Scientific Software International

Multilevel Generalized Linear Modeling Guide

Du Toit, S.H.C., & Du Toit, M. (2021). *Multilevel Generalized Linear Modeling Guide*. Chapel Hill, NC: Scientific Software International, Inc.

**Please download the examples from our website at
https://ssicentral.com/index.php/products/lisrel/lisrel-examples and unzip them into a convenient folder location. The examples shown in the manual uses 'C:\LISREL Examples' and you are more than welcome to use the same or a different location. Please note that the actual location may be different on your machine.**

TABLE OF CONTENTS

1.   Introduction

There has been a growing interest in recent years in fitting models to data collected from longitudinal surveys that use complex sample designs. This interest reflects expansion in requirements by policy makers and researchers for in-depth studies of social processes over time. Traditionally, the analyses of complex survey samples have been carried out using specialized software packages. More recently, a number of statistical analyses packages, for example SAS and SPSS, have implemented procedures to handle complex survey data appropriately in the case of regression models with continuous and categorical outcome variables. In this guide we describe techniques currently implemented in LISREL for analyzing complex surveys data. Research on the longitudinal analysis of complex survey data with LISREL was supported by SBIR grant R43 AA014999-01 from NIAAA to Scientific Software International.

A common theme in substance use research is that data are usually obtained from a multi-stage or so-called complex sampling data. A complex sampling design typically entails stratification, often on the basis of geography; defining meaningful clusters of population elements (PSUs); and one or more stages of subsampling within each PSU. While a complex sample has the advantages of being more economical and practical, guarantees a better representative sample of the population, and does not require a complete sampling frame of the population elements, it is generally less efficient than simple random sampling.

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.

The collection of models called Generalized Linear Models (GLIMs) have become important, and practical, statistical tools. The basic idea of GLIMs is an adaption of standard regression to quite different kinds of data. The variables may be dichotomous, ordinal (as with a 5-point Likert scale), counts (number of arrest records), or nominal. The motivation is to tailor the regression relationship connecting the outcome to relevant independent variables so that it is appropriate to the properties of the dependent variable. The statistical theory and methods for fitting Generalized Linear Models (GLIMs) to survey data is described in the *Generalized Linear Modeling Guide*. Researchers from the social and economic sciences are often applying these methods to multilevel data and consequently, inappropriate results are obtained. We describe the LISREL statistical module for the analysis of multilevel data with design weights. Two estimation methods, MAP (maximization of the posterior distribution) and ADAP (adaptive quadrature) for fitting generalized linear models to multilevel data are available. The LISREL module allows for a wide variety of sampling distributions and link functions.

2. Generalized multilevel models

2.1 Background

In this guide, we illustrate and outline the statistical theory and methods that are implemented in the multilevel generalized linear model (MGLIM) module of LISREL. Section 2.2 reviews the options and dialog boxes of the **Multilevel**, **Generalized Linear Model** menu on the LISREL System File (LSF) window of LISREL. MGLIM syntax files are reviewed in Section 2.3. Illustrative examples are provided in Section 2.4. In Section 2.5, the results of the MGLIM module are assessed by simulation study. The MGLIM statistical theory for complex survey data is outlined in Section 2.6.

2.2 Graphical User Interface

2.2.1 The Multilevel, Generalized Linear Model menu

The **Multilevel** menu provides you access to three options: **Linear Model**, **Generalized Linear Model** and **Non-Linear Regression**. In this guide, the **Generalized Linear Model** option is introduced. As shown below, the **Generalized Linear Model** option provides a sequence of five dialog boxes that can be used to create a MGLIM syntax file interactively. It is located on the LSF (LISREL System File) window of LISREL which is used to display, manipulate and process raw data. In other words, you must create a LSF file and open it in a LSF window before syntax can be generated interactively. To illustrate this, the LSF window for the file **ASPART.lsf** (**Multilevel Generalized Linear Model examples** folder) with the **Multilevel**, **Generalized Linear Model** menu expanded is shown below.
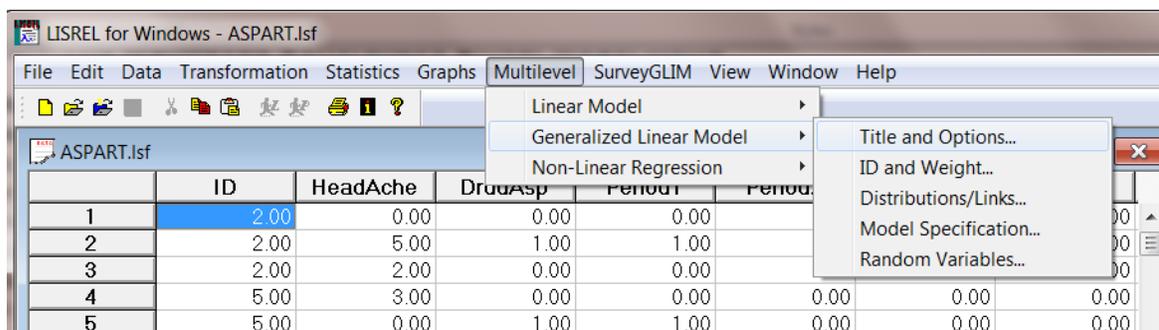
The typical next step would be to click on the **Title and Options** option to load the **Title and Options** dialog box (see Section 2.2.2). However, you can directly click on the **ID and Weight**, **Distributions/Links**, **Model Specification** or **Random Variables** option to go to the **ID and Weight** (see Section 2.2.3), the **Distributions and Links** (see Section 2.2.4), the **Model Specification** (see Section 2.2.5), or the **Random Variables** dialog box (see Section 2.2.6) respectively.

2.2.2 The Title and Options dialog box

The **Title and Options** dialog box allows you to specify a title and the options of the MGLIM analysis interactively and is accessed by selecting the **Title and Options** option on the **Multilevel**, **Generalized Linear Model** menu. This selection loads the following **Title and Options** dialog box.
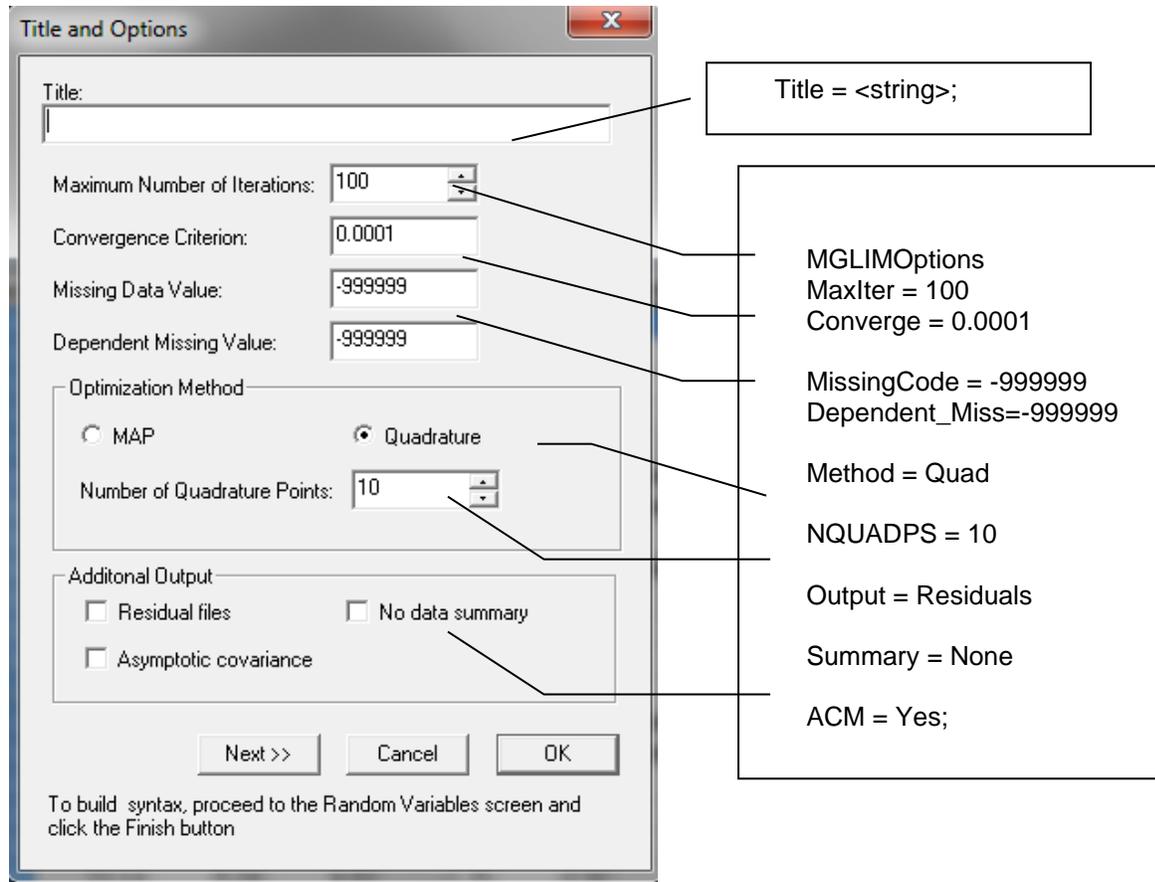
Note that the **Title and Options** dialog box corresponds with the Title and MGLIMOptions commands as indicated on the image above.

If desired, you can enter a descriptive title in the **Title** string field. If the raw data include missing values with a global missing value other than -999999, you need to enter the global missing value in the **Missing Data Value** number field. In order to specify missing data values on the outcome variable, the **Dependent Missing Value** field should be used.

Since the MGLIM estimation equations do not have a closed form solution, MGLIM uses an iterative algorithm to estimate the parameters of the MGLIM. In this regard, the MAP algorithm and the Quadrature algorithm are available in the **Optimization Method** option. The default algorithm is MAP; click the **Quadrature** radio button to choose that algorithm instead. You can then enter the number of quadrature points in the **Number of Quadrature Points** field if the default of 10 is not appropriate. Enter the appropriate convergence criterion in the **Convergence Criterion** number field if the default value of 0.0001 is not to be used.

In the **Additional Ouput** section, three options are available upon request. If the **Residual Files** option is selected, the residuals will be saved to a file with extensions **.ba2** and **ba3** for level-2 and level-3 residuals respectively. These residual files will have the same name than the output file except that **.out** is replaced with **.ba2** and **.ba3**. When the **No data summary** check box is checked, the data summary will not be written to the output file. The **Asymptotic covariance** option provides the user the option to save the asymptotic covariance file for the fixed and random part of the model to additional files with extension **_fixed.acm** and **_random.acm**. Once you are done with the **Title and Options** dialog box, click on the **Next** button to go to the **Distributions and Links** dialog box.

2.2.3 The ID and Weight Variables dialog box

The **Level-2 ID variable** and **Level-3 ID variable** boxes on the **ID and Weight Variables** dialog box is used to select the variables in the LISREL data file (*.**lsf**) that identify the various levels of the hierarchy. The **Weight** variable box is used to define the weight variable to be used in the analysis. The image below shows the default settings for this dialog box and, to the right, the corresponding syntax commands. See the alphabetical list of syntax commands for details on the options available other than the default settings: IDn command (section 2.3.6) and WEIGHT command (section 2.3.15).

The **Next** button provides access to the **Distributions and Links** dialog box.

2.2.4 The Distributions and Links dialog box

The **Distributions and Links** dialog box allows you to specify the sampling distribution and the link function of the MGLIM interactively. A summary of the combinations of sampling distributions and link functions that are available in this module is listed in Table 1.

The **Distributions and Links** dialog box is, as shown below, usually accessed by clicking on the **Next** button of the **ID and Weight Variables** dialog box. It can also be accessed by selecting the **Distributions / Links** option from the **Multilevel**, **Generalized Linear Model** menu.

Table 1: Sampling Distribution and Link Functions

Link / Distribution	CLL	Identity	Log	Logit	OCLL	OLogit	OProbit	Power	Probit
Bernoulli	x			x					x
Binomial	x			x					x
Gamma			x					x	
Inverse Gaussian			x					x	
Multinomial				x	x	x	x		
Negative binomial			x						
Normal		x							
Poisson			x						

Note that the **Distributions and Links** dialog box corresponds with the Distribution, Link, Intercept, Dispersion and Scale commands as indicated on the image below.

Use the **Distribution type** and the **Link function** drop-down list boxes to select the distribution and link function for your MGLIM. If an intercept for the mean model of the MGLIM is not required, you should activate the **No** radio button.
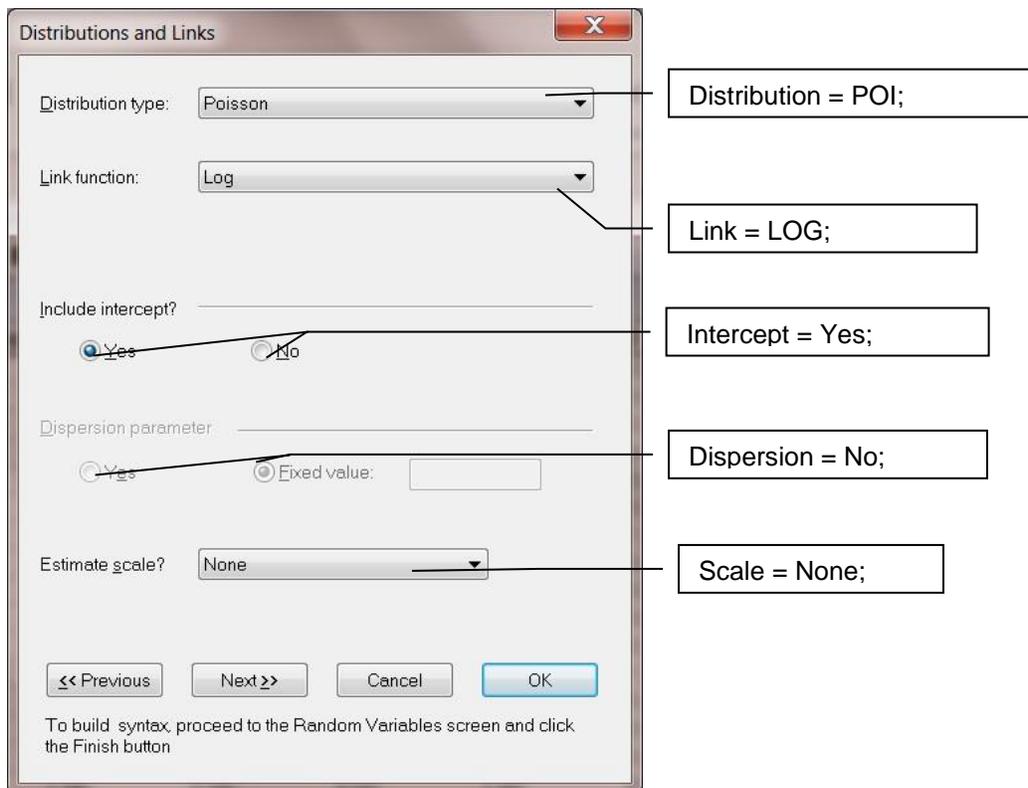
Some MGLIMs include dispersion or scale parameters. These MGLIMs are listed in Table 2. If a scale parameter is desired, you can select the appropriate scale parameter from the **Estimate scale?** drop-down list box. In the case of a dispersion parameter, you can fix its value by activating the **Fixed value** radio button.

Once the **Distributions and Links** dialog box has been completed, the **Next** button is clicked to go to the **Dependent and Independent Variables** dialog box.

Table 2: Scale and Dispersion Parameters

Parameters / Distribution	Scale	Dispersion	Maximum Likelihood	Pearson	Deviance
Binomial	x			x	x
Gamma	x		x	x	x
Inverse Gaussian	x		x	x	x
Negative binomial		x			
Normal	x			x	x
Poisson	x			x	x

2.2.5 The Dependent and Independent Variables dialog box

Access to this dialog box is obtained by clicking on the **Next** button of the **Distributions and Links** dialog box or by selecting the **Model Specification** option from the **Multilevel**, **Generalized Linear Model** menu. Examples of the **Dependent and Independent Variables** dialog box are shown above.

The model for the means of the outcome variable is a function of a set of covariates. You specify the outcome variable by first selecting it from the **Variables in data** list box and then by clicking on the **Add** button of the **Dependent variable** section. The covariates of the model can either be categorical or continuous variables. Dummy variables are also regarded as continuous variables. Categorical covariates are specified by first selecting the covariates from the **Variables in data** list box and then by clicking on the **Categorical** button. In a similar fashion, the **Continuous** button is used to specify the continuous covariates and dummy variables of the model.

The **Dependent and Independent Variables** dialog box allows you to specify the model for the means of the outcome variable and, if applicable, a frequency variable. In the case of a multinomial model with ordinal link functions, the dialog box contains the **Event Variable** field. This field corresponds to the optional CENSOR_VAR command. Note that the **Event Variable** cannot have the same name as the dependent and independent variables selected for inclusion in the model.

The **Dependent and Independent Variables** dialog box for a multinomial model with ordinal link function corresponds with the DEPVAR, COVARS and CENSOR_VAR commands as shown on the image below.
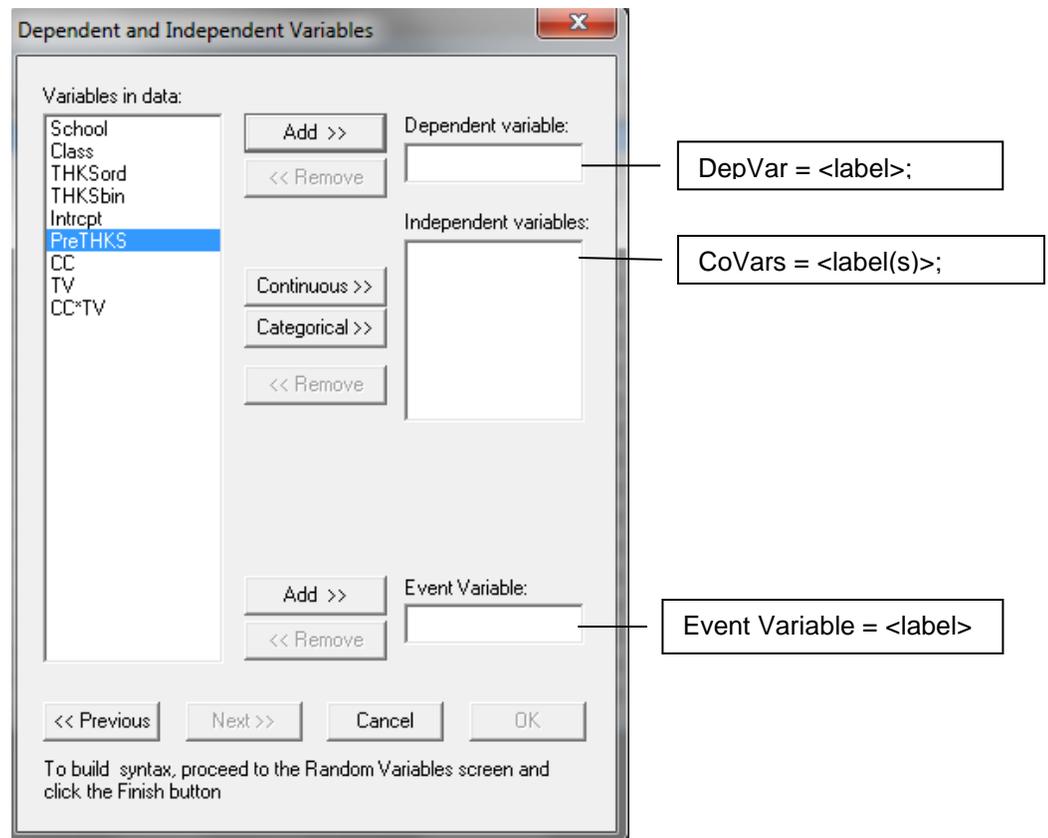
MGLIM can process models for count outcomes with offset variable. Specify the offset variable by first selecting it from the **Variables in data** list box and then by clicking on the **Add** button of the **Offset variable** section. This option is only available if a Poisson model is selected. This field corresponds to the optional OFFSET command.

The **Dependent and Independent Variables** dialog box for a Poisson model corresponds with the DEPVAR, COVARS and OFFSET commands as shown on the image below.
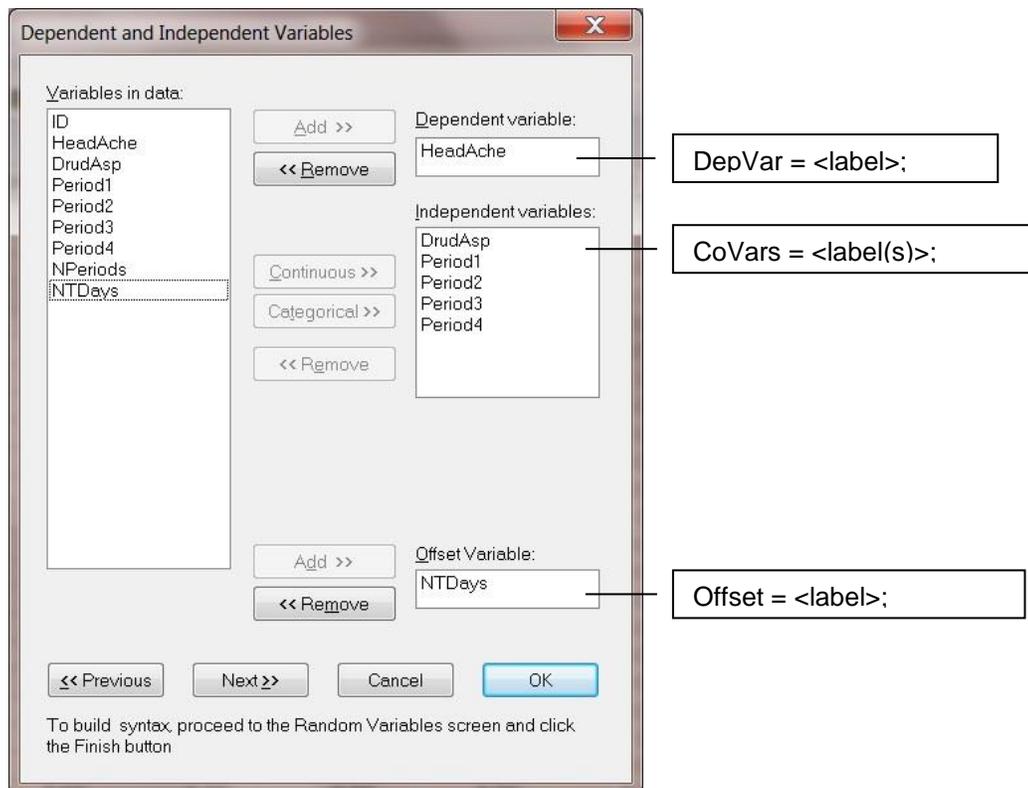
Once the variables have been selected, click the **Next** button to load the **Random Variables** dialog box.

2.2.6 The Random Variables dialog box

The **Random Variables** dialog box is used to select the variables for which coefficients are assumed to be random from the LISREL data file (*.**lsf**). Default settings for this dialog box are shown in the image below. To the right, the corresponding syntax commands are given. See the alphabetical list of syntax commands for details on the options available other than the default settings: RANDOMn command (section 2.3.11). Note that the number of interactions cannot exceed the toatl number of predictors selected. The default number of interactions is 0.

Random Variables

Variables in data

School
Class
THKSord
THKSbin
Intrcpt
PreTHKS
CC
TV
CC*TV

☑ Intercept

Add >>

<< Remove

Random Level 2

☑ Intercept

Add >>

<< Remove

Random Level 3

Number of interactions: 0

<< Previous   Finish   Cancel   OK

To build syntax, click the Finish button.

RANDOM2 = <names of variables random on level-2 of the model>;

RANDOM3 = <names of variables random on level-3 of the model>;

Interactions=<number>;
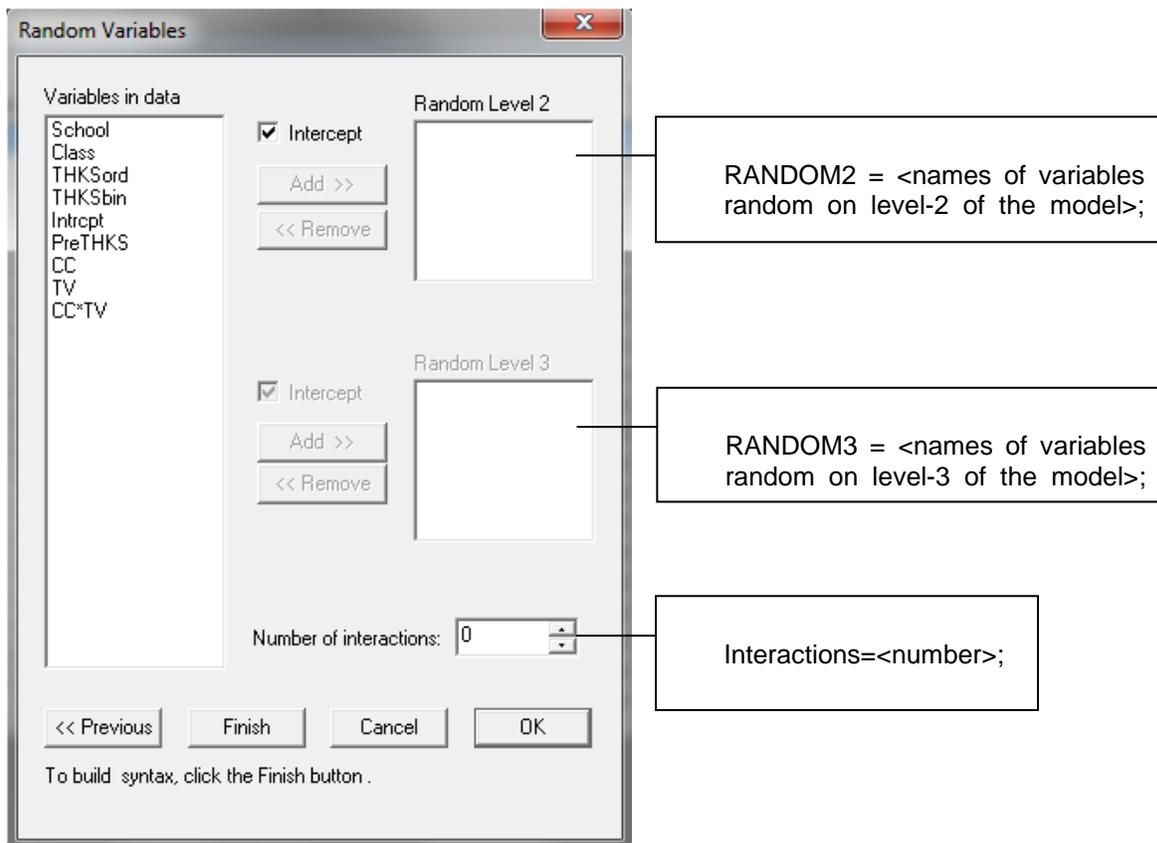
Once all the options are set as desired, click the **Finish** button to generate the syntax.

2.3   Syntax

2.3.1 The structure of the syntax file

The syntax file, which is generated by the interface of the MGLIM module, can also be prepared by using the LISREL text editor or any other text editor such as Notepad and WordPad. The structure of the syntax file follows.

MGLIMOPTIONS <options>;	Required
TITLE = <string>;	Optional
SY = '<filename>';	Required
IDn = <label>; name of variable identifying level n units	Required
DEPENDENT_MISS = <value>;	Optional
WEIGHT = <label>;	Optional
DISTRIBUTION = <name>;	Required
LINK = <function>;	Required
INTERCEPT = <option>;	Required
DISPERSION = <option>;	Optional
SCALE = <type>;	Required
DEPVAR = <label>;	Required
COVARS = <label(s)>;	Required
OFFSET = <label>;	Optional

```
RANDOMn = <label>; names of variables included as random effects
                 on level n of the model                              Required
CENSOR_VAR = <name>;                                                  Optional
INTERACTIONS=<value>;                                                 Optional
```

where <label> denotes a case sensitive variable name used in the raw data file, <filename> denotes a complete name (including the drive and folder names) of a file, <option> is either Yes or No, <type> is one of None, Pearson, Deviance or ML (see Section 2.3.5), <name> is one of BER, BIN, GAM, INVG MUL, NBIN, NOR or POI (see Section 2.3.6) and <function> is one of CLL, IDEN, LOG, LOGIT, OLOGIT, OCLL, OPROBIT, POW[n] and PROBIT (see Section 2.3.8). <options> denotes a list of options for the analysis, each of which has the following syntax:

<keyword> = <selection>

where <keyword> is one or more of ACM, CONVERGE, MAXITER, METHOD, MISSINGCODE, NQUADPTS, OUTPUT, REFCAT or SUMMARY and <selection> denotes a number, an option or a name (see Section 2.3.9). In many applications, optional commands and keywords can be left out if there are program default values available.

The TITLE, WEIGHT, DISPERSION, DEPENDENT_MISS, CENSOR_VAR, INTERACTIONS and OFFSET commands are **optional** commands while the other ten commands are all **required**. The MGLIMOPTIONS 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 seventeen MGLIM commands are discussed separately in alphabetical order.

2.3.2   COVARS command

The purpose of the COVARS command is to specify the covariates of the model for the means of the outcome variable and it is a **required** command. The COVARS command corresponds with the **Independent variables** section on the **Dependent and Independent Variables** dialog box (see Section 2.5).

Syntax

COVARS = <label(s)>;

where <label(s)> denotes the case sensitive label(s) of the covariates of the model. In the case of a categorical variable, the label should be augmented with a $ symbol. Dummy variables are regarded as continuous variables. Consequently, dummy variable labels are not augmented with a $ symbol.

Suppose that the covariates of the model consist of a dummy variable, sex, a categorical variable, edu, and a continuous variable, age. For this example, the corresponding COVARS command is given by

    COVARS = sex edu$ age;

2.3.3 CENSOR_VAR command

The purpose of the CENSOR_VAR command is to specify the frequency variable in the case of multinomial models with ordinal link functions and it is an **optional** command. The CENSOR_VAR command corresponds with the **Event Variable** section on the **Dependent and Independent Variables** dialog box (see Section 2.5). Note that the **Event Variable** cannot have the same name as the dependent and independent variables selected for inclusion in the model.

Syntax

    CENSOR_VAR = <name>;

where <name)> denotes the case sensitive label(s) of the offset variable.

2.3.4 DEPENDENT_MISS command

The DEPENDENT_MISS command is used to specify the value of missing values, if any, on the outcome variable of the model and it is an **optional** command. It corresponds with the **Dependent_Miss** section on the **Dependent and Independent Variables** dialog box (see Section 2.5).

Syntax

    DEPENDENT_MISS = <value>;

where <value> denotes the value of the missing data indicator on the outcome variable of the model. By default, the value is set to – 999999.

2.3.5 DEPVAR command

The DEPVAR command is used to specify the outcome variable of the model and it is a **required** command. It corresponds with the **Dependent variable** section on the **Dependent and Independent Variables** dialog box (see Section 2.5).

Syntax

    DEPVAR = <label>;

where <label> denotes the label of the outcome variable of the model. Note that variable names are case sensitive.

Example

Suppose that the variable, depr, is the outcome variable to be used. In this case, the corresponding DEPVAR command would be

    DEPVAR = depr;

2.3.6   DISPERSION command

The Negative Binomial sampling distribution, for example, has a dispersion parameter. This parameter is specified by using the DISPERSION command. Since not all sampling distributions involve a dispersion parameter, the command is **optional** with default of no dispersion to be estimated. The DISPERSION command corresponds with the **Estimate dispersion?** section on the **Distributions and Links** dialog box (see Section 2.4).

Syntax

    DISPERSION = <option>;

where <option> is either Yes or No.

Default

    DISPERSION = No;

2.3.7   DISTRIBUTION command

Each MGLIM involves the sampling distribution of the outcome variable. The sampling distribution is specified by means of the DISTRIBUTION command, which is **optional**. The DISTRIBUTION command corresponds with the **Distribution type** drop-down list box on the **Distributions and Links** dialog box (see Section 2.4) as shown below.

Syntax

    DISTRIBUTION = <name>;

where <name> is one of BER (Bernoulli), BIN (Binomial), GAM (Gamma), INVG (Inverse Gaussian), MUL (Multinomial), NBIN (Negative Binomial), NOR (Normal) or POI (Poisson).

Default

    DISTRIBUTION = MUL;
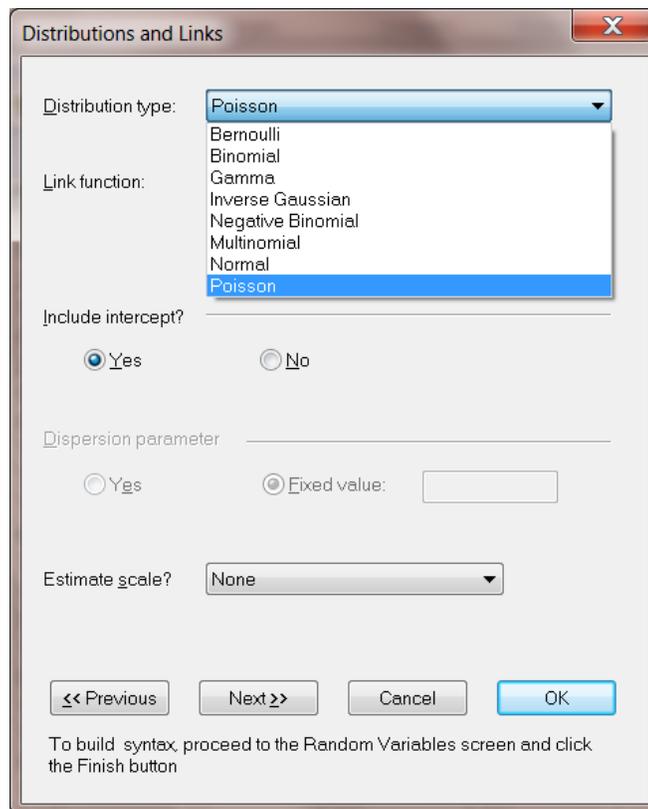
2.3.8 IDn command

The ID command(s) are used to indicate the variable(s) identifying the units on the different levels of the hierarchy. ID command(s) are **required** command(s).

If the model specified is a 2-level model, the command ID2 is required. Likewise, if a level-3 model is to be considered, the ID2 and ID3 commands are required in the syntax file.

Variables listed in the ID commands must be included in the data file (*.**lsf** file). Variable names are case sensitive, therefore the spelling and case in which they are given need to correspond to that given in the spreadsheet.

Syntax

IDn = <variable name identifying level-n units>;

Example

Suppose the raw data file contains information on the test scores, age and gender of pupils belonging to classes within schools, and the variables school, class, pupil, age, gender and score are contained in the data file. The following ID commands may be used to identify the levels of the hierarchical structure:

```
        ID3 = school;
        ID2 = class;
```

The **Identification Variables** dialog box shown below shows the settings needed to obtain these commands.
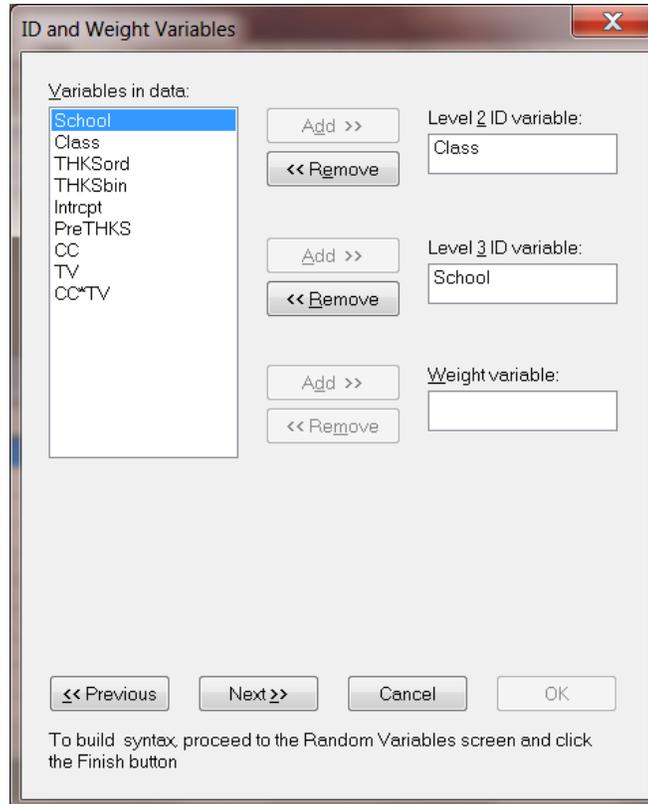
2.3.9 INTERACTIONS command

The purpose of the INTERACTIONS command is to allow specification of the number of interactions. It is an **optional** command. The INTERACTIONS command corresponds with the **Number of Interactions** section on the **Random Variables** dialog box. The number of interactions cannot exceed the number of selected predictors.

Syntax

```
        INTERACTIONS = <number>;
```

Default

```
        INTERACTIONS=0;
```

2.3.10　INTERCEPT command

Many MGLIMs can either include or exclude an intercept parameter for the model for the means of the outcome variable. The purpose of the INTERCEPT command is to allow you to either include or exclude an intercept parameter and it is an **optional** command. The INTERCEPT command corresponds with the **Include intercept?** section on the **Distributions/Links** dialog box (see Section 2.3).

Syntax

INTERCEPT = <option>;

where <option> is either Yes or No.

Default

INTERCEPT = Yes;

2.3.11　LINK command

The link function of a MGLIM describes the relationships between the means of the outcome variable and the means of the corresponding linear model. The LINK command is used to specify the link of the MGLIM, and corresponds with the **Link function** drop-down list box on the **Distributions/Links** dialog box as shown below. It is an **optional** command.

Syntax

LINK = <name>;

where <name> is one of CLL (complementary log-log), IDEN (identity), LOG (log), LOGIT (logit), OCLL (proportional hazards), OLOGIT (cumulative logit), OPROBIT (cumulative probit), and PROBIT (probit).

Default

LINK = Iden;

2.3.12　MGLIMOPTIONS command

The purpose of the MGLIMOPTIONS command is to select the iterative algorithm to be used and to specify options for the selected iterative algorithm. In addition, it is used to specify a global missing data value and the output to be generated. Finally, it allows you to specify the variance adjustment proposed by Morel (1989) if the estimated asymptotic covariance matrix of the parameter estimators is not positive definite. The MGLIMOPTIONS command must always be the first command and is a **required** command. It corresponds with the **Title and Options** dialog box (see Section 2.2).

MGLIMOPTIONS <options>;

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

<keyword> = <selection>

where **<keyword>** is one or more of ACM, CONVERGE, MAXITER, METHOD, MISSINGCODE, NQUADPTS, OUTPUT, REFCAT or SUMMARY and **<selection>** refers to a name, a number or an option.

2.3.12.1 ACM keyword

The ACM keyword is used to print the large-sample covariance matrices of the estimated parameters in the fixed part and random part of the model. This keyword is controlled from the **Title and Options** dialog box.

Standard errors of the estimated parameters are equal to the square roots of the diagonal elements. The non-duplicated elements of these asymptotic covariance matrices are written to external files with the following default names:

<Outputfilename>_fixed.acm
<Outputfilename>_random.acm

If the output file name is, for example, **aspart.out**, then the large-sample covariance matrices are saved to the files **aspart_fixed.acm** and **aspart_random.acm**.

Syntax

ACM = <option>

where <option> is either Yes or No.

Default

ACM = No

No asymptotic covariance matrices will be printed by default.

Example

In the MGLIMOPTIONS command below, the ACM keyword is used to request the printing of the asymptotic covariance matrices at convergence. A convergence criterion of 0.0001 is set as the requirement for convergence, and 30 iterations is indicated as the maximum number of iterations to be performed.

MGLIMOPTIONS MAXITER = 30 CONV = 0.0001 ACM = Yes;

2.3.12.2   CONVERGE keyword

The tolerance limit of the convergence criterion of the selected iterative algorithm is specified by using the CONVERGE keyword which is an **optional** keyword. The CONVERGE keyword corresponds with the **Convergence Criterion** section on the **Title and Options** dialog box (see Section 2.5.2.2).

Syntax

      CONVERGE = <number>

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

Default

      CONVERGE = 0.0001

2.3.12.3   MAXITER keyword

You can control the maximum number of iterations of the selected iterative algorithm by means of the MAXITER keyword which is an **optional** keyword. The MAXITER keyword corresponds with the **Maximum Number of Iterations** section on the **Title and Options** dialog box (see Section 2.5.2.2).

Syntax

      MAXITER = <number>

where $<number>$ denotes a positive integer.

Default

      MAXITER = 100

2.3.12.4   METHOD keyword

MGLIM implements the MAP and quadrature iterative algorithms to obtain the estimates and standard error estimates of the MGLIM parameters. The METHOD keyword is used to select one of these algorithms and it is an **optional** keyword. It corresponds with the **Optimization Method** section on the **Title and Options** dialog box (see Section 2.2).

Syntax

      METHOD = <method>

where <method> is either MAP or QUAD.

Default

      METHOD = MAP

2.3.12.5   MISSINGCODE keyword

Raw data often include missing values. MGLIM uses list-wise deletion for handling data with missing values if you specify a global missing value by means of the MISSINGCODE option, which is **optional**. The MISSINGCODE keyword corresponds with the **Missing Data Value** section on the **Title and Options** dialog box (see Section 2.2).

Syntax

     MISSINGCODE = <number>

where <number> denotes a real number.

Default

     MISSINGCODE = -999999

2.3.12.6   NQUADPTS keyword

The NQUADPTS keyword is used to define the number of quadrature points (per random-effect dimension) to use in the evaluation of the log-likelihood function and derivatives using numerical integration. It is usually between 10 and 20 for 1 random effect and 5 to 10 for 2 or 3 effects.

Syntax

     NQUADPTS = <number>

where number is a positive integer.

Default

     NQUADPTS = 10

2.3.12.7   OUTPUT option

MGLIM can write the residuals of the MGLIM analysis to separate CSV files. The OUTPUT keyword is used to request this file and is an **optional** keyword. The OUTPUT keyword corresponds with the **Residual files** option in the **Additional Output** section on the **Title and Options** dialog box (see Section 2.2.2).

Syntax

     OUTPUT = Residuals

When residual file is requested, MGLIM will save the file to a separate CSV file, which has the same name as the LSF, except that **.lsf** is replaced with **_RES.CSV**.

Default

No residual file is generated.

2.3.12.8 REFCAT keyword

The REFCAT is an option to select whether the first or last category of the outcome should be used as the reference category. It is to be used with nominal outcomes only.

Syntax

REFCAT = <keyword>

The keyword is either LAST or FIRST.

Default

REFCAT = Last

2.3.12.9 SUMMARY keyword

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

Syntax

SUMMARY = <option>

where <option> is either Yes or No.

Default

SUMMARY = Yes;

the summary table containing sample sizes of units within the various levels of the hierarchy is printed.

