Chapter 337

Multi-Arm Superiority by a Margin Tests for the Ratio of Treatment and Control Means (Log-Normal Data)

Introduction

This module computes power and sample size for multiple superiority by a margin tests of treatment means versus a control mean when the data are assumed to follow the log-normal distribution and the statistical hypotheses are expressed in terms of mean ratios.

The details of this t-test are given in Julious (2004). The multiple comparison aspect of this procedure is based on the results in Machin, Campbell, Tan, and Tan (2018).

In this parallel-group design, there are *k* treatment groups and one control group. A mean is measured in each group. A total of *k* hypothesis tests are anticipated each comparing a treatment group with the common control group using a t-test based on the difference of two means of the log-transformed data.

The Bonferroni adjustment of the type I error rate may be optionally made because several comparisons are being tested using the same data. Making a multiplicity adjustment is usually recommended, but not always. In fact, Saville (1990) advocates not applying it and Machin, Campbell, Tan, and Tan (2018) include omitting it as a possibility.

Background

Whether you want to test several doses of a single treatment or several types of treatments, good research practice requires that each treatment be compared with a control. A popular three-arm design consists of three groups: control, treatment A, and treatment B. Two tests are run: treatment A versus control and treatment B versus the same control. This design avoids having to obtain a second control group for treatment B. Besides the obvious efficiency in subjects, it may be easier to recruit subjects if their chances of receiving the new treatment are better than 50-50.

Technical Details

Suppose you want to compare *k* treatment groups with means μ_i and sample sizes N_i and one control group with mean μ_C and sample size N_C. The total sample size is $N = N_1 + N_2 + \cdots + N_k + N_C$.

Superiority by a Margin Tests

A *superiority by a margin test* tests that the treatment mean is better than the control mean by more than the superiority margin. The actual direction of the hypothesis depends on the response variable being studied. Define $R = \mu_i/\mu_C$.

Case 1: High Values Good

In this case, higher response values are better. The hypotheses are arranged so that rejecting the null hypothesis implies that the treatment mean is greater than the control mean by at least a small amount. This results in a superiority boundary called R_U . The null and alternative hypotheses with are

$$H_0: R \leq R_U \quad \text{vs.} \quad H_1: R > R_U$$

where $R_U > 1$.

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 treatment mean is less than the control mean by at least a small amount. This results in a superiority boundary called R_L. The null and alternative hypotheses with are

$$H_0: R \geq R_L \quad \text{vs.} \quad H_1: R < R_L$$

where $R_L > 1$.

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 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 the ratio of the means.
2. Transform this into hypotheses about a difference 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 two-sided null hypothesis are as follows:

$$H_0: R = R_0 \quad \Rightarrow \quad H_0: \frac{\mu_i}{\mu_C} = R_0 \quad \Rightarrow \quad H_0: \ln(\mu_i) - \ln(\mu_C) = \ln(R_0)$$

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 can be 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

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

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

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 the power can be analyzed in the transformed (X) scale. For parallel-group designs, $\sigma_X^2 = \sigma_d^2$, the average variance used in the t-test of the logged data.

Power Calculation

As is shown above, the hypotheses can be stated in the original (Y) scale using ratios or the logged (X) scale using differences. In either case, the power and sample size calculations are made using the formulas for testing the difference in two means. These formulas are presented in another chapter and are not duplicated here.

Multiplicity Adjustment

Because *k* t-tests between treatment groups and the control group are run when analyzing the results of this study, many statisticians recommend that the Bonferroni adjustment be applied. This adjustment is easy to apply: the value of alpha that is used in the test is found by dividing the original alpha by the number of tests. For example, if the original alpha is set at 0.05 and the number of treatment (not including the control) groups is five, the individual tests will be conduction using an alpha of 0.01.

The main criticism of this procedure is that if there are many tests, the value of alpha becomes very small. To mitigate against this complaint, some statisticians recommend separating the treatment groups into those that are of primary interest and those that are of secondary interest. The Bonferroni adjustment is made by the using the number of primary treatments rather than the total number of treatments.

There are some who advocate ignoring the adjustment entirely in the case of randomized clinical trials. See for example Saville (1990) and the discussion in chapter 14 of Machin, Campbell, Tan, and Tan (2018).

Size of the Control Group

Because the control group is used over and over, some advocate increasing the number of subjects in this group. The standard adjustment is to include $\sqrt{k}$ subjects in the control group for each subject in one of the treatment groups. See Machin, Campbell, Tan, and Tan (2018, pages 231-232). Note that usually, the treatment groups all have the same size.

