BÀI TẬP ƯỚC LƯỢNG + KIỂM ĐỊNH CHO MỘT TỔNG THỂ

Bài toán 1: Kiểm định, tìm KTC cho một trung bình, khi σ đã biết: sử dụng hàm `zsum.test` hoặc `z.test`

Chú ý: **+ Nếu dữ liệu cho ở dạng số liệu quan sát (dữ liệu sơ cấp): Sử dụng hàm** `z.test`

+ Dữ liệu dạng thứ cấp (đã qua xử lí): Sử dụng hàm`zsum.test`

Usage

```
zsum.test(mean.x, sigma.x = NULL, n.x = NULL, mean.y = NULL,
  sigma.y = NULL, n.y = NULL, alt="t", mu = 0,
  conf.level = 0.95)
```

trong đó:

alt="t" (two-side): kiểm định 2 phía và cho ước lượng khoảng

alt="g" (greater): kiểm định lớn hơn

alt="l" (less): kiểm định nhỏ hơn

Ví dụ 1(KĐ 2 phía, tìm KTC) :

a) Một nhà sản xuất khẳng định rằng khối lượng trung bình của dây câu có thể chịu là 8 kg, biết khối lượng dây có thể chịu có phân phối chuẩn với độ lệch chuẩn của tổng thểlà 0,5 kg. Để kiểm định giả thiết $\mu = 8$ kg với đối thiết $\mu \neq 8$ kg, 50 dây ngẫu nhiên được kiểm tra và khối lượng trung bình dây có thể chịu là 7,8 kg. Hãy kiểm định khẳng định của nhà sản xuất với mức ý nghĩa 0,01.

b)Cho giả thiết như câu a): khối lượng dây có thể chịu có phân phối chuẩn với độ lệch chuẩn của tổng thể là 0,5 kg. 50 dây ngẫu nhiên được kiểm tra và khối lượng trung bình dây có thể chịu là 7,8 kg.Tìm khoảng tin cậy 99% cho trung bình khối lượng dây của nhà sản xuất có thể chịu.

c) Kiểm tra lại p-value = 0.004678 trong ví dụ này bằng công thức $P-value = 2P(Z > |z|)$; với

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

HD câu c):

```
> pnorm(alpha,lower.tail = FALSE)=P(z>alpha)
> pnorm(alpha,lower.tail = TRUE)=P(z<alpha)
```

GIẢI:

a) Từ giả thiết ta có: $n = 50,\quad \bar{x} = 7,8,\quad \sigma = 0,5,\quad \mu_0 = 8$

Đây là bài toán kiểm định giả thiết cho một giá trị trung bình, khi σ đã biết

Đặt bài toán: $\begin{cases} H_0 : \mu = 8 \\ H_1 : \mu \neq 8 \end{cases}$.

+Gọi thư viện BSDA:

```
>library(BSDA)
Loading required package: lattice

Attaching package: ‘BSDA’

The following object is masked from ‘package:datasets’:

    Orange
```

+Sử dụng hàm zsum.test trong R

```
>zsum.test(mean.x=7.8,sigma.x=0.5, n.x=50, alt="t", mu=8,conf.level = 0.99)

        One-sample z-Test

data:  Summarized x
z = -2.8284, p-value = 0.004678
alternative hypothesis: true mean is not equal to 8
99 percent confidence interval:
 7.617861 7.982139
sample estimates:
mean of x
      7.8
```

+Có thể kết luận theo 3 cách

Cách 1:p-value = 0.004678< mứcý nghĩa alpha=0.01 nênbác bỏ giả thiết H_0

Cách 2:Chỉ tiêu kiểm định z = -2.8284 thuộc D là miền bác bỏ giả thiết H_0, nên bác bỏ giả thiết H_0. Trong đó

```
>z0.005=qnorm(0.005,lower.tail = FALSE)
>z0.005
[1] 2.575829
```

$$\Rightarrow D = \left(-\infty; -z_{\alpha/2}\right) \cup \left(z_{\alpha/2}; +\infty\right) = \left(-\infty; -2.576\right) \cup \left(2.576; +\infty\right)$$

(Ghi chú:

```
> qnorm(alpha,lower.tail = FALSE)=z thỏa mãn P(z>alpha)
> qnorm(alpha,lower.tail = TRUE)= z thỏa mãn P(z<alpha)
```

)

Cách 3: Khoảng tin cậy 99% cho trọng lượng trung bình là (7.617861 7.982139) không chứa 8 nên bác bỏ giả thiết H_0

Kết luận: Với mức ý nghĩa 0,01 ta bác bỏ giả thiết H_0 tức là trọng lượng trung bình dây có thể chịu là khác 8kg.

b)Từ kết quả phân tích câu (a), khoảng tin cậy 99% cho trung bình khối lượng dây của nhà sản xuất có thể chịu là

```
99 percent confidence interval:
 7.617861 7.982139
```

Hay (7.617861; 7.982139).

Hoặc ta làm như sau (bỏ đối thiết kiểm định mu=8):

```
>zsum.test(mean.x=7.8,sigma.x=0.5, n.x=50, alt="t",conf.level = 0.99)

        One-sample z-Test

data:  Summarized x
z = 110.31, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
99 percent confidence interval:
 7.617861 7.982139
sample estimates:
mean of x
      7.8
```

```
c)>z=(7.8-8)/(0.5/sqrt(50))
>p_value=2*pnorm(abs(z),lower.tail = F)
>p_value
[1] 0.004677735
Hoặc
>p_value=2*pnorm(abs(-2.8284),lower.tail = F)
>p_value
[1] 0.004678131
```

Ví dụ2(KĐ 1 phía) :Một nhà sản xuất khẳng định rằng khối lượng trung bình của dây câu có thể chịu là nhỏ hơn 8 kg, biết khối lượng dây có thể chịu có PP chuẩn với độ lệch chuẩn của tổng thể là 0,5 kg. Để kiểm định giả thuyết μ = 8 kg với đối thuyết μ <8 kg, 50 dây ngẫu nhiên được kiểm tra và khối lượng trung bình dây có thể chịu là 7,8 kg. Hãy kiểm định khẳng định của nhà sản xuất với mức ý nghĩa 0,01.

