear relationship. While the average CBC score is expected to decrease with 0.94119 units between two successive visits, a smaller increase in score of 0.13671 is associated with the squared value of the time of measurement.



Random effects results

The output for the **random part** of the model follows, and is shown in the image below.



There is significant variation in the average estimated total health expenditure at all levels, with the most variation over the patients (level-2), and the least variation over therapists (level-3).

3.4.3.4 Interpreting the results

Estimated outcomes for different groups

A typical patient at the start of the study is expected to have a transformed CBC score of

$$\begin{aligned}\widehat{\text{SQR\_CBC}}_{ij0} &= \hat{\beta}_0 + \hat{\beta}_1 * \text{VISIT}_{ij0} + \hat{\beta}_2 * \text{SQ\_VISIT}_{ij0} \\ &= 7.59246 - 0.94119(0) + 0.13671(0^2) \\ &= 7.59246,\end{aligned}$$

that is, the estimated intercept. Similar equations for expected transformed scores at subsequent measurements (visits) are obtained in the same way:

$$\text{VISIT } 1: \widehat{\text{SQR\_CBC}}_{ij1} = 7.59245 - 0.94119(1) + 0.13671(1^2)$$
$$= 7.59246 - 0.94119 + 0.13671$$
$$= 6.78798$$

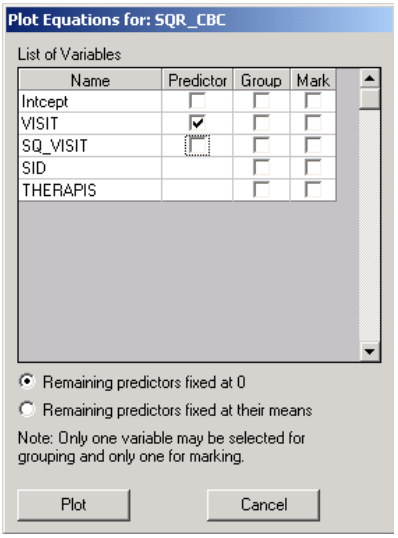
$$\text{VISIT } 2: \widehat{\text{SQR\_CBC}}_{ij2} = 7.59245 - 0.94119(2) + 0.13671(2^2)$$
$$= 7.59246 - 1.88238 + 0.54684$$
$$= 6.25692$$

$$\text{VISIT } 3: \widehat{\text{SQR\_CBC}}_{ij3} = 7.59245 - 0.94119(3) + 0.13671(3^2)$$
$$= 7.59245 - 2.82357 + 1.23039$$
$$= 5.99928$$

The effect of the positive estimate for SQ_VISIT in slowing down the expected decrease in CBC scores over successive measurement occasions is clear from the equations above: without this estimate, the expected CBC scores at visits 1, 2, and 3 would have been 6.65126, 5.71007, and 4.76888 respectively. In terms of the actual rather than the square root of the CBC scores, the expected scores at the 4 measurement occasions under the current model are 57.6453, 46.0765, 39.1489, and 35.9914 respectively.

Model-based graphs

Using the **Plot Equations for: SQR_CBC** dialog box that appears when the **File**, **Model-based Graphs**, **Equations** option is selected, we can graphically depict the trend in expected average squared score for the predictors VISIT and SQ_VISIT. The dialog box below shows the selection of the predictor VISIT, and in the graph requested, SQ_VISIT will be fixed at a value of 0.



The graph below shows the result obtained when the **Plot** button is clicked after completion of the **Plot Equations for: SQR_CBC** dialog box as shown above.

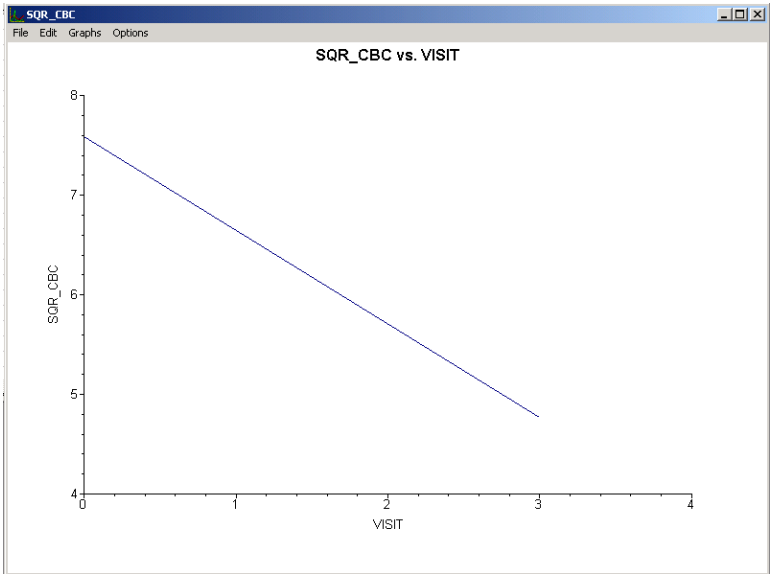


Figure 3.23: Plot of SQR_CBC vs. VISIT

A similar plot for the predictor SQ_VISIT is given in Figure 3.24. Note that, in the second graph, the increase in expected score seems larger than implied by the estimate of 0.13671. This is due in part to the difference in the ranges of the two predictors, as reflected in the tick marks on the X-axes of the graphs.

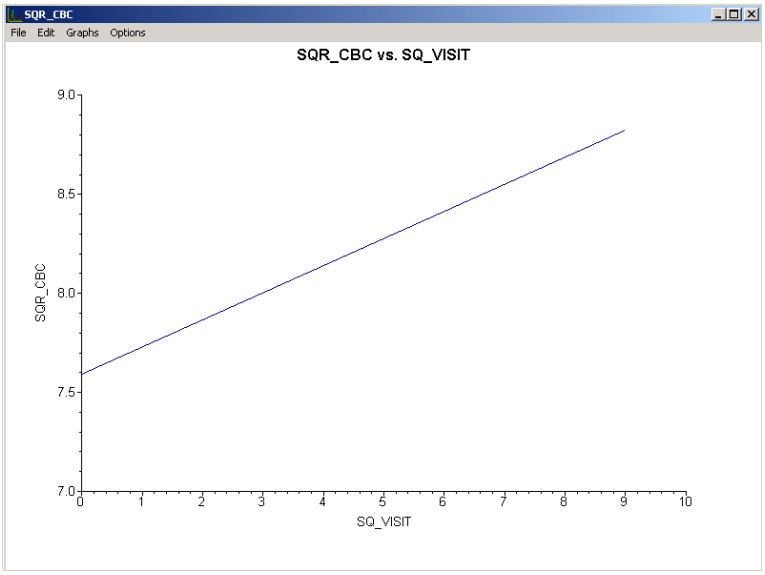


Figure 3.24: Plot of SQR_CBC vs. SQ_VISIT

Fit statistics and ICC

From the output for the **random part** of the model it is clear that there is significant variation in the average estimated total health expenditure at all levels, with the most variation over the patients (level-2), and the least variation over therapists (level-3).

An estimate of the percentage of variation in the outcome at a patient level, for example, is obtained as

$$\frac{3.08097}{0.58201+3.08097+2.34083}\times 100\% = 51.32\%$$

indicating that 51.32% of the total variance in scores is at the patient level. In contrast,

$$\frac{0.58201}{0.58201+3.08097+2.34083}\times 100\% = 9.69\%$$

is at the therapist level, with the remainder over measurements nested within patients.

3.4.4 Fitting a random intercept model with 3 predictors and interaction term to the data

3.4.4.1 The model

From the exploratory analysis, we are aware of a possibly nonlinear relationship between the transformed outcome variable SQR_CBC and the visit number, as represented by the variables VISIT and SQ_VISIT. Differences in the distributions of the transformed scores of the two gender groups also lead us to suspect that the outcome may depend to some extent on the gender of the patient. The possibility of an interaction between the time elapsed since the start of the study, as represented by VISIT and SQ_VISIT, and the gender of a patient cannot be ruled out.

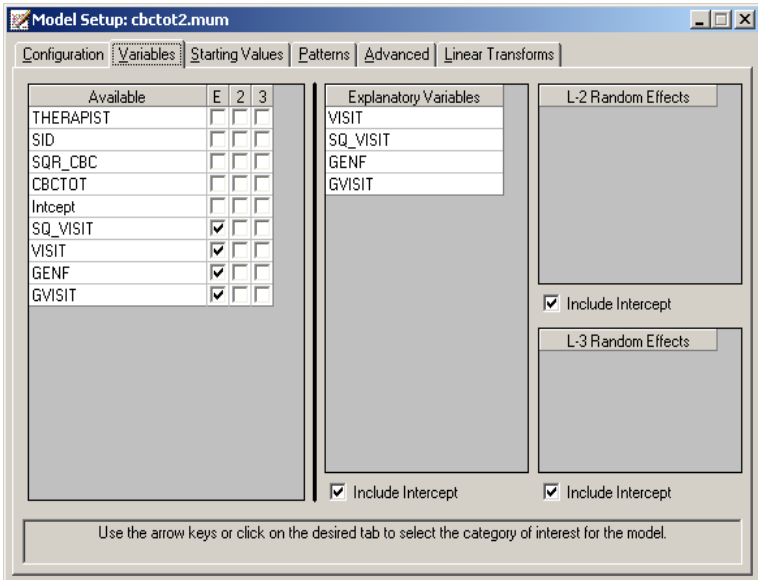
The model considered in this section uses the participant's gender, visit number, squared value of the visit number, and the interaction between visit number and gender (represented by the variable GVISIT in the data spreadsheet) to predict the square root of the total score on the Child Behavior Checklist. This second order growth curve with gender and the interaction term as covariates can be expressed as follows:

$$\begin{aligned}\mathrm{SQR\_CBC}_{ijk} = {} & \beta_0 + \beta_1 * \mathrm{GENF}_{ij} + \beta_2 * \mathrm{VISIT}_{ijk} + \beta_3 * \mathrm{SQ\_VISIT}_{ijk} + \beta_4 * \mathrm{GENF}_{ij} * \mathrm{VISIT}_{ijk} \\ & + v_{i0} + v_{ij0} + e_{ijk}\end{aligned}$$

As before, β_0 denotes the average expected total score, $\beta_1, \beta_2, \ldots, \beta_4$ indicate the estimated coefficients associated with the fixed part of the model, and v_{i0}, v_{ij0} and e_{ijk} represent the random part of the model.

3.4.4.2 Setting up the analysis

The SuperMix spreadsheet **cbtot.ss3** and the model specification file **cbtot.mum** discussed in the previous example are used a point of departure. With the model specification file open, click on the **Variables t**ab of the **Model Setup** window. Add the predictors GENF and GVISIT to the model by checking the boxes next to these variables in the **E** column, as shown below.

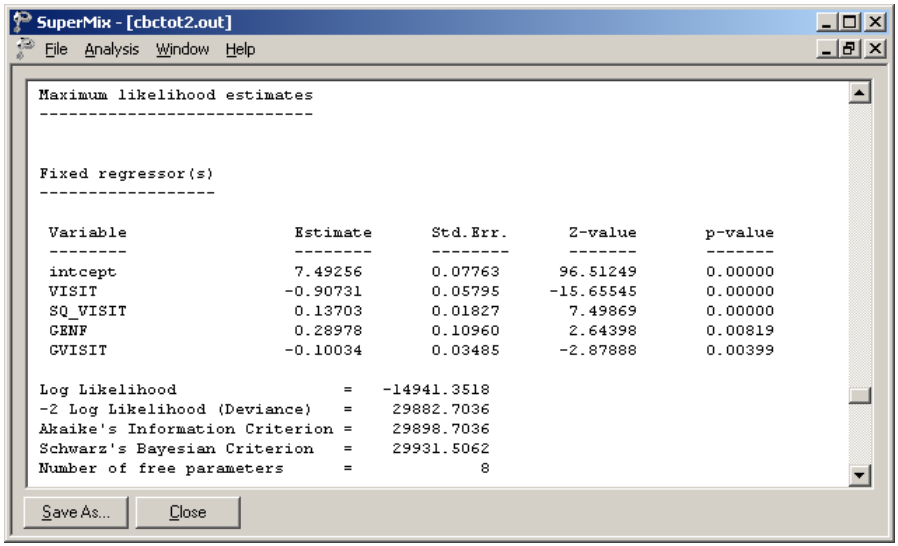


Save the modified model specification file as **cbtot2.mum**, then select the **Run** option from the **Analysis** menu to perform the analysis.

3.4.4.3 Discussion of results

Fixed effects results

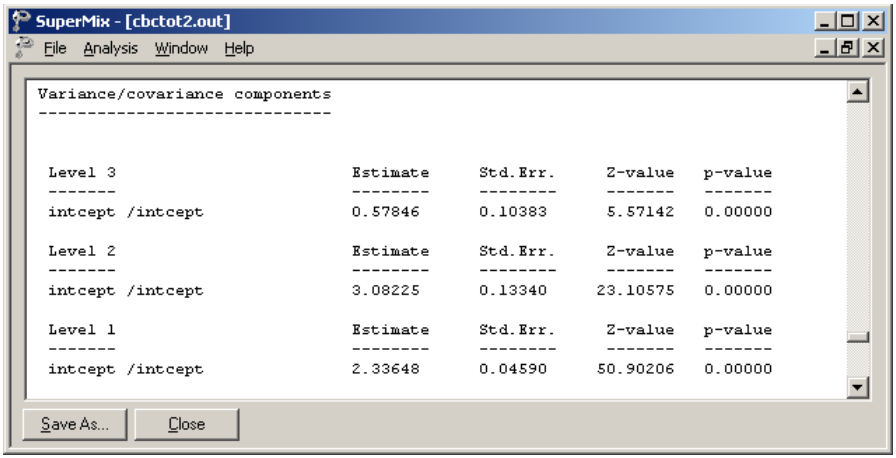
The maximum likelihood estimates of the coefficients in the fixed part of the model are shown below.



The statistical significance of all the effects confirm our suspicion that the CBC scores measured over time not only depend on the time of measurement and the squared value thereof, but also on the gender of the patient. A significant interaction between gender and the time of measurement is also observed. Recall that for male patients GENF was coded 0, and for females GENF was assigned a value of 1. The positive estimate of 0.28977 for the effect of gender indicates that males tended to have a lower score on average than females: the expected average male score is 0.28977 units lower on the transformed CBC scores than for females. This effect is offset by the negative estimate of the interaction effect. For males, the interaction term GVISIT assumes the value 0, but for females GVISIT is equal to 0, 1, 2, and 3 respectively at the 4 measurement occasions. The transformed score of a female patient is thus expected to be 0.10034 units lower at the second visit than the score of a male patient, or a female patient at the beginning of the study.

Random effects results

The output for the **random part** of the model is given next.



As before, most variation in scores is found at a patient level, and the least variation at the therapist level. The estimated percentages of variation in outcome at patient and therapist level are

$$\frac{3.08225}{0.57846+3.08225+2.33648}\times 100\% = 51.39\%$$

and

$$\frac{0.57846}{0.57846+3.08225+2.33648}\times 100\% = 9.65\%$$

respectively. When compared to the similar percentages for the growth curve model, changes observed are negligible. The addition of the variables GENF and GVISIT did not contribute significantly to the explanation of remaining variation in the outcome at the various levels of the model.

3.4.4.4 Interpretation of the results

Estimated outcomes for different groups

For a typical patient, the expected CBC score can be calculated as

$$\begin{aligned}\widehat{\text{SQR\_CBC}}_{ijk} &= \hat{\beta}_0 + \hat{\beta}_1 * \text{GENF}_{ij} + \hat{\beta}_2 * \text{VISIT}_{ijk} + \hat{\beta}_3 * \text{SQ\_VISIT}_{ijk} + \hat{\beta}_4 * \text{GVISIT}_{ijk} \\ &= 7.49255 + 0.28977 * \text{GENF}_{ij} - 0.90730 * \text{VISIT}_{ijk} \\ &\quad + 0.13703 * \text{SQ\_VISIT}_{ijk} - 0.10034 * \text{GVISIT}_{ijk}.\end{aligned}$$

For male patients GENF = 0, and thus the formula used to predict their CBC scores reduces to

$$\begin{aligned}\widehat{\text{SQR\_CBC}}_{ijk} &= 7.49255 + 0.28977 * (0) - 0.90730 * \text{VISIT}_{ijk} \\ &\quad + 0.13703 * \text{SQ\_VISIT}_{ijk} - 0.10034 * (0) \\ &= 7.49255 - 0.90730 * \text{VISIT}_{ijk} + 0.13703 * \text{SQ\_VISIT}_{ijk}.\end{aligned}$$

For female patients GENF = 1, and thus the formula used to predict their CBC scores can be expressed as

$$\begin{aligned}\widehat{\text{SQR\_CBC}}_{ijk} &= 7.49255 + 0.28977 * (1) - 0.90730 * \text{VISIT}_{ijk} \\ &\quad + 0.13703 * \text{SQ\_VISIT}_{ijk} - 0.10034 * \text{GVISIT}_{ijk} \\ &= 7.78232 - 0.90730 * \text{VISIT}_{ijk} + 0.13703 * \text{SQ\_VISIT}_{ijk} - 0.10034 * \text{GVISIT}_{ijk}.\end{aligned}$$

Table 3.9 below shows the expected square roots of CBC scores for the various groups formed by the gender groups and interaction term at all measurement occasions. In Table 3.10, the same expected scores are given in the scale of the original total score on the Child Behavioral Checklist. The initial impression, based on the positive coefficient of GENF, that females had higher expected CBC scores than males, seems to hold at the onset of the study. However, these tables show that,

after the effects of the other variables are also taken into account, females are likely to have a slightly lower score than males at the end of the study period.

Table 3.9: Expected square root of CBC scores

Gender	Visit			
	0	1	2	3
Male	7.4926	6.7223	6.2261	6.0039
Female	7.7823	6.9117	6.3152	5.9927

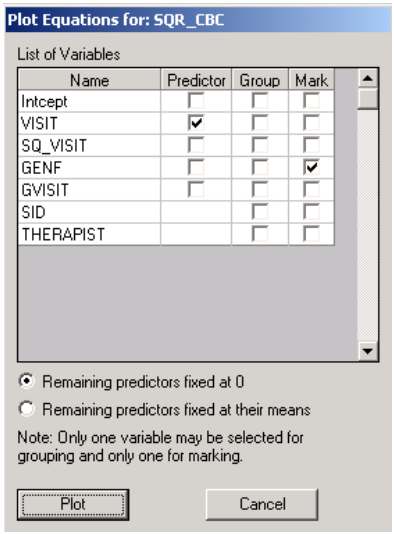
Table 3.10: Expected CBC scores in original scale

Gender	Visit			
	0	1	2	3
Male	56.1385	45.1890	38.7638	36.0468
Female	60.5648	47.7719	39.8812	35.9120

The results in these tables can also be depicted graphically using the **File**, **Model-based Graphs** menu. This menu offers three options, namely equation modeling, residual plots and confidence intervals for random effects.

Equation modeling

To plot the trends in CBC scores for gender groups over successive visits, make sure the **Model Setup** window is activated by clicking on it before select the **Equations** option from the **File**, **Model-based Graphs** menu.



This activates the **Plot Equations for: SQR_CBC** dialog box. Select VISIT as the predictor, and request marking by gender as shown in Figure 3.25. Note that, by default, remaining predictors are fixed at 0. Click **Plot** to display the graphing window.

By default, graphs using a two-category marking variable such as GENF will be displayed using blue and green to indicate the categories. To make the distinction between the groups of interest more clear, and create a graph that can be included in a report or paper to be printed in black and white, the plotting symbols can be changed. Here, we opt to change the line for female patients to a black, dotted line. Double-click on the line for this group in the legend box shown at the top right of the graph to activate the **Plot Parameters** dialog box. Next, click the **Line Attributes** button to load the **Line Parameters** dialog box.

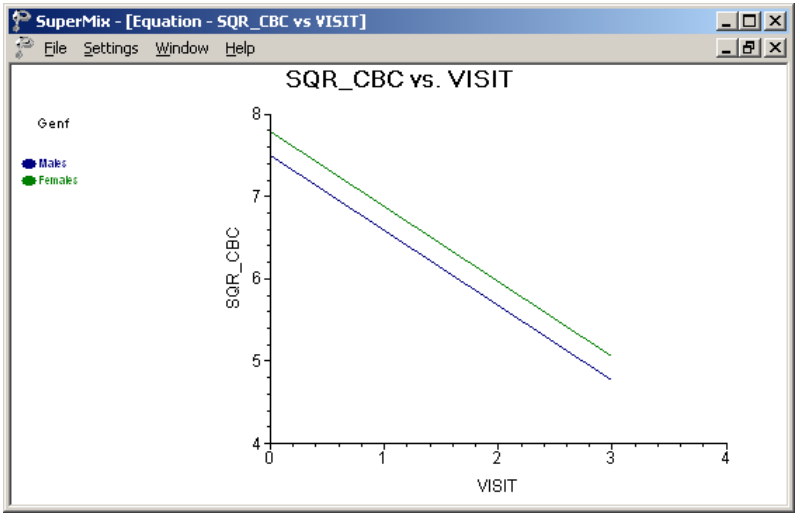
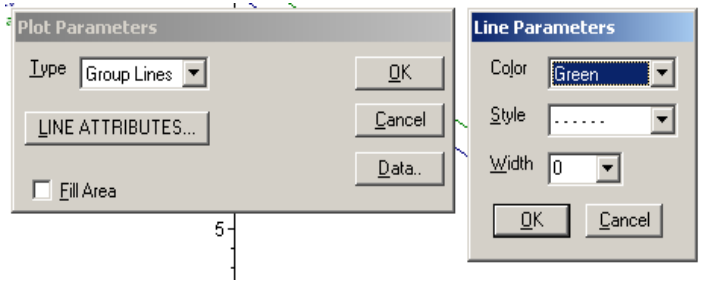


Figure 3.25: Plot of SQR_CBC vs. VISIT for gender groups

Change the line style to dotted using the **Style** drop-down list box, and select black from the **Color** drop-down list box. Click **OK** on both the **Line Parameters** and **Plot Parameters** dialog boxes to obtain the final graph shown below.



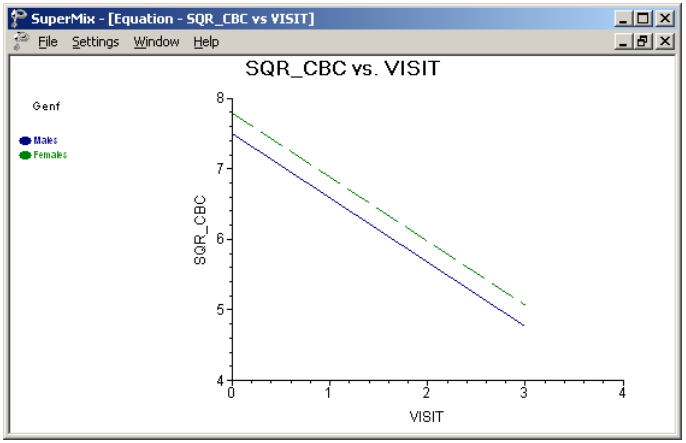


Figure 3.26: Modified plot of SQR_CBC vs. VISIT for gender groups

The predicted decrease in CBC score echoes the results of the maximum likelihood estimation of the fixed effects, where a negative coefficient of –0.9073 was reported for the predictor VISIT. While the predicted intercept for males at the beginning of the study is approximately 7.5 as indicated in the graph at the top-left of the graphing window, the predicted intercept for the same group has decreased to approximately 4.75 by the final visit. This is lower than reported in Table 3.10, where calculations showed an expected CBC score of 6.00 for males by the final visit. The reason for this difference can be found in the formula used to produce the graph: recall that all remaining predictors were fixed to a value of 0. Whereas the result for males at the final visit shown in Table 3.10 was based on the calculation

$$\begin{aligned}\widehat{\text{SQR\_CBC}}_{ijk} &= 7.49255 + 0.28977 * (0) - 0.90730 * \text{VISIT}_{ijk} \\ &\quad +0.13703 * \text{SQ\_VISIT}_{ijk} - 0.10034 * (0) \\ &= 7.49255 - 0.90730 * \text{VISIT}_{ijk} + 0.13703 * \text{SQ\_VISIT}_{ijk},\end{aligned}$$

the line shown for this group in the graph above is based on the formula

$$\widehat{\text{SQR\_CBC}}_{ijk} = 7.49255 + 0.28977 * (0) - 0.90730 * \text{VISIT}_{ijk}.$$

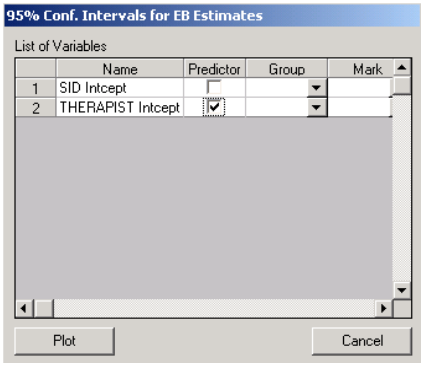
As a result, the predicted outcome shown in the graph for males at the end of the study will be $(0.13703)(9) = 1.2333$ units lower than reported in Table 3.10. This

difference underlines the fact that care should be taken when selecting the treatment for remaining predictors in the model. In this case, both SQ_VISIT and GVISIT can assume the value of 0, and thus using the remaining predictors fixed at zero option is permissible. In cases where predictors cannot assume a value of zero, the better choice would be to fix remaining predictors at their mean values instead when completing the **Plot Equations for:** dialog box.

Confidence intervals for random coefficients

The **Confidence Intervals** option on the **File**, **Model-based Graphs** menu provides the option to display confidence intervals for the empirical Bayes estimates of the random effects specified in a given model. This option is now used to examine the confidence intervals of the random intercepts for the therapists, who represent the highest level of the hierarchy in the current example.

Select the **Confidence Intervals** option on the **File**, **Model-based Graphs** menu to activate the **95% Conf. Intervals for EB estimates** dialog box. A simple graph of the confidence intervals for the empirical Bayes estimates of the intercepts at the therapist level is obtained by selecting THERAPIST Intcept in the **Predictor** column before clicking **Plot**. Note that it is also possible to select both a grouping and marking variable to be used in the graph.



The graph obtained, as seen below, shows that, in general, the range of the confidence intervals for the level-3 empirical Bayes estimates of the intercepts is $(-2; 2)$. The deviations from the estimated population intercept over therapists are also apparent.

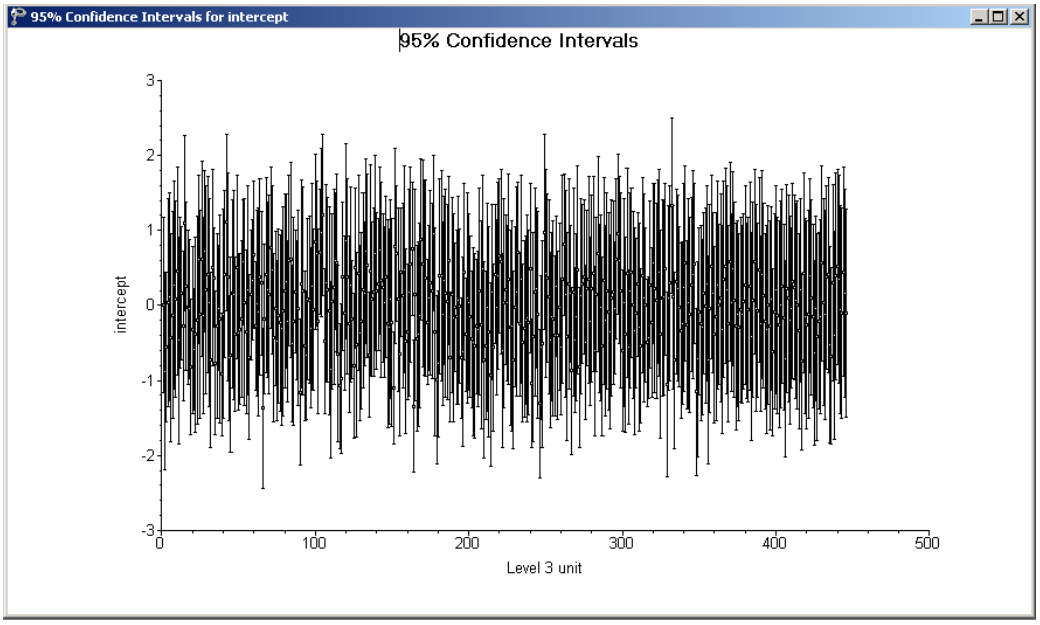


Figure 3.27: 95% confidence intervals for level-3 units

Each confidence interval is obtained using

$$Empirical\ Bayes\ residuals \pm 1.96\sqrt{var\left(Empirical\ Bayes\ residuals\right)}\ .$$

Fit statistics

Recall that for the growth curve model the following indices were obtained:

- Log Likelihood: –14946.6389
- –2 Log Likelihood (Deviance): 29893.2778

- Akaike's Information Criterion: 29905.2778
- Schwarz's Bayesian Criterion: 29929.8797
- Number of free parameters: 6

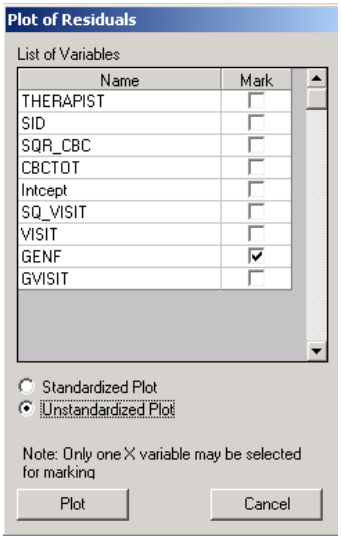
When the deviances of the two models are compared, a χ^2-statistic of 29893.2778 – 29882.7036 = 10.57 with 8 – 6 = 2 degrees of freedom is obtained. This indicates that the current model fits the data better than the growth curve model. The AIC decreased from 29905.2778 to 29898.7036, and also favors the use of the current model. The SBC, however, increased slightly, from 29929.8797 to 29931.5062, and thus favors the growth curve model previously fitted as the more parsimonious. Note, however, that the changes in all three criteria are rather small.

3.4.4.5 Residuals

Residual plots

The **Residuals** option on the **File**, **Model-based Graphs** menu is used to examine the residuals obtained for a fitted model. It is useful for examining the fit of the model, and also as a check for possible distributional assumption violations. As residuals are defined as the difference between the observed and predicted outcomes, trends in residuals, for example over the course of a study in a longitudinal data set, may indicate that an important predictor was not included in the model fitted to the data.

Select the **Residuals** option on the **File**, **Model-based Graphs** menu to activate the **Plot of Residuals** dialog box. To simultaneously check for any differences in residuals for the gender groups, select GENF as marking variable. Opt to create an unstandardized plot of the residuals by selecting the **Unstandardized Plot** option rather than the default **Standardized Plot** option. Click **Plot**.



The graph below shows the residuals for the gender groups in the default colors of blue and green. To make the distinction between the groups more clear, click on the plotting symbol for the female group in the legend box.

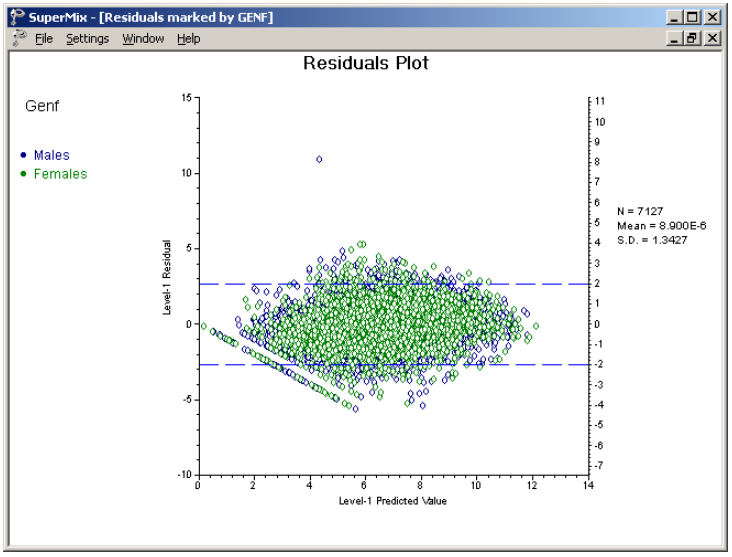
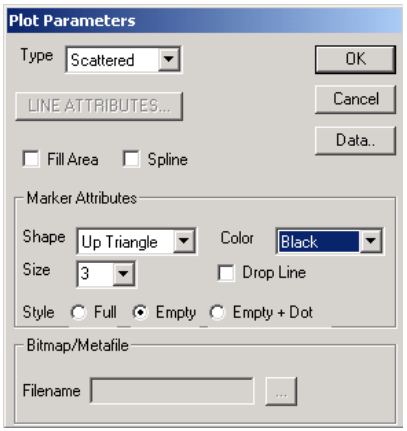
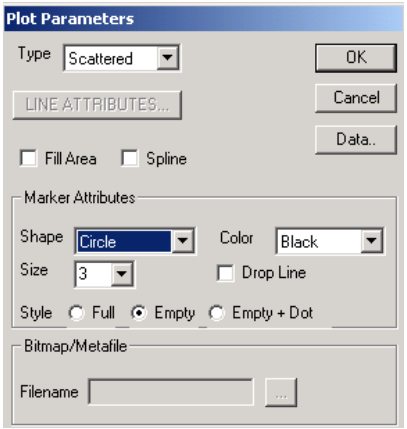


Figure 3.28: Level-1 residual plot by gender group

The **Plot Parameters** dialog box appears. Change the **Shape** of the symbol to "Up Triangle," adjust the **Size** to 3 and change the display **Color** to black as shown below. Click **OK** when done.



Click on the symbol for the male group next, and change the parameters for this group to those shown in the dialog box below. Click **OK** to return to the graphing window.



The final plot is shown below. The residuals are clustered reasonably symmetrically around the 0 tick mark on the Y-axis, and no gender pattern can be discerned for the larger residuals. A single residual, for a male respondent, has a value larger than 10. This potential outlier can be identified using the **Data** option on the **Plot Parameters** dialog box (see above).

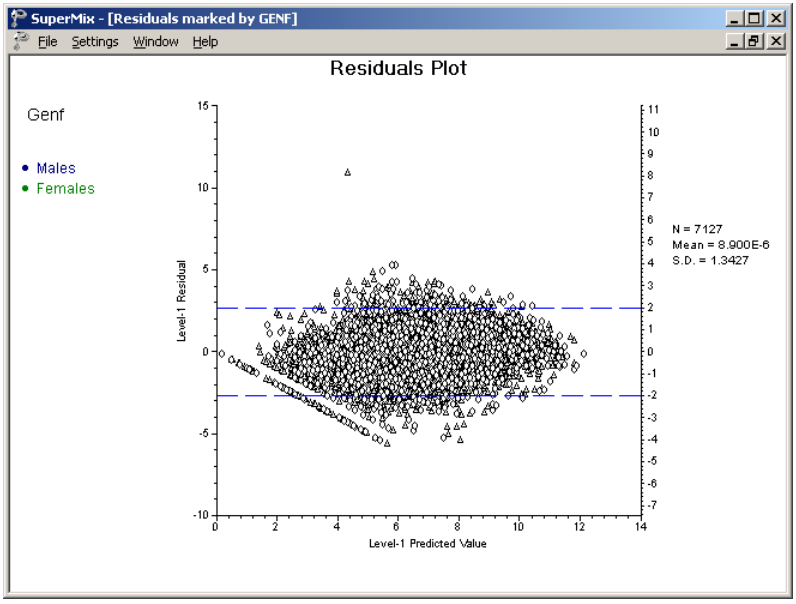


Figure 3.29: Modified level-1 residual plot by gender group

3.4.5 Fitting a random intercepts and slopes model

3.4.5.1 The model

The graphs obtained during the exploratory analysis of the CBC data showed that the change in total CBC score over the course of the study differed from patient to patient. Because of this, the models fitted in Sections 3.4.3 and 3.4.4 allowed for the intercepts to vary randomly at both patient and therapist level. In effect, we assumed that each patient may have a different starting point. These models indicated a statistically significant relationship between the observed CBC score and the measurement occasion. To test whether there is significant variation in the way

individual patients' scores change over the study period, a model in which both intercepts and slopes of the predictor VISIT are allowed to vary randomly can be used.

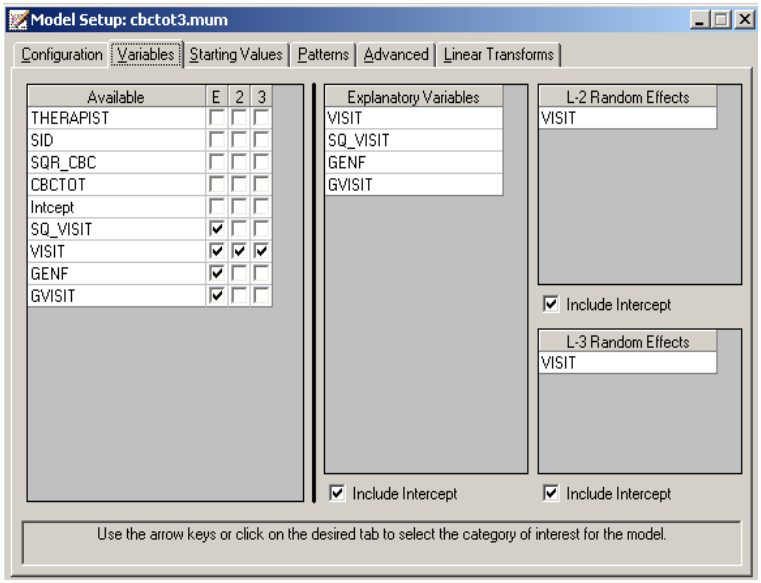
The model can be formulated as

$$\begin{aligned}\text{SQR\_CBC}_{ijk} = {} & \beta_0 + \beta_1 * \text{GENF}_{ij} + \beta_2 * \text{VISIT}_{ijk} + \beta_3 * \text{SQ\_VISIT}_{ijk} + \beta_4 * \text{GENF}_{ij} * \text{VISIT}_{ijk} \\ & + (v_{i0} + v_{i1} * \text{VISIT}_{ijk}) + (v_{ij0} + v_{ij1}\text{VISIT}_{ijk}) + e_{ijk}\end{aligned}$$

At level 2, two random coefficients are now included: v_{ij0} represents the random intercept and v_{ij1} the random coefficient for the slope of the predictor VISIT. The random coefficients v_{i0} and v_{i1} serve the same purpose at level 3 (the therapist level) of the model.

3.4.5.2 Setting up the analysis

Again, we use the SuperMix spreadsheet **cbtot.ss3** and the model specification file **cbtot.mum** discussed in the previous example as the starting point. With the model specification file open, click on the **Variables** tab of the **Model Setup** window. Add random coefficients for the predictor VISIT to levels 2 and 3 of the model by checking the boxes next to these variables in the **2** and **3** column, as shown below. Save the changes to the model specification file, using the **File**, **Save** option to overwrite the previous specification file or the **File**, **Save as** option to assign a new filename. Click **Run** on the **Analysis** menu to perform the analysis.

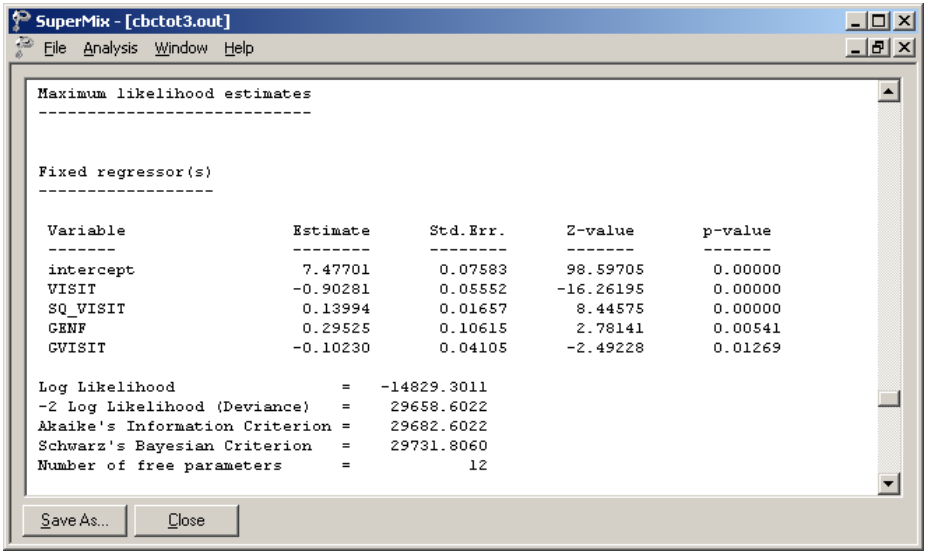


3.4.5.3 Discussion of results

Partial output is given below. We focus on those parts of the output that differ from the output obtained for the previous analysis, and conclude with a discussion of the additional output files containing the empirical Bayes residuals.

Fixed effects results

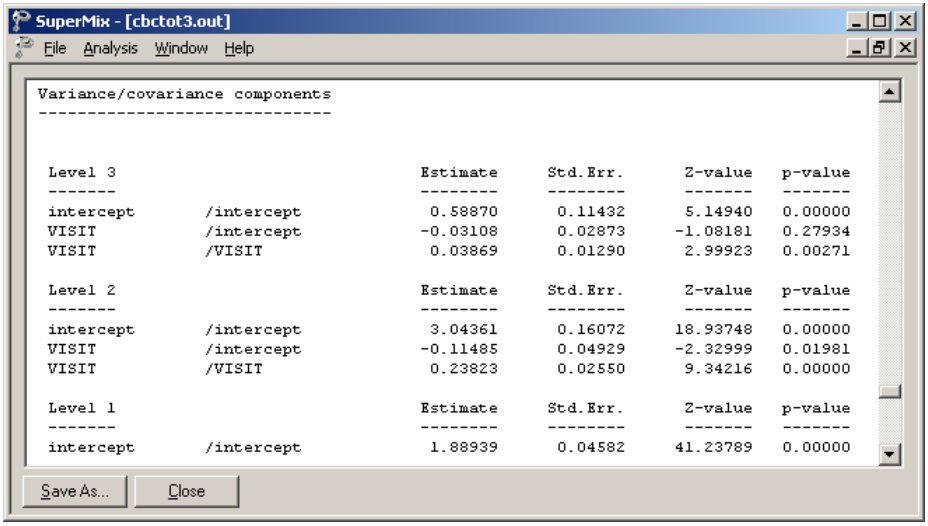
The inclusion of random VISIT slopes at levels 2 and 3 of the hierarchy has very little impact the estimated fixed coefficients. Results for the fixed part are shown below.



Random effects results

Turning to the estimated coefficients in the random part of the model, we note a change in the between measurement (level-1) variation, which has decreased from 2.33648 to 1.88939. This illustrates that the addition of a random coefficient at any level of a model can affect the random effect(s) at another level.

At levels 2 and 3 we find evidence of significant variation in the VISIT slopes. While not of the same magnitude as the intercept variation, it is clear that it is more realistic to allow the slopes to vary from patient to patient, and from therapist to therapist, than to assume that the VISIT slope can be described adequately by a common fixed effect as was done in the previous model.



3.4.5.4 Interpreting the results

Fit statistics and ICC

Model fit

When the measures of fit are compared to those of the random-intercepts-only model, it becomes clear that the current model fits the data better. Recall that for the random intercepts model the following fit measures were obtained:

- Log likelihood: –14941.3518
- –2 log Likelihood (Deviance): 29882.7036
- Akaike's Information Criterion: 29898.7036
- Schwarz's Bayesian Criterion: 29931.5062
- Number of free parameters: 8

While four more parameters had to be estimated in the random intercepts and slopes model, the deviance decreased significantly. The χ^2-statistic for comparing these models is 29882.7036 – 29658.6022 = 223.9185, with 4 degrees of freedom. The improved fit of the current model is also clear from the other fit measures: both the

AIC and the SBC clearly favor the current model, and have decreased substantially from those reported for the random-intercepts-only model.

Percentage variation explained

To take a closer look at the amount of variation explained at the levels of the hierarchy, the total variation at each level has to be calculated. At level 3, we have three variance/covariance components to take into account. Recall that the model included two random effects, namely v_{i0} and $v_{i1}(VISIT_{ijk})$. The total variation at this level follows as

$$\begin{aligned} Var(level-3) &= \text{var}(v_{i0} + v_{i1}(\text{VISIT}_{ijk})) \\ &= \text{var}(v_{i0}) + \text{var}(v_{i1}(\text{VISIT}_{ijk})) + \text{cov}(v_{i0}, v_{i1}(\text{VISIT}_{ijk})) \\ &= \text{var}(v_{i0}) + (\text{VISIT}_{ijk})^2\ \text{var}(v_{i1}) + 2(\text{VISIT}_{ijk})\,\text{cov}(v_{i0}, v_{i1}) \\ &= 0.58870 + 0.03869(\text{VISIT}_{ijk})^2 - 2(0.03108)(\text{VISIT}_{ijk}) \\ &= 0.58870 + 0.03869(\text{VISIT}_{ijk})^2 - 0.06216(\text{VISIT}_{ijk}) \end{aligned}$$

At level 2, the total variation can be expressed in similar fashion as

$$\begin{aligned} Var(level-2) &= \text{var}(u_{ij0} + u_{ij1}(\text{VISIT}_{ijk})) \\ &= \text{var}(u_{ij0}) + (\text{VISIT}_{ijk})^2\ \text{var}(u_{ij1}) + 2(\text{VISIT}_{ijk})\,\text{cov}(u_{ij0}, u_{ij1}) \\ &= 3.04361 + 0.23823(\text{VISIT}_{ijk})^2 - 0.2297(\text{VISIT}_{ijk}) \end{aligned}$$

The total variation in the model is

$$\begin{aligned} Total\ Var &= \text{var}(level-1) + \text{var}(level-2) + \text{var}(level-3) \\ &= 5.5217 + 0.27692(\text{VISIT}_{ijk})^2 - 0.29186(\text{VISIT}_{ijk}) \end{aligned}$$

The variation at the higher levels and, consequently, the total variation are a function of the measurement occasion, as represented by the predictor VISIT. For example, at the start of the study we find that the total variation is equal to 5.5217, with 0.58870 at level 3 and 3.04361 at level 2. This indicates that at the time of the first visit,

$$\frac{0.58870}{5.5217} \times 100 = 10.66\%$$

of the total variation explained by this model is at a therapist level. By the end of the study, VISIT assumes a value of 3, and thus the total variation is equal to 7.1384. The total variation at the therapist level at the last measurement occasion is 0.75043, and thus the percentage of variation at therapist level at the end of the study is

$$\frac{0.75043}{7.1384} \times 100 = 10.51\%.$$

At a patient level, the corresponding percentages of variation at the first and last visit are

$$\frac{3.04361}{5.5217} \times 100 = 55.12\%.$$

and

$$\frac{4.49858}{7.1384} \times 100 = 63.02\%$$

respectively. While the total variation explained at a therapist level declines over visits, there is an increase of approximately the same size in the total variation explained at a patient level over visits. The variation over patients is consistently much higher than over therapists or over measurements nested within patients.

4 Models for binary outcomes

4.1 Introduction

The nominal and ordinal outcome models can be seen as generalizations of the binary outcome model. In order to understand these models, an understanding of the binary outcome model is required.

A binary random variable is a discrete random variable that has only two possible values, such as whether a subject dies (event) or lives (non-event). Such events are often described as success versus failure, and coded using the values 0 or 1. Consequently, the assumption that this type of outcome variable has a normal distribution does not hold anymore.

The most common distribution used for a binary outcome is the Bernoulli distribution, which takes a value 1 with probability of success p and a value 0 with probability of failure $q = 1 - p$. The selection of the distribution for the outcome variable is not fixed. For example, if the occurrence is very rare, the Poisson distribution can be used.

4.1.1 Link functions

In the case of a binary variable, observed values are usually assigned as either 0 or 1. When such a variable is treated as if it were continuous, predicted values, indicating the probability of the event occurring, can fall outside the (0,1) interval. Moreover, the assumption of normality at level 1 is not realistic as the random effects can no longer be assumed to have a normal distribution or to have homogeneous variance.

The multilevel generalized linear model (MGLM) generalizes the multilevel model for continuous outcomes by additionally allowing for error distributions from the

exponential family (see, for example, McCullagh & Nelder, 1989). Let y denote the outcome variable, and $E(y)$ the expected value of y. The key to MGLM models is that a nonlinear relationship between $E(y)$ and β is allowed, with the aid of a link function.

Suppose that $\mathbf{x} = (x_1 \ldots x_n)$ is the vector of all the predictors and that $\boldsymbol{\beta} = (\beta_1 \ldots \beta_n)$ is the vector of unknown regression parameters. In the models discussed up to now, it was assumed that the outcomes were normally distributed variables and that a model of the form

$$y_{ij} = \mathbf{x}_{ij}^{'}\boldsymbol{\beta} + \mathbf{z}_{ij}^{'}\mathbf{v}_i + e_{ij}, \quad j = 1, 2, ..., n_i$$

could be used to describe the relationship between the outcome and predictor variables. The vector $\mathbf{z}_{ij}^{'}$ denotes a design vector for the random effects contained in the vector $\mathbf{v}_i$, and $\mathbf{x}_{ij}^{'}$ is the design vector for the predictors in the fixed part of the model with corresponding vector $\boldsymbol{\beta}$ of regression parameters. The covariance matrix of $\mathbf{v}_i$ is denoted by $\boldsymbol{\Phi}_{(2)}$ and the variance of e_{ij} by σ_e^2.

The link function specifies a nonlinear transformation between the linear predictor η and the assumed distribution function. These link functions transform the observed outcome value to a function $\eta = \mathbf{x}'\boldsymbol{\beta}$ and ensure that the predicted probability lies within the (0,1) interval. Instead of y, η is being analyzed. For the binary outcome, the probability of success η is the predictor of interest.

The most commonly used link functions are the log, logit, probit and complementary log-log link functions. The log link generally is used for the count variable with Poisson distribution, which will be discussed in the next chapter. The link functions available in SuperMix include the logit, probit and complementary log-log functions for models with an ordinal dependent variable, and the logit link function for models with a nominal dependent variable. Table 4.1 shows these link functions, along with their distribution functions (CDF), means and variances.

Table 4.1: Link functions for the Bernoulli distribution

Link name	Link function $F^{-1}(p)$, $0 < p < 1$	CDF $-\infty < w < \infty$	Mean	Variance
logit (logistic)	$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right)$	$\frac{e^w}{1+e^w}$	0	$\frac{\pi^2}{3}$
probit	$\Phi^{-1}(p)$, where Φ^{-1} is the inverse of the standard normal cumulative distribution	$\Phi(w)$	0	1
complementary log-log	$\log(-\log(1-p))$	$1\text{-}\exp(-\exp(w))$	-0.577	$\frac{\pi^2}{6}$

These link functions map the probability η with an open interval (0,1) to the entire set of real numbers $\mathbb{R}$. Figure 4.1 illustrates how a real number *w* is transformed to the probability η.

As shown below, the logit and probit link functions are both symmetric around a value of 0. The logit function has a larger variance. The complementary log-log link function is asymmetric. When the probability of a successful outcome (p) is extremely small or large, the linear relationship does not hold. Understanding the nature of the link function used in an analysis is essential to the correct interpretation of the results.

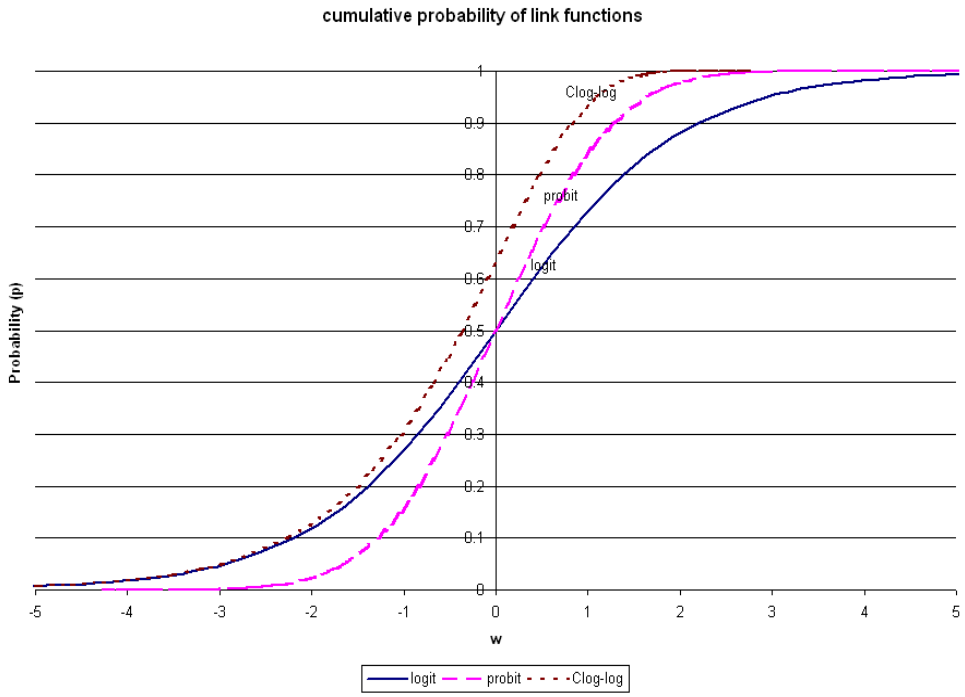


Figure 4.1: Cumulative density of link functions

In this chapter, we will consider examples of two- and three-level models based on two data sets, both with binary outcome variables. These data will also be used to illustrate the ordinal outcome model in Section 6.2.

4.1.2 Methods of estimation

For models with binary, ordinal, count, and nominal outcomes, SuperMix offers two methods of estimation: maximization of the posterior distribution (MAP) and numerical integration (adaptive and non-adaptive quadrature) to obtain parameter and standard error estimates.

The MAP method of estimation can be used to obtain a point estimate of an unobserved quantity on the basis of empirical data. It is closely related to Fisher's

method of maximum likelihood (ML), but employs an augmented optimization objective which incorporates a prior distribution over the quantity one wants to estimate.

Quadrature is a numeric method for evaluating multi-dimensional integrals. For mixed effect models with count and categorical outcomes, the log-likelihood function is expressed as the sum of the logarithm of integrals, where the summation is over higher-level units, and the dimensionality of the integrals equals the number of random effects.

Typically, MAP estimates are used as starting values for the quadrature procedure. When the number of random effects is large, the quadrature procedures can become computationally intensive. In such cases, MAP estimation is usually selected as the final method of estimation. Numerical quadrature, as implemented in SuperMix, offers users a choice between adaptive and non-adaptive quadrature. Quadrature uses a quadrature rule, *i.e.*, an approximation of the definite integral of a function, usually stated as a weighted sum of function values at specified points within the domain of integration.

Adaptive quadrature generally requires fewer points and weights to yield estimates of the model parameters and standard errors that are as accurate as would be obtained with more points and weights in non-adaptive quadrature. The reason for that is that the adaptive quadrature procedure uses the empirical Bayes means and covariances, updated at each iteration to essentially shift and scale the quadrature locations of each higher-level unit in order to place them under the peak of the corresponding integral.

A full description of these methods is given in Chapter 10. A brief description of MAP estimation and quadrature follows below.

MAP estimation

For level-2 unit i, let $v_{i1}, v_{i2}, ..., v_{ir}$ denote the random effects and $y_{i1}, y_{i2}, ..., y_{in_n}$ the outcomes. Let $f(\mathbf{v}_i, \mathbf{y}_i)$ denote the joint distribution of $\mathbf{v}_i = (v_{i1}, v_{i2}, ..., v_{ir})$ and $\mathbf{y}_i = (y_{i1}, y_{i2}, ..., y_{in_i})$.

Using standard results for conditional distributions, it follows that

$$f(\mathbf{v}_i \mid \mathbf{y}_i) = f(\mathbf{y}_i \mid \mathbf{v}_i) f(\mathbf{v}_i) / f(\mathbf{y}_i).$$

By taking logarithms on both sides of the equation, the following density function is obtained:

$$\ln f(\mathbf{v}_i \mid \mathbf{y}_i) = \ln f(\mathbf{y}_i \mid \mathbf{v}_i) + \ln f(\mathbf{v}_i) - K$$

where K is a constant. Mode estimates $\hat{v}_i$ of the random effects and estimates $\hat{\boldsymbol{\beta}}$ of the fixed parameters are obtained by iteratively solving the equations

$$\frac{\partial}{\partial \gamma_k} \ln f(\mathbf{v}_i \mid \mathbf{y}_i) = 0,$$

where γ_k is a typical element of the unknown parameters $v_{i1}, v_{i2}, ..., v_{ir}$ and $\beta_1, \beta_2, ..., \beta_p$.

As a by-product of the iterative procedure, estimates of $\operatorname{cov}\left(\hat{v}_i\right), i = 1, 2, ..., N$ are obtained and these, in turn, are used to estimate $\Phi_{(2)} = \operatorname{cov}(\mathbf{v}_i)$.

Numerical quadrature

Since

$$f\left(\mathbf{y}_i, \mathbf{v}_i\right) = f\left(\mathbf{y}_i \mid \mathbf{v}_i\right) f\left(\mathbf{v}_i\right)$$

it follows that the marginal distribution of $\mathbf{y}_i$ can be obtained as the solution to the multi-dimensional integral

$$f\left(\mathbf{y}_i\right) = \int_{v_1} \ldots \int_{v_r} f\left(\mathbf{y}_i \mid \mathbf{v}_i\right) f\left(\mathbf{v}\right) dv_1 \ldots . dv_r .$$

Since it is assumed that $\mathbf{v}_i \sim N\left(\mathbf{0}, \mathbf{\Phi}_{(2)}\right)$ it follows, for example, that

$$f\left(\mathbf{v}_i\right) = \left(2\pi\right)^{-r/2} \mid \mathbf{\Phi}_{(2)} \mid^{-1/2} \exp\left[-\frac{1}{2} \mathbf{v}_i^{'} \mathbf{\Phi}_{(2)}^{-1} \mathbf{v}_i\right].$$

In general, a closed-form solution to this integral does not exist. To evaluate integrals of the type described above, we use a direct implementation of Gauss-Hermite quadrature (see, *e.g.*, Krommer & Ueberhuber, 1994, Section 4.2.6 and Stroud & Sechrest, 1966, Section 1).

With this rule, an integral of the form

$$I(t) = \int f(t) \exp\left[-t^2\right] dt$$

is approximated by the sum

$$I(t) \approx \sum_{u=1}^{Q} w_u f\left(z_u\right),$$

where w_u and z_u are weights and nodes of the Hermite polynomial of degree Q. A Q-point adaptive quadrature rule is a quadrature rule constructed to yield an exact result for polynomials of degree $2Q-1$, by a suitable choice of the n points x_i and n weights w_i.

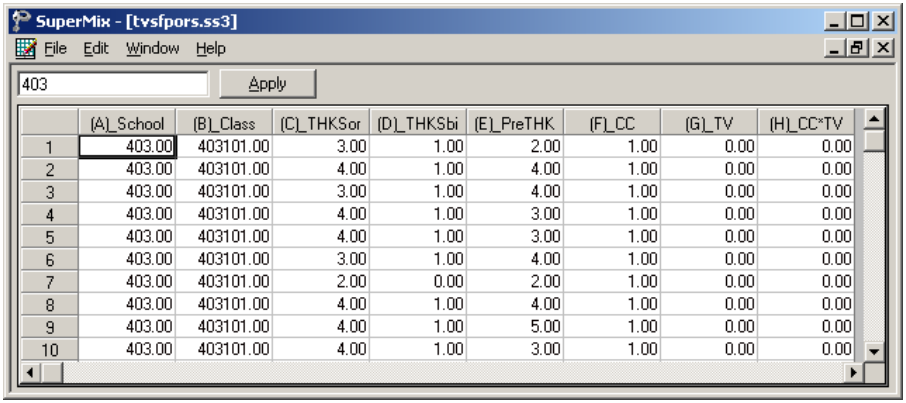
4.2 Models based on a subset of the TVSFP data

4.2.1 The data

The data are from the Television School and Family Smoking Prevention and Cessation Project (TVSFP) study (Flay, *et. al.*, 1988) described in Section 3.3. The study was designed to test the independent and combined effects of a school-based social-resistance curriculum and a television-based program in terms of tobacco use and cessation.

A tobacco and health knowledge scale was used in classifying subjects as knowledgeable or not. In its original form, the student's score was defined as the number of correct answers to seven items on tobacco and health knowledge. The structure of this study indicates a three-level hierarchical structure. However, we will first consider two two-level structures. In the first, students are nested within schools; in the second, students are nested within classrooms. Finally, a three-level model recognizing the role of both classroom and school in the hierarchical structure of the data will be considered.

Data for the first 10 participants on most of the variables used in this section are shown below in the form of a SuperMix spreadsheet file, named **tvsfpors.ss3**, located in the **Examples\Binary** subfolder.



	(A)_School	(B)_Class	(C)_THKSor	(D)_THKSbi	(E)_PreTHK	(F)_CC	(G)_TV	(H)_CC*TV
1	403.00	403101.00	3.00	1.00	2.00	1.00	0.00	0.00
2	403.00	403101.00	4.00	1.00	4.00	1.00	0.00	0.00
3	403.00	403101.00	3.00	1.00	4.00	1.00	0.00	0.00
4	403.00	403101.00	4.00	1.00	3.00	1.00	0.00	0.00
5	403.00	403101.00	4.00	1.00	3.00	1.00	0.00	0.00
6	403.00	403101.00	3.00	1.00	4.00	1.00	0.00	0.00
7	403.00	403101.00	2.00	0.00	2.00	1.00	0.00	0.00
8	403.00	403101.00	4.00	1.00	4.00	1.00	0.00	0.00
9	403.00	403101.00	4.00	1.00	5.00	1.00	0.00	0.00
10	403.00	403101.00	4.00	1.00	3.00	1.00	0.00	0.00

The variables of interest are:

- School indicates the school a student is from (28 schools in total).
- Class identifies the classroom (135 classrooms in total).
- THKSord represents the tobacco and health knowledge scale, with 4 categories ranging between 1 and 4. The frequency distribution of the post-intervention THKS scores indicated that approximately half the students had scores of 2 or less, and half of 3 or greater. In terms of quartiles, four ordinal classifications were suggested corresponding to 0 – 1, 2, 3, and 4 – 7 correct responses.
- THKSbin is a recoded version of the ordinal variable THKSord, but in binary form: a value of "0" indicates an original scale score of 1 or 2, while a value of "1" indicates an scale score of 3 or 4. This variable will serve as our outcome variable in the current chapter.
- PreTHKS indicates a student's score prior to intervention, *i.e.*, the number correct of 7 items.
- CC is a binary variable indicating whether a social-resistance classroom curriculum was introduced, with 0 indicating "no" and 1 "yes."
- TV is an indicator variable for the use of media (television) intervention, with a "1" indicating the use of media intervention, and "0" the absence thereof.
- CC*TV is the product of the variables TV and CC, and represents the CC by TV interaction.

In this chapter, we consider models for binary outcomes, using quadrature as method of estimation.

4.2.1.1 Exploring the data

Crosstabulation

The focus is on the influence of the intervention on the tobacco health knowledge scores of the students, as represented by the binary outcome variable THKSbin. A cross-tabulation of the variables CC and TV for the two categories of the binary variable THKSbin is given in Table 4.1 below.

Table 4.1: Crosstabulation of CC, TV and THKSbin

THKSbin			CC		Total
			0	1	
0	TV	0	246	140	386
		1	215	152	367
Total			461	292	753
1	TV	0	175	240	415
		1	201	231	432
Total			376	471	847

The proportion of students with high scale scores (THKSbin = 1) in each of the four cells formed by the categories of CC and TV can be derived from Table 4.1. For example, 246 students in the category CC = 0, TV = 0 had a low score (THKSbin = 0), while 175 students had a high score (THKSbin = 1). The proportion of students in this cell with a high score is thus $\frac{175}{175+246} = 0.4157$. The observed proportions of high scores are summarized in Table 4.2 below.

Table 4.2: Observed proportion of high post–intervention scores

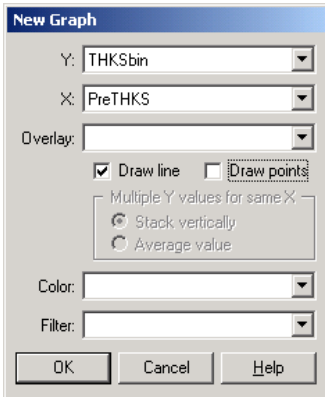
Study condition	Proportion	odds	logits
CC = 0, TV = 0	0.4157	0.711	–0.340
CC = 0, TV = 1	0.4832	0.935	–0.067
CC = 1, TV = 0	0.6316	1.714	0.539
CC = 1, TV = 1	0.6031	1.520	0.418

Proportions less than 0.5 indicate odds less than one and negative logits, while proportions above 0.5 yield odds greater than one and positive logit values. We note that, based on the observed proportion of high post-intervention scores, the use of only the social-resistance classroom curriculum (as represented by the variable CC) seemed the most successful, followed by the use of both curriculum and media intervention (CC = 1, TV = 1).

Exploratory graphs

The pre-intervention scores of the students may be useful as a covariate in the analysis. To get an idea of the relationship between the scale score PreTHKS and the post-intervention score THKSbin, an exploratory graph is created. Select the **Data-based Graphs**, **Exploratory** option from the **File** menu.

The **New Graph** dialog box is activated. Select the binary outcome variable THKSbin as the **Y** variable and the pre-intervention score PreTHKS as the **X** variable. Uncheck the **Draw points** check box, which is checked by default, to obtain the settings as shown.



Click **OK** to obtain Figure 4.1. The value associated with the tick marks on the X-axis represents the proportion of students with that PreTHKS score that had a value of 1 on the THKSbin variable, in other words the proportion of students with a post-intervention score of 3 or 4. We note that the relationship is reasonably linear, and that higher post-intervention scores are more often observed for students with high pre-intervention scores, which is what one intuitively would expect.

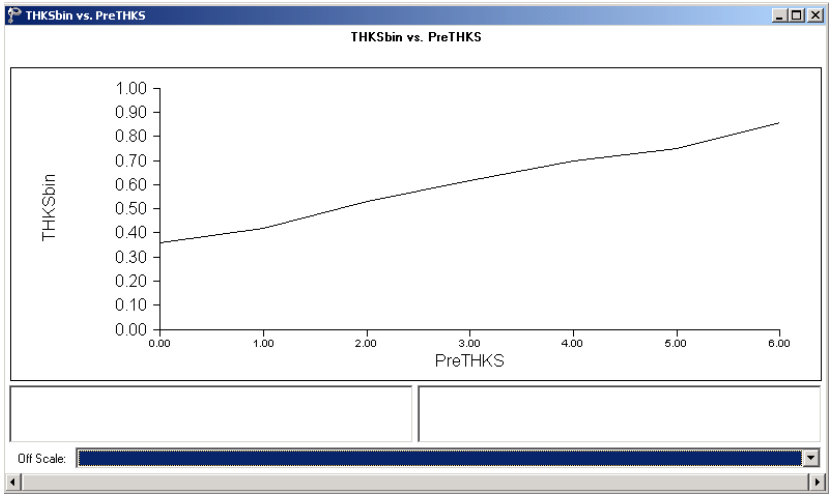
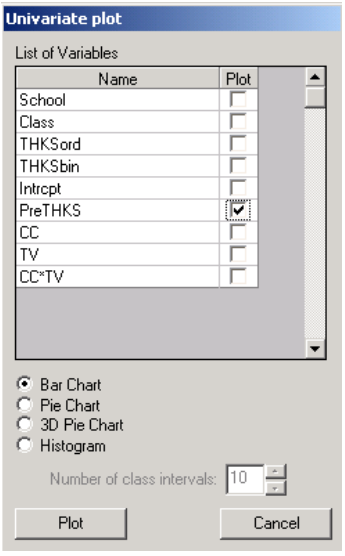


Figure 4.1: Exploratory graph of THKSbin vs. PreTHKS

Univariate graphs

We now take a closer look at the distribution of the pre-intervention scores by utilizing the **Data-based Graphs**, **Univariate** option on the **File** menu. By default, a bar chart will be produced. Select the variable PreTHKS in the **Plot** column, and click **Plot**.



By clicking anywhere in the bars, the **Bar Graph Parameters** dialog box is activated. Click the **Data** button and then **OK** to display the data used to construct the bar chart.

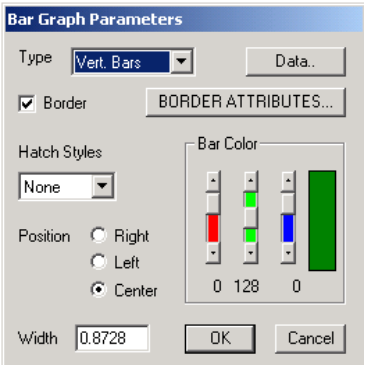


Figure 4.2 below shows both the graphing window with bar chart and the data in spreadsheet format. Note that only 55 of the 1600 observations showed a score of 5 or higher, and that no student obtained a post-intervention score of 7 out of 7.

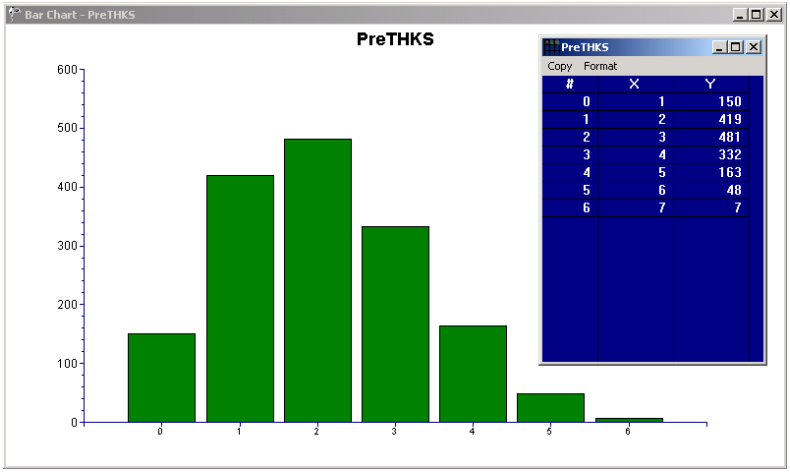


Figure 4.2: Bar chart of PreTHKS values

Finally, we also take a look at the mean pre-intervention scores of the students for each of the four subgroups. These are summarized in Table 4.3 below, and show that the mean pre-intervention scores do not differ much.

Table 4.3: Mean pre-intervention scores

Study condition	Mean
CC = 0, TV = 0	2.152
CC = 0, TV = 1	2.087
CC = 1, TV = 0	2.050
CC = 1, TV = 1	1.979

4.2.2 A 2-level random intercept logistic model with 2 predictors

4.2.2.1 The model

The outcome variable THKSbin used here is binary. It assumes a value of "0" when the original scale score was either 1 or 2, and a value of "1" for an original scale score of 3 or 4. The predicted value of the outcome can be viewed as the predicted probability that THKSbin is 1. As explained in Section 4.1.1, predicted values outside the interval [0,1] would not be meaningful and a model constraining predicted values to lie in this interval would be appropriate, in contrast with the model for a continuous outcome (see above) where effect sizes outside this interval would be interpretable. In addition, the assumption of normality at level 1 is not realistic, as the level-1 random effect can only assume one of two values: 0 or 1. This random effect can thus not have homogeneous variance.

In order to insure that the predicted values lie within the (0,1) interval, a transformation of the level-1 predicted probability can be used. For the binary case considered here, the following link function is used:

$$\text{Prob}(\text{THKSbin}_{ij} = 1 \mid \boldsymbol{\beta}, \mathbf{v}) = \frac{e^{\eta_{ij}}}{1 + e^{\eta_{ij}}}$$

where η_{ij} represents the log of the odds of success.

For the current model, we want to explore the relationship between the post-intervention scores and the type of intervention applied. This relationship can be expressed as

Level 1 model:

$$\eta_{ij} = b_{0i} + b_{1i} \times \mathrm{CC}_i + b_{2i} \times \mathrm{TV}_i + b_{3i} \times \mathrm{CC}_i * \mathrm{TV}_{ij} + b_{4i} \times \mathrm{PreTHKS}_{ij} + e_{ij}$$

Level 2 model:

$$\begin{aligned} b_{0i} &= \beta_0 + v_{0i} \\ b_{1i} &= \beta_1 \\ b_{2i} &= \beta_2 \\ b_{3i} &= \beta_3 \\ b_{4i} &= \beta_4 \end{aligned}$$

An equivalent expression for the model is

$$\eta_{ij} = \beta_0 + \beta_1 \times \mathrm{CC}_i + \beta_2 \times \mathrm{TV}_i + \beta_3 \times \mathrm{CC}_i * \mathrm{TV}_{ij} + \beta_4 \times \mathrm{PreTHKS}_{ij} + v_{0i} + e_{ij}.$$

The interpretation of the logistic regression model is made in terms of the logits, as the model is linear in terms of the logits. Thus the coefficients $\beta_1, \beta_2, \ldots, \beta_4$ can be interpreted as follows:

- β_0 is the THKS logit for CC = 0, TV = 0, that is the log odds of a positive outcome for an individual from the control group where no intervention was made and with a pre-intervention score of 0. One could also refer to β_0 as the PreTHKS adjusted logit for the CC = 0, TV = 0 subgroup.
- β_1 = the logit difference between (CC = 1, TV = 0) and (CC = 0, TV = 0) for the case where PreTHKS = 0:

$$\eta_{ij} = \beta_0 + (\beta_1 + \beta_3 \mathrm{TV}_i)\mathrm{CC}_i + \beta_2 \mathrm{TV}_i + \beta_4 \mathrm{PreTHKS}_{ij} + v_{0i},$$

in other words, the PreTHKS adjusted logit difference between these two subgroups.

- β_2 = the logit difference between (TV = 1, CC = 0) and (TV = 0, CC = 0) with PreTHKS = 0:

$$\eta_{ij} = \beta_0 + (\beta_2 + \beta_3 \mathrm{CC}_i)\mathrm{TV}_i + \beta_1 \mathrm{CC}_i + \beta_4 \mathrm{PreTHKS}_{ij} + v_{0i}\,.$$

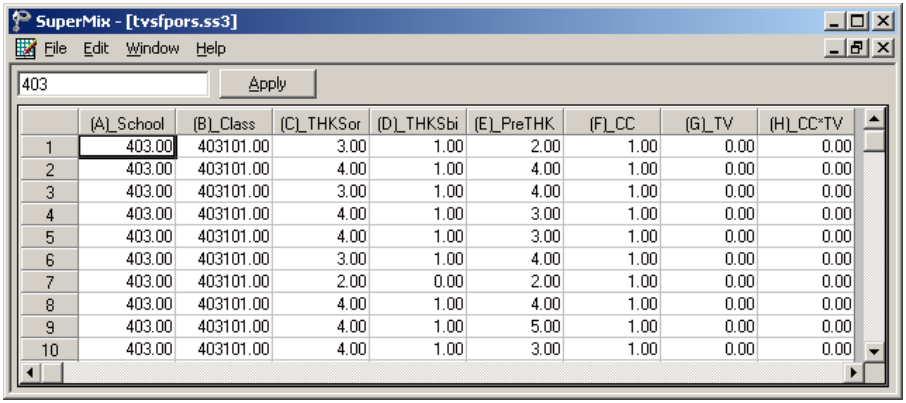
- β_3 is the difference in logit attributable to the interaction between the two intervention methods.

The interpretation of the coefficients is dependent on the coding of the variables used in the model.

4.2.2.2 Setting up the analysis

Using the data in **tvsfpors.ss3**, we consider the situation where students are nested within schools, and fit a two-level model with the binary variable THKSbin as outcome. We wish to examine the relationships between the outcome and the two intervention methods employed, simultaneously taking students' pre-intervention scores into account. To do so, we use the model described above with schools as the level-2 units.

Use the **File**, **Open Spreadsheet** option to activate the **Open** dialog box. Browse for the file **tvsfpors.ss3** in the **Examples\Binary** folder. Select the file and click the **Open** button to return to the main SuperMix window, where the contents of the SuperMix system file are displayed.

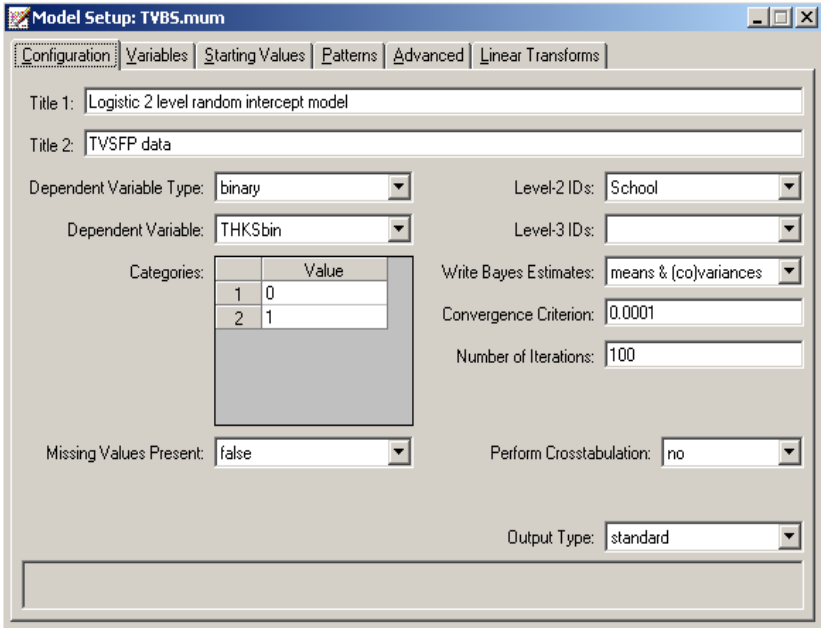


	(A)_School	(B)_Class	(C)_THKSor	(D)_THKSbi	(E)_PreTHK	(F)_CC	(G)_TV	(H)_CC*TV
1	403.00	403101.00	3.00	1.00	2.00	1.00	0.00	0.00
2	403.00	403101.00	4.00	1.00	4.00	1.00	0.00	0.00
3	403.00	403101.00	3.00	1.00	4.00	1.00	0.00	0.00
4	403.00	403101.00	4.00	1.00	3.00	1.00	0.00	0.00
5	403.00	403101.00	4.00	1.00	3.00	1.00	0.00	0.00
6	403.00	403101.00	3.00	1.00	4.00	1.00	0.00	0.00
7	403.00	403101.00	2.00	0.00	2.00	1.00	0.00	0.00
8	403.00	403101.00	4.00	1.00	4.00	1.00	0.00	0.00
9	403.00	403101.00	4.00	1.00	5.00	1.00	0.00	0.00
10	403.00	403101.00	4.00	1.00	3.00	1.00	0.00	0.00

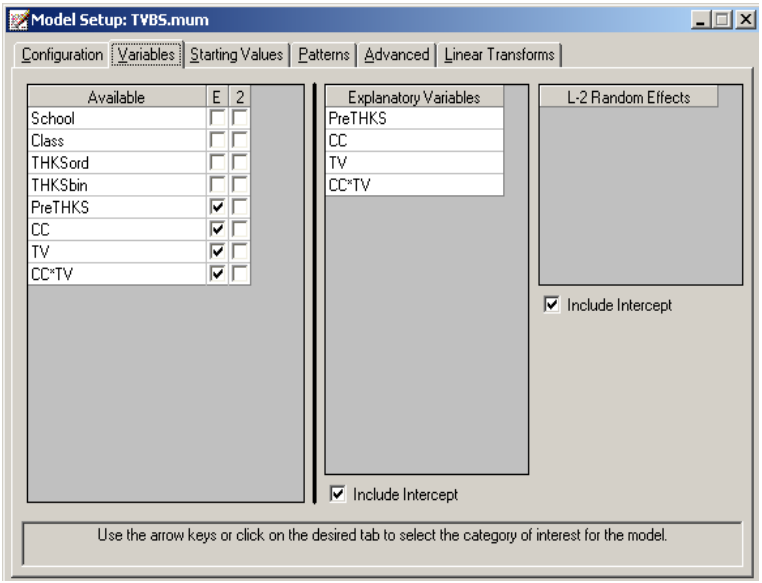
Next, we use the SuperMix interface to provide the model specifications. From the main menu bar, select the **File, New Model Setup** option.

The **Model Setup** dialog box that appears has six tabs: **Configuration**, **Variables**, **Starting Values**, **Patterns**, **Advanced**, and **Linear Transforms**. In this example, only three of the tabs are used.

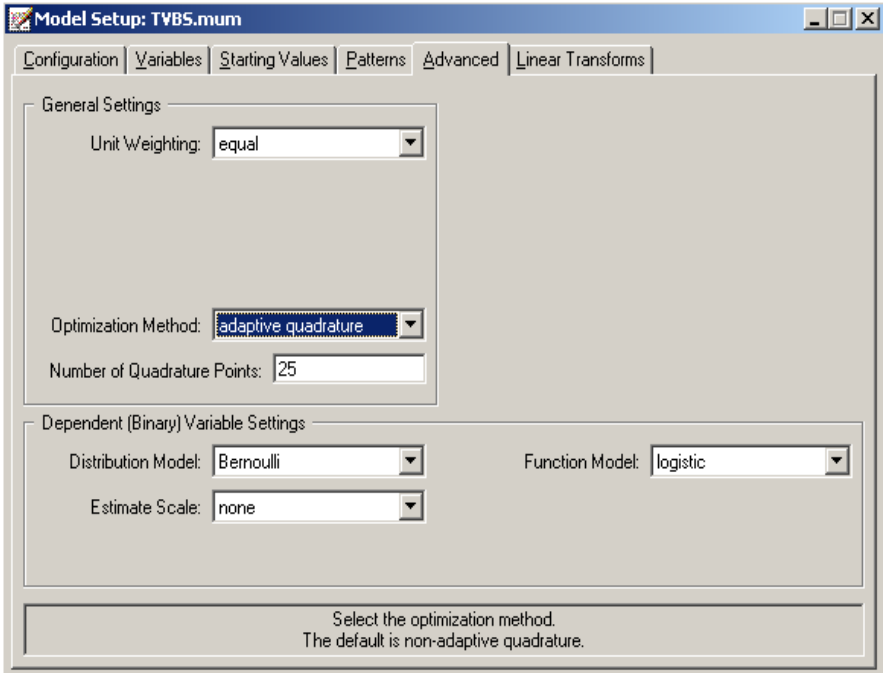
As a first step, the binary outcome variable THKSbin is selected from the **Dependent Variable** drop-down list box. The type of outcome is specified as binary using the drop-down list box in the **Dependent Variable Type** field. Once this selection is made, the **Categories** field is displayed. The school identification variable 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. A title for the analysis (optional) is entered in the **Title** fields. A convergence criterion of 0.0001 is requested. By default, the maximum number of iterations performed is set to 100. Empirical Bayes residuals, written to additional output files, are requested by setting the **Write Bayes Estimates** option to **means and (co)variances.** Default settings for all other options associated with this tab are used. Proceed to the **Variables** tab by clicking on this tab.



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 first column of boxes in the **Available** group field. The first variable selected is PreTHKS, followed by CC, TV, and the interaction term CC*TV. After selecting these explanatory variables, the random effect(s) at level 2 must be selected. In this case, we wish to allow only the intercept to vary randomly over the schools. By default, the intercept is assumed to vary randomly over higher levels of the hierarchy as indicated by the checked box for the **Include Intercept** option in the **L-2 Random Effects** group field. A common fixed intercept coefficient is also included, as shown in the **Explanatory Variables** group field.



We opt to increase the number of quadrature points to be used during estimation. To do so, click the **Advanced** tab. First select **adaptive quadrature** from the **Optimization Method** drop-down list box, then change the **Number of Quadrature Points** field to 25. The default distribution for a binary outcome variable is **Bernoulli** and the default link function is **probit**. Change **probit** to **logistic** by using the drop-down list box in the **Function Model** field.



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

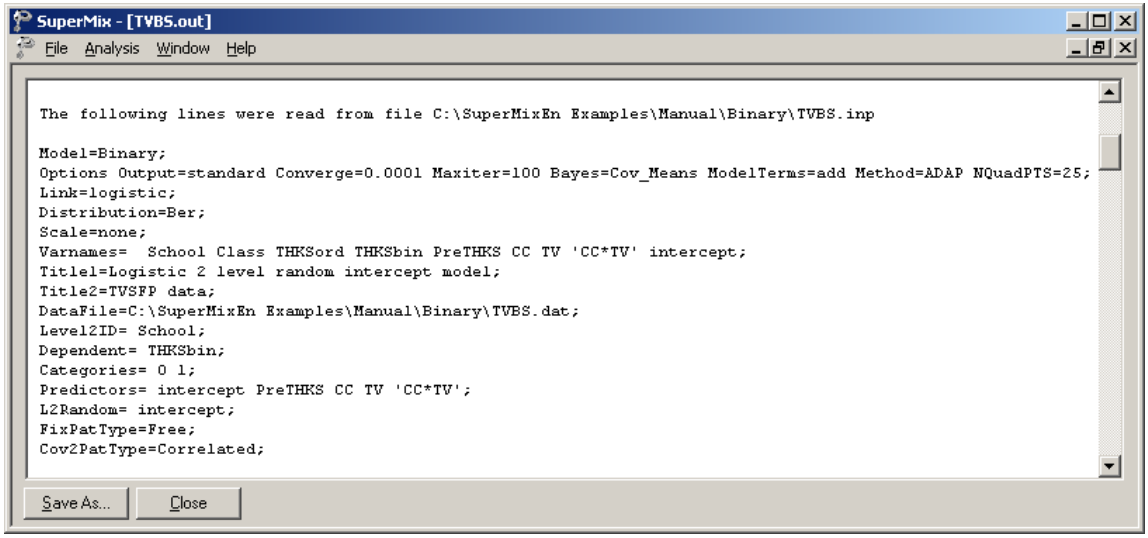
4.2.2.3 Discussion of results

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

Syntax

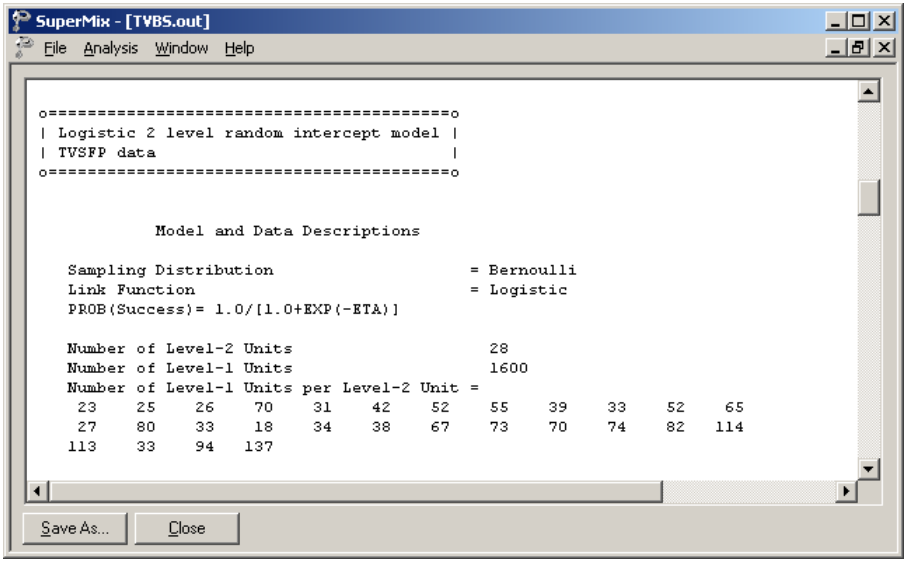
At the top of the file, the syntax saved to the **TVBS.mum** file is shown. The first part states the selection of iteration control options, requests for Bayes residuals, and the specifications necessary to define the model fitted as an binary model with a logistic link function. The second part of the syntax provides information on the structure of the data, the name and structure of the outcome variable, and the predictors included

in the model. Text to the left of the equal sign in each line denote keywords recognized by the program; text to the right are either keywords (for example, in the case of Cov2PatType = Correlated) or variable names as given in the **ss3** file (for example, Level2ID = School).



Model and data description

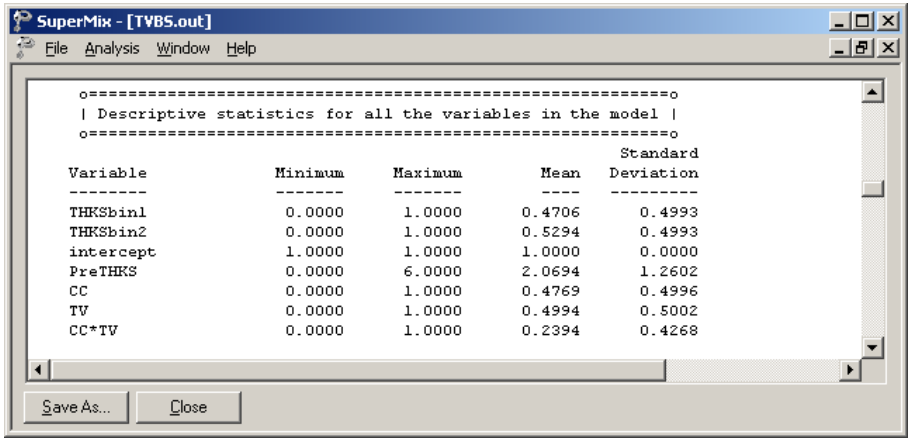
The next section of the output file contains a description of the hierarchical structure and model specifications.



The use of a logistic response function (logit link function) with the assumption of a Bernoulli distribution is indicated. This is followed by a summary of the number of students nested within each school. The number of students per school (level-2 unit) ranges between 23 and 137.

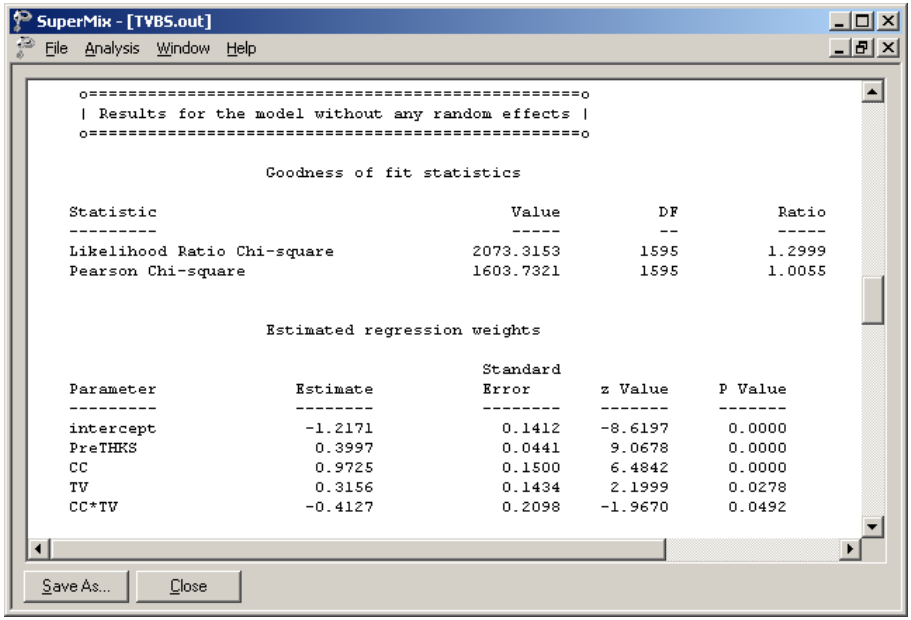
Descriptive statistics

The data summary is followed by descriptive statistics for all variables included in the model. We note that 47% of the students had a value of 0 on the binary knowledge score outcome variable THKSbin, and 53% a value of 1.

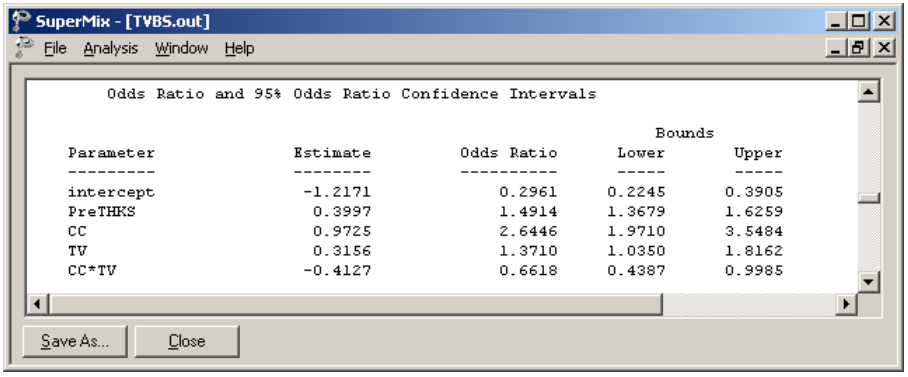


Results for the model without any random effects

Descriptive statistics are followed by parameter estimates obtained under the assumption that all random effects are zero. The parameter values for the predictors CC, TV, CC*TV and PreTHKS are given in the first column, followed by the standard errors and *z*- and *p*-values.

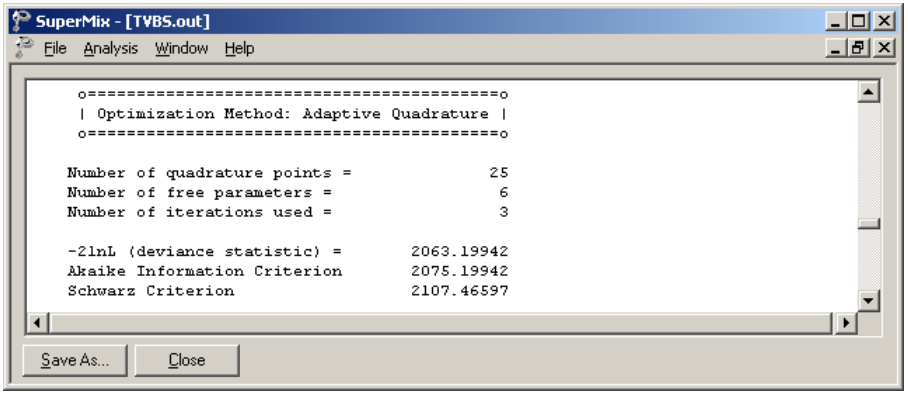


This is followed by the odds ratios and associated confidence intervals. The odds ratios are the exponents ($e^{\hat{\beta}}$) of the estimated regression coefficients.



