

➤ Chương 7. Kiểm định Giả thuyết Thống kê

§1. Khái niệm về kiểm định giả thuyết thống kê

§2. Kiểm định so sánh đặc trưng với một số

§3. Kiểm định so sánh hai đặc trưng

§1. KHÁI NIỆM VỀ KIỂM ĐỊNH GIẢ THUYẾT THỐNG KÊ

1.1. Khái niệm chung

- Mô hình tổng quát của bài toán kiểm định là: ta nêu lên hai mệnh đề trái ngược nhau, một mệnh đề được gọi là *giả thuyết H* và mệnh đề còn lại được gọi là *ngịch thuyết* (hay *đối thuyết*) \bar{H} .
- Giải quyết một bài toán kiểm định là: bằng cách dựa vào quan sát mẫu, ta nêu lên một quy tắc hành động, ta chấp nhận giả thuyết H hay bác bỏ giả thuyết H .

➤ Chương 7. Kiểm định Giả thuyết Thống kê

- Khi ta chấp nhận giả thuyết H , nghĩa là ta tin rằng H đúng; khi bác bỏ H , nghĩa là ta tin rằng H sai. Do chỉ dựa trên một mẫu quan sát ngẫu nhiên, nên ta không thể khẳng định chắc chắn điều gì cho tổng thể.
- Trong chương này, ta chỉ xét loại kiểm định tham số (so sánh đặc trưng với 1 số, so sánh hai đặc trưng của hai tổng thể).

1.2. Các loại sai lầm trong kiểm định

Khi thực hiện kiểm định giả thuyết, ta dựa vào quan sát ngẫu nhiên một số trường hợp rồi suy rộng ra cho tổng thể. Sự suy rộng này có khi đúng, có khi sai.

Thống kê học phân biệt 2 loại sai lầm sau:

a) Sai lầm loại I

- Sai lầm loại 1 là loại sai lầm mà ta phạm phải trong việc bác bỏ giả thuyết H khi H đúng.
- Xác suất của việc bác bỏ H khi H đúng là xác suất của sai lầm loại 1 và được ký hiệu là α .

b) Sai lầm loại II

- Sai lầm loại 2 là loại sai lầm mà ta phạm phải trong việc chấp nhận giả thuyết H khi H sai.
- Xác suất của việc chấp nhận giả thuyết H khi H sai là xác suất của sai lầm loại 2 và được ký hiệu là β .

c) Mỗi liên hệ giữa hai loại sai lầm

- Khi thực hiện kiểm định, ta luôn muốn xác suất phạm phải sai lầm càng ít càng tốt. Tuy nhiên, nếu hạ thấp α thì β sẽ tăng lên và ngược lại.

Trong thực tế, giữa hai loại sai lầm này, loại nào tác hại hơn thì ta nên tránh.

- Trong thống kê, người ta quy ước rằng sai lầm loại 1 tác hại hơn loại 2 nên cần tránh hơn. Do đó, ta chỉ xét các phép kiểm định có α không vượt quá một giá trị ấn định trước, thông thường là 1%; 3%; 5%;...
Giá trị α còn được gọi là *mức ý nghĩa* của kiểm định.

1.3. Cơ sở lý thuyết của kiểm định

- Để giải quyết bài toán kiểm định, ta quan sát mẫu ngẫu nhiên X_1, \dots, X_n và đưa ra giả thuyết H .
- Từ mẫu trên, ta chọn thống kê $T = f(X_1, \dots, X_n; \theta_0)$ sao cho nếu khi H đúng thì phân phối xác suất của T hoàn toàn xác định.
- Với mức ý nghĩa α , ta tìm được khoảng tin cậy (hay khoảng ước lượng) $[a; b]$ cho T ở độ tin cậy $1 - \alpha$.

Khi đó:

- nếu $t \in [a; b]$ thì ta chấp nhận giả thuyết H ;
- nếu $t \notin [a; b]$ thì ta bác bỏ giả thuyết H .

➤ Chương 7. Kiểm định Giả thuyết Thống kê

- Nếu hàm mật độ của T đối xứng qua trục Oy thì ta chọn khoảng đối xứng $[-t_\alpha; t_\alpha]$, với:

$$P(T \leq -t_\alpha) = P(T \geq t_\alpha) = \frac{\alpha}{2}.$$

Vậy, khi xét nửa bên phải của trục Oy thì ta được:

- nếu $t \leq t_\alpha$ thì ta chấp nhận giả thuyết H ;
 - nếu $t > t_\alpha$ thì ta bác bỏ giả thuyết H .
- Nếu hàm mật độ của T không đối xứng qua trục Oy thì ta chọn khoảng tin cậy $[0; C]$, với $P(T \geq C) = \alpha$.
 - Nếu $t \leq C$ thì ta chấp nhận giả thuyết H , và
 - nếu $t > C$ thì ta bác bỏ giả thuyết H .

.....

