ELSEVIER

Computer Methods and Programs in Biomedicine 49 (1996) 157–176

computer methods and programs in biomedicine

MIXOR: a computer program for mixed-effects ordinal regression analysis

Donald Hedeker*[a], Robert D. Gibbons[b]

[a]*Division of Epidemiology and Biostatistics (M/C 922), School of Public Health and Prevention Research Center, University of Illinois at Chicago, 2121 West Taylor Street, Room 510, Chicago, IL, 60612-7260, USA*

[b]*Department of Psychiatry and School of Public Health, University of Illinois at Chicago, Chicago, IL, 60612-7260, USA*

Received 25 July 1995; revised 17 January 1996; accepted 18 January 1996

Abstract

MIXOR provides maximum marginal likelihood estimates for mixed-effects ordinal probit, logistic, and complementary log-log regression models. These models can be used for analysis of dichotomous and ordinal outcomes from either a clustered or longitudinal design. For clustered data, the mixed-effects model assumes that data within clusters are dependent. The degree of dependency is jointly estimated with the usual model parameters, thus adjusting for dependence resulting from clustering of the data. Similarly, for longitudinal data, the mixed-effects approach can allow for individual-varying intercepts and slopes across time, and can estimate the degree to which these time-related effects vary in the population of individuals. MIXOR 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 MIXOR are provided.

Keywords: Longitudinal data; Clustered data; Random effects; Correlated responses; Multilevel data; Random coefficients models; Dichotomous outcomes; Graded responses; Categorical data

1. Introduction

Models for dichotomous and ordinal responses are important in many areas of research, since subjects are often classified or may respond on a dichotomous or ordinal scale. In biomedical studies, for example, subjects may be classified in terms of response versus non-response, or exhibiting definite, mild, or no symptomatology of a given disease or condition. Additionally, it is often the case that subjects are nested within clusters (i.e. schools, firms, clinics) or are repeatedly assessed over time, making use of ordinal regression models [1,2] that assume independence of observations problematic.

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 [3–5]. Common to both clustered and longitudinal data is the idea of

* Corresponding author, Tel: +1 312 9964896.

0169-2607/96/$15.00 © 1996 Elsevier Science Ireland Ltd. All rights reserved
PII S0169-2607(96)01720-8

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 the cluster effects, while for longitudinal data the random effects represent the subject effects.

Much of the work in this area has focused on models for continuous response data, and variants of these mixed-effects regression models for continuous data have been developed under a variety of names: random-effects models [6,7], variance component models [8], hierarchical linear models [4], multilevel models [3], two-stage models [9], random coefficient models [10], mixed models [11], empirical Bayes models [12], unbalanced repeated-measures models [13], and random regression models [14–18]. The approaches presented in these articles generally involve linear regression models that allow the possibility that some parameters besides the residuals are random, and not fixed.

Thus, a mixed-effects regression model generally contains some fixed effects in the model in addition to the random effects. For longitudinal data, mixed-effects models allow for presence of missing data (i.e. subjects are not assumed to be measured at the same number of timepoints), time-varying or invariant covariates, and subjects measured at different timepoints. In the analysis of clustered data, outcomes at the individual level are modeled in terms of both individual and cluster level variables, while concurrently estimating and adjusting for amount of intraclass correlation present in the data. Further, these models make no assumption regarding cluster sample size, allowing for a varying number of subjects within each cluster.

An increasing amount of work has focused on random-effects models for non-continuous response data. Mixed-effect models for both dichotomous [19–23] and ordinal [24–27] outcomes have recently been proposed. This paper describes the FORTRAN program MIXOR (mixed-effects ordinal regression) for the analysis of longitudinal or clustered responses which are either dichotomous or ordinal. MIXOR can accommodate multiple random effects, and allows for a general form for model covariates. Assuming either a probit, logistic, or complementary log-log response function, a maximum marginal likelihood solution is implemented using multidimensional 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 ordinal response data will illustrate features of MIXOR.

Some commercially-based software exists to perform mixed-effects regression analysis for non-continuous outcomes. The MLn [28], VARCL [29], and EGRET [30] software programs have facilities for dichotomous response data. MIXOR, though, can also handle ordinal outcomes and is available both for the MACINTOSH and MS-DOS environments. Also, MIXOR is unique in allowing either a probit, logistic or complementary log-log response function.

2. Computational methods

Hedeker and Gibbons [27] described the statistical development of the random-effects ordinal regression model; here we will present the key computational features. To motivate the ordinal regression model, it is often assumed that there is an unobservable latent variable (y) which is related to the actual response through the 'threshold concept'. For the dichotomous model, one threshold value is assumed, while for the ordinal model, a series of threshold values $\gamma_1,\gamma_2,\ldots,\gamma_{J-1}$ is assumed, where J equals the number of ordered categories, $\gamma_0 = -\infty$, and $\gamma_J = \infty$. Here, a response occurs in category j ($Y = j$) if the latent response process y exceeds the threshold value γ_{j-1}, but does not exceed the threshold value γ_j.

Using the terminology of multilevel analysis [3] let i denote the level-2 units (clusters in the clustered data context, or subjects in the longitudinal data context), and let k denote the level-1 units (subjects in the clustered data context or repeated observations in the longitudinal data context).

Assume that there are $i = 1,...N$ level-2 units and $k = 1,...,n_i$ level-1 units nested within each level-2 unit. The mixed-effects regression model for the latent response strength y_{ik} can be written as follows:

$$y_{ik} = x'_{ik}\beta_i + w'_{ik}\alpha + \epsilon_{ik} \quad (1)$$

where w_{ik} is the $p \times 1$ covariate vector and x_{ik} is the design vector for the r random effects, both vectors being for the kth level-1 unit nested within level-2 unit i. Also, α is the $p \times 1$ vector of unknown fixed regression parameters, β_i is the $r \times 1$ vector of unknown random effects for the level-2 unit i, and ϵ_{ik} are the model residuals. The distribution of the random effects is assumed to be multivariate normal with mean vector μ and covariance matrix Σ_β, and the residuals are assumed to be independently normally distributed with mean 0 and variance σ^2.

With the above mixed-effects regression model for the underlying and unobservable variable y_{ik}, the probability, for a given level-2 unit i, that $y_k = j$ (a response occurs in category j), conditional on β and α, is given by the following equation:

$$P(y_k = j|\beta,\alpha) = \Phi[(\gamma_j - z_k/\sigma] - \Phi[(\gamma_{j-1} - z_k)/\sigma]$$

where $z_k = x'_k\beta + w'_k\alpha$ and $\Phi(\cdot)$ represents the cumulative standard normal density function. Without loss of generality, the origin and unit of z may be chosen arbitrarily. For convenience, let $\gamma_1 = 0$ and $\sigma = 1$.

