Minitab

4 New Methods to Improve Your Estimation of a Single Population Proportion in Minitab

A common problem in basic statistics is the estimation of the proportion of individuals with a certain characteristic of interest in a population. For example, a quality engineer may want to estimate the proportion of defects in a large batch of mass-produced units on a given day; a medical scientist may want to investigate the proportion of individuals in some community who were vaccinated against a specific pathogen but experienced the related disease nonetheless; a campaign manager may be interested in the proportion of registered voters who intend to vote for their candidate.

The best-known interval estimation methods for this problem are the textbook normal approximation method referred to as the Wald confidence interval (CI) and the Clopper-Pearson exact (1934) CI. On one hand, the Wald CI is extremely liberal in that the actual confidence level (or coverage probability) of the CI is well below the targeted nominal level, particularly when the true proportion is close to 0 or 1 (see Figure 1). On the other hand, the exact Clopper-Pearson CI is excessively conservative in that the actual confidence level (or coverage probability) of the CI is well above the targeted nominal level. Both of these methods should no longer be used for any practical applications (see Agresti-Coull, 1998; Brown et al., 2001).

In recent years, however, they have played a major role in the development of better CI methods with better intermediate coverage probabilities. For example, Agresti-Coull approximate CI is an adjustment on the Wald CI; the Blaker (2000, 2001) exact CI uses Clopper-Pearson confidence bounds as starting estimates in an iterative numerical algorithm. Mindful of these newly improved methods, Minitab has updated the statistical tool for estimating a single population proportion to include the following 4 methods: the adjusted Blaker CI and test methods, the Wilson/score CI and test methods (with and without a continuity correction), and the Agresti-Coull CI and test methods. In addition, for each of these methods, Minitab ensures that the CI and test yield consistent results.

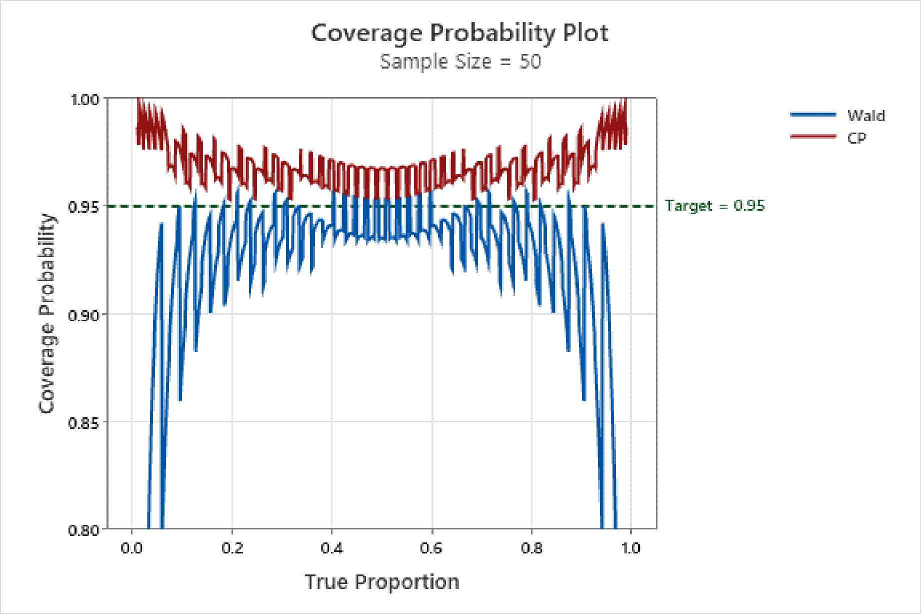

Figure 1: Comparison of coverage probabilities for the Wald CIs and Clopper-Pearson (CP) CIs as functions of the true proportion when the sample size is 50. The graph illustrates that the Wald CIs and the Clopper Pearson CIs are excessively liberal and conservative, respectively, particularly when the true proportion is near 0 or 1. Assuming that the true proportions are uniformly distributed in the interval (0, 1), the mean coverage probabilities based on a sample of size 50 are 0.901 and 0.969 for the Wald CI and Clopper-Pearson CI, respectively.

