

## Chapter 154

# Non-Inferiority Tests for Two Means in a Cluster-Randomized Design with Clustering in Only One Arm

---

## Introduction

This procedure computes power and sample size for a *non-inferiority* test in cluster-randomized designs in which the outcome is a continuous normal random variable and clustering occurs in only one arm.

Cluster-randomized designs are those in which whole clusters of subjects (classes, hospitals, communities, etc.) are randomized into a treatment group or a control group. In this case, the means of the two arms (groups), where the second arm is made up of individuals and the first arm is made up of  $K_i$  clusters of  $M_{ij}$  individuals each, are to be tested using a modified z test.

For example, one arm might receive individual intervention (such as medicine) while the other receives this medicine plus some type of group therapy session. The fact that they attend this therapy session implies that they are “clustered”. Thus the clustering occurs in only one arm.

It should be noted that we could not find any published results about non-inferiority testing with cluster-randomized designs with clustering in only one group. 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.

---

## The Statistical Hypotheses

Non-inferiority tests are examples of directional (one-sided) tests. This program module provides the input and output in a format that is convenient for these types of tests. This section will review the specifics of non-inferiority testing.

Remember that in the usual z-test setting, the null ( $H_0$ ) and alternative ( $H_1$ ) hypotheses for one-sided tests are defined as follows, assuming that  $\delta = \mu_T - \mu_C$  is to be tested. For an *upper-tailed test* we have

$$H_0: \delta \leq 0 \quad \text{versus} \quad H_1: \delta > 0$$

Rejecting this test implies that the mean difference is larger than the value  $\delta$ . 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  $D$ .

Following is an example of a *lower-tailed test*.

$$H_0: \delta \geq 0 \quad \text{versus} \quad H_1: \delta < 0$$

## Non-Inferiority Tests for Two Means in a Cluster-Randomized Design with Clustering in Only One Arm

*Non-inferiority* tests are special cases of the above directional tests. It will be convenient to adopt the following specialized notation for the discussion of these tests.

<b>Parameter</b>	<b>PASS Input/Output</b>	<b>Interpretation</b>
$\mu_T$	Not used	<i>Mean</i> of the treatment population. This population is assumed to consist of those who have received the new treatment.
$\mu_C$	Not used	<i>Mean</i> of the control population. This population is assumed to consist of those who have received the control (reference) treatment.
$\varepsilon$	NIM	<i>Margin of non-inferiority</i> . This is a tolerance value that defines the magnitude of the amount that is not of practical importance. This may be thought of as the largest change from the baseline that is considered to be trivial. The absolute value is shown to emphasize that this is a magnitude. The sign of the value will be determined by the specific design that is being used.
$\delta$	$\delta$	<i>True difference</i> . This is the value of $\mu_T - \mu_C$ , the difference between the means.

Note that the actual values of  $\mu_T$  and  $\mu_C$  are not needed. Only their difference is needed for power and sample size calculations.

---

## Non-Inferiority Tests

A *non-inferiority test* tests that the treatment mean is not worse than the control mean by more than the non-inferiority margin. The actual direction of the hypothesis depends on the response variable being studied.

**Note that the following definitions work equally well whether the clustered arm is assigned to the treatment group or the control group.**

### Case 1: High Values Good, Non-Inferiority Test

In this case, higher values are better. The hypotheses are arranged so that rejecting the null hypothesis implies that the treatment mean is no less than a small amount below the reference mean,  $\varepsilon$ . The value of  $\delta$  is often set to zero. The following are equivalent sets of hypotheses.

$$H_0: \delta \leq -\varepsilon \quad \text{versus} \quad H_1: \delta > -\varepsilon, \quad \varepsilon > 0$$

$$H_0: \mu_T - \mu_C \leq -\varepsilon \quad \text{versus} \quad H_1: \mu_T - \mu_C > -\varepsilon, \quad \varepsilon > 0$$

$$H_0: \mu_T \leq \mu_C - \varepsilon \quad \text{versus} \quad H_1: \mu_T > \mu_C - \varepsilon, \quad \varepsilon > 0$$

