Chapter 460

Two-Sample T-Tests for Equivalence Assuming Equal 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 equal-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

$$H_0: \delta \leq E_L \text{ or } \delta \geq E_U.$$

The alternative hypothesis of equivalence is

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

Two-Sample Equal-Variance T-Test Statistics

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.

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

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

where

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

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

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)},$$

$$df = n_1 + n_2 - 2.$$

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

Power Calculation

When $\sigma_1 = \sigma_2 = \sigma$, the power of the equal-variance equivalence *t*-test is calculated as

$$\Pr\left(t_L \geq t_{1-\alpha,n_1+n_2-2} \text{ and } t_U \leq -t_{1-\alpha,n_1+n_2-2} | \mu_1, \mu_2, \sigma\right)$$

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

$$\Delta_L = \frac{\delta - E_L}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$\Delta_U = \frac{\delta - E_U}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Example 1 – Finding Power

A parallel-group design 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

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

Design Tab	
Solve For	**Power**
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**
σ (Standard Deviation)	**18**

Output

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

Numeric Reports

Numeric Results

Solve For: Power
Difference: δ = μ1 - μ2 = μT - μR
Hypotheses: H0: δ ≤ EL or δ ≥ EU vs. H1: EL < δ < EU
Test Type: Two One-Sided Equal-Variance T-Tests

	Sample Size			Equivalence Limits				
Power	N1	N2	N	Lower EL	Upper EU	Actual Difference δ	Standard Deviation σ	Alpha
0.03856	3	3	6	-19.2	19.2	-4	18	0.05
0.09277	5	5	10	-19.2	19.2	-4	18	0.05
0.28871	8	8	16	-19.2	19.2	-4	18	0.05
0.43913	10	10	20	-19.2	19.2	-4	18	0.05
0.69339	15	15	30	-19.2	19.2	-4	18	0.05
0.82662	20	20	40	-19.2	19.2	-4	18	0.05
0.94326	30	30	60	-19.2	19.2	-4	18	0.05
0.98205	40	40	80	-19.2	19.2	-4	18	0.05
0.99458	50	50	100	-19.2	19.2	-4	18	0.05
0.99843	60	60	120	-19.2	19.2	-4	18	0.05

Power	The probability of rejecting a false null hypothesis when the alternative hypothesis is true.
N1 and N2	The number of items sampled from each population.
N	The total sample size. N = N1 + N2.
EL and EU	The lower and upper equivalence limits, respectively, and are the maximum allowable differences that still result in equivalence.
δ	The difference between the treatment and reference means at which power and sample size calculations are made. δ = μ1 - μ2 = μT - μR.
σ	The assumed population standard deviation for each of the two groups.
Alpha	The probability of rejecting a true null hypothesis.

Summary Statements

A parallel two-group design will be used to test whether the Group 1 (treatment) mean (μ1) is equivalent to the Group 2 (reference) mean (μ2), with difference equivalence bounds of -19.2 and 19.2 (H0: δ ≤ -19.2 or δ ≥ 19.2 versus H1: -19.2 < δ < 19.2, δ = μ1 - μ2 = μT - μR). The comparison will be made using two one-sided, two-sample, equal-variance t-tests, with an overall Type I error rate (α) of 0.05. The common standard deviation for both groups is assumed to be 18. To detect a mean difference (δ = μ1 - μ2 = μT - μR) of -4 with sample sizes of 3 for Group 1 (treatment) and 3 for Group 2 (reference), the power is 0.03856.

Dropout-Inflated Sample Size

	Sample Size			Dropout-Inflated Enrollment Sample Size			Expected Number of Dropouts		
Dropout Rate	N1	N2	N	N1'	N2'	N'	D1	D2	D
20%	3	3	6	4	4	8	1	1	2
20%	5	5	10	7	7	14	2	2	4
20%	8	8	16	10	10	20	2	2	4
20%	10	10	20	13	13	26	3	3	6
20%	15	15	30	19	19	38	4	4	8
20%	20	20	40	25	25	50	5	5	10
20%	30	30	60	38	38	76	8	8	16
20%	40	40	80	50	50	100	10	10	20
20%	50	50	100	63	63	126	13	13	26
20%	60	60	120	75	75	150	15	15	30

Dropout Rate	The percentage of subjects (or items) that are expected to be lost at random during the course of the study and for whom no response data will be collected (i.e., will be treated as "missing"). Abbreviated as DR.
N1, N2, and N	The evaluable sample sizes at which power is computed (as entered by the user). If N1 and N2 subjects are evaluated out of the N1' and N2' subjects that are enrolled in the study, the design will achieve the stated power.
N1', N2', and N'	The number of subjects that should be enrolled in the study in order to obtain N1, N2, and N evaluable subjects, based on the assumed dropout rate. N1' and N2' are calculated by inflating N1 and N2 using the formulas N1' = N1 / (1 - DR) and N2' = N2 / (1 - DR), with N1' and N2' always rounded up. (See Julious, S.A. (2010) pages 52-53, or Chow, S.C., Shao, J., Wang, H., and Lokhnygina, Y. (2018) pages 32-33.)
D1, D2, and D	The expected number of dropouts. D1 = N1' - N1, D2 = N2' - N2, and D = D1 + D2.

Dropout Summary Statements

Anticipating a 20% dropout rate, 4 subjects should be enrolled in Group 1, and 4 in Group 2, to obtain final group sample sizes of 3 and 3, respectively.

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.

