How confident are you, really?

Simulated vs. Stated Confidence in One-Proportion Intervals

Devyn Willey

Mentor: Dr. Lindsey Bell

Department of Mathematics and Statistics
Coastal Carolina University

Introduction

According to the CCU *Fact Book*, we know that 49.84% of students were out-of-state and 6.52% of students were graduate students in the year of 2021. These values are known for the entire population of CCU students and therefore called **parameters** (p). However, if we take a sample of say $n = 50$ students, we would compute a **statistic** $(\widehat{p})$ to estimate the parameter. It is unlikely that the value for the statistic would equal the value for the population parameter. In fact, even different samples would yield different statistics. Instead of using this one estimate from our sample for the population, we create an estimate with an interval of values, called a **confidence interval**.

Confidence intervals estimate population parameters with a certain level of confidence. There are several options for constructing a confidence interval for the parameter p. With the help of statistical computing software, e.g., *RStudio*, one can simulate samples from a known population and compute different confidence intervals for a given problem simultaneously and significantly faster than computing them by hand. Using this computational power, we can compare the performance of the different intervals for the population proportion.

Simulation

In order to study achieved vs. stated confidence we utilize the following steps for simulation:

1. Generate data x from the known population, $X \sim \text{Binomial}(n, p)$.
2. Compute the desired confidence interval(s) for p, using the simulated data, $\widehat{p} = x/n$.
3. Determine whether or not the known parameter exists within the interval's bounds.
4. Repeat steps 1 - 3 thousands of times to estimate what percentage of time the interval contains the population parameter.
5. Compare this simulated level of confidence to the stated level of confidence (95%).

We explore the impact of p and n on the performance of the different intervals by considering variations of p and n.

Confidence Intervals

Confidence intervals compared in this simulation study follow:

- **Wald (Asymptotic):** $\widehat{p} \pm z_{1-\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$, where $\widehat{p} = \frac{x}{n}$
- **Agresti-Coull:** $\widetilde{p} \pm z_{1-\alpha/2}\sqrt{\frac{\widetilde{p}(1-\widetilde{p})}{n+z^2_{1-\alpha/2}}}$, where $\widetilde{p} = \frac{x+(z^2_{1-\alpha/2})/2}{n+z^2_{1-\alpha/2}}$
- **Wilson:** $\widetilde{p} \pm \frac{z_{1-\alpha/2}\sqrt{n}}{n+z^2_{\alpha/2}}\sqrt{\widehat{p}(1-\widehat{p}) + \frac{z^2_{\alpha/2}}{4n}}$, where $\widetilde{p} = \frac{x+(z^2_{1-\alpha/2})/2}{n+z^2_{1-\alpha/2}}$
- **Bayesian:** Highest probability density interval for posterior $p|x \sim \text{Beta}(x+0.5, n-x+0.5)$
- **Clopper-Pearson (Exact):** End points are p's that are the solutions to $\sum_{i=0}^{x}\binom{n}{i}p^i(1-p)^{n-i} = \alpha/2$ and $\sum_{i=x}^{n}\binom{n}{i}p^i(1-p)^{n-i} = \alpha/2$

Results

Table 1: *Simulated level of confidence for various values of n and p. Overall performance for each value of n is given across all values of p.*

		AC	*A*	*B*	*E*	*W*
n = 20	p = 0.15	0.9777	0.8190	0.9393	0.9777	0.9777
	p = 0.50	0.9586	0.9586	0.9586	0.9586	0.9586
	p = 0.75	0.9555	0.8971	0.9631	0.9631	0.9351
	Overall	0.9639	0.8916	0.9537	0.9665	0.9571
n = 50	p = 0.15	0.9563	0.9408	0.9408	0.9729	0.9563
	p = 0.50	0.9368	0.9368	0.9368	0.9679	0.9368
	p = 0.75	0.9519	0.9406	0.9256	0.9670	0.9519
	Overall	0.9483	0.9394	0.9344	0.9693	0.9483
n = 100	p = 0.15	0.9666	0.9333	0.9513	0.9666	0.9350
	p = 0.50	0.9436	0.9436	0.9436	0.9647	0.9436
	p = 0.75	0.9526	0.9471	0.9364	0.9633	0.9526
	Overall	0.9543	0.9413	0.9438	0.9649	0.9437

Figure 1:

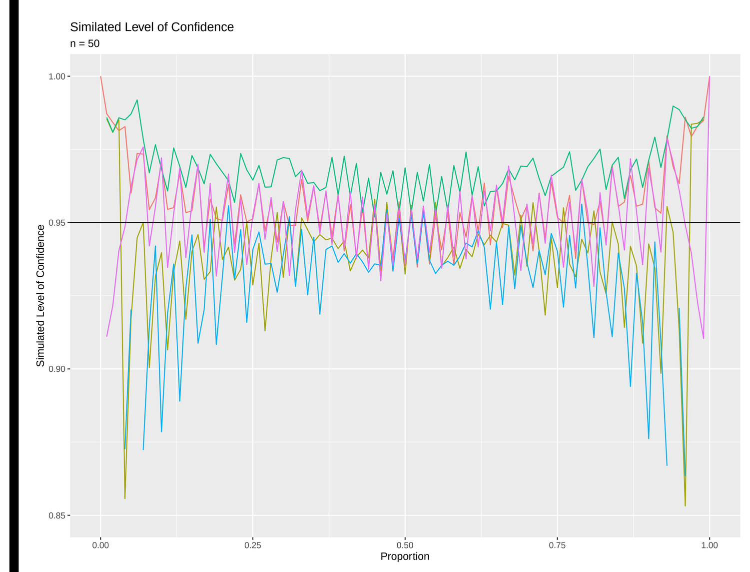
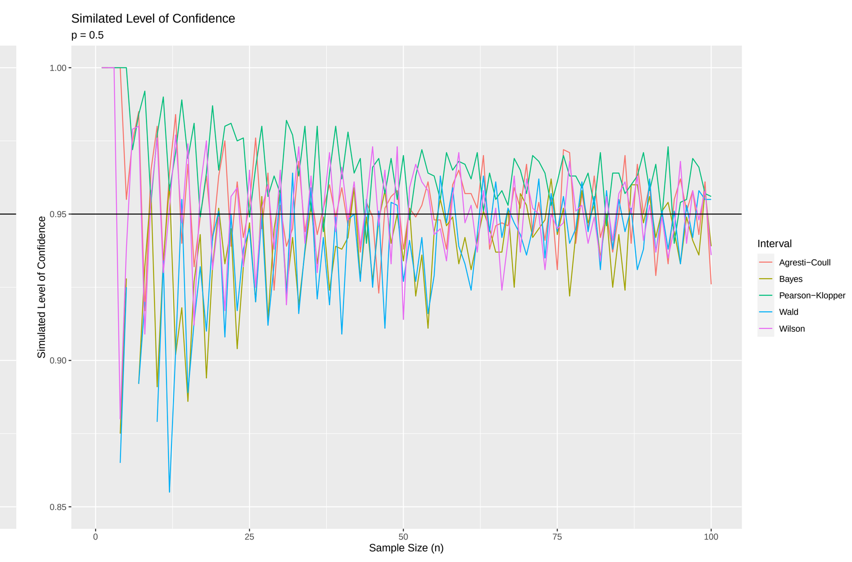


Figure 2:



Discussion

Observing the performance of these five confidence intervals compared to one another, we can see that the results of these simulations agree with established literature. The confidence level of the ***Wald (Asymptotic)*** CI results in a level that is far lower than what is stated. Results fall well below 95%, especially when n is small or p is close to zero or one. We say Wald CI's are too "liberal", or actually much lower in confidence, than what is stated.

The ***Exact (Clopper-Pearson)*** CI appears to contain intervals with the highest level of confidence compared to the others. We say this interval is too "conservative", in that it provides results that are actually *more* confident than what is stated. The trade-off for an increased level of confidence is a wider, less precise estimate of the population proportion.

The ***Agresti-Coull***, ***Wilson***, and ***Bayes*** confidence intervals all perform fairly well, in that the simulated level of confidence is much closer to 95%, the *actual* stated level of confidence.

In general, intervals tend to perform better for larger values of n and values of p closer to 0.50.

References

[1] Brown, Lawrence D., et. al. "Interval Estimation for a Binomial Proportion". *Statistical Science*. (2001). Vol. 16, No. 2, 101-133.

[2] Coastal Carolina University. "2021 Fact Book". (2021). *www.coastal.edu/iraa/factbooks*.

[3] R Core Team. "R: A Language and Environment for Statistical Computing". R Foundation for Statistical Computing. Vienna, Austria. (2020). *https://www.R-project.org/*.