Chapter 550

Tests for Paired Means (Simulation)

Introduction

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

Paired data may occur because two measurements are made on the same subject or because measurements are made on two subjects that have been matched according to other, often demographic, variables. Hypothesis tests on paired data can be analyzed by considering the differences between the paired items. The distribution of differences is usually symmetric. In fact, the distribution must be symmetric if the individual distributions of the two items are identical. Hence, the paired t-test and the Wilcoxon signed-rank test are appropriate for paired data even when the distributions of the individual items are not normal.

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

Technical Details

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

The steps to a simulation study are as follows:

1. Specify the method by which the test is to be carried out. This includes specifying how the test statistic is calculated and how the significance level is specified.

2. Generate a random sample, $X_1, X_2, \ldots, X_n$, from the distribution specified by the <u>alternative</u> hypothesis. Calculate the test statistic from the simulated data and determine if the null hypothesis is accepted or rejected. Each of these samples is used to calculate the power of the test.

3. Generate a random sample, $Y_1, Y_2, \ldots, Y_n$, from the distribution specified by the <u>null</u> hypothesis. Calculate the test statistic from the simulated data and determine if the null hypothesis is accepted or rejected. Each of these samples is used to calculate the significance level of the test.

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

Data Distributions

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

Test Statistics

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

Paired T-Test

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

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

where

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

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

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

Wilcoxon Signed-Rank Test

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

1. Subtract the hypothesized difference, $D0$, from each data value. Rank the values according to their absolute values.

2. Compute the sum of the positive ranks Sp and the sum of the negative ranks Sn. The test statistic, W_R, is the minimum of Sp and Sn.

3. Compute the mean and standard deviation of W_R using the formulas

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

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

where t represents the number of times the ith 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 comparing the z value to a normal probability table. If this p-value is less than a specified level (often 0.05), the null hypothesis is rejected. Otherwise, no conclusion can be reached.

Sign Test

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

1. Count the number of values strictly greater than $D0$. Call this value X.

2. Count the number of values strictly less than $D0$. Call this value Y.

3. Set $m = X + Y$.

4. Under the null hypothesis, X is distributed as a binomial random variable with a proportion of 0.5 and sample size of m.

The significance of X is calculated using binomial probabilities.

Bootstrap Test

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

1. Compute the mean of the sample. Call it $\bar{X}$.

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

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

where $D0$ is the hypothesized difference.

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

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

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

5. For a two-tailed test, if $|t_Y| > |t_X|$ then add one to a counter variable, A.

6. Repeat steps 3 – 5 B times. B may be anywhere from 100 to 10,000.

7. Compute the p-value of the bootstrap test as $(A + 1) / (B + 1)$

8. Steps 1 – 7 complete one simulation iteration. Repeat these steps M times, where M is the number of simulations. The power and significance level are equal to the percent of the time the p-value is less than the nominal alpha of the test in their respective simulations.

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

Standard Deviations

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

The Standard Deviation of Paired Differences (σ) when using Simple Input

If you have results from a previous (or pilot) study, use the estimate of the standard deviation of paired differences, σ, from the study. Another reasonable (but somewhat rough) estimate of σ may be obtained using the range of paired differences as

$$\sigma = \frac{Range}{4}$$

If you have estimates of the expected standard deviations of the paired variables (σ_1 and σ_2) and the Pearson correlation between the paired variables (ρ), the standard deviation of paired differences (σ) may be calculated using the equation

$$\sigma^2 = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2$$

such that

$$\sigma = \sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}\,.$$

If $\sigma_1 = \sigma_2 = \sigma_x$, then this formula reduces to

$$\sigma^2 = 2\sigma_x^2(1 - \rho)$$

such that

$$\sigma = \sqrt{2\sigma_x^2(1 - \rho)}\,.$$

If you have an estimate of the within-subject population standard deviation (σ_w), then σ may be calculated using the equation

$$\sigma^2 = 2\sigma_w^2$$

such that

$$\sigma = \sqrt{2\sigma_w^2}\,.$$

σ_w is often estimated by the square root of the within mean square error (WMSE) from a repeated measures ANOVA.

