Chapter 508

Superiority by a Margin Tests for the Difference of Two Means in a 2x2 Cross-Over Design

Introduction

This procedure computes power and sample size for non-zero null 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-zero null tests. Sample size formulas for non-zero null 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 no 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

$$\mathrm{H}_0: \mu_X \leq A \text{ versus } \mathrm{H}_1: \mu_X > A$$

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*.

$$\mathrm{H}_0: \mu_X \geq A \text{ versus } \mathrm{H}_1: \mu_X < A$$

Non-inferiority and *non-zero null* 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.

Parameter	PASS Input/Output	Interpretation
μ_T	Not used	*Treatment mean*. This is the treatment mean.
μ_R	Not used	*Reference mean*. This is the mean of a reference population.
M_s	SM	*Margin of superiority*. This is a tolerance value that defines the magnitude of difference that is required for practical importance. This may be thought of as the smallest difference from the reference that is considered to be different.
δ	D	*True difference*. 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-Zero Null Tests

A *non-zero null test* tests that the treatment mean is better than the reference mean by more than the superiority margin. The actual direction of the hypothesis depends on the response variable being studied.

Case 1: High Values Good

In this case, higher values are better. The hypotheses are arranged so that rejecting the null hypothesis implies that the treatment mean is greater than the reference mean by at least the margin of superiority. The value of δ must be greater than $|M_S|$. The following are equivalent sets of hypotheses.

$$\mathrm{H}_0: \mu_1 \leq \mu_2 + |M_S| \quad \text{versus} \quad \mathrm{H}_1: \mu_1 > \mu_2 + |M_S|$$

$$\mathrm{H}_0: \mu_1 - \mu_2 \leq |M_S| \quad \text{versus} \quad \mathrm{H}_1: \mu_1 - \mu_2 > |M_S|$$

$$\mathrm{H}_0: \delta \leq |M_S| \quad \text{versus} \quad \mathrm{H}_1: \delta > |M_S|$$

Case 2: High Values Bad

In this case, lower values are better. The hypotheses are arranged so that rejecting the null hypothesis implies that the treatment mean is less than the reference mean by at least the margin of superiority. The value of δ must be less than $-|M_S|$. The following are equivalent sets of hypotheses.

$$\mathrm{H}_0: \mu_1 \geq \mu_2 - |M_S| \quad \text{versus} \quad \mathrm{H}_1: \mu_1 < \mu_2 - |M_S|$$

$$\mathrm{H}_0: \mu_1 - \mu_2 \geq -|M_S| \quad \text{versus} \quad \mathrm{H}_1: \mu_1 - \mu_2 < -|M_S|$$

$$\mathrm{H}_0: \delta \geq -|M_S| \quad \text{versus} \quad \mathrm{H}_1: \delta < -|M_S|$$

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.

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 will have an estimate of the variance of the period differences from the t-test, $\hat{\sigma}_d^2$. These variances are functionally related by $\sigma_w^2 = 2\sigma_d^2$. Either variance can be entered.

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.

If Higher Means Are Better the null hypothesis is Diff ≤ SM and the alternative hypothesis is Diff > SM. If Higher Means Are Worse the null hypothesis is Diff ≥ -SM and the alternative hypothesis is Diff < -SM.

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 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 when in fact the treatment is not superior.

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

SM (Superiority Margin)

This is the magnitude of the margin of superiority. It must be entered as a positive number.

When higher means are better, this value is the distance above the reference mean that is required to be considered superior. When higher means are worse, this value is the distance below the reference mean that is required to be considered superior.

D (True Difference)

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

When higher means are better, this value should be greater than SM. When higher means are worse, this value should be negative and greater in magnitude than SM.

Effect Size – Standard Deviation

Specify S as Sw, SdPeriod, or SdPaired

Specify the form of the standard deviation that is entered in the box below.

- **Sw**

  Specify S 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 ANOVA.

- **SdPeriod**

  Specify the standard deviation S as the the standard deviation of the period differences for each subject within each sequence. Note SdPeriod^2 = var((Yi2k - Yi1k)/2) = Sw^2 / 2.

