

Chapter 550

Tests for Paired Means (Simulation)

Introduction

This procedure allows you to study the power and sample size of several statistical tests of the null hypothesis that the difference between two correlated means is equal to a specific value versus the alternative that it is greater than, less than, or not equal to that value. The paired t-test is commonly used in this situation. Other tests have been developed for the case when the data are not normally distributed. These additional tests include the Wilcoxon signed-ranks test, the sign test, and the computer-intensive bootstrap test.

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 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 and the Wilcoxon signed-rank test are appropriate for paired data even when the distributions of the individual items are not normal.

The details of the power analysis of the paired t-test using analytic techniques are presented in another **PASS** chapter and they won't be duplicated here. This chapter will only consider power analysis using computer simulation.

Technical Details

Computer simulation allows one to estimate the power and significance level that is actually achieved by a test procedure in situations that are not mathematically tractable. Computer simulation was once limited to mainframe computers. Currently, due to increased computer speeds, simulation studies can be completed on desktop and laptop computers in a reasonable period of time.

The steps to a simulation study are as follows.

1. Specify the method by which the test is to be carried out. This includes specifying how the test statistic is calculated and how the significance level is specified.
2. Generate a random sample, X_1, X_2, \dots, X_n , from the distribution specified by the alternative hypothesis. Calculate the test statistic from the simulated data and determine if the null hypothesis is accepted or rejected. Each of these samples is used to calculate the power of the test.
3. Generate a random sample, Y_1, Y_2, \dots, Y_n , from the distribution specified by the null hypothesis. Calculate the test statistic from the simulated data and determine if the null hypothesis is accepted or rejected. Each of these samples is used to calculate the significance level of the test.
4. Repeat steps 2 and 3 several thousand times, tabulating the number of times the simulated data lead to a rejection of the null hypothesis. The power is the proportion of simulation samples in step 2 that lead to rejection. The significance level is the proportion of simulated samples in step 3 that lead to rejection.

Tests for Paired Means (Simulation)

Data Distributions

A wide variety of distributions may be studied. These distributions can vary in skewness, elongation, or other features such as bimodality. A detailed discussion of the distributions that may be used in the simulation is provided in the chapter 'Data Simulator'.

Test Statistics

This section describes the test statistics that are available in this procedure.

Paired T-Test

The paired t -test assumes that the paired differences, X_i , are a simple random sample from a population of normally-distributed difference values that all have the same mean and variance. This assumption implies that the data are continuous, and their distribution is symmetric. The calculation of the t -test proceeds as follows

$$t_{n-1} = \frac{\bar{X} - D0}{s_{\bar{X}}/\sqrt{n}}$$

where

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n},$$

$$s_{\bar{X}} = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}},$$

The significance of the test statistic is determined by computing the p-value. If this p-value is less than a specified level (usually 0.05), the hypothesis is rejected. Otherwise, no conclusion can be reached.

Wilcoxon Signed-Rank Test

The Wilcoxon signed-rank test is a popular, nonparametric substitute for the t -test. It assumes that the data follow a symmetric distribution. The test is computed using the following steps.

1. Subtract the hypothesized difference, $D0$, from each data value. Rank the values according to their absolute values.
2. Compute the sum of the positive ranks Sp and the sum of the negative ranks Sn . The test statistic, W_R , is the minimum of Sp and Sn .
3. Compute the mean and standard deviation of W_R using the formulas

$$\mu_{W_R} = \frac{n(n+1)}{4}$$

$$\sigma_{W_R} = \sqrt{\frac{n(n+1)(2n+1)}{24} - \frac{\sum t^3 - \sum t}{48}}$$

where t represents the number of times the i^{th} value occurs.

4. Compute the z -value using

$$z_W = \frac{W_R - \mu_{W_R}}{\sigma_{W_R}}$$

Tests for Paired Means (Simulation)

For cases when n is less than 38, the significance level is found from a table of exact probabilities for the Wilcoxon test. When n is greater than or equal to 38, the significance of the test statistic is determined by comparing the z value to a normal probability table. If this p -value is less than a specified level (often 0.05), the null hypothesis is rejected. Otherwise, no conclusion can be reached.

Sign Test

The sign test is popular because it is simple to compute. This test assumes that the data all follow the same distribution. The test is computed using the following steps.

1. Count the number of values strictly greater than $D0$. Call this value X .
2. Count the number of values strictly less than $D0$. Call this value Y .
3. Set $m = X + Y$.
4. Under the null hypothesis, X is distributed as a binomial random variable with a proportion of 0.5 and sample size of m .

The significance of X is calculated using binomial probabilities.

