Chapter 484

Superiority by a Margin Tests for Two Means in a Cluster-Randomized Design

Introduction

This procedure computes power and sample size for *superiority by a margin* tests in cluster-randomized designs in which the outcome is a continuous normal random variable.

Cluster-randomized designs are those in which whole clusters of subjects (classes, hospitals, communities, etc.) are put into the treatment group or the control group. In this case, the means of two groups, made up of K_i clusters of M_{ij} individuals each, are to be tested using a modified *z* test, or t-test, in which the clusters are treated as subjects. Generally speaking, the larger the cluster sizes and the higher the correlation among subjects within the same cluster, the larger will be the overall sample size necessary to detect an effect with the same power.

It should be noted that we could not find any published results about superiority testing with cluster-randomized designs. What we could find were Schuirmann's TOST procedure and a discussion of how to adjust the t-test sample size results given by Campbell and Walters (2014). So, we applied the Campbell and Walters adjustment to Schuirmann's test. We look forward to results that substantiate our approach.

The Statistical Hypotheses

Superiority tests are examples of directional (one-sided) tests. This program module provides the input and output in formats that are convenient for these types of tests. This section will review the specifics of superiority testing.

Remember that in the usual t-test setting, the null (H_0) and alternative (H_a) hypotheses for one-sided tests are defined as follows, assuming that $\delta = \mu_1 - \mu_2$ is to be tested.

$$H_0: \delta \leq 0 \quad \text{versus} \quad H_a: \delta > 0$$

Rejecting this test implies that the mean difference is larger than the value δ. This test is called an *upper-tailed test* because it is rejected in samples in which the difference between the sample means is larger than *D*.

Following is an example of a *lower-tailed test*.

$$H_0: \delta \geq 0 \quad \text{versus} \quad H_a: \delta < 0$$

Superiority tests are special cases of the above directional tests. It will be convenient to adopt the following specialized notation for the discussion of these tests.

Parameter	PASS Input/Output	Interpretation
μ_1	Not used	*Mean* of population 1. Population 1 is assumed to consist of those who have received the new treatment.
μ_2	Not used	*Mean* of population 2. Population 2 is assumed to consist of those who have received the reference treatment.
ε	SM	*Margin of superiority.* This is a tolerance value that defines the magnitude of the amount that is not of practical importance. This may be thought of as the smallest change from the baseline that is considered to be different.
δ	δ	*True difference*. This is the value of $\mu_1 - \mu_2$, the difference between the means.

Note that the actual values of μ_1 and μ_2 are not needed. Only their difference is needed for power and sample size calculations.

Superiority Tests

A *superiority test* tests that the treatment mean is not worse than the reference mean by more than the 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 no less than a small amount below the reference mean. The value of δ must be greater than ε. The following are equivalent sets of hypotheses.

$H_0: \delta \leq \varepsilon$ versus $H_a: \delta > \varepsilon,\ \ \varepsilon > 0$

$H_0: \mu_1 - \mu_2 \leq \varepsilon$ versus $H_a:\ \mu_1 - \mu_2 > \varepsilon,\ \ \varepsilon > 0$

$H_0: \mu_1 \leq \mu_2 + \varepsilon$ versus $H_a:\ \ \mu_1 > \mu_2 + \varepsilon,\ \ \varepsilon > 0$

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 value of δ must be less than $-\varepsilon$. The following are equivalent sets of hypotheses.

$H_0: \delta \geq -\varepsilon$ versus $H_a: \delta < -\varepsilon,\ \ \varepsilon > 0$

$H_0: \mu_1 - \mu_2 \geq -\varepsilon$ versus $H_a:\ \mu_1 - \mu_2 < -\varepsilon,\ \ \varepsilon > 0$

$H_0: \mu_1 \geq \mu_2 - \varepsilon$ versus $H_a:\ \ \mu_1 < \mu_2 - \varepsilon,\ \ \varepsilon > 0$

Technical Details

Our formulation is a combination of superiority formulas given by Chow et al. (2003) pages 57-59 and the cluster-randomized design formulas given in Campbell and Walters (2014) and Ahn, Heo, and Zhang (2015). Denote an observation by Y_{ijk} where i = 1, 2 gives the group, j = 1, 2, ..., K_i gives the cluster within group i, and k = 1, 2, ..., m_{ij} denotes an individual in cluster j of group i.

