

## Chapter 502

# Paired Wilcoxon Signed-Rank Tests for Superiority by a Margin

---

### Introduction

This procedure computes power and sample size for superiority by a margin tests in paired-data designs in which the paired  $t$ -test assumptions are violated and is the *Paired Wilcoxon Signed-Rank Test* is used.

---

### Paired Designs

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 variables. Hypothesis tests on paired data can be analyzed by considering the difference between the paired items as the response. 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 is appropriate for paired data even when the distributions of the individual items are not normal.

---

### The Statistical Hypotheses

Both non-inferiority and superiority tests are examples of directional (one-sided) tests and their power and sample size could be calculated using the One-Sample T-Tests procedure. However, at the urging of our users, we have developed this procedure which provides the input and output options that are convenient for superiority hypotheses. This section will review the specifics of superiority testing.

Remember that in the usual  $t$ -test setting with  $\delta$  defined as the mean paired difference, the null ( $H_0$ ) and alternative ( $H_1$ ) hypotheses for one-sided upper-tail tests are defined as

$$H_0: \delta \leq \delta_0 \quad \text{versus} \quad H_1: \delta > \delta_0$$

Rejecting  $H_0$  implies that the mean is larger than the value  $\delta_0$ . This test is called an *upper-tail test* because  $H_0$  is rejected in samples in which the sample mean is larger than  $\delta_0$ .

The *lower-tail test* is

$$H_0: \delta \geq \delta_0 \quad \text{versus} \quad H_1: \delta < \delta_0$$

*Superiority by a margin* tests are special cases of the above directional tests. It will be convenient to adopt the following specialize notation for the discussion of these tests.

## Paired Wilcoxon Signed-Rank Tests for Superiority by a Margin

<u>Parameter</u>	<u>PASS Input/Output</u>	<u>Interpretation</u>
$\delta$	$\delta$	<i>Population mean paired difference.</i> This is the mean of paired differences in the population. This parameter will be estimated by the study.
$\delta_1$	$\delta_1$	<i>Actual paired difference at which power is calculated.</i> This is the value of the mean paired difference at which power is calculated.
$M_S$	SM	<i>Margin of superiority.</i> This is a tolerance value that defines the magnitude of difference that is required for practical importance. This may be thought of as the largest difference that is considered to be trivial. The sign is determined by the specific design. This value is used to compute the bound, $\delta_0$ .

## Superiority by a Margin Tests

A *superiority by a margin test* tests that the mean of paired differences is greater than (or less than) zero by more than a small superiority margin. The actual direction of the hypothesis depends on the whether higher values of the response are good or bad.

### Case 1: High Values Good

In this case, higher values are better. The hypotheses are arranged so that rejecting the null hypothesis implies that the mean paired difference is greater than zero by at least the margin of superiority. The value of  $\delta$  at which power is calculated must be greater than  $|M_S|$ . The null and alternative hypotheses with  $\delta_0 = |M_S|$  are

$$H_0: \delta \leq |M_S| \quad \text{versus} \quad H_1: \delta > |M_S|$$

### Case 2: High Values Bad

In this case, lower values are better. The hypotheses are arranged so that rejecting the null hypothesis implies that the mean is less than the reference value by at least the margin of superiority. The value of  $\delta$  at which power is calculated must be less than  $-|M_S|$ . The null and alternative hypotheses with  $\delta_0 = -|M_S|$  are

$$H_0: \delta \geq -|M_S| \quad \text{versus} \quad H_1: \delta < -|M_S|$$

## 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. Subtract the hypothesized mean,  $\delta_0$ , 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.

## Paired Wilcoxon Signed-Rank Tests for Superiority by a Margin

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 Wilcoxon Signed-Rank Test

The power calculation for the Wilcoxon signed-rank test is the same as that for the one-sample  $t$ -test except that an adjustment is made to the sample size based on an assumed data distribution as described in Al-Sunduqchi 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:

<b>Distribution</b>	<b><math>W</math></b>
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  $\delta_1 > \delta_0$ .

1. Find  $t_\alpha$  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:  $X_1 = \delta_0 + t_\alpha \frac{\sigma}{\sqrt{n'}}$ .
3. Calculate the noncentrality parameter:  $\lambda = \frac{\delta_1 - \delta_0}{\frac{\sigma}{\sqrt{n'}}}$ .
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  $\lambda$ .

---

## 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 will be of interest.

---

#### 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.

---

#### Test

##### Higher Means Are

This option defines whether higher values of the response variable are to be considered better or worse. The choice here determines the direction of the superiority by a margin test. Note that  $\delta$  is the population mean of paired differences.