Bootstrap Test

The one-sample bootstrap procedure for testing whether the mean is equal to a specific value is given in Efron & Tibshirani (1993), pages 224-227. The bootstrap procedure is as follows.

1. Compute the mean of the sample. Call it \bar{X} .
2. Compute the t-value using the standard t-test. The formula for this computation is

$$t_x = \frac{\bar{X} - D0}{s_{\bar{X}}/\sqrt{n}}$$

where $D0$ is the hypothesized difference.

3. Draw a random, with-replacement sample of size n from the original X values. Call this sample Y_1, Y_2, \dots, Y_n .
4. Compute the t-value of this bootstrap sample using the formula

$$t_y = \frac{\bar{Y} - \bar{X}}{s_{\bar{Y}}/\sqrt{n}}$$

5. For a two-tailed test, if $|t_y| > |t_x|$ then add one to a counter variable, A .
6. Repeat steps 3 – 5 B times. B may be anywhere from 100 to 10,000.
7. Compute the p -value of the bootstrap test as $(A + 1) / (B + 1)$
8. Steps 1 – 7 complete one simulation iteration. Repeat these steps M times, where M is the number of simulations. The power and significance level are equal to the percent of the time the p -value is less than the nominal alpha of the test in their respective simulations.

Note that the bootstrap test is a time-consuming test to analyze, especially if you set B to a value much larger than 100.

Standard Deviations

Care must be used when either the null or alternative distribution is not normal. In these cases, the standard deviation is usually not specified directly. For example, you might use a gamma distribution with a shape parameter of 1.5 and a mean of 4 as the null distribution and a gamma distribution with the same shape parameter and a mean of 5 as the alternative distribution. This allows you to compare the two means. However, note that although the shape parameters are constant, the standard deviations are not. In cases such as this, the null and alternatives not only have different means, but different standard deviations!

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 most of the parameters and options that you will be concerned with.

Solve For

Solve For

This option specifies the parameter to be calculated using the values of the other parameters. Under most conditions, you would select either *Power* or *Sample Size*.

Select *Power* when you want to estimate the power for a specific scenario.

Select *Sample Size* when you want to determine the sample size needed to achieve a given power and alpha error level. This option can be very computationally intensive and may take considerable time to complete.

Test and Simulations

Alternative Hypothesis

Specify the alternative hypothesis of the test. Since the null hypothesis is the opposite of the alternative, specifying the alternative is all that is needed. Usually, the two-tailed (\neq) option is selected.

The options containing only $<$ or $>$ are one-tailed tests. When you choose one of these options, you must be sure that the input parameters match this selection.

Possible selections are:

- **Two-Sided (H1: $\delta \neq \delta_0$)**
This is the most common selection. It yields the *two-tailed t-test*. Use this option when you are testing whether the means are different, but you do not want to specify beforehand which mean is larger. Many scientific journals require two-tailed tests.
- **One-Sided (H1: $\delta < \delta_0$)**
This option yields a *one-tailed t-test*. Use it when you are only interested in the case in which δ is less than δ_0 .
- **One-Sided (H1: $\delta > \delta_0$)**
This option yields a *one-tailed t-test*. Use it when you are only interested in the case in which δ is greater than δ_0 .

Tests for Paired Means (Simulation)

Test Type

Specify which test statistic (t-test, Wilcoxon signed-rank test, sign test, or bootstrap test) is to be simulated. Although the t-test is the most commonly used test statistic, it is based on assumptions that may not be viable in many situations. For your data, you may find that one of the other tests is more accurate (actual alpha = target alpha) and more precise (higher power).

Note that the bootstrap test is computationally intensive, so it can be very slow to evaluate.

Bootstrap Iterations

Displayed when Test Type = Bootstrap Test

Specify the number of iterations used in the bootstrap hypothesis test. This value is only used if the bootstrap test is displayed on the reports. The running time of the procedure depends heavily on the number of iterations specified here.

Recommendations by authors of books discussing the bootstrap range from 100 to 10,000. If you enter a large (greater than 500) value, the procedure may take several hours to run.

Simulations

This option specifies the number of iterations, M , used in the simulation. Larger numbers of iterations result in longer running time and more accurate results.

The precision of the simulated power estimates can be determined by recognizing that they follow the binomial distribution. Thus, confidence intervals may be constructed for power estimates. The following table gives an estimate of the precision that is achieved for various simulation sizes when the power is either 0.50 or 0.95. The table values are interpreted as follows: a 95% confidence interval of the true power is given by the power reported by the simulation plus and minus the 'Precision' amount given in the table.

