Chapter 415

Non-Inferiority Tests for One Mean

Introduction

This module computes power and sample size for non-inferiority tests in one-sample designs in which the outcome is distributed as a normal random variable. This includes the analysis of the differences between paired values.

The details of sample size calculation for the one-sample design are presented in the Inequality Tests for One Mean 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. (2003) page 50.

The *one-sample t-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 t-test*. This module also calculates the power of the nonparametric analog of the t-test, the *Wilcoxon 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 t-test and the Wilcoxon signed-rank test are appropriate for paired data even when the distributions of the individual items are not normal.

In paired designs, the variable of interest is the difference between two individual measurements. Although the non-inferiority hypothesis refers to the difference between two individual means, the actual values of those means are not needed. All that is needed is their difference.

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 T-Test procedure. However, at the urging of our users, we have developed this module which provides the input and output options that are convenient for non-inferiority tests. This section will review the specifics of non-inferiority and superiority testing.

Remember that in the usual t-test setting, the null (H0) and alternative (H1) hypotheses for one-sided tests are defined as

$$H_0: \mu_X \leq A \text{ versus } H_1: \mu_X > A$$

Rejecting H0 implies that the mean is larger than the value *A*. This test is called an *upper-tail test* because H0 is rejected in samples in which the sample mean is larger than *A*.

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

$$H_0: \mu_X \geq A \text{ versus } H_1: \mu_X < A$$

Non-inferiority tests are special cases of the above directional tests. It will be convenient to adopt the following specialize notation for the discussion of these tests.

Parameter	PASS Input/Output	Interpretation
μ_T	Not used	*Population mean*. If the data are paired differences, this is the mean of those differences.
μ_R	Not used	*Reference value*. Usually, this is the mean of a reference population. If the data are paired differences, this is the hypothesized value of the mean difference.
M_{NI}	NIM	*Margin of non-inferiority*. 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 from the reference value that is considered to be trivial. The absolute value symbols are used to emphasize that this is a magnitude. The sign is determined by the specific design.
δ	D	*True difference*. This is the value of $\mu_T - \mu_R$, the difference between the mean and the reference value, at which the power is calculated.

Note that the actual values of μ_T and μ_R are not needed. Only their difference is needed for power and sample size calculations.

Non-Inferiority Tests

A *non-inferiority test* tests that the mean is not worse than that of the baseline (reference) population by more than a small non-inferiority margin. The actual direction of the hypothesis depends on the 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 of the treatment group is no less than a small amount below the reference value. The value of δ is often set to zero. Equivalent sets of the null and alternative hypotheses are

$$H_0 : \mu_T \leq \mu_R - |M_{NI}| \qquad \text{versus} \qquad H_1 : \mu_T > \mu_R - |M_{NI}|$$

$$H_0 : \mu_T - \mu_R \leq -|M_{NI}| \qquad \text{versus} \qquad H_1 : \mu_T - \mu_R > -|M_{NI}|$$

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

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 δ is often set to zero. Equivalent sets of the null and alternative hypotheses are

$$H_0 : \mu_T \geq \mu_R + |M_{NI}| \qquad \text{versus} \qquad H_1 : \mu_T < \mu_R + |M_{NI}|$$

$$H_0 : \mu_T - \mu_R \geq |M_{NI}| \qquad \text{versus} \qquad H_1 : \mu_T - \mu_R < |M_{NI}|$$

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

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 the 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 AMBD in the treated group is greater than 0.002300-0.000115 = 0.002185. The statistical test will be set up so that if the null hypothesis that the AMBD is less than or equal to 0.002185 is rejected, the conclusion will be that the new treatment is non-inferior, at least in terms of AMBD. The value 0.000115 gm/cm is called the *margin of non-inferiority*.

Test Statistics

This section describes the test statistics that are available in this procedure.

One-Sample T-Test

The one-sample t-test assumes that the data are a simple random sample from a 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 t-test proceeds as follow

$$t_{n-1} = \frac{\overline{X} - D0}{s_{\overline{X}}}$$

where

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

$$s_{\overline{X}} = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1}},$$

and $D0$ is the value of the mean hypothesized by the null hypothesis.

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.

Wilcoxon Signed-Rank Test

The Wilcoxon signed-rank test is a popular, nonparametric substitute for the t-test. It assumes that the data follow a symmetric distribution. The test is computed using the following steps.

