

Chapter 510

Non-Inferiority Tests for the Difference Between Two Means in a 2x2 Cross-Over Design

Introduction

This procedure computes power and sample size for non-inferiority tests in 2x2 cross-over designs in which the outcome is a continuous normal random variable. The details of sample size calculation for the 2x2 cross-over design are presented in the 2x2 Cross-Over Designs chapter and they will not be duplicated here. This chapter only discusses those changes necessary for non-inferiority tests. Sample size formulas for non-inferiority tests of cross-over designs are presented in Chow et al. (2003) pages 63-68.

Cross-Over Designs

Senn (2002) defines a *cross-over* design as one in which each subject receives all treatments and the objective is to study differences among the treatments. The name *cross-over* comes from the most common case in which there are only two treatments. In this case, each subject *crosses over* from one treatment to the other. It is assumed that there is a *washout* period between treatments during which the response returns back to its baseline value. If this does not occur, there is said to be a *carry-over* effect.

A 2x2 cross-over design refers to two treatments (periods) and two *sequences* (treatment orderings). One sequence receives treatment A followed by treatment B. The other sequence receives B and then A. The design includes a washout period between responses to make certain that the effects of the first drug do not carry-over to the second. Thus, the groups in this design are defined by the sequence in which the two drugs are administered, not by the treatments they receive.

Cross-over designs are employed because, if the no-carryover assumption is met, treatment differences are measured within a subject rather than between subjects—making a more precise measurement. Examples of the situations that might use a cross-over design are the comparison of anti-inflammatory drugs in arthritis and the comparison of hypotensive agents in essential hypertension. In both of these cases, symptoms are expected to return to their usual baseline level shortly after the treatment is stopped.

The Statistical Hypotheses

Both non-inferiority and superiority tests are examples of directional (one-sided) tests and their power and sample size can be calculated using the 2x2 Cross-Over Design 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 reviews the specifics of non-inferiority and superiority testing.

Remember that in the usual t-test setting, the null (H0) and alternative (H1) hypotheses for one-sided tests are defined as

$$H_0: \mu_X \leq A \text{ versus } H_1: \mu_X > A$$

Rejecting H0 implies that the mean is larger than the value A. This test is called an *upper-tailed test* because it is rejected in samples in which the difference in sample means is larger than A.

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

$$H_0: \mu_X \geq A \text{ versus } H_1: \mu_X < A$$

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

<u>Parameter</u>	<u>PASS Input/Output</u>	<u>Interpretation</u>
μ_T	Not used	<i>Treatment mean.</i> This is the treatment mean.
μ_R	Not used	<i>Reference mean.</i> This is the mean of a reference population.
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	D1	<i>True difference.</i> This is the value of $\mu_T - \mu_R$, the difference between the treatment and reference means. This is the value at which the power is calculated.

Note that the actual values of μ_T and μ_R 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.

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

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

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

Non-Inferiority Tests for the Difference Between Two Means in a 2x2 Cross-Over Design

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.

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

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

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

Test Statistics

This section describes the test statistic that is used to perform the hypothesis test.

T-Test

A t -test is used to analyze the data. When the data are balanced between sequences, the two-sided t -test is equivalent to an analysis of variance F-test. The test assumes that the data are a simple random sample from a population of normally-distributed values that have the same variance. This assumption implies that the differences are continuous and normal. The calculation of the t -statistic proceeds as follow

$$t_d = \frac{(\bar{x}_T - \bar{x}_R) - \varepsilon}{\hat{\sigma}_w \sqrt{\frac{2}{N}}}$$

where $\hat{\sigma}_w^2$ is the within mean square error from the appropriate ANOVA table.

The significance of the test statistic is determined by computing the p-value. If this p-value is less than a specified level (usually 0.05), the hypothesis is rejected. That is, the one-sided null hypothesis is rejected at the α significance level if $t_d > t_{\alpha, N-2}$. Otherwise, no conclusion can be reached.