Simulation Size M	Precision when Power = 0.50	Precision when Power = 0.95
100	0.100	0.044
500	0.045	0.019
1000	0.032	0.014
2000	0.022	0.010
5000	0.014	0.006
10000	0.010	0.004
50000	0.004	0.002
100000	0.003	0.001

Notice that a simulation size of 1000 gives a precision of plus or minus 0.014 when the true power is 0.95. Also note that as the simulation size is increased beyond 5000, there is only a small amount of additional precision achieved.

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.

Tests for Paired Means (Simulation)

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. This value must be an integer greater than one. Note that you may enter a list of values using the syntax *50 100 150 200 250* or *50 to 250 by 50*.

Effect Size

Input Type

Select which type of input method to use to specify the simulated distributions and the effect size.

The choices are

- **Simple**
Provides the simplest possible input for specifying the distributions and effect size. This entry type is restricted to a set of distributions that can be specified using a mean and standard deviation. All simulated values come from one named distribution.
- **General**
Provides all options for specifying the distributions and parameters. This allows for complete flexibility in specifying the simulated distributions but is more complicated.

Effect Size – Simulated Distribution

Displayed when Input Type = Simple

Distribution to Simulate

Specify the distribution that represents the type of data you expect from your study.

The possible distributions are Beta, Exponential, Gamma, Gumbel, Laplace, Logistic, Lognormal, Normal, Poisson, TukeyGH, Uniform, Weibull.

All these distributions can be specified with a mean and standard deviation. The Beta and TukeyGH distributions each require two additional parameters. The Exponential and Poisson distributions require only the mean to be specified since the standard deviation can be computed from the mean.

Normal vs. TukeyGH

Tukey's distribution can be used to generate values that are nearly Normal, with departure from normality controlled by entering skewness (G) and kurtosis (H) parameters. Tukey's distribution with $G = H = 0$ is equivalent to the Normal distribution.

Tests for Paired Means (Simulation)

Effect Size – Means

Displayed when Input Type = Simple

δ_0 (Null Mean of Paired Differences)

Enter a value (or range of values) for the mean of paired differences under the null hypothesis.

δ_1 (Actual Mean of Paired Differences)

Enter a value (or range of values) for the mean of paired differences at which power and sample size are calculated. This value indicates the minimum detectible paired difference.

Effect Size – Standard Deviation

Displayed when Input Type = Simple

σ (Std Dev of Paired Differences)

Enter an estimate of the standard deviation of the paired differences (must be positive). Use results from a previous (or pilot) study or the range divided by 5.

When this value is not known, you must supply an estimate of it. **PASS** includes a special tool for estimating the standard deviation. This tool may be loaded by pressing the *SD* button. Refer to the Standard Deviation Estimator chapter for further details.

Effect Size – Simulated Distributions

Displayed when Input Type = General

Distribution|H0 (Null Hypothesis)

Specify the distribution under the null hypothesis, H0.

The parameters of the distribution can be specified using numbers or letters. If letters are used, their values are specified in the boxes below. The value “D0” is reserved for the value of the mean under the null hypothesis.

Following is a list of the distributions that are available, and the syntax used to specify them. Each of the parameters should be replaced with a number or parameter name.

Distributions with Common Parameters

Beta(Shape1, Shape2, Min, Max)

Binomial(P, N)

Cauchy(Mean, Scale)

Constant(Value)

Exponential(Mean)

Gamma(Shape, Scale)

Gumbel(Location, Scale)

Laplace(Location, Scale)

Logistic(Location, Scale)

Lognormal(Mu, Sigma)

Multinomial(P1, P2, P3, ..., Pk)

Normal(Mean, Sigma)

Poisson(Mean)

TukeyGH(Mu, S, G, H)

Tests for Paired Means (Simulation)

Uniform(Min, Max)

Weibull(Shape, Scale)

Distributions with Mean and SD Parameters

BetaMS(Mean, SD, Min, Max)

BinomialMS(Mean, N)

GammaMS(Mean, SD)

GumbelMS(Mean, SD)

LaplaceMS(Mean, SD)

LogisticMS(Mean, SD)

LognormalMS(Mean, SD)

UniformMS(Mean, SD)

WeibullMS(Mean, SD)

Details of writing mixture distributions, combined distributions, and compound distributions are found in the chapter on *Data Simulation* and will not be repeated here.

Finding the Value of the Mean of a Specified Distribution

The mean of a distribution created as a linear combination of other distributions is found by applying the linear combination to the individual means. However, the mean of a distribution created by multiplying or dividing other distributions is not necessarily equal to applying the same function to the individual means. For example, the mean of $4 \text{ Normal}(4, 5) + 2 \text{ Normal}(5, 6)$ is $4*4 + 2*5 = 26$, but the mean of $4 \text{ Normal}(4, 5) * 2 \text{ Normal}(5, 6)$ is not exactly $4*4*2*5 = 160$ (although it is close).

