

MIXNO: a computer program for mixed-effects nominal logistic regression

Donald Hedeker

Division of Epidemiology and Biostatistics

& Health Policy and Research Centers

School of Public Health

University of Illinois at Chicago

Correspondence to Donald Hedeker, Division of Epidemiology & Biostatistics (M/C 922), School of Public Health, University of Illinois at Chicago, 2121 West Taylor Street, Room 525, Chicago, IL, 60612-7260. Phone: (312) 996-4896.

E-MAIL: hedeker@uic.edu

URL: <http://www.uic.edu/~hedeker/mix.html>

Abstract

MIXNO provides maximum marginal likelihood estimates for mixed-effects nominal logistic regression analysis. These models can be used for analysis of correlated nominal response data, for example, data arising from a clustered or longitudinal design. For such data, the mixed-effects model assumes that data within clusters or subjects are dependent. The degree of dependency is jointly estimated with the usual model parameters, thus adjusting for dependence resulting from nesting of the data. MIXNO uses marginal maximum likelihood estimation, utilizing a Fisher-scoring solution. For the scoring solution, the Cholesky factor of the random-effects variance-covariance matrix is estimated along with the effects of model covariates. Examples illustrating usage and features of MIXNO are provided.

Keywords: nominal responses; polytomous responses; multinomial regression; heterogeneity; clustering; multilevel data; random effects; correlated responses; categorical data

1 Introduction

Nominal or polytomous response data are common in many fields of research. For example, the variable “type of service use” is a nominal response variable which is often used in services research. Types of health services utilization can include medical provider visit, hospital outpatient visit, emergency room visit, hospital inpatient stay, and home health care visit. When observations are independent (*i.e.*, not longitudinal or clustered), a statistical model which is appropriate for assessing the influence of explanatory variables on a nominal response variable is the multinomial logistic regression model, or multinomial logit model ([1], [2], [3], [4], [5]). This model is sometimes called a polychotomous logistic regression model. It is often the case, however, that subjects are observed nested within clusters (*i.e.*, schools, firms, clinics), or are repeatedly measured. In this case, use of the ordinary multinomial logistic regression model assuming independence of observations is problematic since observations from the same cluster or subject are usually correlated.

For data that are clustered and/or longitudinal, mixed-effects regression models are becoming increasingly popular, and several books have recently been written on this topic [6, 7, 8, 9]. Common to both clustered and longitudinal data is the idea of nesting. In clustered data, subjects are clustered or nested within a larger context, for example, a hospital, school, clinic, or firm. In longitudinal data where individuals are repeatedly assessed, measures are clustered or nested within individuals. In order to take the nesting of data into account, models with random effects are typically employed. For clustered data the random effects represent cluster effects, while for longitudinal data the random effects represent subject effects.

Though much of the work on mixed-effects models has been for continuous responses, an increasing amount of work has emerged on mixed-effects models for non-continuous response data. In this regard, mixed-effect models for both dichotomous [10, 11, 12, 13, 14, 15] and ordinal [16, 17, 18, 19, 20]. responses have been described. This paper describes the FORTRAN program MIXNO (mixed-effects nominal logistic regression) for the analysis of repeated or clustered nominal responses. MIXNO can accommodate multiple random effects, and allows for a general form for model covariates. Assuming a logistic response function, a maximum marginal likelihood solution is implemented using multi-dimensional quadrature to numerically integrate over the distribution of random-effects. A Fisher scoring solution provides relatively quick convergence and standard errors for the model parameters. Examples of analysis of both clustered and longitudinal

data will illustrate features of MIXNO for nominal response data.

Very little commercially-based software exists to perform mixed-effects regression analysis for nominal responses. The MLn [21] software program, which uses an approximate Taylor-series expansion to linearize the non-linear model, can be used to estimate models for multilevel nominal response data. However, biased results have been reported using the approximate Taylor-series methods in certain situations for binary data [22, 23]. Instead, MIXNO, by using numerical integration, implements a full-information maximum likelihood approach that does not suffer from these reported biases.

The organization of this manual is as follows. Section 2 describes the computational and statistical features of the model that is implemented in MIXNO. Although MIXNO has a WINDOWS interface that greatly enhances its ease of use, it is difficult to describe MIXNO use via the interface in a document like this. As a result, Section 3 describes the procedure for running MIXNO in DOS, and as such, serves as a guide to the necessary MIXNO syntax for DOS use, and a reference (to help understand what the menus and boxes of the interface refer to) for WINDOWS use. Section 4 presents an overview of the three examples in this manual. The first example is described in Section 5. This example focuses on an analysis of a clustered dataset where students are observed nested within classrooms. Section 6 describes MIXNO analysis of a longitudinal dataset where psychiatric homeless individuals were followed across time and assessed regarding their housing status. The last example, in Section 7, shows how MIXNO can be used to fit various psychometric latent trait models. This example uses survey data where multiple items are nested within subjects. For all examples, listings of program syntax and abbreviated output are provided. Additionally, listings of SAS IML code is given for the last two examples to illustrate how MIXNO estimates can be used to examine model fit. Finally, Sections 8, 9, and 10 briefly list hardware and software specifications, availability, and acknowledgements.

2 Computational Methods

The statistical development of the model is described in [24]. Here, an overview of the computational features will be provided. Using the terminology of multilevel analysis [6] let i denote the level-2 units (clusters) and let j denote the level-1 units (nested observations). Assume that there are $i = 1, \dots, N$ level-2 units and $j = 1, \dots, n_i$ level-1 units nested within each level-2 unit. Let y_{ij} be the value of the nominal variable associated with level-2 unit i and level-1 unit j . In the nominal case, we need to consider the values corresponding to the unordered multiple categories of the response variable. For this, let us assume that the $K + 1$ response categories are coded as $0, 1, 2, \dots, K$.

Adding random effects to the multinomial logistic regression model of Bock [2], Nerlove and Press [3], and others, we get that the probability, for a given level-2 unit i , that $y_{ij} = k$ (a response occurs in category k), conditional on $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, is:

$$P_{ijk} = P(y_{ij} = k \mid \boldsymbol{\beta}, \boldsymbol{\alpha}) = \frac{\exp(z_{ijk})}{1 + \sum_{h=1}^K \exp(z_{ijh})} \quad \text{for } k = 1, 2, \dots, K \quad (1)$$

$$P_{ij0} = P(y_{ij} = 0 \mid \boldsymbol{\beta}, \boldsymbol{\alpha}) = \frac{1}{1 + \sum_{h=1}^K \exp(z_{ijh})} \quad (2)$$

where $z_{ijk} = \mathbf{x}'_{ij}\boldsymbol{\beta}_{ik} + \mathbf{w}'_{ij}\boldsymbol{\alpha}_k$. Here, \mathbf{w}_{ij} is the $p \times 1$ covariate vector and \mathbf{x}_{ij} is the design vector for the r random effects, both vectors being for the j th level-1 unit nested within level-2 unit i . Correspondingly, $\boldsymbol{\alpha}_k$ is a $p \times 1$ vector of unknown fixed regression parameters, and $\boldsymbol{\beta}_{ik}$ is a $r \times 1$ vector of unknown random effects for the level-2 unit i . The distribution of the random effects is assumed to be multivariate normal with mean vector $\boldsymbol{\mu}_k$ and covariance matrix $\boldsymbol{\Sigma}_k$. Notice, that the regression coefficient vectors $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ carry the k subscript. Thus, for each of the p covariates and r random effects, there will be K parameters to be estimated. Additionally, the random effect variance-covariance matrix $\boldsymbol{\Sigma}_k$ is allowed to vary with k .

It is convenient to standardize the random effects by letting $\boldsymbol{\beta}_{ik} = \mathbf{T}_k\boldsymbol{\theta}_i + \boldsymbol{\mu}_k$, where $\mathbf{T}_k\mathbf{T}'_k = \boldsymbol{\Sigma}_k$ is the Cholesky decomposition of $\boldsymbol{\Sigma}_k$. The model is now given as

$$z_{ijk} = \mathbf{x}'_{ij}(\mathbf{T}_k\boldsymbol{\theta}_i + \boldsymbol{\mu}_k) + \mathbf{w}'_{ij}\boldsymbol{\alpha}_k. \quad (3)$$

In this form, it is clear to see that this generalizes Bock's model for educational test data [25] by including covariates \mathbf{w}_{ij} , and by allowing a general random-effects design vector \mathbf{x}_{ij} including the

possibility of multiple random effects θ_i .

2.1 Parameter Estimation

Let \mathbf{y}_i denote the vector of nominal responses from level-2 unit i (for the n_i level-1 units nested within). Then the probability of any \mathbf{y}_i , conditional on the random effects θ and given α_k, μ_k , and \mathbf{T}_k , is equal to the product of the probabilities of the level-1 responses:

$$\ell(\mathbf{y}_i | \theta; \alpha_k, \mu_k, \mathbf{T}_k) = \prod_{j=1}^{n_i} \prod_{k=0}^K [P(y_{ij} = k | \theta; \alpha_k, \mu_k, \mathbf{T}_k)]^{d_{ijk}} \quad (4)$$

where $d_{ijk} = 1$ if $y_{ij} = k$, and 0 otherwise. Thus, associated with the response from a particular level-1 unit, $d_{ijk} = 1$ for only one of the $K + 1$ categories and zero for all others. The marginal density of the response vector \mathbf{y}_i in the population is expressed as the following integral of the likelihood, $\ell(\cdot)$, weighted by the prior density $g(\cdot)$:

$$h(\mathbf{y}_i) = \int_{\theta} \ell(\mathbf{y}_i | \theta; \alpha_k, \mu_k, \mathbf{T}_k) g(\theta) d\theta \quad (5)$$

where $g(\theta)$ represents the population distribution of the random effects.

For parameter estimation, the marginal log-likelihood from the N level-2 units can be written as: $\log L = \sum_i^N \log h(\mathbf{y}_i)$. Then, using η_k to represent an arbitrary parameter vector,

$$\frac{\partial \log L}{\partial \eta_k} = \sum_{i=1}^N h^{-1}(\mathbf{y}_i) \int_{\theta} \left[\sum_{j=1}^{n_i} (d_{ijk} - P_{ijk}) \frac{\partial z_{ijk}}{\partial \eta_k} \right] \ell(\mathbf{y}_i | \theta; \alpha_k, \mu_k, \mathbf{T}_k) g(\theta) d\theta \quad (6)$$

where

$$\frac{\partial z_{ijk}}{\partial \alpha_k} = \mathbf{w}_{ij}, \quad \frac{\partial z_{ijk}}{\partial \mu_k} = \mathbf{x}_{ij}, \quad \frac{\partial z_{ijk}}{\partial \mathbf{v}(\mathbf{T}_k)} = \mathbf{J}_r(\theta \quad \mathbf{x}_{ij}), \quad (7)$$

\mathbf{J}_r is a transformation matrix eliminating elements above the main diagonal (see [26]), and $\mathbf{v}(\mathbf{T}_k)$ is the vector containing the unique elements of the Cholesky factor \mathbf{T}_k . If \mathbf{T}_k is a $r \times 1$ vector of independent random effect variance terms, then $\partial z_{ijk} / \partial \mathbf{T}_k = \mathbf{x}_{ij} \theta$ in the equation above.

Fisher's method of scoring can be used to provide the solution to these likelihood equations. For this, provisional estimates for the vector of parameters Θ , on iteration ι are improved by

$$\Theta_{\iota+1} = \Theta_{\iota} - \mathcal{E} \left[\frac{\partial^2 \log L}{\partial \Theta_{\iota} \partial \Theta'_{\iota}} \right]^{-1} \frac{\partial \log L}{\partial \Theta_{\iota}} \quad (8)$$

where the empirical information matrix is given by:

$$\mathcal{E} \left[\frac{\partial^2 \log L}{\partial \boldsymbol{\Theta}_l \partial \boldsymbol{\Theta}_l'} \right] = - \sum_{i=1}^N h^{-2}(\mathbf{y}_i) \frac{\partial h(\mathbf{y}_i)}{\partial \boldsymbol{\Theta}_l} \left(\frac{\partial h(\mathbf{y}_i)}{\partial \boldsymbol{\Theta}_l} \right)'. \quad (9)$$

In general, the total number of parameters equals the $K \times p$ fixed regression coefficients ($\boldsymbol{\alpha}_k$; $k = 1, \dots, K$), plus the $K \times r$ means of the random effects ($\boldsymbol{\mu}_k$; $k = 1, \dots, K$), and the $K \times r \times (r-1)/2$ random effect variance-covariance terms ($\mathbf{v}[\mathbf{T}_k]$; $k = 1, \dots, K$). In certain cases, restrictions on the parameters (*e.g.*, options `DIAG`, `NOMU`, `VGRP`, `VCAT` presented later) can be made to reduce the total number of estimated parameters. Notice that the parameter vector $\mathbf{v}(\mathbf{T}_k)$, which indicates the degree of level-2 population variance, is what distinguishes the mixed-effects model from the ordinary fixed-effects multinomial logistic regression model.

At convergence, the MML estimates and their accompanying standard errors can be used to construct asymptotic z -statistics by dividing the parameter estimate by its standard error [27]. The computed z -statistic can then be compared with the standard normal table to test whether the parameter is significantly different from zero. While this use of the standard errors to perform hypothesis tests (and construct confidence intervals) for the fixed effects $\boldsymbol{\mu}_k$ and $\boldsymbol{\alpha}_k$ is generally reasonable, for the variance and covariance components $\mathbf{v}(\mathbf{T}_k)$ this practice is problematic (see Bryk and Raudenbush [7], page 55).

2.2 Numerical Quadrature

In order to solve the above likelihood equations, numerical integration on the transformed $\boldsymbol{\theta}$ space can be performed. If the assumed random-effect distribution is normal, Gauss-Hermite quadrature can be used to approximate the above integrals to any practical degree of accuracy. In Gauss-Hermite quadrature, the integration is approximated by a summation on a specified number of quadrature points Q for each dimension of the integration; thus, for the transformed $\boldsymbol{\theta}$ space, the summation goes over Q^r points. For the standard normal univariate density, optimal points and weights (denoted B_q and $A(B_q)$, respectively) are given in Stroud and Secrest [28]. For the multivariate density, the r -dimensional vector of quadrature points is denoted by $\mathbf{B}_q' = (B_{q1}, B_{q2}, \dots, B_{qr})$, with its associated (scalar) weight given by the product of the corresponding univariate weights,

$$A(\mathbf{B}_q) = \prod_{h=1}^r A(B_{qh}) . \quad (10)$$

If another distribution is assumed, other points may be chosen and density weights substituted for $A(B_q)$ or $A(B_{qh})$ above (note, the weights must be normalized to sum to unity). For example, if a rectangular or uniform distribution is assumed, then Q points may be set at equal intervals over an appropriate range (for each dimension) and the quadrature weights are then set equal to $1/Q$. Other distributions are possible; Bock and Aitkin [29] discuss the possibility of empirically estimating the random-effect distribution. In MIXNO, users can select either a normal or uniform distribution for the random effects via the `PRIOR` option. Since the latter distribution represents vague information about the shape of the random-effects distribution, it can be used to get some idea about the sensitivity of the results to the assumed normal distribution.

For models with few random effects the quadrature solution is relatively fast and computationally tractable. In particular, if there is only one random effect in the model, there is only one additional summation over Q points relative to the fixed effects solution. As the number of random effects r is increased, the terms in the summation (Q^r) increases exponentially in the quadrature solution. Fortunately, as is noted by Bock, Gibbons and Muraki [30] in the context of a dichotomous factor analysis model, the number of points in each dimension can be reduced as the dimensionality is increased without impairing the accuracy of the approximations; they indicated that for a five-dimensional solution as few as three points per dimension were sufficient to obtain adequate accuracy. In general, specifying between 10 to 20 quadrature points for a unidimensional solution and 7 to 10 points for a two-dimensional solution is usually reasonable.

2.3 Solution incorporating Level-2 Weights

The above solution can be modified to accommodate weighted data, which occurs when the same response pattern \mathbf{y}_i and covariate vector \mathbf{w}_i is observed for a number of level-2 units. As an example of weighted data, consider the frequencies listed in Table 2.1 that are published in Ezzet and Whitehead [18]. This table lists the frequencies corresponding to categorical responses at two timepoints for two groups of subjects: group 1 and (group 2). Each cell of this table represents the frequency of individuals (level-2 units) with the same response pattern \mathbf{y}_i (the categorical responses

at two timepoints) and covariate vector \mathbf{w}_i (*e.g.*, group).

Pre-Test Rating	Rating at Post-Test			
	1	2	3	4
1	59 (63)	35 (13)	3 (0)	2 (0)
2	11 (40)	27 (15)	2 (0)	1 (0)
3	0 (7)	0 (2)	0 (1)	0 (0)
4	1 (2)	1 (0)	0 (1)	0 (0)

As such, these frequencies represent weights which can be incorporated into a modeling of the categorical response across the two timepoints. To select a weighted solution in MIXNO, the WT option is used. MIXNO usage for frequency weighted data will be illustrated in section 7.

2.4 Varying Random-effect Variance Terms

MIXNO can also accommodate separate random-effect variance terms for groups of either i or j units. For example, suppose that there is interest in allowing varying random-effect variance terms by gender. For this, \mathbf{x}_{ij} is specified as a 2×1 vector of dummy codes indicating male and female membership, respectively. \mathbf{T}_k is then a 2×1 vector of independent random-effect standard deviations for males and females, and the subject effect θ_i is a scalar that is pre-multiplied by the vector \mathbf{T}_k . Section 6.3 illustrates how to accomplish this using the VGRP option in MIXNO.

This option can also be used to estimate psychometric latent trait models [25] where n item responses ($j = 1, 2, \dots, n$) are nested within N subjects ($i = 1, 2, \dots, N$). Here, a separate random-effect standard deviation (*i.e.*, an element of the $n \times 1$ vector \mathbf{T}_k) can be estimated for each test item (*i.e.*, each j unit). Again, this is accomplished by specifying \mathbf{x}_{ij} as a $n \times 1$ vector of dummy codes indicating the repeated items. This use of the VGRP option is presented in Section 7. Notice that in terms of the statistical model, for both cases, \mathbf{T}_k is a $r \times 1$ vector that is pre-multiplied by the transpose of a $r \times 1$ vector of indicator variables \mathbf{x}_{ij} , and so \mathbf{T}_k pre-multiplies a scalar random effect θ_i (instead of a $r \times 1$ vector of random effects $\boldsymbol{\theta}_i$).