- **SdPaired**

  Specify the standard deviation S as the the standard deviation of the paired differences. Note SdPaired^2 = var(Yi2k - Yi1k) = 2 * Sw^2.

S (Value of Sw, SdPeriod, or SdPaired)

Specify the value(s) of the standard deviation S. The interpretation of this value depends on the entry in "Specify S as Sw, SdPeriod or SdPaired" above.

If S=Sw is selected, this is the value of Sw which is sqrt(WMSE), where WMSE is the within mean square error from the ANOVA table used to analyze the Cross-Over design. Note Sw^2 = var(Yijk).

IF S=SdPeriod is selected, this is the value of SdPeriod, which is the standard deviation of the period differences for each subject within each sequence. Note SdPeriod^2 = var((Yi2k - Yi1k)/2) = Sw^2 / 2.

IF S=SdPaired is selected, this is the value of Sd which is the standard deviation of the paired differences. Note SdPaired^2 = var(Yi2k - Yi1k) = 2 * Sw^2.

These values must be positive. A list of values may be entered.

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 superiority is either 5 or 10 at several sample sizes between 5 and 50. The true difference between the means under H0 is assumed to be 15. Similar experiments have had a value for *Sw* 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 **Superiority by a Margin Tests for the Difference of Two Means in a 2x2 Cross-Over Design** procedure window by expanding **Means**, then **Cross-Over (2x2) Design**, then clicking on **Superiority by a Margin**, and then clicking on **Superiority by a Margin Tests for the Difference of 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**.

Option	Value
Design Tab	
Solve For	**Power**
Higher Means Are	**Better**
Alpha	**0.025**
N (Total Sample Size)	**5 10 15 20 30 40 50**
SM (Superiority Margin)	**5 10**
D (True Difference)	**15**
Specify S as Sw or Sd	**Sw**
S (Value of Sw or Sd)	**10**

Annotated Output

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

Numeric Results and Plots

Numeric Results for Superiority T-Test (H0: Diff ≤ SM; H1: Diff > SM)
Higher Means are Better