Alternatively, if the logistic response function is assumed, then the logistic function $\Psi(\cdot)$ replaces $\Phi(\cdot)$ in the conditional probability, where,

$$\Psi(\gamma_j - z_k) = \frac{1}{1 + \exp[-(\gamma_j - z_k)]}$$

Another choice for the response function is the complementary log-log function, in which case, $\Upsilon(\cdot)$ replaces $\Phi(\cdot)$, where

$$\Upsilon(\gamma_j - z_k) = 1 - \exp[-\exp(\gamma_j - z_k)]$$

For either the logistic or complementary log-log response function, again, we let $\gamma_1 = 0$, however, the residual variance equals $\pi^2/3$ and $\pi^2/6$, respectively, for the standard logistic and complementary log-log distributions. In what follows, the normal response function will be assumed and modifications for the logistic and complementary log-log response functions will be indicated.

2.1. Maximum marginal likelihood estimation

Letting $\mathbf{y}_i$ denote the vector pattern of ordinal item responses from level-2 unit i for the n_i level-1 units nested within, the probability of any pattern $\mathbf{y}_i$, given β and α, is equal to the product of the probabilities of the level-1 responses:

$$\ell(\mathbf{y}_i|\beta,\alpha) = \prod_{k=1}^{n_i} \prod_{j=1}^{J} [\Phi((\gamma_j - z_{ik}) - \Phi(\gamma_{j-1} - z_{ik})]^{d_{ikj}} \quad (2)$$

where $d_{ikj} = \begin{cases} 1 \text{ if } Y_{ik} = j \\ 0 \text{ if } Y_{ik} \neq j \end{cases}$

It is convenient to orthogonally transform the response model, so that $\beta = T\theta + \mu$, where $TT' = \Sigma_\beta$ is the Cholesky decomposition of Σ_β. The reparameterized model is then

$$z_{ik} = x'_{ik}(T\theta + \mu) + w'_{ik}\alpha$$

Then the marginal density of $\mathbf{y}_i$ in the population is expressed as the following integral of the likelihood, $\ell(\cdot)$, weighted by the prior density $g(\cdot)$:

$$h(y_i) = \int_\theta \ell(y_i|\theta,\alpha)\, g(\theta)\, d\theta$$

where $g(\theta)$ represents the multivariate standard normal density.

For the estimation of the covariate coefficients α, the population parameters μ and T, and the $\gamma_j (j = 2,...,J-1)$ threshold parameters, the marginal log-likelihood for the patterns from the N level-2 units,

$$\log L = \sum_{i}^{N} \log h(\mathbf{y}_i)$$

is differentiated with respect to each parameter vector (see Hedeker and Gibbons [27]. Fisher's method of scoring can then be used to provide the solution to these likelihood equations. Provisional estimates for the vector containing all parameters $\boldsymbol{\Theta}$, on iteration ι are improved by

$$\boldsymbol{\Theta}_{\iota+1} = \boldsymbol{\Theta}_{\iota} - \mathrm{E}\left[\frac{\partial^2 \log L}{\partial \boldsymbol{\Theta}_{\iota}\, \partial \boldsymbol{\Theta}'_{\iota}}\right]^{-1} \frac{\partial \log L}{\partial \boldsymbol{\Theta}_{\iota}} \qquad (3)$$

where the information matrix, or expectation of the matrix of second derivatives, is given by

$$\mathrm{E}\left[\frac{\partial^2 \log L}{\partial \boldsymbol{\Theta}_{\iota}\, \partial \boldsymbol{\Theta}'_{\iota}}\right] = \sum_{i=1}^{N} h^{-2}(\mathbf{y}_i) \frac{\partial h(\mathbf{y}_i)}{\partial \boldsymbol{\Theta}_{\iota}} \left(\frac{\partial h(\mathbf{y}_i)}{\partial \boldsymbol{\Theta}_{\iota}}\right)'$$

In order to solve the above likelihood equations, numerical integration on the transformed θ space can be used. For this, Gauss-Hermite quadrature can approximate the above integrals to any practical degree of accuracy. 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 θ space, the summation goes over Q^r points. 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 [31] 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 necessary to obtain adequate accuracy.

Hedeker and Gibbons [27] discuss how the above solution can be modified to accommodate weighted data, for example, when the same response pattern $\mathbf{y}_i$ and covariate vector $\mathbf{w}_i$ is observed for a number of level-2 units. To modify the solution for the logistic regression formulation, the logistic function $\Psi(\cdot)$ replaces the normal response function $\Phi(\cdot)$ and the product $\Psi(\cdot) \times (1 - \Psi(\cdot))$ replaces the standard normal density function $\phi(\cdot)$ in the above derivatives. A similar substitution using $\Upsilon(\cdot)$ is used for the complementary log-log response function.

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 [32]. While this use of the standard errors to perform hypothesis tests (and construct confidence intervals) for the fixed effects (the coefficient vector $\boldsymbol{\alpha}$ and mean vector of the random effects $\boldsymbol{\mu}$) is generally reasonable for the variance and covariance components ($\boldsymbol{T}$) this practice is problematic (see Bryk and Raudenbush [4] p. 55). Instead, in order to test hypotheses related to the variance and covariance components, as well as the fixed effects, the likelihood-ratio χ^2 test can be used for comparison of nested models.

3. Program description and usage

MIXOR is currently available in executable form for both MS-DOS and MACINTOSH computers. In the MS-DOS environment, MIXOR can be run in either batch (MIXORB.EXE) or interactive mode (MIXOR.EXE), while in the MACINTOSH environment only the batch mode (MIXORB) is possible. For batch processing, the MIXOR instructions must be stored in the file MIXOR.DEF (described below), while in interactive mode the user can specify the various options using the menu-orientated user interface. Here, we will discuss the procedure for running the program primarily in batch mode. In either batch or interactive processing, MIXOR makes use of the following files:

- input data file
- MIXOR.DEF — main definition file for analysis options and settings

used in interactive mode only, and is described more fully below.

3.1. Structure of the input data file

This file contains all data (i.e. responses 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, for example, in the longitudinal data setting, there are repeated observations (level 1) within individuals (level 2). There must be a level-2 ID variable for each record and the data must be sorted by this level-2 ID variable. The repeated measurements of an individual take up as many records in this file as there are measurements for that individual. Thus, some individuals will have, for example, four records while others may have two or five records. Alternatively, if missing value codes are utilized, each individual may have data on the same number of records, but some records will contain missing value codes for some (or all) of the variables. Similarly, in the clustered data setting there are nested observations (students, employees, patients) within clusters (schools, firms, clinics). 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 will have, for example, 40 records while others may have 20–50 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 DepVar Xvector Wvector

where, in the longitudinal context, *ID* refers to an individual ID number which does not change across timepoints, *DepVar* is the value of the dependent measure at the particular timepoint, *Xvector* is the part of the design matrix for the random effects at the given timepoint, and *Wvector* is the covariate vector at the timepoint. In the clustered data context, *ID* refers to the cluster ID number which does not change across nested observations, *DepVar* is the value of the dependent measure, *Xvector* is the part of the design matrix for the random effects, and *Wvector* is the covariate vector; all given for each clustered 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 in particular, missing values left as blank fields will definitely cause problems.