GIẢI:

+Từ giả thiết ta có: $n = 50, \quad \bar{x} = 7,8, \quad \sigma = 0,5, \quad \mu_0 = 8$

Đây là bài toán kiểm định giả thiết cho một giá trị trung bình, khi σ đã biết

Đặt bài toán: $\begin{cases} H_0 : \mu = 8 \\ H_1 : \mu < 8 \end{cases}$.

+Sử dụng hàm zsum.test trong R

```
> zsum.test(mean.x=7.8,sigma.x=0.5, n.x=50, alt="less", mu=8,conf.level = 0.99)

        One-sample z-Test

data:  Summarized x
z = -2.8284, p-value = 0.002339
alternative hypothesis: true mean is less than 8
99 percent confidence interval:
       NA 7.964498
sample estimates:
mean of x
      7.8
```

+Có thể kết luận theo 3 cách

Cách 1: `p-value = 0.002339< mứcý nghĩa alpha=0.01 nên` bác bỏ giả thiết H_0

Cách 2: Chỉ tiêu kiểm định `z = -2.8284` thuộc D nên bác bỏ giả thiết H_0

Cách 3: Các giá trị trong khoảng tin cậy 99% là: `NA 7.964498` nhỏ hơn 8 nên bác bỏ giả thiết H_0.

Kết luận: Với mức ý nghĩa 0,01 ta bác bỏ giả thuyết H_0 tức là có thể xem trọng lượng trung bình dây có thể chịu là nhỏ hơn 8kg.

***Ví dụ 3*(dữ liệu sơ cấp (hay dữ liệu quan sát)):** Một nhà sản xuất khẳng định rằng khối lượng trung bình của dây câu có thể chịu là 8 kg, biết khối lượng dây có thể chịu có PP chuẩn với độ lệch chuẩn của tổng thể là 0,5 kg. Để kiểm định giả thiết $\mu = 8$ kg với đối thiết $\mu \neq 8$ kg, 11 dây ngẫu nhiên được kiểm tra và cho số liệu như sau:7.8, 6.6, 6.5, 7.4, 7.3, 7.0, 6.4, 7.1, 6.7, 7.6, 6.8.

Hãy kiểm định khẳng định của nhà sản xuất với mức ý nghĩa 0,01.

GIẢI:

Đặt bài toán: $\begin{cases} H_0 : \mu = 8 \\ H_1 : \mu \neq 8 \end{cases}$.

```
> x=c(7.8, 6.6, 6.5, 7.4, 7.3, 7.0, 6.4, 7.1, 6.7, 7.6, 6.8)
>  zsum.test(mean(x),sigma.x=0.5, n.x=11, alt="t", mu=8,conf.level = 0.99)

        One-sample z-Test

data:  Summarized x
z = -6.5126, p-value = 7.384e-11
```

```
alternative hypothesis: true mean is not equal to 8
99 percent confidence interval:
 6.629861 7.406503
sample estimates:
mean of x
 7.018182
```

Kết luận: Với mức ý nghĩa 0,01 ta bác bỏ giả thiết H_0 tức là có thể xem trọng lượng trung bình dây có thể chịu là khác 8kg.

Nhận xét : VD này cho dữ liệu sơ cấp, ta cũng có thể dùng hàm z.test để giải quyết bài toán. Khi đó không cần nhập `mean(x), n.x`.

+Sử dụng hàm z.test trong R

```
> x=c(7.8, 6.6, 6.5, 7.4, 7.3, 7.0, 6.4, 7.1, 6.7, 7.6, 6.8)
> z.test(x,sigma.x=0.5, alt="t", mu=8,conf.level = 0.99)

        One-sample z-Test

data:  x
z = -6.5126, p-value = 7.384e-11
alternative hypothesis: true mean is not equal to 8
99 percent confidence interval:
 6.629861 7.406503
sample estimates:
mean of x
 7.018182
```

Kết luận: Với mức ý nghĩa 0,01 ta bác bỏ giả thiết H_0 tức là có thể xem trọng lượng trung bình dây có thể chịu là khác 8kg.

Bài toán 2: Kiểm định, tìm KTC về một trung bình, khi σ chưa biết, cỡ mẫu lớn (n ≥ 30):

Nếu dữ liệu cho ở dạng số liệu quan sát (dữ liệu sơ cấp): sử dụng hàm *`t.test`*, dữ liệu cho dạng thứ cấp(thu gọn): sử dụng hàm *`tsum.test`*.

Ví dụ 1: Chiều cao của một mẫu ngẫu nhiên 50 sinh viên TLU cho thấy có giá trị trung bình 174.5cm và độ lệch chuẩn 6.9cm.Xác định khoảng tin cậy 98% cho chiều cao trung bình của tất cả sinh viên TLU.

GIẢI:

```
>tsum.test(mean.x=174.5,s.x=6.9, n.x=50, alt="t",conf.level = 0.98)

       One-sample t-Test
```

```
data:  Summarized x
t = 178.83, df = 49, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
98 percent confidence interval:
 172.1533 176.8467
sample estimates:
mean of x
    174.5
```

Vậy khoảng tin cậy 98% cho chiều cao tổng thể của sinh viên toàn trường là: (172.1533; 176.8467)

Chú ý: + Vì cỡ mẫu lớn (*n* ≥ 30) nên ta có thể coi $s \approx \sigma$ và đưa về như trường hợp σ đã biết(Trường hợp 1), do đó ta sử dụng các hàm *zsum.test*, *z.test* thì kết quả gần như xấp xỉ. Ta ưu tiên sử dụng hàm *tsum.test, t.test hơn.*

```
>zsum.test(mean.x=174.5,sigma.x=6.9, n.x=50, alt="t",conf.level = 0.98)

        One-sample z-Test

data:  Summarized x
z = 178.83, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
98 percent confidence interval:
 172.2299 176.7701
sample estimates:
mean of x
    174.5
```

Vậy khoảng tin cậy 98% cho chiều cao tổng thể của sinh viên toàn trường là: (172.2299; 176.7701)

Câu hỏi

Dữ liệu ChiTieu2010.csv là mẫu điều tra ngẫu nhiên vài chục nghìn hộ gia đình ở nước ta. Từ đó hãy:

