

Chapter 482

Tests for Two Means in a Cluster-Randomized Design

Introduction

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.

Technical Details

Our formulation comes from Campbell and Walters (2014) and Ahn, Heo, and Zhang (2015). Denote an observation by Y_{ijk} where $i = 1, 2$ gives the group, $j = 1, 2, \dots, K_i$ gives the cluster within group i , and $k = 1, 2, \dots, m_{ij}$ 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 intraclass correlation coefficient (ICC) which is $\sigma_{Between}^2 / (\sigma_{Between}^2 + \sigma_{Within}^2)$. This correlation is the simple 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.

Tests for Two Means in a Cluster-Randomized Design

Assume that $\delta = \mu_1 - \mu_2$ is to be tested using a t-test (small sample) or z-test (large sample). The statistical hypotheses are $H_0: \delta = 0$ vs. $H_a: \delta \neq 0$. The test statistic

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

has an approximate t distribution with degrees of freedom $DF = K_1M_1 + K_2M_2 - 2$ for a *subject-level* analysis or $K_1 + K_2 - 2$ for a *cluster-level* analysis.

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

$$X_1 = t_{\frac{\alpha}{2}, DF}$$

$$X_2 = t_{1-\frac{\alpha}{2}, DF}$$

The power can be found from the following to probabilities

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

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

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

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.

Procedure Options

This section describes the options that are specific to this procedure. These are located on the Design tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Design Tab

The Design tab contains most of the parameters and options that you will be concerned with.

Solve For

Solve For

This option specifies the parameter to be solved for from the other parameters. The parameters that may be selected are δ , *Power*, *KI*, and *MI*.

Under most situations, you will select either *Power* to calculate power or *KI* to calculate the number of clusters. Occasionally, you may want to fix the number of clusters and find the necessary cluster size.

Note that the value selected here always appears as the vertical axis on the charts.

The program is set up to calculate power directly. To find appropriate values of the other parameters, a binary search is made using an iterative procedure until an appropriate value is found.

Tests for Two Means in a Cluster-Randomized Design

Test

Alternative Hypothesis

Specify whether the test is one-sided or two-sided. A two-sided hypothesis states that the means are not equal without specifying which is greater. If you do not have any special reason to do otherwise, you should use the two-sided option.

The one-sided options $\delta < 0$ and $\delta > 0$ select one-tailed tests. When you choose one of these, you should be careful that you enter a δ value that is consistent with this alternative hypothesis direction. For example, if you select $\delta < 0$ here, the values of δ that you enter below should be negative.

Test Statistic

Specify which t-test statistic you are going to use: a t-test based on the number of subjects or a t-test in which the cluster means are treated as subjects.

- **T-Test Based on Number of Subjects**

This uses the methodology shown in the recent books by Campbell and Walters (2014) and Ahn, Heo, and Zhang (2015). In this case, power is based on a t-test in which the variance is inflated to adjust for the clustering and the degrees of freedom is based on to the number of subjects.

- **T-Test Based on Number of Clusters**

This uses the original methodology of Donner and Klar (1996). In this case, power is based on a t-test in which the variance is also inflated to adjust for the clustering, but the degrees of freedom are based on the number of clusters. Donner and Klar ignored the impact of COV, so, if you want to match their results, you should set the COV to zero.

Power and Alpha

Power

This option specifies one or more values for power. Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered or a range of values such as *0.8 to 0.95 by 0.05* may be entered.

If your only interest is in determining the appropriate sample size for a confidence interval, set power to 0.5.

Alpha

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected.

Values must be between zero and one. Usually, the value of 0.05 is used for alpha and this has become a standard. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as *0.01 0.05 0.10* or *0.01 to 0.10 by 0.01*.

Tests for Two Means in a Cluster-Randomized Design

Sample Size – Number of Clusters & Cluster Size**Group 1 (Treatment)****K1 (Number of Clusters)**

Enter a value (or range of values) for the number of clusters in the treatment group. You may enter a range of values such as *10 to 20 by 2*. The sample size for this group is equal to the number of clusters times the average cluster size.

M1 (Average Cluster Size)

This is the average number of subjects per cluster in group one. This value must be a positive number that is at least one. You can use a list of values such as *100 150 200*.

Group 2 (Control)**K2 (Number of Clusters)**

This is the number of clusters in group two. This value must be a positive number. The sample size for this group is equal to the number of clusters times the number of subjects per cluster.

If you simply want a multiple of the value for group one, you would enter the multiple followed by *K1*, with no blanks. If you want to use *K1* directly, you do not have to premultiply by *1*. For example, all of the following are valid entries: *10 K1 2K1 0.5K1 K1*.

You can use a list of values such as *10 20 30* or *K1 2K1 3K1*.

