SAS PROC MIXED

http://www.id.unizh.ch/software/unix/statmath/sas/sasdoc/stat/chap41/index.htm

Overview

The MIXED procedure fits a variety of mixed linear models to data and enables you to use these fitted models to make statistical inferences about the data. A *mixed linear model* is a generalization of the standard linear model used in the GLM procedure, the generalization being that the data are permitted to exhibit correlation and nonconstant variability. The mixed linear model, therefore, provides you with the flexibility of modeling not only the means of your data (as in the standard linear model) but their variances and covariances as well.

The primary assumptions underlying the analyses performed by PROC MIXED are as follows:

- The data are normally distributed (Gaussian).
- The means (expected values) of the data are linear in terms of a certain set of parameters.
- The variances and covariances of the data are in terms of a different set of parameters, and they exhibit a structure matching one of those available in PROC MIXED.

Since Gaussian data can be modeled entirely in terms of their means and variances/covariances, the two sets of parameters in a mixed linear model actually specify the complete probability distribution of the data. The parameters of the mean model are referred to as *fixed-effects parameters*, and the parameters of the variance-covariance model are referred to as *covariance parameters*.

The fixed-effects parameters are associated with known explanatory variables, as in the standard linear model. These variables can be either qualitative (as in the traditional analysis of variance) or quantitative (as in standard linear regression). However, the covariance parameters are what distinguishes the mixed linear model from the standard linear model.

The need for covariance parameters arises quite frequently in applications, the following being the two most typical scenarios:

- The experimental units on which the data are measured can be grouped into clusters, and the data from a common cluster are correlated.
- Repeated measurements are taken on the same experimental unit, and these repeated measurements are correlated or exhibit variability that changes.

The first scenario can be generalized to include one set of clusters nested within another. For example, if students are the experimental unit, they can be clustered into classes, which in turn can be clustered into schools. Each level of this hierarchy can introduce an additional source of variability and correlation. The second scenario occurs in longitudinal studies, where repeated measurements are taken over time. Alternatively, the repeated measures could be spatial or multivariate in nature.

PROC MIXED provides a variety of covariance structures to handle the previous two scenarios. The most common of these structures arises from the use of *random-effects parameters*, which are additional unknown random variables assumed to impact the variability of the data. The variances of the random-effects parameters, commonly known as *variance components*, become the covariance parameters for this particular structure. Traditional mixed linear models contain both fixed- and random-effects parameters, and, in fact, it is the combination of these two types of effects that led to the name *mixed model*. PROC MIXED fits not only these traditional variance component models but numerous other covariance structures as well.

PROC MIXED fits the structure you select to the data using the method of *restricted maximum likelihood (REML)*, also known as *residual maximum likelihood*. It is here that the Gaussian assumption for the data is exploited. Other

estimation methods are also available, including *maximum likelihood* and *MIVQUE0*. The details behind these estimation methods are discussed in subsequent sections.

Once a model has been fit to your data, you can use it to draw statistical inferences via both the fixed-effects and covariance parameters. PROC MIXED computes several different statistics suitable for generating hypothesis tests and confidence intervals. The validity of these statistics depends upon the mean and variance-covariance model you select, so it is important to choose the model carefully. Some of the output from PROC MIXED helps you assess your model and compare it with others.

Basic Features

PROC MIXED provides easy accessibility to numerous mixed linear models that are useful in many common statistical analyses. In the style of the GLM procedure, PROC MIXED fits the specified mixed linear model and produces appropriate statistics.

Some basic features of PROC MIXED are

- covariance structures, including variance components, compound symmetry, unstructured, AR(1), Toeplitz, spatial, general linear, and factor analytic
- GLM-type grammar, using MODEL, RANDOM, and REPEATED statements for model specification and CONTRAST, ESTIMATE, and LSMEANS statements for inferences
- appropriate standard errors for all specified estimable linear combinations of fixed and random effects, and corresponding *t*- and *F*-tests
- subject and group effects that enable blocking and heterogeneity, respectively
- REML and ML estimation methods implemented with a Newton-Raphson algorithm
- capacity to handle unbalanced data
- ability to create a SAS data set corresponding to any table

PROC MIXED uses the Output Delivery System (ODS), a SAS subsystem that provides capabilities for displaying and controlling the output from SAS procedures. ODS enables you to convert any of the output from PROC MIXED into a SAS data set. See the "Changes in Output" section.

Notation for the Mixed Model

This section introduces the mathematical notation used throughout this chapter to describe the mixed linear model. You should be familiar with basic matrix algebra (refer to Searle 1982). A more detailed description of the mixed model is contained in the "Mixed Models Theory" section.

A statistical model is a mathematical description of how data are generated. The standard linear model, as used by the GLM procedure, is one of the most common statistical models:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

In this expression, **y** represents a vector of observed data, $\boldsymbol{\beta}$ is an unknown vector of fixed-effects parameters with known design matrix **X**, and $\boldsymbol{\epsilon}$ is an unknown random error vector modeling the statistical noise around $\mathbf{X}\boldsymbol{\beta}$. The

focus of the standard linear model is to model the mean of **y** by using the fixed-effects parameters β. The residual errors ϵ are assumed to be independent and identically distributed Gaussian random variables with mean 0 and variance σ^2.

The mixed model generalizes the standard linear model as follows:

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

Here, γ is an unknown vector of random-effects parameters with known design matrix **Z**, and ϵ is an unknown random error vector whose elements are no longer required to be independent and homogeneous.

To further develop this notion of variance modeling, assume that γ and ϵ are Gaussian random variables that are uncorrelated and have expectations **0** and variances **G** and **R**, respectively. The variance of **y** is thus

$$\mathbf{V} = \mathbf{ZGZ}' + \mathbf{R}$$

Note that, when $\mathbf{R} = \sigma^2\mathbf{I}$ and **Z** = **0**, the mixed model reduces to the standard linear model.

You can model the variance of the data, **y**, by specifying the structure (or form) of **Z**, **G**, and **R**. The model matrix **Z** is set up in the same fashion as **X**, the model matrix for the fixed-effects parameters. For **G** and **R**, you must select some *covariance structure.* Possible covariance structures include

- variance components
- compound symmetry (common covariance plus diagonal)
- unstructured (general covariance)
- autoregressive
- spatial
- general linear
- factor analytic

By appropriately defining the model matrices **X** and **Z**, as well as the covariance structure matrices **G** and **R**, you can perform numerous mixed model analyses.

PROC MIXED Contrasted with Other SAS Procedures

PROC MIXED is a generalization of the GLM procedure in the sense that PROC GLM fits standard linear models, and PROC MIXED fits the wider class of mixed linear models. Both procedures have similar CLASS, MODEL, CONTRAST, ESTIMATE, and LSMEANS statements, but their RANDOM and REPEATED statements differ (see the following paragraphs). Both procedures use the nonfull-rank model parameterization, although the sorting of classification levels can differ between the two. PROC MIXED computes only Type I -Type III tests of fixed effects, while PROC GLM offers Types I - IV. The RANDOM statement in PROC MIXED incorporates random effects constituting the γ vector in the mixed model. However, in PROC GLM, effects specified in the RANDOM statement are still treated as fixed as far as the model fit is concerned, and they serve only to produce corresponding

expected mean squares. These expected mean squares lead to the traditional ANOVA estimates of variance components. PROC MIXED computes REML and ML estimates of variance parameters, which are generally preferred to the ANOVA estimates (Searle 1988; Harville 1988; Searle, Casella, and McCulloch 1992). Optionally, PROC MIXED also computes MIVQUE0 estimates, which are similar to ANOVA estimates.

The REPEATED statement in PROC MIXED is used to specify covariance structures for repeated measurements on subjects, while the REPEATED statement in PROC GLM is used to specify various transformations with which to conduct the traditional univariate or multivariate tests. In repeated measures situations, the mixed model approach used in PROC MIXED is more flexible and more widely applicable than either the univariate or multivariate approaches. In particular, the mixed model approach provides a larger class of covariance structures and a better mechanism for handling missing values (Wolfinger and Chang 1995).

PROC MIXED subsumes the VARCOMP procedure. PROC MIXED provides a wide variety of covariance structures, while PROC VARCOMP estimates only simple random effects. PROC MIXED carries out several analyses that are absent in PROC VARCOMP, including the estimation and testing of linear combinations of fixed and random effects.

The ARIMA and AUTOREG procedures provide more time series structures than PROC MIXED, although they do not fit variance component models. The CALIS procedure fits general covariance matrices, but it does not allow fixed effects as does PROC MIXED. The LATTICE and NESTED procedures fit special types of mixed linear models that can also be handled in PROC MIXED, although PROC MIXED may run slower because of its more general algorithm. The TSCSREG procedure analyzes time-series cross-sectional data, and it fits some structures not available in PROC MIXED.

Syntax

The following statements are available in PROC MIXED.

PROC MIXED < options > **;**
BY variables **;**
CLASS variables **;**
ID variables **;**
MODEL dependent = < fixed-effects > < / options > **;**
RANDOM random-effects < / options > **;**
REPEATED < repeated-effect > < / options > **;**
PARMS (value-list) ... < / options > **;**
PRIOR < distribution > < / options > **;**
CONTRAST 'label' < fixed-effect values ... >
< | random-effect values ... > , ... < / options > **;**
ESTIMATE 'label' < fixed-effect values ... >
< | random-effect values ... >< / options > **;**
LSMEANS fixed-effects < / options > **;**
MAKE 'table' **OUT=**SAS-data-set **;**
WEIGHT variable **;**

Items within angle brackets (< >) are optional. The CONTRAST, ESTIMATE, LSMEANS, MAKE, and RANDOM statements can appear multiple times; all other statements can appear only once.

The PROC MIXED and MODEL statements are required, and the MODEL statement must appear after the CLASS statement if a CLASS statement is included. The CONTRAST, ESTIMATE, LSMEANS, RANDOM, and REPEATED statements must follow the MODEL statement. The CONTRAST and ESTIMATE statements must also follow any RANDOM statements.

Table 41.1 summarizes the basic functions and important options of each PROC MIXED statement. The syntax of each statement in Table 41.1 is described in the following sections in alphabetical order after the description of the PROC MIXED statement.

Table 41.1: Summary of PROC MIXED Statements

Statement	Description	Important Options
PROC MIXED	invokes the procedure	DATA= specifies input data set, METHOD= specifies estimation method
BY	performs multiple PROC MIXED analyses in one invocation	none
CLASS	declares qualitative variables that create indicator variables in design matrices	none
ID	lists additional variables to be included in predicted values tables	none
MODEL	specifies dependent variable and fixed effects, setting up **X**	S requests solution for fixed-effects parameters, DDFM= specifies denominator degrees of freedom method, OUTP= outputs predicted values to a data set
RANDOM	specifies random effects, setting up **Z** and **G**	SUBJECT= creates block-diagonality, TYPE= specifies covariance structure, S requests solution for random-effects parameters, G displays estimated **G**
REPEATED	sets up **R**	SUBJECT= creates block-diagonality, TYPE= specifies covariance structure, R displays estimated blocks of **R**, GROUP= enables between-subject heterogeneity, LOCAL adds a diagonal matrix to **R**
PARMS	specifies a grid of initial values for the covariance parameters	HOLD= and NOITER hold the covariance parameters or their ratios constant, PDATA= reads the initial values from a SAS data set
PRIOR	performs a sampling-based Bayesian analysis for variance component models	NSAMPLE= specifies the sample size, SEED= specifies the starting seed
CONTRAST	constructs custom hypothesis tests	E displays the **L** matrix coefficients
ESTIMATE	constructs custom scalar estimates	CL produces confidence limits
LSMEANS	computes least squares means for classification fixed effects	DIFF computes differences of the least squares means, ADJUST= performs multiple comparisons adjustments, AT changes covariates, OM changes weighting, CL produces confidence limits, SLICE= tests simple effects
MAKE	converts any displayed table into a SAS data set	none. Has been superceded by the Output Delivery System (ODS)
WEIGHT	specifies a variable by which to weight **R**	none

PROC MIXED Statement

PROC MIXED < *options* >**;**

The PROC MIXED statement invokes the procedure. You can specify the following options.

ABSOLUTE
makes the convergence criterion absolute. By default, it is relative (divided by the current objective function value). See the CONVF, CONVG, and CONVH options in this section for a description of various convergence criteria.

ALPHA=*number*
requests that confidence limits be constructed for the covariance parameter estimates with confidence level 1-number. The value of *number* must be between 0 and 1; the default is 0.05.

ASYCORR
produces the asymptotic correlation matrix of the covariance parameter estimates. It is computed from the corresponding asymptotic covariance matrix (see the description of the ASYCOV option, which follows). For ODS purposes, the label of the "Asymptotic Correlation" table is "AsyCorr."

ASYCOV
requests that the asymptotic covariance matrix of the covariance parameters be displayed. By default, this matrix is the observed inverse Fisher information matrix, which equals $2\mathbf{H}^{-1}$, where **H** is the Hessian (second derivative) matrix of the objective function. See the "Covariance Parameter Estimates" section for more information about this matrix. When you use the SCORING= option and PROC MIXED converges without stopping the scoring algorithm, PROC MIXED uses the expected Hessian matrix to compute the covariance matrix instead of the observed Hessian. For ODS purposes, the label of the "Asymptotic Covariance" table is "AsyCov."

CL<=WALD>
requests confidence limits for the covariance parameter estimates. A Satterthwaite approximation is used to construct limits for all parameters that have a default lower boundary constraint of zero. These limits take the form

$$\frac{\nu\hat{\sigma}^2}{\chi^2_{\nu,1-\alpha/2}} \leq \sigma^2 \leq \frac{\nu\hat{\sigma}^2}{\chi^2_{\nu,\alpha/2}}$$

where $\nu = 2Z^2$, Z is the Wald statistic $\hat{\sigma}^2/\mathrm{se}(\hat{\sigma}^2)$, and the denominators are quantiles of the χ^2-distribution with ν degrees of freedom. Refer to Milliken and Johnson (1992) and Burdick and Graybill (1992) for similar techniques.

For all other parameters, Wald *Z*-scores and normal quantiles are used to construct the limits. The optional =WALD specification requests Wald limits for all parameters.

The confidence limits are displayed as extra columns in the "Covariance Parameter Estimates" table. The confidence level is $1-\alpha = 0.95$ by default; this can be changed with the ALPHA= option.

CONVF<=*number*>
requests the relative function convergence criterion with tolerance *number*. The relative function convergence criterion is

$$\frac{|f_k - f_{k-1}|}{|f_k|} \leq number$$

where f_k is the value of the objective function at iteration k. To prevent the division by $|f_k|$, use the ABSOLUTE option. The default convergence criterion is CONVH, and the default tolerance is 1E-8.

CONVG <=*number*>
requests the relative gradient convergence criterion with tolerance *number*. The relative gradient convergence criterion is

$$\frac{\max_j |g_{jk}|}{|f_k|} \leq number$$

where f_k is the value of the objective function, and g_{jk} is the jth element of the gradient (first derivative) of the objective function, both at iteration k. To prevent division by $|f_k|$, use the ABSOLUTE option. The default convergence criterion is CONVH, and the default tolerance is 1E-8.

CONVH<=*number*>
requests the relative Hessian convergence criterion with tolerance *number*. The relative Hessian convergence criterion is

$$\frac{\mathbf{g}_k' \mathbf{H}_k^{-1} \mathbf{g}_k}{|f_k|} \leq number$$

where f_k is the value of the objective function, $\mathbf{g}_k$ is the gradient (first derivative) of the objective function, and $\mathbf{H}_k$ is the Hessian (second derivative) of the objective function, all at iteration k.

If $\mathbf{H}_k$ is singular, then PROC MIXED uses the following relative criterion:

$$\frac{\mathbf{g}_k' \mathbf{g}_k}{|f_k|} \leq number$$

To prevent the division by $|f_k|$, use the ABSOLUTE option. The default convergence criterion is CONVH, and the default tolerance is 1E-8.

COVTEST
produces asymptotic standard errors and Wald *Z*-tests for the covariance parameter estimates.

DATA=*SAS-data-set*

names the SAS data set to be used by PROC MIXED. The default is the most recently created data set.

DFBW

has the same effect as the DDFM=BW option in the MODEL statement.

EMPIRICAL

computes the estimated variance-covariance matrix of the fixed-effects parameters by using the asymptotically consistent estimator described in Huber (1967), White (1980), Liang and Zeger (1986), and Diggle, Liang, and Zeger (1994). This estimator is commonly referred to as the "sandwich" estimator, and it is computed as follows:

$$(\mathbf{X}'\widehat{\mathbf{V}}^{-1}\mathbf{X})^{-}\left(\sum_{i=1}^{S}\mathbf{X}_i'\widehat{\mathbf{V}}_i^{-1}\hat{\boldsymbol{\epsilon}}_i\hat{\boldsymbol{\epsilon}}_i'\widehat{\mathbf{V}}_i^{-1}\mathbf{X}_i\right)(\mathbf{X}'\widehat{\mathbf{V}}^{-1}\mathbf{X})^{-}$$

$$\hat{\boldsymbol{\epsilon}}_i = \mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}$$

Here, , *S* is the number of subjects, and matrices with an *i* subscript are those for the *i*th subject. You must include the SUBJECT= option in either a RANDOM or REPEATED statement for this option to take effect.

When you specify the EMPIRICAL option, PROC MIXED adjusts all standard errors and test statistics involving the fixed-effects parameters. This changes output in the following tables (listed in Table 41.7): Contrast, CorrB, CovB, Diffs, Estimates, InvCovB, LSMeans, MMEq, MMEqSol, Slices, SolutionF, Tests1 -Tests3. The OUTP= and OUTPM= data sets are also affected. Finally, the Satterthwaite and Kenward-Roger degrees of freedom methods are not available if you specify EMPIRICAL.

IC

displays a table of various information criteria. Four different criteria are computed in four different ways, producing 16 values in all. Table 41.2 displays the four criteria in both larger-is-better and smaller-is-better forms.

Table 41.2: Information Criteria

Criteria	Larger-is-better	Smaller-is-better	Reference
AIC	*l- d*	-2*l*+ 2*d*	Akaike (1974)
HQIC	*l- d loglogn*	-2*l*+ 2*d loglogn*	Hannan and Quinn (1979)
BIC	*l- d*/2 *logn*	-2*l*+ *d logn*	Schwarz (1978)
CAIC	*l- d*(*logn* + 1)/2	-2*l*+ *d*(*logn* + 1)	Bozdogan (1987)

Here *l* denotes the maximum value of the (possibly restricted) log likelihood, *d* the dimension of the model, and *n* the number of effective observations. In Version 6 of SAS/STAT software, *n* equals the number of valid observations for maximum likelihood estimation and *n-p* for restricted maximum likelihood estimation, where *p* equals the rank of **X**. In later versions, *n* equals the number of effective subjects as displayed in the "Dimensions" table, unless this value equals 1, in which case *n* reverts to the Version 6 values.

PROC MIXED evaluates the criteria for both forms using *d* equal to both *q* and *q+p*, where *q* is the

effective number of estimated covariance parameters. In Version 6, when a parameter estimate lies on a boundary constraint, then it is still included in the calculation of *d*, but in later versions it is not. The most common example of this behavior is when a variance component is estimated to equal zero.

For ODS purposes, the name of the "Information Criteria" table is "InfoCrit."

INFO
is a default option. The creation of the "Model Information" and "Dimensions" tables can be suppressed using the NOINFO option.

Note that, in Version 6, this option displays the "Model Information" and "Dimensions" tables.

ITDETAILS
displays the parameter values at each iteration and enables the writing of notes to the SAS log pertaining to "infinite likelihood" and "singularities" during Newton-Raphson iterations.

LOGNOTE
writes periodic notes to the log describing the current status of computations. It is designed for use with analyses requiring extensive CPU resources.

MAXFUNC=*number*
specifies the maximum number of likelihood evaluations in the optimization process. The default is 150.

MAXITER=*number*
specifies the maximum number of iterations. The default is 50.

METHOD=REML
METHOD=ML
METHOD=MIVQUE0
METHOD=TYPE1
METHOD=TYPE2
METHOD=TYPE3
specifies the estimation method for the covariance parameters. The REML specification performs residual (restricted) maximum likelihood, and it is the default method. The ML specification performs maximum likelihood, and the MIVQUE0 specification performs minimum variance quadratic unbiased estimation of the covariance parameters.

The METHOD=TYPE*n* specifications apply only to variance component models with no SUBJECT= effects and no REPEATED statement. An analysis of variance table is included in the output, and the expected mean squares are used to estimate the variance components (refer to Chapter 30, "The GLM Procedure," for further explanation). The resulting method-of-moment variance component estimates are used in subsequent calculations, including standard errors computed from ESTIMATE and LSMEANS statements. For ODS purposes, the new table names are "Type1," "Type2," and "Type3," respectively.

MMEQ
requests that coefficients of the mixed model equations be displayed. These are

$$\begin{bmatrix} \mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{X} & \mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{Z} \\ \mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{X} & \mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{Z}+\hat{\mathbf{G}}^{-1} \end{bmatrix}, \begin{bmatrix} \mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{y} \\ \mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{y} \end{bmatrix}$$

assuming that $\hat{\mathbf{G}}$ is nonsingular. If $\hat{\mathbf{G}}$ is singular, PROC MIXED produces the following coefficients

$$\begin{bmatrix} \mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{X} & \mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{Z}\hat{\mathbf{G}} \\ \hat{\mathbf{G}}\mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{X} & \hat{\mathbf{G}}\mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{Z}\hat{\mathbf{G}}+\hat{\mathbf{G}} \end{bmatrix}, \begin{bmatrix} \mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{y} \\ \hat{\mathbf{G}}\mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{y} \end{bmatrix}$$

See the "Estimating and in the Mixed Model" section for further information on these equations.

MMEQSOL
requests that a solution to the mixed model equations be produced, as well as the inverted coefficients matrix. Formulas for these equations are provided in the preceding description of the MMEQ option.

When $\hat{\mathbf{G}}$ is singular, $\hat{\boldsymbol{\tau}}$ and a generalized inverse of the left-hand-side coefficient matrix are transformed using $\hat{\mathbf{G}}$ to produce $\hat{\boldsymbol{\gamma}}$ and $\hat{\mathbf{C}}$, respectively, where $\hat{\mathbf{C}}$ is a generalized inverse of the left-hand-side coefficient matrix of the original equations.

NAMELEN<=*number*>
specifies the length to which long effect names are shortened. The default and minimum value is 20.

NOBOUND
has the same effect as the NOBOUND option in the PARMS statement.

NOCLPRINT<=*number*>
suppresses the display of the "Class Level Information" table if you do not specify *number*. If you do specify *number*, only levels with totals that are less than *number* are listed in the table.

NOINFO
suppresses the display of the "Model Information" and "Dimensions" tables.

NOITPRINT
suppresses the display of the "Iteration History" table.

NOPROFILE
includes the residual variance as part of the Newton-Raphson iterations. This option applies only to models that have a residual variance parameter. By default, this parameter is profiled out of the likelihood calculations, except when you have specified the HOLD= or NOITER option in the PARMS statement.

ORD
displays ordinates of the relevant distribution in addition to *p*-values. The ordinate can be viewed as an approximate odds ratio of hypothesis probabilities.

ORDER=DATA
ORDER=FORMATTED
ORDER=FREQ
ORDER=INTERNAL
specifies the sorting order for the levels of all CLASS variables. This ordering determines which parameters in the model correspond to each level in the data, so the ORDER= option may be useful when you use CONTRAST or ESTIMATE statements.

The default is ORDER=FORMATTED, and its behavior has been modified for Version 8. Now, for numeric variables for which you have supplied no explicit format (that is, for which there is no

corresponding FORMAT statement in the current PROC MIXED run or in the DATA step that created the data set), the levels are ordered by their internal (numeric) value. In releases previous to Version 8, numeric class levels with no explicit format were ordered by their BEST12. formatted values. In order to revert to the previous method you can specify this format explicitly for the CLASS variables. The change was implemented because the former default behavior for ORDER=FORMATTED often resulted in levels not being ordered numerically and required you to use an explicit format or ORDER=INTERNAL to get the more natural ordering.

