SAS/STAT® 15.1 User's Guide
Introduction to Power and Sample Size Analysis

This document is an individual chapter from *SAS/STAT® 15.1 User's Guide*.

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2018. *SAS/STAT® 15.1 User's Guide*. Cary, NC: SAS Institute Inc.

SAS/STAT® 15.1 User's Guide

SAS Institute Inc., SAS Campus Drive, Cary, NC 27513-2414

November 2018

Chapter 18
Introduction to Power and Sample Size Analysis

Contents

Overview

Power and sample size analysis optimizes the resource usage and design of a study, improving chances of conclusive results with maximum efficiency. The standard statistical testing paradigm implicitly assumes that Type I errors (mistakenly concluding significance when there is no true effect) are more costly than Type II errors (missing a truly significant result). This might be appropriate for your situation, or the relative costs of the two types of error might be reversed. For example, in screening experiments for drug development, it is often less damaging to carry a few false positives forward for follow-up testing than to miss potential leads. Power and sample size analysis can help you achieve your desired balance between Type I and Type II errors. By using optimal designs and sample sizes, you can improve your chances of detecting effects that might otherwise have been ignored, save money and time, and perhaps minimize risks to subjects.

Relevant tools for power and sample size analysis include the following:

- the GLMPOWER procedure in SAS/STAT
- the POWER procedure in SAS/STAT
- the %POWTABLE macro
- the power and sample size tasks in SAS Studio
- various procedures, statements, and functions in Base SAS and SAS/STAT for developing customized formulas and simulations

These tools, discussed in detail in the section “SAS Tools for Power and Sample Size Analysis” on page 379, deal exclusively with *prospective* analysis—that is, planning for a future study. This is in contrast to *retrospective* analysis for a past study, which is not supported by the main tools. Although retrospective analysis is more convenient to perform, it is often uninformative or misleading, especially when power is computed directly based on observed data.

The goals of prospective power and sample size analysis include the following:

- 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
- computing the probability of achieving the desired precision of a confidence interval, or the sample size required to ensure this probability
- conducting what-if analyses to assess how sensitive the power or required sample size is to other factors

The phrase *power analysis* is used for the remainder of this document as a shorthand to represent any or all of these goals. For more information about the GLMPOWER procedure, see Chapter 52, “The GLMPOWER Procedure.” For more information about the POWER procedure, see Chapter 93, “The POWER Procedure.” For more information about SAS Studio, see *SAS Studio: Task Reference Guide*.

Coverage of Statistical Analyses

The GLMPOWER procedure covers power analysis for Type III F tests and contrasts of fixed effects in univariate and multivariate linear models. For univariate models, you can specify covariates, which can be continuous or categorical. For multivariate models, you can choose among Wilks’ likelihood ratio, Hotelling-Lawley trace, and Pillai’s trace F tests for multivariate analysis of variance (MANOVA) and among uncorrected, Greenhouse-Geisser, Huynh-Feldt, and Box conservative F tests for the univariate approach to repeated measures. Tests and contrasts that involve random effects are not supported.

The POWER procedure covers power analysis for the following:

- t tests, equivalence tests, and confidence intervals for means
- tests, equivalence tests, and confidence intervals for binomial proportions

- multiple regression
- tests of correlation and partial correlation
- one-way analysis of variance
- rank tests for comparing two survival curves
- Cox proportional hazards regression
- logistic regression with binary response
- Wilcoxon Mann-Whitney rank-sum test
- extensions of existing analyses that involve the chi-square, F, t, or normal distribution, or the distribution of the correlation coefficient under multivariate normality

The extensions of existing analyses consist of scalar multipliers and custom values for primary noncentrality and degrees of freedom. Important use cases include the following:

- Wald and likelihood ratio tests in generalized linear models that have nominal, count, or ordinal responses
- sample size inflation that is caused by correlated predictors
- sample size deflation that is caused by correlation between the covariates and the response

Examples of generalized linear models include Poisson regression, logistic regression, and zero-inflated models.

The power and sample size tasks in SAS Studio cover a large subset of the analyses in the GLMPOWER and POWER procedures.

Statistical Background

Hypothesis Testing, Power, and Confidence Interval Precision

Standard Hypothesis Tests

In statistical hypothesis testing, you typically express the belief that some effect exists in a population by specifying an alternative hypothesis H_1. You state a null hypothesis H_0 as the assertion that the effect does *not* exist and attempt to gather evidence to reject H_0 in favor of H_1. Evidence is gathered in the form of sample data, and a statistical test is used to assess H_0. If H_0 is rejected but there really is *no* effect, this is called a *Type I error*. The probability of a Type I error is usually designated “alpha” or α, and statistical tests are designed to ensure that α is suitably small (for example, less than 0.05).

