

Chapter 272

Confidence Intervals for One-Sample Specificity

Introduction

This procedure calculates the (whole table) sample size necessary for a single-sample specificity confidence interval, based on a specified specificity, interval width, confidence level, and prevalence.

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

Specificity (True Negative Rate)

The specificity (or true negative rate) is the proportion of the individuals with a known negative condition for which the predicted condition is negative.

		Predicted Condition		
		Positive	Negative	
True Condition	Positive	True Positive (A)	False Negative (C)	Specificity = $D / (B + D)$
	Negative	False Positive (B)	True Negative (D)	

Prevalence

The prevalence is the overall proportion of individuals with a positive condition.

		Predicted Condition		
		Positive	Negative	
True Condition	Positive	True Positive (A)	False Negative (C)	Prevalence = $(A + C) / (A + B + C + D)$
	Negative	False Positive (B)	True Negative (D)	

Technical Details

In general terms, the required sample size is determined by first calculating the sample size needed for the specificity proportion confidence interval, followed by a prevalence adjustment. The initial sample size calculation for the specificity confidence interval gives the number of individuals with a negative condition that are needed. The prevalence adjustment is used to add the number of individuals with a positive condition that are needed. The resulting sample size is the total number of individuals needed to obtain a table where the number of negative condition individuals will give the needed confidence interval width for the specificity.

Similarly, when calculating the confidence interval width for a given sample size, the given sample size is first used to produce the number of negative condition individuals, according to the given prevalence, and then the width based on the resulting negative condition count is then calculated.

If prevalence is to be ignored, a value of 0 may be used for prevalence, or the Confidence Intervals for One Proportion procedure may be used, as the scenario has been reduced to a simple confidence interval of a single proportion.

Confidence Interval Formulas

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).

For each of the following methods, let p be the population specificity, and let r represent the number of true negatives with n total negatives. 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 $100(1-\alpha)\%$ 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

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

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

Confidence Intervals for One-Sample Specificity

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 $100(1-\alpha)\%$ limits are calculated by

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

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

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 $100(1-\alpha)\%$ 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 $100(1-\alpha)\%$ 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 specificity 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 specificity 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.

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, go to the Procedure Window chapter.

Design Tab

The Design tab contains most of the parameters and options that you will be concerned with.

Solve For

Solve For

This option specifies the parameter to be solved for from the other parameters.

Confidence Interval Method

Confidence Interval Formula

Specify the formula to be in used in calculation of confidence intervals.

- **Exact (Clopper-Pearson)**
The exact formula uses the binomial probabilities directly.
 - **Score (Wilson)**
This formula is based on inverting a critical ratio test.
 - **Score (Continuity Correction)**
This formula is based on inverting a critical ratio test with continuity correction.
 - **Simple Asymptotic**
The simple asymptotic formula is based on the normal approximation to the binomial.
 - **Simple Asymptotic (Continuity Correction)**
This formula is based on the normal approximation to the binomial with continuity correction.
-

One-Sided or Two-Sided Interval

Interval Type

Specify whether the interval to be used will be a two-sided confidence interval, an interval that has only an upper limit, or an interval that has only a lower limit.

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 specificity.

Often, the values 0.95 or 0.99 are used. You can enter single values or a range of values such as *0.90, 0.95* or *0.90 to 0.99 by 0.01*.

Confidence Intervals for One-Sample Specificity

Sample Size

N (Sample Size)

Enter one or more values for the sample size. This is the number of individuals selected at random from the population to be in the study.

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

Precision

Confidence Interval Width (Two-Sided)

This is the distance from the lower confidence limit to the upper confidence limit.

You can enter a single value or a list of values. The value(s) must be between 0 and 1.

Distance from Specificity to Limit (One-Sided)

This is the distance from the sample specificity to the lower or upper limit of the confidence interval, depending on whether the Interval Type is set to Lower One-Sided or Upper One-Sided.