The following table shows how PROC MIXED interprets values of the ORDER= option.

Value of ORDER=	Levels Sorted By
DATA	order of appearance in the input data set
FORMATTED	external formatted value, except for numeric
	variables with no explicit format, which are
	sorted by their unformatted (internal) value
FREQ	descending frequency count; levels with the
	most observations come first in the order
INTERNAL	unformatted value

For FORMATTED and INTERNAL, the sort order is machine dependent. For more information on sorting order, see the chapter on the SORT procedure in the *SAS Procedures Guide* and the discussion of BY-group processing in *SAS Language Reference: Concepts*.

RATIO
produces the ratio of the covariance parameter estimates to the estimate of the residual variance when the latter exists in the model.

RIDGE=*number*
specifies the starting value for the minimum ridge value used in the Newton-Raphson algorithm. The default is 0.3125.

SCORING<=*number*>
requests that Fisher scoring be used in association with the estimation method up to iteration *number*, which is 0 by default. When you use the SCORING= option and PROC MIXED converges without stopping the scoring algorithm, PROC MIXED uses the expected Hessian matrix to compute approximate standard errors for the covariance parameters instead of the observed Hessian. The output from the ASYCOV and ASYCORR options is similarly adjusted.

SIGITER
is an alias for the NOPROFILE option.

UPDATE
is an alias for the LOGNOTE option.

BY Statement

BY *variables* **;**

You can specify a BY statement with PROC MIXED to obtain separate analyses on observations in groups defined by the BY variables. When a BY statement appears, the procedure expects the input data set to be sorted in order of the BY variables. The *variables* are one or more variables in the input data set.

If your input data set is not sorted in ascending order, use one of the following alternatives:

- Sort the data using the SORT procedure with a similar BY statement.
- Specify the BY statement options NOTSORTED or DESCENDING in the BY statement for the MIXED procedure. The NOTSORTED option does not mean that the data are unsorted but rather means that the data are arranged in groups (according to values of the BY variables) and that these groups are not necessarily in alphabetical or increasing numeric order.
- Create an index on the BY variables using the DATASETS procedure (in base SAS software).

Since sorting the data changes the order in which PROC MIXED reads observations, the sorting order for the levels of the CLASS variable may be affected if you have specified ORDER=DATA in the PROC MIXED statement. This, in turn, affects specifications in the CONTRAST statements.

For more information on the BY statement, refer to the discussion in *SAS Language Reference: Concepts*. For more information on the DATASETS procedure, refer to the discussion in the *SAS Procedures Guide*.

CLASS Statement

CLASS *variables* ***;***

The CLASS statement names the classification variables to be used in the analysis. If the CLASS statement is used, it must appear before the MODEL statement. Classification variables can be either character or numeric. The procedure uses only the first 16 characters of a character variable. Class levels are determined from the formatted values of the CLASS variables. Thus, you can use formats to group values into levels. Refer to the discussion of the FORMAT procedure in the *SAS Procedures Guide* and to the discussions of the FORMAT statement and SAS formats in *SAS Language Reference: Dictionary*. You can adjust the display order of CLASS variable levels with the ORDER= option in the PROC MIXED statement.

CONTRAST Statement

CONTRAST *'label'* < *fixed-effect values ...*>
< | *random-effect values ...*> , ...< / *options* > ***;***

The CONTRAST statement provides a mechanism for obtaining custom hypothesis tests. It is patterned after the CONTRAST statement in PROC GLM, although it has been extended to include random effects. This enables you to select an appropriate inference space (McLean, Sanders, and Stroup 1991).

$$\mathbf{L}'\boldsymbol{\phi} = 0 \qquad \boldsymbol{\phi}' = (\boldsymbol{\beta}' \; \boldsymbol{\gamma}')$$

You can test the hypothesis , where **L**' = (**K**' **M**') and , in several inference spaces. The inference space corresponds to the choice of **M**. When **M** = **0**, your inferences apply to the entire population from which the random effects are sampled; this is known as the *broad* inference space. When all elements of **M** are nonzero, your inferences apply only to the observed levels of the random effects. This is known as the *narrow* inference space, and you can also choose it by specifying all of the random effects as fixed. The GLM procedure uses the narrow inference space. Finally, by zeroing portions of **M** corresponding to selected main effects and interactions, you can choose *intermediate* inference spaces. The broad inference space is usually the most appropriate, and it is used when you do not specify any random effects in the CONTRAST statement.

In the CONTRAST statement,
label

identifies the contrast in the table. A label is required for every contrast specified. Labels can be up to 20 characters and must be enclosed in single quotes.

fixed-effect
identifies an effect that appears in the MODEL statement. The keyword INTERCEPT can be used as an effect when an intercept is fitted in the model. You do not need to include all effects that are in the MODEL statement.

random-effect
identifies an effect that appears in the RANDOM statement. The first random effect must follow a vertical bar (|); however, random effects do not have to be specified.

values
are constants that are elements of the **L** matrix associated with the fixed and random effects.

The rows of **L**' are specified in order and are separated by commas. The rows of the **K**' component of **L**' are specified on the left side of the vertical bars (|). These rows test the fixed effects and are, therefore, checked for estimability. The rows of the **M**' component of **L**' are specified on the right side of the vertical bars. They test the random effects, and no estimability checking is necessary.

If PROC MIXED finds the fixed-effects portion of the specified contrast to be nonestimable (see the SINGULAR= option), then it displays "Non-est" for the contrast entries.

The following CONTRAST statement reproduces the *F*-test for the effect A in the split-plot example (see Example 41.1):

```
contrast 'A broad'
      A  1 -1 0    A*B  .5 .5 -.5 -.5 0 0 ,
      A  1 0 -1    A*B  .5 .5 0 0 -.5 -.5 / df=6;
```

Note that no random effects are specified in the preceding contrast; thus, the inference space is broad. The resulting *F*-test has two numerator degrees of freedom because **L**' has two rows. The denominator degrees of freedom is, by default, the residual degrees of freedom (9), but the DF= option changes the denominator degrees of freedom to 6. The following CONTRAST statement reproduces the *F*-test for A when Block and A*Block are considered fixed effects (the narrow inference space):

```
contrast 'A narrow'
      A       1 -1 0
      A*B     .5 .5 -.5 -.5 0 0 |
      A*Block .25  .25  .25  .25
            -.25 -.25 -.25 -.25
              0    0    0    0 ,
      A       1 0 -1
      A*B     .5 .5 0 0 -.5 -.5 |
      A*Block .25  .25  .25  .25
              0    0    0    0
            -.25 -.25 -.25 -.25 ;
```

The preceding contrast does not contain coefficients for B and Block because they cancel out in estimated differences between levels of A. Coefficients for B and Block are necessary when estimating the mean of one of the levels of A in the narrow inference space (see Example 41.1).

If the elements of **L** are not specified for an effect that contains a specified effect, then the elements of the specified effect are automatically "filled in" over the levels of the higher-order effect. This feature is designed to preserve estimability for cases when there are complex higher-order effects. The coefficients for the higher-order effect are determined by equitably distributing the coefficients of the lower-level effect as in the construction of least squares means. In addition, if the intercept is specified, it is distributed over all classification effects that are not contained by any other specified effect. If an effect is not specified and does not contain any specified effects, then all of its

coefficients in **L** are set to 0. You can override this behavior by specifying coefficients for the higher-order effect.

If too many values are specified for an effect, the extra ones are ignored; if too few are specified, the remaining ones are set to 0. If no random effects are specified, the vertical bar can be omitted; otherwise, it must be present. If a SUBJECT effect is used in the RANDOM statement, then the coefficients specified for the effects in the RANDOM statement are equitably distributed across the levels of the SUBJECT effect. You can use the E option to see exactly what **L** matrix is used.

The SUBJECT and GROUP options in the CONTRAST statement are useful for the case when a SUBJECT= or GROUP= variable appears in the RANDOM statement, and you want to contrast different subjects or groups. By default, CONTRAST statement coefficients on random effects are distributed equally across subjects and groups.

PROC MIXED handles missing level combinations of classification variables similarly to the way PROC GLM does. Both procedures delete fixed-effects parameters corresponding to missing levels in order to preserve estimability. However, PROC MIXED does not delete missing level combinations for random-effects parameters because linear combinations of the random-effects parameters are always estimable. These conventions can affect the way you specify your CONTRAST coefficients.

The CONTRAST statement computes the statistic

$$F = \frac{\begin{bmatrix} \hat{\beta} \\ \hat{\gamma} \end{bmatrix}' \mathbf{L}'(\mathbf{L}'\hat{\mathbf{C}}\mathbf{L})^{-1}\mathbf{L} \begin{bmatrix} \hat{\beta} \\ \hat{\gamma} \end{bmatrix}}{\mathrm{rank}(\mathbf{L})}$$

and approximates its distribution with an *F*-distribution. In this expression, $\hat{\mathbf{C}}$ is an estimate of the generalized inverse of the coefficient matrix in the mixed model equations. See the "Inference and Test Statistics" section for more information on this *F*-statistic.

The numerator degrees of freedom in the *F*-approximation is *rank*(**L**), and the denominator degrees of freedom is taken from the "Tests of Fixed Effects" table and corresponds to the final effect you list in the CONTRAST statement. You can change the denominator degrees of freedom by using the DF= option.

You can specify the following options in the CONTRAST statement after a slash (/).

CHISQ
requests that χ^2-tests be performed in addition to any *F*-tests. A χ^2-statistic equals its corresponding *F*-statistic times the associate numerator degrees of freedom, and this same degrees of freedom is used to compute the *p*-value for the χ^2-test. This *p*-value will always be less than that for the *F*-test, as it effectively corresponds to an *F*-test with infinite denominator degrees of freedom.

DF=*number*
specifies the denominator degrees of freedom for the *F*-test. The default is the denominator degrees of freedom taken from the "Tests of Fixed Effects" table and corresponds to the final effect you list in the CONTRAST statement.

E
requests that the **L** matrix coefficients for the contrast be displayed. For ODS purposes, the label of this "L Matrix Coefficients" table is "Coefficients".

GROUP ***coeffs***
GRP ***coeffs***

sets up random-effect contrasts between different groups when a GROUP= variable appears in the RANDOM statement. By default, CONTRAST statement coefficients on random effects are distributed equally across groups.

SINGULAR=*number*

tunes the estimability checking. If **v** is a vector, define ABS(**v**) to be the absolute value of the element of **v** with the largest absolute value. If ABS(**K**'-**K**'**T**) is greater than C***number* for any row of **K**' in the contrast, then **K** is declared nonestimable. Here **T** is the Hermite form matrix (**X**'**X**)⁻**X**'**X**, and C is ABS(**K**') except when it equals 0, and then C is 1. The value for *number* must be between 0 and 1; the default is 1E-4.

SUBJECT ***coeffs***
SUB ***coeffs***

sets up random-effect contrasts between different subjects when a SUBJECT= variable appears on the RANDOM statement. By default, CONTRAST statement coefficients on random effects are distributed equally across subjects.

ESTIMATE Statement

ESTIMATE *'label' < fixed-effect values ...>*
< | random-effect values ...> , ...< / options > ***;***

The ESTIMATE statement is exactly like a CONTRAST statement, except only one-row **L** matrices are permitted. $\mathbf{L}'\hat{\mathbf{p}}$ $\mathbf{L}'\hat{\mathbf{p}}$ The actual estimate, , is displayed along with its approximate standard error. An approximate *t*-test that = 0 is also produced.

PROC MIXED selects the degrees of freedom to match those displayed in the "Tests of Fixed Effects" table for the final effect you list in the ESTIMATE statement. You can modify the degrees of freedom using the DF= option.

If PROC MIXED finds the fixed-effects portion of the specified estimate to be nonestimable, then it displays "Non-est" for the estimate entries.

The following examples of ESTIMATE statements compute the mean of the first level of A in the split-plot example (see Example 41.1) for various inference spaces.

```
estimate 'A1 mean narrow'   intercept 1
                            A 1 B .5 .5 A*B .5 .5 |
                            block .25 .25 .25 .25
                            A*Block .25 .25 .25 .25
                                  0 0 0 0
                                  0 0 0 0;
estimate 'A1 mean intermed' intercept 1
                            A 1 B .5 .5 A*B .5 .5 |
                            Block .25 .25 .25 .25;
estimate 'A1 mean broad'    intercept 1
                            A 1 B .5 .5 A*B .5 .5;
```

The construction of the **L** vector for an ESTIMATE statement follows the same rules as listed under the CONTRAST statement.

You can specify the following options in the ESTIMATE statement after a slash (/).

ALPHA=*number*

requests that a *t*-type confidence interval be constructed with confidence level 1-number. The value of *number* must be between 0 and 1; the default is 0.05.

CL

requests that *t*-type confidence limits be constructed. The confidence level is 0.95 by default; this can be changed with the ALPHA= option.

DF=*number*

specifies the degrees of freedom for the *t*-test and confidence limits. The default is the denominator degrees of freedom taken from the "Tests of Fixed Effects" table and corresponds to the final effect you list in the ESTIMATE statement.

DIVISOR=*number*

specifies a value by which to divide all coefficients so that fractional coefficients can be entered as integer numerators.

E

requests that the **L** matrix coefficients be displayed. For ODS purposes, the label of this "L Matrix Coefficients" table is "Coefficients".

GROUP *coeffs*
GRP *coeffs*

sets up random-effect contrasts between different groups when a GROUP= variable appears in the RANDOM statement. By default, ESTIMATE statement coefficients on random effects are distributed equally across groups.

LOWER
LOWERTAILED

requests that the *p*-value for the *t*-test be based only on values less than the *t*-statistic. A two-tailed test is the default. A lower-tailed confidence limit is also produced if you specify the CL option.

SINGULAR=*number*

tunes the estimability checking as documented for the CONTRAST statement.

SUBJECT *coeffs*
SUB *coeffs*

sets up random-effect contrasts between different subjects when a SUBJECT= variable appears in the RANDOM statement. By default, ESTIMATE statement coefficients on random effects are distributed equally across subjects.

For example, the ESTIMATE statement in the following code from Example 41.5 constructs the difference between the random slopes of the first two batches.

```
proc mixed data=rc;
  class batch;
  model y = month / s;
  random int month / type=un sub=batch s;
  estimate 'slope b1 - slope b2' | month 1 / subject 1 -1;
run;
```

UPPER
UPPERTAILED

requests that the *p*-value for the *t*-test be based only on values greater than the *t*-statistic. A two-tailed test is the default. An upper-tailed confidence limit is also produced if you specify the CL option.

ID Statement

ID *variables* ***;***

The ID statement specifies which variables from the input data set are to be included in the OUTP= and OUTPM= data sets from the MODEL statement. If you do not specify an ID statement, then all variables are included in these data sets. Otherwise, only the variables you list in the ID statement are included. Specifying an ID statement with no variables prevents any variables from being included in these data sets.

LSMEANS Statement

LSMEANS *fixed-effects* < */ options* > ***;***

The LSMEANS statement computes least-squares means (LS-means) of fixed effects. As in the GLM procedure, LS-means are *predicted population margins* -that is, they estimate the marginal means over a balanced population. In a sense, LS-means are to unbalanced designs as class and subclass arithmetic means are to balanced designs. The **L** matrix constructed to compute them is the same as the **L** matrix formed in PROC GLM; however, the standard errors are adjusted for the covariance parameters in the model.

Each LS-mean is computed as $\mathbf{L}\hat{\beta}$ where **L** is the coefficient matrix associated with the least-squares mean and $\hat{\beta}$ is the estimate of the fixed-effects parameter vector (see the "Estimating and in the Mixed Model" section). The approximate standard errors for the LS-mean is computed as the square root of $\mathbf{L}(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-}\mathbf{L}'$.

LS-means can be computed for any effect in the MODEL statement that involves CLASS variables. You can specify multiple effects in one LSMEANS statement or in multiple LSMEANS statements, and all LSMEANS statements must appear after the MODEL statement. As in the ESTIMATE statement, the **L** matrix is tested for estimability, and if this test fails, PROC MIXED displays "Non-est" for the LS-means entries.

Assuming the LS-mean is estimable, PROC MIXED constructs an approximate *t*-test to test the null hypothesis that the associated population quantity equals zero. By default, the denominator degrees of freedom for this test are the same as those displayed for the effect in the "Tests of Fixed Effects" table (see the "Default Output" section). You can specify the following options in the LSMEANS statement after a slash (/).

ADJUST=BON
ADJUST=DUNNETT
ADJUST=SCHEFFE
ADJUST=SIDAK
ADJUST=SIMULATE<*(simoptions)*>
ADJUST=SMM | GT2
ADJUST=TUKEY

requests a multiple comparison adjustment for the *p*-values and confidence limits for the differences of LS-means. By default, PROC MIXED adjusts all pairwise differences unless you specify ADJUST=DUNNETT, in which case PROC MIXED analyzes all differences with a control level. The ADJUST= option implies the DIFF option.

The BON (Bonferroni) and SIDAK adjustments involve correction factors described in Chapter 30, "The GLM Procedure," and Chapter 43, "The MULTTEST Procedure"; also refer to Westfall and Young (1993). When you specify ADJUST=TUKEY and your data are unbalanced, PROC MIXED uses the approximation described in Kramer (1956). Similarly, when you specify ADJUST=DUNNETT and the LS-

means are correlated, PROC MIXED uses the factor-analytic covariance approximation described in Hsu (1992). The preceding references also describe the SCHEFFE and SMM adjustments.

The SIMULATE adjustment computes adjusted *p*-values and confidence limits from the simulated distribution of the maximum or maximum absolute value of a multivariate *t* random vector. All covariance parameters except the residual variance are fixed at their estimated values throughout the simulation, potentially resulting in some underdispersion. The simulation estimates *q*, the true $(1-\alpha)$ th quantile, where $1-\alpha$ is the confidence coefficient. The default α is 0.05, and you can change this value with the ALPHA= option in the LSMEANS statement.

The number of samples is set so that the tail area for the simulated *q* is within γ of $1-\alpha$ with $100(1-\epsilon)$% confidence. In equation form,

$$P\left(\left|F(\hat{q}) - (1-\alpha)\right| \leq \gamma\right) \quad = \quad 1-\epsilon$$

where $\hat{q}$ is the simulated *q* and *F* is the true distribution function of the maximum; refer to Edwards and Berry (1987) for details. By default, γ = 0.005 and ϵ = 0.01, placing the tail area of $\hat{q}$ within 0.005 of 0.95 with 99% confidence. The ACC= and EPS= *simoptions* reset γ and ϵ, respectively; the NSAMP= *simoption* sets the sample size directly; and the SEED= *simoption* enables you to control the beginning of the random number sequence (the clock time is used by default). For additional description of these and other simulation options, see the "LSMEANS Statement" section in Chapter 30, "The GLM Procedure."

ALPHA=*number*
requests that a *t*-type confidence interval be constructed for each of the LS-means with confidence level 1-number. The value of *number* must be between 0 and 1; the default is 0.05.

AT *variable* = *value*
AT *(variable-list) = (value-list)*
AT MEANS
enables you to modify the values of the covariates used in computing LS-means. By default, all covariate effects are set equal to their mean values for computation of standard LS-means. The AT option enables you to assign arbitrary values to the covariates. Additional columns in the output table indicate the values of the covariates.

If there is an effect containing two or more covariates, the AT option sets the effect equal to the product of the individual means rather than the mean of the product (as with standard LS-means calculations). The AT MEANS option sets covariates equal to their mean values (as with standard LS-means) and incorporates this adjustment to cross products of covariates.

As an example, consider the following invocation of PROC MIXED:

```
proc mixed;
  class A;
  model Y = A X1 X2 X1*X2;
```

```
  lsmeans A;
  lsmeans A / at means;
  lsmeans A / at X1=1.2;
  lsmeans A / at (X1 X2)=(1.2 0.3);
run;
```

For the first two LSMEANS statements, the LS-means coefficient for X1 is $\overline{x_1}$ (the mean of X1) and for X2 is $\overline{x_2}$ (the mean of X2). However, for the first LSMEANS statement, the coefficient for X1*X2 is $\overline{x_1 x_2}$, but for the second LSMEANS statement, the coefficient is $\overline{x_1} \cdot \overline{x_2}$. The third LSMEANS statement sets the coefficient for X1 equal to 1.2 and leaves it at $\overline{x_2}$ for X2, and the final LSMEANS statement sets these values to 1.2 and 0.3, respectively.

If a WEIGHT variable is present, it is used in processing AT variables. Also, observations with missing dependent variables are included in computing the covariate means, unless these observations form a missing cell and the FULLX option in the MODEL statement is not in effect. You can use the E option in conjunction with the AT option to check that the modified LS-means coefficients are the ones you desire.

The AT option is disabled if you specify the BYLEVEL option.

BYLEVEL
requests PROC MIXED to process the OM data set by each level of the LS-mean effect (LSMEANS effect) in question. For more details, see the OM option later in this section.

CL
requests that *t*-type confidence limits be constructed for each of the LS-means. The confidence level is 0.95 by default; this can be changed with the ALPHA= option.

CORR
displays the estimated correlation matrix of the least-squares means as part of the "Least Squares Means" table.

COV
displays the estimated covariance matrix of the least-squares means as part of the "Least Squares Means" table.

DF=*number*
specifies the degrees of freedom for the *t*-test and confidence limits. The default is the denominator degrees of freedom taken from the "Tests of Fixed Effects" table corresponding to the LS-means effect.

DIFF<=*difftype*>
PDIFF<=*difftype*>
requests that differences of the LS-means be displayed. The optional *difftype* specifies which differences to produce, with possible values being ALL, CONTROL, CONTROLL, and CONTROLU. The *difftype* ALL requests all pairwise differences, and it is the default. The *difftype* CONTROL requests the differences with a control, which, by default, is the first level of each of the specified LSMEANS effects.

To specify which levels of the effects are the controls, list the quoted formatted values in parentheses after the keyword CONTROL. For example, if the effects A, B, and C are class variables, each having two levels, 1 and 2, the following LSMEANS statement specifies the (1,2) level of A*B and the (2,1) level of B*C as controls:

lsmeans A*B B*C / diff=control('1' '2' '2' '1');

For multiple effects, the ordering of the list is significant, and you should check the output to make sure that the controls are correct.

Two-tailed tests and confidence limits are associated with the CONTROL *difftype*. For one-tailed results, use either the CONTROLL or CONTROLU *difftype*. The CONTROLL *difftype* tests whether the noncontrol levels are significantly smaller than the control; the upper confidence limits for the control minus the noncontrol levels are considered to be infinity and are displayed as missing. Conversely, the CONTROLU *difftype* tests whether the noncontrol levels are significantly larger than the control; the upper confidence limits for the noncontrol levels minus the control are considered to be infinity and are displayed as missing.

If you want to perform multiple comparison adjustments on the differences of LS-Means, use the ADJUST= option. For DIFF=ALL (the default), ADJUST=TUKEY is the default method, and in all other instances, the default ADJUST= option is DUNNETT. If there is a conflict between the DIFF= and ADJUST= options, the ADJUST= option takes precedence.

The differences of the LS-means are displayed in a table titled "Differences of Least Squares Means." For ODS purposes, the table name is "Diffs."

E

requests that the **L** matrix coefficients for all LSMEANS effects be displayed. For ODS purposes, the label of this "**L** Matrix Coefficients" table is "Coefficients".

OM<**=***OM-data-set*>
OBSMARGINS<**=***OM-data-set*>

specifies a potentially different weighting scheme for the computation of LS-means coefficients. The standard LS-means have equal coefficients across classification effects; however, the OM option changes these coefficients to be proportional to those found in *OM-data-set*. This adjustment is reasonable when you want your inferences to apply to a population that is not necessarily balanced but has the margins observed in *OM-data-set*.

By default, *OM-data-set* is the same as the analysis data set. You can optionally specify another data set that describes the population for which you want to make inferences. This data set must contain all model variables except for the dependent variable (which is ignored if it is present). In addition, the levels of all CLASS variables must be the same as those occurring in the analysis data set. Specifying an *OM-data-set* enables you to construct arbitrarily weighted LS-means.

