

## Chapter 495

# Equivalence Tests for Paired Means (Simulation)

---

## Introduction

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

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.

## Equivalence Tests for Paired Means (Simulation)

- 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 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.

---

## 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.

## Equivalence Tests for Paired Means (Simulation)

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.

---

## Simulating Data for an Equivalence Test

Simulating equivalence data is more complex than simulating data for a regular two-sided test. An equivalence test essentially reverses the roles of the null and alternative hypothesis. In so doing, the null hypothesis becomes

$$H_0: (\mu_1 - \mu_2) \leq -D \text{ or } (\mu_1 - \mu_2) \geq D$$

where  $D$  is the margin of equivalence. Thus, the null hypothesis is made up of two simple hypotheses:

$$H_{0_1}: (\mu_1 - \mu_2) \leq -D$$

$$H_{0_2}: (\mu_1 - \mu_2) \geq D$$

The additional complexity comes in deciding which of the two simple null hypotheses are used to simulate data for the null hypothesis situation. The choice becomes more problematic when asymmetric equivalence limits are chosen. In that case, you may want to try simulating using each simple null hypothesis in turn.

To generate data for the null hypotheses, you generate data for each group. The difference in the means of these two groups will become one of the equivalence limits. The other equivalence limit will be determined by symmetry and will always have a sign that is the negative of the first equivalence limit.

---

## 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

$$t_{n-1} = \frac{\bar{X} - M_0}{s_{\bar{X}}}$$

## Equivalence Tests for Paired Means (Simulation)

where

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

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

and  $M_0$  is the value of the difference 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  $S_p$  and the sum of the negative ranks  $S_n$ . The test statistic,  $W$ , is the minimum of  $S_p$  and  $S_n$ .
3. Compute the mean and standard deviation of  $W$  using the formulas

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

where  $t_i$  represents the number of times the  $i^{\text{th}}$  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.

## Equivalence Tests for Paired Means (Simulation)

## 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  $\bar{X}$ .
2. Compute the t-value using the standard t-test. The formula for this computation is

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

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

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

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 an experiment to determine if the response to a new drug is equivalent to the response to the standard drug. The average response level to the standard drug is 63 with a standard deviation of 5. The researchers decide that if the average response level to the new drug is between 60 and 66, they will consider it to be equivalent to the standard drug.

The researchers decide to use a paired design so that each subject can serve as their own control. The response level for the standard drug will be measured for each subject. Then, followed by an appropriate wash-out period of two days, the response level to the new drug will be measured. From previous studies, they know that the correlation between the two response levels will be between 0.1 and 0.20.

The researchers will analyze the data using an equivalence test based on the paired t-test with an alpha level of 0.05. They want to compare the power at sample sizes of 10, 30, 50, and 70. They assume that the data are normally distributed and that the true difference between the response level of the two drugs is zero. 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>
Equivalence Limit.....	<b>Symmetric</b>
Test Type.....	<b>T-Test</b>
Simulations .....	<b>2000</b>
Random Seed.....	<b>5139011</b> (for Reproducibility)
Alpha.....	<b>0.05</b>
N (Sample Size).....	<b>10 30 50 70</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>63</b>
M1 (Mean H1) Parameter Value(s) .....	<b>66</b>
Parameter 1 Label .....	<b>S</b>
Parameter 1 Value(s).....	<b>5</b>
R (Correlation of Items A & B) .....	<b>0.1 0.2</b>

Equivalence Tests for Paired Means (Simulation)

# Output

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

### Numeric Results

Solve For: [Power](#)  
 Hypotheses: H0: Diff  $\geq$  |Diff0| vs. H1: Diff < |Diff0|  
 H0 Distribution: Normal(M0 S) - Normal(M1 S)  
 H1 Distribution: Normal(M0 S) - Normal(M0 S)  
 Test Statistic: Paired T-Test

