

Chapter 12

Sample Size and Power Calculations

Chapter Table of Contents

Introduction	253
Hypothesis Testing	255
Confidence Intervals	260
Equivalence Tests	264
One-Way ANOVA	269
Power Computation Details	272
Hypothesis Tests	272
One-Way ANOVA	274
Confidence Intervals	275
Equivalence Tests	276
References	279

Chapter 12

Sample Size and Power Calculations

Introduction



Figure 12.1. Sample Size Menu

When you are planning a study or experiment, you often need to know how many units to sample to obtain a certain power, or you may want to know the power you would obtain with a specific sample

size. The *power* of a hypothesis test is the probability of rejecting the null hypothesis when the alternative hypothesis is true. With an inadequate sample size, you may not reach valid conclusions with your work; with an excessive sample size, you may waste valuable resources. Thus, performing sample size and power computations is often quite important.

The power and sample size calculations depend on the planned data analysis strategy. That is, if the primary hypothesis test is a two-sample *t*-test, then the power calculations must be based on that test. Otherwise, if the sample size calculations and data analyses are not aligned, the results may not be correct.

Determining sample size requirements ahead of the experiment is a prospective exercise. Then, you proceed to select the appropriate number of sampling units and perform data collection and analysis. However, power and sample size calculations are also useful retrospectively. For a given analysis, you may want to calculate what level of power you achieved or what sample size would have been needed for a given power.

Power and sample size calculations are a function of the specific alternative hypothesis of interest, in addition to other parameters. That is, the power results will vary depending on which value of the alternative hypothesis you specify, so sometimes it is useful to do these analyses for a range of values to see how sensitive the power analysis is to changes in the alternative hypothesis value. Often, you produce plots of power versus sample size, called *power curves*, to see how sample size and power affect each other.

The Sample Size tasks provide prospective sample size and power calculations for several types of analyses: *t*-tests, confidence intervals, and tests of equivalence. Each of these calculations is available for one-sample, paired-sample, and two-sample study designs. Power and sample size calculations are also available for the one-way ANOVA design. Multiple parameter values can be input, and results and power curves are produced for each combination of values. Note that retrospective power computations are also available in a number of the statistical tasks in the Analyst Application such as the Hypothesis Test, Regression, and ANOVA tasks.

Hypothesis Testing

Sample size and power calculations are available for one-sample and two-sample paired and independent designs where the proposed analysis is hypothesis testing of a mean or means via a t -test. These computations assume equally sized groups.

Suppose you want to compute the power for a one-sample t -test. The alternative hypothesis mean and the standard deviation have the values 8.6137 and 2.0851, respectively. You are interested in testing whether the null mean has the value 8, at an alpha level of 0.05, and you are interested in looking at a range of sample sizes from 11 to 211. The study for which these statistics were computed had a sample size of 51.

Requesting Power Computations for the One-Sample t -test

To access this task, select

Statistics → **Sample Size** → **One-Sample t -test** . . .

Figure 12.2 displays the resulting dialog. Note that, unlike the other statistical tasks that require a data set for analysis, performing one of the Sample Size tasks requires only entering information in the appropriate dialog. The data table is not involved.

The image shows the 'One-Sample t-test' dialog box in SAS. It is divided into several sections:

- Calculate:** Two radio buttons, 'Power' (selected) and 'N'.
- Test specifications:** A group box containing:
 - Null mean: []
 - Alternate mean: []
 - Standard deviation: []
 - Alpha: 0.05
 - N: From: [], To: [], By: []
- Plot:** A group box containing:
 - Power vs. N
 - Power ref line: []
 - N ref line: []
- Tails:** Two radio buttons, '1-sided' and '2-sided' (selected).
- Buttons:** OK, Cancel, Reset, Save Options, Help, and Titles.

Figure 12.2. Sample Size Dialog for One-Sample t-test

In this task, you specify whether you want to compute sample size or power, enter values for the test hypothesis and parameters, specify the alpha level (0.05 is the default), specify whether you want a power curve produced, and specify a range of power values or sample sizes depending on whether you are computing sample size or power.

To enter the information for this example, follow these steps:

1. Select **Power**.
2. Enter 8 as the **Null mean:** value.
3. Enter 8.6137 as the **Alternate mean:**
4. Enter 2.0851 as the **Standard deviation:**
5. Make sure that the **Alpha:** value is 0.05.
6. Enter 11 as the value for the **From:** field in the line for **N:**
7. Enter 211 and 20 as the values under **To:** and **By:**, respectively, in the line for **N:**
8. Select **Power vs. N** to produce a plot.

9. Enter 51 as the value for **N ref line**:
10. Select **2-sided** for **Tails** if it is not already selected.

Note that you can enter multiple values in fields such as for **Alpha**: and **Null mean**:, separated by commas or blanks, and the analysis will be performed for all combinations of the entered values. Here, power will be computed for sample sizes ranging from 11 to 211 in multiples of 20.

