

Chapter 493

Paired Wilcoxon Signed-Rank Tests

Introduction

The paired t -test may be used to test whether the mean difference of two populations is greater than, less than, or not equal to 0. Because the t distribution is used to calculate critical values for the test, this test is often called the paired t -test. The paired t -test assumes that the population standard deviation of paired differences is unknown and will be estimated by the data.

The nonparametric analog of the t -test is the *Wilcoxon Signed-Rank Test* and may be used when the one-sample t -test assumptions are violated.

Other PASS Procedures for Testing One Mean or Median from Paired Data

Procedures in **PASS** are primarily built upon the testing methods, test statistic, and test assumptions that will be used when the analysis of the data is performed. You should check to identify that the test procedure described below in the Test Procedure section matches your intended procedure. If your assumptions or testing method are different, you may wish to use one of the other one-sample paired-data procedures available in **PASS**—the Paired Z-Tests and the Paired T-Tests. The methods, statistics, and assumptions for those procedures are described in the associated chapters.

If you wish to show that the mean of a population is larger (or smaller) than a reference value by a specified amount, you should use one of the clinical superiority procedures for comparing means. Non-inferiority, equivalence, and confidence interval procedures are also available.

Assumptions for Paired Tests

This section describes the assumptions that are made when you use one of these tests. The key assumption relates to normality or non-normality of the data. One of the reasons for the popularity of the t -test is its robustness in the face of assumption violation. However, if the assumptions are not met, the significance levels and the power of the t -test may be invalidated. Unfortunately, in practice it often happens that several assumptions are not met. Take the steps to check the assumptions before you make important decisions based on these tests.

Paired Z-Test Assumptions

The assumptions of the paired z -test are:

1. The data are continuous (not discrete).
2. The data, i.e., the differences for the matched-pairs, follow a normal probability distribution.

Paired Wilcoxon Signed-Rank Tests

3. The sample of pairs is a simple random sample from its population. Each individual in the population has an equal probability of being selected in the sample.
4. The population standard deviation of paired differences is known.

Paired T-Test Assumptions

The assumptions of the paired t -test are:

1. The data are continuous (not discrete).
2. The data, i.e., the differences for the matched-pairs, follow a normal probability distribution.
3. The sample of pairs is a simple random sample from its population. Each individual in the population has an equal probability of being selected in the sample.

Wilcoxon Signed-Rank Test Assumptions

The assumptions of the Wilcoxon signed-rank test are as follows (note that the difference is between a data value and the hypothesized median or between the two data values of a pair):

1. The differences are continuous (not discrete).
2. The distribution of each difference is symmetric.
3. The differences are mutually independent.
4. The differences all have the same median.
5. The measurement scale is at least interval.

Limitations

There are few limitations when using these tests. Sample sizes may range from a few to several hundred. If your data are discrete with at least five unique values, you can often ignore the continuous variable assumption. Perhaps the greatest restriction is that your data come from a random sample of the population. If you do not have a random sample, your significance levels will probably be incorrect.

Paired Wilcoxon Signed-Rank Test Statistic

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. Rank the paired differences 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^{th} value occurs.

Paired Wilcoxon Signed-Rank Tests

4. Compute the z -value using

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

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.

Power Calculation for the Paired Wilcoxon Signed-Rank Test

The power calculation for the Wilcoxon signed-rank test is the same as that for the paired t -test except that an adjustment is made to the sample size based on an assumed data distribution as described in Al-Sundugchi and Guenther (1990). The sample size n' used in power calculations is equal to

$$n' = n/W,$$

where W is the Wilcoxon adjustment factor based on the assumed data distribution.

The adjustments are as follows:

Distribution	W
Uniform	1
Double Exponential	2/3
Logistic	$9/\pi^2$
Normal	$\pi/3$

The power is calculated as follows for a directional alternative (one-tailed test) in which $\delta_1 > 0$.