- **Better (H1:  $\delta > SM$ )**

If higher means are Better the null hypothesis is  $H_0: \delta \leq SM$ , and the alternative hypothesis is  $H_1: \delta > SM$ .

- **Worse (H1:  $\delta < -SM$ )**

If higher means are Worse the null hypothesis is  $H_0: \delta \geq -SM$ , and the alternative hypothesis is  $H_1: \delta < -SM$ .

##### Data Distribution

This option makes appropriate sample size adjustments for the Wilcoxon Signed-Rank test. Results by Al-Sunduqchi 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/π<sup>2</sup>”.

## Paired Wilcoxon Signed-Rank Tests for Superiority by a Margin

- **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,  $\sigma$  is the original standard deviation, and  $\sigma_1$  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.

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

## Paired Wilcoxon Signed-Rank Tests for Superiority by a Margin

---

### Effect Size – Means

#### **SM (Superiority Margin)**

This is the magnitude of the margin of superiority for the mean of paired differences. It should be entered as a positive number. If a negative value is entered, the absolute value is used.

When higher means are better, this value is the distance above zero that is required for the new treatment to be considered superior to the reference.

When higher means are worse, this value is the distance below zero that is required for the new treatment to be considered superior to the reference.

#### **$\delta_1$ (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.

When higher means are better, this value should be greater than SM. When higher means are worse, this value should be less than -SM.

---

### Effect Size – Standard Deviation

#### **$\sigma$ (Std Dev of Paired Differences)**

Enter an estimate of the standard deviation of the paired differences (must be positive).

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.

## Example 1 – Power Analysis

Suppose that a test is to be conducted to determine if a new cancer treatment improves the mean bone density. The adjusted mean bone density (AMBD) in the population of interest is 0.002300 gm/cm. Clinicians decide that if the treatment increases AMBD by more than 5% (0.000115 gm/cm), it generates a significant health benefit. They also want to consider what would happen if the margin of superiority is set to 2.5% (0.0000575 gm/cm). The standard deviation of paired differences is 0.000300 gm/cm.

The analysis will be a superiority test using the Wilcoxon signed-rank test at the 0.025 significance level. Power is to be calculated assuming that the new treatment has 7.5% improvement on AMBD. Several sample sizes between 20 and 300 will be analyzed. The researchers want to achieve a power of at least 90%. All numbers have been multiplied by 10000 to make the reports and plots easier to read.

## 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 for Superiority by a Margin** procedure window by expanding **Means**, then **Paired Means**, then clicking on **Superiority by a Margin**, and then clicking on **Paired Wilcoxon Signed-Rank Tests for Superiority by a Margin**. 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>
<b>Design Tab</b>	
Solve For .....	<b>Power</b>
Higher Means Are.....	<b>Better (H1: <math>\delta &gt; SM</math>)</b>
Data Distribution .....	<b>Normal</b>
Population Size.....	<b>Infinite</b>
Alpha.....	<b>0.025</b>
N (Sample Size).....	<b>20 40 60 80 100 150 200 300</b>
SM (Superiority Margin).....	<b>0.575 1.15</b>
$\delta_1$ (Mean of Paired Differences).....	<b>1.725</b>
$\sigma$ (Std Dev of Paired Differences).....	<b>3</b>

## Annotated Output

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

<b>Numeric Results</b>							
Higher Means are Better							
Hypotheses: $H_0: \delta \leq SM$ vs. $H_1: \delta > SM$							
Data Distribution: Normal							
Power	N	Superiority Margin SM	Mean of Paired Differences $\delta_1$	Standard Deviation $\sigma$	Alpha	Beta	
0.35274	20	0.575	1.725	3.0	0.025	0.64726	
0.63360	40	0.575	1.725	3.0	0.025	0.36640	
0.81170	60	0.575	1.725	3.0	0.025	0.18830	
0.90968	80	0.575	1.725	3.0	0.025	0.09032	
0.95888	100	0.575	1.725	3.0	0.025	0.04112	
0.99524	150	0.575	1.725	3.0	0.025	0.00476	
0.99951	200	0.575	1.725	3.0	0.025	0.00049	
1.00000	300	0.575	1.725	3.0	0.025	0.00000	
0.12134	20	1.150	1.725	3.0	0.025	0.87866	
(report continues)							

## Paired Wilcoxon Signed-Rank Tests for Superiority by a Margin

### 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.

SM is the magnitude of the margin of superiority. Since higher means are better, this value is positive and is the distance above the reference value that is required to be considered superior.

$\delta$  is the mean of paired differences.

$\delta_1$  is the value of the mean of paired differences at which power and sample size are calculated.