Results for the model fitted with adaptive quadrature

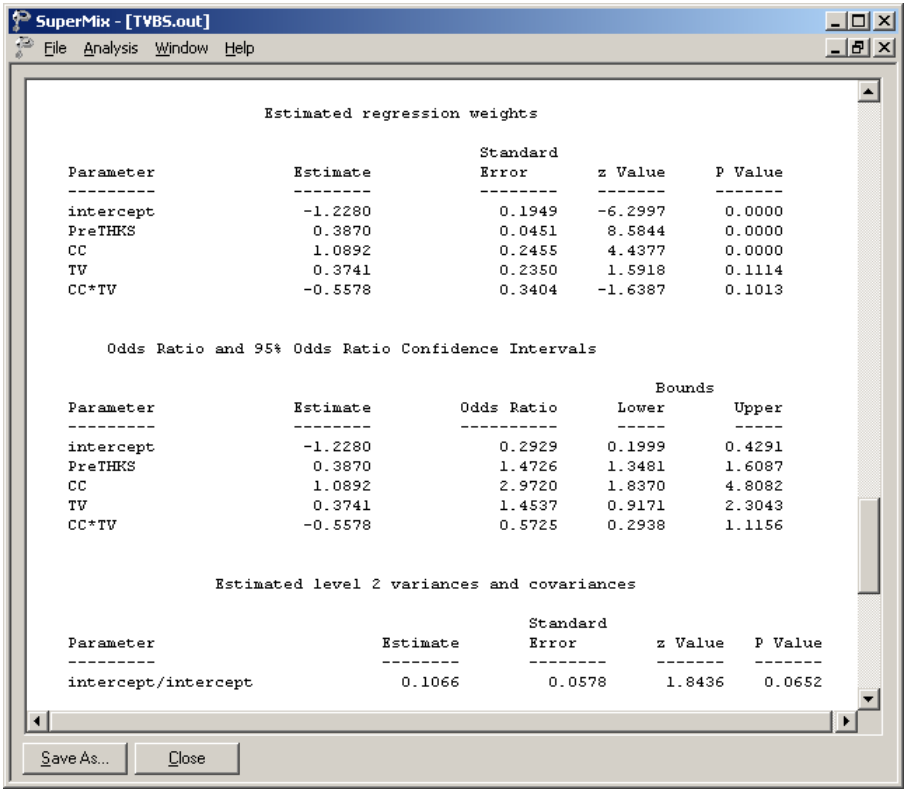
The output describing the estimated parameters after convergence is shown next. Three iterations were required to obtain convergence. The number of quadrature points used per dimension was 25. The likelihood function value at convergence as well as the deviance are also given, and may be used to compare a set of nested models.



The estimates are shown in the column with heading Estimate, and correspond to the coefficients $\beta_0, \beta_1, \ldots, \beta_4$ in the model specification. Significant effects of PreTHKS and CC are observed. The variation in the intercept over the schools is estimated as 0.1065, and from the associated p-value we conclude that there is significant variation, at a 10% level of significance, in the intercept between the schools included in this analysis.

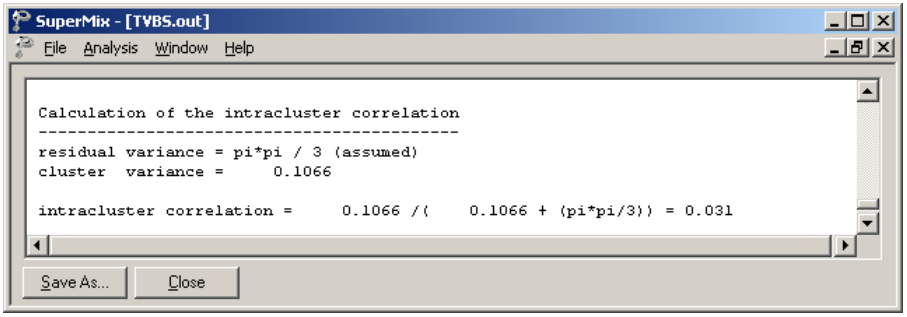
In the case of the fixed effects, a 2-tailed p-value is used, as the alternative hypothesis considered here is of the form $H_1 : \beta \neq 0$. As variances are constrained to be elements of the interval $[0, +\infty)$, the p-values used for these effects are 1-tailed.

If the model is true, it is assumed that the level-1 error variance is equal to $\pi^2 / 3 = 3.29895$ for the logistic link function (see, *e.g.*, Hedeker & Gibbons (2006), p. 157), where π represents the constant 3.141592654.



Thus the estimated ratio between level-2 variation and the total variation is calculated as

$$ICC = \frac{0.1065}{0.1065 + \pi^2 / 3} = 0.031$$



This indicates that almost all variation is attributable to students, rather than to the schools.

4.2.2.4 Interpreting the adaptive quadrature results

The expected log-odds of having a high post-intervention knowledge score (THKSbin score of 1) for a student with a zero value on all the predictors (that is, no social-resistance curriculum, no media intervention, and a pre-intervention knowledge score of 0) is represented by the estimated intercept of –1.2281. When a social-resistance curriculum was in place (CC = 1), or a mass-media intervention was performed (TV = 1), the log-odds of a typical student is expected to increase, as indicated by the positive estimated coefficients for CC and TV. Similarly, a higher score on the pre-intervention knowledge test is associated with higher log-odds of a higher post-intervention knowledge score. It can be concluded from the results that the implementation of a classroom curriculum was more likely to lead to a higher post-intervention knowledge score than was the case when mass-media intervention was used. In contrast, the log-odds of a high post-intervention knowledge score was expected to be lower for a typical student from a school where both social resistance classroom curriculum and mass-media intervention defined the study condition for that school, as the estimated coefficient for the interaction term CC*TV was negative.

Estimated outcomes for different groups: unit-specific results

To evaluate the expected effect of CC, TV, CC*TV, and PreTHKS on the predicted probability that the post-intervention score is equal to 1, we use the following expression for the predicted log odds of success

$$\hat{\eta}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 \times \mathrm{CC}_i + \hat{\beta}_2 \times \mathrm{TV}_i + \hat{\beta}_3 \times \mathrm{CC}_i \times \mathrm{TV}_i + \hat{\beta}_4 \times \mathrm{PreTHKS}_{ij}$$

for the four groups defined by the categories of CC and TV. Note the similarity of this equation with that given for η_{ij} earlier: random coefficients are not included, as their expected value is 0.

For a typical student with a PreTHKS score of 0 from any school where no media television intervention and no social-resistance classroom curriculum was implemented, CC = TV = 0, and thus

$$\hat{\eta}_{ij} = \hat{\beta}_0$$

In the case of a typical student with a PreTHKS score of 0 from any school where only media television intervention was implemented (TV = 1),

$$\hat{\eta}_{ij} = \hat{\beta}_0 + \hat{\beta}_2 \times \mathrm{TV}_i.$$

The equations for similar students from a school with only a social–resistance classroom curriculum implemented (CC = 1, TV = 0), and from a school with both interventions implemented (TV = 1, CC = 1) are

$$\hat{\eta}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 \times \mathrm{CC}_i + \hat{\beta}_4 \times \mathrm{PreTHKS}_{ij}$$

and

$$\hat{\eta}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 \times \mathrm{CC}_i + \hat{\beta}_2 \times \mathrm{TV}_i + \hat{\beta}_3 \times \mathrm{CC}_i \times \mathrm{TV}_i + \hat{\beta}_4 \times \mathrm{PreTHKS}_{ij}$$

respectively.

For a student with an average PreTHKS score (2.152, see exploratory analysis) from any school with similar values of CC and TV we find that

$$\hat{\eta}_{ij} = \hat{\beta}_0 + \hat{\beta}_4 * \text{PreTHKS}_{ij}$$
$$= \hat{\beta}_0 + \hat{\beta}_4 * 2.152.$$

Using the $\hat{\beta}_0$ and $\hat{\beta}_4$ estimates of –1.2280 and 0.3870 respectively as obtained for the current analysis, we can calculate the estimated probability of a THKSbin score of 1 for typical students with PreTHKS scores of 2.152 and 0 respectively as

$$\text{Prob}(\text{THKSbin}_{ij} = 1 \mid \text{CC} = \text{TV} = 0; \text{PreTHKS} = 2.152) = \frac{e^{-1.2280+0.3870(2.152)}}{1+e^{-1.2280+0.3870(2.152)}}$$
$$= \frac{e^{-0.39518}}{1+e^{0.39518}}$$
$$= 0.40247$$

and

$$\text{Prob}(\text{THKSbin}_{ij} = 1 \mid \text{CC} = \text{TV} = \text{PreTHKS} = 0) = \frac{e^{-1.2280}}{1+e^{-1.2280}}$$
$$= 0.22653.$$

A student with an average observed score of PreTHKS is almost twice as likely to have a THKSbin score of 1 as a student with the lowest observed score on the same variable. Note that we opted to use the mean pre-intervention score for this specific subgroup.

On the other end of the scale in terms of intervention, we have schools where both a social-resistance classroom curriculum and a mass-media intervention were implemented (CC = TV = 1). For two typical students from these schools, an observed PreTHKS score of 0 or the mean score of 1.979 will imply a predicted probability of a THKSbin score of 1 of 0.4201 for the first and 0.6091 for the second. Again, the higher the pre-intervention score, the higher the predicted probability of a high post-intervention score.

In Table 4.4, the estimated probabilities of high post-intervention scores on the tobacco and health questionnaire are given for typical students with high or low pre-

intervention scores for each of the subpopulations formed by mass-media intervention and implementation of social-resistance classroom curriculum.

Table 4.4: Estimated unit-specific probability of a high post-intervention knowledge score

Group	prescore	prob.	prescore	prob.
CC = 0, TV = 0	0	22.65%	2.152	40.25%
CC = 1, TV = 0	0	46.54%	2.05	65.81%
CC = 0, TV = 1	0	29.86%	2.87	48.85%
CC = 1, TV = 1	0	42.01%	1.979	60.91%

These estimated probabilities can also be presented graphically, as shown in the bar chart below.

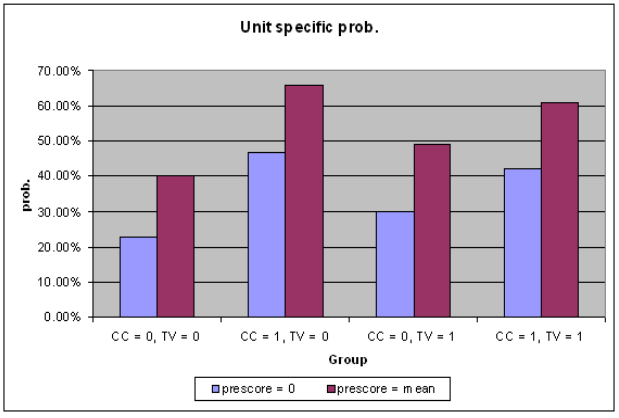


Figure 4.3: Bar chart of estimated unit-specific probabilities

Students with a high pre-intervention score were predicted to have a high post-intervention score too, regardless of the study conditions. Similarly, students with a low pre-intervention score were generally likely to have a low post-intervention score too. If only curriculum intervention (CC = 1) was used, scores for students were likely to be higher regardless of their pre-intervention scores. On both ends of the pre-intervention knowledge score scale, in groups where mass-media intervention was used (TV = 1), scores were predicted to be higher than where media

intervention was not used, except when both mass-media and curriculum intervention were used. For these groups, with CC = TV = 1, the estimated probabilities of a high post-intervention score were actually lower than for the group where only a classroom curriculum was used (42.01% vs. 46.54%, and 60.91% vs. 65.80%).

We conclude that for most students, the implementation of a social-resistance classroom curriculum is more likely to be effective in increasing their knowledge (predicted probabilities of a high score being 46.54% and 65.80% respectively) than mass-media intervention (predicted probabilities of a high score being 29.86% and 48.85% respectively). The control group, where neither method was implemented, had the lowest predicted knowledge scores (22.65% and 40.25% respectively). While the implementation of both procedures was associated with higher probabilities than either the control group or the group where only mass-media intervention was used, its predicted gain was disappointing when compared to the use of only social-resistance curriculum implementation. Generally speaking, the implementation of a curriculum only seems to be most effective in increasing the predicted knowledge of students on the tobacco and health questionnaire.

Estimated outcomes for different groups: population-average results

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

$$y_{ij} = \mathbf{x}_{ij}^{'}\boldsymbol{\beta} + \mathbf{z}_{ij}^{'}\mathbf{v}_i + e_{ij}, \quad j = 1, 2, ..., n_i$$

where $\mathbf{z}_{ij}^{'}$ denotes a design vector for the random effects contained in the vector $\mathbf{v}_i$, and $\mathbf{x}_{ij}^{'}$ the design vector for the predictors in the fixed part of the model with corresponding vector $\boldsymbol{\beta}$ of regression parameters. The covariance matrix of $\mathbf{v}_i$ is denoted by $\boldsymbol{\Phi}_{(v)}$ and the variance of e_{ij} by σ_e^2.

For a probit link function $\sigma_e^2 = 1$, and for a logistic link function it is assumed to be $\sigma_e^2 = \pi^2 / 3$. Under the assumption that $\mathbf{v}_i$ and e_{ij} are independently distributed, it follows that

$$\sigma^2_{y_{ij}} = \mathbf{z}'_{ij}\boldsymbol{\Phi}_{v_i}\mathbf{z}_{ij} + \sigma^2_e.$$

The design effect d_{ij} is defined in terms of σ^2_e and $\sigma^2_{y_{ij}}$:

$$d_{ij} = \frac{\sigma^2_{y_{ij}}}{\sigma^2_e}.$$

This design effect may be used to obtain the estimated population-average probabilities in a similar fashion as the unit-specific probabilities, but with replacing $\hat{\eta}_{ij}$ with $\hat{\eta}^*_{ij} = \hat{\eta}_{ij}/\sqrt{d_{ij}}$ (Hedeker & Gibbons, 2006).

We can compare these estimated population-average probabilities with the observed data for the four groups formed by the categories of TV and CC as shown in Table 4.5. To illustrate, we calculate the estimated population-average probabilities for a few of the subgroups.

From the output, we have $\text{var}(v_{i0}) = 0.1065$, where v_{i0} denotes the random intercept coefficient. In this case, $\mathbf{z}'_{ik} = \mathbf{1}$ and hence, with $\sigma^2_e = \pi^2/3$ for the logistic link,

$$\sigma^2_{y_{ij}} = 1 \times 0.1065 \times 1 + 3.2899 = 3.3964.$$

Therefore

$$d_{ij} = \frac{3.3964}{3.2899} = 1.0324.$$

To obtain the population-average probability estimates, we now replace the $\hat{\eta}_{ij}$ values calculated for the unit-specific case with $\hat{\eta}^*_{ij} = \hat{\eta}_{ij}/\sqrt{d_{ij}}$.

For the subgroup where TV = CC = 0 and the mean PreTHKS value is equal to 2.152, for example, we find that

$$\begin{aligned}\hat{\eta}_{ij} &= -1.2281 + 0.3871(2.152) \\ &= -0.39506\end{aligned}$$

so that

$$\begin{aligned}\hat{\eta}_{ij}^{*} &= -0.39506 / \sqrt{1.0324} \\ &= -0.38881\end{aligned}$$

and

$$\begin{aligned}P(\text{THKSbin}_{ij} = 1 \mid \text{CC} = \text{TV} = 0, \text{PreTHKS} = 2.152) &= \frac{e^{\eta_{ij}^{*}}}{1 + e^{\eta_{ij}^{*}}} \\ &= \frac{0.67786}{1.67786} = 40.40\%.\end{aligned}$$

Similarly, for the group where TV = CC = 0 and PreTHKS = 0, we find that

$$\hat{\eta}_{ij} = -1.2281$$

$$\begin{aligned}\hat{\eta}_{ij}^{*} &= -1.2281 / 1.01606 \\ &= -1.2087.\end{aligned}$$

Table 4.5: Estimated population-average probabilities

Group	prescore	prob.	prescore	prob.
CC = 0, TV = 0	0	22.99%	2.15	40.40%
CC = 1, TV = 0	0	46.59%	2.05	65.57%
CC = 0, TV = 1	0	30.14%	2.87	48.87%
CC = 1, TV = 1	0	42.13%	1.98	60.74%

A comparison of these probabilities with the observed ratios given in Table 4.6 for the control group at the end of the study indicates that the population-average results are slightly closer to the observed ratios than is the case for the unit-specific results. Recall that $\sqrt{d_{ij}} = 1.0161$. The extent of differences between unit-specific and population-average results is highly dependent on the "scaling" induced by dividing the $\hat{\eta}_{ij}$ s by $\sqrt{d_{ij}}$.

Table 4.6: Observed and predicted proportions of high post–intervention scores

Group	Proportion observed	Unit-specific predicted prob.	Population-average predicted prob.
CC = 0, TV = 0	41.57%	40.25%	40.40%
CC = 1, TV = 0	63.16%	65.80%	65.57%
CC = 0, TV = 1	48.32%	48.84%	48.86%
CC = 1, TV = 1	60.31%	60.91%	60.74%

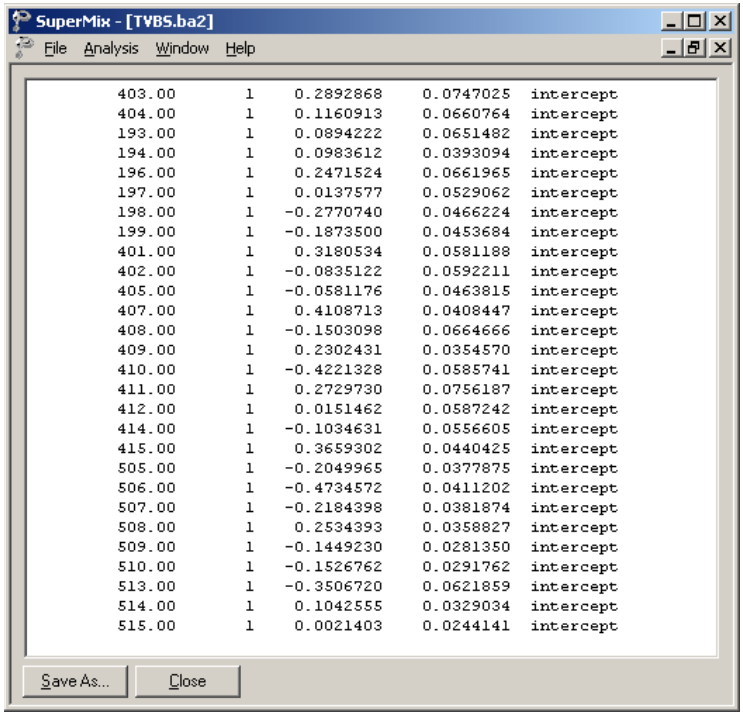
4.2.2.5 Interpreting the contents of the level-2 residual file

In addition to the standard output file, the **Write Bayes Estimates** field on the **Configuration** tab of the **Model Setup** dialog was used to request Bayes estimates for the individual random terms. These estimates are written to the file **TVBS.ba2**. The first few lines of this file are shown below.

Four pieces of information per school are given:

- all 1s for the level-2 model,
- the school's ID,
- the value of random intercept,
- the empirical Bayes estimate,
- the associated posterior variance for the school estimate, and

- the name of the associated random coefficient.



The mean of the empirical Bayes estimates is – 0.0002. The estimates ranged from –0.473614 for school 506 to 0.4110043 for school 407. In both cases a mass-media intervention procedure was applied, and thus TV = 1, but CC = CC*TV = 0. For students with a PreTHKS score of 3 from each of these schools, this implies

$$\text{Prob}(\text{THKSbin}_{ij} = 1 \mid \text{CC} = 0, \text{PreTHKS} = 3, \text{ID} = 506) = \frac{e^{-0.473614+0.3741+0.3870(3)}}{1+e^{-0.473614+0.3741+0.3870(3)}}$$

$$= \frac{e^{1.061486}}{1+e^{1.061486}} = 0.7430$$

and

$$\text{Prob}(\text{THKSbin}_{ij} = 1 \mid \text{CC} = 0, \text{PreTHKS} = 3, \text{ID} = 407) = \frac{e^{0.4110043+0.3741+0.3870(3)}}{1+e^{0.4110043+0.3741+0.3870(3)}}$$
$$= \frac{e^{1.9461043}}{1+e^{1.9461043}} = 0.8750$$

respectively. The fact that the intercept for school 407 lies higher than the average is reflected in the higher probability (87.5%) that a student with average pre-intervention knowledge score will obtain a high post-intervention score. School 506, on the other hand, has an intercept far below the average, and a student from this school has, in effect, a 74.30% chance of obtaining a high post-intervention score.

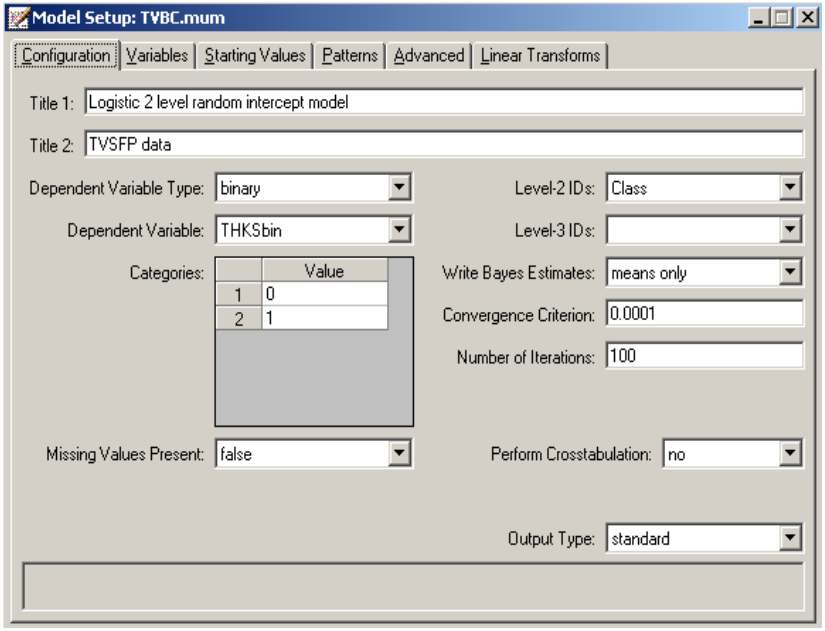
4.2.3 A 2-level random intercept logistic regression model

Using the same data (**tvsfpors.ss3)** and model setup file **TVBS.mum** from the previous example, we now consider the situation where students are nested within classrooms and fit a two-level model of the form described earlier, again with the binary variable THKSbin as outcome.

4.2.3.1 Setting up the analysis

Use the **File**, **Open Spreadsheet** option to re-open the previously used spreadsheet **tvsfpors.ss3** from the **Examples\Binary** folder. Next, use the **File**, **Open Existing Model Setup** option to browse and open the syntax file **TVBS.mum.**

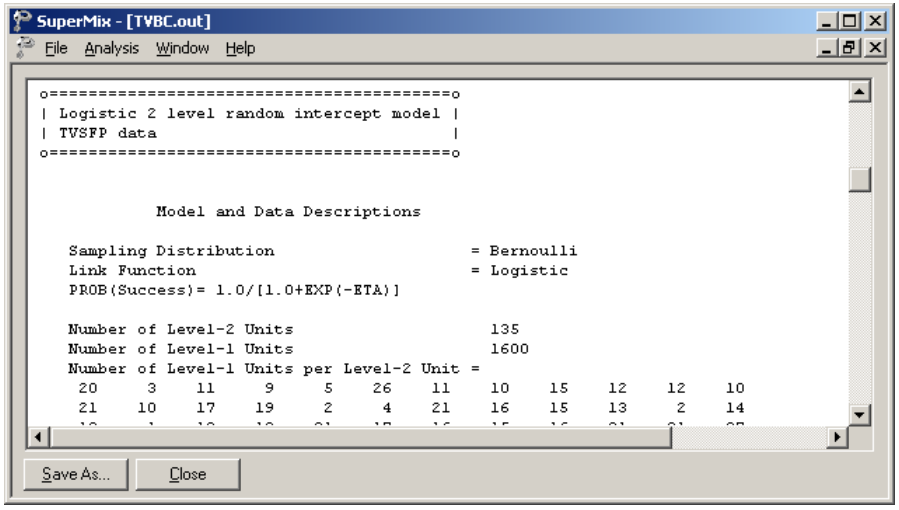
The biggest change to be made to the syntax file is in terms of the ID variable. Change the **Level-2 IDs** field on the **Configuration** tab of the **Model Setup** dialog box from School to Class, as shown below. Also, turn of the writing out of Bayes estimates by setting the **Write Bayes Estimates** field to **no**.



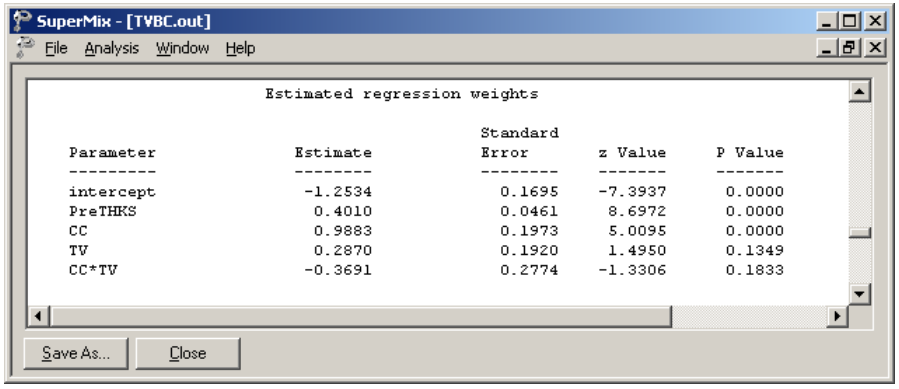
Save the revised syntax file under a new name such as **TVBC.mum** and run the analysis.

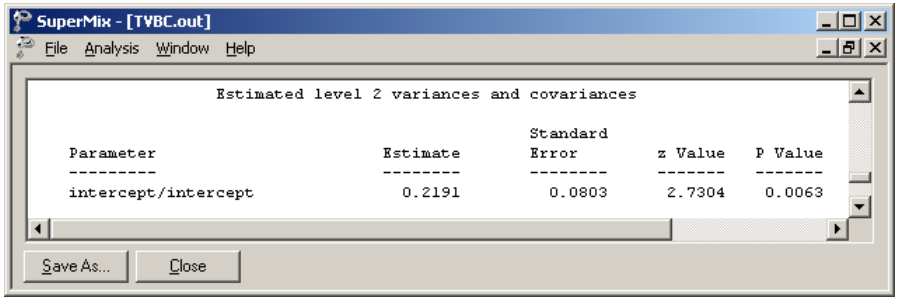
4.2.3.2 Discussion of results

Partial output for this run is provided below. The summary of units now reflects the number of students nested within each classroom. The number of students per class (level-2 unit) ranges between 2 and 28. In this analysis, there were 135 level-2 units, compared to 28 in the previous analysis.



Estimated coefficients with adaptive quadrature and the estimated level-2 variances are given below.





The estimates for the classroom analysis are very similar to those of the school analysis. All estimated fixed coefficients are slightly lower than was the case in the previous analysis. There seems to be more variation between classrooms than between schools, as indicated by the estimated variation in the random intercept of 0.2193, compared to the similar estimate of 0.1065 in the school analysis.

The estimates can again be used to obtain predicted probabilities by first calculating the $\hat{\eta}_{ij}^{*}$ s, using the formulae

$$\hat{\eta}_{ij} = -1.2535 + 0.9883 \times \mathrm{CC}_i + 0.2870 \times \mathrm{TV}_i - 0.369 \times (\mathrm{CC} \times \mathrm{TV})_i + 0.401 \times \mathrm{PreTHKS}_{ij}$$

and $\hat{\eta}_{ij}^{*} = \hat{\eta}_{ij} / \sqrt{d_{ij}}$ where

$$d_{ij} = \frac{\sigma^2_{y_{ij}}}{\sigma^2_e} = \frac{0.2193 + \pi^2/3}{\pi^2/3}$$

$$= \frac{0.2193 + 3.289865}{3.289865} = 1.0666.$$

A comparison of unit-specific and population-average predicted probabilities for the current model are given in Table 4.7. For comparison purposes, similar results for the previous model can be found in Table 4.7.

Table 4.7: Observed and predicted proportions of high post–intervention scores

Group	Proportion observed	Unit-specific predicted prob.	Population-average predicted prob.
CC = 0, TV = 0	41.57%	40.36%	40.66%
CC = 1, TV = 0	63.16%	63.57%	63.16%
CC = 0, TV = 1	48.32%	46.76%	46.87%
CC = 1, TV = 1	60.31%	60.98%	60.64%

4.2.4 A 3-level random intercept logistic regression model

Having fitted 2-level models where students were nested within either classrooms or schools thus far, we now consider a 3-level model with both classroom and school defining levels of the hierarchy.

4.2.4.1 The model

The level-1 and level-2 models are the same as for the previous two models, as shown below.

Level 1 model $(k = 1,\ldots,n_{ij})$:

$$\text{THKSbin}_{ijk} = b_{0ij} + b_{1ij}\text{PRETHKS}_{ijk} + e_{ijk}$$

Level-2 model $(j = 1,\ldots,n_i)$:

$$b_{0ij} = b_{00i} + b_{01i}\text{CC}_{ij} + b_{02i}\text{TV}_{ij} + b_{03i}(\text{CC}_{ij} \times \text{TV}_{ij}) + v_{0ij}$$
$$b_{1ij} = b_{10i}$$

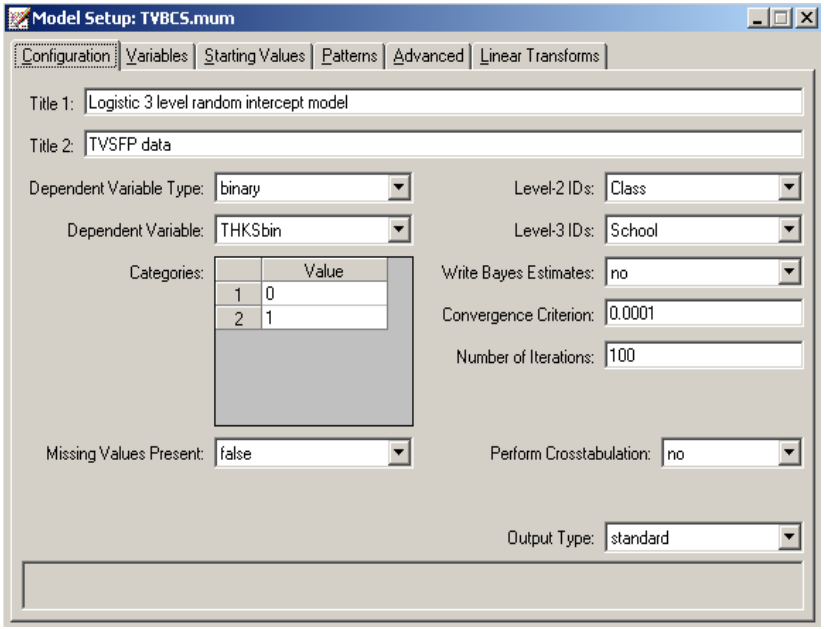
With classrooms nested within schools, however, a third level of the hierarchy is defined. At this level, the level-2 coefficients become outcomes again, and can potentially vary over the schools (level-3 units). In the current model, we allow only the intercept to vary randomly over the schools.

Level-3 model $(i = 1, \ldots, N)$

$$b_{00i} = \beta_0 + v_{0i}$$
$$b_{01i} = \beta_1$$
$$b_{02i} = \beta_2$$
$$b_{03i} = \beta_3$$
$$b_{10i} = \beta_4$$

4.2.4.2 Setting up the analysis

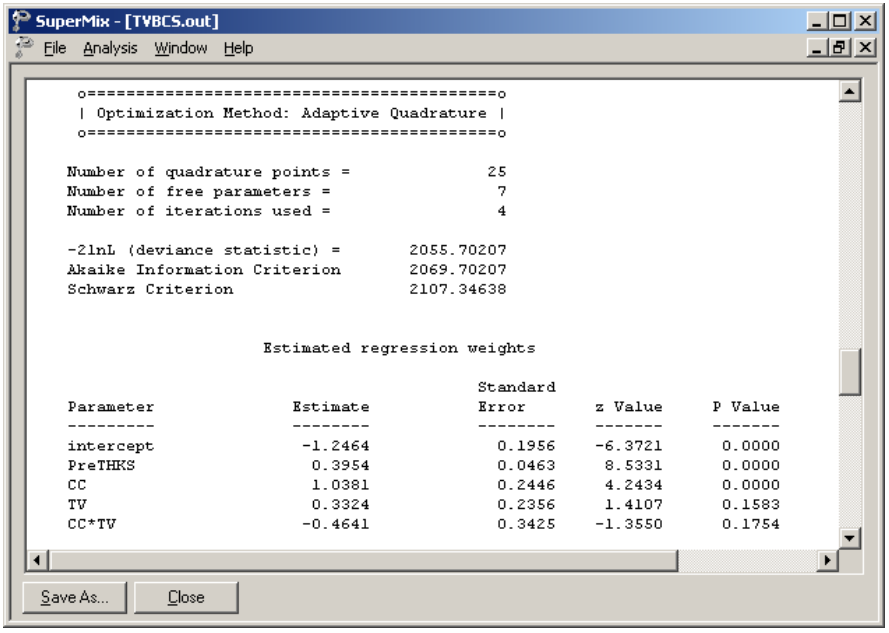
We modify our model setup saved to the syntax file **TVBS.mum** by first using the **Open Existing Model Setup** option on the **File** menu to retrieve the syntax file. Then click on **File, Save** as to save the model setup in a new file, such as **TVBCS.mum**. Next, select CLASS as the **Level-2 ID** and SCHOOL as the **Level-3 IDs** as shown below. We now have both level-2 and level-3 IDs selected.

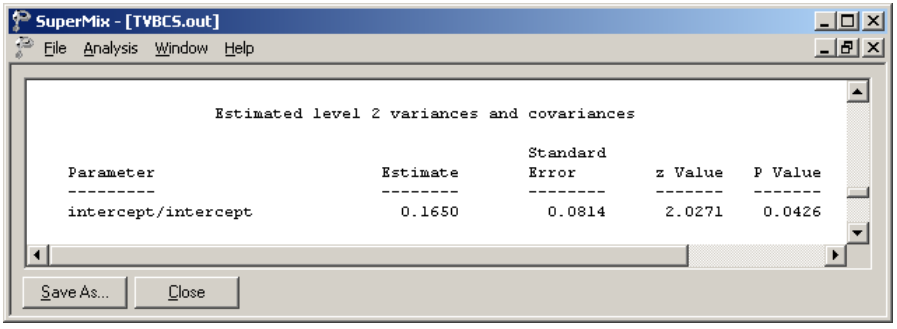


Keep all the other settings unchanged. Save the changes to the file **TVBCS.mum** and select the **Run** option on the **Analysis** menu to run the analysis.

4.2.4.3 Discussion of results

The portions of the output file **TVBCS.out** containing the estimates of the fixed and random coefficients in the current model are shown below.


SuperMix - [TVBCS.out]
File Analysis Window Help
o==o
| Optimization Method: Adaptive Quadrature |
o==o
Number of quadrature points = 25
Number of free parameters = 7
Number of iterations used = 4
-2lnL (deviance statistic) = 2055.70207
Akaike Information Criterion 2069.70207
Schwarz Criterion 2107.34638
Estimated regression weights
Parameter Estimate Standard Error z Value P Value
intercept -1.2464 0.1956 -6.3721 0.0000
PreTHKS 0.3954 0.0463 8.5331 0.0000
CC 1.0381 0.2446 4.2434 0.0000
TV 0.3324 0.2356 1.4107 0.1583
CC*TV -0.4641 0.3425 -1.3550 0.1754
Save As...
Close


SuperMix - [TVBCS.out]
File Analysis Window Help
Estimated level 2 variances and covariances
Parameter Estimate Standard Error z Value P Value
intercept/intercept 0.1650 0.0814 2.0271 0.0426
Save As...
Close

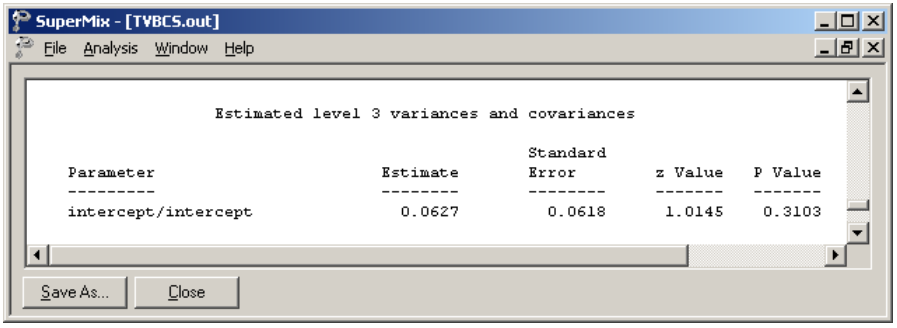

SuperMix - [TVBCS.out]
File Analysis Window Help
Estimated level 3 variances and covariances
Parameter Estimate Standard Error z Value P Value
intercept/intercept 0.0627 0.0618 1.0145 0.3103
Save As...
Close

Table 4.8: Comparison of results for three models with binary variable THKSbin as outcome

Coefficient		2-level: CLASS as ID	2-level: SCHOOL as ID	3-level
Fixed effects:				
Intercept	estimate	–1.2535	–1.228	–1.2465
	standard error	0.1695	0.1949	0.1957
PRETHKS	estimate	0.401	0.3871	0.3954
	standard error	0.0461	0.0451	0.0463
CC	estimate	0.9883	1.0893	1.0383
	standard error	0.1973	0.2454	0.2448
TV	estimate	0.287*	0.3741*	0.3325*
	standard error	0.192	0.235	0.2358
CCxTV	estimate	–0.369*	–0.5578*	–0.4644*
	standard error	0.2774	0.3403	0.3427
Random effects:				
Var(between classrooms)	estimate	0.2193		0.1649
	standard error	0.0802		0.0813
Var(between schools)	estimate		0.1065	0.063*
	standard error		0.0578	0.0616

*: Not significant at 5% level of significance.

Results for this model are compared to those obtained using the two 2-level models in Table 4.8. Generally, there is close agreement between the models in terms of both the sign and size of the effects. Note that the only intervention method that consistently has an estimated coefficient significantly different from zero is CC. While use of the media intervention (TV) can positively influence the post-intervention score, it seems clear that using both methods simultaneously does not have any real benefits.

4.2.4.4 Interpreting the adaptive quadrature results

3-level ICCs

Intraclass correlation coefficients can be obtained for the three-level dichotomous outcome model. As mentioned earlier, it is assumed that the level-1 error variance is equal to $\pi^2/3$ for the logistic link function if the model is true (see, *e.g.*, Hedeker & Gibbons (2006), p. 157). Using this approximation, the formulae for the standard ICCs can be adjusted.

From the output for the random effects, we have

$$\text{Level-1: estimated}\left(\text{error var}\right) = \pi^2/3 = 3.2899$$
$$\text{Level-2: estimated}\left(\text{class var}\right) = 0.1649$$
$$\text{Level-3: estimated}\left(\text{school var}\right) = 0.0630.$$

Based on this information, we can calculate the ICC as shown below.

Similarity of students within the same school:

$$ICC = \frac{\sigma^2_{v(3)}}{\sigma^2_{v(3)} + \sigma^2_{v(2)} + \sigma^2_e} = \frac{0.063}{0.063 + 0.1649 + 3.28986}$$

$$= 0.0179.$$

Similarity of students within the same classrooms (and schools):

$$ICC = \frac{\sigma^2_{v(2)}}{\sigma^2_{v(3)} + \sigma^2_{v(2)} + \sigma^2_e} = \frac{0.1649}{0.063 + 0.1649 + 3.28986}$$

$$= 0.04688.$$

Similarity of classes within the same school:

$$ICC = \frac{\sigma^2_{v(2)}}{\sigma^2_{v(3)} + \sigma^2_{v(2)}} = \frac{0.1649}{0.063 + 0.1649}$$

$$= 0.7236.$$

Estimated unit-specific and population-average probabilities

Under the assumption that $\mathbf{v}_i$, $\mathbf{v}_{ij}$ and ε_{ijk} are independently distributed, it follows that for the three-level model the design effect is defined as

$$d_{ijk} = \frac{(\sigma^2_{v(3)} + \sigma^2_{v(2)} + \sigma^2_e)}{\sigma^2_e} = 1.0692.$$

The estimated unit-specific probabilities are calculated using

$$\hat{\eta}_{ijk} = -1.2465 + 1.0383 \times \mathrm{CC}_i + 0.3325 \times \mathrm{TV}_i - 0.4.644 \times \mathrm{CC}_i \times \mathrm{TV}_i + 0.3954 \times \mathrm{PreTHKS}_{ijk}$$

and

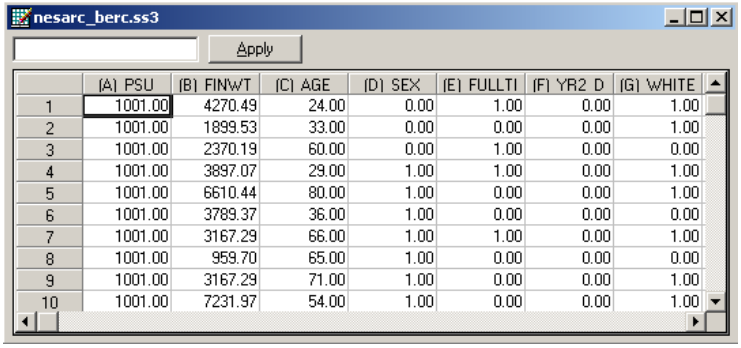
$$\mathrm{Prob}(\mathrm{THKSbin} = 1 \mid \boldsymbol{\beta}) = \frac{1}{1 + e^{-\eta_{ijk}}}$$

The estimated population-average probabilities (Hedeker & Gibbons, 2006) are obtained in a similar fashion as the unit-specific probabilities after replacing $\hat{\eta}_{ijk}$ with $\hat{\eta}^*_{ijk} = \hat{\eta}_{ijk} / \sqrt{d_{ijk}}$ in the second of the equations shown above.

4.3 Models based on the subset of NESARC data

4.3.1 The data

The data set is from the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC). This data file has been used in some of the examples in Section 3.1. Detailed information about the survey is available at the NIAAA website at http://niaaa.census.gov/index.html. We focus on information regarding occurrences of major depression, family history of major depression and dysthymia. This information was used, in combination with the demographic information provided in Section 1 of the study description, to produce the **nesarc_berc.ss3** data set used in this section. The image below shows the first ten records of this data set. There are 2339 dysthymia respondents in the survey; after listwise deletion, the sample size is 1981.



	(A) PSU	(B) FINWT	(C) AGE	(D) SEX	(E) FULLTI	(F) YR2 D	(G) WHITE
1	1001.00	4270.49	24.00	0.00	1.00	0.00	1.00
2	1001.00	1899.53	33.00	0.00	0.00	0.00	1.00
3	1001.00	2370.19	60.00	0.00	1.00	0.00	0.00
4	1001.00	3897.07	29.00	1.00	1.00	0.00	1.00
5	1001.00	6610.44	80.00	1.00	0.00	0.00	1.00
6	1001.00	3789.37	36.00	1.00	0.00	0.00	0.00
7	1001.00	3167.29	66.00	1.00	1.00	0.00	1.00
8	1001.00	959.70	65.00	1.00	0.00	0.00	0.00
9	1001.00	3167.29	71.00	1.00	0.00	0.00	1.00
10	1001.00	7231.97	54.00	1.00	0.00	0.00	1.00

The variables of interest are:

- PSU denotes the Census 2000/2001 Supplementary Survey (C2SS) primary sampling unit.
- FINWT represents the NESARC weights sample results used to form national level estimates. The final weight is the product of the NESARC base weight and other individual weighting factors.
- AGE represents the age of the respondent.

- SEX is the gender of the respondent (1 for male, 0 for female).
- FULLTIME is recoded from question S1Q7A1. It is the response to the statement "present situation includes working full time (35+ hours a week)" with 1 indicating yes and 0 indicating no.
- YR2_DEP is the observed response to the statement that the respondent had a period of at least 2 years with low mood, and being sad or depressed most of day (1 = yes, 0 = no.) It is recoded from S4CQ1 in the source data.
- WHITEOTH represents the origin of white and other ethnicities, excluding Black and Hispanic. It is recoded from items S1Q1C, S1Q1D2, S1Q1D3 and S1Q1D5 in the NESARC source code (1 for white and other, 0 for black and Hispanic).
- BLACK represents African American respondents in the sample. It is recoded from S1Q1C and S1Q1D3 (1 for African American, 0 for others).
- HISPANIC is an indicator for Hispanic respondents in the sample data. It is recoded from S1Q1C, S1Q1D3 and S1Q1D5 (1 for Hispanic, 0 for others).
- YOUNG is recoded from AGE. Respondents younger than 35 have the value 1; otherwise, YOUNG = 0.
- MIDDLE is recoded from AGE. Respondents with $35 \leq \text{AGE} < 50$ have the value 1. Otherwise, MIDDLE = 0.
- OLD is recoded from AGE. Respondents with $\text{AGE} \geq 50$ have the value 1. Otherwise, OLD = 0.

We recoded the ethnicity variables because of the unbalanced numbers of respondents from different ethnicities in the original NESARC data. While weights are supplied with the data and should be used to adjust for the disproportionality of the sample, the use of indicator variables offers the opportunity to obtain estimated coefficients for individual groups while using one of the other ethnic groups as a reference group. The recoding of ethnicity is discussed in detail in Section 3.1.

In this section, we discuss the fitting of three Bernoulli models to these data.

4.3.2 A 2-level random intercept probit model

4.3.2.1 The model

In the previous models (see Section 4.2) the logistic link function was used. We now fit a model by using the probit link function.

The outcome variable of interest is YR2_DEP has the values 0 or 1. For this binary outcome variable

$$\text{Prob}(\text{YR2\_DEP}_{ij} = 1 \mid \boldsymbol{\beta}_i) = \Phi^{-1}\left(\eta_{ij}\right)$$

where η_{ij} represents the log of the odds of success, and can be expressed as

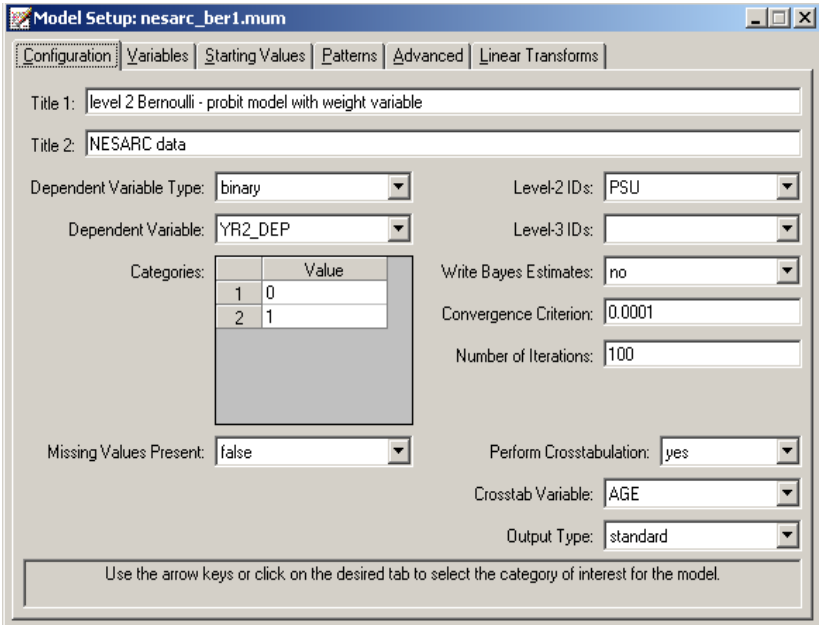
$$\eta_{ij} = b_0 + b_1 \times \text{AGE}_{ij} + b_2 \times \text{SEX}_{ij} + b_3 \times \text{FULLTIME}_{ij} + b_{i0} + v_{i0} + e_{ij}$$

for the intended model. This transformation, commonly referred to as the probit link function, constrains $\text{Prob}(y_{ij} = 1 \mid \boldsymbol{\beta})$ to lie in the interval (0,1).

4.3.2.2 Setting up the analysis

Open the SuperMix spreadsheet **nesarc_berc.ss3**. From the main menu bar, select the **File, New Model Setup** option.

The **Configuration** screen is the first tab on the **Model Setup** dialog box. It is used to define the outcome variable and level-2 and level-3 IDs. Some other settings such as missing values, convergence criterion, number of iterations, etc. can also be specified here. For all the available settings, please refer to Section 2.4.

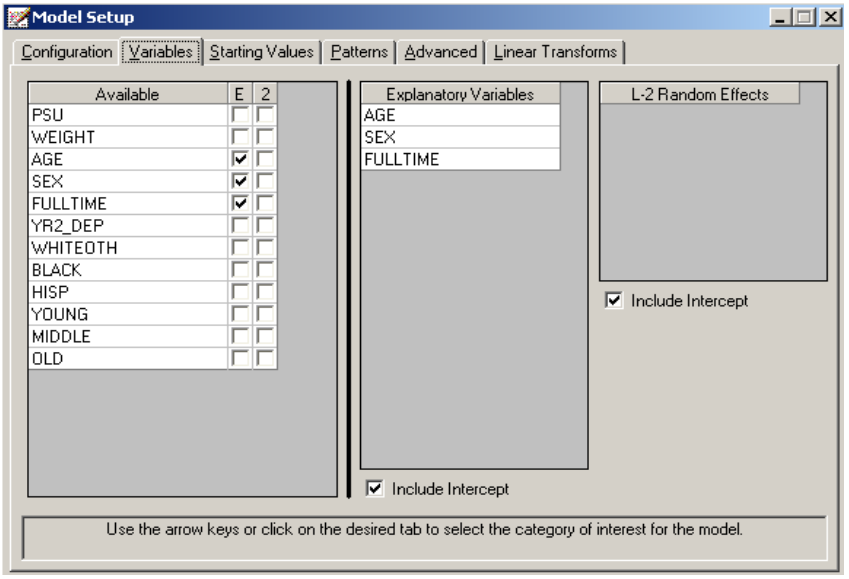


To obtain the model shown above, proceed as follows.

- Select the **binary** option from the **Dependent Variable Type** drop-down list.
- Select the outcome variable YR2_DEP from the **Dependent Variable type** drop-down list box.
- Select PSU from the **Level-2 ID** drop-down list box.
- Enter a title for the analysis in the **Title** text boxes if needed (optional).
- Request a crosstabulation of the outcome variable against AGE by selecting Yes from the **Perform Crosstabulation** drop-down list box, and select AGE as **Crosstab Variable**.
- Keep all the other settings on the **Configuration** screen at their default values. Proceed to the **Variables** screen by clicking on this tab.

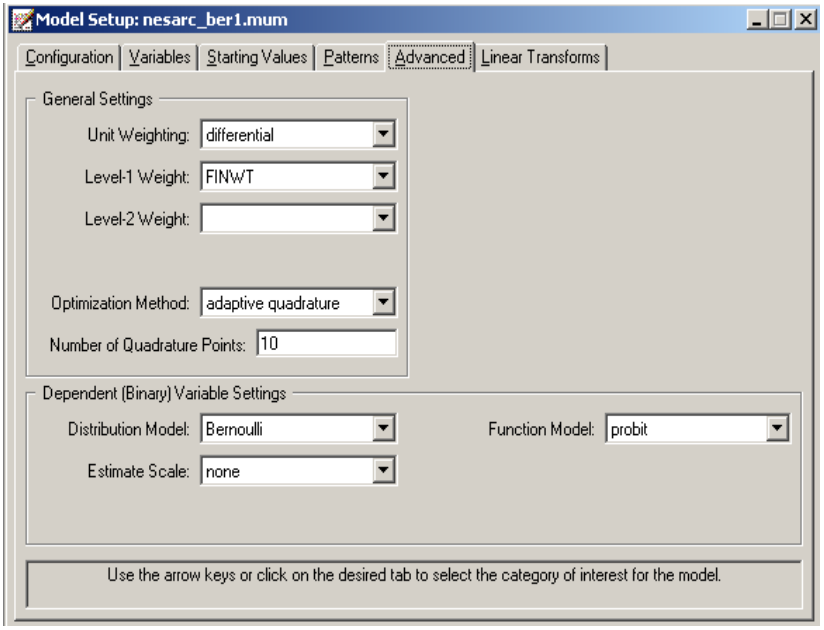
The **Variables** screen is used to specify the fixed and random effects to be included in the model. Select the explanatory (fixed) variables using the **E** check boxes next to the variables AGE, SEX and FULLTIME in the **Available** grid at the left of the

screen. After selecting all the explanatory variables, the screen shown below is obtained. 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.



The **Advanced** tab enables the user to define the weight variable. Weights are often used in complex sampling to adjust the existing sample for known biases. In SuperMix, the weight is normalized by default. To include a weight variable, proceed as follows:

- Select differential from the **Unit Weight** drop-down list to activate the **Assigned Weight**.
- Select FINWT from the drop-down list of the **Level-1 Weight field**.



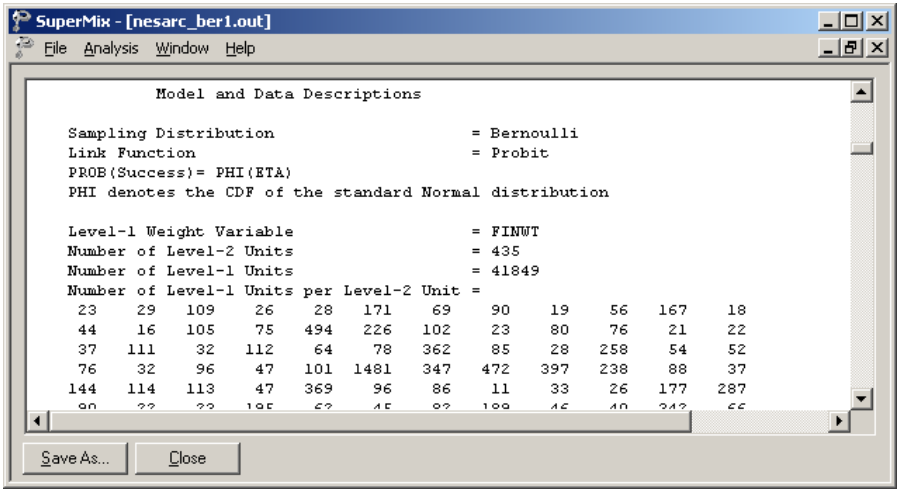
Save the model specifications to the file **nesarc_ber1.mum** and run the analysis.

4.3.2.3 Discussion of results

Portions of the output file **nesarc_ber1.out** are shown below.

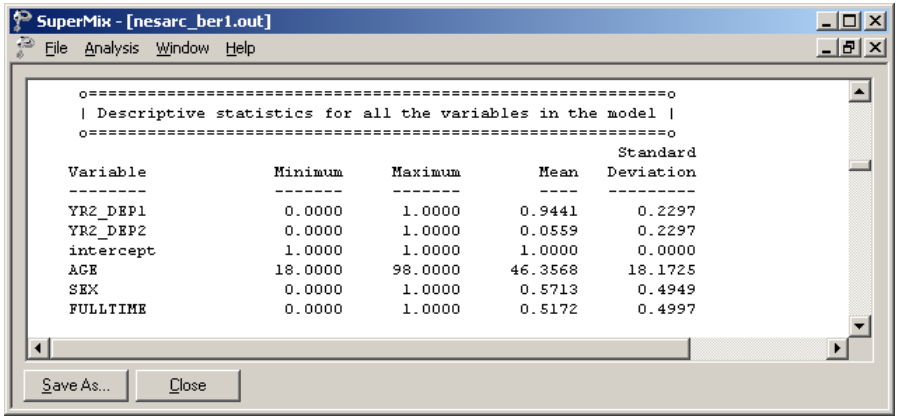
Model and data description

As shown in the model and data description section, the Bernoulli sampling distribution and probit link function are specified. The weight variable FINWT is used to include sampling weight. There are 41,849 observations from 435 PSUs included in the data we are analyzing.



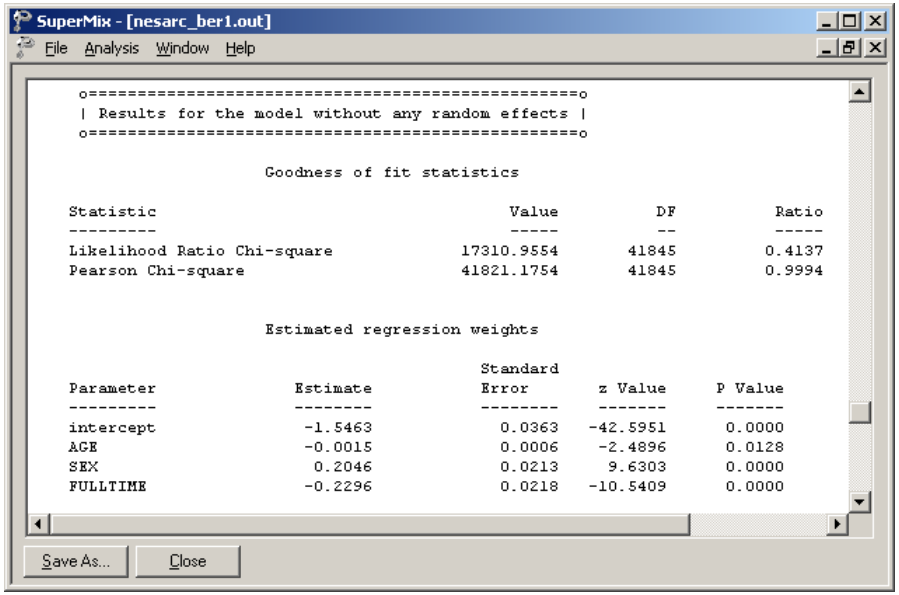
Descriptive statistics

The data summary is followed by descriptive statistics for all the variables included in the model. As shown below, about 94.41% of the respondents did not have a 2+ year period of low moods or being sad or depressed most of day.



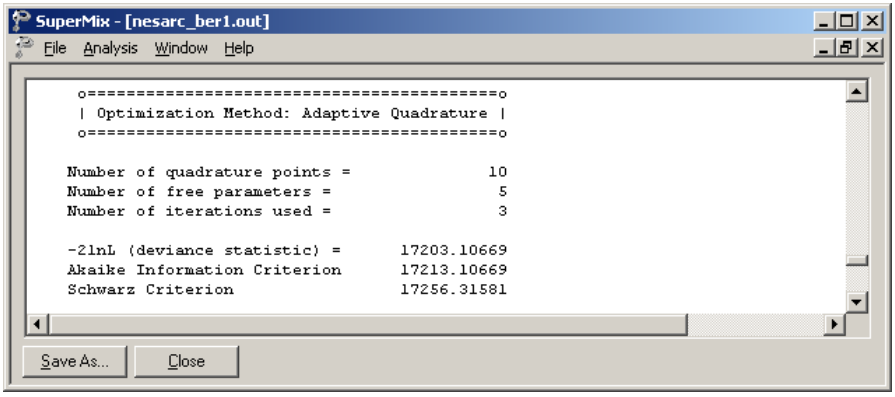
Results for the model without any random effects

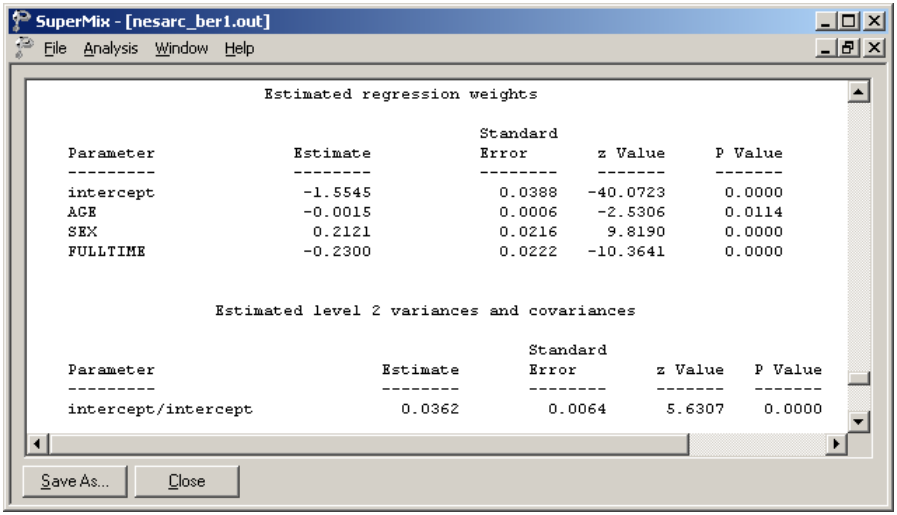
Descriptive statistics are followed by the results for the model without any random effects. These results are used as the starting values for the model with random effects.



Results for the model with random effects

The total number of iterations, the goodness of fit statistics and the estimated regression weights are shown below.





The estimated intercept coefficient is – 1.5544. The estimated coefficient associated with AGE is – 0.0015, which implies that for every year increase in age of a typical

respondent, the estimated probit $\hat{\eta}_{ij}$ is expected to decrease by 0.0015. The coefficient seems small, but keep in mind that age has a wide range, and consequently this estimate may have a big effect on the overall probability. The estimated coefficient associated with gender is 0.2121, which indicates that the male respondents (SEX = 1) have a larger $\hat{\eta}_{ij}$. The estimate for the indicator of FULLTIME shows that respondents with full-time jobs were expected to have a lower $\hat{\eta}_{ij}$ value than respondents with a similar profile in terms of age and gender but without full-time employment.

4.3.2.4 Interpreting the adaptive quadrature results

The probit link function is now used to transform these estimates into probabilities. First, we substitute the regression weights and obtain an expression for $\hat{\eta}_{ij}$:

$$\begin{aligned}\hat{\eta}_{ij} &= \hat{b}_{0i} + \hat{b}_{1i} \times (\text{AGE})_{ij} + \hat{b}_{2i} \times (\text{SEX})_{ij} + \hat{b}_{3i} \times (\text{FULLTIME})_{ij} \\ &= -1.5546 - 0.0015 \times (\text{AGE})_{ij} + 0.2121 \times (\text{SEX})_{ij} - 0.23 \times (\text{FULLTIME})_{ij}.\end{aligned}$$

For a typical 30-year-old male with a full-time job, SEX = 1, FULLTIME = 1 and AGE = 30, and thus

$$\begin{aligned}\hat{\eta}_{ij} &= -1.5546 - 0.0015 \times 30 + 0.2121 - 0.23 \\ &= -1.6025.\end{aligned}.$$

Transform the $\hat{\eta}_{ij}$ into the corresponding probability by using the probit link function:

$$\text{Prob}\left(\hat{\text{YR2\_DEP}}_{ij} = 1\right) = \Phi^{-1}(-1.60937) = 0.0545.$$

In terms of percentages, 5.45% of males with this profile would be expected to suffer from long-term depression episodes. Similarly, the probability of having a depression episode of 2+ years' duration for different gender and age combinations can be calculated. These probabilities, expressed as percentages, are reported in Table 4.9 below.

Table 4.9: % probabilities of having a depression episode

Age	20	30	40	50	60	70
not fulltime, female	5.66%	5.49%	5.32%	5.16%	5.00%	4.85%
not fulltime, male	8.50%	8.27%	8.04%	7.82%	7.60%	7.39%
fulltime, female	3.48%	3.37%	3.26%	3.15%	3.04%	2.94%
fulltime, male	5.45%	5.29%	5.13%	4.97%	4.82%	4.67%

In general, males without full-time employment were more likely to have depression episodes than their female counterparts. Surprisingly, this is also true of males with full-time employment.

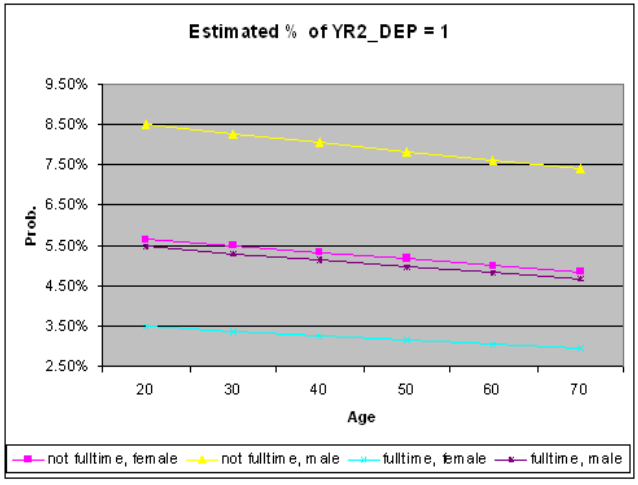


Figure 4.4: Expected probabilities for subgroups

These probabilities can also be depicted in Figure 4.4. The line associated with males without full-time jobs is considerably higher than for any other groups, again illustrating that this group has the highest probability of having 2+ years' period with low mood regardless of their age. For all the correspondents, as they grow older, the probability of having lengthy depression episodes decreased.

4.3.3 A 2-level random intercept model with additional predictors

4.3.3.1 The model

In the previous section, we modeled the outcome variable YR2_DEP in terms of its relationship with the predictors AGE, SEX and FULLTIME. The model discussed in this section takes the ethnicity of patients into consideration by including two dummy variables, BLACK and HISP. Since the group of WHITEOTH is not included, it is automatically regarded as the reference category.

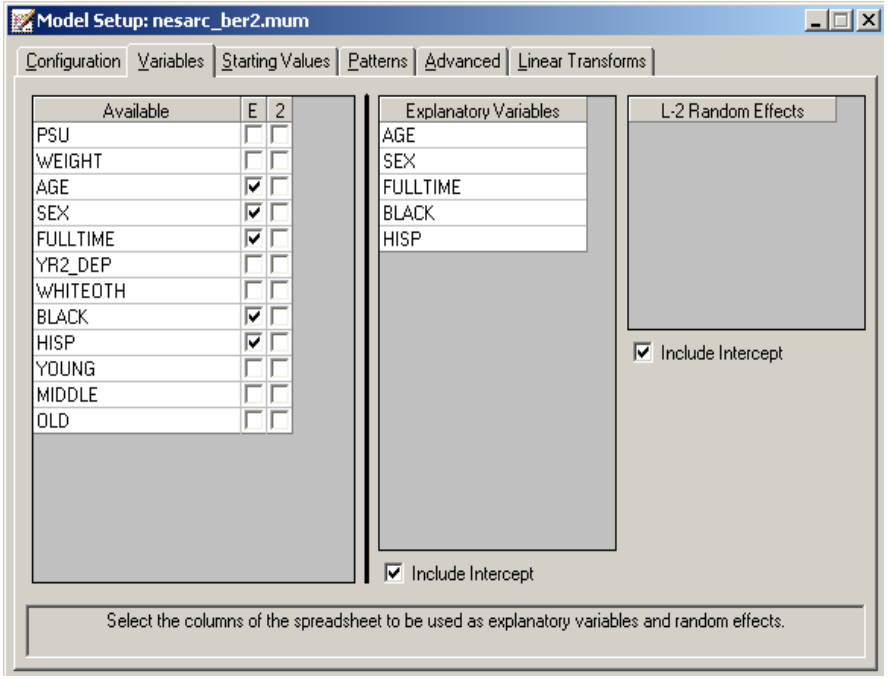
For the current model, the log of the odds of success (η_{ij}) can be expressed as

$$\eta_{ij} = b_0 + b_1 \times \text{AGE}_{ij} + b_2 \times \text{SEX}_{ij} + b_3 \times \text{FULLTIME}_{ij} + b_4 \times \text{BLACK}_{ij} + b_5 \times \text{HISP}_{ij} + v_{0i} + e_{ij}.$$

4.3.3.2 Setting up the analysis

We can modify the model setup file **nesarc_ber1.mum** by opening it and then saving it under a different name, such as **nesarc_ber2.mum**.

Click on the **Variables** tab of the **Model Setup** window. Add the predictors BLACK and HISP to the model by checking the boxes next to these variables in the **E** column, as shown below.



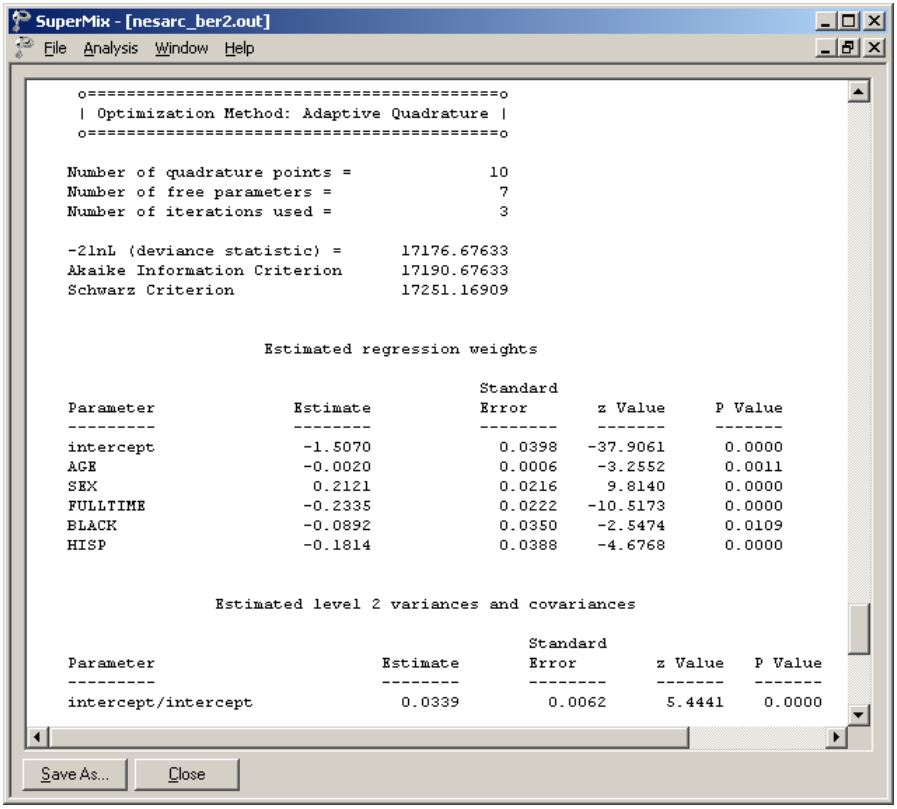
Save the modified model specification file, and select the **Run** option from the **Analysis** menu to perform the analysis.

4.3.3.3 Discussion of results

Portions of the output file **nesarc_berc2.out** are shown below.

Results for the model fitted with adaptive quadrature

The goodness of fit statistics are shown below. Since the previous model can be considered as a submodel of the current model, the deviances of these two models can be used to perform a χ^2 test to evaluate possible improvement in model fit.



The output describing the estimated fixed effects after convergence is shown next. As shown above the estimated logit for the intercept is –1.5069, the estimated logit associated with AGE is –0.002, etc. It is interesting to note that the only positive estimate is for gender. Males are thus more likely to show long-term depression, while it will be less likely in those who are older or fully employed. The ethnicity

indicators' coefficients also indicate that white respondents are most likely to have depression, with the Hispanic population the least likely.

4.3.3.4 Interpreting the adaptive quadrature results

Estimated outcomes for different groups: unit-specific results

To evaluate the simultaneous impact of these estimates on the expected probabilities for respondents from the subgroups formed by the categories of age, gender, and ethnicity, we may use the estimated regression weights and the link function to calculate probabilities of having depression in the same way as for the previous model.

For the current model, $\hat{\eta}_{ij}$ can be expressed as:

$$\hat{\eta}_{ij} = -1.5069 - 0.0020 \times \text{AGE}_{ij} + 0.2121 \times \text{SEX}_{ij} - 0.2335 \times \text{FULLTIME}_{ij} \\ -0.0891 \times \text{BLACK}_{ij} - 0.1814 \times \text{HISP}_{ij}.$$

Table 4.10 contains a subset of these estimated probabilities. Only typical respondents 30 or 50 years old are considered here, and probabilities are expressed as percentages.

Younger white males without full-time employment have the highest risk of having long-term depression, while female Hispanic respondents with full-time employment were least at risk.

Table 4.10: % probabilities of having depression episodes for selected age groups

Age	30			50		
Ethnicity	White	Black	Hispanic	White	Black	Hispanic
not fulltime, female	5.86%	4.89%	4.02%	5.40%	4.49%	3.69%
not fulltime, male	8.77%	7.44%	6.22%	8.15%	6.89%	5.75%
fulltime, female	3.59%	2.94%	2.38%	3.29%	2.68%	2.16%
fulltime, male	5.61%	4.67%	3.84%	5.17%	4.30%	3.52%

The results in Table 4.10 can also be depicted as a bar chart. Figure 4.5 shows that white respondents are more likely to get depressed for a long period than African American or Hispanic respondents.

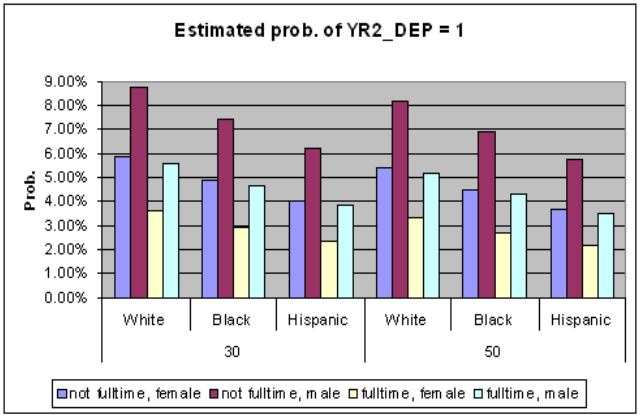


Figure 4.5: Estimated probabilities for subgroups

Model comparison

Since the two models in this section are nested models, the χ^2 difference test can be used. The deviances, AIC, and SBC statistics for these models are summarized in Table 4.11. These statistics suggest that the second model fits the data better.

Table 4.11: Model comparison

Statistic	Model 1	Model 2	difference	Difference in d.f.
$-2\ln L$ (deviance statistic)	17203.107	17176.677	26.430	2
Akaike Information Criterion	17213.107	17190.677	22.430	2
Schwarz Criterion	17256.316	17251.169	5.147	2

5 Models for count outcomes

5.1 Introduction

A count variable counts the number of discrete occurrences of a characteristic of interest that takes place during a time interval. Examples are the occurrence of cancer cases in a hospital during a given period of time, the number of cars that pass through a toll station per day, and the phone calls at a call center. The most common distribution for a count variable is the Poisson distribution. Besides the Poisson distribution, negative binomial distributions may also be used to describe the properties of count variables. In this chapter, models for count data, utilizing both the Poisson and negative binomial distributions, are discussed. For further information regarding these distributions, please refer to Section 4.1.1.

5.1.1 Poisson distribution

The Poisson distribution is a discrete probability distribution. It is appropriate for expressing the probability of a number of events occurring in a fixed time period with a known average rate, under the assumption that the occurrences are independent of one another.

The probability of k occurrences can be expressed as

$$f(k;\lambda) = \frac{e^{-\lambda}\lambda^{k}}{k!} \quad for \quad k = 0, 1, 2, ...$$

where k is a non-negative integer and λ is a positive real number, which equals the expected number of occurrences during the given interval. The cumulative probability function is

$$\Pr(k;\lambda) = \sum_{i=0}^{k} \frac{e^{-\lambda}\lambda^{i}}{i!} \quad for \quad k = 0, 1, 2, ...,$$

with the single parameter λ. A Poisson distribution has an important property: the mean number of occurrences λ is equal to the variance: $E(f) = \text{var}(f) = \lambda$. Figure 5.1 shows Poisson probabilities $f(k)$ and cumulative probabilities $g(k)$ for $\lambda = 0.5$, 2 and 5.

As shown below, the smaller λ is, the more skewed to the right the probability distribution is. When λ is large, the Poisson distribution is close to the normal distribution.

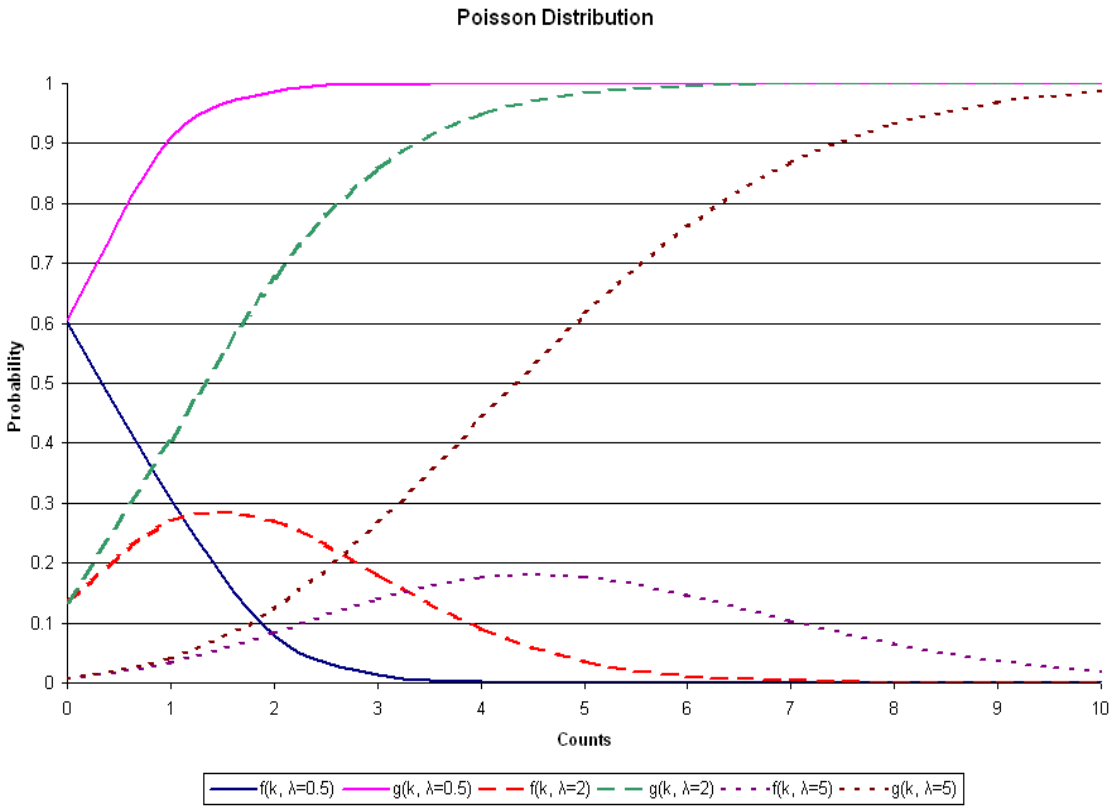


Figure 5.1: Poisson probabilities for various values of λ

The log link function is generally used for the Poisson distribution. Assume the response measurements for a count variable $y_1, ..., y_n$ are independent and

$$y_i \sim Poi(\lambda_i), \qquad where \qquad \lambda_i = e^{\beta_1 x_{i1} + \cdots + \beta_p x_{ip}}$$

The natural logarithm of the above equation is used to define the link function:

$$\log(\lambda_i) = \beta_1 x_{i1} + \cdots + \beta_p x_{ip}$$

As shown in Figure 5.2, using the log link function maps the mean of the count variable λ with an open interval $(0,+\infty)$ to the set of real numbers $(-\infty,+\infty)$.

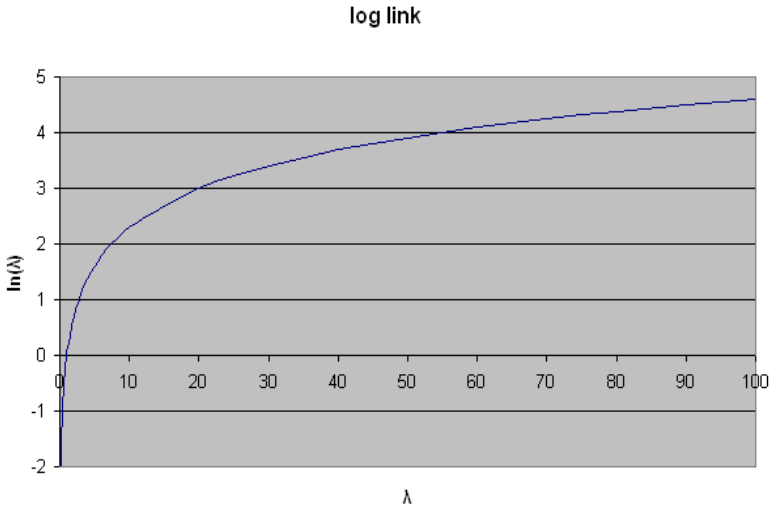


Figure 5.2: Log link function

5.1.2 Negative binomial distribution

The negative binomial distribution is a probability distribution used to describe a certain number of failures and successes in a series of independent and identically distributed Bernoulli trials. Specifically, for $k+r$ Bernoulli trials with success probability p, the negative binomial gives the probability of k failures and r successes, with success on the last trial. In other words, the negative binomial

distribution is the probability distribution of the number of failures before the r^{th} success in a Bernoulli process, with probability p of success on each trial.

The negative binomial distribution can be expressed as

$$f(y_i) = \frac{\Gamma(y_i + 1/\alpha)}{\Gamma(y_i + 1)\Gamma(1/\alpha)} \times \frac{(\alpha\mu_i)^{y_i}}{(1+\alpha\mu_i)^{y_i+1/\alpha}}$$

with $\Sigma(y_i) = \mu_i + \alpha\mu_i^2$, where $\Gamma(\ )$ is the gamma function or generalized factorial from advanced calculus, and where α denotes an additional parameter and it can no longer be assumed that the variance is a known function of the mean. In the example to follow, α is assumed to have a fixed value.

5.1.3 Adaptive versus non-adaptive quadrature

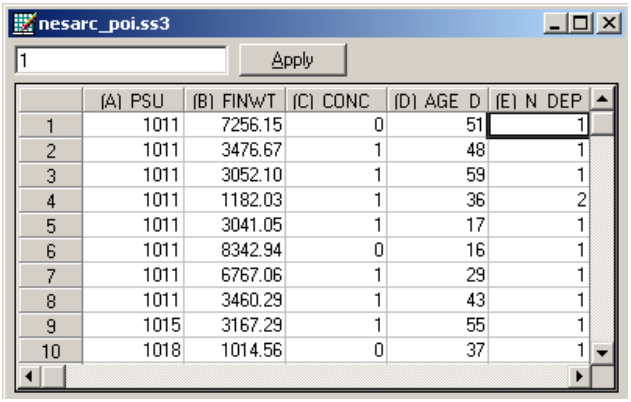
Ordinary quadrature is a numeric method for evaluating multi-dimensional integrals. For mixed-effect models with count and categorical outcomes, the log-likelihood function is expressed as the sum of the logarithm of integrals, where the summation is over higher-level units, and the dimensionality of the integrals equals the number of random effects.

A problem with ordinary quadrature is that it assumes a common location and scale for each level-2 unit. This assumption often requires the use of a large number of quadrature points to calculate the log-likelihood and derivatives to an acceptable level of accuracy. To overcome this problem with ordinary quadrature, SuperMix also offers a numeric integration procedure called adaptive quadrature. The adaptive quadrature procedure uses the empirical Bayes means and covariances, updated at each iteration to essentially shift and scale the quadrature locations of each higher-level unit in order to place them under the peak of the corresponding integral. To distinguish between the two quadrature methods, SuperMix uses the terminology non-adaptive quadrature (ordinary quadrature) and adaptive quadrature. To illustrate this, models in Section 5.2 will be fitted using the default method of adaptive quadrature, while models in Section 5.3 will use non-adaptive quadrature.

5.2 Two-level models for count outcomes from NESARC data

5.2.1 The data

The data set is from the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC), which was designed to be a longitudinal survey with its first wave fielded in 2001–2002. This data file has been used in some of the examples in Chapter 3, and contains information on the occurrences of major depression, family history of major depression and dysthymia of 2339 dysthymia respondents. After listwise deletion, the sample size is 1981.



	(A) PSU	(B) FINWT	(C) CONC	(D) AGE_D	(E) N_DEP
1	1011	7256.15	0	51	1
2	1011	3476.67	1	48	1
3	1011	3052.10	1	59	1
4	1011	1182.03	1	36	2
5	1011	3041.05	1	17	1
6	1011	8342.94	0	16	1
7	1011	6767.06	1	29	1
8	1011	3460.29	1	43	1
9	1015	3167.29	1	55	1
10	1018	1014.56	0	37	1

The variables of interest are:

- PSU denotes the Census 2000/2001 Supplementary Survey (C2SS) primary sampling unit.
- FINWT represents the NESARC weights sample results used to form national level estimates. The final weight is the product of the NESARC base weight and other individual weighting factors.
- CONC_DEP contains the information captured in field S4CQ3A6 of the NESARC data. It represents the response to the statement "Often had trouble

concentrating/keeping mind on things," with 1 indicating "Yes," and 0 indicating "No."

- AGE_DEP is based on field S4CQ7AR of the NESARC data. It represents the age at onset of first episode.
- N_DEP is recoded from field S4CQ6A of the NESARC data, and gives the number of depression/dysthymia episodes. This is the count variable we would like to use as outcome variable in the examples to follow.

5.2.1.1 Exploring the data

Inspecting the distribution of the intended outcome variable, N_DEP, before starting with the model is important. In the case of a count variable, this can easily be done by producing a bar chart of the observed frequencies of occurrence captured by the count variable. Select the **File**, **Data-based Graph**, **Univariate** option from the main SuperMix window and request a bar chart before clicking the **Plot** button.

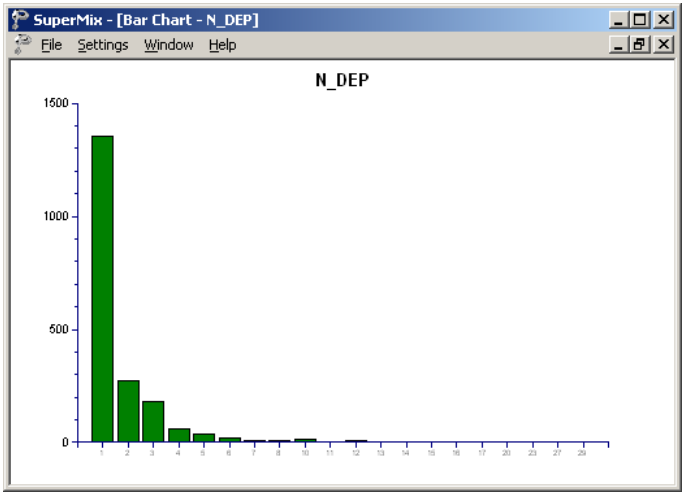


Figure 5.3: Bar chart for count variable N_DEP

The frequency bar chart for the count variable N_DEP shown in Figure 5.3 is obtained. We note that the number of depression episode ranges from 1 to 29, with most respondents having a small number of reported episodes of depression.

5.2.2 A 2-level Poisson model with 2 predictors

5.2.2.1 The model

The first model fitted to the data explores the relationship between N_DEP and the variables indicating concentration (or lack thereof) and age, as represented by the variables CONC_DEP and AGE_DEP.

The level-1 model is

$$\log\left(\lambda_{ij}\right) = \beta_0 + \beta_1 \times \text{CONC\_DEP}_{ij} + \beta_2 \times \text{AGE\_DEP}_{ij}$$

where the expected number of depression episodes is $\lambda_{ij} = E\left(\text{N\_DEP}_{ij}\right)$.

The level-2 model is

$$\beta_0 = b_{00} + v_{i0},\ \beta_1 = b_{10} \text{ and } \beta_2 = b_{20}.$$

Another way of writing the combined model is

$$\log\left(\lambda_{ij}\right) = b_{00} + b_{10} \times \text{CONC\_DEP}_{ij} + b_{20} \times \text{AGE\_DEP}_{ij} + v_{i0}.$$

In this model, $e^{b_{00}}$ denotes the average expected count of depression episodes, and b_{10} represents the estimated coefficient for the respondent's level of concentration.

Taking exponents on both sides, we also have

$$\begin{aligned}\lambda_{ij} &= e^{b_{00} + b_{10} \times \text{CONC\_DEP}_{ij} + b_{20} \times \text{AGE\_DEP}_{ij} + v_{i0}} \\ &= e^{b_{00}} e^{b_{10} \times \text{CONC\_DEP}_{ij}} e^{b_{20} \times \text{AGE\_DEP}_{ij}} e^{v_{i0}}\end{aligned}$$

For a person who had problems concentrating (CONC_DEP = 1), the expected average number of episodes $e^{b_{00}}$ is multiplied by e^{β_1}, compared to an expected count of $e^{b_{00}}$ for a person for whom CONC_DEP = 0. Similarly, an increase of one year in age increases the estimated number of episodes by a factor of $e^{b_{20}}$. For

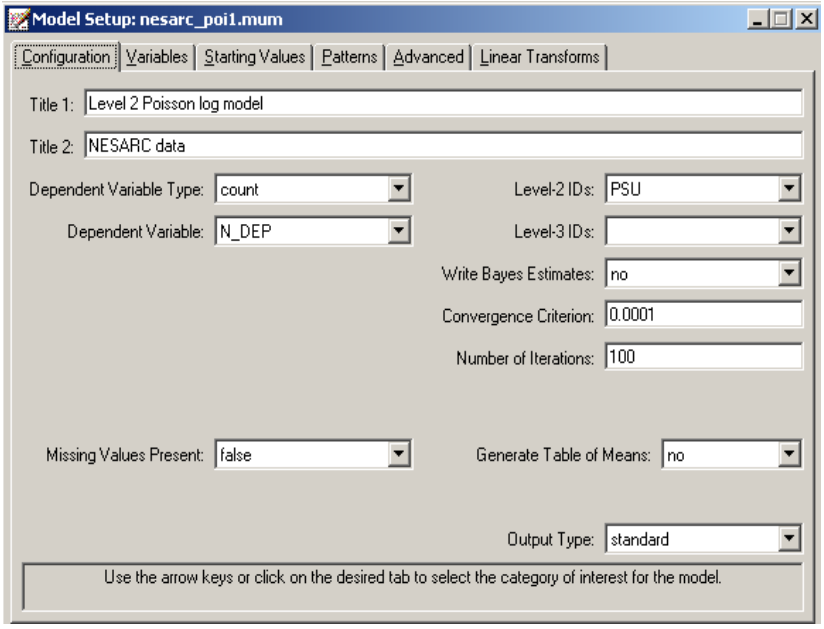
example, a respondent with concentration problems who is two years older than another respondent who had no concentration problems is expected to have $e^{b_{00}} e^{b_{10}} e^{2b_{20}}$ episodes compared to only $e^{b_{00}}$ episodes for the younger person without concentration problems.

The random part of the model is represented by $e^{v_{i0}}$, which denotes the variation in average count of depression episodes over PSU and between respondents (or, in other words, over respondents nested within PSU). For a Poisson distribution, the assumption of normality at level 1 is not realistic, as the level-1 random effect can only assume a number of distinct values. Thus, this random effect cannot have homogeneous variance.

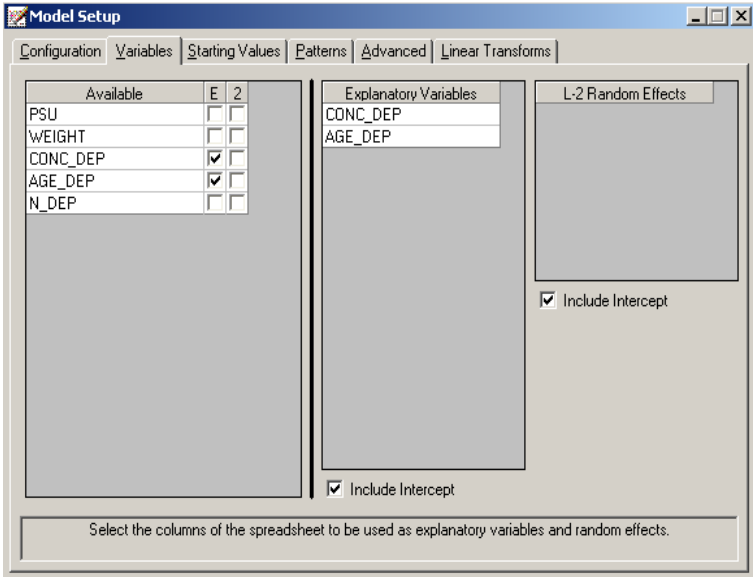
5.2.2.2 Setting up the analysis

Open the SuperMix spreadsheet **nesarc_poi.ss3** used during the exploratory analysis. From the main menu bar, select the **File, New Model Setup** option. The **Model Setup** window that appears has six tabs. In this example, only three tabs are used: the **Configuration**, **Variables,** and **Advanced** tabs.

The **Configuration** screen is the first tab on the **Model Setup** window. It enables the user to define the outcome variable and the level-2 and level-3 IDs. Some other settings such as missing values, convergence criterion, number of iterations, etc. can be specified here. For all the available settings, please refer to Section 2.4. To obtain the model we discussed, start by selecting the outcome variable N_DEP from the **Dependent Variable** drop-down list box. Indicate that it is a count variable by selecting the count option from the **Dependent Variable Type** drop-down list box. Next, describe the hierarchical structure of the data by selecting the level-2 ID, PSU, from the **Level-2 IDs** drop-down list box. Enter a title in the **Title** text boxes, and 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. To include the variables CONC_DEP and AGE_DEP as predictor variables, check the **E** check boxes next to the variables' names. Note that, as the variables are selected, the selected variables are listed in the **Explanatory Variables** grid. After selection, the screen below is obtained. Note that the **Include Intercept** check boxes in the **Explanatory Variables** grid and **L-2 Random Effects** are checked by default, indicating that an intercept term will automatically be included in the fixed and random parts of the model.



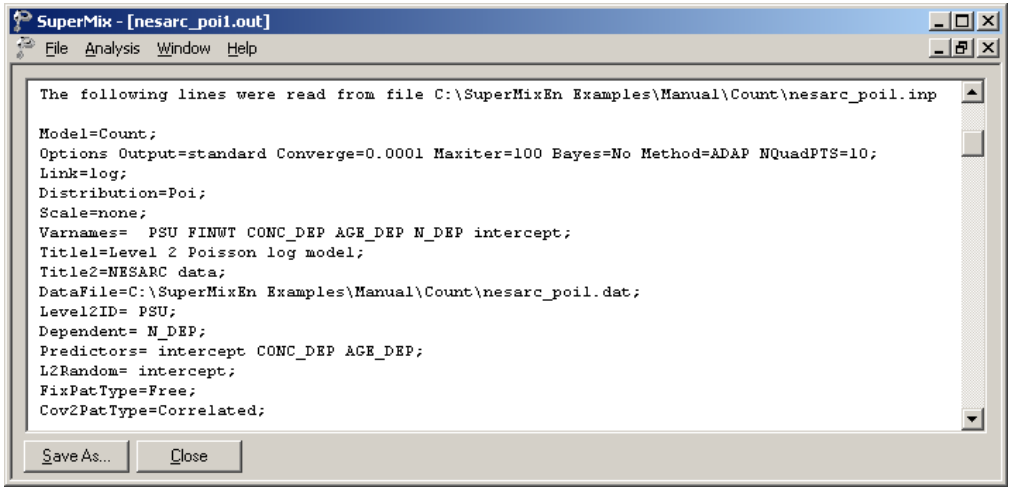
Before running the analysis, the model specifications have to be saved. Select the **File**, **Save As** option, and provide a name (**nesarc_poi1.mum**) for the model specification file. Run the analysis by selecting the **Run** option from the **Analysis** menu.