2.3.12.10 MGLIMOPTIONS example

Suppose that the quadrature algorithm with a 10 quadrature points and a convergence criterion tolerance limit of 0.0001 with printed details is required. Suppose the residual file and asymptotic covariances files are required, no data summary is printed and the global missing value for the raw data is -9. For this example, the GLIMOPTIONS command is given by

MGlimOptions Converge=0.0001 MaxIter=100 MissingCode=-9   Method=Quad NQUADPTS=10
Output=Residuals Summary=None ACM=Yes ;

This MGLIMOPTIONS command corresponds with the following **Title and Options** dialog box.
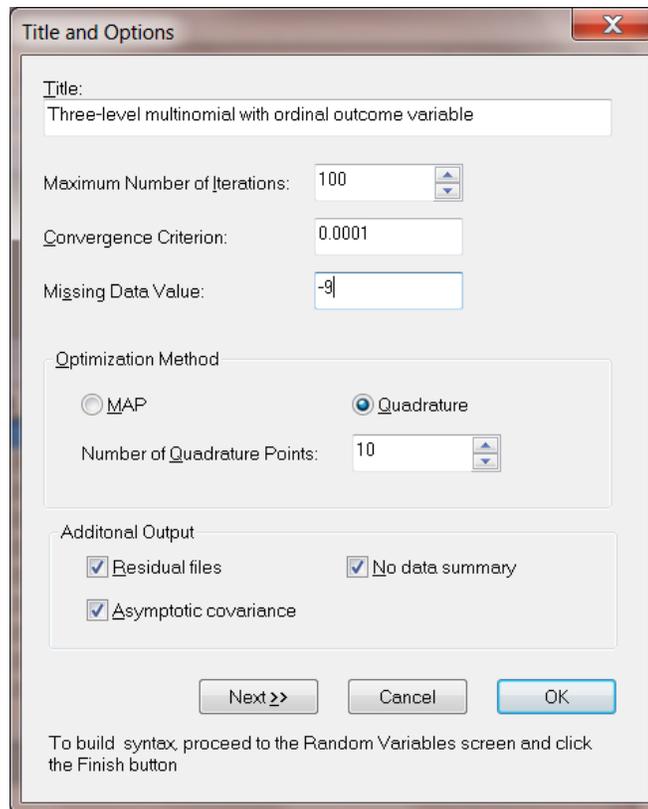
2.3.13   OFFSET command

The OFFSET command is an **optional** command used for count outcomes only. It is used to define the offset (exposure) variable.

Count models are also appropriate for rate data, where the rate is a count of events occurring for a particular unit of observation, divided by a measure of that unit's *exposure*. An example is the death rates in geographic areas as the count of deaths (outcome variable) divided by person-years (exposure). In count models, this is handled by an offset where the exposure variable is a predictor with regression coefficient constrained to 1.

Syntax

      OFFSET = <variable name>;

Example

      OFFSET = Pers_Yrs;

2.3.14   RANDOMn command

The RANDOMn command is used to identify those variables whose coefficients are allowed to vary randomly over level-2 and 3 of the hierarchy. One RANDOM command is allowed for each level of the hierarchy. When the syntax file is created through the interface, the RANDOM command(s) are automatically generated. Variables listed, except for the variable intcept (intercept), must be included in the data spreadsheet (*.**lsf** file). The spelling and case in which they are given need to correspond to that given in the spreadsheet. By default, the intercept is automatically included as a random effect at level-2 and 3 of the model. To exclude the intercept term at any level, the corresponding **Intercept** check box (see the **Random Variables** dialog box below) must be unchecked.

Syntax

> RANDOMn = <list of variables names to be included as random effects on level n> ;

where n = 2 or 3.

Example
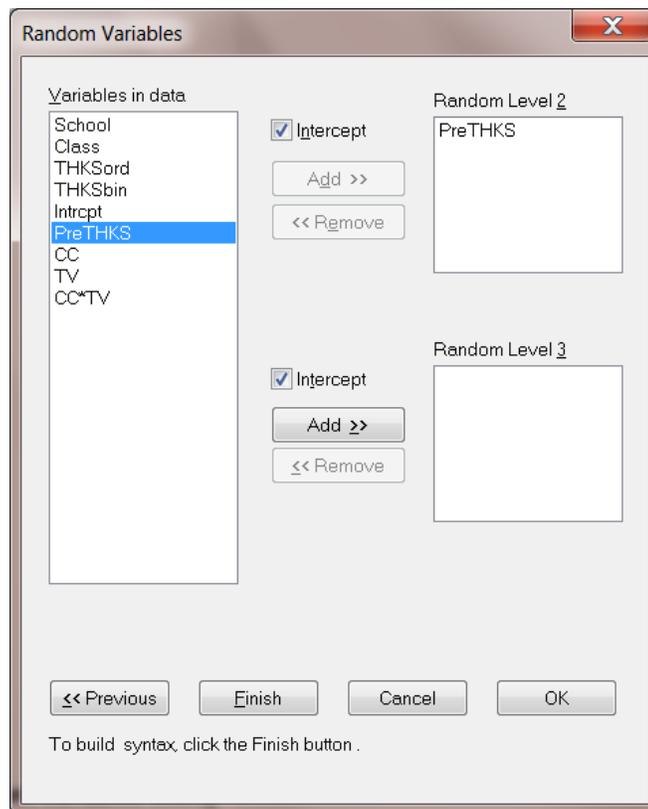
The **Random Variables** dialog box shown above corresponding to the commands

RANDOM2 = intcept PreTHKS;
RANDOM3 = intcept;

The settings corresponding to the following commands are shown on the **Random Variables** dialog box below:

RANDOM3 = intcept X2 X1 ;
RANDOM2 = time;
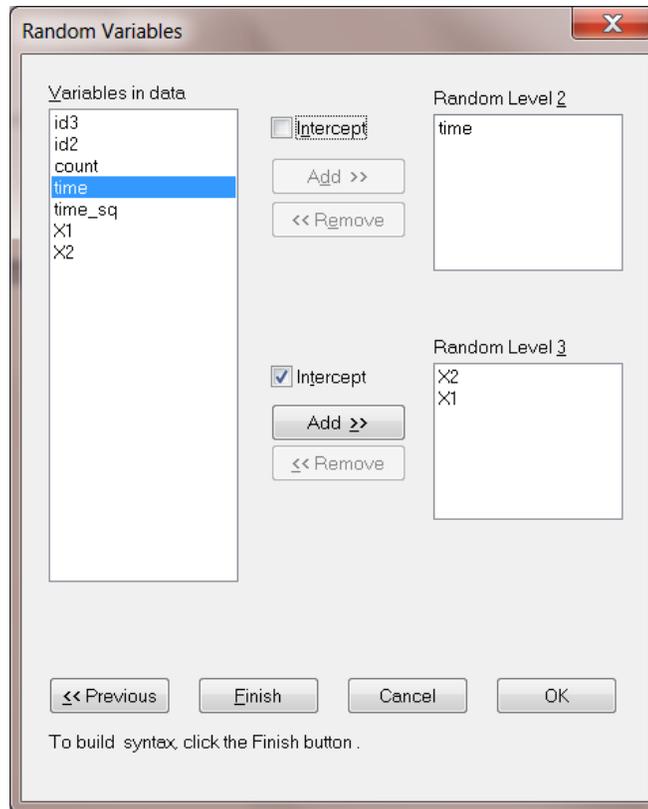
From this hypothetical example the following can be seen:

- o The random variables may be listed in any order.
- o Any or all of the possible predictors may be included in a RANDOM command at any level of the model.

2.3.15   SCALE command

Some sampling distributions such as the Poisson, Binomial, Gamma, Inverse Gaussian and Normal distributions have an **optional** scale parameter. This parameter is specified by using the SCALE command. Since not all sampling distributions involve a scale parameter, the command is **optional**. The SCALE command corresponds with the **Estimate scale?** drop-down list box on the **Distributions and Links** dialog box as shown below.
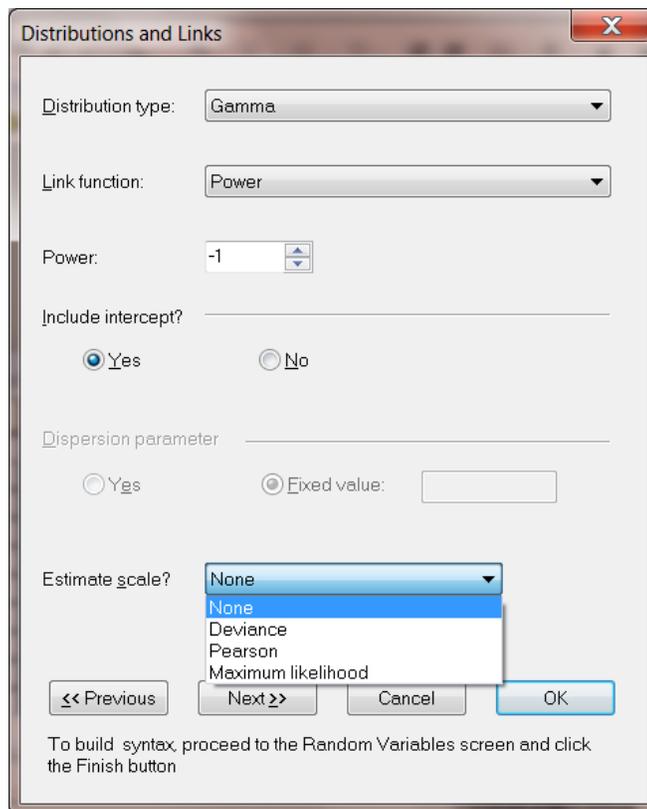
Syntax

SCALE = <type>;

where <type> is one of None, Pearson or Deviance.

Default

SCALE = None;

2.3.16   SY command

MGLIM can process raw data or frequency data that are available in the form of a LSF (LISREL System Data file). The LSF to be processed is specified by means of the SY command. The SY command is a **required** command and must be the **second** command listed in the syntax file. The SY command corresponds with the LSF window.

Syntax

SY = '<filename>';

where <filename> denotes the complete name (including drive and folder names) of the LSF. The drive and folder names may be omitted if the LSF and syntax file are in the same folder. Note the use of single quotes in this command.

Suppose that the data to be processed are listed in the file **ASPART.lsf** which is located in the **SGLIMEX** folder on the C drive. In this case, the SY command is given by

    SY = 'C:\MGLIMEX\ASPART.LSF';

2.3.17   TITLE command

It is often convenient to label a specific analysis to distinguish it from other analyses. This can be accomplished by using the TITLE command which is an **optional** command. The TITLE command corresponds with the **Title** string field on the **Title and Options** dialog box (see Section 2.2.2).

Syntax

    TITLE = <string> ;

where <string> denotes a descriptive title for the analysis.

Example

Consider an analysis in which a Bernoulli-Probit model was fitted to substance abuse data. In this case, one possible TITLE command is given by

    TITLE = MGLIMEX2: Binary outcome variable, logit link function;

2.3.18   WEIGHT command

Design weights are constructed for the ultimate sampling units of complex surveys. The purpose of the WEIGHT command is to allow you to specify the design weight variable. Since surveys without design weights are permitted, the WEIGHT command is an **optional** command. The WEIGHT command corresponds with the **Weight variable** section on the **ID and Weight Variables** dialog box (see Section 2.2.3).

Syntax

    WEIGHT = <label>;

where <label> denotes the case sensitive label of the design weight variable.

Example

Suppose that the variable A2TWA0 is used to capture the design weight for each observation. For this example, the WEIGHT command is given by

    WEIGHT = A2TWA0;

2.4 Examples

2.4.1 Introduction

2.4.1.1 Generalized linear models (GLIM)

Many popular statistical methods are based on mathematical models that assume data follow a normal distribution. The most obvious among these are the analysis of variance for planned experiments and multiple regression for general analyses of independent and dependent variables. In many situations, the normality assumption is not plausible. Consequently, use of methods that assume normality may perform unsatisfactorily. In these cases, other alternatives that do not require data to have a normal distribution are attractive.

A generalized linear model is a useful generalization of ordinary least square (OLS) regression. In OLS, the outcome variable is assumed to follow a normal distribution. However, this assumption is often invalid especially when dealing with categorical outcomes. A GLM assumes that the outcome variable follows a distribution function in the exponential family. It allows a linear transformation of the non-normal outcome variable. And it carries the regression concepts of OLS.

2.4.1.2 Multilevel model

The collection of models called Generalized Linear Models (GLIMs) have become important, and practical, statistical tools. The basic idea of GLIMs is an adaptation of standard regression to quite different kinds of data. The variables may be dichotomous (agree/disagree), categorical (as with a 5-point Likert scale), counts (number of arrest records), or nominal (choose among six candidates for mayor). The motivation is to tailor the regression relationship connecting the outcome to relevant independent variables so that it is appropriate to the properties of the dependent variable. The payoff is an analysis that often is more justifiable for the particular problem than a standard regression model would be.

The statistical theory and methods for fitting Generalized LInear Models (GLIMs) to simple random sample data are described in, amongst others, McCullach & Nelder (1989) and Agresti (2002). However, researchers from the social and economic sciences are often applying these methods to multilevel data.

Consequently, inappropriate results are obtained if these methods are applied to multilevel data. Statistical applications such as HLM (Raudenbush & Bryk 2007) and SAS PROC NLIN (SAS Institute 2004) implement methods to fit generalized linear models to multilevel data.

LISREL includes a statistical module for the analysis of multilevel data and uses two estimation methods, MAP (maximization of the posterior distribution) and ADAP (adaptive quadrature), which fits generalized linear models to multilevel data. The LISREL module allows for a wide variety of sampling distributions and link functions.

In the following sections, examples are provided to illustrate the MAP and ADAP methods to fit generalized linear models for binary, count, ordinal, and nominal response variables to multilevel data.

2.4.2 Models for binary outcomes

2.4.2.1 Binary outcome and its distribution

A binary random variable is a discrete random variable that has only two possible values, such as whether a subject dies (event) or lives (non-event). Such events are often described as success versus failure. The binary variable is often presented by the values 0 or 1. Obviously, the assumption of normality of the outcome variable does not hold anymore. The most common distribution for the binary outcome is Bernoulli distribution, which takes value 1 with success probability p and value 0 with failure probability $q = 1 - p$. The selection of the distribution for the outcome variable is not fixed. For example, if the occurrence is very rare, Poisson distribution can be used.

Linear predictors

The linear predictor is a quantity which relate to the expectation through the link function. For example, y is used as our binary outcome variable; η is the transformed linear predictor; $\mathbf{x} = (x_1 \ldots x_n)$ is the vector of all the predictors; and $\boldsymbol{\beta} = (\beta_1 \ldots \beta_n)$ denotes the unknown regression parameters. Thus, $\eta = \mathbf{x}'\boldsymbol{\beta}$ is the GLIM. Instead of y, η is being analyzed. For the binary outcome, the probability of success η is the predictor of interest.

Link functions

The link function provides the relationship between the linear predictor η and the distribution function. These link functions transform the observed outcome value to η and ensure that the predicted probability lies within the (0, 1) interval.

The available link functions for binary outcomes with Bernoulli distribution in MGLIM include logit, probit and complementary log-log functions. Table 3 gives each of these link function and the corresponding cumulative distribution functions (cdf), which are the inverse functions of each other, and its mean and variance.

Table 3: Link functions for Bernoulli distribution

Link name	Link function $F^{-1}(p)$, $0 < p < 1$	CDF $-\infty < w < \infty$	Mean	Variance
logit (logistic)	$\text{logit}(p) = \ln\left(\dfrac{p}{1-p}\right)$	$\dfrac{e^w}{1+e^w}$	0	$\dfrac{\pi^2}{3}$
probit	$\Phi^{-1}(p)$, where Φ^{-1} is the inverse of the standard normal cumulative distribution	$\Phi(w)$	0	1
complementary log-log	$\log(-\log(1-p))$	$1-\exp(-\exp(w))$	-0.577	$\dfrac{\pi^2}{6}$

These link functions map the probability η with an open interval (0,1) to the entire real numbers $\mathbb{R}$.

The logit and probit link functions are both symmetric. The logit function has a larger variance. The complementary log-log link function is asymmetric. When *p* is extremely small or large, the linear relationship does not hold. Understand the link function is essential when interpreting the outcomes.

2.4.3 Binary models with logit link function

2.4.3.1 The data

The data set forms part of the data library of the Alcohol and Drug Services Study (ADSS). The ADSS is a national study of substance abuse treatment facilities and clients. Background data and data on the substance abuse of a sample of 1752 clients were obtained. The sample was stratified by census region (CENREG) and within each stratum a sample was obtained for each of three facility treatment types (FACTYPE) within each of the four census regions. More information on the ADSS and the data are available at http://webapp.icpsr.umich.edu/cocoon/ICPSR-STUDY/03088.xml.

The specific data set is provided in the **Multilevel Generalized Linear Model examples** folder as the LSF file **Depress.LSF**. The first portion of this file is shown in the following LSF window.

	depr	sex	race_d	SEXxRace	A2TWA0	LEV2ID
1	1.00	0.00	1.00	0.00	316.80	2.00
2	1.00	0.00	0.00	0.00	276.60	2.00
3	1.00	1.00	0.00	0.00	276.60	2.00
4	1.00	0.00	0.00	0.00	276.60	2.00
5	1.00	1.00	0.00	0.00	276.60	2.00
6	0.00	0.00	0.00	0.00	276.60	2.00
7	1.00	1.00	0.00	0.00	276.60	2.00
8	1.00	1.00	0.00	0.00	276.60	2.00
9	1.00	0.00	0.00	0.00	276.60	2.00
10	1.00	0.00	0.00	0.00	276.60	2.00

The variables of interest are:

- o DEPR addresses the question of whether the patient has depression. (1 = Yes; 0 = No)
- o A2TWA0 is the sampling weight
- o SEX is a dummy variable indicating the gender (0 for male and 1 for female) of the client
- o RACE_D is a dummy variable representing the ethnicity (0 for black and 1 for white) of the client
- o SEXxRACE is the interaction term of gender and race.
- o LEV2ID is the variable used to identify the level-2 ID or grouping variable.

2.4.3.2 Importing the data

The data set shown previously is available in the form of a spreadsheet file, named **depress.lsf**. This file contains data for the 2,214 respondents who reported some form of depression.

The first step is to create the LISREL spreadsheet file (**lsf)** from the Excel file. Use the **Import Data File** option on the **File** menu to load the **Open** dialog box. Select **Excel (*.xls)** from the **Files of type** drop-down list. Browse for the file **depress.xls**. Select the file and click on the **Open** button to open the following LISREL spreadsheet window for **depress.lsf**.

	depr	sex	race_d	SEXxRace	A2TWA0	LEV2ID
1	1.00	0.00	1.00	0.00	316.80	2.00
2	1.00	0.00	0.00	0.00	276.60	2.00
3	1.00	1.00	0.00	0.00	276.60	2.00
4	1.00	0.00	0.00	0.00	276.60	2.00
5	1.00	1.00	0.00	0.00	276.60	2.00
6	0.00	0.00	0.00	0.00	276.60	2.00
7	1.00	1.00	0.00	0.00	276.60	2.00
8	1.00	1.00	0.00	0.00	276.60	2.00
9	1.00	0.00	0.00	0.00	276.60	2.00
10	1.00	0.00	0.00	0.00	276.60	2.00

Besides EXCEL data files, LISREL is capable of importing SAS, SPSS, STATA and most of the data files in other formats. The data import processes are similar, and will not be discussed again in this document.

When the external data is imported into LISREL, the default variable type is Ordinal. Variables that have more than 15 categories are treated as continuous. To change the default settings for the variable type, click on the **Data**, **Define** variables menu and change the settings. In this example the default setting is valid, and no further changes are needed.

2.4.3.3 Exploring the data

Graphics are often a useful data-exploring technique through which the researcher may familiarize her- or himself with the data. Relationships and trends may be conveyed in an informal and simplified visual form via graphical displays.

The **Graphs** option provides univariate, bivariate, and multivariate graphs. Univariate graphs are particularly useful in obtaining an overview of the characteristics of a variable.

2.4.3.4 Univariate graphs

As a first step, we take a look at the distribution of depression (DEPR), which is the potential dependent variable in this study.

Pie chart

A pie chart gives a good picture of probability of success. To create a pie chart for DEPR, select the **Univariate** option from the **Graphs** menu as shown below.
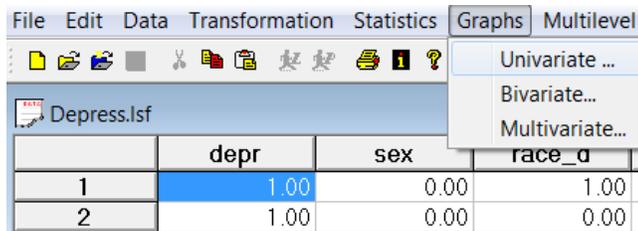
The **Univariate plot** dialog box appears. Select the variable DEPR and indicate that a **Pie chart** is to be graphed. Click the **Plot** button to display the pie chart.
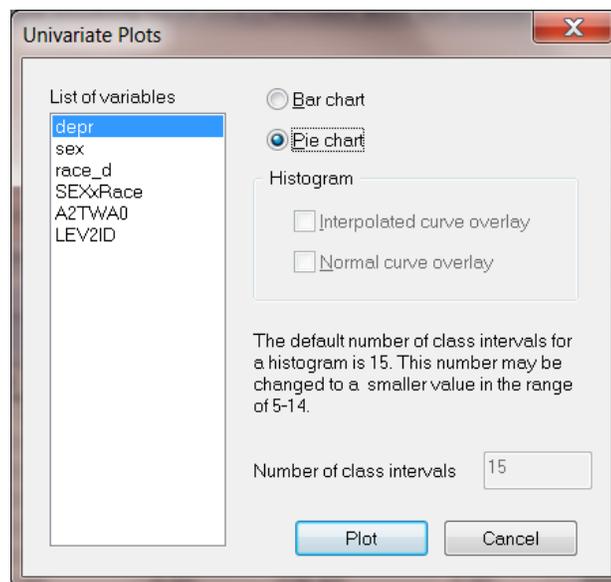
The pie below shows that about 58.9% of the respondents have depression. Since the probability is not extremely large or small, the Bernoulli distribution should be appropriate for our study.
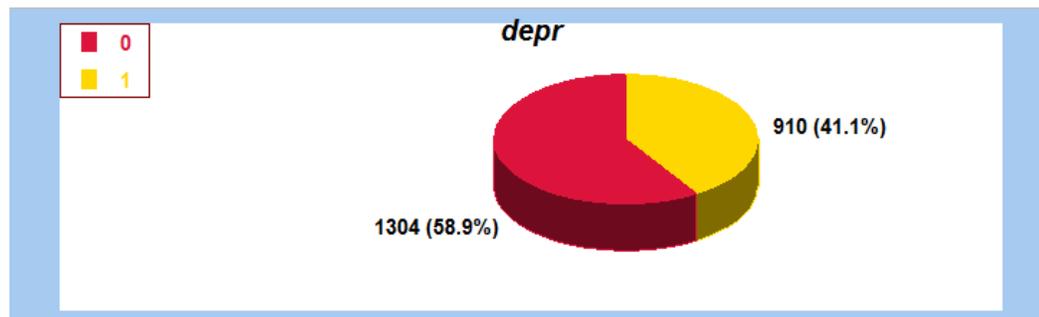
2.4.3.5 The model

The first model fitted to the depression data explores the relationship between DEPR, gender, and race, as represented by the variables SEX and RACE_D. The level-1 model is at a individual level, while the level-2 model is at a PSU level. The model can be expressed as follows.

For the binary case with logit link considered here

$$\text{Prob}(\text{DEPR}_{ij} = 1) = \frac{e^{\eta_{ij}}}{1 + e^{\eta_{ij}}}$$

where η_{ij} represents the log of the odds of success. With the logit link function, the probability $\text{Prob}(y_{ij} = 1 \,|\, \boldsymbol{\beta})$ is transformed to lie in the interval $(0,1)$. And (for the current model) the two level model can be expressed as

Level-1 model:
$$\eta_{ij} = b_{0i} + b_{1i} \times (\text{SEX})_{ij} + b_{2i} \times (\text{RACE\_d})_{ij} + e_{ij}$$

Level-2 model:
$$b_{0i} = \beta_0 + u_{0i}$$
$$b_{1i} = \beta_1 + u_{1i}$$
$$b_{2i} = \beta_2$$

where
$$e_i \sim N\left(0, \sigma^2 \mathbf{I}_i\right)$$
$$\mathbf{u}_i \sim N\left(0, \boldsymbol{\Sigma}_i\right)$$

35

β_0 denotes the average expected η_{ij}, which can be converted to the expected probability of getting depression. β_1 denotes the coefficient of the predictor variable SEX (slope) in the fixed part of the model. The random coefficients u_{i0} and e_{ij} denote the variation in the average expected DEPR value between PSUs and between patients respectively.

2.4.3.6 Setting up the analysis

Open the LISREL spreadsheet **depress.lsf** used during the exploratory analysis discussed previously. The next step is to describe the model to be fitted. We use the LISREL interface to provide the model specifications. From the main menu bar, select the **Multilevel, Generalized Linear Model**, **Title and Options** option.
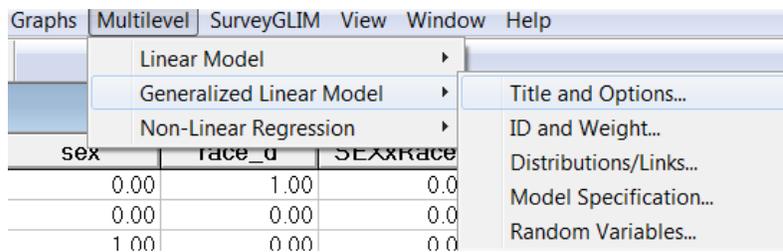
The multilevel generalized linear model contains five consequential dialogs boxes. The **Titles and Options** dialog box as shown below enables the user to input the title, maximum number of iteration, convergence criterion, missing values, and method and request additional output. Enter a title for the analysis in the **Title** text boxes (optional) and keep all the other settings as default.

Proceed to the **ID and Weights** screen by clicking on the **Next** button. Highlight LEV2ID from the **Variables in data** list and click on the upper **Add** button to select is as the **Level-2 ID variable**. Similarly, highlight the variable A2TWA0 and click on the lower **Add** button to select it as the **Weight variable** and obtain the screen shown below.
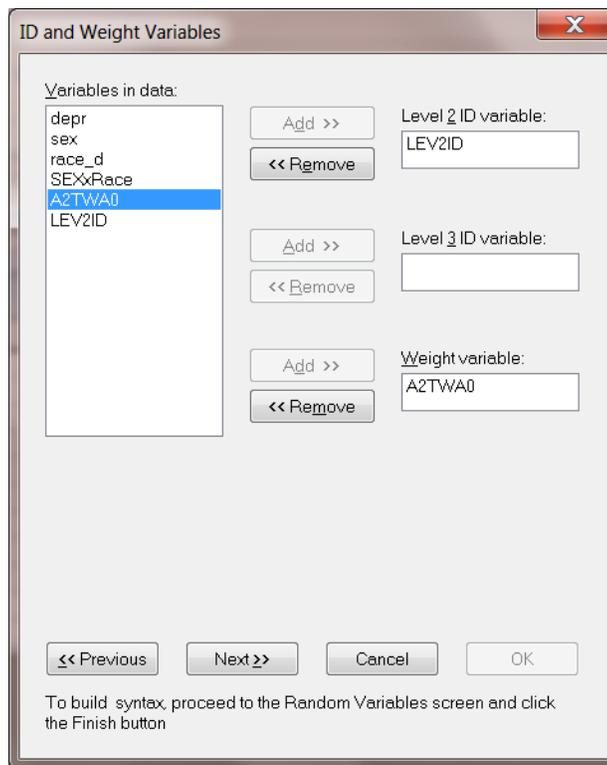
Click on the **Next** button to load the **Distribution and Links** dialog box. Select **Binomial** from the **Distribution type** dropdown list box. By default, the logit link function is selected. Keep the other default settings unchanged as shown below, and click on the **Next** button.
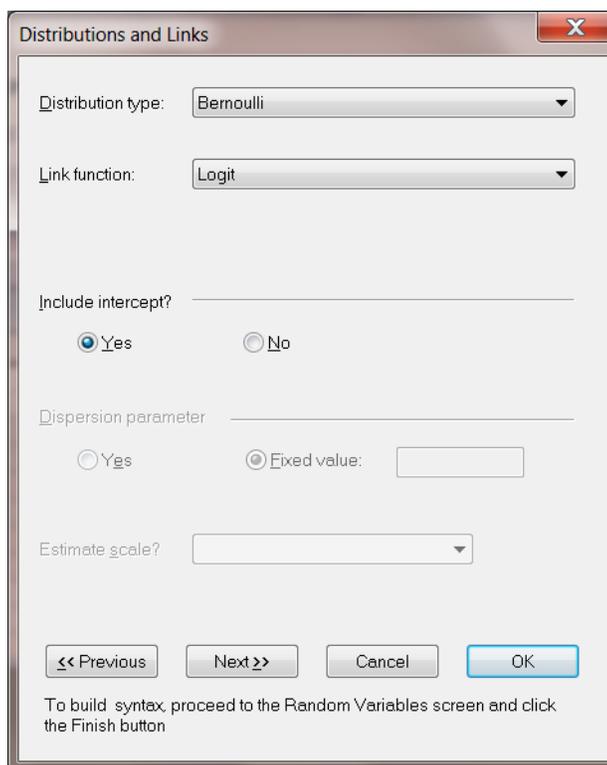
On the **Dependent and Independent Variables** dialog box screen, first select DEPR and click on the upper **Add** button to define it as the **Dependent variable**. Then, select SEX and RACE_D and click on the **Continuous** button to add these variables in the **Independent variables** list box as shown below.
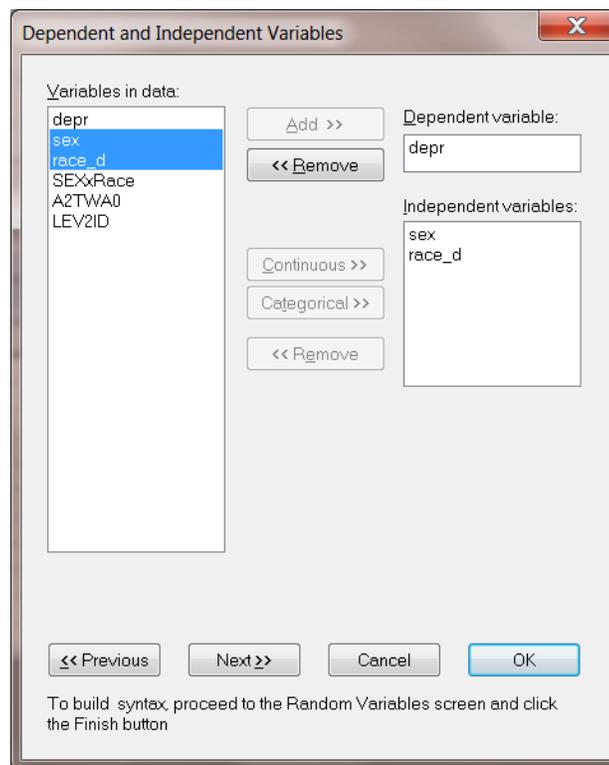
Click on the **Next** button to proceed to the **Random Variables** dialog box once these settings have been defined. Keep the **Intercept** check box checked so as to include a level-2 intercept.
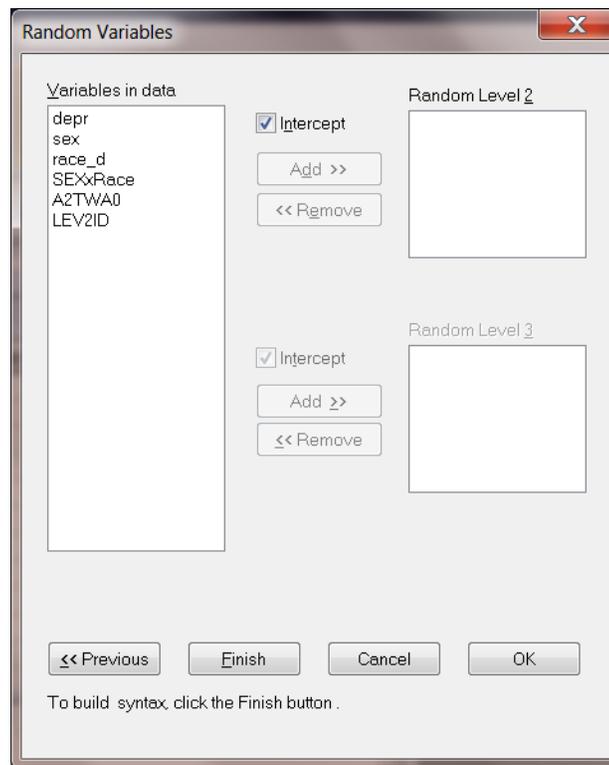
Click on the **Finish** button to generate the PRELIS syntax file (**.prl**) that corresponds to the above settings. Select the **File**, **Save As** option, and provide a name (**depress1.prl**) for the model specification file. The default folder for the syntax to be saved in is the same folder used for the data file.

```
 MGlimOptions Converge=0.0001 MaxIter=100 MissingCode=-999999
             Method=Quad NQUADPTS=8;
Title=Bernoulli-Logit model based on Depression data;
SY=Depress.LSF;
ID2=LEV2ID;
Weight=A2TWA0;
Distribution=BER;
Link=LOGIT;
Intercept=Yes;
DepVar=depr;
CoVars=sex race_d ;
RANDOM2=intcept;
```

2.4.3.7 The syntax file

The syntax file contains the following information:
   o   The MGlimOptions keyword requests the MGLIM module to run. The first two lines, together with the Title line, correspond to the settings entered in the **Title and Options** dialog box.
   o   The SY line indicates the location of the **.lsf** data file.

- o ID2 is the level-2 id variable, while Weight corresponds with the weight variable. These are defined in the **ID and Weights** dialog box.
- o The syntax lines for Distribution, Link and level-1 Intercept are set up in the **Distribution and Links** dialog box.
- o The DepVar line, which represents the dependent variable and the CoVars line, which represents the covariate variable, are defined in the **Dependent and Independent Variables** dialog box.
- o Finally, the RANDOM2 syntax line corresponds to the **Random Variables** dialog box.

Understanding how the syntax works enables the user to make changes directly to the syntax file. Run the analysis by selecting the **Run PRELIS** button to generate the output file **depress.out**. The output file has the same file name as the syntax file with a different extension **.out**. It is saved in the same folder as the syntax file.

2.4.3.8 Discussion of results

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

Program information and syntax

At the top of the output file, program information is given. It states the version number, corporate and technical support information, the date and time of analysis, and the locations of data file and syntax file.

```
Depress.OUT                                                    _ □ ✕

 o- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -o
 |                                                                        |
 |                    MLGLIM module of LISREL 9.1                         |
 |                    Generalized Linear Modeling                         |
 |                               of                                       |
 |                    Multilevel (Hierarchical) Data                      |
 |                                                                        |
 |        Copyright by Scientific Software International, Inc., 2012       |
 |        Scientific Software International, Inc.                          |
 |        Website:  www.ssicentral.com                                    |
 |        Support:  techsupport@ssicentral.com                            |
 |                                                                        |
 |  DATE OF ANALYSIS: August 15, 2012                                     |
 |  TIME OF ANALYSIS: 17H41:22                                            |
 |  DATA FILE:                                                            |
 |  C:\LISREL9 Examples\MGLIMEX\Depress.lsf                               |
 |  INPUT FILE:                                                           |
 |  C:\LISREL9 Examples\MGLIMEX\Depress.PRL                               |
 |                                                                        |
 o- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -o
```

Program information is followed by the Multilevel GLIM syntax. This section echoes the contents of the syntax file **depress.prl**. For more information on syntax and keywords, please see Section 2.2.3.

Model and data description

In the next section of the output file as shown above, descriptions of the distribution, the link function, the weight variable and the hierarchical structure of the data are provided. Data from a total of 10 level-2 units and 2,214 respondents were included at levels 2 and 1 of the model. In addition, a summary of the number of respondents nested within each level-2 unit is provided.
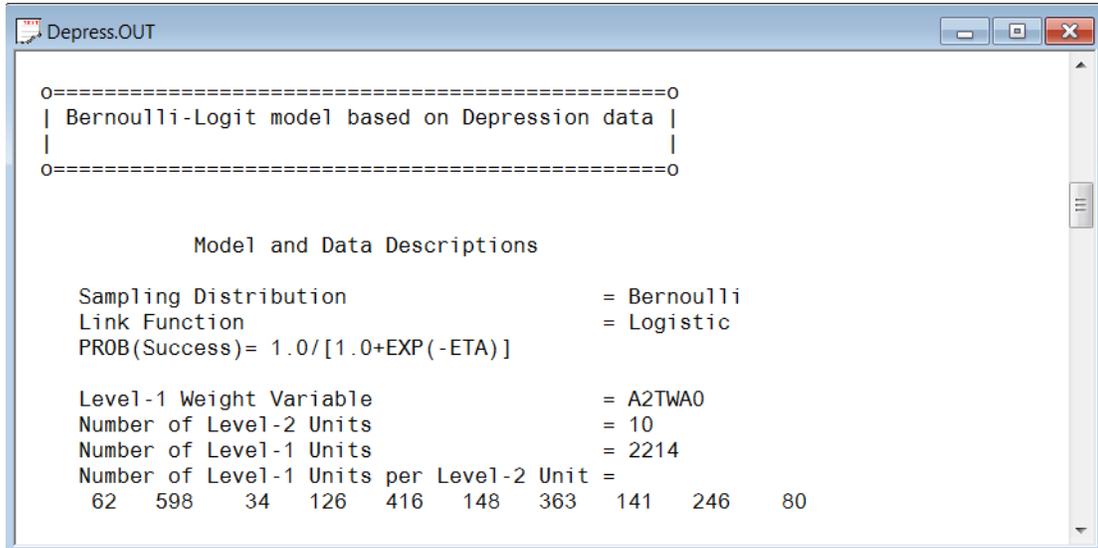
```
Depress.OUT                                                    _  □  ✕

O===============================================O
| Bernoulli-Logit model based on Depression data |
|                                                 |
O===============================================O


            Model and Data Descriptions

   Sampling Distribution                = Bernoulli
   Link Function                        = Logistic
   PROB(Success)= 1.0/[1.0+EXP(-ETA)]

   Level-1 Weight Variable              = A2TWA0
   Number of Level-2 Units              = 10
   Number of Level-1 Units              = 2214
   Number of Level-1 Units per Level-2 Unit =
     62    598     34    126    416    148    363    141    246    80
```

Descriptive statistics

The data summary is followed by descriptive statistics for all the variables included in the model. Since DEPR is defined as a binary variable, it is presented by two dummy variables depr1 and depr2.
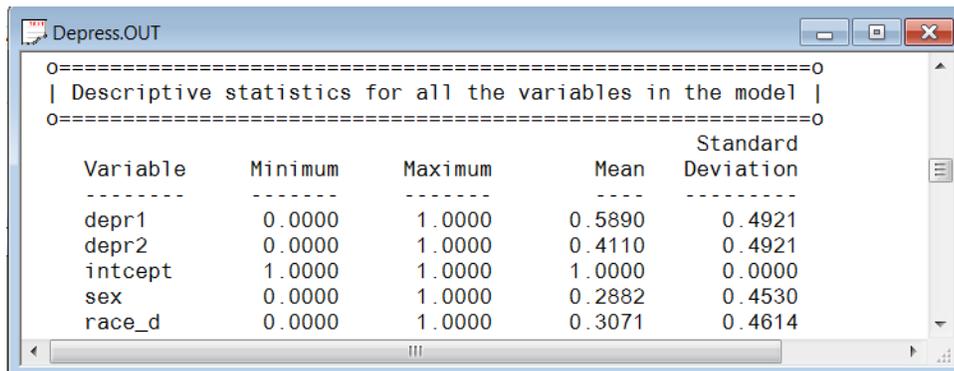
```
Depress.OUT                                                    _  □  ✕

O=========================================================O
| Descriptive statistics for all the variables in the model |
O=========================================================O
                                                    Standard
    Variable    Minimum    Maximum      Mean    Deviation
    --------    -------    -------      ----    ---------
    depr1       0.0000     1.0000     0.5890      0.4921
    depr2       0.0000     1.0000     0.4110      0.4921
    intcept     1.0000     1.0000     1.0000      0.0000
    sex         0.0000     1.0000     0.2882      0.4530
    race_d      0.0000     1.0000     0.3071      0.4614
```

Results for the model without any random effects

Descriptive statistics are followed by the results for the model without any random effects. These parameters are used in the initial step of the iterative algorithm. They are obtained by ordinary weighted least squares (WLS) regression. The goodness of WLS fit statistics are also given as shown below.

```
┌─ Depress.OUT ─────────────────────────────────────────── ─ □ ✕ ┐
│                                                                  │
│ O================================================O              │
│ | Results for the model without any random effects |            │
│ O================================================O              │
│                                                                  │
│                   Goodness of fit statistics                     │
│                                                                  │
│    Statistic                           Value       DF     Ratio  │
│    ---------                           -----       --     -----  │
│    Likelihood Ratio Chi-square       3324.9263    2211    1.5038 │
│    Pearson Chi-square                2575.8286    2211    1.1650 │
│                                                                  │
│                                                                  │
│                 Estimated regression weights                     │
│                                                                  │
│                          Standard                                │
│    Parameter     Estimate      Error    z Value    P Value       │
│    ---------     --------    --------    -------    -------       │
│    intcept        -0.1433      0.0551    -2.6007     0.0093       │
│    sex             0.6949      0.0981     7.0822     0.0000       │
│    race_d         -0.5683      0.1034    -5.4950     0.0000       │
│                                                                  │
└──────────────────────────────────────────────────────────────────┘
```

Results for the model with fixed and random effects
Number of iterations and fit statistics

The total number of (macro) iterations is reported. As shown below, there are 58 iterations to get the converged results.

In addition to the likelihood function value at convergence, a number of related statistical measures for assessing model adequacy are available. The most common of these are the likelihood ratio test, Pearson chi-square, and Akaike's and Schwarz's criteria. Both the Akaike information criterion (AIC) and the Schwarz Bayesian criterion (SBC) are functions of the number of estimated parameters, and therefore "penalize" models with large numbers of parameters. In the LISREL output file, all three of these are reported. A chi-square scale factor, with which a chi-square value obtained from the difference between two deviance statistics should be multiplied to yield a corrected chi-square statistic in the case of a weighted analysis, may also be found in this section.

```
┌─ Depress.OUT ─────────────────────────────────────────── ─ □ ✕ ┐
│                                                                  │
│ O======================================O                        │
│ | Optimization Method: Adaptive Quadrature |                    │
│ O======================================O                        │
│                                                                  │
│    Number of quadrature points =            10                   │
│    Number of free parameters =               4                   │
│    Number of iterations used =               2                   │
│                                                                  │
│    -2lnL (deviance statistic) =      2892.37979                  │
│    Akaike Information Criterion      2900.37979                  │
│    Schwarz Criterion                 2923.19002                  │
│                                                                  │
│  ◄                      III                        ►             │
└──────────────────────────────────────────────────────────────────┘
```

- The Pearson Chi-square is defined as $\chi_P^2 = \sum_{i=1}^{N}\sum_{j=1}^{n_i}\sum_{k=1}^{n_{ij}} \frac{w_{ijk}\left(y_{ijk}-\hat{\mu}_{ijk}\right)^2}{\hat{\sigma}^2\left(y_{ijk}\right)}$.

- The deviance is defined as $-2\ln L$. For a pair of nested models, the difference in $-2\ln L$ values has a χ^2 distribution, with degrees of freedom equal to the difference in number of parameters estimated in the models compared.

- The AIC was originally proposed for time-series models, but is also used in regression. It is defined as $-2\ln L + 2r$, where r denotes the number of parameters estimated in the model. The model with minimum AIC, in a set of nested models, will be the most parsimonious according to this criterion.

- The SBC is defined as $-2\ln L + r\log n$, where n denotes the number of units at the highest level of the hierarchy. A smaller value of this criterion would indicate the most parsimonious of the models being compared.

