Chapter 113

Confidence Intervals for One Proportion in a Stratified Cluster-Randomized Design

Introduction

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

Suppose that the response proportion of a binary 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 know.

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).

Suppose you are interested in estimating the disease response rate of a particular population. Further suppose that response rate 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, $h = 1, \ldots, H$.

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 $N_h = \sum_{k=1}^{K_h} N_{kh} = K_h M_h$.

© NCSS, LLC. All Rights Reserved.

Let Y_{ikh} indicate the binary variable (0 or 1) of subject *i* of cluster *j* of stratum *h*. Let P_h indicate the response probability (or proportion) in stratum *k*. This value is estimated by

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

The variance of p_h is given by

$$\begin{aligned} V(p_h) &= \frac{P_h(1-P_h)}{K_h M_h}\left[\rho M_h\left(C_h^2+1\right)+(1-\rho)\right] \\ &= \frac{P_h(1-P_h)}{K_h M_h} A_h \end{aligned}$$

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

Let ρ 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 response probability (proportion) *P* is given by

$$P = \sum_{h=1}^{H} f_h P_h$$

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

The parameter *P* is estimated by

$$p = \sum_{h=1}^{H} f_h p_h$$

The variance of this estimate is given by

$$\begin{aligned} V(p) &= \sum_{h=1}^{H} f_h^2 \frac{P_h(1-P_h)}{K_h M_h}\left[\rho M_h\left(C_h^2+1\right)+(1-\rho)\right] \\ &= \frac{1}{N^2}\sum_{h=1}^{H} K_h M_h P_h(1-P_h) A_h \\ &= \frac{1}{N}\sum_{h=1}^{H} f_h P_h(1-P_h) A_h \end{aligned}$$

This quantity can be estimated by substituting p_h for P_h.

If the common assumption is made that *p* is asymptotically standard normal, then a confidence interval for *P* can be constructed as follows

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

The lower and upper limits of this confidence interval are denoted as LCL_P and UCL_P. The half-width, *d*, of this interval is given by

$$\begin{aligned} d &= \left|z_{1-\frac{\alpha}{2}}\right| \sqrt{V(p)} \\ &= \left|z_{1-\alpha/2}\right| \sqrt{\frac{1}{N}\sum_{h=1}^{H} f_h P_h(1-P_h) A_h} \end{aligned}$$

This can be rearranged to provide the following formula for the total sample size.

$$N = \frac{z_{1-\alpha/2}^2}{d^2}\sum_{h=1}^{H} f_h P_h(1-P_h) A_h$$

© NCSS, LLC. All Rights Reserved.

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 ρ_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 ρ.

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, refer to the Procedure Window chapter.

Design Tab

The Design tab contain most of the parameters and options of interest for this procedure.

Solve For

Solve For

This option specifies the parameter to be solved for using the other parameters. The parameters that may be selected are *Sample Size* or *Half-Width of C.I*. Select *Sample Size* when you want to find the number of clusters needed. Select *Half-Width of C.I*. when you want to investigate the precision of a certain cluster count.

Confidence and Precision

Confidence Level

Enter the confidence level (or confidence coefficient). This is the proportion of confidence intervals (constructed with this same confidence level, sample size, etc.) that contain the population proportion.

The practical range is between 0.5 and 1. Common values are 0.95 and 0.99. Use 0.9973 if you want *z* to be 3.0 and 0.977249 if you want *z* to be 2.0.

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

d (Precision, Half-Width)

Enter *d*, the precision, margin of error, or confidence interval half-width. This is half the distance between the lower and upper confidence limits of the proportion.

The formula is d = |UCL(P) - LCL(P)|/2.

The range is 0 < d < 0.4999.

Typical values are 0.01, 0.02, 0.03, or 0.05.

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

© NCSS, LLC. All Rights Reserved.

Sample Size (when Solve For = Sample Size)

Cluster Allocation Pattern

Specify how the clusters are allocated to the individual strata during the search for the total number of clusters needed to assure that the half-width requirement is met.