Example 1 – Finding the Sample Size

A parallel-group clinical trial is being designed to compare three treatment therapies against the standard therapy. Higher values of the response are desirable. Suppose the standard therapy has a mean response of 10.0 with a standard deviation of 2.5. The investigators would like a sample size large enough to find statistical significance at the 0.025 level (since this a one-sided test) if the actual mean responses of the three treatments are 13.2, 13.4, and 13.6, the power of each test is 0.80, and the superiority limit is 1.25. They want to consider a range of standard deviations from 2.0 to 3.0. The data are log-normally distributed.

Following standard procedure, the control group multiplier will be set to $\sqrt{k} = \sqrt{3} = 1.732$ since the control group is used for three comparisons in this design.

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	**Sample Size**
Higher Means Are	**Better (H1: R > RU)**
Power of Each Test	**0.8**
Overall Alpha	**0.025**
Bonferroni Adjustment	**Standard Bonferroni**
Group Allocation	**Enter Group Allocation Pattern, solve for group sample sizes**
RU (Upper Superiority Limit)	**1.25**
Control Mean	**10**
Control Sample Size Allocation	**1.732**
Set A Number of Groups	**1**
Set A Mean	**13.2**
Set A Sample Size Allocation	**1**
Set B Number of Groups	**1**
Set B Mean	**13.4**
Set B Sample Size Allocation	**1**
Set C Number of Groups	**1**
Set C Mean	**13.6**
Set C Sample Size Allocation	**1**
Set D Number of Groups	**0**
More	**Unchecked**
σ (Standard Deviation)	**2 2.5 3**

Output

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

Numeric Reports

Numeric Results

Solve For:	Sample Size
Group Allocation:	Enter Group Allocation Pattern, solve for group sample sizes
Test Type:	Equal-Variance T-Test
Higher Means Are:	Better
Hypotheses:	H0: R ≤ RU vs. H1: R > RU
Number of Groups:	4
Bonferroni Adjustment:	Standard Bonferroni (Divisor = 3)

	Power		Sample Size				Upper			Alpha	
Comparison	Target	Actual	Ni	Allocation	Mean μi	Mean Ratio Ri	Superiority Limit RU	Standard Deviation σ	COVi	Overall	Bonferroni-Adjusted
Control			381	1.732	10.00			2.0	0.20000		
vs A	0.8	0.80164	220	1.000	1.32	1.25	1.32	2.0	0.15152	0.025	0.00833
vs B	0.8	0.95925	220	1.000	1.34	1.25	1.34	2.0	0.14925	0.025	0.00833
vs C	0.8	0.99565	220	1.000	1.36	1.25	1.36	2.0	0.14706	0.025	0.00833
Total			1041								
Control			585	1.732	10.00			2.5	0.25000		
vs A	0.8	0.80077	338	1.000	1.32	1.25	1.32	2.5	0.18939	0.025	0.00833
vs B	0.8	0.95880	338	1.000	1.34	1.25	1.34	2.5	0.18657	0.025	0.00833
vs C	0.8	0.99555	338	1.000	1.36	1.25	1.36	2.5	0.18382	0.025	0.00833
Total			1599								
Control			831	1.732	10.00			3.0	0.30000		
vs A	0.8	0.80043	480	1.000	1.32	1.25	1.32	3.0	0.22727	0.025	0.00833
vs B	0.8	0.95869	480	1.000	1.34	1.25	1.34	3.0	0.22388	0.025	0.00833
vs C	0.8	0.99554	480	1.000	1.36	1.25	1.36	3.0	0.22059	0.025	0.00833
Total			2271								

Comparison	The group that is involved in the comparison between the treatment and control displayed on this report line. The comparison is made using the ratio.
Target Power	The power desired. Power is probability of rejecting a false null hypothesis for this comparison. This power is of the comparison shown on this line only.
Actual Power	The power actually achieved.
Ni	Sample Size. The number of subjects in the ith group. The total sample size, N, is shown as the last row of the column.
Allocation	The group sample size allocation pattern. The value on each row represents the relative number of subjects assigned to the group.
μi	The mean of the ith group at which the power is computed. The first row contains μc, the control group mean.
Ri	The ratio of the means at which the power is calculated. Ri = μi / μc.
RU	The upper superiority limit for the mean ratio. RU > 1.
σ	The standard deviation of the responses within each group.
COVi	The coefficient of variation of the ith group. COVi = σ / μi.
Overall Alpha	The probability of rejecting at least one of the comparisons in this experiment when each null hypothesis is true.
Bonferroni Alpha	The adjusted significance level at which each individual comparison is made.

Summary Statements

A parallel, 4-group design (with one control group and 3 treatment groups) will be used to test whether the mean for each treatment group is superior to the control group mean by a margin, with a superiority ratio limit of 1.25 (H0: R ≤ 1.25 versus H1: R > 1.25, R = μi / μc). In this study, higher means are considered to be better. The superiority-by-a-margin hypotheses will be evaluated using 3 one-sided, two-sample, Bonferroni-adjusted, equal-variance t-tests calculated on the log-transformed data, with an overall (experiment-wise) Type I error rate (α) of 0.025. The common standard deviation for all groups is assumed to be 2. The control group mean is assumed to be 10. To detect the treatment means 13.2, 13.4, and 13.6 with at least 80% power for each test, the control group sample size needed will be 381 and the number of needed subjects for the treatment groups will be 220, 220, and 220 (totaling 1041 subjects overall).

