Chapter 112

# Confidence Intervals for One Mean in a Stratified Cluster-Randomized Design

## Introduction

This procedure calculates sample size and half-width for confidence intervals of a mean from a stratified cluster randomization trial (CRT) in which the outcome variable is continuous. It uses the results from elementary sampling theory which are presented in Wang, Zhang, and Ahn (2017) and Xu, Zhu, and Ahn (2019).

Suppose that the mean of a continuous outcome variable of a sample from a population of subjects (or items) is to be estimated with a confidence interval. Further suppose that the population can be separated into a few subpopulations, often called *strata*. Further suppose that each stratum can be separated into a number of clusters and that sampling occurs at the cluster level. That is, a simple random sample of clusters is drawn within a stratum. Next, a simple random sample of subjects is drawn from within each cluster.

Note that this procedure assumes an infinite population in which the size of every cluster and every stratum is not known.

This procedure allows you to determine the appropriate sample size to be taken from each stratum so that width of the confidence interval is guaranteed.

## Technical Details

The following discussion summarizes the results in Xu *et al*. (2019) and Wang *et al.* (2017).

Suppose you are interested in estimating the mean of a particular population. Further suppose that response is known to be related to other covariates (such as age, race, or gender). It may be possible to improve estimation efficiency by stratifying on one or more of these covariates.

In this design, assume clusters are grouped into *H* strata. Let $K_h$ denote the number of clusters sampled in the $h^{th}$ stratum, *h* = 1, ..., *H*.

Let $N_{kh}$ denote the number of subjects sampled (the cluster size) in cluster *k* in stratum *h*, $k = 1, \ldots, K_h, h = 1, \ldots, H$. Assume that the $N_{kh}'s$ are independently and identically distributed with mean $M_h$ and variance $v_h^2$. The total number of subjects in the trial is $N = \sum_{h=1}^{H}\sum_{k=1}^{K_h} N_{kh} = \sum_{h=1}^{H} N_h$ where $N_h$ is the number of subjects sampled from stratum *h*. Note that $N_h = \sum_{k=1}^{K_h} N_{kh} = K_h M_h$.

Let $Y_{ikh}$ indicate the continuous variable of subject *i* of cluster *j* of stratum *h*. Let $\mu_h$ indicate the mean in stratum *k*. This value is estimated by

$$\hat{\mu}_h = \sum_{k=1}^{K_h}\sum_{i=1}^{n_{kh}} Y_{ikh} \,/\, \sum_{k=1}^{K_h} N_{kh}.$$

The variance of $\hat{\mu}_h$ is given by

$$V(\hat{\mu}_h) = \frac{\sigma_h^2}{K_h M_h}\left[\rho M_h\left(C_h^2 + 1\right) + (1-\rho)\right]$$

$$= \frac{\sigma_h^2}{K_h M_h} A_h$$

where $C_h = v_h / M_h$ is the coefficient of variation of the sizes of clusters within stratum $h$.

Let $\rho$ indicate the intracluster correlation coefficient (ICC) give the correlation of subjects within the same cluster. This value is assumed to be constant for all clusters.

The overall mean $\mu$ is given by

$$\mu = \sum_{h=1}^{H} f_h \mu_h$$

where $f_h$ is the fraction of sampled subjects in stratum $h$. Note that $f_h = K_h M_h / N$.

The parameter $\mu$ is estimated by

$$\hat{\mu} = \sum_{h=1}^{H} f_h \hat{\mu}_h$$

The variance of this estimate is given by

$$V(\hat{\mu}) = \sum_{h=1}^{H} f_h^2 \frac{\sigma_h^2}{K_h M_h}\left[\rho M_h\left(C_h^2 + 1\right) + (1-\rho)\right]$$

$$= \sum_{h=1}^{H} \left(\frac{K_h M_h}{N}\right)^2 \frac{\sigma_h^2}{K_h M_h} A_h$$

$$= \frac{1}{N^2} \sum_{h=1}^{H} K_h M_h \sigma_h^2 A_h$$

If the common assumption is made that $\hat{\mu}$ is asymptotically standard normal, then a confidence interval for $\mu$ can be constructed as follows

$$CI(\mu) = \hat{\mu} \pm z_{1-\alpha/2}\sqrt{V(\hat{\mu})}$$

The lower and upper limits of this confidence interval are denoted as $LCL_\mu$ and $UCL_\mu$. The half-width, $d$, of this interval is given by

$$d = \left|z_{1-\alpha/2}\right|\sqrt{V(\hat{\mu})}$$

This formula can be used to determine the sample size $N$ required to achieve a specified value of $d$.

