

## Chapter 492

# Paired Z-Tests

---

## Introduction

The paired z-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 standard normal distribution is used to calculate critical values for the test, this test is often called the paired z-test. The paired z-test assumes that the population standard deviation of paired differences is known.

---

## 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 T-Tests and the nonparametric Wilcoxon Signed-Rank Test procedures. 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.

---

## The Statistical Hypotheses

In the usual paired z-test setting with  $\delta$  defined as the mean paired difference, the null ( $H_0$ ) and alternative ( $H_1$ ) hypotheses for two-sided tests are defined as

$$H_0: \delta = 0 \quad \text{versus} \quad H_1: \delta \neq 0.$$

Rejecting  $H_0$  implies that the mean paired difference is not equal to 0. The hypotheses for one-sided upper-tail tests are

$$H_0: \delta \leq 0 \quad \text{versus} \quad H_1: \delta > 0.$$

Rejecting  $H_0$  implies that the mean is larger than the value  $\mu_0$ . This test is called an *upper-tail test* because  $H_0$  is rejected in samples in which the sample mean is larger than  $\mu_0$ .

The *lower-tail test* is

$$H_0: \delta \geq 0 \quad \text{versus} \quad H_1: \delta < 0.$$

## Paired Z-Tests

It will be convenient to adopt the following specialized notation for the discussion of these tests.

<b>Parameter</b>	<b>PASS Input/Output</b>	<b>Interpretation</b>
$\delta$	$\delta$	<i>Population mean paired difference.</i> This is the mean of paired differences. 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.

---

## 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 Z-Test Statistic

The paired z-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 z-test proceeds as follows

$$z = \frac{\bar{X}}{\sigma_D/\sqrt{n}}$$

where

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

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.

## 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,  $\sigma$  is the original standard deviation, and  $\sigma'$  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 ( $\sigma$ )

If you know the standard deviation of paired differences ( $\sigma$ ), enter it directly. If you know the standard deviations of the paired variables ( $\sigma_1$  and  $\sigma_2$ ) and the Pearson correlation between the paired variables ( $\rho$ ), the standard deviation of paired differences ( $\sigma$ ) 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 know the within-subject population standard deviation ( $\sigma_w$ ), then  $\sigma$  may be calculated using the equation

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

such that

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

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

---

## Power Calculation for the Paired Z-Test

When the standard deviation is known, the power is calculated as follows for a directional alternative (one-tailed test) in which  $\delta_1 > 0$ .

1. Find  $z_\alpha$  such that  $1 - \Phi(z_\alpha) = \alpha$ , where  $\Phi(x)$  is the area to the left of  $x$  under the standardized normal curve.
2. Calculate:  $X_1 = z_\alpha \frac{\sigma}{\sqrt{n}}$ .
3. Calculate:  $z_1 = \frac{X_1 - \delta_1}{\frac{\sigma}{\sqrt{n}}}$ .
4. Power =  $1 - \Phi(z_1)$ .

## 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 are analyzed using a one-sample z-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 z-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.

The standard deviation of paired differences is known to be 15. 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?

### 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 .....	<b>Power</b>
Alternative Hypothesis .....	<b>Two-Sided (H1: <math>\delta \neq 0</math>)</b>
Population Size .....	<b>Infinite</b>
Alpha.....	<b>0.05</b>
N (Sample Size).....	<b>30 to 100 by 10</b>
$\delta_1$ (Mean of Paired Differences).....	<b>-5</b>
Standard Deviation Input Type .....	<b>Enter the SD of Paired Differences</b>
$\sigma$ (SD of Paired Differences).....	<b>15</b>

## Paired Z-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.  $H_1: \delta \neq 0$ 


---

Power	Sample Size N	Paired Differences		Effect Size $ \delta_1 /\sigma$	Alpha	Beta
		Mean $\delta_1$	Standard Deviation $\sigma$			
0.44669	30	-5	15	0.33333	0.05	0.55331
0.55894	40	-5	15	0.33333	0.05	0.44106
0.65435	50	-5	15	0.33333	0.05	0.34565
0.73304	60	-5	15	0.33333	0.05	0.26696
0.79642	70	-5	15	0.33333	0.05	0.20358
0.84648	80	-5	15	0.33333	0.05	0.15352
0.88538	90	-5	15	0.33333	0.05	0.11462
0.91518	100	-5	15	0.33333	0.05	0.08482

---

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.  
 $\delta$  The population mean of paired differences.  
 $\delta_1$  The value of the mean of paired differences at which power and sample size are calculated.  
 $\sigma$  The standard deviation of paired differences for the population.  
 $|\delta_1|/\sigma$  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 mean difference ( $\delta$ ) 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 Z-test, with a Type I error rate ( $\alpha$ ) of 0.05. The (known) standard deviation of paired differences is assumed to be 15. To detect a paired mean difference of -5 with a sample size of 30 pairs, the power is 0.44669.