1. *Người ta cho rằng chi tiêu giáo dục trung bình một năm của các hộ gia đình nước ta vào thời điểm 2010 là ít nhất 250 (nghìn). Hãy kiểm định điều này với mức ý nghĩa 1%.*
2. *Kiểm định tại mức ý nghĩa 5% ý kiến cho rằng chi tiêu ăn uống trung bình một tháng của tổng thể các hộ gia đình sống ở khu vực thành thị (khu vực = 1) ở nước ta vào thời điểm đó là không vượt quá 1000 (nghìn).*

1)

#Goi mu la trung binh tong the

#Bai toan KĐGT cho 1 gia tri trung binh,phuong sai chua biet, n>30

#H0: mu>=250;   H1: mu<250

#(Chu y gia thiet H0 co dau "=")

```
>getwd()
[1] "C:/Users/Phuong/Documents"
#Copy file ChiTieu2010.csv vào thư mục R đang làm việc theo đường dẫn
> ChiTieu2010=read.csv("ChiTieu2010.csv")
>sd(ChiTieu2010$ChiTieuGiaoDucTrongNam)
[1] 720.7803
>x= ChiTieu2010$ChiTieuGiaoDucTrongNam
```

Hoặc dùng lệnh attach

```
>DL=read.csv("ChiTieu2010.csv")
>attach(DL)
>sd(ChiTieuGiaoDucTrongNam)
[1] 720.7803
>x=ChiTieuGiaoDucTrongNam
#Cach 1: Su dung ham t.test
>t.test(x,mu=250,alt="less",conf.level = 0.95)

        One Sample t-test
```

```
data:  x
t = -1.3258, df = 9397, p-value = 0.09247
alternative hypothesis: true mean is less than 250
95 percent confidence interval:
     -Inf 252.3736
sample estimates:
mean of x
 240.1428
```

p-value = 0.09246 >0.05 nên chấp nhận gt H0

2)

#Goi mu la trung binh tong the chi tieu an uong trung binh cac ho thanh thi

#Bai toan KĐGT cho 1 gia tri trung binh,phuong sai chua biet, n>30

#H0: mu>=1000;   H1: mu<1000

```
>y=CTAnUongTrongThang[KhuVuc==1]
>t.test(y,mu=1000,alt="less",conf.level = 0.95)

        One Sample t-test

data:  y
t = 49.939, df = 2646, p-value = 1
alternative hypothesis: true mean is less than 1000
95 percent confidence interval:
     -Inf 2915.493
sample estimates:
mean of x
 2854.393
```

Cho dữ liệu iris.csv. Tìm khoảng tin cậy 95% cho độ dài trung bình đài hoa

```
> getwd()
[1] "C:/Users/Phuong/Documents"
```

#Như vậy R đang làm việc với thư mục "C:/Users/Phuong/Documents"
#Copy file iris.csv vào thư mục có đường dẫn trên (mặc định là Documents)
> t.test(iris$Sepal.Length,conf.level = 0.95)

95 percent confidence interval:
5.709732 5.976934

Bài toán 3.Kiểm định, tìm KTC về một trung bình, khi σ chưa biết, cỡ mẫu nhỏ (n < 30): sử dụng hàmt.test hoặctsum.test (giả thiết phân phối là chuẩn)

Usage

```
tsum.test(mean.x, s.x = NULL, n.x = NULL, mean.y = NULL, s.y = NULL,
  n.y = NULL, alternative = "two.sided", mu = 0, var.equal = FALSE,
  conf.level = 0.95)
```

trong đó:

alt="t" : kiểm định 2 phía và cho KTC

alt="g" : kiểm định lớn hơn

alt="l" : kiểm định nhỏ hơn

Chú ý: + Nếu dữ liệu cho ở dạng số liệu quan sát (dữ liệu sơ cấp): sử dụng hàm `t.test`

+ Nếu dữ liệu cho ở dạng số liệu thu gọn (dữ liệu thứ cấp): sử dụng hàm `tsum.test`

Ví dụ 1: (dữ liệu thứ cấp)a)Một báo cáo khẳng định mỗi máy hút bụi tiêu thụ khoảng 46 kWh/1 năm. Từ một mẫu gồm 12 gia đình được nghiên cứu, cho thấy máy hút bụi tiêu thụ trung bình 42 kWh mỗi năm với độ lệch chuẩn mẫu 11,9 kWh. Liệu có thể nói, với mức ý nghĩa 0,05, trung bình máy hút bụi tiêu thụ không bằng 46 kWh mỗi năm hay không? Giả sử tổng thể đang xét có phân phối chuẩn.

b) Từ một mẫu gồm 12 gia đình được nghiên cứu, cho thấy máy hút bụi tiêu thụ trung bình 42 kWh mỗi năm với độ lệch chuẩn mẫu 11,9 kWh. Tìm khoảng tin cậy 95% cho trung bình tổng thể máy hút bụi tiêu thụ mỗi năm. Giả sử tổng thể đang xét có phân phối chuẩn.

c) Kiểm tra lại p-value = `0.2689` trong ví dụ này bằng công thức $P-value = 2P(t > |t_0|)$ với

$$t_0 = \frac{\overline{x} - \mu_0}{s/\sqrt{n}}$$

Và giá trị trong bảng tra A4

$$t_{\alpha/2;v=n-1} = t_{0.025;11} = 2.201$$

HD câu c):

```
> pt(alpha,so_bac_tu_do,lower.tail = FALSE)=P(t>alpha)
> pt(alpha,so_bac_tu_do,lower.tail = TRUE)=P(t<alpha)
> qt(alpha,so_bac_tu_do,lower.tail = FALSE)=t thỏa mãn P(t>alpha)
> qt(alpha,so_bac_tu_do,lower.tail = TRUE)= t thỏa mãn P(t<alpha)
```

GIẢI:

- Từ giả thiết:

$$n = 12, \quad \overline{x} = 42, \quad s = 11,9, \quad \alpha = 0,05, \quad \mu_0 = 46$$

- Đây là bài toán kiểm định một giá trị trung bình, khi σ chưa biết, $n < 30$.

