Thực hành xác suất thống kê

Bài 9:KIỂM ĐỊNH GIẢ THUYẾT THỐNG KÊ (2 MẪU)

1. So sánh kỳ vọng giữa hai tổng thể độc lập

Bài toán: Giả sử $(X_1, \dots, X_n)$ (mẫu 1) là mẫu ngẫu nhiên được chọn từ một tổng thể có phân phối chuẩn với kỳ vọng μ_X và phương sai σ_X^2; $(Y_1, \dots, Y_m)$ (mẫu 2) là mẫu ngẫu nhiên được chọn từ một tổng thể có phân phối chuẩn với kỳ vọng μ_Y và phương sai σ_Y^2. Trong thực hành, ta giả sử σ_X^2 và σ_Y^2 không biết.

Các giả thuyết:

$\begin{cases} \boldsymbol{H_0}: \boldsymbol{\mu_X = \mu_Y} \\ \boldsymbol{H_1}: \boldsymbol{\mu_X \neq \mu_Y} \end{cases}$	$\begin{cases} \boldsymbol{H_0}: \boldsymbol{\mu_X \geq \mu_Y} \\ \boldsymbol{H_1}: \boldsymbol{\mu_X < \mu_Y} \end{cases}$	$\begin{cases} \boldsymbol{H_0}: \boldsymbol{\mu_X \leq \mu_Y} \\ \boldsymbol{H_1}: \boldsymbol{\mu_X > \mu_Y} \end{cases}$
Hai phía (‘two.sided’)	Một phía, bên trái (‘less’)	Một phía, bên phải (‘greater’)

1. So sánh kỳ vọng giữa hai tổng thể độc lập

Sử dụng hàm t.test để kiểm định

```
t.test(x,y, alternative = “đối thuyết”, var.equal = FALSE, conf.level = 0.95)
```

Trong đó

`x,y`: là véc-tơ dữ liệu tương ứng với hai mẫu.

`alternative`: là đối thuyết gồm hai phía (“`two.sided`”), bên trái (“`less`”), bên phải (“`greater`”), mặc định là “`two.sided`”.

`var.equal`: phương sai có bằng nhau không, mặc định không (FALSE).

`conf.level`: Độ tin cậy (= 1 - α) cho sự sai khác giữa hai kỳ vọng, tức $\mu_X - \mu_Y$, mặc định là 0.95

Bài tập

Bài 1: Hai máy rót sữa tự động được sử dụng để đưa sữa vào hộp giấy có dung tích 1 lít trong một dây chuyền sản xuất. Lượng sữa thực tế hai máy đưa vào hộp có phân phối chuẩn với độ lệch tiêu chuẩn lần lượt là 0.002 và 0.0025 lít. Một thành viên trong số các kỹ sư giám sát dây chuyền cho rằng thể tích sữa trung bình được hai máy đưa vào các hộp là như nhau và không phụ thuộc vào dung tích của các hộp chứa. Một mẫu ngẫu nhiên được lấy từ hai máy này được cho trong file *volume.csv.*

Bài tập

```
# Bai 1
# Doc file "volume.csv" vao R
dat1 <- read.csv("volume.csv", header = T)
attach(dat1); names(dat1)
dat1
    X  machine1  machine2
1   1 0.9986924 0.9968177
2   2 0.9993856 0.9981757
3   3 0.9999283 0.9992946
4   4 0.9983961 0.9982450
5   5 1.0005144 1.0011562
...

# Gan gia tri cho x,y:
x <- dat1[,2]
y <- dat1[,3]
```

Bài tập

Bài 1: Hai máy rót sữa tự động được sử dụng để đưa sữa vào hộp giấy có dung tích 1 lít trong một dây chuyền sản xuất. Lượng sữa thực tế hai máy đưa vào hộp có phân phối chuẩn với độ lệch tiêu chuẩn lần lượt là 0.002 và 0.0025 lít. Một thành viên trong số các kỹ sư giám sát dây chuyền cho rằng thể tích sữa trung bình được hai máy đưa vào các hộp là như nhau và không phụ thuộc vào dung tích của các hộp chứa. Một mẫu ngẫu nhiên được lấy từ hai máy này được cho trong file *volume.csv*.

a) Bạn có cho rằng phán đoán của kỹ sư trên là đúng hay không? Sử dụng $\alpha = 0.05$.

