Chapter 115

Confidence Intervals for One Proportion

Introduction

This routine calculates the sample size necessary to achieve a specified interval width or distance from the sample proportion to the confidence limit at a stated confidence level for a confidence interval for one proportion.

Caution: This procedure assumes that the proportion of the future sample will be the same as the proportion that is specified. If the sample proportion is different from the one specified when running this procedure, the interval width may be narrower or wider than specified.

Technical Details

Many methods have been devised for computing confidence intervals for a single proportion. Five of these methods are available in this procedure. The five confidence interval methods are

1. Exact (Clopper-Pearson)
2. Score (Wilson)
3. Score with continuity correction
4. Simple Asymptotic
5. Simple Asymptotic with continuity correction

For a comparison of methods, see Newcombe (1998a).

Confidence Interval Formulas

For each of the following methods, let *p* be the population proportion, and let *r* represent the number of successes from a sample of size *n*. Let $\hat{p} = r/n$.

Exact (Clopper-Pearson)

Using a mathematical relationship (see Fleiss et al (2003), p. 25) between the *F* distribution and the cumulative binomial distribution, the lower and upper confidence limits of a 100(1-α)% exact confidence interval for the true proportion *p* are given by

$$\left[\frac{r}{r+(n-r+1)F_{1-\alpha/2;2(n-r+1),2r}}, \frac{(r+1)F_{1-\alpha/2;2(r+1),2(n-r)}}{(n-r)+(r+1)F_{1-\alpha/2;2(r+1),2(n-r)}}\right]$$

One-sided limits may be obtained by replacing $\alpha/2$ by α.

Score (Wilson)

The Wilson Score confidence interval, which is based on inverting the z-test for a single proportion, is calculated using

$$\frac{\left(2n\hat{p}+z_{1-\alpha/2}^{2}\right)\pm z_{1-\alpha/2}\sqrt{z_{1-\alpha/2}^{2}+4n\hat{p}(1-\hat{p})}}{2\left(n+z_{1-\alpha/2}^{2}\right)}$$

One-sided limits may be obtained by replacing $\alpha/2$ by α.

Score with Continuity Correction

The Score confidence interval with continuity correction is based on inverting the z-test for a single proportion with continuity correction. The 100(1-α)% limits are calculated by

$$Lower\ Limit=\frac{\left(2n\hat{p}+z_{1-\alpha/2}^{2}-1\right)-z_{1-\alpha/2}\sqrt{z_{1-\alpha/2}^{2}-\{2+(1/n)\}+4\hat{p}\{n(1-\hat{p})+1\}}}{2\left(n+z_{1-\alpha/2}^{2}\right)}$$

$$Upper\ Limit=\frac{\left(2n\hat{p}+z_{1-\alpha/2}^{2}+1\right)+z_{1-\alpha/2}\sqrt{z_{1-\alpha/2}^{2}+\{2-(1/n)\}+4\hat{p}\{n(1-\hat{p})-1\}}}{2\left(n+z_{1-\alpha/2}^{2}\right)}$$

One-sided limits may be obtained by replacing $\alpha/2$ by α.

Simple Asymptotic

The simple asymptotic formula is based on the normal approximation to the binomial distribution. The approximation is close only for very large sample sizes. The 100(1-α)% confidence limits are given by

$$\hat{p}\pm z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

One-sided limits may be obtained by replacing $\alpha/2$ by α.

Simple Asymptotic with Continuity Correction

This formula is identical to the previous one, but with continuity correction. The 100(1-α)% confidence limits are

$$\left(\hat{p}-z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}-\frac{1}{2n},\hat{p}+z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}+\frac{1}{2n}\right)$$

One-sided limits may be obtained by replacing $\alpha/2$ by α.

Interval Widths (One-Sided vs. Two-Sided)

For two-sided intervals, the distance from the sample proportion to each of the limits may be different. Thus, instead of specifying the distance to the limits we specify the width of the interval, *W*.