## Estimating ICC

An often-difficult task necessary in computing the sample size is to estimate the value of the intracluster correlation coefficient (ICC or ρ). Xu et al. (2019) provides guidance in estimating this parameter using the ANOVA method. The **PASS** procedure *Confidence Intervals for Intraclass Correlation* provides methods for estimating ICC within a stratum. These stratum estimates can be averaged to provide an overall estimate.

Step 1. Estimate $\rho_h$ for each stratum using one of the methods given in the **PASS** procedure *Confidence Intervals for Intraclass Correlation.*

Step 2. Compute the average of these estimates and use it as the overall estimate of $\rho$.

# Example 1 – Finding Sample Size

A study using a stratified cluster design is being planned to estimate the effectiveness of a certain drug in treating a certain disease. The strata are four large metropolitan areas. The clusters are doctor's practices.

The average size of the practices in each of the strata are 80, 60, 50, 40. The cluster allocation pattern for the relative frequencies of clusters for the strata are 1, 1.5, 1.75, and 2. The COV for all strata will be set to 0.40. The ICC of similar studies has been 0.02.

Prior studies have shown a standard deviation for this disease is 0.4702. The confidence level is set to 0.95 and *d* is set to three values 0.02, 0.03, 0.04.

## 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**
Confidence Level ..... **0.95**
d (Precision, Half-Width) ..... **0.02 0.03 0.04**
Cluster Allocation Pattern ..... **Proportional (Enter Rh = Cluster Allocation Pattern)**
Mh (Average Cluster Size) ..... **Custom**
Adjust results… ..... **Checked**
Ch (COV of Cluster Sizes) ..... **All Equal**
Ch for All Strata ..... **0.4**
Sh (Standard Deviations) ..... **All Equal**
Sh for All Strata ..... **0.4702**
ρ (Intracluster Correlation, ICC) ..... **0.02**

Set 1 Number of Strata ..... **1**
Set 1 Rh (Cluster Allocation Pattern) ..... **1**
Set 1 Mh (Average Cluster Size) ..... **80**

Set 2 Number of Strata ..... **1**
Set 2 Rh (Cluster Allocation Pattern) ..... **1.5**
Set 2 Mh (Average Cluster Size) ..... **60**

Set 3 Number of Strata ..... **1**
Set 3 Rh (Cluster Allocation Pattern) ..... **1.75**
Set 3 Mh (Average Cluster Size) ..... **50**

Set 4 Number of Strata ..... **1**
Set 4 Rh (Cluster Allocation Pattern) ..... **2**
Set 4 Mh (Average Cluster Size) ..... **40**

Set 5 Number of Strata ..... **0**
Show More Strata Sets ..... **Unchecked**

# Output

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

## Numeric Reports

**Numeric Results**

Solve For: Sample Size
Number of Strata: 4
Allocation: Proportional

| Confidence Interval Half-Width d | Sample Size: Total Number of Subjects N | Sample Size: Total Number of Clusters K | Sample Size: Average Clusters per Stratum K0 | Sample Size: Average Cluster Size M | Sample Size: Average COV of Cluster Sizes C | Standard Deviation S | Intracluster Correlation (ICC) ρ | Confidence Level CL |
|---|---|---|---|---|---|---|---|---|
| 0.0200 | 4930 | 91 | 22.75 | 54 | 0.4 | 0.4702 | 0.02 | 0.95 |
| 0.0297 | 2230 | 41 | 10.25 | 54 | 0.4 | 0.4702 | 0.02 | 0.95 |
| 0.0396 | 1260 | 23 | 5.75 | 54 | 0.4 | 0.4702 | 0.02 | 0.95 |

d The half-width of the confidence interval of the mean. d = [UCL(μ) - LCL(μ)] / 2.
N The total number of subjects.
K The total number of clusters.
K0 The average number of clusters per stratum.
M The average cluster size. The weighted average of the number of subjects per cluster.
C The average COV of cluster sizes. The weighted average COV of all clusters.
S The square root of weighted average of the strata variances. The weights are proportional to the number of subjects.
ρ The intracluster correlation coefficient (ICC) averaged across all strata.
CL The confidence level of the confidence interval for the mean.