3 Program description and usage

MIXNO is currently available in executable form for MS-DOS, WINDOWS, and MACINTOSH computers. For WINDOWS, an interface has been developed that greatly aids in model specification and selection of program options. To use the interface, the MIXNO.EXE file is invoked (issuing the command MIXNO in DOS, or double-clicking on the MIXNO icon in WINDOWS). The MIXNO instructions are then created interactively using the interface and are saved (by the interface) in a file named MIXNO.DEF that is read by the FORTRAN executable (MIXNOB.EXE) at program run-time. To run the program without the interface, the user must create the MIXNO.DEF file using a text editor and then invoke MIXNOB.EXE. Because it is difficult to describe use of MIXNO via the interface in document form, here we will detail the procedure for running the program without the interface. The main difference in using the interface is that the format of the MIXNO.DEF file described in section 3.2 is not important, because the interface creates the MIXNO.DEF file for the user. Thus, for interface users, it is only the content and not the syntax of the MIXNO.DEF file that needs to be considered.

MIXNO makes use of the following files:

- input data file
- MIXNO.DEF - main definition file for analysis options and settings
- main output file

In addition to the main output file, MIXNO produces the following additional output files:

- MIXNO.EST - a file containing the estimated parameters (with labels).
- MIXNO.VAR - a file containing the large-sample variance covariance matrix of the parameter estimates (the inverse of the information matrix). The full rectangular matrix is printed out, row by row, with the order of the parameters identical to that of MIXNO.EST (*i.e.*, no labels are given in MIXNO.VAR).
- MIXNO.RES - a file containing empirical Bayes estimates of the random effect for each level-2 unit. This file lists for each level-2 unit: level-2 ID, the number of level-1 units n_i , the empirical Bayes estimate (posterior mean or mean vector), and the posterior standard

deviation (or posterior variance covariance matrix if there are more than one random effect). Additionally, if the level-2 units have a frequency weight (*i.e.*, if the **WT** option is selected), then this weight is also output to this file immediately following the level-2 ID (and before n_i).

3.1 Structure of the input data file

This file contains all data (*i.e.*, level-2 identifier, nominal response variable, and covariates) to be read in by the program. It is read in free format and must be a standard text (ASCII) file with no hidden characters or word processing format codes. Variable fields must be separated by one or more blanks. The data are assumed to consist of multiple level 1 observations within a higher-order (2nd level) unit. **There must be a level-2 ID variable for each record and the data must be sorted by this level-2 ID variable.** The nested measurements (level 1) of a cluster (level 2) take up as many records in this file as there are level 1 units within that cluster. Thus, some clusters can have, for example, 40 records while others may have 20 to 50 records.

The fields of variables that are read in, separated by one or more blanks, on a line (or lines) are as follows (the order of the variables does not matter):

$$ID \quad NomVar \quad Xvector \quad Wvector$$

where, *ID* refers to the level-2 ID number which does not change across level-1 units, *NomVar* is the value of nominal response for the observation, *Xvector* is the part of the design matrix for the random effects, and *Wvector* is the covariate vector for the observation,. All variables are read as REAL*8 with the exception of the level-2 IDs which are read as INTEGER. **All missing data must have a numeric missing value code.** Thus, missing values cannot be left as blank fields or designated by periods.

3.2 Analysis options and settings - MIXNO.DEF

This file contains the information to determine which statistical model is fit to the data in the input data file. Although a word processor can be used to create this file, it must be saved as a standard text (ASCII) file with no hidden characters or word processing format codes. The analysis options and settings that comprise this file are described in Tables 3.1, 3.2, 3.3, and 3.4 below.

Except where noted, this file is read in free format. This file is created by the user *before* typing the command MIXNOB (*i.e.*, before running the FORTRAN executable). Again, for WINDOWS, the interface can be used (by running the MIXNO.EXE file) to aid in creation of the MIXNO.DEF file and to automatically invoke the FORTRAN executable. This filename and extension (MIXNO.DEF) *must* be used and should be in the same directory as the program (MIXNO.EXE and MIXNOB.EXE) or accessible via appropriate PATH statements.

Line 1	A title of 60 characters.
Line 2	A subtitle of 60 characters.
Line 3	Name of input data file. Any legal filename of 80 characters or less can be specified.
Line 4	Name of main output file. Any legal filename of 80 characters or less can be specified.
Line 5	Name of definition file to be saved or retrieved. Any legal filename of 80 characters or less can be specified. Note that a name for this file <i>must</i> be specified even in batch processing, although in batch processing nothing is done to this file.

Line 6	NPR NF R P CONV MAXK1 MISS START WT CATYX PRIOR UNIF NQUAD LINFN DIAG NOMU VGRP VCAT
NPR	Number of level-2 units whose data will be listed on the screen (<i>usually set to 1</i>).
NF	Number of fields of data to read from the input data file.
R	Number of random effects.
P	Number of fixed effects (not including the mean of the random effects).
CONV	Convergence criterion (<i>usually set to .001 or .0001</i>).
MAXK1	Number of nominal response categories (= K+1).
MISS	Set to 0 if no missing values are present in the data, or 1 if missing values are present (codes will be defined later).
START	Set to 0 if automatic starting values are to be used, or 1 if user-defined starting values are to be used.
WT	Set to 0 if each 2nd level unit is weighted equally, or 1 for differential weighting.
CATYX	Set to 1 if a crosstabulation of the nominal response variable by another variable (defined later) should be produced, and 0 otherwise.

Table 3.2 (continued)	
PRIOR	Set to 0 or 1, respectively, for a specified form (see UNIF) or empirically-determined distributional form for the random-effects distribution. (<i>This is an option that is under development; at present IPRIOR=0 is the only possible choice</i>).
UNIF	Set to 0 or 1 for a normal or uniform distribution, respectively, for the assumed random-effects distribution (<i>ordinarily set to 0</i>).
NQUAD	Number of quadrature points (per random-effect dimension) to use in the numerical integration (<i>usually set between 10 and 20 for models with one random effect, and between 5 and 10 for models with multiple random effects</i>).
LINFN	Number of linear transforms of the estimated parameters to estimate (<i>ordinarily set to 0</i>).
DIAG	Set to 0 for correlated random effects or 1 for independent random effects (<i>ordinarily set to 0</i>).
NOMU	Set to 0 to estimate the mean of the random effects or 1 to fix it to zero (<i>ordinarily set to 0</i>).
VGRP	Set to 0 (no) or 1 (yes) for random-effects grouping variables. Specify yes only if R > 1 and the R random-effect variables are dummy-coded level-1 or level-2 grouping variables; otherwise specify no (<i>ordinarily set to 0</i>). If yes is specified, then R random-effect variance terms are estimated: one for each of the (level-1 or level-2) groups determined by the dummy-codes.
VCAT	Set to 0 if the random-effects variance terms are homogeneous across the K contrasts of the nominal response, or 1 if K separate random-effects variance terms are to be estimated (<i>i.e.</i> , the random-effects variance terms vary across the K contrasts).

Table 3.3 - Analysis options and settings specified in MIXNO.DEF: lines 7-9	
Line 7	Two parameters are to be read on this line: the field of the input data file which contains the (level-2) IDs, followed by the field of the input data file which contains the nominal response variable.
Line 8	R parameters are to be read on this line: the field(s) of the input data file which contain(s) the R random effects.
Line 9	P parameters are to be read on this line: the field(s) of the input data file which contain(s) the P fixed effects.

Table 3.4 - Analysis options and settings specified in MIXNO.DEF: lines after line 9	
next line	If WT=1: the field of the input data file which contains the weight to be assigned to each level-2 unit.
next line	The MAXK1 values of the nominal response variable.
next line	If CATYX=1: two parameters and a list of values: the field of the input data file which contains the variable that is to be crosstabulated with the nominal response variable, followed by (a) the number of categories for this variable, and (b) a list of all category values of this variable.
next line	If MISS=1: missing value code for the nominal response variable.
next line	If MISS=1: R missing value codes for the random-effect variables.
next line	If MISS = 1: P missing value codes for the fixed effects.
next line	An 8 character label for the nominal response variable.
next line	R labels for the random effects in 8 character width fields.
next line	If START=1 and NOMU=0: $R \times K$ starting values for the means of the random effects.
next line	P labels for the covariates in 8 character width fields (a maximum of 10 labels per line).
next line	If START=1: $P \times K$ starting values for the covariate effects.
next line	If START=1: $(R \times (R+1)) / 2$ starting values for the variance and covariance terms of the random effects given in “packed” form, <i>e.g.</i> , for a 2 x 2 covariance matrix, the order of the starting values should be: variance(1), covariance(1,2) and variance(2). If VCAT=1 then $K \times R \times (R+1) / 2$ starting values are specified. Note: if either DIAG=1 or VGRP=1 then only R starting values are needed (or $R \times K$ if VCAT=1 is also specified).

Table 3.4 (continued)	
final lines	<p>If LINFN > 0: LINFN by NPAR coefficients for the linear re-expressions of the estimated parameters. The NPAR terms in order are</p> <ol style="list-style-type: none"> 1) $R \times K$ random-effect mean vector (unless NOMU=1). 2) $P \times K$ effects of explanatory variables. 3) Random-effect variance covariance matrix: <ul style="list-style-type: none"> • if VCAT=0 and (DIAG=0 and VGRP=0): $R \times (R+1)/2$ unique elements of the random-effect variance-covariance matrix (in packed form). • if VCAT=1 and (DIAG=0 and VGRP=0): $K \times (R \times (R+1)/2)$ unique elements of the random-effect variance-covariance matrix (in packed form). • if VCAT=0 and (DIAG=1 or VGRP=1): R variance-covariance terms. • if VCAT=1 and (DIAG=1 or VGRP=1): $R \times K$ variance-covariance terms.
	<p>Each of these LINFN sets of coefficients are multiplied by the “original” parameter estimates according to the order given. Standard errors for these LINFN transforms are also printed out. For an example of LINFN usage, see Section 6.4.</p>

3.3 Main output file

This file contains descriptive information about the variables read in to MIXNO, as well as the main results of the specified analysis. The examples of the output file provided below illustrate the contents of this file. After listing out the titles, the selected form of the random-effects distribution is listed. The numbers of observations are then summarized by a listing of the number of level-2 units, the total number of level-1 units, and the number of level-1 units for each level-2 unit. For each variable read in to the program (except the ID variable) descriptive statistics (minimum, maximum, mean, and standard deviation) are then provided. These descriptive statistics are based on the total number of level-1 observations. For the nominal response variable *NomVar*, a frequency count is provided which lists for each category the number (and proportion) of level-1 observations. An optional listing of the frequencies and proportions of *NomVar* by the levels of another variable may be obtained if CATYX = 1. Starting values, either user-defined or program-generated, are listed

for all model parameters. Finally, MIXNO indicates the number (and percentage) of level-2 units with non varying level-1 responses on *NomVar*.

In terms of program results, the number of iterations required to achieve convergence is listed, followed by the number of quadrature points requested, and the value of the log-likelihood at convergence. Since the log-likelihood value multiplied by -2 (*i.e.*, the deviance) can be used to perform likelihood-ratio tests [31], the deviance value is also listed. Following the deviance value is a listing of the ridge value. The ridge is an incremental adjustment which is made to the diagonal elements of the information matrix if the program encounters a non-increasing likelihood or some other indication of numerical difficulty during the iterations. This adjustment often improves the chances of convergence. At present, the ridge starts at zero, and is increased by 0.1 each time that difficulties are encountered. At convergence, the ridge is set back to zero in order to obtain the correct standard errors for the model parameters, however the listing of the ridge value indicates its value prior to being reset to zero. As such, the listed ridge value is indicative of the degree of computational difficulty that the program encountered.

For each parameter of the model, maximum marginal likelihood estimates, standard errors, z-values, and p-values are then provided. These p-values are two-tailed, except for the variance parameters where one-tailed p-values are given. This use of the standard errors to perform hypothesis tests for the variance parameters is controversial (see Bryk and Raudenbush [7], page 55). Also, it is important to realize that it is the Cholesky factor of the random-effects variance-covariance matrix that is estimated, and not the variance-covariance matrix itself. If only one random effect is requested in the model, the Cholesky factor is simply the square root of the variance, that is, the standard deviation. Analogously, with multiple random effects, the Cholesky factor represents the matrix square root.

Following the parameter estimates (and associated statistics), MIXNO lists a correlation matrix associated with the estimates of all model parameters. This correlation matrix does not contain correlations of the variables themselves, but correlations of the estimated model parameters. This matrix may be helpful in determining the degree to which collinearity is present in terms of the model parameters. Finally, if $LINFN > 0$ is specified, linear transforms or re-expressions of the estimated parameters are listed along with their standard errors.

3.4 Some Common MIXNO Errors

There are a few errors which can prevent MIXNO from running correctly, or even running at all. First, as mentioned, missing values that are not given a specified numeric missing value code, but instead left as blank fields, may cause the program to fail or to estimate a model which is incorrect from the user's perspective. To see if this has occurred, the user can check the correctness of each variable's descriptive statistics (minimum, maximum, mean, and standard deviation) listed in the output file. If these descriptive statistics are incorrect, the data are not being read into the program correctly and a common reason is that missing values are being left as blank fields in the data file.

Second, use of the `CATYX` option (described in Tables 3.2 and 3.4) is somewhat tricky. The values listed by the user for the levels of the crosstabulation variable *must* be exactly the same as the values that are found in the data file. If a strange error prevents MIXNO from running and this option is selected, the user can set `CATYX=0` to avoid this option.

Third, the `NPR` option (described in Table 3.2), which is used to list data to the screen, can cause MIXNO to stop in certain cases. This can happen when the number of digits to be listed for a variable exceeds the format specification of the program. If the program stops after indicating (on the screen) the number of random and fixed effects in the model, but prior to listing any iterative results to the screen, the user can set `NPR=0` and re-run the program.

Fourth, problems can develop if the user tries to fit a model with a single random effect, and that random effect is not the intercept. In this case, the procedure used to generate starting values for the program is poor. Instead, the user can choose the `START` option (described in Table 3.2) and specify "naive" starting values of 0 for the mean of the random effect and for the covariate effects, and some fraction of the assumed residual variance for the random-effect variance term (*e.g.*, .5 or 1).

Finally, if the program "blows up," it may be that the specified model is not estimable. In this case, the user should try fitting a less complicated model by specifying fewer random effects, or fewer covariates, or collapsing some of the response categories if these are sparse. If the number of random effects is 1, and problems still exist, it may be that the random-effect variance cannot be reliably estimated as being different from zero. In this case, a model without random effects may be warranted.

4 Examples of MIXNO usage

MIXNO can estimate a variety of models for correlated nominal response data. Here, we will use three different datasets to illustrate some of the different models MIXNO can fit. For each example, listings of MIXNO.DEF files and selected output files will be provided. For space, the listings of the output files in this manual do not contain the correlation matrix of the parameter estimates that the program provides; otherwise the output listings are complete. The datasets and accompanying MIXNO.DEF files used in this manual can be downloaded at the author's website¹.

An analysis of a clustered dataset where students are clustered within classrooms is presented first to illustrate features of mixed-effects regression analysis of clustered data. A comparison to an analysis using a fixed-effects nominal logistic regression model, which ignores the clustering of the data, will illustrate the importance of taking the clustering of the data into account. In the mixed regression model for clustered data, one random term is included in order to account for the clustering of students within classrooms. This random classroom term describes the way in which students from the same classroom respond similarly, relative to the sample as a whole. Additionally, this example will include model covariates at both the level of the classroom (level 2) and the level of the student (level 1).

To illustrate usefulness of mixed-effects regression analysis for longitudinal nominal data, an analysis of a psychiatric dataset where homeless subjects are measured in terms of their housing status across multiple timepoints will be presented. An analysis utilizing a model with one random subject effect to account for the repeated measurements made on each subject will be given first. For this model, we will specify the random effects distribution first as a normal and then as a uniform distribution. In this way, some idea will be obtained about the sensitivity of the results to the distribution specification of the random effects. Also, allowing the random-effects variance terms to vary by group will be shown. Finally, a trend analysis allowing both a random intercept and a linear trend across time will be presented to illustrate how MIXNO can be used to estimate models with multiple random effects.

The last example will illustrate how MIXNO can be used to estimate an item-response theory (IRT) or latent trait model for nominal responses [25]. The data for this example are taken from Clogg [32] and concern three questions about degree of satisfaction with family, hobbies, and

¹<http://www.uic.edu/~hedeker/mix.html>

residence. 1472 subjects from the 1975 US General Household Survey indicated their satisfaction to each of these three questions on a three-point scale denoting low, medium, or high satisfaction. Although these data could be analyzed using methods for ordinal responses, here we will apply the nominal model to these data. To estimate the item parameters, we will make use of the `VGRP` option that allows the random effect variance to vary across levels of a grouping variable (*i.e.*, the items).

These three examples will highlight some of the results that are obtained from mixed-effects analysis, and will be accompanied by listings of specific file setups that are used to run MIXNO. All of the examples presented here are termed two-level models in the multilevel literature [6]. For the first example, students (level-1) are treated as being nested within classrooms (level-2), while for the second and third examples, repeated observations (level-1) are nested within subjects (level-2). At present, MIXNO does not allow a three-level analysis which would consider, concurrently, students nested within classrooms *and* classrooms nested within schools (or repeated observations nested within subjects who are nested within clusters).

5 Analysis of a Clustered Dataset

Hedeker, Gibbons, and Flay [33] illustrated use of mixed-effects regression for clustered data applied to a dataset where students are clustered within classrooms and schools. In that article, the dependent variable was treated continuously and the mixed-effects approach was compared with both individual-level analysis which ignores the clustering of data, and classroom-level analysis which aggregates individual data. In describing the MIXOR software [34] the same dataset was analyzed treating the response as an ordinal response. Here, we will use the same data considering the outcome variable as a nominal response.

