

Chương 5. Kiểm định giả thuyết thống kê

Mục lục

1	Giả thuyết cơ bản, giả thuyết tuyệt đối	1
2	Thủ tục kiểm định giả thuyết thống kê	1
2.1	Trường hợp đã biết phương sai	1
2.2	Trường hợp chưa biết phương sai, cỡ mẫu $n < 30$	3
2.3	Trường hợp chưa biết phương sai, cỡ mẫu $n > 30$	4
2.4	Kiểm định giả thuyết về tỷ lệ hay xác suất	4
3	So sánh hai kỳ vọng của hai biến ngẫu nhiên phân phối chuẩn	5
3.1	Trường hợp đã biết phương sai	5
3.2	Trường hợp chưa biết phương sai $n_1 < 30, n_2 < 30$	6
3.3	Trường hợp chưa biết phương sai $n_1 > 30, n_2 > 30$	6
3.4	So sánh hai tỷ lệ	7

1 Giả thuyết cơ bản, giả thuyết tuyệt đối

Giả sử cần nghiên cứu tham số θ của biến ngẫu nhiên X và có cơ sở viết lên giả thuyết $\theta = \theta_0$, kí hiệu giả thuyết là H_0 , được gọi là giả thuyết kiểm định. Khi đó, ta có đối thuyết H_1 được gọi là mệnh đề đối lập với giả thuyết H_0 , kí hiệu là H_1 .

- Các cặp giả thuyết thường dùng:

Giả thuyết H_0	$\theta = \theta_0$	$\theta = \theta_0$	$\theta = \theta_0$
Đối thuyết H_1	$\theta \neq \theta_0$	$\theta > \theta_0$	$\theta < \theta_0$

2 Thủ tục kiểm định giả thuyết thống kê

Bước 1: Phát biểu giả thuyết H_0 , đối thuyết H_1 .

Bước 2: Chọn tiêu chuẩn kiểm định với giả thuyết H_0 đúng.

Bước 3: Với mức ý nghĩa α , xác định miền bác bỏ W_α .

Bước 4: Tính giá trị quan sát u_{qs} .

Bước 5: So sánh giá trị quan sát với miền bác bỏ và kết luận.

2.1 Trường hợp đã biết phương sai

Bước 1: Phát biểu giả thuyết $H_0 : \mu = \mu_0$, đối thuyết H_1 .

Bước 2: Chọn tiêu chuẩn $U = \frac{\bar{X} - \mu_0}{\sigma} \sqrt{n}$.

Bước 3: Miền bác bỏ W_α được xác định theo các trường hợp:

H_0	H_1	W_α
$\mu = \mu_0$	$\mu \neq \mu_0$	$\left(-\infty; -u_{1-\frac{\alpha}{2}}\right) \cup \left(u_{1-\frac{\alpha}{2}}; +\infty\right)$
$\mu = \mu_0$	$\mu > \mu_0$	$(u_{1-\alpha}; +\infty)$
$\mu = \mu_0$	$\mu < \mu_0$	$(-\infty; -u_{1-\alpha})$

Bước 4: Tính giá trị quan sát $u_{qs} = \frac{\bar{x} - \mu_0}{\sigma} \sqrt{n}$ với \bar{x}, μ_0, n đã biết.

Bước 5: Xem xét u_{qs} có thuộc W_α hay không

- Nếu thuộc miền bác bỏ \rightarrow bác bỏ H_0 , chấp nhận H_1 .
- Nếu không thuộc miền bác bỏ \rightarrow không bác bỏ H_0

Ví dụ: Một hãng bảo hiểm thông báo rằng số tiền trung bình hãng chi trả cho khách hàng bị tai nạn ô tô là 8500 USD. Để kiểm tra lại, người ta kiểm tra ngẫu nhiên hồ sơ chi trả của 25 khách hàng thì thấy số tiền trung bình chi trả là 8900 USD. Giả sử số tiền chi trả tuân theo luật phân phối chuẩn. với độ lệch chuẩn là 2600 USD. Hãy kiểm định lại thông báo của hãng bảo hiểm trên với mức ý nghĩa 5%.