Meet the 4 New Methods of Estimation

The 4 new methods consist of 1 exact CI and test method called the adjusted Blaker method and 3 approximate CI and test methods including the Wilson/score (Wilson) method, Wilson/score with Yates' continuity correction (Wilson CC) method, and the Agresti-Coull (AC) method. An exact method in this context means that there is no approximation used in the derivation of the method as opposed to approximate methods that are obtained using some forms of normal approximation procedures.

1. The Adjusted Blaker Method

The adjusted Blaker method, due to Klaschka and Reiczigel (2021), is a modification of the Blaker (2000, 2001) exact CI and testing methods. The modification addresses the computation-intensive nature of the original Blaker algorithm and the occasional inconsistencies between its CI and test results. Like the original Blaker CI, the resulting adjusted CI is exact, nested, and is contained in the Clopper-Pearson CI. As a result, the adjusted Blaker CI is less conservative than the Clopper-Pearson CI. The CI is nested in the sense that a CI with a higher confidence level always contains a CI with a lower confidence level. For example, a two-sided 95% (adjusted) Blaker CI always contains the corresponding two-sided 90% CI. Nestedness is an appealing property of exact CI methods derived from a discrete distribution such as the binomial. For example, the Clopper-Pearson CI is nested. There are, however, available exact CI methods that are not necessarily nested. For example, the so-called Blyth-Still-Casella CI (Blyth and Still, 1983; Casella, 1986) is guaranteed to be the shortest exact CI but is not nested. Crow (1956) CI is also not nested. The calculations of CIs based on Blaker or the adjusted Blaker method are more complex than the aforementioned classical CI methods because they require numerical algorithms. With the current innovations in computing technology, however, we should no longer shy away from implementing complex algorithms that yield better results. Figure 2 illustrates the improvements of the adjusted Blaker CI over the Clopper-Pearson CI.

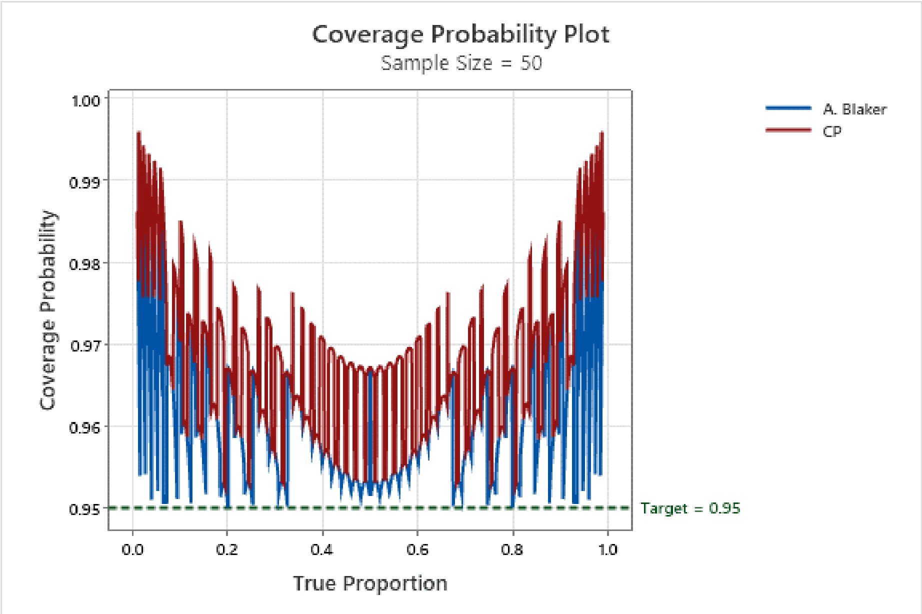

Figure 2: Comparison of coverage probabilities for the adjusted Blaker (A. Blaker) CIs and the Clopper-Pearson (CP) CIs as functions of the true proportion when the sample size is 50. The graph indicates that the coverage probability of the Clopper-Pearson CI is at least that of the adjusted Blaker CI. This is consistent with the fact that adjusted Blaker CIs are contained in Clopper-Pearson CIs. For any given sample of size 50 the average coverage probabilities (assuming that the true proportion is uniformly distributed on the unit interval) are 0.960 and 0.969 for the adjusted Blaker CI and Clopper-Pearson CI, respectively.

