

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.
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.

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.

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:

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

where n is the sample size, N is the population size, σ is the original standard deviation, and σ' 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*.

The Standard Deviation of Paired Differences (σ)

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

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

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.

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-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:

Distribution	W
Uniform	1
Double Exponential	2/3
Logistic	9/ π^2

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 .

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

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	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
Standard Deviation Input Type	Enter the SD of Paired Differences
σ (SD of Paired Differences).....	10 12.5 15

Paired Wilcoxon Signed-Rank Tests

Output

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

Numeric Results

Solve For: [Power](#)
 Hypotheses: $H_0: \delta = 0$ vs. $H_1: \delta \neq 0$
 Data Distribution: Normal

Power	Paired Differences			Effect Size $ \delta /\sigma$	Alpha	Beta
	Sample Size N	Mean δ	Standard Deviation σ			
0.72266	30	-5	10.0	0.50000	0.05	0.27734
0.85114	40	-5	10.0	0.50000	0.05	0.14886
0.91856	50	-5	10.0	0.50000	0.05	0.08144
0.95989	60	-5	10.0	0.50000	0.05	0.04011
0.97939	70	-5	10.0	0.50000	0.05	0.02061
0.99042	80	-5	10.0	0.50000	0.05	0.00958
0.99529	90	-5	10.0	0.50000	0.05	0.00471
0.99790	100	-5	10.0	0.50000	0.05	0.00210
0.53234	30	-5	12.5	0.40000	0.05	0.46766
0.67055	40	-5	12.5	0.40000	0.05	0.32945
0.76567	50	-5	12.5	0.40000	0.05	0.23433
0.84322	60	-5	12.5	0.40000	0.05	0.15678
0.89275	70	-5	12.5	0.40000	0.05	0.10725
0.93085	80	-5	12.5	0.40000	0.05	0.06915
0.95404	90	-5	12.5	0.40000	0.05	0.04596
0.97119	100	-5	12.5	0.40000	0.05	0.02881
0.39800	30	-5	15.0	0.33333	0.05	0.60200
0.51654	40	-5	15.0	0.33333	0.05	0.48346
0.60926	50	-5	15.0	0.33333	0.05	0.39074
0.69613	60	-5	15.0	0.33333	0.05	0.30387
0.76048	70	-5	15.0	0.33333	0.05	0.23952
0.81821	80	-5	15.0	0.33333	0.05	0.18179
0.85946	90	-5	15.0	0.33333	0.05	0.14054
0.89537	100	-5	15.0	0.33333	0.05	0.10463

Power The probability of rejecting a false null hypothesis when the alternative hypothesis is true.
 N The sample size, the number of subjects (or pairs) in the study.
 δ The population mean of paired differences.
 δ The value of the mean of paired differences at which power and sample size are calculated.
 σ The standard deviation of paired differences for the population.
 $|\delta|/\sigma$ The Effect Size, i.e., the relative magnitude of the effect.
 Alpha The probability of rejecting a true null hypothesis.
 Beta The probability of failing to reject the null hypothesis when the alternative hypothesis is true.

Summary Statements

A paired design will be used to test whether the paired difference in distributions (δ) is different from 0 ($H_0: \delta = 0$ versus $H_1: \delta \neq 0$). The comparison will be made using a two-sided, paired-difference Wilcoxon Signed-Rank test, with a Type I error rate (α) of 0.05. The underlying data distribution of paired differences is assumed to be Normal. The underlying standard deviation of the paired difference distribution is assumed to be 10. To detect a paired mean difference of -5 with a sample size of 30 pairs, the power is 0.72266.

Paired Wilcoxon Signed-Rank Tests

Dropout-Inflated Sample Size

Dropout Rate	Sample Size N	Dropout- Inflated Enrollment Sample Size N'	Expected Number of Dropouts D
20%	30	38	8
20%	40	50	10
20%	50	63	13
20%	60	75	15
20%	70	88	18
20%	80	100	20
20%	90	113	23
20%	100	125	25