The choices are

- **All Equal (Search for Clusters per Stratum, K0)**

  The smallest value of the number of clusters per stratum is found that still assures that the half-width requirement is met. The value of Kh is constant across all strata. That is, all Kh = K0.

- **Proportional (Enter Rh = Cluster Allocation Pattern and K)**

  Enter values that express the cluster allocation pattern across strata in the 'Custom Strata Information' section below. Search for a value for K.

  The Rh values will be scaled as proportions which give the proportion of the clusters allocated to the corresponding cluster. The formula is sRh = Rh/ΣRh. The number of clusters is then calculated using Kh = K x sRh. Finally, Kh is rounded to an integer.

  The search for K continues until the minimum value is found that assures that the half-width requirement is met.

Sample Size (when Solve For = Half-Width of C.I.)

Cluster Allocation Pattern

Specify how the clusters are allocated to the individual strata.

The choices are

- **All Equal (Kh = K0)**

  Enter a value for the number of clusters (K0) to be allocated to each stratum.

- **Proportional (Enter Rh = Cluster Allocation Pattern and K)**

  Enter the total number of clusters in the trial, K, in the box below. Also enter values that express the cluster allocation pattern across strata in the 'Custom Strata Information' section below.

  The Rh values will be scaled as proportions which give the proportion of the clusters allocated to the corresponding stratum. The formula is sRh = Rh/ΣRh. The number of clusters is then calculated using Kh = K x sRh. Finally, Kh is rounded to an integer.

- **Custom Kh**

  Enter the number of clusters in each stratum in the 'Kh (Number of Clusters)' column in the 'Custom Strata Information' section below.

K0 (Clusters per Stratum)

Enter one or more values of K0, the number of clusters per stratum to be used for all strata. Thus, the total number of clusters in the study is K0 x (number of strata).

Range

K0 > 1.

You can enter a single value such as 10 or a series of values such as ‘10 25 50’ or ‘10 to 50 by 5’.

© NCSS, LLC. All Rights Reserved.

K (Total Number of Clusters)

Enter one or more values of K, the total number of clusters in the study. These clusters are allocated to individual clusters using Rh, the cluster allocation pattern, given in the 'Custom Strata Information' section.

K > number of strata + 1. You can enter a single value such as *100* or a series of values such as *50 100 200* or *100 to 500 by 100*.

Sample Size – Cluster Size

Mh (Average Cluster Size)

Specify how the average cluster size per stratum is to be entered.

Note that Mh, average cluster size, is the average number of subjects per cluster in stratum h.

The choices are

- **All Equal**

  Enter a single value to be used for all Mh. Hence, all Mh will be equal.

  Specify a value for Mh, average cluster size (average number of subjects per cluster in stratum h), to be used for all stratum.

  This value must be a positive number that is at least 1. It can be a decimal number such as '2.7'.

  You can enter a single value such as '5' or a list of values such as “10 50 100”. If a list is entered, a separate analysis will be conducted for each value.

- **Custom**

  Enter a unique value for each Mh in the Mh (Average Cluster Size) column of the 'Custom Strata Information' section below.

Adjust results for variable cluster sizes with a stratum

In most trials, the cluster sizes (number of subjects per cluster) vary from cluster to cluster. If this variation is ignored by considering only the average cluster size, the calculated number of clusters will underestimate the actual number of clusters that is needed. This can be corrected by adjusting for variation in the cluster sizes. This adjustment is based on the coefficient of variation of the cluster sizes (COV).

Check this option to enter the COV for each stratum, Ch.

Ch (COV of Cluster Sizes)

Specify how the COV of cluster sizes per stratum, Ch, is to be entered.

The choices are

- **All Equal**

  Enter a single value to be used for all Ch. Hence, all Ch will be equal.

  Enter the coefficient of variation of the cluster sizes (number of subjects). This value must be zero or a positive number. You can use a list of values such as “0.4 0.6 0.8”.

- **Custom**

  Enter a unique value for each Ch in the Ch (COV of Cluster Sizes) column of the 'Custom Strata Information' section below.