3.2. Analysis options and settings — MIXOR.DEF

This file contains the information to determine which statistical model should be 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 1a, 1b, 1c, and 1d.

Except where noted, this file is read in free

Table 1a
Analysis options and settings specified in MIXOR.DEF: lines 1–5

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 must be specified even in batch processing, although in batch processing nothing is done to this file. In interactive mode, after a filename is entered in the appropriate menu field, the program settings and options of the specified file will be retrieved if that file exists. Prior to running the statistical procedure from interactive mode, the current MIXOR options and settings selected by the user will be saved into the specified file.

Table 1b
Analysis options and settings specified in MIXOR.DEF: line 6

Line 6	—	NPR NF R P CONV MAXJ MISS START WT CATYX NQUAD FUNC
NPR	=	number of level-2 units whose data will be listed on the screen (usually set to 1).
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 vector of the random effects).
CONV	=	convergence criterion (usually set to 0.001 or 0.0001).
MAXJ	=	number of ordered dependent variable categories.
MISS	=	0 if no missing values are present in the data or 1 if missing values are present (codes for which will later be defined).
START	=	0 if automatic starting values are to be used, or 1 if user-defined starting values are to be used.
WT	=	0 if each 2nd level unit (person or cluster) is weighted equally, or 1 for differential weighting.
CATYX	=	0 if a crosstab of any variable by the dependent variable is not requested, and 1 if such a crosstab is requested.
NQUAD	=	number of quadrature points (per random-effect dimension) to use in the numerical integration (usually set between 10 and 20 for models with one random effect, and between 5 and 10 for models with multiple random effects).
FUNC	=	0 for the probit response function, 1 for the logistic response function, or 2 for the complementary log-log response function.

Table 1c
Analysis options and settings specified in MIXOR.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 dependent variable.
Line 8	—	R parameters are to be read on this line: the field(s) of the input data file which contains(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 contains(s) the P fixed effects.

Table 1d
Analysis options and settings specified in MIXOR.DEF: remaining 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 MAXJ values of the ordinal dependent 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 dependent variable, followed by the number of levels of this variable, and a list of the values for all of these levels.
next line (if MISS = 1)	— Missing value code for the dependent 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 covariates.
next line	— An 8-character label for the dependent variable
next line	— R labels for the random effects in eight-character width fields.
next line (if START = 1)	— R starting values for the means of the random effects.
next line	— P labels for the covariates in eight-character width fields (a maximum of ten labels per line).
next line (if START = 1)	— P starting values for the covariate effects.
next line (if START = 1)	— ((R × (R + 1))/2) starting values for the variance and covariance terms of the random effects given in 'packed' form, e.g. for a 2 × 2 covariance matrix, the order of the starting values should be: variance (1), covariance (1,2) and variance (2).
final line	— (if START = 1 and MAXJ > 2) — MAXJ −2 starting values for thresholds.

format. In batch processing, this file is created by the user directly before typing the command MIXORB (or on the MACINTOSH, double-clicking on the MIXORB file), while in interactive mode (typing the command MIXOR), this file is created using the menu-orientated user interface. For batch mode, this filename and extension (MIXOR.DEF) must be used and should be in the same directory as the program MIXORB.EXE or accessible via appropriate PATH statements. In interactive mode, the creation and storage of this file is done by the program.

3.3. Main output file

This file contains descriptive information about the variables read in to MIXOR as well as the main results of the specified analysis. The examples of the output file provided below illustrate the contents of this file. In terms of numbers of observations, 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 are listed. For each variable (except the ID variable) read in to the program, the following descriptive statistics are provided: minimum, maximum, mean, and standard deviation. These descriptive statistics are based on the total number of level-1 observations. For the dependent variable, 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 the dependent variable by the levels of one of the model covariates may be obtained. Starting values, either user-defined or program-generated, are listed for all model parameters. Finally, MIXOR indicates the number (and percentage) of level-2 units with non-varying level-1 responses on the dependent variable. As this percentage approaches the maximum value of 100%, computational difficulties can arise, since, in this case, for every level-2 unit, all of the nested level-1 responses for that level-2 unit are the same.

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. As mentioned, the log-likelihood value can be used to perform likelihood-ratio tests. Following the log-likelihood 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 and threshold parameters where one-tailed P-values are given. As noted earlier, this use of the standard errors to perform hypothesis tests for the variance and covariance parameters is controversial (see Bryk and Raudenbush [4], p. 55). Also, it is important to realize that for the variance terms, 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), MIXOR lists re-expressions of the estimated random-effects variance terms, depending on the type of model specified. If a random-intercepts model is specified then MIXOR calculates and lists the value of the estimated intra-cluster correlation. If a model with more than one random effect is specified, MIXOR first expresses the estimated Cholesky factor as a variance-covariance matrix, and then as a correlation matrix. Finally, correlation matrices are also provided for the estimates of all model parameters. These correlation matrices are not correlations of the

variables themselves but correlations of the estimated model parameters. These matrices may be helpful in determining the degree to which collinearity is present in terms of the model parameters.

3.4. Some common MIXOR errors

There are a few errors which can prevent MIXOR 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 are 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 is occurring, 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, the CATYX option (described in Tables 1b and 1d) is fairly unforgiving at this point. 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 MIXOR from running and this option is selected, the user can set CATYX = 0 to avoid this option. Third, the NPR option (described in Table 1b), which is used to list data to the screen, can cause MIXOR to stop in certain cases (essentially, 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. Finally, if the program 'blows up,' it may be that the model that is specified 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 ordered outcome categories if these are very 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 mixed-effects regression

MIXOR can estimate a variety of models for clustered and longitudinal data. An analysis of a clustered dataset where students are clustered within classrooms and schools is presented first to illustrate features of mixed regression analysis of clustered data. 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. To illustrate usefulness of mixed regression analysis for longitudinal ordinal data, an analysis of a psychiatric dataset where patients are rated on symptom severity across multiple timepoints will be presented. This analysis includes two random effects in order to account for differential baseline starting points (in terms of severity) and differential trends across time (change in severity across time) for the individual patients. These two examples will serve to 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 MIXOR.