Phát biểu giả thuyết: $\begin{cases} \boldsymbol{H_0}: \boldsymbol{\mu_X = \mu_Y} \\ \boldsymbol{H_1}: \boldsymbol{\mu_X \neq \mu_Y} \end{cases}$

Kiểm định hai phía.

Bài tập

```
t.test(x,y,alternative = "two.sided", var.equal = F, conf.level =
0.95)

        Welch Two Sample t-test

data:  x and y
t = -1.48, df = 87.191, p-value = 0.1425
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.0014222030  0.0002081663
sample estimates:
mean of x mean of y
0.9993200 0.9999271
```

Bài tập

a) Bạn có cho rằng phán đoán của kỹ sư trên là đúng hay không? Sử dụng $\alpha = 0.05$.

Phát biểu giả thuyết: $\begin{cases} \boldsymbol{H_0}: \boldsymbol{\mu_X = \mu_Y} \\ \boldsymbol{H_1}: \boldsymbol{\mu_X \neq \mu_Y} \end{cases}$

p - giá trị = 0.1425 > 0.05 = 5% = alpha.

⇒Chưa đủ cơ sở để bác bỏ $H_0: \mu_X = \mu_Y$.

Với mức ý nghĩa 5%, thể tích sữa trung bình được hai máy đưa vào các hộp là như nhau, hay phán đoán của kỹ sư trên là đúng.

Bài tập

Bài 1: Hai máy rót sữa tự động được sử dụng để đưa sữa vào hộp giấy có dung tích 1 lít trong một dây chuyền sản xuất. Lượng sữa thực tế hai máy đưa vào hộp có phân phối chuẩn với độ lệch tiêu chuẩn lần lượt là 0.002 và 0.0025 lít. Một thành viên trong số các kỹ sư giám sát dây chuyền cho rằng thể tích sữa trung bình được hai máy đưa vào các hộp là như nhau và không phụ thuộc vào dung tích của các hộp chứa. Một mẫu ngẫu nhiên được lấy từ hai máy này được cho trong file *volume.csv.*

b) P-giá trị của kiểm định trên là bao nhiêu?

Bài tập

```
t.test(x,y,alternative = "two.sided", var.equal = F, conf.level = 
0.95)

        Welch Two Sample t-test

data:  x and y
t = -1.48, df = 87.191, p-value = 0.1425
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.0014222030  0.0002081663
sample estimates:
mean of x mean of y 
0.9993200 0.9999271 
```

Bài tập

b) P-giá trị của kiểm định trên là bao nhiêu?

p – giá trị = 0.1425

Bài tập

c) Hãy tìm khoảng tin cậy 95% cho sự khác biệt về trung bình lượng sữa hai máy đưa vào các hộp.

```
t.test(x,y,alternative = "two.sided", var.equal = F, conf.level =
0.95)

        Welch Two Sample t-test

data:  x and y
t = -1.48, df = 87.191, p-value = 0.1425
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.0014222030  0.0002081663
sample estimates:
mean of x mean of y
0.9993200 0.9999271
```

Bài tập

c) Khoảng tin cậy 95% cho sự khác biệt về trung bình lượng sữa hai máy đưa vào các hộp.

$$-0.0014222030 \leq \mu_X - \mu_Y \leq 0.0002081663$$

Bài tập

d) Viết hàm **test.leq.oneside(x, y, μ_0, $\boldsymbol{\sigma_1}$, $\boldsymbol{\sigma_2}$,$\boldsymbol{\alpha}$)** để kiểm định giả thiết $\boldsymbol{H_0}$: $\boldsymbol{\mu} = \boldsymbol{\mu_0}$và đối thuyết $\boldsymbol{H_1}$: $\boldsymbol{\mu} < \boldsymbol{\mu_0}$ trong đó $\boldsymbol{\mu} = \boldsymbol{\mu_1} - \boldsymbol{\mu_2}$, với $\boldsymbol{\mu_1}$, $\boldsymbol{\mu_2}$ lần lượt là thể tích sữa trung bình hai máy 1 và 2 đưa vào các hộp; **x, y** là hai vector dữ liệu chứa thể tích sữa do máy số 1 và máy số 2 đưa vào các hộp giấy; $\boldsymbol{\alpha}$ là mức ý nghĩa. Hàm xuất ra kết quả chấp nhận hoặc bác bỏ và cho biết p – giá trị.

```
# d)
# Viet ham test.leq.oneside
test.leq.oneside <- function(x,y,mu0,sig1,sig2,alpha){
            n = length(x)
            m = length(y)
            x.bar = mean(x)
            y.bar = mean(y)
            z0 = (x.bar - y.bar - mu0)/(sqrt(sig1^2/n + sig2^2/m))
            p.value = pnorm(z0)
            cat('Voi muc y nghia alpha =',alpha*100,'% :\n')
            if(p.value < alpha)
               cat('Bac bo H0 voi p-gia tri = ', p.value, '\n')
            else
            cat('Chua du co so de bac bo H0 voi p-gia tri = ',
p.value, '\n')
}

# Ap dung H1: mu > mu0 <=> mu1 - mu2 > 0 <=> mu1 > mu2
test.leq.oneside(x,y,0,0.002,0.0025,0.05)

# Kiem tra lai bang t.test
t.test(x,y,alternative = "less", var.equal = F, conf.level = 0.95)
```