The ANOVA F-test is calculated using a standard repeated-measures analysis of variance table in which the between factor is the sequence and the within factor is the treatment. The within mean square error provides an estimate of the within-subject variance, σ_w^2 . If prior studies used a t -test rather than an ANOVA to analyze the data, you may not have a direct estimate of σ_w^2 . Instead, you may have an estimate of the variance of the period differences from the t -test, $\hat{\sigma}_{period\ diff}^2$, or an estimate of the variance of the paired differences, $\hat{\sigma}_{paired\ diff}^2$. The three variances, σ_w^2 , $\sigma_{period\ diff}^2$, and $\sigma_{paired\ diff}^2$, are functionally related by

$$\begin{aligned} \sigma_w^2 &= \text{Variance}(Y_{ijk}) \\ &= 2\sigma_{period\ diff}^2 = \frac{\sigma_{paired\ diff}^2}{2} \\ \sigma_{period\ diff}^2 &= \text{Variance}\left(\frac{Y_{i2k} - Y_{i1k}}{2}\right) \\ &= \frac{\sigma_w^2}{2} = \frac{\sigma_{paired\ diff}^2}{4} \\ \sigma_{paired\ diff}^2 &= \text{Variance}(Y_{i2k} - Y_{i1k}) \\ &= 2\sigma_w^2 = 4\sigma_{period\ diff}^2 \end{aligned}$$

Non-Inferiority Tests for the Difference Between Two Means in a 2x2 Cross-Over Design

Any of the three variances can be used in this procedure.

Computing the Power

The power is calculated as follows.

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

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 *Power* or *Sample Size*.

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

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 test.

- **Better (H1: Diff > -NIM)**

If higher means are Better the null hypothesis is $H_0: \text{Diff} \leq -\text{NIM}$, and the alternative hypothesis is $H_1: \text{Diff} > -\text{NIM}$.

- **Worse (H1: Diff < NIM)**

If higher means are Worse the null hypothesis is $H_0: \text{Diff} \geq \text{NIM}$, and the alternative hypothesis is $H_1: \text{Diff} < \text{NIM}$.

Non-Inferiority Tests for the Difference Between Two Means in a 2x2 Cross-Over Design

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 in fact the treatment mean is non-inferior.

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 different means when in fact the means are different.

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

N (Total Sample Size)

This option specifies one or more values of the sample size, the number of individuals in the study. This value must be an integer greater than one. Note that you may enter a list of values using the syntax *50,100,150,200,250* or *50 to 250 by 50*.

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.

D1 (Actual Difference to Detect)

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

For non-inferiority tests, this value is often set to zero.

When higher means are better, this value should be greater than -NIM.

When higher means are worse, this value should be less than NIM.

Effect Size – Standard Deviation

Standard Deviation Input Type

Choose what type of value to enter for the standard deviation.

This option does not change the test but is given for compatibility with various analysis result formats to which the user may have access. The choices all have functional relationships with the each other.

- **Sw (Sqrt[Within MSE from RM ANOVA])**

Enter the standard deviation as the square root of the within mean square error from a repeated measures ANOVA. This is the most common method since cross-over designs are usually analyzed using repeated measures ANOVA.

$$S_w = \sqrt{[\text{var}(Y_{ijk})]}$$

$$S_w = \sqrt{[\text{Within MSE}]}$$

$$S_w = S_{d\text{Period}} \times \sqrt{2}$$

$$S_w = S_{d\text{Paired}} / \sqrt{2}$$

- **SdPeriod (Std Dev of Period Differences)**

Enter the standard deviation of the period differences for each subject within each sequence.

$$S_{d\text{Period}} = \sqrt{[\text{var}((Y_{i2k} - Y_{i1k}) / 2)]}$$

$$S_{d\text{Period}} = S_w / \sqrt{2}$$

$$S_{d\text{Period}} = S_{d\text{Paired}} / 2$$

- **SdPaired (Std Dev of Paired Differences)**

Enter the standard deviation of the paired differences.

$$S_{d\text{Paired}} = \sqrt{[\text{var}(Y_{i2k} - Y_{i1k})]}$$

$$S_{d\text{Paired}} = S_w \times \sqrt{2}$$

$$S_{d\text{Paired}} = S_{d\text{Period}} \times 2$$

Sw

Enter the standard deviation as the square root of the within mean square error from a repeated measures ANOVA analysis of a prior cross-over design with $S_w > 0$.

$$S_w = \sqrt{[\text{var}(Y_{ijk})]} = \sqrt{[\text{Within MSE}]} = S_{d\text{Period}} \times \sqrt{2} = S_{d\text{Paired}} / \sqrt{2}$$

SdPeriod

Enter the standard deviation of the period differences for each subject within each sequence with $S_{d\text{Period}} > 0$.

$$S_{d\text{Period}} = \sqrt{[\text{var}((Y_{i2k} - Y_{i1k}) / 2)]} = S_w / \sqrt{2} = S_{d\text{Paired}} / 2$$

Sw

Enter the standard deviation of the paired differences with $S_{d\text{Paired}} > 0$.

