BÀI 7
KIỂM ĐỊNH GIẢ THUYẾT THỐNG KÊ

TS. Nguyễn Mạnh Thế

v1.0012107210

TÌNH HUỐNG KHỞI ĐỘNG BÀI

Tình huống

Công ty Hoàng Lâm sản xuất mỳ chính theo dây chuyền của Đức. Theo tiêu chuẩn thì trọng lượng các gói mỳ chính được đóng trên một máy tự động là 453g. Nghi ngờ máy tự động làm việc không còn đủ chính xác, công ty Hoàng Lâm tiến hành kiểm tra ngẫu nhiên 81 gói ta thấy trọng lượng trung bình là 448g. Với mức ý nghĩa 0.05 có thể cho rằng trọng lượng các gói mỳ chính không đạt tiêu chuẩn hay không? Biết rằng trọng lượng gói mỳ chính là biến ngẫu nhiên có phân phối chuẩn với độ lệch chuẩn là 36g.

Tiêu chuẩn kiểm định:

Câu hỏi gợi mở

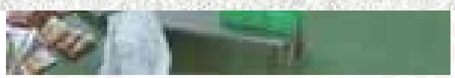
Câu 1: Trọng lượng trung bình của 01 gói mỳ chính theo điều tra là bao nhiêu?

Câu 2: Để bác bỏ giả thuyết "dây chuyền vẫn hoạt động tốt, trọng lượng mỳ chính đúng tiêu chuẩn" thì tiêu chuẩn kiểm định phải không nằm trong khoảng nào?

Câu 3: Dây chuyền còn hoạt động tốt không?

tiêu chuẩn đối với
phối chuẩn.
):

biến X thực sự $\neq \mu_0$



TÌNH HUỐNG KHỞI ĐỘNG BÀI (tiếp theo)

Kết luận

- Kiểm định so sánh kỳ vọng của biến ngẫu nhiên có phân phối chuẩn (với một giá trị cho trước của kỳ vọng).

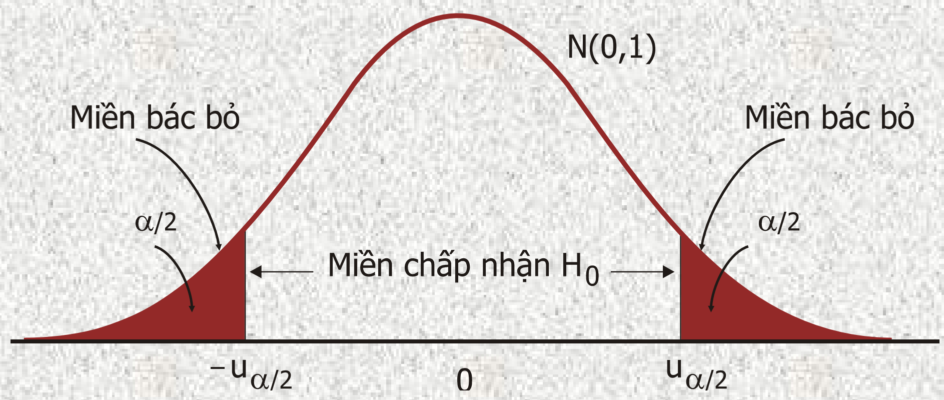
Trường hợp σ^2 đã biết:

Bài toán 1:

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$$

Tiêu chuẩn kiểm định:

$$U = \frac{(\overline{X} - \mu_0)}{\sigma}\sqrt{n} \sim N(0,1)$$



Hình 1. Miền tiêu chuẩn đối với phân phối chuẩn.

- Với mức ý nghĩa a cho trước, ta có miền bác bỏ:

$\left(-\infty; -U_{\alpha/2}\right) \cup \left(U_{\alpha/2}; +\infty\right)$ với $P\left\{|U| > U_{\alpha/2}\right\} = \alpha$

Nếu giá trị $U_{qs} = \dfrac{\overline{x} - \mu_0}{\sigma}\sqrt{n} \in W \Rightarrow$ kỳ vọng của biến X thực sự $\neq \mu_0$

MỤC TIÊU

Khái niệm giả thuyết thống kê	• Miền bác bỏ; • Các bước làm bài toán kiểm định.
Kiểm định tham số	• Tham số kỳ vọng; • Tham số phương sai; • Tham số tỉ lệ.
Một số tiêu chuẩn kiểm định phi tham số	• Kiểm định giả thuyết phân phối; • So sánh nhiều tỉ lệ; • Kiểm định tính độc lập.

1. KHÁI NIỆM GIẢ THUYẾT THỐNG KÊ

Khái niệm:

Giả thuyết thống kê là một mệnh đề về tham số của tổng thể.

Ký hiệu H_0 là giả thuyết của tham số tổng thể, đi kèm với giả thuyết là mệnh đề đối lập được gọi là đối thuyết, ký hiệu là H_1.

