

## Chapter 485

# Tests for Paired Means

---

### Introduction

The one-sample paired  $t$ -test may be used to test whether the mean difference of two populations is greater than, less than, or not equal to a specific value. Because the  $t$  distribution is used to calculate critical values for the test, this test is often called the paired  $t$ -test. If the standard deviation of the differences is known, the normal distribution is used instead of the  $t$  distribution and the test is officially known as the  $z$  test.

This module also calculates the power of the nonparametric analog of the  $t$ -test, the *Wilcoxon test*.

---

### Test Procedure

1. **Find the critical value.** Assume that the true mean difference is  $D_0$ . Choose a value  $T_a$  so that the probability of rejecting  $H_0$  when  $H_0$  is true is equal to a specified value called  $\alpha$ . Using the  $t$  distribution, select  $T_a$  so that  $\Pr(t > T_a) = \alpha$ . This value is found using a  $t$  probability table or a computer program (like *PASS*).
2. **Select a sample of  $n$  items from the population and compute the  $t$  statistic.** Call this value  $T$ . If  $T > T_a$  reject the null hypothesis that the mean equals  $D_0$  in favor of an alternative hypothesis that the mean equals  $D_1$  where  $D_1 > D_0$ .

Following is a specific example. Suppose we want to test the hypothesis that a variable,  $X$ , which is made up of paired differences, has a mean of 0 versus the alternative hypothesis that the mean is greater than 0. Suppose that previous studies have shown that the standard deviation of the paired differences,  $\sigma$ , is 40. A random sample of 100 pairs is used.

We first compute the critical value,  $T_a$ . The value of  $T_a$  that yields  $\alpha = 0.05$  is 6.6. If the paired difference mean computed from a sample is greater than 6.6, reject the hypothesis that the mean is 0. Otherwise, do not reject the hypothesis. We call the region greater than 6.6 the *Rejection Region* and values less than or equal to 6.6 the *Acceptance Region* of the significance test.

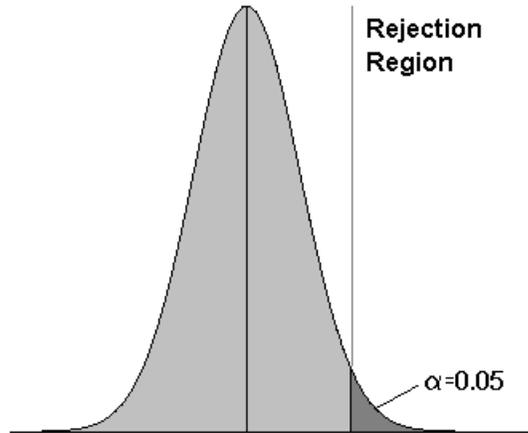
Tests for Paired Means

Now suppose that you want to compute the *power* of this testing procedure. In order to compute the power, we must specify an alternative value for the mean.

We decide to compute the power if the true mean difference were 10. Figure 2 shows how to compute the power in this case.

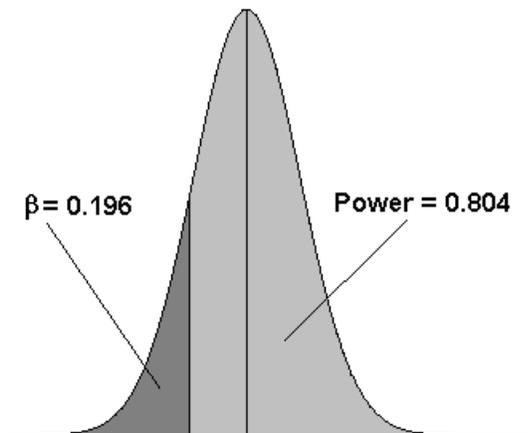
The *power* is the probability of rejecting  $H_0$  when the true mean is 10. Since we reject  $H_0$  when the calculated mean is greater than 6.6, the probability of a Type-II error (called  $\beta$ ) is given by the dark, shaded area of the second graph. This value is 0.196. The power is equal to  $1 - \beta$  or 0.804.

Figure 1 - Finding Alpha



Note that there are six parameters that may be varied in this situation: two means, standard deviation, alpha, beta, and the sample size.

Figure 2 - Finding Beta



## Tests for Paired Means

---

### Assumptions

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 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.

---

### Technical Details

---

#### Standard Deviation Known

When the standard deviation is known, the power is calculated as follows for a directional alternative (one-tailed test) in which  $D1 > D0$ .

