# 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)

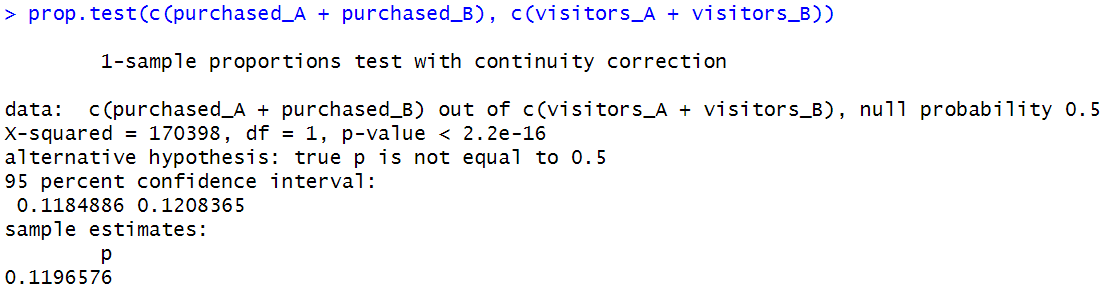
ci\_A <- c(phat\_A - qnorm(0.975) \* se\_hat\_A, phat\_A + qnorm(0.975) \* se\_hat\_A)

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

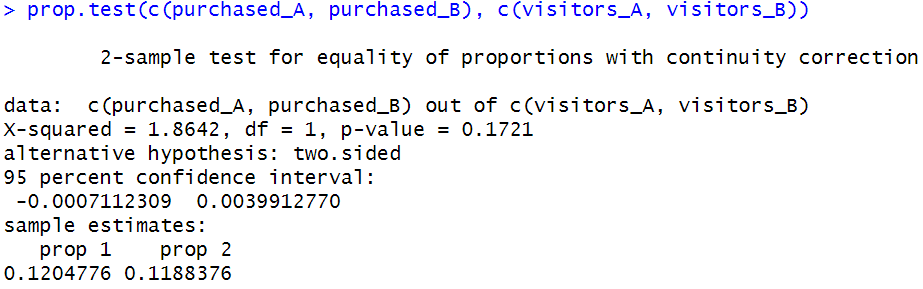
prop.test(c(purchased\_A + purchased\_B), c(visitors\_A + visitors\_B))



**整體接受狀況結果為顯著**

# 2-sample test

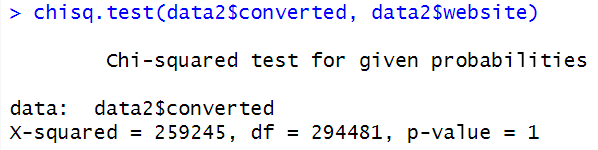
prop.test(c(purchased\_A, purchased\_B), c(visitors\_A, visitors\_B))



**p-value 為0.17無顯著差異**

# Chi-squared test

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



**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**

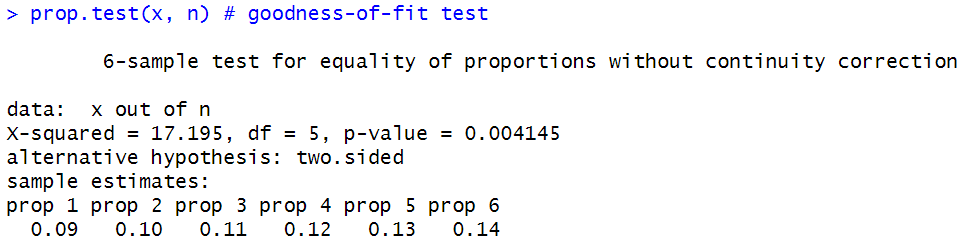
**若有更多資料，可再針對地區、年齡、性別等資料作分析**

# Fake data

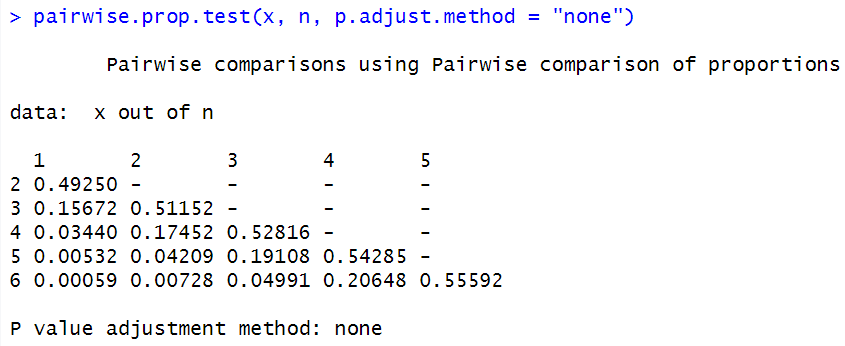
x <- seq(from = 90, by = 10, length.out = 6)

n <- rep(1000, 6)

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

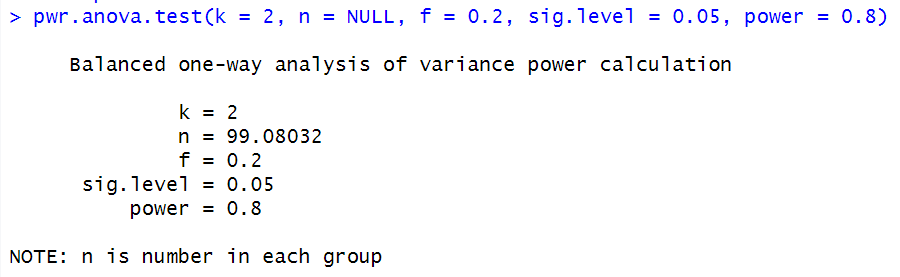


pairwise.prop.test(x, n, p.adjust.method = "none")



# Sample size determination

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



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

