Chapter 529

Two-Sample T-Tests for Equivalence Allowing Unequal Variance

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 unequal-variance *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 $\mu_1 = \mu_T$ be the test group mean, $\mu_2 = \mu_R$ be the reference group mean, and E_L and E_U be the lower and upper bounds, respectively, on $\delta = \mu_1 - \mu_2 = \mu_T - \mu_R$ that define the region of equivalence.

It will be convenient to adopt the following specialized notation for the discussion of these tests.

Parameter	PASS Input/Output	Interpretation
μ_1 or μ_T	Not used	*Mean* of population 1. Population 1 is assumed to consist of those who have received the new treatment.
μ_2 or μ_R	Not used	*Mean* of population 2. Population 2 is assumed to consist of those who have received the reference treatment.
E_L, E_U	EL, EU	*Lower* and *Upper Equivalence Limits.* If the difference is between these two limits, the new treatment is said to be *equivalent* to the reference.
δ	δ	*Actual difference.* This is the value of $\mu_1 - \mu_2$, the difference between the means. This is the value at which the power is calculated.

Note that the actual values of μ_1 and μ_2 are not needed. Only their difference is needed for power and sample size calculations.

With $E_L < 0$ and $E_U > 0$, the null hypothesis of non-equivalence is

$$H_0: \delta \le E_L \ \text{ or } \ \delta \ge E_U.$$

The alternative hypothesis of equivalence is

$$H_1: E_L < \delta < E_U.$$

Two-Sample Unequal-Variance T-Test (Welch's T-Test) Statistics

This test assumes that the two groups of normally-distributed values have unequal variances. Based on the unequal-variance *t*-statistic proposed in Welch (1938), the calculation of the two one-sided test statistics uses the following equations.

$$t_L^* = \frac{(\bar{X}_1 - \bar{X}_2) - E_L}{s_{\bar{X}_1 - \bar{X}_2}^*}$$

$$t_U^* = \frac{(\bar{X}_1 - \bar{X}_2) - E_U}{s_{\bar{X}_1 - \bar{X}_2}^*}$$

where

$$\bar{X}_k = \frac{\sum_{i=1}^{n_k} X_{ki}}{n_k},$$

$$s_k = \sqrt{\left(\frac{\sum_{i=1}^{n_k}(X_{ki} - \bar{X}_k)^2}{(n_k - 1)}\right)},$$

$$s_{\bar{X}_1 - \bar{X}_2}^* = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}},$$

$$df = \frac{s_{\bar{X}_1-\bar{X}_2}^{*4}}{\dfrac{s_1^4}{n_1^2(n_1-1)} + \dfrac{s_2^4}{n_2^2(n_2-1)}}.$$

The null hypothesis is rejected if t_L^* and $-t_U^*$ are both greater than or equal to $t_{1-\alpha,df}$.

Power Calculation

When $\sigma_1 \neq \sigma_2$, the power of the unequal-variance equivalence *t*-test is calculated as

$$\Pr\left(t_L^* \geq t_{1-\alpha,df} \text{ and } t_U^* \leq -t_{1-\alpha,df} | \mu_1, \mu_2, \sigma_1, \sigma_2\right)$$

where t_L^* and t_U^* are distributed as the bivariate, noncentral *t* distribution with noncentrality parameters Δ_L and Δ_U given by

$$\Delta_L = \frac{\delta - E_L}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

$$\Delta_U = \frac{\delta - E_U}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

and

$$df = \frac{\left(\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}\right)^4}{\dfrac{\sigma_1^4}{n_1^2(n_1-1)} + \dfrac{\sigma_2^4}{n_2^2(n_2-1)}}.$$

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 calculated from the values of the other parameters.

Select *Sample Size* when you want to determine the sample size needed to achieve a given power and alpha.
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 ≥ *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 Not 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 ≥ *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)

Enter one or more values for the upper limit of equivalence. If the actual difference is between EL and EU, the new treatment is said to be equivalent to the standard. You may enter values in the range $EU > 0$ and $EL < \delta < EU$.

EL (Lower Equivalence Limit)

Enter one or more values for the lower limit of equivalence. If the actual difference is between EL and EU, the new treatment is said to be equivalent to the standard. You may enter values in the range $EL < 0$ and $EL < \delta < EU$.

If you want symmetric limits around zero, enter "-Upper Limit" here. The program will set $EL = -|EU|$.

Effect Size – Mean Difference

δ (Actual Difference)

This is the actual difference between the treatment mean and the reference mean at which the power is calculated. For equivalence tests, this value is often set to zero. You may enter values in the range $EL < \delta < EU$.

Effect Size – Standard Deviations

σ1 (Standard Deviation of Group 1)

The standard deviation entered here is the assumed standard deviation for the Group 1 population. $\sigma 1$ must be a positive number.

When $\sigma 1$ is not known, you must supply an estimate. Press the small 'σ' button to the right to obtain calculation options for estimating the standard deviation.

