Chapter 335

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

Introduction

This module computes power and sample size for multiple comparisons 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 ratio of the 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

Statistical Hypotheses Based on Mean Ratios

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 $N = N_1 + N_2 + \cdots + N_k + N_C$.

Let $R = \mu_i/\mu_C$ and $R_0 = \mu_{i|H_0}/\mu_C$. For each treatment, the null and alternative two-sided hypotheses are

$$H_0: R = R_0 \quad \text{vs.} \quad H_1: R \neq R_0.$$

The one-sided hypotheses are

$$H_0: R \leq R_0 \quad \text{vs.} \quad H_1: R > R_0$$

and

$$H_0: R \geq R_0 \quad \text{vs.} \quad H_1: R < R_0.$$

In these formulas, usually R_0 is set to 1, but this is not always the case.

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 often, 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. Lower values of the response is desirable. Suppose the standard therapy has mean response of 9.3 with a standard deviation of 2.5.

The investigators would like a sample size large enough to find statistical significance at the 0.05 level if the actual mean responses of the three treatments are 7.3, 7.6, and 8.1 and the power is 0.80 in each test. They want to consider a range of standard deviations from 2.0 to 3.0. The tests will be two-sided and based on the assumption that the data are log-normal.

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**
Alternative Hypothesis	**Two-Sided (H1: R ≠ R0)**
Power of Each Test	**0.8**
Overall Alpha	**0.05**
Bonferroni Adjustment	**Standard Bonferroni**
Group Allocation	**Enter Group Allocation Pattern, solve for group sample sizes**
R0 (Mean Ratio | H0)	**1**
Control Mean	**9.3**
Control Sample Size Allocation	**1.732**
Set A Number of Groups	**1**
Set A Mean	**7.3**
Set A Sample Size Allocation	**1**
Set B Number of Groups	**1**
Set B Mean	**7.6**
Set B Sample Size Allocation	**1**
Set C Number of Groups	**1**
Set C Mean	**8.1**
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 Variances T-Test
Hypotheses:	H0: R = R0 vs. H1: R ≠ R0
Number of Groups:	4
Bonferroni Adjustment:	Standard Bonferroni (Divisor = 3)

	Power		Sample Size			Mean Ratio				Alpha	
Comparison	Target	Actual	Ni	Allocation	Mean μi	Ratio|H0 R0	Ratio|H1 Ri	Standard Deviation σ	COVi	Overall	Bonferroni-Adjusted
Control			71	1.732	9.3			2.0	0.21505		
vs A	0.8	0.99957	41	1.000	7.3	1	0.78495	2.0	0.27397	0.05	0.01667
vs B	0.8	0.99140	41	1.000	7.6	1	0.81720	2.0	0.26316	0.05	0.01667
vs C	0.8	0.80925	41	1.000	8.1	1	0.87097	2.0	0.24691	0.05	0.01667
Total			194								
Control			107	1.732	9.3			2.5	0.26882		
vs A	0.8	0.99952	62	1.000	7.3	1	0.78495	2.5	0.34247	0.05	0.01667
vs B	0.8	0.99069	62	1.000	7.6	1	0.81720	2.5	0.32895	0.05	0.01667
vs C	0.8	0.80365	62	1.000	8.1	1	0.87097	2.5	0.30864	0.05	0.01667
Total			293								
Control			151	1.732	9.3			3.0	0.32258		
vs A	0.8	0.99950	87	1.000	7.3	1	0.78495	3.0	0.41096	0.05	0.01667
vs B	0.8	0.99045	87	1.000	7.6	1	0.81720	3.0	0.39474	0.05	0.01667
vs C	0.8	0.80189	87	1.000	8.1	1	0.87097	3.0	0.37037	0.05	0.01667
Total			412								

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.
H0	The null hypothesis is H0: R = R0.
H1	The alternative hypothesis is H1: R ≠ R0.
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.
R0	The ratio of means assumed by the null hypothesis. R0 = (μi|H0) / μc.
Ri	The ratio of the means at which the power is calculated. Ri = μi / μc.
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 different from the control group mean, with a null mean ratio of 1 (H0: R = 1 versus H1: R ≠ 1, R = μi / μc). The hypotheses will be evaluated using 3 two-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 0.05. The common standard deviation for all groups is assumed to be 2. The control group mean is assumed to be 9.3. To detect the treatment means 7.3, 7.6, and 8.1 with at least 80% power for each test, the control group sample size needed will be 71 and the number of needed subjects for the treatment groups will be 41, 41, and 41 (totaling 194 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%	71	89	18
2	20%	41	52	11
3	20%	41	52	11
4	20%	41	52	11
Total		194	245	51
1	20%	107	134	27
2	20%	62	78	16
3	20%	62	78	16
4	20%	62	78	16
Total		293	368	75
1	20%	151	189	38
2	20%	87	109	22
3	20%	87	109	22
4	20%	87	109	22
Total		412	516	104

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 89, 52, 52, and 52 subjects should be enrolled to obtain final group sample sizes of 71, 41, 41, and 41 subjects.