The basic equation for determining sample size for a two-sided interval when *W* has been specified is

$$W = U - L$$

For one-sided intervals, the distance from the sample proportion to limit, *D*, is specified.

The basic equation for determining sample size for a one-sided upper limit when *D* has been specified is

$$D = U - \hat{p}$$

The basic equation for determining sample size for a one-sided lower limit when *D* has been specified is

$$D = \hat{p} - L$$

Each of these equations can be solved for any of the unknown quantities in terms of the others.

Example 1 – Calculating Sample Size

Suppose a study is planned in which the researcher wishes to construct a two-sided 95% exact (Clopper-Pearson) confidence interval for the population proportion such that the width of the interval is no wider than 0.06. The anticipated proportion estimate is 0.3, but a range of values from 0.1 to 0.5 will be included to determine the effect of the proportion estimate on necessary sample size. Instead of examining only the interval width of 0.06, widths of 0.04 and 0.10 will also be considered.

The goal is to determine the necessary sample size.

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 Interval Formula.......................... **Exact (Clopper-Pearson)**
Interval Type .... **Two-Sided**
Confidence Level **0.95**
Confidence Interval Width (Two-Sided) **0.04 0.06 0.10**
P (Proportion) .... **0.1 to 0.5 by 0.1**

Output

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

Numeric Reports

Numeric Results

Solve For: Sample Size
Confidence Interval Formula: Exact (Clopper-Pearson)
Interval Type: Two-Sided

Confidence Level	Sample Size N	Confidence Interval Width Target	Actual	If P = 0.5	Proportion P	Confidence Interval Limits Lower	Upper
0.95	914	0.04	0.04	0.066	0.1	0.081	0.121
0.95	1585	0.04	0.04	0.050	0.2	0.181	0.221
0.95	2065	0.04	0.04	0.044	0.3	0.280	0.320
0.95	2353	0.04	0.04	0.041	0.4	0.380	0.420
0.95	2449	0.04	0.04	0.040	0.5	0.480	0.520
0.95	417	0.06	0.06	0.098	0.1	0.073	0.133
0.95	715	0.06	0.06	0.075	0.2	0.171	0.231
0.95	928	0.06	0.06	0.065	0.3	0.271	0.331
0.95	1056	0.06	0.06	0.061	0.4	0.370	0.430
0.95	1098	0.06	0.06	0.060	0.5	0.470	0.530
0.95	158	0.10	0.10	0.161	0.1	0.058	0.158
0.95	264	0.10	0.10	0.124	0.2	0.153	0.253
0.95	341	0.10	0.10	0.109	0.3	0.252	0.352
0.95	387	0.10	0.10	0.102	0.4	0.351	0.451
0.95	402	0.10	0.10	0.100	0.5	0.450	0.550

Confidence Level — The proportion of confidence intervals (constructed with this same confidence level, sample size, etc.) that would contain the population proportion.
N — The size of the sample drawn from the population.
Confidence Interval Width — The distance from the lower limit to the upper limit.
Target Width — The value of the width that is entered into the procedure.
Actual Width — The value of the width that is obtained from the procedure.
If P = 0.5 — The maximum width for a confidence interval with sample size N.
P — The sample proportion.
Confidence Interval Limits — The lower and upper limits of the confidence interval.

Summary Statements

A single-group design will be used to obtain a two-sided 95% confidence interval for a single proportion. The Exact (Clopper-Pearson) formula will be used to calculate the confidence interval. The sample proportion is assumed to be 0.1. To produce a confidence interval with a width of no more than 0.04, 914 subjects will be needed.