4.1. Analysis of a clustered dataset

Hedeker, Gibbons, and Flay [33] describe mixed-effects regression for clustered data using a dataset where students are clustered within classrooms and schools. In that article, the dependent variable is treated continuously and the mixed-effects approach is compared with both individual-level analysis which ignores the clustering of data, and classroom-level analysis which aggregates individual data. Here, we present use of MIXOR for analysis of an ordinal outcome in the clustered data context to perform the mixed-effects analysis. The data for this example is from the Television School and Family Smoking Prevention and Cessation Project (TVSFP) [34].

For this illustration, a subset of the TVSFP data was used. 1600 students from 135 classrooms and

Table 2
Data from example 4.1: first 18 students from five classrooms and two schools

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
404	404101	3	1	1	1	1	1	1
404	404101	4	1	1	2	1	1	1
404	404101	2	0	1	4	1	1	1
404	404101	3	1	1	3	1	1	1
404	404101	2	0	1	1	1	1	1
404	404101	1	0	1	1	1	1	1
404	404101	3	1	1	2	1	1	1
404	404101	4	1	1	0	1	1	1
404	404101	2	0	1	1	1	1	1
404	404101	3	1	1	2	1	1	1
404	404101	4	1	1	2	1	1	1
404	404102	3	1	1	1	1	1	1
404	404102	1	0	1	1	1	1	1
404	404102	1	0	1	0	1	1	1
404	404102	4	1	1	4	1	1	1
404	404102	4	1	1	1	1	1	1
404	404102	3	1	1	2	1	1	1
404	404102	3	1	1	1	1	1	1
404	404102	4	1	1	2	1	1	1
404	404102	2	0	1	2	1	1	1
404	404103	3	1	1	2	1	1	1
404	404103	2	0	1	1	1	1	1
404	404103	2	0	1	2	1	1	1
404	404103	3	1	1	1	1	1	1
404	404103	3	1	1	1	1	1	1

28 schools are included, where 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 × 2 design of social-resistance classroom curriculum (CC = yes or no) by mass-media intervention (TV = yes or no). The outcome variable for this illustration is a tobacco and health knowledge scale (THKS) score. A student's score was defined as the number of correct answers to seven items on tobacco and health knowledge. The frequency distribution of

Table 3a
MIXOR.DEF file for example 4.1: student-level analysis ignoring clustering

```
TVSFP study — Post-test THKS ORDINAL
Students treated as independent observations — 0 random
TVSFPORS.DAT
TVSFPORO.OUT
TVSFPORO.DEF
1  9  0  5   0.0001  4  0  0  0  0   10  1

   2  3

   5  6  7   8   9
   1  2  3   4
THKScore

Intrcpt  PreTHKS CC    TV          CC*TV
```

Table 3b
MIXOR.DEF file for example 4.1: students-within-classrooms analysis

```
TVSFP study — Post-test THKS ORDINAL
Students nested within CLASSROOMS — 1 random effect
TVSFPORS.DAT
TVSFPORC.OUT
TVSFPORC.DEF
1  9  1  4   0.0001  4  0  0  0  0   10  1
   2  3
   5
   6  7  8  9
   1  2  3  4
THKScore
Intrcpt
PreTHKS CC     TV      CC*TV
```

the post-intervention THKS scores indicated that approximately half the students had scores of 2 or less, and half with 3 or greater. In terms of quartiles, four ordinal classifications were suggested corresponding to 0–1, 2, 3, and 4–7 correct responses.

A partial list of these data is given in Table 2. The variables are, in order, school ID, class ID, ordinal 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.

For these data, in addition to performing a mixed-effects analysis (to include and estimate the effect of clustering on the students' outcomes), MIXOR can be used to perform a more basic analysis ignoring data clustering. Tables 3a and 3b list the DEF files for the models ignoring and including data clustering, respectively.

In both, the ordinal post-intervention THKS score (THKScore) is modeled in terms of baseline THKS score (PreTHKS) and effects of CC, TV, and CC by TV interaction. The titles for these two DEF files are slightly different to reflect the different analyses. In the file designations on lines 3–5, the same input data file is specified in both DEF files, however, different OUT files are indicated so that the results of both analyses are saved to different files.

Line 6 in the DEF files contains the primary specifications for these two models. For both, NPR = 1, so the data from the first subject will be listed to the screen prior to the analysis. Nine fields of variables are to be read from the input data file, with the first model having 0 and 5 random and fixed effects, respectively, while the second specifies 1 random and 4 fixed effects. As a result of specifying zero random effects (in Table 3a), blank lines are given below for the field and label specifications of the random effects. A convergence criterion of 0.0001 is chosen, and four ordered categories are indicated for the dependent variable (with actual values of 1, 2, 3, and 4 given below in both DEF files). The number of quadrature points is set to ten for both models, though the model without random effects does not use quadrature for the solution (thus, for this model only, this specification is irrelevant though it must be given). Finally, line 6 of the DEF files indicates that the logistic response function is selected.

In both DEF files, the classroom ID (field 2) is identified as the cluster ID (though again, for the model without random effects this is irrelevant though a specification is necessary). The remaining

Table 4
Output file for example 4.1: students-within-classrooms analysis

MIXOR — The program for mixed-effects ordinal regression analysis
Response function: logistic
TVSFP study — Post-test THKS ORDINAL
Students nested within CLASSROOMS — 1 random effect

Numbers of observations
Level 2 observations = 135
Level 1 observations = 1600

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

20	3	11	9	5	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	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	S.D.
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	0.580			
covariates	0.212	0.451	0.140	−0.201
var. terms	1.040			
thresholds	1.137	2.202		

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

Final results — maximum marginal likelihood estimates

Total iterations = 12
Quad pts. per dim = 10
Log likelihood = −2115.381
Ridge = 0.000

Variable	Estimate	S.E.	*Z*	*P*-value
Intrcpt	0.07573	0.15356	0.49315	0.62191
PreTHKS	0.41480	0.04082	10.16072	0.00000
CC	0.86137	0.18730	4.59878	0.00000
TV	0.20572	0.16810	1.22375	0.22105
CC*TV	−0.30097	0.25186	−1.19502	0.23208
Random effect variance term (S.D.)				
Intrcpt	0.43454	0.07627	5.69709	0.00000
Thresholds				
1	0.00000			
2	1.27344	0.06295	20.22913	0.00000
3	2.47899	0.08021	30.90619	0.00000

note: *P*-values are two-tailed except for thresholds and variance terms which are one-tailed

Calculation of the intracluster correlation

residual variance	= pi × pi/3 (assumed)
cluster variance	= (0.435 × 0.435) = 0.189
intracluster correlation	= 0.189/(0.189 + (pi × pi/3)) = 0.054

Correlation of the maximum marginal likelihood estimates

		1 Intrcpt	2 PreTHKS	3 CC	4 TV	5 CC*TV	6 VarCov1
1	Intrcpt	1.0000					
2	PreTHKS	−0.5333	1.0000				
3	CC	−0.5608	0.0282	1.0000			
4	TV	−0.6027	−0.0147	0.4952	1.0000		
5	CC*TV	0.3688	0.0891	−0.7179	−0.6605	1.0000	
6	VarCov1	−0.0081	0.0997	0.1411	−0.1750	0.0745	1.0000
7	Threshd2	−0.0186	0.3161	0.1145	0.0842	−0.0240	0.1454
8	Threshd3	0.0239	0.3119	0.0464	0.0031	0.0624	0.3201

		7 Threshd2	8 Threshd3
7	Threshd2	1.0000	
8	Threshd3	0.7603	1.0000

lines in the DEF files indicate the fields and labels for the dependent and independent variables, as well as information depending on the options selected on line 6. Table 4 lists the results for the model specified by the DEF file given in Table 3b, that is, from the analysis treating students nested within classrooms.