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

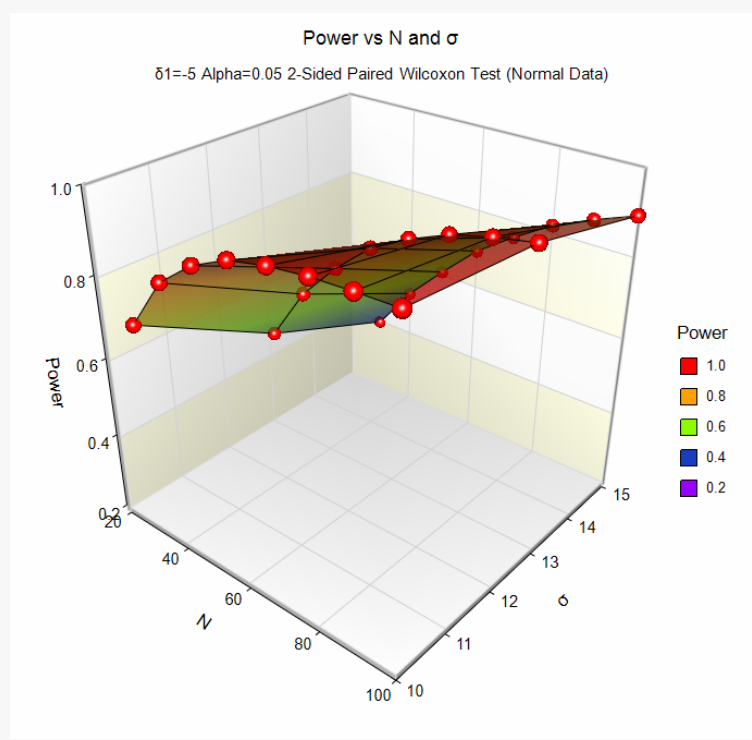
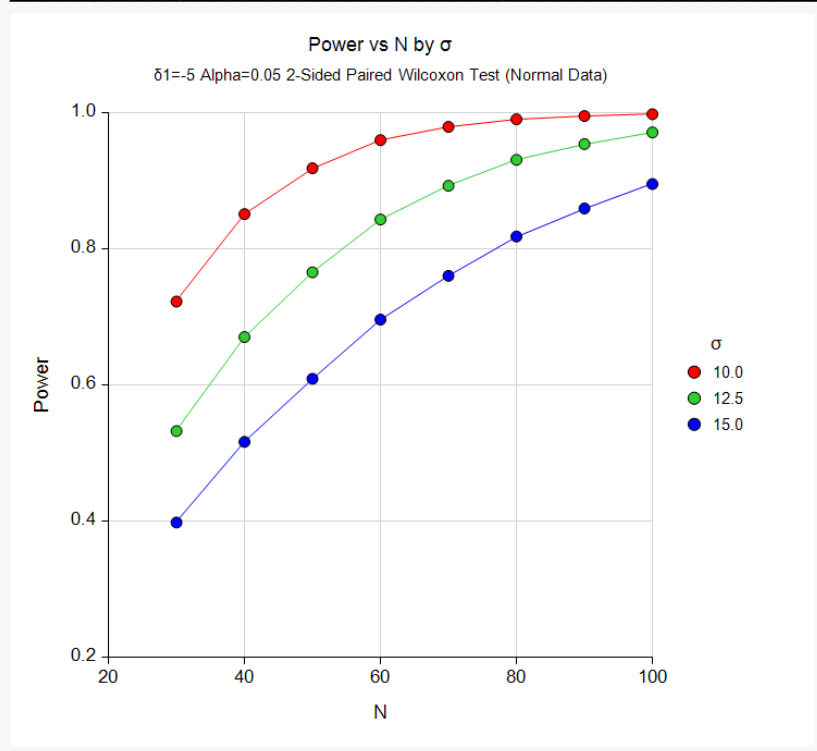
Anticipating a 20% dropout rate, 38 subjects should be enrolled to obtain a final sample size of 30 subjects.

References

- Al-Sundugchi, 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.

Paired Wilcoxon Signed-Rank Tests

Plots



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

Example 2 – Finding the Sample Size

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

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: $\delta \neq 0$)**
 Data Distribution **Normal**
 Population Size **Infinite**
 Power **0.8**
 Alpha **0.05**
 δ_1 (Mean of Paired Differences) **-5**
 Standard Deviation Input Type **Enter the SD of Paired Differences**
 σ (SD of Paired Differences) **10 12.5 15**

Output

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

Numeric Results

Solve For: [Sample Size](#)
 Hypotheses: H0: $\delta = 0$ vs. H1: $\delta \neq 0$
 Data Distribution: Normal

Power	Sample Size N	Paired Differences			Alpha	Beta
		Mean δ_1	Standard Deviation σ	Effect Size $ \delta_1 /\sigma$		
0.80778	36	-5	10.0	0.50000	0.05	0.19222
0.80779	55	-5	12.5	0.40000	0.05	0.19221
0.80230	77	-5	15.0	0.33333	0.05	0.19770

The required sample sizes for each scenario are displayed.