M2 (Average Cluster Size)

This is the average number of subjects per cluster in group two. This value must be at least one.

If you simply want a multiple of the value for group one, you would enter the multiple followed by *M1*, with no blanks. If you want to use *M1* directly, you do not have to premultiply by *1*. For example, all of the following are valid entries: *10M1 2M1 0.5M1 M1*.

You can use a list of values such as *10 20 30* or *M1 2M1 3M1*.

Coefficient of Variation of Cluster Sizes**COV of Cluster Sizes**

Enter the *coefficient of variation* of the cluster sizes (number of subjects). This value must be zero or a positive number. The COV of X is defined as the standard deviation of X divided by the mean of X.

Campbell and Walters (2014) page 71 give guidance on the possible values of COV. They indicate that as the average cluster size increases, COV tends toward 0.65. They say that typical values of COV range from 0.4 to 0.9.

You can use a list of values such as *0.4 0.6 0.8*.

Standard Deviation

The standard deviation, calculated by the sample formula (divide by $n-1$), is a measure of the variability. When no other information is available, Campbell and Walters (2014) page 71 suggest using $(\text{Maximum Cluster Size} - \text{Minimum Cluster Size}) / 4$.

All Cluster Sizes Equal

When all cluster sizes are equal, the coefficient of variation is zero.

Tests for Two Means in a Cluster-Randomized Design

Effect Size

δ (Mean Difference = $\mu_1 - \mu_2$)

Enter a value for the difference between the means of groups 1 and 2 at which the design is powered. That is, the power is the probability of detecting a difference of at least this amount. This value does not represent the true difference between the means. Rather, it is the value at which you want the sample size computed.

δ can be any non-zero value (positive or negative), but it should be consistent with the choice of the alternative hypothesis. Also, it should have the same scale as the standard deviation.

You can enter a single value such as *10* or a series of values such as *10 20 30* or *5 to 50 by 5*.

When a series of values is entered, PASS will generate a separate calculation result for each value of the series.

σ (Standard Deviation)

Enter the subject-to-subject standard deviation. This standard deviation applies for both groups.

Note that σ must be a positive number. You can enter a single value such as *5* or a series of values such as *1 3 5 7 9* or *1 to 9 by 2*.

Press the small ‘ σ ’ button to the right to obtain calculation options for estimating the standard deviation.

ρ (Intracluster Correlation, ICC)

This is the value of the intracluster correlation coefficient. It may be interpreted as the correlation between any two observations in the same cluster. It may also be thought of as the proportion of the variation in response that can be accounted for by the between-cluster variation.

Possible values are from 0 to just below 1. Typical values are between 0.0001 and 0.05.

You may enter a single value or a list of values.

Tests for Two Means in a Cluster-Randomized Design

Example 1 – Calculating Power

Suppose that a cluster randomized study is to be conducted in which $\delta = 1$; $\sigma = 2$; $\rho = 0.01$; $M1$ and $M2 = 5$ or 10 ; $COV = 0.65$; $alpha = 0.05$; and $K1$ and $K2 = 5$ to 20 by 5 . Power is to be calculated for a two-sided test.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Means in a Cluster-Randomized Design** procedure window by expanding **Means**, then **Two Independent Means**, then clicking on **Cluster-Randomized**, and then clicking on **Tests for Two Means in a Cluster-Randomized Design**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Power
Ha (Alternative Hypothesis)	Two-Sided (Ha: $\delta \neq 0$)
Test Statistic	T-Test Based on Number of Subjects
Alpha.....	0.05
K1 (Number of Clusters).....	5 10 15 20
M1 (Average Cluster Size)	5 10
K2 (Number of Clusters).....	K1
M2 (Average Cluster Size)	M1
COV of Cluster Sizes.....	0.65
δ (Mean Difference = $\mu_1 - \mu_2$)	1
σ (Standard Deviation)	2
ρ (Intracluster Correlation, ICC).....	0.01

Annotated Output

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

Numeric Results

Numeric Results for a Test of Mean Difference

Test Statistic: T-Test with DF based on number of subjects

Hypotheses: $H_0: \delta = 0$ vs. $H_a: \delta \neq 0$

	Subj Cnt Gr 1	Subj Cnt Gr 2	Clus Cnt Gr 1	Clus Cnt Gr 2	Clus Size Gr 1	Clus Size Gr 2	COV Clus Sizes COV	Diff $\mu_1 - \mu_2$ δ	Std Dev σ	ICC ρ	Alpha
Power	N1	N2	K1	K2	M1	M2					
0.3908	25	25	5	5	5	5	0.650	1.00	2.00	0.010	0.050
0.6439	50	50	5	5	10	10	0.650	1.00	2.00	0.010	0.050
0.6714	50	50	10	10	5	5	0.650	1.00	2.00	0.010	0.050
0.9115	100	100	10	10	10	10	0.650	1.00	2.00	0.010	0.050
0.8399	75	75	15	15	5	5	0.650	1.00	2.00	0.010	0.050
0.9822	150	150	15	15	10	10	0.650	1.00	2.00	0.010	0.050
0.9274	100	100	20	20	5	5	0.650	1.00	2.00	0.010	0.050
0.9969	200	200	20	20	10	10	0.650	1.00	2.00	0.010	0.050

