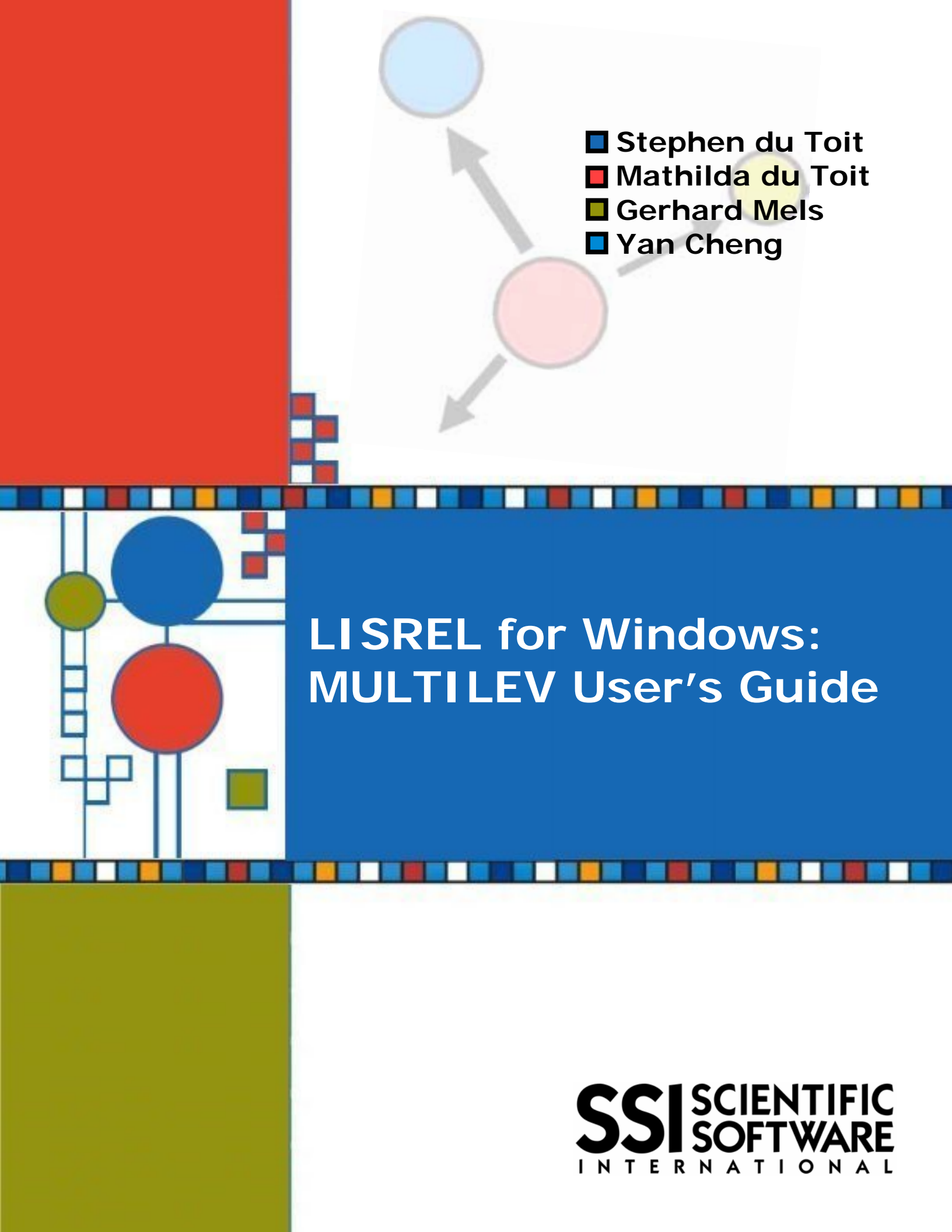


- 
- Stephen du Toit
 - Mathilda du Toit
 - Gerhard Mels
 - Yan Cheng

LISREL for Windows: MULTILEV User's Guide

SSI SCIENTIFIC
SOFTWARE
INTERNATIONAL

Table of contents

INTRODUCTION	1
GRAPHICAL USER INTERFACE	2
The Linear Model popup menu	2
The Title and Options dialog box	3
The Identification Variables dialog box.....	4
The Select Response and Fixed Variables dialog box	5
The Random Variables dialog box	6
The Non-Linear Model popup menu	7
The Title and Options dialog box	8
The ID, Response and Fixed Variables dialog box.....	9
The Select Model dialog box	10
The Select Covariate for First Component dialog box.....	11
The Select Covariate for Second Component dialog box	12
MULTILEV SYNTAX FILES	13
MULTILEV syntax files for linear models	13
The structure of the syntax file	13
CONTRAST command	14
COV1PAT command	15
COV2PAT command	16
COV3PAT command	17
COV1VAL command.....	19
COV2VAL command.....	19
COV3VAL command.....	20
DUMMY command.....	20
FIXED command.....	21
FIXPAT command	21
FIXVAL command.....	22
ID2 command	22
ID3 command	22
MISSING_DAT command.....	23
MISSING_DEP command	23
OPTIONS command	24
RANDOM1 command.....	30
RANDOM2 command.....	30
RANDOM3 command.....	31
RESPONSE command	31
SY command	31
TITLE command	32
WEIGHT1 command.....	32
WEIGHT2 command.....	33
WEIGHT3 command.....	33
MULTILEV syntax files for non-linear models	34
The structure of the syntax file	34
COVARIATES command	35
FIXED command.....	35
ID2 command	36
MISSING_DAT command.....	36

MODEL command	37
OPTIONS command	37
RESPONSE command	39
SY command	40
TITLE command	40
EXAMPLES.....	41
Two-level analysis of mice data	41
The data	41
Variance decomposition model with random intercepts	42
Fitting the variance decomposition model with random intercepts	43
Linear growth model with random intercepts	47
Fitting the linear growth model with random intercepts	48
Linear growth model with random intercepts and slopes	52
Fitting the linear growth model with random intercepts and slopes	53
Non-linear growth model with random intercepts and slopes.....	57
Fitting the non-linear growth model with random intercepts and slopes	58
Non-linear growth model with random intercepts slopes and a covariate	63
Fitting the non-linear growth model with random intercepts slopes and a covariate.....	64
A model with complex variation at level-1 of the model of the hierarchy.....	66
Fitting the model with complex variation at level-1 of the hierarchy	67
Two-level analysis of air traffic control data	73
The data	73
Variance decomposition model with random intercepts	74
Fitting the variance decomposition model with random intercepts	74
Non-linear growth model with random intercepts and slopes.....	79
Fitting the non-linear growth model with random intercepts and slopes	80
Non-linear growth model with random intercepts and slopes and with a covariate.....	85
Fitting the non-linear growth model with random intercepts and slopes and with a covariate	85
Three-level analysis of health expenditure data	91
The data	91
Regression model with random intercepts	92
Fitting the regression model with random intercepts	93
Three-level analysis of simulated data.....	101
The data	101
Linear growth model with random intercepts and slopes and a covariate	102
Fitting the linear growth model with random intercepts and slopes and a covariate	103
Two-level non-linear regression analysis of mice data	109
The data	109
Non-linear regression model.....	109
Fitting the non-linear regression model	110
REFERENCES	115

Introduction

The analysis of data with a hierarchical structure is known in the literature as, amongst others, hierarchical modeling, random coefficient modeling, latent curve modeling, growth curve modeling or multilevel modeling. Here we opt to use "multilevel modeling" to describe models exhibiting nested hierarchical structures.

The basic idea is that units, be it patients or measurements, are nested within units at a higher level of the hierarchy. For example, multiple blood pressure measurements may be “nested” within patients, where patients form the next, higher, level of the hierarchy. Alternatively, duration of stay within a hospital for each individual may form measurements nested within a hospital. Here the individuals are the lower-level units, nested within the hospitals that serve as the higher-level units. No matter which of these structures applies, the outcome measured at the lowest level may be described using regression coefficients at some or all the levels of the hierarchy. Variance components at different levels of the hierarchy can be included for study. This allows the researcher to evaluate the variation in outcome at various levels of the hierarchy, while inclusion of any moderating effects is optional. In addition, the dependence of repeated measurements belonging to one experimental unit in a typical growth curve analysis, for example, is taken into account with this approach. Multilevel models are also suited to the analysis of unbalanced data, and thus estimates can be obtained for units for which a very limited amount of information is available.

Multilevel models are particularly useful in the modeling of data from complex surveys. Cluster or multi-stage samples designs are frequently used for populations with an inherent hierarchical structure. Ignoring the hierarchical structure of data has serious implications. The use of alternatives such as aggregation and disaggregation of information to another level can induce an increase in collinearity among predictors and large or biased standard errors for the estimates.

LISREL for Windows (Jöreskog & Sörbom 2005) includes the application MULTILEV, which implements the methods in Jöreskog et al. (2001) and du Toit & du Toit (2001), to fit linear and non-linear multilevel models to multilevel data obtained from complex and simple random samples.

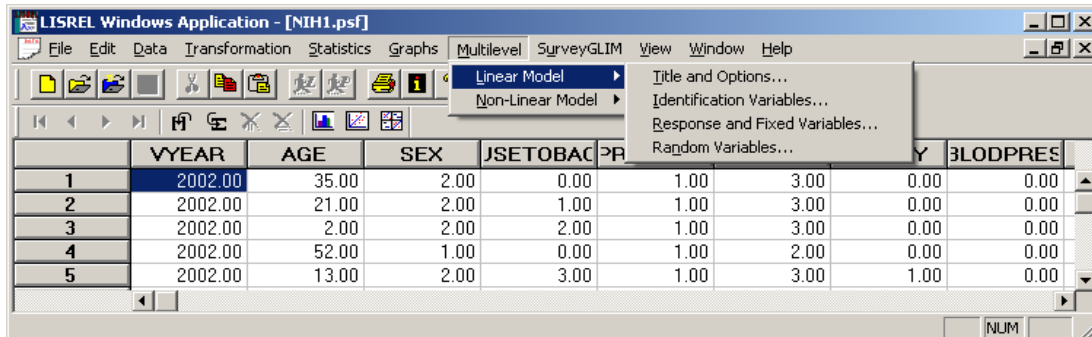
In this document, we describe and illustrate the features of MULTILEV. Section 2 is an overview of the MULTILEV graphical user interface (GUI). An overview of the MULTILEV syntax file is given in Section 3. Section 4 consists of illustrative examples of 2-level and 3-level linear multilevel analyses and a 2-level non-linear multilevel analysis.

Graphical user interface

The GUI of MULTILEV consists of the **Linear Model** and **Non-Linear Model** popup menus and the corresponding dialog boxes on the **Multilevel** menu on the PSF (PRELIS System File) window of LISREL for Windows. Both these popup menus and the corresponding dialog boxes are discussed next.

The Linear Model popup menu

The **Linear Model** popup menu on the **Multilevel** menu provides you access to a sequence of four dialog boxes that can be used to create a MULTILEV syntax file for a linear multilevel model interactively. It is located on the PSF window of LISREL for Windows, which is used to display, manipulate and process raw data. In other words, this menu is only available when a PSF window is opened. To illustrate this, the PSF window for the file **NIH1.psf** in the **TUTORIAL** subfolder is shown below with the **Linear Model** popup menu expanded.

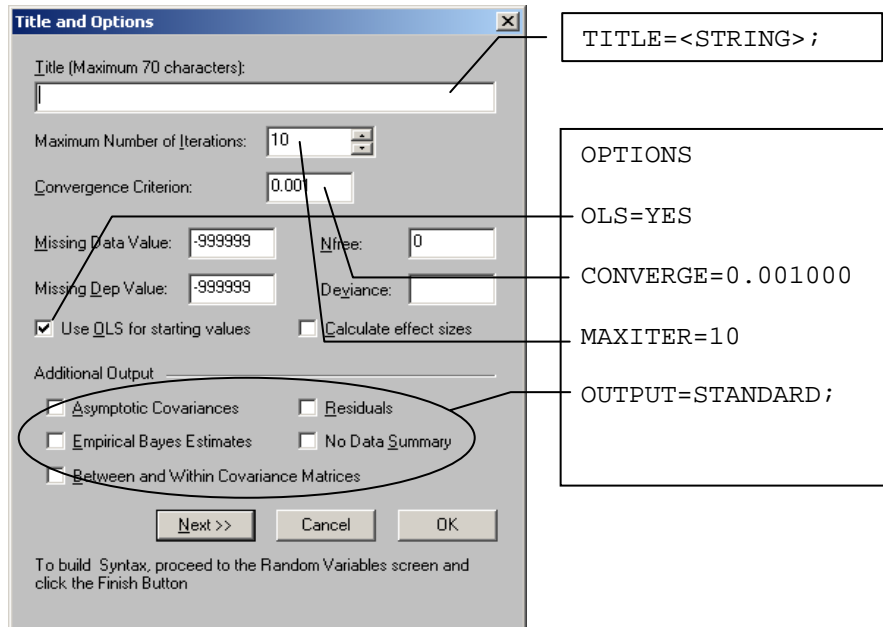


The **Linear Model** popup menu on the **Multilevel** menu has four options that can be used to perform basic linear multilevel analyses. Advanced options that enable the user to specify more complex models must be inserted manually once a syntax file has been generated.

The typical first step for using the **Linear Model** popup menu on the **Multilevel** menu would be to click on the **Title and Options** option to activate the **Title and Options** dialog box. However, you can click on the other options to go directly to the **Identification Variables**, the **Response and Fixed Variables** or the **Random Variables** dialog box.

The Title and Options dialog box

The **Title and Options** dialog box allows you to specify a title and the options of a linear multilevel analysis interactively and is accessed by selecting the **Title and Options** option on the **Linear Model** popup menu of the **Multilevel** menu. This selection loads the following **Title and Options** dialog box.



Note that the **Title and Options** dialog box corresponds with the `TITLE` and `OPTIONS` commands as indicated on the image above.

If desired, you can enter a descriptive title in the **Title** string field.

Since the linear multilevel estimation equations do not have a closed form solution, the Fisher scoring algorithm is used to obtain estimates of the unknown parameters. You can enter the maximum number of iterations for the algorithm in the **Maximum Number of Iterations** number field if the default of 10 is not appropriate. Enter the appropriate convergence criterion for the algorithm in the **Convergence Criterion** number field if the default value of 0.001 is not to be used.

If the raw data include missing values with a global missing value other than -999999, you need to enter the corresponding global missing value in the **Missing Data Value** number field. Similarly, if the global missing value for the response variable is not -999999, the corresponding global missing value needs to be entered in the **Missing Dep Value** number field.

The **NFree** number field allows you to specify the number of free parameters of the model of a previous analysis to be used for comparing nested models. You may provide the corresponding deviance statistic value by entering it in to the **Deviance** number field.

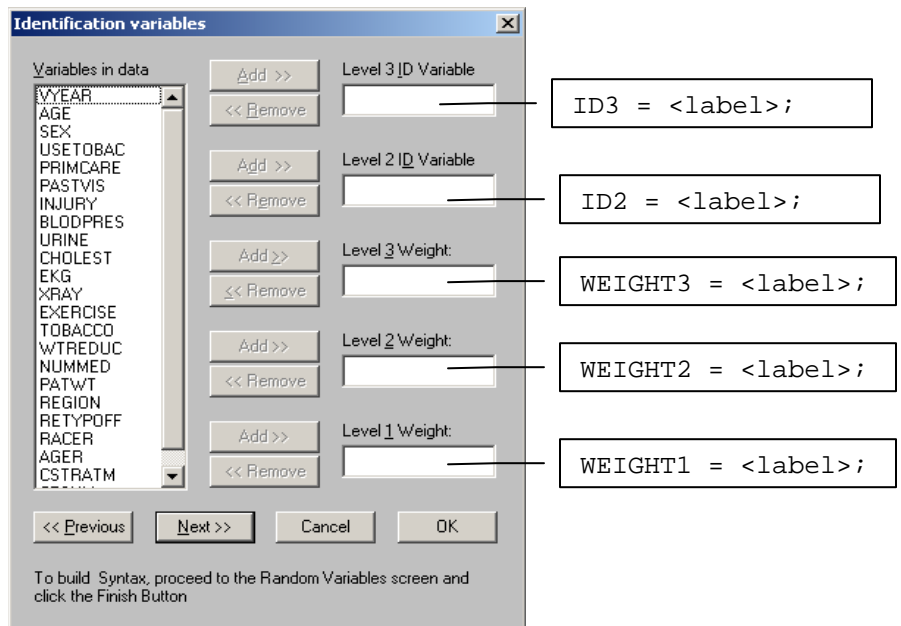
The **Use OLS for starting values** check box is checked by default. If this is not desired, the check box needs to be cleared. To request the calculation of effect sizes, you need to check the **Calculate effect sizes** check box.

The standard output can be changed by checking the desired check box(es) in the **Additional Output** section. If the estimated asymptotic covariance matrices of the parameter estimators of the fixed and random parts of the model are to be written to the external text files, you need to check the **Asymptotic Covariances** checkbox. Similarly, you need to check the **Empirical Bayes Estimates**, the **Between and Within Covariance Matrices** and/or the **Residuals** check box(es) to write the corresponding values to external text file(s). The data summary is listed in the output file by default. If this is not desired, check the **No Data Summary** check box to suppress it.

Once you are done with the **Title and Options** dialog box, click on the **Next** button to go to the **Identification variables** dialog box.

The Identification Variables dialog box

The **Identification variables** dialog box is used to specify the variables in the PSF that identify the various levels of the hierarchy. It is accessed by clicking on the **Next** button of the **Title and Options** dialog box or by selecting the **Identification Variables** option on the **Linear Model** popup menu. An example of this dialog box follows.



Note that the **Identification variables** dialog box corresponds with the ID2, ID3, WEIGHT1, WEIGHT2 and WEIGHT3 commands as indicated on the image above.

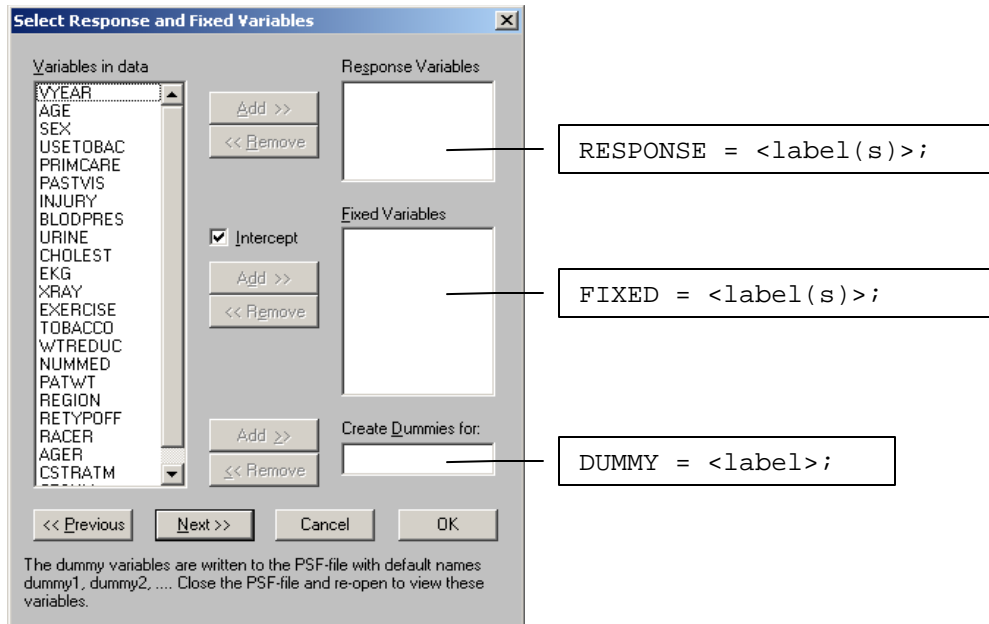
A level-2 ID variable is required when you set up a two or a three level model. To define a level-2 ID variable, first select the desired variable from the **Variables in data** list box and then click on the second **Add** button to add the variable to the **Level 2 ID Variable** grid box. Similarly, if a three level model is desired, the level-3 ID variable must be selected and added to the **Level 3 ID Variable** grid box by clicking on the top **Add** button.

The user can specify weight variables for each level of the model. If a level-1 weight variable is desired, the user may specify it by first selecting the desired weight variable from the **Variables in data** list box and then click on the corresponding **Add** button to add the variable to the **Level 1 Weight** grid box. Level 2 and level 3 weight variables are specified in a similar way.

Once the **Identification variables** dialog box has been completed, click on the **Next** button to access the **Select Response and Fixed Variables** dialog box.

The Select Response and Fixed Variables dialog box

The **Select Response and Fixed Variables** dialog box is used to select the outcome and fixed variables to be included in the model from the PSF. It is accessed by clicking on the **Next** button of the **Identification of Variables** dialog box or by selecting the **Select Response and Fixed Variables** option on the **Linear Model** popup menu. An example of this dialog box follows.



Note that the **Select Response and Fixed Variables** dialog box corresponds with the RESPONSE, FIXED and DUMMY commands as indicated on the image above.

The outcome variable(s) is specified by first selecting the variable(s) from the **Variables in data** list box and then clicking on the upper **Add** button to add the variable to the **Response Variables** list box.

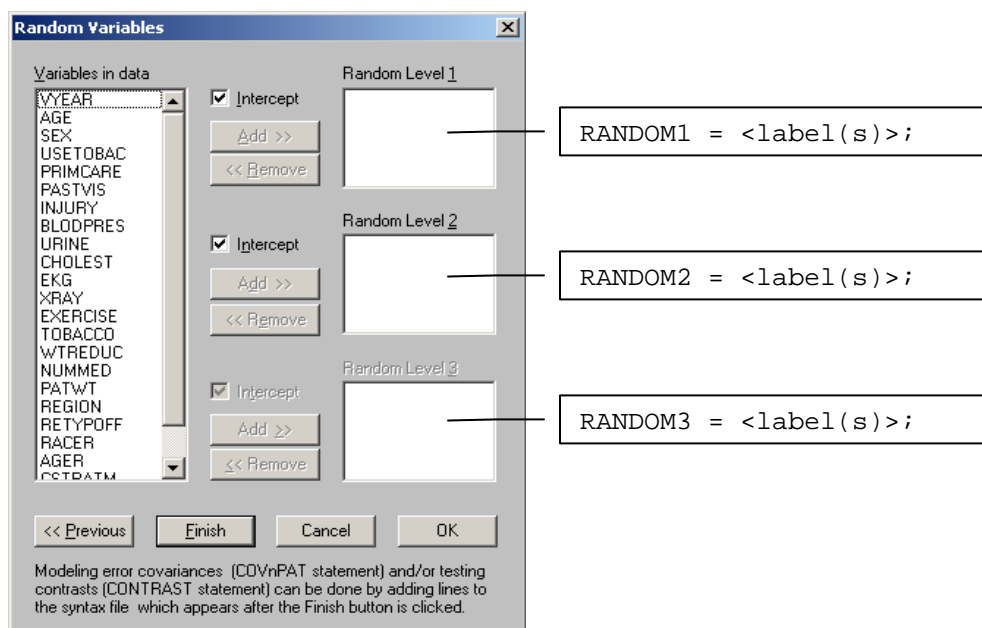
To define the independent variable(s) for the fixed part of the model, first select the variable(s) from the **Variables in data** list box and then click on the middle **Add** button to add the variable(s) to the **Fixed Variables** list box. Note that the intercept is included in the fixed part of the model by default. If this is not desired, clear the **Intercept** check box.

Dummy variables can be created for a categorical variable. To create the dummy variables for a categorical variable, first select the variable from the **Variables in data** list box and then click on the lower **Add** button to add the variable to the **Create Dummies for** grid box.

Once you are done with the **Select Response and Fixed Variables** dialog box, click on the **Next** button to go to the **Random Variables** dialog box.

The Random Variables dialog box

The **Random Variables** dialog box is used to specify the variables for which coefficients are assumed to be random. It is accessed by clicking on the **Next** button of the **Select Response and Fixed Variables** dialog box or by selecting the **Random Variables** option on the **Linear Model** popup menu. An example of this dialog box follows.



Note that the **Random Variables** dialog box corresponds with the RANDOM1, RANDOM2 and RANDOM3 commands as indicated on the image above.

To specify the independent variable(s) for the random part of the model at level-1, first select the variable(s) from the **Variables in data** list box and then click on the upper **Add** button to add the variable(s) to the **Random Level 1** list box. Note that the intercept is included in the random part of the model by default. If this is not desired, clear the **Intercept** check box.

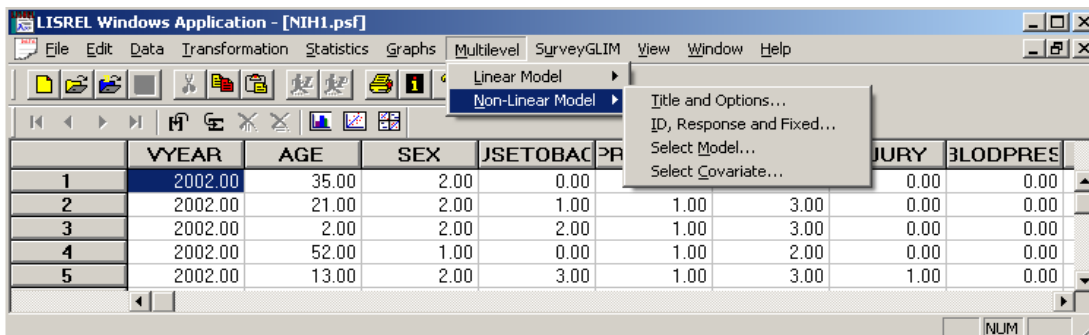
Similarly, the user may specify the level-2 independent variable(s) by first selecting them from the **Variables in data** list box and then by clicking on the middle **Add** button to add the variable(s) to the **Random Level 2** list box.

The level-3 independent variable(s) is specified by first selecting them from the **Variables in data** list box and then by clicking of the lower **Add** button.

Once the random part of the model is specified, click the **Finish** button to generate the text editor window for the corresponding MULTILEV syntax file.

The Non-Linear Model popup menu

The **Non-Linear Model** popup menu on the **Multilevel** menu of the PSF window provides you access to a sequence of four or five dialog boxes that can be used to create a MULTILEV syntax file for a nonlinear multilevel model interactively. This menu is only available when a PSF window is opened. To illustrate this, the PSF window for the file **NIH1.psf** in the **TUTORIAL** subfolder is shown below with the **Non-Linear Model** popup menu expanded.



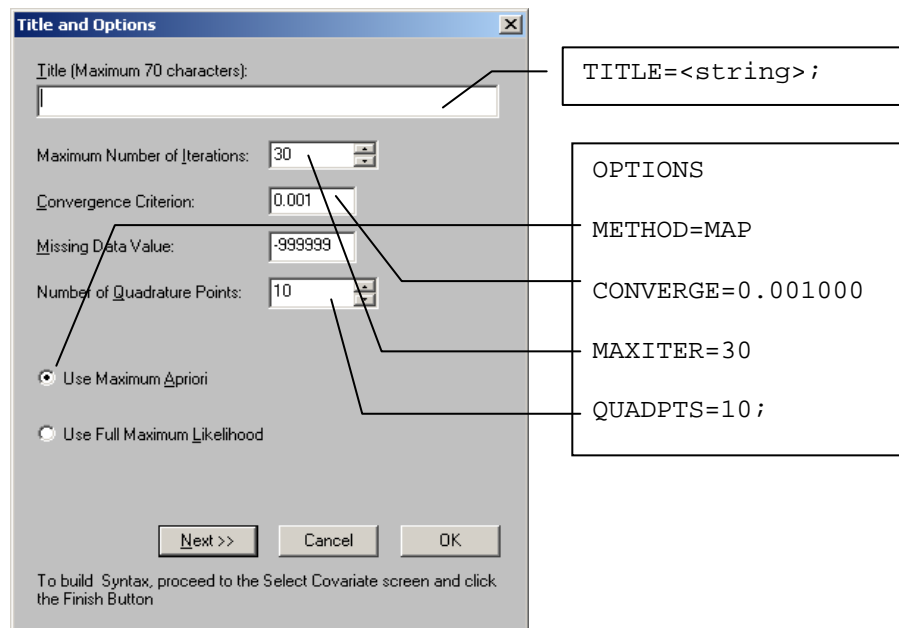
The **Non-Linear Model** popup menu on the **Multilevel** menu has four options that can be used to perform nonlinear multilevel analyses.

The typical first step for using the **Non-Linear Model** popup menu on the **Multilevel** menu would be to click on the **Title and Options** option to activate the **Title and Options** dialog box. However, you

can click on the other options to go directly to the **ID, Response and Fixed**, the **Select Model** or the **Select Covariate for First Component** dialog box.

The Title and Options dialog box

The **Title and Options** dialog box allows you to specify a title and the options of a non-linear multilevel analysis interactively and is accessed by selecting the **Title and Options** option on the **Linear Model** popup menu of the **Multilevel** menu. This selection loads the following **Title and Options** dialog box.



Note that the **Title and Options** dialog box corresponds with the **TITLE** and **OPTIONS** commands as indicated on the image above.

If desired, you can enter a descriptive title in the **Title** string field.

Since the non-linear multilevel estimation equations do not have a closed form solution, the Fisher scoring algorithm is used to obtain estimates of the unknown parameters. You can enter the maximum number of iterations for the algorithm in the **Maximum Number of Iterations** number field if the default of 30 is not appropriate. Enter the appropriate convergence criterion for the algorithm in the **Convergence Criterion** number field if the default value of 0.001 is not to be used.

