A Lack of Confidence

The *Wald Interval*, commonly called the 'Standard Interval', is the confidence interval most commonly presented in introductory statistics textbooks for the population proportion *p*. It is the confidence interval prescribed in the VCE Study Design for Mathematical Methods and takes the form

$$\left(\hat{p}-k\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},\ \hat{p}+k\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

Here, n is the sample size, $\hat{p}$ is the observed sample proportion and $\Pr(-k<Z<k)=\frac{C}{100}$ where C is the confidence level and Z is the standard normal random variable.

Problems with the Wald interval

The Wald Interval possesses a number of defects:

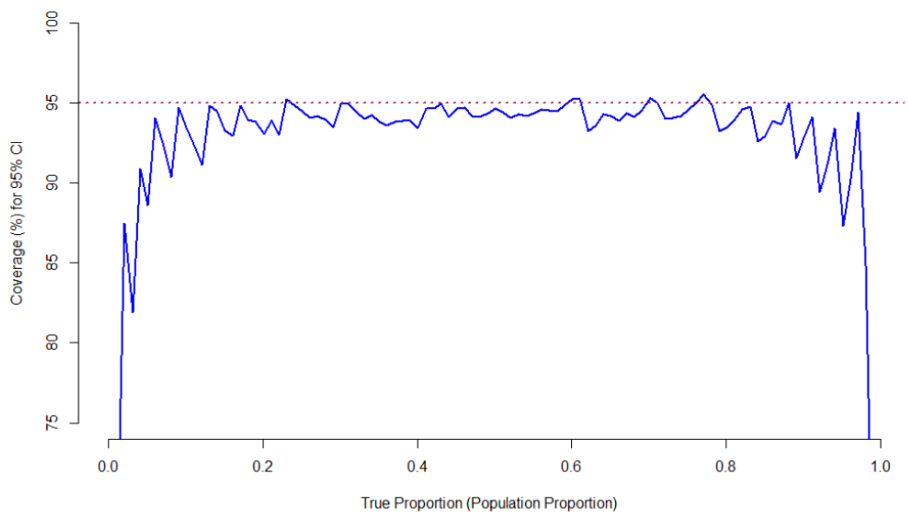
1. In the special cases $\hat{p}=0$ or $\hat{p}=1$, the Wald Interval has zero width and thus disappears. The Wald Interval also performs very poorly for values of $\hat{p}$ close to 0 and 1.

2: Intervals can have 'overshoot'. For example, when $n=30$ and $\hat{p}=0.9$ the approximate 95% Wald interval is (0.793, 1.007).

3. The Wald Interval often performs very poorly in practical scenarios, in that the *coverage probability* (see below) is often less than the nominal confidence level (e.g. the coverage probability of the nominal 95% Wald Interval is often less than 95%). That is not good since we hope to have a reasonable 'coverage' when constructing a confidence interval.

Note: The probability that an interval contains or *covers* the true value of an unknown parameter is called the *coverage probability*. It is a property of the procedure that produces the interval. The interval produced for a particular sample, using a procedure with coverage probability $\frac{C}{100}$, is said to have a confidence level of C.

The coverage probability of intervals such as the Wald Interval can be investigated by simulating random sampling from a population with a known value of *p*. A confidence interval is constructed for each of the random samples to see how many such confidence intervals actually 'cover' (include) *p*.



Simulated coverage probability of the nominal 95% Wald Interval

(From *Five Confidence Intervals For Proportions You Should Know About* - Dr Dennis Robert Aug 2020)

Alternatives to the Wald interval

Given the imposition of statistical inference onto Mathematical Methods by VCAA, the ubiquity of CAS technology in VCE mathematics and the 'black box' (or should thst be 'black CAS' approach to calculating confidence intervals, it is puzzling that the Wald Interval is the only confidence interval mentioned in the VCE Study Design for Mathematical Methods and even more puzzling that its defects are not mentioned.

