

MIXPREG: a computer program for mixed-effects Poisson regression

Donald Hedeker

Division of Epidemiology and Biostatistics

& Health Policy and Research Centers

School of Public Health

University of Illinois at Chicago

Correspondence to Dr. Donald Hedeker, Division of Epidemiology & Biostatistics (M/C 922), School of Public Health, University of Illinois at Chicago, 2121 West Taylor Street, Room 510, Chicago, IL, 60612-7260. Phone: (312) 996-4896.

E-MAIL: hedeker@uic.edu

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

Abstract

MIXPREG provides maximum marginal likelihood estimates for mixed-effects Poisson regression analysis. These models can be used for analysis of correlated count data, for example, data arising from a clustered 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. MIXPREG uses marginal maximum likelihood estimation, utilizing a Newton-Raphson iterative solution. Specifically, 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 MIXPREG are provided.

Keywords: count data; poisson regression; heterogeneity; clustering; multilevel data; random effects; correlated responses

1 Introduction

Count data are common in some fields of research. For example, outcome variables such as “frequency of service use” or “number of hospitalizations” are typical of count data that are obtained in mental health services research. While one could analyze these outcomes using models for normally-distributed outcomes, the preponderance of zeros and small values and the clearly discrete nature of the dependent variable suggest that a model that accounts for these characteristics is preferable and more appropriate. The Poisson regression model provides an attractive solution for the analysis of count data if observations are independent, *i.e.*, not longitudinal or clustered, ([1]).

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 Poisson 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 [2, 3, 4]. 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.

There has been much work done on mixed-effects models for continuous responses, and an increasing amount of work has focused on mixed-effects models for non-continuous response data. In this regard, mixed-effect models for both dichotomous [5, 6, 7, 8, 9, 10] and ordinal [11, 12, 13, 14, 15, 16] responses have been described. For count data, various types of Poisson random-effects models have been proposed. A recent review of some of these methods applied to longitudinal

Poisson data is given in Stukel [17]. Thall [18] proposed a mixed Poisson-gamma regression model for analyzing longitudinal interval count data. In this model, a single random effect is included to account for the correlated responses within an individual, and this random effect is distributed as a gamma distribution in the population of subjects. Other similar Poisson random-effects models are discussed by Albert [19], and Lawless [20]. These procedures generally do not allow for multiple random effects, and they do not add the random effects on the same scale as the fixed effects. As mentioned by Diggle, Liang, and Zeger [21] adding multiple normally distributed random effects on the same scale as the fixed effects of the Poisson regression model provides a more general and flexible model, however, these authors only outline an approximate maximum likelihood solution for this model.

This paper describes the FORTRAN program MIXPREG (mixed-effects Poisson regression) for the analysis of repeated or clustered count data. MIXPREG can accommodate multiple random effects, and allows for a general form for model covariates. A maximum marginal likelihood solution is implemented for parameter estimation using multi-dimensional quadrature to numerically integrate over the distribution of random-effects. A Newton-Raphson iterative 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 MIXPREG for count data.

Very little commercially-based software exists to perform mixed-effects regression analysis for count data. Recently, the SAS macro program GLIMMIX has been made available to perform such analysis. Unlike GLIMMIX which uses an approximate Taylor expansion to linearize the non-linear model, MIXPREG uses a full-information maximum likelihood approach.

2 Computational Methods

Siddiqui and Hedeker [22] describe the statistical development of the mixed-effects Poisson regression model. Here, we will present the key computational features.

Using the terminology of multilevel analysis [2] 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 t_{ik} represent the follow-up time associated with units i and k . Let y_{ik} be the value of the count variable (where y_{ik} can equal $0, 1, \dots$) associated with level-2 unit i and level-1 unit k . If this count is assumed to be drawn from a Poisson distribution, then the random-effects Poisson regression model indicates the expected number of counts in t_{ik} as:

$$\lambda_{ik} = t_{ik} \exp z_{ik} \quad (1)$$

with

$$z_{ik} = \mathbf{x}'_{ik} \boldsymbol{\beta}_i + \mathbf{w}'_{ik} \boldsymbol{\alpha} \quad (2)$$

where \mathbf{w}_{ik} is the $p \times 1$ covariate vector and \mathbf{x}_{ik} is the design vector for the r random effects, both vectors being for the k th level-1 unit nested within level-2 unit i . Also, $\boldsymbol{\alpha}$ is the $p \times 1$ vector of unknown fixed regression parameters, and $\boldsymbol{\beta}_i$ is the $r \times 1$ 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 $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}_\beta$. Since the level-2 subscript i is present for the \mathbf{x} vector, not all level-2 units are assumed to have the same number of level-1 observations nested within.

Assuming the Poisson process for the count y_{ik} , the probability, for a given level-2 unit i , that $Y_{ik} = y_{ik}$ (Y is a random variable whose realization in data is y), conditional on $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, is given as:

$$P(Y_{ik} = y_{ik} \mid \boldsymbol{\beta}, \boldsymbol{\alpha}) = \exp(-\lambda_{ik}) \frac{(\lambda_{ik})^{y_{ik}}}{y_{ik}!} \quad (3)$$

