THE POWER TO KNOW.

SAS/STAT® 12.3 User's Guide

Introduction to Mixed Modeling Procedures (Chapter)

This document is an individual chapter from *SAS/STAT® 12.3 User's Guide*.

The correct bibliographic citation for the complete manual is as follows: SAS Institute Inc. 2013. *SAS/STAT® 12.3 User's Guide*. Cary, NC: SAS Institute Inc.

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513.

July 2013

SAS® Publishing provides a complete selection of books and electronic products to help customers use SAS software to its fullest potential. For more information about our e-books, e-learning products, CDs, and hard-copy books, visit the SAS Publishing Web site at **support.sas.com/bookstore** or call 1-800-727-3228.

Chapter 6
Introduction to Mixed Modeling Procedures

Contents

Overview: Mixed Modeling Procedures

A mixed model is a model that contains fixed and random effects. Since all statistical models contain some stochastic component and many models contain a residual error term, the preceding sentence deserves some clarification. The classical linear model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ contains the parameters $\boldsymbol{\beta}$ and the random vector $\boldsymbol{\epsilon}$. The vector $\boldsymbol{\beta}$ is a vector of fixed-effects parameters; its elements are unknown constants to be estimated from the data. A mixed model in the narrow sense also contains random effects, which are unobservable random variables. If the vector of random effects is denoted by $\boldsymbol{\gamma}$, then a linear mixed model can be written as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon}$$

In a broader sense, mixed modeling and mixed model software is applied to special cases and generalizations of this model. For example, a purely random effects model, $\mathbf{Y} = \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon}$, or a correlated-error model, $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, is subsumed by mixed modeling methodology.

Over the last few decades virtually every form of classical statistical model has been enhanced to accommodate random effects. The linear model has been extended to the linear mixed model, generalized linear models have been extended to generalized linear mixed models, and so on. In parallel with this trend,

SAS/STAT software offers a number of classical and contemporary mixed modeling tools. The aim of this chapter is to provide a brief introduction and comparison of the procedures for mixed model analysis (in the broad sense) in SAS/STAT software. The theory and application of mixed models are discussed at length in many monographs, including Milliken and Johnson (1992); Diggle, Liang, and Zeger (1994); Davidian and Giltinan (1995); Verbeke and Molenberghs (1997, 2000); Vonesh and Chinchilli (1997); Demidenko (2004); Molenberghs and Verbeke (2005); and Littell et al. (2006).

The following procedures in SAS/STAT software can perform mixed and random effects analysis to various degrees:

GLM — is primarily a tool for fitting linear models by least squares. The GLM procedure has some capabilities for including random effects in a statistical model and for performing statistical tests in mixed models. Repeated measures analysis is also possible with the GLM procedure, assuming unstructured covariance modeling. Estimation methods for covariance parameters in PROC GLM are based on the method of moments, and a portion of its output applies only to the fixed-effects model.

GLIMMIX — fits generalized linear mixed models by likelihood-based techniques. As in the MIXED procedure, covariance structures are modeled parametrically. The GLIMMIX procedure also has built-in capabilities for mixed model smoothing and joint modeling of heterocatanomic multivariate data.

HPMIXED — fits linear mixed models by sparse-matrix techniques. The HPMIXED procedure is designed to handle large mixed model problems, such as the solution of mixed model equations with thousands of fixed-effects parameters and random-effects solutions.

LATTICE — computes the analysis of variance and analysis of simple covariance for data from an experiment with a lattice design. PROC LATTICE analyzes balanced square lattices, partially balanced square lattices, and some rectangular lattices. Analyses performed with the LATTICE procedure can also be performed as mixed models for complete or incomplete block designs with the MIXED procedure.

MIXED — performs mixed model analysis and repeated measures analysis by way of structured covariance models. The MIXED procedure estimates parameters by likelihood or moment-based techniques. You can compute mixed model diagnostics and influence analysis for observations and groups of observations. The default fitting method maximizes the restricted likelihood of the data under the assumption that the data are normally distributed and any missing data are missing at random. This general framework accommodates many common correlated-data methods, including variance component models and repeated measures analyses.

NESTED — performs analysis of variance and analysis of covariance for purely nested random-effects models. Because of its customized algorithms, PROC NESTED can be useful for large data sets with nested random effects.

