



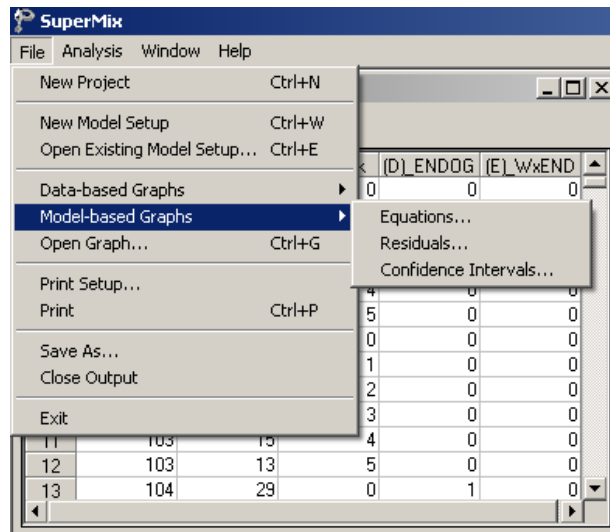
Contents

<i>Model-based graphs</i>	1
Graphing model equations	2
Residual plots.....	11
Confidence interval plots.....	14

Model-based graphs

Three types of model-based graphs can be produced with **SuperMix**. Plots of predicted outcomes based on the model equations can be made, and these graphs can be displayed by group or marked by a third variable. Residual plots (normal outcomes only), again marked by an additional variable, can also be made. Finally, an option to plot model-based confidence intervals of random coefficients is available. Again, these plots can be displayed by group or by marking variable.

Model-based graphing options are accessed via the **File, Model-based Graphs** option. The pop-up menu activated through this selection has three options – **Equations**, **Residuals** and **Confidence Intervals** – each associated with one of the type of plots described above.



Graph parameters that may be changed include the axes and descriptions thereof, the symbols used, and the colors assigned to the symbols/text. To change any of these, simply double-click on the symbol/text to be changed to activate a dialog box in which changes can be made. Use of such dialog boxes will be illustrated in the course of the discussion of the various graphs in the sections to follow.

Graphing model equations

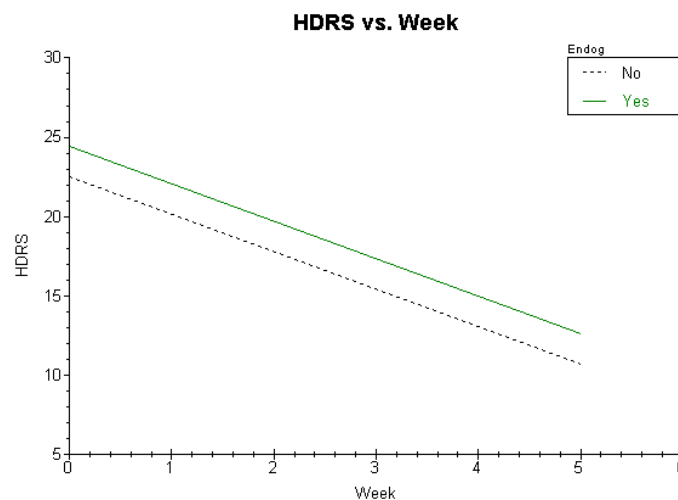
A graph frequently used as part of initial exploratory analysis of data is the scatter and/or line plot. This type of plot is used to examine potential relationships between an outcome variable and possible predictor variables, using the observed data. After fitting a model, one can plot the relationship between the outcome variable and predictor variables, subject to the model fitted to the data.

Creating an equation based graph for a two-level model

Recall that, for the Reisby data, we looked at the patients' HDRS ratings at six time points. In Section 3.2, we fitted a model with random intercept and slope for the response variable HDRS with the fixed predictors Week, ENDOG and the interaction effect between Week and ENDOG to the data. In this model

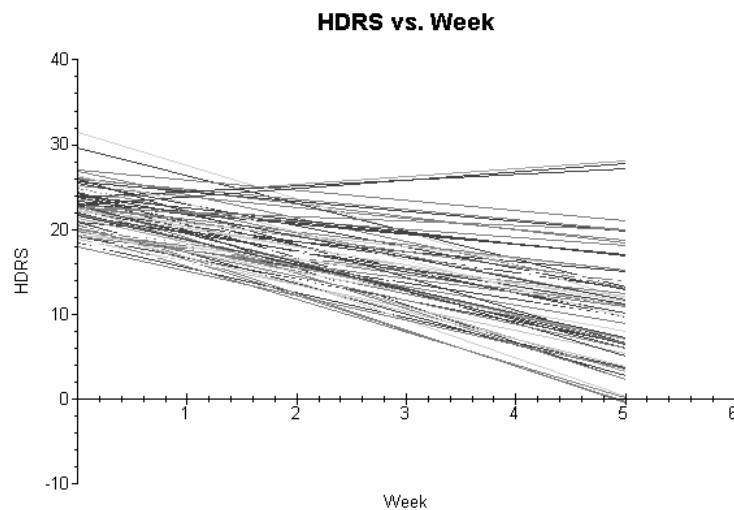
$$\text{HDRS}_{ij} = \beta_0 + \beta_1 * (\text{Week})_{ij} + \beta_2 * (\text{ENDOG})_i + \beta_3 * (\text{WxENDOG})_{ij} + u_{i0} + u_{i1} (\text{Week})_{ij} + e_{ij}$$

β_0 denotes the average expected depression rating scale value, β_1 and β_2 denote the coefficients of the predictor variables Week and ENDOG in the fixed part of the model, β_3 denotes the coefficient of the interaction between Week and ENDOG in the fixed part of the model, u_{i1} denotes the variation in the Week slopes over patients, and u_{i0} and e_{ij} denote the variation in the average expected HDRS value over patients and over measurements (*i. e.*, between patients) respectively.