© NCSS, LLC. All Rights Reserved.

Coefficient of Variation

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.

Standard Deviation

The standard deviation, calculated by the sample formula (divide by M-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.

Proportion and Intracluster Correlation

Ph (Response Proportions)

Specify how the response proportion (probability) per stratum, Ph, is to be entered.

Note that in 'Ph', the 'P' represents *proportion* and 'h' designates the *stratum*.

The choices are

- **All Equal**

  Enter a single value to be used for all Ph. Hence, all Ph will be equal.

  Enter the average response proportion of the subjects in all strata. This is the probability that a subject has the event of interest.

  0 < Ph < 1.

  If you have no idea what the proportion is, you can enter 0.5 since this value will result in the largest sample size.

- **Custom**

  Enter a unique value for each Ph in the Ph (Response Proportion) column of the 'Custom Strata Information' section below.

ρ (Intracluster Correlation, ICC)

This is the value of the intracluster (or intraclass) 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.

The documentation presents details of estimating this value.

Range

Possible values are from 0 to just below 1. Typical values are between 0.0001 and 0.3.

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

© NCSS, LLC. All Rights Reserved.

Custom Strata Information

This section lets you enter settings for each of the *H* individual strata. Each line on the report represent one or more strata. You can save time by entering settings for groups of strata that will have identical parameter values.

Set

This is an identification number for a set of equal strata used on the reports.

Number of Strata

Specify the number of strata specified on this line. Usually, you will enter a “1” to specify a single stratum, or you will enter a “0” to ignore this line. However, this option lets you specify several strata that have the same parameter values.

The total number of strata is equal to the sum of these values.

Examples

0 which means ‘ignore this line’.

1 which means ‘one stratum defined by this line’.

2 which means ‘two strata defined by this line’.

Kh (Number of Clusters)

Enter a value for the number of clusters in stratum h.

Range

Kh ≥ 1. At least one stratum must have a Kh value greater than 1.

Rh (Cluster Allocation Pattern)

Enter an allocation ratio value for this stratum. This value represents the relative frequency of clusters in this stratum.

Rescaling

These values are rescaled so that they sum to one. The formula is sRh = Rh/ΣRh. The number of clusters is then calculated using Kh = K x sRh, where K is the total number of clusters in the study.

Note that this value applies to the number of CLUSTERS, not the number of SUBJECTS.

For example, if there are four strata, the following sets of Rh would result in identical stratum cluster proportions:

2, 4, 6, 8

10, 20, 30, 40

0.1, 0.2, 0.3, 0.4

The resulting cluster proportions are

0.1 in stratum 1.

0.2 in stratum 2.

0.3 in stratum 3.

0.4 in stratum 4.

Note

Only enter one number, even if there are more than one stratum being defined by the line.

© NCSS, LLC. All Rights Reserved.

Range

These values can be any positive values. The values will be rescaled so that the resulting proportions sum to 1.

Ph (Response Proportion)

Enter the average response proportion of the subjects in stratum *h*. This is the probability that a subject has the event of interest.

The range is 0 < Ph < 1.

If you have no idea what the proportion is, you can enter 0.5 since this value will result in the largest sample size.

Mh (Average Cluster Size)

Specify a value for Mh, the average cluster size (average number of subjects per cluster) in stratum *h*.

This value must be a positive number that is at least 1. It can be a decimal number such as '2.7'.

Ch (COV of Cluster Sizes)

Enter the coefficient of variation of the cluster sizes (number of subjects) in stratum h. This value must be zero or a positive number. It is used to find the standard deviation of the cluster sizes.

Coefficient of Variation

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.

Standard Deviation of Cluster Sizes

The standard deviation of the cluster sizes, calculated by the sample formula (divide by Kh - 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.

Show More Strata Sets

Check this box to show ten more Strata Information sets. If this option is not checked, any active strata sets (Number of Strata > 0) with set identification numbers > 5 will be ignored.

© NCSS, LLC. All Rights Reserved.