Estimated regression weights

The output describing the estimated regression weights after fit statistics is shown next. The estimates are shown in the column with heading Estimate and correspond to the coefficients β_0, β_1 and β_2 in the model specification. From the z-values and associated exceedance probabilities, we see that all three estimates are highly significant at 10% level.
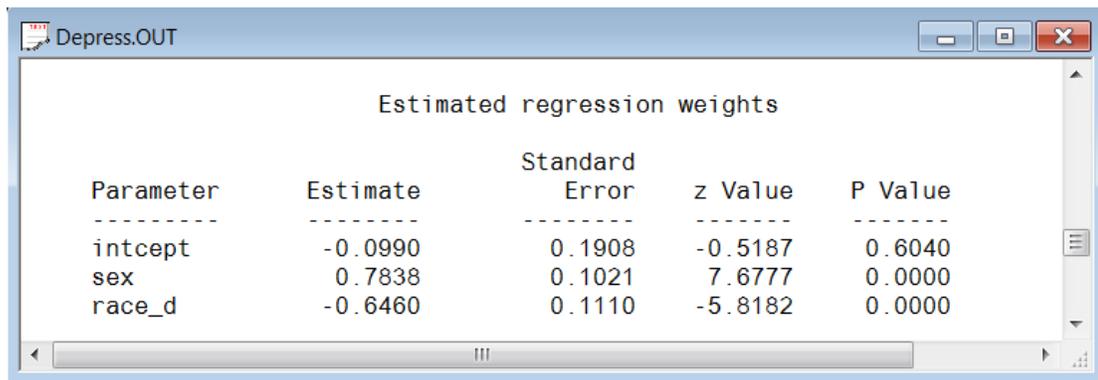
```
Depress.OUT                                               ─ ▫ ✕

                    Estimated regression weights

                              Standard
     Parameter      Estimate     Error    z Value    P Value
     ---------      --------   --------    -------    -------
     intcept        -0.0990     0.1908    -0.5187     0.6040
     sex             0.7838     0.1021     7.6777     0.0000
     race_d         -0.6460     0.1110    -5.8182     0.0000
```

The estimated intercept is -0.0990, which is the average logit. The estimated coefficients associated with gender (SEX) is -0.7838, which indicates that the female respondents (SEX = 1) have a smaller $\hat{\eta}$. The estimate for the indicator of race (RACE_D) shows that white clients have higher $\hat{\eta}$ value. To describe the η's in an more accessible way to readers of reports, we need the link functions to transform them into probabilities.

Interpreting estimated regression weights by using link function

First, we substitute the regression weights and obtain the function for $\hat{\eta}_{ij}$

$$\hat{\eta}_{ij} = \hat{b}_{0i} + \hat{b}_{1i} \times (\text{SEX})_{ij} + \hat{b}_{2i} \times (\text{RACE\_d})_{ij}$$
$$= -0.0990 + 0.7838 \times (\text{SEX})_{ij} - 0.6460 \times (\text{RACE\_d})_{ij}.$$

For a black male, we have SEX = 0, RACE_d = 0, thus

$$\hat{\eta}_{ij} = -0.0990 + 0.7838 \times 0 - 0.6460 \times 0$$
$$= -0.0990 \qquad .$$

Similarly, the calculation of $\hat{\eta}_{ij}$ for a BLACK female (SEX = 1, RACE_d = 0) is

$$\hat{\eta}_{ij} = -0.0990 + 0.7838 \times 1 - 0.6460 \times 0$$
$$= 0.6848 \qquad .$$

Next, we transform the $\hat{\eta}_{ij}$'s into corresponding probabilities by using the logit link function. Take the black male as the example, the probability is calculated as below.

$$\text{Prob}(\text{DEPR}_{ij} = 1) = \frac{e^{\hat{\eta}_{ij}}}{1 + e^{\hat{\eta}_{ij}}} = \frac{1}{1 + e^{-\hat{\eta}_{ij}}} = \frac{1}{1 + e^{0.0990}} = 47.53\%$$

Thus we can conclude that the estimated probability for a black male who has depression is about 47.53%. Similarly, the probability of having depression for different gender and ethnicity are calculated in the following table.

Group	Code	$\hat{\eta}$	Prob (DEPR = 1)
Black, male	sex = 0, race_d = 0	0.1018	47.53%
Black, female	sex = 1, race_d = 0	0.6848	66.48%
White, male	sex = 0, race_d = 1	-0.7450	32.19%
White, female	sex = 1, race_d = 1	0.0388	50.97%

Estimated level-2 variance

The output for the estimated level-2 variance is shown in the image below. The *p* value of intercept shows the probability of getting depression differs significantly from PSU to PSU (the level-2 units).

```
Depress.OUT                                              [_][□][✕]
              Estimated level 2 variances and covariances

                                        Standard
      Parameter              Estimate      Error    z Value   P Value
      ---------              --------    --------    -------   -------
      intcept/intcept          0.3204      0.1633    1.9621    0.0498


   Level 2 covariance matrix

                  intcept
      intcept     0.320389

   Calculation of the intracluster correlation
   -------------------------------------------
   residual variance = pi*pi / 3 (assumed)
   cluster  variance =    0.3204

   intracluster correlation =    0.3204 / (    0.3204 + (pi*pi/3)) = 0.089
```
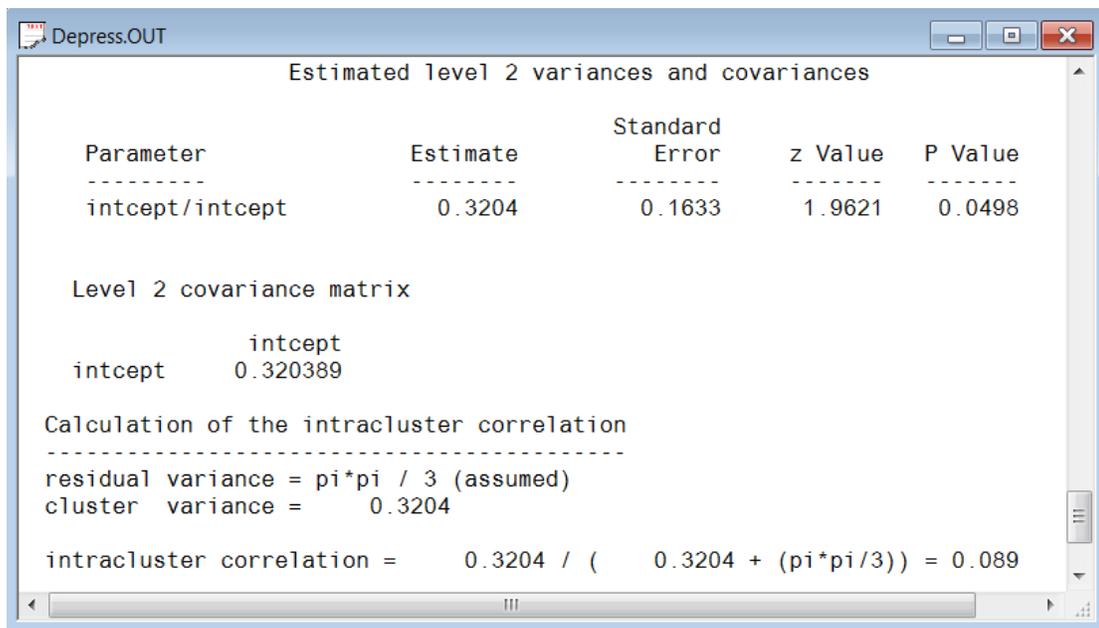
ICCs and % variance explained

The intraclass coefficient (ICC), or say the percentage of variance explained by level-2 unit is calculated by

$$ICC = \frac{\text{level 2 variation}}{\text{level 1 variation} + \text{level 2 variation}}$$

In the case of a model with only a random intercept, the variation in the random intercept at the level-2 unit, and the residual variation at level-1. The intracluster coefficient is defined as

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

Level-1 variation

As mentioned earlier, for the dichotomous outcome model, it is assumed that the level-1 error variance is equal to $\pi^2/3$ for the logistic link function if the model is true (see, e.g., Hedeker & Gibbons (2006), p. 157). Thus,

$$\text{Var(level 1)} = \text{var}(e_{i0}) = \frac{\pi^2}{3} = 3.2865$$

46

2.4.4 Binary models with probit link function

2.4.4.1 The model

In the model just discussed the logistic link function was used. We now will fit another model by using the probit link function. We will only include a random intercept at level-2. On the other hand, the interaction between gender and race will be considered at level-1. The model can be expressed as follows.

The probit link function is

$$\text{Prob}(\text{DEPR}_{ij} = \gamma) = \Phi^{-1}(\eta)$$

where η_{ij} represents the log of the odds of success. With the probit link function, the probability $\text{Prob}(y_{ij} = 1 \mid \boldsymbol{\beta})$ is transformed to lie in the interval (0,1). And (for the current model) the two level model can be expressed as

Level-1 model:
$$\eta_{ij} = b_{0i} + b_{1i} \times (\text{SEX})_{ij} + b_{2i} \times (\text{RACE\_d})_{ij} + b_{3i} \times (\text{SEXxRACE})_{ij} + e_{ij}$$

Level-2 model:
$$b_{0i} = \beta_0 + u_{0i}$$
$$b_{1i} = \beta_1$$
$$b_{2i} = \beta_2$$
$$b_{3i} = \beta_3$$

The current model only assumes the intercepts differ between PSUs, but the slopes are the same for all the level-2 units.

2.4.4.2 Setting up the analysis

Open the LISREL spreadsheet **depress.lsf** used previously, Select the **Multilevel, Generalized Linear Model**, **Title and Options** option. Input the new analysis title in the **Titles and Options** dialog box and keep all the other settings as default.
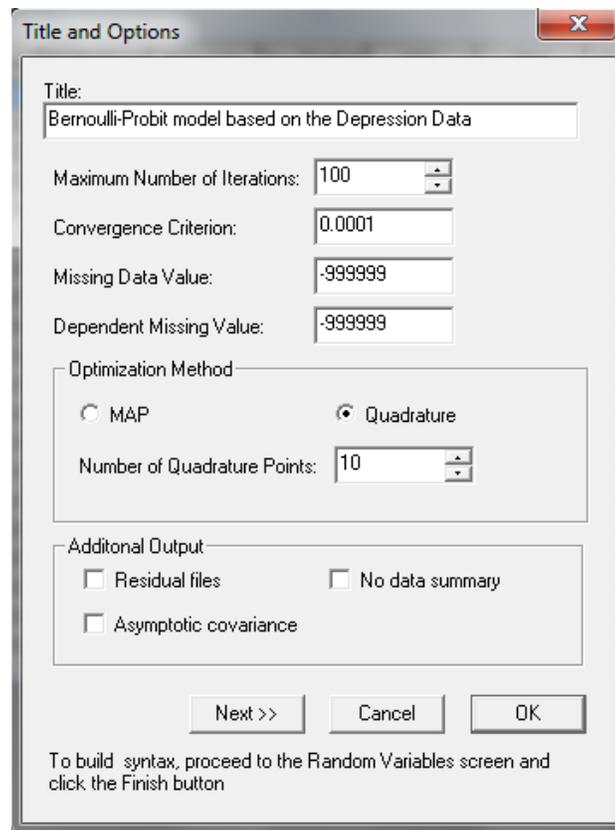
Proceed to the **ID and Weights** screen by clicking on the **Next** button. The settings of this screen are exactly the same as the previous model. Highlight LEV2ID from the **Variables in data** list and click on the upper **Add** button to select is as the **Level-2 ID variable**. Similarly, highlight the A2TWA0 and click on the lower **Add** button to select it as the **Weight variable** to obtain the screen shown below.

Click on the **Next** button to load the **Distribution and Links** dialog box. Select **Binomial** from the **Distribution type** dropdown list box. Select **Probit** from the **Link function** drop-down list. Keep the other default settings unchanged as shown below, and click on the **Next** button.

On the **Dependent and Independent Variables** dialog box screen, first select DEPR and click on the upper **Add** button to define it as the **Dependent variable**. Then select the variables SEX, RACE_D and SEXxRace and click on the **Continuous** button to add them to the **Independent variables** list box as shown below.
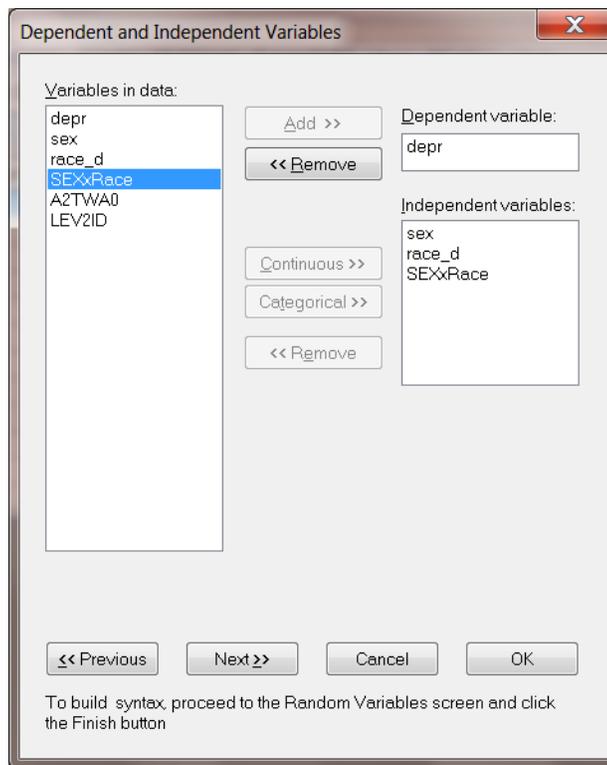
Click on the **Next** button to proceed to the **Random Variables** dialog box. Keep the **Intercept** check box checked to include the random effect for the level-2 intercept.
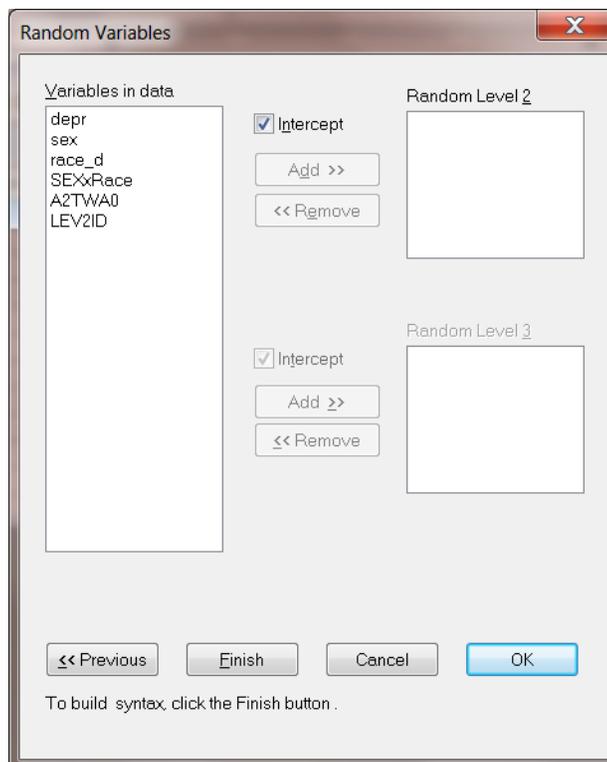
Click the **Finish** button to generate the PRELIS syntax file (**.prl**) that corresponds to the above settings. Select the **File**, **Save As** option, and provide a name (**depress2.prl**) for the model specification file. The default folder to which the the syntax file will be save is the folder in which the data file is located.
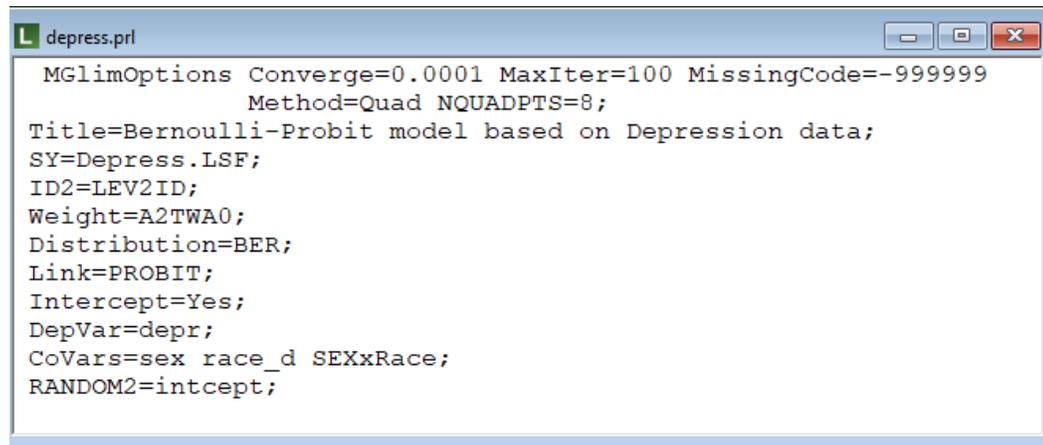
```
L depress.prl                                                    _ □ ✕
 MGlimOptions Converge=0.0001 MaxIter=100 MissingCode=-999999
              Method=Quad NQUADPTS=8;
Title=Bernoulli-Probit model based on Depression data;
SY=Depress.LSF;
ID2=LEV2ID;
Weight=A2TWA0;
Distribution=BER;
Link=PROBIT;
Intercept=Yes;
DepVar=depr;
CoVars=sex race_d SEXxRace;
RANDOM2=intcept;
```

2.4.4.3 The syntax file

Note that the following syntax lines are different from the previous model:

- o Link = PROBIT defines the probit link function, which is defined in the **Distribution and Links** dialog box.
- o CoVars = sex racd_d SEXxRace, now includes one more independent variable than was the case in the previous model.

Run the analysis by selecting the **Run PRELIS** button to generate the output file **depress2.out**, which is saved to the same folder as the syntax file.

2.4.4.4 Discussion of results

Portions of the output file are shown below.

Model and data descriptions

The program info and the syntax are printed on the top of the output file. In the next section of the output file as shown below, descriptions of the distribution, the link function, the weight variable and the hierarchical structure of the data is provided. Note that probit link function is used as the link function.

```
o==================================================o
| Bernoulli-Probit model based on Depression data |
|                                                  |
o==================================================o


             Model and Data Descriptions

   Sampling Distribution                  = Bernoulli
   Link Function                          = Probit
   PROB(Success)= PHI(ETA)
   PHI denotes the CDF of the standard Normal distribution

   Level-1 Weight Variable                = A2TWA0
   Number of Level-2 Units                = 10
   Number of Level-1 Units                = 2214
   Number of Level-1 Units per Level-2 Unit =
     62    598    34    126    416    148    363    141    246    80
```

Descriptive statistics

The data summary is followed by descriptive statistics for all the variables included in the model.

```
o==========================================================o
| Descriptive statistics for all the variables in the model |
o==========================================================o
                                                   Standard
     Variable    Minimum    Maximum       Mean    Deviation
     --------    -------    -------       ----    ---------
     depr1       0.0000     1.0000      0.5890     0.4921
     depr2       0.0000     1.0000      0.4110     0.4921
     intcept     1.0000     1.0000      1.0000     0.0000
     sex         0.0000     1.0000      0.2882     0.4530
     race_d      0.0000     1.0000      0.3071     0.4614
     SEXxRace    0.0000     1.0000      0.0980     0.2974
```

Results for the model without any random effects

The descriptive statistics is followed by the results for the model without any random effects. The deviance can be used for comparison tests of nested models. The estimated regression weights without any random effect are given below.

```
Depress2.OUT

O================================================O
| Results for the model without any random effects |
O================================================O

                    Goodness of fit statistics

    Statistic                              Value          DF
    ---------                              -----          --
    Likelihood Ratio Chi-square           3325.8392       2210
    Pearson Chi-square                    2578.1501       2210


                    Estimated regression weights

                                Standard
    Parameter      Estimate       Error    z Value   P Value
    ---------      --------     --------   -------   -------
    intcept        -0.0857       0.0357    -2.4036    0.0162
    sex             0.4175       0.0703     5.9361    0.0000
    race_d         -0.3712       0.0762    -4.8706    0.0000
    SEXxRace        0.0610       0.1396     0.4368    0.6623
```

Results for the model with fixed and random effects
Number of iterations and fit statistics
　　As shown below, 1iteration was needed to obtain convergence. The likelihood ratio test, Akaike's and Schwarz's criteria are given after the iteration number. For detailed information about these statistics, please refer to the previous section.

```
Depress2.OUT

O========================================O
| Optimization Method: Adaptive Quadrature |
O========================================O

    Number of quadrature points =           10
    Number of free parameters =              5
    Number of iterations used =              1

    -2lnL (deviance statistic) =      2892.01070
    Akaike Information Criterion      2902.01070
    Schwarz Criterion                 2930.52348
```

Estimated regression weights
　　The output describing the estimated regression weights for the multilevel model is shown next. The estimates are shown in the column with heading Estimate and correspond to the coefficients β_0, β_1, β_2 and β_3 in the model specification. From the z-values and associated exceedance probabilities, we see that the intercept and the regression weight for SEXxRace are not significant at a 10% level of significance.

```
Depress2.OUT                                          — ▢ ✕
                    Estimated regression weights

                              Standard
        Parameter      Estimate      Error    z Value   P Value
        ---------     --------   --------    -------   -------
        intcept        -0.0544     0.1186    -0.4582    0.6468
        sex             0.4612     0.0723     6.3824    0.0000
        race_d         -0.4207     0.0805    -5.2282    0.0000
        SEXxRace        0.0805     0.1430     0.5628    0.5736

                              III
```

The estimated intercept has changed to -0.0544 in this second model. The estimated coefficient associated with gender (sex) is now 0.4612, which indicates that the female respondents (sex = 1) have a smaller $\hat{\eta}$. The estimate for the indicator of race (race_d) shows that white clients have a higher $\hat{\eta}$ value in the current model. The interaction term of gender and race is positive, but not significant. The probit link function is needed to transform these results into probabilities.

Interpreting estimated regression weights by using link function

First, we substitute the regression weights and obtain the function for $\hat{\eta}_{ij}$

$$\hat{\eta}_{ij} = \hat{b}_{0i} + \hat{b}_{1i} \times (\text{sex})_{ij} + \hat{b}_{2i} \times (\text{race\_d})_{ij} + \hat{b}_{3i} \times (\text{SEXxRace})_{ij}$$
$$= -0.0544 + 0.4612 \times (\text{sex})_{ij} - 0.4207 \times (\text{race\_d})_{ij} + 0.0805 \times (\text{SEXxRace})_{ij} .$$

For a black male, we have SEX = 0, RACE_d = 0 and SEXxRace = 0 thus

$$\hat{\eta}_{ij} = -0.0544 .$$

Similarly, the calculation of $\hat{\eta}_{ij}$ for a black female (SEX = 1, RACE_d = 0 and SEXxRace = 0) is

$$\hat{\eta}_{ij} = -0.0544 + 0.4612 \times 1$$
$$= 0.4068 .$$

Next, we transform the $\hat{\eta}_{ij}$'s into corresponding probabilities by using the probit link function. Taking black males as the example, the probability is calculated as shown below.

$$\text{Prob}(\text{DEPR}_{ij} = 1) = \Phi(-0.0544) = 47.83\%$$

Similarly, the probabilities of having depression for different gender and ethnicity groups are reported in the following table.

Group	Code	$\hat{\eta}$	Prob (DEPR = 1)
Black, male	sex = 0, race_d = 0	-0.0544	47.83%
Black, female	sex = 1, race_d = 0	0.4068	65.79%
White, male	sex = 0, race_d = 1	-0.4751	31.74%
White, female	sex = 1, race_d = 1	0.0666	52.65%

Estimated level-2 variance and covariance

The output for the estimated level-2 variance and covariance is shown in the image below. In the previous model, we have discussed the interpretation of these in detail. In this model, the random part of the current model is simpler than in the previous model.
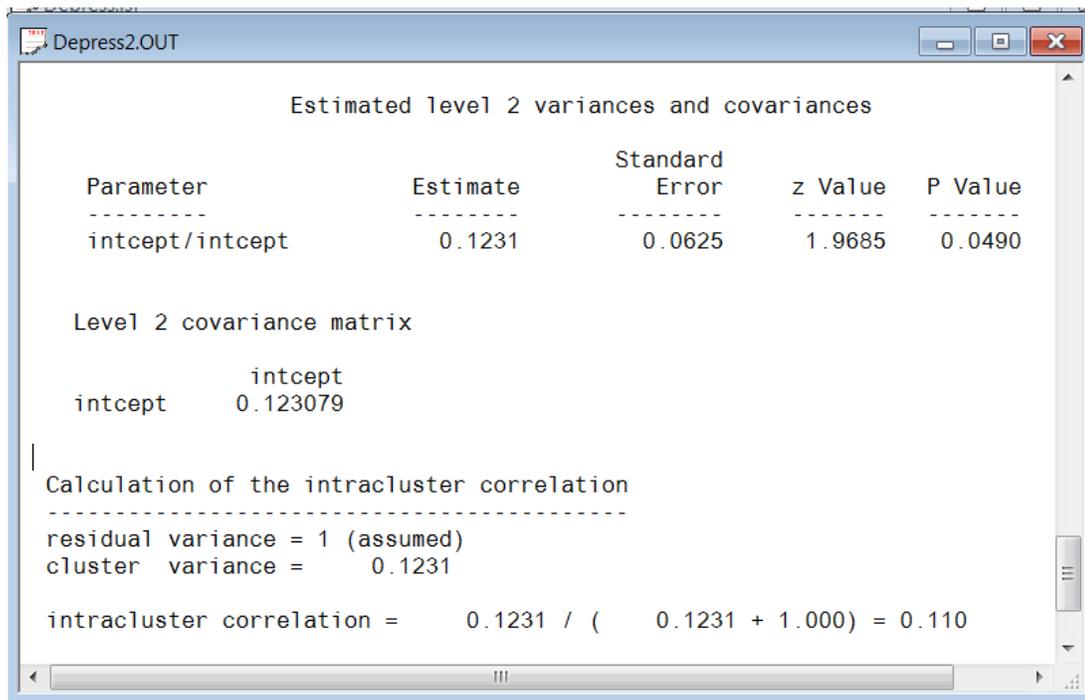
2.4.5 Bernoulli distribution with complementary log-log link function

2.4.5.1 The model

In the previous model, the probit link function is used. We now fit the same model by using a complementary log-log link function.

The complementary log-log link function is defined as

$$1\text{-}\exp\left(-\exp\left(\eta_{ij}\right)\right)$$

The level-1 and level-2 models are unchanged.

Level-1 model:

$$\eta_{ij} = b_{0i} + b_{1i} \times \left(\text{SEX} \right)_{ij} + b_{2i} \times \left(\text{RACE\_d} \right)_{ij} + b_{3i} \times \left(\text{SEXxRACE} \right)_{ij} + e_{ij}$$

Level-2 model:

$$b_{0i} = \beta_0 + u_{0i}$$
$$b_{1i} = \beta_1$$
$$b_{2i} = \beta_2$$
$$b_{3i} = \beta_3$$

2.4.5.2 Edit the existing syntax file

To obtain the model discussed above, we can either go through the multilevel generalized linear model dialog boxes as we did in the previous models or modify the existing syntax file directly. Here, we will illustrate how to modify the syntax file generated in the previous example.

First, open the syntax file for the previous model and save it under a different name so as not to overwrite the syntax file associated with the previous analysis. To do so, click on the **File, Open** option on the LISREL main window. Keep the **Syntax Only (*.spl, *.lis, *.prl)** dropdown list unchanged. Browse for the saved syntax file (**depress2.prl**) of the previous model. Double click on the **depress2.prl** to open it. Select the **File**, **Save As** option, and provide a new name (**depress3.prl**) for the model specification file.

As a first step, give an appropriate title by changing the Title line. Next, change the probit link function to a complementary log-log link function by modifying the syntax line Link=PROBIT to Link = CLL to produce the following syntax file. The CLL keyword refers to the complementary log-log link. For all the syntax available in MGLIM, please refer to Section 2.2.3.

Run the analysis by selecting the **Run PRELIS** button to generate the output file **depress3.out**, which is saved to the same folder as the syntax file.

2.4.5.3 Discussion of results

Portions of the output file **depress3.out** are shown below.

Model and data description

The model and data descriptions clearly show that complementary log-log (CLL) link function is used in the model.
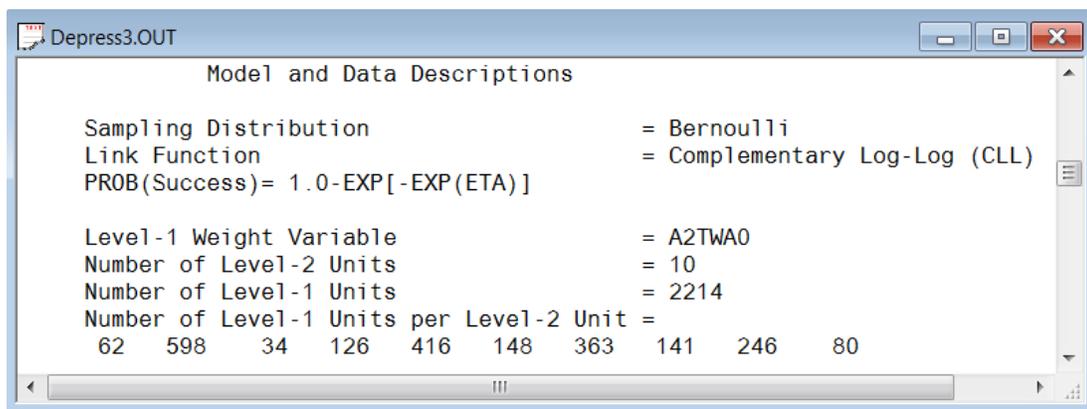
```
Depress3.OUT                                                    _  □  ✕
            Model and Data Descriptions

   Sampling Distribution                    = Bernoulli
   Link Function                            = Complementary Log-Log (CLL)
   PROB(Success)= 1.0-EXP[-EXP(ETA)]

   Level-1 Weight Variable                  = A2TWA0
   Number of Level-2 Units                  = 10
   Number of Level-1 Units                  = 2214
   Number of Level-1 Units per Level-2 Unit =
      62   598    34   126   416   148   363   141   246    80

◄                               III                            ►
```

Results for the model with fixed and random effects

Number of iterations and fit statistics

Six iterations were needed to obtain convergence. The likelihood ratio test, Pearson chi-square, and Akaike's and Schwarz's criteria are given after to the iteration number.

```
Depress3.OUT                                               _  □  ✕
  o=======================================o
  | Optimization Method: Adaptive Quadrature |
  o=======================================o

     Number of quadrature points =               10
     Number of free parameters =                  5
     Number of iterations used =                  2

     -2lnL (deviance statistic) =          2891.16332
     Akaike Information Criterion          2901.16332
     Schwarz Criterion                     2929.67610

◄                       III                               ►
```
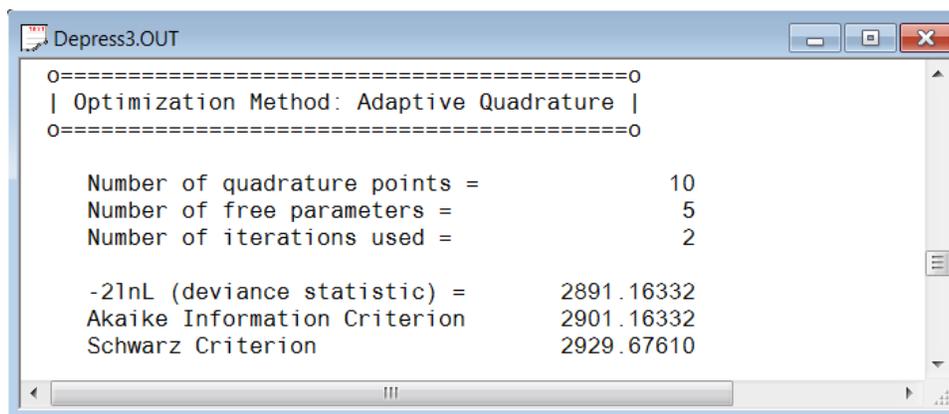
Estimated regression weights

The output describing the estimated regression weights is shown next. The estimates are shown in the column with heading Estimate and correspond to the coefficients β_0, β_1, β_2 and β_3 in the model specification. From the z-values and associated exceedance probabilities, we see that the intercept and the regression weight for SEXxRace is not significant at a 10% level of significance.

```
Depress3.OUT                                               [ - ][ □ ][ X ]

                    Estimated regression weights

                                Standard
      Parameter      Estimate      Error      z Value    P Value
      ---------      --------    --------     -------    -------
      intcept        -0.4403       0.1397     -3.1525     0.0016
      sex             0.5005       0.0771      6.4902     0.0000
      race_d         -0.5452       0.1048     -5.2029     0.0000
      SEXxRace        0.1646       0.1678      0.9808     0.3267


             Estimated level 2 variances and covariances

                                          Standard
      Parameter               Estimate      Error     z Value   P Value
      ---------               --------    --------    -------   -------
      intcept/intcept           0.1714      0.0852     2.0124    0.0442


    | Calculation of the intracluster correlation
    -------------------------------------------------
    residual variance = pi*pi / 6 (assumed)
    cluster  variance =    0.1714

    intracluster correlation =    0.1714 / (    0.1714 + (pi*pi/6)) = 0.094

                              |||
```

The estimated coefficient for the intercept is negative in value in this model. The estimated coefficient associated with gender (SEX) is – 0.5508. The estimate for the ethnicity indicator (RACE_D) also shows that white clients are likely to have a higher $\hat{\eta}$ value. The coefficient representing the iinteraction of gender and ethnicity is also negative and implies a decrease in $\hat{\eta}$. To transform these results into probabilities, we use the complementary log-log link function.

Interpreting estimated regression weights by using the link function

First, we substitute the regression weights and obtain an expression for $\hat{\eta}_{ij}$:

$$\hat{\eta}_{ij} = \hat{b}_{0i} + \hat{b}_{1i} \times (\text{SEX})_{ij} + \hat{b}_{2i} \times (\text{RACE\_d})_{ij} + \hat{b}_{3i} \times (\text{SEXxRACE})_{ij}$$
$$= -0.4403 + 0.5005 \times (\text{SEX})_{ij} - 0.5452 \times (\text{RACE\_d})_{ij} + 0.1646 \times (\text{SEXxRACE})_{ij} .$$

For a black males, we have SEX = 0, RACE_d = 0 and SEXxRace = 0 thus

$$\hat{\eta}_{ij} = -0.4403 .$$

Similarly, the calculation of $\hat{\eta}_{ij}$ for black females (SEX = 1, RACE_d = 0 and SEXxRace = 0) is

$$\hat{\eta}_{ij} = -0.4403 + 0.5005$$
$$= 0.0602 \qquad .$$

Next, we transform the $\hat{\eta}_{ij}$'s into the corresponding probabilities by using the complementary log-log link function. Taking the black males as an example, we calculate their probability of being depressed as

$$\text{Prob}(\text{DEPR}_{ij} = 1) = 1\text{-}\exp\left(-\exp\left(\eta_{ij}\right)\right) = 61.18\%$$

The probabilities of having depression for different gender and ethnicity groups are reported in the following table.

Group	Code	$\hat{\eta}$	Prob (DEPR = 1)
Black, male	sex = 0, race_d = 0	-0.4403	47.47%
Black, female	sex = 1, race_d = 0	0.0602	65.43%
White, male	sex = 0, race_d = 1	-0.9855	31.15%
White, female	sex = 1, race_d = 1	-0.3204	51.61%

In all three binary models, even though the estimated $\hat{\eta}$'s are different, the estimated probabilities are close to each other in all groups.

The three models we discussed in this section demonstrated the use of link functions available in multilevel analysis with MGLIM. In the next section, we will discuss the MGLIM functions available for count outcomes.

2.4.6 Models for count outcomes from the NESARC data

2.4.6.1 Count variable and its distributions

A count variable is used to count a number of discrete occurrences that take place during a time interval. For example, the occurrence of cancer cases in a hospital during a given period of time, the number cars that pass through a toll station per day and the phone calls at a call center are all count variables.

The most common distribution for a count variable is the Poisson distribution. Besides the Poisson distribution, the negative binomial distribution is also used to model count variables. In this guide models for the Poisson and negative binomial distributions will be discussed.

Poisson distribution

Poisson distribution is a discrete probability distribution. It is an appropriate distribution to express the probability of a number of events occurring in a fixed time period with a known average rate, and are independent of time. The probability with k occurrences is

$$f(k;\lambda) = \frac{e^{-\lambda}\lambda^k}{k!} \quad for \quad k = 0,1,2\ldots$$

where k is a non-negative integer and λ is a positive real number, which equals the expected number of occurence during the given interval. The cumulative probability function is

$$\Pr(k;\lambda) = \sum_{i=0}^{k} \frac{e^{-\lambda}\lambda^i}{i!} \quad for \quad k = 0,1,2\ldots$$

with the single parameter λ. A Poisson distribution has an important property: the mean number of occurrences λ equals the variance $E(f) = \mathrm{var}(f) = \lambda$.

The smaller the value of λ, the more skewed the probability distribution becomes. When λ is large, the Poisson distribution is close to the normal distribution.

Negative binomial distribution

The negative binomial distribution is the probability distribution of the number of failures before the r-th success in a Bernoulli process, with probability p of success on each trial.

Log link function

The log link function is generally used for the Poisson distribution. Assume the response measurements for a count variable $y_1,\ldots,y_n$ are independent and

$$y_i \sim Poi(\lambda_i), \quad where \quad \lambda_i = e^{\beta_1 x_{i1}+\cdots+\beta_p x_{ip}}$$

To make inference on the unknown parameters, we take the natural logarithm on the above equation.

$$\log(\lambda_i) = \beta_1 x_{i1} + \ldots + \beta_p x_{ip}$$

2.4.6.2 The data

The data set is from the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC), which was designed to be a longitudinal survey with its first wave fielded in 2001–2002. This data contains information on the occurrences of major depression, family history of major depression and dysthymia of 2339 dysthymia respondents. After list-wise deletion, the sample size is 1981.

The variables of interest are:

- o PSU denotes the Census 2000/2001 Supplementary Survey (C2SS) primary sampling unit.
- o FINWT represents the NESARC weights sample results used to form national level estimates. The final weight is the product of the NESARC base weight and other individual weighting factors.
- o CONCENTR contains the information captured in field S4CQ3A6 of the NESARC data. It represents the response to the statement "Often had trouble concentrating/keeping mind on things," with 1 indicating "Yes," and 0 indicating "No."
- o AGE_ONS is based on field S4CQ7AR of the NESARC data. It represents the age at onset of first episode.
- o N_DEP is recoded from field S4CQ6A of the NESARC data, and gives the number of depression/dysthymia episodes. This is the count variable we would like to use as outcome variable in the examples to follow.

2.4.6.3 Exploring the data

Inspecting the distribution of the intended outcome variable, N_DEP, before starting with the model is important. The number of depression episode ranges from 1 to 29, with most respondents having a small number of reported episodes of depression.

2.4.6.4 The model

The first model fitted to the data explores the relationship between N_DEP and the variables indicating concentration (or lack thereof) and age, as represented by the variables CONCENTR and AGE_ONS.

The level-1 model is

$$\log\left(\lambda_{ij}\right) = \beta_0 + \beta_1 \times \text{CONC\_DEP}_{ij} + \beta_2 \times \text{AGE\_DEP}_{ij}$$

where the expected number of depression episodes is $\lambda_{ij} = E\left(\text{N\_DEP}_{ij}\right)$.

The level-2 model is

$$\beta_0 = b_{00} + v_{i0}, \ \beta_1 = b_{10} \quad \text{and} \quad \beta_2 = b_{20}.$$

Another way of writing the combined model is

$$\log(\lambda_{ij}) = b_{00} + b_{10} \times \text{CONC\_DEP}_{ij} + b_{20} \times \text{AGE\_DEP}_{ij} + v_{i0}.$$

In this model, $e^{b_{00}}$ denotes the average expected count of depression episodes, and b_{10} represents the estimated coefficient for the respondent's level of concentration.

Taking exponents on both sides, we also have

$$\lambda_{ij} = e^{b_{00} + b_{10} \times \text{CONC\_DEP}_{ij} + b_{20} \times \text{AGE\_DEP}_{ij} + v_{i0}}$$
$$= e^{b_{00}} e^{b_{10} \times \text{CONC\_DEP}_{ij}} e^{b_{20} \times \text{AGE\_DEP}_{ij}} e^{v_{i0}}$$

For a person who had problems concentrating ($\text{CONCENTR} = 1$), the expected average number of episodes $e^{b_{00}}$ is multiplied by e^{β_1}, compared to an expected count of $e^{b_{00}}$ for a person for whom $\text{CONCENTR} = 0$. Similarly, an increase of one year in age increases the estimated number of episodes by a factor of $e^{b_{20}}$. For example, a respondent with concentration problems who is two years older than another respondent who had no concentration problems is expected to have $e^{b_{00}} e^{b_{10}} e^{2b_{20}}$ episodes compared to only $e^{b_{00}}$ episodes for the younger person without concentration problems.

The random part of the model is represented by $e^{v_{i0}}$, which denotes the variation in average count of depression episodes over PSU and between respondents (or, in other words, over respondents nested within PSU). For a Poisson distribution, the assumption of normality at level 1 is not realistic, as the level-1 random effect can only assume a number of distinct values. Thus, this random effect cannot have homogeneous variance.

