

Chapter 117

Confidence Intervals for One Proportion in a Stratified Design

Introduction

This procedure calculates sample size and half-width for confidence intervals of a proportion from a stratified design in which the outcome variable is binary. 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 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*. 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 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.

Population Proportions

In this design, assume that a simple random sample is drawn from each stratum. Let Y_{hi} indicate the binary outcome (0 or 1) 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 $P = \sum_{h=1}^L \sum_{i=1}^{N_h} \frac{Y_{hi}}{N}$, where $N = \sum_{h=1}^L N_h$, represent the population proportion that is to be estimated. This formula can be rearranged using strata proportions as follows.

$$P = \sum_{h=1}^L \frac{N_h}{N} \sum_{i=1}^{N_h} \frac{Y_{hi}}{N_h} = \sum_{h=1}^L \frac{N_h}{N} P_h$$

Confidence Intervals for One Proportion in a Stratified Design

where

$$P_h = \sum_{i=1}^{N_h} \frac{Y_{hi}}{N_h}$$

is the response proportion within stratum h .

Sample Proportions

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

$$p = \sum_{h=1}^L \frac{N_h}{N} \sum_{i=1}^{n_h} \frac{Y_{hi}}{n_h} = \sum_{h=1}^L \frac{N_h}{N} p_h$$

where

$$p_h = \sum_{i=1}^{n_h} \frac{Y_{hi}}{n_h}$$

is the sample response proportion within stratum h . Thus, p estimates P .

It can be shown that the expected value and variance of p are as follows.

$$E(p) = P$$

$$V(p) = \sum_{h=1}^L \left(\frac{N_h}{N} \right)^2 \left(\frac{N_h - n_h}{N_h n_h} \right) \left(\frac{N_h}{N_h - 1} \right) p_h (1 - p_h)$$

An unbiased estimator of $V(p)$ is

$$\hat{V}(p) = \sum_{h=1}^L \left(\frac{N_h}{N} \right)^2 \left(\frac{N_h - n_h}{N_h n_h} \right) \left(\frac{N_h}{N_h - 1} \right) p_h (1 - p_h)$$

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

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

The lower and upper limits of this confidence interval are denoted as LCL_P and UCL_P .

Sample Size Estimation

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.

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

where $D = d/z_{1-\alpha/2}$ and $d = (UCL_P - LCL_P)/2$ which is the *half width* of the confidence interval.

Optimum Allocation

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

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

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

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

where $D = d/z_{1-\alpha/2}$ and $d = (UCL_P - LCL_P)/2$ 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 had shown the probability of this disease continuing after receiving the baseline treatment was 0.25, 0.20, 0.15, and 0.10, respectively, among the four age groups.

The confidence level is set to 0.95 and d is set to three values 0.01, 0.03, 0.05.

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)	0.01 0.03 0.05
Sample Size Allocation	Proportional
Set 1 Number of Strata	1
Set 1 Stratum Population Size	14000
Set 1 Stratum Proportion	0.25
Set 2 Number of Strata	1
Set 2 Stratum Population Size	18000
Set 2 Stratum Proportion	0.2
Set 3 Number of Strata	1
Set 3 Stratum Population Size	6000
Set 3 Stratum Proportion	0.15
Set 4 Number of Strata	1
Set 4 Stratum Population Size	10000
Set 4 Stratum Proportion	0.10
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 n	Proportion P	Confidence Interval Limits		Standard Error of P SE(P)	Confidence Level
Actual d(A)	Target d(T)			Lower LCL(P)	Upper UCL(P)		
0.0100	0.01	5126	0.1875	0.1775	0.1975	0.0051	0.95
0.0300	0.03	630	0.1875	0.1575	0.2175	0.0153	0.95
0.0499	0.05	229	0.1875	0.1376	0.2374	0.0255	0.95

d(A) The actual half-width of the confidence interval of P. $d(A) = [UCL(p) - LCL(p)] / 2$.
 d(T) The target half-width of the confidence interval of P. 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.
 P The population proportion based on all strata. This is the weighted average of the individual strata proportions.
 LCL(P) The lower limit of a confidence interval for P.
 UCL(P) The upper limit of a confidence interval for P.
 SE(P) The standard error of the estimate of P.
 Confidence Level The confidence level of the confidence interval for P.