The data for this example is from the Television School and Family Smoking Prevention and Cessation Project (TVSFP) [35] which examined a school-based social-resistance curriculum and a television-based program on tobacco use prevention and cessation. For this illustration, a subset of the TVSFP data was used: students from 28 Los Angeles schools, where the schools were randomized to one of four study conditions: (a) a social-resistance classroom curriculum, (b) a media (television) intervention, (c) a social-resistance classroom curriculum combined with a mass-media intervention, and (d) a no-treatment control group. These conditions form a 2 x 2 design of social-resistance classroom curriculum (CC = yes or no) by mass-media intervention (TV = yes or no). The 1600 students in this subset are from 135 classrooms and 28 schools; there is a range of 1 to 13 classrooms per school, and 2 to 28 students per classroom.

A tobacco and health knowledge scale (THKS) score was one of the primary study response variables and the one used here. The scale consisted of seven items used to assess student tobacco and health knowledge. The frequency distribution of post-intervention THKS total scores indicated four ordinal classifications corresponding to 0-1, 2, 3, and 4-7 correct responses.

A partial list of these data is given in Table 5.1. The variables are, in order, school ID, class ID, post-intervention THKS (with 1=0-1, 2=2, 3=3, and 4=4-7), dichotomous post-intervention THKS (with 0=0-2, and 1=3-7), a column of ones for the intercept, pre-intervention THKS (from 0 to 7), CC (with yes=1 and no=0), TV (with yes=1 and no=0), and the product of CC and TV.

403	403101	3	1	1	2	1	0	0
403	403101	4	1	1	4	1	0	0
403	403101	3	1	1	4	1	0	0
403	403101	4	1	1	3	1	0	0
403	403101	4	1	1	3	1	0	0
403	403101	3	1	1	4	1	0	0
403	403101	2	0	1	2	1	0	0
403	403101	4	1	1	4	1	0	0
403	403101	4	1	1	5	1	0	0
403	403101	4	1	1	3	1	0	0
403	403101	3	1	1	3	1	0	0
403	403101	4	1	1	3	1	0	0
403	403101	3	1	1	1	1	0	0
403	403101	4	1	1	2	1	0	0
403	403101	2	0	1	2	1	0	0
403	403101	4	1	1	1	1	0	0
403	403101	4	1	1	4	1	0	0
403	403101	3	1	1	3	1	0	0
403	403101	3	1	1	0	1	0	0
403	403101	4	1	1	3	1	0	0
403	403102	2	0	1	0	1	0	0
403	403102	4	1	1	1	1	0	0
403	403102	3	1	1	5	1	0	0

5.1 Fixed-effects Regression Ignoring Data Clustering

Before proceeding with the mixed-effects analysis of these data, we will present a fixed-effects analysis which ignores clustering of students. Using MIXNO for this type of analysis is equivalent to performing a multinomial logistic regression analysis treating all observations as independent observations. As noted by Hedeker, Gibbons, and Flay [33] and others, ignoring data clustering often results in statistical tests which are too liberal for level-2 covariates, resulting in falsely rejecting the null hypothesis too often. For this first analysis, the post-intervention THKS score is modeled in terms of baseline THKS score and effects of CC, TV, and CC by TV interaction. Table 5.2 lists the MIXNO.DEF file for this model. Note that zero random effects are specified on line 6 of the DEF file, and that blank lines are present for the records that define fields and labels of random effects. In terms of options, a crosstabulation of post-intervention THKS by baseline levels is being requested. The results from this analysis are listed in Table 5.3 (in Section 5.3).

As seen in the output in Table 5.3, when zero random effects are requested, MIXNO indicates the number of observations as number of level-1 observations. For these data, this number is simply the number of students. Descriptive statistics are listed for all variables. Response frequencies and a crosstabulation table are then listed. Since this outcome variable has 4 categories, 3 sets of

Table 5.2 - MIXNO.DEF file for student-level analysis ignoring clustering									
TVSFP study - Post-Test Tobacco and Health Knowledge Scale									
Students nested within classrooms - Independence model									
TVSFPORS.DAT									
TVSFPNO1.OUT									
TVSFPNO1.DEF									
1	9	0	5	0.00010	4	0	0	0	1 0 0 10 0 0 0 0 0
2	3								
5	6	7	8	9					
1	2	3	4						
6	7	0	1	2	3	4	5	6	
THKScore									
Intrcpt	PreTHKS	CC	TV	CC*TV					

regression coefficients are estimated. These coefficients are obtained contrasting responses in the first listed category value with each remaining category (*i.e.*, 2 vs 1, 3 vs 1, and 4 vs 1). Note that MIXNO uses the first category value listed in the MIXNO.DEF file as the reference category. In Table 5.2, this list is on the 10th line and specifies the nominal outcome category value of 1 as the reference category. To have a different category as the reference category, the list of category values (on line 10) can simply be reordered.

This analysis, which ignores the data clustering, indicates increasingly significant results across the response code comparisons. For example, contrasting responses in the two extreme categories, 4 vs. 1, yields significant or marginally significant results for of all model covariates: PreTHKS, CC, TV, and CC by TV.

5.2 Mixed-effects Regression Including Data Clustering

Two mixed-effects regression models can be considered for these data: students within schools, and students within classrooms. As mentioned, MIXNO does not allow a three-level analysis which would consider students nested within both classrooms *and* schools concurrently. To perform the students-within-classrooms analysis the class ID (the second variable field in the datafile) is indicated as the cluster ID on line 7 of the DEF file, while to perform the students-within-schools analysis the school ID (the first variable field) would be indicated as the cluster ID. Table 5.4 lists the MIXNO.DEF file for the students-within-classrooms analysis. Again, the THKS score is modeled in terms of baseline THKS score and effects of CC, TV, and CC by TV interaction.

However, in contrast to the analysis of the previous section, a random classroom effect is included to account for the data clustering.

Table 5.4 - MIXNO.DEF file for first students-within-classrooms analysis																
TVSFP study - Post-Test Tobacco and Health Knowledge Scale																
Students nested w/in CLASSROOMS - Random Intercept																
TVSFPORS.DAT																
TVSFPNO2.OUT																
TVSFPNO2.DEF																
1	9	1	4	0.00010	4	0	0	0	0	0	0	10	0	0	0	0
2 3																
5																
6 7 8 9																
1 2 3 4																
THKScore																
Intrcpt																
PreTHKS	CC		TV		CC*TV											

Comparing Tables 5.4 and 5.2 illustrates the necessary specifications for a random-intercepts model. Specifically, Table 5.4 indicates a random intercept that varies by classrooms, and so, each classroom has its own intercept. Since these classroom intercepts are considered random effects, they are representative of a larger population of classroom effects. MIXNO estimates the variance of this population distribution, and additionally provides empirical Bayes estimates of the classroom effects. The empirical Bayes estimates are provided in the file MIXNO.RES (not shown).

Table 5.5 lists the MIXNO results from the students-within classroom analysis. The output indicates that there are 135 classrooms with 1600 students nested within. The classroom size varies between 1 to 27 students per classroom. Following the listing of the starting values, MIXNO indicates that 6 of the 135 classrooms had response vectors that were non-varying. Thus, students within each of six schools gave identical responses.

The random-effect standard deviation is estimated as .511 and a Wald test rejects the hypothesis that this parameter equals 0. As mentioned earlier, use of the Wald test for testing whether variance parameters equal zero has been questioned [7]. Expressing the estimate of the random-effect standard deviation in terms of an estimated intraclass correlation (in this case, the *intra*classroom correlation) yields .074. Regarding the regression coefficients, in contrast to the analysis ignoring the clustering of the students (in Table 5.3), mixed-effects regression analysis indicates that neither the TV effect nor the interaction of CC by TV are statistically significant at the $p < .10$ level for the comparison of responses 4 vs 1. Thus, conclusions regarding model terms can change if

the clustering of the data is not appropriately accounted for. Regarding the correlation matrix of the MML estimates (not shown), the order of parameter estimates is: K intercept terms, P covariate terms for the first of K response category comparisons, P covariate terms for the second of K response category comparisons, ... P covariate terms for the Kth response category comparison.

An additional feature of MIXNO allows the random-effect variance term(s) to vary across the K response category comparisons. This feature uses the VCAT option on line 6 of the MIXNO.DEF file. The MIXNO.DEF file given in Table 5.6 illustrates usage of this option.

Table 5.6 - MIXNO.DEF file for second students-within-classrooms analysis															
TVSFP study - Post-Test Tobacco and Health Knowledge Scale															
Students nested w/in CLASSROOMS - Varying Random Intercept															
TVSFPORS.DAT															
TVSFPNO3.OUT															
TVSFPNO3.DEF															
1	9	1	4	0.00010	4	0	0	0	0	10	0	0	0	0	1
2	3														
5															
6	7	8	9												
1	2	3	4												
THKScore															
Intrcpt															
PreTHKS	CC	TV	CC*TV												

Enabling this option is achieved simply by specifying VCAT=1 on line 6 of the MIXNO.DEF file. Results from this analysis are given in Table 5.7. These results are similar to the results from the model assuming homogeneous random-effect variance across response code comparisons (*i.e.*, the results in Table 5.5), though the standard errors do change, especially for the school-varying condition effects. Many fewer iterations were required for this model that relaxes the homogeneity of variance assumption across the response code comparisons. Also of interest, the estimated random-effect standard deviation increases across the response code comparisons. Expressed as intraclass correlations, they are given as .026, .099, and .125 for the three comparisons. This suggests that the school effect is more pronounced when contrasting the more extreme response categories (*i.e.*, category 4 vs. 1). A likelihood-ratio test of $\sigma_{\beta}^{(2 \text{ vs } 1)} = \sigma_{\beta}^{(3 \text{ vs } 1)} = \sigma_{\beta}^{(4 \text{ vs } 1)}$ yields $\chi_2^2 = 4234.684 - 4223.020 = 11.664$, $p < .01$, supporting rejection of the homogeneous variance assumption.

5.3 Section 5 MIXNO Output

Table 5.3 - output file for student-level analysis ignoring clustering

MIXNO - The program for mixed-effects nominal logistic regression analysis

TVSFP study - Post-Test Tobacco and Health Knowledge Scale
 Students nested within classrooms - Independence model

Numbers of observations

Level 1 observations = 1600

Descriptive statistics for all variables

Variable	Minimum	Maximum	Mean	Std.Dev.
THKScore	1.00000	4.00000	2.58688	1.11612
Intrcpt	1.00000	1.00000	1.00000	0.00000
PreTHKS	0.00000	6.00000	2.06937	1.26018
CC	0.00000	1.00000	0.47687	0.49962
TV	0.00000	1.00000	0.49938	0.50016
CC*TV	0.00000	1.00000	0.23938	0.42684

Categories of the response variable THKScore

Category	Frequency	Proportion
1.00	355.00	0.22187
2.00	398.00	0.24875
3.00	400.00	0.25000
4.00	447.00	0.27937

Crosstabulation of variable PreTHKS by the response variable THKScore

PreTHKS	THKScore				Total
	1.00	2.00	3.00	4.00	
0.00	55.0 (0.37)	41.0 (0.27)	35.0 (0.23)	19.0 (0.13)	150.0
1.00	117.0 (0.28)	126.0 (0.30)	89.0 (0.21)	87.0 (0.21)	419.0
2.00	106.0 (0.22)	119.0 (0.25)	135.0 (0.28)	121.0 (0.25)	481.0
3.00	54.0 (0.16)	73.0 (0.22)	96.0 (0.29)	109.0 (0.33)	332.0
4.00	15.0 (0.09)	34.0 (0.21)	38.0 (0.23)	76.0 (0.47)	163.0
5.00	8.0 (0.17)	4.0 (0.08)	7.0 (0.15)	29.0 (0.60)	48.0
6.00	0.0 (0.00)	1.0 (0.14)	0.0 (0.00)	6.0 (0.86)	7.0
Total	355.0	398.0	400.0	447.0	1600.0

Starting values

```

-----
covariates -1.255 0.948 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
covariates 0.000 0.000 0.000 0.000 0.000

```

 * Final Results - Maximum Marginal Likelihood Estimates *

Total Iterations = 8
 Log Likelihood = -2122.808
 Deviance (-2logL) = 4245.617
 Ridge = 0.000

Variable	Estimate	Stand.Error	Z	p-value
----------	----------	-------------	---	---------

RESPONSE CODE 2. vs CODE 1.

fixed effects

Intrcpt	-0.21907	0.17564	-1.24721	0.21232 (2)
PreTHKS	0.16427	0.06321	2.59888	0.00935 (2)
CC	0.17864	0.21440	0.83321	0.40473 (2)
TV	-0.11896	0.18820	-0.63206	0.52735 (2)
CC*TV	0.15998	0.30306	0.52788	0.59758 (2)

RESPONSE CODE 3. vs CODE 1.

fixed effects

Intrcpt	-0.96134	0.19483	-4.93435	0.00000 (2)
PreTHKS	0.33728	0.06388	5.28012	0.00000 (2)
CC	0.90177	0.21699	4.15584	0.00003 (2)
TV	0.13087	0.20322	0.64399	0.51958 (2)
CC*TV	-0.10633	0.30367	-0.35015	0.72622 (2)

RESPONSE CODE 4. vs CODE 1.

fixed effects

Intrcpt	-1.72156	0.20095	-8.56727	0.00000 (2)
PreTHKS	0.63227	0.06176	10.23803	0.00000 (2)
CC	1.23329	0.21982	5.61054	0.00000 (2)
TV	0.37443	0.20446	1.83129	0.06706 (2)
CC*TV	-0.54476	0.30424	-1.79055	0.07337 (2)

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

Table 5.5 - output file for first students-within-classrooms analysis

MIXNO - The program for mixed-effects nominal logistic regression analysis

TVSFP study - Post-Test Tobacco and Health Knowledge Scale

Students nested w/in CLASSROOMS - Random Intercept

Random-effects distribution: normal

Numbers of observations

Level 1 observations = 1600

Level 2 observations = 135

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

```

26 11 10 15 12 12 10 21 10 17 19 2 4 21 16 15 13 2 14
13 1 12 18 21 17 16 20 3 11 9 5 15 16 21 21 27 17 3
2 15 7 24 22 15 19 7 12 8 6 11 7 7 8 3 5 8 3
8 9 8 2 11 9 21 13 12 12 14 9 6 11 10 12 11 6 6
14 10 14 2 3 2 4 3 6 10 14 11 6 22 4 7 22 18 23
19 14 5 14 28 15 15 11 12 11 11 15 17 24 20 15 6 8 14
5 11 9 17 14 11 17 15 6 7 14 10 14 18 4 9 7 12 15
11 10

```

Descriptive statistics for all variables

Variable	Minimum	Maximum	Mean	Std.Dev.
THKScore	1.00000	4.00000	2.58688	1.11612
Intrcpt	1.00000	1.00000	1.00000	0.00000
PreTHKS	0.00000	6.00000	2.06937	1.26018
CC	0.00000	1.00000	0.47687	0.49962
TV	0.00000	1.00000	0.49938	0.50016
CC*TV	0.00000	1.00000	0.23938	0.42684

Categories of the response variable THKScore

Category	Frequency	Proportion
1.00	355.00	0.22187
2.00	398.00	0.24875
3.00	400.00	0.25000
4.00	447.00	0.27937

Starting values

```

-----
mean          -3.481 -2.344 -1.278
covariates    -0.212 -0.451 -0.140 0.201 -0.233 -0.496 -0.154 0.221 -0.254 -0.541
covariates    -0.168 0.241
var. terms   0.574

```

```

==> The number of level 2 observations with non-varying responses
     = 6 ( 4.44 percent )

```

```

-----
* Final Results - Maximum Marginal Likelihood Estimates *
-----

```

```

Total Iterations   =      152
Quad Pts per Dim   =        10
Log Likelihood     =  -2117.342
Deviance (-2logL)  =   4234.684
Ridge              =        0.100

```

```

-----
Variable           Estimate      Stand.Error      Z      p-value
-----

```

RESPONSE CODE 2. vs CODE 1.

fixed effects

```

-----
Intrcpt           -0.17170      0.22197      -0.77353      0.43921 (2)
PreTHKS           0.16410      0.07428       2.20934      0.02715 (2)
CC                 0.19437      0.27060       0.71830      0.47257 (2)
TV                -0.18887      0.27547      -0.68563      0.49295 (2)
CC*TV             0.23236      0.41112       0.56520      0.57194 (2)

```

RESPONSE CODE 3. vs CODE 1.

fixed effects

```

-----
Intrcpt           -0.91385      0.23343      -3.91493      0.00009 (2)
PreTHKS           0.33714      0.07675       4.39277      0.00001 (2)
CC                 0.91732      0.24924       3.68045      0.00023 (2)
TV                 0.06047      0.26974       0.22419      0.82261 (2)
CC*TV            -0.03328      0.36414      -0.09140      0.92717 (2)

```

Variable	Estimate	Stand. Error	Z	p-value
----------	----------	--------------	---	---------

RESPONSE CODE 4. vs CODE 1.

fixed effects

Intrcpt	-1.67408	0.22201	-7.54058	0.00000 (2)
PreTHKS	0.63221	0.06959	9.08454	0.00000 (2)
CC	1.24830	0.26206	4.76348	0.00000 (2)
TV	0.30358	0.23061	1.31640	0.18804 (2)
CC*TV	-0.47039	0.35820	-1.31320	0.18911 (2)

random effect variance term: expressed as a standard deviation

Intrcpt	0.51098	0.13859	3.68693	0.00011 (1)
---------	---------	---------	---------	-------------

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

(2) = 2-tailed p-value

Calculation of the intracluster correlation

residual variance = $\pi \cdot \pi / 3$ (assumed)

1 cluster variance = $(0.511 * 0.511) = 0.261$

intracluster correlation = $0.261 / (0.261 + (\pi \cdot \pi / 3)) = 0.074$

Table 5.7 - output file for second students-within-classrooms analysis

MIXNO - The program for mixed-effects nominal logistic regression analysis

TVSFP study - Post-Test Tobacco and Health Knowledge Scale

Students nested w/in CLASSROOMS - Varying Random Intercept

Random-effects distribution: normal

Numbers of observations

Level 1 observations = 1600

Level 2 observations = 135

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

```

26 11 10 15 12 12 10 21 10 17 19 2 4 21 16 15 13 2 14
13 1 12 18 21 17 16 20 3 11 9 5 15 16 21 21 27 17 3
 2 15 7 24 22 15 19 7 12 8 6 11 7 7 8 3 5 8 3
 8 9 8 2 11 9 21 13 12 12 14 9 6 11 10 12 11 6 6
14 10 14 2 3 2 4 3 6 10 14 11 6 22 4 7 22 18 23
19 14 5 14 28 15 15 11 12 11 11 15 17 24 20 15 6 8 14
 5 11 9 17 14 11 17 15 6 7 14 10 14 18 4 9 7 12 15
11 10

