

## Chapter 500

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

---

### Introduction

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 contains two *sequences* (treatment orderings) and two time periods (occasions). 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 drugs are administered, not by the treatments they receive. Indeed, higher-order cross-over designs have been used in which the same treatment is used at both occasions.

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.

---

### Advantages of Cross-Over Designs

A comparison of treatments on the same subject is expected to be more precise. The increased precision often translates into a smaller sample size. Also, patient enrollment into the study may be easier because each patient will receive both treatments. Finally, it is often more difficult to obtain a subject than to obtain a measurement.

## Disadvantages of Cross-Over Designs

The statistical analysis of a cross-over experiment is more complex than a parallel-group experiment and requires additional assumptions. It may be difficult to separate the treatment effect from the period effect, the carry-over effect of the previous treatment, and the interaction between period and treatment.

The design cannot be used when the treatment (or the measurement of the response) alters the subject permanently. Hence, it should not be used to compare treatments that are intended to provide a cure.

Because subjects must be measured at least twice, it is often more difficult to keep patients enrolled in the study. It is arguably simpler to measure a subject once than to obtain their measurement twice. This is particularly true when the measurement process is painful, uncomfortable, embarrassing, or time consuming.

## Technical Details

The 2x2 crossover design may be described as follows. Randomly assign the subjects to one of two sequence groups so that there are  $N_1$  subjects in sequence one and  $N_2$  subjects in sequence two. In order to achieve design balance, the sample sizes  $N_1$  and  $N_2$  are assumed to be equal so that  $N_1 = N_2 = N / 2$ .

Sequence one is given treatment A followed by treatment B. Sequence two is given treatment B followed by treatment A. The sequence is replicated  $m$  times. So, if  $m = 3$ , the sequences are ABABAB and BABABA.

The usual method of analysis is the analysis of variance. However, the power and sample size formulas that follow are based on the t-test, not the F-test. This is done because, in the balanced case, the t-test and the analysis of variance F-test are equivalent. Also, the F-test is limited to a two-sided hypothesis, while the t-test allows both one-sided and two-sided hypotheses. This is important because one-sided hypotheses are used for non-inferiority and equivalence testing.

## 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

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

where  $i$  represents a subject (1 to  $N_k$ ),  $j$  represents the period (1 or 2),  $k$  represents the sequence (1 or 2), and  $l$  represents the replicate. 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.

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

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.

<i>Sequence</i>	<i>Period 1</i>	<i>Period 2</i>
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  $C_1 + C_2 = 0$ .

### Test Statistic

The presence of a treatment effect can be studied by testing whether  $\mu_1 - \mu_2 = \delta$  using a  $t$ -test or an F-test. If the F-test is used, only a two-sided test is possible. The  $t$  statistic is calculated as follows

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

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

The two-sided null hypothesis is rejected at the  $\alpha$  significance level if  $|t_d| > t_{\alpha/2, N-2}$ . Similar results are available for a one-sided hypothesis test.

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,  $\sigma_w^2$ . If prior studies used a  $t$ -test rather than an ANOVA to analyze the data, you may not have a direct estimate of  $\sigma_w^2$ . 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,  $\sigma_w^2$ ,  $\sigma_{period\ diff}^2$ , and  $\sigma_{paired\ diff}^2$ , are functionally related by

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

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

## Computing the Power

The power is calculated as follows for a directional alternative (one-sided test).

1. Find  $t_\alpha$  such that  $1 - T_{df}(t_\alpha) = \alpha$ , where  $T_{df}(x)$  is the area left of  $x$  under a central- $t$  curve and  $df = N - 2$ .
2. Calculate the noncentrality parameter:  $\lambda = \frac{\delta\sqrt{N}}{\sigma_w\sqrt{2}}$ .
3. 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  $\lambda$ .

## 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

##### Alternative Hypothesis

This option specifies the alternative hypothesis. This implicitly specifies the direction of the hypothesis test. Note that the alternative hypothesis enters into power calculations by specifying the rejection region of the hypothesis test. Its accuracy is critical.

Possible selections are:

- **H1: Diff  $\neq$  D0 (Two-Sided)**

This is the most common selection. It yields the *two-sided* t-test. Use this option when you are testing whether the means are different but you do not want to specify beforehand which mean is larger. Many scientific journals require two-sided tests.

- **H1: Diff  $<$  D0 (One-Sided)**

This option yields a *one-sided* t-test. Use it when you are only interested in the case in which the actual difference is less than D0.

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

- **H1: Diff > D0 (One-Sided)**

This option yields a *one-sided* t-test. Use it when you are only interested in the case in which actual difference is greater than D0.

