Chapter 490

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 us 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. But, in recent years, as computer speeds have increased, simulation studies can be completed on desktop and laptop computers in a reasonable period of time.

The steps to a simulation study are

1. Specify the test procedure and the test statistic. This includes the significance level, sample size, and underlying data distributions.

2. Generate a random sample $X_1, X_2, \ldots, X_n$ 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. These samples are used to calculate the power of the test. In the case of paired data, the individual values are simulated as the difference between two other random variables. These samples are constructed so that they exhibit a certain amount of correlation.

3. Generate a random sample $Y_1, Y_2, \ldots, Y_n$ 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. These samples are used to calculate the significance-level of the test. In the case of paired data, the individual values are simulated as the difference between two other random variables. These samples are constructed so that they exhibit a certain amount of correlation.

4. Repeat steps 2 and 3 several thousand times, tabulating the number of times the simulated data leads to a rejection of the null hypothesis. The power is the proportion of simulated samples in step 2 that lead to rejection. The significance level is the proportion of simulated samples in step 3 that lead to rejection.

Simulating Paired Distributions

Paired data occur when two observations are correlated. Examples of paired designs are pre – post designs, cross-over designs, and matched pair designs.

In order to simulate paired data, the simulation should mimic the actual data generation process as closely as possible. Since paired data are analyzed by creating the individual difference between each pair, the simulation should also create data as the difference between two variates. Paired data exhibit a correlation between the two variates. As this correlation between the variates increases, the variance of the difference decreases. Thus it is important not only to specify the distributions of the two variates that will be differenced, but to also specify their correlation.

Obtaining paired samples from arbitrary distributions with a set correlation is difficult because the joint, bivariate distribution must be specified and simulated. Rather than specify the bivariate distribution, *PASS* requires the specification of the two marginal distributions and the correlation between them.

Monte Carlo samples with given marginal distributions and correlation are generated using the method suggested by Gentle (1998). The method begins by generating a large population of random numbers from the two distributions. Each of these populations is evaluated to determine if their means are within a small relative tolerance (0.0001) of the target mean. If the actual mean is not within the tolerance of the target mean, individual members of the population are replaced with new random numbers if the new random number moves the mean towards its target. Only a few hundred such swaps are required to bring the actual mean to within tolerance of the target mean.

The next step is to obtain the target correlation. This is accomplished by permuting one of the populations until they have the desired correlation.

The above steps provide a large pool of random numbers that exhibit the desired characteristics. This pool is then sampled at random using the uniform distribution to obtain the random numbers used in the simulation.

This algorithm may be stated as follows.

1. Draw individual samples of size M from the two distributions where M is a large number, usually over 10,000. Adjust these samples so that they have the specified mean and standard deviation. Label these samples A and B. Create an index of the values of A and B according to the order in which they are generated. Thus, the first value of A and the first value of B are indexed as one, the second values of A and B are indexed as two, and so on up to the final set which is indexed as M.

2. Compute the correlation between the two generated variates.

3. If the computed correlation is within a small tolerance (usually less than 0.001) of the specified correlation, go to step 7.

4. Select two indices (I and J) at random using uniform random numbers.

5. Determine what will happen to the correlation if B_I is swapped with B_J. If the swap will result in a correlation that is closer to the target value, swap the indices and proceed to step 6. Otherwise, go to step 4.

6. If the computed correlation is within the desired tolerance of the target correlation, go to step 7. Otherwise, go to step 4.

7. End with a population with the required marginal distributions and correlation.

Now, to complete the simulation, random samples of the designated size are drawn from this population.

Test Statistics

This section describes the test statistics that are available in this procedure. Note that these test statistics are computed on the differences. Thus, when the equation refers to an X value, this X value is assumed to be a difference between two individual variates.

One-Sample t-Test

The one-sample t-test assumes that the data are a simple random sample from a population of normally-distributed 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 follow

$$t_{n-1} = \frac{\overline{X} - M0}{s_{\overline{X}}}$$

where

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

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

and *M*0 is the value of the <u>difference</u> hypothesized by the null hypothesis.

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 mean, *D*0, 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*, is the minimum of *Sp* and *Sn*.