NLMIXED — fits mixed models in which the fixed or random effects enter nonlinearly. The NLMIXED procedure requires that you specify components of your mixed model via programming statements. Some built-in distributions enable you to easily specify the conditional distribution of the data, given the random effects.

VARCOMP — estimates variance components for random or mixed models.

The focus in the remainder of this chapter is on procedures designed for random effects and mixed model analysis: the GLIMMIX, HPMIXED, MIXED, NESTED, NLMIXED, and VARCOMP procedures. The

important distinction between fixed and random effects in statistical models is addressed in the section "Fixed, Random, and Mixed Models" on page 31, in Chapter 3, "Introduction to Statistical Modeling with SAS/STAT Software."

Types of Mixed Models

Linear, Generalized Linear, and Nonlinear Mixed Models

The linear model shown at the beginning of this chapter was incomplete because the distributional properties of the random variables and their relationship were not specified. In this section the specification of the models is completed and the three model classes, linear mixed models (LMM), generalized linear mixed models (GLMM), and nonlinear mixed models (NLMM), are delineated.

Linear Mixed Model

It is a defining characteristic of the class of linear mixed models (LMM), the class of generalized linear mixed models (GLMM), and the class of nonlinear mixed models (NLMM) that the random effects are normally distributed. In the linear mixed model, this also applies to the error term; furthermore, the errors and random effects are uncorrelated. The standard linear mixed model (LMM) is thus represented by the following assumptions:

$$\begin{aligned} \mathbf{Y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon} \\ \boldsymbol{\gamma} &\sim N(\mathbf{0}, \mathbf{G}) \\ \boldsymbol{\epsilon} &\sim N(\mathbf{0}, \mathbf{R}) \\ \mathrm{Cov}[\boldsymbol{\gamma}, \boldsymbol{\epsilon}] &= \mathbf{0} \end{aligned}$$

The matrices $\mathbf{G}$ and $\mathbf{R}$ are covariance matrices for the random effects and the random errors, respectively. A *G-side* random effect in a linear mixed model is an element of $\boldsymbol{\gamma}$, and its variance is expressed through an element in $\mathbf{G}$. An *R-side* random variable is an element of $\boldsymbol{\epsilon}$, and its variance is an element of $\mathbf{R}$. The GLIMMIX, HPMIXED, and MIXED procedures express the $\mathbf{G}$ and $\mathbf{R}$ matrices in parametric form—that is, you structure the covariance matrix, and its elements are expressed as functions of some parameters, known as the *covariance parameters* of the mixed models. The NLMIXED procedure also parameterizes the covariance structure, but you accomplish this with programming statements rather than with predefined syntax.

Since the right side of the model equation contains multiple random variables, the stochastic properties of $\mathbf{Y}$ can be examined by conditioning on the random effects, or through the marginal distribution. Because of the linearity of the G-side random effects and the normality of the random variables, the conditional and the marginal distribution of the data are also normal with the following mean and variance matrices:

$$\begin{aligned} \mathbf{Y}|\boldsymbol{\gamma} &\sim N(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma}, \mathbf{R}) \\ \mathbf{Y} &\sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V}) \\ \mathbf{V} &= \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R} \end{aligned}$$

Parameter estimation in linear mixed models is based on likelihood or method-of-moment techniques. The default estimation method in PROC MIXED, and the only method available in PROC HPMIXED, is restricted (residual) maximum likelihood, a form of likelihood estimation that accounts for the parameters in the fixed-effects structure of the model to reduce the bias in the covariance parameter estimates. Moment-based estimation of the covariance parameters is available in the MIXED procedure through the METHOD= option in the PROC MIXED statement. The moment-based estimators are associated with sums of squares, expected mean squares (EMS), and the solution of EMS equations.

Parameter estimation by likelihood-based techniques in linear mixed models maximizes the marginal (restricted) log likelihood of the data—that is, the log likelihood is formed from $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$. This is a model for $\mathbf{Y}$ with mean $\mathbf{X}\boldsymbol{\beta}$ and covariance matrix $\mathbf{V}$, a correlated-error model. Such *marginal models* arise, for example, in the analysis of time series data, repeated measures, or spatial data, and are naturally subsumed into the linear mixed model family. Furthermore, some mixed models have an equivalent formulation as a correlated-error model, when both give rise to the same marginal mean and covariance matrix. For example, a mixed model with a single variance component is identical to a correlated-error model with compound-symmetric covariance structure, provided that the common correlation is positive.