Distribution|H1

This option specifies the mean and distribution under the alternative hypothesis, H1. That is, this is the actual (true) value of the mean at which the power is computed. Usually, the mean is specified by entering 'D1' for the mean parameter in the distribution expression and then entering values for the D1 parameter.

The parameters of the distribution can be specified using numbers or letters. If letters are used, their values are specified in the boxes below. The value "D1" is reserved for the value of the mean under the alternative hypothesis.

A list of the distributions that are available and the syntax used to specify them is given above.

Finding the Value of the Mean under H1

The distributions have been parameterized in terms of their means, since this is the parameter being tested. The mean of a distribution created as a linear combination of other distributions is found by applying the linear combination to the individual means. However, the mean of a distribution created by multiplying or dividing other distributions is not necessarily equal to applying the same function to the individual means. For example, the mean of $4 \text{ Normal}(4, 5) + 2 \text{ Normal}(5, 6)$ is $4*4 + 2*5 = 26$, but the mean of $4 \text{ Normal}(4, 5) * 2 \text{ Normal}(5, 6)$ is not exactly $4*4*2*5 = 160$ (although it is close).

Tests for Paired Means (Simulation)

Effect Size – Distribution Parameters

Displayed when Input Type = General

D0 (Mean Diff|H0)

These values are substituted for the D0 in the distribution specifications given above. D0 is intended to be the value of the mean hypothesized by the null hypothesis, H0.

You can enter a list of values using syntax such as *0 1 2 3* or *0 to 3 by 1*.

Note that whether D0 is the mean of the simulated distribution depends on the formula you have entered. For example, $\text{Normal}(D0, S)$ has a mean of D0, but $\text{Normal}(D0, S) - \text{Normal}(D0, S)$ has a mean of zero.

D1 (Mean Diff|H1)

These values are substituted for the D1 in the distribution specifications given above. D1 is intended to be the value of the mean hypothesized by the alternative hypothesis, H1.

You can enter a list of values using syntax such as *0 1 2 3* or *0 to 3 by 1*.

Note that whether D1 is the mean of the simulated distribution depends on the formula you have entered. For example, $\text{Normal}(D1, S)$ has a mean of D1, but $\text{Normal}(D1, S) - \text{Normal}(D0, S)$ has a mean of D1-D0.

Parameter Values (S, A, B, C)

Enter the numeric value(s) of parameter listed above. These values are substituted for the corresponding letter in the distribution specifications for H0 and H1.

You can enter a list of values using syntax such as *0 1 2 3* or *0 to 3 by 1*.

You can also change the letter that is used as the name of this parameter.

Tests for Paired Means (Simulation)

Example 1 – Power at Various Sample Sizes

Researchers are planning a pre-post experiment to test whether the difference in response to a certain drug is different from zero. The researchers will use a paired t-test with an alpha level of 0.05. They want to compare the power at sample sizes of 50, 100, and 150 when the shift in the means is 0.6 from pre-test to post-test. They assume that the data are normally distributed with a standard deviation of paired differences of 2.53. Since this is an exploratory analysis, they set the number of simulation iterations to 2000.

A researcher is planning an experiment to test whether the mean response level to a certain drug is significantly different from zero. The researcher wants to use a t-test with an alpha level of 0.05. He wants to compute the power at various sample sizes from 5 to 40, assuming the true mean is one. He assumes that the data are normally distributed with a standard deviation of 2.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Paired Means (Simulation)** procedure window. You may then make the appropriate entries as listed below, or open **Example 1a** by going to the **File** menu and choosing **Open Example Template**. To analyze the same problem using the General input type, load the Example 1b settings template.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Power
Alternative Hypothesis	Two-Sided (H1: $\delta \neq \delta_0$)
Test Type	T-Test
Simulations	2000
Alpha	0.05
N (Sample Size)	50 100 150
Input Type	Simple
Distribution to Simulate	Normal
δ_0 (Null Mean of Paired Diffs)	0
δ_1 (Actual Mean of Paired Diffs)	0.6
σ (Std Dev of Paired Diffs)	2.53

Annotated Output

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

Numeric Results

Numeric Results									
Hypotheses: H0: $\delta = \delta_0$ vs. H1: $\delta \neq \delta_0$									
Test Statistic: Paired T-Test									
Simulated Distribution: Normal									
Power	N	δ_0	δ_1	Diff $\delta_1 - \delta_0$	σ	Effect Size	Target Alpha	Actual Alpha	
0.3895 (0.0214) [0.3681 0.4109]	50	0.0	0.6	0.6	2.5	0.237	0.050	0.050 (0.010) [0.040 0.060]	
0.6565 (0.0208) [0.6357 0.6773]	100	0.0	0.6	0.6	2.5	0.237	0.050	0.052 (0.010) [0.042 0.062]	
0.8270 (0.0166) [0.8104 0.8436]	150	0.0	0.6	0.6	2.5	0.237	0.050	0.052 (0.010) [0.042 0.062]	

Tests for Paired Means (Simulation)

Notes:

Simulations: 2000. Run Time: 2.42 seconds.

