Final Project

In Progress

Three Awesome Intelligent Genius

0.1 Abstract

0.2 Keywords

0.3 Introduction

In 2019, human beings experiences an unprecedented outbreak of respiratory infections. This is a worldwide disease known as coronavirus disease 2019 (Covid-19). The first cases of novel coronavirus are detected in China and then spread to other countries across the world. During more than 3 years pandemic, "over 2 million people in the European Region have died from the disease." [1] At the beginning of the outbreak, people didn't expect it to last for a long time and treated it like a common flu. However, because the infected person is exposed to the crowd for a long time, the development of new coronary pneumonia is out of control and gradually forms a global disease.

How to effectively control the epidemic? Isolation is often the first and most effective measure that comes to mind when attempting to control an epidemic. It is a logical approach since, without person-to-person contact, the virus cannot spread. However, it is essential to understand that this approach requires a sustained effort until the government can confidently ensure the quarantine of all infected individuals. It is undeniable that Quarantine and social distancing are one of the most effective ways to prevent the virus from spreading, however, this method also lead to "elevated levels of loneliness and social isolation, which in turn produce physical- and mental-health related

---

[1](Coronavirus Disease (COVID-19) Pandemic)

repercussions." [2]

So it becomes meaningful to control the spread of the virus from the human body itself, and vaccination is one of them. The effective vaccines can greatly reduce "diseases that once routinely harmed or killed babies, children, and adults." [3]

BNT162b2 vaccine, a Covid 19 vaccine developed by Pfizer, has entered public's view. BNT162b2 is a vaccine containing modified RNA, encapsulated in lipid nanoparticles, that carries the genetic instructions for producing a stable form of the SARS-CoV-2 spike protein, which is anchored to the viral membrane. And in this research, individuals aged 16 years or older were randomly divided into two groups. One group received two doses of either a placebo or the BNT162b2 vaccine. The two doses were rigorously administered 21 days apart. And then, researchers measure the effectiveness of vaccine by record the number of occurrence of laboratory-confirmed COVID-19. By analyzing the data calculated, we could use many statistics model to illustrates the effectiveness of the BNT162b2 vaccine.

How does this work relate to other work in the scientific literature?

The following sections will analyse whether BNT162b2 vaccine is efficient and safety enough to against laboratory-confirmed Covid 19 by three methods: likelihood, Bayesian, and bootstrapping.

0.4 Statistical Method

0.5 Likelihood

Let X denote the number of Covid 19 cases among $n_1 = 17411$ subjects assigned to the BNT162b2 group while use Y denote the number of Covid 19 cases among $n_2 = 17511$ subjects assigned to the Placebo group.

Based on these, we can use π_1 denote the probability of Covid 19 cases happened in the BNT162b2 group and use π_2 denote the probability of Covid 19 cases happened in the Placebo group.

Therefore, we have $X \sim Binomial(\pi_1, n_1 = 17411)$ and $Y \sim Binomial(\pi_2, n_2 = 17511)$.

Then, we have $\psi = 1 - \frac{\pi_1}{\pi_2} = \frac{\pi_2 - \pi_1}{\pi_2} = \frac{\pi_p - \pi_v}{\pi_p}$

---

[2](Hwang et al.)

[3]("Adult Vaccination - Reasons to Vaccinate.")

```
##        Group Cases Normal No_subjects
## 1 BNT162b2     8  17403        17411
## 2  Placebo   162  17349        17511
## 3    Total   170  34752        34922
```

Based on the table provided above, we have $\pi_v = \frac{8}{17411}$ and $\pi_p = \frac{162}{17511}$. Therefore, we have

$$\psi = \frac{\pi_p - \pi_v}{\pi_p} = \frac{\frac{162}{17511} - \frac{8}{17411}}{\frac{162}{17511}} = 1 - \frac{8 \cdot 17511}{17411 \cdot 162} = 1 - \frac{140088}{2820582} = 0.9503337$$

This means vaccine BNT162b2 can reduce risk of having Covid 19 by 95% compare to placebo.