Generalized Linear Mixed Model

In a generalized linear mixed model (GLMM) the G-side random effects are part of the linear predictor, $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma}$, and the predictor is related nonlinearly to the conditional mean of the data

$$\mathrm{E}[\mathbf{Y}|\boldsymbol{\gamma}] = g^{-1}(\boldsymbol{\eta}) = g^{-1}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma})$$

where $g^{-1}(\cdot)$ is the inverse link function. The conditional distribution of the data, given the random effects, is a member of the exponential family of distributions, such as the binary, binomial, Poisson, gamma, beta, or chi-square distribution. Because the normal distribution is also a member of the exponential family, the class of the linear mixed models is a subset of the generalized linear mixed models. In order to completely specify a GLMM, you need to do the following:

1. Formulate the linear predictor, including fixed and random effects.
2. Choose a link function.
3. Choose the distribution of the response, conditional on the random effects, from the exponential family.

As an example, suppose that s pairs of twins are randomly selected in a matched-pair design. One of the twins in each pair receives a treatment and the outcome variable is some binary measure. This is a study with s clusters (subjects) and each cluster is of size 2. If Y_{ij} denotes the binary response of twin $j = 1, 2$ in cluster i, then a linear predictor for this experiment could be

$$\eta_{ij} = \beta_0 + \tau x_{ij} + \gamma_i$$

where x_{ij} denotes a regressor variable that takes on the value 1 for the treated observation in each pair, and 0 otherwise. The γ_i are pair-specific random effects that model heterogeneity across sets of twins and that induce a correlation between the members of each pair. By virtue of random sampling the sets of twins,

it is reasonable to assume that the γ_i are independent and have equal variance. This leads to a diagonal **G** matrix,

$$\mathrm{Var}[\boldsymbol{\gamma}] = \mathrm{Var}\begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \vdots \\ \gamma_s \end{bmatrix} = \begin{bmatrix} \sigma^2_\gamma & 0 & 0 & \cdots & 0 \\ 0 & \sigma^2_\gamma & 0 & \cdots & 0 \\ 0 & 0 & \sigma^2_\gamma & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma^2_\gamma \end{bmatrix}$$

A common link function for binary data is the logit link, which leads in the second step of model formulation to

$$\mathrm{E}\left[Y_{ij}|\gamma_i\right] = \mu_{ij}|\gamma_i = \frac{1}{1+\exp\{-\eta_{ij}\}}$$
$$\mathrm{logit}\left\{\frac{\mu_{ij}|\gamma_i}{1-\mu_{ij}|\gamma_i}\right\} = \eta_{ij}$$

The final step, choosing a distribution from the exponential family, is automatic in this example; only the binary distribution comes into play to model the distribution of $Y_{ij}|\gamma_i$.

As for the linear mixed model, there is a marginal model in the case of a generalized linear mixed model that results from integrating the joint distribution over the random effects. This marginal distribution is elusive for many GLMMs, and parameter estimation proceeds by either approximating the model or by approximating the marginal integral. Details of these approaches are described in the section "Generalized Linear Mixed Models Theory" on page 3052, in Chapter 41, "The GLIMMIX Procedure."

A marginal model, one that models correlation through the **R** matrix and does not involve G-side random effects, can also be formulated in the GLMM family; such models are the extension of the correlated-error models in the linear mixed model family. Because nonnormal distributions in the exponential family exhibit a functional mean-variance relationship, fully parametric estimation is not possible in such models. Instead, estimating equations are formed based on first-moment (mean) and second-moment (covariance) assumptions for the marginal data. The approaches for modeling correlated nonnormal data via generalized estimating equations (GEE) fall into this category (see, for example, Liang and Zeger 1986; Zeger and Liang 1986).

Nonlinear Mixed Model

In a nonlinear mixed model (NLMM), the fixed and/or random effects enter the conditional mean function nonlinearly. If the mean function is a general, nonlinear function, then it is customary to assume that the conditional distribution is normal, such as in modeling growth curves or pharmacokinetic response. This is not a requirement, however.

