Studying Efficacy of COVID-19 Vaccine: Brief Analysis Using Statistical Methods

Cheeson Lau, Edison Lu, and Noah McMahon

June 3, 2024

1 Abstract

In this paper, we will be analyzing the efficacy of the BNT162b2 COVID-19 vaccine (referred to as ψ_0), using the following two methods: maximum likelihood estimator and method of moments estimator.

From calculating our maximum likelihood estimator, we arrived at a 95% confidence interval of [0.9156, 0.9857], which is congruous with the 95% confidence interval from our method of moments estimator ([0.9411, 0.9601]) as well as the 95% confidence interval from the paper itself ([0.903, 0.976]). We also calculated the P-Values from both of these estimators and got values of approximately zero, meaning that in the hypothesis test of H_0: $\psi_0 = 0.3$ vs. H_1: $\psi_0 > 0.3$ to see if the FDA would approve the vaccine, we rejected the null hypothesis as both P-Values are below the presumed α level, meaning that the BNT162b2 vaccine would be approved by the FDA for general use.

2 Keywords

COVID-19, Vaccine, Maximum Likelihood Estimator, Method of Moments Estimator, Confidence Interval, P-Value

3 Introduction

The global outbreak of the COVID-19 pandemic in 2020 caused catastrophic damage on public health and the economy. Hundreds of millions of people were severed impacted, losing friends, families and jobs. A solution was and is still urgently needed. Treatments such as acetaminophen and ibuprofen could soften the symptoms (COVID-19 Treatment and Preventive Medication), which include fever, fatigue, cough, loss of taste or smell, etc (Symptoms of COVID-19). Masks, quarantine and lockdowns are deterrents to the pandemic, and could slow down infection speed. However, all these solutions are only temporary and not sustainable in the long run. This is why an effective, well-tested COVID-19 vaccine is urgently needed.

In December 2020, Pfizer and BioNTech successfully obtained a US FDA Emergency Use Authorization (EUA) to begin distributing their two-dose vaccine (BNT162b2) for SARS-CoV-2 ("COVID-19") (FDA Approves First COVID-19 Vaccine). In order to obtain the authorization to release the vaccine to the populace, Pfizer and BioNTech need well-designed research to advocate their vaccine. The BNT162b2 vaccine is special in the sense that it uses the mRNA approach, contrary to the traditional style vaccines developed by AstraZeneca, and Johnson & Johnson. While traditional vaccines require injecting weakened virus into human body, mRNA vaccines teach

human cells to create a tiny piece of virus to develop immunity (MRNA VACCINES FOR COVID-19). The discovery of mRNA dated back to early 1960s. It took decades for scientists to achieve breakthroughs. In 2017, by studying a Zika virus DNA-based vaccine, scientists from National Institutes of Health discovered that mRNA vaccines are safe (Decades in the Making: mRNA COVID-19 Vaccines). It is safe to assume that mRNA technology had been somewhat tested before COVID-19 hit humanity in 2020, so it made sense that Pfizer and BioNTech aimed to develop a COVID-19 vaccine using such technology.

This study not only contributes to tackling COVID-19, but also to the broader body of research on mRNA vaccines and preparing for future pandemics. Given that these findings are crucial to humanity in many facets, the statistical methods chosen to analyze and verify vaccine efficacy have to be robust. The goal of this paper is to compare probability of getting infected by COVID-19 between the vaccine and placebo groups, acquiring supportive evidence on vaccine efficacy.

4 Statistical Methods

The main objective of the study is to apply two main statistical methods, Maximum Likelihood Inference as well as Method of Moments Estimation to analyze the efficacy of COVID-19 Vaccine by studying the sample data obtained from the placebo-controlled, observer-blinded, randomization-sampled experiment. Let ψ be our parameter of interest, the BNT162b2 Vaccine efficacy. The efficacy measures how much COVID-19 infection risk

Group	COVID-19 Cases	No. of Subjects
BNT162b2	8	17411
Placebo	162	17511
Total	170	34922

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

the vaccine reduces compared to the placebo.

From Table 1: Vaccine Efficacy against COVID-19 at least 7 days after second dose in patients without evidence of infection, we learned that there were 8 COVID-19 Cases from a total of 17,411 subjects who received BNT162b2 Vaccine, and 162 COVID-19 Cases from a total of 17,511 subjects who received the placebo.