Letting \mathbf{y}_i denote the vector of counts from level-2 unit i (for the n_i level-1 units nested within), the probability of any \mathbf{y}_i , given $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, is equal to the product of the probabilities of the level-1 responses:

$$\begin{aligned}
\ell(\mathbf{y}_i | \boldsymbol{\beta}; \boldsymbol{\alpha}) &= \prod_{k=1}^{n_i} \left[\exp(-\lambda_{ik}) \frac{(\lambda_{ik})^{y_{ik}}}{y_{ik}!} \right] \\
&= \exp \left[- \sum_{k=1}^{n_i} \lambda_{ik} + \sum_{k=1}^{n_i} y_{ik} \log t_{ik} + \sum_{k=1}^{n_i} y_{ik} z_{ik} - \sum_{k=1}^{n_i} \log(y_{ik}!) \right] \quad (4)
\end{aligned}$$

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 $g(\cdot)$:

$$h(\mathbf{y}_i) = \int_{\boldsymbol{\beta}} \ell(\mathbf{y}_i | \boldsymbol{\beta}; \boldsymbol{\alpha}) g(\boldsymbol{\beta}) d\boldsymbol{\beta}$$

where $g(\boldsymbol{\beta})$ represents the distribution of the $\boldsymbol{\beta}$ vector in the population. Following Gibbons and Bock [8], it is useful to orthogonally transform the response model. For this, let $\boldsymbol{\beta}_i = \mathbf{T}\boldsymbol{\theta}_i + \boldsymbol{\mu}$, where $\mathbf{T}\mathbf{T}' = \boldsymbol{\Sigma}_\beta$ is the Cholesky decomposition of $\boldsymbol{\Sigma}_\beta$. The reparameterized model is then

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

The marginal density is then $h(\mathbf{y}_i) = \int_{\boldsymbol{\theta}} \ell(\mathbf{y}_i | \boldsymbol{\theta}; \boldsymbol{\alpha}, \boldsymbol{\mu}, \mathbf{T}) g(\boldsymbol{\theta}) d\boldsymbol{\theta}$, where $g(\boldsymbol{\theta})$ represents the multivariate standard normal density.

For estimation of the covariate coefficients $\boldsymbol{\alpha}$, and the population parameters $\boldsymbol{\mu}$ and \mathbf{T} , the marginal log-likelihood for the patterns from the N level-2 units can be written as $\log L = \sum_{i=1}^N \log h(\mathbf{y}_i)$. Let $\boldsymbol{\eta}$ represent the vector of parameters obtained by stacking $\boldsymbol{\alpha}$, $\boldsymbol{\mu}$, and $\mathbf{v}(\mathbf{T})$ (the vector containing the unique elements of the Cholesky factor \mathbf{T}). Differentiating the marginal log-likelihood yields:

$$\begin{aligned}
\frac{\partial \log L}{\partial \boldsymbol{\eta}} &= \sum_{i=1}^N h^{-1}(\mathbf{y}_i) \int_{\boldsymbol{\theta}} \frac{\partial \log \ell(\mathbf{y}_i | \boldsymbol{\theta}; \boldsymbol{\alpha}, \boldsymbol{\mu}, \mathbf{T})}{\partial \boldsymbol{\eta}} \ell(\mathbf{y}_i | \boldsymbol{\theta}; \boldsymbol{\alpha}, \boldsymbol{\mu}, \mathbf{T}) g(\boldsymbol{\theta}) d\boldsymbol{\theta} \\
&= \sum_{i=1}^N h^{-1}(\mathbf{y}_i) \int_{\boldsymbol{\theta}} \left[\sum_{k=1}^{n_i} (y_{ik} - \lambda_{ik}) \frac{\partial z_{ik}}{\partial \boldsymbol{\eta}} \right] \ell(\mathbf{y}_i | \boldsymbol{\theta}; \boldsymbol{\alpha}, \boldsymbol{\mu}, \mathbf{T}) g(\boldsymbol{\theta}) d\boldsymbol{\theta}
\end{aligned}$$

where $\partial z_{ik} / \partial \boldsymbol{\eta}$ is obtained by stacking the following vectors:

$$\frac{\partial z_{ik}}{\partial \boldsymbol{\alpha}} = \mathbf{w}_{ik} \quad \frac{\partial z_{ik}}{\partial \boldsymbol{\mu}} = \mathbf{x}_{ik} \quad \frac{\partial z_{ik}}{\partial (\mathbf{v}(\mathbf{T}))} = \mathbf{J}_r(\boldsymbol{\theta}_i \quad \mathbf{x}_{ik})$$

noting that \mathbf{J}_r is a transformation matrix [23] eliminating elements above the main diagonal.

For a Newton-Raphson solution, the matrix of second derivatives must also be derived. For this, denote $h_i = h(\mathbf{y}_i)$, $\ell_i = \ell(\mathbf{y}_i \mid \boldsymbol{\theta}; \boldsymbol{\alpha}, \boldsymbol{\mu}, \mathbf{T})$, and

$$\frac{\partial \log L}{\partial \boldsymbol{\eta}} = \sum_{i=1}^N h_i^{-1} \mathbf{d}_i$$

where

$$\mathbf{d}_i = \int_{\boldsymbol{\theta}} \frac{\partial \log \ell_i}{\partial \boldsymbol{\eta}} \ell_i g(\boldsymbol{\theta}) d\boldsymbol{\theta} .$$

