SAS/STAT® 9.3 User's Guide
The GLMPOWER Procedure
(Chapter)

Chapter 43

The GLMPOWER Procedure

Contents

Overview: GLMPOWER Procedure

Power and sample size analysis optimizes the resource usage and design of a study, improving chances of conclusive results with maximum efficiency. The GLMPOWER procedure performs prospective power and sample size analysis for linear models, with a variety of goals:

- determining the sample size required to get a significant result with adequate probability (power)

- characterizing the power of a study to detect a meaningful effect

- conducting what-if analyses to assess sensitivity of the power or required sample size to other factors

Here *prospective* indicates that the analysis pertains to planning for a future study. This is in contrast to *retrospective* analysis for a past study, which is not supported by this procedure.

The statistical analyses that are covered include Type III tests and contrasts of fixed effects in univariate linear models, optionally with covariates. The covariates can be continuous or categorical. Tests and contrasts involving random effects are not supported. For power and sample size analyses in a variety of other statistical situations, see Chapter 70, "The POWER Procedure."

Input for PROC GLMPOWER includes the components considered in study planning:

- design (including subject profiles and their allocation weights)

- statistical model

- contrasts of class effects

- significance level (alpha)

- surmised response means for subject profiles (often called "cell means")

- surmised variability

- power

- sample size

In order to identify power or sample size as the result parameter, you designate it by a missing value in the input. The procedure calculates this result value over one or more scenarios of input values for all other components.

You specify the design and the cell means by using an *exemplary data set*, a data set of artificial values constructed to represent the intended sampling design and the surmised response means in the underlying population. You specify the model and contrasts by using MODEL and CONTRAST statements similar to those in the GLM, ANOVA, and MIXED procedures. You specify the remaining parameters with the POWER statement, which is similar to analysis statements in the POWER procedure.

In addition to tabular results, PROC GLMPOWER produces graphs. You can produce the most common types of plots easily with default settings and use a variety of options for more customized graphics. For

example, you can control the choice of axis variables, axis ranges, number of plotted points, mapping of graphical features (such as color, line style, symbol, and panel) to analysis parameters, and legend appearance.

If ODS Graphics is enabled, then PROC GLMPOWER uses ODS Graphics to create graphs; otherwise, traditional graphs are produced.

For more information about enabling and disabling ODS Graphics, see the section "Enabling and Disabling ODS Graphics" on page 609 in Chapter 21, "Statistical Graphics Using ODS."

For specific information about the statistical graphics and options available with the GLMPOWER procedure, see the PLOT statement and the section "ODS Graphics" on page 3362.

The GLMPOWER procedure is one of several tools available in SAS/STAT software for power and sample size analysis. PROC POWER covers a variety of other analyses such as *t* tests, equivalence tests, confidence intervals, binomial proportions, multiple regression, one-way ANOVA, survival analysis, logistic regression, and the Wilcoxon rank-sum test. The Power and Sample Size application provides a user interface and implements many of the analyses supported in the procedures. See Chapter 70, "The POWER Procedure," and Chapter 71, "The Power and Sample Size Application," for details.

The following sections of this chapter describe how to use PROC GLMPOWER and discuss the underlying statistical methodology. The section "Getting Started: GLMPOWER Procedure" on page 3339 introduces PROC GLMPOWER with examples of power computation for a two-way analysis of variance. The section "Syntax: GLMPOWER Procedure" on page 3345 describes the syntax of the procedure. The section "Details: GLMPOWER Procedure" on page 3357 summarizes the methods employed by PROC GLMPOWER and provides details on several special topics. The section "Examples: GLMPOWER Procedure" on page 3363 illustrates the use of the GLMPOWER procedure with several applications.

For an overview of methodology and SAS tools for power and sample size analysis, see Chapter 18, "Introduction to Power and Sample Size Analysis." For more discussion and examples for linear models, see Castelloe and O'Brien (2001), O'Brien and Shieh (1992), Muller et al. (1992), and O'Brien and Muller (1993). For additional discussion of general power and sample size concepts, see O'Brien and Castelloe (2007), Castelloe (2000), Muller and Benignus (1992), and Lenth (2001).

Getting Started: GLMPOWER Procedure

Simple Two-Way ANOVA

This example demonstrates how to use PROC GLMPOWER to compute and plot power for each effect test in a two-way analysis of variance (ANOVA).

Suppose you are planning an experiment to study the effect of light exposure at three levels on the growth of two varieties of flowers. The planned data analysis is a two-way ANOVA with flower height (measured at two weeks) as the response and a model consisting of the effects of light exposure, flower variety, and their interaction. You want to calculate the power of each effect test for a balanced design with a total of 60

specimens (10 for each combination of exposure and variety) with $\alpha = 0.05$ for each test.

As a first step, create an *exemplary data set* describing your conjectures about the underlying population means. You believe that the mean flower height for each combination of variety and exposure level (that is, for each design profile, or for each *cell* in the design) roughly follows Table 43.1.

Table 43.1 Mean Flower Height (in cm) by Variety and Exposure

	Exposure		
Variety	1	2	3
1	14	16	21
2	10	15	16

The following statements create a data set named Exemplary containing these cell means.

```
data Exemplary;
   do Variety = 1 to 2;
      do Exposure = 1 to 3;
         input Height @@;
         output;
      end;
   end;
   datalines;
      14 16 21
      10 15 16
   ;
run;
```

You also conjecture that the error standard deviation is about 5 cm.

Use the DATA= option in the PROC GLMPOWER statement to specify Exemplary as the exemplary data set. Identify the classification variables (Variety and Exposure) by using the CLASS statement. Specify the model by using the MODEL statement. Use the POWER statement to specify power as the result parameter and provide values for the other analysis parameters, error standard deviation and total sample size. The following SAS statements perform the power analysis:

```
proc glmpower data=Exemplary;
   class Variety Exposure;
   model Height = Variety | Exposure;
   power
      stddev = 5
      ntotal = 60
      power  = .;
run;
```

The MODEL statement defines the full model including both main effects and the interaction. The POWER= option in the POWER statement identifies power as the result parameter with a missing value (POWER=.). The STDDEV= option specifies an error standard deviation of 5, and the NTOTAL= option specifies a total sample size of 60. The default value for the ALPHA= option sets the significance level to $\alpha = 0.05$.

Figure 43.1 shows the output.

Figure 43.1 Sample Size Analysis for Two-Way ANOVA

```
                         The GLMPOWER Procedure

                         Fixed Scenario Elements

               Dependent Variable              Height
               Error Standard Deviation             5
               Total Sample Size                   60
               Alpha                             0.05
               Error Degrees of Freedom            54


                            Computed Power

                                        Test
               Index        Source        DF     Power

                  1     Variety           1     0.718
                  2     Exposure          2     0.957
                  3     Variety*Exposure  2     0.191
```

The power is about 0.72 for the test of the Variety effect. In other words, there is a probability of 0.72 that the test of the Variety effect will produce a significant result (given the assumptions for the means and error standard deviation). The power is 0.96 for the test of the Exposure effect and 0.19 for the interaction test.

Now, suppose you want to account for some of your uncertainty in conjecturing the true error standard deviation by evaluating the power at reasonable low and high values, 4 and 6.5. You also want to plot power for sample sizes between 30 and 90. The following statements perform the analysis:

```
ods listing style=htmlbluecml;
ods graphics on;

proc glmpower data=Exemplary;
   class Variety Exposure;
   model Height = Variety | Exposure;
   power
      stddev = 4 6.5
      ntotal = 60
      power  = .;
   plot x=n min=30 max=90;
run;

ods graphics off;
```

The PLOT statement with the X=N option requests a plot with sample size on the X axis. (The result parameter—in this case, power—is always plotted on the other axis.) The MIN= and MAX= options in the PLOT statement specify the sample size range. The ODS GRAPHICS ON statement enables ODS Graphics. The ODS LISTING STYLE=HTMLBLUECML statement specifies the HTMLBLUECML style, which is suitable for use with PROC GLMPOWER because it allows both marker symbols and line styles to vary. See the section "ODS Styles Suitable for Use with PROC GLMPOWER" on page 3363 for more information.

Figure 43.2 shows the output, and Figure 43.3 shows the plot.

Figure 43.2 Sample Size Analysis for Two-Way ANOVA with Input Ranges

```
                    The GLMPOWER Procedure

                  Fixed Scenario Elements

        Dependent Variable                Height
        Total Sample Size                     60
        Alpha                               0.05
        Error Degrees of Freedom              54


                     Computed Power

                                 Std    Test
        Index          Source    Dev      DF     Power

          1      Variety         4.0       1     0.887
          2      Variety         6.5       1     0.496
          3      Exposure        4.0       2     0.996
          4      Exposure        6.5       2     0.793
          5      Variety*Exposure 4.0      2     0.280
          6      Variety*Exposure 6.5      2     0.130
```

Figure 43.3 Plot of Power versus Sample Size for Two-Way ANOVA with Input Ranges
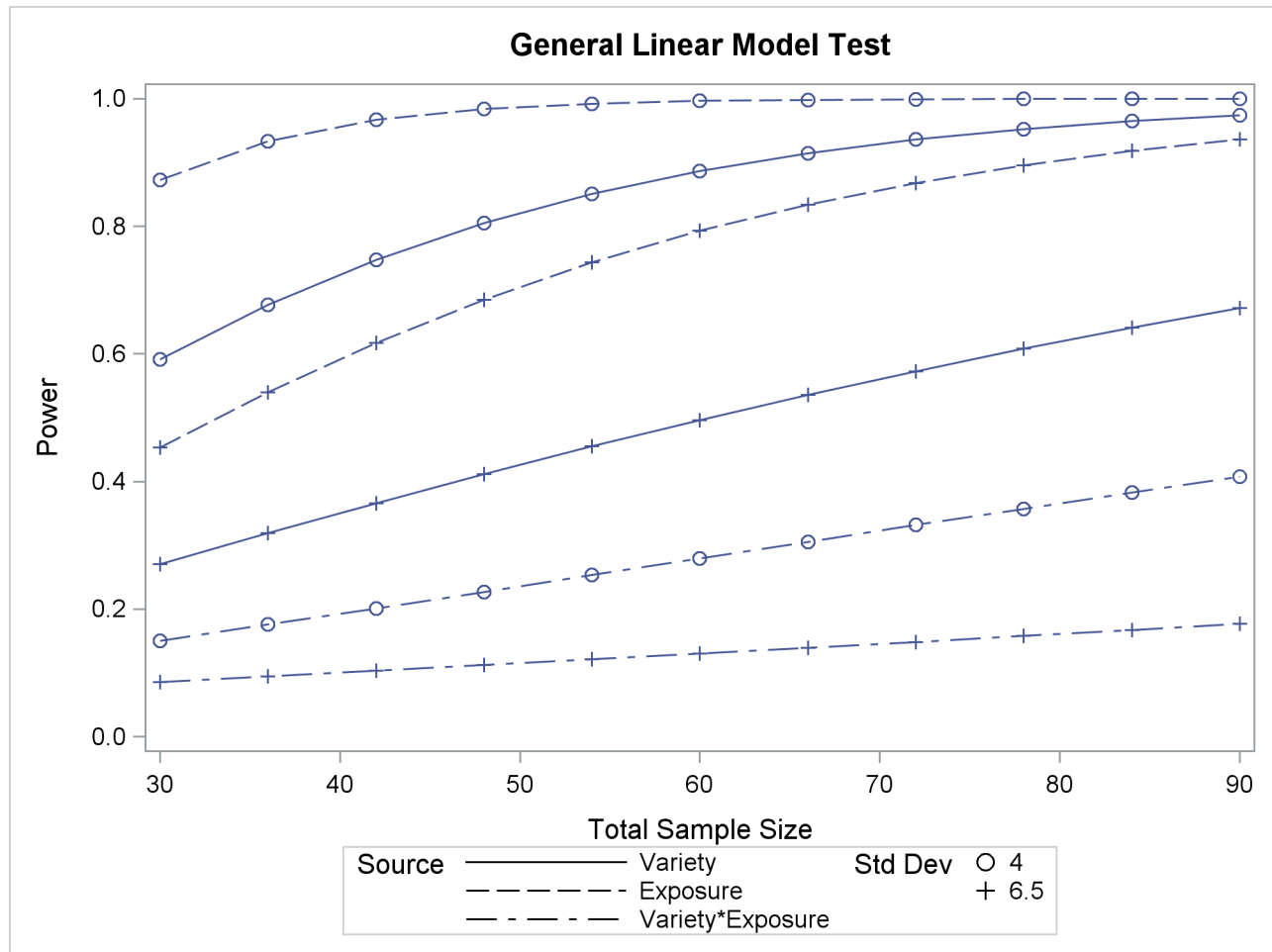
Figure 43.2 reveals that the power ranges from about 0.130 to 0.996 for the different effect tests and scenarios for standard deviation, with a sample size of 60. In Figure 43.3, the line style identifies the effect test, and the plotting symbol identifies the standard deviation. The locations of the plotting symbols identify actual computed powers; the curves are linear interpolations of these points. Note that the computed points in the plot occur at sample size multiples of 6, because there are 6 cells in the design (and by default, sample sizes are rounded to produce integer cell sizes).

