

## Chapter 237

# Superiority by a Margin Tests for the Difference of Two Proportions in a Cluster-Randomized Design

## Introduction

This module provides power analysis and sample size calculation for superiority by a margin tests of the difference in two-sample, cluster-randomized designs in which the outcome is binary.

## Technical Details

Our formulation comes from Donner and Klar (2000). Denote a binary observation by  $Y_{gkm}$  where  $g = 1$  or  $2$  is the group,  $k = 1, 2, \dots, K_g$  is a cluster within group  $g$ , and  $m = 1, 2, \dots, M_g$  is an individual in cluster  $k$  of group  $g$ . The results that follow assume an equal number of individuals per cluster. When the number of subjects 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 subjects per cluster can be used.

The statistical hypothesis that is tested concerns the difference between the two group proportions,  $p_1$  and  $p_2$ . When necessary, we assume that group 1 is the treatment group and group 2 is the control group. With a simple modification, all of the large-sample sample size formulas that are listed in the module for testing two proportions can be used here.

When the individual subjects are randomly assigned to one of the two groups, the variance of the sample proportion is

$$\sigma_{S,g}^2 = \frac{p_g(1 - p_g)}{n_g}$$

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

$$\begin{aligned}\sigma_{C,g}^2 &= \frac{p_g(1 - p_g)(1 + (m_g - 1)\rho)}{k_g m_g} \\ &= \sigma_{S,g}^2 [1 + (m_g - 1)\rho] \\ &= F_{g,\rho} \sigma_{S,g}^2\end{aligned}$$

## Superiority by a Margin Tests for the Difference of Two Proportions in a Cluster-Randomized Design

The factor  $[1 + (m_g - 1)\rho]$  is called the *inflation factor*. The Greek letter  $\rho$  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  $\rho$  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 195, "Superiority by a Margin Tests for the Difference Between Two Proportions") may be used with cluster-randomized designs as well, as long as an adjustment is made for the inflation factor.

---

## Power Calculations

A large sample approximation may be used that is most accurate when the values of  $n_1$  and  $n_2$  are large. The large approximation is made by replacing the values of  $\hat{p}_1$  and  $\hat{p}_2$  in the  $z$  statistic with the corresponding values of  $p_1$  and  $p_2$  under the alternative hypothesis, and then computing the results based on the normal distribution.

Note that in this case, exact calculations are not possible.

## Example 1 – Finding Power

A study is being designed to establish the superiority of a new treatment compared to the current treatment. Historically, the current treatment has enjoyed a 60% cure rate. The new treatment is hoped to perform better than the current treatment. Thus, the new treatment will be adopted if it is more effective than the current treatment by a clinically significant amount. The researchers will recommend adoption of the new treatment if it has a cure rate of at least 70%.

The researchers will recruit patients from various hospitals. All patients at a particular hospital will receive the same treatment. They anticipate an average of 100 patients per hospital. Based on similar studies, they estimate the intracluster correlation to be 0.002.

The researchers plan to use the Farrington and Manning likelihood score test statistic to analyze the data. They want to study the power of the one-sided Farrington and Manning test at group cluster sizes ranging from 2 to 10 when the superiority difference is 0.1 and the actual cure rate of the new treatment ranges from 71% to 80%. The significance level will be 0.05.

### 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 .....	<b>Power</b>
Higher Proportions Are .....	<b>Better (H1: P1 - P2 &gt; D0)</b>
Test Type .....	<b>Likelihood Score (Farr. &amp; Mann.)</b>
Alpha .....	<b>0.05</b>
K1 (Clusters in Group 1) .....	<b>2 4 6 8 10</b>
M1 (Average Cluster Size) .....	<b>100</b>
K2 (Clusters in Group 2) .....	<b>K1</b>
M2 (Average Cluster Size) .....	<b>M1</b>
Input Type .....	<b>Differences</b>
D0 (Superiority Difference) .....	<b>0.1</b>
D1 (Actual Difference) .....	<b>0.11 0.14 0.17 0.2</b>
P2 (Group 2 Proportion) .....	<b>0.6</b>
ICC (Intracluster Correlation) .....	<b>0.002</b>

## Superiority by a Margin Tests for the Difference of Two Proportions in a Cluster-Randomized Design

## Output

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

## Numeric Reports

## Numeric Results

Solve For: **Power**  
 Groups: 1 = Treatment, 2 = Reference  
 Test Statistic: Likelihood Score Test (Farrington & Manning)  
 Hypotheses:  $H_0: P_1 - P_2 \leq D_0$  vs.  $H_1: P_1 - P_2 > D_0$

