Chapter 486

Equivalence Tests for Two Means in a Cluster-Randomized Design

Introduction

This procedure computes power and sample size for Schuirmann's (1987) two one-sided (TOST) test of equivalence when the data come from a cluster-randomized design in which the outcome is a continuous normal random variable. Only a brief introduction to the subject of equivalence testing will be given here. For a comprehensive discussion, refer to Chow and Liu (1999).

It should be noted that we could not find any published results about equivalence testing in 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.

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. 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.

The Statistical Hypotheses

PASS follows the *two one-sided tests* approach described by Schuirmann (1987) and Phillips (1990). Remember that when testing equivalence, the null and alternative hypotheses are defined as follows.

$$H_0: \delta \leq EL \text{ or } \delta \geq EU \text{ versus. } H_a: EL < \delta < EU$$

Rejecting H_0 in favor of H_a infers that the group means are equivalent.

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*.

Technical Details

Our formulation is a combination of equivalence formulas of Chow and Liu (199) 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 $\sigma^2_{Between} + \sigma^2_{Within}$, where $\sigma^2_{Between}$ is the variation between clusters and σ^2_{Within} is the variation within clusters. Also, let ρ denote the intracluster correlation coefficient (ICC) which is $\sigma^2_{Between}/\left(\sigma^2_{Between} + \sigma^2_{Within}\right)$. This correlation is the simply 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 two modified two-sample t-tests. The test statistics are

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

and

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

We assume these statistics have approximate t distributions 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.

Define the noncentrality parameters as

$$\Delta_{\mathrm{L}} = (\delta - EL)/\sigma_d$$

and

$$\Delta_{\mathrm{U}} = (\delta - EU)/\sigma_d$$

where

$$\sigma_d = \sqrt{V_1 + V_2}.$$

The power of this test procedure is given by

$$\text{Power} = \Pr\left(T_L \geq t_{1-\alpha,DF} \text{ and } T_U \leq -t_{1-\alpha,DF}\right)$$

where $\mathrm{T_L}$ and $\mathrm{T_U}$ are distributed as the bivariate, noncentral t distribution with noncentrality parameters Δ_{L} and Δ_{U}.

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*, *K1*, and *M1*.

Under most situations, you will select either *Power* to calculate power or *K1* 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.

Test

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 0.025 is used for alpha of one-sided test and this has become a standard.

You may enter a range of values such as *0.025 0.05 0.10* or *0.01 to 0.10 by 0.01*.

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 (Maximum Cluster Size - Minimum Cluster Size) / 4.

All Cluster Sizes Equal

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

Effect Size

EU (Upper Equivalence Limit)

Enter one or more values for the upper limit of equivalence. If the actual difference is between EL and EU, the new treatment is said to be equivalent to the standard. Enter values in the range $EU > 0$ and $EL < \delta < EU$.

EL (Lower Equivalence Limit)

Enter one or more values for the lower limit of equivalence. If the actual difference is between EL and EU, the new treatment is said to be equivalent to the standard. Enter values in the range $EL < 0$ and $EL < \delta < EU$.

If you want symmetric limits around zero, enter "-Upper Limit" here. The program will set EL = -|EU|.

δ (Mean Difference = μ1 – μ2)

This is the actual difference between the treatment group mean and the reference group mean.

Usually, this value is set to zero for equivalence tests. However, you may use any value between EL and EU.

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.

Example 1 – Calculating Power

Suppose an equivalence test is to be conducted on data obtained from a cluster-randomized design in which *EU*= 1; EL = -Upper Limit; $\delta = 0$; $\sigma = 2$; $\rho = 0.02$; *M1* and *M2* = 5, 10; *COV* = 0.65; *alpha* = 0.025; and *K1* and *K2* = 5, 10, 15, and 20.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Equivalence Test 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 **Equivalence Test 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**.