Đặt bài toán $\begin{cases} H_0 : \mu = 46 \\ H_1 : \mu \neq 46 \end{cases}$.

```
> tsum.test(mean.x=42, s.x=11.9, n.x=12, mu=46, alternative="t",conf.level = 0.95)

        One-sample t-Test

data:  Summarized x
t = -1.1644, df = 11, p-value = 0.2689
alternative hypothesis: true mean is not equal to 46
95 percent confidence interval:
 34.4391 49.5609
sample estimates:
mean of x
       42
```

+Kết luận: Do p-value = 0.2689 > alpha=0.05 nên chấp nhận giả thiết H_0.

Vậy có thể khẳng định mỗi máy hút bụi tiêu thụ TB khoảng 46 kWh/1 năm.

b) Khoảng tin cậy 95% cho trung bình tổng thể là

```
95 percent confidence interval:
 34.4391 49.5609
Hay (34.4391; 49.5609)
```

c)

```
> 2*pt(abs(-1.1644),11,lower.tail = FALSE)
[1] 0.2688943
> qt(0.025,11,lower.tail = FALSE)
[1] 2.200985
```

Ví dụ 2(dữ liệu sơ cấp): Kiểm định giả thuyết rằng thể tích của các hộp đựng loại dầu nhờn nào đó là 10 lít, nếu từ mẫu ngẫu nhiên gồm 10 hộp ta có các thể tích là: 10,2 9,7 10,1 10,3 10,1 9,8 9,9 10,4 10,3 9,8. Sử dụng mức ý nghĩa 0,01 và giả sử phân phối của thể tích là chuẩn.

Giải:

- Đây là bài toán kiểm định một giá trị trung bình, khi σ chưa biết, $n < 30$.

Đặt bài toán $\begin{cases} H_0 : \mu = 10 \\ H_1 : \mu \neq 10 \end{cases}$

+Nhập dữ liệu và dùng hàm `tsum.test`

```
> x=c(10.2, 9.7, 10.1, 10.3, 10.1, 9.8, 9.9, 10.4, 10.3, 9.8)
> tsum.test(mean(x), sd(x), n.x=10, mu=10, alternative="t",conf.level = 0.99)
```

```
        One-sample t-Test

data:  Summarized x
t = 0.77174, df = 9, p-value = 0.46
alternative hypothesis: true mean is not equal to 10
99 percent confidence interval:
  9.807338 10.312662
sample estimates:
mean of x
    10.06
```

Vậy có thể xem thể tích của các hộp đựng loại dầu nhờn nào đó là 10 lít.

Nhận xét: Ví dụ này cho dữ liệu sơ cấp, ta có thể sử dụng hàm t.test để giải quyết bài toán:

```
> x=c(10.2, 9.7, 10.1, 10.3, 10.1, 9.8, 9.9, 10.4, 10.3, 9.8)
> t.test(x, mu=10, alternative="t",conf.level = 0.99)

        One Sample t-test

data:  x
t = 0.77174, df = 9, p-value = 0.46
alternative hypothesis: true mean is not equal to 10
99 percent confidence interval:
  9.807338 10.312662
sample estimates:
mean of x
    10.06
```

Vậy có thể xem thể tích của các hộp đựng loại dầu nhờn nào đó là 10 lít.

Ví dụ 3 (dữ liệu có tần số):Điều tra ngẫu nhiên lương tháng (triệu đồng) của giám đốc công ty địa ốc, ta có bảng số liệu sau:

Lương tháng	90	95	105	110	115	130
Số lượng	2	5	2	3	1	3

Tìm khoảng tin cậy 99% cho mức lương tháng trung bình giám đốc công ty địa ốc. Giả sử PP làxấp xỉ chuẩn.

GIẢI:

<u>Nhận xét</u>: VD này phải dùng lệnh `rep( )` ghép Luongthang và Tanso.

+ Bước 1: Nhập dữ liệu:

+ Bước 2: Dùng hàm `t.test` trong trường hợp này

```
> Luongthang=c(90,95, 105, 110, 115, 130)
> Tanso=c(2, 5, 2, 3, 1, 3)
> Dulieu=rep(Luongthang,Tanso)
> Dulieu
 [1]  90  90  95  95  95  95  95 105 105 110 110 110 115 130 130 130
> t.test(Dulieu, alternative="t",conf.level = 0.99)
```

```
        One Sample t-test

data:  Dulieu
t = 30.178, df = 15, p-value = 7.619e-15
alternative hypothesis: true mean is not equal to 0
99 percent confidence interval:
  95.8753 116.6247
sample estimates:
mean of x
   106.25
```

Kết luận: khoảng tin cậy 99% cho mức lương tháng trung bình là (`95.8753 116.6247`)

Bài toán 4: Kiểm định giả thiết và tìmKhoảng tin cậy cho 1 tỷ lệ, cỡ mẫu lớn:

Cách 1: Sử dụng hàm prop.test

Usage

```
+prop.test(x, n, p = NULL,
          alternative = c("two.sided", "less", "greater"),
          conf.level = 0.95, correct = TRUE)
```

<u>trong đó:</u>

alt="t" : kiểm định 2 phía và cho KTC

alt="g" : kiểm định lớn hơn

alt="l" : kiểm định nhỏ hơn

`correct:` tham số dạng logic chỉ xem có hay không sự điều chỉnh liên tục Yate, mặc định là`correct = TRUE.`

Nếu $n\hat{p} \geq 5$ và $n(1-\hat{p}) \geq 5 \Leftrightarrow 5 \leq x \leq n-5$ ko cần hiệu chỉnh liên tục, chọn `correct = FALSE.`

Cách 2: Do x có phân phối nhị thức nên hi tiêu kiểm định là: $Z = \dfrac{\hat{P} - p_0}{\sqrt{\dfrac{p_0 q_0}{n}}} = \dfrac{X - np_0}{\sqrt{np_0 q_0}}$

Như vậy, ta có thể làm cách khác là sử dụng hàm binom.test:

Usage

```
binom.test(x, n, p = 0.5,
           alternative = c("two.sided", "less", "greater"),
           conf.level = 0.95)
```

Ví dụ 1(KĐ 1 phía): Loại thuốc an thần cũ chỉ có tác động tới 60% người sử dụng. Kết quả thử nghiệm loại thuốc mới với 100 người thì thấy thuốc có tác dụng với 70 người. Có thể tin được hay không rằng loại thuốc mới tốt hơn loại thường dùng? Sử dụng mức ý nghĩa 0,05.

GIẢI

Cách 1:Gọi p là tỷ lệ người nhận được sự tác động của loại thuốc mới.