In computing the observed margins, PROC MIXED uses all observations for which there are no missing or invalid independent variables, including those for which there are missing dependent variables. Also, if *OM-data-set* has a WEIGHT variable, PROC MIXED uses weighted margins to construct the LS-means coefficients. If *OM-data-set* is balanced, the LS-means are unchanged by the OM option.

The BYLEVEL option modifies the observed-margins LS-means. Instead of computing the margins across all of *OM-data-set*, PROC MIXED computes separate margins for each level of the LSMEANS effect in question. The resulting LS-means are actually equal to raw means in this case, but their estimated standard errors account for the covariance structure that you have specified. If the AT option is specified, the BYLEVEL option disables it.

You can use the E option in conjunction with either the OM or BYLEVEL option to check that the modified LS-means coefficients are the ones you desire. It is possible that the modified LS-means are not estimable when the standard ones are, or vice versa. Nonestimable LS-means are noted as "Non-est" in the output.

PDIFF

is the same as the DIFF option.

SINGULAR=*number*
tunes the estimability checking as documented on the "CONTRAST Statement" section.

SLICE= *fixed-effect*
SLICE= *(fixed-effects)*
specifies effects by which to partition interaction LSMEANS effects. This can produce what are known as tests of simple effects (Winer 1971). For example, suppose that A*B is significant, and you want to test the effect of A for each level of B. The appropriate LSMEANS statement is

```
lsmeans A*B / slice=B;
```

This code tests for the simple main effects of A for B, which are calculated by extracting the appropriate rows from the coefficient matrix for the A*B LS-means and using them to form an *F*-test. See the "Inference and Test Statistics" section for more information on this *F*-test.

The SLICE option produces a table titled "Tests of Effect Slices." For ODS purposes, the table name is "Slices."

MODEL Statement

MODEL *dependent = < fixed-effects >< / options >;*
The MODEL statement names a single dependent variable and the fixed effects, which determine the **X** matrix of the mixed model (see the "Parameterization of Mixed Models" section for details). The specification of effects is the same as in the GLM procedure; however, unlike PROC GLM, you do not specify random effects in the MODEL statement. The MODEL statement is required.

An intercept is included in the fixed-effects model by default. If no fixed effects are specified, only this intercept term is fit. The intercept can be removed by using the NOINT option.

You can specify the following options in the MODEL statement after a slash (/).

ALPHA=*number*
requests that a *t*-type confidence interval be constructed for each of the fixed-effects parameters with confidence level 1-number. The value of *number* must be between 0 and 1; the default is 0.05.

ALPHAP=*number*
requests that a *t*-type confidence interval be constructed for the predicted values with confidence level 1-number. The value of *number* must be between 0 and 1; the default is 0.05.

CHISQ
requests that χ^2-tests be performed for all specified effects in addition to the *F*-tests. Type III tests are the default; you can produce the Type I and Type II tests using the HTYPE= option.

CL
requests that *t*-type confidence limits be constructed for each of the fixed-effects parameter estimates. The confidence level is 0.95 by default; this can be changed with the ALPHA= option.

CONTAIN
has the same effect as the DDFM=CONTAIN option.

CORRB

produces the approximate correlation matrix of the fixed-effects parameter estimates. For ODS purposes, the label for this table is "CorrB."

COVB

produces the approximate variance-covariance matrix of the fixed-effects parameter estimates $\hat{\beta}$. By default, this matrix equals $(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-}$ and results from sweeping $(\mathbf{X}\ \mathbf{y})'\hat{\mathbf{V}}^{-1}(\mathbf{X}\ \mathbf{y})$ on all but its last pivot and removing the *y* border. The EMPIRICAL option in the PROC MIXED statement changes this matrix into "empirical sandwich" form. For ODS purposes, the label for this table is "CovB."

COVBI

produces the inverse of the approximate variance-covariance matrix of the fixed-effects parameter estimates. For ODS purposes, the label for this table is "InvCovB."

DDF=*value-list*

enables you to specify your own denominator degrees of freedom for the fixed effects. The *value-list* specification is a list of numbers or missing values (.) separated by commas. The degrees of freedom should be listed in the order in which the effects appear in the "Tests of Fixed Effects" table. If you want to retain the default degrees of freedom for a particular effect, use a missing value for its location in the list. For example,

```
model Y = A B A*B / ddf=3,.,4.7;
```

assigns 3 denominator degrees of freedom to A and 4.7 to A*B, while those for B remain the same.

DDFM=CONTAIN
DDFM=BETWITHIN
DDFM=RESIDUAL
DDFM=SATTERTH
DDFM=KENWARDROGER

specifies the method for computing the denominator degrees of freedom for the tests of fixed effects resulting from the MODEL, CONTRAST, ESTIMATE, and LSMEANS statements.

The DDFM=CONTAIN option invokes the *containment method* to compute denominator degrees of freedom, and it is the default when you specify a RANDOM statement. The containment method is carried out as follows: Denote the fixed effect in question A, and search the RANDOM effect list for the effects that *syntactically* contain A. For example, the RANDOM effect B(A) contains A, but the RANDOM effect C does not, even if it has the same levels as B(A).

Among the RANDOM effects that contain A, compute their rank contribution to the (**X Z**) matrix. The DDF assigned to A is the smallest of these rank contributions. If no effects are found, the DDF for A is set equal to the residual degrees of freedom, *N* - rank(**X Z**). This choice of DDF matches the tests performed for balanced split-plot designs and should be adequate for moderately unbalanced designs.

Caution: If you have a **Z** matrix with a large number of columns, the overall memory requirements and the computing time after convergence can be substantial for the containment method. If it is too large, you may want to use the DDFM=BETWITHIN option.

The DDFM=BETWITHIN option is the default for REPEATED statement specifications (with no RANDOM statements). It is computed by dividing the residual degrees of freedom into between-subject and within-subject portions. PROC MIXED then checks whether a fixed effect changes within any subject.

If so, it assigns within-subject degrees of freedom to the effect; otherwise, it assigns the between-subject degrees of freedom to the effect (refer to Schluchter and Elashoff 1990). If there are multiple within-subject effects containing classification variables, the within-subject degrees of freedom is partitioned into components corresponding to the subject-by-effect interactions.

One exception to the preceding method is the case when you have specified no RANDOM statements and a REPEATED statement with the TYPE=UN option. In this case, all effects are assigned the between-subject degrees of freedom to provide for better small-sample approximations to the relevant sampling distributions.

The DDFM=RESIDUAL option performs all tests using the residual degrees of freedom, n - rank(**XZ**), where n is the number of observations.

The DDFM=SATTERTH option performs a general Satterthwaite approximation for the denominator degrees of freedom, computed as follows. Let $C = (X'V^{-1}X)^{-}$, where $^{-}$ denotes a generalized inverse, and let θ be the vector of unknown parameters in V. Let $\hat{C}$ and $\hat{\theta}$ be the corresponding estimates.

We first consider the one-dimensional case, and consider l to be a vector defining an estimable linear combination of $\hat{\beta}$ .The Satterthwaite degrees of freedom for the t-statistic

$$t = \frac{l\hat{\beta}}{\sqrt{l\hat{C}l'}}$$

is computed as

$$\nu = \frac{2(l\hat{C}l')^2}{g'Ag}$$

where g is the gradient of $lC\,l'$ with respect to θ ,evaluated at $\hat{\theta}$,and A is the asymptotic variance-covariance matrix of $\hat{\theta}$ obtained from the second derivative matrix of the likelihood equations.

For the multi-dimensional case, let L be an estimable contrast matrix of rank $q > 1$. The Satterthwaite denominator degrees of freedom for the F-statistic

$$F = \frac{\hat{\beta}'L'(L\hat{C}L')^{-1}L\hat{\beta}}{q}$$

is computed by first performing the spectral decomposition $L\hat{C}L' = P'DP$ where P is an orthogonal matrix of eigenvectors and D is a diagonal matrix of eigenvalues, both of dimension $q \times q$. Define l_m to be the mth row of PL, and let

$$\nu_m = \frac{2(D_m)^2}{g_m' A g_m}$$

where D_m is the mth diagonal element of D and g_m is the gradient of $l_m C l_m'$ with respect to θ ,evaluated at $\hat{\theta}$. Then let

$$E = \sum_{m=1}^{q} \frac{\nu_m}{\nu_m - 2} I(\nu_m > 2)$$

where the indicator function eliminates terms for which $\nu_m \leq 2$ .The degrees of freedom for F are then computed as

$$\nu = \frac{2E}{E - q}$$

provided $E > q$; otherwise ν is set to zero.

This method is a generalization of the techniques described in Giesbrecht and Burns (1985), McLean and Sanders (1988), and Fai and Cornelius (1996). The method can also include estimated random effects. In this case, append $\hat{\gamma}$ to $\hat{\beta}$ and change $\hat{C}$ to be the inverse of the coefficient matrix in the mixed model equations. The calculations require extra memory to hold c matrices that are the size of the mixed model equations, where c is the number of covariance parameters. In the notation of Table 41.9, this is approximately $8q(p+g)(p+g)/2$ bytes. Extra computing time is also required to process these matrices. The Satterthwaite method implemented here is intended to produce an accurate F-approximation; however, the results may differ from those produced by PROC GLM. Also, the small sample properties of this approximation have not been extensively investigated for the various models available with PROC MIXED.

The DDFM=KENWARDROGER option performs the degrees-of-freedom calculations detailed by Kenward and Roger (1997). This approximation involves inflating the estimated variance-covariance matrix of the fixed and random effects by the method proposed by Prasad and Rao (1990) and Harville and Jeske (1992); refer also to Kackar and Harville (1984). Satterthwaite-type degrees of freedom are then computed based on this adjustment. By default, the observed information matrix of the covariance parameter estimates is used in the calculations.

This method changes output in the following tables (listed in Table 41.7): Contrast, CorrB, CovB, Diffs, Estimates, InvCovB, LSMeans, MMEq, MMEqSol, Slices, SolutionF, SolutionR, Tests1 -Tests3. The OUTP= and OUTPM= data sets are also affected.

E

requests that Type I, Type II, and Type III **L** matrix coefficients be displayed for all specified effects. For ODS purposes, the labels of the tables are "Coefficients".

E1

requests that Type I **L** matrix coefficients be displayed for all specified effects. For ODS purposes, the label of this table is "Coefficients".

E2

requests that Type II **L** matrix coefficients be displayed for all specified effects. For ODS purposes, the label of this table is "Coefficients".

E3

requests that Type III **L** matrix coefficients be displayed for all specified effects. For ODS purposes, the label of this table is "Coefficients".

FULLX

requests that columns of the **X** matrix that consist entirely of zeros not be eliminated from **X**; they are eliminated by default. For a column corresponding to a missing cell to be added to **X**, its particular levels must be present in at least one observation in the analysis data set along with a missing dependent variable. The use of the FULLX option can impact coefficient specifications in the CONTRAST and ESTIMATE statements, as well as covariate coefficients from LSMEANS statements specified with the AT MEANS option.

HTYPE=*value-list*

indicates the type of hypothesis test to perform on the fixed effects. Valid entries for *value* are 1, 2, and 3; the default value is 3. You can specify several types by separating the values with a comma or a space. The ODS table names are "Tests1" for the Type 1 tests, "Tests2" for the Type 2 tests, and "Tests3" for Type 3 tests.

NOCONTAIN

has the same effect as the DDFM=RESIDUAL option.

NOINT

requests that no intercept be included in the model. An intercept is included by default.

NOTEST

specifies that no hypothesis tests be performed for the fixed effects.

OUTP=*SAS-data-set*
OUTPRED=*SAS-data-set*

specifies an output data set containing predicted values and related quantities. This option replaces the P option from Version 6.

Predicted values are formed by using the rows from (**X Z**) as **L** matrices. The predicted values from the

$$\mathbf{X}\hat{\beta} + \mathbf{Z}\hat{\gamma}$$

original data are, thus, . Their approximate standard errors of prediction are formed from the quadratic form of **L** with $\hat{\mathbf{C}}$ defined in the "Statistical Properties" section. The L95 and U95 variables provide a *t*-type confidence interval for the predicted values, and they correspond to the L95M and U95M variables from the GLM and REG procedures for fixed-effect models. The residuals are the observed minus the predicted values. Predicted values for data points other than those observed can be obtained by using missing dependent variables in your input data set.

Specifications that have a REPEATED statement with the SUBJECT= option and missing dependent variables compute predicted values using empirical best linear unbiased prediction (EBLUP). Using hats $(\hat{\ })$ to denote estimates, the EBLUP formula is

$$\hat{\mathbf{m}} = \mathbf{X}_{\mathbf{m}}\hat{\boldsymbol{\beta}} + \hat{\mathbf{C}}_{\mathbf{m}}\hat{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

where **m** represents a hypothetical realization of a missing data vector with associated design matrix $\mathbf{X}_{\mathbf{m}}$. The matrix $\mathbf{C}_{\mathbf{m}}$ is the model-based covariance matrix between **m** and the observed data **y**, and other notation is as presented in the "Mixed Models Theory" section.

The estimated prediction variance is as follows:

$$\hat{\mathrm{Var}}(\hat{\mathbf{m}} - \mathbf{m}) = \hat{\mathbf{V}}_{\mathbf{m}} - \hat{\mathbf{C}}_{\mathbf{m}}\hat{\mathbf{V}}^{-1}\hat{\mathbf{C}}_{\mathbf{m}}^{T} + [\mathbf{X}_{\mathbf{m}} - \hat{\mathbf{C}}_{\mathbf{m}}\hat{\mathbf{V}}^{-1}\mathbf{X}](\mathbf{X}^{T}\hat{\mathbf{V}}^{-1}\mathbf{X})^{-}[\mathbf{X}_{\mathbf{m}} - \hat{\mathbf{C}}_{\mathbf{m}}\hat{\mathbf{V}}^{-1}\mathbf{X}]^{T}$$

where $\mathbf{V}_{\mathbf{m}}$ is the model-based variance matrix of **m**. For further details, refer to Henderson (1984) and Harville (1990). This feature can be useful for forecasting time series or for computing spatial predictions.

By default, all variables from the input data set are included in the OUTP= data set. You can select a subset of these variables using the ID statement.

OUTPM=*SAS-data-set*
OUTPREDM=*SAS-data-set*

specifies an output data set containing predicted means and related quantities. This option replaces the PM option from Version 6.

The output data set is of the same form as that resulting from the OUTP= option, except that the predicted values do not incorporate the EBLUP values $\mathbf{Z}\hat{\boldsymbol{\gamma}}$ nor do they use the EBLUPs for specifications that have a REPEATED statement with the SUBJECT= option and missing dependent variables. The predicted values are formed as $\mathbf{X}\hat{\boldsymbol{\beta}}$ in the OUTPM= data set, and standard errors are quadratic forms in the approximate variance-covariance matrix of $\hat{\boldsymbol{\beta}}$ as displayed by the COVB option.

By default, all variables from the input data set are included in the OUTPM= data set. You can select a subset of these variables using the ID statement.

SINGULAR=*number*

tunes the sensitivity in sweeping. If a diagonal pivot element is less than D**number* as PROC MIXED sweeps a matrix, the associated column is declared to be linearly dependent upon previous columns, and the associated parameter is set to 0. The value D is the original diagonal element of the matrix. The default is 1E4 times the machine epsilon; this product is approximately 1E-12 on most computers.

SINGCHOL=*number*

tunes the sensitivity in computing Cholesky roots. If a diagonal pivot element is less than D**number* as PROC MIXED performs the Cholesky decomposition on a matrix, the associated column is declared to be linearly dependent upon previous columns and is set to **0**. The value D is the original diagonal element of the matrix. The default for *number* is 1E4 times the machine epsilon; this product is approximately 1E-12 on most computers.

SINGRES=*number*
sets the tolerance for which the residual variance is considered to be zero. The default is 1E4 times the machine epsilon; this product is approximately 1E-12 on most computers.

SOLUTION
S
requests that a solution for the fixed-effects parameters be produced. Using notation from the "Mixed Models Theory" section, the fixed-effects parameter estimates are $\hat{\mathbf{b}}$ and their approximate standard errors are the square roots of the diagonal elements of $(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-}$. You can output this approximate variance matrix with the COVB option or modify it with the EMPIRICAL option in the PROC MIXED statement.

Along with the estimates and their approximate standard errors, a *t*-statistic is computed as the estimate divided by its standard error. The degrees of freedom for this *t*-statistic matches the one appearing in the "Tests of Fixed Effects" table under the effect containing the parameter. The "Pr > |t|" column contains the two-tailed *p*-value corresponding to the *t*-statistic and associated degrees of freedom. You can use the CL option to request confidence intervals for all of the parameters; they are constructed around the estimate by using a radius of the standard error times a percentage point from the *t*-distribution.

XPVIX
is an alias for the COVBI option.

XPVIXI
is an alias for the COVB option.

ZETA=*number*
tunes the sensitivity in forming Type III functions. Any element in the estimable function basis with an absolute value less than *number* is set to 0. The default is 1E-8.

PARMS Statement

PARMS *(value-list) ...< / options >* ***;***

The PARMS statement specifies initial values for the covariance parameters, or it requests a grid search over several values of these parameters. You must specify the values in the order in which they appear in the "Covariance Parameter Estimates" table.

The *value-list* specification can take any of several forms:
m
a single value

$m_1, m_2, \ldots, m_n$
several values

m to *n*
a sequence where *m* equals the starting value, *n* equals the ending value, and the increment equals 1

m to *n* by *i*
a sequence where *m* equals the starting value, *n* equals the ending value, and the increment equals *i*

m_1, m_2 to m_3

mixed values and sequences

You can use the PARMS statement to input known parameters. Referring to the split-plot example (Example 41.1), suppose the three variance components are known to be 60, 20, and 6. The SAS code to fix the variance components at these values is as follows:

```
proc mixed data=sp noprofile;
  class Block A B;
  model Y = A B A*B;
  random Block A*Block;
  parms (60) (20) (6) / noiter;
run;
```

The NOPROFILE option requests PROC MIXED to refrain from profiling the residual variance parameter during its calculations, thereby enabling its value to be held at 6 as specified in the PARMS statement. The NOITER option prevents any Newton-Raphson iterations so that the subsequent results are based on the given variance components. You can also specify known parameters of **G** using the GDATA= option in the RANDOM statement.

If you specify more than one set of initial values, PROC MIXED performs a grid search of the likelihood surface and uses the best point on the grid for subsequent analysis. Specifying a large number of grid points can result in long computing times. The grid search feature is also useful for exploring the likelihood surface. See Example 41.3.

The results from the PARMS statement are the values of the parameters on the specified grid (denoted by CovP1 - CovP *n*), the residual variance (possibly estimated) for models with a residual variance parameter, and various functions of the likelihood.

For ODS purposes, the label of the "Parameter Search" table is "ParmSearch."

You can specify the following options in the PARMS statement after a slash (/).

HOLD=*value-list*
EQCONS=*value-list*
specifies which parameter values PROC MIXED should hold to equal the specified values. For example, the statement

```
parms (5) (3) (2) (3) / hold=1,3;
```

constrains the first and third covariance parameters to equal 5 and 2, respectively.

LOGDETH
evaluates the log determinant of the Hessian matrix for each point specified in the PARMS statement. A Log Det H column is added to the "Parameter Search" table.

LOWERB=*value-list*
enables you to specify lower boundary constraints on the covariance parameters. The *value-list* specification is a list of numbers or missing values (.) separated by commas. You must list the numbers in the order that PROC MIXED uses for the covariance parameters, and each number corresponds to the lower boundary constraint. A missing value instructs PROC MIXED to use its default constraint, and if you do not specify numbers for all of the covariance parameters, PROC MIXED assumes the remaining ones are missing.

An example for which this option is useful is when you want to constrain the **G** matrix to be positive definite in order to avoid the more computationally intensive algorithms required when **G** becomes singular. The corresponding code for a random coefficients model is as follows:

```
proc mixed;
  class person;
  model y = time;
  random int time / type=fa0(2) sub=person;
  parms / lowerb=1e-4,.,1e-4;
run;
```

Here the FA0(2) structure is used in order to specify a Cholesky root parameterization for the 2 ×2 unstructured blocks in **G**. This parameterization ensures that the **G** matrix is nonnegative definite, and the PARMS statement then ensures that it is positive definite by constraining the two diagonal terms to be greater than or equal to 1E-4.

NOBOUND

requests the removal of boundary constraints on covariance parameters. For example, variance components have a default lower boundary constraint of 0, and the NOBOUND option allows their estimates to be negative.

NOITER

requests that no Newton-Raphson iterations be performed and that PROC MIXED use the best value from the grid search to perform inferences. By default, iterations begin at the best value from the PARMS grid search.

NOPROFILE

specifies a different computational method for the residual variance during the grid search. By default, PROC MIXED estimates this parameter using the profile likelihood when appropriate, and this estimate is displayed in the Variance column of the "Parameter Search" table. The NOPROFILE option suppresses the profiling and uses the actual value of the specified variance in the likelihood calculations.

OLS

requests starting values corresponding to the usual general linear model. Specifically, all variances and covariances are set to zero except for the residual variance, which is set equal to its ordinary least-squares (OLS) estimate. This option is useful when the default MIVQUE0 procedure produces poor starting values for the optimization process.

PARMSDATA=*SAS-data-set*
PDATA=*SAS-data-set*

reads in covariance parameter values from a SAS data set.. The data set should contain the EST or COVP1 -COVP*n* variables. (Note: In releases prior to 6.12, the data set should contain EST or COL1 -COL*n* variables.)

RATIOS

indicates that ratios with the residual variance are specified instead of the covariance parameters themselves. The default is to use the individual covariance parameters.

UPPERB=*value-list*

enables you to specify upper boundary constraints on the covariance parameters. The *value-list* specification is a list of numbers or missing values (.) separated by commas. You must list the numbers in the order that PROC MIXED uses for the covariance parameters, and each number corresponds to the upper boundary constraint. A missing value instructs PROC MIXED to use its default constraint, and if you do not specify numbers for all of the covariance parameters, PROC MIXED assumes that the remaining ones are missing.

PRIOR Statement

PRIOR < *distribution* > < */ options* > ***;***

The PRIOR statement enables you to carry out a sampling-based Bayesian analysis in PROC MIXED. It currently operates only with variance component models. The analysis produces a SAS data set containing a pseudo-random sample from the joint posterior density of the variance components and other parameters in the mixed model.

The posterior analysis is performed after all other PROC MIXED computations. It begins with the "Posterior Sampling Information" table, which provides basic information about the posterior sampling analysis, including the prior densities, sampling algorithm, sample size, and random number seed. For ODS purposes, the name of this table is "Posterior."

By default, PROC MIXED uses an independence chain algorithm in order to generate the posterior sample (Tierney 1994). This algorithm works by generating a pseudo-random proposal from a convenient base distribution, chosen to be as close as possible to the posterior. The proposal is then retained in the sample with probability proportional to the ratio of weights constructed by taking the ratio of the true posterior to the base density. If a proposal is not accepted, then a duplicate of the previous observation is added to the chain.

In selecting the base distribution, PROC MIXED makes use of the fact that the fixed-effects parameters can be analytically integrated out of the joint posterior, leaving the marginal posterior density of the variance components. In order to better approximate the marginal posterior density of the variance components, PROC MIXED transforms them using the MIVQUE(0) equations. You can display the selected transformation with the PTRANS option or specify your own with the TDATA= option. The density of the transformed parameters is then approximated by a product of inverted gamma densities (refer to Gelfand et al. 1990).

To determine the parameters for the inverted gamma densities, PROC MIXED evaluates the logarithm of the posterior density over a grid of points in each of the transformed parameters, and you can display the results of this search with the PSEARCH option. PROC MIXED then performs a linear regression of these values on the logarithm of the inverted gamma density. The resulting base densities are displayed in the "Base Densities" table; for ODS purposes, the name of this table is "BaseDen." You can input different base densities with the BDATA= option.