Notice that \mathbf{d}_i is equal to $\partial h_i / \partial \boldsymbol{\eta}$, so that

$$\begin{aligned} \frac{\partial^2 \log L}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'} &= \sum_{i=1}^N h_i^{-2} \left[h_i \frac{\partial \mathbf{d}_i}{\partial \boldsymbol{\eta}} - \mathbf{d}_i \frac{\partial h_i}{\partial \boldsymbol{\eta}'} \right] \\ &= \sum_{i=1}^N h_i^{-2} \left[h_i \frac{\partial \mathbf{d}_i}{\partial \boldsymbol{\eta}} - \mathbf{d}_i \mathbf{d}_i' \right] , \end{aligned}$$

where

$$\frac{\partial \mathbf{d}_i}{\partial \boldsymbol{\eta}} = \int_{\boldsymbol{\theta}} \left[\left(\frac{\partial \log \ell_i}{\partial \boldsymbol{\eta}} \right) \left(\frac{\partial \log \ell_i}{\partial \boldsymbol{\eta}} \right)' + \frac{\partial^2 \log \ell_i}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'} \right] \ell_i g(\boldsymbol{\theta}) d\boldsymbol{\theta} .$$

Since the first derivatives are given as

$$\frac{\partial \log \ell_i}{\partial \boldsymbol{\eta}} = \sum_{k=1}^{n_i} (y_{ik} - \lambda_{ik}) \frac{\partial z_{ik}}{\partial \boldsymbol{\eta}} ,$$

the second derivatives are obtained as

$$\frac{\partial^2 \log \ell_i}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'} = - \sum_{k=1}^{n_i} \lambda_{ik} \frac{\partial z_{ik}}{\partial \boldsymbol{\eta}} \left(\frac{\partial z_{ik}}{\partial \boldsymbol{\eta}} \right)' .$$

The Newton-Raphson iterative procedure can now be implemented. Specifically, estimates for the vector of parameters $\boldsymbol{\eta}$, on iteration ι are improved by

$$\boldsymbol{\eta}_{\iota+1} = \boldsymbol{\eta}_{\iota} - \left[\frac{\partial^2 \log L}{\partial \boldsymbol{\eta}_{\iota} \partial \boldsymbol{\eta}_{\iota}'} \right]^{-1} \frac{\partial \log L}{\partial \boldsymbol{\eta}_{\iota}} \quad (5)$$

At convergence, the large-sample variance covariance matrix of the maximum likelihood estimator is then obtained as the inverse of the matrix of second derivatives. The MML estimates and their accompanying standard errors can then be used to construct asymptotic z -statistics by dividing the parameter estimate by its standard error [24]. 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 ($\boldsymbol{\mu}$ and $\boldsymbol{\alpha}$) is generally reasonable, for the variance and covariance components (\boldsymbol{T}) this practice is problematic (see Bryk and Raudenbush [3] page 55).

2.1 Numerical Quadrature

In order to solve the above likelihood equations, numerical integration on the transformed $\boldsymbol{\theta}$ space must be performed. For this, Gauss-Hermite quadrature can be used to approximate the above integrals to any practical degree of accuracy [25]. 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. 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 [26] 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.

2.2 Solution incorporating Level-2 Weights

The above solution can be modified to accommodate weighted data, which occurs when the same response pattern \boldsymbol{y}_i and covariate vector \boldsymbol{w}_i is observed for a number of

level-2 units. As an example of weighted data, consider the frequency table published by Ezzet and Whitehead [13] which is given below. This table lists the frequencies corresponding to categorical responses at two timepoints for two groups of subjects: group 1 and (group 2).

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)

Each cell of this table represents a number (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). As such, these frequencies represent weights which can be incorporated into a modeling of the responses across the two timepoints.

Another instance where a weighted solution is useful is when one wants to differentially weight subjects (in a longitudinal study) or clusters (in a clustered study) in order to make the sample more representative of a specified population. Variants of this type of weighting of observations has been used extensively in survey research [27].

For the weighted solution, express the weight for level-2 unit i as s_i . In the summation over i , the log-likelihood is then written as: $\log L = \sum_i^N s_i \log h(\mathbf{y}_i)$. The derivatives and matrix of second derivatives are then multiplied by s_i . Notice, that the unweighted solution is simply a restricted version of the weighted solution where $s_i = 1$.

3 Program description and usage

MIXPREG is currently available in executable form for MS-DOS, WINDOWS-95, and MACINTOSH computers. The MIXPREG instructions *must* be stored in the

file MIXPREG.DEF (described below), and the user begins program execution using the MIXPREG.EXE file (issuing the command MIXPREG in DOS, or double-clicking on the MIXPREG icon in WINDOWS-95 or on the MACINTOSH). Here, we will detail the procedure for running the program. MIXPREG makes use of the following files:

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

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

- MIXPREG.EST - a file containing the estimated parameters (with labels).
- MIXPREG.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 MIXPREG.EST (*i.e.*, no labels are given in MIXPREG.VAR).
- MIXPREG.RES - if there is only one random effect (either $R=1$, or $R>1$ and $VGRP=1$, see below) then a file containing empirical Bayes estimates of the random effect for each level-2 unit is produced. 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), and the posterior standard deviation. Additionally, if each level-2 unit has a frequency weight, 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, count 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 Count Xvector Wvector

where, *ID* refers to an the level-2 ID number which does not change across level-1 units, *Count* is the value of count variable 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, in particular, *missing values left as blank fields will definitely cause problems.*

3.2 Analysis options and settings - MIXPREG.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.

Insert Tables 1a, 1b, 1c, and 1d about here

Except where noted, this file is read in free format. This file is created by the user directly *before* typing the command MIXPREG (or for WINDOWS-95 or

MACINTOSH, double-clicking on the MIXPREG file). This filename and extension (MIXPREG.DEF) *must* be used and should be in the same directory as the program MIXPREG.EXE or accessible via appropriate PATH statements.

3.3 Main output file

This file contains descriptive information about the variables read in to MIXPREG, 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. An optional listing of the mean and standard deviation of *Count* by 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, MIXPREG indicates the number (and percentage) of level-2 units with non varying level-1 responses on *Count*.