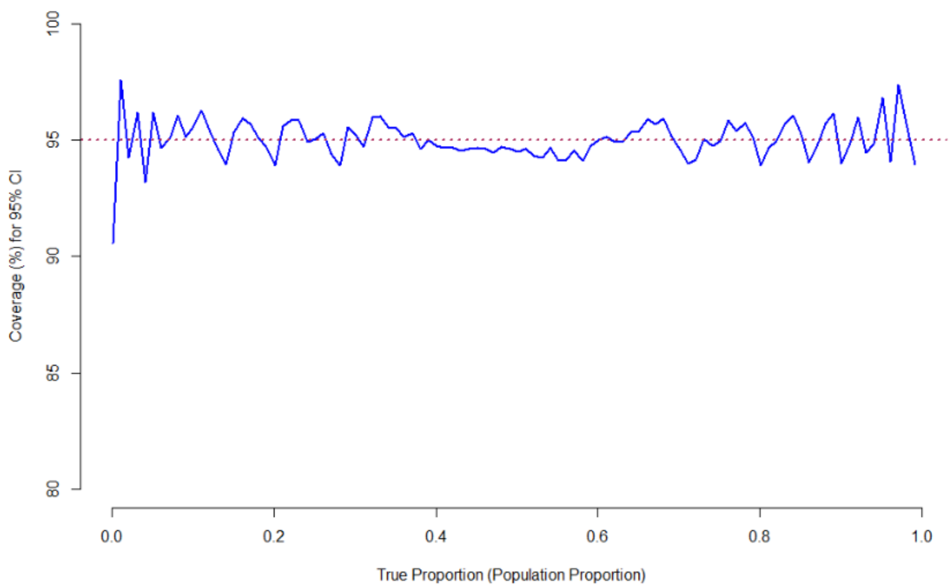
Elsewhere, it has been strongly recommended that instructors present the *Wilson Interval* (see Appendix) as a better alternative:

$$\left(\frac{\hat{p}+\frac{k^2}{2n}}{1+\frac{k^2}{n}} - \frac{k\sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}+\frac{k^2}{4n^2}}}{1+\frac{k^2}{n}},\ \frac{\hat{p}+\frac{k^2}{2n}}{1+\frac{k^2}{n}} + \frac{k\sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}+\frac{k^2}{4n^2}}}{1+\frac{k^2}{n}} \right)$$

where the midpoint of the Wilson Interval is given by $\dfrac{\hat{p}+\frac{k^2}{2n}}{1+\frac{k^2}{n}}$.

See, for example, *Approximate is Better Than "Exact" for Interval Estimation of Binomial Proportions*, by Alan Agresti and Brent Coull (The American Statistician, 52: 2, 119-126, 1998).

The Wilson Interval does not suffer from any of the above-stated defects of the Wald Interval, and in particular its coverage probability is superior:



Simulated coverage probability of the nominal 95% Wilson Interval

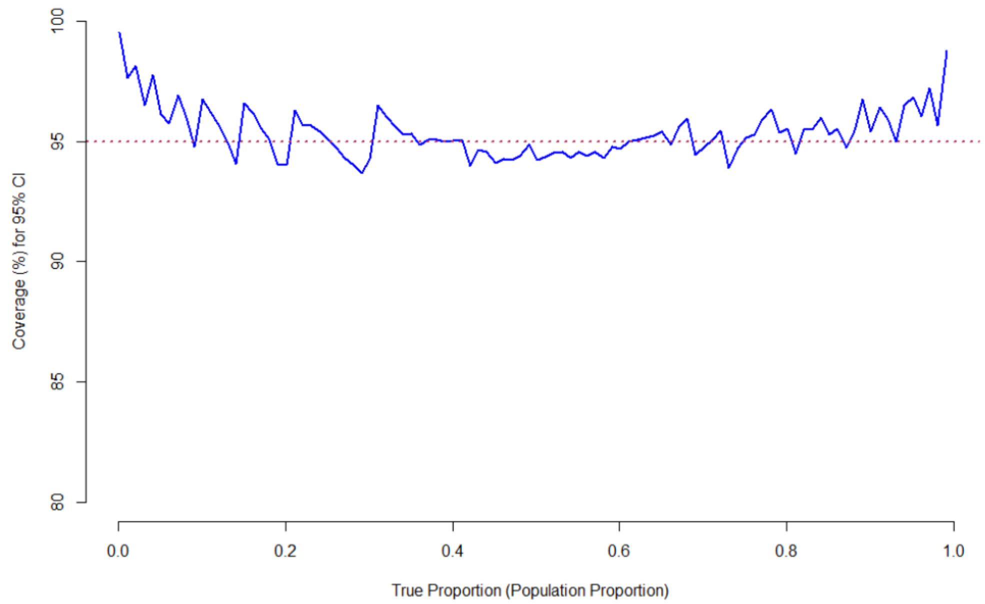
(From *Five Confidence Intervals For Proportions You Should Know About* - Dr Dennis Robert Aug 2020)

Nonetheless, many instructors might hesitate to present such a complicated formula in elementary statistics courses. A simpler alternative is the *Agresti-Coull Interval*:

$$\left(\tilde{p}-k\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n}},\ \tilde{p}+k\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n}}\right)$$

where $\tilde{p}=\dfrac{\hat{p}+\dfrac{k^2}{2n}}{1+\dfrac{k^2}{n}}$ is again the midpoint of the Wilson Interval, given above.