```

Descriptive statistics for all variables

Variable	Minimum	Maximum	Mean	Std.Dev.
THKScore	1.00000	4.00000	2.58688	1.11612
Intrcpt	1.00000	1.00000	1.00000	0.00000
PreTHKS	0.00000	6.00000	2.06937	1.26018
CC	0.00000	1.00000	0.47687	0.49962
TV	0.00000	1.00000	0.49938	0.50016
CC*TV	0.00000	1.00000	0.23938	0.42684

Categories of the response variable THKScore

Category	Frequency	Proportion
1.00	355.00	0.22187
2.00	398.00	0.24875
3.00	400.00	0.25000
4.00	447.00	0.27937

Starting values

```

-----
mean          -3.481 -2.344 -1.278
covariates    -0.212 -0.451 -0.140 0.201 -0.233 -0.496 -0.154 0.221 -0.254 -0.541
covariates    -0.168 0.241
var. terms   0.574 0.631 0.688

```

```

==> The number of level 2 observations with non-varying responses
     = 6 ( 4.44 percent )

```

```

-----
* Final Results - Maximum Marginal Likelihood Estimates *
-----

```

```

Total Iterations   =          38
Quad Pts per Dim   =          10
Log Likelihood     =   -2111.510
Deviance (-2logL)  =    4223.020
Ridge              =          0.100

```

```

-----
Variable           Estimate      Stand.Error      Z           p-value
-----

```

RESPONSE CODE 2. vs CODE 1.

fixed effects

```

-----
Intrcpt           -0.17829      0.19585      -0.91037     0.36263 (2)
PreTHKS           0.16539      0.07384       2.23971     0.02511 (2)
CC                0.20052      0.22003       0.91136     0.36211 (2)
TV               -0.14241      0.22035      -0.64631     0.51808 (2)
CC*TV            0.18532      0.33990       0.54522     0.58561 (2)

```

random effect variance term: expressed as a standard deviation

```

Intrcpt           0.29519      0.14212       2.07706     0.01890 (1)

```

RESPONSE CODE 3. vs CODE 1.

fixed effects

```

-----
Intrcpt           -0.95871      0.25323      -3.78592     0.00015 (2)
PreTHKS           0.34054      0.07615       4.47201     0.00001 (2)
CC                0.93555      0.27710       3.37627     0.00073 (2)
TV                0.06972      0.28570       0.24404     0.80720 (2)
CC*TV            -0.03491      0.39831      -0.08763     0.93017 (2)

```

random effect variance term: expressed as a standard deviation

```

Intrcpt           0.60260      0.15590       3.86536     0.00006 (1)

```

Variable	Estimate	Stand. Error	Z	p-value
----------	----------	--------------	---	---------

RESPONSE CODE 4. vs CODE 1.

fixed effects

Intrcpt	-1.74341	0.26148	-6.66753	0.00000 (2)
PreTHKS	0.63614	0.07113	8.94370	0.00000 (2)
CC	1.26668	0.30773	4.11621	0.00004 (2)
TV	0.30216	0.27787	1.08743	0.27685 (2)
CC*TV	-0.45566	0.42021	-1.08436	0.27821 (2)

random effect variance term: expressed as a standard deviation

Intrcpt	0.68647	0.13900	4.93872	0.00000 (1)
---------	---------	---------	---------	-------------

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

(2) = 2-tailed p-value

Calculation of the intracluster correlation

residual variance = $\pi^2 / 3$ (assumed)

1 cluster variance = $(0.295 * 0.295) = 0.087$

intracluster correlation = $0.087 / (0.087 + (\pi^2/3)) = 0.026$

2 cluster variance = $(0.603 * 0.603) = 0.363$

intracluster correlation = $0.363 / (0.363 + (\pi^2/3)) = 0.099$

3 cluster variance = $(0.686 * 0.686) = 0.471$

intracluster correlation = $0.471 / (0.471 + (\pi^2/3)) = 0.125$

6 Analysis of a Longitudinal Dataset

The McKinney Homeless Research Project study ([37], [38]) was designed to evaluate the effectiveness of using Section 8 certificates to provide independent housing to the severely mentally ill homeless. These housing certificates, which require clients to pay 30% of their income toward rent, are designed to make it possible for low income individuals to choose and obtain independent housing in the community. Three hundred sixty-two clients took part in this longitudinal study employing a randomized factorial design. Clients were randomly assigned to one of two types of supportive case management (comprehensive vs. traditional) and to one of two level of access to independent housing (using Section 8 certificates). Eligibility for the project was restricted to individuals diagnosed with a severe and persistent mental illness who were either homeless or at high risk of becoming homeless at the start of the study. Individuals' housing status was classified at baseline and at 6, 12, and 24 month follow-ups. Here, we focus on examining the effect of access to Section 8 certificates on housing outcomes across time. Specifically, at each timepoint subjects' housing status was classified as either streets/shelters, community housing, or independent housing. A partial list of these data is given in Table 6.1.

9	1	1	1	0	0	0	0	0	0	0	0	0
9	1	1	1	1	0	0	1	0	0	0	1	1
9	2	1	1	0	1	0	0	1	0	0	2	2
9	2	1	1	0	0	1	0	0	1	0	3	3
10	2	1	1	0	0	0	0	0	0	0	0	0
10	999	1	1	1	0	0	1	0	0	0	1	1
10	999	1	1	0	1	0	0	1	0	0	2	2
10	999	1	1	0	0	1	0	0	1	0	3	3
11	1	1	1	0	0	0	0	0	0	0	0	0
11	2	1	1	1	0	0	1	0	0	0	1	1
11	999	1	1	0	1	0	0	1	0	0	2	2
11	1	1	1	0	0	1	0	0	1	0	3	3
361	0	1	0	0	0	0	0	0	0	1	0	0
361	0	1	0	1	0	0	0	0	0	1	1	0
361	999	1	0	0	1	0	0	0	0	1	2	0
361	999	1	0	0	0	1	0	0	0	1	3	0
362	1	1	0	0	0	0	0	0	0	1	0	0
362	1	1	0	1	0	0	0	0	0	1	1	0
362	1	1	0	0	1	0	0	0	0	1	2	0
362	1	1	0	0	0	1	0	0	0	1	3	0

The 13 variables are: (in order) subject ID, housing status (1=street, 2=community, 3=independent), a column of ones for the intercept, Section 8 group (0=no, 1=yes), 3 dummy-codes for time effects (Time1 = 1 for the 6 month follow-up and 0 otherwise, Time2 = 1 for the 12

month follow-up and 0 otherwise, and Time3 = 1 for the 24 month follow-up and 0 otherwise), the three Section 8 by time interaction terms (*i.e.*, the products of Section 8 by Time1, Section 8 by Time2, and Section 8 by Time3), Non Section 8 group (0=no, 1=yes), linear time contrast (0 = baseline, 1 = 6-month, 2 = 12-month, and 3 = 24-month followup), and the product of Section 8 by linear time. Notice that values of 999 are sometimes present for the housing status variable. These represent missing value codes. Thus, some subjects are measured at all four timepoints and others (*e.g.*, subjects 10, 11, and 361) at fewer timepoints. The data from these timepoints with missing value codes will not be used in the analysis, however data from these subjects at other timepoints 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 timepoint must be complete. The number of repeated observations per subject then depends on the number of timepoints for which there are non-missing data for that subject.

Based on these data, we can obtain the observed sample sizes and response proportions by group that are presented below in Table 6.2.

Table 6.2 - Response proportions across time by group					
group	status	timepoint			
		baseline	6-months	12-months	24-months
control	street	.555	.186	.089	.124
	community	.339	.578	.582	.455
	independent	.106	.236	.329	.421
	<i>n</i>	180	161	146	145
section 8	street	.442	.093	.121	.120
	community	.414	.280	.146	.228
	independent	.144	.627	.732	.652
	<i>n</i>	181	161	157	158

These observed proportions indicate a general decrease in street living and an increase in independent living across time for both groups. The increase in independent housing, however, appears to occur sooner for the section 8 group relative to the control group. Regarding community living, across time there is an increase for the control group and a decrease for the section 8 group.

Regarding missing data, further inspection of Table 6.2 indicates that there is some attrition across time; attrition rates of 19.4% and 12.7% are observed at the final timepoint for the control

and section 8 groups, respectively. Also, one subject provided no housing data at any of the four measurement timepoints. Since estimation of model parameters is based on a full-likelihood approach, the missing data are assumed to be “ignorable” conditional on both the model covariates and the observed nominal responses [39]. In longitudinal studies, ignorable nonresponse falls under Rubin’s [40] “missing at random” (MAR) assumption, in which the missingness depends only on observed data. In what follows, since the focus is on describing the mixed-effects multinomial regression model, we will make the MAR assumption. A further approach, however, that does not rely on the MAR assumption (*e.g.*, a mixed-effects pattern-mixture model as described in [41]) could be used.

In preparation for the subsequent analyses, the marginal response proportions can be converted to the two logits of the nominal regression model (*i.e.*, $\log[p(C)/p(S)]$ and $\log[p(I)/p(S)]$, where S =street, C =community, and I =independent housing).² These logits are given in Table 6.3.

group	status	timepoint			
		baseline	6-months	12-months	24-months
control	community vs street	-.49	1.13	1.88	1.30
	independent vs street	-1.66	.24	1.31	1.22
section 8	community vs street	-.07	1.10	.19	.64
	independent vs street	-1.12	1.91	1.80	1.69
<i>difference</i>	community vs street	.42	-.03	-1.69	-.66
	independent vs street	.54	1.67	.49	.47

These logits clearly show the increase in community and independent housing, relative to street housing, at all follow-up timepoints (6, 12, and 24 months). In terms of group-related differences, these appear most pronounced at 6-months for independent housing and 12-months for community housing. While examination of these logits is helpful in getting a general feel for the data, the subsequent MIXNO analyses will more rigorously assess the degree to which these logits vary by time and group.

A series of mixed-effects multinomial logistic regression models were fit to these data using MIXNO. The first analysis assumes one random subject effect (*i.e.*, a random subject intercept) and models the repeated housing status classifications in terms of the dummy-coded time effects (6,

²Again, street housing is treated as the reference category because its code (0) is listed as the first response category. A different reference category can be selected by altering the listing of the response categories in MIXNO.

12, and 24 month follow-ups compared to baseline), a group effect (section 8 versus control), and group by time interaction terms. By varying choices on MIXNO options, this random intercepts is modified to allow first for a uniform distribution for the random effects (instead of the default normal distribution), and then varying random-effect variance terms by treatment groups. Finally, a trend analysis is also performed which uses the linear time contrast instead of the dummy-coded time effects. For the trend analysis two random subject effects, the intercept and linear trend, are included.

6.1 Random intercept model with dummy-coded time effects

Table 6.4 lists the MIXNO.DEF file that specifies a random subject intercept and fixed effects of group (Section 8), time (dummy-coded time effects), and group by time interaction.

Table 6.4 - MIXNO.DEF file for random intercept model						
San Diego Homeless Project - Housing Outcome across time						
random intercepts model - Section8 & Time effects						
sdhouse.dat						
sdhouse.out						
sdhouse.def						
1	13	1	7	0.00010	3	1 0 0 1 0 0 20 0 0 0 0 1
1	2					
3						
4	5	6	7	8	9	10
0	1	2				
4	2	0	1			
999						
9						
9	9	9	9	9	9	9
Housing						
Intercpt						
Section8Time1	Time2	Time3	Sect8T1	Sect8T2	Sect8T3	

In terms of options, three are indicated on line 6: missing values are present in the data (MISS=1), a table of housing status by Section 8 is being requested (CATYX=1), and separate variance terms will be estimated for the K contrasts of the nominal outcome (VCAT=1). The missing value codes are indicated as 999 for housing status and 9 for all other variables. The results from this analysis are listed in Table 6.5. Comparing the log-likelihood value from this analysis to one where there are no random effects (not shown) clearly supports inclusion of the random subject effect (likelihood ratio $\chi_1^2 = 134.3$). Expressed as intraclass correlations, MIXNO indicates $r_1 = .19$ and $r_2 = .62$ for

community versus street and independent versus street, respectively. Thus, the subject influence is much more pronounced in terms of distinguishing independent versus street living, relative to community versus street living. This is borne out by contrasting models with separate versus a common random-effect variance across the two category contrasts (*i.e.*, $\text{VCAT}=0$, not shown) which yields a highly significant likelihood ratio $\chi^2_1 = 49.2$ favoring the model with separate variance terms.

In terms of statistical significance of the fixed-effects, the Time effects are observed to be highly significant. Since the Time by Section 8 interaction terms are included in the model, the Time effects reflect comparisons between timepoints for the control group (*i.e.*, Section 8 coded as 0). Thus, subjects in the control group increase both independent and community housing relative to street housing at all three follow-ups, as compared to baseline. Similarly, because of the inclusion of the interaction terms, the Section 8 effect is the group difference at baseline (*i.e.*, when all Time effects are 0). Using a .05 cutoff, there is no statistical evidence of group differences at baseline. Turning to the interaction terms, these indicate how the two groups differ in terms of comparisons between timepoints. Compared to controls, the increase in community versus street housing is less pronounced for section 8 subjects at 12 months (the estimate equals -1.92 in terms of the logit), but not statistically different at 6 months and only marginally different at 24 months. Conversely, as compared to controls, the increase in independent versus street housing is more pronounced for section 8 subjects at 6 months (the estimate equals 2.00 in terms of the logit), but not statistically different at 12 or 24 months.

In terms of community versus street housing (*i.e.*, response code 1 versus 0), there is an increase across time for the control group relative to the Section 8 group. As the statistical test indicated, these groups differ most at 12 months. For the independent versus street housing comparison (*i.e.*, response code 2 versus 0) there is a beneficial effect of Section 8 certificates at 6 months. Thereafter, the nonsignificant interaction terms indicates that the control group catches up to some degree. Considering these results of the mixed-effects analysis, it is seen that both groups reduce the degree of street housing, but do so in somewhat different ways. The control group subjects are shifted more towards community housing, whereas Section 8 subjects are more quickly shifted towards independent housing.

This differential effect of Section 8 certificates over time is completely missed if one simply analyzes the outcome variable as a binary indicator of street versus non-street housing (*i.e.*, collapsing

community and independent housing categories). In this case (not shown), none of the section 8 by time interaction terms are observed to be statistically significant. Thus, analysis of the three category nominal outcome is important in uncovering the beneficial effect of Section 8 certificates.

6.2 Uniform-distributed random effects

By default, MIXNO assumes that the random effects are normally distributed. The program does allow, however, users to select a uniform distribution as an alternative to the normal. Since the uniform distribution can be thought to represent vague information about the form of the random-effects distribution, contrasting the results from the normal to those from the uniform provides some idea about the sensitivity of the results to the random-effects distribution form. Table 6.6 lists the MIXNO.DEF file for the analysis selecting the uniform distribution.

Table 6.6 - MIXNO.DEF file for uniform distributed random subject effects						
San Diego Homeless Project - Housing Outcome across time						
random intercepts model - Section8 & Time effects						
sdhouse.dat						
sdhouseu.out						
sdhouseu.def						
1	13	1	7	0.00010	3	1 0 0 1 0 1 20 0 0 0 0 1
1	2					
3						
4	5	6	7	8	9	10
0	1	2				
4	2	0	1			
999						
9						
9	9	9	9	9	9	9
Housing						
Intercpt						
Section8Time1	Time2	Time3	Sect8T1	Sect8T2	Sect8T3	

In terms of options, notice that, relative to Table 6.4, there is one additional option requested on line 6: selection of the uniform distribution for the random effects (*i.e.*, UNIF=1). Also, the output and definition files have been given different names, so that the previous results are not overwritten. Besides these differences, the two MIXNO.DEF files are identical. Comparing the results of this analysis (not shown) to the previous one yields very similar estimates and the same conclusions. Thus, the random-effects distributional form does not appear to play an important role for these data.

6.3 Varying random-effects variance terms by groups

In some cases it may be reasonable to allow the random-effects variance terms to vary across groups of subjects. For example, there may be interest in estimating separate intraclass correlations by gender or racial groups. To accomplish this in MIXNO, the DAT file must contain M dummy-codes (with codings of 0 or 1) for the M groups, which are then specified as random effects representing dummy-coded grouping variables by indicating $VGRP = 1$. For example, to allow the random-effects variance terms to vary by gender, two dummy-codes, one for males and another for females, must be included in the DAT file. For the San Diego Homeless data, two such dummy-codes are included in the DAT file (in fields 4 and 11) to indicate the two treatment groups: Section 8 and Non-Section 8. Table 6.7 lists the MIXNO.DEF file that can be used to allow the random-effects variance terms to vary across these two groups.

Table 6.7 - MIXNO.DEF file for group-varying random-effects variance terms																	
San Diego Homeless Project - Housing Outcome across time																	
Varying random intercepts by groups																	
sdhouse.dat																	
sdhouseg.out																	
sdhouseg.def																	
1	13	2	8	0.00010	3	1	0	0	1	0	0	20	0	0	1	1	1
1	2																
4	11																
3	4	5	6	7	8	9	10										
0	1	2															
4	2	0	1														
999																	
9	9																
9	9	9	9	9	9	9	9										
Housing																	
IntSect8IntCont																	
IntercptSection8Time1	Time2	Time3	Sect8T1	Sect8T2	Sect8T3												

Notice that, relative to Table 6.4, there are a few differences in the MIXNO.DEF file. First, both group indicator variables (fields 4 and 11) are specified as random effects ($R=2$), while the intercept (field 3) is included as a fixed effect ($P=8$). Also, both $NOMU=1$ and $VGRP=1$ are indicated on line 6. This instructs MIXNO that the two random effects are grouping variables and that the mean vector of the random effects (*i.e.*, the separate intercepts for the two groups) are not to be estimated. Without the latter specification, the model would be over-identified since the intercept and group effect (Section8) are included as fixed effects.