Example 1 – Power at Various Sample Sizes

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

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

Setup

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 (a or b)** 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 ....**Power**
Alternative Hypothesis ....**Two-Sided (H1: δ ≠ δ0)**
Test Type..**T-Test**
Simulations **2000**
Random Seed..**5379518** (for Reproducibility)
Alpha...**0.05**
N (Sample Size)..**50 100 150**
Input Type..**Simple**
Distribution to Simulate ....**Normal**
δ0 (Null Mean of Paired Diffs).........................**0**
δ1 (Actual Mean of Paired Diffs)....................**0.6**
Standard Deviation Input Type ....**Enter the SD of Paired Differences**
σ (SD of Paired Differences)..........................**2.53**

Output

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

Numeric Results

Numeric Results

Solve For: Power
Hypotheses: H0: δ = δ0   vs.   H1: δ ≠ δ0
Test Statistic: Paired T-Test
Simulated Distribution: Normal

Power	Sample Size N	Null δ0	Actual δ1	Difference δ1 - δ0	Standard Deviation σ	Effect Size |δ1 - δ0| / σ	Alpha Target	Alpha Actual
0.4015 (0.02148) [0.38002 0.42298]	50	0	0.6	0.6	2.53	0.23715	0.05	0.048 (0.009) [0.038 0.057]
0.6365 (0.02108) [0.61542 0.65758]	100	0	0.6	0.6	2.53	0.23715	0.05	0.046 (0.009) [0.036 0.055]
0.8290 (0.0165) [0.8125 0.8455]	150	0	0.6	0.6	2.53	0.23715	0.05	0.054 (0.01) [0.044 0.064]

Simulations: 2000. Run Time: 2.55 seconds.
User-Entered Random Seed: 5379518

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.
δ0	The value of the population mean of paired differences under the null hypothesis.
δ1	The actual value of the population mean of paired differences at which power and sample size are calculated.
δ1 - δ0	The difference between the actual and null mean of paired differences.
σ	The standard deviation of paired differences for the population. It measures the variability in the population.
|δ1 - δ0| / σ	The Effect Size, which is the relative magnitude of the effect.
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.

Summary Statements

A paired design will be used to test whether the paired mean difference (δ) is different from 0 (H0: δ = 0 versus H1: δ ≠ 0). The comparison will be made using a two-sided, paired-sample t-test of the paired differences, with a Type I error rate (α) of 0.05. The standard deviation of the paired differences is assumed to be 2.53. To detect a paired mean difference of 0.6 with a sample size of 50 pairs, the power is 0.4015. These results are based on 2000 simulations (Monte Carlo samples) from the Normal distribution.

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 Lokhnygina, 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., Wang, H., and Lokhnygina, Y. 2018. Sample Size Calculations in Clinical Research, Third Edition. Taylor & Francis/CRC. Boca Raton, Florida.

Devroye, Luc. 1986. Non-Uniform Random Variate Generation. Springer-Verlag. New York.

Machin, D., Campbell, M., Fayers, P., and Pinol, A. 1997. Sample Size Tables for Clinical Studies, 2nd Edition. Blackwell Science. Malden, MA.