We let σ^2 denote the variance of Y_{ijk}, which is $\sigma_{Between}^2 + \sigma_{Within}^2$, where $\sigma_{Between}^2$ is the variation between clusters and σ_{Within}^2 is the variation within clusters. Also, let ρ denote the intracluster correlation coefficient (ICC) which is $\sigma_{Between}^2 / \left(\sigma_{Between}^2 + \sigma_{Within}^2\right)$. This correlation is simply the correlation between any two observations in the same cluster.

For sample size calculation, we assume that the m_{ij} are distributed with a mean cluster size of M_i and a coefficient of variation cluster sizes of *COV*. The variance of the two group means, $\bar{Y}_i$, are approximated by

$$V_i = \frac{\sigma^2 (DE_i)(RE_i)}{K_i M_i}$$

$$DE_i = 1 + (M_i - 1)\rho$$

$$RE_i = \frac{1}{1 - (COV)^2 \lambda_i (1 - \lambda_i)}$$

$$\lambda_i = M_i \rho / (M_i \rho + 1 - \rho)$$

DE is called the *Design Effect* and RE is the *Relative Efficiency* of unequal to equal cluster sizes. Both are greater than or equal to one, so both inflate the variance.

Assume that $\delta = \mu_1 - \mu_2$ is to be tested using a modified two-sample t-test. Assuming that higher values are better, the superiority test statistic is

$$t = \frac{\bar{Y}_1 - \bar{Y}_2 - \varepsilon}{\sqrt{\hat{V}_1 + \hat{V}_2}}$$

We assume this statistic has an approximate t distribution with degrees of freedom $DF = K_1 M_1 + K_2 M_2 - 2$ for a *subject-level* analysis or $K_1 + K_2 - 2$ for a *cluster-level* analysis.

Define the noncentrality parameter as $\Delta = (\delta - \varepsilon)/\sigma_d$, where $\sigma_d = \sqrt{V_1 + V_2}$. We can define the critical value based on a central t-distribution with DF degrees of freedom as follows.

$$X = t_{\alpha, DF}$$

The power can be found from the following to probabilities

$$\text{Power} = 1 - H_{X, DF, \Delta}$$

where $H_{X,DF,\Delta}$ is the cumulative probability distribution of the noncentral-t distribution.

Example 1 – Calculating Power

Suppose a superiority test is to be conducted on data obtained from a cluster-randomized design in which *SM*= 1; *δ* = 2; *σ* = 4; *ρ* = 0.0, 0.05, and 0.10; *M1* and *M2* = 10; *COV* = 0.65; *alpha* = 0.025; and *K1* and *K2* = 20, 40, or 60. Power is to be calculated assuming higher means are better.

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**
Higher Means Are .... **Better (Ha: δ > SM)**
Test Statistic .... **T-Test Based on Number of Subjects**
Alpha.. **0.025**
K1 (Number of Clusters) **20 40 60**
M1 (Average Cluster Size)............................ **10**
K2 (Number of Clusters) **K1**
M2 (Average Cluster Size)............................ **M1**
COV of Cluster Sizes.................................... **0.65**
SM (Superiority Margin)................................ **1**
δ (Mean Difference = μ1 - μ2)....................... **2**
σ (Standard Deviation).................................. **4**
ρ (Intracluster Correlation, ICC)..................... **0 0.05 0.1**

Output

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

Numeric Reports

Numeric Results for a Test of Mean Difference

Solve For: Power
Groups: 1 = Treatment, 2 = Control
Test Statistic: T-Test with DF based on number of subjects
Higher Means Are: Better
Hypotheses: H0: δ ≤ SM vs. H1: δ > SM

	Number of Clusters			Cluster Size			Sample Size			Mean Difference	Superiority Margin	Standard Deviation	ICC	
Power	K1	K2	K	M1	M2	COV	N1	N2	N	δ	SM	σ	ρ	Alpha
0.7033	20	20	40	10	10	0.65	200	200	400	2	1	4	0.00	0.025
0.9423	40	40	80	10	10	0.65	400	400	800	2	1	4	0.00	0.025
0.9911	60	60	120	10	10	0.65	600	600	1200	2	1	4	0.00	0.025
0.5039	20	20	40	10	10	0.65	200	200	400	2	1	4	0.05	0.025
0.7973	40	40	80	10	10	0.65	400	400	800	2	1	4	0.05	0.025
0.9278	60	60	120	10	10	0.65	600	600	1200	2	1	4	0.05	0.025
0.4018	20	20	40	10	10	0.65	200	200	400	2	1	4	0.10	0.025
0.6795	40	40	80	10	10	0.65	400	400	800	2	1	4	0.10	0.025
0.8440	60	60	120	10	10	0.65	600	600	1200	2	1	4	0.10	0.025

