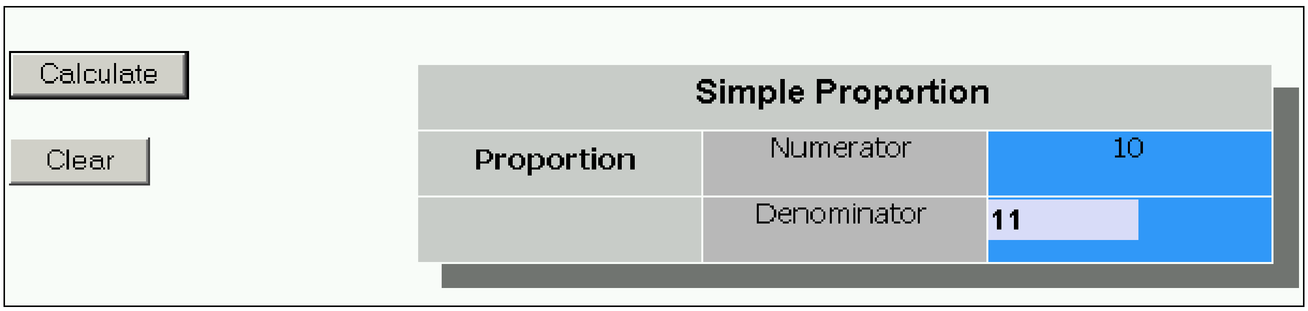
January 5, 2006

Documentation for
Confidence Intervals for a Proportion

Kevin M. Sullivan, PhD, MPH, MHA   cdckms@sph.emory.edu
Minn M. Soe, MD, MCTM, MPH   msoe@sph.emory.edu

This module calculates various confidence intervals for a proportion. First, the user is prompted to enter a numerator and denominator value:



Default values are provided: a numerator of 10 and denominator of 11. The user can change these to any other values. The output from the from the above example is as follows:

Results

95% Confidence Limits for Proportion 10/11

	Lower CL	Proportion (Percent)	Upper CL
		90.9091%	
Mid-P Exact	62.66%		99.55%
Fisher's Exact	58.72%		99.77%
Wald (Normal Approximation)	73.92%		100%
Modified Wald	60.1%		100%
Score(Wilson)	62.27%		98.38%
Score with Continuity Correction (Fleiss Quadratic)	57.12%		99.52%

npq=0.9091
The Wald method (Normal Approximation) is not recommended when npq < 5.

LookFirst items: Editor's choice of items to examine first.

The interpretation is that the observed proportion is 0.909 or 90.9% and several methods are used to calculate the confidence interval around this point estimate: Mid-P exact, Fisher's exact, Wald, modified Wald, Score, and Score with continuity correction. Which confidence interval method should you use? There is some debate concerning the best confidence interval method, but our preference is for the exact mid-p method. Refer to Agresti and Coull (1998) and Newcombe (1998) for discussion and comparisons of various confidence interval methods for a proportion.

For estimates > 1.0, the values of 1.0 is presented; for estimates < 0.0, the value 0.0 is presented. Currently all confidence intervals calculated are two-sided 95% confidence intervals. The formulae for the methods are provided in the following section.

Formulae

The notation for the formulae are:

a = the observed numerator
n = the observed denominator

$\hat{p}$ = the estimated proportion which is a/n
= $\hat{p}$ - 1
$Z_{1-\alpha/2}$ = the two-sided Z value, 1.96 for a 95% confidence interval
α = the tail probability, e.g., 0.05 for a 95% confidence interval
p_{LB} = the lower bound proportion
p_{UB} = the upper bound proportion

Mid-p Exact Limits:
For the Mid-p exact limits for the lower and upper bounds (Rothman and Boice, 1979) are:

$$\alpha/2 = \frac{1}{2}\binom{n}{a} p_{LB}^{a}(1-p_{LB})^{n-a} + \sum_{k=a+1}^{n}\binom{n}{k} p_{LB}^{k}(1-p_{LB})^{n-k}$$

$$\alpha/2 = \frac{1}{2}\binom{n}{a} p_{UB}^{a}(1-p_{UB})^{n-a} + \sum_{k=0}^{a-1}\binom{n}{k} p_{UB}^{k}(1-p_{UB})^{n-k}$$

