

Chapter 520

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

Introduction

This procedure calculates power and sample size of statistical tests of equivalence of the means of a 2x2 cross-over design which is analyzed with a t-test. Schuirmann's (1987) two one-sided tests (TOST) approach is used to test equivalence. Only a brief introduction to the subject will be given here. For a comprehensive discussion on the subject, refer to Chow and Liu (1999) and Julious (2004).

Measurements are made on individuals that have been randomly assigned to one of two sequences. The first sequence receives the treatment followed by the reference (AB). The second sequence receives the reference followed by the treatment (BA). This *cross-over* design may be analyzed by a TOST equivalence test to show that the two means do not differ by more than a small amount, called the margin of equivalence.

The definition of equivalence has been refined in recent years using the concepts of prescribability and switchability. *Prescribability* refers to ability of a physician to prescribe either of two drugs at the beginning of the treatment. However, once prescribed, no other drug can be substituted for it. *Switchability* refers to the ability of a patient to switch from one drug to another during treatment without adverse effects. Prescribability is associated with equivalence of location and variability. Switchability is associated with the concept of individual equivalence. This procedure analyzes average equivalence. Thus, it partially analyzes prescribability. It does not address equivalence of variability or switchability.

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 *carryover* 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

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

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.

Outline of an Equivalence Test

PASS follows the *two one-sided tests* approach described by Schuirmann (1987) and Phillips (1990). It will be convenient to adopt the following specialized 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.
ϵ_U, ϵ_L	EU, EL	<i>Equivalence Limits.</i> These are the tolerance values that define the maximum changes that are not of practical importance. These may be thought of as the largest changes from the baseline that are considered to be trivial.
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.

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

The null hypothesis of non-equivalence is

$$H_0: \delta \leq \varepsilon_L \text{ or } \delta \geq \varepsilon_U \text{ where } \varepsilon_L < 0, \varepsilon_U > 0.$$

and the alternative hypothesis of equivalence is

$$H_1: \varepsilon_L < \delta < \varepsilon_U$$

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. 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 two, one-sided t -tests proceeds as follow

$$T_L = \frac{(\bar{x}_T - \bar{x}_R) - \varepsilon_L}{\hat{\sigma}_w \sqrt{\frac{2}{N}}} \text{ and } T_U = \frac{(\bar{x}_T - \bar{x}_R) - \varepsilon_U}{\hat{\sigma}_w \sqrt{\frac{2}{N}}}$$

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

The significance of each test statistic is determined by computing the p -value. If this p -value is less than a specified level (usually 0.05), the null hypothesis is rejected.

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

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

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

Computing the Power

The power of this test is given by

$$\Pr(T_L \geq t_{1-\alpha, N-2} \text{ and } T_U \leq -t_{1-\alpha, N-2})$$

where T_L and T_U are distributed as the bivariate, noncentral t distribution with noncentrality parameters Δ_L and Δ_U given by

$$\Delta_L = \frac{\delta - \varepsilon_L}{\sigma_w \sqrt{\frac{2}{N}}}$$

$$\Delta_U = \frac{\delta - \varepsilon_U}{\sigma_w \sqrt{\frac{2}{N}}}$$

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 the parameters associated with this test such as the means, sample sizes, alpha, and power.

Solve For

Solve For

This option specifies the parameter to be solved for from the other parameters. Under most situations, you will select either *Power* for a power analysis or *Sample Size* for sample size determination.

Select *Sample Size* when you want to calculate 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.

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 unequal means when in fact the means are equal.

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.

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

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 unequal means when in fact the means are unequal.

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.

When N is even, it is split evenly between the two sequences. When N is odd, the first sequence has one more subject than the second sequence.

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 – Equivalence Limits

EU (Upper Equivalence Limit)

This value is the upper limit of equivalence. Differences greater than this limit are not equivalent. You can enter values in the range $EU > 0$ and $EL < D1 < EU$.

EL (Lower Equivalence Limit)

This value is the lower limit of equivalence. Differences less than this limit are not equivalent. You can enter values in the range $EL < 0$ and $EL < D1 < EU$. If you want symmetric limits, enter “-Upper Limit” so that $EL = -EU$.

Effect Size – Mean Difference

D1 (Actual Difference to Detect)

This is the actual mean difference under the alternative hypothesis at which the power is calculated. You can enter values in the range $EL < D1 < EU$. Often this value is set to zero.

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 $SdPaired > 0$.