Incorporating Contrasts, Unbalanced Designs, and Multiple Means Scenarios

Suppose you want to compute power for the two-way ANOVA described in the section "Simple Two-Way ANOVA" on page 3339, but you want to additionally perform the following tasks:

- try an unbalanced sample size allocation with respect to Exposure, using twice as many samples for levels 2 and 3 as for level 1

- consider an additional, less optimistic scenario for the cell means, shown in Table 43.2

● test a contrast of Exposure comparing levels 1 and 3

Table 43.2 Additional Cell Means Scenario

Variety	Exposure 1	2	3
1	15	16	20
2	11	14	15

To specify the unbalanced design and the additional cell means scenario, you can add two new variables to the exemplary data set (Weight for the sample size weights, and HeightNew for the new cell means scenario). Change the name of the original cell means scenario to HeightOrig. The following statements define the exemplary data set:

```
data Exemplary;
   input Variety $ Exposure $ HeightOrig HeightNew Weight;
   datalines;
         1   1   14  15  1
         1   2   16  16  2
         1   3   21  20  2
         2   1   10  11  1
         2   2   15  14  2
         2   3   16  15  2
   ;
run;
```

In PROC GLMPOWER, specify the name of the weight variable by using the WEIGHT statement, and specify the name of the cell means variables as dependent variables in the MODEL statement. Use the CONTRAST statement to specify the contrast as you would in PROC GLM. The following statements perform the sample size analysis.

```
proc glmpower data=Exemplary;
   class Variety Exposure;
   model HeightOrig HeightNew = Variety | Exposure;
   weight Weight;
   contrast 'Exposure=1 vs Exposure=3' Exposure 1 0 -1;
   power
      stddev = 5
      ntotal = 60
      power  = .;
run;
```

Figure 43.4 shows the output.

Figure 43.4 Sample Size Analysis for More Complex Two-Way ANOVA

```
                        The GLMPOWER Procedure

                        Fixed Scenario Elements

                Weight Variable                 Weight
                Error Standard Deviation              5
                Total Sample Size                    60
                Alpha                              0.05
                Error Degrees of Freedom             54


                          Computed Power

                                                       Test
      Index     Dependent      Type            Source         DF     Power

        1      HeightOrig     Effect     Variety               1     0.672
        2      HeightOrig     Effect     Exposure              2     0.911
        3      HeightOrig     Effect     Variety*Exposure      2     0.217
        4      HeightOrig     Contrast   Exposure=1 vs Exposure=3   1     0.951
        5      HeightNew      Effect     Variety               1     0.754
        6      HeightNew      Effect     Exposure              2     0.633
        7      HeightNew      Effect     Variety*Exposure      2     0.137
        8      HeightNew      Contrast   Exposure=1 vs Exposure=3   1     0.705
```

The power of the contrast of Exposure levels 1 and 3 is about 0.95 for the original cell means scenario (HeightOrig) and only 0.71 for the new one (HeightNew). The power is higher for the test of Variety, but lower for the tests of Exposure and of Variety*Exposure for the new cell means scenario compared to the original one. Note also for the HeightOrig scenario that the power for the unbalanced design (Figure 43.4) compared to the balanced design (Figure 43.1) is slightly lower for the tests of Variety and Exposure, but slightly higher for the test of Variety*Exposure.

Syntax: GLMPOWER Procedure

The following statements are available in PROC GLMPOWER:

PROC GLMPOWER < *options* > ;
   BY *variables* ;
   CLASS *variables* ;
   CONTRAST *'label' effect values* < *... effect values* > < */ options* > ;
   MODEL *dependents = independents* ;
   PLOT < *plot-options* > < */ graph-options* > ;
   POWER < *options* > ;
   WEIGHT *variable* ;

The PROC GLMPOWER statement, the MODEL statement, and the POWER statement are required. 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. In addition, CONTRAST and POWER statements must appear after the MODEL statement. PLOT statements must appear after the POWER statement that defines the analysis for the plot.

You can use multiple CONTRAST, POWER, and PLOT statements. Each CONTRAST statement defines a separate contrast. Each POWER statement produces a separate analysis and uses the information contained in the CLASS, MODEL, WEIGHT, and all CONTRAST statements. Each PLOT statement refers to the previous POWER statement and generates a separate graph (or set of graphs).

Table 43.3 summarizes the basic functions of each statement in PROC GLMPOWER. The syntax of each statement in Table 43.3 is described in the following pages.

Table 43.3 Statements in the GLMPOWER Procedure

Statement	Description
PROC GLMPOWER	Invokes procedure and specifies exemplary data set
BY	Specifies variables to define subgroups for the analysis
CLASS	Declares classification variables
CONTRAST	Defines linear tests of model parameters
MODEL	Defines model and specifies dependent variable(s) used for cell means scenarios
PLOT	Displays graphs for preceding POWER statement
POWER	Identifies parameter to solve for and provides one or more scenarios for values of other analysis parameters
WEIGHT	Specifies variable for allocating sample sizes to different subject profiles

PROC GLMPOWER Statement

> **PROC GLMPOWER** < *options* > ;

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

DATA=*SAS-data-set*
> names a SAS data set to be used as the exemplary data set, which is an artificial data set constructed to represent the intended sampling design and the conjectured response means for the underlying population.

ORDER=DATA | FORMATTED | FREQ | INTERNAL
> specifies the order in which to sort the levels of the classification variables (which are specified in the

CLASS statement).

This option applies to the levels for all classification variables, except when you use the (default) ORDER=FORMATTED option with numeric classification variables that have no explicit format. With this option, the levels of such variables are ordered by their internal value.

The ORDER= option can take the following values:

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

By default, ORDER=FORMATTED. For ORDER=FORMATTED and ORDER=INTERNAL, the sort order is machine-dependent.

For more information about sorting order, see the chapter on the SORT procedure in the *Base SAS Procedures Guide* and the discussion of BY-group processing in *SAS Language Reference: Concepts*.

PLOTONLY
 specifies that only graphical results from the PLOT statement be produced.

BY Statement

 BY *variables* **;**

You can specify a BY statement with PROC GLMPOWER to obtain separate analyses on observations in groups that are 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. If you specify more than one BY statement, only the last one specified is used.

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

- Sort the data by using the SORT procedure with a similar BY statement.

- Specify the NOTSORTED or DESCENDING option in the BY statement for the GLMPOWER procedure. The NOTSORTED option does not mean that the data are unsorted but rather 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 by using the DATASETS procedure (in Base SAS software).

Because sorting the data changes the order in which PROC GLMPOWER reads observations, the sorting order for the levels of the classification variables might be affected if you have also specified ORDER=DATA in the PROC GLMPOWER statement. This, in turn, affects specifications in CONTRAST statements.

For more information about BY-group processing, see the discussion in *SAS Language Reference: Concepts*. For more information about the DATASETS procedure, see the discussion in the *Base SAS Procedures Guide*.

CLASS Statement

CLASS *variables* **;**

The CLASS statement names the classification variables to be used in the analysis. If you use the CLASS statement, it must appear before the MODEL statement.

Classification variables can be either character or numeric. By default, class levels are determined from the entire set of formatted values of the CLASS variables.

CONTRAST Statement

CONTRAST *'label' effect values < . . . effect values > < / options >* **;**

The CONTRAST statement enables you to define custom hypothesis tests by specifying an $\mathbf{L}$ vector or matrix for testing the hypothesis $\mathbf{L}\boldsymbol{\beta} = 0$. Thus, to use this feature you must be familiar with the details of the model parameterization used in PROC GLM. For more information, see the section "Parameterization of PROC GLM Models" on page 3190 of Chapter 41, "The GLM Procedure." All of the elements of the $\mathbf{L}$ vector can be given, or if only certain portions of the $\mathbf{L}$ vector are given, the remaining elements are constructed by PROC GLMPOWER from the context (in a manner similar to rule 4 discussed in the section "Construction of Least Squares Means" on page 3226 of Chapter 41, "The GLM Procedure").

There is no limit to the number of CONTRAST statements you can specify. Each sample size analysis includes tests for all CONTRAST statements.

In the CONTRAST statement,

label identifies the contrast on the output. A label is required for every contrast specified. Labels must be enclosed in quotes.

effect identifies an effect that appears in the MODEL statement, or the INTERCEPT effect. You do not need to include all effects that are in the MODEL statement.

values are constants that are elements of the $\mathbf{L}$ vector associated with the effect.

You can specify the following option in the CONTRAST statement after a slash (/):

SINGULAR=*number*

tunes the estimability checking. If ABS($\mathbf{L} - \mathbf{LH}$) > $C \times number$ for any row in the contrast, then $\mathbf{L}$ is declared nonestimable. $\mathbf{H}$ is the $(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{X}$ matrix, and C is ABS($\mathbf{L}$) except for rows where $\mathbf{L}$ is zero, and then it is 1. The default value for the SINGULAR= option is 10^{-4}. Values for the SINGULAR= option must be between 0 and 1.

The CONTRAST statement enables you to perform custom hypothesis tests. If the hypothesis is estimable, then the sum of squares due to it, SS($H_0 : \mathbf{L\beta} = 0$), is computed as

$$(\mathbf{Lb})'(\mathbf{L}(\mathbf{X}'\mathbf{X})^-\mathbf{L}')^{-1}(\mathbf{Lb})$$

where $\mathbf{b} = (\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{y}$ is the estimated solution vector.

The degrees of freedom associated with the hypothesis are equal to the row rank of $\mathbf{L}$. The sum of squares computed in this situation is equivalent to the sum of squares computed using an $\mathbf{L}$ matrix with any row deleted that is a linear combination of previous rows.

Multiple-degrees-of-freedom hypotheses can be specified by separating the rows of the $\mathbf{L}$ matrix with commas.

MODEL Statement

MODEL *dependents = independents* **;**

The MODEL statement serves two basic purposes:

- The *dependents* specify scenarios for the cell means.

- The *independents* specify the independent effects.

The *independents* can involve classification variables, continuous variables, or both. You can include main effects and interactions by using the effects notation of PROC GLM; see the section "Specification of Effects" on page 3186 in Chapter 41, "The GLM Procedure" for further details. For any model effect involving classification variables (interactions as well as main effects), the number of levels cannot exceed 32,767. If no independent effects are specified, only an intercept term is fit. The MODEL statement must appear before the POWER statement if the EFFECTS option is used in the POWER statement.

You can account for covariates in the model by using the NCOVARIATES= option and either the CORRXY= or PROPVARREDUCTION= option in the POWER statement.

