

Bài 7: ƯỚC LƯỢNG THAM SỐ THỐNG KÊ

1. Khoảng tin cậy cho trung bình

Bài toán: Xét biến ngẫu nhiên $X \sim N(\mu; \sigma^2)$, cần tìm khoảng tin cậy (KTC) cho μ với độ tin cậy $100(1 - \alpha)\%$.

Với mẫu cụ thể $(X_1 = x_1, \dots, X_n = x_n)$, KTC với độ tin cậy $100(1 - \alpha)\%$ cho kỳ vọng μ có dạng

$$\bar{x} - \varepsilon \leq \mu \leq \bar{x} + \varepsilon$$

Trong đó, ε : sai số hay độ chính xác.

Phương sai σ^2	Cỡ mẫu	
	$n \leq 30$	$n > 30$
Đã biết	$\epsilon = z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$	
Chưa biết	$\epsilon = t_{1-\alpha/2}^{n-1} \frac{s}{\sqrt{n}}$	$\epsilon = z_{1-\alpha/2} \frac{s}{\sqrt{n}}$

1. Khoảng tin cậy cho trung bình

Trong đó, ϵ : sai số hay độ chính xác.

<div>Phương sai σ^2</div> <div>Cỡ mẫu</div>	$n \leq 30$	$n > 30$
Đã biết	$\epsilon = z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$	
Chưa biết	$\epsilon = t_{1-\alpha/2}^{n-1} \frac{s}{\sqrt{n}}$	$\epsilon = z_{1-\alpha/2} \frac{s}{\sqrt{n}}$

$z_{1-\frac{\alpha}{2}}$: phân vị mức $1 - \frac{\alpha}{2}$ của biến ngẫu nhiên có phân phối chuẩn hóa,
 $Z \sim N(0; 1)$.

$t_{1-\frac{\alpha}{2}}^{n-1}$: phân vị mức $1 - \frac{\alpha}{2}$ của biến ngẫu nhiên có phân phối Student với
 $n - 1$ bậc tự do.

1. Khoảng tin cậy cho trung bình

Trong đó,

- ε : sai số hay độ chính xác.
- $z_{1-\frac{\alpha}{2}}$: phân vị mức $1 - \frac{\alpha}{2}$ của biến ngẫu nhiên có phân phối chuẩn hóa, $Z \sim N(0; 1)$.
- $t_{1-\frac{\alpha}{2}}^{n-1}$: phân vị mức $1 - \frac{\alpha}{2}$ của biến ngẫu nhiên có phân phối Student với $n - 1$ bậc tự do.

Trong **R**

$$z_{1-\frac{\alpha}{2}} = \text{qnorm}(1 - \text{alpha}/2)$$

$$t_{1-\frac{\alpha}{2}}^{n-1} = \text{qt}(1 - \text{alpha}/2, n - 1)$$

Khi n đủ lớn thì $t_{1-\frac{\alpha}{2}}^{n-1} \approx z_{1-\frac{\alpha}{2}}$.

2. Khoảng tin cậy cho tỷ lệ

Bài toán: Cần tìm khoảng tin cậy cho tỷ lệ phần tử thỏa tính chất A trong tổng thể với độ tin cậy $100(1 - \alpha)\%$, chẳng hạn, tỷ lệ sản phẩm kém chất lượng trong một lô hàng, tỷ lệ người ủng hộ một ứng cử viên nào đó trong một cuộc bầu cử, tỷ lệ người dân có thu nhập trên 7 triệu đồng/tháng tại 1 thành phố,...

- Lấy mẫu cỡ n .
- Gọi Y = số phần tử thỏa tính chất A trong n phần tử khảo sát, thì $Y \sim B(n; p)$.
 - Với p là tỷ lệ cần tìm. Từ mẫu khảo sát, ta có $Y = y$.
 - Tỷ lệ mẫu $\hat{p} = \frac{y}{n}$.

2. Khoảng tin cậy cho tỷ lệ

Khoảng tin cậy $100(1 - \alpha)\%$ cho tỷ lệ p có dạng

$$\hat{p} - \varepsilon \leq p \leq \hat{p} + \varepsilon$$

Trong đó

- $\varepsilon = z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$: sai số hay độ chính xác.
- $z_{1-\frac{\alpha}{2}}$: phân vị mức $1 - \frac{\alpha}{2}$ của biến ngẫu nhiên có phân phối chuẩn hóa, $Z \sim N(0; 1)$.

