

Chapter 450

Non-Inferiority Tests for the Difference Between Two Means

Introduction

This procedure computes power and sample size for *non-inferiority* tests in two-sample designs in which the outcome is a continuous normal random variable. Measurements are made on individuals that have been randomly assigned to one of two groups. This is sometimes referred to as a *parallel-groups* design. This design is used in situations such as the comparison of the income level of two regions, the nitrogen content of two lakes, or the effectiveness of two drugs.

The two-sample t-test is commonly used with this situation. When the variances of the two groups are unequal, Welch's t-test may be used. When the data are not normally distributed, the Mann-Whitney (Wilcoxon signed-ranks) U test may be used.

The details of sample size calculation for the two-sample design are presented in the Two-Sample T-Test chapter and they will not be duplicated here. This chapter only discusses those changes necessary for non-inferiority and superiority tests. Sample size formulas for non-inferiority and superiority tests of two means are presented in Chow et al. (2003) pages 57-59.

The Statistical Hypotheses

Both non-inferiority and superiority tests are examples of directional (one-sided) tests and their power and sample size could be calculated using the *Two-Sample T-Test* procedure. However, at the urging of our users, we have developed this module, which provides the input and output in formats that are convenient for these types of tests. This section will review the specifics of non-inferiority and superiority testing.

Remember that in the usual t-test setting, the null (H_0) and alternative (H_1) hypotheses for one-sided tests are defined as

$$H_0: \mu_1 - \mu_2 \leq D \text{ versus } H_1: \mu_1 - \mu_2 > D$$

Rejecting this test implies that the mean difference is larger than the value D . This test is called an *upper-tailed test* because it is rejected in samples in which the difference between the sample means is larger than D .

Following is an example of a *lower-tailed test*.

$$H_0: \mu_1 - \mu_2 \geq D \text{ versus } H_1: \mu_1 - \mu_2 < D$$

Non-inferiority and *superiority* tests are special cases of the above directional tests. It will be convenient to adopt the following specialized notation for the discussion of these tests.

Non-Inferiority Tests for the Difference Between Two Means

<u>Parameter</u>	<u>PASS Input/Output</u>	<u>Interpretation</u>
μ_1	Not used	<i>Mean</i> of population 1. Population 1 is assumed to consist of those who have received the new treatment.
μ_2	Not used	<i>Mean</i> of population 2. Population 2 is assumed to consist of those who have received the reference treatment.
M_{NI}	NIM	<i>Margin of non-inferiority</i> . This is a tolerance value that defines the magnitude of the amount that is not of practical importance. This may be thought of as the largest change from the baseline that is considered to be trivial. The absolute value is shown to emphasize that this is a magnitude. The sign of the value will be determined by the specific design that is being used.
δ	D	<i>True difference</i> . This is the value of $\mu_1 - \mu_2$, the difference between the means. This is the value at which the power is calculated.

Note that the actual values of μ_1 and μ_2 are not needed. Only their difference is needed for power and sample size calculations.

Non-Inferiority Tests

A *non-inferiority test* tests that the treatment mean is not worse than the reference mean by more than the equivalence margin. The actual direction of the hypothesis depends on the response variable being studied.

Case 1: High Values Good, Non-Inferiority Test

In this case, higher values are better. The hypotheses are arranged so that rejecting the null hypothesis implies that the treatment mean is no less than a small amount below the reference mean. The value of δ is often set to zero. The following are equivalent sets of hypotheses.

$$H_0 : \mu_1 \leq \mu_2 - |M_{NI}| \quad \text{versus} \quad H_1 : \mu_1 > \mu_2 - |M_{NI}|$$

$$H_0 : \mu_1 - \mu_2 \leq -|M_{NI}| \quad \text{versus} \quad H_1 : \mu_1 - \mu_2 > -|M_{NI}|$$

$$H_0 : \delta \leq -|M_{NI}| \quad \text{versus} \quad H_1 : \delta > -|M_{NI}|$$

Case 2: High Values Bad, Non-Inferiority Test

In this case, lower values are better. The hypotheses are arranged so that rejecting the null hypothesis implies that the treatment mean is no more than a small amount above the reference mean. The value of δ is often set to zero. The following are equivalent sets of hypotheses.

$$H_0 : \mu_1 \geq \mu_2 + |M_{NI}| \quad \text{versus} \quad H_1 : \mu_1 < \mu_2 + |M_{NI}|$$

$$H_0 : \mu_1 - \mu_2 \geq |M_{NI}| \quad \text{versus} \quad H_1 : \mu_1 - \mu_2 < |M_{NI}|$$

$$H_0 : \delta \geq |M_{NI}| \quad \text{versus} \quad H_1 : \delta < |M_{NI}|$$

Non-Inferiority Tests for the Difference Between Two Means

Example

A non-inferiority test example will set the stage for the discussion of the terminology that follows. Suppose that a test is to be conducted to determine if a new cancer treatment adversely affects mean bone density. The adjusted mean bone density (AMBD) in the population of interest is 0.002300 gm/cm with a standard deviation of 0.000300 gm/cm. Clinicians decide that if the treatment reduces AMBD by more than 5% (0.000115 gm/cm), it poses a significant health threat.