Allowing the random-effects variance terms to vary by groups produces near-identical results (not shown). A likelihood-ratio test comparing this model to that of Table 6.5 yields $\chi^2 = 2218.725 - 2218.431 = .294$ (on 2 degrees of freedom) which is clearly not significant. In terms of the estimates of the random-effects variance terms, Table 6.8 lists the results that are obtained for these two models.

	logit 1 vs. 0	logit 2 vs. 0
common variance	.871	2.34
separate variance		
Section 8	.961	2.43
Non Section 8	.771	2.23

As can be seen, the estimates do not vary greatly between the two groups and are similar to the pooled estimates.

6.4 Trend analysis

For longitudinal data, a common approach is to posit a model where changes across time are represented by low-order polynomial trends, for example, linear or quadratic trends. Further, these trends are often allowed to vary across subjects, leading to models with multiple random subject effects. For continuous data, this approach was introduced by Laird and Ware [42], Bock [43], Strenio et al. [44], Jennrich and Schluchter [45] and others. Similarly, this approach has been developed and described for dichotomous [13] and ordinal outcomes [19].

For nominal data, trend analysis of this type can be done using MIXNO by embedding the polynomial contrasts in the DAT file read by the program. For a linear contrast, a simple approach is to include a variable that starts at 0 for the first timepoint and is incremented by 1 for each subsequent timepoint. This assumes that the change in the logits is linear across the time intervals. More sophisticated approaches can be used to center and/or orthogonalize the polynomials, or to take into account unequal time intervals (see [46] for a complete treatment on this subject; for an example see [47]). Orthogonalization of the polynomials is particularly useful when fitting models with multiple polynomials (*i.e.*, linear and quadratic; or linear, quadratic, and cubic). For the current example, only a linear time effect will be considered. For this, a linear time variable and

a group by linear time interaction are included as the 12th and 13th fields, respectively (see Table 6.1). Table 6.9 lists the MIXNO.DEF file for a trend analysis considering both subject-varying intercepts and linear trends, and fixed effects of group and group by linear time.

Table 6.9 - MIXNO.DEF file for trend analysis																	
San Diego Homeless Project - Housing Outcome across time																	
Random intercepts and slopes model - SECT8 & TIME EFFECTS																	
sdhouse.dat																	
sdhouset.out																	
sdhouset.def																	
1	13	2	2	0.00010	3	1	0	0	1	0	0	10	6	0	0	0	1
1	2																
3	12																
4	13																
0	1	2															
12	4	0	1	2	3												
999																	
9 9																	
9 9																	
Housing																	
IntercptTime																	
Section8Sec8Time																	
0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	
0	0	0	0	1	2	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	1	2	0	0	0	0	0	0	0	0	0	
0	0	0	0	1	3	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	1	3	0	0	0	0	0	0	0	0	0	

Notice that for the selected CATYX option, a crosstabulation table of Housing by Time (the variable in the 12th field with 4 values: 0, 1, 2, and 3) is requested. Also notice that 10 quadrature points per dimension are indicated on line 6. Since there are two random effects, the quadrature will go over $10^2 = 100$ points at each iteration. Finally, notice that LINFN=6 is indicated on line 6, requesting six re-expressions of the estimated parameters. The coefficients for these are given as the last six lines of the MIXNO.DEF file. Each of these lines indicate coefficients for the 14 parameters of the model, which are (in order):

- Intercept and Time (response code 1 vs. 0): 2 terms;
- Intercept and Time (response code 2 vs. 0): 2 terms;
- Section8 and Sec8Time (response code 1 vs. 0): 2 terms;

- Section8 and Sec8Time (response code 2 vs. 0): 2 terms;
- Elements of the Cholesky matrix corresponding to the Intercept variance, Intercept-Time covariance, Time covariance (response code 1 vs. 0): 3 terms;
- Elements of the Cholesky matrix corresponding to the Intercept variance, Intercept-Time covariance, Time covariance (response code 2 vs. 0): 3 terms.

Thus, re-expressions 1, 3, and 5 represent the Section8 effects at Time=1, 2, and 3, respectively, for the response code 1 vs. 0 comparison. Similarly, re-expressions 2, 4, and 6 are the Section8 effects at Time=1, 2, and 3, respectively, for response code 2 vs. 0. Note that these three values of time correspond to the codings for the three follow-up timepoints.

Table 6.10 lists the output for this trend analysis. Notice first the marginal crosstabulation table of Housing by Time that is produced. This clearly shows the overall decline in street housing (response category 0) after the first timepoint. What is of more interest, though, is whether the trend across time in housing status differs between the two groups. Turning to the results of the analysis, it is apparent that, while there is a highly significant effect of time, no group-related statistical differences emerge when comparing community to street housing (response code 1 vs 0). However, turning to the comparison of independent to street housing (response code 2 vs 0) reveals highly significant effects of Time and Section 8, and a marginally significant Section 8 by Time interaction. This suggests that there are more independent versus street housing responses for the Section 8 group at baseline, relative to the control group, and that this Section 8 increase in independent responses tends to become more pronounced over time. However, as indicated in Table 6.3, the group differences in the logits do not appear to be linear across time.

Inspection of the transforms of the estimated parameters indicates what the linear model estimates for the group differences across time. As mentioned above, transforms 1, 3, and 5 represent the group difference at each follow-up timepoint for the 1 vs 0 response code comparison (*i.e.*, community vs street), and transforms 2, 4, and 6 provide similar estimates for the 2 vs 0 response code comparison (independent vs street). These transforms, which are estimated based on this linear model, indicate no group difference in terms of the community vs street comparison at any follow-up timepoint, and large increasing differences for the independent vs street comparison. This does not agree well with the previous analyses, mainly because the observed group differences, as given in Table 6.3, do not follow a linear pattern across time. Thus, the linear representation of the

time effect which uses only 1 degree of freedom for comparing outcomes across time (as opposed to the 3 degrees of freedom used by the time dummy-codes used previously) is an over-simplification for these data.

6.5 Model fit of marginal proportions

To examine the degree of model fit to the marginal proportions in Table 6.2 requires a bit of work that is greatly facilitated using a software program with matrix algebra routines. Here SAS-IML will be used, however other programs can also be used for this purpose (*e.g.*, GAUSS, S-PLUS, or the matrix routines in SPSS). We will demonstrate this using the results provided in Tables 6.5 and 6.10 to illustrate the procedure for models with one or multiple random effects.

6.5.1 Fit of Table 6.5 results

This model included a single random effect, and so using Equations (1) and (2) yields

$$P_{ij2} = \frac{\exp(\sigma_2\theta_i + \mu_2 + \mathbf{w}'_{ij}\boldsymbol{\alpha}_2)}{1 + \exp(\sigma_1\theta_i + \mu_1 + \mathbf{w}'_{ij}\boldsymbol{\alpha}_1) + \exp(\sigma_2\theta_i + \mu_2 + \mathbf{w}'_{ij}\boldsymbol{\alpha}_2)} \quad (11)$$

$$P_{ij1} = \frac{\exp(\sigma_1\theta_i + \mu_1 + \mathbf{w}'_{ij}\boldsymbol{\alpha}_1)}{1 + \exp(\sigma_1\theta_i + \mu_1 + \mathbf{w}'_{ij}\boldsymbol{\alpha}_1) + \exp(\sigma_2\theta_i + \mu_2 + \mathbf{w}'_{ij}\boldsymbol{\alpha}_2)} \quad (12)$$

$$P_{ij0} = \frac{1}{1 + \exp(\sigma_1\theta_i + \mu_1 + \mathbf{w}'_{ij}\boldsymbol{\alpha}_1) + \exp(\sigma_2\theta_i + \mu_2 + \mathbf{w}'_{ij}\boldsymbol{\alpha}_2)} \quad (13)$$

for the subject-specific response probabilities of the three categories at a particular timepoint j . These are referred to as “subject-specific” probabilities because they indicate response probabilities for particular values of the random subject effect θ_i ([50], [51]). Replacing the parameters with their estimates and denoting the resulting subject-specific probabilities as \hat{P}_{ss} , marginal probabilities \hat{P}_m are then obtained by integrating over the random-effect distribution, namely $\hat{P}_m = \int_{\theta} \hat{P}_{ss} g(\theta) d\theta$. Numerical quadrature can be used for this. Table 6.11 (at the end of this section) provides a listing of SAS-IML code that implements a 10 point quadrature solution for this. The estimated response probabilities that are listed in Table 6.12 were obtained using this SAS-IML code.

group	status	baseline	timepoint		
			6-months	12-months	24-months
control	street	.556	.189	.091	.130
	community	.336	.579	.587	.465
	independent	.108	.233	.322	.405
section 8	street	.444	.090	.127	.121
	community	.411	.284	.155	.235
	independent	.145	.626	.719	.644

These agree well with the observed marginal proportions in Table 6.2. This is not too surprising since the model included one degree of freedom for the group difference, and three each for both the time and group by time interaction terms.

6.5.2 Fit of Table 6.10 results

Because this model included two random effects, the formulas for the probabilities are a little more complicated. Specifically, we get

$$P_{ij2} = \frac{\exp[\mathbf{x}'_{ij}(\mathbf{T}_2\boldsymbol{\theta}_i + \boldsymbol{\mu}_2) + \mathbf{w}'_{ij}\boldsymbol{\alpha}_2]}{1 + \exp[\mathbf{x}'_{ij}(\mathbf{T}_1\boldsymbol{\theta}_i + \boldsymbol{\mu}_1) + \mathbf{w}'_{ij}\boldsymbol{\alpha}_1] + \exp[\mathbf{x}'_{ij}(\mathbf{T}_2\boldsymbol{\theta}_i + \boldsymbol{\mu}_2) + \mathbf{w}'_{ij}\boldsymbol{\alpha}_2]} \quad (14)$$

$$P_{ij1} = \frac{\exp[\mathbf{x}'_{ij}(\mathbf{T}_1\boldsymbol{\theta}_i + \boldsymbol{\mu}_1) + \mathbf{w}'_{ij}\boldsymbol{\alpha}_1]}{1 + \exp[\mathbf{x}'_{ij}(\mathbf{T}_1\boldsymbol{\theta}_i + \boldsymbol{\mu}_1) + \mathbf{w}'_{ij}\boldsymbol{\alpha}_1] + \exp[\mathbf{x}'_{ij}(\mathbf{T}_2\boldsymbol{\theta}_i + \boldsymbol{\mu}_2) + \mathbf{w}'_{ij}\boldsymbol{\alpha}_2]} \quad (15)$$

$$P_{ij0} = \frac{1}{1 + \exp[\mathbf{x}'_{ij}(\mathbf{T}_1\boldsymbol{\theta}_i + \boldsymbol{\mu}_1) + \mathbf{w}'_{ij}\boldsymbol{\alpha}_1] + \exp[\mathbf{x}'_{ij}(\mathbf{T}_2\boldsymbol{\theta}_i + \boldsymbol{\mu}_2) + \mathbf{w}'_{ij}\boldsymbol{\alpha}_2]} \quad (16)$$

for the subject-specific response probabilities. Because there are two random effects the quadrature must go over two dimensions to obtain the marginal probabilities. Table 6.13 (at the end of this section) lists SAS-IML code utilizing 10 quadrature points for each of the two dimensions. Using this SAS-IML code yields the estimated response probabilities given in Table 6.14.

Table 6.14 - Estimated response probabilities based on Table 6.10 results					
group	status	baseline	timepoint		
			6-months	12-months	24-months
control	street	.534	.207	.132	.107
	community	.373	.555	.543	.522
	independent	.094	.238	.325	.371
section 8	street	.403	.152	.115	.103
	community	.393	.314	.238	.204
	independent	.204	.535	.647	.692

The agreement of these with the observed marginal proportions in Table 6.2 is not very good. In particular, because the model specifies the trend to be linear across time (in terms of the logits), it is only response patterns that are approximately linear across time that are fit well by the model (*e.g.*, Control independent housing). Since many of the other patterns do not follow a linear trend across time, these are not fit well by the model (*e.g.*, Section 8 independent housing).

6.6 Section 6 MIXNO Output

Table 6.5 - output file for normally distributed random subject effects

MIXNO - The program for mixed-effects nominal logistic regression analysis

San Diego Homeless Project - Housing Outcome across time
 random intercepts model - SECTION8 & TIME EFFECTS

Random-effects distribution: normal

Numbers of observations

Level 1 observations = 1289

Level 2 observations = 361

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

```

4 4 4 4 4 4 4 4 4 1 3 4 3 4 4 1 4 4 4
4 4 4 4 4 4 4 3 4 4 1 4 2 1 4 4 4 3 3
4 4 4 4 4 4 4 4 4 4 4 1 4 1 3 4 4 3 4
4 4 4 4 4 1 4 3 3 4 3 4 4 4 4 3 4 4 3
4 4 4 4 4 4 4 4 4 4 3 4 4 4 1 1 4 4 4
4 2 4 4 4 4 4 4 4 4 4 4 4 4 4 1 4 4 4
4 1 4 4 3 4 4 4 4 4 4 4 4 4 4 4 4 4 3
3 3 4 4 4 4 4 4 4 4 4 2 3 4 4 4 4 4 4
4 2 2 3 4 4 4 4 4 4 3 4 4 4 4 4 4 4 4
3 4 4 3 4 4 3 3 4 4 4 3 4 4 4 4 4 3 4
4 4 2 2 4 4 4 4 4 4 4 4 4 4 4 4 1 3 2
4 4 4 4 4 4 4 4 4 4 1 4 3 2 3 2 4 4 3
4 3 4 4 4 4 4 3 3 3 3 3 4 4 4 4 4 4 4
4 1 4 4 4 3 4 4 4 4 4 1 4 4 4 4 4 4 4
4 3 4 4 4 4 4 4 3 4 4 3 4 4 4 4 3 4 1
4 1 2 4 4 4 4 4 4 4 4 4 4 4 4 1 4 4 4
4 4 4 4 4 4 3 3 4 4 4 4 4 2 4 4 4 1
3 4 4 4 2 2 4 1 4 2 3 1 4 4 4 4 1 2 4
1 4 4 4 3 4 4 4 2 3 4 4 3 2 4 4 4 2 4
    
```

Descriptive statistics for all variables

Variable	Minimum	Maximum	Mean	Std.Dev.
Housing	0.00000	2.00000	1.16835	0.77242
Intercpt	1.00000	1.00000	1.00000	0.00000
Section8	0.00000	1.00000	0.50970	0.50010
Time1	0.00000	1.00000	0.24981	0.43307
Time2	0.00000	1.00000	0.23507	0.42420
Time3	0.00000	1.00000	0.23507	0.42420
Sect8T1	0.00000	1.00000	0.12490	0.33074
Sect8T2	0.00000	1.00000	0.12180	0.32718
Sect8T3	0.00000	1.00000	0.12258	0.32808

Categories of the response variable Housing

Category	Frequency	Proportion
0.00	294.00	0.22808
1.00	484.00	0.37548
2.00	511.00	0.39643

Crosstabulation of variable Section8 by the response variable Housing

Section8	Housing			Total
	0.00	1.00	2.00	
0.00	161.0 (0.25)	305.0 (0.48)	166.0 (0.26)	632.0
1.00	133.0 (0.20)	179.0 (0.27)	345.0 (0.53)	657.0
Total	294.0	484.0	511.0	1289.0

Starting values

```

mean          -2.900 -1.261
covariates    -0.196 -0.647 -0.893 -0.967 -0.431 -0.285 -0.108 -0.216 -0.712 -0.982
covariates    -1.063 -0.474 -0.313 -0.119
var. terms    0.574 0.631

```

==> The number of level 2 observations with non-varying responses
= 78 (21.61 percent)

 * Final Results - Maximum Marginal Likelihood Estimates *

Total Iterations = 29
 Quad Pts per Dim = 20
 Log Likelihood = -1109.363
 Deviance (-2logL) = 2218.725
 Ridge = 0.200

Variable	Estimate	Stand.Error	Z	p-value
----------	----------	-------------	---	---------

RESPONSE CODE 1. vs CODE 0.

fixed effects

Intercpt	-0.45189	0.19230	-2.34991	0.01878 (2)
Section8	0.52099	0.26831	1.94171	0.05217 (2)
Time1	1.94152	0.31197	6.22350	0.00000 (2)
Time2	2.81975	0.46605	6.05035	0.00000 (2)
Time3	2.25949	0.37835	5.97201	0.00000 (2)
Sect8T1	-0.13527	0.49003	-0.27604	0.78252 (2)
Sect8T2	-1.91710	0.61096	-3.13785	0.00170 (2)
Sect8T3	-0.95229	0.53542	-1.77858	0.07531 (2)

random effect variance term: expressed as a standard deviation

Intercpt	0.87071	0.13764	6.32613	0.00000 (1)
----------	---------	---------	---------	-------------

RESPONSE CODE 2. vs CODE 0.

fixed effects

Intercpt	-2.67540	0.36749	-7.28030	0.00000 (2)
Section8	0.78108	0.49088	1.59118	0.11157 (2)
Time1	2.68209	0.42491	6.31213	0.00000 (2)
Time2	4.08805	0.55933	7.30887	0.00000 (2)
Time3	4.09897	0.46938	8.73268	0.00000 (2)
Sect8T1	2.00273	0.61389	3.26236	0.00111 (2)
Sect8T2	0.54838	0.69357	0.79065	0.42915 (2)
Sect8T3	0.30403	0.61473	0.49458	0.62090 (2)

random effect variance term: expressed as a standard deviation

Intercpt	2.33380	0.19642	11.88189	0.00000 (1)
----------	---------	---------	----------	-------------

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

(2) = 2-tailed p-value

Calculation of the intracluster correlation

residual variance = $\pi^2 / 3$ (assumed)

1 cluster variance = $(0.871 * 0.871) = 0.758$

intracluster correlation = $0.758 / (0.758 + (\pi^2/3)) = 0.187$

2 cluster variance = $(2.334 * 2.334) = 5.447$

intracluster correlation = $5.447 / (5.447 + (\pi^2/3)) = 0.623$

Table 6.10 - output file for trend analysis

MIXNO - The program for mixed-effects nominal logistic regression analysis

San Diego Homeless Project - Housing Outcome across time
 Random intercepts and slopes model - SECT8 & TIME EFFECTS

