ARTICLES FROM THE SCIP CONFERENCE

Fitting mixed-effects models for repeated ordinal outcomes with the NLMIXED procedure

CHING-FAN SHEU
DePaul University, Chicago, Illinois

This paper presents an analysis of repeated ordinal outcomes arising from two psychological studies. The first case is a repeated measures analysis of variance; the second is a mixed-effects regression in a longitudinal design. In both, the subject-specific variation is modeled by including random effects in the linear predictor (inside a link function) of a generalized linear model. The NLMIXED procedure in SAS is used to fit the mixed-effects models for the categorical response data. The presentation emphasizes the parallel between the model specifications and the SAS statements. The purpose of this paper is to facilitate the use of mixed-effects models in the analysis of repeated ordinal outcomes.

Ordered categorical variables play an important role in psychological studies in which precise measurement is not always possible. For example, in a sensory experiment on the bitter taste of wine, the subject is asked to give his or her rating on a 5-point scale ranging from *least* to *most bitter*. A more exact measurement of the bitterness of wine may not be meaningful or necessary. It is commonly assumed that the subject can employ the same rating scale consistently from one trial to another. Less plausible is the assumption that all subjects have the same discriminability for the bitter taste of wine or that subjects use the same criteria for responding. What is most bitter for one subject may be only moderately so for another.

A within-subjects design in which subjects are asked to provide responses across several experimental conditions offers the investigator the opportunity to examine between-subjects variation in setting response thresholds. For example, the thresholds of each subject for the bitterness of the wine can be modeled by adding a subject-specific term into a model that relates the ordinal scale of measurement to the treatment factors. These subject-specific terms can be conceptualized as fixed effects or as random effects. In fixed-effects models, each subject is allowed to have a constant but unknown amount of shift in thresholds with respect to the reference subject. By contrast, random-effects models account for the between-subjects variation in thresholds by assuming that these unobserved subject-specific terms vary randomly among a sample of subjects. In other words, one assumes that

the subjects are a random sample out of a certain population, and the scope of the inferences is about the overall characteristics of this population.

Owing to the flexibility of the normal distribution, the use of random effects in linear mixed models for continuous responses is well established (Ware, 1985)—for example, random-effects analyses of variance. For categorical response data, the heterogeneous subject effects are often ignored, because introducing random effects greatly complicates model fitting. If the subject gives only a single response or the number of subjects is small, one can either ignore the subject effect or consider it as a fixed effect. However, when the subjects are nested within clusters or are repeatedly measured in a within-subjects design, a random-effects approach may be more appropriate.

For continuous normal responses, it is increasingly popular to use models incorporating cluster-specific or subject-specific heterogeneous effects in the form of hierarchical linear models (Bryk & Raudenbusch, 1992). For dichotomous or ordinal responses, such approaches to random-effects models are newer and have only recently seen much use in the social sciences (Fielding, 1999; Guo & Zhao, 2000). Specialized computer programs—notably, MIXOR (Hedeker & Gibbons, 1996), BUGS (Spiegelhalter, Thomas, Best, & Gilks, 1997), MLwiN (Yang, Rashbash, & Goldstein, 1998), and HLM5 (Raudenbusch, Bryk, Cheong, & Congdon, 2000)—have been developed to permit analyses of ordered categorical outcomes from either a clustered or a longitudinal design. Unfortunately, learning to use a specialized software often takes time and effort and may not appeal to users who do not encounter repeated categorical data on a routine basis but want to become familiar with the potential application of mixed-effects models. This paper advocates the use of a new SAS procedure, NLMIXED, to fit these models. The purpose is to facilitate the use of

mixed-effects ordinal regression models through the wide availability of SAS in both academic and industrial computing environments.

The NLMIXED procedure is designed to fit nonlinear mixed models in which both fixed and random effects (parameters) are allowed to relate nonlinearly to the response variable. The mixed-effects ordinal regression models discussed here belong to the class of generalized linear mixed models (McCulloch & Searle, 2001). A recent survey of random-effects modeling of dichotomous and count data is given by Agresti, Booth, Hobert, and Caffo (2000). Supplementing their discussion, this paper presents two examples of mixed-effects regression models for repeated ordinal outcomes that are of substantive interest to psychologists. It is hoped that the examples will stimulate further application of mixed-effects modeling in psychological research.