$$S_{d\text{Paired}} = \sqrt{[\text{var}(Y_{i2k} - Y_{i1k})]} = S_w \times \sqrt{2} = S_{d\text{Period}} \times 2$$

Non-Inferiority Tests for the Difference Between Two Means in a 2x2 Cross-Over Design

Example 1 – Power Analysis

Suppose you want to consider the power of a balanced, cross-over design that will be analyzed using the t-test approach. You want to compute the power when the margin of equivalence is either 5 or 10 at several sample sizes between 5 and 50. The true difference between the means under H_0 is assumed to be 0. Similar experiments have had an S_w of 10. The significance level is 0.025.

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 in a 2x2 Cross-Over Design** procedure window by expanding **Means**, then **Cross-Over (2x2) Design**, then clicking on **Non-Inferiority**, and then clicking on **Non-Inferiority Tests for the Difference Between Two Means in a 2x2 Cross-Over Design**. 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 (H1: Diff > -NIM)
Alpha.....	0.025
N (Total Sample Size).....	5 10 15 20 30 40 50
NIM (Non-Inferiority Margin)	5 10
D1 (Actual Difference to Detect).....	0
Standard Deviation Input Type	Sw (Sqrt[Within MSE from RM ANOVA])
Sw	10

Annotated Output

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

Numeric Results and Plots

Numeric Results for 2x2 Cross-Over Design Non-Inferiority T-Test

Higher Means are Better

Hypotheses: $H_0: \text{Diff} \leq -\text{NIM}$ vs. $H_1: \text{Diff} > -\text{NIM}$

Power	N	Non-Inferiority Margin -NIM	Actual Difference D1	Sw	Alpha	Beta
0.08310	5	-5.000	0.000	10.000	0.025	0.91690
0.16563	10	-5.000	0.000	10.000	0.025	0.83437
0.24493	15	-5.000	0.000	10.000	0.025	0.75507
0.32175	20	-5.000	0.000	10.000	0.025	0.67825
0.46414	30	-5.000	0.000	10.000	0.025	0.53586
0.58682	40	-5.000	0.000	10.000	0.025	0.41318
0.68785	50	-5.000	0.000	10.000	0.025	0.31215
0.20131	5	-10.000	0.000	10.000	0.025	0.79869
0.50245	10	-10.000	0.000	10.000	0.025	0.49755
0.71650	15	-10.000	0.000	10.000	0.025	0.28350
0.84845	20	-10.000	0.000	10.000	0.025	0.15155
0.96222	30	-10.000	0.000	10.000	0.025	0.03778
0.99173	40	-10.000	0.000	10.000	0.025	0.00827
0.99835	50	-10.000	0.000	10.000	0.025	0.00165

Non-Inferiority Tests for the Difference Between Two Means in a 2x2 Cross-Over Design

References

- Chow, S.C. and Liu, J.P. 1999. Design and Analysis of Bioavailability and Bioequivalence Studies. Marcel Dekker. New York
- Chow, S.C., Shao, J., and 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.
- Senn, Stephen. 2002. Cross-over Trials in Clinical Research. Second Edition. John Wiley & Sons. New York.

Report Definitions

Power is the probability of rejecting H_0 when it is false. It should be close to one.

N is the total sample size drawn from all sequences. The sample is divided equally among sequences.

$-NIM$ 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.

Diff = treatment mean - reference mean.

$D1$ is the actual mean difference under the alternative hypothesis at which the power is computed.

