Chapter 535

Non-Inferiority Tests for the Ratio of Two Means in a Higher-Order Cross-Over Design (Log-Normal Data)

Introduction

This procedure calculates power and sample size for non-inferiority tests which use the ratio of the two means of a higher-order cross-over design. Measurements are made on individuals that have been randomly assigned to one of several treatment sequences. Only a brief introduction to the subject will be given here. For a comprehensive discussion on the subject, refer to Chen et al. (1997) and Chow et al. (2003).

Cross-Over Designs

Senn (2002) defines a *cross-over* design as one in which each subject receives all treatments at least once and the object is to study differences among the treatments. The name *cross-over* comes from the most common case in which there are only two treatments. In this case, each subject *crosses over* from one treatment to another. It is assumed that there is a *washout* period between treatments during which the response returns to its baseline value. If this does not occur, there is said to be a *carryover* effect.

A 2x2 cross-over design refers to two treatments (periods) and two *sequences* (treatment orderings). One sequence of treatments is treatment A followed by treatment B. The other sequence is B and then A. The design includes a washout period between responses to make certain that the effects of the first drug do no carryover to the second. Thus, the groups of subjects in this design are defined by the sequence in which the two treatments are administered, not by the treatments they receive.

Higher-Order Cross-Over Designs

Chen et al. (1997) present the results for four cross-over designs that are more complicated than the 2x2 design. Assume that the two treatments are labeled A and B. The available designs are defined by the order and number of times the two treatments are administered.

Balaam's Design

Balaam's design has four sequences with two treatments each. It is popular because it allows the intrasubject variabilities to be estimated. The design is

Sequence	Period 1	Period 2
1	A	A
2	B	B
3	A	B
4	B	A

Two-Sequence Dual Design

This design has two sequences with three periods each. It is popular because it allows the intrasubject variabilities to be estimated. The design is

Sequence	Period 1	Period 2	Period 3
1	A	B	B
2	B	A	A

Four-Period Design with Two Sequences

This design has two sequences of four periods each. The design is

Sequence	Period 1	Period 2	Period 3	Period 4
1	A	B	B	A
2	B	A	A	B

Four-Period Design with Four Sequences

This design has four sequences of four periods each. The design is

Sequence	Period 1	Period 2	Period 3	Period 4
1	A	A	B	B
2	B	B	A	A
1	A	B	B	A
2	B	A	A	B

Advantages of Cross-Over Designs

A comparison of treatments on the same subject is expected to be more precise. The increased precision often translates into a smaller sample size. Also, patient enrollment may be easier to obtain because each patient will receive both treatments.

Disadvantages of Cross-Over Designs

The statistical analysis of a cross-over experiment is more complex than a parallel-group experiment and requires additional assumptions. In a cross-over experiment, it may be difficult to separate the treatment effect from the time effect and the carry-over effect of the previous treatment.

These cross-over designs cannot be used when the treatment (or the measurement of the response) alters the subject permanently. Hence, it cannot be used to compare treatments that are intended to provide a cure.

Because subjects must be measured at least twice, it may be more difficult to keep patients enrolled in the study. This is particularly true when the measurement process is painful, uncomfortable, embarrassing, or time consuming.

The Statistical Hypotheses

Both non-inferiority and non-zero null tests are examples of directional (one-sided) tests. Remember that in the usual t-test setting, the null (H0) and alternative (H1) hypotheses for one-sided tests are defined as

$$\mathrm{H}_0{:}\phi \le A \ \text{ versus } \ \mathrm{H}_1{:}\phi > A$$

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

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

$$\mathrm{H}_0{:}\phi \ge A \ \text{ versus } \ \mathrm{H}_1{:}\phi < A$$

Non-inferiority and *non-zero null* 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	*Treatment mean*. This is the treatment mean.
μ_R	Not used	*Reference mean*. This is the mean of a reference population.
M_{NI}	NIM	*Margin of non-inferiority*. This is a tolerance value that defines the maximum amount that is not of practical importance. This is the largest change in the mean ratio from the baseline value (usually one) that is still considered to be trivial.
ϕ	R1	*True ratio*. This is the value of $\phi = \mu_T / \mu_R$ at which the power is calculated.