At the end of the sampling, the "Acceptance Rates" table displays the acceptance rate computed as the number of accepted samples divided by the total number of samples generated. For ODS purposes, the label of the "Acceptance Rates" table is "AcceptanceRates."

The OUT= option specifies the output data set containing the posterior sample. PROC MIXED automatically includes all variance component parameters in this data set (labeled COVP1 - COVP*n*), the Type III *F*-statistics constructed as in Ghosh (1992) discussing Schervish (1992) (labeled T3F*n*), the log values of the posterior (labeled LOGF), the log of the base sampling density (labeled LOGG), and the log of their ratio (labeled LOGRATIO). If you specify the SOLUTION option in the MODEL statement, the data set also contains a random sample from the posterior density of the fixed-effects parameters (labeled BETA*n*), and if you specify the SOLUTION option in the RANDOM statement, the table contains a random sample from the posterior density of the random-effects parameters (labeled GAM*n*). PROC MIXED also generates additional variables corresponding to any CONTRAST, ESTIMATE, or LSMEANS statement that you specify.

Subsequently, you can use SAS/INSIGHT, or the UNIVARIATE, CAPABILITY, or KDE procedures to analyze the posterior sample.

The prior density of the variance components is, by default, a noninformative version of Jeffreys' prior (Box and Tiao 1973). You can also specify informative priors with the DATA= option or a flat (equal to 1) prior for the variance components. The prior density of the fixed-effects parameters is assumed to be flat (equal to 1), and the resulting posterior is conditionally multivariate normal (conditioning on the variance component parameters) with mean $(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$ and variance $(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}$.

The *distribution* argument in the PRIOR statement determines the prior density for the variance component parameters of your mixed model. Valid values are as follows.

DATA=
enables you to input the prior densities of the variance components used by the sampling algorithm. This data set must contain the TYPE and PARM1 -PARM*n* variables, where *n* is the largest number of parameters among each of the base densities. The format of the DATA= data set matches that created by PROC MIXED in the "Base Densities" table, so you can output the densities from one run and use them as input for a subsequent run.

JEFFREYS
specifies a noninformative reference version of Jeffreys' prior constructed using the square root of the determinant of the expected information matrix as in (1.3.92) of Box and Tiao (1973). This is the default prior.

FLAT
specifies a prior density equal to 1 everywhere, making the likelihood function the posterior.

You can specify the following options in the PRIOR statement after a slash (/).

ALG=IC | INDCHAIN
ALG=IS | IMPSAMP
ALG=RS | REJSAMP
ALG=RWC | RWCHAIN
specifies the algorithm used for generating the posterior sample. The ALG=IC option requests an independence chain algorithm, and it is the default. The option ALG=IS requests importance sampling, ALG=RS requests rejection sampling, and ALG=RWC requests a random walk chain. For more information on these techniques, refer to Ripley (1987), Smith and Gelfand (1992), and Tierney (1994).

BDATA=
enables you to input the base densities used by the sampling algorithm. This data set must contain the TYPE and PARM1 - PARM*n* variables, where *n* is the largest number of parameters among each of the base densities. The format of the BDATA= data set matches that created by PROC MIXED in the "Base Densities" table, so you can output the densities from one run and use them as input for a subsequent run.

GRID=*(value-list)*
specifies a grid of values over which to evaluate the posterior density. The *value-list* syntax is the same as in the PARMS statement, and you must specify an output data set name with the OUTG= option.

GRIDT=*(value-list)*
specifies a transformed grid of values over which to evaluate the posterior density. The *value-list* syntax is the same as in the PARMS statement, and you must specify an output data set name with the OUTGT= option.

IFACTOR=*number*
is an alias for the SFACTOR= option.

LOGNOTE=*number*
instructs PROC MIXED to write a note to the SAS log after it generates the sample corresponding to each multiple of *number*. This is useful for monitoring the progress of CPU-intensive runs.

LOGRBOUND=*number*
specifies the bounding constant for rejection sampling. The value of *number* equals the maximum of log(*f*/*g*) over the variance component parameter space, where *f* is the posterior density and *g* is the product

inverted gamma densities used to perform rejection sampling.

When performing the rejection sampling, you may encounter the message
WARNING: The log ratio bound of LL was violated at sample XX.

When this occurs, PROC MIXED reruns an optimization algorithm to determine a new log upper bound and then restarts the rejection sampling. The resulting OUT= data set contains all observations that have been generated; therefore, assuming that you have requested N samples, you should retain only the final N observations in this data set for analysis purposes.

NSAMPLE=*number*
specifies the number of posterior samples to generate. The default is 1000, but more accurate results are obtained with larger samples such as 10000.

NSEARCH=*number*
specifies the number of posterior evaluations PROC MIXED makes for each transformed parameter in determining the parameters for the inverted gamma densities. The default is 20.

OUT=*SAS-data-set*
creates an output data set containing the sample from the posterior density.

OUTG=*SAS-data-set*
creates an output data set from the grid evaluations specified in the GRID= option.

OUTGT=*SAS-data-set*
creates an output data set from the transformed grid evaluations specified in the GRIDT= option.

PSEARCH
displays the search used to determine the parameters for the inverted gamma densities. For ODS purposes, the name of the table is "Search."

PTRANS
displays the transformation of the variance components. For ODS purposes, the name of the table is "Trans."

SEED=*number*
specifies a starting value for the random number generation in PROC MIXED. The computer clock time is the default. You should use a seed whenever you want to duplicate the sample in another run of PROC MIXED.

SFACTOR=*number*
enables you to adjust the range over which PROC MIXED searches the transformed parameters in order to determine the parameters for the inverted gamma densities. PROC MIXED determines the range by first transforming the estimates from the standard PROC MIXED analysis (REML, ML, or MIVQUE0, depending upon which estimation method you select). It then multiplies and divides the transformed estimates by 2**number* to obtain upper and lower bounds, respectively. Transformed values that produce negative variance components in the original scale are not included in the search. The default value is 1; *number* must be greater than 0.5.

TDATA=
enables you to input the transformation of the covariance parameters used by the sampling algorithm. This data set should contain the CovP1 -CovP*n* variables. The format of the TDATA= data set matches that created by PROC MIXED in the "Trans" table, so you can output the transformation from one run and use is as input for a subsequent run.

TRANS=EXPECTED
TRANS=MIVQUE0
TRANS=OBSERVED
specifies the particular algorithm used to determine the transformation of the covariance parameters. The default is MIVQUE0, indicating a transformation based on the MIVQUE(0) equations. The other two options indicate the type of Hessian matrix used in constructing the transformation via a Cholesky root.

UPDATE=*number*
is an alias for the LOGNOTE= option.

RANDOM Statement

RANDOM *random-effects* < / *options* > **;**

The RANDOM statement defines the random effects constituting the γ vector in the mixed model. It can be used to specify traditional variance component models (as in the VARCOMP procedure) and to specify random coefficients. The random effects can be classification or continuous, and multiple RANDOM statements are possible.

Using notation from the "Mixed Models Theory" section, the purpose of the RANDOM statement is to define the **Z** matrix of the mixed model, the random effects in the γ vector, and the structure of **G**. The **Z** matrix is constructed exactly as the **X** matrix for the fixed effects, and the **G** matrix is constructed to correspond with the effects constituting **Z**. The structure of **G** is defined by using the TYPE= option.

You can specify INTERCEPT (or INT) as a random effect to indicate the intercept. PROC MIXED does not include the intercept in the RANDOM statement by default as it does in the MODEL statement.

You can specify the following options in the RANDOM statement after a slash (/).

ALPHA=*number*
requests that a *t*-type confidence interval be constructed for each of the random effect estimates with confidence level 1-number. The value of *number* must be between 0 and 1; the default is 0.05.

CL

requests that *t*-type confidence limits be constructed for each of the random effect estimates. The confidence level is 0.95 by default; this can be changed with the ALPHA= option.

G

requests that the estimated **G** matrix be displayed. PROC MIXED displays blanks for values that are 0. If you specify the SUBJECT= option, then the block of the **G** matrix corresponding to the first subject is displayed. For ODS purposes, the name of the table is "G."

GC

displays the lower-triangular Cholesky root of the estimated **G** matrix according to the rules listed under the G option. For ODS purposes, the name of the table is "CholG."

GCI
displays the inverse Cholesky root of the estimated **G** matrix according to the rules listed under the G option. For ODS purposes, the name of the table is "InvCholG."

GCORR

displays the correlation matrix corresponding to the estimated **G** matrix according to the rules listed under the G option. For ODS purposes, the name of the table is "GCorr."

GDATA=*SAS-data-set*

requests that the **G** matrix be read in from a SAS data set. This **G** matrix is assumed to be known; therefore, only **R**-side parameters from effects in the REPEATED statement are included in the Newton-Raphson iterations. If no REPEATED statement is specified, then only a residual variance is estimated.

The information in the GDATA= data set can appear in one of two ways. The first is a sparse representation for which you include ROW, COL, and VALUE variables to indicate the row, column, and value of **G**. All unspecified locations are assumed to be 0. The second representation is for dense matrices. In it you include ROW and COL1 -COL*n* variables to indicate the row and columns of **G**, which is a symmetric matrix of order *n*. For both representations, you must specify effects in the RANDOM statement that generate a **Z** matrix that contains *n* columns. See Example 41.4.

If you have more than one RANDOM statement, only one GDATA= option is required on any one of them, and the data set you specify must contain the entire **G** matrix defined by all of the RANDOM statements.

If the GDATA= data set contains variance ratios instead of the variances themselves, then use the RATIOS option.

Known parameters of **G** can also be input using the PARMS statement with the HOLD= option.

GI

displays the inverse of the estimated **G** matrix according to the rules listed under the G option. For ODS purposes, the name of the table is "InvG."

GROUP=*effect*
GRP=*effect*

defines an effect specifying heterogeneity in the covariance structure of **G**. All observations having the same level of the group effect have the same covariance parameters. Each new level of the group effect produces a new set of covariance parameters with the same structure as the original group. You should exercise caution in defining the group effect, as strange covariance patterns can result with its misuse. Also, the group effect can greatly increase the number of estimated covariance parameters, which may adversely affect the optimization process.

Continuous variables are permitted as arguments to the GROUP= option. PROC MIXED does not sort by the values of the continuous variable; rather, it considers the data to be from a new subject or group whenever the value of the continuous variable changes from the previous observation. Using a continuous variable decreases execution time for models with a large number of subjects or groups and also prevents the production of a large "Class Levels Information" table.

LDATA=*SAS-data-set*

reads the coefficient matrices associated with the TYPE=LIN(*number*) option. The data set must contain the variables PARM, ROW, COL1 -COLn, or PARM, ROW, COL, VALUE. The PARM variable denotes which of the *number* coefficient matrices is currently being constructed, and the ROW, COL1 -COLn, or ROW, COL, VALUE variables specify the matrix values, as they do with the GDATA= option. Unspecified values of these matrices are set equal to 0.

NOFULLZ

eliminates the columns in **Z** corresponding to missing levels of random effects involving CLASS variables. By default, these columns are included in **Z**.

RATIOS

indicates that ratios with the residual variance are specified in the GDATA= data set instead of the covariance parameters themselves. The default GDATA= data set contains the individual covariance parameters.

SOLUTION
S

requests that the solution for the random-effects parameters be produced. Using notation from the "Mixed Models Theory" section, these estimates are the empirical best linear unbiased predictors (EBLUPs) $\hat{\gamma} = \hat{\mathbf{G}}\mathbf{Z}'\hat{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$. They can be useful for comparing the random effects from different experimental units and can also be treated as residuals in performing diagnostics for your mixed model.

The numbers displayed in the SE Pred column of the "Solution for Random Effects" table are not the standard errors of the $\hat{\gamma}$ displayed in the Estimate column; rather, they are the standard errors of predictions $\hat{\gamma}_i - \gamma_i$, where $\hat{\gamma}_i$ is the *i*th EBLUP and γ_i is the *i*th random-effect parameter.

SUBJECT=*effect*
SUB=*effect*

identifies the subjects in your mixed model. Complete independence is assumed across subjects; thus, for the RANDOM statement, the SUBJECT= option produces a block-diagonal structure in **G** with identical blocks. The **Z** matrix is modified to accommodate this block-diagonality. In fact, specifying a subject effect is equivalent to nesting all other effects in the RANDOM statement within the subject effect.

Continuous variables are permitted as arguments to the SUBJECT= option. PROC MIXED does not sort by the values of the continuous variable; rather, it considers the data to be from a new subject or group whenever the value of the continuous variable changes from the previous observation. Using a continuous variable decreases

execution time for models with a large number of subjects or groups and also prevents the production of a large "Class Levels Information" table.

When you specify the SUBJECT= option and a classification random effect, computations are usually much quicker if the levels of the random effect are duplicated within each level of the SUBJECT= effect.

TYPE=*covariance-structure*

specifies the covariance structure of **G**. Valid values for *covariance-structure* and their descriptions are listed in Table 41.3 and Table 41.4. Although a variety of structures are available, most applications call for either TYPE=VC or TYPE=UN. The TYPE=VC (variance components) option is the default structure, and it models a different variance component for each random effect.

The TYPE=UN (unstructured) option is useful for correlated random coefficient models. For example,

```
random intercept age / type=un subject=person;
```

specifies a random intercept-slope model that has different variances for the intercept and slope and a covariance between them. You can also use TYPE=FA0(2) here to request a **G** estimate that is constrained to be nonnegative definite.

If you are constructing your own columns of **Z** with continuous variables, you can use the TYPE=TOEP(1) structure to group them together to have a common variance component. If you desire to have different covariance structures in different parts of **G**, you must use multiple RANDOM statements with different TYPE= options.

V<=*value-list*>
requests that blocks of the estimated **V** matrix be displayed. The first block determined by the SUBJECT= effect is the default displayed block. PROC MIXED displays entries that are 0 as blanks in the table.

You can optionally use the *value-list* specification, which indicates the subjects for which blocks of **V** are to be displayed. For example, the statement

```
random int time / type=un subject=person v=1,3,7;
```

displays block matrices for the first, third, and seventh persons. The table name for ODS purposes is "V".

VC<=*value-list*>
displays the Cholesky root of the blocks of the estimated **V** matrix. The *value-list* specification is the same as in the V= option. The table name for ODS purposes is "CholV".

VCI<=*value-list*>
displays the inverse of the Cholesky root of the blocks of the estimated **V** matrix. The *value-list* specification is the same as in the V= option. The table name for ODS purposes is "InvCholV".

VCORR<=*value-list*>
displays the correlation matrix corresponding to the blocks of the estimated **V** matrix. The *value-list* specification is the same as in the V= option. The table name for ODS purposes is "VCorr".

VI<=*value-list*>
displays the inverse of the blocks of the estimated **V** matrix. The *value-list* specification is the same as in the V= option. The table name for ODS purposes is "InvV".

REPEATED Statement

REPEATED < *repeated-effect* > < / *options* > ***;***

The REPEATED statement is used to specify the **R** matrix in the mixed model. Its syntax is different from that of the REPEATED statement in PROC GLM. If no REPEATED statement is specified, **R** is assumed to be equal to $\sigma^2\mathbf{I}$.

For many repeated measures models, no repeated effect is required in the REPEATED statement. Simply use the SUBJECT= option to define the blocks of **R** and the TYPE= option to define their covariance structure. In this case, the repeated measures data must be similarly ordered for each subject, and you must indicate all missing response variables with periods in the input data set unless they all fall at the end of a subject's repeated response profile. These requirements are necessary in order to inform PROC MIXED of the proper location of the observed repeated responses.

Specifying a repeated effect is useful when you do not want to indicate missing values with periods in the input data set. The repeated effect must contain only classification variables. Make sure that the levels of the repeated effect are different for each observation within a subject; otherwise, PROC MIXED constructs identical rows in **R** corresponding to the observations with the same level. This results in a singular **R** and an infinite likelihood. Whether you specify a REPEATED effect or not, the rows of **R** for each subject are constructed in the order that they appear in the input data set.

You can specify the following options in the REPEATED statement after a slash (/).

GROUP=*effect*
GRP=*effect*
defines an effect specifying heterogeneity in the covariance structure of **R**. All observations having the same level of the GROUP effect have the same covariance parameters. Each new level of the GROUP

effect produces a new set of covariance parameters with the same structure as the original group. You should exercise caution in properly defining the GROUP effect, as strange covariance patterns can result with its misuse. Also, the GROUP effect can greatly increase the number of estimated covariance parameters, which may adversely affect the optimization process.

Continuous variables are permitted as arguments to the GROUP= option. PROC MIXED does not sort by the values of the continuous variable; rather, it considers the data to be from a new subject or group whenever the value of the continuous variable changes from the previous observation. Using a continuous variable decreases execution time for models with a large number of subjects or groups and also prevents the production of a large "Class Levels Information" table.

HLM

produces a table of Hotelling-Lawley-McKeon statistics (McKeon 1974) for all fixed effects whose levels change across data having the same level of the SUBJECT= effect (the *within-subject* fixed effects). This option applies only when you specify a REPEATED statement with the TYPE=UN option and no RANDOM statements. For balanced data, this model is equivalent to the multivariate model for repeated measures in PROC GLM.

The Hotelling-Lawley-McKeon statistic has a slightly better *F* approximation than the Hotelling-Lawley-Pillai-Samson statistic (see the description of the HLPS option, which follows). Both of the Hotelling-Lawley statistics can perform much better in small samples than the default *F* statistic (Wright 1994).

Separate tables are produced for Type I, II, and III tests, according to the ones you select. For ODS purposes, the labels for these tables are "HLM1," "HLM2," and "HLM3," respectively.

HLPS

produces a table of Hotelling-Lawley-Pillai-Samson statistics (Pillai and Samson 1959) for all fixed effects whose levels change across data having the same level of the SUBJECT= effect (the *within-subject* fixed effects). This option applies only when you specify a REPEATED statement with the TYPE=UN option and no RANDOM statements. For balanced data, this model is equivalent to the multivariate model for repeated measures in PROC GLM, and this statistic is the same as the Hotelling-Lawley Trace statistic produced by PROC GLM.

Separate tables are produced for Type I, II, and III tests, according to the ones you select. For ODS purposes, the labels for these tables are "HLPS1," "HLPS2," and "HLPS3," respectively.

LDATA=*SAS-data-set*

reads the coefficient matrices associated with the TYPE=LIN(*number*) option. The data set must contain the variables PARM, ROW, COL1 -COLn, or PARM, ROW, COL, VALUE. The PARM variable denotes which of the *number* coefficient matrices is currently being constructed, and the ROW, COL1 - COLn, or ROW, COL, VALUE variables specify the matrix values, as they do with the RANDOM statement option GDATA=. Unspecified values of these matrices are set equal to 0.

LOCAL
LOCAL=EXP(<*effects*>)
LOCAL=POM(*POM-data-set*)

requests that a diagonal matrix be added to **R**. With just the LOCAL option, this diagonal matrix equals $\sigma^2\mathbf{I}$, and σ^2 becomes an additional variance parameter that PROC MIXED profiles out of the likelihood provided that you do not specify the NOPROFILE option in the PROC MIXED statement. The LOCAL option is useful

if you want to add an observational error to a time series structure (Jones and Boadi-Boateng 1991) or a nugget effect to a spatial structure (Cressie 1991).

The LOCAL=EXP(<*effects*>) option produces exponential local effects, also known as dispersion effects, in a log-linear variance model. These local effects have the form

$$\sigma^2 \mathrm{diag}[\exp(\mathbf{U}\delta)]$$

where **U** is the full-rank design matrix corresponding to the effects that you specify and δ are the parameters that PROC MIXED estimates. An intercept is not included in **U** because it is accounted for by σ^2. PROC MIXED constructs the full-rank **U** in terms of 1s and -1s for classification effects. Be sure to scale continuous effects in **U** sensibly.

The LOCAL=POM(*POM-data-set*) option specifies the power-of-the-mean structure. This structure possesses a variance of the form $\sigma^2 |\mathbf{x}_i'\beta^*|^\theta$ for the *i*th observation, where $\mathbf{x}_i$ is the *i*th row of **X** (the design matrix of the fixed effects), and β^* is an estimate of the fixed-effects parameters that you specify in *POM-data-set*.

The SAS data set specified by *POM-data-set* contains the numeric variable Estimate (in previous releases, the variable name was required to be EST), and it has at least as many observations as there are fixed-effects parameters. The first *p* observations of the Estimate variable in *POM-data-set* are taken to be the elements of β^*, where *p* is the number of columns of **X**. You must order these observations according to the nonfull-rank parameterization of the MIXED procedure. One easy way to set up *POM-data-set* for a β^* corresponding to ordinary least squares is illustrated by the following code:

```
ods output SolutionF=sf;
proc mixed;
  class a;
  model y = a x / s;
run;

proc mixed;
  class a;
  model y = a x;
  repeated / local=pom(sf);
run;
```

Note that the generalized least-squares estimate of the fixed-effects parameters from the second PROC MIXED step usually is not the same as your specified β^*. However, you can iterate the POM fitting until the two estimates agree. Continuing from the previous example, the code for performing one step of this iteration is as follows.

```
ods output SolutionF=sf1;
proc mixed;
  class a;
  model y = a x / s;
  repeated / local=pom(sf);
run;
```

```
proc compare brief data=sf compare=sf1;
  var estimate;
run;

data sf;
  set sf1;
run;
```

Unfortunately, this iterative process does not always converge. For further details, refer to the description of pseudo-likelihood in Chapter 3 of Carroll and Ruppert (1988).

LOCALW
specifies that only the local effects and no others be weighted. By default, all effects are weighted. The LOCALW option is used in connection with the WEIGHT statement and the LOCAL option in the REPEATED statement

NONLOCALW
specifies that only the nonlocal effects and no others be weighted. By default, all effects are weighted. The NONLOCALW option is used in connection with the WEIGHT statement and the LOCAL option in the REPEATED statement

R<=*value-list*>
requests that blocks of the estimated **R** matrix be displayed. The first block determined by the SUBJECT= effect is the default displayed block. PROC MIXED displays blanks for value-lists that are 0.

The *value-list* indicates the subjects for which blocks of **R** are to be displayed. For example,

```
repeated / type=cs subject=person r=1,3,5;
```

displays block matrices for the first, third, and fifth persons. See the "PARMS Statement" section for the possible forms of *value-list*. The table name for ODS purposes is "R".

RC<=*value-list*>
produces the Cholesky root of blocks of the estimated **R** matrix. The *value-list* specification is the same as with the R option. The table name for ODS purposes is "CholR".

RCI<=*value-list*>
produces the inverse Cholesky root of blocks of the estimated **R** matrix. The *value-list* specification is the same as with the R option. The table name for ODS purposes is "InvCholR".

RCORR<=*value-list*>
produces the correlation matrix corresponding to blocks of the estimated **R** matrix. The *value-list* specification is the same as with the R option. The table name for ODS purposes is "RCorr".

RI<=*value-list*>
produces the inverse of blocks of the estimated **R** matrix. The *value-list* specification is the same as with the R option. The table name for ODS purposes is "InvR".

SSCP
requests that an unstructured **R** matrix be estimated from the sum-of-squares-and-crossproducts matrix of the residuals. It applies only when you specify TYPE=UN and have no RANDOM statements. Also, you must have a sufficient number of subjects for the estimate to be positive definite.

This option is useful when the size of the blocks of **R** are large (for example, greater than 10) and you want to use or inspect an unstructured estimate that is much quicker to compute than the default REML estimate. The two estimates will agree for certain balanced data sets when you have a classification fixed effect defined across all time points within a subject.

SUBJECT=*effect*
SUB=*effect*

identifies the subjects in your mixed model. Complete independence is assumed across subjects; therefore, the SUBJECT= option produces a block-diagonal structure in **R** with identical blocks. When the SUBJECT= effect consists entirely of classification variables, the blocks of **R** correspond to observations sharing the same level of that effect. These blocks are sorted according to this effect as well.