Now, want to treat the problem together. Let S denotes the total cases in two groups and let T denotes the number of cases in vaccine group from S total cases.

Then, we have $T \sim Binom(s, \pi)$ where $\pi = P(\text{BNT162b2} \mid \text{Covid 19})$. And we have $\pi = \frac{n_1 \cdot \pi_1}{n_1 \cdot \pi_1 + n_2 \cdot \pi_2}$. Because $n_1 = 17411$ and $n_2 = 17511$ are approximately equal, we have

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

$$T \sim Binom(s, \pi = \frac{\pi_v}{\pi_v + \pi_p})$$

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

$$\pi_v = \pi \cdot (\pi_v + \pi_p)$$

$$\pi_v = \pi \cdot \pi_v + \pi \cdot \pi_p$$

$$\pi \cdot \pi_p = (1 - \pi) \cdot \pi_v$$

$$\pi_p = \frac{(1 - \pi) \cdot \pi_v}{\pi}$$

Because known that: $\psi = \dfrac{\pi_p - \pi_v}{\pi_p}$

Plug π_p into ψ, we have: $\psi = \dfrac{\frac{(1-\pi)\cdot\pi_v}{\pi} - \pi_v}{\frac{(1-\pi)\cdot\pi_v}{\pi}}$

$$= \frac{1 - 2\pi}{\pi} \cdot \pi_v \cdot \frac{\pi}{(1 - \pi) \cdot \pi_v}$$

$$= \frac{1 - 2\pi}{1 - \pi}$$

$$(1 - \pi)\psi = 1 - 2\pi$$

$$\psi - \pi\,\psi = 1 - 2\pi$$

$$(2 - \psi)\,\pi = 1 - \psi$$

$$\pi = \frac{1 - \psi}{2 - \psi}$$

Based on the data given, $s = 170$, so we have $T \sim Binom(170, \pi_0)$

The likelihood function is then

4

$$f(t) = \binom{170}{t} \cdot \pi_0^t \cdot (1 - \pi_0)^{170-t}$$

$$L(\psi) = \binom{170}{t} \cdot (\frac{1-\psi}{2-\psi})^t \cdot (1 - \frac{1-\psi}{2-\psi})^{170-t}$$

$$= \binom{170}{t} \cdot (\frac{1-\psi}{2-\psi})^t \cdot (\frac{1}{2-\psi})^{170-t}$$

$$\ell(\psi) = log(\binom{170}{t} \cdot (\frac{1-\psi}{2-\psi})^t \cdot (\frac{1}{2-\psi})^{170-t})$$

$$= ln(\binom{170}{t}) + t \cdot (ln(1-\psi) - ln(2-\psi)) + (170-t) \cdot (ln(1) - ln(2-\psi))$$

$$= ln(\binom{170}{t}) + t \cdot (ln(1-\psi) - ln(2-\psi)) + (t-170) \cdot ln(2-\psi)$$

$$\frac{d}{d\psi}\ell(\psi) = \frac{d}{d\psi}(ln(\binom{170}{t}) + t \cdot (ln(1-\psi) - ln(2-\psi)) + (t-170) \cdot ln(2-\psi))$$

$$= -\frac{t}{1-\psi} + \frac{t}{2-\psi} - \frac{t-170}{2-\psi}$$

$$= -\frac{t}{1-\psi} + \frac{170}{2-\psi}$$

$$= \frac{-t(2-\psi) + 170(1-\psi)}{(1-\psi)(2-\psi)}$$

$$= \frac{-2t - 2\psi + 170 - 170\psi}{(1-\psi)(2-\psi)}$$

$$= \frac{-2t + 170 - 172\psi}{(1-\psi)(2-\psi)} = 0$$

$$-2t + 170 - 172\psi = 0$$

$$\psi = \frac{170 - 2t}{172}$$

Given that 8 cases in Vaccine group: $\psi = \dfrac{170 - 16}{172}$

$$= 0.8953488$$

Therefore, $\widehat{\psi^{mle}} = 0.8953488$ is a candidate for MLE

$$
\frac{d^2}{d\psi^2}\ell(\psi) = \ell''(\psi)
$$

$$
= \frac{d^2}{d\psi^2}\left(-\frac{t}{1-\psi} + \frac{170}{2-\psi}\right)
$$

$$
= \frac{-t}{(1-\psi)^2} + \frac{170}{(2-\psi)^2}
$$

Given that 8 cases in Vaccine group:
$$
= \frac{-8}{(1-\psi)^2} + \frac{170}{(2-\psi)^2}
$$