Report Definitions

Power is the probability of rejecting a false null hypothesis. Second Row: (Power Inc.) [95% LCL and UCL Power].

N is the size of the sample drawn from the population.

δ_0 is the value of the population mean of paired differences under the null hypothesis.

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

$\delta_1 - \delta_0$ is the difference between the actual and null mean of paired differences.

σ is the standard deviation of paired differences for the population. It measures the variability in the population.

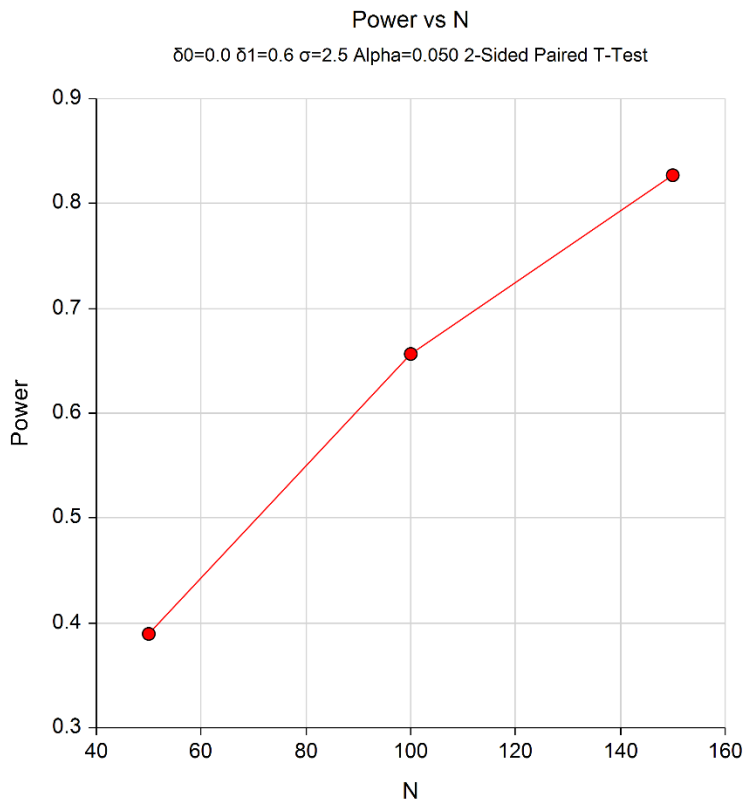
Effect Size = $|\delta_1 - \delta_0|/\sigma$ is the relative magnitude of the effect.

Target Alpha is the probability of rejecting a true null hypothesis. It is set by the user.

Actual Alpha is the alpha level that was actually achieved by the experiment. Second Row: (Alpha Inc.) [95% LCL and UCL Alpha].

This report shows the estimated power for each scenario. The first row shows the parameter settings and the estimated power and significance level (Actual Alpha). Note that because these are results of a simulation study, the computed power and alpha will vary from run to run. Thus, another report obtained using the same input parameters will be slightly different than the one above.

The second row shows two 95% confidence intervals in brackets: the first for the power and the second for the significance level. Half the width of each confidence interval is given in parentheses as a fundamental measure of the accuracy of the simulation. As the number of simulations is increased, the width of the confidence interval will decrease.

Plots Section

This plot shows the relationship between sample size and power.

Tests for Paired Means (Simulation)

Example 2 – Finding the Sample Size

Continuing with Example 1, the researchers want to determine how large a sample is needed to obtain a power of 0.90?

