Chapter 15
Mixed Models

Chapter Table of Contents

Chapter 15
Mixed Models

Introduction

The Mixed Models task provides facilities for fitting a number of basic mixed models. These models enable you to handle both fixed effects and random effects in a linear model for a continuous response. Numerous experimental designs produce data for which mixed models are appropriate, including split-plot experiments, multilocation trials, and hierarchical designs.

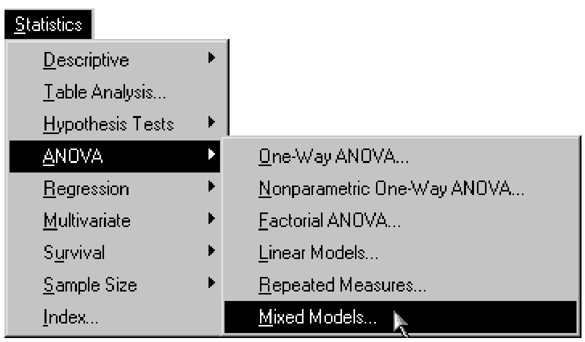

Figure 15.1. Mixed Models Menu

A standard linear model is designed to handle *fixed effects*, in which the levels of the factor represent all possible levels for that factor or at least all levels about which inference is to be made. Factor effects are *random effects* if the levels of the factor in a study or experiment are randomly selected from a population of possible levels of that factor. The population of possible levels of a random effect has a probability distribution with a mean and a variance. By modeling both fixed and random effects, the mixed model provides you with the flexibility of modeling not only means (as in the standard linear model) but variances and covariances as well.

The mixed model is written

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

where $\mathbf{y}$ denotes the vector of observed values, $\mathbf{X}$ is the known fixed effects design matrix, and β is the unknown fixed effects parameter vector. $\mathbf{Z}\gamma$ represents the additional random component of the mixed model. Here, $\mathbf{Z}$ is the known random effects design matrix and γ is a vector of unknown random-effects parameters. $\mathbf{Z}$ contains indicator variables constructed from the random effects, just as $\mathbf{X}$ contains variables constructed for fixed effects. Finally, ϵ is the unobserved vector of independent and identically distributed Gaussian random errors.

Assume that γ and ϵ are Gaussian random variables that are uncorrelated and have expectations 0 and variances $\mathbf{G}$ and $\mathbf{R}$, respectively.

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

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

The variance of $\mathbf{y}$ is therefore $\mathbf{V} = \mathbf{ZGZ}' + \mathbf{R}$.

Note that this is a general specification of the mixed model. The Mixed Models task enables you to specify classification random effects that are a special case of the general specification. You can specify that $\mathbf{Z}$ contains dummy variables, $\mathbf{G}$ contains 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 Mixed Models task enables you to specify a mixed model that incorporates fixed effects and random classification effects and includes interactions and nested terms. You can select from six estimation methods, including maximum likelihood, restricted maximum likelihood (REML), and MIVQUE. You can also compute least-squares means, produce Type 1, 2, and 3 tests for fixed effects, and output predicted values and means to a SAS data set. Plots include means plots for fixed effects, predicted plots, and residual plots.

The examples in this chapter demonstrate how you can use the Mixed Models task in the Analyst Application to analyze linear models data that contain fixed and random effects.

Split Plot Experiment

One of the most common mixed models is the split-plot design. The split-plot design involves two experimental factors, A and B. Levels of A are randomly assigned to whole plots (main plots), and levels of B are randomly assigned to split plots (subplots) within each whole plot. 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. The design provides more precise information about B than about A, and it often arises when A can be applied only to large experimental units.

The hypothetical data set analyzed in this example was created as a balanced split-plot design with the whole plots arranged in a randomized complete-block design (Stroup 1989). The response variable Y represents crop growth measurements. The variable A is a whole plot factor that represents irrigation levels for large plots, and the subplot variable B represents different crop varieties planted in each large plot. The levels of B are randomly assigned to split plots (subplots) within each whole plot. The data set Split contains the whole plot factor A, split plot factor B, response Y, and blocking factor Block. Using the Mixed Models task, you can estimate variance components for Block, A*Block, and the residual and automatically incorporate correct error terms into the tests for fixed effects.

Open the Split Data Set

These data are provided as the Split data set in the Analyst Sample Library. To open the Split data set, follow these steps:

1. Select **Tools** → **Sample Data** . . .
2. Select Split.
3. Click **OK** to create the sample data set in your Sasuser directory.
4. Select **File** → **Open By SAS Name** . . .
5. Select Sasuser from the list of **Libraries**.
6. Select Split from the list of members.
7. Click **OK** to bring the Split data set into the data table.