The graph above shows the predicted HDRS ratings over the study period for the two groups of patients. The average predicted HDRS rating of patient with endogenous depression is consistently higher than that of the patients with non-endogenous depression.

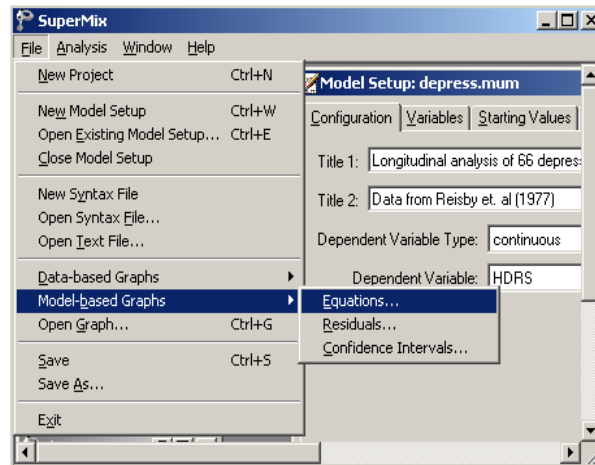
This trend holds for most of the patients, as illustrated in the graph below, where the predicted HDRS ratings of individual patients were plotted. In this graph, the predictors not used in the graph (i.e., ENDOG and WxENDOG) were held at the mean values, in contrast to the graph shown above where WxENDOG was held constant at a value of 0.



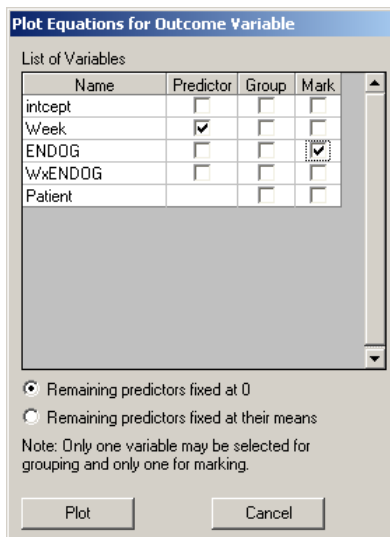
Care should be taken when making these graphs to verify that, if the default value of 0 is assumed for predictors not included, the value 0 is within the range of possible predictor values. In our case, the variables ENDOG and WxENDOG both include observed values/categories equal to zero, and the choice between holding the predictors constant at zero or the mean is a question of personal preference. However, should the variable ENDOG for example be recoded to have values 1 and 2 for the depression categories rather than the current 0,1 coding, interpretation of the graph shown above would be problematic. The variables ENDOG and ENDOG would then both be held constant at nonexistent values in terms of their ranges of observation/coding.

Creating an equation based graph

To create the model-based graph of HDRS ratings over the study period for the two groups of patients shown previously, open both the data and model files in the Graphics folder, in this case **reisby.ss3** and **depress.mum**, and run the model. Next, select the **File, Model-based Graphs, Equations** option from the main menu bar. The **Plot Equations for: HDRS** dialog box is displayed. Note that it is assumed that the outcome variable used in the model, in this case HDRS, is the variable that will be displayed on the vertical axis of the graph.



Select **Week** as the predictor variable that will be displayed on the X-axis of the graph by checking the box next to this variable in the **Predictor** column. To obtain separate lines for patients with and without endogenous depression, select the variable **ENDO** in the **Mark** column. Note that only one grouping variable and one marking variable may be selected.

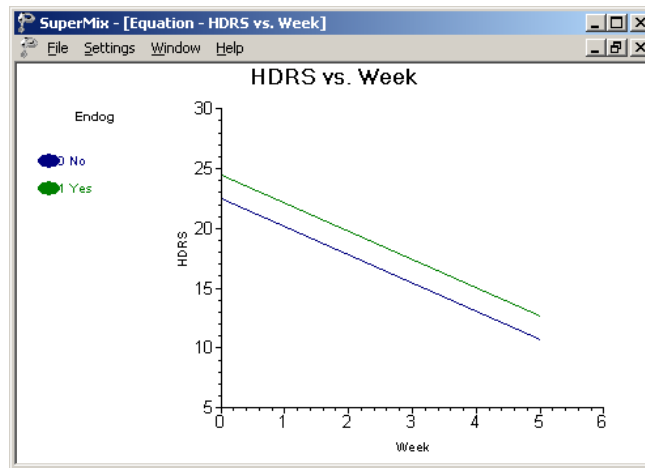


By default, predictors not selected will be held constant at zero, as indicated by the default selection of the **Remaining predictors fixed at 0** option. Click **Plot** to display the graph shown below.

Note that the two lines shown in this graph correspond to the model

$$\begin{aligned}\widehat{\text{HDRS}}_{ij} &= \hat{\beta}_0 + \hat{\beta}_1 * (\text{Week})_{ij} + \hat{\beta}_2 * (\text{ENDOG})_i + \hat{\beta}_3 * (\text{WxENDOG})_{ij} \\ &= \hat{\beta}_0 + \hat{\beta}_1 * (\text{Week})_{ij} + \hat{\beta}_2 * (\text{ENDOG})\end{aligned}$$

as WxENDOG is held constant to zero. This variable can assume a value of 0 when a patient exhibits non-endogenous depression (ENDOG = 0), or alternatively at the beginning of the study period when WEEK = 0. Consequently, the graph reflects the predicted HDRS ratings over time for patients meeting these requirements – in effect a subset of all the patients.