As can be seen from Table 4, the nesting of the 1600 students within the 135 classrooms is associated with classroom sizes between 1 and 28. Descriptive statistics are listed for all variables considered in the analysis, including a listing of the frequencies in each of the ordered categories. Following the descriptive statistics, the program lists the number and percentage of classrooms (i.e. the level-2 units) with non-varying responses across students; six classrooms (4.44%) had identical THKScore values from students. In terms of results, significant effects of PreTHKS and CC are observed. Also, the intracluster correlation (in this case, the intraclassroom correlation) from this analysis equals 0.054. Comparing the fit of this model to a model that does not include a random classroom effect (i.e. the model specified by the DEF file in Table 3a), yields a likelihood-ratio $\chi^2 = -2[-2125.103 - (-2115.381)] = 19.44$ on 1 degree of freedom ($P < 0.001$), indicating that there is clear evidence of a non-zero intraclassroom correlation.

It should be noted that two mixed-effects regression models could be considered for these data: students within schools, and students within classrooms. At present, MIXOR does not allow a three-level analysis which would consider the students as nested within both classrooms and schools concurrently. For the students-within-classrooms analysis performed, the class ID (the second variable field in the datafile) was indicated as the cluster ID on line 7 of the DEF file. To perform a students-within-schools analysis, the school ID (the first variable field) would be indicated as the cluster ID.

Table 5
Data from example 4.2: first eight subjects

1103	5.500	1	4	1	1	0	0.0000	0.0000
1103	3.000	0	2	1	1	1	1.0000	1.0000
1103	−9.000	−9	−9	1	1	2	1.4142	1.4142
1103	2.500	0	2	1	1	3	1.7321	1.7321
1103	−9.000	−9	−9	1	1	4	2.0000	2.0000
1103	−9.000	−9	−9	1	1	5	2.2361	2.2361
1103	4.000	1	2	1	1	6	2.4495	2.4495
1104	6.000	1	4	1	1	0	0.0000	0.0000
1104	3.000	0	2	1	1	1	1.0000	1.0000
1104	−9.000	−9	−9	1	1	2	1.4142	1.4142
1104	1.500	0	1	1	1	3	1.7321	1.7321
1104	−9.000	−9	−9	1	1	4	2.0000	2.0000
1104	−9.000	−9	−9	1	1	5	2.2361	2.2361
1104	2.500	0	2	1	1	6	2.4495	2.4495
1105	4.000	1	2	1	1	0	0.0000	0.0000
1105	3.000	0	2	1	1	1	1.0000	1.0000
1105	−9.000	−9	−9	1	1	2	1.4142	1.4142
1105	1.000	0	1	1	1	3	1.7321	1.7321
1105	−9.000	−9	−9	1	1	4	2.0000	2.0000
1105	−9.000	−9	−9	1	1	5	2.2361	2.2361
1105	−9.000	−9	−9	1	1	6	2.4495	2.4495
1106	3.000	0	2	1	1	0	0.0000	0.0000
1106	1.000	0	1	1	1	1	1.0000	1.0000
1106	−9.000	−9	−9	1	1	2	1.4142	1.4142
1106	1.500	0	1	1	1	3	1.7321	1.7321
1106	−9.000	−9	−9	1	1	4	2.0000	2.0000
1106	−9.000	−9	−9	1	1	5	2.2361	2.2361
1106	1.000	0	1	1	1	6	2.4495	2.4495
1107	5.000	1	3	1	0	0	0.0000	0.0000
1107	5.000	1	3	1	0	1	1.0000	0.0000
1107	−9.000	−9	−9	1	0	2	1.4142	0.0000
1107	5.000	1	3	1	0	3	1.7321	0.0000
1107	−9.000	−9	−9	1	0	4	2.0000	0.0000
1107	−9.000	−9	−9	1	0	5	2.2361	0.0000
1107	5.500	1	4	1	0	6	2.4495	0.0000
1108	6.000	1	4	1	1	0	0.0000	0.0000
1108	6.000	1	4	1	1	1	1.0000	1.0000
1108	−9.000	−9	−9	1	1	2	1.4142	1.4142
1108	3.500	1	2	1	1	3	1.7321	1.7321
1108	−9.000	−9	−9	1	1	4	2.0000	2.0000
1108	−9.000	−9	−9	1	1	5	2.2361	2.2361
1108	4.500	1	3	1	1	6	2.4495	2.4495
1109	4.000	1	2	1	1	0	0.0000	0.0000
1109	2.000	0	1	1	1	1	1.0000	1.0000
1109	−9.000	−9	−9	1	1	2	1.4142	1.4142
1109	2.000	0	1	1	1	3	1.7321	1.7321
1109	−9.000	−9	−9	1	1	4	2.0000	2.0000
1109	−9.000	−9	−9	1	1	5	2.2361	2.2361
1109	2.500	0	2	1	1	6	2.4495	2.4495
1110	4.000	1	2	1	1	0	0.0000	0.0000
1110	4.500	1	3	1	1	1	1.0000	1.0000
1110	−9.000	−9	−9	1	1	2	1.4142	1.4142
1110	4.000	1	2	1	1	3	1.7321	1.7321
1110	−9.000	−9	−9	1	1	4	2.0000	2.0000
1110	−9.000	−9	−9	1	1	5	2.2361	2.2361
1110	3.500	1	2	1	1	6	2.4495	2.4495

4.2. Longitudinal data

To illustrate application of the mixed-effects ordinal probit regression model to longitudinal data, we examined data collected in the NIMH Schizophrenia Collaborative Study on treatment-related changes in overall severity. Specifically, we examined Item 79 of the Inpatient Multidimensional Psychiatric Scale (IMPS; [35]). Item 79, 'Severity of Illness,' (IMPS79) was scored as: 1 = normal, not at all ill; 2 = borderline mentally ill; 3 = mildly ill; 4 = moderately ill; 5 = markedly ill; 6 = severely ill; and 7 = among the most extremely ill. Previously, we have analyzed these data both assuming a continuous scale for these seven ordered response categories using mixed-effects regression [15] and also dichotomizing responses using mixed-effects binary probit regression [36]. For the present illustration of the ordinal mixed-effects model, we recoded the seven ordered categories into four: (1) normal or borderline mentally ill; (2) mildly or moderately ill; (3) markedly ill; and (4) severely or among the most extremely ill. In this study, patients were randomly assigned to receive one of four medications: placebo, chlorpromazine, fluphenazine, or thioridazine. Since our previous analyses revealed similar effects for the three anti-psychotic drug groups, they were combined in the present analysis. Finally, again based on previous analysis, to linearize the relationship of the IMPS79 scores over time a square root transformation of time was chosen. Table 5 lists the data from the first eight subjects.