Power	Sample Size N	Paired Difference Diff1	Equivalence Limits		Correlation of A and B R	Alpha		M0	M1	S
			Lower	Upper		Target	Actual			
0.031 (0.008) [0.023 0.038]	10	0	-3	3	0.1	0.05	0.008 (0.004) [0.004 0.012]	63	66	5
0.052 (0.01) [0.042 0.062]	10	0	-3	3	0.2	0.05	0.019 (0.006) [0.013 0.024]	63	66	5
0.520 (0.022) [0.498 0.541]	30	0	-3	3	0.1	0.05	0.047 (0.009) [0.038 0.056]	63	66	5
0.639 (0.021) [0.618 0.66]	30	0	-3	3	0.2	0.05	0.045 (0.009) [0.036 0.054]	63	66	5
0.855 (0.015) [0.84 0.87]	50	0	-3	3	0.1	0.05	0.046 (0.009) [0.037 0.055]	63	66	5
0.895 (0.013) [0.881 0.908]	50	0	-3	3	0.2	0.05	0.054 (0.01) [0.044 0.063]	63	66	5
0.958 (0.009) [0.949 0.967]	70	0	-3	3	0.1	0.05	0.056 (0.01) [0.046 0.066]	63	66	5
0.976 (0.007) [0.969 0.983]	70	0	-3	3	0.2	0.05	0.043 (0.009) [0.034 0.051]	63	66	5

Pool Size: 10000. Simulations: 2000. Run Time: 1.77 seconds.  
 User-Entered Random Seed: 5139011

- 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.
- Diff1: The paired-difference mean (A-B) assuming the alternative hypothesis, H1. This is the true value.
- Lower Equivalence Limit: The lower limit on a difference (A-B) that is considered as equivalent.
- Upper Equivalence Limit: The upper limit on a difference (A-B) that is considered as equivalent.
- Diff0: The paired-difference mean (A-B) assuming the null hypothesis, H0. This is one of the equivalence limits.
- 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.
- Parameters: The additional columns in the report represent the distribution parameters used (if any) in the calculations.

### Summary Statements

A paired design will be used to test whether the Item A mean ( $\mu_A$ ) is equivalent to the Item B mean ( $\mu_B$ ), with lower and upper mean difference equivalence limits of -3 and 3 (H0:  $\delta \leq -3$  or  $\delta \geq 3$  versus H1:  $-3 < \delta < 3$ ,  $\delta = \mu_A - \mu_B$ ). The comparison will be made using two one-sided, paired-difference t-tests, with an overall Type I error rate ( $\alpha$ ) of 0.05. The correlation between the two items (A and B) is assumed to be 0.1. To detect a mean difference ( $\mu_A - \mu_B$ ) of 0 with a sample size of 10 pairs, the power is 0.031. These results are based on 2000 simulations (Monte Carlo samples) from the null distribution: Normal(M0 S) - Normal(M1 S), and the alternative distribution: Normal(M0 S) - Normal(M0 S), with M0 = 63, M1 = 66, S = 5.



## Equivalence Tests for Paired Means (Simulation)

**Dropout-Inflated Sample Size**

Dropout Rate	Sample Size N	Dropout- Inflated Enrollment Sample Size N'	Expected Number of Dropouts D
20%	10	13	3
20%	30	38	8
20%	50	63	13
20%	70	88	18

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**

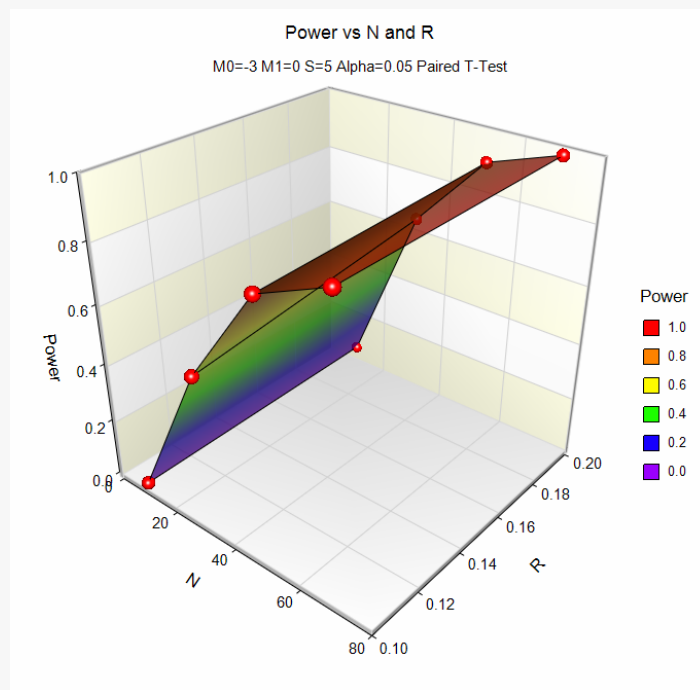
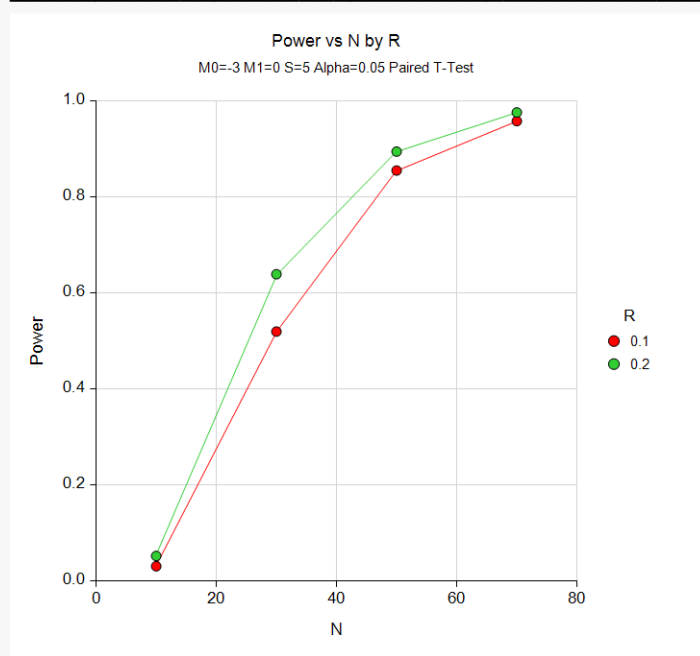
Anticipating a 20% dropout rate, 13 subjects should be enrolled to obtain a final sample size of 10 subjects.