1. Subtract the hypothesized mean, $D0$, from each data value. Rank the values according to their absolute values.

2. Compute the sum of the positive ranks Sp and the sum of the negative ranks Sn. The test statistic, W, is the minimum of Sp and Sn.

3. Compute the mean and standard deviation of W using the formulas

$$\mu_{W_n} = \frac{n(n+1)}{4} \quad \text{and} \quad \sigma_{W_n} = \sqrt{\frac{n(n+1)(2n+1)}{24} - \frac{\sum t^3 - \sum t}{48}}$$

   where t represents the number of times the ith value occurs.

4. Compute the z value using

$$z_W = \frac{W - \mu_{W_n}}{\sigma_{W_n}}$$

The significance of the test statistic is determined by computing the p-value using the standard normal distribution. If this p-value is less than a specified level (usually 0.05), the null hypothesis is rejected in favor of the alternative hypothesis. Otherwise, no conclusion can be reached.

Computing the Power

The power is calculated as follows for a directional alternative (one-tailed test) in which $D1 > D0$. $D1$ is the value of the mean at which the power is computed.

1. Find t_α such that $1 - T_{n-1}(t_\alpha) = \alpha$, where $T_{n-1}(t_\alpha)$ is the area to the left of x under a central-t curve with $n - 1$ degrees of freedom.

2. Calculate $x_a = D0 + t_\alpha \dfrac{\sigma}{\sqrt{n}}$.

3. Calculate the noncentrality parameter $\lambda = \dfrac{D1 - D0}{\dfrac{\sigma}{\sqrt{n}}}$.

4. Calculate $t_a = \dfrac{x_a - D1}{\dfrac{\sigma}{\sqrt{n}}} + \lambda$.

5. Calculate the power $= 1 - T'_{n-1,\lambda}(t_a)$, where $T'_{n-1,\lambda}(x)$ is the area to the left of x under a noncentral-t curve with degrees of freedom $n - 1$ and noncentrality parameter λ.

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 will be of interest.

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.

Test

Higher Means Are

This option defines whether higher values of the response variable are to be considered better or worse.

The choice here determines the direction of the non-inferiority test.

If Higher Means Are Better the null hypothesis is Diff ≤ -NIM and the alternative hypothesis is Diff > -NIM. If Higher Means Are Worse the null hypothesis is Diff ≥ NIM and the alternative hypothesis is Diff < NIM.

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.

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, σ is the original standard deviation, and σ_1 is the new standard deviation.

The quantity n/N is often called the sampling fraction. The quantity $\left(1 - \dfrac{n}{N}\right)$ 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 individuals in the study. This value must be an integer greater than one. You may enter a list of values using the syntax *50 100 150 200 250* or *50 to 250 by 50*.

Effect Size – Mean Difference

NIM (Non-Inferiority Margin)

This is the magnitude of the margin of non-inferiority. It must be entered as a positive number.

When higher means are better, this value is the distance below the reference value that is still considered non-inferior. When higher means are worse, this value is the distance above the reference value that is still considered non-inferior.

D (True Value)

This is the actual difference between the mean and the reference value.

For non-inferiority tests, this value is often set to zero. When this value is non-zero, care should be taken that this value is consistent with whether higher means are better or worse.

Effect Size – Standard Deviation

Standard Deviation

This option specifies one or more values of the standard deviation. This must be a positive value. **PASS** includes a special module for estimating the standard deviation. This module may be loaded by pressing the *SD* button. Refer to the Standard Deviation Estimator chapter for further details.

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 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 non-inferiority 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 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

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 One Mean** procedure window by expanding **Means**, then **One Mean**, then clicking on **Non-Inferiority**, and then clicking on **Non-Inferiority Tests for One Mean**. 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**
Higher Means Are....................................	**Better**
Nonparametric Adjustment	**Ignore**
Population Size..	**Infinite**
Alpha...	**0.025**
N (Sample Size).......................................	**20 40 60 80 100 150 200 300**
NIM (Non-Inferiority Margin)	**0.575 1.15**
D (True Difference)	**0**
S (Standard Deviation)	**3**

Annotated Output

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

Numeric Results for Non-Inferiority Test (H0: Diff ≤ -NIM; H1: Diff > -NIM)
Higher Means are Better
Test Statistic: T-Test