3. Compute the mean and standard deviation of *W* using the formulas

$$\mu_{W_n} = \frac{n(n+1)}{4} \text{ and } \sigma_{W_n} = \sqrt{\frac{n(n+1)(2n+1)}{24} - \frac{\sum t^3 - \sum t}{48}}$$

    where t_i represents the number of times the ith value occurs.

4. Compute the *z* value using

$$z_W = \frac{W - \mu_{W_n}}{\sigma_{W_n}}$$

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 computing the p-value using the standard normal distribution. If this p-value is less than a specified level (usually

0.05), the null hypothesis is rejected in favor of the alternative hypothesis. Otherwise, no conclusion can be reached.

Sign Test

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

1. Count the number of values strictly greater than $M0$. Call this value X.

2. Count the number of values strictly less than $M0$. 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 $\overline{X}$.

2. Compute the t-value using the standard t-test. The formula for this computation is

$$t_X = \frac{\overline{X} - M0}{s_{\overline{X}}}$$

3. Draw a random, with-replacement sample of size n from the original X values. Call this sample $Y_1, Y_2, \cdots, Y_n$.

4. Compute the t-value of this bootstrap sample using the formula

$$t_Y = \frac{\overline{Y} - \overline{X}}{s_{\overline{Y}}}$$

5. For a two-tailed test, if $\left| t_Y \right| > \left| t_x \right|$ 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 is equal to the percent of the time the p-value is less than the nominal alpha of the test.

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

The Problem of Differing 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. Thus 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 and Options tabs. 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 will be of interest.

Solve For

Solve For

This option specifies whether you want to find *Power* or *Sample Size* from the simulation. Select *Power* when you want to estimate the power of a certain scenario. Select *Sample Size* when you want to determine the sample size needed to achieve a given power and alpha error level. Finding *Sample Size* is very computationally intensive, and so it may take a long time to complete.

Test and Simulations

Alternative Hypothesis

This option specifies the alternative hypothesis, H1. This implicitly specifies the direction of the hypothesis test. The null hypothesis is always H0: Diff = Diff0.

Note that the alternative hypothesis enters into power calculations by specifying the rejection region of the hypothesis test. Its accuracy is critical.

Possible selections are:

- **Difference ≠ Diff0**

  This is the most common selection. It yields a *two-tailed test*. Use this option when you are testing whether the mean is different from a specified value Diff0, but you do not want to specify beforehand whether it is smaller or larger. Most scientific journals require two-tailed tests.

- **Difference < Diff0**

  This option yields a *one-tailed test*. Use it when you want to test whether the true mean is less than Diff0.

- **Difference > Diff0**

  This option yields a *one-tailed test*. Use it when you want to test whether the true mean is greater than Diff0. Note that this option could be used for a **non-inferiority test**.

Test Type

Specify which test statistic (t-test, Wilcoxon 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 (better power).

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

Simulations

This option specifies the number of iterations, M, used in the simulation. As the number of iterations is increased, the accuracy and running time of the simulation will be increased also.

The precision of the simulated power estimates are calculated from the binomial distribution. Thus, confidence intervals may be constructed for various power values. 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 0.01 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 accuracy 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. In this procedure, a type-II error occurs when you fail to reject the null hypothesis of equal means when in fact the means are different.

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 subjects in the study. The paired design assumes that a pair of observations will be obtained from each subject. Thus there will be 2N observations simulated, resulting in N differences.

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

Item A & B Distributions|H0 (Null)

These options specify the distributions of the two items making up the pair under the null hypothesis, H0. The difference between the means of these two distributions is the difference that is tested, Diff0.

Usually, you will want Diff0 = 0. This zero difference is specified by entering *M0* for the mean parameter in each of the distributions and then entering an appropriate value for the M0 parameter below.

Several of the distributions are parameterized so that the mean is entered first. For example, if you wanted to test whether the mean of a normal distributed variable is five, you could enter Normal(5, S) or Normal(M0, S) here.

The parameters of each distribution are specified using numbers or letters. If letters are used, their values are specified in the boxes below. The value *M0* 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)

Uniform(Min, Max)

Weibull(Shape, Scale)

Tests for Paired Means (Simulation)

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

Except for the multinomial distribution, 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 Normal(4, 5) + 2 Normal(5, 6) is 4*4 + 2*5 = 26, but the mean of 4 Normal(4, 5) * 2 Normal(5, 6) is not exactly 4*4*2*5 = 160 (although it is close).