1. Find  $z_\alpha$  such that  $1 - \Phi(z_\alpha) = \alpha$ , where  $\Phi(x)$  is the area under the standardized normal curve to the left of  $x$ .
2. Calculate:  $x_a = D_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$ .

## Tests for Paired Means

3. Calculate: 
$$z_\alpha = \frac{x_a - M1}{\frac{\sigma}{\sqrt{n}}}.$$
4. Power =  $1 - \Phi(z_\alpha).$

## Standard Deviation Unknown

When the standard deviation is unknown, the power is calculated as follows for a directional alternative (one-tailed test) in which  $D1 > D0$ .

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: 
$$x_a = D_0 + t_\alpha \frac{\sigma}{\sqrt{n}}.$$
3. Calculate the noncentrality parameter: 
$$\lambda_\alpha = \frac{D1 - D0}{\frac{\sigma}{\sqrt{n}}}.$$
4. Calculate: 
$$t_a = \frac{x_a - D1}{\frac{\sigma}{\sqrt{n}}} + \lambda_\alpha.$$
5. Calculate: Power =  $1 - T'_{df,\lambda}(t_a)$ , where  $T'_{df,\lambda}(x)$  is the area under a noncentral- $t$  curve with degrees of freedom  $df$  and noncentrality parameter  $\lambda$  to the left of  $x$ .

## 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.

## Tests for Paired Means

---

**Test****Alternative Hypothesis**

This option specifies the alternative hypothesis. This implicitly specifies the direction of the hypothesis test. The null hypothesis is always  $H_0$ : Mean of Paired Differences = 0.

Note that the alternative hypothesis enters into power calculations by specifying the rejection region of the hypothesis test. Its accuracy is critical.

Possible selections are:

- **Ha: Mean of Paired Diffs  $\neq 0$**   
This is the most common selection. It yields the *two-tailed t-test*. Use this option when you are testing whether the means are different but you do not want to specify beforehand which mean is larger. Many scientific journals require two-tailed tests.
- **Ha: Mean of Paired Diffs  $< 0$**   
This option yields a *one-tailed t-test*.
- **Ha: Mean of Paired Diffs  $> 0$**   
This options yields a *one-tailed t-test*.

**Nonparametric Adjustment**

This option makes appropriate sample size adjustments for the Wilcoxon test. Results by Al-Sunduqchi and Guenther (1990) indicate that power calculations for the Wilcoxon 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. These are 1 for the uniform distribution,  $2/3$  for the double exponential distribution,  $9/\pi^2$  for the logistic distribution, and  $\pi/3$  for the normal distribution.

The options are as follows:

- **Ignore**  
Do not make a Wilcoxon adjustment. This indicates that you want to analyze a *t-test*, not the Wilcoxon test.
- **Uniform**  
Make the Wilcoxon sample size adjustment assuming the uniform distribution. Since the factor is one, this option performs the same as Ignore. It is included for completeness.
- **Double Exponential**  
Make the Wilcoxon sample size adjustment assuming that the data actually follow the double exponential distribution.
- **Logistic**  
Make the Wilcoxon sample size adjustment assuming that the data actually follow the logistic distribution.
- **Normal**  
Make the Wilcoxon sample size adjustment assuming that the data actually follow the normal distribution.

## Tests for Paired Means

### 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 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 – Means

### Mean of Paired Differences (Alternative)

Enter a value (or range of values) for the mean of paired differences under the alternative hypothesis. This value indicates the minimum detectable difference for the corresponding power and sample size.

## Tests for Paired Means

---

**Effect Size – Standard Deviation****S (Standard Deviation of Paired Differences)**

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

Use results from a previous (or pilot) study or the range divided by 5.

You can enter a range of values such as *1 2 3* or *1 to 10 by 1*.

Press the Estimator button to load the Standard Deviation Estimator window.

**Known Standard Deviation**

This option specifies whether the standard deviation ( $\sigma$ ) is known or unknown. In almost all experimental situations, the standard deviation is not known. However, great calculation efficiencies are obtained if the standard deviation is assumed to be known.

When this box is checked, the program performs its calculations assuming that the standard deviation is known. This results in the use of the normal distribution in all probability calculations. Calculations using this option will be much faster than for the unknown standard deviation case. The results for either case will be close when the sample size is over 30.

