

Chapter 508

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

Introduction

This procedure computes power and sample size for superiority by a margin 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 superiority by a margin tests. Sample size formulas for superiority by a margin 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.

Cross-Over Analysis

The following discussion summarizes the presentation of Chow and Liu (1999). The general linear model for the standard 2x2 cross-over design is

$$Y_{ijk} = \mu + S_{ik} + P_j + \mu_{(j,k)} + C_{(j-1,k)} + e_{ijk}$$

where i represents a subject (1 to N_k), j represents the period (1 or 2), and k represents the sequence (1 or 2). The S_{ik} represent the random effects of the subjects. The P_j represent the effects of the two periods. The $\mu_{(j,k)}$ represent the means of the two treatments. In the case of the 2x2 cross-over design

$$\mu_{(j,k)} = \begin{cases} \mu_1 & \text{if } k = j \\ \mu_2 & \text{if } k \neq j \end{cases}$$

where the subscripts 1 and 2 represent treatments A and B, respectively.

The $C_{(j-1,k)}$ represent the carry-over effects. In the case of the 2x2 cross-over design

$$C_{(j-1,k)} = \begin{cases} C_1 & \text{if } j = 2, k = 1 \\ C_2 & \text{if } j = 2, k = 2 \\ 0 & \text{otherwise} \end{cases}$$

where the subscripts 1 and 2 represent treatments A and B, respectively.

Assuming that the average effect of the subjects is zero, the four means from the 2x2 cross-over design can be summarized using the following table.

Sequence	Period 1	Period 2
1 (AB)	$\mu_{11} = \mu + P_1 + \mu_1$	$\mu_{21} = \mu + P_2 + \mu_2 + C_1$
2 (BA)	$\mu_{12} = \mu + P_1 + \mu_2$	$\mu_{22} = \mu + P_2 + \mu_1 + C_2$

where $P_1 + P_2 = 0$ and $C_1 + C_2 = 0$.

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 (H_0) and alternative (H_1) hypotheses for one-sided tests are defined as

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

Rejecting H_0 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 .

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

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

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

Non-inferiority and *superiority by a margin* 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	<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_S	SM	<i>Margin of superiority.</i> 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	$\delta 1$	<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.

Superiority by a Margin Tests

A *superiority by a margin 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.

$$H_0: \mu_1 \leq \mu_2 + |M_S| \quad \text{versus} \quad H_1: \mu_1 > \mu_2 + |M_S|$$

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

$$H_0: \delta \leq |M_S| \quad \text{versus} \quad H_1: \delta > |M_S|$$

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.

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

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

$$H_0: \delta \geq -|M_S| \quad \text{versus} \quad H_1: \delta < -|M_S|$$

Test Statistic

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.

Computing the Within-Subject Variance (σ_w^2)

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 , where

$$\sigma_w^2 = \text{Variance}(e_{ijk})$$

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 ($\hat{\sigma}_p^2$), an estimate of the variance of the paired differences ($\hat{\sigma}_D^2$), or an estimate of the variances of the paired variables ($\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$) and the correlation between the paired variables ($\hat{\rho}$). The within-subject variance, σ_w^2 , is functionally related to these other variances as described below. Any of these different variances may be entered directly into this procedure.

Using the Variance of the Period Differences (σ_p^2)

The variance of the period differences for each subject within each sequence (σ_p^2) is defined as

$$\sigma_p^2 = \text{Variance}\left(\frac{Y_{i2k} - Y_{i1k}}{2}\right).$$

σ_p^2 has a functional relationship with the within-subject population variance (σ_w^2), namely

$$\sigma_p^2 = \frac{\sigma_w^2}{2},$$

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

such that

$$\sigma_w^2 = 2\sigma_p^2 .$$

The within-subject standard deviation (σ_w) is then

$$\sigma_w = \sqrt{2\sigma_p^2} .$$

Using the Variance of the Paired Differences (σ_D^2)

The variance of the paired differences (σ_D^2) is defined as

$$\sigma_D^2 = \text{Variance}(Y_{i2k} - Y_{i1k}) .$$

σ_D^2 has a functional relationship with the within-subject population variance (σ_w^2), namely

$$\sigma_D^2 = 2\sigma_w^2 ,$$

such that

$$\sigma_w^2 = \frac{\sigma_D^2}{2} .$$

The within-subject standard deviation (σ_w) is then

$$\sigma_w = \sqrt{\frac{\sigma_D^2}{2}} .$$

Using the Variances of the Paired Variables (σ_1^2 and σ_2^2) and the Correlation Between the Paired Variables (ρ)

The variances of the paired variables (σ_1^2 and σ_2^2) and the correlation between the paired variables (ρ) are defined as

$$\sigma_1^2 = \text{Variance}(Y_{i1k})$$

$$\sigma_2^2 = \text{Variance}(Y_{i2k})$$

$$\rho = \text{Correlation}(Y_{i1k}, Y_{i2k})$$

The variance of paired differences (σ_D^2) can be computed from σ_1^2 , σ_2^2 and ρ as

$$\sigma_D^2 = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2 ,$$

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

such that the within-subject population variance (σ_w^2) can be computed as

$$\sigma_w^2 = \frac{\sigma_D^2}{2} = \frac{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}{2}.$$

The within-subject standard deviation (σ_w) is then

$$\sigma_w = \sqrt{\frac{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}{2}}.$$

If $\sigma_1^2 = \sigma_2^2 = \sigma_x^2$, then with

$$\sigma_x^2 = \text{Variance}(Y_{ijk}),$$

the formula for σ_w^2 reduces to

$$\sigma_w^2 = \sigma_x^2(1 - \rho).$$

The within-subject standard deviation (σ_w) is then

$$\sigma_w = \sqrt{\sigma_x^2(1 - \rho)}.$$

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 .

