

Chapter 519

Paired T-Tests for Equivalence

Introduction

This procedure allows you to study the power and sample size of tests of equivalence of means of two correlated (paired) variables. Schuirmann's (1987) two one-sided tests (TOST) approach is used to test equivalence. The paired t-test is commonly used in this situation.

Paired data may occur because two measurements are made on the same subject or because measurements are made on two subjects that have been matched according to other, often demographic, variables. Hypothesis tests on paired data can be analyzed by considering the differences between the paired items. The distribution of these differences is usually symmetric. In fact, the distribution must be symmetric if the individual distributions of the two items are identical. Hence, the paired t-test is appropriate for paired data even when the distributions of the individual items are not normal.

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.

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 specialize notation for the discussion of these tests.

Parameter	PASS Input/Output	Interpretation
δ	δ	<i>Population mean of paired differences.</i> This is the mean of paired differences, $X_1 - X_2$, in the population. This parameter will be estimated by the study.
δ_1	$\delta 1$	<i>Actual paired difference at which power is calculated.</i> This is the value of the mean paired difference at which power is calculated.
E_L, E_U	EL, EU	<i>Lower and Upper Equivalence Limits.</i> If the average paired difference is between these two limits, the treatment is said to be <i>equivalent</i> to the reference.

Paired T-Tests for Equivalence

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

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

The alternative hypothesis of equivalence is

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

Paired T-Test Statistic

A paired t -test is used to analyze the data when the standard deviation of paired differences is unknown and estimated by 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

$$t_L = \frac{\bar{d} - E_L}{s_d/\sqrt{N}}$$

$$t_U = \frac{\bar{d} - E_U}{s_d/\sqrt{N}}$$

where s_d is the sample standard deviation of the paired differences. The test is usually calculated using a $100(1 - 2\alpha)\%$ confidence interval of the mean difference. If both limits of this confidence interval are between E_L and E_U , equivalence is concluded.

The Standard Deviation of Paired Differences (σ)

If you have results from a previous (or pilot) study, use the estimate of the standard deviation of paired differences, σ , from the study. Another reasonable (but somewhat rough) estimate of σ may be obtained using the range of paired differences as

$$\sigma = \frac{\text{Range}}{4}$$

If you have estimates of the expected standard deviations of the paired variables (σ_1 and σ_2) and the Pearson correlation between the paired variables (ρ), the standard deviation of paired differences (σ) may be calculated using the equation

$$\sigma^2 = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2$$

such that

$$\sigma = \sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}.$$

Paired T-Tests for Equivalence

If $\sigma_1 = \sigma_2 = \sigma_x$, then this formula reduces to

$$\sigma^2 = 2\sigma_x^2(1 - \rho)$$

such that

$$\sigma = \sqrt{2\sigma_x^2(1 - \rho)}.$$

If you have an estimate of the within-subject population standard deviation (σ_w), then σ may be calculated using the equation

$$\sigma^2 = 2\sigma_w^2$$

such that

$$\sigma = \sqrt{2\sigma_w^2}.$$

σ_w is often estimated by the square root of the within mean square error (WMSE) from a repeated measures ANOVA.

Power Calculation for the Paired T-Test

The power of this test is

$$\Pr(t_L \geq t_{1-\alpha, N-1} \text{ and } t_U \leq t_{\alpha, N-1})$$

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

$$\Delta_L = \frac{\delta - E_L}{\sigma\sqrt{1/N}}$$

$$\Delta_U = \frac{\delta - E_U}{\sigma\sqrt{1/N}}$$

Example 1 – Finding Power

A paired 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 (σ_w^2) of similar studies is 324. Its square root (σ_w) is 18. Therefore, the estimate of σ may be calculated as

$$\sigma = \sqrt{2 \times 324} = 25.45584$$

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 5 and 50. 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
N (Sample Size).....	5 10 15 20 30 40 50
EU (Upper Equivalence Limit).....	19.2
EL (Lower Equivalence Limit)	-Upper Limit
δ_1 (Mean of Paired Differences)	-4
Standard Deviation Input Type	Enter the Within-Subject Population SD
σ_w (Within-Subject Population SD).....	18

Paired T-Tests for Equivalence

Output

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

Numeric Reports

Numeric Results

Solve For: **Power**

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

Power	Sample Size N	Equivalence Limits		Mean of Paired Differences δ_1	Standard Deviation		Alpha
		Lower EL	Upper EU		σ_w	σ	
0.10599	5	-19.2	19.2	-4	18	25.45584	0.05
0.39640	10	-19.2	19.2	-4	18	25.45584	0.05
0.66629	15	-19.2	19.2	-4	18	25.45584	0.05
0.81214	20	-19.2	19.2	-4	18	25.45584	0.05
0.93855	30	-19.2	19.2	-4	18	25.45584	0.05
0.98051	40	-19.2	19.2	-4	18	25.45584	0.05
0.99410	50	-19.2	19.2	-4	18	25.45584	0.05

Power The probability of rejecting a false null hypothesis when the alternative hypothesis is true.