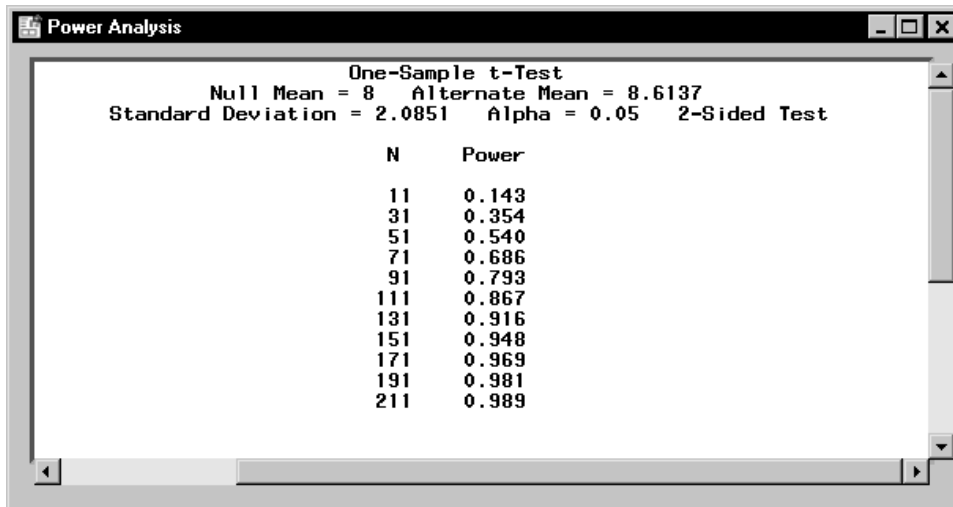
Figure 12.3 contains the completed dialog.

The screenshot shows the 'One-Sample t-test' dialog box with the following settings:

- Calculate:** Power, N
- Test specifications:**
 - Null mean: 8
 - Alternate mean: 8.6137
 - Standard deviation: 2.0851
 - Alpha: 0.05
- N:** From: 11, To: 211, By: 20
- Plot:**
 - Power vs. N
 - Power ref line: (empty)
 - N ref line: 51
- Tails:** 1-sided, 2-sided
- Buttons:** OK, Cancel, Reset, Save Options, Help, Titles

Figure 12.3. Sample Size Dialog for One-Sample t-test

Figure 12.4 contains the power computations for the sample sizes ranging from 11 to 211.



The screenshot shows a window titled "Power Analysis" with the following text and table:

One-Sample t-Test
Null Mean = 8 Alternate Mean = 8.6137
Standard Deviation = 2.0851 Alpha = 0.05 2-Sided Test

N	Power
11	0.143
31	0.354
51	0.540
71	0.686
91	0.793
111	0.867
131	0.916
151	0.948
171	0.969
191	0.981
211	0.989

Figure 12.4. Sample Size Results for One-Sample t-test

The interpretation of a power of 0.540 for $n = 51$ is as follows: suppose the true mean and standard deviation are 8.6137 and 2.0851, and suppose a random sample of 51 observations is taken. Then the probability that the hypothesis test will reject the null hypothesis ($H_0: \mu = 8.0$) and conclude (correctly) that the alternative hypothesis ($H_A: \mu = 8.6137$) is true is 0.540.

The requested plot is shown in Figure 12.5 with a reference line at $n = 51$.

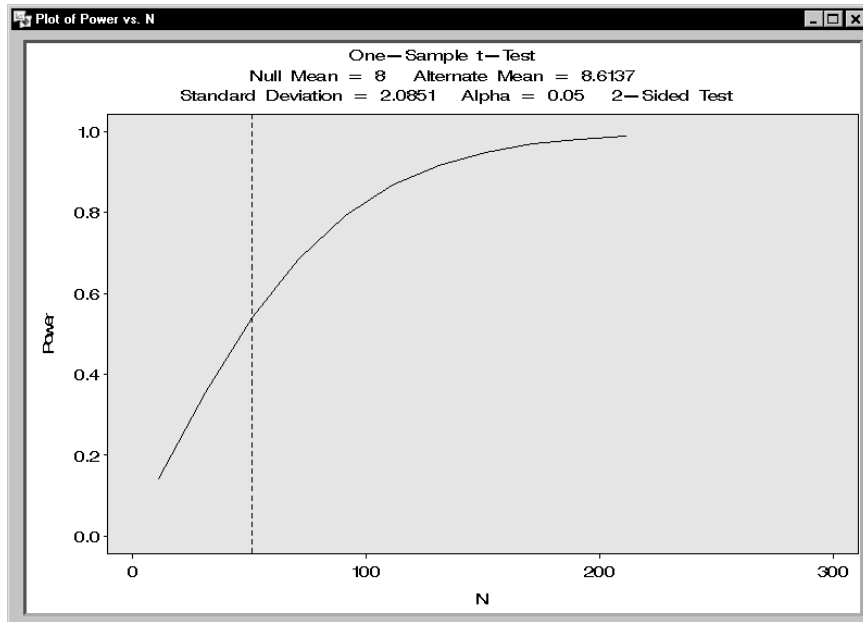


Figure 12.5. Plot of Power versus Sample Size

More on Hypothesis Tests

In the two-sample cases, you must enter the null means of each group and the standard deviation. In the paired case, the standard deviation entered is the standard deviation of the differences between the two groups. In the independent case, the standard deviation is the pooled standard deviation, which is calculated as follows:

$$S_p = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{(n_1 + n_2 - 2)}}$$

Confidence Intervals

Sample size and power calculations are available for one-sample and two-sample paired and independent designs, when the proposed analysis is construction of confidence intervals of a mean (one-sample) or difference of two means (two-sample), via the t -test.