---

## 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 equal means when in fact the means are different.

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 equal means when in fact the means are equal.

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 (total subjects in both sequences). 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 Differences

### D0 (Mean Difference|H0)

Enter the difference between the treatment means under the null (H0) hypothesis. This is the value that is to be rejected when the t-test is significant. This value is commonly set to zero.

You may enter a range of values such as *10 20 30* or *0 to 100 by 25*.

### D1 (Mean Difference|H1)

Enter the difference between the population means under the alternative (H1) hypothesis. This is the value of the difference at which the power is calculated.

You may enter a range of values such as *10 20 30* or *0 to 100 by 25*.

## 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.

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

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

$$Sw = SdPeriod \times \sqrt{2}$$

$$Sw = SdPaired / \sqrt{2}$$

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

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

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

$$SdPeriod = Sw / \sqrt{2}$$

$$SdPeriod = SdPaired / 2$$

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

Enter the standard deviation of the paired differences.

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

$$SdPaired = Sw \times \sqrt{2}$$

$$SdPaired = SdPeriod \times 2$$

### 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  $Sw > 0$ .

$$Sw = \sqrt{[\text{var}(Y_{ijk})]} = \sqrt{[\text{Within MSE}]} = SdPeriod \times \sqrt{2} = SdPaired / \sqrt{2}$$

### SdPeriod

Enter the standard deviation of the period differences for each subject within each sequence with  $SdPeriod > 0$ .

$$SdPeriod = \sqrt{[\text{var}((Y_{i2k} - Y_{i1k}) / 2)]} = Sw / \sqrt{2} = SdPaired / 2$$

### Sw

Enter the standard deviation of the paired differences with  $SdPaired > 0$ .

$$SdPaired = \sqrt{[\text{var}(Y_{i2k} - Y_{i1k})]} = Sw \times \sqrt{2} = SdPeriod \times 2$$

## 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 two-sided t-test approach. The difference between the treatment means under  $H_0$  is 0. Similar experiments have had a standard deviation of the period differences (SdPeriod) of 10. Compute the power when the true differences are 5 and 10 at sample sizes between 5 and 50. The significance level is 0.05.

### Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **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 **Test (Inequality)**, and then clicking on **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>
<b>Design Tab</b>	
Solve For .....	<b>Power</b>
Alternative Hypothesis .....	<b>H1: Diff <math>\neq</math> D0 (Two-Sided)</b>
Alpha .....	<b>0.05</b>
N (Total Sample Size).....	<b>5 10 15 20 30 40 50</b>
D0 (Mean Difference H0).....	<b>0</b>
D1 (Mean Difference H1).....	<b>5 10</b>
Standard Deviation Input Type .....	<b>SdPeriod (Std Dev of Period Differences)</b>
SdPeriod .....	<b>10</b>

### Annotated Output

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

### Numeric Results

#### Numeric Results for 2x2 Cross-Over Design T-Test

Hypotheses:  $H_0$ : Diff = D0 vs.  $H_1$ : Diff  $\neq$  D0

Power	N	Null Difference D0	Actual Difference D1	SdPeriod	Effect Size	Alpha	Beta
0.06912	5	0.000	5.000	10.000	0.354	0.050	0.93088
0.10769	10	0.000	5.000	10.000	0.354	0.050	0.89231
0.14630	15	0.000	5.000	10.000	0.354	0.050	0.85370
0.18510	20	0.000	5.000	10.000	0.354	0.050	0.81490
0.26244	30	0.000	5.000	10.000	0.354	0.050	0.73756
0.33794	40	0.000	5.000	10.000	0.354	0.050	0.66206
0.41010	50	0.000	5.000	10.000	0.354	0.050	0.58990
0.12657	5	0.000	10.000	10.000	0.707	0.050	0.87343
0.28630	10	0.000	10.000	10.000	0.707	0.050	0.71370
0.43392	15	0.000	10.000	10.000	0.707	0.050	0.56608
0.56201	20	0.000	10.000	10.000	0.707	0.050	0.43799
0.75292	30	0.000	10.000	10.000	0.707	0.050	0.24708
0.86895	40	0.000	10.000	10.000	0.707	0.050	0.13105
0.93371	50	0.000	10.000	10.000	0.707	0.050	0.06629