Summary Statements

A stratified design with 4 strata will be used to obtain a two-sided 95% confidence interval for a single proportion. The average response proportion is assumed to be 0.1875. From a combined-strata population of 48000 subjects, to produce a confidence interval with a half-width of no more than 0.01, a combined sample size (across the 4 strata) of 5126 subjects will be needed.

This report gives the results for each of the three values of p .

Strata-Detail Report

Strata-Detail Report for Row 1

Stratum h	Population Size		Sample Size		Proportion Ph
	Value Nh	Percent Pct(Nh)	Value nh	Percent Pct(nh)	
1	14000	29.2	1495	29.2	0.25
2	18000	37.5	1922	37.5	0.20
3	6000	12.5	641	12.5	0.15
4	10000	20.8	1068	20.8	0.10

Confidence Intervals for One Proportion in a Stratified Design

Strata-Detail Report for Row 2

Stratum h	Population Size		Sample Size		Proportion Ph
	Value Nh	Percent Pct(Nh)	Value nh	Percent Pct(nh)	
1	14000	29.2	184	29.2	0.25
2	18000	37.5	236	37.5	0.20
3	6000	12.5	79	12.5	0.15
4	10000	20.8	131	20.8	0.10

Strata-Detail Report for Row 3

Stratum h	Population Size		Sample Size		Proportion Ph
	Value Nh	Percent Pct(Nh)	Value nh	Percent Pct(nh)	
1	14000	29.2	67	29.3	0.25
2	18000	37.5	86	37.6	0.20
3	6000	12.5	28	12.2	0.15
4	10000	20.8	48	21.0	0.10

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

Dropout-Inflated Sample Size

Dropout Rate	Sample Size n	Dropout- Inflated Enrollment Sample Size n'	Expected Number of Dropouts D
20%	5126	6408	1282
20%	630	788	158
20%	229	287	58

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

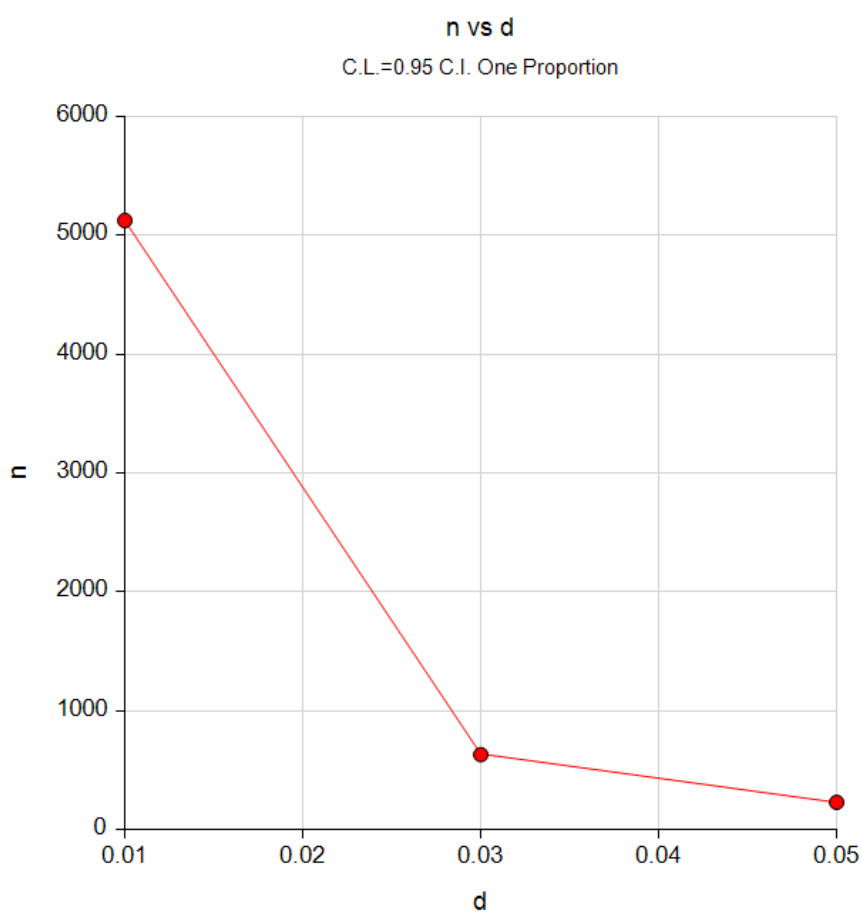
Anticipating a 20% dropout rate, 6408 subjects should be enrolled to obtain a final sample size of 5126 subjects.

Confidence Intervals for One Proportion in a Stratified Design

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.