1. Find t_α such that $1 - T_{df}(t_\alpha) = \alpha$, where $T_{df}(t_\alpha)$ is the area under a central- t curve to the left of x and $df = n' - 1$.
2. Calculate: $X_1 = t_\alpha \frac{\sigma}{\sqrt{n'}}$.
3. Calculate the noncentrality parameter: $\lambda = \frac{\delta_1}{\frac{\sigma}{\sqrt{n'}}}$.
4. Calculate: $t_1 = \frac{X_1 - \delta_1}{\frac{\sigma}{\sqrt{n'}}} + \lambda$.
5. Power = $1 - T'_{df,\lambda}(t_1)$, where $T'_{df,\lambda}(x)$ is the area to the left of x under a noncentral- t curve with degrees of freedom df and noncentrality parameter λ .

Procedure Options

This section describes the options that are specific to this procedure. These are located on the Design tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Design Tab

The Design tab contains most of the parameters and options that you will be concerned with.

Solve For

Solve For

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

Select *Sample Size* when you want to determine the sample size needed to achieve a given power and alpha error level.

Select *Power* when you want to calculate the power of an experiment that has already been run.

Test

Alternative Hypothesis

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

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

Possible selections are:

- **Two-Sided ($H_1: \delta \neq 0$)**
This is the most common selection. It yields the *two-tailed t-test*. Use this option when you do not want to specify beforehand the direction of the test. Many scientific journals require two-tailed tests.
- **One-Sided ($H_1: \delta < 0$)**
This option yields a *one-tailed t-test*. Use it when you are only interested in the case in which δ is less than 0.
- **One-Sided ($H_1: \delta > 0$)**
This option yields a *one-tailed t-test*. Use it when you are only interested in the case in which δ is greater than 0.

Paired Wilcoxon Signed-Rank Tests

Data Distribution

This option makes appropriate sample size adjustments for the Wilcoxon Signed-Rank test. Results by Al-Sundugchi and Guenther (1990) indicate that power calculations for the Wilcoxon Signed-Rank test may be made using the standard t -test formulations with a simple adjustment to the sample size. The size of the adjustment depends upon the actual distribution of the data. They give sample size adjustment factors for four distributions.

The options are as follows:

- **Uniform**

The sample size adjustment factor, W , is equal to “1”. This selection gives the same result as the one-sample t -test.

- **Double Exponential**

The sample size adjustment factor, W , is equal to “2/3”.

- **Logistic**

The sample size adjustment factor, W , is equal to “ $9/\pi^2$ ”.

- **Normal**

The sample size adjustment factor, W , is equal to “ $\pi/3$ ”.

Population Size

This is the number of subjects in the population. Usually, you assume that samples are drawn from a very large (infinite) population. Occasionally, however, situations arise in which the population of interest is of limited size. In these cases, appropriate adjustments must be made.

When a finite population size is specified, the standard deviation is reduced according to the formula:

$$\sigma_1^2 = \left(1 - \frac{n}{N}\right) \sigma^2$$

where n is the sample size, N is the population size, σ is the original standard deviation, and σ_1 is the new standard deviation.

The quantity n/N is often called the sampling fraction. The quantity $\left(1 - \frac{n}{N}\right)$ is called the *finite population correction factor*.

Power and Alpha

Power

This option specifies one or more values for power. Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as *0.8 to 0.95 by 0.05* may be entered.

Paired Wilcoxon Signed-Rank Tests

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

Effect Size

δ_1 (Mean of Paired Differences)

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

σ (Std Dev of Paired Differences)

This option specifies one or more values of the standard deviation. This must be a positive value. Be sure to use the standard deviation of the paired differences and not the standard deviation of the mean paired difference (the standard error).

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

Paired Wilcoxon Signed-Rank Tests

Example 1 – Computing Power

Usually, a researcher designs a study to compare two or more groups of subjects, so the one sample case described in this chapter occurs infrequently. However, there is a popular research design that does lead to the single mean test: *paired observations*.

For example, suppose researchers want to study the impact of an exercise program on the individual's weight. To do so they randomly select N individuals, weigh them, put them through the exercise program, and weigh them again. The variable of interest is not their actual weight, but how much their weight changed.

