20210423

```r
#####109 學年度第二學期R 語言#####
####2021/04/23####
rm(list=ls()) #Removes all items in Environment!
## set environment & import data
setwd("C:/Users/User/Desktop/R/R-project/R Statistics/data")
##
library(asbio)
babies <- read.table("babies.txt", header = T)
b1 <- na.omit(babies) #去除NA 值
bwt <- b1$bwt
### 母體平均數 ###(alternative="two.sided","less","greater")
## 母體變異數已知(常態母體或大樣本)：Z
## 假設母體標準差已知(18.236)，a = 0.05(預設)，雙尾檢定
one.sample.z(bwt, null.mu = 120, sigma = 18.236,
             alternative = "two.sided") #故不拒絕 H0

##
## One sample z-test
##         z*   P-value
##  -1.009871 0.3125573

# 區間估計
width <- qnorm(0.975)*(18.236/sqrt(length(bwt))) #qnorm(常態分配機率)
mean(bwt) + c(-width, width)

## [1] 118.4194 120.5057

## 母體變異數未知(常態母體)：t
# t.test(x, mu, alternative, conf.level)
t.test(bwt, mu = 120, alternative = "two.sided")

##
##  One Sample t-test
##
## data:  bwt
## t = -1.0048, df = 1173, p-value = 0.3152
## alternative hypothesis: true mean is not equal to 120
## 95 percent confidence interval:
##  118.413 120.512
## sample estimates:
## mean of x
##  119.4625

t.test(bwt, mu = 120, alternative = "two.sided",
       conf.level = 0.92) #其他信賴水準
```

```
##
##  One Sample t-test
##
## data:  bwt
## t = -1.0048, df = 1173, p-value = 0.3152
## alternative hypothesis: true mean is not equal to 120
## 92 percent confidence interval:
##  118.5252 120.3998
## sample estimates:
## mean of x
##  119.4625
```

```
####################
### 母體比例p ###
## prop.test(x, n, p, alt, conf)
prop.test(x = 46, n = 150, p = 1/3, alternative = "greater")
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  46 out of 150, null probability 1/3
## X-squared = 0.3675, df = 1, p-value = 0.7278
## alternative hypothesis: true p is greater than 0.3333333
## 95 percent confidence interval:
##  0.2455298 1.0000000
## sample estimates:
##         p
## 0.3066667
```

```
## 講義Ch9, p.22, 範例14
```

Cliff Obermeyer is running for Congress from the 6th District of New Jersey. Suppose 500 voters are contacted upon leaving the polls and 275 indicate they voted for Mr. Obermeyer. We will assume that the exit poll of 500 voters is a random sample of th ose voting in the 6th District. Construct a 95% confidence interval for the population proportion. Should Mr. Obermeyer be elected

```
prop.test(x = 275, n = 500, p = 0.5, alternative = "two.sided")
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  275 out of 500, null probability 0.5
## X-squared = 4.802, df = 1, p-value = 0.02843
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
##  0.5051762 0.5940439
## sample estimates:
```

```
##    p
## 0.55
```

```
####################
### 兩母體比例差
```

假設台北市與高雄市各隨機抽出 200 人與 150 人，其中台北市民有兩張以上信用卡者佔 128 人，高雄市民則為 86 人。我們可以算出兩個城市擁有兩張以上信用卡者的比例差 $p_1 - p_2$ 的 95% 信賴區間。

```
#方法1
p1 <- 128/200
p2 <- 86/150
c(p1, p2)
```

```
## [1] 0.6400000 0.5733333
```

```
center <- p1 - p2
width <- qnorm(0.975)*sqrt( p1*(1 - p1)/200 + p2*(1 - p2)/150 )
c(lower <- center - width, upper <- center + width)
```

```
## [1] -0.03672612  0.17005945
```

```
#方法2
```

某次民調從台北市抽出 3205 人，其中男性 1586 人、女性 1619 人，並且得知這兩群選民當中大專院校畢業者，男性有 214 人，女性有 219 人。我們可以檢定看看男、女兩個性別中，大專院校畢業的比例是否相等。($\alpha = 0.05$)

```
graduated = c(214, 219) #x
citizen  = c(1586, 1619) #n
prop.test(graduated, citizen) #不拒絕H0
```

```
##
##  2-sample test for equality of proportions with continuity correctio
n
##
## data:  graduated out of citizen
## X-squared = 1.9168e-29, df = 1, p-value = 1
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.02434590  0.02366981
## sample estimates:
##    prop 1    prop 2
## 0.1349306 0.1352687
```

```
### 兩母體平均數差：樣本獨立###
##假設兩母體變異數已知 25, 36, 大樣本
```

兩班學生期末成績母體變異數各為 25 跟 36，兩班個隨機抽出 30 人，則以下程式可以算出 $\mu_1 - \mu_2$ 的 95% 信賴區間。

```
x <- c(62, 57, 70, 55, 55, 64, 55, 67, 60, 64, 51, 55,
       57, 68, 61, 61, 61, 64, 57, 56, 67, 59, 60, 61,
       71, 65, 63, 63, 58, 58)
y <- c(61, 62, 68, 54, 60, 69, 53, 60, 61, 62, 67, 61,
       60, 58, 65, 66, 61, 62, 66, 71, 61, 62, 55, 58,
       66, 60, 58, 60, 63, 48)
center <- mean(x) - mean(y)
width <- qnorm(0.975)*sqrt(25/30 + 36/30)
c(lower <- center - width, upper <- center + width)
```

```
## [1] -3.228144  2.361477
```

```
##假設兩母體變異數未知，大樣本
c(var(x), var(y))
```

```
## [1] 23.93678 23.99540
```

```
center <- mean(x) - mean(y)
width <- qnorm(0.975)*sqrt(var(x)/30 + var(y)/30)
c(lower <- center - width, upper <- center + width)
```

```
## [1] -2.910762  2.044095
```

```
##小樣本，假設母體為常態分配，兩母體變異數相等
```