## Case 2: High Values Bad, Non-Inferiority Test

In this case, lower values are better. The hypotheses are arranged so that rejecting the null hypothesis implies that the treatment mean is no more than a small amount above the reference mean,  $\varepsilon$ . The value of  $\delta$  is often set to zero. The following are equivalent sets of hypotheses.

$$\begin{aligned} H_0: \delta &\geq \varepsilon && \text{versus} && H_1: \delta < \varepsilon, \quad \varepsilon > 0 \\ H_0: \mu_T - \mu_C &\geq \varepsilon && \text{versus} && H_1: \mu_T - \mu_C < \varepsilon, \quad \varepsilon > 0 \\ H_0: \mu_T &\geq \mu_C + \varepsilon && \text{versus} && H_1: \mu_T < \mu_C + \varepsilon, \quad \varepsilon > 0 \end{aligned}$$

---

## Technical Details

Our formulation comes from Moerbeek and Wong (2008). They combine two mixed models: one for the arm made up of individuals and a second, more, complicated model for the arm which accounts for the clustering.

The model for the first arm (group 2) is

$$y_i = \beta_0 + x_i' \beta + r_i$$

where  $y_i$  is a continuous outcome,  $x_i$  is a vector of covariates for subject  $i$ ,  $\beta$  is a vector of effects of these covariates, and  $r_i$  is a random error term with expectation zero and variance  $\sigma_r^2$ .

The mixed model for the second (clustered) arm (group 1) is

$$y_{ij} = \beta_0 + \delta + x_{ij}' \beta + u_i + e_{ij}$$

where  $y_{ij}$  is a continuous outcome for subject  $i = 1, \dots, m_j$  within cluster  $j = 1, \dots, k$ ,  $\delta$  is the treatment effect to be tested,  $u_i$  is a random error term for clusters with expectation zero and variance  $\sigma_u^2$ , and  $e_{ij}$  is a random error term for subjects within a cluster with expectation zero and variance  $\sigma_e^2$ .

Note that the variance of  $y_{ij}$  is  $\sigma_u^2 + \sigma_e^2$ . The intra-class correlation (ICC) is

$$\rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}$$

These models can be combined into a single mixed model using

$$y_{ij} = \beta_0 + I_{ij} \delta + x_{ij}' \beta + I_{ij} u_i + e_{ij}$$

where  $I_{ij}$  is an indicator variable which is equal to one for subjects in the cluster arm and zero for subjects in the other, non-clustered, arm. Note that subjects in the non-clustered arm are assigned to a single "cluster"  $j = k + 1$ . Here  $k$  is the number of clusters. Note that  $\sigma_r^2 \neq \sigma_u^2 + \sigma_e^2$ , so the model is heteroscedastic (has unequal variances).

In this model, the treatment effect is measured by the difference in the two treatment means

$$\hat{\delta} = \bar{y}_T - \bar{y}_C$$

## Non-Inferiority Tests for Two Means in a Cluster-Randomized Design with Clustering in Only One Arm

The variance of this estimator is

$$\text{var}(\hat{\delta}) = \frac{\bar{m}\sigma_u^2 + \sigma_e^2}{\bar{m}k} + \frac{\sigma_r^2}{N_2} = \sigma_r^2 \left( \theta \frac{(\bar{m} - 1)\rho + 1}{\bar{m}k} + \frac{1}{N_2} \right)$$

where  $\bar{m}$  is the average cluster size,  $N_2$  is the number of subjects in the non-clustered arm, and

$$\theta = \frac{\sigma_u^2 + \sigma_e^2}{\sigma_r^2}.$$

Using the above, a large sample formula for computing the power of a one-sided test of the significance of the treatment effect at significance level  $\alpha$  is given by

$$\text{var}(\hat{\delta}) = \left( \frac{\delta - \varepsilon}{z_{1-\alpha} + z_{\text{power}}} \right)^2$$

For sample size calculation, we assume that the cluster sizes  $m_j$  are distributed with a mean cluster size of  $\bar{m}$  and a coefficient of variation cluster sizes of  $COV$ . The accuracy of the variance of the two group means,  $\bar{Y}_i$ , can be enhanced by replacing  $\bar{m}$  in the numerator of the first term of  $\text{var}(\hat{\delta})$  with  $\bar{m}(COV^2 + 1)$ . See Julious (2023) page 86.