Let n be the total number of COVID-19 cases. We can find n easily, $n = 162+8 = 170$. Suppose we denote the random variable T as the number in the COVID-19 group. The distribution of T can then be approximated by a binomial distribution, assuming that each participant is independent of the other participants. Then,

$$T \sim Binom(n = 170, \pi)$$

where π = P(Vaccine group | COVID-19 infected) and

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

where π_v and π_p are the probabilities of having a COVID-19 infection in the vaccine and placebo group respectively. n_v and n_p are then the number of participants in the vaccine and placebo group respectively.

We know that $n_v = 17411$ and $n_p = 17511$. Since $n_v \approx n_p$, we can simplify π further to get

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

From our definition of our parameter of interest ψ, we can express it as

$$\psi = 1 - \frac{\pi_v}{\pi_p} = \frac{\pi_p - \pi_v}{\pi_p}$$

Now we can write ψ in terms of the binomial probability π:

$$\begin{aligned}
\psi &= \frac{\pi_p - \pi_v}{\pi_p} \\
&= \frac{\pi_v + \pi_p - 2\pi_v}{\pi_v + \pi_p - \pi_v} \\
&= \frac{1 - 2(\frac{\pi_v}{\pi_v + \pi_p})}{1 - \frac{\pi_v}{\pi_v + \pi_p}} \\
&= \frac{1 - 2\pi}{1 - \pi}
\end{aligned}$$

We can also write π in terms of ψ:

$$\begin{aligned}
\psi &= \frac{1 - 2\pi}{1 - \pi} \\
\psi - \psi\pi &= 1 - 2\pi \\
2\pi - \psi\pi &= 1 - \psi \\
\pi &= \frac{1 - \psi}{2 - \psi}
\end{aligned}$$

We will then analyze these data using two above-mentioned methods, the Maximum Likelihood Inference as well as the Method of Moments Estimation.

4.1 Maximum Likelihood Inference

Given above that

$$T \sim Binom(170, \pi)$$

where we observed $t = 8$ as t is the observed value of patients who were infected by the COVID-19 virus after receiving 2 doses of the vaccine.

The PMF of binomial distribution is given by

$$\binom{n}{x}\pi^x(1-\pi)^{n-x}$$

where $x = 0, 1, 2, \ldots$

Thus, we know that the likelihood function of π is given by

$$L(\pi) = \binom{n}{t}\pi^t(1-\pi)^{n-t}$$

where t = observed value of number of COVID-19 cases from the vaccine group.

Because the vaccine efficacy is given by $\psi = \frac{1-2\pi}{1-\pi}$, we can rewrite the expression in terms of ψ, which is

$$g(\pi) = \frac{1-\psi}{2-\psi}$$

Therefore, we can write the likelihood function $L(\psi)$.

$$\begin{aligned} L(\psi) &= L(g(\pi)) \\ &= \binom{n}{t}(\frac{1-\psi}{2-\psi})^t(1-\frac{1-\psi}{2-\psi})^{n-t} \\ &= \binom{n}{t}(\frac{1-\psi}{2-\psi})^t(\frac{1}{2-\psi})^{n-t} \\ &= \binom{n}{t}\frac{(1-\psi)^t}{(2-\psi)^n} \end{aligned}$$

We can find the MLE for ψ by applying the technique of differentiation.

So, the log-likelihood function is:

$$\ell(\psi) = \ln(L(\psi)) \propto t \times \ln(1-\psi) - n \times \ln(2-\psi)$$

Therefore candidates for the MLE of ψ satisfy the equation:

$$\tfrac{d}{d\psi}\ell(\psi) = \tfrac{-t}{1-\psi} + \tfrac{n}{2-\psi} = 0$$

Hence we can solve the above equation:

$$\begin{aligned}
\frac{d}{d\psi}\ell(\psi) = \frac{-t}{1-\psi} + \frac{n}{2-\psi} &= 0 \\
-t(2-\psi) + n(1-\psi) &= 0 \\
-2t + t\psi + n - n\psi &= 0 \\
(t-n)\psi &= 2t - n \\
\hat{\psi_0}^{MLE} &= \frac{2t-n}{t-n} \\
&= \frac{-(n-2t)}{-(n-t)} \\
&= \frac{n-2t}{n-t} \\
&= \frac{1-2\frac{t}{n}}{1-\frac{t}{n}} \\
&= \frac{1-2\hat{\pi_0}^{MLE}}{1-\hat{\pi_0}^{MLE}}
\end{aligned}$$

Thus, we get: $\hat{\psi_0}^{MLE} = \frac{1-2\hat{\pi_0}^{MLE}}{1-\hat{\pi_0}^{MLE}}$.