5.2.2.3 Discussion of results

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

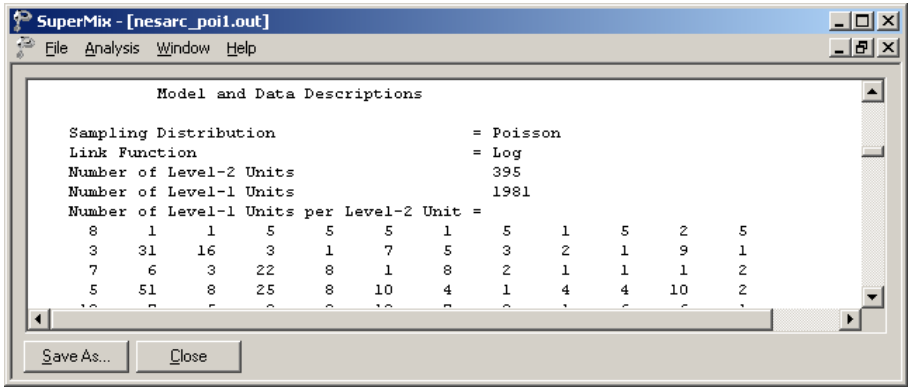
Program information and syntax

As shown below, the syntax for the model setup is printed in the output file. The first line of the syntax shows the option Model = Count, which indicates the outcome variable is a count variable. The Options syntax line corresponds to the settings on the **Configuration** screen. The Link = log and Distribution = Poi options specify the use of a Poisson distribution with a log link function for the fitted model.



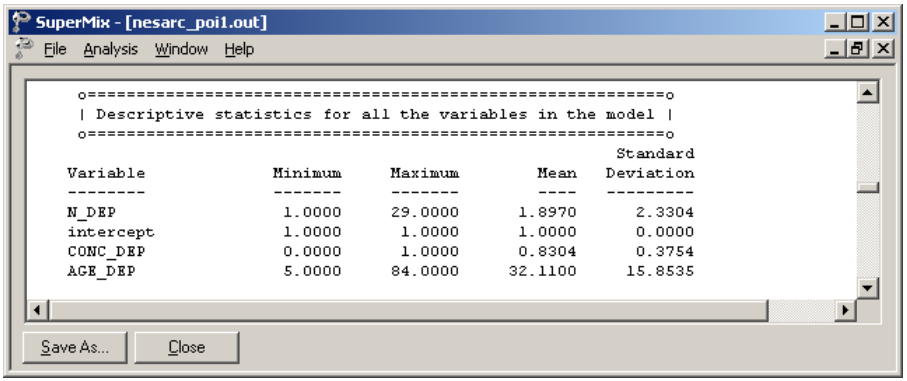
Model and data description

A description of the hierarchical structure follows the syntax: data from a total of 395 PSU and 1981 respondents were included at levels 2 and 1 of the model. In addition, an enumeration of the number of respondents nested within each of the 395 PSUs is provided.



Descriptive statistics

The data summary is followed by descriptive statistics for all the variables included in the model. The mean of 1.8970 and standard deviation of 2.3304 are reported for the outcome N_DEP indicating that, on average, 1.8970 episodes of depression were recorded.

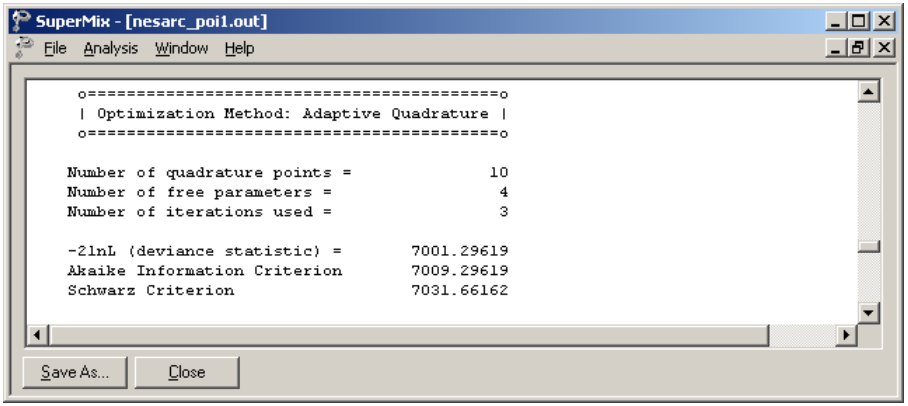


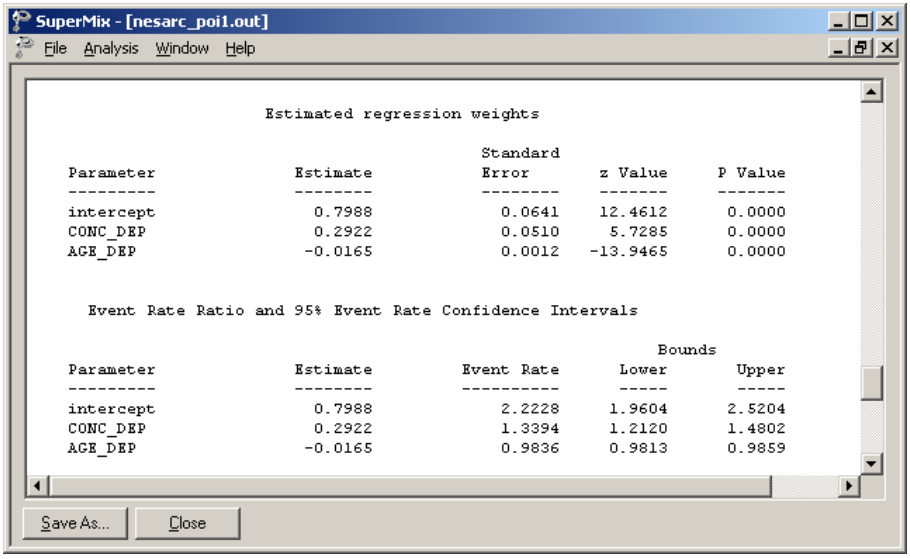
Descriptive statistics are followed by the results for a fixed-effects-only model, *i.e.* a model without random coefficients.

Fixed effects results

At the top of the final results, the number of iterations required for convergence of the iterative procedure is given.

Next, the number of quadrature points per dimension is reported which, in this case, is the default number of points. The log likelihood and the deviance, which is defined as $-2\ln L$, are listed next. For a pair of nested models, the difference in $-2\ln L$ values has a χ^2 distribution, with degrees of freedom equal to the difference in number of parameters estimated in the models compared.



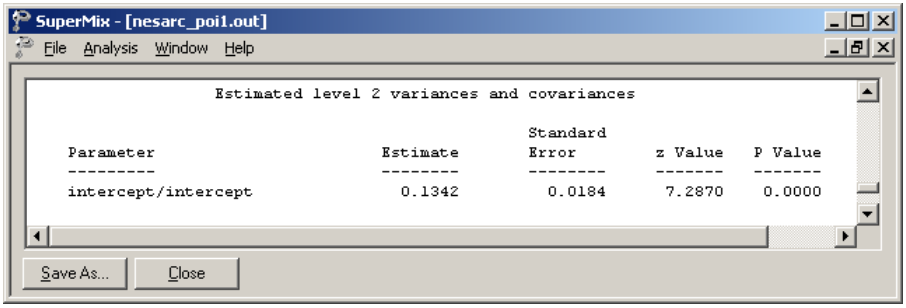


The estimated intercept is 0.7982, which means that the average number of depression episodes is $e^{0.7982}=2.2215$, implying that on average the number of episodes is about two. The estimated coefficient for CONC_DEP is 0.2922, which indicates that respondents who had trouble concentrating on things tended to have $2.2215e^{0.2922} = (2.2215)(1.3394)=2.9754$ episodes at the same age as respondents without concentration problems. The estimate of the effect of age at the onset of the first episode (AGE_DEP) shows that the onset age does not affect the number of episodes much, since $e^{-0.0165}=0.98$. A slight reduction in the expected number of

episodes is expected with increasing age. If one compares two typical respondents with reported concentration problems, but with one respondent ten years older than the other, one would expect the older respondent to have $(2.2215)(1.3394)e^{10(-0.0165)}$ =2.5229 episodes, compared to 2.9268 expected episodes for the younger respondent. In other words, the longer it takes for the first episode to occur, the fewer episodes a respondent is expected to have. Of course, it has to be kept in mind that the younger a respondent is at the first episode, the longer that person must live with the condition and thus the more time there is for subsequent episodes to occur.

Random effects results

The output for the level-2 random effect variance term follows next. The estimated variation in the average estimated N_DEP at level 2 is 0.1347, which is highly significant. Respondents are different in terms of their average expected number of episodes, holding all other variables constant.



Level-1 variation for Poisson distribution

The variance-to-mean ratio is a measure of the dispersion of a probability distribution:

$$R = \text{variance-to-mean ratio} = \frac{\sigma^2}{\mu}$$

For the Poisson distribution, where the variance equals the mean, this implies $R = 1$. Thus, we use a value of one as our level-1 variation. In the cases when over-dispersion ($R > 1$) or under-dispersion ($R < 1$) is assumed, different level-1 variation values will apply. The details of these scenarios are not discussed in this guide.

5.2.2.4 Interpreting the results

Estimated outcomes for groups: unit-specific results

First, we substitute the regression weights and obtain the following function for $\log\left(\text{N\_DEP}_{ij}\right)$:

$$\begin{aligned}\log\left(\hat{\text{N\_DEP}}_{ij}\right) &= \hat{b}_{00} + \hat{b}_{10} \times \text{CONC\_DEP}_{ij} + \hat{b}_{20} \times \text{AGE\_DEP}_{ij} \\ &= 0.7982 + 0.2922 \times \text{CONC\_DEP}_{ij} - 0.0165 \times \text{AGE\_DEP}_{ij}.\end{aligned}$$

For example, at age 40, the estimated $\log\left(\text{N\_DEP}_{ij}\right)$ for a typical respondent who does not often have trouble concentrating (CONC_DEP = 0), we find that

$$\begin{aligned}\log\left(\hat{\text{N\_DEP}}_{ij}\right) &= \hat{\beta}_0 + \hat{\beta}_1 \times \text{CONC\_DEP}_{ij} + \hat{\beta}_2 \times \text{AGE\_DEP}_{ij} \\ &= 0.7982 + 0.2922 \times \text{CONC\_DEP}_{ij} - 0.0165 \times \text{AGE\_DEP}_{ij} \\ &= 0.7982 + 0.2922 \times 0 - 0.0165 \times 40 \\ &= 0.1382.\end{aligned}$$

Keeping in mind that we defined the relationship between λ and the predictors as

$$\log\left(\lambda_{ij}\right) = \beta_1 x_{i1} + \cdots + \beta_p x_{ip},$$

it follows that

$$\hat{\lambda}_{ij} = e^{0.1382} = 1.1482.$$

We can estimate the count of the occurrence of depression episodes for typical individuals of different ages in the same way. Results are summarized in Table 5.1. The results show a decrease in the expected number of episodes with increasing age, regardless of whether they had concentration problems or not.

Table 5.1: Estimated number of episodes under the Poisson log model

AGE_DEP	10	20	30	40	50	60	70
CONC_DEP = 1	2.5229	2.1391	1.8138	1.5379	1.3040	1.1056	0.9374
CONC_DEP = 0	1.8836	1.5971	1.3542	1.1482	0.9736	0.8255	0.6999

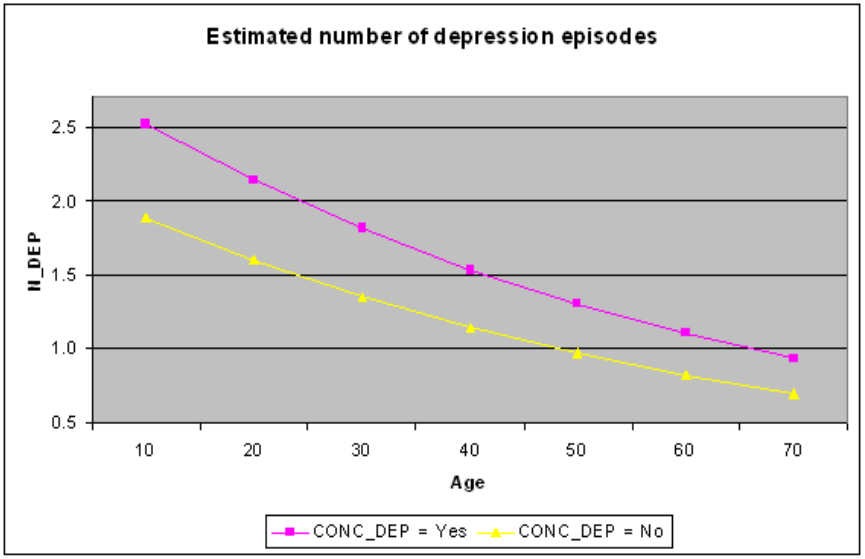


Figure 5.4: Expected number of episodes for two groups

The results in Table 5.1 can also be presented graphically, as shown in Figure 5.4. We clearly see that the correspondents who often had trouble concentrating

(CONC_DEP = 1) have a higher estimated number of depression episodes. It also shows that the number of episodes is expected to decrease as people get older.

Level 2 ICC

The percentage of variance explained over level-2 units, or intraclass correlation coefficient (ICC), is calculated as

$$ICC = \frac{\text{level-2 variation}}{\text{level-1 variation} + \text{level-2 variation}}$$

In this example, under the assumption that the level-1 variation is fixed at a value of one, we have

$$ICC = \frac{0.1347}{1 + 0.1347} \times 100\% = 11.8\%$$

We can conclude that most of the unexplained variation in the outcome (approximately 78%) is between measurements at the lowest level of the hierarchy.

5.2.3 A 2-level negative binomial model with 2 predictors

5.2.3.1 The model

In Section 5.2.2, a Poisson model was fitted to the data. It was also noted that a Poisson distribution has an important property: the mean number of occurrences is equal to the variance. The negative binomial distribution is an alternative distribution that may also be used to describe the properties of count variables. If the assumption of a Poisson distribution is reasonable, one would expect a model using a negative binomial distribution with a very small dispersion parameter to produce results that correspond closely to those obtained for the Poisson model. In this section, we fit a negative binomial model, utilizing the same predictors and a small dispersion parameter, to the NESARC data. Again, adaptive quadrature is used as the method of optimization.

Recall that the negative binomial distribution can be expressed as

$$f\left(y_i\right)=\frac{\Gamma\left(y_i+1/\alpha\right)}{\Gamma\left(y_i+1\right)\Gamma\left(1/\alpha\right)}\times\frac{\left(\alpha\mu_i\right)^{y_i}}{\left(1+\alpha\mu_i\right)^{y_i+1/\alpha}}$$

with $\Sigma\left(y_i\right)=\mu_i+\alpha\mu_i^2$ where α denotes an additional parameter and it can no longer be assumed that the variance is a known function of the mean. We assume α to be a fixed parameter.

The model fitted to the data explores the relationship between N_DEP and the variables indicating concentration (or lack thereof) and age, as represented by the variables CONC_DEP and AGE_DEP.

The level-1 model is

$$\log\left[\mathrm{E}\left(\mathrm{N\_DEP}_{ij}\right)\right]=\beta_0+\beta_1\times\mathrm{CONC\_DEP}_{ij}+\beta_2\times\mathrm{AGE\_DEP}_{ij}$$

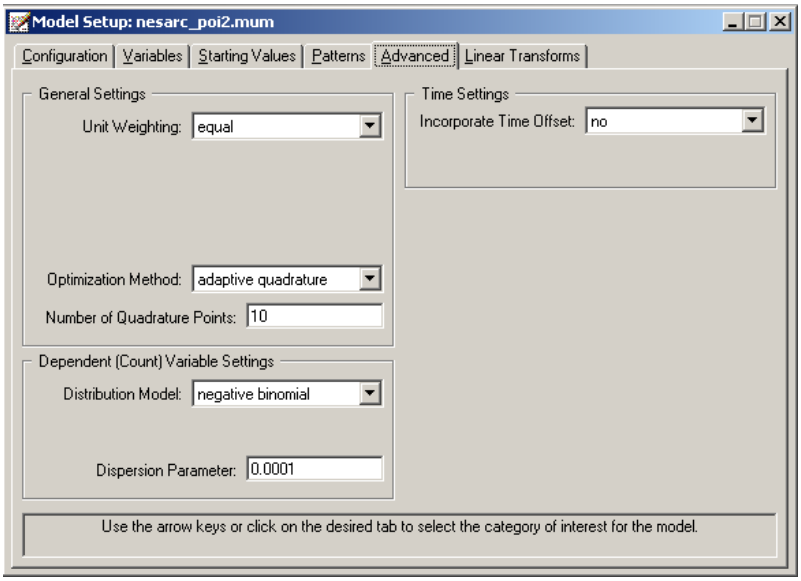
The level-2 model is

$$\beta_0=b_{00}+v_{i0},\ \beta_1=b_{10}\ \text{and}\ \beta_2=b_{20}.$$

5.2.3.2 Setting up the analysis

Using the SuperMix spreadsheet **nesarc_poi.ss3** and model specification file **nesarc_poi1.mum** from the previous section, we now set up a negative binomial model for these data.

Start by saving the model specification file under the new name **nesarc_poi2.mum** using the **File, Save As** option. Next, click on the **Advanced** tab of the **Model Setup** window. This is the only tab on which modifications have to be made to change the previously specified Poisson model to a negative binomial model. Set the **Distribution Model** to **negative binomial**, and the **Dispersion Parameter** to 0.0001 to obtain an **Advanced** tab as shown below.



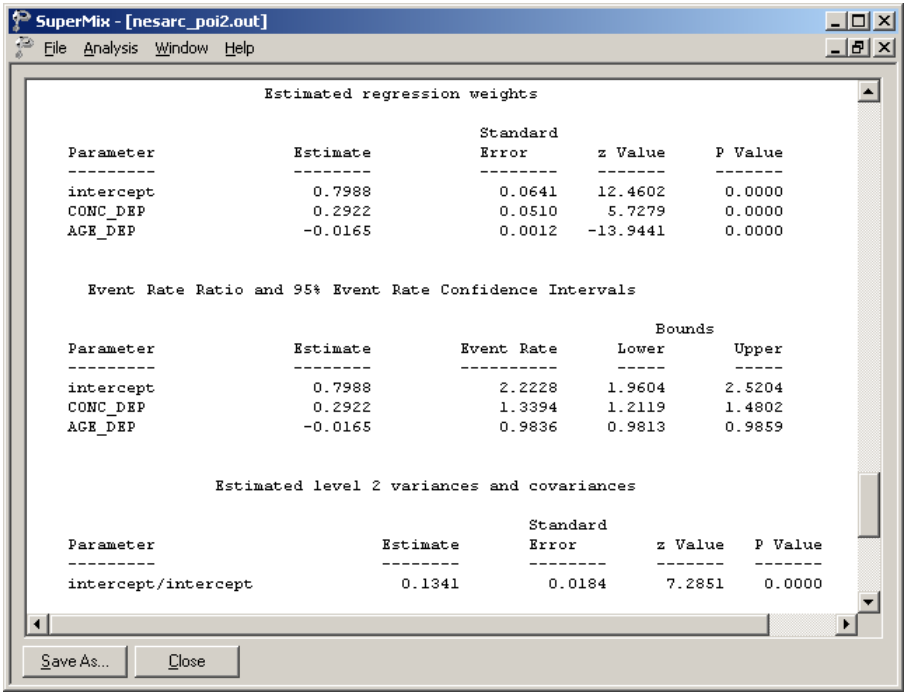
Save the revised model specification file, and click the **Analysis**, **Run** option to start the iterative process.

5.2.3.3 Discussion of results

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

Fixed and random effect results

The estimated regression coefficients for fixed effects in the model are shown below. Recall that the estimated coefficients of the intercept, CONC_DEP, and AGE_DEP under the Poisson model in Section 5.2.2 were 0.7982, 0.2922, and −0.0165 respectively. The estimated variation in the average estimated N_DEP at level-2 was 0.1347, and highly significant. The similarity of the results obtained under these two models indicate that the specification of a Poisson distribution model is reasonable for this data.



5.2.4 Weighted 2-level models

5.2.4.1 The data

The sampling frame of many multistage surveys frequently entails selection of units with known, but unequal, selection probabilities. This situation is the result of a number of design factors, of which the cost of doing the survey is an important consideration. When this is the case, it is appropriate to weight observations in order to produce unbiased estimates of population parameters.

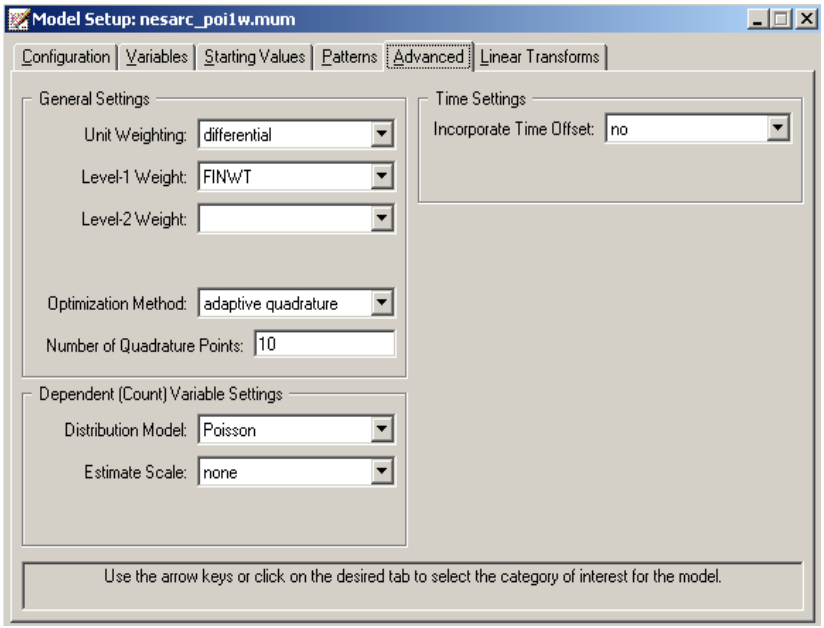
Recall from Section 5.2.1 that the data also included a weight variable. The variable FINWT represents the NESARC weights sample results used to form national-level estimates. The final weight is the product of the NESARC base weight and other

individual weighting factors. In this section, we explore the effect of inclusion of the weights on the results obtained in Sections 5.2.2 and 5.2.3.

5.2.4.2 Setting up the analysis

The models remain the same, with only the selection of the weight variable on the **Advanced** tab of the **Model Specification** screen to be added. Below, we show how this is done in the case of the Poisson distribution model.

Open the model specification file for the Poisson distribution model (**nesarc_poi1.mum**) and click on the **Advanced** tab. Change the **Unit Weighting** field from its default value of equal to differential. Next, select the variable FINWT from the **Assigned Weight** drop-down list box that appears when the selection has been made in the **Unit Weighting** field. The completed **Advanced** tab is shown below.



Save the specification file as **nesarc_poi1w.mum**, and run the analysis.

5.2.4.3 Discussion of results

Results for this analysis are reported in Table 5.2 below. The results from the unweighted Poisson distribution model are included in order to facilitate evaluation of the impact of the weights on the results.

Table 5.2: Comparison of results for weighted and unweighted Poisson models

Parameter	Unweighted model		Weighted model	
	Estimate	Standard error	Estimate	Standard error
intcept	0.7982	0.0641	0.7229	0.0659
CONC_DEP	0.2922	0.0510	0.3055	0.0532
AGE_DEP	–0.0165	0.0012	–0.0156	0.0013
Level-2 variance	0.1347	0.0184	0.1373	0.0189

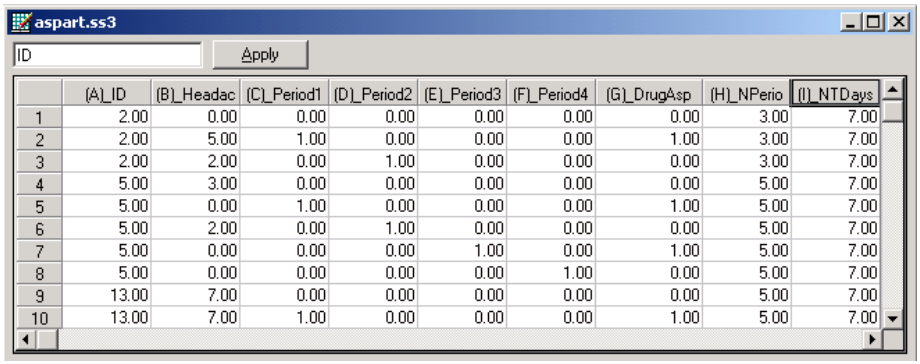
Results for the two models are very similar, and interpretation of the results of both models would lead to the same conclusions, both in terms of significance and in terms of the expected number of depression episodes. However, this is more the exception than the rule – users are cautioned to use weight variables whenever they are available in order to prevent skewed or biased results that may occur when weights are excluded in the analysis of a disproportionally drawn sample.

5.3 Two-level models for count outcomes from ASPART data

5.3.1 The data

The data for this example are taken from a paper by McKnight and Van Den Eeden (1993), who reported on the number of headaches in a two treatment, multiple period crossover trial. Specifically, the number of headaches per week was repeatedly measured for 27 patients. Following a seven day placebo run-in period, subjects received either aspartame or placebo in four seven-day treatment periods

according to the double-blind crossover treatment design. Each treatment period was separated by a washout day. The sample size is 122. Data for the first 10 observations of all the variables used in this section are shown below in the form of a SuperMix spreadsheet window for **aspart.ss3**.


aspart.ss3

ID | Apply

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

The variables of interest are:

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

5.3.2 A 2 level Poisson model with random intercept

5.3.2.1 The model

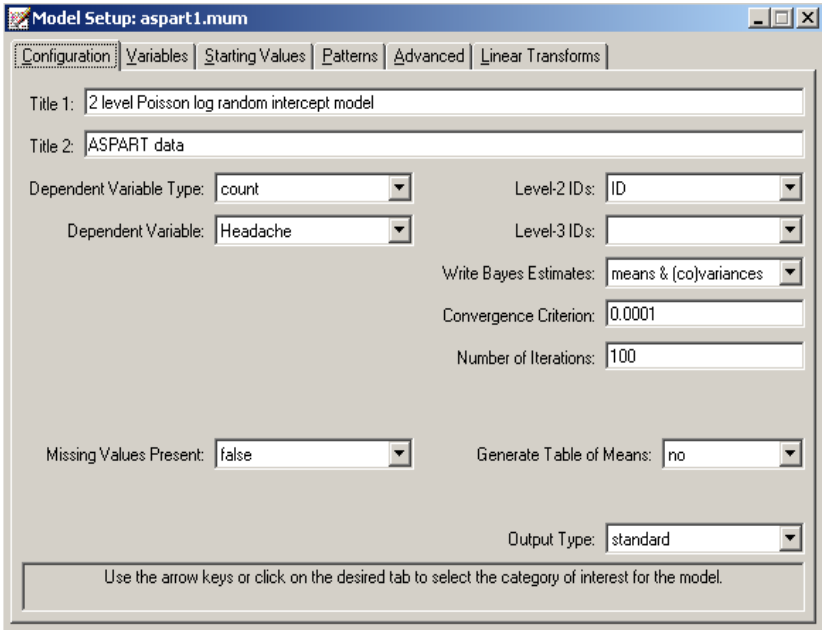
To model the relationship between the number of headaches during the week (Headache) and the treatment indicators (Period1 to Period4) and the type of drug administered (DrugAsp), the following Poisson regression model with a random intercept may be used:

$$\log\left(\lambda_{ij}\right) = \beta_0 + \beta_1 \times \text{Period1}_{ij} + \beta_2 \times \text{Period2}_{ij} + \beta_3 \times \text{Period3}_{ij} + \beta_4 \times \text{Period4}_{ij} + \beta_5 \times \text{DrugAsp}_{ij} + v_{i0}$$

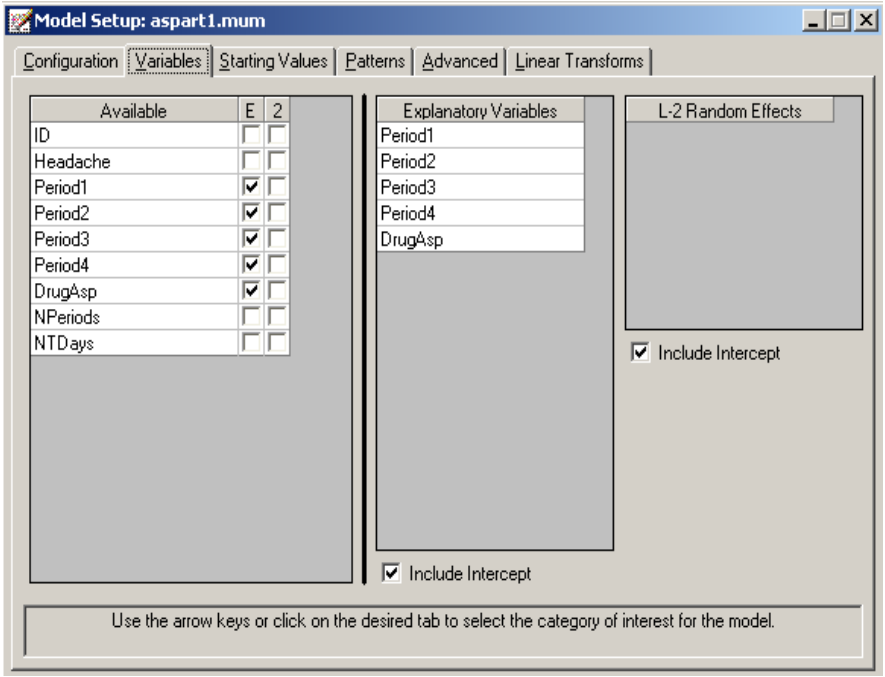
where λ_{ij} denotes the mean number of headaches of patient i for treatment period j; Period1_{ij}, Period2_{ij}, Period3_{ij} and Period4_{ij} denote the values of the dummy variables Period1, Period2, Period3 and Period4 for patient i for treatment period j respectively; DrugAsp_{ij} denotes the value of the DrugAsp for patient i for treatment period j; β_0, β_1, β_2, β_3, β_4 and β_5 denote unknown parameters; and v_{i0} denotes the random intercept for patient i for $i = 1, 2, \ldots, 27$ and $j = 0, 1, 2, 3$. This model is fitted to the data in **aspart.ss3** as described below.

5.3.2.2 Setting up the analysis

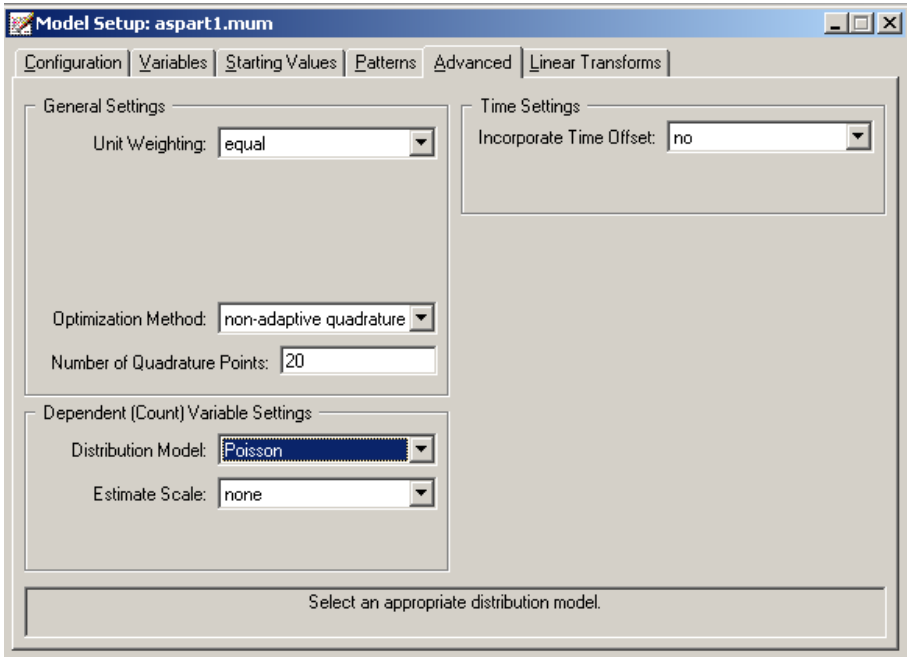
Start by opening the SuperMix spreadsheet **aspart.ss3**. Select the **New Model Setup** option on the **File** menu to load the **Model Setup** window. On the **Configuration** tab, enter the titles 2 level Poisson log random intercept model and ASPART data for the analysis in the **Title 1** and **Title 2** text boxes respectively. The count outcome variable Headache is selected from the **Dependent Variable** drop-down list box. The **Dependent Variable Type** drop-down list box is used to indicate that the outcome variable is a count. The variable ID, which defines the levels of the hierarchy, is selected as the **Level-2 ID** from the **Level-2 IDs** drop-down list box.



Next, click on the **Variables** tab to proceed with variable selection. The variables Period1, Period2, Period3, Period4, and DrugAsp are specified as the fixed effects of the model by checking the **E** check boxes for Period1, Period2, Period3, Period4, and DrugAsp in the **Available** grid. These actions produce the following **Variables** tab. By default, an intercept model is included in the fixed part of the model, along with a random intercept at level 2.



Finally, we click on the **Advanced** screen and keep all the default settings as shown below, except for those concering the method of estimation. Select **non-adaptive quadrature**, and set the quadrature points to 20. Before we can run the analysis, we have to save the model specifications to a file. This is accomplished by using the **Save** option on the **File** menu to open a **Save Mixed Up Model** dialog box. First enter the name **aspart1.mum** in the **File name** text box and then click on the **Save** button to save the file. The analysis is run by selecting the **Run** option from the **Analysis** menu. This produces the corresponding output file **aspart1.out**.

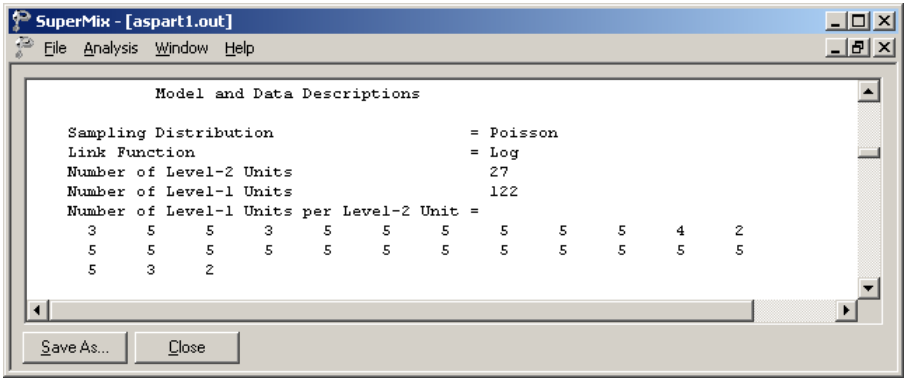


5.3.2.3 Discussion of results

Portions of this output file are shown below.

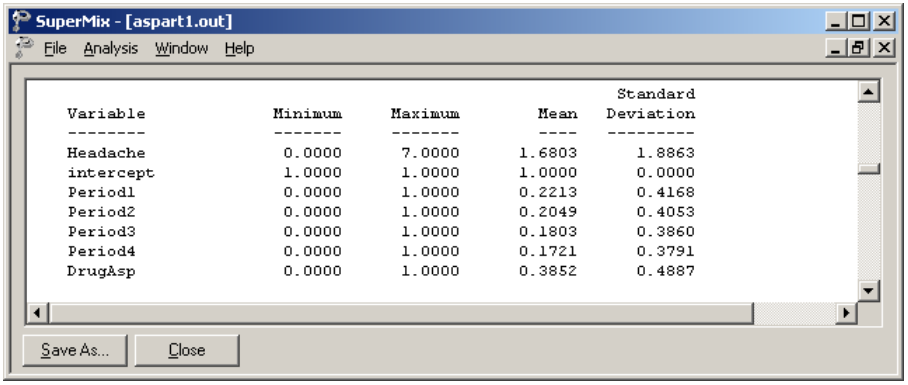
Model and data description

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



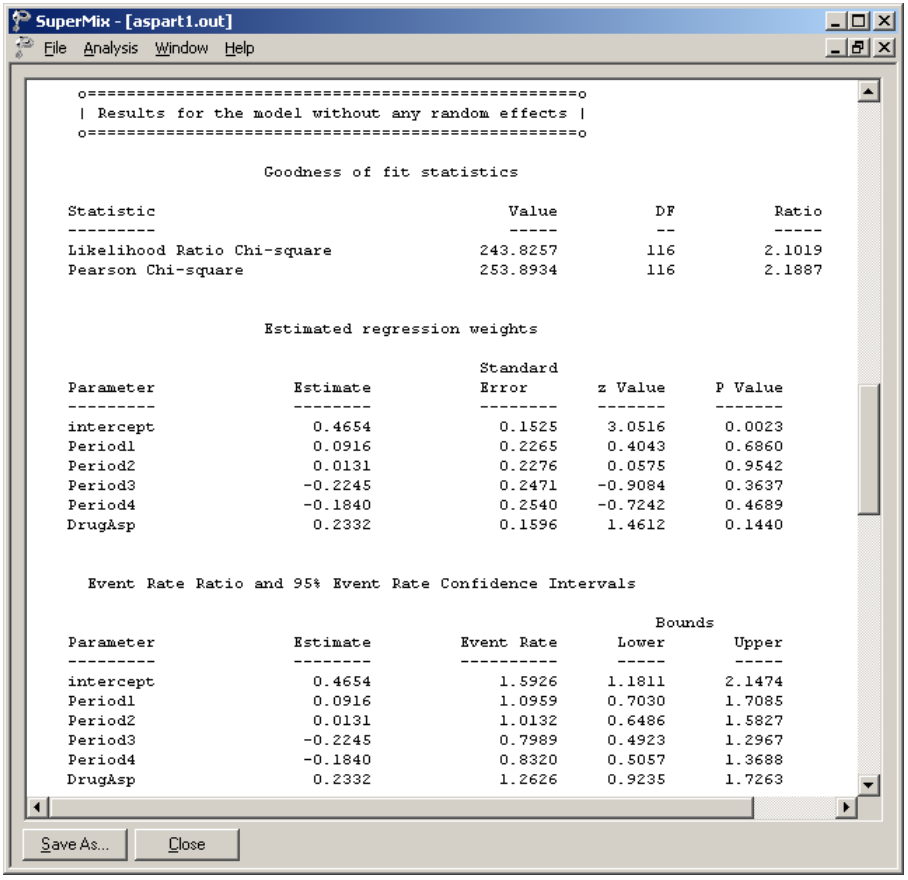
Descriptive statistics

The descriptive statistics for all the variables is shown next. The variance of Headache is $1.8863^2 = 3.5581$, which is substantially larger than the mean 1.6803. This might conflict with our assumption that the Poisson distribution is an appropriate choice for these data. As pointed out in Section 5.2.3, this can be verified by fitting a negative binomial model with a small dispersion parameter.



Results for the model without any random effects

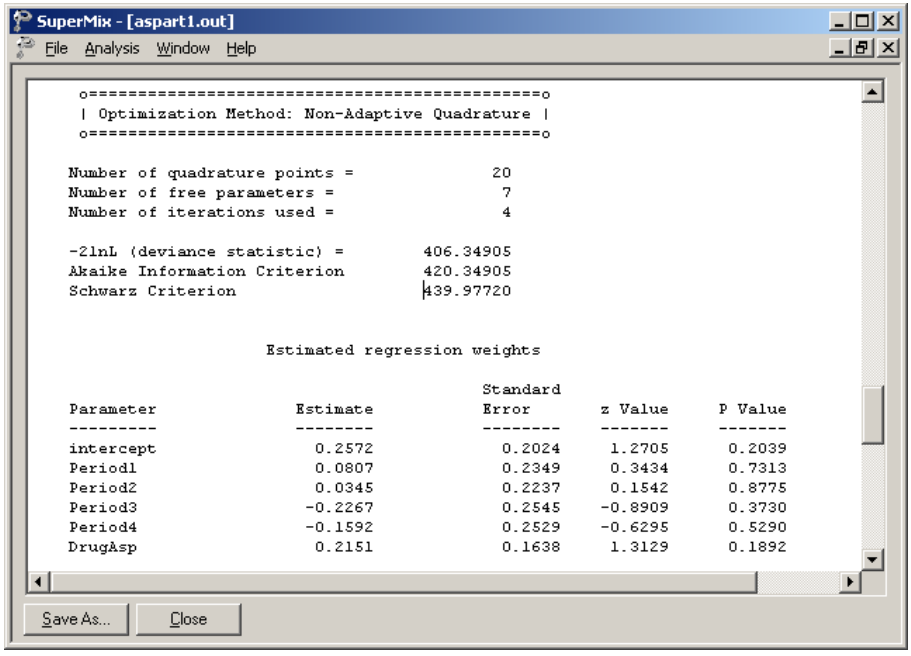
The results for the model without any random effects are shown below. In this section the goodness of fit statistics, estimated regression weights and event rate ratio and 95% event rate confidence intervals are included.



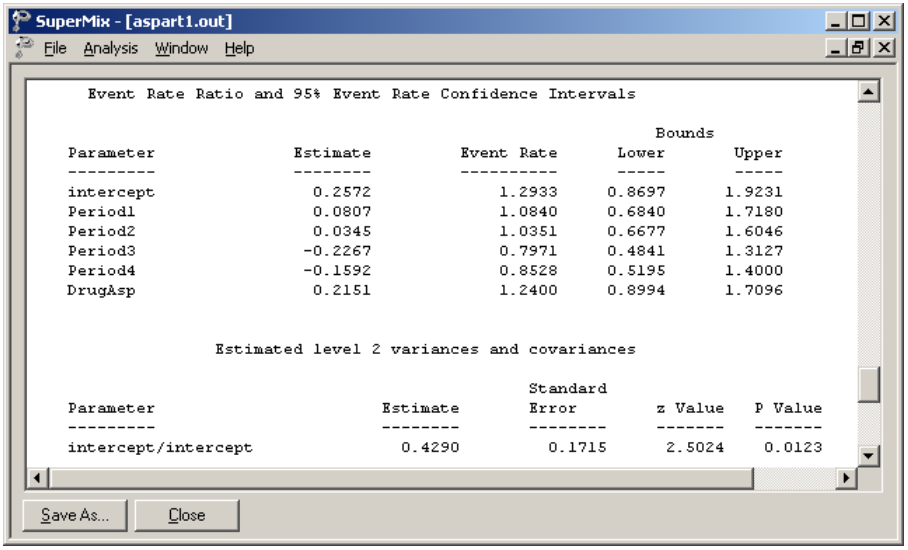
Fixed and random effect results

The final results are shown next. The number of iterations needed for convergence and the deviance information are given first, followed by the estimates.

The random-effect standard deviation is estimated as .643, and although a Wald test rejects the hypothesis that this parameter equals 0, use of the Wald test for testing whether variance parameters equal zero is questionable, since the Wald test is based on the assumption that parameters can assume any real value. Regarding the regression coefficients, all effects are non-significant. The results indicate that the model does not fit the data very well.



The event ratio and 95% event rate confidence interval and estimated level-2 variances and covariances are shown next to the estimated regression weights. The event ratios are the exponents ($e^{\hat{\beta}}$) of the estimated regression coefficients.



The random-effect variance is estimated as 0.429, and although a Wald test rejects the hypothesis that this parameter equals 0, use of the Wald test for testing whether variance parameters equal zero is questionable, since the Wald test is based on the assumption that parameters can assume any real value. Regarding the regression coefficients, all effects are non-significant. The results indicate that the model does not fit the data very well.

5.3.2.4 Interpreting the results

Estimated outcomes for groups: unit-specific results

The expected number of headaches can be obtained in the following fashion. First, we substitute the estimated coefficients in the model formulation

$$\log\left(\hat{\text{Headache}}_{ij}\right) = \hat{\beta}_0 + \hat{\beta}_1 \times \text{Period1}_{ij} + \hat{\beta}_2 \times \text{Period2}_{ij}$$
$$+ \hat{\beta}_3 \times \text{Period3}_{ij} + \hat{\beta}_4 \times \text{Period4}_{ij} + \hat{\beta}_5 \times \text{DrugAsp}_{ij}$$
$$= 0.2572 + 0.0807 \times \text{Period1}_{ij} + 0.0345 \times \text{Period2}_{ij}$$
$$-0.2267 \times \text{Period3}_{ij} - 0.1592 \times \text{Period4}_{ij} + 0.2151 \times \text{DrugAsp}_{ij}.$$

or, after taking exponents on both sides, as

$$\hat{\text{Headache}}_{ij} = \exp(0.2572 + 0.0807 \times \text{Period1}_{ij} + 0.0345 \times \text{Period2}_{ij}$$
$$-0.2267 \times \text{Period3}_{ij} - 0.1592 \times \text{Period4}_{ij} + 0.2151 \times \text{DrugAsp}_{ij}).$$

As an example, we calculate the expected number of headaches for a typical patient to whom aspartame was administered (DrugAsp = 1). During the first treatment period, we find that for such a patient

$$\hat{\text{Headache}}_{ij} = \exp(0.2572 + 0.0807 + 0.2151)$$
$$= 1.7385.$$

The expected numbers of headaches for a typical patient (again with DrugAsp = 1) for the second, third, and fourth treatment periods are calculated as

$$\hat{\text{Headache}}_{ij} = \exp(0.2572 + 0.0345 + 0.2151)$$
$$= 1.6600,$$
$$\hat{\text{Headache}}_{ij} = \exp(0.2572 - 0.2267 + 0.2151)$$
$$= 1.2784,$$

and

$$\begin{aligned}\hat{\text{Headache}}_{ij} &= \exp(0.2752 - 0.1592 + 0.2151) \\ &= 1.3677\end{aligned}$$

respectively. Complete results for all groups are given in Table 5.2.

Estimated outcomes for groups: population-average results

The latent response variable model,

$$y_{ij} = \mathbf{z}'_{(1)ij}\mathbf{b}_i + \mathbf{x}'_{(1)ij}\boldsymbol{\beta}_{(1)} + e_{ij},$$

makes the assumption that e_{ij} : $LID(0, \sigma_e^2)$. For a Poisson distribution it is assumed that $\sigma_e^2 = 1$. Under the assumption that $\mathbf{v}_i$ and e_{ij} are independently distributed, it follows that

$$\sigma_{y_{ij}}^2 = \mathbf{z}'_{ij}\boldsymbol{\Phi}_{v_i}\mathbf{z}_{ij} + \sigma_e^2.$$

The design effect d_{ij} is defined as

$$d_{ij} = \frac{\sigma_{y_{ij}}^2}{\sigma_e^2},$$

which, for the current model, may be calculated as

$$d_{ij} = \frac{\sigma_{y_{ij}}^2}{\sigma_e^2} = \frac{\text{var}(v_{i0}) + 1}{1} = 1.4290$$

where $\text{var}(v_{i0}) = 0.4290$, with v_{i0} denoting the random intercept coefficient. The estimated population-average probabilities (Hedeker & Gibbons, 2006) are obtained in a similar fashion as the unit-specific probabilities, after replacing the exponent in

the formula used for calculation of the estimated unit-specific probabilities with $\exp = \exp/\sqrt{d_{ij}}$ as shown below.

$$\bar{H}\text{eadache}_{ij} = \exp[(0.2572 + 0.0807 \times \text{Period1}_{ij} + 0.0345 \times \text{Period2}_{ij} - 0.2267 \times \text{Period3}_{ij} - 0.1592 \times \text{Period4}_{ij} + 0.2151 \times \text{DrugAsp}_{ij}) / \sqrt{1.4290}].$$

The expected unit-specific and population average probabilities are summarized in Table 5.3. We see that there is very little difference in the estimated number of headaches. This result is to be expected as the design effect is $\sqrt{1.4290} = 1.1954$.

Table 5.3: Estimated unit-specific and population average results for groups

DRUGASP	Period	Estimated headaches (unit-specific)	Estimated headaches (population-average)
0	1	1.4020	1.1728
0	2	1.3387	1.1199
0	3	1.0310	0.8624
0	4	1.1030	0.9227
1	1	1.7385	1.4543
1	2	1.6600	1.3886
1	3	1.2784	1.0694
1	4	1.3677	1.1441

5.3.3 A 2-level Poisson log model with an offset variable

5.3.3.1 The model

The previous analysis assumed that the counts were all observed for the same number of days. However, this was not the case since the number of treatment days in the period did vary to some degree. Most of the counts were based on the full seven days in the week; however, some observations were made only for 1 day in the given week. To take this into account, we need to specify a so-called OFFSET variable. The offset variable indicates the amount of time that each count is based

on. If OFFSET = no is specified, as was the case in the previous example, SuperMix assumes that all counts are based on the same amount of time.

The offset variable is introduced into the Poisson model in the following way:

$$\log\left(\hat{\lambda}_{ij}\right) = \log(\text{offset variable}) + \left[\mathbf{x}_{ij}^{'}\mathbf{b}_{i}\right]$$

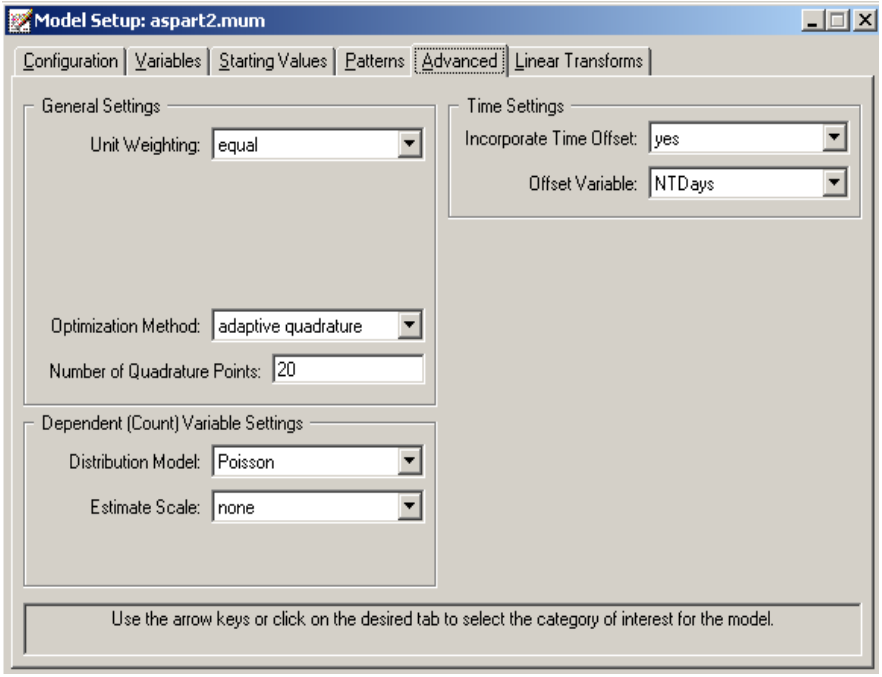
where $\mathbf{x}_{ij}$ represent the values of the covariates corresponding to level-1 unit j nested within level-2 unit i and $\mathbf{b}_i$ denotes the coefficient vector containing both fixed and random effects.

In the current situation, the variable NTDays is the appropriate choice as the OFFSET variable. The model to be fitted to the data now changes to:

$$\log\left(\text{Headache}_{ij}\right) = \log\left(\text{NTDays}\right) + (\beta_0 + \beta_1 \times \text{Period1}_{ij} + \beta_2 \times \text{Period2}_{ij} + \beta_3 \times \text{Period3}_{ij} + \beta_4 \times \text{Period4}_{ij} + \beta_5 \times \text{DrugAsp}_{ij} + v_{i0}).$$

5.3.3.2 Setting up the analysis

To create the model specifications for this model, start by opening **aspart.ss3** in a SuperMix spreadsheet window and using the **Open Existing Model Setup** option on the **File** menu to open the **Model Setup** window for **aspart1.mum**. On the **Configuration** screen, extend the title in the **Title 1** text box by adding the string "with Offset Variable." Next, click on the **Advanced** tab of the **Model Setup** window. Select yes from the **Incorporate Time Offset** drop-down list to activate the **Offset Variable** drop-down list box. Select the variable NTDays from the drop-down list of **Offset Variable** to produce the following **Advanced** tab.

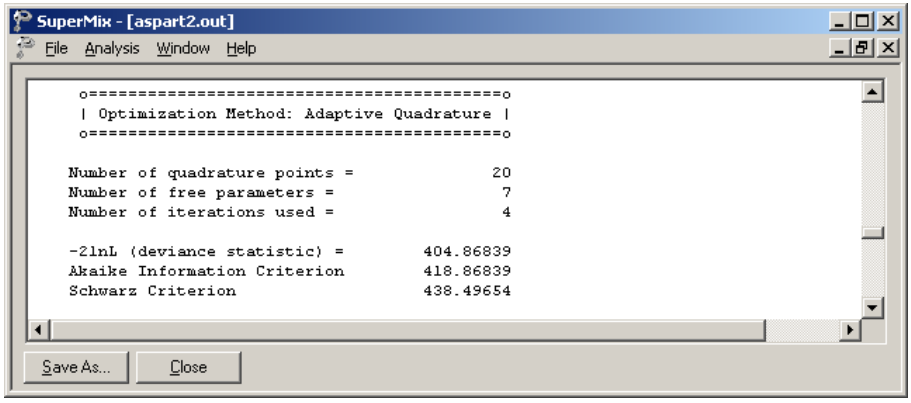


Save the changes to the file **aspart2.mum** by using the **Save As** option on the **File** menu and select the **Run** option on the **Analysis** menu to produce the output file **aspart2.out**.

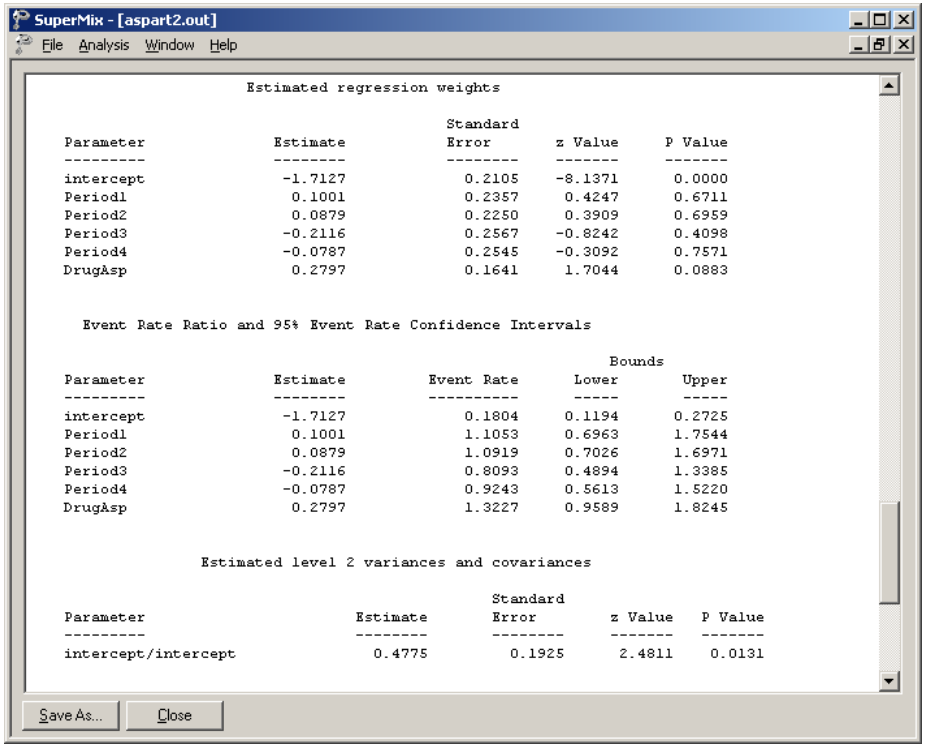
5.3.3.3 Discussion of results

Fixed and random effect results

Portions of this output file are shown below.



Results for this model differ from those obtained for the model without offset variable discussed in the previous section. While the overall trend in predictor coefficient estimates is similar, the intercept is now estimated as –1.7127, compared to 0.2572 previously. The variance in intercept over patients for this model is estimated as 0.4775 compared to 0.4290 previously.



5.3.3.4 Interpreting the results

Estimated outcomes for groups: unit-specific results

The expected number of headaches can be obtained in the following fashion. First, we substitute the estimated coefficients in the model formulation

$$\log\left(\hat{\text{Headache}}_{ij}\right) = \log\left(\text{NTDays}_{ij}\right) + (\hat{\beta}_0 + \hat{\beta}_1 \times \text{Period1}_{ij} + \hat{\beta}_2 \times \text{Period2}_{ij}$$

$$+ \hat{\beta}_3 \times \text{Period3}_{ij} + \hat{\beta}_4 \times \text{Period4}_{ij} + \hat{\beta}_5 \times \text{DrugAsp}_{ij})$$

$$= \log\left(\text{NTDays}_{ij}\right) + (-1.7127 + 0.1001 \times \text{Period1}_{ij} + 0.0879 \times \text{Period2}_{ij}$$

$$-0.2116 \times \text{Period3}_{ij} - 0.0787 \times \text{Period4}_{ij} + 0.2797 \times \text{DrugAsp}_{ij}),$$

or, after taking exponents on both sides, as

$$\begin{aligned}\hat{\text{Headache}}_{ij} = \text{NTDays}_{ij} \times \exp(&-1.7127 + 0.1001 \times \text{Period1}_{ij} + 0.0879 \times \text{Period2}_{ij} \\ &-0.2116 \times \text{Period3}_{ij} - 0.0787 \times \text{Period4}_{ij} + 0.2797 \times \text{DrugAsp}_{ij}).\end{aligned}$$

As most observations had a value of NTDays = 7, we start by considering typical patients with a full set of treatment days. We also assume that DrugAsp = 1, in other words, that aspartame rather than a placebo was administered.

During the first treatment period, we find that for such a patient

$$\begin{aligned}\hat{\text{Headache}}_{ij} &= 7\exp(-1.7127 + 0.1001 + 0.2797) \\ &= 7\exp(-1.3329) \\ &= 1.8460.\end{aligned}$$

The expected numbers of headaches for a typical patient (again with NTDays = 7 and DrugAsp = 1) for the second, third, and fourth treatment periods are calculated as

$$\begin{aligned}\hat{\text{Headache}}_{ij} &= 7\exp(-1.7127 + 0.0879 + 0.2797) \\ &= 1.8236,\end{aligned}$$

$$\begin{aligned}\hat{\text{Headache}}_{ij} &= 7\exp(-1.7127 - 0.2116 + 0.2797) \\ &= 1.3516,\end{aligned}$$

and

$$\begin{aligned}\hat{\text{Headache}}_{ij} &= 7\exp(-1.7127 - 0.0787 + 0.2797) \\ &= 1.5437\end{aligned}$$

respectively.

For a typical patient with only 5 treatment days, the expected numbers of headaches in each of the four treatment periods are 1.3186, 1.3026, 0.9654, and 1.1027 respectively.

When the expected numbers of headaches for a typical patient receiving aspartame under the Poisson model without offset variable (see previous section) and the Poisson model with offset variable are compared, we clearly see the impact of the inclusion of the offset variable on the estimated coefficients. These results are shown in Table 5.4.

Table 5.4: Comparison of results for Poisson models

Period	Without offset variable	With offset variable (NTDays = 7)	With offset variable (NTDays = 5)
1	1.7385	1.846	1.3186
2	1.6600	1.8236	1.3026
3	1.2784	1.3516	0.9654
4	1.3677	1.5437	1.1027

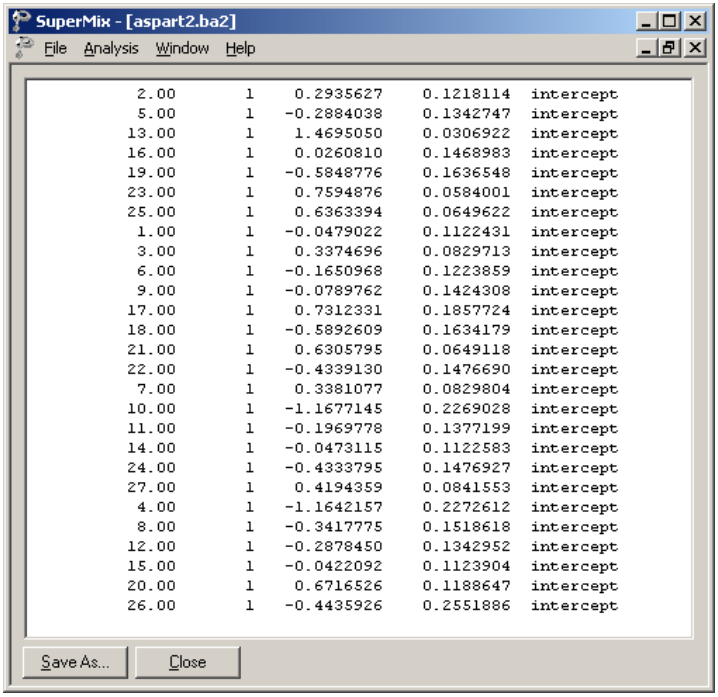
Level 2 Bayes results

As requested during the model specification stage, the empirical Bayes estimates of the random effects are written to the file **aspart2.ba2**. The first few lines of this file are shown below.

The file **aspart.ba2** contains five pieces of information per individual:

- the individual's ID,
- the number of repeated observations for that individual,
- the empirical Bayes estimate for that individual (which is the mean of the posterior distribution),

- the associated posterior standard deviation, and
- the name of the relevant random coefficient.



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

Graphical displays

Figure 5.5 is a comparison (represented by a dotted line) of the predicted average number of headaches reported by each patient when taking a placebo (left axis) as opposed to the predicted average number when the treatment is aspartame (right axis). From the graphical display, it appears as if all of the lines (each representing a patient) have a positive slope. The slopes become steeper as the number of headaches increases. This suggests an increase, albeit small, in the expected average number of headaches when aspartame is used. Note that patient 13, who reported a consistently high number of headaches at all occasions, was excluded from this graph.

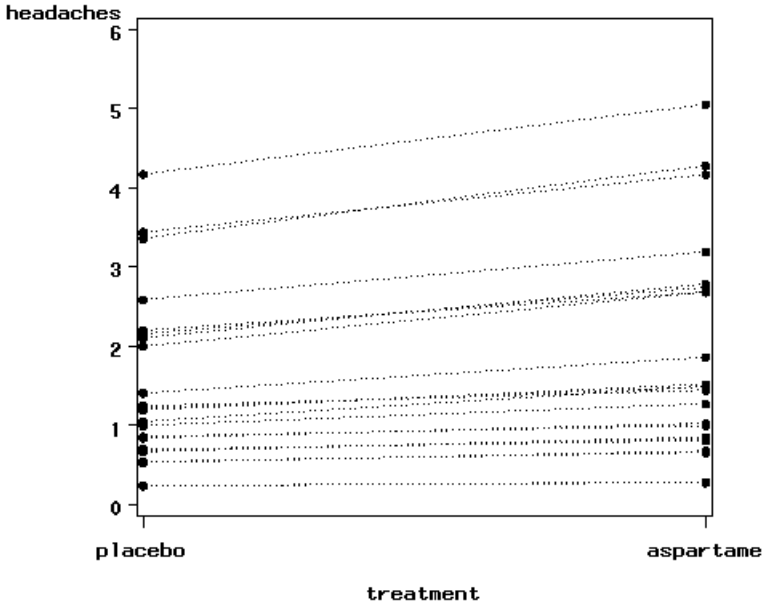
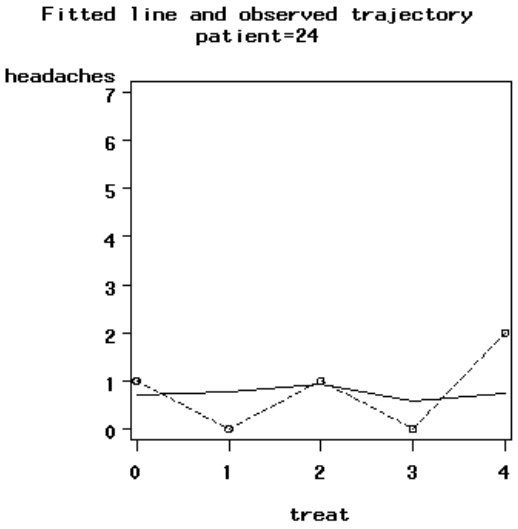
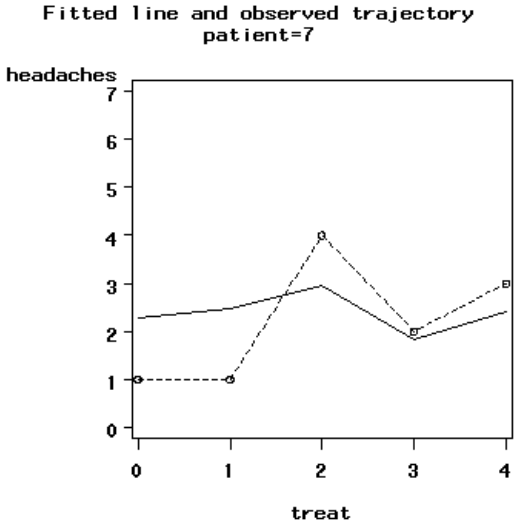
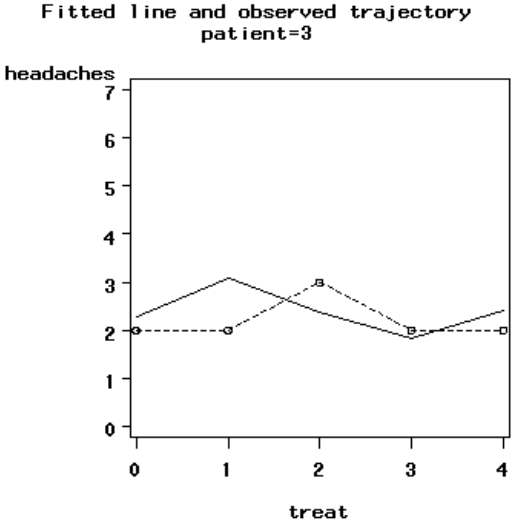
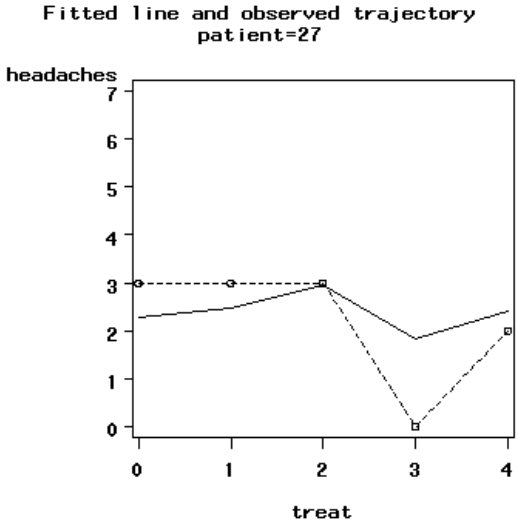


Figure 5.5: Predicted average number of headaches for placebo and aspartame


Fitted line and observed trajectory
patient=24
headaches
7
6
5
4
3
2
1
0
0
1
2
3
4
treat


Fitted line and observed trajectory
patient=7
headaches
7
6
5
4
3
2
1
0
0
1
2
3
4
treat


Fitted line and observed trajectory
patient=3
headaches
7
6
5
4
3
2
1
0
0
1
2
3
4
treat


Fitted line and observed trajectory
patient=27
headaches
7
6
5
4
3
2
1
0
0
1
2
3
4
treat

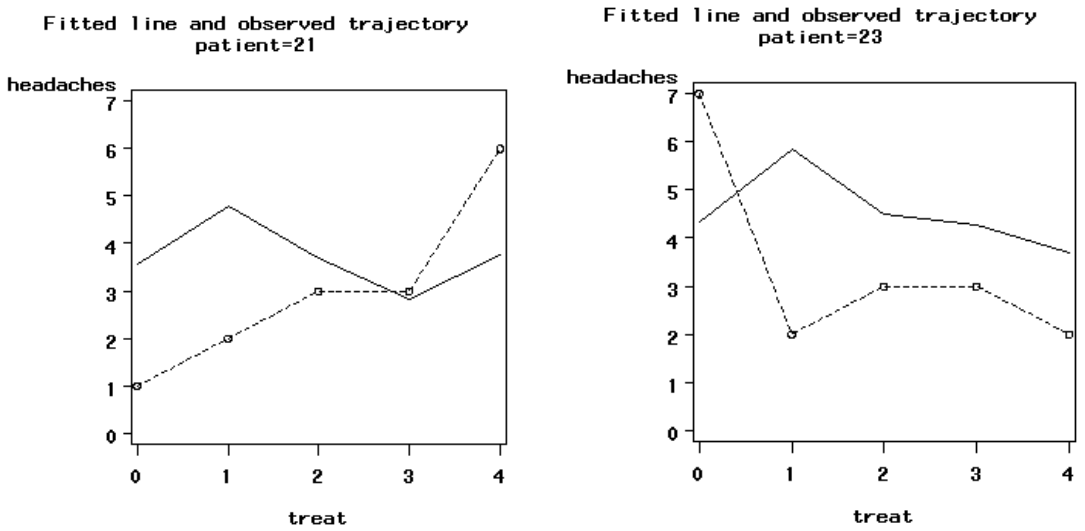


Figure 5.6: Fitted and observed trajectories

Figure 5.6 is a graphical display of the fitted trajectory (solid line) and observed trajectory (dotted line) for a sample of 6 patients. These displays are ordered from a patient who reported a relatively small number of headaches at the different treatment occasions to one who reported a relatively high number of headaches at the treatment occasions. A study of the fitted and observed trajectories reveals that, in general, the model fit is best when the number of headaches is smaller and becomes less accurate as the number of headaches increases. For patient 13, who is not represented in the graphical display, the number of predicted headaches is almost twice the number observed.

The fitted lines were obtained as

$$\hat{\text{Headache}}_{ij} = \text{NTDays}_{ij} \times \exp(-1.7127 + 0.1001 \times \text{Period1}_{ij} + 0.0879 \times \text{Period2}_{ij}$$
$$-0.2116 \times \text{Period3}_{ij} - 0.0787 \times \text{Period4}_{ij} + 0.2797 \times \text{DrugAsp}_{ij}) + \hat{b}_{i0}\,.$$

where $\hat{b}_{i0}$ is obtained from the **aspart2.ba2** file, shown previously in this section.

6 Models for ordinal outcomes

6.1 Introduction

The term "ordinal" is applied to variables where the response measure of interest is measured in a series of ordered categories. Examples of such variables include Likert scales and psychiatric ratings of severity. Nominal and ordinal outcome models can be seen as generalizations of the binary outcome model. The ordinal model becomes important when the outcome variable is not dichotomous, or not truly continuous. If an ordinal outcome is analyzed within a continuous model, such a model can yield predicted values outside the range of the ordinal variable. As with binary data, some transformation or link function becomes necessary to prevent this from happening. The continuous model can also yield correlated residuals and regressors when applied to ordinal outcomes because the continuous model does not take the ceiling and floor effects of the ordinal outcome into account. This can then result in biased estimates of regression coefficients, and is most critical when the ordinal variable in question is highly skewed. Armstrong & Sloan (1989) also report efficiency losses between 89% and 99% when comparing an ordinal to a continuous outcome, depending on the number of categories and distribution within the ordinal categories.

Extensive work on the development of methods for the analysis of ordinal response data has been undertaken by numerous researchers, including Hedeker & Gibbons (1994). These developments have focused on the extension of methods for dichotomous variables to ordinal response data, and have been mainly in terms of logistic and probit regression models. The proportional odds model proposed by McCullagh (1980) is a common choice for analysis of ordinal data. This model, which is described in detail in Section 6.2.2, is based on the logistic regression formulation.

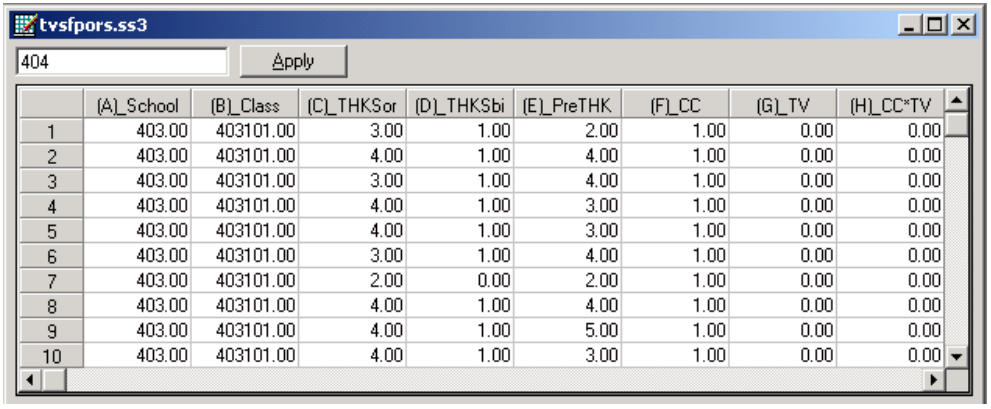
In this chapter we will now build on the dichotomous model discussed earlier and introduce the ordinal model, illustrating the use of this model using the TVSFP (Flay, *et. al.*, 1988) data previously used in this manual.

6.2 Two-level ordinal analysis of TVSFP data

6.2.1 The data

The study was designed to test independent and combined effects of a school-based social-resistance curriculum and a television-based program in terms of tobacco use and cessation.

The structure of this study indicates a three-level hierarchical structure. However, for illustration purposes in this chapter we will consider a two-level structure in which students are nested within schools. Data for the first 10 participants on most of the variables used in this section are shown below in the form of a **SuperMix** spreadsheet file, named **tvsfpors.ss3**, located in the **Examples\Ordinal** subfolder.


tvsfpors.ss3

404 Apply

	(A)_School	(B)_Class	(C)_THKSor	(D)_THKSbi	(E)_PreTHK	(F)_CC	(G)_TV	(H)_CC*TV
1	403.00	403101.00	3.00	1.00	2.00	1.00	0.00	0.00
2	403.00	403101.00	4.00	1.00	4.00	1.00	0.00	0.00
3	403.00	403101.00	3.00	1.00	4.00	1.00	0.00	0.00
4	403.00	403101.00	4.00	1.00	3.00	1.00	0.00	0.00
5	403.00	403101.00	4.00	1.00	3.00	1.00	0.00	0.00
6	403.00	403101.00	3.00	1.00	4.00	1.00	0.00	0.00
7	403.00	403101.00	2.00	0.00	2.00	1.00	0.00	0.00
8	403.00	403101.00	4.00	1.00	4.00	1.00	0.00	0.00
9	403.00	403101.00	4.00	1.00	5.00	1.00	0.00	0.00
10	403.00	403101.00	4.00	1.00	3.00	1.00	0.00	0.00

The variables of interest are:

- School indicates the school a student is from (28 schools in total).
- Class identifies the classroom (135 classrooms in total).
- THKSord represents the post-intervention tobacco and health knowledge scaled score, with 4 categories ranging between 1 and 4. The frequency distribution of the post-intervention THKS scores indicated that approximately half the students had scores of 2 or less, and half of 3 or

greater. In terms of quartiles, four ordinal classifications were suggested corresponding to 0 – 1, 2, 3, and 4 – 7 correct responses.

- PreTHKS indicates a student's score prior to intervention, *i.e.* the number correct of 7 items.
- CC is a binary variable indicating whether a social-resistance classroom curriculum was introduced, where 0 indicates "no" and 1 "yes."
- TV is an indicator variable for the use of media (television) intervention, with a "1" indicating the use of media intervention, and "0" the absence thereof.
- CC*TV was constructed by multiplying the variables TV and CC, and represents the CC by TV interaction.

In this chapter we will explore a random intercept model using the ordinal variable THKSord as outcome. In Section 3.3, the post-intervention score was assumed to be a continuous variable. In contrast, here categories are created and the implied data collapse may lead to a loss of information and thus results may differ from those obtained previously.

6.2.1.1 Exploring the data

The focus in this chapter is on the influence of the intervention on the tobacco health knowledge scores of the students, as represented by the ordinal outcome variable THKSord. A cross-tabulation of the variables CC, TV, and THKSord are given in Table 6.1 below.

In general, students not exposed to the social-resistance classroom curriculum (CC = 0) seem to have less health knowledge than those students exposed to the social-resistance classroom curriculum (CC = 1), regardless of their exposure to media intervention. The opposite is true for students from groups assigned the social-resistance classroom curriculum (CC = 1).

Table 6.1: Crosstabulation of CC, TV and THKSord

TV			CC		Total
			0	1	
0	THKSord	1	117	62	179
		2	129	78	207
		3	89	106	195
		4	86	134	220
	Total		421	380	801
1	THKSord	1	110	66	176
		2	105	86	191
		3	91	114	205
		4	110	117	227
	Total		416	383	799

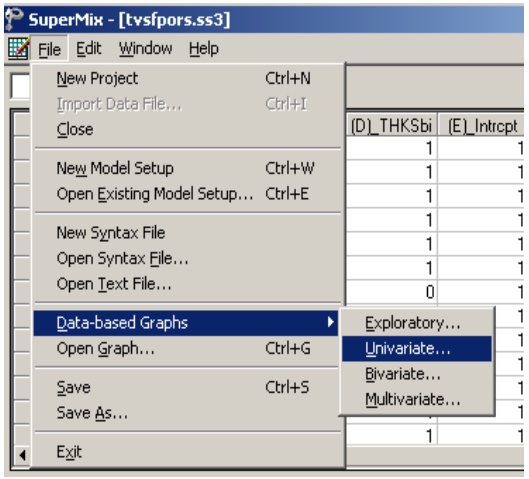
The trend is also apparent when the post-intervention scores are expressed as proportions (see Table 6.2).

First, notice that the outcome variable THKSord has a skewed distribution. By combining the proportions per category over interventions, we find that 0.2219 of the 1600 students had a value of 1 for THKSord, 0.2488 had a value of 2, 0.25 had a value of 3, and 0.2794 a value of 4 for THKSord. The monotonic increase in the proportion observed in each category of THKSord indicates that it would be inappropriate to try to fit a continuous model to the data.

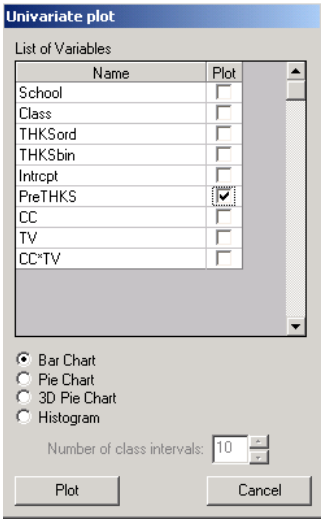
The pre-intervention scores of the students may be used as a covariate in the analysis. To get some idea of the relationship between the scale score PreTHKS and the post-intervention score THKSord, an exploratory graph may be useful. To take a closer look at the distribution of PreTHKS, select the **Data-based Graphs**, **Univariate…** option from the **File** menu after opening the SuperMix spreadsheet **tvsfpors.ss3**.

Table 6.2: Observed proportion of high post–intervention scores

TV			CC		Total
			0	1	
0	THKSord	1	0.0731	0.0388	0.1119
		2	0.0806	0.0488	0.1294
		3	0.0556	0.0663	0.1219
		4	0.0538	0.0838	0.1375
	Total		0.2631	0.2375	0.5006
1	THKSord	1	0.0688	0.0413	0.1100
		2	0.0656	0.0538	0.1194
		3	0.0569	0.0713	0.1281
		4	0.0688	0.0731	0.1419
	Total		0.2600	0.2394	0.4994



The **Univariate plot** dialog box is activated. Select the variable PreTHKS, and request a **Bar Chart**. Click **Plot**.



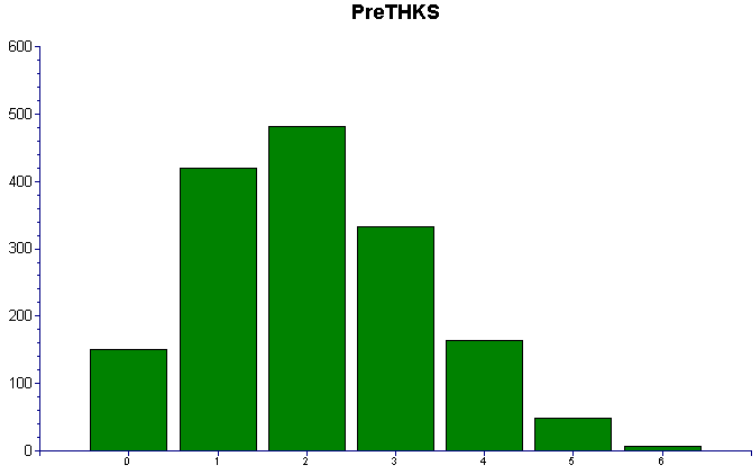
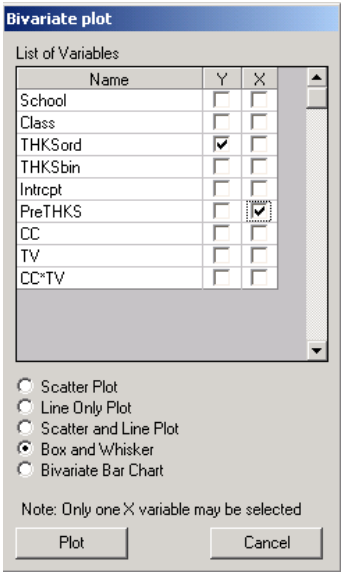


Figure 6.1: Distribution of the PreTHKS scores

Figure 6.1 is obtained. In contrast to the outcome variable THKSord, the distribution of the PreTHKS score has a lower mean, with very few students exhibiting extensive knowledge on the subject matter (PreTHKS = 5 or PreTHKS = 6).

We now take a closer look at the distribution of the outcome variable at each distinct pre-intervention score value by utilizing the **Data-based Graphs**, **Bivariate** option on the **File** menu. By default, a bar chart will be produced. Select the variable THKSord in the **Y** column and the variable PreTHKS in the **X** column, and request a **Box and Whisker** plot before clicking the **Plot** button.



The figure below shows a reasonably steady increase in the mean THKSord with increasing PreTHKS scores. This seems to be expected: students with more initial knowledge ending up having higher post-intervention scores as well. Note that only 55 of the 1600 observations showed a score of 5 or higher on the pre-intervention score, and that no student obtained a post-intervention score of 7 out of 7.

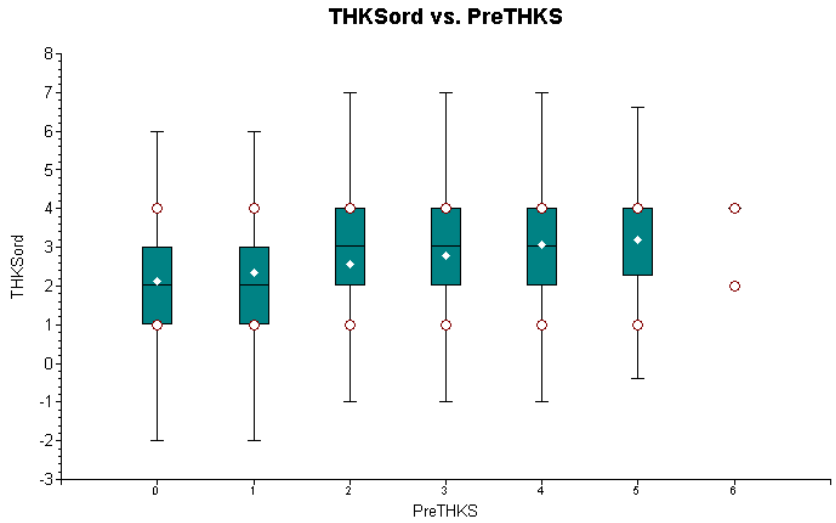


Figure 6.2: Box-and-whisker plot of THKSord for values of PreTHKS

Finally, we also take a look at the mean pre-intervention scores of the students for each of the four subgroups. These are summarized in Table 6.3 below, and show that the mean pre-intervention scores do not differ much.

Table 6.3: Mean pre-intervention scores

Study condition	Mean
CC = 0, TV = 0	2.152
CC = 0, TV = 1	2.087
CC = 1, TV = 0	2.050
CC = 1, TV = 1	1.979

6.2.2 A multilevel ordinal model with logistic link function

6.2.2.1 The proportional odds model

The model we use for the analysis of ordinal data is based on McCullagh's (1980) proportional odds model, which characterizes the ordinal responses in C categories in terms of $C-1$ cumulative category comparisons, specifically $C-1$ cumulative logits. The McCullagh model can be written as

$$\log\left[\frac{P(y \le c)}{1-P(y \le c)}\right] = \gamma_c - \mathbf{x}'\boldsymbol{\beta}$$

where

- $c = 1,...,C-1$ for the C categories of the ordinal outcome
- $\mathbf{x}$ is the vector of explanatory variables, plus the intercept
- γ_c represent the threshold parameter(s); and reflect the cumulative odds when $\mathbf{x} = 0$.

The positive association between a predictor variable x and the ordinal outcome variable y is reflected by β. It is assumed that the effect of x is the same for each of the cumulative odds ratios.

To illustrate, consider a model with a single predictor x.The odds that the response is less than or equal to c (for any fixed c) is divided by e^{β} for every unit change in x, as shown below:

$$\left[\frac{P(y \le c)}{1-P(y \le c)}\right] = \exp\left(\gamma_c - x\beta\right) = \frac{e^{\gamma_c}}{\left(e^{\beta}\right)^x}.$$

On the other hand, the odds that the response is greater than or equal to c (again for a fixed c) is multiplied by e^{β} for every unit change in x:

$$\left[\frac{1-P(y \le c)}{P(y \le c)}\right] = e^{-\gamma_c} \times \left(e^{\beta}\right)^x.$$

It can be illustrated that the ordinal model, when used for a dichotomous variable (coded 0 or 1), is equivalent to the model discussed in Chapter 4. In that model, however, no thresholds were introduced. 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." An example of this is when respondents are asked to rate their agreement with a given statement using the categories "Disagree," "Neutral," "Agree." These three options leave no room for

any other response, though one can argue that these are three possibilities along a continuous scale of agreement that would also make provision for "Strongly Agree" and "Disagree somewhat." The ordinal responses capture in y and the latent continuous variable y^* are linked through some fixed, but unknown, thresholds.

For the dichotomous model, one threshold value is assumed, while for the ordinal model, a series of threshold values $\gamma_0, \gamma_1, \gamma_2, ..., \gamma_C$, where C equals the number of ordered categories, $\gamma_0 = -\infty$, and $\gamma_C = \infty$, is assumed. Here, a response occurs in category c ($Y = c$) if the latent response process y exceeds the threshold value γ_{C-1}, but does not exceed the threshold value γ_c. The cumulative probabilities are given in terms of the cumulative logits with $C-1$ strictly increasing model thresholds $\gamma_1, \gamma_2, ..., \gamma_{C-1}$. In the current case, we will thus have $C-1 = 3$ cumulative probabilities, given in terms of 3 thresholds γ_1, γ_2 and γ_3. The thresholds represent the marginal response probabilities in the C categories. We will illustrate the use of the logistic link function in this example.

To set the location of the latent variable, it is common to set a threshold to zero. Usually, the first of the threshold parameters (γ_1) is set to zero. Alternatively, the model intercept (β_0) is set to zero and $C-1$ thresholds are estimated.

6.2.2.2 The mixed-effect ordinal logistic regression model

A limitation of the model specified in the previous section is that it is assumed that the effect of covariates is the same across the cumulative logits. To overcome this limitation, an extension of the mixed-effects ordinal logistic regression model to allow for nonproportional odds for a set of regressors was developed by Hedeker & Mermelstein (1998). This generalization of the proportional odds model can be formulated as

$$\log\left[\frac{P(y \le c)}{1 - P(y \le c)}\right] = \gamma_c - \left[\mathbf{x}_{ij}'\boldsymbol{\beta} + \mathbf{z}_{ij}'\mathbf{v}_i\right].$$

In this model, as in the proportional odds model, the origin of the latent variable y is set by setting the first threshold, γ_1, equal to zero. It is assumed that $\mathbf{v}_i : NID(\mathbf{0}, \mathbf{\Sigma}_v)$. The unit of measurement is $\sigma = \pi / \sqrt{3}$.

For this model, the category probabilities are defined as

$$P(y_{ij} \le c) = \psi\left(\gamma_c - \left(\mathbf{x}_{ij}'\boldsymbol{\beta} + \mathbf{z}_{ij}'\mathbf{v}_i\right)\right)$$

and

$$P\left(y_{ij} = c\right) = \psi\left(\gamma_c - \left(\mathbf{x}_{ij}'\boldsymbol{\beta} + \mathbf{z}_{ij}'\mathbf{v}_i\right)\right) - \psi\left(\gamma_{c-1} - \left(\mathbf{x}_{ij}'\boldsymbol{\beta} + \mathbf{z}_{ij}'\mathbf{v}_i\right)\right)$$

where the cumulative standard logistic distribution function is

$$\psi\left(\gamma_c - \left(\mathbf{x}_{ij}'\boldsymbol{\beta} + \mathbf{z}_{ij}'\mathbf{v}_i\right)\right) = \frac{1}{1 + \exp\left[-\gamma_c - \left(\mathbf{x}_{ij}'\boldsymbol{\beta} + \mathbf{z}_{ij}'\mathbf{v}_i\right)\right]}.$$

Various link functions may be used with this model. If we define $G^{-1}\left[P\left(y_{ij} \le c\right)\right]$ as

$$G^{-1}\left[P\left(y_{ij} \le c\right)\right] = \gamma_c - \left(\mathbf{x}_{ij}'\boldsymbol{\beta} + \mathbf{z}_{ij}'\mathbf{v}_i\right),$$

or, equivalently,

$$P\left(y_{ij} \le c\right) = G\left[\gamma_c - \left(\mathbf{x}_{ij}'\boldsymbol{\beta} + \mathbf{z}_{ij}'\mathbf{v}_i\right)\right],$$

three types of models can easily be fitted:

- Using $G^{-1}(P) = \log\left[P/(1-P)\right]$ will give a cumulative logit model, *i.e.* a proportional odds model,
- using $G^{-1}(P) = \mathbf{\Phi}^{-1}\left[P/(1-P)\right]$ will produce a cumulative probit model, and

- using $G^{-1}(P) = \log\left[-\log(1-P)\right]$, the so-called complementary log-log link, will give a proportional hazards model.

For more on the use of link functions, please see Section 4.1.1.

6.2.2.3 A general multilevel ordinal model

The multilevel ordinal model is defined in terms of the cumulative probability $P(y_{ij} \le c)$ where c denotes the category of interest. The level-1 model is written in terms of the cumulative logits, as shown below.

Level-1 model:

$$\log\left[\frac{P(y_{ij} \le c)}{1-P(y_{ij} \le c)}\right] = \gamma_c - \left[\mathbf{x}_{ij}'\mathbf{b}_i\right].$$

where $\mathbf{x}_{ij}$ represent the values of the covariates corresponding to level-1 unit j nested within level-2 unit i.

Level-2 model:

If all the elements of the coefficient vector $\mathbf{b}_i$ are allowed to vary randomly across level-2 units, then

$$\mathbf{b}_i = \boldsymbol{\beta} + \mathbf{v}_i,$$

which models the level-2 effects as a function of an overall mean $\boldsymbol{\beta}$ and a unique random component $\mathbf{v}_i$: $NID(\mathbf{0}, \boldsymbol{\Sigma}_v)$. The latter is also referred to as the level-2 residuals and indicates the extent to which a given level-2 unit differs from the average, as estimated by the first part of the level-2 model.

Note that the level-2 model does not depend on the response variable. As the regression coefficients $\beta_0, \beta_1, \beta_2$ and β_3 are without subscript, it is assumed that they do not vary across the categories and hence that the relationship between the predictor variables and the cumulative logits is not dependent on c. McCullagh (1980) referred to this as the assumption of identical odds ratios across the $C-1$ categories.

In practice, a subset of the coefficients $\mathbf{b}_i$ are assumed to have fixed, but unknown, values. For example, a random intercept-and-slope model with 2 predictors of which the first has a random slope would have a level-2 model of the form

$$\begin{aligned} b_{0i} &= \beta_0 + v_{0i} \\ b_{1i} &= \beta_1 + v_{1i} \\ b_{2i} &= \beta_2 \end{aligned}$$

In this model, only the first two coefficients are assumed to vary randomly across the level-2 units.

Another characteristic of the current model is that a positive coefficient for a regressor indicates that the odds that the response is greater than or equal to c increases with an increase in regressor values. However, another formulation as shown below, in which the regression parameters $\boldsymbol{\beta}$ are identical but of opposite sign, is commonly used in survival analysis models (see Chapter 8):

$$\log\left[\frac{P_{ijc}}{1-P_{ijc}}\right] = \gamma_c + \left[\mathbf{x}_i'\mathbf{b}_i\right] \qquad (c = 1,\ldots,C-1).$$

6.2.2.4 An ordinal model with 2 covariates and an interaction term

As in the case of the binary variable THKSbin, we intend to explore the relationship between the type of intervention, the pre-intervention scores of students and the ordinal outcome variable THKSord. We do so using a 2-level model, with students nested within schools.

Level-1 model:

At the first level, the pre-intervention score is used as predictor.

$$\log\left[\frac{P\left(\text{THKSord}_{ij} \le c\right)}{1-P\left(\text{THKSord}_{ij} \le c\right)}\right] = \gamma_c - \left[b_{0i} + b_{1i}\text{PreTHKS}_{ij}\right] \qquad (j = 1,...,n_i \text{ subjects})$$

Level-2 model:

At the school level, the types of intervention (represented by the dummy variables CC and TV) are used to explain differences in the intercepts of the groups. In addition, the interaction between CC and TV is included in the model.

$$b_{0i} = \beta_0 + \beta_2\text{CC}_i + \beta_3\text{TV}_i + \beta_4\left(\text{CC*TV}\right)_i + v_{0i} \quad (i = 1,...,N \text{ groups})$$

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

It is assumed that v_{0i} : $NID(0, \sigma_v^2)$.

The model can also be formulated in a single expression as:

$$\log\left[\frac{P\left(\text{THKSord}_{ij} \le c\right)}{1-P\left(\text{THKSord}_{ij} \le c\right)}\right]$$

$$= \gamma_c - [\beta_0 + \beta_1\text{PreTHKS}_{ij} + \beta_2\text{CC}_i + \beta_3\text{TV}_i + \beta_4\left(\text{CC*TV}\right)_i + v_{0i}]$$

Recall that the outcome variable has 4 categories. There are thus 3 thresholds. In this model

- $0 - \beta_0$ (remember that $\gamma_1 = 0$ for identification purposes) is the first logit (category 1 vs. categories 2 to 4) for groups with no intervention (CC = TV = 0). This logit is adjusted for the effect of PreTHKS.