Bài tập

e) Viết hàm **test.geq.oneside(x, y, μ_0, σ_1, σ_2,α)** để kiểm định giả thiết $H_0: \mu = \mu_0$ và đối thuyết $H_1: \mu > \mu_0$ trong đó $\mu = \mu_1 - \mu_2$, với μ_1, μ_2 lần lượt là thể tích sữa trung bình hai máy 1 và 2 đưa vào các hộp; **x, y** là hai vector dữ liệu chứa thể tích sữa do máy số 1 và máy số 2 đưa vào các hộp giấy; α là mức ý nghĩa. Hàm xuất ra kết quả chấp nhận hoặc bác bỏ và cho biết p – giá trị.

```
# e)
# Viet ham test.geq.oneside
test.geq.oneside <- function(x,y,mu0,sig1,sig2,alpha){
            n = length(x)
            m = length(y)
            x.bar = mean(x)
            y.bar = mean(y)
            z0 = (x.bar - y.bar - mu0)/(sqrt(sig1^2/n + sig2^2/m))
            p.value = 1 - pnorm(z0)
            cat('Voi muc y nghia alpha =',alpha*100,'% :\n')
            if(p.value < alpha)
               cat('Bac bo H0 voi p-gia tri = ', p.value, '\n')
            else
            cat('Chua du co so de bac bo H0 voi p-gia tri = ',
p.value, '\n')
}

# Ap dung H1: mu > mu0 <=> mu1 - mu2 > 0 <=> mu1 > mu2
test.geq.oneside(x,y,0,0.002,0.0025,0.05)

# Kiem tra lai bang t.test
t.test(x,y,alternative = "greater", var.equal = F, conf.level =
0.95)
```

2. So sánh theo cặp

Bài toán: Giả sử $(X_1, \dots, X_n)$, $(Y_1, \dots, Y_n)$ là hai mẫu ngẫu nhiên cùng cỡ được chọn từ cùng một tổng thể có phân phối chuẩn với kỳ vọng lần lượt là μ_X và μ_Y; như vậy, sẽ có một mối liên hệ giữa các giá trị X_i và Y_i. Do đó, các cặp giá trị $(X_i, Y_i), i = 1, \dots, n$ sẽ không độc lập. Khi đó, ta sử dụng phương pháp kiểm định theo cặp để kiểm tra các giả thuyết liên quan đến μ_X và μ_Y.

2. So sánh theo cặp

Với (X_1, Y_1), (X_2, Y_2),..., (X_n, Y_n) là n cặp quan trắc, đặt

$$D_i = X_i - Y_i, i = 1, \dots, n$$

Biến ngẫu nhiên D_i giả sử có phân phối chuẩn với kỳ vọng là

$$\mu_D = E(D_i) = E(X_i - Y_i) = E(X_i) - E(Y_i) = \mu_X - \mu_Y$$

Với phương sai σ_D^2.