Now we can calculate the 95% confidence interval of our parameter of interest ψ as well as the Likelihood Ratio Test.

By Theorem 25.1, the Wald Confidence interval of ψ is given by

$$\hat{\psi_0}^{MLE} \pm z_{\alpha/2}\sqrt{\frac{-1}{\frac{-t}{(1-\psi)^2} + \frac{n}{(2-\psi)^2}}}$$

since we are using large values of n and regularity conditions are ensured.

We then conduct a likelihood ratio test of $H_0 : \psi_0 = 0.3$ versus H_1: $\psi_0 \neq 0.3$.

Since the likelihood ratio test statistic is:

$$W = 2ln\left[\frac{L(\hat{\psi_0}^{MLE})}{L(\psi_0^{null})}\right]$$

So under $H_0 : \psi_0 = \psi_0^{null}$ for large samples, $W \approx \chi_1^2$.

Then:

$$\begin{aligned} W &= 2 \times ln\left[\frac{L(\hat{\psi_0}^{MLE})}{L(\psi_0^{null})}\right] \\ &= 2 \times ln\left[\frac{\binom{n}{t}\frac{(1-\hat{\psi_0}^{MLE})^t}{(2-\hat{\psi_0}^{MLE})^n}}{\binom{n}{t}\frac{(1-\psi_0^{null})^t}{(2-\psi_0^{null})^n}}\right] \\ &= 2 \times ln\left[\frac{(1-\hat{\psi_0}^{MLE})^t}{(2-\hat{\psi_0}^{MLE})^n} \times \frac{(2-\psi_0^{null})^n}{(1-\psi_0^{null})^t}\right] \end{aligned}$$

4.2 Method of Moments Estimation

We know that $T \sim Binom(n = 170, \pi)$ and we observed $t = 8$. Now we have to find $\hat{\psi}_0^{MOM}$.

$$\begin{aligned} E[T] &= \bar{t} \\ n\pi &= t \\ n \cdot \frac{1-\psi}{2-\psi} &= t \\ \frac{n - n\psi}{2-\psi} &= t \\ n - n\psi &= 2t - t\psi \\ n - 2t &= n\psi - t\psi \\ \psi(n-t) &= n - 2t \\ \psi &= \frac{n-2t}{n-t} \end{aligned}$$

Then, we can construct confidence intervals for ψ_0 using the parametric bootstrap method. First, we need to generate m samples from $Binom(n, \pi)$ independently. Second, from the newly generated samples, we calculate a new estimate for ψ from it.

$$\hat{\psi_0^\star}^{MOM} = \tfrac{n-2\bar{t^\star}}{n-t^\star}$$

where $\bar{t^\star} = \frac{1}{m}\sum_{i=1}^{m} t_i^\star$ is the sample mean. $\hat{\psi_0^\star}^{MOM}$ is the bootstrapped estimate of ψ_0.

Next, we will repeat the first and second steps B times, so we will obtain B bootstrapped estimates of ψ_0. We will have a bootstrapped sampling distribution of $\hat{\psi_0}^{MOM}$ as a result. We have to find the mean and standard error

of the bootstrapped estimates, so that we can build a bootstrapped confidence interval for ψ_0 by using the standard bootstrap method. In addition, we need to evaluate bias like so:

$$\text{Bias} = \text{Bootstrap Mean - } \hat{\psi_0}^{MOM}$$

Our bias-corrected 95% confidence interval is

$$\text{Bias-corrected estimate} \pm 1.96 \times \text{Bootstrap Standard Error}$$

where the bias-corrected estimate is equal to $\hat{\psi_0}^{MOM}$ - Bias.

It is complied to figure out the sampling distribution, so we can find the empirical p-value by using simulation.