To understand the power of a confidence interval, first define the *precision* to be half the length of a two-sided confidence interval (or the distance between the endpoint and the parameter estimate in a one-sided interval). The power can then be considered to be the probability that the desired precision is achieved, that is, the probability that the length of the two-sided interval is no more than twice the desired precision. Here, a slight modification of this concept is used. The power is considered to be the conditional probability that the desired precision is achieved, given that the interval includes the true value of the parameter of interest. The reason for the modification is that there is no reason for the interval to be particularly small if it does not contain the true value of the parameter.

These computations assume equally sized groups.

Requesting Power Computations for a Confidence Interval in a Paired t -test

To perform this task, select

Statistics → **Sample Size** → **Paired Confidence Interval . . .**

Figure 12.6 displays the resulting dialog.

Figure 12.6. Sample Size Dialog for Paired Confidence Interval

You specify whether you want to compute sample sizes or power, enter values for desired precision and standard deviation, enter the alpha levels, enter the sample sizes or power, and select if you want a power curve.

To request power for a paired confidence interval where the desired precision is 0.5 and the standard deviation is 2.462, follow these steps:

1. Select **Power**.
2. Enter 0.5 as the **Desired precision**:
3. Enter 2.462 as the **Std dev of diff**:
4. Enter 0.01, 0.025, 0.05, and 0.1 as values in the field for **Alpha**:
5. Enter 11 as the value for the **From**: field in the line for **N**:
6. Enter 211 and 5 as the values under **To**: and **By**:, respectively, in the line for **N**:
7. Select **Power vs. N**.
8. Select **2-sided** for **Interval** if it is not already selected.

Note that you can enter multiple values in these fields, for example, for **Alpha:** and **Desired precision:**, separated by commas or blanks, and the analysis will be performed for all combinations of the input values. Here, power will be computed for sample sizes ranging from 11 to 211 in multiples of 5.

Figure 12.7 contains the completed dialog. Note that, because multiple alpha values were entered, sets of results will be created for each one.

Figure 12.7. Completed Sample Size Dialog for Paired Confidence Interval

Figure 12.8 contains the power computations for the sample sizes ranging from 36 to 171.

Paired Confidence Interval		
Desired Precision = 0.5 Std Dev of Differences = 2.462 2-Sided Test		
Alpha	N	Power
0.025	36	< .01
	41	< .01
	46	< .01
	51	< .01
	56	< .01
	61	< .01
	66	< .01
	71	< .01
	76	< .01
	81	< .01
	86	0.010
	91	0.020
	96	0.040
	101	0.073
	106	0.125
	111	0.199
	116	0.298
	121	0.415
126	0.543	
131	0.668	
136	0.777	
141	0.863	
146	0.924	
151	0.961	
156	0.982	
161	> .99	
166	> .99	
171	> .99	

Figure 12.8. Sample Size Results for Paired Confidence Interval

The power analysis results in Figure 12.8 show that, for a two-sided paired confidence interval where the standard deviation of the differences is 2.462, the significance level is $\alpha = 0.025$, the sample size is 121, and the power is 0.415. That is, 0.415 represents the probability that a confidence interval containing the true parameter value has a length of no more than twice the desired precision of 0.5.

The requested plot is displayed in Figure 12.9 and includes one power curve for each specified alpha value.

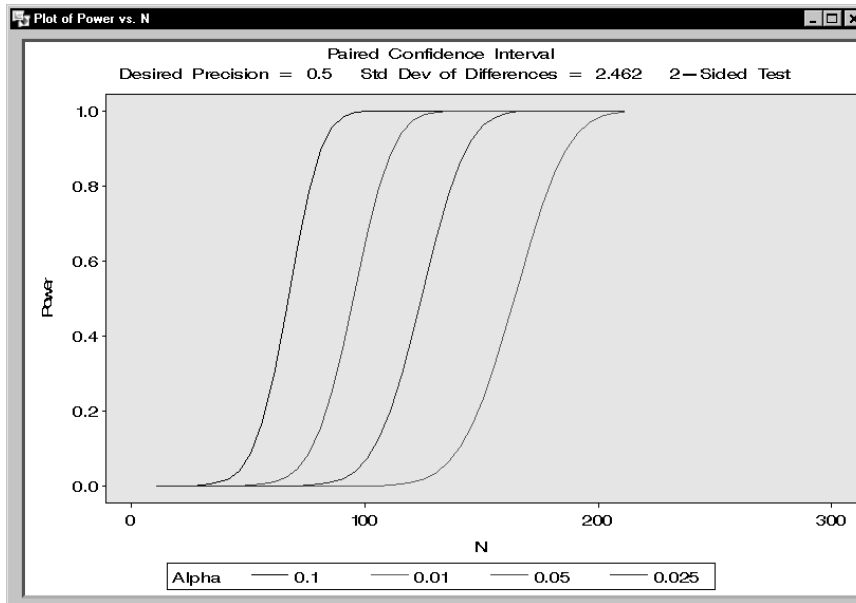


Figure 12.9. Plot for Paired Confidence Interval

Equivalence Tests

In a test of equivalence, a treatment mean and a reference mean are compared to each other. Equivalence is taken to be the alternative hypothesis, and the null hypothesis is nonequivalence. The model assumed may be additive or multiplicative. In the additive model (Phillips 1990), the focus is on the difference between the treatment mean and the reference mean, while in the multiplicative model (Diletti, Hauschke, and Steinijans 1991), the focus is on the ratio of the treatment mean to the reference mean.