2.4.6.5 Setting up the analysis

Open the LISREL data spreadsheet file **nesarc_poi.lsf** and select the **Multilevel, Generalized Linear Model** option from the main menu bar as shown below.
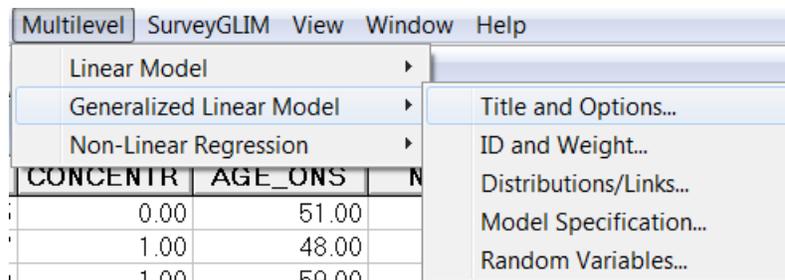
Proceed to fill in the **Title and Options** (number of quadrature points is 6); **ID and Weight** (Level-2 ID is PSU); **Distributions/Links** (Poisson, log-link); **Model Specification** (Dependent variable is N_DEP, predictors are intcept, CONCENTR and AGE_ONS); and the **Random Variables** dialog (Intercepts only). When done, click the **Finish** button to create the syntax file **nesarc_poi.prl**. Save this file as **nesarc_poi1.prl** using the **File, Save As** option.

```
L nesarc_poi1.PRL                                          _ □ ✖

MGlimOptions Converge=0.0001 MaxIter=100 MissingCode=-999999
             Method=Quad NQUADPTS=10 ;
Title=Random Intercept Poisson Model fitted to NESARC data;
SY='nesarc_poi.LSF';
ID2=PSU;
Distribution=POI;
Link=LOG;
Intercept=Yes;
Scale=None;
DepVar=N_DEP;
CoVars=CONCENTR AGE_ONS;
RANDOM2=intcept;
```

2.4.6.6 Discussion of results

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

Model and data description

A description of the hierarchical structure follows the syntax: data from a total of 395 PSUs and 1981 respondents were included at levels 2 and 1 of the model. In addition, an enumeration of the number of respondents nested within each of the 395 PSUs is provided.

```
nesarc_poi1.OUT                                          _ □ ✖

o====================================================o
| Random Intercept Poisson Model fitted to NESARC data |
o====================================================o
|
           Model and Data Descriptions

   Sampling Distribution                = Poisson
   Link Function                        = Log
   Number of Level-2 Units                395
   Number of Level-1 Units                1981
   Number of Level-1 Units per Level-2 Unit =
     8    1    1    5    5    5    1    5    1    5    2    5
     3   31   16    3    1    7    5    3    2    1    9    1
     7    6    3   22    8    1    8    2    1    1    1    2
     5   51    8   25    8   10    4    1    4    4   10    2
    19    7    5    2    2   10    7    3    1    6    6    1
     3    8    4    3   10    2    4    2    1    6    2    1
    16   18    5    3    7    3    1    6    4    8    5    3
```

Descriptive statistics

The data summary is followed by descriptive statistics for all the variables included in the model. The mean of 1.8970 and standard deviation of 2.3304 are reported for the outcome N_DEP indicating that, on average, 1.8970 episodes of depression were recorded.
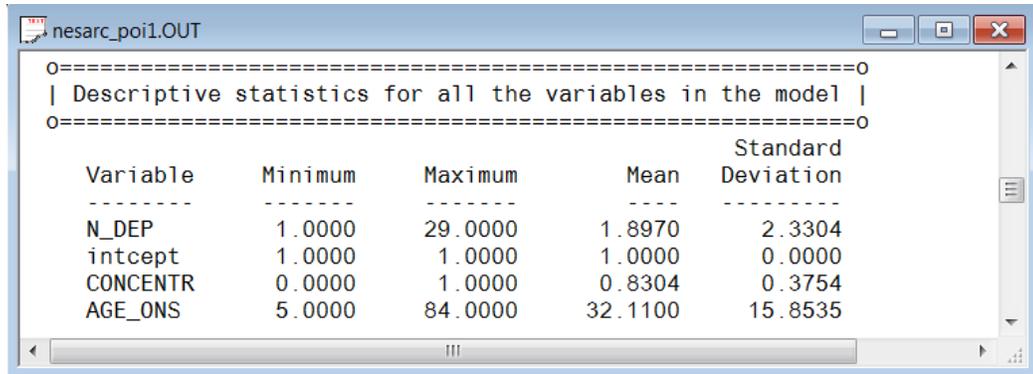
```
nesarc_poi1.OUT                                                  ▢ ▣ ✖

O=============================================================O
| Descriptive statistics for all the variables in the model |
O=============================================================O
                                                      Standard
     Variable      Minimum      Maximum       Mean    Deviation
     --------      -------      -------       ----    ---------
     N_DEP          1.0000      29.0000      1.8970      2.3304
     intcept        1.0000       1.0000      1.0000      0.0000
     CONCENTR       0.0000       1.0000      0.8304      0.3754
     AGE_ONS        5.0000      84.0000     32.1100     15.8535
```

Descriptive statistics are followed by the results for a fixed-effects-only model, *i.e.* a model without random coefficients.

Fixed effects results

At the top of the final results, the number of iterations required for convergence of the iterative procedure is given. Next, the number of quadrature points per dimension is reported which, in this case, is the default number of points. The log likelihood and the deviance, which is defined as $-2\ln L$, are listed next. For a pair of nested models, the difference in $-2\ln L$ values has a χ^2 distribution, with degrees of freedom equal to the difference in number of parameters estimated in the models compared.
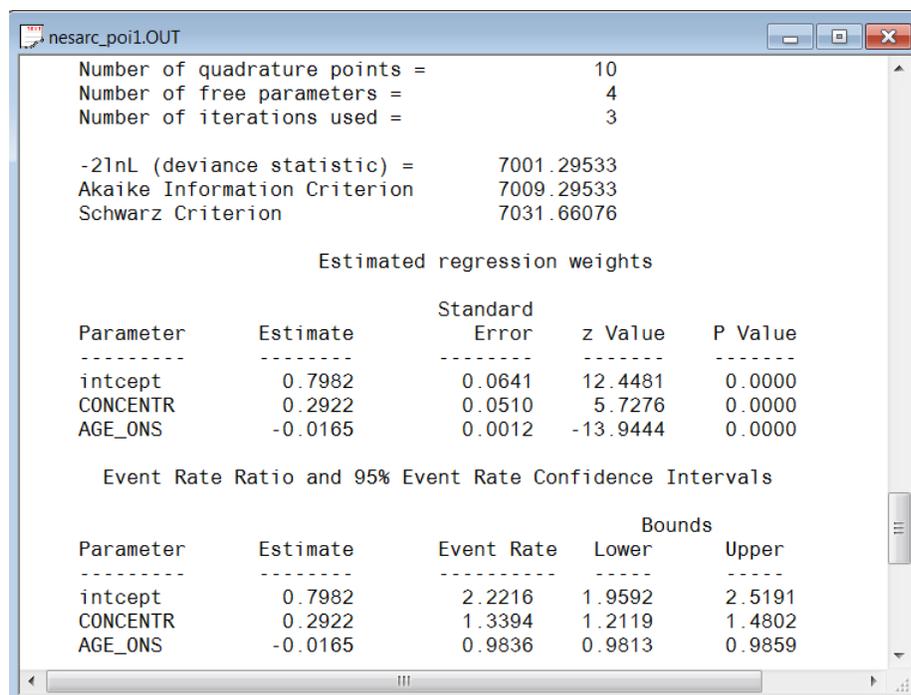
```
nesarc_poi1.OUT                                                  ▢ ▣ ✖

     Number of quadrature points =              10
     Number of free parameters =                 4
     Number of iterations used =                 3

     -2lnL (deviance statistic) =        7001.29533
     Akaike Information Criterion        7009.29533
     Schwarz Criterion                   7031.66076

                    Estimated regression weights

                                Standard
     Parameter     Estimate       Error     z Value     P Value
     ---------     --------     --------     -------     -------
     intcept         0.7982       0.0641     12.4481      0.0000
     CONCENTR        0.2922       0.0510      5.7276      0.0000
     AGE_ONS        -0.0165       0.0012    -13.9444      0.0000

        Event Rate Ratio and 95% Event Rate Confidence Intervals

                                                    Bounds
     Parameter     Estimate    Event Rate    Lower      Upper
     ---------     --------    ----------    -----      -----
     intcept         0.7982       2.2216     1.9592     2.5191
     CONCENTR        0.2922       1.3394     1.2119     1.4802
     AGE_ONS        -0.0165       0.9836     0.9813     0.9859
```

The estimated intercept is 0.7982, which means that the average number of depression episodes is $e^{0.7982}=2.2215$, implying that on average the number of episodes is about two. The estimated coefficient for CONCENTR is 0.2922, which indicates that respondents who had trouble concentrating on things tended to have $2.2215e^{0.2922}=(2.2215)(1.3394)=2.9754$ episodes at the same age as respondents without concentration problems. The estimate of the effect of age at the onset of the first episode (AGE_ONS) shows that the onset age does not affect the number of episodes much, since $e^{-0.0165}=0.98$. A slight reduction in the expected number of episodes is expected with increasing age. If one compares two typical respondents with reported concentration problems, but with one respondent ten years older than the other, one would expect the older respondent to have $(2.2215)(1.3394)e^{10(-0.0165)}=2.5229$ episodes, compared to 2.9268 expected episodes for the younger respondent. In other words, the longer it takes for the first episode to occur, the fewer episodes a respondent is expected to have. Of course, it has to be kept in mind that the younger a respondent is at the first episode, the longer that person must live with the condition and thus the more time there is for subsequent episodes to occur.

Random effects results

The output for the level-2 random effect variance term follows next. The estimated variation in the average estimated N_DEP at level 2 is 0.1347, which is highly significant. Respondents are different in terms of their average expected number of episodes, holding all other variables constant.

```
nesarc_poi1.OUT                                                    [ _ ][ □ ][ X ]
                    Estimated level 2 variances and covariances

                                        Standard
        Parameter              Estimate     Error    z Value   P Value
        ---------              --------  --------    -------   -------
        intcept/intcept          0.1347    0.0184     7.3058    0.0000


    Level 2 covariance matrix

                    intcept
        intcept    0.134671
```

Level-1 variation for Poisson distribution

The variance-to-mean ratio is a measure of the dispersion of a probability distribution:

$$R = \text{variance-to-mean ratio} = \frac{\sigma^2}{\mu}$$

For the Poisson distribution, where the variance equals the mean, this implies $R=1$. Thus, we use a value of one as our level-1 variation. In the cases when over-dispersion ($R>1$) or under-dispersion ($R<1$) is assumed, different level-1 variation values will apply. The details of these scenarios are not discussed in this guide.

2.4.6.7 Interpreting the results

Estimated outcomes for groups: unit-specific results

First, we substitute the regression weights and obtain the following function for $\log\left(\text{N\_DEP}_{ij}\right)$:

$$\log\left(\text{N\_DEP}_{ij}\right) = \hat{b}_{00} + \hat{b}_{10}\times\text{CONC\_DEP}_{ij} + \hat{b}_{20}\times\text{AGE\_DEP}_{ij}$$
$$= 0.7982 + 0.2922\times\text{CONC\_DEP}_{ij} - 0.0165\times\text{AGE\_DEP}_{ij}.$$

For example, at age 40, the estimated $\log\left(\text{N\_DEP}_{ij}\right)$ for a typical respondent who does not often have trouble concentrating (CONCENTR = 0), we find that

$$\log\left(\text{N\_DEP}_{ij}\right) = \hat{b}_{00} + \hat{\beta}_{1}\times\text{CONC\_DEP}_{ij} + \hat{\beta}_{2}\times\text{AGE\_DEP}_{ij}$$
$$= 0.7982 + 0.2922\times\text{CONC\_DEP}_{ij} - 0.0165\times\text{AGE\_DEP}_{ij}$$
$$= 0.7982 + 0.2922\times 0 - 0.0165\times 40$$
$$= 0.1382.$$

Keeping in mind that we defined the relationship between λ and the predictors as

$$\log\left(\lambda_{ij}\right) = \beta_{1}x_{i1} + \ldots + \beta_{p}x_{ip},$$

it follows that

$$\hat{\lambda}_{ij} = e^{0.1382} = 1.1482.$$

We can estimate the count of the occurrence of depression episodes for typical individuals of different ages in the same way. Results are summarized in the table below. The results show a decrease in the expected number of episodes with increasing age, regardless of whether they had concentration problems or not.

Estimated number of episodes under the Poisson log model

AGE_ONS	10	20	30	40	50	60	70
CONCENTR = 1	2.5229	2.1391	1.8138	1.5379	1.3040	1.1056	0.9374
CONCENTR = 0	1.8836	1.5971	1.3542	1.1482	0.9736	0.8255	0.6999

We clearly see that the correspondents who often had trouble concentrating (CONCENTR = 1) have a higher estimated number of depression episodes. It also shows that the number of episodes is expected to decrease as people get older.

Level 2 ICC

The percentage of variance explained over level-2 units, or intraclass correlation coefficient (ICC), is calculated as

$$ICC = \frac{\text{level-2 variation}}{\text{level-1 variation} + \text{level-2 variation}}$$

In this example, under the assumption that the level-1 variation is fixed at a value of one, we have

$$ICC = \frac{0.1347}{1 + 0.1347} \times 100\% = 11.8\%$$

We can conclude that most of the unexplained variation in the outcome (approximately 78%) is between measurements at the lowest level of the hierarchy.

2.4.7 Negative binomial model for the NESARC data

2.4.7.1 The model

In the previous section a Poisson model was fitted to the data. It was also noted that a Poisson distribution has an important property: the mean number of occurrences is equal to the variance. The negative binomial distribution can be used as an alternative to the Poisson distribution. It is especially useful for discrete data that assumes values 0, 1, 2, 3… whose sample variance exceeds the sample mean. In such cases, the observations are over-dispersed with respect to a Poisson distribution, for which the mean is equal to the variance. Since the negative binomial distribution has one more parameter than the Poisson, the second parameter can be used to adjust the variance independently of the mean. It can be shown that a model based on the negative binomial distribution with a dispersion parameter close to zero will produce results that correspond closely to those obtained for the Poisson model. In this section, we fit a negative binomial model, utilizing the same predictors and a small dispersion parameter, to the NESARC data. Subsequently larger values of the dispersion parameter will be used to study the impact on parameter estimates and the deviance statistic. Again, adaptive quadrature is used as the method of optimization.

The negative binomial distribution can be expressed as

$$f(y_i) = \frac{\Gamma(y_i + 1/\alpha)}{\Gamma(y_i + 1)\Gamma(1/\alpha)} \times \frac{(\alpha\mu_i)^{y_i}}{(1 + \alpha\mu_i)^{y_i + 1/\alpha}}$$

with $\sigma^2(y_i) = \mu_i + \alpha\mu_i^2$ where α denotes an additional parameter and it can no longer be assumed that the variance is a known function of the mean. We assume α to be a fixed parameter.

The model fitted to the data explores the relationship between N_DEP and the variables indicating concentration (or lack thereof) and age, as represented by the variables CONCENTR and AGE_ONS.

The level-1 model is

$$\log\left[E\left(N\_DEP_{ij}\right)\right] = \beta_0 + \beta_1 \times CONC\_DEP_{ij} + \beta_2 \times AGE\_DEP_{ij}$$

The level-2 model is

$$\beta_0 = b_{00} + v_{i0}, \; \beta_1 = b_{10} \; \text{ and } \; \beta_2 = b_{20}.$$

2.4.7.2 Setting up the analysis

Make sure that the spreadsheet **nesarc_poi.lsf** is the active window and keeping all the other settings for a **Multilevel**, **Generalized Linear Model** unchanged, set the **Distribution Model** to **negative binomial**, and the **Dispersion Parameter** to 0.0001 as shown below.

Save the revised syntax file as **nesarc_poi2.prl**, and click the **PRELIS**, **Run** icon button to start the iterative process.

2.4.7.3 Discussion of results

Portions of the output file **nesarc_poi2.out** are shown below.

Fixed and random effect results

The estimated regression coefficients for fixed effects in the model are shown below. Recall that the estimated coefficients of the intercept, CONCENTR, and AGE_ONS under the Poisson model reported in the previous section were 0.7982, 0.2922, and -0.0165 respectively. The estimated variation in the average estimated N_DEP at level-2 was 0.1347, and highly significant. Note that with the dispersion parameter set at 0.0001, results of the two models are almost identical.

```
nesarc_poi2.OUT

   Number of quadrature points =          10
   Number of free parameters =            4
   Number of iterations used =            3

   -2lnL (deviance statistic) =       7000.81243
   Akaike Information Criterion       7008.81243
   Schwarz Criterion                  7031.17786

                 Estimated regression weights

                              Standard
   Parameter     Estimate       Error    z Value   P Value
   ---------     --------     --------   -------   -------
   intcept        0.7982       0.0641    12.4471    0.0000
   CONCENTR       0.2922       0.0510     5.7270    0.0000
   AGE_ONS       -0.0165       0.0012   -13.9420    0.0000

      Event Rate Ratio and 95% Event Rate Confidence Intervals

                                           Bounds
   Parameter     Estimate    Event Rate   Lower     Upper
   ---------     --------    ----------   -----     -----
   intcept        0.7982       2.2216     1.9592    2.5191
   CONCENTR       0.2922       1.3394     1.2119    1.4802
   AGE_ONS       -0.0165       0.9836     0.9813    0.9859

         Estimated level 2 variances and covariances

                                 Standard
   Parameter          Estimate     Error    z Value   P Value
   ---------          --------   --------   -------   -------
   intcept/intcept      0.1346     0.0184    7.3039    0.0000
```

2.4.7.4 Changing the value of the dispersion parameter

Save the syntax file for the negative binomial model as **nesarc_poi3.prl** and change the dispersion value to 0.1 before running the analysis and making a note of the deviance value (6760.73 in this case). Repeat the procedure for dispersion values of 0.2, 0.3, 0.4, 0.5 and 0.6 respectively. The list of (dispersion; deviance)-values are shown below in the LISREL data file **dispersion.lsf**.

dispersion.LSF	DISP	DEVIANCE	
1	0.000	7000.810	
2	0.100	6760.730	
3	0.200	6715.200	
4	0.300	6737.190	
5	0.400	6791.270	
6	0.500	6862.620	
7	0.600	6943.790	

From the main menu-bar select the **Graphs, Bivariate Plots** option to request a line plot of the deviance-statistic against dispersion.
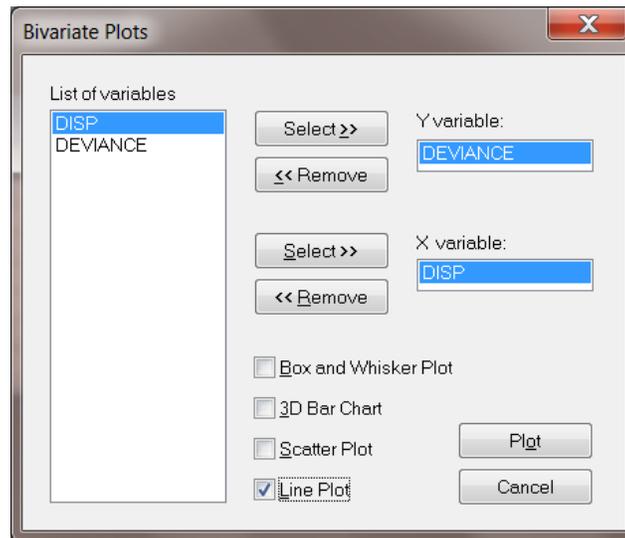
The plot shown below, shows that the best model fit is obtained for a dispersion-value of 0.2 (best fit corresponds to smallest deviance value).
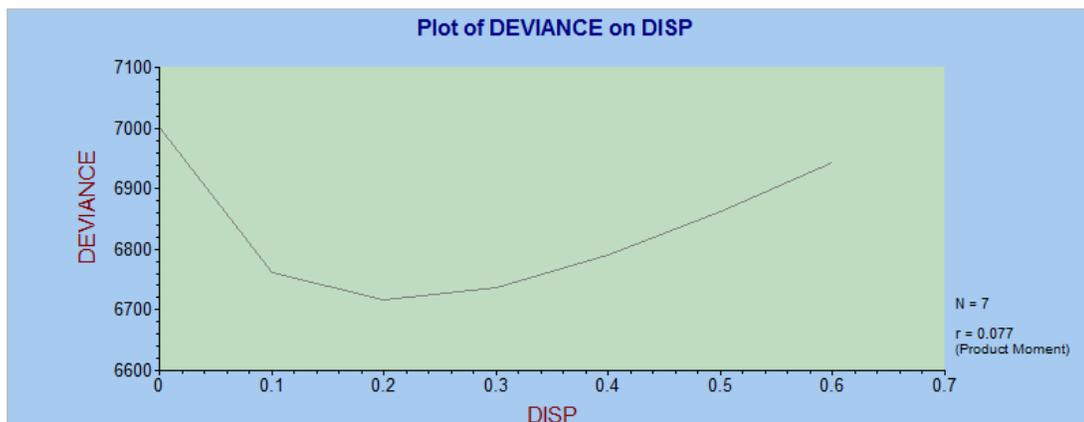
The output for the fixed and random effects is listed below. All the parameter estimates are significant, but different from those reported for the Poisson model.

```
nesarc_poi2.OUT

    Number of quadrature points =           10
    Number of free parameters =              4
    Number of iterations used =              2

    -2lnL (deviance statistic) =      6715.20779
    Akaike Information Criterion      6723.20779
    Schwarz Criterion                 6745.57322

                  Estimated regression weights

                           Standard
    Parameter     Estimate      Error    z Value    P Value
    ---------    ---------  ---------    -------    -------
    intcept         0.8075     0.0709    11.3889     0.0000
    CONCENTR        0.2836     0.0577     4.9117     0.0000
    AGE_ONS        -0.0154     0.0014   -11.3724     0.0000

      Event Rate Ratio and 95% Event Rate Confidence Intervals

                                          Bounds
    Parameter     Estimate   Event Rate   Lower      Upper
    ---------    ---------   ----------   -----      -----
    intcept         0.8075       2.2422   1.9513     2.5765
    CONCENTR        0.2836       1.3279   1.1858     1.4870
    AGE_ONS        -0.0154       0.9847   0.9821     0.9873

            Estimated level 2 variances and covariances

                                 Standard
    Parameter          Estimate     Error   z Value   P Value
    ---------         ---------  --------   -------   -------
    intcept/intcept      0.0709    0.0155    4.5810    0.0000
```

2.4.8 Weighted 2-level models

2.4.8.1 The data

The sampling frame of many multistage surveys frequently entails selection of units with known, but unequal, selection probabilities. This situation is the result of a number of design factors, of which the cost of doing the survey is an important consideration. When this is the case, it is appropriate to weigh observations in order to produce unbiased estimates of the population parameters.

The variable FINWT represents the NESARC weights sample results used to form national-level estimates. The final weight is the product of the NESARC base weight and other individual weighting factors. In this section, we explore the effect of inclusion of the weights on the results obtained in the section describing the unweighted Poisson model.

2.4.8.2 Setting up the analysis

The models remain the same, with only the selection of the weight variable on the **Advanced** tab of the **Model Specification** screen to be added. Below, we show how this is done in the case of the Poisson distribution model.

Open the syntax file for the Poisson distribution model (**nesarc_poi1.prl**). Save the syntax file file as **nesarc_poi3.prl** and add Weight = FINWT below the **ID2** paragraph as shown below.

```
MGlimOptions Converge=0.0001 MaxIter=100 MissingCode=-999999
             Method=Quad NQUADPTS=10 ;
Title=Random Intercept Poisson Model fitted to NESARC data;
SY='nesarc_poi.LSF';
ID2=PSU;
Weight=FINWT;
Distribution=POI;
Link=LOG;
Intercept=Yes;
Scale=None;
DepVar=N_DEP;
CoVars=CONCENTR AGE_ONS;
RANDOM2=intcept;
```

Save the file and run the analysis.

2.4.8.3 Discussion of results

Results for this analysis are reported in the table below. The results from the unweighted Poisson distribution model are included in order to facilitate evaluation of the impact of the weights on the results.

Comparison of results for weighted and unweighted Poisson models

Parameter	Unweighted model		Weighted model	
	Estimate	Standard error	Estimate	Standard error
intcept	0.7982	0.0641	0.7225	0.0660
CONCENTR	0.2922	0.0510	0.3055	0.0532
AGE_ONS	–0.0165	0.0012	–0.0156	0.0013
Level-2 variance	0.1347	0.0184	0.1378	0.1089

Results for the two models are very similar, and interpretation of the results of both models would lead to the same conclusions, both in terms of significance and in terms of the expected number of depression episodes. However, this is more the exception than the rule – users are cautioned to use weight variables whenever they are available in order to prevent skewed or biased results that may occur when weights are excluded in the analysis of a disproportionally drawn sample.

2.4.9 Models for count outcomes using ASPART data

2.4.9.1 The data

The data for this example are taken from a paper by McKnight and Van Den Eeden (1993), who reported on the number of headaches in a two treatment, multiple period crossover trial. Specifically, the number of headaches per week was repeatedly measured for 27 patients. Following a seven day placebo run-in period, subjects received either aspartame or placebo in four seven-day treatment periods according to the double-blind crossover treatment design. Each treatment period was separated by a washout day. The sample size is 122. Data for the first 10 observations of all the variables used in this section, contained in the file **aspart.lsf** are shown below in the form of an LSF spreadsheet window.

ASPART.lsf							
	ID	HeadAche	DrugAsp	Period1	Period2	Period3	Period4
1	2.00	0.00	0.00	0.00	0.00	0.00	0.00
2	2.00	5.00	1.00	1.00	0.00	0.00	0.00
3	2.00	2.00	0.00	0.00	1.00	0.00	0.00
4	5.00	3.00	0.00	0.00	0.00	0.00	0.00
5	5.00	0.00	1.00	1.00	0.00	0.00	0.00
6	5.00	2.00	0.00	0.00	1.00	0.00	0.00
7	5.00	0.00	1.00	0.00	0.00	1.00	0.00
8	5.00	0.00	0.00	0.00	0.00	0.00	1.00
9	13.00	7.00	0.00	0.00	0.00	0.00	0.00
10	13.00	7.00	1.00	1.00	0.00	0.00	0.00

The variables of interest are:

- o ID is the patient ID (27 patients in total).
- o Headache is the number of headaches during the week (from 0 to 7).
- o Period1 is a period 1 treatment indicator (1 for the first treatment period and 0 otherwise).
- o Period2 is a period 2 treatment indicator (1 for the second treatment period and 0 otherwise).
- o Period3 is a period 3 treatment indicator (1 for the third treatment period and 0 otherwise).
- o Period4 is a period 4 treatment indicator (1 for the fourth treatment period and 0 otherwise).
- o DrugAsp indicates the type of drug being used for the treatment, (0 = placebo and 1 = aspartame). 75 observations used placebo and 47 used aspartame.
- o Nperiods is the number of periods the individual was observed (from 2 to 5).
- o NTDays is the number of treatment days in the period (from 1 to 7).

2.4.9.2 The model

To model the relationship between the number of headaches during the week (Headache) and the treatment indicators (Period1 to Period4) and the type of drug administered (DrugAsp), the following Poisson regression model with a random intercept may be used:

$$\log(\lambda_{ij}) = \beta_0 + \beta_1 \times \text{Period1}_{ij} + \beta_2 \times \text{Period2}_{ij} + \beta_3 \times \text{Period3}_{ij}$$
$$+ \beta_4 \times \text{Period4}_{ij} + \beta_5 \times \text{DrugAsp}_{ij} + v_{i0}$$

where λ_{ij} denotes the mean number of headaches of patient i for treatment period j; Period1$_{ij}$, Period2$_{ij}$, Period3$_{ij}$ and Period4$_{ij}$ denote the values of the dummy variables Period1, Period2, Period3 and Period4 for patient i for treatment period j respectively; DrugAsp$_{ij}$ denotes the value of the DrugAsp for patient i for treatment period j; β_0, β_1, β_2, β_3, β_4 and β_5 denote unknown parameters; and v_{i0} denotes the random intercept for patient i for $i = 1, 2, \ldots, 27$ and $j = 0, 1, 2, 3$. This model is fitted to the data in **aspart.ss3** as described below.

2.4.9.3 Setting up the analysis

Start by opening **aspart.lsf**. Select the **Multilevel, Generalized Linear Model** option on the main menu bar.
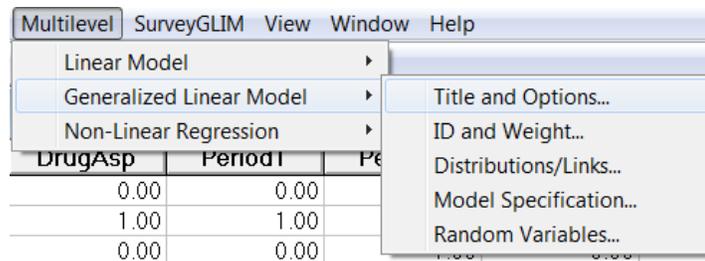
Using the **Titles and Options** dialog enter the title: 2-level Poisson log random intercept model, ASPART data. Select **Quadrature** and set the number of quadrature points to 20.
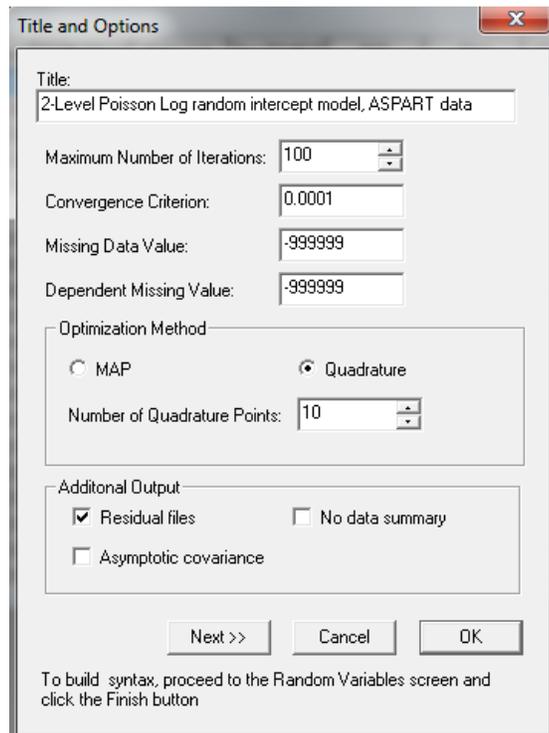
Click the **Next** button to activate the **ID and Weight** dialog and select the variable ID, which defines the second level of the hierarchy, as the **Level-2 ID.** To proceed to the **Distributions/Links** dialog, click the **Next** button and enter the values shown below.

Proceed to the **Dependent and Independent variable** dialog and select count outcome variable Headache as the dependent variable.

The variables Period1, Period2, Period3, Period4, and DrugAsp are specified as the independent variables of the model. By default, an intercept model is included in the fixed part of the model, along with a random intercept at level 2.
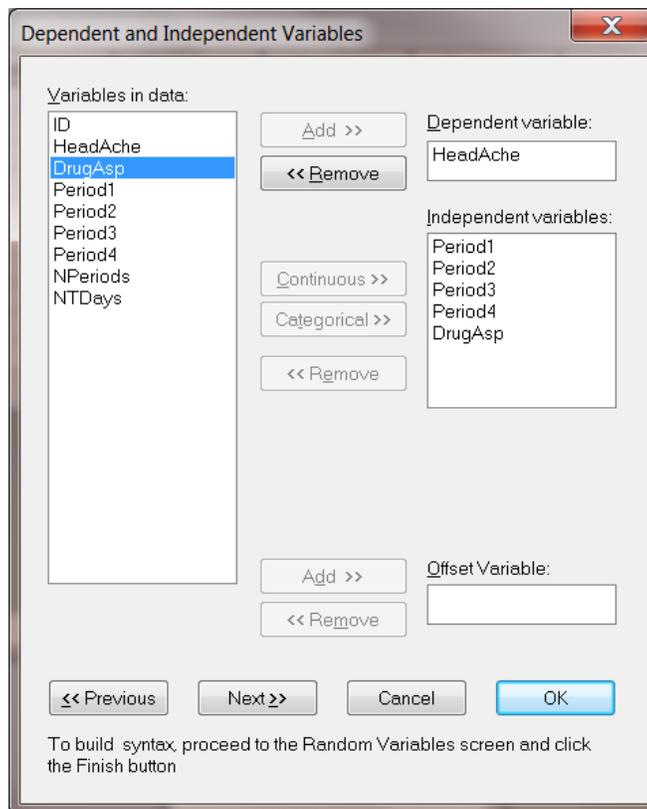
Click the **Finish** button on the **Random Variables** dialog to produce the syntax file shown below.

```
MGlimOptions Converge=0.0001 MaxIter=100 MissingCode=-999999
IterDetails=No Method=Quad NQUADPTS=10 Output=Residuals;
Title=Poisson - Log random intercept model, ASPART data;
SY=aspart.lsf;
ID2=ID;
Distribution=POI;
Link=LOG;
Intercept=Yes;
Scale=None;
DepVar=HeadAche;
CoVars=DrugAsp Period1 Period2 Period3 Period4 DrugAsp;
RANDOM2=intcept;
```

Before running the analysis, save the file as **aspart1.prl** and click the **RUN PRELIS** icon button to start the analysis.

2.4.9.4 Discussion of results

Portions of this output file are shown below.

Model and data description

The output file indicates that there are 27 subjects with 122 observations nested within them. The number of observations per subject varies between 2 and 5.

```
┌─────────────────────────────────────────────────────────────────┐
│ ⬚ ASPART1.OUT                                        ⬚ ▢ ✕       │
├─────────────────────────────────────────────────────────────────┤
│              Model and Data Descriptions                     ▲   │
│                                                                  │
│    Sampling Distribution                    = Poisson            │
│    Link Function                            = Log            ▤   │
│    Number of Level-2 Units                    27                 │
│    Number of Level-1 Units                    122                │
│    Number of Level-1 Units per Level-2 Unit =                    │
│       3    5    5    3    5    5    5    5    5    5    4    2     │
│       5    5    5    5    5    5    5    5    5    5    5    5     │
│       5    3    2                                                 │
│                                                             ▼    │
└─────────────────────────────────────────────────────────────────┘
```
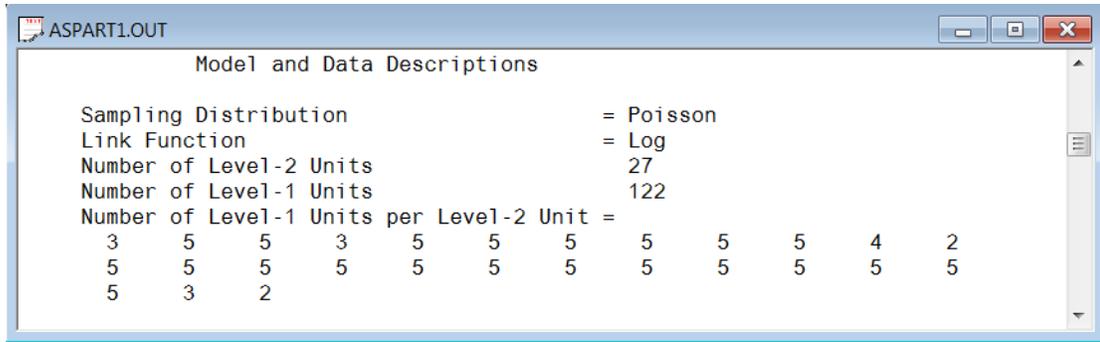
Descriptive statistics

The descriptive statistics for all the variables is shown next. The variance of Headache is $1.8863^2 = 3.5581$, which is substantially larger than the mean 1.6803. This might conflict with our assumption that the Poisson distribution is an appropriate choice for these data. As pointed out in Section 2.4.7, this can be verified by fitting a negative binomial model with a small dispersion parameter.

```
┌─────────────────────────────────────────────────────────────────┐
│ ⬚ ASPART1.OUT                                        ─ ▢ ✕       │
├─────────────────────────────────────────────────────────────────┤
│  0===========================================================0 ▲ │
│  | Descriptive statistics for all the variables in the model |   │
│  0===========================================================0   │
│                                                  Standard        │
│     Variable    Minimum    Maximum       Mean    Deviation   ▤   │
│     --------    -------    -------       ----    ---------        │
│     HeadAche    0.0000     7.0000      1.6803    1.8863           │
│     intcept     1.0000     1.0000      1.0000    0.0000           │
│     Period1     0.0000     1.0000      0.2213    0.4168           │
│     Period2     0.0000     1.0000      0.2049    0.4053           │
│     Period3     0.0000     1.0000      0.1803    0.3860           │
│     Period4     0.0000     1.0000      0.1721    0.3791           │
│     DrugAsp     0.0000     1.0000      0.3852    0.4887       ▼   │
│  ◄                        III                             ►      │
└─────────────────────────────────────────────────────────────────┘
```
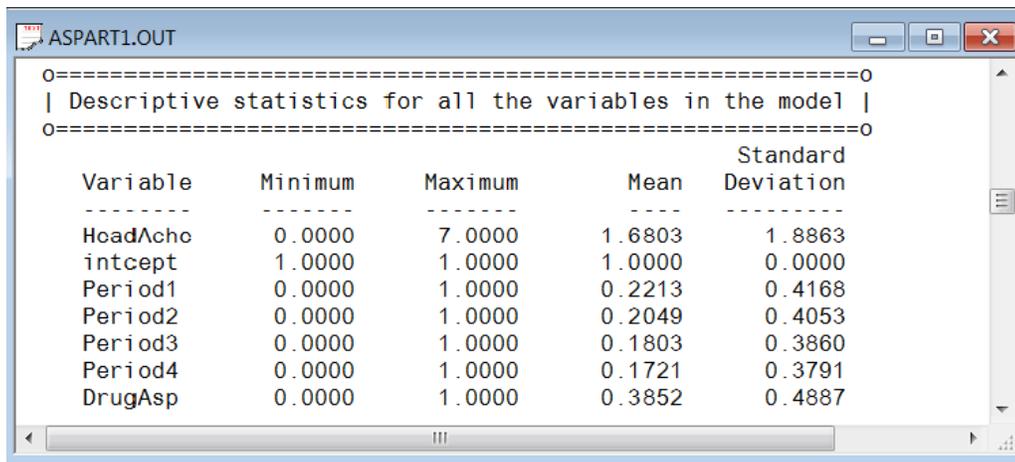
Results for the model without any random effects

The results for the model without any random effects are shown below. In this section the goodness of fit statistics, estimated regression weights and event rate ratio and 95% event rate confidence intervals are included.

```
ASPART1.OUT                                                           [ - ] [ □ ] [ X ]

   Statistic                                  Value        DF          Ratio
   ---------                                  -----        --          -----
   Likelihood Ratio Chi-square              243.8257       116         2.1019
   Pearson Chi-square                       253.8934       116         2.1887

                          Estimated regression weights

                               Standard
   Parameter      Estimate       Error     z Value     P Value
   ---------      --------     --------     -------     -------
   intcept         0.4654       0.1525      3.0516      0.0023
   Period1         0.0916       0.2265      0.4043      0.6860
   Period2         0.0131       0.2276      0.0575      0.9542
   Period3        -0.2245       0.2471     -0.9084      0.3637
   Period4        -0.1840       0.2540     -0.7242      0.4689
   DrugAsp         0.2332       0.1596      1.4612      0.1440

      Event Rate Ratio and 95% Event Rate Confidence Intervals

                                                Bounds
   Parameter      Estimate     Event Rate    Lower       Upper
   ---------      --------     ----------    -----       -----
   intcept         0.4654       1.5926      1.1811      2.1474
   Period1         0.0916       1.0959      0.7030      1.7085
   Period2         0.0131       1.0132      0.6486      1.5827
   Period3        -0.2245       0.7989      0.4923      1.2967
   Period4        -0.1840       0.8320      0.5057      1.3688
   DrugAsp         0.2332       1.2626      0.9235      1.7263
```

Fixed and random effect results

The final results are shown next. The number of iterations needed for convergence and the deviance information are given first, followed by the estimates.

The random-effect variance component is estimated as .4327 which is significant at the 5% level. Regarding the regression coefficients, all effects are non-significant. The results indicate that the model does not fit the data very well.

The event ratio and 95% event rate confidence interval and estimated level-2 variances and covariances are shown next to the estimated regression weights. The event ratios are the exponents ($e^{\hat{\beta}}$) of the estimated regression coefficients.

```
ASPART1.OUT                                                    - □ ✕

    Number of quadrature points =            20
    Number of free parameters =              7
    Number of iterations used =              3

    -2lnL (deviance statistic) =        406.35386
    Akaike Information Criterion        420.35386
    Schwarz Criterion                   439.98201

                    Estimated regression weights

                               Standard
    Parameter      Estimate      Error     z Value    P Value
    ---------      --------    --------    -------    -------
    intcept         0.2507      0.2051     1.2222     0.2216
    Period1         0.0806      0.2350     0.3429     0.7317
    Period2         0.0345      0.2236     0.1541     0.8775
    Period3        -0.2273      0.2546    -0.8927     0.3720
    Period4        -0.1594      0.2528    -0.6306     0.5283
    DrugAsp         0.2151      0.1639     1.3123     0.1894

       Event Rate Ratio and 95% Event Rate Confidence Intervals

                                              Bounds
    Parameter      Estimate    Event Rate   Lower      Upper
    ---------      --------    ----------   -----      -----
    intcept         0.2507      1.2849      0.8596     1.9208
    Period1         0.0806      1.0839      0.6838     1.7181
    Period2         0.0345      1.0351      0.6677     1.6045
    Period3        -0.2273      0.7967      0.4837     1.3122
    Period4        -0.1594      0.8526      0.5194     1.3995
    DrugAsp         0.2151      1.2399      0.8993     1.7096

            Estimated level 2 variances and covariances

                                   Standard
    Parameter              Estimate    Error     z Value   P Value
    ---------              --------   --------   -------   -------
    intcept/intcept         0.4327     0.1737    2.4917    0.0127
```

2.4.9.5 Interpreting the results

Estimated outcomes for groups: unit-specific results

The expected number of headaches can be obtained in the following fashion. First, we substitute the estimated coefficients in the model formulation

$$\log\left(\text{Head\hat{a}che}_{ij}\right) = \hat{\beta}_0 + \hat{\beta}_1 \times \text{Period1}_{ij} + \hat{\beta}_2 \times \text{Period2}_{ij}$$

$$+ \hat{\beta}_3 \times \text{Period3}_{ij} + \hat{\beta}_4 \times \text{Period4}_{ij} + \hat{\beta}_5 \times \text{DrugAsp}_{ij}$$

$$= 0.2572 + 0.0807 \times \text{Period1}_{ij} + 0.0345 \times \text{Period2}_{ij}$$

$$- 0.2267 \times \text{Period3}_{ij} - 0.1592 \times \text{Period4}_{ij} + 0.2151 \times \text{DrugAsp}_{ij}.$$

or, after taking exponents on both sides, as

$$\text{Head\hat{a}che}_{ij} = \exp(0.2572 + 0.0807 \times \text{Period1}_{ij} + 0.0345 \times \text{Period2}_{ij}$$

$$- 0.2267 \times \text{Period3}_{ij} - 0.1592 \times \text{Period4}_{ij} + 0.2151 \times \text{DrugAsp}_{ij}).$$

As an example, we calculate the expected number of headaches for a typical patient to whom aspartame was administered (DrugAsp = 1). During the first treatment period, we find that for such a patient

$$\hat{\text{Headache}}_{ij} = \exp(0.2572 + 0.0807 + 0.2151)$$
$$= 1.7385.$$

The expected numbers of headaches for a typical patient (again with DrugAsp = 1) for the second, third, and fourth treatment periods are calculated as

$$\hat{\text{Headache}}_{ij} = \exp(0.2572 + 0.0345 + 0.2151)$$
$$= 1.6600,$$

$$\hat{\text{Headache}}_{ij} = \exp(0.2572 - 0.2267 + 0.2151)$$
$$= 1.2784,$$

and

$$\hat{\text{Headache}}_{ij} = \exp(0.2752 - 0.1592 + 0.2151)$$
$$= 1.3677$$

respectively. Complete results for all groups are given in Table 5.2.

Estimated outcomes for groups: population-average results

The latent response variable model,

$$y_{ij} = \mathbf{z}_{(1)ij}' \mathbf{b}_i + \mathbf{x}_{(1)ij}' \boldsymbol{\beta}_{(1)} + e_{ij},$$

makes the assumption that $e_{ij} \sim LID(0, \sigma_e^2)$. For a Poisson distribution it is assumed that $\sigma_e^2 = 1$. Under the assumption that $\mathbf{v}_i$ and e_{ij} are independently distributed, it follows that

$$\sigma_{y_{ij}}^2 = \mathbf{z}_{ij}' \boldsymbol{\Phi}_{v_i} \mathbf{z}_{ij} + \sigma_e^2.$$

The design effect d_{ij} is defined as

$$d_{ij} = \frac{\sigma_{y_{ij}}^2}{\sigma_e^2},$$

which, for the current model, may be calculated as

$$d_{ij} = \frac{\sigma_{y_{ij}}^2}{\sigma_e^2} = \frac{\text{var}(v_{i0}) + 1}{1} = 1.4290$$

where $\text{var}(v_{i0}) = 0.4290$, with v_{i0} denoting the random intercept coefficient. The estimated population-average probabilities (Hedeker & Gibbons, 2006) are obtained in a similar fashion as the unit-specific probabilities, after replacing the exponent in the formula used for calculation of the estimated unit-specific probabilities with $\exp = \exp / \sqrt{d_{ij}}$ as shown below.

$$\hat{\text{Headache}}_{ij} = \exp[(0.2572 + 0.0807 \times \text{Period1}_{ij} + 0.0345 \times \text{Period2}_{ij} - 0.2267 \times \text{Period3}_{ij}$$
$$- 0.1592 \times \text{Period4}_{ij} + 0.2151 \times \text{DrugAsp}_{ij}) / \sqrt{1.4290}].$$

The expected unit-specific and population average probabilities are summarized in Table 5.3. We see that there is very little difference in the estimated number of headaches. This result is to be expected as the design effect is $\sqrt{1.4290} = 1.1954$.