**Summary Statements**

A stratified cluster-randomized design will be used to obtain a two-sided 95% confidence interval for a single mean. The overall (weighted average) standard deviation is assumed to be 0.4702 and the intracluster correlation coefficient is assumed to be 0.02. The average coefficient of variation of cluster sizes is assumed to be 0.4. With an average cluster size of 54, and with clusters allocated to 4 strata, to produce a confidence interval with a half-width of no more than 0.02, a total of 91 clusters (an average of 22.75 per stratum) will be needed. The total sample size needed is 4930.

This report gives the results for each of the three values of *d*.

## Strata-Detail Report

**Strata-Detail Report for Row 1**

| Strata h | Sample Size: Number of Subjects Nh | Sample Size: Number of Clusters Kh | Sample Size: Average Cluster Size Mh | Sample Size: COV of Cluster Sizes Ch | Proportion of Total: Subjects Fh | Proportion of Total: Clusters sRh | Standard Deviation Sh |
|---|---|---|---|---|---|---|---|
| 1 | 1200 | 15 | 80 | 0.4 | 0.243 | 0.16 | 0.4702 |
| 2 | 1320 | 22 | 60 | 0.4 | 0.268 | 0.24 | 0.4702 |
| 3 | 1250 | 25 | 50 | 0.4 | 0.254 | 0.28 | 0.4702 |
| 4 | 1160 | 29 | 40 | 0.4 | 0.235 | 0.32 | 0.4702 |

**Strata-Detail Report for Row 2**

| | Sample Size | | | | Proportion of Total | | |
|---|---|---|---|---|---|---|---|
| Strata h | Number of Subjects Nh | Number of Clusters Kh | Average Cluster Size Mh | COV of Cluster Sizes Ch | Subjects Fh | Clusters sRh | Standard Deviation Sh |
| 1 | 560 | 7 | 80 | 0.4 | 0.251 | 0.16 | 0.4702 |
| 2 | 600 | 10 | 60 | 0.4 | 0.269 | 0.24 | 0.4702 |
| 3 | 550 | 11 | 50 | 0.4 | 0.247 | 0.28 | 0.4702 |
| 4 | 520 | 13 | 40 | 0.4 | 0.233 | 0.32 | 0.4702 |

**Strata-Detail Report for Row 3**

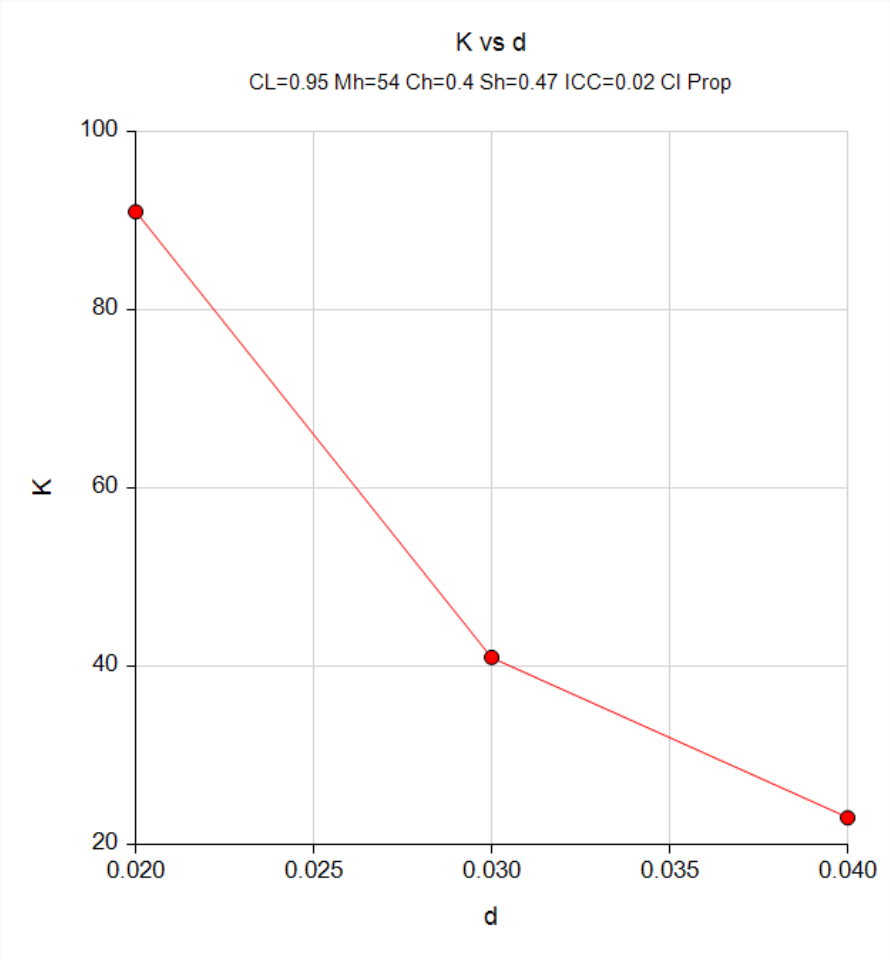
| | Sample Size | | | | Proportion of Total | | |
|---|---|---|---|---|---|---|---|
| Strata h | Number of Subjects Nh | Number of Clusters Kh | Average Cluster Size Mh | COV of Cluster Sizes Ch | Subjects Fh | Clusters sRh | Standard Deviation Sh |
| 1 | 320 | 4 | 80 | 0.4 | 0.254 | 0.16 | 0.4702 |
| 2 | 360 | 6 | 60 | 0.4 | 0.286 | 0.24 | 0.4702 |
| 3 | 300 | 6 | 50 | 0.4 | 0.238 | 0.28 | 0.4702 |
| 4 | 280 | 7 | 40 | 0.4 | 0.222 | 0.32 | 0.4702 |