In the additive model, the null hypothesis is that the difference between the treatment mean and the reference mean is not near zero. That is, the difference is less than the lower equivalence bound or greater than the upper equivalence bound and thus nonequivalent.

The alternative is that the difference is between the equivalence bounds; therefore, the two means are considered to be equivalent.

In the multiplicative model, the null hypothesis is that the ratio of the treatment mean to the reference mean is not near one. That is, the ratio is below the lower equivalence bound or above the upper equivalence bound, and thus the two means are not equivalent. The alternative is that the ratio is between the bounds; thus, the two means are considered to be equivalent.

The power of a test is the probability of rejecting the null hypothesis when the alternative is true. In this case, the power is the probability of accepting equivalence when the treatments are in fact equivalent, that is, the treatment difference or ratio is within the prespecified boundaries.

Often, the null difference is specified to be 0; the null hypothesis is that the treatment difference is less than the lower bound or greater than the upper bound, and the alternative is that the difference is not outside the bounds specified. However, in a case where you suspect that the treatments differ slightly (for example, $\mu_1 = 6$, $\mu_2 = 5$, $\mu_1 - \mu_2 = 1$), but you want to rule out a larger difference (for example, $|\mu_1 - \mu_2| > 2$) with probability equal to the power you select, you would specify the null difference to be 1 and the lower and upper bounds to be -2 and 2 , respectively. Note that the null difference must lie within the bounds you specify.

Requesting Sample Sizes for Paired Samples In Equivalence

As an example of computing sample sizes for an equivalence test, consider determining sample sizes for an additive model. The coefficient of variation is 0.2, and the null differences of interest are 0, 0.05, 0.10, and 0.15. The significance level under investigation is 0.05, and the power of interest is 0.80. The lower and upper equivalence bounds are -0.2 and 0.2 , respectively.

To perform this computation, select

Statistics → Sample Size → Paired Equivalence . . .

Figure 12.10 displays the resulting dialog.

Figure 12.10. Sample Size Dialog for Paired Equivalence

For this analysis, you need to input the model type, null difference, coefficient of variation, and the usual alpha level. In addition, you need to specify the equivalence bounds.

These bounds should be chosen to be the minimum difference so that, if the treatments differed by at least this amount, you would consider them to be different. For the multiplicative model, enter the bioequivalence lower and upper limits. For the additive model, enter the bioequivalence lower and upper limits as percentages of the reference mean $\frac{\text{lowerbound}}{\mu_R}$ and $\frac{\text{upperbound}}{\mu_R}$.

For the null difference or ratio, specify one or more values for the null hypothesis difference between the treatment and reference means (additive model) or the ratio of means (multiplicative model). The null difference/ratio value must lie within the equivalence bounds you specify. For the additive model, specify the null difference as a percentage of the reference mean $\frac{|\mu_T - \mu_R|}{\mu_R}$, where μ_T is the hypothesized treatment mean, and μ_R is the hypothesized reference mean. For the multiplicative model, calculate the null ratio as $\frac{\mu_T}{\mu_R}$.

You must also input one or more values for the coefficient of variation (c.v.). For the additive model, enter this as a percentage of the reference mean $\frac{\sigma}{\mu_R}$, which can be estimated by $\frac{\sqrt{MSE}}{\mu_R}$. For the multiplicative model, the coefficient of variation is defined as $\sqrt{e^{(\sigma^2)} - 1}$. You can estimate σ by $\hat{\sigma}$, where $\hat{\sigma}^2$ is the residual variance of the logarithmically transformed observations. That is, σ can be estimated by \sqrt{MSE} from the ANOVA of the transformed observations.

To produce sample size computations for the preceding problem, follow these steps:

1. Select **N**.
2. Select **Additive**.
3. Enter 0, 0.05, 0.10, and 0.15 as values for **Null difference**:
4. Enter 0.20 for **Coeff of variation**:
5. Enter 0.05 as the **Alpha**:
6. Enter 0.80 as the **Power**:
7. Enter -0.2 and 0.2 as the values for **Lower:** and **Upper:**, respectively, for the **Equivalence bounds**.
8. Click **OK** to perform the analysis.

Figure 12.11 displays the completed dialog.

