

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	<i>Mean of population 1.</i> Population 1 is assumed to consist of those who have received the new treatment.
μ_2 or μ_R	Not used	<i>Mean of population 2.</i> Population 2 is assumed to consist of those who have received the reference treatment.
E_L, E_U	EL, EU	<i>Lower and Upper Equivalence Limits.</i> If the difference is between these two limits, the new treatment is said to be <i>equivalent</i> to the reference.
δ	δ	<i>Actual difference.</i> 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 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.

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

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

Two-Sample T-Tests for Equivalence Allowing Unequal Variance

where

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

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

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

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

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

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

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

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

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

and

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

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 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

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_1 (Standard Deviation of Group 1).....	18
σ_2 (Standard Deviation of Group 2).....	15

Two-Sample T-Tests for Equivalence Allowing Unequal Variance

Output

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

Numeric Reports

Numeric Results

Solve For: **Power**

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

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

Test Type: Two One-Sided Unequal-Variance T-Tests

Power	Sample Size			Equivalence Limits		Actual Difference δ	Standard Deviation		Alpha
	N1	N2	N	Lower EL	Upper EU		σ_1	σ_2	
0.10733	3	3	6	-19.2	19.2	-4	18	15	0.05
0.17783	5	5	10	-19.2	19.2	-4	18	15	0.05
0.40940	8	8	16	-19.2	19.2	-4	18	15	0.05
0.55247	10	10	20	-19.2	19.2	-4	18	15	0.05
0.77233	15	15	30	-19.2	19.2	-4	18	15	0.05
0.88100	20	20	40	-19.2	19.2	-4	18	15	0.05
0.96794	30	30	60	-19.2	19.2	-4	18	15	0.05
0.99238	40	40	80	-19.2	19.2	-4	18	15	0.05
0.99818	50	50	100	-19.2	19.2	-4	18	15	0.05
0.99959	60	60	120	-19.2	19.2	-4	18	15	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. $N = N_1 + N_2$.

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

σ_1 and σ_2 The assumed population standard deviations for groups 1 and 2, respectively.

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 (μ_1) is equivalent to the Group 2 (reference) mean (μ_2), with difference equivalence bounds of -19.2 and 19.2 ($H_0: \delta \leq -19.2 \text{ or } \delta \geq 19.2$ versus $H_1: -19.2 < \delta < 19.2$, $\delta = \mu_1 - \mu_2 = \mu_T - \mu_R$). The comparison will be made using two one-sided, two-sample, unequal-variance (Welch's) t-tests, with an overall Type I error rate (α) of 0.05. The standard deviation for Group 1 is assumed to be 18 and the standard deviation for Group 2 is assumed to be 15. To detect a mean difference ($\delta = \mu_1 - \mu_2 = \mu_T - \mu_R$) of -4 with sample sizes of 3 for Group 1 (treatment) and 3 for Group 2 (reference), the power is 0.10733.

Two-Sample T-Tests for Equivalence Allowing Unequal Variance

Dropout-Inflated Sample Size

Dropout Rate	Sample Size			Dropout-Inflated Enrollment Sample Size			Expected Number of Dropouts		
	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 Lohknygina, 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