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 H_0 is assumed to be 15. Similar experiments have had a value for σ_w of 10. The significance level is 0.025.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 1** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Solve For	Power
Higher Means Are	Better ($H_1: \delta > SM$)
Alpha.....	0.025
N (Total Sample Size).....	5 10 15 20 30 40 50
SM (Superiority Margin)	5 10
δ_1 (Actual Difference to Detect)	15
Standard Deviation Input Type	Enter the Within-Subject Population SD
σ_w (Within-Subject Population SD).....	10

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

Output

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

Numeric Results

Solve For: **Power**
 Higher Means Are: **Better**
 Hypotheses: **H0: $\delta \leq SM$ vs. H1: $\delta > SM$**

Power	Total Sample Size N	Superiority Margin SM	Actual Difference δ_1	Standard Deviation σ_w	Alpha	Beta
0.20131	5	5	15	10	0.025	0.79869
0.50245	10	5	15	10	0.025	0.49755
0.71650	15	5	15	10	0.025	0.28350
0.84845	20	5	15	10	0.025	0.15155
0.96222	30	5	15	10	0.025	0.03778
0.99173	40	5	15	10	0.025	0.00827
0.99835	50	5	15	10	0.025	0.00165
0.08310	5	10	15	10	0.025	0.91690
0.16563	10	10	15	10	0.025	0.83437
0.24493	15	10	15	10	0.025	0.75507
0.32175	20	10	15	10	0.025	0.67825
0.46414	30	10	15	10	0.025	0.53586
0.58682	40	10	15	10	0.025	0.41318
0.68785	50	10	15	10	0.025	0.31215

Power The probability of rejecting a false null hypothesis when the alternative hypothesis is true.
 N The total sample size drawn from all sequences. The sample is divided equally among sequences.
 SM 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.
 δ The difference in means. Difference (δ) = Treatment Mean (μ_T) - Reference Mean (μ_R).
 δ_1 The actual mean difference under the alternative hypothesis at which the power is computed.
 σ_w The within-subject population standard deviation. $\sigma_w = \sqrt{[var(e_{ijk})]}$. σ_w is estimated as the square root of the within mean square error (WMSE) (i.e., $\sigma_w = \sqrt{[WMSE]}$) from a repeated measures ANOVA analysis of a prior cross-over design.
 Alpha The probability of rejecting a true null hypothesis.
 Beta The probability of failing to reject the null hypothesis when the alternative hypothesis is true.

Summary Statements

A 2x2 cross-over design (where higher means are considered to be better) will be used to test whether the treatment mean (μ_T) is superior to the reference mean (μ_R) by a margin, with a superiority margin of 5 (H0: $\mu_T - \mu_R \leq 5$ versus H1: $\mu_T - \mu_R > 5$). The comparison will be made using a one-sided t-test, with a Type I error rate (α) of 0.025. The within-subject population standard deviation is assumed to be 10. To detect a difference in means ($\mu_T - \mu_R$) of 15, with a total sample size of 5 (allocated equally to the two sequences), the power is 0.20131.

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

Dropout-Inflated Sample Size

Dropout Rate	Sample Size N	Dropout- Inflated Enrollment Sample Size N'	Expected Number of Dropouts D
20%	5	7	2
20%	10	13	3
20%	15	19	4
20%	20	25	5
20%	30	38	8
20%	40	50	10
20%	50	63	13

Dropout Rate	The percentage of subjects (or items) that are expected to be lost at random during the course of the study and for whom no response data will be collected (i.e., will be treated as "missing"). Abbreviated as DR.
N	The evaluable sample size at which power is computed (as entered by the user). If N subjects are evaluated out of the N' subjects that are enrolled in the study, the design will achieve the stated power.
N'	The total number of subjects that should be enrolled in the study in order to obtain N evaluable subjects, based on the assumed dropout rate. N' is calculated by inflating N using the formula $N' = N / (1 - DR)$, with N' always rounded up. (See Julious, S.A. (2010) pages 52-53, or Chow, S.C., Shao, J., Wang, H., and Lokenygina, Y. (2018) pages 32-33.)
D	The expected number of dropouts. $D = N' - N$.