The paper is organized as follows. In the next section, the standard fixed-effects models are reviewed, and the mixed-effects ordinal regression models are introduced, using a wine-tasting data set reported by Randall (1989). In the succeeding section, model fits of the wine-tasting example are presented, using the cumulative logistic regression model and the mixed-effects logit-normal ordinal regression model. Some of the primary features of the NLMIXED procedure are presented, and the results of the two model fits are compared. In the final section, an application of the mixed-effects probit-normal regression model for a set of longitudinal ordinal data from the NIMH Schizophrenia Collaborative Study (Hedeker & Gibbons, 1996) is considered. The partial data, SAS code segments, and parameter estimates of the two examples are provided in tables and listings.

EXAMPLE 1
Bitterness of Wine

Randall (1989) has reported results from a within-subjects experiment in which 9 judges were asked to determine the bitterness of white wines. The experiment was a $2 \times 2 \times 2$ factorial design. The three factors were bottle (first or second), contact (yes or no), and temperature (cold or warm). Two bottles of the same wine were used. *Contact* referred to whether there was contact of the juice with the skins when the grapes were crushed. *Temperature* referred to the controlled temperature during crushing. Each bottle of wine was presented in random order to each of the 9 judges for tasting. The judges were asked to determine the bitterness of the wine on a 5-point rating scale ranging from *least* to *most bitter*. Listing 1 shows the data from the 1st and 9th judges. The variables are, in order, judge identification number, bottle (bottle 1 = 1, bottle 2 = −1), contact (no = 1, yes = −1), temperature (cold = 1, warm = −1), and bitterness rating (1 = *least bitter* and 5 = *most bitter*). There were two research questions: which of the experimental factors determine the bitterness of wines and whether or not the judges have varying scales of bitterness.

Listing 1
The Data of the First and Last Judges
in the Wine-Tasting Experiment

1	1	1	1	2
1	−1	1	1	3
1	1	−1	1	3
1	−1	−1	1	4
1	1	1	−1	4
1	−1	1	−1	4
1	1	−1	−1	5
1	−1	−1	−1	5
9	1	1	1	1
9	−1	1	1	2
9	1	−1	1	3
9	−1	−1	1	2
9	1	1	−1	3
9	−1	1	−1	2
9	1	−1	−1	4
9	−1	−1	−1	4

Note—The variables are, from left to right by column, judge identification number, bottle (first = 1, second = −1), contact (no = 1, yes = −1), temperature (cold = 1, warm = −1), bitterness rating (1 = *least bitter* and 5 = *most bitter*).

The Cumulative Logit Model

We start with a simple model in which all the judges in the example have the same sensitivity for the bitter taste of wine and use the same criteria (thresholds) for responding. Let *Y* denote the response variable for an observation of a judge on a 5-point rating scale. The cumulative probability $P\{Y \leq k\}$ is the probability that a response variable will take on a value less than or equal to a particular value $k = 1, \ldots 5$. Because $P\{Y \leq 5\}$ must equal one, there are only four uniquely defined cumulative probabilities. The probability of a particular ordered categorical response is defined through the cumulative probabilities—for example, $P\{Y = 4\} = P\{Y \leq 4\} - P\{Y \leq 3\}$. For all the judges in the study, the level of response is determined by the three treatment factors: bottle, contact, and temperature. The linear predictor of the model is

$$\eta = \beta_1 \text{ bottle} + \beta_2 \text{ contact} + \beta_3 \text{ temperature},$$

where β_1, β_2, and β_3 are the regression coefficients and the variable η can be thought of as taking its value on a continuous latent scale that determines each judge's perception of bitterness. McCullagh (1980) has proposed a cumulative logit model in which the conditional probabilities of the ordered responses are related to the linear predictor by the following representation:

$$P\{Y = 1 \mid x\} = \frac{1}{1 + \exp\left[-\left(c_1 + \eta\right)\right]},$$

$$P\{Y = k \mid x\} = \frac{1}{1 + \exp\left[-\left(c_k + \eta\right)\right]} - \frac{1}{1 + \exp\left[-\left(c_{k-1} + \eta\right)\right]},$$

for $k = 2, 3, 4,$

$$P\{Y = 5 \mid x\} = 1 - \frac{1}{1 + \exp\left[-\left(c_4 + \eta\right)\right]},$$

where

$$F(z) = \frac{1}{1 + \exp(-z)}$$

is the standard logistic cumulative distribution function, x stands for the three treatment factors, and the cutpoints (thresholds) $c_1 \ldots, c_4$ are unknown parameters satisfying the condition $c_1 \leq c_2 \leq \ldots \leq c_4$. Equivalently, we can define the thresholds by setting $c_1 = c$, $c_2 = c + d_1$, $c_3 = c + d_1 + d_2$, and $c_4 = c + d_1 + d_2 + d_3$, with $d_1 \geq 0$, $d_2 \geq 0$, and $d_3 \geq 0$. The latter constraints can be implemented in SAS with a BOUNDS statement. Fitting the fixed-effects cumulative logit models to ordinal data can be performed in SAS or S-PLUS (for examples, see Bender & Benner, 2000; Sheu & Heathcote, 2001). Sheu and Heathcote have discussed an extension of the cumulative logit model and the connection between this class of models and models of signal detectability (Green & Swets, 1974).

Mixed-Effects Cumulative Logit-Normal Models

The model described above assumes that all the judges in the study adopted a constant set of thresholds in producing the ratings. This assumption is often unrealistic and can be removed in two different ways. The first approach is to treat the judges as a blocking factor and to introduce eight subject-specific parameters for the 9 judges in the study. In this approach, a judge's initial threshold (first cutpoint) can be set as the baseline against which the thresholds of the other judges will be compared. A potential problem with this approach is that with a large number of subjects, the number of parameters to be estimated also becomes large. A second approach is to allow the thresholds to vary across judges by adding a subject-specific random variable U to the thresholds of each judge. In other words, for judge j, the original (fixed) thresholds $c_1, \ldots, c_4$ are simultaneously shifted (at random) to yield $c_1 + U_j, c_2 + U_j, \ldots, c_4 + U_j$ for $j = 1, 2, \ldots, 9$. For simplicity, the random effects U_j are assumed to be independent and identically distributed normal distributions with a mean of zero and a standard deviation of σ. In this formulation, the fixed thresholds are interpreted as the average thresholds across judges. From judge to judge, however, the thresholds are shifted simultaneously and randomly, with the amount of variation determined by the parameter σ. Unlike the fixed-effects approach, the number of parameters in a random-effects analysis does not grow linearly with the number of subjects in the sample. Thus, the random effects may behave like parameters; formally, they are just another level of random variation in the model.

For the wine example, the conditional probability of the ordered response, given the experimental factors and the random judge-specific effects, has the same representation as the equations defining the cumulative logit model, except for the linear predictor terms:

$$\eta_j = \beta_1 \text{ bottle} + \beta_2 \text{ contact} + \beta_3 \text{ temperature} + U_j,$$

for $j = 1, \ldots, 9$. This is a mixed-effects cumulative logit-normal model for ordinal responses. In general, the likelihood function is the marginal density function of the observed data, viewed as a function of the parameters. Here, the parameters to be investigated are the three regression coefficients, β_1, β_2, β_3; the first threshold parameter c; the nonnegative increment parameters d_1, d_2, d_3; and the random-effect parameter σ. In a mixed-effects analysis, estimating parameters requires maximizing the likelihood, which, in turn, requires integrating the joint probability function of the responses with respect to the random-effects distributions. Unfortunately, the likelihood nearly always involves intractable integrals. For cumulative logit and probit models, approximate integrated likelihoods can be found by using numerical integration methods, such as Gaussian quadrature (Abramowitz & Stegun, 1972). The SAS procedure NLMIXED (SAS Institute, 2000, chap. 46; Wolfinger, 1999) and the computer program MIXOR (Hedeker & Gibbons, 1996) directly fit the specified model by maximizing an approximation to the likelihood integrated over the random effects. In NLMIXED, an adaptive version of Gauss–Hermite quadrature is used to approximate the likelihood, and the default maximization routine is a dual quasi-Newton algorithm (Pinheiro & Bates, 1995).