Để tìm $z_{1-\frac{\alpha}{2}}$ trong R, sử dụng: $z_{1-\frac{\alpha}{2}} = \text{qnorm}(1 - \alpha/2)$

3. Ý nghĩa KTC cho tham số cần ước lượng

Nếu chúng ta lấy mẫu nhiều lần, mỗi mẫu ta tính 1 KTC thì sẽ có xấp xỉ $100(1 - \alpha)\%$ KTC chứa tham số cần ước lượng.

Ví dụ: Đoạn chương trình sau mô phỏng 50 KTC cho tỷ lệ xuất hiện mặt sấp khi tung một đồng xu, trong đó $\alpha = 5\%$.

```
m = 50; n=30; p = .5; alpha = 0.05 # Tung 30 đồng xu cân đối 50 lần
y = rbinom(m,n,p); y
p.hat = y/n # Tính tỷ lệ mẫu

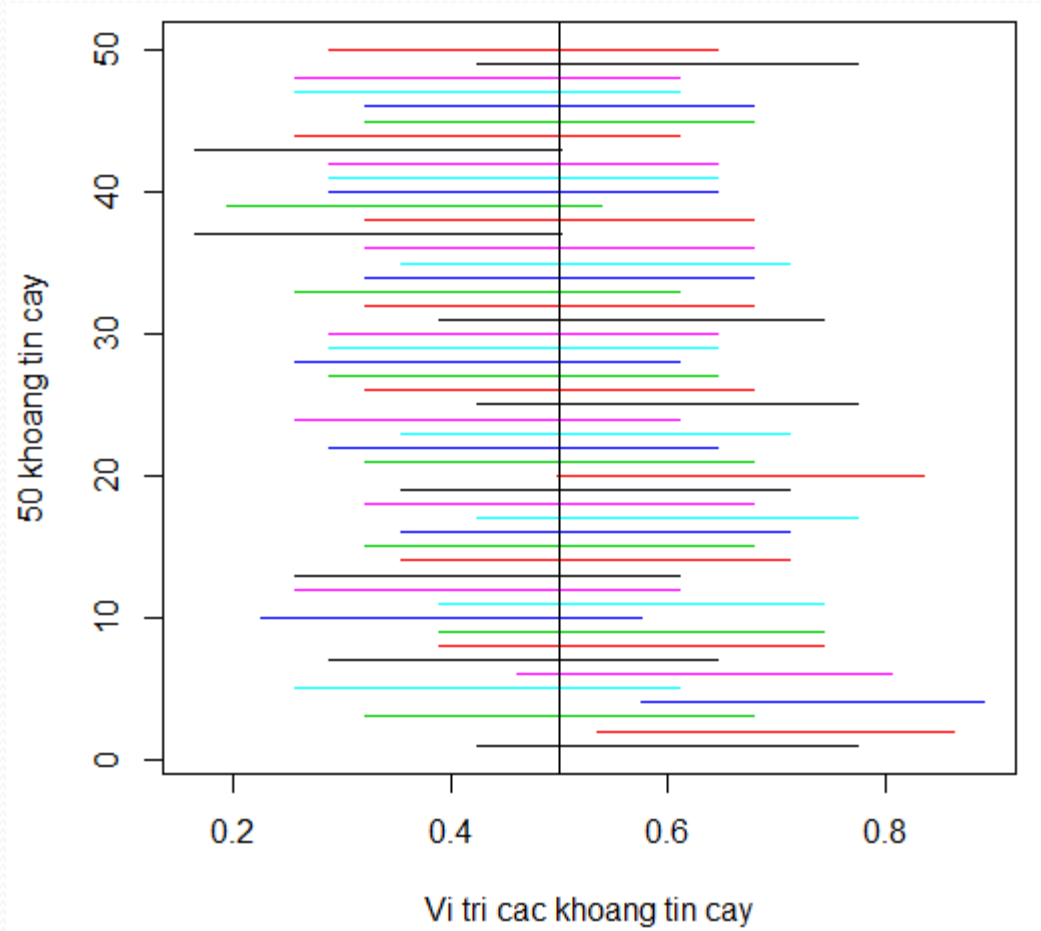
epsilon = qnorm(1-alpha/2)*sqrt(p.hat*(1-p.hat)/n) # Dung sai

matplot(rbind(p.hat - epsilon, p.hat + epsilon),
        rbind(1:m,1:m),type="l",lty=1,
        xlab = "Vi tri cac khoang tin cay",
        ylab = "50 khoang tin cay")

# Ve 50 khoang tin cay
abline(v=p) # Ve duong thang p = 0.5
```