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 [28], 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 [3] 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), MIXPREG 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.

3.4 Some Common MIXPREG Errors

There are a few errors which can prevent MIXPREG 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 MEANYX option (described

in Tables 1b and 1d) is fairly unforgiving. The values listed by the user for the levels of the covariate *must* be exactly the same as the values that are found in the data file. If a strange error prevents MIXPREG from running and this option is selected, the user can set MEANYX=0 to avoid this option. Third, the NPR option (described in Table 1b), which is used to list data to the screen, can cause MIXPREG 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. 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 1b) 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 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 covariates. 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 MIXPREG usage

MIXPREG can estimate a variety of models for correlated count data. Here, we will present an analysis of a longitudinal dataset where observations are clustered within subjects. A comparison to an analysis using a fixed-effects Poisson regression model, which ignores the clustering of the data, will illustrate the importance of taking the clustering of the data into account. In the first mixed-effects regression analysis, one random term is included to account for the clustering of observations within

subjects. This random subject term describes the way in which responses from the same subject are similar, relative to the sample as a whole. The possibility of including multiple random subject effects will be discussed as well. This example 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 MIXPREG.

The example presented here is termed a two-level model in the multilevel literature [2]. Specifically, repeated observations (level-1) are treated as being nested within subjects (level-2). At present, MIXPREG does not allow a three-level analysis which would consider, concurrently, observations nested within subjects *and* subjects nested within clusters (*e.g.*, clinics, hospitals, classrooms, or firms).

4.1 Analysis of a Longitudinal Dataset

The data for this example are taken from a paper by McKnight and Van Den Eeden [30] who report on the number of headaches in a two treatment, multiple period crossover trial. Specifically, the number of headaches per week was repeatedly measured for 27 subjects. Following a seven-day placebo run-in period, subjects received either aspartame or placebo in four seven-day treatment periods according to the double-blind crossover treatment design. Each treatment period was separated by a washout day. A partial list of these data is given in Table 2. The variables are, in order, subject ID, number of headaches during the week (from 0 to 7), a column of ones for the intercept, period1 (1 for the first treatment period and 0 otherwise), period2 (1 for the second treatment period and 0 otherwise), period3 (1 for the third treatment period and 0 otherwise), period4 (1 for the fourth treatment period and 0 otherwise), drug (0=placebo and 1=aspartame), the number of periods the individual was observed (from 2 to 5), and the number of treatment days in the period.

Insert Table 2 about here

4.1.1 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 data clustering. Using MIXPREG for this type of analysis is equivalent to performing a Poisson regression analysis treating all observations as independent observations. In this analysis, the number of headaches in a week is modeled in terms of four period effects and a drug effect. Table 3 lists the MIXPREG.DEF file for this analyses.

Insert Table 3 about here

Note that zero random effects are specified in the DEF file, and that blank lines are present for the records that define fields and labels of random effects. The results from this analysis are listed in Table 4.

Insert Table 4 about here

As seen in the output in Table 4, when zero random effects are requested, MIXPREG indicates the number of observations as number of level-1 observations. For these data, this number is simply the total number of observations. Descriptive statistics are listed for all variables. This analysis, which ignores the data clustering, indicates non-significant effects for all model covariates.

4.1.2 Mixed-effects Regression Including Data Clustering

To perform the mixed-effects analysis the subject ID (the first variable field in the datafile) is indicated as the level-2 ID on line 7 of the DEF file. **Note that the data must be sorted by this ID variable.** Tables 5 lists the MIXPREG.DEF file for the random-intercepts analysis. Again, the number of headaches in a week is modeled in terms of four period effects and a drug effect. However, in contrast to

the analysis of the previous section, a random subject effect (*i.e.*, the intercept) is included to account for the data clustering.

Insert Table 5 about here

Comparing Tables 5 and 3 illustrates the proper specifications for a random-intercepts model. In this case, the random intercept varies by subjects, that is, each subject has their own intercept. Since these intercepts are considered random effects, they are representative of a larger population of subject effects. MIXPREG estimates the variance of this population distribution, and provides empirical Bayes estimates of the random effects. At the conclusion of each MIXPREG run, the empirical Bayes estimates are output to the file MIXPREG.RES (not shown). Table 6 lists the MIXPREG results from the random-intercepts analysis.

Insert Table 6 about here

The output file given in Table 6 indicates that there are 27 subjects with 122 observations nested within. The numbers of observation per subject varies between 2 to 5. Following the listing of the starting values, MIXPREG indicates that of the 27 subjects, 2 had response vectors that were non-varying. Thus, 2 subjects gave identical responses at all timepoints that they were measured on.

The random-effect standard deviation is estimated as .643, and although 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 is questionable [3]. A likelihood ratio χ^2 test compares the difference in deviance ($-2\log L$) between nested models. For this, we get $462.323 - 406.205 = 56.118$, which on 1 degree of freedom is highly significant. Thus, there is ample evidence that the repeated observations within subjects are correlated. Regarding the regression coefficients, the conclusions are the same as was previously observed in the the analysis ignoring the clustering of the data (in Table 4), namely, all effects are non-significant. Thus,

for these data the conclusions do not change when the random subject effect is added to the model. Finally, the output provides the correlation matrix of the MML estimates. It is important to realize that these are not the correlations of the variables. These correlations of the parameter estimates can be used to assess collinearity problems in estimation.