Power The probability of rejecting a false null hypothesis when the alternative hypothesis is true.
K1, K2, and K The number of clusters in groups 1 and 2, and their total.
M1 and M2 The average number of items (subjects) per cluster in groups 1 and 2, respectively.
COV The coefficient of variation of the cluster sizes.
N1, N2, and N The number of subjects in groups 1 and 2, and their total.
SM The superiority margin. Since higher means are better, this value is positive and is the distance above the group 2 (reference) mean that is still considered superior.
δ The mean difference in the response at which the power is calculated. δ = μ1 - μ2.
σ The standard deviation of the subject responses.
ρ The intra-cluster correlation (ICC). The correlation between a pair of subjects within a cluster.
Alpha The probability of rejecting a true null hypothesis.

Summary Statements

A parallel, two-group cluster-randomized design (where higher means are considered to be better) will be used to test whether the Group 1 (treatment) mean (μ1) is superior to the Group 2 (control) mean (μ2) by a margin, with a superiority margin of 1 (H0: δ ≤ 1 versus H1: δ > 1, δ = μ1 - μ2). The comparison will be made using a one-sided t-test with the degrees of freedom based on the total number of subjects (see Campbell and Walters, 2014, and Ahn, Heo, and Zhang, 2015), with a Type I error rate (α) of 0.025. The common subject-to-subject standard deviation for both groups is assumed to be 4, the intracluster correlation coefficient is assumed to be 0, and the coefficient of variation of cluster sizes is assumed to be 0.65. To detect a mean difference (μ1 - μ2) of 2, with 20 clusters of 10 subjects per cluster in Group 1 (totaling 200 subjects) and 20 clusters of 10 subjects per cluster in Group 2 (totaling 200 subjects), the power is 0.7033.