Assume that  $\delta = \mu_T - \mu_C$  is to be tested using a z-test. The statistical hypotheses are  $H_0: \delta \leq \varepsilon$  vs.  $H_a: \delta > \varepsilon$ . The test statistic

$$z = \frac{\bar{y}_T - \bar{y}_C - \varepsilon}{\sqrt{\text{var}(\hat{\delta})}}$$

has an approximate normal distribution for a *subject-level* analysis.

Using the above, a large sample formula for computing the power of a one-sided test of the significance of the treatment effect at significance level  $\alpha$  can be derived from

$$\text{var}(\hat{\delta}) = \left( \frac{\delta - \varepsilon}{z_{1-\alpha} + z_{\text{power}}} \right)^2$$

Note that Julious (2023) uses a homoscedastic model that gives the same power values as these when  $\theta = 1$ . Candlish *et al.* (2018) recommend that the more complicated heteroscedastic model be used because of their findings in a large simulation study.

## Example 1 – Calculating Power

Suppose a non-inferiority, cluster-randomized study is to be conducted in which one arm (group 2) will receive an individual medical intervention while the other arm (group 1) receives this medicine plus a special group therapy session conducted by a trained therapist. These therapy sessions will be treated as clusters. Here, group 2 (the non-clustered subjects) is the 'control group' and group 1 (the clustered subjects) is the treatment group. The therapy sessions will involve 10 patients with a cluster size COV of 0.65. The point of the study is to determine if the treatment group is non-inferior to the control group.

At the time of the design of the trial, little is known about the variances, so the study will be powered in terms of standardized values. This is accomplished by the setting  $\sigma_2 = 1$ . The NIM is set to 0.5.

The parameter values are set as follows:  $\alpha = 0.025$ ,  $N2 = 70$ ,  $K1 = 4\ 6\ 8\ 10$ ,  $M1 = 10$ ,  $COV = 0.65$ ,  $NIM = 0.5$ ,  $\delta = 0.0$ ,  $\sigma_2 = 1$ ,  $\theta = 0.9, 1.0, 1.1$ , and  $\rho = 0.05$ . Power is to be calculated using the standard normal distribution.

### 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>
Assign Group 1 (Clusters) to be the.....	<b>Treatment Group</b>
Higher Means Are.....	<b>Better (H1: <math>\delta &gt; -NIM</math>, <math>\delta = \mu_T - \mu_C</math>)</b>
Alpha.....	<b>0.025</b>
Group Allocation .....	<b>Enter K1 and N2 individually</b>
K1 (Number of Clusters) .....	<b>4 6 8 10</b>
M1 (Average Cluster Size).....	<b>10</b>
COV (Coefficient of Variation of Cluster Sizes).....	<b>0.65</b>
N2 (Number of Subjects) .....	<b>70</b>
NIM (Non-Inferiority Margin) .....	<b>0.5</b>
$\delta$ (Mean Difference = $\mu_T - \mu_C$ ) .....	<b>0</b>
$\sigma_2$ (Standard Deviation in Group 2) .....	<b>1</b>
$\Theta$ (Group Variance Ratio = $\sigma_1^2 / \sigma_2^2$ ) .....	<b>0.9 1 1.1</b>
$\rho$ (Intracluster Correlation, ICC).....	<b>0.05</b>

## Output

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

### Numeric Reports

#### Numeric Results

Solve For: Power  
 Group 1 Assignment: Group 1 has Clusters and is the Treatment Group  
 Group 2 Assignment: Group 2 does not have Clusters and is the Control Group  
 Group Allocation: Enter K1 and N2 individually  
 Higher Means Are: Better  
 Hypotheses: H0:  $\delta \leq -NIM$  vs. H1:  $\delta > -NIM$