The likelihood function for non-linear multilevel models is also not in closed form. Hence, the adaptive quadrature method for numerical iteration is used to compute the value of the likelihood function. The corresponding number of quadrature points is specified by using the **Number of Quadrature Points** number field unless the default of 10 quadrature points is sufficient.

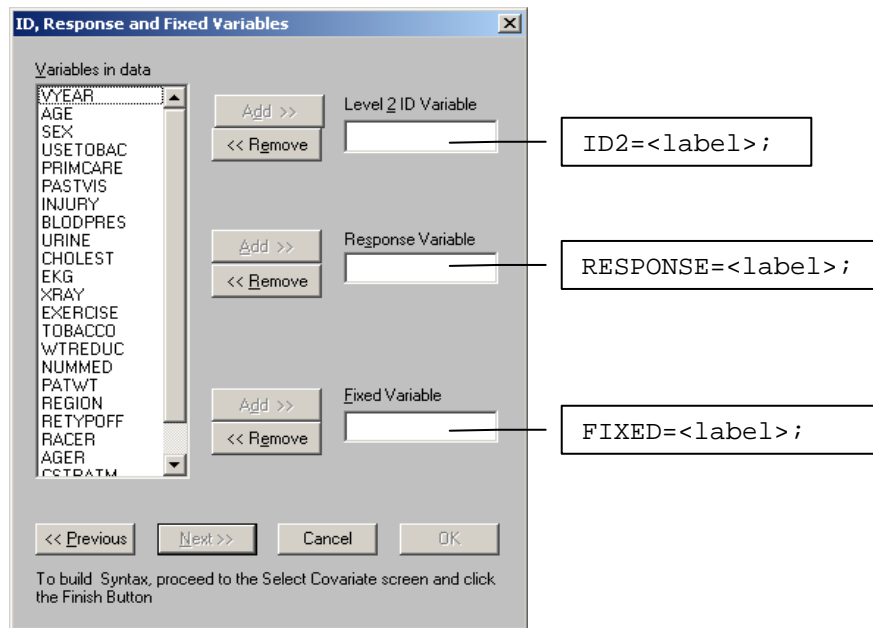
If the raw data include missing values with a global missing value other than -999999, you need to enter the corresponding global missing value in the **Missing Data Value** number field.

The **Use Maximum Apriori** radio button is activated by default to specify the Maximum Apriori estimation method. If the Maximum likelihood method is desired instead, you need to select the **Use Full Maximum Likelihood** radio button.

Once you are done with the **Title and Options** dialog box, click on the **Next** button to go to the **ID, Response and Fixed Variables** dialog box.

The ID, Response and Fixed Variables dialog box

The **ID, Response and Fixed Variables** dialog box is used to specify the level-2 ID variable, the response variable and the fixed variable in the PSF. It is accessed by clicking on the **Next** button of the **Title and Options** dialog box or by selecting the **ID, Response and Fixed Variables** option on the **Non-Linear Model** popup menu. An example of this dialog box follows.



Note that the **ID, Response and Fixed Variables** dialog box corresponds with the ID2, RESPONSE and FIXED commands as indicated on the image above.

To define a level-2 ID variable, first select the desired ID variable from the **Variables in data** list box and then click on the top **Add** button to add the variable to the **Level 2 ID Variable** grid box.

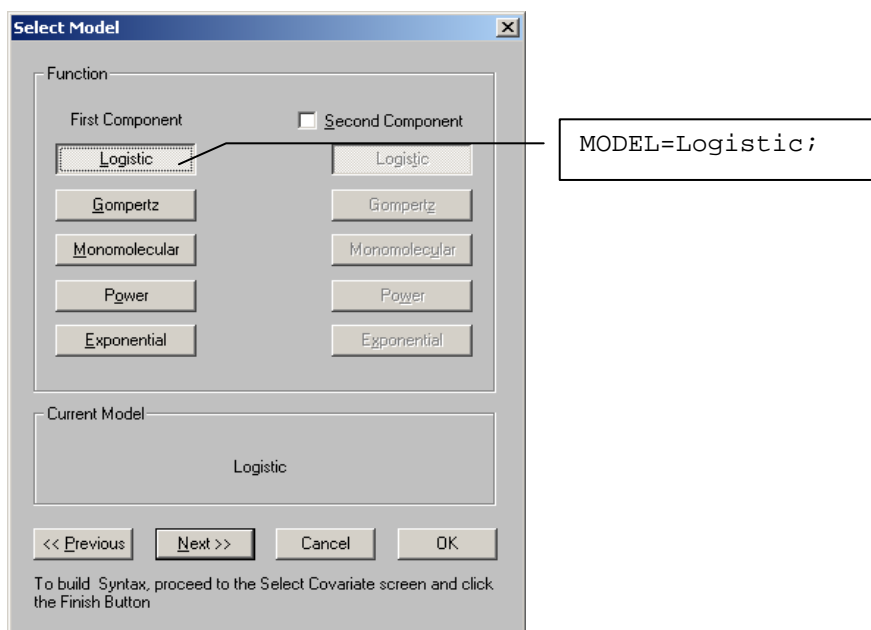
The response variable is specified by first selecting the desired variable from the **Variables in data** list box and then by clicking on the middle **Add** button to add the variable to the **Response Variable** grid box.

Similarly, the fixed variable is specified by first selecting the desired variable from the **Variables in data** list box and then by clicking on the bottom **Add** button.

Once the **ID, Response and Fixed Variables** dialog box has been completed, click on the **Next** button to access the **Select Model** dialog box.

The Select Model dialog box

The **Select Model** dialog box is used to select the functions for the first and second components of the model. It is accessed by clicking on the **Next** button of the **ID, Response and Fixed Variables** dialog box or by selecting the **Select Model** option on the **Non-Linear Model** popup menu. These selections load the following dialog box.



Note that the **Select Model** dialog box corresponds with the **MODEL** command as indicated on the image above.

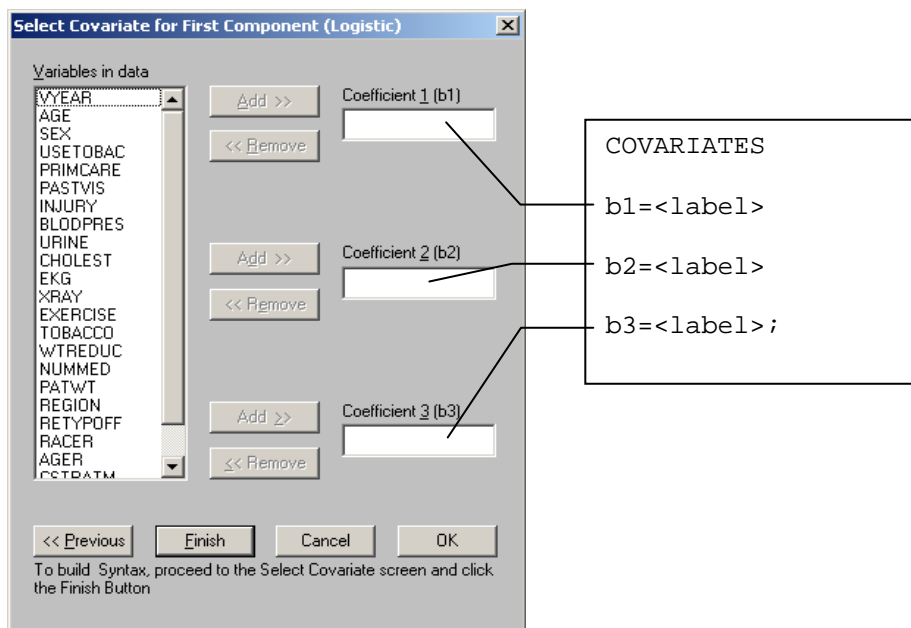
The available functions are **Logistic**, **Gompertz**, **Monomolecular**, **Power** and **Exponential**. **Logistic** is the default function for the **First Component**. If this is not desired, you may choose another function by clicking on the corresponding button.

To specify a function for the second component of the model, first check the **Second Component** check box to activate all the five available function buttons and then click on the button of the desired function name.

Once you are done with the **Select Model** dialog box, click on the **Next** button to go to the **Select Covariate for First Component** dialog box.

The Select Covariate for First Component dialog box

The **Select Covariate for First Component** dialog box is used to specify covariates for the random part of the model. It is accessed by clicking on the **Next** button of the **Select Model** dialog box or by selecting the **Select Covariate** option on the **Non-Linear Model** popup menu. An example of this dialog box follows.



Note that the **Select Covariate for First Component** dialog box corresponds COVARIATES command as indicated on the image above.

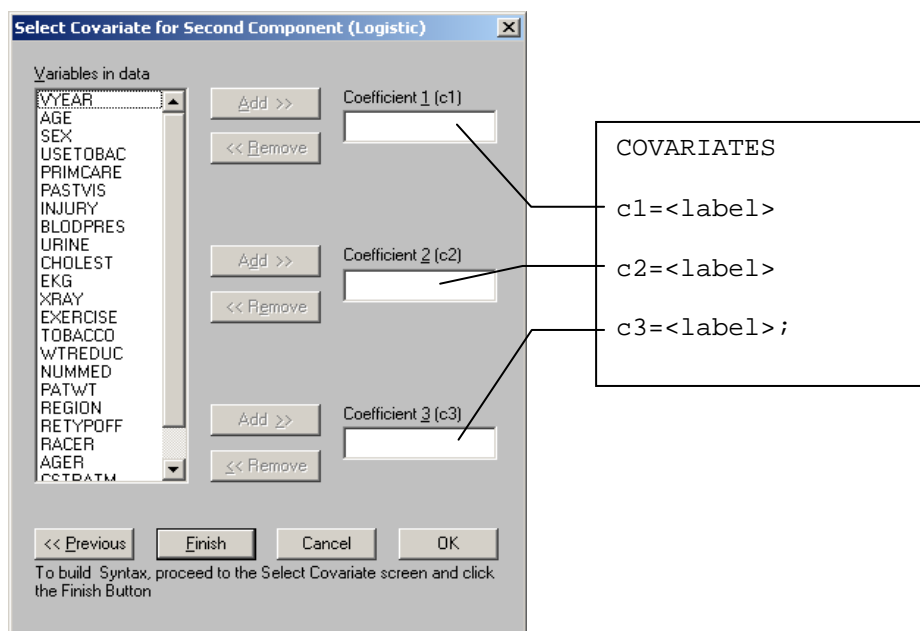
To specify the covariate for the first coefficient, first select the variable from the **Variables in data** list box and then click on the upper **Add** button to add the variable to the **Coefficient 1 (b1)** grid box. Similarly, you can specify covariates for the second and third coefficients.

Once the covariate(s) is specified, click the **Finish** button to generate the text editor window for the corresponding MULTILEV syntax file.

Note that if a function for the second component is specified in the **Select Model** dialog box, a **Next** button instead of **Finish** button is displayed on the **Select Covariate for First Component** dialog box. The user needs to click on the **Next** button to open the **Select Covariate for Second Component** dialog box and the specify covariates of the second component before generating the syntax file.

The Select Covariate for Second Component dialog box

The **Select Covariate for Second Component** dialog box is used to specify covariates for the second component of the model. It is only activated if a second component model is specified, in which case, it is accessed by clicking on the **Next** button of the **Select Covariate for First Component** dialog box. An example of this dialog box follows.



Note that the **Select Covariate for Second Component** dialog box corresponds COVARIATES command as indicated on the image above.

To specify the covariate for the first component, first select the variable from the **Variables in data** list box and then click on the upper **Add** button to add the variable to the **Coefficient 1 (c1)** grid box. Similarly, you can specify the covariates for the second and third coefficients.

Once the covariate(s) is specified, click the **Finish** button to generate the text editor window for the corresponding MULTILEV syntax file.

MULTILEV syntax files

MULTILEV syntax files for linear models

The structure of the syntax file

The MULTILEV syntax file generated by the MULTILEV GUI for a linear model can also be created by using the text editor of LISREL for Windows or any other text editor such as Notepad or WordPad. The structure of the MULTILEV syntax file for a linear model follows.

```
OPTIONS <options>;
SY = '<filename>';
ID2 = <label>;
ID3 = <label>;
WEIGHT1 = <label>;
WEIGHT2 = <label>;
WEIGHT3 = <label>;
MISSING_DAT = <number>;
MISSING_DEP = <number>;
RESPONSE= <label(s)>;
FIXED = <label(s)>;
RANDOM1 = <label(s)>;
RANDOM2 = <label(s)>;
RANDOM3 = <label(s)>;
TITLE = <string>;
CONTRAST = <filename>;
COV1VAL = <number(s)>;
COV2VAL = <number(s)>;
COV3VAL = <number(s)>;
COV1PAT = <pattern>;
COV2PAT = <pattern>;
COV3PAT = <pattern>;
FIXVAL = <number(s)>;
FIXPAT = < number(s)>;
DUMMY = <label>;
```

where <label> denotes a case sensitive variable name used in the raw data file, <label(s)> denotes a list of one or more variable names used in the raw data file, <number> denotes a specific number, <number(s)> denotes a list of one or more numbers, <string> represents a character string, <filename> denotes a complete name (including the drive and folder names) of a file, <pattern> is one of DIAG, TOEPLITZ, INTRA, MA1 or a matrix pattern (see the COV1PAT, COV2PAT and COV3PAT command sections) and <options> denotes a list of options for the analysis (see the OPTIONS command section).

The OPTIONS, SY, ID2, RESPONSE and FIXED commands and at least one of the RANDOM1, RANDOM2 and RANDOM3 commands are **required** commands while the other commands are all **optional**. The OPTIONS and SY commands should be the first two commands respectively, but the

other commands can be entered in any order. Except for variable labels, the contents of the syntax file are not case-sensitive. Blank lines can be inserted in any section of the syntax file.

In the following sections, the MULTILEV commands for a linear model are discussed separately in alphabetical order.

CONTRAST command

The CONTRAST command is used to specify the name of the text file containing information on any contrast(s) between the fixed effects in the model to be tested. It is an **optional** command.

Syntax

CONTRAST = <filename>;

where <filename> denotes the complete name (including drive and folder names) of the text file containing information on the fixed effects contrasts among fixed effects to be tested.

Example

Suppose that there are six fixed effects in a particular model, these being INTERCEPT, GENDER, MATHS, READING, SCIENCE and WRITING.

Suppose that one wishes to test the null hypothesis

$$H_0 : \beta_{READING} - \beta_{WRITING} = 0$$

$$\beta_{MATHS} - \beta_{SCIENCE} = 0$$

The null hypothesis above can also be expressed as

$$H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$$

where

$$\mathbf{0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 & -1 & 0 \end{bmatrix}$$

and

$$\boldsymbol{\beta} = [\beta_{INTERCEPT} \quad \beta_{GENDER} \quad \beta_{MATHS} \quad \beta_{READING} \quad \beta_{SCIENCE} \quad \beta_{WRITING}]'$$

Note that each row of \mathbf{C} has six elements, corresponding to the six fixed effects. The first contrast between fixed effects is specified in the first row. Since the fourth element in the first equals 1, and the sixth element is -1, this denotes a contrast between the effects $\beta_{READING}$ and $\beta_{WRITING}$.

The corresponding text file, MLEVEL.CTR, will have the following form.

```
2
0 0 0 1 0 -1
0 0 1 0 -1 0
```

The first row indicates the number of contrasts and the second and third rows the actual contrasts to be tested. The corresponding CONTRAST command is

```
CONTRAST=C:\MLEVEL\EXAMPLES\MLEVEL.CTR;
```

COV1PAT command

The COV1PAT command is used to place constraints on the elements of the level-1 covariance matrix $\Phi_{(1)}$. It is an **optional** command.

Syntax

```
COV1PAT= <pattern>;
```

where <pattern> is either DIAG for a diagonal matrix or a diagonal matrix pattern in lower triangular form.

Examples

```
COV1PAT = 1
          0 2
          0 0 2
          0 0 0 3;
```

```
COV1PAT = 1
          0 1
          0 0 1
          0 0 0 1;
```

Note

The zero entries off the diagonal indicate a covariance of zero while zero entries on the diagonal indicate fixed level-1 variances.

COV2PAT command

The purpose of the COV2PAT command is to impose constraints on the elements of the level-2 covariance matrix $\Phi_{(2)}$. It is an **optional** command.

Syntax

COV2PAT= <pattern>;

where <pattern> is one of

- DIAG In this case the level-2 covariance matrix of the model will be constrained to be a diagonal matrix.
- TOEPLITZ The level-2 covariance matrix will be constrained to be of the form of a so-called Toeplitz matrix, that is

$$\Phi_{(2)} = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & & \\ \gamma_1 & \gamma_0 & \gamma_1 & \ddots & \\ \gamma_2 & \gamma_1 & \gamma_0 & \ddots & \gamma_2 \\ & \ddots & \ddots & \ddots & \gamma_1 \\ & & \gamma_2 & \gamma_1 & \gamma_0 \end{bmatrix}$$

- INTRA The level-2 covariance matrix will be constrained to have an intra-class structure, that is

$$\Phi_{(2)} = \begin{bmatrix} \alpha & \beta & \dots & \dots & \beta \\ \beta & \alpha & \beta & \dots & \vdots \\ \vdots & \beta & \alpha & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \beta \\ \beta & \dots & \dots & \beta & \alpha \end{bmatrix}$$

- MA1 Constrains the level-2 covariance matrix to be similar to that of a time series process of order MA1. The form of the covariance matrix will then be

$$\Phi_{(2)} = \begin{bmatrix} \gamma & \beta & 0 & \dots & 0 \\ \beta & \gamma & \beta & \ddots & \vdots \\ 0 & \beta & \gamma & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \beta \\ 0 & \dots & 0 & \beta & \gamma \end{bmatrix}$$

or a symmetric matrix pattern in lower triangular form.

Example

```
COV2PAT = 1
      2 0
      2 2 0
      2 0 0 3;
```

Note

The zero entries indicate the fixed values which are specified in the corresponding COV2VAL command.

COV3PAT command

The COV3PAT command is used to impose constraints on the elements of the level-3 covariance matrix $\Phi_{(3)}$ and is an **optional** command.

Syntax

```
COV3PAT= <pattern>;
```

where <pattern> is one of

- | | |
|----------|---|
| DIAG | In this case the level-3 covariance matrix of the model will be constrained to be a diagonal matrix. |
| TOEPLITZ | The level-3 covariance matrix will be constrained to be of the form of a so-called Toeplitz matrix, that is |

$$\Phi_{(3)} = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & & \\ \gamma_1 & \gamma_0 & \gamma_1 & \ddots & \\ \gamma_2 & \gamma_1 & \gamma_0 & \ddots & \gamma_2 \\ & \ddots & \ddots & \ddots & \gamma_1 \\ & & \gamma_2 & \gamma_1 & \gamma_0 \end{bmatrix}$$

- | | |
|-------|---|
| INTRA | The level-3 covariance matrix will be constrained to have an intra-class structure, that is |
|-------|---|

$$\Phi_{(3)} = \begin{bmatrix} \alpha & \beta & \dots & \dots & \beta \\ \beta & \alpha & \beta & \dots & \vdots \\ \vdots & \beta & \alpha & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \beta \\ \beta & \dots & \dots & \beta & \alpha \end{bmatrix}$$

MA1 Constrains the level-3 covariance matrix to be similar to that of a time series process of order MA1. The form of the covariance matrix will then be

$$\Phi_{(3)} = \begin{bmatrix} \gamma & \beta & 0 & \dots & 0 \\ \beta & \gamma & \beta & \ddots & \vdots \\ 0 & \beta & \gamma & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \beta \\ 0 & \dots & 0 & \beta & \gamma \end{bmatrix}$$

or a symmetric matrix pattern in lower triangular form.

Examples

```
COV3PAT = 1
    0 3
    0 0 6
    0 0 0 10;
```

```
COV3PAT = 1
    2 0
    4 2 0
    0 0 2 0;
```

```
COV3PAT = 0
    2 0
    4 2 0
    0 0 2 0;
```

Note

The zero entries indicate fixed values specified in the corresponding COV3VAL command.

COV1VAL command

The COV1VAL command is used to provide initial values for the elements of the level-1 covariance matrix of the model and is an **optional** command.

Syntax

```
COV1VAL = <number(s)>;
```

where <number(s)> denotes a list of variances and covariances in the form of a diagonal matrix in lower triangular form.

Example

```
COV1VAL = 1.00  
          0.00 0.85  
          0.00 0.00 0.78  
          0.00 0.00 0.00 0.99;
```

COV2VAL command

The purpose of the COV2VAL command is to provide initial values for the elements of the level-2 covariance matrix of the model. The COV2VAL command is **optional**.

Syntax

```
COV2VAL = <number(s)>;
```

where <number(s)> denotes a list of variances and covariances in the form of a lower triangular matrix.

Example

```
COV2VAL = 1.00  
          0.32 0.85  
          0.63 0.62 0.78  
          0.19 0.00 0.25 0.99;
```

COV3VAL command

The COV3VAL command is used to provide initial values for the elements of the level-3 covariance matrix of the model and is an **optional** command.

Syntax

```
COV3VAL = <number(s)>;
```

where <number(s)> denotes a list of variances and covariances in the form of a lower triangular matrix.

Example

```
COV3VAL = 1.00  
          0.32 0.85  
          0.63 0.62 0.78  
          0.19 0.00 0.25 0.99;
```

DUMMY command

The DUMMY command is used to create dummy variables for a categorical variable. Names for the dummy variables are denoted by default by dummy1, dummy2, ..., dummyk, where k denotes the number of distinct values of the categorical variable. The DUMMY command is an **optional** command.

Syntax

```
DUMMY = <label1> PREFIX = <label2>;
```

where <label1> denotes the label of the categorical variable and <label2> denotes the prefix to be used for the labels of the dummy variables if dummy1, dummy2, etc. are not desired.

Example

```
DUMMY = TIME PREFIX=TIM;
```

FIXED command

The purpose of FIXED command is to specify the fixed effects for the model to be analyzed and is a **required** command.

Syntax

```
FIXED = <label(s)>;
```

where <label(s)> denotes a list of variable labels and interaction labels (if applicable).

Examples

```
FIXED = intcept AGE AGESQ GENDER GENDER*AGE GENDER*AGESQ;
```

Note

The label intcept is used to specify an intercept term and the symbol '*' is used to specify an interaction effect.

FIXPAT command

To specify a patterned structure for the vector of fixed parameters, the FIXPAT command may be used, with or without an additional FIXVAL command (see the FIXVAL command section). It is an **optional** command.

Syntax

```
FIXPAT = < number(s)>;
```

where <number(s)> denotes a list of positive integers separated by blank spaces.

Examples

```
FIXPAT = 0 0 3;
```

Note

The zero entries indicate fixed values specified in the corresponding FIXVAL command.

FIXVAL command

It is also possible to provide initial values for the parameters of the fixed part of the model to be analyzed. This may be achieved with the FIXVAL command, which allows you to provide starting values for these parameters. The use of the FIXVAL command and the OLS = NO option of the OPTIONS command may be particularly effective when good starting values of these parameters are available. The FIXVAL command is **optional**.

Syntax

```
FIXVAL = <number(s)>;
```

where <number(s)> denotes a list of value(s) for the parameter(s) of the fixed part of the model.

Example

```
FIXVAL = 0.151 0.355 0.654;
```

ID2 command

The ID2 command is used to indicate the variable identifying the units on the second level of the hierarchy. It is a **required** command for 2-level and 3-level models.

Syntax

```
ID2 = <label>;
```

where <label> denotes the label of the variable that identifies the units on level-2 of the model.

Example

```
ID2 = class;
```

ID3 command

The purpose of ID3 command is to specify the variable identifying the level-3 units of the model. The ID3 command is **required** for 3-level models.

Syntax

```
ID3 = <label>;
```

where <label> denotes the label of the variable that identifies the units on level-3 of the model.

Example

```
ID3 = school;
```

MISSING_DAT command

The MISSING_DAT command is used when missing values are present in the raw data file. The MISSING_DAT command allows you to specify a numeric value, which will represent a missing value on any of the variables used in the analysis. It is an **optional** command.

Syntax

```
MISSING_DAT = <number>;
```

where <number> denotes the global missing value in the raw data file.

Default

```
MISSING_DAT = -999999.0;
```

Note

All records with data values equal to the code specified in the MISSING_DAT command will subsequently be removed from the analysis.

Examples

```
MISSING_DAT = 99;  
MISSING_DAT = -998.0;  
MISSING_DAT = 0;
```

MISSING_DEP command

The purpose of the MISSING_DEP command is to specify a code assigned to missing values on the response variable(s) only. The consequence of using the MISSING_DEP command is that only records with response variable values equal to the code assigned through the MISSING_DEP command will be removed from the analysis. This command is **optional**.

Syntax

```
MISSING_DEP = <number>;
```

where <number> denotes the global missing value for the response variable(s).

Default

```
MISSING_DEP = -999999.0;
```

Notes

- The MISSING_DEP command is recommended for use in the case of a multivariate analysis. If only one of the response variables to be used in the multivariate analysis has a missing response, only that particular response will be considered missing while the remaining responses will still be used.
- All records with response variable values equal to the code specified in the MISSING_DEP command will subsequently be removed from the analysis.

Examples

```
MISSING_DEP = -999;  
MISSING_DEP = 0;
```

OPTIONS command

Each MULTILEV syntax file for a linear multilevel analysis should start 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. The OPTIONS command is a **required** command.

Syntax

```
OPTIONS <options>;
```

where <options> is a list of options each of which has the following syntax

```
<keyword> = <selection>
```

where <keyword> is one of ACM, Converge, CovBW, Deviance, Effects, Maxiter, Nfree, OLS, Output and Summary and <selection> refers to a name, a number or an option.

ACM keyword

The ACM keyword is used to request the writing of the large-sample covariance matrices of the estimators of the parameters of the fixed part and random part of the model to external text files. The corresponding standard error estimates are equal to the square roots of the diagonal elements.

The non-duplicated elements of these asymptotic covariance matrices are written to external text files with the following default names:

```
<filename>_fixed.acm  
<filename>_random.acm
```

where <filename> denotes the name of the MULTILEV syntax file.

Syntax

```
ACM = <answer>  
where <answer> is one of Yes or No.
```

Default

```
ACM = No
```

CONVERGE Keyword

A test for convergence is made at the end of each iteration of the iterative algorithm. 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. This criterion is specified by using the CONVERGE keyword.

Syntax

```
CONVERGE = <number>
```

where <number> denotes a real number greater than zero.

Default

```
CONVERGE = 0.001
```

COVBW keyword

The COVBW keyword is used to request the writing of the within-clusters and between-clusters matrices of the random effects to external text files. The non-duplicated elements of these matrices are written to external text files with the following default names:

```
<filename>_between.cov  
<filename>_within.cov
```

where <filename> denotes the name of the MULTILEV syntax file.

Syntax

COVBW = <answer>

where <answer> is one of Yes or No.

Default

COVBW = No

DEVIANCE keyword

The DEVIANCE keyword is used to provide the $-2 \log$ likelihood value as reported in a previous analysis, in order to obtain a χ^2 test statistic value 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 (see the NFREE keyword section).

Syntax

DEVIANCE = <number>

where <number> is the deviance ($-2 \log L$) value at convergence printed to the MULTILEV output file of the previous analysis.

EFFECTS keyword

The EFFECTS keyword is used to specify the estimation and printing of the indirect effects of coefficients in the fixed part of the model.

Syntax

EFFECTS = <answer>

where <answer> is one of Yes or No.

Default

EFFECTS = No

MAXITER keyword

The MAXITER keyword is used to indicate the maximum number of iterations for the iterative algorithm.

Syntax

MAXITER = <number>

where <number> denotes a positive integer.

Default

MAXITER =10

Note

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. If, on the other hand, you wish to obtain only the OLS estimates calculated in the first iteration, MAXITER may be set equal to 1.

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 value 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, which is used to provide the $-2 \log$ likelihood value as reported in the previous analysis (see the DEVIANCE keyword section).

Syntax

NFREE = <number>

where <number> is the number of free parameters, that is, the total number of parameters estimated in the previous analysis, as reported in the MULTILEV output file.

OLS keyword

OLS estimates of the fixed effects are calculated as a first step of the iterative procedure unless otherwise specified. The OLS keyword is used to indicate whether the OLS estimates are to be calculated during the first iteration or not.

Syntax

OLS = <answer>

where <answer> is one of Yes or No.

Default

OLS = Yes

Note

If starting values (see the FIXVAL command section) are provided, specify the OLS = No option.

OUTPUT keyword

The OUTPUT keyword determines the amount of output produced by MULTILEV.

Syntax

OUTPUT = <amount>

where <amount> is one of STANDARD, BAYES, RESIDUAL or ALL. Details on each of these options follow.

OUTPUT = STANDARD;

The following information is written to the default output file.

1. Input specifications as supplied by you in the MULTILEV syntax file.
2. A summary of the hierarchical structure of the raw data.
3. Details of the iterative procedure at iteration 1 and at convergence, or MAXITER if convergence was not attained. For each iteration, aside from the first iteration, these details include the estimates and the corresponding standard error estimates, z-values and exceedance probabilities.
4. The covariance and correlation matrices of the random parameters on the different levels of the model.
5. The -2 log likelihood (deviance) value at each iteration and the number of parameters estimated.
6. The CPU time used by the iterative procedure and the writing of the required results to the output file.

OUTPUT = BAYES;

If OUTPUT = BAYES is specified, the standard default results are written to the output file and the empirical Bayes estimates are written to external text file(s). The empirical Bayes estimates on levels 2 and 3 of the model are calculated and written to the files

<filename>.ba2
<filename>.ba3

where <filename> denotes the name of the MULTILEV syntax file.

