

Bootstrap

2024-05-22

```
data <- read_csv("data.csv", show_col_types = F)

vaccine <- data %>%
  filter(Test == "Vaccine")

placebo <- data %>%
  filter(Test == "Placebo")

n_vaccine <- vaccine$COVID + vaccine$No_COVID
n_placebo <- placebo$COVID + placebo$No_COVID

prop_vaccine <- vaccine$COVID[1] / n_vaccine
prop_placebo <- placebo$COVID[1] / n_placebo

observed_pi <- prop_vaccine / (prop_vaccine + prop_placebo)

observed_psi <- (1 - 2*observed_pi) / (1 - observed_pi)

n_bootstrap <- 10000
bootstrap_psis <- numeric(n_bootstrap)
set.seed(123)

for (i in 1:n_bootstrap) {
  vaccine_sample <- sample(c(0, 1), size = n_vaccine, replace = TRUE, prob = c(1 - prop_vaccine, prop_vaccine))
  placebo_sample <- sample(c(0, 1), size = n_placebo, replace = TRUE, prob = c(1 - prop_placebo, prop_placebo))

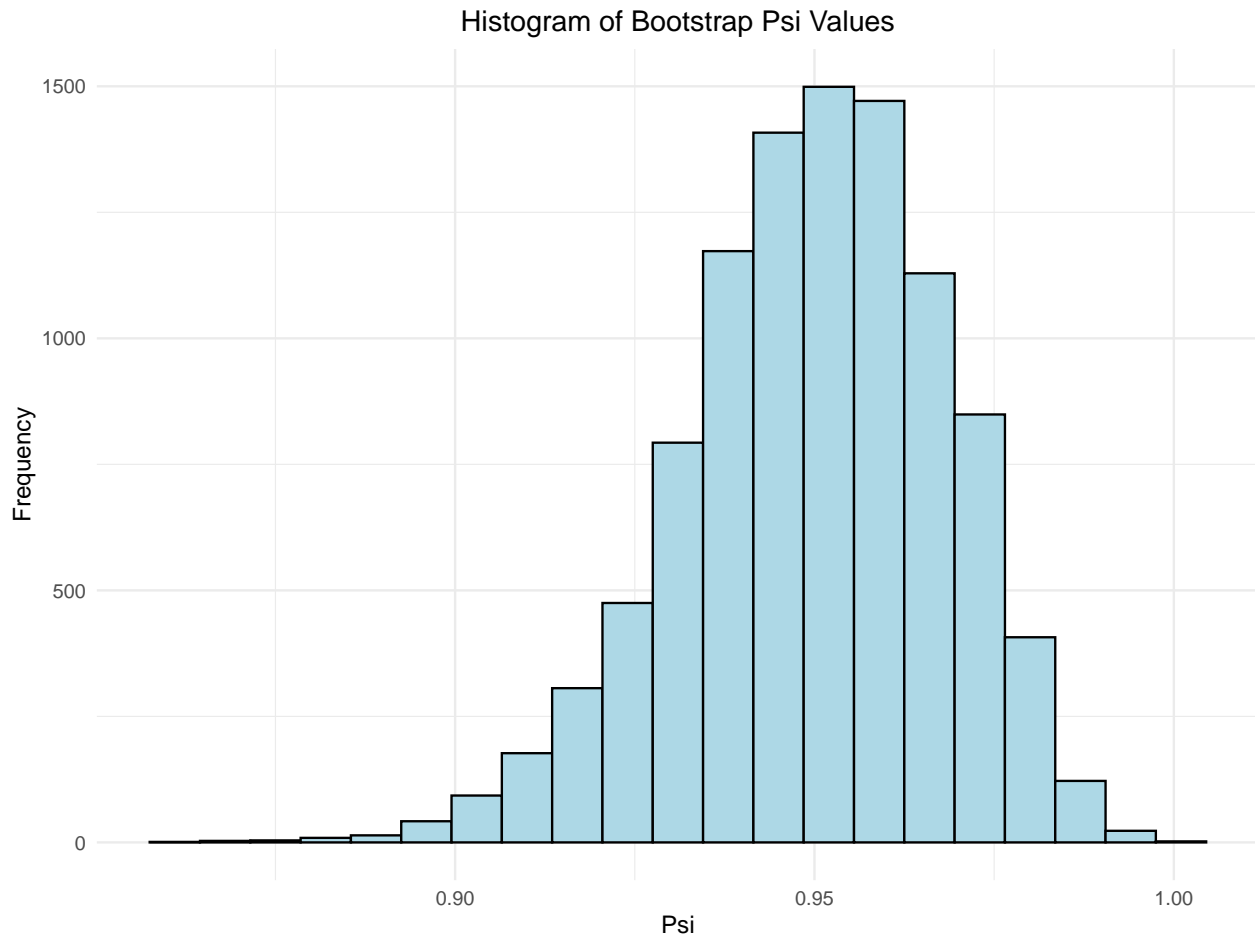
  prop_vaccine_boot <- mean(vaccine_sample)
  prop_placebo_boot <- mean(placebo_sample)

  bootstrap_pi <- prop_vaccine_boot / (prop_vaccine_boot + prop_placebo_boot)

  bootstrap_psis[i] <- (1 - 2 * bootstrap_pi) / (1 - bootstrap_pi)
}

bootstrap_df <- data.frame(psi = bootstrap_psis)

ggplot(bootstrap_df, aes(x = psi)) +
  geom_histogram(binwidth = 0.007, fill = "lightblue", color = "black") +
  labs(title = "Histogram of Bootstrap Psi Values", x = "Psi", y = "Frequency") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))
```



```
overall_ci <- quantile(bootstrap_psis, c(0.025, 0.975))

ci_data <- data.frame(
  Iteration = 1:n_bootstrap,
  Lower = numeric(n_bootstrap),
  Upper = numeric(n_bootstrap)
)

for (i in 1:n_bootstrap) {
  sample_psis <- sample(bootstrap_psis, n_bootstrap, replace = TRUE)
  ci_data$Lower[i] <- quantile(sample_psis, 0.025)
  ci_data$Upper[i] <- quantile(sample_psis, 0.975)
}

print(overall_ci)

##      2.5%      97.5%
## 0.9102751 0.9817740

plot_data <- ci_data[seq(1, n_bootstrap, by = 300), ]

ggplot(plot_data, aes(y = Iteration)) +
  geom_vline(xintercept = observed_psi, linetype = "solid", color = "red") +
  geom_text(aes(x = observed_psi-0.0125, y = max(Iteration) + 300, label = "Observed Psi"), color = "red",
    size = 12) +
  geom_segment(aes(yend = Iteration, x = Lower, xend = Upper), color = "navy") +
```

```

geom_vline(xintercept = overall_ci[1], linetype = "dashed", color = "red") +
geom_vline(xintercept = overall_ci[2], linetype = "dashed", color = "red") +
geom_text(aes(x = overall_ci[1] + 0.0129, y = max(Iteration) + 300, label = "Lower Bound"), color = "red") +
geom_text(aes(x = overall_ci[2] - 0.0129, y = max(Iteration) + 300, label = "Upper Bound"), color = "red") +
labs(title = "Confidence Intervals of Bootstrap Samples",
      y = "Iteration",
      x = "Confidence Interval",
      subtitle = "Each line segment represents a 95% confidence interval for a bootstrap sample") +
theme_minimal() +
theme(plot.title = element_text(hjust = 0.5),
      plot.subtitle = element_text(hjust = 0.5, size = 10),
      axis.text.y = element_text(hjust = 1))

```