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Paired Means (Simulation)** procedure window. 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**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Sample Size
Alternative Hypothesis	Two-Sided (H1: $\delta \neq \delta_0$)
Test Type	T-Test
Simulations	2000
Power	0.90
Alpha	0.05
Input Type	Simple
Distribution to Simulate	Normal
δ_0 (Null Mean of Paired Diffs)	0
δ_1 (Actual Mean of Paired Diffs)	0.6
σ (Std Dev of Paired Diffs)	2.53

Output

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

Numeric Results of Search for N

Numeric Results								
Hypotheses: H0: $\delta = \delta_0$ vs. H1: $\delta \neq \delta_0$								
Test Statistic: Paired T-Test								
Simulated Distribution: Normal								
Power	N	δ_0	δ_1	Diff $\delta_1 - \delta_0$	σ	Effect Size	Target Alpha	Actual Alpha
0.8855 (0.0140)	188 [0.8715 0.8995]	0.0	0.6	0.6	2.5	0.237	0.050	0.049 (0.009) [0.040 0.058]

The required sample size of 188 achieved a power of 0.8855. The power of 0.8855 is less than the target value of 0.900 because the sample size search algorithm re-simulates the power for the final sample size. Thus, it is possible for the search algorithm to converge to a sample size which exhibits the desired power, but then on a succeeding simulation to achieve a power that is slightly less than the target. To achieve more accuracy, a reasonable strategy would be to run simulations to obtain the powers using N's from 180 to 200 using a simulation size of 5000 or greater.

Tests for Paired Means (Simulation)

Example 3 – Comparative Results

Continuing with Example 2, the researchers want to study the characteristics of alternative test statistics.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Paired Means (Simulation)** procedure window. 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**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Power
Alternative Hypothesis	Two-Sided (H1: $\delta \neq \delta_0$)
Test Type	T-Test
Simulations	2000
Alpha	0.05
N (Sample Size)	50 100 150 200
Input Type	Simple
Distribution to Simulate	Normal
δ_0 (Null Mean of Paired Diffs)	0
δ_1 (Actual Mean of Paired Diffs)	0.6
σ (Std Dev of Paired Diffs)	2.53
Reports Tab	
Show Comparative Reports	Checked
Include T-Test Results	Checked
Include Wilcoxon & Sign Test	Checked
Comparative Plots Tab	
Show Comparative Plots	Checked

Output

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

Numeric Results

Power Comparison

Hypotheses: $H_0: \delta = \delta_0$ vs. $H_1: \delta \neq \delta_0$

Simulated Distribution: Normal

N	δ_0	δ_1	σ	Target Alpha	T-Test Power	Wilcxn Power	Sign Power
50	0.0	0.6	2.5	0.050	0.3685	0.3600	0.2120
100	0.0	0.6	2.5	0.050	0.6535	0.6260	0.4150
150	0.0	0.6	2.5	0.050	0.8405	0.8130	0.6295
200	0.0	0.6	2.5	0.050	0.9140	0.8980	0.7290

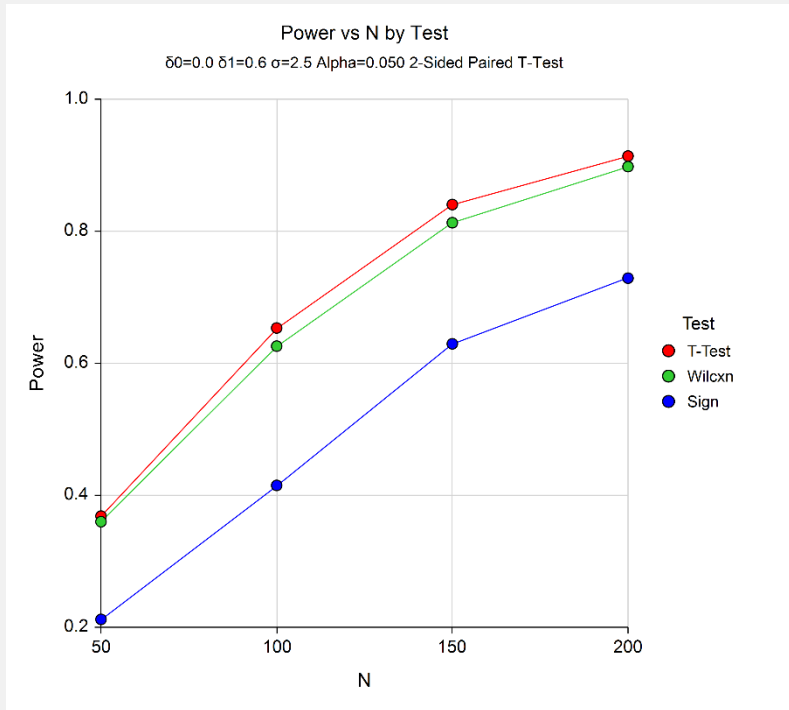
Tests for Paired Means (Simulation)

Alpha Comparison

Hypotheses: $H_0: \delta = \delta_0$ vs. $H_1: \delta \neq \delta_0$
 Simulated Distribution: Normal

N	δ_0	δ_1	σ	Target Alpha	T-Test Alpha	Wilcxn Alpha	Sign Alpha
50	0.0	0.6	2.5	0.050	0.064	0.062	0.046
100	0.0	0.6	2.5	0.050	0.053	0.048	0.036
150	0.0	0.6	2.5	0.050	0.041	0.037	0.036
200	0.0	0.6	2.5	0.050	0.056	0.052	0.035

Chart Section



These results show that for paired data, the t-test and Wilcoxon test have very similar power and alpha values. The sign test is both less accurate and less powerful.