So suppose $T \sim Binom(n = 170, \pi_0 = \frac{1-\psi_0}{2-\psi_0})$. We wish to test $H_0 : \psi_0 = 0.3$ vs. $H_1 : \psi_0 > 0.3$. Set $\alpha = 0.01$. When $\psi_0 = 0.3$, $\pi_0 = \frac{1-0.3}{2-0.3} = 0.4118$. Our observed $t_{obs} = 8$.

We decide to use T itself as the test statistic. Note that $E[T] = n\pi_0 = n \cdot \frac{1-\psi_0}{2-\psi_0}$. Assuming that the vaccine does not make a patient more prone to COVID-19, the range of ψ_0 is [0, 1]. Within this range, π_0 decreases when ψ_0 increases. Therefore, small values of T gives evidence against the null hypothesis and support the alternative hypothesis more.

To calculate our empiricial P-value, we first have to generate B samples t^* from $Binom(n = 170, \pi_0 = 0.4118)$. Let $B = 10000$.

5 Results

5.1 Maximum Likelihood Inference

We then derive the second-order derivative to obtain the observed information:

$$\frac{d^2}{d\psi^2} = \frac{-t}{(1-\psi)^2} + \frac{n}{(2-\psi)^2}.$$

Thus, the estimated standard error of $\hat{\psi_0}^{MLE}$ is $\sqrt{\frac{-1}{\frac{-t}{(1-\psi)^2}+\frac{n}{(2-\psi)^2}}}$.
In order to check if we can apply Theorem 25.1, we first plot the log-likelihood function along with the second-order Taylor approximation around the MLE and evaluate the quadratic approximation.
The quadratic approximation to the log-likelihood suggests that we can rely on the confidence interval of the Wald intervals.

$$\hat{\psi_0}^{MLE} \pm z_{\alpha/2}\sqrt{\frac{1}{\frac{-t}{(1-\psi)^2}+\frac{n}{(2-\psi)^2}}}$$

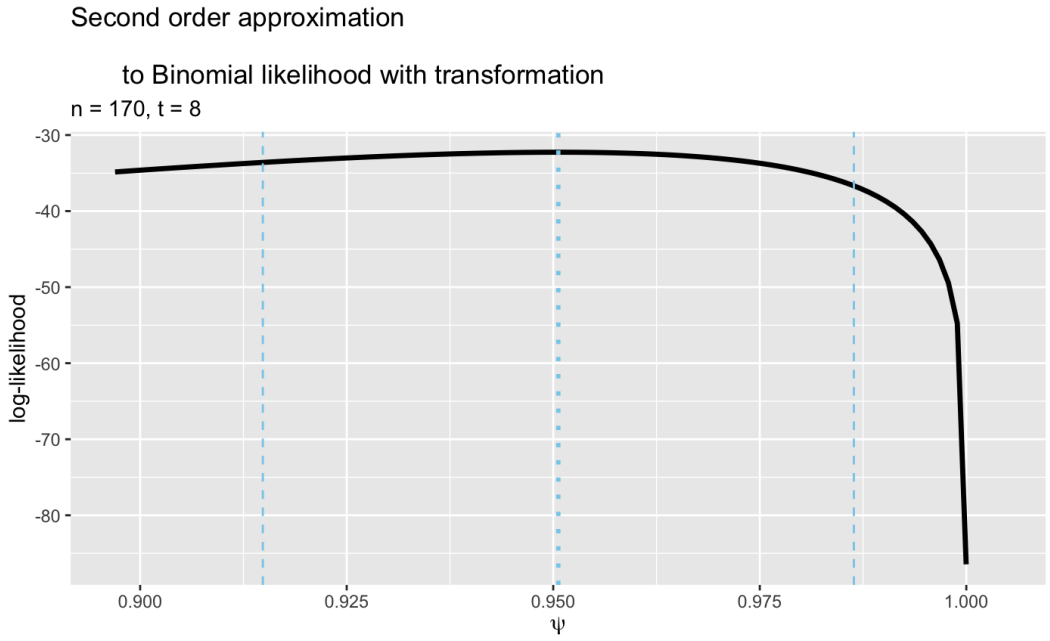
Hence, the 95% confidence interval of ψ_0 is [0.9156, 0.9857].
Since we know that $\hat{\psi_0}^{MLE} \approx 0.9506$ and $\psi_0^{null} = 0.3$, and we have the likelihood ratio test statistic $W = 2 \times ln\left[\frac{(1-\hat{\psi_0}^{MLE})^t}{(2-\hat{\psi_0}^{MLE})^n} \times \frac{(2-\psi_0^{null})^n}{(1-\psi_0^{null})^t}\right]$,