The hypothesis of interest is whether the mean AMBD in the treated group is more than 0.000115 below that of the reference group. The statistical test will be set up so that if the null hypothesis is rejected, the conclusion will be that the new treatment is non-inferior. The value 0.000115 gm/cm is called the *margin of non-inferiority*.

Test Statistics

This section describes the test statistics that are available in this procedure.

Two-Sample T-Test

Under the null hypothesis, this test assumes that the two groups of data are simple random samples from a single population of normally-distributed values that all have the same mean and variance. This assumption implies that the data are continuous and their distribution is symmetric. The calculation of the test statistic for the case when higher response values are good is as follows.

$$t_{df} = \frac{(\bar{X}_1 - \bar{X}_2) - |\varepsilon|}{s_{\bar{X}_1 - \bar{X}_2}}$$

where

$$\bar{X}_k = \frac{\sum_{i=1}^{N_k} X_{ki}}{N_k}$$

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sum_{i=1}^{N_1} (X_{1i} - \bar{X}_1)^2 + \sum_{i=1}^{N_2} (X_{2i} - \bar{X}_2)^2}{N_1 + N_2 - 2} \left(\frac{1}{N_1} + \frac{1}{N_2} \right)}$$

$$df = N_1 + N_2 - 2$$

The null hypothesis is rejected if the computed p-value is less than a specified level (usually 0.05). Otherwise, no conclusion can be reached.

Non-Inferiority Tests for the Difference Between Two Means

Welch's T-Test

Welch (1938) proposed the following test when the two variances are not assumed to be equal.

$$t_f^* = \frac{(\bar{X}_1 - \bar{X}_2) - |\epsilon|}{s_{\bar{X}_1 - \bar{X}_2}^*}$$

where

$$s_{\bar{X}_1 - \bar{X}_2}^* = \sqrt{\left(\frac{\sum_{i=1}^{N_1} (X_{1i} - \bar{X}_1)^2}{N_1(N_1 - 1)} \right) + \left(\frac{\sum_{i=1}^{N_2} (X_{2i} - \bar{X}_2)^2}{N_2(N_2 - 1)} \right)}$$

$$f = \frac{\left(\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2} \right)^2}{\frac{s_1^4}{N_1^2(N_1 - 1)} + \frac{s_2^4}{N_2^2(N_2 - 1)}}$$

$$s_1 = \sqrt{\left(\frac{\sum_{i=1}^{N_1} (X_{1i} - \bar{X}_1)^2}{N_1 - 1} \right)}, \quad s_2 = \sqrt{\left(\frac{\sum_{i=1}^{N_2} (X_{2i} - \bar{X}_2)^2}{N_2 - 1} \right)}$$

Mann-Whitney U Test

This test is the nonparametric substitute for the equal-variance t-test. Two key assumptions are that the distributions are at least ordinal and that they are identical under H₀. This means that ties (repeated values) are not acceptable. When ties are present, you can use approximations, but the theoretic results no longer hold.

The Mann-Whitney test statistic is defined as follows in Gibbons (1985).

$$z = \frac{W_1 - \frac{N_1(N_1 + N_2 + 1)}{2} + C}{s_W}$$

where

$$W_1 = \sum_{k=1}^{N_1} \text{Rank}(X_{1k})$$

The ranks are determined after combining the two samples. The standard deviation is calculated as

$$s_W = \sqrt{\frac{N_1 N_2 (N_1 + N_2 + 1)}{12} - \frac{N_1 N_2 \sum_{i=1} (t_i^3 - t_i)}{12(N_1 + N_2)(N_1 + N_2 - 1)}}$$

where t_i is the number of observations tied at value one, t_2 is the number of observations tied at some value two, and so forth.

The correction factor, C , is 0.5 if the rest of the numerator is negative or -0.5 otherwise. The value of z is then compared to the normal distribution.

Computing the Power

Standard Deviations Equal

When $\sigma_1 = \sigma_2 = \sigma$, the power of the t test is calculated as follows.