Estimated unit-specific and population average results for groups

DRUGASP	Period	Estimated headaches (unit-specific)	Estimated headaches (population-average)
0	1	1.4020	1.1728
0	2	1.3387	1.1199
0	3	1.0310	0.8624
0	4	1.1030	0.9227
1	1	1.7385	1.4543
1	2	1.6600	1.3886
1	3	1.2784	1.0694
1	4	1.3677	1.1441

2.4.10 Poisson log model with an offset variable

2.4.10.1 The model

The previous analysis assumed that the counts were all observed for the same number of days. However, this was not the case since the number of treatment days in the period did vary to some degree. Most of the counts were based on the full seven days in the week; however, some observations were made only for 1 day in the given week. To take this into account, we need to specify a so-called OFFSET variable. The offset variable indicates the amount of time that each count is based on. If OFFSET = no is specified, as was the case in the previous example, it is assumed that all counts are based on the same amount of time.

The offset variable is introduced into the Poisson model in the following way:

$$\log\left(\hat{\lambda}_{ij}\right) = \log(\text{offset variable}) + \left[\mathbf{x}_{ij}'\mathbf{b}_i\right]$$

where $\mathbf{x}_{ij}$ represent the values of the covariates corresponding to level-1 unit j nested within level-2 unit i and $\mathbf{b}_i$ denotes the coefficient vector containing both fixed and random effects.

In the current situation, the variable NTDays is the appropriate choice as the OFFSET variable. The model to be fitted to the data now changes to:

$$\log\left(\text{Headache}_{ij}\right) = \log\left(\text{NTDays}\right) + (\beta_0 + \beta_1 \times \text{Period1}_{ij} + \beta_2 \times \text{Period2}_{ij}$$
$$+ \beta_3 \times \text{Period3}_{ij} + \beta_4 \times \text{Period4}_{ij} + \beta_5 \times \text{DrugAsp}_{ij} + v_{i0}).$$

2.4.10.2　　　　Setting up the analysis

To create the model specifications for this model, ensure that **aspart.lsf** is displayed as the active window. Select the variable NTDays from the **Variables in data:** list as the **Offset Variable** by proceeding to the **Dependent and Independent Variables** dialog.
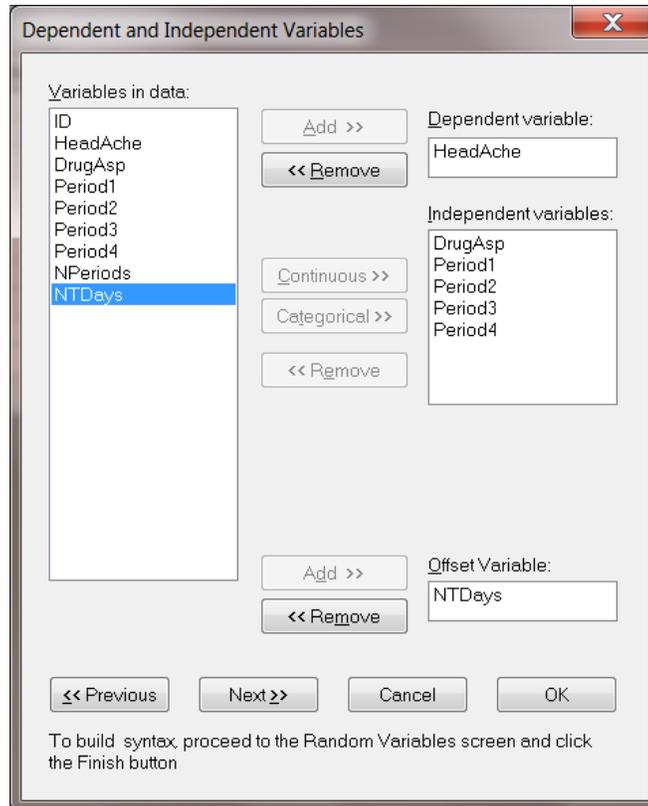
Click the **Next** button to proceed to the **Random Variables** dialog and then click **Finish** to create the syntax file (saved **as Aspart2.prl**) shown below.

.

```
MGlimOptions Converge=0.0001 MaxIter=100 MissingCode=-999999
IterDetails=No Method=Quad NQUADPTS=10 Output=Residuals;
Title=Poisson - Log random intercept model, ASPART data;
SY=aspart.lsf;
ID2=ID;
Distribution=POI;
Link=LOG;
Intercept=Yes;
Scale=None;
DepVar=HeadAche;
CoVars=DrugAsp Period1 Period2 Period3 Period4 DrugAsp;
Offset=NTDays;
RANDOM2=intcept;
```

2.4.10.3 Discussion of results

Fixed and random effect results

```
ASPART2.OUT                                                            ─ ◻ ✖

   Number of quadrature points =              20
   Number of free parameters =                7
   Number of iterations used =                3

   -2lnL (deviance statistic) =          404.86839
   Akaike Information Criterion          418.86839
   Schwarz Criterion                     438.49654

                       Estimated regression weights

                               Standard
    Parameter      Estimate       Error     z Value    P Value
    ---------      --------    --------     -------    -------
    intcept        -1.7127      0.2105      -8.1371     0.0000
    DrugAsp         0.2797      0.1641       1.7044     0.0883
    Period1         0.1001      0.2357       0.4247     0.6711
    Period2         0.0879      0.2250       0.3909     0.6959
    Period3        -0.2116      0.2567      -0.8242     0.4098
    Period4        -0.0787      0.2545      -0.3092     0.7571

      Event Rate Ratio and 95% Event Rate Confidence Intervals

                                              Bounds
    Parameter      Estimate    Event Rate    Lower      Upper
    ---------      --------    ----------    -----      -----
    intcept        -1.7127       0.1804      0.1194     0.2725
    DrugAsp         0.2797       1.3227      0.9589     1.8245
    Period1         0.1001       1.1053      0.6963     1.7544
    Period2         0.0879       1.0919      0.7026     1.6971
    Period3        -0.2116       0.8093      0.4894     1.3385
    Period4        -0.0787       0.9243      0.5613     1.5220

             Estimated level 2 variances and covariances

                                         Standard
    Parameter            Estimate          Error     z Value    P Value
    ---------            --------        --------     -------    -------
    intcept/intcept        0.4775          0.1924      2.4811     0.0131
```

Results for this model differ from those obtained for the model without offset variable discussed in the previous section. While the overall trend in predictor coefficient estimates is similar, the intercept is now estimated as –1.7127, compared to 0.2572 previously. The variance in intercept over patients for this model is estimated as 0.4775 compared to 0.4290 previously.

2.4.10.4 Interpreting the results

Estimated outcomes for groups: unit-specific results

The expected number of headaches can be obtained in the following fashion. First, we substitute the estimated coefficients in the model formulation

$$\log\left(\widehat{\text{Headache}}_{ij}\right) = \log\left(\text{NTDays}_{ij}\right) + (\hat{\beta}_1 \times \text{Period1}_{ij} + \hat{\beta}_2 \times \text{Period2}_{ij}$$
$$+ \hat{\beta}_3 \times \text{Period3}_{ij} + \hat{\beta}_4 \times \text{Period4}_{ij} + \hat{\beta}_5 \times \text{DrugAsp}_{ij})$$
$$= \log\left(\text{NTDays}_{ij}\right) + (-1.7127 + 0.1001 \times \text{Period1}_{ij} + 0.0879 \times \text{Period2}_{ij}$$
$$- 0.2116 \times \text{Period3}_{ij} - 0.0787 \times \text{Period4}_{ij} + 0.2797 \times \text{DrugAsp}_{ij}),$$

or, after taking exponents on both sides, as

$$\widehat{\text{Headache}}_{ij} = \text{NTDays}_{ij} \times \exp(-1.7127 + 0.1001 \times \text{Period1}_{ij} + 0.0879 \times \text{Period2}_{ij}$$
$$- 0.2116 \times \text{Period3}_{ij} - 0.0787 \times \text{Period4}_{ij} + 0.2797 \times \text{DrugAsp}_{ij}).$$

As most observations had a value of NTDays = 7, we start by considering typical patients with a full set of treatment days. We also assume that DrugAsp = 1, in other words, that aspartame rather than a placebo was administered.

During the first treatment period, we find that for such a patient

$$\widehat{\text{Headache}}_{ij} = 7 \exp(-1.7127 + 0.1001 + 0.2797)$$
$$= 7 \exp(-1.3329)$$
$$= 1.8460.$$

The expected numbers of headaches for a typical patient (again with NTDays = 7 and DrugAsp = 1) for the second, third, and fourth treatment periods are calculated as

$$\widehat{\text{Headache}}_{ij} = 7 \exp(-1.7127 + 0.0879 + 0.2797)$$
$$= 1.8236,$$

$$\widehat{\text{Headache}}_{ij} = 7 \exp(-1.7127 - 0.2116 + 0.2797)$$
$$= 1.3516,$$

and

$$\widehat{\text{Headache}}_{ij} = 7 \exp(-1.7127 - 0.0787 + 0.2797)$$
$$= 1.5437$$

respectively.

For a typical patient with only 5 treatment days, the expected numbers of headaches in each of the four treatment periods are 1.3186, 1.3026, 0.9654, and 1.1027 respectively.

When the expected numbers of headaches for a typical patient receiving aspartame under the Poisson model without offset variable (see previous section) and the Poisson model with offset variable are compared, we clearly see the impact of the inclusion of the offset variable on the estimated coefficients. These results are shown in Table 5.4.

Comparison of results for Poisson models

Period	Without offset variable	With offset variable (NTDays = 7)	With offset variable (NTDays = 5)
1	1.7385	1.846	1.3186
2	1.6600	1.8236	1.3026
3	1.2784	1.3516	0.9654
4	1.3677	1.5437	1.1027

Level 2 Bayes results

As requested during the model specification stage, the empirical Bayes estimates of the random effects are written to the file **aspart2.ba2**. The first few lines of this file are shown below.

The file **aspart.ba2** contains five pieces of information per individual:

- o the individual's ID,
- o the number of the random effect (only intercept in this case),
- o the empirical Bayes estimate for that individual (which is the mean of the posterior distribution),
- o the associated posterior variance, and
- o the name of the relevant random coefficient.

Since they are estimates of b_{i0} for each individual, the empirical Bayes estimates are expressed on the standard normal scale. Inspection of these estimates indicates that subject 13 has a very high score. This person's estimate of 1.469 (with variance .031) suggests a very high level of headaches. This agrees well with the raw data, which reveals that this person recorded 7 headaches on four occasions and 6 on the only other occasion.

2.4.11 Models for ordinal outcomes using NIMH data

2.4.11.1 Description of the NIMH data

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

Data for the first 10 observations are shown below in the form of a LISREL System File (LSF) spreadsheet file, named **nimh_study.lsf**.

NIHM_Study.LSF

	ID	IMPS79	IMPS79D	IMPS79O	DRUG	WEEK	SQRTWEEK
1	1103.00	5.50	1.00	4.00	1.00	0.00	0.00
2	1103.00	3.00	0.00	2.00	1.00	1.00	1.00
3	1103.00	-9.00	-9.00	-9.00	1.00	2.00	1.41
4	1103.00	2.50	0.00	2.00	1.00	3.00	1.73
5	1103.00	-9.00	-9.00	-9.00	1.00	4.00	2.00
6	1103.00	-9.00	-9.00	-9.00	1.00	5.00	2.24
7	1103.00	4.00	1.00	2.00	1.00	6.00	2.45
8	1104.00	6.00	1.00	4.00	1.00	0.00	0.00
9	1104.00	3.00	0.00	2.00	1.00	1.00	1.00
10	1104.00	-9.00	-9.00	-9.00	1.00	2.00	1.41

The variables of interest are:

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

- o WEEK represents the time during the course of the study when a specific measurement was made, and ranges between 0 and 6.
- o SQRTWEEK is the square root of WEEK. This variable is generated within the LISREL spreadsheet.
- o WSQRTxDR is the product of the treatment group and the square root of WEEK.

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

2.4.11.2 Defining variables

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

Select variable IMPS79O as shown below to activate all the options on the **Define Variables** dialog box.
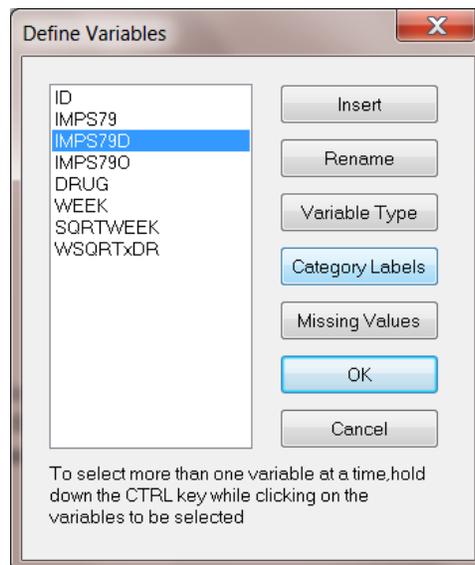
Click on the **Category Labels** option to activate the **Category Labels for IMPS79O** dialog box. Enter 1 in the **Value** box and corresponding **Label** of Norm, click on the **Add** button to add the first label for IMPS79O, with 1 = Norm. Similarly, add the next three labels: 2 = Mod, 3 = Mark, and 4 = Sev as shown below.
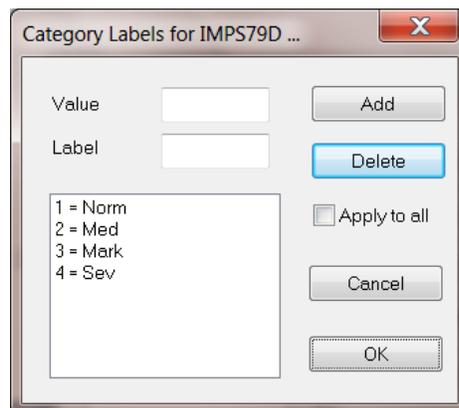
Click on the **OK** button to return to the **Define Variables** dialog box. Click on **OK** button to return to the LSF window. Save the change to the data set by clicking on the **File**, **Save** option.

2.4.11.3   Univariate graphs

As a first step, we take a look at the ordinal variable IMPS79O which is the potential dependent variable in this study.

Pie chart

To generate a pie chart for IMPS79O, first open the **nimh_study.lsf** spreadsheet. Next, select the **Graphs**, **Univariate** option to load the **Univariate plot** dialog box. Select the variable IMPS79O and indicate that a **Pie Chart** is to be graphed as shown below.
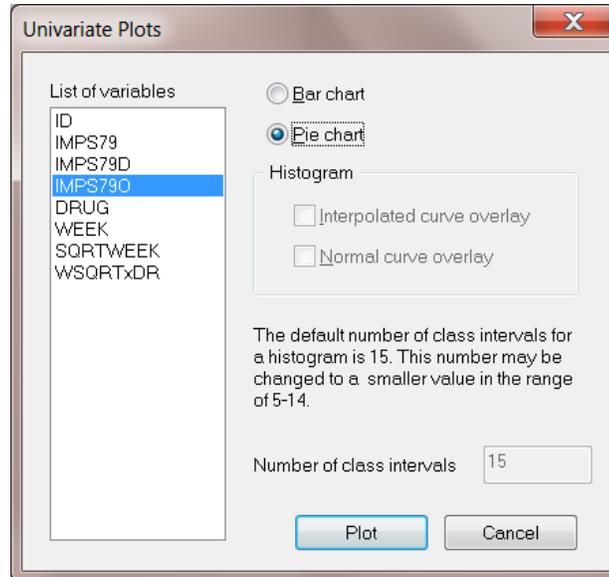
Click the **Plot** button to display the following pie chart.
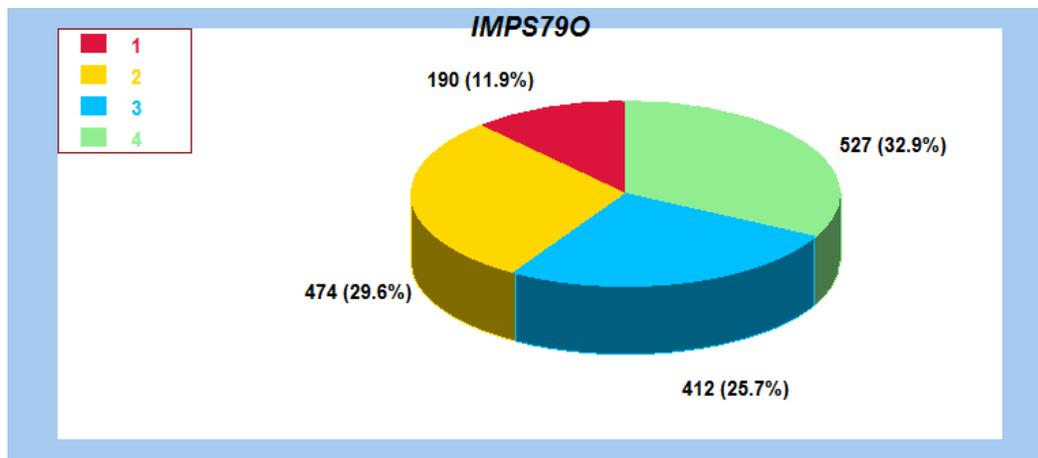
Note that most of the observations fall into the severe illness category. Keep in mind that the pie chart takes all observations, regardless of the time of measurement, into account. As such, it is informative about the distribution of all observed values of the potential outcome, but does not provide any information on possible trends in illness level over time.

2.4.11.4   Bivariate graphs

It is hoped that the severity of the illness (IMPS79O) will decrease over the treatment period. Before considering fitting a model to these data, we would like to explore the relationship between IMPS79O and WEEK using a bivariate bar chart.

Bivariate bar chart
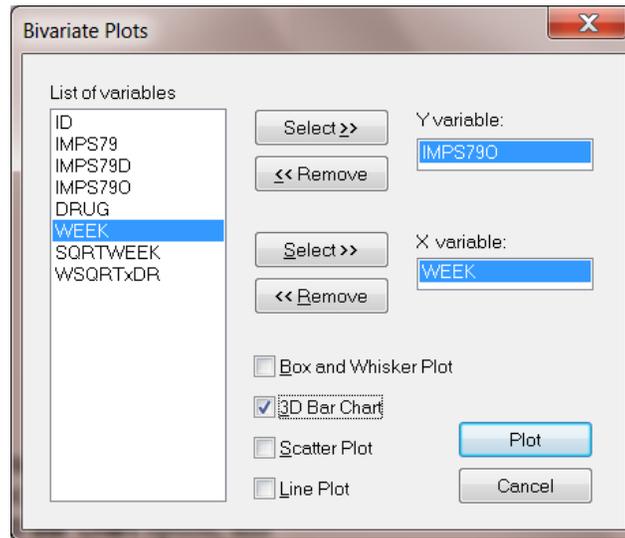
A bivariate bar chart is accessed via the **Graphs**, **Bivariate** option on the **File** menu. The **Bivariate** dialog box is completed as below: select the outcome variable IMPS79O as the **Y**-variable of interest, and the predictor WEEK to be plotted on the **X**-axis. Check the **3D Bar Chart** option, and click **Plot**.
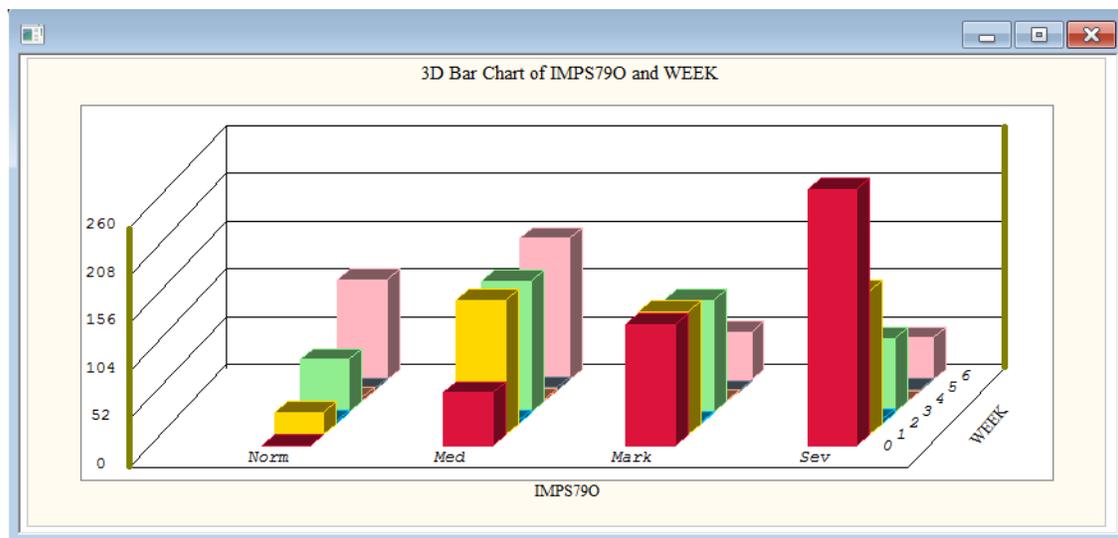
As shown above, most patients did not participate in the study at weeks 2, 4 and 5. At the beginning of the study (week 0), a large percentage of patients are markedly or severely ill. By the end of the study (week 6), most patients are reported as normal or moderate.

2.4.12 An ordinal regression model with random intercept

2.4.12.1 Introduction

An ordinal variable is a categorical variable where there is a logical ordering to the categories. In most cases, treating an ordinal outcome as a continuous variable is inadvisable. As in the case of a binary outcome variable, a link function is used in order to take the ceiling and floor effects of the ordinal outcome into account. The available link functions in LISREL include probit, logistic, complementary log-log and log-log. Detailed information on these link functions are given in Section 2.2.6.

2.4.12.2 The model

Let the outcome variable be coded into c categories, where $c = 1, 2, ..., C$. In this example, the ordinal variable IMPS79O defines the severity of the illness in terms of four categories, and thus $C = 4$. As ordinal models utilize cumulative comparisons of the categories, define the cumulative probabilities for the C categories of the outcome Y as $P_{ijc} = \Pr(Y_{ij} \leq c) = \sum_{k=1}^{c} p_{ijk}$, where p_{ijk} represents the probability that the response of the j-th measurement on patient i occurs in category k.

The type of drug, time elapsed since start of treatment, and the interaction between drug taken and time elapsed are of interest as predictors. The logistic regression model with IMPS79O as outcome can then be written as

Level-1 model:

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

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

Level-2 model:

$$b_{0i} = \beta_0 + v_{0i}, \quad i = 1, \ldots, N$$
$$b_{1i} = \beta_1$$
$$b_{2i} = \beta_2$$
$$b_{3i} = \beta_3$$

The cumulative probability can be expressed by

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

To obtain the probability for category c,

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

As shown above, the intercept b_{0i} is estimated by a level-2 equation. It indicates that patient i's initial IMPS79O value is not only determined by the population average β_0, but also by the patient difference v_{0i}. In other words, patients may have different average intercepts, and the model makes provision for this eventuality. The slopes are assumed to be the same for all the patients, which imply that each patient's trend line is parallel to the population trend.

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

$$y = c \leftrightarrow \gamma_{j-1} \le y^* \le \gamma_j, \ c = 1, 2, ..., C$$

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

2.4.12.3   Setting up the analysis

Open the LISREL spreadsheet **nimh_study.lsf** and select **Title and Options** option on the **Multilevel**, **Generalized Linear Model** menu.
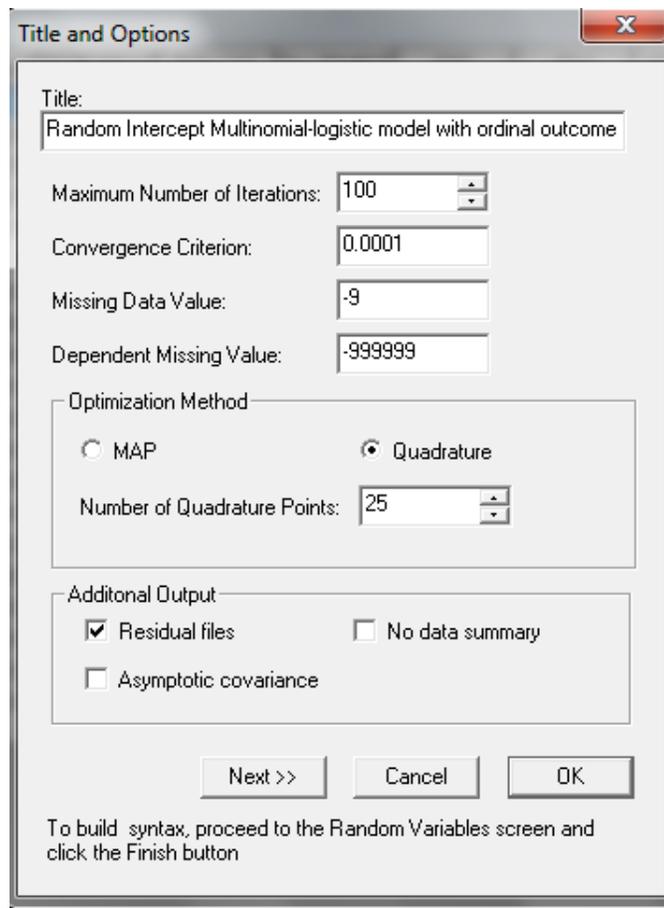
In the **Title and Options** dialog box, enter a title for the analysis in the **Title** text boxes. Keep the default settings for the **Maximum Number of Iterations** and **Convergence Criterion**. The **Missing Data Value** text box is used to specify the values of missing data for both outcome and predictors. We notice that the missing value –9 is presented in the data. Define the missing value by entering the number -9 in the **Missing Data Value** text box as shown above. Activate **Quadrature** radio button in the **Optimization Method** section and change the **Number of Quadrature Points** to 25 to obtain the above screen. Proceed to the **ID and Weight Variables** dialog box by clicking on the **Next** button.

Select ID from the **Variables in data** list box. Click on the **Add** button of the **Level-2 ID variable** section to obtain the following dialog box.
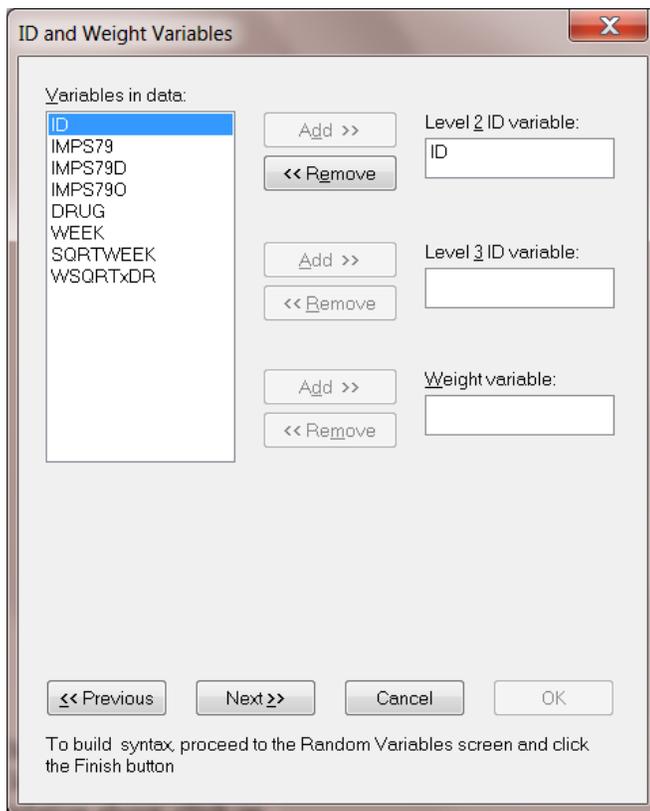
Proceed to the **Distribution and Links** dialog box by clicking on the **Next** button. Use the default **Distribution type**, which is **Multinomial**. The default link function is the logit link function. To change it to the ordinal logit link function corresponding to the model formulation above, click on the **Link function** drop-down list and select the **Ordinal logit** link function. Select **Subtract** from the **Model terms** drop-down list box as shown below.
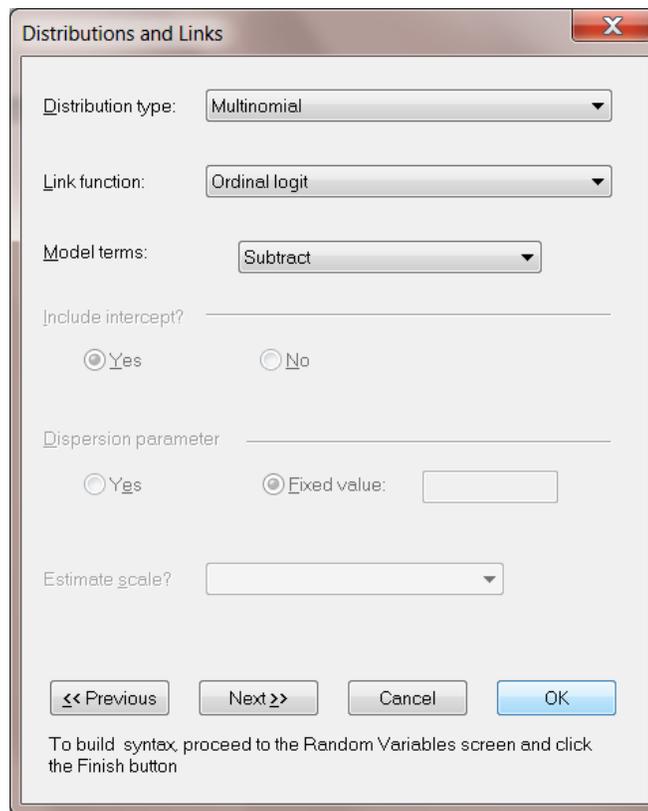
Click on the **Next** button to proceed to the **Dependent and Independent Variables** dialog box.

The **Dependent and Independent Variables** dialog box is used to specify the dependent and independent variables. First, select the dependent variable IMPS79O from the **Variables in data** list box and then click on the **Add** button to define it as the **Dependent variable**. Next, select DRUG, SQRTWEEK and WSQRTxDR one at a time and click on the **Continuous** button to add them as **Independent variables** as shown below.
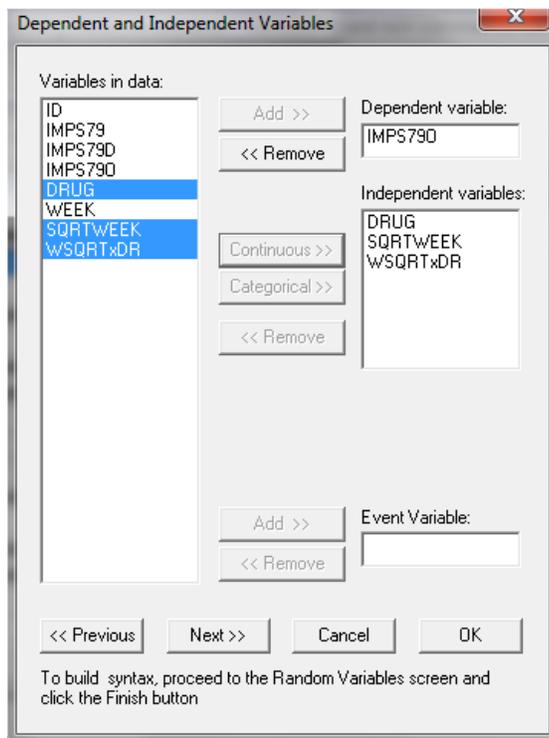
Click on the **Next** button to activate the Random Variables dialog box. By default, the **Intercept** check box in the **Random Level-2** is checked, indicating the inclusion of a random intercept at this level in the model. Keep the default settings as shown below and click on the **Finish** button to generate the PRELIS syntax (**prl**) file.
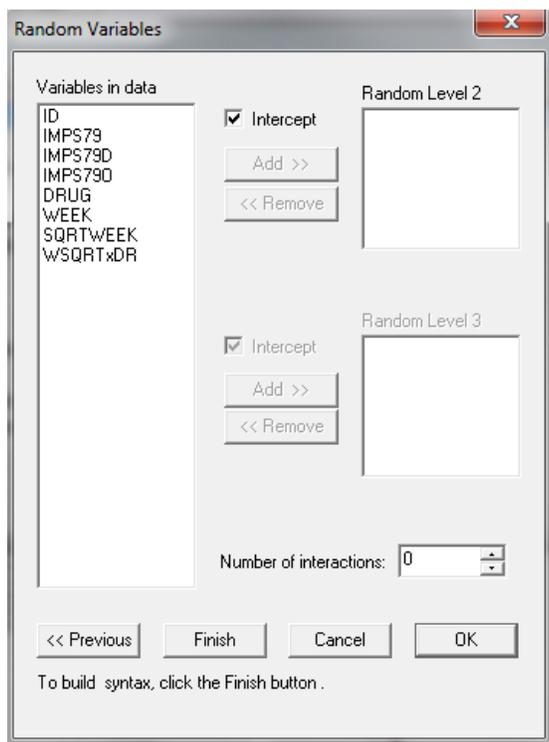
Before running the analysis, the PRELIS syntax file could be saved under a different file name. Select the **File**, **Save As** option, and provide a name (**nimh_study1.prl**) for the syntax file. Run the analysis by selecting the **Run PRELIS** icon as shown below.

```
MGlimOptions Converge=0.0001 MaxIter=100 MissingCode=-9
           Method=Quad NQUADPTS=25 Output=Residuals ModelTerms=sub;
Title=Random Intercept Multinomial-logistic model with ordinal outcome;
SY=NIHM_Study.lsf;
ID2=ID;
Distribution=MUL;
Link=OLOGIT;
DepVar=IMPS79O;
CoVars=DRUG SQRTWEEK WSQRTxDR;
RANDOM2=intcept;
```

2.4.12.4   Discussion of results

Syntax

The syntax lines are repeated in the output file corresponding to the PRELIS syntax (***.prl**) file we saved. Refer to Section 2.2.3 for detailed information about the syntax.

Model and data description

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

```
NIHM_Study1.OUT

O==================================================================O
| Random intercept multinomial logistic model with ordinal outcome |
O==================================================================O

            Model and Data Descriptions

   Sampling Distribution                 = Multinomial
   Link Function                         = Cumulative Logit
   Number of Level-2 Units                 437
   Number of Level-1 Units                 1603
   Number of Level-1 Units per Level-2 Unit =
      4    4    3    4    4    4    4    4    4    3    4
      4    2    3    4    3    4    3    4    4    4    3
      2    4    4    4    4    4    3    4    4    4    4
      4    4    4    4    2    3    4    3    4    4    4
      4    4    2    2    4    5    4    2    4    4    3
```

Descriptive statistics and starting values

Next, the descriptive statistics for all the variables are given. Notice that the variable name WSQRTxDR is truncated to WSQRTxDR. This is because LISREL only recognizes the first 8 characters of a variable name.
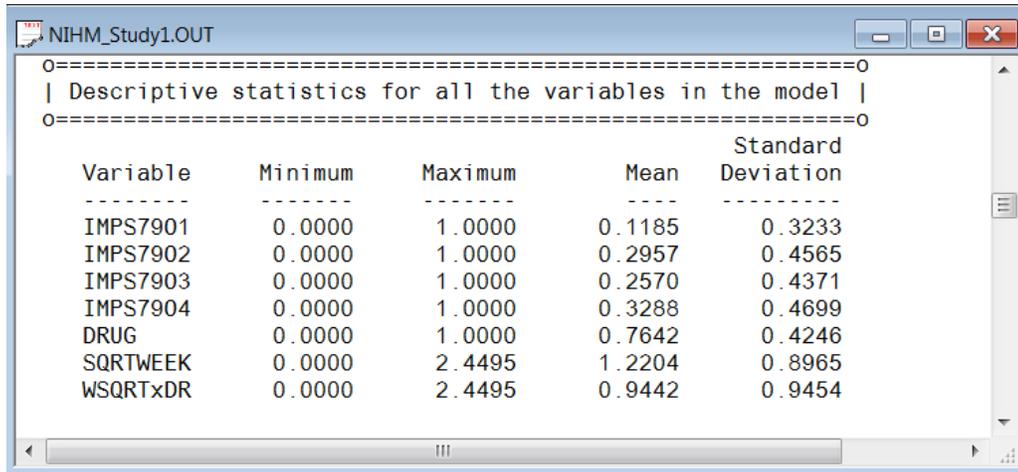
```
NIHM_Study1.OUT                                                    — □ ✕
O===========================================================O
| Descriptive statistics for all the variables in the model |
O===========================================================O
                                                    Standard
      Variable      Minimum      Maximum      Mean   Deviation
      --------      -------      -------      ----   ---------
      IMPS7901      0.0000       1.0000      0.1185    0.3233
      IMPS7902      0.0000       1.0000      0.2957    0.4565
      IMPS7903      0.0000       1.0000      0.2570    0.4371
      IMPS7904      0.0000       1.0000      0.3288    0.4699
      DRUG          0.0000       1.0000      0.7642    0.4246
      SQRTWEEK      0.0000       2.4495      1.2204    0.8965
      WSQRTxDR      0.0000       2.4495      0.9442    0.9454
```

Descriptive statistics are followed by the parameter estimates of a model with no random effects.
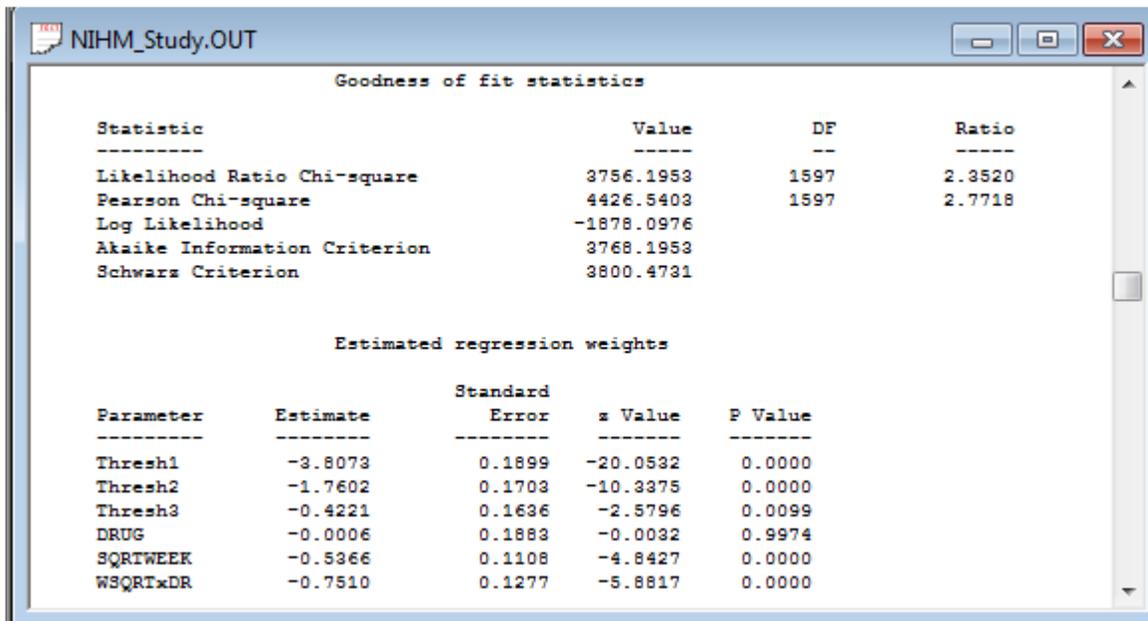
```
NIHM_Study.OUT                                                    — □ ✕
                   Goodness of fit statistics

Statistic                                Value        DF      Ratio
---------                                -----        --      -----
Likelihood Ratio Chi-square           3756.1953      1597     2.3520
Pearson Chi-square                    4426.5403      1597     2.7718
Log Likelihood                       -1878.0976
Akaike Information Criterion          3768.1953
Schwarz Criterion                     3800.4731


                   Estimated regression weights

                          Standard
Parameter     Estimate      Error    z Value    P Value
---------     --------    --------    -------    -------
Thresh1       -3.8073      0.1899    -20.0532    0.0000
Thresh2       -1.7602      0.1703    -10.3375    0.0000
Thresh3       -0.4221      0.1636     -2.5796    0.0099
DRUG          -0.0006      0.1883     -0.0032    0.9974
SQRTWEEK      -0.5366      0.1108     -4.8427    0.0000
WSQRTxDR      -0.7510      0.1277     -5.8817    0.0000
```

The final results after 4 iterations are shown next. The estimates are shown in the column with heading Estimate and correspond to the coefficients $\beta_0, \beta_1, \ldots, \beta_3$ in the model specification. The standard error, z-value and p-value are also printed.

```
NIHM_Study1.OUT                                              [ - ] [ □ ] [ X ]

    Number of quadrature points =                    25
    Number of free parameters =                       7
    Number of iterations used =                       4

    -2lnL (deviance statistic) =            3402.75922
    Akaike Information Criterion            3416.75922
    Schwarz Criterion                      3454.41665

                    Estimated regression weights

                                    Standard
    Parameter         Estimate        Error     z Value     P Value
    ---------         --------     --------     -------     -------
    Thresh1            -5.8593       0.3318    -17.6565      0.0000
    Thresh2            -2.8264       0.2900     -9.7458      0.0000
    Thresh3            -0.7085       0.2750     -2.5766      0.0100
    DRUG               -0.0585       0.3137     -0.1863      0.8522
    SQRTWEEK           -0.7658       0.1308     -5.8561      0.0000
    WSQRTxDR           -1.2061       0.1527     -7.9005      0.0000

        Odds Ratio and 95% Odds Ratio Confidence Intervals

                                                    Bounds
    Parameter         Estimate    Odds Ratio    Lower       Upper
    ---------         --------    ----------    -----       -----
    Thresh1            -5.8593       0.0029     0.0015      0.0055
    Thresh2            -2.8264       0.0592     0.0335      0.1046
    Thresh3            -0.7085       0.4924     0.2873      0.8440
    DRUG               -0.0585       0.9432     0.5100      1.7445
    SQRTWEEK           -0.7658       0.4650     0.3598      0.6008
    WSQRTxDR           -1.2061       0.2994     0.2219      0.4038

                              III
```

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

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

Intraclass correlation (ICC)

An estimate of the level-2 variance of the intercept and of the intracluster correlation (ICC) is given in the next section of the output. The residual variance for the logistic link function is assumed to be $\pi^2/3$.

```
NIHM_Study1.OUT                                                          _ □ ✕
                    Estimated level 2 variances and covariances

                                        Standard
        Parameter               Estimate       Error     z Value   P Value
        ---------               --------     --------     -------   -------
        intcept/intcept          3.7739       0.4652      8.1132    0.0000


    Calculation of the intracluster correlation
    -------------------------------------------
    residual variance = pi*pi / 3 (assumed)
    cluster  variance =    3.7739

    intracluster correlation =    3.7739 / (    3.7739 + (pi*pi/3)) = 0.534
```

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

Interpreting the output
Estimated outcomes for groups: unit-specific probabilities

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

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

or, in the notation introduced in Section 2.4.12.2,

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

When $c = 1$, we find that, for a patient from the control group (DRUG = 0, SQRTWEEK = WSQRTxDR = 0),

$$\log\left(\frac{\hat{P}_{ij1}}{1-\hat{P}_{ij1}}\right) = \hat{\eta}_{ij1} = -5.8593$$

$$\hat{P}_{ij1} = \frac{e^{\hat{\eta}_{ij1}}}{1+e^{\hat{\eta}_{ij1}}} = 0.0028$$

100

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

The cumulative probabilities we calculated are

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

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

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

$$\hat{p}_{ij1} = 0.0028 - 0 = 0.0028$$
$$\hat{p}_{ij2} = 0.0559 - 0.0028 = 0.0531$$
$$\hat{p}_{ij3} = 0.3299 - 0.059 = 0.2740$$
$$\hat{p}_{ij4} = 1 - 0.3299 = 0.6701.$$

For this group of patients (DRUG = 0) at the starting week, the expected percentages of patients in each of the categories are as follows: 0.3% of the patients are normal or borderline mentally ill; 5.3% of the patients are mildly or moderately ill; 27.4% are markedly ill and 67% are severely or extremely ill. Similarly, we can calculate the estimated percentages for both groups at all the time points as shown in Table 8.