Power	Group 1 (Treatment) Clusters			Sample Size (Subjects)			Non-Inferiority Margin -NIM	Mean Difference $\delta$	Group Variance Ratio		Standard Deviation Group 2	
	Number of Clusters K1	Average Cluster Size M1	Coefficient of Variation of Cluster Sizes COV	Group 1 Treatment	Group 2 Control	Total			ICC $\rho$	$\theta$	$\sigma_2$	Alpha
				N1	N2	N						
0.57481	4	10	0.65	40	70	110	-0.5	0	0.9	0.05	1	0.025
0.56222	4	10	0.65	40	70	110	-0.5	0	1.0	0.05	1	0.025
0.55012	4	10	0.65	40	70	110	-0.5	0	1.1	0.05	1	0.025
0.69611	6	10	0.65	60	70	130	-0.5	0	0.9	0.05	1	0.025
0.68459	6	10	0.65	60	70	130	-0.5	0	1.0	0.05	1	0.025
0.67335	6	10	0.65	60	70	130	-0.5	0	1.1	0.05	1	0.025
0.77114	8	10	0.65	80	70	150	-0.5	0	0.9	0.05	1	0.025
0.76132	8	10	0.65	80	70	150	-0.5	0	1.0	0.05	1	0.025
0.75163	8	10	0.65	80	70	150	-0.5	0	1.1	0.05	1	0.025
0.81996	10	10	0.65	100	70	170	-0.5	0	0.9	0.05	1	0.025
0.81173	10	10	0.65	100	70	170	-0.5	0	1.0	0.05	1	0.025
0.80354	10	10	0.65	100	70	170	-0.5	0	1.1	0.05	1	0.025

- Power The probability of rejecting a false null hypothesis when the alternative hypothesis is true.
- K1 The number of clusters in group 1.
- M1 The average cluster size (number of subjects) of the clusters in group 1.
- COV The coefficient of variation of the cluster sizes. This is the standard deviation of cluster sizes divided by the average cluster size.
- N1, N2, and N The number of subjects in groups 1 and 2, and their total. Note that N1 = K1 × M1.
- NIM The margin of non-inferiority. Since higher means are better, this value is negative and is the maximum distance below the control group mean that the treatment mean can be and still be considered as non-inferior.
- $\delta$  The mean difference in the response at which the power is calculated.  $\delta = \mu_T - \mu_C$ . Note that it does not matter whether the treatment mean ( $\mu_T$ ) is associated with the group with clustering or the group with no clustering. Both configurations lead to the same statistical hypotheses.
- $\theta$  The ratio of the group 1 and group 2 variances. In this case,  $\theta = (\sigma_u^2 + \sigma_e^2) / \sigma_2^2$ . Hence,  $\theta = 1$  implies that  $\sigma_1^2 = \sigma_2^2 = (\sigma_u^2 + \sigma_e^2)$ .
- $\rho$  The intraclass correlation (ICC). The correlation between a (randomly selected) pair of responses within a cluster. In this case,  $\rho = \sigma_u^2 / (\sigma_u^2 + \sigma_e^2)$ , where  $\sigma_u$  is the between-cluster standard deviation and  $\sigma_e$  is the within-cluster standard deviation of the responses in group 1
- $\sigma_2$  The standard deviation of the subject responses in the non-cluster arm, group 2.
- Alpha The probability of rejecting a true null hypothesis.