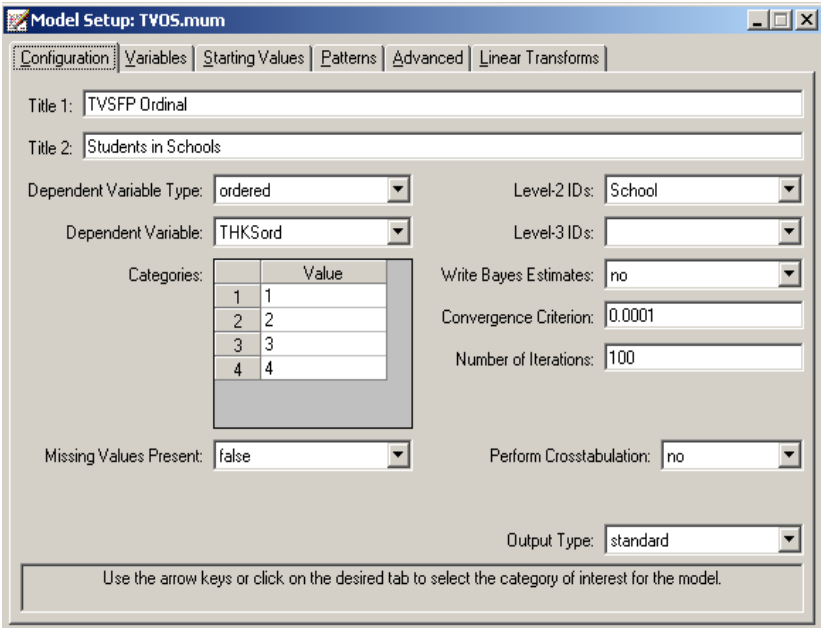
- $\gamma_2 - \beta_0$ is the second logit, representing categories 1 and 2 vs. categories 3 and 4, for groups with no intervention (CC = TV = 0). This logit is also adjusted for the effect of PreTHKS.
- $\gamma_3 - \beta_0$ is the third logit, representing categories 1 to 3 vs. category 4, for the same groups and again adjusted for the effect of PreTHKS.
- The coefficient β_1 represents the effect of PreTHKS on THKSord.
- The coefficient β_2 denotes the PreTHKS adjusted logit differences between CC = 1 and CC = 0 (for TV = 0).
- The coefficient β_3 denotes the PreTHKS adjusted logit differences between TV = yes and TV =no (for CC = 0).
- The coefficient β_4 is the adjusted difference in logit attributable to interaction between CC and TV (CC * TV).
- The random school deviation is represented by v_{0i}. Note that we assume a single, fixed and thus common PreTHKS slope over the level-2 units.
- The interpretation of the coefficients is dependent on the coding of the variables used in the model.

6.2.2.5 Setting up the analysis

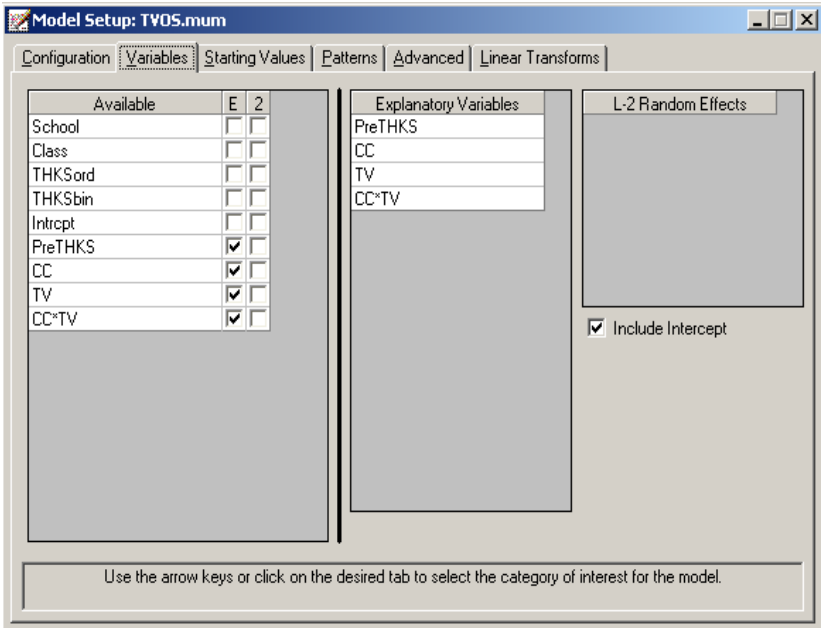
Using the data in **tvsfpors.ss3**, we consider the situation where students are nested within schools and fit a two-level model with the ordinal variable THKSord as outcome. We wish to examine the relationships between the outcome and the two intervention methods employed, simultaneously taking students' pre-intervention scores into account. To do so, we use the model described above with schools as the level-2 units.

Use the **File**, **Open Spreadsheet** option to activate the display of an **Open** dialog box. Browse for the file **tvsfpors.ss3** in the **Examples\Ordinal** folder. Select the file and click the **Open** button to return to the main SuperMix window, where the contents of the SuperMix system file are displayed. We are now ready to provide model specifications.

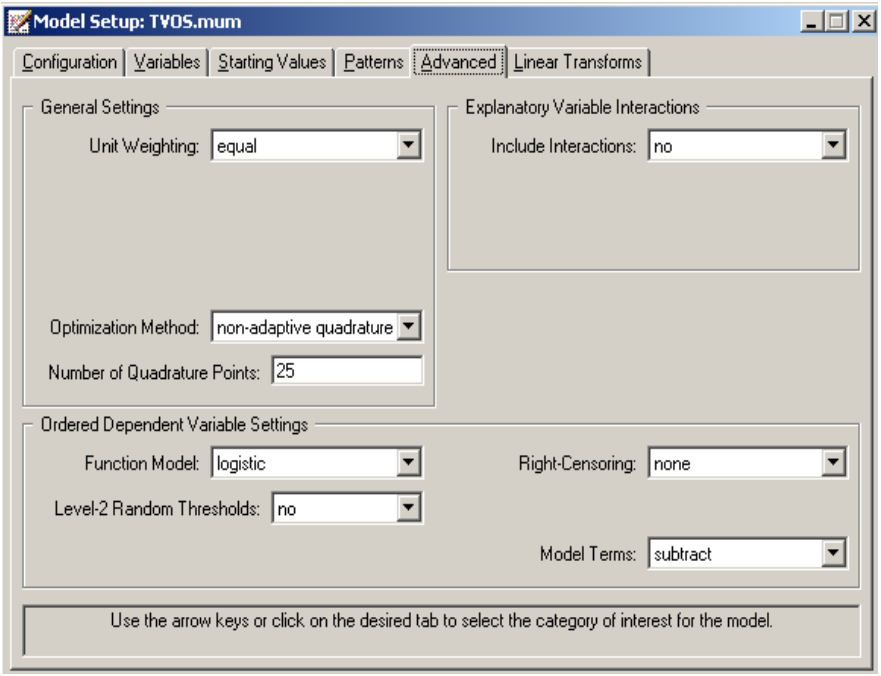
We use the SuperMix interface to provide the model specifications. From the main menu bar, select the **File, New Model Setup** option. The **Configuration** tab of the **Model Setup** dialog box is displayed by default.



Start by selecting the ordinal outcome variable THKSord from the **Dependent Variable** drop-down list box. The type of outcome is specified as ordered using the drop-down list box in the **Dependent Variable Type** field. Once this selection is made, the **Categories** field is displayed. The School identification variable 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. A title for the analysis is entered in the **Title** fields. A convergence criterion of 0.0001 is requested. By default, the maximum number of iterations allowed is 100. Default settings for all other options associated with this tab are used. Proceed to the **Variables** tab by clicking on this tab.



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 drop-down list box next to the first row in the **Explanatory Variables** box. After selecting all the explanatory variables, the random effect(s) at level 2 must be selected. In this case, we wish to allow only the intercept to vary randomly over the schools. By default, the intercept is assumed to vary randomly over higher levels of the hierarchy as indicated by the checked boxes for the **Include Intercept** options.



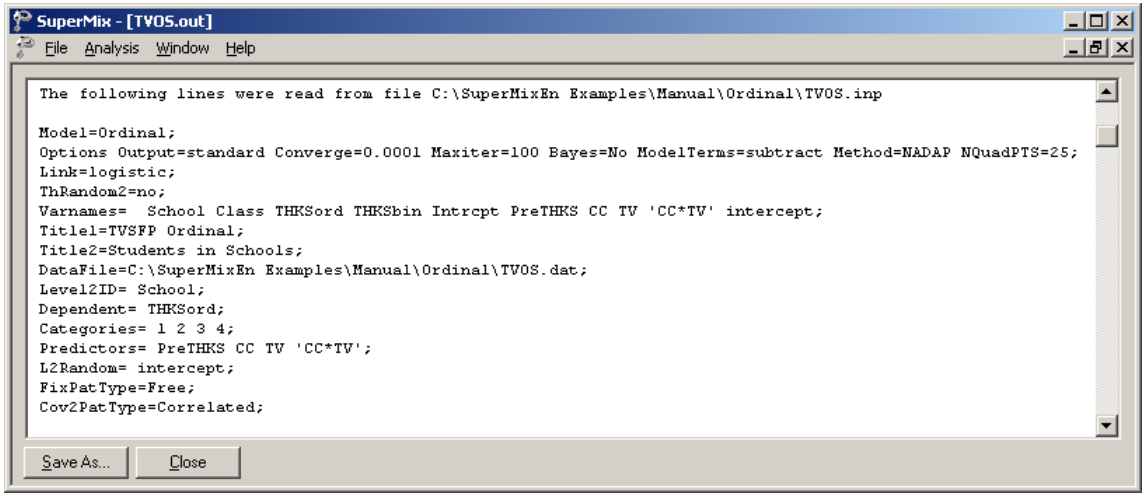
We opt to increase the number of quadrature points to be used during estimation. To do so, select the **Advanced** tab and change the **Number of Quadrature Points** field to 25. We also request the use of a **logistic link** function from the **Function model** drop-down list box.

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

6.2.2.6 Discussion of results

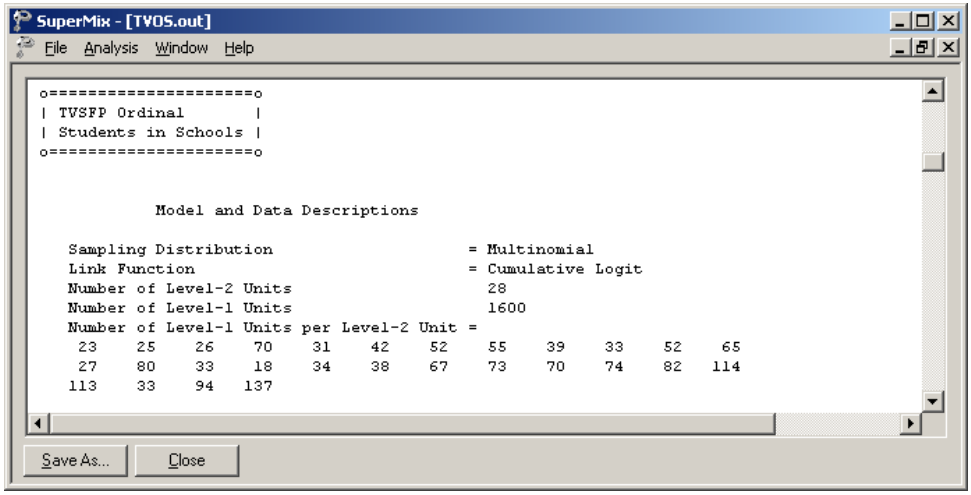
Portions of the output file **TVOS.out** are shown below.

Program information and syntax



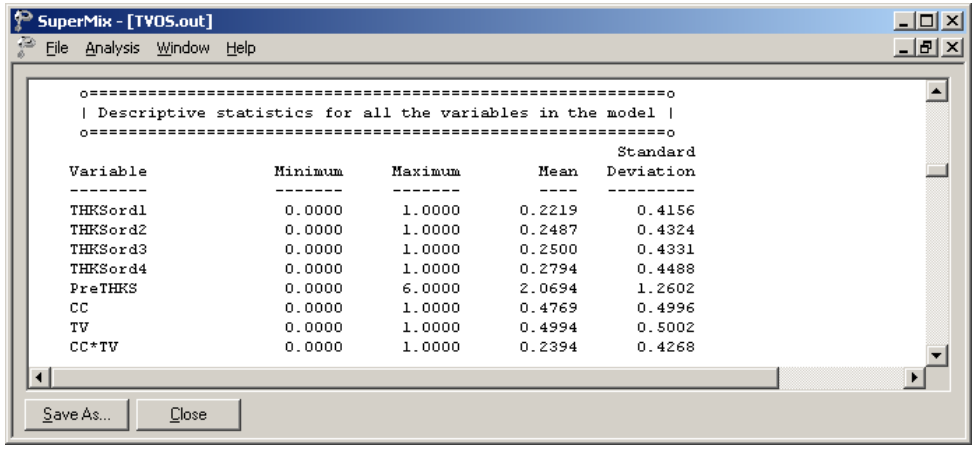
At the top of the file, the syntax saved to the **TVOS.mum** file is shown. The first part states the selection of iteration control options, requests for Bayes residuals, and the specifications necessary to define the model fitted as an ordinal model with logistic link function. The second part of the syntax provides information on the structure of the data, the name and structure of the outcome variable, and the predictors included in the model. Note that this part now also includes information on the categories of the outcome variable and the link function selected.

The next section contains a description of the model specifications. The use of a logistic response function (logit link function), with the assumption of a normal distribution of random effects is indicated. It is also noted that covariate and random effect means are subtracted from the thresholds, implying that a positive coefficient indicates a positive association between the outcome and the predictor in question. To add the covariate and random effect means instead of using the default subtract setting, the **add** option must be selected in the **Model Terms** field on the **Advanced** tab of the **Model Setup** dialog box.



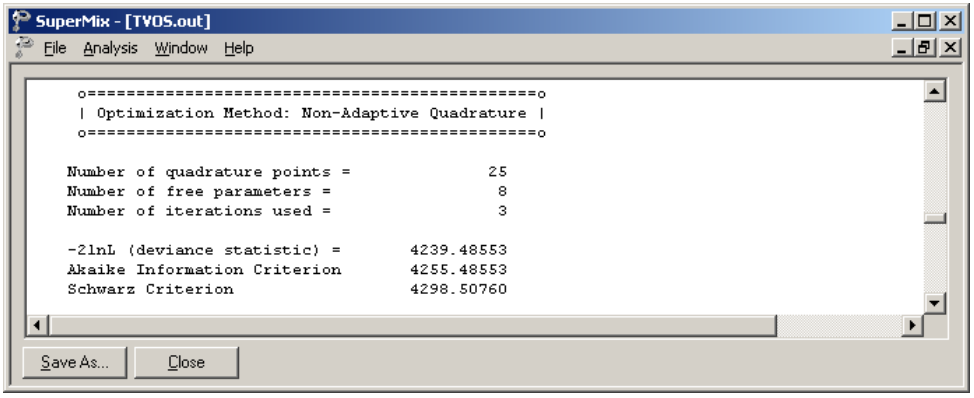
Descriptive statistics

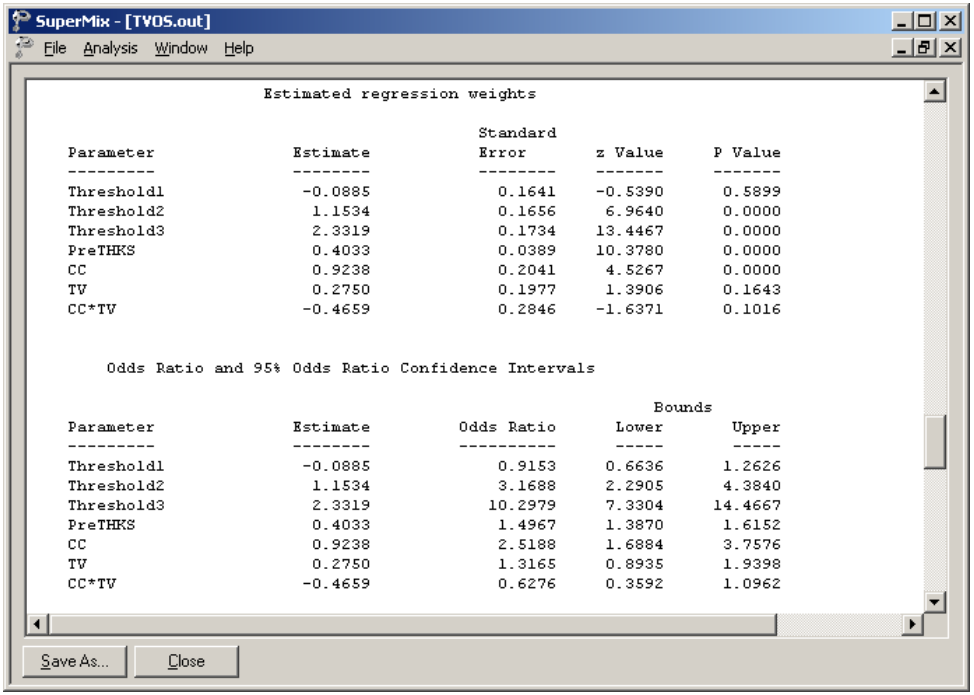
After the observation counts, descriptive statistics for all variables included in the model are followed by a frequency table for the categories of the outcome variable.



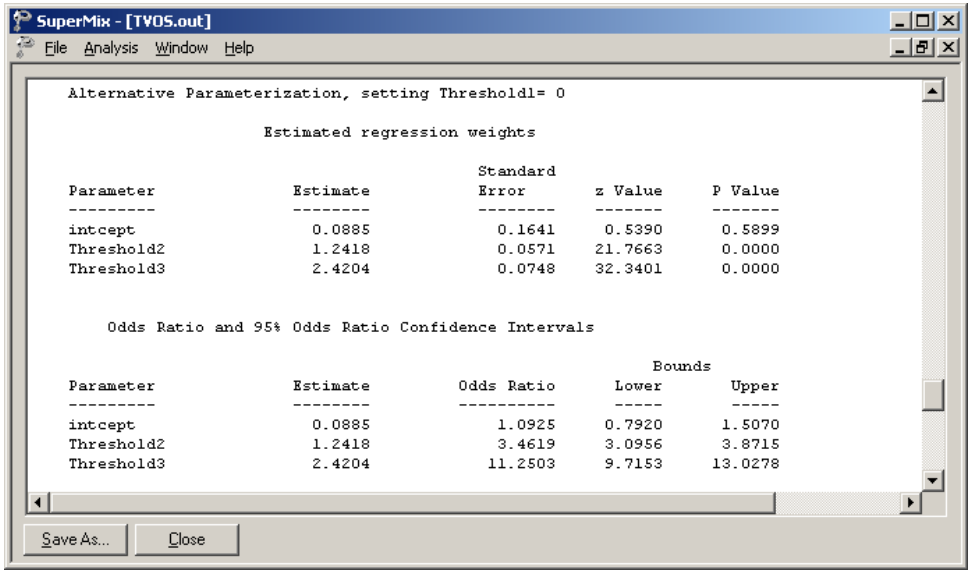
Fixed effects results

The output describing the estimated parameters after convergence is shown next. Two iterations were required to obtain convergence, using 25 quadrature points per dimension. The likelihood function value at convergence as well as the deviance are also given, and may be used to compare a set of nested models. The estimates are shown in the column with heading Estimate, and correspond to the coefficients $\beta_0, \beta_1, \ldots, \beta_4$ in the model specification. Significant effects of PreTHKS and CC are observed. With the exception of the CC *TV interaction term, positive relationships between the predictors and the ordinal outcome variable are indicated by these results. We also note that the coefficient associated with the curriculum-based intervention (CC) is almost three times the size of the estimated coefficient for media intervention (TV).



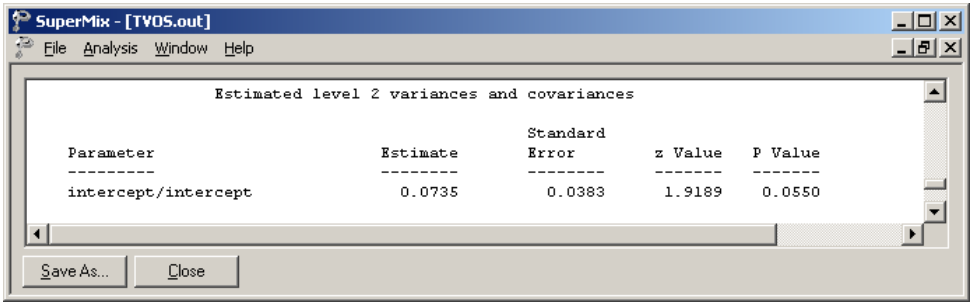


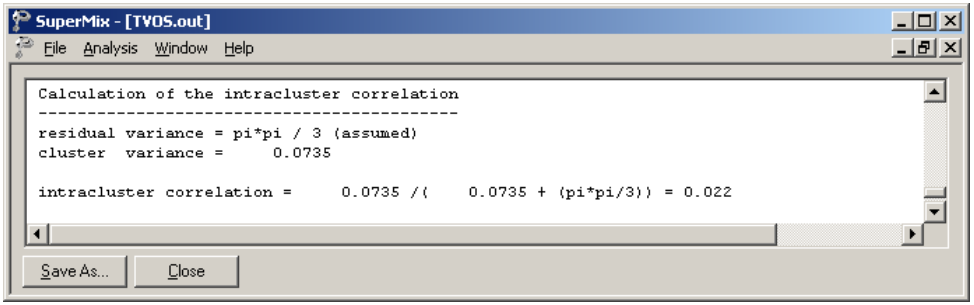
The alternative parameterization, setting threshold = 0 is shown next. The estimates of γ_2 and γ_3 are 1.242 and 2.420 respectively – recall that for identification purposes γ_1 was set to zero.



Random effects results

The last part of the output file contains information on the random effects and calculation of the intracluster correlation coefficient. The variation in intercept over schools is estimated at 0.0735, with the associated *p*-value of 0.055 indicating its statistical significance.





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. If the model is true, it is assumed that the level-1 error variance, σ_e^2, is equal to $\pi^2/3$ for the logistic link function, where π represents the constant 3.141592654 (see, *e.g.*, Hedeker & Gibbons (2006), p. 157).

Finally, the calculation of the intracluster correlation is shown. In this calculation it is assumed that the residual variation, σ_e^2, is equal to $\pi^2/3$. The value of 0.022 indicates that almost all variation is attributable to students, rather than to the schools.

6.2.2.7 Interpreting the output

Model-based graphs

Activate the **Model Setup** window by clicking on it. Using the **Plot Equations for Outcome Variable** dialog box that appears when the **File, Model-based Graphs, Equations** option is selected, we can graphically depict the trend in post-intervention scores as a function of pre-intervention scores, taking the type of intervention into account. The dialog box below shows the selection of the predictor PreTHKS. Grouping of plots by the categories of CC is requested, while marking of the plots by TV is indicated by the selection in the **Mark** column. Two graphs will thus be displayed on the same set of axes: one for each value of the indicator variable TV.

By default, all variables present in the model, but not selected for inclusion in the graph, will be assumed to have a value of 0. In the current situation, this means that CC*TV is kept constant at zero. In effect, the graphs are for students from schools where only one of the interventions was administered; students from schools where both were implemented would have a value of 1 on the variable CC*TV.

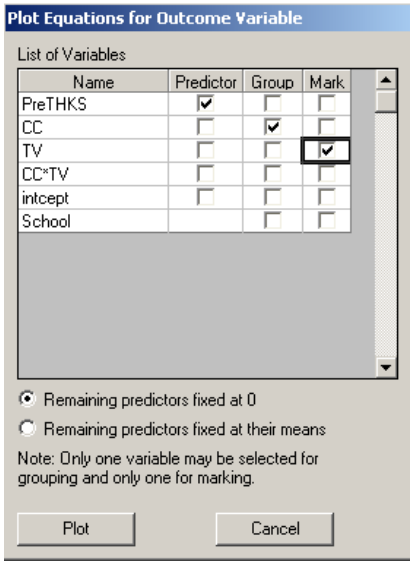


Figure 6.3 shows slightly modified versions of the graphs obtained when the **Plot** button is clicked. For publication purposes, the line type associated with the value TV = 1 was changed to a dotted line. This was accomplished by clicking on the top line to activate the **Plot Parameters** dialog box and changing the line parameters so that the color is green and the style is dotted rather than solid. The plots show that the curriculum-based intervention had a larger effect on the post-intervention score: the intercept in the case where CC = 1 is appreciably higher than when CC = 0. In both graphs, the solid line indicates the absence of media-based intervention. The use of media-based intervention seems to have had a positive, albeit small, effect on the outcome.

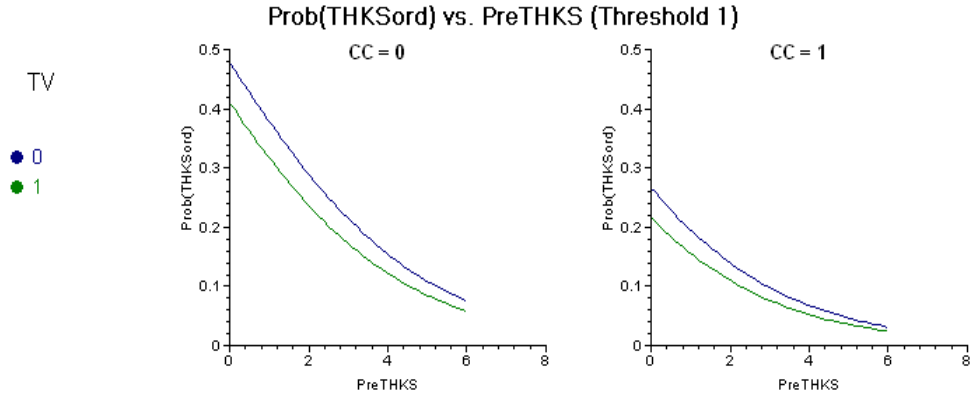


Figure 6.3: Model-based graphs of THKSord by PreTHKS for groups

Interpretation of fixed effect estimates

The outcome variable has four categories, and there are thus 3 thresholds. The coefficient β_1, representing the effect of PreTHKS on THKSord, is estimated as 0.4033. The PreTHKS adjusted logit differences between CC = 1 and CC = 0 (keeping TV = 0) is estimated as $\hat{\beta}_2 = 0.9238$, in contrast with the PreTHKS adjusted logit differences between TV = yes and TV = no (keeping CC = 0) which is estimated as $\hat{\beta}_3 = 0.2750$. The coefficient β_4 denotes the adjusted difference in logit attributable to the interaction between CC and TV (CC * TV) and is estimated at -0.4659, which diminishes the combined effects of CC and TV.

Logits for groups with no intervention

The first logit for groups with no intervention, for category 1 vs. categories 2 to 4, is $\hat{\gamma}_1 = -0.0885$. The second logit for the same group, for categories 1 and 2 vs. 3 and 4, can be calculated as $\hat{\gamma}_2 = 1.1534$. The third and final logit for this group, for

categories 1 to 3 vs. 4, is $\hat{\gamma}_3 = 2.3319$. All of the logits are adjusted for the effect of the pre-intervention score PreTHKS.

Logits for groups with classroom curriculum intervention (CC = 1)

Turning to the groups with classroom curriculum intervention (CC = 1), logits can be obtained in similar fashion: $\hat{\gamma}_1 - \hat{\beta}_2 = -0.0885 - 0.9238 = -1.0123$ $\hat{\gamma}_2 - \hat{\beta}_2 = 1.1534 - 0.9238 = 0.2296$, and $\hat{\gamma}_3 - \hat{\beta}_2 = 1.4081$.

Logits for groups with media intervention (TV = 1)

For the groups where media intervention was employed, the logits are:
$\hat{\gamma}_1 - \hat{\beta}_3 = -0.3635$, $\hat{\gamma}_2 - \hat{\beta}_3 = 0.8784$, and $\hat{\gamma}_3 - \hat{\beta}_3 = 2.0569$.

Estimated outcomes for groups: unit-specific results

To evaluate the expected effect of the CC, TV, CC*TV, and PreTHKS variables we use the expression below:

$$\log\left[\frac{\widehat{P}\left(\text{THKSord}_{ij} \le c\right)}{1-\widehat{P}\left(\text{THKSord}_{ij} \le c\right)}\right] = \hat{\gamma}_c - \left[\hat{\beta}_2 \text{CC}_i + \hat{\beta}_3 \text{TV}_i + \hat{\beta}_4 (\text{CC} * \text{TV})_i + \hat{\beta}_1 \text{PreTHKS}_{ij}\right]$$

The variable PreTHKS is treated as a continuous variable in this example, although it too is originally a scale score. In order to facilitate comparison of treatment groups, the mean PreTHKS score for groups can be used to obtain the logits. The mean PreTHKS scores for each of the four treatment groups were given in Table 6.3. This table is reproduced below.

Table 6.4: Mean pre-intervention scores

Study condition	Proportion
CC = 0, TV = 0	2.152
CC = 0, TV = 1	2.087
CC = 1, TV = 0	2.050
CC = 1, TV = 1	1.979

The probabilities for the responses of typical subjects from the group with no intervention (TV = CC = 0) can be obtained using the modified equation

$$\log\left[\frac{\widehat{P}\left(\text{THKSord}_{ij} \le c\right)}{1-\widehat{P}\left(\text{THKSord}_{ij} \le c\right)}\right] = \hat{\gamma}_c - \left[0.4033(2.1520)\right]$$
$$= \hat{\gamma}_c - 0.8679.$$

Let

$$\hat{\eta}_{ijc} = \log\left[\frac{\widehat{P}\left(\text{THKSor}_{ij} \le c\right)}{1-\widehat{P}\left(\text{THKSor}_{ij} \le c\right)}\right].$$

Similar equations for the groups with classroom curriculum intervention and media intervention respectively are then

$$\hat{\eta}_{ijc} = \hat{\gamma}_c - \left[+0.9238 + 0.4033(2.050)\right]$$
$$= \hat{\gamma}_c - 1.7506$$

and

$$\hat{\eta}_{ijc} = \hat{\gamma}_c - \left[0.2750 + 0.4033(2.087)\right]$$
$$= \hat{\gamma}_c - 1.1167.$$

When both intervention methods were employed and thus TV = CC = CC*TV = 1, we have

$$\hat{\eta}_{ijc} = \hat{\gamma}_c - \left[0.9238 + 0.2750 - 0.4659 + 0.4033(1.979)\right]$$
$$= \hat{\gamma}_c - 1.5310.$$

In this example, the logistic link function was specified, and we can rewrite any formula of the form $\hat{\eta}_{ijc} = \gamma_c - a$ in the alternative form

$$\hat{P}\left(\text{THKSord}_{ij} \le c\right) = \frac{e^{\gamma_c - a}}{1 + e^{\gamma_c - a}} = \frac{e^{\eta_{ijc}}}{1 + e^{\eta_{ijc}}}.$$

Table 6.5 contains the cumulative response probabilities obtained through substitution in the above formulae for the first three categories of the ordinal outcome THKSord.

Table 6.5: Cumulative response probabilities for various groups and categories

Category	CC	TV	$\hat{\eta}_{ijc} = \gamma_c - \mathbf{x}'\boldsymbol{\beta}$	Probability of response
1	0	0	–0.9564	0.2776
1	1	0	–1.8390	0.1372
1	0	1	–1.2052	0.2306
1	1	1	–1.6195	0.1653
1 or 2	0	0	0.2855	0.5709
1 or 2	1	0	–0.5972	0.3550
1 or 2	0	1	0.0367	0.5092
1 or 2	1	1	–0.3776	0.4067
1, 2 or 3	0	0	1.4640	0.8121
1, 2 or 3	1	0	0.5813	0.6414
1, 2 or 3	0	1	1.2152	0.7712
1, 2 or 3	1	1	0.8009	0.6902

The probabilities reported in Table 6.5 are cumulative: for example, the probability of a response in either category 1 or 2 for the group with CC = TV = 0 is equal to 0.5709. The probability of a response in category 1 is 0.2776, and therefore the probability of a response in category 2 is 0.5709 – 0.2776 = 0.2933. Similarly, the

estimated response probability of a category 3 response for a respondent from the same group is 0.8121 – 0.5709 = 0.2412. To obtain the category 4 response for a respondent from the first group, the value of the estimated response in categories 1, 2, or 3 has to be subtracted from 1, so that the probability of a response in category 4 for a typical respondent with CC = TV = 0 is 1 – 0.8121 = 0.1879. The cumulative probabilities for the first 3 categories of the ordinal outcome are plotted in Figure 6.4.

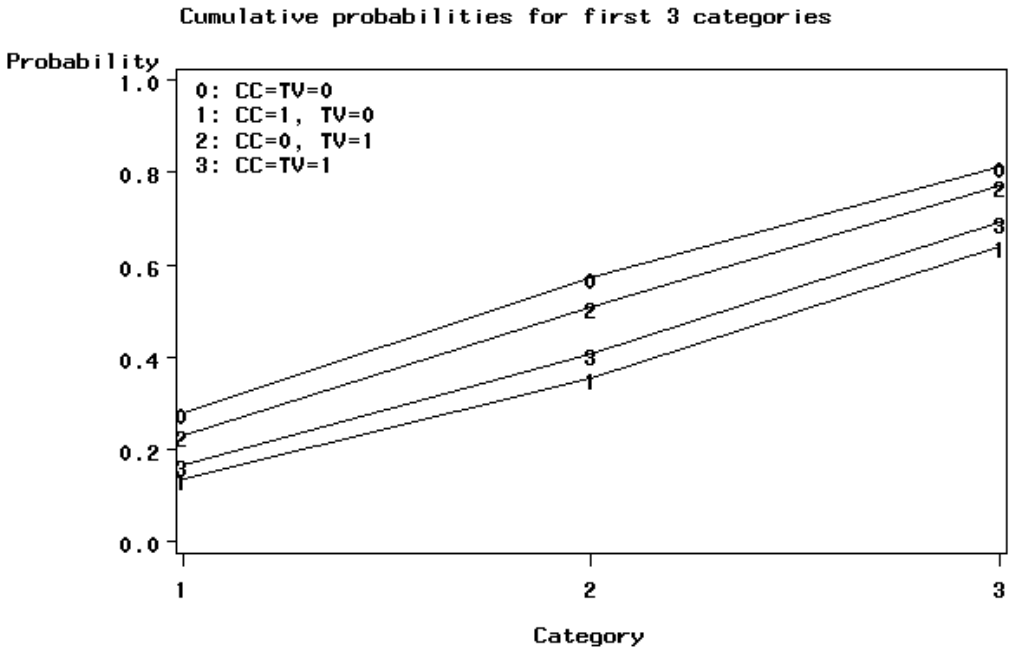


Figure 6.4: Cumulative response probabilities for categories 1 to 3 of THKSord

The graph shows two groupings: one representing CC = 0, regardless of the value of TV; and the other CC = 1, again regardless of the value of TV. The smallest probability to fall in categories other than category 1 (normal) is for the combination CC = TV = 1. The fact that the plotted cumulative probability lines for CC = 1 and TV = 1 are close to the line for CC = TV = 1 suggests that the implementation of media intervention (TV = 1 if implemented) has less impact on the outcome than the use of a classroom curriculum (CC = 1 if implemented).

To obtain category probabilities, differences between the cumulative probabilities obtained above are calculated. In other words,

$$\hat{P}\left(\text{THKSord}_{ij} = c\right) = \hat{P}\left(\text{THKSord}_{ij} \le c\right) - \hat{P}\left(\text{THKSord}_{ij} \le c - 1\right)$$

The category probabilities are reported in Table 6.6 and are graphically displayed in Figure 6.5.

A typical respondent from the control group (no intervention) was less likely to respond in categories 3 or 4 of the ordinal post-intervention outcome variable. For both this group and the group which was assigned to media intervention only, the most likely response was in category 2 and the least likely response in category 4. In contrast, groups that were subjected to the classroom curriculum intervention, with or without media intervention, were most likely to display a high level of knowledge (*i.e.*, a response in categories 3 or 4), and least likely to respond in the first category. From this graph we conclude that the classroom curriculum intervention was key – groups subjected to the intervention tended to increase in knowledge over the study period.

Table 6.6: Estimated unit-specific probabilities for THKSord categories

Category	CC	TV	$\hat{P}\left(\text{THKSord}_{ij} = c\right)$
1	0	0	0.2776
	1	0	0.1372
	0	1	0.2306
	1	1	0.1653
2	0	0	0.2933
	1	0	0.2178
	0	1	0.2786
	1	1	0.2414
3	0	0	0.2413
	1	0	0.2864
	0	1	0.2620
	1	1	0.2835

4	0	0	0.1879
	1	0	0.3586
	0	1	0.2287
	1	1	0.3098

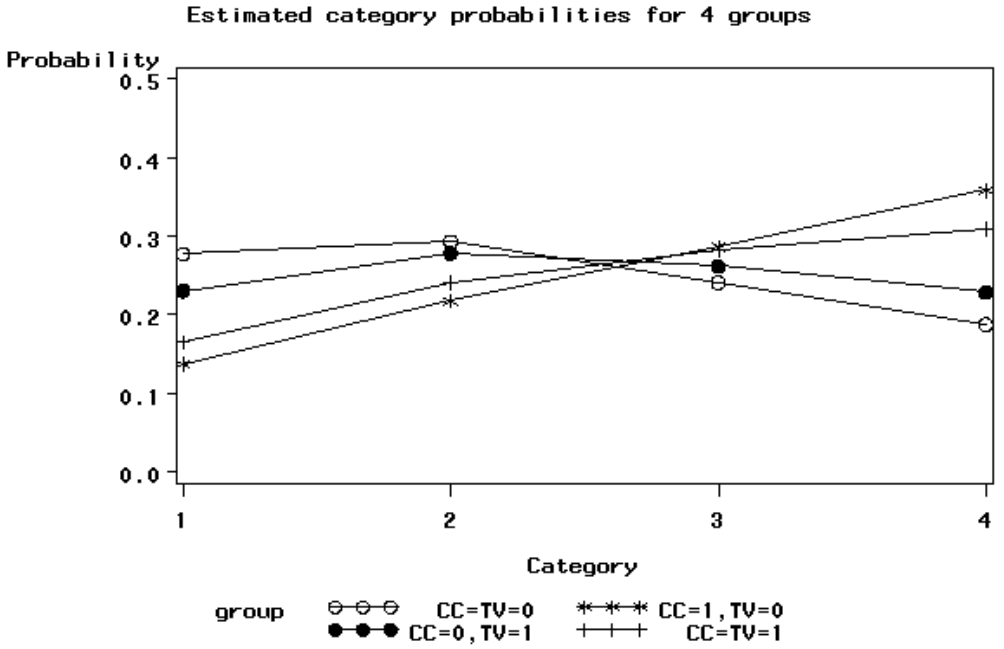


Figure 6.5: Estimated category probabilities for THKSord

Estimated outcomes for different groups: population-average results

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

$$y_{ij} = \mathbf{z}'_{(1)ij}\mathbf{b}_i + \mathbf{x}'_{(1)ij}\boldsymbol{\beta}_{(1)} + e_{ij},$$

making the assumption that $e_{ij} : i.i.d.(0, \sigma^2)$. For a probit link function $\sigma_e^2 = 1$, and for a logistic link function it is assumed that $\sigma_e^2 = \pi^2/3$, as indicated in the final

lines of the output file. Under the assumption that $\mathbf{v}_i$ and e_{ij} are independently distributed, it follows that

$$\sigma^2_{y_{ij}} = \mathbf{z}'_{ij}\boldsymbol{\Phi}_{v_i}\mathbf{z}_{ij} + \sigma^2_e.$$

Let

$$d_{ij} = \frac{\sigma^2_{y_{ij}}}{\sigma^2_e}.$$

The quantity d_{ij} 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 replacing $\hat{\eta}_{ijc}$ with $\hat{\eta}^*_{ijc} = \hat{\eta}_{ijc}/\sqrt{d_{ij}}$.

From the output, we have $\text{var}(u_{i0}) = 0.074$, where u_{i0} denotes the random intercept coefficient. In this case, $\mathbf{z}'_{ij} = \mathbf{1}$ and hence, with $\sigma^2_e = \pi^2/3$ for the logistic link,

$$\sigma^2_{y_{ij}} = 1 \times 0.074 + (3.1416)^2/3 = 3.3639.$$

Therefore

$$d_{ij} = \frac{3.3639}{3.2899} = 1.0225.$$

To obtain the population-average probability estimates, we now replace the η_{ijc} values calculated for the unit-specific case with $\hat{\eta}^*_{ijc} = \hat{\eta}_{ijc}/\sqrt{d_{ij}}$.

We can compare these estimated population-average probabilities with the observed data for the four groups formed by the categories of TV and CC as shown in Table 6.5 previously. Table 6.7 shows the estimated population-average probabilities.

A comparison of these probabilities with those reported in Table 6.5 shows little difference between the unit-specific and population-average category probabilities for treatment groups. The population-average category probabilities for the first two categories are slightly smaller than the corresponding unit-specific probabilities, while those for category 3 are slightly larger. The extent of differences between unit-specific and population-average results are highly dependent on the "scaling" induced by dividing the $\eta_{ijc}s$ by $\sqrt{d_{ij}}$. In the current example, $\sqrt{d_{ij}} = 1.0112$ and thus no large differences could be expected. To obtain category probabilities, differences between the cumulative probabilities may be calculated, as illustrated in the case of the unit-specific results.

Table 6.7: Cumulative response probabilities for various groups and categories

Category	CC	TV	$\hat{\eta}^{*}_{ijc} = \hat{\eta}_{ijc} / \sqrt{d_{ij}}$	Probability of response
1	0	0	–0.9564/1.0112	0.2797
1	1	0	–1.8391/1.0112	0.1396
1	0	1	–1.2052/1.0112	0.2330
1	1	1	–1.6195/1.0112	0.1678
1 or 2	0	0	0.2855/1.0112	0.5701
1 or 2	1	0	–0.5972/1.0112	0.3565
1 or 2	0	1	0.0367/1.0112	0.5092
1 or 2	1	1	–0.3776/1.0112	0.4077
1, 2 or 3	0	0	1.4640/1.0112	0.8097
1, 2 or 3	1	0	0.5813/1.0112	0.6399
1, 2 or 3	0	1	1.2152/1.0112	0.7689
1, 2 or 3	1	1	0.8009/1.0112	0.6883

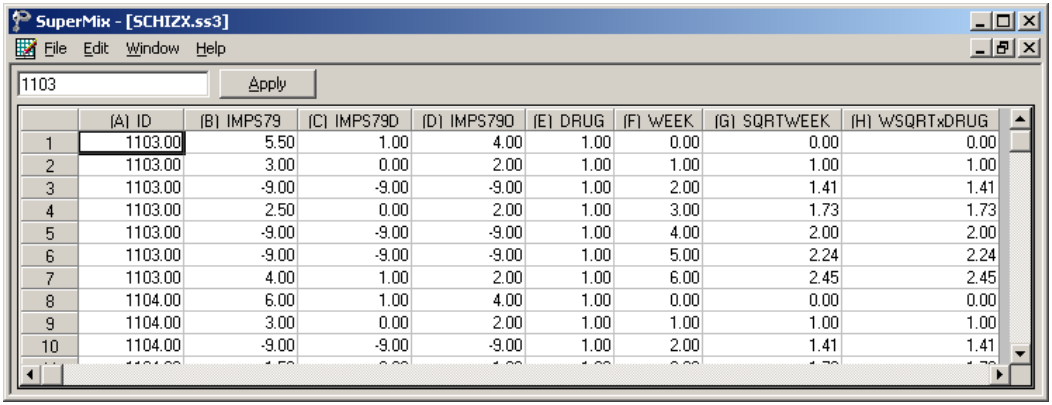
6.3 Two-level ordinal analysis of NIMH data

6.3.1 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	0.00
2	1103.00	3.00	0.00	2.00	1.00	1.00	1.00	1.00
3	1103.00	-9.00	-9.00	-9.00	1.00	2.00	1.41	1.41
4	1103.00	2.50	0.00	2.00	1.00	3.00	1.73	1.73
5	1103.00	-9.00	-9.00	-9.00	1.00	4.00	2.00	2.00
6	1103.00	-9.00	-9.00	-9.00	1.00	5.00	2.24	2.24
7	1103.00	4.00	1.00	2.00	1.00	6.00	2.45	2.45
8	1104.00	6.00	1.00	4.00	1.00	0.00	0.00	0.00
9	1104.00	3.00	0.00	2.00	1.00	1.00	1.00	1.00
10	1104.00	-9.00	-9.00	-9.00	1.00	2.00	1.41	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

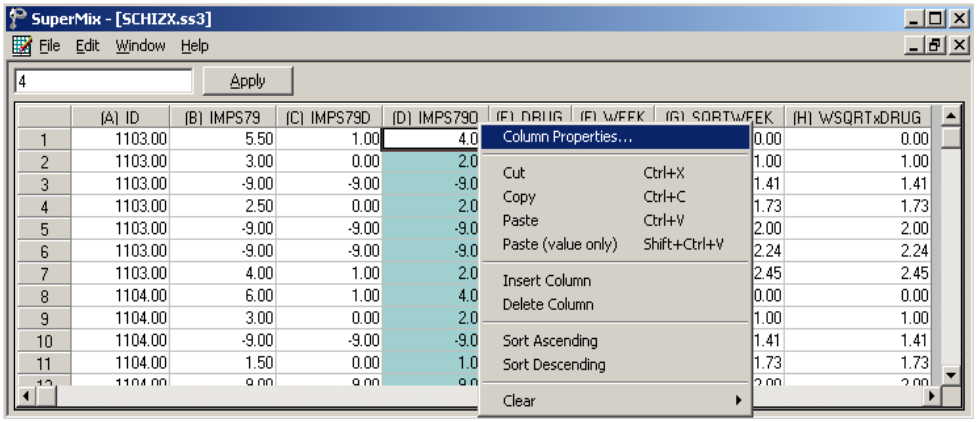
normal, borderline, mildly, or moderately mentally ill, while the value "1" was assigned to measurements corresponding to "markedly ill" through "most extremely ill."

- 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. For more information on data manipulation, please refer to Section 2.5.
- WSQRTxDRUG is the product of the treatment group and the square root of WEEK.

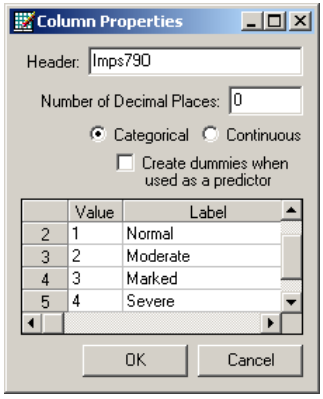
In this data file, each subject's data consist of seven lines, these being the repeated measurements on seven occasions. Notice that there are missing value codes (–9) for some subjects at specific time points. The data from these time points will not be used in the analysis, but data from these subjects 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.

6.3.1.1 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.



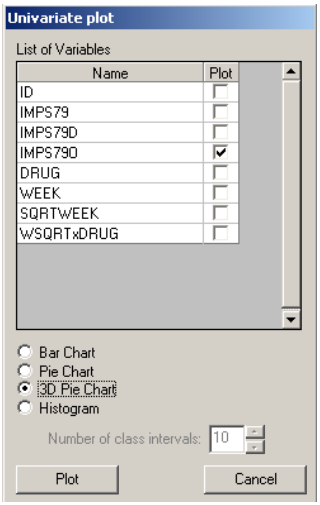
Click on the **OK** button and save the change to the data set by clicking on the **File**, **Save** option.

The outcome variable: univariate graphs

As a first step, we take a look at the ordinal variable IMPS79O which is the potential dependent variable in this study.

Pie chart

To generate a pie chart for IMPS79O, 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 IMPS79O and indicate that a **3D Pie Chart** is to be graphed as shown below.



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

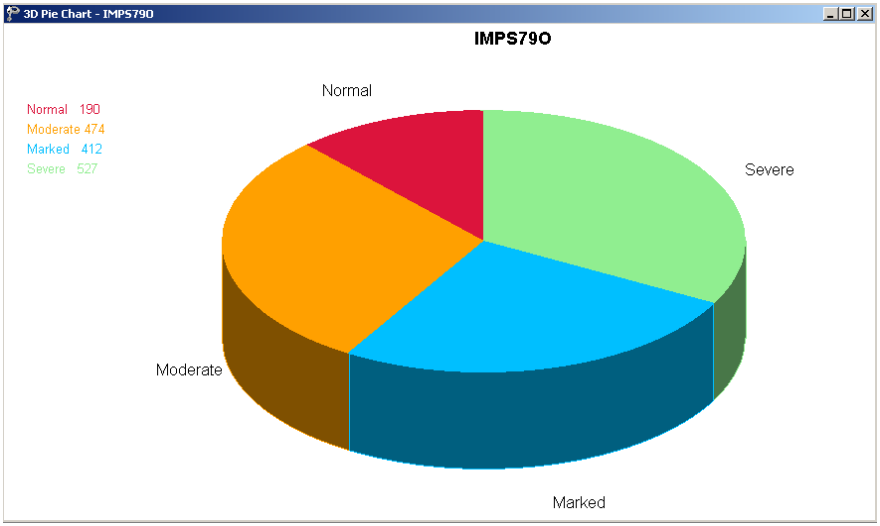


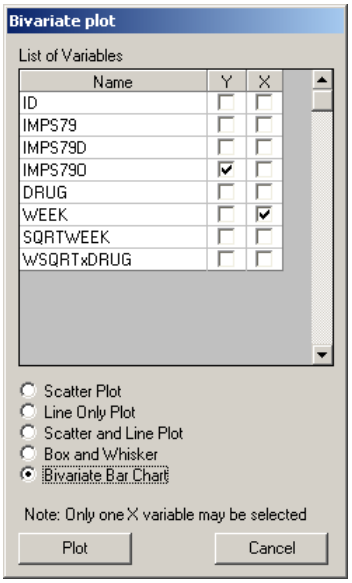
Figure 6.6: Pie chart of IMPS79O values

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 (IMPS79O) will decrease over the treatment period. Before considering fitting a model to these data, we would like to explore the relationship between IMPS79O 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**.

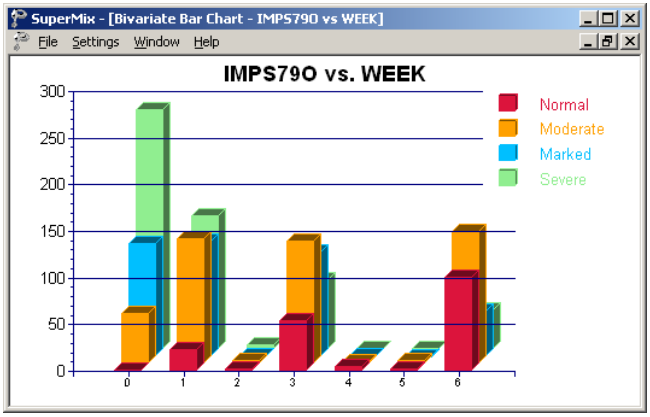


Figure 6.7: Bar chart of IMPS79O vs. WEEK

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.

6.3.2 An ordinal regression model with random intercept

6.3.2.1 Introduction

As discussed in Section 6.1, 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, due to the reasons discussed in Section 4.1. 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. Detailed information on these link functions are given in section 4.1.1.

6.3.2.2 The model

Let the outcome variable be coded into c categories, where $c = 1, 2, ..., C$. In this example, the ordinal variable IMPS79O 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 $P_{ijc} = \Pr\left(Y_{ij} \le c\right) = \sum_{k=1}^{c} p_{ijk}$, 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 IMPS79O as outcome can then be written as

Level 1 model:

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

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

Level 2 model:

$$b_{0i} = \beta_0 + v_{0i}, \quad i = 1,\ldots,N$$
$$b_{1i} = \beta_1$$
$$b_{2i} = \beta_2$$
$$b_{3i} = \beta_3$$

The cumulative probability can be expressed by

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

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 IMPS79O 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.

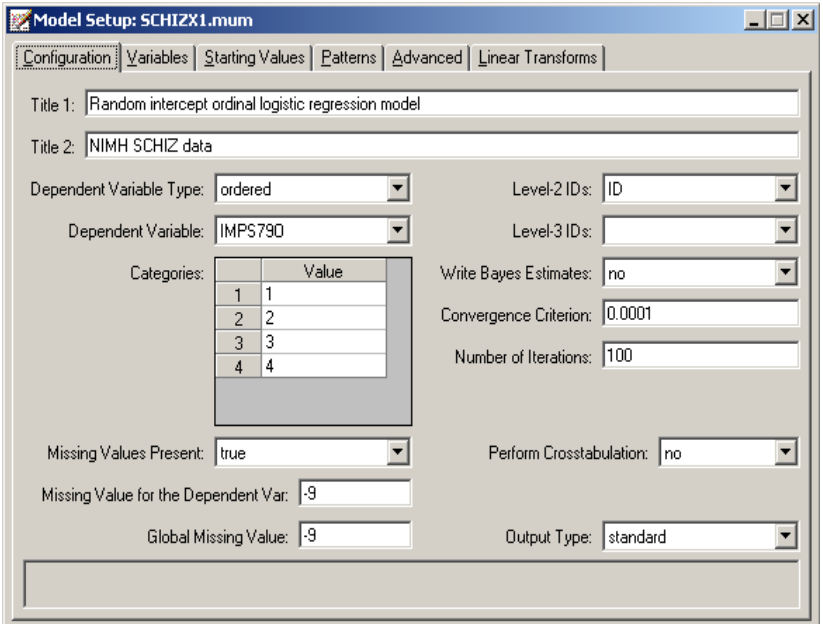
The connection between an ordinal outcome variable y with C categories and an underlying continuous variable y^* is

$$y = c \leftrightarrow \gamma_{j-1} \le y^* \le \gamma_j,\ \ c = 1,2,...,C$$

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.

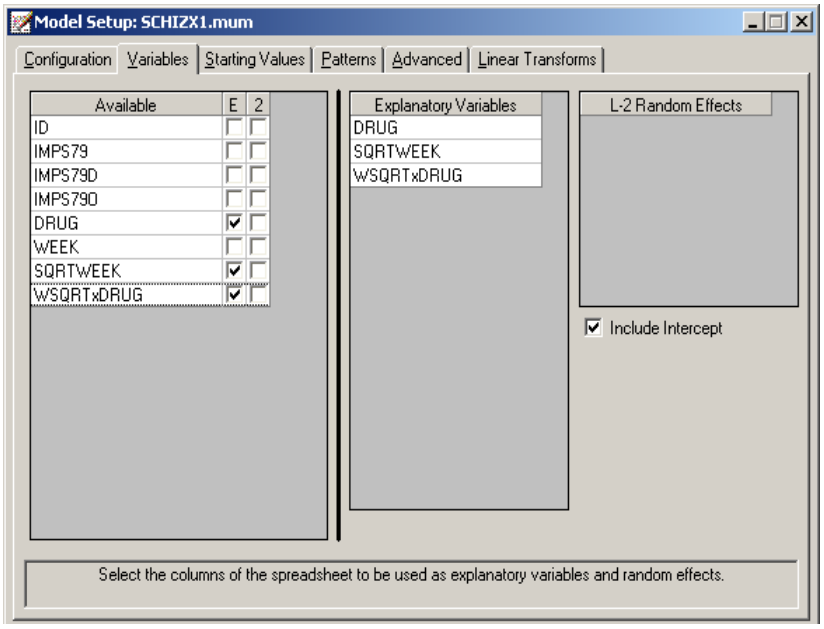
6.3.2.3 Setting up the analysis

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



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 IMPS79O 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.

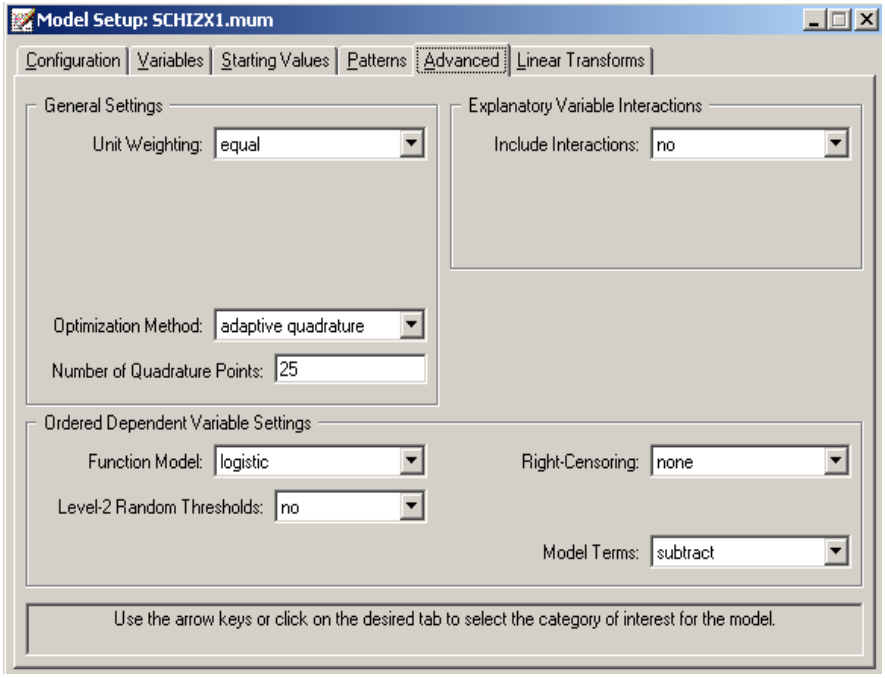
We notice that the missing value –9 is also included as a category. The **Missing Values Present** drop-down list box is used to specify the values of missing data for both outcome and predictors. As a first step, set the value of the **Missing Values Present** drop-down list box to **True**. The appearance 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 below. Use 25 quadrature points.

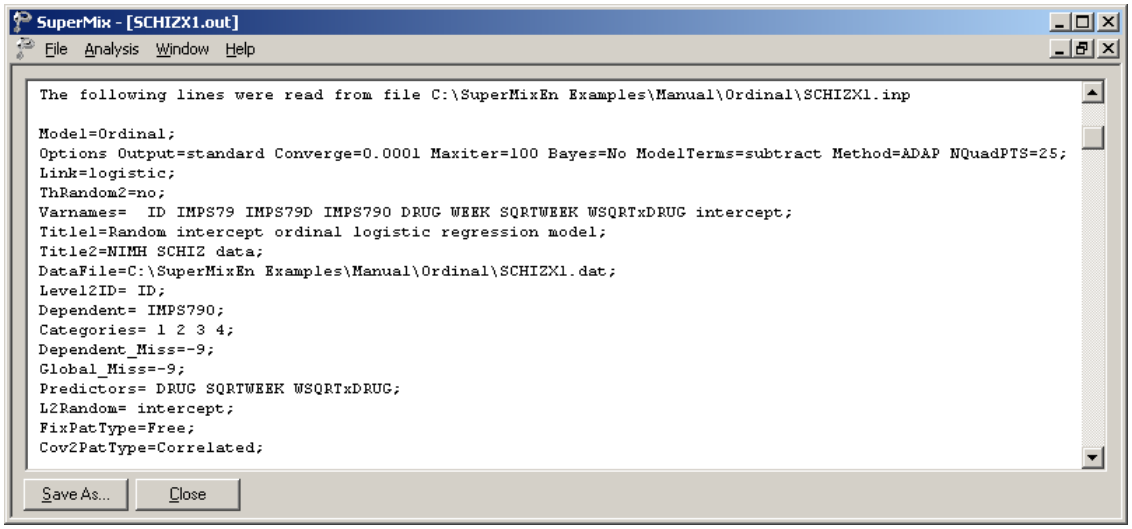


Before running the analysis, the model specifications have to be saved. Select the **File**, **Save As** option, and provide a name (**SCHIZX1.mum**) for the model specification file. Run the analysis by selecting the **Run** option from the **Analysis** menu.

6.3.2.4 Discussion of results

Syntax

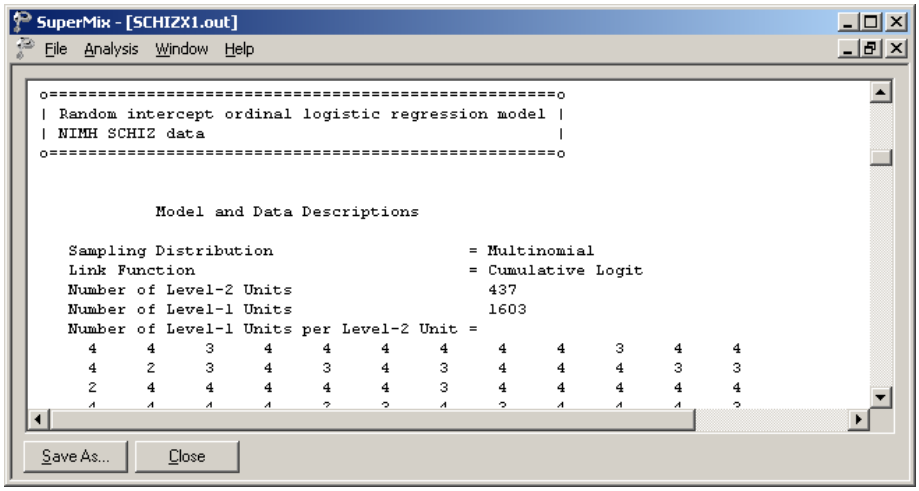
The syntax corresponding to the model setup is given in the **model specifications**. These lines of SuperMix syntax are saved as a ***.inp** file with the same name as the model setup file (***.mum**). At the top of the output file, the syntax lines are printed as shown below.



The 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.

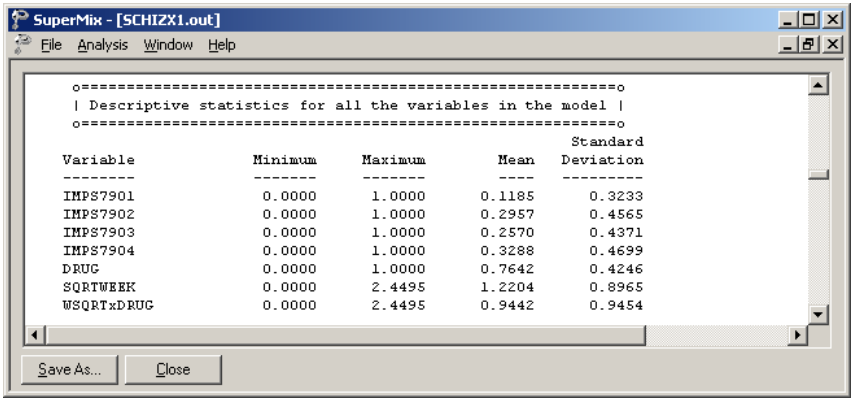
Model and data description

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.

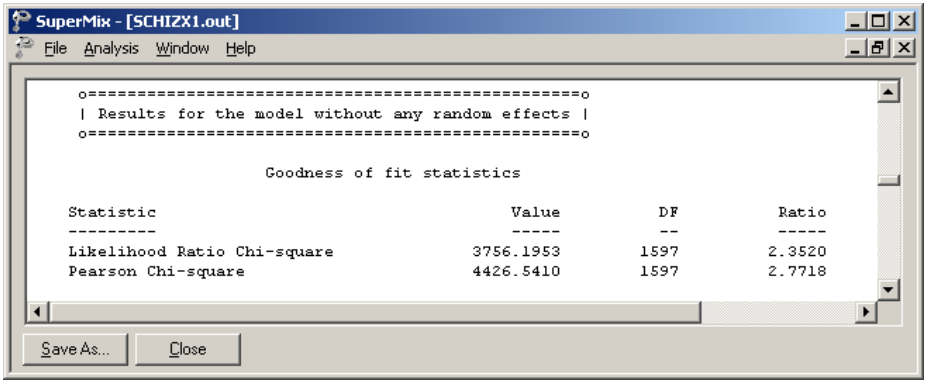


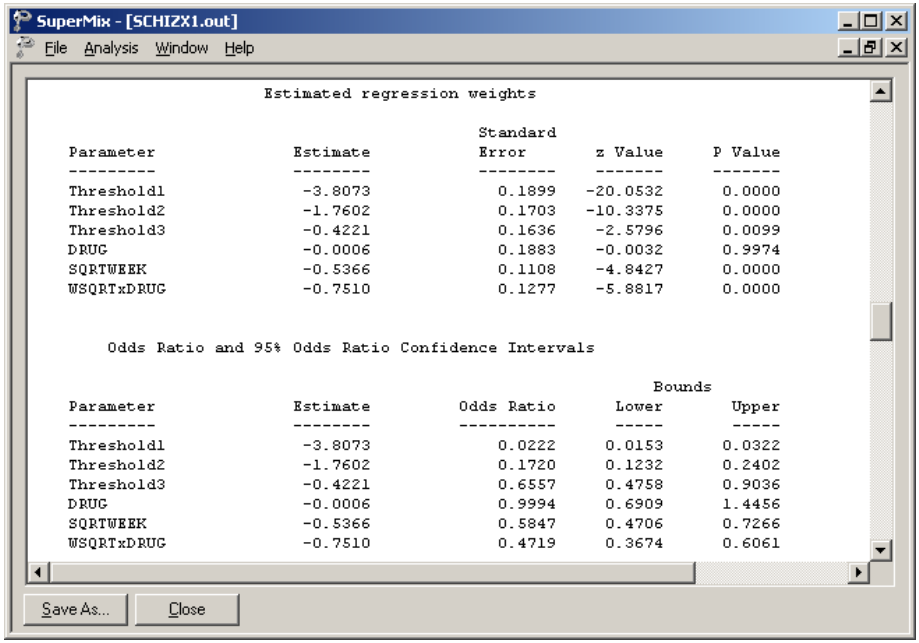
Descriptive statistics and starting values

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.



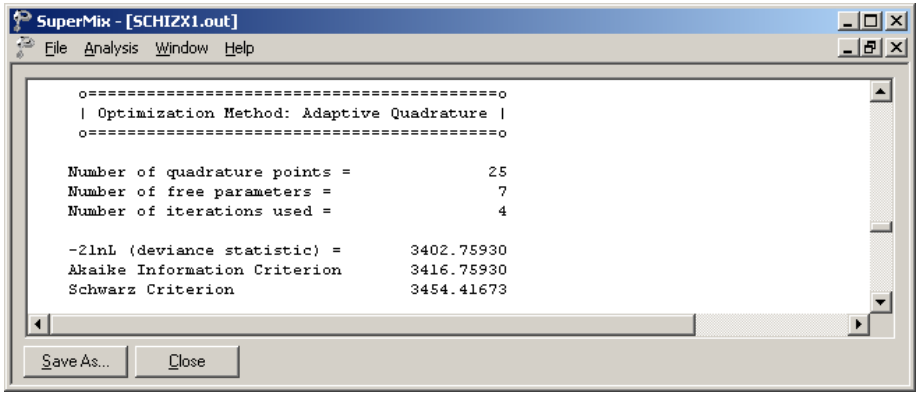
Descriptive statistics are followed by the starting values of parameters.

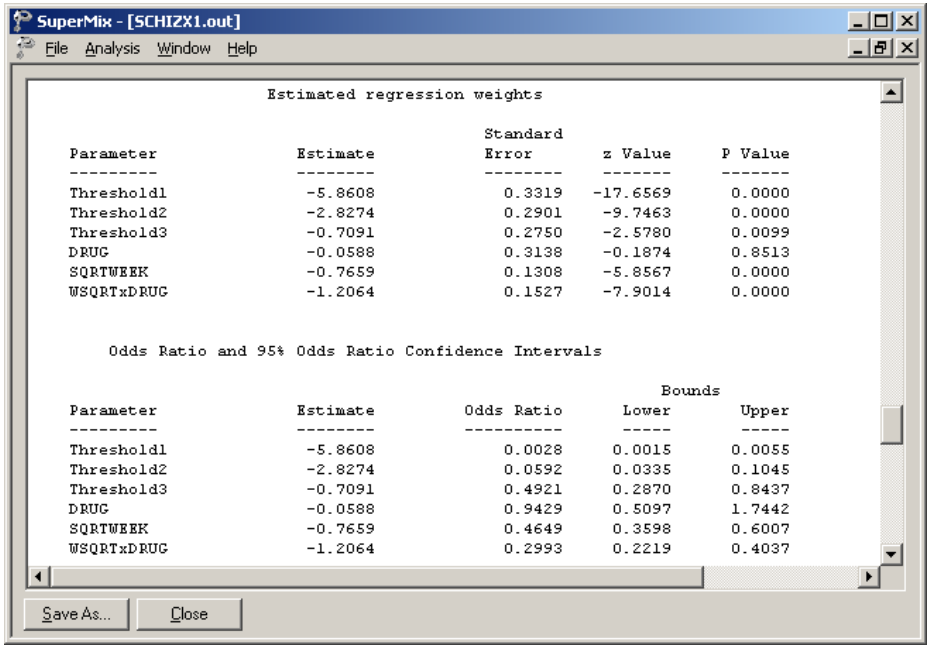


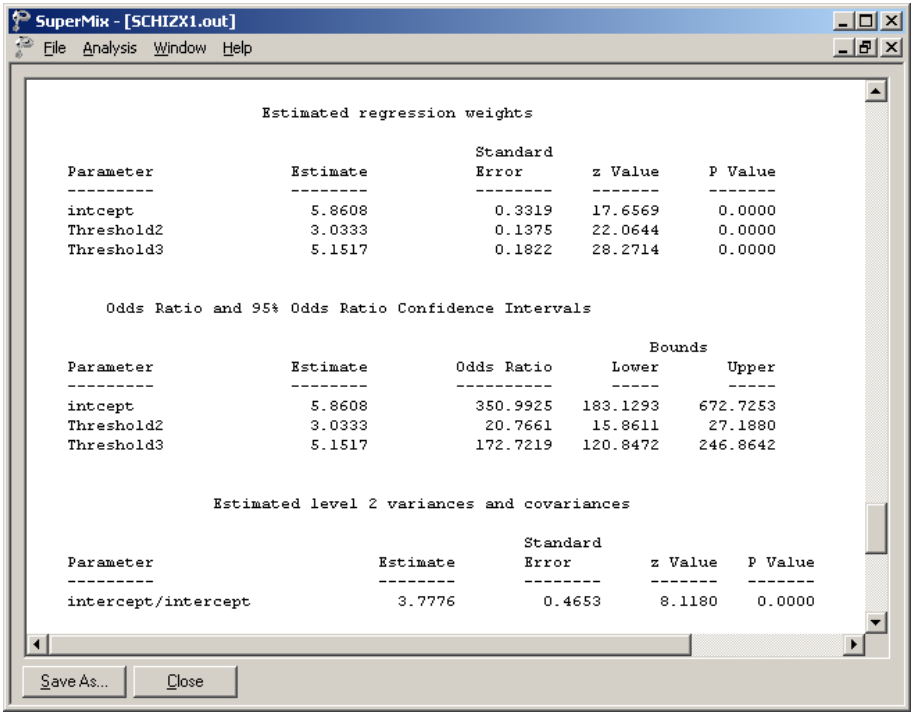


Fixed effects estimates

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


SuperMix - [SCHIZX1.out]
File Analysis Window Help
o==========o
| Optimization Method: Adaptive Quadrature |
o==========o
Number of quadrature points = 25
Number of free parameters = 7
Number of iterations used = 4
-2lnL (deviance statistic) = 3402.75930
Akaike Information Criterion 3416.75930
Schwarz Criterion 3454.41673
Save As...
Close


SuperMix - [SCHIZX1.out]
File Analysis Window Help
Estimated regression weights
Parameter Estimate Standard Error z Value P Value
Threshold1 -5.8608 0.3319 -17.6569 0.0000
Threshold2 -2.8274 0.2901 -9.7463 0.0000
Threshold3 -0.7091 0.2750 -2.5780 0.0099
DRUG -0.0588 0.3138 -0.1874 0.8513
SQRTWEEK -0.7659 0.1308 -5.8567 0.0000
WSQRTxDRUG -1.2064 0.1527 -7.9014 0.0000
Odds Ratio and 95% Odds Ratio Confidence Intervals
Parameter Estimate Odds Ratio Bounds Lower Upper
Threshold1 -5.8608 0.0028 0.0015 0.0055
Threshold2 -2.8274 0.0592 0.0335 0.1045
Threshold3 -0.7091 0.4921 0.2870 0.8437
DRUG -0.0588 0.9429 0.5097 1.7442
SQRTWEEK -0.7659 0.4649 0.3598 0.6007
WSQRTxDRUG -1.2064 0.2993 0.2219 0.4037
Save As...
Close

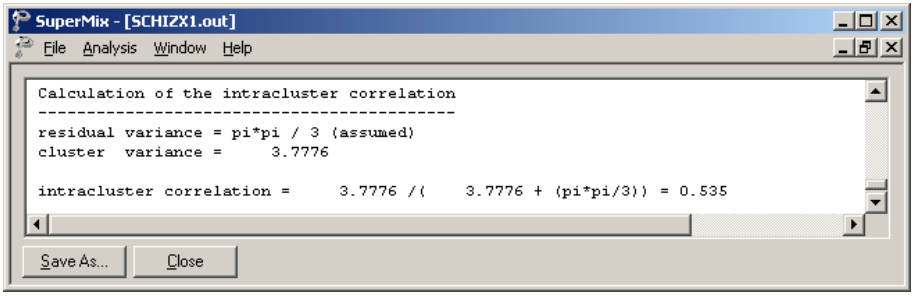


The variation in the intercept over the subjects is estimated as 3.7739, and from the associated *p*-value we conclude that there is significant variation in the (random) intercept between the patients included in this analysis. In the case of the fixed effects, a 2-tailed *p*-value is used, as the alternative 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.

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



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.

6.3.2.5 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:

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

or, in the notation introduced in Section 6.3.2.2,

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

When $c = 1$, we find that, for a patient from the control group (DRUG = 0, SQRTWEEK = WSQRTxDR = 0),

$$\log\left(\frac{\hat{P}_{ij1}}{1-\hat{P}_{ij1}}\right)=\hat{\eta}_{ij1}=-5.8593$$

$$\hat{P}_{ij1}=\frac{e^{\hat{\eta}_{ij1}}}{1+e^{\hat{\eta}_{ij1}}}=0.0028$$

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

	Drug patients (drug = 1)				Placebo patients (drug = 0)			
severity	**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%

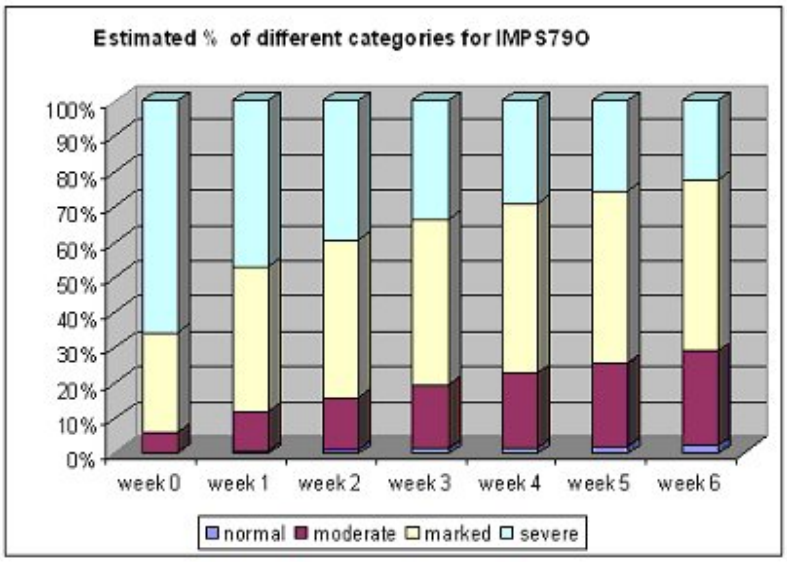


Figure 6.8: Estimated percentage of patients over time (treatment group)

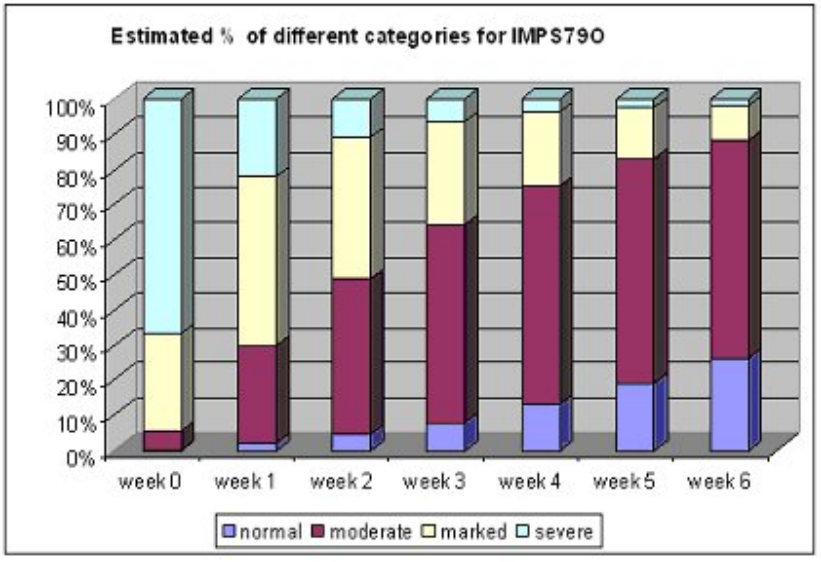


Figure 6.9: Estimated percentage of patients over time (control group)

6.3.3 A 2-level random intercept model and trend model

In this section, we fit a model with random intercept and slope. To do this, the level-1 model is unchanged; only the level-2 model is modified.

6.3.3.1 The model

Level 1 model:

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

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

Level 2 model:

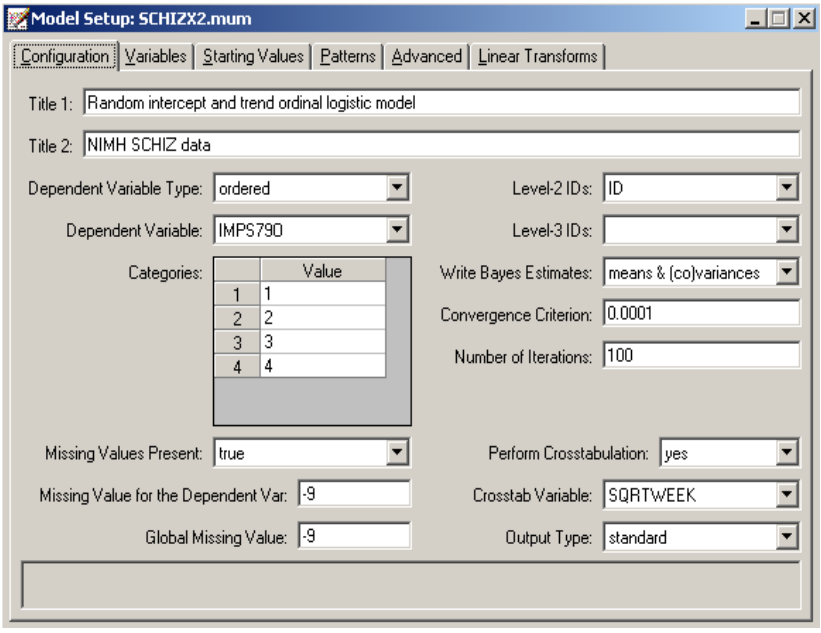
$$b_{0i} = \beta_0 + v_{0i}, \quad i = 1,\ldots,N$$
$$b_{1i} = \beta_1$$
$$b_{2i} = \beta_2 + v_{2i}$$
$$b_{3i} = \beta_3$$

As shown above, the slope of the time variable b_{2i} is now estimated by a level-2 equation containing both a fixed and a random effect. It indicates that patients are now not only assumed to have different intercepts, but may also exhibit different responses to the treatment over time.

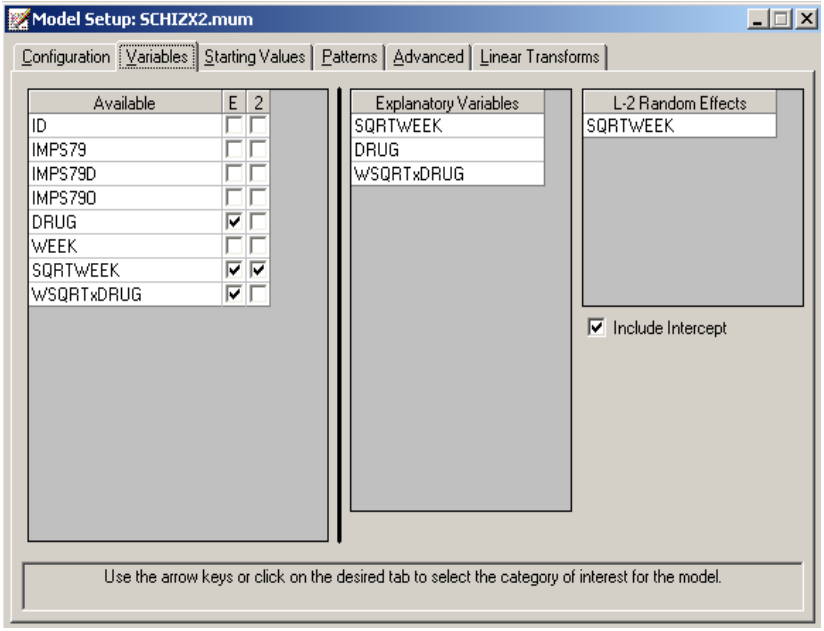
6.3.3.2 Setting up the analysis

Use the **File**, **Open Spreadsheet** option to re-open the previously used spreadsheet **schizx.ss3** from the **Examples\Ordinal** folder. Next, use the **File**, **Open Existing Model Setup** option to locate and open the syntax file **SCHIZX1.mum**. Click on **File**, **Save as** to save the model setup in a new file, such as **SCHIZX2.mum**. Next, we change the string in the **Title 1** text box on the **Configuration** screen (optional). Request a crosstabulation of the variable SQRTWEEK by the response variable

IMPS79O by selecting the **yes** option from the **Perform Crosstabulation** drop-down list box, followed by the selection of **SQRTWEEK** as the **Crosstab Variable**.



Proceed to the **Variables** tab, and check the **2** check box for SQRTWEEK to select it as a level-2 random variable as shown below.

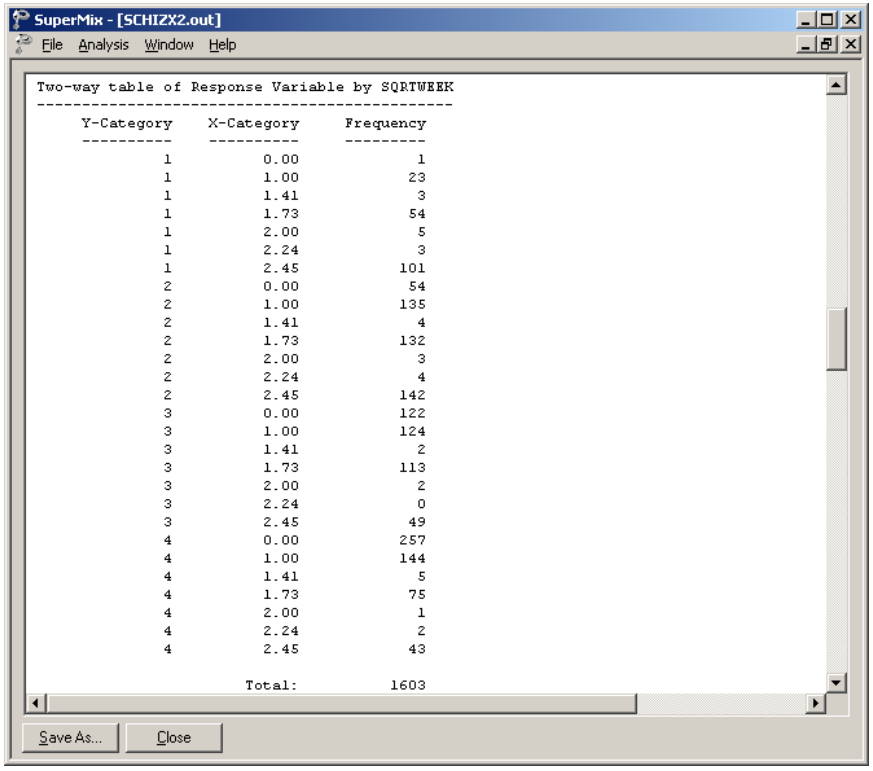


Keep all the other settings unchanged. Save the changes to the file **SCHIZX2.mum** and click the **Analysis, Run** option to produce the output file **SCHIZX2.out**.

6.3.3.3 Discussion of results

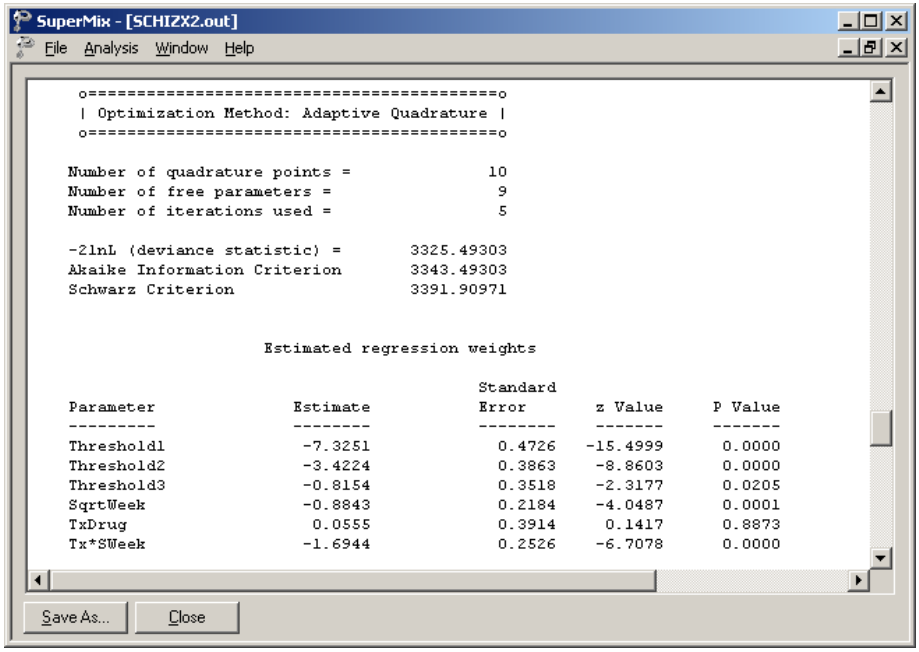
Crosstabulation

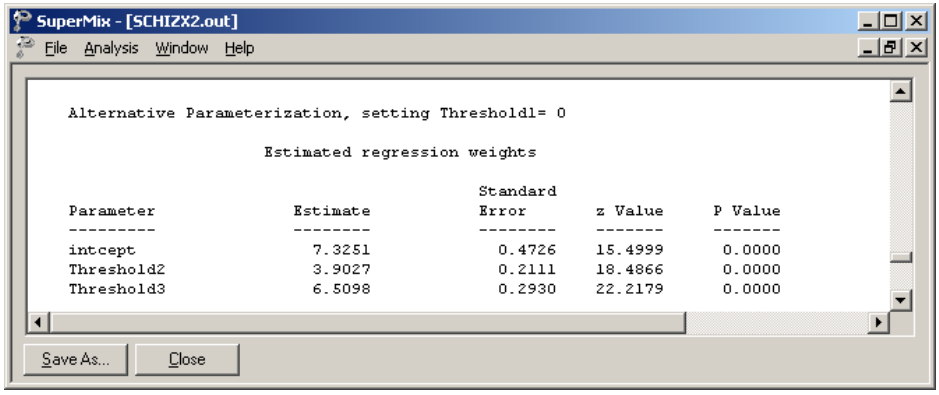
The following portion of the output is a crosstabulation of the seven distinct values of the variable SQRTWEEK by the four categories of the outcome variable IMPS79O. We note that there are relatively few observations for the third, fifth and sixth weeks. For example, for week 5 (SQRTWEEK = 2.24), measurements on only 9 of the 437 patients are available. Looking down the columns (SQRTWEEK) we see the severity of symptoms (IMPS79O) declining.



Fixed effect results

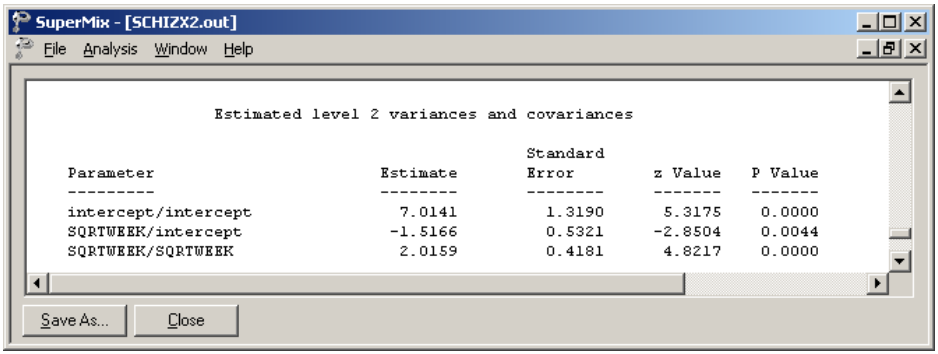
The final results after 23 iterations are listed below. While the values of the estimated coefficients differ from those in the random-intercept-only model, the overall picture remains very similar. The decline in severity over time noticed in the crosstabulation is captured by the significant fixed effect coefficient of –0.88295 for SQRTWEEK.





Random effects results

Note that the estimated coefficient for the random SQRTWEEK slope is highly significant, indicating that patients not only start at different points but follow different paths during the treatment period.



6.3.3.4 Interpreting the output

Estimated outcomes for groups: unit-specific results

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 root of time, we use the expression below:

$$\log\left(\frac{\hat{P}_{ijc}}{1-\hat{P}_{ijc}}\right)=\hat{\gamma}_c-\left[\hat{b}_{0i}+\hat{b}_{1i}\mathrm{DRUG}_i+\hat{b}_{2i}\mathrm{SQRTWEEK}_i+\hat{b}_{3i}\left(\mathrm{WSQRT}\times\mathrm{DRUG}\right)_i\right]$$

so that

$$\hat{\eta}_{ijc}=\hat{\gamma}_c-7.3793+0.0553\times\mathrm{DRUG}_i+0.8841\times\mathrm{SQRTWEEK}_i$$
$$+1.6940\times\left(\mathrm{WSQRT}\times\mathrm{DRUG}\right)_i$$

As illustrated in the previous example, by substituting the values for DRUG, SQRTWEEK and WSQRTxDRUG, the results shown in Table 6.9 can be obtained.

Table 6.9: Estimated unit-specific results for random intercept & slope model

	Placebo patients (drug = 0)				Drug patients (drug = 1)			
severity	**normal**	**moderate**	**marked**	**severe**	**normal**	**moderate**	**marked**	**severe**
week 0	0.06%	2.96%	26.90%	70.08%	0.07%	3.13%	27.90%	68.91%
week 1	0.15%	6.87%	43.81%	49.17%	0.86%	29.42%	55.32%	14.40%
week 2	0.22%	9.61%	50.03%	40.15%	2.47%	51.90%	39.98%	5.81%
week 3	0.29%	12.32%	53.77%	33.62%	5.42%	68.72%	23.37%	2.49%
week 4	0.36%	15.09%	55.99%	28.55%	10.27%	74.85%	13.62%	1.26%
week 5	0.45%	17.94%	57.12%	24.49%	17.38%	73.94%	7.99%	0.69%
week 6	0.54%	20.84%	57.44%	21.17%	26.72%	68.08%	4.80%	0.40%

We can again represent the results from the above table graphically, as shown in Figures 6.10 and 6.11. The graphs tell us the same story as the previous model: patients from the treatment group showed more improvement over time than patients from the control group. While a very small proportion of treatment patients were still diagnosed as being severely ill at the end of the treatment period (0.42% according to table 6.9), 20% of the control group were still classified as being severely ill by week 6.

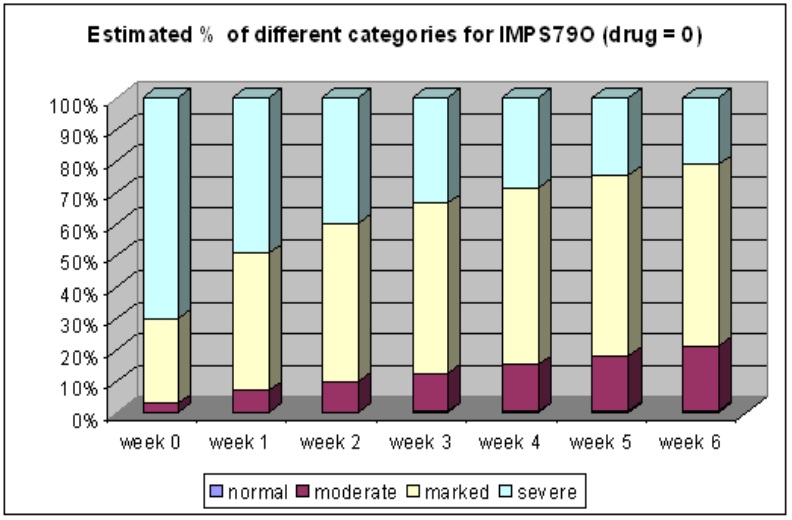


Figure 6.10: Estimated percentage of patients over time (treatment group)

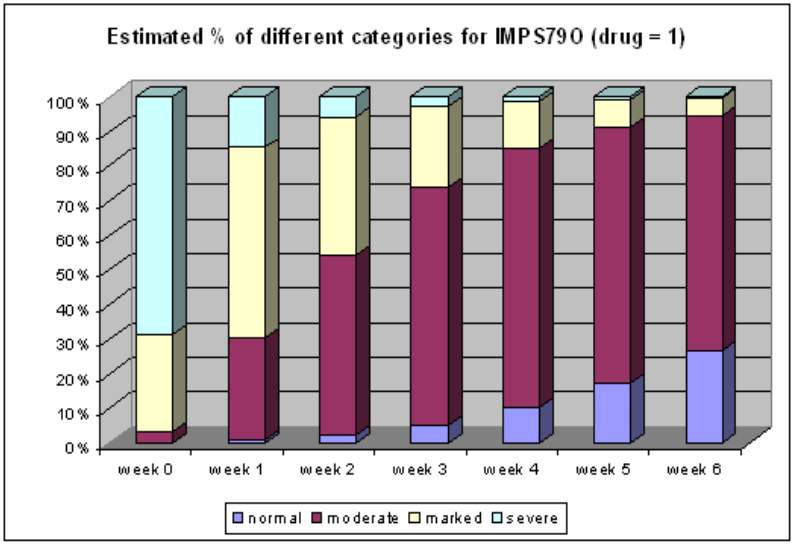


Figure 6.11: Estimated percentage of patients over time (treatment group)

Estimated time trend variance

When we consider the heterogeneity in responses across time, we notice that the estimated variance in the time trend is $\sigma_{v_1}^2 = (1.29774)^2 + (-0.57054)^2 = 2.0096$. The estimates for the time trends are -0.88295 for SQRTWEEK and -1.69416 for WSQRTxDR respectively. Thus the estimated trends for the placebo and drug groups are -0.88295 and -0.88295 -1.69416 = -2.57711. Thus the 95% confidence interval of the time trend for the placebo group is $-0.88295 \pm \left(1.96 \times \sqrt{2.0096}\right)$ $= \left(-3.6615, 1.896\right)$. Similarly, the confidence interval for the drug group is $\left(-5.3556, 0.2014\right)$. Notice that both intervals are fairly large and include negative and positive slopes, which reflects the wide heterogeneity in trends. The estimated correlation value is –0.402, which is moderately large. This indicates that the patients who are initially less severely ill improve at a smaller rate. The more severely ill patients improve at a greater rate.