You can enter a single value or a list of values. The value(s) must be between 0 and 1.

Specificity

Specificity

Enter an estimate of the specificity (or true negative rate). The sample size and width calculations assume that the value entered here is the specificity estimate that is obtained from the sample. If the sample specificity is different from the one specified here, the width may be narrower or wider than specified.

You can enter a range of values such as *0.1 0.2 0.3* or *0.1 to 0.5 by 0.1*.

Prevalence

Prevalence

Enter the value of the assumed prevalence. Since the sample size calculation for the specificity is concerned only with negative condition individuals, the prevalence is used to produce the total sample size needed to generate the proper number of negative condition individuals.

You can enter a range of values such as *0.1 0.2 0.3* or *0.1 to 0.5 by 0.1*.

Confidence Intervals for One-Sample Specificity

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 specificity such that the width of the interval is no wider than 0.06. The anticipated specificity estimate is 0.7, but a range of values from 0.5 to 0.9 will be included to determine the effect of the specificity estimate on necessary sample size. Instead of examining only the interval width of 0.06, widths of 0.04, 0.08, and 0.10 will also be considered.

The goal is to determine the total sample size needed when also accounting for 30% prevalence.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Confidence Intervals for One-Sample Specificity** procedure. You may make the appropriate entries as listed below or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
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 0.08 0.10
Specificity	0.5 to 0.9 by 0.05
Prevalence	0.3

Annotated Output

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

Numeric Results

Numeric Results for Two-Sided Confidence Intervals for One-Sample Specificity
 Confidence Interval Formula: Exact (Clopper-Pearson)

Confidence Level	Sample Size (N)	Target Width	Actual Width	Specificity	Lower Limit	Upper Limit	Prevalence	Number of Negatives
0.950	3499	0.040	0.040	0.500	0.480	0.520	0.300	2449
0.950	3465	0.040	0.040	0.550	0.530	0.570	0.300	2425
0.950	3362	0.040	0.040	0.600	0.580	0.620	0.300	2353
0.950	3190	0.040	0.040	0.650	0.630	0.670	0.300	2233
0.950	2950	0.040	0.040	0.700	0.680	0.720	0.300	2065
0.950	2642	0.040	0.040	0.750	0.730	0.770	0.300	1849
0.950	2265	0.040	0.040	0.800	0.779	0.819	0.300	1585
0.950	1819	0.040	0.040	0.850	0.829	0.869	0.300	1273
0.950	1306	0.040	0.040	0.900	0.879	0.919	0.300	914
0.950	1569	0.060	0.060	0.500	0.470	0.530	0.300	1098
0.950	1555	0.060	0.060	0.550	0.520	0.580	0.300	1088
0.950	1509	0.060	0.060	0.600	0.570	0.630	0.300	1056
0.950	1432	0.060	0.060	0.650	0.620	0.680	0.300	1002
0.950	1326	0.060	0.060	0.700	0.669	0.729	0.300	928
0.950	1189	0.060	0.060	0.750	0.719	0.779	0.300	832
0.950	1022	0.060	0.060	0.800	0.769	0.829	0.300	715
.
.
.

[report continues]

Confidence Intervals for One-Sample Specificity

References

- Hajian-Tilaki, K. 2014. 'Sample size estimation in diagnostic test studies of biomedical informatics.' Journal of Biomedical Informatics, 48, pp. 193-204.
- 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.

Report Definitions

Confidence level is the proportion of confidence intervals (constructed with this same confidence level, sample size, etc.) that would contain the population specificity.

N is the size of the sample drawn from the population.

Width is the distance from the lower limit to the upper limit.

Target Width is the value of the width that is entered into the procedure.

Actual Width is the value of the width that is obtained from the procedure.

Specificity is the assumed sample specificity, or true negative rate.

Lower Limit is the lower limit of the confidence interval.

Upper Limit is the upper limit of the confidence interval.

