

BÀI THỰC HÀNH SỐ 6

A. CÁC LOẠI SAI LẦM TRONG KIỂM ĐỊNH THỐNG KÊ VÀ P_VALUE

Sai lầm loại I: Nếu ta bác bỏ H_0 khi H_0 đúng thì sai lầm đó gọi là sai lầm loại I, ký hiệu α . α còn được gọi là *mức ý nghĩa* (significance level) của kiểm định.

Sai lầm loại II: Nếu H_0 sai mà ta không bác bỏ H_0 thì sai lầm đó gọi là sai lầm loại II, ký hiệu β .

Quyết định	H_0 đúng	H_0 sai
Không bác bỏ H_0	Đúng	Sai lầm loại II
Bác bỏ H_0	Sai lầm loại I	Đúng

P value (P_giá trị): Với một giả thuyết không H_0 và mẫu cỡ n cho trước, P – giá trị là mức ý nghĩa nhỏ nhất dẫn đến việc bác bỏ giả thuyết H_0 , P – giá trị được tính dựa theo giá trị thống kê kiểm định.

- Quy tắc:** Khi P_{value} bé hơn α ta bác bỏ giả thuyết H_0

B. KIỂM ĐỊNH GIẢ THUYẾT THỐNG KÊ CHO 1 MẪU

I. Kiểm định giả thuyết cho kỳ vọng

(1) Trường hợp 1: Biết phương sai σ^2

B1: Phát biểu giả thuyết không và đối thuyết

B2: Xác định mức ý nghĩa α

B3: Lấy mẫu ngẫu nhiên cỡ n : X_1, \dots, X_n và tính thống kê kiểm định

$$z_0 = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$$

B4: Xác định miền bác bỏ W_α

Giả thuyết	Miền bác bỏ	P-giá trị
$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}$	$W_\alpha = \{z_0: z_0 > z_{\alpha/2}\}$	$P_{value} = 2[1 - \Phi(z_0)]$
$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu < \mu_0 \end{cases}$	$W_\alpha = \{z_0: z_0 < -z_\alpha\}$	$P_{value} = \Phi(z_0)$
$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$	$W_\alpha = \{z_0: z_0 > z_\alpha\}$	$P_{value} = 1 - \Phi(z_0)$

B5: Kết luận: Bác bỏ H_0 / Chưa đủ cơ sở để bác bỏ H_0

(2) Trường hợp 2: Không biết phương sai σ^2 và mẫu lớn ($n > 30$)

B1: Phát biểu giả thuyết không và đối thuyết

B2: Xác định mức ý nghĩa α

B3: Tính ước lượng $\hat{\sigma} = s$

B4: Lấy mẫu ngẫu nhiên cỡ $n: X_1, \dots, X_n$ và tính thống kê kiểm định

$$z_0 = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$$

B5: Xác định miền bác bỏ W_α

Giả thuyết	Miền bác bỏ	P-giá trị
$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}$	$W_\alpha = \{z_0: z_0 > z_{\alpha/2}\}$	$P_{Value} = 2[1 - \Phi(z_0)]$
$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu < \mu_0 \end{cases}$	$W_\alpha = \{z_0: z_0 < -z_\alpha\}$	$P_{Value} = \Phi(z_0)$
$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$	$W_\alpha = \{z_0: z_0 > z_\alpha\}$	$P_{Value} = 1 - \Phi(z_0)$

B6: Kết luận: Bác bỏ H_0 / Chưa đủ cơ sở để bác bỏ H_0

(3) Trường hợp 3: Không biết phương sai σ^2 và mẫu nhỏ ($n \leq 30$)

B1: Phát biểu giả thuyết không và đối thuyết

B2: Xác định mức ý nghĩa α

B3: Tính ước lượng $\hat{\sigma} = s$

B4: Lấy mẫu ngẫu nhiên cỡ $n: X_1, \dots, X_n$ và tính thống kê kiểm định

$$t_0 = \frac{\bar{X} - \mu_0}{s / \sqrt{n}} \sim t(n-1)$$

B5: Xác định miền bác bỏ W_α

Giả thuyết	Miền bác bỏ	P-giá trị
$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}$	$W_\alpha = \{t_0: t_0 > t_{\alpha/2}^{n-1}\}$	$P_{Value} = 2\mathbb{P}(T_{n-1} \geq t_0)$
$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu < \mu_0 \end{cases}$	$W_\alpha = \{t_0: t_0 < -t_\alpha^{n-1}\}$	$P_{Value} = \mathbb{P}(T_{n-1} \leq t_0)$
$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$	$W_\alpha = \{t_0: t_0 > t_\alpha^{n-1}\}$	$P_{Value} = \mathbb{P}(T_{n-1} \geq t_0)$