7 Models for nominal outcomes

7.1 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:

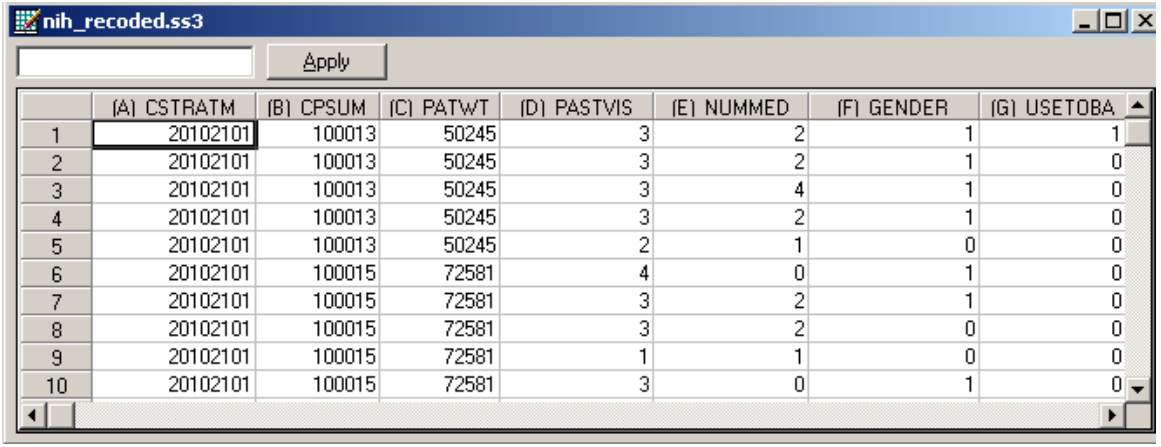
- Annual Income (1 = not employed, 2 = less or equal to $20,000, 3 = more than $20,000 but less than or equal to $50,000, 4 = more than $50,000 but less than or equal to $100,000, 5 = more than $100,000)
- Age when diagnosed (1 = not applicable, 2 = younger than 25 years, 3 = 25 years or older but less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

7.1.1 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

7.1.2 The model

7.1.2.1 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 $1, 2, 3, ..., C$.

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

within level-2 unit j, $j = 1, 2, \ldots n_i$, which in turn is nested within the i-th level-3 unit, where $i = 1, 2, \ldots, N$. The probabilities P_{ijkc} are computed as

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

where

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

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

7.1.2.2 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

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

where

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

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

Remarks:

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

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

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

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

7.1.3 A random intercept model with fourteen predictors

7.1.3.1 Preparing the data

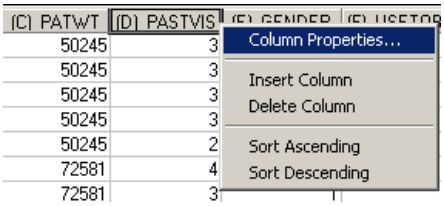
The model to be fitted to the data is contained in **nihs_recoded.ss3**. This file was created from the SPSS data file **nihs_recoded.sav** as follows.

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**.

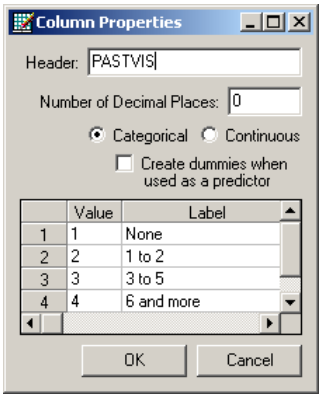
7.1.3.2 Exploring the data

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

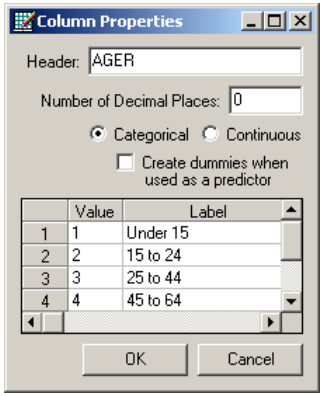
the **Column Properties** dialog. Select the **Nominal** or **Ordinal** option to obtain the list of values assigned to the categories of PASTVIS.



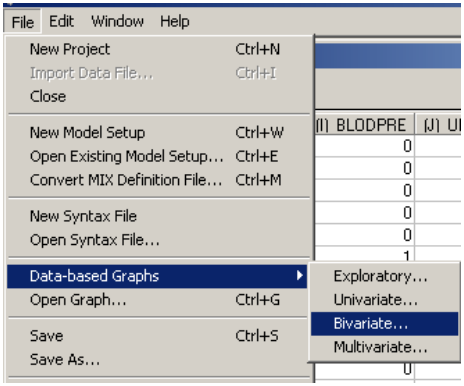
Enter the labels None, 1 to 2, 3 to 5 and 6 and more as shown below and click **OK**.



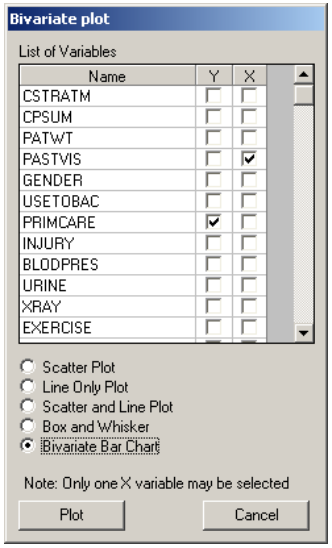
Category labels can be assigned to the variables GENDER and AGER in a similar way. For example, right-click on the AGER header and enter labels as shown below.



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.

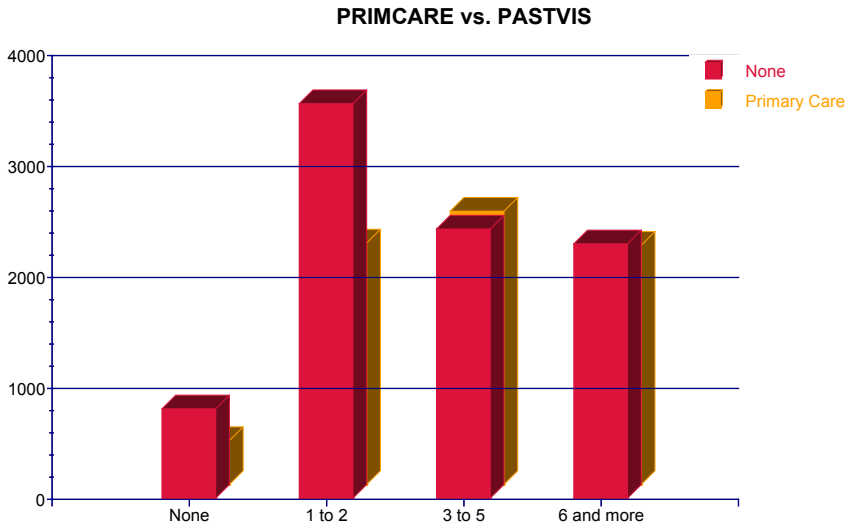


Figure 7.1: Bivariate bar chart of PRIMCARE vs PASTVIS

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

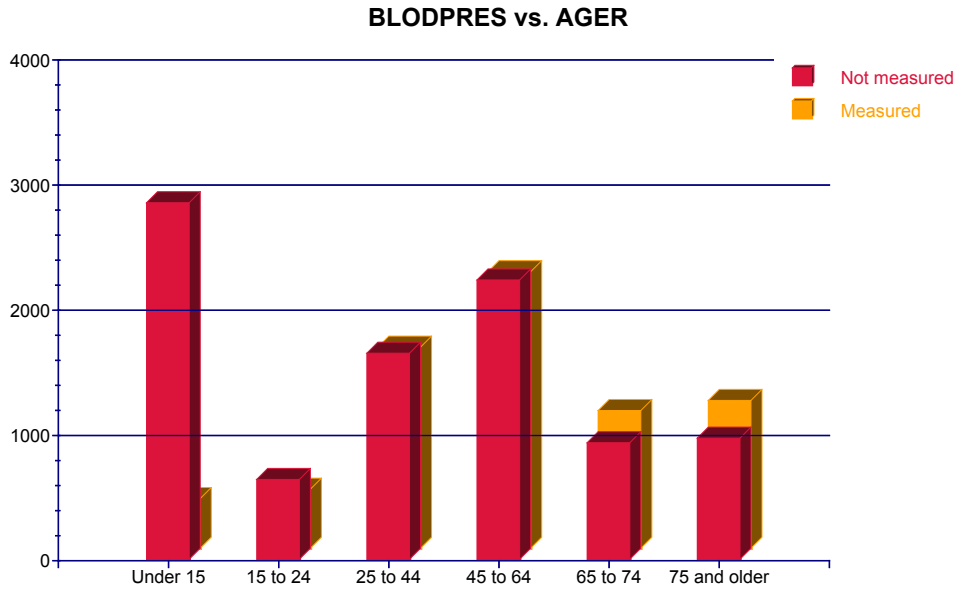


Figure 7.2: Bivariate bar chart of BLODPRES vs AGER

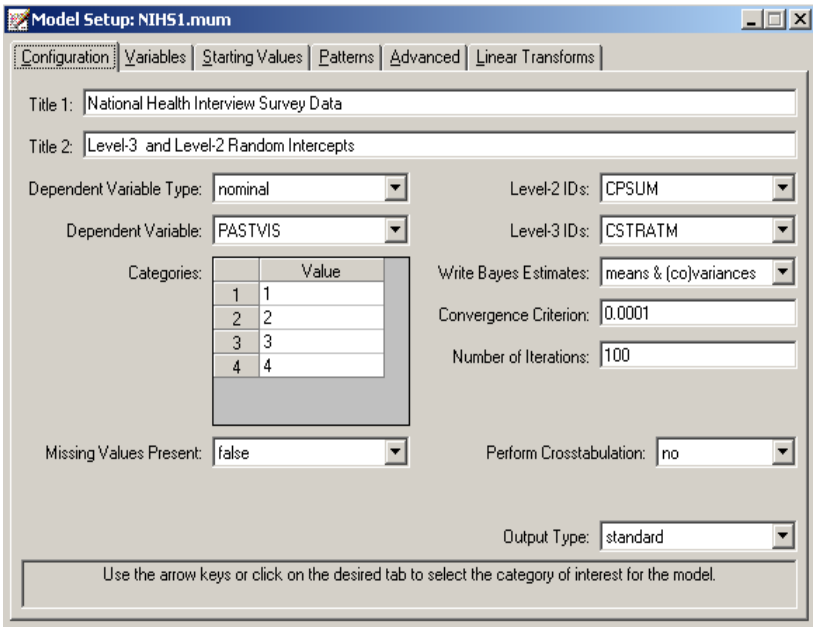
From Figure 7.2 it is evident that it is much more likely that blood pressure will be measured for patients older than 45 than for younger patients.

7.1.3.3 Setting up the analysis

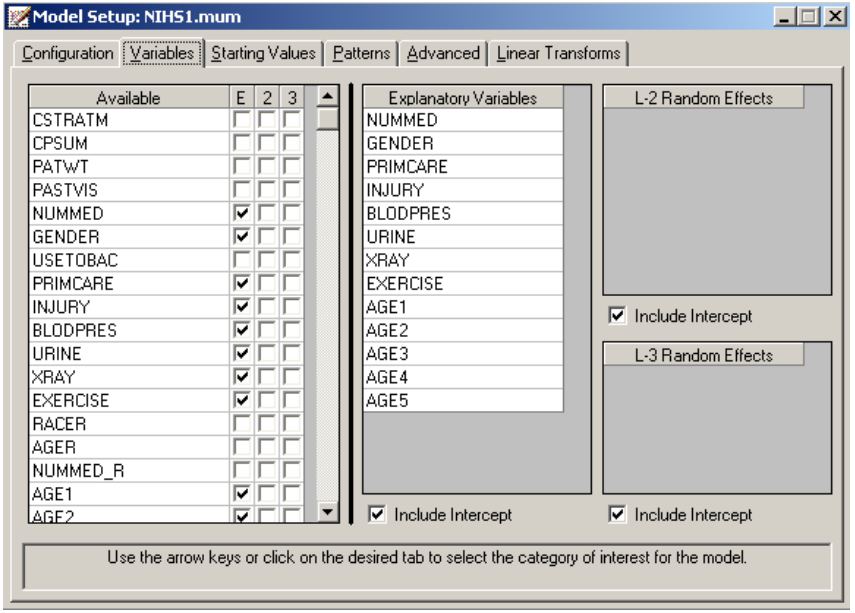
From the main menu bar, select the **File**, **New Model Setup** option. In this example, only the **Configuration**, **Variables** and **Advanced** tabs of the **Model Setup** window are used. By default, the **Configuration** tab is displayed first.

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.

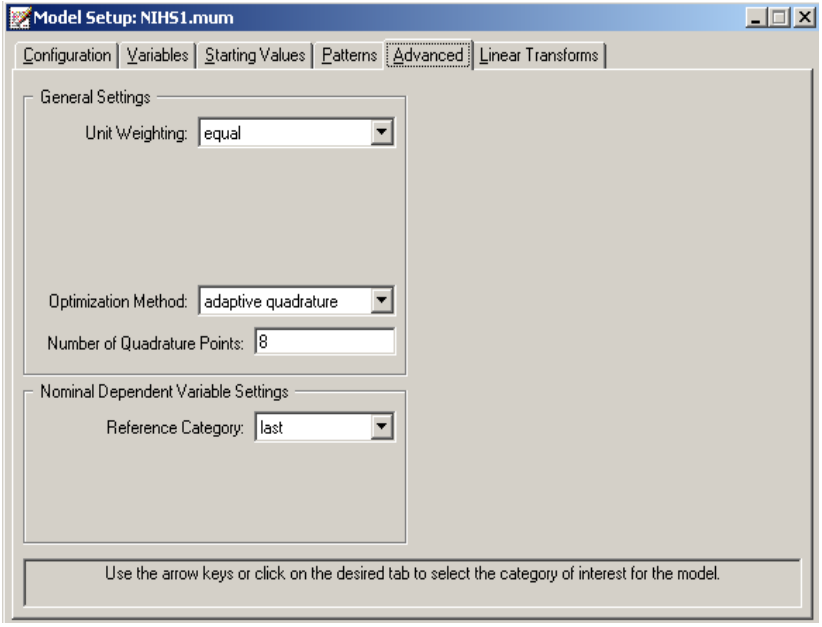


The **Variables** tab is used to specify the fixed and random effects to be included in the model. Start by selecting the explanatory (fixed) variables, using the check boxes in the **Available** grid. The image below shows the completed selection of all predictors.



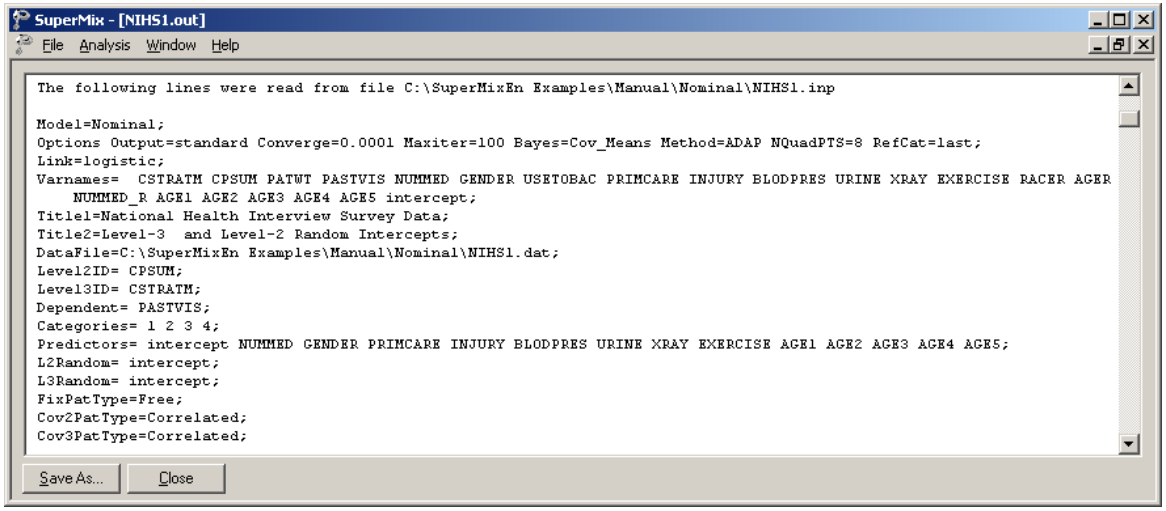
By default, the inclusion of a fixed intercept coefficient and random intercepts at level 2 and level 3 is assumed. As these selections correspond to the model we intend fitting to the data, no further changes are made on this tab. Click on the **Advanced** tab and request adaptive quadrature as the **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 **NHIS1.mum**, and then run the analysis using the **Analysis**, **Run** option.

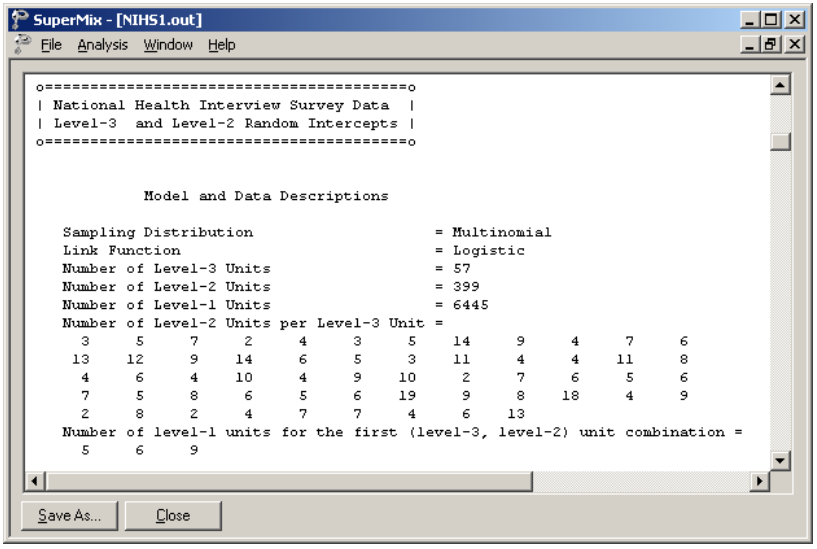


7.1.3.4 Discussion of results

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.



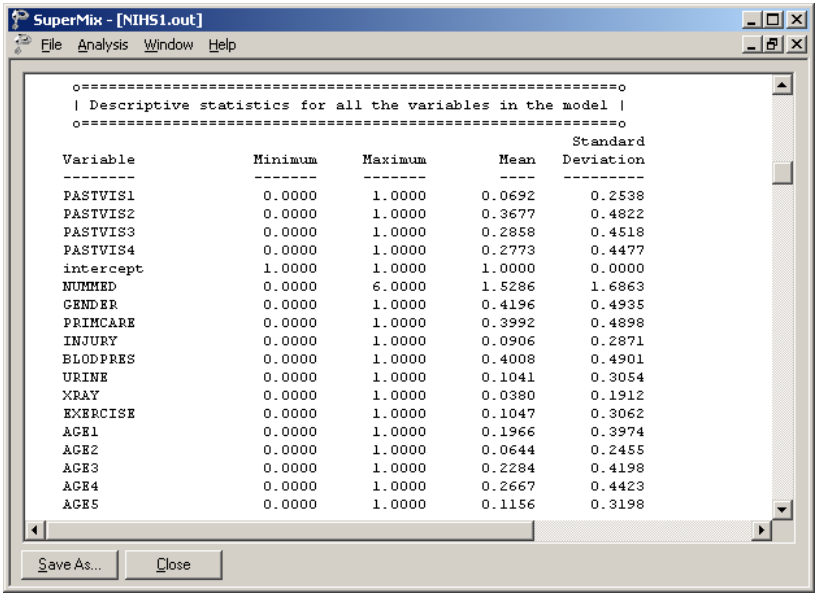
Model and data description



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.

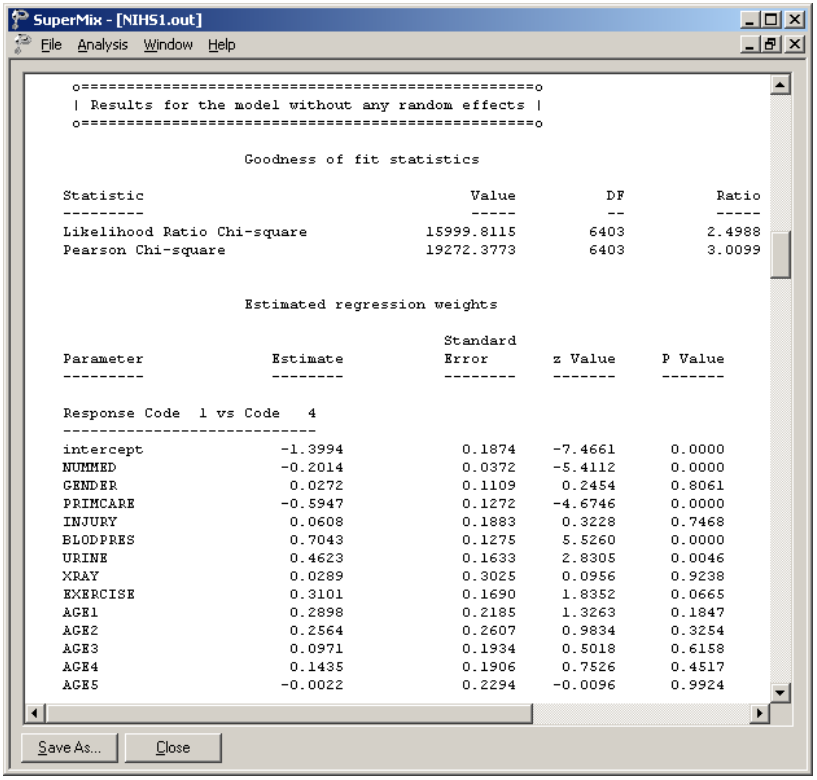


Each category of the nominal outcome variable is denoted as PASTVIS_i, $i = 1, 2, 3, 4$. From the output it can be seen that the distribution of respondents over these categories are 6.9%, 36.8%, 28.6%, and 27.7% respectively. The age distribution is given in Table 7.2.

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*

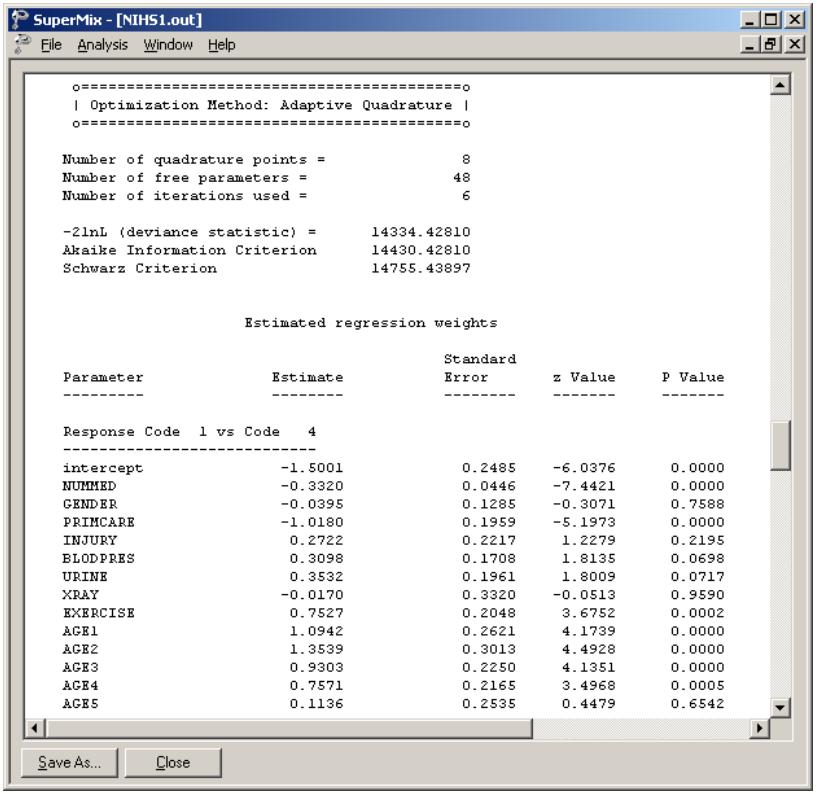
*: calculated as 100 – (19.7+6.4+22.8+26.7+11.6)



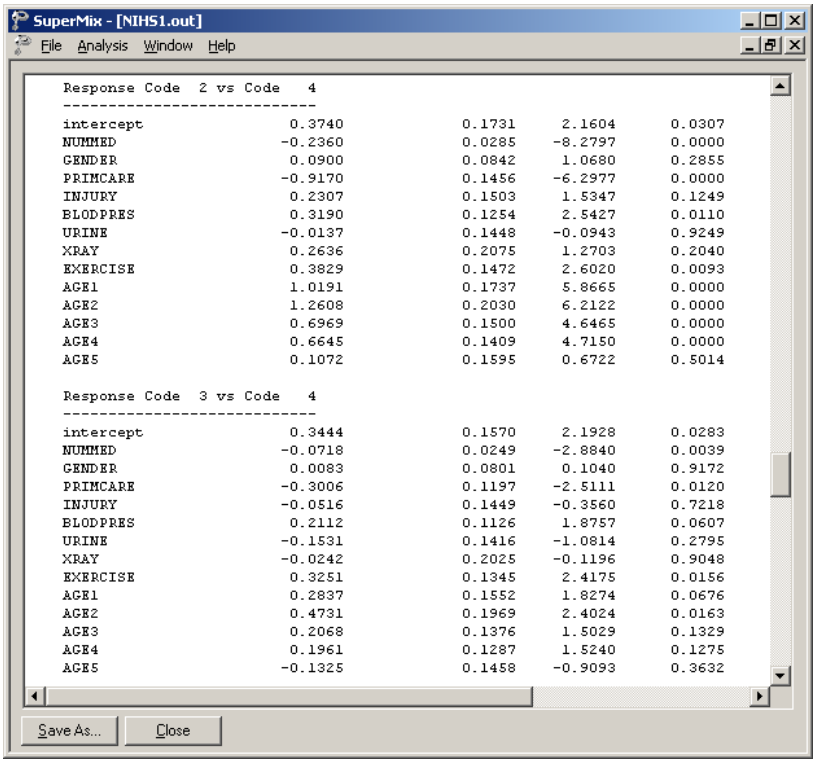
The estimated parameters for the model, assuming no random effects, are reported next. For each response code i versus code 4, $i = 1, 2, 3$, there are 14 parameter estimates. Only the estimates for response code 1 versus response code 4 are displayed. Comparing these estimates with those obtained when 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.

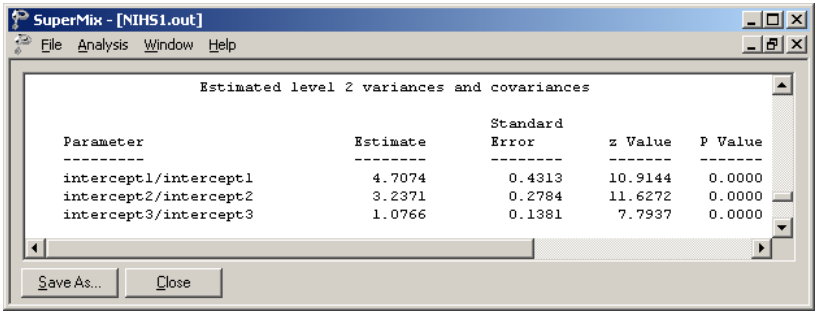


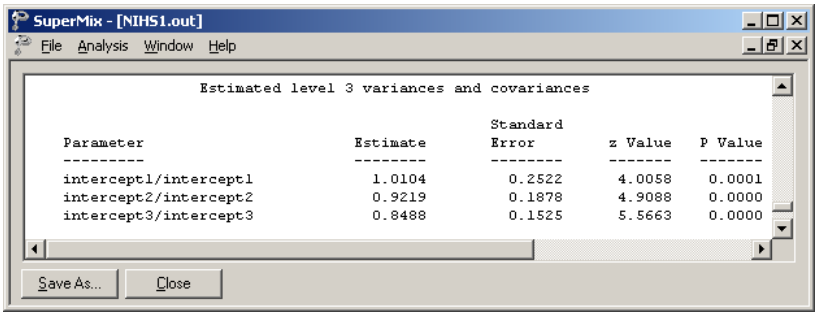
A study of the p-values associated with the parameter estimates indicates that the estimated GENDER, INJURY, URINE, and XRAY coefficients are not significant, regardless of the values of the category of the outcome variable.



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.

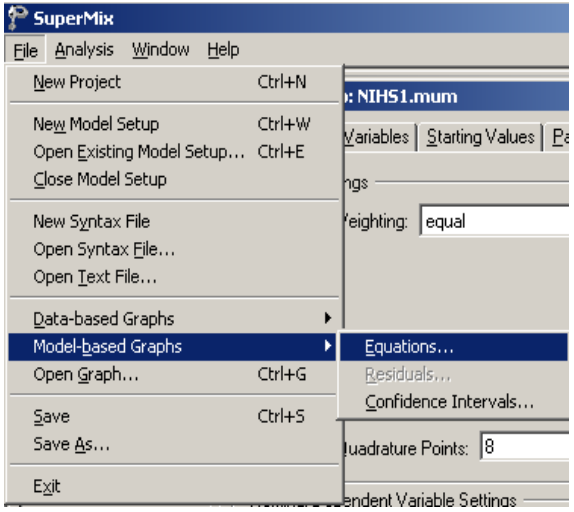




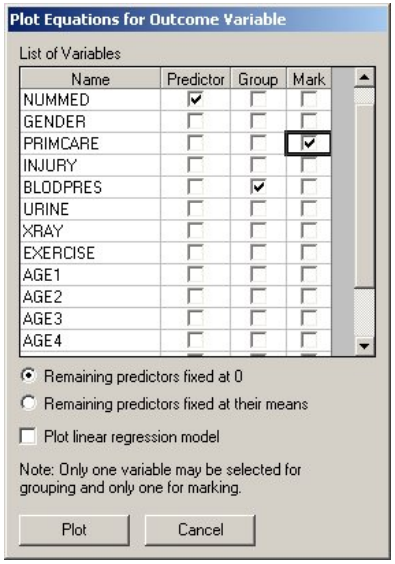
7.1.3.5 Interpreting the output

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.

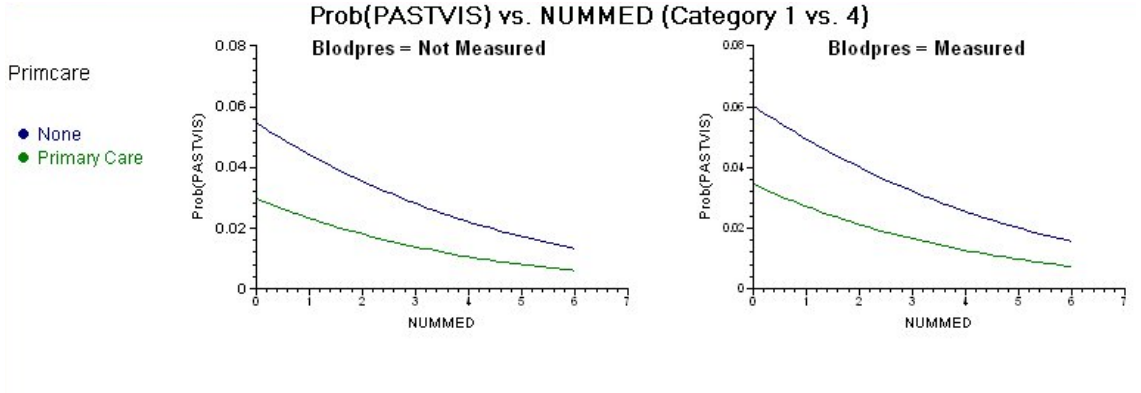


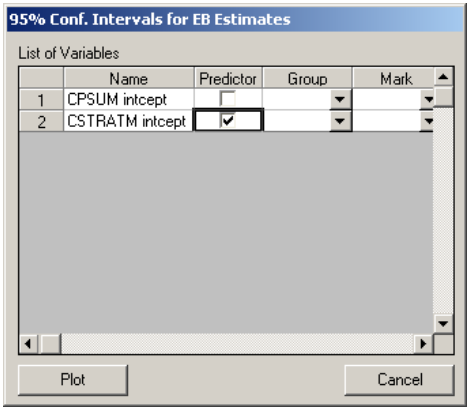
Figure 7.3: Plot of Prob(PASTVIS) vs. NUMMED

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 $\hat{v}_{0i} \pm 1.96\widehat{\sigma}_{v_{0i}}$, where $\widehat{\sigma}_{v0i}$ is the square root of the variance of $\hat{v}_{0i}$. Figure 7.4 shows that the estimated deviances from $\widehat{\beta}_0$ 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.

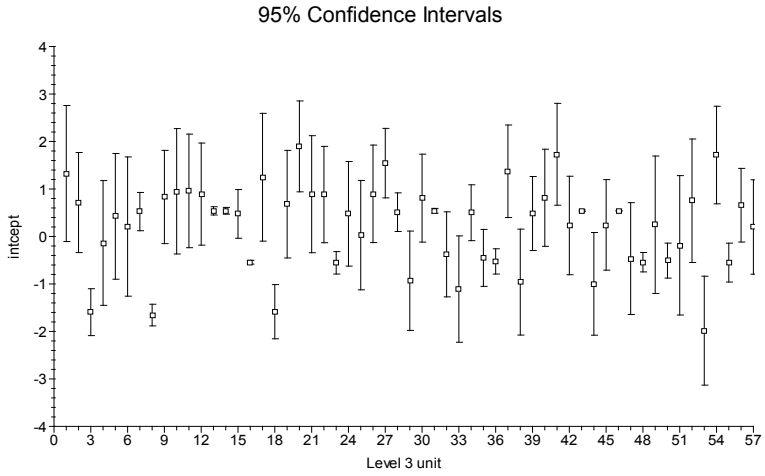
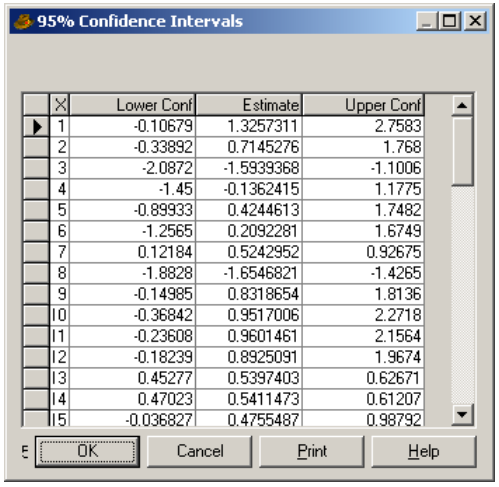


Figure 7.4: 95% Confidence intervals for level-3 EB estimates

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.



X	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:

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

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

For response code 2 vs. 4, we find that

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

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

For response code 3 vs. code 4

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

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

Using these values, it follows that

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

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

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

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

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

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

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

Estimated population-average probabilities

The population-average probabilities are obtained by dividing the ETA1, ETA2 and ETA3 values given in 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

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

For the logistic model it is assumed that

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

Therefore

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

Similarly,

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

and

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

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

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

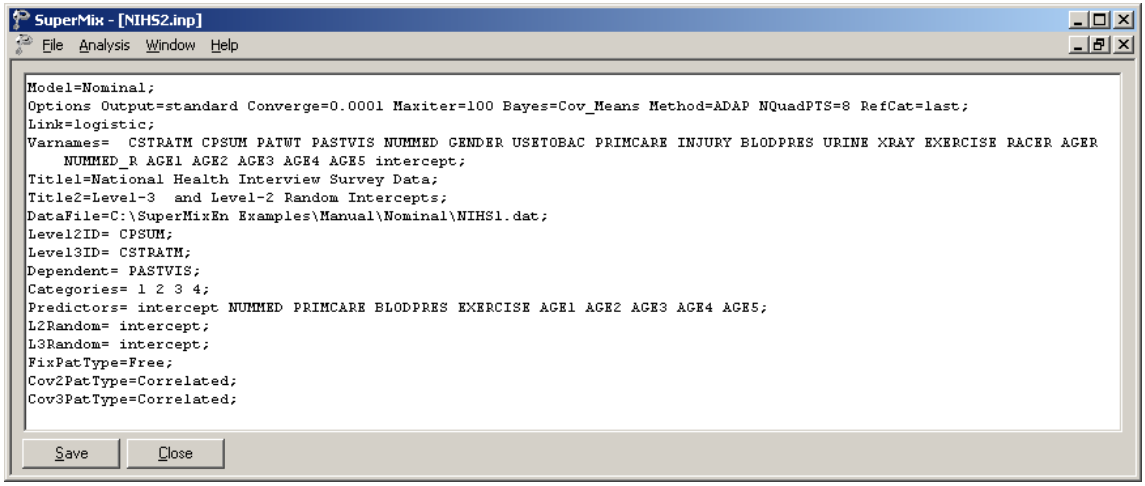
NUMMED	PRIM CARE	EXER CISE	AGE	ETA1	ETA2	ETA3	PROB1	PROB2	PROB3	PROB4
none	No	no	< 15	−0.058	1.138	0.667	0.135	0.445	0.278	0.143
none	No	no	25:44	−0.157	0.924	0.605	0.138	0.406	0.295	0.161
none	No	no	>= 75	−0.720	0.460	0.441	0.105	0.343	0.336	0.216
none	No	yes	< 15	0.397	1.392	0.925	0.165	0.446	0.279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

7.1.4 A random intercept model with ten predictors

7.1.4.1 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.

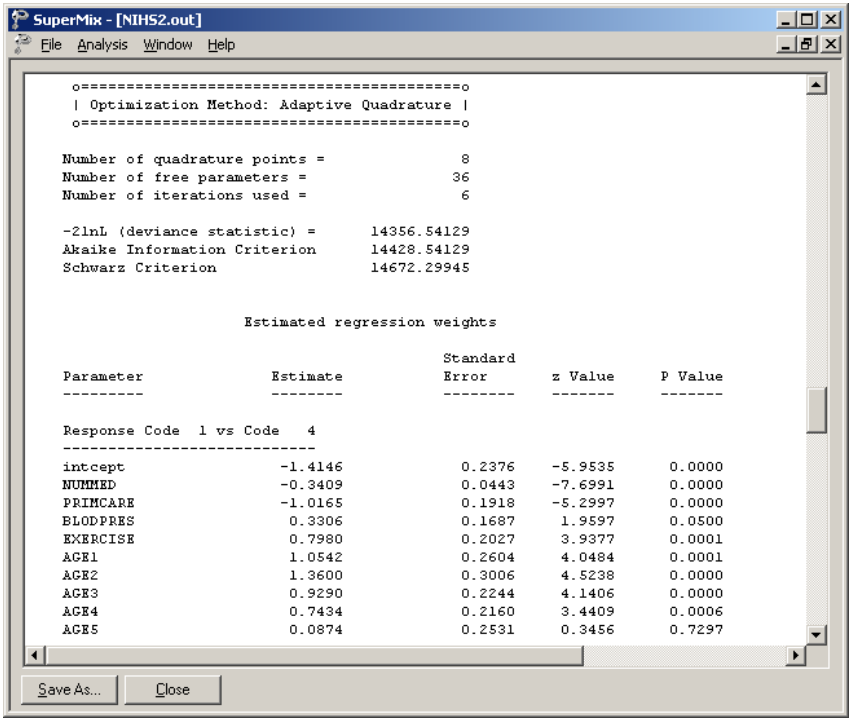
To run the model without these fixed effects, use the **File**, **Open Syntax File** option and select the command syntax previously saved to the file **nihs1.inp**. Delete the variables GENDER, INJURY, URINE and XRAY from the Predictors paragraph and save the modified syntax file as **nihs2.inp**. To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



7.1.4.2 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.

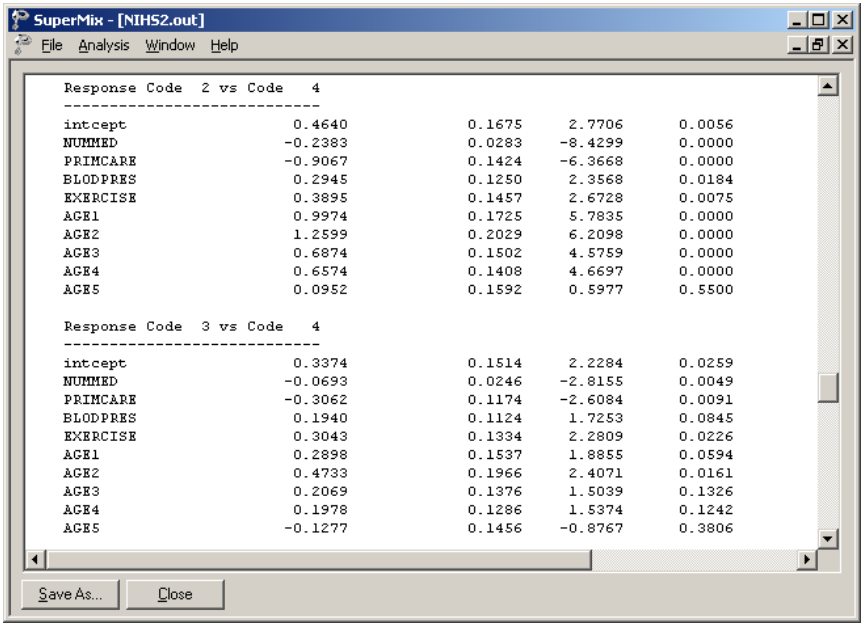


The χ^2-value obtained for this test is 14356.34 – 14334.43 = 21.91, with 12 degrees of freedom. Since the associated p-value equals 0.0385, the χ^2-value is significant at the 5% level, but not at the 1% level of significance. We therefore conclude that, based on the χ^2-difference test, we do not have 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.

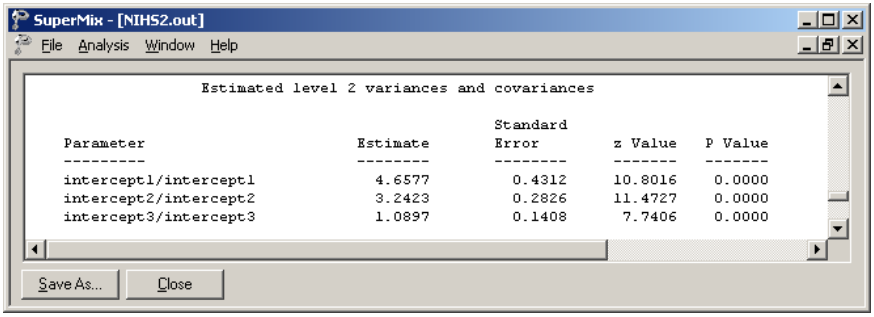
Table 7.7: Akaike and Schwarz fit criteria for two nested models

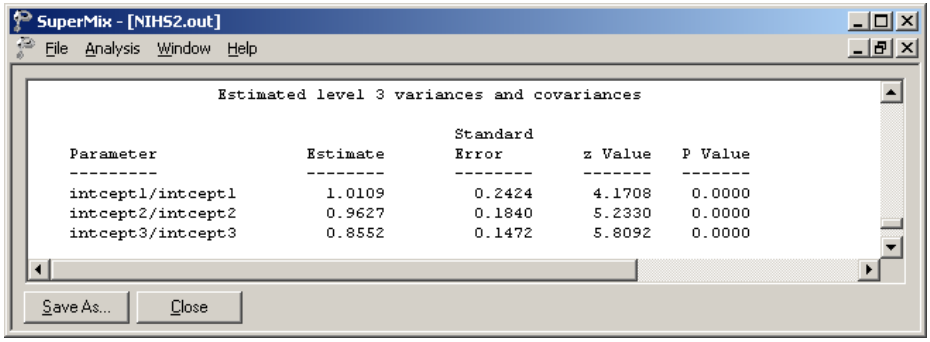
Fit statistic	14 predictors	10 predictors
Akaike	14430.43	14428.54
Schwarz	14755.44	14672.30

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.



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.

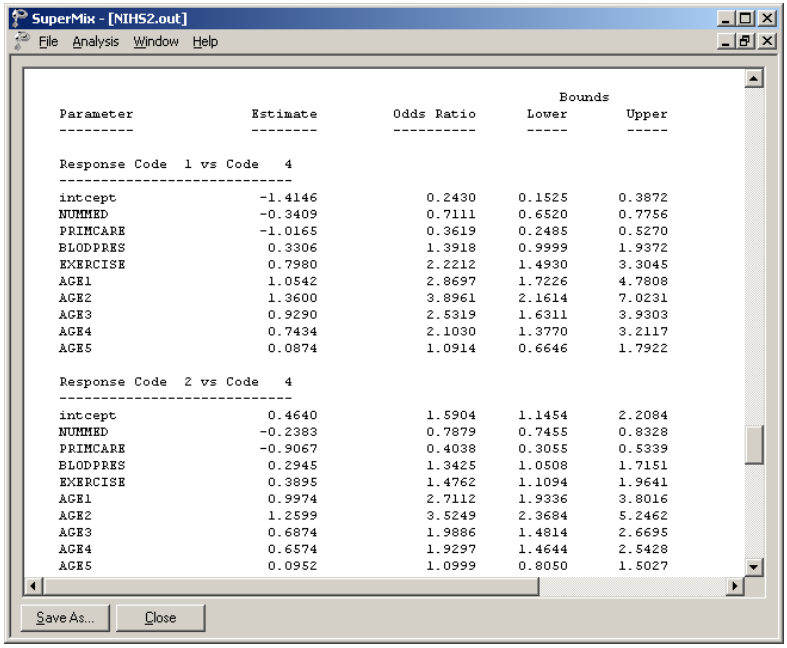




Odds ratios and 95% confidence intervals for the odds ratios

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.



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

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

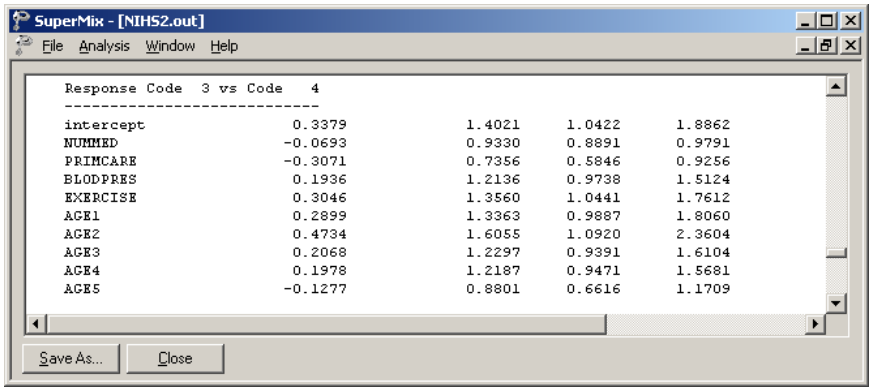
From the output, it follows that this interval is

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

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

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

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.


SuperMix - [NIH52.out]
File Analysis Window Help
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
Save As...
Close

8 Models for grouped- and discrete-time survival data

8.1 Introduction

Models for grouped-time survival data are useful for analysis of failure time data when subjects are measured repeatedly at fixed intervals in terms of the occurrence of some event, or when determination of the exact time of the event is only known within grouped intervals of time. Additionally, it is often the case that subjects are observed nested within clusters (*i.e.*, schools, firms, clinics), or are repeatedly measured in terms of recurrent events. In this case, use of grouped-time models that assume independence of observations (Thompson, 1977; Allison, 1982; Prentice & Gloeckler, 1978) is problematic since observations from the same cluster or subject are usually correlated.

For data that are clustered and/or repeated, models including random effects provide a convenient way of accounting for association in correlated survival data. In terms of continuous-time survival data, several authors have developed survival analysis models including random effects that are usually assumed to be distributed as a gamma distribution. These models are often termed frailty models or survival models including heterogeneity, and recent review articles describe many of these models (Pickles & Crouchley, 1995; Hougaard, 1995).

Several authors have noted the relationship between ordinal regression models (using complementary log-log and logistic link functions) and survival analysis models for grouped and discrete time. Hedeker, Siddiqui, and Hu (2000) described a generalization of an ordinal random-effects regression model to handle correlated grouped-time survival data. This model accommodates multivariate normally-distributed random effects, and additionally, allows for a general form for model covariates.

Assuming a proportional, or partial proportional, hazards or odds model, a maximum marginal likelihood solution is implemented using multi-dimensional quadrature to numerically integrate over the distribution of random-effects.

In this chapter, we explore various survival analysis models using the TVSFP data discussed in Sections 3.3, 4.2, and 6.2. Two analysis approaches are considered for these data in the examples to follow. The first treats survival time as a set of dichotomous indicators of whether the event occurred for time periods up to the period of the event or censoring. This analysis, shown in Section 8.4, uses the data set mentioned above. The second approach treats survival time as an ordinal outcome, which is either right-censored or not. The same data, but in different format, is used for this second analysis (see Section 8.6).

8.2 Choosing between binary and ordinal outcome models

8.2.1 The data for a binary approach

An analysis of a data set where students are clustered within schools is used to illustrate features of random-effects analysis of clustered grouped-time survival data.

As described in previous chapters, the TVSFP study was designed to test independent and combined effects of a school-based social-resistance curriculum and a television-based program in terms of tobacco use prevention and cessation. In previous chapters the focus was on the pre- and post-intervention knowledge of students of the dangers of smoking. Here, we focus on actual usage of tobacco products and on subsequent data collected from the respondents.

As mentioned previously, schools were randomized to one of four study conditions: (a) a social-resistance classroom curriculum (CC); (b) a media (television) intervention (TV); (c) a combination of curriculum and TV conditions; and (d) a no-treatment control group. These conditions form a 2 x 2 factorial design of CC (yes or no) by TV (yes or no).

The outcome variable of interest in this chapter is the response the question "Have you ever tried a cigarette?." Students were assessed at 4 occasions:

- pre-intervention (January 1986, also referred to as Wave A)
- post-intervention (April 1986, *i.e.* Wave B)
- year follow-up (April 1987, *i.e.* Wave C)
- year follow-up (April 1988, *i.e.* Wave D)

As the intervention procedures were implemented following the pretest, we focus in the analyses to follow on the three post-intervention time points and include only those students who had not answered yes to this question at pretest. Of the original 1,600 respondents, 1,556 are included in the data considered here. Thus, our analysis examines the degree to which the intervention prevented or delayed students from initiating smoking experimentation. Because the intervention was also aimed at smoking cessation for individuals who had initiated smoking, here we are examining only a part of the intervention aims.

The first few lines of the SuperMix spreadsheet **SMKBCD2.ss3** used in this section are shown below. Note that there is a maximum of 3 observations associated with each student – not all students have data at all 3 occasions.

SMKBCD2.ss3

Apply

	(A) School	(B) Class	(C) Student	(D) Event	(E) TimeC	(F) TimeD	(G) SexM	(H) CC	(I) TV	(J) CC*TV
1	193	193101	193101105	0	0	0	1	0	0	0
2	193	193101	193101105	0	1	0	1	0	0	0
3	193	193101	193101105	0	0	1	1	0	0	0
4	193	193101	193101108	0	0	0	0	0	0	0
5	193	193101	193101108	0	1	0	0	0	0	0
6	193	193101	193101108	0	0	1	0	0	0	0
7	193	193101	193101113	1	0	0	1	0	0	0
8	193	193101	193101117	1	0	0	1	0	0	0
9	193	193101	193101121	0	0	0	1	0	0	0
10	193	193101	193101121	0	1	0	1	0	0	0

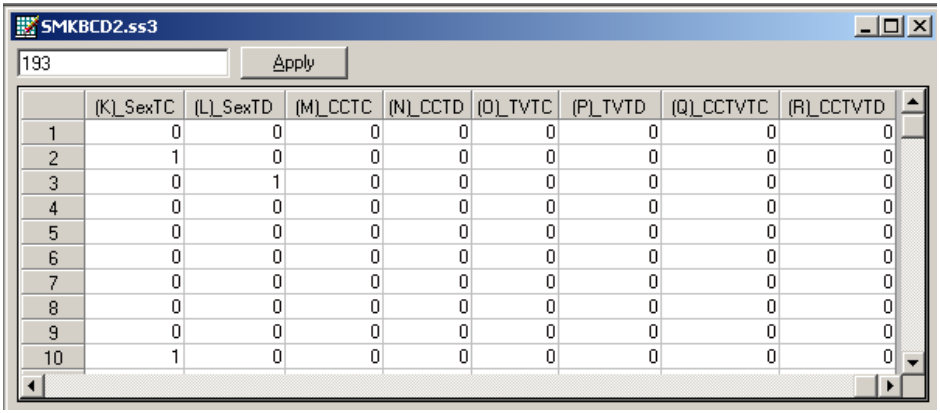
The variables of interest are:

- School indicates the school a student is from.
- Class identifies the classroom to which a student belongs.
- Student represents the student identification number.
- Event indicates occurrence of the event (1 indicating "yes" and 0 "no.").
- TimeC is an indicator variable indicating the first follow-up occasion after the post-intervention measurement occasion. It assumes a value of 1 if a measurement was made at the first follow-up occasion, and 0 otherwise.
- TimeD is the indicator variable for the second follow-up occasion. It assumes a value of 1 if a measurement was made at the second follow-up occasion, and 0 otherwise.
- SexM is an indicator variable for gender, with "1" indicating male respondents, and "0" female respondents.
- CC is a binary variable indicating whether a social-resistance classroom curriculum was introduced, with 0 indicating "no" and 1 "yes."
- TV is an indicator variable for the use of media (television) intervention, with a "1" indicating the use of media intervention, and "0" the absence thereof.

The post-intervention measurement, which is the first of the three measurement occasions in this data set, serves as the reference cell. In terms of the indicator variables TimeC and TimeD it would be a measurement for which TimeC = TimeD = 0.

In addition to these variables, **SMKBCD2.ss3** includes a number of interaction terms:

- CCTV was constructed by multiplying the variables TV and CC, and represents the CC by TV interaction.
- SexTC denotes the SexM by TimeC interaction.
- SexTD denotes the SexM by TimeD interaction.
- CCTC denotes the interaction between classroom curriculum intervention CC and TimeC.
- CCTD denotes the interaction between CC and TimeD.



	(K)_SexTC	(L)_SexTD	(M)_CCTC	(N)_CCTD	(O)_TVTC	(P)_TVTD	(Q)_CCTVTC	(R)_CCTVTD
1	0	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0	0
3	0	1	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0
10	1	0	0	0	0	0	0	0

- TVTC denotes the interaction between media intervention TV and TimeC.
- TVTD denotes the interaction between TV and TimeD.
- CCTVTC represents the interaction between the CC by TV interaction at the TimeC.
- CCTVTD represents the interaction between the CC by TV interaction at the TimeD.

In all, there were 1556 students included in the analysis of smoking initiation. Of these students, approximately 40% (n = 634) answered yes to the smoking question at one of the three post-intervention time points, while the other 60% (n = 922) either answered no at the last time point or were censored prior to the last time point.

Consider a level-2 model, with schools as the level-2 units. In general, for $i = 1,\ldots,N$ N level-2 units, containing $j = 1,\ldots,n_i$ level-1 units (subjects or multiple failure times) the concept of a censoring or event indicator can be expressed as follows. First, we assume that the assessment time takes on discrete positive values $t = 1,2,\ldots,m$ representing time points or intervals and that each ij unit is observed until time t_{ij}. The censor/event indicator δ_{ij} is coded depending on what happens at time t_{ij}:

- an event occurs $\left(t_{ij} = t \quad and \quad \delta_{ij} = 1\right)$
- the observation is censored $\left(t_{ij} = t \quad and \quad \delta_{ij} = 0\right)$

The term censoring is used when a unit is observed at t_{ij}, but not at $t_{ij} + 1$ (and we know that the event has not occurred up to time t_{ij}).

As mentioned previously, the dichotomous variable EVENT indicated the occurrence of an event. Occurrence of an event was recorded at three time points (WaveB, WaveC, and WaveD), though some subjects dropped out of the study and were not measured at all three time points. To model the time until the event as the outcome variable in a binary analysis of the data, person-time indicators are created (Singer & Willett, 1993). For this, the number of records for each person depends on the timing of the event or censoring for that person. For example, if there were two follow-up points, the two person-time indicators T1 and T2 would be coded as follows:

- T1 = 1: event occurred at T1 (or in interval between T0 and T1)
- T1 = 0: event did not occur at T1 (or in interval between T0 and T1) and T1 was the subject's last measured time point
- T1 = 0 and T2 = 1: event did not occur at T1 but did occur at T2 (or in the interval between T1 and T2)
- T1 = 0 and T2 = 0: individual was censored at T2 (the subject did not experience the event at either T1 or T2)

Note that for the first two scenarios above, subjects would contribute a single record in the data set (for the T1 indicator), whereas they would contribute two records (one for each person-time indicator T1 and T2) for the latter two scenarios. These indicators would represent the dependent variable in the analysis, akin to the variable named EVENT in our TVSFP data.

For the TVSFP data, there were three follow-up occasions, and thus three person-time indicators are necessary to describe the occurrence of event/censoring. The

three person-time indicators form the EVENT variable in the data set, and the timing of the event/censoring is represented by the two variables TimeC and TimeD in the data set. The coding of the person-time indicators (T1, T2, T3) that form the EVENT variable are given in Table 8.1.

Table 8.1: Three time points with censoring

Outcome	Up to 3 records per person
Censor at T1	T1 = 0
Event at T1	T1 = 1
Censor at T2	T1 = 0; T2 = 0
Event at T2	T1 = 0; T2 = 1
Censor at T3	T1 = 0; T2 = 0; T3 = 0
Event at T3	T1 = 0; T2 = 0; T3 = 1

Table 8.2: Coding of time and event indicators for binary TVSFP analysis

EVENT records	Time indicators		Outcome description
	TimeC	TimeD	
T1 =0	0	0	Censor at T1
T1 = 0	0	0	No event at T1
T2 = 0	1	0	Censor at T2
T1 = 0	0	0	No event at T1
T2 = 0	1	0	No event at T2
T3 = 0	0	1	Censor at T3
T1 =1	0	0	Event at T1
T1 = 0	0	0	No event at T1
T2 = 1	1	0	Event at T2
T1 = 0	0	0	No event at T1
T2 = 0	1	0	No event at T2
T3 = 1	0	1	Event at T3

Note that each person would contribute from one to three records in the data set depending on their outcome. For example, for the current data, the EVENT records and their corresponding time indicators are coded as shown in Table 8.2.

The breakdown of cigarette onset for gender and condition subgroups is presented in Table 8.3. Percentages given in the table are calculated relative to the totals for that subgroup at the time of response. At Wave B (post-intervention time point; TimeC = 0 and TimeD = 0), 130 females (SexM = 0) and 156 males (SexM = 1) reported an event (Event = 1), while 105 females and 83 males were censored (Event = 0). These censored subjects did not experience the event at Wave B and were not measured at subsequent waves. The total numbers of females and males that provided data at Wave B were 814 and 742 respectively. The totals at Wave C (TimeC =1) are only 579 and 503 females and males, respectively because the numbers of Wave B event and censored subjects are removed from the Wave C totals. For example, the total number of females at Wave C equals 814 (the number at Wave B) – 130 (females experiencing the event at Wave B) – 105 (censored females at Wave B) = 579. The male total of 503 is obtained in the same way. Of the 579 females, 117 experienced the event at Wave C and 154 were censored at Wave C. Similar calculations for Wave D (TimeD =1) yield the total of 308 females (= 579 – 117 – 154), where 79 females experienced the event and 229 did not and were censored at this last time point. Regarding the differences between males and females, it can be seen that the proportion of males who experienced the event is relatively similar across the three waves. Alternatively, females were initially lower than males (16% versus 21% at Wave B) but increasingly experienced the event across the waves. At the end, the total proportion of males who experienced the event is 41.5% (156 + 89 + 63 of 742), and similarly it is 40.0% for females (130 + 117 + 79 of 814). Thus, the initial gender difference is largely gone by the end of the study.

In terms of the invention groups, the differences do not appear to be very large. If anything, there is some suggestion that control subjects have lower rates of the event, but this difference is not striking.

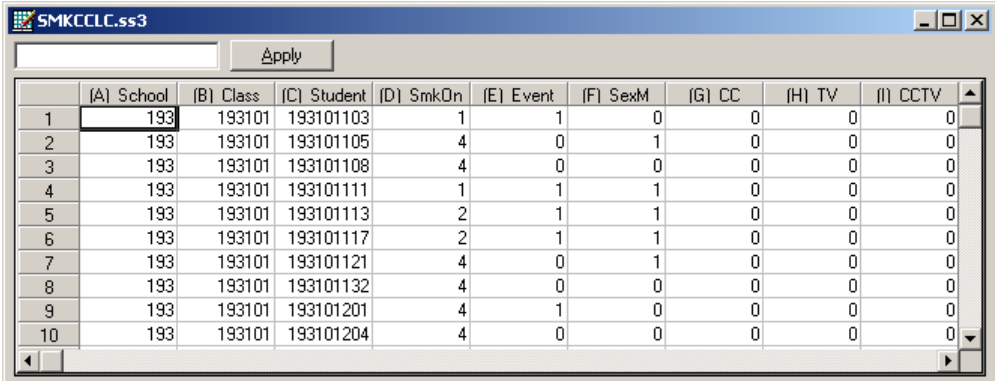
Table 8:3: Onset of cigarette experimentation across three time points

	TimeB			TimeC			TimeD		
	with event	censored	total	with event	censored	total	with event	censored	total
Males	156 (21.0)	83 (11.2)	742	89 (17.7)	134 (26.6)	503	63 (22.5)	217 (77.5)	280
Females	130 (16.0)	105 (12.9)	814	117 (20.2)	154 (26.6)	579	79 (25.6)	229 (74.4)	308
Control	66 (16.5)	60 (15.0)	401	53 (19.3)	69 (25.1)	275	34 (22.2)	119 (77.8)	153
CC only	75 (19.1)	27 (6.9)	392	53 (18.3)	61 (21.0)	290	49 (27.8)	127 (72.2)	176
TV only	71 (17.3)	54 (13.2)	410	60 (21.1)	79 (27.7)	285	38 (26.0)	108 (74.0)	146
CC & TV	74 (21.0)	47 (13.3)	353	40 (17.2)	79 (34.1)	232	21 (18.6)	92 (81.4)	113

In terms of clustering, these 1556 students were from 28 schools with between 13 and 151 students per school ($\bar{n}$ = 56, S.D. = 38) Thus, the data are highly unbalanced with large variation in the number of clustered observations.

8.2.2 The data for an ordinal approach

The ordinal analysis illustrated in this chapter is again based on the TVSFP data. As shown in the previous section, one can also fit grouped-time survival models using dichotomous indicators of event/censoring across the study time points. To do so, requires additional data manipulation. The data set used for the ordinal approach differs from that previously discussed, and is represented by the SuperMix spreadsheet file **SMKCCLC.ss3.** The first 10 records of this data set are shown below.


SMKCCLC.ss3

Apply

	(A) School	(B) Class	(C) Student	(D) SmkOn	(E) Event	(F) SexM	(G) CC	(H) TV	(I) CCTV
1	193	193101	193101103	1	1	0	0	0	0
2	193	193101	193101105	4	0	1	0	0	0
3	193	193101	193101108	4	0	0	0	0	0
4	193	193101	193101111	1	1	1	0	0	0
5	193	193101	193101113	2	1	1	0	0	0
6	193	193101	193101117	2	1	1	0	0	0
7	193	193101	193101121	4	0	1	0	0	0
8	193	193101	193101132	4	0	0	0	0	0
9	193	193101	193101201	4	1	0	0	0	0
10	193	193101	193101204	4	0	0	0	0	0

The variables of interest are:

- School indicates the school a student is from.
- Class identifies the classroom to which a student belongs.
- Student represents the student identification number.
- SmkOnset indicates the time at which an event occurred. It assumes a value of 1 for a WaveA measurement (*i.e.*, the event occurred at Wave A), 2 for a WaveB measurement, 3 for a WaveC measurement, and 4 for a WaveD measurement.
- Event is an indicator variable indicating whether the subject experienced the event or was censored. A value of 1 indicates that the student did experience the event (*i.e.*, onset of cigarette experimentation) at one of the time points, while a value of 0 indicates that the subject was censored and never experienced the event (*i.e.*, no onset of cigarette experimentation) at any time point that they were assessed at.
- SexM is an indicator variable for gender, with "1" indicating male respondents, and "0" female respondents.
- CC is a binary variable indicating whether a social-resistance classroom curriculum was introduced, with 0 indicating "no" and 1 "yes."
- TV is an indicator variable for the use of media (television) intervention, with a "1" indicating the use of media intervention, and "0" the absence thereof.

- CC*TV was constructed by multiplying the variables TV and CC, and represents the CC by TV interaction.

Survival data as ordinal outcomes

Assume 4 time points with no intermittent censoring and let y denote the ordinal outcome variable. Let us first consider subjects who initiated smoking at some point in the study. For these subjects, the variable Event will be coded as 1 and the coding of the SmkOnset variable will be as follows.

SmkOnset:

- $y_{ij} = 1$: Student first started to smoke at $t = 1$.
- $y_{ij} = 2$: Student did not smoke at $t = 1$, but first smoked at $t = 2$.
- $y_{ij} = 3$: Student did not smoke at $t = 1$ or 2, but first smoked at $t = 3$.
- $y_{ij} = 4$: Student did not smoke at $t = 1$, 2, or 3, but first smoked at $t = 4$.

Similarly, subjects who were censored would have the variable Event coded as 0, and the following codes for the SmkOnset variable.

SmkOnset:

- $y_{ij} = 1$: Student did not smoke at $t = 1$ and no data beyond $t = 1$.
- $y_{ij} = 2$: Student did not smoke at $t = 1$ or 2, and no data beyond $t = 2$.
- $y_{ij} = 3$: Student did not smoke at $t = 1$, 2, or 3, and no data beyond $t = 3$ (*i.e.*, no data at $t = 4$).
- $y_{ij} = 4$: Student did not smoke at $t = 1$, 2, 3, or 4.

Here, the phrase "did not smoke" is more precisely "did not answer yes to the question have you ever smoked a cigarette."

Table 8.4: Three time points with censoring

Outcome	Ordinal dep. Variable	Event indicator
Censor at T_1	1	0
Event at T_1	1	1
Censor at T_2	2	0
Event at T_2	2	1
Censor at T_3	3	0
Event at T_3	3	1
Censor at T_4	4	0
Event at T_4	4	1

Table 8.4 shows how values are assigned to y_{ij}, and the relationship between the y_{ij} outcomes and the event indicator.

8.3 The models

8.3.1 Binary case: a 2-level model

In the binary case, the survival time of individual i at occasion j is treated as a set of dichotomous observations indicating whether or not an individual failed in each time unit until a person either experiences the event or is censored. Thus, each survival time is represented as a $t_{ij} \times 1$ vector of zeros for censored individuals, while for individuals experiencing the event the last element of this $t_{ij} \times 1$ vector of zeros is changed to a one. These multiple person-time indicators are then treated as distinct observations in a dichotomous regression model. In the case of clustered data, a random-effects dichotomous regression model is used. This method has been called the pooling of repeated observations method by Cupples (1985). It is particularly useful for handling time-dependent covariates and fitting non-

proportional hazards models because the covariate values can change across each individuals' t_{ij} time points.

For this approach, define p_{ijt} to be the probability of failure in time interval *t*, conditional on survival prior to *t*:

$$p_{ijt} = \Pr\left[t_{ij} = t \mid t_{ij} \geq t \right]$$

Similarly, $1 - p_{ijt}$ is the probability of survival beyond time interval t, conditional on survival prior to *t*. The proportional hazards model is then written as

$$\log\left[-\log\left(1 - p_{ijt}\right) \right] = \alpha_{0t} + \mathbf{x}'_{ijt}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{v}_i$$

and the corresponding proportional odds model is

$$\log\left[p_{ijt} \Big/ \left(1 - p_{ijt}\right) \right] = \alpha_{0t} + \mathbf{x}'_{ijt}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{v}_i$$

where now the covariates **x** can vary across time and so are denoted as $\mathbf{x}_{ijt}$. Augmenting the model intercept, which we will denote α_{01}, the remaining intercept terms $\alpha_{0t}\ (t = 2, \ldots, m)$ are obtained by including as regressors $m-1$ time indicators representing deviations from the first time point. Because the covariate vector **x** now varies with *t*, this approach automatically allows for time-dependent covariates, and relaxing the proportional hazards assumption only involves including interactions of covariates with the $m-1$ time point dummy codes. It is further assumed that the random effects vector has a $\boldsymbol{N}\left(\mathbf{0}, \boldsymbol{\Phi}_{(2)}\right)$ distribution.

In the examples to follow, two random intercept models are fitted to the data described in Section 8.2.2. The type of intervention (CC and/or TV), the gender of the student and the interactions between gender and time (SexTC and SexTD) are included as fixed effects, along with indicators of the time of assessment (TimeC and TimeD).

8.3.2 Ordinal case: 2-level model

Let y_{ij} denote an ordinal outcome variable that takes on discrete positive values $t = 1, 2, \ldots, m$. In previous examples we assumed that y_{ij} has C categories or distinct values, however here to be consistent with the survival analysis notation we will use m to represent the number of ordinal categories. The subscript (i, j) denotes subject j, $j = 1, 2, \ldots n_i$ nested within level-2 unit i, $i = 1, 2, \ldots, N$. In the present context the level-1 units j indicates students and the level-2 unit i indicates schools. Note, that as another example of this type of model, one could have multiple failure times nested within individuals.

Let δ_{ij} denote the censor/event indicator, then $\delta_{ij} = 1$ if the event occurs and $\delta_{ij} = 0$ if an observation is censored. In survival analysis each ij is observed until time t_{ij} and if an event occurs $t_{ij} = t$ and $\delta_{ij} = 1$. If the observation is censored at $t_{ij} = t$ then $\delta_{ij} = 0$.

In the case of censoring it is assumed that a unit is observed at t_{ij} but not at t_{ij+1}. As described in Hedeker, Siddiqui & Hu (2000), if events occur within continuous time intervals (*i.e.*, grouped-time), for example, a student initiates smoking experimentation in the past year, use of the complementary log-log link for an ordinal outcome is equivalent to a proportional hazards model in continuous time. Therefore, the grouped-time proportional hazards mixed model can be written as:

$$\log\left[-\log\left(1 - P_{ijt}\right)\right] = \gamma_t + \mathbf{x}_{ij}'\boldsymbol{\beta} + \mathbf{z}_{ij}'\mathbf{v}_i$$

where $\mathbf{x}_{ij}$ is a vector of explanatory variables and $\mathbf{z}_{ij}$ a vector of random effects. Typically, the elements of $\mathbf{z}_{ij}$ are a subset of $\mathbf{x}_{ij}$. For example, the elements of $\mathbf{z}_{ij}$ might correspond to the intercept and age, whereas $\mathbf{x}_{ij}$ would include these two terms plus any additional model covariates. It is assumed that the random effects $\mathbf{v}_i$ are from a normal distribution with mean zero and covariance matrix $\boldsymbol{\Phi}_{(2)}$.

P_{ijt} denotes the probability that an event takes place up to and including the interval designated at time t_{ij}. Thus, P_{ijt} represents a cumulative probability of failure, whereas p_{ijt} is the interval-specific failure probability. Also, γ_t represent threshold values, and in the present context these reflect the baseline hazard (*i.e.*, the hazard when all covariates equal 0). These threshold parameters are akin to the intercept parameters α_{0t} in the dichotomous version of the model. The plus sign following γ_t means that a positive regression coefficient for a covariate indicates an increased hazard (*i.e.*, the event occurs sooner) as values of the covariate increase.

8.4 Example: A proportional hazards model- Binary case

8.4.1 Introduction

The first model fitted to the data will use the binary case and is of the form

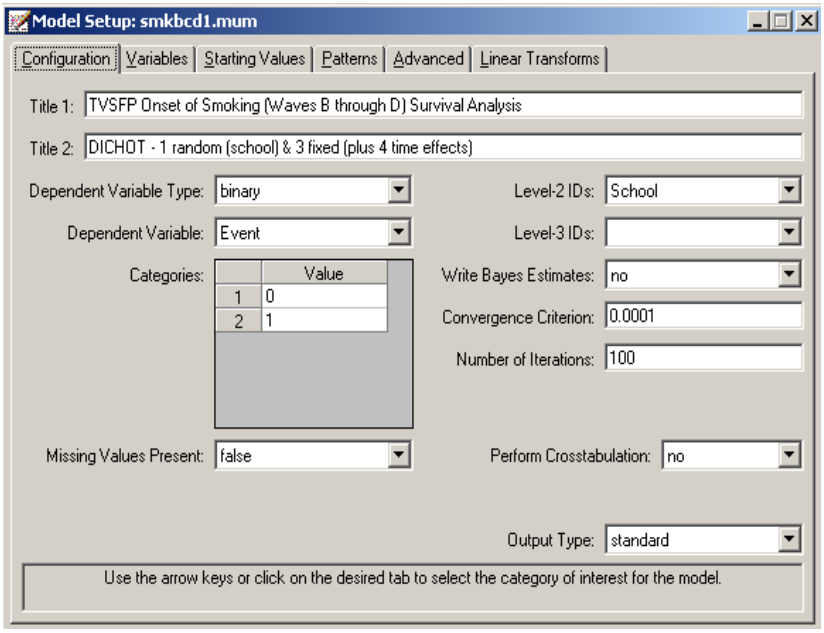
$$\log\left[-\log\left(1-p_{ijt}\right)\right] = \alpha_{01} + (TimeC_{ij})\alpha_{02} + (TimeD_{ij})\alpha_{03} + (SexM_{ij})\beta_1 + (CC_j)\beta_2 + (TV_j)\beta_3 + v_{0i}.$$

In the current model specification, the baseline hazard is a function of the model intercept and the coefficients for the time indicators. Specifically, the baseline hazard estimate at the first time point equals the estimated model intercept, the baseline hazard estimate at the second time point is the sum of the model intercept and the estimated coefficient for the TimeC indicator, the baseline hazard at the third time point is the sum of the model intercept and the estimated coefficient for the TimeD indicator. Thus, two of these baseline hazard estimates involve sums of the estimated parameters.

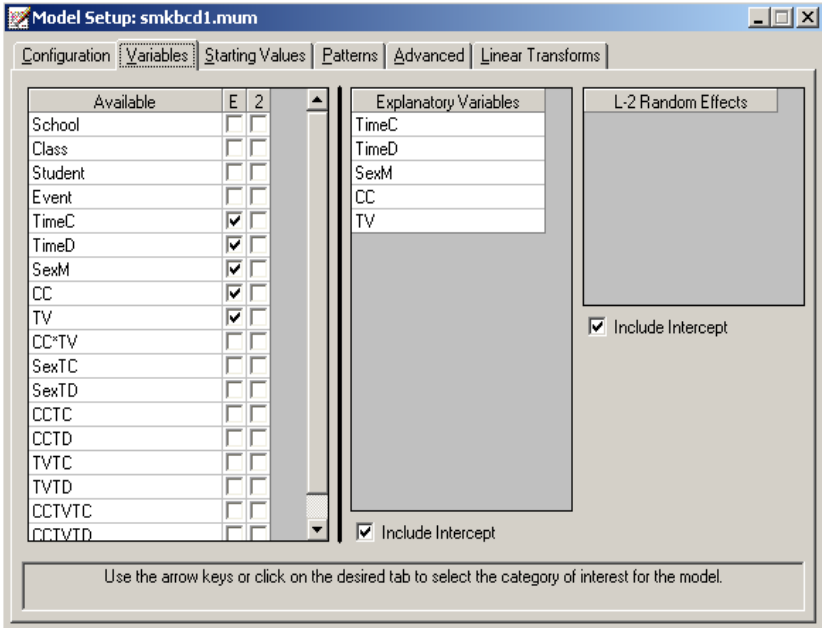
8.4.2 Setting up the analysis

Start by selecting the **New Model Setup** option on the **File** menu to open the **Model Setup** window. Enter (optional) titles in the **Title 1** and **Title 2** text boxes. Select the

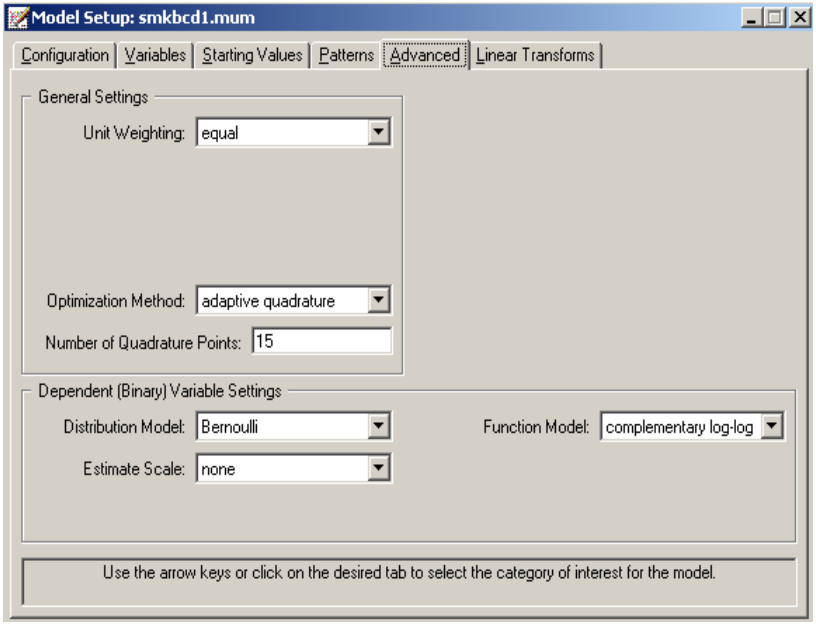
binary outcome variable Event from the **Dependent Variable** drop-down list box. The variable School, which defines the units within which students are nested, is selected as the Level-2 ID from the **Level-2 IDs** drop-down list box.



Next, click on the **Variables** tab of the **Model Setup** window. TimeC, TimeD, SexM, CC, and TV are specified as the predictors (explanatory variables) of the fixed part of the model by checking the corresponding boxes in the **E** column of the **Available** grid on the **Variables** screen. By default, it is assumed that the intercept is allowed to vary randomly over the level-2 units (*i.e.*, the schools), as indicated by the checked box in the **Include Intercept** field.



To specify the number of quadrature points and link function (**Function Model**), we proceed to the **Advanced** screen by clicking on the **Advanced** tab. Select **complementary log-log** as the **Function Model** in order to yield the proportional hazards model. Note that the default **Number of Quadrature Points** of **10** is replaced by **25**. Here, more quadrature points are used because it is thought that the School effect on the student outcomes (*i.e.*, the clustering effect) is likely to be small, resulting in a near-zero random effect variance parameter. In such cases, for computational purposes it is beneficial to use a relatively larger number of quadrature points.

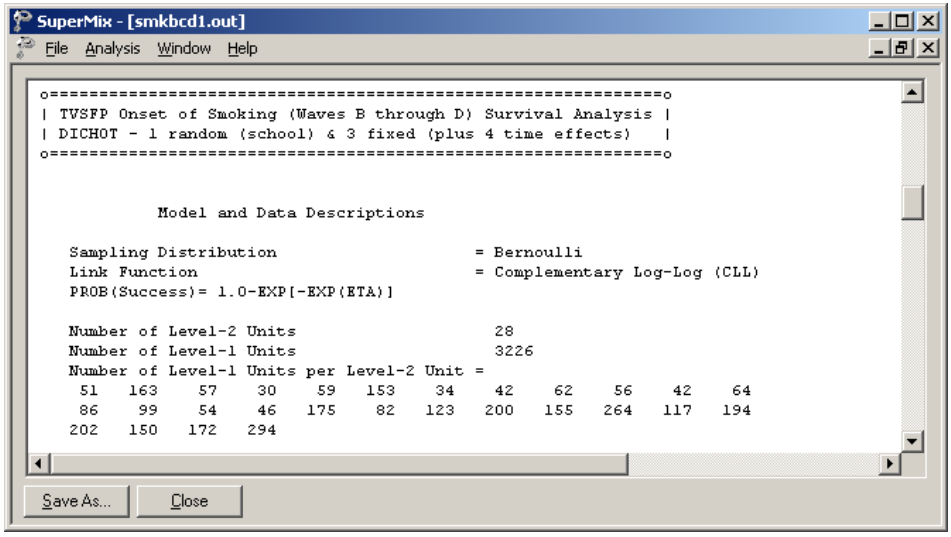


This step completes the model set-up. Use the **File**, **Save** option to save the model setup to a file named **smkbcd1.mum**. Next, use the **Analysis**, **Run** option on the main menu bar to run the analysis.

8.4.3 Discussion of results

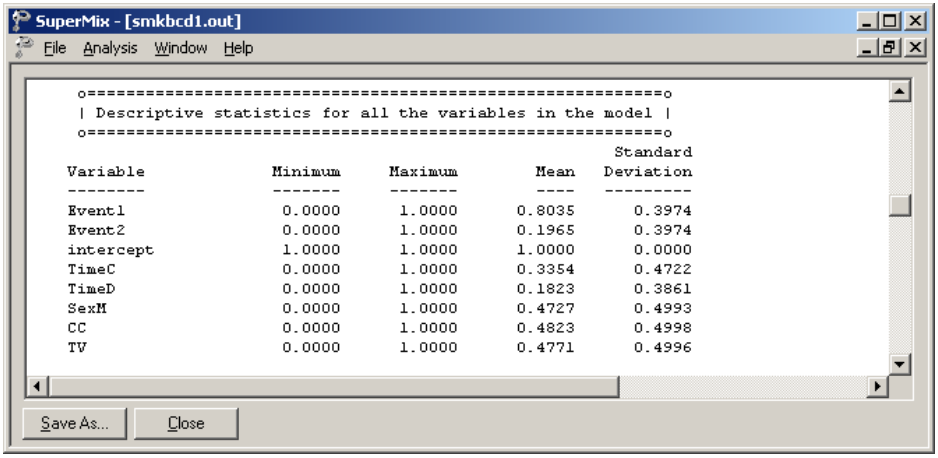
Data summary

The portion of the output file shown below indicates that there are 28 schools. Nested within these level-2 units are 3226 measurements (note: this is not equal; to the number of students because of the creation of person-time indicators in this binary version of the survival analysis model). A summary of the number of level-1 observations per level-2 unit is also given.



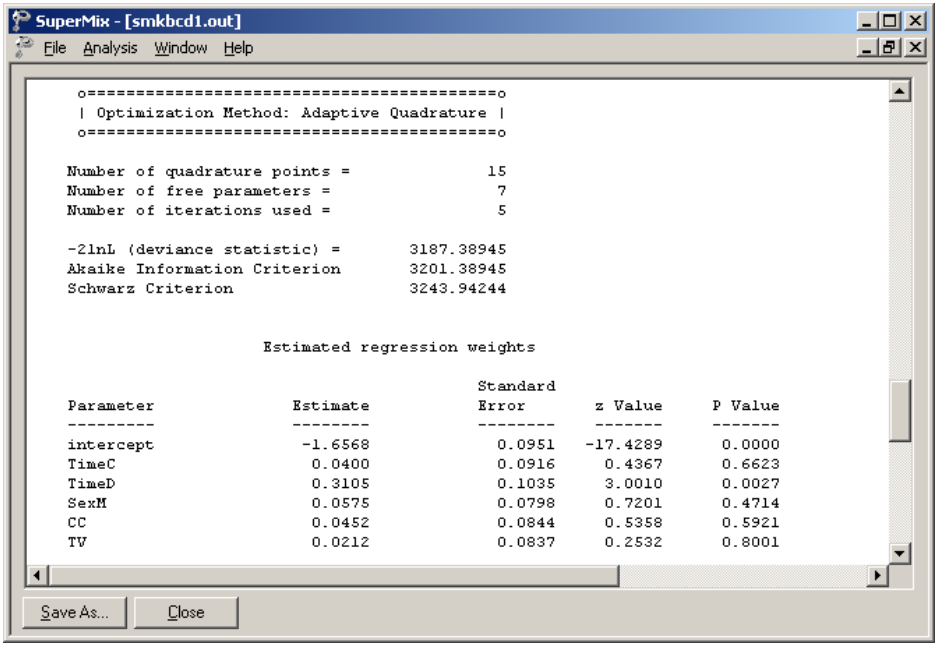
Descriptive statistics

This is followed by descriptive statistics for all the variables. Except for the intercept term, the variables are all dichotomous. The proportions of subjects assigned a value of 0 or 1 are 0.80347 and 0.19653 respectively. In approximately 20% of the person-time indicators, an event occurred.



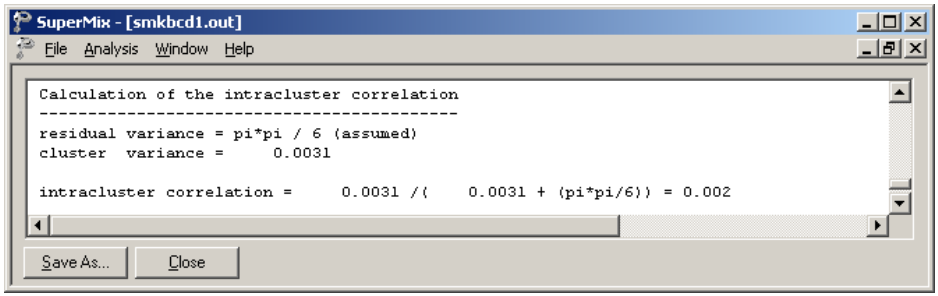
Fixed effects estimates

Parameter estimates are given in the next part of the output. The effect of SexM is positive and indicates that boys have a slightly, but non-significant, increased hazard (*i.e.*, a shorter time to the first occurrence), relative to girls. The coefficients associated with the TimeD indicator variable is significant at a 5% level. In contrast, the corresponding TimeC coefficient is not significant. These indicate that the baseline hazard does not significantly change between Waves B and C, however there is significant change between Waves B and D as relatively more students experiment with smoking at Wave D. Finally, the effects of the intervention variables CC and TV are not seen to be statistically significant, though the direction of their effects is positive (*i.e.*, increased hazard relative to the control group).



Intraclass correlation (ICC)

The last part of the output contains an estimate of the intracluster correlation. This estimate indicates a very modest school effect, and we also note that the random effect variance term is not significant. From this, we conclude that the time until the occurrence of an event does not vary significantly across schools. However, from a design point of view, because schools were randomized to the intervention conditions in this study, one can argue that the clustering attributable to schools is an important part of the model regardless of its significance.



8.4.4 Interpreting the output

Estimated unit-specific probabilities

We now use the estimated coefficients from the fitted model

$$\log\left[-\log(1-\hat{p}_{ijt}\right] = \hat{\alpha}_{01} + (TimeC_{ij})\hat{\alpha}_{02} + (TimeD_{ij})\hat{\alpha}_{03} + (SexM_{ij})\hat{\beta}_1 + (CC_j)\hat{\beta}_2 + (TV_j)\hat{\beta}_3$$

$$= -1.6564 + (TimeC_{ij})0.0399 + (TimeD_{ij})0.3103 + (SexM_{ij})0.0574$$

$$+(CC_j)0.0449 + (TV_j)0.0213$$

and the inverse cumulative log-log link function

$$P(z) = 1 - \exp[-\exp(z)]$$

to calculate the probability of Event = 1 at various time points and for different covariate values.

At the first time point (Wave B), $TimeC_{ij} = TimeD_{ij} = 0$, and thus the relevant part of the fitted model (see above) is

$$\log\left[-\log(1-\hat{p}_{ijt})\right] = \hat{\alpha}_{01} + (SexM_{ij})\hat{\beta}_1 + (CC_j)\hat{\beta}_2 + (TV_j)\hat{\beta}_3$$
$$= -1.6564 + (SexM_{ij})0.0574 + (CC_j)0.0449 + (TV_j)0.0213$$

For female students (SexM = 0) from the control group (CC = TV = 0) the probability of smoking experimentation (Event = 1) at the point of post-intervention can be expressed as

$$P(Event = 1\, at\, WaveB, female) = 1 - \exp\left[-\exp(-1.6564)\right]$$
$$= 0.1737.$$

For male students in the control group adding the intercept with the SexM estimate together yields $z = -1.6564 + 0.0574 = -1.599$, and so

$$P(Event = 1\, at\, WaveB, male) = 1 - \exp\left[-\exp(-1.599)\right]$$
$$= .1830.$$

Results for all groups are summarized in Table 8.5. The probability of smoking experimentation at the time of post-intervention is larger for males than for females. The results also indicate an increased probability of failure with an increase of time. In the current model, it is assumed that the ratio of the estimated hazards over time will be constant for two individuals with the same values on the covariates. To check whether the effect of gender is dependent on time, and thus to check on the proportional hazards assumption, interactions with time indicators should be included in the model.

Table 8.5: Unit-specific probabilities for groups

Gender	CC	TV	WaveB (TimeC = 0, TimeD = 0)	WaveC (TimeC = 1, TimeD = 0)	WaveD (TimeC = 0, TimeD = 1)
Female	0	0	0.1737	0.1801	0.2291
	1	0	0.1809	0.1876	0.2383
	0	1	0.1771	0.1836	0.2335
	1	1	0.1844	0.1912	0.2428
Male	0	0	0.1830	0.1897	0.2409
	1	0	0.1905	0.1975	0.2505
	0	1	0.1865	0.1933	0.2454
	1	1	0.1942	0.2012	0.2551

Table 8.6 shows the differences between the estimated unit-specific probabilities and the observed proportions for each of the 24 subgroups formed by crossing all predictors currently in the model.

Looking at the direction of the differences, we note that for females all the estimated probabilities are larger in size than the observed ratios at WaveB, but consistently lower than the observed ratios at the next two time points, with the exception of the situation where TimeD = CC = TV = 1. It seems as if the model is overestimating the probabilities of failure at the first time point, but underestimating probabilities at the last time of measurement. However, the pattern for males is almost the opposite. At the first wave, only one estimated probability is larger than the observed proportion, at WaveC this is true for 2 of the four cells, and at WaveD for three of the four cells.

Table 8.6: Differences between unit-specific probabilities and observed proportions

Gender	CC	TV	Difference at WaveB (estimated – observed)	Difference at WaveC (estimated – observed)	Difference at WaveD (estimated – observed)
Female	0	0	0.0227	–0.0419	–0.0179
	1	0	0.0149	0.0016	–0.0117
	0	1	0.0091	–0.0174	–0.0875
	1	1	0.0204	–0.0058	0.0568
Male	0	0	–0.0150	–0.0073	–0.0361
	1	0	0.0165	–0.0025	0.0625
	0	1	–0.0275	0.0303	0.0064
	1	1	–0.0678	0.0613	0.0710

This trend could be the result of a gender effect (which we know to be non-significant in the current model) or from an interaction between gender and time. While only TimeD had a significant estimated coefficient, this apparent trend leads us to conclude that testing of the assumption of proportional hazards is appropriate. Specifically, the interaction between gender and the time of measurement will be explored.

Estimated population-average probabilities

Table 8.5 contains the estimated unit (school) specific probabilities. To obtain population-average probabilities, the estimated z_{ij} – values are divided by the square root of the design effect. For the current example, we know that $\hat{\sigma}_v^2 = 0.0028$, and that the underlying variance (*i.e.*, level-1 variance) associated with the complementary log-log link is $\sigma^2 = \pi^2 / 6$ (see Agresti, 2002, pp. 248-250). The design effect follows as

$$\hat{d} = \frac{0.0028 + \pi^2/6}{\pi^2/6} = 1.0017.$$

Since $\hat{d} \approx 1.0$, the estimated population-average probabilities for this model would thus be interchangeable with the unit-specific probabilities.

8.5 Example: Checking the proportional hazards assumption in a binary model

8.5.1 Introduction

In a proportional hazards model such as the model fitted previously, it is assumed that the hazard function for an observation in the analysis depends on the values of the covariates and the value of the baseline hazard. This implies that the ratio of the estimated hazards over time will be constant for two individuals with the same values on the covariates. To test the validity of this assumption using the current data, interactions with time indicators are included in the model. Doing so allows us to check whether the impact of the covariates in the model are dependent on time.

The model fitted to the data is of the form

$$\log\left[-\log(1-p_{ijt})\right] = \alpha_{01} + (TimeC_{ij})\alpha_{02} + (TimeD_{ij})\alpha_{03} + (SexM_{ij})\beta_1 + (CC_j)\beta_2 + (TV_j)\beta_3 + (SexTC_{ij})\beta_4 + (SexTD_{ij})\beta_5 + v_{0i},$$

and includes two interaction terms: SexTC represents the SexM by TimeC interaction, while SexTD represents the SexM by TimeD interaction. Thus, in this model, β_1 represents the gender effect at Wave B, while β_4 and β_5 indicate how the gender effect varies at Waves C and D, respectively, relative to Wave B. Linear transforms will be used to obtain the specific gender effects at Wave C ($\beta_1 + \beta_4$) and Wave D ($\beta_1 + \beta_5$). The baseline hazard would be as shown in Table 8.7, while the two linear transforms used in the model are described in Table 8.8.