Dropout Summary Statements

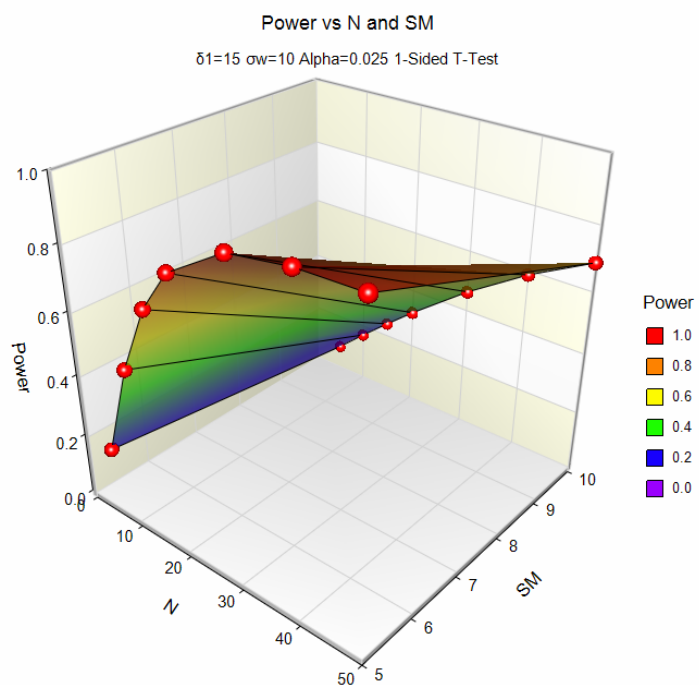
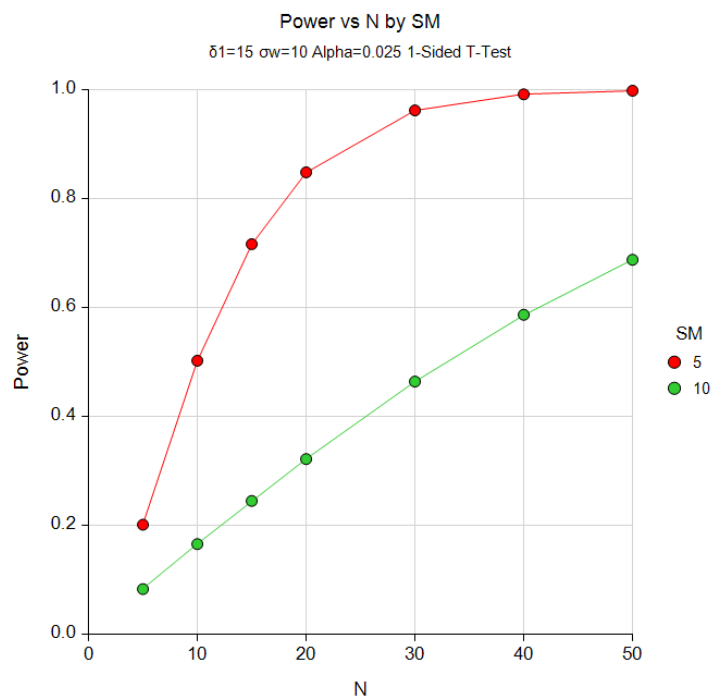
Anticipating a 20% dropout rate, 7 subjects should be enrolled to obtain a final sample size of 5 subjects.

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.

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

Plots



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 δ .

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 2** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Solve For **Sample Size**
 Higher Means Are **Better (H1: $\delta > SM$)**
 Power..... **0.90**
 Alpha..... **0.025**
 SM (Superiority Margin) **5 10**
 δ 1 (Actual Difference to Detect) **15**
 Standard Deviation Input Type **Enter the Within-Subject Population SD**
 σ_w (Within-Subject Population SD)..... **10**

Output

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

Numeric Results

Solve For: [Sample Size](#)
 Higher Means Are: Better
 Hypotheses: $H_0: \delta \leq SM$ vs. $H_1: \delta > SM$

Power	Total Sample Size N	Superiority Margin SM	Actual Difference δ 1	Standard Deviation σ_w	Alpha	Beta
0.91139	24	5	15	10	0.025	0.08861
0.90648	88	10	15	10	0.025	0.09352

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 a non-inferiority example in which $\delta_1 = 0.0$, $NIM = 10$, $\sigma_w = 20.00$, $\alpha = 0.025$, and $\beta = 0.10$. Julious obtains a sample size of 86. We can use this example to validate this procedure with $SM = 10$ and $\delta_1 = 20$ since $\delta_1 - SM$ has the same magnitude of difference as $\delta_1 - NIM$.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 3** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Solve For **Sample Size**
 Higher Means Are **Better (H1: $\delta > SM$)**
 Power **0.90**
 Alpha **0.025**
 SM (Superiority Margin) **10**
 δ_1 (Actual Difference to Detect) **20**
 Standard Deviation Input Type **Enter the Within-Subject Population SD**
 σ_w (Within-Subject Population SD) **20**

Output

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

Numeric Results

Solve For: [Sample Size](#)
 Higher Means Are: Better
 Hypotheses: $H_0: \delta \leq SM$ vs. $H_1: \delta > SM$

Power	Total Sample Size N	Superiority Margin SM	Actual Difference δ_1	Standard Deviation σ_w	Alpha	Beta
0.90648	88	10	20	20	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.