Estimating Vaccine Efficacy: A Comparative Analysis of Method of Moments and Maximum Likelihood Estimation

Authors: Katherine Fulton, Kreslyn Hinds, Tracia Pan

1 Abstract

The abstract provides a brief summary of the entire paper (background, methods, results and conclusions). The suggested length is no more than 150 words. This allows you approximately 1 sentence (and likely no more than two sentences) summarizing each of the following sections. Typically, abstracts are the last thing you write.

adding words: The abstract provides a brief summary of the entire paper (background, methods, results and conclusions). The suggested length is no more than 150 words. This allows you approximately 1 sentence (and likely no more than two sentences) summarizing each of the following sections. Typically, abstracts are the last thing you write.The abstract provides a brief summary of the entire paper (background, methods, results and conclusions). The suggested length is no more than 150 words. This allows you approximately 1 sentence (and likely no more than two sentences) summarizing each of the following sections. Typically, abstracts are the last thing you write.

Assessment: Are the main points of the paper clearly and succinctly described?

2 Keywords

- Maximum Likelihood Estimation: Using the likelihood function of a parameter, the maximum likelihood estimation approach for estimating a parameter finds the maximum of the function after a log transformation by using it's first and second derivative. After finding the maximum of the function we can solve for the estimated parameter [6.1].

- Method of Moments Estimation: With the assumption that the expected value of a random variable will be a function of a parameter and the sample average will be close to the expected value, methods of moments estimation sets the sample average equal to the expected value and solves for the estimated parameter [6.2].
- Food and Drug Administration (FDA) Emergency Use Authorization (EUA): A mechanism of the FDA to enable the usage of medical countermeasures during public health emergencies. This may include allowing use of unapproved medical products or allowing the usage of already approved medical products for an unapproved purpose [5].
- Coronavirus (COVID 19): A contagious infectious disease caused by SARS-CoV-2, first originating in Wuhan, China December 2019 [1].
- Pfizer Vaccine (BNT162b2):A two-dose vaccine for SARS-CoV-2, safe for all individuals 6 months and above [3].

3 Introduction

With a total of 775 million cases and 7 million deaths reported from January 2020-May 2024 [4], Coronavirus has affected people around the globe and continues to be an ongoing concern. After being declared a pandemic by the World Health Organization on March 11, 2020 [1] emergency use authorizations were issued to combat the virus. In December 2020, Pfizer and BioNTech successfully obtained a US FDA Emergency Use Authorization (EUA) to begin distributing their vaccine (BNT162b2) for COVID 19 [6.3].

Using data as reported by Polack et al. [2] specifically the relevant data from Table 2 of the paper referenced above (summarized below):

Table 1: Vaccine Efficacy against Covid 19 at least 7 days after second dose in patients without evidence of infection

Group	Covid-19 Cases	No. of subjects
BNT162b2	8	17,411
Placebo	162	17,511
Total	170	34,922

We will use maximum likelihood estimation and method of moments estimation to report the estimated efficacy rate of the Pfizer vaccine. Using these estimates we will preform hypothesis testing to determine if the efficacy rate meets the FDA EUA requirements of at least 30% [6.4].

```
##      Group Covid19_Cases No_of_Subjects Non_Covid19_Cases
## 1 BNT162b2             8          17411             17403
## 2  Placebo           162          17511             17349
```

4 Statistical Methods

Using a single binomial variable model, let T denote the number of subjects in the BNT162b2 group out of all the $n = 170$ Covid-19 cases.

$$
\begin{aligned}
T &\sim (n = 170, \pi) \\
\text{Where } \pi = P(BNT162b2|Covid-19case) &\text{ and can be shown to be equal to} \\
\pi &= \frac{n_1\pi_1}{n_1\pi_1 + n_2\pi_2} \\
\text{Since } n_1 = 17411 \text{ and } n_2 = 17511, &\text{ we can say that} \\
n_1 \approx n_2, \text{ the randomization is } 1:1, &\text{ and can simplify } \pi \text{ to} \\
\pi &= \frac{\pi_1}{\pi_1 + \pi_2} \\
\text{We then have our parameter of interest} \\
\psi &= \frac{\pi_2 - \pi_1}{\pi_2} \\
&= \frac{1 - 2\pi}{1 - \pi} \\
&= 1 - \frac{\pi_1}{\pi_2}
\end{aligned}
\tag{1}
$$

To make inferences, we decided to use likelihood inference and method of moments estimation.