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 the response proportion for this disease is 0.67.

The confidence level is set to 0.95 and *d* is set to three values 0.02, 0.03, 0.04.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load this procedure. 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	**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**
Ph (Response Proportions)	**All Equal**
Ph for All Strata	**0.67**
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**

© NCSS, LLC. All Rights Reserved.

Annotated Output

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

Numeric Results

Numeric Results

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

C.I. Half-Width d	Total Number Subjects N	Total Number Clusters K	Average Clusters per Strata K0	Average Cluster Size	Average COV of Cluster Sizes	Average Proportion P	ICC ρ	Conf Level
0.0200	4930	91	22.75	54.0	0.4000	0.6700	0.0200	0.950
0.0297	2230	41	10.25	54.0	0.4000	0.6700	0.0200	0.950
0.0396	1260	23	5.75	54.0	0.4000	0.6700	0.0200	0.950

References

Xu, X., Zhu, H., and Ahn, C. 2019. 'Sample size considerations for stratified cluster randomization design with binary outcomes and varying cluster size', Statistics in Medicine, 38(18), pages 3395-3404.

Report Definitions

d is the half-width of the confidence interval of P. d = [UCL(P) - LCL(P)] / 2.
N is the total number of subjects.
K is the total number of clusters.
K0 is the average number of clusters per stratum.
The Average Cluster Size is the weighted average of the number of subjects per cluster.
The Average COV of Cluster Sizes is the weighted average COV of all clusters.
P is the weighted average of the strata proportions. The weights are proportional to the number of subjects.
ρ is the intracluster correlation coefficient (ICC) average across all strata.
Conf Level is the confidence level of the confidence interval for P.

Summary Statements

A confidence interval for P will be computed from a stratified cluster design, which allocates the 91 clusters among 4 strata resulting in an overall sample size of 4930. The average cluster size is 54.0. This scenario has a confidence interval half-width of 0.0200 when the confidence level is 0.950, the average COV of cluster sizes is 0.4000, the intracluster correlation coefficient is 0.0200, and the average response proportion is 0.6700.

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

Strata-Detail Report

Strata-Detail Report for Row 1

Strata h	Number Subjects Nh	Number Clusters Kh	Average Cluster Size Mh	COV of Cluster Sizes Ch	Prop of Total Subjects Fh	Prop of Total Clusters sRh	Response Proportion Ph
1	1200	15	80.0	0.4000	0.243	0.160	0.6700
2	1320	22	60.0	0.4000	0.268	0.240	0.6700
3	1250	25	50.0	0.4000	0.254	0.280	0.6700
4	1160	29	40.0	0.4000	0.235	0.320	0.6700

Strata-Detail Report for Row 2

Strata h	Number Subjects Nh	Number Clusters Kh	Average Cluster Size Mh	COV of Cluster Sizes Ch	Prop of Total Subjects Fh	Prop of Total Clusters sRh	Response Proportion Ph
1	560	7	80.0	0.4000	0.251	0.160	0.6700
2	600	10	60.0	0.4000	0.269	0.240	0.6700
3	550	11	50.0	0.4000	0.247	0.280	0.6700
4	520	13	40.0	0.4000	0.233	0.320	0.6700

© NCSS, LLC. All Rights Reserved.

