

Chapter 603

Multi-Arm Superiority by a Margin Tests for the Difference of Treatment and Control Proportions in a Cluster-Randomized Design

Introduction

This module computes power and sample size for multiple superiority by a margin tests for treatment proportions versus a single control proportion when the binary data are gathered from a cluster-randomized design. We could not find any published results about superiority testing with cluster-randomized designs. The formulas are based on results in Donner and Klar (2000) and Machin, Campbell, Tan, and Tan (2018).

A *cluster (group) randomized design* is one in which whole units, or clusters, of subjects are randomized to the groups rather than the individual subjects in those clusters. The conclusions of the study concern individual subjects rather than the clusters. Examples of clusters are families, school classes, neighborhoods, hospitals, and doctor's practices.

Cluster-randomized designs are often adopted when there is a high risk of contamination if cluster members were randomized individually. For example, it may be difficult for doctors to use two treatment methods in their practice. The price of randomizing by clusters is a loss of efficiency--the number of subjects needed to obtain a certain level of precision in a cluster-randomized trial is usually much larger than the number needed when the subjects are randomized individually. Hence, standard methods of sample size estimation cannot be used.

In this multi-arm design, there are G treatment groups and one control group. A mean is measured in each group. A total of G hypothesis tests are anticipated each comparing a treatment group with the common control group using a 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.

Example

Suppose that the current treatment for a disease works 50% of the time. Two promising new treatments with better performance have been developed and are now ready to be tested. Two superiority by a margin hypotheses need to be tested in this study: whether each new treatment is better than the current treatment. Hence, at least three groups are needed to complete this study of the two new treatments.

Because of the many costs of switching to a new treatment, clinicians are willing to adopt the new treatment only if it is shown to be more effective by substantial amount (margin). They must determine, however, how much more effective the new treatment must be and still be adopted. Should it be adopted if 69% respond? 68%? 65%? 60%? There is a percentage above 50% at which the difference between the two treatments cannot be ignored. After thoughtful discussion with several clinicians, it was decided that if a response of at least 60% is achieved, the new treatment will be adopted. The difference between these two percentages is called the *superiority margin*. The superiority margin in this example is 10 percentage points.

The developers must design an experiment to test the hypothesis that the response rate of the new treatment is at least 0.6. The statistical hypotheses to be tested are

$$H_0: P_A - P_C \leq \delta_0 \quad \text{vs.} \quad H_1: P_A - P_C > \delta_0$$

$$H_0: P_B - P_C \leq \delta_0 \quad \text{vs.} \quad H_1: P_B - P_C > \delta_0$$

where $\delta_0 = 0.10$.

Notice that when the null hypothesis is rejected, the conclusion is that the response rate of the treatment group is at least 0.6. Note that a rejection of the null hypothesis results in the conclusion of interest.

Multiple Treatments Versus a Single Control

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 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

Our formulation for cluster-randomized designs comes from Donner and Klar (2000). Suppose you have G treatment groups with response probabilities P_i . These groups have samples of size N_i , and one control group with response probability P_C has a sample of size N_C . The total sample size is $N = N_1 + N_2 + \dots + N_G + N_C$.

Superiority by a Margin Test Hypotheses

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 $\delta_i = P_i - P_C$ and δ_0 as the superiority difference.

Case 1: High Values Better

In this case, higher response proportions are better. The hypotheses are arranged so that rejecting the null hypothesis implies that the treatment proportion is better by at least a small amount (SM) above the control proportion. The null and alternative hypotheses with are

$$H_{0i}: P_i - P_C \leq \delta_0 \quad \text{vs.} \quad H_{1i}: P_i - P_C > \delta_0$$

$$H_{0i}: P_i \leq P_C + \delta_0 \quad \text{vs.} \quad H_{1i}: P_i > P_C + \delta_0$$

$$H_{0i}: \delta_i \leq \delta_0 \quad \text{vs.} \quad H_{1i}: \delta_i > \delta_0$$

where $\delta_0 > 0$.