Random-effects distribution: normal

Numbers of observations

Level 1 observations = 1289

Level 2 observations = 361

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

```

4 4 4 4 4 4 4 4 4 1 3 4 3 4 4 1 4 4 4
4 4 4 4 4 4 4 4 3 4 4 1 4 2 1 4 4 4 3 3
4 4 4 4 4 4 4 4 4 4 4 1 4 1 3 4 4 3 4
4 4 4 4 4 1 4 3 3 4 3 4 4 4 4 3 4 4 3
4 4 4 4 4 4 4 4 4 4 4 3 4 4 4 1 1 4 4 4
4 2 4 4 4 4 4 4 4 4 4 4 4 4 4 1 4 4 4
4 1 4 4 3 4 4 4 4 4 4 4 4 4 4 4 4 4 3
3 3 4 4 4 4 4 4 4 4 4 2 3 4 4 4 4 4 4
4 2 2 3 4 4 4 4 4 4 3 4 4 4 4 4 4 4 4
3 4 4 3 4 4 3 3 4 4 4 3 4 4 4 4 4 3 4
4 4 2 2 4 4 4 4 4 4 4 4 4 4 4 1 3 2
4 4 4 4 4 4 4 4 4 4 1 4 3 2 3 2 4 4 3
4 3 4 4 4 4 4 3 3 3 3 3 4 4 4 4 4 4 4
4 1 4 4 4 3 4 4 4 4 4 1 4 4 4 4 4 4 4
4 3 4 4 4 4 4 4 3 4 4 3 4 4 4 4 3 4 1
4 1 2 4 4 4 4 4 4 4 4 4 4 4 4 1 4 4 4
4 4 4 4 4 4 3 3 4 4 4 4 4 2 4 4 4 4 1
3 4 4 4 2 2 4 1 4 2 3 1 4 4 4 4 1 2 4
1 4 4 4 3 4 4 4 2 3 4 4 3 2 4 4 4 2 4
    
```

Descriptive statistics for all variables

Variable	Minimum	Maximum	Mean	Std.Dev.
Housing	0.00000	2.00000	1.16835	0.77242
Intercpt	1.00000	1.00000	1.00000	0.00000
Time	0.00000	3.00000	1.42514	1.12944
Section8	0.00000	1.00000	0.50970	0.50010
Sec8Time	0.00000	3.00000	0.73623	1.08359

Categories of the response variable Housing

Category	Frequency	Proportion
0.00	294.00	0.22808
1.00	484.00	0.37548
2.00	511.00	0.39643

Crosstabulation of variable Time by the response variable Housing

Time	Housing			Total
	0.00	1.00	2.00	
0.00	180.0 (0.50)	136.0 (0.38)	45.0 (0.12)	361.0
1.00	45.0 (0.14)	138.0 (0.43)	139.0 (0.43)	322.0
2.00	32.0 (0.11)	108.0 (0.36)	163.0 (0.54)	303.0
3.00	37.0 (0.12)	102.0 (0.34)	164.0 (0.54)	303.0
Total	294.0	484.0	511.0	1289.0

Starting values

```

mean      -2.607 -0.968 0.000 0.000
covariates -0.363 -0.022 -0.399 -0.024
var. terms 1.814 0.000 0.907 1.995 0.000 1.088

```

=> The number of level 2 observations with non-varying responses
= 78 (21.61 percent)

```
-----
* Final Results - Maximum Marginal Likelihood Estimates *
-----
```

```
Total Iterations   =      112
Quad Pts per Dim   =       10
Log Likelihood      = -1093.582
Deviance (-2logL)  =   2187.163
Ridge               =       0.400
```

```
-----
Variable           Estimate      Stand.Error      Z           p-value
-----
```

```
RESPONSE CODE 1.  vs CODE 0.
-----
```

```
fixed effects
```

```
Intercpt          -0.38250      0.15906      -2.40467      0.01619 (2)
Time              2.12595      0.33459       6.35395      0.00000 (2)
Section8          0.33888      0.22849       1.48311      0.13805 (2)
Sec8Time         -0.41221      0.33492      -1.23078      0.21841 (2)
```

```
random effect variance & covariance terms:  cholesky of var-covariance matrix
```

```
Intercpt          0.44661      0.27370       1.63174      0.05137 (1)
covariance        1.10929      0.41051       2.70225      0.00689 (2)
Time              0.96018      0.44047       2.17989      0.01463 (1)
```

```
RESPONSE CODE 2.  vs CODE 0.
-----
```

```
fixed effects
```

```
Intercpt          -1.73350      0.22403      -7.73794      0.00000 (2)
Time              1.99880      0.42324       4.72258      0.00000 (2)
Section8          1.05881      0.29502       3.58898      0.00033 (2)
Sec8Time          0.93802      0.48430       1.93686      0.05276 (2)
```

```
random effect variance & covariance terms:  cholesky of var-covariance matrix
```

```
Intercpt          0.28216      0.28602       0.98650      0.16194 (1)
covariance        0.98178      0.53926       1.82060      0.06867 (2)
Time              2.63438      0.48839       5.39403      0.00000 (1)
```

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

Calculation of the random effects variance-covariance matrix

```
-----
1 Intercpt variance = (0.447 * 0.447) = 0.199
   covariance       = (0.447 * 1.109) = 0.495
   Time variance    = (1.109 * 1.109) + (0.960 * 0.960) = 2.152
```

Covariance expressed as a correlation = 0.756

```
2 Intercpt variance = (0.282 * 0.282) = 0.080
   covariance       = (0.282 * 0.982) = 0.277
   Time variance    = (0.982 * 0.982) + (2.634 * 2.634) = 7.904
```

Covariance expressed as a correlation = 0.349

```
-----
* Transforms of parameter estimates *
-----
```

Transpose of the Transform Matrix (parameters by transforms)

	1	2	3	4	5	6
1 Intercpt	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2 Time	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3 Intercpt	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
4 Time	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
5 Section8	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000
6 Sec8Time	1.0000	0.0000	2.0000	0.0000	3.0000	0.0000
7 Section8	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000
8 Sec8Time	0.0000	1.0000	0.0000	2.0000	0.0000	3.0000
9 VarCov1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
10 VarCov2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
11 VarCov3	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
12 VarCov1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
13 VarCov2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
14 VarCov3	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Transform	Estimate	Stand. Error	Z	p-value
1	-0.07334	0.35165	-0.20855	0.83480
2	1.99683	0.48224	4.14073	0.00003
3	-0.48555	0.64765	-0.74971	0.45343
4	2.93485	0.92042	3.18861	0.00143
5	-0.89777	0.96933	-0.92618	0.35435
6	3.87288	1.38956	2.78712	0.00532

note: p-values are 2-tailed

6.7 Section 6 SAS IML listings

Table 6.12 - computing marginal probabilities for Table 6.5 results

```

TITLE1 'San Diego Homeless Data - Estimated Marginal Probabilities';
PROC IML;
/* Results from MIXNO analysis - Table 6.5:  output file SDHOUSE.OUT */;

w0 = { 0 0 0 0 0 0 0,
       0 1 0 0 0 0 0,
       0 0 1 0 0 0 0,
       0 0 0 1 0 0 0};
w1 = { 1 0 0 0 0 0 0,
       1 1 0 0 1 0 0,
       1 0 1 0 0 1 0,
       1 0 0 1 0 0 1};
mua = {-0.45189};
sda = { 0.87071};
alphaa= { 0.52099, 1.94152, 2.81975, 2.25949, -0.13527, -1.91710, -0.95229};

mub = {-2.67540};
sdb = { 2.33380};
alphab= { 0.78108, 2.68209, 4.08805, 4.09897, 2.00273, 0.54838, 0.30403};

/* Now get the estimated marginal probabilities */;
/* number of quadrature points, quadrature nodes & weights */
nq = 10;
bq = { 3.8869246, 2.9630366, 2.0883447, 1.2426890, 0.4125905,
       -0.4125905, -1.2426890, -2.0883447, -2.9630366, -3.8869246};
aq = { .0002003, .0044289, .0386501, .1540833, .3026346,
       .3026346, .1540833, .0386501, .0044289, .0002003};

/* initialize to zero */
grp0 = J(4,1,0);
grp0a = J(4,1,0);
grp0b = J(4,1,0);
grp1 = J(4,1,0);
grp1a = J(4,1,0);
grp1b = J(4,1,0);

```

```

DO q = 1 to nq;
  za0 = w0*alphaa + mua + sda*bq[q];
  zb0 = w0*alphab + mub + sdb*bq[q];
  za1 = w1*alphaa + mua + sda*bq[q];
  zb1 = w1*alphab + mub + sdb*bq[q];

  grp0 = grp0 + 1 / (1+(EXP(za0)+EXP(zb0)))*aq[q];
  grp0a = grp0a + EXP(za0) / (1+(EXP(za0)+EXP(zb0)))*aq[q];
  grp0b = grp0b + EXP(zb0) / (1+(EXP(za0)+EXP(zb0)))*aq[q];

  grp1 = grp1 + 1 / (1+(EXP(za1)+EXP(zb1)))*aq[q];
  grp1a = grp1a + EXP(za1) / (1+(EXP(za1)+EXP(zb1)))*aq[q];
  grp1b = grp1b + EXP(zb1) / (1+(EXP(za1)+EXP(zb1)))*aq[q];

END;

print 'Quadrature method - 10 points';
print 'marginal probability for group 0 - category 0' grp0 [FORMAT=8.4];
print 'marginal probability for group 0 - category 1' grp0a [FORMAT=8.4];
print 'marginal probability for group 0 - category 2' grp0b [FORMAT=8.4];
print 'marginal probability for group 1 - category 0' grp1 [FORMAT=8.4];
print 'marginal probability for group 1 - category 1' grp1a [FORMAT=8.4];
print 'marginal probability for group 1 - category 2' grp1b [FORMAT=8.4];

```

Table 6.14 - computing marginal probabilities for Table 6.10 results

```

TITLE1 'San Diego Homeless Data - Estimated Marginal Probabilities';
PROC IML;
/* Results from MIXNO analysis - Table 6.10:  output file SDHOUSET.OUT */;
x   = { 1 0,
        1 1,
        1 2,
        1 3};
w0  = { 0 0,
        0 0,
        0 0,
        0 0};
w1  = { 1 0,
        1 1,
        1 2,
        1 3};
mua = {- .38250, 2.12595};
sda = { .44661 0, 1.10929 .96018};
alphaa= { .33888, -.41221};

mub = {-1.73350, 1.99880};
sdb = { .028216 0, .98178 2.63438};
alphab= { 1.05881, .93802};

/* Now get the estimated marginal probabilities */;
/* number of quadrature points, quadrature nodes & weights */
nq = 10;
bq = { 3.8869246, 2.9630366, 2.0883447, 1.2426890, 0.4125905,
       -0.4125905, -1.2426890, -2.0883447, -2.9630366, -3.8869246};
aq = { .0002003, .0044289, .0386501, .1540833, .3026346,
       .3026346, .1540833, .0386501, .0044289, .0002003};

/* initialize to zero */
grp0  = J(4,1,0);
grp0a = J(4,1,0);
grp0b = J(4,1,0);
grp1  = J(4,1,0);
grp1a = J(4,1,0);
grp1b = J(4,1,0);

```

```

DO q1 = 1 to nq;
DO q2 = 1 to nq;
  quadvec = bq[q1]/bq[q2];
  za0 = w0*alphaa + x*mua + x*sda*quadvec;
  zb0 = w0*alphab + x*mub + x*sdb*quadvec;
  za1 = w1*alphaa + x*mua + x*sda*quadvec;
  zb1 = w1*alphab + x*mub + x*sdb*quadvec;

  grp0 = grp0 + 1 / (1+(EXP(za0)+EXP(zb0)))*aq[q1]*aq[q2];
  grp0a = grp0a + EXP(za0) / (1+(EXP(za0)+EXP(zb0)))*aq[q1]*aq[q2];
  grp0b = grp0b + EXP(zb0) / (1+(EXP(za0)+EXP(zb0)))*aq[q1]*aq[q2];

  grp1 = grp1 + 1 / (1+(EXP(za1)+EXP(zb1)))*aq[q1]*aq[q2];
  grp1a = grp1a + EXP(za1) / (1+(EXP(za1)+EXP(zb1)))*aq[q1]*aq[q2];
  grp1b = grp1b + EXP(zb1) / (1+(EXP(za1)+EXP(zb1)))*aq[q1]*aq[q2];
END;
END;

print 'Quadrature method - 10 points per dimension';
print 'marginal probability for group 0 - category 0' grp0 [FORMAT=8.4];
print 'marginal probability for group 0 - category 1' grp0a [FORMAT=8.4];
print 'marginal probability for group 0 - category 2' grp0b [FORMAT=8.4];
print 'marginal probability for group 1 - category 0' grp1 [FORMAT=8.4];
print 'marginal probability for group 1 - category 1' grp1a [FORMAT=8.4];
print 'marginal probability for group 1 - category 2' grp1b [FORMAT=8.4];

```

7 Latent trait model for item responses

This illustration uses data from Clogg [32] on responses to three life satisfaction questions. 1472 subjects from the 1975 US General Household Survey indicated their degree of satisfaction with family (F), hobbies (H), and residence (R) on a three-point scale (1=low, 2=medium, or 3=high). These data have been further described and analyzed by Masters [48] and Bartholomew [49]; the observed frequencies that are presented in Table 7.1 can be found in either of these sources.

response pattern		observed frequency	response pattern		observed frequency	response pattern		observed frequency		
F	H	R	F	H	R	F	H	R		
1	1	1	15	1	2	1	3	1	5	
1	1	2	11	1	2	2	1	3	2	14
1	1	3	7	1	2	3	1	3	3	16
2	1	1	16	2	2	1	2	3	1	18
2	1	2	26	2	2	2	2	3	2	38
2	1	3	12	2	2	3	2	3	3	27
3	1	1	23	3	2	1	3	3	1	64
3	1	2	49	3	2	2	3	3	2	191
3	1	3	54	3	2	3	3	3	3	466

The observed marginal response proportions and logits for the three items (x-axis) are plotted in Figures 7.1 and 7.2, respectively. The fitted response proportions and logits are also presented in the figures; it will be explained later how these were obtained. For all items the response proportions increase with the satisfaction level, however the proportion of high satisfaction responses for the Family item is very high. Figure 7.2 expresses these proportions in terms of the two logits that are modeled by the nominal logistic regression, namely, $\log[P(\text{high})/P(\text{low})]$ and $\log[P(\text{med})/P(\text{low})]$. These logits indicate the higher proportion of high and medium satisfaction responses relative to low satisfaction responses for all three items. Again, the pronounced level of high satisfaction, relative to low satisfaction, is noted for the Family item.

Figure 7.1 – Satisfaction marginal response proportions by Item

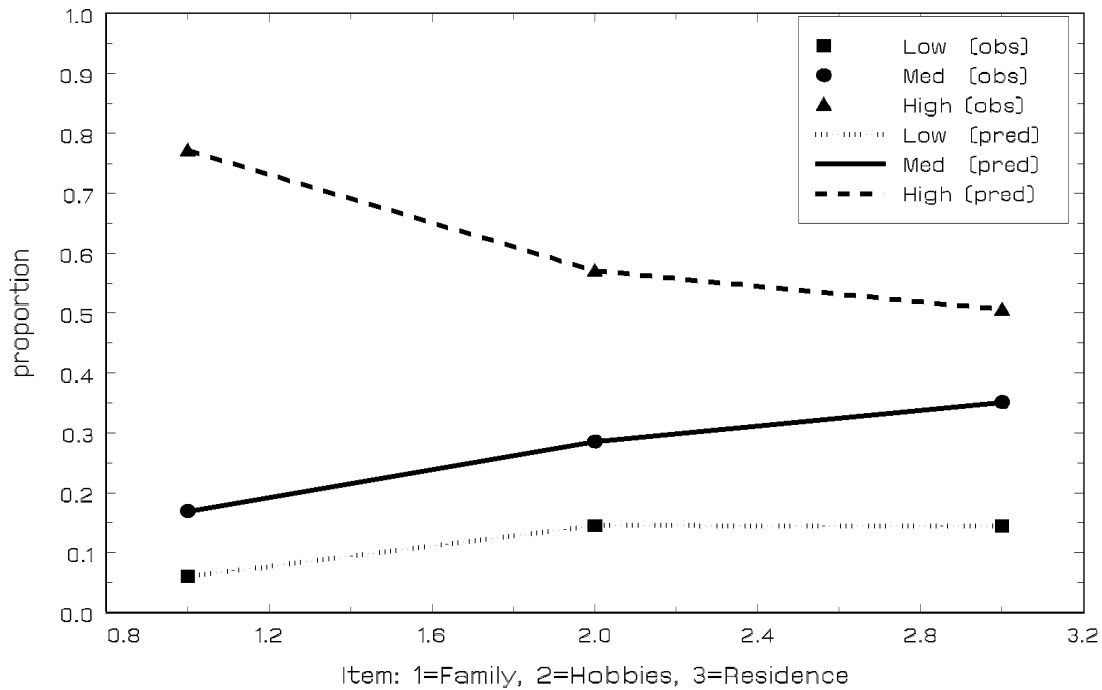
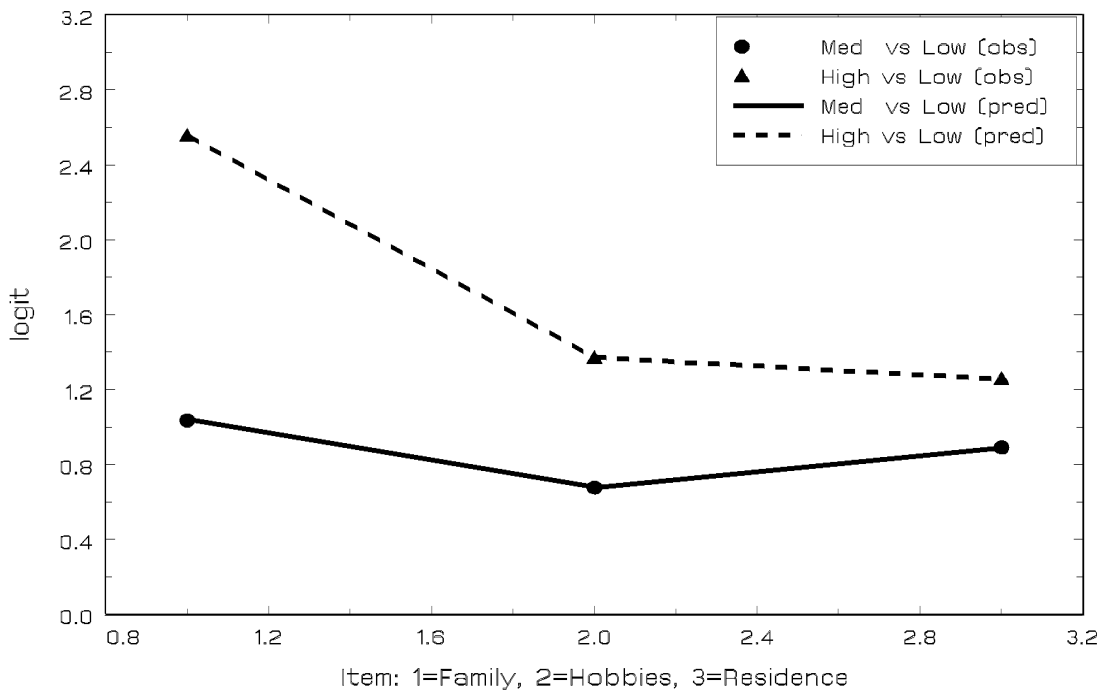


Figure 7.2 – Satisfaction marginal logits by Item



To analyze these data using MIXNO, the WT option is used to input the data of the $3^3 = 27$ response patterns. With three questions per response pattern, the data for each pattern occupies

three physical lines in the dataset to be created for MIXNO use. Each line includes the response for one of the three questions. The partial list of the resulting dataset is given in Table 7.2.