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

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

Example 1 – Finding Power

A cross-over design is to be used to compare the impact of two drugs on diastolic blood pressure. The average diastolic blood pressure after administration of the reference drug is known to be 96 mmHg. Researchers believe this average may drop to 92 mmHg with the use of a new drug. The within mean square error of similar studies is 324. Its square root is 18.

Following FDA guidelines, the researchers want to show that the diastolic blood pressure with the new drug is within 20% of the diastolic blood pressure with the reference drug. Thus, the equivalence limits of the mean difference of the two drugs are -19.2 and 19.2. They decide to calculate the power for a range of sample sizes between 6 and 100. 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 **Equivalence 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 **Equivalence**, and then clicking on **Equivalence 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
Alpha.....	0.05
N (Total Sample Size).....	6 10 16 20 40 60 80 100
EU (Upper Equivalence Limit)	19.2
EL (Lower Equivalence Limit)	-Upper Limit
D1 (Actual Difference to Detect)	-4
Standard Deviation Input Type	Sw (Sqrt[Within MSE from RM ANOVA])
Sw	18

Annotated Output

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

Numeric Results

Numeric Results for Testing Equivalence Using a 2x2 Cross-Over Design								
Hypotheses: $H_0: \text{Diff} \leq \text{EL or } \text{Diff} \geq \text{EU}$ vs. $H_1: \text{EL} < \text{Diff} < \text{EU}$								
Power	Total Sample Size	Lower Equivalence Limit	Upper Equivalence Limit	Actual Difference	Sw	Alpha	Beta	
	N	EL	EU	D1				
0.14704	6	-19.200	19.200	-4.000	18.000	0.050	0.85296	
0.38731	10	-19.200	19.200	-4.000	18.000	0.050	0.61269	
0.69965	16	-19.200	19.200	-4.000	18.000	0.050	0.30035	
0.81045	20	-19.200	19.200	-4.000	18.000	0.050	0.18955	
0.98042	40	-19.200	19.200	-4.000	18.000	0.050	0.01958	
0.99828	60	-19.200	19.200	-4.000	18.000	0.050	0.00172	
0.99987	80	-19.200	19.200	-4.000	18.000	0.050	0.00013	
0.99999	100	-19.200	19.200	-4.000	18.000	0.050	0.00001	

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

References

- Blackwelder, W.C. 1998. 'Equivalence Trials.' In Encyclopedia of Biostatistics, John Wiley and Sons. New York. Volume 2, 1367-1372.
- 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

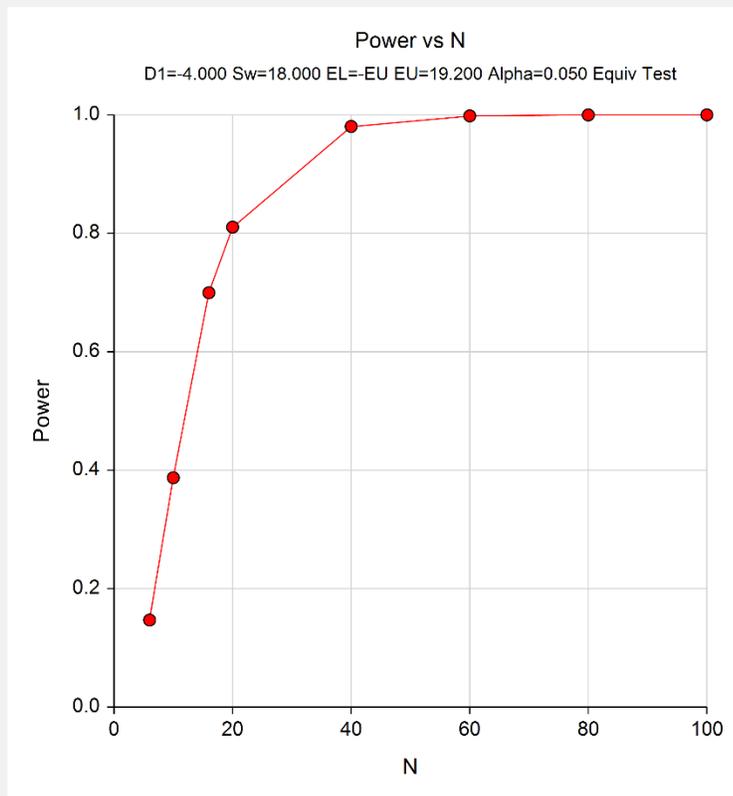
- Power is the probability of rejecting H_0 when it is false. It should be close to one.
- N is the total sample size drawn from all sequences. The sample is divided equally among sequences.
- EL & EU lower and upper equivalence limits, respectively, which are the maximum allowable differences that still result in equivalence.
- Diff = treatment mean - reference mean.
- D_1 is the actual mean difference under the alternative hypothesis at which the power is computed.
- S_w 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

An equivalence test of means using two one-sided tests on data from a 2x2 cross-over design with a total sample size of 6 achieves 15% power at a 5.0% significance level when the true difference between the means is -4.000, the square root of the within mean square error is 18.000, and the equivalence limits are -19.200 and 19.200.