An example of a nonlinear mixed model is the following logistic growth curve model for the *j*th observation of the *i*th subject (cluster):

$$f(\boldsymbol{\beta}, \boldsymbol{\gamma}_i, x_{ij}) = \frac{\beta_1 + \gamma_{i1}}{1 + \exp[-(x_{ij} - \beta_2)/(\beta_3 + \gamma_{i2})]}$$

$$Y_{ij} = f(\boldsymbol{\beta}, \boldsymbol{\gamma}_i, x_{ij}) + \epsilon_{ij}$$

$$\begin{bmatrix} \gamma_{i1} \\ \gamma_{i2} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} \right)$$

$$Y_{ij} | \gamma_{i1}, \gamma_{i2} \sim N(0, \sigma_\epsilon^2)$$

The inclusion of R-side covariance structures in GLMM and NLMM models is not as straightforward as in linear mixed models for the following reasons:

- The normality of the conditional distribution in the LMM enables straightforward modeling of the covariance structure because the mean structure and covariance structure are not functionally related.
- The linearity of the random effects in the LMM leads to a marginal distribution that incorporates the **R** matrix in a natural and meaningful way.

To incorporate R-side covariance structures when random effects enter nonlinearly or when the data are not normally distributed requires estimation approaches that rely on linearizations of the mixed model. Among such estimation methods are the pseudo-likelihood methods that are available with the GLIMMIX procedure. Generalized estimating equations also solve this marginal estimation problem for nonnormal data; these are available with the GENMOD procedure.

Models for Clustered and Hierarchical Data

Mixed models are often applied in situations where data are clustered, grouped, or otherwise hierarchically organized. For example, observations might be collected by randomly selecting schools in a school district, then randomly selecting classrooms within schools, followed by selecting students within the classroom. A longitudinal study might randomly select individuals and take repeatedly measurements on them. In the first example, a school is a cluster of observations, which consists of smaller clusters (classrooms) and so on. In the longitudinal example the observations for a particular individual form a cluster. Mixed models are popular analysis tools for hierarchically organized data for the following reasons:

- The selection of groups is often performed randomly, so that the associated effects are random effects.
- The data from different clusters are independent by virtue of the random selection or by assumption.
- The observations from the same cluster are often correlated, such as the repeated observations in a repeated measures or longitudinal study.
- It is often believed that there is heterogeneity in model parameters across subjects; for example, slopes and intercepts might differ across individuals in a longitudinal growth study. This heterogeneity, if due to stochastic sources, can be modeled with random effects.

A linear mixed models with clustered, hierarchical structure can be written as a special case of the general linear mixed model by introducing appropriate subscripts. For example, a mixed model with one type of clustering and s clusters can be written as

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{\gamma}_i + \boldsymbol{\epsilon}_i \quad i = 1, \cdots, s$$

In SAS/STAT software, the clusters are referred to as *subjects*, and the effects that define clusters in your data can be specified with the SUBJECT= option in the GLIMMIX, HPMIXED, MIXED, and NLMIXED procedures. The vector $\mathbf{Y}_i$ collects the n_i observations for the ith subject. In certain disciplines, the organization of a hierarchical model is viewed in a *bottom-up* form, where the measured observations represent the first level, these are collected into units at the second level, and so forth. In the school data example, the bottom-up approach considers a student's score as the level-1 observation, the classroom as the level-2 unit, and the school district as the level-3 unit (if these were also selected from a population of districts).

The following points are noteworthy about mixed models with SUBJECT= specification:

- A SUBJECT= option is available in the RANDOM statements of the GLIMMIX, HPMIXED, MIXED, and NLMIXED procedures and in the REPEATED statement of the MIXED and HPMIXED procedures.
- A SUBJECT= specification is required in the NLMIXED and HPMIXED procedures. It is not required with any other mixed modeling procedure in SAS/STAT software.
- Specifying models with subjects is usually more computationally efficient in the MIXED and GLIMMIX procedures, especially if the SUBJECT= effects are identical or contained within each other. The computational efficiency of the HPMIXED procedure is not dependent on SUBJECT= effects in the manner in which the MIXED and GLIMMIX procedures are affected.
- There is no limit to the number of SUBJECT= effects with the MIXED, HPMIXED, and GLIMMIX procedures—that is, you can achieve an arbitrary depth of the nesting.

Models with Subjects and Groups

The concept of a subject as a unit of clustering observations in a mixed model has been described in the preceding section. This concept is important for mixed modeling with the GLIMMIX, HPMIXED, MIXED, and NLMIXED procedures. Observations from two subjects are considered uncorrelated in the analysis. Observations from the same subject are potentially correlated, depending on your specification of the covariance structure. Random effects at the subject level always lead to correlation in the marginal distribution of the observations that belong to the subject.

