

## Chapter 490

# Tests for Paired Means (Simulation) (Legacy)

---

## 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  $X_1, X_2, \dots, X_n$  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. 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  $Y_1, Y_2, \dots, Y_n$  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. 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.

## Tests for Paired Means (Simulation) (Legacy)

- 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:

- 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  $M$ .
- Compute the correlation between the two generated variates.
- If the computed correlation is within a small tolerance (usually less than 0.001) of the specified correlation, go to step 7.
- Select two indices ( $I$  and  $J$ ) at random using uniform random numbers.
- 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.

## Tests for Paired Means (Simulation) (Legacy)

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.

### 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

$$t_{n-1} = \frac{\bar{X} - M0}{s_{\bar{X}}/\sqrt{n}}$$

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.

## Tests for Paired Means (Simulation) (Legacy)

**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,  $M_0$ , from each data value. Rank the values according to their absolute values.
2. Compute the sum of the positive ranks  $S_p$  and the sum of the negative ranks  $S_n$ . The test statistic,  $W_R$ , is the minimum of  $S_p$  and  $S_n$ .
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  $i^{\text{th}}$  value occurs.

4. Compute the z-value using

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

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  $M_0$ . Call this value  $X$ .
2. Count the number of values strictly less than  $M_0$ . 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.

## Tests for Paired Means (Simulation) (Legacy)

## 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

$$t_x = \frac{\bar{X} - M0}{s_{\bar{X}}/\sqrt{n}}$$

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 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!

## 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

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 .....	<b>Power</b>
Alternative Hypothesis .....	<b>Diff ≠ Diff0</b>
Test Type.....	<b>T-Test</b>
Simulations .....	<b>2000</b>
Random Seed.....	<b>3827024</b> (for Reproducibility)
Alpha.....	<b>0.05</b>
N (Sample Size).....	<b>50 100 150</b>
Item A Distribution H0.....	<b>Normal(M0 S)</b>
Item A Distribution H1 .....	<b>Normal(M0 S)</b>
Item B Distribution H0.....	<b>Normal(M0 S)</b>
Item B Distribution H1 .....	<b>Normal(M1 S)</b>
M0 (Mean H0) Parameter Value(s).....	<b>0</b>
M1 (Mean H1) Parameter Value(s).....	<b>0.6</b>
Parameter 1 Label .....	<b>S</b>
Parameter 1 Value(s).....	<b>2</b>
R (Correlation of Items A & B) .....	<b>0.2</b>

## Output

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

### Numeric Results

Solve For: [Power](#)  
 Hypotheses: H0: Diff1 = Diff0 vs. H1: Diff1 ≠ Diff0  
 H0 Distribution: Normal(M0 S) - Normal(M0 S)  
 H1 Distribution: 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
0.396 (0.021) [0.374 0.417]	50	0	-0.6	0.2	0.05	0.053 (0.01) [0.043 0.063]	0.605	0	0.6	2
0.653 (0.021) [0.632 0.674]	100	0	-0.6	0.2	0.05	0.054 (0.01) [0.044 0.063]	0.347	0	0.6	2
0.827 (0.017) [0.81 0.844]	150	0	-0.6	0.2	0.05	0.054 (0.01) [0.044 0.063]	0.173	0	0.6	2

Pool Size: 10000. Simulations: 2000. Run Time: 1.19 seconds.  
 User-Entered Random Seed: 3827024

- Power The probability of rejecting a false null hypothesis when the alternative hypothesis is true. The second row provides the precision and a 95% confidence interval for Power, (Power Precision) [95% LCL and UCL], based on the size of the simulation.
- N The size of the sample drawn from the population.
- Diff0 The paired-difference mean (A-B) assuming the null hypothesis, H0. This is the value being tested.
- Diff1 The paired-difference mean (A-B) assuming the alternative hypothesis, H1. This is the true value.
- R The correlation between the paired items.
- Target Alpha The probability of rejecting a true null hypothesis. It is set by the user.
- Actual Alpha The alpha level that was actually achieved by the experiment. The second row provides the precision and a 95% confidence interval for Alpha, (Alpha Precision) [95% LCL and UCL], based on the size of the simulation.
- Beta The probability of failing to reject the null hypothesis when the alternative hypothesis is true.