If there is an effect in the population but H_0 is *not* rejected in the statistical test, then a *Type II error* has been committed. The probability of a Type II error is usually designated “beta” or β. The probability $1 - \beta$ of avoiding a Type II error (that is, correctly rejecting H_0 and achieving statistical significance) is called the *power* of the test.

Most, but not all, of the power analyses in the GLMPOWER and POWER procedures are based on such standard hypothesis tests.

Equivalence and Noninferiority

Whereas the standard two-sided hypothesis test for a parameter μ (such as a mean difference) aims to demonstrate that it is significantly different than a null value μ_0:

$$H_0{:}\mu = \mu_0$$
$$H_1{:}\mu \neq \mu_0$$

an equivalence test instead aims to demonstrate that it is *significantly similar* to some value, expressed in terms of a range θ_L, θ_U around that value:

$$H_0{:}\mu < \theta_L \quad \text{or} \quad \mu > \theta_U$$
$$H_1{:}\theta_L \leq \mu \leq \theta_U$$

Whereas the standard one-sided hypothesis test for μ (say, the upper one-sided test) aims to demonstrate that it is significantly greater than μ_0:

$$H_0{:}\mu \leq \mu_0$$
$$H_1{:}\mu > \mu_0$$

a corresponding noninferiority test aims to demonstrate that it is *not significantly less* than μ_0, expressed in terms of a margin $\delta > 0$:

$$H_0{:}\mu \leq \mu_0 - \delta$$
$$H_1{:}\mu > \mu_0 - \delta$$

Corresponding forms of these hypotheses with the inequalities reversed apply to lower one-sided noninferiority tests (sometimes called *nonsuperiority* tests).

The POWER procedure performs power analyses for equivalence tests for one-sample, paired, and two-sample tests of normal and lognormal mean differences and ratios. It also supports noninferiority tests for a variety of analyses of means, proportions, and correlation, both directly (with a MARGIN= option representing δ) and indirectly (with an option for a custom null value representing the sum or difference of μ_0 and δ).

Confidence Interval Precision

An analysis of confidence interval precision is analogous to a traditional power analysis, with *CI Half-Width* taking the place of effect size and *Prob(Width)* taking the place of power. The *CI Half-Width* is the margin of error associated with the confidence interval, the distance between the point estimate and an endpoint. The *Prob(Width)* is the probability of obtaining a confidence interval with *at most* a target half-width.

The POWER procedure performs confidence interval precision analyses for *t*-based confidence intervals for one-sample, paired, and two-sample designs, and for several varieties of confidence intervals for a binomial proportion.

Computing Power and Sample Size

For some statistical models and tests, power analysis calculations are exact—that is, they are based on a mathematically accurate formula that expresses power in terms of the other components. Such formulas typically involve either enumeration or noncentral versions of the distribution of the test statistic.

When a power computation is based on a noncentral t, F, or chi-square distribution, the noncentrality parameter generally has the same form as the test statistic, with the conjectured population parameters in place of their corresponding estimators.

For example, the test statistic for a two-sample t test is computed as follows:

$$t = N^{\frac{1}{2}}(w_1 w_2)^{\frac{1}{2}} \left(\frac{\bar{x}_2 - \bar{x}_1 - \mu_0}{s_p} \right)$$

where N is the total sample size, w_1 and w_2 are the group allocation weights, $\bar{x}_1$ and $\bar{x}_2$ are the sample means, μ_0 is the null mean difference, and s_p is the pooled standard deviation. Under the null hypothesis, the statistic $F = t^2$ is distributed as $F(1, N-2)$. In general, F has a noncentral F distribution $F(1, N-2, \delta^2)$ where

$$\delta = N^{\frac{1}{2}}(w_1 w_2)^{\frac{1}{2}} \left(\frac{\mu_{\text{diff}} - \mu_0}{\sigma} \right)$$

and μ_{diff} and σ are the (unknown) true mean difference and common group standard deviation, respectively. Note that the square-root noncentrality δ is exactly the same as the t statistic except that the estimators of mean difference and standard deviation are replaced by their corresponding true population values.

The power for the two-sided two-sample t test with significance level α is computed as

$$P\left(F \geq F_{1-\alpha}(1, N-2)\right)$$

where F is distributed as $F(1, N-2, \delta^2)$ and $F_{1-\alpha}(1, N-2)$ is the $100(1-\alpha)\%$ quantile of the central F distribution with 1 and N – 2 degrees of freedom. See the section “Customized Power Formulas (DATA Step)” on page 384 for an example of the implementation of this formula in the DATA step.