The nine variables are, in order, subject ID, IMPS79 score, dichotomous IMPS79 score, ordinal IMPS79 score, a column of ones for the intercept, treatment group, week, the square root of week, and the product of treatment group by square root of week. Not all of these variables need to be included in a given analysis, since MIXOR allows the user to choose among the variables in an input data file for a particular model. In this data file, each subject's data consist of seven lines. Notice that there are missing value codes (–9) for some subjects at specific timepoints — the data from these timepoints 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. The use of missing value codes is not the only way of dealing with missing data. An alternative way of handling the missing data is for the user to physically remove the records with

Table 6a
MIXOR.DEF file for example 4.2: random-intercepts probit model

```
NIMH Schiz data — two groups — seven timepoints
IMPS79 (ordinal) across SQRT week — one random effect
SCHIZX1.DAT
SCHIZo1.OUT
SCHIZo1.DEF
1 9 1 3   0.00010   4 1 0 0 1   20 0
1 4
5
6 8 9
1 2 3 4
8 7 0.0000   1.0000   1.4142   1.7321   2.0000   2.2361   2.4495
-9
-9
- 9  -9  -9
Imps790
Intercpt
TxDrug SqrtWeekTx*SWeek
```

Table 6b
MIXOR.DEF file for example 4.2: random intercepts and slopes probit model

```
NIMH Schiz data — two groups — seven timepoints
IMPS79 (ordinal) across SQRT week — two random effects
SCHIZX1.DAT
SCHIZo2.OUT
SCHIZo2.DEF
1 9 2 2   0.00010   4 1 0 0 1   10 0
1 4
5 8
6 9
1 2 3 4
8 7 0.0000   1.0000   1.4142   1.7321   2.0000   2.2361   2.4495
-9
-9  -9
-9  -9
Imps790
InterceptSqrtWeek
TxDrug Tx*Sweek
```

missing data, so that each subject would have a varying number of records in the file.

Tables 6a and 6b list the MIXOR.DEF files for two mixed-effects ordinal probit regression models of these data. For both models, the repeated ordinal IMPS score (IMPS79O) is modeled in terms of a dummy-coded drug effect (TxDrug: placebo = 0 and drug = 1), a time effect (SqrtWeek: square root of week) and a drug by time interaction (Tx*Sweek). In terms of the random effects, the first DEF file specifies a random-intercepts model, while the second requests a model allowing patients to vary in terms of both their intercept and their trend across time (random intercepts and slopes). Again, the titles for these two DEF files reflect the different analyses, and different OUT files are indicated so that the results of both analyses are saved to different files.

From line 6 of the DEF files, one random and three fixed effects are specified for the first model, while two random and two fixed are specified for the second. Four outcome categories (with values 1, 2, 3, and 4) are indicated, and the missing value (MISS) option is requested. Missing value codes of −9 are given subsequently for all variables. The CATYX option is specified, and subsequent information in the DEF files indicates that a crosstabulation of IMPS79O by SqrtWeek (the square root of week) is requested; the number of levels of SqrtWeek and all actual values that this variable assumes in the input data file are listed subsequently in the DEF files.

The number of quadrature points is 20 for the univariate random-effects normal distribution (random-intercepts model), and ten in each dimension for the bivariate random-effects normal distribution (i.e. a total of $10 \times 10 = 100$ points). Finally, line 6 of the DEF files indicates that the probit response function is selected.

In terms of the missing value specification, notice that even though missing values are coded only for the dependent variable in the input data file, numeric missing value codes must be specified in the MIXOR.DEF file for all model terms (if MISS = 1). In this case, the value −9 was specified for all variables since for the dependent variable this value is the correct missing value code, while for all other model terms (intercpt, TxDrug, SqrtWeek, and Tx*Sweek) this value was never observed.

Table 7 lists the results from the random intercepts and slopes model (the model specified by the DEF file in Table 6b).

Following the descriptive information provided

Table 7
Output file for example 4.2: random intercepts and slopes probit model

MIXOR — The program for mixed-effects ordinal regression analysis
Response function: normal
NIMH Schiz data — two groups — seven timepoints
IMPS79 (ordinal) across SQRT week — two random effects
Numbers of observations
Level 2 observations = 437
Level 1 observations = 1603

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

4	4	3	4	4	4	4	4	4	3	4	4	4	2	3	4	3	4	3
4	4	4	3	3	2	4	4	4	4	4	3	4	4	4	4	4	4	4
4	4	2	3	4	3	4	4	4	3	4	4	2	2	4	5	4	2	4
4	3	4	4	3	2	3	4	4	4	4	4	4	2	4	4	4	5	4
4	2	2	4	2	4	4	3	3	4	4	4	4	4	4	4	4	3	3
4	2	3	4	4	4	2	5	3	4	4	2	4	4	4	2	4	4	4
4	4	4	4	4	4	5	2	4	3	4	4	2	2	4	4	4	4	4
2	4	4	4	4	4	4	4	4	4	4	4	4	4	2	4	4	2	4
4	4	3	4	2	4	4	3	2	3	4	4	3	3	4	3	4	4	4
4	4	4	4	4	4	4	4	4	4	4	2	3	3	5	4	3	4	4
3	2	4	4	4	4	4	3	3	4	4	4	4	4	4	4	4	4	4
4	4	4	4	4	3	4	4	4	4	4	4	2	3	4	4	4	2	4
4	4	4	3	4	4	4	4	4	4	4	4	3	4	4	3	4	4	2
4	4	4	4	2	4	4	4	2	4	4	4	3	3	4	3	4	2	4
4	4	3	3	4	4	4	4	3	3	4	3	4	4	4	4	4	3	4
4	4	4	4	3	3	4	2	4	4	4	4	4	4	4	4	3	4	4
3	3	4	2	4	3	3	3	4	4	4	4	4	4	4	3	2	3	4
4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
4	4	4	4	4	2	4	4	4	4	4	4	4	4	4	4	4	4	4
4	4	2	4	4	4	4	3	4	4	4	4	3	4	4	4	4	4	4
3	3	3	4	4	4	4	4	2	3	4	2	4	2	2	4	4	3	4
4	4	2	4	4	4	4	3	4	4	4	4	4	4	4	4	4	4	4
4	4	4	4	4	4	4	3	4	4	4	4	2	4	4	4	4	4	4