1. Find t_α such that $1 - T_{df}(t_\alpha) = \alpha$, where $T_{df}(t_\alpha)$ is the area under a central- t curve to the left of x and $df = N_1 + N_2 - 2$.
2. Calculate: $\sigma_{\bar{x}} = \sigma \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}$
3. Calculate the noncentrality parameter: $\lambda = \frac{|\varepsilon| - \delta}{\sigma_{\bar{x}}}$
4. Calculate: Power = $1 - T'_{df,\lambda}(t_\alpha)$, where $T'_{df,\lambda}(x)$ is the area to the left of x under a noncentral- t curve with degrees of freedom df and noncentrality parameter λ .

Standard Deviations Unequal

This case often recommends Welch's test. When $\sigma_1 \neq \sigma_2$, the power is calculated as follows.

1. Calculate: $\sigma_{\bar{x}} = \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}$.
2. Calculate: $f = \frac{\sigma_{\bar{x}}^4}{\frac{\sigma_1^4}{N_1^2(N_1+1)} + \frac{\sigma_2^4}{N_2^2(N_2+1)}} - 2$

which is the adjusted degrees of freedom. Often, this is rounded to the next highest integer. Note that this is not the value of f used in the computation of the actual test. Instead, this is the expected value of f .

3. Find t_α such that $1 - T_f(t_\alpha) = \alpha$, where $T_f(t_\alpha)$ is the area to the left of x under a central- t curve with f degrees of freedom.
4. Calculate: $\lambda = \frac{|\varepsilon|}{\sigma_{\bar{x}}}$, the noncentrality parameter.
5. Calculate: Power = $1 - T'_{f,\lambda}(t_\alpha)$, where $T'_{f,\lambda}(x)$ is the area to the left of x under a noncentral- t curve with degrees of freedom f and noncentrality parameter λ .

Nonparametric Adjustment

When using the Mann-Whitney test rather than the t test, results by Al-Sundugchi and Guenther (1990) indicate that power calculations for the Mann-Whitney test may be made using the standard t test formulations with a simple adjustment to the sample sizes. The size of the adjustment depends on the actual distribution of the data. They give sample size adjustment factors for four distributions. These are 1 for uniform, $2/3$ for double exponential, $9/\pi^2$ for logistic, and $\pi/3$ for normal distributions.

Procedure Options

This section describes the options that are specific to this procedure. These are located on the Design tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Design Tab

The Design tab contains most of the parameters and options that you will be concerned with.

Solve For

Solve For

This option specifies the parameter to be calculated from the values of the other parameters. Under most conditions, you would select either *Sample Size (NI)*.

Select *Sample Size (NI)* when you want to determine the sample size needed to achieve a given power and alpha.

Select *Power* when you want to calculate the power of an experiment that has already been run.

Test

Higher Means Are

This option defines whether higher values of the response variable are to be considered better or worse. The choice here determines the direction of the non-inferiority test.

If Higher Means Are Better the null hypothesis is $\text{Diff} \leq -\text{NIM}$ and the alternative hypothesis is $\text{Diff} > -\text{NIM}$. If Higher Means Are Worse the null hypothesis is $\text{Diff} \geq \text{NIM}$ and the alternative hypothesis is $\text{Diff} < \text{NIM}$.

Nonparametric Adjustment (Mann-Whitney Test)

This option makes appropriate sample size adjustments for the Mann-Whitney test. Results by Al-Sundugchi and Guenther (1990) indicate that power calculations for the Mann-Whitney test may be made using the standard t test formulations with a simple adjustment to the sample size. The size of the adjustment depends upon the actual distribution of the data. They give sample size adjustment factors for four distributions. These are 1 for the uniform distribution, $2/3$ for the double exponential distribution, $9 / \pi^2$ for the logistic distribution, and $\pi / 3$ for the normal distribution.

The options are as follows:

- **Ignore**
Do not make a Mann-Whitney adjustment. This indicates that you want to analyze a t test, not the Wilcoxon test.
- **Uniform**
Make the Mann-Whitney sample size adjustment assuming the uniform distribution. Since the factor is one, this option performs the same function as Ignore. It is included for completeness.
- **Double Exponential**
Make the Mann-Whitney sample size adjustment assuming that the data actually follow the double exponential distribution.

Non-Inferiority Tests for the Difference Between Two Means

- **Logistic**
Make the Mann-Whitney sample size adjustment assuming that the data actually follow the logistic distribution.
- **Normal**
Make the Mann-Whitney sample size adjustment assuming that the data actually follow the normal distribution.

