

## Chapter 494

# Paired Z-Tests for Non-Inferiority

---

## Introduction

This procedure computes power and sample size for non-inferiority tests in paired designs in which the paired differences are distributed as a normal random variable and the standard deviation is known.

The details of sample size calculation for the paired design are presented in the Paired Z-Tests chapter and they will not be duplicated here. This chapter only discusses those changes necessary for non-inferiority tests. Sample size formulas for non-inferiority tests of a single mean are presented in Chow et al. (2018) page 43.

The *one-sample z-test* is used to test whether a population mean is different from a specific value. When the data are differences between paired values, this test is known as the *paired z-test*.

---

## 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 z-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 Z-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 non-inferiority tests for paired data. This section will review the specifics of non-inferiority testing.

Remember that in the usual z-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 mean of paired differences is larger than  $\delta_0$ .

The *lower-tail test* is

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

## Paired Z-Tests for Non-Inferiority

*Non-inferiority* tests are special cases of the above directional 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 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.
$M_{NI}$	NIM	<i>Margin of non-inferiority.</i> This is a tolerance value that defines the magnitude of difference that is not of 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, $\delta_0$ .

---

## Non-Inferiority Tests

A *non-inferiority test* tests that the mean difference is not less than (or greater than) zero by more than a small non-inferiority margin. The actual direction of the hypothesis depends on 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 difference is no less than a small amount below zero. The value of  $\delta$  at which power is calculated is often set to zero. The null and alternative hypotheses with  $\delta_0 = -|M_{NI}|$  are

$$H_0: \delta \leq -|M_{NI}| \quad \text{versus} \quad H_1: \delta > -|M_{NI}|$$

### 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 of the treatment group is no more than a small amount above the reference value. The value of  $\delta$  at which power is calculated is often set to zero. The null and alternative hypotheses with  $\delta_0 = |M_{NI}|$  are

$$H_0: \delta \geq |M_{NI}| \quad \text{versus} \quad H_1: \delta < |M_{NI}|$$

---

## 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 (known). This assumption implies that the data are continuous, and their distribution is symmetric. The calculation of the z-test proceeds as follows

$$z = \frac{\bar{X} - \delta_0}{\sigma/\sqrt{n}}$$

where

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

and  $\delta_0$  is the value of the mean hypothesized by the null hypothesis which depends on the magnitude and sign of  $M_{NI}$ .

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  $\delta_1 > \delta_0$ .

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:  $X_1 = \delta_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$ .
3. Calculate:  $z_1 = \frac{X_1 - \delta_1}{\frac{\sigma}{\sqrt{n}}}$ .
4. Power =  $1 - \Phi(z_1)$ .

## Example 1 – Power Analysis

Suppose that a test is to be conducted to determine if a new cancer treatment adversely affects 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 reduces AMBD by more than 5% (0.000115 gm/cm), it poses a significant health threat. They also want to consider what would happen if the margin of non-inferiority is set to 2.5% (0.0000575 gm/cm). The standard deviation of paired differences is 0.000300 gm/cm.

Following accepted procedure, the analysis will be a non-inferiority test using the paired z-test at the 0.025 significance level. Power is to be calculated assuming that the new treatment has no effect 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

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>
Higher Means Are .....	<b>Better (H1: <math>\delta &gt; -NIM</math>)</b>
Population Size .....	<b>Infinite</b>
Alpha.....	<b>0.025</b>
N (Sample Size).....	<b>20 40 60 80 100 150 200 300</b>
NIM (Non-Inferiority Margin) .....	<b>0.575 1.15</b>
$\delta_1$ (Mean of Paired Differences).....	<b>0</b>
Standard Deviation Input Type .....	<b>Enter the SD of Paired Differences</b>
$\sigma$ (SD of Paired Differences).....	<b>3</b>

## Paired Z-Tests for Non-Inferiority

## Output

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

---

**Numeric Results**


---

Solve For: **Power**  
 Higher Means Are: **Better**  
 Hypotheses: **H0:  $\delta \leq$  -NIM vs. H1:  $\delta >$  -NIM**