4.1.3 Mixed-effects Analysis with an Offset Variable

The previous analyses have assumed that the counts were all observed for the same number of days. However, this was not the case since the the number of treatment days in the period (the 10th field of the input data file) did vary to some degree. Most of the counts were based on the full seven days in the week, however, some observations were made only for 1 day in the given week. To take this into account, we need to specify OFFSET=1 on line 7 of the DEF file. The offset variable indicates the amount of time that each count is based on. If OFFSET=0 is specified, then MIXPREG assumes that all counts are based on the same amount of time. Table 7 lists the MIXPREG.DEF file for an analysis including an offset variable.

Insert Table 7 about here

In addition to specifying OFFSET=1, the field for the offset variable (*i.e.*, field 10) is indicated, in this case, after indicating the fields for the fixed covariates.

Table 8 lists results from the analysis including the offset variable, number of treatment days in the period.

Insert Table 8 about here

Here, we see a marginally significant positive relationship between drug treatment and number of headaches. All time effects are again non-significant.

As mentioned, the empirical Bayes estimates of the random effects are written to the file MIXPREG.RES at the conclusion of the MIXPREG run. Table 9 lists

the contents of MIXPREG.RES for the model just considered (*i.e.*, the model with output in Table 8).

Insert Table 9 about here

The file MIXPREG.RES contains four pieces of information per individual: 1) the individual's ID, 2) the number of repeated observations for that individual, 3) the empirical Bayes estimate for that individual (which is the mean of the posterior distribution), and 4) the associated posterior standard deviation. Since they are estimates of θ for each individual, the empirical Bayes estimates are expressed on the standard normal scale. Inspection of these estimates indicates that subject 13 has a very high score. This person's estimate of 1.043 (with standard deviation .016) suggests a very high level of headaches. This agrees well with the raw data presented in Table 2 which reveals that this person scored a 7 on four occasions and a 6 on the only other occasion.

5 Hardware and software specifications

MIXPREG 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. MIXPREG utilizes some MATCAL subroutines [29] for matrix algebra operations.

6 Availability

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

7 Acknowledgements

Thanks are in order to Drs. R. Darrell Bock, Ohidul Siddiqui, Robert D. Gibbons, Paul Levy, 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 the MATCAL [29] subroutines. The development of MIXPREG was supported by the National Institutes of Mental Health Grant MH56146-01.

References

- [1] W.H. Greene, *Econometric Analysis*, Second Edition (Prentice Hall, Englewood Cliffs, NJ, 1993).
- [2] H. Goldstein, *Multilevel Statistical Models* (Halsted Press, New York, 1995).
- [3] A.S. Bryk and S.W. Raudenbush, *Hierarchical Linear Models: Applications and Data Analysis Methods* (Sage Publications, Inc., Newbury Park CA, 1992).
- [4] N.T. Longford, *Random Coefficient Models*, (Oxford University Press, New York, 1993).
- [5] R. Stiratelli, N.M. Laird, and J.H. Ware, Random-effects models for serial observations with binary response, *Biometrics* 40, (1984) 961-971.
- [6] 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.
- [7] 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.
- [8] R.D. Gibbons and R.D. Bock, Trend in correlated proportions, *Psychometrika* 52, (1987) 113-124.
- [9] M.R. Conaway, Analysis of repeated categorical measurements with conditional likelihood methods, *Journal of the American Statistical Association* 84, (1989) 53-61.
- [10] H. Goldstein, Nonlinear multilevel models, with an application to discrete response data, *Biometrika* 78, (1991) 45-51.
- [11] D.A. Harville and R.W. Mee, A mixed-model procedure for analyzing ordered categorical data, *Biometrics* 40, (1984) 393-408.

- [12] J. Jansen, On the statistical analysis of ordinal data when extravariation is present, *Applied Statistics* 39, (1990) 75-84.
- [13] F. Ezzet and J. Whitehead, A random effects model for ordinal responses from a crossover trial, *Statistics in Medicine* 10, (1991) 901-907.
- [14] D. Hedeker and R.D. Gibbons, A random-effects ordinal regression model for multilevel analysis, *Biometrics* 50, (1994) 933-944.
- [15] G. Molenberghs and E. Lesaffre, Marginal modeling of correlated ordinal data using a multivariate Plackett distribution, *Journal of the American Statistical Association* 89, (1994) 633-644.
- [16] T.R. Ten Have, A mixed effects model for multivariate ordinal response data including correlated failure times with ordinal responses, *Biometrics* 52, (1996) 473-491.
- [17] T.A. Stukel, Comparison of methods for the analysis of longitudinal interval count data, *Statistics in Medicine* 12, (1993) 1339-1351.
- [18] P.F. Thall, Mixed Poisson likelihood regression models for longitudinal interval count data, *Biometrics* 44, (1988) 197-209.
- [19] J. Albert, A Bayesian analysis of a Poisson random effects model for home run hitters, *The American Statistician* 46, (1992) 246-253.
- [20] J.F. Lawless, Negative binomial and mixed Poisson regression, *Canadian Journal of Statistics* 15, (1987) 209-225.
- [21] P.J. Diggle, K.Y. Liang, and S.L. Zeger, *Analysis of Longitudinal Data* (Oxford University Press, New York, 1994).
- [22] O. Siddiqui and D. Hedeker, Poisson random-effects regression models for correlated count data with applications, Technical Report (1997).
- [23] J.R. Magnus, *Linear Structures* (Charles Griffin, London UK, 1988)

- [24] 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.
- [25] A.H. Stroud and D. Secrest, *Gaussian Quadrature Formulas* (Prentice Hall, Englewood Cliffs, NJ, 1966).
- [26] R.D. Bock, R.D. Gibbons, and E. Muraki, Full-information item factor analysis, *Applied Psychological Measurement* 12, (1988) 261-280.
- [27] D. Pfeffermann, The role of sampling weights when modeling survey data, *International Statistical Review* 61, (1993) 317-337.
- [28] S.D. Silvey, *Statistical Inference* (Chapman and Hall, New York, 1975).
- [29] R.D. Bock and B.H. Repp, *MATCAL: Double Precision Matrix Operations Subroutines*, (National Educational Resources, Chicago, 1974).
- [30] B. McKnight and S.K. Van Den Eeden, A conditional analysis for two-treatment multiple period crossover designs with binomial or Poisson outcomes and subjects who drop out, *Statistics in Medicine* 12, (1993) 825-834.