B6: Kết luận: Bác bỏ H_0 / Chưa đủ cơ sở để bác bỏ H_0

II. Kiểm định giả thuyết cho tỉ lệ

Điều kiện: Cỡ mẫu n lớn và $np_0 \geq 5$ và $n(1 - p_0) \geq 5$

B1: Phát biểu giả thuyết không và đối thuyết

B2: Xác định mức ý nghĩa α

B3: Lấy mẫu ngẫu nhiên cỡ $n: X_1, \dots, X_n$ và tính thống kê kiểm định

$$z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

$$\text{với } \hat{p} = \frac{\text{số lần xuất hiện biến cố } Y}{n}$$

B4: Xác định miền bác bỏ W_α

Giả thuyết	Miền bác bỏ	P-giá trị
$\begin{cases} H_0: p = p_0 \\ H_1: p \neq p_0 \end{cases}$	$W_\alpha = \{z_0: z_0 > z_{\alpha/2}\}$	$P_{value} = 2[1 - \Phi(z_0)]$
$\begin{cases} H_0: p = p_0 \\ H_1: p < p_0 \end{cases}$	$W_\alpha = \{z_0: z_0 < -z_\alpha\}$	$P_{value} = \Phi(z_0)$
$\begin{cases} H_0: p = p_0 \\ H_1: p > p_0 \end{cases}$	$W_\alpha = \{z_0: z_0 > z_\alpha\}$	$P_{value} = 1 - \Phi(z_0)$

B5: Kết luận: Bác bỏ H_0 / Chưa đủ cơ sở để bác bỏ H_0

C. KIỂM ĐỊNH GIẢ THUYẾT THỐNG KÊ CHO 2 MẪU

I. So sánh hai kỳ vọng

(1) Trường hợp 1: Biết phương sai σ^2

B1: Phát biểu giả thuyết không và đối thuyết

B2: Xác định mức ý nghĩa α

B3: Tính thống kê kiểm định

$$z_0 = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}$$

B4: Xác định miền bác bỏ W_α

Giả thuyết	Miền bác bỏ	P-giá trị
$\begin{cases} H_0: \mu_1 - \mu_2 = D_0 \\ H_1: \mu_1 - \mu_2 \neq D_0 \end{cases}$	$W_\alpha = \{z_0: z_0 > z_{\alpha/2}\}$	$P_{value} = 2[1 - \Phi(z_0)]$
$\begin{cases} H_0: \mu_1 - \mu_2 = D_0 \\ H_1: \mu_1 - \mu_2 < D_0 \end{cases}$	$W_\alpha = \{z_0: z_0 < -z_\alpha\}$	$P_{value} = \Phi(z_0)$
$\begin{cases} H_0: \mu_1 - \mu_2 = D_0 \\ H_1: \mu_1 - \mu_2 > D_0 \end{cases}$	$W_\alpha = \{z_0: z_0 > z_\alpha\}$	$P_{value} = 1 - \Phi(z_0)$

B5: Kết luận: Bác bỏ H_0 / Chưa đủ cơ sở để bác bỏ H_0

(2) Trường hợp 2: Không biết phương sai σ^2 và mẫu lớn ($n > 30$)

B1: Phát biểu giả thuyết không và đối thuyết

B2: Xác định mức ý nghĩa α

B3: Tính ước lượng $\widehat{\sigma_1^2} = s_1^2$ và $\widehat{\sigma_2^2} = s_2^2$