When this box is not checked, the program assumes that the standard deviation is not known and will be estimated from the data when the  $t$ -test is run. This results in probability calculations using the noncentral- $t$  distribution. This distribution requires a lot more calculations than does the normal distribution.

The calculation speed comes into play whenever the Find option is set to something besides *Beta*. In these cases, the program uses a special searching algorithm which requires numerous iterations. You will note a real difference in calculation speed depending on whether this option is checked.

A reasonable strategy would be to leave this option checked while you are experimenting with the parameters and then turn it off when you are ready for your final results.

## Example 1 – Paired T-Test

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 are analyzed using a one-sample  $t$ -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  $t$ -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. Alpha values of 0.01 and 0.05 will be tried. Beta is set to 0.20 so that the power is 80%. How large of a sample must the researchers take?

### Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for One Mean** procedure window by expanding **Means**, then **Paired Means**, then clicking on **Test (Inequality)**, and then clicking on **Tests for Paired Means**. 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>Sample Size</b>
Alternative Hypothesis .....	<b>Ha: Mean of Paired Diffs <math>\neq</math> 0</b>
Nonparametric Adjustment .....	<b>Ignore</b>
Population Size .....	<b>Infinite</b>
Power .....	<b>0.80</b>
Alpha .....	<b>0.01 0.05</b>
Mean of Paired Diffs (Alternative) .....	<b>-5</b>
S (SD of Paired Differences) .....	<b>10 12.5 15</b>
Known Standard Deviation .....	<b>Not checked</b>

## Tests for Paired Means

## Output

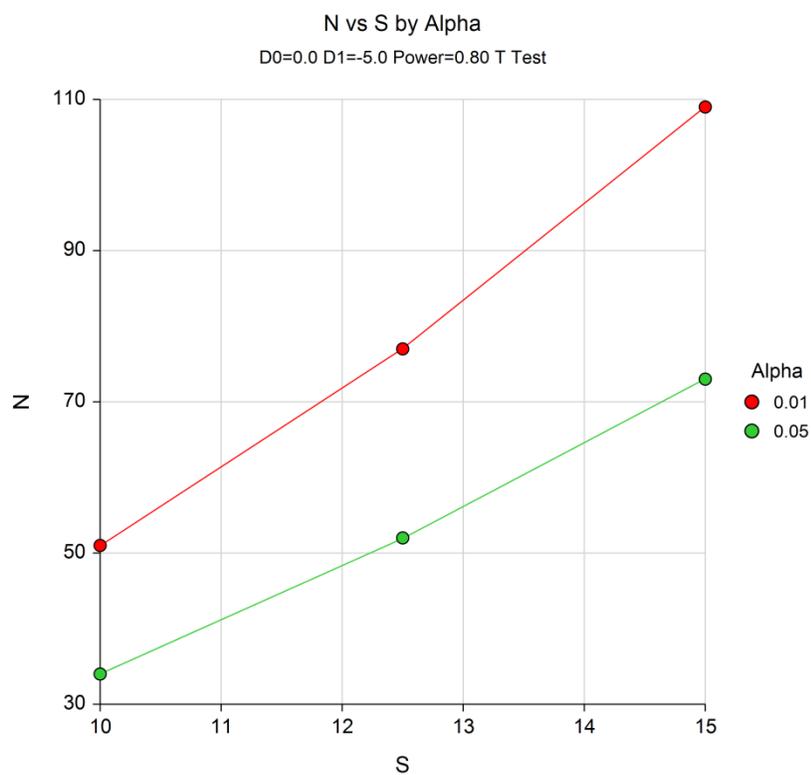
Click the Calculate button to perform the calculations and generate the following output.