Dropout-Inflated Sample Size

Group	Dropout Rate	Sample Size Ni	Dropout-Inflated Enrollment Sample Size Ni'	Expected Number of Dropouts Di
1	20%	381	477	96
2	20%	220	275	55
3	20%	220	275	55
4	20%	220	275	55
Total		1041	1302	261
1	20%	585	732	147
2	20%	338	423	85
3	20%	338	423	85
4	20%	338	423	85
Total		1599	2001	402
1	20%	831	1039	208
2	20%	480	600	120
3	20%	480	600	120
4	20%	480	600	120
Total		2271	2839	568

Group	Lists the group numbers.
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.
Ni	The evaluable sample size for each group at which power is computed (as entered by the user). If Ni subjects are evaluated out of the Ni' subjects that are enrolled in the study, the design will achieve the stated power.
Ni'	The number of subjects that should be enrolled in each group in order to obtain Ni evaluable subjects, based on the assumed dropout rate. Ni' is calculated by inflating Ni using the formula Ni' = Ni / (1 - DR), with Ni' 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.)
Di	The expected number of dropouts in each group. Di = Ni' - Ni.

Dropout Summary Statements

Anticipating a 20% dropout rate, group sizes of 477, 275, 275, and 275 subjects should be enrolled to obtain final group sample sizes of 381, 220, 220, and 220 subjects.

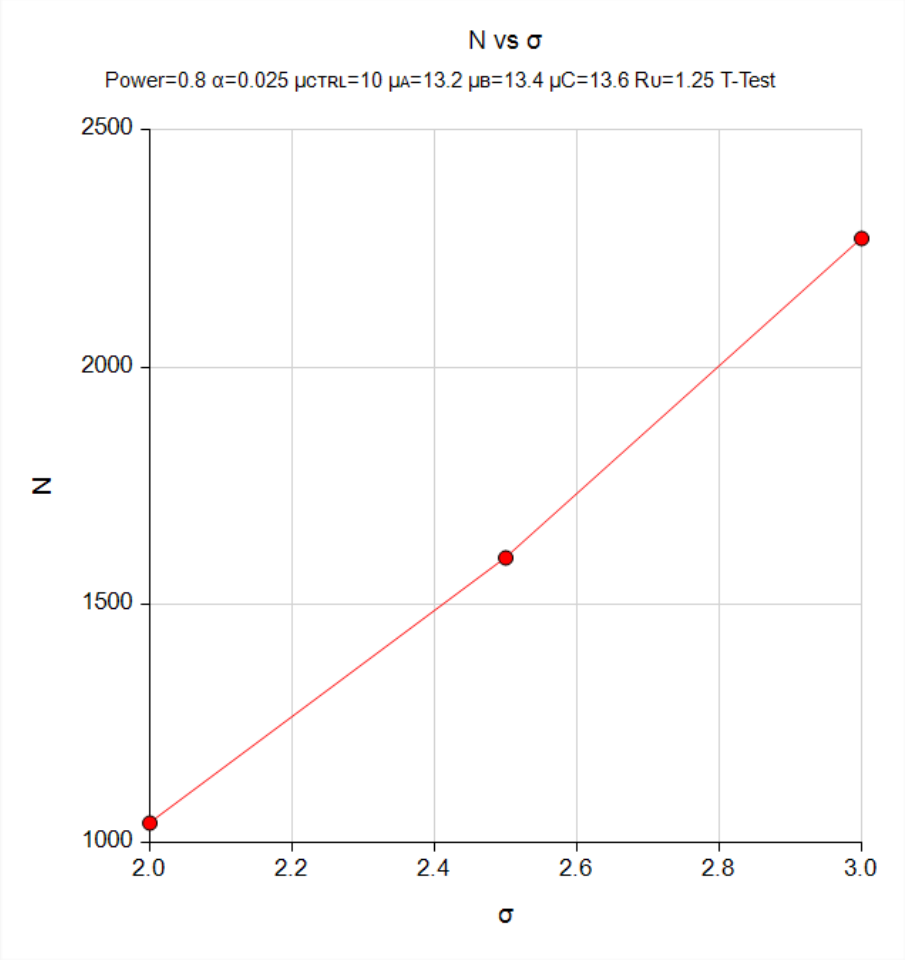
References

Blackwelder, W.C. 1998. 'Equivalence Trials.' In Encyclopedia of Biostatistics, John Wiley and Sons. New York. Volume 2, 1367-1372.
Chow, S.C., Shao, J., Wang, H., and Lokhnygina, Y. 2018. Sample Size Calculations in Clinical Research, 3rd Edition. Chapman & Hall/CRC. Boca Raton, FL. Pages 86-88.
Julious, Steven A. 2004. 'Tutorial in Biostatistics. Sample sizes for clinical trials with Normal data.' Statistics in Medicine, 23:1921-1986.
Machin, D., Campbell, M.J., Tan, S.B, and Tan, S.H. 2018. Sample Sizes for Clinical, Laboratory, and Epidemiology Studies, 4th Edition. Wiley Blackwell.