Power and Alpha

Power

This option specifies one or more values for power. Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected. In this procedure, a type-II error occurs when you fail to reject the null hypothesis of inferiority when the null hypothesis should be rejected.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as *0.8 to 0.95 by 0.05* may be entered.

Alpha

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected. In this procedure, a type-I error occurs when you reject the null hypothesis of inferiority when in fact the mean is not non-inferior.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as *0.01 0.05 0.10* or *0.01 to 0.10 by 0.01*.

Sample Size (When Solving for Sample Size)

Group Allocation

Select the option that describes the constraints on $N1$ or $N2$ or both.

The options are

- **Equal ($N1 = N2$)**
This selection is used when you wish to have equal sample sizes in each group. Since you are solving for both sample sizes at once, no additional sample size parameters need to be entered.
- **Enter $N2$, solve for $N1$**
Select this option when you wish to fix $N2$ at some value (or values), and then solve only for $N1$. Please note that for some values of $N2$, there may not be a value of $N1$ that is large enough to obtain the desired power.
- **Enter $R = N2/N1$, solve for $N1$ and $N2$**
For this choice, you set a value for the ratio of $N2$ to $N1$, and then PASS determines the needed $N1$ and $N2$, with this ratio, to obtain the desired power. An equivalent representation of the ratio, R , is

$$N2 = R * N1.$$

Non-Inferiority Tests for the Difference Between Two Means

- **Enter percentage in Group 1, solve for N1 and N2**

For this choice, you set a value for the percentage of the total sample size that is in Group 1, and then PASS determines the needed $N1$ and $N2$ with this percentage to obtain the desired power.

N2 (Sample Size, Group 2)

This option is displayed if Group Allocation = "Enter N2, solve for N1"

$N2$ is the number of items or individuals sampled from the Group 2 population.

$N2$ must be ≥ 2 . You can enter a single value or a series of values.

R (Group Sample Size Ratio)

This option is displayed only if Group Allocation = "Enter R = N2/N1, solve for N1 and N2."

R is the ratio of $N2$ to $N1$. That is,

$$R = N2 / N1.$$

Use this value to fix the ratio of $N2$ to $N1$ while solving for $N1$ and $N2$. Only sample size combinations with this ratio are considered.

$N2$ is related to $N1$ by the formula:

$$N2 = [R \times N1],$$

where the value $[Y]$ is the next integer $\geq Y$.

For example, setting $R = 2.0$ results in a Group 2 sample size that is double the sample size in Group 1 (e.g., $N1 = 10$ and $N2 = 20$, or $N1 = 50$ and $N2 = 100$).

R must be greater than 0. If $R < 1$, then $N2$ will be less than $N1$; if $R > 1$, then $N2$ will be greater than $N1$. You can enter a single or a series of values.

Percent in Group 1

This option is displayed only if Group Allocation = "Enter percentage in Group 1, solve for N1 and N2."

Use this value to fix the percentage of the total sample size allocated to Group 1 while solving for $N1$ and $N2$. Only sample size combinations with this Group 1 percentage are considered. Small variations from the specified percentage may occur due to the discrete nature of sample sizes.

The Percent in Group 1 must be greater than 0 and less than 100. You can enter a single or a series of values.

Sample Size (When Not Solving for Sample Size)

Group Allocation

Select the option that describes how individuals in the study will be allocated to Group 1 and to Group 2.

The options are

- **Equal ($N1 = N2$)**

This selection is used when you wish to have equal sample sizes in each group. A single per group sample size will be entered.

- **Enter N1 and N2 individually**

This choice permits you to enter different values for $N1$ and $N2$.

- **Enter N1 and R, where $N2 = R * N1$**

Choose this option to specify a value (or values) for $N1$, and obtain $N2$ as a ratio (multiple) of $N1$.

Non-Inferiority Tests for the Difference Between Two Means

- **Enter total sample size and percentage in Group 1**

Choose this option to specify a value (or values) for the total sample size (N), obtain $N1$ as a percentage of N , and then $N2$ as $N - N1$.

Sample Size Per Group

This option is displayed only if Group Allocation = "Equal ($N1 = N2$)."

The Sample Size Per Group is the number of items or individuals sampled from each of the Group 1 and Group 2 populations. Since the sample sizes are the same in each group, this value is the value for $N1$, and also the value for $N2$.

The Sample Size Per Group must be ≥ 2 . You can enter a single value or a series of values.

N1 (Sample Size, Group 1)

*This option is displayed if Group Allocation = "Enter $N1$ and $N2$ individually" or "Enter $N1$ and R , where $N2 = R * N1$."*

$N1$ is the number of items or individuals sampled from the Group 1 population.