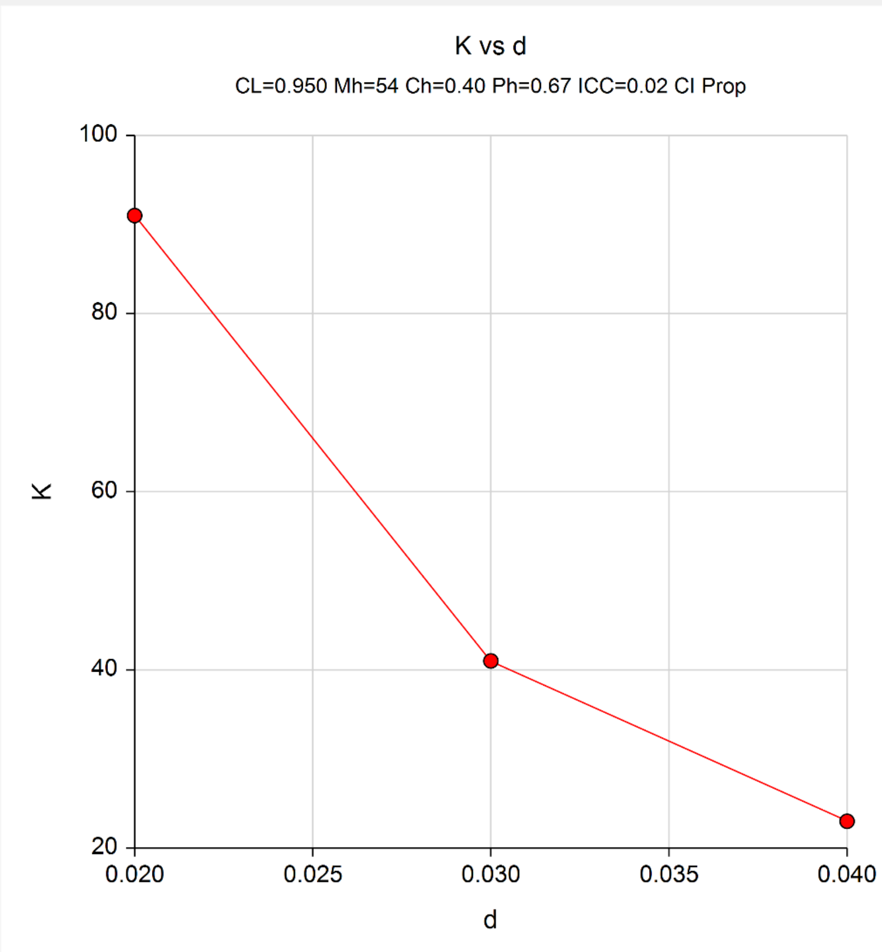
Strata-Detail Report for Row 3

Strata h	Number Subjects Nh	Number Clusters Kh	Average Cluster Size Mh	COV of Cluster Sizes Ch	Prop of Total Subjects Fh	Prop of Total Clusters sRh	Response Proportion Ph
1	320	4	80.0	0.4000	0.254	0.160	0.6700
2	360	6	60.0	0.4000	0.286	0.240	0.6700
3	300	6	50.0	0.4000	0.238	0.280	0.6700
4	280	7	40.0	0.4000	0.222	0.320	0.6700

Strata-Detail Report Definitions
h is an arbitrary sequence number for each stratum.
Nh is the number of subjects in stratum h. Nh = Kh x Mh.
Kh is the number of clusters in stratum h.
Mh is the average cluster size in stratum h.
Ch is the COV of the cluster sizes in stratum h.
Fh is the proportion of the total subjects in stratum h.
sRh is the proportion of the total clusters in stratum h.
Ph is the response proportion in stratum h of the event of interest.

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

Chart Section



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

© NCSS, LLC. All Rights Reserved.

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 response proportions in A and B are 0.4 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	Response Proportion Ph
A	10	20	0.4	0.4
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

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

$$= 3.22\left[\frac{1}{9}\left(\frac{0.4(0.6)}{200}\right) + \frac{4}{9}\left(\frac{0.5(0.5)}{400}\right)\right]$$

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

$$= 0.00132377778$$

Finally, the half-width is calculated as

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

$$= 1.95996398\sqrt{0.00132377778}$$

$$= 0.07131085$$

© NCSS, LLC. All Rights Reserved.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load this procedure. 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	**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**
Ph (Response Proportions)	**Custom**
ρ (Intracluster Correlation, ICC)	**0.1**
Set 1 Number of Strata	**1**
Set 1 Kh (Number of Clusters)	**10**
Set 1 Ph (Response Proportion)	**0.4**
Set 2 Number of Strata	**1**
Set 2 Kh (Number of Clusters)	**20**
Set 2 Ph (Response Proportion)	**0.5**