HƯỚNG DẪN

Theo bài ra:

- Số tiền hãng thông báo: $\mu_0 = 8500$.
- 25 khách hàng: $n = 25$.
- Số tiền trung bình/25 người: $\bar{x} = 8900$.
- Độ lệch chuẩn $\sigma = 2600$.
- Mức ý nghĩa 5%: $\alpha = 0.05$.

Bước 1: $H_0 : \mu = \mu_0; H_1 : \mu \neq \mu_0$.

Bước 2: Chọn tiêu chuẩn kiểm định: $U = \frac{\bar{X} - \mu_0}{\sigma} \sqrt{n}$.

Bước 3: Với $\alpha = 0.05, u_{1-\frac{\alpha}{2}} = u_{0.975} = 1.96$. Miền bác bỏ $W_\alpha = (-\infty; -1.96) \cup (1.96; +\infty)$

Bước 4: $u_{qs} = \frac{\bar{x} - \mu_0}{\sigma} \sqrt{n} = \frac{8900 - 8500}{2600} \sqrt{25} = 0.77 \notin W_\alpha \Rightarrow$ chấp nhận H_0

Bài tập:

Bài 1: Giám đốc một xí nghiệp cho biết lương trung bình của 1 công nhân thuộc xí nghiệp là 380 ngàn đ/tháng. Chọn ngẫu nhiên 36 công nhân thấy lương trung bình là 350 ngàn đ/tháng, với độ lệch chuẩn $\sigma = 40$. Lời báo cáo của giám đốc có tin cậy được không, với mức có ý nghĩa là $\alpha = 5\%$. (ĐS: $u_{qs} = -4.5$)

Bài 2: Nếu máy móc hoạt động bình thường thì trọng lượng sản phẩm là biến ngẫu nhiên có phân phối chuẩn $N(\mu, \sigma^2)$ với trọng lượng trung bình $\mu_0 = 100\text{gam}$, độ lệch tiêu chuẩn $\sigma = 2\text{gam}$. Qua một thời gian sản xuất người ta nghi ngờ trọng lượng sản phẩm có xu hướng tăng lên, cân thử 100 sản phẩm thì trọng lượng trung bình của chúng là 100,4 gam. Với mức ý nghĩa $\alpha = 5\%$ Hãy kết luận về điều nghi ngờ trên. (ĐS: $u_{qs} = 2$)

Bài 3: Một cửa hàng thực phẩm nhận thấy thời gian vừa qua trung bình một khách hàng mua 25 ngàn đồng thực phẩm trong ngày. Nay cửa hàng chọn ngẫu nhiên 15 khách hàng thấy trung bình một khách hàng mua 24 ngàn đồng trong ngày và độ lệch chuẩn là 2. Với mức ý nghĩa là 5%, kiểm định xem có phải sức mua của khách hàng hiện nay thực sự giảm sút hay không. Biết rằng sức mua của khách hàng có phân phối chuẩn. (ĐS: $u_{qs} = -1.761$)

2.2 Trường hợp chưa biết phương sai, cỡ mẫu $n < 30$

Bước 1: Phát biểu giả thuyết $H_0: \mu = \mu_0$, đối thuyết H_1 .

Bước 2: Chọn tiêu chuẩn $T = \frac{\bar{X} - \mu_0}{s} \sqrt{n}$.

Bước 3: Miền bác bỏ W_α được xác định theo các trường hợp:

H_0	H_1	W_α
$\mu = \mu_0$	$\mu \neq \mu_0$	$\left(-\infty; -t_{1-\frac{\alpha}{2}}^{(n-1)}\right) \cup \left(t_{1-\frac{\alpha}{2}}^{(n-1)}; +\infty\right)$
$\mu = \mu_0$	$\mu > \mu_0$	$\left(t_{1-\alpha}^{(n-1)}; +\infty\right)$
$\mu = \mu_0$	$\mu < \mu_0$	$\left(-\infty; -t_{1-\alpha}^{(n-1)}\right)$

Bước 4: Tính giá trị quan sát $t_{qs} = \frac{\bar{x} - \mu_0}{s} \sqrt{n}$ với \bar{x}, μ_0, s, n đã biết.