h An arbitrary sequence number for each stratum.
Nh The number of subjects in stratum h. Nh = Kh x Mh.
Kh The number of clusters in stratum h.
Mh The average cluster size in stratum h.
Ch The COV of the cluster sizes in stratum h.
Fh The proportion of the total subjects in stratum h.
sRh The proportion of the total clusters in stratum h.
Sh The standard deviation of the response in stratum h.

This report shows the values of the individual, strata-level parameters.

## Plots Section

**Plots**




The values from the Numerical Results report are displayed in this plot.

# Example 2 – Validation using Hand Calculations

We could not find an example of this procedure in the literature, so we will validate it using hand calculations. To do this, we will use the following example.

Suppose a stratified cluster design has two clusters: A and B. Suppose the number of clusters per stratum is 10 for stratum A and 20 for stratum B. Suppose the average cluster sizes are 20 in both strata and the COV of cluster sizes are 0.4 in both strata. Suppose the standard deviations in A and B are 0.4899 and 0.5, respectively. Further suppose that the ICC is 0.1 and the confidence level is 0.95.

These strata values are summarized in the following table.

| Strata h | Number Clusters Kh | Average Cluster Size Mh | COV of Cluster Sizes Ch | Standard Deviation Sh |
|---|---|---|---|---|
| A | 10 | 20 | 0.4 | 0.4899 |
| B | 20 | 20 | 0.4 | 0.5 |

First, calculate N = 10(20) + 20(20) = 600.

Next, calculate $f_A = \frac{K_A M_A}{N} = \frac{10 \times 20}{600} = \frac{1}{3}$.

Similarly, calculate $f_B = \frac{K_B M_B}{N} = \frac{20 \times 20}{600} = \frac{2}{3}$.

Next, calculate $A_A = A_B = \rho M_h\left(C_h^2 + 1\right) + (1 - \rho) = 0.1(20)(0.4^2 + 1) + (1 - 0.1) = 3.22$.

The variance can then be calculated as

$$V(\hat{\mu}) = \sum_{h=1}^{H} f_h^2 \frac{\sigma_h^2}{K_h M_h}\left[\rho M_h\left(C_h^2 + 1\right) + (1 - \rho)\right]$$

$$= 3.22\left[\frac{1}{9}\left(\frac{0.4899^2}{200}\right) + \frac{4}{9}\left(\frac{0.5^2}{400}\right)\right]$$

$$= 3.22\left[\frac{0.0012}{9} + \frac{0.0025}{9}\right]$$

$$= 0.00132377778$$

Finally, the half-width is calculated as

$$d = \left|z_{1-\alpha/2}\right|\sqrt{V(p)}$$

$$= 1.95996398\sqrt{0.00132377778} = 0.07131085$$

## 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 ..... **Half-Width of C.I.**
Confidence Level ..... **0.95**
Cluster Allocation Pattern ..... **Custom Kh**
Mh (Average Cluster Size) ..... **All Equal**
Mh for All Strata ..... **20**
Adjust results… ..... **Checked**
Ch (COV of Cluster Sizes) ..... **All Equal**
Ch for All Strata ..... **0.4**
Sh (Standard Deviations) ..... **Custom**
ρ (Intracluster Correlation, ICC) ..... **0.1**

Set 1 Number of Strata ..... **1**
Set 1 Kh (Number of Clusters) ..... **10**
Set 1 Sh (Standard Deviation) ..... **0.4899**