Prevalence is the assumed overall proportion of individuals with a positive condition.

Number of Negatives is calculated as the Sample Size times the Prevalence with appropriate rounding. The

Number of Negatives is the count upon which the confidence interval width calculation is based.

Summary Statements

A sample size of 8164 produces a two-sided 95% confidence interval with a width equal to 0.040 when the sample specificity is 0.500 and the prevalence is 0.300.

Dropout-Inflated Sample Size

Dropout Rate	Sample Size N	Dropout- Inflated Enrollment Sample Size N'	Expected Number of Dropouts D
20%	3499	4374	875
20%	3465	4332	867
20%	3362	4203	841
20%	3190	3988	798
20%	2950	3688	738
20%	2642	3303	661
20%	2265	2832	567
20%	1819	2274	455
20%	1306	1633	327
20%	1569	1962	393
20%	1555	1944	389
20%	1509	1887	378
20%	1432	1790	358
20%	1326	1658	332
.	.	.	.
.	.	.	.
.	.	.	.

Definitions

Dropout Rate (DR) is 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").

N is 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.

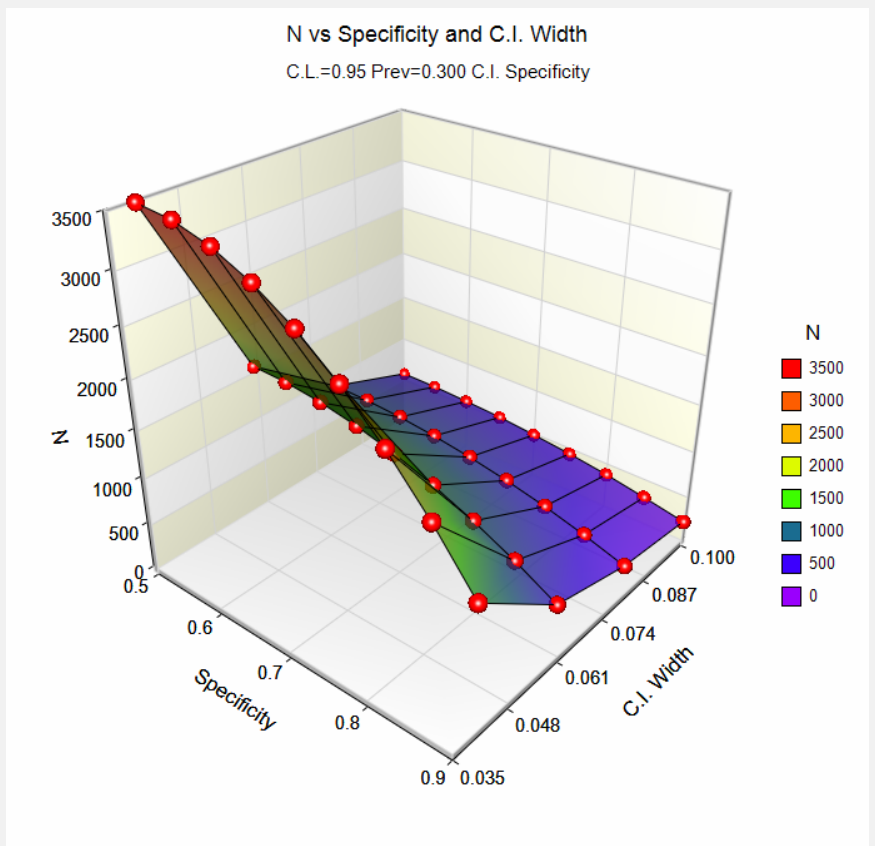
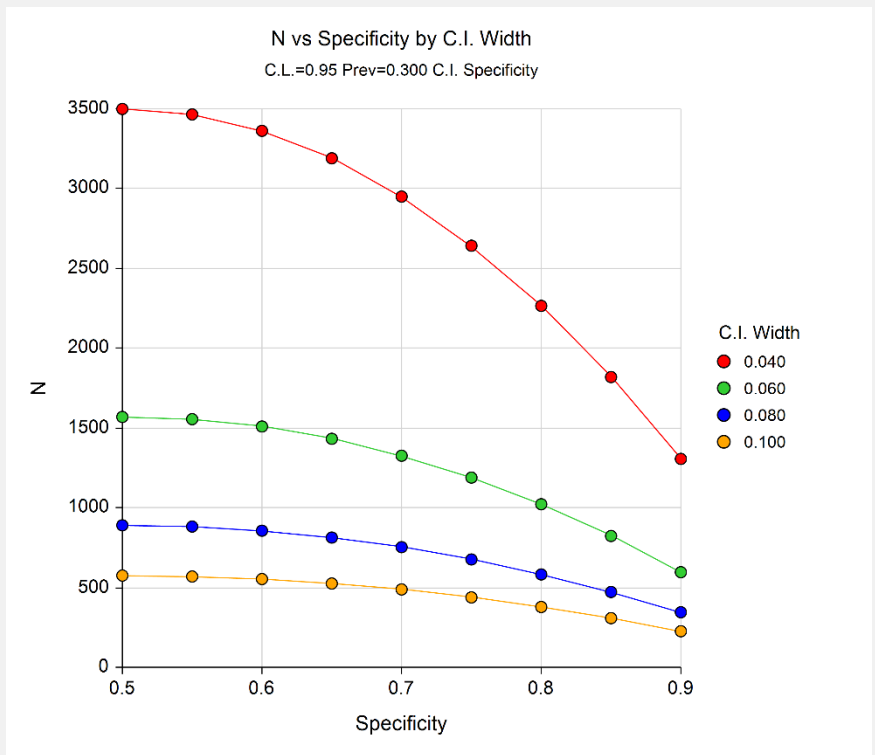
N' is the total number of subjects that should be enrolled in the study in order to end up with N evaluable subjects, based on the assumed dropout rate. After solving for N, N' is calculated by inflating N using the formula $N' = N / (1 - DR)$, with N' always rounded up. (See Julious, S.A. (2010) pages 52-53, or Chow, S.C., Shao, J., and Wang, H. (2008) pages 39-40.)