Fitting Mixed-Effects Logit-Normal Models with NLMIXED

Listing 2 illustrates the use of PROC NLMIXED in Version 8 of SAS to fit the mixed-effects cumulative logistic-normal model to the wine data. The program takes about 3 sec to complete on a Pentium III 600-MHz personal computer with 128 MB of RAM running on the Windows 98 operating system. The PROC NLMIXED statement invokes the procedure and inputs the *wine* data set. The QPOINTS option specifies the number of quadrature points to use in approximating the likelihood with Gauss–Hermite quadrature. It is recommended that the number of quadrature points be increased until the results of the parameter estimates are stabilized. This often requires using about 10 quadrature points for each random effect in the model. Using a larger number of quadrature points requires longer computing time. The PARMS statement defines parameters and sets initial values for them. Starting values can be obtained from estimates of the corresponding fixed-effects model. Accurate starting values speed up convergence.

Within the NLMIXED procedure, we define the conditional probabilities for ordered responses as a function of the linear predictor *eta*. The categorical response probabilities are then specified according to the cumulative logit model. The BOUNDS statement provides the appropriate constraints on the increment variables for the thresholds. The MODEL statement defines the response variable and its conditional distribution, given the random effects. The multinomial distribution is not directly supported by the MODEL statement in the NLMIXED pro-

Listing 2
SAS Code to Fit the Mixed-Effects Cumulative Logit-Normal Model
for Wine Bitterness Rating Data

```
PROC NLMIXED DATA = wine QPOINTS = 10;
  PARMS c=−4 d1=1 d2=1 d3=1 beta1=0 beta2=1 beta3=1.5 sigma=1;
  BOUNDS d1 > 0, d2 > 0, d3 > 0;
  eta = beta1*bottle+beta2*contact+beta3*temper+u;
  IF (resp=1) THEN z = 1/(1+exp(−c−eta));
  ELSE IF(resp=2) THEN z=1/(1+exp(−c−d1−eta))−1/(1+exp(−c−eta));
  ELSE IF(resp=3) THEN z=1/(1+exp(−c−d1−d2−eta))−1/(1+exp(−c−d1−eta));
  ELSE IF(resp=4) THEN z=1/(1+exp(−c−d1−d2−d3−eta))−1/(1+exp(−c−d1−d2−eta));
  ELSE z = 1 − 1/(1+exp(−c−d1−d2−d3−eta));
  IF ( z > 1E−8) THEN llkhd = log(z);
  ELSE llkhd = −1E100;
  MODEL resp ~ GENERAL(llkhd);
  RANDOM u ~ NORMAL(0, sigma*sigma) SUBJECT=judge OUT=new;
  ESTIMATE 'thresh2' c+d1;
  ESTIMATE 'thresh3' c+d1+d2;
  ESTIMATE 'thresh4' c+d1+d2+d3;
  PROC PRINT DATA=new;
RUN;
```

Note—The input file wine.dat contains the ratings of 9 judges and the values of covariates in the format of Listing 1.

cedure. Instead, a general likelihood function is specified, using SAS statements. The likelihood (z) is checked to see whether it is numerically too close to zero and then is converted to the log likelihood (*llkhd*). The log likelihood is set to a large negative value (−*1E100*) if the likelihood is close to zero. The RANDOM statement defines *u* to be the random effect and specifies that it follows a normal distribution with mean zero and variance σ^2. Agresti et al. (2000) recommended specifying the variance component in terms of the standard deviation in the program code. This improves the stability of numerical solutions in cases in which the estimated variance components are very close to zero. The SUBJECT argument is set to the variable *judge*, indicating that the random effect changes according to the value of the subject variable *judge*. The OUT option in the RANDOM statement requests the predicted values of the random effect to be stored in a SAS data set named *new*. These are the estimates of judge effects for the bitterness of wine. The PRINT procedure sends the predicted values of the judge effects to the output file. The ESTIMATE statements re-

quest the threshold estimates to be computed from the first cutpoint and the increment variables.

