Chapter 533

Superiority by a Margin 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-zero null 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

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

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*.

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

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_S	SM	*Margin of superiority.* This is a tolerance value that defines the magnitude of difference that is required for practical importance. This may be thought of as the smallest difference from the reference that is considered to be different.
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 non-superiority is

$$H_0: \phi \leq \phi_L, \text{ where } \phi_L > 1.$$

and the alternative hypothesis of superiority is

$$H_1: \phi > \phi_L$$

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. Also, taking the logarithm of the data can make skewed data closer to normal. 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:

$$\phi_L \leq \phi$$

$$\Rightarrow \phi_L \leq \left\{ \frac{\mu_T}{\mu_R} \right\}$$

$$\Rightarrow \ln(\phi_L) \leq \{\ln(\mu_T) - \ln(\mu_R)\}$$

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

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

$$= \sqrt{e^{\sigma_X^2} - 1}$$

$$= \sqrt{e^{\sigma_W^2} - 1}$$

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.

Superiority Tests

A *superiority by a margin test* tests that the treatment mean is better than the reference mean by more than a small superiority margin. The actual direction of the hypothesis depends on the response variable being studied.

Case 1: High Values Good, Superiority Test

In this case, higher values are better. The hypotheses are arranged so that rejecting the null hypothesis implies that the treatment mean is greater than the reference mean by at least the margin of superiority. The null and alternative hypotheses are

$$H_0\colon\frac{\mu_T}{\mu_R} \le (1 + \varepsilon) \quad \text{versus} \quad H_1\colon\frac{\mu_T}{\mu_R} > (1 + \varepsilon)$$

$$H_0\colon\ln(\mu_T) - \ln(\mu_R) \le \ln(1 + \varepsilon) \quad \text{versus} \quad H_1\colon\ln(\mu_T) - \ln(\mu_R) > \ln(1 + \varepsilon)$$

Case 2: High Values Bad, Superiority Test

In this case, lower values are better. The hypotheses are arranged so that rejecting the null hypothesis implies that the treatment mean is less than the reference mean by at least the margin of superiority. The null and alternative hypotheses are

$$H_0: \frac{\mu_T}{\mu_R} \geq (1 - \varepsilon) \qquad \text{versus} \qquad H_1: \frac{\mu_T}{\mu_R} < (1 - \varepsilon)$$

$$H_0: \ln(\mu_T) - \ln(\mu_R) \geq \ln(1 - \varepsilon) \quad \text{versus} \quad H_1: \ln(\mu_T) - \ln(\mu_R) < \ln(1 - \varepsilon)$$

Test Statistics

The analysis for assessing non-inferiority and non-zero null hypotheses 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

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

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	$V = 12n - 5$, $b = 1/4$.

Example 1 – Finding Power

A company has developed a generic drug for treating rheumatism and wants to show that it is superior to a standard drug by a small amount. Balaam's cross-over design will be used.

Researchers have decided to set the margin of superiority 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 1.40. 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 not a problem here because our 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

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**
Design Type..**2x4 (Balaam: AA|BB|AB|BA)**
Higher Means Are ....**Better**
Alpha..**0.05**
N (Total Sample Size)...................................**50 to 550 by 100**
SM (Superiority Margin) **0.2**
R1 (Mean Ratio) **1.4**
COV (Coefficient of Variation).......................**0.4**

Output

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

Numeric Reports

Numeric Results

―――

Solve For: Power
Design Type: Two-Period, Four-Sequence (Balaam)
Treatment Sequences: AA | BB | AB | BA
Higher Means Are: Better
Hypotheses: H0: μT/μR ≤ 1 + SM   vs.   H1: μT/μR > 1 + SM

―――

Power	Total Sample Size N	Superiority Margin SM	Mean Ratio (μT/μR) R1	Coefficient of Variation COV	Alpha
0.25067	50	0.2	1.4	0.4	0.05
0.53076	150	0.2	1.4	0.4	0.05
0.72071	250	0.2	1.4	0.4	0.05
0.84035	350	0.2	1.4	0.4	0.05
0.91158	450	0.2	1.4	0.4	0.05
0.95226	550	0.2	1.4	0.4	0.05

―――

Power   The probability of rejecting a false null hypothesis when the alternative hypothesis is true.
N   The total number of subjects. They are divided evenly among all sequences.
μT   The treatment mean. It is usually associated with the letter "A" in the design.
μR   The reference mean. It is usually associated with the letter "B" in the design.
SM   The magnitude of the margin of superiority. Since higher means are better, this value is positive and is the distance
         above one that is required to be considered superior.
R1   The ratio of the means at which the power is computed.
COV   The coefficient of variation on the original scale.
Alpha   The probability of falsely rejecting H0 (falsely concluding superiority).

Summary Statements

―――

A two-period, four-sequence (Balaam) cross-over (AA | BB | AB | BA) design (where higher means are considered to be better) will be used to test whether the treatment mean (μT) is superior to the reference mean (μR) by a margin, by testing whether the ratio of means (μT / μR) is greater than one plus the superiority margin 0.2 (H0: μT / μR ≤ 1 + 0.2 versus H1: μT / μR > 1 + 0.2). The comparison will be made using a one-sided t-test using a log-transformation, with a Type I error rate (α) of 0.05. The coefficient of variation on the original scale is assumed to be 0.4. To detect a ratio of means (μT / μR) of 1.4, with a total sample size of 50 (allocated equally to the 4 sequences), the power is 0.25067.