Case 2: High Values Worse

In this case, lower values are better. The hypotheses are arranged so that rejecting the null hypothesis implies that the treatment proportion is more than a small amount (δ_0) below the control mean. The null and alternative hypotheses with are

$$H_{0i}: P_i - P_C \geq \delta_0 \quad \text{vs.} \quad H_{1i}: P_i - P_C < \delta_0$$

$$H_{0i}: P_i \geq P_C + \delta_0 \quad \text{vs.} \quad H_{1i}: P_i < P_C + \delta_0$$

$$H_{0i}: \delta_i \geq \delta_0 \quad \text{vs.} \quad H_{1i}: \delta_i < \delta_0$$

where $\delta_0 < 0$.

The Model

Denote a binary (0, 1) observation by Y_{ikj} where i is the group, $k = 1, 2, \dots, K_i$ is a cluster within group i , and $j = 1, 2, \dots, M_i$ is an item (often a subject) in cluster k of group i . The results that follow assume an equal number of items per cluster per group. When the number of items from cluster to cluster are about the same, the power and sample size values should be fairly accurate. In these cases, the average number of items per cluster can be used.

The statistical hypothesis that is tested concerns the difference between a treatment group proportion and the control group proportion: P_i and P_C . With a simple modification, the large-sample sample size formulas that are listed in the module for testing two proportions can be used here.

When the items are randomly assigned to one of the $G + 1$ groups, the variance of the sample proportion is

$$\sigma_{S,i}^2 = \frac{P_i(1 - P_i)}{N_i}$$

When the randomization is by clusters of items, the variance of the sample proportion is

$$\begin{aligned}\sigma_{C,i}^2 &= \frac{P_i(1 - P_i)(1 + (M_i - 1)\rho)}{K_i M_i} \\ &= \sigma_{S,i}^2 [1 + (M_i - 1)\rho] \\ &= F_{i,\rho} \sigma_{S,i}^2\end{aligned}$$

The factor $[1 + (M_i - 1)\rho]$ is called the *inflation factor* or *design effect*. The Greek letter ρ is used to represent the *intracluster correlation coefficient (ICC)*. This correlation may be thought of as the simple correlation between any two subjects within the same cluster. If we stipulate that ρ is positive, it may also be interpreted as the proportion of total variability that is attributable to differences between clusters. This value is critical to the sample size calculation

The asymptotic formulas that were used in comparing two proportions (see Chapter 200, "Tests for Two Proportions") may be used with cluster-randomized designs as well, as long as an adjustment is made for the inflation factor. The basic form of the z-test becomes

$$z = \frac{|\hat{D} - \delta_0|}{\hat{\sigma}_{\hat{D}}(\delta_0)}$$

where

$$\hat{D} = \hat{p}_i - \hat{p}_C$$

$$\delta_0 = p_i - p_C | H_0$$

$$\hat{\sigma}_{\hat{D}}(\delta_0) = \sqrt{\frac{\tilde{p}_i(1 - \tilde{p}_i)F_{i,\rho}}{N_i} + \frac{\tilde{p}_C(1 - \tilde{p}_C)F_{C,\rho}}{N_C}}$$

The quantities \tilde{p}_i and \tilde{p}_C are the maximum likelihood estimates constrained by $\tilde{p}_i - \tilde{p}_C = \delta_0$.

Test Statistics

Three test statistics are available in this routine. These are

Z-Test (or Chi-Square Test) (Pooled and Unpooled)

This test statistic was first proposed by Karl Pearson in 1900. Although this test is usually expressed directly as a Chi-Square statistic, it is expressed here as a z statistic so that it can be more easily used for one-sided hypothesis testing.

Both *pooled* and *unpooled* versions of this test have been discussed in the statistical literature. The pooling refers to the way in which the standard error is estimated. In the pooled version, the two proportions are averaged, and only one proportion is used to estimate the standard error. In the unpooled version, the two proportions are used separately.

The formula for the test statistic is

$$z_i = \frac{\hat{p}_i - \hat{p}_C}{\hat{\sigma}_D}$$

Pooled Version

$$\hat{\sigma}_D = \sqrt{\hat{p}(1 - \hat{p}) \left(\frac{F_{i,\rho}}{N_i} + \frac{F_{C,\rho}}{N_C} \right)}$$

$$\hat{p} = \frac{N_i \hat{p}_i + N_C \hat{p}_C}{N_i + N_C}$$

Unpooled Version

$$\hat{\sigma}_D = \sqrt{\frac{\hat{p}_i(1 - \hat{p}_i)F_{i,\rho}}{N_i} + \frac{\hat{p}_C(1 - \hat{p}_C)F_{C,\rho}}{N_C}}$$

Power

The following normal approximation to the binomial is used as presented in Chow et al. (2008).