Từ giả thiết ta có: $n = \mathbf{100},\ p = \mathbf{70},\ p_0 = 0.6,\ \alpha = 0.05$.

Đây là bài toán kiểm định một phía cho một tỷ lệ với cỡ mẫu lớn.

Đặt bài toán: $\begin{cases} H_0 : p = 0,6 \\ H_1 : p > 0,6 \end{cases}$

Do $n\hat{p} = 70 \geq 5$ và $n(1-\hat{p}) = 30 \geq 5 \Leftrightarrow 5 \leq x \leq n-5$,chọn `correct = FALSE`

```
> prop.test(70, 100, p = 0.6, alternative = "greater",conf.level = 0.95, correct = F)


        1-sample proportions test without continuity correction


data:  70 out of 100, null probability 0.6
X-squared = 4.1667, df = 1, p-value = 0.02061
alternative hypothesis: true p is greater than 0.6
95 percent confidence interval:
 0.6201679 1.0000000
sample estimates:
  p
0.7
```

Kết luận: Do `p-value = 0.02061< 0.05`,nên bác bỏ H_0. Với mức ý nghĩa 0,05 ta có thể nói rằng loại thuốc mới tốt hơn loại thuốc thường dùng.

Cách 2:

Sử dụng hàm `binom.test`

```
>binom.test(70, 100, p = 0.6, alternative = "greater",conf.level = 0.95)
        Exact binomial test

data:  70 and 100
number of successes = 70, number of trials = 100, p-value = 0.02478
alternative hypothesis: true probability of success is greater than 0.6
95 percent confidence interval:
 0.6157794 1.0000000
sample estimates:
probability of success
                   0.7
```

Kết luận: Do `p-value = 0.02478< 0.05`, nên bác bỏ H_0. Với mức ý nghĩa 0,05 ta có thể nói rằng loại thuốc mới tốt hơn loại thuốc thường dùng.

Ví dụ 2(ƯL, KĐ 2 phía): Điều tra ngẫu nhiên 500 gia đình có tivi ở thành phố Hamilton, Canada, thấy rằng có 340 gia đình thuê bao chương trình HBO. Hãy tìm khoảng tin cậy 95% cho tỷ lệ gia đình thuê bao chương trình HBO trong số những gia đình có tivi trong thành phố này.

GIẢI

- Đặt p là tỷ lệ cần ước lượng.
- Từ giả thiết: ta có: $x = 340,\ n = 500,\ \alpha = 0.05$.
- Đây là bài toán ước lượng cho 1 tỷ lệ, cỡ mẫu lớn

Do $n\hat{p} = 340 \geq 5$ và $n(1-\hat{p}) = 160 \geq 5$, chọn `correct = FALSE`

-Chọn alt="t" cho ước lượng KTC.

```
> prop.test(340, 500, alternative = "t",conf.level = 0.95, correct = F)

        1-sample proportions test without continuity correction

data:  340 out of 500, null probability 0.5
X-squared = 64.8, df = 1, p-value = 8.29e-16
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.6378730 0.7193822
sample estimates:
   p
0.68
```

KL: 95 percent confidence interval:

0.6378730 0.7193822

Example

Có ý kiến cho rằng các sinh viên đại học thì có chỉ số IQ cao hơn mức bình thường 100 của toàn xã hội. Một nhóm sinh viên muốn kiểm định xem ý kiến nói trên có đúng với sinh viên TLU hay không. Họ tiến hành lấy mẫu hệ thống 25 sinh viên và test IQ của họ. Kết quả như sau:

126 105 120 104 120 99 91 87

116 119 103 117 115 115 143 138 109

108 126 127 126 107 70 105 121

Giả sử tổng thể chỉ số IQ của sinh viên TLU có phân bố chuẩn. Hãy kiểm định khẳng định nói trên tại mức ý nghĩa 5%

HD:

#Goi mu la trung binh tong the

#Bai toan KĐGT cho 1 gia tri trung binh,phuong sai chua biet, n<30

#Su dung ham t.test

Example

Một chủ hồ nuôi khi thương thảo bán lứa cá chim cho rằng lứa cá của họ có trọng lượng trung bình là 2.5kg. Công ty thu mua cá đã bắt ngẫu nhiên 20 con cá và cân nặng của chúng như sau (kg):

1.7 1.8 2.8 1.8 2.1 2.9 1.0 1.5 1.4 2.7 1.3 0.9 1.4 1.3 2.3 1.6 1.8 3.0 1.0 1.2

Giả sử rằng trọng lượng của tổng thể lứa cá có phân phối chuẩn với phương sai 0.36 (kg). Hỏi tại mức ý nghĩa 5% có thể chấp nhận khẳng định của chủ hồ nuôi không?

HD:

#Goi mu la trung binh tong the

#Bai toan KĐGT cho 1 gia tri trung binh,phuong sai da biet

#Su dung ham z.test

Example

Một hãng sản xuất ti vi công bố rằng 95% số sản phẩm của họ không phải sửa chữa trong 5 năm đầu sử dụng. Một một kê ngẫu nhiên 200 gia đình sử dụng ti vi của hãng này cho thấy có 18 gia đình nói rằng họ đã phải sửa ti vi trong vòng 5 năm sử dụng. Có thể bác bỏ khẳng định của hãng ti vi trên không?

HD:

#Goi p la ti le chung cua tong the

#Bai toan KĐGT cho 1 ti le, n lon

#H0: p=0.95,   H1:p # 0.95

```
>prop.test(200-18, 200, p = 0.95, alternative ="t", conf.level = 0.95, correct = FALSE)

        1-sample proportions test without continuity correction

data:  200 - 18 out of 200, null probability 0.95
X-squared = 6.7368, df = 1, p-value = 0.009444
alternative hypothesis: true p is not equal to 0.95
95 percent confidence interval:
 0.8622343 0.9423126
sample estimates:
   p
0.91
Do p-value = 0.009444 < 0.05 nên bác bỏ gt H0.
```

Cho dữ liệu hoa iris file iris.csv gồm độ dài đài hoa (Sepal.Length), độ rộng đài hoa(Sepal.Width), độ dài cánh hoa (Petal.Length), độ rộng cánh hoa(Petal.Width) chia thành 3 loài(Species).

