Chapter 512

Paired Z-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 z-test may be used in this situation when the standard deviation of paired differences is known.

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

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

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

Paired Z-Test Statistic

A paired z-test is used to analyze the data when the standard deviation of paired differences is known. The test assumes that the data are a simple random sample from a population of normally-distributed values that have the same known variance. This assumption implies that the differences are continuous and normal. The calculation of the two, one-sided z-tests proceeds as follow

$$z_L = \frac{\bar{d} - E_L}{\sigma/\sqrt{N}}$$

$$z_U = \frac{\bar{d} - E_U}{\sigma/\sqrt{N}}$$

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

Power Calculation

The power of this test is

$$\Pr(z_L \geq z_{1-\alpha} \ \text{ and } \ z_U \leq z_\alpha)$$

where z_L and z_U are distributed as standard normal.

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* for a power analysis or *Sample Size* for sample size determination.

Select *Sample Size* 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. 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. 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

N (Sample Size)

This option specifies one or more values of the sample size, the number of individuals in the study (total subjects). This value must be an integer greater than one.

You may enter a list of values using the syntax *50,100,150,200,250* or *50 to 250 by 50*.

Effect Size – Equivalence Limits

EU (Upper Equivalence Limit)

Enter one or more values for the upper limit of equivalence. If the actual mean of paired differences is between EL and EU, the new treatment is said to be equivalent to the standard. Enter values in the range EU > 0 and $EL < \delta 1 < EU$.

EL (Lower Equivalence Limit))

Enter one or more values for the lower limit of equivalence. If the actual mean of paired differences is between EL and EU, the new treatment is said to be equivalent to the standard. Enter values in the range EL < 0 and $EL < \delta 1 < EU$. If you want symmetric limits around zero, enter "-Upper Limit" here. The program will set EL = -|EU|.

Effect Size – Mean

δ1 (Mean of Paired Differences)

Enter one or more values for the actual value of the mean of paired differences at which power and sample size will be calculated. This value is often set to zero for paired data, but it can be non-zero as long as $EL < \delta 1 < EU$.

Effect Size – Standard Deviation

σ (Std Dev of Paired Differences)

This option specifies one or more values of the known standard deviation. This must be a positive value.

PASS includes a special module for estimating the standard deviation. This module may be loaded by pressing the *SD* button. Refer to the Standard Deviation Estimator chapter for further details.

These values must be positive.

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 known σ is 25.

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

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Paired Z-Tests for Equivalence** procedure window by expanding **Means**, then **Paired Means**, then **Equivalence**, and then clicking on **Paired Z-Tests for Equivalence**. 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**
N (Sample Size)......................................	**5 10 15 20 30 40 50**
EU (Upper Equivalence Limit)	**19.2**
EL (Lower Equivalence Limit)................	**-Upper Limit**
δ1 (Mean of Paired Differences).............	**-4**
σ (Std Dev of Paired Differences)..........	**25**

Annotated Output

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

Numeric Results

Numeric Results ──
Hypotheses: H0: δ ≤ EL or δ ≥ EU   vs.   H1: EL < δ < EU

Power	Sample Size N	Lower Equiv Limit EL	Upper Equiv Limit EU	Mean of Paired Differences δ1	Standard Deviation σ	Alpha
0.05418	5	-19.200	19.200	-4.000	25.000	0.050
0.51085	10	-19.200	19.200	-4.000	25.000	0.050
0.73549	15	-19.200	19.200	-4.000	25.000	0.050
0.85252	20	-19.200	19.200	-4.000	25.000	0.050
0.95374	30	-19.200	19.200	-4.000	25.000	0.050
0.98610	40	-19.200	19.200	-4.000	25.000	0.050
0.99603	50	-19.200	19.200	-4.000	25.000	0.050

Paired Z-Tests for Equivalence

References
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.
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.

Report Definitions
Power is the probability of rejecting the null hypothesis when it is false. It should be close to one.
N is the sample size, the number of subjects (or pairs) in the study.
EL is the minimum allowable mean that still results in equivalence.
EU is the maximum allowable mean that still result in equivalence.
$\delta 1$ is the value of the mean of paired differences at which power and sample size are calculated.
σ is the standard deviation of paired differences for the population.
Alpha is the probability of rejecting the null hypothesis when it is true. It should be small.