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

The null hypothesis of inferiority is

$$\mathrm{H}_0{:}\phi \le \phi_L \quad where\ \phi_L < 1.$$

and the alternative hypothesis of non-inferiority is

$$\mathrm{H}_1{:}\phi > \phi_L$$

Log-Transformation

In many cases, hypotheses stated in terms of ratios are more convenient than hypotheses stated in terms of differences. This is because ratios can be interpreted as scale-less percentages, but differences must be interpreted as actual amounts in their original scale. Hence, it has become a common practice to take the following steps in hypothesis testing.

1. State the statistical hypotheses in terms of ratios.
2. Transform these into hypotheses about differences by taking logarithms.
3. Analyze the logged data—that is, do the analysis in terms of the difference.
4. Draw the conclusion in terms of the ratio.

The details of step 2 for the alternative hypothesis are as follows.

$$\phi_L \le \phi$$
$$\Rightarrow \phi_L \le \left\{\frac{\mu_T}{\mu_R}\right\}$$
$$\Rightarrow \ln(\phi_L) \le \{\ln(\mu_T) - \ln(\mu_R)\}$$

Thus, a hypothesis about the ratio of the means on the original scale can be translated into a hypothesis about the difference of two means on the logged scale.

Coefficient of Variation

The coefficient of variation (COV) is the ratio of the standard deviation to the mean. This parameter is used to represent the variation in the data because of a unique relationship that it has in the case of log-normal data.

Suppose the variable *X* is the logarithm of the original variable *Y*. That is, *X* = ln(*Y*) and *Y* = exp(*X*). Label the mean and variance of *X* as μ_X and σ_X^2, respectively. Similarly, label the mean and variance of *Y* as μ_Y and σ_Y^2, respectively. If *X* is normally distributed, then *Y* is log-normally distributed. Julious (2004) presents the following well-known relationships between these two variables

$$\mu_Y = e^{\mu_X + \frac{\sigma_X^2}{2}}$$
$$\sigma_Y^2 = \mu_Y^2\left(e^{\sigma_X^2} - 1\right)$$

From this relationship, the coefficient of variation of *Y* can be found to be

$$COV_Y = \frac{\sqrt{\mu_Y^2\left(e^{\sigma_X^2} - 1\right)}}{\mu_Y}$$
$$= \sqrt{e^{\sigma_X^2} - 1}$$
$$= \sqrt{e^{\sigma_W^2} - 1}$$

where σ_W^2 is the within mean square error from the analysis of variance of the logged data. Solving this relationship for σ_X^2, the standard deviation of *X* can be stated in terms of the coefficient of variation of *Y* as

$$\sigma_X = \sqrt{\ln(COV_Y^2 + 1)}$$

Similarly, the mean of *X* is

$$\mu_X = \ln\left(\frac{\mu_Y}{\sqrt{COV_Y^2 + 1}}\right)$$

Thus, the hypotheses can be stated in the original (Y) scale and then power can be analyzed in the transformed (X) scale.

Non-Inferiority Tests

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

Case 1: High Values Good, Non-Inferiority Test

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 null and alternative hypotheses are

$$\mathrm{H}_0\text{:}\frac{\mu_T}{\mu_R} \le (1-\varepsilon) \quad \text{versus} \quad \mathrm{H}_1\text{:}\frac{\mu_T}{\mu_R} > (1-\varepsilon)$$

$$\mathrm{H}_0\text{:}\ln(\mu_T) - \ln(\mu_R) \le \ln(1-\varepsilon) \quad \text{versus} \quad \mathrm{H}_1\text{:}\ln(\mu_T) - \ln(\mu_R) > \ln(1-\varepsilon)$$

Case 2: High Values Bad, Non-Inferiority Test

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 null and alternative hypotheses are

$$\mathrm{H}_0\text{:}\frac{\mu_T}{\mu_R} \ge (1+\varepsilon) \quad \text{versus} \quad \mathrm{H}_1\text{:}\frac{\mu_T}{\mu_R} < (1+\varepsilon)$$

$$\mathrm{H}_0\text{:}\ln(\mu_T) - \ln(\mu_R) \ge \ln(1+\varepsilon) \quad \text{versus} \quad \mathrm{H}_1\text{:}\ln(\mu_T) - \ln(\mu_R) < \ln(1+\varepsilon)$$

Test Statistics

The analysis for assessing non-inferiority using higher-order cross-over designs is discussed in detail in Chapter 9 of Chow and Liu (2000). Unfortunately, their presentation is too lengthy to give here. Their method involves the computation of an analysis of variance to estimate the error variance. It also describes the construction of confidence limits for appropriate contrasts. One-sided confidence limits can be used for non-inferiority tests. Details of this approach are given in Chapter 3 of Chow et al. (2003). We refer you to these books for details.