$$
= \frac{-8\cdot(2-\psi)^2 + 170\cdot(1-\psi)^2}{(1-\psi)^2\cdot(2-\psi)^2}
$$

Because $0 < \psi < 1$: $(1-\psi)^2 \cdot (2-\psi)^2 > 0$

$$
-8\cdot(2-\psi)^2 + 170\cdot(1-\psi)^2 < 0
$$

we have $\ell''(\psi) < 0$

Therefore, $\widehat{\psi^{mle}} = 0.8953488$ is a MLE for ψ. Based on the given information in Chapter 25, we

have the estimated variance of the MLE of ψ is $-\frac{1}{\ell''(\widehat{\psi^{mle}})}$

$$Var[\widehat{\psi^{mle}}] = -\frac{1}{\ell''(\widehat{\psi^{mle}})}$$

$$= -\frac{1}{\frac{-8}{(1-0.8953488)^2} + \frac{170}{(2-0.8953488)^2}}$$

$$= -\frac{1}{\frac{-8}{0.01095187} + \frac{170}{1.220254}}$$

$$= -\frac{1}{-730.4689 + 139.3153}$$

$$= -\frac{1}{-591.1536}$$

$$= 0.001691608$$

$$SD[\widehat{\psi^{mle}}] = \sqrt{Var[\widehat{\psi^{mle}}]}$$

$$= \sqrt{0.001691608}$$

$$= 0.04112916$$

Then, to find the confidence interval, we have:

$$\widehat{\psi^{mle}} \pm Z \cdot \widehat{SE}$$

$$\widehat{\psi^{mle}} \pm Z \cdot \sqrt{-\frac{1}{\ell''(\widehat{\psi^{mle}})}}$$

$$0.8953488 \pm 1.96 \cdot 0.04112916$$

$$0.8953488 \pm 0.08061315$$

$$[0.8147356, 0.9759619]$$

The null hypothesis is $\psi = \frac{1-2\pi}{1-\pi} = 0.3$ and $\pi = \frac{1-\psi}{2-\psi} = \frac{0.7}{1.7} = 0.4117647$.

And the $\widehat{\psi^{mle}} = 0.8953488$, we have $\pi = \frac{1-\psi}{2-\psi} = \frac{1-0.8953488}{2-0.8953488} = \frac{0.1046512}{1.104651} = 0.09473689$

7

```r
pbinom(q = 0.09473689* 170,size = 170, prob = 0.4117647, lower.tail = F)
```

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

0.6 Bootstrap

```r
set.seed(666)


case_vaccine <- 8
total_vaccine <- 17411
case_placebo <- 162
total_placebo <- 17511
total_case <- case_vaccine + case_placebo


pi_v <- case_vaccine / total_vaccine
pi_p <- case_placebo / total_placebo
psi_hat <- (pi_p - pi_v) / pi_p
pi_hat <- pi_v / (pi_p + pi_v)


B <- 10000
```

