

Chapter 160

Non-Inferiority Tests for the Difference Between Two Correlated Proportions

Introduction

This module computes power and sample size for non-inferiority tests of the difference in which two dichotomous responses are measured on each subject. When one is interested in showing that the true proportions are different, the data are often analyzed with McNemar's test. However, we are interested in showing non-inferiority rather than difference. For example, suppose a diagnostic procedure is accurate, but is expensive to apply or has serious side effects. A replacement procedure is sought which is no less accurate, but is less expensive or has fewer side effects. In this case, we are not interested in showing that the two diagnostic procedures are different, but rather that the second is no worse than the first. *Non-inferiority tests* were designed for this situation.

These tests are often divided into two categories: *equivalence* (two-sided) tests and *non-inferiority* (one-sided) tests. Here, the term *equivalence tests* means that we want to show that two diagnostic procedures are equivalent—that is, their accuracy is about the same. This requires a two-sided hypothesis test. On the other hand, *non-inferiority tests* are used when we want to show that a new (experimental) procedure is no worse than the existing (reference or gold-standard) one. This requires a one-sided hypothesis test. The procedures discussed in this chapter deal with the non-inferiority (one-sided) case.

Technical Details

The results of a study in which two dichotomous responses are measured on each subject can be displayed in a 2-by-2 table in which one response is shown across the columns and the other is shown down the rows. In the discussion to follow, the columns of the table represent the standard (reference or control) response and the rows represent the treatment (experimental) response. The outcome probabilities can be classified into the following table.

| Experimental Diagnosis | Standard Yes | Diagnosis No | Total |
|---------------------------|-----------------|-----------------|-----------|
| Yes | p_{11} | p_{10} | P_T |
| No | p_{01} | p_{00} | $1 - P_T$ |
| Total | P_S | $1 - P_S$ | 1 |

Non-Inferiority Tests for the Difference Between Two Correlated Proportions

In this table, $p_{ij} = p_{Treatment, Standard}$. That is, the first subscript represents the response of the new, experimental procedure while the second subscript represents the response of the standard procedure. Thus, p_{01} represents the proportion having a negative treatment response and a positive standard response.

Sensitivity, Specificity, and Prevalence

To aid in interpretation, analysts have developed a few proportions that summarize the table. Three of the most popular ratios are *sensitivity*, *specificity*, and *prevalence*.

Sensitivity

Sensitivity is the proportion of subjects with a positive standard response who also have a positive experimental response. In terms of proportions from a 2-by-2 table,

$$\text{Sensitivity} = p_{11} / (p_{01} + p_{11}) = p_{11} / P_S$$

Specificity

Specificity is the proportion of subjects with a negative standard response who also have a negative experimental response. In terms of proportions from a 2-by-2 table,

$$\text{Specificity} = p_{00} / (p_{10} + p_{00})$$

Prevalence

Prevalence is the overall proportion of individuals with the disease (or feature of interest). In terms of proportions from a 2-by-2 table,

$$\text{Prevalence} = P_T$$

Table Probabilities

The outcome counts from a sample of n subjects can be classified into the following table.

| Experimental Diagnosis | Standard Yes | Diagnosis No | Total |
|---------------------------|-----------------|-----------------|-----------|
| Yes | n_{11} | n_{10} | n_T |
| No | n_{01} | n_{00} | $n - n_T$ |
| Total | n_S | $n - n_S$ | n |

Note that $n_{11} + n_{00}$ is the number of matches (*concordant pairs*) and $n_{10} + n_{01}$ is the number of *discordant pairs*.

The hypothesis of interest concerns the two marginal probabilities P_T and P_S . P_S represents the accuracy or success of the standard test and P_T represents the accuracy or success of the new, experimental test. Non-inferiority is defined in terms of either the difference, $D = P_T - P_S$, or the relative risk ratio, $R = P_T / P_S$, of these two proportions. The choice between D and R will usually lead to different sample sizes to achieve the same power.

Non-Inferiority Hypotheses using Differences

This section is based on Liu, Hsueh, Hsieh and Chen (2002). Refer to that paper for complete details.