1. Find the critical value (or values in the case of a two-sided test) using the standard normal distribution. The critical value is that value of z that leaves exactly the target value of alpha in the tail.
2. Use the normal approximation to binomial distribution to compute binomial probabilities, compute the power for the pooled and unpooled tests, respectively, using

$$\text{Pooled: } 1 - \beta = \Pr\left(Z < \frac{z_\alpha \sigma_{D,p} + (p_i - p_C)}{\sigma_{D,u}}\right) \quad \text{Unpooled: } 1 - \beta = \Pr\left(Z < \frac{z_\alpha \sigma_{D,u} + (p_i - p_C)}{\sigma_{D,u}}\right)$$

where

$$\sigma_{D,u} = \sqrt{\frac{p_i q_i}{N_i / F_{i,\rho}} + \frac{p_C q_C}{N_C F_{C,\rho}}} \quad (\text{unpooled standard error})$$

$$\sigma_{D,p} = \sqrt{\bar{p}\bar{q} \left(\frac{F_{i,\rho}}{N_i} + \frac{F_{C,\rho}}{N_C} \right)} \quad (\text{pooled standard error})$$

$$\text{with } \bar{p} = \frac{N_i p_i + N_C p_C}{N_i + N_C} \quad \text{and} \quad \bar{q} = 1 - \bar{p}$$

Farrington and Manning's Likelihood Score Test

Farrington and Manning (1990) proposed a test statistic for testing whether the difference is equal to a specified value δ_0 . The regular MLE's, \hat{p}_i and \hat{p}_C , are used in the numerator of the score statistic while MLE's \tilde{p}_i and \tilde{p}_C , constrained so that $\tilde{p}_i - \tilde{p}_C = \delta_0$, are used in the denominator. The significance level of the test statistic is based on the asymptotic normality of the score statistic.

The formula for computing the test statistic is

$$z_{FMD} = \frac{\hat{p}_i - \hat{p}_C - \delta_0}{\sqrt{\left(\frac{\tilde{p}_i \tilde{q}_i}{N_i / F_{i,\rho}} + \frac{\tilde{p}_C \tilde{q}_C}{N_C / F_{C,\rho}} \right)}}$$

where the estimates \tilde{p}_i and \tilde{p}_C are computed as in the corresponding test of Miettinen and Nurminen (1985).

Multiplicity Adjustment

Because G z-tests between treatment groups and the control group are run when analyzing the results of this study, many statisticians recommend that a 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 should 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).

The power can be found from the following to probabilities

$$P_1 = H_{X_1, DF, \Delta_i}$$

$$P_2 = H_{X_2, DF, \Delta_i}$$

$$\text{Power} = 1 - (P_2 - P_1)$$

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

The power of a one-sided test can be calculated similarly.

Multiplicity Adjustment

Because G t-tests between treatment groups and the control group are run when analyzing the results of this study, many statisticians recommend that a 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 should 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 clusters in this group. The standard adjustment is to include \sqrt{G} clusters in the control group for each cluster 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 sample size.

Example 1 – Finding the Sample Size

A cluster-randomized, multi-arm trial is being designed to compare three treatments against the standard drug in patients with a specific type of disease. They plan to use the Farrington and Manning likelihood score test to analyze the data.

Historically, the standard treatment has enjoyed a 50% cure rate. The new treatments increase the cure rate over the standard treatment and cost more to produce. Thus, the new treatments will be adopted if they are more effective than the standard treatment by a substantial amount. The researchers will recommend adoption of the either of the new treatments that exhibit a cure rate of at least 60%. That is, the margin of superiority is 10%. They want a study that can detect a cure rate of 65%. The researchers will recruit patients from various hospitals. All patients at a particular hospital will receive the same treatment. They anticipate an average of 20 patients per hospital. They want to see the impact on cluster count of having cluster sizes ranging for 10 to 30. The investigators would like a sample size large enough to find statistical significance at the 0.025 level and a power of 0.80 in each test. Based on similar studies, they estimate the intracluster correlation to be 0.01. Since the control group will be used three times, they set the control group allocation ratio to $\sqrt{G} = \sqrt{3} = 1.732$. The three treatment allocation ratios are set to 1.0.