**References**

- Blackwelder, W.C. 1998. 'Equivalence Trials.' In Encyclopedia of Biostatistics, John Wiley and Sons. New York. Volume 2, 1367-1372.
- 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.

Equivalence Tests for Paired Means (Simulation)

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.

We see that a sample size of about 50 is needed to obtain a reasonable power level.

## 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? They decide to use a correlation of 0.10, since that will result in a larger, more conservative, sample size.

### 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>
Equivalence Limit.....	<b>Symmetric</b>
Test Type.....	<b>T-Test</b>
Simulations .....	<b>2000</b>
Random Seed.....	<b>8464844</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(M1 S)</b>
Item B Distribution H1 .....	<b>Normal(M0 S)</b>
M0 (Mean H0) Parameter Value(s) .....	<b>63</b>
M1 (Mean H1) Parameter Value(s) .....	<b>66</b>
Parameter 1 Label .....	<b>S</b>
Parameter 1 Value(s).....	<b>5</b>
R (Correlation of Items A & B) .....	<b>0.1</b>

## Equivalence Tests for Paired Means (Simulation)

## Output

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

### Numeric Results

Solve For: [Sample Size](#)  
 Hypotheses:  $H_0: \text{Diff} \geq |\text{Diff}_0|$  vs.  $H_1: \text{Diff} < |\text{Diff}_0|$   
 H0 Distribution: Normal(M0 S) - Normal(M1 S)  
 H1 Distribution: Normal(M0 S) - Normal(M0 S)  
 Test Statistic: Paired T-Test

Power	Sample Size N	Paired Difference Diff1	Equivalence Limits		Correlation of A and B R	Alpha		M0	M1	S
			Lower	Upper		Target	Actual			
0.919 (0.012) [0.907 0.93]	57	0	-3	3	0.1	0.05	0.061 (0.01) [0.051 0.071]	63	66	5

Pool Size: 10000. Simulations: 2000. Run Time: 1.89 seconds.  
 User-Entered Random Seed: 8464844

The required sample size was 57, which achieved a power of 0.919.

It is possible for the search algorithm to converge to a sample size which exhibits the desired power, but then on the second simulation, achieves a power that is slightly less than the target. To obtain more accuracy, a reasonable strategy would be to run simulations to obtain the powers using N's from 50 to 60 using a simulation size of 5000.