0.6.1 Parametric bootstrap

```r
para_bootstrap_samples <- lapply(1:B, function(X) {
  bootstrap_t <- rbinom(1, size = total_case, prob = pi_hat)
  bootstrap_pi <- bootstrap_t / total_case
  bootstrap_psi_hat <- (1 - 2 * bootstrap_pi) / (1 - bootstrap_pi)
  data.frame(sample_pi = bootstrap_pi,
             sample_psi_hat = bootstrap_psi_hat)
})
```

```r
para_sample_summary <- do.call(rbind, para_bootstrap_samples)


# Calculate the binwidth using the method we have learned in class:

# Here, n = B = 10000

# log2(10000) + 1 = 14.28771

# Here, I took a number of 15.

# Now, I need the range from my data.

# para_sample_summary$sample_psi_hat %>% range()

# From the summary, I got the range for the data is: 1.0000000 - 0.8741722= 0.1258278

# Thus, here, I resize the binwidth into 0.1258278/15 and breaks the bar by (0.8741722, 1.0000
ggplot(data = para_sample_summary,

       mapping = aes(x = sample_psi_hat)) +

  geom_histogram(binwidth = 0.1258278/15,

                 breaks = seq(0.8741722, 1.0000000, 0.1258278/15),

                 alpha = 0.8,

                 color = 'gray') +

  labs(title = "The distribution of bootstrapped sample ")
```
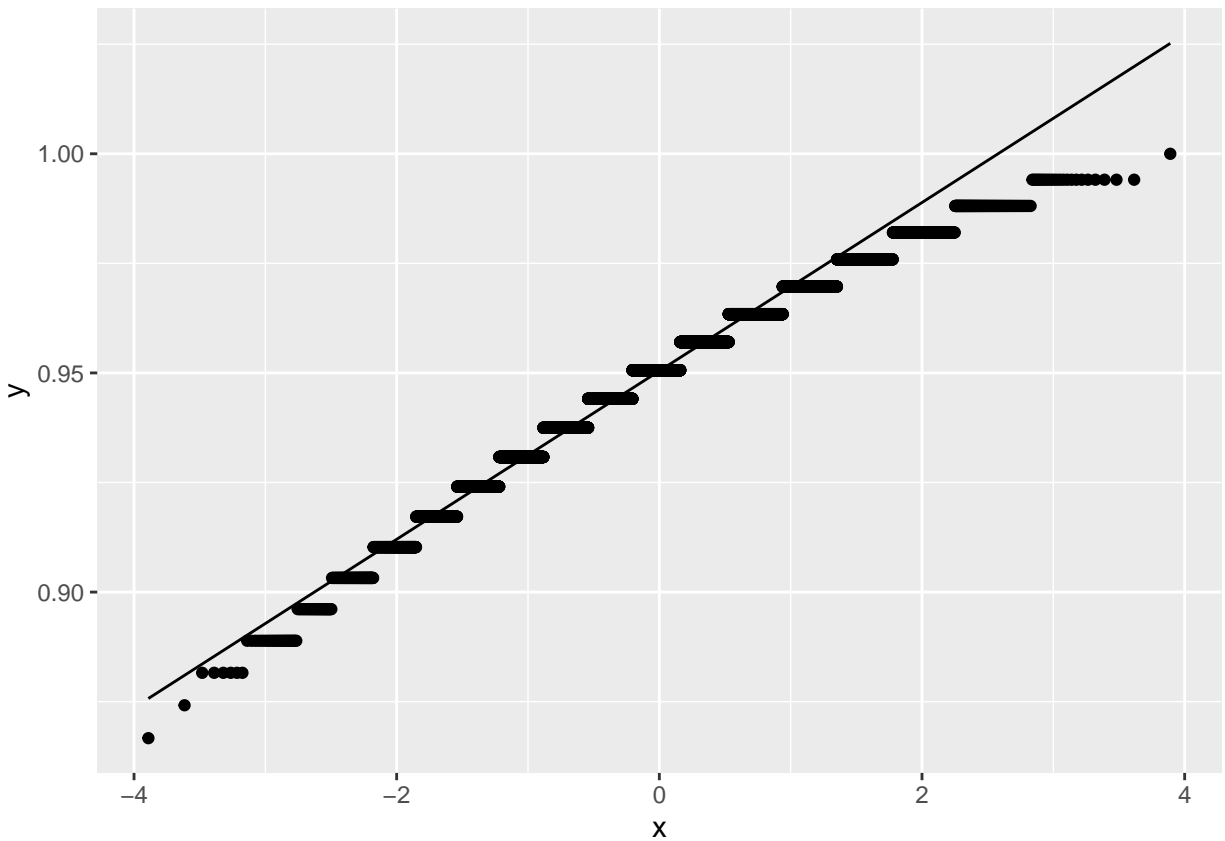
The distribution of bootstrapped sample
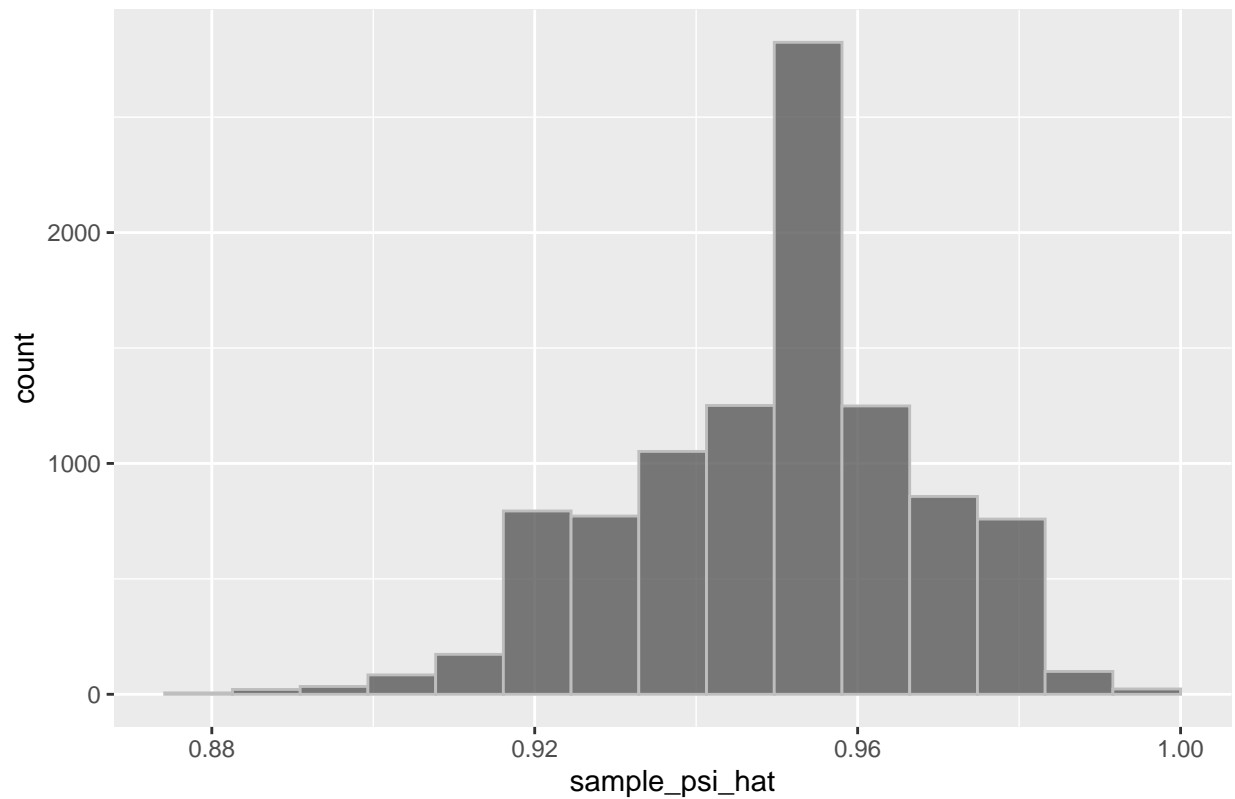

```
ggplot(data = para_sample_summary, mapping = aes(sample = sample_psi_hat)) +

  stat_qq(distribution = qnorm) +

  stat_qq_line(distribution = qnorm)
```

```
# without bias correction
a = 0.05
z <- qnorm(p = 1 - a/2)
para_ci <- para_sample_summary %>% summarize(psi_hat = mean(sample_psi_hat),
                                             s = sd(sample_psi_hat),
                                             lower = psi_hat - z * s,
                                             upper = psi_hat + z * s)
para_ci$lower
```

```
## [1] 0.9145186
```

```
para_ci$upper
```

```
## [1] 0.985043
```

```r
# Perform hypothesis testing
h_0_psi <- 0.3
para_p_value <- sum(para_sample_summary$sample_psi_hat > h_0_psi) / B
para_p_value
```