OUTPUT = RESIDUAL;

If OUTPUT = RESIDUAL is specified, the standard default results are written to the output file. An additional file, <filename>.res, where <filename> denotes the name of the MULTILEV syntax file, is created, and contains the residuals as at convergence.

OUTPUT = ALL;

The standard default results are written to the MULTILEV output file, the empirical Bayes estimates are written to external text file(s) and the residuals are written to an external text file if OUTPUT = ALL is specified.

Default

OUTPUT = STANDARD;

SUMMARY keyword

The SUMMARY keyword is used to suppress or request the printout of the data summary table.

Syntax

SUMMARY = <answer>

where <answer> is one of Yes or No.

Default

SUMMARY = Yes

RANDOM1 command

The RANDOM1 command is used to identify those variables whose coefficients are allowed to vary randomly over level-1. Either RANDOM1 or RANDOM2 or both commands should be included for a 2-level model.

Syntax

```
RANDOM1 = <label(s) > ;
```

where <label(s)> denotes label(s) of the variable(s) whose coefficients are allowed to vary randomly over level-1.

Example

```
RANDOM1 = intcept ;
```

RANDOM2 command

The purpose of the RANDOM2 command is to identify those variables whose coefficients are allowed to vary randomly over level-2. Either RANDOM1 or RANDOM2 or both commands should be included for a 2-level model.

Syntax

```
RANDOM2 = <label(s) > ;
```

where <label(s)> denotes label(s) of the variable(s) whose coefficients are allowed to vary randomly over level-2.

Example

```
RANDOM2 = intcept Age Agesq;
```

RANDOM3 command

The RANDOM3 command is used to identify those variables whose coefficients are allowed to vary randomly over level-3. At least 2 of the RANDOM1, RANDOM2 and RANDOM3 commands should be included for a 3-level model.

Syntax

```
RANDOM3 = <label(s) > ;
```

where <label(s)> denotes label(s) of the variable(s) whose coefficients are allowed to vary randomly over level-3.

Examples

```
RANDOM3 = intcpt;  
RANDOM3 = X1 X2 X3 X4 ;
```

RESPONSE command

The RESPONSE command is used to specify the response variable(s) to be used in the analysis and is a **required** command.

Syntax

```
RESPONSE = <label(s)>;
```

where <label(s)> denotes a list of the label(s) of the response variable(s).

Examples

```
RESPONSE = Y1;  
RESPONSE = Math1 Math2 Math3 Eng1 Eng2 Eng3;
```

SY command

The SY command is used to specify the PSF to be processed and is a **required** command.

Syntax

```
SY = '<filename>';
```

where <filename> denotes the complete name (including drive and folder names) of the PSF.

Example

```
SY = 'C:\MLEVELEX\kanfer.psf';
```

Note

The drive and folder names may be omitted if the PSF and MULTILEV syntax file are in the same folder.

TITLE command

The TITLE command allows you to provide a description for the analysis to be performed and is an **optional** command.

Syntax

```
TITLE = <string>;
```

where <string> denotes a character string consisting of at most 60 characters.

Example

```
TITLE = Level-3 model with design weights;
```

WEIGHT1 command

The WEIGHT1 command is used to specify design weights for level-1 of the model. It is an **optional** command.

Syntax

```
WEIGHT1 = <label>;
```

where <label> denotes the case sensitive name of the level-1 weight variable.

Example

```
WEIGHT1 = SPWT;
```

WEIGHT2 command

The purpose of WEIGHT2 command is to specify design weights for level-2 of the model and it is an **optional** command.

Syntax

```
WEIGHT2 = <label>;
```

where <label> denotes the case sensitive name of the level-2 weight variable.

Example

```
WEIGHT2 = L2WT;
```

WEIGHT3 command

The WEIGHT3 command is used to specify design weights for level-3 of the model. It is an **optional** command.

Syntax

```
WEIGHT3 = <label>;
```

where <label> denotes the case sensitive name of the level-3 weight variable.

Example

```
WEIGHT3 = WTL3;
```

MULTILEV syntax files for non-linear models

The structure of the syntax file

The MULTILEV syntax file generated by the MULTILEV GUI for a non-linear model can also be created by using the text editor of LISREL for Windows or any other text editor such as Notepad or WordPad. The structure of the MULTILEV syntax file for a non-linear model follows.

```
OPTIONS <options>;
TITLE = <string>;
SY = '<filename>';
ID2 = <label>;
RESPONSE= <label>;
FIXED = <label>;
MODEL = <selection1> + <selection2>;
MISSING_DAT = <number>;
COVARIATES    b1 = <label>
               b2 = <label>
               b3 = <label>
               c1 = <label>
               c2 = <label>
               c3 = <label>;
```

where <label> denotes a case sensitive variable name used in the raw data file, <string> represents a character string, <filename> denotes a complete name (including the drive and folder names) of a file, <number> denotes a specific number, both <selection1> and <selection2> are one of Logistic, Gompertz, Monomolecular, Power or Exponential and <options> denotes a list of options for the analysis (see the OPTIONS command section).

The OPTIONS, SY, ID2, RESPONSE, FIXED and MODEL commands are **required** while the other commands are all **optional**. The OPTIONS and SY commands should be the first two commands respectively, but the other commands can be entered in any order. Except for variable labels, the contents of the syntax file are not case-sensitive. Blank lines can be inserted in any section of the syntax file.

In the following sections, the MULTILEV commands for a non-linear model are discussed separately in alphabetical order.

COVARIATES command

The COVARIATES command is used to specify covariates for the level-2 random coefficients. It is an **optional** command.

Syntax

```
COVARIATES    b1 = <label1>
               b2 = <label2>
               b3 = <label3>
               c1 = <label4>
               c2 = <label5>
               c3 = <label6>;
```

where <label1>, <label2>, ... and <label6> denote the names of the covariates.

Examples

```
COVARIATES    b1 = age
               b2 = gender;
COVARIATES    b1 = age
               c1 = gender
               c2 = income;
```

Note

The c1, c2 and c3 keywords are only available if a second component is specified in the MODEL command.

FIXED command

The purpose of FIXED command is to specify the fixed effect for the model to be analyzed and is a **required** command.

Syntax

```
FIXED = <label>;
```

where <label> denotes the name of the variable of the fixed part of the model.

Examples

```
FIXED = AGE;
```

Note

The label intcept is used to specify an intercept.

ID2 command

The ID2 command is used to indicate the variable identifying the units on the second level of the hierarchy. It is a **required** command.

Syntax

```
ID2 = <label>;
```

where <label> denotes the label of the variable that identifies the units on level-2 of the model.

Example

```
ID2 = class;
```

MISSING_DAT command

The MISSING_DAT command is used when missing values are present in the raw data file. The MISSING_DAT command allows you to specify a numeric value, which will represent a missing value on any of the variables used in the analysis. It is an **optional** command.

Syntax

```
MISSING_DAT = <number>;
```

where <number> denotes the global missing value in the raw data file.

Default

```
MISSING_DAT = -999999.0;
```

Note

All records with data values equal to the code specified in the MISSING_DAT command will subsequently be removed from the analysis.

Examples

```
MISSING_DAT = 99;  
MISSING_DAT = -998.0;  
MISSING_DAT = 0;
```

MODEL command

The purpose of the MODEL command is to specify the non-linear functions for the component(s) of the model. It is an **optional** command.

Syntax

```
MODEL = <selection1> + <selection2>;
```

where <selection1> refers to the non-linear function specified for the first component and <selection2> denotes the non-linear function specified for the second component. Both <selection1> and <selection2> are one of Logistic, Gompertz, Monomolecular, Power or Exponential.

Default

```
MODEL = Logistic;
```

Examples

```
MODEL = Power;  
MODEL = Logistic + Gompertz;
```

OPTIONS command

Each MULTILEV syntax file for a non-linear multilevel analysis should start 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. The OPTIONS command is a **required** command.

Syntax

```
OPTIONS <options>;
```

where <options> is a list of options each of which has the following syntax

```
<keyword> = <selection>
```

where <keyword> is one of METHOD, CONVERGE, MAXITER and QUADPTS and <selection> refers to a name, a number or an option.

METHOD keyword

The METHOD keyword is used to specify the estimation method as either Maximum Apriori (MAP) or Full Maximum Likelihood (ML).

Syntax

METHOD = <selection>

where <selection> is one of MAP or ML.

Default

METHOD = MAP

CONVERGE Keyword

A test for convergence is made at the end of each iteration of the iterative algorithm. 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. This criterion is specified by using the CONVERGE keyword.

Syntax

CONVERGE = <number>

where <number> denotes a real number greater than zero.

Default

CONVERGE = 0.001

MAXITER keyword

The MAXITER keyword is used to indicate the maximum number of iterations for the iterative algorithm.

Syntax

MAXITER = <number>

where <number> denotes a positive integer.

Default

MAXITER = 30

Note

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.

QUADPTS keyword

The purpose of the QUADPTS keyword is to specify the number of quadrature points to be used in the numerical integration procedure.

Syntax

QUADPTS = <number>

where <number> is the number of quadrature points.

Default

QUADPTS = 10

RESPONSE command

The RESPONSE command is used to specify the response variable to be used in the analysis and is a **required** command.

Syntax

RESPONSE = <label>;

where <label> denotes the label of the response variable.

Examples

RESPONSE = Y1;

SY command

The SY command is used to specify the PSF to be processed and is a **required** command.

Syntax

```
SY = '<filename>;'
```

where <filename> denotes the complete name (including drive and folder names) of the PSF.

Example

```
SY = 'C:\MLEVELEX\kanfer.psf';
```

Note

The drive and folder names may be omitted if the PSF and MULTILEV syntax file are in the same folder.

TITLE command

The TITLE command allows you to provide a description for the analysis to be performed and is an **optional** command.

Syntax

```
TITLE = <string>;
```

where <string> denotes a character string consisting of at most 60 characters.

Example

```
TITLE = Non-linear multilevel model for traffic data;
```

Examples

MULTILEV fits multilevel models to multilevel data emanating from complex surveys or from simple random sample surveys. This is accomplished by using the popup menus on the **Multilevel** menu of the PSF window of LISREL for Windows and the corresponding dialog boxes. This feature is illustrated by fitting models to both real and simulated data in the sections to follow.

Two-level analysis of mice data

In this section, we illustrate how MULTILEV may be used to explicitly recognize the hierarchical structure of repeated measurement data. This is accomplished by fitting the following six models to repeated weight measurements on 82 striped mice.

- A variance decomposition model with random intercepts
- A linear growth model with random intercepts
- A linear growth model with random intercepts and slopes
- A non-linear growth model with random intercepts and slopes
- A non-linear growth model with random intercepts and slopes and a covariate
- A model with complex variation at level-1 of the hierarchy

The data

The data contain repeated measurements on 82 striped mice and were obtained from the Department of Zoology at the University of Pretoria, South Africa (Du Toit 1979). A number of male and female mice were released in an outdoor enclosure with nest boxes and sufficient food and water. They were allowed to multiply freely. Occurrence of birth was recorded daily and newborn mice were weighed weekly, from the end of the second week after birth until physical maturity was reached. The data set consists of the weights of 42 male and 40 female mice. For male mice, 9 repeated weight measurements are available and for the female mice 8 repeated measurements. The data are provided as the PSF **mouse1.psf** in the **TUTORIAL** subfolder. The first portion of this PSF is shown in the following PSF window.

	mouseID	measID	weight	time	timesq	gender
1	1.0	1.0	15.0	1.0	1.0	1.0
2	1.0	2.0	17.0	2.0	4.0	1.0
3	1.0	3.0	23.0	3.0	9.0	1.0
4	1.0	4.0	24.0	4.0	16.0	1.0
5	1.0	5.0	26.0	5.0	25.0	1.0
6	1.0	6.0	31.0	6.0	36.0	1.0
7	1.0	7.0	37.0	7.0	49.0	1.0
8	1.0	8.0	42.0	8.0	64.0	1.0
9	1.0	9.0	46.0	9.0	81.0	1.0
10	2.0	1.0	11.0	1.0	1.0	1.0

The variables in the data set are:

- mouseID denotes the mouse for which the measurement was obtained.
- measID is the occasion on which the measurement for the particular mouse was made.
- weight denotes the weight (in grams) of the mouse.
- time is the time point at which the measurement was made.
- timesq is the squared value of the time point.
- gender is the gender of the mouse.

Variance decomposition model with random intercepts

The simplest multilevel model is equivalent to a one-way ANOVA with random intercepts. Although this model is not interesting in itself, it is useful as a preliminary step in a multilevel analysis as it provides important information about the outcome variability at each of the levels of the hierarchy. It may also function as a baseline model with which more sophisticated models may be compared.

Let the subscripts i and j refer to the i -th mouse and the j -th weight measurement respectively. Using this notation, the one-way ANOVA model can be expressed as

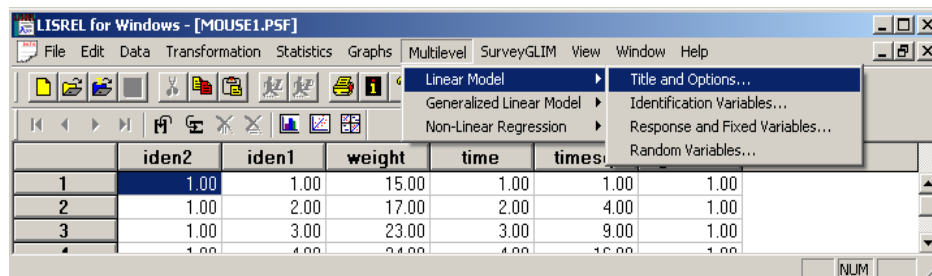
$$\text{weight}_{ij} = \beta_0 + u_{0i} + e_{ij}$$

where weight_{ij} denotes the j -th weight measurement for mouse i , β_0 denotes the intercept of the fixed part of the model, u_{0i} represents the random variation in intercepts at level-2 of the model and e_{ij} denotes the random variation at level-1 of the model. It is assumed that u_{0i} has mean 0 and variance $\Phi_{(2)}$. The variance $\Phi_{(2)}$ may be interpreted as the ‘between-group’ variability. Likewise, it is assumed that e_{ij} follows a Normal distribution with mean 0 and variance $\Phi_{(1)}$. Thus, $\Phi_{(1)}$ may be interpreted as the ‘within-group’ variability.

Fitting the variance decomposition model with random intercepts

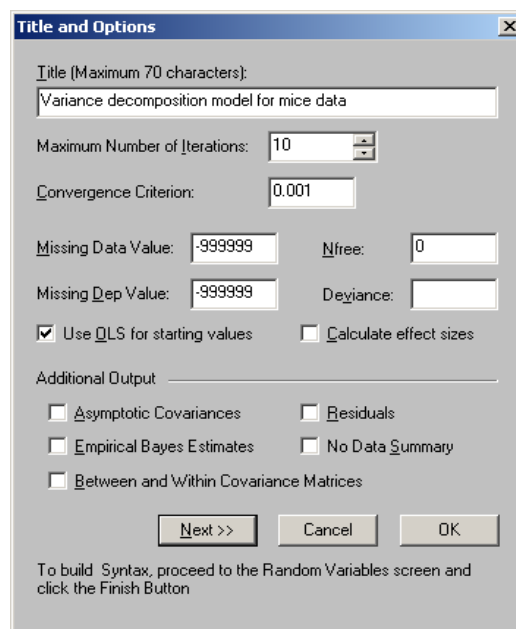
To fit the model above to the data in **mouse1.psf**, we proceed as follows.

- Select the **Open** option on the **File** menu to load the **Open** dialog box.
- Select the **PRELIS Data (*.psf)** option on the **Files of type** drop-down list box.
- Browse for the file **mouse1.psf** in the **TUTORIAL** subfolder and select it.
- Click on the **Open** button to open the PSF window for **mouse1.psf**.
- Select the **Title and Options** option from the **Linear Model** popup menu on the **Multilevel** menu as shown below



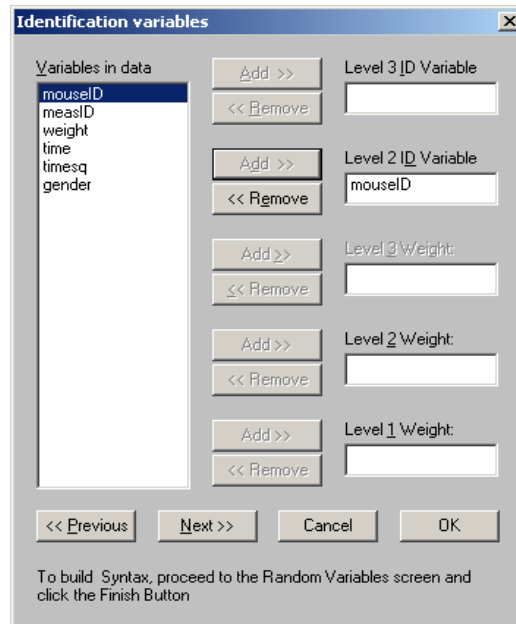
to activate the **Title and Options** dialog box.

- Enter the title Variance decomposition model for mice data in the **Title** string box to produce the following dialog box.

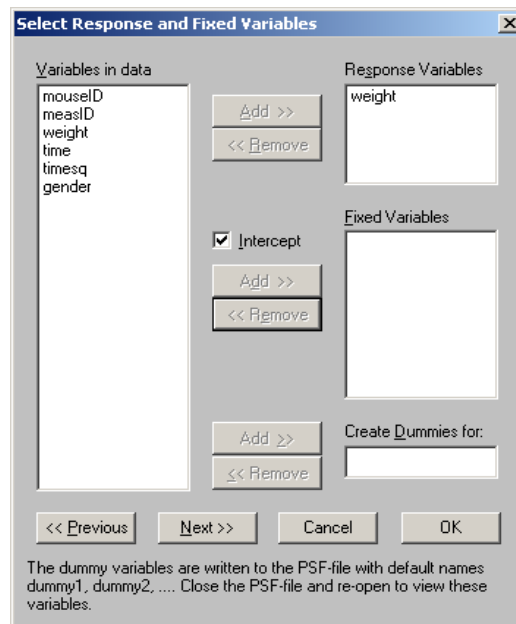


- Click on the **Next** button to go to the **Identification Variables** dialog box.

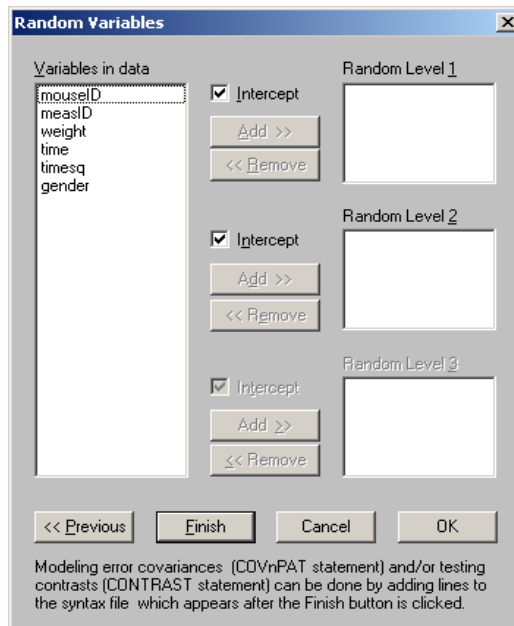
- Select the variable mouseID from the **Variables in data** list box and click on the second **Add** button to produce the following dialog box.



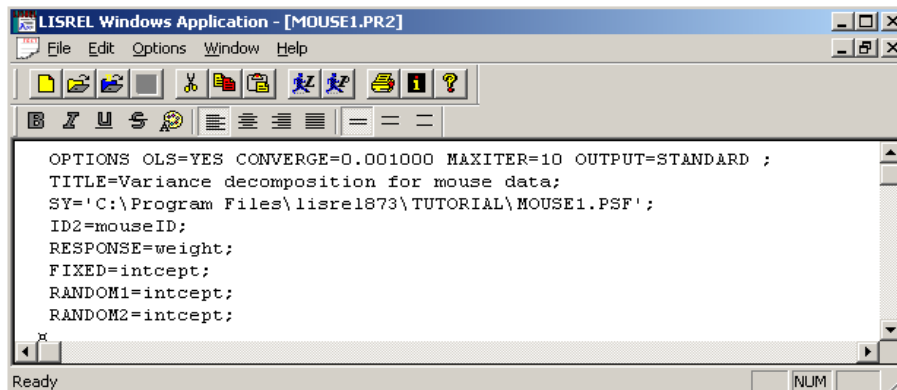
- Click on the **Next** button to go to the **Select Response and Fixed Variables** dialog box.
- Select the variable weight from the **Variables in data** list box and click on the upper **Add** button to obtain the following dialog box.



- Click on the **Next** button to go to the following **Random Variables** dialog box.



- o Click on the **Finish** button to generate the following text editor window for **mouse1.pr2**.



- o Click the **Run PRELIS** button to open the text editor window for the output file **mouse1.out**.

Discussion of results

Portions of the output file **mouse1.out** are shown below.

In the first selection of the output file, a description of the hierarchical structure of the data is provided as shown below.

```

mouse1.OUT
+-----+
| DATA SUMMARY |
+-----+

NUMBER OF LEVEL 2 UNITS :      82
NUMBER OF LEVEL 1 UNITS :     698

N2 :      1      2      3      4      5      6      7      8
N1 :      9      9      9      9      9      9      9      9

N2 :      9     10     11     12     13     14     15     16
N1 :      9      9      9      9      9      9      9      9

N2 :     17     18     19     20     21     22     23     24
N1 :      9      9      9      9      9      9      9      9

N2 :     25     26     27     28     29     30     31     32
N1 :      9      9      9      9      9      9      9      9

```

The data summary above indicates that a total of 698 observations were obtained from 82 mice. In addition, a summary of the number of measurements per mouse is provided. For example, for mouse number 1 (N2: 1), there are 9 observations (N1: 9) available.

The portion of the output file that contains the estimation results for the fixed part of the model is shown next.

```

mouse1.OUT
Variance decomposition for mouse data
ITERATION NUMBER      3

+-----+
| FIXED PART OF MODEL |
+-----+

-----
COEFFICIENTS          BETA-HAT      STD.ERR.      Z-VALUE      PR > |Z|
-----
intcept                28.63410     0.57021      50.21634     0.00000

+-----+
|  -2 LOG-LIKELIHOOD  |
+-----+

DEVIANCE= -2*LOG(LIKELIHOOD) =  5425.490015929897
NUMBER OF FREE PARAMETERS =      3

```

The results above indicate that $\hat{\beta}_0 = 28.63410$ which is statistically significant ($p < 0.000001$).

The estimation results for the random part of the model are shown below.

RANDOM PART OF MODEL				
LEVEL 2	TAU-HAT	STD.ERR.	Z-VALUE	PR > Z
intcept /intcept	11.32910	4.25185	2.66451	0.00771
LEVEL 1	TAU-HAT	STD.ERR.	Z-VALUE	PR > Z
intcept /intcept	130.32083	7.42514	17.55130	0.00000

It is evident from the results above that $\hat{\Phi}_{(2)} = 11.32910$ and $\hat{\Phi}_{(1)} = 130.32083$. Both these estimates are highly significant.

An estimate of the level-2 effect (differences between the 82 mice), for example, is obtained as

$$\frac{11.32910}{11.32910 + 130.32083} = 8.00\%$$

indicating that about 8% of the total variation in the mice weights is explained by the differences between mice.

Linear growth model with random intercepts

The variance decomposition model with random intercepts is now extended by including the variable time as a fixed effect to obtain the following linear growth model with random intercepts

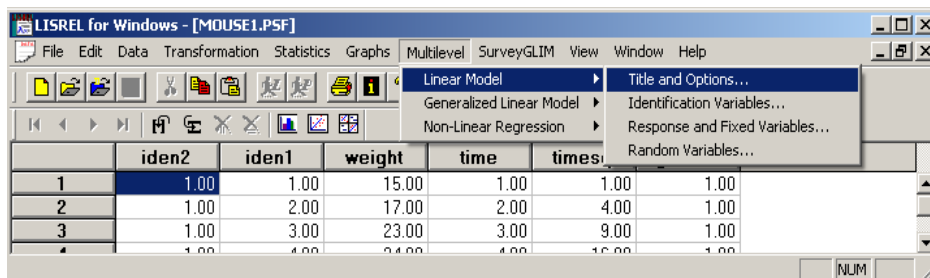
$$\text{weight}_{ij} = \beta_0 + \beta_1 \times \text{time}_{ij} + u_{0i} + e_{ij}$$

where weight_{ij} denotes the j -th weight measurement for mouse i , β_0 denotes the intercept of the fixed part of the model, β_1 denotes the regression weight for time in the fixed part of the model, time_{ij} denotes the time point for the j -th weight measurement for mouse i , u_{0i} represents the random variation in intercepts at level-2 of the model and e_{ij} denotes the random variation at level-1 of the model. It is assumed that u_{0i} has mean 0 and variance $\Phi_{(2)}$. The variance $\Phi_{(2)}$ may be interpreted as the ‘between-group’ variability. Likewise, it is assumed that e_{ij} follows a Normal distribution with mean 0 and variance $\Phi_{(1)}$. Thus, $\Phi_{(1)}$ may be interpreted as the ‘within-group’ variability.

Fitting the linear growth model with random intercepts

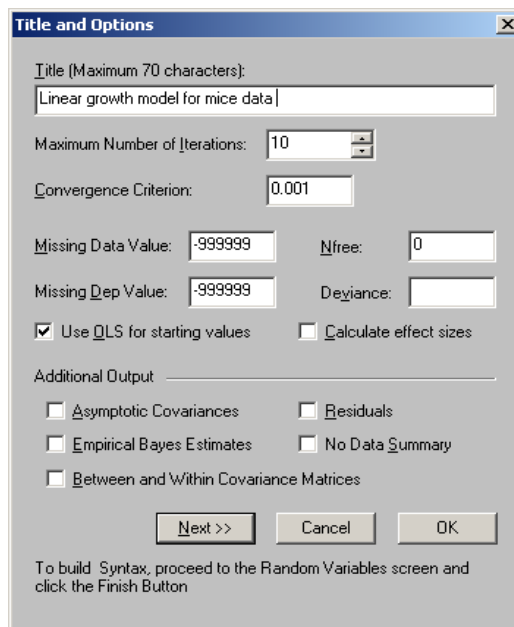
To fit the model above to the data in **mouse1.psf**, we proceed as follows.

- Select the **Open** option on the **File** menu to load the **Open** dialog box.
- Select the **PRELIS Data (*.psf)** option on the **Files of type** drop-down list box.
- Browse for the file **mouse1.psf** in the **TUTORIAL** subfolder and select it.
- Click on the **Open** button to open the PSF window for **mouse1.psf**.
- Select the **Title and Options** option from the **Linear Model** popup menu on the **Multilevel** menu as shown below



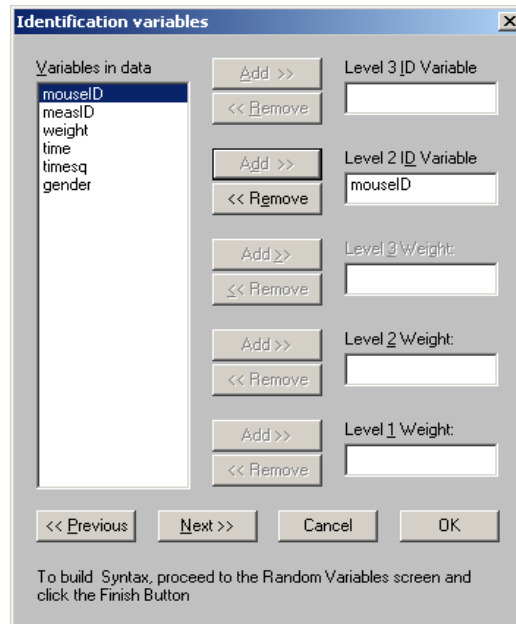
to activate the **Title and Options** dialog box.

- Enter the title **Linear growth model for mice data** in the **Title** string box to produce the following dialog box.

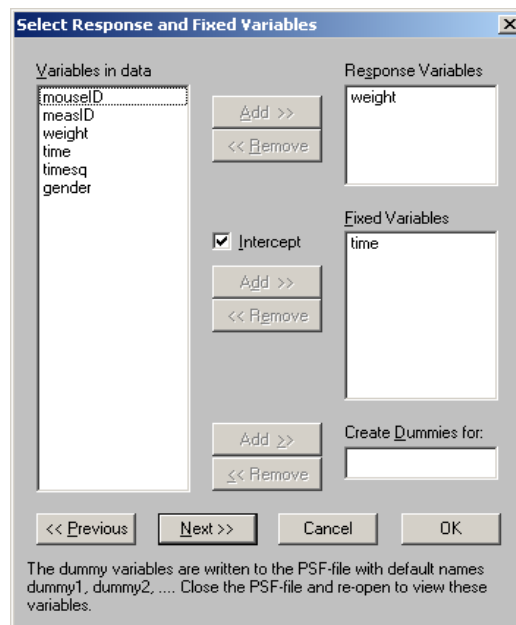


- Click on the **Next** button to go to the **Identification Variables** dialog box.