Bài toán kiểm định giả thuyết thống kê gồm một cặp giả thuyết H_0 và đối thuyết H_1.

- Sai lầm loại I: Bác bỏ giả thuyết H_0 nhưng thực tế H_0 là đúng.
  α là xác suất mắc sai lầm loại 1.

- Sai lầm loại II: Chấp nhận giả thuyết H_0 nhưng thực tế H_0 là sai.
  β là xác suất mắc sai lầm loại 2.

α được gọi là mức ý nghĩa, thường được lấy nhỏ: 0,05; 0,02; 0,01.

1.1. MIỀN BÁC BỎ

Để giải quyết bài toán kiểm định giả thuyết ta xây dựng một thống kê.

G gọi là tiêu chuẩn thống kê.

Định nghĩa 1: Thống kê $T = G(X_1 + X_2 + ... + X_n)$ được gọi là một tiêu chuẩn thống kê nếu giá trị của nó được dùng để xem xét bác bỏ hay chấp nhận giả thuyết H_0.

Định nghĩa 2: Miền này được dùng cùng với tiêu chuẩn thống kê T và giá trị cụ thể t_{qs} của tiêu chuẩn đó để đưa ra kết luận về giả thuyết H_0.

$t_{qs} \in W$ bác bỏ giả thuyết H_0.

$t_{qs} \in W^c$ chấp nhận giả thuyết H_0.

1.2. CÁC BƯỚC LÀM BÀI TOÁN KIỂM ĐỊNH

Bước 1	Xác định tham số kiểm định, đặt giả thuyết và đối thuyết
Bước 2	Xác định tiêu chuẩn và giá trị tiêu chuẩn với giá trị mẫu đã cho
Bước 3	Xác định miền bác bỏ W
Bước 4	So sánh giá trị của tiêu chuẩn thống kê với miền bác bỏ, kết luận bác bỏ hay chấp nhận giả thuyết

2. KIỂM ĐỊNH THAM SỐ

Kiểm định bằng miền tiêu chuẩn (miền bác bỏ)

Kiểm định bằng xác suất ý nghĩa

Kiểm định bằng khoảng tin cậy

- Kiểm định giả thuyết kì vọng
- Kiểm định giả thuyết phương sai
- Kiểm định giả thuyết xác suất

Kiểm định giả thuyết hai phía

Kiểm định giả thuyết một phía

Ta coi tất cả các biến ngẫu nhiên được xét tới đều có phân phối chuẩn.

2.1. KIỂM ĐỊNH GIẢ THUYẾT KỲ VỌNG

Trường hợp σ đã biết

Bài toán 1.

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$$

Nếu giả thuyết H_0 là đúng thì:

$$U = \frac{(\overline{X} - \mu_0)}{\sigma}\sqrt{n} \sim N(0,1)$$

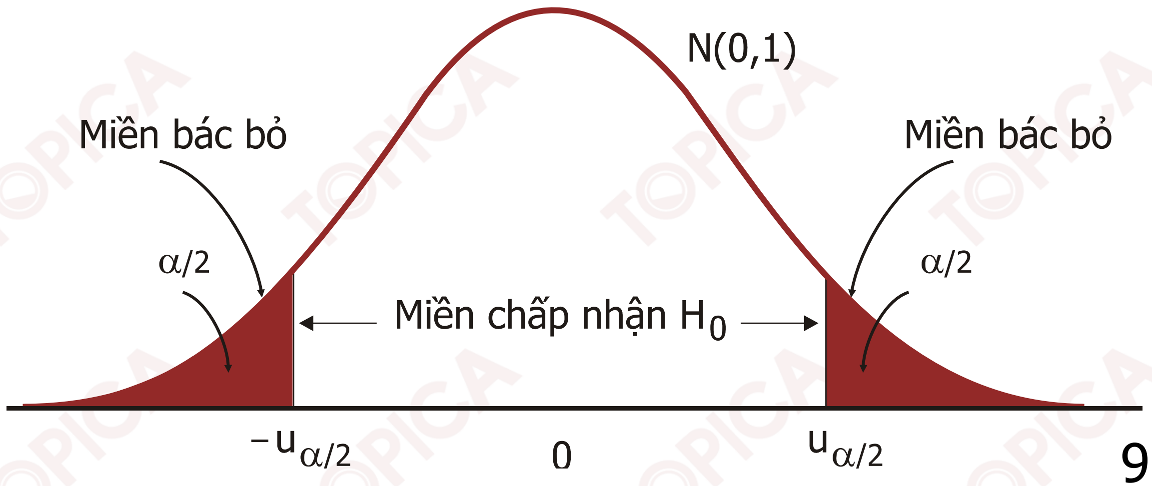
Giả thuyết H_0 bị bác bỏ nếu $P\{|U| > u_{\alpha/2}\} = \alpha$