The null and alternative hypotheses of non-inferiority in terms of the difference are

$$H_0 : P_T - P_S \leq -D_E \text{ versus } H_1 : P_T - P_S > -D_E$$

To demonstrate non-inferiority, one desires to reject the null hypothesis and thus conclude that the experimental treatment is not worse than the standard by as much or more than D_E . In the context of the preceding statement, D_E is defined to be positive. The choice of an appropriate D_E may be difficult. It should be clinically meaningful so that clinicians are willing to conclude that the experimental treatment is acceptable if the difference is less than D_E . From a statistical perspective, D_E should be less than the effect, if known, of the standard treatment compared to placebo.

Liu et al. (2002) discuss the RMLE-based (score) method for constructing these confidence intervals. This method is based on (developed by, described by) Nam (1997).

Asymptotic Tests

An asymptotic test is given by

$$Z_L = \frac{\hat{D} + D_E}{\hat{\sigma}} = \frac{c + nD_E}{\sqrt{d - n\hat{D}^2}} \geq z_\alpha$$

where

$$\hat{D} = \frac{n_T}{n} - \frac{n_S}{n} = \frac{n_{10}}{n} - \frac{n_{01}}{n}$$

$$d = n_{10} + n_{01}$$

$$c = n_{10} - n_{01}$$

and z_α is the standard normal deviate having α in the right tail.

An estimate of $\hat{\sigma}$ based on the RMLE-based (score) procedure of Nam (1997) uses the estimates

$$\tilde{\sigma}_L = \sqrt{\frac{\tilde{p}_{L,10} + \tilde{p}_{L,01} - D_E^2}{n}}$$

where

$$\tilde{p}_{L,10} = \frac{-\tilde{a}_L + \sqrt{\tilde{a}_L^2 - 8\tilde{b}_L}}{4}$$

$$\tilde{p}_{L,01} = \tilde{p}_{L,10} - D_E$$

$$\tilde{a}_L = -\hat{D}(1 - D_E) - 2(\hat{p}_{01} + D_E)$$

$$\tilde{b}_L = D_E(1 + D_E)\hat{p}_{01}$$

Non-Inferiority Tests for the Difference Between Two Correlated Proportions

Power Formula

The power when the actual difference is D_A can be evaluated exactly using the multinomial distribution. However, when the sample size is above a user-set level, we use a normal approximation to this distribution which leads to

$$1 - \beta(D_A) = \begin{cases} 1 - \Phi(c_L) & \text{if } D_A > -D_E \\ 0 & \text{otherwise} \end{cases}$$

where

$$c_L = \frac{-D_A}{\sigma} - \frac{D_E}{\sigma} + \frac{z_\alpha}{w_L}$$

$$\sigma = \sqrt{\frac{p_{01} + p_{10} - D_A^2}{n}}$$

$$w_L = \sqrt{\frac{2p_{01} + D_A - D_A^2}{2\bar{p}_{L,01} - D_E - D_E^2}}$$

$$\bar{p}_{L,01} = \frac{-a_L + \sqrt{a_L^2 - 8b_L}}{4}$$

$$a_L = -D_A(1 - D_E) - 2(p_{01} + D_E)$$

$$b_L = D_E(1 + D_E)p_{01}$$

Nuisance Parameter

Unfortunately, the 2-by-2 table includes four parameters p_{11} , p_{10} , p_{01} , and p_{00} , but the power specifications above only specify two parameters: P_S and D_A or R_A . A third parameter is defined implicitly since the sum of the four parameters is one. One parameter, known as a nuisance parameter, remains unaccounted for. This parameter must be addressed to fully specify the problem. This fourth parameter can be specified by specifying any one of the following: p_{11} , p_{10} , p_{01} , p_{00} , $p_{10} + p_{01}$, $p_{11} + p_{00}$, or the sensitivity of the experimental response, p_{11} / P_S .