Power	N	Non-Inferiority Margin (-NIM)	Actual Difference (D)	Significance Level (Alpha)	Beta	Standard Deviation (S)
0.12601	20	-0.575	0.000	0.02500	0.87399	3.000
0.21844	40	-0.575	0.000	0.02500	0.78156	3.000
0.30873	60	-0.575	0.000	0.02500	0.69127	3.000
0.39493	80	-0.575	0.000	0.02500	0.60507	3.000
0.47532	100	-0.575	0.000	0.02500	0.52468	3.000
0.64517	150	-0.575	0.000	0.02500	0.35483	3.000
0.76959	200	-0.575	0.000	0.02500	0.23041	3.000
0.91262	300	-0.575	0.000	0.02500	0.08738	3.000
0.36990	20	-1.150	0.000	0.02500	0.63010	3.000
0.65705	40	-1.150	0.000	0.02500	0.34295	3.000
0.83164	60	-1.150	0.000	0.02500	0.16836	3.000
(report continues)						

Non-Inferiority Tests for One Mean

Report Definitions
Power is the probability of rejecting H0 when it is false.
N is the sample size, the number of subjects (or pairs) in the study.
-NIM is 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.
D is the mean difference (treatment - reference value) at which the power is computed.
Alpha is the probability of rejecting H0 when it is true, which is the probability of a false positive.
Beta is the probability of accepting H0 when it is false, which is the probability of a false negative.
S is the standard deviation of the response. It measures the variability in the population.

Summary Statements
A sample size of 20 achieves 13% power to detect non-inferiority using a one-sided t-test when
the margin of equivalence is -0.575 and the true difference between the mean and the reference
value is 0.000. The data are drawn from a single population with a standard deviation of 3.000.
The significance level (alpha) of the test is 0.02500.