§2. KIỂM ĐỊNH SO SÁNH ĐẶC TRƯNG CỦA TỔNG THỂ VỚI MỘT SỐ

2.1. Kiểm định so sánh trung bình với một số

Với số μ_0 cho trước, ta đặt giả thuyết $H : \mu = \mu_0$.

a) Trường hợp 1. Với $n \geq 30$, σ^2 đã biết.

- Từ mức ý nghĩa $\alpha \Rightarrow \frac{1-\alpha}{2} = \varphi(t_\alpha) \xrightarrow{B} t_\alpha$.

- Tính giá trị thống kê $t = \frac{|\bar{x} - \mu_0|}{\sigma} \sqrt{n}$.

- Nếu $t \leq t_\alpha$ thì ta chấp nhận H , nghĩa là $\mu = \mu_0$;
nếu $t > t_\alpha$ thì ta bác bỏ H , nghĩa là $\mu \neq \mu_0$.

➤ Chương 7. Kiểm định Giả thuyết Thống kê

b) Trường hợp 2. Với $n \geq 30$, σ^2 chưa biết.

Ta làm như trường hợp 1 nhưng thay σ bằng s .

c) Trường hợp 3. Với $n < 30$, σ^2 đã biết và X có phân phối chuẩn, ta làm như trường hợp 1.

d) Trường hợp 4. Với $n < 30$, σ^2 chưa biết và X có phân phối chuẩn.

- Từ cỡ mẫu n và mức ý nghĩa $\alpha \xrightarrow{\text{tra bảng C}} t_{\alpha}^{n-1}$.
- Tính giá trị thống kê $t = \frac{|\bar{x} - \mu_0|}{s} \sqrt{n}$.
- Nếu $t \leq t_{\alpha}^{n-1}$ thì ta chấp nhận giả thuyết H ;
 $t > t_{\alpha}^{n-1}$ thì ta bác bỏ giả thuyết H .

Chú ý

Trong tất cả các trường hợp bác bỏ, ta so sánh \bar{x} và μ_0 :

- Nếu $\bar{x} > \mu_0$ thì ta kết luận $\mu > \mu_0$.
- Nếu $\bar{x} < \mu_0$ thì ta kết luận $\mu < \mu_0$.

➤ Chương 7. Kiểm định Giả thuyết Thống kê

VD 1. Sở Điện lực A báo cáo rằng: trung bình một hộ hàng tháng phải trả 250 ngàn đồng tiền điện, với độ lệch chuẩn là 20 ngàn. Người ta khảo sát ngẫu nhiên 500 hộ thì tính được trung bình hàng tháng một hộ trả 252 ngàn đồng tiền điện.

Trong kiểm định giả thuyết H : “**trung bình một hộ phải trả hàng tháng là 250 ngàn đồng tiền điện**” với mức ý nghĩa $\alpha = 1\%$, hãy cho biết giá trị thống kê t và kết luận ?

Giải

Do $n = 500$, $\sigma = 20$ nên bài toán thuộc TH1.

➤ Chương 7. Kiểm định Giả thuyết Thống kê

Ta có: $\bar{x} = 252$ (ngàn đồng), và

$$\alpha = 0,01 \Rightarrow \frac{1-\alpha}{2} = 0,495 \Rightarrow t_{\alpha} = 2,58.$$

Giá trị thống kê:

$$t = \frac{|\bar{x} - \mu_0|}{\sigma} \sqrt{n} = \frac{|252 - 250|}{20} \sqrt{500} = 2,2361.$$

Vì $t < t_{\alpha}$ nên ta chấp nhận giả thuyết H .

Vậy $t = 2,2361$ và thực tế, trung bình một hộ hàng tháng phải trả là 250 ngàn đồng tiền điện (với mức ý nghĩa 1%).

➤ Chương 7. Kiểm định Giả thuyết Thống kê

VD 2. Nhà Giáo dục học B muốn nghiên cứu xem số giờ tự học trung bình hàng ngày của sinh viên có thay đổi không so với mức 1 giờ/ngày cách đây 10 năm.

Ông B khảo sát ngẫu nhiên 120 sinh viên và tính được trung bình là 0,82 giờ/ngày với $\hat{s} = 0,75$ giờ/ngày.

Với mức ý nghĩa 3%, hãy cho biết kết luận của ông B ?

Giải

Do $n = 120$, σ^2 chưa biết nên bài toán thuộc TH2.

Đặt giả thuyết $H : \mu = \mu_0 = 1$ (giờ/ngày).

Ta có: $\bar{x} = 0,82$ và $s = \hat{s} \cdot \sqrt{\frac{n}{n-1}} = 0,7531$ (giờ/ngày).

➤ Chương 7. Kiểm định Giả thuyết Thống kê

Giá trị thống kê:

$$t = \frac{|\bar{x} - \mu_0|}{s} \sqrt{n} = \frac{|0,82 - 1|}{0,7531} \sqrt{120} = 2,6182.$$

Vì $t > t_{3\%} = 2,17$ nên ta bác bỏ giả thuyết H .

Mặt khác, $\bar{x} < 1$ (giờ/ngày) nên kết luận của ông B là:
Thực tế, số giờ tự học trung bình hàng ngày của sinh viên hiện nay thấp hơn so với cách đây 10 năm.