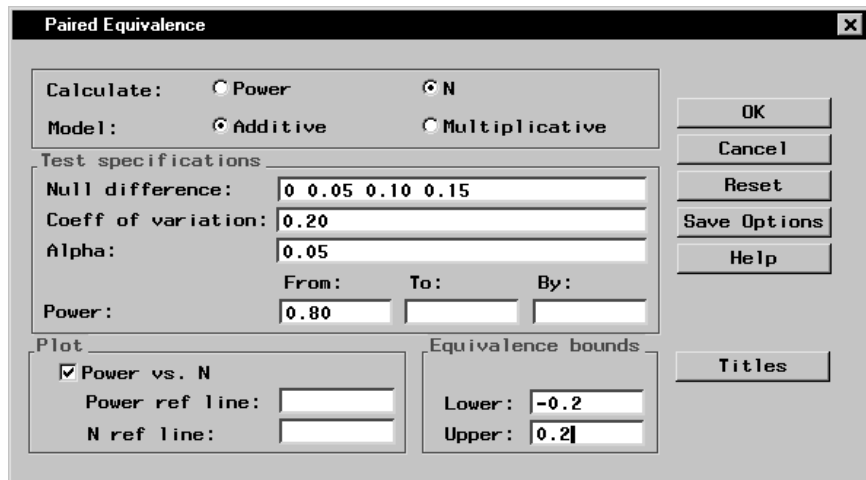


Figure 12.11. Sample Size Dialog for Paired Equivalence

The results are displayed in Figure 12.12.

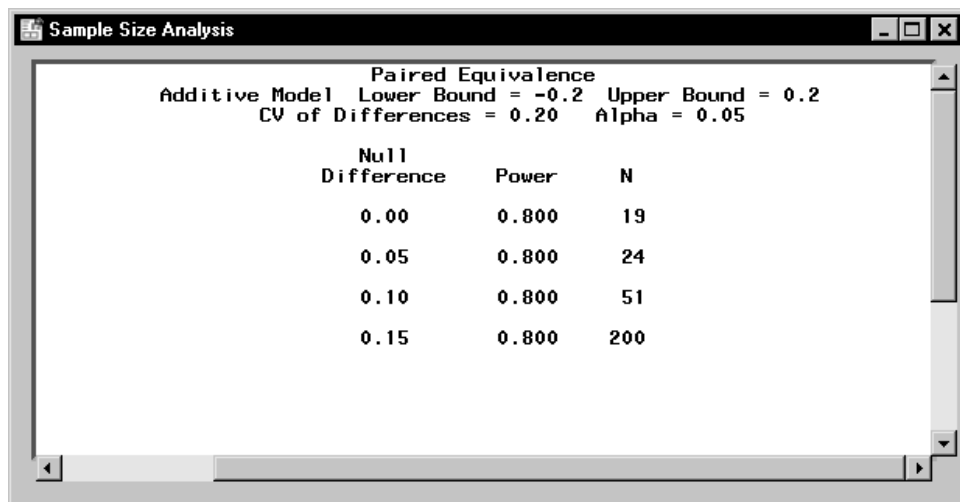


Figure 12.12. Results for Paired Equivalence

The results consist of the sample sizes for a power of 0.80 for the values of the null difference, as displayed in Figure 12.12. These results are for the alpha level of 0.05. For a null difference of 0.10, the sample size is 51. For a null difference of 0.15, the sample size jumps to 200.

One-Way ANOVA

When you are planning to analyze data from more than two groups with a one-way ANOVA, you need to calculate your sample size and power accordingly. These computations are available, prospectively, for use in the planning stages of the study, using the Sample Size task. Retrospective calculations are available, for use in the analysis stage, from the One-Way ANOVA task. This section discusses the prospective computations available in the Analyst Application, which assume equally sized groups.

You must supply two quantities in order to produce these computations: the corrected sum of squares of means (CSS) and the standard deviation. CSS is calculated as

$$CSS = \sum_{g=1}^G (\mu_g - \mu_{.})^2$$

where

μ_g = mean of the g th group

$\mu_{.}$ = overall mean

You must enter one or more values for the standard deviation, which in this case is the square root of the Mean Squared Error (MSE).

Requesting Power Computations for ANOVA

The following is an example of calculating the power for a one-way ANOVA with specified values of sample size. Suppose that you are comparing three groups, the overall mean is 5.5, and the group means

are 4.5, 5.5, and 6.5. Therefore, the corrected sum of squares of means (CSS) is

$$(4.5 - 5.5)^2 + (5.5 - 5.5)^2 + (6.5 - 5.5)^2 = 2$$

The standard deviation is the square root of the MSE, which is 1.4142. You are interested in studying sample sizes that range from 6 to 20.

To perform these computations, select

Statistics → **Sample Size** → **One-Way ANOVA . . .**

Figure 12.13 displays the resulting dialog. For this analysis, you need to enter the number of treatments, or factor levels, the CSS of means, the standard deviation, and the alpha level.

Figure 12.13. Sample Size Dialog for One-Way ANOVA

To produce power computations for the preceding problem, follow these steps:

1. Select **Power**.
2. Enter 3 for **# of treatments**:
3. Enter 2 for **CSS of means**:
4. Enter 1.4142 for **Standard deviation**:
5. Enter 0.05 for **Alpha**:
6. Enter 6, 20, and 1 for the fields **N per group**:, for **From**:, **To**:, and **By**:, respectively.
7. Click **OK** to perform the analysis.