Each dependent variable refers to a set of surmised cell means in the exemplary data set (named by the DATA= option in the PROC GLMPOWER statement). These cell means are response means for all of the subject profiles. Multiple dependent variables correspond to multiple scenarios for these cell means. All models are univariate; the GLMPOWER procedure currently does not support multivariate analyses.

The MODEL statement is required. You can specify only one MODEL statement.

PLOT Statement

> **PLOT** < *plot-options* > < / *graph-options* > ;

The PLOT statement produces a graph or set of graphs for the sample size analysis defined by the previous POWER statement. The *plot-options* define the plot characteristics, and the *graph-options* are like those in SAS/GRAPH software. If ODS Graphics is enabled, then the PLOT statement uses ODS Graphics to create graphs. For example:

```
ods listing style=htmlbluecml;
ods graphics on;

proc glmpower data=Exemplary;
   class Variety Exposure;
   model Height = Variety | Exposure;
   power
      stddev = 4 6.5
      ntotal = 60
      power  = .;
   plot x=n min=30 max=90;
run;

ods graphics off;
```

Otherwise, traditional graphics are produced. For example:

```
ods graphics off;

proc glmpower data=Exemplary;
   class Variety Exposure;
   model Height = Variety | Exposure;
   power
      stddev = 4 6.5
      ntotal = 60
      power  = .;
   plot x=n min=30 max=90;
run;
```

For more information about enabling and disabling ODS Graphics, see the section "Enabling and Disabling ODS Graphics" on page 609 in Chapter 21, "Statistical Graphics Using ODS."

The ODS LISTING STYLE=HTMLBLUECML statement specifies the HTMLBLUECML style, which is suitable for use with PROC GLMPOWER because it allows both marker symbols and line styles to vary. See the section "ODS Styles Suitable for Use with PROC GLMPOWER" on page 3363 for more information.

Options

You can specify the following *plot-options* in the PLOT statement.

INTERPOL=JOIN | NONE

>specifies the type of curve to draw through the computed points. The INTERPOL=JOIN option connects computed points with straight lines. The INTERPOL=NONE option leaves computed points unconnected.

KEY=BYCURVE <(*bycurve-options* **)>**

KEY=BYFEATURE <(*byfeature-options* **)>**

KEY=ONCURVES

>specifies the style of key (or "legend") for the plot. The default is KEY=BYFEATURE, which specifies a key with a column of entries for each plot feature (line style, color, and/or symbol). Each entry shows the mapping between a value of the feature and the value(s) of the analysis parameter(s) linked to that feature. The KEY=BYCURVE option specifies a key with each row identifying a distinct curve in the plot. The KEY=ONCURVES option places a curve-specific label adjacent to each curve.

>You can specify the following *byfeature-options* in parentheses after the KEY=BYCURVE option.

>**NUMBERS=OFF | ON**

>>specifies how the key should identify curves. If NUMBERS=OFF, then the key includes symbol, color, and line style samples to identify the curves. If NUMBERS=ON, then the key includes numbers matching numeric labels placed adjacent to the curves. The default is NUMBERS=ON.

>**POS=BOTTOM | INSET**

>>specifies the position of the key. The POS=BOTTOM option places the key below the X axis. The POS=INSET option places the key inside the plotting region and attempts to choose the least crowded corner. The default is POS=BOTTOM.

>You can specify the following *byfeature-options* in parentheses after KEY=BYFEATURE option.

>**POS=BOTTOM | INSET**

>>specifies the position of the key. The POS=BOTTOM option places the key below the X axis. The POS=INSET option places the key inside the plotting region and attempts to choose the least crowded corner. The default is POS=BOTTOM.

MARKERS=ANALYSIS | COMPUTED | NICE | NONE

>specifies the locations for plotting symbols.

>The MARKERS=ANALYSIS option places plotting symbols at locations corresponding to the values of the relevant input parameter from the POWER statement preceding the PLOT statement.

>The MARKERS=COMPUTED option (the default) places plotting symbols at the locations of actual computed points from the sample size analysis.

>The MARKERS=NICE option places plotting symbols at tick mark locations (corresponding to the argument axis).

>The MARKERS=NONE option disables plotting symbols.

MAX=*number* **| DATAMAX**

>specifies the maximum of the range of values for the parameter associated with the "argument" axis (the axis that is *not* representing the parameter being solved for). The default is DATAMAX, which

specifies the maximum value that occurs for this parameter in the POWER statement that precedes the PLOT statement.

MIN=number | **DATAMIN**

specifies the minimum of the range of values for the parameter associated with the "argument" axis (the axis that is *not* representing the parameter being solved for). The default is DATAMIN, which specifies the minimum value that occurs for this parameter in the POWER statement that precedes the PLOT statement.

NPOINTS=number

NPTS=number

specifies the number of values for the parameter associated with the "argument" axis (the axis that is *not* representing the parameter being solved for). You cannot use the NPOINTS= and STEP= options simultaneously. The default value for typical situations is 20.

STEP=number

specifies the increment between values of the parameter associated with the "argument" axis (the axis that is *not* representing the parameter being solved for). You cannot use the STEP= and NPOINTS= options simultaneously. By default, the NPOINTS= option is used instead of the STEP= option.

VARY (*feature* **< BY** *parameter-list* **> < , . . . ,** *feature* **< BY** *parameter-list* **> >)**

specifies how plot features should be linked to varying analysis parameters. Available *features* are COLOR, LINESTYLE, PANEL, and SYMBOL. A "panel" refers to a separate plot with a heading identifying the subset of values represented in the plot.

The *parameter-list* is a list of one or more names separated by spaces. Each name must match the name of an analysis option used in the POWER statement preceding the PLOT statement, *or* one of the following keywords: SOURCE (for the tests) and DEPENDENT (for the cell means scenarios). Also, the name must be the *primary* name for the analysis option—that is, the one listed first in the syntax description.

If you omit the < BY *parameter-list* > portion for a feature, then one or more multivalued parameters from the analysis will be automatically selected for you.

X=N | POWER

specifies a plot with the requested type of parameter on the X axis and the parameter being solved for on the Y axis. When X=N, sample size is assigned to the X axis. When X=POWER, power is assigned to the X axis. You cannot use the X= and Y= options simultaneously. The default is X=POWER, unless the result parameter is power, in which case the default is X=N.

XOPTS= (*x-options* **)**

specifies plot characteristics pertaining to the X axis.

You can specify the following *x-options* in parentheses.

CROSSREF=NO | YES

specifies whether the reference lines defined by the REF= *x-option* should be crossed with a reference line on the Y axis that indicates the solution point on the curve.

REF=number-list

specifies locations for reference lines extending from the X axis across the entire plotting region.

See the section "Specifying Value Lists in the POWER Statement" on page 3357 for information about specifying the *number-list*.

Y=N | POWER

specifies a plot with the requested type of parameter on the Y axis and the parameter being solved for on the X axis. When Y=N, sample size is assigned to the Y axis. When Y=POWER, power is assigned to the Y axis. You cannot use the Y= and X= options simultaneously. By default, the X= option is used instead of the Y= option.

YOPTS= (*y-options* **)**

specifies plot characteristics pertaining to the Y axis.

You can specify the following *y-options* in parentheses.

CROSSREF=NO | YES

specifies whether the reference lines defined by the REF= *y-option* should be crossed with a reference line on the X axis that indicates the solution point on the curve.

REF=ptn*number-list*

specifies locations for reference lines extending from the Y axis across the entire plotting region. See the section "Specifying Value Lists in the POWER Statement" on page 3357 for information about specifying the *number-list*.

You can specify the following *graph-options* in the PLOT statement after a slash (/).

DESCRIPTION='*string* **'**

specifies a descriptive string of up to 40 characters that appears in the "Description" field of the graphics catalog. The description does not appear on the plots. By default, PROC GLMPOWER assigns a description either of the form "*Y* versus *X*" (for a single-panel plot) or of the form "*Y* versus *X* (*S*)," where *Y* is the parameter on the Y axis, *X* is the parameter on the X axis, and *S* is a description of the subset represented on the current panel of a multipanel plot.

NAME='*string* **'**

specifies a name of up to eight characters for the catalog entry for the plot. The default name is PLOT*n*, where *n* is the number of the plot statement within the current invocation of PROC GLMPOWER. If the name duplicates the name of an existing entry, SAS/GRAPH software adds a number to the duplicate name to create a unique entry—for example, PLOT11 and PLOT12 for the second and third panels of a multipanel plot generated in the first PLOT statement in an invocation of PROC GLMPOWER.

POWER Statement

> **POWER** < *options* > **;**

The POWER statement performs power and sample size analyses for the Type III test of each effect in the model defined by the MODEL statements and for the contrasts defined by all CONTRAST statements. The MODEL statement must appear before the POWER statement if the EFFECTS option is used in the POWER statement.

Summary of Options

Table 43.4 summarizes categories of options available in the POWER statement.

Table 43.4 Summary of Options in the POWER Statement

Task	Options
Specify effects	EFFECTS
Specify significance level	ALPHA=
Specify covariates	CORRXY=
	NCOVARIATES=
	PROPVARREDUCTION=
Specify error standard deviation	STDDEV=
Specify sample size	NTOTAL=
Specify power	POWER=
Control sample size rounding	NFRACTIONAL
Control ordering in output	OUTPUTORDER=
	DEPENDENT

Table 43.5 summarizes the valid result parameters.

Table 43.5 Summary of Result Parameters in the POWER Statement

Solve for	Syntax
Power	POWER = .
Sample size	NTOTAL = .

Dictionary of Options

ALPHA=_number-list_

specifies the level of significance of each test. The default is 0.05, corresponding to the usual 0.05 × 100% = 5% level of significance. Note that this is a test-wise significance level with the same value for all tests, not incorporating any corrections for multiple testing. See the section "Specifying Value Lists in the POWER Statement" on page 3357 for information about specifying the _number-list_.

CORRXY=_number-list_

specifies the multiple correlation (ρ) between all covariates and the response. The error standard deviation given by the STDDEV= option is consequently reduced by multiplying it by a factor of $(1 - \rho^2)^{\frac{1}{2}}$, provided that the number of covariates (as determined by the NCOVARIATES= option) is greater than zero. You cannot use the CORRXY= and the PROPVARREDUCTION= options simultaneously. See the section "Specifying Value Lists in the POWER Statement" on page 3357 for information about specifying the _number-list_.

DEPENDENT

specifies the location of the Dependent column in the output when the OUTPUTORDER=REVERSE option or OUTPUTORDER=SYNTAX option is used, according to its relative position in the POWER statement.

EFFECTS < = <(effect ... effect) >>

specifies the model effects to include in the power analysis. By default, or if the EFFECTS keyword is specified without the equal sign (=), all model effects are included. Specify EFFECTS=() to exclude all model effect tests from the power analysis. You can include main effects and interactions by using the effects notation of PROC GLM; see the section "Specification of Effects" on page 3186 in Chapter 41, "The GLM Procedure" for further details. The MODEL statement must appear before the POWER statement if the EFFECTS option is used.

NCOVARIATES=*number-list*

NCOVARIATE=*number-list*

NCOVS=*number-list*

NCOV=*number-list*

specifies the number of additional degrees of freedom to accommodate covariate effects—both class and continuous—not listed in the MODEL statement. The error degrees of freedom are consequently reduced by the value of the NCOVARIATES= option, and the error standard deviation (whose unadjusted value is provided with the STDDEV= option) is reduced according to the value of the COR-RXY= or PROPVARREDUCTION= option. See the section "Specifying Value Lists in the POWER Statement" on page 3357 for information about specifying the *number-list*.

NFRACTIONAL

NFRAC

enables fractional input and output for sample sizes. See the section "Sample Size Adjustment Options" on page 3357 for information about the ramifications of the presence (and absence) of the NFRACTIONAL option.

