

Chapter 333

Multi-Arm Superiority by a Margin Tests for the Ratio of Treatment and Control Means (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 normal distribution and the statistical hypotheses are expressed in terms of mean ratios. Note that when the data follow a log-normal distribution rather than the normal distribution so that a log transformation is used, you should use other PASS procedures that assume a log-normal data distribution.

The details of this t-test are given in Rothmann, Wiens, and Chan (2012) and, to lesser extent, in Kieser and Hauschke (1999). 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 two means.

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 $R_L > 1$.

Equal-Variances T-Test Statistic

The ratio hypotheses are rearranged as from

$$H_0: \frac{\mu_i}{\mu_C} \leq R_U \quad \text{vs.} \quad H_1: \frac{\mu_i}{\mu_C} > R_U$$

to

$$H_0: \mu_i - R_U \mu_C \leq 0 \quad \text{vs.} \quad H_1: \mu_i - R_U \mu_C > 0$$

The null hypothesis is tested using the test statistic

$$T_1 = \frac{\bar{X}_i - R_U \bar{X}_C}{\sqrt{S \left(\frac{1}{N_i} + \frac{R_U^2}{N_C} \right)}}$$

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

where \bar{X}_i and \bar{X}_C are the sample means of the treatment and control groups, and S is the pooled estimate of the standard deviation, σ which is given by

$$S^2 = \frac{(N_i - 1)s_i^2 + (N_C - 1)s_C^2}{N_i + N_C - 2}$$

It is assumed that T_1 is distributed as a central t distribution with degrees of freedom given by $N_i + N_C - 2$. For a specified alternative R_1 , T_1 follows the noncentral t distribution with $N_i + N_C - 2$ degrees of freedom and noncentrality

$$\left(\frac{R_1 - R_U}{CV} \right) \sqrt{\frac{N_C}{\frac{N_C}{N_i} + R_U^2}}$$

Hence, the power of this test is given by noncentral t distribution as follows

$$(1 - \beta) = \Pr(T_1 \geq t_{1-\alpha, N_i + N_C - 2} | R_1, R_U, CV, N_C, N_i)$$

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 conducted 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. Higher values of the response are desirable. Suppose the standard therapy has a 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 0.025 level (since this is a one-sided test) if the actual mean responses of the three treatments are 12.2, 12.4, and 12.6, 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.

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	9.3
Control Sample Size Allocation	1.732
Set A Number of Groups	1
Set A Mean	12.2
Set A Sample Size Allocation	1
Set B Number of Groups	1
Set B Mean	12.4
Set B Sample Size Allocation	1
Set C Number of Groups	1
Set C Mean	12.6
Set C Sample Size Allocation	1
Set D Number of Groups	0
More	Unchecked
σ (Standard Deviation)	2 2.5 3

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

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: $H_0: R \leq RU$ vs. $H_1: R > RU$
 Number of Groups: 4
 Bonferroni Adjustment: Standard Bonferroni (Divisor = 3)

Comparison	Power		Sample Size		Mean μ_i	Mean Ratio R_i	Upper Superiority Limit RU	Standard Deviation σ	Coefficient of Variation COV_i	Alpha	
	Target	Actual	N_i	Allocation						Overall	Bonferroni-Adjusted
Control			419	1.732	9.3			2.0	0.21505		
vs A	0.8	0.80169	242	1.000	12.2	1.31183	1.25	2.0	0.16393	0.025	0.00833
vs B	0.8	0.97576	242	1.000	12.4	1.33333	1.25	2.0	0.16129	0.025	0.00833
vs C	0.8	0.99903	242	1.000	12.6	1.35484	1.25	2.0	0.15873	0.025	0.00833
Total			1145								
Control			653	1.732	9.3			2.5	0.26882		
vs A	0.8	0.80052	377	1.000	12.2	1.31183	1.25	2.5	0.20492	0.025	0.00833
vs B	0.8	0.97548	377	1.000	12.4	1.33333	1.25	2.5	0.20161	0.025	0.00833
vs C	0.8	0.99901	377	1.000	12.6	1.35484	1.25	2.5	0.19841	0.025	0.00833
Total			1784								
Control			940	1.732	9.3			3.0	0.32258		
vs A	0.8	0.80055	543	1.000	12.2	1.31183	1.25	3.0	0.24590	0.025	0.00833
vs B	0.8	0.97551	543	1.000	12.4	1.33333	1.25	3.0	0.24194	0.025	0.00833
vs C	0.8	0.99901	543	1.000	12.6	1.35484	1.25	3.0	0.23810	0.025	0.00833
Total			2569								