#### Summary Statements

A parallel, two-group design (where higher means are considered to be better), with cluster-randomized subjects in Group 1 only (the treatment group), and no clustering in Group 2 (the control group), will be used to test whether the treatment mean ( $\mu_T$ ) of Group 1 is non-inferior to the control mean ( $\mu_C$ ) of Group 2, with a non-inferiority margin of -0.5 (H0:  $\delta \leq -0.5$  versus H1:  $\delta > -0.5$ ,  $\delta = \mu_T - \mu_C$ ). The comparison will be made using a one-sided mixed model test, with a Type I error rate ( $\alpha$ ) of 0.025. For Group 2 (the control group), the subject-to-subject standard deviation is assumed to be 1. The Group 1 (treatment) to Group 2 (control) variance ratio is assumed to be 0.9. For Group 1 (the treatment group), the intraclass correlation coefficient is assumed to be 0.05, and the coefficient of variation of cluster sizes is assumed to be 0.65. To detect a mean difference ( $\mu_T - \mu_C$ ) of 0, with 4 clusters of 10 subjects per cluster in the treatment group (Group 1) (totaling 40 subjects), and 70 subjects in the control group (Group 2), the power is 0.57481.

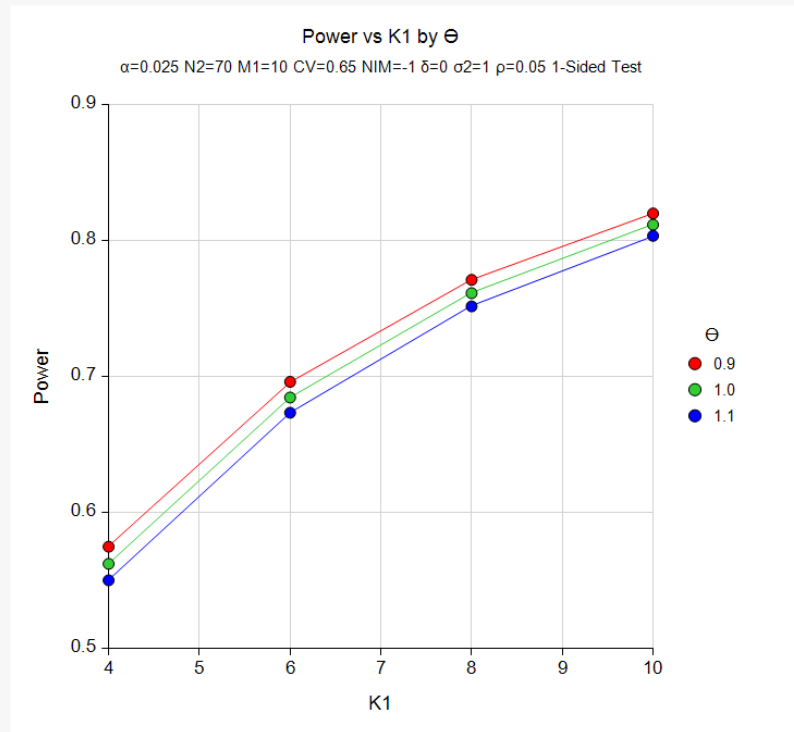
**References**

Moerbeek, M. and Wong, W.K. 2008. 'Sample size formulae for trials comparing group and individual treatments in a multilevel model.' *Statistics in Medicine*, Vol. 27, pages 2850-2864.  
 Julious, Steven A. 2023. *Sample Sizes for Clinical Trials*. Second Edition. CRC Press. New York.  
 Donner, A. and Klar, N. 2000. *Design and Analysis of Cluster Randomization Trials in Health Research*. Arnold. London.