$N1$ must be ≥ 2 . You can enter a single value or a series of values.

N2 (Sample Size, Group 2)

This option is displayed only if Group Allocation = "Enter $N1$ and $N2$ individually."

$N2$ is the number of items or individuals sampled from the Group 2 population.

$N2$ must be ≥ 2 . You can enter a single value or a series of values.

R (Group Sample Size Ratio)

*This option is displayed only if Group Allocation = "Enter $N1$ and R , where $N2 = R * N1$."*

R is the ratio of $N2$ to $N1$. That is,

$$R = N2/N1$$

Use this value to obtain $N2$ as a multiple (or proportion) of $N1$.

$N2$ is calculated from $N1$ using the formula:

$$N2 = [R \times N1],$$

where the value $[Y]$ is the next integer $\geq Y$.

For example, setting $R = 2.0$ results in a Group 2 sample size that is double the sample size in Group 1.

R must be greater than 0. If $R < 1$, then $N2$ will be less than $N1$; if $R > 1$, then $N2$ will be greater than $N1$. You can enter a single value or a series of values.

Total Sample Size (N)

This option is displayed only if Group Allocation = "Enter total sample size and percentage in Group 1."

This is the total sample size, or the sum of the two group sample sizes. This value, along with the percentage of the total sample size in Group 1, implicitly defines $N1$ and $N2$.

The total sample size must be greater than one, but practically, must be greater than 3, since each group sample size needs to be at least 2.

You can enter a single value or a series of values.

Non-Inferiority Tests for the Difference Between Two Means

Percent in Group 1

This option is displayed only if Group Allocation = "Enter total sample size and percentage in Group 1."

This value fixes the percentage of the total sample size allocated to Group 1. Small variations from the specified percentage may occur due to the discrete nature of sample sizes.

The Percent in Group 1 must be greater than 0 and less than 100. You can enter a single value or a series of values.

Effect Size – Mean Difference

NIM (Non-Inferiority Margin)

This is the magnitude of the margin of non-inferiority. It must be entered as a positive number.

When higher means are better, this value is the distance below the reference mean that is still considered non-inferior. When higher means are worse, this value is the distance above the reference mean that is still considered non-inferior.

D (True Difference)

This is the actual difference between the treatment mean and the reference mean at which power is calculated.

For non-inferiority tests, this value is often set to zero. When this value is non-zero, care should be taken that this value is consistent with whether higher means are better or worse.

Effect Size – Standard Deviations

S1 and S2 (Standard Deviations)

These options specify the values of the standard deviations for each group. When the S2 is set to S1, the EQUAL VARIANCE test is used and only S1 needs to be specified. The value of S1 will be used for S2. Otherwise, the UNEQUAL VARIANCE test is used (even if the value entered for S2 equals S1).

When these values are not known, you must supply estimates of them. Press the *SD* button to display the Standard Deviation Estimator window. This procedure will help you find appropriate values for the standard deviation.

Non-Inferiority Tests for the Difference Between Two Means

Example 1 – Power Analysis

Suppose that a test is to be conducted to determine if a new cancer treatment adversely affects bone density. The adjusted mean bone density (AMBD) in the population of interest is 0.002300 gm/cm with a standard deviation of 0.000300 gm/cm. Clinicians decide that if the treatment reduces AMBD by more than 5% (0.000115 gm/cm), it poses a significant health threat. They also want to consider what would happen if the margin of equivalence is set to 2.5% (0.0000575 gm/cm).

Following accepted procedure, the analysis will be a non-inferiority test using the t-test at the 0.025 significance level. Power to be calculated assuming that the new treatment has no effect on AMBD. Several sample sizes between 10 and 800 will be analyzed. The researchers want to achieve a power of at least 90%. All numbers have been multiplied by 10000 to make the reports and plots easier to read.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Non-Inferiority Tests for the Difference Between Two Means** procedure window by expanding **Means**, then **Two Independent Means**, then clicking on **Non-Inferiority**, and then clicking on **Non-Inferiority Tests for the Difference Between Two Means**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Power
Higher Means Are.....	Better
Nonparametric Adjustment	Ignore
Alpha.....	0.025
Group Allocation	Equal (N1 = N2)
Sample Size Per Group	10 50 100 200 300 500 600 800
NIM (Non-Inferiority Margin)	0.575 1.15
D (True Difference)	0
S1 (Standard Deviation Group 1)	3
S2 (Standard Deviation Group 2)	S1