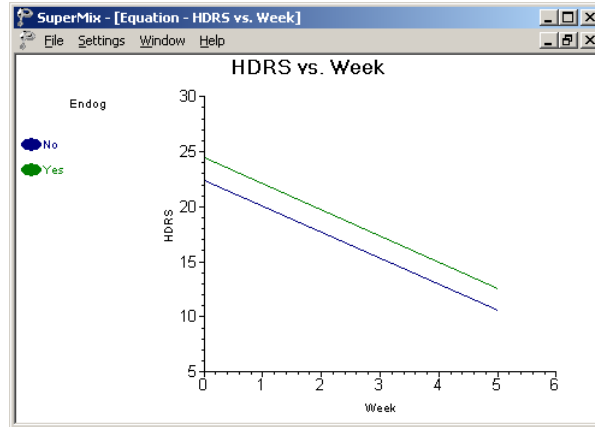
If, instead of holding the predictor WxENDOG equal to zero, this variable is held constant at its mean of 1.745, the same plot would usually look different.

For this case, the two lines shown in this graph correspond to the model

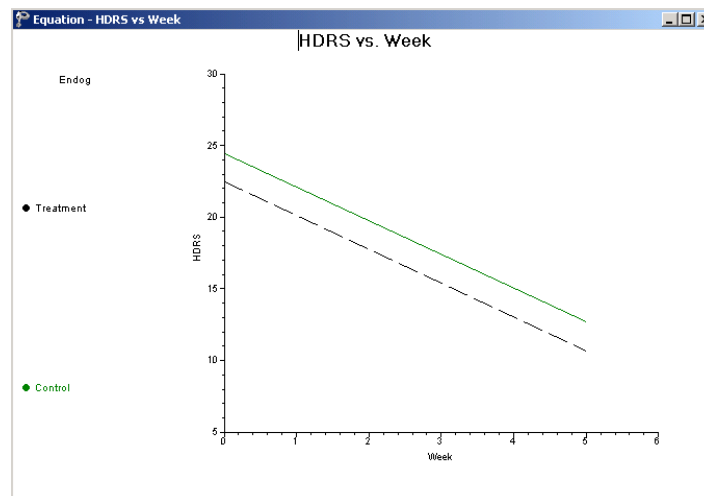
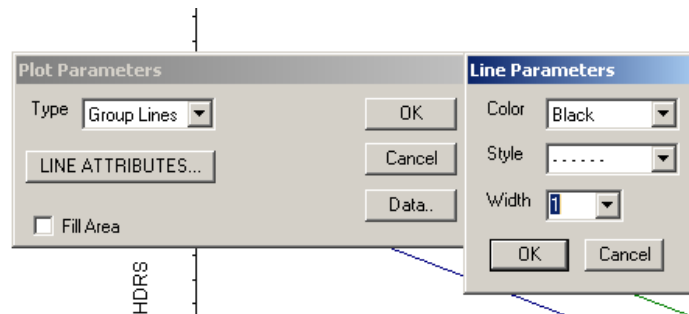
$$\begin{aligned}\widehat{\text{HDRS}}_{ij} &= \hat{\beta}_0 + \hat{\beta}_1 * (\text{Week})_{ij} + \hat{\beta}_2 * (\text{ENDOG})_i + \hat{\beta}_3 * (\text{WxENDOG})_{ij} \\ &= \hat{\beta}_0 + \hat{\beta}_1 * (\text{Week})_{ij} + \hat{\beta}_2 * (\text{ENDOG}) + \hat{\beta}_3 (1.745)\end{aligned}$$

and thus corresponds to the predicted outcome assuming a mean value for the interaction term.

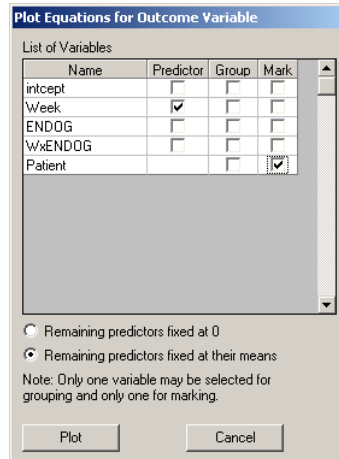
From Section 3.2, we have $\hat{\beta}_3 = -0.02$, and therefore $\hat{\beta}_3 (1.745) = -0.035$. The difference between this value and zero is too small to see the effect on the graphical display below, which appears to be identical to the "remaining predictors fixed at 0" selection. However, it gives support to the finding of a non-significant interaction effect.



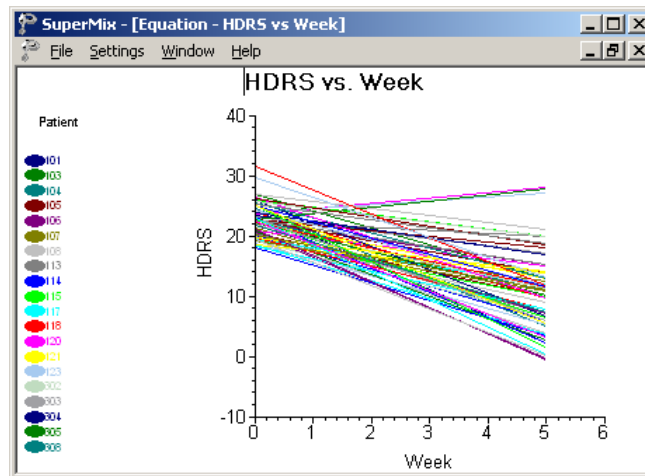
The graph shown above distinguishes between the two depression groups by using blue and green lines (green for "yes", blue for "no"). For publication purposes, the difference may be clearer if both lines are shown in black, but with different line styles. To change the graph, double-click on the lower of the two lines to activate the **Plot Parameters** dialog box. Click the **Line Attributes** button to display the **Line Parameters** dialog box and set the **Color** to black, the style to "...." and the width to 1 as shown below. Click **OK** on both dialog boxes to display the modified graph.



