Chapter 329

Multi-Arm Superiority by a Margin Tests for the Difference Between Treatment and Control Means Assuming Equal Variance

Introduction

This module computes power and sample size for multiple superiority by a margin tests of treatment means versus a control mean. This chapter is based on the results in Machin, Campbell, Tan, and Tan (2018). In this 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 superiority by a margin t-test of the difference between 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. For example, a popular three-arm design consists of three groups: control, treatment A, and treatment B. Two superiority by a margin 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_i. 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 (*SM*). The actual direction of the hypothesis depends on the response variable being studied.

In the following sections, define $\delta_i = \mu_i - \mu_C$.

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 a small amount (*SM*) above the control mean. The null and alternative hypotheses with are

$H_{0i}: \mu_i - \mu_C \leq SM$ vs. $H_{1i}: \mu_i - \mu_C > SM$

$H_{0i}: \mu_i \leq \mu_C + SM$ vs. $H_{1i}: \mu_i > \mu_C + SM$

$H_{0i}: \delta_i \leq SM$ vs. $H_{1i}: \delta_i > SM$

where $SM > 0$.

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 more than a small amount (*SM*) below the control mean. The null and alternative hypotheses with are

$H_{0i}: \mu_i - \mu_C \geq SM$ vs. $H_{1i}: \mu_i - \mu_C < SM$

$H_{0i}: \mu_i \geq \mu_C + SM$ vs. $H_{1i}: \mu_i < \mu_C + SM$

$H_{0i}: \delta_i \geq SM$ vs. $H_{1i}: \delta_i < SM$

where $SM < 0$.

Two-Sample Equal-Variance T-Test Statistic

Under the null hypothesis, this test assumes that the two groups of data are simple random samples from a single population of normally distributed values that all have the same mean and variance. This assumption implies that the data are continuous, and their distribution is symmetric. The calculation of the test statistic for the case when higher response values are better is as follows.

$$t_{df} = \frac{(\bar{x}_i - \bar{x}_C) - SM}{\sqrt{\frac{(N_i - 1)s_i^2 + (N_C - 1)s_C^2}{N_i + N_C - 2}\left(\frac{1}{N_i} + \frac{1}{N_C}\right)}}$$

where

$$\bar{X}_i = \frac{\sum_{j=1}^{N_i} X_{ij}}{N_i}$$

$$s_i = \sqrt{\left(\frac{\sum_{j=1}^{N_i}\left(X_{ij} - \bar{X}_i\right)^2}{(N_i - 1)}\right)}$$

$$df = N_i + N_C - 2$$

This *t*-statistic follows a *t* distribution with $N_i + N_C - 2$ degrees of freedom.

Power Calculation

The power of this test is computed using the noncentral t distribution with $N_i + N_C - 2$ degrees of freedom and non-centrality parameter

$$\lambda = \frac{\mu_i - \mu_C - SM}{\sigma\sqrt{\frac{1}{N_i} + \frac{1}{N_C}}}$$

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 0.05 and the number of treatment (not including the control) groups is five, the individual tests will be conduction using an alpha of 0.05 / 5 = 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 determine if any or all of three treatment therapies are better than the standard therapy. Higher values of the response are 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 10.6, 10.9, and 11.2. The power of each test is 0.80. The superiority margin is 10% of 9.3 = 0.93. The standard deviation ranges 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: δ > SM)**
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**
SM (Superiority Margin)	**0.93**
Control Mean	**9.3**
Control Sample Size Allocation	**1.732**
Set A Number of Groups	**1**
Set A Mean	**10.6**
Set A Sample Size Allocation	**1**
Set B Number of Groups	**1**
Set B Mean	**10.9**
Set B Sample Size Allocation	**1**
Set C Number of Groups	**1**
Set C Mean	**11.2**
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: T-Test
Higher Means Are: Better
Hypotheses: H0: δ ≤ SM vs. H1: δ > SM
Number of Groups: 4
Bonferroni Adjustment: Standard Bonferroni (Divisor = 3)