Annotated Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results and Plots

Numeric Results for Non-Inferiority Test (H0: Diff \leq -NIM; H1: Diff $>$ -NIM)

Higher Means are Better

Test Statistic: T-Test

Power	N1	N2	N	-NIM	D	S1	S2	Alpha
0.06013	10	10	20	-0.575	0.0	3.0	3.0	0.025
0.15601	50	50	100	-0.575	0.0	3.0	3.0	0.025
0.27052	100	100	200	-0.575	0.0	3.0	3.0	0.025
0.48089	200	200	400	-0.575	0.0	3.0	3.0	0.025
0.64940	300	300	600	-0.575	0.0	3.0	3.0	0.025
0.85769	500	500	1000	-0.575	0.0	3.0	3.0	0.025
0.91295	600	600	1200	-0.575	0.0	3.0	3.0	0.025
0.96943	800	800	1600	-0.575	0.0	3.0	3.0	0.025

(report continues)

Non-Inferiority Tests for the Difference Between Two Means

References

Chow, S.C.; Shao, J.; Wang, H. 2003. *Sample Size Calculations in Clinical Research*. Marcel Dekker. New York.
 Julious, Steven A. 2004. 'Tutorial in Biostatistics. Sample sizes for clinical trials with Normal data.'
Statistics in Medicine, 23:1921-1986.

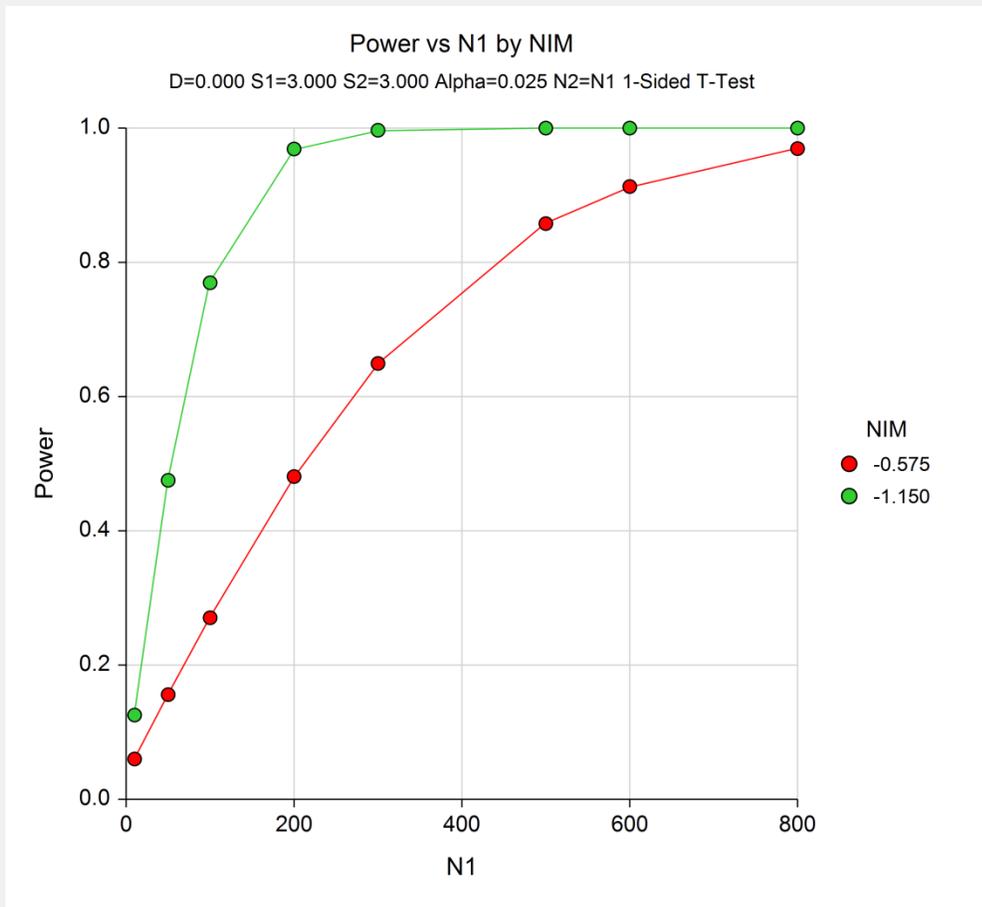
Report Definitions

Power is the probability of rejecting a false null hypothesis.
 N1 and N2 are the number of items sampled from each population.
 N is the total sample size, $N1 + N2$.
 -NIM is the magnitude and direction of the margin of non-inferiority. Since higher means are better, this value is negative and is the distance below the reference mean that is still considered non-inferior.
 D is the mean difference at which the power is computed. $D = \text{Mean1} - \text{Mean2}$, or Treatment Mean - Reference Mean.
 S1 and S2 are the assumed population standard deviations for groups 1 and 2, respectively.
 Alpha is the probability of rejecting a true null hypothesis.