To create a graph displaying the predicted outcomes over time by individual patient, close the graphing window to return to the **Plot Equations for: HDRS** dialog box. Select the variable Patient as marking variable. Ensure that the **Remaining predictors fixed at their means** option is enabled. Click **Plot** to display this graph, as shown below.



Apart from a few patients for whom an increase in the HDRS ratings is predicted over the course of the study, the general trend is for HDRS to decrease over time.

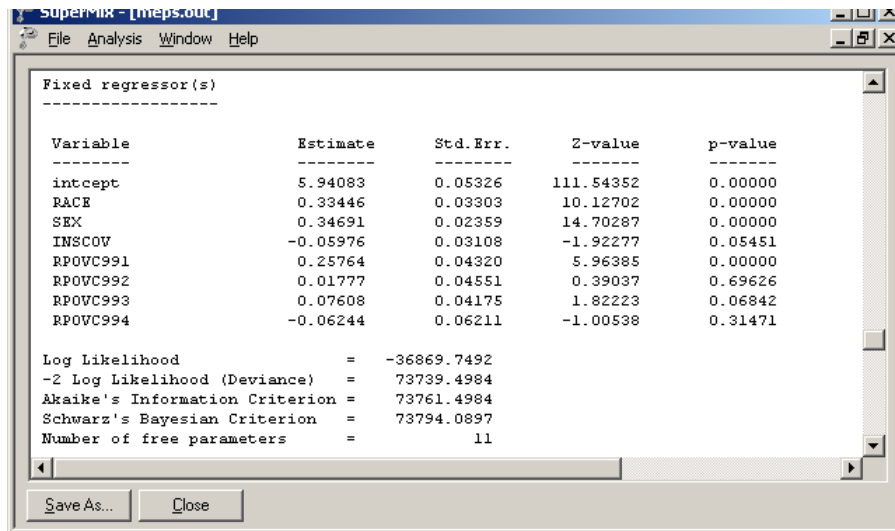


Creating an equation based graph for a three-level model

In Section 3.3 of the primer a model using the participant's gender, ethnicity, type of health insurance coverage, and measure of income relative to poverty level was fitted to predict the total expenditure on health care in 1999, with expenditure transformed to the natural logarithm of the actual expenses incurred. The model fitted was of the form

$$\begin{aligned} \text{TOTEXP99}_{ijk} = & \beta_0 + \beta_1 * \text{SEX}_{ijk} + \beta_2 * \text{RACE}_{ijk} + \beta_3 * \text{INSCOV}_{ijk} + \\ & \beta_4 * \text{RPOVC991}_{ijk} + \beta_5 * \text{RPOVC992}_{ijk} + \beta_6 * \text{RPOVC993}_{ijk} + \\ & \beta_7 * \text{RPOVC994}_{ijk} + v_{i0} + u_{ij0} + e_{ijk} \end{aligned}$$

For this model, the output shown below was obtained.



The screenshot shows the output of a fixed regressor model in SuperMIX. The window title is 'SuperMIX - [meps.oua]'. The menu bar includes 'File', 'Analysis', 'Window', and 'Help'. The output is displayed in a text area with the following content:

Variable	Estimate	Std. Err.	Z-value	p-value
intcept	5.94083	0.05326	111.54352	0.00000
RACE	0.33446	0.03303	10.12702	0.00000
SEX	0.34691	0.02359	14.70287	0.00000
INSCOV	-0.05976	0.03108	-1.92277	0.05451
RPOVC991	0.25764	0.04320	5.96385	0.00000
RPOVC992	0.01777	0.04551	0.39037	0.69626
RPOVC993	0.07608	0.04175	1.82223	0.06842
RPOVC994	-0.06244	0.06211	-1.00538	0.31471

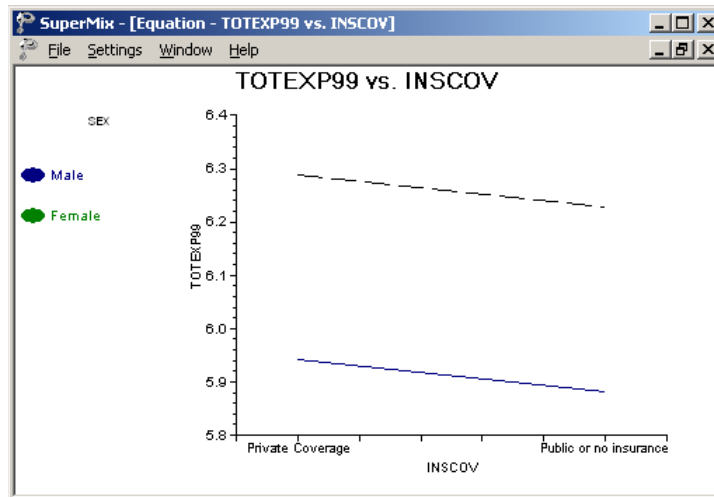
Below the table, the following statistics are listed:

- Log Likelihood = -36869.7492
- 2 Log Likelihood (Deviance) = 73739.4984
- Akaike's Information Criterion = 73761.4984
- Schwarz's Bayesian Criterion = 73794.0897
- Number of free parameters = 11

At the bottom of the window, there are buttons for 'Save As...' and 'Close'.

Recall that a value of 1 for the gender indicator variable SEX indicated that a participant was female, with a value of 0 assigned to male participants. Female participants are expected to have a total health expenditure 0.93063 higher than male participants do if all other variables are held constant. In contrast, participants with public coverage or no coverage have a lower expected total expenditure, as indicated by the negative estimated coefficient -0.61785 for the predictor INSCOV.

