Chapter 460

Equivalence Tests for the Difference Between Two Means

Introduction

This procedure allows you to study the power and sample size of equivalence tests of the means of two independent groups using the two-sample 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, refer to Chow and Liu (1999).

Measurements are made on individuals that have been randomly assigned to one of two groups. This *parallel-groups* design may be analyzed by a TOST equivalence test to show that the means of the two groups 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.

Parallel-Group Design

In a parallel-group design, subjects are assigned at random to either of two groups. Group 1 is the treatment group and group 2 is the reference group.

Outline of an Equivalence Test

PASS follows the *two one-sided tests* approach described by Schuirmann (1987) and Phillips (1990). Let $\mu_2 = \mu_T$ be the test group mean, $\mu_1 = \mu_R$ the reference group mean, and ε_L and ε_U the lower and upper bounds on $D = \mu_2 - \mu_1 = \mu_T - \mu_R$ that define the region of equivalence. The null hypothesis of non-equivalence is

$$H_0: D \le \varepsilon_L \quad or \quad H_0: D \ge \varepsilon_U$$

and the alternative hypothesis of equivalence is

$$H_1: \varepsilon_L < D < \varepsilon_U.$$

Two-Sample T-Test

This test assumes that the two groups of normally-distributed values have the same variance. The calculation of the two one-sided test statistics uses the following equations.

$$T_L = \frac{\left(\overline{X}_2 - \overline{X}_1\right) - \varepsilon_L}{s_{\overline{X}_1 - \overline{X}_2}} \quad \text{and} \quad T_U = \frac{\left(\overline{X}_2 - \overline{X}_1\right) - \varepsilon_U}{s_{\overline{X}_1 - \overline{X}_2}}$$

where

$$\overline{X}_k = \frac{\sum\limits_{i=1}^{N_k} X_{ki}}{N_k}$$

$$s_{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{\sum\limits_{i=1}^{N_1}\left(X_{1i} - \overline{X}_1\right)^2 + \sum\limits_{i=1}^{N_2}\left(X_{2i} - \overline{X}_2\right)^2}{N_1 + N_2 - 2}\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}$$

$$df = N_1 + N_2 - 2$$

The null hypothesis is rejected if T_L and $-T_U$ are greater than or equal to $t_{1-\alpha, N_1 + N_2 - 2}$.

The power of this test is given by

$$\Pr(T_L \geq t_{1-\alpha, N_1 + N_2 - 2} \ \text{and} \ T_U \leq -t_{1-\alpha, N_1 + N_2 - 2} | \mu_T, \mu_R, \sigma^2) \, 1$$

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

$$\Delta_L = \frac{D - \varepsilon_L}{\sigma\sqrt{\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}}$$

$$\Delta_U = \frac{D - \varepsilon_U}{\sigma\sqrt{\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}}$$

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* or *Sample Size (N1)*.

Select *Sample Size (N1)* 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 nonequivalent means when in fact the means are equivalent.

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 non-equivalent means when in fact the means are nonequivalent.

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 (When Solving for Sample Size)

Group Allocation

Select the option that describes the constraints on *N1* or *N2* or both.

The options are

- **Equal (N1 = N2)**

  This selection is used when you wish to have equal sample sizes in each group. Since you are solving for both sample sizes at once, no additional sample size parameters need to be entered.

- **Enter N1, solve for N2**

  Select this option when you wish to fix *N1* at some value (or values), and then solve only for *N2*. Please note that for some values of *N1*, there may not be a value of *N2* that is large enough to obtain the desired power.

- **Enter N2, solve for N1**

  Select this option when you wish to fix *N2* at some value (or values), and then solve only for *N1*. Please note that for some values of *N2*, there may not be a value of *N1* that is large enough to obtain the desired power.

- **Enter R = N2/N1, solve for N1 and N2**

  For this choice, you set a value for the ratio of *N2* to *N1*, and then PASS determines the needed *N1* and *N2*, with this ratio, to obtain the desired power. An equivalent representation of the ratio, *R*, is

  $$N2 = R * N1.$$

- **Enter percentage in Group 1, solve for N1 and N2**

  For this choice, you set a value for the percentage of the total sample size that is in Group 1, and then PASS determines the needed *N1* and *N2* with this percentage to obtain the desired power.

N1 (Sample Size, Group 1)

This option is displayed if Group Allocation = "Enter N1, solve for N2"

N1 is the number of items or individuals sampled from the Group 1 population.

N1 must be ≥ 2. You can enter a single value or a series of values.

N2 (Sample Size, Group 2)

This option is displayed if Group Allocation = "Enter N2, solve for N1"

N2 is the number of items or individuals sampled from the Group 2 population.