Table 8.7: Definition of baseline hazard

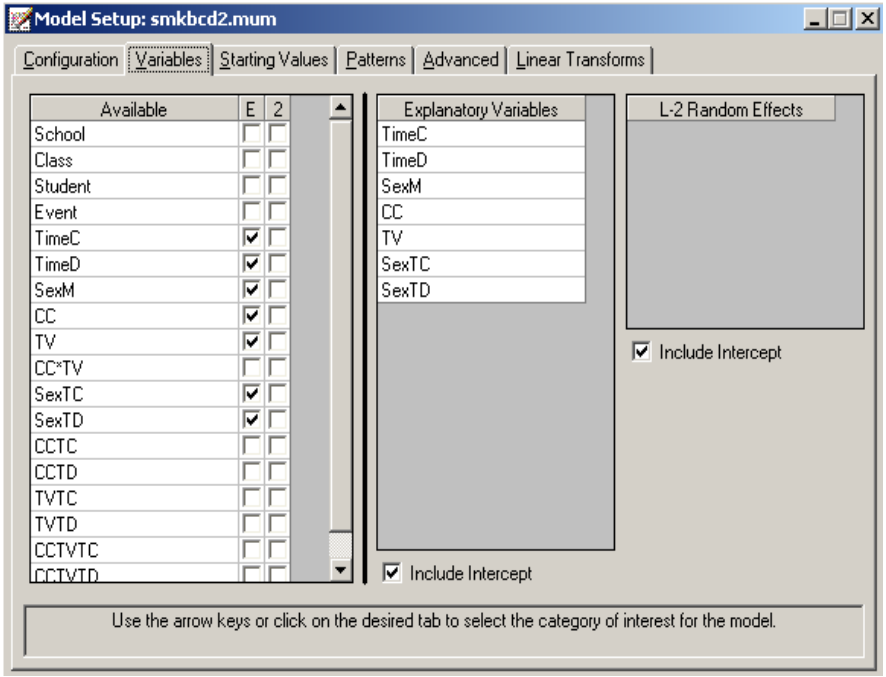
Intercpt	TimeC	TimeD
1 1	1 0	0 1

Table 8.8: Description of linear transforms

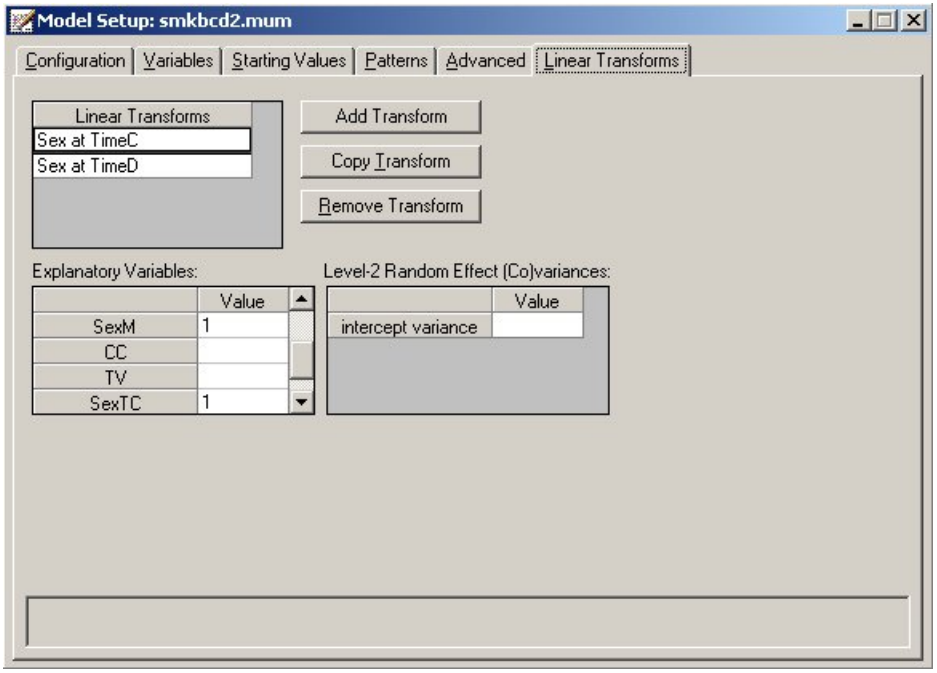
Intercpt	TimeC	TimeD	SexM	CC	TV	SexTC	SexTD
0 0	0 0	0 0	1 1	0 0	0 0	1 0	0 1

8.5.2 Setting up the analysis

Using the same data as in the previous example, start by selecting the **Open Existing Model Setup** option on the **File** menu to open the model setup file named **smkbcd1.mum**. Next, click on the **Variables** tab and add SexTC and SexTD to the list of predictors by checking the corresponding boxes in the **E** column of the **Available** grid on the **Variables** screen.



To complete the model setup, we use the **Linear Transforms** option to enter the information given in Table 8.8. This will provide estimates of the gender effect at Waves C and D. The screen below shows the values entered for the first transform. To enter the first linear transform, click **Add Transform** and enter the name of the transform, in this case Sex at TimeC in the **Linear transform** text field. Next, enter the value 1 next to the variables SexM and SexTC in the **Explanatory Variables** field. The screen below shows the values entered for the first transform.



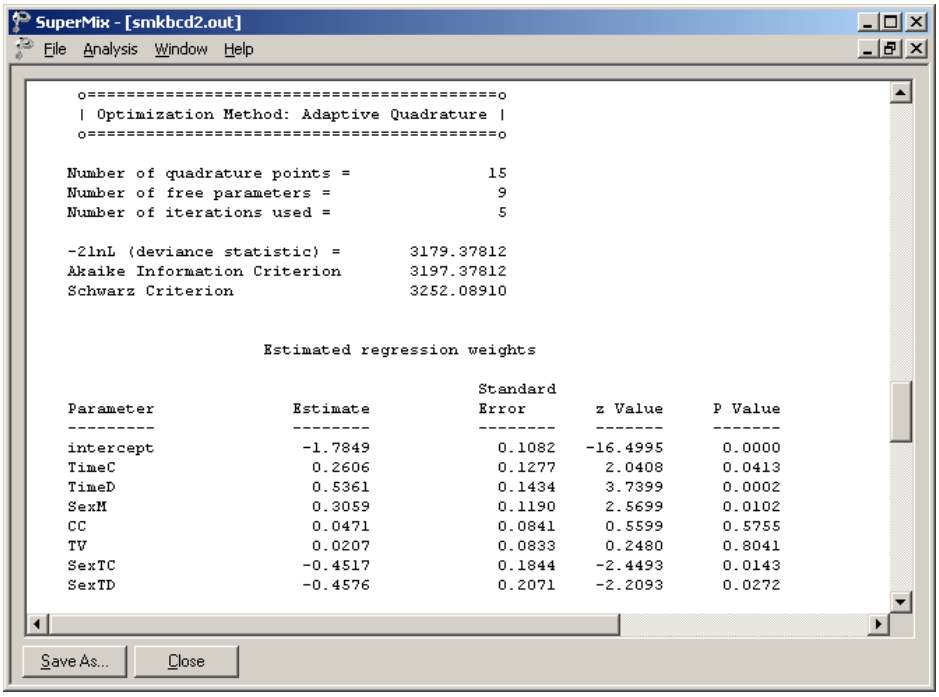
Values for the second transform are entered in the same way. All other input remains the same. Use the **File**, **Save** option to save the model setup to a file named **smkbcd2.mum**. Next, use the **Analysis**, **Run** option on the main menu bar to run the analysis.

8.5.3 Discussion of results

Fixed effects estimates

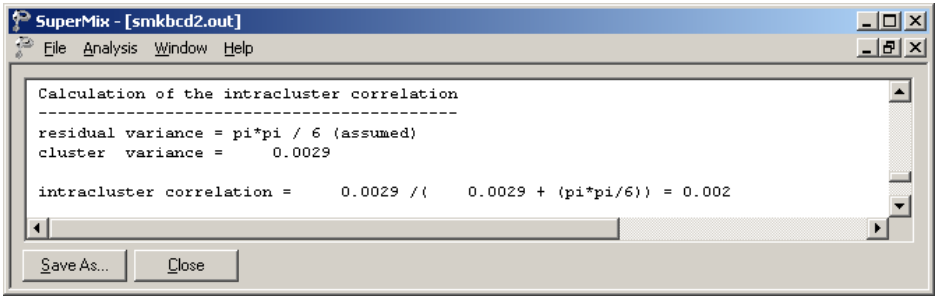
Parameter estimates are given in the next part of the output. The effect of SexM is positive and highly significant, indicating that boys have a significantly increased hazard (*i.e.*, a shorter time to the first occurrence), relative to girls at Wave B (*i.e.*, the post-intervention time point). The coefficients associated with the TimeC and TimeD indicator variables and the interaction terms SexTC and SexTD are also significant at a 5% level. The latter two indicate that the gender difference at Waves C and D, respectively, are different than the gender difference at Wave B. Recall the

deviance statistic for the first model was 3187.20. The addition of the two predictors SexTC and SexTD have led to a decrease of 8 in this statistic, at the cost of predicting an additional 2 parameters. This χ^2 statistic is significant at a 5% level, and we conclude that the addition of the interaction terms have contributed significantly to the overall explanation of variation in the outcome variable. Thus, the proportional hazards assumption is rejected for the gender effect.



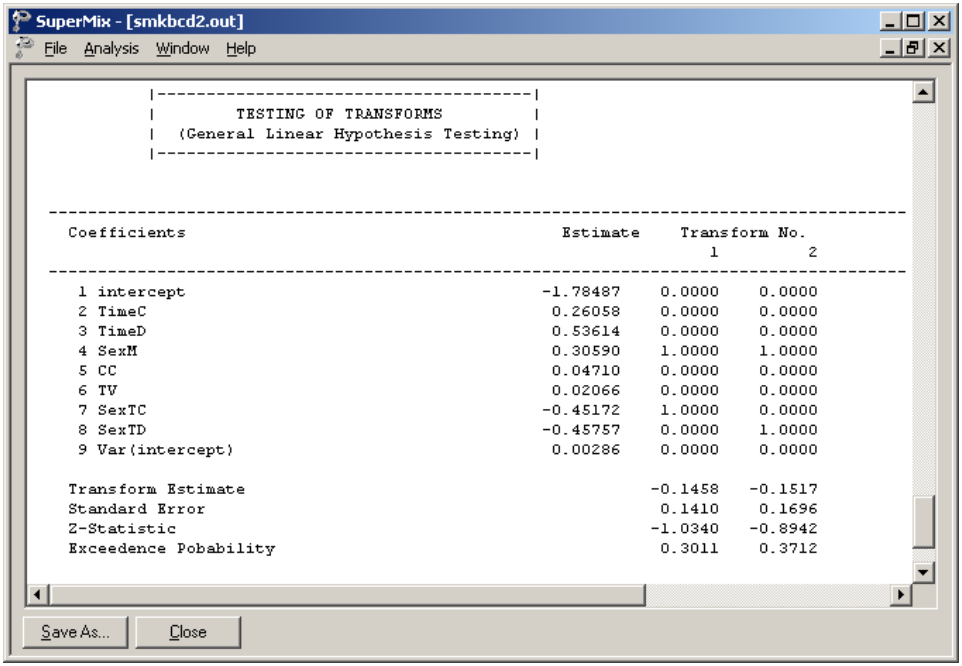
Intraclass correlation (ICC) and transforms

The last part of the output contains an estimate of the intracluster correlation. We see little change here compared to the results of the model without interaction terms.



Finally, a summary of the transforms (given in transposed form) is given followed by a significance test for each transform. These two transforms indicate the gender effect at Waves C and D, respectively. Notice that neither is significant. Thus, whereas there was a significant gender effect at Wave B, with boys having increased hazard of cigarette experimentation, this difference is no longer significant at the subsequent Waves.

In combination with the intercept and time indicator estimates, these provide estimates of the hazard. Specifically, the hazard estimates for the three study time points for girls (*i.e.*, when SexM = 0) are –1.7841, –1.7841 + 0.2605, –1.7841 + 0.5357, and –1.7841 + 0.3058, –1.7841 + 0.2605 – 0.1461, –1.7841 + 0.5357 – 0.1518 for boys (*i.e.*, when SexM = 1). As shown in the next section, these can be converted to the probability scale using the inverse of the complementary log-log function.



8.5.4 Interpreting the output

Estimated unit-specific probabilities

We now use the estimated coefficients from the fitted model

$$\log\left[-\log(1-\hat{p}_{ijt})\right] = \hat{\alpha}_{01}\hat{\alpha}_{01} + (TimeC_{ij})\hat{\alpha}_{02} + (TimeD_{ij})\hat{\alpha}_{03} + (SexM_{ij})\hat{\beta}_1 + (CC_j)\hat{\beta}_2 + (TV_j)\hat{\beta}_3$$

$$+(SexTC_{ij})\hat{\beta}_4 + (SexTD_{ij})\hat{\beta}_5$$

$$= -1.7841 + (TimeC_{ij})0.2605 + (TimeD_{ij})0.5357 + (SexM_{ij})0.3058$$

$$+(CC_j)0.0465 + (TV_j)0.0209 - (SexTC_{ij})0.4518 - (SexTD_{ij})0.4576$$

and the inverse cumulative log-log link function $P(z)=1-\exp[-\exp(z)]$ to calculate the probability of smoking experimentation across the three waves for boys and girls.

In order to calculate the probabilities, we set the values of CC and TV to the mean values as observed in the sample, *i.e.* 0.4823 and 0.4771 respectively. Note that these values can be found in the descriptive statistics section of the output file. Alternatively, if we had not done so, but set these predictors to zero, this would have implied that all estimated probabilities were for the groups where CC = TV = 0 (*i.e.*, the control group).

We again start by calculating the probabilities at Wave B (post-intervention). For all respondents, this implies that $TimeC_{ij} = TimeD_{ij} = 0$, and thus the relevant part of the fitted model (see above) is

$$\log\left[-\log(1-\hat{p}_{ijt})\right] = -1.7841 + (SexM_{ij})0.3058 + (0.4823)0.0465 + (0.4771)0.0209$$
$$= -1.7517 + (SexM_{ij})0.3058$$

For female students (SexM = 0) the probability of smoking experimentation at the point of post-intervention can be expressed as

$$P(Event = 1\, at\, WaveB, female) = 1-\exp\left[-\exp(-1.7517)\right]$$
$$= 0.1593.$$

and for male students

$$P(Event = 1\, at\, WaveB, male) = 1-\exp\left[-\exp(-1.7517+0.3058)\right]$$
$$= 0.2099.$$

Results for all waves are summarized in Table 8.9.

The probability of smoking experimentation at the time of post-intervention is larger for males than for females. This was reflected by the significant main effect of SexM in the analysis. However, this gender difference changes across time, as indicated by

the significant gender by time interaction terms, as females exhibit relatively higher rates of smoking experimentation at the latter two waves.

Table 8.9: Unit-specific probabilities for gender groups across waves

Gender	WaveB (TimeC = 0, TimeD = 0)	WaveC (TimeC = 1, TimeD = 0)	WaveD (TimeC = 0, TimeD = 1)
Female (SexM = 0)	0.1593	0.2016	0.2565
Male (SexM = 1)	0.2099	0.1768	0.2248

Estimated population-average probabilities

Table 8.9 contains estimated unit (school) specific probabilities. These are sometimes referred to as conditional estimates, conditional on the school effects. In other words, they are estimates controlling for the effect of school on the individual student outcomes. To obtain population-average probabilities, adjusted z_{ij} – values are used in the computation of the probabilities.

For the current example the design effect is equal to

$$\hat{d} = \frac{0.0023 + \pi^2 / 6}{\pi^2 / 6} = 1.0014.$$

The estimated population-average probabilities are obtained in a similar fashion as the unit-specific probabilities, but with replacing $\hat{z}_{ik}$ by $\hat{z}_{ik} / \sqrt{d_{ik}}$. For this example, due to the fact that $\hat{d} \approx 1.000$, the estimated unit-specific and population-average probabilities are, for all purposes, identical.

Table 8.10 shows the estimated population-average probabilities for all of the 24 subgroups. These probabilities were calculated using the observed data values on all included predictors.

Table 8.10: Population-average probabilities for all groups

Gender	CC	TV	WaveB (TimeC = 0, TimeD = 0)	WaveC (TimeC = 1, TimeD = 0)	WaveD (TimeC = 0, TimeD = 1)
Female	0	0	0.1547	0.1960	0.2496
	1	0	0.1615	0.2043	0.2598
	0	1	0.1577	0.1997	0.2542
	1	1	0.1646	0.2081	0.2645
Male	0	0	0.2040	0.1718	0.2187
	1	0	0.2126	0.1792	0.2278
	0	1	0.2079	0.1751	0.2227
	1	1	0.2166	0.1826	0.2320

With the interaction terms included in the model, the trend in the differences between the estimated probabilities and observed proportions have disappeared to a large extent. The differences between estimated probabilities and observed proportions are slightly smaller for the larger model when results of Tables 8.11 and 8.6 are compared. We conclude that there is evidence of an interaction between the gender of respondents and the time of measurement, and that it would not be appropriate for these data to assume that the ratio of the estimated hazards over time will be constant for the two gender groups.

Table 8.11: Difference between estimated probabilities and observed proportions of failure for all subgroups

Gender	CC	TV	Difference at WaveB (estimated – observed)	Difference at WaveC (estimated – observed)	Difference at WaveD (estimated – observed)
Female	0	0	–0.0012	0.0100	–0.0004
	1	0	–0.0065	0.0033	–0.06117
	0	1	0.0068	–0.0223	0.0072
	1	1	0.0006	0.0111	0.0785
Male	0	0	0.0307	–0.0282	0.0307
	1	0	–0.0113	0.01622	–0.0112
	0	1	0.0099	–0.0219	–0.0542
	1	1	–0.0454	0.0426	0.0470

8.6 Example: Survival analysis model for an ordinal outcome

8.6.1 Introduction

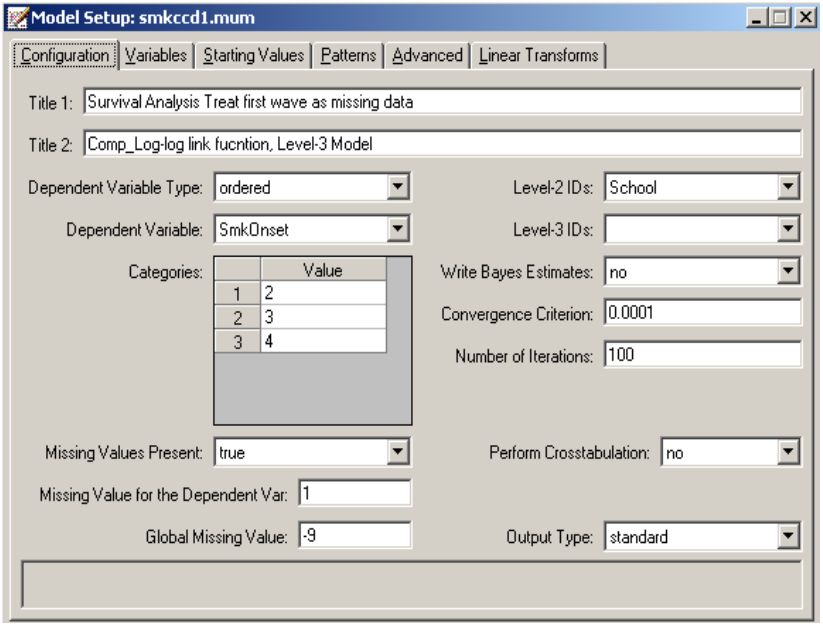
In this section, the re-formatted form of the data, as captured in **smkcclc.ss3** is used to fit a model to the data with the ordinal variable SmkOnset as outcome.

The model fitted to the data is of the form

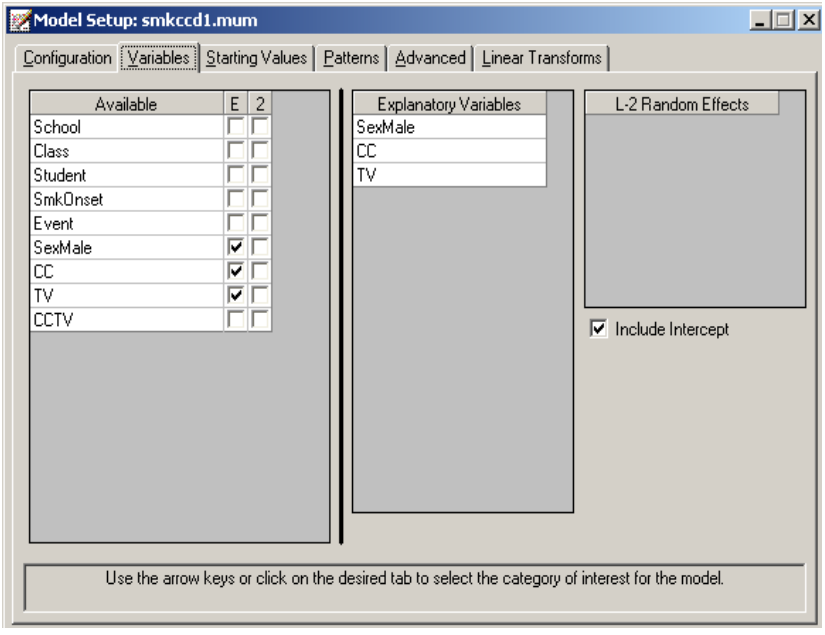
$$\log\left[-\log(1-P_{ijt})\right] = \gamma_t + (SexM_{ij})\beta_1 + (CC_j)\beta_2 + (TV_j)\beta_3 + v_{0i}.$$

8.6.2 Setting up the analysis

Using the data in the SuperMix spreadsheet **SMKCCLC.ss3**, we start by selecting the **New Model Setup** option on the **File** menu to open the **Model Setup** window. Enter (optional) titles in the **Title 1** and **Title 2** text boxes. Select the ordinal outcome variable SmkOnset from the **Dependent Variable** drop-down list box. Note that when the variable is selected, the **Categories** field is populated with values 1 through 4. In these data, the value "1" represents missing data because this value indicates failure or censoring at Wave A (*i.e.*, the pre-intervention time point). As previously noted, the intent was to focus on the post-intervention time points only (*i.e.*, Waves B, C, and D). Indicate this by setting the **Missing Values Present** field to true, and entering the value "1" in the **Missing value for the Dependent Var** field. The **Categories** field now shows the remaining three categories only. The variable School, which defines the units within which students are nested, is selected as the Level-2 ID from the **Level-2 IDs** drop-down list box. The completed dialog box is shown below.

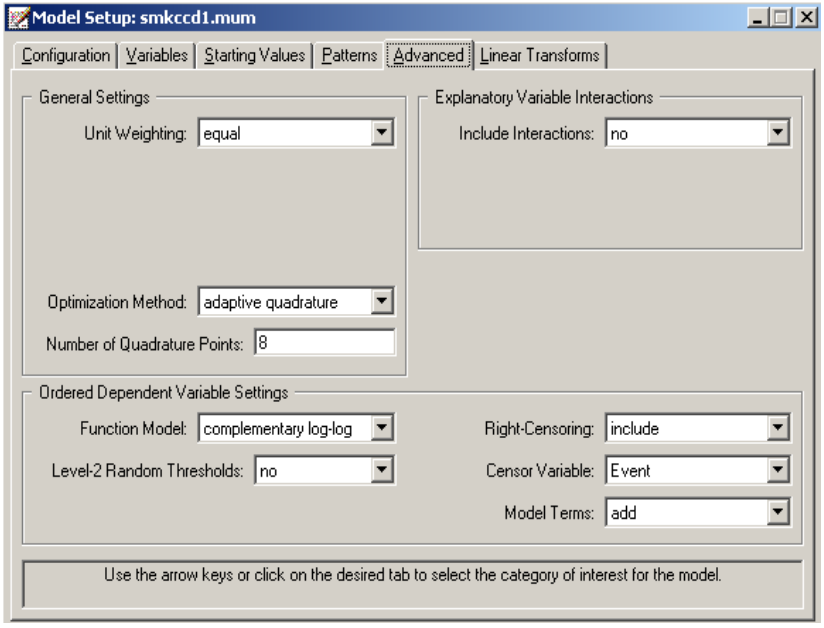


Click on the **Variables** tab of the **Model Setup** window. SexM, CC, and TV are specified as the predictors (explanatory variables) of the fixed part of the model by checking the corresponding boxes in the **E** column of the **Available** grid on the **Variables** screen. By default, it is assumed that the intercept is allowed to vary randomly over the level-2 units, as indicated by the checked box in the **Include Intercept** field.



To specify the number of quadrature points, link function (**Function Model**), and right censoring, we proceed to the **Advanced** screen by clicking on the **Advanced** tab. Change **Model Terms** from subtract to add (so that the model terms are added to the thresholds as specified in the ordinal version of the survival analysis model) and select **complementary log-log** as the **Function Model** (to yield a proportional hazards model). Note that the default **Number of Quadrature Points** of **10** is replaced by **8**. Only 8 quadrature points were used here since the values of the estimated parameters and $-2\ln L$ statistic remain unchanged, up to 5 decimal places, for this or a larger number of quadrature points. Finally, we indicate that **Right Censoring** is

to be included and that the variable for this is Event (which is coded 0 = censor and 1 = event).



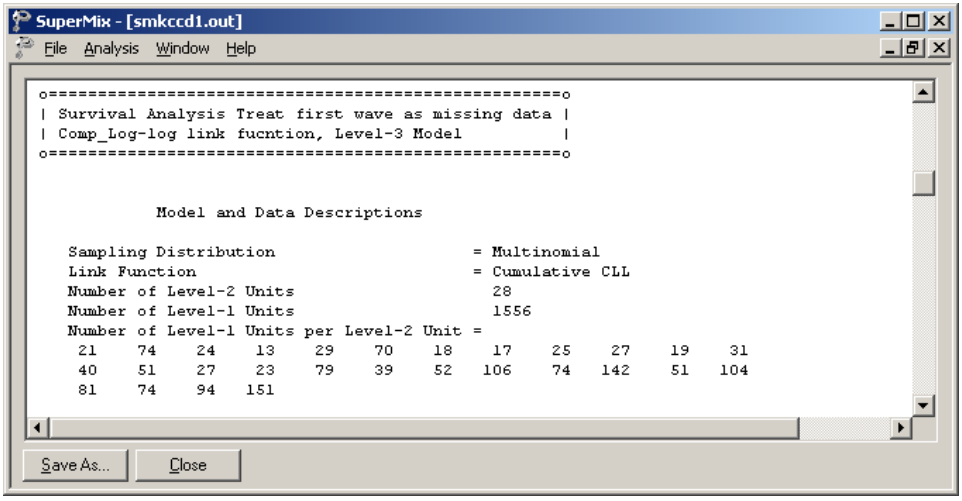
Use the **File**, **Save** option to save the model setup to a file named **smkccd1.mum**. Next, use the **Analysis**, **Run** option on the main menu bar to run the analysis.

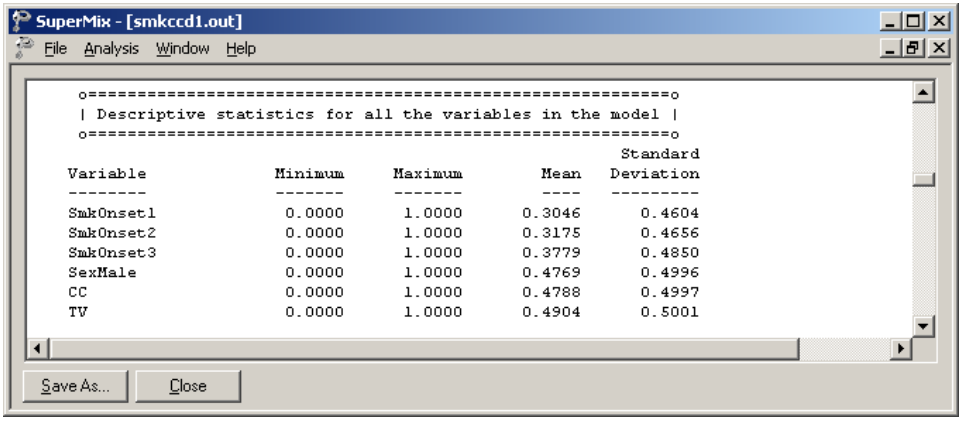
8.6.3 Discussion of results

Selected portions of the output file **smkccd1.out** are shown below.

Data summary and descriptive statistics

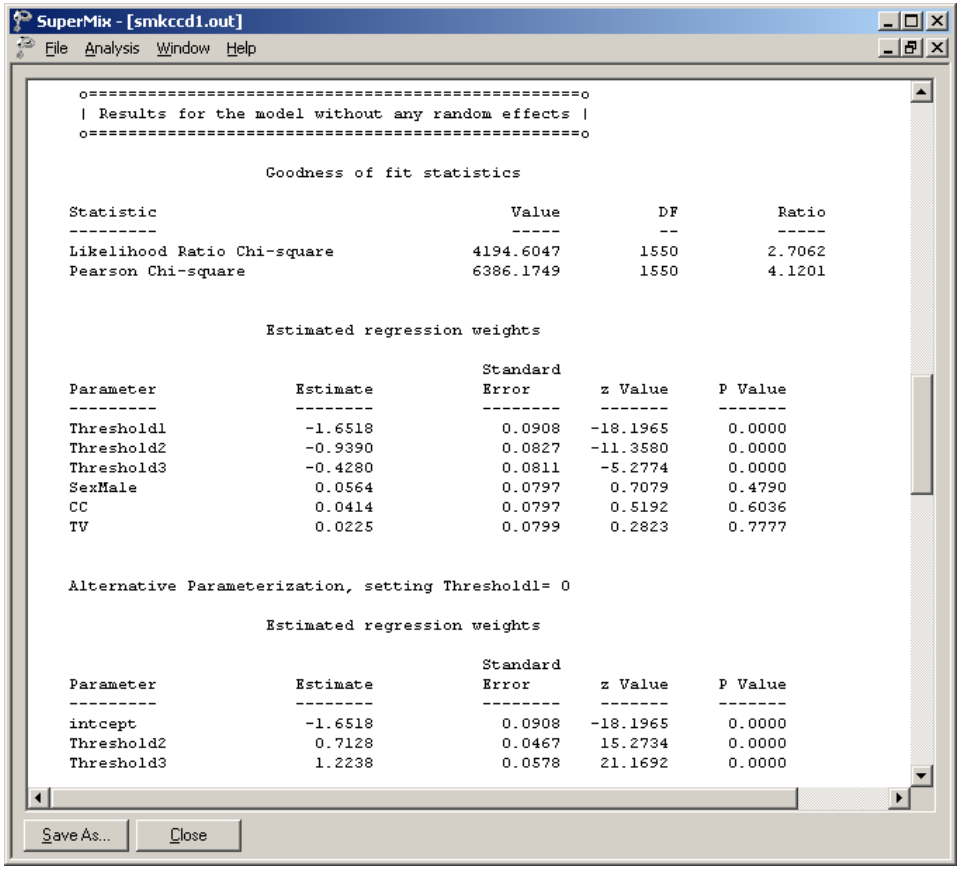
The portion of the output file shown below indicates that there are 28 schools, with 1556 students nested within these. This is followed by descriptive statistics for all the variables. Note that all three predictor variables are dichotomous in nature.

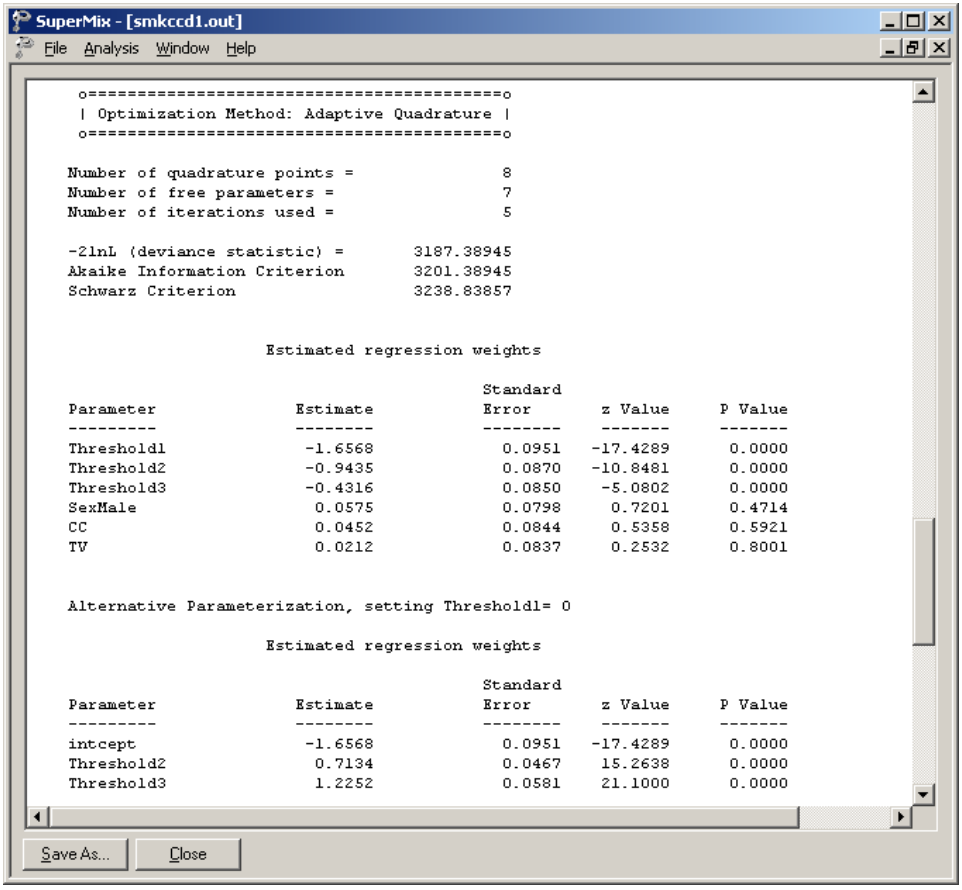




Fixed effects estimates

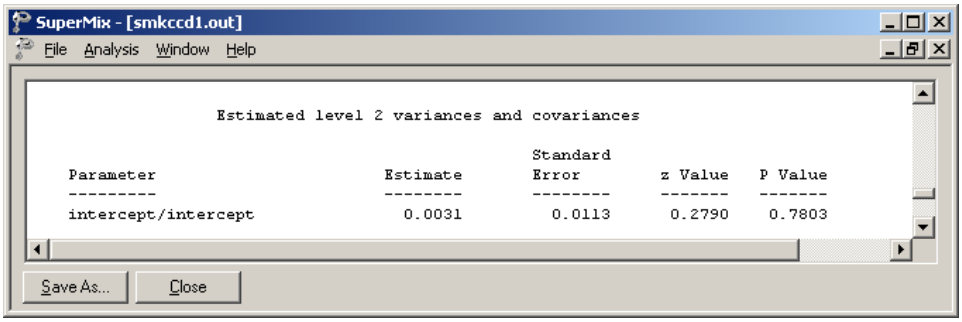
This is followed by the results for the model specified, but without any random effects. In this format, none of the included predictors are significant. It will be interesting to compare these results with those obtained once the hierarchical structure of the data has been taken into account.

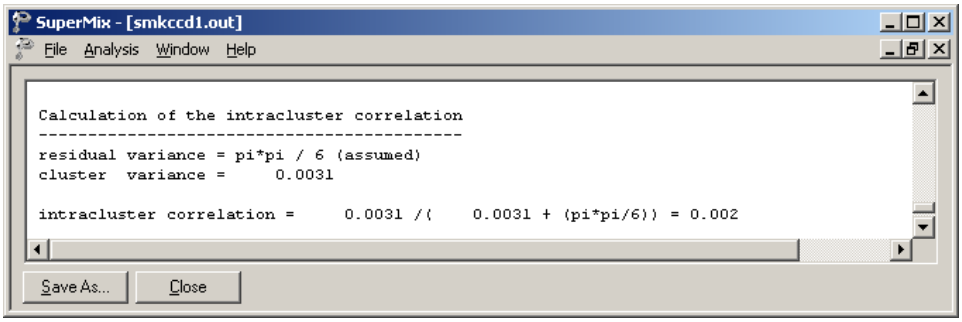




Parameter estimates are given in the next part of the output. Taking the hierarchical structure into account and allowing for the intercept to vary randomly over the schools had little effect on the significance level of the 3 covariates: all are still non-significant. We note that the three thresholds, which represent the cumulative baseline hazard, are estimated as –1.6564, –0.9431, and –0.4313 respectively. An alternative parameterization is also given. Here, the first threshold has been set to zero and as a result, the intercept and second and third threshold estimates are calculated as –1.6564, 0.7133, and 1.2251 respectively.

Random effects estimates and intraclass correlation (ICC)





The last part of the output shows the estimates of the random effects and an estimate of the intracluster correlation. There is no evidence of significant random variation in the intercept over the schools ($p = 0.8120$). The intracluster correlation coefficient shown is based on the use of the complementary log-log link function for these data, which results in a residual variance of $\pi^2/6$ (see Agresti, 2002).

8.6.4 Interpreting the output

Comparing binary and ordinal models

When the number of measurement occasions is not too large, the binary outcome model utilizing dummy variables to represent the measurement occasions can be useful in fitting survival analysis models. Additionally, the binary model easily

allows relaxation of the proportional hazards assumption for model covariates through inclusion of interaction terms with the time point indicators. Finally, though not illustrated here, the binary model can also handle time-dependent covariates in the same manner as the covariate by time interactions. When the number of occasions is very large, however, the number of time point indicators that must be created for the binary model, and the resulting size of the data set, can get very large and unwieldy. In this case, the ordinal outcome model such as the model discussed in this section is perhaps the better analysis option (though covariates must follow the proportional hazards assumptions and time-dependent covariates are not allowed). If the complementary log-log link function is selected (*i.e.*, the model is specified as a proportional hazards model), the binary and ordinal outcome models yield identical estimates for parameters that do not depend on time (Laara & Matthews, 1985). This is shown in Table 8.12. The regression coefficients are exactly the same for Male, CC, and TV. This is also true of their standard errors and so the p-values for both sets are identical. However, the intercept and threshold parameters, which do represent time-related information, are not the same with the exception of the first intercept. The reason for this is that the intercepts in the binary model represent the *interval-specific* baseline hazard, whereas their corresponding threshold parameters in the ordinal model represent the *cumulative* baseline hazard across the time intervals. These are only equivalent only for the first time interval and thereafter diverge in value and meaning. Finally, it should be mentioned that if one uses the logit link, in place of the complementary log-log link, the estimates (of the parameters not involving time) from the binary and ordinal models are not equivalent, though similar.

Notice also that the likelihood values for the two representations are identical, as are the AIC values. The Schwarz values are not the same because the numbers of observations in the two representations are different. That is, because the binary-case data set consists of multiple person-time indicators for each outcome, the numbers of observations in the binary-case data set is inflated, relative to the ordinal case.

Table 8.12: Comparison of results of binary and ordinal outcome models

Term	Binary outcome (EVENT)	Ordinal outcome (SmkOnset)
Wave B baseline hazard binary α_{01} or ordinal γ_1	–1.6564	–1.6564
Wave C baseline hazard binary $\alpha_{01} + \alpha_{02}$ or ordinal γ_2	–1.65654+0.0399 = –1.6165	–0.9431
Wave D baseline hazard binary $\alpha_{01} + \alpha_{03}$ or ordinal γ_3	–1.6564 +0.3103 = –1.3461	–0.4313
Male β_1	0.0574	0.0574
CC β_2	0.0449	0.0449
TV β_3	0.0213	0.0213
$-2 \ln L$	3187.38817	3187.38817
AIC	3201.38817	3201.38817
Schwarz	3243.94116	3238.83729
No. of parameters	7	7

9 Syntax

9.1 Introduction and notes

SuperMix syntax files can be generated either through the interface or by inputting the commands in Notepad and then saving it as an ***.inp** file. The structures of the syntax files and the interfaces vary slightly for the different types of outcome variables.

When syntax is generated through the interface, the commands are generated and saved to a ***.inp** file. When the input file is constructed or edited outside the interface, the following guidelines should be kept in mind:

- All commands start with a keyword and conclude with a semi-colon.
- There is no specific required order in which commands have to be given, with the exception of the MODELS and OPTIONS commands, which must always be the first two lines in the input file.
- Lines may be left blank between commands.
- Commands and keywords are not case-sensitive, but variable names are.
- When data is imported to an **ss3** file, a **mum** or **inp** file can be created.
- If an **inp** (syntax file) contains variable names that exceeds 16 characters, these names are truncated and only the first 16 characters are displayed in the output file.
- When variable names contain blank(s) or arithmetic symbols, quotation marks are needed. Examples are “%Fat”, “CC*TV” and “VISIT 1” etc.
- Line length is restricted to 128 characters in the syntax (.inp) file, but a command can continue over several lines. For example, the list of predictor names when there are a large number of predictors in the model.

For new users, generating the syntax file through the interface is highly recommended.

In this chapter, the dialog boxes, with corresponding syntax, are first discussed for continuous outcomes and then for other outcome variable types. When a dialog box that has been illustrated in one case is also used for another type of outcome variable, the image or screenshot will not be repeated, and the user is referred to the original image. Finally, each of the commands is explained in detail, in alphabetical order.

9.2 Syntax file for continuous outcomes

9.2.1 Structure

The basic structure of the syntax file for the continuous outcome is as given below, and the **required** commands are indicated.

Model = Continuous;	Required
Options;	Required
Link = name of the link function;	Optional
Distribution = name of the distribution;	Optional
Varnames = names of the variables used in the model;	Required
Title1 = job title;	Optional
Title2 = job title;	Optional
DataFile = name of the system data file with data to be analyzed;	Required
Level2ID = name of the variable identifying level-2 units;	Optional
Level3ID = name of the variable identifying level-3 units;	Optional
Dependent = name of the outcome variable;	Required
MeansTable = name of the variable to generate a means by outcome table;	Optional
Dependent_Miss = missing value for the outcome variable;	Optional
Global_Miss = global missing value;	Optional
Predictors = names of predictors in the fixed part of the model;	Required
L1Random = names of the level-1 random effects;	Optional
L2Random = names of the level-2 random effects;	Optional
L3Random = names of the level-3 random effects;	Optional
ErrStart = starting value(s) of the error variance(s) ;	Optional
FixStart = starting value(s) for the parameters in the fixed part of the model;	Optional
Cov2Start = starting value(s) for the level-2 random effects (co)variance(s);	Optional
Cov3Start = starting value(s) for the level-3 random effects (co)variance(s);	Optional

AutoStart = starting values for the autocorrelation terms;	Optional
FixPatType = free or user-defined patterns for the fixed parameters;	Optional
FixPat = patterns for the fixed parameters;	Optional
Cov2PatType = free or user-defined level-2 covariance structure;	Optional
Cov2Pat = pattern of the level-2 random coefficient covariance matrix;	Optional
Cov3PatType = free or user-defined level-3 covariance structure;	Optional
Cov3Pat = pattern of the level-2 random coefficient covariance matrix;	Optional
AutoCor = the autocorrelation terms;	Optional
ErrorForm = the autocorrelated error form for the time series analysis;	Optional
TimeVar = name of the 'time' variable;	Optional
Weight1 = level-1 weight variable;	Optional
Weight2 = level-2 weight variable;	Optional
Weight3 = level-3 weight variable;	Optional
TransformNames = names of the linear transformations;	Optional
Transf_Start = name of a linear transformation;	Optional
FixTransf = list of values;	Optional
Cov2Transf = list of values;	Optional
Cov3Transf = list of values;	Optional
Transf_End = name of the linear transformations given in Transf_Start;	Optional

Not all of the available commands have to be included in the input file.

9.2.2 Interface with corresponding syntax

9.2.2.1 The Configuration tab

The fields on the Configuration tab include the DEPENDENT, DEPENDENT_MISS, GLOBAL_MISS, LEVELnID, MEANSTABLE, MODEL, OPTIONS, and TITLEn commands. The Configuration tab for count outcomes is structured in the same way.

The required commands are listed as shown below. Corresponding explanations of these commands are given in Section 9.7.

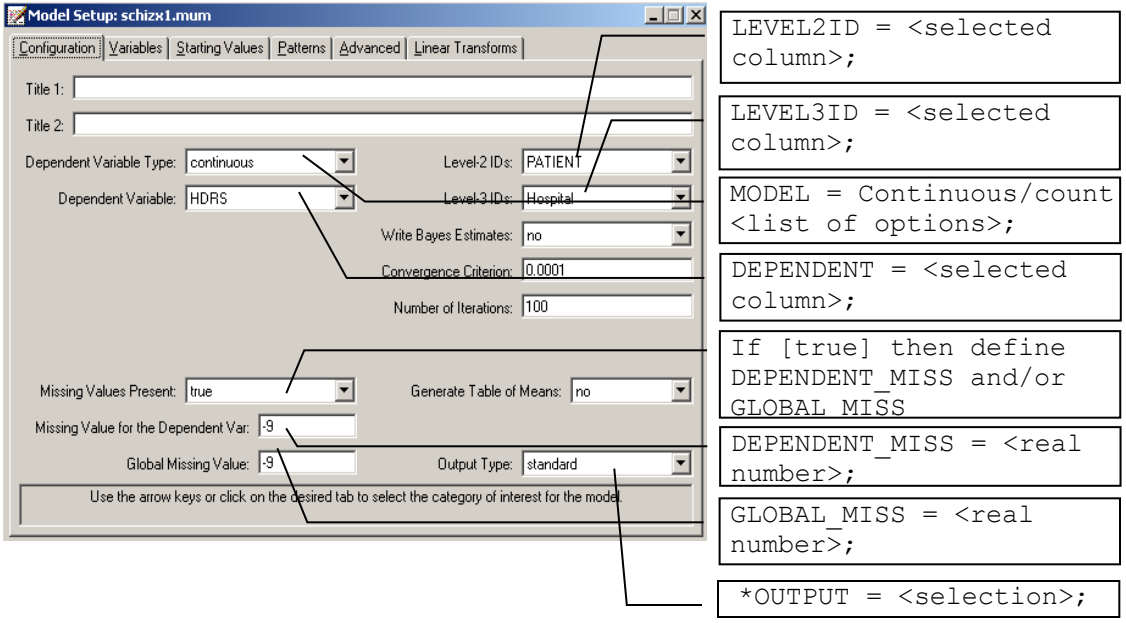


Figure 9.1(a): Configuration tab for continuous and count outcomes – required fields

Besides the required commands as shown above, a number of options are also available. The optional fields are shown in the following image.

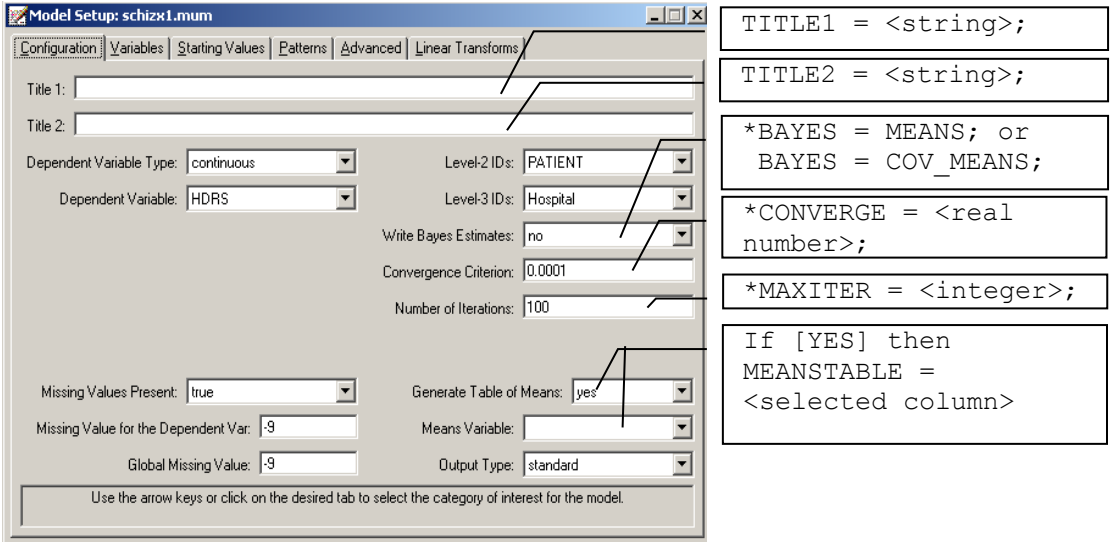


Figure 9.1(b): Configuration tab for continuous and count outcomes – optional fields

9.2.2.2 The Variables tab

The **Variables** tab is identical for all types of outcome variables, with the exception of the ordered outcome. The LnRANDOM and PREDICTORS commands are defined through this tab.

The syntax associated with this tab is shown below. The commands are explained in Section 9.7.

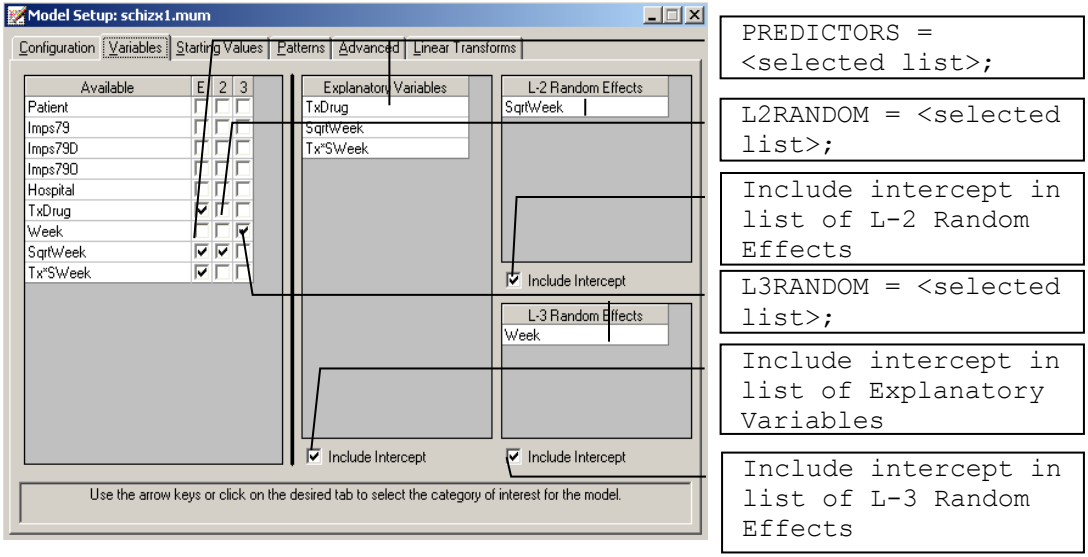


Figure 9.2: The Variables tab for continuous / count / nominal / binary outcomes

9.2.2.3 The Starting Values tab

The ERRSTART, COVnPAT, and FIXSTART commands are defined through the **Starting Values** tab. The **Starting Values** tab for the count outcome is the same as for the continuous outcome, except that the ERRSTART command is not available for count outcomes.

The syntax associated with this tab is shown below. The **Starting Values** commands are explained in Section 9.7.

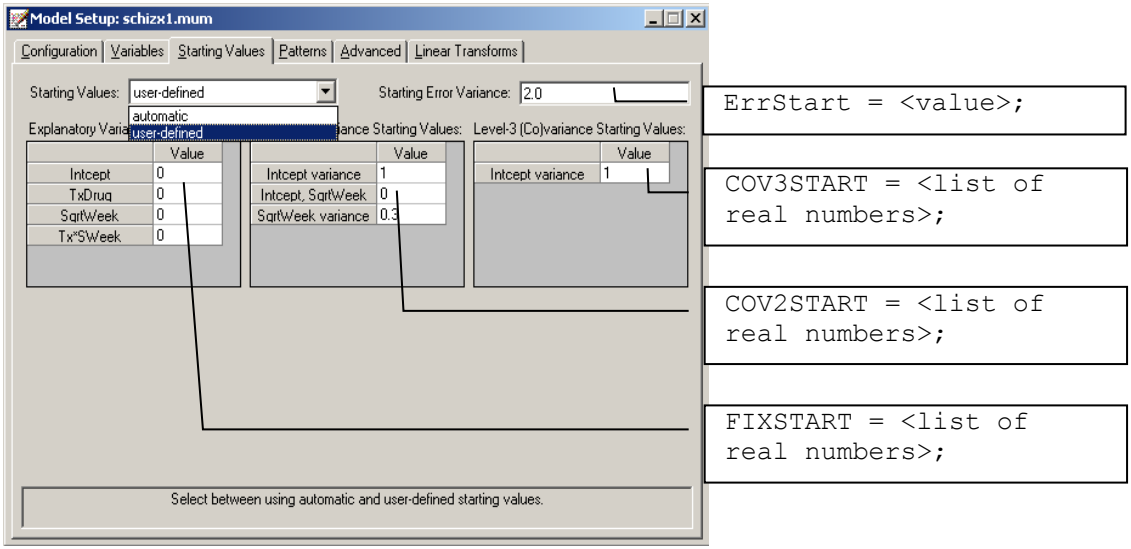


Figure 9.3: The Starting Values tab for continuous / count outcomes

9.2.2.4 The Patterns tab

The **Patterns** tab is the same for all outcomes. The COVnPATTYPE and FIXPATTYPE commands are defined here.

The syntax associated with the Patterns tab is shown below. Each command is explained in detail in Section 9.7.

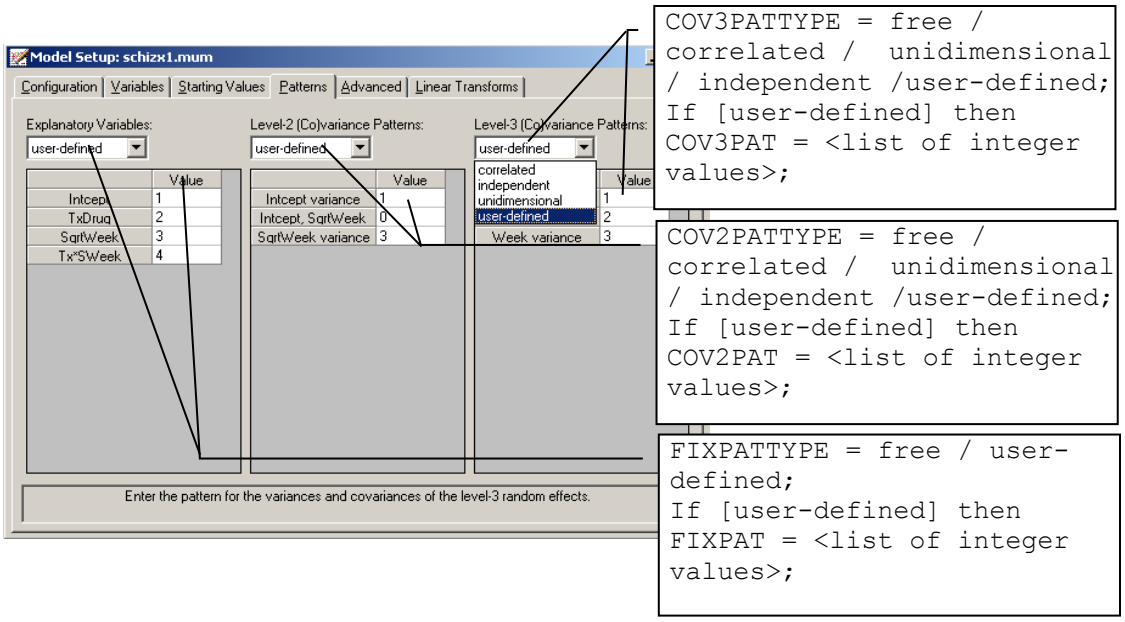


Figure 9.4: The Patterns tab for continuous / ordered / count / nominal / binary outcomes

9.2.2.5 The Advanced tab

The general part of the **Advanced** tab, which is the same for all outcomes, is used to set up the weight variables at different levels of the model. The other fields on the tab vary according to the type of outcome variable. The AUTOCOR, AUTOSTART, ERRORFORM, DISTRIBUTION, TIMEVAR, and WEIGHTn commands are defined through the **Advanced** tab.

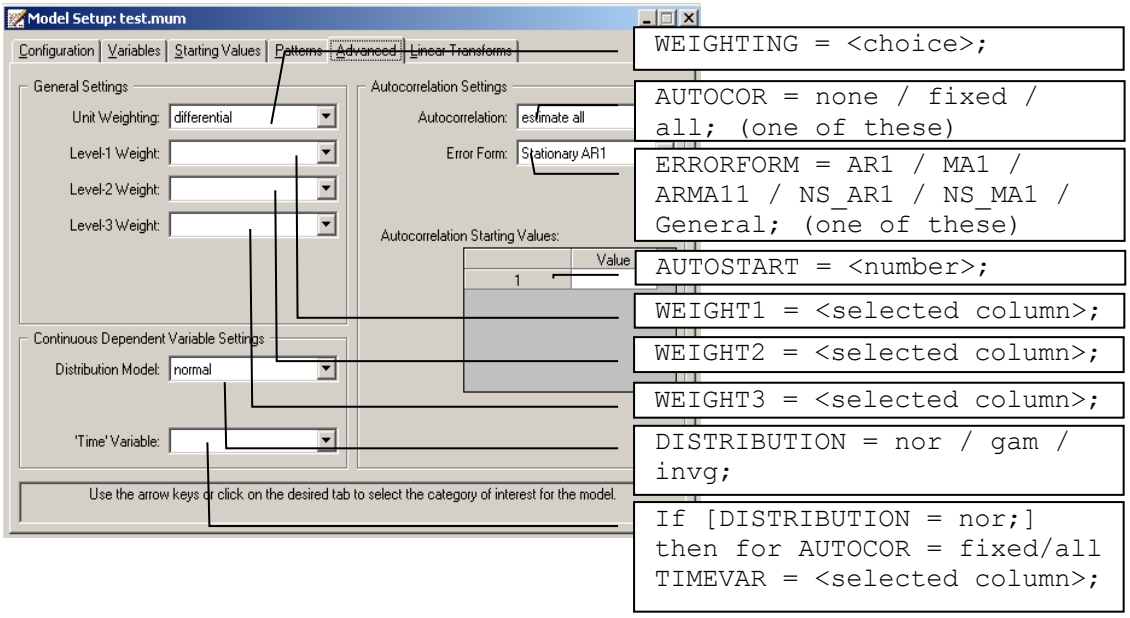


Figure 9.5(a): The Advanced tab for the normal distribution

When either the gamma or inverse Gaussian distribution is selected, an additional SCALE command is activated as shown below.

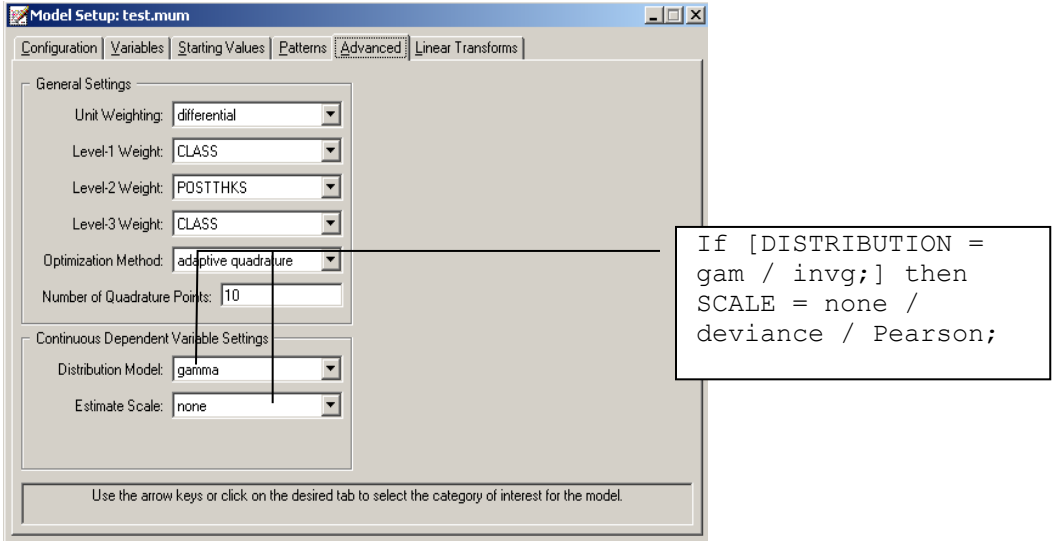


Figure 9.5(b): The Advanced tab for the gamma and inverse Gaussian distribution

The corresponding syntax associated with this tab is shown above. Each command is explained in detail in Section 9.7.

9.2.2.6 The Linear Transforms tab

The **Linear Transforms** tab includes the COVnTRANSF, FIXTRANSF, TRANSF_END, TRANSF_START, and TRANSFORMNAMES commands.

The syntax associated with this tab is shown below. All the commands are explained in Section 9.7.

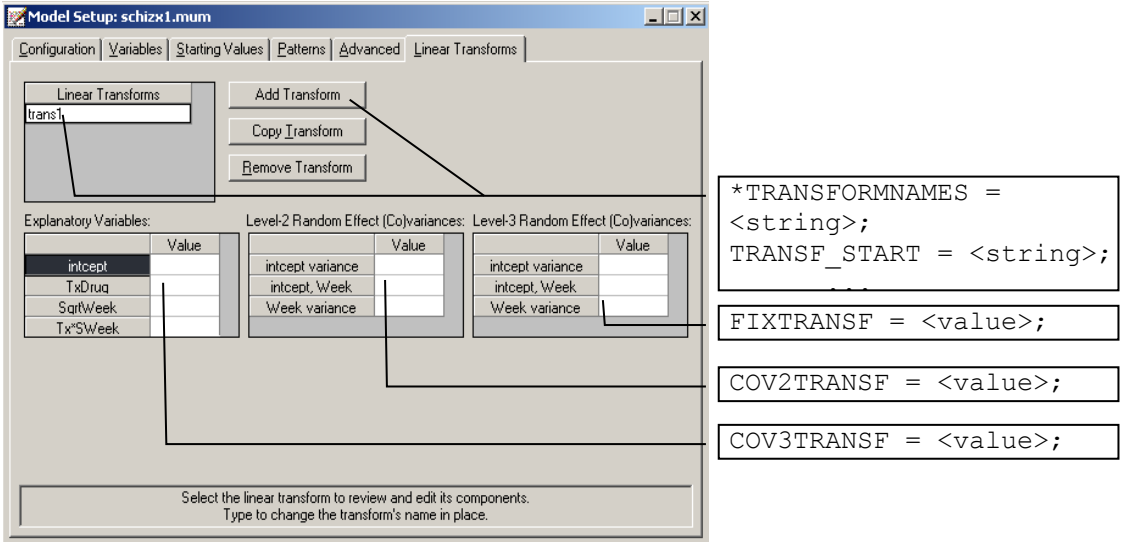


Figure 9.6: The Linear Transforms tab for continuous / count outcomes

9.3 Syntax file for ordered outcomes

9.3.1 Structure

The basic structure of the syntax file for an ordered outcome variable is as given below, and the **required** commands are indicated. Most of the commands are the same as the ones for the continuous outcome. The different/new ones are listed below in bold face.

Model = Ordered;	Required
Options;	Required
Link = name of the link function;	Required
ThRandom2 = specify random thresholds at level-2 (yes or no);	Optional
ThRandom3 = specify random thresholds at level-3 (yes or no);	Optional
Varnames = names of the variables used in the model;	Required
Title1 = first job title;	Optional
Title2 = second job title;	Optional
DataFile = name of the system data file with data to be analyzed;	Required

Level2ID = name of the variable identifying level-2 units;	Optional
Level3ID = name of the variable identifying level-3 units;	Optional
Dependent = name of the outcome variable;	Required
Categories = list of distinct values of the outcome variable;	Required
Crosstab = name of the variable to generate cross-tabulation with the outcome variable;	Optional
Dependent_Miss = missing value for the outcome variable;	Optional
Global_Miss = global missing value;	Optional
Predictors = names of **the** predictors in the fixed part of the model;	Required
L2Random = names of the level-2 random effects;	Optional
L3Random = names of the level-3 random effects;	Optional
FixStart = starting value(s) for the parameters in the fixed part of the model;	Optional
Cov2Start = starting value(s) for the level-2 random effects (co)variance(s);	Optional
Cov3Start = starting value(s) for the level-3 random effects (co)variance(s);	Optional
ThresholdStart = starting values for the threshold parameters;	Optional
FixPatType = free or user-defined patterns for the fixed parameters;	Optional
FixPat = patterns for the fixed parameters;	Optional
Cov2PatType = free or user-defined level-2 covariance structure;	Optional
Cov2Pat = pattern of the level-2 random coefficient covariance matrix;	Optional
Cov3PatType = free or user-defined level-3 covariance structure;	Optional
Cov3Pat = pattern of the level-2 random coefficient covariance matrix;	Optional
Interactions = number of predictor*threshold interactions;	Optional
Censor = name of the censor variable;	Optional
Weight1 = level-1 weight variable;	Optional
Weight2 = level-2 weight variable;	Optional
Weight3 = level-3 weight variable;	Optional
TransformNames = names of the linear transformations;	Optional
Transf_Start = name of a linear transformation;	Optional
FixTransf = list of values;	Optional
Cov2Transf = list of values;	Optional
Cov3Transf = list of values;	Optional
ThreshTransf = list of values for thresholds for the specified transformation;	Optional
FixbyThresh = list of values for the threshold interactions;	Optional
Transf_End = name of linear transformation specified in Transf_Start;	Optional

Not all of the available commands have to be included in the input file.

9.3.2 Interface with corresponding syntax

9.3.2.1 The Configuration tab

The **Configuration** tab for ordered outcomes is shown below. The DEPENDENT, DEPENDENT_MISS, GLOBAL_MISS, LEVELnID, and TITLEn commands are the same as for the continuous case, which is discussed in Section 9.2. The different or new commands included for the ordered outcome on this tab are the MODEL, CATEGORIES, and CROSSTAB commands.

Detailed information on these commands is given in Section 9.7.

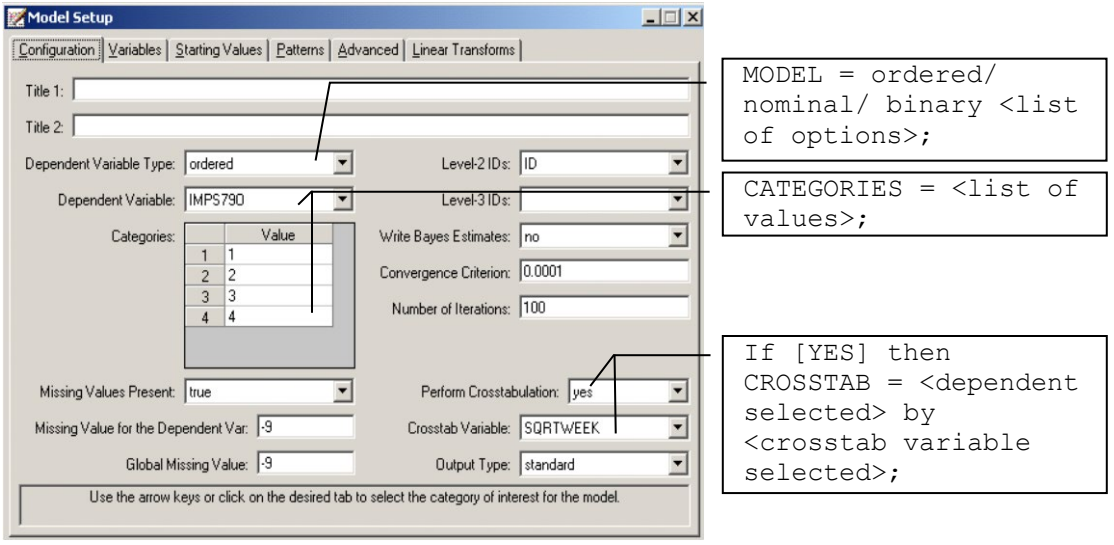


Figure 9.7: The Configuration tab for ordered / nominal / binary outcomes

9.3.2.2 The Variable tab

The **Variable** tab includes the LnRANDOM and PREDICTORS commands as discussed in Section 9.2.2.2 for the continuous outcome. The only difference for the ordered outcome is that no option to include the intercept as an explanatory variable is provided, as shown below.

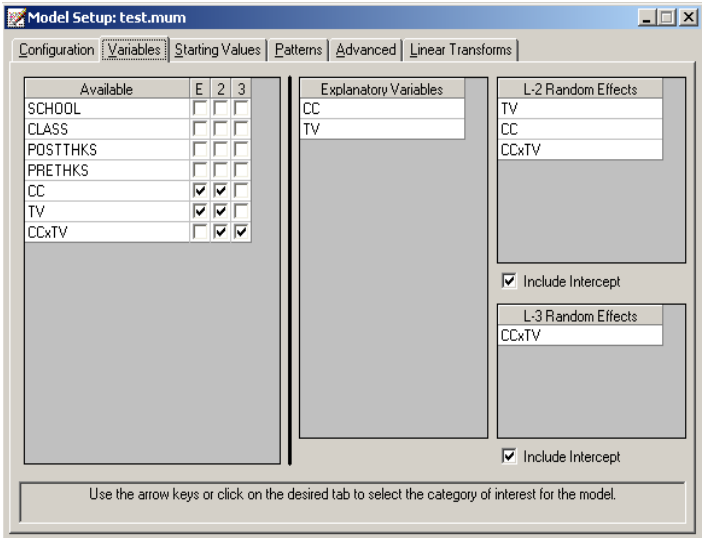


Figure 9.8: The Variables tab for ordered outcomes

9.3.2.3 The Starting Values tab

Besides the ERRSTART, COVnPAT, and FIXSTART commands, which are the same as the commands for the continuous outcome, the THRESHOLDSTART command is defined through this tab.

The syntax associated with this tab is shown below. Command syntax is explained in Section 9.7.

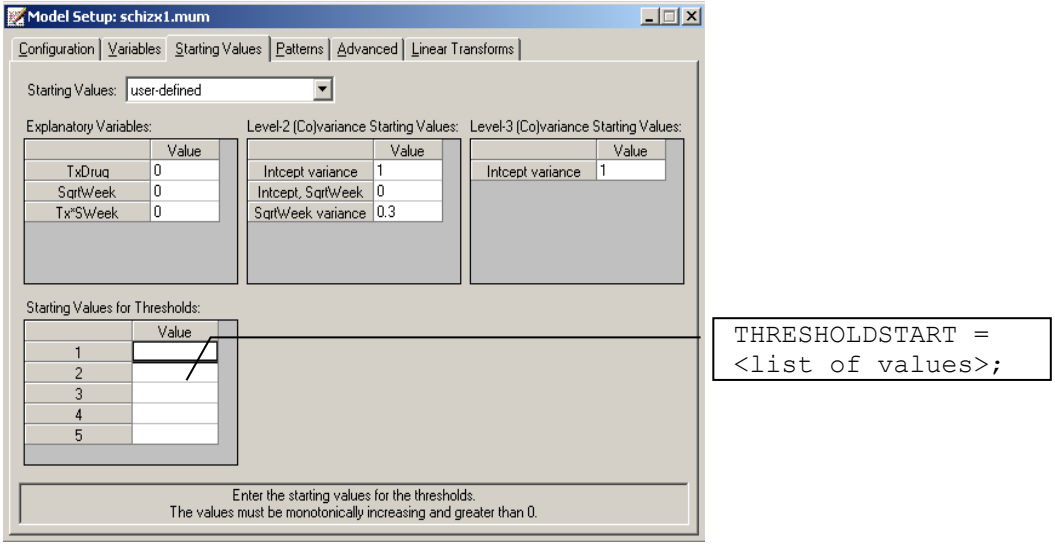


Figure 9.9: The Starting Values tab for ordered outcomes

9.3.2.4 The Patterns tab

The **Patterns** tab for the ordered outcome is identical to that for continuous outcomes, which is discussed in Section 9.2.2.4.

9.3.2.5 The Advanced tab

The general settings for weights, which includes the WEIGHTn ($n = 1, 2, 3$)command, are the same as on the **Advanced** tab for the continuous outcome. Fields used in the case of ordered outcomes, corresponding to the CENSOR, INTERACTIONS, LINK commands and METHOD, MODELTERMS, and NQUADPTS keywords, are illustrated below. The syntax associated with this tab is shown below. The syntax commands are discussed in Section 9.7.

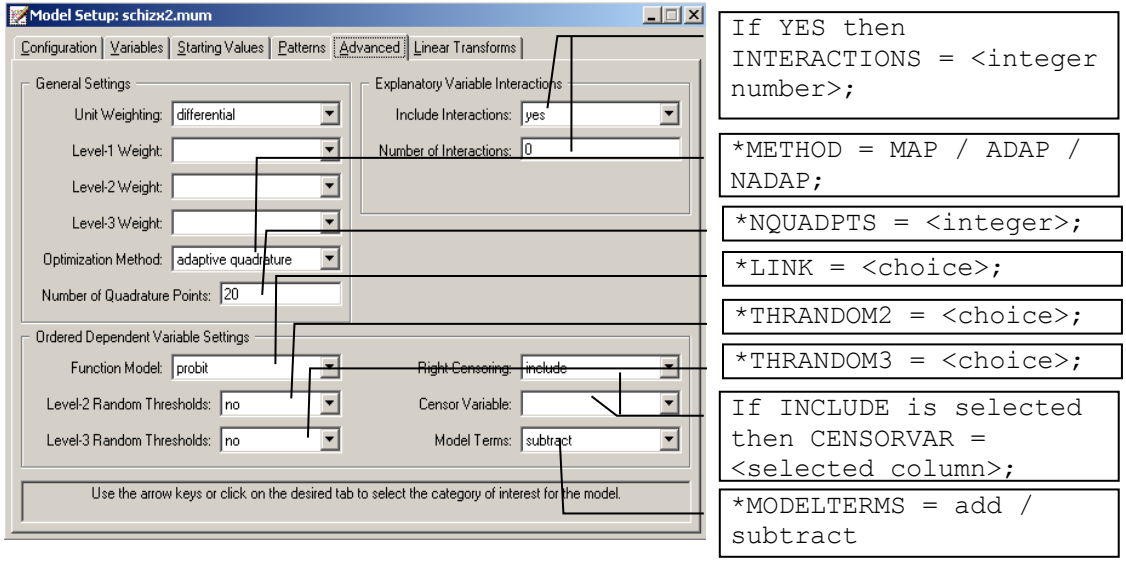


Figure 9.10: The Advanced tab for ordered outcomes

9.3.2.6 The Linear Transforms tab

Besides the COVnTRANSF, FIXTRANSF, TRANSF_END, TRANSF_START, and TRANSFORMNAMES commands, which are the same as for continuous outcomes, two new commands, FIXBYTHRESH and THRESHTRANSF, are defined here for ordered outcomes.

The syntax associated with this tab is shown below. Command syntax is explained in Section 9.7.

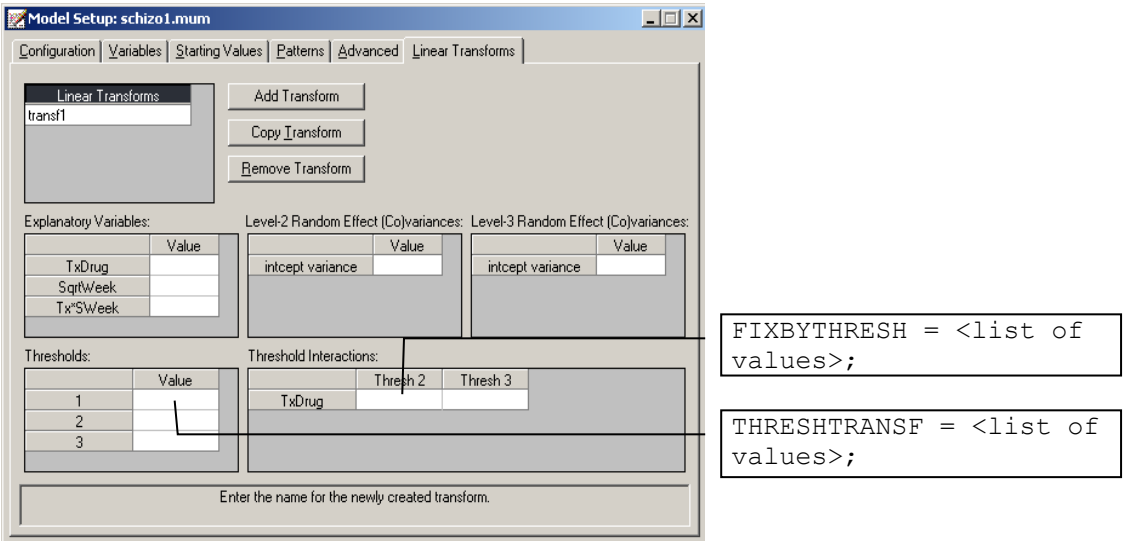


Figure 9.11: The Linear Transforms tab for ordered outcomes

9.4 Syntax file for nominal outcomes

9.4.1 Structure

The basic structure of the syntax file for a nominal outcome variable is as given below, and the **required** commands are indicated. Most of the commands are the same as the ones for the continuous outcome. The different/new ones are listed below in bold face.

Model = Nominal;	Required
Options = list of options;	Required
Link = name of the link function;	Required
Varnames = names of the variables used in the model;	Required
Title1 = job title;	Optional
Title2 = job title;	Optional
DataFile = name of the system data file with data to be analyzed;	Required
Level2ID = name of the variable identifying level-2 units;	Optional
Level3ID = name of the variable identifying level-3 units;	Optional

Dependent = name of the outcome variable;	Required
Categories = list of distinct values of the outcome variable;	Required
Crosstab = name of the **variable to generate cross-tabulation with the outcome variable;**	Optional
Dependent_Miss = missing value for the outcome variable;	Optional
Global_Miss = global missing value;	Optional
Predictors = names of the predictors in the fixed part of the model;	Required
L2Random = names of the level-2 random effects;	Optional
L3Random = names of the level-3 random effects;	Optional
FixStart = starting value(s) for the parameters in the fixed part of the model;	Optional
Cov2Start = starting value(s) for the level-2 random effects (co)variance(s);	Optional
Cov3Start = starting value(s) for the level-3 random effects (co)variance(s);	Optional
FixPatType = free or user-defined patterns for the fixed parameters;	Optional
FixPat = patterns for the fixed parameters;	Optional
Cov2PatType = free or user-defined level-2 covariance structure;	Optional
Cov2Pat = pattern of the level-2 random coefficient covariance matrix;	Optional
Cov3PatType = free or user-defined level-3 covariance structure;	Optional
Cov3Pat = pattern of the level-2 random coefficient covariance matrix;	Optional
Weight1 = level-1 weight variable;	Optional
Weight2 = level-2 weight variable;	Optional
Weight3 = level-3 weight variable;	Optional
TransformNames = names of the linear transformations;	Optional
Transf_Start = name of a linear transformation;	Optional
FixTransf = list of values;	Optional
Cov2Transf = list of values;	Optional
Cov3Transf = list of values;	Optional
Transf_End = name of the linear transformation given in Transf_Start;	Optional

Not all of the available commands have to be included in the input file.

9.4.2 Interface with corresponding syntax

The **Configuration** tab for nominal outcome variables is identical to that of ordered outcomes, which is discussed in Section 9.3.2.1. The **Variable** tab, which includes both the LnRANDOM and PREDICTORS commands, is the same as for continuous outcomes, as discussed in Section 9.2.2.2.

9.4.2.1 The Starting Values tab

The FIXSTART and COVnSTART commands are the same as for a continuous outcome, except that starting values for the predictors and random effects are requested in terms of categories versus a reference (first or last) category. This tab is shown below.

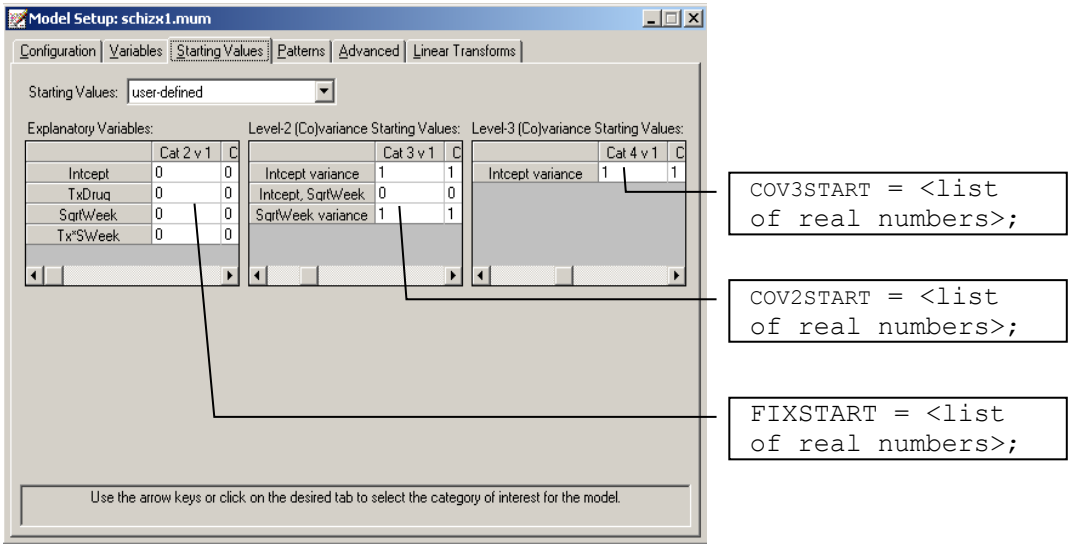


Figure 9.12: The Starting Values tab for nominal outcomes

9.4.2.2 The Patterns tab

The Patterns tab for the nominal outcome is used for specifying the COVnPATTYPE and FIXPATTYPE commands as shown below.

Each field of the **Patterns** tab is explained in detail in Sections 9.7.5, 9.7.6, 9.7.17, and 9.7.18.

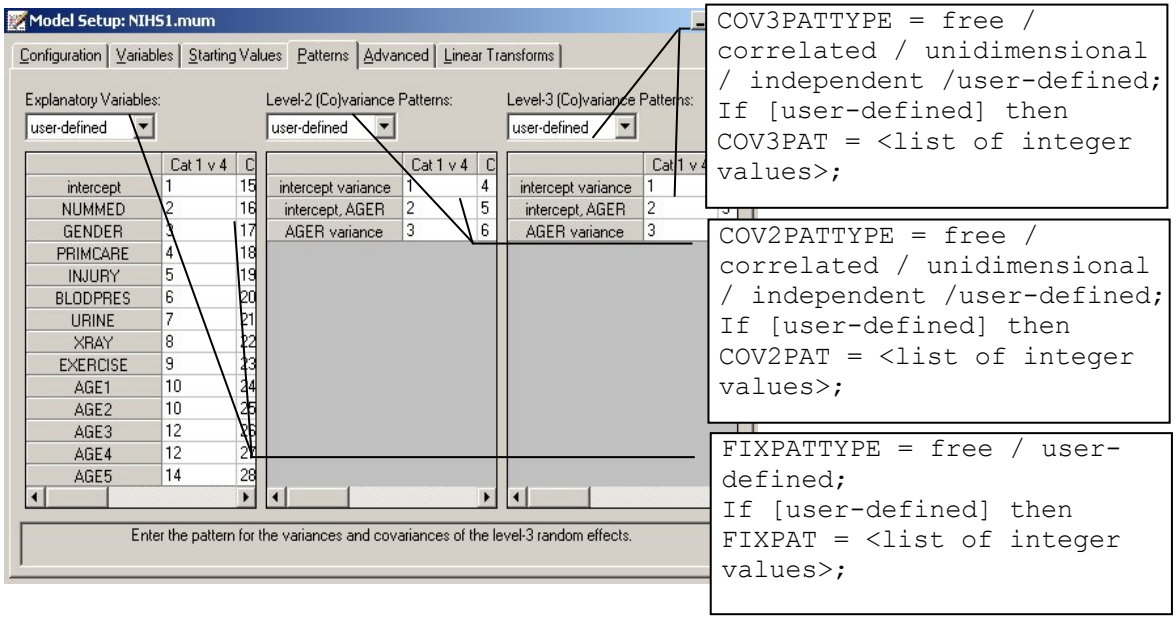


Figure 9.13: The Patterns tab for nominal outcome variables

9.4.2.3 The Advanced tab

The general settings of weights, which includes the WEIGHTn commands, are the same as for continuous outcomes, while the METHOD and NQUADPTS commands are the same as for ordered outcomes. The unique field for the nominal outcome is the the REFCAT keyword on the OPTIONS command.

The syntax associated with this tab is shown below. Command syntax is explained in Section 9.7.

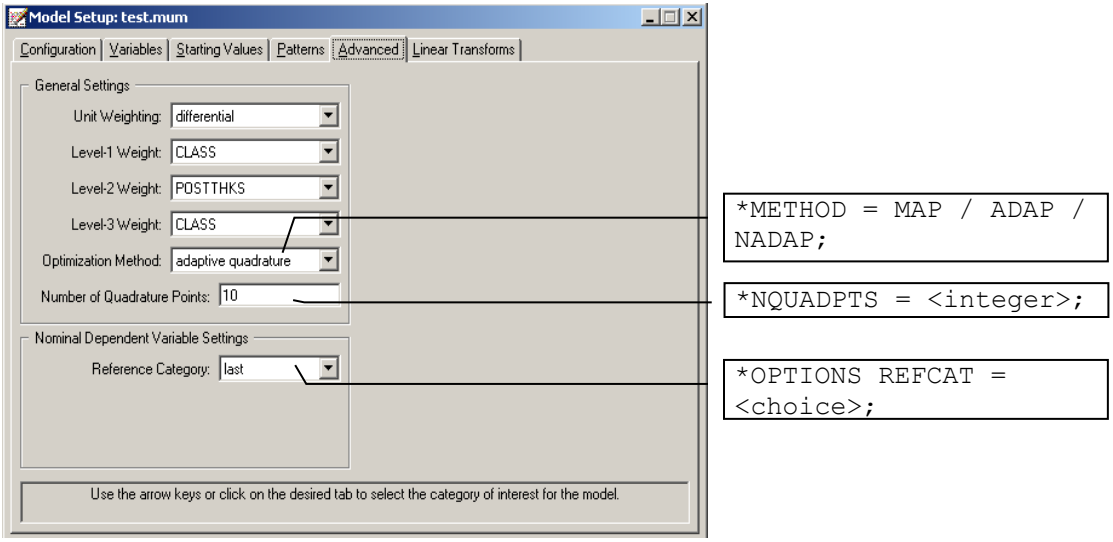


Figure 9.14: The Advanced tab for nominal outcomes

9.4.2.4 The Linear Transforms tab

The **Linear Transform** tab includes the same commands as for the continuous outcome, these being the COVnTRANSF, FIXTRANSF, TRANSF_END, TRANSF_START, and TRANSFORMNAMES commands. The appearance of this tab is slightly different from the one for the continuous outcome, as shown below.

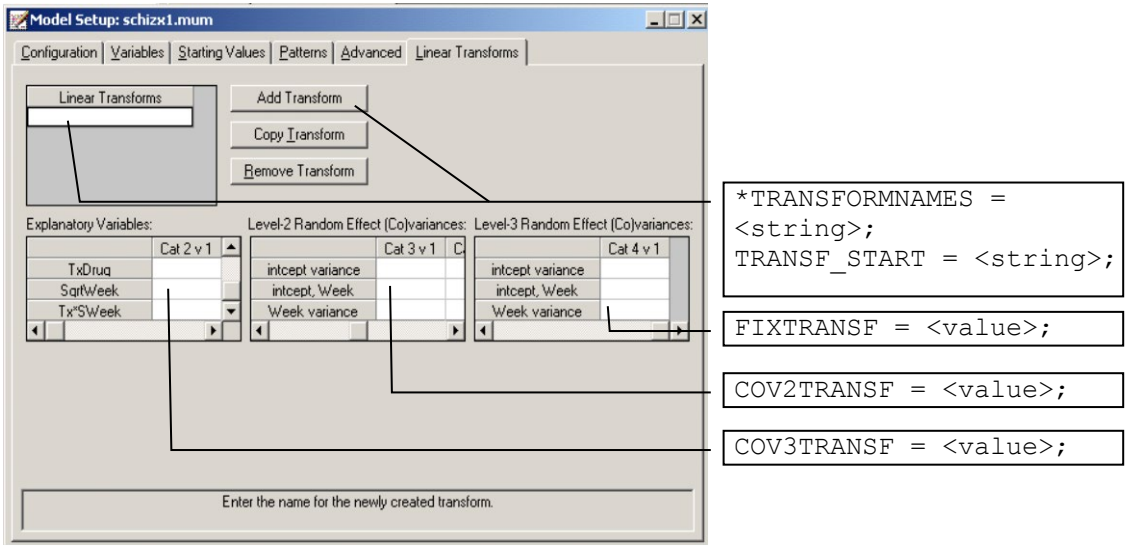


Figure 9.15: The Linear Transforms tab for nominal outcomes

9.5 Syntax file for count outcomes

9.5.1 Structure

The basic structure of the syntax file for the count outcome is as given below, and the **required** commands are indicated. Most of the commands are the same as the ones for the continuous outcome. The different/new ones are listed below in bold face.

Model = Count;	Required
Options;	Required
Distribution = name of the distribution;	Required
Scale = scaling method; / Dispersion = value of the dispersion parameter;	Optional
Varnames = names of the variables used in the model;	Required
Title1 = job title;	Optional
Title2 = job title;	Optional
DataFile = name of the system data file with data to be analyzed;	Required

Level2ID = name of the variable identifying level-2 units;	Optional
Level3ID = name of the variable identifying level-3 units;	Optional
Dependent = name of the outcome variable;	Required
MeansTable = name of the variable to generate a means by outcome table;	Optional
Dependent_Miss = missing value for the outcome variable;	Optional
Global_Miss = global missing value;	Optional
Predictors = names of the predictors in the fixed part of the model;	Required
L2Random = names of the level-2 random effects;	Optional
L3Random = names of the level-3 random effects;	Optional
FixStart = starting value(s) for the parameters in the fixed part of the model;	Optional
Cov2Start = starting value(s) for the level-2 random effects (co)variance(s);	Optional
Cov3Start = starting value(s) for the level-3 random effects (co)variance(s);	Optional
FixPatType = free or user-defined patterns for the fixed parameters;	Optional
FixPat = patterns for the fixed parameters;	Optional
Cov2PatType = free or user-defined level-2 covariance structure;	Optional
Cov2Pat = pattern of the level-2 random coefficient covariance matrix;	Optional
Cov3PatType = free or user-defined level-3 covariance structure;	Optional
Cov3Pat = pattern of the level-2 random coefficient covariance matrix;	Optional
Offset = name of the offset (exposure) variable;	Optional
Weight1 = level-1 weight variable;	Optional
Weight2 = level-2 weight variable;	Optional
Weight3 = level-3 weight variable;	Optional
TransformNames = names of the linear transformations;	Optional
Transf_Start = name of a linear transformation;	Optional
FixTransf = list of values;	Optional
Cov2Transf = list of values;	Optional
Cov3Transf = list of values;	Optional
Transf_End = name of the linear transformation given in Transf_Start;	Optional

Not all of the available commands have to be included in the input file.

9.5.2 Interface with corresponding syntax

The **Configuration**, **Variables**, **Starting Values**, **Patterns** and **Linear Transformations** tabs for count outcomes are identical to those for continuous outcomes, which are

discussed in Section 9.2. The only tab that will be discussed in this section is the **Advanced** tab for the count outcome.

9.5.2.1 The Advanced tab

The general settings for weights, which includes the WEIGHTn commands, are the same as the **Advanced** tab for the continuous outcome. The METHOD and NQUADPTS keywords are the same as the ones in the **Advanced** tab for the ordered outcome. The unique fields for the count outcome are the SCALE and OFFSET commands when Poisson distribution is selected, and the DISPERSION and OFFSET commands when the negative binomial distribution is selected. The syntax associated with this tab is shown below. Command syntax is explained in Section 9.7.

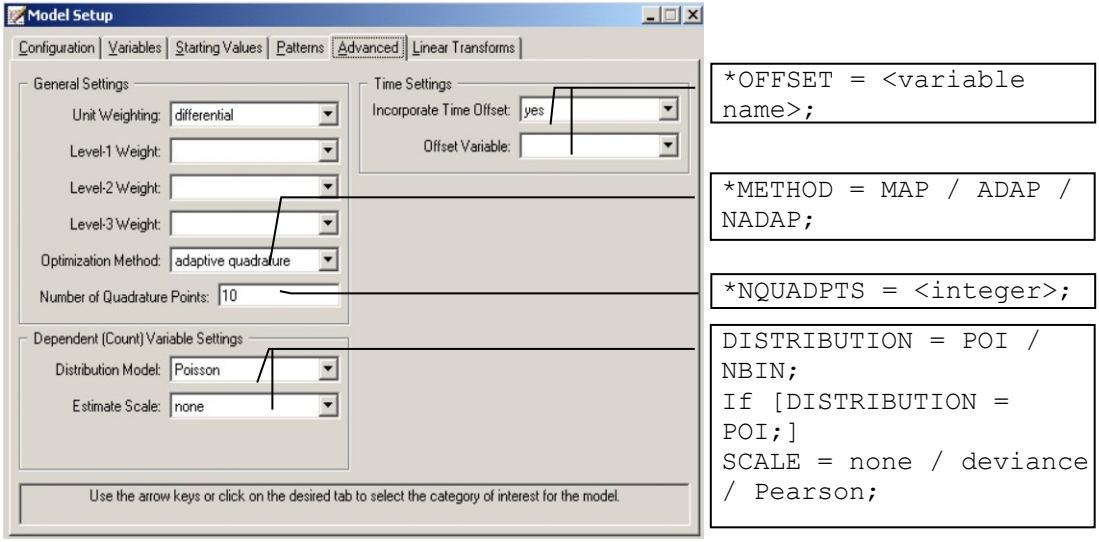


Figure 9.16: The Advanced tab for count outcomes – Poisson distribution

When the negative binomial distribution is selected, the **Advanced** tab with corresponding commands is as shown below.

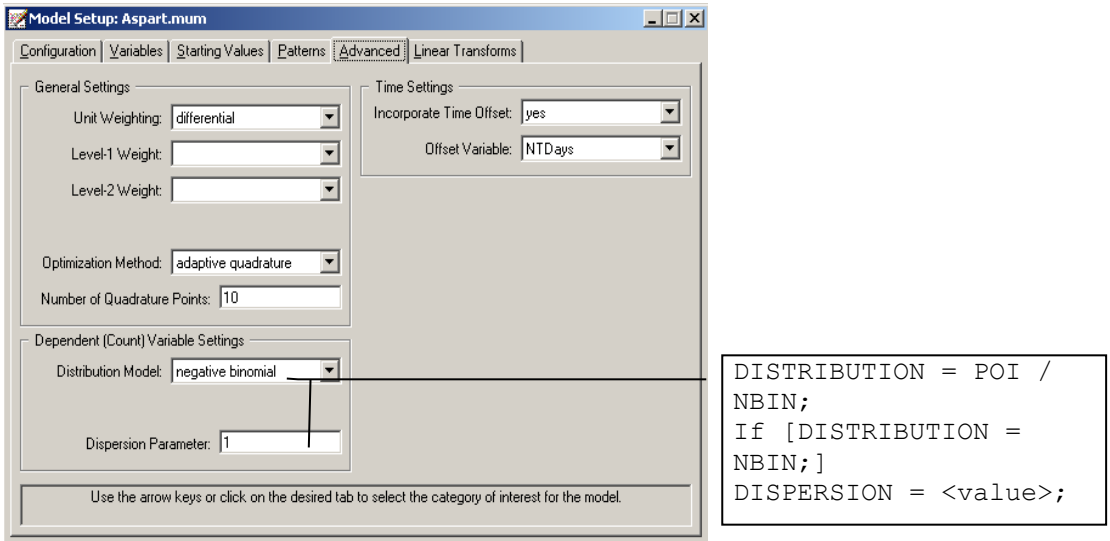


Figure 9.17: The Advanced tab for count outcomes – negative binomial distribution

9.6 Syntax file for binary outcomes

9.6.1 Structure

The basic structure of the syntax file for the binary outcome is as given below, and the **required** commands are indicated. Most of the commands are the same as the ones for the continuous outcome. The different/new ones are listed below in bold face.