Request the Mixed Models Analysis

To specify the split plot analysis, follow these steps:

1. Select **Statistics** → **ANOVA** → **Mixed Models** . . .
2. Select Y as the dependent variable.
3. Select A, B, and Block as classification variables.

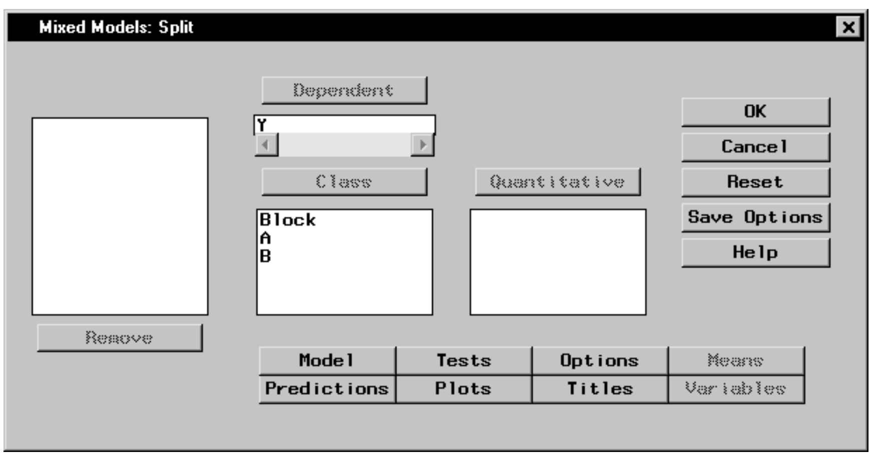

Figure 15.2. Mixed Models Dialog

Figure 15.2 displays the dialog with Y specified as the dependent variable and A, B, and Block specified as classification effects in the mixed model.

Specify the Mixed Model

You can define fixed and random effects, create nested terms, and specify interactions in the Model dialog. The Analyst Application adds terms to the **Fixed effects** list or the **Random effects** list depending on whether the check box at the top of each list is checked. Check the appropriate box for each term you add. Only classification variables can be specified as random effects, and once a term has been specified as a random effect, all higher-order interactions that include that effect must also be specified as random effects.

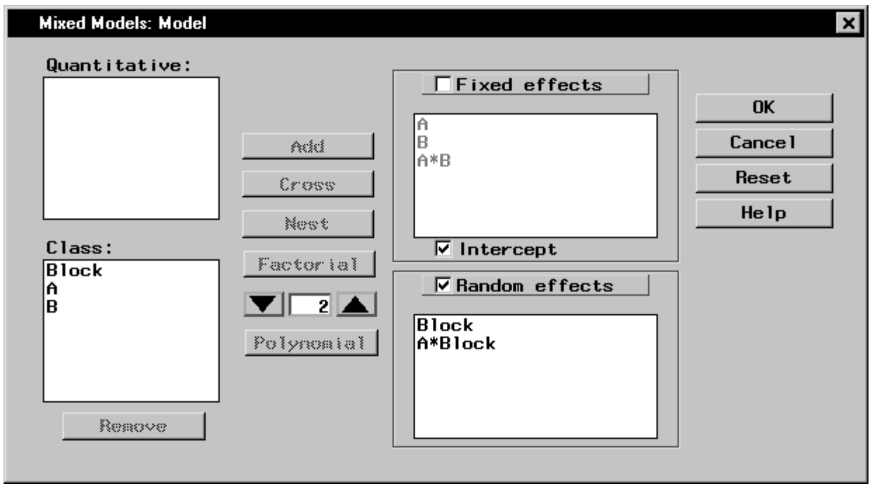

Figure 15.3. Mixed Models: Model Dialog

To specify the mixed model, follow these steps:

1. Click **Model** in the main dialog.
2. Ensure that the **Fixed effects** check box is selected.
3. Select A and B and click **Factorial**.
4. Select the **Random effects** check box, and then select Block and click **Add**.
5. Select Block and A and click **Cross**.

These selections create a factorial structure that contains the A and B main effects and the A*B interaction as fixed effects, and Block and A*Block as random effects. Since you specified the random effects, the columns of the model matrix **Z** now consist of indicator variables corresponding to the levels of Block and A*Block. The **G** matrix is diagonal and contains the variance components of Block and A*Block; the **R** matrix is also diagonal and contains residual variance.