Bước 5: Xem xét t_{qs} có thuộc W_α hay không

- Nếu thuộc miền bác bỏ \rightarrow bác bỏ H_0 , chấp nhận H_1 .
- Nếu không thuộc miền bác bỏ \rightarrow không bác bỏ H_0

Bài tập:

Bài 1: Quan sát số hoa hồng bán ra trong một ngày của một cửa hàng bán hoa sau một thời gian, người ta ghi được số liệu sau:

Số hoa hồng (đóa)	12	13	15	16	17	18	19
Số ngày	3	2	7	7	3	2	1

Sau khi tính toán, ông chủ cửa hàng nói rằng nếu trung bình một ngày không bán được 15 đóa hoa thì chẳng thà đóng cửa còn hơn. Dựa vào số liệu trên, anh (chị) hãy kết luận giúp ông chủ cửa hàng xem có nên tiếp tục bán hay không ở mức ý nghĩa $\alpha = 0.05$. (ĐS: $t_{qs} = 1.0691$)

Bài 2: Một công ty sản xuất hạt giống tuyên bố rằng một loại giống mới của họ có năng suất trung bình là 21,5 tạ/ha. Gieo thử hạt giống mới này tại 16 vườn thí nghiệm và thu được kết quả:

19, 2; 18, 7; 22, 4; 20, 3; 16, 8; 25, 1; 17, 0; 15, 8; 21, 0; 18, 6; 23, 7; 24, 1; 23, 4; 19, 8; 21, 7; 18, 9

Dựa vào kết quả này hãy xác nhận xem quảng cáo của công ty có đúng không với mức ý nghĩa $\alpha = 0,05$. (ĐS: $t_{qs} = -1.44$)

Bài 3: Đo cholesterol (đơn vị mg%) cho một nhóm người, ta ghi nhận lại được:

Chol.	150 – 160	160 – 170	170 – 180	180 – 190	190 – 200	200 – 210
Số người	3	9	11	3	2	1

Có tài liệu cho biết lượng cholesterol trung bình là $\mu_0 = 175$ mg%. Giá trị này có phù hợp với mẫu quan sát không? (kết luận với $\alpha = 0.05$). (ĐS: $u_{qs} = 2.048$)

2.3 Trường hợp chưa biết phương sai, cỡ mẫu $n > 30$

Bước 1: Phát biểu giả thuyết $H_0: \mu = \mu_0$, đối thuyết H_1 .

Bước 2: Chọn tiêu chuẩn $U = \frac{\bar{X} - \mu_0}{s} \sqrt{n}$.

Bước 3: Miền bác bỏ W_α được xác định theo các trường hợp:

H_0	H_1	W_α
$\mu = \mu_0$	$\mu \neq \mu_0$	$\left(-\infty; -u_{1-\frac{\alpha}{2}}\right) \cup \left(u_{1-\frac{\alpha}{2}}; +\infty\right)$
$\mu = \mu_0$	$\mu > \mu_0$	$(u_{1-\alpha}; +\infty)$
$\mu = \mu_0$	$\mu < \mu_0$	$(-\infty; -u_{1-\alpha})$

Bước 4: Tính giá trị quan sát $u_{qs} = \frac{\bar{x} - \mu_0}{s} \sqrt{n}$ với \bar{x}, μ_0, n đã biết.

Bước 5: Xem xét u_{qs} có thuộc W_α hay không

- Nếu thuộc miền bác bỏ \rightarrow bác bỏ H_0 , chấp nhận H_1 .
- Nếu không thuộc miền bác bỏ \rightarrow không bác bỏ H_0

Bài tập:

Bài 1: Một máy đóng gói các sản phẩm có khối lượng 1 kg. Nghi ngờ máy hoạt động không bình thường, người ta chọn ra một mẫu ngẫu nhiên gồm 100 sản phẩm thì thấy như sau:

Khối lượng	0.95	0.97	0.99	1.01	1.03	1.05
Số gói	9	31	40	15	3	2

Với mức ý nghĩa 0.05, hãy kết luận về nghi ngờ trên. (ĐS: $u_{qs} = -6.9204$).