Output

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

Numeric Results

Numeric Results

Number of Strata: 2
Solve for: Half-Width
Allocation: Custom

C.I. Half-Width d	Total Number Subjects N	Total Number Clusters K	Average Clusters per Strata K0	Average Cluster Size	Average COV of Cluster Sizes	Average Proportion P	ICC ρ	Conf Level
0.0713	600	30	15.00	20.0	0.4000	0.4667	0.1000	0.950

Strata-Detail Report

Report Row: 1

Strata h	Number Subjects Nh	Number Clusters Kh	Average Cluster Size Mh	COV of Cluster Sizes Ch	Prop of Total Subjects Fh	Prop of Total Clusters sRh	Response Proportion Ph
1	200	10	20.0	0.4000	0.333	0.333	0.4000
2	400	20	20.0	0.4000	0.667	0.667	0.5000

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

© NCSS, LLC. All Rights Reserved.

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 response proportion for this disease is 0.67. 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

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

Option	Value
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**
Ph (Response Proportions)	**All Equal**
Ph for All Strata	**0.67**
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**

© NCSS, LLC. All Rights Reserved.

Output

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

Numeric Results

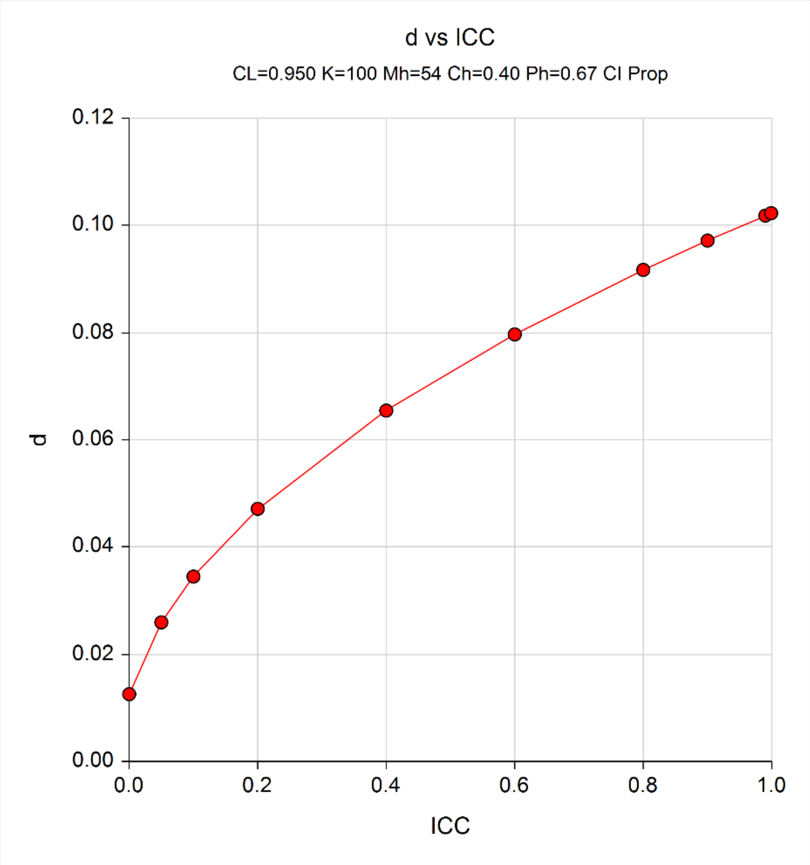
Numeric Results

Number of Strata: 4
Solve for: Half-Width
Allocation: Proportional

C.I. Half-Width d	Total Number Subjects N	Total Number Clusters K	Average Clusters per Strata K0	Average Cluster Size	Average COV of Cluster Sizes	Average Proportion P	ICC ρ	Conf Level
0.0125	5400	100	25.00	54.0	0.4000	0.6700	0.0000	0.950
0.0259	5400	100	25.00	54.0	0.4000	0.6700	0.0500	0.950
0.0345	5400	100	25.00	54.0	0.4000	0.6700	0.1000	0.950
0.0471	5400	100	25.00	54.0	0.4000	0.6700	0.2000	0.950
0.0655	5400	100	25.00	54.0	0.4000	0.6700	0.4000	0.950
0.0797	5400	100	25.00	54.0	0.4000	0.6700	0.6000	0.950
0.0917	5400	100	25.00	54.0	0.4000	0.6700	0.8000	0.950