This report shows the numeric results of this power study. Notice that the results are shown in blocks of three rows at a time. Each block represents a single design.

Plots Section

Plots



This plot gives a visual presentation to the results in the Numeric Report. We can quickly see the impact on the sample size of changing the standard deviation.

Example 2 – Validation using a Previously Validated Procedure

We could not find a validation result in the statistical literature, so we will use the previously validated **PASS** procedure **Superiority by a Margin Tests for the Ratio of Two Means (Log-Normal Data)** to produce the results for the following example.

A parallel-group clinical trial is being designed to compare three treatment therapies against the standard therapy. Higher values of the response are desirable. Suppose the standard therapy has a mean response of 10 with a standard deviation of 2.5. The investigators would like a sample size large enough to find statistical significance at the 0.025 level (since this a one-sided test) if the actual mean responses of the three treatments are 13.2, 13.4, and 13.6, the power of each test is 0.80, and the superiority limit is 1.25. The alpha level used is 0.5 / 3 = 0.0083333. They want to consider a standard deviation of 2.5.

The sample sizes of all groups will be equal.

The **Superiority by a Margin Tests for the Ratio of Two Means (Log-Normal Data)** procedure is set up as follows.

Design Tab

Solve For	**Sample Size**
Higher Means Are	**Better (H1: R > 1 + SM)**
Power	**0.8**
Alpha	**0.00833** (which is Alpha / k)
Group Allocation	**Enter R = N2/N1, solve for N1 and N2**
R	**1.0**
SM (Superiority Margin)	**0.25**
R1 (Actual Ratio)	**1.32 1.34 1.36**
COV (Coefficient of Variation)	**0.25**

This set of options generates the following report.

Numeric Results for a T-Test

Solve For:	Sample Size
Ratio:	R = Treatment Mean / Reference Mean = μ1 / μ2
Higher Means Are:	Better
Hypotheses:	H0: R ≤ 1 + SM vs. H1: R > 1 + SM

Target Power	Actual Power	N1	N2	N	Target R	Actual R	SM	Bound R0	R1	COV	Alpha
0.8	0.80024	428	428	856	1	1	0.25	1.25	1.32	0.25	0.00833
0.8	0.80161	265	265	530	1	1	0.25	1.25	1.34	0.25	0.00833
0.8	0.80025	180	180	360	1	1	0.25	1.25	1.36	0.25	0.00833

A sample size of 428 in all groups will maintain a power of at least 80% for all three tests. This table contains the validation values. We will now run these values through the current procedure and compare the results with these values.

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: R > RU)**
Power of Each Test	**0.8**
Overall Alpha	**0.025**
Bonferroni Adjustment	**Standard Bonferroni**
Group Allocation	**Equal (Nc = N1 = N2 = ...)**
RU (Upper Superiority Limit)	**1.25**
Control Mean	**10**
Set A Number of Groups	**1**
Set A Mean	**13.2**
Set B Number of Groups	**1**
Set B Mean	**13.4**
Set C Number of Groups	**1**
Set C Mean	**13.6**
Set D Number of Groups	**0**
More	**Unchecked**
σ (Standard Deviation)	**2.5**

Output

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

Numeric Results

Solve For:	Sample Size
Group Allocation:	Equal (Nc = N1 = N2 = ...)
Test Type:	Equal-Variance T-Test
Higher Means Are:	Better
Hypotheses:	H0: R ≤ RU vs. H1: R > RU
Number of Groups:	4
Bonferroni Adjustment:	Standard Bonferroni (Divisor = 3)

	Power								Alpha	
Comparison	Target	Actual	Sample Size Ni	Mean μi	Mean Ratio Ri	Upper Superiority Limit RU	Standard Deviation σ	COVi	Overall	Bonferroni-Adjusted
Control			428	10.0			2.5	0.25000		
vs A	0.8	0.80028	428	13.2	1.32	1.25	2.5	0.18939	0.025	0.00833
vs B	0.8	0.95859	428	13.4	1.34	1.25	2.5	0.18657	0.025	0.00833
vs C	0.8	0.99552	428	13.6	1.36	1.25	2.5	0.18382	0.025	0.00833
Total			1712							

As you can see, the power values match to within rounding. The procedure is validated.