Power Calculation

The power of the non-inferiority and superiority tests for the case in which higher values are better is given by

$$Power = T_V\left(\left(\frac{\ln(1-\varepsilon)}{\sigma_W\sqrt{b/n}}\right) - t_{V,1-\alpha}\right)$$

where T represents the cumulative t distribution, V and b depend on the design, n is the average number of subjects per sequence, and

$$\sigma_W = \sqrt{\ln\left(COV_Y^2 + 1\right)}$$

The power of the non-inferiority and superiority tests for the case in which higher values are worse is given by

$$Power = 1 - T_V\left(t_{V,1-\alpha} - \left(\frac{-\ln(1+\varepsilon)}{\sigma_W\sqrt{b/n}}\right)\right)$$

The constants V and b depend on the design as follows:

Design Type	Parameters (V,b)
Balaam's Design	V = 4n – 3, b = 2.
Two-Sequence Dual Design	V = 4n – 4, b = 3/4.
Four-Period Design with Two Sequences	V = 6n – 5, b = 11/20.
Four-Period Design with Four Sequences	V = 12n – 5, b = 1/4.

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 means, sample sizes, alpha, and power.

Solve For

Solve For

This option specifies the parameter to be solved for from the other parameters. Under most situations, you will select either *Power* for a power analysis or *Sample Size* for sample size determination.

Select *Sample Size* when you want to calculate the sample size needed to achieve a given power and alpha level. Note that there are two choices for finding *Sample Size*. Select *Sample Size (Equal Per Sequence)* when you want the design to have an equal number of subjects per sequence. Select *Sample Size (Exact)* when you want to find the exact sample size even though the number of subjects cannot be dividing equally among the sequences.

Select *Power* when you want to calculate the power of an experiment.

Test

Design Type

Specify which cross-over design that you want to analyze. These designs allow you to compare two means: a treatment (A) and a reference (B). The designs assume that each subject is measured two or more times (or periods) with ample time in between to wash-out the effect of the treatment.

A sequence is an ordering of the how A and B are applied. For example, 'ABA' means that the subjects are measured three times: first treatment A, then B, then A again. Other possible sequences are BAB, AAB, and BBA. Each design includes several sequences.

The design tries to balance the experiment so that each treatment occurs an equal number of times and after each other treatment an equal number of times.

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 R <= 1-NIM and the alternative hypothesis is R > 1-NIM. If Higher Means Are Worse the null hypothesis is R >= 1+NIM and the alternative hypothesis is R < 1+NIM.

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 (Total Sample Size)

This option specifies one or more values of the sample size, the number of individuals in the study (total subjects in all sequences). These values must be integers greater than one.

Effect Size – Ratios

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 one for which the mean ratio (Treatment Mean / Reference Mean) still indicates non-inferiority of the treatment mean. E.g., a value of 0.2 here specifies that mean ratios greater than 0.8 indicate non-inferiority of the treatment mean.

When higher means are worse, this value is the distance above one for which the mean ratio (Treatment Mean / Reference Mean) still indicates non-inferiority of the treatment mean. E.g., a value of 0.2 here specifies that mean ratios less than 1.2 indicate non-inferiority of the treatment mean.

Recommended values: 0.20 is a common value for the parameter.

R1 (Mean Ratio)

This is the value of the ratio of the two means (Treatment Mean / Reference Mean) at which the power is to be calculated.

Often, the ratio will be set to one. However, some authors recommend using a ratio slightly different than one, such as 0.95 (when higher means are "better") or 1.05 (when higher means are "worse"), since this will require a larger sample size.

Effect Size – Coefficient of Variation

COV (Coefficient of Variation of Y)

A response variable Y has mean μ_Y and standard deviation σ_Y. The coefficient of variation of Y, σ_Y/μ_Y, is used to specify the variation in Y.

Suppose that Y is skewed, so a logarithmic transformation will be applied. That is, the variable X = Ln(Y) will be analyzed using an appropriate ANOVA procedure. The analysis assumes that X is normally distributed with mean μ_X and standard deviation σ_X. So Y is lognormal and X is normal. Power and sample size calculations are made on X. However, interpretation may be easier in the Y scale.