## Example 3 – Comparing Test Statistics

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>
Equivalence Limit.....	<b>Symmetric</b>
Test Type.....	<b>T-Test</b>
Simulations .....	<b>2000</b>
Random Seed.....	<b>9846168</b> (for Reproducibility)
Alpha.....	<b>0.05</b>
N (Sample Size).....	<b>10 30 50 70</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>63</b>
M1 (Mean H1) Parameter Value(s) .....	<b>66</b>
Parameter 1 Label .....	<b>S</b>
Parameter 1 Value(s).....	<b>5</b>
R (Correlation of Items A & B) .....	<b>0.1</b>

#### Reports Tab

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

#### Comparative Plots Tab

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

## Equivalence Tests for Paired Means (Simulation)

## Output

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

## Power Comparison

Hypotheses:  $H_0: \text{Diff} \geq |\text{Diff}_0|$  vs.  $H_1: \text{Diff} < |\text{Diff}_0|$   
 H0 Distribution: Normal(M0 S) - Normal(M1 S)  
 H1 Distribution: Normal(M0 S) - Normal(M0 S)

Sample Size N	Paired Difference Diff1	Equivalence Limits		Correlation of A and B R	Alpha	Power			M0	M1	S
		Lower	Upper			T-Test	Wilcoxon	Sign			
10	0	-3	3	0.1	0.05	0.035	0.031	0.004	63	66	5
30	0	-3	3	0.1	0.05	0.521	0.469	0.293	63	66	5
50	0	-3	3	0.1	0.05	0.859	0.843	0.521	63	66	5
70	0	-3	3	0.1	0.05	0.961	0.954	0.770	63	66	5

Pool Size: 10000. Simulations: 2000. Run Time: 3.03 seconds.  
 User-Entered Random Seed: 9846168

## Alpha Comparison

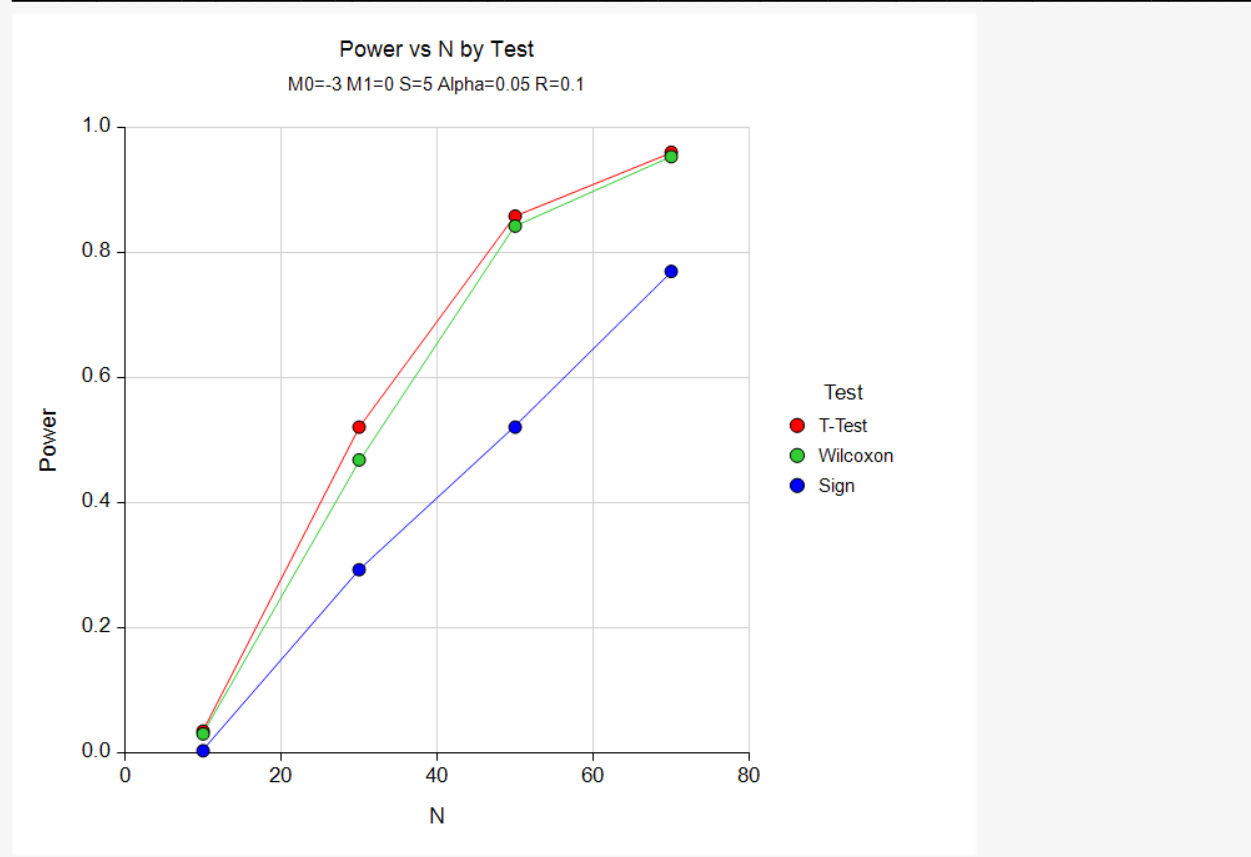
Hypotheses:  $H_0: \text{Diff} \geq |\text{Diff}_0|$  vs.  $H_1: \text{Diff} < |\text{Diff}_0|$   
 H0 Distribution: Normal(M0 S) - Normal(M1 S)  
 H1 Distribution: Normal(M0 S) - Normal(M0 S)

