Chapter 355

Multi-Arm Tests for Treatment and Control Proportions in a Cluster-Randomized Design

Introduction

This module computes power and sample size for multiple comparisons of treatment proportions versus a control proportion when the data are gathered from a cluster-randomized design. The formulas are based on results in Machin, Campbell, Tan, and Tan (2018) and Donner and Klar (2000).

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 proportion 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 simple z-test of the difference between two proportions.

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 tests are run: treatment A versus control and treatment B versus the same control. This 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 cluster-randomized designs comes from Donner and Klar (2000). Suppose you have G treatment groups with response probabilities P_i that have samples of size N_i and one control group with response probability P_C that has a sample of size N_C. The total sample size is $N = N_1 + N_2 + ... + N_G + N_C$.

The hypotheses for two-sided tests are

$$H_{0i}: P_i = P_C \quad \text{versus} \quad H_{1i}: P_i \neq P_C$$

and for one-sided tests the hypotheses are

$$H_{0i}: P_i \leq P_C \quad \text{versus} \quad H_{1i}: P_i > P_C$$

or

$$H_{0i}: P_i \geq P_C \quad \text{versus} \quad H_{1i}: P_i < P_C$$

Denote a binary observation by Y_{ikj} where i is the group, k = 1, 2, ..., K_i is a cluster within group i, and j = 1, 2, ..., M_i is an item 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

$$\sigma_{C,i}^2 = \frac{P_i(1 - P_i)(1 + (M_i - 1)\rho)}{K_i M_i}$$

$$= \sigma_{S,i}^2 [1 + (M_i - 1)\rho]$$

$$= F_{i,\rho} \sigma_{S,i}^2$$

The factor $[1 + (M_i - 1)\rho]$ is called the *inflation factor*. 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

$$z = \frac{|\hat{D} - 0|}{\hat{\sigma}_{\hat{D}}(\delta_0)}$$

where

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

$$\hat{\sigma}_{\hat{D}}(0) = \sqrt{\frac{\tilde{p}_i(1 - \tilde{p}_i)F_{i,\rho}}{N_i} + \frac{\tilde{p}_C(1 - \tilde{p}_C)F_{C,\rho}}{N_C}}$$

The quantities $\tilde{p}_i$ and $\tilde{p}_C$ are the maximum likelihood estimates constrained by $\tilde{p}_i - \tilde{p}_C = 0$.

Test Statistics

Two 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

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

where

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

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

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

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).

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, parallel-group clinical trial is being designed to compare three doses of a test compound against the standard drug in patients with a specific type of disease. Suppose the standard drug has a response rate of 60%. The investigators would like a sample size large enough to find statistical significance at the 0.05 level if the actual response rate is at least 70% and the power is 0.80 in each test. They want to see the impact on cluster count of having cluster sizes ranging for 10 to 50.

The researchers want to examine effect of the number of clusters used in each group and the average number of patients in each cluster on the power of the test. They plan to use the two-sided unpooled Z-test to analyze the data. Based on similar studies, the intracluster correlation is estimated to be 0.02. The researchers want to calculate the required sample size when ICC is either 0.01 or 0.02.

The control group multiplier will be set to $\sqrt{G} = \sqrt{2} = 1.414$ since the control group is used for two 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**
Alternative Hypothesis ....**Two-Sided (H1: Pi ≠ Pc)**
Test Type...**Z-Test (Unpooled)**
Power of Each Test **0.90**
Overall Alpha **0.05**
Bonferroni Adjustment ....**Standard Bonferroni**
Group Allocation ....**Enter Group Allocation Pattern, solve for group numbers of clusters**
M (Items Per Cluster)....................................**10 20 30 40 50**
Control Proportion...**0.6**
Control Items Per Cluster...............................**M**
Control Cluster Allocation **1.414**
Set A Number of Groups................................**1**
Set A Proportion **0.7**
Set A Items Per Cluster ....**M**
Set A Cluster Allocation **1**
Set B Number of Groups................................**1**
Set B Proportion **0.7**
Set B Items Per Cluster ....**M**
Set B Cluster Allocation **1**
Set C Number of Groups **0**
Set D Number of Groups **0**
More...**Unchecked**
ρ (Intracluster Correlation) **0.01 0.02**

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:	Unpooled Z-Test
Hypotheses:	H0: δ = 0 vs. H1: δ ≠ 0
Number of Groups:	3
Bonferroni Adjustment:	Standard Bonferroni (Divisor = 2)