Tests for Two Means in a Cluster-Randomized Design

References

- 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.
- Ahn, C., Heo, M., and Zhang, S. 2015. Sample Size Calculations for Clustered and Longitudinal Outcomes in Clinical Research. CRC Press. New York.
- 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.
- Chow, S.C., Shao, J., Wang, H., and Lohknygina, Y. 2018. Sample Size Calculations in Clinical Research, Third Edition. Taylor & Francis/CRC. Boca Raton, Florida.

Report Definitions

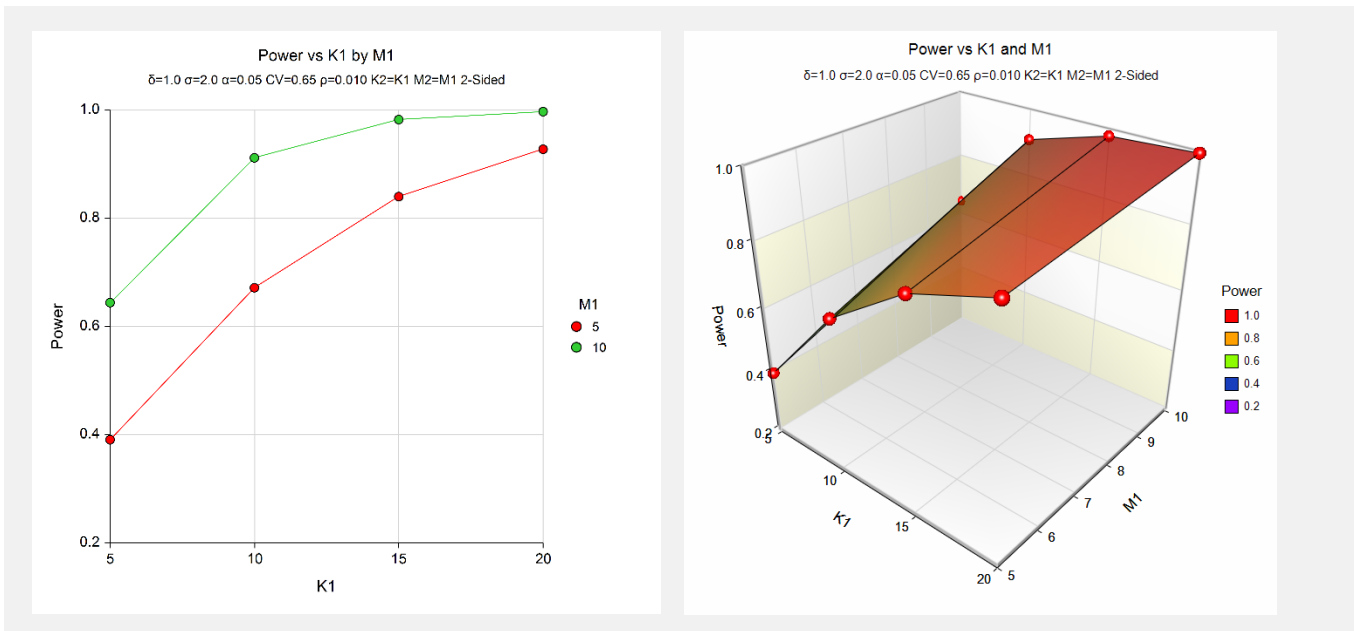
- Power is the probability of rejecting a false null hypothesis. It should be close to one.
- N_1 and N_2 are the number of subjects in groups 1 and 2, respectively.
- K_1 and K_2 are the number of clusters in groups 1 and 2, respectively.
- M_1 and M_2 are the average number of items (subjects) per cluster in groups 1 and 2, respectively.
- COV is the coefficient of variation of the cluster sizes.
- δ is the mean difference ($\mu_1 - \mu_2$) in the response at which the power is calculated.
- σ is the standard deviation of the subject responses.
- ρ (ICC) is the intracluster correlation.
- Alpha is the probability of rejecting a true null hypothesis, that is, rejecting when the means are actually equal.

Summary Statements

Sample sizes of 25 in group one and 25 in group two, which were obtained by sampling 5 clusters with an average of 5 subjects each in group one and 5 clusters with an average of 5 subjects each in group two, achieve 39% power to detect a difference between the group means of at least 1. The standard deviation of subjects is 2.00. The intracluster correlation coefficient is 0.010. The coefficient of variation of cluster sizes is 0.650. A two-sided t-test was used with a significance level of 0.050. This test used degrees of freedom based on the number of subjects.