Power	Number of Clusters			Cluster Size		Total Sample Size N	Proportions			Difference		Intraclass Correlation ICC	Alpha
	K1	K2	K	M1	M2		Superiority P1.0	Actual P1.1	Reference P2	Superiority D0	Actual D1		
0.07313	2	2	4	100	100	400	0.7	0.71	0.6	0.1	0.11	0.002	0.05
0.08492	4	4	8	100	100	800	0.7	0.71	0.6	0.1	0.11	0.002	0.05
0.09489	6	6	12	100	100	1200	0.7	0.71	0.6	0.1	0.11	0.002	0.05
0.10394	8	8	16	100	100	1600	0.7	0.71	0.6	0.1	0.11	0.002	0.05
0.11243	10	10	20	100	100	2000	0.7	0.71	0.6	0.1	0.11	0.002	0.05
0.19326	2	2	4	100	100	400	0.7	0.74	0.6	0.1	0.14	0.002	0.05
0.29446	4	4	8	100	100	800	0.7	0.74	0.6	0.1	0.14	0.002	0.05
0.38568	6	6	12	100	100	1200	0.7	0.74	0.6	0.1	0.14	0.002	0.05
0.46812	8	8	16	100	100	1600	0.7	0.74	0.6	0.1	0.14	0.002	0.05
0.54203	10	10	20	100	100	2000	0.7	0.74	0.6	0.1	0.14	0.002	0.05
0.39804	2	2	4	100	100	400	0.7	0.77	0.6	0.1	0.17	0.002	0.05
0.62615	4	4	8	100	100	800	0.7	0.77	0.6	0.1	0.17	0.002	0.05
0.77839	6	6	12	100	100	1200	0.7	0.77	0.6	0.1	0.17	0.002	0.05
0.87328	8	8	16	100	100	1600	0.7	0.77	0.6	0.1	0.17	0.002	0.05
0.92957	10	10	20	100	100	2000	0.7	0.77	0.6	0.1	0.17	0.002	0.05
0.64643	2	2	4	100	100	400	0.7	0.80	0.6	0.1	0.20	0.002	0.05
0.88913	4	4	8	100	100	800	0.7	0.80	0.6	0.1	0.20	0.002	0.05
0.96934	6	6	12	100	100	1200	0.7	0.80	0.6	0.1	0.20	0.002	0.05
0.99221	8	8	16	100	100	1600	0.7	0.80	0.6	0.1	0.20	0.002	0.05
0.99814	10	10	20	100	100	2000	0.7	0.80	0.6	0.1	0.20	0.002	0.05

H0	An abbreviation for the null hypothesis. This is the hypothesis being evaluated by the statistical test.
H1	An abbreviation for the alternative hypothesis. This hypothesis gives the "true" parameter values.
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, 2, and both, respectively.
M1 and M2	The average number of items (subjects) per cluster in groups 1 and 2, respectively.
N	The total number of subjects in the study. $N = (K1 \times M1) + (K2 \times M2)$ .
P1.0	The superiority limit for the group 1 proportion.
P1.1	The actual group 1 proportion assumed by the alternative hypothesis. This is the value at which the power is calculated.
P2	The proportion for group 2, the standard, reference, baseline, or control group.
D0	The superiority difference. This is the smallest difference between P1 and P2 for which group 1 will be considered superior to group 2.
D1	The actual proportion difference at which the power is calculated. $D1 = P1.1 - P2$ .
ICC	The intraclass correlation. This may be interpreted as the correlation between any two observations in the same cluster.
Alpha	The probability of rejecting a true null hypothesis.

## Summary Statements

A parallel two-group cluster-randomized design will be used to test whether the Group 1 (treatment) proportion (P1) is superior to the Group 2 (reference) proportion (P2) by a margin, with a superiority margin of 0.1 ( $H_0: P_1 - P_2 \leq 0.1$  versus  $H_1: P_1 - P_2 > 0.1$ ). The comparison will be made using a one-sided Likelihood Score Test (Farrington & Manning) based on the proportion difference ( $P_1 - P_2$ ), with a Type I error rate ( $\alpha$ ) of 0.05. The reference group proportion (P2) is assumed to be 0.6. The intraclass correlation is assumed to be 0.002. To detect a proportion difference ( $P_1 - P_2$ ) of 0.11 (or P1 of 0.71), with 2 clusters of 100 subjects per cluster in Group 1 and 2 clusters of 100 subjects per cluster in Group 2, the power is 0.07313.