Using the **Equations** option on the **File, Model-based Graphs** menu, the plot shown below was obtained. This graph shows the predicted total expenditure on health care for the two insurance coverage groups, marked by gender. As indicated by the output shown above, the total expenditure for females is appreciably higher than for males, and respondents with private insurance coverage tend to spend less than those without coverage or with public insurance coverage.



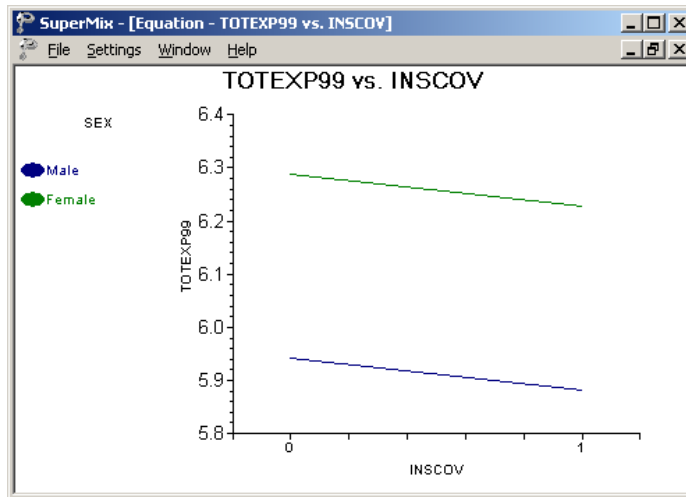
Creating an equation based graph

To create the graph shown above, start by opening both the data and model files, in this case **Examples\Primer\Graphics\meqs.ss3** and **meqs.mum**. After running the analysis, select the **File, Model-based Graphs, Equations** option from the main menu bar. The **Plot Equations for: TOTEXP99** dialog box is displayed. Select the predictor INSCOV for display on the horizontal axis by checking the box next to the variable name in the **Predictor** column, and request separate plots for the gender groups by selecting SEX in the **Mark** column. Click **OK**.

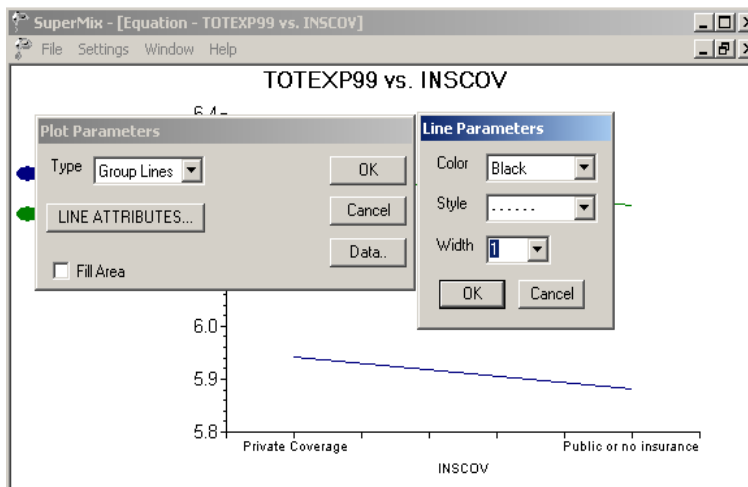
Name	Predictor	Group	Mark
intcept	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
RACE	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
SEX	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
INSCOV	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
RPOVC991	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
RPOVC992	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
RPOVC993	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
RPOVC994	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
VARPSU99	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
VARSTR99	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

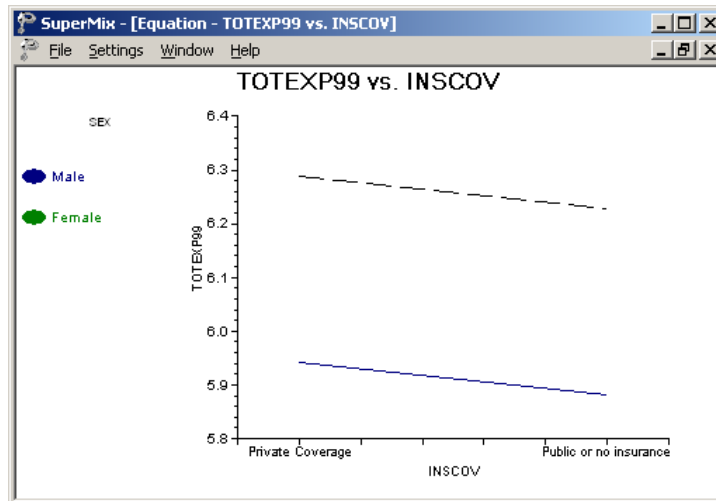
Remaining predictors fixed at 0
 Remaining predictors fixed at their means
 Note: Only one variable may be selected for grouping and only one for marking.

The graph shown below is now displayed. For publication purposes, the distinction between the two lines shown may be emphasized by changing the style of one of the lines. To change the attributes of the line for female respondents, *i.e.* the higher of the two lines, double-click on this line to activate the **Plot Parameters** dialog box.



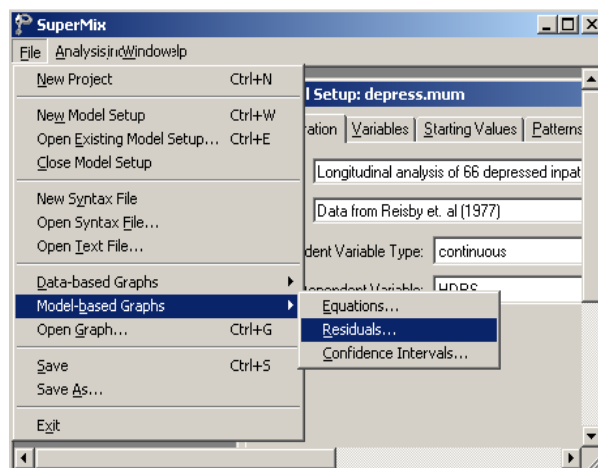
Click the **Line Attributes** button to open the **Line Parameters** dialog box. Change the **Color** of the line to black, and select a dotted line from the **Style** list box. Adjust the width of the line to 1 in the **Width** list box, and click **OK** on both the **Line Parameters** and **Plot Parameters** dialog boxes to obtain the revised graph.