---

## Paired Z-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.

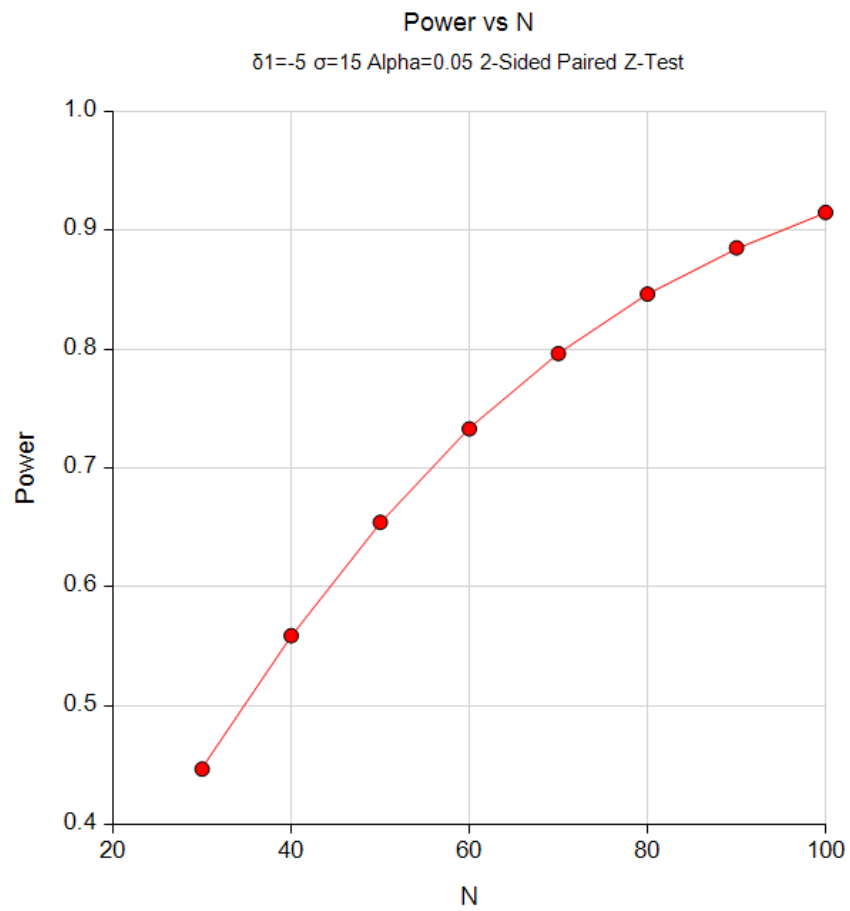
**References**

- 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 Z-Tests

## Plots



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

## Example 2 – Finding the Sample Size

Continuing with Example 1, how many pairs are required 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:  $\delta \neq 0$ )**  
 Population Size ..... **Infinite**  
 Power ..... **0.8**  
 Alpha ..... **0.05**  
 $\delta_1$  (Mean of Paired Differences) ..... **-5**  
 Standard Deviation Input Type ..... **Enter the SD of Paired Differences**  
 $\sigma$  (SD of Paired Differences) ..... **15**

### 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.  $H_1: \delta \neq 0$

Power	Sample Size N	Paired Differences			Alpha	Beta
		Mean $\delta_1$	Standard Deviation $\sigma$	Effect Size $ \delta_1 /\sigma$		
0.80199	71	-5	15	0.33333	0.05	0.19801

A sample size of 71 is required to achieve 80% power.

## 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 z-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 32.

If we set  $\delta_1 = 2.0 - 1.5 = 0.5$  then we should get the same result because the one-sample z-test and the paired z-test use the same fundamental calculations.

### 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 .....	<b>Sample Size</b>
Alternative Hypothesis .....	<b>Two-Sided (H1: <math>\delta \neq 0</math>)</b>
Population Size .....	<b>Infinite</b>
Power.....	<b>0.8</b>
Alpha.....	<b>0.05</b>
$\delta_1$ (Mean of Paired Differences).....	<b>0.5</b>
Standard Deviation Input Type .....	<b>Enter the SD of Paired Differences</b>
$\sigma$ (SD of Paired Differences).....	<b>1</b>

### Output

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

Numeric Results						
Solve For: <a href="#">Sample Size</a>						
Hypotheses: $H_0: \delta = 0$ vs. $H_1: \delta \neq 0$						
Power	Sample Size N	Paired Differences		Effect Size $ \delta_1 /\sigma$	Alpha	Beta
		Mean $\delta_1$	Standard Deviation $\sigma$			
0.80743	32	0.5	1	0.5	0.05	0.19257

The sample size of 32 matches Chow, Shao, Wang, and Lokhnygina (2018) exactly.

## Example 4 – 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. The z-test should give a similar but slightly lower result because the normal distribution approximates the  $t$  distribution very well at this sample size.

### 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 ..... **Sample Size**  
 Alternative Hypothesis ..... **Two-Sided (H1:  $\delta \neq 0$ )**  
 Population Size ..... **Infinite**  
 Power ..... **0.8**  
 Alpha ..... **0.05**  
 $\delta_1$  (Mean of Paired Differences) ..... **0.2**  
 Standard Deviation Input Type ..... **Enter the SD of Paired Differences**  
 $\sigma$  (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.  $H_1: \delta \neq 0$

---

Power	Sample Size N	Paired Differences		Effect Size $ \delta_1 /\sigma$	Alpha	Beta
		Mean $\delta_1$	Standard Deviation $\sigma$			
0.80155	197	0.2	1	0.2	0.05	0.19845

The sample size of 197 is very close to Machin’s result for the  $t$ -test.