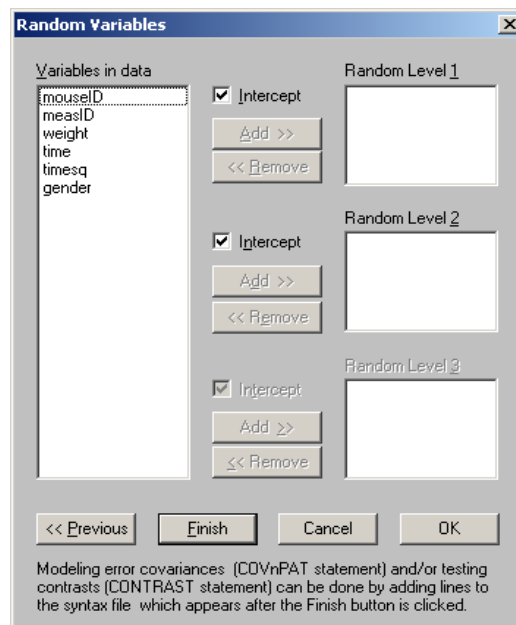
- Select the variable mouseID from the **Variables in data** list box and click on the second **Add** button to produce the following dialog box.



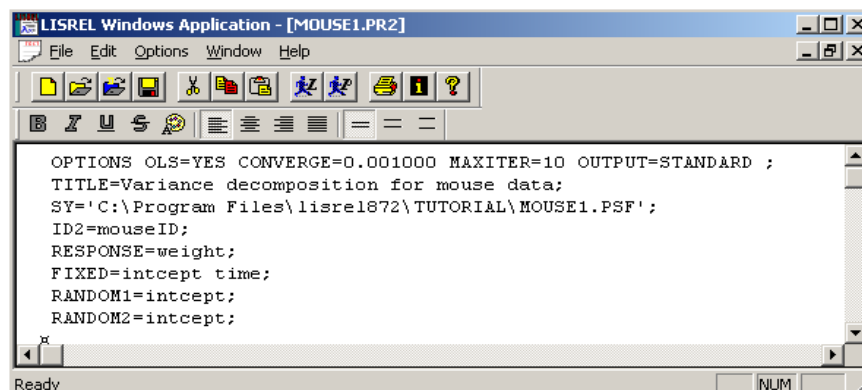
- Click on the **Next** button to go to the **Select Response and Fixed Variables** dialog box.
- Select the variable weight from the **Variables in data** list box and click on the upper **Add** button to add the variable to the **Response Variables** list box.
- Select the variable time from the **Variables in data** list box and click on the middle **Add** button to obtain the following dialog box.



- Click on the **Next** button to go to the following **Random Variables** dialog box.



- Click on the **Finish** button to generate the following text editor window for **mouse1.pr2**.



- Click the **Run PRELIS** button to open the text editor window for the output file **mouse1.out**.

Discussion of results

Portions of the output file **mouse1.out** are shown below.

In the first selection of the output file a description of the hierarchical structure of the data is provided as shown below.

```

+-----+
| DATA SUMMARY |
+-----+

NUMBER OF LEVEL 2 UNITS :      82
NUMBER OF LEVEL 1 UNITS :     698

N2 :      1      2      3      4      5      6      7      8
N1 :      9      9      9      9      9      9      9      9

N2 :      9     10     11     12     13     14     15     16
N1 :      9      9      9      9      9      9      9      9

N2 :     17     18     19     20     21     22     23     24
N1 :      0      0      0      0      0      0      0      0

```

The data summary above indicates that a total of 698 observations were obtained from 82 mice. In addition, a summary of the number of measurements per mouse is provided. For example, for mouse number 1 (N2: 1), there are 9 observations (N1: 9) available.

The portion of the output file that contains the estimation results for the fixed part of the model is shown next.

```

Linear growth model for mice data
ITERATION NUMBER      3
+-----+
| FIXED PART OF MODEL |
+-----+

-----
COEFFICIENTS          BETA-HAT      STD.ERR.      Z-VALUE      PR > |Z|
-----
intcept              9.09586      0.60387      15.06258     0.00000
time                 4.09218      0.06258      65.39108     0.00000

+-----+
|  -2 LOG-LIKELIHOOD  |
+-----+

DEVIANCE= -2*LOG(LIKELIHOOD) = 4137.578760208256
NUMBER OF FREE PARAMETERS = 4

```

The results above indicate that the weight of mice increases significantly over time. More specifically, it indicates an average gain of weight is approximately 4g for each additional time period.

The estimation results for the random part of the model are shown below.

RANDOM PART OF MODEL				
LEVEL 2	TAU-HAT	STD.ERR.	Z-VALUE	PR > Z
intcept / intcept	20.69397	3.53655	5.85146	0.00000
LEVEL 1	TAU-HAT	STD.ERR.	Z-VALUE	PR > Z
intcept / intcept	16.46288	0.93806	17.54996	0.00000

It is evident from the results above that $\hat{\Phi}_{(2)} = 20.69397$ and $\hat{\Phi}_{(1)} = 16.46288$. Both these estimates are highly significant.

An estimate of the level-2 effect (differences between the 82 mice), for example, is obtained as

$$\frac{20.69397}{20.69397 + 16.46288} = 55.69\%$$

indicating that about 55.69% of the total variation in the mice weights is explained by the differences between mice.

Linear growth model with random intercepts and slopes

It is expected that the linear growth rate may vary from mouse to mouse around its mean value instead of being fixed. The corresponding model that allows for this variation may be expressed as

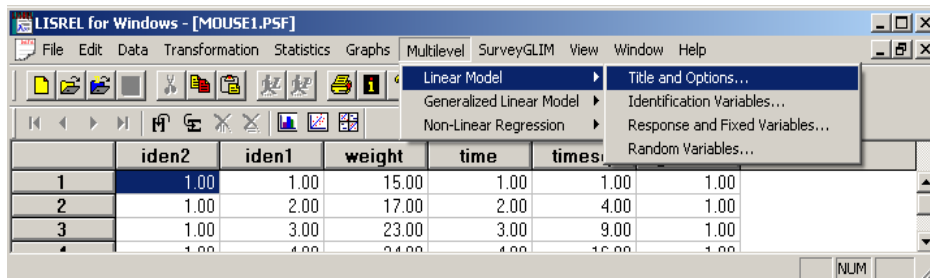
$$\text{weight}_{ij} = \beta_0 + \beta_1 \times \text{time}_{ij} + u_{0i} + u_{1i} \times \text{time}_{ij} + e_{ij}$$

where weight_{ij} denotes the j -th weight measurement for mouse i , β_0 denotes the intercept of the fixed part of the model, β_1 denotes the regression weight for time in the fixed part of the model, time_{ij} denotes the time point for the j -th weight measurement for mouse i , u_{0i} represents the random variation in intercepts at level-2 of the model, u_{1i} denotes the random variation in slopes for time at level-2 of the model and e_{ij} denotes the random variation at level-1 of the model. It is assumed that $\mathbf{u}_i = (u_{0i} \ u_{1i})'$ has mean $\mathbf{0}$ and covariance matrix $\Phi_{(2)}$ and that e_{ij} follows a Normal distribution with mean 0 and variance $\Phi_{(1)}$.

Fitting the linear growth model with random intercepts and slopes

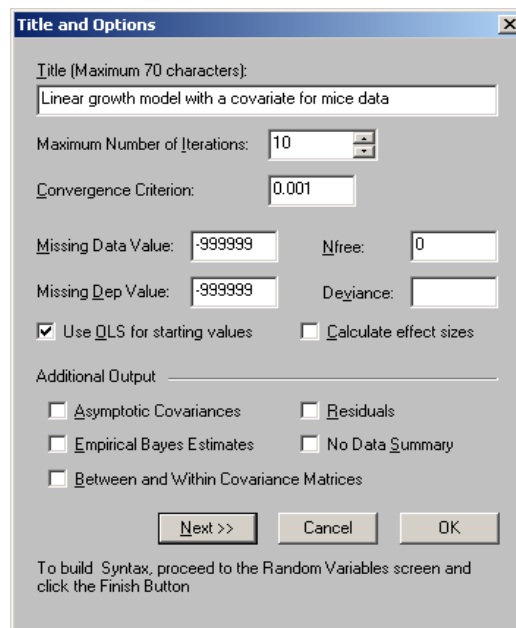
To fit the model above to the data in **mouse1.psf**, we proceed as follows.

- Select the **Open** option on the **File** menu to load the **Open** dialog box.
- Select the **PRELIS Data (*.psf)** option on the **Files of type** drop-down list box.
- Browse for the file **mouse1.psf** in the **TUTORIAL** subfolder and select it.
- Click on the **Open** button to open the PSF window for **mouse1.psf**.
- Select the **Title and Options** option from the **Linear Model** popup menu on the **Multilevel** menu as shown below



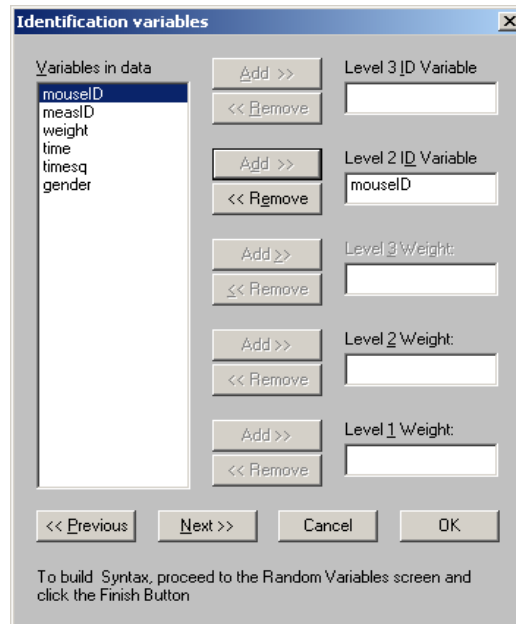
to activate the **Title and Options** dialog box.

- Enter the title Linear growth model with a covariate for mice data in the **Title** string box to produce the following dialog box.

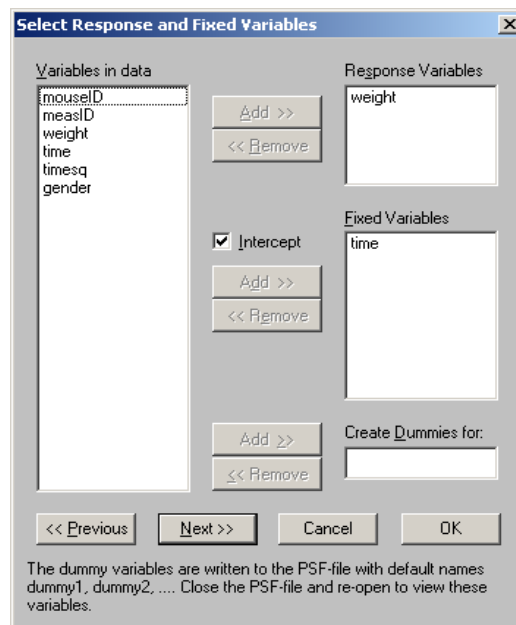


- Click on the **Next** button to go to the **Identification Variables** dialog box.

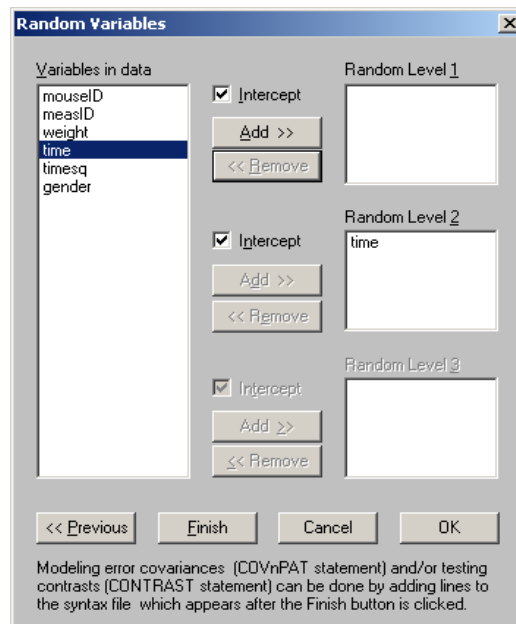
- Select the variable mouseID from the **Variables in data** list box and click on the second **Add** button to produce the following dialog box.



- Click on the **Next** button to go to the **Select Response and Fixed Variables** dialog box.
- Select the variable weight from the **Variables in data** list box and click on the upper **Add** button to add the variable to the **Response Variables** list box.
- Select the variable time from the **Variables in data** list box and click on the middle **Add** button to obtain the following dialog box.



- Click on the **Next** button to go to the **Random Variables** dialog box.
- Select the variable **time** from the **Variables in data** list box and click on the middle **Add** button to obtain the following dialog box.



- Click on the **Finish** button to generate the following text editor window for **mouse1.pr2**.

```

MOUSE1.PR2
OPTIONS OLS=YES CONVERGE=0.001000 MAXITER=10 OUTPUT=STANDARD ;
TITLE=Linear growth model with a covariate for mice data;
SY='C:\Program Files\lisrel872\TUTORIAL\MOUSE.PSF';
ID2=mouseID;
RESPONSE=weight;
FIXED=intcept time;
RANDOM1=intcept;
RANDOM2=intcept time;

```

- Click the **Run PRELIS** button to open the text editor window for the output file **mouse1.out**.

Discussion of results

Portions of the output file **mouse1.out** are shown below.

In the first selection of the output file a description of the hierarchical structure of the data is provided as shown below.

```

+-----+
| DATA SUMMARY |
+-----+

NUMBER OF LEVEL 2 UNITS :      82
NUMBER OF LEVEL 1 UNITS :     698

N2 :      1      2      3      4      5      6      7      8
N1 :      9      9      9      9      9      9      9      9

N2 :      9     10     11     12     13     14     15     16
N1 :      9      9      9      9      9      9      9      9

N2 :     17     18     19     20     21     22     23     24
N1 :      0      0      0      0      0      0      0      0

```

The data summary above indicates that a total of 698 observations were obtained from 82 mice. In addition, a summary of the number of measurements per mouse is provided. For example, for mouse number 1 (N2: 1), there are 9 observations (N1: 9) available.

The portion of the output file that contains the estimation results for the fixed part of the model is shown next.

```

Linear growth model with a covariate for mice data
ITERATION NUMBER      4

+-----+
| FIXED PART OF MODEL |
+-----+

-----
COEFFICIENTS          BETA-HAT      STD.ERR.      Z-VALUE      PR > |Z|
-----
intcept                9.20384      0.46702      19.70746     0.00000
time                   4.05978      0.12358      32.85144     0.00000

+-----+
|  -2 LOG-LIKELIHOOD  |
+-----+

DEVIANCE= -2*LOG(LIKELIHOOD) = 3873.660547828703
NUMBER OF FREE PARAMETERS =      6

```

The results above indicate that the weight of mice increases significantly over time. In particular, it indicates an average gain of weight is approximately 4g for each additional time period.

The estimation results for the random part of the model of are shown below.

RANDOM PART OF MODEL				
LEVEL 2	TAU-HAT	STD.ERR.	Z-VALUE	PR > Z
intcept /intcept	12.85320	2.80949	4.57492	0.00000
time /intcept	-1.64430	0.59236	-2.77584	0.00551
time /time	1.07389	0.19582	5.48415	0.00000
LEVEL 1	TAU-HAT	STD.ERR.	Z-VALUE	PR > Z
intcept /intcept	8.90088	0.54469	16.34126	0.00000

The estimated total variance on level-2 follows as

$$\begin{aligned}
 & \hat{\sigma}^2(u_0 + \text{time} \times u_1) \\
 &= \hat{\sigma}^2(u_0) + 2 \times \text{time} \times \hat{\sigma}(u_0, u_1) + \text{time}^2 \times \hat{\sigma}^2(u_1) \\
 &= \hat{\Phi}_{(2)1,1} + 2 \times \text{time} \times \hat{\Phi}_{(2)2,1} + \text{time}^2 \times \hat{\Phi}_{(2)2,2} \\
 &= 12.85320 - 3.2886 \times \text{time} + 1.07389 \times \text{time}^2
 \end{aligned}$$

We can thus express the total estimated variation on level-2 as a quadratic function of the variable time. When time is one, an estimate of the level-2 effect (differences between the 82 mice), for example, is obtained as

$$\frac{10.63849}{10.63849 + 8.9008} = 54.45\%$$

indicating that about 54.45% of the total variation in the mice weights is explained by the differences between mice at the first time point.

Non-linear growth model with random intercepts and slopes

In data of this nature, it is unlikely that the increase in weight measurement will be linear for all mice over the time period concerned. A non-linear component may be introduced in the model discussed in the previous section by adding a quadratic term to the model. The corresponding model is given by

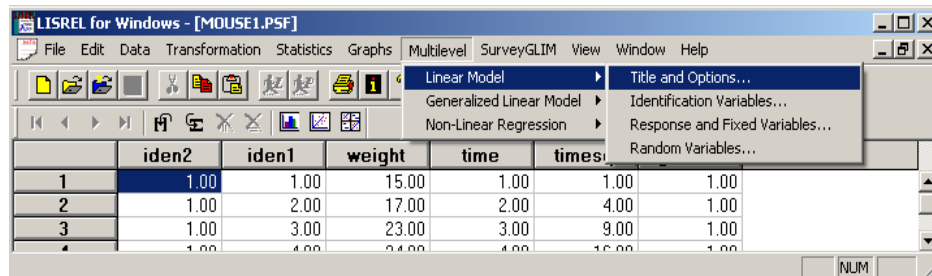
$$\text{weight}_{ij} = \beta_0 + \beta_1 \times \text{time}_{ij} + \beta_2 \times \text{timesq}_{ij} + u_{0i} + u_{1i} \times \text{time}_{ij} + u_{2i} \times \text{timesq}_{ij} + e_{ij}$$

where $weight_{ij}$ denotes the j -th weight measurement for mouse i , β_0 denotes the intercept of the fixed part of the model, β_1 denotes the regression weight for time in the fixed part of the model, $time_{ij}$ denotes the time point for the j -th weight measurement for mouse i , β_2 denotes the regression weight for timesq in the fixed part of the model, $timesq_{ij}$ denotes the squared value of time for the j -th weight measurement for mouse i , u_{0i} represents the random variation in intercepts at level-2 of the model, u_{1i} denotes the random variation in linear slopes at level-2 of the model, u_{2i} denotes the random variation in the nonlinear slopes at level-2 of the model and e_{ij} denotes the random variation at level-1 of the model. It is assumed that $\mathbf{u}_i = (u_{0i} \ u_{1i} \ u_{2i})'$ has mean $\mathbf{0}$ and covariance matrix $\Phi_{(2)}$ and that e_{ij} follows a Normal distribution with mean 0 and variance $\Phi_{(1)}$.

Fitting the non-linear growth model with random intercepts and slopes

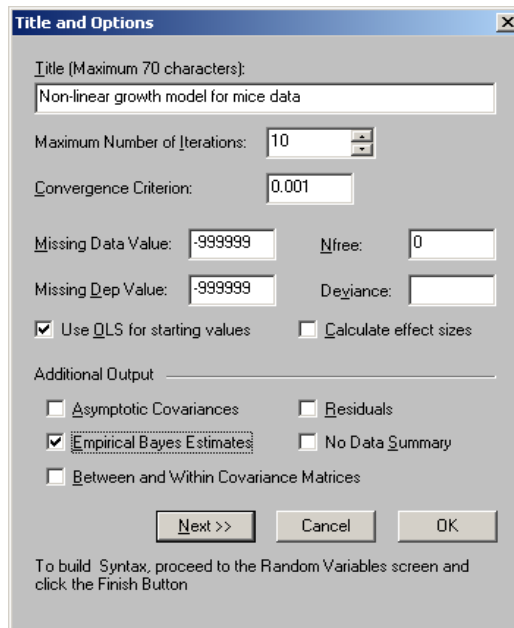
To fit model above to the data in **mouse1.psf**, we proceed as follows.

- Select the **Open** option on the **File** menu to load the **Open** dialog box.
- Select the **PRELIS Data (*.psf)** option on the **Files of type** drop-down list box.
- Browse for the file **mouse1.psf** in the **TUTORIAL** subfolder and select it.
- Click on the **Open** button to open the PSF window for **mouse1.psf**.
- Select the **Title and Options** option from the **Linear Model** popup menu on the **Multilevel** menu as shown below

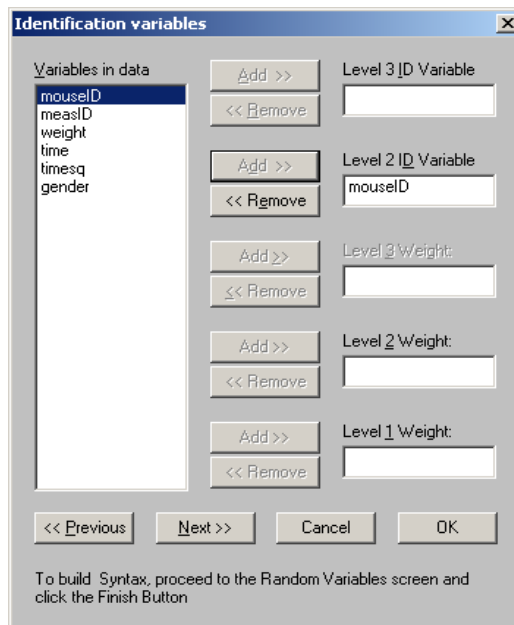


to activate the **Title and Options** dialog box.

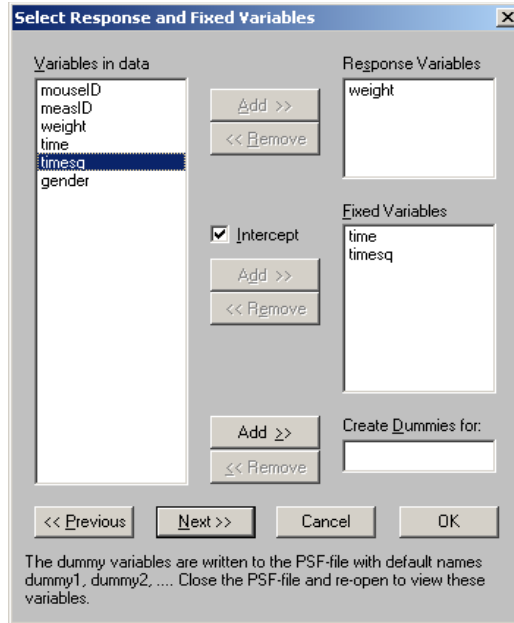
- Enter the title Non-linear growth model for mice data in the **Title** string box.
- Check the **Empirical Bayes Estimates** check box to produce the following dialog box.



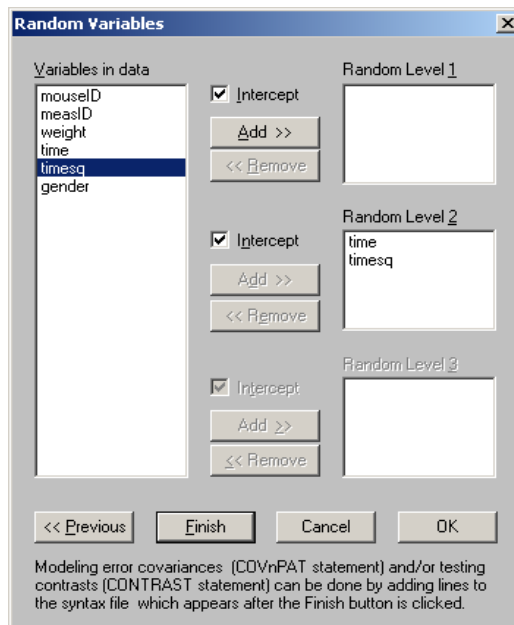
- Click on the **Next** button to go to the **Identification Variables** dialog box.
- Select the variable mouseID from the **Variables in data** list box and click on the second **Add** button to produce the following dialog box.



- Click on the **Next** button to go to the **Select Response and Fixed Variables** dialog box.
- Select the variable weight from the **Variables in data** list box and click on the upper **Add** button to add the variable to the **Response Variables** list box.
- Select the variables time and timesq from the **Variables in data** list box and click on the middle **Add** button to obtain the following dialog box.



- Click on the **Next** button to go to the **Random Variables** dialog box.
- Select the variables **time** and **timesq** from the **Variables in data** list box and click on the middle **Add** button to obtain the following dialog box.



- Click on the **Finish** button to generate the following text editor window for **mouse1.pr2**.

```

MOUSE1.PR2
OPTIONS OLS=YES CONVERGE=0.001000 MAXITER=10 OUTPUT=BAYES ;
TITLE=Non-linear growth model for mice data ;
SY='C:\Program Files\lisrel872\TUTORIAL\MOUSE1.PSF';
ID2=mouseID;
RESPONSE=weight;
FIXED=intcept time timesq;
RANDOM1=intcept;
RANDOM2=intcept time timesq;

```

- Click the **Run PRELIS** button to open the text editor window for the output file **mouse1.out**.

Discussion of results

Portions of the output file **mouse1.out** are shown below.

In the first selection of the output file a description of the hierarchical structure of the data is provided as shown below.

```

MOUSE1.OUT
+-----+
| DATA SUMMARY |
+-----+
NUMBER OF LEVEL 2 UNITS :      82
NUMBER OF LEVEL 1 UNITS :     698

N2 :      1      2      3      4      5      6      7      8
N1 :      9      9      9      9      9      9      9      9

N2 :      9     10     11     12     13     14     15     16
N1 :      9      9      9      9      9      9      9      9

N2 :     17     18     19     20     21     22     23     24
N1 :      0      0      0      0      0      0      0      0

```

The data summary above indicates that a total of 698 observations were obtained from 82 mice. In addition, a summary of the number of measurements per mouse is provided. For example, for mouse number 1 (N2: 1), there are 9 observations (N1: 9) available.

The portion of the output file that contains the estimation results for the fixed part of the model is shown next.

```

MOUSE1.OUT
Non-linear growth model for mice data
ITERATION NUMBER      5
+-----+
| FIXED PART OF MODEL |
+-----+
-----
COEFFICIENTS          BETA-HAT      STD.ERR.      Z-VALUE      PR > |Z|
-----
intcept               4.16213      0.45748      9.09788      0.00000
time                  6.90560      0.30980      22.29056     0.00000
timesq                -0.29629     0.02960     -10.01009    0.00000
+-----+
| -2 LOG-LIKELIHOOD  |
+-----+
DEVIANCE= -2*LOG(LIKELIHOOD) = 3400.932487897844
NUMBER OF FREE PARAMETERS = 10

```

The results above indicate that the weight of mice increases significantly over time. In addition, this increase is quadratic.

The estimation results for the random part of the model of are shown below.

```

MOUSE1.OUT
+-----+
| RANDOM PART OF MODEL |
+-----+
-----
LEVEL 2              TAU-HAT      STD.ERR.      Z-VALUE      PR > |Z|
-----
intcept /intcept     11.72906     2.70296      4.33933     0.00001
time /intcept        -5.86552     1.58228     -3.70699    0.00021
time /time           6.59372     1.23130      5.35509    0.00000
timesq /intcept      0.38927     0.14067      2.76733    0.00565
timesq /time         -0.55909     0.11275     -4.95880    0.00000
timesq /timesq       0.05814     0.01124      5.17424    0.00000
-----
LEVEL 1              TAU-HAT      STD.ERR.      Z-VALUE      PR > |Z|
-----
intcept /intcept     3.07863     0.20465     15.04328    0.00000

```

The total estimated variance on level-2 follows as

$$\begin{aligned}
& \hat{\sigma}^2(u_0 + \text{time} \times u_1 + \text{time}^2 \times u_2) \\
&= \hat{\sigma}^2(u_0) + \text{time}^2 \times \hat{\sigma}^2(u_1) + \text{time}^4 \times \hat{\sigma}^2(u_2) + 2 \times \text{time} \times \hat{\sigma}(u_0, u_1) \\
&+ 2 \times \text{time}^2 \times \hat{\sigma}(u_0, u_2) + 2 \times \text{time}^3 \times \hat{\sigma}(u_1, u_2) \\
&= \hat{\Phi}_{(2)1,1} + \text{time}^2 \times \hat{\Phi}_{(2)2,2} + \text{time}^4 \times \hat{\Phi}_{(2)3,3} + 2 \times \text{time} \times \hat{\Phi}_{(2)2,1} \\
&+ 2 \times \text{time}^2 \times \hat{\Phi}_{(2)3,1} + 2 \times \text{time}^3 \times \hat{\Phi}_{(2)3,2} \\
&= 11.72906 + 6.59372 \times \text{time}^2 + 0.05814 \times \text{time}^4 - 11.73104 \times \text{time} \\
&+ 0.77854 \times \text{time}^2 - 1.11818 \times \text{time}^3
\end{aligned}$$

We can thus express the total estimated variation on level-2 as a nonlinear function of the variable time.

Non-linear growth model with random intercepts slopes and a covariate

In this example, we want to determine whether there is a significant difference between the growth pattern of the male and female mice, as modeled in the non-linear growth model discussed previously. This can be determined by adding the gender of the mice as covariate to the model fitted in the previous section. The corresponding model is given by

$$\begin{aligned}
\text{weight}_{ij} = & \beta_0 + \beta_1 \times \text{time}_{ij} + \beta_2 \times \text{timesq}_{ij} + \beta_3 \times \text{gender}_{ij} \\
& + \beta_4 \times \text{gender}_{ij} \times \text{time}_{ij} + \beta_5 \times \text{gender}_{ij} \times \text{time}_{ij} \\
& + u_{0i} + u_{1i} \times \text{time}_{ij} + u_{2i} \times \text{timesq}_{ij} + e_{ij}
\end{aligned}$$

where weight_{ij} denotes the j -th weight measurement for mouse i , β_0 denotes the intercept of the fixed part of the model, β_1 denotes the regression weight for time in the fixed part of the model, time_{ij} denotes the time point for the j -th weight measurement for mouse i , β_2 denotes the regression weight for timesq in the fixed part of the model, timesq_{ij} denotes the squared value of time for the j -th weight measurement for mouse i , β_3 denotes the regression weight for gender in the fixed part of the model, gender_{ij} represents gender for mouse i , β_4 denotes the regression weight for the interaction between gender and time in the fixed part of the model, $\text{gender}_{ij} \times \text{time}_{ij}$ denotes the product of gender and time for the j -th weight measurement for mouse i , u_{0i} represents the random variation in intercepts at level-2 of the model, u_{1i} denotes the random variation in linear slopes at level-2 of the model, u_{2i} denotes the random variation in the nonlinear slopes at level-2 of the model and e_{ij} denotes the random variation at level-1 of the model. It is assumed that

$\mathbf{u}_i = (u_{0i} \ u_{1i} \ u_{2i})'$ has mean $\mathbf{0}$ and covariance matrix $\Phi_{(2)}$ and that e_{ij} follows a Normal distribution with mean 0 and variance $\Phi_{(1)}$.

