

hw6

2024-12-27

Problem Overview

The company has developed a new webpage with the goal of increasing the number of users who convert (i.e., users who decide to pay for the company's product). We need to perform an A/B test to determine whether there is a statistically significant difference in conversion rates between the old webpage (control group) and the new webpage (treatment group).

Objective

define a function `analyze_ab_test()` that takes two datasets: one for the control group (old webpage) and one for the test group (new webpage), along with the significance level (α). The function will perform a hypothesis test to evaluate the difference in conversion rates and provide a p-value to inform the decision of whether to use the new webpage or maintain the old one.

A/B Test Explanation

- **Control Group:** Old webpage (group = "control")
- **Treatment Group:** New webpage (group = "treatment")
- **Conversion Variable:** Whether a user converted (0 = did not purchase, 1 = purchased)

```
# Load the data
library(readr)
ecommerce <- read_csv("C:/Users/Ava/Desktop/R/HW6/ecommerce.csv")
```

```
## New names:
## Rows: 294482 Columns: 6
## — Column specification
## _____ Delimiter: "," chr
## (2): group, country dbl (3): ...1, user_id, converted time (1): timestamp
## i Use `spec()` to retrieve the full column specification for this data. i
## Specify the column types or set `show_col_types = FALSE` to quiet this message.
## • `` -> `...1`
```

```
View(ecommerce)
```

```

# Make a function
analyze_ab_test <- function(control_data, treatment_data, alpha = 0.05) {

  control_conversion_rate <- mean(control_data$converted)
  treatment_conversion_rate <- mean(treatment_data$converted)

  combined_data <- rbind(
    data.frame(group = "control", converted = control_data$converted),
    data.frame(group = "treatment", converted = treatment_data$converted)
  )

  # Perform a two-sample z-test for proportions
  # Calculate proportions and sample sizes
  n_control <- nrow(control_data)
  n_treatment <- nrow(treatment_data)
  p_control <- control_conversion_rate
  p_treatment <- treatment_conversion_rate

  # Pooled proportion
  p_pool <- (sum(control_data$converted) + sum(treatment_data$converted)) / (n_control + n_treatment)

  # Standard error
  se_pool <- sqrt(p_pool * (1 - p_pool) * (1/n_control + 1/n_treatment))

  # Z-score
  z_score <- (p_treatment - p_control) / se_pool

  # P-value (two-tailed test)
  p_value <- 2 * pnorm(-abs(z_score))

  result <- ifelse(p_value < alpha,
    paste("There is a significant difference (p-value =", round(p_value, 4),
    ")"),
    paste("There is no significant difference (p-value =", round(p_value, 4),
    ")"))

  return(list(
    p_value = p_value,
    result = result,
    control_conversion_rate = control_conversion_rate,
    treatment_conversion_rate = treatment_conversion_rate
  ))
}

```

```

# Split the data (way1:select randomly)
set.seed(123)
split <- sample(2, nrow(ecommerce), replace = TRUE, prob = c(0.5, 0.5))
control_data <- ecommerce[split == 1, ]
test_data <- ecommerce[split == 2, ]

```

```
# Applied the function
result <- analyze_ab_test(control_data, test_data, alpha = 0.05)
print(result)
```

```
## $p_value
## [1] 0.6077559
##
## $result
## [1] "There is no significant difference (p-value = 0.6078 )"
##
## $control_conversion_rate
## [1] 0.11935
##
## $treatment_conversion_rate
## [1] 0.1199639
```

Suggestion1

Based on the results:

(1.)P-value: The p-value is greater than 0.05. The result shows that there is no significant difference between the conversion rates of the old webpage (control group) and the new webpage (treatment group).

(2.)Conversion Rates: The conversion rates for both groups are very close

Accroding to these results, the company should maintain the old webpage since there is no evidence to suggest that the new webpage leads to a significantly higher conversion rate. The conversion rates are almost identical, and the difference observed is not statistically significant.

```
#Split the data (Way2: based on the 'group' variable)
control_data <- ecommerce[ecommerce$group == "control", ]
test_data <- ecommerce[ecommerce$group == "treatment", ]
```

```
# Applied the function
result <- analyze_ab_test(control_data, test_data, alpha = 0.05)
print(result)
```

```
## $p_value
## [1] 0.2151095
##
## $result
## [1] "There is no significant difference (p-value = 0.2151 )"
##
## $control_conversion_rate
## [1] 0.1203992
##
## $treatment_conversion_rate
## [1] 0.1189163
```

Suggestions2

Based on the results:

(1.)P-value: The p-value is greater than 0.05. The result shows that there is no significant difference between the conversion rates of the old webpage (control group) and the new webpage (treatment group).

(2.)Conversion Rates: The conversion rates for both groups are very close

According to these results, the company should maintain the old webpage since there is no evidence to suggest that the new webpage leads to a significantly higher conversion rate. The conversion rates are almost identical, and the difference observed is not statistically significant.

Conclusion:

The company should maintain the old webpage.