It may be difficult to specify a reasonable value for the nuisance parameter since its value may not be even approximately known until after the study is conducted. Because of this, we suggest that you calculate power or sample size for a range of values of the nuisance parameter. This will allow you to determine how sensitive the results are to its value.

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 the parameters associated with this test such as the proportions, sample sizes, alpha, and power.

Solve For

Solve For

This option specifies the parameter to be solved for from the other parameters. The parameters that may be selected are *Power* or *Sample Size*.

Power Calculation

Power Calculation Method

Select the method to be used to calculate power.

The choices are

- **Multinomial Enumeration**

Power is computed using multinomial enumeration of all possible outcomes when $N \leq \text{Max } N$ for Multinomial Enumeration (otherwise, the normal approximation is used). Multinomial enumeration of all outcomes is possible because of the discrete nature of the data.

- **Normal Approximation**

Approximate power is computed using the normal approximation to the multinomial distribution.

The exact calculation using the multinomial distribution becomes very time consuming for $N > 500$. When $N > 500$, the difference between the multinomial and approximate calculations is small.

For small values of N (less than 100), the Multinomial Enumeration power may be overly optimistic because the discrete nature of the trinomial distribution results in the actual alpha value being higher than its target. To be on the safe side, we recommend that you use the approximate calculation.

Max N for Multinomial Enumeration

Only shown when Power Calculation Method = "Multinomial Enumeration"

Specify the maximum value of N (sample size) that uses the exact power calculation based on the multinomial distribution. N 's greater than this value will use the asymptotic approximation.

Non-Inferiority Tests for the Difference Between Two Correlated Proportions

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. Here, a type-II error occurs when you fail to conclude non-inferiority when in fact it is true.

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. Here, a type-I error occurs when you falsely conclude non-inferiority.

Sample Size

N (Sample Size)

Enter a value for the sample size. This value must be greater than two. You may enter a range of values such as *10 to 100 by 10*.

Effect Size – Differences

|Dni| (Non-Inferiority Difference)

|Dni| is the maximum allowable difference between the standard and treatment proportions that will still result in the conclusion of non-inferiority.

Only positive values can be entered here. Typical values for this difference are 0.05, 0.10, and 0.20. For 1-sided tests, you must have $D_a > -|D_{ni}|$.

Da (Actual Difference)

Enter D_a , the actual difference between the standard and treatment proportions ($P_t - P_s$). This value is used to generate the value of P_t using the formula $P_t = P_s + D_a$.

Often this value is set equal to zero, but this is not necessary. For 1-sided tests, you must have $D_a > -|D_{ni}|$.

Effect Size – Standard Proportion

Ps (Standard Proportion)

This is the proportion of yes's (or successes), P_s , when subjects received the standard treatment. This value or a good estimate is often available from previous studies.

You may enter a set of values separated by blank spaces. For example, you could enter '0.50 0.60 0.70'. Values between, but not including, 0 and 1 are permitted.

Non-Inferiority Tests for the Difference Between Two Correlated Proportions

Effect Size – Nuisance Parameter

Nuisance Parameter Type

Enter the type of nuisance parameter here. Unfortunately, the 2-by-2 table cannot be completely specified by using only the parameters P_s and D_a or P_s and R_a . One other parameter must be specified. This additional parameter is called a “nuisance” parameter. It will be assumed to be a known quantity. Several possible choices are available. This option lets you specify which parameter you want to use. In all cases, the value you specify is a proportion.

- **P11 (% Positive Matches)**
The proportion of subjects that are positive on both tests.
- **P00 (% Negative Matches)**
The proportion of subjects that are negative on both tests.
- **P01 (% -Trt +Std)**
The proportion of subjects that are negative on the treatment, but positive on the standard.
- **P10 (% +Trt -Std)**
The proportion of subjects that are positive on the treatment, but negative on the standard.
- **P11+P00 (% Matches)**
The proportion of matches (concordant pairs).
- **P01+P10 (% Disagree)**
The proportion of non-matches (discordant pairs).
- **P11/ P_s (Sensitivity)**
The sensitivity.