Non-Inferiority Tests for One Mean

Power vs N and NIM
D=0.000 S=3.000 Alpha=0.025 T Test
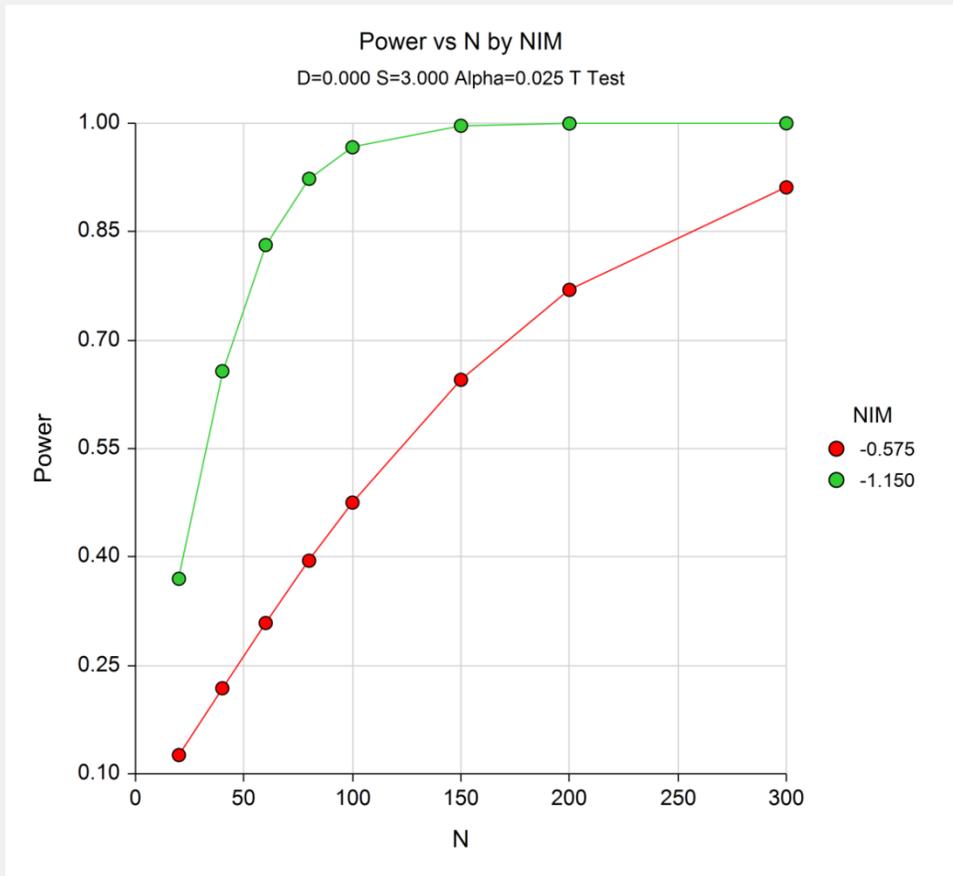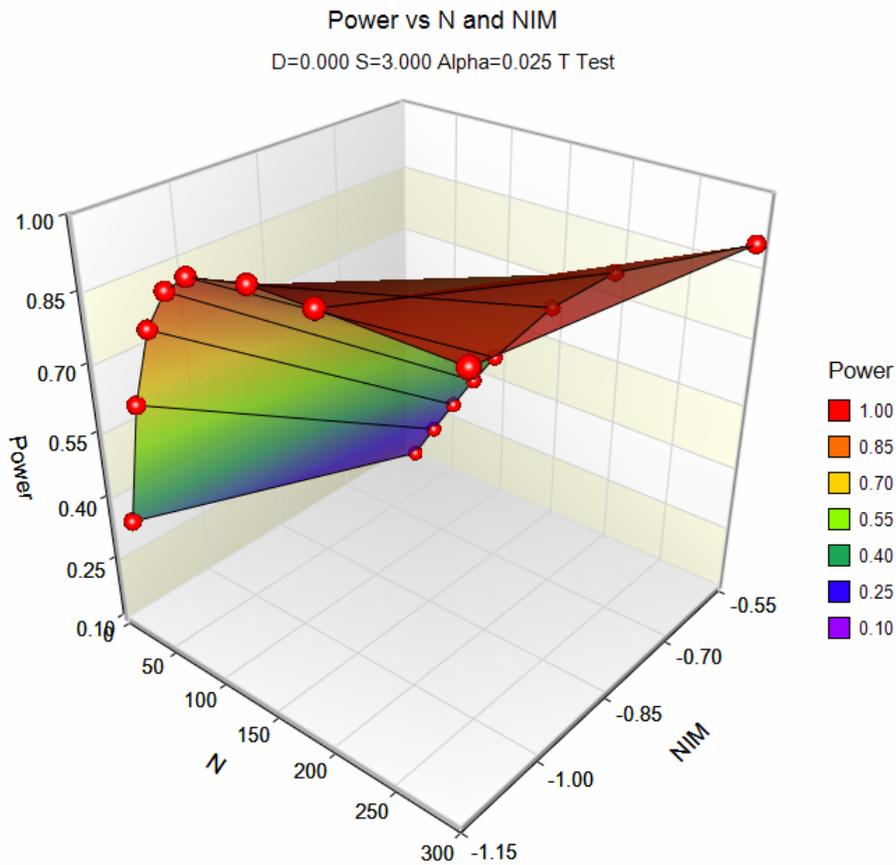
The above report shows that for NIM = 1.15, the sample size necessary to obtain 90% power is just under 80. However, if NIM = 0.575, the required sample size is about 300.

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

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 One Mean** procedure window by expanding **Means**, then **One Mean**, then clicking on **Non-Inferiority**, and then clicking on **Non-Inferiority Tests for One Mean**. 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**.

Option	**Value**
Design Tab	
Solve For	**Sample Size**
Higher Means Are....................................	**Better**
Nonparametric Adjustment	**Ignore**
Population Size......................................	**Infinite**
Power...	**0.90**
Alpha..	**0.025**
NIM (Non-Inferiority Margin)	**0.575 1.15**
D (True Difference)	**0**
S (Standard Deviation)	**3**

Output

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

Numeric Results

Numeric Results for Non-Inferiority Test (H0: Diff ≤ -NIM; H1: Diff > -NIM)
Higher Means are Better
Test Statistic: T-Test

Power	N	Non-Inferiority Margin (-NIM)	Actual Difference (D)	Significance Level (Alpha)	Beta	Standard Deviation (S)
0.90005	288	-0.575	0.000	0.02500	0.09995	3.000
0.90215	74	-1.150	0.000	0.02500	0.09785	3.000

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

Example 3 – Validation using Chow

Chow, Shao, Wang (2003) pages 54-55 has an example of a sample size calculation for a non-inferiority trial. Their example obtains a sample size of 8 when D = 0.5, NIM = 0.5, S = 1, Alpha = 0.05, and Beta = 0.20.