0.0972	5400	100	25.00	54.0	0.4000	0.6700	0.9000	0.950
0.1018	5400	100	25.00	54.0	0.4000	0.6700	0.9900	0.950
0.1023	5400	100	25.00	54.0	0.4000	0.6700	0.9990	0.950

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

Chart Section



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

© NCSS, LLC. All Rights Reserved.

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 response proportion for this disease is 0.67. 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

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load this procedure. You may then make the appropriate entries as listed below, or open **Example 4** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
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**
Ph (Response Proportions)	**All Equal**
Ph for All Strata	**0.67**
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**

© NCSS, LLC. All Rights Reserved.

Output

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

Numeric Results

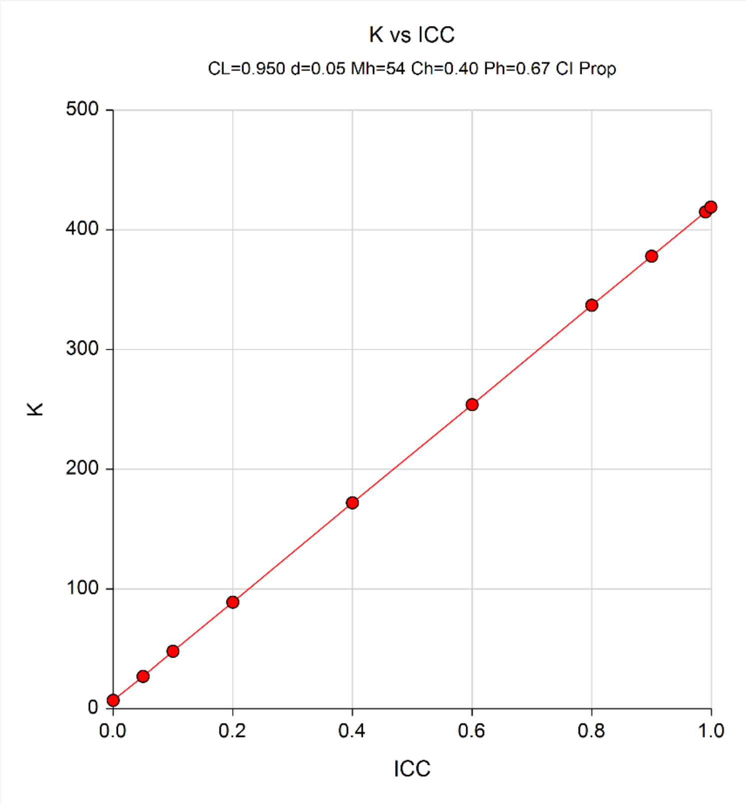
Numeric Results

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

C.I. Half-Width d	Total Number Subjects N	Total Number Clusters K	Average Clusters per Strata K0	Average Cluster Size	Average COV of Cluster Sizes	Average Proportion P	ICC ρ	Conf Level
0.0473	380	7	1.75	54.0	0.4000	0.6700	0.0000	0.950
0.0500	1440	27	6.75	54.0	0.4000	0.6700	0.0500	0.950
0.0498	2610	48	12.00	54.0	0.4000	0.6700	0.1000	0.950
0.0500	4790	89	22.25	54.0	0.4000	0.6700	0.2000	0.950
0.0499	9300	172	43.00	54.0	0.4000	0.6700	0.4000	0.950
0.0500	13730	254	63.50	54.0	0.4000	0.6700	0.6000	0.950
0.0500	18200	337	84.25	54.0	0.4000	0.6700	0.8000	0.950
0.0500	20400	378	94.50	54.0	0.4000	0.6700	0.9000	0.950
0.0500	22400	415	103.75	54.0	0.4000	0.6700	0.9900	0.950
0.0500	22630	419	104.75	54.0	0.4000	0.6700	0.9990	0.950

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

Chart Section



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.