Nuisance Parameter Value

Enter the value of the nuisance parameter that you specified in the “Nuisance Parameter Type” box. This value is a proportion, so it must be between 0 and 1.

Non-Inferiority Tests for the Difference Between Two Correlated Proportions

Example 1 – Finding Power

A clinical trial will be conducted to show that a non-invasive MRI test is not inferior to the invasive CTAP reference test. Historical data suggest that the CTAP test is 80% accurate. After careful discussion, the researchers decide that if the MRI test is 75% accurate or better, it will be considered non-inferior. They decide to use a difference test statistic. Thus, the non-inferiority difference is 0.05. They want to study the power for various sample sizes between 20 and 1000 at the 5% significance level.

They use P01 as the nuisance parameter and look at two values: 0.05 and 0.10.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Non-Inferiority Tests for the Difference Between Two Correlated Proportions** procedure window by expanding **Proportions**, then **Two Correlated Proportions**, then clicking on **Non-Inferiority**, and then clicking on **Non-Inferiority Tests for the Difference Between Two Correlated Proportions**. 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**.

| Option | Value |
|---|--|
| Design Tab | |
| Solve For | Power |
| Power Calculation Method..... | Normal Approximation |
| Alpha..... | 0.05 |
| N (Sample Size)..... | 20 100 200 300 450 600 800 1000 |
| Dni (Non-Inferiority Difference) | 0.05 |
| Da (Actual Difference) | 0.0 |
| Ps (Standard Proportion) | 0.80 |
| Nuisance Parameter Type | P01 (% -Trt +Std) |
| Nuisance Parameter Value..... | 0.05 0.10 |

Annotated Output

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

Numeric Results

| Numeric Results for a Non-Inferiority (One-Sided) Test of a Difference | | | | | | | |
|--|-------------|---------------------|-------------------|----------------------|---------------------|--------------------|-------|
| | Sample Size | Non-Inf. Difference | Actual Difference | Treatment Proportion | Standard Proportion | Nuisance Parameter | Alpha |
| Power | N | Dni | Da | Pt | Ps | P01 | |
| 0.14284 | 20 | 0.050 | 0.000 | 0.800 | 0.800 | 0.050 | 0.050 |
| 0.12028 | 20 | 0.050 | 0.000 | 0.800 | 0.800 | 0.100 | 0.050 |
| 0.42323 | 100 | 0.050 | 0.000 | 0.800 | 0.800 | 0.050 | 0.050 |
| 0.28926 | 100 | 0.050 | 0.000 | 0.800 | 0.800 | 0.100 | 0.050 |
| 0.67771 | 200 | 0.050 | 0.000 | 0.800 | 0.800 | 0.050 | 0.050 |
| 0.46318 | 200 | 0.050 | 0.000 | 0.800 | 0.800 | 0.100 | 0.050 |
| 0.83244 | 300 | 0.050 | 0.000 | 0.800 | 0.800 | 0.050 | 0.050 |
| 0.60369 | 300 | 0.050 | 0.000 | 0.800 | 0.800 | 0.100 | 0.050 |
| 0.94287 | 450 | 0.050 | 0.000 | 0.800 | 0.800 | 0.050 | 0.050 |
| 0.75745 | 450 | 0.050 | 0.000 | 0.800 | 0.800 | 0.100 | 0.050 |
| 0.98206 | 600 | 0.050 | 0.000 | 0.800 | 0.800 | 0.050 | 0.050 |
| 0.85657 | 600 | 0.050 | 0.000 | 0.800 | 0.800 | 0.100 | 0.050 |

Non-Inferiority Tests for the Difference Between Two Correlated Proportions

| | | | | | | | |
|---------|------|-------|-------|-------|-------|-------|-------|
| 0.99651 | 800 | 0.050 | 0.000 | 0.800 | 0.800 | 0.050 | 0.050 |
| 0.93172 | 800 | 0.050 | 0.000 | 0.800 | 0.800 | 0.100 | 0.050 |
| 0.99937 | 1000 | 0.050 | 0.000 | 0.800 | 0.800 | 0.050 | 0.050 |
| 0.96870 | 1000 | 0.050 | 0.000 | 0.800 | 0.800 | 0.100 | 0.050 |

* Power was computed using the normal approximation method.