Các giả thuyết

$\begin{cases} \boldsymbol{H_0}: \boldsymbol{\mu_D} = \mathbf{0} \\ \boldsymbol{H_1}: \boldsymbol{\mu_D} \neq \mathbf{0} \end{cases}$	$\begin{cases} \boldsymbol{H_0}: \boldsymbol{\mu_D} \geq \mathbf{0} \\ \boldsymbol{H_1}: \boldsymbol{\mu_D} < \mathbf{0} \end{cases}$	$\begin{cases} \boldsymbol{H_0}: \boldsymbol{\mu_D} \leq \mathbf{0} \\ \boldsymbol{H_1}: \boldsymbol{\mu_D} > \mathbf{0} \end{cases}$
Hai phía (‘two.sided’)	Một phía, bên trái (‘less’)	Một phía, bên phải (‘greater’)

2. So sánh theo cặp

Thống kê kiểm định

$$T_0 = \frac{\overline{D}}{S_D/\sqrt{n}} \sim t(n-1)$$

Trong đó

$$\overline{D} = \frac{1}{n}\sum_{i=1}^{n} D_i \,; S_D^2 = \frac{1}{n-1}\sum_{i=1}^{n}(D_i - \overline{D})^2$$

Miền bác bỏ - p-giá trị:

Giả thuyết H_0	Đối thuyết H_1	Miền bác bỏ với mức ý nghĩa α	P – giá trị
$\mu_D = 0$	$\mu_D \neq 0$	$\lvert T_0 \rvert > t_{1-\alpha/2}^{n-1}$	$2P\{T_{n-1} \geq \lvert T_0 \rvert\}$
$\mu_D \leq 0$	$\mu_D > 0$	$T_0 > t_{1-\alpha}^{n-1}$	$P\{T_{n-1} \geq T_0\}$
$\mu_D \geq 0$	$\mu_D < 0$	$T_0 < -t_{1-\alpha}^{n-1}$	$P\{T_{n-1} \leq T_0\}$

2. So sánh theo cặp

Sử dụng hàm t.test để kiểm định

```
t.test(x,y, alternative = “đối thuyết”, var.equal = FALSE, conf.level = 0.95,
paired = TRUE)
```

Trong đó

`x,y`: là véc-tơ dữ liệu tương ứng với hai mẫu.

`alternative`: là đối thuyết gồm hai phía (“`two.sided`”), bên trái (“`less`”), bên phải (“`greater`”), mặc định là “`two.sided`”.

`var.equal`: phương sai có bằng nhau không, mặc định không (`FALSE`).

`conf.level`: Độ tin cậy (= 1 - α) cho sự sai khác giữa hai kỳ vọng, tức $\mu_X - \mu_Y$, mặc định là 0.95

`paired = TRUE`: so sánh theo cặp.

Bài tập

Bài 4: Hai mươi người nam với độ tuổi từ 35 đến 50 tham gia vào một nghiên cứu để đánh giá sự ảnh hưởng của chế độ ăn uống và luyện tập thể thao lên hàm lượng cholesterol trong máu. Mỗi người tham gia được đo lượng cholesterol trước khi bắt đầu chương trình tập luyện thể dục và chuyển sang ăn uống với chế độ ít chất béo. Dữ liệu được cho trong file *cholesterol.txt*.

```
cholesterol = read.table('cholesterol.txt',header=T)
cholesterol
attach(cholesterol)
   Subject Before After
1        1    265   229
2        2    240   231
3        3    258   227
4        4    295   240
...
```

Bài tập

Bài 4: Hai mươi người nam với độ tuổi từ 35 đến 50 tham gia vào một nghiên cứu để đánh giá sự ảnh hưởng của chế độ ăn uống và luyện tập thể thao lên hàm lượng cholesterol trong máu. Mỗi người tham gia được đo lượng cholesterol trước khi bắt đầu chương trình tập luyện thể dục và chuyển sang ăn uống với chế độ ít chất béo. Dữ liệu được cho trong file *cholesterol.txt*.

a) Hỏi rằng dữ liệu có ủng hộ kết luận rằng chế độ ăn kiêng và luyện tập thể dục đã có tác dụng trong việc giảm lượng cholesterol trong máu hay không? Hãy tiến hành ở mức ý nghĩa 0.05.

Bài tập

```
cholesterol = read.table('cholesterol.txt',header=T)
cholesterol
attach(cholesterol)
# Kiem dinh phuong sai:
var.test(Before,After)

         F test to compare two variances

data:  Before and After
F = 5.6691, num df = 14, denom df = 14, p-value = 0.002529
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
  1.903292 16.885974
sample estimates:
ratio of variances
          5.669121
```

Bài tập

```
# Lay ket qua tu Kiem dinh phuong sai
v.equal <- ifelse(var.test(Before,After)$p.value < alpha, FALSE, 
TRUE); v.equal
[1] FALSE

# Kiem dinh bang t.test
t.test(Before,After,alternative = "greater", var.equal = v.equal, 
conf.level = 0.95,paired = TRUE)

           Paired t-test

data:  Before and After
t = 5.4659, df = 14, p-value = 4.158e-05
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 18.20922      Inf
sample estimates:
mean of the differences 
           26.86667
```

Bài tập

p – giá trị = `4.158e-05` $<$ 0.05 = alpha

⇒ Bác bỏ $H_0: \mu_1 - \mu_2 = 0$.