In this design, the data will be analyzed using a Wilcoxon Signed-Rank test on the differences between the paired observations. The null hypothesis is that the average difference is zero. The alternative hypothesis is that the average difference is some nonzero value.

To study the impact of an exercise program on weight loss, the researchers decide to conduct a study that will be analyzed using the paired test. A sample of individuals will be weighed before and after a specified exercise program that will last three months. The difference in their weights will be analyzed.

Past experiments of this type have had standard deviations in the range of 10 to 15 pounds. The researcher wants to detect a difference of 5 pounds or more with an alpha of 0.05. What is the power for sample sizes between 30 and 100 if we assume a normal distribution for the paired differences?

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Paired Wilcoxon Signed-Rank Tests** procedure window by expanding **Means**, then **Paired Means**, then clicking on **Nonparametric**, and then clicking on **Paired Wilcoxon Signed-Rank Tests**. 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**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Power
Alternative Hypothesis	Two-Sided (H1: $\delta \neq 0$)
Data Distribution	Normal
Population Size.....	Infinite
Alpha.....	0.05
N (Sample Size).....	30 to 100 by 10
δ_1 (Mean of Paired Differences).....	-5
σ (Std Dev of Paired Differences).....	10 12.5 15

Paired Wilcoxon Signed-Rank Tests

Annotated Output

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

Numeric Results and Plots

Numeric Results

Hypotheses: $H_0: \delta = 0$ vs. $H_1: \delta \neq 0$

Data Distribution: Normal

Power	N	δ_1	σ	Effect Size	Alpha	Beta
0.72266	30	-5.0	10.0	0.500	0.050	0.27734
0.85114	40	-5.0	10.0	0.500	0.050	0.14886
0.91856	50	-5.0	10.0	0.500	0.050	0.08144
0.95989	60	-5.0	10.0	0.500	0.050	0.04011
0.97939	70	-5.0	10.0	0.500	0.050	0.02061
0.99042	80	-5.0	10.0	0.500	0.050	0.00958
0.99529	90	-5.0	10.0	0.500	0.050	0.00471
0.99790	100	-5.0	10.0	0.500	0.050	0.00210
0.53234	30	-5.0	12.5	0.400	0.050	0.46766
0.67055	40	-5.0	12.5	0.400	0.050	0.32945
0.76567	50	-5.0	12.5	0.400	0.050	0.23433
0.84322	60	-5.0	12.5	0.400	0.050	0.15678
0.89275	70	-5.0	12.5	0.400	0.050	0.10725
0.93085	80	-5.0	12.5	0.400	0.050	0.06915
0.95404	90	-5.0	12.5	0.400	0.050	0.04596
0.97119	100	-5.0	12.5	0.400	0.050	0.02881
0.39800	30	-5.0	15.0	0.333	0.050	0.60200
0.51654	40	-5.0	15.0	0.333	0.050	0.48346
0.60926	50	-5.0	15.0	0.333	0.050	0.39074
0.69613	60	-5.0	15.0	0.333	0.050	0.30387
0.76048	70	-5.0	15.0	0.333	0.050	0.23952
0.81821	80	-5.0	15.0	0.333	0.050	0.18179
0.85946	90	-5.0	15.0	0.333	0.050	0.14054
0.89537	100	-5.0	15.0	0.333	0.050	0.10463

References

- Al-Sunduqchi, Mahdi S. 1990. Determining the Appropriate Sample Size for Inferences Based on the Wilcoxon Statistics. Ph.D. dissertation under the direction of William C. Guenther, Dept. of Statistics, University of Wyoming, Laramie, Wyoming.
- Chow, S.C., Shao, J., Wang, H., and Lohknygina, Y. 2018. Sample Size Calculations in Clinical Research, Third Edition. Taylor & Francis/CRC. Boca Raton, Florida.
- Machin, D., Campbell, M., Fayers, P., and Pinol, A. 1997. Sample Size Tables for Clinical Studies, 2nd Edition. Blackwell Science. Malden, MA.
- Zar, Jerrold H. 1984. Biostatistical Analysis (Second Edition). Prentice-Hall. Englewood Cliffs, New Jersey.

