

A SAS® Program to Calculate and Plot Widths of (1- α)100% Confidence Intervals for Binomial Proportions

Maribeth Johnson and Mark Litaker
Medical College of Georgia, Augusta, GA

ABSTRACT

The width of a confidence interval around the estimated proportion may be used as a measure of the precision of a measurement obtained in a study with a given sample size. A SAS® program is presented which calculates the (1- α)100% confidence intervals of binomial proportions, using the F-distribution, for a range of sample sizes and proportions. Plots of confidence interval widths across the expected range of proportions which may be observed are an aid in determining the sample size necessary to achieve a desired level of study precision.

INTRODUCTION

Medical researchers are sometimes interested in estimating proportions such as the prevalence of a disease or the sensitivity or specificity of a test in detecting a certain condition or disease. A confidence interval around these estimated proportions gives an indication of the limits for the true population proportion. Researchers who have an idea of the size of these proportions prior to the start of a study may need to know the sample size necessary to estimate the proportion within a certain amount of error. This program not only calculates the exact binomial confidence interval for proportions but may help researchers in determining sample sizes for future studies.

BINOMIAL CONFIDENCE INTERVAL

The proportion of individuals in a sample of n observations that fall into one of two categories is estimated by $\hat{p} = X/n$, where X is the number of observations in one category. Using a relationship between the F-distribution and the binomial distribution, an exact (1- α)100% confidence interval may be computed for this estimate of the binomial parameter p (Zar, 1984).

The lower confidence limit is

$$\text{Lower} = X / \{ X + (n-X+1) F_{\alpha/2, v1, v2} \},$$

where

$$v1 = 2(n-X+1)$$

$$\text{and } v2 = 2X.$$

The upper confidence limit is

$$\text{Upper} = \{ (X+1) F_{\alpha/2, v1', v2'} \} / \{ n - X + (X+1) F_{\alpha/2, v1', v2'} \},$$

where

$$v1' = 2(X+1)$$

$$\text{and } v2' = 2(n-X).$$

Use of the exact binomial method avoids the inaccuracy that may occur when using the normal approximation.

The normal approximation usually does not give accurate confidence limits, it is especially poor when np or $n(1-p)$ is less than 5 or when p is near 0 or 1 (say $p < .20$ or $p > .80$). Also, the normal approximation gives confidence limits symmetrical around p , which can result in the computation of a lower limit less than 0 or an upper limit greater than 1.

THE PROGRAM

The program produces the confidence intervals for incremental values of p at different sample sizes (n). The value of X (number of successes) is calculated from these two parameters. X is not necessarily an integer because of this but the FINV function that returns the quantile from the F-distribution will accept noninteger degrees of freedom parameters.

The print of the confidence limits and widths by each level of sample size enables direct comparison of the precision associated with each. The plot of the width of the confidence intervals across the values selected for p , with the point labels specified as n , enables one to see the decrease in width across the levels of sample size for a given p . The print of the transposed widths displays the same information in tabular form that is shown graphically in the plot.

CONCLUSION

This program calculates the exact (1- α)100% binomial confidence interval for one or more proportions. When used prior to the beginning of a study where the objective is to estimate a proportion, it enables the user to examine the level of precision associated with different sample sizes across the range of proportions anticipated.

REFERENCES

Zar, J. H. 1984. *Biostatistical Analysis*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey 07632.

SAS is a registered trademark or trademark of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

```

*****;
* This program calculates (1-alpha)*100% confidence *;
* limits for the binomial parameter p, based on the *;
* F distribution. Zar, 1984 *;
*****;
options nodate nonumber;

*Binomial parameters of interest;
%let p1 = .05; *Lowest;
%let p2 = .95; *Highest;
%let p_incr = .05; *Increment;

*Sample sizes of interest;
%let n1 = 50; *Smallest;
%let n2 = 200; *Largest;
%let n_incr = 50; *Increment;

*Error rate and confidence level;
%let alpha = .05; *Rate of error;
%let pct_ci = 95; *(1-alpha)*100;

data ci;

do p = &p1 to &p2 by &p_incr;
do n = &n1 to &n2 by &n_incr;

x=n*p; *Number of successes;
percent = (x/n)*100;

*Numerator df for lower limit;
lv1 = 2*(n-x+1);
*Denominator df for lower limit;
lv2 = 2*x;
*Numerator df for upper limit;
uv1 = 2*(x + 1);
*Denominator df for upper limit;
uv2 = 2*(n - x);

*Determine values from F distribution;
f_l = finv(1-&alpha/2,lv1,lv2);
f_u = finv(1-&alpha/2,uv1,uv2);

*Calculate confidence limits and width;
lower = (x/(x+((n - x + 1)*f_l)))*100;
if lower = . then lower = 0;
upper =
(((x + 1)*f_u)/(n - x + ((x + 1)*f_u)))*100;
if upper = . then upper = 100;
width = upper-lower;

output;
end; *do n;
end; *do p;

proc sort data=ci; by n percent;
run;
proc print data=ci label noobs;
var x percent lower upper width;
label
x = 'Number of successes'
n = 'Sample size'
percent = 'Percentage of successes (p*100)'
lower = 'Lower confidence limit'
upper = 'Upper confidence limit'
width = 'Width of confidence interval' ;
by n; pageby n;
title1 "&pct_ci% Confidence Intervals and Widths";

```

```

title2 "Limits Expressed as Percentages";
run;

proc plot data=ci;
plot width*percent=' ' $ n;
title1 "Plot of &pct_ci% Confidence Width
by Percentage of Successes";
title2 " ";
run; quit;

proc sort; by percent;
run;

proc transpose out=tran;
var width;
id n;
by percent;
run;

proc print data=tran label noobs;
var percent _&n1--&n2;
label
percent = 'Percentage of successes (p*100)';
title "&pct_ci% Confidence Widths for All Proportions
and Sample Sizes";
run;
*****;

```

USE OF THE BETA DISTRIBUTION

There is a relationship between the F and the beta distributions such that if $X \sim F_{v_1, v_2}$ then $\{(v_1/v_2)X\}/\{1 + (v_1/v_2)X\} \sim \text{beta}(v_1/2, v_2/2)$. Using this relationship the following code may be substituted in the program and the results are identical.

```

*First shape parameter for lower limit;
lv1 = x;
*Second shape parameter for lower limit;
lv2 = (n-x+1);
*First shape parameter for upper limit;
uv1 = (x+1);
*Second shape parameter for upper limit;
uv2 = (n-x);

*Determine values from beta distribution;
beta_l = betainv(&alpha/2,lv1,lv2);
beta_u = betainv(1-&alpha/2,uv1,uv2);

*Calculate confidence limits and width;
lower = beta_l*100;
if lower = . then lower = 0;
upper = beta_u*100;
if upper = . then upper = 100;
width = upper-lower;

```

CONTACT INFORMATION

Maribeth Johnson
Office of Biostatistics AE-3035
Medical College of Georgia
Augusta, GA 30912-4900

maribeth@stat.mcga.edu