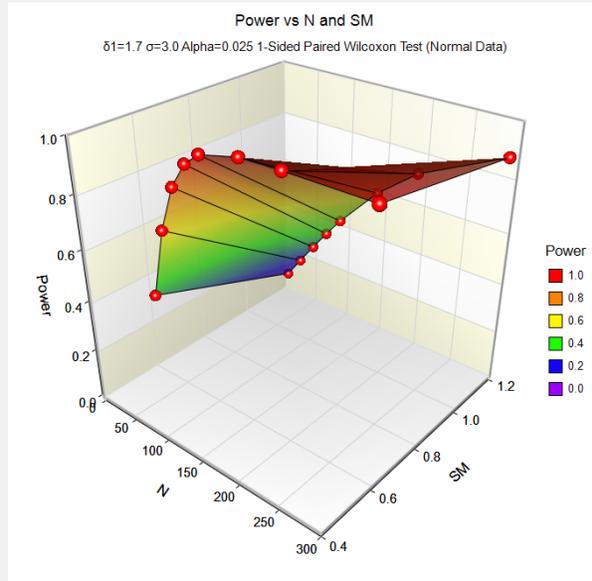
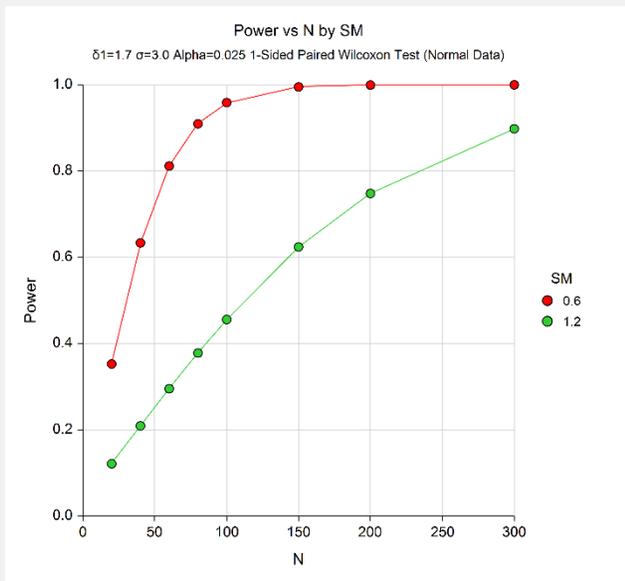
$\sigma$  is the standard deviation of paired differences for the population.

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 20 achieves 35% power to detect superiority using a one-sided paired Wilcoxon Signed-Rank test assuming that the actual distribution of paired differences is normal when the superiority margin is 0.575 and the actual mean of paired differences is 1.725. The estimated standard deviation of paired differences is 3.0. The significance level (alpha) of the test is 0.025.



The above report shows that for SM = 1.15, the sample size necessary to obtain 90% power is just under 300. However, if SM = 0.575, the required sample size is only about 75.

## Example 2 – Finding the Sample Size

Continuing with Example 1, the researchers want to know the exact sample size for each value of SM to achieve 90% 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 for Superiority by a Margin** procedure window by expanding **Means**, then **Paired Means**, then clicking on **Superiority by a Margin**, and then clicking on **Paired Wilcoxon Signed-Rank Tests for Superiority by a Margin**. 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>
<b>Design Tab</b>	
Solve For .....	<b>Sample Size</b>
Higher Means Are .....	<b>Better (H1: <math>\delta &gt; SM</math>)</b>
Data Distribution .....	<b>Normal</b>
Population Size .....	<b>Infinite</b>
Power .....	<b>0.90</b>
Alpha .....	<b>0.025</b>
SM (Superiority Margin) .....	<b>0.575 1.15</b>
$\delta_1$ (Mean of Paired Differences) .....	<b>1.725</b>
$\sigma$ (Std Dev of Paired Differences) .....	<b>3</b>

### Output

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

#### Numeric Results

<b>Numeric Results</b>							
Higher Means are Better							
Hypotheses: $H_0: \delta \leq SM$ vs. $H_1: \delta > SM$							
Data Distribution: Normal							
		Superiority Margin	Mean of Paired Differences	Standard Deviation			
Power	N	SM	$\delta_1$	$\sigma$	Alpha	Beta	
0.90215	78	0.575	1.725	3.0	0.025	0.09785	
0.90005	302	1.150	1.725	3.0	0.025	0.09995	

This report shows the exact sample size requirement for each value of SM.

## Example 3 – Validation

This procedure uses the same mechanics as the Paired Wilcoxon Signed-Rank Tests for Non-Inferiority procedure. We refer the user to Example 3 of Chapter 499 for the validation.