This report shows the values of the individual, strata-level parameters, as well as the dropout and references sections.

Plots Section**Plots**

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

Example 2 – Validation using Yamane (1967)

Yamane (1967) page 155 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 proportion of families with two or more TV sets in a population comprised of three cities. These cities will form the strata. The number of families in the three cities is 2000, 3000, and 5000. Previous studies place the estimated proportions within the cities at 0.10, 0.15, and 0.20. The value of d is 0.03. The confidence level is set to 0.9973 which results in a z value of 2.99999.

They calculated the sample size as 1210. However, we found that this is a typo. It is easy to see that the sample size should be 1200. The individual strata sizes should be 240, 360, and 600.

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)	0.03
Sample Size Allocation	Proportional
Set 1 Number of Strata	1
Set 1 Stratum Population Size	2000
Set 1 Stratum Proportion	0.1
Set 2 Number of Strata	1
Set 2 Stratum Population Size	3000
Set 2 Stratum Proportion	0.15
Set 3 Number of Strata	1
Set 3 Stratum Population Size	5000
Set 3 Stratum Proportion	0.2
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): 10000
 Allocation: Proportional

Confidence Interval Half-Width		Total Sample Size n	Proportion P	Confidence Interval Limits		Standard Error of P SE(P)	Confidence Level
Actual d(A)	Target d(T)			Lower LCL(P)	Upper UCL(P)		
0.03	0.03	1200	0.165	0.135	0.195	0.01	0.9973

This report shows that **PASS** also obtains an n of 1200 which validates the procedure.

Strata-Detail Report

Strata-Detail Report

Stratum h	Population Size		Sample Size		Proportion Ph
	Value Nh	Percent Pct(Nh)	Value nh	Percent Pct(nh)	
1	2000	20	240	20	0.10
2	3000	30	360	30	0.15
3	5000	50	600	50	0.20

This report shows the values of the individual, strata-level parameters. They match those given above.

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) **0.01 0.03 0.05**
 Sample Size Allocation **Optimal**

Set 1 Number of Strata **1**
 Set 1 Stratum Population Size **14000**
 Set 1 Stratum Proportion **0.25**

Set 2 Number of Strata **1**
 Set 2 Stratum Population Size **18000**
 Set 2 Stratum Proportion **0.2**

Set 3 Number of Strata **1**
 Set 3 Stratum Population Size **6000**
 Set 3 Stratum Proportion **0.15**

Set 4 Number of Strata **1**
 Set 4 Stratum Population Size **10000**
 Set 4 Stratum Proportion **0.10**

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

Confidence Intervals for One Proportion in a Stratified Design

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 n	Proportion P	Confidence Interval Limits		Standard Error of P SE(P)	Confidence Level
Actual d(A)	Target d(T)			Lower LCL(P)	Upper UCL(P)		
0.01	0.01	5045	0.1875	0.1775	0.1975	0.0051	0.95
0.03	0.03	620	0.1875	0.1575	0.2175	0.0153	0.95
0.05	0.05	225	0.1875	0.1375	0.2375	0.0255	0.95

Strata-Detail Report for Row 1

Stratum h	Population Size		Sample Size		Proportion Ph
	Value Nh	Percent Pct(Nh)	Value nh	Percent Pct(nh)	
1	14000	29.2	1662	32.9	0.25
2	18000	37.5	1974	39.1	0.20
3	6000	12.5	587	11.6	0.15
4	10000	20.8	822	16.3	0.10

Strata-Detail Report for Row 2

Stratum h	Population Size		Sample Size		Proportion Ph
	Value Nh	Percent Pct(Nh)	Value nh	Percent Pct(nh)	
1	14000	29.2	204	32.9	0.25
2	18000	37.5	243	39.2	0.20
3	6000	12.5	72	11.6	0.15
4	10000	20.8	101	16.3	0.10

Strata-Detail Report for Row 3

Stratum h	Population Size		Sample Size		Proportion Ph
	Value Nh	Percent Pct(Nh)	Value nh	Percent Pct(nh)	
1	14000	29.2	74	32.9	0.25
2	18000	37.5	88	39.1	0.20
3	6000	12.5	26	11.6	0.15
4	10000	20.8	37	16.4	0.10

Confidence Intervals for One Proportion in a Stratified Design

For proportional allocation, the sample sizes are: 5126, 630, 229.

For optimal allocation, the sample sizes are: 5045, 620, 225.