Residual plots

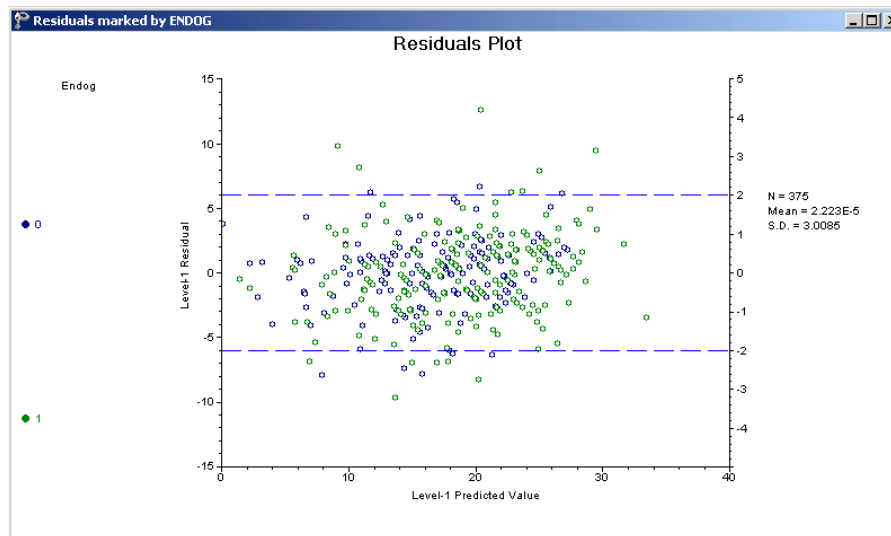
The **Residuals** option on the **File, Model-based Graphs** menu is used to examine the residuals obtained for a fitted model. It is useful for examining the fit of the model, and also as a check for possible distributional assumption violations. As residuals are defined as the difference between the observed and predicted outcome, 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. To request the creation of a residual file, set the **Write Bayes Estimates** field on the **Configuration** tab to yes.



The image below shows the residuals for a model fitted to the Reisby data, in which the time of measurement, depression status, and interaction between the depression status and time of measurement, were used to predict the HDRS ratings of patients. The model fitted to the data was of the form

$$\text{HDRS}_{ij} = \beta_0 + \beta_1 * (\text{Week})_{ij} + \beta_2 * (\text{ENDO G})_i + \beta_3 * (\text{WxENDO G})_{ij} + u_{i0} + u_{i1} (\text{Week})_{ij} + e_{ij} .$$

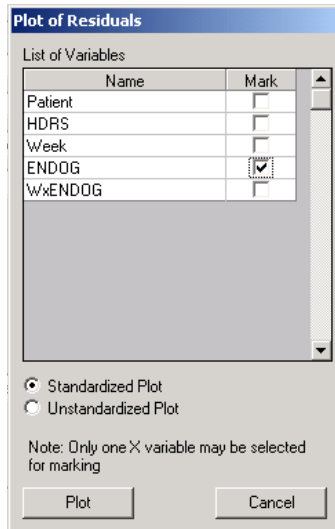
If a perfect fit is obtained, the residuals would all be equal to zero. In practice, this is not the case, but clustering of the residuals around 0 is usually observed. Using the 0 tick mark on the vertical axis as guideline, we observe that approximately half of the residuals are above this mark, and half below. It is interesting to note that the largest positive residuals occur for measurements associated with patients with endogenous depression (indicated by circles in the graph), and that most of the largest negative residuals are those for patients with non-endogenous depression. This indicates that, under the fitted model, some of the predicted measurements for endogenously depressed patients were smaller than the observed measurements, while the opposite is true for non-endogenously depressed patients. This occurs for a small percentage of the measurements, most of which are considerably closer to zero and tightly clustered. As such, we do not suspect systematic deviations from the fitted model over the range of the outcome variable.



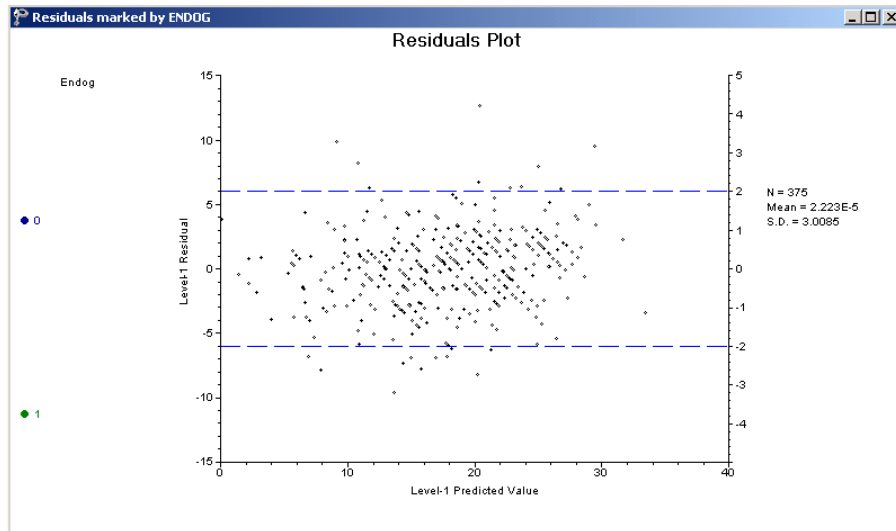
Creating a residual plot

To create the model-based graph of HDRS rating residuals shown above, start by opening the data and model files for the Reisby data (**Examples\Primer\Graphics\reisby.ss3, depress.mum**) and run the analysis. Next, select the **Residuals** option on the **File, Model-based Graphs** menu to open the **Plot of Residuals** dialog box.