The GLIMMIX, HPMIXED, and MIXED procedures also support the notion of a GROUP= effect in the specification of the covariance structure. Like a subject effect, a G-side group effect identifies independent random effects. In addition to a subject effect, the group effect assumes that the realizations of the random effects correspond to draws from different distributions; in other words, each level of the group effect is associated with a different set of covariance parameters. For example, the following statements in any of these procedures fit a random coefficient model with fixed intercept and slope and subject-specific random intercept and slope:

```
class id;
model y = x;
random intercept x / subject=id;
```

The interpretation of the RANDOM statement is that for each ID an independent draw is made from a bivariate normal distribution with zero mean and a diagonal covariance matrix. In the following statements (in any of these procedures) these independent draws come from different bivariate normal distributions depending on the value of the grp variable.

```
class id grp;
model y = x;
random intercept x / subject=id group=grp;
```

Adding GROUP= effects in your model increases the flexibility to model heterogeneity in the covariance parameters, but it can add numerical complexity to the estimation process.

Linear Mixed Models

You can fit linear mixed models in SAS/STAT software with the GLM, GLIMMIX, HPMIXED, LATTICE, MIXED, NESTED, and VARCOMP procedures.

The procedure specifically designed for statistical estimation in linear mixed models is the MIXED procedure. To fit the linear mixed model

$$\begin{aligned} \mathbf{Y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon} \\ \boldsymbol{\gamma} &\sim N(\mathbf{0}, \mathbf{G}) \\ \boldsymbol{\epsilon} &\sim N(\mathbf{0}, \mathbf{R}) \\ \mathrm{Cov}[\boldsymbol{\gamma}, \boldsymbol{\epsilon}] &= \mathbf{0} \end{aligned}$$

with the MIXED procedure, you specify the fixed-effects design matrix $\mathbf{X}$ in the MODEL statement, the random-effects design matrix $\mathbf{Z}$ in the RANDOM statement, the covariance matrix of the random effects $\mathbf{G}$ with options (SUBJECT=, GROUP=, TYPE=) in the RANDOM statement, and the $\mathbf{R}$ matrix in the REPEATED statement.

By default, covariance parameters are estimated by restricted (residual) maximum likelihood. In supported models, the METHOD=TYPE1, METHOD=TYPE2, and METHOD=TYPE3 options lead to method-of-moment-based estimators and analysis of variance. The MIXED procedure provides an extensive list of diagnostics for mixed models, from various residual graphics to observationwise and groupwise influence diagnostics.

The NESTED procedure performs an analysis of variance in nested random effects models. The VARCOMP procedure can be used to estimate variance components associated with random effects in random and mixed models. The LATTICE procedure computes analysis of variance for balanced and partially balanced square lattices. You can fit the random and mixed models supported by these procedures with the MIXED procedure as well. Some specific analyses, such as the analysis of Gauge R & R studies in the VARCOMP procedure (Burdick, Borror, and Montgomery 2005), are unique to the specialized procedures.

The GLIMMIX procedure can fit most of the models that you can fit with the MIXED procedure, but it does not offer method-of-moment-based estimation and analysis of variance in the narrow sense. Also, PROC

GLIMMIX does not support the same array of covariance structures as the MIXED procedure and does not support a sampling-based Bayesian analysis. An in-depth comparison of the GLIMMIX and MIXED procedures can be found in the section "Comparing the GLIMMIX and MIXED Procedures" on page 3101, in Chapter 41, "The GLIMMIX Procedure."

Comparing the MIXED and GLM Procedures

Random- and mixed-effects models can also be fitted with the GLM procedure, but the philosophy is different from that of PROC MIXED and other dedicated mixed modeling procedures. The following lists important differences between the GLM and MIXED procedures in fitting random and mixed models:

- The default estimation method for covariance parameters in the MIXED procedure is restricted maximum likelihood. Covariance parameters are estimated by the method of moments by solving expressions for expected mean squares.
- In the GLM procedure, fixed and random effects are listed in the MODEL statement. Only fixed effects are listed in the MODEL statement of the MIXED procedure. In the GLM procedure, random effects must be repeated in the RANDOM statement.
- You can request tests for model effects by adding the TEST option in the RANDOM statement of the GLM procedure. PROC GLM then constructs exact tests for random effects if possible and constructs approximate tests if exact tests are not possible. For details on how the GLM procedure constructs tests for random effects, see the section "Computation of Expected Mean Squares for Random Effects" on page 3376, in Chapter 42, "The GLM Procedure." Tests for fixed effects are constructed by the MIXED procedure as Wald-type F tests, and the degrees of freedom for these tests can be determined by a variety of methods.
- Some of the output of the GLM procedure applies only to the fixed effects part of the model, whether a RANDOM statement is specified or not.
- Variance components are independent in the GLM procedure and covariance matrices are generally unstructured. The default covariance structure for variance components in the MIXED procedure is also a variance component structure, but the procedure offers a large number of parametric structures to model covariation among random effects and observations.

Comparing the MIXED and HPMIXED Procedures

The HPMIXED procedure is designed to solve large mixed model problems by using sparse matrix techniques. The largeness of a mixed model can take many forms: a large number of observations, large number of columns in the $\mathbf{X}$ matrix, a large number of random effects, or a large number of covariance parameters. The province of the HPMIXED procedure is parameter estimation, inference, and prediction in mixed models with large $\mathbf{X}$ and/or $\mathbf{Z}$ matrices, many observations, but relatively few covariance parameters.

The models that you can fit with the HPMIXED procedure are a subset of the models available with the MIXED procedure. The HPMIXED procedure supports only a limited number of types of covariance structure in the RANDOM and REPEATED statements in order to balance performance and generality.

To some extent, the generality of the MIXED procedure precludes it from serving as a high-performance computing tool for all the model-data scenarios that the procedure can potentially estimate parameters for. For example, although efficient sparse algorithms are available to estimate variance components in large mixed models, the computational configuration changes profoundly when, for example, standard error adjustments and degrees of freedom by the Kenward-Roger method are requested.

Generalized Linear Mixed Models

Generalized linear mixed models can be fit with the GLIMMIX and NLMIXED procedures in SAS/STAT software. The GLIMMIX procedure is specifically designed to fit this class of models and offers syntax very similar to the syntax of other linear modeling procedures, such as the MIXED procedure. Consider a generalized linear model with linear predictor and link function

$$\mathrm{E}[\mathbf{Y}|\boldsymbol{\gamma}] = g^{-1}(\boldsymbol{\eta}) = g^{-1}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma})$$

and distribution in the exponential family. The fixed-effects design matrix $\mathbf{X}$ is specified in the MODEL statement of the GLIMMIX procedure, and the random-effects design matrix $\mathbf{Z}$ is specified in the RANDOM statement, along with the covariance matrix of the random effects and the covariance matrix of R-side random variables. The link function and (conditional) distribution are determined by defaults or through options in the MODEL statement.

The GLIMMIX procedure can fit heterocatanomic multivariate data—that is, data that stem from different distributions. For example, one measurement taken on a patient might be a continuous, normally distributed outcome, whereas another measurement might be a binary indicator of medical history. The GLIMMIX procedure also provides capabilities for mixed model smoothing and mixed model splines.

The GLIMMIX procedure offers an extensive array of postprocessing features to produce output statistics and to perform linear inference. The ESTIMATE and LSMESTIMATE statements support multiplicity-adjusted p-values for the protection of the familywise Type-I error rate. The LSMEANS statement supports the slicing of interactions, simple effect differences, and ODS statistical graphs for group comparisons.

The default estimation technique in the GLIMMIX procedure depends on the class of models fit. For linear mixed models, the default technique is restricted maximum likelihood, as in the MIXED procedure. For generalized linear mixed models, the estimation is based on linearization methods (pseudo-likelihood) or on integral approximation by adaptive quadrature or Laplace methods.

The NLMIXED procedure facilitates the fitting of generalized linear mixed models through several built-in distributions from the exponential family (binary, binomial, gamma, negative binomial, and Poisson). You have to code the linear predictor and link function with SAS programming statements and assign starting values to all parameters, including the covariance parameters. Although you are not required to specify starting values with the NLMIXED procedure (because the procedure assigns a default value of 1.0 to every parameter not explicitly given a starting value), it is highly recommended that you specify good starting values. The default estimation technique of the NLMIXED procedure, an adaptive Gauss-Hermite quadrature, is also available in the GLIMMIX procedure through the METHOD=QUAD option in the PROC GLIMMIX statement. The Laplace approximation that is available in the NLMIXED procedure by setting QPOINTS=1 is available in the GLIMMIX procedure through the METHOD=LAPLACE option.