Model = Binary;	Required
Options;	Required
Link = name of the link function;	Required
Distribution = name of the distribution;	Required
Scale = scaling method;	Optional
Ntrials = variable contains the number of trials (binomial);	Optional
Varnames = names of the variables used in the model;	Required
Title1 = job title;	Optional

Title2 = job title;	Optional
DataFile = name of the system data file with data to be analyzed;	Required
Level2ID = name of the variable identifying level-2 units;	Optional
Level3ID = name of the variable identifying level-3 units;	Optional
Dependent = name of outcome variable;	Required
Categories = list of distinct values of the outcome variable;	Required
Crosstab = name of variable to generate cross-tabulation with outcome variable;	Optional
Dependent_Miss = missing value for the outcome variable;	Optional
Global_Miss = global missing value;	Optional
Predictors = names of the predictors in the fixed part of the model;	Required
L2Random = names of the level-2 random effects;	Optional
L3Random = names of the level-3 random effects;	Optional
FixStart = starting value(s) for the parameters in the fixed part of the model;	Optional
Cov2Start = starting value(s) for the level-2 random effects (co)variance(s);	Optional
Cov3Start = starting value(s) for the level-3 random effects (co)variance(s);	Optional
FixPatType = free or user-defined patterns for the fixed parameters;	Optional
FixPat = patterns for the fixed parameters;	Optional
Cov2PatType = free or user-defined level-2 covariance structure;	Optional
Cov2Pat = pattern of the level-2 random coefficient covariance matrix;	Optional
Cov3PatType = free or user-defined level-3 covariance structure;	Optional
Cov3Pat = pattern of the level-2 random coefficient covariance matrix;	Optional
Weight1 = level-1 weight variable;	Optional
Weight2 = level-2 weight variable;	Optional
Weight3 = level-3 weight variable;	Optional
TransformNames = names of the linear transformations;	Optional
Transf_Start = name of a linear transformation;	Optional
FixTransf = list of values;	Optional
Cov2Transf = list of values;	Optional
Cov3Transf = list of values;	Optional
Transf_End = name of the linear transformation given in Transf_Start;	Optional

Not all of the available commands have to be included in the input file.

9.6.2 Interface with corresponding syntax

The **Configuration** tab for the binary outcome is the same as for the ordinal outcome (see Section 9.3.2.1). The **Variables**, **Patterns**, **Starting Values**, and **Linear Transforms** tabs are the same as the continuous case, and are discussed in Section 9.2. The only unique tab is the **Advanced** tab for the binary outcome.

9.6.2.1 The Advanced tab

The general settings for weights, which include the WEIGHTn commands, are the same as for the continuous outcome. The DISTRIBUTION, LINK, and SCALE commands and METHOD and NQUADPTS keywords were all discussed previously. The unique field for the binary outcome is the NTRIALS command. The syntax associated with this tab is shown below. Command syntax is explained in Section 9.7.

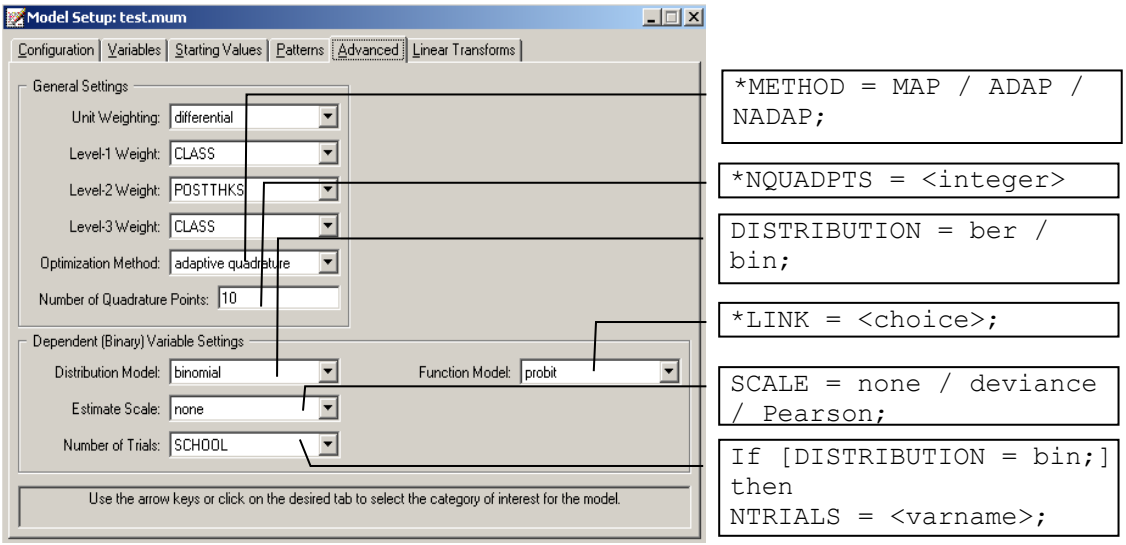


Figure 9.18: The Advanced tab for binary outcomes

9.7 Commands

9.7.1 AUTOCOR command

The AUTOCOR command is used to specify the inclusion of autocorrelated level-1 errors. This is an **optional** command for the continuous outcome, and presently only available for 2-level models.

Syntax:

```
AUTOCOR = <keywords>;
```

Valid keywords are as follows:

NONE	In this case no autocorrelation terms are considered.
FIXED	Fixed autocorrelation terms are considered in such a case. The ErrorForm command will be included to complete this selection.
ALL	When the ALL keyword is used, all the autocorrelation terms will be estimated. The ErrorForm command will be included to complete this selection.

Example:

```
AutoCor = NONE;
```

9.7.2 AUTOSTART command

The AUTOSTART command is used to specify the starting values for the autocorrelation terms. This is an **optional** command for the continuous outcome, and presently only available for 2-level models.

Syntax:

AUTOSTART = <values>;

The values must be between –0.999 and 0.999 inclusive.

Example:

AUTOSTART = 0.75;

9.7.3 CATEGORIES command

The CATEGORIES command is used to specify each of the distinct values of the outcome variable. This is a **required** command for models with ordinal, nominal and binary outcomes. These values are automatically generated by the GUI.

Syntax:

CATEGORIES = <list of integer values>;

Examples:

Categories = 1 2 3 4 5;
Categories = 0 1;

9.7.4 CENSOR command

The CENSOR command is used to define the censor variable, which is coded 0 for censor, 1 for event. This is an **optional** command for use with ordinal outcomes.

Syntax:

CENSOR = <variable name>;

Example:

CENSOR = cen_var;

9.7.5 COVnPAT command

The COVnPAT commands are used to place constraints on the covariance matrices of random coefficients on the different levels of the model. We denote these covariance matrices by $\Phi_{(1)}$, $\Phi_{(2)}$, and $\Phi_{(3)}$ or, in general, by $\Phi_{(n)}$, $n = 1, 2, 3$.

One COVnPAT command is allowed for each level of the hierarchy. If, for instance, a 3-level linear model with random components on all three levels of the hierarchy is to be fitted, up to three COVnPAT commands may be included in the syntax file.

Note that on level 1 (continuous outcome), only structures pertaining to the diagonal elements of the level-1 random effects covariance matrix are permissible. The use of COVnPAT commands is **optional**.

Syntax:

COVnPAT= <keywords>;

Valid keywords are as follows:

DIAG	In this case the covariance matrix of random parameters on level *n* of the model will be constrained to be a diagonal matrix.
TOEPLITZ	The covariance matrix on levels 2 or 3 will be constrained to be of the form of a so-called Toeplitz matrix, that is

$$\mathbf{\Phi}_{(n)} = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & & \\ \gamma_1 & \gamma_0 & \gamma_1 & O & \\ \gamma_2 & \gamma_1 & \gamma_0 & O & \gamma_2 \\ & O & O & O & \gamma_1 \\ & & \gamma_2 & \gamma_1 & \gamma_0 \end{bmatrix}$$

INTRA — The covariance matrix of random parameters on levels 2 or 3 will be constrained to have an intra-class structure, that is

$$\mathbf{\Phi}_{(n)} = \begin{bmatrix} \alpha & \beta & K & K & \beta \\ \beta & \alpha & \beta & K & M \\ M & \beta & \alpha & O & M \\ M & O & O & O & \beta \\ \beta & K & K & \beta & \alpha \end{bmatrix}$$

MA1 — Constrains the covariance matrix on level *n* to be similar to that of a time series process of order MA1. The form of the covariance matrix will then be

$$\mathbf{\Phi}_{(n)} = \begin{bmatrix} \gamma & \beta & 0 & K & 0 \\ \beta & \gamma & \beta & O & M \\ 0 & \beta & \gamma & O & 0 \\ M & O & O & O & \beta \\ 0 & K & 0 & \beta & \gamma \end{bmatrix}$$

USER_SPECIFIED

To constrain the elements of the covariance matrix to be of a form other than those discussed above, the user may specify this required structure with the COVnPAT command. This can be done by entering a lower-triangular matrix with the required structure on the COVnPAT command. If, for example, the covariance matrix corresponding to the LnRANDOM command

```
LnRANDOM = X1 X2 X3 X4;
```

is to be constrained, it can be accomplished by following a row-wise numbering convention for the lower triangular elements of the covariance matrix as shown below.

```
1
2  3
4  5  6
7  8  9  10
```

The elements to be fixed are then replaced with "0". If, for example, the matrix is constrained to be diagonal, the command to be used is as follows:

```
COVnPAT = 1
          0  3
          0  0  6
          0  0  0  10;
```

The structure as specified indicates that there are four parameters to be estimated (*i.e.* numbers 1, 3, 6 and 10, corresponding to the variances) and six fixed parameters (corresponding to the covariances), indicated by 0. The values which the fixed parameters are to be set equal to can be supplied using the COVnVAL command. If the COVnVAL command is omitted, the fixed parameters will be constrained to be equal to zero, as the initial structure of all covariance matrices is assumed to be diagonal at the start of the iterative procedure.

The following conventions apply to the use of the COVnPAT command:

- Any line of input may not exceed 127 characters. Thus, if a large COVnPAT matrix is entered, a row of the lower-triangular matrix may be continued on the next line of the syntax file if the number of characters in that row of the matrix exceeds 127 characters.

- If elements of the covariance matrix to be estimated are constrained to be equal in value, the number assigned to those elements must be the same.
- As with all other commands in the syntax file, the command should end with a semi-colon that may be placed directly after the last element of the matrix as specified or on the next line of the syntax file.
- The matrix specified must have the same number of elements as implied by the LnRANDOM command. That is, if there are p variables listed in the LnRANDOM command, a total number of $\frac{1}{2} p (p + 1)$ elements must be entered.
- In order to assign initial values to elements of the covariance matrix on level n or to set fixed elements of the matrix to user specified values, the COVnPAT command must be used in conjunction with the COVnVAL command.

Examples:

1. In the case of an MA1 process, for example, the command will be as follows:

```
COVnPAT = 1
          2  1
          0  2  1
          0  0  2  1;
```

   From this structure it follows that there are only two parameters to be estimated (numbers 1 and 2) while all other parameters are constrained to be equal to zero, unless otherwise specified using the COVnVAL command (see Section 9.7.7 for information on the COVnSTART command).

2. It is permissible to constrain diagonal elements of the level-n covariance matrix to be fixed through the use of the COVnPAT command.

The following commands, for example, are permissible:

```
COVnPAT = 1
          2  0
          3  2  0
          0  0  2  0;

COVnPAT = 0
          2  0
          3  2  0
          0  0  2  0;
```

Note that 0-values indicate that the corresponding elements remain fixed at the values specified in the COVnSTART paragraphs.

9.7.6 COVnPATTYPE command

The COVnPATTYPE commands are used to specify specific structures for the covariance matrices of the random effects on the different levels of the hierarchy. This is an **optional** command.

Note that on level 1 (continuous case), only structures pertaining to the diagonal elements of the level-1 random effects covariance matrix are permissible.

Syntax:

```
COVnPATTYPE = <keywords>;
```

Valid keywords are as follows:

CORRELATED — This is the default option. In this case the covariance matrix of random parameters on level *n* of the model will be constrained to be a symmetric matrix.

INDEPENDENT The covariance matrix on levels 2 or 3 will be constrained to be independent, that is

$$\Phi_{(n)} = \begin{pmatrix} \gamma_1 & 0 & \mathsf{L} & 0 \\ 0 & \gamma_2 & \mathsf{L} & 0 \\ 0 & 0 & \mathsf{O} & 0 \\ 0 & 0 & \mathsf{L} & \gamma_n \end{pmatrix}$$

UNIDIMENSIONAL When the UNIDIMENSIONAL keyword is used, the covariance matrix of random parameters on levels 2 or 3 is replaced by a scalar, expressed as a linear combination of the random effects.

- Unidimensional selected at level 3

$$\Phi_{(3)} = \psi_1 v_{i1} + \psi_2 v_{i2} + ... + \psi_r v_{ir},$$

where $\psi_1, \psi_2, ..., \psi_r$ are unknown parameters to be estimated.

- Unidimensional selected at level 2

$$\Phi_{(2)} = \theta_1 v_{ij1} + \theta_2 v_{ij2} + ... + \theta_q v_{ijq},$$

where $\theta_1, \theta_2, ..., \theta_q$ are unknown parameters to be estimated.

USER-DEFINED To constrain the elements of the covariance matrix to be of a form other than those discussed above, the user may specify this required structure through use of the USER-DEFINED keyword together with the COVnPAT command. This can be done by entering a lower-triangular matrix with the required structure, using the COVnPAT command.

The following convention applies to the use of the COVnPATTYPE command:

- As with all other commands in the input file, the command should end with a semi-colon that may be placed directly after the last element of the matrix as specified or on the next line of the input file.

Examples:

```
Cov2PatType = Correlated;
Cov2PatType = User-Defined;
Cov3PatType = Independent;
```

9.7.7 COVnSTART command

COVnSTART commands may be used to provide either initial values for elements of the covariance matrix on level *n* of the model or to provide values for elements fixed through the use of keywords of the COVnPAT command. Note that the use of COVnSTART commands is **optional**.

One COVnSTART command is allowed for each level of the hierarchy. If, for instance, a 3-level linear model with random coefficients on all three levels of the hierarchy is to be fitted, up to three COVnSTART commands may be included in the syntax file.

The values to be used for the elements of the covariance matrix must be entered in the form of a lower-triangular matrix. The number of values entered must be the same as the number of elements implied by the relevant LnRANDOM command. If there are p variables listed in the LnRANDOM command, $\frac{1}{2}\, p\, (p + 1)$ values must be entered. If a large number of values is entered, a row of the lower-triangular matrix may be continued on the next line of the syntax file if the number of characters in that row of the matrix exceeds 127 characters. The command must end with a semi-colon, which may be entered on the last line of the values given or on the next line of the syntax file.

Syntax:

COVnSTART = <values specified by user>;

Examples:

```
COV2START = 1.00
            0.32  0.85
            0.63  0.62  0.78
            0.19  0.00  0.25  0.99;
```

The above command can also be written as

```
COV2START = 1 0.32 0.85 0.63 0.62 0.78 0.19 0.00 0.25 0.99;
```

If an accompanying COVnPAT command is not used, these values will function as starting values for the level-*n* covariance matrix. When good starting values for the elements of this covariance matrix are known, the use of the command as shown could decrease the number of iterations required to obtain convergence.

When the command

```
COVnPAT = DIAG;
```

is used together with the COVnSTART command given in the previous example, the values specified on the diagonal of the lower-triangular matrix will be used as initial values for the parameters which are to be estimated. The off-diagonal elements of the covariance matrix will then be constrained to be equal to the values of the off-diagonal elements of the matrix given above.

9.7.8 COVnTRANSF command

The COVnTRANSF command is used together with the TRANSF_START and TRANSF_END commands to specify values corresponding to elements of the level-2 and level-3 random effect covariance matrices that are used to test linear contrasts. It is an **optional** command for all the types of the outcome variables.

Syntax:

COVnTRANSF = <values>;

Example:

Suppose a specific model has two level-2 random effects, and that we wish to test the null hypothesis

$$H_0 : \left[\boldsymbol{\Phi}_{(2)}\right]_{1,1} = \left[\boldsymbol{\Phi}_{(2)}\right]_{2,2},$$

that is,

$$H_0 : \left[\boldsymbol{\Phi}_{(2)}\right]_{1,1} - \left[\boldsymbol{\Phi}_{(2)}\right]_{2,2} = \mathbf{0}.$$

This is accomplished by using the command

COV2TRANSF = 1 0 –1;

9.7.9 CROSSTAB command

The purpose of the CROSSTAB command is to select a categorical variable to be cross-tabulated with the outcome variable. A $(C \times K)$ frequency table is produced, where C denotes the number of distinct values of the outcome variable and K the

number of distinct values of the variable selected in the CROSSTAB command. It is an **optional** command for the ordered, nominal and binary outcomes.

Syntax:

```
CROSSTAB = <variable name>;
```

Example:

```
CROSSTAB = week;
```

9.7.10 DATAFILE command

The DATAFILE command is used to specify the name of the data file (space-delimited text file) to be analyzed, and is automatically generated if the multilevel model specifications are built via the dialog boxes. Note that the data file does not refer to the ***.ss3** file, but refers to the ***.dat** file which is generated when the data analysis is run using a ***.mum** file. The DATAFILE command is a **required** command.

Syntax:

```
DATAFILE = <file name>;
```

where <file name> denotes the complete name (including folder name) of the data file. The folder name may be omitted if the data file and SuperMix syntax file are in the same folder.

Example:

The command shown below is used to set up the model by using the DAT file **reisby.dat**, located in the **continuous** folder.

```
DATAFILE = 'C:\SUPERMIX\CONTINUOUS\REISBY.DAT';
```

9.7.11 DEPENDENT command

The DEPENDENT command is used to select the name of the outcome variable(s). Since variable names are case sensitive, spelling, etc. of the names of the outcome variables must be the same as that used in the data spreadsheet (***.ss3** file). The DEPENDENT command is a **required** command.

Syntax:

```
DEPENDENT = <outcome variable(s)>;
```

In the case of a multivariate model, more than one outcome variable may be listed in the DEPENDENT command. Response variables may be entered in any order. This command is presently only available for normally distributed continuous outcomes.

Examples:

In the DEPENDENT command below, the outcome variable is indicated as the variable Y1:

```
DEPENDENT = Y1;
```

The DEPENDENT command for a multivariate model (continuous outcomes only), in which 6 response variables are listed, would look like this:

```
DEPENDENT = Math1 Math2 Math3 Eng1 Eng2 Eng3;
```

9.7.12 DEPENDENT_MISS command

The DEPENDENT_MISS command may be used to specify a code assigned to missing values on the outcome variable(s) only. The consequence of using the DEPENDENT_MISS command is that only records with outcome variable values equal to the code assigned through the DEPENDENT_MISS command will be removed from the analysis. The DEPENDENT_MISS command is an **optional** command.

Syntax:

```
DEPENDENT_MISS = <value>;
```

Any positive or negative value may be used. Only one value is allowed in this command. All records containing data values equal to the code specified in this command will subsequently be removed from the analysis.

Example:

Consider the observations

Outcome variable	Predictor variables
99	1 10 14.5 48.7
7.7	3 12 13.7 53.2
6.5	4 11 12.6 999
8.3	2 15 18.1 55.4

and the command

```
DEPENDENT_MISS = 99;
```

If the code 99 is identified as the code for missing data values on the dependent variables, this will imply that all the values for case 1 will be deleted, and all the values for cases 2, 3, and 4 will be retained.

If, additionally, the code 999 is specified (GLOBAL_MISS = 999, see Section 9.7.21) as the code for missing data values on all the variables included in the analysis, the third record as given above will be deleted from the data set to be analyzed. The second and fourth observations will be retained.

This is accomplished by using both the DEPENDENT_MISS and GLOBAL_MISS commands as follows:

```
DEPENDENT_MISS = 99;
GLOBAL_MISS = 999;
```

Note that if only the DEPENDENT_MISS command is used, the value of 999 for the last predictor variable for the third observation will be considered valid data and will be used as such in the analysis.

9.7.13 DISTRIBUTION command

The DISTRIBUTION command includes different distributions for each type of outcome variable. The keywords and the distribution names (given in parentheses) for different outcomes are shown below.

Table 9.1: List of distributions and associated keywords

outcome variable type	default distribution	other available distribution(s)	
Continuous	nor (normal)	gam (gamma)	invg (inverse Gaussian)
Ordered	mul (multinomial)	–	–
Nominal	mul (multinomial)	–	–
Count	poi (Poisson)	nbin (negative binomial)	–
Binary	ber (Bernoulli)	bin (binomial)	–

The DISTRIBUTION command is a **required** command for count and binary outcomes.

Syntax:

```
DISTRIBUTION = <keywords the distribution>;
```

Examples:

```
DISTRIBUTION = NOR;
DISTRIBUTION = poi;
```

9.7.14 ERRORFORM command

The ERRORFORM command is used to specify the autocorrelated error form for the time series analysis. It is an **optional** command for continuous outcomes and is presently only available for 2-level models.

Syntax:

ERRORFORM = <keyword>;

The available keywords for different error forms are listed below.

AR1 The first-order autoregressive disturbance of order 1, or AR(1) process, has the following error variance matrix

$$\mathbf{\Phi}_{(1)} = \begin{pmatrix} \gamma & \gamma\rho & \gamma\rho^2 & \mathsf{L} & \gamma\rho^{t-1} \\ \gamma\rho & \gamma & \gamma\rho & \mathsf{L} & \gamma\rho^{t-2} \\ \mathsf{M} & \mathsf{M} & \mathsf{M} & \mathsf{M} & \mathsf{M} \\ \gamma\rho^{t-1} & \gamma\rho^{t-2} & \gamma\rho^{t-3} & \mathsf{L} & \gamma \end{pmatrix}$$

The keyword AR1 refers to the stationary AR(1) process, which is the autoregressive process with $|\rho| < 1$.

NSAR1 The keyword NSAR1 refers to the non-stationary AR(1) process.

MA1 The moving average process of order 1 is obtained by using the MA1 keyword. In order to constrain the covariance matrix on level *n* to be similar to that of a time series process of order MA1, The form of the covariance matrix will then be

$$\mathbf{\Phi}_{(1)} = \begin{pmatrix} \gamma & \beta & 0 & \mathsf{L} & 0 \\ \beta & \gamma & \beta & \mathsf{L} & 0 \\ 0 & \beta & \gamma & \mathsf{L} & 0 \\ \mathsf{O} & \mathsf{O} & \mathsf{O} & \mathsf{O} & \mathsf{O} \\ 0 & 0 & \mathsf{L} & \beta & \gamma \end{pmatrix}$$

ARMA11 The mixed process, ARMA, is more complicated since it is a mixture of the AR and MA forms. For the ARMA(1, 1) process,

$$y_t = \gamma y_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1}$$

9.7.15 ERRSTART command

The ERRSTART command is used to specify starting values for the level-1 error covariance matrix. It is an **optional** command, used only for the normally distributed continuous outcome.

Syntax:

```
ERRSTART = <values>;
```

Example:

```
ERRSTART = 0.85;
```

9.7.16 FIXBYTHRESH command

The FIXBYTHRESH command is used together with the TRANSF_START and TRANSF_END commands, which define linear transformations. FIXBYTHRESH allows the users to test a null hypothesis of the type $H_0 : \mathbf{C}'\boldsymbol{\gamma} = \mathbf{0}$ involving threshold interaction terms. It is an **optional** command for ordered outcomes.

Syntax:

```
FIXBYTHRESH = <value(s)>;
```

Example:

```
Transf_Start= t;
FixTransf= 0 0 0 0;
Cov2Transf= 0 0 0;
Cov3Transf= 0 0 0;
ThreshTransf= 0 0;
FixbyThresh= 1 0 –1 0;
Transf_End= t;
```

9.7.17 FIXPAT command

To specify a patterned structure for the vector of fixed parameters, the FIXPAT command may be used, with or without an additional FIXSTART command (see Section 9.7.19). Use of this command is **optional**.

Syntax:

```
FIXPAT = <list of numbers>;
```

where <list of numbers> denotes a list of positive integers separated by blank spaces. The number of values entered must equal the number of predictors in the model.

Examples:

1. Constraining fixed effects to be equal:

```
FIXPAT = 1 1 3 3 5 6;
```

This statement specifies that the vector of six parameters in the fixed part of the model are constrained as follows: BETA1 = BETA2; BETA3 = BETA4 while BETA5 and BETA6 are estimated freely. In the command shown above, the actual numbers correspond to the order of the parameter in question: "1" denotes the

first parameter, "3" the third and "5" and "6" the fifth and sixth of the parameters in the fixed part of the model.

2. Fixing fixed effects to user-specified values:

```
FIXPAT = 0 0 3;
```

If '0' values are in the list of numbers, then the FIXPAT command should be used in conjunction with the FIXSTART command. If, for example, FIXSTART = 10 2.5 0.15; then BETA1 and BETA2 are fixed at their initial values (10 and 2.5 respectively) while BETA3 is estimated freely, using a starting value of 0.15.

9.7.18 FIXPATTYPE command

The FIXPATTYPE command is used to specify that all the parameters in the fixed part of the model are free to be estimated, or that parameter estimation will be user-defined using the FIXPAT command. The FIXPATTYPE command is an **optional** command.

Syntax:

```
FIXPATTYPE = <keywords>;
```

Valid keywords are as follows:

FREE	This is the default setting for the FIXPATTYPE command.
USER-DEFINED	Allows the user to fix specific elements (see FIXPAT command) of the vector **β** of fixed parameters to values specified by the FIXSTART command and/or to constrain elements to be equal.

The following convention applies to the use of the FIXPATTYPE command:

- The command should end with a semi-colon.

Example:

```
FIXPATTYPE  = User-Defined;
```

9.7.19 FIXSTART command

It is also possible to provide initial values for the fixed parameters in the model to be analyzed. This may be achieved with the FIXSTART command, which allows the user to provide starting values for these parameters. Use of this command is **optional**.

Syntax:

```
FIXSTART = <as specified by user>;
```

The number of values entered using this command must be equal to the number of fixed parameters to be estimated. There is no specific format in which the values have to be entered.

Example:

The commands

```
FIXSTART = 0.151 0.355 0.654;
FIXSTART = 0.151
           0.355
           0.654;
```

and

```
FIXSTART = 0.151  0.355
           0.654
           ;
```

are all permissible. If the first of these commands is used and the number of characters in the user-specified string exceeds 127 characters, the next line of the syntax file should be used.

9.7.20 FIXTRANSF command

The FIXTRANSF command is used together with the TRANSF_START and TRANSF_END commands to specify a linear hypothesis of the form $H_0 : \mathbf{C}'\boldsymbol{\gamma} = \mathbf{0}$, invoking elements of the fixed parameters. It is an **optional** command for all the types of outcome variables.

Syntax:

```
FIXTRANSF = <list of values equal to number of fixed parameters>;
```

Example:

Suppose the vector $\boldsymbol{\beta}$ of fixed parameters has four elements and we wish to test the hypothesis

$$H_0 : 0.5\beta_1 + 0.5\beta_2 = \beta_3,$$

that is,

$$0.5\beta_1 + 0.5\beta_2 - \beta_3 = 0.$$

Then, the FIXTRANSF command will be:

```
FIXTRANSF = 0.5  0.5  –1.0  0;
```

9.7.21 GLOBAL_MISS command

The GLOBAL_MISS and commands may be used when missing data is present in the raw data file. The GLOBAL_MISS command allows the user to specify a numerical value, which will represent a missing value on any of the variables used in the analysis. This command may also be used in conjunction with the DEPENDENT_MISS command, as described in Section 9.7.12. Note that use of the GLOBAL_MISS command is **optional**.

Syntax:

```
GLOBAL_MISS = <value>;
```

Any positive or negative value may be used. Only one value is allowed in this command. All records with data values equal to the code specified in this command will subsequently be removed from the analysis.

Examples:

Valid examples of the use of this command include the following:

```
GLOBAL_MISS = 99;
GLOBAL_MISS = –998.0;
GLOBAL_MISS = 0;
```

9.7.22 INTERACTIONS command

The INTERACTIONS command allows the user to enter the number of predictors that are multiplied with the thresholds to form interactions. It is an **optional** command for the ordered outcome variables.

Syntax:

```
INTERACTIONS = <number>;
```

The number refers to a positive integer. The maximum value is equal to the number of predictors in the model.

Example:

Suppose that there are 4 predictors: AGE, GENDER, WEIGHT and BLODPRES. Further assume that the outcome variable has 3 categories, so that there are 2 threshold parameters to be estimated, denoted as THRESH1 and THRESH2. SuperMix also prints out an alternative parameterization denoted as intcept and THRESH2 by assuming THRESH1 = 0. The command

```
INTERACTIONS = 2;
```

instructs the program to estimate the interaction terms AGE*THRESH2 and GENDER*THRESH2.

To ensure the estimability of the interaction coefficients, there are no interactions with the first threshold or, equivalently, the intercept. If the predictors were selected in the order BLODPRES, WEIGHT, AGE, GENDER, the interactions terms would be given by BLODPRES *THRESH2 and WEIGHT *THRESH2.

9.7.23 LEVELnID command

The LEVELnID command(s) are used to indicate the variable(s) identifying the units on the different levels of the hierarchy.

If the model specified by the user is a level-2 model, the command LEVEL2ID is required. Likewise, if a level-3 model is to be considered, the LEVEL2ID and LEVEL3ID commands are required in the input file.

Variables listed in the LEVELnID commands must be included in the spreadsheet (***.ss3** file). Variable names are case sensitive, therefore the spelling and case in which they are given need to correspond to that given in the spreadsheet. LEVELnID command(s) are **required** command(s).

Syntax:

LEVELnID = <variable name identifying level-n units>;

Example:

Suppose the raw data file contains information on the test scores, age and gender of pupils belonging to classes within schools, and the variables school, class, age, gender, and score are contained in the spreadsheet. The following LEVELnID commands may be used to identify the levels of the hierarchical structure:

```
LEVEL2ID = CLASS;
LEVEL3ID = SCHOOL;
```

9.7.24 LINK command

The LINK command is used to indicate the link functions. It is an **optional** command for continuous, count, and nominal outcomes.

Syntax:

```
LINK = <keyword>;
```

The link functions available for the different types of outcome variables are listed in the table below. Note that the complete link names are listed in parentheses if different from the keywords.

Table 9.2: Outcome variable types and available link functions

Outcome variable type	Distribution	Default link function	Other available link functions		
Continuous	Normal	Identity	–	–	–
	Inverse Gaussian	log	–	–	–
	Gamma	log	–	–	–
Ordered		probit	logistic	comp_log-log (complementary log-log)	log-log
Nominal		logistic	–	–	–
Count	Poisson	log	–	–	–
	Negative binomial	log	–	–	–
Binary	Bernoulli	probit	logistic	comp_log-log (complementary log-log)	log-log
	Binomial	probit	logistic	comp_log-log (complementary log-log)	log-log

Examples:

```
LINK = probit;
LINK = comp_log-log;
```

9.7.25 LnRANDOM command

The LnRANDOM command is used to identify those variables whose coefficients are allowed to vary randomly over a given level of the hierarchy. One LnRANDOM command is allowed for each level of the hierarchy. When the input file is created through the interface, the LnRANDOM command(s) are automatically generated. Variables listed, except for the variable intcept (intercept), must be included in the data spreadsheet (***.ss3** file). The spelling and case in which they are given need to correspond to that given in the spreadsheet. By default, the intercept is automatically included as a random effect at level 2 and level 3 of the hierarchy and, in the case of normally distributed continuous outcomes, also at level 1 of the hierarchy. To exclude the intercept term at any level, the corresponding **Intercept** check box must be unchecked. The LnRANDOM command is an **optional** command.

Syntax:

```
LnRANDOM = <list of variables names to be included as random effects on level n>;
```

Example:

```
L2RANDOM = intcept PreTHKS;
L3RANDOM = intcept;
```

Note that only a model with a continuous, normally distributed outcome variable allows for random effects (usually only a random intercept) on level 1 of the hierarchy.

It is possible to place constraints on elements of the random coefficient covariance matrices. Information on the constraints permitted and on the provision of initial

values for elements of these matrices are discussed elsewhere (see Sections 9.7.5 and 9.7.7 for the COVnPAT and COVnSTART commands respectively).

9.7.26 MEANSTABLE command

The MEANSTABLE command is used to compute the mean of the selected dependent variable for each category (distinct value) of the variable selected in the MEANSTABLE command. It is an **optional** command for the continuous and count outcomes.

Syntax:

```
MEANSTABLE = <variable name>;
```

Example:

```
MEANSTABLE = Gender;
```

This command requests mean values of the outcome variables to be computed for males and females.

9.7.27 MODEL command

The MODEL command is used to define the type of dependent (outcome) variable. It is **required** for all the syntax files.

Syntax:

```
MODEL = <keyword>;
```

Valid keywords are:

CONTINUOUS	For the continuous dependent variable
ORDINAL	For the ordered dependent variable
NOMINAL	For the nominal dependent variable
COUNT	For the count dependent variable
BINARY	For the binary dependent variable

Example:

```
MODEL = BINARY;
```

9.7.28 NTRIALS command

The NTRIALS command is an **optional** command used only for the binomial distribution when a binary outcome variable is selected. It is used to define the variable specifying the number of trials corresponding to a specific number of successes. Each trial in a binomial experiment can have one of two outcomes; one is classified as a success, and the other as a failure. The number of trials refers to the number of attempts in a binomial experiment and is equal to the number of successes plus the number of failures.

Syntax:

```
NTRIALS = <variable name>;
```

Example:

```
NTRIALS = ntrials;
```

9.7.29 OFFSET command

The OFFSET command is an **optional** command used for count outcomes only. It is used to define the offset (exposure) variable.

Count models are also appropriate for rate data, where the rate is a count of events occurring for a particular unit of observation, divided by a measure of that unit's *exposure*. An example is the death rates in geographic areas as the count of deaths (outcome variable) divided by person-years (exposure). In count models, this is handled by an offset where the exposure variable is a predictor with regression coefficient constrained to 1.

Syntax:

```
OFFSET = <variable name>;
```

Example:

```
OFFSET = Pers_Yrs;
```

9.7.30 OPTIONS command

Each SuperMix analysis starts with an OPTIONS command. The keywords of the OPTIONS command are used to control the estimation procedure and the amount of output to be written at convergence of the iterative procedure. Inclusion of an OPTIONS command in a syntax file is **required**, even if it contains no keywords.

Syntax:

```
OPTIONS <keywords>;
```

If no OPTIONS keywords are given, the default values for these keywords are used. Not all the keywords are available for the different types of dependent variables. The table below summarizes the available keywords and the availability.

Table 9.3: Keywords associated with the OPTIONS command

Keyword	function	Conti-nuous	Ordinal	Nominal	Count	Binary
ACM *	Requests printing of asymptotic covariance matrix.	X	X	X	X	X
BAYES	Requests printing of Bayes results.	X	X	X	X	X
CONVERGE	Sets a value for the test for convergence made at the end of each iteration.	X	X	X	X	X
DEVIANCE *	Provides value of $-2 \ln L$ from a previous analysis to compare fit of nested models.	X				
MAXITER	Indicates the maximum number of iterations to be performed.	X	X	X	X	X
METHOD	Defines the optimization method.	X	X	X	X	X
MODELTERMS	Selects subtracting or adding the model terms from the thresholds.		X			
NFREE *	Indicates the number of free parameters from a previous analysis to compare fit of nested models.	X				

Table 9.3: Keywords associated with the OPTIONS command (continued)

NQUADPTS	Sets the number of quadrature points used for numeric integration.	X	X	X	X	X
REFCAT	Defines whether the first or last category of the outcome should be used as the reference category.			X		
SUMMARY *	Requests printing of summary of hierarchical data structure	X	X	X	X	X

*: Keywords cannot be generated via the GUI. Insert manually in syntax (***.inp**) file.

9.7.30.1 ACM keyword

The ACM keyword is used to print the large-sample covariance matrices of the estimated parameters in the fixed part and random part of the model. Standard errors of the estimated parameters are equal to the square roots of the diagonal elements. The non-duplicated elements of these asymptotic covariance matrices are written to external files with the following default names:

<Output filename>_params.acm

If the output file name is, for example, **kanfer1.out**, then the large-sample covariance matrices are saved to the file **kanfer1_params.acm**.

Syntax:

ACM = <Yes/No>

Default:

No: asymptotic covariance matrices will not be printed.

Example:

In the OPTIONS command below, the ACM keyword is used to request the printing of the asymptotic covariance matrices at convergence. A convergence criterion of 0.0001 is set as the requirement for convergence, and 30 iterations is indicated as the maximum number of iterations to be performed.

```
OPTIONS MAXITER = 30 CONVERGE = 0.0001 ACM = Yes;
```

9.7.30.2 BAYES keyword

This option allows the user to select between writing the (i) Empirical Bayes estimates to an external file, (ii) Empirical Bayes estimates and covariances to an external file, or (iii) suppressing this output.

Syntax:

```
BAYES = <keyword>
```

The keywords include NO, which suppresses the Bayes results; MEANS, which requests the printout of the Bayes estimates of the random effects and COV_MEANS, which requests printing of both the E.B. estimates and (co)variances to external files with extensions ***.bay2** (level-2) and ***.bay3** (level-3). These files may be imported and saved as ***.ss3** files.

Default:

```
BAYES = NO
```

9.7.30.3 CONVERGE keyword

A test for convergence is made at the end of each iteration. If the absolute difference between the estimated parameters and their previous values are all smaller than the convergence criterion, convergence is said to have been reached.

Syntax:

CONVERGE = <value>

Default:

0.0001.

Example:

In order to use a value of, for example, 0.00001 as convergence criterion, the keyword CONVERGE = 0.00001 must be included as part of the OPTIONS command, as shown in the following example:

OPTIONS CONVERGE = 0.00001 MAXITER = 20;

The iterative procedure will terminate if this requirement is met, or if 20 iterations (set with the MAXITER keyword described in Section 9.7.30.5) have been performed without meeting this requirement.

9.7.30.4 DEVIANCE keyword

The DEVIANCE keyword is used to provide the value of – 2 log likelihood as reported in a previous analysis, in order to obtain a χ^2 test statistic for comparing two nested models. The χ^2 statistic is defined as the difference in the deviance statistics for the two models, and has as associated degrees of freedom the difference in the number of parameters estimated in the models compared. It must be accompanied by the NFREE keyword, which is used to indicate the number of parameters estimated in the previous model.

Syntax:

DEVIANCE = <value>

where value equals the deviance ($-2\ln L$) value at convergence printed to the output file of the previous analysis.

Default:

None: no – 2 log likelihood value provided.

Example:

In the OPTIONS command below, the DEVIANCE keyword indicates that a – 2 log likelihood value of 22735.524 was obtained in the previous analysis, and that 44 parameters were estimated (NFREE = 44).

```
OPTIONS NFREE = 44 DEVIANCE = 22735.524;
```

9.7.30.5 MAXITER keyword

The keyword MAXITER is used to indicate the maximum number of iterations to be performed. To change the value via the interface, click in the box and enter the required maximum number of iterations.

Syntax:

```
MAXITER = <value>
```

Default:

100.

The default number of iterations should be sufficient for convergence to be reached in most cases. If, however, a more stringent convergence criterion is used or

previous experience with a particular data set indicates slow convergence, this keyword may be used to increase the maximum number of iterations.

Example:

In the OPTIONS command below, the MAXITER keyword is set to 30, indicating that a maximum number of 30 iterations should be performed. The iterative procedure may terminate before this number is reached if the convergence criterion of 0.00001 (CONVERGE = 0.00001) is met.

```
OPTIONS CONVERGE = 0.00001 MAXITER = 30;
```

9.7.30.6 METHOD keyword

The METHOD option defines the optimization method on the **Advanced** tab.

Syntax:

```
METHOD = <keyword>
```

The keyword is one of the following:

MAP	maximum posterior method
ADAP	adaptive quadrature method
NADAP	non-adaptive quadrature method

Default:

```
METHOD = ADAP
```

Example:

The keyword METHOD = MAP is included as part of the OPTIONS command to request the maximum posterior method, as shown in the following example:

```
OPTIONS CONVERGE = 0.00001 MAXITER = 100 BAYES = Cov_Means
METHOD = MAP;
```

9.7.30.7 MODELTERMS keyword

The MODELTERMS option allows the user to select subtracting or adding the model terms from the thresholds. This option is only available for ordinal outcomes.

Syntax:

```
MODELTERMS = <keyword>
```

The keyword is either SUBTRACT or ADD.

Default:

```
MODELTERMS = SUBTRACT
```

Example:

Consider a level-2 model with a random intercept and assume the outcome variable has 4 categories. For the subtract option

$$\eta_{ijc} = \tau_c - (\mathbf{x}_{ij}^{'}\boldsymbol{\beta} + u_{i0}), \quad i = 1, 2, ..., N; \; j = 1, 2, ..., n_i,$$

and for the add option

$$\eta_{ijc} = \tau_c + (\mathbf{x}_{ij}{}'\boldsymbol{\beta} + u_{i0}), \quad i = 1, 2, ..., N; \; j = 1, 2, ..., n_i.$$

9.7.30.8 NFREE keyword

The NFREE keyword is used to denote the number of free parameters as reported in a previous analysis, in order to obtain a χ^2 test statistic for comparing two nested models. The χ^2 statistic is defined as the difference in the deviance statistics for the two models, and has as associated degrees of freedom the difference in the number of parameters estimated in the models compared. It must be accompanied by the DEVIANCE keyword (see Section 9.7.30.4) which is used to provide the value of – 2 log likelihood as reported in the previous analysis.

Syntax:

NFREE = <number>

where number is the number of free parameters, that is, the total number of parameters estimated during the previous analysis, as reported in the output file.

Default:

None: no parameters indicated for previous model.

Example:

In the OPTIONS command below, the NFREE keyword indicates that 44 parameters were estimated in the previous model, with a – 2 log likelihood value of 22735.524 (DEVIANCE = 22735.524).

```
OPTIONS NFREE = 44 DEVIANCE = 22735.524;
```

9.7.30.9 NQUADPTS keyword

The NQUADPTS keyword is used to define the number of quadrature points (per random-effect dimension) to use in the evaluation of the log-likelihood function and

derivatives using numerical integration. It is usually between 10 and 20 for 1 random effect and 5 to 10 for 2 or 3 effects.

Syntax:

NQUADPTS = <number>

where number is a positive integer.

Default:

NQUADPTS = 10

9.7.30.10 REFCAT keyword

The REFCAT is an option to select whether the first or last category of the outcome should be used as the reference category. It is to be used with nominal outcomes only.

Syntax:

REFCAT = <keyword>

The keyword is either LAST or FIRST.

Default:

REFCAT = first

1.7.1.1 SUMMARY keyword

The SUMMARY keyword is used to suppress the printout of the data summary table.

Syntax:

SUMMARY = <Yes/No>

Default:

Yes: the summary table containing sample sizes of units within the various levels of the hierarchy is printed.

Example:

The OPTIONS command below requests use of the default values for the MAXITER, and CONVERGE keywords, along with suppression of the printing of the summary table as indicated the SUMMARY=NO keyword.

```
OPTIONS SUMMARY=NO MAXITER=10 CONVERGE=0.0001;
```

9.7.31 PREDICTORS command

The PREDICTORS command is used to identify the fixed effects for the model to be analyzed. When the input file is created using the interface dialogs, the PREDICTORS command is automatically generated. This command is entered in the **Variables** tab. The PREDICTORS command is a **required** command for all model types with the exception of the ordinal model. In the ordinal case, only thresholds are estimated if the PREDICTORS command is omitted.

Syntax:

PREDICTORS = <list of covariates>;

Example:

PREDICTORS = intcept;
PREDICTORS = intcept AGE GENDER;

9.7.32 SCALE command

Some sampling distributions, such as the Binomial, Poisson, Gamma, and Inverse Gaussian distributions, have an *optional* scale parameter. Estimation of this parameter is specified by using the SCALE keyword, available form the **Advanced** tab. This is an **optional** command.

Syntax:

SCALE = <keyword>;

The keyword is one of the following:

NONE	No scale estimated
DEVIANCE	Scale estimate based on the deviance statistic
PEARSON	Scale estimate based on the Pearson statistic

Default:

SCALE = none;

9.7.33 THRANDOMn command

The THRANDOMn, n = 2 or 3, command is an **optional** command for ordinal outcomes. It is used to allow for the threshold parameters to vary randomly across the level-2 and level-3 units.

Syntax:

THRANDOMn = <keyword>;

The keyword is one of the following:

NO	Threshold parameters are fixed values
YES	Threshold parameters vary randomly

Default:

THRANDOMn = no;

9.7.34 THRESHOLDSTART command

The THRESHOLDSTART command is an **optional** command for ordinal outcomes. It is used to provide the starting values for the thresholds. The values must be monotonically increasing. The number of thresholds to be estimated is equal to $C-1$ for ordinal outcomes and C if a censor variable is additionally selected. C denotes the number of categories of the outcome variable.

Syntax:

THRESHOLDSTART = <list of values>;

Example:

THRESHOLDSTART = –0.5 1.0 2.0;

9.7.35 THRESHTRANSF command

The THRESHTRANSF command is an **optional** command for ordinal outcomes. It is used to together with the TRANSF_START and TRANSF_END commands to test the null hypothesis that a linear combination of the estimated parameters is equal to zero.

Syntax:

THRESHTRANSF = <list of values>;

Example:

THRESHTRANSF = 1 0 –1;

9.7.36 TITLEn command

The TITLEn command, where n = 1 or 2, allows the user to provide a description of the analysis to be performed. The maximum permissible length of the title for this **optional** command is 70 characters.

Syntax:

TITLEn = <title as provided by the user>;

Default:

No title.

Example:

```
TITLE1 = Level-3 model with design weights;
TITLE2 = Random intercepts;
```

9.7.37 TRANSF_END command

TRANSF_END, together with the TRANSFORMNAMES and TRANSF_START commands, is used to test that a linear combination of the parameters is equal to zero. The command is **optional**.

Syntax:

```
TRANSF_END = <string>;
```

where string denotes the name of the transformation.

Example:

```
TransformNames = H01;
Transf_Start = H01;
FixTransf = 1  0  –1  0;
*Cov2Transf = 0  0  0 ;
*Cov3Transf = 0 ;
*ThreshTransf = 0  0 ;
*FixbyThresh = 0  0  0  0;
Transf_End = H01;
```

*Commands can be omitted if all values in the list are equal to zero.

9.7.38 TRANSF_START command

TRANSF_START, together with the TRANSFORMNAMES and TRANSF_END commands, is used to test that a linear combination of parameters is equal to zero. The command is **optional**.

Syntax:

TRANSF_START = <string>;

where string denotes the name of the transformation.

Example:

```
TransformNames = H01 H02 H03;
Transf_Start = H03;
FixTransf = 0.3  –0.3  –0.3  0.3;
Transf_End= H03;
```

9.7.39 TRANSFORMNAMES command

The TRANSFORMNAMES command is used together with the TRANSF_START and TRANSF_END commands to test that a linear combination of parameters is equal to zero. The command is **optional**.

Syntax:

TRANSFORMNAMES = <list of names>;

where list of names denotes the names of the transformations to be tested. A name should not exceed 8 characters, and should not include blank spaces between characters.

Example:

The TRANSFORMNAMES command below indicates that there are two linear transformations to be tested.

```
TransformNames= transf1 transf2;
Transf_Start= transf1;
Cov2Transf= 1  0  –1 ;
Transf_End= transf1;
Transf_Start= transf2;
Cov2Transf= 1  0  0 ;
Transf_End= transf2;
```

9.7.40 VARNAMES command

The VARNAMES command lists all the variables used in the model. It is a **required** command.

Syntax:

```
VARNAMES = <variable names>;
```

Note that all the variables used in the model, including the outcome variable, response variable(s), IDs, and weight variable(s) must be included in the VARNAMES command. Variable names are case-sensitive.

Example:

```
Varnames =   SCHOOL CLASS POSTTHKS PRETHKS CC TV CCxTV intcept;
```

9.7.41 WEIGHTn command

The WEIGHT command is used to specify design weights for each level of the multilevel model. One WEIGHT command for each level of the hierarchy may be included in the syntax file. For a 2-level model, either or both level-1 and level-2 weights, if available, can be used. Likewise, any combination of weights can be selected for a 3-level model. Use of the command is **optional**.

Syntax:

```
WEIGHTn = <variable name>;
```

where n denotes a positive integer, (1,2,3), for the weight level and <variable name> denotes the name of the weight variable.

Default:

No weights.

Example:

The WEIGHT command shown below indicates the use of the level-1 weighting variable SPWT.

```
WEIGHT1 = SPWT;
```

10 Theory

10.1 A general framework for level-3 linear mixed-effects models

Suppose that y_{ijk} denotes a level-1 outcome variable, where *i* denotes level-3 units ($i = 1, 2, ..., N$), *j* denotes level-2 units (*j* = 1, 2, …, n_i), and *k* denotes level-1 units ($k = 1, 2, ..., n_{ij}$).

Let

$$y_{ijk} = \mathbf{x}'_{ijk}\boldsymbol{\beta} + \mathbf{z}'_{(3)ijk}\mathbf{v}_i + \mathbf{z}'_{(2)ijk}\mathbf{v}_{ij} + \mathbf{z}'_{(1)ijk}\mathbf{e}_{ijk}, \qquad (10.1)$$

where $\boldsymbol{\beta}$ is an $(m \times 1)$ vector of regression coefficients, and where $\mathbf{v}_i$, $\mathbf{u}_{ij}$ and $\mathbf{e}_{ijk}$ denote level-3, level-2, and level-1 random effects respectively. We assume that $\mathbf{v}_1, \mathbf{v}_1, ..., \mathbf{v}_N$ are i.i.d. $N(\mathbf{0}, \boldsymbol{\Phi}_{(3)})$, independent of $\mathbf{v}_{11}, \mathbf{v}_{12}, ..., \mathbf{v}_{Nn_i}$ which are i.i.d. $N(\mathbf{0}, \boldsymbol{\Phi}_{(2)})$. We further assume that the $\mathbf{v}_i$ and $\mathbf{v}_{ij}$ effects are independent of $\mathbf{e}_{111}, \mathbf{e}_{112}, ..., \mathbf{e}_{Nn_i n_{ij}}$ which are i.i.d. $N(\mathbf{0}, \boldsymbol{\Phi}_{(1)})$.

The set of regression equations, *k* = 1, 2, …, n_{ij} defined by (10.1) can be written as

$$\mathbf{y}_{ij} = \mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_{(3)ij}\mathbf{v}_i + \mathbf{Z}_{(2)ij}\mathbf{v}_{ij} + \begin{bmatrix} \mathbf{z}'_{(1)ij1}\mathbf{e}_{ij1} \\ \text{M} \\ \mathbf{z}'_{(1)ijk}\mathbf{e}_{ijk} \\ \text{M} \\ \mathbf{z}'_{(1)ijn_{ij}}\mathbf{e}_{ijn_{ij}} \end{bmatrix}, \qquad (10.2)$$

where y_{ijk}, $\mathbf{x}'_{ijk}$, $\mathbf{z}'_{(3)ijk}$ and $\mathbf{z}'_{(2)ijk}$ are typical rows of $\mathbf{y}_{ij}$, $\mathbf{X}_{ij}$, $\mathbf{Z}_{(3)ij}$ and $\mathbf{Z}_{(2)ij}$. In turn, the set of regression equations, *j* = 1, 2, …, n_i, can be written as

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_{(3)i}\mathbf{v}_i + \sum_{j=1}^{n_i}\mathbf{Z}^*_{(2)ij}\mathbf{v}_{ij} + \sum_{k=1}^{n_{ij}}\mathbf{Z}^*_{(1)is}\mathbf{e}_{is}, \qquad (10.3)$$

where

$$\mathbf{X}_i = \begin{bmatrix} \mathbf{X}_{i1} \\ \mathbf{X}_{i2} \\ \mathsf{M} \\ \mathbf{X}_{in_i} \end{bmatrix}, \quad \mathbf{Z}_{(3)i} = \begin{bmatrix} \mathbf{Z}_{(3)i1} \\ \mathbf{Z}_{(3)i2} \\ \mathsf{M} \\ \mathbf{Z}_{(3)in_i} \end{bmatrix},$$

$$\mathbf{Z}^*_{(2)ij} = \begin{bmatrix} \mathbf{0} \\ \mathsf{M} \\ \mathbf{0} \\ \mathbf{Z}_{(2)ij} \\ \mathbf{0} \\ \mathsf{M} \\ \mathbf{0} \end{bmatrix} \quad \text{and} \quad \mathbf{Z}^*_{(1)is} = \begin{bmatrix} \mathbf{0} \\ \mathsf{M} \\ \mathbf{0} \\ \mathbf{z}'_{(1)is} \\ \mathbf{0} \\ \mathsf{M} \\ \mathbf{0} \end{bmatrix}. \qquad (10.4)$$

Note that for the level-1 part of the model, the double subscript *jk* is replaced by the single subscript *s*, where $s = 1, 2, \ldots, n_i^*$ and $n_i^* = \sum_{j=1}^{n_i} n_{ij}$.

From (10.3) and the distributional assumptions given above, it follows that

$$\mathbf{y}_i : N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i),$$

where

$$\boldsymbol{\mu}_i = \mathbf{X}_i\boldsymbol{\beta}, \qquad (10.5)$$

$$\boldsymbol{\Sigma}_i = \mathbf{Z}_{(3)i}\boldsymbol{\Phi}_{(3)}\mathbf{Z}'_{(3)i} + \boldsymbol{\Lambda}_{(2)i}, \qquad (10.6)$$

$$\boldsymbol{\Lambda}_{(2)i} = Diag\left[\mathbf{Z}_{(2)i1}\boldsymbol{\Phi}_{(2)}\mathbf{Z}'_{(2)i1} + \boldsymbol{\Lambda}_{(1)i1}, \ldots, \mathbf{Z}_{(2)in_i}\boldsymbol{\Phi}_{(2)}\mathbf{Z}'_{(2)in_i} + \boldsymbol{\Lambda}_{(1)in_i}\right] \qquad (10.7)$$

and

$$\mathbf{\Lambda}_{(1)ij} = Diag\left[\mathbf{z}_{(1)ij1}\mathbf{\Phi}_{(1)}\mathbf{z}'_{(1)ij1},...,\mathbf{z}_{(1)ij2}\mathbf{\Phi}_{(1)}\mathbf{z}'_{(1)ijn_{ij}}\right] \tag{10.8}$$

In practice, the number of level-1 units within a specific level-2 unit may be quite large, which leads to $\mathbf{\Sigma}_i$ matrices of very high order. If, for example, there are 100 level-2 units (such as clinics) and nested within each of these units there are 10 level-1 units (for example patients), the order of $\mathbf{\Sigma}_i$ is 1,000. Note that the inversion of a general symmetric matrix or order n requires operations of order n^3 (see *e.g.* Press, *et al.* 2002). It is therefore apparent that further simplification of the likelihood function, derivatives, and Hessian is required if the goal is to implement the theoretical results in SuperMix (see Section 10.1.1 and 10.1.2 where this issue is addressed).

The log-likelihood function of $\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_N$ is

$$\ln L = -\frac{1}{2}\sum_{i=1}^{N}\{n_i \ln 2\pi + \ln\left|\mathbf{\Sigma}_i\right| + tr\mathbf{\Sigma}_i^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i)(\mathbf{y}_i - \boldsymbol{\mu}_i)'\}. \tag{10.9}$$

Instead of maximizing $\ln L$, maximum normal likelihood estimates of the unknown parameters are obtained by minimizing $-\ln L$ with the constant term omitted, *i.e.* by minimizing the function

$$F(\boldsymbol{\gamma}) = \frac{1}{2}\sum_{i=1}^{N}\{\ln\left|\mathbf{\Sigma}_i\right| + tr\mathbf{\Sigma}_i^{-1}\mathbf{G}_y\}, \tag{10.10}$$

where

$$\mathbf{G}_{y_i} = (\mathbf{y}_i - \boldsymbol{\mu}_i)(\mathbf{y}_i - \boldsymbol{\mu}_i)'. \tag{10.11}$$

Its minimum $\frac{\partial F(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} = \mathbf{0}$ yields the normal maximum likelihood estimator $\hat{\boldsymbol{\gamma}}$ of the unknown vector of parameters $\boldsymbol{\gamma}$. We subsequently give a general framework for maximum likelihood estimation of the unknown parameters.

10.1.1 A general optimization framework

Unless the model yields maximum likelihood estimates in closed form, it will be necessary to make use of an iterative procedure to minimize the discrepancy function. The optimization procedure described next (see Browne & du Toit, 1992) is based on the so-called Fisher scoring algorithm, that in the case of structured means and covariances may be regarded as a sequence of Gauss-Newton steps with quantities to be fitted as well as the weight matrix changing at each step. Fisher scoring algorithms require the gradient vector and an approximation to the Hessian matrix. Elements of the gradient vector, $g(\boldsymbol{\gamma})$, and approximate Hessian matrix, $\mathbf{H}(\boldsymbol{\gamma})$, of $F(\boldsymbol{\gamma})$ are given by

$$\frac{\partial F}{\partial \gamma_r} = [g(\boldsymbol{\gamma})]_r = -\sum_{i=1}^{N} \{tr\, \mathbf{Q}_i \frac{\partial \mu_i}{\partial \gamma_r} + \frac{1}{2} tr\, \mathbf{P}_i \frac{\partial \boldsymbol{\Sigma}_i}{\partial \gamma_r}\}, \tag{10.12}$$

where

$$\mathbf{Q}_i = (y_i - \mu_i)' \boldsymbol{\Sigma}_i^{-1} \tag{10.13}$$

$$\mathbf{P}_i = \boldsymbol{\Sigma}_i^{-1} (\mathbf{G}_{y_i} - \boldsymbol{\Sigma}_i) \boldsymbol{\Sigma}_i^{-1}. \tag{10.14}$$

Let

$$[H(\boldsymbol{\gamma})]_{r,s} = -E(\frac{\partial^2 \ln L}{\partial \gamma_r \partial \gamma_s}).$$

Browne and du Toit (1992) showed that

$$[H(\boldsymbol{\gamma})]_{r,s} = \sum_{i=1}^{N} \{tr\left(\frac{\partial \mu_i'}{\partial \gamma_r} \boldsymbol{\Sigma}_i^{-1} \frac{\partial \mu_i}{\partial \gamma_s} \right) + \frac{1}{2}\left(\boldsymbol{\Sigma}_i^{-1} \frac{\partial \boldsymbol{\Sigma}_i}{\partial \gamma_r} \boldsymbol{\Sigma}_i^{-1} \frac{\partial \boldsymbol{\Sigma}_i}{\partial \gamma_s} \right)\}. \tag{10.15}$$

Suppose that $\hat{\gamma}_k$ is the k-th approximation to the $\hat{\boldsymbol{\gamma}}$ that minimizes $F(\boldsymbol{\gamma})$.

Let $\mathbf{g}_k = \mathbf{g}(\boldsymbol{\gamma}_k), \mathbf{H}_k = \mathbf{H}(\boldsymbol{\gamma}_k)$, and $F_k = F(\gamma_k)$. The next approximation is obtained from

$$\hat{\boldsymbol{\gamma}}_{k+1} = \hat{\boldsymbol{\gamma}}_k + \alpha_k \mathbf{d}_k , \tag{10.16}$$

where

$$\mathbf{d}_k = -\mathbf{H}_k^{-1}\mathbf{g}_k \tag{10.17}$$

and α_k is a step size parameter chosen initially as 1 and then successively halved until $F_{k+1} \leq F_k$.

Agresti (1990) pointed out that the Fisher scoring method resembles the Newton-Raphson method, the distinction being that the Fisher scoring uses the expected value of the second derivative matrix.

A convenient feature of the Fisher scoring algorithm is that an estimate, $\{\mathbf{H}(\hat{\boldsymbol{\gamma}})\}^{-1}$ of the asymptotic covariance matrix of estimators γ is available on convergence as a by-product of the calculations.

It can happen that the matrix to be inverted in (10.17) is singular or near singular. An adaptation of the Jennrich and Sampson (1968) stepwise regression procedure may be used to obtain an appropriate conditional inverse. Their procedure for imposing bounds on the estimates may also be employed.

10.1.2 Efficient algorithms for the calculation of derivatives in linear-mixed effects models

The vector $\boldsymbol{\gamma}$ of unknown parameters is

$$\boldsymbol{\gamma} = \begin{bmatrix} vecs\, \boldsymbol{\Phi}_{(3)} \\ vecs\, \boldsymbol{\Phi}_{(2)} \\ vecs\, \boldsymbol{\Phi}_{(1)} \\ \boldsymbol{\beta} \end{bmatrix},$$

where $vecs\, \mathbf{S}$ is a vector of order $p(p+1)/2$ of non-duplicated elements of the $(p \times p)$ matrix $\mathbf{S}$. Suppose, for example, that $p = 3$, then $vecs\, \mathbf{S} = \left(s_{11}, s_{21}, s_{22}, s_{31}, s_{32}, s_{33}\right)'$.

We next illustrate how to obtain computationally efficient expressions for the elements of the gradient vector and information matrix.

For this purpose, we derive $\dfrac{\partial \ln L_i}{\partial \beta_r}$ and $-E\left[\dfrac{\partial^2 \ln L}{\partial \boldsymbol{\Phi}_{(3)rs} \partial \boldsymbol{\Phi}_{(3)uv}}\right]$.

From (10.6), using a well-known matrix identity (see *e.g.* Khatri (1966)), it follows that

$$\boldsymbol{\Sigma}_i^{-1} = \boldsymbol{\Lambda}_{(2)i}^{-1} - \boldsymbol{\Lambda}_{(2)i}^{-1} \mathbf{Z}_{(3)i} \mathbf{C}_{(3)i} \mathbf{Z}_{(3)i}' \boldsymbol{\Lambda}_{(2)i}^{-1}, \tag{10.18}$$

where

$$\mathbf{C}_{(3)i} = \left(\mathbf{Z}_{(3)i}' \boldsymbol{\Lambda}_{(2)i}^{-1} \mathbf{Z}_{(3)i} + \boldsymbol{\Phi}_{(3)}^{-1}\right)^{-1}. \tag{10.19}$$

Note that $\mathbf{C}_{(3)i}$ is of order p_3, where p_3 denotes the number of level-3 random effects.

From (10.12) it follows that

$$\frac{\partial \ln L}{\partial \beta_r} = \sum_{i=1}^{N} tr\left(\mathbf{y}_i - \boldsymbol{\mu}_i\right)' \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \frac{\partial \boldsymbol{\beta}}{\partial \beta_r}. \tag{10.20}$$

Let $\mathbf{y}_i - \boldsymbol{\mu}_i = \mathbf{e}_i$, then (cf. (10.18))

$$\frac{\partial \ln L}{\partial \beta_r} = \sum_{i=1}^{N} tr\left(\mathbf{e}_i' \boldsymbol{\Lambda}_{(2)i}^{-1} \mathbf{X}_i \mathbf{J}_{r1} + \mathbf{e}_i' \boldsymbol{\Lambda}_{(2)i}^{-1} \mathbf{Z}_{(3)i} \mathbf{C}_{(3)i} \mathbf{Z}_{(3)i}' \boldsymbol{\Lambda}_{(2)i}^{-1} \mathbf{X}_i \mathbf{J}_{r1}\right),$$

where $\mathbf{J}_{r1}$ is a column vector with all elements equal to zero, with the exception of the *i*-th element, which is equal to unity.

A well-known result for the trace operator is that $tr\mathbf{A}\mathbf{J}_{r1} = \left[\mathbf{A}\right]_{1,r}$. Use of this result and (10.7) gives

$$\frac{\partial \ln L}{\partial \boldsymbol{\beta}'} = \sum_{i=1}^{N} \left(\sum_{j=1}^{n_{ij}} \mathbf{e}_{ij}' \boldsymbol{\Lambda}_{(2)ij}^{-1} \mathbf{X}_{ij} + \left[\sum_{j=1}^{n_{ij}} \mathbf{e}_{ij}' \boldsymbol{\Lambda}_{(2)ij}^{-1} \mathbf{Z}_{(3)ij} \right] \mathbf{C}_{(3)i} \left[\sum_{j=1}^{n_{ij}} \mathbf{Z}_{(3)ij}' \boldsymbol{\Lambda}_{(2)ij}^{-1} \mathbf{X}_{ij} \right] \right). \tag{10.21}$$

Each of the terms in (10.21) can be further simplified by noting that

$$\boldsymbol{\Lambda}_{(2)ij}^{-1} = \boldsymbol{\Lambda}_{(1)ij}^{-1} - \boldsymbol{\Lambda}_{(1)ij}^{-1} \mathbf{Z}_{(2)ij} \mathbf{C}_{(2)ij} \mathbf{Z}_{(2)ij}' \boldsymbol{\Lambda}_{(1)ij}^{-1}, \tag{10.22}$$

where the $p_2 \times p_2$ matrix $\mathbf{C}_{(2)ij}$ is given by

$$\mathbf{C}_{(2)ij} = \left[\mathbf{Z}_{(2)ij}' \boldsymbol{\Lambda}_{(1)ij}^{-1} \mathbf{Z}_{(2)ij} + \boldsymbol{\Phi}_{(2)}^{-1} \right]^{-1}, \tag{10.23}$$

and where p_2 denotes the number of level-2 random effects.

For example,

$$\mathbf{Z}'_{(3)ij}\boldsymbol{\Lambda}^{-1}_{(2)ij}\mathbf{X}_{ij} = \sum_{k=1}^{n_{ij}} \mathbf{z}_{(3)ijk}\boldsymbol{\Lambda}^{-1}_{(1)ijk}\mathbf{x}'_{ijk} - \left[\sum_{j=1}^{n_{ijk}} \mathbf{z}_{(3)ijk}\boldsymbol{\Lambda}^{-1}_{(1)ijk}\mathbf{z}'_{(2)ijk}\right]\mathbf{C}_{(2)ij}\left[\sum_{j=1}^{n_{ijk}} \mathbf{z}_{(2)ijk}\boldsymbol{\Lambda}^{-1}_{(1)ijk}\mathbf{x}'_{ijk}\right]. \tag{10.24}$$

It is therefore essential to obtain computationally efficient expressions for the first and second order derivatives. As illustrated above, only matrix inversions of order p_k, where p_k denotes the number of random effects on level k, k = 1, 2, 3 are required.

From (10.15) it follows that

$$-E\left[\frac{\partial^2 \ln L}{\partial\boldsymbol{\Phi}_{(3)rs}\partial\boldsymbol{\Phi}_{(3)uv}}\right] = \sum_{i=i}^{N} tr\left\{\boldsymbol{\Sigma}_i^{-1}\frac{\partial\boldsymbol{\Sigma}_i}{\partial\boldsymbol{\Phi}_{(3)rs}}\boldsymbol{\Sigma}_i^{-1}\frac{\partial\boldsymbol{\Sigma}_i}{\partial\boldsymbol{\Phi}_{(3)uv}}\right\},$$

where

$$tr\left\{\boldsymbol{\Sigma}_i^{-1}\frac{\partial\boldsymbol{\Sigma}_i}{\partial\boldsymbol{\Phi}_{(3)rs}}\boldsymbol{\Sigma}_i^{-1}\frac{\partial\boldsymbol{\Sigma}_i}{\partial\boldsymbol{\Phi}_{(3)uv}}\right\} = tr\left\{\mathbf{Z}'_{(3)i}\boldsymbol{\Sigma}_i^{-1}\mathbf{Z}_{(3)i}\mathbf{D}_{rs}\mathbf{Z}_{(3)i}\boldsymbol{\Sigma}_i^{-1}\mathbf{Z}'_{(3)i}\mathbf{D}_{uv}\right\}. \tag{10.25}$$

The result (10.25) follows since

$$\frac{\partial\boldsymbol{\Sigma}_i}{\partial\boldsymbol{\Phi}_{(3)rs}} = \mathbf{Z}_{(3)i}\frac{\partial\boldsymbol{\Phi}_{(3)}}{\partial\boldsymbol{\Phi}_{(3)rs}}\mathbf{Z}'_{(3)i},$$

where

$$\mathbf{D}_{rs} = \frac{\partial\boldsymbol{\Phi}_{(3)}}{\partial\boldsymbol{\Phi}_{(3)rs}} = \mathbf{J}_{rs} + (1-\delta_{rs})\mathbf{J}_{sr},$$

and where $tr\mathbf{AB} = tr\mathbf{BA}$. The scalar δ_{rs} equals 1 if $r = s$ and 0 otherwise.

$\mathbf{Z}'_{(3)i}\boldsymbol{\Sigma}_i^{-1}\mathbf{Z}_{(3)i}$ can be evaluated if we substitute $\boldsymbol{\Sigma}_i^{-1}$ by the right-hand side of (10.18). That is

$$\mathbf{Z}'_{(3)i}\boldsymbol{\Sigma}_i^{-1}\mathbf{Z}_{(3)i} = \mathbf{Z}'_{(3)i}\boldsymbol{\Lambda}_{(2)i}^{-1}\mathbf{Z}_{(3)i} - \mathbf{Z}'_{(3)i}\boldsymbol{\Lambda}_{(2)i}^{-1}\mathbf{Z}_{(3)i}\mathbf{C}_{(3)i}\mathbf{Z}'_{(3)i}\boldsymbol{\Lambda}_{(2)i}^{-1}\mathbf{Z}_{(3)i}. \qquad (10.26)$$

Similarly

$$\mathbf{Z}'_{(3)i}\boldsymbol{\Lambda}_{(2)i}^{-1}\mathbf{Z}_{(3)i} = \sum_{j=1}^{n_i}\mathbf{Z}'_{(3)ij}\boldsymbol{\Lambda}_{(2)ij}^{-1}\mathbf{Z}_{(3)ij},$$

where

$$\mathbf{Z}'_{(3)ij}\boldsymbol{\Lambda}_{(2)ij}^{-1}\mathbf{Z}_{(3)ij} = \mathbf{Z}'_{(3)ij}\boldsymbol{\Lambda}_{(1)ij}^{-1}\mathbf{Z}_{(3)ij} - \mathbf{Z}'_{(3)ij}\boldsymbol{\Lambda}_{(1)ij}^{-1}\mathbf{Z}_{(3)ij}\mathbf{C}_{(2)ij}\mathbf{Z}'_{(3)ij}\boldsymbol{\Lambda}_{(1)ij}^{-1}\mathbf{Z}_{(3)ij}, \qquad (10.27)$$

and where

$$\mathbf{Z}'_{(3)ij}\boldsymbol{\Lambda}_{(1)ij}^{-1}\mathbf{Z}_{(3)ij} = \sum_{k=1}^{n_{ij}}\mathbf{z}_{(3)ijk}\boldsymbol{\Lambda}_{(1)ij}^{-1}\mathbf{z}'_{(3)ijk}. \qquad (10.28)$$

10.1.3 Patterned structures for random effects covariance matrices

Suppose y_{ijk} is an outcome variable corresponding to the *k*-th level-1 unit (k = 1, 2, …, n_{ij}) nested within the *j*-th level-2 unit (j = 1, 2, …, n_i), which, in turn, is nested within the *i*-th level-3 unit (i = 1, 2, …, N).

A general formulation for the linear mixed-effects model is

$$y_{ijk} = \mathbf{x}'_{(f)ijk}\boldsymbol{\beta} + \mathbf{z}'_{(3)ijk}\mathbf{v}_i + \mathbf{z}'_{(2)ijk}\mathbf{v}_{ij} + \mathbf{z}'_{(1)ijk}\mathbf{e}_{ijk},$$

where $\boldsymbol{\beta}$ is a $r\times 1$ vector of population parameters, $\mathbf{v}_i$ a $p\times 1$ vector of level-3 random effects, $\mathbf{v}_{ij}$ a $q\times 1$ vector of level-2 random effects, and $\mathbf{e}_{ijk}$ an $s\times 1$ vector of level-1 random effects.

We denote the random effects covariance matrices by $\mathbf{\Phi}_{(3)}$, $\mathbf{\Phi}_{(2)}$ and $\mathbf{\Phi}_{(1)}$ respectively. Let $\boldsymbol{\gamma}' = \left(\boldsymbol{\beta}', \boldsymbol{\sigma}'_{(3)}, \boldsymbol{\sigma}'_{(2)}, \boldsymbol{\sigma}'_{(1)}\right)$ denote the vector of unknown parameters of order *k*, where $\boldsymbol{\sigma}_{(l)}$, l = 1, 2, 3 is a vector formed from the non-duplicated elements of $\mathbf{\Phi}_{(l)}$. In general, the number of parameters is $k = r + p^* + q^* + s^*$, where $p^* = \frac{1}{2} p(p+1)$, $q^* = \frac{1}{2} q(q+1)$ and $s^* = \frac{1}{2} s(s+1)$.

In typical level-3 models

$$\mathbf{\Phi}_{(1)} = \begin{bmatrix} \sigma_e^2 & & & \\ & \sigma_e^2 & & \\ & & \text{O} & \\ & & & \sigma_e^2 \end{bmatrix},$$

that is, we assume that the level-1 error variances are homogeneous. In practical applications, this need not be the case and it may be more realistic to assume that the level-1 error variances are heterogeneous, that is,

$$\mathbf{\Phi}_{(1)} = \begin{bmatrix} \sigma_{11}^2 & & & \\ & \sigma_{22}^2 & & \\ & & \text{O} & \\ & & & \sigma_{ss}^2 \end{bmatrix}.$$

In the first case, the number of unknown parameters reduces to $k = r + p^* + q^* + 1$ and in the second case to $k = r + p^* + q^* + s$.

The ability to impose constraints on the elements of the random effects covariance matrices offers users great flexibility in the type of models that can be analyzed, and moreover, can lead to substantial savings in computational time.

Example

Suppose that the number of random effects at level-2 of a mixed-effects model equals 4, that is $\mathbf{\Phi}_{(2)}$ has 10 non-duplicated elements. Suppose further that there is reason to believe that the effects have equal variances and equal covariances (compound symmetry model),

$$\mathbf{\Phi}_{(2)} = \begin{bmatrix} \phi_{11} & & & \\ \phi_{22} & \phi_{11} & & \\ \phi_{22} & \phi_{22} & \phi_{11} & \\ \phi_{22} & \phi_{22} & \phi_{22} & \phi_{11} \end{bmatrix},$$

then $\mathbf{\Phi}_{(2)}$ have only 2 unique elements. It follows from Section 10.1.1 that at iteration n, the increment vector $\mathbf{d}_n$ is obtained as a solution to a set of k simultaneous equations

$$\mathbf{H}_n \mathbf{d}_n = \mathbf{g}_n, \tag{10.29}$$

where $\mathbf{H}_n$ and $\mathbf{g}_n$ are the approximate Hessian matrix and gradient vector with respect to the vector of unknown parameters $\boldsymbol{\gamma}$ respectively.

Equality constraints

Consider an arbitrary set of equations $\mathbf{Ax} = \mathbf{b}$,

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + a_{14}x_4 &= b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + a_{24}x_4 &= b_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + a_{34}x_4 &= b_3 \\ a_{41}x_1 + a_{42}x_2 + a_{43}x_3 + a_{44}x_4 &= b_4. \end{aligned} \tag{10.30}$$

Suppose we constrain x_2 to be equal to x_3, then

$$\begin{aligned} &a_{11}x_1 + (a_{12} + a_{13})x_2 + a_{14}x_4 = b_1 \\ &(a_{21} + a_{31})x_1 + (a_{22} + a_{23} + a_{32} + a_{33})x_2 + (a_{24} + a_{34})x_4 = b_2 + b_3 \\ &a_{41}x_1 + (a_{42} + a_{43})x_3 + a_{44}x_4 = b_4. \end{aligned} \qquad (10.31)$$

Likewise, it follows that constraining elements of $\boldsymbol{\gamma}$ to be equal involves the addition of rows and corresponding columns of $\mathbf{H}_n$ and the addition of rows of $\mathbf{g}_n$. The result is a reduction in the order of $\mathbf{H}_n$ and $\mathbf{g}_n$, so that the inversion of $\mathbf{H}_n$ (see (10.17), Section 10.1.1) and the calculation of the increment vector $\mathbf{d}_n = \mathbf{H}_n^{-1}\mathbf{g}_n$ are computationally efficient.

Constraining parameters to be equal to zero

Suppose, on the other hand, that we wish to impose a stationary moving average process of order 1 on $\boldsymbol{\Phi}_{(2)}$, where $q = 5$

$$\boldsymbol{\Phi}_{(2)} = \begin{bmatrix} \phi_{11} & & & & \\ \phi_{22} & \phi_{11} & & & \\ 0 & \phi_{22} & \phi_{11} & & \\ 0 & 0 & \phi_{22} & \phi_{11} & \\ 0 & 0 & 0 & \phi_{22} & \phi_{11} \end{bmatrix},$$

then $\boldsymbol{\Phi}_{(2)}$ contains 6 zero elements.

To illustrate how one would handle the pattern described above, consider the set of simultaneous equations (10.30), but suppose that x_2 and x_4 are constrained to be equal to 0. Elimination of rows 2, 4, and columns 2 and 4 from the coefficient matrix $\mathbf{A}$ gives

$$\begin{aligned} a_{11}x_1 + a_{13}x_3 &= b_1 \\ a_{31}x_1 + a_{33}x_3 &= b_3. \end{aligned}$$

Likewise, constraining elements of $\boldsymbol{\gamma}$ to be equal to zero involves the elimination of rows and corresponding columns from the approximate Hessian matrix **H** and elements from the gradient vector **g**.

Advantage of the element-wise calculation of H and g

In situations where elements of the covariance matrices $\boldsymbol{\Phi}_{(3)}$, $\boldsymbol{\Phi}_{(2)}$, and $\boldsymbol{\Phi}_{(1)}$ are constrained to be equal, significant reductions in computation time and storage requirements can be obtained if typical elements of **H** and **g** are computed rather than matrix expressions for **H** and **g**.

Example

Consider the MA(1) process for $\mathbf{v}_{ij}$ described in Section 10.1.3.3. If we compute the gradient element-wise, we need not compute

$$\frac{\partial \ln L}{\partial \phi_{31}}, \frac{\partial \ln L}{\partial \phi_{41}}, \frac{\partial \ln L}{\partial \phi_{42}}, \frac{\partial \ln L}{\partial \phi_{51}}, \frac{\partial \ln L}{\partial \phi_{52}}, \text{ and } \frac{\partial \ln L}{\partial \phi_{53}}.$$

Using $-E\left[\frac{\partial^2 \ln L}{\partial \gamma_\alpha \partial \gamma_\beta}\right]$ to approximate the Hessian, we need not compute the 21 elements $-E\left[\frac{\partial^2 \ln L}{\partial \phi_{31} \partial \phi_{31}}\right]$, $-E\left[\frac{\partial^2 \ln L}{\partial \phi_{31} \partial \phi_{32}}\right]$, …, $-E\left[\frac{\partial^2 \ln L}{\partial \phi_{43} \partial \phi_{43}}\right]$.

To impose equality constraints, and to constrain parameters to be equal to zero or to a fixed value, SuperMix uses CovnPat and FixPat commands (see Chapter 9). To illustrate, suppose that we wish to impose a structure on the covariance matrix of the level-3 random effects and that there are 4 random effects at this level. Using the convention of numbering the non-duplicated elements of a symmetric matrix row-wise, it follows that the elements of $\boldsymbol{\Phi}_{(3)}$ are numbered as shown below.

```
1
2 3
4 5 6
7 8 9 10
```

The elements to be fixed are then replaced by a "0". If, for example, the matrix is constrained to be diagonal, the covariance pattern is

```
1
0 3
0 0 6
0 0 0 10
```

The structure as specified indicates that there are four parameters to be estimated (that is, the variances of the random effects) and five parameters that are fixed (that is, the covariances). By default, parameters elements indicated by a zero are set equal to zero, unless the user overrides these values by specifying his/her own set of fixed values.

Parameters constrained to be equal all have the same number starting with the smallest number. *E.g.* if $\phi_{22} = \phi_{33}$, the number corresponding to ϕ_{33} (that is 6) is replaced with "3".

Examples

1. Toeplitz: number of level-3 random effects equals 5

```
 1
 2 1
 4 2 1
 7 4 2 1
11 7 4 2 1
```

This pattern is equivalent to the covariance structure

$$\mathbf{\Phi}_{(3)} = \begin{bmatrix} \gamma_0 & & & & \\ \gamma_1 & \gamma_0 & & & \\ \gamma_2 & \gamma_1 & \gamma_0 & & \\ \gamma_3 & \gamma_2 & \gamma_1 & \gamma_0 & \\ \gamma_4 & \gamma_3 & \gamma_2 & \gamma_1 & \gamma_0 \end{bmatrix},$$

that is $\text{cov}(v_{ir}, v_{is}) = \gamma_{(1r-s1)}$.

The Toeplitz structure is a general representation of a stationary ARMA process (see *e.g.* Box & Jenkins, 1976).

2. Block diagonal: Number of level-2 random effects equals 6.

Consider the covariance pattern

```
1
2 3
0 0 6
0 0 9 10
0 0 0  0 15
0 0 0  0 20 21
```

The pattern shown above specifies that (v_{ij1}, v_{ij2}), (v_{ij3}, v_{ij4}) and (v_{ij5}, v_{ij6}) are 3 sets of uncorrelated random variables. Within each set, the covariance term is non-zero. This pattern is equivalent to the covariance structure

$$\mathbf{\Phi}_{(2)} = \begin{bmatrix} \phi_{11} & & & & & \\ \phi_{21} & \phi_{22} & & & & \\ 0 & 0 & \phi_{33} & & & \\ 0 & 0 & \phi_{43} & \phi_{44} & & \\ 0 & 0 & 0 & 0 & \phi_{55} & \\ 0 & 0 & 0 & 0 & \phi_{65} & \phi_{66} \end{bmatrix}.$$

10.1.4 Use of dummy variables in longitudinal studies

To simplify the presentation, we consider growth-curves with only intercepts and slopes. We begin with a consideration of growth curves for individuals and then extend the discussion to the more complex case in which growth curves for both members of a dyad, in which the members of the dyad are distinguished from each other, (*e.g.*, husband or wife) are of interest.

Consider the first few records of a typical multilevel data set for the analysis of growth in which the systolic blood pressure (y) of wives are measured over five annual assessments (x, ranging from 0 at the first assessment to 4 at the fifth assessment so that the intercept reflects average blood pressure at the first assessment)

Case	Occasion	*y*	Intercep	*x(slope)*
1	1	110	1	0
1	2	112	1	1
1	3	115	1	2
1	4	118	1	3
2	1	98	1	0
2	2	102	1	1
2	3	103	1	2
2	4	105	1	3
2	5	106	1	4
3	1	103	1	0
3	3	108	1	2
3	5	111	1	4

Note that occasions 1, 2, 3, 4, and 5 correspond to time-linked x-values of 0, 1, 2, 3, and 4, respectively. Furthermore, the data are unbalanced, in the sense that some wives have missing measurements. For example, wife 1 has missing data from the fifth occasion, and wife 3 has missing data from the second and fourth occasions.

Let N denote the number of cases and n_i the number of occasions for case i. The standard intercept-and-slopes-as-outcome-variables model is

$$y_{ij} = \beta_{0i} + \beta_{1i} x_{ij} + e_{ij}, \quad \begin{matrix} i = 1, 2, \mathsf{K}\ , N; \\ j = 1, 2, \mathsf{K}\ , n_i \end{matrix} \tag{10.32}$$

where it is assumed that on level 2 (cases 1, 2, …, N)

$$\beta_{0i} = \beta_0 + v_{0i}$$
$$\beta_{1i} = \beta_1 + v_{1i}.$$

It is further assumed that $e_{i1}, e_{i2}, \ldots, e_{in_i}$ are independently and identically distributed as $N(0, \sigma_e^2)$ random variables. It is also assumed that $(v_{01}, v_{11})', (v_{02}, v_{12})', \ldots, (v_{0N}, v_{1N})'$ are independently and identically distributed as $N(\mathbf{0}, \mathbf{\Phi})$ and that the level-2 random effects (v_{01}, v_{1i}) and e_{ij} are independent.

For case *i*, the set of regression equations (10.32) can be written in matrix notation as

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta}_i + \mathbf{e}_i, \tag{10.33}$$

where $\mathbf{y}_i' = (y_{i1,} y_{i2,} \ldots, y_{in_i})$, $\boldsymbol{\beta}_i' = (\beta_{0i}\ \beta_{1i})'$, $\mathbf{e}_i' = (e_{i1}, e_{i2}, \ldots, e_{in_i})$, and

$$\mathbf{X}_i = \begin{bmatrix} 1 & x_{i1} \\ 1 & x_{i2} \\ \mathsf{M} & \mathsf{M} \\ 1 & x_{in_i} \end{bmatrix}.$$

From the distributional assumptions above, it follows that

$$\boldsymbol{\mu}_i = E(\mathbf{y}_i) = \mathbf{X}_i \boldsymbol{\beta} \tag{10.34}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1)'$ and

$$\mathbf{\Sigma}_i = Cov(\mathbf{y}_i) = \mathbf{X}_i \mathbf{\Phi} \mathbf{X}_i' + \sigma^2 \mathbf{I}_{n_i}. \tag{10.35}$$

From (10.34) and (10.35) one can compute the likelihood function (L_i) for case *i*, and hence

$$-2 \log L = -2 \sum_{i=1}^{N} \log L_i. \tag{10.36}$$

Usually, a χ^2- test statistic for testing the fit of this model against the saturated model is not available for unbalanced data.

Next, we show how one can introduce dummy variables so that a χ^2-test statistic can be obtained for testing the hypothesis that a linear growth model provides an adequate fit to the data versus the alternative hypothesis of saturated means and covariances. The alternative model implies that no structure is imposed on the elements of $\boldsymbol{\mu}_i$ (see (10.34)) or on the elements of $\boldsymbol{\Sigma}_i$. For balanced data, it is well known that $\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$ and $\hat{\boldsymbol{\Sigma}} = \mathbf{S}$ (sample covariance matrix). We show how $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ can be estimated in longitudinal models with unbalanced data. We also discuss additional covariance structures that can be imposed on the error variances $e_{i1}, e_{i2}, \ldots, e_{in_i}$.

The data set below shows the original dataset augmented with five dummy variables OCC1, OCC2, …, OCC5 created as follows:

Occasion	OCC1	OCC2	OCC3	OCC4	OCC5
1	1	0	0	0	0
2	0	1	0	0	0
3	0	0	1	0	0
4	0	0	0	1	0
5	0	0	0	0	1

Case	Occ.	y	Intercep	x	OCC1	OCC2	OCC3	OCC4	OCC5
1	1	110	1	0	1	0	0	0	0
1	2	112	1	1	0	1	0	0	0
1	3	115	1	2	0	0	1	0	0
1	4	118	1	3	0	0	0	1	0
2	1	98	1	0	1	0	0	0	0
2	2	102	1	1	0	1	0	0	0
2	3	103	1	2	0	0	1	0	0
2	4	105	1	3	0	0	0	1	0
2	5	106	1	4	0	0	0	0	1
3	1	103	1	0	1	0	0	0	0
3	3	108	1	2	0	0	1	0	0
3	5	111	1	4	0	0	0	0	1

The following model is equivalent to model (10.32)

$$y_{ij} = \beta_{0i} + \beta_{1i} x_{ij} + \mathrm{OCC}_{ij} \times e_{ij},$$

and since

$$\begin{aligned} \beta_{0i} &= \beta_0 + v_{0i} \\ \beta_{1i} &= \beta_1 + v_{1i} \end{aligned}$$

it follows that

$$\mathbf{v}_i^{*'} = (v_{0i}, v_{1i}, e_{i1}, e_{i2}, e_{i3}, e_{i4}, e_{i5})'. \qquad (10.37)$$

This model can be written in matrix notation as a level-2 model with no random component on level-1:

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i^* \mathbf{v}_i^*, \qquad (10.38)$$

where

$$\mathbf{Z}_i^* = \begin{bmatrix} 1 & z_{i1} & 1 & 0 & 0 & 0 & 0 \\ 1 & z_{i2} & 0 & 1 & 0 & 0 & 0 \\ 1 & z_{i3} & 0 & 0 & 1 & 0 & 0 \\ 1 & z_{i4} & 0 & 0 & 0 & 1 & 0 \\ 1 & z_{i5} & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

and $\mathbf{v}_i^{*'} = (v_{0i}, v_{1i}, e_{i1}, e_{i2}, e_{i3}, e_{i4}, e_{i5})'$.

If a specific measurement was unavailable on occasion *j*, then $\mathbf{Z}_i^*$ is defined as above, with the corresponding row removed. Rewriting model (10.33) as model (10.38) enables one to impose more general covariance structures on the error variances. From (10.38) it follows that

$$\boldsymbol{\Sigma}_i = Cov(\mathbf{y}_i) = \mathbf{Z}_i^* \boldsymbol{\Phi}^* \mathbf{Z}_i^{*'}. \qquad (10.39)$$

It should be noted that $\mathbf{\Sigma}_i$ is a $n_i \times n_i$ matrix, and therefore has $n_i(n_i+1)/2$ non-duplicated elements. However, $\mathbf{\Phi}^*$ is a $(n_i \times k)(n_i \times k)$ matrix, where k = 2, and therefore has $(n_i+k)(n_i+k+1)/2$ non-duplicated elements. These elements cannot be uniquely estimated unless constraints are imposed on the elements of $\mathbf{\Phi}^*$ so that the number of free parameters is less or equal to $n_i(n_i+1)/2$.

Homogeneous level-1 variances

Model (10.38) is exactly equivalent to model (10.33) if

$$\mathbf{\Phi}^* = \begin{bmatrix} \phi_{11} & & & & & & \\ \phi_{21} & \phi_{22} & & & & & \\ 0 & 0 & \phi_{33} & & & & \\ 0 & 0 & 0 & \phi_{44} & & & \\ 0 & 0 & 0 & 0 & \phi_{55} & & \\ 0 & 0 & 0 & 0 & 0 & \phi_{66} & \\ 0 & 0 & 0 & 0 & 0 & 0 & \phi_{77} \end{bmatrix}$$

where

$$\begin{aligned} \phi_{11} &= \text{var}(v_{0i}) \\ \phi_{21} &= \text{cov}(v_{0i}, v_{1i}) \\ \phi_{22} &= \text{var}(v_{1i}) \end{aligned}$$

and where

$$\phi_{33} = \phi_{44} = ... = \phi_{77} = \text{var}(e_{ij})$$

This model can be fitted using the following pattern for the level-2 random effects covariance matrix. The number of free parameters is 2 (fixed part) + 4 (random part) = 6.

```
1
2 3
0 0 6
0 0 0 6
0 0 0 0 6
0 0 0 0 0 6
0 0 0 0 0 0 6
```

Note that the 6-th element of the covariance matrix is ϕ_{33} and therefore the 10-th element (ϕ_{44}), 15-th element (ϕ_{55}), etc. are constrained to be equal to ϕ_{33}. The $-2\ln L$ value for the saturated model (also called the deviance statistic) allows one to calculate χ^2-statistics for testing models against a model with saturated mean and covariance structure. This model is defined by

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i^*\mathbf{v}_i^*$$

where $\mathbf{Z}_i^*$ is an $n_{ij} \times n_{ij}$ identity matrix with rows deleted if no measurements for the occasions corresponding to these rows are available. The number of free parameters is 5 (fixed part) + 5(5 + 1) / 2 (random part) = 20. Use of dummy variables as described above yields identical estimators of the population mean and population covariance matrix than those obtained using multiple imputation (see *e.g.* Schafer, 1997).

Heterogeneous level-1 error variances

Use of the following covariance structure enables one to estimate heterogeneous level-1 error variances:

```
1
2 3
0 0 6
0 0 0 10
0 0 0 0 15
0 0 0 0 0 21
0 0 0 0 0  0 28
```

The number of free parameters is 2 (fixed part) + 8 (random part) = 10.

Intra-class correlation structure for the level-1 variances

To specify equal level-1 variances and covariances over time, one can use the following covariance pattern:

```
1
2 3
0 0 6
0 0 9 6
0 0 9 9 6
0 0 9 9 9 6
0 0 9 9 9 9 6
```

In the examples above, we assumed that the $\mathbf{e}_i$ and $\mathbf{v}_i$ are uncorrelated. Clearly, one can use the covariance patterns to introduce correlation between these variables as well. This property is important when dealing with crossed and nested designs.

Structure of the data set

Consider the first few records of a typical multilevel data set for the analysis of growth in which blood pressure measurements (y) for both husbands and wives are measured over five annual assessments (x, ranging from 0 at the first assessment to 4 at the fifth assessment).

		y	Husband		Wife	
Couple	Occ.	Pressure	Intcep	x(slope)	Intcep	x(slope)
1	1	110	1	0	0	0
1	2	108	1	1	0	0
1	3	109	1	2	0	0
1	4	118	1	3	0	0
1	5	117	1	4	0	0
1	6	111	0	0	1	0
1	7	106	0	0	1	1
1	8	101	0	0	1	2
1	9	109	0	0	1	3
1	10	104	0	0	1	4
2	1	117	1	0	0	0
2	6	141	0	0	1	1
3	1	119	1	0	0	0

3	2	84	1	1	0	0
3	6	107	0	0	1	0
3	7	89	0	0	1	1

Because the couple is the unit of analysis and both spouses are nested within couple, there are now 10 possible occasions for each couple, five for the husband, and five for the wife. Further, each blood pressure count is defined as belonging to either the husband or to the wife so that a husband's intercept and slope score will have entries of 0 for the wife, and the wife's intercept and slope will have entries of 0 for the husband. As in the earlier example, the data here are also unbalanced because spouses from some couples had data at only some assessments. For example, for couple 2, husband and wife contributed data at only the first assessment, and for couple 3, husband and wife contributed data for only the first two assessments.

Because the couple now has 10 occasions, 10 dummy variables (H1 to H5 for the husbands' five occasions and W1 to W5 for the wives' five occasions) are needed to define each occasion. Thus, the full data set, augmented with the set of dummy variables looks as follows:

			Husband		Wife		Husband					Wife				
Couple	Occ.	Pres.	Intcep	x	Intcep	x	H1	H2	H3	H4	H5	W1	W2	W3	W4	W5
1	1	110	1	0	0	0	1	0	0	0	0	0	0	0	0	0
1	2	108	1	1	0	0	0	1	0	0	0	0	0	0	0	0
1	3	109	1	2	0	0	0	0	1	0	0	0	0	0	0	0
1	4	118	1	3	0	0	0	0	0	1	0	0	0	0	0	0
1	5	117	1	4	0	0	0	0	0	0	1	0	0	0	0	0
1	6	111	0	0	1	0	0	0	0	0	0	1	0	0	0	0
1	7	106	0	0	1	1	0	0	0	0	0	0	1	0	0	0
1	8	101	0	0	1	2	0	0	0	0	0	0	0	1	0	0
1	9	109	0	0	1	3	0	0	0	0	0	0	0	0	1	0
1	10	104	0	0	1	4	0	0	0	0	0	0	0	0	0	1
2	1	117	1	0	0	0	1	0	0	0	0	0	0	0	0	0
2	6	141	0	0	1	1	0	0	0	0	0	1	0	0	0	0
3	1	119	1	0	0	0	1	0	0	0	0	0	0	0	0	0
3	2	84	1	1	0	0	0	1	0	0	0	0	0	0	0	0
3	6	107	0	0	1	0	0	0	0	0	0	1	0	0	0	0
3	7	89	0	0	1	1	0	0	0	0	0	0	1	0	0	0

The covariance pattern below allows for correlation between wives' and husbands' intercept and slope coefficients but restricts the model to homogeneous level-1 error variances and uncorrelated level-1 errors.

```
1
2  3
4  5  6
7  8  9 10
0  0  0  0 15
0  0  0  0  0 15
0  0  0  0  0  0 15
0  0  0  0  0  0  0 15
0  0  0  0  0  0  0  0 15
0  0  0  0  0  0  0  0  0 15
0  0  0  0  0  0  0  0  0  0 15
0  0  0  0  0  0  0  0  0  0  0 15
0  0  0  0  0  0  0  0  0  0  0  0 15
```

The number of free parameters is 4 (fixed part) + 11 (random part) = 15.

