MATH 445 - Final Project Research Plan

Project Title:
Finding Reliable Confidence Interval for the Odds Ratio through Simulation and Visualization

Group Members: Jiyoung Cheong, Phuong Dinh, Reagan Billingsley

Simulation Study Plan
We will conduct a comprehensive simulation study to evaluate the performance of confidence interval (CI) methods for the odds ratio (OR). The simulation will involve varying four key parameters:p1, OR(θ), n1d, and n2d. These parameters will be combined systematically, resulting in 10 * 11 * 4 * 4 = 1760 unique scenarios.

p1 = [0.05, 0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, 0.85, 0.95]
OR(θ) = [1, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100]
n1d = [5, 10, 25, 50]
n2d = [5, 10, 25, 50]

Where:
- p1: the probability of success for group 1.
- OR(θ): true odds ratio, and is defined by the formula $\theta = \frac{p1/(1-p1)}{p2/(1-p2)}$.
- n1d, n2d: sample sizes for groups 1 and 2 respectively.

Rationale

Vary p1
The parameter p1 represents the probability of success for group 1. By selecting values between 0.05 and 0.95, we ensure that our simulations cover both extremes (low and high probabilities) as well as intermediate cases. Since the odds ratio depends on the relationship between p1and p2 , varying p1 allows us to assess CI performance across diverse probability distributions.

Expanding the Range of OR (θ)
For OR(θ), in Fagerland et al. (2011), the range of the true odds ratio (θ) was restricted to [0.1,10], which encompasses moderate associations between groups. However, to thoroughly investigate the performance of confidence interval (CI) methods under extreme conditions, we have extended the range of θ to [1,100]. This expanded range allows us to examine the behavior of CIs in scenarios involving both highly unequal probabilities (p1 and p2) and equal probabilities. By including these extreme values, we aim to identify edge cases where CI methods may exhibit notable variations in their reliability and coverage properties. The values are chosen including 1, from 10 to 100 in increments of 10.

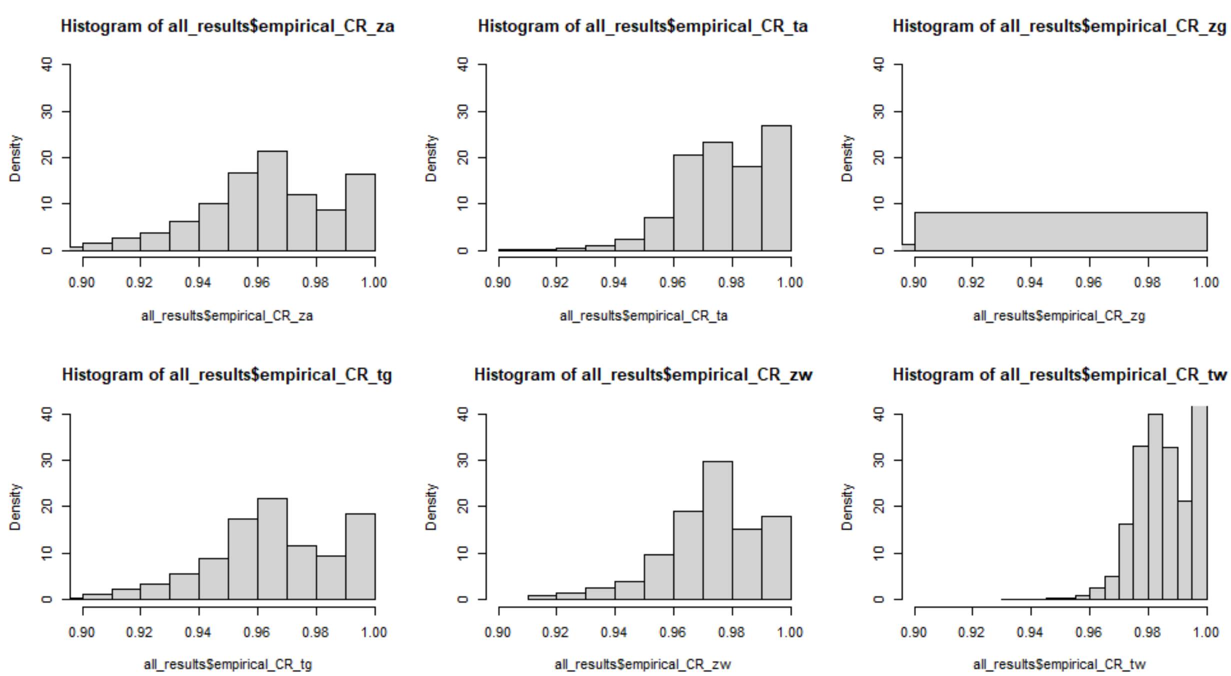
Sample Size Variations

The sample sizes of n1d and n2d are chosen to reflect a wide range of practical scenarios, including equal and unequal group sizes, Specifically, we consider n1d, n2d $\in$ [5, 10, 25, 50], allowing us to examine cases with small and equal sample sizes (e.g., n1d =n2d=5 or n1d =n2d=10), large and equal sample sizes (e.g., n1d=n2d=50), and highly unequal sample sizes (e.g., n1d=5,n2d=50). In the literature, population sizes commonly range between 10 and 20. By including sample sizes of 10 and 25 in our simulation, we enable a direct comparison of our results with those reported in prior studies. Furthermore, the selected sample sizes provide an opportunity to systematically investigate the impact of parity configurations (e.g., odd-odd, odd-even, even-odd, and even-even) on confidence interval performance, potentially revealing patterns or systematic effects related to sample size composition.

Example simulation result

The histograms below display the empirical coverage rates obtained from 1000 Monte Carlo simulations using our chosen parameters. A substantial number of simulations yield coverage rates close to 0.95, suggesting that our parameter choices may provide reliable confidence intervals. However, further analysis is needed to confirm consistency across different conditions and to optimize parameter selection if necessary

```
simulations <- expand.grid(n1d=c(5,10,25,50), n2d=c(5,10,25,50),
                           p1=seq(0.05,0.95,0.1),
                           OR=c(1, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100),
                           alpha=0.05)
```



Reference

National Library of Medicine - Explaining Odds Ratios
https://pmc.ncbi.nlm.nih.gov/articles/PMC2938757/