Chapter 118

Confidence Intervals for One Mean in a Stratified Design

Introduction

This procedure calculates sample size and half-width for confidence intervals of a mean from a stratified design in which the outcome variable is continuous. It uses the results from elementary sampling theory which are presented in many works including Yamane (1967) and Levy and Lemeshow (2008).

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*. If these strata are created so that items are similar within a particular stratum, but quite different between strata, then a *stratified design* might be adopted for a number of reasons. Note that the population may be finite or infinite.

This procedure allows you to determine the appropriate sample size to be taken from each stratum so that various parameters of the confidence interval are guaranteed. These parameters include the confidence level and width of the interval.

Technical Details

The following discussion summarizes the results in Yamane (1967).

Suppose you are interested in estimating a certain outcome in a particular population. Further suppose that outcome 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.

Population Mean

In this design, assume that a simple random sample is drawn from each stratum. Let X_{hi} indicate the continuous outcome of the i^{th} subject in stratum h. Denote the total number of subjects in this stratum as N_h. Let the number of strata be denoted by L.

Let $\mu = \sum_{h=1}^{L} \sum_{i=1}^{N_h} \frac{X_{hi}}{N}$, where $N = \sum_{h=1}^{L} N_h$, represent the population mean that is to be estimated. This formula can be rearranged using strata proportions as follows.

$$\mu = \sum_{h=1}^{L} \frac{N_h}{N} \sum_{i=1}^{N_h} \frac{X_{hi}}{N_h} = \sum_{h=1}^{L} \frac{N_h}{N} \mu_h$$

where

$$\mu_h = \sum_{i=1}^{N_h} \frac{X_{hi}}{N_h}$$

is the mean within stratum *h*.

Sample Means

Let the size of the sample from stratum *h* be n_h. The sample mean is estimated as follows.

$$\bar{x} = \sum_{h=1}^{L} \frac{N_h}{N} \sum_{i=1}^{n_h} \frac{X_{hi}}{n_h} = \sum_{h=1}^{L} \frac{N_h}{N} \bar{x}_h$$

where

$$\bar{x}_h = \sum_{i=1}^{n_h} \frac{X_{hi}}{n_h}$$

is the sample mean within stratum *h*. Thus, $\bar{x}$ estimates μ.

It can be shown (Yamane 1967 page 115) that the expected value and variance of $\bar{x}$ assuming without replacement sampling are

$$E(\bar{x}) = \mu$$

$$V(\bar{x}) = \sum_{h=1}^{L} \left(\frac{N_h}{N}\right)^2 \left(\frac{N_h - n_n}{N_h n_h}\right) S_h^2$$

where

$$S_h^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (X_{hi} - \mu_h)^2$$

Note that the stratum variance is given by

$$\sigma_h^2 = \frac{1}{N_h} \sum_{i=1}^{N_h} (X_{hi} - \mu_h)^2$$

Usually the value of N_h will be large enough so that $S_h^2 = \sigma_h^2$ for all intents and purposes.

An unbiased estimator of $V(\bar{x})$ is

$$\hat{V}(\bar{x}) = \sum_{h=1}^{L} \left(\frac{N_h}{N}\right)^2 \left(\frac{N_h - n_n}{N_h n_h}\right) s_h^2$$

where

$$s_h^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (X_{hi} - \bar{x}_h)^2$$

Using the central limit theorem, it is known that $\bar{x}$ is asymptotically standard normal. Therefore, a confidence interval for μ can be constructed as follows.

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

The lower and upper limits of this confidence interval are denoted as LCL_μ and UCL_μ.

Sample Size Estimation

Equal Allocation

Equal allocation assumes that the overall sample size is allocated across strata using $n_h = \frac{n}{L}$. Using this allocation method, the overall sample size is estimated as follows.

$$n = \frac{L\sum_{h=1}^{L} N_h^2 S_h^2}{N^2 D^2 + \sum_{h=1}^{L} N_h S_h^2}$$

where $D = d/z_{1-\frac{\alpha}{2}}$ and $d = \left(UCL_\mu - LCL_\mu\right)/2$ which is the *half width* of the confidence interval.