The fact that the variable gender is a level-2 variable, thus having the same value for all level-1 observations nested within each level-2 unit, is perhaps best illustrated by rewriting the model as:

$$\text{weight}_{ij} = b_{0i} + b_{1i} \times \text{time}_{ij} + b_{2i} \times \text{timesq}_{ij} + e_{ij}$$

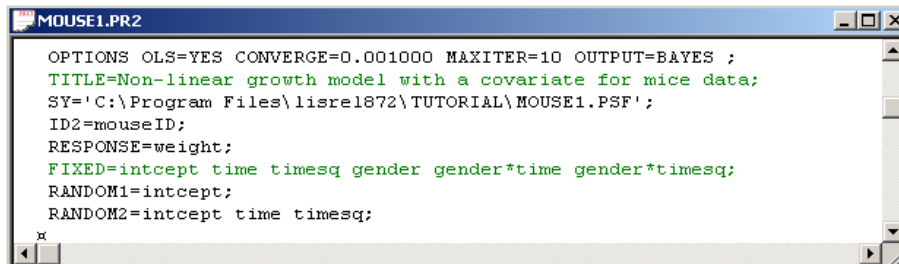
where

$$\begin{aligned} b_{0i} &= \beta_0 + \beta_3 \times \text{gender}_{ij} + u_{0i} \\ b_{1i} &= \beta_1 + \beta_4 \times \text{gender}_{ij} + u_{1i} \\ b_{2i} &= \beta_2 + \beta_5 \times \text{gender}_{ij} + u_{2i} \end{aligned}$$

Fitting the non-linear growth model with random intercepts slopes and a covariate

To fit the model shown above to the data in **mouse1.psf**, the MULTILEV syntax file generated in the previous section needs to be modified. This is accomplished as follows.

- Select the **Open** option on the **File** menu to load the **Open** dialog box.
- Browse for the file **mouse1.pr2**, which is generated in the previous example, in the **TUTORIAL** subfolder and select it.
- Click on the **Open** button to open the text editor window for **mouse1.pr2**.
- Change the TITLE command to TITLE=Non-linear growth model with a covariate for mice data;
- Modify the FIXED command to FIXED=intcept time timesq gender gender*time gender*timesq; to generate the following MULTILEV syntax file.



```

MOUSE1.PR2
OPTIONS OLS=YES CONVERGE=0.001000 MAXITER=10 OUTPUT=BAYES ;
TITLE=Non-linear growth model with a covariate for mice data;
SY='C:\Program Files\lisrel872\TUTORIAL\MOUSE1.PSF';
ID2=mouseID;
RESPONSE=weight;
FIXED=intcept time timesq gender gender*time gender*timesq;
RANDOM1=intcept;
RANDOM2=intcept time timesq;

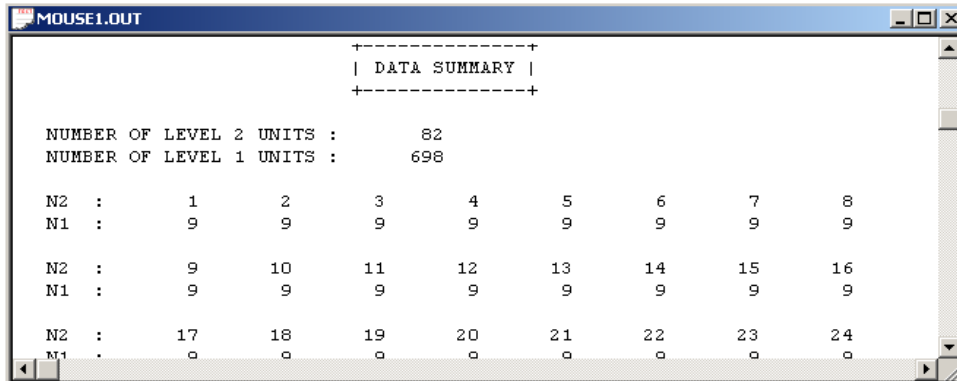
```

- Click the **Run PRELIS** button to open the text editor window for the output file **mouse1.out**.

Discussion of results

Portions of the output file **mouse1.out** are shown below.

In the first selection of the output file a description of the hierarchical structure of the data is provided as shown below.



```
MOUSE1.OUT
+-----+
| DATA SUMMARY |
+-----+
NUMBER OF LEVEL 2 UNITS :      82
NUMBER OF LEVEL 1 UNITS :     698

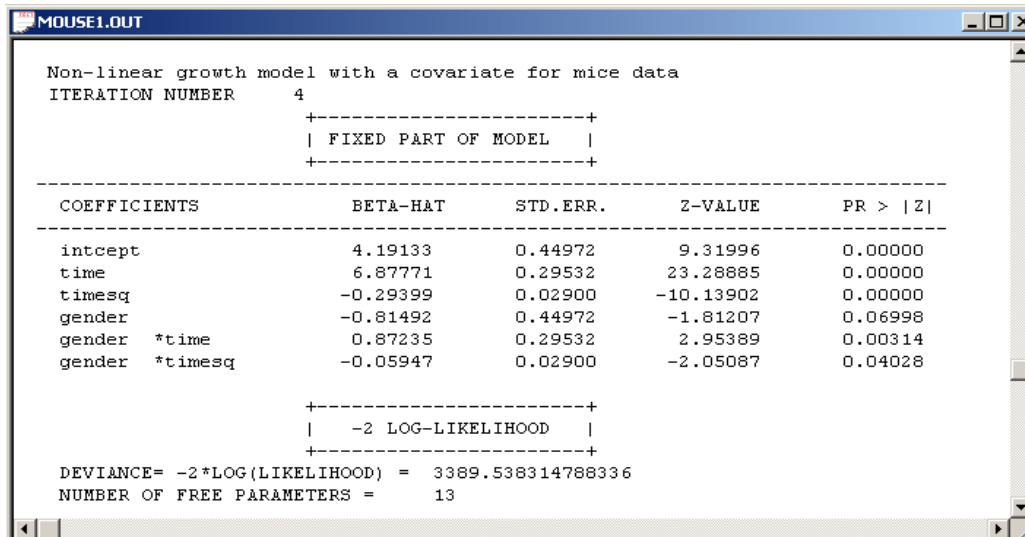
N2 :      1      2      3      4      5      6      7      8
N1 :      9      9      9      9      9      9      9      9

N2 :      9     10     11     12     13     14     15     16
N1 :      9      9      9      9      9      9      9      9

N2 :     17     18     19     20     21     22     23     24
N1 :      0      0      0      0      0      0      0      0
```

The data summary above indicates that a total of 698 observations were obtained from 82 mice. In addition, a summary of the number of measurements per mouse is provided. For example, for mouse number 1 (N2: 1), there are 9 observations (N1: 9) available.

The portion of the file that contains the estimation results for the fixed part of the model is shown next.



```
MOUSE1.OUT
Non-linear growth model with a covariate for mice data
ITERATION NUMBER      4
+-----+
| FIXED PART OF MODEL |
+-----+
-----
COEFFICIENTS          BETA-HAT      STD.ERR.      Z-VALUE      PR > |Z|
-----
intcept              4.19133      0.44972      9.31996      0.00000
time                 6.87771      0.29532     23.28885     0.00000
timesq              -0.29399      0.02900     -10.13902    0.00000
gender              -0.81492      0.44972     -1.81207     0.06998
gender *time         0.87235      0.29532      2.95389     0.00314
gender *timesq      -0.05947      0.02900     -2.05087     0.04028

+-----+
| -2 LOG-LIKELIHOOD |
+-----+
DEVIANCE= -2*LOG(LIKELIHOOD) = 3389.538314788336
NUMBER OF FREE PARAMETERS = 13
```

The results above indicate that the interaction effects between gender and time and gender and time squared are both statistically significant if we use a 5% level of significance. In other words, gender exerts a significant influence on the linear and non-linear slopes for time .

The estimation results for the random part of the model of are shown below.

RANDOM PART OF MODEL				
LEVEL 2	TAU-HAT	STD. ERR.	Z-VALUE	PR > Z
intcept /intcept	11.07339	2.60146	4.25660	0.00002
time /intcept	-5.16047	1.46703	-3.51763	0.00044
time /time	5.83687	1.11334	5.24265	0.00000
timesq /intcept	0.34138	0.13349	2.55735	0.01055
timesq /time	-0.50761	0.10447	-4.85897	0.00000
timesq /timesq	0.05464	0.01069	5.11098	0.00000
LEVEL 1	TAU-HAT	STD. ERR.	Z-VALUE	PR > Z
intcept /intcept	3.07844	0.20464	15.04338	0.00000

The estimated total variance on level-2 follows as

$$\begin{aligned}
 & \hat{\sigma}^2(u_0 + \text{time} \times u_1 + \text{time}^2 \times u_2) \\
 &= \hat{\sigma}^2(u_0) + \text{time}^2 \times \hat{\sigma}^2(u_1) + \text{time}^4 \times \hat{\sigma}^2(u_2) + 2 \times \text{time} \times \hat{\sigma}(u_0, u_1) \\
 & \quad + 2 \times \text{time}^2 \times \hat{\sigma}(u_0, u_2) + 2 \times \text{time}^3 \times \hat{\sigma}(u_1, u_2) \\
 &= \hat{\Phi}_{(2)1,1} + \text{time}^2 \times \hat{\Phi}_{(2)2,2} + \text{time}^4 \times \hat{\Phi}_{(2)3,3} + 2 \times \text{time} \times \hat{\Phi}_{(2)2,1} \\
 & \quad + 2 \times \text{time}^2 \times \hat{\Phi}_{(2)3,1} + 2 \times \text{time}^3 \times \hat{\Phi}_{(2)3,2} \\
 &= 11.07339 + 5.83687 \times \text{time}^2 + 0.05464 \times \text{time}^4 - 10.32094 \times \text{time} \\
 & \quad + 0.68272 \times \text{time}^2 - 1.01522 \times \text{time}^3
 \end{aligned}$$

We can thus write the estimated total variation on level-2 as a function of the variable time.

A model with complex variation at level-1 of the model of the hierarchy

In the final example of a level-2 model, the nonlinear growth model is extended to include complex variation on both levels of the hierarchy. The term “complex variation” refers to the existence of two or more random variables at the same level of the hierarchy. We include the variable time in this model to illustrate such a model. The corresponding model may be expressed as

$$\text{weight}_{ij} = \beta_0 + \beta_1 \times \text{time}_{ij} + \beta_2 \times \text{timesq}_{ij} + u_{0i} + u_{1i} \times \text{time}_{ij} + u_{2i} \times \text{timesq}_{ij} + e_{1ij} + e_{2ij} \times \text{time}_{ij}$$

where weight_{ij} denotes the j -th weight measurement for mouse i , β_0 denotes the intercept of the fixed part of the model, β_1 denotes the regression weight for time in the fixed part of the model, time_{ij} denotes the time point for the j -th weight measurement for mouse i , β_2 denotes the regression weight for timesq in the fixed part of the model, timesq_{ij} denotes the squared value of time for the j -th weight measurement for mouse i , u_{0i} represents the random variation in intercepts at level-2 of the model, u_{1i} denotes the random variation in linear slopes at level-2 of the model, u_{2i} denotes the random variation in the nonlinear slopes at level-2 of the model, e_{1ij} denotes the random intercept variation at level-1 of the model and e_{2ij} denotes the random slope variation at level-1 of the model. It is assumed that $\mathbf{u}_i = (u_{0i} \ u_{1i} \ u_{2i})'$ has mean $\mathbf{0}$ and covariance matrix $\Phi_{(2)}$ and that $\mathbf{e}_{ij} = (e_{1ij} \ e_{2ij})'$ follows a Normal distribution with mean $\mathbf{0}$ and covariance matrix $\Phi_{(1)}$.

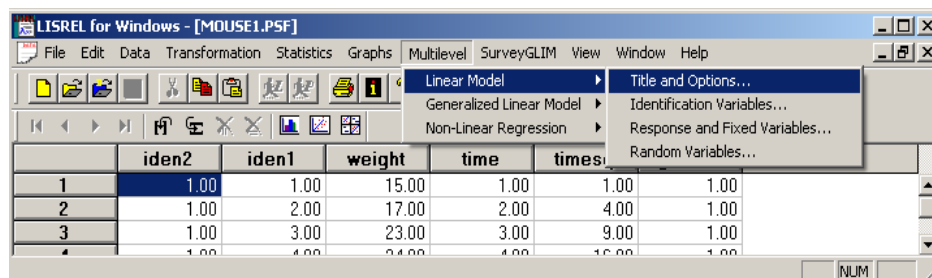
This change in the level-1 covariance structure implies that the estimated total variation at this level of the model can now be expressed as:

$$\begin{aligned} & \hat{\sigma}^2(e_{1ij} + \text{time} \times e_{2ij}) \\ &= \hat{\Phi}_{(1)1,1} + 2 \times \text{time} \times \hat{\Phi}_{(1)2,1} + \text{time}^2 \times \hat{\Phi}_{(1)2,2} \end{aligned}$$

Fitting the model with complex variation at level-1 of the hierarchy

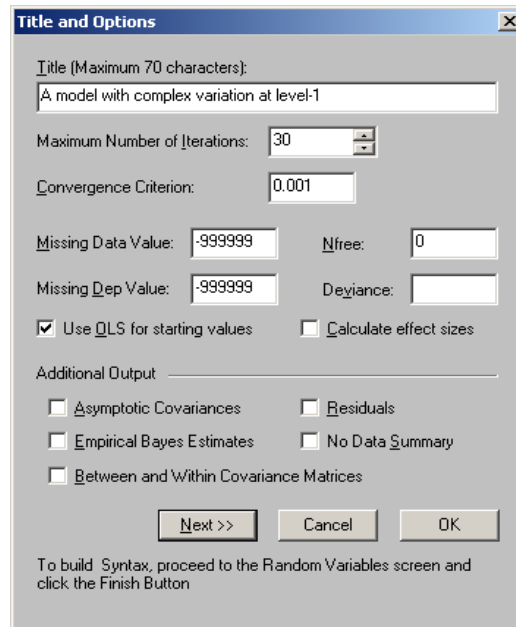
To fit the model above to the data in **mouse1.psf**, we proceed as follows.

- Select the **Open** option on the **File** menu to load the **Open** dialog box.
- Select the **PRELIS Data (*.psf)** option on the **Files of type** drop-down list box.
- Browse for the file **mouse1.psf** in the **TUTORIAL** subfolder and select it.
- Click on the **Open** button to open the PSF window for **mouse1.psf**.
- Select the **Title and Options** option from the **Linear Model** popup menu on the **Multilevel** menu as shown below

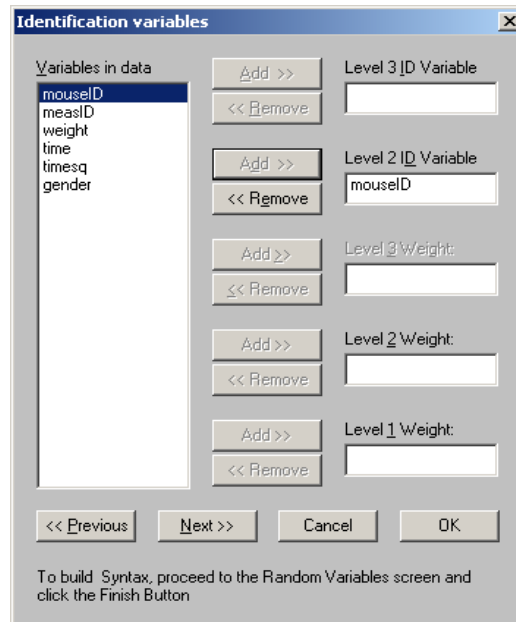


to activate the **Title and Options** dialog box.

- Enter the title A model with complex variation at level-1 in the **Title** string box.
- Change the **Maximum Number of Iterations** to 30 to obtain the following **Title and Options** dialog box.

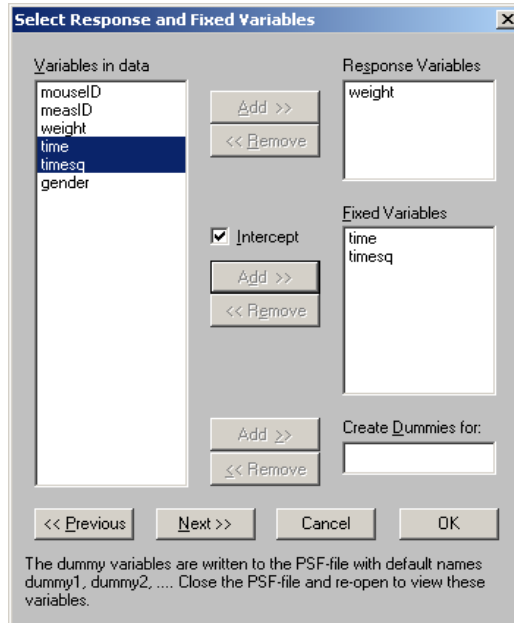


- Click on the **Next** button to go to the **Identification Variables** dialog box.
- Select the variable mouseID from the **Variables in data** list box and click on the second **Add** button to produce the following dialog box.

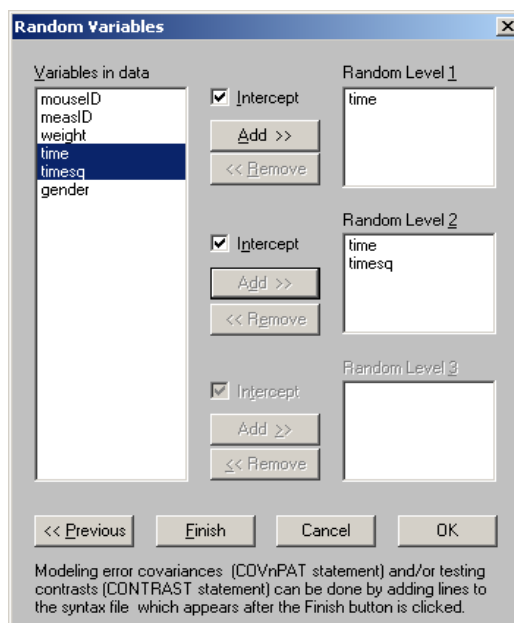


- Click on the **Next** button to go to the **Select Response and Fixed Variables** dialog box.

- Select the variable weight from the **Variables in data** list box and click on the upper **Add** button to add the variable to the **Response Variables** list box.
- Select the variables time and timesq from the **Variables in data** list box and click on the middle **Add** button to obtain the following dialog box.



- Click on the **Next** button to go to the **Random Variables** dialog box.
- Select the variable time from the **Variables in data** list box and click on the upper **Add** button to add the variable to the **Random Level 1** variable list box.
- Select the variables time and timesq from the **Variables in data** list box and click on the middle **Add** button to obtain the following dialog box.



- Click on the **Finish** button to generate the following text editor window for **mouse1.pr2**.

```

MOUSE1.PR2
OPTIONS OLS=YES CONVERGE=0.001000 MAXITER=30 OUTPUT=STANDARD ;
TITLE=A model with complex variation at level-1;
SY='C:\Program Files\lisrel872\TUTORIAL\MOUSE1.PSF';
ID2=mouseID;
RESPONSE=weight;
FIXED=intcept time timesq;
RANDOM1=intcept time;
RANDOM2=intcept time timesq;
x

```

- Click the **Run PRELIS** button to open the text editor window for the output file **mouse1.out**.

Discussion of results

Portions of the output file **mouse1.out** are shown below.

In the first selection of the output file a description of the hierarchical structure of the data is provided as shown below.

```

MOUSE1.OUT
+-----+
| DATA SUMMARY |
+-----+
NUMBER OF LEVEL 2 UNITS :      82
NUMBER OF LEVEL 1 UNITS :     698

N2 :      1      2      3      4      5      6      7      8
N1 :      9      9      9      9      9      9      9      9

N2 :      9     10     11     12     13     14     15     16
N1 :      9      9      9      9      9      9      9      9

N2 :     17     18     19     20     21     22     23     24
N1 :      0      0      0      0      0      0      0      0

```

The data summary above indicates that a total of 698 observations were obtained from 82 mice. In addition, a summary of the number of measurements per mouse is provided. For example, for mouse number 1 (N2: 1), there are 9 observations (N1: 9) available.

The portion of the file that contains the estimation results for the fixed part of the model is shown next.

MOUSE1.OUT

A model with complex variation at level-1
 ITERATION NUMBER 16

-----+
 | FIXED PART OF MODEL |
 -----+

COEFFICIENTS	BETA-HAT	STD.ERR.	Z-VALUE	PR > Z
intcept	4.67447	0.41483	11.26827	0.00000
time	6.58601	0.32377	20.34167	0.00000
timesq	-0.26114	0.03616	-7.22095	0.00000

-----+
 | -2 LOG-LIKELIHOOD |
 -----+

DEVIANCE= -2*LOG(LIKELIHOOD) = 3714.931628448757
 NUMBER OF FREE PARAMETERS = 12

The results above indicate that the weight of mice increases significantly over time and that this increase is quadratic in nature.

The estimation results for the random part of the model of are shown below.

MOUSE1.OUT

-----+
 | RANDOM PART OF MODEL |
 -----+

LEVEL 2	TAU-HAT	STD.ERR.	Z-VALUE	PR > Z
intcept /intcept	8.78969	2.78993	3.15050	0.00163
time /intcept	-4.31318	1.67904	-2.56883	0.01020
time /time	5.92342	1.39225	4.25458	0.00002
timesq /intcept	0.31565	0.17401	1.81396	0.06968
timesq /time	-0.55258	0.14830	-3.72595	0.00019
timesq /timesq	0.06508	0.01722	3.78032	0.00016

LEVEL 1	TAU-HAT	STD.ERR.	Z-VALUE	PR > Z
intcept /intcept	-0.30329	1.67844	-0.18070	0.85661
time /intcept	1.01779	0.59775	1.70269	0.08863
time /time	-0.22900	0.16669	-1.37382	0.16950

The estimated total variance on level-2 follows as

$$\begin{aligned}
& \hat{\sigma}^2(u_0 + \text{time} \times u_1 + \text{time}^2 \times u_2) \\
&= \hat{\sigma}^2(u_0) + \text{time}^2 \times \hat{\sigma}^2(u_1) + \text{time}^4 \times \hat{\sigma}^2(u_2) + 2 \times \text{time} \times \hat{\sigma}(u_0, u_1) \\
& \quad + 2 \times \text{time}^2 \times \hat{\sigma}(u_0, u_2) + 2 \times \text{time}^3 \times \hat{\sigma}(u_1, u_2) \\
&= \hat{\Phi}_{(2)1,1} + \text{time}^2 \times \hat{\Phi}_{(2)2,2} + \text{time}^4 \times \hat{\Phi}_{(2)3,3} + 2 \times \text{time} \times \hat{\Phi}_{(2)2,1} \\
& \quad + 2 \times \text{time}^2 \times \hat{\Phi}_{(2)3,1} + 2 \times \text{time}^3 \times \hat{\Phi}_{(2)3,2} \\
&= 8.78969 + 5.92342 \times \text{time}^2 + 0.06508 \times \text{time}^4 - 8.62636 \times \text{time} \\
& \quad + 0.6313 \times \text{time}^2 - 1.10516 \times \text{time}^3
\end{aligned}$$

We can thus express the estimated total variation on level-2 as a nonlinear function of the variable time. Similarly, the estimated level-1 variation can be expressed as a quadratic function of time as well.

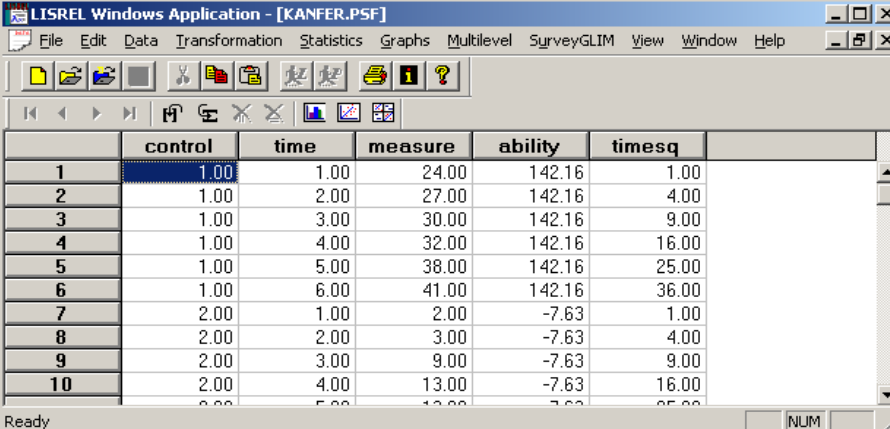
Two-level analysis of air traffic control data

The data

The data used in this example are described by Kanfer and Ackerman (1989). The data consist of information for 141 U.S. Air Force enlisted personnel. The personnel carried out a computerized air traffic controller task developed by Kanfer and Ackerman (1989).

The subjects were instructed to accept planes into their hold pattern and land them safely and efficiently on one of four runways, varying in length and compass directions, according to rules governing plane movements and landing requirements. For each subject, the success of a series of between three and six 10-minute trials was recorded. The measurement employed was the number of correct landings per trial. The Armed Services Vocational Battery (ASVB) was also administered to each subject. A global measure of cognitive ability, obtained from the sum of scores on 10 subscales, is included in the data.

The data are provided as the PSF **kanfer.psf** in the **TUTORIAL** subfolder. The first portion of this PSF is shown in the following PSF window.



The screenshot shows a window titled "LISREL Windows Application - [KANFER.PSF]". The window contains a menu bar (File, Edit, Data, Transformation, Statistics, Graphs, Multilevel, SurveyGLIM, View, Window, Help) and a toolbar with various icons. Below the toolbar is a data table with the following columns: control, time, measure, ability, timesq. The data is as follows:

	control	time	measure	ability	timesq
1	1.00	1.00	24.00	142.16	1.00
2	1.00	2.00	27.00	142.16	4.00
3	1.00	3.00	30.00	142.16	9.00
4	1.00	4.00	32.00	142.16	16.00
5	1.00	5.00	38.00	142.16	25.00
6	1.00	6.00	41.00	142.16	36.00
7	2.00	1.00	2.00	-7.63	1.00
8	2.00	2.00	3.00	-7.63	4.00
9	2.00	3.00	9.00	-7.63	9.00
10	2.00	4.00	13.00	-7.63	16.00

The variables in the data set are:

- control is the identifying number of the air traffic controller.
- time is the number of the trial (ranging from 1 to 6).
- measure indicates the number of successful landings recorded for the trial.
- ability is the cognitive ability score (combined ASVB score).
- timesq is the squared value of the number of the trial.

Note:

Permission for SSI to use the copyrighted raw data was provided by R. Kanfer and P.L. Ackerman, and are from experiments reported in: Kanfer, R., and Ackerman, P.L. (1989). Motivation and

cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition. *Journal of Applied Psychology. Monograph*, **74**, 657-690. The data remain the copyrighted property of Ruth Kanfer and Phillip L. Ackerman. Further publication or further dissemination of these data is not permitted without the expressed consent of the copyright owners.

Variance decomposition model with random intercepts

The simplest multilevel model is equivalent to a one-way ANOVA with random intercepts. Although this model is not interesting in itself, it is useful as a preliminary step in a multilevel analysis as it provides important information about the outcome variability at each of the levels of the hierarchy. It may also function as a baseline with which more sophisticated models may be compared.

Let the subscripts i and j refer to the i -th subject and the j -th trial respectively. Using this notation, the one-way ANOVA model can be expressed as

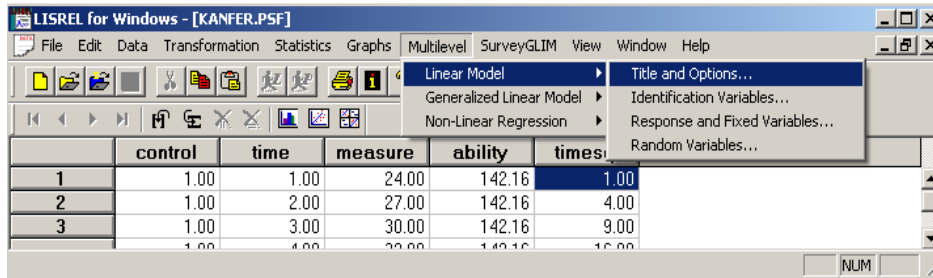
$$\text{measure}_{ij} = \beta_0 + u_{0i} + e_{ij}$$

where measure_{ij} denotes the number of successes recorded for the j -th trial for subject i , β_0 denotes the intercept of the fixed part of the model, u_{0i} represents the random variation in intercepts at level-2 of the model and e_{ij} denotes the random variation at level-1 of the model. It is assumed that u_{0i} has mean 0 and variance $\Phi_{(2)}$. The variance $\Phi_{(2)}$ may be interpreted as the ‘between-group’ variability. Likewise, it is assumed that e_{ij} follows a Normal distribution with mean 0 and variance $\Phi_{(1)}$. Thus, $\Phi_{(1)}$ may be interpreted as the ‘within-group’ variability.

