Application Of Wald And Score Confidence Interval For A Population Proportion

James Issos[1], Rishi Raj Subedi[2], Daniel Osborne[3], Elise Simmons[4]
[1]*Department of Mathematics, Florida A&M University, USA*
[2]*Department of Mathematics and Statistics, Georgia State University, USA*
[3]*Department of Mathematics, Florida A&M University, USA*
[4]*Department of Mathematics, Florida A&M University, USA*

Abstract: *Interval estimation of a population proportion causes problems because some procedures such as the Score or Exact are complicated and the commonly used Wald interval is too inaccurate in certain situations. However, in other instances the Wald procedure is reasonably accurate so that it can be used in these cases. We are presenting a simple method for determining when and when not to use the Wald method.*
Keywords: *Confidence interval; Discrete distribution; Binomial distribution; Score test; Wald procedure.*

I. Introduction

A basic analysis in statistical inference is constructing a confidence interval for a binomial parameter p, Cai [1]. The simplest interval which is almost universally used is,

$$\hat{p} \pm z_{\frac{\alpha}{2}}[\frac{\hat{p}(1-\hat{p})}{n}]^{\frac{1}{2}} \tag{1.1}$$

where $\hat{p}$ is the sample proportion, n is the sample size and $z_{\frac{\alpha}{2}}$ denotes the $1-\frac{\alpha}{2}$ quantile of the standard normal distribution. For instance α = 0.05 for a 95 % confidence interval, $\alpha = 0.10$ for a 90% confidence interval, etc. This interval is derived from the Wald large sample confidence interval and is commonly referred to as the Wald interval.

So it seems at first glance that the problem is simple and has a clear solution. Actually the problem is a difficult one with several complexities. It is widely recognized that Wald interval coverage probability is poor for p near 0 or 1. It is known that the Wald interval performs poorly unless n is large, Blyth and Still [2] . Most statistics books and journals take this into account by requiring that this interval should be used only when $\min(np, n(1-p))$ is at least 5 or 10, Brown, Cai, and Das Gupta [3], Agresti [4], and Triola [5].

A considerable literature exists about this and other less common methods for constructing a confidence interval for p. Santner and Duffy [6] and Vollset [7] reviewed a variety of methods, Newcombe [8]. One of the methods is the Clopper-Pearson "exact" interval [9]. This method is widely used and has the advantage of a coverage probability of at least $1-\alpha$ for every possible value of p. The Score method, Wilson [10] discussed by Agresti and Coull [11], is arguably the best procedure for constructing a confidence interval for a population proportion. Guan [12] introduced the generalized score method which computes easily and reduces the spike fluctuations of the score method. Also Bayesian methods are effective for constructing confidence intervals for a population proportion. In addition other effective procedures such as the Arcsin, Logit and Jeffres prior intervals are discussed in Brown, Cai, Das Gupta [3]. The Jeffres prior interval is a special case of a Bayes procedure with a non-informative prior. And Bayes procedures with a non- informative prior have a good track record in constructing confidence intervals for p; see Wasserman [13]. Wang [14] discusses methods for constructing the smallest exact confidence intervals. Zao, Huang and Zhang [15] use the Score interval to construct a confidence interval for a linear function of binomial proportions.

However, most effective procedures are too complicated to use in statistics. Therefore Agresti and Coull [11] introduced the Adjusted Wald (AC) procedure. The AC method consists of adding two successe and two failures to the data and then proceding as in the Wald interval. This method is simple, easy to use and accurate. The accuracy of the AC procedure is due to its midpoint and width being almost the same as those of the Score procedure. Actually the AC interval is a simplified version of the Score interval, Subedi and Issos [16] and Wei, Xu, and Wangli [17].

At the present time the Wald interval is almost exclusively used in everyday practical statistics. Some reasons for its popularity are that it is easy to motivate and easy to use. Under the right conditions such as $np(1-p) \geq 10$, it is reasonably accurate.

II. The Wald And Score Intervals

The Wald confidence interval for a population p is by far the most commonly used confidence interval for a population proportion. It is

$$\hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \tag{2.1}$$

Here $\hat{p}$ is the sample proportion, n is the sample size and z is the z-value for the confidence level. That is, z = 1.960 for a 95 % confidence level, z = 1.645 for a 90 % confidence level, etc. It is well known that the Wald interval performs poorly in many situations. In fact, many people think the Wald Interval should be discarded entirely.

