

107-1 Statistics
LAB7: ESTIMATING MEANS WITH CONFIDENCE

助教：廖皓宇、吳家禎、賴冠宇
2018/11/09

1109實習：數量資料的信賴區間估計

■ Student distribution (t dist.) & standard normal distribution (Z dist.)

Lab #1: 1. Plotting probability distributions (t & Z)

Lab #2: 2. Given a probability (with df), find the t-value and Z-value

3. Given a t-value (with df) & a Z-value, find the Probabilities

Lab #3: 4. Estimate confidence interval for a population mean

Lab #4: 5. Simulate confidence interval for a population mean (μ)

THIS WEEK LAB

Basics:

- Lab #1: Comparing z-distribution and t-distribution with different degrees of freedom
- Lab #2: Introducing t-distribution functions (`dt()`, `pt()`, `qt()`, `rt()` series)

Confidence interval:

- Lab #3: Finding the confidence interval
- Lab #4: **Simulating** multiple confidence intervals to **check the idea of confidence level**

LAB #4:

Simulating multiple confidence intervals to check the idea of confidence level



THE IDEA OF CONFIDENCE INTERVAL

(Example 11.6 in textbook p.423-425; lecture slide p.19-21)

μ = Penn State的學生平均一天看幾小時的電視

根據一次的抽樣結果： $n = 175$, $\bar{x} = 2.09$, $s = 1.644$

95%信心水準下，CI = 1.84 ~ 2.34小時

● Interpretation:

- ① 若重複抽樣100次，約有95次抽樣的信賴區間會包含母體 μ
- ② 我們有95%的信心說：「所有Penn State的學生平均一天看電視的時數在1.84 ~ 2.34小時之間。」

WORK FLOW

- 1) Simulate one confidence interval
- 2) Simulate multiple confidence intervals
- 3) Check whether each confidence interval captures the population μ
- 4) Plot and return the simulation results

```
48 Sim_CI <- function (n, mu, sd, m, conf.level){
49
50   prob<-(1-conf.level)/2
51   t.value=qt(prob, df = n-1, lower.tail = FALSE)
52
53   mean.x = c()
54   se.x = c()
55
56   CI.half = c()
57   CI.Up = c()
58   CI.Lower = c()
59   Check = c()
60
61   # Multiple sampling
62   for (i in 1:m) {
63
64     # One sampling
65     x <- rnorm(n, mean = mu, sd = sd)
66     mean.x[i]<-mean(x)
67     se.x[i]<-sd(x)/sqrt(n)
68
69     # Calculate one confidence interval
70     CI.half[i] <- t.value * se.x[i]
71     CI.Up[i] <- mean.x[i] + CI.half[i]
72     CI.Lower[i] <-mean.x[i] - CI.half[i]
73     if (CI.Up[i]>mu & CI.Lower[i]<mu) Check[i]<-1 else Check[i]<-0
74   }
```

```
76 plot(c(CI.Up,CI.Lower),type="n",pch=19, xlim=c(1,m),
77       xlab="Trial", ylab=expression(mu))
78
79 abline(h = mu, col = "blue")
80
81 # Check whether each confidence interval captures the population mean (mu)
82 for (i in 1:m) {
83   if (Check[i]==1) {
84
85     points(i, mean.x[i], col = "green", pch = 10)
86     points(i, CI.Up[i], col = "green", pch = 20)
87     points(i, CI.Lower[i], col = "green", pch = 20)
88     lines(c(i,i), c(CI.Lower[i],CI.Up[i]), col = "green", pch = 19)
89
90   } else {
91
92     points(i, mean.x[i], col = "red", pch = 10)
93     points(i, CI.Up[i], col = "red", pch = 20)
94     points(i, CI.Lower[i], col = "red", pch = 20)
95     lines(c(i,i), c(CI.Lower[i],CI.Up[i]), col = "red", pch = 19)
96   }
97 }
98
99 title(expression(paste("Simulating confidence interval for ", mu)))
100
101 legend("bottomright", cex=0.6, bty = "n", ncol = 2,
102       c(expression(paste(mu," Captured")), expression(paste(mu," Not Captured"))),
103       fill = c("green","red"))
```



```
104  
105 No.Caputured <- m-sum(Check) #Check(captured)  
106 RESULT <- list(Trial=m, Sample.Size=n,  
107               Population.mean = mu, Population.sd = sd,  
108               Confidence=conf.level,  
109               No.Caputured=No.Caputured)  
110  
111 return(RESET)  
112 }
```

```
114 Sim_CI(n = 30, m = 100, mu = 50, sd = 5, conf.level =0.95)
```