Continuous variables are permitted as arguments to the SUBJECT= option. PROC MIXED does not sort by the values of the continuous variable; rather, it considers the data to be from a new subject or group whenever the value of the continuous variable changes from the previous observation. Using a continuous variable decreases execution time for models with a large number of subjects or groups and also prevents the production of a large "Class Levels Information" table.

If you want to model nonzero covariance among all of the observations in your SAS data set, specify SUBJECT=INTERCEPT to treat the data as if they are all from one subject. Be aware though that, in this case, PROC MIXED manipulates an **R** matrix with dimensions equal to the number of observations. If no SUBJECT= effect is specified, then every observation is assumed to be from a different subject and **R** is assumed to be diagonal. For this reason, you usually want to use the SUBJECT= option in the REPEATED statement.

TYPE=*covariance-structure*

specifies the covariance structure of the **R** matrix. The SUBJECT= option defines the blocks of **R**, and the TYPE= option specifies the structure of these blocks. Valid values for *covariance-structure* and their descriptions are provided in Table 41.3 and Table 41.4. The default structure is VC.

Table 41.3: Covariance Structures

Structure	Description	Parms	(*i*,*j*)th element
ANTE(1)	Ante-Dependence	2*t*-1	$\sigma_i\sigma_j\prod_{k=i}^{j-1}\rho_k$
AR(1)	Autoregressive(1)	2	$\sigma^2\rho^{\lvert i-j\rvert}$
ARH(1)	Heterogeneous AR(1)	*t*+1	$\sigma_i\sigma_j\rho^{\lvert i-j\rvert}$
ARMA(1,1)	ARMA(1,1)	3	$\sigma^2[\gamma\rho^{\lvert i-j\rvert-1}1(i\neq j)+1(i=j)]$
CS	Compound Symmetry	2	$\sigma_1^2+\sigma^2 1(i=j)$
CSH	Heterogeneous CS	*t*+1	$\sigma_i\sigma_j[\rho 1(i\neq j)+1(i=j)]$

FA(*q*)	Factor Analytic	[*q*/2](2*t* -*q* + 1) + *t*	$\Sigma_{k=1}^{\min(i,j,q)}\lambda_{ik}\lambda_{jk}+\sigma_i^2 1(i=j)$
FA0(*q*)	No Diagonal FA	[*q*/2](2*t* -*q* + 1)	$\Sigma_{k=1}^{\min(i,j,q)}\lambda_{ik}\lambda_{jk}$
FA1(*q*)	Equal Diagonal FA	[*q*/2](2*t* -*q* + 1) + 1	$\Sigma_{k=1}^{\min(i,j,q)}\lambda_{ik}\lambda_{jk}+\sigma^2 1(i=j)$
HF	Huynh-Feldt	*t*+1	$(\sigma_i^2+\sigma_j^2)/2+\lambda 1(\mathrm{i}\neq \mathrm{j})$
LIN(*q*)	General Linear	*q*	$\Sigma_{k=1}^{q}\theta_k \mathbf{A}_{ij}$
TOEP	Toeplitz	*t*	$\sigma_{\lvert i-j\rvert+1}$
TOEP(*q*)	Banded Toeplitz	*q*	$\sigma_{\lvert i-j\rvert+1}1(\lvert i-j\rvert<q)$
TOEPH	Heterogeneous TOEP	2*t*-1	$\sigma_i\sigma_j\rho_{\lvert i-j\rvert}$
TOEPH(*q*)	Banded Hetero TOEP	*t*+*q*-1	$\sigma_i\sigma_j\rho_{\lvert i-j\rvert}1(\lvert i-j\rvert<q)$
UN	Unstructured	*t*(*t*+1)/2	σ_{ij}
UN(*q*)	Banded	[*q*/2](2*t*-*q*+1)	$\sigma_{ij}1(\lvert i-j\rvert<q)$
UNR	Unstructured Corrs	*t*(*t*+1)/2	$\sigma_i\sigma_j\rho_{\max(i,j)\min(i,j)}$
UNR(*q*)	Banded Correlations	[*q*/2](2*t*-*q*+1)	$\sigma_i\sigma_j\rho_{\max(i,j)\min(i,j)}$
UN@AR(1)	Direct Product AR(1)	$t_1(t_1+1)/2+1$	$\sigma_{i_1j_1}\rho^{\lvert i_2-j_2\rvert}$
UN@CS	Direct Product CS	$t_1(t_1+1)/2+1$	$\sigma_{i_1j_1}(1-\sigma^2 1(i_2\neq j_2)), 0\leq\sigma^2\leq 1$
UN@UN	Direct Product UN	$t_1(t_1+1)/2+$	$\sigma_{1,i_1j_1}\sigma_{2,i_2j_2}$
		$t_2(t_2+1)/2-1$	
VC	Variance Components	*q*	$\sigma_k^2 1(i=j)$ and *i* corresponds to *k*th effect

In Table 41.3, "Parms" is the number of covariance parameters in the structure, *t* is the overall dimension of the covariance matrix, and 1(*A*) equals 1 when *A* is true and 0 otherwise. For example, 1(*i*=*j*) equals 1 when *i*=*j* and 0 otherwise, and 1(|*i*-*j*|<*q*) equals 1 when |*i*-*j*|<*q* and 0 otherwise. For the TOEPH structures, $\rho_0 = 1$, and for the UNR structures, $\rho_{ii} = 1$ for all *i*. For the direct product structures, the subscripts "1" and "2" refer to the first and second structure in the direct product, respectively, and $i_1 = \text{int}((i+t_2-1)/t_2)$, $j_1 = \text{int}((j+t_2-1)/t_2)$, $i_2 = \text{mod}(i-1,t_2)+1$, and $j_2 = \text{mod}(j-1,t_2)+1$.

Table 41.4: Spatial Covariance Structures

Structure	Description	Parms	(*i*,*j*)th element
SP(EXP)(*c-list*)	Exponential	2	$\sigma^2[\exp(-d_{ij}/\theta)]$
SP(EXPA)(*c-list*)	Anisotropic Exponential	2*c* + 1	$\sigma^2 \prod_{k=1}^{c} \exp[-\theta_k d(i,j,k)^{p_k}]$
SP(GAU)(*c-list*)	Gaussian	2	$\sigma^2[\exp(-d_{ij}^2/\rho^2)]$
SP(LIN)(*c-list*)	Linear	2	$\sigma^2(1-\rho d_{ij})\,1(\rho d_{ij} \le 1)$
SP(LINL)(*c-list*)	Linear log	2	$\sigma^2(1-\rho \log(d_{ij}))\,1(\rho \log(d_{ij}) \le 1)$
SP(POW)(*c-list*)	Power	2	$\sigma^2 \rho^{d_{ij}}$
SP(POWA)(*c-list*)	Anisotropic Power	*c*+1	$\sigma^2 \rho_1^{d(i,j,1)} \rho_2^{d(i,j,2)} \ldots \rho_c^{d(i,j,c)}$
SP(SPH)(*c-list*)	Spherical	2	$\sigma^2[1-(\frac{3d_{ij}}{2\rho})+(\frac{d_{ij}^3}{2\rho^3})]1(d_{ij} \le \rho)$

In Table 41.4, *c-list* contains the names of the numeric variables used as coordinates of the location of the observation in space, and d_{ij} is the Euclidean distance between the *i*th and *j*th vectors of these coordinates, which correspond to the *i*th and *j*th observations in the input data set. For SP(POWA) and SP(EXPA), *c* is the number of coordinates, and *d*(*i*,*j*,*k*) is the absolute distance between the *k*th coordinate, *k* = 1, ... ,*c*, of the *i*th and *j*th observations in the input data set.

Table 41.5 lists some examples of the structures in Table 41.3 and Table 41.4.

Table 41.5: Covariance Structure Examples

Description	Structure	Example

Variance Components	VC (default)	$\begin{bmatrix} \sigma_B^2 & 0 & 0 & 0 \\ 0 & \sigma_B^2 & 0 & 0 \\ 0 & 0 & \sigma_{AB}^2 & 0 \\ 0 & 0 & 0 & \sigma_{AB}^2 \end{bmatrix}$
Compound Symmetry	CS	$\begin{bmatrix} \sigma^2 - \sigma_1 & \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma^2 + \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma^2 + \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma_1 & \sigma^2 + \sigma_1 \end{bmatrix}$
Unstructured	UN	$\begin{bmatrix} \sigma_1^2 & \sigma_{21} & \sigma_{31} & \sigma_{41} \\ \sigma_{21} & \sigma_2^2 & \sigma_{32} & \sigma_{42} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{43} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{bmatrix}$
Banded Main Diagonal	UN(1)	$\begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 \\ 0 & 0 & 0 & \sigma_4^2 \end{bmatrix}$
First-Order Autoregressive	AR(1)	$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$
Toeplitz	TOEP	$\begin{bmatrix} \sigma^2 & \sigma_1 & \sigma_2 & \sigma_3 \\ \sigma_1 & \sigma^2 & \sigma_1 & \sigma_2 \\ \sigma_2 & \sigma_1 & \sigma^2 & \sigma_1 \\ \sigma_3 & \sigma_2 & \sigma_1 & \sigma^2 \end{bmatrix}$
Toeplitz with Two Bands	TOEP(2)	$\begin{bmatrix} \sigma^2 & \sigma_1 & 0 & 0 \\ \sigma_1 & \sigma^2 & \sigma_1 & 0 \\ 0 & \sigma_1 & \sigma^2 & \sigma_1 \\ 0 & 0 & \sigma_1 & \sigma^2 \end{bmatrix}$

Spatial Power	SP(POW)(*c*)	$\sigma^2 \begin{bmatrix} 1 & \rho^{d_{12}} & \rho^{d_{13}} & \rho^{d_{14}} \\ \rho^{d_{21}} & 1 & \rho^{d_{23}} & \rho^{d_{24}} \\ \rho^{d_{31}} & \rho^{d_{32}} & 1 & \rho^{d_{34}} \\ \rho^{d_{41}} & \rho^{d_{42}} & \rho^{d_{43}} & 1 \end{bmatrix}$
Heterogeneous AR(1)	ARH(1)	$\begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho & \sigma_1\sigma_3\rho^2 & \sigma_1\sigma_4\rho^3 \\ \sigma_2\sigma_1\rho & \sigma_2^2 & \sigma_2\sigma_3\rho & \sigma_2\sigma_4\rho^2 \\ \sigma_3\sigma_1\rho^2 & \sigma_3\sigma_2\rho & \sigma_3^2 & \sigma_3\sigma_4\rho \\ \sigma_4\sigma_1\rho^3 & \sigma_4\sigma_2\rho & \sigma_4\sigma_3\rho & \sigma_4^2 \end{bmatrix}$
First-Order Autoregressive Moving-Average	ARMA(1,1)	$\sigma^2 \begin{bmatrix} 1 & \gamma & \gamma\rho & \gamma\rho^2 \\ \gamma & 1 & \gamma & \gamma\rho \\ \gamma\rho & \gamma & 1 & \gamma \\ \gamma\rho^2 & \gamma\rho & \gamma & 1 \end{bmatrix}$
Heterogeneous CS	CSH	$\begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho & \sigma_1\sigma_3\rho & \sigma_1\sigma_4\rho \\ \sigma_2\sigma_1\rho & \sigma_2^2 & \sigma_2\sigma_3\rho & \sigma_2\sigma_4\rho \\ \sigma_3\sigma_1\rho & \sigma_3\sigma_2\rho & \sigma_3^2 & \sigma_3\sigma_4\rho \\ \sigma_4\sigma_1\rho & \sigma_4\sigma_2\rho & \sigma_4\sigma_3\rho & \sigma_4^2 \end{bmatrix}$
First-Order Factor Analytic	FA(1)	$\begin{bmatrix} \lambda_1^2+d_1 & \lambda_1\lambda_2 & \lambda_1\lambda_3 & \lambda_1\lambda_4 \\ \lambda_2\lambda_1 & \lambda_2^2+d_2 & \lambda_2\lambda_3 & \lambda_2\lambda_4 \\ \lambda_3\lambda_1 & \lambda_3\lambda_2 & \lambda_3^2+d_3 & \lambda_3\lambda_4 \\ \lambda_4\lambda_1 & \lambda_4\lambda_2 & \lambda_4\lambda_3 & \lambda_4^2+d_4 \end{bmatrix}$
Huynh-Feldt	HF	$\begin{bmatrix} \sigma_1^2 & \frac{\sigma_1^2+\sigma_2^2}{2}-\lambda & \frac{\sigma_1^2+\sigma_3^2}{2}-\lambda \\ \frac{\sigma_2^2+\sigma_1^2}{2}-\lambda & \sigma_2^2 & \frac{\sigma_2^2+\sigma_3^2}{2}-\lambda \\ \frac{\sigma_3^2+\sigma_1^2}{2}-\lambda & \frac{\sigma_3^2+\sigma_2^2}{2}-\lambda & \sigma_3^2 \end{bmatrix}$
First-Order Ante-dependence	ANTE(1)	$\begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_1 & \sigma_1\sigma_3\rho_1\rho_2 \\ \upsilon_2\upsilon_1\mu_1 & \upsilon_2^2 & \upsilon_2\upsilon_3\mu_2 \\ \sigma_3\sigma_1\rho_2\rho_1 & \sigma_3\sigma_2\rho_2 & \sigma_3^2 \end{bmatrix}$

Heterogeneous Toeplitz	TOEPH	$\begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_1 & \sigma_1\sigma_3\rho_2 & \sigma_1\sigma_4\rho_3 \\ \sigma_2\sigma_1\rho_1 & \sigma_2^2 & \sigma_2\sigma_3\rho_1 & \sigma_2\sigma_4\rho_2 \\ \sigma_3\sigma_1\rho_2 & \sigma_3\sigma_2\rho_1 & \sigma_3^2 & \sigma_3\sigma_4\rho_1 \\ \sigma_4\sigma_1\rho_3 & \sigma_4\sigma_2\rho_2 & \sigma_4\sigma_3\rho_1 & \sigma_4^2 \end{bmatrix}$
Unstructured Correlations	UNR	$\begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_{21} & \sigma_1\sigma_3\rho_{31} & \sigma_1\sigma_4\rho_{41} \\ \sigma_2\sigma_1\rho_{21} & \sigma_2^2 & \sigma_2\sigma_3\rho_{32} & \sigma_2\sigma_4\rho_{42} \\ \sigma_3\sigma_1\rho_{31} & \sigma_3\sigma_2\rho_{32} & \sigma_3^2 & \sigma_3\sigma_4\rho_{43} \\ \sigma_4\sigma_1\rho_{41} & \sigma_4\sigma_2\rho_{42} & \sigma_4\sigma_3\rho_{43} & \sigma_4^2 \end{bmatrix}$
Direct Product AR(1)	UN@AR(1)	$\begin{bmatrix} \sigma_1^2 & \sigma_{21} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} \otimes \begin{bmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{bmatrix} =$
		$\begin{bmatrix} \sigma_1^2 & \sigma_1^2\rho & \sigma_1^2\rho^2 & \sigma_{21} & \sigma_{21}\rho & \sigma_{21}\rho^2 \\ \sigma_1^2\rho & \sigma_1^2 & \sigma_1^2\rho & \sigma_{21}\rho & \sigma_{21} & \sigma_{21}\rho \\ \sigma_1^2\rho^2 & \sigma_1^2\rho & \sigma_1^2 & \sigma_{21}\rho^2 & \sigma_{21}\rho & \sigma_{21} \\ \sigma_{21} & \sigma_{21}\rho & \sigma_{21}\rho^2 & \sigma_2^2 & \sigma_2^2\rho & \sigma_2^2\rho^2 \\ \sigma_{21}\rho & \sigma_{21} & \sigma_{21}\rho & \sigma_2^2\rho & \sigma_2^2 & \sigma_2^2\rho \\ \sigma_{21}\rho^2 & \sigma_{21}\rho & \sigma_{21} & \sigma_2^2\rho^2 & \sigma_2^2\rho & \sigma_2^2 \end{bmatrix}$

The following provides some further information about these covariance structures:

TYPE=ANTE(1)
specifies the first-order antedependence structure (refer to Kenward 1987, Patel 1991, and Macchiavelli and Arnold 1994). In Table 41.3, σ_i^2 is the *i*th variance parameter, and ρ_k is the *k*th autocorrelation parameter satisfying $|\rho_k| < 1$.

TYPE=AR(1)
specifies a first-order autoregressive structure. PROC MIXED imposes the constraint $|\rho| < 1$ for stationarity.

TYPE=ARH(1)
specifies a heterogeneous first-order autoregressive structure. As with TYPE=AR(1), PROC MIXED imposes the constraint $|\rho| < 1$ for stationarity.

TYPE=ARMA(1,1)

specifies the first-order autoregressive moving average structure. In Table 41.3, ρ is the autoregressive parameter, γ models a moving average component, and σ^2 is the residual variance. In the notation of Fuller (1976, p. 68), $\rho = \theta_1$ and

$$\gamma = \frac{(1 + b_1\theta_1)(\theta_1 + b_1)}{1 + b_1^2 + 2b_1\theta_1}$$

The example in Table 41.5 and $|b_1| < 1$ imply that

$$b_1 = \frac{\beta - \sqrt{\beta^2 - 4\alpha^2}}{2\alpha}$$

where $\alpha = \gamma - \rho$ and $\beta = 1 + \rho^2 - 2\gamma\rho$.PROC MIXED imposes the constraints $|\rho| < 1$ and $|\gamma| < 1$ for stationarity, although for some values of ρ and γ in this region the resulting covariance matrix is not positive definite. When the estimated value of ρ becomes negative, the computed covariance is multiplied by $\cos(\pi d_{ij})$ to account for the negativity.

TYPE=CS
specifies the compound-symmetry structure, which has constant variance and constant covariance.

TYPE=CSH
specifies the heterogeneous compound-symmetry structure. This structure has a different variance parameter for each diagonal element, and it uses the square roots of these parameters in the off-diagonal entries. In Table 41.3, σ_i^2 is the *i*th variance parameter, and ρ is the correlation parameter satisfying $|\rho| < 1$.

TYPE=FA(*q*)
specifies the factor-analytic structure with *q* factors (Jennrich and Schluchter 1986). This structure is of the form $\mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{D}$, where $\mathbf{\Lambda}$ is a $t \times q$ rectangular matrix and **D** is a $t \times t$ diagonal matrix with *t* different parameters. When $q > 1$, the elements of $\mathbf{\Lambda}$ in its upper right-hand corner (that is, the elements in the *i*th row and *j*th column for $j > i$) are set to zero to fix the rotation of the structure.

TYPE=FA0(*q*)
is similar to the FA(*q*) structure except that no diagonal matrix **D** is included. When $q < t$, that is, when the number of factors is less than the dimension of the matrix, this structure is nonnegative definite but not of full rank. In this situation, you can use it for approximating an unstructured **G** matrix in the RANDOM

statement or for combining with the LOCAL option in the REPEATED statement. When $q = t$, you can use this structure to constrain **G** to be nonnegative definite in the RANDOM statement.

TYPE=FA1(q)
is similar to the FA(q) structure except that all of the elements in **D** are constrained to be equal. This offers a useful and more parsimonious alternative to the full factor-analytic structure.

TYPE=HF
specifies the Huynh-Feldt covariance structure (Huynh and Feldt 1970). This structure is similar to the CSH structure in that it has the same number of parameters and heterogeneity along the main diagonal. However, it constructs the off-diagonal elements by taking arithmetic rather than geometric means.

You can perform a likelihood ratio test of the Huynh-Feldt conditions by running PROC MIXED twice, once with TYPE=HF and once with TYPE=UN, and then subtracting their respective values of -2 times the maximized likelihood.

If PROC MIXED does not converge under your Huynh-Feldt model, you can specify your own starting values with the PARMS statement. The default MIVQUE(0) starting values can sometimes be poor for this structure. A good choice for starting values is often the parameter estimates corresponding to an initial fit using TYPE=CS.

TYPE=LIN(q)
specifies the general linear covariance structure with q parameters (Helms and Edwards 1991). This structure consists of a linear combination of known matrices that are input with the LDATA= option. This structure is very general, and you need to make sure that the variance matrix is positive definite. By default, PROC MIXED sets the initial values of the parameters to 1. You can use the PARMS statement to specify other initial values.

TYPE=SIMPLE
is an alias for TYPE=VC.

TYPE=SP(EXPA)(*c-list*)
specifies the spatial anisotropic exponential structure, where *c-list* is a list of variables indicating the coordinates. This structure has (i,j)th element equal to

$$\sigma^2 \prod_{k=1}^{c} \exp[-\theta_k d(i,j,k)^{p_k}]$$

where c is the number of coordinates and $d(i,j,k)$ is the absolute distance between the kth coordinate (k = 1, ... ,c) of the ith and jth observations in the input data set. There are $2c + 1$ parameters to be estimated: θ_k, p_k (k = 1, ... ,c), and σ^2.

You may want to constrain some of the EXPA parameters to known values. For example, suppose you have three coordinate variables C1, C2, and C3 and you want to constrain the powers p_k to equal 2, as in Sacks et al. (1989). Suppose further that you want to model covariance across the entire input data set and you suspect the θ_k and σ^2 estimates are close to 3, 4, 5, and 1, respectively. Then specify

```
repeated / type=sp(expa)(c1 c2 c3)
  subject=intercept;
parms (3) (4) (5) (2) (2) (2) (1) /
```

hold=4,5,6;

TYPE=SP(POW)(*c-list*)
TYPE=SP(POWA)(*c-list*)

specifies the spatial power structures. When the estimated value of ρ becomes negative, the computed covariance is multiplied by $\cos(\pi d_{ij})$ to account for the negativity.

TYPE=TOEP<(*q*)>
specifies a banded Toeplitz structure. This can be viewed as a moving-average structure with order equal to *q*-1. The TYPE=TOEP option is a full Toeplitz matrix, which can be viewed as an autoregressive structure with order equal to the dimension of the matrix. The specification TYPE=TOEP(1) is the same as $\sigma^2 I$, where *I* is an identity matrix, and it can be useful for specifying the same variance component for several effects.

TYPE=TOEPH<(*q*)>

specifies a heterogeneous banded Toeplitz structure. In Table 41.3, σ_i^2 is the *i*th variance parameter and ρ_j is the *j*th correlation parameter satisfying $|\rho_j| < 1$. If you specify the order parameter *q*, then PROC MIXED estimates only the first *q* bands of the matrix, setting all higher bands equal to 0. The option TOEPH(1) is equivalent to both the UN(1) and UNR(1) options.

TYPE=UN<(*q*)>
specifies a completely general (unstructured) covariance matrix parameterized directly in terms of variances and covariances. The variances are constrained to be nonnegative, and the covariances are unconstrained. This structure is not constrained to be nonnegative definite in order to avoid nonlinear constraints; however, you can use the FA0 structure if you want this constraint to be imposed by a Cholesky factorization. If you specify the order parameter *q*, then PROC MIXED estimates only the first *q* bands of the matrix, setting all higher bands equal to 0.

TYPE=UNR<(*q*)>
specifies a completely general (unstructured) covariance matrix parameterized in terms of variances and correlations. This structure fits the same model as the TYPE=UN(*q*) option but with a different parameterization. The *i*th variance parameter is σ_i^2. The parameter ρ_{jk} is the correlation between the *j*th and *k*th measurements; it satisfies $|\rho_{jk}| < 1$.If you specify the order parameter *r*, then PROC MIXED estimates only the first *q* bands of the matrix, setting all higher bands equal to zero.

TYPE=UN@AR(1)
TYPE=UN@CS
TYPE=UN@UN
specify direct (Kronecker) product structures designed for multivariate repeated measures (refer to Galecki 1994). These structures are constructed by taking the Kronecker product of an unstructured matrix (modeling covariance across the multivariate observations) with an additional covariance matrix (modeling covariance across time or another factor). The upper left value in the second matrix is constrained to equal 1 to identify the model. Refer to *SAS/IML User's Guide, First Edition,* for more details on direct products.