NTOTAL=*number-list*

specifies the sample size or requests a solution for the sample size with a missing value (NTOTAL=.). Values for the sample size must be no smaller than the model degrees of freedom (counting the covariates). See the section "Specifying Value Lists in the POWER Statement" on page 3357 for information about specifying the *number-list*.

OUTPUTORDER=INTERNAL | REVERSE | SYNTAX

controls how the input and default analysis parameters are ordered in the output. OUT-PUTORDER=INTERNAL (the default) arranges the parameters in the output according to the following order of their corresponding options:

- DEPENDENT
- EFFECTS
- weight variable (from the WEIGHT statement)
- ALPHA=
- NCOVARIATES=
- CORRXY=

- PROPVARREDUCTION=
- STDDEV=
- NTOTAL=
- POWER=

The OUTPUTORDER=SYNTAX option arranges the parameters in the output in the same order in which their corresponding options are specified in the POWER statement. The OUTPUTORDER=REVERSE option arranges the parameters in the output in the reverse of the order in which their corresponding options are specified in the POWER statement.

POWER=_number-list_

specifies the desired power of each test or requests a solution for the power with a missing value (POWER=.). The power is expressed as a probability (for example, 0.9) rather than a percentage. Note that this is a test-wise power with the same value for all tests, without any correction for multiple testing. See the section "Specifying Value Lists in the POWER Statement" on page 3357 for information about specifying the _number-list_.

PROPVARREDUCTION=_number-list_
PVRED=_number-list_

specifies the proportional reduction (r) in total R^2 incurred by the covariates—in other words, the amount of additional variation explained by the covariates. The error standard deviation given by the STDDEV= option is consequently reduced by multiplying it by a factor of $(1 - r)^{\frac{1}{2}}$, provided that the number of covariates (as determined by the NCOVARIATES= option) is greater than zero. You cannot use the PROPVARREDUCTION= and the CORRXY= options simultaneously. See the section "Specifying Value Lists in the POWER Statement" on page 3357 for information about specifying the _number-list_.

STDDEV=_number-list_

specifies the error standard deviation, or root MSE. If covariates are specified using the NCOVARIATES= option, then the STDDEV= option denotes the error standard deviation before accounting for these covariates. See the section "Specifying Value Lists in the POWER Statement" on page 3357 for information about specifying the _number-list_.

Restrictions on Option Combinations

For the relationship between covariates and response, specify either the multiple correlation (by using the CORRXY= option) or the proportional reduction in total R^2 (by using the PROPVARREDUCTION= option).

WEIGHT Statement

WEIGHT _variable_ **;**

The WEIGHT statement names a variable that provides a profile weight ("cell weight") for each observation in the exemplary data set specified by the DATA= option in the PROC GLMPOWER statement.

If the WEIGHT statement is not used, then a balanced design is assumed with default cell weights of 1.

Details: GLMPOWER Procedure

Specifying Value Lists in the POWER Statement

To specify one or more scenarios for an analysis parameter (or set of parameters) in the POWER statement, you provide a list of values for the option that corresponds to the parameter(s). To identify the parameter you want to solve for, you place a missing value in the appropriate list.

Scenarios for scalar-valued parameters, such as power, are represented by a *number-list*.

Number-Lists

A *number-list* can be one of two things: a series of one or more numbers expressed in the form of one or more DOLISTs, or a missing value indicator (.).

The DOLIST format is the same as in the DATA step. For example, you can specify four scenarios (30, 50, 70, and 100) for a total sample size in either of the following ways:

```
NTOTAL = 30 50 70 100
NTOTAL = 30 to 70 by 20 100
```

A missing value identifies a parameter as the result parameter; it is valid only with options representing parameters you can solve for in a given analysis. For example, you can request a solution for NTOTAL:

```
NTOTAL = .
```

Sample Size Adjustment Options

By default, PROC GLMPOWER rounds sample sizes conservatively (down in the input, up in the output) so that all total sizes *and* sample sizes for individual design profiles are integers. This is generally considered conservative because it selects the closest realistic design providing *at most* the power of the (possibly fractional) input or mathematically optimized design. In addition, all design profile sizes are adjusted to be multiples of their corresponding weights. If a design profile is present more than once in the exemplary data set, then the weights for that design profile are summed. For example, if a particular design profile is present twice in the exemplary data set with weight values 2 and 6, then all sample sizes for this design profile become multiples of $2 + 6 = 8$.

With the NFRACTIONAL option, sample size input is not rounded, and sample size output is reported in two versions, a raw "fractional" version and a "ceiling" version rounded up to the nearest integer.

Whenever an input sample size is adjusted, both the original ("nominal") and adjusted ("actual") sample sizes are reported. Whenever computed output sample sizes are adjusted, both the original input ("nominal") power and the achieved ("actual") power at the adjusted sample size are reported.

Error and Information Output

The Error column in the main output table explains reasons for missing results and flags numerical results that are bounds rather than exact answers.

The Info column provides further information about Error entries, warnings about any boundary conditions detected, and notes about any adjustments to input. Note that the Info column is hidden by default in the main output. You can view it by using the ODS OUTPUT statement to save the output as a data set and the PRINT procedure. For example, the following SAS statements print both the Error and Info columns for a power computation in a one-way ANOVA:

```
data MyExemp;
   input A $ Y1 Y2;
   datalines;
        1   10 11
        2   12 11
        3   15 11
   ;
run;

proc glmpower data=MyExemp;
   class A;
   model Y1 Y2 = A;
   power
      stddev = 2
      ntotal = 3 10
      power  = .;
   ods output output=Power;
run;

proc print noobs data=Power;
   var NominalNTotal NTotal Dependent Power Error Info;
run;
```

The output is shown in Figure 43.5.

Figure 43.5 Error and Information Columns

Nominal NTotal	NTotal	Dependent	Power	Error	Info
3	3	Y1	.	Invalid input	Error DF=0
10	9	Y1	0.557		Input N adjusted
3	3	Y2	.	Invalid input	Error DF=0 / No effect
10	9	Y2	0.050		Input N adjusted / No effect

The sample size of 3 specified with the NTOTAL= option causes an "Invalid input" message in the Error column and an "Error DF=0" message in the Info column, because a sample size of 3 is so small that there are no degrees of freedom left for the error term. The sample size of 10 causes an "Input N adjusted" message in the Info column, because it is rounded down to 9 to produce integer group sizes of 3 per cell. The cell means scenario represented by the dependent variable Y2 causes a "No effect" message to appear in the Info column, because the means in this scenario are all equal.

Displayed Output

If you use the PLOTONLY option in the PROC GLMPOWER statement, the procedure displays only graph-ical output. Otherwise, the displayed output of the GLMPOWER procedure includes the following:

- the "Fixed Scenario Elements" table, which shows all applicable single-valued analysis parameters, in the following order: the dependent variable representing the cell means, the source of the test, the weight variable, parameters input explicitly, parameters supplied with defaults, and ancillary results

- an output table showing the following when applicable (in order): the index of the scenario, the de-pendent variable representing the cell means, the type of the test, the source of the test, all multivalued input, ancillary results, the primary computed result, and error descriptions

- plots (if requested)

The exception to these ordering conventions is that the DEPENDENT and EFFECTS options may be used along with the OUTPUTORDER=SYNTAX or OUTPUTORDER=REVERSE option in the POWER state-ment to specify the relative location of the output for dependent variable and type and source of test.

Ancillary results include the following:

- Actual Power, the achieved power, if it differs from the input (Nominal) power value

- fractional sample size, if the NFRACTIONAL option is used in the POWER statement

If sample size is the result parameter and the NFRACTIONAL option is used in the POWER statement, then both "Fractional" and "Ceiling" sample size results are displayed. Fractional sample sizes correspond to the "Nominal" values of power. Ceiling sample sizes are simply the fractional sample sizes rounded up to the nearest integer; they correspond to "Actual" values of power.

The noncentrality parameter is computed and stored in a hidden column called Noncentrality in the "Output" table.

ODS Table Names

PROC GLMPOWER assigns a name to each table that it creates. You can use these names to reference the table when using the Output Delivery System (ODS) to select tables and create output data sets. These names are listed in Table 43.6. For more information about ODS, see Chapter 20, "Using the Output Delivery System."

Table 43.6 ODS Tables Produced by PROC GLMPOWER

ODS Table Name	Description	Statement
FixedElements	Factoid with single-valued analysis parameters	Default

Table 43.6 *continued*

ODS Table Name	Description	Statement
Output	All input and computed analysis parameters, error messages, and information messages for each scenario	Default
PlotContent	Data contained in plots, including analysis parameters and indices identifying plot features. (**NOTE:** This table is saved as a data set and not displayed in PROC GLMPOWER output.)	PLOT

Computational Methods and Formulas

This section describes the approaches used in PROC GLMPOWER to compute power and sample size.

Contrasts in Fixed-Effect Univariate Models

The univariate linear model has the form

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

where $\mathbf{y}$ is the $N \times 1$ vector of responses, $\mathbf{X}$ is the $N \times p$ design matrix, $\boldsymbol{\beta}$ is the $p \times 1$ vector of model parameters corresponding to the columns of $\mathbf{X}$, and $\boldsymbol{\epsilon}$ is an $N \times 1$ vector of errors with

$$\epsilon_1, \ldots, \epsilon_N \sim \mathrm{N}(0, \sigma^2) \quad \text{(i.i.d.)}$$

In PROC GLMPOWER, the model parameters $\boldsymbol{\beta}$ are not specified directly, but rather indirectly as $\mathbf{y}^\star$, which represents either conjectured response means or typical response values for each design profile. The $\mathbf{y}^\star$ values are manifested as the dependent variable in the MODEL statement. The vector $\boldsymbol{\beta}$ is obtained from $\mathbf{y}^\star$ according to the least squares equation,

$$\boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}^\star$$

Note that, in general, there is not a 1-to-1 mapping between $\mathbf{y}^\star$ and $\boldsymbol{\beta}$. Many different scenarios for $\mathbf{y}^\star$ might lead to the same $\boldsymbol{\beta}$. If you specify $\mathbf{y}^\star$ with the intention of representing cell means, keep in mind that PROC GLMPOWER allows scenarios that are *not* valid cell means according to the model specified in the MODEL statement. For example, if $\mathbf{y}^\star$ exhibits an interaction effect but the corresponding interaction term is left out of the model, then the cell means ($\mathbf{X}\boldsymbol{\beta}$) derived from $\boldsymbol{\beta}$ differ from $\mathbf{y}^\star$. In particular, the cell means thus derived are the projection of $\mathbf{y}^\star$ onto the model space.

It is convenient in power analysis to parameterize the design matrix $\mathbf{X}$ in three parts, $\{\ddot{\mathbf{X}}, \mathbf{w}, N\}$, defined as follows:

1. The $q \times p$ essence design matrix $\ddot{\mathbf{X}}$ is the collection of unique rows of $\mathbf{X}$. Its rows are sometimes referred to as "design profiles." Here, $q \leq N$ is defined simply as the number of unique rows of $\mathbf{X}$.

2. The $q \times 1$ weight vector $\mathbf{w}$ reveals the relative proportions of design profiles. Row i of $\ddot{\mathbf{X}}$ is to be included in the design w_i times for every w_j times row j is included. The weights are assumed to be standardized (that is, sum up to 1).

3. The total sample size is N. This is the number of rows in $\mathbf{X}$. If you gather $Nw_i = n_i$ copies of the ith row of $\ddot{\mathbf{X}}$, for $i = 1, \ldots, q$, then you end up with $\mathbf{X}$.

It is useful to express the crossproduct matrix $\mathbf{X'X}$ in terms of these three parts,

$$\mathbf{X'X} = N\ddot{\mathbf{X}}'\text{diag}(\mathbf{w})\ddot{\mathbf{X}}$$

since this factors out the portion (N) depending on sample size and the portion $(\ddot{\mathbf{X}}'\text{diag}(\mathbf{w})\ddot{\mathbf{X}})$ depending only on the design structure.