➤ Chương 7. Kiểm định Giả thuyết Thống kê

VD 3. Trong một nhà máy gạo, trọng lượng đóng bao theo quy định của một bao gạo là 50 kg và độ lệch chuẩn là 0,3 kg. Cân thử 296 bao gạo của nhà máy này thì thấy trọng lượng trung bình là 49,97 kg. Kiểm định giả thuyết H : “**trọng lượng mỗi bao gạo của nhà máy này là 50 kg**” có giá trị thống kê t và kết luận là:

- A. $t = 1,7205$; chấp nhận H với mức ý nghĩa 6%.
- B. $t = 1,7205$; bác bỏ H , trọng lượng thực tế của bao gạo nhỏ hơn 50 kg với mức ý nghĩa 6%.
- C. $t = 1,9732$; chấp nhận H với mức ý nghĩa 4%.
- D. $t = 1,9732$; bác bỏ H , trọng lượng thực tế của bao gạo nhỏ hơn 50 kg với mức ý nghĩa 4%.

Đáp án đúng: A.

VD 4. Một công ty cho biết mức lương trung bình của một kỹ sư ở công ty là 5,7 triệu đồng/tháng với độ lệch chuẩn 0,5 triệu đồng/tháng. Kỹ sư A dự định xin vào làm ở công ty này và đã thăm dò 18 kỹ sư thì thấy lương trung bình là 5,45 triệu đồng/tháng.

Kỹ sư A quyết định rằng: nếu mức lương trung bình bằng với mức công ty đưa ra thì nộp đơn xin làm.

Với mức ý nghĩa 2%, cho biết kết luận của kỹ sư A ?

Giải

Do $n = 18$, $\sigma = 0,5$ nên bài toán thuộc TH3.

Đặt giả thuyết $H : \mu = \mu_0 = 5,7$ (triệu đồng/tháng).

➤ Chương 7. Kiểm định Giả thuyết Thống kê

Ta có: $\bar{x} = 5,45$ và

$$\alpha = 0,02 \Rightarrow \frac{1 - \alpha}{2} = 0,49 \Rightarrow t_{\alpha} = 2,33.$$

$$t = \frac{|\bar{x} - \mu_0|}{\sigma} \sqrt{n} = \frac{|5,45 - 5,7|}{0,5} \sqrt{18} = 2,1213.$$

Vì $t < t_{\alpha}$ nên ta chấp nhận giả thuyết H .

Vậy kỹ sư A quyết định nộp đơn xin vào công ty trên.

➤ Chương 7. Kiểm định Giả thuyết Thống kê

VD 5. Người ta kiểm tra ngẫu nhiên 38 cửa hàng của công ty A và có bảng doanh thu trong 1 tháng là:

X (triệu đồng/tháng)	200	220	240	260
Số cửa hàng	8	16	12	2

Kiểm định giả thuyết H : “**doanh thu trung bình hàng tháng của một cửa hàng công ty là 230 triệu đồng**”, mức ý nghĩa tối đa để giả thuyết H được chấp nhận là:

- A. 3,4%; B. 4,2%; C. 5,6%; D. 7,8%.

Giải

Do $n = 38$, σ^2 chưa biết nên bài toán thuộc TH2.

Từ bảng số liệu ta tính được:

$$\bar{x} = 224,2105 \text{ và } s = 16,867.$$

➤ Chương 7. Kiểm định Giả thuyết Thống kê

Giá trị thống kê:

$$t = \frac{|\bar{x} - \mu_0|}{s} \sqrt{n} = \frac{|224,2105 - 230| \cdot \sqrt{38}}{16,867} \approx 2,12.$$

Do giả thuyết H được chấp nhận nên:

$$t \leq t_{\alpha} \Rightarrow \varphi(t_{\alpha}) \geq \varphi(2,12)$$

$$\Rightarrow \frac{1 - \alpha}{2} \geq 0,483 \Rightarrow \alpha \leq 0,034 \Rightarrow A.$$

➤ Chương 7. Kiểm định Giả thuyết Thống kê

VD 6. Điểm trung bình môn Toán của sinh viên năm trước là 5,72. Năm nay, theo dõi 100 SV được số liệu:

Điểm	3	4	5	6	7	8	9
Số sinh viên	3	5	27	43	12	6	4

Kiểm định giả thuyết H : “**điểm trung bình môn Toán của sinh viên năm nay bằng năm trước**”, mức ý nghĩa tối đa để H được chấp nhận là:

A. 13,94%; B. 13,62%; C. 11,74%; D. 11,86%.

Hướng dẫn

$\bar{x} = 5,9$ và $s = 1,2102$.

Đáp án đúng là B.

➤ Chương 7. Kiểm định Giả thuyết Thống kê

VD 7. Thời gian X (phút) giữa hai chuyến xe bus trong một thành phố là biến ngẫu nhiên có phân phối chuẩn. Công ty xe bus nói rằng: trung bình cứ 5 phút lại có 1 chuyến xe bus. Người ta chọn ngẫu nhiên 8 thời điểm và ghi lại thời gian (phút) giữa hai chuyến xe bus là:

5,3; 4,5; 4,8; 5,1; 4,3; 4,8; 4,9; 4,7.