### Summary Statements

A sample size of 50 pairs with a correlation of 0.2 achieves 40% power to detect a difference of -0.6 between the null hypothesis mean difference of 0 and the actual mean difference of -0.6 at the 0.05 significance level (alpha) using a two-sided Paired T-Test. These results are based on 2000 simulations (Monte Carlo samples) from the null distribution: Normal(M0 S) - Normal(M0 S) and the alternative distribution: Normal(M0 S) - Normal(M1 S).

## Tests for Paired Means (Simulation) (Legacy)

**Dropout-Inflated Sample Size**

Dropout Rate	Sample Size N	Dropout- Inflated Enrollment Sample Size N'	Expected Number of Dropouts D
20%	50	63	13
20%	100	125	25
20%	150	188	38

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, 63 subjects should be enrolled to obtain a final sample size of 50 subjects.

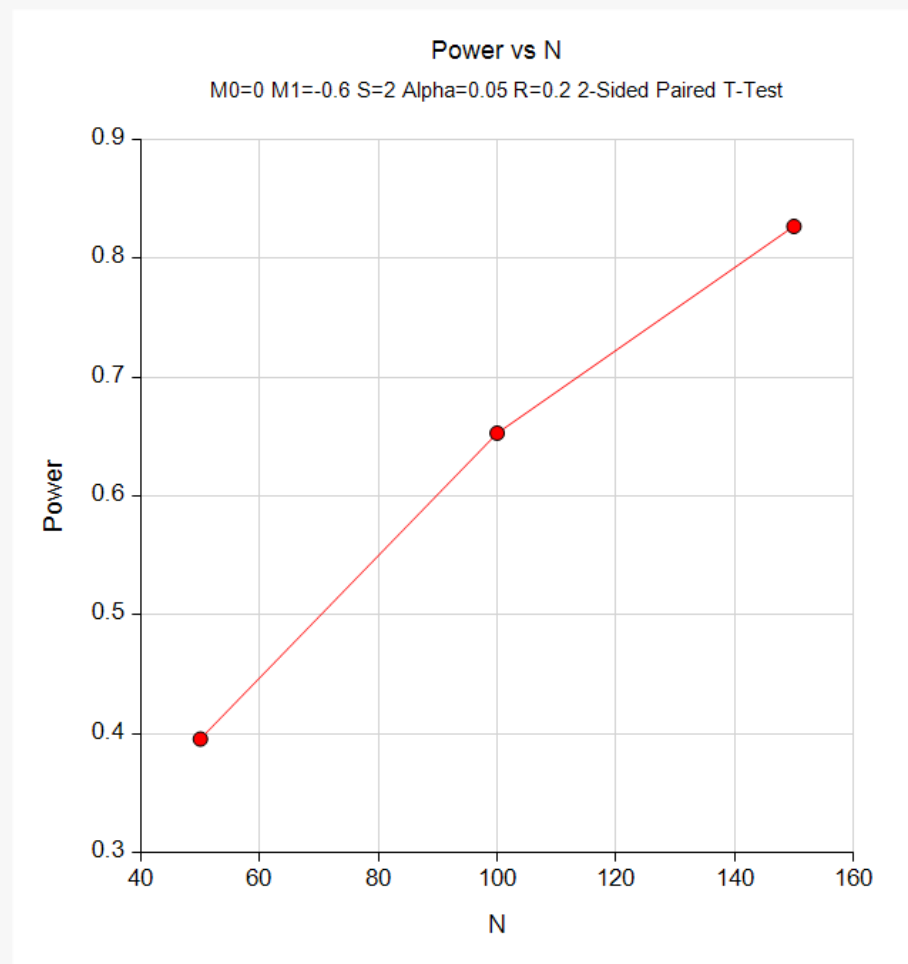
**References**

- Chow, S.C., Shao, J., and Wang, H. 2003. Sample Size Calculations in Clinical Research. Marcel Dekker. New York.
- Devroye, Luc. 1986. Non-Uniform Random Variate Generation. Springer-Verlag. New York.
- Matsumoto, M. and Nishimura, T. 1998. 'Mersenne twister: A 623-dimensionally equidistributed uniform pseudorandom number generator.' ACM Trans. On Modeling and Computer Simulations.