3. Ý nghĩa KTC cho tham số cần ước lượng

Ta được kết quả



4. Bài tập

Bài 1: Tạo ngẫu nhiên 35 giá trị của biến ngẫu nhiên có phân phối chuẩn với trung bình bằng 10 và độ lệch chuẩn 5. Tìm khoảng tin cậy 95% cho kỳ vọng của biến ngẫu nhiên chuẩn dựa vào số liệu vừa tạo.

```
# Bai 1
# Tao mau 35 gia tri cua bien ngau nhien  $X \sim N(10;5^2)$ 
n = 35
x <- rnorm(n,10,5);x
#Tim KTC 95% cho ky vong, alpha = 5%
x.bar = mean(x)
alpha = 0.05
s = 5
epsilon = qnorm(1-alpha/2)*s/sqrt(n)
cat('KTC',100*(1-alpha),'% cho ky vong mu la:\n')
cat('[', x.bar - epsilon, ';' , x.bar + epsilon , ']\n')

# Xoa du lieu duoc gan
rm(list=ls(all=T))
```

4. Bài tập

Bài 2: Số liệu thống kê về doanh số bán hàng của một siêu thị cho ở file data31.xls:

a) Đọc dữ liệu từ file data31.xls vào R.

```
# Bai 2
# a) Nhap du lieu tu file data31.xls vao R
dat31 <- read.csv("data31.csv",header = T)
attach(dat31); names(dat31)
```

4. Bài tập

Bài 2: Số liệu thống kê về doanh số bán hàng của một siêu thị cho ở file data31.xls:

- b) Viết hàm **ci.mean(x, alpha)** xuất ra khoảng tin cậy cho kỳ vọng, với x là véc-tơ dữ liệu, (1-alpha) là độ tin cậy. Áp dụng để tìm khoảng tin cậy 95% và 99% cho doanh số bán hàng trung bình ở siêu thị.

```
# b)
# Viet ham ci.mean
ci.mean <- function(x,alpha){
  n = length(x)          # Co mau
  x.bar = mean(x)         # Trung binh mau
  s = sd(x)               # Do lech tieu chuan mau
  t = qt(1-alpha/2,n-1)  # Phan vi muc (1-alpha/2) cua T ~ t(n-
1)
  ep = t*s/sqrt(n)       # Sai so uoc luong
  cat('KTC',100*(1-alpha),'% cho ky vong mu la:\n')
  cat('[',x.bar - ep, ';' , x.bar + ep, ']\n')
}
```

4. Bài tập

```
# Bai 2
# a) Nhap du lieu tu file data31.csv vao R
dat31 <- read.csv("data31.csv",header = T)
attach(dat31); names(dat31)
# b)
# Viet ham ci.mean
ci.mean <- function(x,alpha){
  n = length(x)          # Co mau
  x.bar = mean(x)         # Trung binh mau
  s = sd(x)               # Do lech tieu chuan mau
  t = qt(1-alpha/2,n-1)   # Phan vi muc (1-alpha/2) cua T ~ t(n-
1)

  ep = t*s/sqrt(n)        # Sai so uoc luong
  cat('KTC',100*(1-alpha),'% cho ky vong mu la:\n')
  cat('[',x.bar - ep, ';' , x.bar + ep, ']\n')
}

# Ap dung
ci.mean(profit,0.05)
ci.mean(profit,0.01)
```

4. Bài tập

Bài 3: File data32.xls chứa số liệu về thời gian tự học của 120 sinh viên trường ĐH Khoa học Tự nhiên.

- a) Hãy ước lượng thời gian học nhóm trung bình của sinh viên trường ĐH KHTN, độ tin cậy là 95%. (Dùng hàm **ci.mean(x, alpha)**).

```
# Bai 3
# Nhap du lieu tu file data32.csv vao R
dat32 <- read.csv("data32.csv", header = T)
attach(dat32); names(dat32)

# a) Uoc luong khoang cho thoi gian tu hoc trung binh cua sv truong
KHTN
ci.mean(KHTN, 0.05)
```

4. Bài tập

Bài 3: File data32.xls chứa số liệu về thời gian tự học của 120 sinh viên trường ĐH Khoa học Tự nhiên.