Tests for Paired Means (Simulation)

Example 4 – Validation using Zar (1984)

Zar (1984), pages 111-112, presents an example in which $\delta_0 = 0.0$, $\delta_1 = 1.0$, $\sigma = 1.25$, $\alpha = 0.05$, and $N = 12$. Zar obtains an approximate power of 0.72. We will validate this procedure by running this example. To make certain that the results are very accurate, the number of simulations will be set to 10,000.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Paired Means (Simulation)** procedure window. You may then make the appropriate entries as listed below, or open **Example 4** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Power
Alternative Hypothesis	Two-Sided (H1: $\delta \neq \delta_0$)
Test Type	T-Test
Simulations	10000
Alpha	0.05
N (Sample Size)	12
Input Type	Simple
Distribution to Simulate	Normal
δ_0 (Null Mean of Paired Diffs)	0
δ_1 (Actual Mean of Paired Diffs)	1
σ (Std Dev of Paired Diffs)	1.25

Output

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

Numeric Results

Numeric Results									
Hypotheses: H0: $\delta = \delta_0$ vs. H1: $\delta \neq \delta_0$									
Test Statistic: Paired T-Test									
Simulated Distribution: Normal									
Power	N	δ_0	δ_1	Diff $\delta_1 - \delta_0$	σ	Effect Size	Target Alpha	Actual Alpha	
0.7189	12	0.0	1.0	1.0	1.3	0.800	0.050	0.052	
(0.0088)	[0.7101	0.7277]						(0.004)	[0.048 0.057]

This simulation obtained a power of 0.7189 which rounds to the 0.72 computed by Zar. Note that another repetition of this same analysis will probably be slightly different since a different set of random numbers will be used.

Tests for Paired Means (Simulation)

Example 5 – Validation using Machin (1997)

Machin, et. al. (1997), page 37, presents an example in which $\mu_0 = 0.0$, $\mu_1 = 0.2$, $\sigma = 1.0$, $\alpha = 0.05$, and power = 0.80. They obtain a sample size of 199.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Paired Means (Simulation)** procedure window. You may then make the appropriate entries as listed below, or open **Example 5** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Sample Size
Alternative Hypothesis	Two-Sided (H1: $\mu \neq \mu_0$)
Test Type	T-Test
Simulations	2000
Power	0.80
Alpha	0.05
Input Type	Simple
Distribution to Simulate	Normal
μ_0 (Null or Baseline Mean)	0
μ_1 (Actual Mean)	0.2
σ (Standard Deviation)	1

Output

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

Numeric Results

Numeric Results									
Hypotheses: $H_0: \delta = \delta_0$ vs. $H_1: \delta \neq \delta_0$									
Test Statistic: Paired T-Test									
Simulated Distribution: Normal									
Power	N	δ_0	δ_1	Diff $\delta_1 - \delta_0$	σ	Effect Size	Target Alpha	Actual Alpha	
0.8125 (0.0171)	199 [0.7954 0.8296]	0.0	0.2	0.2	1.0	0.200	0.050	0.063 (0.011)	[0.052 0.073]

The sample size result matches the value of Machin (1997). If you run the simulation multiple times, you'll come up with values right around 199.

Tests for Paired Means (Simulation)

Example 6 – Non-Inferiority Test

A non-inferiority test is appropriate when you want to show that a new treatment is no worse than the standard. For example, suppose that a standard diagnostic test has an average score of 70. Unfortunately, this diagnostic test is expensive. A promising new diagnostic test must be compared to the standard. Researchers want to show that it is no worse than the standard.

Because of many benefits from the new test, clinicians are willing to adopt it even if it is slightly less accurate than the current test. How much less can the score of the new treatment be and still be adopted? Should it be adopted if the difference is -1? -2? -5? -10? There is an amount below 0 at which the difference between the two treatments is no longer considered ignorable. After thoughtful discussion with several clinicians, the *margin of non-inferiority* is set to -5.

The developers decided to use a paired t-test. They must design an experiment to test the hypothesis that the average difference between the two tests is greater than -5. The statistical hypothesis to be tested is

$$H_0: A - B \leq -5 \text{ versus } H_1: A - B > -5$$

where A represents the mean of the new test and B represents the mean of the standard test. Notice that when the null hypothesis is rejected, the conclusion is that the average difference is greater than -5.