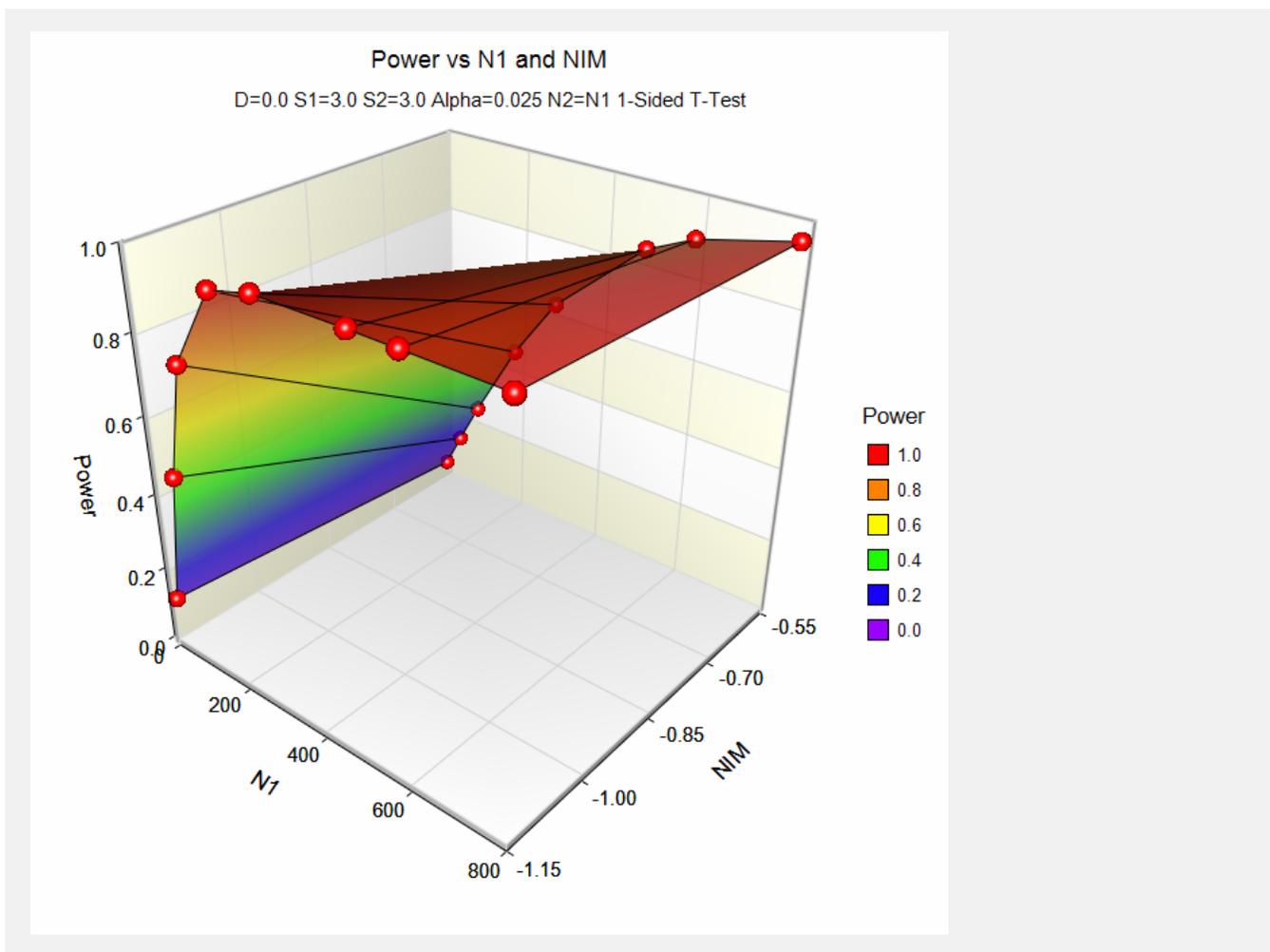
Summary Statements

Group sample sizes of 10 and 10 achieve 6% power to detect non-inferiority using a one-sided, two-sample t-test. The margin of non-inferiority is -0.575. The true difference between the means is assumed to be 0.0. The significance level (alpha) of the test is 0.025. The data are drawn from populations with standard deviations of 3.0 and 3.0.