2. The Wilson and Wilson CC methods

The Wilson (1927) CI method is derived as the CI that corresponds to the score test, the test that uses the null standard error, $\sqrt{p_0(1-p_0)/n}$, as opposed to the classical standard error $\sqrt{\hat{p}(1-\hat{p})/n}$, on the denominator of the test statistic. For this reason, it is also referred to as the Wilson/score CI. Its actual coverage probability can be less or greater than the targeted nominal level, but remain close to it, except when the true proportion is close to 0 or 1 (see Figure 3). An adjustment can be made to make the Wilson/score CI conservative by using Yates' continuity correction. Minitab provides both CI versions (with and without Yates' continuity correction) and their matching hypothesis tests.

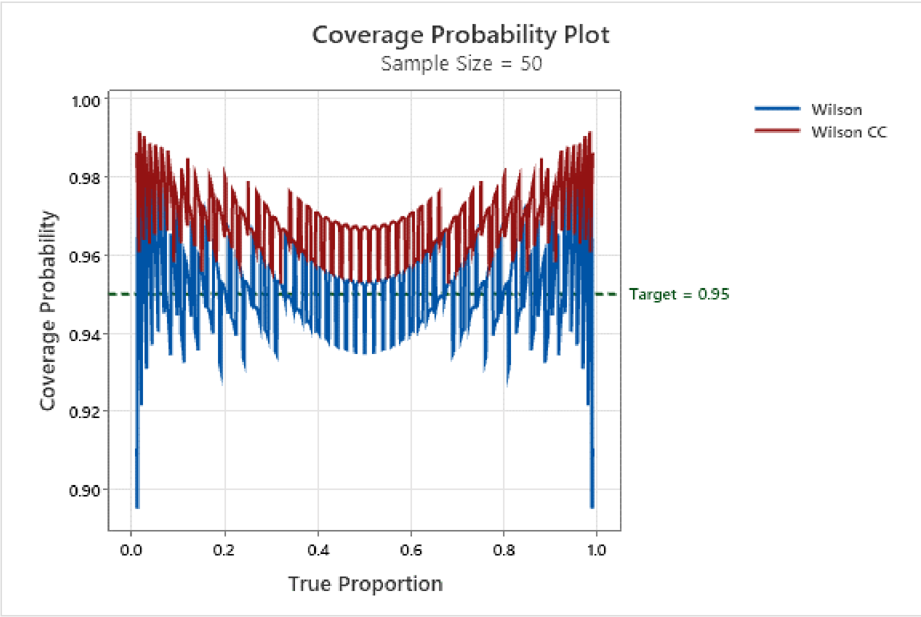

Figure 3: Comparison of coverage probabilities for the Wilson/score (Wilson) CIs and the Wilson/score with Yates' continuity correction (Wilson CC) as functions of the true proportion when the sample size is 50. The graph shows that Wilson CC CIs are always conservative while Wilson CIs are conservative and liberal depending on the magnitude of the true proportion. In particular, Wilson CIs tend to be too liberal when the true proportion is very close to 0 or 1. For any given sample of size 50, the mean coverage probabilities are 0.952 and 0.969 for the Wilson CI and Wilson CC CI, respectively.