假設兩班學生各抽出 17 人與 15 人的成績，已知母體服從於常態分配，σ_1 與 σ_2 未知，但可以確定兩者相等。

```
#方法1
x <- c(61, 65, 57, 56, 64, 64, 57, 65, 55, 67, 67, 60,
       56, 58, 55, 60, 62)
y <- c(56, 66, 57, 56, 63, 61, 66, 61, 62, 55, 55, 62,
       55, 70, 51)
s2.A <- var(x)
s2.B <- var(y)
s2.p <- ((17 - 1)*s2.A + (15 - 1)*s2.B)/(17 + 15 - 2)
center <- mean(x) - mean(y)
width <- qt(0.975, 17 + 15-2)*sqrt(s2.p*(1/17 + 1/15))
c(lower <- center - width, upper = center + width)
```

```
##               upper
## -2.625872  4.218029
```

```
#方法2
t.test(x, y, var.equal = TRUE, alternative = "two.sided", mu=0.0, conf.
level = 0.95 )
```

```
##
##  Two Sample t-test
##
## data:  x and y
## t = 0.47511, df = 30, p-value = 0.6381
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.625872  4.218029
## sample estimates:
## mean of x mean of y
##  60.52941  59.73333
```

```
##小樣本, 假設母體為常態分配, 兩母體變異數不相等
sn1 <- var(x)/17
sn2 <- var(y)/15
v <- trunc((sn1 + sn2)^2/(sn1^2/(17 - 1) + sn2^2/(15 - 1)))
v
```

```
## [1] 26
```

```
center <- mean(x) - mean(y)
width <- qt(0.975, 26)*sqrt(var(x)/17 + var(y)/15)
c(lower <- center - width, upper <- center + width)
```

```
## [1] -2.696823  4.288980
```

兩母體平均數差：相依樣本

以下兩組資料是 30 個眼科病人經過手術後的左、右眼度數。求 $\mu_1 - \mu_2$ 的 95%信賴區間。

```
##大樣本
x <- c(129, 314, 294, 261, 188, 222, 230, 165, 117,
       248, 267, 286, 216, 292, 175, 120, 269, 195,
       248, 174, 245, 197, 100, 121, 168, 239, 207,
       180, 160, 288)
y <- c(380, 402, 374, 306, 374, 423, 344, 363, 371,
       350, 294, 318, 338, 333, 349, 350, 333, 339,
       359, 338, 347, 340, 268, 346, 425, 362, 281,
       353, 313, 346)
D <- x - y
center <- mean(D)
width <- qnorm(0.975,)*sd(D)/sqrt(30)
c(lower <- center - width, upper <- center + width)
```

```
## [1] -161.5259 -112.0741
```

```
##小樣本
```

有 19 對雙胞胎個服用一種品牌的感冒藥，觀察值為感冒復原時間(小時)。求 $\mu_1 - \mu_2$ 的 95%信賴區間。

```
x <- c(8, 5, 7, 12, 7, 5, 7, 6, 7, 8, 9,
       6, 8, 9, 8, 5, 6, 5, 5)
y <- c(11, 4, 3, 5, 7, 4, 6, 2, 1, 0, 11,
       5, 6, 7, 8, 6, 3, 2, 2)
D <-    x - y
center <- mean(D)
width <- qt(0.975, 19 - 1)*sd(D)/sqrt(19)
c(lower <- center - width, upper <- center + width)
```

```
## [1] 0.7148703 3.4956560
```

```
### 變異數檢定 ###
cancers <- read.table("cancers.txt", header = T)
x1 <- cancers$stomach
x2 <- cancers$bronchus
var.result <- var.test(x1, x2, alternative = "two.sided")
var.result$p.value #不拒絕H0
```

```
## [1] 0.05509047
```

```
t.test(x1, x2, alternative = "two.sided", mu = 0.0,
       paired = FALSE, var.equal = TRUE, conf.level = 0.95) #不拒絕H0
```

```
##
##  Two Sample t-test
##
## data:  x1 and x2
## t = 1.0383, df = 26, p-value = 0.3087
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -108.0962  328.7629
## sample estimates:
## mean of x mean of y
##  306.3333  196.0000
```

```
##講義Ch10, p.28
Welles <- c(2, 4, 9, 3, 2)
Atkins <- c(3, 7, 5, 8, 4, 3)
t.test(Welles, Atkins, alternative = "two.sided", mu = 0.0,
       paired = FALSE, var.equal = TRUE, conf.level = 0.9)
```

```
##
##  Two Sample t-test
##
## data:  Welles and Atkins
## t = -0.66205, df = 9, p-value = 0.5245
## alternative hypothesis: true difference in means is not equal to 0
## 90 percent confidence interval:
```

```
##  -3.768839  1.768839
## sample estimates:
## mean of x mean of y
##         4         5

##講義Ch10, p.30
Schadek <- c(235, 210, 231, 242, 205, 230,
             231, 210, 225, 249)
Bowyer <- c(228, 205, 219, 240, 198, 223,
            227, 215, 222, 245)
t.test(Schadek, Bowyer, alternative = "two.sided", mu = 0.0,
       paired = TRUE, var.equal = TRUE, conf.level = 0.95)

##
##  Paired t-test
##
## data:  Schadek and Bowyer
## t = 3.3045, df = 9, p-value = 0.009164
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.450985 7.749015
## sample estimates:
## mean of the differences
##                     4.6
```