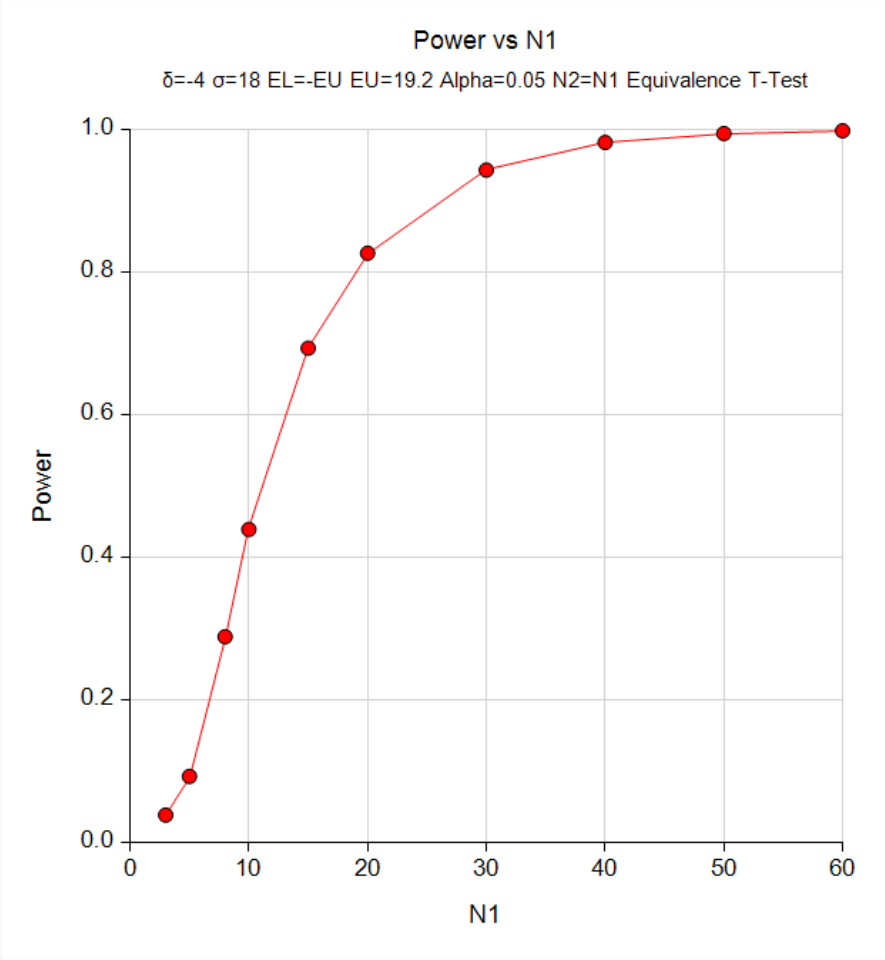
Phillips, Kem F. 1990. 'Power of the Two One-Sided Tests Procedure in Bioequivalence', Journal of Pharmacokinetics and Biopharmaceutics, Volume 18, No. 2, pages 137-144.

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.

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.

Plots Section

Plots



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

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

Design Tab

Solve For	**Sample Size**
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**
σ (Standard Deviation)	**18**

Output

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

Numeric Results

Solve For: Sample Size
Difference: δ = μ1 - μ2 = μT - μR
Hypotheses: H0: δ ≤ EL or δ ≥ EU vs. H1: EL < δ < EU
Test Type: Two One-Sided Equal-Variance T-Tests

Power		Sample Size			Equivalence Limits				
Target	Actual	N1	N2	N	Lower EL	Upper EU	Actual Difference δ	Standard Deviation σ	Alpha
0.8	0.80601	19	19	38	-19.2	19.2	-4	18	0.05

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

Example 3 – 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

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

Design Tab

Solve For	**Sample Size**
Power	**0.90**
Alpha	**0.025**
Group Allocation	**Equal (N1 = N2)**
EU (Upper Equivalence Limit)	**10**
EL (Lower Equivalence Limit)	**-Upper Limit**
δ (Actual Difference)	**0 2**
σ (Standard Deviation)	**100**

Output

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

Numeric Results

Solve For: Sample Size
Difference: δ = μ1 - μ2 = μT - μR
Hypotheses: H0: δ ≤ EL or δ ≥ EU vs. H1: EL < δ < EU
Test Type: Two One-Sided Equal-Variance T-Tests

Power		Sample Size			Equivalence Limits				
Target	Actual	N1	N2	N	Lower EL	Upper EU	Actual Difference δ	Standard Deviation σ	Alpha
0.9	0.90011	2600	2600	5200	-10	10	0	100	0.025
0.9	0.90006	3305	3305	6610	-10	10	2	100	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 4 – Validation using Machin et al. (1997)

Machin *et al.* (1997) page 107 presents 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

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 4** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab	
Solve For	**Sample Size**
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**
σ (Standard Deviation)	**8**

Output

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

Numeric Results

Solve For: Sample Size
Difference: δ = μ1 - μ2 = μT - μR
Hypotheses: H0: δ ≤ EL or δ ≥ EU vs. H1: EL < δ < EU
Test Type: Two One-Sided Equal-Variance T-Tests

Power		Sample Size			Equivalence Limits				
Target	Actual	N1	N2	N	Lower EL	Upper EU	Actual Difference δ	Standard Deviation σ	Alpha
0.8	0.80151	89	89	178	-5	5	-2	8	0.05

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