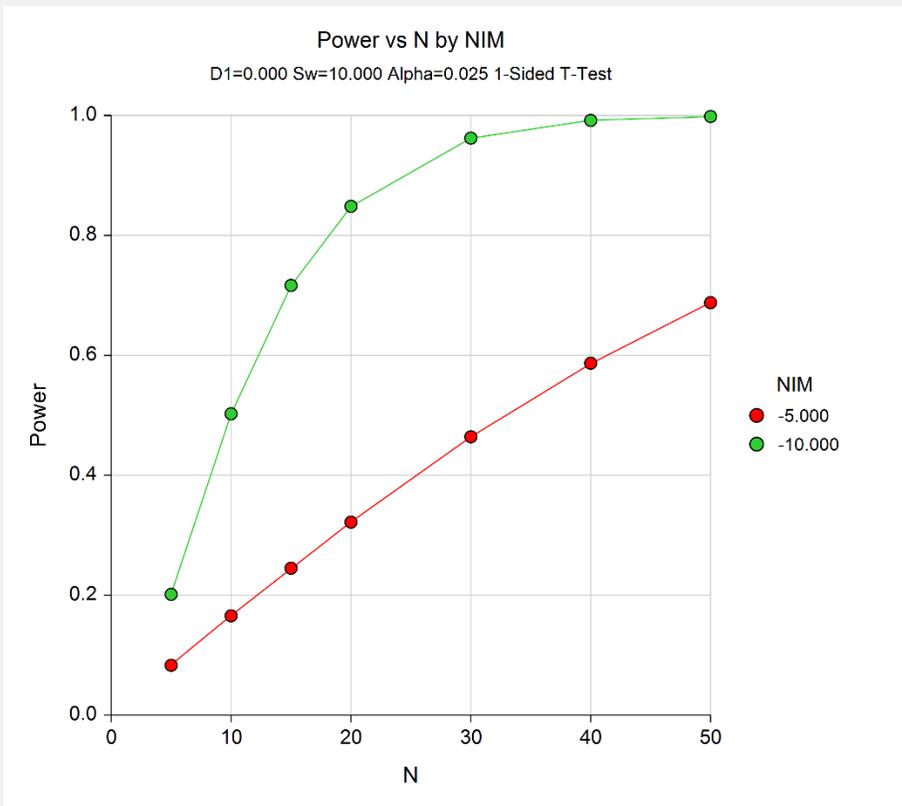
Sw is the square root of the within mean square error from a repeated measures ANOVA table.

Alpha is the probability of rejecting H_0 when it is true, which is the probability of a false positive.

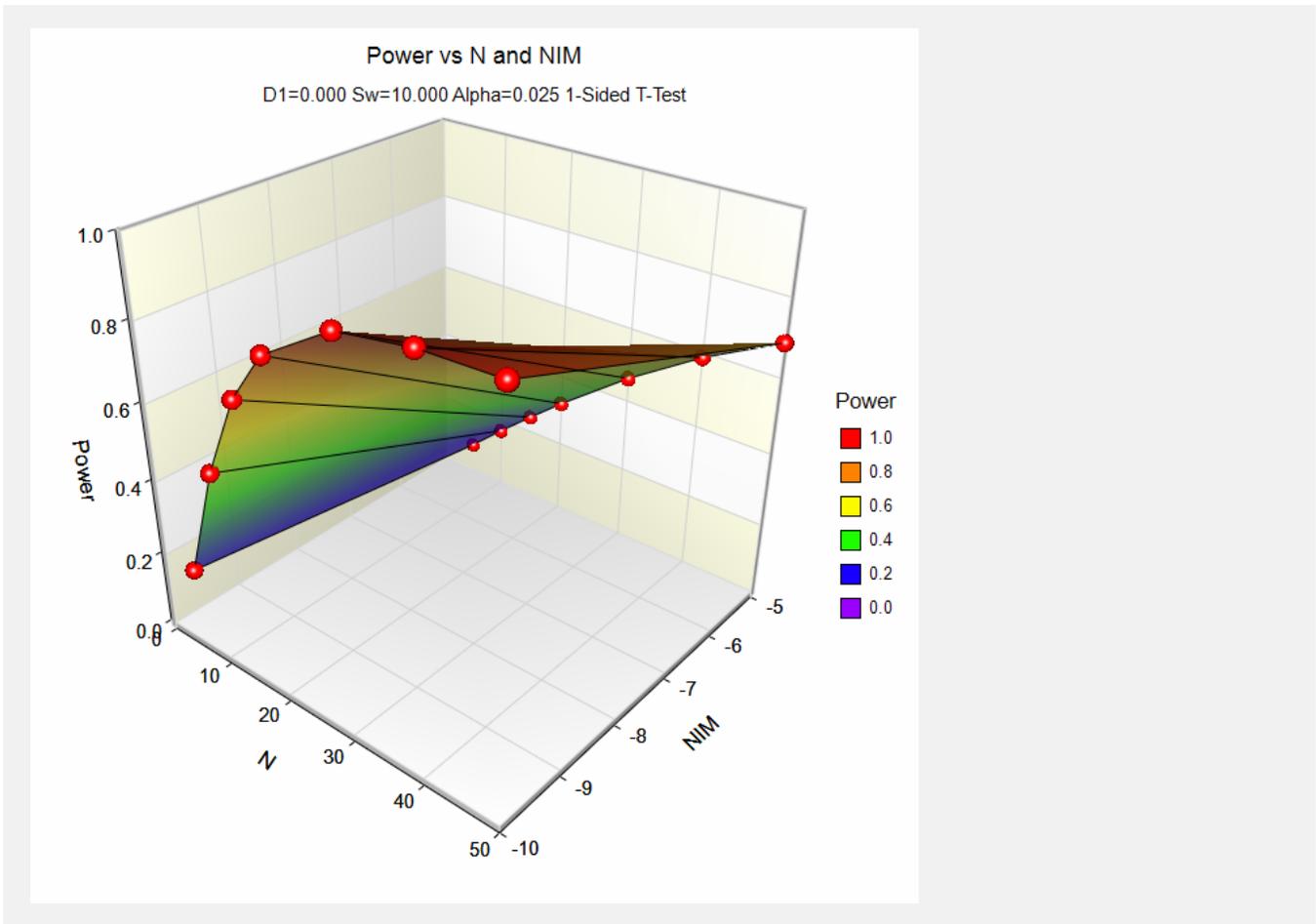
Beta is the probability of accepting H_0 when it is false, which is the probability of a false negative.