Fisher's Exact Limits:
For the exact limits for the lower and upper bounds (Rothman and Boice, 1979) are:

$$\alpha/2 = \sum_{k=a}^{n}\binom{n}{k} p_{LB}^{k}(1-p_{LB})^{n-k}$$

$$\alpha/2 = \sum_{k=0}^{a}\binom{n}{k} p_{UB}^{k}(1-p_{UB})^{n-k}$$

Wald (Normal Approximation):
The lower and upper bounds for the Wald method (Rosner, 2000) are as follows:

$$\hat{p} \pm Z_{1\text{-}a/2}\sqrt{\hat{p}\hat{q}/n}$$

In the Rosner text it is recommended that the Wald method be used only if $n\hat{p}\hat{q} \geq 5$. If $n\hat{p}\hat{q} < 5$, a message will be displayed at the bottom of the output. In the above example, note the message:

"npq=0.9090909. The Wald method (Normal Approximation) is not recommended when npq < 5."

Modified Wald:
This method has additional notation (Agresti and Coull, 1998):

$$\hat{p}' = \frac{a + \frac{Z^2_{1-\alpha/2}}{2}}{n + Z^2_{1-\alpha/2}} \quad \text{and} \quad \hat{q}' = 1 - \hat{p}'$$

with the confidence interval:

$$\hat{p}' \pm Z_{1-\alpha/2}\sqrt{\frac{\hat{p}'\hat{q}'}{n + Z^2_{1-\alpha/2}}}$$

Score (Wilson):
The lower and upper bounds (Rothman and Boice, 1979) are:

$$p_{LB} = \frac{n}{n + Z^2_{1-\alpha/2}}\left[\frac{a}{n} + \frac{Z^2_{1-\alpha/2}}{2n} - Z^2_{1-\alpha/2}\sqrt{\frac{a(n-a)}{n^3} + \frac{Z^2_{1-\alpha/2}}{4n^2}}\right]$$

$$p_{UB} = \frac{n}{n + Z^2_{1-\alpha/2}}\left[\frac{a}{n} + \frac{Z^2_{1-\alpha/2}}{2n} + Z^2_{1-\alpha/2}\sqrt{\frac{a(n-a)}{n^3} + \frac{Z^2_{1-\alpha/2}}{4n^2}}\right]$$

Score with Continuity Correction (Fleiss Quadratic):
The lower and upper bounds (Fleiss, 1981) are:

$$\frac{(2n\hat{p}+Z^{2}_{1\text{-}a/2}-1)-Z_{1\text{-}a/2}\sqrt{Z^{2}_{1\text{-}a/2}-(2+1/n)+4\hat{p}(n\hat{q}+1)}}{2(n+Z^{2}_{1\text{-}a/2})}$$

$$\frac{(2n\hat{p}+Z^{2}_{1\text{-}a/2}+1)+Z_{1\text{-}a/2}\sqrt{Z^{2}_{1\text{-}a/2}+(2\text{-}1/n)+4\hat{p}(n\hat{q}\text{-}1)}}{2(n+Z^{2}_{1\text{-}a/2})}$$

References

Agresti A, Coull BA. Approximate is better than "Exact" for interval estimation of binomial proportions. *The American Statistician* 1988;52(2):119-126.

Fleiss JL. Statistical Methods for Rates and Proportions, 2nd Ed. John Wiley & Sons, New York, 1981.

Newcombe RG. Two-sided confidence intervals for the single proportion: comparison of seven methods. *Statistics in Medicine* 1988;17:857-872.

Rosner B. Fundamentals of Biostatistics, 5th Edition. Duxbury Press, 2000.

Rothman KJ, Boice JD Jr: Epidemiologic analysis with a programmable calculator. NIH Pub No. 79-1649. Bethesda, MD: National Institutes of Health, 1979;31-32.

Updates

Updated June 18 2004: Typo in the numerator for confidence interval formula for the modified Wald which was changed from "4" to Z^2; this is consistent with the Javascript code. There was also a typo in the Mid-p upper bound confidence limit formula.

Updated December 12 2005: the Mid-p exact confidence interval procedure added. Ability to specify a confidence interval other than 95% added.