This report shows the power for each of the scenarios.

Plots Section



These plots show the power versus the cluster size for the two alpha values.

Tests for Two Means in a Cluster-Randomized Design

Example 2 – Validation using Donner and Klar (1996)

Donner and Klar (1996) page 436 provide a table in which several power values are calculated for a two-sided test. When α is 0.05, δ is 0.2, ρ is 0.001, σ is 1.0, and $K1 = K2 = 3$, they calculate a power of 0.43 for an $M1=M2$ of 100, 0.79 for an $M1=M2$ of 300, and 0.91 for an $M1=M2$ of 500. $COV = 0$.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Means in a Cluster-Randomized Design** procedure window by expanding **Means**, then **Two Independent Means**, then clicking on **Cluster-Randomized**, and then clicking on **Tests for Two Means in a Cluster-Randomized Design**. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Power
Ha (Alternative Hypothesis)	Two-Sided (Ha: $\delta \neq 0$)
Test Statistic	T-Test Based on Number of Clusters
Alpha	0.05
K1 (Number of Clusters)	3
M1 (Average Cluster Size)	100 300 500
K2 (Number of Clusters)	K1
M2 (Average Cluster Size)	M1
COV of Cluster Sizes	0
δ (Mean Difference = $\mu_1 - \mu_2$)	0.2
σ (Standard Deviation)	1
ρ (Intracluster Correlation, ICC)	0.001

Output

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

Numeric Results

Numeric Results for a Test of Mean Difference											
Test Statistic: T-Test with DF based on number of clusters											
Hypotheses: $H_0: \delta = 0$ vs. $H_a: \delta \neq 0$											
	Subj Cnt	Subj Cnt	Clus Cnt	Clus Cnt	Clus Size	Clus Size	COV Clus Sizes	Diff $\mu_1 - \mu_2$	Std Dev	ICC ρ	Alpha
Power	N1	N2	K1	K2	M1	M2	COV	δ	σ	ρ	Alpha
0.4301	300	300	3	3	100	100	0.000	0.20	1.00	0.001	0.050
0.7924	900	900	3	3	300	300	0.000	0.20	1.00	0.001	0.050
0.9091	1500	1500	3	3	500	500	0.000	0.20	1.00	0.001	0.050

PASS calculates the same power values as Donner and Klar (1996).

Example 3 – Validation using Campbell and Walters (2014)

Campbell and Walters (2014) page 71 give an example of the effect of variable cluster size when a subject-level analysis is run. In this example, they set alpha to 0.05, and ρ to 0.05. They don't specifically give values for the mean and standard deviation. By trial and error, we found that for a power of 0.9, δ of 0.3247 and σ of 1 gave results that matched theirs for a two-sided test. They indicate that when COV = 0, K1 = 29. When COV = 0.725, K1 = 33.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Means in a Cluster-Randomized Design** procedure window by expanding **Means**, then **Two Independent Means**, then clicking on **Cluster-Randomized**, and then clicking on **Tests for Two Means in a Cluster-Randomized Design**. You may then make the appropriate entries as listed below, or open **Example 3** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	K1 (Number of Clusters)
Ha (Alternative Hypothesis)	Two-Sided (Ha: $\delta \neq 0$)
Test Statistic	T-Test Based on Number of Subjects
Power	0.90
Alpha	0.05
M1 (Average Cluster Size)	10
K2 (Number of Clusters)	K1
M2 (Average Cluster Size)	M1
COV of Cluster Sizes	0 0.725
δ (Mean Difference = $\mu_1 - \mu_2$)	0.3247
σ (Standard Deviation)	1
ρ (Intraclass Correlation, ICC)	0.05

Output

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

Numeric Results

Numeric Results for a Test of Mean Difference											
Test Statistic: T-Test with DF based on number of subjects											
Hypotheses: H0: $\delta = 0$ vs. Ha: $\delta \neq 0$											
	Subj Cnt	Subj Cnt	Clus Cnt	Clus Cnt	Clus Size	Clus Size	COV Clus Sizes	Diff $\mu_1 - \mu_2$	Std Dev	ICC ρ	Alpha
Power	Gr 1 N1	Gr 2 N2	Gr 1 K1	Gr 2 K2	Gr 1 M1	Gr 2 M2	COV	δ	σ		
0.9000	290	290	29	29	10	10	0.000	0.32	1.00	0.050	0.050
0.9016	330	330	33	33	10	10	0.725	0.32	1.00	0.050	0.050

PASS calculates the same values of K1 as Campbell and Walters (2014).