Set 2 Number of Strata ..... **1**
Set 2 Kh (Number of Clusters) ..... **20**
Set 2 Sh (Standard Deviation) ..... **0.5**

## Output

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

**Numeric Results**

Solve For: Half-Width of C.I.
Number of Strata: 2
Allocation: Custom

| | Sample Size | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Confidence Interval Half-Width d** | **Total Number of Subjects N** | **Total Number of Clusters K** | **Average Clusters per Stratum K0** | **Average Cluster Size M** | **Average COV of Cluster Sizes C** | **Standard Deviation S** | **Intracluster Correlation (ICC) ρ** | **Confidence Level CL** |
| 0.0713 | 600 | 30 | 15 | 20 | 0.4 | 0.4966 | 0.1 | 0.95 |

This report shows that **PASS** has also computed d = 0.0713. Thus, the procedure is validated.

# Example 3 – Looking at the Impact of ICC on the Half-Width

We will continue with the scenario began in Example 1 to show the impact of the intracluster correlation coefficient (ICC) on half-width.

From Example 1: a study using a stratified cluster design is being planned to estimate the effectiveness of a certain drug in treating a certain disease. The strata are four large metropolitan areas. The clusters are doctor's practices. The total number of clusters will be set to 100. The average size of the practices in each of the strata are 80, 60, 50, 40. The cluster allocation pattern for the relative frequencies of clusters for the strata are 1, 1.5, 1.75, and 2. The COV for all strata will be set to 0.40. Prior studies have shown the standard deviation for this disease is 0.4702. The confidence level is set to 0.95 and *d* will be solved for.

The values of ICC will be 0, 0.05, 0.1, 0.2, 0.4, 0.6, 0.8, 0.9, 0.99, 0.999.

## 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 3** 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 ..... **Half-Width of C.I.**
Confidence Level ..... **0.95**
Cluster Allocation Pattern ..... **Proportional (Enter Rh = Cluster Allocation Pattern and K)**
K (Total Number of Clusters) ..... **100**
Mh (Average Cluster Size)..... **Custom**
Adjust results… ..... **Checked**
Ch (COV of Cluster Sizes) ..... **All Equal**
Ch for All Strata ..... **0.4**
Sh (Standard Deviations)..... **All Equal**
Sh for All Strata..... **0.4702**
ρ (Intracluster Correlation, ICC) ..... **0 0.05 0.1 0.2 0.4 0.6 0.8 0.9 0.99 0.999**

Set 1 Number of Strata ..... **1**
Set 1 Rh (Cluster Allocation Pattern) ..... **1**
Set 1 Mh (Average Cluster Size) ..... **80**

Set 2 Number of Strata ..... **1**
Set 2 Rh (Cluster Allocation Pattern) ..... **1.5**
Set 2 Mh (Average Cluster Size) ..... **60**

Set 3 Number of Strata ..... **1**
Set 3 Rh (Cluster Allocation Pattern) ..... **1.75**
Set 3 Mh (Average Cluster Size) ..... **50**

Set 4 Number of Strata ..... **1**
Set 4 Rh (Cluster Allocation Pattern) ..... **2**
Set 4 Mh (Average Cluster Size) ..... **40**

# Output

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

## Numeric Results

**Numeric Results**

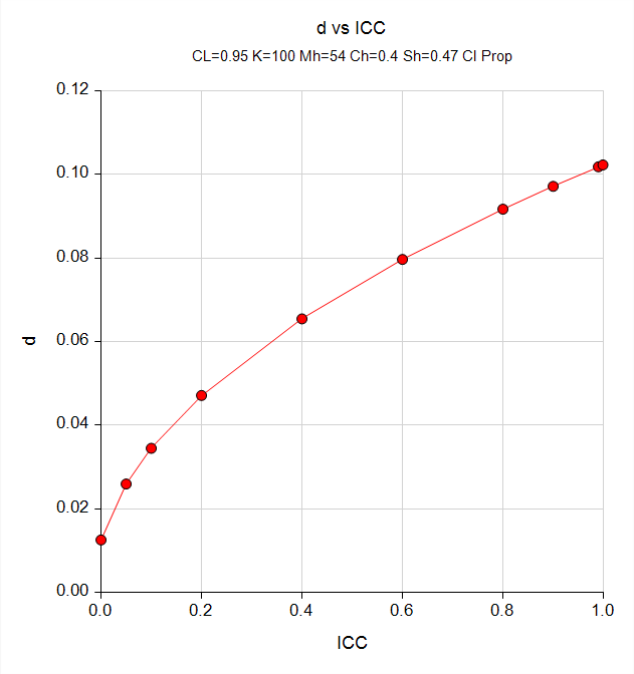
Solve For: Half-Width of C.I.
Number of Strata: 4
Allocation: Proportional