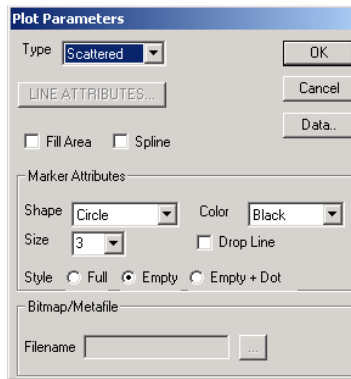
The **Plot of Residuals** dialog box offers the option to obtain either standardized or unstandardized residual plots. By default, a standardized plot of the residuals will be displayed. All the variables included in the model specification, including the level-2 identification variable Patient, are shown in the **List of Variables**. Any single variable on this list can be used as marking variable, to request different legends for its categories. Check the box for ENDOG, the dichotomous variable indicating the depression classification of a patient and click the **Plot** button.



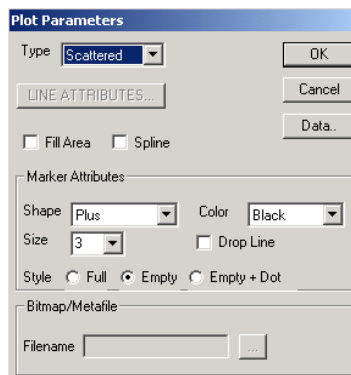
The plot obtained is the one shown above. By default, the same plotting symbol is used for both categories of the marking variable, but is shown in different colors (blue and green) for the two groups. If the plot is to be used in a report, typically produced in black and white format, it may be advisable to change the plotting symbols to make it easier to distinguish between the groups. To do so, double-click on any of the green circles towards the top of the graph.



The **Plot Parameters** dialog box is activated. While retaining a circle as the **Shape** of choice for this group, change the **Color** to black, and reduce the size of the plotting symbol to 3 by selecting this option from the **Size** list box as shown below. Click **OK** to return to the graphing window.

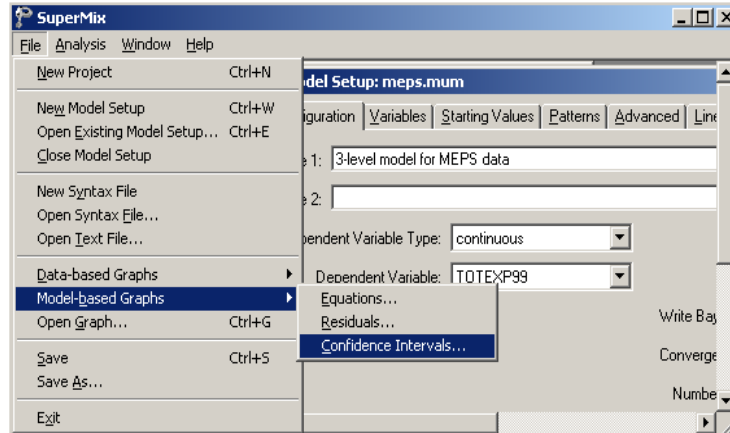


Next, click on any of the blue plot symbols at the bottom of the graph. The **Plot Parameters** dialog box is again activated, but now shows the details of the symbols used for the second category of the marking variable ENDOG. Set the **Shape** of the symbol to a plus sign instead of a circle, and use the **Color** and **Size** list boxes to request the display of these in black and of size 3. Click **OK**. The plot shown previously is obtained.



Confidence interval plots

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.



Using the three-level model discussed in Section 3.3 of the primer, graphs showing the confidence intervals for the empirical Bayes estimates of the random intercepts for level-2 and level-3 units may be obtained.

Recall that the model fitted to the MEPS data, formulated as

$$\begin{aligned} \text{TOTEXP99}_{ijk} = & \beta_0 + \beta_1 * \text{SEX}_{ijk} + \beta_2 * \text{RACE}_{ijk} + \beta_3 * \text{INSCOV}_{ijk} + \\ & \beta_4 * \text{RPOVC991}_{ijk} + \beta_5 * \text{RPOVC992}_{ijk} + \beta_6 * \text{RPOVC993}_{ijk} + \\ & \beta_7 * \text{RPOVC994}_{ijk} + v_{i0} + u_{ij0} + e_{ijk}, \end{aligned}$$

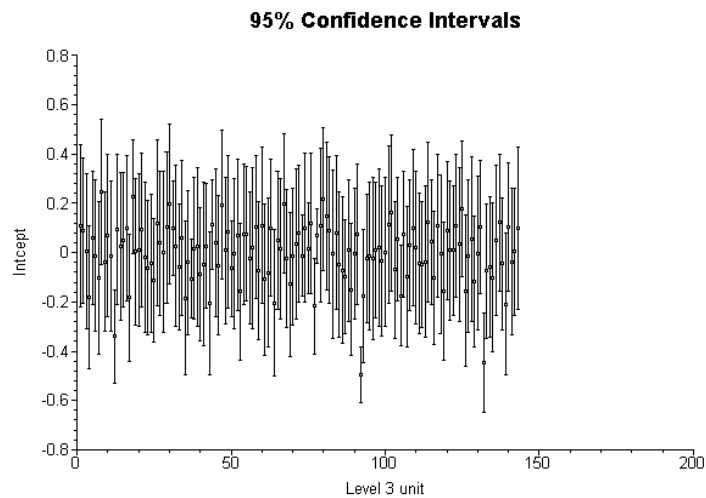
included a random intercept coefficient at both the PSU and stratum levels of the hierarchy (levels 2 and 3 respectively). Under this model, the expected total expenditure on health for a typical male respondent ($\text{SEX} = 0$) with no insurance coverage ($\text{INSCOV} = 0$) from the "negative or poor" income level ($\text{RPOVC991} = \text{RPOVC992} = \text{RPOVC993} = \text{RPOVC994} = 0$) can be calculated as