In the absence of exact mathematical results, approximate formulas can sometimes be used. When neither exact power computations nor reasonable approximations are possible, simulation provides an increasingly viable alternative. You specify values for model parameters and use them to randomly generate a large number of hypothetical data sets. Applying the statistical test to each data set, you estimate power with the percentage of times the null hypothesis is rejected. While the simulation approach is computationally intensive, faster computing makes this less of an issue. A simulation-based power analysis is always a valid option, and, with a large number of data set replications, it can often be more accurate than approximations. See the section “Empirical Power Simulation (DATA Step, SAS/STAT Software)” on page 385 for an example of an empirical power simulation.

Sample size is usually computed by iterative numerical methods because it often cannot be expressed in closed form as a function of the other parameters. Sample size tends to appear in both a noncentrality parameter and a degrees of freedom term for the critical value.

Power and Study Planning

Power analysis is most effective when performed at the study planning stage, and as such it encourages early collaboration between researcher and statistician. It also focuses attention on effect sizes and variability in the underlying scientific process, concepts that both researcher and statistician should consider carefully at this stage.

There are many factors involved in a power analysis, such as the research objective, design, data analysis method, power, sample size, Type I error, variability, and effect size. By performing a power analysis, you can learn about the relationships between these factors, optimizing those that are under your control and exploring the implications of those that are fixed or unknown.

Components of Study Planning

Even when the research questions and study design seem straightforward, the ensuing power analysis can seem technically daunting. It is often helpful to break the process down into five components:

- **Study Design**: What is the structure of the planned design? This must be clearly and completely specified. What groups and treatments ("cells" and "factors" of the design) are going to be assessed, and what will be the relative sizes of those cells? How is each case going to be studied—that is, what is the primary outcome measure ("dependent variable")? Will covariates be measured and included in the statistical model?

- **Scenario Model**: What are your beliefs about patterns in the data? Imagine that you had unlimited time and resources to execute the study design, so that you could gather an "infinite data set." Characterize that infinite data set as best you can using a mathematical model, realizing that it will be a simplification of reality. Alternatively, as is common with complex linear models, you may decide to construct an "exemplary" data set that mimics the infinite data set. However you do this, your scenario model should capture the key features of the study design and the main relationships among the primary outcome variables and study factors.

- **Effects and Variability**: What exactly are the "signals and noises" in the patterns you suspect? Set specific values for the parameters of your scenario model, keeping at most one unspecified. It is often enlightening to consider a variety of realistic possibilities for the key values by performing a sensitivity analysis, to explore the consequences of competing views on what the infinite data set might look like.

- **Statistical Method**: How will you cast your model in statistical terms and conduct the eventual data analysis? Define the statistical models and procedures that will be used to embody the study design and estimate and/or test the effects central to the research question. What significance levels will be used? Will one- or two-sided tests be used?

- **Aim of Assessment**: Finally, what needs to be determined in the power analysis? Most often you want to examine the statistical powers obtained across the various scenarios for the effects, variability, alternative varieties of the statistical procedures to be used, and the feasible total sample sizes. Sometimes the goal is to find sample size values that provide given levels of power, say 85%, 90%, or 95%.

Effect Size

There is some confusion in practice about how to postulate the effect size. One alternative is to specify the effect size that represents minimal clinical significance; then the result of the power analysis reveals the chances of detecting a minimally meaningful effect size. Often this minimal effect size is so small that it requires excessive resources to detect. Another alternative is to make an educated guess of the true underlying effect size. Then the power analysis determines the chance of detecting the effect size that is believed to be true. The choice is ultimately determined by the research goals. Finally, you can specify a collection of possible values, perhaps spanning the range between minimally meaningful effects and larger surmised effects.

You can arrive at values for required quantities in a power analysis, such as effect sizes and measures of variability, in many different ways. For example, you can use pilot data, results of previous studies reported in literature, educated guesses derived from theory, or educated guesses derived from partial data (a small sample or even just quantiles).

Uncertainty and Sensitivity Analysis

Uncertainty is a fact of life in any power analysis, because at least some of the numbers used are best guesses of unknown values. The result of a power calculation, whether it be achieved power or required sample size or something else, serves only as a point estimate, conditional on the conjectured values of the other components. It is not feasible in general to quantify the variability involved in using educated guesses or undocumented results to specify these components. If observed data are used, relevant adjustments for variability in the data tend to be problematic in the sense of producing confidence intervals for power that are too wide for practical use. But there is a useful way for you to characterize the uncertainty in your power analysis, and also discover the extent to which statistical power is affected by each component. You can posit a reasonable range for each input component, vary each one within its range, and observe the variety of results in the form of tables or graphs.