Table 1a

Analysis options and settings specified in MIXPREG.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.

Table 1b

Analysis options and settings specified in MIXPREG.DEF: line 6

Line 6 - NPR NF R P CONV MISS START WT MEANS PRIOR UNIF NQUAD
LINFN DIAG NOMU VGRP OFFSET

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 of the random effects).

CONV = convergence criterion (*usually set to .001 or .0001*).

MISS = 0 if no missing values are present in the data, or 1 if missing values are present (codes 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 is weighted equally, or 1 for differential weighting.

MEANS = 0 if a table of means of the dependent variable broken down by levels of another variable is not requested, and 1 if such a table is requested.

PRIOR = 0 or 1, respectively, for a specified form (see UNIF) or empirically-determined distributional form for the random-effects distribution. (*This is an option that is under development; at present IPRIOR=0 is the only possible choice*).

UNIF = 0 or 1 for a normal or uniform distribution, respectively, for the assumed random-effects distribution (*ordinarily set to 0*).

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*).

Table 1b (continued)

LINFN = number of linear transforms of the estimated parameters to estimate
(*ordinarily set to 0*).

DIAG = 0 for correlated random effects or 1 for independent random effects (*ordinarily set to 0*).

NOMU = 0 to estimate the mean of the random effects or 1 to fix them to zero
(*ordinarily set to 0*).

VGRP = 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 (*ordinarily set to 0*). 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.

OFFSET = 0 if there is no offset variable, or 1 if there is an offset variable. An offset variable is used if the counts are not all based on the same period of time - in this case the offset variable indicates the period of time that each of the counts is based on (*ordinarily set to 0*).

Table 1c

Analysis options and settings specified in MIXPREG.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 (count) 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 1d

Analysis options and settings specified in MIXPREG.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 - (*if OFFSET = 1*) - the field of the input data file which contains the offset variable.

next line - (*if MEANS = 1*) - two parameters and a list of values: the field of the input data file which contains the variable for which means of the dependent variable are to be broken down by, followed by (a) the number of categories for this variable, and (b) a list of the category values.

next line (*if MISS = 1*) - missing value code for the dependent (count) 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 dependent (count) variable.

next line - R labels for the random effects in 8 character width fields.

next line (*if START = 1 and NOMU = 0*) - R 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 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, *e.g.*, for a 2 x 2 covariance matrix, the order of the starting values should be: variance(1), covariance(1,2) and variance(2). Note: if either DIAG=1 or VGRP=1 then only R starting values are needed.

Table 1d (continued)

final lines - (if *LINFN* > 0) - *LINFN* by *NPAR* coefficients for the linear re-expressions of the estimated parameters. The *NPAR* terms in order are

- *R* elements of the random-effect mean vector (unless *NOMU*=1).
- *P* effects of explanatory variables.
- random-effect variance covariance matrix:
 - if (*DIAG*=0 and *VGRP*=0): ($R \times (R+1)/2$) unique elements of the random-effect variance-covariance matrix (in packed form).
 - if (*DIAG*=1 or *VGRP*=1): *R* variance-covariance terms.

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.

Table 2

Data from example 4.1: First 12 subjects with 2 to 5 repeated observations

2	0	1	0	0	0	0	0	3	7
2	5	1	1	0	0	0	1	3	7
2	2	1	0	1	0	0	0	3	7
5	3	1	0	0	0	0	0	5	7
5	0	1	1	0	0	0	1	5	7
5	2	1	0	1	0	0	0	5	7
5	0	1	0	0	1	0	1	5	7
5	0	1	0	0	0	1	0	5	7
13	7	1	0	0	0	0	0	5	7
13	7	1	1	0	0	0	1	5	7
13	7	1	0	1	0	0	0	5	7
13	6	1	0	0	1	0	1	5	7
13	7	1	0	0	0	1	0	5	7
16	1	1	0	0	0	0	0	3	7
16	3	1	1	0	0	0	1	3	7
16	1	1	0	1	0	0	0	3	7
19	0	1	0	0	0	0	0	5	7
19	1	1	1	0	0	0	1	5	7
19	1	1	0	1	0	0	0	5	7
19	0	1	0	0	1	0	1	5	7
19	1	1	0	0	0	1	0	5	7
23	7	1	0	0	0	0	0	5	7
23	2	1	1	0	0	0	1	5	7
23	3	1	0	1	0	0	0	5	7
23	3	1	0	0	1	0	1	5	7
23	2	1	0	0	0	1	0	5	7
25	1	1	0	0	0	0	0	5	7
25	6	1	1	0	0	0	1	5	7
25	1	1	0	1	0	0	0	5	7
25	7	1	0	0	1	0	1	5	7
25	0	1	0	0	0	1	0	5	7
1	3	1	0	0	0	0	0	5	7
1	0	1	1	0	0	0	1	5	7
1	3	1	0	1	0	0	0	5	7
1	1	1	0	0	1	0	0	5	7
1	0	1	0	0	0	1	1	5	7
3	2	1	0	0	0	0	0	5	7
3	2	1	1	0	0	0	1	5	7
3	3	1	0	1	0	0	0	5	7
3	2	1	0	0	1	0	0	5	7
3	2	1	0	0	0	1	1	5	7
6	1	1	0	0	0	0	0	5	7
6	1	1	1	0	0	0	1	5	7
6	0	1	0	1	0	0	0	5	7
6	3	1	0	0	1	0	0	5	7
6	1	1	0	0	0	1	1	5	7
9	2	1	0	0	0	0	0	4	7
9	2	1	1	0	0	0	1	4	7
9	0	1	0	1	0	0	0	4	7
9	1	1	0	0	1	0	0	4	5
17	4	1	0	0	0	0	0	2	7
17	1	1	1	0	0	0	1	2	1