If Y is truly lognormal, there is a direct relationship between μ_Y and σ_Y and μ_X and σ_X. Using this relationship, it can be shown that

$$COV_Y = \sqrt{Exp(\sigma_X^2) - 1}$$

Solving for σ_X, we obtain

$$\sigma_X = \sqrt{\ln(COV_Y^2 + 1)}$$

There are two ways to obtain a value for COV_Y.

1. Calculate COV_Y directly from the mean and SD of a set of Y values obtained from a previous (or pilot) study.

2. Obtain σw_X as the square root of the within mean square error in an ANOVA table of X. Then use the fact that $\sigma_X = \sigma w_X$.

We have found that method 1 **does not** provide a reliable estimate of σ_X, **so we recommend method 2**.

Example 1 – Finding Power

A company has developed a generic drug for treating rheumatism and wants to show that it is not inferior to standard drug. Balaam's cross-over design will be used.

Researchers have decided to set the margin of equivalence at 0.20. Past experience leads the researchers to set the COV to 0.40. The significance level is 0.05. The power will be computed assuming that the true ratio is one. Sample sizes between 50 and 550 will be included in the analysis. Note that several of these sample size values are not divisible by 4. This is note a problem here because are main goal is to get an overview of power versus sample size. When searching for the sample size, we can request that only designs divisible by 4 be considered.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the procedure window. 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**
Design Type	**2x4 (Balaam: AA|BB|AB|BA)**
Higher Means Are	**Better**
Alpha	**0.05**
N (Total Sample Size)	**50 to 550 by 100**
NIM (Non-Inferiority Margin)	**0.20**
R1 (Mean Ratio)	**1.0**
COV (Coefficient of Variation)	**0.40**

Annotated Output

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

Numeric Results

Numeric Results for a Non-Inferiority Test of the Mean Ratio in a Two-Period, Four-Sequence (Balaam) Design
Treatment Sequences: AA | BB | AB | BA
Higher Means are Better
Hypotheses: H0: μT/μR ≤ 1 - NIM; H1: μT/μR > 1 - NIM

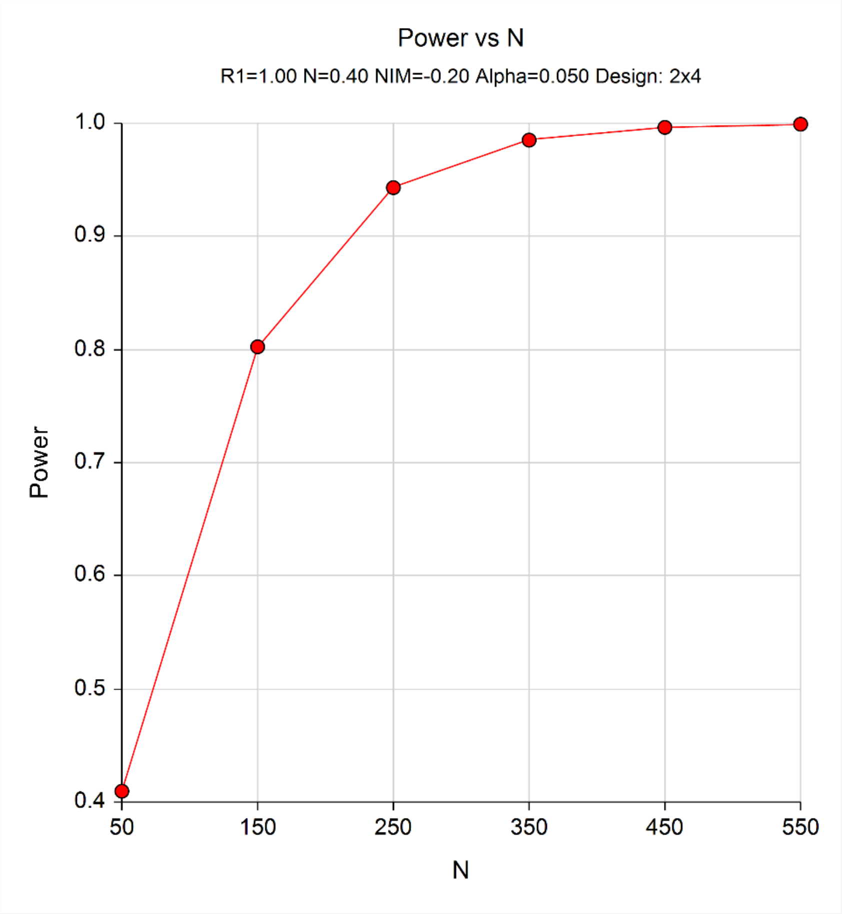
Power	Total Sample Size N	Non-Inferiority Margin NIM	Mean Ratio R1 (μT / μR)	Within Standard Deviation σw	Alpha
0.4096	50	0.200	1.000	0.4000	0.050
0.8024	150	0.200	1.000	0.4000	0.050
0.9431	250	0.200	1.000	0.4000	0.050
0.9851	350	0.200	1.000	0.4000	0.050
0.9964	450	0.200	1.000	0.4000	0.050
0.9992	550	0.200	1.000	0.4000	0.050