Với mức ý nghĩa 5%, chế độ ăn kiêng và tập thể dục có hiệu quả trong việc giảm hàm lượng cholesterol trong máu những người béo phì tham gia thử nghiệm.

Bài tập

b) Giả sử hàm lượng cholesterol ở mỗi người tham gia nghiên cứu trước và sau khi bắt đầu chương trình luyện tập thể dục thể thao có phân phối chuẩn với phương sai khác nhau. Viết hàm **test.leq.oneside(x, y, $\boldsymbol{\mu_0}$,alpha)** để kiểm định giả thiết $H_0: \mu = \mu_0$ và đối thiết $H_1: \mu < \mu_0$ trong đó $\mu = \mu_1 - \mu_2$ với $\mu_1; \mu_2$ lần lượt là trung bình hàm lượng cholesterol trước và sau khi tham gia chương trình luyện tập, **alpha** là mức ý nghĩa. Hàm xuất ra kết quả chấp nhận hoặc bác bỏ và cho biết p – giá trị.

```
# Cau b)
# Viet ham test.leq.oneside
test.leq.oneside <- function(x,y,mu0,alpha){
         D = x - y
         n = length(D)
         D.bar = mean(D)
         D.sd = sd(D)
         t0 = (D.bar-mu0)*sqrt(n)/D.sd
         p.value = pt(t0, n-1)
         cat('Voi muc y nghia',100*alpha,'% \n')
          if(p.value < alpha)
               cat('Bac bo H0 voi p-gia tri =', p.value,'\n')
          else
               cat('Chua du co so de bac bo H0 voi p-gia tri =',
p.value,'\n')
}
x = Before; y = After
# Ap dung mu0 = 0
test.leq.oneside(x,y, mu0 = 0, alpha = 0.05)

# Kiem tra lai bang t.test
t.test(x,y,alternative = "less", var.equal = v.equal, conf.level =
0.95,paired = T)
```

Bài tập

c) Giả sử hàm lượng cholesterol ở mỗi người tham gia nghiên cứu trước và sau khi bắt đầu chương trình luyện tập thể dục thể thao có phân phối chuẩn với phương sai khác nhau. Viết hàm **test.geq.oneside(x, y, $\boldsymbol{\mu_0}$,alpha)** để kiểm định giả thiết $H_0: \mu = \mu_0$ và đối thiết $H_1: \mu > \mu_0$ trong đó $\mu = \mu_1 - \mu_2$ với $\mu_1; \mu_2$ lần lượt là trung bình hàm lượng cholesterol trước và sau khi tham gia chương trình luyện tập, **alpha** là mức ý nghĩa. Hàm xuất ra kết quả chấp nhận hoặc bác bỏ và cho biết p – giá trị.

```
# Cau c)
# Viet ham test.geq.oneside
test.leq.oneside <- function(x,y,mu0,alpha){
         D = x - y
         n = length(D)
         D.bar = mean(D)
         D.sd = sd(D)
         t0 = D.bar*sqrt(n)/D.sd
         p.value = 1 - pt(t0, n-1)
         cat('Voi muc y nghia',100*alpha,'% \n')
          if(p.value < alpha)
               cat('Bac bo H0 voi p-gia tri =', p.value,'\n')
          else
               cat('Chua du co so de bac bo H0 voi p-gia tri =',
p.value,'\n')
}
x = Before; y = After
# Ap dung mu0 = 0
test.geq.oneside(x,y, mu0 = 0, alpha = 0.05)

# Kiem tra lai bang t.test
t.test(x,y,alternative = "greater", var.equal = v.equal, conf.level
= 0.95,paired = T)
```

3. So sánh hai tỷ lệ

Bài toán: Gọi p_1 và p_2 lần lượt là tỷ lệ của các phần tử thỏa tính chất quan tâm tương ứng với hai tổng thể. Từ hai tổng thể chọn ra hai mẫu độc lập với cỡ mẫu lần lượt là n_1 và n_2 (n_1 và n_2 lớn)