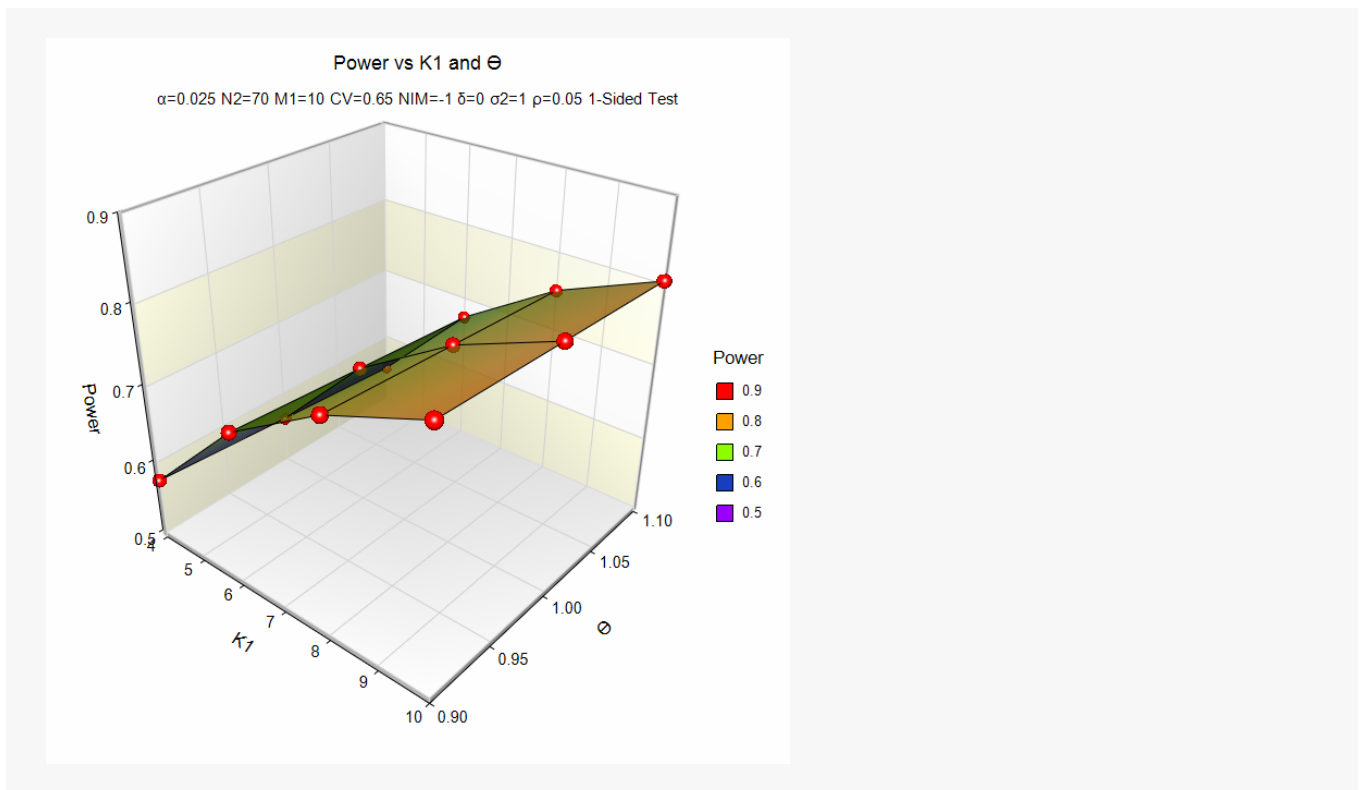
This report shows the power for each of the scenarios.

**Plots Section**

**Plots**



Non-Inferiority Tests for Two Means in a Cluster-Randomized Design with Clustering in Only One Arm



These plots show the power versus the number of clusters by the group variance ratio.



## Example 2 – Validation using a Previously Validated Procedure

This procedure will be validated using the results of the previously validated procedure: **Non-Inferiority Tests for Two Means in a Cluster-Randomized Design**. In order to create a validation example, the cluster size of one group will be set to one.

Note that the previous procedure uses the non-central t distribution for calculations while this procedure uses the normal distribution. This will make the power values slightly different.

The parameter values are set as follows:  $\alpha = 0.025$ ,  $K1 = 70$ ,  $M1 = 1$ ,  $K2 = 8$ ,  $M2 = 10$ ,  $COV = 0$ ,  $NIM = 0.5$ ,  $\delta = 0.0$ ,  $\sigma = 1$ , and  $\rho = 0.05$ . The resulting power is calculated to be 0.78794.