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

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

Plots Section
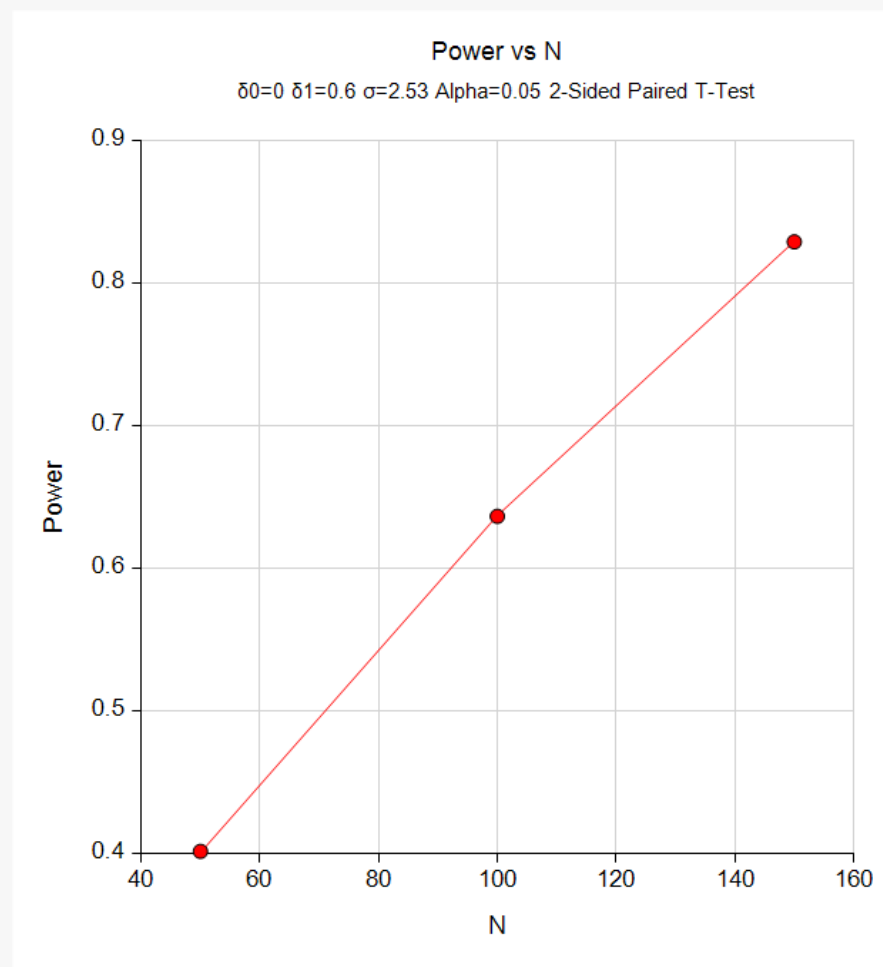
Plots
_____

This plot shows the relationship between sample size and power.

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 ....**Sample Size**
Alternative Hypothesis ....**Two-Sided (H1: δ ≠ δ0)**
Test Type...**T-Test**
Simulations **2000**
Random Seed..**2919111** (for Reproducibility)
Power..**0.90**
Alpha...**0.05**
Input Type..**Simple**
Distribution to Simulate ....**Normal**
δ0 (Null Mean of Paired Diffs).......................**0**
δ1 (Actual Mean of Paired Diffs)...................**0.6**
Standard Deviation Input Type ....**Enter the SD of Paired Differences**
σ (SD of Paired Differences).........................**2.53**

Output

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

Numeric Results

Solve For: Sample Size
Hypotheses: H0: δ = δ0   vs.   H1: δ ≠ δ0
Test Statistic: Paired T-Test
Simulated Distribution: Normal

| | | Mean of Paired Differences | | | Standard Deviation | Effect Size | Alpha | |
| | Sample Size | Null | Actual | Difference | | | | |
Power	N	δ0	δ1	δ1 - δ0	σ	|δ1 - δ0| / σ	Target	Actual
0.891	187	0	0.6	0.6	2.53	0.23715	0.05	0.046
(0.01366) [0.87734 0.90466]								(0.009) [0.037 0.055]

Simulations: 2000. Run Time: 12.54 seconds.
User-Entered Random Seed: 2919111

The required sample size of 187 achieved a power of 0.891. The power of 0.891 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