Table 1 shows the results of both the fixed-effects and the mixed-effects models. The standard errors of the estimates are given in parentheses. Irrespective of constant or varying thresholds across judges, both models support the conclusions that the two experimental factors, contact and temperature, are significant in determining the bitterness of white wines, whereas the two bottles of the same wine elicit the same level of bitterness from the judges. Notice, however, that the random-effects parameter σ is statistically different from zero, since the estimated value of σ is about 2.86 times its estimated standard error. This suggests that the judges did not use a common set of constant thresholds to evaluate the wines. Also, according to the predicted values of the judge effects of the mixed-effects model displayed in Table 2, the approximate z values for estimates of Judges 1 and 7 are −2.262 and 2.364, respectively, supporting the heterogeneity of the judges. The intrajudge correlations are estimated by

$$\frac{\sigma^2}{\sigma^2 + \frac{\pi^2}{3}} = \frac{1.1453^2}{1.1453^2 + \frac{\pi^2}{3}} = .2851,$$

where $\pi^2/3$ is the variance of the standard logistic distribution.

EXAMPLE 2
The NIMH Schizophrenia Collaborative Study

In this section, a longitudinal data set in which 437 psychiatric patients were rated repeatedly up to seven times as to the severity of their illness is examined (Hedeker & Gibbons, 1996). The data were collected in the NIMH Schizophrenia Collaborative Study on treatment-related changes in overall severity of illness. The response vari-

Table 1
Parameter Estimates of the Two Models for the Bitterness
of Wine Example (With Standard Errors)

Parameter	Fixed Model		Mixed Model	
	Estimate	SE	Estimate	SE
c	−3.3588	0.5255	−4.0823	0.7595
$c+d_1$	−0.7617	0.2940	−0.9300	0.5035
$c+d_1+d_2$	1.4558	0.3310	1.7970	0.5525
$c+d_1+d_2+d_3$	2.9934	0.4707	3.6570	0.6924
β_1	0.0476	0.2209	0.1219	0.2321
β_2	0.7624	0.2368	0.9164	0.2561
β_3	1.2506	0.2660	1.5357	0.2982
σ	–	–	1.1453	0.4010
Log-likelihood	−86.469		−81.4	

Table 2
Predicted Values of Judge Effects of the
Mixed-Effects Logit-Normal Model for Bitterness of Wine Data
(With Standard Errors)

Judge	Estimate	SE
1	−1.71674	0.75900
2	0.59808	0.71036
3	−0.99176	0.72498
4	0.05435	0.69297
5	−0.23442	0.68742
6	−0.47293	0.68941
7	1.92946	0.81614
8	0.27335	0.65310
9	0.55228	0.67203

able is Item 79 of the Inpatient Multidimensional Psychiatric Scale (IMPS; Lorr & Klett, 1966). For this item, "Severity of Illness" 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*. The authors recoded these seven ordered categories into four by collapsing Categories 1 and 2, 3 and 4, and 6 and 7. Patients were randomly assigned to receive either a placebo medication or one of the three antipsychotic drugs (Chloropromazine, Fluphenazine, or Thioridazine), which were combined as one treatment group in this analysis. Moreover, a square root transformation of time (week) was chosen to linearize the relationship of the IMPS Item 79 scores over time. The main research question is whether there was differential improvement for the treatment group relative to the control group. Listing 3 shows the data from the first and last subjects. The original version of the data can be found at the Web address http://tigger.uic.edu/~hedeker/ long.html. Notice that neither of the 2 subjects has complete observations on all seven occasions. In addition to the outcome severity measure's being ordinal, the varying numbers of observations per subject and the need to account for the within-subjects variability (over time) pose difficulties for the traditional multivariate techniques.