N The sample size, the number of subjects (or pairs) in the study.

EL The minimum allowable mean that still results in equivalence.

EU The maximum allowable mean that still result in equivalence.

δ The population mean of paired differences.

δ_1 The value of the mean of paired differences at which power and sample size are calculated.

σ_w The within-subject population standard deviation. σ_w is estimated as the square root of the within mean square error (WMSE) in a repeated measures ANOVA.

σ The standard deviation of paired differences, which was calculated from σ_w using the equation, $\sigma = \sqrt{2\sigma_w^2}$.

Alpha The probability of rejecting a true null hypothesis.

Summary Statements

A paired design will be used to test whether the treatment 1 mean (μ_1) is equivalent to the treatment 2 mean (μ_2), with lower and upper mean difference equivalence limits of -19.2 and 19.2 ($H_0: \delta \leq -19.2 \text{ or } \delta \geq 19.2$ versus $H_1: -19.2 < \delta < 19.2, \delta = \mu_1 - \mu_2$). The comparison will be made using two one-sided, paired-difference t-tests, with an overall Type I error rate (α) of 0.05. The underlying standard deviation of the paired difference distribution is assumed to be 25.45584 (calculated using $\sigma_w = 18$ and the equation, $\sigma = \sqrt{2\sigma_w^2}$). To detect a paired mean difference of -4 with a sample size of 5 pairs, the power is 0.10599.

Paired T-Tests for Equivalence

Dropout-Inflated Sample Size

Dropout Rate	Sample Size N	Dropout- Inflated Enrollment Sample Size N'	Expected Number of Dropouts D
20%	5	7	2
20%	10	13	3
20%	15	19	4
20%	20	25	5
20%	30	38	8
20%	40	50	10
20%	50	63	13

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.
N	The evaluable sample size at which power is computed (as entered by the user). If N subjects are evaluated out of the N' subjects that are enrolled in the study, the design will achieve the stated power.
N'	The total number of subjects that should be enrolled in the study in order to obtain N evaluable subjects, based on the assumed dropout rate. N' is calculated by inflating N using the formula $N' = N / (1 - DR)$, with N' 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.)
D	The expected number of dropouts. $D = N' - N$.

Dropout Summary Statements

Anticipating a 20% dropout rate, 7 subjects should be enrolled to obtain a final sample size of 5 subjects.