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 ....**Power**
Alternative Hypothesis ....**Two-Sided (H1: δ ≠ δ0)**
Test Type..**T-Test**
Simulations **2000**
Random Seed...**4985023** (for Reproducibility)
Alpha..**0.05**
N (Sample Size)...**50 100 150 200**
Input Type..**Simple**
Distribution to Simulate ....**Normal**
δ0 (Null Mean of Paired Diffs).......................**0**
δ1 (Actual Mean of Paired Diffs)...................**0.6**
Standard Deviation Input Type ....**Enter the SD of Paired Differences**
σ (SD of Paired Differences).........................**2.53**

Reports Tab

Show Comparative Reports...........................**Checked**
Include T-Test Results..................................**Checked**
Include Wilcoxon & Sign Test Results ....**Checked**

Comparative Plots Tab

Show Comparative Plots................................**Checked**

Output

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

Power Comparison

Hypotheses: H0: δ = δ0   vs.   H1: δ ≠ δ0
Simulated Distribution: Normal

Sample Size N	Mean of Paired Differences		Standard Deviation σ	Target Alpha	Power		
	Null δ0	Actual δ1			T-Test	Wilcoxon	Sign
50	0	0.6	2.53	0.05	0.3505	0.3310	0.1930
100	0	0.6	2.53	0.05	0.6495	0.6310	0.4130
150	0	0.6	2.53	0.05	0.8215	0.8030	0.6015
200	0	0.6	2.53	0.05	0.9080	0.8935	0.7185

Simulations: 2000. Run Time: 4.15 seconds.
User-Entered Random Seed: 4985023

Alpha Comparison

Hypotheses: H0: δ = δ0   vs.   H1: δ ≠ δ0
Simulated Distribution: Normal

Sample Size N	Mean of Paired Differences		Standard Deviation σ	Alpha			
	Null δ0	Actual δ1		Target	T-Test	Wilcoxon	Sign
50	0	0.6	2.53	0.05	0.053	0.051	0.035
100	0	0.6	2.53	0.05	0.047	0.046	0.035
150	0	0.6	2.53	0.05	0.053	0.053	0.048
200	0	0.6	2.53	0.05	0.058	0.054	0.040

Simulations: 2000. Run Time: 4.15 seconds.
User-Entered Random Seed: 4985023

Comparative Plots

Power vs N by Test
δ0=0 δ1=0.6 σ=2.53 Alpha=0.05 2-Sided Paired T-Test
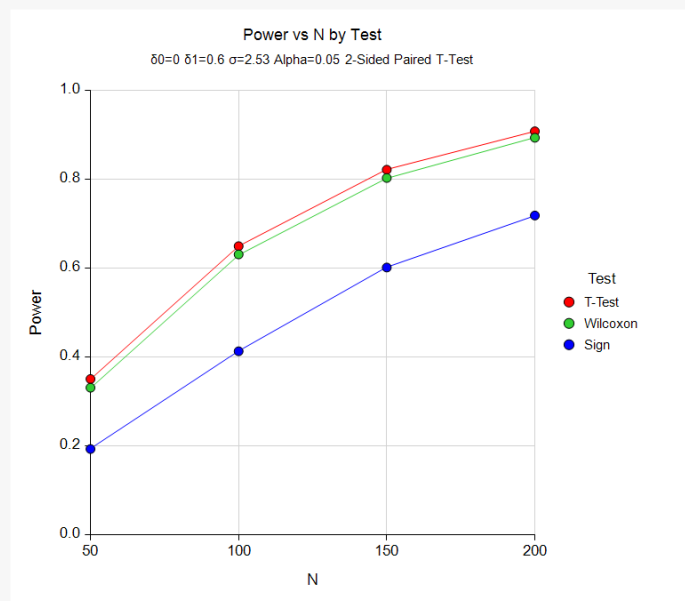
These results show that for paired data, the t-test and Wilcoxon test have very similar power and alpha values. The sign test is both less accurate and less powerful.

