

## Chapter 450

# Two-Sample T-Tests for Non-Inferiority Assuming Equal Variance

## Introduction

This procedure computes power and sample size for *non-inferiority* tests in two-sample designs in which the outcome is a continuous normal random variable and the variances of the two groups (populations) are assumed to be equal. Measurements are made on individuals that have been randomly assigned to one of two groups. This is sometimes referred to as a *parallel-groups* design. This design is used in situations such as the comparison of the income level of two regions, the nitrogen content of two lakes, or the effectiveness of two drugs.

The details of sample size calculation for the two-sample design are presented in the Two-Sample T-Tests Assuming Equal Variance 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 two means are presented in Chow et al. (2018) pages 50-51.

## The Statistical Hypotheses

Remember that in the usual *t*-test setting, the null ( $H_0$ ) and alternative ( $H_1$ ) hypotheses for one-sided tests are defined as

$$H_0: \mu_1 - \mu_2 \leq \delta_0 \quad \text{versus} \quad H_1: \mu_1 - \mu_2 > \delta_0$$

or equivalently

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

Rejecting this test implies that the mean difference is larger than the value  $\delta_0$ . This test is called an *upper-tailed test* because it is rejected in samples in which the difference between the sample means is larger than  $\delta_0$ .

Following is an example of a *lower-tailed test*.

$$H_0: \mu_1 - \mu_2 \geq \delta_0 \quad \text{versus} \quad H_1: \mu_1 - \mu_2 < \delta_0$$

or equivalently

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

*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.

## Two-Sample T-Tests for Non-Inferiority Assuming Equal Variance

<b>Parameter</b>	<b>PASS Input/Output</b>	<b>Interpretation</b>
$\mu_1$	Not used	<i>Mean</i> of population 1. Population 1 is assumed to consist of those who have received the new treatment.
$\mu_2$	Not used	<i>Mean</i> of population 2. Population 2 is assumed to consist of those who have received the reference treatment.
$M_{NI}$	NIM	<i>Margin of non-inferiority</i> . This is a tolerance value that defines the magnitude of the amount that is not of practical importance. This may be thought of as the largest change from the baseline that is considered to be trivial. The sign of the value will be determined by the specific design that is being used.
$\delta$	$\delta$	<i>Actual difference</i> . This is the value of $\mu_1 - \mu_2$ , the difference between the means. This is the value at which the power is calculated.

Note that the actual values of  $\mu_1$  and  $\mu_2$  are not needed. Only their difference is needed for power and sample size calculations.

## Non-Inferiority Tests

A *non-inferiority test* tests that the treatment mean is not worse than the reference mean by more than the non-inferiority margin. The actual direction of the hypothesis depends on the response variable being studied.

### 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 treatment mean is no less than a small amount below the reference mean. 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

$$H_0: \mu_1 \leq \mu_2 - |M_{NI}| \quad \text{versus} \quad H_1: \mu_1 > \mu_2 - |M_{NI}|$$

$$H_0: \mu_1 - \mu_2 \leq -|M_{NI}| \quad \text{versus} \quad H_1: \mu_1 - \mu_2 > -|M_{NI}|$$

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

### 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 treatment mean is no more than a small amount above the reference mean. 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

$$H_0: \mu_1 \geq \mu_2 + |M_{NI}| \quad \text{versus} \quad H_1: \mu_1 < \mu_2 + |M_{NI}|$$

$$H_0: \mu_1 - \mu_2 \geq |M_{NI}| \quad \text{versus} \quad H_1: \mu_1 - \mu_2 < |M_{NI}|$$

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

## Example

A non-inferiority test example will set the stage for the discussion of the terminology that follows. Suppose that a test is to be conducted to determine if a new cancer treatment adversely affects mean bone density. The adjusted mean bone density (AMBD) in the population of interest is 0.002300 gm/cm with a standard deviation of 0.000300 gm/cm. Clinicians decide that if the treatment reduces AMBD by more than 5% (0.000115 gm/cm), it poses a significant health threat.