## Superiority by a Margin Tests for the Difference of Two Proportions in a Cluster-Randomized Design

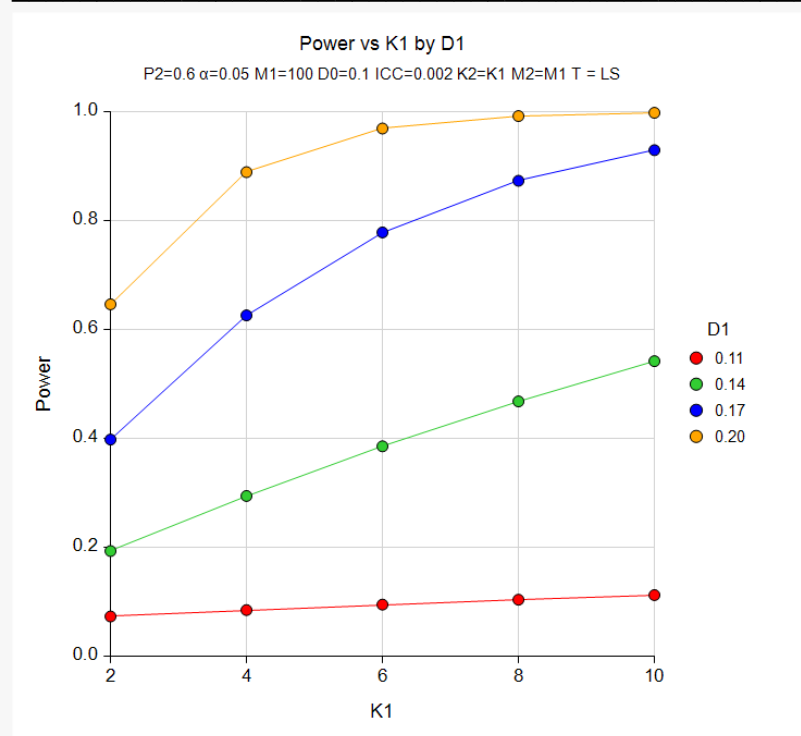
## References

- 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.
- 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.
- Gart, John J. and Nam, Jun-mo. 1988. 'Approximate Interval Estimation of the Ratio in Binomial Parameters: A Review and Corrections for Skewness.' Biometrics, Volume 44, Issue 2, 323-338.
- Gart, John J. and Nam, Jun-mo. 1990. 'Approximate Interval Estimation of the Difference in Binomial Parameters: Correction for Skewness and Extension to Multiple Tables.' Biometrics, Volume 46, Issue 3, 637-643.
- Miettinen, O.S. and Nurminen, M. 1985. 'Comparative analysis of two rates.' Statistics in Medicine 4: 213-226.

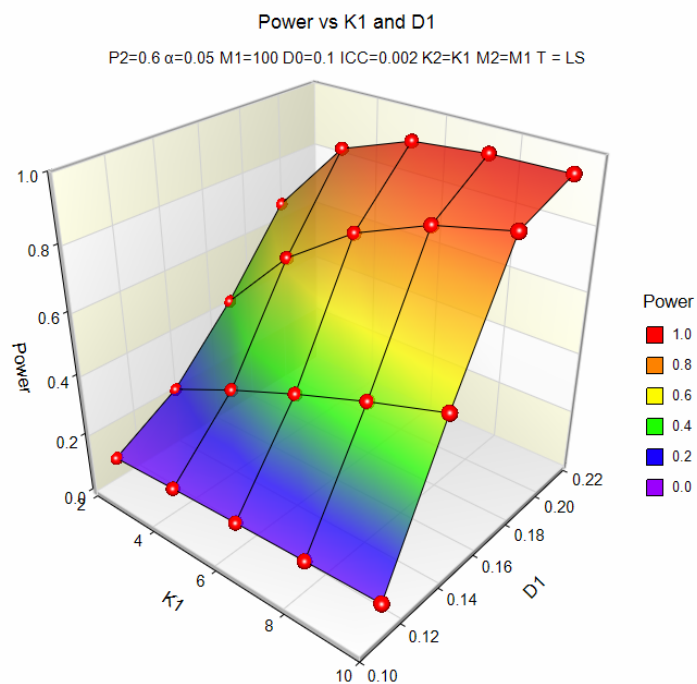
This report shows the values of each of the parameters, one scenario per row. The total number of items sampled in group 1 is  $N1 = K1 \times M1$ . The total number of items sampled in group 2 is  $N2 = K2 \times M2$ .

## Plots Section

## Plots



## Superiority by a Margin Tests for the Difference of Two Proportions in a Cluster-Randomized Design



The values from the table are displayed on the above plots. These plots give a quick look at the sample sizes that will be required for various values of D1.

## Example 2 – Finding the Sample Size (Number of Clusters)

Continuing with the scenario given in Example 1, the researchers want to determine the number of clusters necessary for each value of D1 when the target power is set to 0.80.