To use these structures in the REPEATED statement, you must specify two distinct REPEATED effects, both of which must be included in the CLASS statement. The first effect indicates the multivariate

observations, and the second identifies the levels of time or some additional factor. Note that the input data set must still be constructed in "univariate" format; that is, all dependent observations are still listed observation-wise in one single variable. Although this construction provides for general modeling possibilities, it forces you to construct variables indicating both dimensions of the Kronecker product.

For example, suppose your observed data consist of heights and weights of several children measured over several successive years. Your input data set should then contain variables similar to the following:

- Y, all of the heights and weights, with a separate observation for each
- Var, indicating whether the measurement is a height or a weight
- Year, indicating the year of measurement
- Child, indicating the child on which the measurement was taken

Your PROC MIXED code for a Kronecker AR(1) structure across years would then be

```
proc mixed;
  class Var Year Child;
  model Y = Var Year Var*Year;
  repeated Var Year / type=un@ar(1)
                subject=Child;
run;
```

You should nearly always want to model different means for the multivariate observations, hence the inclusion of Var in the MODEL statement. The preceding mean model consists of cell means for all combinations of VAR and YEAR.

TYPE=VC
specifies standard variance components and is the default structure for both the RANDOM and REPEATED statements. In the RANDOM statement, a distinct variance component is assigned to each effect. In the REPEATED statement, this structure is usually used only with the GROUP= option to specify a heterogeneous variance model.

Jennrich and Schluchter (1986) provide general information about the use of covariance structures, and Wolfinger (1996) presents details about many of the heterogeneous structures. Marx and Thompson (1987), Cressie (1991), and Zimmerman and Harville (1991) discuss spatial structures.

WEIGHT Statement

WEIGHT *variable* ***;***

If you do not specify a REPEATED statement, the WEIGHT statement operates exactly like the one in PROC GLM. In this case PROC MIXED replaces **X'X** and **Z'Z** with **X'WX** and **Z'WZ**, where **W** is the diagonal weight matrix. If you specify a REPEATED statement, then the WEIGHT statement replaces **R** with **LRL**, where **L** is a diagonal matrix with elements $\mathbf{W}^{-1/2}$. Observations with nonpositive or missing weights are not included in the PROC MIXED analysis.

Mixed Models Theory

This section provides an overview of a likelihood-based approach to general linear mixed models. This approach simplifies and unifies many common statistical analyses, including those involving repeated measures, random effects, and random coefficients. The basic assumption is that the data are linearly related to unobserved multivariate normal random variables. Extensions to nonlinear and nonnormal situations are possible but are not discussed here. Additional theory with examples is provided in Littell et al. (1996) and Verbeke and Molenberghs (1997).

Matrix Notation

Suppose that you observe n data points $y_1, \ldots, y_n$ and that you want to explain them using n values for each of p explanatory variables $x_{11}, \ldots, x_{1p}, x_{21}, \ldots, x_{2p}, \ldots, x_{n1}, \ldots, x_{np}$. The x_{ij} values may be either regression-type continuous variables or dummy variables indicating class membership. The standard linear model for this setup is

$$y_i = \sum_{j=1}^{p} x_{ij}\beta_j + \epsilon_i \qquad i = 1, \ldots, n$$

where $\beta_1, \ldots, \beta_p$ are unknown *fixed-effects* parameters to be estimated and $\epsilon_1, \ldots, \epsilon_n$ are unknown independent and identically distributed normal (Gaussian) random variables with mean 0 and variance σ^2.

The preceding equations can be written simultaneously using vectors and a matrix, as follows:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1p} \\ x_{21} & x_{22} & \ldots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \ldots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

For convenience, simplicity, and extendibility, this entire system is written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where $\mathbf{y}$ denotes the vector of observed y_i's, $\mathbf{X}$ is the known matrix of x_{ij}'s, $\boldsymbol{\beta}$ is the unknown fixed-effects parameter vector, and $\boldsymbol{\epsilon}$ is the unobserved vector of independent and identically distributed Gaussian random errors.

In addition to denoting data, random variables, and explanatory variables in the preceding fashion, the subsequent development makes use of basic matrix operators such as transpose ('), inverse ($^{-1}$), generalized inverse ($^{-}$), determinant ($|\cdot|$), and matrix multiplication. Refer to Searle (1982) for details on these and other matrix techniques.

Formulation of the Mixed Model

The previous general linear model is certainly a useful one (Searle 1971), and it is the one fitted by the GLM procedure. However, many times the distributional assumption about $\boldsymbol{\epsilon}$ is too restrictive. The mixed model extends the general linear model by allowing a more flexible specification of the covariance matrix of $\boldsymbol{\epsilon}$. In other words, it allows for both correlation and heterogeneous variances, although you still assume normality.

The mixed model is written as

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

where everything is the same as in the general linear model except for the addition of the known design matrix, **Z**, and the vector of unknown *random-effects parameters*, $\boldsymbol{\gamma}$. The matrix **Z** can contain either continuous or dummy variables, just like **X**. The name *mixed model* comes from the fact that the model contains both fixed-effects parameters, $\boldsymbol{\beta}$, and random-effects parameters, $\boldsymbol{\gamma}$. Refer to Henderson (1990) and Searle, Casella, and McCulloch (1992) for historical developments of the mixed model.

A key assumption in the foregoing analysis is that $\boldsymbol{\gamma}$ and $\boldsymbol{\epsilon}$ are normally distributed with

$$\mathbf{E}\begin{bmatrix} \boldsymbol{\gamma} \\ \boldsymbol{\epsilon} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}$$

$$\mathbf{Var}\begin{bmatrix} \boldsymbol{\gamma} \\ \boldsymbol{\epsilon} \end{bmatrix} = \begin{bmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix}$$

The variance of **y** is, therefore, **V** = **ZGZ**' + **R**. You can model **V** by setting up the random-effects design matrix **Z** and by specifying covariance structures for **G** and **R**.

Note that this is a general specification of the mixed model, in contrast to many texts and articles that discuss only simple random effects. Simple random effects are a special case of the general specification with **Z** containing dummy variables, **G** containing variance components in a diagonal structure, and $\mathbf{R} = \sigma^2\mathbf{I}_n$, where $\mathbf{I}_n$ denotes the $n \times n$ identity matrix. The general linear model is a further special case with **Z** = **0** and $\mathbf{R} = \sigma^2\mathbf{I}_n$.

The following two examples illustrate the most common formulations of the general linear mixed model.

Example: Growth Curve with Compound Symmetry

Suppose that you have three growth curve measurements for *s* individuals and that you want to fit an overall linear trend in time. Your **X** matrix is as follows:

$$\mathbf{X} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ \vdots & \vdots \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$$

The first column (coded entirely with 1s) fits an intercept, and the second column (coded with times of 1,2,3) fits a slope. Here, $n = 3s$ and $p = 2$.

Suppose further that you want to introduce a common correlation among the observations from a single individual, with correlation being the same for all individuals. One way of setting this up in the general mixed model is to eliminate the **Z** and **G** matrices and let the **R** matrix be block diagonal with blocks corresponding to the individuals and with each block having the *compound-symmetry* structure. This structure has two unknown parameters, one modeling a common covariance and the other a residual variance. The form for **R** would then be as follows:

$$\mathbf{R} = \begin{bmatrix} \sigma_1^2 + \sigma^2 & \sigma_1^2 & \sigma_1^2 & & & \\ \sigma_1^2 & \sigma_1^2 + \sigma^2 & \sigma_1^2 & & & \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 + \sigma^2 & & & \\ & & & \ddots & & \\ & & & & \sigma_1^2 + \sigma^2 & \sigma_1^2 & \sigma_1^2 \\ & & & & \sigma_1^2 & \sigma_1^2 + \sigma^2 & \sigma_1^2 \\ & & & & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 + \sigma^2 \end{bmatrix}$$

where blanks denote zeroes. There are 3*s* rows and columns altogether, and the common correlation is $\sigma_1^2/(\sigma_1^2 + \sigma^2)$.

The PROC MIXED code to fit this model is as follows:

```
proc mixed;
  class indiv;
  model y = time;
  repeated / type=cs subject=indiv;
run;
```

Here, indiv is a classification variable indexing individuals. The MODEL statement fits a straight line for time; the intercept is fit by default just as in PROC GLM. The REPEATED statement models the **R** matrix: TYPE=CS specifies the compound symmetry structure, and SUBJECT=INDIV specifies the blocks of **R**.

An alternative way of specifying the common intra-individual correlation is to let

$$\mathbf{Z} = \begin{bmatrix} \mathbf{1} & & & \\ \mathbf{1} & & & \\ \mathbf{1} & & & \\ & 1 & & \\ & 1 & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \\ & & & 1 \\ & & & 1 \end{bmatrix}$$

$$\mathbf{G} = \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_1^2 & & \\ & & \ddots & \\ & & & \sigma_1^2 \end{bmatrix}$$

$\mathbf{R} = \sigma^2 \mathbf{I}_n$

and . The **Z** matrix has 3*s* rows and *s* columns, and **G** is *s* ×*s*.

You can set up this model in PROC MIXED in two different but equivalent ways:

```
proc mixed;
  class indiv;
  model y = time;
  random indiv;
run;

proc mixed;
  class indiv;
  model y = time;
  random intercept / subject=indiv;
run;
```

Both of these specifications fit the same model as the previous one that used the REPEATED statement; however, the RANDOM specifications constrain the correlation to be positive whereas the REPEATED specification leaves the correlation unconstrained.

Example: Split-Plot Design

The split-plot design involves two experimental treatment factors, A and B, and two different sizes of experimental units to which they are applied (refer to Winer 1971, Snedecor and Cochran 1980, and Milliken and Johnson 1992).

The levels of A are randomly assigned to the larger sized experimental unit, called *whole plots*, whereas the levels of B are assigned to the smaller sized experimental unit, the *subplots*. The subplots are assumed to be nested within the whole plots, so that a whole plot consists of a cluster of subplots and a level of A is applied to the entire cluster.

Such an arrangement is often necessary by nature of the experiment, the classical example being the application of fertilizer to large plots of land and different crop varieties planted in subdivisions of the large plots. For this example, fertilizer is the whole plot factor A and variety is the subplot factor B.

The first example is a split-plot design for which the whole plots are arranged in a randomized block design. The appropriate PROC MIXED code is as follows:

```
proc mixed;
  class a b block;
  model y = a|b;
  random block a*block;
run;
```

Here

$$\mathbf{R} = \sigma^2 \mathbf{I}_{24}$$

and **X**, **Z**, and **G** have the following form:

$$\mathbf{X} = \begin{bmatrix}
1 & 1 & & & 1 & & 1 & & & & & \\
1 & 1 & & & & 1 & & 1 & & & & \\
1 & & 1 & & 1 & & & & 1 & & & \\
1 & & 1 & & & 1 & & & & 1 & & \\
1 & & & 1 & 1 & & & & & & 1 & \\
1 & & & 1 & & 1 & & & & & & 1 \\
\vdots & & \vdots & & \vdots & & & & & \vdots & & \\
1 & 1 & & & 1 & & 1 & & & & & \\
1 & 1 & & & & 1 & & 1 & & & & \\
1 & & 1 & & 1 & & & & 1 & & & \\
1 & & 1 & & & 1 & & & & 1 & & \\
1 & & & 1 & 1 & & & & & & 1 & \\
1 & & & 1 & & 1 & & & & & & 1
\end{bmatrix}$$

$$\mathbf{Z} = \begin{bmatrix}
1 & & & 1 & & & & & & & & & & & & & & & & & \\
1 & & & 1 & & & & & & & & & & & & & & & & & \\
1 & & & & 1 & & & & & & & & & & & & & & & & \\
1 & & & & 1 & & & & & & & & & & & & & & & & \\
1 & & & & & 1 & & & & & & & & & & & & & & & \\
1 & & & & & 1 & & & & & & & & & & & & & & & \\
 & 1 & & & & & 1 & & & & & & & & & & & & & & \\
 & 1 & & & & & 1 & & & & & & & & & & & & & & \\
 & 1 & & & & & & 1 & & & & & & & & & & & & & \\
 & 1 & & & & & & 1 & & & & & & & & & & & & & \\
 & 1 & & & & & & & 1 & & & & & & & & & & & & \\
 & 1 & & & & & & & 1 & & & & & & & & & & & & \\
 & & 1 & & & & & & & 1 & & & & & & & & & & & \\
 & & 1 & & & & & & & 1 & & & & & & & & & & & \\
 & & 1 & & & & & & & & 1 & & & & & & & & & & \\
 & & 1 & & & & & & & & 1 & & & & & & & & & & \\
 & & 1 & & & & & & & & & 1 & & & & & & & & & \\
 & & 1 & & & & & & & & & 1 & & & & & & & & & \\
 & & & 1 & & & & & & & & & 1 & & & & & & & & \\
 & & & 1 & & & & & & & & & 1 & & & & & & & & \\
 & & & 1 & & & & & & & & & & 1 & & & & & & & \\
 & & & 1 & & & & & & & & & & 1 & & & & & & & \\
 & & & 1 & & & & & & & & & & & 1 & & & & & & \\
 & & & 1 & & & & & & & & & & & 1 & & & & & &
\end{bmatrix}$$

$$\mathbf{G} = \begin{bmatrix}
\sigma^2_B & & & & & & & & \\
 & \sigma^2_B & & & & & & & \\
 & & \sigma^2_B & & & & & & \\
 & & & \sigma^2_B & & & & & \\
 & & & & \sigma^2_{AB} & & & & \\
 & & & & & \sigma^2_{AB} & & & \\
 & & & & & & \ddots & & \\
 & & & & & & & \sigma^2_{AB} &
\end{bmatrix}$$

where σ^2_B is the variance component for Block and σ^2_{AB} is the variance component for A*Block. Changing the RANDOM statement to

random int a / subject=block;

fits the same model, but with **Z** and **G** sorted differently.

$$\mathbf{Z} = \begin{bmatrix}
1 & 1 & & & & & & & & & & & & & \\
1 & 1 & & & & & & & & & & & & & \\
1 & & 1 & & & & & & & & & & & & \\
1 & & 1 & & & & & & & & & & & & \\
1 & & & 1 & & & & & & & & & & & \\
1 & & & 1 & & & & & & & & & & & \\
 & & & & 1 & 1 & & & & & & & & & \\
 & & & & 1 & 1 & & & & & & & & & \\
 & & & & 1 & & 1 & & & & & & & & \\
 & & & & 1 & & 1 & & & & & & & & \\
 & & & & 1 & & & 1 & & & & & & & \\
 & & & & 1 & & & 1 & & & & & & & \\
 & & & & & & & & 1 & 1 & & & & & \\
 & & & & & & & & 1 & 1 & & & & & \\
 & & & & & & & & 1 & & 1 & & & & \\
 & & & & & & & & 1 & & 1 & & & & \\
 & & & & & & & & 1 & & & 1 & & & \\
 & & & & & & & & 1 & & & 1 & & & \\
 & & & & & & & & 1 & & & & 1 & & \\
 & & & & & & & & 1 & & & & 1 & &
\end{bmatrix}$$

$$\mathbf{G} = \begin{bmatrix}
\sigma^2_B & & & & & & & & & & & & & & \\
 & \sigma^2_{AB} & & & & & & & & & & & & & \\
 & & \sigma^2_{AB} & & & & & & & & & & & & \\
 & & & \sigma^2_{AB} & & & & & & & & & & & \\
 & & & & \ddots & & & & & & & & & & \\
 & & & & & \sigma^2_B & & & & & & & & & \\
 & & & & & & \sigma^2_{AB} & & & & & & & & \\
 & & & & & & & \sigma^2_{AB} & & & & & & & \\
 & & & & & & & & \sigma^2_{AB} & & & & & & \\
 & & & & & & & & & \sigma^2_{AB} & & & & & \\
 & & & & & & & & & & \sigma^2_{AB} & & & & \\
 & & & & & & & & & & & \sigma^2_{AB} & & & \\
 & & & & & & & & & & & & \sigma^2_{AB} & & \\
 & & & & & & & & & & & & & \sigma^2_{AB} &
\end{bmatrix}$$

Estimating G and R in the Mixed Model

Estimation is more difficult in the mixed model than in the general linear model. Not only do you have β as in the general linear model, but you have unknown parameters in γ, **G**, and **R** as well. Least squares is no longer the best method. *Generalized least squares* (GLS) is more appropriate, minimizing

$$(\mathbf{y} - \mathbf{X}\beta)'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\beta)$$

However, it requires knowledge of **V** and, therefore, knowledge of **G** and **R**. Lacking such information, one approach is to use *estimated* GLS, in which you insert some reasonable estimate for **V** into the minimization problem. The goal thus becomes finding a reasonable estimate of **G** and **R**.

In many situations, the best approach is to use *likelihood-based* methods, exploiting the assumption that γ and ϵ are normally distributed (Hartley and Rao 1967; Patterson and Thompson 1971; Harville 1977; Laird and Ware 1982; Jennrich and Schluchter 1986). PROC MIXED implements two likelihood-based methods: *maximum likelihood* (ML) and *restricted/residual maximum likelihood* (REML). A favorable theoretical property of ML and REML is that they accommodate data that are missing at random (Rubin 1976; Little 1995). PROC MIXED constructs an objective function associated with ML or REML and maximizes it over all unknown parameters. Using calculus, it is possible to reduce this maximization problem to one over only the parameters in **G** and **R**. The corresponding log-likelihood functions are as follows:

$$\begin{aligned}
\textbf{ML:} \quad l(\mathbf{G},\mathbf{R}) &= -\frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}\mathbf{r}'\mathbf{V}^{-1}\mathbf{r} - \frac{n}{2}\log(2\pi) \\
\textbf{REML:} \quad l_R(\mathbf{G},\mathbf{R}) &= -\frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}\log|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} \\
&\quad -\frac{1}{2}\mathbf{r}'\mathbf{V}^{-}\mathbf{r} - \frac{n-p}{2}\log(2\pi)\}
\end{aligned}$$

where $\mathbf{r} = \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$ and p is the rank of **X**. PROC MIXED actually minimizes -2 times these functions using a ridge-stabilized Newton-Raphson algorithm. Lindstrom and Bates (1988) provide reasons for preferring Newton-Raphson to the Expectation-Maximum (EM) algorithm described in Dempster, Laird, and Rubin (1977) and Laird, Lange, and Stram (1987), as well as analytical details for implementing a QR-decomposition approach to the problem. Wolfinger, Tobias, and Sall (1994) present the sweep-based algorithms that are implemented in PROC MIXED.

One advantage of using the Newton-Raphson algorithm is that the second derivative matrix of the objective function evaluated at the optima is available upon completion. Denoting this matrix **H**, the asymptotic theory of maximum likelihood (refer to Serfling 1980) shows that $2\mathbf{H}^{-1}$ is an asymptotic variance-covariance matrix of the estimated parameters of **G** and **R**. Thus, tests and confidence intervals based on asymptotic normality can be obtained. However, these can be unreliable in small samples, especially for parameters such as variance components which have sampling distributions that tend to be skewed to the right.

If a residual variance σ^2 is a part of your mixed model, it can usually be *profiled* out of the likelihood. This means solving analytically for the optimal σ^2 and plugging this expression back into the likelihood formula (refer to Wolfinger, Tobias, and Sall 1994). This reduces the number of optimization parameters by one and can improve convergence properties. PROC MIXED profiles the residual variance out of the log likelihood whenever it appears reasonable to do so. This includes the case when **R** equals $\sigma^2\mathbf{I}$ and when it has blocks with a compound symmetry, time series, or spatial structure. PROC MIXED does not profile the log likelihood when **R** has unstructured

blocks, when you use the HOLD= or NOITER option in the PARMS statement, or when you use the NOPROFILE option in the PROC MIXED statement.

Instead of ML or REML, you can use the noniterative MIVQUE0 method to estimate **G** and **R** (Rao 1972; LaMotte 1973; Wolfinger, Tobias, and Sall 1994). In fact, by default PROC MIXED uses MIVQUE0 estimates as starting values for the ML and REML procedures. For variance component models, another estimation method involves equating Type I, II, or III expected mean squares to their observed values and solving the resulting system. However, Swallow and Monahan (1984) present simulation evidence favoring REML and ML over MIVQUE0 and other method-of-moment estimators.

Estimating β and γ in the Mixed Model

ML, REML, MIVQUE0, or Type1 -Type3 provide estimates of **G** and **R**, which are denoted $\hat{\mathbf{G}}$ and $\hat{\mathbf{R}}$,respectively. To obtain estimates of β and γ, the standard method is to solve the *mixed model equations* (Henderson 1984):

$$\begin{bmatrix} \mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{X} & \mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{Z} \\ \mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{X} & \mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{Z}+\hat{\mathbf{G}}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\gamma}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{y} \\ \mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{y} \end{bmatrix}$$

The solutions can also be written as

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-}\mathbf{X}'\hat{\mathbf{V}}^{-}\mathbf{y} \\ \hat{\boldsymbol{\gamma}} &= \hat{\mathbf{G}}\mathbf{Z}'\hat{\mathbf{V}}^{-1}(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}) \end{aligned}$$

and have connections with empirical Bayes estimators (Laird and Ware 1982).

Note that the mixed model equations are extended normal equations and that the preceding expression assumes that $\hat{\mathbf{G}}$ is nonsingular. For the extreme case when the eigenvalues of $\hat{\mathbf{G}}$ are very large, $\hat{\mathbf{G}}^{-1}$ contributes very little to the equations and $\hat{\gamma}$ is close to what it would be if γ actually contained fixed-effects parameters. On the other hand,

when the eigenvalues of $\hat{\mathbf{G}}$ are very small, $\hat{\mathbf{G}}^{-1}$ dominates the equations and $\hat{\gamma}$ is close to 0. For intermediate cases, $\hat{\mathbf{G}}^{-1}$ can be viewed as shrinking the fixed-effects estimates of γ towards 0 (Robinson 1991).

If $\hat{\mathbf{G}}$ is singular, then the mixed model equations are modified (Henderson 1984) as follows:

$$\begin{bmatrix} \mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{X} & \mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{Z}\hat{\mathbf{L}} \\ \hat{\mathbf{L}}'\mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{X} & \hat{\mathbf{L}}'\mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{Z}\hat{\mathbf{L}}+\mathbf{I} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\tau} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{y} \\ \hat{\mathbf{L}}'\mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{y} \end{bmatrix}$$

where $\hat{\mathbf{L}}$ is the lower-triangular Cholesky root of $\hat{\mathbf{G}}$, satisfying $\hat{\mathbf{G}} = \hat{\mathbf{L}}\hat{\mathbf{L}}'$. Both $\hat{\tau}$ and a generalized inverse of the left-hand-side coefficient matrix are then transformed using $\hat{\mathbf{L}}$ to determine $\hat{\gamma}$.

An example of when the singular form of the equations is necessary is when a variance component estimate falls on the boundary constraint of 0.

Model Selection

The previous section on estimation assumes the specification of a mixed model in terms of **X**, **Z**, **G**, and **R**. Even though **X** and **Z** have known elements, their specific form and construction is flexible, and several possibilities may present themselves for a particular data set. Likewise, several different covariance structures for **G** and **R** might be reasonable.

Space does not permit a thorough discussion of model selection, but a few brief comments and references are in order. First, subject matter considerations and objectives are of great importance when selecting a model; refer to Diggle (1988) and Lindsey (1993).

Second, when the data themselves are looked to for guidance, many of the graphical methods and diagnostics appropriate for the general linear model extend to the mixed model setting as well (Christensen, Pearson, and Johnson 1992).

Finally, a likelihood-based approach to the mixed model provides several statistical measures for model adequacy as well. The most common of these are the likelihood ratio test and Akaike's and Schwarz's criteria (Bozdogan 1987; Wolfinger 1993).

Statistical Properties

If **G** and **R** are known, $\hat{\beta}$ is the *best linear unbiased estimator* (BLUE) of β, and $\hat{\gamma}$ is the *best linear unbiased predictor* (BLUP) of γ (Searle 1971; Harville 1988, 1990; Robinson 1991; McLean, Sanders, and Stroup 1991). Here, "best" means minimum mean squared error. The covariance matrix of $(\hat{\beta}-\beta, \hat{\gamma}-\gamma)$ is

$$\mathbf{C} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix}^{-}$$

where $^{-}$ denotes a generalized inverse (refer to Searle 1971).