Với mức ý nghĩa 5%, hãy kiểm định lời nói trên ?

Giải

Do $n = 8$, σ^2 chưa biết nên bài toán thuộc TH4.

Đặt giả thuyết $H : \mu = \mu_0 = 5$ (phút).

Từ bảng số liệu ta tính được:

$$\bar{x} = 4,8 \text{ (phút)}, s = 0,3162 \text{ (phút)}.$$

➤ Chương 7. Kiểm định Giả thuyết Thống kê

Ta có: $\alpha = 0,05 \xrightarrow{\text{tra bảng C}} t_{0,05}^7 = 2,365.$

$$t = \frac{|\bar{x} - \mu_0|}{s} \sqrt{n} = \frac{|4,8 - 5|}{0,3162} \sqrt{8} = 1,789.$$

Do $t < t_{0,05}^7$ nên ta chấp nhận giả thuyết H .

Vậy lời nói trên là đúng thực tế.

➤ Chương 7. Kiểm định Giả thuyết Thống kê

VD 8. Chiều cao cây giống X (m) trong một vườn ươm là biến ngẫu nhiên có phân phối chuẩn. Người ta đo ngẫu nhiên 25 cây giống này và có bảng số liệu:

X (m)	0,8	0,9	1,0	1,1	1,2	1,3
Số cây	1	2	9	7	4	2

Theo quy định của vườn ươm, khi nào cây cao hơn 1 m thì đem ra trồng. Với mức ý nghĩa 5%, kiểm định giả thuyết H : “**cây giống của vườn ươm cao 1 m**” có giá trị thống kê và kết luận là:

- A. $t = 2,7984$, không nên đem cây ra trồng.
- B. $t = 2,7984$, nên đem cây ra trồng.
- C. $t = 1,9984$, không nên đem cây ra trồng.
- D. $t = 1,9984$, nên đem cây ra trồng.

Giải

Do $n < 30$, σ^2 chưa biết nên bài toán thuộc TH4.

Từ bảng số liệu ta tính được:

$$\bar{x} = 1,068 \text{ m}, s = 0,1215 \text{ m}.$$

Ta có: $\alpha = 0,05 \xrightarrow{\text{tra bảng C}} t_{0,05}^{24} = 2,064.$

$$t = \frac{|\bar{x} - \mu_0|}{s} \sqrt{n} = \frac{|1,068 - 1| \cdot 5}{0,1215} = 2,7984 > t_{0,05}^{24}.$$

Giả thuyết H bị bác bỏ, do $\bar{x} > \mu_0$ nên cây giống đã cao hơn 1 m, và theo quy định nên đem cây ra trồng.
Vậy chọn đáp án B.

2.2. Kiểm định so sánh tỉ lệ với một số

- Với số p_0 cho trước, ta đặt giả thuyết $H : p = p_0$.
- Từ mức ý nghĩa $\alpha \Rightarrow \frac{1-\alpha}{2} = \varphi(t_\alpha) \xrightarrow{B} t_\alpha$.
- Từ mẫu cụ thể, ta tính tỉ lệ mẫu $f = \frac{m}{n}$ và

giá trị thống kê $t = \frac{|f - p_0|}{\sqrt{p_0 q_0}} \sqrt{n}$, $q_0 = 1 - p_0$.

- Nếu $t \leq t_\alpha$ thì chấp nhận H , nghĩa là $p = p_0$.
- Nếu $t > t_\alpha$ thì bác bỏ H , nghĩa là $p \neq p_0$.

Khi đó: $f > p_0 \Rightarrow p > p_0$; $f < p_0 \Rightarrow p < p_0$.

➤ Chương 7. Kiểm định Giả thuyết Thống kê

VD 9. Một báo cáo cho biết có 58% người tiêu dùng Việt Nam quan tâm đến hàng Việt. Khảo sát ngẫu nhiên 1.000 người dân Việt Nam thấy có 536 người được hỏi là có quan tâm đến hàng Việt. Với mức ý nghĩa 5%, hãy kiểm định lại báo cáo trên ?

Giải

Đặt giả thuyết $H : p = p_0 = 0,58$.

Mức ý nghĩa: $\alpha = 0,05 \Rightarrow t_{\alpha} = 1,96$.

Tỉ lệ mẫu là: $f = \frac{536}{1.000} = 0,536$.

➤ Chương 7. Kiểm định Giả thuyết Thống kê

Giá trị thống kê:

$$t = \frac{|f - p_0|}{\sqrt{p_0 q_0}} \sqrt{n} = \frac{|0,536 - 0,58|}{\sqrt{0,58 \times 0,42}} \sqrt{1.000} = 2,8191.$$

Do $t > t_{\alpha}$ nên ta bác bỏ giả thuyết H .

Vì $f < p_0$ nên thực tế là, tỉ lệ người tiêu dùng quan tâm đến hàng Việt thấp hơn 58%.