The Agresti-Coull Interval also has none of the above-stated defects of the Wald Interval. In particular its coverage probability is superior, although not as good as the Wilson Interval:



Simulated coverage probability of the nominal 95% Agresti-Coull Interval

(From *Five Confidence Intervals For Proportions You Should Know About* - Dr Dennis Robert Aug 2020)

Conclusion

VCAA has imposed statistical inference onto Mathematical Methods. However, the Wald Interval is the only confidence interval included in the VCE Study Design. Given the obvious defects of the Wald Interval and the ubiquity of CAS technology in VCE mathematics, it is bewildering that superior intervals such as the Wilson Interval and the Agresti-Coull Interval are not considered.

Appendix: Derivation of the Confidence Interval Formulae

Here, we derive some $C\%$ confidence intervals for the population proportion p. The following two assumptions on the sample size n are made:

1. The sample size is 'small' *relative to the population size*. Under this assumption, the distribution of the sample proportion, $\hat{P}$ which is a random variable, can be approximated by the binomial distribution.

2. The sample size is 'large' enough (see below) that the Normal approximation to the Binomial distribution can be used:

$$\hat{P} \sim \text{Norm}\left(\mu = p,\quad \sigma = \sqrt{\frac{p(1-p)}{n}}\right).$$

Note: The standard conditions for the Normal distribution to be a good approximation to the Binomial distribution are $np > 5$ and $n(1-p) > 5$ (or even better, $np > 10$ and $n(1-p) > 10$). It will not be known if these conditions are met, however, because the population proportion p is not known.

In summary, the sample size is assumed to be simultaneously small enough that the binomial approximation can be used, and large enough so that the normal approximation to the binomial distribution can be used.

Now, let

$$\Pr(-k < Z < k) = \frac{C}{100}$$

where Z is the standard normal random variable. Then, with the assumptions above, we can substitute

$$Z = \frac{\hat{P} - \mu}{\sigma} = \frac{\hat{P} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

giving

$$\Pr\left(-k < \frac{\hat{P} - p}{\sqrt{\frac{p(1-p)}{n}}} < k\right) = \frac{C}{100}.$$

The idea now is to somehow 'invert' the inequalities

$$-k < \frac{\hat{P} - p}{\sqrt{\frac{p(1-p)}{n}}} < k \qquad \ldots. (1)$$

in order to 'trap' the population proportion p between lower and upper values L and U:

$$\Pr\left(L\left(\hat{P}, k, n\right) < p < U\left(\hat{P}, k, n\right)\right) = \frac{C}{100}. \qquad \ldots. (2)$$

(2) is not a standard probability statement, because the population proportion p is not a random variable. Rather, (2) defines a *random interval*

$$\left(L\left(\hat{P}, k, n\right),\ U\left(\hat{P}, k, n\right)\right)$$

which contains the fixed but unknown population proportion p with probability $\frac{C}{100}$.

The substitution into this random interval of an observed value $\hat{p}$ of $\hat{P}$ (calculated from a sample) gives the C% confidence interval for p. This constitutes the *realisation* of this random interval. The differing methods of 'inverting' (1), which underlies this realisation, results in differing confidence interval formulas.

The Wald interval

Substitute the approximation $p(1-p) = \hat{p}\left(1-\hat{p}\right)$ into inequality (1) **before** 'inverting' to 'trap' *p*.

Note: This is **not** a realisation of a random interval. It is an approximation that is used solely to avoid the cumbersome algebra in solving exactly for *p*, and is unnecessary when CAS technology is so ubiquitous.

$$-k < \frac{\hat{P}-p}{\sqrt{\dfrac{\hat{p}\left(1-\hat{p}\right)}{n}}} < k \qquad \Rightarrow -k\sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}} < \hat{P}-p < k\sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}}$$

$$\Rightarrow \hat{P}-k\sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}} < p < \hat{P}+k\sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}}.$$

$$L = \hat{P}-k\sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}} \qquad U = \hat{P}+k\sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}}.$$

The *realisation* of this random interval, by substituting the observed value $\hat{p}$ for $\hat{P}$, produces the Wald or 'Standard' *C%* confidence interval met in every introductory statistics textbook:

$$\left(\hat{p}-k\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},\ \hat{p}+k\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right).$$

The Wilson interval

Inequality (1) is exactly inverted to 'trap' *p*:

$$-k < \frac{\hat{P}-p}{\sqrt{\dfrac{p(1-p)}{n}}} < k \qquad \Rightarrow -k\sqrt{\frac{p(1-p)}{n}} < \hat{P}-p < k\sqrt{\frac{p(1-p)}{n}} \qquad \Rightarrow \left(\hat{P}-p\right)^2 < k^2\,\frac{p(1-p)}{n}$$

$$\Rightarrow \left(\hat{P}\right)^2 - 2\hat{P}\,p + p^2 < \frac{k^2 p - k^2 p^2}{n} \qquad \Rightarrow \left(\hat{P}\right)^2 - 2\hat{P}\,p + p^2 - \frac{k^2 p}{n} + \frac{k^2 p^2}{n} < 0$$

$$\Rightarrow \left(1+\frac{k^2}{n}\right)p^2 - \left(2\hat{P}+\frac{k^2}{n}\right)p + \hat{P}^2 < 0.$$