A general linear hypothesis for the univariate model has the form

$$H_0{:}\mathbf{L}\boldsymbol{\beta} = \boldsymbol{\theta}_0$$
$$H_A{:}\mathbf{L}\boldsymbol{\beta} \neq \boldsymbol{\theta}_0$$

where $\mathbf{L}$ is an $r_L \times p$ contrast matrix (assumed to be full rank) and $\boldsymbol{\theta}_0$ is the null value (usually just a vector of zeros). Note that effect tests are just contrasts that use special forms of $\mathbf{L}$. Thus, this scheme covers both effect tests and custom contrasts.

The test statistic is

$$F = \frac{\left(\frac{\text{SS}_\text{H}}{r_L}\right)}{\hat{\sigma}^2}$$

where

$$\text{SS}_\text{H} = \frac{1}{N}\left(\mathbf{L}\hat{\boldsymbol{\beta}} - \boldsymbol{\theta}_0\right)'\left(\mathbf{L}\left(\mathbf{X'X}\right)^{-1}\mathbf{L'}\right)^{-1}\left(\mathbf{L}\hat{\boldsymbol{\beta}} - \boldsymbol{\theta}_0\right)$$
$$\hat{\boldsymbol{\beta}} = (\mathbf{X'X})^{-1}\mathbf{X'y}$$
$$\hat{\sigma}^2 = \frac{1}{\text{DF}_\text{E}}\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)'\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)$$

where $\text{DF}_\text{E} = N - \text{rank}(\mathbf{X})$. Note that $\text{DF}_\text{E} = N - p$ if $\mathbf{X}$ has full rank.

Under H_0, $F \sim F(r_L, \text{DF}_\text{E})$. Under H_A, F is distributed as $F(r_L, \text{DF}_\text{E}, \lambda)$ with noncentrality

$$\lambda = N\,(\mathbf{L}\boldsymbol{\beta} - \boldsymbol{\theta}_0)'\left(\mathbf{L}\left(\ddot{\mathbf{X}}'\text{diag}(\mathbf{w})\ddot{\mathbf{X}}\right)^{-1}\mathbf{L'}\right)^{-1}(\mathbf{L}\boldsymbol{\beta} - \boldsymbol{\theta}_0)\,\sigma^{-2}$$

Muller and Peterson (1984) give the exact power of the test as

$$\text{power} = P\left(F(r_L, \text{DF}_\text{E}, \lambda) \geq F_{1-\alpha}(r_L, \text{DF}_\text{E})\right)$$

Sample size is computed by inverting the power equation.

See Muller et al. (1992) and O'Brien and Shieh (1992) for additional discussion.

Adjustments for Covariates

If you specify covariates in the model (whether continuous or categorical), then two adjustments are made in order to compute approximate power in the presence of the covariates. Let n_v denote the number of covariates (counting dummy variables for categorical covariates individually). In other words, n_v is the total degrees of freedom used by the covariates. The adjustments are as follows:

1. The error degrees of freedom decrease by n_v.

2. The error standard deviation σ shrinks by a factor of $(1-\rho^2)^{\frac{1}{2}}$ (if the CORRXY= option to specify the correlation ρ between covariates and response) or $(1-r)^{\frac{1}{2}}$ (if the PROPVARREDUCTION= option is used to specify the proportional reduction in total R^2 incurred by the covariates). Let $\sigma^\star$ represent the updated value of σ.

As a result of these changes, the power is computed as

$$\text{power} = P\left(F(r_L, \text{DF}_E - n_v, \lambda^\star) \geq F_{1-\alpha}(r_L, N - r_x - n_v)\right)$$

where $\lambda^\star$ is calculated using $\sigma^\star$ rather than σ:

$$\lambda^\star = N\,(\mathbf{L}\boldsymbol{\beta} - \boldsymbol{\theta}_0)'\left(\mathbf{L}\left(\ddot{\mathbf{X}}'\text{diag}(\mathbf{w})\ddot{\mathbf{X}}\right)^{-1}\mathbf{L}'\right)^{-1}(\mathbf{L}\boldsymbol{\beta} - \boldsymbol{\theta}_0)\,(\sigma^\star)^{-2}$$

ODS Graphics

Statistical procedures use ODS Graphics to create graphs as part of their output. ODS Graphics is described in detail in Chapter 21, "Statistical Graphics Using ODS."

Before you create graphs, ODS Graphics must be enabled (for example, with the ODS GRAPHICS ON statement). For more information about enabling and disabling ODS Graphics, see the section "Enabling and Disabling ODS Graphics" on page 609 in Chapter 21, "Statistical Graphics Using ODS."

The overall appearance of graphs is controlled by ODS styles. Styles and other aspects of using ODS Graphics are discussed in the section "A Primer on ODS Statistical Graphics" on page 608 in Chapter 21, "Statistical Graphics Using ODS."

If ODS Graphics is not enabled, then PROC GLMPOWER creates traditional graphics.

You can reference every graph produced through ODS Graphics with a name. The names of the graphs that PROC GLMPOWER generates are listed in Table 43.7, along with the required statements and options.

Table 43.7 Graphs Produced by PROC GLMPOWER

ODS Graph Name	Plot Description	Option
PowerPlot	Plot with power and sample size on the axes	PLOT
PowerAbort	Empty plot that shows an error message when a plot could not be produced	PLOT

ODS Styles Suitable for Use with PROC GLMPOWER

ODS styles control the appearance of graphs produced by PROC GLMPOWER. ODS provides over 50 styles, but most are not suitable for use in PROC GLMPOWER. PROC GLMPOWER requires a style that distinguishes curves based on a combination of color, line style, and symbol marker. Styles that are well-suited for use in PROC GLMPOWER include: STATISTICAL, ANALYSIS, DEFAULT, LISTING, and HTMLBLUECML. The HTMLBLUE and PLATEAU styles are commonly used, but they are not well-suited for use with PROC GLMPOWER because they rely primarily on color to distinguish curves rather than a combination of color, line style, and symbol marker.

In this chapter, a style is explicitly specified at the start of each example that uses ODS Graphics to remind you to use one of the suitable styles. Styles are specified in an ODS destination statement. Destinations include LISTING, HTML, RTF, PDF, and many others. You can set the style and the destination as follows:

```
ods html style=htmlbluecml;
ods graphics on;

proc glmpower data=Exemplary;
   class Variety Exposure;
   model Height = Variety | Exposure;
   power
      stddev = 4 6.5
      ntotal = 60
      power  = .;
   plot x=n min=30 max=90;
run;

ods graphics off;
ods html close;
```

For more information about ODS and ODS destinations, see Chapter 20, "Using the Output Delivery System." For more information ODS styles, see Chapter 21, "Statistical Graphics Using ODS."

Examples: GLMPOWER Procedure

Example 43.1: One-Way ANOVA

This example deals with the same situation as in Example 70.1 in Chapter 70, "The POWER Procedure."

Hocking (1985, p. 109) describes a study of the effectiveness of electrolytes in reducing lactic acid buildup for long-distance runners. You are planning a similar study in which you will allocate five different fluids to runners on a 10-mile course and measure lactic acid buildup immediately after the race. The fluids consist of water and two commercial electrolyte drinks, EZDure and LactoZap, each prepared at two concentrations, low (EZD1 and LZ1) and high (EZD2 and LZ2).

You conjecture that the standard deviation of lactic acid measurements given any particular fluid is about 3.75, and that the expected lactic acid values will correspond roughly to Table 43.8. You are least familiar with the LZ1 drink and hence decide to consider a range of reasonable values for that mean.

Table 43.8 Mean Lactic Acid Buildup by Fluid

Water	EZD1	EZD2	LZ1	LZ2
35.6	33.7	30.2	29 or 28	25.9

You are interested in four different comparisons, shown in Table 43.9 with appropriate contrast coefficients.

Table 43.9 Planned Comparisons

	Contrast Coefficients				
Comparison	Water	EZD1	EZD2	LZ1	LZ2
Water versus electrolytes	4	-1	-1	-1	-1
EZD versus LZ	0	1	1	-1	-1
EZD1 versus EZD2	0	1	-1	0	0
LZ1 versus LZ2	0	0	0	1	-1

For each of these contrasts you want to determine the sample size required to achieve a power of 0.9 for detecting an effect with magnitude in accord with Table 43.8. You are not yet attempting to choose a single sample size for the study, but rather checking the range of sample sizes needed for individual contrasts. You plan to test each contrast at $\alpha = 0.025$. In the interests of reducing costs, you will provide twice as many runners with water as with any of the electrolytes; that is, you will use a sample size weighting scheme of 2:1:1:1:1.

Before calling PROC GLMPOWER, you need to create the *exemplary data set* to specify means and weights for the design profiles:

```
data Fluids;
   input Fluid $ LacticAcid1 LacticAcid2 CellWgt;
   datalines;
        Water      35.6        35.6        2
        EZD1       33.7        33.7        1
        EZD2       30.2        30.2        1
        LZ1        29          28          1
        LZ2        25.9        25.9        1
   ;
run;
```

The variable LacticAcid1 represents the cell means scenario with the larger LZ1 mean (29), and LacticAcid2 represents the scenario with the smaller LZ1 mean (28). The variable CellWgt contains the sample size allocation weights.

Use the DATA= option in the PROC GLMPOWER statement to specify Fluids as the exemplary data set. The following statements perform the sample size analysis:

```
proc glmpower data=Fluids;
   class Fluid;
   model LacticAcid1 LacticAcid2 = Fluid;
   weight CellWgt;
   contrast "Water vs. others" Fluid  -1 -1 -1 -1 4;
   contrast "EZD vs. LZ"       Fluid   1  1 -1 -1 0;
   contrast "EZD1 vs. EZD2"    Fluid   1 -1  0  0 0;
   contrast "LZ1 vs. LZ2"      Fluid   0  0  1 -1 0;
   power
      stddev = 3.75
      alpha  = 0.025
      ntotal = .
      power  = 0.9;
run;
```

The CLASS statement identifies Fluid as a classification variable. The MODEL statement specifies the model and the two cell means scenarios LacticAcid1 and LacticAcid2. The WEIGHT statement identifies CellWgt as the weight variable. The CONTRAST statement specifies the contrasts. Since PROC GLM-POWER by default processes class levels in order of formatted values, the contrast coefficients correspond to the following order: EZD1, EZD2, LZ1, LZ2, Water. (**NOTE:** You could use the ORDER=DATA option in the PROC GLMPOWER statement to achieve the same ordering as in Table 43.9 instead.) The POWER statement specifies total sample size as the result parameter and provides values for the other analysis parameters (error standard deviation, alpha, and power).

Output 43.1.1 displays the results.

Output 43.1.1 Sample Sizes for One-Way ANOVA Contrasts

```
                        The GLMPOWER Procedure

                       Fixed Scenario Elements

           Weight Variable                 CellWgt
           Alpha                             0.025
           Error Standard Deviation           3.75
           Nominal Power                       0.9


                       Computed N Total


                                           Test  Error  Actual      N
    Index   Dependent    Type       Source   DF     DF   Power  Total

        1  LacticAcid1  Effect    Fluid         4     25  0.958     30
        2  LacticAcid1  Contrast  Water vs. others  1     25  0.947     30
        3  LacticAcid1  Contrast  EZD vs. LZ        1     55  0.929     60
        4  LacticAcid1  Contrast  EZD1 vs. EZD2     1    169  0.901    174
        5  LacticAcid1  Contrast  LZ1 vs. LZ2       1    217  0.902    222
        6  LacticAcid2  Effect    Fluid         4     25  0.972     30
        7  LacticAcid2  Contrast  Water vs. others  1     19  0.901     24
        8  LacticAcid2  Contrast  EZD vs. LZ        1     43  0.922     48
        9  LacticAcid2  Contrast  EZD1 vs. EZD2     1    169  0.901    174
       10  LacticAcid2  Contrast  LZ1 vs. LZ2       1    475  0.902    480
```

The sample sizes range from 24 for the comparison of water versus electrolytes to 480 for the comparison of LZ1 versus LZ2, both assuming the smaller LZ1 mean. The sample size for the latter comparison is relatively large because the small mean difference of $28 - 25.9 = 2.1$ is hard to detect. PROC GLMPOWER also includes the effect test for Fluid. Note that, in this case, it is equivalent to TEST=OVERALL_F in the ONEWAYANOVA statement of PROC POWER, since there is only one effect in the model.