$$P(\psi_0 \neq \psi_0^{null}) = 2.822294 \times 10^{-28}$$

5.2 Method of Moments Estimation

So we have $\hat{\psi}_0^{MOM} = \frac{n-2T}{n-T}$. By substituting the value of n and T in the equation, we get that:

$$\hat{\psi_0}^{MOM} = \frac{170-2\times8}{170-8} = \frac{154}{162}$$



To find the confidence interval for ψ_0, we use parametric bootstrap method. We select $m = 15$ and $B = 1000$.

After obtaining 1000 bootstrapped estimates of ψ_0, we find that the mean of the bootstrap estimates is 0.950608 and the standard error is 0.004844. So we can evaluate our bias:

$$\text{Bias} = 0.950608 - \tfrac{154}{162} = -9.3849 \times 10^{-6}$$

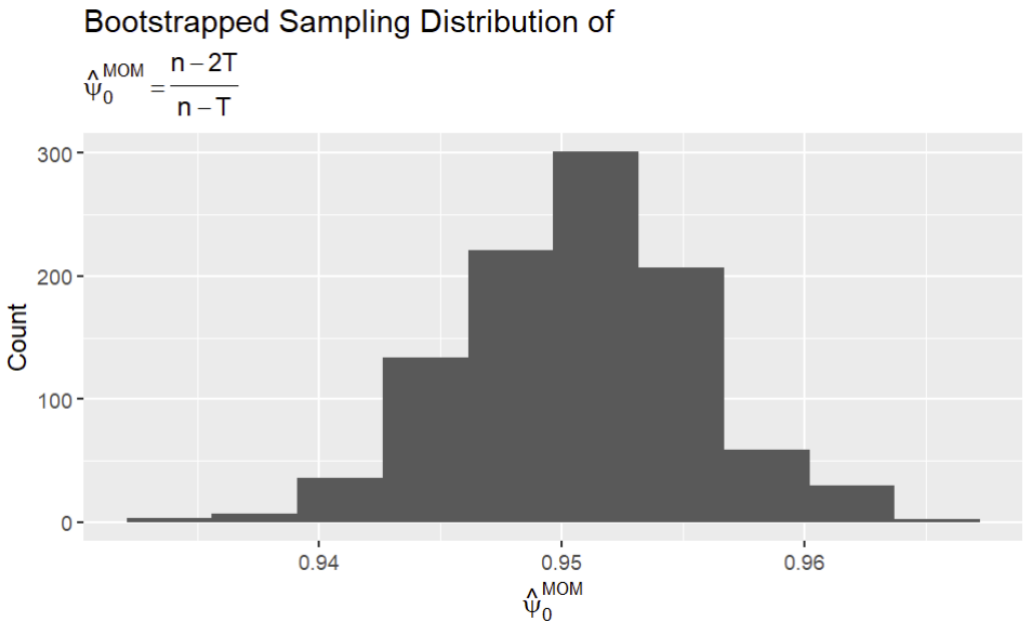
Our bias-corrected estimate is equal to

$$\tfrac{154}{162} - (-9.3849 \times 10^{-6}) = 0.950627$$

Having all these information, we can finally calculate our confidence interval for ψ_0, which is

$$[0.950627 - 1.96 \times 0.004844, 0.950627 + 1.96 \times 0.004844]$$

$$[0.9411, 0.9601]$$



After generating $B = 10000$ samples from $Binom(n = 170, \pi_0 = 0.4118)$ where our observed value $t = 8$, we found that the empirical P-value is 0, which is obviously below $\alpha = 0.01$. Therefore, we can reject $H_0 : \psi = 0.3$ and favour $H_1 : \psi_0 > 0.3$. The FDA is highly likely to approve the vaccine.