Proportional Allocation

Proportional allocation assumes that the overall sample size is allocated across strata using $n_h = \frac{N_h}{N}n$. Using this allocation method, the overall sample size is estimated as follows.

$$n = \frac{N\sum_{h=1}^{L} N_h S_h^2}{N^2 D^2 + \sum_{h=1}^{L} N_h S_h^2}$$

where $D = d/z_{1-\frac{\alpha}{2}}$ and $d = \left(UCL_\mu - LCL_\mu\right)/2$ which is the *half width* of the confidence interval.

Optimum Allocation

Optimum allocation assumes that the overall sample size is allocated across strata using

$$n_h = n\left(\frac{N_h S_h^2}{\sum_{h=1}^{L} N_h S_h^2}\right)$$

Using this allocation method, the overall sample size is estimated as follows.

$$n = \frac{\left(\sum_{h=1}^{L} N_h S_h\right)^2}{N^2 D^2 + \sum_{h=1}^{L} N_h S_h^2}$$

where $D = d/z_{1-\frac{\alpha}{2}}$ and $d = \left(UCL_\mu - LCL_\mu\right)/2$ which is the *half width* of the confidence interval.

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 calculate the sample size needed. Select *Half-Width of C.I.* when you want to investigate the precision of a certain sample size.

Confidence

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 mean.

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.

Precision

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 mean. It is also the distance from the mean to either confidence limit.

The formula is $d = |UCL(\mu) - LCL(\mu)|/2$.

The range is $0 < d$.

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

Sample Size (when Solve For = Sample Size)

Sample Size Allocation

Specify the way the total sample size is to be specified and then allocated to the individual strata. The choices are

- **Proportional**

  The program will determine the overall sample size using the formula given in the help file. This value will be allocated to the individual strata so that the strata sample size proportions match the strata population size proportions (within rounding).

- **Optimal**

  The program will determine the overall sample size using the formula given in the help file. This value will be allocated to the individual strata so that the width of the confidence interval of the proportion is as narrow as possible.

- **Equal**

  The program will determine the overall sample size using equal strata sample sizes. This value will be just large enough that the half-width requirement is met.

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

Sample Size Allocation

Specify the way the total sample size is to be entered and then allocated to the individual strata. The choices are

- **Proportional**

  Enter the total sample, n, in the box below. This value will be allocated to the individual strata so that the strata sample size proportions match the strata population size proportions (within rounding).

- **Optimal**

  Enter the total sample, n, in the box below. This value will be allocated to the individual strata so that the width of the confidence interval of the proportion is as narrow as possible.

- **Equal**

  Enter a single strata size that will be used for all strata.

- **Custom**

  Enter custom stratum sample sizes in the 'Stratum Sample Size' column in the 'Strata Information' section below.

n (Total Sample Size)

Enter one or more values of *n*, the total number of items in the sample. This parameter is the sum of the subjects in all strata. These subjects are then divided among the individual strata.

The total sample size you enter must be large enough so that there is at least one sample item available for each stratum. Also, the total sample size must be smaller than the total population size.

nh (Strata Sample Size)

Enter a single stratum sample size that is to be used for all strata. The strata sample sizes will all be equal.

The value must be an integer greater than zero and less than the smallest strata population size.

Strata Information

This section lets you enter the settings for each of the *L* individual strata.

Set

This is an identification number 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'.

Stratum Population Size

Enter the total population sizes of each of the strata in this column. If this line defines more than one stratum, this is the amount used FOR EACH stratum.

These values can be any positive integers greater than zero.

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

If the stratum population size very large and unknown, just enter a large value such as 100000.

Stratum Std Deviation

Enter the standard deviation for this stratum. This value may come from previous studies or preliminary results.

The range is any value greater than zero.

Stratum Sample Size

Enter a custom sample size the strata defined on this line.

This value must be an integer greater than zero and less than the corresponding stratum population size.

Show More Strata Sets

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

Example 1 – Finding Sample Size with Proportional Allocation

A study using a stratified design is being planned to estimate the effectiveness of a certain drug in treating a certain disease. Since age is known to affect the disease rates, the population is stratified into four age groups. The sizes of these four age groups are 14000, 18000, 6000, and 10000. The overall sample size will be allocated across strata proportional to the strata population size.