_____

										Alpha	
	Power		**Number of Clusters**	**Cluster**	**Items Per Cluster**	**Sample Size**	**Proportion**	**Difference**	**ICC**		**Bonferroni-**
Comparison	**Target**	**Actual**	**Ki**	**Allocation**	**Mi**	**Ni**	**Pi**	**δi**	**ρ**	**Overall**	**Adjusted**
Control			74	1.414	10	740	0.6		0.01		
vs A	0.9	0.90458	52	1.000	10	520	0.7	0.1	0.01	0.05	0.025
vs B	0.9	0.90458	52	1.000	10	520	0.7	0.1	0.01	0.05	0.025
Total			178			1780					
Control			79	1.414	10	790	0.6		0.02		
vs A	0.9	0.90182	56	1.000	10	560	0.7	0.1	0.02	0.05	0.025
vs B	0.9	0.90182	56	1.000	10	560	0.7	0.1	0.02	0.05	0.025
Total			191			1910					
Control			40	1.414	20	800	0.6		0.01		
vs A	0.9	0.90095	28	1.000	20	560	0.7	0.1	0.01	0.05	0.025
vs B	0.9	0.90095	28	1.000	20	560	0.7	0.1	0.01	0.05	0.025
Total			96			1920					
Control			47	1.414	20	940	0.6		0.02		
vs A	0.9	0.90545	33	1.000	20	660	0.7	0.1	0.02	0.05	0.025
vs B	0.9	0.90545	33	1.000	20	660	0.7	0.1	0.02	0.05	0.025
Total			113			2260					
Control			30	1.414	30	900	0.6		0.01		
vs A	0.9	0.91198	21	1.000	30	630	0.7	0.1	0.01	0.05	0.025
vs B	0.9	0.91198	21	1.000	30	630	0.7	0.1	0.01	0.05	0.025
Total			72			2160					
Control			35	1.414	30	1050	0.6		0.02		
vs A	0.9	0.90084	25	1.000	30	750	0.7	0.1	0.02	0.05	0.025
vs B	0.9	0.90084	25	1.000	30	750	0.7	0.1	0.02	0.05	0.025
Total			85			2550					
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.

_____

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.
δi	The difference between the ith group proportion (Pi) and the control group proportion (Pc) at which the power is calculated. The formula is δi = Pi - Pc.
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, 3-group cluster-randomized design (with one control group and 2 treatment groups) will be used to test whether the proportion for each treatment group is different from the control group proportion (H0: Pi = Pc versus H1: Pi ≠ Pc). The hypotheses will be evaluated using 2 two-sided, two-sample, Bonferroni-adjusted unpooled Z-tests, with an overall (experiment-wise) Type I error rate (α) of 0.05. The control group proportion is assumed to be 0.6. 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 and 10. To detect the treatment proportions 0.7 and 0.7 with at least 90% power for each test, the control group cluster count needed will be 74 and the number of needed clusters for the treatment groups will be 52 and 52 (totaling 178 clusters overall).

───

References

───

Donner, A. and Klar, N. 2000. Design and Analysis of Cluster Randomization Trials in Health Research. Arnold. London.
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.
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
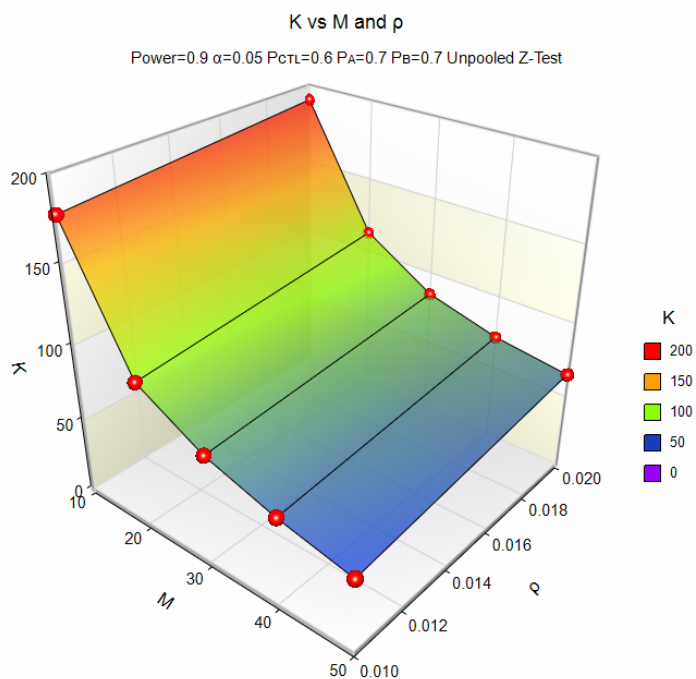
Plots

