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.

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

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 simply the 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

$$\Delta_L = (\delta - EL)/\sigma_d$$

and

$$\Delta_U = (\delta - EU)/\sigma_d$$

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 T_L and T_U are distributed as the bivariate, noncentral *t* distribution with noncentrality parameters Δ_L and Δ_U.

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; δ = 0; σ = 2; ρ = 0.02; *M1* and *M2* = 5, 10; *COV* = 0.65; *alpha* = 0.025; and *K1* and *K2* = 5, 10, 15, and 20.

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 .... **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**

Output

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

Numeric Reports

Numeric Results for a Test of Mean Difference

Solve For: Power
Groups: 1 = Treatment, 2 = Control
Test Statistic: T-Test with DF based on number of subjects
Hypotheses: H0: δ ≤ EL or δ ≥ EU vs. H1: EL < δ < EU

	Number of Clusters			Cluster Size			Sample Size		Mean Difference	Equivalence Limits		Standard Deviation	ICC	
Power	K1	K2	K	M1	M2	COV	N1	N2	δ	Lower EL	Upper EU	σ	ρ	Alpha
0.0547	5	5	10	5	5	0.65	25	25	0	-1	1	2	0.02	0.05
0.4324	5	5	10	10	10	0.65	50	50	0	-1	1	2	0.02	0.05
0.5169	10	10	20	5	5	0.65	50	50	0	-1	1	2	0.02	0.05
0.8666	10	10	20	10	10	0.65	100	100	0	-1	1	2	0.02	0.05
0.7833	15	15	30	5	5	0.65	75	75	0	-1	1	2	0.02	0.05
0.9730	15	15	30	10	10	0.65	150	150	0	-1	1	2	0.02	0.05
0.9080	20	20	40	5	5	0.65	100	100	0	-1	1	2	0.02	0.05
0.9951	20	20	40	10	10	0.65	200	200	0	-1	1	2	0.02	0.05

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 and 2, and their total.
M1 and M2	The average number of items (subjects) per cluster in groups 1 and 2, respectively.
COV	The coefficient of variation of the cluster sizes.
N1 and N2	The number of subjects in groups 1 and 2, respectively.
δ	The mean difference in the response at which the power is calculated. δ = μ1 - μ2.
EL and EU	The lower and upper equivalence limits. Means whose difference is outside these limits are not "equivalent".
σ	The standard deviation of the subject responses.
ρ	The intra-cluster correlation (ICC). The correlation between a pair of subjects within a 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) mean (μ1) is equivalent to the Group 2 (control) mean (μ2), with mean difference equivalence limits of -1 and 1 (H0: δ ≤ -1 or δ ≥ 1 versus H1: -1 < δ < 1, δ = μ1 - μ2). The comparison will be made using two one-sided t-tests with the degrees of freedom based on the total number of subjects (see Campbell and Walters, 2014, and Ahn, Heo, and Zhang, 2015), with an overall Type I error rate (α) of 0.05. The common subject-to-subject standard deviation for both groups is assumed to be 2, the intracluster correlation coefficient is assumed to be 0.02, and the coefficient of variation of cluster sizes is assumed to be 0.65. To detect a mean difference (μ1 - μ2) of 0, with 5 clusters of 5 subjects per cluster in Group 1 (totaling 25 subjects) and 5 clusters of 5 subjects per cluster in Group 2 (totaling 25 subjects), the power is 0.0547.