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 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 Proportions Are	Better (H1: $\delta > \delta_0$)
Test Type	Likelihood Score (Farr. & Mann.)
Power of Each Test	0.8
Overall Alpha	0.025
Bonferroni Adjustment	Standard Bonferroni
Group Allocation	Enter Group Allocation Pattern, solve for group numbers of clusters
M (Items Per Cluster).....	10 20 30
δ_0 (Superiority Difference).....	0.1
Control Proportion.....	0.5
Control Items Per Cluster.....	M
Control Cluster Allocation	1.732
Set A Number of Groups.....	3
Set A Proportion	0.65
Set A Items Per Cluster	M
Set A Cluster Allocation	1
Set B Number of Groups.....	0
Set C Number of Groups	0
Set D Number of Groups	0
More.....	Unchecked
ρ (Intracluster Correlation)	0.01

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 numbers of clusters
 Test Type: Farrington and Manning Likelihood Score Test
 Hypotheses: $H_0: \delta \leq \delta_0$ vs. $H_1: \delta > \delta_0$
 Number of Groups: 4
 Bonferroni Adjustment: Standard Bonferroni (Divisor = 3)

Comparison	Power		Number of Clusters Ki	Cluster Allocation	Items Per Cluster Mi	Sample Size Ni	Proportion Pi	Difference			Alpha	
	Target	Actual						Superiority δ_0	Actual δ_i	ICC ρ	Overall	Bonferroni-Adjusted
Control			300	1.732	10	3000	0.50	0.1		0.01		
vs A1	0.8	0.80160	173	1.000	10	1730	0.65	0.1	0.15	0.01	0.025	0.00833
vs A2	0.8	0.80160	173	1.000	10	1730	0.65	0.1	0.15	0.01	0.025	0.00833
vs A3	0.8	0.80160	173	1.000	10	1730	0.65	0.1	0.15	0.01	0.025	0.00833
Total			819			8190						
Control			165	1.732	20	3300	0.50	0.1		0.01		
vs A1	0.8	0.80457	95	1.000	20	1900	0.65	0.1	0.15	0.01	0.025	0.00833
vs A2	0.8	0.80457	95	1.000	20	1900	0.65	0.1	0.15	0.01	0.025	0.00833
vs A3	0.8	0.80457	95	1.000	20	1900	0.65	0.1	0.15	0.01	0.025	0.00833
Total			450			9000						
Control			118	1.732	30	3540	0.50	0.1		0.01		
vs A1	0.8	0.80006	68	1.000	30	2040	0.65	0.1	0.15	0.01	0.025	0.00833
vs A2	0.8	0.80006	68	1.000	30	2040	0.65	0.1	0.15	0.01	0.025	0.00833
vs A3	0.8	0.80006	68	1.000	30	2040	0.65	0.1	0.15	0.01	0.025	0.00833
Total			322			9660						

- 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.
- Ki** The number of clusters in the ith group. The total number of clusters is reported in the last row of the column.
- Allocation** The cluster allocation ratio of the ith group. The value on each row represents the relative number of clusters assigned to the group.
- Mi** The average number of items per cluster (or average cluster size) in the ith group.
- Ni** The number of items in the ith group. The total sample size is shown as the last row of the column.
- Pi** The response proportion in the ith group at which the power is calculated.
- δ_0** The superiority difference or superiority margin (SM). It is the minimum size of the difference ($P_i - P_c$) that must be achieved in order for the treatment to be designated as significant.
- δ_i** The difference between the ith group proportion (P_i) and the control group proportion (P_c) at which the power is calculated. The formula is $\delta_i = P_i - P_c$.
- ρ** The intracluster correlation (ICC). The correlation between subjects within a cluster.
- 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 cluster-randomized design (with one control group and 3 treatment groups) will be used to test whether the proportion for each treatment group is superior to the control group proportion by a margin, with a superiority difference of 0.1 ($H_0: \delta \leq 0.1$ versus $H_1: \delta > 0.1$, $\delta = P_i - P_c$). The hypotheses will be evaluated using 3 one-sided, two-sample, Bonferroni-adjusted (divisor = 3) Farrington and Manning likelihood score tests, with an overall (experiment-wise) Type I error rate (α) of 0.025. The control group proportion is assumed to be 0.5. The intracluster correlation is assumed to be 0.01. The average cluster size (number of subjects or items per cluster) for the control group is assumed to be 10, and the average cluster size for each of the treatment groups is assumed to be 10, 10, and 10. To detect the treatment proportions 0.65, 0.65, and 0.65 with at least 80% power for each test, the control group cluster count needed will be 300 and the number of needed clusters for the treatment groups will be 173, 173, and 173 (totaling 819 clusters overall).