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 One Mean** procedure window by expanding **Means**, then **One Mean**, then clicking on **Non-Inferiority**, and then clicking on **Non-Inferiority Tests for One Mean**. 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**.

Option	Value
Design Tab	
Solve For	**Sample Size**
Higher Means Are	**Better**
Nonparametric Adjustment	**Ignore**
Population Size	**Infinite**
Power	**0.80**
Alpha	**0.05**
NIM (Non-Inferiority Margin)	**0.5**
D (True Difference)	**0.5**
S (Standard Deviation)	**1**

Output

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

Numeric Results for Non-Inferiority Test (H0: Diff ≤ -NIM; H1: Diff > -NIM)
Higher Means are Better
Test Statistic: T-Test

Power	N	Non-Inferiority Margin (-NIM)	Actual Difference (D)	Significance Level (Alpha)	Beta	Standard Deviation (S)
0.81502	8	-0.500	0.500	0.05000	0.18498	1.000

PASS also obtains a sample size of 8.

Example 4 – Validation of a Cross-Over Design given in Julious

Julious (2004) page 1953 gives an example of a sample size calculation for a cross-over design. His example obtains a sample size of 87 when D = 0, NIM = 10, S = 28.28427, Alpha = 0.025, and Beta = 0.10. When D is changed to 2, the resulting sample size is 61.

Note that in Julius's example, the population standard deviation is given as 20. Assuming that the correlation between items in a pair is 0, the standard deviation of the difference is calculated to be

$S = \sqrt{20^2 + 20^2 - (0)(20)(20)} = 28.284271$. Actually, the value of S probably should be less because the correlation is usually greater than 0 (at least 0.2).

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 One Mean** procedure window by expanding **Means**, then **One Mean**, then clicking on **Non-Inferiority**, and then clicking on **Non-Inferiority Tests for One Mean**. You may then make the appropriate entries as listed below, or open **Example 4** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Design Tab	
Solve For	**Sample Size**
Higher Means Are.....................................	**Better**
Nonparametric Adjustment	**Ignore**
Population Size...	**Infinite**
Power...	**0.90**
Alpha..	**0.025**
NIM (Non-Inferiority Margin)	**10**
D (True Difference)...................................	**0 2**
S (Standard Deviation)	**28.284271**

Output

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

Numeric Results for Non-Inferiority Test (H0: Diff ≤ -NIM; H1: Diff > -NIM)
Higher Means are Better
Test Statistic: T-Test

Power	N	Non-Inferiority Margin (-NIM)	Actual Difference (D)	Significance Level (Alpha)	Beta	Standard Deviation (S)
0.90332	87	-10.000	0.000	0.02500	0.09668	28.284
0.90323	61	-10.000	2.000	0.02500	0.09677	28.284

PASS also obtains sample sizes of 87 and 61.

Example 5 – Validation of a Cross-Over Design given in Chow, Shao, and Wang

Chow, Shao, and Wang (2004) page 67 give an example of a sample size calculation for a cross-over design. Their example calculates sample sizes of 13 and 14 (13 by formula and 14 from their table) in each sequence (26 or 28 total) when D = -0.1, NIM = 0.2, S = 0.2, Alpha = 0.05, and Beta = 0.20.

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 One Mean** procedure window by expanding **Means**, then **One Mean**, then clicking on **Non-Inferiority**, and then clicking on **Non-Inferiority Tests for One Mean**. You may then make the appropriate entries as listed below, or open **Example 5** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Design Tab	
Solve For	**Sample Size**
Higher Means Are...................................	**Better**
Nonparametric Adjustment	**Ignore**
Population Size.......................................	**Infinite**
Power..	**0.80**
Alpha...	**0.05**
NIM (Non-Inferiority Margin)	**0.2**
D (True Difference).................................	**-0.1**
S (Standard Deviation)	**0.2**

Output

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

Numeric Results for Non-Inferiority Test (H0: Diff ≤ -NIM; H1: Diff > -NIM)
Higher Means are Better
Test Statistic: T-Test

Power	N	Non-Inferiority Margin (-NIM)	Actual Difference (D)	Significance Level (Alpha)	Beta	Standard Deviation (S)
0.81183	27	-0.200	-0.100	0.05000	0.18817	0.200

PASS obtains a sample size of 27, which is between the values of 26 and 28 that were obtained by Chow et al.