Tìm khoảng tin cậy 95% cho tỷ lệ hoa có chiều dài đài hoa >5.4

```
> getwd()
[1] "C:/Users/Phuong/Documents"
```

#Như vậy R đang làm việc với thư mục "C:/Users/Phuong/Documents"
#Copy file iris.csv vào thư mục có đường dẫn trên (mặc định là Documents)

```
>x=sum(iris$Sepal.Length>5.4)
>x
>n=length(iris$Sepal.Length)
```

```
>n
>prop.test(x, n, conf.level = 0.95, correct = F)
```

95 percent confidence interval:

0.5742042 0.7248049

Kiểm định giả thuyết "tỷ lệ hoa có chiều dài đài hoa >5.4" là 0.6

$H_0 : p = 0.6$

$H_1 : p \neq 0.6$

```
x=sum(iris$Sepal.Length >5.4)
x
n=150
prop.test(x, n, p = 0.6, alternative = "two.sided", conf.level = 0.95,
correct = F)
```

Ta có p-value = $0.1824 > 0.05$ nên chấp nhận giả thuyết H_0. Vậy có thể xem tỷ lệ là 0.6.

Kiểm định giả thuyết "tỷ lệ hoa có chiều dài đài hoa >5.4" lớn hơn 0.55

$H_0 : p = 0.55$

$H_1 : p > 0.55$

```
x=sum(iris$Sepal.Length >5.4)
x
98
n=150
prop.test(x, n, p = 0.55, alternative = "greater", conf.level = 0.95,
correct = F)
```

Ta có p-value = $0.005481 < 0.05$ nên bác bỏ giả thuyết H_0. Vậy có thể xem tỉ lệ > 0.55.

Kiểm định giả thuyết

Kiểm định giả thiết chiều dài của đài hoa có phân phối chuẩn hay không với mức ý nghĩa 0.01
H_0: chiều dài của đài hoa có phân phối chuẩn
H_1: chiều dài của đài hoa không có phân phối chuẩn
Cách 1: Dùng hàm jarque.bera.test
library(tseries)
jarque.bera.test(iris$Sepal.Length)
Ta có p-value $= 0.1061 > 0.01$ nên chấp nhận giả thiết H_0. Vậy chiều dài của đài hoa có phân phối chuẩn với mức ý nghĩa 0.01.

```
>library(tseries)
>jarque.bera.test(iris$Sepal.Length)
```

Kiểm định giả thiết chiều dài của đài hoa có phân phối chuẩn hay không với mức ý nghĩa 0.01.
H_0: chiều dài của đài hoa có phân phối chuẩn
H_1: chiều dài của đài hoa không có phân phối chuẩn
Cách 2: Dùng hàm Shapiro-Wilk
shapiro.test(iris$Sepal.Length)
Ta có p-value $= 0.01018 > 0.01$ nên chấp nhận giả thiết H_0.
Vậy chiều dài của đài hoa có phân phối chuẩn với mức ý nghĩa 0.01.

```
>shapiro.test(iris$Sepal.Length)
```

Kiểm định giả thiết chiều dài của đài hoa có phân phối chuẩn hay không với mức ý nghĩa 0.01.
H_0: có phân phối chuẩn
H_1: không có phân phối chuẩn
Cách 3: Dùng hàm Kolmogorov-Smirnov test
ks.test(iris$Sepal.Length,'pnorm',mean(iris$Sepal.Length), sd(iris$Sepal.Length))
Ta có p-value = 0.1891 > 0.01 nên chấp nhận giả thiết H_0. Vậy chiều dài của đài hoa có phân phối chuẩn với mức ý nghĩa 0.01.

>ks.test(iris$Sepal.Length,"pnorm",mean(iris$Sepal.Length),sd(iris$Sepal.Length))

BÀI TẬP LUYỆN TẬP
3. ƯỚC LƯỢNG KHOẢNG TIN CẬY + KIỂM ĐỊNH CHO MỘT TỔNG THỂ

Bài toán 1: Kiểm định, tìm KTC về một trung bình, khi σ đã biết

Nếu dữ liệu cho ở dạng số liệu quan sát (dữ liệu sơ cấp): sử dụng hàm *z.test*, dữ liệu cho dạng thứ cấp(thu gọn): sử dụng hàm z*sum.test*.

trong đó:

alt="t" : kiểm định 2 phía và cho KTC

alt="g" : kiểm định lớn hơn

alt="l" : kiểm định nhỏ hơn

1.a) Một nhà sản xuất khẳng định rằng khối lượng trung bình của dây câu có thể chịu là 9 kg, biết khối lượng dây có thể chịu có phân phối chuẩn với độ lệch chuẩn của tổng thểlà 0,6 kg. Để kiểm định giả thiết $\mu = 9$ kg với đối thiết $\mu \neq 9$ kg, 60 dây ngẫu nhiên được kiểm tra và khối lượng trung bình dây có thể chịu là 7,8 kg. Hãy kiểm định khẳng định của nhà sản xuất với mức ý nghĩa 0,05.

b)Giả thiết khối lượng dây có thể chịu có phân phối chuẩn với độ lệch chuẩn của tổng thể là 0,6 kg. 60 dây ngẫu nhiên được kiểm tra và khối lượng trung bình dây có thể chịu là 8 kg.Tìm khoảng tin cậy 99% cho trung bình khối lượng dây của nhà sản xuất có thể chịu.

c) Kiểm tra lại p-value = ? trong ví dụ này bằng công thức $P-value = 2P(Z > |z|)$; với

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

2. Một nhà sản xuất khẳng định rằng khối lượng trung bình của dây câu có thể chịu là nhỏ hơn 9 kg, biết khối lượng dây có thể chịu có PP chuẩn với độ lệch chuẩn của tổng thể là 0,7 kg. Để kiểm định giả thuyết μ = 9 kg với đối thuyết μ<9 kg, 100 dây ngẫu nhiên được kiểm tra và khối lượng trung bình dây có thể chịu là 8,8 kg. Hãy kiểm định khẳng định của nhà sản xuất với mức ý nghĩa 0,01.