## Numeric Results and Plots

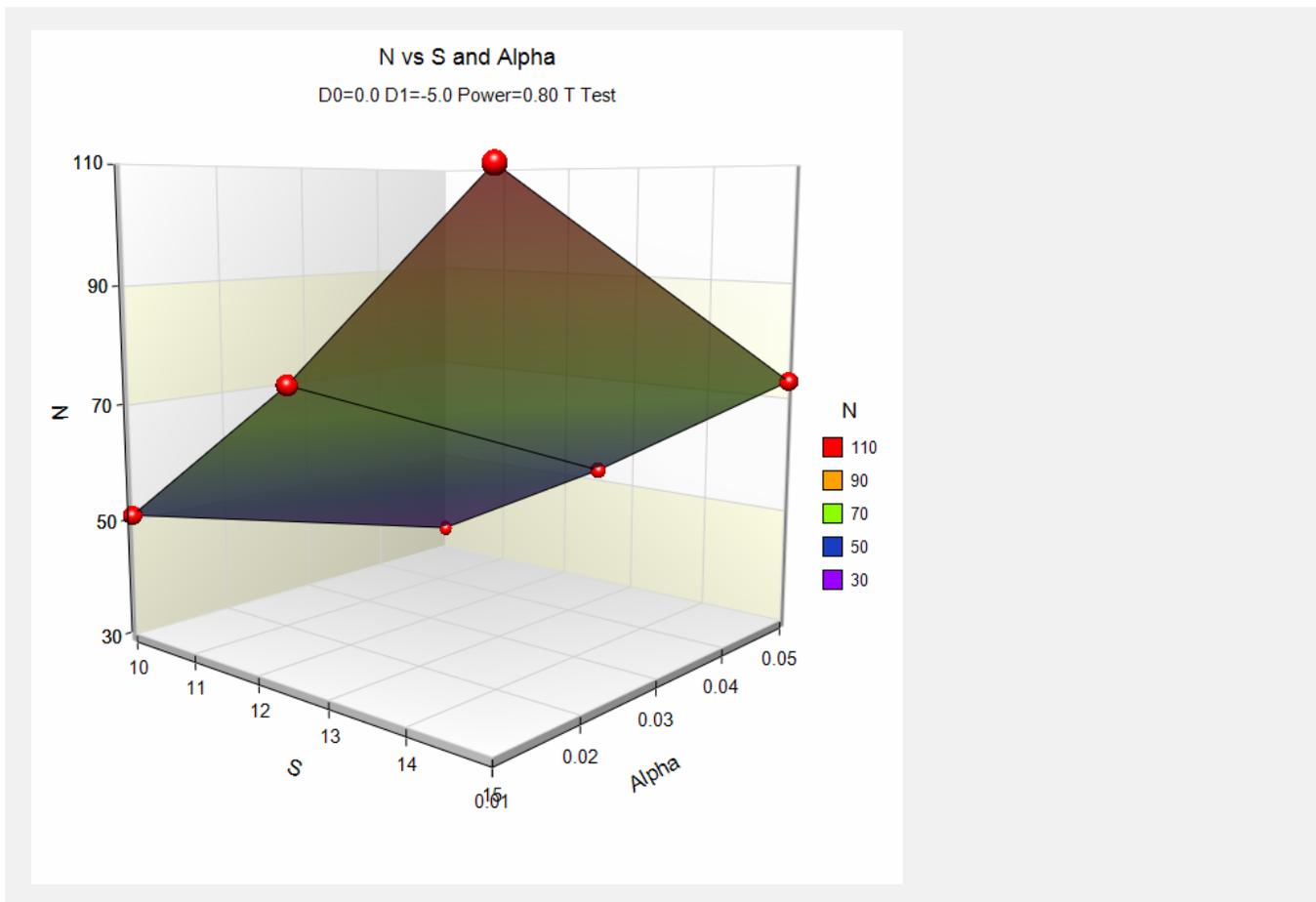
## Numeric Results for Paired T-Test

Null Hypothesis: Mean of Paired Differences = 0, Alternative Hypothesis: Mean of Paired Differences  $\neq$  0  
Unknown standard deviation.

Power	N	Alpha	Beta	Mean of Paired Differences	S	Effect Size
0.80939	51	0.01000	0.19061	-5.0	10.0	0.500
0.80778	34	0.05000	0.19222	-5.0	10.0	0.500
0.80434	77	0.01000	0.19566	-5.0	12.5	0.400
0.80779	52	0.05000	0.19221	-5.0	12.5	0.400
0.80252	109	0.01000	0.19748	-5.0	15.0	0.333
0.80230	73	0.05000	0.19770	-5.0	15.0	0.333



## Tests for Paired Means



The report shows the values of each of the parameters, one scenario per row. The parameter solved for was the sample size,  $N$ .

Note that depending on our choice of assumptions, the sample size ranges from 34 to 109. Hence, the researchers have to make a careful determination of which standard deviation and significance level should be used.

---

## Example 2 – Validation using Zar

This procedure uses the same mechanics as the Tests for One Mean procedure. The validation for this procedure is obtained through Examples 6 and 7 of Chapter 400.