$$ \tag{2} $$

5 Likelihood inference

In this section, the $\hat{\psi_0^{mle}}$ was calculated, and large sample confidence interval and a likelihood ratio test were performed to calculate the P-value that tests $H_0 : \psi_0 = 0.3$ versus $H_a : \psi_0 > 0.3$.
parameter of interest: $\psi = \frac{1-2\pi}{1-\pi} \iff \pi = \frac{1-\psi}{2-\psi}$

$$
\begin{aligned}
L(\psi) &= (\binom{170}{t})(\frac{1-\psi}{2-\psi})^t(1-\frac{1-\psi}{2-\psi})^{170-t}, x = 0, 1, \ldots, 170 \\
\ell(\psi) &= ln(\binom{170}{t}) + tln(\frac{1-\psi}{2-\psi}) + (170-t)ln(1-\frac{1-\psi}{2-\psi}) \\
\ell'(\psi) &= -\frac{t}{(-\psi+1)(-\psi+2)} + \frac{170-t}{\psi+2}
\end{aligned}
\tag{3}
$$

$$
\begin{aligned}
&\text{To find } \hat{\psi_0^{mle}}\text{, we set } \ell'(\psi) = 0 \\
-\frac{t}{(-\psi+1)(-\psi+2)} + \frac{170-t}{\psi+2} &= 0 \\
\frac{t}{(-\psi+1)(-\psi+2)} &= \frac{170-t}{\psi+2} \\
t(-\psi+2) &= (170-t)(-\psi+1)(-\psi+2) \\
t &= (170-t)(-\psi+1) \\
t &= -170\psi + \psi t - t + 170 \\
170\psi - \psi t &= -2t + 170 \\
\psi(170-t) &= -2t + 170 \\
\hat{\psi_0^{mle}} &= \frac{-2t+170}{170-t} \\
t &= 8 \\
\hat{\psi_0^{mle}} &= \frac{-2(8)+170}{170-(8)} = 0.95
\end{aligned}
\tag{4}
$$

To find the 95% confidence interval, we need to find $SE = \sqrt{\frac{-1}{\ell''(\psi)}}$

$$\ell'(\psi) = -\frac{t}{(-\psi+1)(-\psi+2)} + \frac{170-t}{\psi+2}$$
$$\ell''(\psi) = \frac{t(2\psi-3)}{(-\psi+1)^2(-\psi+2)^2} + \frac{170-t}{(-\psi+2)^2}$$

Plugging in $t = 8$ and $\psi = 0.95$

$$= -3045.805 \tag{5}$$
$$-\ell''(\psi) = 3045.805$$
$$SE = \sqrt{\frac{-1}{\ell''(\psi)}} = \sqrt{\frac{1}{-\ell''(\psi)}}$$
$$= \sqrt{\frac{1}{3045.805}}$$

Assuming that n is large and the regularity conditions are met, we have

$\psi_0^{mle} \approx Norm(\psi_0, SE(\psi_0^{mle}) = \sqrt{\frac{1}{3045.805}})$

The 95% Confidence interval is

$\psi_0^{mle} \pm 1.96 \times SE$

$0.95 \pm 1.96 \times \sqrt{\frac{1}{3045.805}}$

Using the Likelihood ratio test to find the p-value, we have the hypothesis

$H_0 : \psi_0 = 0.3$ $H_a : \psi_0 > 0.3$.

$\psi_0^{\hat{mle}} = 0.95$

$W = 2ln(\frac{L(\psi_0^{mle})}{L(\psi_0^{mom})}) \approx 121.6$

When H_0 is true, $W \approx \chi_1^2$

P-value $= P(\chi_1^2 \geq 121.6)$ = pchisq(121.6, df = 1, lower.tail=F)

6 Method of moments estimation

In this section, we found the $\psi_0^{\hat{mom}}$, confidence intervals for ψ_0 using bootstrap, and the empirical P-values to calculate the P-value that tests $H_0 : \psi_0 = 0.3$ versus $H_a : \psi_0 > 0.3$.

We observed $t = 8$ and the distribution is $T \sim (n = 170, \pi)$

parameter of interest: $\psi = \frac{1-2\pi}{1-\pi} \Longleftrightarrow \pi = \frac{1-\psi}{2-\psi}$

Assuming the estimator is unbiased, M.O.M estimator satisfies

$$\begin{aligned}
E[T] &= t \\
n\pi &= t \\
170\pi &= t \\
170 \times \frac{1-\psi}{2-\psi} &= t \\
170 \times \frac{1-\psi}{2-\psi} &= 8 \\
\frac{1-\psi}{2-\psi} = \frac{8}{170} \\
8(2-\psi) &= 170(1-\psi) \\
16 - 8\psi &= 170 - 170\psi \\
162\psi &= 154 \\
\hat{\psi_0} &= \frac{154}{162} \approx 0.95 \\
\psi_0^{\hat{mom}} &= 0.95
\end{aligned} \tag{6}$$

To get the confidence intervals, the parametric bootstrap was used because we repeatedly draw new samples from a binomial distribution.