3 Một nhà sản xuất khẳng định rằng khối lượng trung bình của dây câu có thể chịu là 8.2 kg, biết khối lượng dây có thể chịu có PP chuẩn với độ lệch chuẩn của tổng thể là 0,5 kg. Để kiểm định giả thiết μ = 8.2 kg với đối thiết $\mu \neq$ 8.2 kg, 11 dây ngẫu nhiên được kiểm tra và cho số liệu như sau:7.9, 6.6, 6.5, 7.4, 7.5, 7.0, 6.4, 7.1, 6.7, 7.6, 6.9.

Hãy kiểm định khẳng định của nhà sản xuất với mức ý nghĩa 0,05.

Bài toán 2: Kiểm định, tìm KTC về một trung bình, khi σ chưa biết, cỡ mẫu lớn (n ≥ 30):

Nếu dữ liệu cho ở dạng số liệu quan sát (dữ liệu sơ cấp): sử dụng hàm `t.test`, dữ liệu cho dạng thứ cấp(thu gọn): sử dụng hàm `tsum.test`.

trong đó:

alt="t" : kiểm định 2 phía và cho KTC

alt="g" : kiểm định lớn hơn

alt="l" : kiểm định nhỏ hơn

1. Chiều cao của một mẫu ngẫu nhiên 60 sinh viên TLU cho thấy có giá trị trung bình 175.5cm và độ lệch chuẩn 6.8 cm.Xác định khoảng tin cậy 96% cho chiều cao trung bình của tất cả sinh viên TLU.

Câu hỏi

Dữ liệu ChiTieu2010.csv là mẫu điều tra ngẫu nhiên vài chục nghìn hộ gia đình ở nước ta. Từ đó hãy:

1. *Người ta cho rằng chi tiêu giáo dục trung bình một năm của các hộ gia đình nước ta vào thời điểm 2010 là ít nhất 250 (nghìn). Hãy kiểm định điều này với mức ý nghĩa 1%.*
2. *Kiểm định tại mức ý nghĩa 5% ý kiến cho rằng chi tiêu ăn uống trung bình một tháng của tổng thể các hộ gia đình sống ở khu vực thành thị (khu vực = 1) ở nước ta vào thời điểm đó là không vượt quá 1000 (nghìn).*

Bài toán 3.Kiểm định, tìm KTC về một trung bình, khi σ chưa biết, cỡ mẫu nhỏ (n < 30) (giả thiết phân phối là chuẩn)

Nếu dữ liệu cho ở dạng số liệu quan sát (dữ liệu sơ cấp): sử dụng hàm *t.test*, dữ liệu cho dạng thứ cấp(thu gọn): sử dụng hàm *tsum.test*.

trong đó:

alt="t" : kiểm định 2 phía và cho KTC

alt="g" : kiểm định lớn hơn

alt="l" : kiểm định nhỏ hơn

1 a)Một báo cáo khẳng định mỗi máy hút bụi tiêu thụ khoảng 47 kWh/1 năm. Từ một mẫu gồm 15 gia đình được nghiên cứu, cho thấy máy hút bụi tiêu thụ trung bình 48 kWh mỗi năm với độ lệch chuẩn mẫu 11,8 kWh. Liệu có thể nói, với mức ý nghĩa 0,04, trung bình máy hút bụi tiêu thụ không bằng 47 kWh mỗi năm hay không? Giả sử tổng thể đang xét có phân phối chuẩn.

b) Từ một mẫu gồm 17 gia đình được nghiên cứu, cho thấy máy hút bụi tiêu thụ trung bình 46 kWh mỗi năm với độ lệch chuẩn mẫu 11,5 kWh. Tìm khoảng tin cậy 98% cho trung bình tổng thể máy hút bụi tiêu thụ mỗi năm. Giả sử tổng thể đang xét có phân phối chuẩn.

c) Kiểm tra lại p-value = ? trong ví dụ này bằng công thức $P-value = 2P(t > |t_0|)$ với

$$t_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

Và giá trị trong bảng tra A4

$$t_{\alpha/2;v=n-1} = ?$$

2(dữ liệu sơ cấp):Kiểm định giả thuyết rằng thể tích của các hộp đựng loại dầu nhờn nào đó là 10 lít, nếu từ mẫu ngẫu nhiên gồm 10 hộp ta có các thể tích là: 10,5 9,8 10,1 10,3 10,1 9,8 9,9 10,4 10,3 9,9. Sử dụng mức ý nghĩa 0,02 và giả sử phân phối của thể tích là chuẩn.

3 (dữ liệu có tần số):Điều tra ngẫu nhiên lương tháng (triệu đồng) của giám đốc công ty địa ốc, ta có bảng số liệu sau:

Lương tháng	96	95	105	110	115	120
Số lượng	3	5	2	3	1	4

Tìm khoảng tin cậy 95% cho mức lương tháng trung bình giám đốc công ty địa ốc. Giả sử PP làxấp xỉ chuẩn.

Bài toán 4: Kiểm định giả thiết và tìmKhoảng tin cậy cho 1 tỷ lệ, cỡ mẫu lớn:

Sử dụng hàm prop.test

Usage

```
+prop.test(x, n, p = NULL,
          alternative = c("two.sided", "less", "greater"),
          conf.level = 0.95, correct = TRUE)
```

trong đó:

alt="t" : kiểm định 2 phía và cho KTC

alt="g" : kiểm định lớn hơn

alt="l" : kiểm định nhỏ hơn

`correct`: tham số dạng logic chỉ xem có hay không sự điều chỉnh liên tục Yate, mặc định là`correct = TRUE`.

Nếu $n\hat{p} \geq 5$ và $n(1-\hat{p}) \geq 5 \Leftrightarrow 5 \leq x \leq n-5$ ko cần hiệu chỉnh liên tục, chọn `correct = FALSE`.

1(KĐ 1 phía): Loại thuốc an thần cũ chỉ có tác động tới 70% người sử dụng. Kết quả thử nghiệm loại thuốc mới với 105 người thì thấy thuốc có tác dụng với 80 người. Có thể tin được hay không rằng loại thuốc mới tốt hơn loại thường dùng? Sử dụng mức ý nghĩa 0,05.

2(ƯL, KĐ 2 phía): Điều tra ngẫu nhiên 600 gia đình có tivi ở thành phố Hamilton, Canada, thấy rằng có 350 gia đình thuê bao chương trình HBO. Hãy tìm khoảng tin cậy 96% cho tỷ lệ gia đình thuê bao chương trình HBO trong số những gia đình có tivi trong thành phố này.