3. The Agresti-Coull Method

The Agresti-Coull CI is obtained from an adjustment of the excessively liberal classical Wald CI. The resulting CI has similar coverage properties as the Wilson CIs, but a bit more conservative in general. Furthermore, the two types of CIs have the same midpoint, but Wilson CIs are always contained in Agresti-Coull CIs. As illustrated in Figure 4, they have essentially the same coverage probabilities when the true proportion is moderate. The Agresti Coull CI, however, is generally less liberal when the true proportion is close to 0 or 1. As shown in Figure 4, for a sample of size 50 the Agresti-Coull CI becomes conservative as the true proportion approaches 0 or 1. Another appeal of the Agresti-Coull CI is its implementation simplicity inherited from the Wald CI. Moreover, it is easy to teach and remember, particularly when the confidence level is 95%. For this confidence level, it is commonly referred to as the "add 2 successes and 2 failures" CI method as a memorandum to the adjustment made to the Wald CI to derive it.

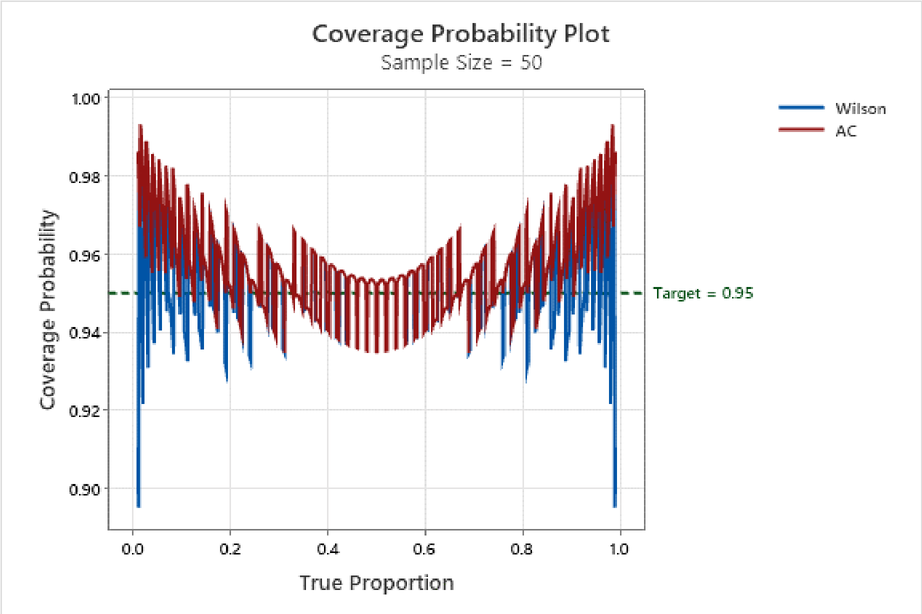

Figure 4: Comparison of coverage probabilities for the Wilson/score (Wilson) CIs and the Agresti-Coull (AC) CIs as functions of the true proportion when the sample size is 50. The graph indicates that the two methods yield essentially the same coverage probabilities for moderate values of the true proportion in the interval (0.3, 0.75). For values of the true proportion close to 0 or 1, however, Wilson CIs are liberal while Agresti-Coull CIs are conservative. This is consistent with the fact that Wilson CIs are contained in Agresti-Coull CIs. For any given sample of size 50 the mean coverage probabilities are 0.952 and 0.958 for the Wilson CI and Agresti-Coull CI, respectively.

Some Simple Illustrative Examples

A quality engineer in a mass production factory has selected a random sample of 1465 mass-produced goods on a given day. After independent testing of the 1465 items, 53 were found to be defective. The engineer wants to know if the proportion of defective items produced on the given day is significantly different from 2.5%.

Although this problem is cast as a statistical hypothesis test question, applied statisticians are increasingly encouraged to include in the analysis result a point estimate and a confidence interval along with the p-value of the test. Minitab follows this convention as much as possible, particularly in the Basic Statistics Modules. For example, using Minitab, the analysis results for the above question based on the adjusted Blaker method is as follows.

Test and CI for One Proportion

Method

p	Event proportion
Method	Adjusted Blaker's exact method

Descriptive Statistics

N	Event	Sample p	95% CI for p
1465	53	0.036177	(0.027353, 0.046822)

Test

Null hypothesis	H_0: p = 0.025
Alternative hypothesis	H_1: p ≠ 0.025

P-Value
0.009

A similar output can be generated for each of the 4 methods. The analyses results for all these methods are summarized in the following table.