───
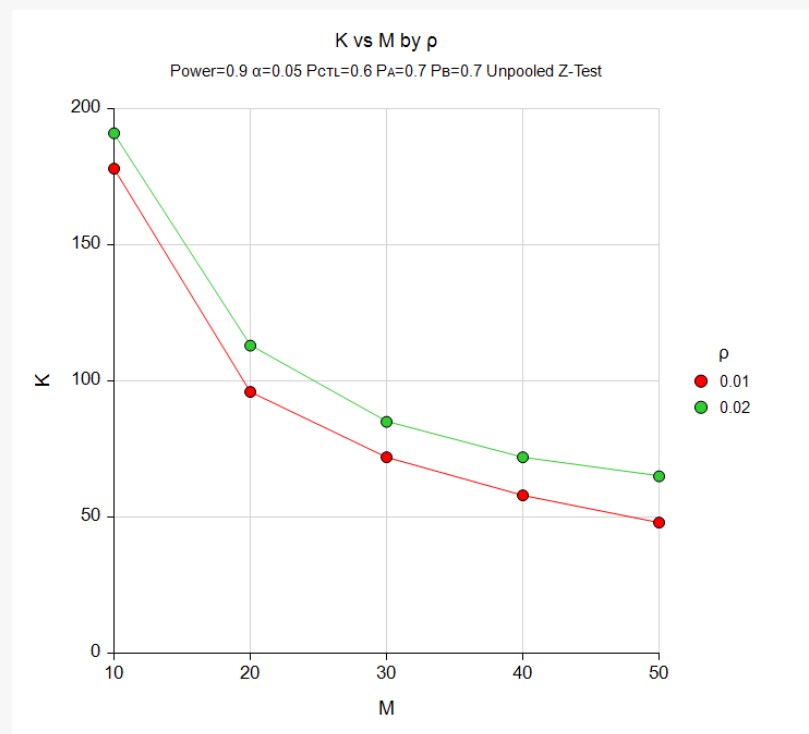
Multi-Arm Tests for Treatment and Control Proportions in a Cluster-Randomized Design

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

A cluster-randomized, parallel-group clinical trial is being designed to compare three doses of a test compound against the standard therapy in patients with a specific type of disease. Suppose the standard therapy has a response rate of 60%. The investigators would like to compute the power when the significance level is 0.05, the treatment response rate is at least 70%, the cluster size is 30, the number of clusters per group is 25 for all groups, and the ICC is 0.02. They plan to use the two-sided unpooled Z-test to analyze the data. Since there are two tests, the significance level will be reset to 0.025.

The **Tests for Two Proportions in a Cluster-Randomized Design** procedure is set up as follows.

Design Tab

Solve For ....**Power**
Alternative Hypothesis ....**Two-Sided (H1: P1 - P2 ≠ 0)**
Test Type..**Z-Test (Unpooled)**
Alpha...**0.025**
K1 (Clusters in Group 1) **25**
M1 (Average Cluster Size)..............................**30**
K2 (Clusters in Group 2) **K1**
M2 (Average Cluster Size)..............................**M1**
Input Type...**Proportions**
P1 (Group 1 Proportion|H1) **0.7**
P2 (Group 2 Proportion)..................................**0.6**
ICC (Intracluster Correlation) **0.02**

This set of options generates the following report.

Numeric Results for Testing Two Proportions (Cluster-Randomized)

Solve For: Power
Groups: 1 = Treatment, 2 = Control
Test Statistic: Z-Test (Unpooled)
Hypotheses: H0: P1 - P2 = 0   vs.   H1: P1 - P2 ≠ 0

	Number of Clusters			Cluster Size		Sample Size			Proportion		Difference	Intracluster Correlation	
Power	K1	K2	K	M1	M2	N1	N2	N	P1	P2	D1	ICC	Alpha
0.8429	25	25	50	30	30	750	750	1500	0.7	0.6	0.1	0.02	0.025

The power is computed to be 0.8429.

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**
Alternative Hypothesis	**Two-Sided (H1: Pi ≠ Pc)**
Test Type...	**Z-Test (Unpooled)**
Overall Alpha	**0.05**
Bonferroni Adjustment	**Standard Bonferroni**
Group Allocation	**Equal (Kc = K1 = K2 = ...)**
Ki (Group Number of Clusters)	**25**
M (Items Per Cluster).....................................	**30**
Control Proportion..	**0.6**
Control Items Per Cluster...............................	**M**
Set A Number of Groups.................................	**1**
Set A Proportion	**0.7**
Set A Items Per Cluster	**M**
Set B Number of Groups.................................	**1**
Set B Proportion	**0.7**
Set B Items Per Cluster	**M**
Set C Number of Groups	**0**
Set D Number of Groups	**0**
More...	**Unchecked**
ρ (Intracluster Correlation)	**0.02**

Output

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

Numeric Results

Solve For:	Power
Test Type:	Unpooled Z-Test
Hypotheses:	H0: δ = 0 vs. H1: δ ≠ 0
Number of Groups:	3
Bonferroni Adjustment:	Standard Bonferroni (Divisor = 2)

								Alpha	
Comparison	Power	Number of Clusters Ki	Items Per Cluster Mi	Sample Size Ni	Proportion Pi	Difference δi	ICC ρ	Overall	Bonferroni-Adjusted
Control		25	30	750	0.6		0.02		
vs A	0.8429	25	30	750	0.7	0.1	0.02	0.05	0.025
vs B	0.8429	25	30	750	0.7	0.1	0.02	0.05	0.025
Total		75		2250					

As you can see, the power is 0.8429 for both treatment groups which matches the power found in the validation run above. The procedure is validated.