The hypothesis of interest is whether the mean AMBD in the treated group is more than 0.000115 below that of the reference group. The statistical test will be set up so that if the null hypothesis is rejected, the conclusion will be that the new treatment is non-inferior. The value 0.000115 gm/cm is called the *margin of non-inferiority*.

## Two-Sample Equal-Variance T-Test Statistic

Under the null hypothesis, this test assumes that the two groups of data are simple random samples from a single population of normally distributed 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 test statistic for the case when higher response values are better is as follows.

$$t_{df} = \frac{(\bar{X}_1 - \bar{X}_2) - \delta_0}{s_{\bar{X}_1 - \bar{X}_2}}$$

where

$$\bar{X}_k = \frac{\sum_{i=1}^{n_k} X_{ki}}{n_k},$$

$$s_k = \sqrt{\left( \frac{\sum_{i=1}^{n_k} (X_{ki} - \bar{X}_k)^2}{(n_k - 1)} \right)},$$

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)},$$

$$df = n_1 + n_2 - 2,$$

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

The null hypothesis is rejected if the computed p-value is less than a specified level (usually 0.05). Otherwise, no conclusion can be reached.

---

## Computing the Power

When  $\sigma_1 = \sigma_2 = \sigma$ , the power of the equal-variance  $t$ -test is calculated as follows.

1. Find  $t_\alpha$  such that  $1 - T_{df}(t_\alpha) = \alpha$ , where  $T_{df}(x)$  is the area to the left of  $x$  under a central- $t$  distribution with degrees of freedom,  $df = n_1 + n_2 - 2$ .
2. Calculate:  $\sigma_{\bar{x}} = \sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ .
3. Calculate the noncentrality parameter:  $\lambda = \frac{\delta - \delta_0}{\sigma_{\bar{x}}}$ .
4. Calculate:  $Power = 1 - T'_{df,\lambda}(t_\alpha)$ , where  $T'_{df,\lambda}(x)$  is the area to the left of  $x$  under a noncentral- $t$  distribution with degrees of freedom,  $df = n_1 + n_2 - 2$ , and noncentrality parameter,  $\lambda$ .

When solving for something other than power, **PASS** uses this same power calculation formulation, but performs a search to determine that parameter.

## Example 1 – Power Analysis

Suppose that a test is to be conducted to determine if a new cancer treatment adversely affects bone density. The adjusted mean bone density (AMBD) in the population of interest is 0.002300 gm/cm with a standard deviation of 0.000300 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 equivalence is set to 2.5% (0.0000575 gm/cm).

Following accepted procedure, the analysis will be a non-inferiority test using the  $t$ -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 10 and 800 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>
Alpha.....	<b>0.025</b>
Group Allocation .....	<b>Equal (N1 = N2)</b>
Sample Size Per Group .....	<b>10 50 100 200 300 500 600 800</b>
NIM (Non-Inferiority Margin) .....	<b>0.575 1.15</b>
$\delta$ (Actual Difference to Detect).....	<b>0</b>
$\sigma$ (Standard Deviation).....	<b>3</b>

## Two-Sample T-Tests for Non-Inferiority Assuming Equal Variance

## Output

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

---

**Numeric Results**


---

Solve For: **Power**  
 Test Type: Two-Sample Equal-Variance T-Test  
 Difference:  $\delta = \mu_1 - \mu_2 = \mu_T - \mu_R$   
 Higher Means Are: Better  
 Hypotheses:  $H_0: \delta \leq -NIM$  vs.  $H_1: \delta > -NIM$