The contents of Table 8 can be graphically represented as shown in Figures 3 and 4. It clearly shows that the numbers of markedly and severely ill patients decrease dramatically over time. The improvement for the drug patients is larger than the placebo patients.

Table 6: Estimated % for both groups at 7 time points

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

A 2-level random intercept model and trend model

In this section, we fit a model with random intercept and slope. To do this, the level-1 model is unchanged; only the level-2 model is modified.

The model

Level-1 model:

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

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

Level-2 model:

$$b_{0i} = \beta_0 + v_{0i}, \quad i = 1, \ldots, N$$
$$b_{1i} = \beta_1$$
$$b_{2i} = \beta_2 + v_{2i}$$
$$b_{3i} = \beta_3$$

As shown above, the slope of the time variable b_{2i} is now estimated by a level-2 equation containing both a fixed and a random effect. It indicates that patients are now not only assumed to have different intercepts, but may also exhibit different responses to the treatment over time.

Setting up the analysis

In this example, we want to use 10 quadrature points and include SQRTWEEK as level-2 random effect. We modify the commands syntax previously saved to **nimh_study1.prl** to obtain the new model setup.

First, click on **File**, **Open** to browse and open **nimh_study1.prl**. Next, we change the string in the **NQUADPTS = 10** in the MGLIM command. Change **RANDOM2 = intcept SQRTWEEK** and save the syntax file to **nimh_study2.prl**.

```
MGlimOptions Converge=0.0001 MaxIter=100 MissingCode=-9
          Method=Quad NQUADPTS=10 Output=Residuals ModelTerms=sub;
Title=Random Intercept Multinomial-logistic model with ordinal outcome;
SY=NIHM_Study.lsf;
ID2=ID;
Distribution=MUL;
Link=OLOGIT;
DepVar=IMPS79O;
CoVars=DRUG SQRTWEEK WSQRTxDR;
RANDOM2=intcept SQRTWEEK;
```

Click on the **Run PRELIS** icon to produce the output file **nimh_study2.out**.

```
O===================================================O
| Results for the model without any random effects  |
O===================================================O

                  Goodness of fit statistics

   Statistic                              Value        DF       Ratio
   ---------                              -----        --       -----
   Likelihood Ratio Chi-square          3756.1953      1597     2.3520
   Pearson Chi-square                   4426.5410      1597     2.7718

                   Estimated regression weights

                                 Standard
   Parameter     Estimate         Error      z Value   P Value
   ---------     --------       --------     -------   -------
   Thresh1        -3.8073         0.1899     -20.0532   0.0000
   Thresh2        -1.7602         0.1703     -10.3375   0.0000
   Thresh3        -0.4221         0.1636      -2.5796   0.0099
   DRUG           -0.0006         0.1883      -0.0032   0.9974
   SQRTWEEK       -0.5366         0.1108      -4.8427   0.0000
   WSQRTxDR       -0.7510         0.1277      -5.8817   0.0000

       Odds Ratio and 95% Odds Ratio Confidence Intervals

                                                 Bounds
   Parameter     Estimate      Odds Ratio     Lower     Upper
   ---------     --------      ----------     -----     -----
   Thresh1        -3.8073        0.0222       0.0153    0.0322
   Thresh2        -1.7602        0.1720       0.1232    0.2402
   Thresh3        -0.4221        0.6557       0.4758    0.9036
   DRUG           -0.0006        0.9994       0.6909    1.4456
   SQRTWEEK       -0.5366        0.5847       0.4706    0.7266
   WSQRTxDR       -0.7510        0.4719       0.3674    0.6061
```

Discussion of results
Fixed effect results, adaptive quadrature

The final results after 7 iterations are listed below. While the values of the estimated coefficients differ from those in the random-intercept-only model, the overall picture remains very similar. The decline in severity over time noticed in the crosstabulation is captured by the significant fixed effect coefficient of –0.8840 for SQRTWEEK.

```
NIHM_Study2.OUT                                              [ _ ][ □ ][ ✕ ]
o==========================================o
| Optimization Method: Adaptive Quadrature |
o==========================================o

    Number of quadrature points =              10
    Number of free parameters =                 9
    Number of iterations used =                 7

    -2lnL (deviance statistic) =        3325.49276
    Akaike Information Criterion        3343.49276
    Schwarz Criterion                  3391.90945


                     Estimated regression weights

                                  Standard
    Parameter       Estimate         Error    z Value    P Value
    ---------       --------      --------    -------    -------
    Thresh1          -7.3234        0.4726   -15.4954     0.0000
    Thresh2          -3.4214        0.3862    -8.8597     0.0000
    Thresh3          -0.8151        0.3517    -2.3177     0.0205
    DRUG              0.0553        0.3912     0.1414     0.8875
    SQRTWEEK         -0.8840        0.2183    -4.0495     0.0001
    WSQRTxDR         -1.6939        0.2525    -6.7091     0.0000
                              III
```

Random effects results

Note that the estimated coefficient for the random SQRTWEEK slope is highly significant, indicating that patients not only start at different points but follow different paths during the treatment period.

```
NIHM_Study2.OUT                                              [ _ ][ □ ][ ✕ ]
            Estimated level 2 variances and covariances

                                  Standard
    Parameter            Estimate     Error    z Value    P Value
    ---------            --------  --------    -------    -------
    intcept/intcept        7.0058    1.3190     5.3114     0.0000
    SQRTWEEK/intcept      -1.5122    0.5320    -2.8424     0.0045
    SQRTWEEK/SQRTWEEK      2.0118    0.4181     4.8112     0.0000
                              III
```

Interpreting the output
Estimated outcomes for groups: unit-specific results

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

$$\log\left(\frac{\hat{P}_{ijc}}{1-\hat{P}_{ijc}}\right) = \hat{\gamma}_c - \left[\hat{b}_{0i} + \hat{b}_{1i}\text{DRUG}_i + \hat{b}_{2i}\text{SQRTWEEK}_i + \hat{b}_{3i}\left(\text{WSQRT} \times \text{DRUG}\right)_i\right]$$

so that

$$\hat{\eta}_{ijc} = \hat{\gamma}_c - 7.3793 + 0.0553 \times \text{DRUG}_i + 0.8841 \times \text{SQRTWEEK}_i$$
$$+ 1.6940 \times (\text{WSQRT} \times \text{DRUG})_i$$

As illustrated in the previous example, by substituting the values for DRUG, SQRTWEEK and WSQRTxDR, the results shown in Table 7 can be obtained.

Table 7: Estimated unit-specific results for random intercept & slope model

severity	Placebo patients (drug = 0)				Drug patients (drug = 1)			
	normal	moderate	marked	severe	normal	moderate	marked	severe
week 0	0.06%	2.96%	26.90%	70.08%	0.07%	3.13%	27.90%	68.91%
week 1	0.15%	6.87%	43.81%	49.17%	0.86%	29.42%	55.32%	14.40%
week 2	0.22%	9.61%	50.03%	40.15%	2.47%	51.90%	39.98%	5.81%
week 3	0.29%	12.32%	53.77%	33.62%	5.42%	68.72%	23.37%	2.49%
week 4	0.36%	15.09%	55.99%	28.55%	10.27%	74.85%	13.62%	1.26%
week 5	0.45%	17.94%	57.12%	24.49%	17.38%	73.94%	7.99%	0.69%
week 6	0.54%	20.84%	57.44%	21.17%	26.72%	68.08%	4.80%	0.40%

We can again represent the results from the above table graphically, as shown in Figures 5 and 6. The graphs tell us the same story as the previous model: patients from the treatment group showed more improvement over time than patients from the control group. While a very small proportion of treatment patients were still diagnosed as being severely ill at the end of the treatment period (0.42% according to table 9), 20% of the control group were still classified as being severely ill by week 6.

Estimated time trend variance

When we consider the heterogeneity in responses across time, we notice that the estimated variance in the time trend is $\sigma_{v_1}^2 = (1.29774)^2 + (-0.57054)^2 = 2.0096$. The estimates for the time trends are -0.88295 for SQRTWEEK and -1.69416 for WSQRTxDR respectively. Thus the estimated trends for the placebo and drug groups are -0.88295 and $-0.88295 -1.69416 = -2.57711$. Thus the 95% confidence interval of the time trend for the placebo group is $-0.88295 \pm (1.96 \times \sqrt{2.0096})$ $= (-3.6615, 1.896)$. Similarly, the confidence interval for the drug group is $(-5.3556, 0.2014)$. Notice that both intervals are fairly large and include negative and positive slopes, which reflects the wide heterogeneity in trends. The estimated correlation value is –0.402, which is moderately large. This indicates that the patients who are initially less severely ill improve at a smaller rate. The more severely ill patients improve at a greater rate.

2.4.13 Models for nominal outcomes using NHIS data

2.4.13.1 Introduction

In statistics, the kinds of significance tests and model fitting procedures that are appropriate depend on the level of measurement of the variables concerned. A widely accepted classification scheme, proposed by Stevens (1946), is listed below and consists of four levels of measurement:

o   nominal (also categorical or discrete)
o   ordinal
o   interval
o   ratio

Interval and ratio variables are usually grouped together as continuous variables.

In the case of nominal variables there are no "less than" or "greater than" relations among the categories of the variable and operations such as addition or multiplication do not exist.

Examples of nominal variables are

o   Cancer Type (1 = breast, 2 = lung, 3 = brain, 4 = leukemia, 5 = liver, 6 = colon, 7 = other),
o   Smoking Status (1 = never smoked, 2 = former smoker, 3 = current smoker),
o   Preference for U.S. President (1 = Democrat, 2 = Republican, 3 = Independent),
o   Type of Sweetener (1 = sugar, 2 = saccharin, 3 = aspartame, 4 = other),
o   Pain Reliever (1 = Acetaminophen, 2 = Aspirin, 3 = Ibuprofen, 4 = Ketoprofen,
     5 = Naproxen, 6 = other).

In many research situations, the underlying variable type is continuous. However, to ensure anonymity of respondents, information is obtained by categorizing variables. For example:

o   Annual Income (1 = not employed, 2 = less or equal to $20,000, 3 = more than $20,000 but less than or equal to $50,000, 4 = more than $50,000 but less than or equal to $100,000, 5 = more than $100,000)
o   Age when diagnosed (1 = not applicable, 2 = younger than 25 years, 3 = 25 years or older but less than 50 years, 4 = 50 years or older but less than 70 years, 5 = 70 years and older).

In both the cases above, the available data values are coded 1, 2, 3, 4 and 5. Arithmetic operation with these codes will not provide accurate estimates of the actual age and income characteristics and in both cases the first category makes "less than" and "more than" comparisons less feasible.

In this guide we illustrate the analysis of a nominal outcome variable by fitting a three-level model to health related data.

2.4.13.2   The data

The data set comes from the data library of the National Health Interview Survey (NHIS). The NHIS is a national longitudinal health survey. During 2002, background data and data on the health conditions of a sample of 28,737 participants were obtained. The 2002 sample was stratified into 64 strata and into 601 PSUs. Using this data, we created a subset consisting of 57 strata (the level-3

units), 399 PSUs (the level-2 units) and 6445 participants. A partial list of the data is given below in the form of a LISREL spreadsheet file, named **nihs_subset.lsf**.

	CSTRATM	CPSUM	PATWT	PASTVIS	NUMMED	GENDER	USETOBAC
1	20102101.00	100013.00	50245.00	3.00	2.00	1.00	1.00
2	20102101.00	100013.00	50245.00	3.00	2.00	1.00	0.00
3	20102101.00	100013.00	50245.00	3.00	4.00	1.00	0.00
4	20102101.00	100013.00	50245.00	3.00	2.00	1.00	0.00
5	20102101.00	100013.00	50245.00	2.00	1.00	0.00	0.00
6	20102101.00	100015.00	72581.00	4.00	0.00	1.00	0.00
7	20102101.00	100015.00	72581.00	3.00	2.00	1.00	0.00
8	20102101.00	100015.00	72581.00	3.00	2.00	0.00	0.00
9	20102101.00	100015.00	72581.00	1.00	1.00	0.00	0.00
10	20102101.00	100015.00	72581.00	3.00	0.00	1.00	0.00

A description of the variables is as follows:

- CSTRATM is the stratum used as level-3 ID (57 strata).
- CPSUM is the primary sampling unit (PSU) and is used as level-2 ID (399 clusters).

- PATWT is the participant design weight.
- PASTVIS is the value of the nominal variable for the number of visits to a medical doctor during the past 12 months (1 = none or unknown, 2 = 1 to 2, 3 = 3 to 5, 4 = 6 medications and more).
- NUMMED is the number of medications.
- GENDER, where 0 = Female and 1 = Male.
- USETOBAC indicates whether a participant smoked cigarettes or not, where 0 = no and 1 = yes.
- PRIMCARE, where 0 = none and 1 = participant has primary care.
- INJURY indicates whether a participant suffered from an injury or not (0 = no, 1 = yes).
- BLODPRES, where 0 = blood pressure not measured and 1 = blood pressure measured.
- URINE, where 0 = no urine tested, 1 = tested.
- XRAY, where 0 = no X rays taken and 1 = X ray taken.
- EXERCISE, where 0 = no exercise and 1 = participant does exercise.
- RACER indicates the ethnicity of a participant where 1 = White, 2 = Black and 3 = Other.
- AGER indicates in which age category a participant belongs. Coded as follows: 1 = Under 15, 2 = 15 to 24, 3 = 25 to 44, 4 = 45 to 64, 5 = 65 to 74, 6 = 75 and older.
- AGE1 to AGE5 are five dummy variables coded as follows:

Table 8: Dummy variables

Age	AGE1	AGE2	AGE3	AGE4	AGE5
Under 15	1	0	0	0	0
15 to 24	0	1	0	0	0
25 to 44	0	0	1	0	0
45 to 64	0	0	0	1	0
65 to 74	0	0	0	0	1
75 and older	0	0	0	0	0

2.4.13.3   The model

A general multilevel nominal model

In the nominal case we need to consider the values corresponding to the unordered multiple categories of the response variable. We thus assume that the C response categories are coded as $1, 2, 3, ..., C$.

Let $P_{ijkc} = P\left(y_{ijk} = c \mid \boldsymbol{\beta}_c, \boldsymbol{\upsilon}_{ic}, \boldsymbol{\upsilon}_{ijc}\right)$ denote the probability that a response occurs in category c, conditional on a $(p \times 1)$ vector of fixed regression parameters $\boldsymbol{\beta}_c$, the $(m \times 1)$ vector of level-2 random effects $\boldsymbol{\upsilon}_{ijc}$ and the $(r \times 1)$ vector of level-3 random effects $\boldsymbol{\upsilon}_{ic}$. It is further assumed that the level-2 random effects $\boldsymbol{\upsilon}_{ijc}$ are independent and identically distributed (i.i.d.) as a $N\left(\mathbf{0}, \boldsymbol{\Phi}_{(2)}\right)$ random variable. Uncorrelated with $\boldsymbol{\upsilon}_{ijc}$, the level-3 random effects are i.i.d. $N\left(\mathbf{0}, \boldsymbol{\Phi}_{(3)}\right)$. The scalar y_{ijk} denotes the value of the nominal variable associated with level-1 unit k, $k = 1, 2, \ldots, n_{ij}$, nested within level-2 unit j, $j = 1, 2, \ldots n_i$, which in turn is nested within the i-th level-3 unit, where $i = 1, 2, \ldots, N$. The probabilities P_{ijkc} are computed as

$$
\begin{aligned}
P_{ijkc} &= P\left(y_{ijk} = c \mid \boldsymbol{\beta}_c, \boldsymbol{\upsilon}_{ic}, \boldsymbol{\upsilon}_{ijc}\right) \\
&= \frac{\exp\left(\eta_{ijkc}\right)}{1 + \sum_{h=1}^{C-1} \exp\left(\eta_{ijkc}\right)}, \qquad c = 1, 2, \ldots, C-1
\end{aligned}
$$

where

$$
\eta_{ijkc} = \mathbf{x}'_{ijk} \boldsymbol{\beta}_c + \mathbf{z}'_{ijk(2)} \boldsymbol{\upsilon}_{ijc} + \mathbf{z}'_{ijk(3)} \boldsymbol{\upsilon}_{ic}
$$

Note that $\mathbf{x}'_{ijk}$, $\mathbf{z}'_{ijk(2)}$ and $\mathbf{z}'_{ijk(3)}$ are design vectors for the explanatory variables and the level-2 and level-3 random effects respectively.

Random intercept model with two explanatory variables

For the **nihs_subset.lsf** data set considered earlier, let PASTVIS denote the outcome variable. Assume further that GENDER and EXERCISE are the only predictors and that only intercepts are allowed to vary randomly across level-3 and level-2 units. The corresponding estimated probability model is given by

$$P\left(\text{PASTVIS}_k = c\right) = \frac{\exp\left(\eta_{ijkc}\right)}{1 + \sum_{h=1}^{3} \exp\left(\eta_{ijkh}\right)}, \qquad c = 1, 2, 3$$

where

$$\eta_{ijkh} = \beta_{0h} + \beta_{1h} \times \text{GENDER}_k + \beta_{1h} \times \text{EXERCISE}_k + \upsilon_{ijh} + \upsilon_{ih}$$

and where PASTVIS_k, GENDER_k and EXERCISE_k denote values of the variables for client k nested within unit (i, j). Note that for PASTVIS the number of categories is $C = 4$.

Remarks:

The probability $P\left(\text{PASTVIS}_k = 4\right)$ is obtained as $1 - \sum_{c=1}^{3} P\left(\text{PASTVIS}_k = c\right)$. In the formulation above, we used the last category as the so-called reference category.

MGLIM allows the user to select the first or the last category as the reference category. If the first category is selected as reference category, then

$$P\left(\text{PASTVIS}_k = c\right) = \frac{\exp\left(\eta_{ijkc}\right)}{1 + \sum_{h=2}^{4} \exp\left(\eta_{ijkh}\right)}, \qquad c = 2, 3, 4.$$

$$P\left(\text{PASTVIS}_k = 1\right) = 1 - \sum_{c=2}^{4} P\left(\text{PASTVIS}_k = c\right)$$

2.4.13.4 A random intercept model with fourteen predictors

Preparing the data

The model to be fitted to the data is contained in **nihs_subset1.lsf**. This file was created from the SPSS data file **nihs_subset1.sav** as follows.

Use the **File**, **Import Data File** option to activate the display of an **Open** dialog box. From the **Files of type** drop-down list, select **SPSS Data File (*.sav)**. Browse for the file **nih_subset.lsf**. Select the file and click the **Open** button to activate **Save As** dialog box. Enter the file name **nih_subset.lsf** and click on the Save button to display **nih_subset.lsf** as a LISREL spreadsheet.

Exploring the data

To obtain some insight into the distributional properties and possible relationships between variables, it is useful to present these properties graphically using the **Graphs** option. Prior to making visual presentations, it is a good idea to assign labels to the categories of the nominal and ordinal variables.

First, highlight the column of PASTVIS by clicking on its header. Then right click and select the **Define Variables** option as shown below to open the **Define Variables** dialog box.

Select variable PASTVIS as shown below to activate all the options on the **Define Variables** dialog box.
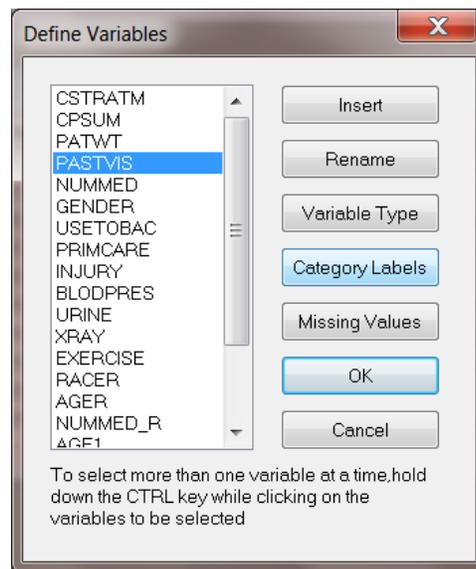
Click on the **Category Labels** option to activate the **Category Labels for PASTVIS** dialog box. Enter the labels None, 1 to 2, 3 to 5 and >5 as shown below and click **OK**.
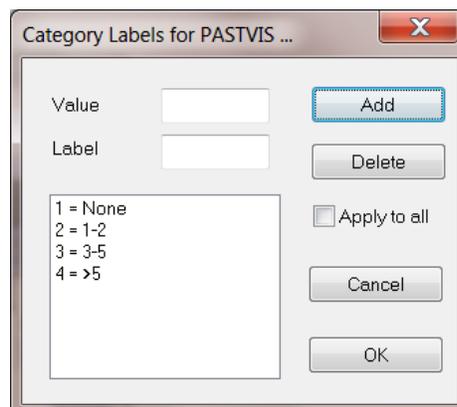
Click on the **OK** button to return to the **Define Variables** dialog box. Click on **OK** button to return to the LSF window. Save the change to the data set by clicking on the **File**, **Save** option.

From the main menu bar, select the **Graphs**, **Bivariate** option. By clicking on the **Bivariate** tab of the pop-up menu, the **Bivariate plot** dialog box is invoked. Select PASTVIS as the Y variable and PRIMCARE as the X variable.
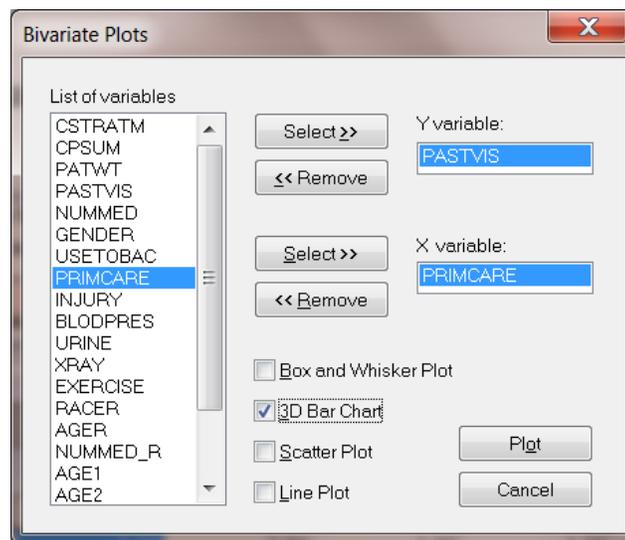
Next, check the **3D Bar Chart** check box and then the **Plot** button to obtain the bivariate bar chart of PRIMCARE versus PASTVIS. The graph below shows that there is an increase in the use of primary care with the number of visits to a medical doctor.
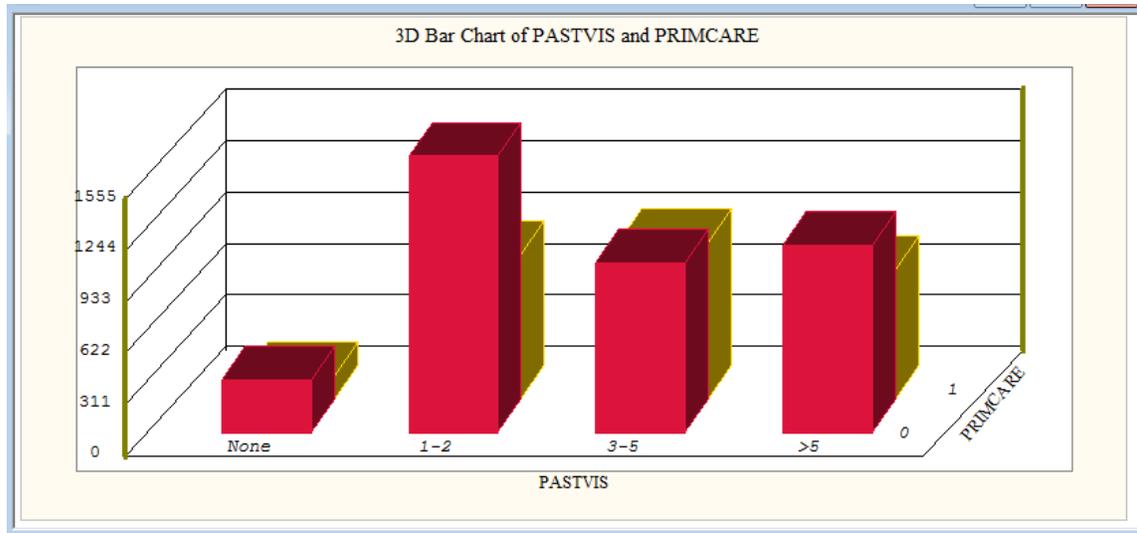
Setting up the analysis

From the main menu bar of the LSF window, select the **Multilevel**, **Generalized Linear Model**, **Title and Options** option. Enter a title for the analysis in the **Title** text boxes. Keep the default settings for the **Maximum Number of Iterations**, **Convergence Criterion** and the **Missing Data Value**. Activate **Quadrature** radio button in the **Optimization Method** section and change the **Number of Quadrature Points** to 8 to obtain the above screen. Proceed to the **ID and Weight Variables** dialog box by clicking on the **Next** button.

Select CPSUM from the **Variables in data** list box. Click on the upper **Add** button of the **Level-2 ID variable** section to define the level-2 ID. Similarly define CSTRATM and click on the middle **Add** button to define it as **Level-3 ID variable**.
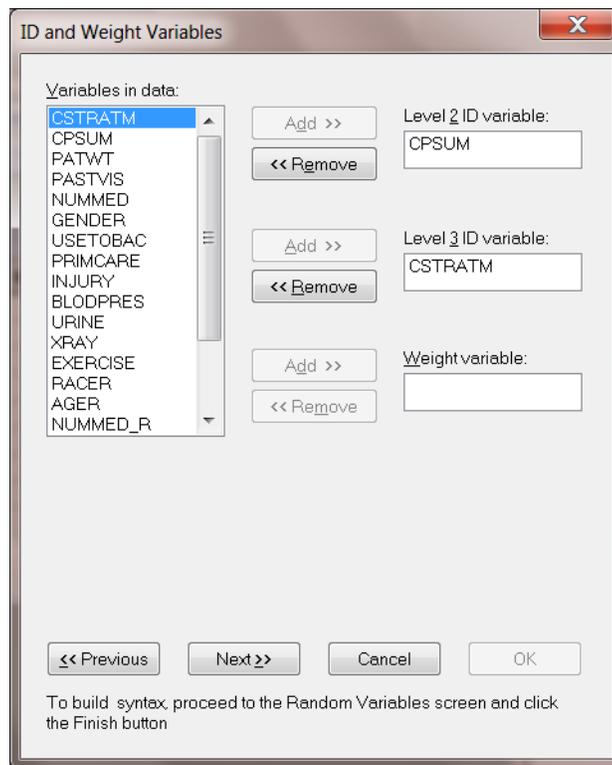
Proceed to the **Distribution and Links** dialog box by clicking on the **Next** button. Keep all the default settings on this dialog box as shown below.
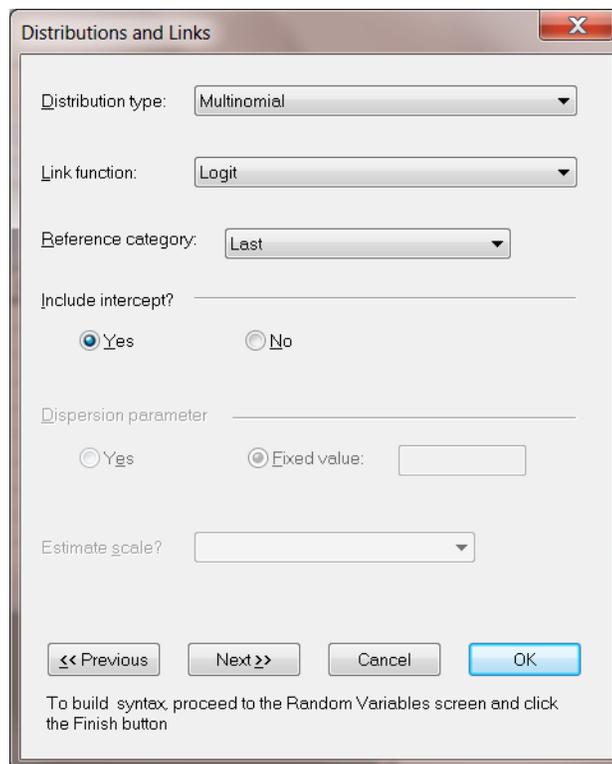
Click on the **Next** button to proceed to the **Dependent and Independent Variables** dialog box.

The **Dependent and Independent Variables** dialog box is used to specify the dependent and independent variables. First, select the dependent variable PASTVIS from the **Variables in data** list box and then click on the **Add** button to define it as the **Dependent variable**. Next, select NUMMED, GENDER, PRIMCARE, INJURY, BLODPRES, URINE, XRAY, EXERCISE, AGE1, AGE2, AGE3, AGE4 and AGE5 and click on the **Continuous** button to add them as **Independent variables** as shown below.
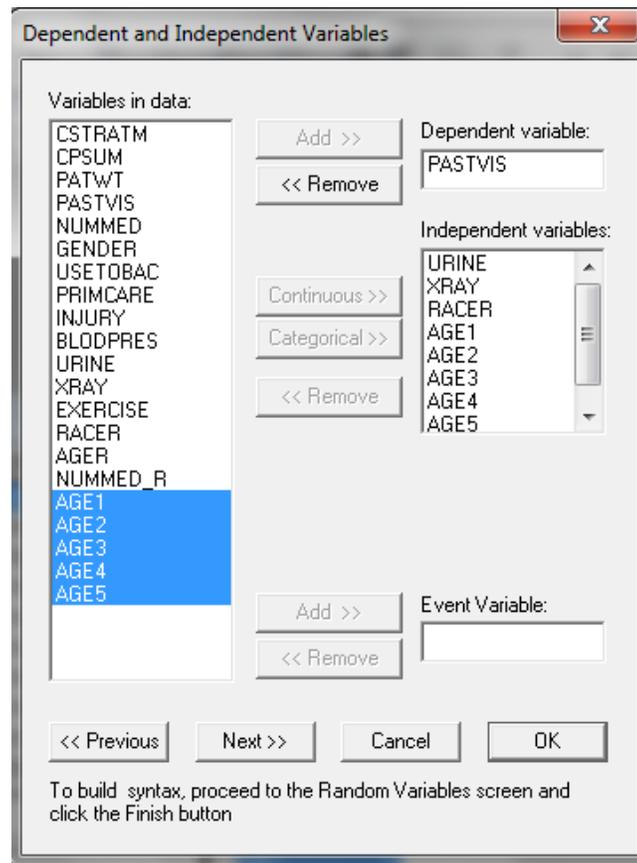
Click on the **Next** button to activate the **Random Variables** dialog box. By default, the **Intercept** check box in the **Random Level-2** and **Random Level-3** are checked, indicating the inclusion of a random intercept at level-2 and 3 in the model. Keep the default settings as shown below and click on the **Finish** button to generate the PRELIS syntax (**prl**) file.
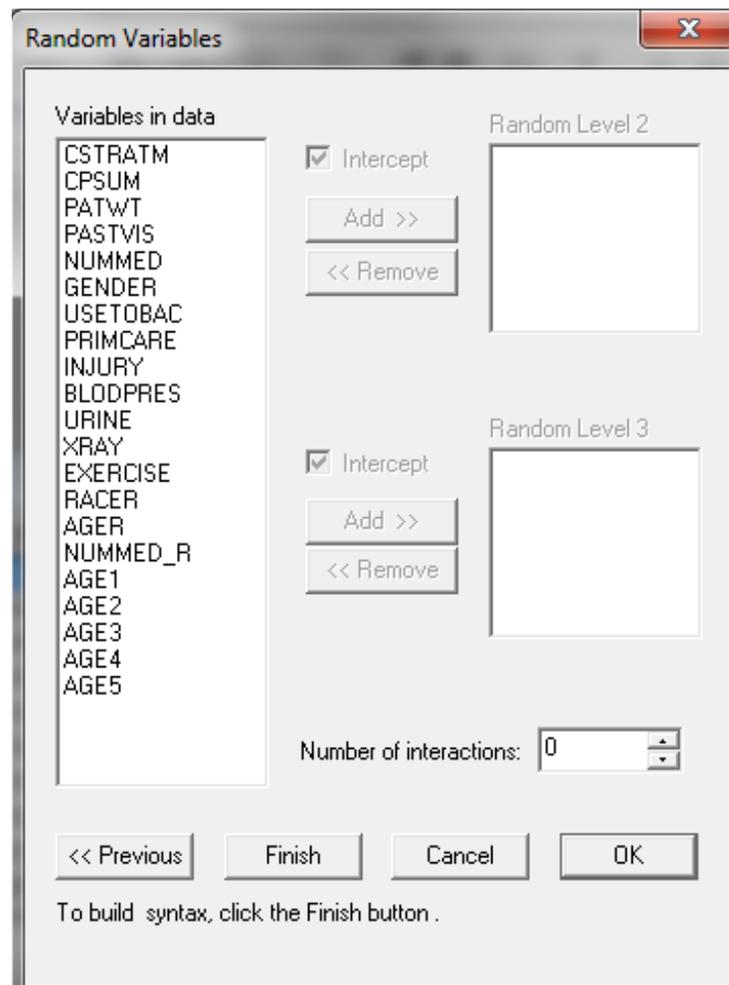
Before running the analysis, the PRELIS syntax file should be saved. Select the **File**, **Save As** option, and provide a name (**nhis1.prl**) for the syntax file. Run the analysis by selecting the **Run PRELIS** icon as shown below.

```
MGlimOptions Converge=0.0001 MaxIter=100 MissingCode=-999999
          Method=Quad NQUADPTS=8 RefCat=last;
Title= National Health Interview Data Nominal model ;
SY=nih_subset.LSF;
ID2=CPSUM;
ID3=CSTRATM;
Distribution=MUL;
Link=LOGIT;
Intercept=Yes;
DepVar=PASTVIS;
CoVars=NUMMED GENDER PRIMCARE INJURY BLODPRES URINE XRAY EXERCISE AGE1
 AGE2 AGE3 AGE4 AGE5;
RANDOM2=intcept;
RANDOM3=intcept;
```

Discussion of results
Model and data description

```
O=============================================O
| National Health Interview Data Nominal Model |
O=============================================O

            Model and Data Descriptions

    Sampling Distribution                   = Multinomial
    Link Function                           = Logistic
    Number of Level-3 Units                 = 57
    Number of Level-2 Units                 = 399
    Number of Level-1 Units                 = 6444
    Number of Level-2 Units per Level-3 Unit =
      3     5     7     2     4     3     5    14     9     4     7     6
     13    12     9    14     6     5     3    11     4     4    11     8
      4     6     4    10     4     9    10     2     7     6     5     6
      7     5     8     6     5     6    19     9     8    18     4     9
      2     8     2     4     7     7     4     6    13
    Number of level-1 units for the first (level-3, level-2) unit combination =
      5     6     9
```

The first part of the output file gives a description of the model specifications. This is followed by a data summary of the number of observations nested within each subject.

Descriptive statistics and starting values

The data summary is followed by descriptive statistics for all the variables included in the model.

```
O==========================================================O
| Descriptive statistics for all the variables in the model |
O==========================================================O
                                                  Standard
       Variable    Minimum    Maximum      Mean   Deviation
       --------    -------    -------      ----   ---------
       PASTVIS1     0.0000     1.0000     0.0691     0.2536
       PASTVIS2     0.0000     1.0000     0.3678     0.4822
       PASTVIS3     0.0000     1.0000     0.2858     0.4519
       PASTVIS4     0.0000     1.0000     0.2773     0.4477
       intcept      1.0000     1.0000     1.0000     0.0000
       NUMMED       0.0000     6.0000     1.5286     1.6864
       GENDER       0.0000     1.0000     0.4196     0.4935
       PRIMCARE     0.0000     1.0000     0.3993     0.4898
       INJURY       0.0000     1.0000     0.0906     0.2871
       BLODPRES     0.0000     1.0000     0.4008     0.4901
       URINE        0.0000     1.0000     0.1041     0.3055
       XRAY         0.0000     1.0000     0.0380     0.1913
       EXERCISE     0.0000     1.0000     0.1047     0.3063
       AGE1         0.0000     1.0000     0.1966     0.3975
       AGE2         0.0000     1.0000     0.0644     0.2455
       AGE3         0.0000     1.0000     0.2284     0.4199
       AGE4         0.0000     1.0000     0.2668     0.4423
       AGE5         0.0000     1.0000     0.1155     0.3196
```

Each category of the nominal outcome variable is denoted as PASTVIS_i, $i = 1, 2, 3, 4$. From the output it can be seen that the distribution of respondents over these categories are 6.9%, 36.8%, 28.6%, and 27.7% respectively. The age distribution is given in Table 9.

Table 9: Age distribution of respondents

Age	Percentage
Younger than 15 (AGE1)	19.7
15 to 24 (AGE2)	6.4
25 to 44 (AGE3)	22.8
45 to 64 (AGE4)	26.7
65 to 74 (AGE5)	11.6
75 and older	12.8*

*: calculated as $100 - (19.7+6.4+22.8+26.7+11.6)$

```
nihs1.OUT                                                         ─  □  ✕

O====================================================O
| Results for the model without any random effects |
O====================================================O

                      Goodness of fit statistics

   Statistic                                    Value          DF
   ---------                                    -----          --
   Likelihood Ratio Chi-square               15994.0450        6402
   Pearson Chi-square                        19269.4871        6402

                     Estimated regression weights

                              Standard
   Parameter      Estimate       Error    z Value    P Value
   ---------      --------    --------    -------    -------

   Response Code  1 vs Code   4
   -----------------------------
   intcept        -1.4038       0.1876    -7.4847     0.0000
   NUMMED         -0.2022       0.0373    -5.4238     0.0000
   GENDER          0.0325       0.1110     0.2926     0.7698
   PRIMCARE       -0.5933       0.1273    -4.6618     0.0000
   INJURY          0.0623       0.1883     0.3311     0.7406
   BLODPRES        0.7097       0.1276     5.5633     0.0000
   URINE           0.4633       0.1634     2.8359     0.0046
   XRAY            0.0299       0.3026     0.0988     0.9213
   EXERCISE        0.3110       0.1690     1.8399     0.0658
   AGE1            0.2907       0.2186     1.3298     0.1836
   AGE2            0.2566       0.2607     0.9842     0.3250
   AGE3            0.0970       0.1934     0.5016     0.6160
   AGE4            0.1434       0.1906     0.7523     0.4519
   AGE5           -0.0250       0.2305    -0.1086     0.9135

                              III
```

The estimated parameters for the model, assuming no random effects, are reported next. For each response code i versus code 4, $i = 1, 2, 3,$ there are 14 parameter estimates. Only the estimates for response code 1 versus response code 4 are displayed. Comparing these estimates with those obtained when allowance is made for the hierarchical structure of the data, a considerable difference is apparent.

Fixed effects estimates and fit statistics

The final results obtained using adaptive quadrature are given next. Using 8 quadrature points, 6 iterations were required to reach convergence. The deviance statistic $(-2\ln L)$ allows the user to compare the current model with other nested models.

```
nihs1.OUT                                                    □ ▣ ✕

O==========================================O
| Optimization Method: Adaptive Quadrature |
O==========================================O

    Number of quadrature points =            8
    Number of free parameters =             48
    Number of iterations used =              5

    -2lnL (deviance statistic) =      14329.74536
    Akaike Information Criterion      14425.74536
    Schwarz Criterion                 14750.74879


                    Estimated regression weights

                               Standard
    Parameter     Estimate       Error     z Value     P Value
    ---------     --------     --------     -------     -------

    Response Code  1 vs Code   4
    ----------------------------
    intcept       -1.5012       0.2483     -6.0447      0.0000
    NUMMED        -0.3325       0.0446     -7.4481      0.0000
    GENDER        -0.0355       0.1285     -0.2764      0.7822
    PRIMCARE      -1.0197       0.1954     -5.2190      0.0000
    INJURY         0.2728       0.2221      1.2282      0.2194
    BLODPRES       0.3133       0.1707      1.8351      0.0665
    URINE          0.3527       0.1962      1.7979      0.0722
    XRAY          -0.0141       0.3320     -0.0423      0.9662
    EXERCISE       0.7546       0.2047      3.6868      0.0002
    AGE1           1.0955       0.2621      4.1803      0.0000
    AGE2           1.3540       0.3013      4.4941      0.0000
    AGE3           0.9306       0.2249      4.1374      0.0000
    AGE4           0.7571       0.2165      3.4969      0.0005
    AGE5           0.0916       0.2545      0.3598      0.7190

                              III
```

A study of the p-values associated with the parameter estimates indicates that the estimated GENDER, INJURY, URINE, and XRAY coefficients are not significant, regardless of the values of the category of the outcome variable.

Random effect estimates

The last part of the output file shows the variance estimates for the level-2 and level-3 random effects. Both effects are highly significant.

```
nihs1.OUT
                Estimated level 2 variances and covariances

                                    Standard
        Parameter           Estimate    Error     z Value   P Value
        ---------           --------   --------   -------   -------
        intcept1/intcept1    4.7004     0.4312    10.8997   0.0000
        intcept2/intcept2    3.2406     0.2791    11.6099   0.0000
        intcept3/intcept3    1.0800     0.1386     7.7938   0.0000

                Estimated level 3 variances and covariances

                                    Standard
        Parameter           Estimate    Error     z Value   P Value
        ---------           --------   --------   -------   -------
        intcept1/intcept1    1.0122     0.2513     4.0287   0.0001
        intcept2/intcept2    0.9227     0.1875     4.9221   0.0000
        intcept3/intcept3    0.8511     0.1523     5.5881   0.0000
```

Interpreting the output
Estimated unit-specific probabilities

The estimated regression coefficients given in the adaptive quadrature portion of the output provide the information necessary to compute unit-specific probabilities for a typical participant that is associated with each possible combination of the predictor variables. For example, consider a typical female patient (GENDER $= 0$) that received 3 medications (NUMMED $= 3$), has primary care (PRIMCARE $= 1$), had no injuries (INJURY $= 0$), did not have a blood pressure or urine test (BLODPRES $=$ URINE $= 0$), does not exercise (EXERCISE $= 0$), and is in the age group 25 to 44 (AGE3 $= 1$).