We then used the Simple Percentile Method to find the confidence intervals because the $\psi_0^{\hat{mom}}$ we calculated which is 0.95 indicates that the sampling distribution would be skewed.

To find the P-value that tests $H_0 : \psi_0 = 0.3$ $H_a : \psi_0 > 0.3$, we used empirical P-values where we calculated the p value by determining the proportion of time the test statistic value was at least as large as our observed value $t = 8$.

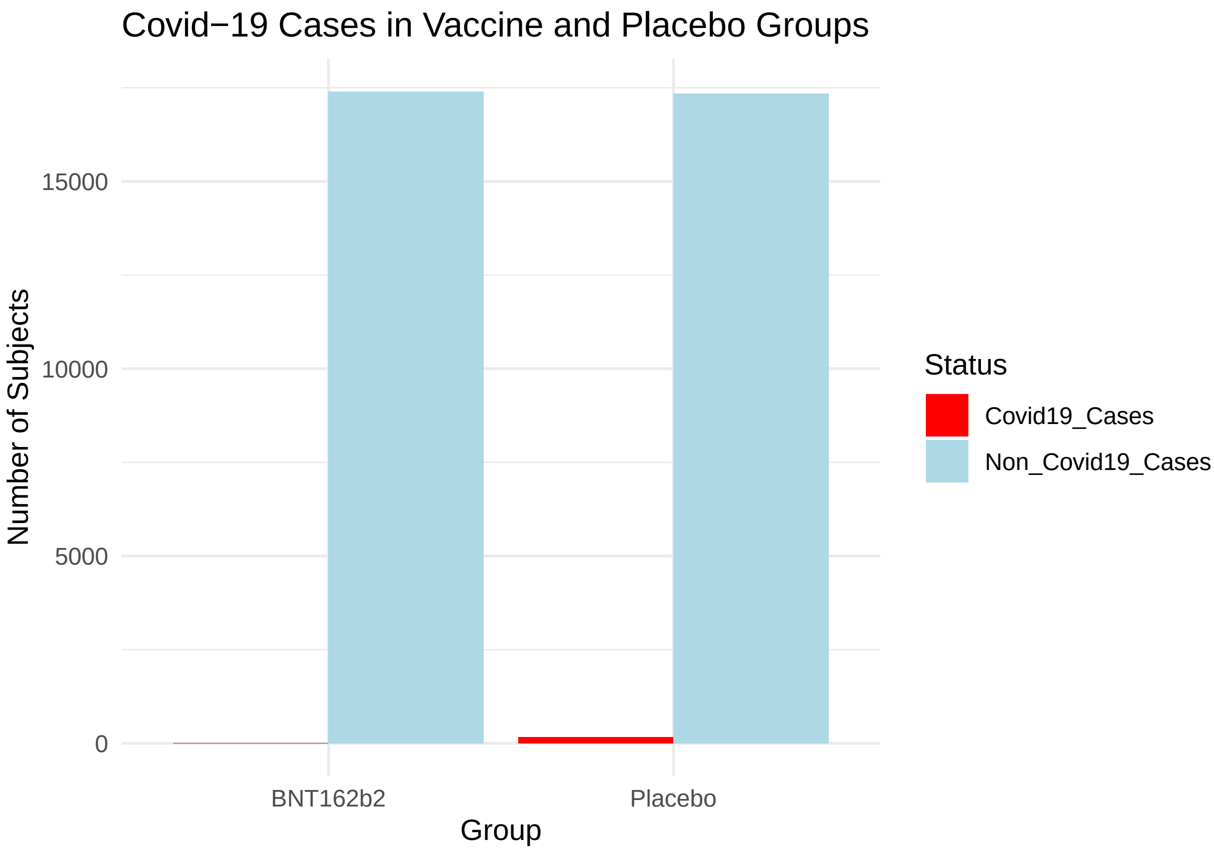
7 Results:

To begin, we examined the descriptive statistics of our sample to provide a foundational understanding of the data. Figure 1 presents a segmented bar plot illustrating the distribution of key categorical variables.