The Nominal Power of 0.9 in the "Fixed Scenario Elements" table in Output 43.1.1 represents the input target power, and the Actual Power column in the "Computed N Total" table is the power at the sample size (N Total) adjusted to achieve the specified sample weighting. Note that all of the sample sizes are rounded up to multiples of 6 to preserve integer group sizes (since the group weights add up to 6). You can use the NFRACTIONAL option in the POWER statement to compute raw fractional sample sizes.

Suppose you want to plot the required sample size for the range of power values from 0.5 to 0.95. First, define the analysis by specifying the same statements as before, but add the PLOTONLY option to the PROC GLMPOWER statement to disable the nongraphical results. Next, specify the PLOT statement with X=POWER to request a plot with power on the X axis. (The result parameter—here sample size—is always plotted on the other axis.) Use the MIN= and MAX= options in the PLOT statement to specify the power range. The following statements produce the plot:

```
ods listing style=htmlbluecml;
ods graphics on;

proc glmpower data=Fluids plotonly;
   class Fluid;
   model LacticAcid1 LacticAcid2 = Fluid;
   weight CellWgt;
   contrast "Water vs. others" Fluid  -1 -1 -1 -1 4;
   contrast "EZD vs. LZ"       Fluid   1  1 -1 -1 0;
   contrast "EZD1 vs. EZD2"    Fluid   1 -1  0  0 0;
   contrast "LZ1 vs. LZ2"      Fluid   0  0  1 -1 0;
   power
      stddev = 3.75
      alpha  = 0.025
      ntotal = .
      power  = 0.9;
   plot x=power min=.5 max=.95;
run;
```

The ODS LISTING STYLE=HTMLBLUECML statement specifies the HTMLBLUECML style, which is suitable for use with PROC GLMPOWER because it allows both marker symbols and line styles to vary. See the section "ODS Styles Suitable for Use with PROC GLMPOWER" on page 3363 for more information. See Output 43.1.2 for the resulting plot.

Output 43.1.2 Plot of Sample Size versus Power for One-Way ANOVA Contrasts
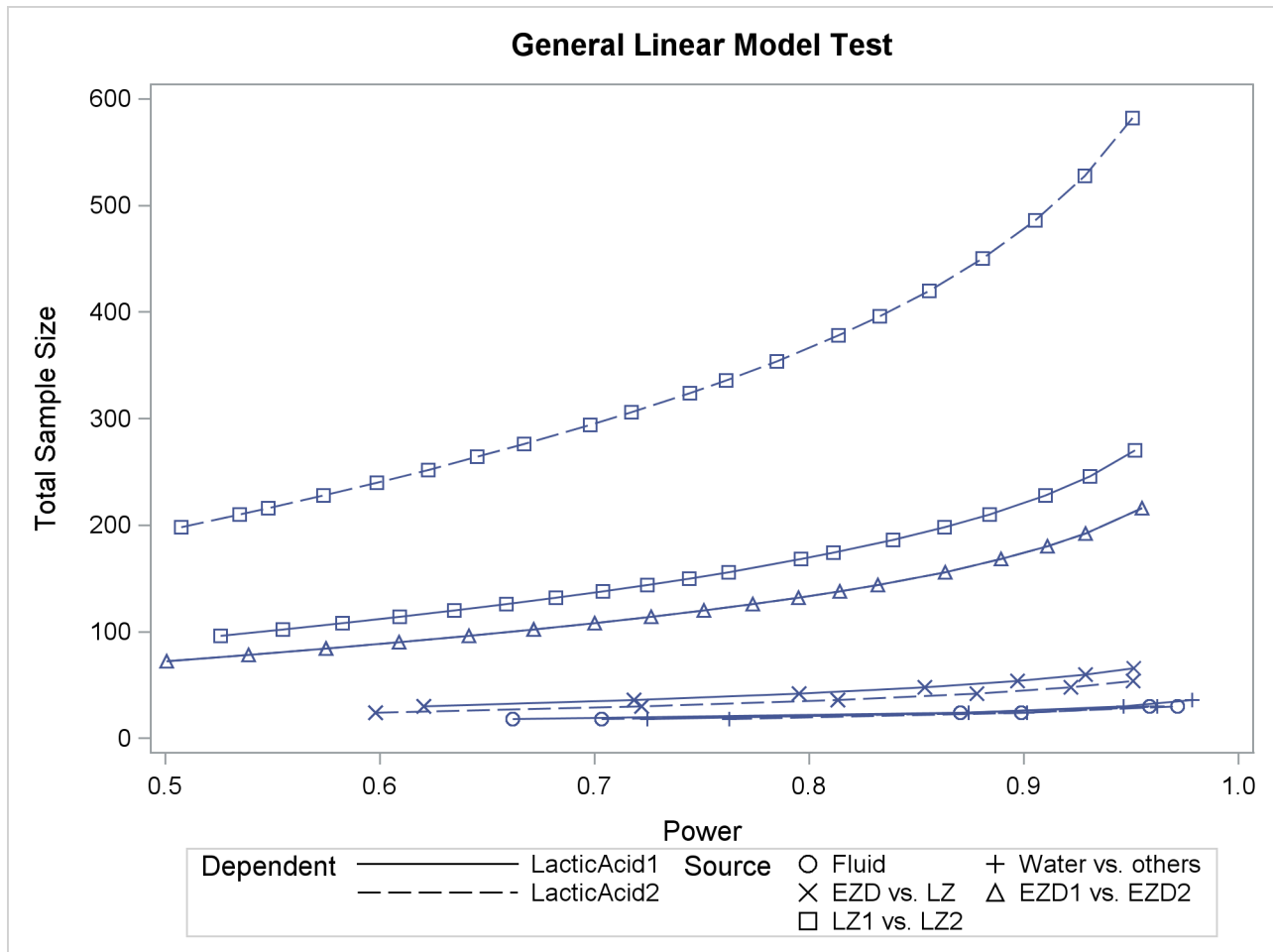
In Output 43.1.2, the line style identifies the cell means scenario, and the plotting symbol identifies the test. The plotting symbol locations identify actual computed powers; the curves are linear interpolations of these points. The plot shows that the required sample size is highest for the test of LZ1 versus LZ2, which was previously found to require the most resources.

Note that some of the plotted points in Output 43.1.2 are unevenly spaced. This is because the plotted points are the *rounded* sample size results at their corresponding *actual* power levels. The range specified with the MIN= and MAX= values in the PLOT statement corresponds to *nominal* power levels. In some cases, actual power is substantially higher than nominal power. To obtain plots with evenly spaced points (but with *fractional* sample sizes at the computed points), you can use the NFRACTIONAL option in the POWER statement preceding the PLOT statement.

Finally, suppose you want to plot the power for the range of sample sizes you will likely consider for the study (the range of 24 to 480 that achieves 0.9 power for different comparisons). In the POWER statement, identify power as the result (POWER=.), and specify any total sample size value (say, NTOTAL=100). Specify the PLOT statement with X=N to request a plot with sample size on the X axis.

The following statements produce the plot:

```
proc glmpower data=Fluids plotonly;
   class Fluid;
   model LacticAcid1 LacticAcid2 = Fluid;
   weight CellWgt;
   contrast "Water vs. others" Fluid  -1 -1 -1 -1 4;
   contrast "EZD vs. LZ"       Fluid   1  1 -1 -1 0;
   contrast "EZD1 vs. EZD2"    Fluid   1 -1  0  0 0;
   contrast "LZ1 vs. LZ2"      Fluid   0  0  1 -1 0;
   power
      stddev = 3.75
      alpha  = 0.025
      ntotal = 24
      power  = .;
   plot x=n min=24 max=480;
run;

ods graphics off;
```

Note that the value 100 specified with the NTOTAL=100 option is not used. It is overridden in the plot by the MIN= and MAX= options in the PLOT statement, and the PLOTONLY option in the PROC GLMPOWER statement disables nongraphical results. But the NTOTAL= option (along with a value) is still needed in the POWER statement as a placeholder, to identify the desired parameterization for sample size.

See Output 43.1.3 for the plot.

Output 43.1.3 Plot of Power versus Sample Size for One-Way ANOVA Contrasts
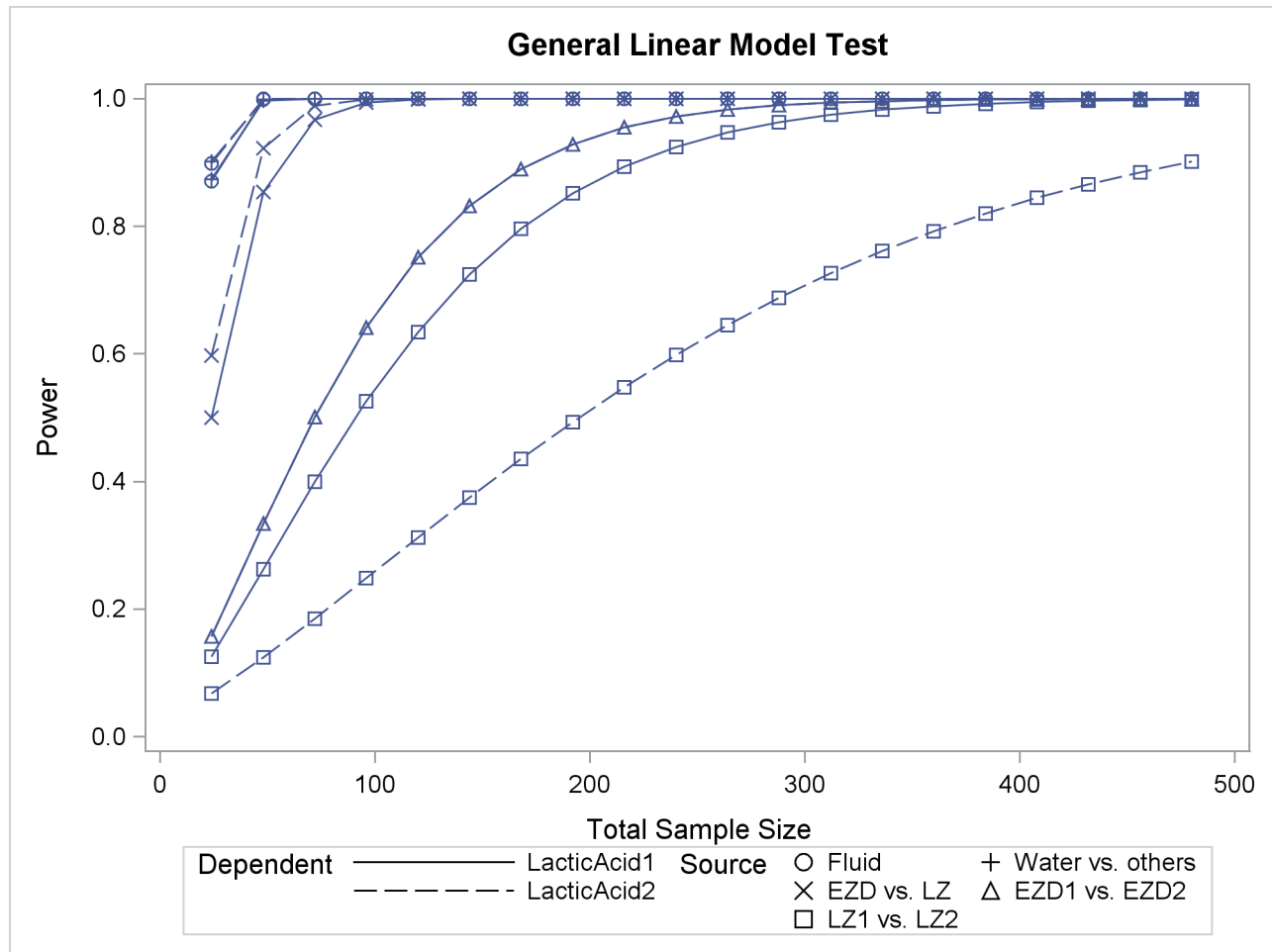
Although Output 43.1.2 and Output 43.1.3 surface essentially the same computations for practical power ranges, they each provide a different quick visual assessment. Output 43.1.2 reveals the range of required sample sizes for powers of interest, and Output 43.1.3 reveals the range of achieved powers for sample sizes of interest.