- b) Viết hàm **ci.prop(f, n, alpha)** xuất ra khoảng tin cậy cho tỷ lệ, với n là cỡ mẫu; f : số các phần tử thỏa yêu cầu (với tỷ lệ p cần tìm); $(1-\alpha)$ là độ tin cậy. Áp dụng để tìm khoảng tin cậy 90%; 95% và 99% cho tỷ lệ sinh viên có thời gian tự học trên 5 giờ mỗi ngày.

```
# b)
# Viet ham ci.prop
ci.prop <- function(f,n,alpha){
  p.hat = f/n                                     #Ty le mau
  ep = qnorm(1 - alpha/2)*sqrt(p.hat*(1-p.hat)/n) #Sai so
  cat('KTC', 100*(1-alpha), '% cho ty le p la:\n')
  cat('[', p.hat - ep , ';', p.hat + ep , ']\n')}
```

4. Bài tập

```
# b)
# Viet ham ci.prop
ci.prop <- function(f,n,alpha){
  p.hat = f/n                                #Ty le mau
  ep = qnorm(1 - alpha/2)*sqrt(p.hat*(1-p.hat)/n) #Sai so
  cat('KTC', 100*(1-alpha), '% cho ty le p la:\n')
  cat('[', p.hat - ep , ';' , p.hat + ep , ']\n')
}

# Tinh f,n
n = length(KHTN)
f = length(KHTN[KHTN > 5])

# Ap dung
ci.prop(f,n,0.1)
ci.prop(f,n,0.05)
ci.prop(f,n,0.01)
```

```
# Bai 3
```

```
# Nhap du lieu tu file data32.xls vao R
```

```
dat32 <- read.csv("data32.csv", header = T)
```

```
attach(dat32); names(dat32)
```

```
# a) Uoc luong khoang cho thoi gian tu hoc trung binh cua sv truong  
KHTN
```

```
ci.mean(KHTN, 0.05)
```

```
# b)
```

```
# Viet ham ci.prop
```

```
ci.prop <- function(f,n,alpha){
```

```
  p.hat = f/n
```

```
#Ty le mau
```

```
  ep = qnorm(1 - alpha/2)*sqrt(p.hat*(1-p.hat)/n)
```

```
#Sai so
```

```
  cat('KTC', 100*(1-alpha), '% cho ty le p la:\n')
```

```
  cat('[', p.hat - ep , ';', p.hat + ep , ']\n')}
```

```
# Tinh f,n
```

```
n = length(KHTN)
```

```
f = length(KHTN[KHTN > 5])
```

```
# Ap dung
```

```
ci.prop(f,n,0.1)
```

```
ci.prop(f,n,0.05)
```

```
ci.prop(f,n,0.01)
```


4. Bài tập

Bài 4: Bảng sau thống kê chiều cao (Đơn vị: m) của 125 thanh niên 18 tuổi trong một khu vực:

Chiều cao	[1.2;1.4)	[1.4;1.6)	[1.6;1.8)	[1.8;2.0)	[2.0;2.2)
Số thanh niên	6	34	31	42	12

- a) Chuyển bảng tần số dạng khoảng ở trên thành dữ liệu dạng véc-tơ cột. Áp dụng hàm **ci.mean** đã ở bài 2 để tìm khoảng tin cậy 95% cho chiều cao trung bình của thanh niên trong khu vực.

4. Bài tập

Bài 4: Chuyển bảng tần số dạng khoảng ở trên thành dữ liệu dạng véc-tơ cột.

X	$[a_1; b_1)$	$[a_2; b_2)$...	$[a_k; b_k)$
n	n_1	n_2	...	n_k



X	$\frac{a_1 + b_1}{2}$	$\frac{a_2 + b_2}{2}$...	$\frac{a_k + b_k}{2}$
n	n_1	n_2	...	n_k

4. Bài tập

Bài 4: Bảng sau thống kê chiều cao (Đơn vị: m) của 125 thanh niên 18 tuổi trong một khu vực:

Chiều cao	[1.2;1.4)	[1.4;1.6)	[1.6;1.8)	[1.8;2.0)	[2.0;2.2)
Số thanh niên	6	34	31	42	12

```
# Bai 4
# a) Dưa dữ liệu bảng về dạng vectơ
a <- seq(1.2,2.0,by = 0.2)
b <- seq(1.4,2.2,by = 0.2)
m <- (a+b)/2
N <- c(6,34,31,42,12)
x <- rep(m,N);x
# Áp dụng ci.mean để tìm KTC 95% cho chiều cao trung bình của thanh
# niên ở trong khu vực
ci.mean(x,0.05)
```

4. Bài tập

Bài 4: Bảng sau thống kê chiều cao (Đv: m) của 125 thanh niên 18 tuổi trong một khu vực:

Chiều cao	[1.2;1.4)	[1.4;1.6)	[1.6;1.8)	[1.8;2.0)	[2.0;2.2)
Số thanh niên	6	34	31	42	12

- b) Những người có chiều cao từ 1.7 m trở lên được xếp vào sức khỏe loại A. Sử dụng hàm **ci.prop** ở bài 3 để tìm khoảng tin cậy 95% cho tỷ lệ thanh niên đạt sức khỏe loại A.

```
# b)
f = length(x[x>=1.7])
n = length(x)
ci.prop(f,n,0.05)
```