B4: Tính thống kê kiểm định

$$z_0 = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{m}}}$$

B5: Xác định miền bác bỏ W_α

Giả thuyết	Miền bác bỏ	P-giá trị
------------	-------------	-----------

$\begin{cases} H_0: \mu_1 - \mu_2 = D_0 \\ H_1: \mu_1 - \mu_2 \neq D_0 \end{cases}$	$W_\alpha = \{z_0: z_0 > z_{\alpha/2}\}$	$P_{Value} = 2[1 - \Phi(z_0)]$
$\begin{cases} H_0: \mu_1 - \mu_2 = D_0 \\ H_1: \mu_1 - \mu_2 < D_0 \end{cases}$	$W_\alpha = \{z_0: z_0 < -z_\alpha\}$	$P_{Value} = \Phi(z_0)$
$\begin{cases} H_0: \mu_1 - \mu_2 = D_0 \\ H_1: \mu_1 - \mu_2 > D_0 \end{cases}$	$W_\alpha = \{z_0: z_0 > z_\alpha\}$	$P_{Value} = 1 - \Phi(z_0)$

B6: Kết luận: Bác bỏ H_0 / Chưa đủ cơ sở để bác bỏ H_0

(3) Trường hợp 3: Không biết phương sai σ^2 và mẫu nhỏ ($n \leq 30$)

So sánh hai phương sai

B1: Phát biểu giả thuyết không và đối thuyết

B2: Xác định mức ý nghĩa α

B3: Tính ước lượng $\widehat{\sigma_1^2} = s_1^2$ và $\widehat{\sigma_2^2} = s_2^2$

B4: Tính thống kê kiểm định

$$F_0 = \frac{s_1^2}{s_2^2} \sim \mathcal{F}(n-1, m-1)$$

B5: Xác định miền bác bỏ:

Giả thuyết	Miền bác bỏ
$\begin{cases} H_0: \sigma_1^2 = \sigma_2^2 \\ H_1: \sigma_1^2 \neq \sigma_2^2 \end{cases}$	$W_\alpha = \{F_0: F_0 > f_{\alpha/2, n_1-1, n_2-1} \text{ hoặc } F_0 < f_{(1-\alpha/2), n_1-1, n_2-1}\}$
$\begin{cases} H_0: \sigma_1^2 = \sigma_2^2 \\ H_1: \sigma_1^2 > \sigma_2^2 \end{cases}$	$W_\alpha = \{F_0: F_0 > f_{\alpha, n_1-1, n_2-1}\}$
$\begin{cases} H_0: \sigma_1^2 = \sigma_2^2 \\ H_1: \sigma_1^2 < \sigma_2^2 \end{cases}$	$W_\alpha = \{F_0: F_0 > f_{\alpha, n_1-1, n_2-1}\}$

B6: Kết luận: Bác bỏ H_0 / Chưa đủ cơ sở để bác bỏ H_0

(3.1) Xét trường hợp phương sai bằng nhau $\sigma_1^2 = \sigma_2^2$

B1: Phát biểu giả thuyết không và đối thuyết

B2: Xác định mức ý nghĩa α

B3: Tính ước lượng $\widehat{\sigma_1^2} = s_1^2$ và $\widehat{\sigma_2^2} = s_2^2$

B4: Tính phương sai mẫu chung

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

B5: Tính thống kê kiểm định

$$t_0 = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

có phân phối Student với bậc tự do $df = n_1 + n_2 - 2$

B6: Xác định miền bác bỏ W_α

Giả thuyết	Miền bác bỏ	P-giá trị
------------	-------------	-----------

$\begin{cases} H_0: \mu_1 - \mu_2 = D_0 \\ H_1: \mu_1 - \mu_2 \neq D_0 \end{cases}$	$W_\alpha = \{t_0: t_0 > t_{\alpha/2}^{df}\}$	$P_{Value} = 2\mathbb{P}(T_{df} \geq t_0)$
$\begin{cases} H_0: \mu_1 - \mu_2 = D_0 \\ H_1: \mu_1 - \mu_2 < D_0 \end{cases}$	$W_\alpha = \{t_0: t_0 < -t_\alpha^{df}\}$	$P_{Value} = \mathbb{P}(T_{df} \leq t_0)$
$\begin{cases} H_0: \mu_1 - \mu_2 = D_0 \\ H_1: \mu_1 - \mu_2 > D_0 \end{cases}$	$W_\alpha = \{t_0: t_0 > t_\alpha^{df}\}$	$P_{Value} = \mathbb{P}(T_{df} \geq t_0)$