© NCSS, LLC. All Rights Reserved.

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 response proportion for this disease is 0.67. 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

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load this procedure. You may then make the appropriate entries as listed below, or open **Example 5** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
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**
Ph (Response Proportions)	**All Equal**
Ph for All Strata	**0.67**
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**

© NCSS, LLC. All Rights Reserved.

Output

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

Numeric Results

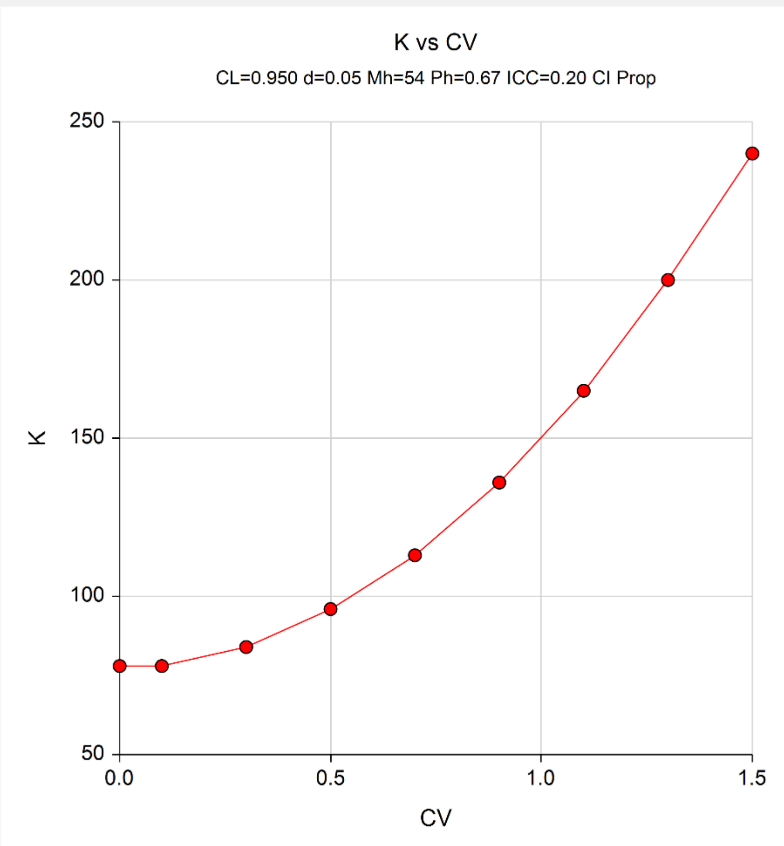
Numeric Results

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

C.I. Half-Width d	Total Number Subjects N	Total Number Clusters K	Average Clusters per Strata K0	Average Cluster Size	Average COV of Cluster Sizes	Average Proportion P	ICC ρ	Conf Level
0.0497	4200	78	19.50	54.0	0.0000	0.6700	0.2000	0.950
0.0500	4200	78	19.50	54.0	0.1000	0.6700	0.2000	0.950
0.0499	4520	84	21.00	54.0	0.3000	0.6700	0.2000	0.950
0.0498	5170	96	24.00	54.0	0.5000	0.6700	0.2000	0.950
0.0499	6100	113	28.25	54.0	0.7000	0.6700	0.2000	0.950
0.0500	7360	136	34.00	54.0	0.9000	0.6700	0.2000	0.950
0.0499	8900	165	41.25	54.0	1.1000	0.6700	0.2000	0.950
0.0499	10800	200	50.00	54.0	1.3000	0.6700	0.2000	0.950
0.0500	12950	240	60.00	54.0	1.5000	0.6700	0.2000	0.950

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.

Chart Section



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

© NCSS, LLC. All Rights Reserved.