Report Definitions

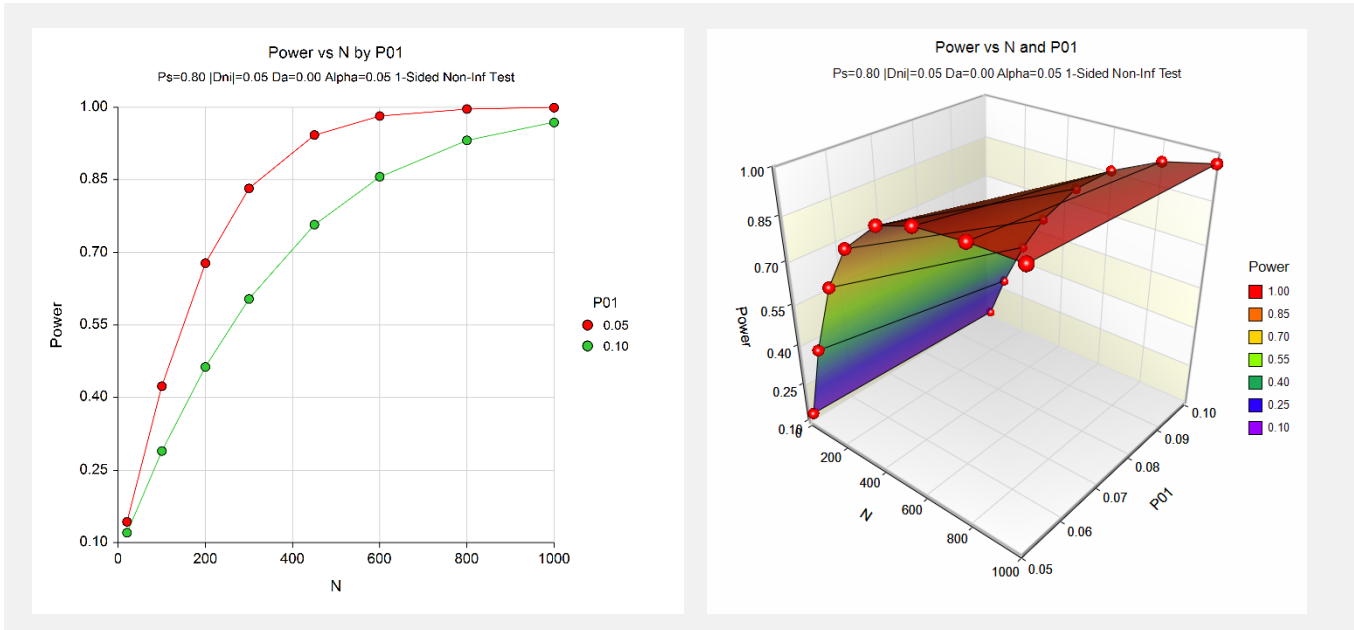
Power is the probability of rejecting a false null hypothesis.
 N is the number of subjects, the sample size.
 |Dni| is the maximum difference between the two proportions that is still called 'non-inferior.'
 Da is the actual difference between Pt and Ps. That is, $Da = Pt - Ps$.
 Pt is the response proportion to the treatment (experimental or new) test.
 Ps is the response proportion to the standard (reference or old) test.
 The Nuisance Parameter is a value that is needed, but is not a direct part of the hypothesis.
 Alpha is the probability of rejecting a true null hypothesis.

Summary Statements

A sample size of 20 subjects achieves 14.284% power at a significance level of 0.050 using a one-sided non-inferiority test of correlated proportions when the standard proportion is 0.800, the maximum allowable difference between these proportions that still results in non-inferiority (the range of non-inferiority) is 0.050, and the actual difference of the proportions is 0.000.

This report shows the power for the indicated scenarios. All of the columns are defined in the 'Report Definitions' section.