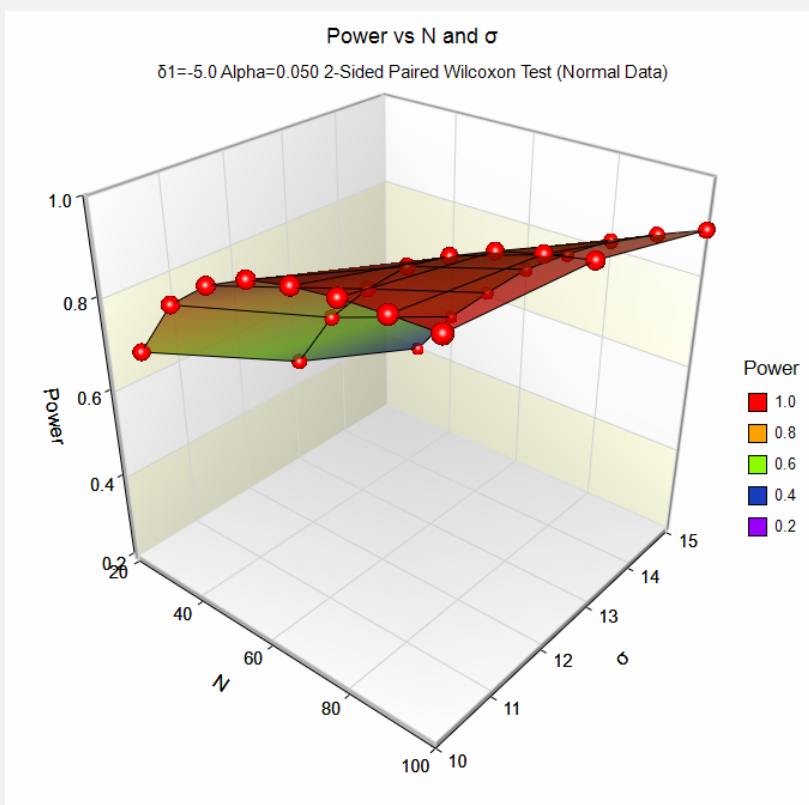
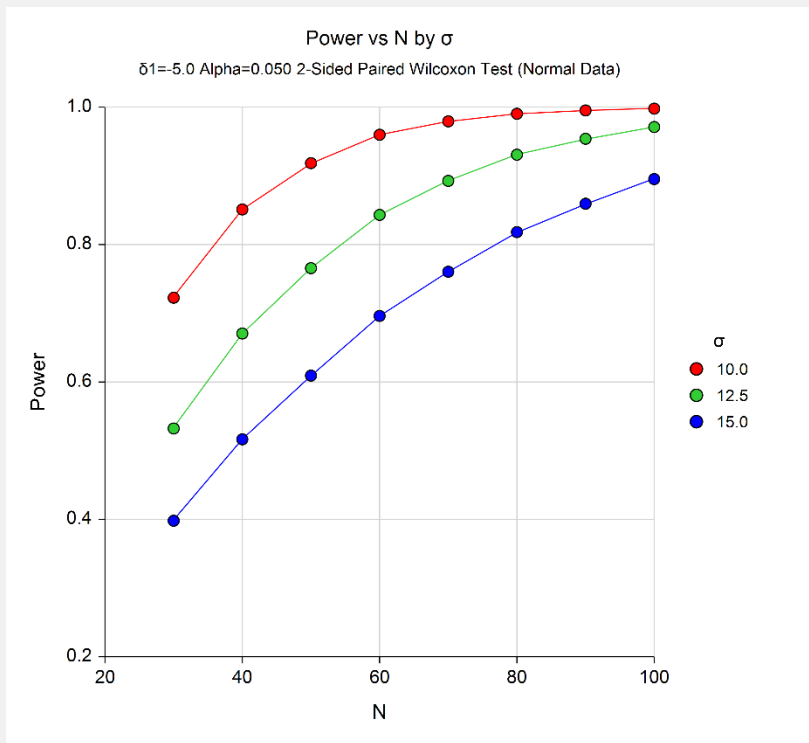
Report Definitions

- Power is the probability of rejecting a false null hypothesis. It should be close to one.
- N is the sample size, the number of subjects (or pairs) in the study.
- δ is the mean of paired differences.
- δ_1 is the value of the mean of paired differences at which power and sample size are calculated.
- σ is the standard deviation of paired differences for the population.
- Effect Size = $|\delta_1|/\sigma$ is the relative magnitude of the effect.
- Alpha is the probability of rejecting a true null hypothesis. It should be small.
- Beta is the probability of accepting a false null hypothesis. It should be small.