Method	95% CI	P-Value of the matching test
Adjusted Blaker	(2.74%,4.68%)	0.009
Wilson with Yates' correction	(2.75%,4.74%)	0.008
Wilson/score	(2.78%,4.70%)	0.006
Agresti-Coull	(2.77%,4.74%)	0.007

In this example, all the methods yield the same conclusion that the % defective differs from 2.5% at the 0.05 level of significance since all the p-value are less than 0.05. The confidence intervals and corresponding p-values for all the methods are similar, in part, because the size of the sample is very large. In addition, the CI for each method does not cover the hypothesized proportion value (2.5%), which is consistent with the p-value of each of the matching hypothesis tests.

In the above example, suppose now that the quality engineer tested only 50 items and found that 2 were defective. Moreover, suppose that the engineer wanted to know if the proportion of defectives was significantly different from 1.0%. The analyses results for each of the methods are summarized in the following table.

Method	95% CI	P-Value of the corresponding test
Adjusted Blaker	(0.72%,13.35%)	0.089
Wilson with Yates' correction	(0.70%,14.86%)	0.155
Wilson/score	(1.10%,13.46%)	0.033
Agresti-Coull	(0.34%,14.22%)	0.124

In this case, only the Wilson/score method yields a significant conclusion in that the % defective differs from 1.0% at the 0.05 level of significance. At the same level of significance, all the other methods yield the opposite conclusion of insufficient evidence to decide that there is a difference. The inconsistencies in the results across methods are, in large part, due to the size of the sample being moderate. On average, the coverage probabilities of these methods approach the nominal level as the sample size increases (see Figure 5 below). For small to moderate sample designs, however, the disparities in the mean coverage probabilities associated with each method are more pronounced causing the corresponding CIs to have noticeably different widths. An important question, however, is which result to report to the boss? There is no straightforward answer to this question because a better one may depend on factors such as prior knowledge of the magnitude of the true proportion or even the area of applications. We will provide some rough general guidelines in the next section.

Conclusion

Figures 2, 3, and 4 show that the 4 CI methods, Adjusted Blaker, Wilson, Wilson CC, and the Agresti-Coull have different coverage probability properties. The Wilson CC is the most conservative followed by the adjusted Blaker. The Agresti Coull and the Wilson are often liberal and conservative depending on the magnitude of the true proportion. Overall, the Wilson CI method is the most liberal of all 4 methods. Moreover, Figure 5 indicates that, on average, all 4 methods are conservative with the Wilson CC being the most conservative followed by the adjusted Blaker, the Agresti-Coull and the Wilson methods. For a given problem, the appropriate method to choose may depend upon the particular application, the sample size, and whether or not some prior knowledge of the magnitude of the true proportion is available.

For example, regulatory agencies are often interested in conservative methods so as to protect consumers. A too conservative method, however, may yield stringent regulations while a too liberal method may yield loose regulations. In general, for moderate to large sample designs the adjusted Blaker or the Willson CC method may be good choices. For small sample designs the Wilson or the Agresti-Coull may be appropriate if a prior knowledge of the true proportion value is available. Such prior knowledge is often based on prior similar experiences or pilot small study specifically designed to obtain a rough estimate of the proportion. For example, in the quality control area of applications where the proportion of defects are typically close to 0, the method to choose depends on whether or not the investigator wants a conservative result. A conservative result can be based on Agresti-Coull method while a liberal result can be based on Wilson/score method. Lastly but more importantly, it is always good practice to plan ahead for sample size so as to guard against the type II error or to control the width of CIs. Minitab has “an app for that too”. Tools are available in Minitab to determine adequate sample size to control the type II error associated with hypothesis tests or to control the width of CIs.