Fitting the variance decomposition model with random intercepts

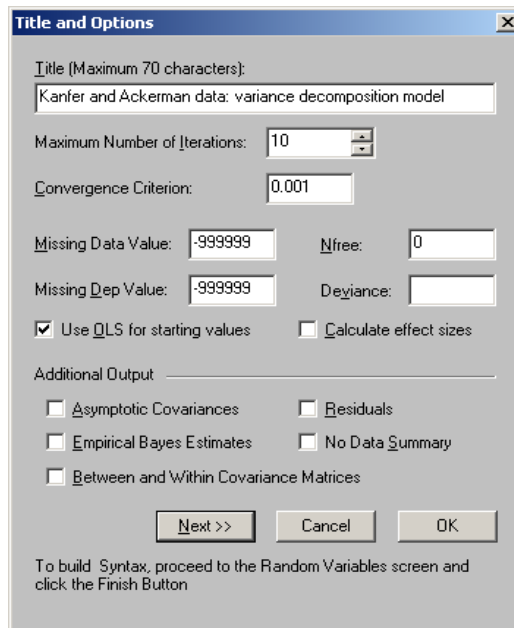
To fit the model above to the data in **kanfer.psf**, we proceed as follows.

- Select the **Open** option on the **File** menu to load the **Open** dialog box.
- Select the **PRELIS Data (*.psf)** option on the **Files of type** drop-down list box.
- Browse for the file **kanfer.psf** in the **TUTORIAL** subfolder and select it.
- Click on the **Open** button to open the PSF window for **kanfer.psf**.
- Select the **Title and Options** option from the **Linear Model** popup menu on the **Multilevel** menu as shown below

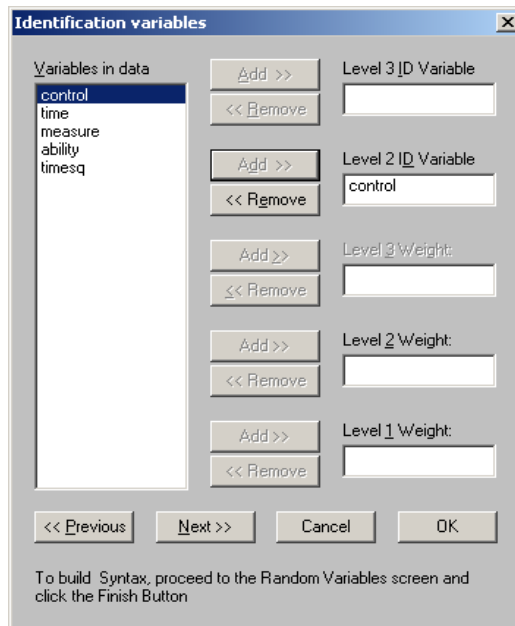


to activate the **Title and Options** dialog box.

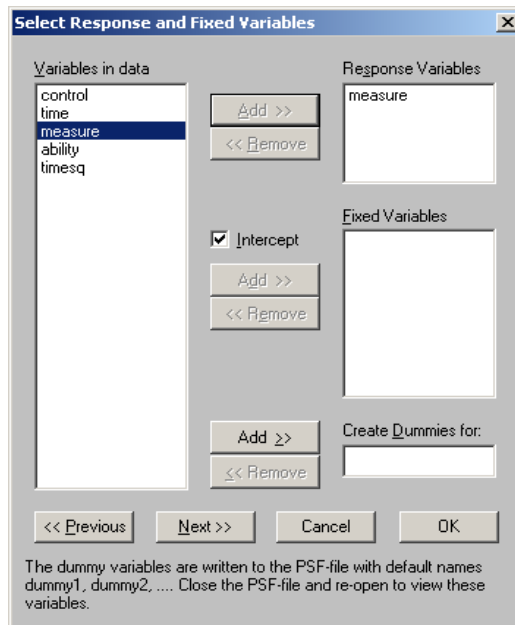
- o Enter the title Kanfer and Ackerman data: variance decomposition model in the **Title** string box to produce the following dialog box.



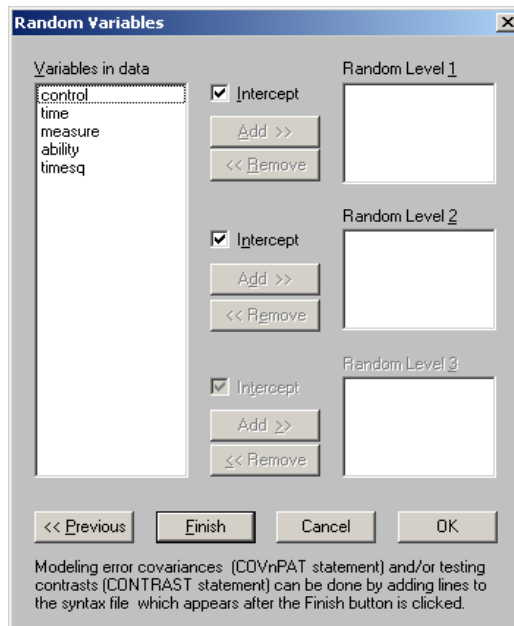
- o Click on the **Next** button to go to the **Identification Variables** dialog box.
- o Select the variable control from the **Variables in data** list box and click on the second **Add** button to produce the following dialog box.



- Click on the **Next** button to go to the **Select Response and Fixed Variables** dialog box.
- Select the variable **measure** from the **Variables in data** list box and click on the upper **Add** button to obtain the following dialog box.



- Click on the **Next** button to go to the following **Random Variables** dialog box.



- o Click on the **Finish** button to generate the following text editor window for **kanfer.pr2**.

```

KANFER.PR2
OPTIONS OLS=YES CONVERGE=0.001000 MAXITER=10 OUTPUT=STANDARD ;
TITLE=Kanfer and Ackerman data: variance decomposition model;
SY='C:\Program Files\lisrel872\TUTORIAL\KANFER.PSF';
ID2=control;
RESPONSE=measure;
FIXED=intcept;
RANDOM1=intcept;
RANDOM2=intcept;

```

- o Click the **Run PRELIS** button to open the text editor window for the output file **kanfer.out**.

Discussion of results

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

In the first selection of the output file, a description of the hierarchical structure of the data is provided as shown below.

```

KANFER.OUT
+-----+
| DATA SUMMARY |
+-----+

NUMBER OF LEVEL 2 UNITS :    141
NUMBER OF LEVEL 1 UNITS :    840

N2 :      1      2      3      4      5      6      7      8
N1 :      6      6      6      6      6      6      6      6

N2 :      9     10     11     12     13     14     15     16
N1 :      6      6      6      6      6      6      6      6

N2 :     17     18     19     20     21     22     23     24

```

The data summary above indicates that a total of 840 measurements scores were obtained from 141 personnel. In addition, a summary of the number of measurements per subject is provided. For example, for subject number 1 (N2: 1), there are 6 measurements (N1: 6) available.

The portion of the output file that contains the estimation results for the fixed part of the model is shown next.

```

KANFER.OUT
Kanfer and Ackerman data: variance decomposition model
ITERATION NUMBER      4
+-----+
| FIXED PART OF MODEL |
+-----+

-----
COEFFICIENTS          BETA-HAT      STD.ERR.      Z-VALUE      PR > |Z|
-----
intcept                26.15108      0.66707      39.20293      0.00000

+-----+
| -2 LOG-LIKELIHOOD |
+-----+

DEVIANCE= -2*LOG(LIKELIHOOD) = 6380.914367903761
NUMBER OF FREE PARAMETERS = 3

```

The results above indicate that $\hat{\beta}_0 = 26.15108$ which is statistically significant ($p < 0.000001$).

The estimation results for the random part of the model are shown below.

RANDOM PART OF MODEL				
LEVEL 2	TAU-HAT	STD.ERR.	Z-VALUE	PR > Z
intcept / intcept	47.22948	7.51714	6.28290	0.00000
LEVEL 1	TAU-HAT	STD.ERR.	Z-VALUE	PR > Z
intcept / intcept	92.17651	4.93041	18.69552	0.00000

It is evident from the results above that $\hat{\Phi}_{(2)} = 47.22948$ and $\hat{\Phi}_{(1)} = 92.17651$. Both these estimates are highly significant.

An estimate of the level-2 effect (differences between the 141 personnel), for example, is obtained as

$$\frac{47.22948}{47.22948 + 92.17651} = 33.88\%$$

indicating that about 33.88% of the total variation is explained by the differences between personnel.

Non-linear growth model with random intercepts and slopes

In data of this nature, it can be expected that the number of successful landings per trial may be influenced by previous experience. In order to take this into account, we now introduce the order of the measurements into the model. The variable time indicates the number of the trial, and timesq the quadratic term equal to time \times time. The corresponding model may be expressed as

$$\text{measure}_{ij} = \beta_0 + \beta_1 \times \text{time}_{ij} + \beta_2 \times \text{timesq}_{ij} + u_{0i} + u_{1i} \times \text{time}_{ij} + u_{2i} \times \text{timesq}_{ij} + e_{ij}$$

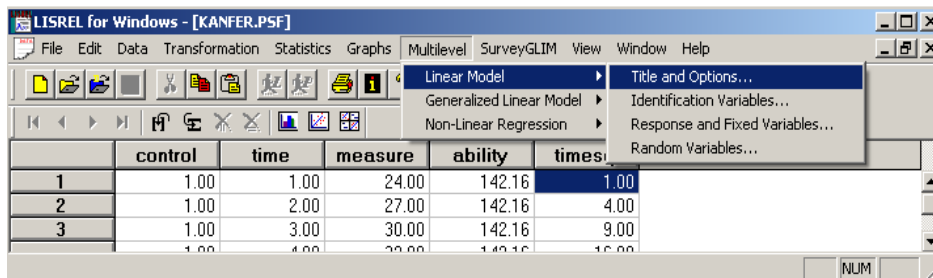
where measure_{ij} denotes the number of successful landings for the j -th trial for subject i , β_0 denotes the intercept of the fixed part of the model, β_1 denotes the regression weight for time in the fixed part of the model, time_{ij} denotes the time point for the j -th trail for subject i , β_2 denotes the regression weight for timesq in the fixed part of the model, timesq_{ij} denotes the squared value of time for the j -th trail for subject i , u_{0i} represents the random variation in intercepts at level-2 of the model, u_{1i} denotes the random variation in linear slopes at level-2 of the model, u_{2i} denotes the

random variation in the nonlinear slopes at level-2 of the model and e_{ij} denotes the random variation at level-1 of the model. It is assumed that $\mathbf{u}_i = (u_{0i} \ u_{1i} \ u_{2i})'$ has mean $\mathbf{0}$ and covariance matrix $\Phi_{(2)}$ and that e_{ij} follows a Normal distribution with mean 0 and variance $\Phi_{(1)}$.

Fitting the non-linear growth model with random intercepts and slopes

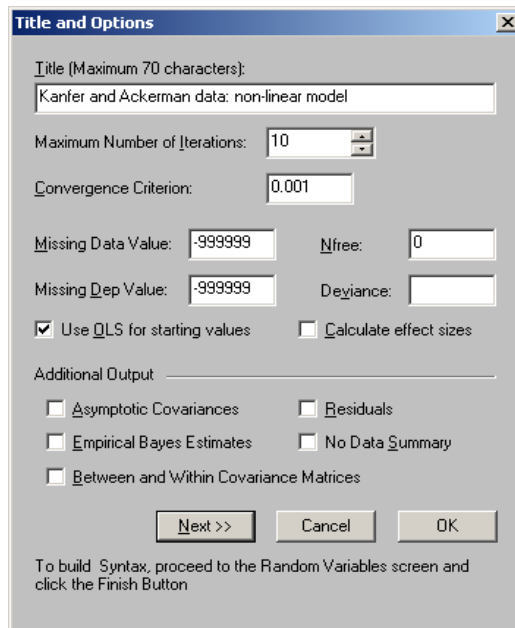
To fit the model above to the data in **kanfer.psf**, we proceed as follows.

- Select the **Open** option on the **File** menu to load the **Open** dialog box.
- Select the **PRELIS Data (*.psf)** option on the **Files of type** drop-down list box.
- Browse for the file **kanfer.psf** in the **TUTORIAL** subfolder and select it.
- Click on the **Open** button to open the PSF window for **kanfer.psf**.
- Select the **Title and Options** option from the **Linear Model** popup menu on the **Multilevel** menu as shown below

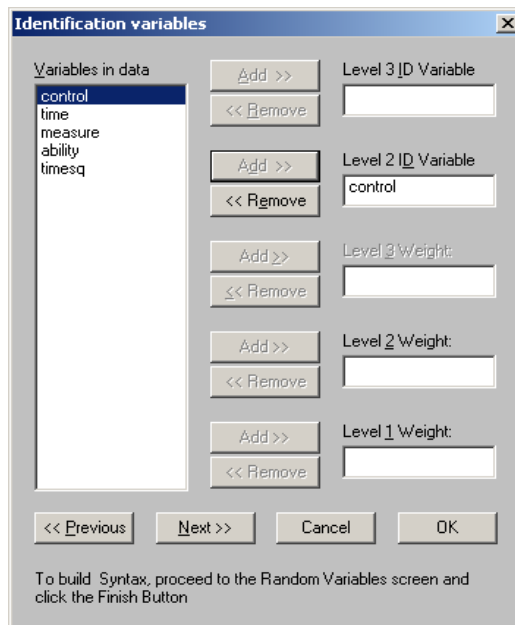


to activate the **Title and Options** dialog box.

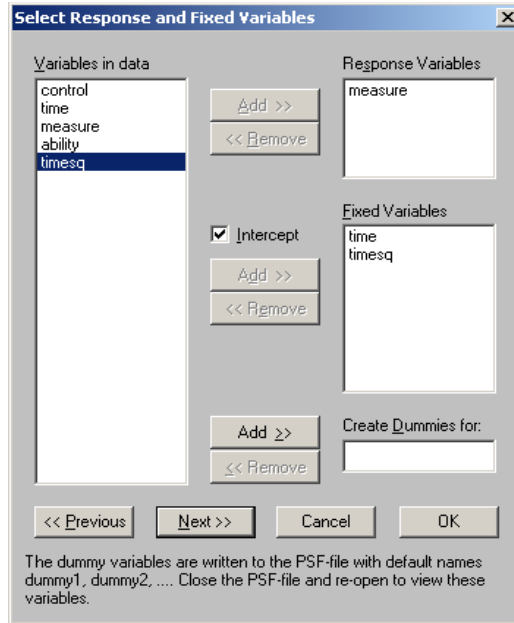
- Enter the title Kanfer and Ackerman data: non-linear model in the **Title** string box to produce the following dialog box.



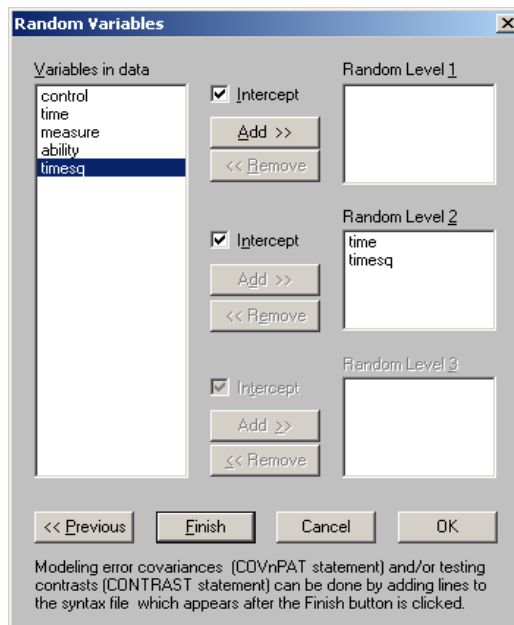
- Click on the **Next** button to go to the **Identification Variables** dialog box.
- Select the variable control from the **Variables in data** list box and click on the second **Add** button to produce the following dialog box.



- Click on the **Next** button to go to the **Select Response and Fixed Variables** dialog box.
- Select the variable measure from the **Variables in data** list box and click on the upper **Add** button to add the variable measure to the **Response Variables** list box.
- Select the variables time and timesq from the **Variables in data** list box and click on the middle **Add** button to obtain the following dialog box.



- Click on the **Next** button to go to the **Random Variables** dialog box.
- Select the variables **time** and **timesq** from the **Variables in data** list box and click on the middle **Add** button to obtain the following dialog box.



- Click on the **Finish** button to generate the following text editor window for **kanfer.pr2**.

```

KANFER.PR2
OPTIONS OLS=YES CONVERGE=0.001000 MAXITER=10 OUTPUT=STANDARD ;
TITLE=Kanfer and Ackerman data: non-linear model;
SY='C:\Program Files\lisrel873\TUTORIAL\KANFER.PSF';
ID2=control;
RESPONSE=measure;
FIXED=intcept time timesq;
RANDOM1=intcept;
RANDOM2=intcept time timesq;

```

- Click the **Run PRELIS** button to open the text editor window for the output file **kanfer.out**.

Discussion of results

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

In the first selection of the output file, a description of the hierarchical structure of the data is provided as shown below.

```

KANFER.OUT
|
|          +-----+
|          | DATA SUMMARY |
|          +-----+
|
| NUMBER OF LEVEL 2 UNITS :      141
| NUMBER OF LEVEL 1 UNITS :      840
|
| N2 :      1      2      3      4      5      6      7      8
| N1 :      6      6      6      6      6      6      6      6
|
| N2 :      9     10     11     12     13     14     15     16
| N1 :      6      6      6      6      6      6      6      6
|
| N2 :      17     18     19     20     21     22     23     24

```

The data summary above indicates that a total of 840 measurements were obtained from 141 personnel. In addition, a summary of the number of measurements per subject is provided. For example, for subject number 1 (N2: 1), there are 6 measurements (N1: 6) available.

The portion of the file that contains the estimation results for the fixed part of the model is shown next.

Kanfer and Ackerman data: non-linear model
 ITERATION NUMBER 4

```

+-----+
| FIXED PART OF MODEL |
+-----+

```

COEFFICIENTS	BETA-HAT	STD.ERR.	Z-VALUE	PR > Z
intcept	1.67806	0.87343	1.92123	0.05470
time	11.48966	0.50397	22.79838	0.00000
timesq	-1.03583	0.06141	-16.86761	0.00000

```

+-----+
| -2 LOG-LIKELIHOOD |
+-----+

```

DEVIANCE= -2*LOG(LIKELIHOOD) = 5095.465854309439
 NUMBER OF FREE PARAMETERS = 10

The results above indicate that the measurement score increases significantly over time. In addition, this increase is quadratic.

The estimation results for the random part of the model are shown below.

```

+-----+
| RANDOM PART OF MODEL |
+-----+

```

LEVEL 2	TAU-HAT	STD.ERR.	Z-VALUE	PR > Z
intcept /intcept	77.02126	12.98193	5.93296	0.00000
time /intcept	-24.73085	6.50474	-3.80197	0.00014
time /time	22.66462	4.35831	5.20033	0.00000
timesq /intcept	2.26771	0.76607	2.96017	0.00307
timesq /time	-2.38073	0.52370	-4.54595	0.00001
timesq /timesq	0.27282	0.06566	4.15473	0.00003

LEVEL 1	TAU-HAT	STD.ERR.	Z-VALUE	PR > Z
intcept /intcept	9.49299	0.65567	14.47823	0.00000

The total estimated variance on level-2 follows as

$$\begin{aligned}
 & \hat{\sigma}^2(u_0 + \text{time} \times u_1 + \text{time}^2 \times u_2) \\
 &= \hat{\sigma}^2(u_0) + \text{time}^2 \times \hat{\sigma}^2(u_1) + \text{time}^4 \times \hat{\sigma}^2(u_2) + 2 \times \text{time} \times \hat{\sigma}(u_0, u_1) \\
 & \quad + 2 \times \text{time}^2 \times \hat{\sigma}(u_0, u_2) + 2 \times \text{time}^3 \times \hat{\sigma}(u_1, u_2) \\
 &= \hat{\Phi}_{(2)1,1} + \text{time}^2 \times \hat{\Phi}_{(2)2,2} + \text{time}^4 \times \hat{\Phi}_{(2)3,3} + 2 \times \text{time} \times \hat{\Phi}_{(2)2,1} \\
 & \quad + 2 \times \text{time}^2 \times \hat{\Phi}_{(2)3,1} + 2 \times \text{time}^3 \times \hat{\Phi}_{(2)3,2} \\
 &= 77.02126 + 22.66462 \times \text{time}^2 + 0.27282 \times \text{time}^4 - 49.4617 \times \text{time} \\
 & \quad + 4.53542 \times \text{time}^2 - 4.76146 \times \text{time}^3
 \end{aligned}$$

We can thus express the total estimated variation on level-2 as a nonlinear function of the variable time. When time is one, an estimate of the level-2 effect (differences between the 141 subjects), for example, is obtained as

$$\frac{50.27096}{50.27096+9.49299} = 84.16\%$$

indicating that about 84.16% of the total variation in the number of successful landings is explained by the differences between personnel at the first time point.

Non-linear growth model with random intercepts and slopes and with a covariate

Recall that the data set also includes a measure of cognitive ability (composite ASVB score) for each of the 141 air traffic controllers, denoted by ability in **kanfer.psf**. We now include this variable as a fixed effect (covariate) in the model. The corresponding model is given by

$$\text{measure}_{ij} = \beta_0 + \beta_1 \times \text{time}_{ij} + \beta_2 \times \text{timesq}_{ij} + \beta_3 \times \text{ability}_{ij} + u_{0i} + u_{1i} \times \text{time}_{ij} + u_{2i} \times \text{timesq}_{ij} + e_{ij}$$

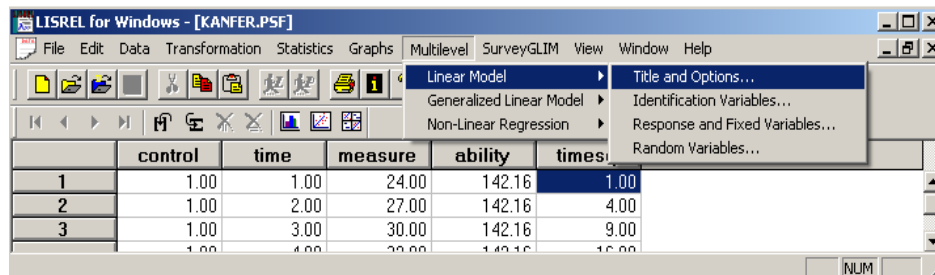
where measure_{ij} denotes the number of successful landings for the j -th trial for subject i , β_0 denotes the intercept of the fixed part of the model, β_1 denotes the regression weight for time in the fixed part of the model, time_{ij} denotes the time point for the j -th trail for subject i , β_2 denotes the regression weight for timesq in the fixed part of the model, timesq_{ij} denotes the squared value of time for the j -th trail for subject i , β_3 denotes the regression weight for ability in the fixed part of the model, ability_{ij} denotes the ability score for the j -th trail for subject i , u_{0i} represents the random variation in intercepts at level-2 of the model, u_{1i} denotes the random variation in linear slopes at level-2 of the model, u_{2i} denotes the random variation in the nonlinear slopes at level-2 of the model and e_{ij} denotes the random variation at level-1 of the model. It is assumed that $\mathbf{u}_i = (u_{0i} \ u_{1i} \ u_{2i})'$ has mean $\mathbf{0}$ and covariance matrix $\Phi_{(2)}$ and that e_{ij} follows a Normal distribution with mean 0 and variance $\Phi_{(1)}$.

Fitting the non-linear growth model with random intercepts and slopes and with a covariate

To fit the non-linear model above to the data in **kanfer.psf**, we proceed as follows.

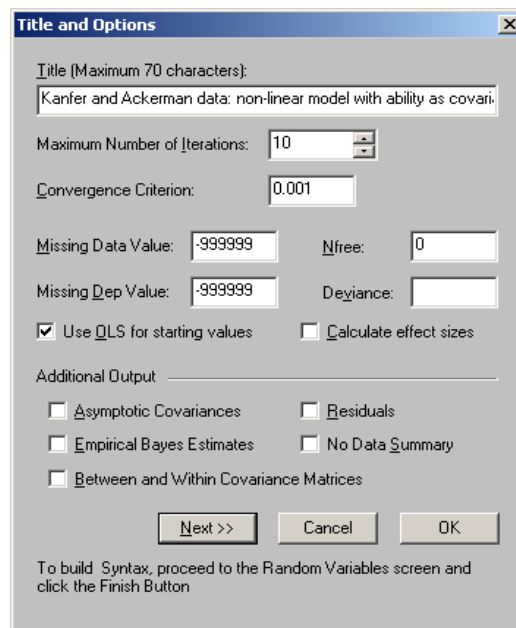
- Select the **Open** option on the **File** menu to load the **Open** dialog box.

- Select the **PRELIS Data (*.psf)** option on the **Files of type** drop-down list box.
- Browse for the file **kanfer.psf** in the **TUTORIAL** subfolder and select it.
- Click on the **Open** button to open the PSF window for **kanfer.psf**.
- Select the **Title and Options** option from the **Linear Model** popup menu on the **Multilevel** menu as shown below

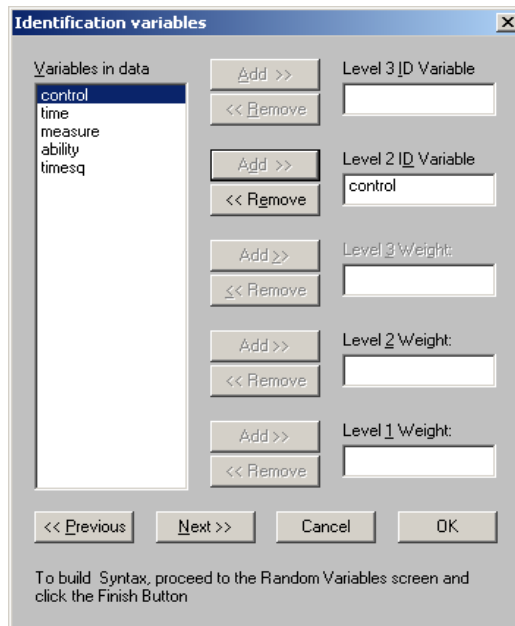


to activate the **Title and Options** dialog box.

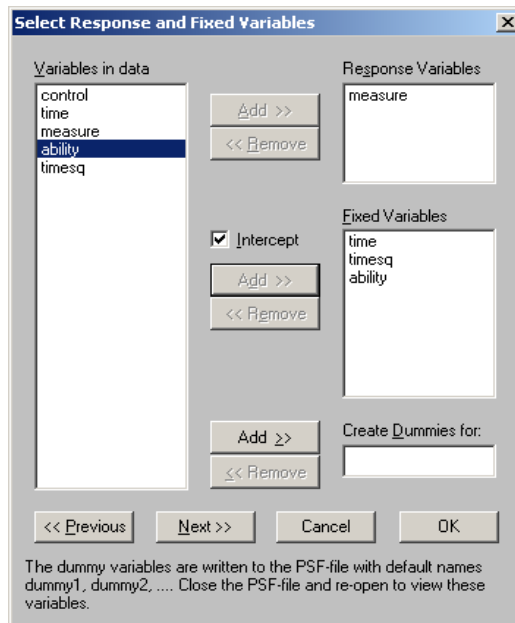
- Enter the title Kanfer and Ackerman data: non-linear model with ability as covariate in the **Title** string box to produce the following dialog box.



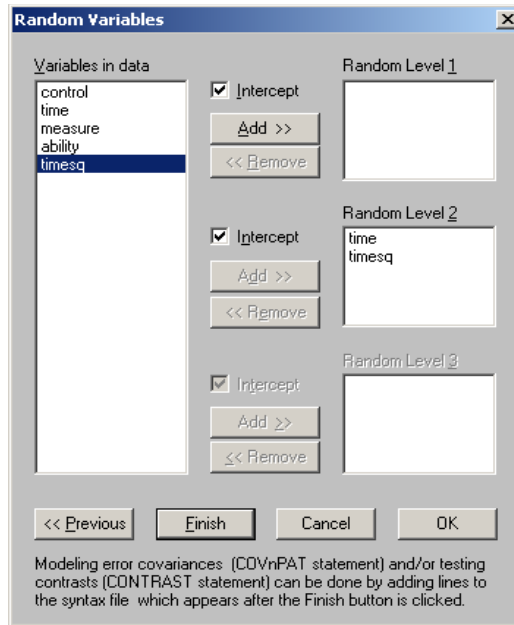
- Click on the **Next** button to go to the **Identification Variables** dialog box.
- Select the variable control from the **Variables in data** list box and click on the second **Add** button to produce the following dialog box.



- o Click on the **Next** button to go to the **Select Response and Fixed Variables** dialog box.
- o Select the variable **measure** from the **Variables in data** list box and click on the upper **Add** button to add the variable **measure** to the **Response Variables** list box.
- o Select the variables **time**, **timesq** and **ability** from the **Variables in data** list box and click on the middle **Add** button to obtain the following dialog box.



- o Click on the **Next** button to go to the **Random Variables** dialog box.
- o Select the variables **time** and **timesq** from the **Variables in data** list box and click on the middle **Add** button to obtain the following dialog box.



- o Click on the **Finish** button to generate the following text editor window for **kanfer.pr2**.

```

KANFER.PR2
OPTIONS OLS=YES CONVERGE=0.001000 MAXITER=10 OUTPUT=STANDARD ;
TITLE=Kanfer and Ackerman data: non-linear model with ability as covariate ;
SY='C:\Program Files\lisrel872\TUTORIAL\KANFER.PSF';
ID2=control;
RESPONSE=measure;
FIXED=intcept time timesq ability;
RANDOM1=intcept;
RANDOM2=intcept time timesq;

```

- o Click the **Run PRELIS** button to open the text editor window for the output file **kanfer.out**.