Các giả thuyết

$\begin{cases} H_0: \mathrm{p}_1 = p_2 \\ H_1: p_1 \neq p_2 \end{cases}$	$\begin{cases} H_0: \mathrm{p}_1 \geq p_2 \\ H_1: p_1 < p_2 \end{cases}$	$\begin{cases} H_0: \mathrm{p}_1 \leq p_2 \\ H_1: p_1 > p_2 \end{cases}$
Hai phía (‘two.sided’)	Một phía, bên trái (‘less’)	Một phía, bên phải (‘greater’)

3. So sánh hai tỷ lệ

Gọi Y_1 và Y_2 lần lượt là số phần tử thỏa tính chất quan tâm trong hai mẫu độc lập được chọn từ hai tổng thể.

Thống kê kiểm định

$$Z_0 = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

trong đó

$$\hat{p}_1 = \frac{Y_1}{n_1}; \hat{p}_2 = \frac{Y_2}{n_2}; \hat{p} = \frac{Y_1 + Y_2}{n_1 + n_2}$$

Miền bác bỏ - p-giá trị:

Giả thuyết H_0	Đối thuyết H_1	Miền bác bỏ với mức ý nghĩa α	P – giá trị
$p_1 = p_2$	$p_1 \neq p_2$	$\lvert Z_0 \rvert > z_{1-\alpha/2}$	$2P\{Z \geq \lvert Z_0 \rvert\}$
$p_1 \leq p_2$	$p_1 > p_2$	$Z_0 > z_{1-\alpha}$	$P\{Z \geq Z_0\}$

3. So sánh hai tỷ lệ

Sử dụng hàm prop.test để kiểm định

```
prop.test(y,n, p = p0, alternative = "đối thuyết", conf.level = 0.95)
```

Trong đó

$y = (y_1; y_2)$: y_i là số phần tử thỏa tính chất A trong n_i phần tử khảo sát.

$n = (n_1; n_2)$: n_i là cỡ mẫu.

`alternative`: là đối thuyết gồm hai phía ("`two.sided`"), bên trái ("`less`"), bên phải ("`greater`"), mặc định là "`two.sided`".

`p0`: giá trị cần kiểm định.

`conf.level`: Độ tin cậy (= 1 - α), mặc định là 0.95

Bài tập

Bài 7: Hai loại máy đúc khác nhau được sử dụng để gia công các chi tiết bằng plastic. Một chi tiết được xem là phế phẩm nếu nó có sự thay đổi về hình dạng và khối lượng hoặc sai biệt về màu sắc so với chi tiết mẫu vượt quá mức cho phép. Hai mẫu ngẫu nhiên được thu thập, mỗi mẫu có cỡ 300. Trong mẫu thu được từ máy số 1 người ta nhận thấy có 15 chi tiết là phế phẩm và đối với mẫu thu được từ máy thứ 2 có 8 chi tiết là phế phẩm. Liệu có cơ sở để kết luận rằng hai loại máy có tỉ lệ phế phẩm như nhau hay không (mức ý nghĩa 0.025)? Tìm p – giá trị cho kiểm định vừa rồi.

Gọi

p_1 là tỷ lệ phế phẩm của máy đúc thứ nhất.
p_2 là tỷ lệ phế phẩm của máy đúc thứ hai.

Giả thuyết cần kiểm định: $\begin{cases} \boldsymbol{H_0}: \mathbf{p_1} = \boldsymbol{p_2} \\ \boldsymbol{H_1}: \boldsymbol{p_1} \neq \boldsymbol{p_2} \end{cases}$

Bài tập

```
# Bai 7
y = c(15,8)
n = c(300,300)
prop.test(y,n, alternative = "two.sided", conf.level = 0.975)

  2-sample test for equality of proportions with continuity correction

data:  y out of n
X-squared = 1.6276, df = 1, p-value = 0.202
alternative hypothesis: two.sided
97.5 percent confidence interval:
 -0.01507289  0.06173956
sample estimates:
    prop 1     prop 2
0.05000000 0.02666667
```

Bài tập

p – giá trị = 0.202 > 0.025 = alpha

⇒ Chưa đủ cơ sở để bác bỏ $\boldsymbol{H_0}: \mathbf{p_1} = \boldsymbol{p_2}$

Với mức ý nghĩa 2,5%, hai loại máy có tỉ lệ phế phẩm như nhau.