Chapter 447

One-Sample Z-Tests for Non-Inferiority

Introduction

This procedure computes power and sample size for non-inferiority tests in one-sample designs in which the outcome is distributed as a normal random variable and the standard deviation is known. 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 One-Sample 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.

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 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. This section will review the specifics of non-inferiority testing.

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

$$H_0 : \mu \leq \mu_0 \quad \text{versus} \quad H_1 : \mu > \mu_0$$

Rejecting H_0 implies that the mean is larger than the value μ_0. This test is called an *upper-tail test* because H_0 is rejected in samples in which the sample mean is larger than μ_0.

The *lower-tail test* is

$$H_0: \mu \geq \mu_0 \quad \text{versus} \quad H_1: \mu < \mu_0$$

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	μ	*Population mean*. If the data are paired differences, this is the mean of those differences. This parameter will be estimated by the study.
μ_1	$\mu 1$	*Actual population mean at which power is calculated*. This is the assumed population mean used in all calculations.
μ_0	$\mu 0$	*Non-Inferiority Mean*. This is the smallest (or largest) value of the mean for which the new treatment will still be considered non-inferior to the reference.
μ_R	μR	*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. This value is assumed to be a positive number.

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 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 μ_1 at which power is calculated is often set to zero for paired tests.

Equivalent sets of the null and alternative hypotheses are

$H_0: \mu \leq \mu_0$	versus	$H_1: \mu > \mu_0$
$H_0: \mu \leq \mu_R - M_{NI}$	versus	$H_1: \mu > \mu_R - M_{NI}$
$H_0: \mu - \mu_R \leq -M_{NI}$	versus	$H_1: \mu - \mu_R > -M_{NI}$

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 μ_1 at which power is calculated is often set to zero for paired tests.

Equivalent sets of the null and alternative hypotheses are

$H_0: \mu \geq \mu_0$ versus $H_1: \mu < \mu_0$

$H_0: \mu \geq \mu_R + M_{NI}$ versus $H_1: \mu < \mu_R + M_{NI}$

$H_0: \mu - \mu_R \geq M_{NI}$ versus $H_1: \mu - \mu_R < M_{NI}$

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.*

One-Sample Z-Test Statistic

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

$$z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

where

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

and μ_0 is the value of the mean hypothesized by the null hypothesis that incorporates both μ_R and 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.

Power Calculation for the One-Sample Z-Test

When the standard deviation is known, the power is calculated as follows for a directional alternative (one-tailed test) in which $\mu_1 > \mu_0$.

1. Find z_α 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: $X_1 = \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$.

3. Calculate: $z_1 = \frac{X_1 - \mu_1}{\frac{\sigma}{\sqrt{n}}}$.

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 (μ_R) 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% (M_{NI} = 0.000115 gm/cm such that $\mu_0 = \mu_R - M_{NI}$ = 0.002185), 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 such that $\mu_0 = \mu_R - M_{NI}$ = 0.0022425).

Following accepted procedure, the analysis will be a non-inferiority test using the *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 ....**Power**
Higher Means Are..**Better (H1: μ > μ0)**
Population Size..**Infinite**
Alpha..**0.025**
N (Sample Size)...**20 40 60 80 100 150 200 300**
μ0 (Non-Inferiority Mean)................................**21.85 22.425**
μ1 (Actual Mean) **23**
σ (Standard Deviation)....................................**3**

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: μ ≤ μ0   vs.   H1: μ > μ0

Power	Sample Size N	Mean Non-Inferiority μ0	Mean Actual μ1	Standard Deviation σ	Alpha
0.40298	20	21.9	23	3	0.025
0.67884	40	21.9	23	3	0.025
0.84359	60	21.9	23	3	0.025
0.92904	80	21.9	23	3	0.025
0.96949	100	21.9	23	3	0.025
0.99688	150	21.9	23	3	0.025
0.99973	200	21.9	23	3	0.025
1.00000	300	21.9	23	3	0.025
0.13506	20	22.4	23	3	0.025
0.22730	40	22.4	23	3	0.025
0.31728	60	22.4	23	3	0.025
0.40298	80	22.4	23	3	0.025
0.48273	100	22.4	23	3	0.025
0.65079	150	22.4	23	3	0.025
0.77356	200	22.4	23	3	0.025
0.91305	300	22.4	23	3	0.025

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.
μ0   The non-inferiority mean. Since higher means are better, μ0 = μR - NIM, where μR is the baseline, standard, or
        reference mean and NIM is the margin of non-inferiority.
μ1   The actual value of the population mean at which power and sample size are calculated.
σ   The standard deviation of the response (or standard deviation of differences for paired data). It measures the
        variability in the population.
Alpha   The probability of rejecting a true null hypothesis.

Summary Statements

A single-group design will be used to test whether the mean is non-inferior, with a non-inferiority margin subtracted from a reference mean totaling 21.9 (H0: μ ≤ 21.9 versus H1: μ > 21.9). The comparison will be made using a one-sided, one-sample Z-test, with a Type I error rate (α) of 0.025. The (known) standard deviation is assumed to be 3. To detect a mean of 23 with a sample size of 20, the power is 0.40298.

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 Lokhnygina, Y. (2018) pages 32-33.)

