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.

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

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) **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**

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
Population Size (N): 48000
Allocation: Proportional

Confidence Interval Half-Width		Total Sample Size	Sampling Fraction	Standard Error of the Estimated Mean	Confidence Level
Actual d(A)	Target d(T)	n	f	SE	
1.0002	1	1312	0.0273	0.5103	0.95
2.9982	3	150	0.0031	1.5297	0.95
5.0163	5	54	0.0011	2.5594	0.95

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

Summary Statements

A stratified design with 4 strata will be used to obtain a two-sided 95% confidence interval for a single mean. The standard error of the mean, based on the assumed stratum standard deviations, is assumed to be 0.5103. From a combined-strata population of 48000 subjects, to produce a confidence interval with a half-width of no more than 1, a combined sample size (across the 4 strata) of 1312 subjects will be needed.

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

Strata-Detail Report

Strata-Detail Report for Row 1

Stratum	Population Size		Sample Size		Standard Deviation
h	Value Nh	Percent Pct(Nh)	Value nh	Percent Pct(nh)	SDh
1	14000	29.2	383	29.2	10
2	18000	37.5	492	37.5	15
3	6000	12.5	164	12.5	20
4	10000	20.8	273	20.8	30

Strata-Detail Report for Row 2

	Population Size		Sample Size		
Stratum h	Value Nh	Percent Pct(Nh)	Value nh	Percent Pct(nh)	Standard Deviation SDh
1	14000	29.2	44	29.3	10
2	18000	37.5	56	37.3	15
3	6000	12.5	19	12.7	20
4	10000	20.8	31	20.7	30

Strata-Detail Report for Row 3

	Population Size		Sample Size		
Stratum h	Value Nh	Percent Pct(Nh)	Value nh	Percent Pct(nh)	Standard Deviation SDh
1	14000	29.2	16	29.6	10
2	18000	37.5	20	37.0	15
3	6000	12.5	7	13.0	20
4	10000	20.8	11	20.4	30

Strata h An arbitrary sequence number for each stratum.
Nh The population size of stratum h.
Pct(Nh) The percentage of the population size that is comes from stratum h.
nh The sample size for stratum h.
Pct(nh) The percentage of the total sample size that comes from stratum h.
SDh The response standard deviation of the event of interest in stratum h.

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

Dropout and References

Dropout-Inflated Sample Size

Dropout Rate	Sample Size n	Dropout-Inflated Enrollment Sample Size n'	Expected Number of Dropouts D
20%	1312	1640	328
20%	150	188	38
20%	54	68	14

Dropout Rate The percentage of subjects (or items) that are expected to be lost at random during the course of the study and for whom no response data will be collected (i.e., will be treated as "missing"). Abbreviated as DR.
n The evaluable sample size at which power is computed. If n subjects are evaluated out of the n' subjects that are enrolled in the study, the design will achieve the stated power.
n' The total number of subjects that should be enrolled in the study in order to obtain n evaluable subjects, based on the assumed dropout rate. After solving for n, n' is calculated by inflating n using the formula n' = n / (1 - DR), with n' always rounded up. (See Julious, S.A. (2010) pages 52-53, or Chow, S.C., Shao, J., Wang, H., and Lokhnygina, Y. (2018) pages 32-33.)
D The expected number of dropouts. D = n' - n.