Figure 12.14 displays the completed dialog.

The screenshot shows the 'One-Way ANOVA' dialog box. The 'Calculate:' section has 'Power' selected with a radio button. The 'Test specifications' section contains the following fields: '# of treatments:' with value 3, 'CSS of means:' with value 2, 'Standard deviation:' with value 1.4142, and 'Alpha:' with value 0.05. The 'N per group:' section has three sub-fields: 'From:' with value 6, 'To:' with value 20, and 'By:' with value 1. The 'Plot' section has a checked box for 'Power vs. N per group' and empty fields for 'Power ref line:' and 'N ref line:'. On the right side, there are buttons for 'OK', 'Cancel', 'Reset', 'Save Options', 'Help', and 'Titles'.

Figure 12.14. Sample Size Dialog for One-Way ANOVA

Requested are power computations for sample sizes ranging from 6 to 20.

The results are displayed in Figure 12.15.

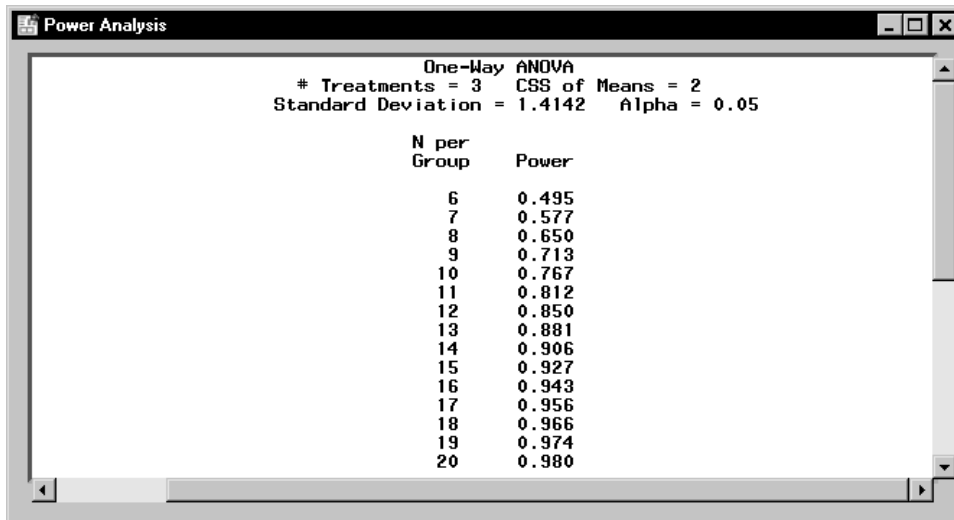


Figure 12.15. Results for Power Computations for One-Way ANOVA

Note that, to achieve a minimum of 80% power, 11 units per group would be needed.

Power Computation Details

This section provides information on how the power is computed in the Analyst Application. When you request that sample size be computed, the computations produce the smallest sample size that provides the specified power.

Hypothesis Tests

The power for the one-sample t -test, the paired t -test, and the two-sample t -test is computed in the usual fashion. That is, power is the probability of correctly rejecting the null hypothesis when the alternative is true. The sample size is the number per group; these calculations assume equally sized groups. To compute the power of a t -test, you make use of the noncentral t distribution. The formula

(O'Brien and Lohr 1984) is given by

$$\text{Power} = \text{Prob}(t < t_{crit}, \nu, NC)$$

for a one-sided alternative hypothesis and

$$\text{Power} = \text{Prob}(t < t_{critu}, \nu, NC) + 1 - \text{Prob}(t < t_{critl}, \nu, NC)$$

for a two-sided alternative hypothesis where t is distributed as non-central $t(NC, \nu)$.

$t_{crit} = t_{(1-\alpha, \nu)}$ is the $(1 - \alpha)$ quantile of the t distribution with ν df

$t_{critu} = t_{(1-\alpha/2, \nu)}$ is the $(1-\alpha/2)$ quantile of the t distribution with ν df

$t_{critl} = t_{(\alpha/2, \nu)}$ is the $(\alpha/2)$ quantile of the t distribution with ν df

For one sample and paired samples,

$$\begin{aligned} \nu &= n - 1 && \text{are the df} \\ NC &= \delta \sqrt{n} && \text{is the noncentrality parameter} \end{aligned}$$

For two samples,

$$\begin{aligned} \nu &= 2(n - 1) && \text{df} \\ NC &= \frac{\delta}{\sqrt{2/n}} && \text{the noncentrality parameter} \end{aligned}$$

Note that n equals the sample size (number per group).