D   The expected number of dropouts. D = N' - N.
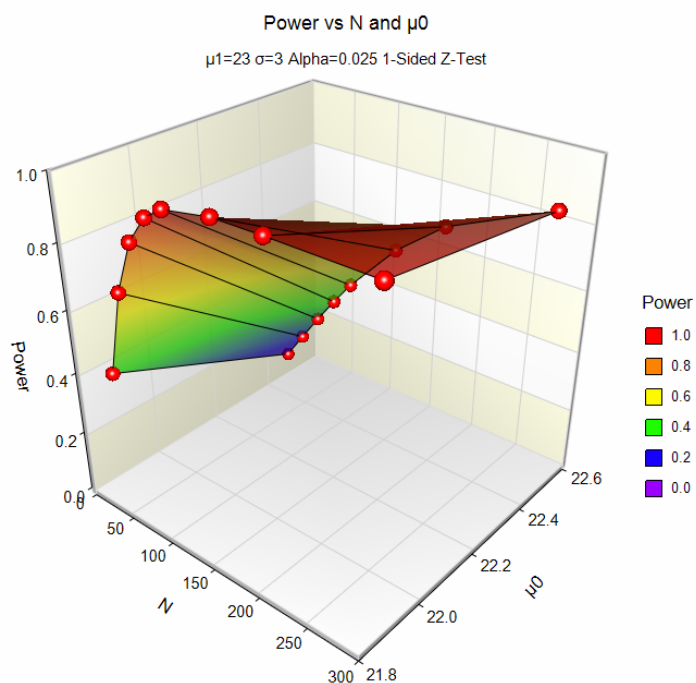
Dropout Summary Statements

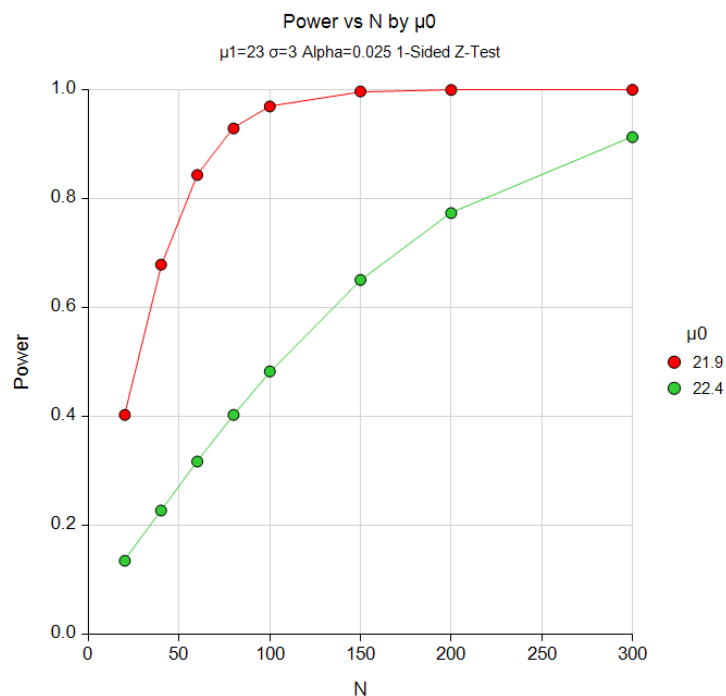Anticipating a 20% dropout rate, 25 subjects should be enrolled to obtain a final sample size of 20 subjects.

References

Chow, S.C., Shao, J., Wang, H., and Lokhnygina, 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.

Plots

Power vs N by μ0
μ1=23 σ=3 Alpha=0.025 1-Sided Z-Test

Power vs N and μ0
μ1=23 σ=3 Alpha=0.025 1-Sided Z-Test

The above report shows that for μ_0 = 21.85 (NIM = 1.15), the sample size necessary to obtain 90% power is just under 80. However, if μ_0 = 22.425 (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 ....**Sample Size**
Higher Means Are...**Better (H1: μ > μ0)**
Population Size...**Infinite**
Power..**0.90**
Alpha...**0.025**
μ0 (Non-Inferiority Mean)..............................**21.85 22.425**
μ1 (Actual Mean) **23**
σ (Standard Deviation)..................................**3**

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: μ ≤ μ0   vs.   H1: μ > μ0

| | | Mean | | | |
| | Sample Size N | Non-Inferiority μ0 | Actual μ1 | Standard Deviation σ | Alpha |
Power					
0.90195	72	21.9	23	3	0.025
0.90097	287	22.4	23	3	0.025

This report shows the exact sample size requirement for each value of μ_0.

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 non-inferiority z-test where μ_R = 1.5 and M_{NI} = 0.5. Their example obtains a sample size of 7 when μ_1 = 2, μ_0 = 1, σ = 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: μ > μ0)
Population Size..............................................Infinite
Power..........................................................0.80
Alpha..........................................................0.05
μ0 (Non-Inferiority Mean)..............................1
μ1 (Actual Mean) .........................................2
σ (Standard Deviation)..................................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: μ ≤ μ0   vs.   H1: μ > μ0
_____

| | | Mean | | | |
| | Sample Size | Non-Inferiority | Actual | Standard Deviation | |
Power	N	μ0	μ1	σ	Alpha
0.84156	7	1	2	1	0.05
_____

PASS also obtains a sample size of 7.