Discussion of results

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

In the first selection of the output file, a description of the hierarchical structure of the data is provided as shown below.

```

KANFER.OUT
|
+-----+
| DATA SUMMARY |
+-----+

NUMBER OF LEVEL 2 UNITS :    141
NUMBER OF LEVEL 1 UNITS :    840

N2 :      1      2      3      4      5      6      7      8
N1 :      6      6      6      6      6      6      6      6

N2 :      9     10     11     12     13     14     15     16
N1 :      6      6      6      6      6      6      6      6

N2 :     17     18     19     20     21     22     23     24

```

The data summary above indicates that a total of 840 measurements were obtained from 141 personnel. In addition, a summary of the number of measurements per subject is provided. For example, for subject number 1 (N2: 1), there are 6 measurements (N1: 6) available.

The portion of the output file that contains the estimation results for the fixed part of the model is shown next.

```

KANFER.OUT
Kanfer and Ackerman data: non-linear model with ability
ITERATION NUMBER      4
+-----+
| FIXED PART OF MODEL |
+-----+

-----
COEFFICIENTS          BETA-HAT      STD.ERR.      Z-VALUE      PR > |Z|
-----
intcept                1.67596      0.78914       2.12379      0.03369
time                   11.49203     0.50394      22.80458     0.00000
timesq                 -1.03639     0.06140     -16.87917    0.00000
ability                0.03703     0.00519       7.12990     0.00000

+-----+
|  -2 LOG-LIKELIHOOD  |
+-----+

DEVIANCE= -2*LOG(LIKELIHOOD) = 5052.791958786890
NUMBER OF FREE PARAMETERS = 11

```

The results above indicate that the measurement score increases significantly over time. In addition, this increase is quadratic. Ability has significant positive effect on measure.

The estimation results for the random part of the model are shown below.

RANDOM PART OF MODEL					
LEVEL 2		TAU-HAT	STD. ERR.	Z-VALUE	PR > Z
intcept	/intcept	57.26989	10.66634	5.36921	0.00000
time	/intcept	-22.58400	5.99981	-3.76412	0.00017
time	/time	22.66330	4.35771	5.20074	0.00000
timesq	/intcept	2.09446	0.70778	2.95921	0.00308
timesq	/time	-2.38017	0.52358	-4.54592	0.00001
timesq	/timesq	0.27273	0.06565	4.15457	0.00003
LEVEL 1		TAU-HAT	STD. ERR.	Z-VALUE	PR > Z
intcept	/intcept	9.49057	0.65551	14.47819	0.00000

The total estimated variance on level-2 follows as

$$\begin{aligned}
& \hat{\sigma}^2(u_0 + \text{time} \times u_1 + \text{time}^2 \times u_2) \\
&= \hat{\sigma}^2(u_0) + \text{time}^2 \times \hat{\sigma}^2(u_1) + \text{time}^4 \times \hat{\sigma}^2(u_2) + 2 \times \text{time} \times \hat{\sigma}(u_0, u_1) \\
& \quad + 2 \times \text{time}^2 \times \hat{\sigma}(u_0, u_2) + 2 \times \text{time}^3 \times \hat{\sigma}(u_1, u_2) \\
&= \hat{\Phi}_{(2)1,1} + \text{time}^2 \times \hat{\Phi}_{(2)2,2} + \text{time}^4 \times \hat{\Phi}_{(2)3,3} + 2 \times \text{time} \times \hat{\Phi}_{(2)2,1} \\
& \quad + 2 \times \text{time}^2 \times \hat{\Phi}_{(2)3,1} + 2 \times \text{time}^3 \times \hat{\Phi}_{(2)3,2} \\
&= 57.26989 + 22.66330 \times \text{time}^2 + 0.27273 \times \text{time}^4 - 45.16800 \times \text{time} \\
& \quad + 4.18892 \times \text{time}^2 - 4.76034 \times \text{time}^3
\end{aligned}$$

We can thus express the total estimated variation on level-2 as a nonlinear function of the variable time.

Three-level analysis of health expenditure data

The data

The data set used here forms part of the data library of the Medical Expenditure Panel Survey (MEPS). Collected in 1999, these data from a longitudinal national survey were used to obtain regional and national estimates of health care use and expenditure based on the health expenditures of a sample of U.S. civilian non-institutionalized participants. The survey sample design utilized stratification, clustering, multiple stages of selection, and disproportionate sampling. The sample was drawn from 143 strata, divided into 460 PSUs. Information on 23,565 participants included positive person-level weights and forms the data set used here, excluding the 1,053 participants in the original data with zero person-level weights. Data for the first 10 participants on most of the variables used in this section are shown in the PSF window below.

	VARSTR99	VARPSU99	PERWT99F	TOTEXP99	RACE	SEX	INSCOV	RPOVC991	RPOVC992	RPOV
1	1.00	1.00	3551.61	0.00	0.00	0.00	1.00	0.00	0.00	
2	1.00	1.00	4495.33	4.20	0.00	1.00	1.00	0.00	0.00	
3	1.00	1.00	7689.24	4.98	0.00	0.00	1.00	0.00	0.00	
4	1.00	2.00	11069.78	5.18	0.00	1.00	0.00	1.00	0.00	
5	1.00	2.00	9288.98	8.11	0.00	0.00	0.00	1.00	0.00	
6	1.00	2.00	12651.61	10.09	1.00	0.00	0.00	0.00	0.00	
7	1.00	2.00	6156.33	6.64	1.00	1.00	0.00	0.00	0.00	
8	1.00	2.00	31359.46	8.29	0.00	0.00	0.00	1.00	0.00	
9	1.00	2.00	29082.10	7.85	0.00	1.00	0.00	1.00	0.00	
10	1.00	2.00	14474.96	8.47	0.00	1.00	1.00	0.00	1.00	

The variables of interest are:

- VARSTR99 is the stratum variable (143 strata in total).
- VARPSU99 is the PSU (cluster) variable (460 PSUs in total).
- PERWT99F represents the final sample weight, with weights ranging between 307.16 and 80061.61, correcting for both non-response and adjustments to population control totals from the Current Population Survey.
- TOTEXP99 is the natural logarithm of the total health expenditure of a respondent in 1999, ranging between 0 and 12.24 and representing actual expenditure of between \$0 and \$206,721.
- RACE is an ethnicity indicator, with a value of 1 indicating white respondents, and 0 denoting all other ethnic groups as well as respondents for which ethnicity is not known. This variable was recoded from the original MEPS variable RACEX.
- SEX is a gender indicator, with a value of 0 indicating a male participant and 1 a female participant; recoded from the original MEPS variable RSEX.
- INSCOV is an indicator of the level of insurance coverage, where 0 indicates private coverage any time during 1999, and 1 indicates public coverage or no insurance at all during 1999.

- RPOVC991 to RPOVC995 are five indicator variables, each associated with a category of the original MEPS variable RPOVC99 which was constructed by dividing family income by the applicable poverty line (selection of which depended on family size and composition), expressed as a percentage.

Income is a variable that is often transformed using its natural logarithm. Doing so in effect causes the impact of each additional dollar to decrease as income increases. Logarithmic transformation is also useful in lessening the influence of outliers, as the natural logarithm of a variable is much less sensitive to extreme observations than is the variable itself.

The original MEPS variable RPOVC99 assumed a value of 1 for a family with "high" income level where family income was equal to or greater than 400% of the applicable poverty line, and a value of 2 for those with a "low income" level (associated with 125% to 200% of the poverty line). Families with "middle income", "near poor" and "negative or poor" levels of income relative to poverty line income were coded 3, 4 and 5 respectively. For the "middle income" category, the ratio (as percentage) of family income to poverty line was 200% to less than 400%. In the case of "near poor" families, the percentages ranged between 100% and 125%, and for "negative or poor", the family income was less than 100% of the relevant poverty line. Thus, a value of 1 on the indicator variable RPOVC991 indicates a family with income at the "high" level, while a value of 1 on the variable RPOVC995 indicates a family with "negative or poor" income level. The variables RPOVC992, RPOVC993, and RPOVC994 are associated with the categories "low income", "middle income" and "near poor" respectively.

Note that as each of the five indicator variables for categories of RPOVC99 is coded 1 if a participant responded in that category and 0 otherwise, only four of the five indicator variables can be used in a model where an intercept is included.

Regression model with random intercepts

A multilevel model does not make provision for the specification of design related variables such as stratum or PSU. Instead, these design variables are used to define the hierarchical structure of the data. In this example, the stratum identification variable VARSTR99 is used as the level-3 identifier and the PSU identification variable VARPSU99 serves to identify level-2 units (*i.e.*, PSUs) nested within a given stratum. We thus use the design variables to define a three-level hierarchical structure, with participants as level-1 observations nested within PSUs, in turn nested within strata. While not explicitly acknowledging the survey design or offering a conventional design effect estimate to measure the difference in estimates obtained when implementing this design compared to estimates obtained under a simple random sample, a multilevel model offers the advantage of estimating the variation in total health care expenditure within and between PSUs.

A general three-level regression model for a response variable y depending on a set of r predictors x_1, x_2, \dots, x_r can be expressed as

$$y_{ijk} = \mathbf{x}'_{(f)ijk} \boldsymbol{\beta} + \mathbf{x}'_{(3)ijk} \mathbf{v}_i + \mathbf{x}'_{(2)ijk} \mathbf{u}_{ij} + \mathbf{x}'_{(1)ijk} \mathbf{e}_{ijk}$$

where $i = 1, 2, \dots, N$ denotes the level-3 units, $j = 1, 2, \dots, n_i$ the level-2 units, and $k = 1, 2, \dots, n_{ij}$ the level-1 units. In this context, y_{ijk} represents the response of individual k , nested within level-2 unit j and level-3 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 $\mathbf{x}'_{(f)ijk} \boldsymbol{\beta}$, where $\mathbf{x}'_{(f)ijk}$ is a typical row of the design matrix of the fixed part of the model with, as elements, a subset of the r predictors. The vector $\boldsymbol{\beta}$ contains the fixed, but unknown parameters to be estimated. The vector products $\mathbf{x}'_{(3)ijk} \mathbf{v}_i$, $\mathbf{x}'_{(2)ijk} \mathbf{u}_{ij}$, and $\mathbf{x}'_{(1)ijk} \mathbf{e}_{ijk}$ denote the random part of the model at levels 3, 2, and 1 respectively. For example, $\mathbf{x}'_{(3)ijk}$ represents a typical row of the design matrix of the random part at level 3, and \mathbf{v}_i the vector of random level-3 coefficients to be estimated. The products $\mathbf{x}'_{(2)ijk} \mathbf{u}_{ij}$ and $\mathbf{x}'_{(1)ijk} \mathbf{e}_{ijk}$ serve the same purpose at levels 2 and 1 respectively. It is assumed that $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N$ are independently and identically distributed (i.i.d) with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Phi}_{(3)}$. Similarly, $\mathbf{u}_{i1}, \mathbf{u}_{i2}, \dots, \mathbf{u}_{in_i}$ are assumed i.i.d., with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Phi}_{(2)}$, and $\mathbf{e}_{ij1}, \mathbf{e}_{ij2}, \dots, \mathbf{e}_{ijn_{ij}}$ are assumed i.i.d., with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Phi}_{(1)}$.

Within this hierarchical framework, the model fitted to the data uses the participant's gender, ethnicity, type of health insurance cover, and measure of income relative to poverty level to predict the total expenditure on health care in 1999, the latter transformed to the natural logarithm of the actual expenses incurred.

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

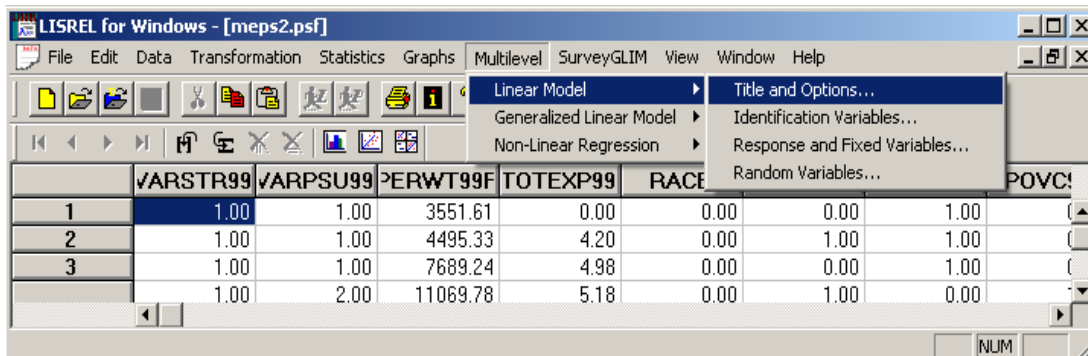
where β_0 denotes the average expected total expenditure on health care in 1999, and $\beta_1, \beta_2, \dots, \beta_7$ indicate the regression weights associated with the fixed part of the model which contains the predictor variables SEX, RACE, INSCOV and the indicator variables for categories of income relative to the poverty level. The random part of the model is represented by v_{i0} , u_{ij0} and e_{ijk} , which denote the variation in average total health related expenditure over strata, between PSUs (or, in other words, over PSUs nested within strata) and between participants at the lowest level of the hierarchy respectively.

Fitting the regression model with random intercepts

To fit the 3-level model above to the data in **meps2.psf**, we proceed as follows.

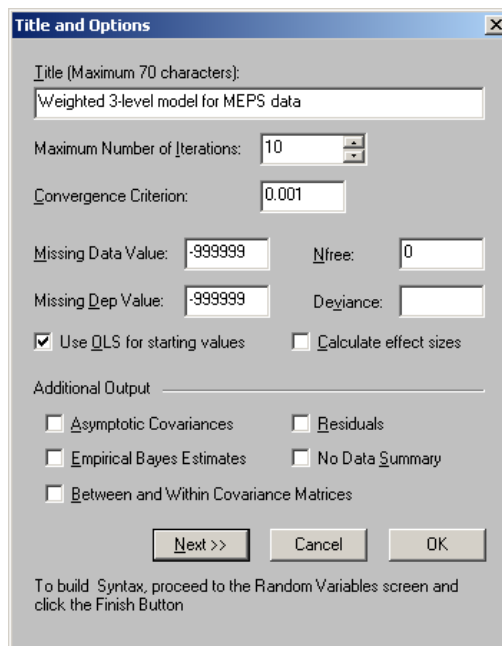
- Select the **Open** option on the **File** menu to load the **Open** dialog box.
- Select the **PRELIS Data (*.psf)** option on the **Files of type** drop-down list box.
- Browse for the file **meps2.psf** in the **TUTORIAL** subfolder and select it.

- Click on the **Open** button to open the PSF window for **meps2.psf**.
- Select the **Title and Options** option from the **Linear Model** popup menu on the **Multilevel** menu as shown below

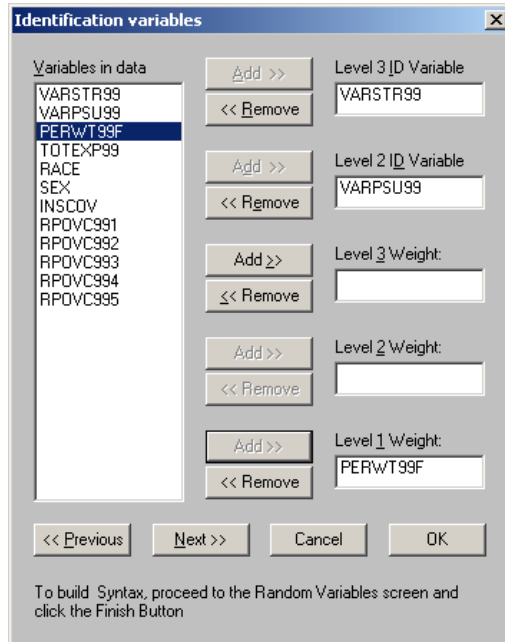


to activate the **Title and Options** dialog box.

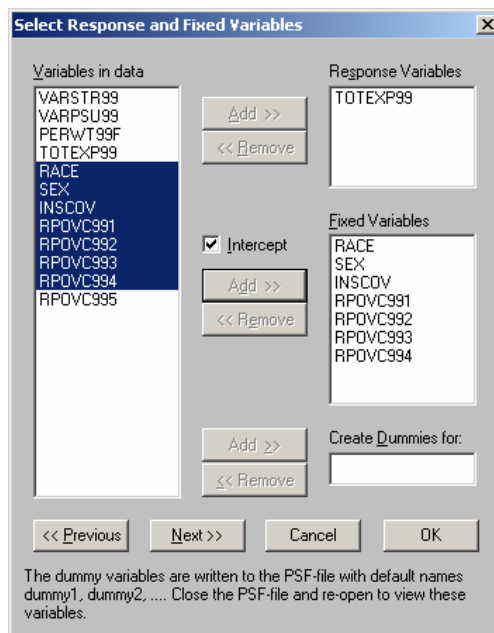
- Enter the title **Weighted 3-level model for MEPS data** in the **Title** string box to produce the following dialog box.



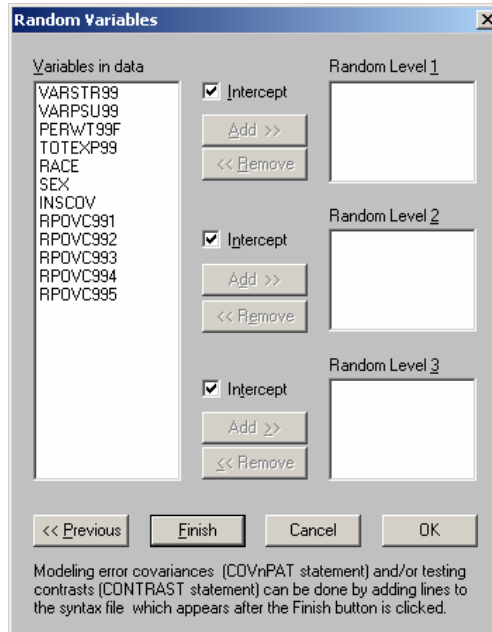
- Click on the **Next** button to go to the **Identification Variables** dialog box.
- Select the variable **VARSTR99** from the **Variables in data** list box and click on the first **Add** button to add the variable in the **Level 3 ID Variable** grid box.
- Select the variable **VARPSU99** from the **Variables in data** list box and click on the second **Add** button to add the variable in the **Level 2 ID Variable** grid box.
- Select the variable **PERWT99** from the **Variables in data** list box and click on the final **Add** button to produce the following dialog box.



- Click on the **Next** button to go to the **Select Response and Fixed Variables** dialog box.
- Select the variable TOTEXP99 from the **Variables in data** list box and click on the upper **Add** button to add the variable TOTEXP99 to the **Response Variables** list box.
- Select the variables RACE, SEX, INSCOV, RPOVC991, RPOVC992, RPOVC993 and RPOVC994 from the **Variables in data** list box and click on the middle **Add** button to obtain the following dialog box.



- Click on the **Next** button to go to the **Random Variables** dialog box.



- Click on the **Finish** button to generate the following text editor window for **meps2.pr2**.

```

meps2.PR2
|
|  OPTIONS OLS=YES CONVERGE=0.001000 MAXITER=10 OUTPUT=STANDARD ;
|  TITLE=Weighted 3-level model for MEPS data;
|  SY='C:\Program Files\lisrel88\TUTORIAL\meps2.psf';
|  ID3=VARSTR99;
|  ID2=VARPSU99;
|  WEIGHT1=PERWT99F;
|  RESPONSE=TOTEXP99;
|  FIXED=intcept RACE SEX INSCOV RPOVC991 RPOVC992 RPOVC993 RPOVC994;
|  RANDOM1=intcept;
|  RANDOM2=intcept;
|  RANDOM3=intcept;
|
|  x

```

- Click the **Run PRELIS** button to open the text editor window for the output file **meps2.out**.

Discussion of results

Portions of the output file **meps2.out** are shown below. In the first section of the output file a description of the hierarchical structure is provided as shown below.

```

+-----+
| DATA SUMMARY |
+-----+

NUMBER OF LEVEL 3 UNITS :      143
NUMBER OF LEVEL 2 UNITS :      460
NUMBER OF LEVEL 1 UNITS :    23564

ID3 :      1      2      3      4      5      6      7      8
N2 :      2      2      2      2      2      3      2      2
N1 :      29     85     55     86    159     48     48     78

ID3 :      9     10     11     12     13     14     15     16
N2 :      2      2      2     11      2      2      2      3
N1 :     114     23     68    408     68     62    168     59

ID3 :     17     18     19     20     21     22     23     24
N2 :      2      5      2      2      2      2      2      2
N1 :     364    215     69     44     40     59    171    151

ID3 :     25     26     27     28     29     30     31     32
N2 :      3      2      2      2      2      2     10      2
N1 :     225     18     86     25     58     26    417     26

```

The data summary above indicate that a total of 143 strata, 460 PSUs and information from 23,564 individual participants were included at levels 3, 2 and 1 of the multilevel model. This corresponds to the survey design described earlier. In addition, a summary of the number of PSUs and participants nested within each stratum is provided. For stratum number 1 (ID3: 1), data are available from only 29 participants nested within 2 primary sampling units (N2: 2). By contrast, for stratum number 12 (ID3: 12), data are available from 408 participants (N1: 408) nested within 11 primary sampling units (N2: 11).

The next section of the output file contains the results of the fixed part of the model as shown below.

```

Weighted 3-level model for MEPS data
ITERATION NUMBER      6

+-----+
| FIXED PART OF MODEL |
+-----+

-----+-----+-----+-----+-----+
| COEFFICIENTS      | BETA-HAT | STD.ERR. | Z-VALUE | PR > |Z| |
+-----+-----+-----+-----+-----+
| intcept           | 4.39123  | 0.11369  | 38.62367| 0.00000|
| RACE              | 0.94298  | 0.08607  | 10.95549| 0.00000|
| SEX               | 0.91057  | 0.03641  | 25.00765| 0.00000|
| INSCOV           | -0.65109 | 0.07548  | -8.62642| 0.00000|
| RPOVC991         | 0.35750  | 0.11408  | 3.13361 | 0.00173|
| RPOVC992         | -0.13832 | 0.10421  | -1.32736| 0.18439|
| RPOVC993         | 0.07036  | 0.11668  | 0.60302 | 0.54649|
| RPOVC994         | -0.32929 | 0.14044  | -2.34470| 0.01904|
+-----+-----+-----+-----+-----+

+-----+
| -2 LOG-LIKELIHOOD |
+-----+

DEVIANCE= -2*LOG(LIKELIHOOD) = 118256.6158143171
NUMBER OF FREE PARAMETERS = 11

```

The estimates are shown in the column with heading BETA-HAT, and correspond to the coefficients $\beta_0, \beta_2, \dots, \beta_7$ in the model specification. From the z-values and associated exceedance probabilities, we see that the coefficients associated with gender, ethnicity and insurance coverage type were all highly significant. Recall that a value of 1 for the ethnicity indicator variable RACE indicated that a participant was white, with a value of 0 assigned to participants from all other ethnic groups. The positive estimated coefficient for this variable indicates an increase of 0.94298 units in the logarithm of total health expenditure, holding all other predictors constant. Similarly, female participants (coded "1" on the gender indicator SEX), are expected to have a total health expenditure 0.91057 higher than male participants 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 -0.65109.

Turning to the indicator variables associated with income relative to the poverty line income, it can be seen that only two of the indicator variables, RPOVC991 and RPOVC994, have estimated coefficients that are significantly different from zero at a 5% level of significance. In the case of families with a "high" income, the estimate of 0.35750 for RPOVC991 indicates an expected increase in expenditure, while for "near poor" families, the estimate of -0.32939 indicates an expected decrease in expenditure, holding all other variables constant.

Estimated outcomes for different groups

To evaluate the expected effect of the measure of a family's income to the corresponding poverty line income, suppose that the variables RACE, SEX, and INSCOV are held at zero, as would be the case for a nonwhite male participant with private insurance coverage. If such a participant originates from a family with "high" income, the logarithm of total health expenditure is expected to be

$$\begin{aligned} & \beta_0 + \beta_4 (\text{RPOVC991}) + \beta_5 (\text{RPOVC992}) + \beta_6 (\text{RPOVC993}) + \beta_7 (\text{RPOVC994}) \\ &= \beta_0 + \beta_4 \\ &= 4.39123 + 0.35750 \\ &= 4.74873 \end{aligned}$$

which translates to a projected total expenditure of $e^{4.74873} = \$115.437$. In contrast, for a participant with similar demographic background and coverage from a "near poor" family, we obtain a projected total expenditure of

$$\begin{aligned} & e^{\beta_0 + \beta_7} \\ &= e^{4.39123 - 0.32929} \\ &= \$58.086 \end{aligned}$$

The predicted total expenditure (as natural logarithm) for similar participants from "low", "middle" or "negative or poor" families are similarly obtained by calculating $e^{\beta_0 + \beta_5}$, $e^{\beta_0 + \beta_6}$ and e^{β_0} respectively.

Predicted total health expenditure for various subgroups

Respondents with high family income (RPOVC991 = 1)	Male (SEX = 0)		Female (SEX = 1)	
	Insurance coverage:		Insurance coverage:	
	Private (INSCOV=0)	Public/none (INSCOV = 1)	Private (INSCOV=0)	Public/none (INSCOV = 1)
Nonwhite (RACE = 0)	\$115	\$60	\$287	\$150
White (RACE = 1)	\$296	\$155	\$737	\$384

Respondents with near poor income (RPOVC994 = 1)

Nonwhite (RACE = 0)	\$58	\$30	\$145	\$75
White (RACE = 1)	\$149	\$78	\$370	\$193

In the table above, the predicted total health expenditure is given for respondents with high or near poor family income, for each of the subpopulations formed by gender, ethnicity and insurance coverage. For purposes of the comparison, results are expressed in U.S. dollars, rather than in the natural logarithmic units of the outcome variable TOTEXP99. Respondents from families with high income consistently outspend their near poor counterparts by approximately 100%, regardless of gender, ethnicity or level of insurance coverage. In families with high income, female respondents spent more in 1999 than their male counterparts, regardless of ethnicity. This is generally also true for near poor respondents. It is also apparent that the total health expenditure in 1999 was higher for respondents with private insurance than for respondents with public or no coverage, and that white respondents spent more than respondents from other ethnic groups, regardless of gender or the level of family income. From exploratory analyses, we know that the outcome variable TOTEXP99 is highly skewed, with median 1999 expenditure of \$ 377.41. When this is taken in account, we can conclude that, generally speaking, white females spent more on health in 1999 than 50% of all respondents in the sample. The results for the random part of the model as shown below.

```

+-----+
| RANDOM PART OF MODEL |
+-----+

-----
LEVEL 3          TAU-HAT      STD.ERR.      Z-VALUE      PR > |Z|
-----
intcept /intcept      0.07305      0.02849      2.56353      0.01036

-----
LEVEL 2          TAU-HAT      STD.ERR.      Z-VALUE      PR > |Z|
-----
intcept /intcept      0.17706      0.03663      4.83310      0.00000

-----
LEVEL 1          TAU-HAT      STD.ERR.      Z-VALUE      PR > |Z|
-----
intcept /intcept      7.00628      0.19563      35.81370      0.00000

```

There is significant variation in the average estimated total health expenditure at all levels, with the most variation over the participants (level-1), and the least variation over strata (level-3).

An estimate of the level-2 cluster effect, for example, is obtained as

$$\frac{0.17706}{0.07305 + 0.17706 + 7.00628} \times 100\% = 2.41\%$$

indicating that only 2.41% of the total variance explained is at level-2 of the model.

Three-level analysis of simulated data

Unlike real data sets, simulated data sets have the advantage that the true population parameters are known. Consequently, it is possible to evaluate how closely particular estimates approaches these values.

The data

It is assumed that the level-3 units are institutions. Within each of 100 institutions, 10 patients are selected on the basis of their initial achievement in a test of short term memory (Score1) and measurements were repeated over six time intervals for five patients from each institution and over 4 time intervals for the remaining 5. In the table below, (Weight3) shows the level-3 weight calculations based on standardized initial scores.

Interval	Lower	Upper	% Expected	% Selected	Weight3
1	-Inf	-1.00	15.87	10.00	1.587
2	-1.00	-0.70	8.33	10.00	0.833
3	-0.70	-0.20	17.88	10.00	1.788
4	-0.20	0.00	7.93	10.00	0.793
5	0.00	0.30	11.79	10.00	1.179
6	0.30	1.00	22.34	10.00	2.234
7	1.00	1.30	6.19	10.00	0.619
8	1.30	1.80	6.09	10.00	0.609
9	1.80	2.30	2.52	10.00	0.252
10	2.30	Inf	1.07	10.00	0.107

Ten patients were selected from each institution as follows:

- Four from ethnic group 1 with Weight2 = 7.0/4.0
- Three from ethnic group 2 with Weight2 = 2.0/3.0
- Three from ethnic group 3 with Weight2 = 1.0/3.0

The first 10 records of the corresponding dataset in **surveyhlm.psf** in the **TUTORIAL** subfolder are shown below.