Prior studies have shown stratum standard deviations of 10, 15, 20, and 20, respectively, among the four age groups.

The confidence level is set to 0.95 and *d* is set to three values 1, 3, and 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 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)	**1 3 5**
Sample Size Allocation	**Proportional**
Set 1 Number of Strata	**1**
Set 1 Stratum Population Size	**14000**
Set 1 Stratum Standard Deviation	**10**
Set 2 Number of Strata	**1**
Set 2 Stratum Population Size	**18000**
Set 2 Stratum Standard Deviation	**15**
Set 3 Number of Strata	**1**
Set 3 Stratum Population Size	**6000**
Set 3 Stratum Standard Deviation	**20**
Set 4 Number of Strata	**1**
Set 4 Stratum Population Size	**10000**
Set 4 Stratum Standard Deviation	**30**
Set 5 Number of Strata	**0**
Show More Strata Sets	**Unchecked**

Annotated Output

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

Numeric Results

Numeric Results

Number of Strata: 4
Population Size (N): 48000
Solve for: Sample Size
Allocation: Proportional

Actual C.I. Half-Width d(A)	Target C.I. Half-Width d(T)	Sample Size n	Sampling Fraction f	Standard Deviation of the Est Mean SE	Confidence Level
1.0002	1.0000	1312	0.0273	0.5103	0.950
2.9982	3.0000	150	0.0031	1.5297	0.950
5.0163	5.0000	54	0.0011	2.5594	0.950

References

Yamane, Taro. 1967. Elementary Sampling Theory. Prentice-Hall, Inc. Englewood Cliffs, New Jersey.
Levy, P.S. and Lemeshow, S. 2008. Sampling of Populations. Fourth Edition. John Wiley & Sons. New York.
Cochran, William G. 1977. Sampling Techniques. Third Edition. John Wiley & Sons. New York.

Report Definitions

d(A) is the actual half-width of the confidence interval of μ. d(A) = [UCL(μ) - LCL(μ)] / 2.
d(T) is the target half-width of the confidence interval of μ. It may be slightly different from d(A) because of rounding.
n is the total sample size, i.e., the total number of subjects summed across all strata.
f is the sampling fraction. This is n/N.
SE is the standard error of the estimated mean.
Confidence Level is the confidence level of the confidence interval for μ.

Summary Statements

A confidence interval for μ will be computed from a stratified design, which divides the sample among 4 strata. A sample of 1312 subjects is obtained from the 48000 subjects in the population. This scenario has a confidence interval half-width of 1.0002 when the confidence level is 0.950 and the standard error of the mean is 0.5103.

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

Strata-Detail Report

Strata-Detail Report for Row 1

Strata h	Pop Size Nh	Percent of Pop Size Pct(Nh)	Sample Size nh	Percent of Sample Size Pct(nh)	Std Dev Sdh
1	14000	29.2	383	29.2	10.0000
2	18000	37.5	492	37.5	15.0000
3	6000	12.5	164	12.5	20.0000
4	10000	20.8	273	20.8	30.0000

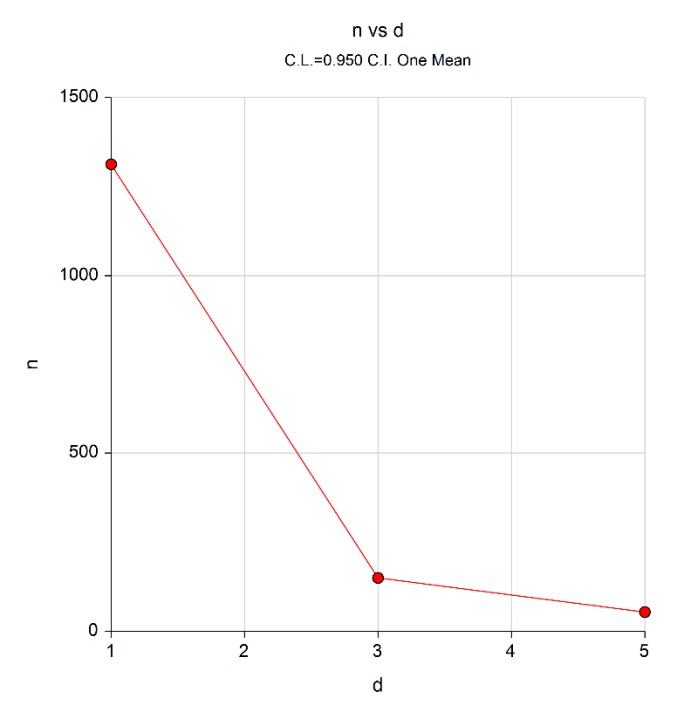
(report continues)