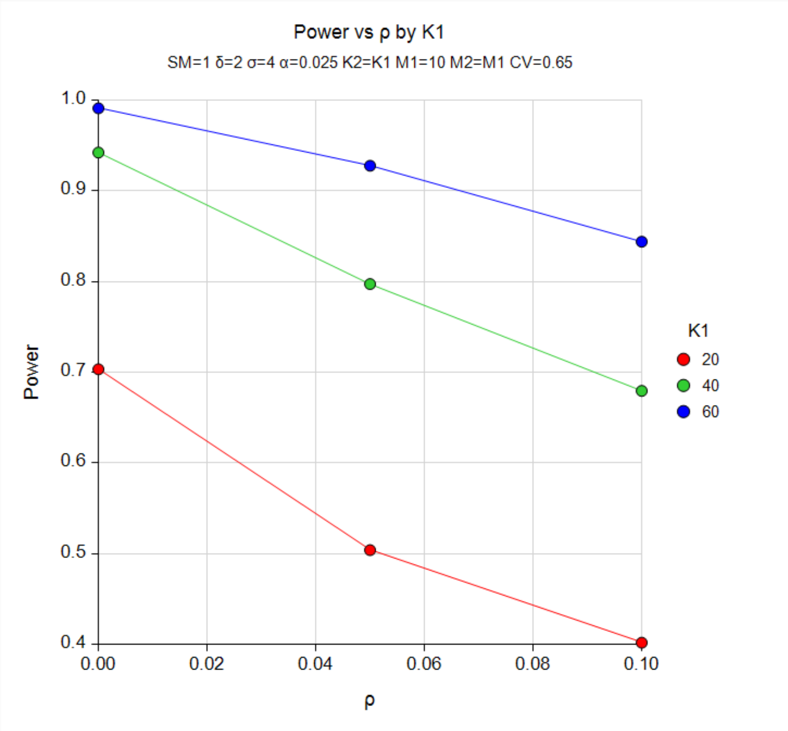
References

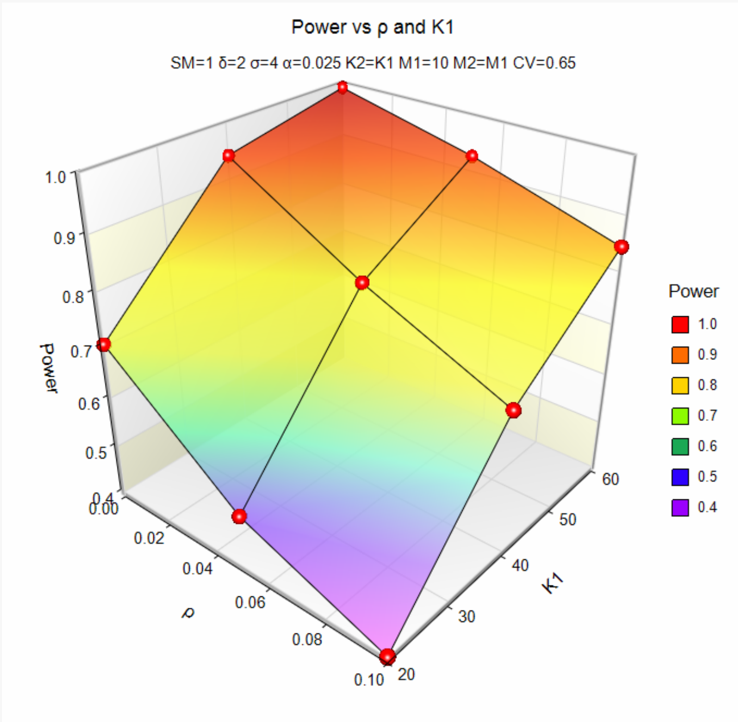
Ahn, C., Heo, M., and Zhang, S. 2015. Sample Size Calculations for Clustered and Longitudinal Outcomes in Clinical Research. CRC Press. New York.
Campbell, M.J. and Walters, S.J. 2014. How to Design, Analyse and Report Cluster Randomised Trials in Medicine and Health Related Research. Wiley. New York.
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.
Donner, A. and Klar, N. 1996. 'Statistical Considerations in the Design and Analysis of Community Intervention Trials'. J. Clin. Epidemiol. Vol 49, No. 4, pages 435-439.
Donner, A. and Klar, N. 2000. Design and Analysis of Cluster Randomization Trials in Health Research. Arnold. London.
Julious, Steven A. 2010. Sample Sizes for Clinical Trials. CRC Press. New York.

This report shows the power for each of the scenarios.

Plots Section

Plots





These plots show the results of the various scenarios specified.

Example 2 – Validation using Another PASS Procedure

We could not find a validation example for this procedure, so we will compare the results with those of the *Two-Sample T-Tests for Superiority by a Margin Assuming Equal Variance* procedure. The results should be identical when M1 = M2 = 1. Use the following scenario: find K1 when δ = 2, SM = 1, σ = 3, alpha = 0.025, and power = 0.90. That procedure obtained a value of 191 for K1 and K2.

Because M1 = 1, the values of ρ and COV are set to 0.

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 .... **K1 (Number of Clusters)**
Higher Means Are .... **Better (Ha: δ > SM)**
Test Statistic .... **T-Test Based on Number of Subjects**
Power... **0.90**
Alpha.. **0.025**
M1 (Average Cluster Size)............................ **1**
K2 (Number of Clusters) **K1**
M2 (Average Cluster Size)............................ **M1**
COV of Cluster Sizes.................................... **0**
SM (Superiority Margin)................................ **1**
δ (Mean Difference = μ1 - μ2)....................... **2**
σ (Standard Deviation).................................. **3**
ρ (Intracluster Correlation, ICC).................... **0**

Output

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

Numeric Results for a Test of Mean Difference

Solve For: K1 (Number of Clusters)
Groups: 1 = Treatment, 2 = Control
Test Statistic: T-Test with DF based on number of subjects
Higher Means Are: Better
Hypotheses: H0: δ ≤ SM vs. H1: δ > SM

	Number of Clusters			Cluster Size			Sample Size			Mean Difference	Superiority Margin	Standard Deviation	ICC	
Power	K1	K2	K	M1	M2	COV	N1	N2	N	δ	SM	σ	ρ	Alpha
0.9013	191	191	382	1	1	0	191	191	382	2	1	3	0	0.025

PASS also calculates K1 to be 191.