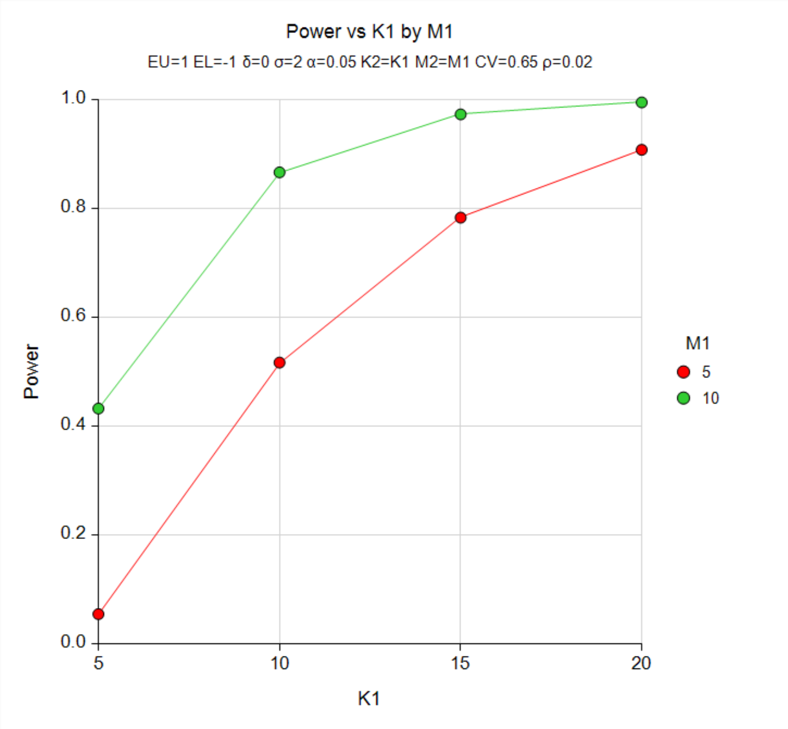
References

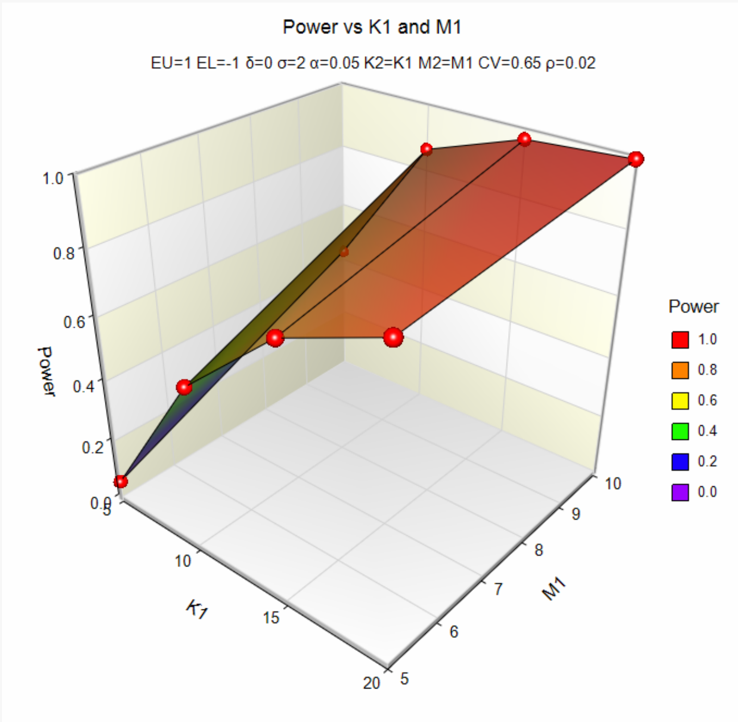
Ahn, C., Heo, M., and Zhang, S. 2015. Sample Size Calculations for Clustered and Longitudinal Outcomes in Clinical Research. CRC Press. New York.
Blackwelder, W.C. 1998. 'Equivalence Trials.' In Encyclopedia of Biostatistics, John Wiley and Sons. New York. Volume 2, 1367-1372.
Campbell, M.J. and Walters, 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, 3rd Edition. Chapman & Hall/CRC. Boca Raton, FL. Pages 86-88.
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.

This report shows the power for each of the scenarios.

Plots Section

Plots





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

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 ....**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 for a Test of Mean Difference

Solve For: K1 (Number of Clusters)
Groups: 1 = Treatment, 2 = Control
Test Statistic: T-Test with DF based on number of subjects
Hypotheses: H0: δ ≤ EL or δ ≥ EU vs. H1: EL < δ < EU

	Number of Clusters			Cluster Size			Sample Size			Equivalence Limits		Standard		
Power	K1	K2	K	M1	M2	COV	N1	N2	Mean Difference δ	Lower EL	Upper EU	Deviation σ	ICC ρ	Alpha
0.8015	89	89	178	1	1	0	89	89	-2	-5	5	8	0	0.05

This procedure also calculates K1 to be 89.