### Setup of the Validation Procedure

If the **Non-Inferiority Tests for Two Means in a Cluster-Randomized Design** 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 **Validation of Only One Clustered Arm Procedure** 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 Means Are .....	<b>Better (Ha: <math>\delta &gt; -NIM</math>)</b>
Test Statistic .....	<b>T-Test Based on Number of Subjects</b>
Alpha.....	<b>0.025</b>
K1 (Number of Clusters) .....	<b>70</b>
M1 (Average Cluster Size).....	<b>1</b>
K2 (Number of Clusters) .....	<b>8</b>
M2 (Average Cluster Size).....	<b>10</b>
COV of Cluster Sizes.....	<b>0</b>
NIM (Non-Inferiority Margin) .....	<b>0.5</b>
$\delta$ (Mean Difference = $\mu_1 - \mu_2$ ).....	<b>0</b>
$\sigma$ (Standard Deviation).....	<b>1</b>
$\rho$ (Intracluster Correlation, ICC).....	<b>0.05</b>

## Validation Output

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

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														
Higher Means Are:	Better														
Hypotheses:	H0: $\delta \leq -NIM$ vs. H1: $\delta > -NIM$														
Power	Number of Clusters			Cluster Size			Sample Size			Mean Difference $\delta$	Non-Inferiority Margin -NIM	Standard Deviation $\sigma$	ICC $\rho$	Alpha	
	K1	K2	K	M1	M2	COV	N1	N2	N						
0.78794	70	8	78	1	10	0	70	80	150	0	-0.5	1	0.05	0.025	

PASS calculates the power value as 0.78794.

## Setup

Now the parameter values of the current procedure are set as follows:  $\alpha = 0.025$ ,  $N2 = 70$ ,  $K1 = 8$ ,  $M1 = 10$ ,  $COV = 0$ ,  $NIM = 0.5$ ,  $\delta = 0.0$ ,  $\sigma^2 = 1$ ,  $\theta = 1$ , and  $\rho = 0.05$ .

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 .....	<b>Power</b>
Assign Group 1 (Clusters) to be the.....	<b>Treatment Group</b>
Higher Means Are.....	<b>Better (H1: <math>\delta &gt; -NIM</math>, <math>\delta = \mu_T - \mu_C</math>)</b>
Alpha.....	<b>0.025</b>
Group Allocation .....	<b>Enter K1 and N2 individually</b>
K1 (Number of Clusters) .....	<b>8</b>
M1 (Average Cluster Size).....	<b>10</b>
COV (Coefficient of Variation of Cluster Sizes).....	<b>0</b>
N2 (Number of Subjects) .....	<b>70</b>
NIM (Non-Inferiority Margin) .....	<b>0.5</b>
$\delta$ (Mean Difference = $\mu_T - \mu_C$ ) .....	<b>0</b>
$\sigma^2$ (Standard Deviation in Group 2) .....	<b>1</b>
$\theta$ (Group Variance Ratio = $\sigma_1^2 / \sigma_2^2$ ) .....	<b>1</b>
$\rho$ (Intracluster Correlation, ICC) .....	<b>0.05</b>

## Output

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

### Numeric Results

Solve For: Power  
 Group 1 Assignment: Group 1 has Clusters and is the Treatment Group  
 Group 2 Assignment: Group 2 does not have Clusters and is the Control Group  
 Group Allocation: Enter K1 and N2 individually  
 Higher Means Are: Better  
 Hypotheses:  $H_0: \delta \leq -NIM$  vs.  $H_1: \delta > -NIM$

Power	Group 1 (Treatment) Clusters			Sample Size (Subjects)			Non-Inferiority Margin -NIM	Mean Difference $\delta$	Group Variance Ratio $\theta$		Standard Deviation Group 2 $\sigma_2$	
	Number of Clusters K1	Average Cluster Size M1	Coefficient of Variation of Cluster Sizes COV	Group 1 Treatment N1	Group 2 Control N2	Total N			ICC $\rho$	Alpha		
0.79308	8	10	0	80	70	150	-0.5	0	1	0.05	1	0.025

PASS calculates a power of 0.79308. The difference between this value and the previous value of 0.78794 is 0.00514. This slight difference occurs because the other procedure uses the non-central t distribution for its power calculation, but this procedure uses the normal distribution.