Item A & B Distributions|H1 (Alternative)

These options specify the distributions of the two items making up the pair under the alternative hypothesis, H1. The difference between the means of these two distributions is the difference that is assumed to be the true value of the difference. That is, this is the difference at which the power is computed.

Usually, the mean difference is specified by entering *M1* for the mean parameter in the distribution expression for item A and *M0* for the mean parameter in the distribution expression for item B. The mean difference under H1 then becomes the value of M1 – M0.

The parameters of each distribution are specified using numbers or letters. If letters are used, their values are specified in the boxes below. The value *M1* 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.

Effect Size – Distribution Parameters

M0 (Mean|H0)

These values are substituted for the *M0* in the four distribution specifications given above. *M0* is intended to be the value of the mean hypothesized by the null hypothesis, H0.

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

M1 (Mean|H1)

These values are substituted for the *M1* in the four distribution specifications given above. *M1* is intended to be the value of the mean hypothesized by the alternative hypothesis, H1.

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

Parameter Values (S, A, B)

Enter the numeric value(s) of the parameters listed above. These values are substituted for the corresponding letter in all four distribution specifications.

You can enter a list of values using the syntax *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 using the pull-down menu to the side.

Effect Size – Distribution Parameters

R (Correlation of Items A & B)

Specify the value of the correlation between items (variates) A and B of the pair.

Since this is a correlation, it must be between -1 and 1. However, some distributions (such as the multinomial distribution) have a maximum possible correlation that is far less than one.

Typical values are between 0 and 0.4.

Options Tab

The Options tab contains some additional computational options.

Bootstrap Iterations

Bootstrap Iterations

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 are from 100 to 10,000. If you enter a large (greater than 500) value, the simulation may take several hours to run.

Random Numbers

Random Number Pool Size

This is the size of the pool of random values from which the random samples will be drawn. Populations of at least 10,000 should be used. Also, the value should be about twice the number of simulations. You can enter *Automatic* and an appropriate value will be calculated.

Note that values over 50,000 may take a long time to permute to achieve the target means and correlation.

Correlation

Maximum Switches

This option specifies the maximum number of index switches that can be made while searching for a permutation of item B that yields a correlation within the specified range. A value near 5,000,000 may be necessary when the correlation is near one.

Correlation Tolerance

Specify the amount above and below the target correlation that will still let a particular index-permutation to be selected for the population. For example, if you have selected a correlation of 0.3 and you set this tolerance to 0.001, then only populations with a correlation between 0.299 and 0.301 will be used. The recommended is 0.001 or smaller. Valid values are between 0 and 0.999.

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 2 and that the correlation between the pre-test and post-test values is 0.20. Since this is an exploratory analysis, they set the number of simulation iterations to 2000.

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 by expanding **Means**, then expanding **Paired Means**, then clicking on **Tests (Inequality)**, and then clicking on **Tests for Paired Means (Simulation)**. 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**
Alternative Hypothesis	**Diff ≠ Diff0**
Test Type	**T-Test**
Simulations	**2000**
Alpha..	**0.05**
N (Sample Size)......................................	**50 100 150**
Item A Distribution|H0	**Normal(M0 S)**
Item B Distribution|H0	**Normal(M0 S)**
Item A Distribution|H1	**Normal(M0 S)**
Item B Distribution|H1	**Normal(M1 S)**
M0 (Mean|H0)	**0**
M1 (Mean|H1)	**0.6**
S...	**2**
R (Correlation of Items A & B)	**0.2**

Annotated Output

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

Numeric Results

Numeric Results for Testing Mean Difference = Diff0.   **Hypotheses: H0: Diff1 = Diff0; H1: Diff1 ≠ Diff0**
H0 Dist'n: Normal(M0 S) - Normal(M0 S)
H1 Dist'n: Normal(M0 S) - Normal(M1 S)
Test Statistic: Paired T-Test

Power	N	H0 Diff0	H1 Diff1	Corr R	Target Alpha	Actual Alpha	Beta	M0	M1	S
0.388	50	0.0	-0.6	0.200	0.050	0.040	0.612	0.0	0.6	2.0
(0.021)	[0.367	0.409]				(0.009)	[0.031	0.048]		
0.642	100	0.0	-0.6	0.200	0.050	0.041	0.359	0.0	0.6	2.0
(0.021)	[0.620	0.663]				(0.009)	[0.032	0.049]		
0.830	150	0.0	-0.6	0.200	0.050	0.042	0.170	0.0	0.6	2.0
(0.016)	[0.814	0.846]				(0.009)	[0.033	0.050]		

Tests for Paired Means (Simulation)

Notes:
Number of Monte Carlo Samples: 2000.   Simulation Run Time: 19.33 seconds.