Sample Size N	Paired Difference Diff1	Equivalence Limits		Correlation of A and B R	Alpha Target	Alpha			M0	M1	S
		Lower	Upper			T-Test	Wilcoxon	Sign			
10	0	-3	3	0.1	0.05	0.012	0.008	0.002	63	66	5
30	0	-3	3	0.1	0.05	0.053	0.051	0.047	63	66	5
50	0	-3	3	0.1	0.05	0.049	0.055	0.037	63	66	5
70	0	-3	3	0.1	0.05	0.049	0.052	0.042	63	66	5

Pool Size: 10000. Simulations: 2000. Run Time: 3.03 seconds.  
 User-Entered Random Seed: 9846168

Equivalence Tests for Paired Means (Simulation)

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 using Chow et al. (2003)

We will validate this procedure by comparing its results to those of Chow et al. (2003) page 55 in which the parameter values are:  $M_0 = 0$ ,  $M_1 = 0.05$ ,  $\alpha = 0.05$ ,  $N = 35$ ,  $R = 0.0$ , and  $S = 0.070711$ . For these parameters, the power is given as 0.800.

Note that they give the standard deviation of the differences as 0.1. Since the correlation is 0.0, the standard deviation of the individual data values is given by  $0.1/\sqrt{2} = 0.070711$ .

In order to understand the accuracy of the simulation, we will re-run the analysis five times.

For reproducibility, we'll use a random seed of 5041929.

### 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>
Equivalence Limit.....	<b>Symmetric</b>
Test Type.....	<b>T-Test</b>
Simulations .....	<b>2000</b>
Random Seed.....	<b>5041929</b> (for Reproducibility)
Alpha.....	<b>0.05</b>
N (Sample Size).....	<b>35 35 35 35 35</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>0.05</b>
Parameter 1 Label .....	<b>S</b>
Parameter 1 Value(s).....	<b>0.070711</b>
R (Correlation of Items A & B) .....	<b>0.0</b>



## Equivalence Tests for Paired Means (Simulation)

## Output

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

## Numeric Results

Solve For: [Power](#)  
 Hypotheses:  $H_0: \text{Diff} \geq |\text{Diff}_0|$  vs.  $H_1: \text{Diff} < |\text{Diff}_0|$   
 H0 Distribution: Normal(M0 S) - Normal(M1 S)  
 H1 Distribution: Normal(M0 S) - Normal(M0 S)  
 Test Statistic: Paired T-Test

Power	Sample Size N	Paired Difference Diff1	Equivalence Limits		Correlation of A and B R	Alpha		M0	M1	S
			Lower	Upper		Target	Actual			
0.790 (0.018) [0.772 0.808]	35	0	-0.1	0.1	0	0.05	0.045 (0.009) [0.035 0.054]	0	0.1	0.1
0.791 (0.018) [0.773 0.809]	35	0	-0.1	0.1	0	0.05	0.051 (0.01) [0.041 0.061]	0	0.1	0.1
0.791 (0.018) [0.773 0.808]	35	0	-0.1	0.1	0	0.05	0.045 (0.009) [0.036 0.054]	0	0.1	0.1
0.794 (0.018) [0.776 0.811]	35	0	-0.1	0.1	0	0.05	0.050 (0.01) [0.04 0.059]	0	0.1	0.1
0.779 (0.018) [0.76 0.797]	35	0	-0.1	0.1	0	0.05	0.061 (0.01) [0.051 0.071]	0	0.1	0.1

Pool Size: 10000. Simulations: 2000. Run Time: 0.93 seconds.  
 User-Entered Random Seed: 5041929

The powers match the analytic value of 0.800 quite well. Note how informative the confidence intervals are.