Dropout Summary Statements

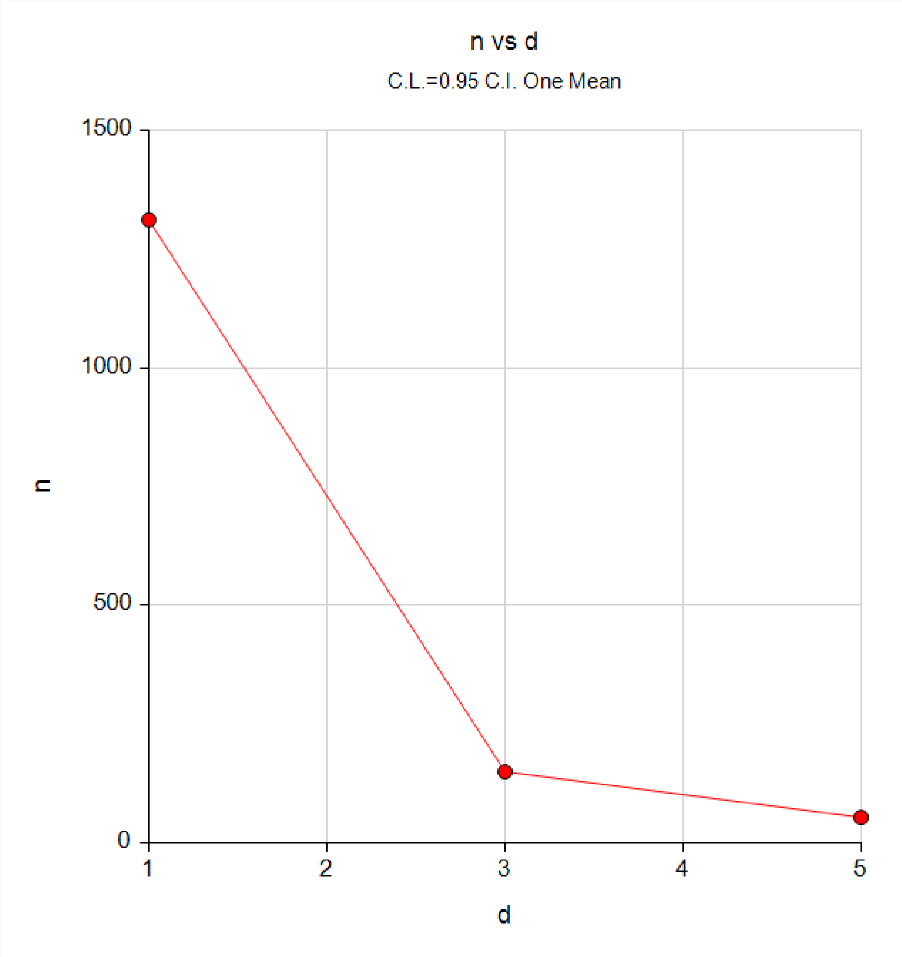
Anticipating a 20% dropout rate, 1640 subjects should be enrolled to obtain a final sample size of 1312 subjects.

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.

Plots Section

Plots



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

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 .... **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 Reports

Numeric Results

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

Confidence Interval Half-Width					
Actual d(A)	Target d(T)	Total Sample Size n	Sampling Fraction f	Standard Error of the Estimated Mean SE	Confidence Level
3.0007	3	432	0.432	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

	Population Size		Sample Size		
Stratum h	Value Nh	Percent Pct(Nh)	Value nh	Percent Pct(nh)	Standard Deviation SDh
1	600	60	259	60.0	20
2	300	30	130	30.1	30
3	100	10	43	10.0	50

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

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 .... **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

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

Confidence Interval Half-Width		Total Sample Size	Sampling Fraction	Standard Error of the Estimated Mean	Confidence Level
Actual d(A)	Target d(T)	n	f	SE	
0.9996	1	1118	0.0233	0.5100	0.95
2.9912	3	128	0.0027	1.5261	0.95
4.9968	5	46	0.0010	2.5494	0.95

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

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) **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

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

Confidence Interval Half-Width Actual d(A)	Confidence Interval Half-Width Target d(T)	Total Sample Size n	Sampling Fraction f	Standard Error of the Estimated Mean SE	Confidence Level
0.9989	1	1280	0.0267	0.5097	0.95
2.9740	3	148	0.0031	1.5174	0.95
4.8396	5	56	0.0012	2.4692	0.95

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

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 ....**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

Solve For:	Half-Width of C.I.
Number of Strata:	3
Population Size (N):	43797
Allocation:	Custom

Confidence Interval Half-Width d	Total Sample Size n	Sampling Fraction f	Standard Error of the Estimated Mean SE	Confidence Level
1.1157	677	0.0155	0.5692	0.95
1.4662	677	0.0155	0.5692	0.99

Strata-Detail Report for Row 1

	Population Size		Sample Size		
Stratum h	Value Nh	Percent Pct(Nh)	Value nh	Percent Pct(nh)	Standard Deviation SDh
1	14257	32.6	215	31.8	10
2	18632	42.5	269	39.7	15
3	10908	24.9	193	28.5	20

Strata-Detail Report for Row 2

	Population Size		Sample Size		
Stratum h	Value Nh	Percent Pct(Nh)	Value nh	Percent Pct(nh)	Standard Deviation SDh
1	14257	32.6	215	31.8	10
2	18632	42.5	269	39.7	15
3	10908	24.9	193	28.5	20

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