Report Definitions
Power is the probability of rejecting a false null hypothesis.
N is the size of the sample drawn from the population.
Diff0 is the paired-difference mean (A-B) assuming the null hypothesis, H0. This is the value being tested.
Diff1 is the paired-difference mean (A-B) assuming the alternative hypothesis, H1. This is the true value.
R is the correlation between the paired items.
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.
Beta is the probability of accepting a false null hypothesis.
Second Row: (Power Inc.) [95% LCL and UCL Power]   (Alpha Inc.) [95% LCL and UCL Alpha]

Summary Statements
A sample size of 50 achieves 39% power to detect a difference of -0.6 between the null
hypothesis mean difference of 0.0 and the actual mean difference of -0.6 at the 0.050
significance level (alpha) using a two-sided Paired T-Test. These results are based on 2000
Monte Carlo samples from the null distribution: Normal(M0 S) - Normal(M0 S) and the alternative
distribution: Normal(M0 S) - Normal(M1 S).

Plots Section
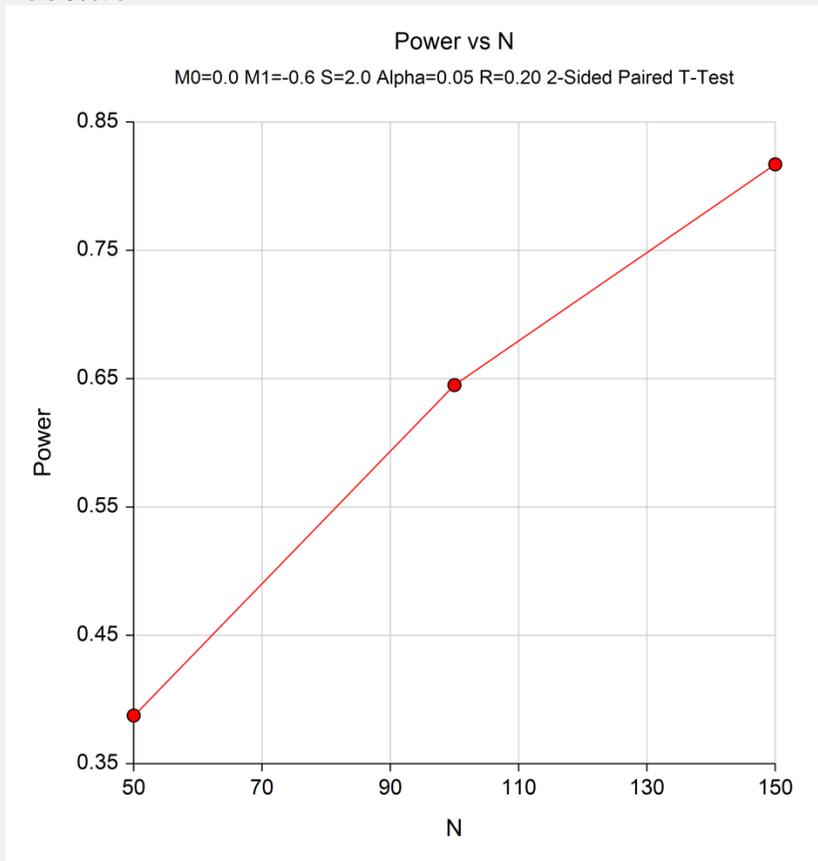
This report shows the estimated power for each scenario. The first row shows the parameter settings and the estimated power and significance level (Actual Alpha).

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 intervals will decrease.