	Power		Sample Size						Alpha	
Comparison	Target	Actual	Ni	Allocation	Mean μi	Difference δi	Superiority Margin SM	Standard Deviation σ	Overall	Bonferroni-Adjusted
Control			705	1.732	9.3			2.0		
vs A	0.8	0.80047	407	1.000	10.6	1.3	0.93	2.0	0.05	0.01667
vs B	0.8	0.99942	407	1.000	10.9	1.6	0.93	2.0	0.05	0.01667
vs C	0.8	1.00000	407	1.000	11.2	1.9	0.93	2.0	0.05	0.01667
Total			1926							
Control			1102	1.732	9.3			2.5		
vs A	0.8	0.80060	636	1.000	10.6	1.3	0.93	2.5	0.05	0.01667
vs B	0.8	0.99943	636	1.000	10.9	1.6	0.93	2.5	0.05	0.01667
vs C	0.8	1.00000	636	1.000	11.2	1.9	0.93	2.5	0.05	0.01667
Total			3010							
Control			1585	1.732	9.3			3.0		
vs A	0.8	0.80020	915	1.000	10.6	1.3	0.93	3.0	0.05	0.01667
vs B	0.8	0.99942	915	1.000	10.9	1.6	0.93	3.0	0.05	0.01667
vs C	0.8	1.00000	915	1.000	11.2	1.9	0.93	3.0	0.05	0.01667
Total			4330							

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 difference.
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.
δi The difference between the ith treatment mean and the control mean (μi - μc) at which the power is computed.
σ The standard deviation of the responses within each group.
SM The margin of superiority in the scale of the mean difference. SM > 0.
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 margin of 0.93 (H0: δ ≤ 0.93 versus H1: δ > 0.93, δ = 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, 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 10.6, 10.9, and 11.2 with at least 80% power for each test, the control group sample size needed will be 705 and the number of needed subjects for the treatment groups will be 407, 407, and 407 (totaling 1926 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%	705	882	177
2	20%	407	509	102
3	20%	407	509	102
4	20%	407	509	102
Total		1926	2409	483
1	20%	1102	1378	276
2	20%	636	795	159
3	20%	636	795	159
4	20%	636	795	159
Total		3010	3763	753
1	20%	1585	1982	397
2	20%	915	1144	229
3	20%	915	1144	229
4	20%	915	1144	229
Total		4330	5414	1084

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 882, 509, 509, and 509 subjects should be enrolled to obtain final group sample sizes of 705, 407, 407, and 407 subjects.