Descriptive statistics for all variables

Variable	Minimum	Maximum	Mean	S.D.
Imps790	1.00000	4.00000	2.79601	1.02840
Intercpt	1.00000	1.00000	1.00000	0.00000
SqrtWeek	0.00000	2.44950	1.22041	0.89651
TxDrug	0.00000	1.00000	0.76419	0.42464
Tx*Sweek	0.00000	2.44950	0.94424	0.94541

Categories of the response variable Imps790

Category	Frequency	Proportion
1.00	190.00	0.11853
2.00	474.00	0.29570
3.00	412.00	0.25702
4.00	527.00	0.32876

Crosstabulation of variable SqrtWeek by the response variable Imps790

Imps790

SqrtWeek	1.00	2.00	3.00	4.00	Total
0.00	1.0 (0.00)	54.0 (0.12)	122.0 (0.28)	257.0 (0.59)	434.0
1.00	23.0 (0.05)	135.0 (0.32)	124.0 (0.29)	144.0 (0.34)	426.0
1.41	3.0 (0.21)	4.0 (0.29)	2.0 (0.14)	5.0 (0.36)	14.0
1.73	54.0 (0.14)	132.0 (0.35)	113.0 (0.30)	75.0 (0.20)	374.0
2.00	5.0 (0.45)	3.0 (0.27)	2.0 (0.18)	1.0 (0.09)	11.0
2.24	3.0 (0.33)	4.0 (0.44)	0.0 (0.00)	2.0 (0.22)	9.0
2.45	101.0 (0.30)	142.0 (0.42)	49.0 (0.15)	43.0 (0.13)	335.0
Total	190.0	474.0	412.0	527.0	1603.0

Starting values

mean	1.673	−0.264	
covariates	−0.028	−0.345	
var. terms	1.000	0.000	0.500
thresholds	1.037	1.700	

→ The number of level 2 observations with non-varying responses = 79 (18.08%)

Final results — maximum marginal likelihood estimates

Total iterations	= 21
Quad pts. per dim	= 10
Log likelihood	= −1663.326
Ridge	= 0.000

Variable	Estimate	S.E.	Z	P-value
Intercpt	4.10961	0.25198	16.30959	0.00000
SqrtWeek	−0.50513	0.13054	−3.86970	0.00011
TxDrug	0.03882	0.22477	0.17271	0.86288
Tx*Sweek	−0.95060	0.14891	−6.38371	0.00000
Random effect variance and covariance terms (Cholesky of var-covariance matrix)				
Intercpt	1.48620	0.14130	10.51838	0.00000
covariance	−0.31464	0.08993	−3.49875	0.00047
SqrtWeek	0.73034	0.06951	10.50678	0.00000
Thresholds				
1	0.00000			
2	2.18421	0.10994	19.86701	0.00000
3	3.65376	0.14426	25.32735	0.00000

note: *P*-values are two-tailed except for thresholds and variance terms which are one-tailed

Calculation of the random effects variance-covariance matrix

```
  Intercpt variance = (1.486 × 1.486)     = 2.209
        covariance = (1.486 × −0.315)     = −0.468
SqrtWeek variance = (−0.315 × −0.315) + (0.730 × 0.730) = 0.632
```

Covariance expressed as a correlation = −0.396

Correlation of the maximum marginal likelihood estimates

		1 Intercpt	2 SqrtWeek	3 TxDrug	4 Tx * Sweek	5 VarCov1	6 VarCov2
1	Intercpt	1.0000					
2	SqrtWeek	−0.5922	1.0000				
3	TxDrug	−0.6265	0.5456	1.0000			
4	Tx * Sweek	0.2380	−0.7911	−0.6254	1.0000		
5	VarCov1	0.3407	−0.1709	0.0424	−0.1611	1.0000	
6	VarCov2	−0.2288	0.0743	−0.1031	0.1516	−0.6576	1.0000
7	VarCov3	0.3442	0.0248	0.1210	−0.2737	0.3396	−0.5122
8	Threshd2	0.5464	−0.0741	0.0981	−0.2880	0.2699	−0.2551
9	Threshd3	0.6302	−0.1499	0.0716	−0.3006	0.4758	−0.3617

		7 VarCov3	8 Threshd2	9 Threshd3
7	VarCov3	1.0000		
8	Threshd2	0.5942	1.0000	
9	Threshd3	0.6435	0.8735	1.0000

by MIXOR, the results indicate that the treatment groups do not significantly differ at baseline (TxDrug term is not significant), the placebo group does improve over time (SqrtWeek term is significant and negative), and that the drug group has greater improvement over time relative to the placebo group (Tx*Sweek term is significant and negative). The user should keep in mind that in this model with the Tx*Sweek interaction, the interpretation of the main effects (SqrtWeek and TxDrug) depends on the coding of these variables. To compare this model to a random-intercepts model (the model specified by the DEF file in Table 6a), the likelihood-ratio χ^2 test can be used. Based on this test, there is clear evidence of significant variation in the linear time-trends (likelihood-ratio $\chi^2 = -2[-1699.739 - (-1663.326)] = 72.83$, d.f. = 2, $P < 0.001$) over and above the individual intercept variation. Significant negative association between the intercept and linear time terms is indicated, suggesting that those patients with the highest initial severity show the greatest improvement across time (e.g. largest negative time-trends).

5. Hardware and software specifications

MIXOR is written in standard FORTRAN-77 with double arithmetic precision. It was originally developed for MS-DOS personal computers and later ported over to the MACINTOSH environment. As a result, it's use on MACINTOSH personal computers does not take advantage of the system's menu-orientated interface. For use in either the MS-DOS or MACINTOSH environment MIXOR requires a math coprocessor. The program stores all necessary matrices and vectors in a single one-dimensional array. Thus, there are no fixed limitations on the numbers of level-2 units, level-1 units, or model variables. MIXOR utilizes some MATCAL subroutines [37] for performing the matrix algebra operations.

6. Availability

The MIXOR program is available at no charge. Those interested in obtaining a copy of the program should contact the first author by electronic mail at HEDEKER@UIC.EDU or send a blank

diskette by normal mail to Ann Hohmann, Ph.D., M.P.H., NIMH Services Research Branch, 5800 Fishers Lane, Room #10C-06, Rockville MD, 20857.