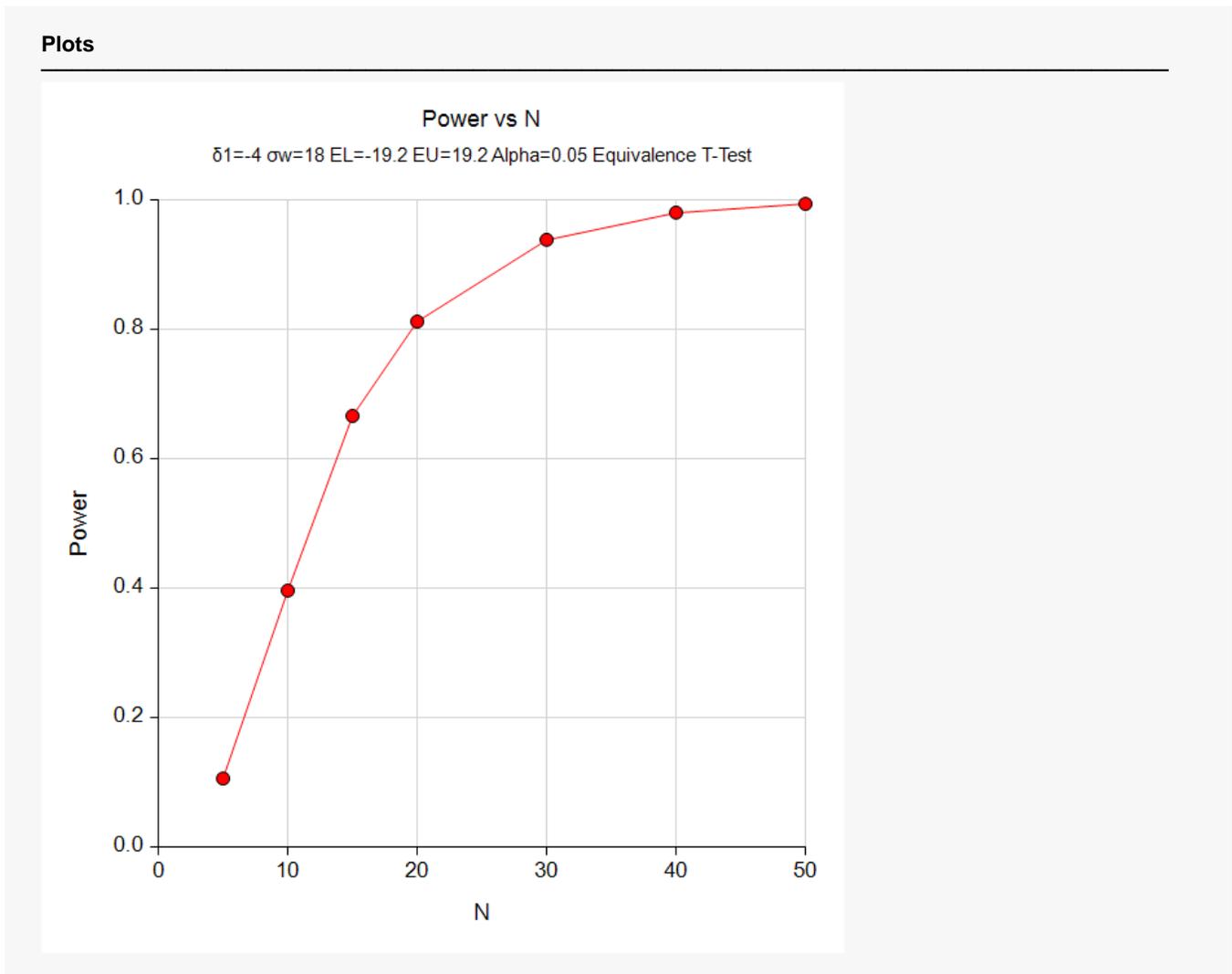
References

- 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.
- Mathews, Paul. 2010. Sample Size Calculations - Practical Methods for Engineers and Scientists. Mathews Malnar and Bailey. Fairport Harbor, OH.
- Blackwelder, W.C. 1998. 'Equivalence Trials.' In Encyclopedia of Biostatistics, John Wiley and Sons. New York. Volume 2, 1367-1372.

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.

Paired T-Tests for Equivalence

Plots Section



This plot shows the power versus the sample size.

Example 2 – Validation using Chow, Shao, Wang, and Lokhnygina (2018)

Chow, Shao, Wang, and Lokhnygina (2018) presents an example on pages 46 and 47 of a paired equivalence *t*-test sample size calculation in which $EU = 0.05$, $EL = -0.05$, $\delta_1 = 0.0$, $\sigma = 0.1$, $\alpha = 0.05$, and $\text{power} = 0.80$. They obtain a sample size of 36.

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
EU (Upper Equivalence Limit).....	0.05
EL (Lower Equivalence Limit)	-Upper Limit
δ_1 (Mean of Paired Differences)	0
Standard Deviation Input Type	Enter the SD of Paired Differences
σ (SD of Paired Differences).....	0.1

Output

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

Numeric Results						
Solve For:	Sample Size					
Hypotheses:	$H_0: \delta \leq EL \text{ or } \delta \geq EU$ vs. $H_1: EL < \delta < EU$					
Paired Differences						
Power	Sample Size N	Equivalence Limits		Mean δ_1	Standard Deviation σ	Alpha
		Lower EL	Upper EU			
0.80515	36	-0.05	0.05	0	0.1	0.05

The sample size of 36 matches Chow, Shao, Wang, and Lokhnygina (2018) exactly.

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 (σ_w) is 20%; the equivalence limits are 20%; the treatment mean is 100%, 95%, 90%, and 85% (δ_1 is 0, -5, -10, -15); the power is 70%; and the significance level is 0.05. A \sqrt{MSE} of 20% equates to a σ of 28.28427%.

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

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..... **0.70**
 Alpha..... **0.05**
 EU (Upper Equivalence Limit)..... **20**
 EL (Lower Equivalence Limit) **-Upper Limit**
 δ_1 (Mean of Paired Differences) **0 -5 -10 -15**
 Standard Deviation Input Type **Enter the Within-Subject Population SD**
 σ_w (Within-Subject Population SD)..... **20**

Output

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

Numeric Results

Solve For: **Sample Size**
 Hypotheses: $H_0: \delta \leq EL \text{ or } \delta \geq EU$ vs. $H_1: EL < \delta < EU$

Power	Sample Size N	Equivalence Limits		Mean of Paired Differences δ_1	Standard Deviation		Alpha
		Lower EL	Upper EU		σ_w	σ	
0.70014	152	-20	20	-15	20	28.28427	0.05
0.70958	40	-20	20	-10	20	28.28427	0.05
0.72395	20	-20	20	-5	20	28.28427	0.05
0.70749	16	-20	20	0	20	28.28427	0.05

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