SAS Tools for Power and Sample Size Analysis

This section demonstrates how you can use the different SAS power analysis tools mentioned in the section “Overview” on page 373 to generate graphs, tables, and narratives; implement your own power formulas; and simulate empirical power.

Suppose you want to compute the power of a two-sample *t* test. You conjecture that the mean difference is between 5 and 6 and that the common group standard deviation is between 12 and 18. You plan to use a significance level between 0.05 and 0.1 and a sample size between 100 and 200. The following SAS statements use the POWER procedure to compute the power for these scenarios:

```
proc power;
   twosamplemeans test=diff
      meandiff = 5 6
      stddev = 12 18
      alpha = 0.05 0.1
      ntotal = 100 200
      power = .;
run;
```

Figure 18.1 shows the results. Depending on the plausibility of the various combinations of input parameter values, the power ranges between 0.379 and 0.970.

Figure 18.1 PROC POWER Tabular Output

The POWER Procedure
Two-Sample t Test for Mean Difference

Computed Power

Index	Alpha	Mean Diff	Std Dev	N Total	Power
1	0.05	5	12	100	0.541
2	0.05	5	12	200	0.834
3	0.05	5	18	100	0.280
4	0.05	5	18	200	0.498
5	0.05	6	12	100	0.697
6	0.05	6	12	200	0.940
7	0.05	6	18	100	0.379
8	0.05	6	18	200	0.650
9	0.10	5	12	100	0.664
10	0.10	5	12	200	0.902
11	0.10	5	18	100	0.397
12	0.10	5	18	200	0.623
13	0.10	6	12	100	0.799
14	0.10	6	12	200	0.970
15	0.10	6	18	100	0.505
16	0.10	6	18	200	0.759

The following seven sections illustrate additional ways of displaying these results using the different SAS tools.

Basic Graphs (PROC POWER, PROC GLMPOWER, and Power and Sample Size Tasks in SAS Studio)

If you include a PLOT statement, the GLMPOWER and POWER procedures produce standard power curves, which represent any multivalued input parameters that have varying line styles, symbols, colors, panels, or any combination of these. The power and sample size tasks in SAS Studio also provide options that enable you to produce power curves. If ODS Graphics is enabled, then graphs are created using ODS Graphics; otherwise, traditional graphs are produced.

To display default power curves for the preceding PROC POWER call, add the PLOT statement with no arguments as follows:

```
ods graphics on;

proc power plotonly;
   twosamplemeans test=diff
      meandiff = 5 6
      stddev = 12 18
      alpha = 0.05 0.1
      ntotal = 100 200
      power = .;
   plot;
run;

ods graphics off;
```

The ODS GRAPHICS ON statement enables ODS Graphics.

Figure 18.2 shows the results. Note that the line style varies by the significance level α, the symbol varies by the mean difference, and the panel varies by standard deviation.