N2 must be ≥ 2. You can enter a single value or a series of values.

R (Group Sample Size Ratio)

This option is displayed only if Group Allocation = "Enter R = N2/N1, solve for N1 and N2."

R is the ratio of *N2* to *N1*. That is,

$$R = N2 / N1.$$

Use this value to fix the ratio of *N2* to *N1* while solving for *N1* and *N2*. Only sample size combinations with this ratio are considered.

N2 is related to *N1* by the formula:

$$N2 = [R \times N1],$$

where the value *[Y]* is the next integer $\geq Y$.

For example, setting $R = 2.0$ results in a Group 2 sample size that is double the sample size in Group 1 (e.g., $N1 = 10$ and $N2 = 20$, or $N1 = 50$ and $N2 = 100$).

R must be greater than 0. If $R < 1$, then $N2$ will be less than $N1$; if $R > 1$, then $N2$ will be greater than $N1$. You can enter a single or a series of values.

Percent in Group 1

This option is displayed only if Group Allocation = "Enter percentage in Group 1, solve for N1 and N2."

Use this value to fix the percentage of the total sample size allocated to Group 1 while solving for $N1$ and $N2$. Only sample size combinations with this Group 1 percentage are considered. Small variations from the specified percentage may occur due to the discrete nature of sample sizes.

The Percent in Group 1 must be greater than 0 and less than 100. You can enter a single or a series of values.

Sample Size (When <u>Not</u> Solving for Sample Size)

Group Allocation

Select the option that describes how individuals in the study will be allocated to Group 1 and to Group 2.

The options are

- **Equal (N1 = N2)**

  This selection is used when you wish to have equal sample sizes in each group. A single per group sample size will be entered.

- **Enter N1 and N2 individually**

  This choice permits you to enter different values for $N1$ and $N2$.

- **Enter N1 and R, where N2 = R * N1**

  Choose this option to specify a value (or values) for $N1$, and obtain $N2$ as a ratio (multiple) of $N1$.

- **Enter total sample size and percentage in Group 1**

  Choose this option to specify a value (or values) for the total sample size (N), obtain $N1$ as a percentage of N, and then $N2$ as $N - N1$.

Sample Size Per Group

This option is displayed only if Group Allocation = "Equal (N1 = N2)."

The Sample Size Per Group is the number of items or individuals sampled from each of the Group 1 and Group 2 populations. Since the sample sizes are the same in each group, this value is the value for $N1$, and also the value for $N2$.

The Sample Size Per Group must be ≥ 2. You can enter a single value or a series of values.

N1 (Sample Size, Group 1)

*This option is displayed if Group Allocation = "Enter N1 and N2 individually" or "Enter N1 and R, where N2 = R * N1."*

$N1$ is the number of items or individuals sampled from the Group 1 population.

$N1$ must be ≥ 2. You can enter a single value or a series of values.

N2 (Sample Size, Group 2)

This option is displayed only if Group Allocation = "Enter N1 and N2 individually."

N2 is the number of items or individuals sampled from the Group 2 population.

N2 must be ≥ 2. You can enter a single value or a series of values.

R (Group Sample Size Ratio)

This option is displayed only if Group Allocation = "Enter N1 and R, where N2 = R * N1."

R is the ratio of *N2* to *N1*. That is,

$$R = N2/N1$$

Use this value to obtain *N2* as a multiple (or proportion) of *N1*.

N2 is calculated from *N1* using the formula:

$$N2=[R \ x \ N1],$$

where the value *[Y]* is the next integer $\geq Y$.

For example, setting *R* = 2.0 results in a Group 2 sample size that is double the sample size in Group 1.

R must be greater than 0. If *R* < 1, then *N2* will be less than N1; if *R* > 1, then *N2* will be greater than *N1*. You can enter a single value or a series of values.

Total Sample Size (N)

This option is displayed only if Group Allocation = "Enter total sample size and percentage in Group 1."

This is the total sample size, or the sum of the two group sample sizes. This value, along with the percentage of the total sample size in Group 1, implicitly defines *N1* and *N2*.

The total sample size must be greater than one, but practically, must be greater than 3, since each group sample size needs to be at least 2.

You can enter a single value or a series of values.

Percent in Group 1

This option is displayed only if Group Allocation = "Enter total sample size and percentage in Group 1."

This value fixes the percentage of the total sample size allocated to Group 1. Small variations from the specified percentage may occur due to the discrete nature of sample sizes.

The Percent in Group 1 must be greater than 0 and less than 100. You can enter a single value or a series of values.

Effect Size – Equivalence Limits

|EU| Upper Equivalence Limit

This value gives upper limit on equivalence. Differences outside EL and EU are not considered equivalent. Differences between them are considered equivalent.