D is the expected number of dropouts. $D = N' - N$.

These reports show the calculated sample size for each of the scenarios.

Confidence Intervals for One-Sample Specificity

Plots Section



These plots show the sample size versus the sample specificity for the four confidence interval widths.

Confidence Intervals for One-Sample Specificity

Example 2 – Validation using Hajian-Tilaki (2014)

Hajian-Tilaki (2014), page 196, gives an example of a calculation for a simple asymptotic two-sided confidence interval for a single specificity when the confidence level is 95%, the specificity is 0.8, the prevalence is 0.1, and the margin of error is 3% (With a margin of error (precision) of 3%, the width is 0.06). The necessary sample size is calculated to be 759.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Confidence Intervals for One-Sample Specificity** procedure. You may make the appropriate entries as listed below or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Sample Size
Confidence Interval Formula	Simple Asymptotic
Interval Type	Two-Sided
Confidence Level	0.95
Confidence Interval Width (Two-Sided) ..	0.06
Specificity	0.8
Prevalence	0.1

Output

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

Numeric Results

Numeric Results for Two-Sided Confidence Intervals for One-Sample Specificity								
Confidence Interval Formula: Simple Asymptotic								
Confidence Level	Sample Size (N)	Target Width	Actual Width	Specificity	Lower Limit	Upper Limit	Prevalence	Number of Negatives
0.950	759	0.060	0.060	0.800	0.770	0.830	0.100	683

PASS also calculates the necessary sample size to be 759. For some entries in the table, the sample size calculated in PASS is slightly different from the article. In the article the sample sizes are calculated directly, while PASS calculates the sample size needed before prevalence is taken into account, and then adjusts for the prevalence.