Figure 18.2 PROC POWER Default Graphical Output

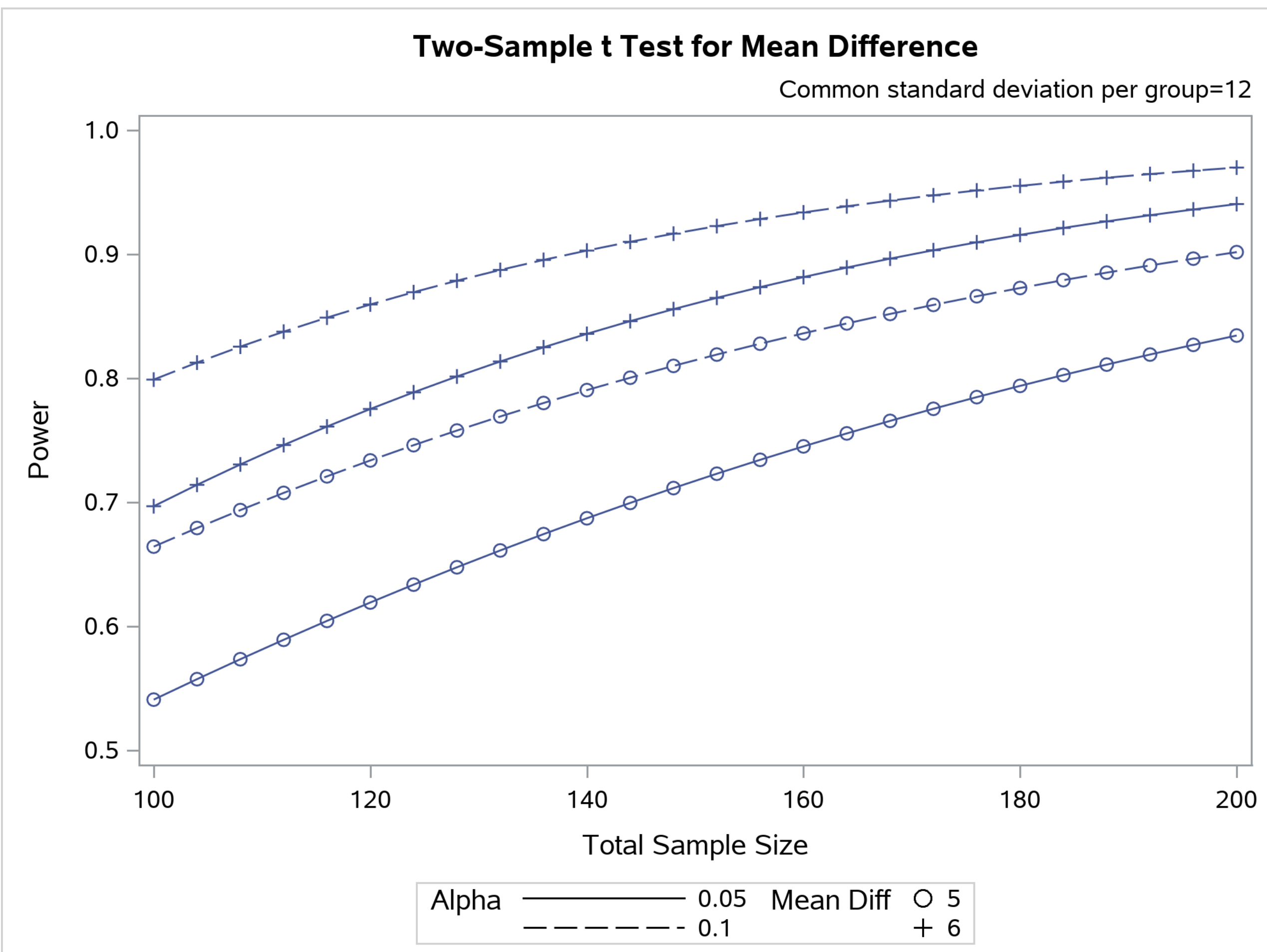
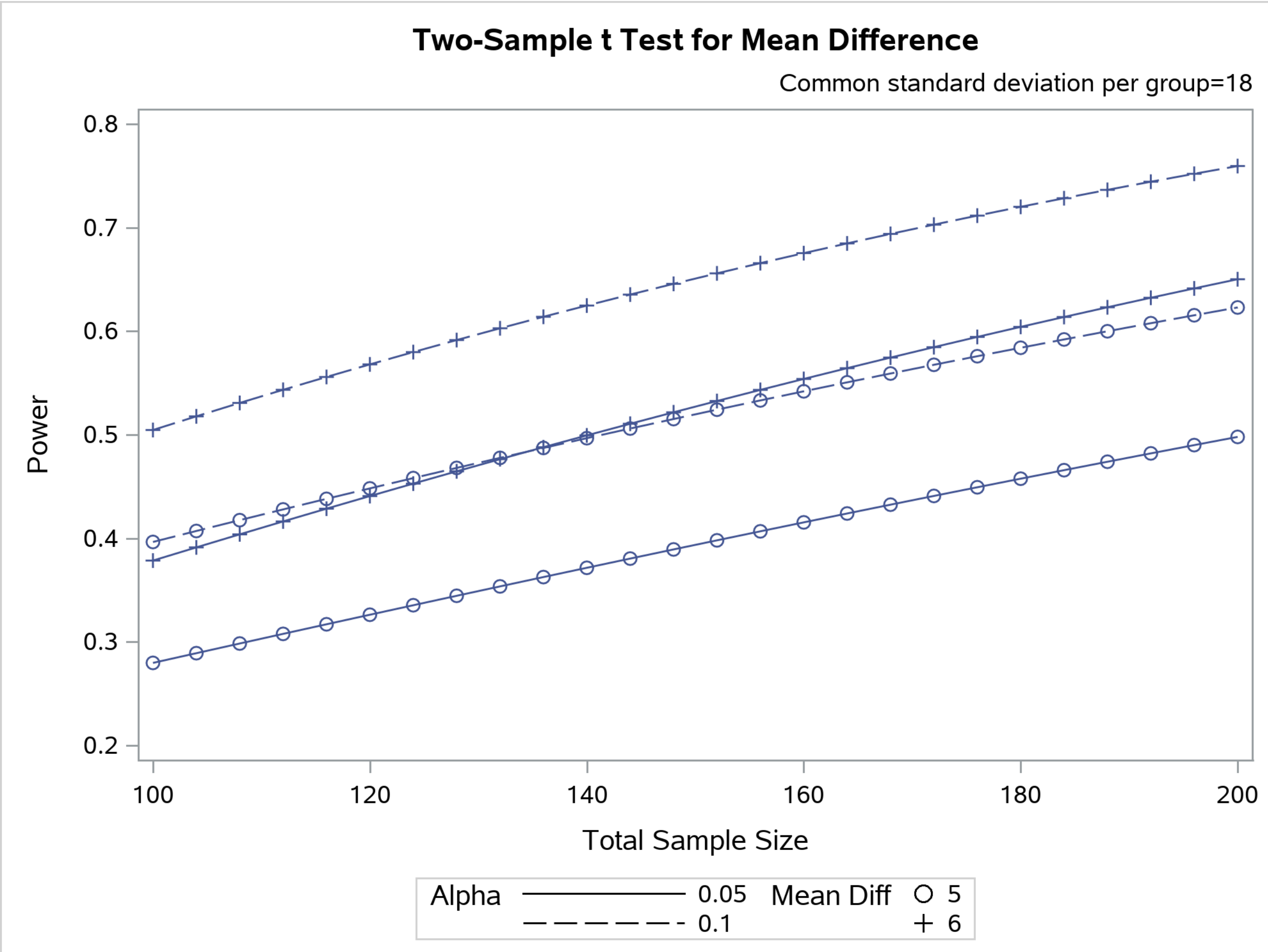


