```r
# install.packages("pwr")
library(pwr)
library(tidyverse)

setwd("C:\\Users\\ASUS\\Desktop\\五 234 R\\HW6") #放你的路徑
data2 <- read.csv("ecommerce.csv")
# Remove missing values
data2 <- na.omit(data2)

# Data2
# Website A: proportion of converted
subset_A <- data2 %>%
  filter(landing_page == "old_page" & converted == 1)
purchased_A <- nrow(subset_A)
visitors_A <- nrow(data2 %>% filter(landing_page == "old_page"))
phat_A <- purchased_A / visitors_A

# Website B: proportion of converted
subset_B <- data2 %>% filter(landing_page == "new_page" & converted ==
1)
purchased_B <- nrow(subset_B)
visitors_B <- nrow(data2 %>% filter(landing_page == "new_page"))
phat_B <- purchased_B / visitors_B

# Uplift calculation
uplift <- (phat_B - phat_A) / phat_A * 100

# Pooled proportion
p_pool <- (purchased_A + purchased_B) / (visitors_A + visitors_B)

# Standard error of the pooled proportion
SE_pool <- sqrt(p_pool * (1 - p_pool) * ((1 / visitors_A) + (1 /
visitors_B)))

# Point Estimate or Difference in proportion
d_hat <- phat_B - phat_A

# Z-score
z_score <- d_hat / SE_pool

# Two-sided p-value
p_value <- pnorm(q = -abs(z_score), mean = 0, sd = 1) * 2

# Confidence interval
ci <- c(d_hat - qnorm(0.975) * SE_pool, d_hat + qnorm(0.975) * SE_pool)
```

```
# SE and CI for website A and B separately
se_hat_A <- sqrt(phat_A * (1 - phat_A) / visitors_A)
ci_A <- c(phat_A - qnorm(0.975) * se_hat_A, phat_A + qnorm(0.975) *
se_hat_A)
se_hat_B <- sqrt(phat_B * (1 - phat_B) / visitors_B)
ci_B <- c(phat_B - qnorm(0.975) * se_hat_B, phat_B + qnorm(0.975) *
se_hat_B)

# 1-sample test
prop.test(c(purchased_A + purchased_B), c(visitors_A + visitors_B))
```

```
> prop.test(c(purchased_A + purchased_B), c(visitors_A + visitors_B))

        1-sample proportions test with continuity correction

data:  c(purchased_A + purchased_B) out of c(visitors_A + visitors_B), null probability 0.5
X-squared = 170398, df = 1, p-value < 2.2e-16
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.1184886 0.1208365
sample estimates:
        p
0.1196576
```

整體接受狀況結果為顯著

```
# 2-sample test
prop.test(c(purchased_A, purchased_B), c(visitors_A, visitors_B))
```

```
> prop.test(c(purchased_A, purchased_B), c(visitors_A, visitors_B))

        2-sample test for equality of proportions with continuity correction

data:  c(purchased_A, purchased_B) out of c(visitors_A, visitors_B)
X-squared = 1.8642, df = 1, p-value = 0.1721
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.0007112309  0.0039912770
sample estimates:
   prop 1    prop 2
0.1204776 0.1188376
```

p-value 為 0.17 無顯著差異

```
# Chi-squared test
chisq.test(data2$converted, data2$website)
```

```
> chisq.test(data2$converted, data2$website)

        Chi-squared test for given probabilities

data:  data2$converted
X-squared = 259245, df = 294481, p-value = 1
```

p-value = 1 ，無顯著差異，兩個變數是獨立的
無論改網頁與否對客戶的差異不大
若有更多資料，可再針對地區、年齡、性別等資料作分析

```
# Fake data
x <- seq(from = 90, by = 10, length.out = 6)
n <- rep(1000, 6)
prop.test(x, n) # goodness-of-fit test
```

```
> prop.test(x, n) # goodness-of-fit test

        6-sample test for equality of proportions without continuity correction

data:  x out of n
X-squared = 17.195, df = 5, p-value = 0.004145
alternative hypothesis: two.sided
sample estimates:
prop 1 prop 2 prop 3 prop 4 prop 5 prop 6
  0.09   0.10   0.11   0.12   0.13   0.14
```

```
pairwise.prop.test(x, n, p.adjust.method = "none")
```

```
> pairwise.prop.test(x, n, p.adjust.method = "none")

        Pairwise comparisons using Pairwise comparison of proportions

data:  x out of n

  1       2       3       4       5
2 0.49250 -       -       -       -
3 0.15672 0.51152 -       -       -
4 0.03440 0.17452 0.52816 -       -
5 0.00532 0.04209 0.19108 0.54285 -
6 0.00059 0.00728 0.04991 0.20648 0.55592

P value adjustment method: none
```

```
# Sample size determination
pwr.anova.test(k = 2, n = NULL, f = 0.2, sig.level = 0.05, power = 0.8)
```

```
> pwr.anova.test(k = 2, n = NULL, f = 0.2, sig.level = 0.05, power = 0.8)

     Balanced one-way analysis of variance power calculation

              k = 2
              n = 99.08032
              f = 0.2
      sig.level = 0.05
          power = 0.8

NOTE: n is number in each group
```

```
pwr.t.test(n = NULL, d = 0.3, sig.level = 0.05, power = 0.8, type =
"two.sample", alternative = "greater")
```

```
> pwr.t.test(n = NULL, d = 0.3, sig.level = 0.05, power = 0.8, type = "two.sample", alternative = "greater")

     Two-sample t test power calculation

              n = 138.0716
              d = 0.3
      sig.level = 0.05
          power = 0.8
    alternative = greater

NOTE: n is number in *each* group
```