4. Bài tập

Bài 4:

```
# Bai 4
# a) Dua du lieu bang ve dang vecto
a <- seq(1.2,2.0,by = 0.2)
b <- seq(1.4,2.2,by = 0.2)
m <- (a+b)/2
N <- c(6,34,31,42,12)
x <- rep(m,N);x
# Ap dung ci.mean de tim KTC 95% cho chieu cao trung binh cua thanh
#nien o trong khu vuc
ci.mean(x,0.05)

# b)
f = length(x[x>=1.7])
n = length(x)
ci.prop(f,n,0.05)

# Xoa du lieu duoc gan
rm(list=ls(all=T))
```

4. Bài tập

Bài 5: Viết hàm *ktc.tb()* để tìm khoảng tin cậy cho trung bình biết:

- **Input:** là trung bình mẫu \bar{x} , độ lệch chuẩn của tổng thể σ (có thể biết trước hoặc không), trường hợp không biết thì phải nhập độ lệch chuẩn của mẫu s , kích thước mẫu n , và mức ý nghĩa α .
- **Output:** khoảng tin cậy cho trung bình.

4. Bài tập

#Bai 5.

Lưu ý: Giá trị độ lệch chuẩn sigma, $s \geq 0$

```
ktc.tb <- function(x.bar, s=-1, sigma=-1, n, alpha){  
  if(sigma >= 0){  
    ep = qnorm(1 - alpha/2)*sigma/sqrt(n)  
  }else{  
    if(n>30){  
      ep = qnorm(1 - alpha/2)*s/sqrt(n)  
    }else{  
      ep = qt(1 - alpha/2, n-1)*s/sqrt(n)  
    }  
  }  
  L <- x.bar - ep  
  U <- x.bar + ep  
  cat('Khoang tin cay',100*(1-alpha),'% cho ky vong mu la:\n')  
  round(c(lower = L , upper = U),2)  
}
```

#Bai 5.

Luu y: Gia tri do lech chuan sigma, $s \geq 0$

```
ktc.tb <- function(x.bar, s=-1, sigma=-1, n, alpha){  
  if(sigma >= 0){  
    ep = qnorm(1 - alpha/2)*sigma/sqrt(n)  
  }else{  
    if(n>30){  
      ep = qnorm(1 - alpha/2)*s/sqrt(n)  
    }else{  
      ep = qt(1 - alpha/2, n-1)*s/sqrt(n)  
    }  
  }  
  L <- x.bar - ep  
  U <- x.bar + ep  
  cat('Khoang tin cay',100*(1-alpha),'% cho ky vong mu la:\n')  
  round(c(lower = L , upper = U),2)  
}  
#Kiem tra  
n <- 35  
x <- rnorm(n,10,5)  
alpha<-0.05  
x.bar <- mean(x); s <- sd(x)  
ktc.tb(x.bar, s, n=n, alpha=alpha)
```


4. Bài tập

Bài 6: Từ hàm được viết trong câu 5) hãy viết hàm *ktc.tb.mau()* để tìm khoảng tin cậy cho trung bình biết:

- **Input:** vecto dữ liệu mẫu x , độ lệch chuẩn của tổng thể σ (có thể biết trước hoặc không) và mức ý nghĩa α .
- **Output:** khoảng tin cậy cho trung bình.

```
#Bai 6.
ktc.tb.mau<-function(x,alpha,sigma=-1){
  n<-length(x)
  x.bar<-mean(x); s<-sd(x)
  ktc.tb(x.bar=x.bar, s=s, sigma=sigma, n=n, alpha=alpha)
}
#Kiem tra
alpha<-0.05
ktc.tb.mau(x=x, alpha=alpha)
```

4. Bài tập

Bài 7: Đo đường kính của một chi tiết máy do một máy tiện tự động sản xuất, ta ghi nhận được số liệu như sau:

X	12.00	12.05	12.10	12.15	12.20	12.25	12.30	12.35	12.40
n	2	3	7	9	10	8	6	5	3

Bằng cách sử dụng hàm *ktc.tb.mau()* trong câu 6), hãy ước lượng khoảng tin cậy 95% cho đường kính trung bình.

```
#Bai 7.
```

```
y<-c(12.00, 12.05, 12.10, 12.15, 12.20, 12.25, 12.30, 12.35,12.40)
```

```
N<-c(2,3,7,9,10,8,6,5,3)
```

```
x <- rep(y,N); x
```

```
ktc.tb.mau(x=x, alpha=0.05)
```

```
rm(list=ls(all=T))
```