B7: Kết luận: Bác bỏ H_0 / Chưa đủ cơ sở để bác bỏ H_0

(3.2) Xét trường hợp phương sai khác nhau $\sigma_1^2 \neq \sigma_2^2$

B1: Phát biểu giả thuyết không và đối thuyết

B2: Xác định mức ý nghĩa α

B3: Tính ước lượng $\widehat{\sigma_1^2} = s_1^2$ và $\widehat{\sigma_2^2} = s_2^2$

B4: Tính thống kê kiểm định

$$t_0 = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

có phân phối Student với bậc tự do

$$df = \frac{\left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right]^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$$

B5: Xác định miền bác bỏ W_α

Giả thuyết	Miền bác bỏ	P-giá trị
$\begin{cases} H_0: \mu_1 - \mu_2 = D_0 \\ H_1: \mu_1 - \mu_2 \neq D_0 \end{cases}$	$W_\alpha = \{t_0: t_0 > t_{\alpha/2}^{df}\}$	$P_{Value} = 2\mathbb{P}(T_{df} \geq t_0)$
$\begin{cases} H_0: \mu_1 - \mu_2 = D_0 \\ H_1: \mu_1 - \mu_2 < D_0 \end{cases}$	$W_\alpha = \{t_0: t_0 < -t_\alpha^{df}\}$	$P_{Value} = \mathbb{P}(T_{df} \leq t_0)$
$\begin{cases} H_0: \mu_1 - \mu_2 = D_0 \\ H_1: \mu_1 - \mu_2 > D_0 \end{cases}$	$W_\alpha = \{t_0: t_0 > t_\alpha^{df}\}$	$P_{Value} = \mathbb{P}(T_{df} \geq t_0)$

B6: Kết luận: Bác bỏ H_0 / Chưa đủ cơ sở để bác bỏ H_0

II. So sánh hai tỉ lệ

B1: Phát biểu giả thuyết không và đối thuyết

B2: Xác định mức ý nghĩa α

B3: Tính thống kê kiểm định

$$z_0 = \frac{\widehat{p}_1 - \widehat{p}_2 - (\mu_1 - \mu_2)}{\sqrt{\widehat{p}(1-\widehat{p})\left(\frac{1}{n} + \frac{1}{m}\right)}}$$

với

$$\widehat{p}_1 = \frac{X}{n_1}; \quad \widehat{p}_2 = \frac{Y}{n_2}; \quad \widehat{p} = \frac{X+Y}{n_1+n_2}$$

B4: Xác định miền bác bỏ W_α

Giả thuyết	Miền bác bỏ	P-giá trị
$\begin{cases} H_0: \mu_1 - \mu_2 = D_0 \\ H_1: \mu_1 - \mu_2 \neq D_0 \end{cases}$	$W_\alpha = \{z_0: z_0 > z_{\alpha/2}\}$	$P_{value} = 2[1 - \Phi(z_0)]$
$\begin{cases} H_0: \mu_1 - \mu_2 = D_0 \\ H_1: \mu_1 - \mu_2 < D_0 \end{cases}$	$W_\alpha = \{z_0: z_0 < -z_\alpha\}$	$P_{value} = \Phi(z_0)$
$\begin{cases} H_0: \mu_1 - \mu_2 = D_0 \\ H_1: \mu_1 - \mu_2 > D_0 \end{cases}$	$W_\alpha = \{z_0: z_0 > z_\alpha\}$	$P_{value} = 1 - \Phi(z_0)$

B5: Kết luận: Bác bỏ H_0 / Chưa đủ cơ sở để bác bỏ H_0

- Chú ý các lệnh trong R

```

z $\alpha$ /2: qnorm(1-alpha/2)
z $\alpha$ : qnorm(1-alpha)
t $\alpha$ /2n-1: qt(1-alpha/2, df=n-1)
t $\alpha$ n-1: qt(1-alpha, df=n-1)
 $\mathbb{P}(T_{n-1} \leq t_0)$ : pt(t0, n-1)
 $\mathbb{P}(T_{n-1} \geq t_0)$ : pt(t0, n-1, lower.tail=FALSE)

```