Chart Section



Non-Inferiority Tests for the Difference Between Two Means



The above report shows that for NIM = 1.15, the sample size necessary to obtain 90% power is about 150 per group. However, if NIM = 0.575, the required sample size is about 600 per group.

Non-Inferiority Tests for the Difference Between Two Means

Example 2 – Finding the Sample Size

Continuing with Example 1, the researchers want to know the exact sample size for each value of NIM to achieve 90% power.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Non-Inferiority Tests for the Difference Between Two Means** procedure window by expanding **Means**, then **Two Independent Means**, then clicking on **Non-Inferiority**, and then clicking on **Non-Inferiority Tests for the Difference Between Two Means**. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Sample Size
Higher Means Are	Better
Nonparametric Adjustment	Ignore
Power	0.90
Alpha	0.025
Group Allocation	Equal (N1 = N2)
NIM (Non-Inferiority Margin)	0.575 1.15
D (True Difference)	0
S1 (Standard Deviation Group 1)	3
S2 (Standard Deviation Group 2)	S1

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Numeric Results for Non-Inferiority Test (H0: Diff \leq -NIM; H1: Diff $>$ -NIM)									
Higher Means are Better									
Test Statistic: T-Test									
Target Power	Actual Power	N1	N2	N	-NIM	D	S1	S2	Alpha
0.90	0.90036	573	573	1146	-0.575	0.0	3.0	3.0	0.025
0.90	0.90004	144	144	288	-1.150	0.0	3.0	3.0	0.025

This report shows the exact sample size requirement for each value of NIM.

Non-Inferiority Tests for the Difference Between Two Means

Example 3 – Validation using Chow, Shao, and Wang (2003)

Chow, Shao, and Wang (2003) page 62 has an example of a sample size calculation for a non-inferiority trial. Their example obtains a sample size of 51 in each group when $D = 0$, $NIM = 0.05$, $S = 0.1$, $\alpha = 0.05$, and $\beta = 0.20$.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Non-Inferiority Tests for the Difference Between Two Means** procedure window by expanding **Means**, then **Two Independent Means**, then clicking on **Non-Inferiority**, and then clicking on **Non-Inferiority Tests for the Difference Between Two Means**. You may then make the appropriate entries as listed below, or open **Example 3** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Sample Size
Higher Means Are	Better
Nonparametric Adjustment	Ignore
Power	0.80
Alpha	0.05
Group Allocation	Equal (N1 = N2)
NIM (Non-Inferiority Margin)	0.05
D (True Difference)	0
S1 (Standard Deviation Group 1)	0.1
S2 (Standard Deviation Group 2)	S1

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

**Numeric Results for Non-Inferiority Test (H0: Diff \leq -NIM; H1: Diff $>$ -NIM)
Higher Means are Better
Test Statistic: T-Test**

Target Power	Actual Power	N1	N2	N	-NIM	D	S1	S2	Alpha
0.80	0.80590	51	51	102	-0.050	0.0	0.1	0.1	0.050

PASS also obtains a sample size of 51 per group.

Non-Inferiority Tests for the Difference Between Two Means

Example 4 – Validation using Julious (2004)

Julious (2004) page 1950 gives an example of a sample size calculation for a parallel, non-inferiority design. His example obtains a sample size of 336 when $D = 0$, $NIM = 10$, $S = 40$, $\text{Alpha} = 0.025$, and $\text{Beta} = 0.10$.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Non-Inferiority Tests for the Difference Between Two Means** procedure window by expanding **Means**, then **Two Independent Means**, then clicking on **Non-Inferiority**, and then clicking on **Non-Inferiority Tests for the Difference Between Two Means**. You may then make the appropriate entries as listed below, or open **Example 4** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Sample Size
Higher Means Are	Better
Nonparametric Adjustment	Ignore
Power	0.90
Alpha	0.025
Group Allocation	Equal (N1 = N2)
NIM (Non-Inferiority Margin)	10
D (True Difference)	0
S1 (Standard Deviation Group 1)	40
S2 (Standard Deviation Group 2)	S1

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Numeric Results for Non-Inferiority Test ($H_0: \text{Diff} \leq -\text{NIM}$; $H_1: \text{Diff} > -\text{NIM}$)
 Higher Means are Better
 Test Statistic: T-Test

Target Power	Actual Power	N1	N2	N	-NIM	D	S1	S2	Alpha
0.90	0.90045	337	337	674	-10.000	0.0	40.0	40.0	0.025

PASS obtained sample sizes of 337 in each group. The difference between 336 that Julious received and 337 that PASS calculated is likely caused by rounding.