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

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 3a** 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: $\delta \neq 0$)
Data Distribution	Uniform
Population Size	Infinite
Power.....	0.8
Alpha.....	0.05
δ_1 (Mean of Paired Differences).....	0.5
Standard Deviation Input Type	Enter the SD of Paired Differences
σ (SD of Paired Differences).....	1

Paired Wilcoxon Signed-Rank Tests

Output

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

Numeric Results

Solve For: [Sample Size](#)
 Hypotheses: $H_0: \delta = 0$ vs. $H_1: \delta \neq 0$
 Data Distribution: Uniform

Power	Sample Size N	Paired Differences		Effect Size $ \delta /\sigma$	Alpha	Beta
		Mean δ_1	Standard Deviation σ			
0.80778	34	0.5	1	0.5	0.05	0.19222

The sample size of 34 matches Chow, Shao, Wang, and Likhnygina (2018) exactly.

If we now assume a Normal data distribution (**Example 3b** settings file), the results match our expected outcome.

Numeric Results

Solve For: [Sample Size](#)
 Hypotheses: $H_0: \delta = 0$ vs. $H_1: \delta \neq 0$
 Data Distribution: Normal

Power	Sample Size N	Paired Differences		Effect Size $ \delta /\sigma$	Alpha	Beta
		Mean δ_1	Standard Deviation σ			
0.80778	36	0.5	1	0.5	0.05	0.19222

The sample size of 36 matches the expected result.

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

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 4a** 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: $\delta \neq 0$)**
 Data Distribution **Uniform**
 Population Size **Infinite**
 Alpha **0.05**
 N (Sample Size) **12**
 δ_1 (Mean of Paired Differences) **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: $\delta = 0$ vs. H1: $\delta \neq 0$**
 Data Distribution: **Uniform**

Power	Sample Size N	Paired Differences		Effect Size $ \delta_1 /\sigma$	Alpha	Beta
		Mean δ_1	Standard Deviation σ			
0.71366	12	1	1.25	0.8	0.05	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.

Paired Wilcoxon Signed-Rank Tests

If we now assume a double exponential data distribution and set $N = 8$ (**Example 4b** settings file), the results match our expected outcome.

Numeric Results

Solve For: **Power**
 Hypotheses: $H_0: \delta = 0$ vs. $H_1: \delta \neq 0$
 Data Distribution: Double Exponential

Power	Paired Differences			Effect Size $ \delta /\sigma$	Alpha	Beta
	Sample Size N	Mean δ_1	Standard Deviation σ			
0.71366	8	1	1.25	0.8	0.05	0.28634

The power of 0.71366 matches the expected result.

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

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 5a** 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: $\delta \neq 0$)**
 Data Distribution **Uniform**
 Population Size **Infinite**
 Power..... **0.8**
 Alpha..... **0.05**
 δ_1 (Mean of Paired Differences)..... **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: $H_0: \delta = 0$ vs. $H_1: \delta \neq 0$
 Data Distribution: Uniform

Power	Sample Size N	Paired Differences		Effect Size $ \delta_1 /\sigma$	Alpha	Beta
		Mean δ_1	Standard Deviation σ			
0.80169	199	0.2	1	0.2	0.05	0.19831

The sample size of 199 matches Machin's result.

Paired Wilcoxon Signed-Rank Tests

If we now assume a double exponential data distribution (**Example 5b** settings file), the results match our expected outcome.

Numeric Results

Solve For: [Sample Size](#)
 Hypotheses: $H_0: \delta = 0$ vs. $H_1: \delta \neq 0$
 Data Distribution: Double Exponential

Power	Sample Size N	Paired Differences			Alpha	Beta
		Mean δ_1	Standard Deviation σ	Effect Size $ \delta_1 /\sigma$		
0.80169	133	0.2	1	0.2	0.05	0.19831

The sample size of 133 matches the expected result.