| Confidence Interval Half-Width d | Sample Size: Total Number of Subjects N | Sample Size: Total Number of Clusters K | Sample Size: Average Clusters per Stratum K0 | Sample Size: Average Cluster Size M | Sample Size: Average COV of Cluster Sizes C | Standard Deviation S | Intracluster Correlation (ICC) ρ | Confidence Level CL |
|---|---|---|---|---|---|---|---|---|
| 0.0125 | 5400 | 100 | 25 | 54 | 0.4 | 0.4702 | 0.000 | 0.95 |
| 0.0259 | 5400 | 100 | 25 | 54 | 0.4 | 0.4702 | 0.050 | 0.95 |
| 0.0345 | 5400 | 100 | 25 | 54 | 0.4 | 0.4702 | 0.100 | 0.95 |
| 0.0471 | 5400 | 100 | 25 | 54 | 0.4 | 0.4702 | 0.200 | 0.95 |
| 0.0655 | 5400 | 100 | 25 | 54 | 0.4 | 0.4702 | 0.400 | 0.95 |
| 0.0797 | 5400 | 100 | 25 | 54 | 0.4 | 0.4702 | 0.600 | 0.95 |
| 0.0917 | 5400 | 100 | 25 | 54 | 0.4 | 0.4702 | 0.800 | 0.95 |
| 0.0972 | 5400 | 100 | 25 | 54 | 0.4 | 0.4702 | 0.900 | 0.95 |
| 0.1018 | 5400 | 100 | 25 | 54 | 0.4 | 0.4702 | 0.990 | 0.95 |
| 0.1023 | 5400 | 100 | 25 | 54 | 0.4 | 0.4702 | 0.999 | 0.95 |

This report gives the results for each of the various values of *ICC*.

## Plots Section

**Plots**




This plot shows the impact on half-width of increasing *ICC*. The value of *d* increases from 0.0125 to 0.1023.

# Example 4 – Looking at the Impact of ICC on the Sample Size

We will continue with the scenario began in Example 1 to show the impact of the intracluster correlation coefficient (ICC) on sample size.

From Example 1: a study using a stratified cluster design is being planned to estimate the effectiveness of a certain drug in treating a certain disease. The strata are four large metropolitan areas. The clusters are doctor's practices. The average size of the practices in each of the strata are 80, 60, 50, 40. The cluster allocation pattern for the relative frequencies of clusters for the strata are 1, 1.5, 1.75, and 2. The COV for all strata will be set to 0.40. Prior studies have shown the standard deviation for this disease is 0.4702. The confidence level is set to 0.95 and *d* will be set to 0.05. The total number of clusters, K, will be solved for.

The values of ICC will be 0, 0.05, 0.1, 0.2, 0.4, 0.6, 0.8, 0.9, 0.99, 0.999.

## 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 4** 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**
Confidence Level ..... **0.95**
d (Precision, Half-Width) ..... **0.05**
Cluster Allocation Pattern ..... **Proportional (Enter Rh = Cluster Allocation Pattern)**
Mh (Average Cluster Size) ..... **Custom**
Adjust results… ..... **Checked**
Ch (COV of Cluster Sizes) ..... **All Equal**
Ch for All Strata ..... **0.4**
Sh (Standard Deviations) ..... **All Equal**
Sh for All Strata ..... **0.4702**
ρ (Intracluster Correlation, ICC) ..... **0 0.05 0.1 0.2 0.4 0.6 0.8 0.9 0.99 0.999**

Set 1 Number of Strata ..... **1**
Set 1 Rh (Cluster Allocation Pattern) ..... **1**
Set 1 Mh (Average Cluster Size) ..... **80**

Set 2 Number of Strata ..... **1**
Set 2 Rh (Cluster Allocation Pattern) ..... **1.5**
Set 2 Mh (Average Cluster Size) ..... **60**

Set 3 Number of Strata ..... **1**
Set 3 Rh (Cluster Allocation Pattern) ..... **1.75**
Set 3 Mh (Average Cluster Size) ..... **50**

Set 4 Number of Strata ..... **1**
Set 4 Rh (Cluster Allocation Pattern) ..... **2**
Set 4 Mh (Average Cluster Size) ..... **40**