## Tests for Paired Means (Simulation) (Legacy)

## Plots



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

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 .....	<b>Sample Size</b>
Alternative Hypothesis .....	<b>Diff <math>\neq</math> Diff0</b>
Test Type.....	<b>T-Test</b>
Simulations .....	<b>2000</b>
Random Seed.....	<b>3846009</b> (for Reproducibility)
Power.....	<b>0.90</b>
Alpha.....	<b>0.05</b>
Item A Distribution H0 .....	<b>Normal(M0 S)</b>
Item A Distribution H1 .....	<b>Normal(M0 S)</b>
Item B Distribution H0 .....	<b>Normal(M0 S)</b>
Item B Distribution H1 .....	<b>Normal(M1 S)</b>
M0 (Mean H0) Parameter Value(s) .....	<b>0</b>
M1 (Mean H1) Parameter Value(s) .....	<b>0.6</b>
Parameter 1 Label .....	<b>S</b>
Parameter 1 Value(s).....	<b>2</b>
R (Correlation of Items A & B) .....	<b>0.2</b>

## Tests for Paired Means (Simulation) (Legacy)

## Output

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

### Numeric Results

Solve For: [Sample Size](#)  
 Hypotheses: H0: Diff1 = Diff0 vs. H1: Diff1  $\neq$  Diff0  
 H0 Distribution: Normal(M0 S) - Normal(M0 S)  
 H1 Distribution: 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
0.903 (0.013) [0.889 0.916]	191	0	-0.6	0.2	0.05	0.049 (0.009) [0.04 0.058]	0.098	0	0.6	2

Pool Size: 10000. Simulations: 2000. Run Time: 5.07 seconds.  
 User-Entered Random Seed: 3846009

The required sample size of 191 achieved a power of 0.903. 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

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 .....	<b>Power</b>
Alternative Hypothesis .....	<b>Diff ≠ Diff0</b>
Test Type .....	<b>T-Test</b>
Simulations .....	<b>2000</b>
Random Seed .....	<b>3855041</b> (for Reproducibility)
Alpha .....	<b>0.05</b>
N (Sample Size) .....	<b>50 100 150 200</b>
Item A Distribution H0 .....	<b>Normal(M0 S)</b>
Item A Distribution H1 .....	<b>Normal(M0 S)</b>
Item B Distribution H0 .....	<b>Normal(M0 S)</b>
Item B Distribution H1 .....	<b>Normal(M1 S)</b>
M0 (Mean H0) Parameter Value(s) .....	<b>0</b>
M1 (Mean H1) Parameter Value(s) .....	<b>0.6</b>
Parameter 1 Label .....	<b>S</b>
Parameter 1 Value(s) .....	<b>2</b>
R (Correlation of Items A & B) .....	<b>0.2</b>

#### Reports Tab

Show Comparative Reports .....	<b>Checked</b>
Include T-Test Results .....	<b>Checked</b>
Include Wilcoxon & Sign Test .....	<b>Checked</b>

#### Comparative Plots Tab

Show Comparative Plots .....	<b>Checked</b>
------------------------------	----------------

## Tests for Paired Means (Simulation) (Legacy)

## Output

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

### Power Comparison

Hypotheses: H0: Diff1 = Diff0 vs. H1: Diff1 ≠ Diff0  
 H0 Distribution: Normal(M0 S) - Normal(M0 S)  
 H1 Distribution: Normal(M0 S) - Normal(M1 S)

N	H0 Diff Diff0	H1 Diff Diff1	Corr R	Target Alpha	T-Test Power	Wilcoxon Power	Sign Power	M0	M1	S
50	0	-0.6	0.2	0.05	0.388	0.375	0.219	0	0.6	2
100	0	-0.6	0.2	0.05	0.630	0.629	0.440	0	0.6	2
150	0	-0.6	0.2	0.05	0.815	0.808	0.609	0	0.6	2
200	0	-0.6	0.2	0.05	0.925	0.925	0.835	0	0.6	2