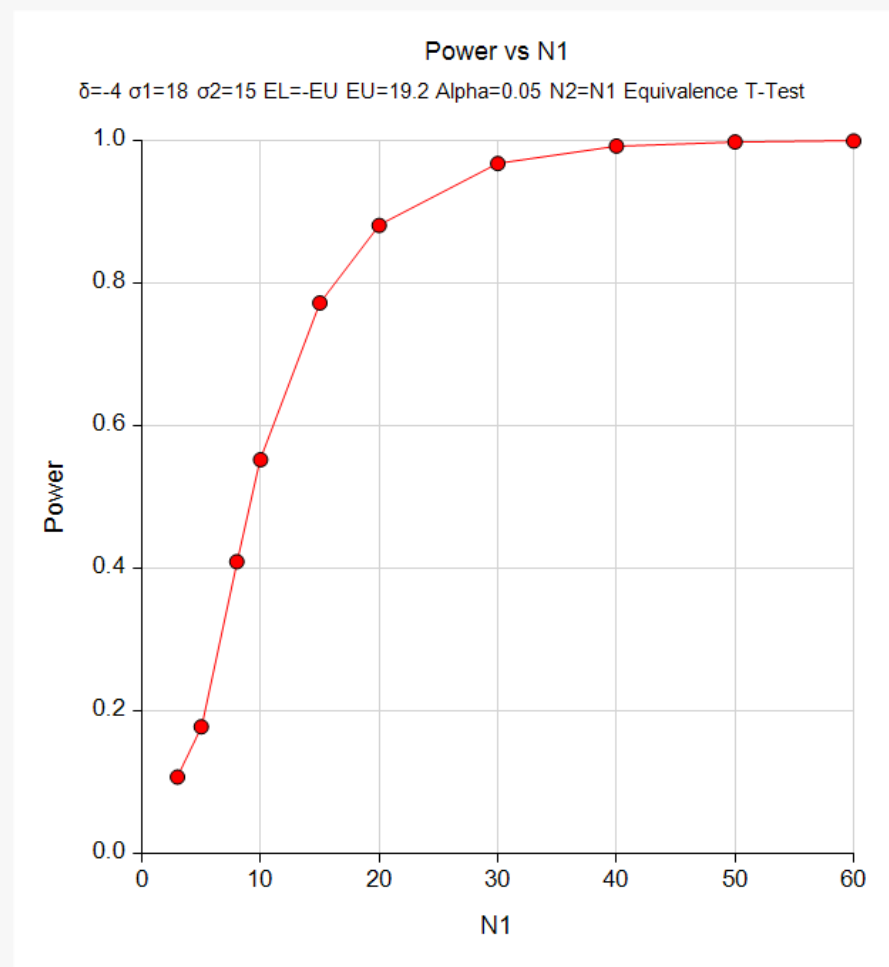
- Blackwelder, W.C. 1998. 'Equivalence Trials.' In Encyclopedia of Biostatistics, John Wiley and Sons. New York. Volume 2, 1367-1372.
- Chow, S.C., Shao, J., Wang, H., and Lohknygina, Y. 2018. Sample Size Calculations in Clinical Research, Third Edition. Taylor & Francis/CRC. Boca Raton, Florida.
- Julious, 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.
- Schuurmann, 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.
- Welch, B.L. 1938. 'The significance of the difference between two means when the population variances are unequal.' Biometrika, 29, 350-362.

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.

Two-Sample T-Tests for Equivalence Allowing Unequal Variance

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_1 (Standard Deviation of Group 1)..... **18**
 σ_2 (Standard Deviation of Group 2)..... **15**

Output

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

Numeric Results

Solve For: [Sample Size](#)
 Difference: $\delta = \mu_1 - \mu_2 = \mu_T - \mu_R$
 Hypotheses: $H_0: \delta \leq EL \text{ or } \delta \geq EU$ vs. $H_1: EL < \delta < EU$
 Test Type: Two One-Sided Unequal-Variance T-Tests

Power		Sample Size			Equivalence Limits		Actual Difference δ	Standard Deviation		Alpha
Target	Actual	N1	N2	N	Lower EL	Upper EU		σ_1	σ_2	
0.8	0.80027	16	16	32	-19.2	19.2	-4	18	15	0.05

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

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..... **0.80**
 Alpha..... **0.05**
 Group Allocation **Equal (N1 = N2)**
 EU (Upper Equivalence Limit)..... **5**
 EL (Lower Equivalence Limit) **-Upper Limit**
 δ (Actual Difference) **-2**
 σ_1 (Standard Deviation of Group 1)..... **8**
 σ_2 (Standard Deviation of Group 2)..... **6**

Output

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

Numeric Results

Solve For: **Sample Size**
 Difference: $\delta = \mu_1 - \mu_2 = \mu_T - \mu_R$
 Hypotheses: $H_0: \delta \leq EL \text{ or } \delta \geq EU$ vs. $H_1: EL < \delta < EU$
 Test Type: Two One-Sided Unequal-Variance T-Tests

Power		Sample Size			Equivalence Limits		Actual Difference δ	Standard Deviation		Alpha
Target	Actual	N1	N2	N	Lower EL	Upper EU		σ_1	σ_2	
0.8	0.80456	70	70	140	-5	5	-2	8	6	0.05

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