However, **G** and **R** are usually unknown and are estimated using one of the aforementioned methods. These estimates, $\hat{\mathbf{G}}$ and $\hat{\mathbf{R}}$, are therefore simply substituted into the preceding expression to obtain

$$\hat{\mathbf{C}} = \begin{bmatrix} \mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{X} & \mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{Z} \\ \mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{X} & \mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{Z} + \hat{\mathbf{G}}^{-1} \end{bmatrix}^{-}$$

$$(\hat{\beta} - \beta, \hat{\gamma} - \gamma$$

as the approximate variance-covariance matrix of). In this case, the BLUE and BLUP acronyms no longer apply, but the word *empirical* is often added to indicate such an approximation. The appropriate acronyms thus become EBLUE and EBLUP. McLean and Sanders (1988) show that $\hat{\mathbf{C}}$ can also be written as

$$\hat{\mathbf{C}} = \begin{bmatrix} \hat{\mathbf{C}}_{11} & \hat{\mathbf{C}}_{21}' \\ \hat{\mathbf{C}}_{21} & \hat{\mathbf{C}}_{22} \end{bmatrix}$$

where

$$\begin{aligned} \hat{\mathbf{C}}_{1:} &= (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-} \\ \hat{\mathbf{C}}_{2:} &= -\hat{\mathbf{G}}\mathbf{Z}'\hat{\mathbf{V}}^{-1}\mathbf{X}\hat{\mathbf{C}}_{11} \\ \hat{\mathbf{C}}_{22} &= (\mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{Z} + \hat{\mathbf{G}}^{-1})^{-1} - \hat{\mathbf{C}}_{21}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{Z}\hat{\mathbf{G}} \end{aligned}$$

$\hat{\mathbf{C}}_{11}$ $\hat{\beta}$

Note that is the familiar estimated generalized least-squares formula for the variance-covariance matrix of .

$\hat{\beta}\ \hat{\gamma}$ $\hat{\mathbf{C}}$

As a cautionary note, tends to underestimate the true sampling variability of ( ) because no account is made for the uncertainty in estimating **G** and **R**. Although inflation factors have been proposed (Kackar and Harville 1984; Kass and Steffey 1989; Prasad and Rao 1990), they tend to be small for data sets that are fairly well balanced. PROC MIXED does not compute any inflation factors by default, but rather accounts for the downward bias by using the approximate *t* and *F* statistics described subsequently. The DDFM=KENWARDROGER option in the

MODEL statement prompts PROC MIXED to compute a specific inflation factor along with Satterthwaite-based degrees of freedom.

Inference and Test Statistics

For inferences concerning the covariance parameters in your model, you can use likelihood-based statistics. One common likelihood-based statistic is the *Wald Z*, which is computed as the parameter estimate divided by its asymptotic standard error. The asymptotic standard errors are computed from the inverse of the second derivative matrix of the likelihood with respect to each of the covariance parameters. The Wald *Z* is valid for large samples, but it can be unreliable for small data sets and for parameters such as variance components, which are known to have a skewed or bounded sampling distribution.

A better alternative is the likelihood ratio χ^2. This statistic compares two covariance models, one a special case of the other. To compute it, you must run PROC MIXED twice, once for each of the two models, and then subtract the corresponding values of -2 times the log likelihoods. You can use either ML or REML to construct this statistic, which tests whether the full model is necessary beyond the reduced model.

As long as the reduced model does not occur on the boundary of the covariance parameter space, the χ^2 statistic computed in this fashion has a large-sample sampling distribution that is χ^2 with degrees of freedom equal to the difference in the number of covariance parameters between the two models. If the reduced model does occur on the boundary of the covariance parameter space, the asymptotic distribution becomes a mixture of χ^2 distributions (Self and Liang 1987). A common example of this is when you are testing that a variance component equals its lower boundary constraint of 0.

A final possibility for obtaining inferences concerning the covariance parameters is to simulate or resample data from your model and construct empirical sampling distributions of the parameters. The SAS macro language and the ODS system are useful tools in this regard.

For inferences concerning the fixed- and random-effects parameters in the mixed model, consider estimable linear combinations of the following form:

$$\mathbf{L}\begin{bmatrix}\beta\\ \gamma\end{bmatrix}$$

The estimability requirement (Searle 1971) applies only to the β-portion of **L**, as any linear combination of γ is estimable. Such a formulation in terms of a general **L** matrix encompasses a wide variety of common inferential procedures such as those employed with Type I -Type III tests and LS-means. The CONTRAST and ESTIMATE statements in PROC MIXED enable you to specify your own **L** matrices. Typically, inference on fixed-effects is the focus, and, in this case, the γ-portion of **L** is assumed to contain all 0s.

Statistical inferences are obtained by testing the hypothesis

$$H : \mathbf{L}\begin{bmatrix}\beta\\ \gamma\end{bmatrix} = 0$$

or by constructing point and interval estimates.

When **L** consists of a single row, a general *t*-statistic can be constructed as follows (refer to McLean and Sanders 1988, Stroup 1989a):

$$t = \frac{\mathbf{L}\begin{bmatrix}\hat{\beta}\\ \hat{\gamma}\end{bmatrix}}{\sqrt{\mathbf{L}\hat{\mathbf{C}}\mathbf{L}'}}$$

Under the assumed normality of γ and ϵ, *t* has an exact *t*-distribution only for data exhibiting certain types of balance and for some special unbalanced cases. In general, *t* is only approximately *t*-distributed, and its degrees of freedom must be estimated. See the DDFM= option for a description of the various degrees-of-freedom methods available in PROC MIXED. With $\hat{\nu}$ being the approximate degrees of freedom, the associated confidence interval is

$$\mathbf{L}\begin{bmatrix}\hat{\beta}\\ \hat{\gamma}\end{bmatrix} \pm t_{\hat{\nu},\alpha/2}\sqrt{\mathbf{L}\hat{\mathbf{C}}\mathbf{L}'}$$

where $t_{\hat{\nu},\alpha/2}$ is the $(1-\alpha/2)100$ th percentile of the $t_{\hat{\nu}}$ -distribution.

When the rank of **L** is greater than 1, PROC MIXED constructs the following general *F*-statistic:

$$F = \frac{\begin{bmatrix}\hat{\beta}\\ \hat{\gamma}\end{bmatrix}'\mathbf{L}'(\mathbf{L}'\hat{\mathbf{C}}\mathbf{L})^{-1}\mathbf{L}\begin{bmatrix}\hat{\beta}\\ \hat{\gamma}\end{bmatrix}}{\mathrm{rank}(\mathbf{L})}$$

Analogous to *t*, *F* in general has an approximate *F*-distribution with rank(**L**) numerator degrees of freedom and $\hat{\nu}$ denominator degrees of freedom.

The *t*- and *F*-statistics enable you to make inferences about your fixed effects, which account for the variance-covariance model you select. An alternative is the χ^2 statistic associated with the likelihood ratio test. This statistic compares two fixed-effects models, one a special case of the other. It is computed just as when comparing different

covariance models, although you should use ML and not REML here because the penalty term associated with restricted likelihoods depends upon the fixed-effects specification.

Parameterization of Mixed Models

Recall that a mixed model is of the form

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

where **y** represents univariate data, β is an unknown vector of fixed effects with known model matrix **X**, γ is an unknown vector of random effects with known model matrix **Z**, and ϵ is an unknown random error vector.

PROC MIXED constructs a mixed model according to the specifications in the MODEL, RANDOM, and REPEATED statements. Each effect in the MODEL statement generates one or more columns in the model matrix **X**, and each effect in the RANDOM statement generates one or more columns in the model matrix **Z**. Effects in the REPEATED statement do not generate model matrices; they serve only to index observations within subjects. This section shows precisely how PROC MIXED builds **X** and **Z**.

Intercept

By default, all models automatically include a column of 1s in **X** to estimate a fixed-effect intercept parameter μ. You can use the NOINT option in the MODEL statement to suppress this intercept. The NOINT option is useful when you are specifying a classification effect in the MODEL statement and you want the parameter estimate to be in terms of the mean response for each level of that effect, rather than in terms of a deviation from an overall mean.

By contrast, the intercept is not included by default in **Z**. To obtain a column of 1s in **Z**, you must specify in the RANDOM statement either the INTERCEPT effect or some effect that has only one level.

Regression Effects

Numeric variables, or polynomial terms involving them, may be included in the model as regression effects (covariates). The actual values of such terms are included as columns of the model matrices **X** and **Z**. You can use the bar operator with a regression effect to generate polynomial effects. For instance, X|X|X expands to X X*X X*X*X, a cubic model.

Main Effects

If a class variable has *m* levels, PROC MIXED generates *m* columns in the model matrix for its main effect. Each column is an indicator variable for a given level. The order of the columns is the sort order of the values of their levels and can be controlled with the ORDER= option in the PROC MIXED statement. The following table is an example.

Data		I	A		B		
A	B	μ	A1	A2	B1	B2	B3
1	1	1	1	0	1	0	0

1	2	1	1	0	0	1	0
1	3	1	1	0	0	0	1
2	1	1	0	1	1	0	0
2	2	1	0	1	0	1	0
2	3	1	0	1	0	0	1

Typically, there are more columns for these effects than there are degrees of freedom for them. In other words, PROC MIXED uses an over-parameterized model.

Interaction Effects

Often a model includes interaction (crossed) effects. With an interaction, PROC MIXED first reorders the terms to correspond to the order of the variables in the CLASS statement. Thus, B*A becomes A*B if A precedes B in the CLASS statement. Then, PROC MIXED generates columns for all combinations of levels that occur in the data. The order of the columns is such that the rightmost variables in the cross index faster than the leftmost variables. Empty columns (that would contain all 0s) are not generated for **X**, but they are for **Z**.

Data		**I**	**A**		**B**			**A*B**					
A	B	μ	A1	A2	B1	B2	B3	A1B1	A1B2	A1B3	A2B1	A2B2	A2B3
1	1	1	1	0	1	0	0	1	0	0	0	0	0
1	2	1	1	0	0	1	0	0	1	0	0	0	0
1	3	1	1	0	0	0	1	0	0	1	0	0	0
2	1	1	0	1	1	0	0	0	0	0	1	0	0
2	2	1	0	1	0	1	0	0	0	0	0	1	0
2	3	1	0	1	0	0	1	0	0	0	0	0	1

In the preceding matrix, main-effects columns are not linearly independent of crossed-effect columns; in fact, the column space for the crossed effects contains the space of the main effect.

When your model contains many interaction effects, you may be able to code them more parsimoniously using the bar operator (|). The bar operator generates all possible interaction effects. For example, A|B|C expands to A B A*B C A*C B*C A*B*C. To eliminate higher-order interaction effects, use the at sign (@) in conjunction with the bar operator. For instance, A|B|C|D@2 expands to A B A*B C A*C B*C D A*D B*D C*D.

Nested Effects

Nested effects are generated in the same manner as crossed effects. Hence, the design columns generated by the following two statements are the same (but the ordering of the columns is different):

model Y=A B(A);

model Y=A A*B;

The nesting operator in PROC MIXED is more a notational convenience than an operation distinct from crossing. Nested effects are typically characterized by the property that the nested variables never appear as main effects. The

order of the variables within nesting parentheses is made to correspond to the order of these variables in the CLASS statement. The order of the columns is such that variables outside the parentheses index faster than those inside the parentheses, and the rightmost nested variables index faster than the leftmost variables.

Data		I	A		B(A)					
A	B	μ	A1	A2	B1A1	B2A1	B3A1	B1A2	B2A2	B3A2
1	1	1	1	0	1	0	0	0	0	0
1	2	1	1	0	0	1	0	0	0	0
1	3	1	1	0	0	0	1	0	0	0
2	1	1	0	1	0	0	0	1	0	0
2	2	1	0	1	0	0	0	0	1	0
2	3	1	0	1	0	0	0	0	0	1

Note that nested effects are often distinguished from interaction effects by the implied randomization structure of the design. That is, they usually indicate random effects within a fixed-effects framework. The fact that random effects can be modeled directly in the RANDOM statement may make the specification of nested effects in the MODEL statement unnecessary.

Continuous-Nesting-Class Effects

When a continuous variable nests with a class variable, the design columns are constructed by multiplying the continuous values into the design columns for the class effect.

Data		I	A		X(A)	
X	A	μ	A1	A2	X(A1)	X(A2)
21	1	1	1	0	21	0
24	1	1	1	0	24	0
22	1	1	1	0	22	0
28	2	1	0	1	0	28
19	2	1	0	1	0	19
23	2	1	0	1	0	23

This model estimates a separate slope for X within each level of A.

Continuous-by-Class Effects

Continuous-by-class effects generate the same design columns as continuous-nesting-class effects. The two models are made different by the presence of the continuous variable as a regressor by itself, as well as a contributor to a compound effect.

Data	I	X	A	X*A

X	A	μ	X	A1	A2	X*A1	X*A2
21	1	1	21	1	0	21	0
24	1	1	24	1	0	24	0
22	1	1	22	1	0	22	0
28	2	1	28	0	1	0	28
19	2	1	19	0	1	0	19
23	2	1	23	0	1	0	23

You can use continuous-by-class effects to test for homogeneity of slopes.

General Effects

An example that combines all the effects is X1*X2*A*B*C(D E). The continuous list comes first, followed by the crossed list, followed by the nested list in parentheses. You should be aware of the sequencing of parameters when you use the CONTRAST or ESTIMATE statements to compute some function of the parameter estimates.

Effects may be renamed by PROC MIXED to correspond to ordering rules. For example, B*A(E D) may be renamed A*B(D E) to satisfy the following:

- Class variables that occur outside parentheses (crossed effects) are sorted in the order in which they appear in the CLASS statement.
- Variables within parentheses (nested effects) are sorted in the order in which they appear in the CLASS statement.

The sequencing of the parameters generated by an effect can be described by which variables have their levels indexed faster:

- Variables in the crossed list index faster than variables in the nested list.
- Within a crossed or nested list, variables to the right index faster than variables to the left.

For example, suppose a model includes four effects - A, B, C, and D -each having two levels, 1 and 2. If the CLASS statement is

class A B C D;

then the order of the parameters for the effect B*A(C D), which is renamed A*B(C D), is

$$\begin{array}{llll} A_1B_1C_1D_1 \rightarrow & A_1B_2C_1D_1 \rightarrow & A_2B_1C_1D_1 \rightarrow & A_2B_2C_1D_1 \rightarrow \\ A_1B_1C_1D_2 \rightarrow & A_1B_2C_1D_2 \rightarrow & A_2B_1C_1D_2 \rightarrow & A_2B_2C_1D_2 \rightarrow \\ A_1B_1C_2D_1 \rightarrow & A_1B_2C_2D_1 \rightarrow & A_2B_1C_2D_1 \rightarrow & A_2B_2C_2D_1 \rightarrow \\ A_1B_1C_2D_2 \rightarrow & A_1B_2C_2D_2 \rightarrow & A_2B_1C_2D_2 \rightarrow & A_2B_2C_2D_2 \end{array}$$

Note that first the crossed effects B and A are sorted in the order in which they appear in the CLASS statement so that A precedes B in the parameter list. Then, for each combination of the nested effects in turn, combinations of A and B appear. The B effect moves fastest because it is rightmost in the cross list. Then A moves next fastest, and D moves next fastest. The C effect is the slowest since it is leftmost in the nested list.

When numeric levels are used, levels are sorted by their character format, which may not correspond to their numeric sort sequence (for example, noninteger levels). Therefore, it is advisable to include a desired format for numeric levels or to use the ORDER=INTERNAL option in the PROC MIXED statement to ensure that levels are sorted by their internal values.

Implications of the Nonfull-Rank Parameterization

For models with fixed-effects involving class variables, there are more design columns in **X** constructed than there are degrees of freedom for the effect. Thus, there are linear dependencies among the columns of **X**. In this event, all of the parameters are not estimable; there is an infinite number of solutions to the mixed model equations. PROC MIXED uses a generalized (g2) inverse to obtain values for the estimates (Searle 1971). The solution values are not displayed unless you specify the SOLUTION option in the MODEL statement. The solution has the characteristic that estimates are 0 whenever the design column for that parameter is a linear combination of previous columns. With this parameterization, hypothesis tests are constructed to test linear functions of the parameters that are estimable.

Some procedures (such as the CATMOD procedure) reparameterize models to full rank using restrictions on the parameters. PROC GLM and PROC MIXED do not reparameterize, making the hypotheses that are commonly tested more understandable. Refer to Goodnight (1978) for additional reasons for not reparameterizing.

Missing Level Combinations

PROC MIXED handles missing level combinations of classification variables similarly to the way PROC GLM does. Both procedures delete fixed-effects parameters corresponding to missing levels in order to preserve estimability. However, PROC MIXED does not delete missing level combinations for random-effects parameters because linear combinations of the random-effects parameters are always estimable. These conventions can affect the way you specify your CONTRAST and ESTIMATE coefficients.

Default Output

The following sections describe the output PROC MIXED produces by default. This output is organized into various tables, and they are discussed in order of appearance.

Model Information

The "Model Information" table describes the model, some of the variables it involves, and the method used in fitting it. It also lists the method (profile, fit, factor, or none) for handling the residual variance in the model. The *profile* method concentrates the residual variance out of the optimization problem, whereas the *fit* method retains it as a parameter in the optimization. The *factor* method keeps the residual fixed, and *none* is displayed when a residual variance is not a part of the model.

The "Model Information" table also has a row labeled Fixed Effects SE Method. This row describes the method used to compute the approximate standard errors for the fixed-effects parameter estimates and related functions of them. The two possibilities for this row are Model-Based, which is the default method, and Empirical, which results from using the EMPIRICAL option in the PROC MIXED statement.

For ODS purposes, the label of the "Model Information" table is "ModelInfo."

Class Level Information

The "Class Level Information" table lists the levels of every variable specified in the CLASS statement. You should check this information to make sure the data are correct. You can adjust the order of the CLASS variable levels with the ORDER= option in the PROC MIXED statement. For ODS purposes, the label of the "Class Level Information" table is "ClassLevels."

Dimensions

The "Dimensions" table lists the sizes of relevant matrices. This table can be useful in determining CPU time and memory requirements. For ODS purposes, the label of the "Dimensions" table is "Dimensions."

Iteration History

The "Iteration History" table describes the optimization of the residual log likelihood or log likelihood. The function to be minimized (the *objective function*) is $-2l$ for ML and $-2l_R$ for REML; the column name of the objective function in the "Iteration History" table is "-2 Log Like" for ML and "-2 Res Log Like" for REML. The minimization is performed using a ridge-stabilized Newton-Raphson algorithm, and the rows of this table describe the iterations that this algorithm takes in order to minimize the objective function.

The Evaluations column of the "Iteration History" table tells how many times the objective function is evaluated during each iteration.

The Criterion column of the "Iteration History" table is, by default, a relative Hessian convergence quantity given by

$$\frac{\mathbf{g}_k' \mathbf{H}_k^{-1} \mathbf{g}_k}{|f_k|}$$

where f_k is the value of the objective function at iteration k, $\mathbf{g}_k$ is the gradient (first derivative) of f_k, and $\mathbf{H}_k$ is the Hessian (second derivative) of f_k. If $\mathbf{H}_k$ is singular, then PROC MIXED uses the following relative quantity:

$$\frac{\mathbf{g}_k' \mathbf{g}_k}{|f_k|}$$

To prevent the division by $|f_k|$, use the ABSOLUTE option in the PROC MIXED statement. To use a relative function or gradient criterion, use the CONVF or CONVG options, respectively. The Hessian criterion is considered superior to function and gradient criteria because it measures orthogonality rather than lack of progress (Bates and Watts 1988). Provided the initial estimate is feasible and the maximum number of iterations is not exceeded, the Newton-Raphson algorithm is considered to have converged when the criterion is less than the tolerance specified with the CONVF, CONVG, or CONVH option in the PROC MIXED statement. The default tolerance is 1E-8. If convergence is not achieved, PROC MIXED displays the estimates of the parameters at the last iteration.

A convergence criterion that is missing indicates that a boundary constraint has been dropped; it is usually not a cause for concern.

If you specify the ITDETAILS option in the PROC MIXED statement, then the covariance parameter estimates at each iteration are included as additional columns in the "Iteration History" table.

For ODS purposes, the label of the "Iteration History" table is "IterHistory."

Covariance Parameter Estimates

The "Covariance Parameter Estimates" table contains the estimates of the parameters in **G** and **R** (see the "Estimating G and R in the Mixed Model" section). Their values are labeled in the "Cov Parm" table along with Subject and Group information if applicable. The estimates are displayed in the Estimate column and are the results of one of the following estimation methods: REML, ML, MIVQUE0, SSCP, Type1, Type2, or Type3.

If you specify the RATIO option in the PROC MIXED statement, the Ratio column is added to the table listing the ratios of each parameter estimate to that of the residual variance.

Requesting the COVTEST option in the PROC MIXED statement produces the Std Error, Z Value, and Pr > |Z| columns. The Std Error column contains the approximate standard errors of the covariance parameter estimates. These are the square roots of the diagonal elements of the observed inverse Fisher information matrix, which equals $2\mathbf{H}^{-1}$, where **H** is the Hessian matrix. The **H** matrix consists of the second derivatives of the objective function with respect to the covariance parameters; refer to Wolfinger, Tobias, and Sall (1994) for formulas. When you use the SCORING= option and PROC MIXED converges without stopping the scoring algorithm, PROC MIXED uses the expected Hessian matrix to compute the covariance matrix instead of the observed Hessian. The observed or expected inverse Fisher information matrix can be viewed as an asymptotic covariance matrix of the estimates.

The Z Value column is the estimate divided by its approximate standard error, and the Pr > |Z| column is the two-tailed area of the standard Gaussian density outside of the *Z*-value. These statistics constitute Wald tests of the covariance parameters, and they are valid only asymptotically.

Caution: Wald tests can be unreliable in small samples.

For ODS purposes, the label of the "Covariance Parameter Estimates" table is "CovParms."

Fitting Information

The "Fitting Information" table provides some statistics about the estimated mixed model. Expressions for the log likelihood are provided in the "Estimating G and R in the Mixed Model" section. If the log likelihood is an extremely large negative number, then PROC MIXED has deemed the estimated **V** matrix to be singular. In this case, all subsequent results should be viewed with caution.

Akaike's Information Criterion (AIC) (Akaike 1974) is computed as

$$\mathbf{AIC} = l(\hat{\theta}) - q$$

where $l(\hat{\theta})$ is the maximized log likelihood or the residual log likelihood, and q is the effective number of covariance parameters (those not estimated to be on a boundary constraint). It can be used to compare models with the same fixed effects but different variance structures; the model having the largest AIC is deemed best.

Schwarz's Bayesian Criterion (BIC) (Schwarz 1978) is computed as

$$\mathrm{BIC} = l(\hat{\theta}) - \frac{1}{2} q \log N^*$$

where N^* is computed as described in the IC option. Again, models with larger BIC are preferred, but note that BIC penalizes models with a greater number of covariance parameters more than AIC does, and the two criteria may not agree as to which covariance model is best.

The IC option in the PROC MIXED statement produces an "Information Criteria" table of these criteria and two others in a variety of different forms.

For ODS purposes, the label of the "Model Fitting Information" table is "FitStatistics."

Null Model Likelihood Ratio Test

If one covariance model is a submodel of another, you can carry out a likelihood ratio test for the significance of the more general model by computing -2 times the difference between their log likelihoods. Then compare this statistic to the χ^2 distribution with degrees of freedom equal to the difference in the number of parameters for the two models.