Produce Least-Squares Means

You can request generalized least-squares means of fixed effects using the Means dialog. The least-squares means are estimators of the class or subclass marginal means that are expected for a balanced design. Each least-squares mean is computed as $\mathbf{L}\hat{\beta}$, where $\mathbf{L}$ is the coefficient matrix associated with the least-squares mean and $\hat{\beta}$ is the estimate of the fixed-effects parameter vector. Least-squares means can be computed for any fixed effect that is composed of only classification variables.

For this analysis, interest lies in comparing response means across combinations of the levels of A and B. To request least-squares means of the A*B interaction, follow these steps:

1. Click **Means** in the main dialog.
2. Select A*B in the candidate list and click **LS Mean**.

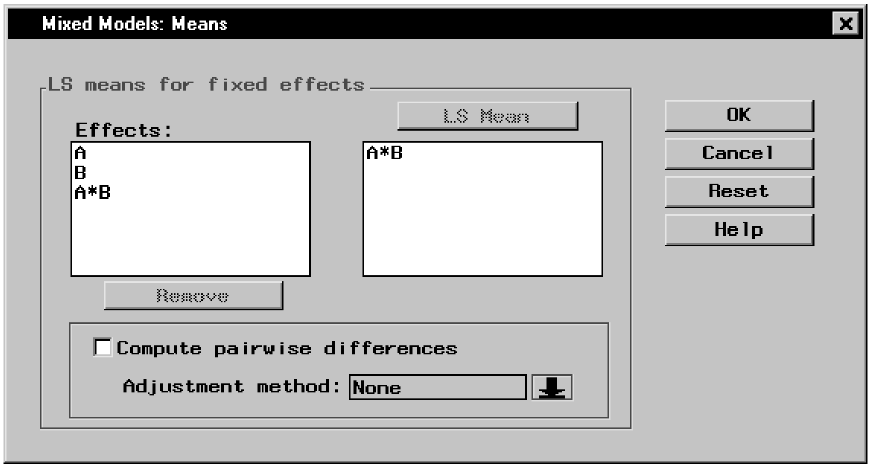

Figure 15.4. Mixed Models: Means Dialog

When you have completed your selections, click **OK** in the main dialog to perform the analysis.

Review the Results

The results are presented in the project tree under the **Mixed Models** folder, as displayed in Figure 15.5. The two nodes represent the mixed models results and the SAS programming statements (labeled **Code**) that generate the output.

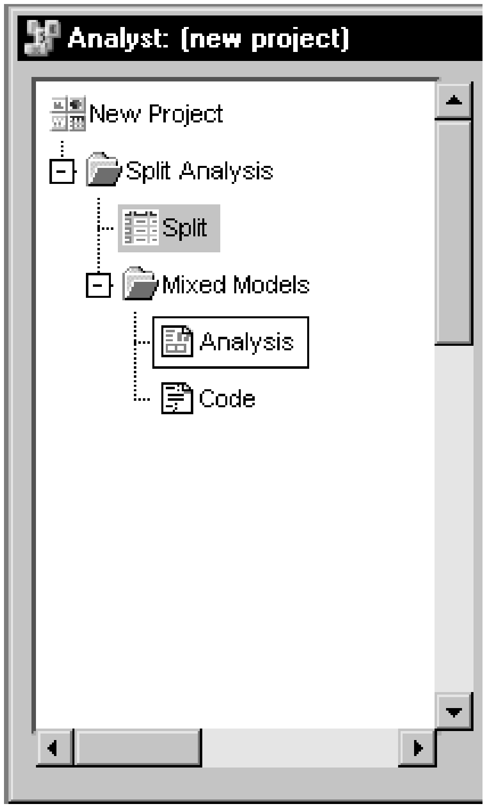

Figure 15.5. Mixed Models: Project Tree

Double-click on the **Analysis** node in the project tree to view the contents in a separate window.

Analysis

Class Level Information

Class	Levels	Values
Block	4	1 2 3 4
A	3	1 2 3
B	2	1 2

Dimensions

Covariance Parameters	3
Columns in X	12
Columns in Z	16
Subjects	1
Max Obs Per Subject	24
Observations Used	24
Observations Not Used	0
Total Observations	24

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	139.81461222	
1	1	119.76184570	0.00000000

Figure 15.6. Mixed Models: Model Information

Figure 15.6 displays class level information, dimensions of model matrices, and the iteration history of the estimated model. The "Class Level Information" table lists the levels of all classification variables included in the model. The "Dimensions" table includes the number of estimated covariance parameters as well as the number of columns in the **X** and **Z** design matrices.