---

Paired Differences						
Power	Sample Size N	Non-Inferiority Margin -NIM	Mean $\delta_1$	Standard Deviation $\sigma$	Alpha	Beta
0.13506	20	-0.575	0	3	0.025	0.86494
0.22730	40	-0.575	0	3	0.025	0.77270
0.31728	60	-0.575	0	3	0.025	0.68272
0.40298	80	-0.575	0	3	0.025	0.59702
0.48273	100	-0.575	0	3	0.025	0.51727
0.65079	150	-0.575	0	3	0.025	0.34921
0.77356	200	-0.575	0	3	0.025	0.22644
0.91305	300	-0.575	0	3	0.025	0.08695
0.40298	20	-1.150	0	3	0.025	0.59702
0.67884	40	-1.150	0	3	0.025	0.32116
0.84359	60	-1.150	0	3	0.025	0.15641
0.92904	80	-1.150	0	3	0.025	0.07096
0.96949	100	-1.150	0	3	0.025	0.03051
0.99688	150	-1.150	0	3	0.025	0.00312
0.99973	200	-1.150	0	3	0.025	0.00027
1.00000	300	-1.150	0	3	0.025	0.00000

---

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.  
 -NIM The magnitude and direction of the margin of non-inferiority. Since higher means are better, this value is negative and is the distance below the reference value that is still considered non-inferior.  
 $\delta$  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.  
 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 (where higher means are considered to be better) will be used to test whether treatment 1 is non-inferior to treatment 2 by testing whether the paired mean difference ( $\delta$ ) is greater than -0.575 (H0:  $\delta \leq$  -0.575 versus H1:  $\delta >$  -0.575). The comparison will be made using a one-sided, paired-difference Z-test, with a Type I error rate ( $\alpha$ ) of 0.025. The (known) standard deviation of paired differences is assumed to be 3. To detect a paired mean difference of 0 with a sample size of 20 pairs, the power is 0.13506.

---

## Paired Z-Tests for Non-Inferiority

**Dropout-Inflated Sample Size**

Dropout Rate	Sample Size N	Dropout- Inflated Enrollment Sample Size N'	Expected Number of Dropouts D
20%	20	25	5
20%	40	50	10
20%	60	75	15
20%	80	100	20
20%	100	125	25
20%	150	188	38
20%	200	250	50
20%	300	375	75

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**

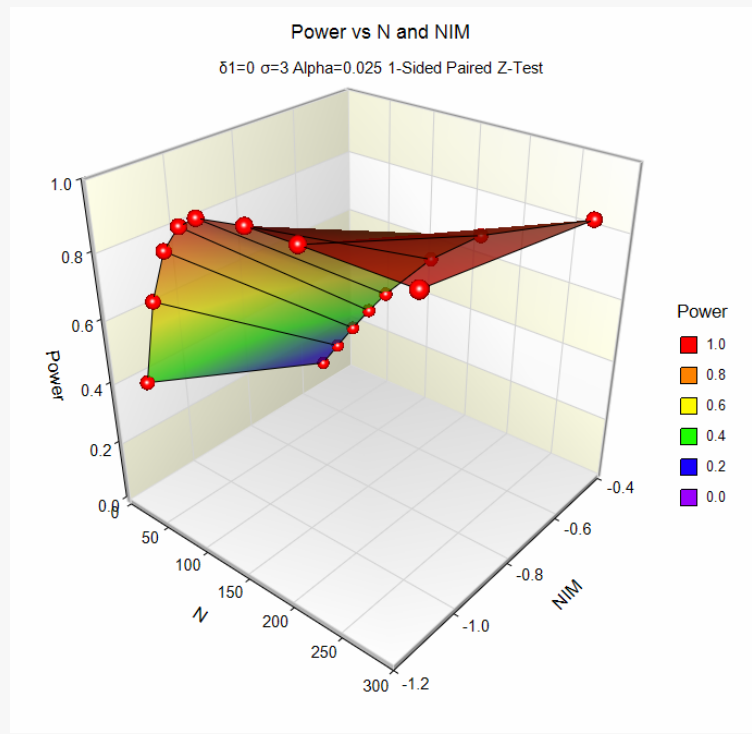
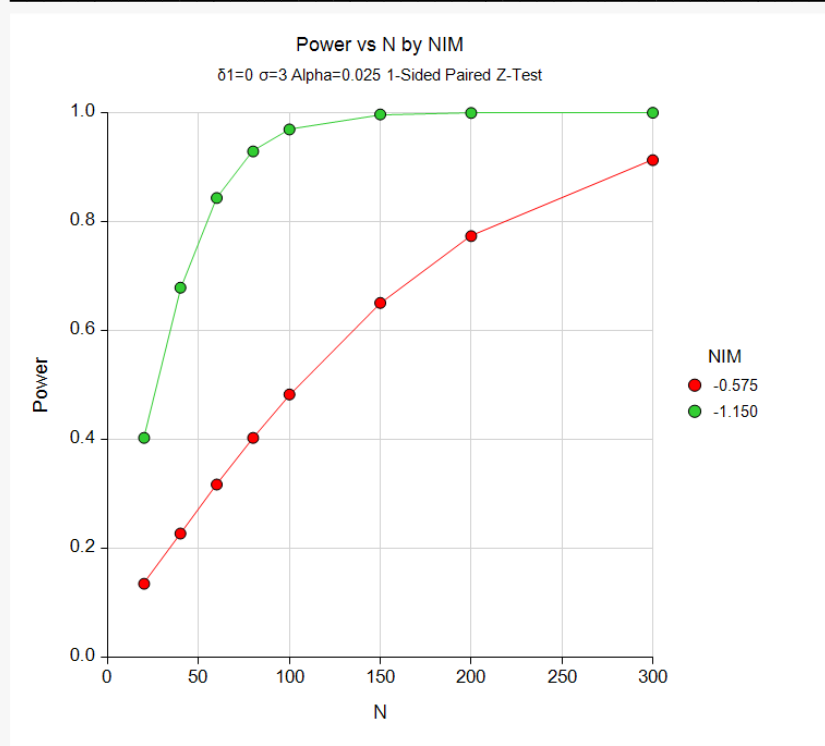
Anticipating a 20% dropout rate, 25 subjects should be enrolled to obtain a final sample size of 20 subjects.