Mixed-Effects Probit-Normal Model

Hedeker and Gibbons (1996) proposed mixed-effects probit regression models to relate the ordinal responses to the treatment and time variables. They implemented a FORTRAN computer program, MIXOR, to fit the models to this data. Here, we replicate their analysis, using the most general form of the models they discussed. For this case, the repeated ordinal IMPS score is modeled in terms of a dummy-coded treatment effect (placebo = 0 and treatment = 1), a time effect (square root of the week), and a treatment × time interaction. For each individual subject i on an occasion (week j) and belonging to a treatment group, the linear predictor of the model is

$$\eta_{i,j} = b_{0,i} + b_{1,i}\sqrt{\text{week}_{i,j}} + \beta_2\,\text{treatment}_i + \beta_3\,\text{treatment}_i \times \sqrt{\text{week}_{i,j}}.$$

This model allows the subjects to vary in terms of both their intercepts and their trends over time (random intercepts $b_{0,i}$ and random slopes $b_{1,i}$). The treatment effect, β_2, and the treatment × time interaction, β_3, are assumed to be the same for all the subjects (fixed effects). For each subject, the random-effects (b_0, b_1) component is assumed to have a bivariate normal distribution with a mean vector of (μ_{b_0}, μ_{b_1}) and a covariance matrix

$$\begin{pmatrix} \sigma_{b_0}^2 & \rho\sigma_{b_0}\sigma_{b_1} \\ \rho\sigma_{b_0}\sigma_{b_1} & \sigma_{b_1}^2 \end{pmatrix},$$

where $\sigma_{b_0}^2$ and $\sigma_{b_1}^2$ are the variances of the intercept and slope parameters, respectively, and ρ is the correlation between the two parameters. With four response categories, three cutpoints (thresholds) are needed. For model identifiability, the first cutpoint is set arbitrarily to zero, since it cannot be distinguished from the mean of the intercept parameter, μ_{b_0}. The third cutpoint is defined as the second cutpoint c plus an increment variable d ($c > 0$ and $d \geq 0$). The categorical response probabilities of subject i at week j, conditional on the mixed effects b and β (omitting the subscripts), are

$$P\{Y = 1 \mid b, \beta\} = \Phi(-\eta),$$
$$P\{Y = 2 \mid b, \beta\} = \Phi(c - \eta) - \Phi(-\eta),$$
$$P\{Y = 3 \mid b, \beta\} = \Phi(c + d - \eta) - \Phi(c - \eta),$$
$$P\{Y = 4 \mid b, \beta\} = 1 - \Phi(c + d - \eta),$$

where Φ is the standard normal cumulative distribution function. This is a mixed-effects probit-normal model for repeated ordinal outcomes.

Fitting Mixed-Effects Probit-Normal Models with NLMIXED

Listing 4 contains the SAS programming statements in the NLMIXED procedure for fitting the mixed-effects

Listing 3
Data of the First and Last Subjects from the
NIMH Schizophrenia Collaborative Study on
Treatment-Related Changes in Overall Severity

1103	4	1	0.0000
1103	2	1	1.0000
1103	.	1	1.4142
1103	2	1	1.7321
1103	.	1	2.0000
1103	.	1	2.2361
1103	2	1	2.4495
9316	4	0	0.0000
9316	4	0	1.0000
9316	.	0	1.4142
9316	4	0	1.7321
9316	.	0	2.0000
9316	.	0	2.2361
9316	4	0	2.4495

Note—The variables are, in order, subject identification number, rating of the severity of illness (1 = *normal or borderline mentally ill*, 2 = *mildly or moderately ill*, 3 = *markedly ill*, 4 = *severely or among the most ill*), treatment group (0 = placebo, 1 = drug), and the square root of the week. Both subjects have three missing values indicated by a period.