Bài 2: Một xí nghiệp đúc một số rất lớn các sản phẩm bằng thép với số khuyết tật trung bình ở mỗi sản phẩm là 3. Người ta cải tiến cách sản xuất và kiểm tra 36 sản phẩm. Kết quả như sau:

Số khuyết tật trên sản phẩm	0	1	2	3	4	5	6
Số sản phẩm tương ứng	7	4	5	7	6	6	1

Hãy cho kết luận về hiệu quả của việc cải tiến sản xuất ở mức ý nghĩa 0.05. (ĐS: $u_{qs} = -1.65$).

2.4 Kiểm định giả thuyết về tỷ lệ hay xác suất

Bước 1: Phát biểu giả thuyết $H_0: p = p_0$, đối thuyết H_1 .

Bước 2: Chọn tiêu chuẩn $U = \frac{f - p_0}{\sqrt{p_0(1-p_0)}} \sqrt{n}$.

Bước 3: Miền bác bỏ W_α được xác định theo các trường hợp:

H_0	H_1	W_α
$p = p_0$	$p \neq p_0$	$\left(-\infty; -u_{1-\frac{\alpha}{2}}\right) \cup \left(u_{1-\frac{\alpha}{2}}; +\infty\right)$
$p = p_0$	$p > p_0$	$(u_{1-\alpha}; +\infty)$
$p = p_0$	$p < p_0$	$(-\infty; -u_{1-\alpha})$

Bước 4: Tính giá trị quan sát $u_{qs} = \frac{f - p_0}{\sqrt{p_0(1-p_0)}} \sqrt{n}$ với $f = \frac{m}{n}, p_0, n$ được tính từ đề bài

Bước 5: Xem xét u_{qs} có thuộc W_α hay không

- Nếu thuộc miền bác bỏ \rightarrow bác bỏ H_0 , chấp nhận H_1 .

- Nếu không thuộc miền bác bỏ \rightarrow không bác bỏ H_0

Bài tập:

Bài 1: Một máy sản xuất tự động với tỷ lệ chính phẩm là 98%. Sau một thời gian hoạt động, người ta nghi ngờ tỷ lệ trên đã bị giảm. Kiểm tra ngẫu nhiên 500 sản phẩm thấy có 28 phế phẩm, với $\alpha = 0.05$ hãy kiểm tra xem chất lượng làm việc của máy có còn được như trước hay không? (ĐS: $u_{qs} = -5.7499$)

Bài 2: Theo một nguồn tin thì tỉ lệ hộ dân thích xem dân ca trên Tivi là 80%. Thăm dò 36 hộ dân thấy có 25 hộ thích xem dân ca. Với mức có ý nghĩa là 5%. Kiểm định xem nguồn tin này có đáng tin cậy không? (ĐS: $u_{qs} = -1.584$)

3 So sánh hai kỳ vọng của hai biến ngẫu nhiên phân phối chuẩn

3.1 Trường hợp đã biết phương sai

Bước 1: Phát biểu giả thuyết $H_0 : \mu_1 = \mu_2$, đối thuyết H_1 .

Bước 2: Chọn tiêu chuẩn $U = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$.

Bước 3: Miền bác bỏ W_α được xác định theo các trường hợp:

H_0	H_1	W_α
$\mu_1 = \mu_2$	$\mu_1 \neq \mu_2$	$\left(-\infty; -u_{1-\frac{\alpha}{2}}\right) \cup \left(u_{1-\frac{\alpha}{2}}; +\infty\right)$
$\mu_1 = \mu_2$	$\mu_1 > \mu_2$	$(u_{1-\alpha}; +\infty)$
$\mu_1 = \mu_2$	$\mu_1 < \mu_2$	$(-\infty; -u_{1-\alpha})$

Bước 4: Tính giá trị quan sát $u_{qs} = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ với $\bar{x}, \bar{y}, \mu_1, \mu_2, n_1, n_2$ đã biết.