Summary Statements ——
In an equivalence test of the mean difference from data in a paired design using two one-sided
z-tests, a sample size of 5 achieves 5% power at a 5.0% significance level when the actual mean
of paired differences is -4.000, the known standard deviation of paired differences is 25.000,
and the equivalence limits are -19.200 and 19.200.

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.

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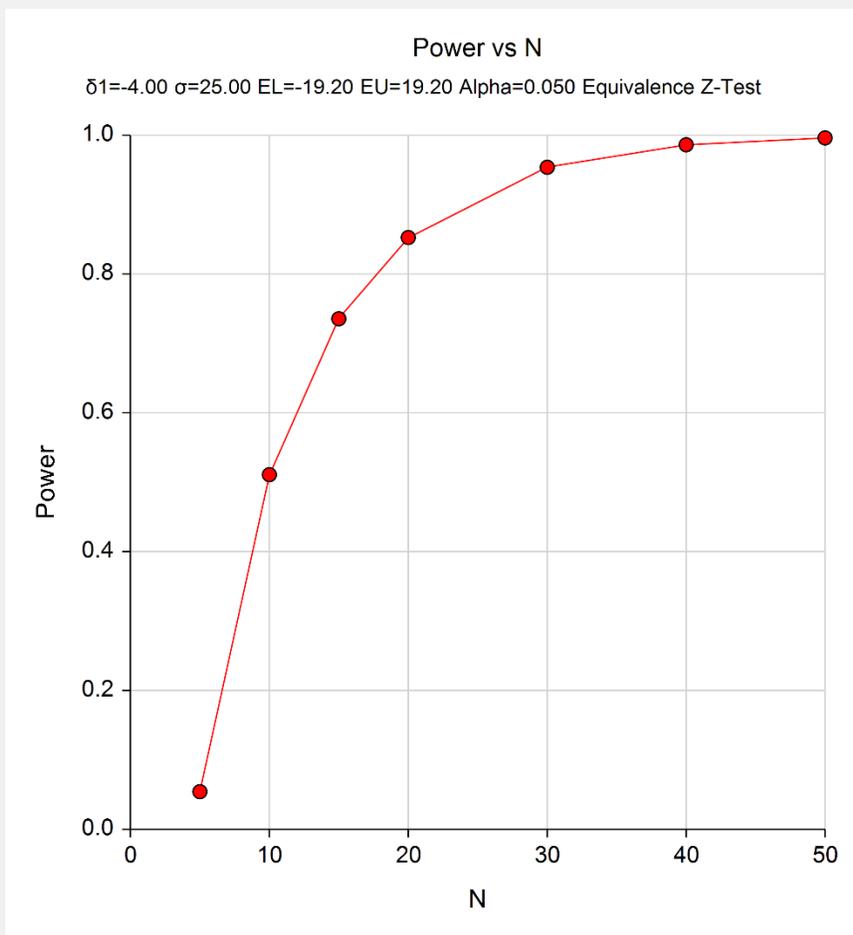
This plot shows the power versus the sample size.

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 *z*-test sample size calculation in which EU = 0.05, EL = -0.05, $\delta 1$ = 0.0, σ = 0.1, alpha = 0.05, and power = 0.80. They obtain a sample size of 35.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Paired Z-Tests for Equivalence** procedure window by expanding **Means**, then **Paired Means**, then **Equivalence**, and then clicking on **Paired Z-Tests for Equivalence**. 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.. **0.8**
Alpha... **0.05**
EU (Upper Equivalence Limit) **0.05**
EL (Lower Equivalence Limit)................. **-Upper Limit**
$\delta 1$ (Mean of Paired Differences)............. **0**
σ (Std Dev of Paired Differences).......... **0.1**

Output

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

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

Power	Sample Size N	Lower Equiv Limit EL	Upper Equiv Limit EU	Mean of Paired Differences $\delta 1$	Standard Deviation σ	Alpha
0.81088	35	-0.050	0.050	0.000	0.100	0.050

The sample size of 35 matches Chow, Shao, Wang, and Lokhnygina (2018) exactly.