Comparing the GENMOD and GLIMMIX Procedures

The GENMOD and GLIMMIX procedures can fit generalized linear models and estimate the parameters by maximum likelihood. For multinomial data, the GENMOD procedure fits cumulative link models for ordinal data. The GLIMMIX procedure fits these models and generalized logit models for nominal data.

When data are correlated, you can use the REPEATED statement in the GENMOD procedure to fit marginal models via generalized estimating equations. A working covariance structure is assumed, and the standard errors of the parameter estimates are computed according to an empirical ("sandwich") estimator that is robust to the misspecification of the covariance structure. Marginal generalized linear models for correlated data can also be fit with the GLIMMIX procedure by specifying the random effects as R-side effects. The empirical covariance estimators are available through the EMPIRICAL= option in the PROC GLIMMIX statement. The essential difference between the estimation approaches taken by the GLIMMIX procedure and generalized estimating equations is that the latter approach estimates the covariance parameters by the method of moments, whereas the GLIMMIX procedure uses likelihood-based techniques.

The GENMOD procedure supports nonsingular parameterizations of classification variables through its CLASS statement. The GLIMMIX procedure supports only the standard, GLM-type singular parameterization of CLASS variables. For the differences between these parameterizations, see the section "Parameterization of Model Effects" on page 383, in Chapter 19, "Shared Concepts and Topics."

Nonlinear Mixed Models: The NLMIXED Procedure

PROC NLMIXED handles models in which the fixed or random effects enter nonlinearly. It requires that you specify a conditional distribution of the data given the random effects, with available distributions including the normal, binomial, and Poisson. You can alternatively code your own distribution with SAS programming statements. Under a normality assumption for the random effects, PROC NLMIXED performs maximum likelihood estimation via adaptive Gaussian quadrature and a dual quasi-Newton optimization algorithm. Besides standard maximum likelihood results, you can obtain empirical Bayes predictions of the random effects and estimates of arbitrary functions of the parameters with delta-method standard errors. PROC NLMIXED has a wide variety of applications; two of the most common applications are nonlinear growth curves and overdispersed binomial data.

References

Burdick, R. K., Borror, C. M., and Montgomery, D. C. (2005), *Design and Analysis of Gauge R&R Studies: Making Decisions with Confidence Intervals in Random and Mixed ANOVA Models*, Philadelphia, PA and Alexandria, VA: SIAM and ASA.

Davidian, M. and Giltinan, D. M. (1995), *Nonlinear Models for Repeated Measurement Data*, New York: Chapman & Hall.

Demidenko, E. (2004), *Mixed Models: Theory and Applications*, New York: John Wiley & Sons.

Diggle, P. J., Liang, K.-Y., and Zeger, S. L. (1994), *Analysis of Longitudinal Data*, Oxford: Clarendon Press.

Laird, N. M. and Ware, J. H. (1982), "Random-Effects Models for Longitudinal Data," *Biometrics*, 38, 963–974.

Liang, K.-Y. and Zeger, S. L. (1986), "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika*, 73, 13–22.

Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., and Schabenberger, O. (2006), *SAS for Mixed Models*, Second Edition, Cary, NC: SAS Press.

Milliken, G. A. and Johnson, D. E. (1992), *Analysis of Messy Data, Volume 1: Designed Experiments*, New York: Chapman & Hall.

Molenberghs, G. and Verbeke, G. (2005), *Models for Discrete Longitudinal Data*, New York: Springer.

Verbeke, G. and Molenberghs, G., eds. (1997), *Linear Mixed Models in Practice: A SAS-Oriented Approach*, New York: Springer.

Verbeke, G. and Molenberghs, G. (2000), *Linear Mixed Models for Longitudinal Data*, New York: Springer.

Vonesh, E. F. and Chinchilli, V. M. (1997), *Linear and Nonlinear Models for the Analysis of Repeated Measurements*, New York: Marcel Dekker.

Zeger, S. L. and Liang, K.-Y. (1986), "Longitudinal Data Analysis for Discrete and Continuous Outcomes," *Biometrics*, 42, 121–130.

Index

Your Turn

We welcome your feedback.

- If you have comments about this book, please send them to **yourturn@sas.com**. Include the full title and page numbers (if applicable).
- If you have comments about the software, please send them to **suggest@sas.com**.