Listing 4
SAS Code to Fit the Mixed-Effects Cumulative Probit-Normal Model
for NIMH Schizophrenic Collaborative Study Data

```
PROC NLMIXED DATA=imps QPOINTS=20;
  PARMS c=2 d=1 mu_b0=3 mu_b1=−.4 beta2=.2 beta3=−.6 s_b01=0, s_b0=1 s_b1=1;
  BOUNDS c > 0, d > 0;
  eta = b0+b1*sqwk+beta2*trt+beta3*trt*sqwk;
  IF (resp=1) THEN z = PROBNORM(−eta);
  ELSE IF (resp=2) THEN z = PROBNORM(c−eta)−PROBNORM(−eta);
  ELSE IF (resp=3) THEN z = PROBNORM(c+d−eta)−PROBNORM(c−eta);
  ELSE z = 1−PROBNORM(c+d−eta);
  IF ( z > 1E−8) THEN llkhd = log(z);
  ELSE llkhd = −1E100;
  MODEL resp ~ GENERAL (llkhd);
  RANDOM b0 b1 ~ NORMAL([mu_b0, mu_b1],[s_b0*s_b0, s_b01, s_b1*s_b1])
                        SUBJECT=sbj;
  ESTIMATE 'thresh2' c+d;
  ESTIMATE 'corr' s_b01/(s_b0*s_b1);
RUN;
```

regression model to the longitudinal data on the treatment of schizophrenics. On the same personal computer as that described previously, the program takes about 28 min to complete. In the following, we comment briefly on the model specification in SAS statements.

We set the number of quadrature point, QPOINTS, to 20 in order to compare the results obtained from NLMIXED with those of Hedeker and Gibbons (1996) using MIXOR. The starting values for the parameters in the PARMS statement are taken from a pilot run of the program using the default number of quadrature points and reasonable guesses of the parameter values. The SAS programming statements begin by defining the linear predictor *eta*. The next statements define the ordinal likelihood according to the response variable *resp*, *eta*, the threshold variables defined by the cutpoint c and the increment variable d, and the standard normal cumulative distribution function PROBNORM. A general log-likelihood is specified in the MODEL statement. The bivariate normal random effects (b_0, b_1) and the mean and covariance structure are specified in the RANDOM statement. The SUBJECT variable in the statement indicates that the random effects are tied to the input variable (*sbj*). The third cutpoint and the correlation are computed in the ESTIMATE statements.

Table 3 compares the results of fitting the mixed-effects probit-normal model to the repeated ordinal IMPS scores, using SAS NLMIXED and the output from MIXOR (Hedeker & Gibbons, 1996). Numerically, the log-likelihood and the parameter estimates from the two programs are in close agreement. Similarly, the conclusions based on the approximate z or t statistics of the model parameters are the same. This is to be expected, because both programs implemented the same statistical procedure, using the same quadrature-based computational technique. Therefore, the conclusions based on the approximate t or z statistics of the model parameters are the same. First, the regression coefficient β_2 for the treatment/placebo group indicator variable is not significantly different from zero. Thus, the treatment group and

the placebo group were not different at the baseline. Second, the regression coefficient β_3 for the treatment × week interaction is negative (and significantly different from zero). Therefore, the subjects in the treatment group showed greater improvement over time, relative to their counterparts in the placebo group. Third, the mean of the random slope parameter μ_{b_1} for the time (week) variable is negative (and significantly different from zero). Hence, the subjects in the placebo group also improved over time. Fourth, a significant correlation of $-.3902$ is observed between the random intercept and the slope parameters for the time variable. This correlation means that those subjects with a higher level of severity at the baseline tended to show greater improvement across time.

CONCLUSIONS

Many response variables in psychological research are ordinal, with three or more ordered categories. Many psychological studies use repeated measures or longitudinal designs. The generalized mixed-effects models account for the between-subjects variation by the subject-specific random effects. Until recently, the applications

Table 3
Parameter Estimates of the Mixed-Effects Probit-Normal Model for IMPS Scores (With Standard Errors)

Parameter	NLMIXED		MIXOR	
	Estimate	SE	Estimate	SE
mu_b0	4.1012	0.2543	4.1096	0.2520
mu_b1	−0.5059	0.1214	−0.5051	0.1305
beta2	0.0324	0.2189	0.0388	0.2248
beta3	−0.9426	0.1400	−0.9506	0.1489
s_b0	1.4738	0.1368	1.4862	0.1413
s_b1	0.7870	0.0819	0.7303	0.0695
s_b01	−0.4526	0.1645	−0.4680	–
corr	−0.3902	0.0876	−0.3969	–
c	2.1791	0.1118	2.1842	0.1099
c+d	3.6434	0.1523	3.6538	0.1443
Log-likelihood	−1,663.524		−1,663.326	

of the generalized mixed-effects models have been hindered by the lack of general purpose software.