Strata-Detail Report Definitions

Strata h is an arbitrary sequence number for each stratum.
Pop Size Nh is the population size of stratum h.
Pct(Nh) is the percentage of the population size that is comes from this stratum.
Sample Size nh is the sample size of stratum h.
Pct(nh) is the percentage of the total sample size that comes from this stratum.
Sdh is the response standard deviation in stratum h of the event of interest.

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

Plots Section



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

Example 2 – Validation using Yamane (1967)

Yamane (1967) page 141-142 provides an example of a stratified design that we will use to validate this procedure.

A study using a stratified design is being planned to estimate the mean number of customers per day per restaurant in a city that is stratified according to the restaurant size: small, medium, and large . These restaurant sizes will form the strata. The number of restaurants in the three strata are 600 small, 300 medium, and 100 large. Previous studies place the estimated standard deviations at 20, 30, and 50. The value of *d* is 3. The confidence level is set to 0.9973 which results in a z value of 2.99999.

They calculated the sample size assuming proportional allocation as 432. The individual strata sizes should be 259, 130, and 43.

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	**Sample Size**
Confidence Level	**0.9973**
d (Precision, Half-Width)	**3**
Sample Size Allocation	**Proportional**
Set 1 Number of Strata	**1**
Set 1 Stratum Population Size	**600**
Set 1 Stratum Standard Deviation	**20**
Set 2 Number of Strata	**1**
Set 2 Stratum Population Size	**300**
Set 2 Stratum Standard Deviation	**30**
Set 3 Number of Strata	**1**
Set 3 Stratum Population Size	**100**
Set 3 Stratum Standard Deviation	**50**
Set 4 Number of Strata	**0**
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 Results

Numeric Results

Number of Strata:	3
Population Size (N):	1000
Solve for:	Sample Size
Allocation:	Proportional

Actual C.I. Half-Width d(A)	Target C.I. Half-Width d(T)	Sample Size n	Sampling Fraction f	Standard Deviation of the Est Mean SE	Confidence Level
3.0007	3.0000	**432**	0.4320	1.0002	0.997

This report shows that PASS also obtains an *n* of 432 which validates the procedure.

Strata-Detail Report

Strata-Detail Report

Strata h	Pop Size Nh	Percent of Pop Size Pct(Nh)	Sample Size nh	Percent of Sample Size Pct(nh)	Std Dev Sdh
1	600	60.0	**259**	60.0	20.0000
2	300	30.0	**130**	30.1	30.0000
3	100	10.0	**43**	10.0	50.0000

This report shows the values of the individual, strata-level parameters. They also match those given in Yamane (1967).

Example 3 – Finding Sample Size with Optimal Allocation

This example will rerun Example 1 with the Sample Size Allocation set to *Optimal*.

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	**Sample Size**
Confidence Level	**0.95**
d (Precision, Half-Width)	**1 3 5**
Sample Size Allocation	**Optimal**
Set 1 Number of Strata	**1**
Set 1 Stratum Population Size	**14000**
Set 1 Stratum Standard Deviation	**10**
Set 2 Number of Strata	**1**
Set 2 Stratum Population Size	**18000**
Set 2 Stratum Standard Deviation	**15**
Set 3 Number of Strata	**1**
Set 3 Stratum Population Size	**6000**
Set 3 Stratum Standard Deviation	**20**
Set 4 Number of Strata	**1**
Set 4 Stratum Population Size	**10000**
Set 4 Stratum Standard Deviation	**30**
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 Results

Numeric Results

Number of Strata:	4
Population Size (N):	48000
Solve for:	Sample Size
Allocation:	Optimal

Actual C.I. Half-Width d(A)	Target C.I. Half-Width d(T)	Sample Size n	Sampling Fraction f	Standard Deviation of the Est Mean SE	Confidence Level
0.9996	1.0000	1118	0.0233	0.5100	0.950
2.9912	3.0000	128	0.0027	1.5261	0.950
4.9968	5.0000	46	0.0010	2.5494	0.950

For proportional allocation, the sample sizes are: 1312, 150, 54.
For optimal allocation, the sample sizes are: 1118, 120, 46.
For equal allocation, the sample sizes are: 1280, 148, 56.