# Output

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

## Numeric Results

**Numeric Results**

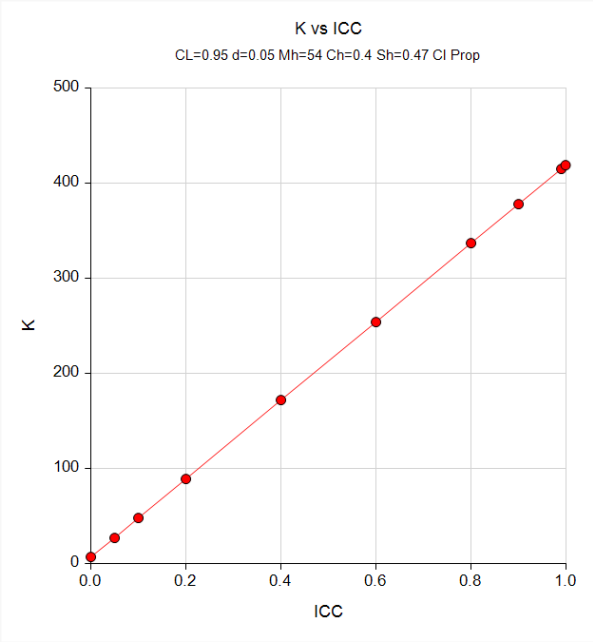
Solve For: Sample Size
Number of Strata: 4
Allocation: Proportional

| Confidence Interval Half-Width d | Sample Size: Total Number of Subjects N | Sample Size: Total Number of Clusters K | Sample Size: Average Clusters per Stratum K0 | Sample Size: Average Cluster Size M | Sample Size: Average COV of Cluster Sizes C | Standard Deviation S | Intracluster Correlation (ICC) ρ | Confidence Level CL |
|---|---|---|---|---|---|---|---|---|
| 0.0473 | 380 | 7 | 1.75 | 54 | 0.4 | 0.4702 | 0.000 | 0.95 |
| 0.0500 | 1440 | 27 | 6.75 | 54 | 0.4 | 0.4702 | 0.050 | 0.95 |
| 0.0498 | 2610 | 48 | 12.00 | 54 | 0.4 | 0.4702 | 0.100 | 0.95 |
| 0.0500 | 4790 | 89 | 22.25 | 54 | 0.4 | 0.4702 | 0.200 | 0.95 |
| 0.0499 | 9300 | 172 | 43.00 | 54 | 0.4 | 0.4702 | 0.400 | 0.95 |
| 0.0500 | 13730 | 254 | 63.50 | 54 | 0.4 | 0.4702 | 0.600 | 0.95 |
| 0.0500 | 18200 | 337 | 84.25 | 54 | 0.4 | 0.4702 | 0.800 | 0.95 |
| 0.0500 | 20400 | 378 | 94.50 | 54 | 0.4 | 0.4702 | 0.900 | 0.95 |
| 0.0500 | 22400 | 415 | 103.75 | 54 | 0.4 | 0.4702 | 0.990 | 0.95 |
| 0.0500 | 22630 | 419 | 104.75 | 54 | 0.4 | 0.4702 | 0.999 | 0.95 |

This report gives the results for each of the various values of ICC.

## Plots Section

**Plots**




This plot shows the impact on sample size (number of clusters) of increasing *ICC*. The value of *K* increases from 7 to 419 and the value of N increases from 380 to 22,630.

# Example 5 – Looking at the Impact of COV on the Sample Size

We will continue with the scenario began in Example 1 to show the impact of the cluster size COV on sample size.

From Example 1: a study using a stratified cluster design is being planned to estimate the effectiveness of a certain drug in treating a certain disease. The strata are four large metropolitan areas. The clusters are doctor's practices. The average size of the practices in each of the strata are 80, 60, 50, 40. The cluster allocation pattern for the relative frequencies of clusters for the strata are 1, 1.5, 1.75, and 2. Prior studies have shown the standard deviation for this disease is 0.4702. The confidence level is set to 0.95 and *d* will be set to 0.05. The total number of clusters, K, will be solved for.

The values of Ch will be 0, 0.1, 0.3, 0.5, 0.7, 0.9, 1.1, 1.3, and 1.5.