**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.
- Julious, Steven A. 2004. 'Tutorial in Biostatistics. Sample sizes for clinical trials with Normal data.' *Statistics in Medicine*, 23:1921-1986.

Paired Z-Tests for Non-Inferiority

Plots



The above report shows that for NIM = 1.15, the sample size necessary to obtain 90% power is just under 80. However, if NIM = 0.575, the required sample size is about 300.



## Example 2 – Finding the Sample Size

Continuing with Example 1, the researchers want to know the exact sample size for each value of NIM.

### 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 .....	<b>Sample Size</b>
Higher Means Are .....	<b>Better (H1: <math>\delta &gt; -NIM</math>)</b>
Population Size .....	<b>Infinite</b>
Power .....	<b>0.90</b>
Alpha .....	<b>0.025</b>
NIM (Non-Inferiority Margin) .....	<b>0.575 1.15</b>
$\delta_1$ (Mean of Paired Differences) .....	<b>0</b>
Standard Deviation Input Type .....	<b>Enter the SD of Paired Differences</b>
$\sigma$ (SD of Paired Differences) .....	<b>3</b>

### Output

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

Numeric Results						
Solve For:	<a href="#">Sample Size</a>					
Higher Means Are:	Better					
Hypotheses:	H0: $\delta \leq -NIM$ vs. H1: $\delta > -NIM$					
Paired Differences						
	Sample Size N	Non-Inferiority Margin -NIM	Mean $\delta_1$	Standard Deviation $\sigma$	Alpha	Beta
0.90097	287	-0.575	0	3	0.025	0.09903
0.90195	72	-1.150	0	3	0.025	0.09805

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

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

Chow, Shao, Wang, and Lokhnygina (2018) page 46 has an example of a sample size calculation for a paired non-inferiority z-test. Their example obtains a sample size of 7 when  $\delta_1 = 0.5$ , NIM = 0.5,  $\sigma = 1$ , Alpha = 0.05, and Power = 0.80.

### 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 ..... Sample Size
Higher Means Are ..... Better (H1:  $\delta > -NIM$ )
Population Size ..... Infinite
Power ..... 0.80
Alpha ..... 0.05
NIM (Non-Inferiority Margin) ..... 0.5
 $\delta_1$  (Mean of Paired Differences) ..... 0.5
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
Higher Means Are: Better
Hypotheses:    H0:  $\delta \leq -NIM$  vs. H1:  $\delta > -NIM$ 
  
```

Power	Sample Size N	Paired Differences			Alpha	Beta
		Non-Inferiority Margin -NIM	Mean $\delta_1$	Standard Deviation $\sigma$		
0.84156	7	-0.5	0.5	1	0.05	0.15844

PASS also obtains a sample size of 7.