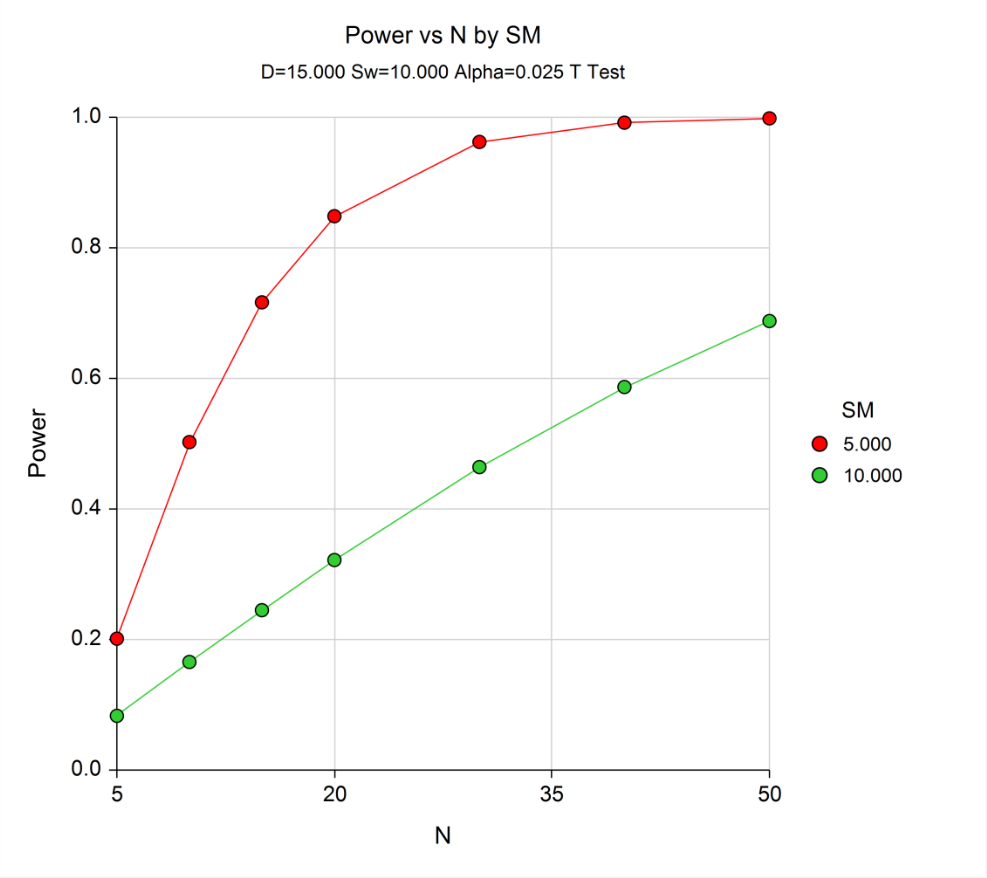
Power	N	Superiority Margin (SM)	Actual Difference (D)	Significance Level (Alpha)	Beta	Standard Deviation (Sw)
0.20131	5	5.000	15.000	0.02500	0.79869	10.000
0.50245	10	5.000	15.000	0.02500	0.49755	10.000
0.71650	15	5.000	15.000	0.02500	0.28350	10.000
0.84845	20	5.000	15.000	0.02500	0.15155	10.000
0.96222	30	5.000	15.000	0.02500	0.03778	10.000
0.99173	40	5.000	15.000	0.02500	0.00827	10.000
0.99835	50	5.000	15.000	0.02500	0.00165	10.000
0.08310	5	10.000	15.000	0.02500	0.91690	10.000
0.16563	10	10.000	15.000	0.02500	0.83437	10.000
0.24493	15	10.000	15.000	0.02500	0.75507	10.000
0.32175	20	10.000	15.000	0.02500	0.67825	10.000
0.46414	30	10.000	15.000	0.02500	0.53586	10.000
0.58682	40	10.000	15.000	0.02500	0.41318	10.000
0.68785	50	10.000	15.000	0.02500	0.31215	10.000