---

Power	Sample Size			Non-Inferiority Margin -NIM	Mean Difference $\delta$	Standard Deviation $\sigma$	Alpha
	N1	N2	N				
0.06013	10	10	20	-0.575	0	3	0.025
0.15601	50	50	100	-0.575	0	3	0.025
0.27052	100	100	200	-0.575	0	3	0.025
0.48089	200	200	400	-0.575	0	3	0.025
0.64940	300	300	600	-0.575	0	3	0.025
0.85769	500	500	1000	-0.575	0	3	0.025
0.91295	600	600	1200	-0.575	0	3	0.025
0.96943	800	800	1600	-0.575	0	3	0.025
0.12553	10	10	20	-1.150	0	3	0.025
0.47524	50	50	100	-1.150	0	3	0.025
0.76957	100	100	200	-1.150	0	3	0.025
0.96885	200	200	400	-1.150	0	3	0.025
0.99681	300	300	600	-1.150	0	3	0.025
0.99998	500	500	1000	-1.150	0	3	0.025
1.00000	600	600	1200	-1.150	0	3	0.025
1.00000	800	800	1600	-1.150	0	3	0.025

---

Power The probability of rejecting a false null hypothesis when the alternative hypothesis is true.  
 N1 The sample size from group 1.  
 N2 The sample size from group 2.  
 N The total sample size from both groups.  $N = N_1 + N_2$ .  
 -NIM The magnitude and direction of the margin of non-inferiority. Since higher means are better, this value is negative and is the maximum distance below  $\mu_2$  that  $\mu_1$  can be and still conclude that group 1 is non-inferior to group 2.  
 $\delta$  The difference between the group means at which power and sample size calculations are made.  $\delta = \mu_1 - \mu_2$ .  
 $\sigma$  The assumed standard deviation for each of the two groups.  
 Alpha The probability of rejecting a true null hypothesis.

---

**Summary Statements**


---

A parallel two-group design will be used to test whether the Group 1 (treatment) mean ( $\mu_1$ ) is non-inferior to the Group 2 (reference) mean ( $\mu_2$ ), with a non-inferiority margin of -0.575 ( $H_0: \delta \leq -0.575$  versus  $H_1: \delta > -0.575$ ,  $\delta = \mu_1 - \mu_2$ ). The comparison will be made using a one-sided, two-sample equal-variance t-test, with a Type I error rate ( $\alpha$ ) of 0.025. The common standard deviation for both groups is assumed to be 3. To detect a difference in means of 0, with sample sizes of 10 in Group 1 and 10 in Group 2, the power is 0.06013.

---

## Two-Sample T-Tests for Non-Inferiority Assuming Equal Variance

## Dropout-Inflated Sample Size

Dropout Rate	Sample Size			Dropout-Inflated Enrollment Sample Size			Expected Number of Dropouts		
	N1	N2	N	N1'	N2'	N'	D1	D2	D
20%	10	10	20	13	13	26	3	3	6
20%	50	50	100	63	63	126	13	13	26
20%	100	100	200	125	125	250	25	25	50
20%	200	200	400	250	250	500	50	50	100
20%	300	300	600	375	375	750	75	75	150
20%	500	500	1000	625	625	1250	125	125	250
20%	600	600	1200	750	750	1500	150	150	300
20%	800	800	1600	1000	1000	2000	200	200	400

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.
N1, N2, and N	The evaluable sample sizes at which power is computed (as entered by the user). If N1 and N2 subjects are evaluated out of the N1' and N2' subjects that are enrolled in the study, the design will achieve the stated power.
N1', N2', and N'	The number of subjects that should be enrolled in the study in order to obtain N1, N2, and N evaluable subjects, based on the assumed dropout rate. N1' and N2' are calculated by inflating N1 and N2 using the formulas $N1' = N1 / (1 - DR)$ and $N2' = N2 / (1 - DR)$ , with N1' and N2' 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.)
D1, D2, and D	The expected number of dropouts. $D1 = N1' - N1$ , $D2 = N2' - N2$ , and $D = D1 + D2$ .