6 Discussion/Conclusion

Our maximum likelihood estimator for ψ_0 exists in the 95% confidence interval [0.9156, 0.9857], which means that the true value of ψ_0 is between 0.9156 and 0.9857 with 95% confidence. Meanwhile, our method of moments estimator for ψ_0 exists in the 95% confidence interval [0.9411, 0.9601], so we know that the true value of ψ_0 is in between 0.9411 and 0.9601 with 95% confidence. We know that the paper returned that the 95% confidence interval for ψ_0 is [0.903, 0.976], meaning that the true value of ψ_0 is in between 0.903 and 0.976 with 95% confidence. Thus, both of our estimators agree with the paper for the true value of ψ_0, which the paper states to be approximately 95%.

Additionally, we then tested the efficacy of the vaccination ψ_0 like so H_0: ψ_0 = 0.3 vs. H_1: $\psi_0 > 0.3$ in order to determine if the FDA would approve of this vaccine. Our maximum likelihood estimator returned a P-Value from the likelihood ratio test statistic of 2.822294×10^{-28}, which is effectively zero. Our method of moments estimator returned a P-Value empirically of 0. Both of these P-Values are very close to zero, meaning that it is lower than presumed α level, which makes us reject the null hypothesis. This means that the BNT162b2 vaccine would be approved by the FDA for general use.

A clear strength with the method of moments estimator is its ease of use, specifically in how it is quite simple to calculate the expected value of a random variable as well as the mean of observed values. This calculations become increasingly more difficult when trying to find the maximum likelihood estimator, as calculating the likelihood function and taking the derivative of said function. However, the maximum likelihood estimator does not rely on any numerical data or mean like the method of moments estimator does, meaning that it can be abstracted regardless of whether or not data is available.

7 References

FDA. "Covid-19 Treatment and Preventive Medication." Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, 23 Aug. 2021, www.cdc.gov/coronavirus/2019-ncov/your-health/treatments for-severe-illness.html.

University of Utah Health. "MRNA Vaccines for COVID-19." University of Utah Health — University of Utah Health, 20 May 2024, healthcare.utah.edu/coronavirus/vaccine/mrna-vaccine.

National Institutes of Health. "Decades in the Making: Mrna Covid-19 Vaccines." National Institutes of Health, U.S. Department of Health and Human Services, 10 Jan. 2023, covid19.nih.gov/ nih-strategic-response-covid-19/decades-making-mrna-covid-19-vaccines.

8 Appendix

8.1 Code for MLE Taylor Approximation Graph

```
loglik.binom <- function(psi, x, n){
  ifelse((psi < 0 | psi > 1), NA,
         x*log(1-psi)-n*log((2-psi)))
}
ml.binom <- maxLik2(loglik=loglik.binom,
                    start = 0.2, x=8, n=170)


plot(ml.binom) +
  labs(title = "Second order approximation\n
       to Binomial likelihood with transformation",
       subtitle = "n = 170, t = 8",
       x = expression(psi))
```

8.2 Code for Confidence Interval (MLE)

```
t <- 8
n <- 170
z_score <- 1.96


pi_mle <- t/n
psi_mle <- (1-2*pi_mle)/(1-pi_mle)
```

```
second_deriv <- -t/((1-psi_mle)^2)+n/((2-psi_mle)^2)
estimated_se <- sqrt(-1/second_deriv)

lower_mle <- psi_mle-z_score*estimated_se
upper_mle <- psi_mle+z_score*estimated_se
```

8.3 Code for P-Value (MLE)

```
t <- 8
n <- 170
psi_0_null <- 0.3

W <- 2*log((((1-psi_mle)^t)/
((2-psi_mle)^n))*(((2-psi_0_null)^n)/((1-psi_0_null)^t)))
p_value <- pchisq(q=W, df=1, lower.tail = FALSE)
p_value
```

8.4 Code for Confidence Interval (MOM)

```
pi_mom <- 8/170
psi_mom <- 154/162

#generate sample of n = 15 from Binom(170, 8) and then
#calculate new estimate of psi_mom. Repeat 1000 times
```

```r
B = 1000

set.seed(8383)

boot_df <- tibble(

  psi_star = replicate(n = B,

                       {sample <- rbinom(n = 15, size = 170, prob = pi_mom)

                        sample_mean <- mean(sample)

                        (170 - 2 * sample_mean) / (170 - sample_mean)})

)


#make a histogram of the bootstrap estimates

ggplot(data = boot_df,

       mapping = aes(x = psi_star) ) +

  geom_histogram(bins = 10) +

  labs(title = "Bootstrapped Sampling Distribution of ",

       subtitle = expression(hat(psi)[0]^{MOM} == frac(n - 2 * T,

       n - T)),

       x = expression(hat(psi)[0]^{MOM}),

       y = "Count")


boot_mean <- mean(boot_df$psi_star)

boot_se <- sd(boot_df$psi_star)

bias <- boot_mean - psi_mom

bias_corrected_estimate <- psi_mom - bias

lower_mom <- bias_corrected_estimate - qnorm(0.975) * boot_se
```