➤ Chương 7. Kiểm định Giả thuyết Thống kê

VD 10. Khảo sát ngẫu nhiên 400 sinh viên về mức độ nghiêm túc trong giờ học thì thấy 13 sinh viên thừa nhận có ngủ trong giờ học. Trong kiểm định giả thuyết H : “**có 2% sinh viên ngủ trong giờ học**”, mức ý nghĩa tối đa là bao nhiêu để H được chấp nhận ?

Giải

Ta có: $p_0 = 0,02 \Rightarrow q_0 = 0,98$.

Tỉ lệ mẫu: $f = \frac{13}{400} = 0,0325$.

➤ Chương 7. Kiểm định Giả thuyết Thống kê

Giá trị thống kê:

$$t = \frac{|f - p_0|}{\sqrt{p_0 q_0}} \sqrt{n} = \frac{|0,0325 - 0,02|}{\sqrt{0,02 \times 0,98}} \sqrt{400} \approx 1,79.$$

Để giả thuyết H được chấp nhận thì:

$$t \leq t_{\alpha} \Rightarrow \varphi(t_{\alpha}) \geq \varphi(1,79) \Rightarrow \frac{1 - \alpha}{2} = 0,4633.$$

Vậy mức ý nghĩa tối đa là 7,34%.

VD 11. Để kiểm tra một loại súng thể thao, người ta cho bắn 1.000 viên đạn vào 1 tấm bia thấy có 670 viên trúng mục tiêu. Sau đó, người ta cải tiến kỹ thuật và kiểm tra lại thì thấy tỉ lệ trúng của súng lúc này là 70%. Trong kiểm định giả thuyết H : “**tỉ lệ bắn trúng của súng thể thao này là 70%**”, với mức ý nghĩa 3% có giá trị thống kê t và kết luận là:

- A. $t = 2,0702$ và cải tiến kỹ thuật là tốt.
- B. $t = 2,0702$ và cải tiến kỹ thuật là chưa tốt.
- C. $t = 2,0176$ và cải tiến kỹ thuật là tốt.
- D. $t = 2,0176$ và cải tiến kỹ thuật là chưa tốt.

Đáp án đúng là B.

VD 12. Công ty A tuyên bố rằng có 40% người tiêu dùng ưa thích sản phẩm của mình. Một cuộc điều tra 400 người tiêu dùng thấy có 179 người ưa thích sản phẩm của công ty A.

Trong kiểm định giả thuyết H : “**có 40% người tiêu dùng thích sản phẩm của công ty A**”, mức ý nghĩa tối đa để H được chấp nhận là:

A. 7,86%; B. 6,48%; C. 5,24%; D. 4,32%.

Đáp án đúng là C.

2.3. Kiểm định so sánh phương sai với một số (tham khảo)

Giả sử tổng thể có phân phối chuẩn. Với số σ_0^2 cho trước, ta thực hiện các bước sau:

- Đặt giả thuyết $H : \sigma^2 = \sigma_0^2$.
- Từ mẫu ta tính thống kê $\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$.
- Từ $\alpha \Rightarrow \frac{\alpha}{2} \xrightarrow{\text{tra bảng } D} \chi_{n-1}^2 \left(\frac{\alpha}{2} \right), \chi_{n-1}^2 \left(1 - \frac{\alpha}{2} \right)$.

➤ Chương 7. Kiểm định Giả thuyết Thống kê

- Nếu $\chi_{n-1}^2 \left(\frac{\alpha}{2} \right) \leq \chi^2 \leq \chi_{n-1}^2 \left(1 - \frac{\alpha}{2} \right)$ thì ta chấp nhận H , ngược lại thì ta bác bỏ H .

Trong trường hợp bác bỏ:

- nếu $s^2 > \sigma_0^2$ thì ta kết luận $\sigma^2 > \sigma_0^2$;
- nếu $s^2 < \sigma_0^2$ thì ta kết luận $\sigma^2 < \sigma_0^2$.

VD 13. Tiến hành 25 quan sát về chỉ tiêu X của 1 loại sản phẩm (phân phối chuẩn), tính được $s^2 = 416,667$. Có tài liệu nói rằng phương sai của chỉ tiêu X là 400. Với mức ý nghĩa 3%, hãy cho nhận xét về tài liệu này ?

.....

§3. KIỂM ĐỊNH SO SÁNH HAI ĐẶC TRƯNG CỦA HAI TỔNG THỂ

3.1. So sánh hai trung bình của hai tổng thể X, Y

Ta có 4 trường hợp và việc chấp nhận hay bác bỏ H ta đều làm như kiểm định so sánh trung bình với 1 số (cả 4 trường hợp ta đều đặt giả thuyết $H : \mu_x = \mu_y$).

a) Trường hợp 1. $n_x, n_y \geq 30$ và σ_x^2, σ_y^2 đã biết.

Ta tính thống kê $t = \frac{|\bar{x} - \bar{y}|}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}$ và so sánh với t_α .