Option	Value
Design Tab	
Solve For	**Power**
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**
EU (Upper Equivalence Limit)	**1**
EL (Lower Equivalence Limit)	**-Upper Limit**
δ (Mean Difference = μ1 - μ2)	**0**
σ (Standard Deviation)	**2**
ρ (Intracluster Correlation, ICC)	**0.02**

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: H0: δ ≤ EL or δ ≥ EU vs. Ha: EL < δ < EU

Power	Subj Cnt Gr 1 N1	Subj Cnt Gr 2 N2	Clus Cnt Gr 1 K1	Clus Cnt Gr 2 K2	Clus Size Gr 1 M1	Clus Size Gr 2 M2	COV Clus Sizes COV	Diff μ1-μ2 δ	Lower Equiv Limit EL	Upper Equiv Limit EU	Std Dev σ	ICC ρ	Alpha
0.0547	25	25	5	5	5	5	0.650	0.0	-1.0	1.0	2.0	0.020	0.050
0.4324	50	50	5	5	10	10	0.650	0.0	-1.0	1.0	2.0	0.020	0.050
0.5169	50	50	10	10	5	5	0.650	0.0	-1.0	1.0	2.0	0.020	0.050
0.8666	100	100	10	10	10	10	0.650	0.0	-1.0	1.0	2.0	0.020	0.050
0.7833	75	75	15	15	5	5	0.650	0.0	-1.0	1.0	2.0	0.020	0.050
0.9730	150	150	15	15	10	10	0.650	0.0	-1.0	1.0	2.0	0.020	0.050
0.9080	100	100	20	20	5	5	0.650	0.0	-1.0	1.0	2.0	0.020	0.050
0.9951	200	200	20	20	10	10	0.650	0.0	-1.0	1.0	2.0	0.020	0.050

References

Ahn, C., Heo, M., and Zhang, S. 2015. Sample Size Calculations for Clustered and Longitudinal Outcomes in Clinical Research. CRC Press. New York.
Blackwelder, W.C. 1998. 'Equivalence Trials.' In Encyclopedia of Biostatistics, John Wiley and Sons. New York. Volume 2, 1367-1372.
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, Third Edition. Taylor & Francis/CRC. Boca Raton, Florida.
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.
Phillips, Kem F. 1990. 'Power of the Two One-Sided Tests Procedure in Bioequivalence', Journal of Pharmacokinetics and Biopharmaceutics, Volume 18, No. 2, pages 137-144.
Schuirmann, Donald. 1987. 'A Comparison of the Two One-Sided Tests Procedure and the Power Approach for Assessing the Equivalence of Average Bioavailability', Journal of Pharmacokinetics and Biopharmaceutics, Volume 15, Number 6, pages 657-680.

Report Definitions

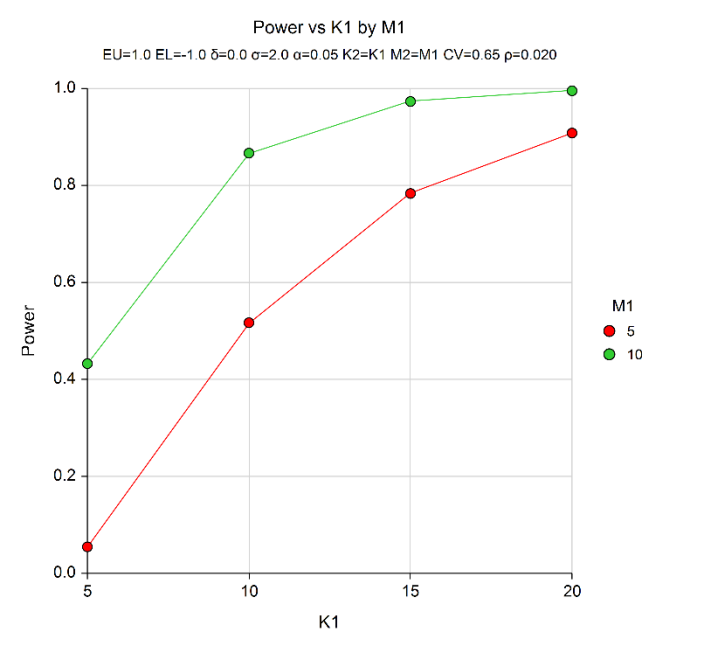
Power is the probability of rejecting a false null hypothesis. It should be close to one.
N1 and N2 are the number of subjects in groups 1 and 2, respectively.
K1 and K2 are the number of clusters in groups 1 and 2, respectively.
M1 and M2 are the average number of subjects per cluster in groups 1 and 2, respectively.
EL is the lower equivalence limit. It is the lower bound on the interval of equivalence. Differences outside EL to EU are not equivalent.
EU is the upper equivalence limit. It is the upper bound on the interval of equivalence. Differences outside EL to EU are not equivalent.
COV is the coefficient of variation of the cluster sizes.
δ is the mean difference (μ1 - μ2) 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.