## Dropout Summary Statements

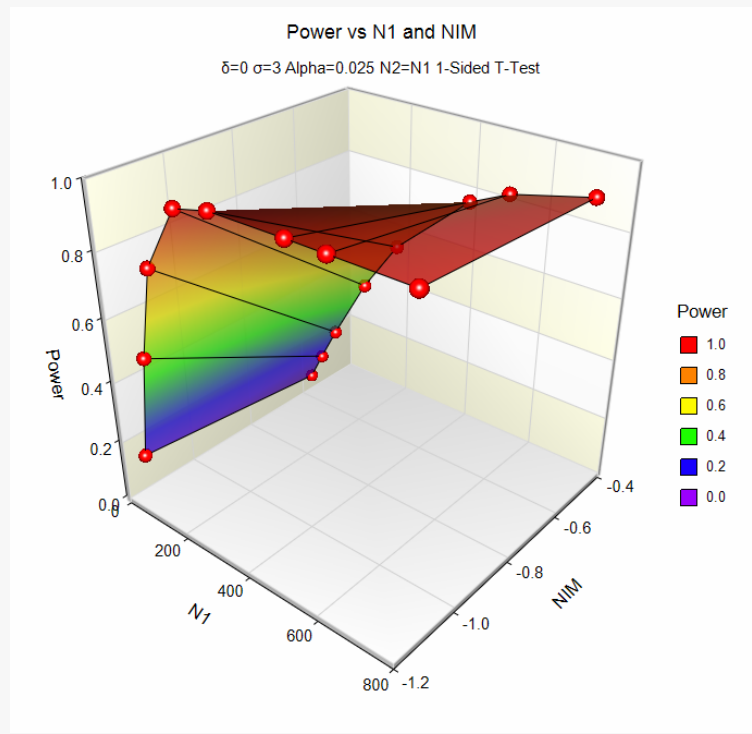
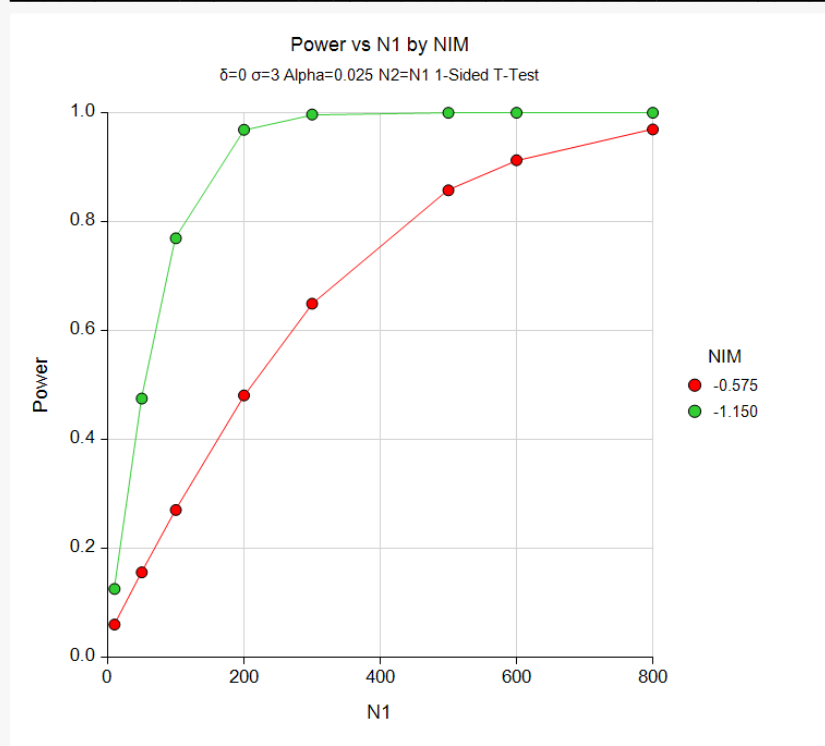
Anticipating a 20% dropout rate, 13 subjects should be enrolled in Group 1, and 13 in Group 2, to obtain final group sample sizes of 10 and 10, respectively.

## 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.

Two-Sample T-Tests for Non-Inferiority Assuming Equal Variance

Plots



The above report shows that for NIM = 1.15, the sample size necessary to obtain 90% power is about 150 per group. However, if NIM = 0.575, the required sample size is about 600 per group.