Log Likelihood

Method of Moments

Descriptive Statistics:



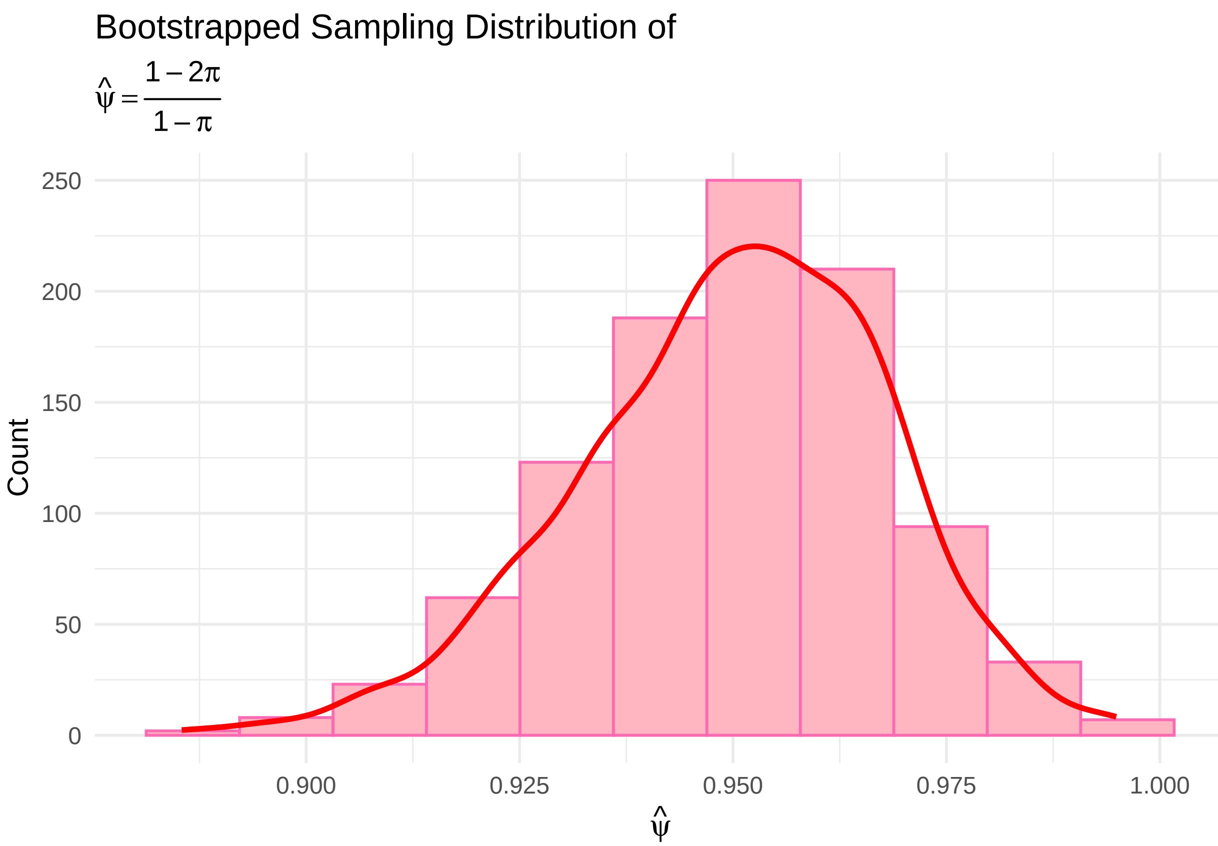
Assessment

- *Is the content appropriate for a results section? Is there a clear description of the results?*
- *Are the results/data analyzed well? Given the data in each figure/table is the interpretation accurate and logical? Is the analysis of the data thorough (anything ignored?)*
- *Are the figures/tables appropriate for the data being discussed? Are the figure legends and*

titles clear and concise?