$$\begin{aligned} \widehat{\text{TOTEXP99}}_{ijk} &= \hat{\beta}_0 + \hat{\beta}_1 * 0 + \hat{\beta}_2 * \text{RACE}_{ijk} + \hat{\beta}_3 * 0 + \hat{\beta}_4 * 0 + \hat{\beta}_5 * 0 + \hat{\beta}_6 * 0 + \hat{\beta}_7 * 0 \\ &= \hat{\beta}_0 + \hat{\beta}_2 * \text{RACE}_{ijk} \end{aligned}$$

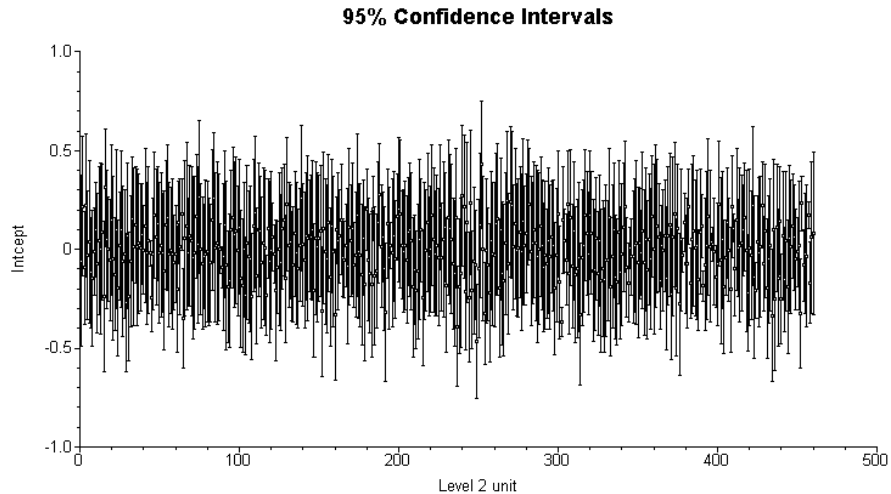
Using the estimates of β_0 and β_2 of 4.5841 and 0.68364 respectively as obtained under this model, a white respondent ($\text{RACE} = 1$) was expected to have an expenditure 0.68364 logarithmic units higher than a similar respondent with different ethnic background. While this result is of interest in own right, it does not provide insight in the extent to which respondents with this background differ in terms of where they are from – that is, in terms of the PSU and stratum they belong to. These characteristics are key to the survey design, and the formulated model makes provision for intercepts to differ from stratum to stratum at the highest level, and from PSU to PSU at level 2 of the model.

To take the structure of the model into account and obtain unique estimates of the predicted total expenditure for each stratum, the extent to which the intercept of a stratum deviates from the average must be taken into account. This unique stratum contribution is represented by the random coefficient ν_{i0} in the model. Estimates of these coefficients for the strata are referred to as the empirical Bayes residuals for the random level-3 intercepts. When the estimate of the ν_{i0} under the model is added to the expected outcome for a typical respondent, the empirical Bayes estimate of the total expenditure specific to a stratum is obtained. Using the empirical Bayes estimates for the strata and their corresponding variances, we can plot confidence intervals for the random intercept of each stratum. A plot showing the 95% confidence intervals for the strata intercepts is shown below. Each confidence interval is obtained using

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

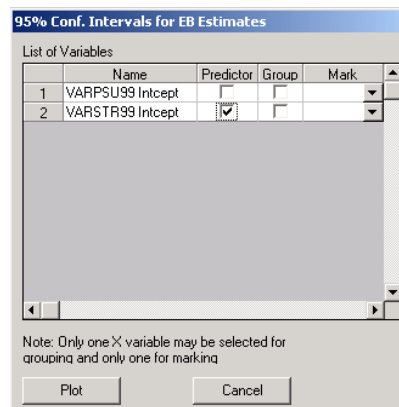


A similar graph can be obtained for the level-2 units, defined by the PSUs in this example.



Creating confidence interval plots

Start by opening the data and model specification files for the MEPS data, named **meps.ss3** and **meps.mum** respectively. After running the analysis, select the **Confidence Intervals** option from the **File, Model-based Graphs** menu to open the **95% Confidence Intervals for EB estimates** dialog box. Check the box associated with **VARSTR99 Intcept** in the **Predictor** column. Note that this box also allows the selection of grouping and marking variables to be used in the graphical display. Click **Plot** to display the graph.



To obtain a similar graph for the empirical Bayes intercepts of level-2 units, close the graph to return to the **95% Confidence Intervals for EB estimates** dialog box. Deselect **VARSTR99 Intcept** in the **Predictor** column and select **VARPSU99 Intcept** instead as shown below. Click **Plot** to display the graph for level-2 units shown earlier. Note that one may check both the **VARSTR99** and **VARPSU99 Intcept** boxes to obtain the two graphical displays simultaneously.

95% Conf. Intervals for EB Estimates

List of Variables

	Name	Predictor	Group	Mark	
1	VARPSU99 Intcept	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	▲
2	VARSTR99 Intcept	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	▼

Note: Only one X variable may be selected for grouping and only one for marking

Plot Cancel