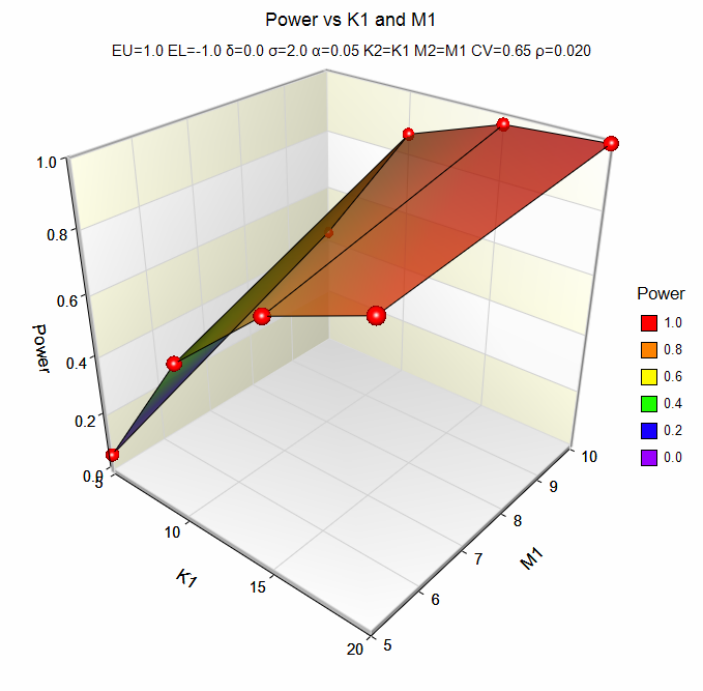
Summary Statements

An equivalence test of means using two one-sided tests on data from a cluster-randomized design is to be conducted. 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 5% power to detect equivalence. The interval of equivalence is between -1.0 and 1.0. The true difference between the means is assumed to be 0.0. The standard deviation of subjects is 2.0. The intracluster correlation coefficient is 0.020. The coefficient of variation of cluster sizes is 0.650. The significance level of the testing procedure is set at 0.050.

This report shows the power for each of the scenarios.

Plots Section





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 the validation example of the *Two-Sample T-Tests for Equivalence Assuming Equal Variance* procedure. The results should be identical when M1 = M2 = 1. Use the following scenario: find K1 when δ = -2, EU = 5, EL=-5, σ = 8, alpha = 0.05, and power = 0.80. That procedure obtained a validated value of 89 for K1 and K2.

Because M1 = 1, the values of ρ and COV are set to 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 **Equivalence Test 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 **Equivalence Test 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**.

Option	Value
Design Tab	
Solve For	**K1 (Number of Clusters)**
Test Statistic	**T-Test Based on Number of Subjects**
Power	**0.8**
Alpha	**0.05**
M1 (Average Cluster Size)	**1**
K2 (Number of Clusters)	**K1**
M2 (Average Cluster Size)	**M1**
COV of Cluster Sizes	**0**
EU (Upper Equivalence Limit)	**5**
EL (Lower Equivalence Limit)	**-Upper Limit**
δ (Mean Difference = μ1 - μ2)	**-2**
σ (Standard Deviation)	**8**
ρ (Intracluster Correlation, ICC)	**0**

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: δ ≤ EL or δ ≥ EU vs. Ha: EL < δ < EU

Power	Subj Cnt Gr 1 N1	Subj Cnt Gr 2 N2	Clus Cnt Gr 1 K1	Clus Cnt Gr 2 K2	Clus Size Gr 1 M1	Clus Size Gr 2 M2	COV Clus Sizes COV	Diff μ1-μ2 δ	Lower Equiv Limit EL	Upper Equiv Limit EU	Std Dev σ	ICC ρ	Alpha
0.8015	89	89	89	89	1	1	0.000	-2.0	-5.0	5.0	8.0	0.000	0.050

This procedure also calculates K1 to be 89.