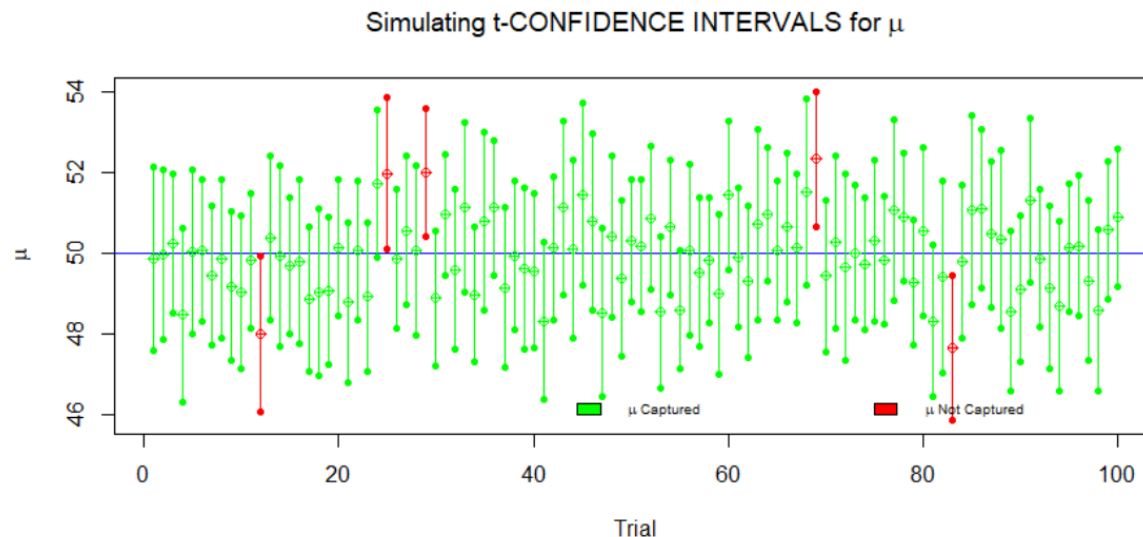
BUILD A SIMULATION FUNCTION

Function Inputs:

- Sample size (n), **population mean (μ)**, **population standard deviation (sd)**, number of sample trial (m), confidence level ($conf.level$)

Function Output: simulation results

1) The plot



2) The number of confidence interval that doesn't include the population μ

```
> Sim_CI(n = 30, m = 100, mu = 50, sd = 5, conf.level = 0.95)
$`Trial`
[1] 100

$Sample.Size
[1] 30

$Population.mean
[1] 50

$Population.sd
[1] 5

$Confidence
[1] 0.95

$No.Captured
[1] 5
```

作業7的R程式練習題

作業7 數量資料的信賴區間估計

■ 練習題6題(Ch. 11)

– 11.18; 11.32; 11.50; 11.60; 11.78; 11.80

■ R程式練習題(繳交程式碼與執行結果)

– 改寫1109實習5程式碼，模擬成對資料相差平均值的信賴區間(**population mean of paired differences**)。

1. 自訂函數**Simd_CI()**
2. **Function Input:** 樣本數、成對資料1的平均值、成對資料1的標準差、成對資料2的平均值、成對資料2的標準差、信賴區間模擬次數、信心水準
3. **Function Output:** 多次信賴區間的模擬圖、沒有包含母體的信賴區間數