We will give our opinion on this by comparing the Wald interval with the Wilson (or Score) confidence interval. The Wilson interval is derived by using

$$\hat{p} + z\sqrt{\frac{p(1-p)}{n}} = p \tag{2.2}$$

Here p is the upper end of the Score interval. If we rewrite this as

$$\hat{p} - p = +z\sqrt{\frac{p(1-p)}{n}}$$

and then square both sides and simplify the result is a complicated quadratic equation in p. The solutions of this quadratic equation yield the end points of the Wilson (score) interval. The reader can find this solution in Agresti and Coull [11]. Here we only summarize the results.

The midpoint of this interval is the weighted average

$$C\hat{p} + \frac{1-C}{2}$$

. Here $C = \frac{n}{n+z^2}$.

The term that is added and subtracted from the midpoint to form the score confidence interval is

$$z\sqrt{\frac{C\hat{p}(1-\hat{p})+\frac{(1-C)}{4}}{n+z^2}}$$

Thus the Score confidence interval can be written as

$$C\hat{p} - (1-C)(0.5) \pm z\sqrt{\frac{C\hat{p}(1-\hat{p})+\frac{(1-C)}{4}}{n+z^2}} \tag{2.3}$$

Here $C = \frac{n}{n+z^2}$ and $1 - C = \frac{z^2}{n+z^2}$

Note that the mid-point of the Wilson (Score) interval is a weighted average of the sample proportion and 0.5. However, the mid-point of the Wald interval is always $\hat{p}$.

This brings us to the main idea of this article. The Wald interval is easy to remember and easy to use. However, it is inaccurate in many situations. But the score interval is always accurate but hard to use, derive and remember. So a way around this dilemma is to use the Wald procedure only when the center is close to the center of the Score interval.

It would be nice if there was a 90 % confidence interval for a population proportion p that covered each value of p with a probability of 0.90. But no such procedure exists. The most one can expect is that the average coverage of different values of p is close to 0.90. The score confidence interval is excellent in this regard.

Perhaps a digression is useful here. The Adjusted Wald (AC) interval is formed by adding 2 successes and two failures to the sample data and then proceeding as in the Wald formula. Why add two successes and two failures? Because for a 95% confidence interval, adding two successes and two failures makes the AC interval have almost the same center and end points as the Score interval, Subedi and Issos [18]. So it seems such that an interval should have good coverage properties. Computer simulations support this.

The reasoning here is similar. Under some conditions the Wald interval should have almost the same center and end points as the Wilson (Score) interval. Thus we present a formula for the difference between the centers of the Wald and Score intervals. This formula is

$$\frac{z^2}{n+z^2} \mid \hat{p} - 0.5 \mid = D \tag{2.4}$$

Equation (2.4) shows that for certain values of z, n, and $\hat{p}$, the midpoints are almost the same. In this case the Wald interval should be reasonably accurate. Also if D is small the width of the Wald interval should be nearly the same as the width of the Score interval. Our reasoning is that under the right conditions (D small, say less than 0.009) the Wald interval will be almost accurate as the Score interval. Since the Wald interval is easier to remember and use than the Score interval why not use the Wald procedure when conditions are right. The following table illustrates this:

Table 1: Average coverage of the Wald, and other procedure with 0.3 ≤ p ≤ 0.7 and selected confidence levels.

Method	Sample size(n)	Confidence level 80%	90%	95%
Wald	20	0.778	0.874	0.924
Score	20	0.799	0.900	0.951
Adjusted Wald	20	0.822	0.912	0.956
Exact	20	0.867	0.939	0.971
T-Wald	20	0.806	0.898	0.945
Wald	30	0.785	0.883	0.933
Score	30	0.799	0.900	0.951
Adjusted Wald	30	0.815	0.909	0.953
Exact	30	0.857	0.932	0.968
T-Wald	30	0.801	0.899	0.947
Wald	40	0.790	0.887	0.938
Score	40	0.799	0.900	0.951
Adjusted Wald	40	0.811	0.906	0.952
Exact	40	0.850	0.928	0.966
T-Wald	40	0.803	0.899	0.947
Wald	70	0.794	0.893	0.943
Score	70	0.799	0.900	0.950
Adjusted Wald	70	0.807	0.904	0.951

Exact	70	0.839	0.922	0.962
T-Wald	70	0.801	0.900	0.949
Wald	100	0.796	0.895	0.945
Score	100	0.800	0.900	0.950
Adjusted Wald	100	0.805	0.902	0.951
Exact	100	0.833	0.919	0.961
T-Wald	100	0.801	0.900	0.949

Table 1: Average coverage of the Wald, and other procedure with $0.3 \leq p \leq 0.7$ and selected confidence levels.