Figure 18.2 *continued*



Highly Customized Graphs (PROC POWER and PROC GLMPOWER)

Example 93.8 of Chapter 93, "The POWER Procedure," demonstrates various ways you can modify and enhance plots created in the GLMPOWER or POWER procedures:

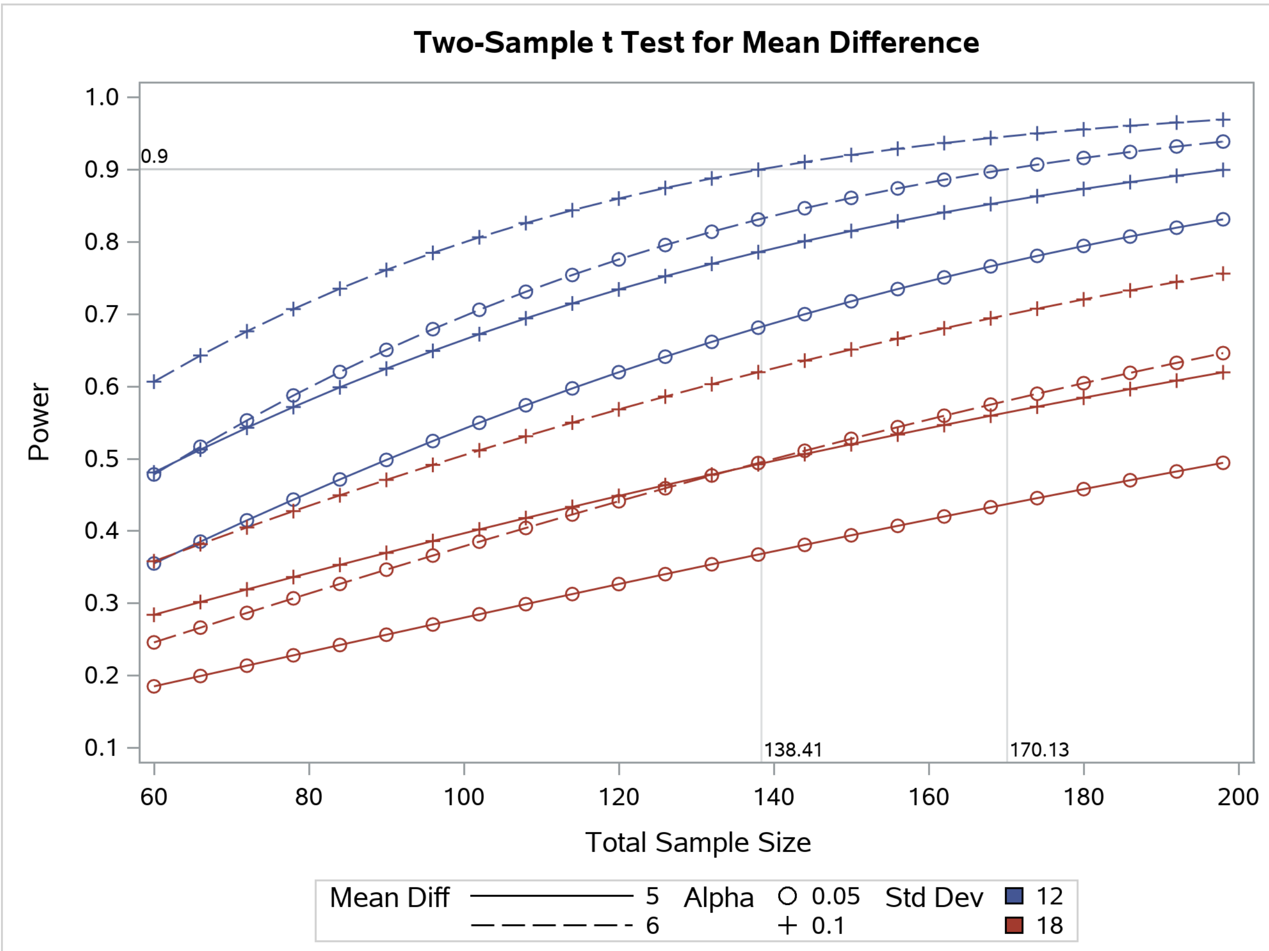
- assigning analysis parameters to axes
- fine-tuning a sample size axis
- adding reference lines
- linking plot features to analysis parameters
- choosing key (legend) styles
- modifying symbol locations