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 by expanding **Means**, then expanding **Paired Means**, then clicking on **Tests (Inequality)**, and then clicking on **Tests for Paired Means (Simulation)**. 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**
Alternative Hypothesis	**Diff ≠ Diff0**
Test Type	**T-Test**
Simulations	**2000**
Power	**0.90**
Alpha	**0.05**
Item A Distribution|H0	**Normal(M0 S)**
Item B Distribution|H0	**Normal(M0 S)**
Item A Distribution|H1	**Normal(M0 S)**
Item B Distribution|H1	**Normal(M1 S)**
M0 (Mean|H0)	**0**
M1 (Mean|H1)	**0.6**
S	**2**
R (Correlation of Items A & B)	**0.2**

Output

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

Numeric Results of Search for N

Power	N	H0 Diff0	H1 Diff1	Corr R	Target Alpha	Actual Alpha	Beta	M0	M1	S
0.897	187	0.0	-0.6	0.200	0.050	0.053	0.104	0.0	0.6	2.0
(0.013)	[0.883	0.910]				(0.010)	[0.043	0.062]		

Notes:
Number of Monte Carlo Samples: 2000.   Simulation Run Time: 95.53 seconds.

The required sample size of 187 achieved a power of 0.897. The power of 0.897 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.

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 by expanding **Means**, then expanding **Paired Means**, then clicking on **Tests (Inequality)**, and then clicking on **Tests for Paired Means (Simulation)**. 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**.

Option	Value
Design Tab	
Solve For	**Power**
Alternative Hypothesis	**Diff ≠ Diff0**
Test Type...	**T-Test**
Simulations	**2000**
Alpha..	**0.05**
N (Sample Size)......................................	**50 100 150 200**
Item A Distribution|H0............................	**Normal(M0 S)**
Item B Distribution|H0............................	**Normal(M0 S)**
Item A Distribution|H1............................	**Normal(M0 S)**
Item B Distribution |H1...........................	**Normal(M1 S)**
M0 (Mean|H0)	**0**
M1 (Mean|H1)	**0.6**
S..	**2**
R (Correlation of Items A & B)	**0.2**
Reports Tab	
Show Comparative Reports....................	**Checked**
Show Comparative Plots	**Checked**

Output

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

Numeric Results

Power Comparison for Testing Mean Difference = Diff0.   Hypotheses: H0: Diff1 = Diff0; H1: Diff1≠ Diff0
H0 Dist'n: Normal(M0 S) - Normal(M0 S)
H1 Dist'n: Normal(M0 S) - Normal(M1 S)

N	H0 Diff (Diff0)	H1 Diff (Diff1)	Corr (R)	Target Alpha	T-Test Power	Wilcxn Power	Sign Power	M0	M1	S
50	0.0	-0.6	0.200	0.050	0.355	0.335	0.187	0.0	0.6	2.0
100	0.0	-0.6	0.200	0.050	0.644	0.610	0.385	0.0	0.6	2.0
150	0.0	-0.6	0.200	0.050	0.843	0.818	0.617	0.0	0.6	2.0
200	0.0	-0.6	0.200	0.050	0.921	0.905	0.734	0.0	0.6	2.0

Number of Monte Carlo Iterations: 2000.   Simulation Run Time: 10.50 seconds.

Alpha Comparison for Testing Mean Difference = Diff0.   Hypotheses: H0: Diff1 = Diff0; H1: Diff1 ≠ Diff0
H0 Dist'n: Normal(M0 S) - Normal(M0 S)
H1 Dist'n: Normal(M0 S) - Normal(M1 S)

N	H0 Diff (Diff0)	H1 Diff (Diff1)	Corr (R)	Target Alpha	T-Test Alpha	Wilcxn Alpha	Sign Alpha	M0	M1	S
50	0.0	-0.6	0.200	0.050	0.059	0.060	0.034	0.0	0.6	2.0
100	0.0	-0.6	0.200	0.050	0.050	0.052	0.033	0.0	0.6	2.0
150	0.0	-0.6	0.200	0.050	0.055	0.050	0.041	0.0	0.6	2.0
200	0.0	-0.6	0.200	0.050	0.058	0.058	0.050	0.0	0.6	2.0

Number of Monte Carlo Iterations: 2000.   Simulation Run Time: 10.50 seconds.

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

Example 4 – Validation

We will validate this procedure by comparing its results to those of the regular one-sample t-test, a procedure that has already by validated. For this run, we will use the settings of Example 1: M0 = 0, M1 = 0.6, alpha = 0.05, N = 50, R = 0.2, and S = 2.