Summary Statements

A sample size of 30 achieves 72% power to detect a mean of paired differences of -5.0 with an estimated standard deviation of paired differences of 10.0 and with a significance level (alpha) of 0.050 using a two-sided paired Wilcoxon Signed-Rank test assuming that the actual distribution of paired differences is normal.

Paired Wilcoxon Signed-Rank Tests



These plots show the relationship between sample size and power for various values of alpha and σ .

Paired Wilcoxon Signed-Rank Tests

Example 2 – Finding the Sample Size

Continuing with Example 1, how many pairs are required for each scenario to achieve 80% power?

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Paired Wilcoxon Signed-Rank Tests** procedure window by expanding **Means**, then **Paired Means**, then clicking on **Nonparametric**, and then clicking on **Paired Wilcoxon Signed-Rank Tests**. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Sample Size
Alternative Hypothesis	Two-Sided (H1: $\delta \neq 0$)
Data Distribution	Normal
Population Size	Infinite
Power	0.80
Alpha	0.05
δ_1 (Mean of Paired Differences)	-5
σ (Std Dev of Paired Differences)	10 12.5 15

Output

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

Numeric Results and Plots

Numeric Results

Hypotheses: $H_0: \delta = 0$ vs. $H_1: \delta \neq 0$
Data Distribution: Normal

Power	N	δ_1	σ	Effect Size	Alpha	Beta
0.80778	36	-5.0	10.0	0.500	0.050	0.19222
0.80779	55	-5.0	12.5	0.400	0.050	0.19221
0.80230	77	-5.0	15.0	0.333	0.050	0.19770

The required sample sizes for each scenario are displayed.

Paired Wilcoxon Signed-Rank Tests

Example 3 – Validation using Chow, Shao, Wang, and Lokhnygina (2018)

Chow, Shao, Wang, and Lokhnygina (2018) presents an example on pages 45 and 46 of a two-sided one-sample t -test sample size calculation in which $\mu_0 = 1.5$, $\mu_1 = 2.0$, $\sigma = 1.0$, $\alpha = 0.05$, and $\text{power} = 0.80$. They obtain a sample size of 34.

The Paired Wilcoxon Signed-Rank test power calculations are the same as the paired t -test except for an adjustment factor for the assumed data distribution. If we set $\delta_1 = 2.0 - 1.5 = 0.5$ and assume a uniform data distribution, we should get this same result because the one-sample t -test and the paired t -test use the same fundamental calculations. If we set the data distribution to Normal, we should get a result of $N = 34 \times \pi/3 = 35.605 = 36$.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Paired Wilcoxon Signed-Rank Tests** procedure window by expanding **Means**, then **Paired Means**, then clicking on **Nonparametric**, and then clicking on **Paired Wilcoxon Signed-Rank Tests**. You may then make the appropriate entries as listed below, or open **Example 3 (a or b)** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Sample Size
Alternative Hypothesis	Two-Sided (H1: $\delta \neq 0$)
Data Distribution	Uniform
Population Size	Infinite
Power	0.80
Alpha	0.05
δ_1 (Mean of Paired Differences)	0.5
σ (Std Dev of Paired Differences)	1

Output

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

Numeric Results						
Hypotheses: $H_0: \delta = 0$ vs. $H_1: \delta \neq 0$						
Data Distribution: Uniform						
Power	N	δ_1	σ	Effect Size	Alpha	Beta
0.80778	34	0.5	1.0	0.500	0.050	0.19222

The sample size of 34 matches Chow, Shao, Wang, and Lokhnygina (2018) exactly. If we now assume a Normal data distribution, the results match our expected outcome.