The Mixed Models task estimates the variance components for Block, A*Block, and the residual by a method known as residual (restricted) maximum likelihood (REML). The REML estimates are the values that maximize the likelihood of a set of linearly independent error contrasts, and they provide a correction for the downward bias found in the usual maximum likelihood estimates.

The "Iteration History" table records the steps of the REML optimization process. The objective function of the process is -2 times the restricted likelihood. The Mixed Models task attempts to minimize this objective function via the Newton-Raphson algorithm, which uses the first and second derivatives of the objective function to iteratively find its minimum. For this example, only one iteration is required to obtain the estimates. The Evaluations column reveals that the restricted likelihood is evaluated once for each iteration, and the criterion of 0 indicates that the Newton-Raphson algorithm has converged.

Analysis

Covariance Parameter Estimates

Cov Parm	Estimate	Alpha	Lower	Upper
Block	62.3958	0.05	18.3875	1404.52
Block*A	15.3819	0.05	5.1893	167.12
Residual	9.3611	0.05	4.4289	31.1992

The Mixed Procedure

Fit Statistics

Res Log Likelihood	-59.9
Akaike's Information Criterion	-62.9
Schwarz's Bayesian Criterion	-62.0
-2 Res Log Likelihood	119.8

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
A	2	6	4.07	0.0764
B	1	9	19.39	0.0017
A*B	2	9	4.02	0.0566

Figure 15.7. Mixed Models: Covariance Estimates and Tests for Fixed Effects

Figure 15.4 displays covariance parameter estimates, information on the model fit, and Type 3 tests of fixed effects. The REML estimates for the variance components of Block, A*Block, and the residual are 62.4, 15.4, and 9.4, respectively. The "Fit Statistics" table lists several pieces of information about the fitted mixed model: the residual log likelihood, Akaike's and Schwarz's criteria, and the -2 residual log likelihood. Akaike's and Schwarz's criteria can be used to compare different models; models with larger values for these criteria are preferred.

The tests of fixed effects are produced using Type 3 estimable functions. The test for the A*B interaction has a *p*-value of 0.0566, indicating that there is moderate evidence of an interaction between crop varieties and irrigation levels.

Analysis

Least Squares Means

Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
37.0000	4.6674	4.68	7.93	0.0007	0.05	24.7495	49.2505
28.7500	4.6674	4.68	6.16	0.0021	0.05	16.4995	41.0005
38.0000	4.6674	4.68	8.14	0.0006	0.05	25.7495	50.2505
30.2500	4.6674	4.68	6.48	0.0017	0.05	17.9995	42.5005
26.0000	4.6674	4.68	5.57	0.0032	0.05	13.7495	38.2505
25.5000	4.6674	4.68	5.46	0.0034	0.05	13.2495	37.7505

Figure 15.8. Mixed Models: Least Squares Means

Figure 15.8 displays the least-squares means for each combination of irrigation levels (A) and crop varieties (B). At each irrigation level, the response is higher for the first crop variety compared to the second variety. The interaction between crop variety and irrigation levels is evident in that variety 1 has a higher mean response than variety 2 at irrigation levels 1 and 2, but the two varieties have nearly the same mean response at irrigation level 3.

Clustered Data

The example in this section contains information on a study investigating the heights of individuals sampled from different families. The response variable Height measures the height (in inches) of 18 individuals that are classified according to Family and Gender. Since the data occurs in clusters (families), it is very likely that observations from the same family are statistically correlated and not independent. In this case, it is inappropriate to analyze the data using a standard linear model.

A simple way to model the correlation is through the use of a Family random effect. The Family effect is assumed to be normally distributed with mean of zero and some unknown variance. Defining Family as a random effect sets up a common correlation among all observations having the same level of family.

In addition, a female within a certain family may exhibit more correlation with other females in that same family than with the males in that family, and likewise for males. Defining Family*Gender as a random effect models an additional correlation for all observations having the same value of both Family and Gender.

Open the Heights Data Set

These data are provided as the Heights data set in the Analyst Sample Library. To open the Heights data set, follow these steps:

1. Select **Tools** → **Sample Data** . . .
2. Select Heights.
3. Click **OK** to create the sample data set in your Sasuser directory.
4. Select **File** → **Open By SAS Name** . . .
5. Select Sasuser from the list of **Libraries**.
6. Select Heights from the list of members.
7. Click **OK** to bring the Heights data set into the data table.