This is a standard quadratic form in the 'variable' p where the value of n is known from the sample size and the quantile k is calculated from the Standard Normal distribution for the particular confidence level $C\%$ required:

$$\Pr(-k < Z < k) = \frac{C}{100}.$$

The upper and lower limits U and L are then given by the solutions to $\left(1+\frac{k^2}{n}\right)p^2 - \left(2\hat{P}+\frac{k^2}{n}\right)p + \hat{P}^2 = 0$:

$$p = \frac{\left(2\hat{P}+\frac{k^2}{n}\right) \pm \sqrt{\left(2\hat{P}+\frac{k^2}{n}\right)^2 - 4\left(1+\frac{k^2}{n}\right)\hat{P}^2}}{2\left(1+\frac{k^2}{n}\right)} = \frac{\left(2\hat{P}+\frac{k^2}{n}\right) \pm \sqrt{4\left(\hat{P}+\frac{k^2}{2n}\right)^2 - 4\left(1+\frac{k^2}{n}\right)\hat{P}^2}}{2\left(1+\frac{k^2}{n}\right)}$$

$$= \frac{\left(2\hat{P}+\frac{k^2}{n}\right)}{2\left(1+\frac{k^2}{n}\right)} \pm \frac{\sqrt{\left(\hat{P}+\frac{k^2}{2n}\right)^2 - \left(1+\frac{k^2}{n}\right)\hat{P}^2}}{1+\frac{k^2}{n}} = \frac{\hat{P}+\frac{k^2}{2n}}{1+\frac{k^2}{n}} \pm \frac{\sqrt{\hat{P}^2 + \hat{P}\frac{k^2}{n} + \frac{k^4}{4n^2} - \hat{P}^2 - \frac{k^2}{n}\hat{P}^2}}{1+\frac{k^2}{n}}$$

$$= \frac{\hat{P}+\frac{k^2}{2n}}{1+\frac{k^2}{n}} \pm \frac{\sqrt{\hat{P}\frac{k^2}{n} + \frac{k^4}{4n^2} - \frac{k^2}{n}\hat{P}^2}}{1+\frac{k^2}{n}} = \frac{\hat{P}+\frac{k^2}{2n}}{1+\frac{k^2}{n}} \pm \frac{\sqrt{\frac{k^2}{n}\hat{P}\left(1-\hat{P}\right) + \frac{k^4}{4n^2}}}{1+\frac{k^2}{n}}$$

$$= \frac{\hat{P}+\frac{k^2}{2n}}{1+\frac{k^2}{n}} \pm \frac{k\sqrt{\frac{\hat{P}\left(1-\hat{P}\right)}{n} + \frac{k^2}{4n^2}}}{1+\frac{k^2}{n}}.$$

$$L = \frac{\hat{P}+\frac{k^2}{2n}}{1+\frac{k^2}{n}} - \frac{k\sqrt{\frac{\hat{P}\left(1-\hat{P}\right)}{n} + \frac{k^2}{4n^2}}}{1+\frac{k^2}{n}}. \qquad U = \frac{\hat{P}+\frac{k^2}{2n}}{1+\frac{k^2}{n}} + \frac{k\sqrt{\frac{\hat{P}\left(1-\hat{P}\right)}{n} + \frac{k^2}{4n^2}}}{1+\frac{k^2}{n}}.$$

The *realisation* of the resulting random interval produces the $C\%$ confidence interval for the population proportion p known as the Wilson Interval:

$$\left(\frac{\hat{p}+\frac{k^2}{2n}}{1+\frac{k^2}{n}} - \frac{k\sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}+\frac{k^2}{4n^2}}}{1+\frac{k^2}{n}},\ \frac{\hat{p}+\frac{k^2}{2n}}{1+\frac{k^2}{n}} + \frac{k\sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}+\frac{k^2}{4n^2}}}{1+\frac{k^2}{n}} \right).$$

The Agresti-Coull Interval

Substitute the midpoint $\tilde{p}=\dfrac{\hat{p}+\frac{k^2}{2n}}{1+\frac{k^2}{n}}$ of the Wilson Interval into inequality (1) for *p* before inverting.

The *realisation* of the resulting random interval produces the Agresti-Coull Interval:

$$\left(\tilde{p}-k\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n}},\ \tilde{p}+k\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n}} \right).$$