## 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.

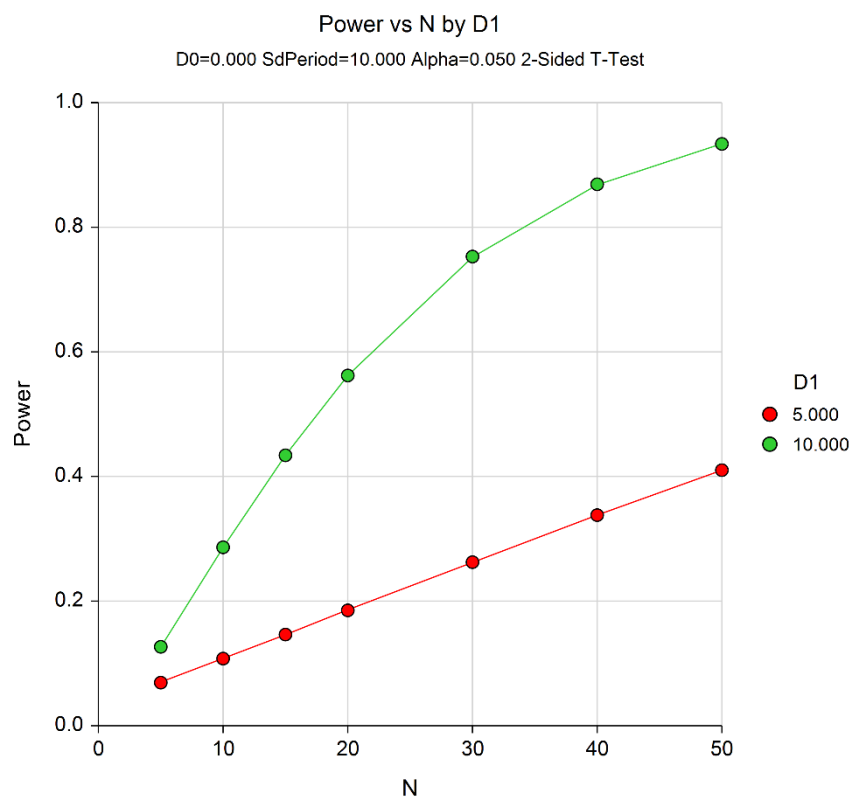
### Report Definitions

- 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.
- Diff = treatment mean - reference mean.
- $D_0$  is the mean difference under the null hypothesis,  $H_0$ .
- $D_1$  is the actual mean difference under the alternative hypothesis at which the power is computed.
- SdPeriod is the standard deviation of the period differences for each subject within each sequence.
- Effect Size =  $|D_1 - D_0| / (SdPeriod \times \sqrt{2})$  is the relative magnitude of the effect under the alternative hypothesis.
- 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 failing to reject  $H_0$  when it is false, which is the probability of a false negative.

### Summary Statements

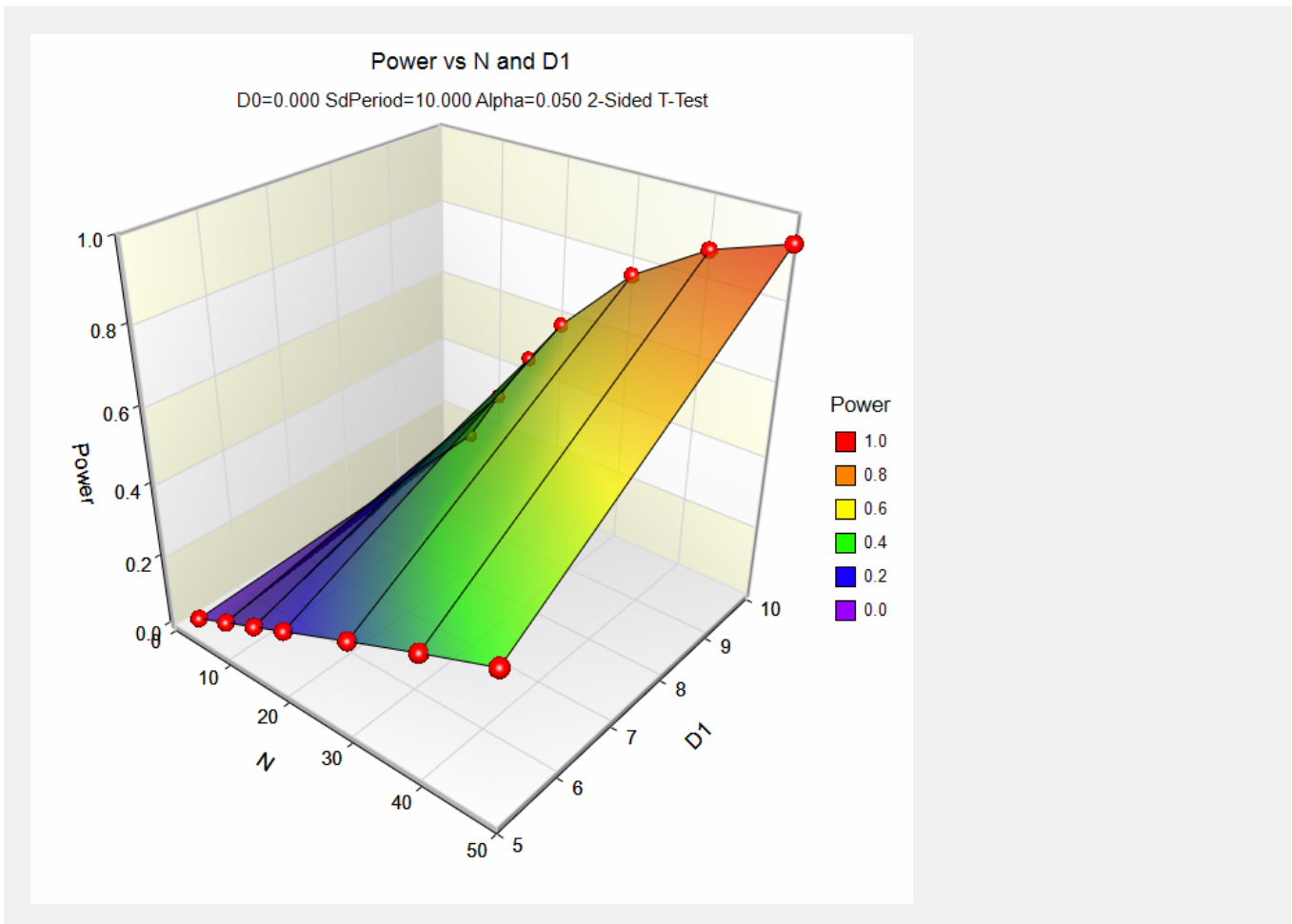
A two-sided t-test achieves 7% power to infer that the mean difference is not 0.000 when the total sample size of a 2x2 cross-over design is 5 with an equal number in each sequence, the actual mean difference is 5.000, the standard deviation of the period differences for each subject within each sequence is 10.000, and the significance level is 0.050.