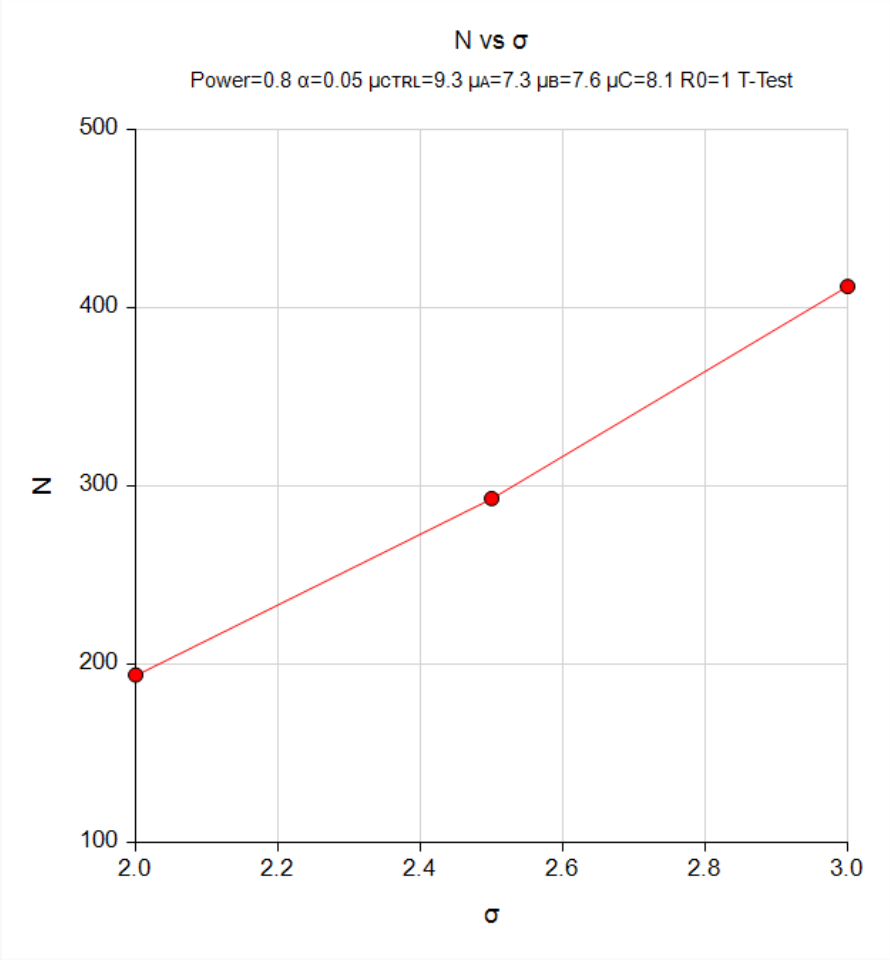
References

Rothmann, M.D., Wiens, B.L., and Chan, I.S.F. 2012. Design and Analysis of Non-Inferiority Trials. Taylor & Francis/CRC Press. Boca Raton, Florida.
Kieser, M. and Hauschke, D. 1999. 'Approximate Sample Sizes for Testing Hypotheses about the Ratio and Difference of Two Means.' Journal of Biopharmaceutical Studies, Volume 9, No. 4, pages 641-650.
Julious, Steven A. 2004. 'Tutorial in Biostatistics. Sample sizes for clinical trials with Normal data.' Statistics in Medicine, 23:1921-1986.
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.
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 a previously validated **PASS** procedure (**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. Suppose the standard therapy has mean response of 9.3 with a standard deviation of 2.5. This standard deviation results in a COV of 0.26882. The investigators would like a sample size large enough to find statistical significance at the 0.05/3 = 0.01667 level if the actual mean responses of the three treatments are 7.3, 7.6, and 8.1 and the power is 0.80 in each test. These three means result in R1 values of 0.78495, 0.81720, and 0.87097. The tests will be two-sided.

The sample sizes of all groups will be equal.

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

Design Tab

Solve For	**Sample Size**
Alternative Hypothesis	**Two-Sided (H1: R ≠ R0)**
Power	**0.8**
Alpha	**0.0166666667** (which is Alpha / k)
Group Allocation	**Equal (N1 = N2)**
R0 (Ratio | H0)	**1**
R1 (Actual Ratio)	**0.78495 0.81720 0.87097**
COV (Coefficient of Variation)	**0.26882**

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
Hypotheses: H0: R1 = R0 vs. H1: R1 ≠ R0

Target Power	Actual Power	N1	N2	N	R0	R1	COV	Effect Size	Alpha
0.8	0.81075	27	27	54	1	0.78495	0.26882	0.91668	0.01667
0.8	0.80853	38	38	76	1	0.81720	0.26882	0.76425	0.01667
0.8	0.80002	78	78	156	1	0.87097	0.26882	0.52300	0.01667

In order to maintain a power of 80% for all three groups, it is apparent that the groups will all need to have a sample size of 78. 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**
Alternative Hypothesis	**Two-Sided (H1: R ≠ R0)**
Power of Each Test	**0.8**
Overall Alpha	**0.05**
Bonferroni Adjustment	**Standard Bonferroni**
Group Allocation	**Equal (Nc = N1 = N2 = ...)**
R0 (Mean Ratio | H0)	**1**
Control Mean	**9.3**
Set A Number of Groups	**1**
Set A Mean	**7.3**
Set B Number of Groups	**1**
Set B Mean	**7.6**
Set C Number of Groups	**1**
Set C Mean	**8.1**
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 Variances T-Test
Hypotheses:	H0: R = R0 vs. H1: R ≠ R0
Number of Groups:	4
Bonferroni Adjustment:	Standard Bonferroni (Divisor = 3)

	Power		Sample Size		Mean Ratio		Standard Deviation		Alpha	
Comparison	Target	Actual	Ni	μi	Ratio|H0 R0	Ratio|H1 Ri	σ	COVi	Overall	Bonferroni-Adjusted
Control			78	9.3			2.5	0.26882		
vs A	0.8	0.99948	78	7.3	1	0.78495	2.5	0.34247	0.05	0.01667
vs B	0.8	0.99020	78	7.6	1	0.81720	2.5	0.32895	0.05	0.01667
vs C	0.8	0.80005	78	8.1	1	0.87097	2.5	0.30864	0.05	0.01667
Total			312							

As you can see, the sample sizes are all 78, which match the largest sample size found in the validation run above. The procedure is validated.