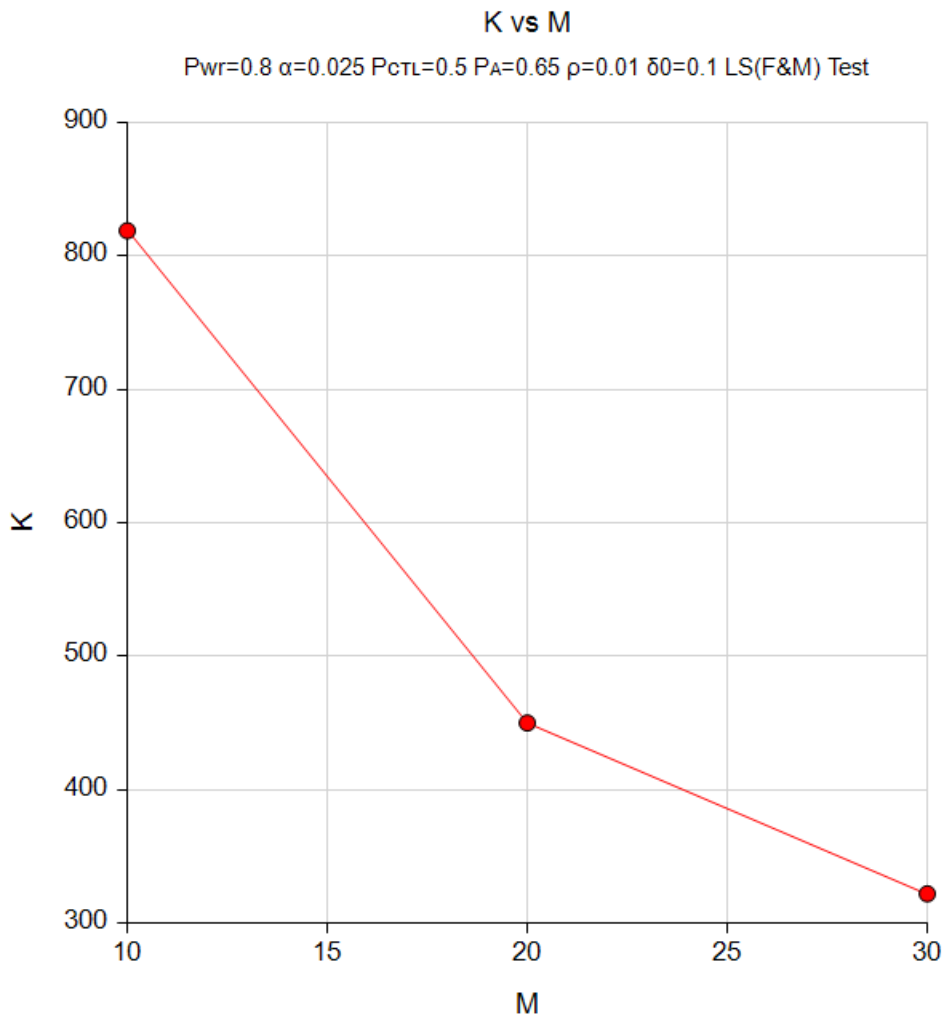
References

- Blackwelder, W.C. 1998. 'Equivalence Trials.' In Encyclopedia of Biostatistics, John Wiley and Sons. New York. Volume 2, 1367-1372.
- Chow, S.C. and Liu, J.P. 1999. Design and Analysis of Bioavailability and Bioequivalence Studies. Marcel Dekker. 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. 2000. Design and Analysis of Cluster Randomization Trials in Health Research. Arnold. London.
- Farrington, C. P. and Manning, G. 1990. 'Test Statistics and Sample Size Formulae for Comparative Binomial Trials with Null Hypothesis of Non-Zero Risk Difference or Non-Unity Relative Risk.' Statistics in Medicine, Vol. 9, pages 1447-1454.
- 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 sample size study. Notice that the results are shown in blocks of five rows at a time. Each block represents an individual treatment.