Note that EL<0 and EU>0. Also, you must have EL<D<EU.

-|EL| Lower Equivalence Limit

This value gives lower limit on equivalence. Differences outside EL and EU are not considered equivalent. Differences between them are.

If you want symmetric limits, enter -UPPER LIMIT for EL to force EL = -|EU|.

Note that EL<0 and EU>0. Also, you must have EL<D<EU. Finally, the scale of these numbers must match the scale of S.

Effect Size – True Mean Difference

D (True Difference)

This is the true difference between the two means at which the power is to be computed. Often this value is set to zero, but it can be non-zero as long as it is between the equivalence limits EL and EU.

Effect Size – Standard Deviation

S (Standard Deviation)

Specify the within-group standard deviation, σ. The standard deviation is assumed to be the same for both groups.

Example 1 – Parallel-Group Design

A parallel-group is to be used to compare influence of two drugs on diastolic blood pressure. The diastolic blood pressure is known to be close to 96 mmHg with the reference drug and is thought to be 92 mmHg with the experimental drug. Based on similar studies, the within-group standard deviation is set to 18mmHg. Following FDA guidelines, the researchers want to show that the diastolic blood pressure with the experimental drug is within 20% of the diastolic blood pressure with the reference drug. Note that 20% of 96 is 19.2. They decide to calculate the power for a range of sample sizes between 3 and 60. 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** procedure window by expanding **Means**, then **Two Independent Means**, then clicking on **Equivalence**, and then clicking on **Equivalence Tests for the Difference Between Two Means**. 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**
Alpha..	**0.05**
Group Allocation	**Equal (N1 = N2)**
Sample Size Per Group	**3 5 8 10 15 20 30 40 50 60**
|EU| Upper Equivalence Limit.................	**19.2**
-|EL| Lower Equivalence Limit	**-Upper Limit**
D (True Difference)	**-4**
S (Standard Deviation)	**18**

Annotated Output

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

Numeric Results

Testing Equivalence of Two Means Using a Parallel-Group Design

Power	N1	N2	N	D	SD	Lower Equiv. Limit	Upper Equiv. Limit	Alpha
0.03856	3	3	6	-4.0	18.0	-19.2	19.2	0.050
0.09277	5	5	10	-4.0	18.0	-19.2	19.2	0.050
0.28871	8	8	16	-4.0	18.0	-19.2	19.2	0.050
0.43913	10	10	20	-4.0	18.0	-19.2	19.2	0.050
0.69339	15	15	30	-4.0	18.0	-19.2	19.2	0.050
0.82662	20	20	40	-4.0	18.0	-19.2	19.2	0.050
0.94326	30	30	60	-4.0	18.0	-19.2	19.2	0.050
0.98205	40	40	80	-4.0	18.0	-19.2	19.2	0.050
0.99458	50	50	100	-4.0	18.0	-19.2	19.2	0.050
0.99843	60	60	120	-4.0	18.0	-19.2	19.2	0.050

References
Blackwelder, W.C. 1998. 'Equivalence Trials.' In Encyclopedia of Biostatistics, John Wiley and Sons. New York.
    Volume 2, 1367-1372.
Chow, S.C.; Shao, J.; 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.
Julious, Steven A. 2010. Sample Sizes for Clinical Trials. Chapman & Hall / CRC. Boca Raton, Florida.
Phillips, Kem F. 1990. 'Power of the Two One-Sided Tests Procedure in Bioequivalence', Journal of
    Pharmacokinetics and Biopharmaceutics, Volume 18, No. 2, pages 137-144.
Schuirmann, Donald. 1987. 'A Comparison of the Two One-Sided Tests Procedure and the Power Approach for
    Assessing the Equivalence of Average Bioavailability', Journal of Pharmacokinetics and Biopharmaceutics,
    Volume 15, Number 6, pages 657-680.

Report Definitions
Power is the probability of rejecting a false null hypothesis.
N1 and N2 are the number of items sampled from each population.
N is the total sample size, N1 + N2.
D is the true difference between the means.
SD is the within-group standard deviation for the two groups.
The lower and upper equivalence limits are the maximum allowable differences that still result in equivalence.
Alpha is the probability of rejecting a true null hypothesis.

Summary Statements
An equivalence test of means using two one-sided tests on data from a parallel-group design
with sample sizes of 3 in the reference group and 3 in the treatment group achieves 4% power at
a 5.0% significance level when the true difference between the means is -4.0, the standard
deviation is 18.0, and the equivalence limits are -19.2 and 19.2.

This report shows the power for the indicated parameter configurations. Note that when the parameters are specified as percentages, they are displayed in the output with percent signs. Note that the desired 80% power occurs for a per group sample size between 15 and 20.