Summary Statements

A total sample size of 5 achieves 8% power to detect non-inferiority using a one-sided t-test when the margin of non-inferiority is -5.000, the true mean difference is 0.000, the significance level is 0.025, the square root of the within mean square error is 10.000, and a 2x2 cross-over design with an equal number in each sequence is used.



Non-Inferiority Tests for the Difference Between Two Means in a 2x2 Cross-Over Design



This report shows the values of each of the parameters, one scenario per row. The plots show the relationship between sample size and power. We see that a sample size of about 20 is needed to achieve 80% power when $NIM = -10$.

Example 2 – Finding the Sample Size

Continuing with Example 1, suppose the researchers want to find the exact sample size necessary to achieve 90% power for both values of D.

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 in a 2x2 Cross-Over Design** procedure window by expanding **Means**, then **Cross-Over (2x2) Design**, then clicking on **Non-Inferiority**, and then clicking on **Non-Inferiority Tests for the Difference Between Two Means in a 2x2 Cross-Over Design**. 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 (H1: Diff > -NIM)
Power.....	0.90
Alpha.....	0.025
NIM (Non-Inferiority Margin)	5 10
D1 (Actual Difference to Detect).....	0
Standard Deviation Input Type	Sw (Sqrt[Within MSE from RM ANOVA])
Sw	10

Output

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

Numeric Results

Numeric Results for 2x2 Cross-Over Design Non-Inferiority T-Test

Higher Means are Better

Hypotheses: H0: Diff \leq -NIM vs. H1: Diff > -NIM

	Non-Inferiority Margin		Actual Difference			
Power	N	-NIM	D1	Sw	Alpha	Beta
0.90648	88	-5.000	0.000	10.000	0.025	0.09352
0.91139	24	-10.000	0.000	10.000	0.025	0.08861

This report shows the exact sample size necessary for each scenario.

Note that the search for N is conducted across only even values of N since the design is assumed to be balanced.

Example 3 – Validation using Julious (2004)

Julious (2004) page 1953 presents an example in which $D = 0.0$, $NIM = 10$, $Sw = 20.00$, $\alpha = 0.025$, and $\beta = 0.10$. Julious obtains a sample size of 86.

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 in a 2x2 Cross-Over Design** procedure window by expanding **Means**, then **Cross-Over (2x2) Design**, then clicking on **Non-Inferiority**, and then clicking on **Non-Inferiority Tests for the Difference Between Two Means in a 2x2 Cross-Over Design**. 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 (H1: Diff > -NIM)
Power.....	0.90
Alpha.....	0.025
NIM (Non-Inferiority Margin)	10
D1 (Actual Difference to Detect).....	0
Standard Deviation Input Type	Sw (Sqrt[Within MSE from RM ANOVA])
Sw	20

Output

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

Numeric Results

Numeric Results for 2x2 Cross-Over Design Non-Inferiority T-Test

Higher Means are Better

Hypotheses: $H_0: \text{Diff} \leq -\text{NIM}$ vs. $H_1: \text{Diff} > -\text{NIM}$

	Non-Inferiority Margin -NIM	Actual Difference D1	Sw	Alpha	Beta
Power	N				
0.90648	88	-10.000	0.000	20.000	0.025
					0.09352

PASS obtained a sample size of 88, two higher than that obtained by Julious (2004). However, if you look at the power achieved by an N of 86, you will find that it is 0.899997—slightly less than the goal of 0.90.