Specify the Mixed Models Analysis

To request a mixed models analysis, follow these steps:

1. Select **Statistics** → **ANOVA** → **Mixed Models** . . .
2. Select Height as the dependent variable.
3. Select Family and Gender as classification variables.
4. Click **Model** to open the **Model** dialog.
5. Ensure that the **Fixed effects** check box is selected.
6. Select Gender and click **Add**.
7. Select the **Random effects** check box, and then select Family and click **Add**.
8. Select Family and Gender, and click **Cross**.
9. Click **OK** to return to the main dialog.

Based on your selections, the Mixed Models task constructs the **X** matrix by creating indicator variables for the Gender effect and including a column of 1s to model the global intercept. The **Z** matrix contains indicator variables for both the Family effect and the Family*Gender interaction.

Produce a Residual Plot

The Mixed Models task can produce means plots for fixed main effects and interactions, plots of predicted values, and residual plots that include or do not include random effects. To produce a plot of residuals versus predicted values that includes random effects, follow these steps:

1. Click **Plots** to open the **Plots** dialog.
2. Click on the **Residual** tab, and select **Plot residuals vs predicted** in the **Residual plots (including random effects)** box.

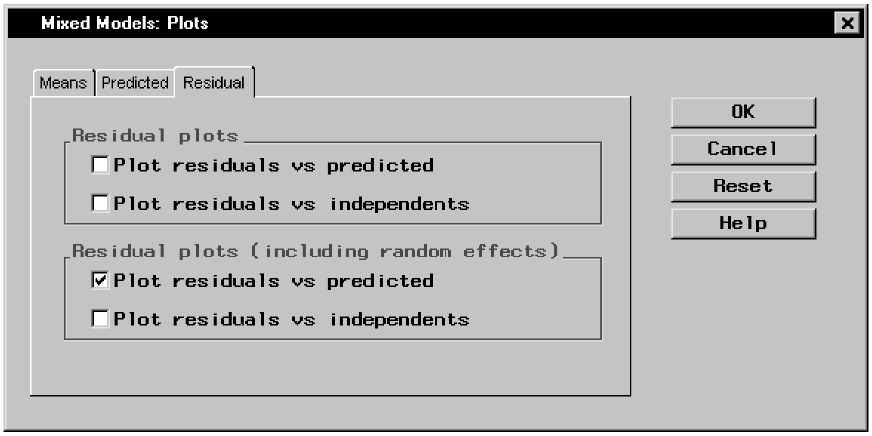

Figure 15.9. Mixed Model: Plots Dialog

When you have completed your selections, click **OK** in the main dialog to perform the analysis.

Review the Results

The results are presented in the project tree under the Heights data in the **Mixed Models** folder, as displayed in Figure 15.10. The three nodes represent the mixed models results, the plot of residuals versus predicted values, and the SAS programming statements (labeled **Code**) that generate the output.

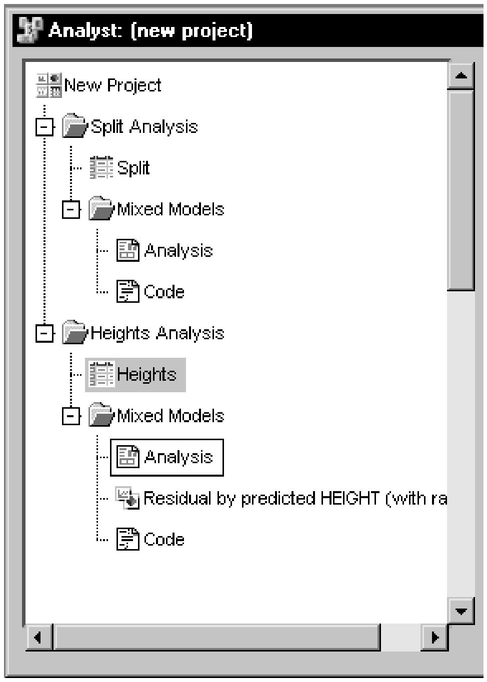

Figure 15.10. Mixed Models: Project Tree

Double-click on the **Analysis** node in the project tree to view the contents in a separate window.