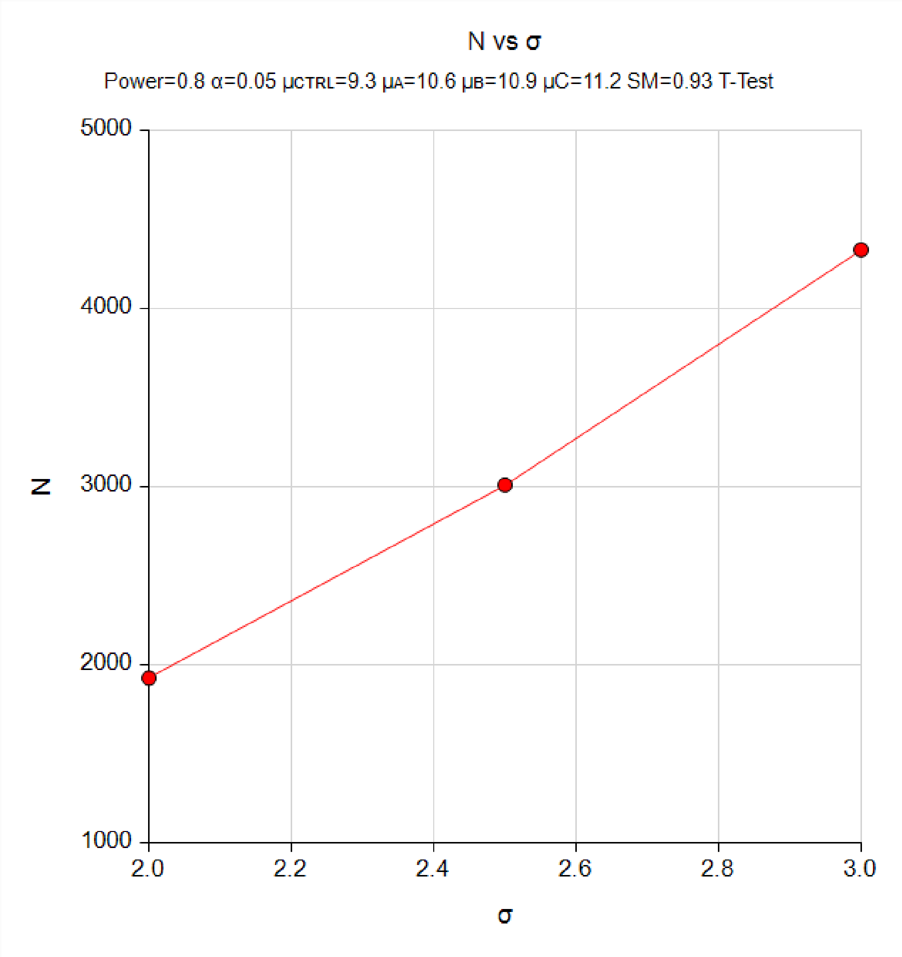
References

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 representation of the results in the Numeric Report. We can quickly see the impact on the sample size of varying 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 (**Two-Sample T-Tests for Superiority by a Margin Assuming Equal Variance**) 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 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 9.1, 9.2 and 9.3, the power of each test is 0.80, and the non-inferiority margin is -10% of 9.3 = -0.93.

The sample sizes of all groups will be equal.

The **Two-Sample T-Tests for Superiority by a Margin Assuming Equal Variance** procedure is set up as follows.

Design Tab

Solve For ....**Sample Size**
Higher Means Are...**Better (H1: δ > SM)**
Power..**0.8**
Alpha..**0.016667** (which is Alpha / k)
Group Allocation ....**Equal (N1 = N2)**
SM (Superiority Margin).................................**0.93**
δ (Actual Difference to Detect)......................**1.3 1.6 1.9**
σ (Standard Deviation)..................................**2.5**

This set of options generates the following report.

Numeric Results for an Equal-Variance T-Test

Solve For: Sample Size
Difference: δ = μ1 - μ2 = μT - μR
Higher Means Are: Better
Hypotheses: H0: δ ≤ SM vs. H1: δ > SM

Target Power	Actual Power	N1	N2	N	SM	δ	σ	Alpha
0.8	0.80032	806	806	1612	0.93	1.3	2.5	0.01667
0.8	0.80050	247	247	494	0.93	1.6	2.5	0.01667
0.8	0.80247	119	119	238	0.93	1.9	2.5	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 806. 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: δ > SM)**
Power of Each Test	**0.8**
Overall Alpha	**0.05**
Bonferroni Adjustment	**Standard Bonferroni**
Group Allocation	**Equal (Nc = N1 = N2 = ...)**
SM (Superiority Margin)	**0.93**
Control Mean	**9.3**
Set A Number of Groups	**1**
Set A Mean	**10.6**
Set B Number of Groups	**1**
Set B Mean	**10.9**
Set C Number of Groups	**1**
Set C Mean	**11.2**
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:	T-Test
Higher Means Are:	Better
Hypotheses:	H0: δ ≤ SM vs. H1: δ > SM
Number of Groups:	4
Bonferroni Adjustment:	Standard Bonferroni (Divisor = 3)

	Power		Sample					Alpha	
Comparison	Target	Actual	Size Ni	Mean μi	Difference δi	Superiority Margin SM	Standard Deviation σ	Overall	Bonferroni-Adjusted
Control			806	9.3			2.5		
vs A	0.8	0.80032	806	10.6	1.3	0.93	2.5	0.05	0.01667
vs B	0.8	0.99942	806	10.9	1.6	0.93	2.5	0.05	0.01667
vs C	0.8	1.00000	806	11.2	1.9	0.93	2.5	0.05	0.01667
Total			3224						

As you can see, the sample sizes are all 806, which match the largest sample size found in the validation run above. The procedure is validated.