Plot Section
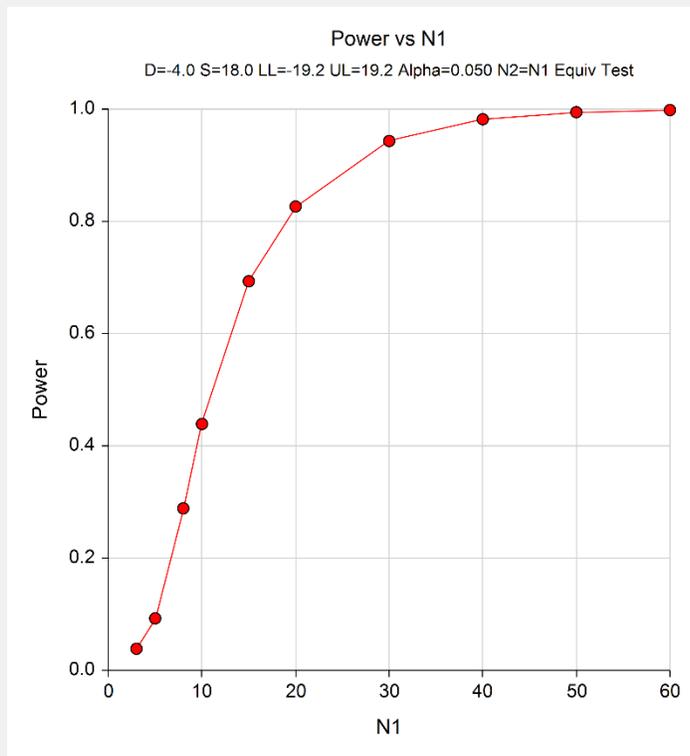
This plot shows the power versus the sample size.

Example 2 –Validation using Julious (2010)

Julious (2010) page 87 presents an example of determining the sample size for testing equivalence in a pain control trial in which the margin of equivalence for the difference is ± 10mm, the actual differences are 0mm (Example 5.1) and 2mm (Example 5.2), the standard deviation is 100mm, the power is 90%, and the significance level is 0.025 with equal sample allocation. Julius (2010) calculates the sample sizes to be 2600 for an actual difference of 0mm and 3306 for an actual difference of 2mm.

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** procedure window by expanding **Means**, then **Two Independent Means**, then clicking on **Equivalence**, and then clicking on **Equivalence Tests for the Difference Between Two Means**. 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**

Alpha **0.025**

Group Allocation .... **Equal (N1 = N2)**

|EU| Upper Equivalence Limit **10**

-|EL| Lower Equivalence Limit .... **-Upper Limit**

D (True Difference) **0 2**

S (Standard Deviation) **100**

Output

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

Numeric Results

Testing Equivalence of Two Means Using a Parallel-Group Design

Target Power	Actual Power	N1	N2	N	D	SD	Lower Equiv. Limit	Upper Equiv. Limit	Alpha
0.90	0.90011	2600	2600	5200	0.0	100.0	-10.0	10.0	0.025
0.90	0.90006	3305	3305	6610	2.0	100.0	-10.0	10.0	0.025

Note that **PASS** obtains sample sizes of 2600 and 3305, which are equal to those calculated by Julious (2010) with slight differences due to rounding.

Example 3 –Validation using Machin et al. (1997)

Machin *et al.* (1997) page 107 present an example of determining the sample size for a parallel-group design in which the reference mean is 96, the treatment mean is 94, the standard deviation is 8, the limits are plus or minus 5, the power is 80%, and the significance level is 0.05. They calculate the sample size to be 88. It is important to note that Machin *et al.* use an approximation, so their results should not be expected to exactly match the results obtained using **PASS**.

We will now set up 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** procedure window by expanding **Means**, then **Two Independent Means**, then clicking on **Equivalence**, and then clicking on **Equivalence Tests for the Difference Between Two Means**. 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.. **0.80**
Alpha... **0.05**
Group Allocation .... **Equal (N1 = N2)**
|EU| Upper Equivalence Limit................. **5**
-|EL| Lower Equivalence Limit .... **-Upper Limit**
D (True Difference)................................. **-2**
S (Standard Deviation) **8**

Output

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

Numeric Results

Testing Equivalence of Two Means Using a Parallel-Group Design									
Target Power	Actual Power	N1	N2	N	D	SD	Lower Equiv. Limit	Upper Equiv. Limit	Alpha
0.80	0.80151	89	89	178	-2.0	8.0	-5.0	5.0	0.050

Note that **PASS** obtains a sample size of 89 which is very close to the approximate value of 88 that Machin calculated.