Example 4 – Finding Sample Size with Equal Allocation

This example will rerun Example 1 with the Sample Size Allocation set to *Equal*.

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)	**1 3 5**
Sample Size Allocation	**Equal**
Set 1 Number of Strata	**1**
Set 1 Stratum Population Size	**14000**
Set 1 Stratum Standard Deviation	**10**
Set 2 Number of Strata	**1**
Set 2 Stratum Population Size	**18000**
Set 2 Stratum Standard Deviation	**15**
Set 3 Number of Strata	**1**
Set 3 Stratum Population Size	**6000**
Set 3 Stratum Standard Deviation	**20**
Set 4 Number of Strata	**1**
Set 4 Stratum Population Size	**10000**
Set 4 Stratum Standard Deviation	**30**
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 Results

Numeric Results

Number of Strata:	4
Population Size (N):	48000
Solve for:	Sample Size
Allocation:	Equal

Actual C.I. Half-Width d(A)	Target C.I. Half-Width d(T)	Sample Size n	Sampling Fraction f	Standard Deviation of the Est Mean SE	Confidence Level
0.9989	1.0000	1280	0.0267	0.5097	0.950
2.9740	3.0000	148	0.0031	1.5174	0.950
4.8396	5.0000	56	0.0012	2.4692	0.950

For proportional allocation, the sample sizes are: 1312, 150, 54.
For optimal allocation, the sample sizes are: 1118, 120, 46.
For equal allocation, the sample sizes are: 1280, 148, 56.

Example 5 – Finding C.I. Half-Width

A study using a stratified design is being conducted. The sizes of the three strata are 14257, 18632, and 10908. The sample sizes drawn from the three strata are 215, 269, and 193. The standard deviations are expected to be 10, 15, and 20. The half-widths of the confidence interval is desired at confidence levels of 0.95 and 0.99.

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	**Half-Width of C.I.**
Confidence Level	**0.95 0.99**
Sample Size Allocation	**Custom**
Set 1 Number of Strata	**1**
Set 1 Stratum Population Size	**14257**
Set 1 Stratum Standard Deviation	**10**
Set 1 Stratum Sample Size	**215**
Set 2 Number of Strata	**1**
Set 2 Stratum Population Size	**18632**
Set 2 Stratum Standard Deviation	**15**
Set 2 Stratum Sample Size	**269**
Set 3 Number of Strata	**1**
Set 3 Stratum Population Size	**10908**
Set 3 Stratum Standard Deviation	**20**
Set 3 Stratum Sample Size	**193**
Set 4 Number of Strata	**0**
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 Results

Numeric Results

Number of Strata:	3
Population Size (N):	43797
Solve for:	Half-Width
Allocation:	Custom

C.I. Half-Width d	Sample Size n	Sampling Fraction f	Standard Deviation of the Est Mean SE	Confidence Level
1.1157	677	0.0155	0.5692	0.950
1.4662	677	0.0155	0.5692	0.990

Strata-Detail Report for Row 1

Strata h	Pop Size Nh	Percent of Pop Size Pct(Nh)	Sample Size nh	Percent of Sample Size Pct(nh)	Std Dev Sdh
1	14257	32.6	215	31.8	10.0000
2	18632	42.5	269	39.7	15.0000
3	10908	24.9	193	28.5	20.0000

Strata-Detail Report for Row 2

Strata h	Pop Size Nh	Percent of Pop Size Pct(Nh)	Sample Size nh	Percent of Sample Size Pct(nh)	Std Dev Sdh
1	14257	32.6	215	31.8	10.0000
2	18632	42.5	269	39.7	15.0000
3	10908	24.9	193	28.5	20.0000

Note that increasing the confidence level from 0.95 to 0.99 has increased the half width from 1.1157 to 1.4662.