The other parameters are

$$\delta = \begin{cases} \frac{|\mu_a - \mu_0|}{s} & \text{for one-sample} \\ \frac{(\mu_1 - \mu_2)}{s} & \text{for two-sample and paired samples} \end{cases}$$

$$s = \begin{cases} \text{standard deviation for one-sample} \\ \text{standard deviation of the differences for paired samples} \\ \text{pooled standard deviation for two samples} \end{cases}$$

One-Way ANOVA

The power for the one-way ANOVA is computed in a similar manner as for the hypothesis tests. That is, power is the probability of correctly rejecting the null (all group means are equal) in favor of the alternative hypothesis (at least one group mean is not equal), when the alternative is true. The sample size is the number per group; these calculations assume equally sized groups. To compute the power, you make use of the noncentral F distribution. The formula (O'Brien and Lohr 1984) is given by

$$\text{Power} = \text{Prob}(F < F_{crit}, \nu_1, \nu_2, NC)$$

where F is distributed as the noncentral $F(NC, -\nu_1, -\nu_2)$ and $F_{crit} = F_{(1-\alpha, \nu_1, \nu_2)}$ is the $(1 - \alpha)$ quantile of the F distribution with ν_1 and ν_2 degrees of freedom.

$\nu_1 = r - 1$ is the numerator df

$\nu_2 = r(n - 1)$ is the denominator df

n is the number per group

r is the number of groups

$NC = \frac{nCSS}{\sigma^2}$ is the noncentrality parameter

$CSS = \sum_{g=1}^G (\mu_g - \mu_{.})^2$ is the corrected sum of squares

μ_g is the mean of the g th group

$\mu_{.}$ is the overall mean

σ^2 is estimated by the mean squared error (MSE)

Confidence Intervals

Power calculations are available when the proposed analysis is construction of confidence intervals of a mean (one-sample) or difference of two means (two-samples or paired-samples). To understand the power of a confidence interval, first define the *precision* to be half the length of a two-sided confidence interval (or the distance between the endpoint and the parameter estimate in a one-sided interval). The power can then be considered to be the probability that the desired precision is achieved, that is, the probability that the length of the two-sided interval is no more than twice the desired precision. Here, a slight modification of this concept is used. The power is considered to be the conditional probability that the desired precision is achieved, given that the interval includes the true value of the parameter of interest. The reason for the modification is that there is no reason to want the interval to be particularly small if it does not contain the true value of the parameter.

To compute the power of a confidence interval or an equivalence test, you make use of Owen's Q formula (Owen 1965). The formula is given by

$$Q_\nu(t, \delta; a, b) = \frac{\sqrt{2\pi}}{\Gamma(\frac{\nu}{2})2^{(\nu-2)/2}} \int_a^b \Phi\left(\frac{tx}{\sqrt{\nu}} - \delta\right) x^{\nu-1} \phi(x) dx$$

where

$$\Phi = \int_{-\infty}^x \phi(t) dt$$

and

$$\phi(x) = \frac{1}{\sqrt{(2\pi)}} e^{(-x^2/2)}$$

The power of a confidence interval (Beal 1989) is given by

$$\text{Power} = \frac{2[Q_\nu(t_c, 0; 0, B) - Q_\nu(0, 0; 0, B)]}{1 - \alpha_s}$$

where

$t_c = t_{(1-\alpha/2, \nu)}$ is the $(1 - \alpha/2)$ quantile of a t distribution with ν df for a two-sided confidence interval

$t_c = t_{(1-\alpha, \nu)}$ is the $(1 - \alpha)$ quantile of a t distribution with ν degrees of freedom for a one-sided confidence interval

α is the confidence level

$$\alpha_s = \begin{cases} \alpha & \text{for a two-sided confidence interval} \\ 2\alpha & \text{for a one-sided confidence interval} \end{cases}$$

$$B = \frac{\delta\sqrt{\nu}}{t_c\kappa}$$

$$\left. \begin{array}{l} \nu = n - 1 \\ \kappa = \sqrt{1/n} \end{array} \right\} \text{for the one-sample and paired confidence intervals}$$

$$\left. \begin{array}{l} \nu = 2(n - 1) \\ \kappa = \sqrt{2/n} \end{array} \right\} \text{for the two-sample confidence interval}$$

$\delta = \frac{\text{desired precision}}{\text{standard deviation}}$ is the upper bound of the interval half-length

Equivalence Tests

In a test of equivalence, a treatment mean and a reference mean are compared to each other. Equivalence is taken to be the alternative hypothesis, and the null hypothesis is nonequivalence. The power of a test is the probability of rejecting the null hypothesis when the alternative is true, so in this case, the power is the probability of failing to reject equivalence when the treatments are in fact equivalent, that is, the treatment difference or ratio is within the prespecified boundaries.

The computational details for the power of an equivalence test (refer to Phillips 1990 for the additive model; Diletti, Hauschke, and Steinijans 1991 for the multiplicative) are as follows:

$$\text{Power} = \text{Prob}(t_1 \geq t_{(1-\alpha, n-2)} \text{ and } t_2 \leq -t_{(1-\alpha, n-2)} | \text{bioequivalence})$$

Owen (1965) showed that (t_1, t_2) has a bivariate noncentral t distribution that can be calculated as the difference of two definite integrals (Owen's Q function):

$$\text{Power} = Q_v(-t_{(1-\alpha, \nu)}, \delta_2; 0, \mathbf{R}) - Q_v(t_{(1-\alpha, \nu)}, \delta_1; 0, \mathbf{R})$$

where $t_{(1-\alpha, \nu)}$ is the $(1 - \alpha)$ quantile of a t distribution with ν df.