Example 43.2: Two-Way ANOVA with Covariate

Suppose you can enhance the planned study discussed in Example 43.1 in two ways:

- incorporate results from races at two different altitudes ("high" and "low")

- measure the body mass index of each runner before the race

This is equivalent to adding a second fixed effect and a continuous covariate to your model.

Since lactic acid buildup is more pronounced at higher altitudes, you will include altitude as a factor in the model along with fluid, extending the one-way ANOVA to a two-way ANOVA. In doing so, you expect to lower the residual standard deviation from about 3.75 to 3.5 (in addition to generalizing the study results). You assume there is negligible interaction between fluid and altitude and plan to use a main-effects-only model. You conjecture that the mean lactic acid buildup follows Table 43.10.

Table 43.10 Mean Lactic Acid Buildup by Fluid and Altitude

| Altitude | Fluid | | | | |
	Water	EZD1	EZD2	LZ1	LZ2
High	36.9	35.0	31.5	30	27.1
Low	34.3	32.4	28.9	27	24.7

By including a measurement of body mass index as a covariate in the study, you hope to further reduce the error variability. The extent of this reduction in variability is commonly expressed in two alternative ways: (1) the correlation between the covariates and the response or (2) the proportional reduction in total R^2 incurred by the covariates. You prefer the former and guess that the correlation between body mass index and lactic acid buildup is between 0.2 and 0.3. You specify these estimates with the NCOVARIATES= and CORRXY= options in the POWER statement. The covariate is not included in the MODEL statement.

You are interested in the same four fluid comparisons as in Example 43.1, shown in Table 43.9, except this time you want to marginalize over the effect of altitude.

For each of these contrasts, you want to determine the sample size required to achieve a power of 0.9 to detect an effect with magnitude according to Table 43.10. You are not yet attempting to choose a single sample size for the study, but rather checking the range of sample sizes needed by individual contrasts. You plan to test each contrast at $\alpha = 0.025$. You will provide twice as many runners with water as with any of the electrolytes, and you predict that you can study approximately two-thirds as many runners at high altitude than at low altitude. The resulting planned sample size weighting scheme is shown in Table 43.11. Since the scheme is only approximate, you use the NFRACTIONAL option in the POWER statement to disable the rounding of sample sizes up to integers satisfying the weights exactly.

Table 43.11 Approximate Sample Size Allocation Weights

| Altitude | Fluid | | | | |
	Water	EZD1	EZD2	LZ1	LZ2
High	4	2	2	2	2
Low	6	3	3	3	3

First, you create the exemplary data set to specify means and weights for the design profiles:

```
data Fluids2;
   input Altitude $ Fluid $ LacticAcid CellWgt;
   datalines;
         High        Water       36.9       4
         High        EZD1        35.0       2
         High        EZD2        31.5       2
         High        LZ1         30         2
         High        LZ2         27.1       2
```

```
        Low          Water        34.3       6
        Low          EZD1         32.4       3
        Low          EZD2         28.9       3
        Low          LZ1          27         3
        Low          LZ2          24.7       3
     ;
  run;
```

The variables Altitude, Fluid, and LacticAcid specify the factors and cell means in Table 43.10. The variable CellWgt contains the sample size allocation weights in Table 43.11.

Use the DATA= option in the PROC GLMPOWER statement to specify Fluids2 as the exemplary data set. The following statements perform the sample size analysis:

```
proc glmpower data=Fluids2;
   class Altitude Fluid;
   model LacticAcid = Altitude Fluid;
   weight CellWgt;
   contrast "Water vs. others" Fluid  -1 -1 -1 -1 4;
   contrast "EZD vs. LZ"       Fluid   1  1 -1 -1 0;
   contrast "EZD1 vs. EZD2"    Fluid   1 -1  0  0 0;
   contrast "LZ1 vs. LZ2"      Fluid   0  0  1 -1 0;
   power
      nfractional
      stddev      = 3.5
      ncovariates = 1
      corrxy      = 0.2 0.3 0
      alpha       = 0.025
      ntotal      = .
      power       = 0.9;
  run;
```

The CLASS statement identifies Altitude and Fluid as classification variables. The MODEL statement specifies the model, and the WEIGHT statement identifies CellWgt as the weight variable. The CONTRAST statement specifies the contrasts in Table 43.9. As in Example 43.1, the order of the contrast coefficients corresponds to the formatted class levels (EZD1, EZD2, LZ1, LZ2, Water). The POWER statement specifies total sample size as the result parameter and provides values for the other analysis parameters. The NCOVARIATES= option specifies the single covariate (body mass index), and the CORRXY= option specifies the two scenarios for its correlation with lactic acid buildup (0.2 and 0.3). Output 43.2.1 displays the results.

Output 43.2.1 Sample Sizes for Two-Way ANOVA Contrasts

```
                    The GLMPOWER Procedure


                    Fixed Scenario Elements


        Dependent Variable                      LacticAcid
        Weight Variable                            CellWgt
        Alpha                                        0.025
        Number of Covariates                             1
        Std Dev Without Covariate Adjustment           3.5
        Nominal Power                                  0.9
```

Output 43.2.1 *continued*

```
                        Computed Ceiling N Total

                                      Adj
                               Corr   Std  Test  Error    Fractional
     Index    Type      Source   XY   Dev   DF    DF       N Total

        1  Effect   Altitude       0.2   3.43   1    84    90.418451
        2  Effect   Altitude       0.3   3.34   1    79    85.862649
        3  Effect   Altitude       0.0   3.50   1    88    94.063984
        4  Effect   Fluid          0.2   3.43   4    16    22.446173
        5  Effect   Fluid          0.3   3.34   4    15    21.687544
        6  Effect   Fluid          0.0   3.50   4    17    23.055716
        7  Contrast Water vs. others 0.2 3.43   1    15    21.720195
        8  Contrast Water vs. others 0.3 3.34   1    14    20.848805
        9  Contrast Water vs. others 0.0 3.50   1    16    22.422381
       10  Contrast EZD vs. LZ      0.2   3.43   1    35    41.657424
       11  Contrast EZD vs. LZ      0.3   3.34   1    33    39.674037
       12  Contrast EZD vs. LZ      0.0   3.50   1    37    43.246415
       13  Contrast EZD1 vs. EZD2   0.2   3.43   1   139   145.613657
       14  Contrast EZD1 vs. EZD2   0.3   3.34   1   132   138.173983
       15  Contrast EZD1 vs. EZD2   0.0   3.50   1   145   151.565917
       16  Contrast LZ1 vs. LZ2     0.2   3.43   1   268   274.055008
       17  Contrast LZ1 vs. LZ2     0.3   3.34   1   253   259.919126
       18  Contrast LZ1 vs. LZ2     0.0   3.50   1   279   285.363976
```

```
                   Computed Ceiling N Total

                         Actual    Ceiling
                Index    Power     N Total

                   1     0.902        91
                   2     0.901        86
                   3     0.903        95
                   4     0.912        23
                   5     0.908        22
                   6     0.919        24
                   7     0.905        22
                   8     0.903        21
                   9     0.910        23
                  10     0.903        42
                  11     0.903        40
                  12     0.906        44
                  13     0.901       146
                  14     0.902       139
                  15     0.901       152
                  16     0.901       275
                  17     0.900       260
                  18     0.901       286
```

The sample sizes in Output 43.2.1 range from 21 for the comparison of water versus electrolytes (assuming a correlation of 0.3 between body mass and lactic acid buildup) to 275 for the comparison of LZ1 versus LZ2 (assuming a correlation of 0.2). PROC GLMPOWER also includes the effect tests for Altitude and Fluid. Note that the required sample sizes for this study are lower than those for the study in Example 43.1.

Note that the error standard deviation has been reduced from 3.5 to 3.43 (when correlation is 0.2) or 3.34 (when correlation is 0.3) in the approximation of the effect of the body mass index covariate. The error degrees of freedom has also been automatically adjusted, lowered by 1 (the number of covariates).

Suppose you want to plot the required sample size for the range of power values from 0.5 to 0.95. First, define the analysis by specifying the same statements as before, but add the PLOTONLY option to the PROC GLMPOWER statement to disable the nongraphical results. Next, specify the PLOT statement with X=POWER to request a plot with power on the X axis. Sample size is automatically placed on the Y axis. Use the MIN= and MAX= options in the PLOT statement to specify the power range. The following statements produce the plot:

```
ods listing style=htmlbluecml;
ods graphics on;

proc glmpower data=Fluids2 plotonly;
   class Altitude Fluid;
   model LacticAcid = Altitude Fluid;
   weight CellWgt;
   contrast "Water vs. others" Fluid  -1 -1 -1 -1 4;
   contrast "EZD vs. LZ"       Fluid   1  1 -1 -1 0;
   contrast "EZD1 vs. EZD2"    Fluid   1 -1  0  0 0;
   contrast "LZ1 vs. LZ2"      Fluid   0  0  1 -1 0;
   power
      nfractional
      stddev     = 3.5
      ncovariates = 1
      corrxy     = 0.2 0.3 0
      alpha      = 0.025
      ntotal     = .
      power      = 0.9;
   plot x=power min=.5 max=.95;
run;
```

The ODS LISTING STYLE=HTMLBLUECML statement specifies the HTMLBLUECML style, which is suitable for use with PROC GLMPOWER because it allows both marker symbols and line styles to vary. See the section "ODS Styles Suitable for Use with PROC GLMPOWER" on page 3363 for more information. See Output 43.2.2 for the resulting plot.