Note that to run this example using the regular one-sample t-test procedure, the variance will have to be altered to account for the correlation of 0.20. The adjusted standard deviation is equal to S times the square root of $2(1 - R)$, which, in this case, is 2.530. Running this through the regular One Mean procedure yields a power of 0.376.

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 by expanding **Means**, then expanding **Paired Means**, then clicking on **Tests (Inequality)**, and then clicking on **Tests for Paired Means (Simulation)**. 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**.

Option   **Value**

Design Tab
Solve For .... **Power**
Alternative Hypothesis .... **Diff ≠ Diff0**
Test Type .... **T-Test**
Simulations **10000**
Alpha **0.05**
N (Sample Size) **50**
Item A Distribution|H0 .... **Normal(M0 S)**
Item B Distribution|H0 .... **Normal(M0 S)**
Item A Distribution|H1 .... **Normal(M0 S)**
Item B Distribution|H1 .... **Normal(M1 S)**
M0 (Mean|H0) **0**
M1 (Mean|H1) **0.6**
S **2**
R (Correlation of Items A & B) **0.2**

Options Tab
Random Number Pool Size .... **50000 (Increase to 5 times Simulations)**

Output

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

Numeric Results

Power	N	H0 Diff0	H1 Diff1	Corr R	Target Alpha	Actual Alpha	Beta	M0	M1	S
0.373	50	0.0	-0.6	0.200	0.050	0.049	0.627	0.0	0.6	2.0
(0.009)	[0.363	0.382]				(0.004)	[0.045	0.053]		

Notes:
Number of Monte Carlo Samples: 10000.   Simulation Run Time: 30.97 seconds.

The power matches the exact value of 0.376 quite well. We re-ran the procedure several times and obtained power values from 0.370 to 0.396.

Example 5 – 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 equivalence* 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

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

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.

Past experience has shown that the standard deviation is 5.0 and the correlation is 0.2. 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 actually 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 by expanding **Means**, then expanding **Paired Means**, then clicking on **Tests (Inequality)**, and then clicking on **Tests for Paired Means (Simulation)**. 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**.

Tests for Paired Means (Simulation)

Option	Value
Design Tab	
Solve For	**Power**
Alternative Hypothesis	**Diff > Diff0**
Test Type..	**T-Test**
Simulations	**2000**
Alpha..	**0.025**
N (Sample Size)......................................	**5 10 15 20 25**
Item A Distribution|H0.............................	**Normal(M0 S)**
Item B Distribution|H0.............................	**Normal(M1 S)**
Item A Distribution|H1	**Normal(M0 S)**
Item B Distribution|H1	**Normal(M0 S)**
M0 (Mean|H0)	**0**
M1 (Mean|H1)	**5**
S...	**5**
R (Correlation of Items A & B)	**0.2**

Output

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

Numeric Results

Numeric Results for Testing Mean Difference = Diff0.   Hypotheses: H0: Diff1 = Diff0; H1: Diff1 > Diff0
H0 Dist'n: Normal(M0 S) - Normal(M1 S)
H1 Dist'n: Normal(M0 S) - Normal(M0 S)
Test Statistic: Paired T-Test

Power	N	H0 Diff0	H1 Diff1	Corr R	Target Alpha	Actual Alpha	Beta	M0	M1	S
0.308	5	-5.0	0.0	0.200	0.025	0.023	0.692	0.0	5.0	5.0
(0.020)	[0.288	0.328]				(0.007)	[0.016	0.030]		
0.617	10	-5.0	0.0	0.200	0.025	0.024	0.383	0.0	5.0	5.0
(0.021)	[0.596	0.638]				(0.007)	[0.017	0.030]		
0.816	15	-5.0	0.0	0.200	0.025	0.027	0.184	0.0	5.0	5.0
(0.017)	[0.799	0.833]				(0.007)	[0.019	0.034]		
0.916	20	-5.0	0.0	0.200	0.025	0.022	0.085	0.0	5.0	5.0
(0.012)	[0.903	0.928]				(0.006)	[0.016	0.028]		
0.968	25	-5.0	0.0	0.200	0.025	0.025	0.032	0.0	5.0	5.0
(0.008)	[0.960	0.976]				(0.007)	[0.018	0.031]		

Notes:
Number of Monte Carlo Samples: 2000.   Simulation Run Time: 13.34 seconds.

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