# CONFIDENCE INTERVALS FOR PROPORTIONS: A COMPARISON BETWEEN TRADITIONAL AND AGRESTI-COULL METHODS

**Authors**: Xavier Carlos, Efren Torres, Kaleb Flores

**Abstract**

Confidence intervals (CIs) are a foundational concept in inferential statistics used to estimate population parameters. This study compares two widely used methods for constructing confidence intervals for proportions: the Traditional Method and the Agresti-Coull Method. Using simulation with a synthetic population, we compute the empirical coverage probability of each method and assess their effectiveness. Our results show that the Agresti-Coull method slightly outperforms the traditional approach in terms of coverage accuracy, particularly in small to moderate sample sizes.

## Introduction

Confidence intervals are a statistical tool used to provide a range of plausible values for a population parameter based on sample data. When estimating proportions, the traditional method uses the normal approximation of the binomial distribution. However, this can be inaccurate for small sample sizes or when proportions are near 0 or 1. The Agresti-Coull method introduces a correction to the traditional formula, leading to better empirical performance under many conditions.

This paper evaluates the empirical performance of both methods using simulations on synthetic data.

## Methods

### Generating a Synthetic Population

We created a synthetic population of size N=5000N = 5000N=5000, where each member has a 67% chance of being a success (coded as 1).

```
create_population <-
function(N){
  prop1 <- 0.67
  population <- rbinom(N, size
= 1, prob = prop1)
  return(population)
}
```

### Sampling Function

From the population, we draw random samples of size n=200n = 200n=200 with replacement to mimic real-world random sampling.

```
f_sample <-
function(population, n) {
  sample_data <-
sample(population, size = n,
replace = TRUE)
  return(sample_data)
}
```

### Confidence Interval Methods

### Traditional Method

The traditional confidence interval for a proportion uses the normal approximation and the sample proportion.

$$p^{\sim} \pm Z\frac{\alpha}{z}\frac{\sqrt{P^{\sim}(1-p^{\sim})}}{n}$$

```
traditional <- function(x, n) {
  p_hat <- x / n
  z <- qnorm(p = .975)  # z ≈ 1.96 for 95% CI
  sd <- sqrt((p_hat * (1 - p_hat)) / n)
  lb <- p_hat - z * sd
  ub <- p_hat + z * sd
  return(c(lb, ub))
}
```

### Agresti-Coull Method

This method adds a small-sample correction: two successes and two failures are added to the observed count, and the sample size is adjusted to n+4n + 4n+4. Here is the formula that we try to replicate in the code.

$$p^{\sim} \pm Z\sqrt{\frac{p^{\sim}(1-p^{\sim})}{n}}$$

```
agresti <- function(x, n) {
  n_hat <- n + 4
  p_hat <- (x + 2) / n_hat
  z <- qnorm(p = .975)
  sd <- sqrt((p_hat * (1 - p_hat)) / n_hat)
  lb <- p_hat - z * sd
  ub <- p_hat + z * sd
  return(c(lb, ub))
}
```

### Simulation Procedure

To evaluate the performance of both methods, we simulate 10,000 random samples from the population. We calculate the CI for each sample and track how often it contains the true population proportion (0.67). If you notice that we use method to state the function that computes a confidence interval.

```
simulation <- function(population, method, n = 200, sim = 10000) {
  true_prop <- mean(population)
  coverage <- 0
  for (i in 1:sim) {
    sample_data <- f_sample(population, n)
    x <- sum(sample_data)
    ci <- method(x, n)
    if (true_prop >= ci[1] && true_prop <= ci[2]) {
      coverage <- coverage + 1
    }
  }
  return(coverage / sim)
}
```

## Results

We generated a population and ran the simulations as follows:

```
population <-
create_population(N = 5000)

coverage_trad <-
simulation(population,
traditional)

coverage_agre <-
simulation(population,
agresti)
```

### Coverage Results

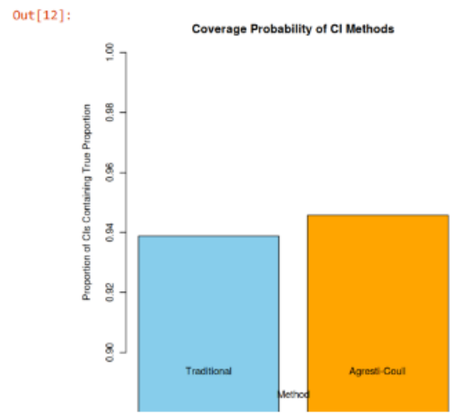
- **Traditional Method**: 0.9388
- **Agresti-Coull Method**: 0.9457

### Visual Comparison

We visualized the coverage probabilities using a bar plot:

```
coverage_rates <-
c(coverage_trad,
coverage_agre)

names(coverage_rates) <-
c("Traditional", "Agresti-
Coull")

barplot(coverage_rates,

        col = c("skyblue",
"orange"),

        ylim = c(0.9, 1),

        main = "Coverage
Probability of CI Methods",

        ylab = "Proportion of
CIs Containing True
Proportion",

        xlab = "Method")
```

*Figure 1: Coverage probabilities for Traditional and Agresti-Coull methods.*




## Discussion

The Agresti-Coull method yielded slightly higher coverage than the Traditional method, aligning more closely with the nominal 95% confidence level. This supports existing literature indicating that the Agresti-Coull method is more robust, particularly in small to moderate sample sizes or when dealing with extreme proportions.

While both methods performed reasonably well, even slight improvements in coverage can have significant implications in statistical decision-making—particularly in fields where precision is critical.

## Conclusion

This simulation study demonstrates that the Agresti-Coull method is a reliable alternative to the traditional normal approximation method for constructing confidence intervals for proportions. Its slight improvement in coverage probability makes it a valuable tool, especially in situations involving limited data or proportions near 0 or 1.

## Source


*Learning statistics with R: A tutorial for psychology students and other beginners. (Version 0.6.1)*

*Danielle Navarro (bookdown translation: Emily Kothe)*