Dropout-Inflated Sample Size

Dropout Rate	Sample Size N	Dropout-Inflated Enrollment Sample Size N'	Expected Number of Dropouts D
20%	914	1143	229
20%	1585	1982	397
20%	2065	2582	517
20%	2353	2942	589
20%	2449	3062	613
20%	417	522	105
20%	715	894	179
20%	928	1160	232
20%	1056	1320	264
20%	1098	1373	275
20%	158	198	40
20%	264	330	66
20%	341	427	86
20%	387	484	97
20%	402	503	101

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 the confidence interval is computed. If N subjects are evaluated out of the N' subjects that are enrolled in the study, the design will achieve the stated confidence interval.
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, 1143 subjects should be enrolled to obtain a final sample size of 914 subjects.

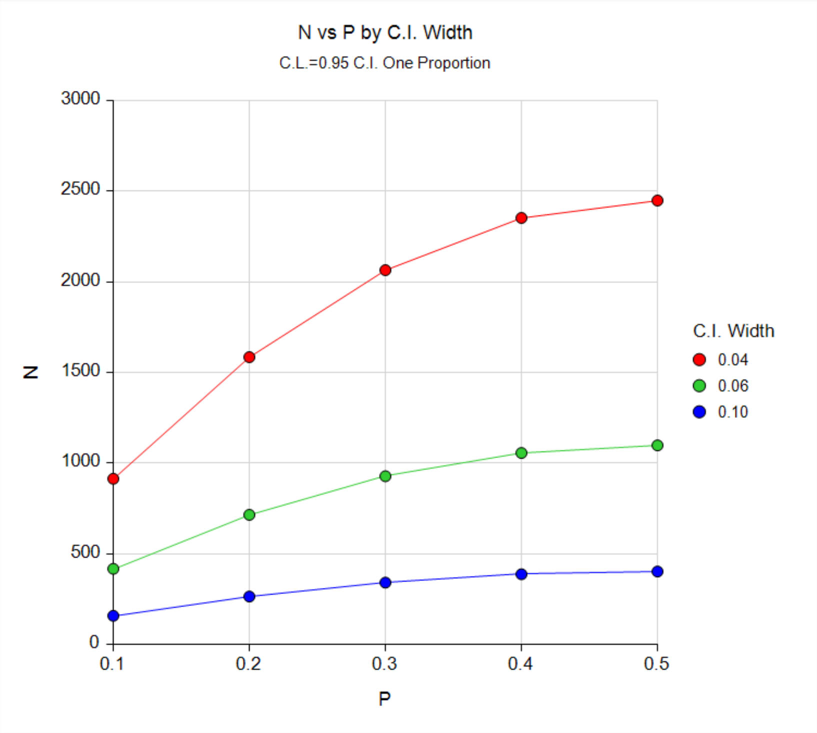
References

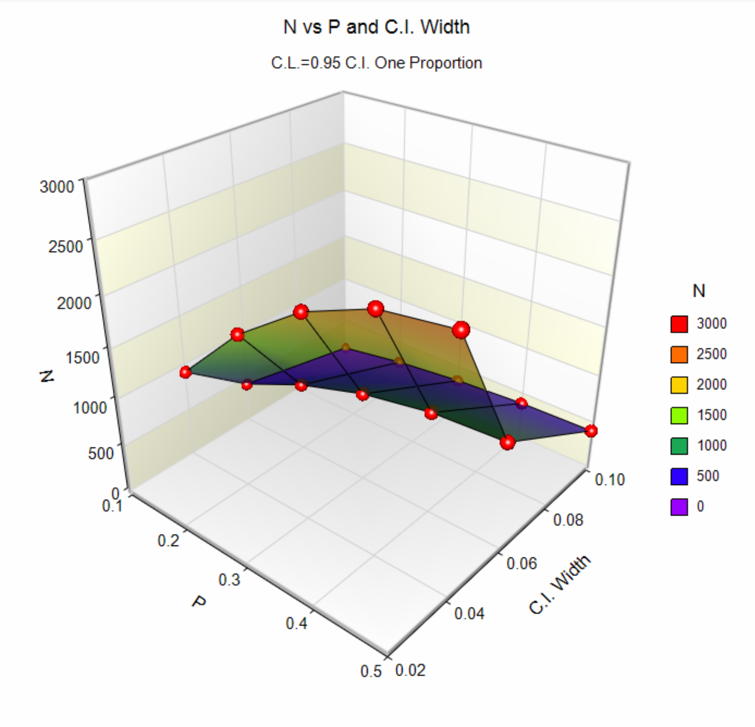
Fleiss, J. L., Levin, B., Paik, M.C. 2003. Statistical Methods for Rates and Proportions. Third Edition. John Wiley & Sons. New York.
Newcombe, R. G. 1998. 'Two-Sided Confidence Intervals for the Single Proportion: Comparison of Seven Methods.' Statistics in Medicine, 17, pp. 857-872.