This test is reported in the "Null Model Likelihood Ratio Test" table to determine whether it is necessary to model the covariance structure of the data at all. The "Chi-Square" value is -2 times the log likelihood from the null model minus -2 times the log likelihood from the fitted model, where the null model is the one with only the fixed effects listed in the MODEL statement and $\mathbf{R} = \sigma^2\mathbf{I}$. This statistic has an asymptotic χ^2-distribution with q-1 degrees of freedom, where q is the effective number of covariance parameters (those not estimated to be on a boundary constraint). The Pr > ChiSq column contains the upper-tail area from this distribution. This p-value can be used to assess the significance of the model fit.

This test is not produced for cases where the null hypothesis lies on the boundary of the parameter space, which is typically for variance component models. This is because the standard asymptotic theory does not apply in this case (Self and Liang 1987, Case 5).

If you specify a PARMS statement, PROC MIXED constructs a likelihood ratio test between the best model from the grid search and the final fitted model and reports the results in the "Parameter Search" table.

For ODS purposes, the label of the "Null Model Likelihood Ratio Test" table is "LRT."

Type 3 Tests of Fixed Effects

The "Type 3 Tests of Fixed Effects" table contains hypothesis tests for the significance of each of the fixed effects, that is, those effects you specify in the MODEL statement. By default, PROC MIXED computes these tests by first constructing a Type III **L** matrix (see Chapter 12, "The Four Types of Estimable Functions,") for each effect. This **L** matrix is then used to compute the following F-statistic:

$$F = \frac{\hat{\beta}'\mathbf{L}'[\mathbf{L}(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-}\mathbf{L}']^{-}\mathbf{L}\hat{\beta}}{\mathrm{rank}(\mathbf{L})}$$

A *p*-value for the test is computed as the tail area beyond this statistic from an *F*-distribution with NDF and DDF degrees of freedom. The numerator degrees of freedom (NDF) is the row rank of **L**, and the denominator degrees of freedom is computed using one of the methods described under the DDFM= option. Small values of the *p*-value (typically less than 0.05 or 0.01) indicate a significant effect. You can use the HTYPE= option in the MODEL statement to obtain tables of Type I (sequential) tests and Type II (adjusted) tests in addition to or instead of the table of Type III (partial) tests.

You can use the CHISQ option in the MODEL statement to obtain Wald χ^2 tests of the fixed effects. These are carried out by using the numerator of the *F*-statistic and comparing it with the χ^2 distribution with NDF degrees of freedom. It is more liberal than the *F*-test because it effectively assumes an infinite denominator degrees of freedom.

For ODS purposes, the label of the "Type 1 Tests of Fixed Effects" through the "Type 3 Tests of Fixed Effects" tables are "Tests1" through "Tests3," respectively.

Changes in Output

The SAS System now features a new Output Delivery System (ODS), replacing the old one used by PROC MIXED in Version 6. The primary changes involve the replacing of the MAKE statement and the _print_ and _disk_ global variables with the new ODS statement. The following table lists some typical conversions you need to make.

Table 41.6: ODS Conversions for PROC MIXED

Version 6 Syntax	Versions 7 and 8 Syntax
make 'covparms' out=cp;	ods output covparms=cp;
make 'covparms' out=cp noprint;	ods listing exclude covparms; ods output covparms=cp;
%global _print_; %let _print_=off;	ods listing close;
%global _print_; %let _print_=on;	ods listing;

Each table created by PROC MIXED has a name associated with it, and you must use this name to reference the table when using ODS statements. These names are listed in Table 41.7.

Table 41.7: ODS Tables Produced in PROC MIXED

Table Name	Description	Required Statement / Option
AccRates	acceptance rates for posterior sampling	PRIOR
AsyCorr	asymptotic correlation matrix of covariance parameters	PROC MIXED ASYCORR
AsyCov	asymptotic covariance matrix of covariance parameters	PROC MIXED ASYCOV

Base	base densities used for posterior sampling	PRIOR
Bound	computed bound for posterior rejection sampling	PRIOR
CholG	Cholesky root of the estimated **G** matrix	RANDOM / GC
CholR	Cholesky root of blocks of the estimated **R** matrix	REPEATED / RC
CholV	Cholesky root of blocks of the estimated **V** matrix	RANDOM / VC
ClassLevels	level information from the CLASS statement	default output
Coef	**L** matrix coefficients	E option on MODEL, CONTRAST, ESTIMATE, or LSMEANS
Contrasts	results from the CONTRAST statements	CONTRAST
ConvergenceStatus	convergence status	default
CorrB	approximate correlation matrix of fixed-effects parameter estimates	MODEL / CORRB
CovB	approximate covariance matrix of fixed-effects parameter estimates	MODEL / COVB
CovParms	estimated covariance parameters	default output
Diffs	differences of LS-means	LSMEANS / DIFF (or PDIFF)
Dimensions	dimensions of the model	default output
Estimates	results from ESTIMATE statements	ESTIMATE
FitStatistics	fit statistics	default
G	estimated **G** matrix	RANDOM / G
GCorr	correlation matrix from the estimated **G** matrix	RANDOM / GCORR
HLM1	Type 1 Hotelling-Lawley-McKeon tests of fixed effects	MODEL / HTYPE=1 and REPEATED / HLM TYPE=UN
HLM2	Type 2 Hotelling-Lawley-McKeon tests of fixed effects	MODEL / HTYPE=2 and REPEATED / HLM TYPE=UN
HLM3	Type 3 Hotelling-Lawley-McKeon tests of fixed effects	REPEATED / HLM TYPE=UN
HLPS1	Type 1 Hotelling-Lawley-Pillai- Samson tests of fixed effects	MODEL / HTYPE=1 and REPEATED / HLPS TYPE=UN
HLPS2	Type 2 Hotelling-Lawley-Pillai- Samson tests of fixed effects	MODEL / HTYPE=1 and REPEATED / HLPS TYPE=UN
HLPS3	Type 3 Hotelling-Lawley-Pillai- Samson tests of fixed effects	REPEATED / HLPS TYPE=UN
InfoCrit	information criteria	PROC MIXED IC

InvCholG	inverse Cholesky root of the estimated **G** matrix	RANDOM / GCI
InvCholR	inverse Cholesky root of blocks of the estimated **R** matrix	REPEATED / RCI
InvCholV	inverse Cholesky root of blocks of the estimated **V** matrix	RANDOM / VCI
InvCovB	inverse of approximate covariance matrix of fixed-effects parameter estimates	MODEL / COVBI
InvG	inverse of the estimated **G** matrix	RANDOM / GI
InvR	inverse of blocks of the estimated **R** matrix	REPEATED / RI
InvV	inverse of blocks of the estimated **V** matrix	RANDOM / VI
IterHistory	iteration history	default output
LRT	likelihood ratio test	default output
LSMeans	LS-means	LSMEANS
MMEq	mixed model equations	PROC MIXED MMEQ
MMEqSol	mixed model equations solution	PROC MIXED MMEQSOL
ModelInfo	model information	default output
ParmSearch	parameter search values	PARMS
Posterior	posterior sampling information	PRIOR
R	blocks of the estimated **R** matrix	REPEATED / R
RCorr	correlation matrix from a blocks of the estimated **R** matrix	REPEATED / RCORR
Search	posterior density search table	PRIOR / PSEARCH
Slices	tests of LS-means slices	LSMEANS / SLICE=
SolutionF	fixed effects solution vector	MODEL / S
SolutionR	random effects solution vector	RANDOM / S
Tests1	Type 1 tests of fixed effects	MODEL / HTYPE=1
Tests2	Type 2 tests of fixed effects	MODEL / HTYPE=2
Tests3	Type 3 tests of fixed effects	default output
Type1	Type 1 analysis of variance	PROC MIXED METHOD=TYPE1
Type2	Type 2 analysis of variance	PROC MIXED METHOD=TYPE2
Type3	Type 3 analysis of variance	PROC MIXED METHOD=TYPE3
Trans	transformation of covariance parameters	PRIOR / PTRANS
V	blocks of the estimated **V** matrix	RANDOM / V

VCorr	correlation matrix from blocks of the estimated **V** matrix	RANDOM / VCORR

In Table 41.7, "Coefficients" refers to multiple tables produced by the E, E1, E2, or E3 options in the MODEL statement and the E option in the CONTRAST, ESTIMATE, and LSMEANS statements. You can create one large data set of these tables with a statement similar to

```
ods output Coefficients=c;
```

Be aware that the number of variables in this data set is determined by the first table created. To create separate data sets, use

```
ods output Coefficients(match_all)=c;
```

Here the resulting data sets are named C1, C2, C3, etc. The same principles apply to data sets created from the "R," "CholR," "InvCholR," "RCorr," "InvR," "V," "CholV," "InvCholV," "VCorr," and "InvV" tables.

In Table 41.7, the following changes have occured from Version 6. The "Predicted," "PredMeans," and "Sample" tables from Version 6 no longer exist and have been replaced by output data sets; see descriptions of the MODEL statement options OUTPRED= and OUTPREDM= and the PRIOR statement option OUT= for more details. The "ML" and "REML" tables from Version 6 have been replaced by the "IterHistory" table. The "Tests," "HLM," and "HLPS" tables from Version 6 have been renamed "Tests3," "HLM3," and "HLPS3."

Table 41.8 lists the variable names associated with the data sets created when you use the ODS OUTPUT option in conjunction with the preceding tables. In Table 41.8, *n* is used to denote a generic number that is dependent upon the particular data set and model you select, and it can assume a different value each time it is used (even within the same table). The phrase *model specific* appears in rows of the affected tables to indicate that columns in these tables depend upon the variables you specify in the model.

Caution: There exists a danger of name collisions with the variables in the *model specific* tables in Table 41.8 and variables in your input data set. You should avoid using input variables with the same names as the variables in these tables.

Table 41.8: Variable Names for the ODS Tables Produced in PROC MIXED

Table Name	Variables
AsyCorr	Row, CovParm, CovP1 -CovP*n*
AsyCov	Row, CovParm, CovP1 -CovP*n*
BaseDen	Type, Parm1 -Parm*n*
Bound	Technique, Converge, Iterations, Evaluations, LogBound, CovP1 -CovP*n*, TCovP1 -TCovP*n*
CholG	*model specific*, Effect, Subject, Sub1 -Sub*n*, Group, Group1 -Group*n*, Row, Col1 -Col*n*
CholR	Index, Row, Col1 -Col*n*
CholV	Index, Row, Col1 -Col*n*
ClassLevels	Class, Levels, Values
Coefficients	*model specific*, LMatrix, Effect, Subject, Sub1 -Sub*n*, Group, Group1 -Group*n*, Row1 -Row*n*

Contrasts	Label, NumDF, DenDF, ChiSquare, FValue, ProbChiSq, ProbF
CorrB	*model specific*, Effect, Row, Col1 -Col*n*
CovB	*model specific*, Effect, Row, Col1 -Col*n*
CovParms	CovParm, Subject, Group, Estimate, StandardError, ZValue, ProbZ, Alpha, Lower, Upper
Diffs	*model specific*, Effect, Margins, ByLevel, AT variables, Diff, StandardError, DF, tValue, Tails, Probt, Adjustment, Adjp, Alpha, Lower, Upper, AdjLow, AdjUpp
Dimensions	Descr, Value
Estimates	Label, Estimate, StandardError, DF, tValue, Tails, Probt, Alpha, Lower, Upper
FitStatistics	Descr, Value
G	*model specific*, Effect, Subject, Sub1 -Sub*n*, Group, Group1 -Group*n*, Row, Col1 -Col*n*
GCorr	*model specific*, Effect, Subject, Sub1 -Sub*n*, Group, Group1 -Group*n*, Row, Col1 -Col*n*
HLM1	Effect, NumDF, DenDF, FValue, ProbF
HLM2	Effect, NumDF, DenDF, FValue, ProbF
HLM3	Effect, NumDF, DenDF, FValue, ProbF
HLPS1	Effect, NumDF, DenDF, FValue, ProbF
HLPS2	Effect, NumDF, DenDF, FValue, ProbF
HLPS3	Effect, NumDF, DenDF, FValue, ProbF
InfoCrit	Better, CovParms, MeanParms, Likelihood, AIC, HQIC, BIC, CAIC
InvCholG	*model specific*, Effect, Subject, Sub1 -Sub*n*, Group, Group1 -Group*n*, Row, Col1 -Col*n*
InvCholR	Index, Row, Col1 -Col*n*
InvCholV	Index, Row, Col1 -Col*n*
InvCovB	*model specific*, Effect, Row, Col1 -Col*n*
InvG	*model specific*, Effect, Subject, Sub1 -Sub*n*, Group, Group1 -Group*n*, Row, Col1 -Col*n*
InvR	Index, Row, Col1 -Col*n*
InvV	Index, Row, Col1 -Col*n*
IterHistory	CovP1 -CovP*n*, Iteration, Evaluations, M2ResLogLike, M2LogLike, Criterion
LRT	DF, ChiSquare, ProbChiSq
LSMeans	*model specific*, Effect, Margins, ByLevel, AT variables, Estimate, StandardError, DF, tValue, Probt, Alpha, Lower, Upper, Cov1 -Cov*n*, Corr1 -Corr*n*
MMEq	*model specific*, Effect, Subject, Sub1 -Sub*n*, Group, Group1 -Group*n*, Row, Col1 -Col*n*
MMEqSol	*model specific*, Effect, Subject, Sub1 -Sub*n*, Group, Group1 -Group*n*, Row, Col1 -Col*n*
ModelInfo	Descr, Value

ParmSearch	CovP1 -CovP*n*, Var, ResLogLike, M2ResLogLike2, LogLike, M2LogLike, LogDetH
Posterior	Descr, Value
R	Index, Row, Col1 -Col*n*
RCorr	Index, Row, Col1 -Col*n*
Search	Parm, TCovP1 -TCovP*n*, Posterior
Slices	*model specific*, Effect, Margins, ByLevel, AT variables, NumDF, DenDF, FValue, ProbF
SolutionF	*model specific*, Effect, Estimate, StandardError, DF, tValue, Probt, Alpha, Lower, Upper
SolutionR	*model specific*, Effect, Subject, Sub1 -Sub*n*, Group, Group1 -Group*n*, Estimate, StdErrPred, DF, tValue, Probt, Alpha, Lower, Upper
Tests1	Effect, NumDF, DenDF, ChiSquare, FValue, ProbChiSq, ProbF
Tests2	Effect, NumDF, DenDF, ChiSquare, FValue, ProbChiSq, ProbF
Tests3	Effect, NumDF, DenDF, ChiSquare, FValue, ProbChiSq, ProbF
Type1	Source, DF, SS, MS, EMS, ErrorTerm, ErrorDF, FValue, ProbF
Type2	Source, DF, SS, MS, EMS, ErrorTerm, ErrorDF, FValue, ProbF
Type3	Source, DF, SS, MS, EMS, ErrorTerm, ErrorDF, FValue, ProbF
Trans	Prior, TCovP, CovP1 -CovP*n*
V	Index, Row, Col1 -Col*n*
VCorr	Index, Row, Col1 -Col*n*

Some of the variables listed in Table 41.8 are created only when you have specified certain options in the relevant PROC MIXED statements.

The following changes have occurred in these variables from Version 6. Nearly all underscores have been removed from variable names in order to be compatible and consistent with other procedures. Some of the variable names have been changed (for example, T has been changed to tValue and PT to Probt) for the same reason. You may have to modify some of your Version 6 code to accommodate these changes.

Computational Issues

Computational Method

In addition to numerous matrix-multiplication routines, PROC MIXED frequently uses the sweep operator (Goodnight 1979) and the Cholesky root (Golub and Van Loan 1989). The routines perform a modified W transformation (Goodnight and Hemmerle 1979) for **G**-side likelihood calculations and a direct method for **R**-side likelihood calculations. For the Type III *F*-tests, PROC MIXED uses the algorithm described in Chapter 30, "The GLM Procedure."

PROC MIXED uses a ridge-stabilized Newton-Raphson algorithm to optimize either a full (ML) or residual (REML) likelihood function. The Newton-Raphson algorithm is preferred to the EM algorithm (Lindstrom and Bates 1988). PROC MIXED profiles the likelihood with respect to the fixed effects and also with respect to the residual variance whenever it appears reasonable to do so. The residual profiling can be avoided by using the

NOPROFILE option of the PROC MIXED statement. PROC MIXED uses the MIVQUE0 method (Rao 1972; Giesbrecht 1989) to compute initial values.

The likelihoods that PROC MIXED optimizes are usually well-defined continuous functions with a single optimum. The Newton-Raphson algorithm typically performs well and finds the optimum in a few iterations. It is a quadratically converging algorithm, meaning that the error of the approximation near the optimum is squared at each iteration. The quadratic convergence property is evident when the convergence criterion drops to zero by factors of ten or more.

Table 41.9: Notation for Order Calculations

Symbol	Number
p	columns of **X**
g	columns of **Z**
N	observations
q	covariance parameters
t	maximum observations per subject
S	subjects

Using the notation from Table 41.9, the following are estimates of the computational speed of the algorithms used in PROC MIXED. For likelihood calculations, the crossproducts matrix construction is of order $N(p+g)^2$ and the sweep operations are of order $(p+g)^3$. The first derivative calculations for parameters in **G** are of order qg^3 for ML and $q(g^3 + pg^2 + p^2g)$ for REML. If you specify a subject effect in the RANDOM statement and if you are not using the REPEATED statement, then replace g by g/S and q by qS in these calculations. The first derivative calculations for parameters in **R** are of order $qS(t^3 + gt^2 + g^2t)$ for ML and $qS(t^3 + (p+g)t^2 + (p^2 + g^2)t)$ for REML. For the second derivatives, replace q by $q(q+1)/2$ in the first derivative expressions. When you specify both **G**- and **R**-side parameters (that is, when you use both the RANDOM and REPEATED statements), then additional calculations are required of an order equal to the sum of the orders for **G** and **R**. Considerable execution times may result in this case.

For further details about the computational techniques used in PROC MIXED, refer to Wolfinger, Tobias, and Sall (1994).

Parameter Constraints

By default, some covariance parameters are assumed to satisfy certain boundary constraints during the Newton-Raphson algorithm. For example, variance components are constrained to be nonnegative and autoregressive parameters are constrained to be between -1 and 1. You can remove these constraints with the NOBOUND option in the PARMS statement, but this may lead to estimates that produce an infinite likelihood. You can also introduce or change boundary constraints with the LOWERB= and UPPERB= options in the PARMS statement.

During the Newton-Raphson algorithm, a parameter may be set equal to one of its boundary constraints for a few iterations and then it may move away from the boundary. You see a missing value in the Criterion column of the "Iteration History" table whenever a boundary constraint is dropped.

For some data sets the final estimate of a parameter may equal one of its boundary constraints. This is usually not a cause for concern, but it may lead you to consider a different model. For instance, a variance component estimate can equal zero; in this case, you may want to drop the corresponding random effect from the model. However, be aware that changing the model in this fashion can impact degrees of freedom calculations.

Convergence Problems

For some data sets, the Newton-Raphson algorithm can fail to converge. Non-convergence can result from a number of causes, including flat or ridged likelihood surfaces and ill-conditioned data.

It is also possible for PROC MIXED to converge to a point that is not the global optimum of the likelihood, although this usually occurs only with the spatial covariance structures. If you experience convergence problems, the following points may be helpful:

- One useful tool is the PARMS statement, which lets you input initial values for the covariance parameters and performs a grid search over the likelihood surface.
- Sometimes the Newton-Raphson algorithm does not perform well when two of the covariance parameters are on a different scale; that is, they are several orders of magnitude apart. This is because the Hessian matrix is processed jointly for the two parameters, and elements of it corresponding to one of the parameters can become close to internal tolerances in PROC MIXED. In this case, you can improve stability by rescaling the effects in the model so that the covariance parameters are on the same scale.
- Data that is extremely large or extremely small can adversely affect results because of the internal tolerances in PROC MIXED. Rescaling it can improve stability.
- For stubborn problems, you may want to specify ODS OUTPUT COVPARMS= data-set-name to output the "CovParms" table as a precautionary measure. That way, if the problem does not converge, you can read the final parameter values back into a new run with the PARMSDATA= option in the PARMS statement.
- Fisher scoring can be more robust than Newton-Raphson to poor MIVQUE(0) starting values. Specifying a SCORING= value of 5 or so may help to recover from poor starting values.
- Tuning the singularity options SINGULAR=, SINGCHOL=, and SINGRES= in the MODEL statement may improve the stability of the optimization process.
- Tuning the MAXITER= and MAXFUNC= options in the PROC MIXED statement can save resources. Also, the ITDETAILS option displays the values of all of the parameters at each iteration.
- Using the NOPROFILE and NOBOUND options in the PROC MIXED statement may help convergence, although they can produce unusual results.
- Although the CONVH convergence criterion usually gives the best results, you may want to try CONVF or CONVG, possibly along with the ABSOLUTE option.
- If the convergence criterion bottoms out at a relatively small value such as 1E-7 but never gets less than 1E-8, you may want to specify CONVH=1E-6 in the PROC MIXED statement to get results; however, interpret the results with caution.
- An infinite likelihood during the iteration process means that the Newton-Raphson algorithm has stepped into a region where either the **R** or **V** matrix is nonpositive definite. This is usually no cause for concern as long as iterations continue. If PROC MIXED stops because of an infinite likelihood, recheck your model to make sure that no observations from the same subject are producing identical rows in **R** or **V** and that you have enough data to estimate the particular covariance structure you have selected. Any time that the final estimated likelihood is infinite, subsequent results should be interpreted with caution.
- A nonpositive definite Hessian matrix can indicate a surface saddlepoint or linear dependencies among the parameters.
- A warning message about the singularities of **X** changing indicates that there is some linear dependency in the estimate of $\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X}$ that is not found in $\mathbf{X}'\mathbf{X}$. This can adversely affect the likelihood calculations and optimization process. If you encounter this problem, make sure that your model specification is reasonable and that you have enough data to estimate the particular covariance structure you have selected. Rearranging effects in the MODEL statement so that the most significant ones are first can help because PROC MIXED sweeps the estimate of $\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}$ in the order of the MODEL effects and the sweep is more stable if larger pivots are dealt with first. If this does not help, specifying starting values with the PARMS statement can place the optimization on a different and possibly more stable path.
- Lack of convergence may indicate model misspecification or a violation of the normality assumption.

Memory

Let p be the number of columns in **X**, and let g be the number of columns in **Z**. For large models, most of the memory resources are required for holding symmetric matrices of order p, g, and $p + g$. The approximate memory requirement in bytes is

$$40(p^2 + g^2) + 32(p+g)^2$$

If you have a large model that exceeds the memory capacity of your computer, see the suggestions listed under "Computing Time."

Computing Time

PROC MIXED is computationally intensive, and execution times can be long. In addition to the CPU time used in collecting sums and cross products and in solving the mixed model equations (as in PROC GLM), considerable CPU time is often required to compute the likelihood function and its derivatives. These latter computations are performed for every Newton-Raphson iteration.

If you have a model that takes too long to run, the following suggestions may be helpful:

- Examine the "Model Information" table to find out the number of columns in the **X** and **Z** matrices. A large number of columns in either matrix can greatly increase computing time. You may want to eliminate some higher order effects if they are too large.
- If you have a **Z** matrix with a lot of columns, use the DDFM=BW option in the MODEL statement to eliminate the time required for the containment method.
- If possible, "factor out" a common effect from the effects in the RANDOM statement and make it the SUBJECT= effect. This creates a block-diagonal **G** matrix and can often speed calculations.
- If possible, use the same or nested SUBJECT= effects in all RANDOM and REPEATED statements.
- If your data set is very large, you may want to analyze it in pieces. The BY statement can help implement this strategy.
- In general, specify random effects with a lot of levels in the REPEATED statement and those with a few levels in the RANDOM statement.
- The METHOD=MIVQUE0 option runs faster than either the METHOD=REML or METHOD=ML option because it is noniterative.
- You can specify known values for the covariance parameters using the HOLD= or NOITER option in the PARMS statement or the GDATA= option in the RANDOM statement. This eliminates the need for iteration.
- The LOGNOTE option in the PROC MIXED statement writes periodic messages to the SAS log concerning the status of the calculations. It can help you diagnose where the slow down is occurring.