―――

Dropout-Inflated Sample Size

Dropout Rate	Sample Size N	Dropout-Inflated Enrollment Sample Size N'	Expected Number of Dropouts D
20%	50	63	13
20%	150	188	38
20%	250	313	63
20%	350	438	88
20%	450	563	113
20%	550	688	138

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, 63 subjects should be enrolled to obtain a final sample size of 50 subjects.

References

Chow, S.C. and Liu, J.P. 1999. Design and Analysis of Bioavailability and Bioequivalence Studies. Marcel Dekker. New York

Chow, S.C., Shao, J., and Wang, H. 2003. Sample Size Calculations in Clinical Research. Marcel Dekker. New York.

Chen, K.W., Chow, S.C., and Li, G. 1997. 'A Note on Sample Size Determination for Bioequivalence Studies with Higher-Order Crossover Designs.' Journal of Pharmacokinetics and Biopharmaceutics, Volume 25, No. 6, pages 753-765.

This report shows the power for the indicated scenarios.

Plots Section

Plots
_____
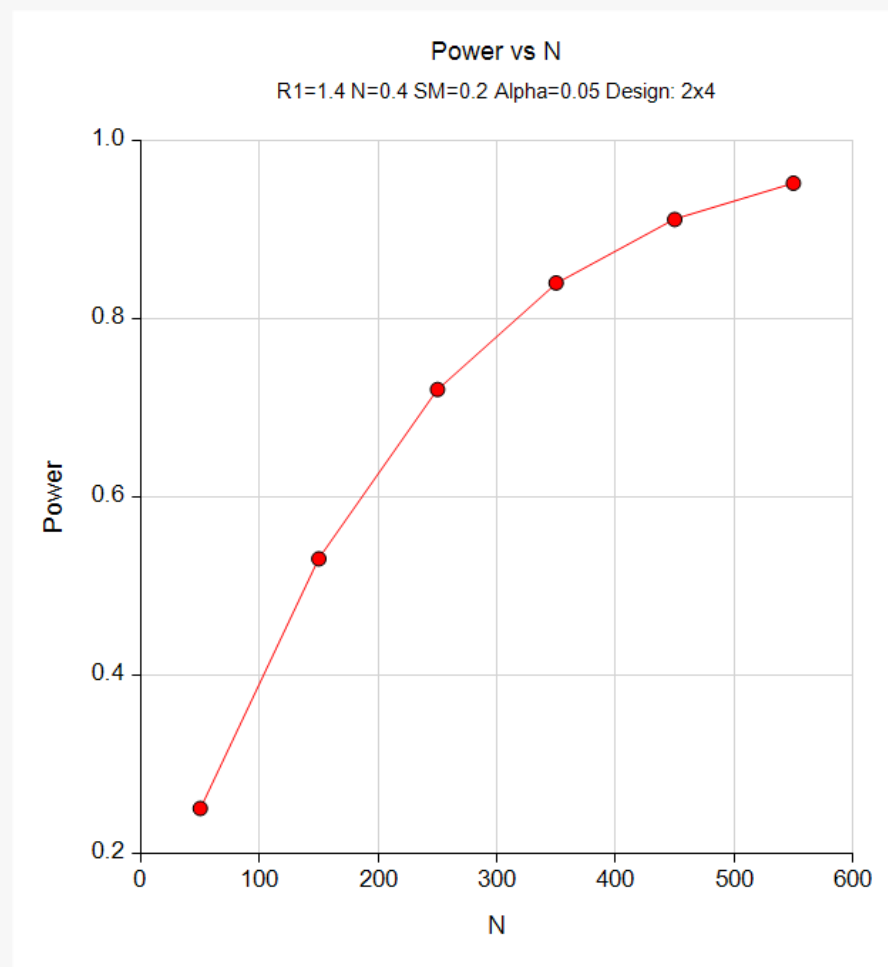
This plot shows the power versus the sample size.

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

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 (Exact)**
Design Type...**2x4 (Balaam: AA|BB|AB|BA)**
Higher Means Are ....**Better**
Power..**0.8 0.9**
Alpha...**0.05**
SM (Superiority Margin) **0.2**
R1 (Mean Ratio) **1.4**
COV (Coefficient of Variation)........................**0.4**

Output

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

Numeric Results

Solve For: Sample Size (Exact)
Design Type: Two-Period, Four-Sequence (Balaam)
Treatment Sequences: AA | BB | AB | BA
Higher Means Are: Better
Hypotheses: H0: μT/μR ≤ 1 + SM   vs.   H1: μT/μR > 1 + SM

Power	Total Sample Size N	Superiority Margin SM	Mean Ratio (μT/μR) R1	Coefficient of Variation COV	Alpha
0.80061	311	0.2	1.4	0.4	0.05
0.90027	430	0.2	1.4	0.4	0.05

When the equivalence margin is set to 0.20, we note that 430 subjects are needed to achieve 90% power and 311 subjects are needed to achieve at least 80% power.

Example 3 – Validation

This procedure uses the same mechanics as the *Non-Inferiority Tests for the Ratio of Two Means in a Higher-Order Cross-Over Design (Log-Normal Data)* procedure. We refer the user to Example 3 of Chapter 535 for the validation.