Bước 5: Xem xét u_{qs} có thuộc W_α hay không

- Nếu thuộc miền bác bỏ \rightarrow bác bỏ H_0 , chấp nhận H_1 .
- Nếu không thuộc miền bác bỏ \rightarrow không bác bỏ H_0

Bài tập:

Bài 1: Một nhà phát triển sản phẩm quan tâm đến việc giảm thời gian khô của sơn. Vì vậy hai công thức sơn được đem thử nghiệm. Công thức 1 là công thức có các thành phần chuẩn và công thức 2 có thêm một thành phần làm khô mới được cho rằng sẽ làm giảm thời gian khô của sơn. Từ các thí nghiệm người ta thấy rằng $\sigma_1 = \sigma_2 = 8$ phút. 10 đồ vật được sơn với công thức 1 và 10 đồ vật khác được sơn với công thức 2. Thời gian khô trung bình của từng mẫu là $x_1 = 121$ phút và $x_2 = 112$ phút. Nhà phát triển sản phẩm có thể rút ra kết luận gì về ảnh hưởng của thành phần làm khô mới? Với mức ý nghĩa 5%.

Bài 2: Tốc độ cháy của hai loại chất nổ lỏng được dùng làm nhiên liệu trong tàu vũ trụ được nghiên cứu. Người ta biết rằng độ lệch chuẩn của tốc độ cháy của hai loại nhiên liệu bằng nhau và bằng 3 cm/s. Hai mẫu ngẫu nhiên kích thước $n_1 = 20$ và $n_2 = 20$ được thử nghiệm; trung bình mẫu tốc độ cháy là $x_1 = 18$ cm/s và $x_2 = 24$ cm/s. Với mức ý nghĩa $\alpha = 0.05$ hãy kiểm định giả thuyết hai loại chất nổ lỏng này có cùng tốc độ đốt cháy.

3.2 Trường hợp chưa biết phương sai $n_1 < 30, n_2 < 30$

Bước 1: Phát biểu giả thuyết $H_0 : \mu_1 = \mu_2$, đối thuyết H_1 .

Bước 2: Chọn tiêu chuẩn $T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$.

Bước 3: Miền bác bỏ W_α được xác định theo các trường hợp:

H_0	H_1	W_α
$\mu_1 = \mu_2$	$\mu_1 \neq \mu_2$	$\left(-\infty; -t_{1-\frac{\alpha}{2}}^{(n-1)} \right) \cup \left(t_{1-\frac{\alpha}{2}}^{(n-1)}; +\infty \right)$
$\mu_1 = \mu_2$	$\mu_1 > \mu_2$	$\left(t_{1-\alpha}^{(n-1)}; +\infty \right)$
$\mu_1 = \mu_2$	$\mu_1 < \mu_2$	$\left(-\infty; -t_{1-\alpha}^{(n-1)} \right)$

Bước 4: Tính giá trị quan sát $t_{qs} = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$

với $\bar{x}, \bar{y}, \mu_1, \mu_2, n_1, n_2$ đã biết.

Bước 5: Xem xét t_{qs} có thuộc W_α hay không

- Nếu thuộc miền bác bỏ \rightarrow bác bỏ H_0 , chấp nhận H_1 .
- Nếu không thuộc miền bác bỏ \rightarrow không bác bỏ H_0

Bài tập:

Bài 1: Để so sánh năng lực học toán và vật lý của học sinh, người ta kiểm tra ngẫu nhiên 8 em bằng hai bài toán và vật lý. Kết quả cho bởi bảng dưới đây (X là điểm toán, Y là điểm lý):

X	15	20	16	22	24	18	20	14
Y	15	22	14	25	19	20	24	16

Giả sử X và Y đều có phân phối chuẩn. Hãy so sánh điểm trung bình giữa X và Y, mức ý nghĩa 5%.