Table 3

MIXPREG.DEF file for example 4.1.1: fixed-effects model ignoring clustering

Aspartame Data - Repeated Headaches across Time
fixed intercept and 5 covariates

aspart0.dat

aspart0.out

aspart0.def

1 10 0 6 0.0001 0 0 0 0 0 0 20 0 0 0 0 0
1 2

3 4 5 6 7 8

HeadAche

IntercptPeriod1 Period2 Period3 Period4 DrugAsp

Table 4
 output file for example 4.1.1: fixed-effects model ignoring clustering

 MIXPREG - The program for mixed-effects poisson regression analysis

Aspartame Data - Repeated Headaces across Time
 fixed intercept and 5 covariates

Numbers of observations

Level 1 observations = 122

Descriptive statistics for all variables

Variable	Minimum	Maximum	Mean	Stand. Dev.
HeadAche	0.00000	7.00000	1.68033	1.88630
Intercpt	1.00000	1.00000	1.00000	0.00000
Period1	0.00000	1.00000	0.22131	0.41684
Period2	0.00000	1.00000	0.20492	0.40531
Period3	0.00000	1.00000	0.18033	0.38605
Period4	0.00000	1.00000	0.17213	0.37905
DrugAsp	0.00000	1.00000	0.38525	0.48866

Starting values

covariates 0.7963 0.0936 0.0165 -0.1728 -0.1477 0.1982

 * Final Results - Maximum Marginal Likelihood Estimates *

Total Iterations = 4
 Log Likelihood = -231.162
 Deviance (-2logL) = 462.323
 Ridge = 0.000

Variable	Estimate	Stand. Error	Z	p-value
Intercpt	0.46536	0.15250	3.05159	0.00228 (2)
Period1	0.09160	0.22653	0.40434	0.68596 (2)
Period2	0.01308	0.22758	0.05746	0.95418 (2)
Period3	-0.22446	0.24709	-0.90842	0.36366 (2)
Period4	-0.18397	0.25402	-0.72421	0.46894 (2)
DrugAsp	0.23318	0.15958	1.46120	0.14396 (2)

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

-

Correlation of the Maximum Marginal Likelihood Estimates

		1	2	3	4	5	6
		Intercpt	Period1	Period2	Period3	Period4	DrugAsp
1	Intercpt	1.0000					
2	Period1	-0.6732	1.0000				
3	Period2	-0.6701	0.6017	1.0000			
4	Period3	-0.6172	0.5454	0.5187	1.0000		
5	Period4	-0.6003	0.5741	0.5399	0.4892	1.0000	
6	DrugAsp	0.0000	-0.4312	-0.3492	-0.3012	-0.3941	1.0000

Table 5

MIXPREG.DEF file for example 4.1.2: random-intercepts model

Aspartame Data - Repeated Headaches across Time
random intercept and 5 covariates

aspart0.dat

aspart.out

aspart.def

1 10 1 5 0.0001 0 0 0 0 0 0 20 0 0 0 0 0

1 2

3

4 5 6 7 8

HeadAche

Intercpt

Period1 Period2 Period3 Period4 DrugAsp

Table 6
 output file for example 4.1.2: random-intercepts model

 MIXPREG - The program for mixed-effects poisson regression analysis

Aspartame Data - Repeated Headaches across Time
 random intercept and 5 covariates

Random-effects distribution: normal

Numbers of observations

Level 1 observations = 122
 Level 2 observations = 27

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

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

Descriptive statistics for all variables

Variable	Minimum	Maximum	Mean	Stand. Dev.
HeadAche	0.00000	7.00000	1.68033	1.88630
Intercpt	1.00000	1.00000	1.00000	0.00000
Period1	0.00000	1.00000	0.22131	0.41684
Period2	0.00000	1.00000	0.20492	0.40531
Period3	0.00000	1.00000	0.18033	0.38605
Period4	0.00000	1.00000	0.17213	0.37905
DrugAsp	0.00000	1.00000	0.38525	0.48866

Starting values

mean 0.7358
 covariates 0.0901 0.0543 -0.1622 -0.0964 0.1364
 var. terms 0.2639

=> The number of level 2 observations with non-varying responses
 = 2 (7.41 percent)

 * Final Results - Maximum Marginal Likelihood Estimates *

Total Iterations = 34
 Quad Pts per Dim = 20
 Log Likelihood = -203.102
 Deviance (-2logL) = 406.205
 Ridge = 0.000