## Example 2 – Finding the Sample Size

Continuing with Example 1, the researchers want to know the exact sample size for each value of NIM to achieve 90% 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 .....	<b>Sample Size</b>
Higher Means Are .....	<b>Better (H1: <math>\delta &gt; -NIM</math>)</b>
Power.....	<b>0.90</b>
Alpha.....	<b>0.025</b>
Group Allocation .....	<b>Equal (N1 = N2)</b>
NIM (Non-Inferiority Margin) .....	<b>0.575 1.15</b>
$\delta$ (Actual Difference to Detect).....	<b>0</b>
$\sigma$ (Standard Deviation).....	<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>						
Test Type:		Two-Sample Equal-Variance T-Test						
Difference:		$\delta = \mu_1 - \mu_2 = \mu_T - \mu_R$						
Higher Means Are:		Better						
Hypotheses:		H0: $\delta \leq -NIM$ vs. H1: $\delta > -NIM$						
Power		Sample Size			Non-Inferiority Margin -NIM	Mean Difference $\delta$	Standard Deviation $\sigma$	Alpha
Target	Actual	N1	N2	N				
0.9	0.90036	573	573	1146	-0.575	0	3	0.025
0.9	0.90004	144	144	288	-1.150	0	3	0.025

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 53 has an example of a sample size calculation for a non-inferiority trial. Their example obtains a sample size of 51 in each group when  $\delta = 0$ , NIM = 0.05,  $\sigma = 0.1$ , 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 .....	<b>Sample Size</b>
Higher Means Are .....	<b>Better (H1: <math>\delta &gt; -NIM</math>)</b>
Power.....	<b>0.80</b>
Alpha.....	<b>0.05</b>
Group Allocation .....	<b>Equal (N1 = N2)</b>
NIM (Non-Inferiority Margin) .....	<b>0.05</b>
$\delta$ (Actual Difference to Detect).....	<b>0</b>
$\sigma$ (Standard Deviation).....	<b>0.1</b>

### Output

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

Numeric Results								
Solve For:		Sample Size						
Test Type:		Two-Sample Equal-Variance T-Test						
Difference:		$\delta = \mu_1 - \mu_2 = \mu_T - \mu_R$						
Higher Means Are:		Better						
Hypotheses:		H0: $\delta \leq -NIM$ vs. H1: $\delta > -NIM$						
Power		Sample Size			Non-Inferiority Margin	Mean Difference	Standard Deviation	Alpha
Target	Actual	N1	N2	N	-NIM	$\delta$	$\sigma$	
0.8	0.8059	51	51	102	-0.05	0	0.1	0.05

PASS also obtains a sample size of 51 per group.

## Example 4 – Validation using Julious (2004)

Julious (2004) page 1950 gives an example of a sample size calculation for a parallel, non-inferiority design. His example obtains a sample size of 336 when  $\delta = 0$ , NIM = 10,  $\sigma = 40$ , Alpha = 0.025, and Power = 0.90.

### 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 .....	<b>Sample Size</b>
Higher Means Are .....	<b>Better (H1: <math>\delta &gt; -NIM</math>)</b>
Power.....	<b>0.90</b>
Alpha.....	<b>0.025</b>
Group Allocation .....	<b>Equal (N1 = N2)</b>
NIM (Non-Inferiority Margin) .....	<b>10</b>
$\delta$ (Actual Difference to Detect).....	<b>0</b>
$\sigma$ (Standard Deviation).....	<b>40</b>

### Output

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

Numeric Results								
Solve For:		<a href="#">Sample Size</a>						
Test Type:		Two-Sample Equal-Variance T-Test						
Difference:		$\delta = \mu_1 - \mu_2 = \mu_T - \mu_R$						
Higher Means Are:		Better						
Hypotheses:		H0: $\delta \leq -NIM$ vs. H1: $\delta > -NIM$						
Power		Sample Size			Non-Inferiority Margin -NIM	Mean Difference $\delta$	Standard Deviation $\sigma$	Alpha
Target	Actual	N1	N2	N				
0.9	0.90045	337	337	674	-10	0	40	0.025

**PASS** obtained sample sizes of 337 in each group. The difference between 336 that Julious received and 337 that **PASS** calculated is likely caused by rounding.