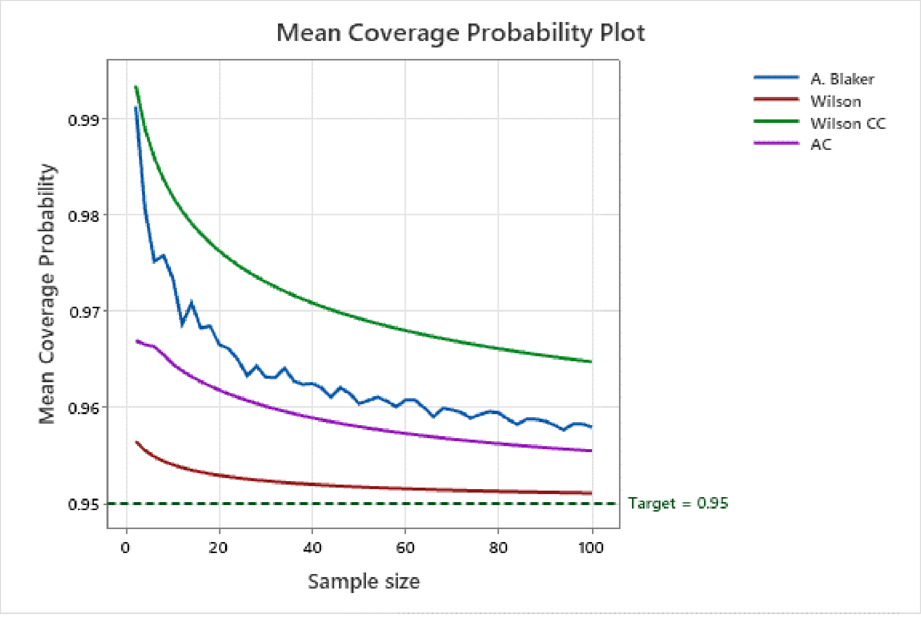

Figure 5: Mean coverage probability of all 4 CI methods as functions of sample size. The mean coverage probabilities are computed assuming that the true proportion is uniformly distributed on the unit interval. The Mean coverage curves illustrate that, on average, the Wilson/score with Yates’ continuity correction (Wilson CC) CI method is the most conservative followed by the adjusted Blaker (A. Blaker), the Agresti-Coull (AC), and the Wilson/score (Wilson) CI methods. The mean coverage probability curves approach the targeted nominal coverage level as the sample size increases. In addition, while the mean coverage curves for the approximate methods (Wilson CC, Wilson, Agresti-Coull) are smooth, the coverage curve for the exact adjusted Blaker have some wiggling movements as it approaches the nominal coverage. This indicates, perhaps, that the adjusted Blaker CI method can be further improved.

Reference

Agresti, A. and Coull, B. A. (1998). Approximate is better than “Exact” for interval Estimation of Binomial Proportion. The American Statistician 52, 119—125

Blaker, H. (2000). Confidence Curves and Improved Exact Confidence Intervals for Discrete Distributions. The Canadian Journal of Statistics, 28, 783—798

Blaker, H. (2001). Corrigenda: Confidence curves and improves exact confidence intervals for discrete distributions. The Canadian Journal of Statistics, 29, 681.

Blyth, C. R. and Still, H. A. (1983). Binomial Confidence Intervals. Journal of the American Statistical Association 78, 108—116.

Brown, L. D., Cai, T. and Das Gupta, A. (2001). Interval Estimation for a Binomial Proportion. Statistical Science 16, 101—133.

Casella, G., 1986. Refining binomial confidence intervals. Canad. J. Statist. 14, 113–129.

Clopper, C. J. and Pearson, E. S. (1934). The Use of Confidence or Fiducial Limits Illustrated in the Case of Binomial. Biometrika 26, 404—413

Crow, E.L., 1956. Confidence intervals for a proportion. Biometrika 43, 423–435.

Klaschka, J. and Reiczigel, J. (2021). On matching confidence intervals and tests for some discrete distributions: methodological and computational aspects, Computational Statistics, Springer, vol. 36(3), 1775-1790.

Wilson E. B. (1927) Probable Inference, the Law of Successions and Statistical Inference. J. Amer. Statist. Assoc. 22, 209—21