Variable	Estimate	Stand. Error	Z	p-value
Intercpt	0.24035	0.14955	1.60713	0.10803 (2)
Period1	0.08031	0.23458	0.34236	0.73208 (2)
Period2	0.03412	0.22142	0.15410	0.87753 (2)
Period3	-0.22923	0.24952	-0.91867	0.35827 (2)
Period4	-0.16071	0.24717	-0.65019	0.51557 (2)
DrugAsp	0.21536	0.16278	1.32303	0.18582 (2)

random effect variance term: expressed as a standard deviation

Intercpt	0.64277	0.05511	11.66395	0.00000 (1)
----------	---------	---------	----------	-------------

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

 Correlation of the Maximum Marginal Likelihood Estimates

	1	2	3	4	5	6
	Intercpt	Period1	Period2	Period3	Period4	DrugAsp
1 Intercpt	1.0000					
2 Period1	-0.6600	1.0000				
3 Period2	-0.6988	0.5885	1.0000			
4 Period3	-0.6063	0.5718	0.5156	1.0000		
5 Period4	-0.6170	0.5780	0.5205	0.4877	1.0000	
6 DrugAsp	0.0003	-0.4913	-0.2852	-0.3550	-0.3596	1.0000
7 VarCov1	-0.2399	-0.0106	0.0080	-0.0239	-0.0040	-0.0027
		7				
		VarCov1				
7 VarCov1		1.0000				

Table 7

MIXPREG.DEF file for example 4.1.3: random-intercepts model with offset

Aspartame Data - Repeated Headaches across Time
random intercept and 5 covariates - WITH OFFSET

asparto.dat

asparto.out

asparto.def

1 10 1 5 0.0001 0 0 0 0 0 0 20 0 0 0 0 1

1 2

3

4 5 6 7 8

10

HeadAche

Intercpt

Period1 Period2 Period3 Period4 DrugAsp

Table 8

output file for example 4.1.3: random-intercepts model with offset

 MIXPREG - The program for mixed-effects poisson regression analysis

Aspartame Data - Repeated Headaches across Time
 random intercept and 5 covariates - WITH OFFSET

Random-effects distribution: normal

Numbers of observations

Level 1 observations = 122
 Level 2 observations = 27

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

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

Descriptive statistics for all variables

Variable	Minimum	Maximum	Mean	Stand. Dev.
HeadAche	0.00000	7.00000	1.68033	1.88630
Intercpt	1.00000	1.00000	1.00000	0.00000
Period1	0.00000	1.00000	0.22131	0.41684
Period2	0.00000	1.00000	0.20492	0.40531
Period3	0.00000	1.00000	0.18033	0.38605
Period4	0.00000	1.00000	0.17213	0.37905
DrugAsp	0.00000	1.00000	0.38525	0.48866
Offset	1.00000	7.00000	6.79508	0.95293

Starting values

mean 0.7358
 covariates 0.0901 0.0543 -0.1622 -0.0964 0.1364
 var. terms 0.2639

=> The number of level 2 observations with non-varying responses
 = 2 (7.41 percent)

 * Final Results - Maximum Marginal Likelihood Estimates *

Total Iterations = 116
 Quad Pts per Dim = 20
 Log Likelihood = -205.425
 Deviance (-2logL) = 410.851
 Ridge = 0.000

Variable	Estimate	Stand. Error	Z	p-value
Intercpt	-1.45244	0.14344	-10.12574	0.00000 (2)
Period1	0.11789	0.23074	0.51092	0.60940 (2)
Period2	0.10988	0.21940	0.50079	0.61652 (2)
Period3	-0.16975	0.24720	-0.68669	0.49228 (2)
Period4	-0.04373	0.24634	-0.17753	0.85910 (2)
DrugAsp	0.28106	0.16146	1.74071	0.08173 (2)

random effect variance term: expressed as a standard deviation
 Intercpt 1.07699 0.08957 12.02414 0.00000 (1)

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

Correlation of the Maximum Marginal Likelihood Estimates

	1	2	3	4	5	6
	Intercpt	Period1	Period2	Period3	Period4	DrugAsp
1 Intercpt	1.0000					
2 Period1	-0.6822	1.0000				
3 Period2	-0.7093	0.5818	1.0000			
4 Period3	-0.6149	0.5659	0.5054	1.0000		
5 Period4	-0.6191	0.5776	0.5142	0.4825	1.0000	
6 DrugAsp	0.0063	-0.4821	-0.2946	-0.3639	-0.3851	1.0000
7 VarCov1	-0.1335	0.0061	0.0593	0.0305	0.0445	0.0304
		7				
		VarCov1				
7 VarCov1		1.0000				

Table 9

Empirical Bayes estimates from example 4.1.3: all 27 subjects

2	3	0.113087	0.348838
5	5	-0.550140	0.367331
13	5	1.043263	0.016471
16	3	-0.203469	0.401517
19	5	-0.917724	0.449395
23	5	0.383281	0.165406
25	5	0.344636	0.118463
1	5	-0.321997	0.308276
3	5	0.166811	0.306927
6	5	-0.427029	0.322937
9	4	-0.334207	0.378246
17	2	0.656546	0.424522
18	5	-0.922774	0.449084
21	5	0.343272	0.117564
22	5	-0.720812	0.415130
7	5	0.167305	0.306656
10	5	-1.788796	0.607237
11	5	-0.460461	0.360827
14	5	-0.321611	0.308419
24	5	-0.720412	0.415103
27	5	0.244487	0.251424
4	5	-1.785018	0.607636
8	5	-0.618974	0.406641
12	5	-0.549779	0.367276
15	5	-0.316687	0.310265
20	3	0.462905	0.335041
26	2	-0.813338	0.595836