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.
N_i	The number of subjects in the i th group. The total sample size shown below the groups is equal to the sum of all individual group sample sizes.
Allocation	The group sample size allocation ratio of the i th group. The value on each row represents the relative number of subjects assigned to the group.
μ_i	The mean of the i th group at which the power is computed. The first row contains μ_c , the control group mean.
R_i	The ratio of the means at which the power is calculated. $R_i = \mu_i / \mu_c$.
RU	The upper superiority limit for the mean ratio. $RU > 1$.
σ	The standard deviation of the responses within each group.
COV_i	The coefficient of variation of the i th group. $COV_i = \sigma / \mu_i$.
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.

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

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 ($H_0: R \leq 1.25$ versus $H_1: R > 1.25$, $R = \mu_i / \mu_c$). 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, ratio-based t-tests, 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 9.3. To detect the treatment means 12.2, 12.4, and 12.6 with at least 80% power for each test, the control group sample size needed will be 419 and the number of needed subjects for the treatment groups will be 242, 242, and 242 (totaling 1145 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%	419	524	105
2	20%	242	303	61
3	20%	242	303	61
4	20%	242	303	61
Total		1145	1433	288
1	20%	653	817	164
2	20%	377	472	95
3	20%	377	472	95
4	20%	377	472	95
Total		1784	2233	449
1	20%	940	1175	235
2	20%	543	679	136
3	20%	543	679	136
4	20%	543	679	136
Total		2569	3212	643

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 524, 303, 303, and 303 subjects should be enrolled to obtain final group sample sizes of 419, 242, 242, and 242 subjects.