Report Definitions
Power is the probability of rejecting H0 (concluding non-inferiority) when H0 is false.
N is the total number of subjects. They are divided evenly among all sequences.
μT is the treatment mean. It is usually associated with the letter 'A' in the design.
μR is the reference mean. It is usually associated with the letter 'B' in the design.
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 one that is still considered non-inferior.
R1 is the ratio of the means at which the power is computed.
COV is the coefficient of variation on the original scale.
Alpha is the probability of falsely rejecting H0 (falsely concluding non-inferiority).

Summary Statements
In a non-inferiority test on data for which higher values are better, drawn from Two-Period, Four-Sequence (Balaam) cross-over design, a total sample size of 50 achieves 41% power at a 5% significance level when the true ratio of the means is 1.000, the coefficient of variation is 0.4000, and the non-inferiority margin is 0.200.

This report shows the power for the indicated scenarios.

Plots Section



This plot shows the power versus the sample size.

Example 2 – Finding Sample Size

Continuing with Example 1, the researchers want to find the exact sample size needed to achieve both 80% and 90% power.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the procedure window. 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 (Exact)**
Design Type	**2x4 (Balaam: AA|BB|AB|BA)**
Higher Means Are	**Better**
Power	**0.8 0.9**
Alpha	**0.05**
NIM (Non-Inferiority Margin)	**0.20**
R1 (Mean Ratio)	**1.0**
COV (Coefficient of Variation)	**0.40**

Output

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

Numeric Results for a Non-Inferiority Test of the Mean Ratio in a Two-Period, Four-Sequence (Balaam) Design
Treatment Sequences: AA | BB | AB | BA
Higher Means are Better
Hypotheses: H0: μT/μR ≤ 1 - NIM; H1: μT/μR > 1 - NIM

Power	Total Sample Size N	Non-Inferiority Margin NIM	Mean Ratio R1 (μT / μR)	Within Standard Deviation σw	Alpha
0.8001	149	0.200	1.000	0.4000	0.050
0.9002	206	0.200	1.000	0.4000	0.050

When the non-inferiority margin is set to 0.20, we note that 206 subjects are needed to achieve 90% power and 149 subjects are needed to achieve at least 80% power.

Example 3 – Validation

We could not find a validation example for this procedure in the statistical literature, so we will have to generate a validated example from within **PASS**. To do this, we use the Higher-Order, Cross-Over Equivalence using Ratios procedure which was validated. By setting the upper equivalence limit to a large value (we used 11), we obtain results for a non-inferiority test that can be used to validate this procedure.

In the other procedure, suppose the coefficient of variation is 0.40, the equivalence limits are 0.80 and 11.0, the true ratio of the means is 1, the power is 90%, and the significance level is 0.05. These settings are stored as Example 4 in that procedure. PASS calculates a sample size of 208.

We will now setup this example in PASS. The only difference is that now we set E to 0.2 instead of RL to 0.8.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the procedure window. 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 (Equal Per Sequence)**
Design Type	**2x4 (Balaam: AA|BB|AB|BA)**
Higher Means Are	**Better**
Power	**0.90**
Alpha	**0.05**
NIM (Non-Inferiority Margin)	**0.20**
R1 (Mean Ratio)	**1.0**
COV (Coefficient of Variation)	**0.40**

Output

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

Numeric Results for a Non-Inferiority Test of the Mean Ratio in a Two-Period, Four-Sequence (Balaam) Design
Treatment Sequences: AA | BB | AB | BA
Higher Means are Better
Hypotheses: H0: μT/μR ≤ 1 - NIM; H1: μT/μR > 1 - NIM

Power	Total Sample Size N	Non-Inferiority Margin NIM	Mean Ratio R1 (μT / μR)	Within Standard Deviation σw	Alpha
0.9027	208	0.200	1.000	0.4000	0.050

PASS has also obtained the sample size of 208.