This report shows the calculated sample size for each of the scenarios.

Plots Section

Plots





These plots show the sample size versus the sample proportion for the three confidence interval widths.

Example 2 – Validation using Fleiss, Levin, and Paik (2003)

Fleiss, Levin, and Paik (2003), pages 22-23, give an example of a calculation for an exact (Clopper-Pearson) one-sided lower limit confidence interval for a single proportion when the confidence level is 95%, the sample proportion is 0.92, and the distance from the lower limit to the sample proportion is 0.15104. The necessary sample size is 25.

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 Interval Formula.......................... **Exact (Clopper-Pearson)**
Interval Type .... **Lower Limit**
Confidence Level **0.95**
Distance from P to Limit (One-Sided) **0.15104**
P (Proportion) **0.92**

Output

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

Numeric Results

Solve For: Sample Size
Confidence Interval Formula: Exact (Clopper-Pearson)
Interval Type: Lower One-Sided

Confidence Level	Sample Size N	Distance from P to Lower Limit: Target	Actual	If P = 0.5	Proportion P	Confidence Interval Limits: Lower	Upper
0.95	25	0.151	0.151	0.177	0.92	0.769	1

PASS also calculated the necessary sample size to be 25.

Example 3 – Validation using Newcombe (1998a)

Newcombe (1998a), pages 860-861, gives an example of a calculation for a two-sided confidence interval for a single proportion for each of the methods when the confidence level is 95%. Here we validate the score method with continuity correction. The sample proportion is 0.034483, and the interval width is 0.1945. The necessary sample size is 29.

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 Interval Formula.......................... **Score (Contin. Correction)**
Interval Type .... **Two-Sided**
Confidence Level **0.95**
Confidence Interval Width (Two-Sided) **0.1945**
P (Proportion) **0.034483**

Output

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

Numeric Results

Solve For: Sample Size
Confidence Interval Formula: Score with Continuity Correction
Interval Type: Two-Sided

		Confidence Interval Width				Confidence Interval Limits	
Confidence Level	Sample Size N	Target	Actual	If P = 0.5	Proportion P	Lower	Upper
0.95	29	0.195	0.194	0.372	0.034	0.002	0.196

PASS also calculated the necessary sample size to be 29.

Example 4 – Zero Events, Validation using Lachin (2000)

Lachin (2000), page 19, gives an example of a calculation for a one-sided upper limit exact confidence interval for a single proportion when the confidence level is 95%, the sample proportion is 0, and the upper bound is 0.01. The necessary sample size is 299.

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 Interval Formula.......................... **Exact (Clopper-Pearson)**
Interval Type .... **Upper Limit**
Confidence Level **0.95**
Distance from P to Limit (One-Sided) **0.01**
P (Proportion) **0**

Output

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

Numeric Results

Solve For: Sample Size
Confidence Interval Formula: Exact (Clopper-Pearson)
Interval Type: Upper One-Sided

Confidence Level	Sample Size N	Distance from P to Upper Limit: Target	Distance from P to Upper Limit: Actual	Distance from P to Upper Limit: If P = 0.5	Proportion P	Confidence Interval Limits: Lower	Confidence Interval Limits: Upper
0.95	299	0.01	0.01	0.049	0	0	0.01

PASS also calculated the necessary sample size to be 299.