1	1	1	1	0	0	15
1	1	1	0	1	0	15
1	1	1	0	0	1	15
2	1	1	1	0	0	11
2	1	1	0	1	0	11
2	2	1	0	0	1	11
3	1	1	1	0	0	7
3	1	1	0	1	0	7
3	3	1	0	0	1	7
4	2	1	1	0	0	16
4	1	1	0	1	0	16
4	1	1	0	0	1	16
5	2	1	1	0	0	26
5	1	1	0	1	0	26
5	2	1	0	0	1	26
6	2	1	1	0	0	12
6	1	1	0	1	0	12
6	3	1	0	0	1	12
7	3	1	1	0	0	23
7	1	1	0	1	0	23
7	1	1	0	0	1	23
8	3	1	1	0	0	49
8	1	1	0	1	0	49
8	2	1	0	0	1	49
9	3	1	1	0	0	54
9	1	1	0	1	0	54
9	3	1	0	0	1	54

The variables are, in order, pattern number (from 1 to 27), degree of satisfaction (from 1 to 3), a column of ones for the intercept, three item indicator variables, and the response pattern frequency. The three item indicator variables (say, x_l , $l = 1, 2, 3$) are coded equal to 1 if $j = l$ and 0 otherwise, where j denotes the level-1 item responses that are nested within subjects (level-2). Unlike our previous examples, here the number of nested level-1 observations is constant (*i.e.*, = 3) across the level-2 units. MIXNO does not require this, and so incomplete response patterns can be included in the analysis.

In the psychometric literature, random or mixed-effects models are often termed “latent trait” models. For these data, MIXNO will first be used to fit a latent trait model for nominal responses as described in Bock [25]. In Bock’s model a single latent trait (*i.e.*, random effect) is assumed, however it’s influence is allowed to vary across items and categories of the nominal items. Allowing the influence of the random effect to vary across items can be accomplished in MIXNO using the VGRP option. For this, the three item indicator variables are all designated as random effects (*i.e.*,

R=3), however VGRP=1 is also specified to indicate that they represent grouping variables. To allow the influence of the random effect to vary across the categories of the nominal responses, VCAT=1 is also specified. Table 7.3 lists the MIXNO.DEF file to be used to estimate this latent trait model for nominal responses.

Table 7.3 - MIXNO.DEF file for Bock nominal latent trait model			
Clogg - life satisfaction data - 3 items & 3 categories			
nominal IRT model			
lifesat.dat			
lifesat.OUT			
lifesat.def			
1	7	3	0
0.0001	3	0	0
1	0	0	0
20	0	0	0
1	1		
1	2		
4	5	6	
7			
1	2	3	
LifeSat			
Item1	Item2	Item3	

Table 7.4 list the results of this analysis. Because the WT option is selected, MIXNO indicates the number of patterns and observations at each of the two levels. The level-2 patterns are the 27 response patterns, the 1472 level-2 observations equal the number of subjects, and the 4416 level-1 observations equal the total number of responses (*i.e.*, 3 responses from 1472 subjects equals 4416). MIXNO then indicates that three responses were observed for each of the 27 response patterns. Prior to listing the estimates of the model, MIXNO indicates that three of the response patterns consisted of non-varying level-1 responses (*i.e.*, 111, 222, and 333); this equals $3/27 = .1111$ or 11.11 percent of the response patterns.

The estimates for the six random effect variance terms (3 items by 2 category comparisons) produced by MIXNO agree almost exactly with those reported in Bartholomew [49] on page 172. As noted by Bartholomew, there is an anomaly in that there is a negative estimate for the first random effect term of the Family item (*i.e.*, the term comparing responses between categories 2 and 1). However, because the standard error of this term greatly exceeds the estimate, it is not significantly different than zero, and so we can designate this variance term to be zero.

Turning to the fixed effects, we see that all are highly significant indicating the increased probability of medium and high satisfaction responses, relative to low satisfaction, for all items. The difference is especially pronounced for the Family item when contrasting high to low satisfaction

responses, where the estimate equals 3.03; this agrees with our previous observations based on Figures 1 and 2. To aid in interpretation of the fixed effects, Bartholomew uses the following parameterization

$$\pi_{jk} = \frac{\exp \mu_{jk}}{\sum_{k=0}^K \exp \mu_{jk}} \quad (17)$$

where $k = 0, 1, \dots, K$ response categories and j denotes the (level-1) items. Additionally, the reference category parameter μ_{j0} is set equal to zero. This yields estimates $\hat{\pi}_{jk}$ that indicate the response probability for a person at the median (in terms of the underlying latent satisfaction distribution). Using this parameterization and the estimates produced by MIXNO yields the following results in Table 7.5.

Table 7.5 - Bartholomew parameterization of MIXNO estimates				
	category 0	category 1	category 2	total
Item 1	$\mu_{10} = 0$	$\hat{\mu}_{11} = .9623$	$\hat{\mu}_{12} = 3.0349$	
exp()	1	2.6177	20.7990	24.4170
$\pi_{1k} = \frac{\exp(\mu_{1k})}{\sum \exp(\mu_{1k})}$.041	.107	.852	
Item 2	$\mu_{20} = 0$	$\hat{\mu}_{21} = .9210$	$\hat{\mu}_{22} = 1.6355$	
exp()	1	2.5118	5.1320	8.6438
$\pi_{2k} = \frac{\exp(\mu_{2k})}{\sum \exp(\mu_{2k})}$.116	.291	.594	
Item 3	$\mu_{30} = 0$	$\hat{\mu}_{31} = 1.0853$	$\hat{\mu}_{32} = 1.4127$	
exp()	1	2.9603	4.1070	8.0674
$\pi_{3k} = \frac{\exp(\mu_{3k})}{\sum \exp(\mu_{3k})}$.124	.367	.509	

These re-expressions of the estimated MIXNO parameters agree exactly with those reported in Bartholomew (page 172), and as noted by Bartholomew, indicate that a median respondent has a greater probability (.85) of being highly satisfied on the Family item than on either of the other two items (.59 and .51, respectively).

Returning to the random-effect variance terms, notice that, for a given category, these are very similar. As noted, the only exception is the negative random-effect variance term for the first item when comparing code 2 vs. code 1 (*i.e.*, medium vs. low). Thus, a simpler model that does not allow the random-effects variance terms to vary across items might be reasonable for these data. For nominal data, such models are generalizations of the Rasch model for dichotomous data. In this vein, Masters [48] describes a model where the influence of the random subject effect does not

vary across items. And while the influence of the random effect does vary across categories, it does so in a predefined way. Specifically, in Masters' model the influence of the random subject effect is fixed to be $k = 0, 1, \dots, K$, respectively, for each of the $K + 1$ successive outcome categories. Here, we will fit a model where the influence of the random subject effect does not vary across items, but its varying influence across categories will be estimated. For this, Table 7.6 lists the MIXNO.DEF file.

Table 7.6 - MIXNO.DEF file for Rasch-type nominal latent trait model																
Clogg - life satisfaction data - 3 items & 3 categories																
nominal IRT model - Rasch-type model																
lifesat.dat																
lifesat1.out																
lifesat1.def																
1	7	1	3	0.0001	3	0	0	1	0	0	20	0	0	1	0	1
1 2																
3																
4 5 6																
7																
1 2 3																
LifeSat																
Intercpt																
Item1	Item2	Item3														

This model simply specifies a random intercept that varies across categories (*i.e.*, VCAT=1). Additionally, the item indicator variables are specified as fixed covariates.

Results for this analysis are given in Table 7.7. A likelihood-ratio test comparing this model to the previous one yields $\chi_4^2 = 7484.464 - 7478.817 = 5.647$ which is not statistically significant at even the .10 level. Thus, the simpler model with the assumption of equal random-effect variances across items is reasonable. Though the estimates change somewhat (mostly for the Family item), the conclusions based on this model are the same as those for the previous model. Based on these estimates, the response probability curves in Figure 7.3 were generated. These curves indicate the response probabilities across values of the underlying latent satisfaction level (*i.e.*, the random subject effect) for each of the three items. For each item j , these were generated using formulas

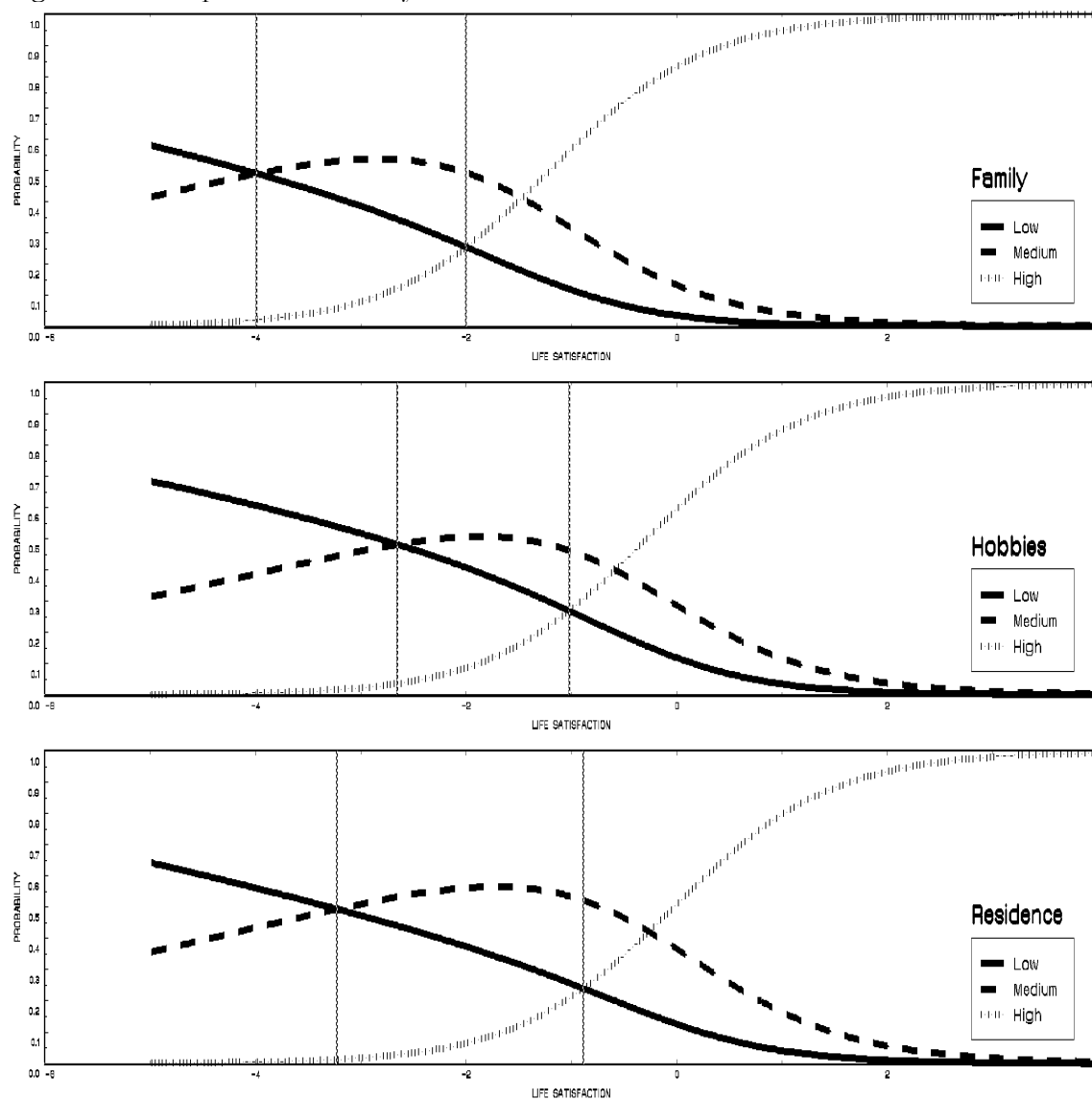
$$P_{ij0} = \frac{1}{1 + \exp(\sigma_1\theta_i + \alpha_{j1}) + \exp(\sigma_2\theta_i + \alpha_{j2})} \quad (18)$$

$$P_{ij1} = \frac{\exp(\sigma_1\theta_i + \alpha_{j1})}{1 + \exp(\sigma_1\theta_i + \alpha_{j1}) + \exp(\sigma_2\theta_i + \alpha_{j2})} \quad (19)$$

$$P_{i,j2} = \frac{\exp(\sigma_2\theta_i + \alpha_{j2})}{1 + \exp(\sigma_1\theta_i + \alpha_{j1}) + \exp(\sigma_2\theta_i + \alpha_{j2})} \quad (20)$$

and substituting the parameter estimates and values of the random effect θ . The vertical lines in the figures indicate the value of θ where the probability of a response in the low category equals the medium and high category, respectively. These are obtained as simple re-expressions of the estimated parameters, namely $-\hat{\alpha}_{j1}/\hat{\sigma}_1$ and $-\hat{\alpha}_{j2}/\hat{\sigma}_2$, respectively.

Figure 7.3 - Response Probability Curves



As noted by Masters [48], who presents a similar figure, the curves show that the probability of responding "low" is modeled to decrease with increasing life satisfaction (*i.e.*, the random effect).

Similarly, the probability of responding "high" is modeled to increase with increasing life satisfaction, and the probability of responding "medium" is modeled to increase and then decrease. The curves also show that the Family item is unique in that even at relatively low levels of life satisfaction, say -1, the "high" response is the most probable response.

To get the estimated marginal probabilities that are presented in Figures 7.1 and 7.2, some additional work is required. Using equations (18)-(20) yields the response probabilities for particular values of the random effect θ (*i.e.*, the subject-specific probabilities mentioned in Section 6.5). Again, these subject-specific probabilities can be used to obtain marginal probabilities by performing numerical integration. Table 7.8 (at the end of this section) contains a listing of SAS-IML code that implements a 10 point quadrature solution for this.

7.1 Estimation of θ

As mentioned in Section 3, MIXNO provides empirical Bayes estimates of the random subject effects θ_i . Upon convergence, these estimates are written to the MIXNO.RES file. Table 7.9 lists these estimates for both models considered here. The empirical Bayes estimates of the random effects are described as "expected a posteriori" (EAP) estimates by Bock and Aitkin [29] because they are obtained as the mean of the posterior distribution of θ given the nominal outcomes \mathbf{y} . The analogous standard deviation of the posterior distribution (PSD) provides information about the degree of precision in the estimation of θ . Also, the columns labeled Bock and Rasch in Table 7.9 refer to the first and second models fit to these data, respectively (*i.e.*, Tables 7.4 and 7.7).

Interestingly, Bock's model gives the lowest EAP score to pattern 211 rather than 111 (and a lower score for 212 rather than 112, etc.). This is due to the negative variance estimate that was obtained under this model for the Family item, response code 2 vs. 1. However, since the estimates for these two patterns are almost the same, this has little practical significance. Because the Rasch model assumes homogeneity of the random-effects variance terms across items, it yields identical EAP and PSD estimates for patterns that share the same response values (*e.g.*, 211, 112, and 121; or 212, 221, and 122).