BÀI TẬP LÀM THÊM

Khoảng tin cậy:

1. Một công ty điện sản xuất các bóng đèn có tuổi thọ tuân theo phân phối chuẩn với độ lệch chuẩn 40 giờ. Nếu một mẫu 30 bóng có tuổi thọ trung bình là 780 giờ, hãy xác định khoảng tin cậy 96% đối với kỳ vọng tổng thể của tất cả các bóng điện do công ty này sản xuất .

2. Chiều cao của một mẫu ngẫu nhiên 50 sinh viên đại học cho thấy có giá trị trung bình 174.5cm và độ lệch chuẩn 6.9cm.

Xác định khoảng tin cậy 98% cho chiều cao trung bình của tất cả sinh viên đó;

3. Một máy sản xuất các mảnh kim loại có hình trụ. Một mẫu các mảnh được lấy ra với các đường kính là 1.01, 0.97, 1.03, 1.04, 0.99, 0.98, 0.99, 1.01 , 1.03cm. Xác định khoảng tin cậy 99% đối với đường kính trung bình của các mảnh được sản xuất ra, giả thiết đường kính có phân phối xấp xỉ chuẩn.

4. Một mẫu ngẫu nhiên 12 chốt nghiền được lấy trong một nghiên cứu về độ cứng Rockwell của đầu trên chốt. Các lần đo được tiến hành lần lượt cho 12 chốt, cho giá trị trung bình 48.50 với độ lệch chuẩn mẫu 1.5. Giả thiết các giá trị đo có phân phối chuẩn, xác định một khoảng tin cậy 90% cho độ cứng Rockwell trung bình.

5. Một mẫu ngẫu nhiên 200 cử tri được lựa chọn và 114 được xác định ủng hộ một ứng viên. Xác định khoảng tin cậy 96% cho tỷ lệ cử tri ủng hộ cho ứng viên đó.

6.Trong một mẫu ngẫu nhiên 1000 hộ gia đình trong 1 thành phố người ta thấy có228 người được cấp dầu để đun. Xác định khoảng tin cậy 99% cho tỷ lệ số gia đình trong thành phố được cấp dầu đun.

7. Theo một thông báo trên Roanoke Times & Word – News, 2/3 trong số 1600 thanh niên được phỏng vấn qua điện thoại cho biết họ cho rằng chương trình tàu con thoi không gian là một khoản đầu tư tốt của Chính phủ. Xác định khoảng tin cậy 95% cho tỷ lệ thanh niên Mỹ nghĩ rằng chương trình này là 1 cách đầu tư tốt của Chính phủ.

Kiểm định giả thiết

1. Một mẫu ngẫu nhiên gồm 100 giấy báo tử ở Mỹ cho thấy tuổi thọ trung bình là 71,8 năm. Giả sử độ lệch chuẩn là 8,9 năm; có thể cho rằng tuổi thọ trung bình hiện nay là hơn 70 năm không? Cho mức ý nghĩa là 0,05.

2. Một hãng sản suất bóng đèn, có tuổi thọ trung bình của bóng là xấp xỉ phân phối chuẩn với kỳ vọng 800 giờ và độ lệch chuẩn 40 giờ. Kiểm định giả thuyết $\mu = 800$ giờ với đối thuyết $\mu \neq 800$ giờ nếu một mẫu ngẫu nhiên gồm 30 bóng có tuổi thọ trung bình là 778 giờ. Mức ý nghĩa 0,04.

3.Chiều cao trung bình của nữ sinh năm thứ nhất tại một trường cao đẳng là 162,5 cm và độ lệch chuẩn 6,9 cm. Có thể tin được hay không rằng có sự thay đổi độ cao trung bình nếu mẫu ngẫu nhiên gồm 50 nữ sinh có chiều cao trung bình 165,2 cm? Cho mức ý nghĩa là 0,01.

4.Kiểm định giả thuyết rằng thể tích của các hộp đựng loại dầu nhờn nào đó là 10 lít, nếu từ mẫu ngẫu nhiên gồm 15 hộp ta có các thể tích là:

9,5 10,2 9,7 10,1 10,3 10,1 9,8 9,9 10,4 10,3 9,8 9,4 10,4 10,6 10,8 9,7.

Sử dụng mức ý nghĩa 0,01 và giả sử phân phối của thể tích là chuẩn.

5.Một chuyên gia phân tích khẳng định thời gian để học sinh phổ thông làm một bài kiểm tra đã chuẩn hóa, là một biến ngẫu nhiên có phân phối chuẩn với kỳ vọng không quá 35 phút. Qua mẫu gồm 20 học sinh, người ta thấy thời gian trung bình để các em hoàn thành bài thi là 33,1 phút với độ lệch 4,3 phút. Với mức ý nghĩa 0,05 hãy kiểm định xem khẳng định đó có cơ sở không?

6. Một chuyên gia marketing của công ty sản xuất mì ống tin rằng, 40% người thích mì ống hơn lasagna (một loại món ăn). Qua phỏng vấn 200 người, thì có 90 người thích mì ống hơn. Có thể kết luận gì về khẳng định của chuyên gia, với mức ý nghĩa 0,05?

7. Giả sử trước đây, có 40% người trưởng thành ủng hộ án tử hình. Có thể tin được hay không, rằng tỷ lệ người ủng hộ án tử hình ngày nay đã tăng lên, nếu trong mẫu ngẫu nhiên gồm 35 người thì có 18 người đồng ý? Sử dụng mức ý nghĩa 0,05.

8. Một công ty xăng dầu khẳng định 1/5 số nhà trong thành phố nào đó được sưởi bằng dầu. Có thể nghi ngờ khẳng định này không, nếu trong một mẫu ngẫu nhiên gồm 1000 ngôi nhà, thì có 136 ngôi nhà được sưởi bằng dầu? Dùng mức ý nghĩa 0,01.

9. Tại một trường cao đẳng nào đó, người ta ước tính rằng nhiều hơn 25% sinh viên tới trường bằng xe đạp. Điều này có hợp lý không, nếu trong mẫu gồm 90 sinh viên có 28 bạn tới trường bằng xe đạp? Sử dụng mức ý nghĩa 0,05.