This report shows the power for the indicated scenarios. Note that if they want 90% power, they will require a sample of around 30 subjects.

Plots Section



This plot shows the power versus the sample size.

Example 2 – Finding Sample Size

Continuing with Example 1, the researchers want to find the exact sample size needed to achieve both 80% power and 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 **Equivalence 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 **Equivalence**, and then clicking on **Equivalence 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
Power.....	0.80 0.90
Alpha.....	0.05
EU (Upper Equivalence Limit)	19.2
EL (Lower Equivalence Limit)	-Upper Limit
D1 (Actual Difference to Detect).....	-4
Standard Deviation Input Type	Sw (Sqrt[Within MSE from RM ANOVA])
Sw	18

Output

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

Numeric Results

Numeric Results for Testing Equivalence Using a 2x2 Cross-Over Design								
Hypotheses: $H_0: \text{Diff} \leq \text{EL or } \text{Diff} \geq \text{EU}$ vs. $H_1: \text{EL} < \text{Diff} < \text{EU}$								
	Total Sample Size	Lower Equivalence Limit	Upper Equivalence Limit	Actual Difference				
Power	N	EL	EU	D1	Sw	Alpha	Beta	
0.81045	20	-19.200	19.200	-4.000	18.000	0.050	0.18955	
0.90321	26	-19.200	19.200	-4.000	18.000	0.050	0.09679	

We note that 20 subjects are needed to achieve 80% power and 26 subjects are needed to achieve 90% power.

Example 3 – Validation using Phillips (1990)

Phillips (1990) page 142 presents a table of sample sizes for various parameter values. In this table, the treatment mean, standard deviation, and equivalence limits are all specified as percentages of the reference mean. We will reproduce the second line of the table in which the square root of the within mean square error is 20%; the equivalence limits are 20%; the treatment mean is 100%, 95%, 90%, and 85%; the power is 70%; and the significance level is 0.05. Phillips reports total sample size as 16, 20, 40, and 152 corresponding to the four treatment mean percentages. We will now setup this example in *PASS*.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Equivalence 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 **Equivalence**, and then clicking on **Equivalence 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
Power.....	0.7
Alpha.....	0.05
EU (Upper Equivalence Limit)	20
EL (Lower Equivalence Limit)	-Upper Limit
D1 (Actual Difference to Detect)	0 -5 -10 -15
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 Testing Equivalence Using a 2x2 Cross-Over Design

Hypotheses: $H_0: \text{Diff} \leq \text{EL or Diff} \geq \text{EU}$ vs. $H_1: \text{EL} < \text{Diff} < \text{EU}$

	Total Sample Size	Lower Equivalence Limit	Upper Equivalence Limit	Actual Difference			
Power	N	EL	EU	D1	Sw	Alpha	Beta
0.70012	152	-20.000	20.000	-15.000	20.000	0.050	0.29988
0.70922	40	-20.000	20.000	-10.000	20.000	0.050	0.29078
0.72205	20	-20.000	20.000	-5.000	20.000	0.050	0.27795
0.70310	16	-20.000	20.000	0.000	20.000	0.050	0.29690

Note that *PASS* obtains the same samples sizes as Phillips (1990).

Example 4 – Validation using Machin et al. (1997)

Machin *et al.* (1997) page 107 present an example of determining the sample size for a cross-over design in which the reference mean is 35.03, the treatment mean is 35.03, the standard deviation, entered as the square root of the within mean square error, is 40% of the reference mean, the limits are plus or minus 20% of the reference mean, the power is 80%, and the significance level is 0.10. Machin *et al.* calculate the total sample size to be 54.