For example, replace the default PLOT statement with the following statement to modify the graphical results in Figure 18.2 to lower the minimum sample size to 60, show a reference line at power=0.9 with corresponding sample size values, distinguish standard deviation by color instead of panel, and swap the roles of α and mean difference:

```
plot
   min=60
   yopts=(ref=0.9 crossref=yes)
   vary(color by stddev, linestyle by meandiff, symbol by alpha);
```

Figure 18.3 shows the results. The plot reveals that only the scenarios with the largest mean difference and smallest standard deviation achieve a power of at least 0.9 for this sample size range.

Figure 18.3 PROC POWER Customized Graphical Output



Formatted Tables (%POWTABLE Macro)

The %POWTABLE macro renders the output of the POWER and GLMPOWER procedures in rectangular form, and it optionally produces simplified results using weighted means across chosen variables. PROC REPORT and the Output Delivery System (ODS) are used to generate the tables. Base SAS and SAS/STAT 9.1 or higher versions are required.

You can run the %POWTABLE macro for the output in Figure 18.1 to display the results in a form more suitable for quickly discerning relationships among parameters. First use the ODS OUTPUT statement to assign the “Output” table produced by the POWER procedure to a data set as follows:

```
ods output output=powdata;
```

Next, specify the same PROC POWER statements that generate Figure 18.1. Finally, use the %POWTABLE macro to assign analysis parameters to table dimensions. To create a table of computed power values with mean difference assigned to rows, sample size and α assigned to columns, and standard deviation assigned to “panels” (rendered by default as rows separated by blank lines), specify the following statements:

```
%powtable ( Data = powdata,
            Entries = power,
            Rows = meandiff,
            Cols = ntotal alpha,
            Panels = stddev )
```

Figure 18.4 shows the results.

Figure 18.4 %POWTABLE Macro Output

The POWTABLE Macro

Entries are Power

		N Total			
		100		200	
		Alpha		Alpha	
Std Dev	**Mean Diff**	**0.05**	**0.10**	**0.05**	**0.10**
12	5	0.541	0.664	0.834	0.902
	6	0.697	0.799	0.940	0.970
18	5	0.280	0.397	0.498	0.623
	6	0.379	0.505	0.650	0.759

Customized Power Formulas (DATA Step)

If you want to perform a power computation for an analysis that is not currently supported directly in SAS/STAT tools, and you have a power formula, then you can program the formula in the DATA step.

For purposes of illustration, here is the power formula in the section “Computing Power and Sample Size” on page 377 implemented in the DATA step to compute power for the *t* test example:

```
data tpow;
   do meandiff = 5, 6;
      do stddev = 12, 18;
         do alpha = 0.05, 0.1;
            do ntotal = 100, 200;
               ncp = ntotal * 0.5 * 0.5 * meandiff**2 / stddev**2;
               critval = finv(1-alpha, 1, ntotal-2, 0);
               power = sdf('f', critval, 1, ntotal-2, ncp);
```

```
                    output;
                 end;
              end;
           end;
        end;
     run;

     proc print data=tpow;
     run;
```

The output is shown in Figure 18.5.