Numeric Results						
Hypotheses: $H_0: \delta = 0$ vs. $H_1: \delta \neq 0$						
Data Distribution: Normal						
Power	N	δ_1	σ	Effect Size	Alpha	Beta
0.80778	36	0.5	1.0	0.500	0.050	0.19222

The sample size of 36 matches the expected result.

Paired Wilcoxon Signed-Rank Tests

Example 4 – Validation using Zar (1984)

Zar (1984) pages 111-112 presents an example in which $\delta_1 = 1.0$, $\sigma = 1.25$, $\alpha = 0.05$, and $N = 12$. Zar obtains an approximate power of 0.72.

The paired Wilcoxon Signed-Rank test power calculations are the same as the paired t -test except for an adjustment factor for the assumed data distribution. If we assume a uniform data distribution, we should get the same power with $N = 12$. If we assume a Double Exponential data distribution with $N = 12 \times 2/3 = 8$, then the adjusted sample size would be $N' = 8/(2/3) = 12$ and we should get the same power as Zar (1984).

Setup

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

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Power
Alternative Hypothesis	Two-Sided (H1: $\delta \neq 0$)
Data Distribution	Uniform
Population Size.....	Infinite
Alpha.....	0.05
N (Sample Size).....	12
δ_1 (Mean of Paired Differences).....	1
σ (Std Dev of Paired Differences).....	1.25

Output

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

Numeric Results						
Hypotheses: $H_0: \delta = 0$ vs. $H_1: \delta \neq 0$						
Data Distribution: Uniform						
Power	N	δ_1	σ	Effect Size	Alpha	Beta
0.71366	12	1.0	1.3	0.800	0.050	0.28634

The difference between the power computed by **PASS** of 0.71366 and the 0.72 computed by Zar is due to Zar's use of an approximation to the noncentral t distribution. Notice that the results for the paired Wilcoxon signed-rank test with the uniform distribution match the paired t -test results.

If we now assume a double exponential data distribution and set $N = 8$, the results match our expected outcome.

Numeric Results						
Hypotheses: $H_0: \delta = 0$ vs. $H_1: \delta \neq 0$						
Data Distribution: Double Exponential						
Power	N	δ_1	σ	Effect Size	Alpha	Beta
0.71366	8	1.0	1.3	0.800	0.050	0.28634

The power of 0.71366 matches the expected result.

Paired Wilcoxon Signed-Rank Tests

Example 5 – Validation using Machin (1997)

Machin, Campbell, Fayers, and Pinol (1997) page 37 presents an example in which $\delta_1 = 0.2$, $\sigma = 1.0$, $\alpha = 0.05$, and $\beta = 0.20$. They obtain a sample size of 199.

As in Example 4, if we set the data distribution to Uniform, we should get this same result. If we set the data distribution to double exponential, we should get a result of $N = 199 \times 2/3 = 132.6667 = 133$.

Setup

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

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Sample Size
Alternative Hypothesis	Two-Sided (H1: $\delta \neq 0$)
Data Distribution	Uniform
Population Size	Infinite
Power	0.80
Alpha	0.05
δ_1 (Mean of Paired Differences)	0.2
σ (Std Dev of Paired Differences)	1

Output

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

Numeric Results						
Hypotheses: $H_0: \delta = 0$ vs. $H_1: \delta \neq 0$						
Data Distribution: Uniform						
Power	N	δ_1	σ	Effect Size	Alpha	Beta
0.80169	199	0.2	1.0	0.200	0.050	0.19831

The sample size of 199 matches Machin's result. If we now assume a double exponential data distribution, the results match our expected outcome.

Numeric Results						
Hypotheses: $H_0: \delta = 0$ vs. $H_1: \delta \neq 0$						
Data Distribution: Double Exponential						
Power	N	δ_1	σ	Effect Size	Alpha	Beta
0.80169	133	0.2	1.0	0.200	0.050	0.19831

The sample size of 133 matches the expected result.