## Plots Sections





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. This plot shows the relationship between sample size and power. We see that a sample size of about 46 is needed when D1 = 10 for 90% power, while D1 = 5 never reaches 90% power in this range of sample sizes.

## 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 D1.

### Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **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 **Test (Inequality)**, and then clicking on **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>
<b>Design Tab</b>	
Solve For .....	<b>Sample Size</b>
Alternative Hypothesis .....	<b>H1: Diff <math>\neq</math> D0 (Two-Sided)</b>
Power .....	<b>0.90</b>
Alpha .....	<b>0.05</b>
D0 (Mean Difference H0) .....	<b>0</b>
D1 (Mean Difference H1) .....	<b>5 10</b>
Standard Deviation Input Type .....	<b>SdPeriod (Std Dev of Period Differences)</b>
SdPeriod .....	<b>10</b>

### Output

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

#### Numeric Results

##### Numeric Results for 2x2 Cross-Over Design T-Test

Hypotheses: H0: Diff = D0 vs. H1: Diff  $\neq$  D0

Power	N	Null	Actual	SdPeriod	Effect	Alpha	Beta
		Difference	Difference				
		D0	D1		Size		
0.90323	172	0.000	5.000	10.000	0.354	0.050	0.09677
0.91250	46	0.000	10.000	10.000	0.707	0.050	0.08750

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 1933 presents an example in which  $D_0 = 0.0$ ,  $D_1 = 10$ ,  $S_w = 20$ ,  $\alpha = 0.05$ , 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 **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 **Test (Inequality)**, and then clicking on **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>
<b>Design Tab</b>	
Solve For .....	<b>Sample Size</b>
Alternative Hypothesis .....	<b>H1: Diff <math>\neq</math> D0 (Two-Sided)</b>
Power .....	<b>0.90</b>
Alpha .....	<b>0.05</b>
D0 (Mean Difference H0) .....	<b>0</b>
D1 (Mean Difference H1) .....	<b>10</b>
Standard Deviation Input Type .....	<b>Sw (Sqrt[Within MSE from RM ANOVA])</b>
Sw .....	<b>20</b>

### Output

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

### Numeric Results

Numeric Results for 2x2 Cross-Over Design T-Test								
Hypotheses: H0: Diff = D0 vs. H1: Diff $\neq$ D0								
Power	N	Null Difference D0	Actual Difference D1	SdPeriod	Effect Size	Alpha	Beta	
0.90648	88	0.000	10.000	20.000	0.500	0.050	0.09352	

PASS obtains 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.