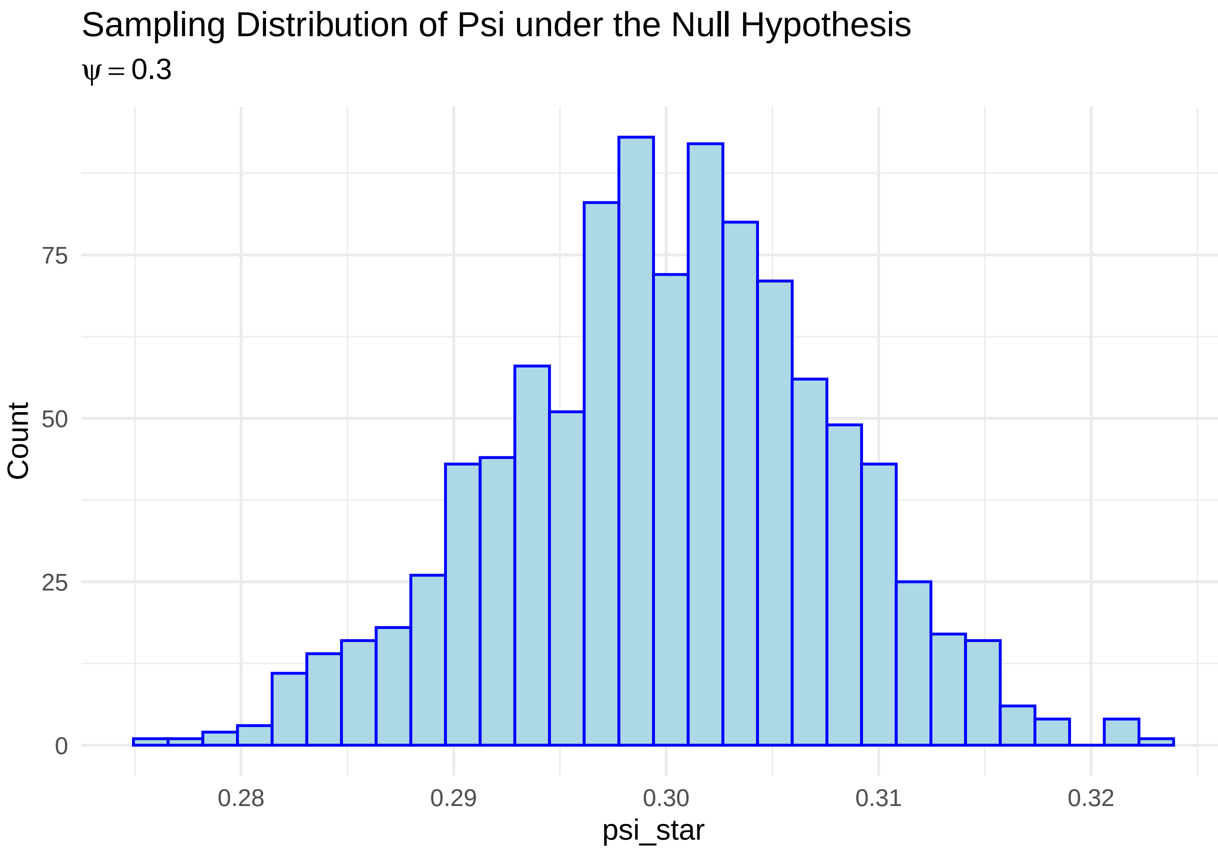
Bootstrap Histogram

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



8 Empirical P Value Histogram

```
## [1] 0
```



9 Discussion/Conclusion

discuss your findings in practical terms, compare and contrast the results with each other and with the Pfizer analysis, identify strengths and weaknesses of the methods used.

Assessment:

- *Do the authors clearly state whether the results answer the question (support or disprove the hypothesis)?*
- *Were specific data cited from the results to support each interpretation? Do the authors clearly articulate the basis for supporting or rejecting each hypothesis?*

Reference the papers results?

In the paper, the authors used a Bayesian beta binomial model with the transformation

$$\pi = \frac{\pi_v}{\pi_v + \pi_p}$$

where π_v and π_p are the probabilities of an infection in the vaccine and placebo groups respectively. They estimated the vaccine efficacy $\psi = \frac{1-2\pi}{1-\pi}$ to be 95%. The 95% Bayesian credible interval for ψ was $[90.3\%, 97.6\%]$, and they reported that the Bayesian estimate for the probability that vaccine efficacy exceeded 30% was larger than 0.9999.

10 References

List at least three references for your paper. These can be other articles on the subject/data, or mathematical references for the statistical techniques that you used. Each item you list here should be cited/referenced at least once in the actual text. For example, here is a blog by Stephen Senn that really helped me understand the problem. He's a great writer and you could read his other blogs on the vaccine trials.

1. COVID-19 Timeline
2. Polack et al.
3. About the vaccine
4. dashboard for covid cases
5. About EUA
6. Stat 342 Materials:

- 6.1 - Slide deck 24, slides 10,34
- 6.2 - Slide deck 17, slide 5
- 6.3 - Assignment instructions
- 6.4 - Slide deck physician-health, slide 7

Assessment

- *Are the references appropriate and of adequate quality?*
- *Are the references cited properly (both in the text and at the end of the paper)?*

11 Appendix

Include important R code, R output and additional plots/figures. Any plots you include in your report should be cited in the body of the report. There should be a reason for the plot – something you want the reader to see. Note: tables and figures which are pertinent to your analysis should be included in the body of your report, not in the appendix. The reader should not have to look at your appendix for main results/figures.

```
For example, you can write your code in chunks interspersed within the document, with chunk op
```