For equal allocation, the sample sizes are: 6088, 748, 272.

We note the equal allocation requires a much larger sample size.

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) **0.01 0.03 0.05**
 Sample Size Allocation **Equal**

Set 1 Number of Strata **1**
 Set 1 Stratum Population Size **14000**
 Set 1 Stratum Proportion **0.25**

Set 2 Number of Strata **1**
 Set 2 Stratum Population Size **18000**
 Set 2 Stratum Proportion **0.2**

Set 3 Number of Strata **1**
 Set 3 Stratum Population Size **6000**
 Set 3 Stratum Proportion **0.15**

Set 4 Number of Strata **1**
 Set 4 Stratum Population Size **10000**
 Set 4 Stratum Proportion **0.10**

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

Confidence Intervals for One Proportion in a Stratified Design

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		Total Sample Size n	Proportion P	Confidence Interval Limits		Standard Error of P SE(P)	Confidence Level
Actual d(A)	Target d(T)			Lower LCL(P)	Upper UCL(P)		
0.0100	0.01	6088	0.1875	0.1775	0.1975	0.0051	0.95
0.0300	0.03	748	0.1875	0.1575	0.2175	0.0153	0.95
0.0499	0.05	272	0.1875	0.1376	0.2374	0.0255	0.95

Strata-Detail Report for Row 1

Stratum h	Population Size		Sample Size		Proportion Ph
	Value Nh	Percent Pct(Nh)	Value nh	Percent Pct(nh)	
1	14000	29.2	1522	25	0.25
2	18000	37.5	1522	25	0.20
3	6000	12.5	1522	25	0.15
4	10000	20.8	1522	25	0.10

Strata-Detail Report for Row 2

Stratum h	Population Size		Sample Size		Proportion Ph
	Value Nh	Percent Pct(Nh)	Value nh	Percent Pct(nh)	
1	14000	29.2	187	25	0.25
2	18000	37.5	187	25	0.20
3	6000	12.5	187	25	0.15
4	10000	20.8	187	25	0.10

Strata-Detail Report for Row 3

Stratum h	Population Size		Sample Size		Proportion Ph
	Value Nh	Percent Pct(Nh)	Value nh	Percent Pct(nh)	
1	14000	29.2	68	25	0.25
2	18000	37.5	68	25	0.20
3	6000	12.5	68	25	0.15
4	10000	20.8	68	25	0.10

Confidence Intervals for One Proportion in a Stratified Design

For proportional allocation, the sample sizes are: 5126, 630, 229.

For optimal allocation, the sample sizes are: 5045, 620, 225.

For equal allocation, the sample sizes are: 6088, 748, 272.

We note the equal allocation requires a much larger sample size.

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 response rates are expected to be 0.23, 0.19, and 0.12. The half-width 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 Proportion **0.23**

Set 1 Stratum Sample Size **215**

Set 2 Number of Strata **1**

Set 2 Stratum Population Size **18632**

Set 2 Stratum Proportion **0.19**

Set 2 Stratum Sample Size **269**

Set 3 Number of Strata **1**

Set 3 Stratum Population Size **10908**

Set 3 Stratum Proportion **0.12**

Set 3 Stratum Sample Size **193**

Set 4 Number of Strata **0**

Set 5 Number of Strata **0**

Show More Strata Sets **Unchecked**

Confidence Intervals for One Proportion in a Stratified Design

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	Proportion P	Confidence Interval Limits		Standard Error of P SE(P)	Confidence Level
			Lower LCL(P)	Upper UCL(P)		
0.0292	677	0.1856	0.1564	0.2147	0.0149	0.95
0.0383	677	0.1856	0.1473	0.2239	0.0149	0.99

Strata-Detail Report for Row 1

Stratum h	Population Size		Sample Size		Proportion Ph
	Value Nh	Percent Pct(Nh)	Value nh	Percent Pct(nh)	
1	14257	32.6	215	31.8	0.23
2	18632	42.5	269	39.7	0.19
3	10908	24.9	193	28.5	0.12

Strata-Detail Report for Row 2

Stratum h	Population Size		Sample Size		Proportion Ph
	Value Nh	Percent Pct(Nh)	Value nh	Percent Pct(nh)	
1	14257	32.6	215	31.8	0.23
2	18632	42.5	269	39.7	0.19
3	10908	24.9	193	28.5	0.12

Increasing the confidence level from 0.95 to 0.99 has increased the half width from 0.029 to 0.038.