Finally, the last column of Table 7.9 lists the latent class assignments reported in Bartholomew [49] for a three class solution. It is interesting to compare the estimated ranking based on the latent trait models to the assignment based on the latent class analysis. To facilitate this, the table is broken at two points where gaps in the EAP estimates are largest. As can be seen, the agreement

between the latent classes and latent trait estimates is very close, especially for class III. However, the latent trait models do provide finer groupings of subjects than the latent class model as well as information about the relative importance of the items to the underlying latent variable. Further discussion about the similarities and differences of these two methods can be found in Bartholomew [49] and Masters [48].

response			observed frequency	EAP		PSD		Latent class
F	H	R		Rasch	Bock	Rasch	Bock	
2	1	1	16	-1.422	-1.535	0.723	0.716	I
1	1	1	15	-1.599	-1.490	0.736	0.712	I
2	1	2	26	-1.251	-1.351	0.712	0.702	I
2	2	1	23	-1.251	-1.332	0.712	0.701	II
1	1	2	11	-1.422	-1.308	0.723	0.699	I
1	2	1	3	-1.422	-1.289	0.723	0.698	II
2	2	2	58	-1.085	-1.155	0.703	0.691	II
1	2	2	12	-1.251	-1.113	0.712	0.689	II
2	1	3	12	-0.641	-0.816	0.691	0.680	I
2	3	1	18	-0.641	-0.793	0.691	0.680	II
1	1	3	7	-0.800	-0.776	0.693	0.680	II
1	3	1	5	-0.800	-0.752	0.693	0.680	II
3	1	1	23	-0.800	-0.748	0.693	0.680	II
2	2	3	31	-0.483	-0.629	0.690	0.680	II
2	3	2	38	-0.483	-0.624	0.690	0.680	II
1	2	3	5	-0.641	-0.589	0.691	0.681	II
1	3	2	14	-0.641	-0.583	0.691	0.681	II
3	1	2	49	-0.641	-0.579	0.691	0.681	II
3	2	1	45	-0.641	-0.561	0.691	0.681	II
3	2	2	117	-0.483	-0.390	0.690	0.686	II
2	3	3	27	0.127	-0.091	0.712	0.700	III
1	3	3	16	-0.039	-0.047	0.703	0.703	III
3	1	3	54	-0.039	-0.043	0.703	0.703	III
3	3	1	64	-0.039	-0.018	0.703	0.705	III
3	2	3	126	0.127	0.161	0.712	0.718	III
3	3	2	191	0.127	0.167	0.712	0.718	III
3	3	3	466	0.809	0.793	0.770	0.776	III

7.2 Section 7 MIXNO Output

Table 7.4 - output file for Bock nominal latent trait model

MIXNO - The program for mixed-effects nominal logistic regression analysis

Clogg - life satisfaction data - 3 items & 3 categories
nominal IRT model

Random-effects distribution: normal

Numbers of observations

Level 1 observations = 4416.00
Level 1 patterns = 81

Level 2 observations = 1472.00
Level 2 patterns = 27

The number of level 1 patterns per level 2 pattern are:

3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
3 3 3 3 3 3 3 3

Descriptive statistics for all variables

Variable	Minimum	Maximum	Mean	Std.Dev.
LifeSat	1.00000	3.00000	2.49932	0.69458
Item1	0.00000	1.00000	0.33333	0.47146
Item2	0.00000	1.00000	0.33333	0.47146
Item3	0.00000	1.00000	0.33333	0.47146

Categories of the response variable LifeSat

Category	Frequency	Proportion
1.00	513.00	0.11617
2.00	1185.00	0.26834
3.00	2718.00	0.61549

Starting values

mean		-1.388	0.170	0.453	0.554	0.000	0.000
var. terms		1.814	0.907	0.907	1.995	1.088	1.088

=> The number of level 2 patterns with non-varying responses
 = 3 (11.11 percent)

 * Final Results - Maximum Marginal Likelihood Estimates *

Total Iterations = 37
 Quad Pts per Dim = 20
 Log Likelihood = -3739.408
 Deviance (-2logL) = 7478.817
 Ridge = 0.100

Variable	Estimate	Stand. Error	Z	p-value
----------	----------	--------------	---	---------

RESPONSE CODE 2. vs CODE 1.

fixed effects

Item1	0.96228	0.26303	3.65838	0.00025 (2)
Item2	0.92095	0.14084	6.53878	0.00000 (2)
Item3	1.08525	0.12659	8.57314	0.00000 (2)

random effect variance & covariance terms: cholesky of var-covariance matrix

Item1	-0.08825	0.26782	-0.32953	0.74175 (2)
Item2	0.40460	0.17429	2.32146	0.02026 (2)
Item3	0.36614	0.16066	2.27892	0.02267 (2)

RESPONSE CODE 3. vs CODE 1.

fixed effects

Item1	3.03490	0.22838	13.28902	0.00000 (2)
Item2	1.63554	0.13300	12.29763	0.00000 (2)
Item3	1.41271	0.12244	11.53793	0.00000 (2)

random effect variance & covariance terms: cholesky of var-covariance matrix

Item1	1.55194	0.29205	5.31391	0.00000 (2)
Item2	1.54380	0.21996	7.01849	0.00000 (2)
Item3	1.49291	0.21146	7.06004	0.00000 (2)

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

Table 7.7 - output file for Rasch-type nominal latent trait model

MIXNO - The program for mixed-effects nominal logistic regression analysis

Clogg - life satisfaction data - 3 items & 3 categories

nominal IRT model - Rasch-type model

Random-effects distribution: normal

Numbers of observations

Level 1 observations = 4416.00

Level 1 patterns = 81

Level 2 observations = 1472.00

Level 2 patterns = 27

The number of level 1 patterns per level 2 pattern are:

3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
3 3 3 3 3 3 3 3

Descriptive statistics for all variables

Variable	Minimum	Maximum	Mean	Std.Dev.
LifeSat	1.00000	3.00000	2.49932	0.69458
Intercpt	1.00000	1.00000	1.00000	0.00000
Item1	0.00000	1.00000	0.33333	0.47146
Item2	0.00000	1.00000	0.33333	0.47146
Item3	0.00000	1.00000	0.33333	0.47146

Categories of the response variable LifeSat

Category	Frequency	Proportion
1.00	513.00	0.11617
2.00	1185.00	0.26834
3.00	2718.00	0.61549

Starting values

covariates	0.000	0.000	0.000	0.000	0.000	0.000
var. terms	0.574	0.631				

=> The number of level 2 patterns with non-varying responses
 = 3 (11.11 percent)

 * Final Results - Maximum Marginal Likelihood Estimates *

Total Iterations = 13
 Quad Pts per Dim = 20
 Log Likelihood = -3742.232
 Deviance (-2logL) = 7484.464
 Ridge = 0.000

Variable	Estimate	Stand.Error	Z	p-value
----------	----------	-------------	---	---------

 RESPONSE CODE 2. vs CODE 1.

fixed effects

Item1	1.32651	0.16055	8.26236	0.00000 (2)
Item2	0.88231	0.10995	8.02481	0.00000 (2)
Item3	1.07369	0.10383	10.34058	0.00000 (2)

random effect variance term: expressed as a standard deviation

Intercpt	0.33254	0.09742	3.41331	0.00032 (1)
----------	---------	---------	---------	-------------

 RESPONSE CODE 3. vs CODE 1.

fixed effects

Item1	3.16593	0.15017	21.08222	0.00000 (2)
Item2	1.61479	0.11019	14.65431	0.00000 (2)
Item3	1.40294	0.10940	12.82365	0.00000 (2)

random effect variance term: expressed as a standard deviation

Intercpt	1.58237	0.10337	15.30775	0.00000 (1)
----------	---------	---------	----------	-------------

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

(2) = 2-tailed p-value

Calculation of the intracluster correlation

residual variance = $\pi \cdot \pi / 3$ (assumed)

1 cluster variance = $(0.333 * 0.333) = 0.111$

intracluster correlation = $0.111 / (0.111 + (\pi \cdot \pi / 3)) = 0.033$

2 cluster variance = $(1.582 * 1.582) = 2.504$

intracluster correlation = $2.504 / (2.504 + (\pi \cdot \pi / 3)) = 0.432$

7.3 Section 7 SAS IML listing

Table 7.8 - computing marginal probabilities for Table 7.7 results

```

TITLE1 'Satisfaction Data - Estimated Marginal Probabilities';
PROC IML;
/* Results from MIXNO analysis - output file: LIFESAT1.OUT */;
w = { 1 0 0, 0 1 0, 0 0 1 };
alpha1 = {1.32651, 0.88231, 1.07369};
sigma1 = .33254;
alpha2 = {3.16593, 1.61479, 1.40294};
sigma2 = 1.58237;
/* Now get the estimated marginal probabilities */;
/* number of quadrature points, quadrature nodes, and weights */;
nq = 10;
bq = { 3.8869246, 2.9630366, 2.0883447, 1.2426890, 0.4125905,
      -0.4125905, -1.2426890, -2.0883447, -2.9630366, -3.8869246};
aq = { .0002003, .0044289, .0386501, .1540833, .3026346,
      .3026346, .1540833, .0386501, .0044289, .0002003};
/* initialize to zero */;
mprb0 = J(3,1,0);
mprb1 = J(3,1,0);
mprb2 = J(3,1,0);
DO q = 1 to nq;
    z1 = sigma1*bq[q] + w*alpha1;
    z2 = sigma2*bq[q] + w*alpha2;
    mprb0 = mprb0 + ( 1.0 / (1.0 + EXP(z1) + EXP(z2)))*aq[q];
    mprb1 = mprb1 + (EXP(z1) / (1.0 + EXP(z1) + EXP(z2)))*aq[q];
    mprb2 = mprb2 + (EXP(z2) / (1.0 + EXP(z1) + EXP(z2)))*aq[q];
END;
/* compute logits */;
mlogit1 = log ( mprb1 / mprb0);
mlogit2 = log ( mprb2 / mprb0);
/* print out results */;
print 'Quadrature method - 10 points';
print 'marginal probability for category 0', mprb0 [FORMAT=8.4];
print 'marginal probability for category 1', mprb1 [FORMAT=8.4];
print 'marginal probability for category 2', mprb2 [FORMAT=8.4];
print 'marginal logit for cat 1 vs cat 0', mlogit1 [FORMAT=8.4];
print 'marginal logit for cat 2 vs cat 0', mlogit2 [FORMAT=8.4];

```

8 Hardware and software specifications

MIXNO is written in standard FORTRAN-77 with double arithmetic precision and requires a math coprocessor. All necessary matrices and vectors are stored in a single one-dimensional array. There are no fixed limitations on the numbers of level-2 units, level-1 units, or model variables. MIXNO utilizes some MATCAL subroutines [52] for matrix algebra operations.

9 Availability

The MIXNO program is available at no charge and can be downloaded from the author's website at <http://www.uic.edu/~hedeker/mix.html>. Specifically, a ZIP file containing the program, manual, and example dataset is provided at this website. Any comments regarding program usage can be e-mailed to the author at hedeker@uic.edu.

10 Acknowledgements

Thanks are in order to Drs. R. Darrell Bock, Robert D. Gibbons, and Ann Hohmann for many valuable discussions, comments, and suggestions that contributed in the preparation of this program. The author is also grateful to Dr. Bock for providing the code for MATCAL [52] subroutines, to Dr. Brian Flay for providing the clustered dataset for the example, and to Drs. Richard Hough and Michael Hurlburt for use of the longitudinal data and for helpful comments in their analysis. A special thanks is due to Dave Patterson and Discerning Systems, Inc., for their work on the WINDOWS interface for MIXNO. The development of MIXNO was supported by the National Institutes of Mental Health Grant MH56146-01.

References

- [1] D.R. Cox, *Analysis of Binary Data* (Chapman and Hall, London, 1970).
- [2] R.D. Bock, Estimating multinomial response relations, in *Contributions to Statistics and Probability*, R.C. Bose (ed.), pp. 453-479 (Univ. of North Carolina Press, Chapel Hill, NC, 1970).
- [3] M. Nerlove and S.J. Press, *Univariate and multivariate log-linear and logistic models* (Rand Corporation Technical Report R-1306-EDA/NIH, Santa Monica, CA, 1973).
- [4] R.L. Plackett, *The Analysis of Categorical Data* (Griffin, London, 1974).
- [5] A. Agresti, *Categorical Data Analysis* (Wiley, New York, 1990).
- [6] H. Goldstein, *Multilevel Statistical Models* (Halsted Press, New York, 1995).
- [7] A.S. Bryk and S.W. Raudenbush, *Hierarchical Linear Models: Applications and Data Analysis Methods* (Sage Publications, Inc., Thousand Oaks CA, 1992).
- [8] N.T. Longford, *Random Coefficient Models*, (Oxford University Press, New York, 1993).
- [9] I.G.G. Kreft and J. de Leeuw, *Introducing Multilevel Modeling* (Sage Publications, Inc., Thousand Oaks CA, 1998).
- [10] R. Stiratelli, N.M. Laird, and J.H. Ware, Random-effects models for serial observations with binary response, *Biometrics* 40, (1984) 961-971.
- [11] D.A. Anderson and M. Aitkin, Variance component models with binary response: interviewer variability, *Journal of the Royal Statistical Society Series B* 47 (1985) 203-210.
- [12] G.Y. Wong and W.M. Mason, The hierarchical logistic regression model for multilevel analysis, *Journal of the American Statistical Association* 80, (1985) 513-524.
- [13] R.D. Gibbons & R.D. Bock, Trend in correlated proportions, *Psychometrika* 52,(1987)113-124.
- [14] M.R. Conaway, Analysis of repeated categorical measurements with conditional likelihood methods, *Journal of the American Statistical Association* 84, (1989) 53-61.
- [15] H. Goldstein, Nonlinear multilevel models, with an application to discrete response data, *Biometrika* 78, (1991) 45-51.

- [16] D.A. Harville and R.W. Mee, A mixed-model procedure for analyzing ordered categorical data, *Biometrics* 40, (1984) 393-408.
- [17] J. Jansen, On the statistical analysis of ordinal data when extravariation is present, *Applied Statistics* 39, (1990) 75-84.
- [18] F. Ezzet and J. Whitehead, A random effects model for ordinal responses from a crossover trial, *Statistics in Medicine* 10, (1991) 901-907.
- [19] D. Hedeker and R.D. Gibbons, A random-effects ordinal regression model for multilevel analysis, *Biometrics* 50, (1994) 933-944.
- [20] T.R. Ten Have, A mixed effects model for multivariate ordinal response data including correlated discrete failure times with ordinal responses, *Biometrics* 52, (1996) 473-491.
- [21] J. Rasbash, M. Wang, G. Woodhouse, and H. Goldstein, *MLn: command reference guide* (Institute of Education, University of London, London UK, 1995).
- [22] G. Rodriguez and N. Goldman, An assessment of estimation procedures for multilevel models with binary response, *Journal of the Royal Statistical Society, Series A*, 158, (1995) 73-89.
- [23] X. Lin & N.E. Breslow, Bias correction in generalized linear mixed models with multiple components of dispersion, *Journal of the American Statistical Association*, 91,(1996)1007-1016.
- [24] D. Hedeker, A mixed-effects multinomial logistic regression model, under review.
- [25] R. D. Bock, Estimating item parameters and latent ability when responses are scored in two or more nominal categories, *Psychometrika* 37, (1972) 29-51.
- [26] J.R. Magnus, *Linear Structures* (Charles Griffin, London UK, 1988)
- [27] A. Wald, Tests of statistical hypotheses concerning several parameters when the number of observations is large, *Transactions of the American Mathematical Society* 54, (1943) 426-482.
- [28] A.H. Stroud and D. Sechrest, *Gaussian Quadrature Formulas* (Prentice Hall, Englewood Cliffs, NJ, 1966).
- [29] R.D. Bock and M. Aitkin, Marginal maximum likelihood estimation of item parameters: an application of the EM algorithm, *Psychometrika* 46, (1981) 443-459.

- [30] R.D. Bock, R.D. Gibbons, and E. Muraki, Full-information item factor analysis, *Applied Psychological Measurement* 12, (1988) 261-280.
- [31] S.D. Silvey, *Statistical Inference* (Chapman and Hall, New York, 1975).
- [32] C.C. Clogg, Some latent structure models for the analysis of Likert-type data, *Social Science Research* 8, (1979), 287-301.
- [33] D. Hedeker, R.D. Gibbons, and B.R. Flay, Random-effects regression models for clustered data: with an example from smoking prevention research, *Journal of Consulting and Clinical Psychology* 62, (1994) 757-765.
- [34] D. Hedeker and R.D. Gibbons, MIXOR: a computer program for mixed-effects ordinal regression analysis, *Computer Methods and Programs in Biomedicine* 49, (1996) 157-176.
- [35] B.R. Flay, B.R. Brannon, C.A. Johnson, *et al.* The television, school and family smoking cessation and prevention project: I. theoretical basis and program development, *Preventative Medicine* 17, (1989) 585-607.
- [36] R.D. Gibbons and D. Hedeker, Application of random-effects probit regression models, *Journal of Consulting and Clinical Psychology* 62, (1994) 285-296.
- [37] Hough, R.L., Harmon, S., Tarke, H., Yamashiro, S., Quinlivan, R., Landau-Cox, P., Hurlburt, M.S., Wood, P.A., Milone, R., Renker, V., Crowell, A. & Morris, E., (1997). Supported independent housing: Implementation issues and solutions in the San Diego McKinney Homeless Demonstration Research Project. In W.R. Breakey & J.W. Thompson (eds.), *Mentally Ill and Homeless: Special Programs for Special Needs*, pp. 95-117. New York: Harwood Academic Publishers.
- [38] Hurlburt, M.S., Wood, P.A., & Hough, R.L. (1996). Providing independent housing for the homeless mentally ill: a novel approach to evaluating long-term longitudinal housing patterns, *Journal of Community Psychology* 24, (1996) 291-310.
- [39] N.M. Laird, Missing data in longitudinal studies, *Statistics in Medicine* 7, (1988) 305-315.
- [40] D.B. Rubin, Inference and missing data, *Biometrika* 63, (1976) 581-592.

- [41] D. Hedeker and R.D. Gibbons, Application of random-effects pattern-mixture models for missing data in longitudinal studies, *Psychological Methods* 2, (1997) 64-78.
- [42] N.M. Laird and J.H. Ware, Random-effects models for longitudinal data, *Biometrics* 40 (1982) 961-971.
- [43] R.D. Bock, Within-subject experimentation in psychiatric research, in *Statistical and Methodological Advances in Psychiatric Research*, Ed. R.D. Gibbons and M. Dysken, pp. 59-90 (Spectrum, New York, 1983).
- [44] J.F. Strenio, H.I. Weisberg, and A.S. Bryk, Empirical Bayes estimation of individual growth curve parameters and their relationship to covariates, *Biometrics* 39 (1983) 71-86.
- [45] R.I. Jennrich and M.D. Schluchter, Unbalanced repeated-measures models with structured covariance matrices, *Biometrics* 42 (1986) 805-820.
- [46] R.D. Bock, *Multivariate Statistical Methods in Behavioral Research*, (McGraw-Hill, New York, 1975).
- [47] D. Hedeker, B.R. Flay, and J. Petraitis, Estimating individual influences of behavioral intentions: an application of random-effects modeling to the Theory of Reasoned Action, *Journal of Consulting and Clinical Psychology* 64, (1996) 109-120.
- [48] G.N. Masters, A comparison of latent trait and latent class analyses of Likert-type data, *Psychometrika*, 50, (1985) 69-82.
- [49] D.J. Bartholomew, *Latent Variable Models and Factor Analysis*, (Oxford University Press, New York, 1987).
- [50] J.M. Neuhaus, J.D. Kalbfleisch, and W.W. Hauck, A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data, *International Statistical Review* 59, (1991) 25-35.
- [51] S.L. Zeger, K.-Y. Liang, and P.S. Albert, Models for longitudinal data: A generalized estimating equation approach, *Biometrics* 44, (1988) 1049-1060.
- [52] R.D. Bock and B.H. Repp, *MATCAL: Double Precision Matrix Operations Subroutines*, (National Educational Resources, Chicago, 1974).