

Airbag Status Analysis

Haochong(Rogers) Yang

03/08/2022

This file mainly analyzed the relationship between fatality rate of passengers and whether the vehicles have airbags or not.

(Adapted from “Biostatistics for the Biological and Health Sciences”)

The table below presents data from a random sample of passengers sitting in the front seat of cars involved in car crashes. Researchers are interested in whether the fatality rates (i.e. death rates) differ for passengers in cars with airbags and passengers in cars without airbags.

	Airbag available	No airbag available
Passenger Fatalities	45	62
Total number of Passengers	10,541	9,867

The code below creates a tidy data frame for this problem, using the R command `rep`. This function creates a vector which replicates its first argument the number of times indicated by its second argument. For example, the `rep("hello", 5)` creates a vector with 5 elements, each of which is “hello”.

```
data <- tibble(group=c(rep("airbag",10541),rep("no_airbag",9867)),
  outcome=c(rep("dead",45), rep("alive",10541-45),
    rep("dead",62), rep("alive",9867-62)))
```

Hypothesis for comparing the proportions of deaths in cars with and without airbags.

The null hypothesis(H_0) is that the proportion of deaths in cars with airbags equals the proportion of deaths in cars without airbags.

$$H_0 : p_{airbags} - p_{noairbags} = 0$$

The alternate hypothesis(H_1) is that the proportion of deaths in cars with airbags is different from the proportion of deaths in cars without airbags.

$$H_1 : p_{airbags} - p_{noairbags} \neq 0$$

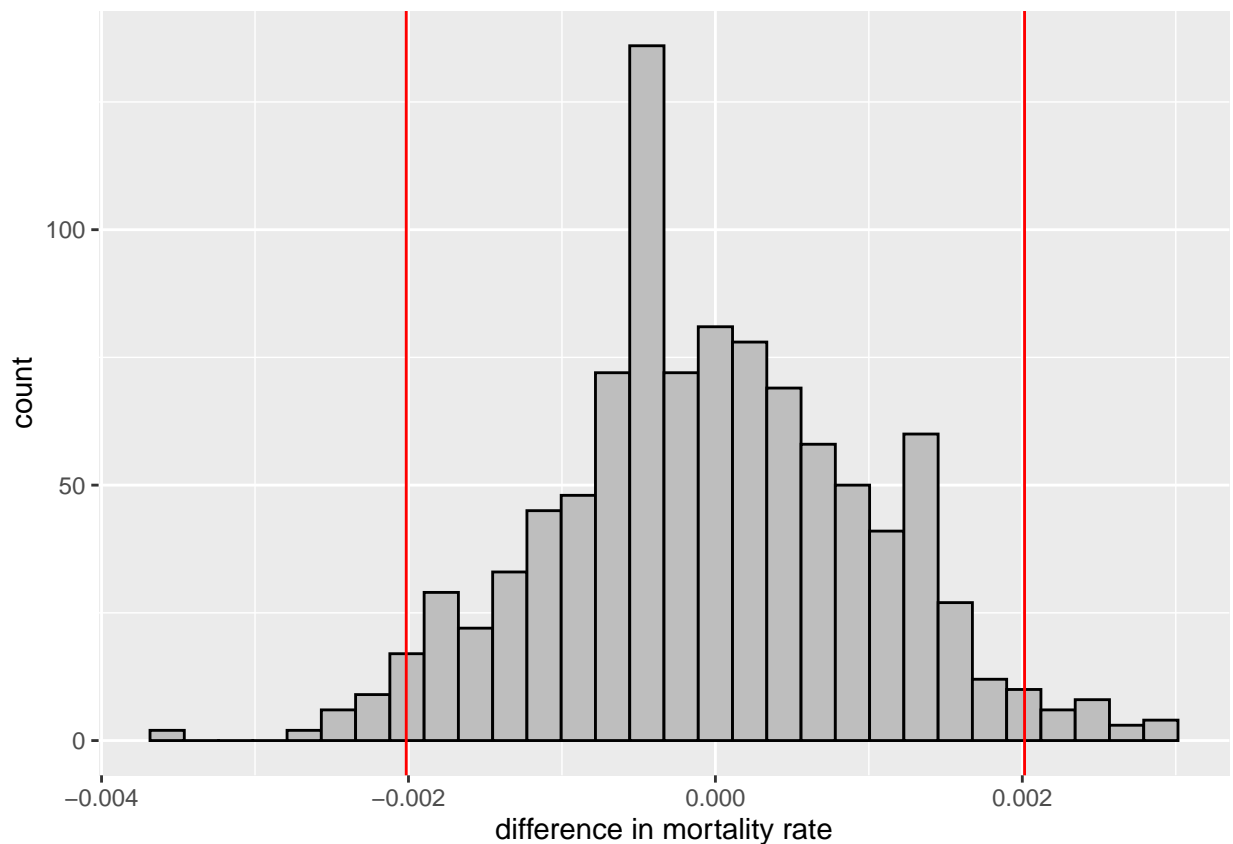
Hypothesis test for the hypotheses stated above.

```
set.seed(107) # Replace the number in the parentheses with the 1st, 3rd, and 5th
# digits in your student number.
test_statistic <- 45 / 10541 - 62 / 9867
n_repetitions <- 1000
simulated_stats <- rep(NA, n_repetitions)
```

```

for (i in 1 : n_repetitions)
{
  sim_data <- data %>%
    mutate(group = sample(group))
  sim_value <- sim_data %>%
    group_by(group) %>%
    summarize(p_mortality = sum(outcome == "dead") / n()) %>%
    summarize(value = diff(p_mortality))
  simulated_stats[i] <- as.numeric(sim_value)
}
simulation <- tibble(p_diff = simulated_stats)
simulation %>% ggplot(aes(x = p_diff)) +
  geom_histogram(color = "black", fill = "gray") +
  labs(x = "difference in mortality rate") +
  geom_vline(xintercept = abs(test_statistic), color = "red") +
  geom_vline(xintercept = -abs(test_statistic), color = "red")

```



```

p_value <- simulation %>%
  filter(p_diff >= abs(test_statistic) | p_diff <= -abs(test_statistic)) %>%
  summarize(pvalue = n() / (10541 + 9867))
as.numeric(p_value)

```

```
## [1] 0.002842023
```

Based on my analysis above, I will reject the null hypothesis at the 0.1 significance level because the p-value is 0.00284, which is smaller than 0.1.

Based on my result in the hypothesis test, since I rejected H_0 and the null hypothesis might be true, so I might made the type I error.