This paper has illustrated the use of a new SAS procedure, NLMIXED, to fit mixed-effects regression models for analyzing repeated ordinal responses. Two applications of the models were illustrated through examples of behavioral research. It is hoped that this paper will be a contribution toward helping psychologists understand the formulation and interpretation of mixed-effects models for categorical response data.

REFERENCES

ABRAMOWITZ, M., & STEGUN, I. A. (1972). *Handbook of mathematical functions*. New York: Dover.

AGRESTI, A., BOOTH, J. G., HOBERT, J. P., & CAFFO, B. (2000). Random effects modeling of categorical response data. *Sociological Methodology*, **30**, 27-80.

BENDER, R., & BENNER, A. (2000). Calculating ordinal regression models in SAS and S-Plus. *Biometrical Journal*, **42**, 677-699.

BRYK, A. S., & RAUDENBUSCH, S. W. (1992). *Hierarchical linear models*. Thousand Oaks, CA: Sage.

FIELDING, A. (1999). Why use arbitrary points scores? Ordered categories in models of educational progress. *Journal of the Royal Statistical Society A*, **162**, 303-328.

GREEN, D. M., & SWETS, J. A. (1974). *Signal detection theory and psychophysics*. Huntington, NY: Krieger.

GUO, G., & ZHAO, H. (2000). Multilevel modeling for binary data. *Annual Review of Sociology*, **26**, 441-462.

HEDEKER, D., & GIBBONS, R. D. (1996). A computer program for mixed-effects ordinal regression analysis. *Computer Methods & Programs in Biomedicine*, **49**, 157-176.

LORR, M., & KLETT, C. J. (1966). *Inpatient multidimensional psychiatric scale: Manual*. Palo Alto, CA: Consulting Psychologist Press.

MCCULLAGH, P. (1980). Regression model for ordinal data. *Journal of the Royal Statistical Society B*, **42**, 109-127.

MCCULLOCH, C. E., & SEARLE, S. R. (2001). *Generalized, linear, and mixed models*. New York: Wiley.

PINHEIRO, J. C., & BATES, D. M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics*, **4**, 12-35.

RANDALL, J. H. (1989). The analysis of sensory data by generalized linear models. *Biometrical Journal*, **31**, 781-793.

RAUDENBUSCH, S. W., BRYK, A. S., CHEONG, Y. F., & CONGDON, R. T. (2000). *HLM5: Hierarchical linear and nonlinear modeling*. Chicago: Scientific Software International.

SAS INSTITUTE (2000). *SAS/STAT user's guide* (Version 8). Cary, NC: Author.

SHEU, C.-F., & HEATHCOTE, A. (2001). A nonlinear regression approach to estimating signal detection models for rating data. *Behavioral Research Methods, Instruments, & Computers*, **33**, 108-114.

SPIEGELHALTER, D. J., THOMAS, A., BEST, N. G., & GILKS, W. (1997). *BUGS: Bayesian inference using Gibbs sampling* (Version 0.60). Cambridge: Medical Research Council Biostatistic Unit.

WARE, J. H. (1985). Linear models for the analysis of longitudinal studies. *American Statistician*, **39**, 95-101.

WOLFINGER, R. (1999). *Fitting nonlinear mixed models with the new NLMIXED procedure* (SUGI 24 Conference Proceedings, Paper 287). Cary, NC: SAS Institute.

YANG, M., RASHBASH, J., & GOLDSTEIN, H. (1998). *MLwiN macros for advanced multilevel modeling* (V2.0). London: Institute of Education.