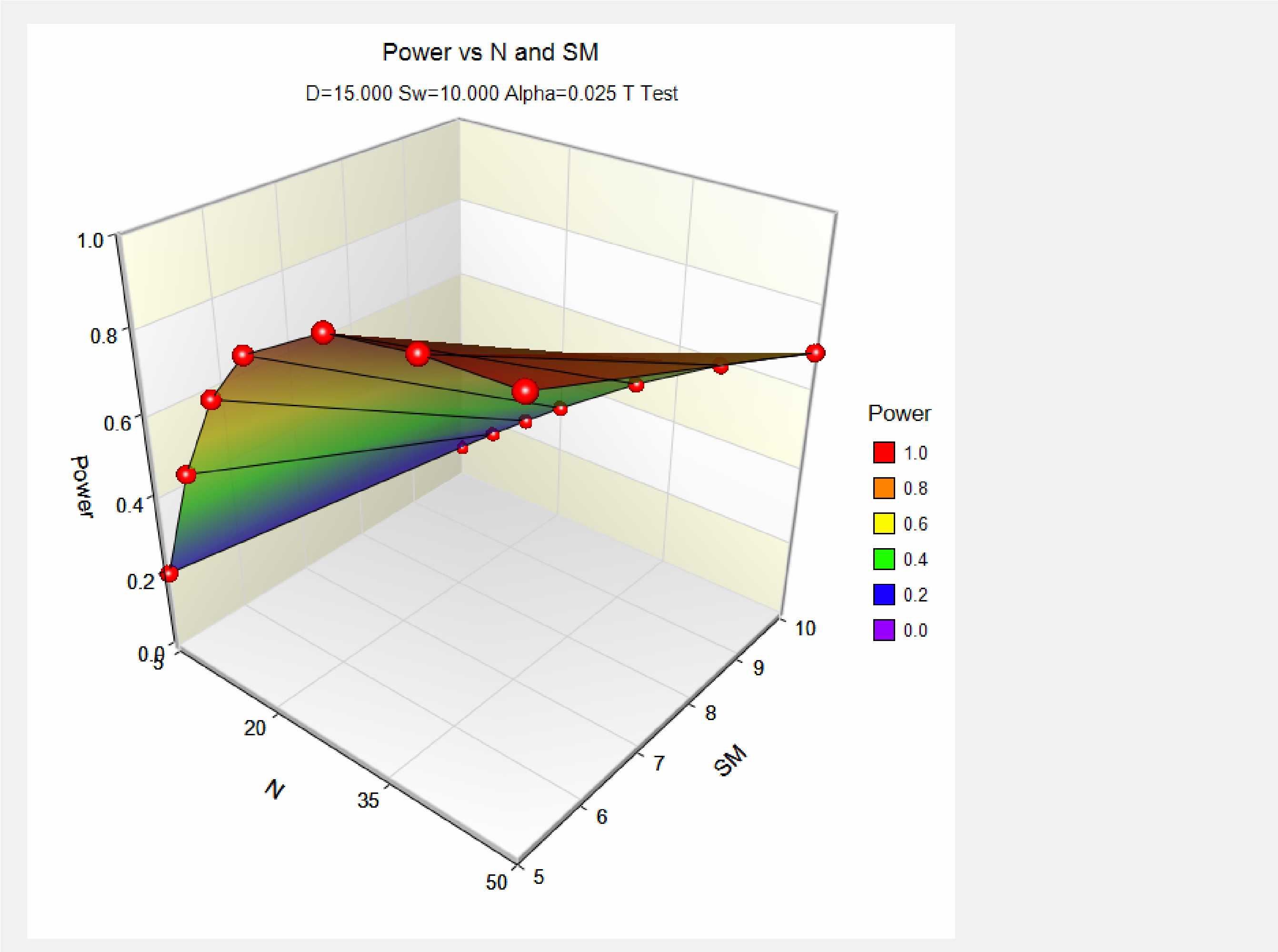
Report Definitions

H0 (null hypothesis) is Diff <= SM, where D = Treatment Mean - Reference Mean.
H1 (alternative hypothesis) is Diff > SM.
Power is the probability of rejecting H0 when it is false.
N is the total sample size drawn from all sequences. The sample is divided equally among sequences.
SM is the magnitude of the margin of superiority. Since higher means are better, this value is positive and is the distance above the reference mean that is required to be considered superior.
D is the mean difference at which the power is computed. D = Mean1 - Mean2 = treatment mean - reference mean.
Alpha is the probability of a false positive.
Beta is the probability of a false negative.
Sw is the square root of the within mean square error from the ANOVA table.

Summary Statements

A total sample size of 5 achieves 20% power to detect superiority using a one-sided t-test when the margin of superiority is 5.000, the true mean difference is 15.000, the significance level is 0.02500, 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.





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 SM = 5.

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 **Superiority by a Margin Tests for the Difference of Two Means in a 2x2 Cross-Over Design** procedure window by expanding **Means**, then **Cross-Over (2x2) Design**, then clicking on **Superiority by a Margin**, and then clicking on **Superiority by a Margin Tests for the Difference of 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**.

Option	Value
Design Tab	
Solve For	**Sample Size**
Power	**0.90**
Higher Means Are	**Better**
Alpha	**0.025**
SM (Superiority Margin)	**5 10**
D (True Difference)	**15**
Specify S as Sw or Sd	**Sw**
S (Value of Sw or Sd)	**10**

Output

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

Numeric Results

Numeric Results for Superiority T-Test (H0: Diff ≤ SM; H1: Diff > SM)
Higher Means are Better

Power	N	Superiority Margin (SM)	Actual Difference (D)	Significance Level (Alpha)	Beta	Standard Deviation (Sw)
0.91139	24	5.000	15.000	0.02500	0.08861	10.000
0.90648	88	10.000	15.000	0.02500	0.09352	10.000

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

This procedure uses the same mechanics as the Non-Inferiority Tests for the Difference Between Two Means in a 2x2 Cross-Over Design procedure. We refer the user to Example 3 of Chapter 510 for the validation.