Output 43.2.2 Plot of Sample Size versus Power for Two-Way ANOVA Contrasts
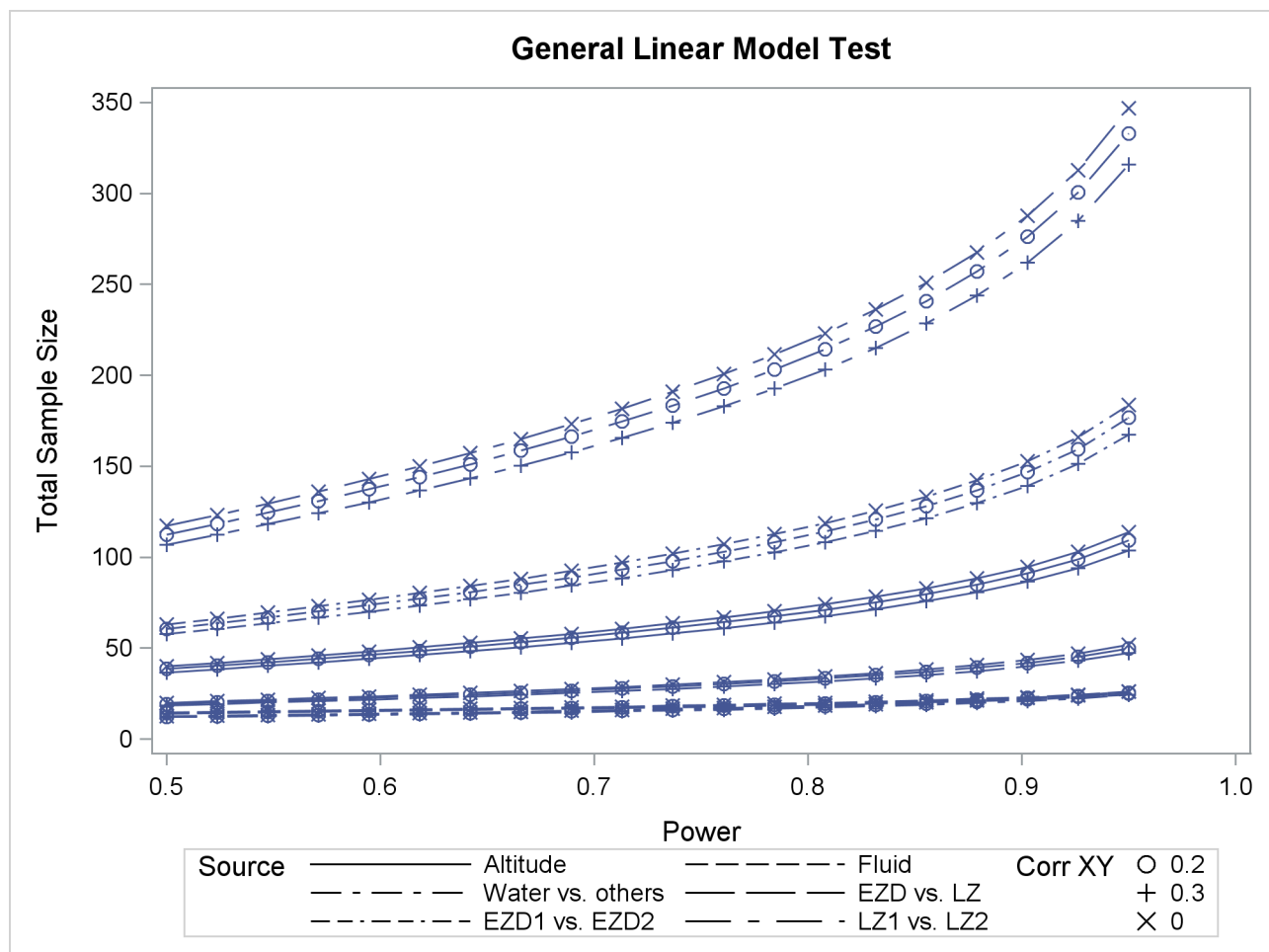
In Output 43.1.2, the line style identifies the test, and the plotting symbol identifies the scenario for the correlation between covariate and response. The plotting symbol locations identify actual computed powers; the curves are linear interpolations of these points. As in Example 43.1, the required sample size is highest for the test of LZ1 versus LZ2.

Finally, suppose you want to plot the power for the range of sample sizes you will likely consider for the study (the range of 21 to 275 that achieves 0.9 power for different comparisons). In the POWER statement, identify power as the result (POWER=.), and specify NTOTAL=21. Specify the PLOT statement with X=N to request a plot with sample size on the X axis.

The following statements produce the plot:

```
proc glmpower data=Fluids2 plotonly;
   class Altitude Fluid;
   model LacticAcid = Altitude Fluid;
   weight CellWgt;
   contrast "Water vs. others" Fluid  -1 -1 -1 -1 4;
   contrast "EZD vs. LZ"       Fluid   1  1 -1 -1 0;
   contrast "EZD1 vs. EZD2"    Fluid   1 -1  0  0 0;
   contrast "LZ1 vs. LZ2"      Fluid   0  0  1 -1 0;
```

```
      power
         nfractional
         stddev      = 3.5
         ncovariates = 1
         corrxy      = 0.2 0.3 0
         alpha       = 0.025
         ntotal      = 21
         power       = .;
      plot x=n min=21 max=275;
   run;


   ods graphics off;
```

The MAX=275 option in the PLOT statement sets the maximum sample size value. The MIN= option automatically defaults to the value of 21 from the NTOTAL= option in the POWER statement.

See Output 43.2.3 for the plot.

Output 43.2.3 Plot of Power versus Sample Size for Two-Way ANOVA Contrasts
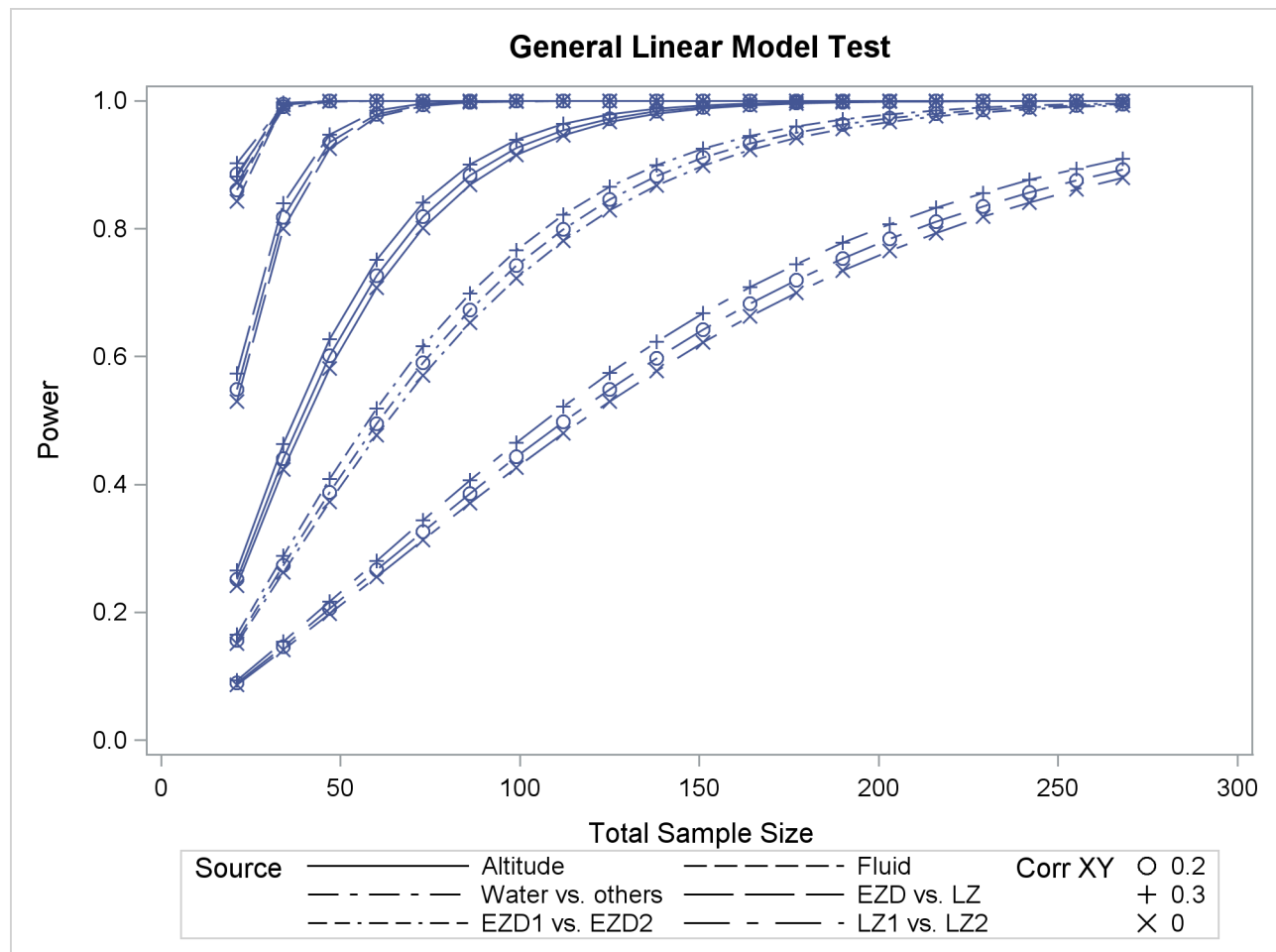
Although Output 43.2.2 and Output 43.2.3 surface essentially the same computations for practical power ranges, they each provide a different quick visual assessment. Output 43.2.2 reveals the range of required

sample sizes for powers of interest, and Output 43.2.3 reveals the range of powers achieved for sample sizes of interest.

References

Castelloe, J. M. (2000), "Sample Size Computations and Power Analysis with the SAS System," *Proceedings of the Twenty-fifth Annual SAS Users Group International Conference,* Paper 265-25, Cary, NC: SAS Institute Inc.

Castelloe, J. M. and O'Brien, R. G. (2001), "Power and Sample Size Determination for Linear Models," *Proceedings of the Twenty-sixth Annual SAS Users Group International Conference,* Paper 240-26, Cary, NC: SAS Institute Inc.

Hocking, R. R. (1985), *The Analysis of Linear Models*, Monterey, CA: Brooks/Cole.

Lenth, R. V. (2001), "Some Practical Guidelines for Effective Sample Size Determination," *The American Statistician*, 55, 187–193.

Muller, K. E. and Benignus, V.A. (1992), "Increasing Scientific Power with Statistical Power," *Neurotoxicology and Teratology*, 14, 211–219.

Muller, K. E., LaVange, L. M., Ramey, S. L., and Ramey, C. T. (1992), "Power Calculations for General Linear Multivariate Models Including Repeated Measures Applications," *Journal of the American Statistical Association*, 87 (420), 1209–1226.

Muller, K. E. and Peterson, B. L. (1984), "Practical Methods for Computing Power in Testing the Multivariate General Linear Hypothesis," *Computational Statistics & Data Analysis*, 2, 143–158.

O'Brien, R. G. and Castelloe, J. (2007), "Sample-Size Analysis for Traditional Hypothesis Testing: Concepts and Issues," in *Pharmaceutical Statistics Using SAS: A Practical Guide*, ed. A. Dmitrienko, C. Chuang-Stein and R. D'Agostino, Cary, NC: SAS Institute Inc., Chapter 10, 237–271.

O'Brien, R. G. and Muller, K. E. (1993), "Unified Power Analysis for *t*-Tests through Multivariate Hypotheses," in *Applied Analysis of Variance in Behavioral Science*, ed. L. K. Edwards, Chapter 8, 297–344, New York: Marcel Dekker.

O'Brien, R. G. and Shieh, G. (1992), "Pragmatic, Unifying Algorithm Gives Power Probabilities for Common F Tests of the Multivariate General Linear Hypothesis," poster presented at the American Statistical Association Meetings, Boston, Statistical Computing Section. Also, paper in review, downloadable in PDF form from www.bio.ri.ccf.org/UnifyPow.

Subject Index

S

V

Syntax Index

Your Turn

We welcome your feedback.

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

SAS® Publishing Delivers!

Whether you are new to the work force or an experienced professional, you need to distinguish yourself in this rapidly changing and competitive job market. SAS® Publishing provides you with a wide range of resources to help you set yourself apart. Visit us online at support.sas.com/bookstore.

SAS® Press

Need to learn the basics? Struggling with a programming problem? You'll find the expert answers that you need in example-rich books from SAS Press. Written by experienced SAS professionals from around the world, SAS Press books deliver real-world insights on a broad range of topics for all skill levels.

support.sas.com/saspress

SAS® Documentation

To successfully implement applications using SAS software, companies in every industry and on every continent all turn to the one source for accurate, timely, and reliable information: SAS documentation. We currently produce the following types of reference documentation to improve your work experience:

- Online help that is built into the software.
- Tutorials that are integrated into the product.
- Reference documentation delivered in HTML and PDF – **free** on the Web.
- Hard-copy books.

support.sas.com/publishing

SAS® Publishing News

Subscribe to SAS Publishing News to receive up-to-date information about all new SAS titles, author podcasts, and new Web site features via e-mail. Complete instructions on how to subscribe, as well as access to past issues, are available at our Web site.

support.sas.com/spn

§sas | THE POWER TO KNOW®