Heteroscedastic level-1 error variances, correlated errors between spouses, and autoregressive errors for wives.

A researcher can also examine improvement in model fit when a lag-1 process is modeled with the aid of dummy variables. This is a common way to model autocorrelated errors (*e.g.*, Sivo & Willson, 1998; Willett & Sayer, 1994) in which errors at one time point influence those from only the next immediate time point (*e.g.*, the error associated with a Time 1 assessment is related to the error in the Time 2 assessment, the error associated with a Time 2 assessment is related to the error in the Time 3 assessment, and so on). As seen in the covariance pattern below, the four errors between adjacent assessments (*i.e.*, between year 1-year 2, year 2-year 3, year 3-year 4, and year 4-year 5) were allowed to correlate. Because there was no reason to expect these lags to differ in strength, they were constrained to be equal.

```
Hint 1
Hslp 2  3
Wint 4  5  6
Wslp 7  8  9 10
H1   0  0  0  0 15
H2   0  0  0  0  20 21
H3   0  0  0  0  0  20 28
H4   0  0  0  0  0  0  20 36
H5   0  0  0  0  0  0  0  20 45
W1   0  0  0  0 50  0  0  0   0 55
W2   0  0  0  0  0 50  0  0   0 65 66
W3   0  0  0  0  0  0 50  0   0  0 65 78
W4   0  0  0  0  0  0  0 50   0  0  0 65 91
W5   0  0  0  0  0  0  0  0  50  0  0  0 65 105
```

The number of free parameters is 4 (fixed part) + 23 (random part) = 27. Again, the deviance statistic associated with this model can be used to compare this model with the previous model.

10.1.5 The use of dummy variables in multivariate response models

In health services research, researchers often have data sets containing more than one response variable. A typical example is counts of inpatient (y_1), outpatient (y_2) and emergency room (y_3) visits for mental health care. There is thus a need for software that can fit multivariate response variables to a linear mixed-effects model. It turns out that, with the use of dummy variables, a multivariate level-2 model can be fitted to the data using a 3-level model with a single response variable and no level-1 random effects.

For the variables y_1 , y_2 , and y_3 considered above, the following represents a typical data set, where x_1 = *depressive severity* and x_2 = *type of insurance coverage* (coded 1 for public and –1 for private). Missing values are coded –9.

Clinic	Patient	y1	y2	y3	x1	x2
1	1	0	2	1	22	-1
1	2	1	3	-9	30	1
1	3	-9	2	1	26	-1
2	1	0	1	1	23	-1
2	2	0	2	0	29	1
2	3	0	1	1	26	1
2	4	1	2	1	33	-1

To create a data set that can be analyzed with the SuperMix level-3 module, dummy variables are created for each response variable in the data set. For the example above this translates to three dummy-coded variables: $d_k = 1$ if y_k is measured, $k = 1, 2, 3$, and 0 otherwise. Using these dummy variables we construct a new data set, shown below for clinic number 1, patients 1, 2, and 3.

Clinic	Patient	y	d1	d2	d3	x1*d1	x1*d2	x1*d3	x2*d1	x2*d2	x2*d3
1	1	0	1	0	0	22	0	0	-1	0	0
1	1	2	0	1	0	0	22	0	0	-1	0
1	1	1	0	0	1	0	0	22	0	0	-1
1	2	1	1	0	0	30	0	0	1	0	0
1	2	3	0	1	0	0	30	0	0	1	0
1	3	2	0	1	0	0	26	0	0	-1	0
1	3	1	0	0	1	0	0	26	0	0	-1

…

In the level-3 framework, *y* is the response variable, d1, d2, d3, x1*d1, …., x2*d3 are typical rows of the fixed-effects design matrix. The fixed effects part consists of intercept coefficients (corresponding to d1, d2, and d3), slope coefficients for depressive severity (corresponding to x1*d1, x1*d2, and x1*d3), and insurance coverage coefficients (corresponding to x2*d1, x2*d2, and x2*d3). Alternatively, one can use depression and insurance as level-2 covariates in which case the data set (shown for Clinic 1, Patient 1 only) has the form

Clinic	Patient	y	d1	d2	d3	x1	x2
1	1	0	1	0	0	22	-1
1	1	2	1	1	0	22	-1
1	1	1	1	0	1	22	-1

The difference between the two approaches is that in the first approach, different slopes are assumed for the three service utilization outcome variables, whereas we assume equal slopes for depression and equal slopes for insurance type in the second approach.

Theoretical Framework

Suppose that there are q response variable and let y_{ijk} denote the k-th response for the (i,j)-the unit.

The multivariate response model to be considered in this section (see du Toit (1995)) is defined by

$$y_{ijk} = \mathbf{x}_{ijk}^{'}\boldsymbol{\beta} + \left[\mathbf{v}_i\right]_k + \left[\mathbf{v}_{ij}\right]_k \qquad (10.40)$$

where $i = 1,2,...,N; \quad j = 1,2,...,n_i$ and $k \in \{1,2,...,q\}$.

Assume that the $q \times 1$ random vectors $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_N$ are i.i.d. with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Phi}_{(3)}$, independently distributed of the $q \times 1$ i.i.d. random vectors $\mathbf{v}_{i1}, \mathbf{v}_{i2}, ..., \mathbf{v}_{in_i}$, which have, mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Phi}_{(2)}$.

$\mathbf{x}_{ijk}^{'} : 1 \times s$ is a typical row of the design matrix of the fixed part of the model, the elements being values of the s predictors. The elements of $\mathbf{v}_i$ and $\mathbf{v}_{ij}$ make provision for variation of responses over level-3 and level-2 units respectively. Note that no allowance is made for level-1 variation, since there are no true experimental units below level-2.

The set of equations given in (10.40) can be written in matrix notation as

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_{(3)i}\mathbf{v}_i + \sum_{j=1}^{n_i} \mathbf{Z}_{(2)i}\mathbf{v}_{ij}$$

where $\mathbf{X}_i$ has typical row $\mathbf{x}_{ijk}^{'}$,

$$\mathbf{Z}_{(3)i} = \begin{bmatrix} \mathbf{S}_{i1} \\ \mathrm{M} \\ \mathbf{S}_{ij} \\ \mathrm{M} \\ \mathbf{S}_{in_i} \end{bmatrix}, \tag{10.41}$$

$\mathbf{Z}_{(2)ij}$ is a $\left(\sum_{j=1}^{n_i} n_{ij} \right) \times m$ matrix partitioned as

$$\mathbf{Z}_{(2)ij} = \begin{bmatrix} \mathbf{0} \\ \mathrm{M} \\ \mathbf{0} \\ \mathbf{S}_{ij} \\ \mathbf{0} \\ \mathrm{M} \\ \mathbf{0} \end{bmatrix} \tag{10.42}$$

and $\mathbf{S}_{ij}$ is a selection matrix consisting of a subset of the rows of the $q \times q$ identity matrix $\mathbf{I}_q$ where the rows of $\mathbf{S}_{ij}$ correspond to the response measurements available for the (i, j)-th unit.

As an example of how the $\mathbf{S}_{ij}$ matrices are constructed, consider the measurement of six plasma lipid variables, $y_1, y_2, ..., y_6$. For the case where all six response measurements are available $\mathbf{S}_{ij} = \mathbf{I}_q$. If, however, only measurements on y_2 and y_4 are available,

$$\mathbf{S}_{ij} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

Under the distributional assumptions given above, it follows that

$$E(\mathbf{y}_i) = \mathbf{X}_i\boldsymbol{\beta}$$

and

$$\begin{aligned} &Cov(\mathbf{y}_i, \mathbf{y}_i^{'}) = \boldsymbol{\Sigma}_i \\ &= \mathbf{Z}_{(3)i}\boldsymbol{\Phi}_{(3)}\mathbf{Z}_{(3)i}^{'} + \sum_{j=1}^{n_i}\mathbf{Z}_{(2)ij}\boldsymbol{\Phi}_{(2)}\mathbf{Z}_{(2)ij}^{'}, \end{aligned} \qquad (10.43)$$

with $\mathbf{X}_i$, $\mathbf{Z}_{(3)i}$ and $\mathbf{Z}_{(2)ij}$ as defined by (10.41) and (10.42).

10.1.6 The use of dummy variables for fitting 4-level regression models

Consider a clinical study designed to measure the impact of hormone therapy on memory and cognition in elderly women. Suppose that 50 hospitals (level-4 units) participated in the study. For each of the hospitals, data are available for 5 types of hormone treatments (level-3 units) obtained from the female patients (level-2 units) who were tested twice a year for a period of up to 6 years (level-1 units).

Let y_{ijkl} denote a cognition score at occasion *l* for patient *k* on treatment *j* at hospital *i*.

A typical mixed-effects model for data of this type is

$$y_{ijkl} = \beta_0 + \beta_1 x1_{ijkl} + \beta_2 x2_{ijkl} + ... + \beta_r xr_{ijkl} + w_i + v_{ij} + u_{ijk} + e_{ijkl}, \qquad (10.44)$$

where w_i denotes a level-4 (hospital level) variance component, v_{ij} a level-3 (treatment level) variance component, u_{ijk} a level-2 (patients) variance component, and e_{ijkl} denotes the level-1 measurement error. It is further assumed that there are *r* covariates $x_1, x_2, ..., x_r$ (such as age, weight and percentage fat) that may influence the cognition score.

The set of regression equations (10.44) can be rewritten as

$$\mathbf{y}_{ijkl} = \mathbf{X}_{ikl}\boldsymbol{\beta} + \mathbf{Z}_{(3)ikl}\mathbf{v}_i^* + \mathbf{Z}_{(2)ikl}\mathbf{v}_{ik} + \mathbf{Z}_{(1)ikl}\mathbf{e}_{ikl}, \tag{10.45}$$

where

$$\mathbf{X}_{ikl} = \begin{bmatrix} 1 & x_{1,i1kl} & \mathsf{K} & x_{r,i1kl} \\ 1 & x_{1,i2kl} & \mathsf{K} & x_{r,i2kl} \\ 1 & x_{1,i3kl} & \mathsf{K} & x_{r,i3kl} \\ 1 & x_{1,i4kl} & \mathsf{K} & x_{r,i4kl} \\ 1 & x_{1,i5kl} & \mathsf{K} & x_{r,i5kl} \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \\ \mathsf{M} \\ \beta_r \end{bmatrix}, \tag{10.46}$$

$$\mathbf{Z}_{(3)ikl}\mathbf{v}_i^* = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} w_i \\ v_{i1} \\ v_{i2} \\ v_{i3} \\ v_{i4} \\ v_{i5} \end{bmatrix}, \tag{10.47}$$

$$\mathbf{Z}_{(2)ikl}\mathbf{v}_{ik} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} v_{i1k} \\ v_{i2k} \\ v_{i3k} \\ v_{i4k} \\ v_{i5k} \end{bmatrix}, \tag{10.48}$$

and

$$\mathbf{Z}_{(1)ijkl}\mathbf{e}_{ikl} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} e_{i1kl} \\ e_{i2kl} \\ e_{i3kl} \\ e_{i4kl} \\ e_{i5kl} \end{bmatrix}. \qquad (10.49)$$

We note that, except for column 1 of the design matrix $\mathbf{Z}_{(3)}$, the remaining columns correspond to dummy variables $T_1, T_2, ..., T_5$ where $T_j = 1$ if treatment number is j and 0 otherwise. If only treatments 2, 3, and 5 are available at hospital i, the design matrices $\mathbf{X}_{ikl}, \mathbf{Z}_{(3)ikl}, \mathbf{Z}_{(2)ikl}$ and $\mathbf{Z}_{(1)ikl}$ are defined by (10.46) to (10.49), but with rows 1 and 4 removed. For example,

$$\mathbf{Z}_{(3)ikl} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

For the level-3 model (10.45) to be equivalent to the level-4 model (10.44), the following patterned covariance specifications (see Chapter 9) are required.

Level-3 covariance pattern

```
1
0 3
0 0 3
0 0 0 3
0 0 0 0 3
0 0 0 0 0 3
```

The advantage of this presentation is that one can allow for cross-level correlation(s). For example, if there is reason to believe that there is differences in the way patients react to the treatments due to some hospital effect, then we may want to assume that $\text{cov}(w_i, v_{ij}) \neq 0$. These covariance terms may be included in the model by using the following covariance pattern:

```
1
2 3
2 0 3
2 0 0 3
2 0 0 0 3
2 0 0 0 0 3
```

Level-2 covariance pattern

```
1
0 1
0 0 1
0 0 0 1
0 0 0 0 1
```

Level-1 covariance pattern

```
1
0 1
0 0 1
0 0 0 1
0 0 0 0 1
```

Since measurement error may be associated with the type of treatment administered, the assumption of homogeneous level-1 error variances may not be realistic and one may want to use a covariance pattern for heterogeneous error variances as described in Section 10.1.3.

10.1.7 Testing of contrasts (linear transforms) in mixed-effects models

Consider a clinical trial in which two types of drugs are administered to 400 obese adults. Adults are randomly assigned to four groups:

- Group 1, Drug A, low dosage (10 mg/day)
- Group 2, Drug A, high dosage (50 mg/day)
- Group 3, Drug B, low dosage (10 mg/day)
- Group 4, Drug B, high dosage (50 mg/day)

Let y_{ij} denote weight loss of subject i on occasion t_j, i = 1, 2, …, 400 and j = 1, 2, …, n_i, and let

$$y_{ij} = \beta_1 \text{AL} + \beta_2 \text{AH} + \beta_3 \text{BL} + \beta_4 \text{BH} + \beta_5 \text{TIJ} + \beta_6 \text{AGE} + \beta_7 \text{GENDER} + \beta_8 \text{INITW} + v_{i1} + \text{TIJ} \times v_{i2} + e_{ij} \tag{10.50}$$

where AL, AH, BL and BH are dummy variables, coded as follows

	AL	AH	BL	BH
Drug A, low dosage	1	0	0	0
Drug A, high dosage	0	1	0	0
Drug B, low dosage	0	0	1	0
Drug B, high dosage	0	0	0	1

In model (10.50), β_1, β_2, β_3, and β_4 represent the average group loss (or gain) in weight over the study period if we control for a subject's age (AGE), gender (GENDER), weight at the onset of the trial (INITW), and time (TIJ) at which the weight loss (y_{ij}) measurement was made.

Visual inspection of the estimated β -coefficients may point to significant differences between the different treatments. The construction of contrasts or linear functions of the parameters is a useful statistical analysis tool and enables the researcher to perform hypothesis testing concerning the equality of subsets of parameters.

In the example above, the fixed part of the model has 8 parameters $\beta_1, \beta_2, \ldots, \beta_8$. We may want to test the following 3 hypotheses:

$$H_{01} : \beta_1 = \beta_2$$
$$H_{02} : \beta_1 = \beta_3$$
$$H_{03} : \beta_1 = \beta_4.$$

Each of these hypotheses can alternatively be written as

$$H_{01}: 1\beta_1 - 1\beta_2 + 0\beta_3 + 0\beta_4 + 0\beta_5 + 0\beta_6 + 0\beta_7 + 0\beta_8 = 0$$
$$H_{02}: 1\beta_1 + 0\beta_2 - 1\beta_3 + 0\beta_4 + 0\beta_5 + 0\beta_6 + 0\beta_7 + 0\beta_8 = 0$$
$$H_{03}: 1\beta_1 + 0\beta_2 + 0\beta_3 - 1\beta_4 + 0\beta_5 + 0\beta_6 + 0\beta_7 + 0\beta_8 = 0$$

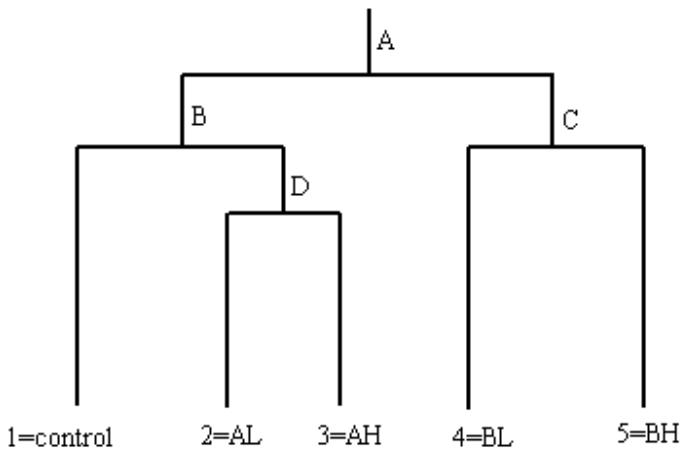
or, in matrix notation,

$$H_0: \mathbf{C\beta} = \mathbf{0},$$

where

$$\mathbf{C} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Suppose that an additional 100 subjects (the control group) are also assigned to the experiment, but each subject from this group receives a placebo. Suppose further that the 5 treatments are hypothesized to be related as described by the tree diagram



Here we can form the orthogonal contrasts:

	Treatments								
Contrast	1	2	3	4	5	TIJ	AGE	GENDER	INITW
A	1/3	1/3	1/3	-1/2	-1/2	0	0	0	0
B	1	-1/2	-1/2	0	0	0	0	0	0
C	0	0	0	1	-1	0	0	0	0
D	0	1	-1	0	0	0	0	0	0

A complex hypothesis about several elements of the vector of fixed coefficients $\boldsymbol{\beta}$ can be tested if use is made of a $p \times m$ contrast matrix $\mathbf{C}$, with p the number of contrasts and m the number of fixed coefficients. The hypothesis is written in the form

$$\mathbf{C}\boldsymbol{\beta} = \mathbf{k},$$

where $\mathbf{k}$ is a known vector, usually $\mathbf{k} = \mathbf{0}$.

For large samples (see *e.g.* du Toit, 1993), $\mathbf{C}\hat{\boldsymbol{\beta}}$ has an approximate $N\left(\mathbf{C}\boldsymbol{\beta}, \mathbf{C}\boldsymbol{\Gamma}^{-1}\mathbf{C}'\right)$ distribution, where $\boldsymbol{\Gamma} = Cov(\hat{\boldsymbol{\beta}})$. The elements of $\boldsymbol{\Gamma}^{-1}$ can be obtained from $-E\left(\dfrac{\partial^2 \log L}{\partial \boldsymbol{\beta}\, \partial \boldsymbol{\beta}'}\right)$ (see Section 10.1.1). If the hypothesis $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{k}$ is true, it follows (see, *e.g.* Anderson (2003)), that

$$U = \left(\mathbf{C}\boldsymbol{\beta} - \mathbf{k}\right)' \left[\mathbf{C}\boldsymbol{\Gamma}^{-1}\mathbf{C}'\right]^{-1} \left(\mathbf{C}\boldsymbol{\beta} - \mathbf{k}\right) \qquad (10.51)$$

has an approximate χ^2-distribution with p degrees of freedom.

A set of $100\,(1-\alpha)\%$ simultaneous confidence intervals for the p elements of $\mathbf{C}\boldsymbol{\beta}$ is given by the p intervals

$$\mathbf{c}_i' \hat{\boldsymbol{\beta}} \pm \left[\mathbf{c}_i' \boldsymbol{\Gamma}^{-1} \mathbf{c}_i \chi^2_{m,\alpha}\right]^{0.5},$$

where $p \le m$, $\mathbf{c}_i^{'}$ denotes the *i*-th row of $\mathbf{C}$ and $\chi^2_{m,\alpha}$ is the critical value of the χ^2 distribution with *m* degrees of freedom.

10.2 Distribution models and link functions

10.2.1 Introduction

It is assumed that y_{ijk} is an outcome variable, where $i = 1, 2, ..., N$ denotes level-3 units and $j = 1, 2, ..., n_i$ denotes level-2 units, nested within each level-3 unit i. The level-1 units $k = 1, 2, ..., n_{ij}$ are nested within the (i, j) -th (level-3; level-2) combination.

For 2-level models, the subscript i is omitted and y_{jk} denotes level-1 unit k nested within level-2 unit j.

A multilevel model with a non-normal outcome variable is transformed to a linear model by using a *link* function which defines the relationship between the dependent variable η_{ijk} of the linear model and the mean μ_{ijk} of the *distribution* selected. More specifically, the linear model of a multilevel generalized linear model is given by

$$\eta_{ijk} = \mathbf{x}_{ijk}^{'}\boldsymbol{\beta} + \mathbf{z}_{(2)ijk}^{'}\mathbf{v}_{ij} + \mathbf{z}_{(3)ijk}^{'}\mathbf{v}_{i},$$

where $\mathbf{x}_{ijk}$ is a $p \times 1$ vector of predictors, $\mathbf{z}_{(2)ijk}$ is a $q \times 1$ design vector associated with the level-2 random effects $\mathbf{v}_{ij}$. Likewise, $\mathbf{z}_{(3)ijk}$ is a $r \times 1$ design vector associated with the level-3 random effects $\mathbf{v}_i$. Typically, the elements of $\mathbf{z}_{(3)ijk}$ and $\mathbf{z}_{(2)ijk}$ are subsets of the elements of $\mathbf{x}_{ijk}$.

It is further assumed that the level-3 and level-2 random effect vectors are uncorrelated and also that $\mathbf{v}_i : N\left(\mathbf{0}, \boldsymbol{\Phi}_{(3)}\right)$ and that $\mathbf{v}_{ij} : N\left(\mathbf{0}, \boldsymbol{\Phi}_{(2)}\right)$.

10.2.2 Link function and derivatives

The link functions available are the log, logistic, complimentary log-log, log-log, and probit. Table 10.1 contains a summary of these link functions and their derivatives. The cumulative distribution for each link is denoted by $CDF(\eta)$ and the corresponding probability distribution function by PDF, where $PDF = \frac{\partial}{\partial \eta} CDF$. The second-order derivatives of η with respect to the link function is denoted by $\frac{\partial}{\partial \eta} PDF$. The CDF of a standardized normal variable is denoted by $\Phi(\cdot)$, while $c1 = \exp(-\eta)$, and $c2 = \exp(\eta) = \frac{1}{c1}$.

Table 10.1: Probability and cumulative distribution functions

Function	$CDF(\eta)$	$PDF(\eta)$	$\frac{\partial}{\partial \eta} PDF.$
Logistic	$\frac{1}{1+c1}$	$CDF(1-CDF)$	$(c1-c2)\times PDF^2$
Probit	$\Phi(\eta)$	$\frac{1}{\sqrt{2\pi}}\exp-\frac{1}{2}\eta^2$	$-\eta \times PDF$
Complementary log-log	$1-\exp(-c2)$	$c2(1-CDF)$	$(1-c2)PDF$
Log-log	$\exp(-c1)$	$c1\times CDF$	$(c1-1)PDF$
Log	$c2$	$c2$	$c2$

In subsequent sections, short descriptions of the different distribution-link type models are given.

10.2.3 The Poisson-log model

Assume y_{ijk} follows a Poisson distribution with mean μ_{ijk}. In other words, the probability density function of y_{ijk} is given by

$$f\left(y_{ijk},\mu_{ijk}\right)=\frac{e^{-\mu_{ijk}}\mu_{ijk}^{y_{ijk}}}{y_{ijk}!}\Rightarrow \ln f\left(y_{ijk},\mu_{ijk}\right)=y_{ijk}\ln\left\{\mu_{ijk}\right\}-\mu_{ijk}-\ln\left\{y_{ijk}!\right\} \quad (10.52)$$

and the variance of y_{ijk} is given by

$$\sigma^2\left(y_{ijk}\right)=\mu_{ijk} \quad (10.53)$$

Suppose further that the following exponential model is imposed on the means of y_{ijk}

$$\mu_{ijk}=\exp\left(\eta_{ijk}\right) \quad (10.54)$$

The model in (10.54) is transformed to a linear model by using the log link function. In other words

$$\eta_{ijk}=\ln\left(\mu_{ijk}\right) \quad (10.55)$$

10.2.4 Models for the Bernoulli sampling distribution

Sampling distribution

$$f(y_{ijk})=\mu_{ijk}^{y_{ijk}}(1-\mu_{ijk})^{1-y_{ijk}} \quad (10.56)$$

Variance

$$\sigma^2(y_{ijk})=\mu_{ijk}(1-\mu_{ijk}) \quad (10.57)$$

10.2.4.1 The logistic model

Model for means

$$\mu_{ijk} = \frac{1}{1+\exp\left(-\eta_{ijk}\right)} \tag{10.58}$$

Link function

$$\eta_{ijk} = \text{logit}(\mu_{ijk}) = \ln\left\{\frac{\mu_{ijk}}{1-\mu_{ijk}}\right\} \tag{10.59}$$

10.2.4.2 The complementary log-log model

Model for means

$$\mu_{ijk} = 1-\exp\left\{-\exp\left(\eta_{ijk}\right)\right\} \tag{10.60}$$

Link function

$$\eta_{ijk} = \ln(-\ln(1-\mu_{ijk})) \tag{10.61}$$

10.2.4.3 The probit model

Model for means

$$\mu_{ijk} = \Phi(\eta_{ijk}) \tag{10.62}$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard Normal distribution.

Link function

$$\eta_{ijk} = \Phi^{-1}(\mu_{ijk}) \tag{10.63}$$

10.2.4.4 The log-log model

Model for means

$$\mu_{ijk} = \exp\left\{-\exp\left(-\eta_{ijk}\right)\right\} \tag{10.64}$$

Link function

$$\eta_{ijk} = -\ln\left\{-\ln\left(\mu_{ijk}\right)\right\} \tag{10.65}$$

10.2.5 Models for the Binomial distribution

Sampling distribution

Let y_{ijk} denote the proportion of successes in n_{ijk} independent trials:

$$f(y_{ijk}) = \begin{pmatrix} n_{ijk} \\ n_{ijk}y_{ijk} \end{pmatrix} \mu_{ijk}^{n_{ijk}y_{ijk}} \left(1-\mu_{ijk}\right)^{n_{ijk}\left(1-y_{ijk}\right)} \tag{10.66}$$

Variance

$$\sigma^2(y_{ijk}) = \frac{\mu_{ijk}\left(1-\mu_{ijk}\right)}{n_{ijk}} \tag{10.67}$$

The models for the means and the link functions are identical to those of the Bernoulli-logit model described in Section 10.2.4.

10.2.6 The Negative Binomial-log model

Sampling distribution

$$f(y_{ijk}) = \frac{\Gamma\left(y_{ijk} + \frac{1}{\psi}\right)}{\Gamma\left(y_{ijk}+1\right)\Gamma\left(\frac{1}{\psi}\right)} \frac{\left(\psi\mu_{ijk}\right)^{y_{ijk}}}{\left(1+\psi\mu_{ijk}\right)^{y_{ijk}+\frac{1}{\psi}}} \tag{10.68}$$

Variance

$$\sigma^2\left(y_{ijk}\right) = \mu_{ijk} + \psi\mu_{ijk}^2 \tag{10.69}$$

The model for means and the link function are identical to those of the Poisson-log model described in Section 10.2.3.

10.2.7 The Gamma-log model

Sampling distribution

$$f(y_{ijk}) = \frac{1}{\Gamma\left(\frac{1}{\psi}\right) y_{ijk}} \left(\frac{y_{ijk}}{\mu_{ijk}\psi}\right)^{\frac{1}{\psi}} \exp\left(-\frac{y_{ijk}}{\mu_{ijk}\psi}\right) \tag{10.70}$$

Variance

$$\sigma^2\left(y_{ijk}\right) = \psi\mu_{ijk}^2 \tag{10.71}$$

The model for means and the link function are identical to those of the Poisson-log model described in Section 10.2.3.

10.2.8 The Inverse Gaussian-log model

Sampling distribution

$$f(y_{ijk}) = \frac{1}{\sqrt{2\pi y_{ijk}^3 \psi}} \exp\left(-\frac{1}{2y_{ijk}} \left(\frac{y_{ijk} - \mu_{ijk}}{\mu_{ijk}} \right)^2 \Big/ \psi \right) \tag{10.72}$$

Variance

$$\sigma^2 \left(y_{ijk} \right) = \psi \mu_{ijk}^3 \tag{10.73}$$

The model for means and the link function are identical to those of the Poisson-log model described in Section 10.2.3.

10.2.9 Models for the Multinomial sampling distribution

Sampling distribution

$$f\left(y_{ijk,1}, y_{ijk,2}, \ldots, y_{ijk,C-1} \right) = \frac{n_{ijk}!}{\left(\prod_{l=1}^{C-1} y_{ijk,l}! \right) \left(n_{ijk} - \sum_{l=1}^{C-1} y_{ijk,l} \right)!} \left(\prod_{l=1}^{C-1} \mu_{ijk,l}^{y_{ijk,l}} \right) \mu_{ijk,C}^{n_{ijk} - \sum_{l=1}^{C-1} y_{ijk,l}} \tag{10.74}$$

Covariance matrix

$$\boldsymbol{\Sigma}(\mathbf{y}_{ijk}^*) = \mathbf{D}_{\mu_{ijk}} - \boldsymbol{\mu}_{ijk} \boldsymbol{\mu}_{ijk}' \tag{10.75}$$

where $\mathbf{y}_{ijk}^* = \left[y_{ijk,1} \quad y_{ijk,2} \quad \ldots \quad y_{ijk,C-1} \right]'$ and $\mathbf{D}_{\mu_{ijk}}$ denotes a $(C-1) \times (C-1)$ diagonal matrix with the elements of $\boldsymbol{\mu}_{ijk} = \left[\mu_{ijk,1} \quad \mu_{ijk,2} \quad \ldots \quad \mu_{ijk,C-1} \right]'$ on the diagonal.

10.2.9.1 The generalized logistic (nominal) Model

Model for means

$$\mu_{ijk,l} = \frac{\exp\left\{\eta_{ijk,l}\right\}}{1+\sum_{l=1}^{C-1}\exp\left\{\eta_{ijk,l}\right\}} \quad \forall \quad l = 1,2,\mathsf{L}\ ,C-1 \tag{10.76}$$

Link function

$$\eta_{ijk,l} = \text{logit}(\mu_{ijk,l}) = \ln\left\{\frac{\mu_{ijk,l}}{\mu_{ijk,C}}\right\} \tag{10.77}$$

10.2.9.2 The cumulative logistic (ordinal) model

Model for means

$$\mu^{*}_{ijk,l} = \sum_{r=1}^{l}\mu_{ijk,r} = \frac{\exp\left\{\tau_l - \eta_{ijk,l}\right\}}{1+\exp\left\{\tau_l - \eta_{ijk,l}\right\}} \quad \forall \quad l = \mathsf{L}\ ,C-1 \tag{10.78}$$

$$= \frac{1}{1+\exp\left\{-\eta^{*}_{ijk}\right\}}$$

where

$$\eta^{*}_{ijk,l} = \tau_l - \eta_{ijk,l}$$

the elements of $\tau_1,\ \tau_2,\ \mathsf{L}\ ,\tau_{C-1}$ denote threshold parameters.

Link function

$$\eta^*_{ijk,l} = \text{clogit}(\mu^*_{ijk,l}) = \ln\left\{\frac{\mu^*_{ijk,l}}{1-\mu^*_{ijk,l}}\right\} \tag{10.79}$$

10.2.9.3 The proportional hazards (cumulative complimentary log-log) model

Model for means

$$\mu^*_{ijk,l} = \sum_{r=1}^{l} \mu_{ijk,r} = 1 - \exp\left(-\exp\left\{\eta^*_{ijk,l}\right\}\right) \quad \forall \quad l = 1,2,\text{L} \ ,C-1 \tag{10.80}$$

$$\eta^*_{ijk,l} = \text{cloglog}(\mu^*_{ijk,l}) = \ln\left(-\ln\left(1-\mu^*_{ijk,l}\right)\right) \tag{10.81}$$

10.2.9.4 The cumulative log-log model

Model for means

$$\mu^*_{ijk,l} = \sum_{r=1}^{l} \mu_{ijk,r} = \exp\left(-\exp\left\{-\eta^*_{ijk,l}\right\}\right) \quad \forall \quad l = 1,2,\text{L} \ ,C-1 \tag{10.82}$$

$$\eta^*_{ijk,l} = \text{loglog}(\mu^*_{ijk,l}) = -\ln\left(-\ln\left(\mu^*_{ijk,l}\right)\right) \tag{10.83}$$

10.2.9.5 The cumulative probit model

Model for means

$$\mu^*_{ijk,l} = \sum_{r=1}^{l} \mu_{ijk,r} = \ \Phi\left(\eta^*_{ijk,l}\right) \quad \forall \quad l = 1,2,\text{L} \ ,C-1 \tag{10.84}$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution.

Link function

$$\eta^*_{ijk,l} = \Phi^{-1}\left(\mu^*_{ijk,l}\right) \tag{10.85}$$

10.2.10 The estimation of scale and dispersion parameters

A number of sampling distributions discussed in the previous sections have a dispersion parameter and/or a scale parameter. A summary of these distributions with respect to dispersion and scale parameters and their estimates is shown in Table 10.2.

Table 10.2: Scale and dispersion parameters

Distribution	Deviance	Dispersion	Pearson	Scale
Binomial	x		x	x
Gamma	x	x	x	x
Inverse Gaussian	x	x	x	x
Negative binomial	x	x	x	
Poisson	x		x	x

10.2.10.1 The deviance χ^2 estimate

$$\hat{\phi}_D = \sqrt{\frac{\chi^2_D}{d}} \tag{10.86}$$

$$\chi^2_D = 2\ln L(\mathbf{y} \mid \mathbf{y}) - 2\ln L(\hat{\boldsymbol{\mu}} \mid \mathbf{y}) \tag{10.87}$$

$$d = \sum_{i=1}^{N} \sum_{j=1}^{n_i} \sum_{k=1}^{n_{ij}} w_{ijk} - q \tag{10.88}$$

10.2.10.2 The Pearson χ^2 estimate

$$\hat{\phi}_P = \sqrt{\frac{\chi_P^2}{d}} \tag{10.89}$$

$$\chi_P^2 = \sum_{i=1}^{N} \sum_{j=1}^{n_i} \sum_{k=1}^{n_{ij}} \frac{w_{ijk} \left(y_{ijk} - \hat{\mu}_{ijk} \right)^2}{\hat{\sigma}^2 \left(y_{ijk} \right)} \tag{10.90}$$

10.3 Theoretical aspects: level-3 generalized linear models

10.3.1 Notation

Let $\mathbf{y}_{ij}$ denote a $n_{ij} \times 1$ vector of outcomes with typical element y_{ijk}, where i denotes the level-3 units, j denotes the level-2 units nested within the i-th level-3 unit and k denotes the level-1 units nested within ij.

Assume further that there are N level-3 units so that $i = 1, 2, ..., N$. Within a typical level-3 unit there are n_i level-2 units, $j = 1, 2, ..., n_i$ and nested within ij there are n_{ij} level-1 units so that $k = 1, 2, ..., n_{ij}$. There are, therefore, $\sum_{i=1}^{N} n_i$ level-2 units and $\sum_{i=1}^{N} \sum_{j=1}^{n_i} n_{ij}$ level-1 units.

Let $\mathbf{y}_i^*$ and $\mathbf{v}_i^*$ denote $\sum_{j=1}^{n_i} n_{ij} \times 1$ vectors partitioned as follows:

$$\mathbf{y}_i^* = \begin{pmatrix} \mathbf{y}_{i1} \\ \mathbf{y}_{i2} \\ \mathsf{M} \\ \mathbf{y}_{in_i} \end{pmatrix}; \quad \mathbf{v}_i^* = \begin{pmatrix} \mathbf{v}_{i1} \\ \mathbf{v}_{i2} \\ \mathsf{M} \\ \mathbf{v}_{in_i} \end{pmatrix}, \quad i = 1, 2, \mathsf{L} \quad N$$

Under the assumption that

$$\mathbf{v}_{i1}, \mathsf{L} \ , \mathbf{v}_{in_i} \qquad i.i.d. \qquad N\left(\mathbf{0}, \mathbf{\Phi}\right) \tag{10.91}$$

and

$$\mathbf{v}_i, \qquad i = 1, \mathsf{L} \ , N \qquad i.i.d. \qquad N\left(\mathbf{0}, \mathbf{\Psi}\right) \tag{10.92}$$

with $\operatorname{cov}\left(\mathbf{v}_{ij}^{'}, \mathbf{v}_i\right) = \mathbf{0}$, it follows that

$$\begin{aligned} f\left(\mathbf{y}_i^*, \mathbf{v}_i^*, \mathbf{v}_i\right) &= f\left(\mathbf{y}_i^* \mid \mathbf{v}_i^*, \mathbf{v}_i\right) \cdot g\left(\mathbf{v}_i^*, \mathbf{v}_i\right) \\ &= f\left(\mathbf{y}_i^* \mid \mathbf{v}_i^*, \mathbf{v}_i\right) \cdot g\left(\mathbf{v}_i^*\right) g\left(\mathbf{v}_i\right) \end{aligned} \tag{10.93}$$

Therefore

$$f\left(\mathbf{y}_i^*\right) = \int_{\mathbf{v}_i} \left\{ \int_{\mathbf{v}_i^*} f\left(\mathbf{y}_i^* \mid \mathbf{v}_i^*, \mathbf{v}_i\right) \cdot g\left(\mathbf{v}_i^*\right) d\mathbf{v}_i^* \right\} g\left(\mathbf{v}_i\right) d\mathbf{v}_i$$

From (10.91), it follows that

$$f\left(\mathbf{y}_i^* \mid \mathbf{v}_i^*, \mathbf{v}_i\right) \cdot g\left(\mathbf{v}_i\right) = \prod_{j=1}^{n_i} f\left(\mathbf{y}_{ij} \mid \mathbf{v}_{ij}, \mathbf{v}_i\right) g\left(\mathbf{v}_{ij}\right).$$

Hence

$$
\begin{aligned}
f\left(\mathbf{y}_i^*\right) &= \int_{\mathbf{v}_i} \prod_{j=1}^{n_i} \left\{ \int_{\mathbf{v}_{ij}} f\left(\mathbf{y}_i \mid \mathbf{v}_i, \mathbf{v}_i\right) \cdot g\left(\mathbf{v}_{ij}\right) d\mathbf{v}_{ij} \right\} g\left(\mathbf{v}_i\right) d\mathbf{v}_i \\
&= \int_{\mathbf{v}_i} \prod_{j=1}^{n_i} \left\{ \int_{\mathbf{v}_{ij}} \left[\prod_{k=1}^{n_{ij}} f\left(y_{ijk} \mid \mathbf{v}_{ij}, \mathbf{v}_i\right) \right] \cdot g\left(\mathbf{v}_{ij}\right) d\mathbf{v}_{ij} \right\} g\left(\mathbf{v}_i\right) d\mathbf{v}_i
\end{aligned}
\tag{10.94}
$$

Using the Poisson distribution model as an example,

$$
f\left(y_{ijk} \mid \mathbf{v}_{ij,}, \mathbf{v}_i\right) = \frac{\exp\left(-\lambda_{ijk}\right) \lambda_{ijk}^{\ y_{ijk}}}{\lambda_{ijk}\,!},
$$

where

$$
\lambda_{ijk} = \exp\left\{ \mathbf{x}_{ijk}^{'} \boldsymbol{\beta} + \mathbf{z}_{ijk(2)}^{'} \mathbf{v}_{ij} + \mathbf{z}_{ijk(3)}^{'} \mathbf{v}_i \right\}.
$$

10.3.2 Log-likelihood function

Let

$$
\begin{aligned}
l_i &= \ln f\left(\mathbf{y}_i^*\right) \\
&= \ln \int_{\mathbf{v}_i} \prod_{j=1}^{n_i} \left\{ h\left(\mathbf{v}_{ij}\right) \right\} g\left(\mathbf{v}_i\right) d\mathbf{v}_i
\end{aligned}
$$

where

$$
h\left(\mathbf{v}_{ij}\right) = \int_{\mathbf{v}_{ij}} \exp \sum_{k=1}^{n_{ij}} \ln f\left(\mathbf{y}_{ijk} \mid \mathbf{v}_{ij}, \mathbf{v}_i\right) g\left(\mathbf{v}_{ij}\right) d\mathbf{v}_i
$$

Note that

$$\prod_{j=1}^{n_i}\left[h\left(\mathbf{v}_{ij}\right)\right]g\left(\mathbf{v}_i\right)=\prod_{j=1}^{n_i}\left\{K_1\int_{\mathbf{u}_{ij}}\exp\left[\sum_{k=1}^{n_{ij}}l_{ijk}-\frac{1}{2}\mathbf{v}_{ij}^{'}\boldsymbol{\Phi}^{-1}\mathbf{v}_{ij}\right]d\mathbf{v}_{ij}\right\}K_2\exp-\frac{1}{2}\mathbf{v}_i'\boldsymbol{\Psi}^{-1}\mathbf{v}_i$$

Let

$$t_{ij}=\sum_{k=1}^{n_{ij}}l_{ijk}-\frac{1}{2}\mathbf{v}_{ij}^{'}\boldsymbol{\Phi}^{-1}\mathbf{v}_{ij}+\ln K_u \tag{10.95}$$

$$l_{ijk}=\ln f\left(y_{ijk}\mid\mathbf{v}_{ij},\mathbf{v}_i\right) \tag{10.96}$$

and

$$q_i=\ln K_v-\frac{1}{2}\mathbf{v}_i'\boldsymbol{\Psi}^{-1}\mathbf{v}_i, \tag{10.97}$$

$$K_u=\left(2\pi\right)^{-r/2}\left|\boldsymbol{\Phi}\right|^{-1/2}$$

$$K_v=\left(2\pi\right)^{-m/2}\left|\boldsymbol{\Psi}\right|^{-1/2}.$$

From (10.95), (10.96), and (10.97) it follows that

$$\begin{aligned}f\left(\mathbf{y}_i^*\right)&=K_2\int_{\mathbf{v}_i}\left\{\prod_{j=1}^{n_i}\int_{\mathbf{v}_{ij}}\exp t_{ij}d\mathbf{v}_{ij}\right\}\exp q_i d\mathbf{v}_i\\&=K_2\int_{\mathbf{v}_i}\exp\left\{\sum_{j=1}^{n_i}\ln\int_{\mathbf{v}\mathbf{u}_{ij}}\exp t_{ij}d\mathbf{v}_{ij}\right\}\exp q_i d\mathbf{v}_i\\&=K_2\int_{\mathbf{v}_i}\exp\left(\sum_{j=1}^{n_i}\ln q_{ij}^*+q_i\right)d\mathbf{v}_i\end{aligned}$$

with $q_{ij}^*=\int_{\mathbf{v}_{ij}}\exp t_{ij}d\mathbf{v}_{ij}$.

Therefore

$$\ln l_i = \ln K_v + \ln f\left(\mathbf{y}_i^*\right) = \ln K_v + \ln \int_{v_i} \exp\left(q_{ij} + q_i\right) dv_i \qquad (10.98)$$

where $q_{ij} = \sum_{j=1}^{n_i} \ln q_{ij}^*$.

10.3.3 Empirical Bayes estimates

Estimates of the random effects are obtained as the conditional expectation of u_{ijk} given the observations $\mathbf{y}_i^*$. More specifically,

$$E\left(u_{ijk} \mid \mathbf{y}_i^*\right) = \left[K_v \int \exp\left(q_i + q_{ij}\right) p_{ijk} d\mathbf{v}_i\right] / f\left(\mathbf{y}_i^*\right) \qquad (10.99)$$

where

$$K_v = \left(2\pi\right)^{-\frac{m}{2}} \left|\boldsymbol{\Psi}\right|^{-\frac{1}{2}},$$

$$q_i = -\frac{1}{2} \mathbf{v}_i \boldsymbol{\Psi}^{-1} \mathbf{v}_i$$

and where

$$p_{ijk} = \int_{\mathbf{v}_{ij}} u_{ijk} f\left(\mathbf{y}_{ij} \mid \mathbf{v}_i, \mathbf{v}_{ij}\right) g\left(\mathbf{v}_{ij}\right) d\mathbf{v}_{ij}$$

Likewise

$$E\left(v_{ijk} \mathbf{g}_{ijl} \mid \mathbf{y}_i^*\right) = \left[K_v \int \exp\left(q_i + q_{ij}\right) c_{ijkl} d\mathbf{v}_i\right] / f\left(\mathbf{y}_i^*\right) \qquad (10.100)$$

where

$$c_{ijkl} = \int_{\mathbf{v}_{ij}} v_{ijk} v_{ijl} f\left(\mathbf{y}_{ij} \mid \mathbf{v}_i, \mathbf{v}_{ij}\right) g\left(\mathbf{v}_{ij}\right) d\mathbf{v}_{ij} .$$

10.3.4 Derivatives of the log-likelihood function

10.3.4.1 Fixed effects: β-derivatives

$$l_i = \ln f\left(\mathbf{y}_i^*\right) = \ln K_v \int \exp\left(q_i + q_{ij}\right) d\mathbf{v}_i$$

Therefore

$$l_i = \ln \int_{\mathbf{v}_i} \exp\left\{\sum_{j=1}^{n_i} \ln \int_{\mathbf{v}_{ij}} f\left(\mathbf{y}_{ij} \mid \mathbf{v}_i, \mathbf{v}_{ij}\right) g\left(\mathbf{v}_{ij}\right)\right\} g\left(\mathbf{v}_i\right) d\mathbf{v}_i$$

$$\begin{aligned} \frac{\partial \ln l_i}{\partial \beta_r} &= \frac{1}{f\left(\mathbf{y}_i^*\right)} K_v \int_{\mathbf{v}_i} \frac{\partial}{\partial \beta_r} \exp\left(q_{ij} + q_i\right) dv_i \\ &= \frac{1}{f\left(\mathbf{y}_i^*\right)} K_v \int_{\mathbf{v}_i} \left[\frac{\partial}{\partial \beta_r} q_{ij}\right] \exp\left(q_{ij} + q_i\right) dv_i \end{aligned} \qquad (10.101)$$

Since

$$q_{ij} = \sum_{j=1}^{n_i} \ln q_{ij}^* ,$$

it follows that

$$\frac{\partial q_{ij}}{\partial \beta_r} = \sum_{j=1}^{n_i} \frac{\frac{\partial}{\partial \beta_r} \ln q_{ij}^*}{q_{ij}^*}$$

and

$$\begin{aligned}\frac{\partial}{\partial \beta_r} q_{ij}^* &= \int_{\mathbf{v}_{ij}} \frac{\partial}{\partial \beta_r} \exp \sum_{k=1}^{n_{ij}} \ln f\left(\mathbf{y}_{ijk} \mid \mathbf{v}_i, \mathbf{v}_{ij}\right) g\left(\mathbf{v}_{ij}\right) d\mathbf{v}_{ij} \\ &= \int_{\mathbf{v}_{ij}} \sum_{k=1}^{n_{ij}} \frac{\partial}{\partial \beta_r} \ln f\left(\mathbf{y}_{ijk} \mid \mathbf{v}_i, \mathbf{v}_{ij}\right) f\left(\mathbf{y}_{ij} \mid \mathbf{v}_i, \mathbf{v}_{ij}\right) g\left(\mathbf{v}_{ij}\right) d\mathbf{v}_{ij}.\end{aligned} \tag{10.102}$$

10.3.4.2 Level-2 variance components: Φ - derivatives

$$\frac{\partial \ln l_i}{\partial \Phi_{rs}} = \frac{1}{f\left(\mathbf{y}_i^*\right)} K_v \int_{\mathbf{v}} \left[\frac{\partial}{\partial \phi_{rs}} q_{ij}\right] \exp\left(q_{ij} + q_i\right) d\mathbf{v}_i \tag{10.103}$$

$$q_{ij} = \sum_{j=1}^{n_i} \ln q_{ij}^*$$

$$\begin{aligned}\frac{\partial q_{ij}}{\partial \Phi_{rs}} &= \sum_{j=1}^{n_i} \frac{1}{q_{ij}^*} \int_{\mathbf{v}_{ij}} \left[\frac{\partial}{\partial \phi_{rs}} \ln g\left(\mathbf{v}_{ij}\right)\right] f\left(\mathbf{y}_{ij} \mid \mathbf{v}_i, \mathbf{v}_{ij}\right) g\left(\mathbf{v}_{ij}\right) d\mathbf{v}_{ij} \\ &= \sum_{j=1}^{n_i} \frac{\mathrm{E}_{\mathbf{v}_{ij}/\mathbf{y}_{ij}}\left[\frac{\partial}{\partial \phi_{rs}} \ln g\left(\mathbf{v}_{ij}\right)\right]}{q_{ij}^*} \\ &= \sum_{j=1}^{n_i} \frac{\mathrm{E}_{\mathbf{v}_{ij}/\mathbf{y}_i}\left[\frac{2-\delta rs}{2}\left(\mathbf{P}_{ij}\right)_{r,s}\right]}{q_{ij}^*}\end{aligned}$$

where $\mathbf{P}_{ij} = \mathbf{\Phi}^{-1}\left(\mathbf{v}_{ij}\mathbf{v}_{ij}' - \mathbf{\Phi}\right)\mathbf{\Phi}^{-1}$.

10.3.4.3 Level-3 variance components: Ψ - derivatives

$$\frac{\partial \ln l_i}{\partial \psi_{rs}} = \frac{1}{f\left(\mathbf{y}_i^*\right)} \int_{\mathbf{v}_i} \left[\frac{\partial}{\partial \psi_{rs}} \left(q_{ij} \right) \right] \exp\left(q_{ij} + q_i \right) dv_i \qquad (10.104)$$

$$\begin{aligned} q_i &= -\frac{1}{2} \mathbf{v}_i \Psi^{-1} \mathbf{v}_i + \ln K_v \\ &= \ln g\left(\mathbf{v}_i \right) \end{aligned}$$

$$\Rightarrow \frac{\partial q_i}{\partial \psi_{rs}} = \left[\frac{2 - \delta rs}{2} \left(\mathbf{P}_i \right)_{r,s} \right] \qquad (10.105)$$

where $\mathbf{P}_i = \Psi^{-1}\left(\mathbf{v}_i \mathbf{v}_i^{'} - \Psi \right) \Psi^{-1}$.

10.3.5 Second order derivatives

The method for obtaining second order partial derivatives is illustrated below for the terms $\frac{\partial^2 \ln l_i}{\partial \psi_{uv} \partial \psi_{rs}}$ and $\frac{\partial^2 \ln l_i}{\partial \phi_{uv} \partial \phi_{rs}}$. The derivatives for $\frac{\partial^2 \ln l_i}{\partial \beta_u \partial \psi_{rs}}$ etc. are obtained in a similar way.

$\underline{\mathrm{I}\left(\psi_{uv}, \psi_{rs} \right)}$.

From (10.104)

$$\begin{aligned} \frac{\partial \ln l_i}{\partial \psi_{rs}} &= \frac{\partial \ln f\left(\mathbf{y}_i^* \right)}{\partial \psi_{rs}} \\ &= \frac{1}{f\left(\mathbf{y}_i^* \right)} \int_{\mathbf{v}_i} \left[\frac{\partial}{\partial \psi_{rs}} \ln g\left(\mathbf{v}_i \right) \right] \exp\left(q_{ij} + q_i \right) d\mathbf{v}_i \end{aligned}$$

Hence

$$\frac{\partial^2 \ln l_i}{\partial \psi_{uv} \partial \psi_{rs}} = \frac{1}{f\left(\mathbf{y}_i^*\right)} \int_{\mathbf{v}_i} \left[\frac{\partial^2}{\partial \psi_{uv} \psi_{rs}} \ln g\left(\mathbf{v}_i\right) \right] \exp\left(q_i + q_{ij}\right) d\mathbf{v}_i$$

$$+ \frac{1}{f\left(\mathbf{y}_i^*\right)} \int_{\mathbf{v}_i} \left[\frac{\partial}{\partial \psi_{rs}} \ln g\left(\mathbf{v}_i\right) \frac{\partial}{\partial \psi_{uv}} \ln g\left(\mathbf{v}_i\right) \right] \exp\left(q_i + q_{ij}\right) d\mathbf{v}_i$$

$$- \frac{1}{f\left(\mathbf{y}_i^*\right)^2} \left\{ \int_{\mathbf{v}_i} \left[\frac{\partial}{\partial \psi_{rs}} \ln g\left(\mathbf{v}_i\right) \right] \exp\left(q_i + q_{ij}\right) d\mathbf{v}_i + \int_{\mathbf{v}_i} \left[\frac{\partial}{\partial \psi_{uv}} \ln g\left(\mathbf{v}_i\right) \right] \exp\left(q_i + q_{ij}\right) d\mathbf{v}_i \right\}$$

$$= \frac{1}{f\left(\mathbf{y}_i^*\right)} \int_{\mathbf{v}_i} \left\{ \frac{\partial^2}{\partial \psi_{uv} \psi_{rs}} \ln g\left(\mathbf{v}_i\right) + \frac{\partial}{\partial \psi_{rs}} \ln g\left(\mathbf{v}_i\right) \frac{\partial}{\partial \psi_{uv}} \ln g\left(\mathbf{v}_i\right) \right\} \times \exp\left(q_i + q_{ij}\right) d\mathbf{v}_i$$

$$- \frac{\partial \ln l_i}{\partial \psi_{uv}} \frac{\partial \ln l_i}{\partial \psi_{rs}}$$

$$= \frac{cons}{f\left(\mathbf{y}_i^*\right)} \int_{\mathbf{v}_i} \left[P_i\right]_{u,v} \left[P_i\right]_{r,s} \times \exp\left(q_i + q_{ij}\right) d\mathbf{v}_i - \frac{\partial \ln l_i}{\partial \psi_{uv}} \frac{\partial \ln l_i}{\partial \psi_{rs}} - cons \psi_{ur}^{-1} \psi_{vs}^{-1} + \psi_{us}^{-1} \psi_{vr}^{-1}$$

(10.106)

$$cons = \frac{\left(2 - \delta uv\right)\left(2 - \delta rs\right)}{2}$$

From (10.106) it follows that

$$\frac{\partial^2 \ln f_i}{\partial \psi_{uv} \psi_{rs}} = cons \left\{ E_{|y_i} \left[\mathbf{P}_i\right]_{uv} E_{|y_i} \left[\mathbf{P}_i\right]_{rs} - E_{|y_i} \left[\mathbf{P}_i\right]_{uv} \cdot \left[\mathbf{P}_i\right]_{rs} + \boldsymbol{\psi}_{ur}^{-1} \boldsymbol{\psi}_{vs}^{-1} + \boldsymbol{\psi}_{us}^{-1} \boldsymbol{\psi}_{vr}^{-1} \right\}.$$

(10.107)

$\mathbf{I}\left(\boldsymbol{\Phi}_{\text{uv}}, \boldsymbol{\Phi}_{\text{rs}}\right)$:

From (10.103) we have

$$\frac{\partial \ln l_i}{\partial \phi_{rs}} = \frac{1}{f\left(\mathbf{y}_i^*\right)} \int_{\mathbf{v}_i} \left[\frac{\partial}{\partial \phi_{rs}} \ln g\left(\mathbf{v}_i\right) \right] \exp\left(q_{ij} + q_i\right) d\mathbf{v}_i$$

and

$$q_{ij} = \sum_{k=1}^{n_{ij}} \ln \int_{\mathbf{v}_{ij}} f\left(\mathbf{y}_{ij} \mid \mathbf{v}_i, \mathbf{v}_{ij}\right) g\left(\mathbf{v}_{ij}\right) d\mathbf{v}_{ij} \ .$$

Thus

$$\frac{\partial}{\partial \phi_{rs}} q_{ij} = \sum_{j=1}^{n_i} \frac{2 - \delta rs}{2} E\_PIJ = EPIJ \ ,$$

where

$$E\_PIJ = \frac{1}{q_{ij}^*} \int \left(P_{ij}\right)_{r,s} \left[\frac{2 - \delta rs}{2} \right] f\left(\mathbf{y}_{ij} \mid \mathbf{v}_i, \mathbf{v}_{ij}\right) g\left(\mathbf{v}_{ij}\right) d\mathbf{v}_{ij} \ ,$$

and

$$\frac{\partial \ln l_i}{\partial \Phi_{rs}} = \frac{1}{f\left(\mathbf{y}_i^*\right)} \int_{\mathbf{v}_i} EPIJ \exp\left(q_{ij} + q_i\right) d\mathbf{v}_i \qquad (10.108)$$

$$EPIJ = \sum_{j=1}^{n_i} E\_PIJ \ ,$$

and therefore

$$\frac{\partial^2 \ln l_i}{\partial \phi_{uv} \partial \phi_{rs}} = \frac{1}{f\left(\mathbf{y}_i^*\right)} \int_{\mathbf{v}_i} \left[\frac{\partial^2}{\partial \phi_{uv} \phi_{rs}} q_{ij} + \left\{ \frac{\partial}{\partial \phi_{uv}} q_{ij} \times \frac{\partial}{\partial \phi_{rs}} q_{ij} \right\} \right] \exp\left(q_i + q_{ij}\right) d\mathbf{v}_i - \frac{\partial \ln l_i}{\partial \Phi_{uv}} \cdot \frac{\partial \ln l_i}{\partial \Phi_{rs}} \ .$$

Hence

$$\frac{\partial^2 \ln f_i}{\partial \Phi_{uv} \Phi_{rs}} = \frac{-cons}{f\left(\mathbf{y}_i^*\right)} \int_{\mathbf{v}_i} \left\{ \sum_{j=1}^{n_i} E_{|y_{ij}} \left[\mathbf{P}_{ij} \right]_{uv} \left[\mathbf{P}_{ij} \right]_{rs} \right\} \exp\left(q_i + q_{ij}\right) d\mathbf{v}_i$$

$$+ \frac{\partial \ln f_i}{\partial \Phi_{uv}} \cdot \frac{\partial \ln f_i}{\partial \Phi_{rs}} + cons \times \left(\sum_{j=1}^{N} n_i \right) \left[\boldsymbol{\Phi}_{ur}^{-1} \boldsymbol{\Phi}_{vs}^{-1} + \boldsymbol{\Phi}_{us}^{-1} \boldsymbol{\Phi}_{vr}^{-1} \right]$$

10.3.6 Evaluation of integrals

In the preceding sections expressions for the log-likelihood function and derivatives are given in terms of multiple integrals. In general, no closed form solution to these multiple integrals exists and therefore use is made of numerical integration to evaluate them.

Consider a general integral of the form

$$I = \int f\left(\mathbf{y}_i \mid \mathbf{v}_i\right) g\left(\mathbf{v}_i\right) d\mathbf{v}_i,$$

where it is assumed that $\mathbf{v}_i : N\left(\mathbf{0}, \boldsymbol{\Phi}\right)$. This integral can equivalently be written as follows:

$$I = \int \phi\left(\mathbf{v}_i \mid \mathbf{y}_i\right) \left\{ \frac{f\left(\mathbf{y}_i \mid \mathbf{v}_i\right) g\left(\mathbf{v}_i\right) d\mathbf{v}_i}{\phi\left(\mathbf{v}_i \mid \mathbf{y}_i\right)} \right\} d\mathbf{v}_i,$$

where

$$\phi\left(\mathbf{v}_i \mid \mathbf{y}_i\right) = k \exp\left(-\frac{1}{2} \left(\mathbf{v}_i - \hat{\mathbf{v}}_i \right) \boldsymbol{\Sigma}_i^{-1} \left(\mathbf{v}_i - \hat{\mathbf{v}}_i \right) \right), \qquad (10.109)$$

$$\hat{\mathbf{v}}_i = E\left(\mathbf{v}_i \mid \mathbf{y}_i\right),$$

$$\boldsymbol{\Sigma}_i = Cov\left(\mathbf{v}_i \mid \mathbf{y}_i\right),$$

and

$$k = \left(2\pi\right)^{-r/2} \mid \hat{\boldsymbol{\Sigma}}_i \mid^{-1/2} . \tag{10.110}$$

Consider the transformation of variables

$$\mathbf{z}_i = \frac{1}{\sqrt{2}} \mathbf{T}_i^{-1} \left(\mathbf{v}_i - \hat{\mathbf{v}}_i \right) \tag{10.111}$$

where

$$\mathbf{T}_i \mathbf{T}_i^{'} = \hat{\boldsymbol{\Sigma}}_i$$

and hence

$$\hat{\boldsymbol{\Sigma}}_i^{\ -1} = \left(\mathbf{T}_i^{'}\right)^{-1} \left(\mathbf{T}_i\right)^{-1} .$$

From (10.111) it follows that

$$\mathbf{v}_i = \mathbf{v}_i \left(\mathbf{z}_i\right) = \sqrt{2} \mathbf{T}_i \mathbf{z}_i + \hat{\mathbf{v}}_i ,$$

The Jacobian of the transformation is given by

$$d\mathbf{v}_i = \mid \mathbf{T}_i^{*} \mid d\mathbf{z}_i ,$$

$$\mathbf{T}_i^{*} = \sqrt{2} \mathbf{T}_i .$$

Using the change in variables, it follows that

$$I = \int k \exp - \mathbf{z}_i \mathbf{z}_i^{'} \left(\frac{f\left(\mathbf{y}_i \mid \mathbf{v}_i\right) g\left(\mathbf{v}_i\right)}{k \exp - \mathbf{z}_i \mathbf{z}_i^{'}} \right) \mid \mathbf{T}_i^{*} \mid d\mathbf{z}_i . \tag{10.112}$$

10.3.7 Adaptive quadrature

To evaluate (10.112), use is made of a direct implementation of Gauss-Hermite quadrature. With this rule

$$\int_{-\infty}^{\infty} \exp\{-z^2\} f(z) dz$$

can be approximated by

$$\sum_{\alpha=1}^{G} w_\alpha f(z_\alpha),$$

where the w_α and z_α denote weights and nodes of the Hermite polynomial of degree G.

Applying this to the multiple integral defined by (10.112), it follows that

$$\begin{aligned} I \; ; \; & C \sum_{g_1=1}^{G} \ldots \sum_{g_r=1}^{G} w_{g_1} \ldots w_{g_r} \exp \mathbf{z}_g \mathbf{z}_g' \, | \, \mathbf{T}_i^* \, | \, f\left(\mathbf{y}_i \, | \, \mathbf{v}_i(\mathbf{z}_g)\right) g\left(\mathbf{v}_i(\mathbf{z}_g)\right) \\ = & C \sum_{g_1=1}^{G} \ldots \sum_{g_r=1}^{G} m_{g_1} \ldots m_{g_r} \, | \, \mathbf{T}_i^* \, | \, f\left(\mathbf{y}_i \, | \, \mathbf{v}_i(\mathbf{z}_g)\right) g\left(\mathbf{v}_i(\mathbf{z}_g)\right) \end{aligned}$$

$$\begin{aligned} I \; ; \; & C \sum_{g_1=1}^{G} \mathsf{L} \sum_{g_r=1}^{G} w_{g_1} \mathsf{L} \, w_{g_r} \exp \mathbf{z}_g \mathbf{z}_g' \left| \mathbf{T}_i^* \right| f\left(\mathbf{y}_i \mid \mathbf{v}_i(\mathbf{z}_g)\right) g\left(\mathbf{v}_i(\mathbf{z}_g)\right) \\ = & C \sum_{g_1=1}^{G} \mathsf{L} \sum_{g_r=1}^{G} m_{g_1} \mathsf{L} \, m_{g_r} \left| \mathbf{T}_i^* \right| f\left(\mathbf{y}_i \mid \mathbf{v}_i(\mathbf{z}_g)\right) g\left(\mathbf{v}_i(\mathbf{z}_g)\right) \end{aligned}$$

where

$$C = (2\pi)^{-r/2} \, | \, \mathbf{\Phi} \, |^{-1/2},$$

$$m_{g_\alpha} = \exp\left(z_{g_\alpha} \cdot \ln w_{g_\alpha}\right),$$

and

$$z_g = \left(z_{g1}, z_{g2}, ..., z_{gr}\right).$$

Values of $\hat{\mathbf{v}}_i$ and $\hat{\mathbf{\Sigma}}_i$ (cf. Section 10.3.3) are iteratively updated. This implies that the location and scale of the area under the integral changes over iterations and depends on the observed values for a particular level-3 or level-2 unit.

10.4 Starting values for generalized linear models

10.4.1 Introduction

SuperMix uses an algorithm based on the maximization of the posterior distribution (MAP) with respect to the random effects.

In the sections to follow, we assume a level-2 model with a count outcome variable. It is also assumed that the Poisson model is appropriate for level-2 data with a subset of the regression coefficients assumed to be random.

10.4.2 Illustration of the procedure for a count outcome variable

Let y_{ij} be a count outcome variable where i denotes level-2 units, $i = 1, 2, \mathsf{K}$, N and j level-1 units nested within the level-2 units $j = 1, 2, \mathsf{K}$, n_i.

Under the assumption of conditional independence

$$f\left(\mathbf{y}_i \mid \mathbf{v}_i\right) = \prod_{j=2}^{n_i} exp\left(-\mu_{ij}\right) \mu_{ij}^{\ y_{ij}} \left(y_{ij}\,!\right)^{-1} \qquad (10.113)$$

Suppose that the following exponential model is imposed on the means of the elements y_{ij} of the $n_i \times 1$ vector $\mathbf{y}_i$

$$\mu_{ij} = exp\left(\eta_{ij}\right) = exp\left(\mathbf{x}_{ij}^{'}\boldsymbol{\beta} + \mathbf{z}_{ij}^{'}\mathbf{v}_i\right), \tag{10.114}$$

where $\mathbf{x}_{ij}$ is a $p \times 1$ vector of covariates and the elements of $\boldsymbol{\beta} = \left[\beta_1, \beta_2, \mathsf{K}\ , \beta_p\right]^{'}$ denote unknown, but fixed, parameters. Generally, the $m \times 1$ vector $\boldsymbol{z}_{ij}$ is a subset of the columns of $\mathbf{x}_{ij}$. Additionally, it is assumed that $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N$ are i.i.d. $N\left(\mathbf{0}, \boldsymbol{\Phi}\right)$.

The model (10.114) is transformed to a linear model by using the log link function. In other words,

$$\eta_{ij} = ln\left(\mu_{ij}\right). \tag{10.115}$$

Using standard results for conditional distributions, it follows that

$$\begin{aligned} f\left(\mathbf{v}_i \mid \mathbf{y}_i\right) &= f\left(\mathbf{v}_i, \mathbf{y}_i\right) / f\left(\mathbf{y}_i\right) \\ &= f\left(\mathbf{y}_i \mid \mathbf{v}_i\right) \cdot g\left(\mathbf{v}_i\right) / f\left(\mathbf{y}_i\right). \end{aligned}$$

Hence

$$\begin{aligned} ln\, f\left(\mathbf{v}_i \mid \mathbf{y}_i\right) &= ln\, f\left(\mathbf{y}_i \mid \mathbf{v}_i\right) + ln\, g\left(\mathbf{v}_i\right) - K \\ &= \sum_{j=1}^{n_i}\left\{-\mu_{ij} + y_{ij}\, ln\, \mu_{ij} - ln\left(y_{ij}\,!\right)\right\} + g\left(\mathbf{v}_i\right) + K, \end{aligned} \tag{10.116}$$

10.4.3 Gradient vector and Hessian matrix

Given $\boldsymbol{\beta}$ and $\boldsymbol{\Phi}$, it follows that

$$\frac{\partial ln\, f\left(\mathbf{v}_i \mid \mathbf{y}_i\right)}{\partial\left[\mathbf{v}_i\right]_r} = \sum_{j=1}^{n_i}\left\{ y_{ij}\frac{\partial}{\partial v_{ir}} ln\,\mu_{ij} - \frac{\partial}{\partial v_{ir}}\mu_{ij}\right\} + \frac{\partial}{\partial v_{ir}} ln\, g\left(\mathbf{v}_i\right), \qquad (10.117)$$

$r = 1, 2, \mathsf{K}\ , m$

Since

$$\frac{\partial \mu_{ij}}{\partial v_{ir}} = \mu_{ij} z_{ijr}, \qquad (10.118)$$

$$g\left(\mathbf{v}_i\right) = \left(2\pi\right)^{-\frac{r}{2}} \left|\boldsymbol{\Phi}\right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\mathbf{v}_i^{'}\boldsymbol{\Phi}^{-1}\mathbf{v}_i\right\}$$

and hence

$$\frac{\partial \ln g\left(\mathbf{v}_i\right)}{\partial\left[\mathbf{v}_i\right]_r} = -\left[\boldsymbol{\Phi}^{-1}\mathbf{v}_i\right]_r, \qquad (10.119)$$

it follows that

$$\frac{\partial \ln f\left(\mathbf{v}_i \mid \mathbf{y}_i\right)}{\partial v_{ir}} = \sum_{j=1}^{n_i} z_{ijr}\left\{y_{ij} - \mu_{ij}\right\} - \left[\boldsymbol{\Phi}^{-1}\mathbf{v}_i\right]_r, \qquad r = 1, 2, \ldots, \ldots, m.$$

Maximization of $\ln f\left(\mathbf{v}_i \mid \mathbf{y}_i\right)$ is equivalent to the minimization of

$$F = \underbrace{-ln\ f\left(\mathbf{v}_i \mid \mathbf{y}_i\right)}_{F2} - \underbrace{-ln\ g\left(\mathbf{v}_i\right)}_{F1}. \qquad (10.120)$$

Hence the gradient vector is defined by

$$\frac{\partial F_i}{\partial v_{ir}} = -\frac{\partial \ln f\left(\mathbf{v}_i \mid \mathbf{y}_i\right)}{\partial v_{ir}}. \qquad (10.121)$$

Furthermore,

$$\frac{\partial^2 F_i}{\partial v_{ir}\partial v_{is}} = \left[\boldsymbol{\Phi}^{-1}\right]_{r,s} + \sum_{j=1}^{n_i} \mu_{ij} z_{ir} z_{ij} s, \qquad r, s = 1, 2, \ldots, \ldots, m. \tag{10.122}$$

Let $\mathbf{H}$ denote the Hessian matrix, where

$$\left[\mathbf{H}_i\right]_{r,s} = \frac{\partial^2 F}{\partial v_{ir}\partial v_{is}}, \tag{10.123}$$

then

$$E\left[\mathbf{H}_i\right]_{r,s} = \boldsymbol{\Phi}^{-1}_{r,s} + \sum_{j=1}^{n_i} z_{ijr} z_{ijs} E\left(\mu_{ij}\right), \tag{10.124}$$

where

$$E\left(\mu_{ij}\right) = E\left(exp\left\{\mathbf{x}_{ij}^{'}\boldsymbol{\beta} + \mathbf{z}_{ij}^{'}\mathbf{v}_{ij}\right\}\right). \tag{10.125}$$

Therefore

$$E\left[\mathbf{H}_i\right]_{r,s} = \boldsymbol{\Phi}^{-1}_{r,s} + \sum_{j=1}^{n_i}\left[exp\left\{\mathbf{x}_{ij}^{'}\boldsymbol{\beta} + \frac{1}{2}\mathbf{z}_{ij}^{'}\boldsymbol{\Phi}\mathbf{z}_{ij}\right\}\right] z_{ijr} z_{ijs}. \tag{10.126}$$

10.4.4 The MAP algorithm

1. Set $\boldsymbol{\Phi} = 0.1 * \mathbf{I}$, $\hat{\mathbf{v}}_1, \hat{\mathbf{v}}_2, \ldots, \hat{\mathbf{v}}_N = \mathbf{0}$
2. Calculate $\hat{\boldsymbol{\beta}}$ given $\hat{\boldsymbol{\Phi}}$ and $\hat{v}_i$
3. Given the current estimates $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ and $\hat{\boldsymbol{\Phi}}$ of $\boldsymbol{\Phi}$, calculate $\hat{v}_i$, $i = 1, 2, \ldots, N$ using the Newton-Raphson method:

$$\hat{v}_i^{k} = \hat{v}_i^{(k-1)} + \mathbf{H}_i^{-1(k)}\mathbf{g}_i^{(k)}, \qquad k = 1, 2, \ldots \tag{10.127}$$

where

$$\left[\mathbf{g}_i\right]_r = \frac{\partial F_i}{\partial v_{ir}}, \qquad r = 1, 2, \ldots, m. \tag{10.128}$$

4. Obtain (see, *e.g.*, du Toit, 1993) a revised estimate $\hat{\boldsymbol{\Phi}}$ of $\boldsymbol{\Phi}$ from

$$\hat{\boldsymbol{\Phi}} = \frac{1}{N}\sum\left\{Cov\left(\hat{\mathbf{v}}_i \mid \mathbf{y}_i\right) + \left(\hat{\mathbf{v}}_i - \hat{\bar{\mathbf{v}}}\right)\left(\hat{\mathbf{v}}_i - \hat{\bar{\mathbf{v}}}\right)'\right\} \tag{10.129}$$

where

$$Cov\left(\hat{\mathbf{v}}_i \mid \mathbf{y}_i\right) = \left[E\left(\mathbf{H}_i\right)\right]^{-1} \tag{10.130}$$

and

$$\hat{\mathbf{v}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} v_{ij}\,. \tag{10.131}$$

5. Repeat steps (2) to (4) until convergence is attained.

10.4.5 Starting values for adaptive quadrature

As initial estimates of the fixed and random parameters, we use the values of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\Phi}}$ at convergence. The $\hat{\mathbf{v}}_i$ and $Cov(\hat{\mathbf{v}}_i \mid \mathbf{y}_i)$ (see (10.130) and (10.131)) are used as initial estimates of the empirical Bayes means and the covariances in the adaptive quadrature procedure described in Section 10.3.7.

10.5 Survival analysis and ordinal models

10.5.1 Introduction

Several authors have noted the connection between survival analysis models and binary and ordinal regression models for survival data that are discrete or grouped within time intervals, see for example Hedeker (2008).

An example that illustrates the binary approach is given in Chapter 8. In Sections 10.5.2 and 10.5.3 we assume that time of assessment can take on only discrete positive values $t = 1, 2, \ldots, T$. To make the connection to ordinal models more direct, in the next sections time will be denoted by c, where $c = 1, 2, \ldots, C$ and where C equals the number of categories of the outcome variable y.

10.5.2 Proportional hazards model

Let y_{ijk} denote the outcome c, where $c = 1, 2, \ldots, C$ for individual k, $k = 1, \ldots, n_{ij}$ nested within level-2 unit j, $j = 1, 2 \ldots, n_i$, which in turn is nested within level-3 unit i, $i = 1, 2, \ldots, N$.

For each level-1 unit, observation continues until time y_{ijk}, at which point either an event occurs, indicated by $d_{ijk} = 1$ or the observation is censored, indicated by $d_{ijk} = 0$. Censoring indicates being observed at c but not at $c + 1$.

Let $P_{ijk,c}$ denote the probability of failure up to and including time interval c, that is

$$P_{ijk,c} = P\left(y_{ijk} \leq c\right) \tag{10.132}$$

From (10.131) it follows that the probability of survival beyond time interval c is $1 - P_{ijk,c}$.

Because $1-P_{ijk,c}$ represents the survival function, Hedeker (2008), based on McCallagh (1980) proposed the following proportional hazards model

$$\begin{aligned}\eta_{ijk,c} &= \log\left[-\log\left(1-P_{ijk,c}\right)\right] \\ &= \gamma_c + \mathbf{x}'_{ijk}\beta + \mathbf{z}'_{(2)ijk}\mathbf{v}_{ij} + \mathbf{z}'_{(3)ijk}\mathbf{v}_i\end{aligned} \tag{10.133}$$

where $\mathbf{x}_{ijk}$, $\mathbf{z}'_{(2)ijk}$ and $\mathbf{z}'_{(3)ijk}$ are design vectors for the fixed, random at level-2 and random at level-3 effects. The threshold terms γ_c represent the logarithm of the integrated baseline hazard (*i.e.* when $\boldsymbol{\beta}=\mathbf{0}$, $\mathbf{v}_{ij}=\mathbf{0}$, and $\mathbf{v}_i=\mathbf{0}$.)

A positive coefficient for a predictor reflects increasing hazard with greater values of the predictor. In the ordinal treatment, survival time is represented by y_{ijk}, which is designated as being censored or not.

10.5.3 Estimation

The probability of a response in category *c*, conditional on the random effects is equal to

$$P\left(y_{ijk}=c \mid \mathbf{v}_i, \mathbf{v}_{ij}\right) = P_{ijk,c} - P_{ijk,c-1} \tag{10.134}$$

where

$$P_{ijk,c} = 1-\exp\left[-\exp\left(\eta_{ijk,c}\right)\right]$$

The likelihood is given by

$$l\left(y_{ijk} \mid \mathbf{v}_i, \mathbf{v}_{ij}\right) = \prod_{j=1}^{n_i}\prod_{k=1}^{n_{ij}}\left[\left(P_{ijk,c}-P_{ijk,c-1}\right)^{d_{ijk}} \times \left(1-P_{ijk,c}\right)^{1-d_{ijk}}\right]^{y_{ijk,c}} \tag{10.135}$$

Consequently, the log-likelihood function is

$$\ln\left(y_{ijk} \mid \mathbf{v}_i, \mathbf{v}_{ij}\right) = \sum_{j=1}^{n_i}\sum_{k=1}^{n_{ij}} y_{ijk,c}\left\{d_{ijk}\ln\left(P_{ijk,c} - P_{ijk,c-1}\right) + \left(1-d_{ijk}\right)\ln\left(1-P_{ijk,c}\right)\right\} \qquad (10.136)$$

The marginal maximum likelihood $\ln\left(y_{ijk}\right)$ and derivatives are obtained using numerical quadrature as described in Section 10.4.

10.6 Level-2 continuous outcome models with autocorrelated level-1 errors

10.6.1 Introduction

It is usually assumed that the errors in linear random coefficient models are conditionally independent (conditional on the random effects). When fitting models to longitudinal data, it is often more realistic to assume that the model errors are autocorrelated over time.

In subsequent sections we describe several models that allow for subject heterogeneity via the level-2 random effects and autocorrelation via time-series structures imposed on the level-1 residuals.

Let

$$y_{ij} = \mathbf{x}_{ij}^{'}\boldsymbol{\beta} + \mathbf{z}_{ij}^{'}\mathbf{v}_i + e_{ij}, \qquad (10.137)$$

where $i = 1, 2, \ldots, N$ and $j = 1, 2, \ldots, n_i$. Level-2 units are denoted by the subscript i and level-1 units by the subscript j.

The set of n_i regression equations (10.137) can be written in matrix notation as

$$\mathbf{y}_i = \mathbf{X}_i^{'}\boldsymbol{\beta} + \mathbf{Z}_i^{'}\mathbf{v}_i + \mathbf{e}_i, \qquad (10.138)$$

where $\mathbf{X}_i$ and $\mathbf{Z}_i$ are design matrices for the fixed and random effects respectively.

It is further assumed that $\mathbf{v}_i : N\left(\mathbf{0}, \mathbf{\Phi}_{(2)}\right)$ and that $\mathbf{e}_i : N\left(\mathbf{0}, \sigma^2 \mathbf{\Omega}_i\right)$ where $\mathbf{v}_i$ and $\mathbf{e}_i$ are uncorrelated. For uncorrelated homogeneous level-1 residuals, $\mathbf{\Omega}_i = \mathbf{I} : (n_i \times 1)$.

From the distributional assumptions it follows that

$$\begin{bmatrix} \mathbf{y}_i \\ \mathbf{v}_i \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbf{X}_i \boldsymbol{\beta} \\ 0 \end{bmatrix}, \begin{bmatrix} \mathbf{Z}_i \mathbf{\Sigma}_v \mathbf{Z}_i' + \sigma^2 \mathbf{\Omega}_i & \mathbf{Z}_i \mathbf{\Sigma}_v \\ \mathbf{\Sigma}_v \mathbf{Z}' & \mathbf{\Sigma}_v \end{bmatrix} \right). \tag{10.139}$$

Also, the mean of the posterior distribution of $\mathbf{v}_i$, given $\mathbf{y}_i$, yields the empirical Bayes (EB) estimator of the random effects,

$$\hat{\mathbf{v}}_i = \left[\mathbf{Z}_i' \left(\sigma^2 \mathbf{\Omega}_i \right)^{-1} \mathbf{Z}_i + \mathbf{\Sigma}_v^{-1} \right]^{-1} \mathbf{Z}_i' \left(\sigma^2 \mathbf{\Omega}_i \right)^{-1} \left(\mathbf{y} - \mathbf{X}_i \boldsymbol{\beta} \right). \tag{10.140}$$

Similarly, the corresponding posterior covariance matrix is given by

$$\mathbf{\Sigma}_{v|y_i} = \left[\mathbf{Z}_i' \left(\sigma^2 \mathbf{\Omega}_i \right)^{-1} \mathbf{Z}_i + \mathbf{\Sigma}_v^{-1} \right]^{-1} \tag{10.141}$$

Further details regarding estimation of multilevel linear models are provided in Section 10.1.

10.6.2 AR(1) errors

The first-order autoregressive process (AR1) for the error e at time point j is given as

$$e_j = \rho e_{j-1} + \xi_j \tag{10.142}$$

where the disturbances ξ_j are assumed to be distributed $N\left(0, \sigma^2\right)$ and ρ is the autocorrelation coefficient that reflects the degree to which the errors are autocorrelated. It is assumed that $|\rho| < 1$ (*i.e.*, that ρ is a correlation parameter).

Under the AR(1) relationship, the variance of the errors at a particular time point is equal to

$$\begin{aligned} V\left(e_j\right) &= V\left(\rho e_{j-1} + \xi_j\right) \\ &= \rho^2 V\left(e_{j-1}\right) + \sigma^2. \end{aligned} \tag{10.143}$$

An assumption that is often made is that of stationarity, which means that the variance of the errors is assumed to be constant across time and that the correlations are the same within a time-lag. With stationarity,

$$Cov\left(e_j, e_{j-s}\right) = \frac{\rho^s \sigma^2}{1-\rho^2}. \tag{10.144}$$

Taken together, this leads to a variance-covariance matrix of the errors

$$\sigma^2 \mathbf{\Omega} = \frac{\sigma^2}{\left(1-\rho^2\right)} \begin{bmatrix} 1 & \rho & \rho^2 & \mathsf{L} & \rho^{n-1} \\ \rho & 1 & \rho & \mathsf{L} & \rho^{n-2} \\ \rho^2 & \rho & 1 & \mathsf{L} & \rho^{n-3} \\ \cdot & \cdot & \cdot & \mathsf{L} & \cdot \\ \cdot & \cdot & \cdot & \mathsf{L} & \cdot \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \mathsf{L} & 1 \end{bmatrix}$$

10.6.3 MA(1) errors

Another common form for autocorrelated errors is the first order moving average process, MA(1), which is given as

$$e_j = \xi_j - \theta \xi_{j-1} \tag{10.145}$$

with disturbances ξ_j assumed to be $N\left(0, \sigma^2\right)$, and θ the autocorrelation coefficient for the moving-average process. Here, the errors at a particular time point equal the

disturbances at that time point plus a correlated part of the disturbances at the previous time point. For the stationary MA(1) process

$$\sigma^2\boldsymbol{\Omega} = \sigma^2\begin{bmatrix} 1+\theta^2 & -\theta & 0 & \mathsf{L} & 0 \\ -\theta & 1+\theta^2 & -\theta & \mathsf{L} & 0 \\ 0 & -\theta & 1+\theta^2 & \mathsf{L} & 0 \\ \cdot & \cdot & \cdot & \mathsf{L} & \cdot \\ 0 & 0 & 0 & \mathsf{L} & 1+\theta^2 \end{bmatrix},$$

that is, a symmetric matrix with $\left(1+\theta^2\right)\sigma^2$ on the main diagonal, $-\theta\sigma^2$ on the first off-diagonal, and 0 everywhere else. This form posits that only the lag-1 errors are correlated. This implies that the errors at a given time point are only correlated with those one time point apart.

While the MA(1) form is generally unreasonable for the variance-covariance matrix of **e** in fixed effects linear models, it might well be reasonable for the variance-covariance matrix of **e** in linear random effect models, which is conditional on both covariates **X** and the random effects $\mathbf{v}_i$.

10.6.4 ARMA(1,1) errors

A more general form for the autocorrelated errors is the first-order mixed autoregressive-moving average process which depends on both the AR parameter ρ and the MA parameter θ, and is given as

$$e_k = \rho e_{k-1} + \varepsilon_k - \theta\varepsilon_{k-1} \qquad (10.146)$$

with all terms as before. The error variance-covariance matrix is now of the form:

$$\sigma^2\boldsymbol{\Omega} = \frac{\sigma^2}{\left(1-\rho^2\right)}\begin{bmatrix} \gamma_0 & \gamma_1 & \rho\gamma_1 & \cdots & \rho^{n-2}\gamma_1 \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \rho^{n-3}\gamma_1 \\ \rho\gamma_1 & \gamma_1 & \gamma_0 & \cdots & \rho^{n-4}\gamma_1 \\ \rho^2\gamma_1 & \rho\gamma_1 & \gamma_1 & \cdots & \rho^{n-5}\gamma_1 \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \rho^{n-2}\gamma_1 & \rho^{n-3}\gamma_1 & \rho^{n-4}\gamma_1 & \cdots & \gamma_0 \end{bmatrix}$$

where $\gamma_0 = 1+\theta^2-2\rho\theta$ and $\gamma_1 = \left(1-\rho\theta\right)\left(\rho-\theta\right)$.

10.6.5 Toeplitz errors

One can assume that each lag (or each off-diagonal in the error variance-covariance matrix) has its own distinct autocorrelation parameter. The error variance-covariance matrix is then of the form

$$\sigma^2\boldsymbol{\Omega} = \sigma^2\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{n-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{n-3} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \cdots & 1 \end{bmatrix}.$$

The matrix $\boldsymbol{\Omega}$ is a symmetric general Toeplitz matrix with $n-1$ unique autocorrelation parameters. It is typical to assume that some of the higher order lags have zero autocorrelation in models with level-2 random effects, and so one can define the s -order symmetric Toeplitz matrix to allow only the first s autocorrelations to be non-zero, with the others equal to zero. For instance, a random-intercepts model can only include at most $n-2$ Toeplitz autocorrelations, since this model is equivalent to a fixed effects model with full Toeplitz structure.

10.6.6 Non-stationary AR(1) errors

All of the above autocorrelated error forms assumed stationarity, and so the error (co)variances are equal within a time-lag in all of these forms. In some cases, it can be advantageous to relax this assumption. For an AR(1) process (see (10.143)) the variance of the errors at a particular time point is given by

$$V\left(e_j\right) = \rho^2 V\left(e_{j-1}\right) + \sigma^2.$$

Instead of assuming stationarity, assume that the errors have zero variance at time 0 (*i.e.*, one time point before the start of the process) namely $V\left(e_0\right) = 0$. Then one gets the following for the error variance at the first four time points:

$$\begin{aligned}
V\left(e_1\right) &= \sigma^2 \\
V\left(e_2\right) &= \left(1+\rho^2\right)\sigma^2 \\
V\left(e_3\right) &= \left(1+\rho^2+\rho^4\right)\sigma^2 \\
V\left(e_4\right) &= \left(1+\rho^2+\rho^4+\rho^6\right)\sigma^2.
\end{aligned}$$

The error covariance matrix is of the form

$$\sigma^2\mathbf{\Omega} = \sigma^2 \begin{bmatrix}
1 & \rho & \rho^2 & \mathsf{L} & \rho^{n-1} \\
\rho & \left(1+\rho^2\right) & \rho\left(1+\rho^2\right) & \mathsf{L} & \rho^{n-2}\left(1+\rho^2\right) \\
\rho^2 & \rho\left(1+\rho^2\right) & \left(1+\rho^2+\rho^4\right) & \mathsf{L} & \rho^{n-3}\left(1+\rho^2+\rho^4\right) \\
\cdot & \cdot & \cdot & \mathsf{L} & \cdot \\
\rho^{n-1} & \rho^{n-2}\left(1+\rho^2\right) & \rho^{n-3}\left(1+\rho^2+\rho^4\right) & \mathsf{L} & 1+\sum_{j=1}^{n-1}\rho^{2j}
\end{bmatrix}$$

which depends only on the non-stationary AR(1) parameter ρ and the error variance σ^2.

11 References

Achenbach, T.M. (1991). *Manual for the Child Behavior Checklist / 4-18 and 1991 Profile*. Burlington, VT: University of Vermont.

Allison, P. D., 1982, Discrete-time methods for the analysis of event histories, editor Leinhardt, S., *Sociological Methodology 1982*, 61-98. San Francisco: Jossey-Bass.

Agresti, A. (1990). *Categorical Data Analysis*. New York: Wiley.

Agresti, A. (2002). *Categorical Data Analysis*, 2nd Edition. New York: Wiley.

Anderson, T.W. (2003). *An introduction to multivariate statistical analysis*. Third edition, New York: Wiley & Sons.

Armstrong, B. G., & Sloan, M. (1989). Ordinal models for epidemiologic data. *Am J Epidemiol*, **129**,191–204.

Barcikowski, R.S. (1981). Statistical power with group mean as the unit of analysis. *Journal of Educational Statistics*, **6**(3), 267-285.

Box, G.E.P., & Jenkins, F.M. (1976)

Time Series Analysis: Forecasting and Control. Revised. California: Holden-Day.

Browne, M.W. & du Toit, S.H.C. (1992). Automated fitting of Nonstandard Models. *Multivariate Behavioral Research*, **27**(**2**), 269-300.

Cupples, L. A., D'Agostino, R. B., Anderson, K. & Kannel, W. B. (1985). Comparison of baseline and repeated measure covariate techniques in the Framingham Heart Study, *Statistics in Medicine*, **7**, 205-218.

Du Toit, M. (1995). *The analysis of hierarchical and unbalanced complex survey data using multilevel models*. PhD Dissertation, University of Pretoria, Pretoria, South Africa.

Du Toit, S.H.C. (1993). *Analysis of multilevel models. Part 1: Theoretical aspects*. Pretoria, South Africa: Human Sciences Research Council.

Flay B., Brannon B., Johnson C. *et. al.* (1988). The Television School and Family Smoking Prevention and Cessation Project: I. Theoretical Basis and Program Development. *Preventive Medicine*, **17**: 585-607.

Hedeker, D. (2008). Multilevel Models for Ordinal and Nominal Variables. in J. de Leeuw & E. Meijer (Eds.), *Handbook of multilevel analysis* (pp. 237-274). New York: Springer.

Hedeker, D. & Gibbons, R.D. (2006). *Longitudinal Data Analysis*. New York: Wiley.

Hedeker, D. & Gibbons, R.D. (1994). A random-effects ordinal regression model for multilevel analysis, *Biometrics*, **50**, 933-944.

Hedeker D., Mermelstein R.J. (1998) A multilevel thresholds of change model for analysis of stages of change data. *Multivariate Behavioral Research*, **33**: 427-55.

Hougaard, P. (1995). *Lifetime Data Analysis*. Netherlands: Springer.

Hedeker, D., Siddiqui, O., & Hu, F.B. (2000). Random-effects regression analysis of correlated grouped-time survival data, *Statistical Methods in Medical Research*, **9**, 161-179.

Jennrich, R., & Sampson, P.F. (1968). Application of stepwise regression to nonlinear estimation. *Technometrics*, **10**, 63-72.

Khatri, C.G. (1966) A note on a MANOVA model applied to problems in growth curves. *Ann. Inst. Statist. Math.*, **18**, 75-86.

Kreft, I. & De Leeuw, J. (1988). *Introducing multilevel modeling*. Sage.

Krommer, A.R. & Ueberhuber, C.W. (1994). *Numerical Integration on advanced computer systems.* New York: Springer-Verlag.

Laara, E. & Matthews, J.N.S. (1985). On the equivalence of two models for ordinal data. *Biometrika*, **72**, 206-207.

Little, R.C., Milliken, G.A., Stroup, W.W. & Wolfinger R.D. (1996). *SAS System for Mixed Models*. Cary: SAS Institute, Inc.

Longford, N.T. (1993). *Random Coefficient Models*, New York: Oxford University Press.

Lorr, M. & Klett, C.J. (1966). *Inpatient multidimensional psychiatric scale: manual*. Palo Alto, CA: Consulting Psychologists Press.

McCullagh, P. & Nelder, J. A., (1989). *Generalized linear models* (2nd edition). New York: Chapman & Hall.

McCallagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society*. Series **B**, **42**:109-142, 1980. (with discussion)

McKnight, B. & Van Den Eeden, S.K. (1993). A conditional analysis for two-treatment multiple period crossover designs with binomial or Poisson outcomes and subjects who drop out, *Statistics in Medicine*, **12**, 825-834.

National Epidemiologic Survey on Alcohol and Related Conditions: NIAAA 2001 - 2002 NESARC.

Prentice R. & Gloeckler L. (1978). Regression analysis of grouped survival data with application to breast cancer data. *Biometrics*, **34**, 57-67.

Pickes A. & Crouchley, R. (1995). A comparison of frailty models for multivariate survival data. *Stat. Med.* **14**(**13**):1447-61.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., & Flannery, B.P. (2002)

Numerical recipes in C++: The art of scientific computing. Second Edition. Cambridge University Press.

Reisby, N., Gram, L.F., Bech, P., Nagy, A., Petersen, G.O., Ortmann, J., Ibsen, J., Dencker, S.J., Jacobsen, O., Krautwald, O., Søndergaard, I., & Christiansen, J. (1977). Imipramine: clinical effects and pharmacokinetic variability, *Psychopharmacology*, **54**, 263-272.

Schafer, J. (1997). *Analysis of Incomplete Multivariate Data*. London: Chapman & Hall.

Schoenwald, S.K. & Henggeler, S.W. (2005). Multisystemic Therapy for Adolescents with Serious Externalizing Problems. In J. Lebow (ed.), *Handbook of Clinical Family Therapy*. Wiley: Jossey-Bass.

Singer, J.D. & Willett, J.B. (1993). It's about time: using discrete-time survival analysis to study duration and the timing of events. *Journal of Educational Statistics*, **18**, 155-195.

Sivo, S.A., & Willson, V.I. (1998). Is parsimony always desirable? Identifying the correct model for a longitudinal panel data set. *J. Experiment. Educ.*, **66**, 249-255.

Snijders, T. & Bosker, R. (1999). *Multilevel Analysis. An introduction to basic and advanced multilevel modeling*. Sage.

Stevens, S.S. (1946). On the theory of scales of measurement. *Science*, **103**, 677-680.

Stroud, A.H. & Secrest, D. (1966). *Gaussian quadrature formulas.* New Jersey: Prentice-Hall.

Thomson, W.A. (1977). On the treatment of grouped observations in life studies. *Biometrics*, **33**, 463-470.

Vonesh, E.F. & Carter, R.L. (1992). Mixed-effects nonlinear regression for unbalanced repeated measures. *Biometrics*, **48**(**1**):1-17.

Willet, J., & Sayer, A. (1994). Using covariance structure analysis to detect correlates and predictors of individual change over time. *Psychol. Bull.*, **116**(2), 363-380.

12 Subject Index

B

C

D

E

F

G

H

I

L

M

N

O

P

Q

R

S

T

U

V

W