Experience has shown that the standard deviation of paired differences is 6.32. Following proper procedure, the researchers decide to use a significance level of 0.025 for this one-sided test to keep it comparable to the usual value of 0.05 for a two-sided test. They decide to look at the power for sample sizes of 5, 10, 15, 20, and 25 subjects. They decide to compute the power for the case when the two tests are equal.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Paired Means (Simulation)** procedure window. You may then make the appropriate entries as listed below, or open **Example 6** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Power
Alternative Hypothesis	One-Sided (H1: $\delta > \delta_0$)
Test Type	T-Test
Simulations	2000
Alpha	0.025
N (Sample Size).....	5 10 15 20 25
Input Type	Simple
Distribution to Simulate.....	Normal
δ_0 (Null Mean of Paired Diffs)	-5
δ_1 (Actual Mean of Paired Diffs).....	0
σ (Std Dev of Paired Diffs).....	6.32

Tests for Paired Means (Simulation)

Output

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

Numeric Results

Numeric Results

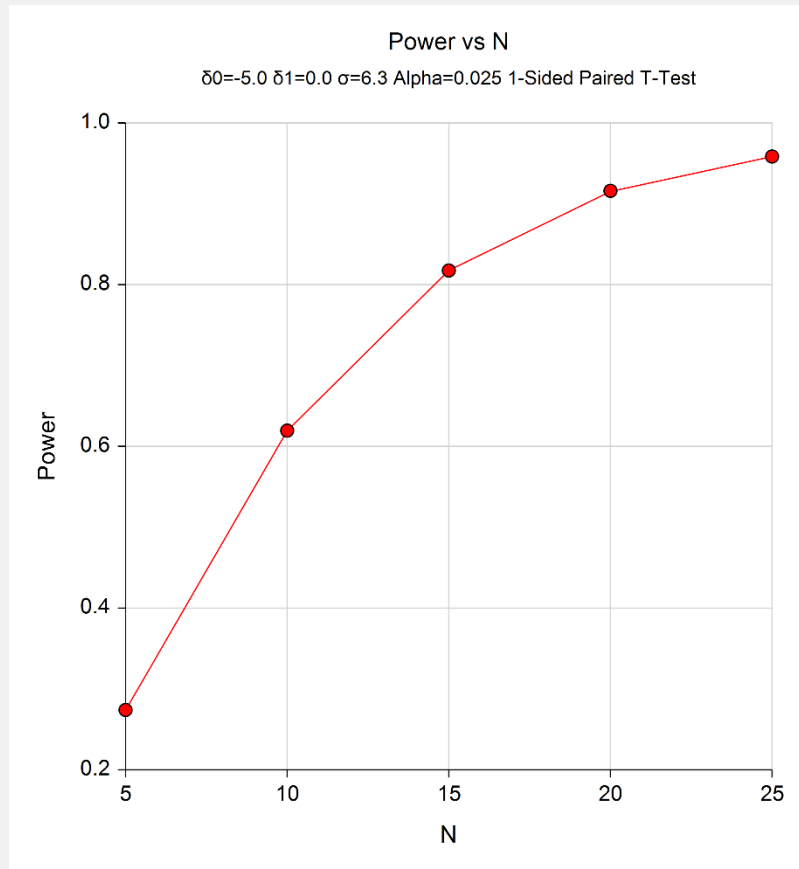
Hypotheses: $H_0: \delta \leq \delta_0$ vs. $H_1: \delta > \delta_0$

Test Statistic: Paired T-Test

Simulated Distribution: Normal

Power	N	δ_0	δ_1	Diff $\delta_1 - \delta_0$	σ	Effect Size	Target Alpha	Actual Alpha
0.2740 (0.0195) [0.2545 0.2935]	5	-5.0	0.0	5.0	6.3	0.791	0.025	0.025 (0.007) [0.018 0.031]
0.6195 (0.0213) [0.5982 0.6408]	10	-5.0	0.0	5.0	6.3	0.791	0.025	0.024 (0.007) [0.017 0.031]
0.8175 (0.0169) [0.8006 0.8344]	15	-5.0	0.0	5.0	6.3	0.791	0.025	0.022 (0.006) [0.016 0.028]
0.9160 (0.0122) [0.9038 0.9282]	20	-5.0	0.0	5.0	6.3	0.791	0.025	0.029 (0.007) [0.022 0.036]
0.9585 (0.0087) [0.9498 0.9672]	25	-5.0	0.0	5.0	6.3	0.791	0.025	0.025 (0.007) [0.018 0.031]

Chart Section



We see that a power of 0.8 is achieved at about 15 subjects, while a power of 0.9 requires about 20 subjects.