Plots Section



These plots show the power versus the sample size for the two values of P01. In this example, we see that the value of the nuisance parameter has a large effect on the calculated power.

Non-Inferiority Tests for the Difference Between Two Correlated Proportions

Example 2 – Finding Sample Size

Continuing with Example 1, the analysts want to determine the exact sample size necessary to achieve 90% power for both values of the nuisance parameter.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Non-Inferiority Tests for the Difference Between Two Correlated Proportions** procedure window by expanding **Proportions**, then **Two Correlated Proportions**, then clicking on **Non-Inferiority**, and then clicking on **Non-Inferiority Tests for the Difference Between Two Correlated Proportions**. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

| <u>Option</u> | <u>Value</u> |
|---|-----------------------------|
| Design Tab | |
| Solve For | Sample Size |
| Power Calculation Method..... | Normal Approximation |
| Power..... | 0.90 |
| Alpha..... | 0.05 |
| Dni (Non-Inferiority Difference) | 0.05 |
| Da (Actual Difference) | 0.0 |
| Ps (Standard Proportion) | 0.80 |
| Nuisance Parameter Type | P01 (% -Trt +Std) |
| Nuisance Parameter Value..... | 0.05 0.10 |

Output

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

Numeric Results

| Numeric Results for a Non-Inferiority (One-Sided) Test of a Difference | | | | | | | |
|--|---------------------|--------------------------------|----------------------------|-------------------------------|------------------------------|------------------------------|-------|
| | Sample Size N | Non-Inf. Difference Dni | Actual Difference Da | Treatment Proportion Pt | Standard Proportion Ps | Nuisance Parameter P01 | Alpha |
| Power | 374 | 0.050 | 0.000 | 0.800 | 0.800 | 0.050 | 0.050 |
| | 699 | 0.050 | 0.000 | 0.800 | 0.800 | 0.100 | 0.050 |
| * Power was computed using the normal approximation method. | | | | | | | |

This report shows that the sample size required nearly doubles when P01 is changed from 0.05 to 0.10.

Non-Inferiority Tests for the Difference Between Two Correlated Proportions

Example 3 – Validation using Liu et al. (2002)

Liu *et al.* (2002) give an example in which P01 is 0.05, P10 is 0.05, Da is 0.00, Dni is 0.05, the significance level is 0.025, and the power is 80%. From their Table III, the sample size is 350.

In this example, the value of Ps is arbitrary. We set it at 0.50.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Non-Inferiority Tests for the Difference Between Two Correlated Proportions** procedure window by expanding **Proportions**, then **Two Correlated Proportions**, then clicking on **Non-Inferiority**, and then clicking on **Non-Inferiority Tests for the Difference Between Two Correlated Proportions**. You may then make the appropriate entries as listed below, or open **Example 3** by going to the **File** menu and choosing **Open Example Template**.

| <u>Option</u> | <u>Value</u> |
|---|-----------------------------|
| Design Tab | |
| Solve For | Sample Size |
| Power Calculation Method..... | Normal Approximation |
| Power..... | 0.80 |
| Alpha..... | 0.025 |
| Dni (Non-Inferiority Difference) | 0.05 |
| Da (Actual Difference) | 0.0 |
| Ps (Standard Proportion) | 0.50 |
| Nuisance Parameter Type | P01 (% -Trt +Std) |
| Nuisance Parameter Value..... | 0.05 |

Output

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

Numeric Results

| Numeric Results for a Non-Inferiority (One-Sided) Test of a Difference | | | | | | | |
|--|---------------------|--------------------------------|----------------------------|-------------------------------|------------------------------|------------------------------|-------|
| Power | Sample Size N | Non-Inf. Difference Dni | Actual Difference Da | Treatment Proportion Pt | Standard Proportion Ps | Nuisance Parameter P01 | Alpha |
| 0.80046 | 350 | 0.050 | 0.000 | 0.500 | 0.500 | 0.050 | 0.025 |
| * Power was computed using the normal approximation method. | | | | | | | |

The calculated sample size of 350 matches the results of Liu et al. (2002).