Acknowledgments

The authors wish to thank Dr. R. Darrell Bock for many valuable discussions, comments, and suggestions during the preparation of this program; the authors also thank Dr. Bock for providing the code for the MATCAL [37] matrix algebra subroutines. The authors also thank Dave Patterson of Discerning Systems, Inc., for developing and creating the user interface, Dr. Brian Flay for providing the clustered data-set, Drs. John Davis and Nina Schooler for the longitudinal data-set, Ohidul Siddiqui for computer programming assistance, Dr. Lawrence DeCarlo for suggesting the complementary log-log response function, and to the associate editor and anonymous referee for their helpful comments. The development of MIXOR was supported by the National Institutes of Mental Health Grant MH44826-01A2, and University of Illinois at Chicago Prevention Research Center Developmental Project, CDC Grant R48/CCR505025.

References

[1] R.D. McKelvey and W. Zavoina, A statistical model for the analysis of ordinal level dependent variables, J. Math. Sociol. 4 (1975) 103–120.

[2] P. McCullagh, Regression models for ordinal data (with discussion), J. R. Stat. Soc. Ser. B 42 (1980) 109–142.

[3] H. Goldstein, Multilevel Statistical Models (Halsted Press, New York, 1995).

[4] A.S. Bryk and S.W. Raudenbush, Hierarchical Linear Models: Applications and Data Analysis Methods (Sage Publications, Inc., Newbury Park, CA, 1992).

[5] N.T. Longford, Random Coefficient Models (Oxford University Press, New York, 1993).

[6] N.M. Laird and J.H. Ware, Random-effects models for longitudinal data, Biometrics 40 (1982) 961-971.

[7] R.D. Bock, Within-subject experimentation in psychiatric research, in: Statistical and Methodological Advances in Psychiatric Research, Eds. R.D. Gibbons and M. Dysken, pp. 59–90 (Spectrum, New York, 1983)

[8] A.P. Dempster, D.B. Rubin and R.K. Tsutakawa, Estimation in covariance components models, J. Am. Stat. Assoc. 76 (1981) 341-353.

[9] R.D. Bock, Measurement of human variation: a two stage model, in: Multilevel Analysis of Educational Data, Ed. R.D. Bock, pp. 319–342 (Academic Press, New York, 1989).

[10] J. DeLeeuw and I. Kreft, Random coefficient models for multilevel analysis, J. Educ. Stat. 11 (1986) 57–85.

[11] N.T. Longford, A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested effects, Biometrika 74 (1987) 817–827.

[12] J.F. Strenio, H.I. Weisberg and A.S. Bryk, Empirical Bayes estimation of individual growth curve parameters and their relationship to covariates, Biometrika 39 (1983) 71–86.

[13] R.I. Jennrich and M.D. Schluchter, Unbalanced repeated-measures models with structured covariance matrices, Biometrics 42 (1986) 805–820.

[14] R.D. Bock, The discrete Bayesian, in: Principles of Modern Psychological Measurement, Eds. H. Wainer and S. Messick, pp. 103–115 (Earlbaum, Hillsdale, NJ, 1983).

[15] R.D. Gibbons, D. Hedeker, C. Waternaux and J.M. Davis, Random regression models: A comprehensive approach to the analysis of longitudinal psychiatric data, Psychopharmacol. Bull. 24 (1988) 438–443.

[16] D. Hedeker, R.D. Gibbons, C. Waternaux and J.M. Davis, Investigating drug plasma levels and clinical response using random regression models, Psychopharmacol. Bull. 25 (1989) 227–231.

[17] D. Hedeker, R.D. Gibbons and J.M. Davis, Random regression models for multicenter clinical trials data, Psychopharmacol. Bull. 27 (1991) 73–77.

[18] R.D. Gibbons, D. Hedeker, I. Elkin, C. Waternaux, H.C. Kraemer, J.B. Greenhouse, M.T. Shea, S.D. Imber, S.M. Sotsky and J.T. Watkins, Some conceptual and statistical issues in analysis of longitudinal psychiatric data, Arch. Gen. Psychiatry 50 (1993) 739–750.

[19] R. Stiratelli, N.M. Laird and J.H. Ware, Random-effects models for serial observations with binary response, Biometrics 40 (1984) 961–971.

[20] D.A. Anderson and M. Aitkin, Variance component models with binary response: interviewer variability, J. R. Stat. Soc. Ser. B 47 (1985) 203–210.

[21] G.Y. Wong and W.M. Mason, The hierarchical logistic regression model for multilevel analysis, J. Am. Stat. Assoc. 80 (1985) 513–524.

[22] R.D. Gibbons and R.D. Bock, Trend in correlated proportions, Psychometrika 52 (1987) 113–124.

[23] M.R. Conaway, Analysis of repeated categorical measurements with conditional likelihood methods, J. Am. Stat. Assoc. 84 (1989) 53–61.

[24] D.A. Harville and R.W. Mee, A mixed-model procedure for analyzing ordered categorical data, Biometrics 40 (1984) 393–408.

[25] J. Jansen, On the statistical analysis of ordinal data when extravariation is present, Appl. Stat. 39 (1990) 75–84.

[26] F. Ezzet and J. Whitehead, A random effects model for ordinal responses from a crossover trial, Stat. Med. 10 (1991) 901–907.

[27] D. Hedeker and R.D. Gibbons, A random-effects ordinal regression model for multilevel analysis, Biometrics 50 (1994) 933–944.

[28] J. Rasbash, M. Wang, G. Woodhouse and H. Goldstein, MLn: command reference guide (Institute of Education, University of London, London, UK, 1995).

[29] N.T. Longford, VARCL — interactive software for variance component analysis: applications for survey data, Prof. Stat. 5 (1986) 28–32.

[30] EGRET computer software program (Statistics and Epidemiology Research Corporation, Seattle, WA).

[31] R.D. Bock, R.D. Gibbons and E. Muraki, Full-information item factor analysis, Appl. Psychol. Meas. 12 (1988) 261–280.

[32] A. Wald, Tests of statistical hypotheses concerning several parameters when the number of observations is large, Trans. Am. Math. Soc. 54 (1943) 426–482.

[33] D. Hedeker, R.D. Gibbons and B.R. Flay, Random-effects regression models for clustered data: with an example from smoking prevention research, J. Consult. Clin. Psychol. 62 (1994) 757–765.

[34] 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, Prev. Med. 17 (1989) 585–607.

[35] M. Lorr and C.J. Klett, Inpatient multidimensional psychiatric scale: manual (Consulting Psychologists Press, Palo Alto, CA, 1966).

[36] R.D. Gibbons and D. Hedeker, Application of random-effects probit regression models, J. Consult. Clin. Psychol. 62 (1994) 285-296.

[37] R.D. Bock and B.H. Repp, MATCAL: Double Precision Matrix Operations Subroutines (National Educational Resources, Chicago, IL, 1974).