σ2 (Standard Deviation of Group 2)

The standard deviation entered here is the assumed standard deviation for the Group 2 population. $\sigma 2$ must be a positive number.

When $\sigma 2$ is not known, you must supply an estimate. Press the small 'σ' button to the right to obtain calculation options for estimating the standard deviation.

Example 1 – Finding Power

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 for the treatment group and 15mmHg for the reference group. 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 **Two-Sample T-Tests for Equivalence Allowing Unequal Variance** procedure window by expanding **Means**, then **Two Independent Means**, then clicking on **Equivalence**, and then clicking on **Two-Sample T-Tests for Equivalence Allowing Unequal Variance**. 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**
δ (Actual Difference)..............................	**-4**
σ1 (Standard Deviation of Group 1)	**18**
σ2 (Standard Deviation of Group 2)	**15**

Annotated Output

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

Numeric Results

Numeric Results for Two One-Sided Unequal-Variance T-Tests ─────────────────────────────

$\delta = \mu_1 - \mu_2 = \mu_T - \mu_R$

Hypotheses: H0: $\delta \le$ EL or $\delta \ge$ EU   vs.   H1: EL $< \delta <$ EU

Power	N1	N2	N	Lower Equiv Limit EL	Upper Equiv Limit EU	δ	σ1	σ2	Alpha
0.10733	3	3	6	-19.2	19.2	-4.0	18.0	15.0	0.050
0.17783	5	5	10	-19.2	19.2	-4.0	18.0	15.0	0.050
0.40940	8	8	16	-19.2	19.2	-4.0	18.0	15.0	0.050
0.55247	10	10	20	-19.2	19.2	-4.0	18.0	15.0	0.050
0.77233	15	15	30	-19.2	19.2	-4.0	18.0	15.0	0.050
0.88100	20	20	40	-19.2	19.2	-4.0	18.0	15.0	0.050
0.96794	30	30	60	-19.2	19.2	-4.0	18.0	15.0	0.050
0.99238	40	40	80	-19.2	19.2	-4.0	18.0	15.0	0.050
0.99818	50	50	100	-19.2	19.2	-4.0	18.0	15.0	0.050
0.99959	60	60	120	-19.2	19.2	-4.0	18.0	15.0	0.050

Two-Sample T-Tests for Equivalence Allowing Unequal Variance

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., and Lokhnygina, Y. 2018. Sample Size Calculations in Clinical Research, Third
   Edition. Taylor & Francis/CRC. Boca Raton, Florida.
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 = N1 + N2 is the total sample size.
EL and EU are the lower and upper equivalence limits, respectively, and are the maximum allowable differences
   that still result in equivalence.
$\delta = \mu_1 - \mu_2 = \mu_T - \mu_R$ is the difference between the treatment and reference means at which power and sample
   size calculations are made.
σ_1 and σ_2 are the assumed population standard deviations for groups 1 and 2, respectively.
Alpha is the probability of rejecting a true null hypothesis.

Summary Statements ───
An equivalence test of means using two one-sided unequal-variance t-tests with sample sizes of
3 in the treatment group and 3 in the reference group achieves 11% power at a 5.0% significance
level when the the equivalence limits are -19.2 and 19.2, the actual difference between the
means is -4.0, and the standard deviations are 18.0 in group 1 and 15.0 in group 2.

This report shows the power for the indicated parameter configurations. Note that the desired 80% power occurs for a per group sample size between 15 and 20.