Pool Size: 10000. Simulations: 2000. Run Time: 2.54 seconds.  
 User-Entered Random Seed: 3855041

### Alpha Comparison

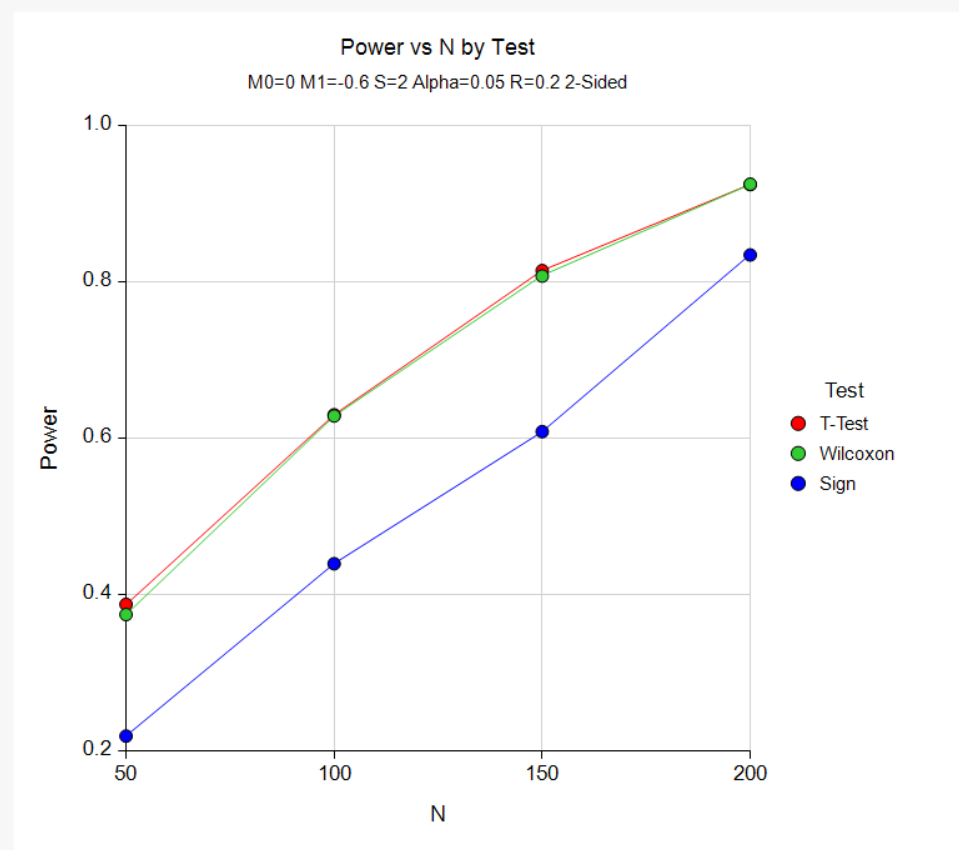
Hypotheses: H0: Diff1 = Diff0 vs. H1: Diff1 ≠ Diff0  
 H0 Distribution: Normal(M0 S) - Normal(M0 S)  
 H1 Distribution: Normal(M0 S) - Normal(M1 S)

N	H0 Diff Diff0	H1 Diff Diff1	Corr R	Target Alpha	T-Test Alpha	Wilcoxon Alpha	Sign Alpha	M0	M1	S
50	0	-0.6	0.2	0.05	0.049	0.049	0.036	0	0.6	2
100	0	-0.6	0.2	0.05	0.040	0.040	0.029	0	0.6	2
150	0	-0.6	0.2	0.05	0.055	0.056	0.040	0	0.6	2
200	0	-0.6	0.2	0.05	0.044	0.041	0.031	0	0.6	2

Pool Size: 10000. Simulations: 2000. Run Time: 2.54 seconds.  
 User-Entered Random Seed: 3855041

## Tests for Paired Means (Simulation) (Legacy)

## Comparative Plots



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 been validated. For this run, we will use the settings of Example 1:  $M_0 = 0$ ,  $M_1 = 0.6$ ,  $\alpha = 0.05$ ,  $N = 50$ ,  $R = 0.2$ , 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