For response code 1 vs. code 4:

$$\hat{\eta}_{ijk1} = -1.5004 - 0.3320(\text{NUMMED}_{ijk}) - 0.0395(\text{GENDER}_{ijk}) - 1.0176(\text{PRIMCARE}_{ijk})$$
$$+ \dots 1.0945(\text{AGE1}_{ijk}) + 1.3539(\text{AGE2}_{ijk}) + 0.9306(\text{AGE3}_{ijk})$$
$$+ 0.7572(\text{AGE4}_{ijk}) + 0.1136(\text{AGE5}_{ijk})$$
$$= -1.5004 - 3(0.3320) - 1(1.0176) + 1(0.9306)$$
$$= -2.5834$$

so that $\exp\left(\hat{\eta}_{ijk1}\right) = 0.0755$.

For response code 2 vs. 4, we find that

$$\hat{\eta}_{ijk2} = 0.3737 - 3(0.2360) - 1(0.9167) + 1(0.6972)$$
$$= -0.5538$$

and thus

$$\exp\left(\hat{\eta}_{ijk2}\right) = 0.5748.$$

For response code 3 vs. code 4

$$\hat{\eta}_{ijk3} = 0.3440 - 3(0.0718) - 1(0.3004) + 1(0.2070)$$
$$= 0.0352$$

and thus

$$\exp\left(\hat{\eta}_{ijk3}\right) = 1.0358.$$

Using these values, it follows that

$$Prob(respondent\ not\ seen\ doctor\ previously)$$

$$= \frac{0.0755}{1 + 0.0755 + 0.5748 + 1.0358}$$

$$= 0.0281.$$

The next two tables contain a selection of unit-specific probabilities for the four categories of PASTVIS for females (GENDER = 0).

Unit-specific probabilities for females with XRAY = no, INJURY = no, URINE = no, and BLODPRES = no

NUMMED	PRIM CARE	EXER CISE	AGE	ETA1	ETA2	ETA3	PROB1	PROB2	PROB3	PROB4
none	no	no	< 15	-0.406	1.393	0.628	0.088	0.532	0.248	0.132
none	no	no	25:44	-0.570	1.071	0.551	0.091	0.469	0.279	0.161
none	no	no	>= 75	-1.500	0.374	0.344	0.055	0.356	0.345	0.245
none	no	yes	< 15	0.347	1.776	0.953	0.130	0.541	0.238	0.092
none	no	yes	25:44	0.183	1.454	0.876	0.135	0.482	0.270	0.113
none	no	yes	>= 75	-0.748	0.757	0.669	0.085	0.384	0.351	0.180
none	yes	no	< 15	-1.424	0.476	0.328	0.057	0.380	0.327	0.236
none	yes	no	25:44	-1.587	0.154	0.251	0.056	0.319	0.351	0.274
none	yes	no	>= 75	-2.518	-0.543	0.044	0.030	0.215	0.386	0.370
none	yes	yes	< 15	-0.671	0.859	0.653	0.088	0.408	0.332	0.173
none	yes	yes	25:44	-0.835	0.537	0.576	0.088	0.348	0.361	0.203
none	yes	yes	>= 75	-1.765	-0.160	0.369	0.049	0.246	0.417	0.288
three	no	no	< 15	-1.402	0.685	0.413	0.052	0.418	0.319	0.211
three	no	no	25:44	-1.566	0.363	0.336	0.052	0.355	0.346	0.247
three	no	no	>= 75	-2.496	-0.334	0.129	0.028	0.244	0.387	0.341
three	no	yes	< 15	-0.649	1.068	0.738	0.080	0.446	0.321	0.153
three	no	yes	25:44	-0.813	0.746	0.661	0.081	0.384	0.353	0.182
three	no	yes	>= 75	-1.744	0.049	0.454	0.046	0.276	0.414	0.263
three	yes	no	< 15	-2.420	-0.232	0.112	0.030	0.264	0.373	0.333
three	yes	no	25:44	-2.584	-0.554	0.035	0.028	0.214	0.386	0.372
three	yes	no	>= 75	-3.514	-1.251	-0.172	0.014	0.133	0.390	0.463
three	yes	yes	< 15	-1.667	0.151	0.437	0.048	0.298	0.397	0.256

```
three   yes   yes  25:44   -1.831  -0.171   0.360  0.047  0.245  0.417  0.291
three   yes   yes  >= 75   -2.761  -0.868   0.153  0.024  0.158  0.440  0.378
```

From these tables we conclude that the proportion of female patients, regardless of age group, that indicated no prior visits to a medical practitioner (PASTVIS = 1) is generally low. Females who exercise have a lower probability of having several past visits when compared to those who do not exercise.

Unit-specific probabilities for females with XRAY = no, INJURY = no, URINE = no, and BLODPRES = yes

NUMMED	PRIM CARE	EXER CISE	AGE	ETA1	ETA2	ETA3	PROB1	PROB2	PROB3	PROB4
none	No	no	< 15	-0.096	1.712	0.839	0.093	0.567	0.237	0.102
none	No	no	25:44	-0.260	1.390	0.762	0.097	0.506	0.270	0.126
none	No	no	>=75	-1.191	0.693	0.555	0.060	0.396	0.345	0.198
none	No	yes	< 15	0.657	2.095	1.164	0.135	0.570	0.225	0.070
none	No	yes	25:44	0.493	1.773	1.088	0.142	0.512	0.258	0.087
none	No	yes	>=75	-0.438	1.076	0.880	0.092	0.420	0.345	0.143
none	Yes	no	< 15	-1.114	0.795	0.539	0.062	0.421	0.326	0.190
none	Yes	no	25:44	-1.278	0.473	0.462	0.062	0.359	0.355	0.224
none	Yes	no	>=75	-2.208	-0.224	0.255	0.034	0.250	0.403	0.313
none	Yes	yes	< 15	-0.361	1.178	0.864	0.095	0.444	0.324	0.137
none	Yes	yes	25:44	-0.525	0.856	0.787	0.096	0.383	0.358	0.163
none	Yes	yes	>=75	-1.456	0.159	0.580	0.056	0.280	0.426	0.239
three	No	no	< 15	-1.092	1.004	0.624	0.057	0.460	0.315	0.169
three	No	no	25:44	-1.256	0.682	0.547	0.057	0.396	0.346	0.200
three	No	no	>=75	-2.187	-0.015	0.340	0.032	0.281	0.401	0.286
three	No	yes	< 15	-0.339	1.387	0.949	0.086	0.482	0.311	0.120
three	No	yes	25:44	-0.503	1.065	0.872	0.088	0.421	0.347	0.145
three	No	yes	>=75	-1.434	0.368	0.665	0.052	0.312	0.420	0.216
three	Yes	no	< 15	-2.110	0.087	0.323	0.034	0.304	0.384	0.278
three	Yes	no	25:44	-2.274	-0.235	0.246	0.032	0.249	0.403	0.315
three	Yes	no	>=75	-3.204	-0.932	0.039	0.016	0.159	0.420	0.404
three	Yes	yes	< 15	-1.357	0.470	0.649	0.054	0.335	0.401	0.210
three	Yes	yes	25:44	-1.521	0.148	0.572	0.053	0.279	0.427	0.241
three	Yes	yes	>=75	-2.452	-0.549	0.365	0.028	0.186	0.464	0.322

Estimated population-average probabilities

The population-average probabilities are obtained by dividing the ETA1, ETA2 and ETA3 values given in the previous two tables by the square root of the corresponding design effects. For the intercepts-only model, this quantity is obtained as

$$\hat{d}_c = \left[\mathrm{var}\left(v_{ij00}\right) + \mathrm{var}\left(v_{ic0}\right) + \mathrm{var}\left(e_{ijk}\right) \right] / \mathrm{var}\left(e_{ijk}\right), \quad c = 1, 2, 3.$$

For the logistic model it is assumed that

$$\mathrm{var}\left(e_{ijk}\right) = \frac{\pi^2}{3} = 3.290.$$

Therefore

$$\sqrt{d_1} = \sqrt{(4.707 + 1.009 + 3.290)/3.290}$$

$$= \sqrt{2.737}$$

$$= 1.6545.$$

Similarly,

$$\sqrt{d_2} = \sqrt{(3.237 + 0.921 + 3.290)/3.290}$$

$$= \sqrt{2.264}$$

$$= 1.5046$$

and

$$\sqrt{d_3} = \sqrt{(1.077 + 0.848 + 3.290)/3.290}$$

$$= \sqrt{1.585}$$

$$= 1.2590.$$

Using these values, we obtain the population-average probabilities for the four categories of PASTVIS for a female respondent. Summaries of a selected number of population-average probabilities are given in the tables below.

Population-average probabilities for females with XRAY = no, INJURY = no, URINE = no, and BLODPRES = no

NUMMED	PRIM CARE	EXER CISE	AGE	ETA1	ETA2	ETA3	PROB1	PROB2	PROB3	PROB4
none	No	no	< 15	−0.245	0.926	0.499	0.131	0.424	0.277	0.168
none	No	no	25:44	−0.344	0.712	0.438	0.134	0.385	0.293	0.189
none	No	no	>= 75	−0.907	0.248	0.273	0.101	0.320	0.329	0.250
none	No	yes	< 15	0.210	1.180	0.757	0.162	0.427	0.280	0.131
none	No	yes	25:44	0.111	0.966	0.696	0.165	0.389	0.297	0.148
none	No	yes	>= 75	−0.452	0.503	0.532	0.128	0.331	0.341	0.200
none	Yes	no	< 15	−0.860	0.317	0.260	0.103	0.335	0.317	0.244
none	Yes	no	25:44	−0.960	0.102	0.199	0.103	0.299	0.329	0.269
none	Yes	no	>= 75	−1.522	−0.361	0.035	0.074	0.236	0.351	0.339
none	Yes	yes	< 15	−0.405	0.571	0.519	0.130	0.346	0.328	0.195
none	Yes	yes	25:44	−0.504	0.357	0.457	0.131	0.310	0.343	0.217
none	Yes	yes	>= 75	−1.067	−0.106	0.293	0.096	0.251	0.374	0.279
three	No	no	< 15	−0.847	0.455	0.328	0.098	0.359	0.316	0.228
three	No	no	25:44	−0.946	0.241	0.267	0.098	0.321	0.329	0.252
three	No	no	>= 75	−1.509	−0.222	0.102	0.071	0.256	0.354	0.320
three	No	yes	< 15	−0.392	0.710	0.586	0.123	0.369	0.326	0.182
three	No	yes	25:44	−0.491	0.496	0.525	0.124	0.332	0.342	0.202
three	No	yes	>= 75	−1.054	0.032	0.360	0.091	0.271	0.376	0.262
three	Yes	no	< 15	−1.462	−0.154	0.089	0.073	0.269	0.344	0.314
three	Yes	no	25:44	−1.562	−0.368	0.028	0.072	0.236	0.351	0.341
three	Yes	no	>= 75	−2.124	−0.831	−0.136	0.049	0.179	0.359	0.412
three	Yes	yes	< 15	−1.007	0.101	0.347	0.094	0.285	0.364	0.257
three	Yes	yes	25:44	−1.107	−0.114	0.286	0.093	0.251	0.375	0.281
three	Yes	yes	>= 75	−1.669	−0.577	0.122	0.065	0.195	0.392	0.347

Population-average probabilities for females with XRAY = no, INJURY = no, URINE = no, and BLODPRES = no

NUMMED	PRIM CARE	EXER CISE	AGE	ETA1	ETA2	ETA3	PROB1	PROB2	PROB3	PROB4
none	No	no	< 15	−0.058	1.138	0.667	0.135	0.445	0.278	0.143
none	No	no	25:44	−0.157	0.924	0.605	0.138	0.406	0.295	0.161
none	No	no	>= 75	−0.720	0.460	0.441	0.105	0.343	0.336	0.216
none	No	yes	< 15	0.397	1.392	0.925	0.165	0.446	0.279	0.111
none	No	yes	25:44	0.298	1.178	0.864	0.169	0.408	0.298	0.126
none	No	yes	>= 75	−0.265	0.715	0.699	0.132	0.351	0.346	0.172
none	Yes	no	< 15	−0.673	0.529	0.428	0.108	0.358	0.324	0.211
none	Yes	no	25:44	−0.772	0.314	0.367	0.108	0.320	0.338	0.234
none	Yes	no	>= 75	−1.335	−0.149	0.202	0.079	0.257	0.366	0.299
none	Yes	yes	< 15	−0.218	0.783	0.686	0.134	0.366	0.332	0.167
none	Yes	yes	25:44	−0.317	0.569	0.625	0.136	0.329	0.348	0.186
none	Yes	yes	>= 75	−0.880	0.106	0.461	0.101	0.270	0.386	0.243
Three	No	no	< 15	−0.660	0.667	0.496	0.101	0.382	0.321	0.196
Three	No	no	25:44	−0.759	0.453	0.434	0.102	0.343	0.337	0.218
Three	No	no	>= 75	−1.322	−0.010	0.270	0.075	0.278	0.367	0.280
Three	No	yes	< 15	−0.205	0.922	0.754	0.126	0.390	0.329	0.155
Three	No	yes	25:44	−0.304	0.708	0.693	0.128	0.352	0.347	0.173
Three	No	yes	>= 75	−0.867	0.244	0.528	0.096	0.291	0.386	0.228
Three	Yes	no	< 15	−1.275	0.058	0.257	0.077	0.292	0.356	0.275
Three	Yes	no	25:44	−1.374	−0.156	0.196	0.076	0.257	0.366	0.301
Three	Yes	no	>= 75	−1.937	−0.619	0.031	0.053	0.198	0.380	0.368
Three	Yes	yes	< 15	−0.820	0.313	0.515	0.098	0.305	0.374	0.223
Three	Yes	yes	25:44	−0.919	0.098	0.454	0.098	0.271	0.386	0.245
Three	Yes	yes	>= 75	−1.482	−0.365	0.290	0.070	0.213	0.410	0.307

2.4.13.5   A random intercept model with ten predictors

Setting up the analysis

In the previous example, we included 14 possible predictors of PASTVIS in the fixed part of the model. The output indicated that the variables GENDER, INJURY, URINE and XRAY did not contribute significantly to explaining the variation in PASTVIS outcomes.

To run the model without these fixed effects, use the **File**, **Open Syntax File** option and select the command syntax previously saved to the file **NHIS1.prl**. Delete the variables GENDER, INJURY, URINE and XRAY from the Predictors paragraph and save the modified syntax file as **NHIS2.prl**. To run this syntax file, select the **Run** option from the **Analysis** menu.

```
MGlimOptions Converge=0.0001 MaxIter=100 MissingCode=-999999
            Method=Quad NQUADPTS=8 RefCat=last;
Title= National Health Interview Data Nominal model ;
SY=nih_subset.LSF;
ID2=CPSUM;
ID3=CSTRATM;
Distribution=MUL;
Link=LOGIT;
Intercept=Yes;
DepVar=PASTVIS;
CoVars=NUMMED PRIMCARE BLODPRES EXERCISE AGE1 AGE2 AGE3 AGE4 AGE5;
RANDOM2=intcept;
RANDOM3=intcept;
```

Interpreting the output
Fit statistics

Only a portion of the output file **NIHS2.out** is shown below. Recall that the deviance statistic for the previous model was 14329.75, with 48 free parameters. For the current model, the deviance statistic is equal to 14351.80 and the number of free parameters is equal to 36. To test whether the removal of GENDER, INJURY, URINE and XRAY made a significant difference to the model fit, we use the fact that the difference in deviance statistics for two nested models follows a χ^2-distribution with degrees of freedom equal to the difference in the number of parameters estimated.

The χ^2-value obtained for this test is 14351.80 – 14329.75 = 22.05, with 12 degrees of freedom. Since the associated p-value equals 0.04, the χ^2-value is significant at the 5% level, but not at the 1% level of significance. We therefore conclude that, based on the χ^2-difference test, we do not have a definitive answer to the question of whether the 4 predictors should remain in the model or not. A summary of the Akaike and Schwarz criteria is shown in Table 14.

Table 14: Akaike and Schwarz fit criteria for two nested models

Fit statistic	14 predictors	10 predictors
Akaike	14425.75	14423.80
Schwarz	14750.75	14667.55

Each of these criteria states that the model with the smallest value is the model to be selected. Based on this decision rule, we conclude that the model without the four predictors should be used, since it is more parsimonious and very little information regarding the explanation of variation in PASTVIS is lost.

Odds ratios and 95% confidence intervals for the odds ratios

An odds ratio of 1 indicates the event under study is equally likely in both the outcome category of interest and in the reference category. An odds ratio greater than 1 indicates that the event is more likely to occur in the category of interest.

The intercept coefficient is the expected log-odds that a participant in the present study indicated no past visits (PASTVIS = 1) relative to the category PASTVIS = 4 (6 or more visits), given that the remaining predictors are held constant at zero. The estimated conditional expected log-odds is – 1.4156, corresponding to an odds ratio of exp(–1.4156)=0.2428. This implies that a qualifying participant (a participant with NUMMED = 0, GENDER = 0, …, AGE5 = 0) has 0.2427 times the odds of having had no previous visits, as opposed to 6 or more visits.

```
nihs2.OUT                                                    ▢ ▣ ✖
     Response Code  1 vs Code   4
     ----------------------------
     intcept          -1.4149      0.2429      0.1525      0.3871
     NUMMED           -0.3414      0.7108      0.6517      0.7753
     PRIMCARE         -1.0147      0.3625      0.2490      0.5278
     BLODPRES          0.3338      1.3963      1.0033      1.9431
     EXERCISE          0.7986      2.2225      1.4941      3.3060
     AGE1              1.0546      2.8707      1.7234      4.7817
     AGE2              1.3594      3.8940      2.1604      7.0186
     AGE3              0.9288      2.5316      1.6309      3.9296
     AGE4              0.7429      2.1021      1.3765      3.2101
     AGE5              0.0656      1.0678      0.6491      1.7566


     Response Code  2 vs Code   4
     ----------------------------
     intcept           0.4636      1.5898      1.1447      2.2080
     NUMMED           -0.2382      0.7881      0.7456      0.8329
     PRIMCARE         -0.9058      0.4042      0.3057      0.5345
     BLODPRES          0.2925      1.3398      1.0492      1.7108
     EXERCISE          0.3891      1.4757      1.1091      1.9635
     AGE1              0.9978      2.7122      1.9344      3.8027
     AGE2              1.2602      3.5261      2.3691      5.2483
     AGE3              0.6881      1.9899      1.4823      2.6713
     AGE4              0.6576      1.9302      1.4647      2.5435
     AGE5              0.0956      1.1003      0.8053      1.5033
```

The 95% confidence interval for the odds ratio is obtained by first computing a 95% confidence interval for the intercept coefficient. This confidence interval is given by

$$\hat{\beta}_0 \pm 1.96 \; std.error\left(\hat{\beta}_0\right).$$

From the output, it follows that this interval is

$$\left(-1.4161 - 1.96 \times 0.2377; -1.4161 + 1.96 \times 0.2377\right)$$
$$= \left(-1.8822; -0.9500\right).$$

Using these values, we obtain the 95% confidence interval for the odds ratio as

$$\big(\exp(-1.8822); \exp(-0.9500)\big)$$
$$= (0.1523; 0.3867).$$

2.5 Evaluation

2.5.1 Introduction

In this section we report the results of simulation studies for the evaluation of the various estimation procedures and outcome variable types available in the LISREL MGLIM (multilevel generalized linear models) module. These simulation studies were also done to determine the effect of fitting models to data that are missing at random (MAR). We started by doing a comparison of the performance of a continuous outcome variable versus a binary outcome variable. The same datasets were used, but for the binary case the outcome variable was dichotomized based on the threshold concept. The results of this study are reported next.

2.5.2 Parameter estimation results for normal and binary outcome variables with complete data and with MAR observations

The simulation results given below shows that the ML estimates for continuous normal data have good properties for both the complete data and missing data sets. However, in the binary case parameter estimates using adaptive Gaussian quadrature (10 points) have good properties for the complete data cases but poor properties when there are missing observations.

Simulation results: normal and binary outcome variables

		β_0	β_1	β_2	β_3	σ_0^2	$\sigma_0\sigma_1$	σ_1^2
Normal	True	5.32	0.12	-0.23	-0.8	0.32	0.06	0.28
Complete	AE	5.34	0.07	-0.26	-0.75	0.30	0.06	0.25
	RB	0.02	-0.05	-0.03	0.05	-0.02	0.00	-0.03
	PB	0.38	-44.24	12.22	-6.66	-5.17	-3.16	-11.11
	SB	19.07	-66.31	-19.16	47.76	-13.16	-2.53	-45.30
	RMSE	0.11	0.10	0.15	0.12	0.13	0.08	0.08
	ASE	0.11	0.08	0.15	0.11	0.13	0.07	0.07
	CR	0.94	0.92	0.94	0.94	0.94	0.96	0.90
Missing	AE	5.32	0.12	-0.24	-0.80	0.30	0.07	0.27
	RB	0.00	0.00	-0.01	0.00	-0.02	0.01	-0.01
	PB	0.04	-1.13	3.91	-0.24	-6.63	9.25	-2.92
	SB	2.02	-1.84	-6.23	1.84	-17.75	8.14	-13.13
	RMSE	0.10	0.07	0.14	0.11	0.12	0.07	0.06

	ASE	0.10	0.08	0.15	0.11	0.12	0.07	0.06
	CR	0.95	0.95	0.96	0.95	0.94	0.94	0.95
Binary	True	6.025	0.281	-1.477	-1.587	7.43	-2.27	3.12
Complete	AE	6.215	0.232	-1.711	-1.481	7.429	-2.241	3.207
	RB	0.190	-0.049	-0.234	0.106	-0.001	0.029	0.087
	PB	3.159	-17.500	15.872	-6.708	-0.013	-1.291	2.801
	SB	8.381	-4.670	-11.190	10.830	-0.023	1.607	4.836
	RMSE	2.279	1.054	2.108	0.989	4.308	1.823	1.809
	ASE	1.701	0.873	1.322	0.777	4.458	1.796	1.472
	CR	0.930	0.955	0.957	0.960	0.888	0.906	0.921
Missing	AE	5.280	0.277	-1.624	-0.974	4.277	-1.528	2.139
	RB	-0.745	-0.004	-0.147	0.613	-3.153	0.742	-0.981
	PB	-12.367	-1.460	9.966	-38.654	-42.436	-32.700	-31.439
	SB	-35.743	-0.380	-7.718	61.733	-92.798	47.864	-60.941
	RMSE	2.214	1.080	1.913	1.168	4.636	1.719	1.885
	ASE	1.575	0.949	1.239	0.834	3.284	1.457	1.319
	CR	0.787	0.948	0.949	0.809	0.662	0.776	0.696

TRUE= The parameter values used to generate data
AE= Average Estimate
RB= Raw bias
PB= Percent bias
SB= Standardized bias
RMSE= Root Mean Square Error
ASE= Average Asymptotic Standard Error of an estimate
CR= Proportion Coverage

Table 16: Average percentage missing at time points

T1	T2	T3	T4
0	3.689 %	12.073%	17.606 %

It is speculated that the reason for the poor performance for the MAR binary case may mainly be contributed to the fact that the distribution of the binary outcome is so severely skewed that when missing values are present, some of the level-2 units contained only a single value for the outcome variable. To test this hypothesis, we dichotomized the continuous variable in such a way that the distribution of values is approximately 15% and 85%. The table below is a summary of this simulation study indicating that as the outcome variable distribution becomes more symmetric, the behavior of parameter estimation under MAR improves.

2.5.3 Parameter estimation results for the binary outcome variables with complete data and with MAR observations using a smaller intercept parameter

Table 17: Simulation results: binary outcome variables

		β_0	β_1	β_2	β_3	σ_0^2	$\sigma_0\sigma_1$	σ_1^2
Binary	True	2.247	-1.037	0.236	-0.877	0.36	0.06	0.28
Complete	AE	2.390	-1.131	0.240	-0.887	0.778	-0.274	0.570
	RB	0.143	-0.094	0.004	-0.010	0.458	-0.334	0.290
	PB	6.383	9.027	1.623	1.122	143.269	-557.010	103.430
	SB	36.043	-32.647	0.832	-2.735	51.987	-58.165	55.630
	RMSE	0.423	0.302	0.460	0.360	0.994	0.665	0.596
	ASE	0.398	0.289	0.463	0.359	0.807	0.533	0.495
	CR	0.955	0.962	0.959	0.944	0.862	0.955	0.891
Missing	AE	2.253	-0.850	0.217	-0.974	0.508	-0.140	0.445
	RB	0.006	0.187	-0.019	-0.097	0.188	-0.200	0.165
	PB	0.260	-18.021	-8.019	11.053	58.852	-333.770	59.080
	SB	1.571	64.383	-4.204	-25.893	27.425	-45.933	37.955
	RMSE	0.372	0.345	0.451	0.387	0.712	0.480	0.466
	ASE	0.377	0.295	0.451	0.371	0.620	0.422	0.435
	CR	0.937	0.881	0.957	0.948	0.781	0.968	0.853

Table 18: Average percentage missing at time points

T1	T2	T3	T4
0	5.05	15.85	28.84

2.5.4 Parameter estimation results for count outcome variables with complete data and with MAR observations

The simulation results given below are ML estimates of the parameters for a count outcome variable. The parameter estimates using adaptive Gaussian quadrature (10 points) have good properties for both the complete data and missing observations cases.

Table 19: Simulation results: count outcome variables

Count	True	1.1	0.002	0.05	-0.31	0.51	0.054	0.24
Complete	AE	1.093	-0.003	0.060	-0.306	0.502	0.056	0.238
	RB	-0.007	-0.005	0.010	0.004	-0.008	0.002	-0.002
	PB	-0.621	-246.119	20.278	-1.186	-1.490	4.061	-0.789
	SB	-7.781	-7.547	8.216	4.115	-9.494	5.212	-4.485
	RMSE	0.088	0.065	0.124	0.089	0.080	0.042	0.042
	ASE	0.090	0.065	0.124	0.091	0.079	0.040	0.040
	CR	0.964	0.949	0.941	0.959	0.935	0.933	0.921
Missing	AE	1.087	0.039	0.060	-0.311	0.506	0.037	0.228
	RB	-0.013	0.037	0.010	-0.001	-0.004	-0.017	-0.012
	PB	-1.186	1865.099	19.456	0.462	-0.774	-30.805	-5.071
	SB	-14.655	52.161	7.815	-1.505	-4.884	-36.162	-28.813
	RMSE	0.090	0.081	0.125	0.095	0.081	0.049	0.044
	ASE	0.091	0.071	0.125	0.096	0.080	0.043	0.042
	CR	0.960	0.903	0.943	0.957	0.939	0.925	0.912

Table 20: Average percentage missing at time points

T1	T2	T3	T4
0	4.02	17.25	30.59

2.5.5 Additional Simulation studies

Two-level binary and ordinal models without missing observations have been tested with simulated data. For the non-linear models two methods of estimation, these being ADAP (adaptive quadrature) and MAP (Maximization of the Posterior distribution), were evaluated. Results of those tests are tabulated below. Overall, the ADAP procedure performed quite well as opposed to the MAP estimators that exhibited high bias and lower coverage rate. It was found, however, that the MAP performance improves for larger numbers of level-1 units nested within the higher order units and in general, the MAP estimates provided good starting value for the ADAP procedure.

Note that the MAP (Maximization of the Posterior distribution) method is always used to obtain starting values for models with non-normal outcome variables. This method does not use numerical quadrature and for a very large number of random effects, it might be the only viable method in LISREL for fitting mixed effects models.

Results in the following tables are based on the 1000 simulated datasets; each with 150 subjects followed over five time points.
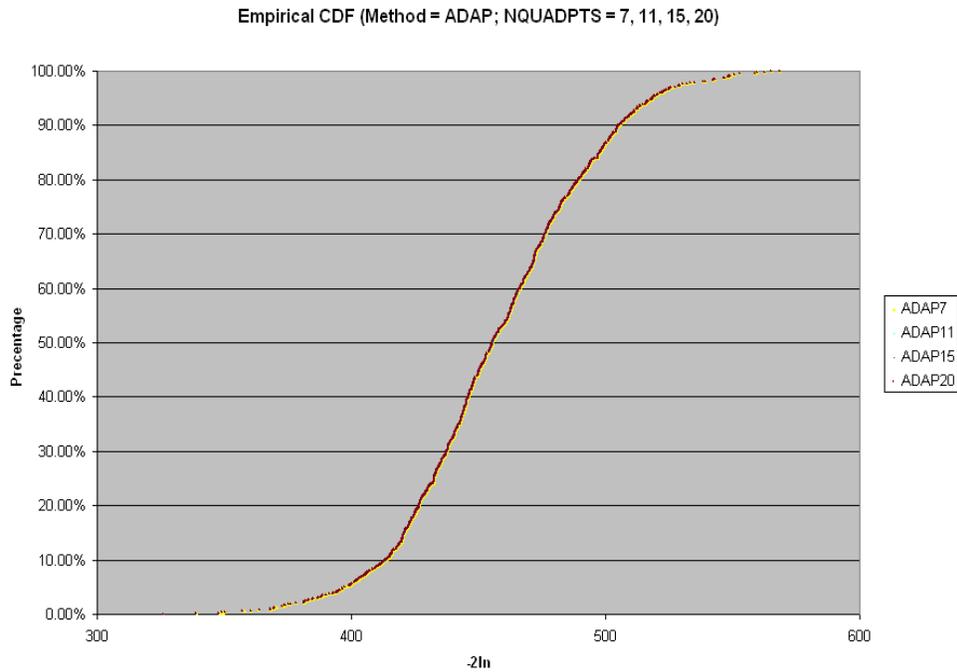
Table 21: Simulation results for a binary outcome using ADAP

	β_0	β_1	β_2	β_3	σ_0^2	$\sigma_0\sigma_1$	σ_1^2
True	6.025	0.281	-1.477	-1.587	7.43	-2.23	3.12
AE	6.282	0.252	-1.760	-1.511	7.814	-2.506	3.430
RB	0.257	-0.029	-0.283	0.076	0.384	-0.236	0.310
PB	4.257	-10.248	19.137	-4.790	5.173	10.410	9.948
SB	12.177	-2.369	-16.641	7.564	6.826	-8.932	14.869
RMSE	2.122	1.216	1.722	1.008	5.644	2.656	2.110
ASE	2.228	1.401	1.742	1.167	5.653	2.620	1.904
CR	0.895	0.947	0.942	0.969	0.853	0.887	0.920

Table 22: Simulation results for a binary outcome using MAP

	β_0	β_1	β_2	β_3	σ_0^2	$\sigma_0\sigma_1$	σ_1^2
True	6.025	0.281	-1.477	-1.587	7.43	-2.23	3.12
AE	4.283	-0.036	-1.193	-0.938	2.440	-0.595	0.851
RB	-1.742	-0.317	0.284	0.649	-4.990	1.675	-2.269
PB	-28.909	-112.769	-19.219	-40.868	-67.155	-73.774	-72.718
SB	-138.787	-42.866	21.867	83.311	-313.142	179.992	-305.950
RMSE	2.148	0.804	1.329	1.013	5.240	1.917	2.388
ASE	0.830	0.592	0.948	0.664	0.429	0.270	0.203
CR	0.284	0.854	0.844	0.760	0.038	0.180	0.052

A plot of the empirical CDF of the –2 log-likelihood function is presented for the adaptive quadrature procedure using 7, 11, 15 and 20 points respectively. For the simulated data these curves are essentially the same and it is therefore concluded that the choice of 10 points in the simulation study is acceptable.

Figure 8: Plot of empirical CDF of -2 ln L function using ADAP

Table 23: Simulation results for an ordinal outcome using ADAP

	T_0	T_1	β_1	β_2	β_3	σ_0^2	$\sigma_0\sigma_1$	σ_1^2
True	3.912	6.528	-0.875	0.111	-1.724	7.13	-0.41	2.06
AE	3.964	6.606	-0.883	0.134	-1.754	7.335	-0.442	2.168
RB	0.052	0.078	-0.008	0.023	-0.030	0.205	-0.032	0.108
PB	1.323	1.197	0.887	20.963	1.726	2.873	7.704	5.234
SB	14.197	15.398	-3.174	4.117	-8.137	9.208	-4.279	15.352
RMSE	0.368	0.513	0.245	0.566	0.367	2.234	0.739	0.711
ASE	0.456	0.409	0.246	0.563	0.357	2.123	0.734	0.669
CR	0.980	0.883	0.950	0.953	0.946	0.929	0.942	0.938

Table 24: Simulation results for an ordinal outcome using MAP

	T_0	T_1	β_1	β_2	β_3	σ_0^2	$\sigma_0\sigma_1$	σ_1^2
True	3.912	6.528	-0.875	0.111	-1.724	7.13	-0.41	2.06
AE	3.389	5.652	-0.750	0.114	-1.453	3.930	0.442	0.902
RB	-0.523	-0.876	0.125	0.003	0.271	-3.200	0.852	-1.158
PB	-13.382	-13.414	-14.325	3.039	-15.699	-44.887	-207.772	-56.190
SB	-196.585	-240.885	63.221	0.717	91.575	-249.059	198.558	-382.912
RMSE	0.588	0.949	0.235	0.471	0.401	3.450	0.954	1.197
ASE	0.330	0.306	0.175	0.420	0.255	0.814	0.344	0.229
CR	0.666	0.227	0.853	0.922	0.778	0.161	0.282	0.044

2.6   Theory

2.6.1 Distribution models and link functions

2.6.1.1 Introduction

It is assumed that y_{ijk} is an outcome variable, where $i = 1, 2, ..., N$ denotes level-3 units and $j = 1, 2, ..., n_i$ denotes level-2 units, nested within each level-3 unit i. The level-1 units $k = 1, 2, ..., n_{ij}$ are nested within the (i, j)-th (level-3; level-2) combination.

For 2-level models, the subscript i is omitted and y_{jk} denotes level-1 unit k nested within level-2 unit j.

A multilevel model with a non-normal outcome variable is transformed to a linear model by using a *link* function which defines the relationship between the dependent variable η_{ijk} of the linear model and the mean μ_{ijk} of the *distribution* selected. More specifically, the linear model of a multilevel generalized linear model is given by

$$\eta_{ijk} = \mathbf{x}_{ijk}^{'}\boldsymbol{\beta} + \mathbf{z}_{(2)ijk}^{'}\mathbf{v}_{ij} + \mathbf{z}_{(3)ijk}^{'}\mathbf{v}_i,$$

where $\mathbf{x}_{ijk}$ is a $p \times 1$ vector of predictors, $\mathbf{z}_{(2)ijk}$ is a $q \times 1$ design vector associated with the level-2 random effects $\mathbf{v}_{ij}$. Likewise, $\mathbf{z}_{(3)ijk}$ is a $r \times 1$ design vector associated with the level-3 random effects $\mathbf{v}_i$. Typically, the elements of $\mathbf{z}_{(3)ijk}$ and $\mathbf{z}_{(2)ijk}$ are subsets of the elements of $\mathbf{x}_{ijk}$.

It is further assumed that the level-3 and level-2 random effect vectors are uncorrelated and also that $\mathbf{v}_i \sim N\left(\mathbf{0}, \boldsymbol{\Phi}_{(3)}\right)$ and that $\mathbf{v}_{ij} \sim N\left(\mathbf{0}, \boldsymbol{\Phi}_{(2)}\right)$.

2.6.1.2 Link function and derivatives

The link functions available are the log, logistic, complimentary log-log, log-log, and probit. Table 1 contains a summary of these link functions and their derivatives. The cumulative distribution for each link is denoted by $CDF(\eta)$ and the corresponding probability distribution function by PDF, where $PDF = \dfrac{\partial}{\partial \eta} CDF$. The second-order derivatives of η with respect to the link function is denoted by $\dfrac{\partial}{\partial \eta} PDF$. The CDF of a standardized normal variable is denoted by $\Phi(\cdot)$, while $c1 = \exp(-\eta)$, and $c2 = \exp(\eta) = \dfrac{1}{c1}$.

Table 25: Probability and cumulative distribution functions

Function	$CDF(\eta)$	$PDF(\eta)$	$\dfrac{\partial}{\partial \eta} PDF.$
Logistic	$\dfrac{1}{1+c1}$	$CDF(1-CDF)$	$(c1-c2)\times PDF^2$
Probit	$\Phi(\eta)$	$\dfrac{1}{\sqrt{2\pi}}\exp-\dfrac{1}{2}\eta^2$	$-\eta \times PDF$
Complementary log-log	$1-\exp(-c2)$	$c2(1-CDF)$	$(1-c2)PDF$
Log-log	$\exp(-c1)$	$c1\times CDF$	$(c1-1)PDF$
Log	$c2$	$c2$	$c2$

In subsequent sections, short descriptions of the different distribution-link type models are given.

2.6.1.3 The Poisson-log model

Assume y_{ijk} follows a Poisson distribution with mean μ_{ijk}. In other words, the probability density function of y_{ijk} is given by

$$f\left(y_{ijk},\mu_{ijk}\right) = \frac{e^{-\mu_{ijk}}\mu_{ijk}^{y_{ijk}}}{y_{ijk}!} \Rightarrow \ln f\left(y_{ijk},\mu_{ijk}\right) = y_{ijk}\ln\{\mu_{ijk}\} - \mu_{ijk} - \ln\{y_{ijk}!\} \tag{1.1}$$

and the variance of y_{ijk} is given by

$$\sigma^2\left(y_{ijk}\right) = \mu_{ijk} \tag{1.2}$$

Suppose further that the following exponential model is imposed on the means of y_{ijk}

$$\mu_{ijk} = \exp\left(\eta_{ijk}\right) \tag{1.3}$$

The model in (2.3) is transformed to a linear model by using the log link function. In other words

$$\eta_{ijk} = \ln\left(\mu_{ijk}\right) \tag{1.4}$$

2.6.1.4 Models for the Bernoulli sampling distribution

Sampling distribution

$$f(y_{ijk}) = \mu_{ijk}^{y_{ijk}}(1 - \mu_{ijk})^{1 - y_{ijk}} \tag{1.5}$$

Variance

$$\sigma^2(y_{ijk}) = \mu_{ijk}(1 - \mu_{ijk}) \tag{1.6}$$

2.6.1.4.1 The logistic model

Model for means

$$\mu_{ijk} = \frac{1}{1 + \exp\left(-\eta_{ijk}\right)} \tag{1.7}$$

Link function

$$\eta_{ijk} = \text{logit}(\mu_{ijk}) = \ln\left\{\frac{\mu_{ijk}}{1 - \mu_{ijk}}\right\} \tag{1.8}$$

2.6.1.4.2 The complementary log-log model

Model for means

$$\mu_{ijk} = 1 - \exp\left\{-\exp\left(\eta_{ijk}\right)\right\} \tag{1.9}$$

Link function

$$\eta_{ijk} = \ln(-\ln(1 - \mu_{ijk})) \tag{1.10}$$

2.6.1.4.3 The probit model

Model for means

$$\mu_{ijk} = \Phi(\eta_{ijk}) \tag{1.11}$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard Normal distribution.

Link function

$$\eta_{ijk} = \Phi^{-1}(\mu_{ijk}) \tag{1.12}$$

2.6.1.4.4 The log-log model

Model for means

$$\mu_{ijk} = \exp\left\{-\exp\left(-\eta_{ijk}\right)\right\} \tag{1.13}$$

Link function

$$\eta_{ijk} = -\ln\left\{-\ln\left(\mu_{ijk}\right)\right\} \tag{1.14}$$

2.6.1.5 Models for the Binomial distribution

Sampling distribution

Let y_{ijk} denote the proportion of successes in n_{ijk} independent trials:

$$f(y_{ijk}) = \binom{n_{ijk}}{n_{ijk}\,y_{ijk}} \mu_{ijk}^{n_{ijk}\,y_{ijk}} \left(1 - \mu_{ijk}\right)^{n_{ijk}\left(1 - y_{ijk}\right)} \tag{1.15}$$

Variance

$$\sigma^2(y_{ijk}) = \frac{\mu_{ijk}\left(1 - \mu_{ijk}\right)}{n_{ijk}} \tag{1.16}$$

The models for the means and the link functions are identical to those of the Bernoulli-logit model described in Section 2.6.1.4.

2.6.1.6 The Negative Binomial-log model

Sampling distribution

$$f(y_{ijk}) = \frac{\Gamma\left(y_{ijk} + \dfrac{1}{\psi}\right)}{\Gamma\left(y_{ijk} + 1\right)\Gamma\left(\dfrac{1}{\psi}\right)} \frac{\left(\psi\mu_{ijk}\right)^{y_{ijk}}}{\left(1 + \psi\mu_{ijk}\right)^{y_{ijk} + \frac{1}{\psi}}} \tag{1.17}$$

Variance

$$\sigma^2\left(y_{ijk}\right) = \mu_{ijk} + \psi\mu_{ijk}^2 \tag{1.18}$$

The model for means and the link function are identical to those of the Poisson-log model described in Section 2.6.1.3.

2.6.1.7 The Gamma-log model

Sampling distribution

$$f(y_{ijk}) = \frac{1}{\Gamma\left(\dfrac{1}{\psi}\right) y_{ijk}} \left(\frac{y_{ijk}}{\mu_{ijk}\psi}\right)^{\frac{1}{\psi}} \exp\left(-\frac{y_{ijk}}{\mu_{ijk}\psi}\right) \tag{1.19}$$

Variance

$$\sigma^2\left(y_{ijk}\right) = \psi\mu_{ijk}^2 \tag{1.20}$$

The model for means and the link function are identical to those of the Poisson-log model described in Section 2.6.1.3.

2.6.1.8 The Inverse Gaussian-log model

Sampling distribution

$$f(y_{ijk}) = \frac{1}{\sqrt{2\pi y_{ijk}^3 \psi}} \exp\left(-\frac{1}{2y_{ijk}}\left(\frac{y_{ijk} - \mu_{ijk}}{\mu_{ijk}}\right)^2 \Big/ \psi\right) \tag{1.21}$$

Variance

$$\sigma^2\left(y_{ijk}\right) = \psi\mu_{ijk}^3 \tag{1.22}$$

The model for means and the link function are identical to those of the Poisson-log model described in Section 2.6.1.3.

2.6.1.9 Models for the Multinomial sampling distribution

Sampling distribution

$$f\left(y_{ijk,1}, y_{ijk,2}, \ldots, y_{ijk,C-1}\right) = \frac{n_{ijk}!}{\left(\displaystyle\prod_{l=1}^{C-1} y_{ijk,l}!\right)\left(n_{ijk} - \displaystyle\sum_{l=1}^{C-1} y_{ijk,l}\right)!} \left(\prod_{l=1}^{C-1} \mu_{ijk,l}^{y_{ijk,l}}\right) \mu_{ijk,C}^{n_{ijk} - \sum_{l=1}^{C-1} y_{ijk,l}} \tag{1.23}$$

Covariance matrix

$$\Sigma(\mathbf{y}_{ijk}^*) = \mathbf{D}_{\mu_{ijk}} - \mathbf{\mu}_{ijk}\mathbf{\mu}_{ijk}' \tag{1.24}$$

where $\mathbf{y}_{ijk}^* = \begin{bmatrix} y_{ijk,1} & y_{ijk,2} \cdots & y_{ijk,C-1} \end{bmatrix}'$ and $\mathbf{D}_{\mu_{ijk}}$ denotes a $(C-1)\times(C-1)$ diagonal matrix with the elements of $\mathbf{\mu}_{ijk} = \begin{bmatrix} \mu_{ijk,1} & \mu_{ijk,2} \cdots & \mu_{ijk,C-1} \end{bmatrix}'$ on the diagonal.