Analysis

Covariance Parameter Estimates

Cov Parm	Estimate	Alpha	Lower	Upper
Family	2.4010	0.05	0.4593	4490.18
Family*Gender	1.7657	0.05	0.3662	1012.90
Residual	2.1668	0.05	1.0373	7.0071

Fit Statistics

Res Log Likelihood	-35.5
Akaike's Information Criterion	-38.5
Schwarz's Bayesian Criterion	-37.6
-2 Res Log Likelihood	71.0

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
Gender	1	2.84	7.95	0.0712

Figure 15.11. Mixed Models: Analysis Results

Figure 15.11 displays the mixed models analysis results for the clustered Heights data. The covariance parameter estimates for Family, Family*Gender, and the residual variance are 2.4, 1.8, and 2.2, respectively. The "Test of Fixed Effects" table contains a significance test for the single fixed effect, Gender. With a *p*-value of 0.0712, the Type 3 test of Gender is not significant at the $\alpha = 0.05$ level of significance. Note that the denominator degrees of freedom for the Type 3 test are computed using a general Satterthwaite approximation. A benefit of performing a random effects analysis using both Family and Family*Gender as random effects is that you can make inferences about gender that apply to an entire population of families, not necessarily to the specific families in this study.

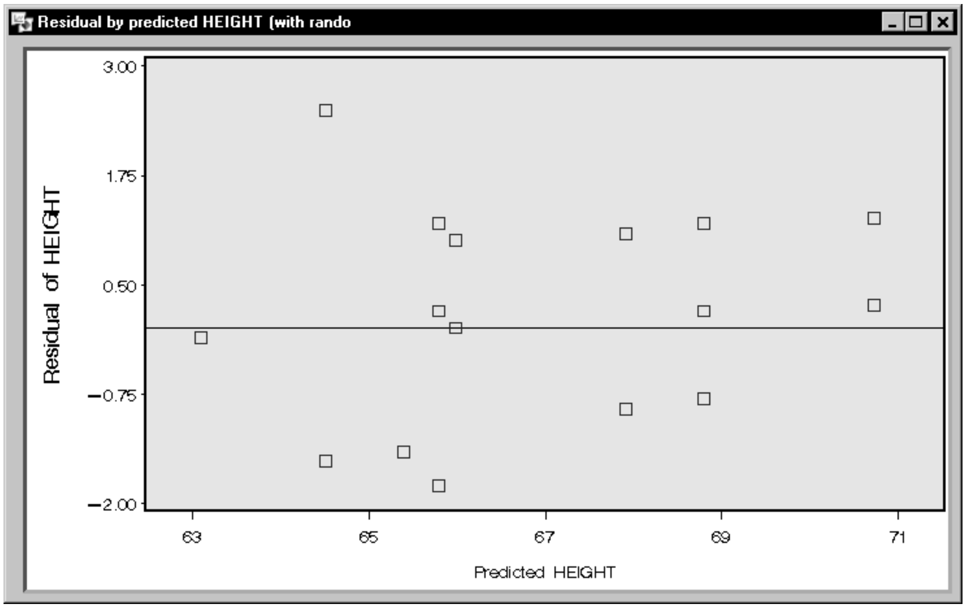

Figure 15.12. Mixed Models: Residuals Plot

Figure 15.12 displays a plot of the residuals versus predicted values that includes random effects, $y - \mathbf{X}\hat{\beta} - \mathbf{Z}\hat{\gamma}$ versus $\mathbf{X}\hat{\beta} + \mathbf{Z}\hat{\gamma}$. Plots are useful for checking model assumptions and identifying potential outlying and influential observations. Based on the plot in Figure 15.12, the data seem to exhibit relatively constant variance across predicted values, and there do not appear to be any outliers or influential observations.

References

Littell, R. C., Milliken, G. A., Stroup, W. W., and Wolfinger, R. D. (1996), *SAS System for Mixed Models*, Cary, NC: SAS Institute Inc.

SAS Institute Inc. (1999), *SAS/STAT User's Guide, Version 7-1*, Cary, NC: SAS Institute Inc.

Stroup, W. W. (1989), "Predictable Functions and Prediction Space in the Mixed Model Procedure," in *Applications of Mixed Models in Agriculture and Related Disciplines*, Southern Cooperative Series Bulletin No. 343, Louisiana Agricultural Experiment Station, Baton Rouge, 39–48.

The correct bibliographic citation for this manual is as follows: SAS Institute Inc., *The Analyst Application, First Edition*, Cary, NC: SAS Institute Inc., 1999. 476 pp.

The Analyst Application, First Edition

ISBN 1–58025–446–2

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

1st printing, October 1999