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

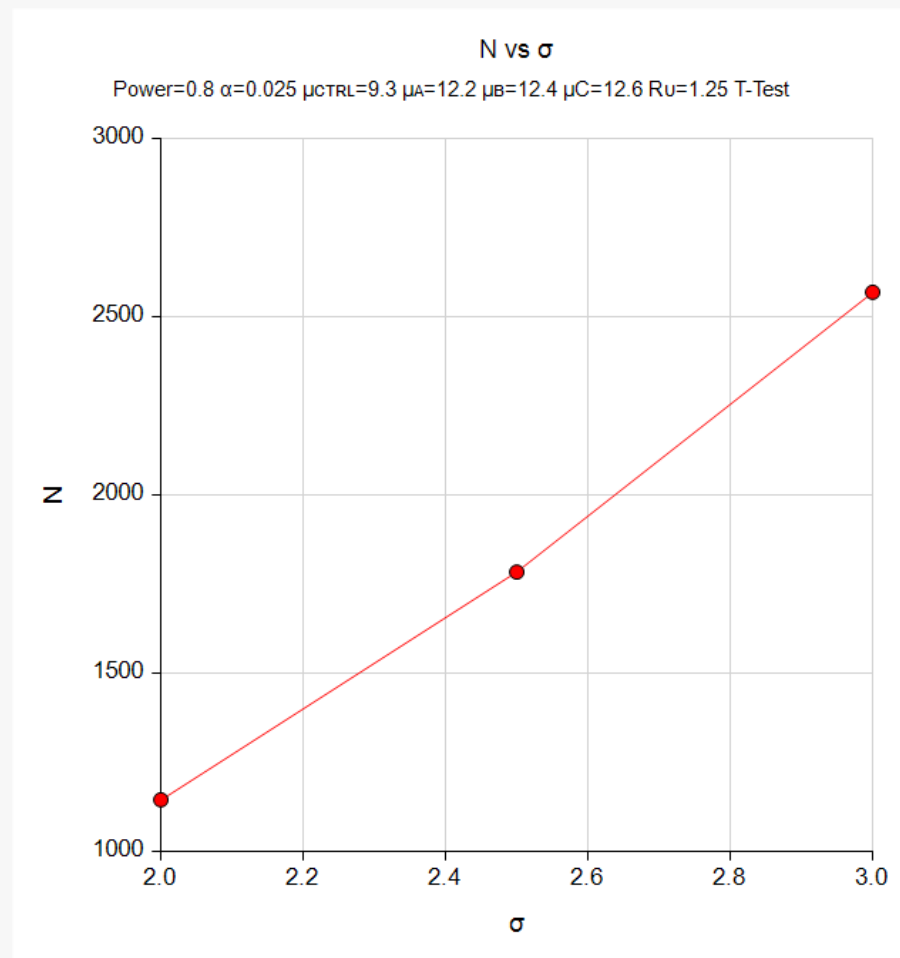
References

- 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 (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 9.3 with a standard deviation of 2.5. The investigators would like a sample size large enough to find statistical significance at the $0.025/3 = 0.0083333$ level (since this a one-sided test) if the actual mean responses of the three treatments are 12.2, 12.4, and 12.6, so the values of R1 are 1.31183, 1.33333, and 1.35484. The power of each test is 0.80 and the superiority limit is 1.25. The standard deviation is 2.5 so the COV is 0.26882.

The sample sizes of all groups will be equal.

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

Design Tab

Solve For **Power**
 Higher Means Are **Better (H1: $R > R_u$, where $R_u > 1$)**
 Alpha **0.00833** (which is Alpha / k)
 Group Allocation **Equal (N1 = N2)**
 Sample Size Per Group **508**
 Test Statistic **Equal Variances T-Test**
 R_u (Upper Superiority Limit) **1.25**
 R1 (Actual Mean Ratio, μ_1 / μ_2) **1.31183 1.33333 1.35484**
 CV (Coef of Variation, σ_2 / μ_2) **0.26882**

This set of options generates the following report.

Numeric Results

Solve For: **Power**
 Groups: 1 = Treatment, 2 = Control
 Ratio: $R = \mu_1 / \mu_2$
 Higher Means Are: Better
 Hypotheses: $H_0: R \leq R_u$ vs. $H_1: R > R_u$
 Test: Equal Variances T-Test

Power	Sample Size			Mean Ratio		Control Group Coefficient of Variation CV	Standard Deviation Ratio λ	Alpha
	N1	N2	N	Upper Superiority Limit R_u	Actual R1			
0.80059	508	508	1016	1.25	1.31183	0.26882	1	0.00833
0.97548	508	508	1016	1.25	1.33333	0.26882	1	0.00833
0.99901	508	508	1016	1.25	1.35484	0.26882	1	0.00833

A sample size of 508 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 **9.3**
 Set A Number of Groups **1**
 Set A Mean **12.2**
 Set B Number of Groups **1**
 Set B Mean **12.4**
 Set C Number of Groups **1**
 Set C Mean **12.6**
 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)

Comparison	Power		Sample Size Ni	Mean μ_i	Mean Ratio Ri	Upper Superiority Limit RU	Standard Deviation σ	Coefficient of Variation COVi	Alpha	
	Target	Actual							Overall	Bonferroni- Adjusted
Control			508	9.3			2.5	0.26882		
vs A	0.8	0.80061	508	12.2	1.31183	1.25	2.5	0.20492	0.025	0.00833
vs B	0.8	0.97550	508	12.4	1.33333	1.25	2.5	0.20161	0.025	0.00833
vs C	0.8	0.99901	508	12.6	1.35484	1.25	2.5	0.19841	0.025	0.00833
Total			2032							

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