Plots Section

Plots



This plot gives a visual presentation to the results in the Numeric Report. We can quickly see the impact on the total cluster count, K , of increasing the cluster size, M .

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 (**Superiority by a Margin Tests for the Difference of Two Proportions in a Cluster-Randomized Design**) to produce the results for the following example.

Suppose that a four-arm, cluster-randomized study is to be conducted in which $P_1 = P_2 = P_3 = 0.65$, $P_C = 0.5$, $D_0 = 0.1$, $\rho = 0.01$, $K_i = 121$, $M_i = 20$, and $alpha = 0.025 / 3 = 0.008333333$. The likelihood score test will be used for the analysis. All groups will have the same number of clusters.

The calculated power is 0.80345.

The **Superiority by a Margin Tests for the Difference of Two Proportions in a Cluster-Randomized Design** procedure is set up as follows.

Design Tab

Solve For **Power**
 Higher Proportions Are **Better (H1: P1 - P2 > D0)**
 Test Type **Likelihood Score (Farr. & Mann.)**
 Alpha **0.008333333**
 K1 (Clusters in Group 1) **121**
 M1 (Average Cluster Size) **20**
 K2 (Clusters in Group 2) **K1**
 M2 (Average Cluster Size) **M1**
 Input Type **Differences**
 D0 (Superiority Difference) **0.1**
 D1 (Actual Difference) **0.15**
 P2 (Group 2 Proportion) **0.5**
 ICC (Intracluster Correlation) **0.01**

This set of options generates the following report.

Numeric Results

Solve For: **Power**
 Groups: 1 = Treatment, 2 = Reference
 Test Statistic: Likelihood Score Test (Farrington & Manning)
 Hypotheses: H0: P1 - P2 ≤ D0 vs. H1: P1 - P2 > D0

Power	Number of Clusters			Cluster Size		Total Sample Size N	Proportions			Difference		Intracluster Correlation ICC	Alpha
	K1	K2	K	M1	M2		Superiority P1.0	Actual P1.1	Reference P2	Superiority D0	Actual D1		
0.80345	121	121	242	20	20	4840	0.6	0.65	0.5	0.1	0.15	0.01	0.00833

The power is computed to be 0.80345.

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	Power
Higher Proportions Are	Better (H1: $\delta > \delta_0$)
Test Type	Likelihood Score (Farr. & Mann.)
Overall Alpha	0.025
Bonferroni Adjustment	Standard Bonferroni
Group Allocation	Equal (Kc = K1 = K2 = ...)
Ki (Group Number of Clusters)	121
M (Items Per Cluster).....	20
δ_0 (Superiority Difference).....	0.1
Control Proportion.....	0.5
Control Items Per Cluster.....	M
Set A Number of Groups.....	3
Set A Proportion	0.65
Set A Items Per Cluster	M
Set B Number of Groups.....	0
Set C Number of Groups	0
Set D Number of Groups	0
More.....	Unchecked
ρ (Intracluster Correlation)	0.01

Output

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

Numeric Results										
Solve For:		Power								
Test Type:		Farrington and Manning Likelihood Score Test								
Hypotheses:		H0: $\delta \leq \delta_0$ vs. H1: $\delta > \delta_0$								
Number of Groups:		4								
Bonferroni Adjustment:		Standard Bonferroni (Divisor = 3)								
Comparison	Power	Number of Clusters Ki	Items Per Cluster Mi	Sample Size Ni	Proportion Pi	Difference			Alpha	
						Superiority δ_0	Actual δ_i	ICC ρ	Overall	Bonferroni-Adjusted
Control		121	20	2420	0.50	0.1		0.01		
vs A1	0.80345	121	20	2420	0.65	0.1	0.15	0.01	0.025	0.00833
vs A2	0.80345	121	20	2420	0.65	0.1	0.15	0.01	0.025	0.00833
vs A3	0.80345	121	20	2420	0.65	0.1	0.15	0.01	0.025	0.00833
Total		484		9680						

As you can see, the power is 0.80345 for all treatment groups which matches the power found in the validation run above. The procedure is validated.