➤ Chương 7. Kiểm định Giả thuyết Thống kê

b) Trường hợp 2. $n_x, n_y \geq 30$ và σ_x^2, σ_y^2 chưa biết.

Ta thay σ_x^2, σ_y^2 bằng s_x^2, s_y^2 trong trường hợp 1.

c) Trường hợp 3. $n_x, n_y < 30$ và σ_x^2, σ_y^2 đã biết
đồng thời X, Y có phân phối chuẩn.
Ta làm như trường hợp 1.

➤ Chương 7. Kiểm định Giả thuyết Thống kê

d) Trường hợp 4. $n_x, n_y < 30$ và σ_x^2, σ_y^2 chưa biết
đồng thời X, Y có phân phối chuẩn.

- Tính phương sai chung của 2 mẫu:

$$s^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}.$$

- Tính giá trị thống kê $t = \frac{|\bar{x} - \bar{y}|}{s \cdot \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}}$.

- Từ $\alpha \xrightarrow{\text{tra bảng } C} t_{\alpha}^{n_x + n_y - 2}$ và so sánh với t .

➤ Chương 7. Kiểm định Giả thuyết Thống kê

VD 1. Người ta tiến hành bón hai loại phân X, Y cho cây cà chua. Với 60 cây được bón phân X thì thu được trung bình 32,2 quả và độ lệch chuẩn 8,5 quả; 72 cây được bón phân Y thu được trung bình 28,4 quả và độ lệch chuẩn 9,3 quả. Với mức ý nghĩa 5%, hãy cho biết kết luận về hai loại phân bón trên ?

Giải

Ta có: $n_x = 60, \bar{x} = 32,2, s_x = 8,5;$

$$n_y = 72, \bar{y} = 28,4, s_y = 9,3.$$

Bài toán thuộc trường hợp thứ 2.

Đặt giả thuyết $H : \mu_x = \mu_y$ (quả).

➤ Chương 7. Kiểm định Giả thuyết Thống kê

Giá trị thống kê:

$$t = \frac{|\bar{x} - \bar{y}|}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}} = \frac{|32,2 - 28,4|}{\sqrt{\frac{(8,5)^2}{60} + \frac{(9,3)^2}{72}}} = 2,4501.$$

Do $t > t_{\alpha} = 1,96$, $\bar{x} > \bar{y}$ nên ta bác bỏ H và kết luận là khi bón phân X thì cây cà chua cho trái nhiều hơn.

➤ Chương 7. Kiểm định Giả thuyết Thống kê

VD 2. Để so sánh mức lương trung bình của nhân viên nữ X (USD/giờ) và nam Y (USD/giờ) ở một công ty đa quốc gia, người ta tiến hành khảo sát ngẫu nhiên 100 nữ và 75 nam thì có kết quả:

$$\bar{x} = 7,23, s_x = 1,64 \text{ và } \bar{y} = 8,06, s_y = 1,85.$$

Với mức ý nghĩa 3%, kiểm định giả thuyết H : “**mức lương trung bình của nữ và nam ở công ty này là như nhau**” có giá trị thống kê và kết luận là:

- A. $t = 4,0957$, mức lương của nữ và nam như nhau.
- B. $t = 4,0957$, mức lương của nữ thấp hơn nam.
- C. $t = 3,0819$, mức lương của nữ và nam như nhau.
- D. $t = 3,0819$, mức lương của nữ thấp hơn nam.

Đáp án đúng là D.

VD 3. Tuổi thọ (năm) của pin là biến ngẫu nhiên có phân phối chuẩn. Một công ty sản xuất thử nghiệm 10 chiếc pin loại X và 12 chiếc pin loại Y thì có kết quả:

$$\bar{x} = 4,8, s_x = 1,1 \text{ và } \bar{y} = 4,3, s_y = 0,3.$$

Với mức ý nghĩa 1%, ta có thể kết luận tuổi thọ của loại pin X cao hơn loại pin Y được không ?

Giải

Bài toán thuộc trường hợp thứ 4.

Đặt giả thuyết $H : \mu_x = \mu_y$ (năm).

➤ Chương 7. Kiểm định Giả thuyết Thống kê

Ta có:

$$s^2 = \frac{(10 - 1)(1,1)^2 + (12 - 1)(0,3)^2}{10 + 12 - 2} = 0,594.$$

$$\alpha = 0,01 \xrightarrow{C} t_{0,01}^{20} = 2,845.$$

$$t = \frac{|\bar{x} - \bar{y}|}{\sqrt{s^2 \left(\frac{1}{n_x} + \frac{1}{n_y} \right)}} = \frac{|4,8 - 4,3|}{\sqrt{0,594 \left(\frac{1}{10} + \frac{1}{12} \right)}} = 1,5152.$$

Do $t < t_{0,01}^{20}$ nên tuổi thọ của hai loại pin là như nhau.

➤ Chương 7. Kiểm định Giả thuyết Thống kê

VD 4. Tuổi thọ (tháng) của thiết bị là biến ngẫu nhiên có phân phối chuẩn. Người ta kiểm tra ngẫu nhiên tuổi thọ của 15 thiết bị loại A , có kết quả:

114; 78; 96; 137; 78; 103; 126; 86;
99; 114; 72; 104; 73; 86; 117.