When the parameters are given as percentages of the reference mean, it is easy enough to calculate the exact amounts by applying those percentages. However, the percentages can all be entered directly as long as all parameters (EU, EL, D, and Sw) are specified as percentages.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Equivalence 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 **Equivalence**, and then clicking on **Equivalence 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 4** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Sample Size
Power.....	0.80
Alpha.....	0.10
EU (Upper Equivalence Limit)	20
EL (Lower Equivalence Limit)	-Upper Limit
D1 (Actual Difference to Detect).....	0
Standard Deviation Input Type	Sw (Sqrt[Within MSE from RM ANOVA])
Sw.....	40

Output

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

Numeric Results

Numeric Results for Testing Equivalence Using a 2x2 Cross-Over Design								
Hypotheses: H0: Diff ≤ EL or Diff ≥ EU vs. H1: EL < Diff < EU								
	Total Sample Size	Lower Equivalence Limit	Upper Equivalence Limit	Actual Difference				
Power	N	EL	EU	D1	Sw	Alpha	Beta	
0.80497	54	-20.000	20.000	0.000	40.000	0.100	0.19503	

Note that **PASS** also has obtained a sample size of 54.

Example 5 – Validation using Chow and Liu (1999)

Chow and Liu (1999) page 153 present an example of determining the sample size for a cross-over design in which the reference mean is 82.559, the treatment mean is 82.559, the standard deviation, entered as the square root of the within mean square error, is 15.66%, the limits are plus or minus 20%, the power is 80%, and the significance level is 0.05. They calculate a sample size of 12. **PASS** calculates a sample size of 13. To see why **PASS** has increased the sample size by one, we will evaluate the power at sample sizes of 10, 12, 13, 14, and 16.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Equivalence 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 **Equivalence**, and then clicking on **Equivalence 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 5** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Power
Alpha.....	0.05
N (Total Sample Size).....	10 12 13 14 16
EU (Upper Equivalence Limit)	20
EL (Lower Equivalence Limit)	-Upper Limit
D1 (Actual Difference to Detect).....	0
Standard Deviation Input Type	Sw (Sqrt[Within MSE from RM ANOVA])
Sw	15.66

Output

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

Numeric Results

Numeric Results for Testing Equivalence Using a 2x2 Cross-Over Design								
Hypotheses: $H_0: \text{Diff} \leq \text{EL or } \text{Diff} \geq \text{EU}$ vs. $H_1: \text{EL} < \text{Diff} < \text{EU}$								
	Total Sample Size	Lower Equivalence Limit	Upper Equivalence Limit	Actual Difference				
Power	N	EL	EU	D1	Sw	Alpha	Beta	
0.66435	10	-20.000	20.000	0.000	15.660	0.050	0.33565	
0.79317	12	-20.000	20.000	0.000	15.660	0.050	0.20683	
0.83634	13	-20.000	20.000	0.000	15.660	0.050	0.16366	
0.87523	14	-20.000	20.000	0.000	15.660	0.050	0.12477	
0.92578	16	-20.000	20.000	0.000	15.660	0.050	0.07422	

The power for $N = 12$ is 0.79317. The power for $N = 13$ is 0.83634. Hence, to achieve better than 80% power, a sample size of 13 is necessary. However, 0.7932 is sufficiently close to 0.800 to make $N = 12$ a reasonable choice (as Chow and Liu did).

Example 6 – Validation using Senn (1993)

Senn (1993) page 217 presents an example of determining the sample size for a cross-over design in which the reference mean is equal to the treatment mean, the standard deviation, entered as the square root of the within mean square error, is 45, the equivalence limits are plus or minus 30, the power is 80%, and the significance level is 0.05. He calculates a sample size of 40.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Equivalence 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 **Equivalence**, and then clicking on **Equivalence 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 6** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Sample Size
Power	0.80
Alpha	0.05
EU (Upper Equivalence Limit)	30
EL (Lower Equivalence Limit)	-Upper Limit
D1 (Actual Difference to Detect)	0
Standard Deviation Input Type	Sw (Sqrt[Within MSE from RM ANOVA])
Sw	45

Output

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

Numeric Results

Numeric Results for Testing Equivalence Using a 2x2 Cross-Over Design								
Hypotheses: $H_0: \text{Diff} \leq \text{EL or } \text{Diff} \geq \text{EU}$ vs. $H_1: \text{EL} < \text{Diff} < \text{EU}$								
	Total Sample Size	Lower Equivalence Limit	Upper Equivalence Limit	Actual Difference				
Power	N	EL	EU	D1	Sw	Alpha	Beta	
0.80045	40	-30.000	30.000	0.000	45.000	0.050	0.19955	

PASS also calculates a sample size of 40.