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 4** 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 .....	<b>Power</b>
Alternative Hypothesis .....	<b>Diff <math>\neq</math> Diff0</b>
Test Type .....	<b>T-Test</b>
Simulations .....	<b>10000</b>
Random Seed .....	<b>3859722</b> (for Reproducibility)
Alpha .....	<b>0.05</b>
N (Sample Size) .....	<b>50</b>
Item A Distribution H0 .....	<b>Normal(M0 S)</b>
Item A Distribution H1 .....	<b>Normal(M0 S)</b>
Item B Distribution H0 .....	<b>Normal(M0 S)</b>
Item B Distribution H1 .....	<b>Normal(M1 S)</b>
M0 (Mean H0) Parameter Value(s) .....	<b>0</b>
M1 (Mean H1) Parameter Value(s) .....	<b>0.6</b>
Parameter 1 Label .....	<b>S</b>
Parameter 1 Value(s) .....	<b>2</b>
Parameter 2 Label .....	<b>A</b>
Parameter 2 Value(s) .....	<b>1</b>
Parameter 3 Label .....	<b>B</b>
Parameter 3 Value(s) .....	<b>1</b>
R (Correlation of Items A & B) .....	<b>0.2</b>

#### Options Tab

Random Number Pool Size .....	<b>50000</b> (Increase to 5 times the number of Simulations)
-------------------------------	--

## Tests for Paired Means (Simulation) (Legacy)

## Output

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

### Numeric Results

Solve For: [Power](#)  
 Hypotheses: H0: Diff1 = Diff0 vs. H1: Diff1  $\neq$  Diff0  
 H0 Distribution: Normal(M0 S) - Normal(M0 S)  
 H1 Distribution: 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
0.38 (0.01) [0.371 0.39]	50	0	-0.6	0.2	0.05	0.053 (0.004) [0.049 0.058]	0.62	0	0.6	2

Pool Size: 50000. Simulations: 10000. Run Time: 1.22 seconds.  
 User-Entered Random Seed: 3859722

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

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

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.

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

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 5** 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 .....	<b>Power</b>
Alternative Hypothesis .....	<b>Diff &gt; Diff0</b>
Test Type.....	<b>T-Test</b>
Simulations .....	<b>2000</b>
Random Seed.....	<b>3866680</b> (for Reproducibility)
Alpha.....	<b>0.025</b>
N (Sample Size).....	<b>5 10 15 20 25</b>
Item A Distribution H0.....	<b>Normal(M0 S)</b>
Item A Distribution H1 .....	<b>Normal(M0 S)</b>
Item B Distribution H0.....	<b>Normal(M1 S)</b>
Item B Distribution H1 .....	<b>Normal(M0 S)</b>
M0 (Mean H0) Parameter Value(s).....	<b>0</b>
M1 (Mean H1) Parameter Value(s).....	<b>5</b>

Tests for Paired Means (Simulation) (Legacy)

Parameter 1 Label ..... **S**  
 Parameter 1 Value(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

Solve For: **Power**  
 Hypotheses: H0: Diff1 = Diff0 vs. H1: Diff1 > Diff0  
 H0 Distribution: Normal(M0 S) - Normal(M1 S)  
 H1 Distribution: 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
0.296 (0.02) [0.276 0.316]	5	-5	0	0.2	0.025	0.029 (0.007) [0.021 0.036]	0.704 [0.021 0.036]	0	5	5
0.611 (0.021) [0.59 0.632]	10	-5	0	0.2	0.025	0.023 (0.006) [0.016 0.029]	0.389 [0.016 0.029]	0	5	5
0.797 (0.018) [0.779 0.814]	15	-5	0	0.2	0.025	0.021 (0.006) [0.015 0.027]	0.204 [0.015 0.027]	0	5	5
0.917 (0.012) [0.905 0.929]	20	-5	0	0.2	0.025	0.025 (0.007) [0.018 0.032]	0.083 [0.018 0.032]	0	5	5
0.963 (0.008) [0.955 0.971]	25	-5	0	0.2	0.025	0.022 (0.006) [0.016 0.028]	0.037 [0.016 0.028]	0	5	5

Pool Size: 10000. Simulations: 2000. Run Time: 0.77 seconds.  
 User-Entered Random Seed: 3866680

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