3.3 Trường hợp chưa biết phương sai $n_1 > 30, n_2 > 30$

Bước 1: Phát biểu giả thuyết $H_0 : \mu_1 = \mu_2$, đối thuyết H_1 .

Bước 2: Chọn tiêu chuẩn $U = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$.

Bước 3: Miền bác bỏ W_α được xác định theo các trường hợp:

H_0	H_1	W_α
$\mu_1 = \mu_2$	$\mu_1 \neq \mu_2$	$\left(-\infty; -u_{1-\frac{\alpha}{2}} \right) \cup \left(u_{1-\frac{\alpha}{2}}; +\infty \right)$
$\mu_1 = \mu_2$	$\mu_1 > \mu_2$	$\left(u_{1-\alpha}; +\infty \right)$
$\mu_1 = \mu_2$	$\mu_1 < \mu_2$	$\left(-\infty; -u_{1-\alpha} \right)$

Bước 4: Tính giá trị quan sát $u_{qs} = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ với $\bar{x}, \bar{y}, \mu_1, \mu_2, n_1, n_2$ đã biết.

Bước 5: Xem xét u_{qs} có thuộc W_α hay không

- Nếu thuộc miền bác bỏ \rightarrow bác bỏ H_0 , chấp nhận H_1 .
- Nếu không thuộc miền bác bỏ \rightarrow không bác bỏ H_0

Bài tập:

Bài 1:

Trồng cùng một giống lúa trên hai thửa ruộng như nhau và bón hai loại phân khác nhau. Đến ngày thu hoạch ta có kết quả như sau:

- Thửa thứ nhất lấy mẫu 1000 bông lúa thấy số hạt trung bình của mỗi bông là $x = 70$ hạt và $s = 10$
- Thửa thứ hai lấy mẫu 500 bông thấy số hạt trung bình mỗi bông là $y = 72$ hạt và $s = 20$.

So sánh giữa X và Y với $\alpha = 0.05$

3.4 So sánh hai tỷ lệ

Bước 1: Phát biểu giả thuyết $H_0 : p_1 = p_2$, đối thuyết H_1 .

Đặt $\bar{f} = \frac{m_1 + m_2}{n_1 + n_2} = \frac{n_1 f_1 + n_2 f_2}{n_1 + n_2}$

Bước 2: Chọn tiêu chuẩn $U = \frac{f_1 - f_2}{\sqrt{\bar{f}(1-\bar{f})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$.

Bước 3: Miền bác bỏ W_α được xác định theo các trường hợp:

H_0	H_1	W_α
$p_1 = p_2$	$p_1 \neq p_2$	$(-\infty; -u_{1-\frac{\alpha}{2}}) \cup (u_{1-\frac{\alpha}{2}}; +\infty)$
$p_1 = p_2$	$p_1 > p_2$	$(u_{1-\alpha}; +\infty)$
$p_1 = p_2$	$p_1 < p_2$	$(-\infty; -u_{1-\alpha})$

Bước 4: Tính giá trị quan sát $u_{qs} = \frac{f_1 - f_2}{\sqrt{\bar{f}(1-\bar{f})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ với $f_1, f_2, \bar{n}, n_1, n_2$ được tính từ đề bài

Bước 5: Xem xét u_{qs} có thuộc W_α hay không

- Nếu thuộc miền bác bỏ \rightarrow bác bỏ H_0 , chấp nhận H_1 .
- Nếu không thuộc miền bác bỏ \rightarrow không bác bỏ H_0

Bài tập:

Bài 1: Người ta điều tra 250 người ở xã A thấy có 140 nữ và điều tra 160 người ở xã B thấy có 80 nữ. Hãy so sánh tỉ lệ nữ ở hai xã với mức ý nghĩa 5%.

Bài 2: Áp dụng hai phương pháp gieo hạt. Theo phương pháp A gieo 180 hạt thì có 150 hạt nảy mầm; theo phương pháp B gieo 256 hạt thì thấy có 160 hạt nảy mầm. Hãy so sánh hiệu quả của hai phương pháp với mức ý nghĩa $\alpha = 5\%$.