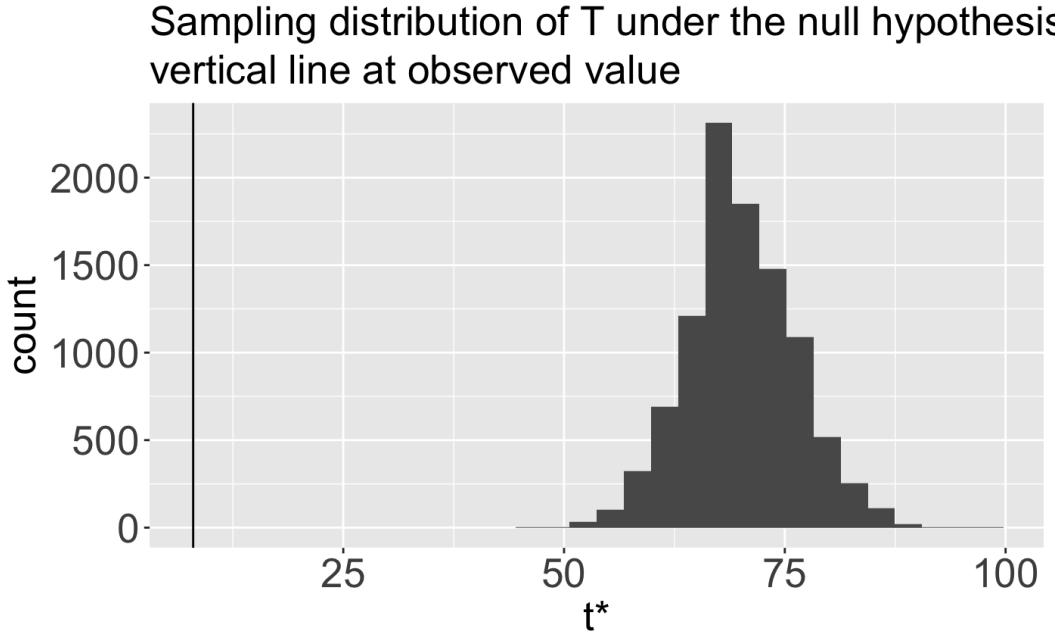
```r
upper_mom <- bias_corrected_estimate + qnorm(0.975) * boot_se


boot_df %>% summarise(boot_mean = round(boot_mean, 6),
                      boot_se = round(boot_se, 6),
                      bias = round(bias, 10),
                      bias_corrected_estimate =
                      round(bias_corrected_estimate, 6),
                      lower = round(lower_mom, 4),
                      upper = round(upper_mom, 4))
```

8.5 Code for Empirical P-Value (MOM)

```r
#Simulate T* ~ Binom(n = 170, prob = 0.4118) 10000 times
B = 10000
set.seed(8383)


null_sim <- tibble(
  tstar = replicate(n = B, expr = rbinom(n = 1, size = 170, prob = 0.4118))
)


#Graph the null sampling distribution of T
ggplot(data = null_sim, mapping = aes(x = tstar)) +
  geom_histogram() +
  geom_vline(xintercept = 8) +
```

```
  labs(x = expression(paste(t, "*")),
       title = "Sampling distribution of T under the null hypothesis",
       subtitle = "vertical line at observed value") +
  theme(plot.title = element_text(size = 20),
        plot.subtitle = element_text(size = 20),
        axis.title = element_text(size = 20),
        axis.text = element_text(size = 20))


#Calculate the empirical P-value if we observe t = 8
obs_t = 8
pval <- sum(null_sim$tstar <= obs_t)/B
```

8.6 Empirical P-Value Histogram (MOM)


Sampling distribution of T under the null hypothesis
vertical line at observed value
count
2000
1500
1000
500
0
25
50
75
100
t*