Example 4 – Validation using Zar (1984)

Zar (1984), pages 111-112, presents an example in which δ0 = 0.0, δ1 = 1.0, σ = 1.25, alpha = 0.05, and N = 12. Zar obtains an approximate power of 0.72. We will validate this procedure by running this example. To make certain that the results are very accurate, the number of simulations will be set to 10,000.

For reproducibility, we'll use a random seed of 6015683.

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 ....**Power**
Alternative Hypothesis ....**Two-Sided (H1: δ ≠ δ0)**
Test Type..**T-Test**
Simulations **10000**
Random Seed...**6015683** (for Reproducibility)
Alpha..**0.05**
N (Sample Size)..**12**
Input Type...**Simple**
Distribution to Simulate.................................**Normal**
δ0 (Null Mean of Paired Diffs).......................**0**
δ1 (Actual Mean of Paired Diffs)...................**1**
Standard Deviation Input Type ....**Enter the SD of Paired Differences**
σ (SD of Paired Differences)..........................**1.25**

Output

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

Numeric Results

Solve For: Power
Hypotheses: H0: δ = δ0   vs.   H1: δ ≠ δ0
Test Statistic: Paired T-Test
Simulated Distribution: Normal

| | | Mean of Paired Differences | | | Standard Deviation | Effect Size | Alpha | |
| | Sample Size | Null | Actual | Difference | | | | |
Power	N	δ0	δ1	δ1 - δ0	σ	$|δ1 - δ0| / σ$	Target	Actual
0.7162	12	0	1	1	1.25	0.8	0.05	0.049
(0.00884) [0.70736 0.72504]								(0.004) [0.045 0.053]

Simulations: 10000. Run Time: 1.43 seconds.
User-Entered Random Seed: 6015683

This simulation obtained a power of 0.7162 which rounds to the 0.72 computed by Zar.

Example 5 – Validation using Machin (1997)

Machin, et. al. (1997), page 37, presents an example in which μ0 = 0.0, μ1 = 0.2, σ = 1.0, alpha = 0.05, and power = 0.80. They obtain a sample size of 199.

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 ....**Sample Size**
Alternative Hypothesis ....**Two-Sided (H1: δ ≠ δ0)**
Test Type..**T-Test**
Simulations **2000**
Random Seed...**6030438** (for Reproducibility)
Power...**0.80**
Alpha..**0.05**
Input Type...**Simple**
Distribution to Simulate ....**Normal**
δ0 (Null Mean of Paired Diffs).......................**0**
δ1 (Actual Mean of Paired Diffs)....................**0.2**
Standard Deviation Input Type ....**Enter the SD of Paired Differences**
σ (SD of Paired Differences)..........................**1**

Output

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

Numeric Results
_____

Solve For: Sample Size
Hypotheses: H0: δ = δ0   vs.   H1: δ ≠ δ0
Test Statistic: Paired T-Test
Simulated Distribution: Normal
_____

| | Sample Size | Mean of Paired Differences | | | Standard Deviation | Effect Size | Alpha | |
Power	N	Null δ0	Actual δ1	Difference δ1 - δ0	σ	|δ1 - δ0| / σ	Target	Actual
0.815	199	0	0.2	0.2	1	0.2	0.05	0.046
(0.01702) [0.79798 0.83202]								(0.009) [0.036 0.055]

_____

Simulations: 2000. Run Time: 13.01 seconds.
User-Entered Random Seed: 6030438

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

Example 6 – Non-Inferiority Test

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

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

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

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

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

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

Setup

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 6** 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 ....**Power**
Alternative Hypothesis ....**One-Sided (H1: δ > δ0)**
Test Type..**T-Test**
Simulations **2000**
Random Seed...**7466448** (for Reproducibility)
Alpha..**0.025**
N (Sample Size)...**5 10 15 20 25**
Input Type...**Simple**
Distribution to Simulate.................................**Normal**
δ0 (Null Mean of Paired Diffs)........................**-5**
δ1 (Actual Mean of Paired Diffs)....................**0**
σ (Std Dev of Paired Diffs)............................**6.32**