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

```r
# with bias correction
bias <- mean(para_sample_summary$sample_psi_hat) - psi_hat
ci_correction <- para_sample_summary %>% summarize(psi_hat = mean(sample_psi_hat),
                                    s = sd(sample_psi_hat),
                                    lower = psi_hat - bias - z * s,
                                    upper = psi_hat - bias + z * s)
ci_correction$lower
```

```
## [1] 0.9150714
```

```r
ci_correction$upper
```

```
## [1] 0.9855959
```

```r
corr_p_value <- sum(para_sample_summary$sample_psi_hat > h_0_psi) / B
corr_p_value
```

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

```r
mosaic::prop.test(x = sum(para_sample_summary$sample_psi_hat > h_0_psi),
                  n = B,
                  p = 0.3,
                  alt = "greater",
                  conf.level = 0.95,
                  correct = FALSE)
```

```
## Registered S3 method overwritten by 'mosaic':
##   method                              from
```

```
##    fortify.SpatialPolygonsDataFrame ggplot2

##
##  1-sample proportions test without continuity correction
##
## data:  sum out of Bpara_sample_summary$sample_psi_hat > h_0_psi out of B
## X-squared = 23333, df = 1, p-value < 2.2e-16
## alternative hypothesis: true p is greater than 0.3
## 95 percent confidence interval:
##  0.9997295 1.0000000
## sample estimates:
## p
## 1
```

0.6.2   Non-parametric bootstrap

Simple percentile method

```r
non_bootstrap_samples <- lapply(1:B, function(X) {
  # Resample from the combined dataset
  bootstrap_data <- sample(c(rep("vaccine", case_vaccine),
                             rep("placebo", case_placebo)),
                           total_case, replace = TRUE)
  bootstrap_case_vaccine <- sum(bootstrap_data == "vaccine")
  bootstrap_case_placebo <- sum(bootstrap_data == "placebo")
  bootstrap_pi_v <- bootstrap_case_vaccine / total_vaccine
  bootstrap_pi_p <- bootstrap_case_placebo / total_placebo


  bootstrap_psi_hat <- (bootstrap_pi_p - bootstrap_pi_v) / bootstrap_pi_p


  data.frame(sample_psi_hat = bootstrap_psi_hat)
})
```

```
non_sample_summary <- do.call(rbind, non_bootstrap_samples)


lower_ci <- quantile(non_sample_summary$sample_psi_hat, a/2)
upper_ci <- quantile(non_sample_summary$sample_psi_hat, 1 - a/2)


# Perform hypothesis testing
non_p_value <- sum(non_sample_summary$sample_psi_hat > h_0_psi) / B


lower_ci
```

```
##     2.5%
## 0.909741
```

```
upper_ci
```

```
##     97.5%
## 0.9819328
```

```
non_p_value
```

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

0.7   Results

0.8   Discussion/Conclusion

0.9   References

Coronavirus disease (COVID-19) pandemic. (2023, May 24).

https://www.who.int/europe/emergencies/situations/covid-19

Hwang TJ, Rabheru K, Peisah C, Reichman W, Ikeda M. Loneliness and social isolation during the COVID-19 pandemic. Int Psychogeriatr. 2020 Oct;32(10):1217-1220. doi: 10.1017/S1041610220000988. Epub 2020 May 26. PMID: 32450943; PMCID: PMC7306546.

"Adult Vaccination - Reasons to Vaccinate." Centers for Disease Control and Prevention, 22 Sept. 2022, www.cdc.gov/vaccines/adults/reasons-to-vaccinate.html.

0.10   Appendix