$$\nu = \begin{cases} n - 1 & \text{for the one-sample test} \\ n - 2 & \text{for the paired test} \\ 2(n - 1) & \text{for the two-sample test} \end{cases}$$

and

$$\left. \begin{array}{l} \delta_1 = \frac{\theta - b_l}{V \cdot \kappa} \\ \delta_2 = \frac{\theta - b_u}{V \cdot \kappa} \\ \theta = \text{null difference} \end{array} \right\} \text{for the additive model}$$

$$\left. \begin{array}{l} \delta_1 = \frac{\log(\theta) - \log(b_l)}{\kappa \sqrt{\log(1+V^2)}} \\ \delta_2 = \frac{\log(\theta) - \log(b_u)}{\kappa \sqrt{\log(1+V^2)}} \\ \theta = \text{null ratio} \end{array} \right\} \text{for the multiplicative model}$$

V is the coefficient of variation
 b_l is the lower equivalence bound
 b_u is the upper equivalence bound

$$\kappa = \begin{cases} \sqrt{1/n} & \text{for the one-sample test} \\ \sqrt{2/n} & \text{for the two-sample and paired tests} \end{cases}$$

$$R = \frac{\sqrt{\nu}(\delta_1 - \delta_2)}{2 \cdot t_{(1-\alpha, \nu)}}$$

For equivalence tests, alpha is usually set to 0.05, and power ranges from 0.70 to 0.90 (often set to 0.80).

For the **additive model** of equivalence, the values you must enter for the null difference, the coefficient of variation (c.v.), and the lower and upper bioequivalence limits must be expressed as percentages of the reference mean. More information on specifications follow:

Calculate the null difference as $\frac{|\mu_T - \mu_R|}{\mu_R}$, where μ_T is the hypothesized treatment mean and μ_R is the hypothesized reference mean. The null difference is often in the range of 0 to 0.20.

For the coefficient of variation value, σ can be estimated by $\hat{\sigma}$, where $\hat{\sigma}^2$ is the residual variance of the observations (MSE). Enter the c.v. as a percentage of the reference mean, so for the c.v., enter $\frac{\hat{\sigma}}{\mu_R}$, or $\frac{\sqrt{MSE}}{\mu_R}$. This value is often in the range of 0.05 to 0.30.

Enter the bioequivalence lower and upper limits as percentages of the reference mean as well. That is, for the bounds, enter $\frac{\text{lowerbound}}{\mu_R}$ and $\frac{\text{upperbound}}{\mu_R}$. These values are often -0.2 and 0.2 , respectively.

For the **multiplicative model** of equivalence, calculate the null ratio as $\frac{\mu_T}{\mu_R}$, where μ_T is the hypothesized treatment mean and μ_R is the hypothesized reference mean. This value is often in the range of 0.80 to 1.20. More information on specifications follow:

The coefficient of variation (c.v.) is defined as $\sqrt{e^{(\sigma^2)} - 1}$. You can estimate σ by $\hat{\sigma}$, where $\hat{\sigma}^2$ is the residual variance of the logarithmically transformed observations. That is, σ can be estimated by \sqrt{MSE} from the ANOVA of the transformed observations. The c.v. value is often in the range of 0.05 to 0.30.

The bioequivalence lower and upper limits are often set to 0.80 and 1.25, respectively.

References

- Beal, S.L. (1989), "Sample Size Determination for Confidence Intervals on the Population Mean and on the Differences Between Two Population Means," *Biometrics*, 45, 969–977.
- Diletti, E., Hauschke, D., and Steinijans, V.W. (1991), "Sample Size Determination for Bioequivalence Assessment by Means of Confidence Intervals," *International Journal of Clinical Pharmacology, Therapy and Toxicology*, Vol. 29, 1–8.
- O'Brien, R., and Lohr, V. (1984), "Power Analysis For Linear Models: The Time Has Come," *Proceedings of the Ninth Annual SAS User's Group International Conference*, 840–846.
- Owen, D.B. (1965), "A Special Case of a Bivariate Non-central t -distribution," *Biometrika*, 52, 437–446.
- Phillips, K.F. (1990), "Power of the Two One-Sided Tests Procedure in Bioequivalence," *Journal of Pharmacokinetics and Biopharmaceutics*, Vol. 18, No. 2, 137–144.

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

The Analyst Application, First Edition

Copyright © 1999 SAS Institute Inc., Cary, NC, USA.

ISBN 1-58025-446-2

All rights reserved. Printed in the United States of America. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, by any form or by any means, electronic, mechanical, photocopying, or otherwise, without the prior written permission of the publisher, SAS Institute, Inc.

U.S. Government Restricted Rights Notice. Use, duplication, or disclosure of the software by the government is subject to restrictions as set forth in FAR 52.227-19 Commercial Computer Software-Restricted Rights (June 1987).

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

1st printing, October 1999

SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries.® indicates USA registration.

IBM®, ACF/VTAM®, AIX®, APPN®, MVS/ESA®, OS/2®, OS/390®, VM/ESA®, and VTAM® are registered trademarks or trademarks of International Business Machines Corporation.® indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.

The Institute is a private company devoted to the support and further development of its software and related services.