2.6.1.10 The generalized logistic (nominal) Model

Model for means

$$\mu_{ijk,l} = \frac{\exp\{\eta_{ijk,l}\}}{1+\sum_{l=1}^{C-1}\exp\{\eta_{ijk,l}\}} \quad \forall \ l = 1,2,\ldots,C-1 \tag{1.25}$$

Link function

$$\eta_{ijk,l} = \text{logit}(\mu_{ijk,l}) = \ln\left\{\frac{\mu_{ijk,l}}{\mu_{ijk,C}}\right\} \tag{1.26}$$

2.6.1.11 The cumulative logistic (ordinal) model

Model for means

$$\mu_{ijk,l}^* = \sum_{r=1}^{l}\mu_{ijk,r} = \frac{\exp\{\tau_l - \eta_{ijk,l}\}}{1+\exp\{\tau_l - \eta_{ijk,l}\}} \quad \forall \ l = \ldots,C-1 \tag{1.27}$$

$$= \frac{1}{1+\exp\{-\eta_{ijk}^*\}}$$

where

$$\eta_{ijk,l}^* = \tau_l - \eta_{ijk,l}$$

the elements of $\tau_1, \ \tau_2, \ \cdots ,\tau_{C-1}$ denote threshold parameters.

Link function

$$\eta_{ijk,l}^* = \text{clogit}(\mu_{ijk,l}^*) = \ln\left\{\frac{\mu_{ijk,l}^*}{1-\mu_{ijk,l}^*}\right\} \tag{1.28}$$

2.6.1.12 The proportional hazards (cumulative complimentary log-log) model

Model for means

$$\mu_{ijk,l}^* = \sum_{r=1}^{l} \mu_{ijk,r} = 1 - \exp\left(-\exp\left\{\eta_{ijk,l}^*\right\}\right) \quad \forall \quad l = 1, 2, \ldots, C-1 \quad (1.29)$$

$$\eta_{ijk,l}^* = \text{cloglog}(\mu_{ijk,l}^*) = \ln\left(-\ln\left(1 - \mu_{ijk,l}^*\right)\right) \quad (1.30)$$

2.6.1.13 The cumulative log-log model

Model for means

$$\mu_{ijk,l}^* = \sum_{r=1}^{l} \mu_{ijk,r} = \exp\left(-\exp\left\{-\eta_{ijk,l}^*\right\}\right) \quad \forall \quad l = 1, 2, \cdots, C-1 \quad (1.31)$$

$$\eta_{ijk,l}^* = \text{loglog}(\mu_{ijk,l}^*) = -\ln\left(-\ln\left(\mu_{ijk,l}^*\right)\right) \quad (1.32)$$

2.6.1.14 The cumulative probit model

Model for means

$$\mu_{ijk,l}^* = \sum_{r=1}^{l} \mu_{ijk,r} = \Phi\left(\eta_{ijk,l}^*\right) \quad \forall \quad l = 1, 2, \ldots, C-1 \quad (1.33)$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution.

Link function

$$\eta_{ijk,l}^* = \Phi^{-1}\left(\mu_{ijk,l}^*\right) \quad (1.34)$$

2.6.1.15 The estimation of scale and dispersion parameters

A number of sampling distributions discussed in the previous sections have a dispersion parameter and/or a scale parameter. A summary of these distributions with respect to dispersion and scale parameters and their estimates is shown in Table 16.

Table 16: Scale and dispersion parameters

Distribution	Deviance	Dispersion	Pearson	Scale
Binomial	x		x	x
Gamma	x	x	x	x
Inverse Gaussian	x	x	x	x
Negative binomial	x	x	x	
Poisson	x		x	x

2.6.1.16   The deviance χ^2 estimate

$$\hat{\phi}_D = \sqrt{\frac{\chi_D^2}{d}} \tag{1.35}$$

$$\chi_D^2 = 2\ln L(\mathbf{y}\,|\,\mathbf{y}) - 2\ln L(\hat{\boldsymbol{\mu}}\,|\,\mathbf{y}) \tag{1.36}$$

$$d = \sum_{i=1}^{N}\sum_{j=1}^{n_i}\sum_{k=1}^{n_{ij}} w_{ijk} - q \tag{1.37}$$

2.6.1.17   The Pearson χ^2 estimate

$$\hat{\phi}_P = \sqrt{\frac{\chi_P^2}{d}} \tag{1.38}$$

$$\chi_P^2 = \sum_{i=1}^{N}\sum_{j=1}^{n_i}\sum_{k=1}^{n_{ij}} \frac{w_{ijk}\left(y_{ijk} - \hat{\mu}_{ijk}\right)^2}{\hat{\sigma}^2\left(y_{ijk}\right)} \tag{1.39}$$

2.6.2 Theoretical aspects: level-3 generalized linear models

2.6.2.1 Notation

Let $\mathbf{y}_{ij}$ denote a $n_{ij} \times 1$ vector of outcomes with typical element y_{ijk}, where i denotes the level-3 units, j denotes the level-2 units nested within the i-th level-3 unit and k denotes the level-1 units nested within ij.

Assume further that there are N level-3 units so that $i = 1, 2, ..., N$. Within a typical level-3 unit there are n_i level-2 units, $j = 1, 2, ..., n_i$ and nested within ij there are n_{ij} level-1 units so that $k = 1, 2, ..., n_{ij}$.

There are, therefore, $\sum_{i=1}^{N} n_i$ level-2 units and $\sum_{i=1}^{N}\sum_{j=1}^{n_i} n_{ij}$ level-1 units.

Let $\mathbf{y}_i^*$ and $\mathbf{v}_i^*$ denote $\sum_{j=1}^{n_i} n_{ij} \times 1$ vectors partitioned as follows:

$$\mathbf{y}_i^* = \begin{pmatrix} \mathbf{y}_{i1} \\ \mathbf{y}_{i2} \\ \vdots \\ \mathbf{y}_{in_i} \end{pmatrix}; \ \mathbf{v}_i^* = \begin{pmatrix} \mathbf{v}_{i1} \\ \mathbf{v}_{i2} \\ \ldots \\ \mathbf{v}_{in_i} \end{pmatrix}, \ i = 1, 2, \ldots, N$$

Under the assumption that

$$\mathbf{v}_{i1},\ldots,\mathbf{v}_{in_i} \qquad i.i.d. \qquad N\left(\mathbf{0},\boldsymbol{\Phi}\right) \tag{1.40}$$

and

$$\mathbf{v}_i, \qquad i=1,\ldots,N \qquad i.i.d. \qquad N\left(\mathbf{0},\boldsymbol{\Psi}\right) \tag{1.41}$$

with $\operatorname{cov}\left(\mathbf{v}_{ij}^{'},\mathbf{v}_i\right)=\mathbf{0}$, it follows that

$$
\begin{aligned}
f\left(\mathbf{y}_i^*,\mathbf{v}_i^*,\mathbf{v}_i\right) &= f\left(\mathbf{y}_i^* \mid \mathbf{v}_i^*,\mathbf{v}_i\right)\cdot g\left(\mathbf{v}_i^*,\mathbf{v}_i\right) \\
&= f\left(\mathbf{y}_i^* \mid \mathbf{v}_i^*,\mathbf{v}_i\right)\cdot g\left(\mathbf{v}_i^*\right)g\left(\mathbf{v}_i\right)
\end{aligned}
\tag{1.42}
$$

Therefore

$$f\left(\mathbf{y}_i^*\right)=\int_{\mathbf{v}_i}\left\{\int_{\mathbf{v}_i^*} f\left(\mathbf{y}_i^* \mid \mathbf{v}_i^*,\mathbf{v}_i\right)\cdot g\left(\mathbf{v}_i^*\right)d\mathbf{v}_i^*\right\}g\left(\mathbf{v}_i\right)d\mathbf{v}_i$$

From (1.40), it follows that

$$f\left(\mathbf{y}_i^* \mid \mathbf{v}_i^*,\mathbf{v}_i\right)\cdot g\left(\mathbf{v}_i\right)=\prod_{j=1}^{n_i} f\left(\mathbf{y}_{ij} \mid \mathbf{v}_{ij},\mathbf{v}_i\right)g\left(\mathbf{v}_{ij}\right).$$

Hence

$$
\begin{aligned}
f\left(\mathbf{y}_i^*\right) &= \int_{\mathbf{v}_i}\prod_{j=1}^{n_i}\left\{\int_{\mathbf{v}_{ij}} f\left(\mathbf{y}_i \mid \mathbf{v}_i,\mathbf{v}_i\right)\cdot g\left(\mathbf{v}_{ij}\right)d\mathbf{v}_{ij}\right\}g\left(\mathbf{v}_i\right)d\mathbf{v}_i \\
&= \int_{\mathbf{v}_i}\prod_{j=1}^{n_i}\left\{\int_{\mathbf{v}_{ij}}\left[\prod_{k=1}^{n_{ij}} f\left(y_{ijk} \mid \mathbf{v}_{ij},\mathbf{v}_i\right)\right]\cdot g\left(\mathbf{v}_{ij}\right)d\mathbf{v}_{ij}\right\}g\left(\mathbf{v}_i\right)d\mathbf{v}_i
\end{aligned}
\tag{1.43}
$$

Using the Poisson distribution model as an example,

$$f\left(y_{ijk} \mid \mathbf{v}_{ij},\mathbf{v}_i\right)=\frac{\exp\left(-\lambda_{ijk}\right)\lambda_{ijk}^{y_{ijk}}}{\lambda_{ijk}!},$$

where

$$\lambda_{ijk}=\exp\left\{\mathbf{x}_{ijk}^{'}\boldsymbol{\beta}+\mathbf{z}_{ijk(2)}^{'}\mathbf{v}_{ij}+\mathbf{z}_{ijk(3)}^{'}\mathbf{v}_i\right\}.$$

2.6.2.2 Log-likelihood function

Let

$$l_i = \ln f\left(\mathbf{y}_i^*\right)$$

$$= \ln \int_{\mathbf{v}_i} \prod_{j=1}^{n_i} \left\{ h\left(\mathbf{v}_{ij}\right)\right\} g\left(\mathbf{v}_i\right) d\mathbf{v}_i$$

where

$$h\left(\mathbf{v}_{ij}\right) = \int_{\mathbf{v}_{ij}} \exp \sum_{k=1}^{n_{ij}} \ln f\left(\mathbf{y}_{ijk} \mid \mathbf{v}_{ij}, \mathbf{v}_i\right) g\left(\mathbf{v}_{ij}\right) d\mathbf{v}_i$$

Note that

$$\prod_{j=1}^{n_i} \left[h\left(\mathbf{v}_{ij}\right)\right] g\left(\mathbf{v}_i\right) = \prod_{j=1}^{n_i} \left\{ K_1 \int_{\mathbf{u}_{ij}} \exp\left[\sum_{k=1}^{n_{ij}} l_{ijk} - \frac{1}{2}\mathbf{v}_{ij}'\mathbf{\Phi}^{-1}\mathbf{v}_{ij} \right] d\mathbf{v}_{ij} \right\} K_2 \exp -\frac{1}{2}\mathbf{v}_i'\mathbf{\Psi}^{-1}\mathbf{v}_i$$

Let

$$t_{ij} = \sum_{k=1}^{n_{ij}} l_{ijk} - \frac{1}{2}\mathbf{v}_{ij}'\mathbf{\Phi}^{-1}\mathbf{v}_{ij} + \ln K_u \tag{1.44}$$

$$l_{ijk} = \ln f\left(y_{ijk} \mid \mathbf{v}_{ij}, \mathbf{v}_i\right) \tag{1.45}$$

and

$$q_i = \ln K_v - \frac{1}{2}\mathbf{v}_i'\mathbf{\Psi}^{-1}\mathbf{v}_i, \tag{1.46}$$

$$K_u = \left(2\pi\right)^{-r/2}\left|\mathbf{\Phi}\right|^{-1/2}$$

$$K_v = \left(2\pi\right)^{-m/2}\left|\mathbf{\Psi}\right|^{-1/2}.$$

From (1.44), (1.45) and (1.46) it follows that

$$f\left(\mathbf{y}_i^*\right) = K_2 \int_{\mathbf{v}_i} \left\{ \prod_{j=1}^{n_i} \int_{\mathbf{v}_{ij}} \exp t_{ij} d\mathbf{v}_{ij} \right\} \exp q_i d\mathbf{v}_i$$

$$= K_2 \int_{\mathbf{v}_i} \exp\left\{ \sum_{j=1}^{n_i} \ln \int_{\mathbf{vu}_{ij}} \exp t_{ij} d\mathbf{v}_{ij} \right\} \exp q_i d\mathbf{v}_i$$

$$= K_2 \int_{\mathbf{v}_i} \exp\left(\sum_{j=1}^{n_i} \ln q_{ij}^* + q_i \right) d\mathbf{v}_i$$

with

$$q_{ij}^* = \int_{\mathbf{v}_{ij}} \exp t_{ij} d\mathbf{v}_{ij}.$$

Therefore

$$\ln l_i = \ln K_v + \ln f\left(\mathbf{y}_i^*\right) = \ln K_v + \ln \int_{v_i} \exp\left(q_{ij} + q_i\right) dv_i \qquad (1.47)$$

where

$$q_{ij} = \sum_{j=1}^{n_i} \ln q_{ij}^*$$

2.6.2.3 Empirical Bayes estimates

Estimates of the random effects are obtained as the conditional expectation of u_{ijk} given the observations $\mathbf{y}_i^*$. More specifically,

$$E\left(u_{ijk} \mid \mathbf{y}_i^*\right) = \left[K_v \int \exp\left(q_i + q_{ij}\right) p_{ijk} d\mathbf{v}_i\right] / f\left(\mathbf{y}_i^*\right) \qquad (1.48)$$

where

$$K_v = \left(2\pi\right)^{-\frac{m}{2}} |\mathbf{\Psi}|^{-\frac{1}{2}},$$

$$q_i = -\frac{1}{2} \mathbf{v}_i \mathbf{\Psi}^{-1} \mathbf{v}_i$$

and where

$$p_{ijk} = \int_{\mathbf{v}_{ij}} u_{ijk} f\left(\mathbf{y}_{ij} \mid \mathbf{v}_i, \mathbf{v}_{ij}\right) g\left(\mathbf{v}_{ij}\right) d\mathbf{v}_{ij}$$

Likewise

$$E\left(v_{ijk} \bullet v_{ijl} \mid \mathbf{y}_i^*\right) = \left[K_v \int \exp\left(q_i + q_{ij}\right) c_{ijkl} d\mathbf{v}_i\right] / f\left(\mathbf{y}_i^*\right) \qquad (1.49)$$

where

$$c_{ijkl} = \int_{\mathbf{v}_{ij}} v_{ijk} v_{ijl} f\left(\mathbf{y}_{ij} \mid \mathbf{v}_i, \mathbf{v}_{ij}\right) g\left(\mathbf{v}_{ij}\right) d\mathbf{v}_{ij}.$$

2.6.2.4 Derivatives of the log-likelihood function

Fixed effects: β -derivatives

$$l_i = \ln f\left(\mathbf{y}_i^*\right) = \ln K_v \int \exp\left(q_i + q_{ij}\right) d\mathbf{v}_i$$

Therefore

$$l_i = \ln \int_{\mathbf{v}_i} \exp \left\{ \sum_{j=1}^{n_i} \ln \int_{\mathbf{v}_{ij}} f\left(\mathbf{y}_{ij} \mid \mathbf{v}_i, \mathbf{v}_{ij}\right) g\left(\mathbf{v}_{ij}\right) \right\} g\left(\mathbf{v}_i\right) d\mathbf{v}_i$$

$$
\begin{aligned}
\frac{\partial \ln l_i}{\partial \beta_r} &= \frac{1}{f\left(\mathbf{y}_i^*\right)} K_v \int_{\mathbf{v}_i} \frac{\partial}{\partial \beta_r} \exp\left(q_{ij} + q_i\right) dv_i \\
&= \frac{1}{f\left(\mathbf{y}_i^*\right)} K_v \int_{\mathbf{v}_i} \left[\frac{\partial}{\partial \beta_r} q_{ij} \right] \exp\left(q_{ij} + q_i\right) dv_i
\end{aligned}
\tag{1.50}
$$

Since

$$q_{ij} = \sum_{j=1}^{n_i} \ln q_{ij}^*,$$

it follows that

$$\frac{\partial q_{ij}}{\partial \beta_r} = \sum_{j=1}^{n_i} \frac{\dfrac{\partial}{\partial \beta_r} \ln q_{ij}^*}{q_{ij}^*}$$

and

$$
\begin{aligned}
\frac{\partial}{\partial \beta_r} q_{ij}^* &= \int_{\mathbf{v}_{ij}} \frac{\partial}{\partial \beta_r} \exp \sum_{k=1}^{n_{ij}} \ln f\left(\mathbf{y}_{ijk} \mid \mathbf{v}_i, \mathbf{v}_{ij}\right) g\left(\mathbf{v}_{ij}\right) d\mathbf{v}_{ij} \\
&= \int_{\mathbf{v}_{ij}} \sum_{k=1}^{n_{ij}} \frac{\partial}{\partial \beta_r} \ln f\left(\mathbf{y}_{ijk} \mid \mathbf{v}_i, \mathbf{v}_{ij}\right) f\left(\mathbf{y}_{ij} \mid \mathbf{v}_i, \mathbf{v}_{ij}\right) g\left(\mathbf{v}_{ij}\right) d\mathbf{v}_{ij}.
\end{aligned}
\tag{1.51}
$$

Level-2 variance components: Φ - derivatives

$$\frac{\partial \ln l_i}{\partial \Phi_{rs}} = \frac{1}{f\left(\mathbf{y}_i^*\right)} K_v \int_{\mathbf{v}} \left[\frac{\partial}{\partial \phi_{rs}} q_{ij} \right] \exp\left(q_{ij} + q_i\right) d\mathbf{v}_i \tag{1.52}$$

$$q_{ij} = \sum_{j=1}^{n_i} \ln q_{ij}^*$$

$$
\begin{aligned}
\frac{\partial q_{ij}}{\partial \Phi_{rs}} &= \sum_{j=1}^{n_i} \frac{1}{q_{ij}^*} \int_{\mathbf{v}_{ij}} \left[\frac{\partial}{\partial \phi_{rs}} \ln g\left(\mathbf{v}_{ij}\right) \right] f\left(\mathbf{y}_{ij} \mid \mathbf{v}_i, \mathbf{v}_{ij}\right) g\left(\mathbf{v}_{ij}\right) d\mathbf{v}_{ij} \\
&= \sum_{j=1}^{n_i} \frac{E_{\mathbf{v}_{ij}/\mathbf{y}_{ij}} \left[\dfrac{\partial}{\partial \phi_{rs}} \ln g\left(\mathbf{v}_{ij}\right) \right]}{q_{ij}^*} \\
&= \sum_{j=1}^{n_i} \frac{E_{\mathbf{v}_{ij}/\mathbf{y}_i} \left[\dfrac{2 - \delta rs}{2} \left(\mathbf{P}_{ij}\right)_{r,s} \right]}{q_{ij}^*}
\end{aligned}
$$

where

$$\mathbf{P}_{ij} = \mathbf{\Phi}^{-1}\left(\mathbf{v}_{ij}\mathbf{v}_{ij}^{'} - \mathbf{\Phi}\right)\mathbf{\Phi}^{-1}.$$

Level-3 variance components: Ψ - derivatives

$$\frac{\partial \ln l_i}{\partial \psi_{rs}} = \frac{1}{f\left(\mathbf{y}_i^*\right)} \int_{\mathbf{v}_i} \left[\frac{\partial}{\partial \psi_{rs}}\left(q_{ij}\right)\right] \exp\left(q_{ij} + q_i\right) dv_i \qquad (1.53)$$

$$q_i = -\frac{1}{2}\mathbf{v}_i\Psi^{-1}\mathbf{v}_i + \ln K_v$$

$$= \ln g\left(\mathbf{v}_i\right)$$

$$\Rightarrow \frac{\partial q_i}{\partial \psi_{rs}} = \left[\frac{2-\delta rs}{2}\left(\mathbf{P}_i\right)_{r,s}\right] \qquad (1.54)$$

where

$$\mathbf{P}_i = \Psi^{-1}\left(\mathbf{v}_i\mathbf{v}_i^{'} - \Psi\right)\Psi^{-1}.$$

Second order derivatives

The method for obtaining second order partial derivatives is illustrated below for the terms $\dfrac{\partial^2 \ln l_i}{\partial \psi_{uv}\partial \psi_{rs}}$

and $\dfrac{\partial^2 \ln l_i}{\partial \phi_{uv}\partial \phi_{rs}}$. The derivatives for $\dfrac{\partial^2 \ln l_i}{\partial \beta_u\partial \psi_{rs}}$ etc. are obtained in a similar way.

$\mathrm{I}\left(\psi_{uv},\psi_{rs}\right)$:

From (1.53)

$$\frac{\partial \ln l_i}{\partial \psi_{rs}} = \frac{\partial \ln f\left(\mathbf{y}_i^*\right)}{\partial \psi_{rs}}$$

$$= \frac{1}{f\left(\mathbf{y}_i^*\right)} \int_{\mathbf{v}_i} \left[\frac{\partial}{\partial \psi_{rs}} \ln g\left(\mathbf{v}_i\right)\right] \exp\left(q_{ij} + q_i\right) d\mathbf{v}_i$$

Hence

$$\frac{\partial^2 \ln l_i}{\partial \psi_{uv} \partial \psi_{rs}} = \frac{1}{f\left(\mathbf{y}_i^*\right)} \int_{\mathbf{v}_i} \left[\frac{\partial^2}{\partial \psi_{uv} \psi_{rs}} \ln g\left(\mathbf{v}_i\right) \right] \exp\left(q_i + q_{ij}\right) d\mathbf{v}_i$$

$$+ \frac{1}{f\left(\mathbf{y}_i^*\right)} \int_{\mathbf{v}_i} \left[\frac{\partial}{\partial \psi_{rs}} \ln g\left(\mathbf{v}_i\right) \frac{\partial}{\partial \psi_{uv}} \ln g\left(\mathbf{v}_i\right) \right] \exp\left(q_i + q_{ij}\right) d\mathbf{v}_i$$

$$- \frac{1}{f\left(\mathbf{y}_i^*\right)^2} \left\{ \int_{\mathbf{v}_i} \left[\frac{\partial}{\partial \psi_{rs}} \ln g\left(\mathbf{v}_i\right) \right] \exp\left(q_i + q_{ij}\right) d\mathbf{v}_i + \int_{\mathbf{v}_i} \left[\frac{\partial}{\partial \psi_{uv}} \ln g\left(\mathbf{v}_i\right) \right] \exp\left(q_i + q_{ij}\right) d\mathbf{v}_i \right\}$$

$$= \frac{1}{f\left(\mathbf{y}_i^*\right)} \int_{\mathbf{v}_i} \left\{ \frac{\partial^2}{\partial \psi_{uv} \psi_{rs}} \ln g\left(\mathbf{v}_i\right) + \frac{\partial}{\partial \psi_{rs}} \ln g\left(\mathbf{v}_i\right) \frac{\partial}{\partial \psi_{uv}} \ln g\left(\mathbf{v}_i\right) \right\} \times \exp\left(q_i + q_{ij}\right) d\mathbf{v}_i$$

$$- \frac{\partial \ln l_i}{\partial \psi_{uv}} \frac{\partial \ln l_i}{\partial \psi_{rs}}$$

$$= \frac{cons}{f\left(\mathbf{y}_i^*\right)} \int_{\mathbf{v}_i} [P_i]_{u,v} [P_i]_{r,s} \times \exp\left(q_i + q_{ij}\right) d\mathbf{v}_i - \frac{\partial \ln l_i}{\partial \psi_{uv}} \frac{\partial \ln l_i}{\partial \psi_{rs}} - cons\psi_{ur}^{-1}\psi_{vs}^{-1} + \psi_{us}^{-1}\psi_{vr}^{-1}$$

$$\tag{1.55}$$

$$cons = \frac{(2 - \delta uv)(2 - \delta rs)}{2}$$

From (1.55) it follows that

$$\frac{\partial^2 \ln f_i}{\partial \psi_{uv} \psi_{rs}} = cons\left\{ E_{|y_i} [\mathbf{P}_i]_{uv} E_{|y_i} [\mathbf{P}_i]_{rs} - E_{|y_i} [\mathbf{P}_i]_{uv} \cdot [\mathbf{P}_i]_{rs} + \mathbf{\psi}_{ur}^{-1}\mathbf{\psi}_{vs}^{-1} + \mathbf{\psi}_{us}^{-1}\mathbf{\psi}_{vr}^{-1} \right\}. \tag{1.56}$$

$\mathbf{I}\left(\mathbf{\Phi}_{uv}, \mathbf{\Phi}_{rs}\right)$:

From (1.52) we have

$$\frac{\partial \ln l_i}{\partial \phi_{rs}} = \frac{1}{f\left(\mathbf{y}_i^*\right)} \int_{\mathbf{v}_i} \left[\frac{\partial}{\partial \phi_{rs}} \ln g\left(\mathbf{v}_i\right) \right] \exp\left(q_{ij} + q_i\right) d\mathbf{v}_i$$

and

$$q_{ij} = \sum_{k=1}^{n_{ij}} \ln \int_{\mathbf{v}_{ij}} f\left(\mathbf{y}_{ij} \mid \mathbf{v}_i, \mathbf{v}_{ij}\right) g\left(\mathbf{v}_{ij}\right) d\mathbf{v}_{ij}$$
.

Thus

$$\frac{\partial}{\partial \phi_{rs}} q_{ij} = \sum_{j=1}^{n_i} \frac{2 - \delta rs}{2} E\_PIJ = EPIJ$$
,

where

145

$$E\_PIJ = \frac{1}{q_{ij}^*} \int \left(P_{ij}\right)_{r,s} \left[\frac{2-\delta rs}{2}\right] f\left(\mathbf{y}_{ij} \mid \mathbf{v}_i, \mathbf{v}_{ij}\right) g\left(\mathbf{v}_{ij}\right) d\mathbf{v}_{ij},$$

and

$$\frac{\partial \ln l_i}{\partial \Phi_{rs}} = \frac{1}{f\left(\mathbf{y}_i^*\right)} \int_{\mathbf{v}_i} EPIJ \exp\left(q_{ij}+q_i\right) d\mathbf{v}_i \qquad (1.57)$$

$$EPIJ = \sum_{j=1}^{n_i} E\_PIJ,$$

and therefore

$$\frac{\partial^2 \ln l_i}{\partial \phi_{uv} \partial \phi_{rs}} = \frac{1}{f\left(\mathbf{y}_i^*\right)} \int_{\mathbf{v}_i} \left[\left[\frac{\partial^2}{\partial \phi_{uv}\phi_{rs}} q_{ij} + \left\{\frac{\partial}{\partial \phi_{uv}} q_{ij} \times \frac{\partial}{\partial \phi_{rs}} q_{ij}\right\}\right] \exp\left(q_i+q_{ij}\right) d\mathbf{v}_i - \frac{\partial \ln l_i}{\partial \Phi_{uv}} \cdot \frac{\partial \ln l_i}{\partial \Phi_{rs}}.$$

Hence

$$\frac{\partial^2 \ln f_i}{\partial \Phi_{uv} \Phi_{rs}} = \frac{-cons}{f\left(\mathbf{y}_i^*\right)} \int_{\mathbf{v}_i} \left\{\sum_{j=1}^{n_i} E_{|y_{ij}} \left[\mathbf{P}_{ij}\right]_{uv} \left[\mathbf{P}_{ij}\right]_{rs}\right\} \exp\left(q_i+q_{ij}\right) d\mathbf{v}_i$$

$$+ \frac{\partial \ln f_i}{\partial \Phi_{uv}} \cdot \frac{\partial \ln f_i}{\partial \Phi_{rs}} + cons \times \left(\sum_{j=1}^{N} n_i\right) \left[\mathbf{\Phi}_{ur}^{-1}\mathbf{\Phi}_{vs}^{-1} + \mathbf{\Phi}_{us}^{-1}\mathbf{\Phi}_{vr}^{-1}\right]$$

2.6.2.5 Evaluation of integrals

In the preceding sections expressions for the log-likelihood function and derivatives are given in terms of multiple integrals. In general, no closed form solution to these multiple integrals exists and therefore use is made of numerical integration to evaluate them.

Consider a general integral of the form

$$I = \int f\left(\mathbf{y}_i \mid \mathbf{v}_i\right) g\left(\mathbf{v}_i\right) d\mathbf{v}_i,$$

where it is assumed that $\mathbf{v}_i \sim N\left(\mathbf{0}, \mathbf{\Phi}\right)$. This integral can equivalently be written as follows:

$$I = \int \phi\left(\mathbf{v}_i \mid \mathbf{y}_i\right) \left\{\frac{f\left(\mathbf{y}_i \mid \mathbf{v}_i\right) g\left(\mathbf{v}_i\right) d\mathbf{v}_i}{\phi\left(\mathbf{v}_i \mid \mathbf{y}_i\right)}\right\} d\mathbf{v}_i,$$

where

$$\phi\left(\mathbf{v}_i \mid \mathbf{y}_i\right) = k \exp\left(-\frac{1}{2}\left(\mathbf{v}_i - \hat{\mathbf{v}}_i\right) \mathbf{\Sigma}_i^{-1}\left(\mathbf{v}_i - \hat{\mathbf{v}}_i\right)\right), \qquad (1.58)$$

$$\hat{\mathbf{v}}_i = E\left(\mathbf{v}_i \mid \mathbf{y}_i\right),$$

$$\hat{\boldsymbol{\Sigma}}_i = Cov\left(\mathbf{v}_i \mid \mathbf{y}_i\right),$$

and

$$k = \left(2\pi\right)^{-r/2} \mid \hat{\boldsymbol{\Sigma}}_i \mid^{-1/2} . \tag{1.59}$$

Consider the transformation of variables

$$\mathbf{z}_i = \frac{1}{\sqrt{2}} \mathbf{T}_i^{-1}\left(\mathbf{v}_i - \hat{\mathbf{v}}_i\right) \tag{1.60}$$

where

$$\mathbf{T}_i \mathbf{T}_i' = \hat{\boldsymbol{\Sigma}}_i$$

and hence

$$\hat{\boldsymbol{\Sigma}}_i = \left(\mathbf{T}_i'\right)^{-1}\left(\mathbf{T}_i\right)^{-1}.$$

From (1.60) it follows that

$$\mathbf{v}_i = \mathbf{v}_i\left(\mathbf{z}_i\right) = \sqrt{2}\mathbf{T}_i \mathbf{z}_i + \hat{\mathbf{v}}_i,$$

The Jacobian of the transformation is given by

$$d\mathbf{v}_i = \mid \mathbf{T}_i^* \mid d\mathbf{z}_i,$$

$$\mathbf{T}_i^* = \sqrt{2}\mathbf{T}_i .$$

Using the change in variables, it follows that

$$I = \int k \exp{-\mathbf{z}_i \mathbf{z}_i'}\left(\frac{f\left(\mathbf{y}_i \mid \mathbf{v}_i\right) g\left(\mathbf{v}_i\right)}{k \exp{-\mathbf{z}_i \mathbf{z}_i'}}\right) \mid \mathbf{T}_i^* \mid d\mathbf{z}_i. \tag{1.61}$$

2.6.2.6 Adaptive quadrature

To evaluate (1.61), use is made of a direct implementation of Gauss-Hermite quadrature. With this rule

$$\int\limits_{-\infty}^{\infty} \exp\{-z^2\} f(z) dz$$

can be approximated by

$$\sum_{\alpha=1}^{G} w_\alpha f(z_\alpha)$$

,

where the w_α and z_α denote weights and nodes of the Hermite polynomial of degree G.

Applying this to the multiple integral defined by (1.61), it follows that

$$I \simeq C\sum_{g_1=1}^{G}...\sum_{g_r=1}^{G} w_{g_1}...w_{g_r} \exp \mathbf{z}_g \mathbf{z}_g^{'} \mid \mathbf{T}_i^* \mid f\left(\mathbf{y}_i \mid \mathbf{v}_i(\mathbf{z}_g)\right) g\left(\mathbf{v}_i(\mathbf{z}_g)\right)$$
$$= C\sum_{g_1=1}^{G}...\sum_{g_r=1}^{G} m_{g_1}...m_{g_r} \mid \mathbf{T}_i^* \mid f\left(\mathbf{y}_i \mid \mathbf{v}_i(\mathbf{z}_g)\right) g\left(\mathbf{v}_i(\mathbf{z}_g)\right)$$

$$I \simeq C \sum_{g_1=1}^{G} \cdots \sum_{g_r=1}^{G} w_{g_1} \cdots w_{g_r} \exp \mathbf{z}_g \mathbf{z}'_g \left|\mathbf{T}_i^*\right| f\left(\mathbf{y}_i \mid \mathbf{v}_i(\mathbf{z}_g)\right) g\left(\mathbf{v}_i(\mathbf{z}_g)\right)$$

$$= C \sum_{g_1=1}^{G} \cdots \sum_{g_r=1}^{G} m_{g_1} \cdots m_{g_r} \left|\mathbf{T}_i^*\right| f\left(\mathbf{y}_i \mid \mathbf{v}_i(\mathbf{z}_g)\right) g\left(\mathbf{v}_i(\mathbf{z}_g)\right)$$

where

$$C = \left(2\pi\right)^{-r/2} \mid \boldsymbol{\Phi} \mid^{-1/2},$$

$$m_{g_\alpha} = \exp\left(z_{g_\alpha} \cdot \ln w_{g_\alpha}\right),$$

and

$$z_g = \left(z_{g1}, z_{g2}, ..., z_{gr}\right).$$

Values of $\hat{\mathbf{v}}_i$ and $\hat{\boldsymbol{\Sigma}}_i$ (cf. Section 2.3.3) are iteratively updated. This implies that the location and scale of the area under the integral changes over iterations and depends on the observed values for a particular level-3 or level-2 unit.

2.6.3 Starting values for generalized linear models

2.6.3.1 Introduction

LISREL uses an algorithm based on the maximization of the posterior distribution (MAP) with respect to the random effects.

In the sections to follow, we assume a level-2 model with a count outcome variable. It is also assumed that the Poisson model is appropriate for level-2 data with a subset of the regression coefficients assumed to be random.

2.6.3.2 Illustration of the procedure for a count outcome variable

Let y_{ij} be a count outcome variable where i denotes level-2 units, $i = 1, 2, \ldots, N$ and j level-1 units nested within the level-2 units $j = 1, 2, \ldots, n_i$.

Under the assumption of conditional independence

$$f\left(\mathbf{y}_i / \mathbf{v}_i\right) = \prod_{j=2}^{n_i} exp\left(-\mu_{ij}\right) \mu_{ij}^{y_{ij}} \left(y_{ij}!\right)^{-1} \qquad (1.62)$$

Suppose that the following exponential model is imposed on the means of the elements y_{ij} of the $n_i \times 1$ vector $\mathbf{y}_i$

$$\mu_{ij} = exp\left(\eta_{ij}\right) = exp\left(\mathbf{x}_{ij}'\boldsymbol{\beta} + \mathbf{z}_{ij}'\mathbf{v}_i\right), \qquad (1.63)$$

where $\mathbf{x}_{ij}$ is a $p \times 1$ vector of covariates and the elements of $\boldsymbol{\beta} = \left[\beta_1, \beta_2, \ldots, \beta_p\right]'$ denote unknown, but fixed, parameters. Generally, the $m \times 1$ vector $\mathbf{z}_{ij}$ is a subset of the columns of $\mathbf{x}_{ij}$. Additionally, it is assumed that $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N$ are i.i.d. $N\left(\mathbf{0}, \boldsymbol{\Phi}\right)$.

The model (1.63) is transformed to a linear model by using the log link function. In other words,

$$\eta_{ij} = ln\left(\mu_{ij}\right). \qquad (1.634)$$

Using standard results for conditional distributions, it follows that

$$f\left(\mathbf{v}_i / \mathbf{y}_i\right) = f\left(\mathbf{v}_i, \mathbf{y}_i\right) / f\left(\mathbf{y}_i\right)$$
$$= f\left(\mathbf{y}_i / \mathbf{v}_i\right) \cdot g\left(\mathbf{v}_i\right) / f\left(\mathbf{y}_i\right).$$

Hence

$$ln\, f\left(\mathbf{v}_i / \mathbf{y}_i\right) = ln\, f\left(\mathbf{y}_i / \mathbf{v}_i\right) + ln\, g\left(\mathbf{v}_i\right) - K$$

$$= \sum_{j=1}^{n_i}\left\{-\mu_{ij} + y_{ij}\, ln\, \mu_{ij} - ln\left(y_{ij}!\right)\right\} + g\left(\mathbf{v}_i\right) + K, \tag{1.65}$$

2.6.3.3 Gradient vector and Hessian matrix

Given β and $\mathbf{\Phi}$, it follows that

$$\frac{\partial\, ln\, f\left(\mathbf{v}_i / \mathbf{y}_i\right)}{\partial[\mathbf{v}_i]_r} = \sum_{j=1}^{n_i}\left\{y_{ij}\frac{\partial}{\partial v_{ir}}\, ln\, \mu_{ij} - \frac{\partial}{\partial v_{ir}}\mu_{ij}\right\} + \frac{\partial}{\partial v_{ir}}\, ln\, g\left(\mathbf{v}_i\right), \tag{1.66}$$

$$r = 1, 2, \ldots, m$$

Since

$$\frac{\partial\mu_{ij}}{\partial v_{ir}} = \mu_{ij}z_{ijr}, \tag{1.67}$$

$$g\left(\mathbf{v}_i\right) = \left(2\pi\right)^{-\frac{r}{2}}\left|\mathbf{\Phi}\right|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\mathbf{v}_i'\mathbf{\Phi}^{-1}\mathbf{v}_i\right\}$$

and hence

$$\frac{\partial\, ln\, g\left(\mathbf{v}_i\right)}{\partial[\mathbf{v}_i]_r} = -\left[\mathbf{\Phi}^{-1}\mathbf{v}_i\right]_r, \tag{1.68}$$

it follows that

$$\frac{\partial\, ln\, f\left(\mathbf{v}_i \mid \mathbf{y}_i\right)}{\partial v_{ir}} = \sum_{j=1}^{n_i}z_{ijr}\left\{y_{ij} - \mu_{ij}\right\} - \left[\mathbf{\Phi}^{-1}\mathbf{v}_i\right]_r, \qquad r = 1, 2, \ldots, \ldots, m.$$

Maximization of $\ln f\left(\mathbf{v}_i \mid \mathbf{y}_i\right)$ is equivalent to the minimization of

$$F = \underbrace{ln\, f\left(\mathbf{v}_i / \mathbf{y}_i\right)}_{F2} - \underbrace{ln\, g\left(\mathbf{v}_i\right)}_{F1}. \tag{1.69}$$

Hence the gradient vector is defined by

$$\frac{\partial F_i}{\partial v_{ir}} = -\frac{\partial\, ln\, f\left(\mathbf{v}_i \mid \mathbf{y}_i\right)}{\partial v_{ir}}. \tag{1.70}$$

Furthermore,

$$\frac{\partial^2 F_i}{\partial v_{ir}\partial v_{is}} = \left[\mathbf{\Phi}^{-1}\right]_{r,s} + \sum_{j=1}^{n_i}\mu_{ij}z_{ir}z_{ij}s, \qquad r, s = 1, 2, \ldots, \ldots, m. \tag{1.71}$$

Let $\mathbf{H}$ denote the Hessian matrix, where

$$\left[\mathbf{H}_i\right]_{r,s} = \frac{\partial^2 F}{\partial v_{ir} \partial v_{is}}, \tag{1.72}$$

then

$$E\left[\mathbf{H}_i\right]_{r,s} = \mathbf{\Phi}_{r,s}^{-1} + \sum_{j=1}^{n_i} z_{ijr} z_{ijs} E\left(\mu_{ij}\right), \tag{1.73}$$

where

$$E\left(\mu_{ij}\right) = E\left(exp\left\{\mathbf{x}_{ij}^{'}\boldsymbol{\beta} + \mathbf{z}_{ij}^{'}\mathbf{v}_{ij}\right\}\right). \tag{1.74}$$

Therefore

$$E\left[\mathbf{H}_i\right]_{r,s} = \mathbf{\Phi}_{r,s}^{-1} + \sum_{j=1}^{n_i}\left[exp\left\{\mathbf{x}_{ij}^{'}\boldsymbol{\beta} + \frac{1}{2}\mathbf{z}_{ij}^{'}\mathbf{\Phi}\mathbf{z}_{ij}\right\}\right] z_{ijr} z_{ijs}. \tag{1.75}$$

2.6.3.4 The MAP algorithm

Set $\hat{\mathbf{\Phi}} = 0.1 * \mathbf{I}$, $\hat{\mathbf{v}}_1, \hat{\mathbf{v}}_2, \ldots, \hat{\mathbf{v}}_N = \mathbf{0}$

Calculate $\hat{\boldsymbol{\beta}}$ given $\hat{\mathbf{\Phi}}$ and $\hat{\mathbf{v}}_i$

Given the current estimates $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ and $\hat{\mathbf{\Phi}}$ of $\mathbf{\Phi}$, calculate $\hat{\mathbf{v}}_i$, $i = 1, 2, \ldots, N$ using the Newton-Raphson method:

$$\hat{\mathbf{v}}_i^{(k)} = \hat{\mathbf{v}}_i^{(k-1)} + \mathbf{H}_i^{-1(k)}\mathbf{g}_i^{(k)}, \qquad k = 1, 2, \ldots \tag{1.76}$$

where

$$\left[\mathbf{g}_i\right]_r = \frac{\partial F_i}{\partial v_{ir}}, \qquad r = 1, 2, \ldots, m. \tag{1.77}$$

Obtain (see, *e.g.*, du Toit, 1993) a revised estimate $\hat{\mathbf{\Phi}}$ of $\mathbf{\Phi}$ from

$$\hat{\mathbf{\Phi}} = \frac{1}{N}\sum\left\{Cov\left(\hat{\mathbf{v}}_i \mid \mathbf{y}_i\right) + \left(\hat{\mathbf{v}}_i - \hat{\bar{\mathbf{v}}}\right)\left(\hat{\mathbf{v}}_i - \hat{\bar{\mathbf{v}}}\right)^{'}\right\} \tag{1.78}$$

where

$$Cov\left(\mathbf{v}_i \mid \mathbf{y}_i\right) = \left[E\left(\mathbf{H}_i\right)\right]^{-1} \tag{1.79}$$

and

$$\hat{\bar{\mathbf{v}}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} v_{ij}. \tag{1.80}$$

Repeat steps (2) to (4) until convergence is attained.

2.6.3.5 Starting values for adaptive quadrature

As initial estimates of the fixed and random parameters, we use the values of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\Phi}}$ at convergence. The $\hat{\mathbf{v}}_i$ and $Cov(\hat{\mathbf{v}}_i | \mathbf{y}_i)$ (see (1.79) and (1.80)) are used as initial estimates of the empirical Bayes means and the covariances in the adaptive quadrature procedure described in Section 2.6.2.

2.7 References

Agresti, A. (2002). Categorical Data Analysis, Second Edition. New York: Wiley.

Binder, D.A. (1983). On the Variances of Asymptotically Normal Estimators from Complex Surveys. International Statistical Review, **51**, 279-292.

Bishop, Y.M.M., Feinberg, S.E. & Holland, P.W. (1988). Discrete Multivariate Analysis: Theory and Practice. Cambridge: MIT Press.

McCullagh, P. & Nelder, J.A. (1989). Generalized Linear Models. Chapman & Hall: London.

Morel, G. (1989). Logistic Regression under Complex Survey Designs. Survey Methodology, **15**, 203-223.

SAS Institute, Inc. (2004). SAS/STAT®: User's Guide. Cary, NC: SAS Institute, Inc.