In table 1, the T-Wald procedure is the Wald procedure with the "Z" replaced with the T-value with n-1 degrees of freedom and "n" replaced by "n-1". Table 1 shows that under favorable conditions the Wald procedure performs well with regard to average coverage. As expected the Score procedure has the best average coverage. Surprisingly, the T-Wald also has good average coverage under the conditons given in Table 1 (i.e. $0.3 \leq p \leq 0.7$), $n \leq 100$ and confidence level ≤ 95%. Table 1 shows that the Score procedure has the best average coverage under all of the conditons given in Table 1. However, in many cases the Wald procedure is almost as accurate as the Score procedure. In these cases the simplicity of the Wald method makes it a reasonable altrnative to the Score or any other method. For instance, for $n \geq 70$, the Wald procedure is a reasonable alternative to any method. Even for $n \geq 30$ and $z \leq 1.645$, the Wald could be used.

Table 1 makes an argument for the use of the Wald interval under certain condition. While some procedure such as the Score are more accurate than the Wald, the Wald procedure has satisfactory accuracy. Also, the simplicity of the Wald makes it a good choice. Alternatively, the T-Wald would not be a bad choice.

III. Conclusion

The Wald procedure is probably the easiest method to use to construct a confidence interval for a population proportion. In practice, there are many situations where almost any method gives at least satisfactory results. So why not use the Wald or alternatively, one could use the T-Wald, which is only slightly more complicated than the Wald.

Certainly, the Wald procedure should not be used in many situations with small sample size n and/or the estimate of p being outside of the interval ($0.3 \leq p \leq 0.7$) however, this applies to most confidence interval procedures for a population proportion. For instance, some point out that the Score 95% procedure does not cover every point with at least 0.95 probability. In the above work, we presented a simple method for determining when and when not to use the Wald method based on the statistic D.

References

[1] Cai, T.T. (2005). *One-sided Confidence Intervals in Discrete Distributions*. *Journal of Statistical Planning and Inference, 131, 63-88.*

[2] Blyth, C. R., & Still, H. A. (1983). *Binomial Confidence Intervals*. Journal of the American Statistical Association. 78, 108-116.

[3] Brown, L. D., Cai, T., & Das Gupta, A. (2001). *Interval Estimation for a Binomial Proportion*. Statistical Science, 16, 101-133.

[4] Agresti, A. (1996). *An introduction to Categorical Data Analysis*. New York: Wiley.

[5] Triola, M. F. (2006). *Elementary Statistics*, 10th ed.

[6] Santner, T. J. (1998). *A Note on Teaching Binomial Confidence Intervals*. Teaching Statistics, 20, 20-23.

[7] Vollset, S. E. (1993). *Confidence intervals for a binomial proportion*, Statistics in Medicine, 12, 809-824.

[8] Newcombe, R. G. (1998). *Two-sided Confidence Intervals for the Single Proportion: Comparison of SevenMethods*. Statistics in Medicine, 17, 857-873.

[9] Clopper, C. J. and Pearson, E. S. (1934). *The Use of Confidence or Fiducial Limits Illustrated in the Cas of the Binomial*. Biometrika, 26, 404-413.

[10] Wilson, E. B. (1927).*Probable Inference the Law of Succession and Statistical Inference*. Journal of the

American Statistical Association, 22, 209-212.

[11] Agresti, A., & Coull, B. A. (1998). *Approximate is Better than Exact for Interval Estimation of Binomial Proportions*. The American Statistician, 52, 119-126.

[12] Guan,Yu (2012). *A generalized score CI for a binomial proportion*. *Journal of Statistical Planning and Inference, 142, 785-793.*

[13] Wasserman, L. (1991). *An inferential interpretation of default priors*. Technical report, Carnegie-Mellon University.

[14] Wang, W. (2006). *Smallest confidence intervals for one binomial proportion. Journal of Statistical Planning and Inference, 136, 4293-4306*

[15] Zou, G. Y., Huang, W., & Zhang, X. (2009). *A note on confidence interval estmation for a linear function of binomial proportions*. Computational Statistics and Data Analysis, 53, 1080-1085.

[16] Subedi, R., & Issos, J. (2019). *An Improved Confidence Interval. American Review of Mathematics and Statistics. 7(1), 8-11*.

[17] Wei, Yu, Xu , Guo, & Wangli (2012). *An Improved Score Interval with a Modified Mid-Point for a Binomial Proportion*. Journal of Statistical Planning and Inference, 142, 785-793.

[18] Subedi, R., & Issos, J. (2019).*Determining the End Points of The Score Confidence Interval Using Computer Program. Asian Journal of Probability and Statistics. 4(1), 1-7.*

.