	Institut	Patient	Score	Time	Lang1	Lang2	WT3	WT2
1	1.00	1.00	-1.84	0.00	0.00	0.00	1.59	1.75
2	1.00	1.00	-0.89	1.00	0.00	0.00	1.59	1.75
3	1.00	1.00	-1.21	2.00	0.00	0.00	1.59	1.75
4	1.00	1.00	-3.24	3.00	0.00	0.00	1.59	1.75
5	1.00	1.00	-1.16	4.00	0.00	0.00	1.59	1.75
6	1.00	1.00	-1.30	5.00	0.00	0.00	1.59	1.75
7	1.00	2.00	-0.07	0.00	0.00	0.00	1.59	1.75
8	1.00	2.00	2.99	1.00	0.00	0.00	1.59	1.75
9	1.00	2.00	0.92	2.00	0.00	0.00	1.59	1.75
10	1.00	2.00	3.63	3.00	0.00	0.00	1.59	1.75

Note that the data were simulated in such a way that odd-numbered patients have six score measurements at time points 0, 1, 2, 3, 4, 5. The even-numbered patients have only four score measurements.

Linear growth model with random intercepts and slopes and a covariate

The hypothetical model is given by

$$Score_{ijk} = \beta_0 + \beta_1 * Time_{ijk} + \gamma_1 * Lang1_{ijk} + \gamma_2 * Lang2_{ijk} + v_{i0} + Time_{ijk} * v_{i1} + u_{ij0} + Time_{ijk} * u_{ij1} + e_{ijk}$$

where i denotes institution i , ($i = 1, 2, \dots, 100$), ij patient j ($j = 1, 2, \dots, 10$) in institution i and ijk the k -th measurement ($k = 1, 2, \dots, 6$) on patient j in institution i . The outcome variable Score denotes a patient's measurement on some test of interest, Time the time of measurement, and Lang1 and Lang2 are indicator variables indicating a patient's first or home language as being English or another language, β_0 denotes the average expected score, while β_1 indicates the estimated coefficients associated with the time of measurement as represented by the fixed effect Time. The fixed part of the model also includes the predictor variables Lang1 and Lang2. The random part of the model is represented by v_{i0} , u_{ij0} and e_{ijk} , which denote the variation in score over institutions, between patients (or, in other words, over patients nested within institutions) and between measurements at the lowest level of the hierarchy respectively.

The data were simulated under the assumption that

$$\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} 1.0 \\ 0.5 \end{pmatrix}, \quad \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} = \begin{pmatrix} 0.5 \\ -1.0 \end{pmatrix}$$

$$\Phi_2 = Cov(u_{ij0}, u_{ij1}) = \begin{pmatrix} 1 & \\ 0.3 & 0.2 \end{pmatrix}$$

$$\Phi_3 = Cov(v_{i0}, v_{i1}) = \begin{pmatrix} 1 & \\ 0.3 & 0.2 \end{pmatrix}$$

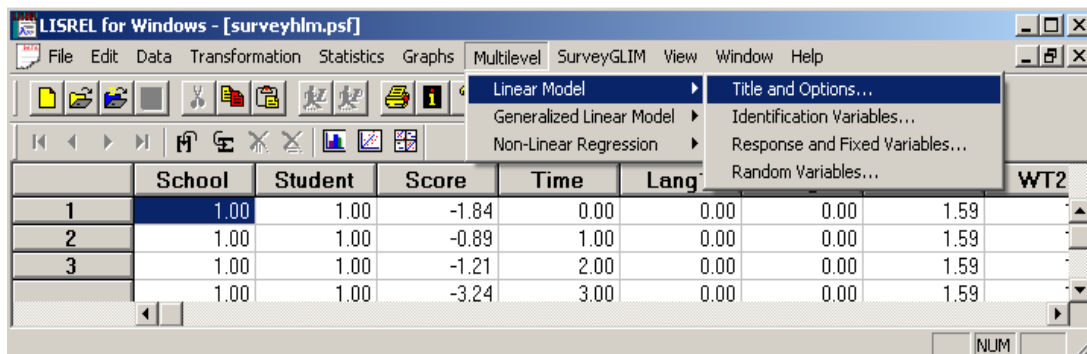
and

$$\sigma^2 = var(e_{ijk}) = 1.0.$$

Fitting the linear growth model with random intercepts and slopes and a covariate

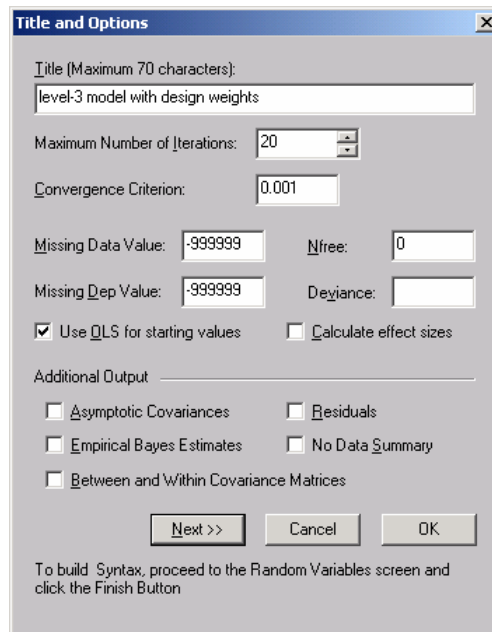
To fit the 3-level model above to the data in **surveyhlm.psf**, we proceed as follows.

- Select the **Open** option on the **File** menu to load the **Open** dialog box.
- Select the **PRELIS Data (*.psf)** option on the **Files of type** drop-down list box.
- Browse for the file **surveyhlm.psf** in the **TUTORIAL** subfolder and select it.
- Click on the **Open** button to open the PSF window for **surveyhlm.psf**.
- Select the **Title and Options** option from the **Linear Model** popup menu on the **Multilevel** menu as shown below

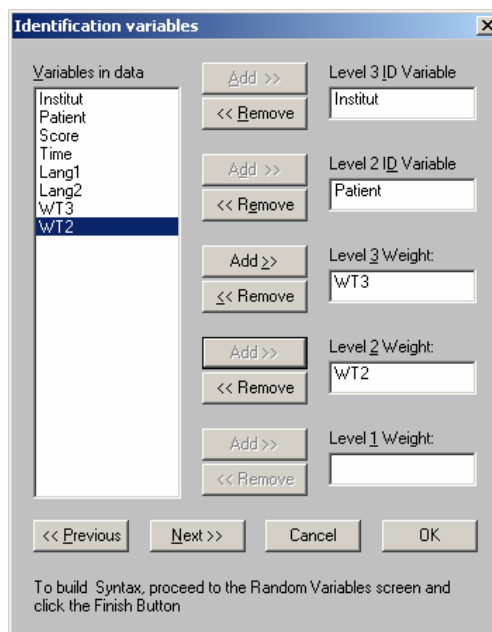


to activate the **Title and Options** dialog box.

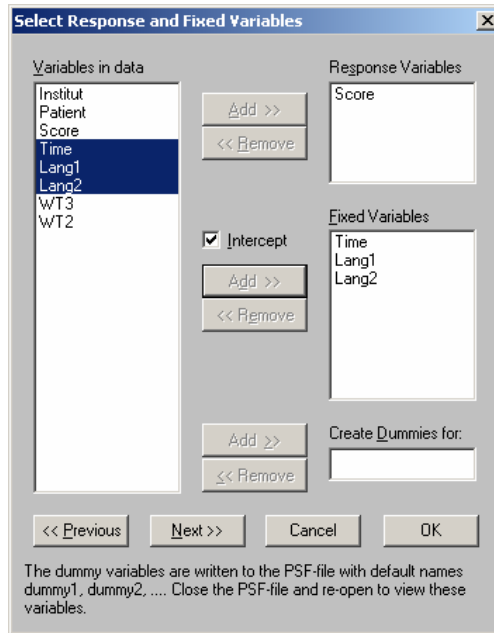
- Enter the title 3-level model with design weights in the **Title** string box to produce the following dialog box.



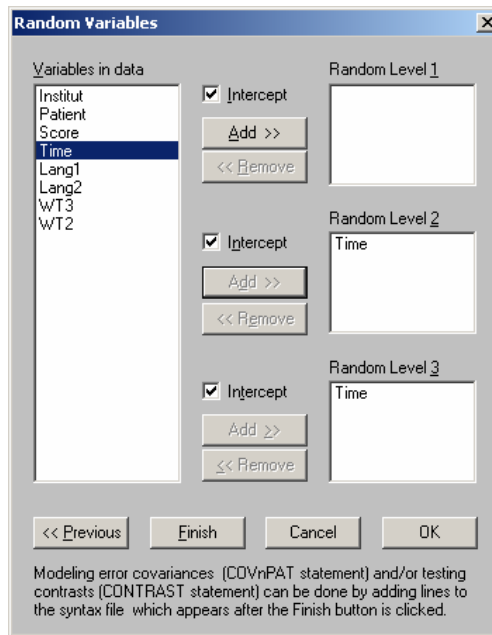
- Click on the **Next** button to go to the **Identification Variables** dialog box.
- Select the variable **Institut** from the **Variables in data** list box and click on the first **Add** button to add the variable in the **Level 3 ID Variable** grid box.
- Select the variable **Patient** from the **Variables in data** list box and click on the second **Add** button to add the variable in the **Level 2 ID Variable** grid box.
- Select the variable **WT3** from the **Variables in data** list box and click on the third **Add** button to add the variable in the **Level 3 Weight** grid box.
- Select the variable **WT2** from the **Variables in data** list box and click on the fourth **Add** button to produce the following dialog box.



- Click on the **Next** button to go to the **Select Response and Fixed Variables** dialog box.
- Select the variable **Score** from the **Variables in data** list box and click on the upper **Add** button to add the variable **Score** to the **Response Variables** list box.
- Select the variables **Time**, **Lang1** and **Lang2** from the **Variables in data** list box and click on the middle **Add** button to obtain the following dialog box.



- Click on the **Next** button to go to the **Random Variables** dialog box.
- Select the variable **Time** from the **Variables in data** list box and click on the middle **Add** button to add the variable **Time** to the **Random Level 2** list box.
- Select the variables **Time** from the **Variables in data** list box and click on the lower **Add** button to obtain the following dialog box.



- o Click on the **Finish** button to generate the following text editor window for **surveyhlm.pr2**.

```

surveyhlm.PR2
OPTIONS OLS=YES CONVERGE=0.001000 MAXITER=10 OUTPUT=STANDARD ;
TITLE=3-level model with design weight ;
SY='C:\Program Files\lisrel88\TUTORIAL\surveyhlm.psf';
ID3=Institut;
ID2=Patient;
WEIGHT3=WT3;
WEIGHT2=WT2;
RESPONSE=Score;
FIXED=intcept Time Lang1 Lang2;
RANDOM1=intcept;
RANDOM2=intcept Time;
RANDOM3=intcept Time;

```

- o Click the **Run PRELIS** button to open the text editor window for the output file **surveyhlm.out**.

Discussion of results

Portions of the output file **meps2.out** are shown below. In the first section of the output file a description of the hierarchical structure is provided as shown below.

```

surveyhlm.OUT
+-----+
| DATA SUMMARY |
+-----+

NUMBER OF LEVEL 3 UNITS :    100
NUMBER OF LEVEL 2 UNITS :   1000
NUMBER OF LEVEL 1 UNITS :   5000

ID3 :    1    2    3    4    5    6    7    8
N2  :   10   10   10   10   10   10   10   10
N1  :   50   50   50   50   50   50   50   50

ID3 :    9   10   11   12   13   14   15   16
N2  :   10   10   10   10   10   10   10   10
N1  :   --   --   --   --   --   --   --   --

```

The data summary above indicate that 5000 observations of 1000 patients from 100 institute are simulated. In addition, a summary of the number of patients nested within each institute is provided. It shows that there are 10 patients (N2: 10) for each of the 100 IDs. And data are available from 50 observations (N1: 50) nested within 10 patients.

The next section of the output file contains the results of the fixed part of the model as shown below.

```

surveyhlm.OUT
|
level-3 model with design weights

ITERATION NUMBER      4

+-----+
| FIXED PART OF MODEL |
+-----+

-----
COEFFICIENTS          BETA-HAT      STD.ERR.      Z-VALUE      PR > |Z|
-----
intcept                0.92865      0.11624       7.98895      0.00000
Time                   0.51411      0.04700      10.93934     0.00000
Lang1                   0.39340      0.10188       3.86158     0.00011
Lang2                  -1.03430     0.12628      -8.19033     0.00000

+-----+
|  -2 LOG-LIKELIHOOD  |
+-----+

DEVIANCE= -2*LOG(LIKELIHOOD) = 20291.45600144901
NUMBER OF FREE PARAMETERS = 11

CHI-SQUARE SCALE FACTOR = 0.68008

```

Recall that the "true" values of the intcept, Time, Lang1 and Lang2 parameters were 1.0, 0.5, 0.5, and -1.0 respectively. To obtain 95% confidence intervals for these estimates, we calculate

$$\text{estimate} \pm 1.96(\text{standard error estimate})$$

and find that the confidence intervals for the estimated intcept, Time, Lang1 and Lang2 parameters are (0.7009; 1.1565), (0.4220; 0.6062), (0.1937; 0.5931) and (-1.2818; -0.7868) respectively. In all four cases, the confidence intervals include the "true" values of corresponding parameter.

Note that a χ^2 scale factor of 0.68008 is reported. This value is used to obtain a corrected χ^2 -statistic for testing one model against another model.

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

RANDOM PART OF MODEL					
LEVEL		TAU-HAT	STD. ERR.	Z-VALUE	PR > Z

LEVEL 3					
	intcept /intcept	0.93174	0.16915	5.50824	0.00000
	Time /intcept	0.25568	0.06072	4.21048	0.00003
	Time /Time	0.17510	0.03408	5.13730	0.00000

LEVEL 2					
	intcept /intcept	0.96301	0.12059	7.98585	0.00000
	Time /intcept	0.36079	0.04117	8.76381	0.00000
	Time /Time	0.20039	0.02207	9.07935	0.00000

LEVEL 1					
	intcept /intcept	1.02326	0.06542	15.64041	0.00000

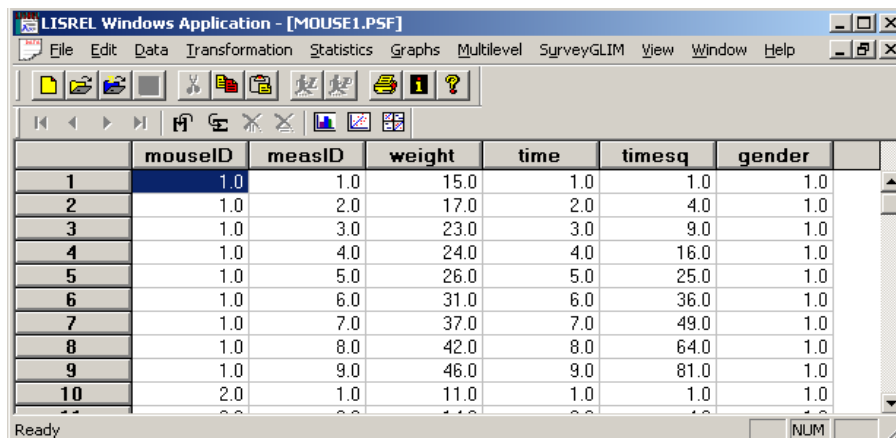
Note that the parameter estimates reported in the output are generally close to the population values which were used to simulate the data. The "true" values for both the level-3 and level-2 variance-covariance components are 1.0, 0.3 and 0.2 respectively.

For the level-3 variance components, 95% confidence intervals can be obtained as shown previously. The confidence intervals corresponding to intcept/intcept, Time/intcept, and Time/Time are (0.6002; 1.2633), (0.1367; 0.3747) and (0.1083; 0.2419) respectively. Again, the "true" values fall within these intervals. This conclusion also holds for confidence intervals for the level-1 and level-2 variance-covariance components, which are calculated in the same way.

Two-level non-linear regression analysis of mice data

The data

The data contain repeated measurements on 82 striped mice and were obtained from the Department of Zoology at the University of Pretoria, South Africa (Du Toit 1979). A number of male and female mice were released in an outdoor enclosure with nest boxes and sufficient food and water. They were allowed to multiply freely. Occurrence of birth was recorded daily and newborn mice were weighed weekly, from the end of the second week after birth until physical maturity was reached. The data set consists of the weights of 42 male and 40 female mice. For male mice, 9 repeated weight measurements are available and for the female mice 8 repeated measurements. The data are provided as the PSF **mouse1.psf** in the **TUTORIAL** subfolder. The first portion of this PSF is shown in the following PSF window.



	mouseID	measID	weight	time	timesq	gender
1	1.0	1.0	15.0	1.0	1.0	1.0
2	1.0	2.0	17.0	2.0	4.0	1.0
3	1.0	3.0	23.0	3.0	9.0	1.0
4	1.0	4.0	24.0	4.0	16.0	1.0
5	1.0	5.0	26.0	5.0	25.0	1.0
6	1.0	6.0	31.0	6.0	36.0	1.0
7	1.0	7.0	37.0	7.0	49.0	1.0
8	1.0	8.0	42.0	8.0	64.0	1.0
9	1.0	9.0	46.0	9.0	81.0	1.0
10	2.0	1.0	11.0	1.0	1.0	1.0

The variables in the data set are:

- mouseID denotes the mouse for which the measurement was obtained.
- measID is the occasion on which the measurement for the particular mouse was made.
- weight denotes the weight (in grams) of the mouse.
- time is the time point at which the measurement was made.
- timesq is the squared value of the time point.
- gender is the gender of the mouse with 1 for male, -1 for female.

Non-linear regression model

The logistic regression model to be fitted to the data in **mouse1.psf** may be expressed as

$$\text{weight}_{ij} = b_{1i} / (1 + \exp(b_{2i} - b_{3i} \cdot \text{time}_{ij})) + e_{ij}$$

where $weight_{ij}$ denotes the j -th weight measurement for mouse i , $time_{ij}$ is the time point for measurement j on mouse i , b_{1i} , b_{2i} and b_{3i} denote unknown coefficients and e_{ij} is the random variation on level-1 of the model. In addition, it is assumed that the 3 regression coefficients depends on the gender of the mouse, i.e.

$$b_{1i} = \beta_1 + \gamma_1 \text{gender}_i + u_{1i}$$

$$b_{2i} = \beta_2 + \gamma_2 \text{gender}_i + u_{2i}$$

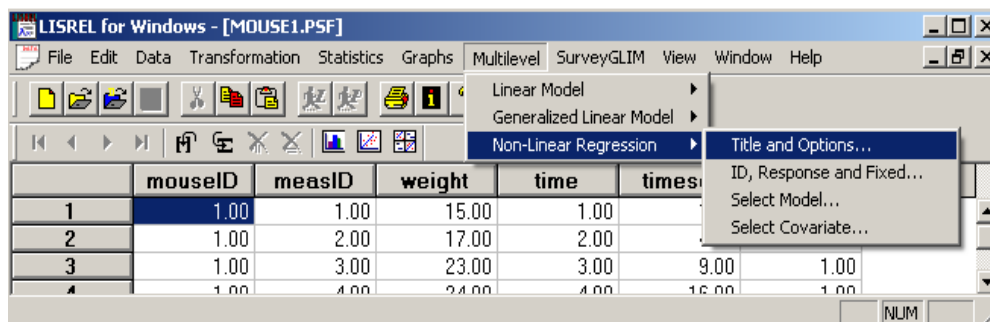
$$b_{3i} = \beta_3 + \gamma_3 \text{gender}_i + u_{3i}$$

where $gender_i$ denotes the gender of mouse i and u_{1i} , u_{2i} and u_{3i} denote the random variations in the three regression coefficients respectively.

Fitting the non-linear regression model

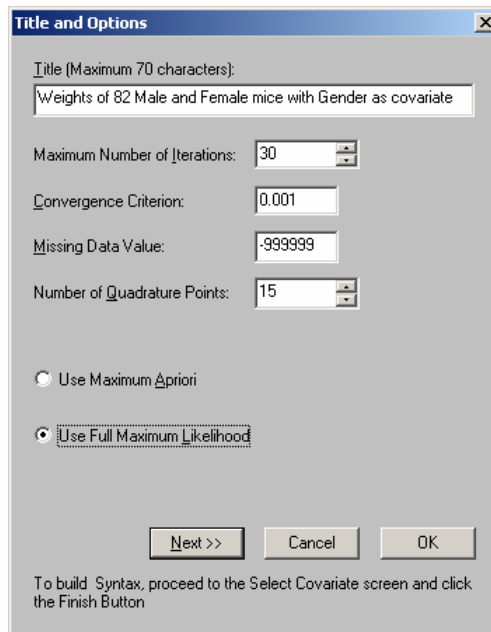
To fit the 2-level non-linear regression model above to the data in **mouse1.psf**, we proceed as follows.

- Select the **Open** option on the **File** menu to load the **Open** dialog box.
- Select the **PRELIS Data (*.psf)** option on the **Files of type** drop-down list box.
- Browse for the file **mouse1.psf** in the **TUTORIAL** subfolder and select it.
- Click on the **Open** button to open the PSF window for **mouse1.psf**.
- Select the **Title and Options** option from the **Non-Linear Regression** popup menu on the **Multilevel** menu as shown below

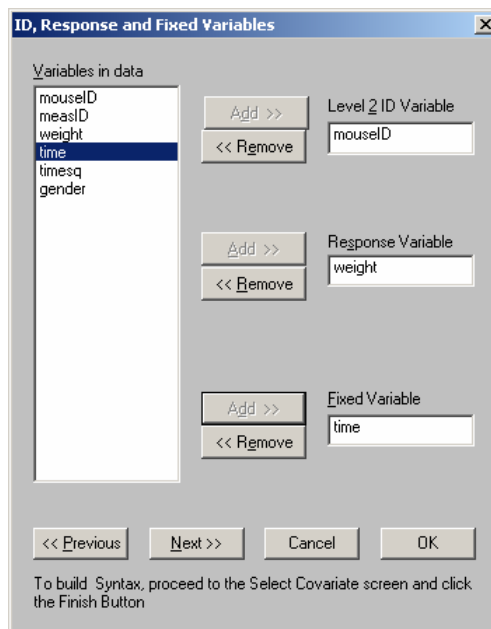


to activate the **Title and Options** dialog box.

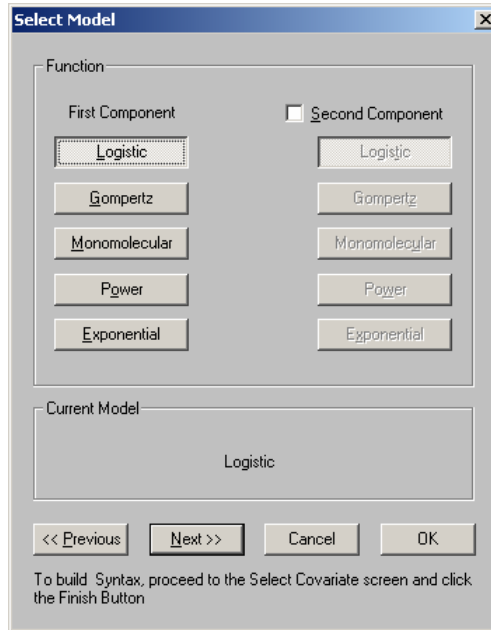
- Enter the title **Weights of 82 Male and Female mice with Gender as covariate** in the **Title** string box.
- Change the **Number of Quadrature Points** to 15.
- Activate the **Use Full Maximum Likelihood** radio button to produce the following dialog box.



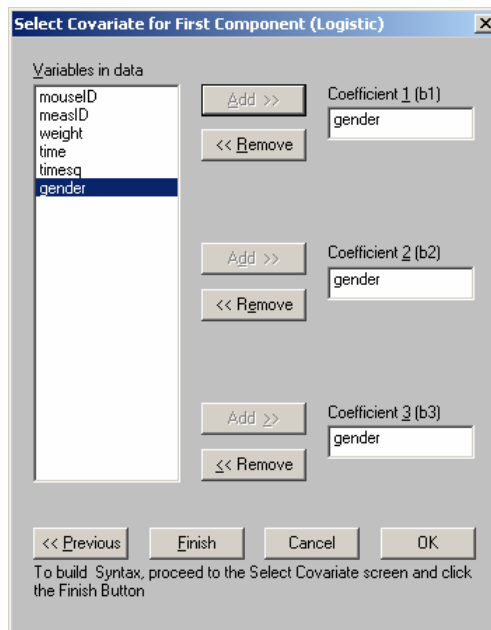
- Click on the **Next** button to go to the **ID, Response and Fixed Variables** dialog box.
- Select the variable mouseID from the **Variables in data** list box and click on the upper **Add** button to add the variable in the **Level 2 ID Variable** grid box.
- Select the variable weight from the **Variables in data** list box and click on the middle **Add** button to add the variable in the **Response Variable** grid box.
- Select the variable time from the **Variables in data** list box and click on the lower **Add** button to produce the following dialog box.



- Click on the **Next** button to obtain the following **Select Model** dialog box.



- Click on the **Next** button to go to the **Select Covariate for First Component (Logistic)** dialog box.
- Select the variable **gender** from the **Variables in data** list box and click on the upper **Add** button to add the variable **gender** to the **Coefficient 1 (b1)** list box.
- Select the variable **gender** from the **Variables in data** list box and click on the middle **Add** button to add the variable **gender** to the **Coefficient 2 (b2)** list box.
- Select the variable **gender** from the **Variables in data** list box and click on the lower **Add** button to obtain the following dialog box.



- Click on the **Finish** button to generate the following text editor window for **mouse1.pr2**.

```

OPTIONS METHOD = ML CONVERGE = 0.001000 MAXITER = 30 QUADPTS = 15;
TITLE = Weights of 82 Male and Female mice with Gender as covariate ;
SY='C:\Program Files\lisrel88\TUTORIAL\MOUSE1.PSF';
ID2 = mouseID;
RESPONSE = weight;
FIXED = time;
MODEL = Logistic;
COVARIATES b1 = gender
            b2 = gender
            b3 = gender;

```

- Click the **Run PRELIS** button to open the text editor window for the output file **mouse1.out**.

Discussion of results

Portions of the output file **meps2.out** are shown below. In the first section of the output file a description of the hierarchical structure is provided as shown below.

```

+-----+
| DATA SUMMARY |
+-----+

NUMBER OF LEVEL 2 UNITS :      82
NUMBER OF LEVEL 1 UNITS :     698

N2 :      1      2      3      4      5      6      7      8
N1 :      9      9      9      9      9      9      9      9

N2 :      9     10     11     12     13     14     15     16
N1 :      9      9      9      9      9      9      9      9

```

The data summary above indicates that a total of 698 observations were obtained from 82 mice. In addition, a summary of the number of measurements per mouse is provided. For example, for mouse number 1 (N2: 1), there are 9 observations (N1: 9) available.

The output for the maximum likelihood estimates of the model is given next as shown in the following image.

MOUSE1.OUT

Maximum Likelihood Solution

Coefficients	Beta	Std.Err.	Z-value	P > z
b1	45.15479	0.87773	51.44478	0.00000
b2	1.57372	0.03242	48.53536	0.00000
b3	0.51629	0.01321	39.09235	0.00000

Covariate Names	Gamma	Std.Err.	Z-value	P > z
gender	2.68072	0.87772	3.05419	0.00226
gender	0.10327	0.03242	3.18588	0.00144
gender	0.04006	0.01320	3.03407	0.00241

Variance estimate	Level 1	Std.Err.	Z-value	P > z
Sigma**2	2.73755	0.10463	26.16436	0.00000

Covariances	Level 2	Std.Err.	Z-value	P > z
u1, u1	100.08828	12.77498	7.83471	0.00000
u2, u1	2.10569	0.38210	5.51082	0.00000
u2, u2	0.13512	0.01840	7.34431	0.00000
u3, u1	-0.60022	0.15246	-3.93681	0.00008
u3, u2	0.00424	0.00526	0.80723	0.41953
u3, u3	0.01881	0.00287	6.55261	0.00000

The results above indicate the three regression coefficients are statistically significant if a 1% level of significance is used. In addition, it is also evident that these three coefficients vary significantly across gender.

The total estimated variance on level-2 follows as

$$\begin{aligned}
 & \hat{\sigma}^2(u_1 + u_2 + \text{time} \times u_3) \\
 &= \hat{\sigma}^2(u_1) + \hat{\sigma}^2(u_2) + \text{time}^2 \times \hat{\sigma}^2(u_3) + 2 \times \hat{\sigma}(u_1, u_2) \\
 & \quad + 2 \times \text{time} \times \hat{\sigma}(u_1, u_3) + 2 \times \text{time} \times \hat{\sigma}(u_2, u_3) \\
 &= 100.08828 + 0.13512 + 0.01881 \times \text{time}^2 + 2 \times 2.10569 \\
 & \quad - 2 \times 0.60022 \times \text{time} + 2 \times 0.00424 \times \text{time} \\
 &= 104.43478 - 1.19196 \times \text{time} + 0.01881 \times \text{time}^2
 \end{aligned}$$

We can thus express the total estimated variation on level-2 as a nonlinear function of the variable time.

References

Du Toit, S.H.C. (1979). *The analysis of growth curves*. Unpublished PhD thesis. Pretoria: University of South Africa.

Du Toit, M. & du Toit, S.H.C. (2001). *Interactive LISREL user's guide*. Lincolnwood, IL: Scientific Software International.

Kanfer, R., and Ackerman, P.L. (1989). Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition. *Journal of Applied Psychology*. Monograph, **74**, 657-690.

Jöreskog, K.G. & Sörbom, D. (2005). *LISREL for Windows [Computer Software]*. Lincolnwood, IL: Scientific Software International, Inc.

Jöreskog, K.G., Sörbom, D., du Toit, S.H.C. & du Toit, M. (2001). *LISREL 8: New Statistical Features*. Lincolnwood, IL: Scientific Software International, Inc.