Figure 18.5 Customized Power Formula (DATA Step)

Obs	meandiff	stddev	alpha	ntotal	ncp	critval	power
1	5	12	0.05	100	4.3403	3.93811	0.54102
2	5	12	0.05	200	8.6806	3.88885	0.83447
3	5	12	0.10	100	4.3403	2.75743	0.66434
4	5	12	0.10	200	8.6806	2.73104	0.90171
5	5	18	0.05	100	1.9290	3.93811	0.27981
6	5	18	0.05	200	3.8580	3.88885	0.49793
7	5	18	0.10	100	1.9290	2.75743	0.39654
8	5	18	0.10	200	3.8580	2.73104	0.62287
9	6	12	0.05	100	6.2500	3.93811	0.69689
10	6	12	0.05	200	12.5000	3.88885	0.94043
11	6	12	0.10	100	6.2500	2.75743	0.79895
12	6	12	0.10	200	12.5000	2.73104	0.96985
13	6	18	0.05	100	2.7778	3.93811	0.37857
14	6	18	0.05	200	5.5556	3.88885	0.65012
15	6	18	0.10	100	2.7778	2.75743	0.50459
16	6	18	0.10	200	5.5556	2.73104	0.75935

Empirical Power Simulation (DATA Step, SAS/STAT Software)

You can obtain a highly accurate power estimate by simulating the power empirically. You need to use this approach for analyses that are not supported directly in SAS/STAT tools and for which you lack a power formula. But the simulation approach is also a viable alternative to existing power approximations. A high number of simulations will yield a more accurate estimate than a non-exact power approximation.

Although exact power computations for the two-sample t test are supported in several of the SAS/STAT tools, suppose for purposes of illustration that you want to simulate power for the continuing t test example. This section describes how you can use the DATA step and SAS/STAT software to do this.

The simulation involves generating a large number of data sets according to the distributions defined by the power analysis input parameters, computing the relevant p-value for each data set, and then estimating the power as the proportion of times that the p-value is significant.

The following statements compute a power estimate along with a 95% confidence interval for power for the first scenario in the two-sample t test example, with 10,000 simulations:

```
%let meandiff =     5;
%let stddev   =    12;
%let alpha    =  0.05;
%let ntotal   =   100;
%let nsim     = 10000;

data simdata;
   call streaminit(123);
   do isim = 1 to &nsim;
      do i = 1 to floor(&ntotal/2);
         group = 1;
         y = rand('normal', 0        , &stddev);
         output;
         group = 2;
         y = rand('normal', &meandiff, &stddev);
         output;
      end;
   end;
run;

ods exclude all;
proc ttest data=simdata;
   ods output ttests=tests;
   by isim;
   class group;
   var y;
run;
ods exclude none;

data tests;
   set tests;
   where method="Pooled";
   issig = probt < &alpha;
run;

proc freq data=tests;
   ods select binomial;
   tables issig / binomial(level='1');
run;
```

First the DATA step is used to randomly generate *nsim* = 10,000 data sets based on the *meandiff*, *stddev*, and *ntotal* parameters and the normal distribution, consistent with the assumptions underlying the two-sample *t* test. These data sets are contained in a large SAS data set called simdata indexed by the variable isim.

The CALL STREAMINIT(123) statement initializes the random number generator with a specific sequence and ensures repeatable results for purposes of this example. (**NOTE:** Skip this step when you are performing actual power simulations.)

The TTEST procedure is run using isim as a BY variable, with the ODS EXCLUDE ALL statement to suppress output. The ODS OUTPUT statement saves the “TTests” table to a data set called tests. The *p*-values are contained in a column called probt.

The subsequent DATA step defines a variable called issig to flag the significant *p*-values.

Finally, the FREQ procedure computes the empirical power estimate as the estimate of P(issig = 1) and provides approximate and exact confidence intervals for this estimate.

Figure 18.6 shows the results. The estimated power is 0.5388 with 95% confidence interval (0.5290, 0.5486). Note that the exact power of 0.541 shown in the first row in Figure 18.1 is contained within this tight confidence interval.

Figure 18.6 Simulated Power (DATA Step, SAS/STAT Software)

The FREQ Procedure

Binomial Proportion issig = 1	
Proportion	0.5388
ASE	0.0050
95% Lower Conf Limit	0.5290
95% Upper Conf Limit	0.5486
Exact Conf Limits	
95% Lower Conf Limit	0.5290
95% Upper Conf Limit	0.5486

References

Castelloe, J. M. (2000). "Sample Size Computations and Power Analysis with the SAS System." In *Proceedings of the Twenty-Fifth Annual SAS Users Group International Conference*. Cary, NC: SAS Institute Inc. `http://www2.sas.com/proceedings/sugi25/25/st/25p265.pdf`.

Castelloe, J. M., and O'Brien, R. G. (2001). "Power and Sample Size Determination for Linear Models." In *Proceedings of the Twenty-Sixth Annual SAS Users Group International Conference*. Cary, NC: SAS Institute Inc. `http://www2.sas.com/proceedings/sugi26/p240-26.pdf`.

Lenth, R. V. (2001). "Some Practical Guidelines for Effective Sample Size Determination." *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.

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

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*, edited by L. K. Edwards, 297–344. New York: Marcel Dekker.

Index