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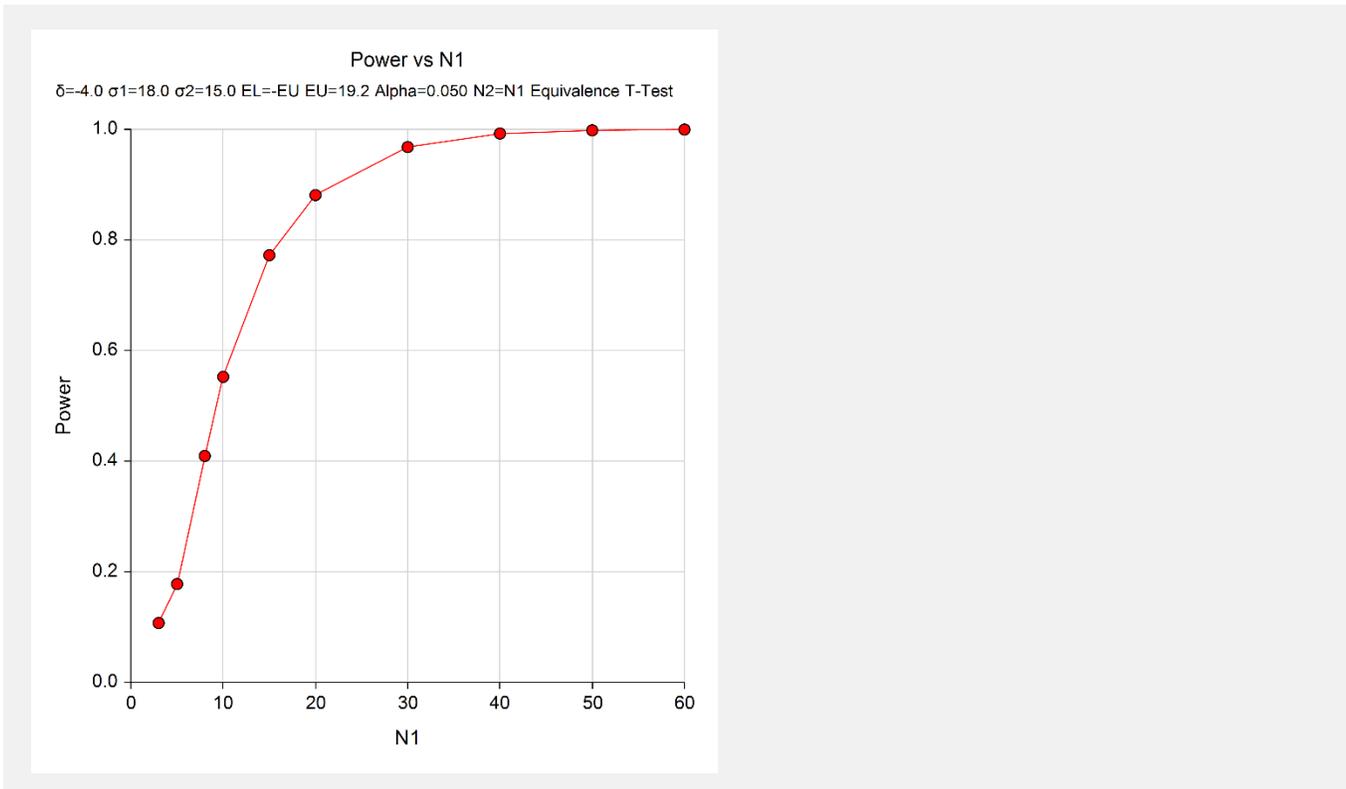
This plot shows the power versus the sample size.

Example 2 – Finding the Sample Size

Continuing with Example 1, the researchers want to know the exact sample size to achieve 80% 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 **Two-Sample T-Tests for Equivalence Allowing Unequal Variance** procedure window by expanding **Means**, then **Two Independent Means**, then clicking on **Equivalence**, and then clicking on **Two-Sample T-Tests for Equivalence Allowing Unequal Variance**. 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)	**19.2**
EL (Lower Equivalence Limit).................	**-Upper Limit**
δ (Actual Difference)...............................	**-4**
σ1 (Standard Deviation of Group 1)	**18**
σ2 (Standard Deviation of Group 2)	**15**

Output

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

Numeric Results

Numeric Results for Two One-Sided Unequal-Variance T-Tests ———————————————————————
$\delta = \mu_1 - \mu_2 = \mu_T - \mu_R$
Hypotheses: H0: $\delta \le$ EL or $\delta \ge$ EU   vs.   H1: EL $< \delta <$ EU

Target Power	Actual Power	N1	N2	N	Lower Equiv Limit EL	Upper Equiv Limit EU	δ	σ1	σ2	Alpha
0.80	0.80027	16	16	32	-19.2	19.2	-4.0	18.0	15.0	0.050

This report shows the exact sample size required for 80% power.

Example 3 – Validation of Sample Size using Simulation

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 for an equal-variance test.

If we assume that the standard deviation of group 1 is 8 and the standard deviation of group 2 is 6 and use Welch's unequal-variance test, then the Equivalence Tests for Two Means (Simulation) procedure, which itself has been validated, calculates a sample size of 70 in each group.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Two-Sample T-Tests for Equivalence Allowing Unequal Variance** procedure window by expanding **Means**, then **Two Independent Means**, then clicking on **Equivalence**, and then clicking on **Two-Sample T-Tests for Equivalence Allowing Unequal Variance**. 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**.

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**
δ (Actual Difference)..............................	**-2**
σ1 (Standard Deviation of Group 1)	**8**
σ2 (Standard Deviation of Group 2)	**6**

Output

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

Numeric Results

Numeric Results for Two One-Sided Unequal-Variance T-Tests ─────────────────────────
$\delta = \mu_1 - \mu_2 = \mu_T - \mu_R$
Hypotheses: H0: $\delta \leq$ EL or $\delta \geq$ EU vs. H1: EL $< \delta <$ EU

Target Power	Actual Power	N1	N2	N	Lower Equiv Limit EL	Upper Equiv Limit EU	δ	σ1	σ2	Alpha
0.80	0.80456	70	70	140	-5.0	5.0	-2.0	8.0	6.0	0.050

PASS obtains a sample size of 70 which matches the simulation result exactly.