Kiểm tra tuổi thọ của 17 thiết bị loại B thấy có trung bình là 84 tháng và độ lệch chuẩn là 19 tháng.

Kiểm định giả thuyết H : “**tuổi thọ của thiết bị loại A và B là như nhau với mức ý nghĩa 3%**” có giá trị thống kê và kết luận là:

- A.** $t = 2,1616$; tuổi thọ của hai loại thiết bị là như nhau.
- B. $t = 2,1616$; tuổi thọ của loại thiết bị A lớn hơn.
- C. $t = 2,4616$; tuổi thọ của hai loại thiết bị là như nhau.
- D. $t = 2,4616$; tuổi thọ của loại thiết bị A lớn hơn.

3.2. So sánh hai tỉ lệ của hai tổng thể X, Y

Ta thực hiện các bước sau:

- Đặt giả thuyết $H : p_x = p_y$.

- Từ 2 mẫu ta tính $f_x = \frac{m_x}{n_x}$, $f_y = \frac{m_y}{n_y}$, $p_0 = \frac{m_x + m_y}{n_x + n_y}$.

- Tính giá trị thống kê $t = \frac{|f_x - f_y|}{\sqrt{p_0 q_0 \left(\frac{1}{n_x} + \frac{1}{n_y} \right)}}$.

➤ Chương 7. Kiểm định Giả thuyết Thống kê

- Kết luận:

- Nếu $t \leq t_{\alpha}$ thì ta chấp nhận $H \Rightarrow p_x = p_y$.
- Nếu $t > t_{\alpha}$ và $f_x < f_y$ thì ta bác bỏ $H \Rightarrow p_x < p_y$.
- Nếu $t > t_{\alpha}$ và $f_x > f_y$ thì ta bác bỏ $H \Rightarrow p_x > p_y$.

➤ Chương 7. Kiểm định Giả thuyết Thống kê

VD 5. Từ hai tổng thể X và Y người ta tiến hành kiểm tra 2 mẫu có kích thước $n_x = 1000$, $n_y = 1200$ về một tính chất A thì được $f_x = 0,27$ và $f_y = 0,3$. Với mức ý nghĩa 9%, hãy so sánh hai tỉ lệ p_x , p_y của hai tổng thể ?

Giải

Đặt giả thuyết $H : p_x = p_y$ (tỉ lệ phần tử có tính chất A ở hai tổng thể là như nhau).

Ta có: $m_x = f_x \cdot n_x = 270$, $m_y = f_y \cdot n_y = 360$

$$\Rightarrow p_0 = \frac{270 + 360}{1000 + 1200} = 0,2864 \Rightarrow q_0 = 0,7136.$$

➤ Chương 7. Kiểm định Giả thuyết Thống kê

$$\alpha = 0,09 \Rightarrow \frac{1 - \alpha}{2} = 0,455 \Rightarrow t_{\alpha} = 1,69.$$

$$t = \frac{|0,27 - 0,3|}{\sqrt{0,2864 \cdot 0,7136 \left(\frac{1}{1000} + \frac{1}{1200} \right)}} = 1,5498 < t_{\alpha}.$$

Vậy tỉ lệ p_x, p_y của hai tổng thể là như nhau.

➤ Chương 7. Kiểm định Giả thuyết Thống kê

VD 6. Kiểm tra 120 sản phẩm ở kho I thấy có 6 phế phẩm; 200 sản phẩm ở kho II thấy có 24 phế phẩm. Hỏi chất lượng hàng ở hai kho có khác nhau không với:

1) mức ý nghĩa 5%; 2) mức ý nghĩa 1%.

Giải

Đặt giả thuyết $H : p_1 = p_2$ (tỉ lệ phế phẩm ở hai kho là như nhau).
Ta có:

$$f_1 = 0,05; f_2 = 0,12; p_0 = 0,0938; q_0 = 0,9062.$$

$$t = \frac{|0,05 - 0,12|}{\sqrt{0,0938 \cdot 0,9062 \left(\frac{1}{120} + \frac{1}{200} \right)}} = 2,0793.$$

Kết luận:

- 1) $t_{0,05} = 1,96 < t, f_1 < f_2$: bác bỏ H và kho I tốt hơn kho II.
- 2) $t_{0,01} = 2,58 > t$: chấp nhận H , chất lượng hai kho là như nhau.

VD 7. Một công ty điện tử tiến hành điều tra thị trường về sở thích xem tivi của cư dân trong 1 thành phố. Điều tra ngẫu nhiên 400 người ở quận X thì thấy có 270 người xem tivi ít nhất 1 giờ trong 1 ngày; 600 người ở quận Y có 450 người xem tivi ít nhất 1 giờ trong 1 ngày. Trong kiểm định giả thuyết H : “**tỉ lệ cư dân xem tivi ít nhất 1 giờ trong 1 ngày ở quận X và Y như nhau**”, mức ý nghĩa tối đa để H được chấp nhận là:
A. 0,96%; B. 2,84%; C. 4,06%; D. 6,14%.