Output

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

Numeric Results
_____

Solve For: Power
Hypotheses: H0: δ ≤ δ0   vs.   H1: δ > δ0
Test Statistic: Paired T-Test
Simulated Distribution: Normal
_____

| | | **Mean of Paired Differences** | | | Standard Deviation | Effect Size | **Alpha** | |
Power	Sample Size N	Null δ0	Actual δ1	Difference δ1 - δ0	σ	|δ1 - δ0| / σ	Target	Actual
0.2880 (0.01985) [0.26815 0.30785]	5	-5	0	5	6.32	0.79114	0.025	0.026 (0.007) [0.019 0.033]
0.6115 (0.02136) [0.59014 0.63286]	10	-5	0	5	6.32	0.79114	0.025	0.020 (0.006) [0.013 0.026]
0.7985 (0.01758) [0.78092 0.81608]	15	-5	0	5	6.32	0.79114	0.025	0.029 (0.007) [0.022 0.036]
0.9210 (0.01182) [0.90918 0.93282]	20	-5	0	5	6.32	0.79114	0.025	0.027 (0.007) [0.02 0.034]
0.9675 (0.00777) [0.95973 0.97527]	25	-5	0	5	6.32	0.79114	0.025	0.029 (0.007) [0.022 0.036]

Simulations: 2000. Run Time: 1.50 seconds.
User-Entered Random Seed: 7466448

Plots
_____

Power vs N
δ0=-5 δ1=0 σ=6.32 Alpha=0.025 1-Sided Paired T-Test
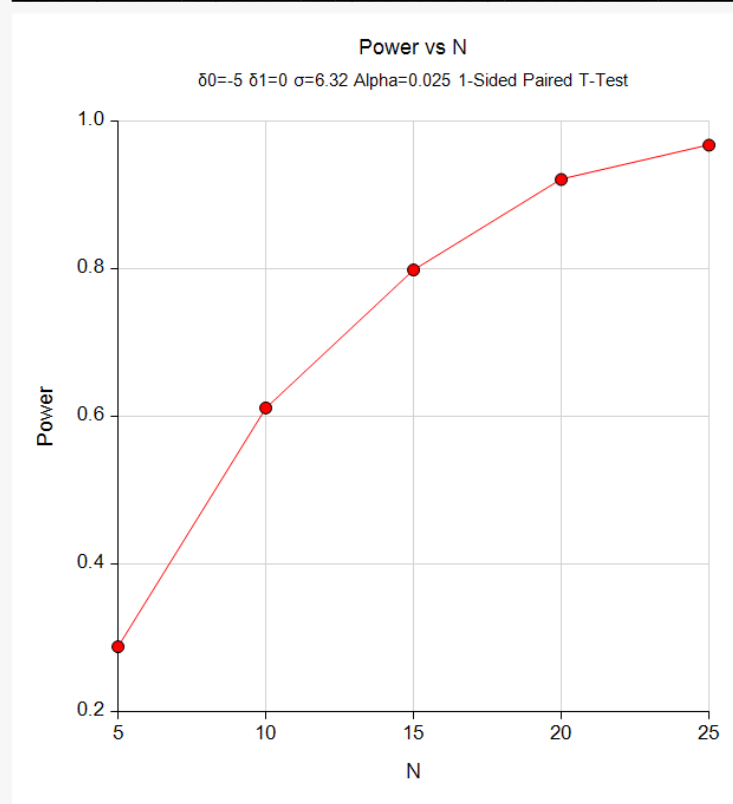
We see that a power of 0.8 is achieved at about 15 subjects, while a power of 0.9 requires about 20 subjects.