### 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 (K1)**  
 Higher Proportions Are ..... **Better (H1: P1 - P2 > D0)**  
 Test Type..... **Likelihood Score (Farr. & Mann.)**  
 Power..... **0.80**  
 Alpha..... **0.05**  
 M1 (Average Cluster Size)..... **100**  
 K2 (Clusters in Group 2) ..... **K1**  
 M2 (Average Cluster Size)..... **M1**  
 Input Type..... **Differences**  
 D0 (Superiority Difference) ..... **0.1**  
 D1 (Actual Difference)..... **0.11 0.14 0.17 0.2**  
 P2 (Group 2 Proportion)..... **0.6**  
 ICC (Intraclass Correlation)..... **0.002**

### Output

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

#### Numeric Results

Solve For: **Sample Size (K1)**  
 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		Intraclass Correlation ICC	Alpha
	K1	K2	K	M1	M2		Superiority P1.0	Actual P1.1	Reference P2	Superiority D0	Actual D1		
0.80034	331	331	662	100	100	66200	0.7	0.71	0.6	0.1	0.11	0.002	0.05
0.81458	21	21	42	100	100	4200	0.7	0.74	0.6	0.1	0.14	0.002	0.05
0.83175	7	7	14	100	100	1400	0.7	0.77	0.6	0.1	0.17	0.002	0.05
0.88913	4	4	8	100	100	800	0.7	0.80	0.6	0.1	0.20	0.002	0.05

The required sample size depends a great deal on the value of D1. The researchers should spend time determining the most appropriate value for D1.

## Example 3 – Finding Power after an Experiment

A group of researchers want to show that a new treatment works better than the current treatment. They believe, in fact, that the new treatment is about 0.20 higher in proportion of success. One hundred patients at each of 10 randomly chosen hospitals were given the current treatment. One hundred patients at each of 10 randomly chosen hospitals were given the new treatment. It was agreed before the experiment that the new treatment needed to be greater than 0.1 in proportion of success above the current treatment to be considered superior. The proportion of patients responding to the current treatment was  $721/1000 = 0.721$ . The proportion of patients responding to the new treatment was  $835/1000 = 0.835$ . This result did not show significant superiority at the 0.05 level. The researchers want to know the power of their superiority by a margin test. They decide to use the intracluster correlation coefficient estimated from the data, which was 0.0068. Although the observed difference in proportions is  $0.835 - 0.721 = 0.114$ , the trivial difference is still 0.1. This value is used in the power calculation.

### 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 3** 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 .....	<b>Power</b>
Higher Proportions Are .....	<b>Better (H1: P1 - P2 &gt; D0)</b>
Test Type .....	<b>Likelihood Score (Farr. &amp; Mann.)</b>
Alpha .....	<b>0.05</b>
K1 (Clusters in Group 1) .....	<b>10</b>
M1 (Average Cluster Size) .....	<b>100</b>
K2 (Clusters in Group 2) .....	<b>K1</b>
M2 (Average Cluster Size) .....	<b>M1</b>
Input Type .....	<b>Differences</b>
D0 (Superiority Difference) .....	<b>0.1</b>
D1 (Actual Difference) .....	<b>0.114 0.2</b>
P2 (Group 2 Proportion) .....	<b>0.721</b>
ICC (Intracluster Correlation) .....	<b>0.0068</b>



## Superiority by a Margin Tests for the Difference of Two Proportions in a Cluster-Randomized Design

## Output

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

### Numeric Results

Solve For: [Power](#)  
 Groups: 1 = Treatment, 2 = Reference  
 Test Statistic: Likelihood Score Test (Farrington & Manning)  
 Hypotheses:  $H_0: P_1 - P_2 \leq D_0$  vs.  $H_1: P_1 - P_2 > D_0$

	Number of Clusters			Cluster Size		Total Sample Size N	Proportions			Difference		Intraclass Correlation ICC	Alpha
	K1	K2	K	M1	M2		Superiority P1.0	Actual P1.1	Reference P2	Superiority D0	Actual D1		
<a href="#">0.14515</a>	10	10	20	100	100	2000	0.821	0.835	0.721	0.1	0.114	0.0068	0.05
<a href="#">0.99877</a>	10	10	20	100	100	2000	0.821	0.921	0.721	0.1	0.200	0.0068	0.05

If indeed the new treatment were 0.20 higher in proportion of success, the power for showing superiority by a margin would be 0.99877. If the true difference is 0.114, the power would be 0.14515.

## Example 4 – Validation

This procedure uses the same mechanics as the Tests for Two Proportions in a Cluster-Randomized Design procedure. We refer the user to Example 4 of Chapter 230 for the validation.