## 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 5** 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**
Confidence Level ..... **0.95**
d (Precision, Half-Width) ..... **0.05**
Cluster Allocation Pattern ..... **Proportional (Enter Rh = Cluster Allocation Pattern)**
Mh (Average Cluster Size) ..... **Custom**
Adjust results… ..... **Checked**
Ch (COV of Cluster Sizes) ..... **All Equal**
Ch for All Strata ..... **0 0.1 0.3 0.5 0.7 0.9 1.1 1.3 1.5**
Sh (Standard Deviations) ..... **All Equal**
Sh for All Strata ..... **0.4702**
ρ (Intracluster Correlation, ICC) ..... **0.2**

Set 1 Number of Strata ..... **1**
Set 1 Rh (Cluster Allocation Pattern) ..... **1**
Set 1 Mh (Average Cluster Size) ..... **80**

Set 2 Number of Strata ..... **1**
Set 2 Rh (Cluster Allocation Pattern) ..... **1.5**
Set 2 Mh (Average Cluster Size) ..... **60**

Set 3 Number of Strata ..... **1**
Set 3 Rh (Cluster Allocation Pattern) ..... **1.75**
Set 3 Mh (Average Cluster Size) ..... **50**

Set 4 Number of Strata ..... **1**
Set 4 Rh (Cluster Allocation Pattern) ..... **2**
Set 4 Mh (Average Cluster Size) ..... **40**

# Output

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

## Numeric Results

**Numeric Results**

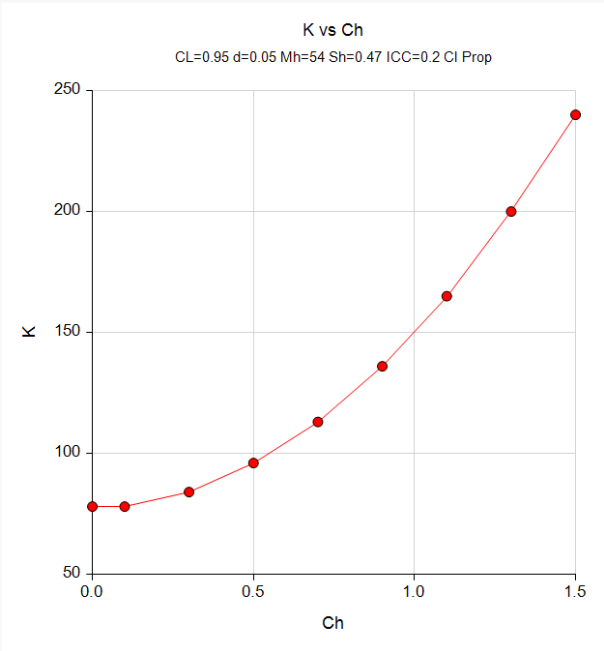
Solve For: Sample Size
Number of Strata: 4
Allocation: Proportional

| Confidence Interval Half-Width d | Sample Size: Total Number of Subjects N | Sample Size: Total Number of Clusters K | Sample Size: Average Clusters per Stratum K0 | Sample Size: Average Cluster Size M | Sample Size: Average COV of Cluster Sizes C | Standard Deviation S | Intracluster Correlation (ICC) ρ | Confidence Level CL |
|---|---|---|---|---|---|---|---|---|
| 0.0497 | 4200 | 78 | 19.50 | 54 | 0.0 | 0.4702 | 0.2 | 0.95 |
| 0.0500 | 4200 | 78 | 19.50 | 54 | 0.1 | 0.4702 | 0.2 | 0.95 |
| 0.0499 | 4520 | 84 | 21.00 | 54 | 0.3 | 0.4702 | 0.2 | 0.95 |
| 0.0498 | 5170 | 96 | 24.00 | 54 | 0.5 | 0.4702 | 0.2 | 0.95 |
| 0.0499 | 6100 | 113 | 28.25 | 54 | 0.7 | 0.4702 | 0.2 | 0.95 |
| 0.0500 | 7360 | 136 | 34.00 | 54 | 0.9 | 0.4702 | 0.2 | 0.95 |
| 0.0499 | 8900 | 165 | 41.25 | 54 | 1.1 | 0.4702 | 0.2 | 0.95 |
| 0.0499 | 10800 | 200 | 50.00 | 54 | 1.3 | 0.4702 | 0.2 | 0.95 |
| 0.0500 | 12950 | 240 | 60.00 | 54 | 1.5 | 0.4702 | 0.2 | 0.95 |

This report gives the results for each of the various values of *COV*. The value of *K* increases from 78 to 240 and the value of *N* increases from 4200 to 12,950.

## Plots Section

**Plots**




This plot shows the impact on sample size (number of clusters) of increasing *ICC*.