# Factorial ANOVA

This analysis technique is used for experimental data in which there is a continuous response variable and one or more independent classification variables. The total variation in the response variable is explained as the sum of the variation due to the effects of the classification variables and the variation due to random error. This analysis is appropriate if you have multiple classification variables.

# Factorial ANOVA Output

The output of the analysis includes information about the levels of the independent variables followed by the standard ANOVA table, which includes the Degrees of Freedom, Sum of Squares and the Mean Square for the model and for random error. "R-Square" is the proportion of variability accounted for by the independent variables. "C.V." is the coefficient of variation. "Root MSE" is the square root of the Mean Square for Error. The F-statistic is the ratio of the Mean Square values, and "Pr > F" is the associated *p*-value. A small *p*-value indicates that the model is significant in explaining the variation in the dependent variable. Below this information, the model sum of squares is broken down by the contributions of the effects in the model, and the individual effects are analyzed for their significance. Additional output may be produced depending on the options you specify.

# PROC GLM: Introduction

The GLM procedure uses the method of least squares to fit general linear models. Among the statistical methods available in PROC GLM are regression, analysis of variance, analysis of covariance, multivariate analysis of variance, and partial correlation

PROC GLM analyzes data within the framework of General linear models. PROC GLM handles models relating one or several continuous dependent variables to one or several independent variables. The independent variables may be either *classification* variables, which divide the observations into discrete groups, or *continuous* variables. Thus, the GLM procedure can be used for many different analyses, including

- simple regression
- multiple regression
- analysis of variance (ANOVA), especially for unbalanced data
- analysis of covariance
- weighted regression
- polynomial regression
- partial correlation
- multivariate analysis of variance (MANOVA)

# PROC GLM: Syntax

The following statements are available in PROC GLM.

**PROC GLM** < options > **;**
**CLASS** variables **;**
**MODEL** dependents=independents < / options > **;**
**BY** variables **;**
**LSMEANS** effects < / options > **;**

To use PROC GLM, the PROC GLM and MODEL statements are required. You can specify only one MODEL statement. If your model contains classification effects, the classification variables must be listed in a CLASS statement, and the CLASS statement must appear before the MODEL statement.

# PROC GLM: LSMEANS Statement

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

Least-squares means (LS-means) are computed for each *effect* listed in the LSMEANS statement. You may specify only classification effects in the LSMEANS statement -that is, effects that contain only classification variables. You may also specify options to perform multiple comparisons.

Multiple effects can be specified in one LSMEANS statement, or multiple LSMEANS statements can be used, but they must all appear after the MODEL statement. For example,

```
proc glm;
   class A B;
   model Y=A B A*B;
   lsmeans A B A*B;
run;
```

LS-means are displayed for each level of the A, B, and A*B effects.

You can specify the following options in the LSMEANS statement after a slash:

STDERR

The STDERR option produces the standard error of the LS-means and the probability level for the hypothesis $H_0$: *LS-mean* = 0.

PDIFF<=*difftype*>

The PDIFF option requests that *p*-values for differences of the LS-means be produced. The optional *difftype* specifies which differences to display. Possible values for *difftype* are ALL, CONTROL, CONTROLL, and CONTROLU. The ALL value requests all pairwise differences, and it is the default. For the other option go to help files.