Đáp án đúng: A.

➤ Chương 7. Kiểm định Giả thuyết Thống kê

VD 8. Trước bầu cử, người ta thăm dò 1000 cử tri thì thấy có 400 người nói rằng sẽ bỏ phiếu cho ông A. Một tuần sau (vẫn chưa bầu cử), người ta tổ chức 1 cuộc thăm dò khác và thấy có 680 trong số 1500 cử tri được hỏi sẽ bỏ phiếu cho ông A. Kiểm định giả thuyết H : “**tỉ lệ cử tri ủng hộ ông A ở hai lần là như nhau**”, với mức ý nghĩa 1% có giá trị thống kê và kết luận là:

- A. $t = 2,6356$; cử tri ngày càng ủng hộ ông A.
- B. $t = 2,6356$; cử tri ủng hộ ông A không thay đổi.
- C. $t = 2,1349$; cử tri ngày càng ủng hộ ông A.
- D. $t = 2,1349$; cử tri ủng hộ ông A không thay đổi.

Đáp án đúng: A.

3.3. So sánh hai phương sai của hai tổng thể X, Y (Tham khảo)

- Đặt giả thuyết $H : \sigma_x^2 = \sigma_y^2$.
- Tính giá trị kiểm định $g = \frac{s_x^2}{s_y^2}$.
- Từ $\alpha \Rightarrow \frac{\alpha}{2} \xrightarrow{\text{tra bảng } E} f = f_{\frac{\alpha}{2}}(n_x - 1, n_y - 1)$.
- Nếu $g \leq f$ thì ta chấp nhận H , nếu $g > f$ ta bác bỏ H .
Trong trường hợp bác bỏ H :
 - nếu $s_x^2 > s_y^2$ thì kết luận $\sigma_x^2 > \sigma_y^2$;
 - nếu $s_x^2 < s_y^2$ thì kết luận $\sigma_x^2 < \sigma_y^2$.

➤ Chương 7. Kiểm định Giả thuyết Thống kê

VD 9. Giá cổ phiếu là biến ngẫu nhiên có phân phối chuẩn. Điều tra ngẫu nhiên giá cổ phiếu của công ty X trong 25 ngày người ta tính được độ lệch tiêu chuẩn mẫu hiệu chỉnh là 7,5 ngàn đồng; của công ty Y trong 22 ngày là 6,2 ngàn đồng. Với mức ý nghĩa 5%, hãy so sánh về độ rủi ro cổ phiếu của hai công ty trên ?

VD 10. Doanh số bán hàng (đơn vị: triệu đồng) của công ty A là biến ngẫu nhiên có phân phối chuẩn. Công ty A cho người theo dõi doanh số bán hàng trong 7 ngày ở vùng X thì tính được phương sai mẫu chưa hiệu chỉnh là 82,1; ở vùng Y trong 6 ngày thì tính được 25,3. Với mức ý nghĩa 3%, hãy so sánh độ rủi ro đầu tư của công ty A ở hai vùng trên ?

3.4. So sánh hai trung bình ở dạng vector (X, Y) (Tham khảo)

- Đặt $d = Y - X$ và giả thuyết $H : \mu_d = 0$.
- Tính thống kê $t = \frac{|\bar{d}|}{s_d} \sqrt{n}$ (n là số cặp có trong mẫu).
- Tùy vào n và phương sai đã biết hay chưa biết, ta xét các trường hợp giống như so sánh trung bình với 1 số.

➤ Chương 7. Kiểm định Giả thuyết Thống kê

VD 11. Giả sử người ta dùng thuốc A cho 10 người. Đo nhịp tim/phút trước và sau khi dùng thuốc của từng người, có bảng kết quả:

Người	1	2	3	4	5	6	7	8	9	10
Trước: X	70	77	78	72	81	78	73	74	79	80
Sau: Y	76	75	78	77	85	81	76	74	85	80
d	6	-2	0	5	4	3	3	0	6	0

Với mức ý nghĩa 5%, thuốc A có làm thay đổi nhịp tim trước khi dùng so với sau khi dùng hay không ?

Giải

Đặt $d = Y - X$ và giả thuyết $H : \mu_d = 0$.

Do $n = 10$, phương sai chưa biết nên bài toán là TH4.

➤ Chương 7. Kiểm định Giả thuyết Thống kê

Từ bảng số liệu, ta tính được: $\bar{d} = 2,5$; $s_d = 2,8382$.

Mức ý nghĩa $\alpha = 0,05 \xrightarrow{C} t_{0,05}^9 = 2,262$.

$$\text{Thống kê } t = \frac{|\bar{d}|}{s_d} \cdot \sqrt{n} = \frac{2,5}{2,8382} \cdot \sqrt{10} = 2,7855.$$

Vì $t > t_{0,05}^9$ nên ta bác bỏ H .

Vậy thuốc A đã làm thay đổi nhịp tim.