Ta có miền bác bỏ $W = (-\infty;\ -u_{\alpha/2}) \cup (u_{\alpha/2};\ +\infty)$

Trong đó $u_{\alpha/2}$ thỏa mãn điều kiện: $\Phi_0(u_{\alpha/2}) = 1 - \alpha/2$

Với mẫu cụ thể giá trị của tiêu chuẩn thống kê U là:

$$u_{qs} = \frac{\overline{x} - \mu_0}{\sigma}\sqrt{n}$$



2.1. KIỂM ĐỊNH GIẢ THUYẾT KỲ VỌNG (tiếp theo)

Ví dụ:

Giải: Gọi X là trọng lượng gói mì chính

Giả thiết cần kiểm định $\begin{cases} H_0 : \mu = 453 \\ H_1 : \mu \neq 453 \end{cases}$

Ta có X ~ N(μ,σ²) trong đó σ=36, với mức ý nghĩa α=0,05.

Tra bảng phân phối chuẩn ta tính được:

$u_{\alpha/2} = u_{0,025} = 1,96$

Vậy miền bác bỏ là: $W = (-\infty;\ -1,96) \cup (1,96;\ +\infty)$

Ta có:

$$\overline{x} = 448 \Rightarrow u_{qs} = \frac{448 - 453}{36}\sqrt{81} = -1,25$$

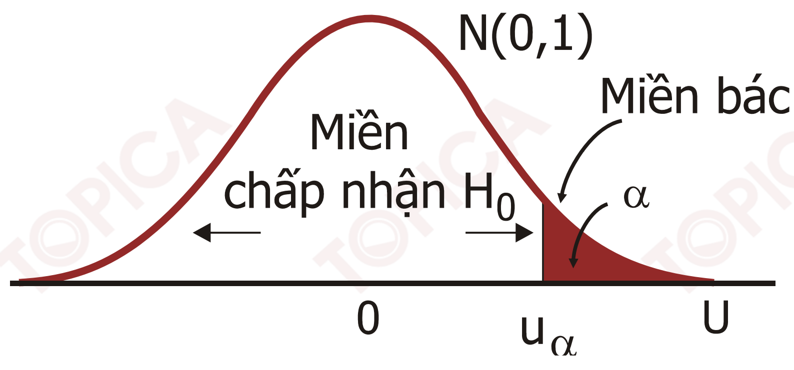
Kiểm tra ta thấy $u_{qs} = -1,25 \notin W$.

Vậy ta chấp nhận giả thuyết H_0, tức là trọng lượng các gói mì chính không đạt tiêu chuẩn .

2.1. KIỂM ĐỊNH GIẢ THUYẾT KỲ VỌNG (tiếp theo)

Bài toán kiểm định một phía

Bài toán 2 $\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu > \mu_0 \end{cases}$



Để bác bỏ giả thuyết H_0 thì giá trị của thống kê U phải đủ lớn để

$$P\{U > u_\alpha\} = \alpha$$

Trong đó u thỏa mãn: $\Phi_0(u_\alpha) = 1 - \alpha$

Ta có miền bác bỏ: $W = (u_\alpha ; +\infty)$.

2.1. KIỂM ĐỊNH GIẢ THUYẾT KỲ VỌNG (tiếp theo)

Ví dụ: Năng suất trung bình của một giống lúa ở các năm trước là 32,5 (tạ/ha). Năm nay người ta đưa vào phương pháp chăm sóc mới và hy vọng năng suất cao hơn năm trước. Điều tra trên 15 thửa ruộng thu được kết quả sau:

33,7	35,4	32,7	36,3	37,3	32,4	30,0		
32,4	31,7	34,5	42,0	33,9	38,1	35,0	33,8	(tạ/ha)

Với mức ý nghĩa 1% có thể chấp nhận hy vọng đó hay không? Biết rằng năng suất lúa là biến ngẫu nhiên có phân phối chuẩn với phương sai là 10.

Giải:

Gọi X là năng suất lúa, ta có X ~ N(μ,σ²) trong đó σ²=10.

Ta cần kiểm định giả thiết: $\begin{cases} H_0 : \mu = 32,5 \\ H_1 : \mu > 32,5 \end{cases}$

Với mức ý nghĩa: $\alpha = 0,01$ nên $\phi_0 = 1 - 0,01 = 0,99$

Tra bảng phân phối chuẩn ta có $u_{0,01} = 2,33$

Vậy miền bác bỏ là: $W = (2,33; +\infty)$

Với mẫu cụ thể đã cho ta tính được $\overline{x} = 34,613$

Ta có: $u_{qs} = \dfrac{\overline{x} - \mu_0}{\sigma}\sqrt{n} = \dfrac{34,613 - 32,5}{\sqrt{10}}\sqrt{15} = 2,587$

$$u_{qs} = 2,587 \in W$$

Vậy bác bỏ giả thuyết H_0 tức là năng suất lúa đã tăng lên.

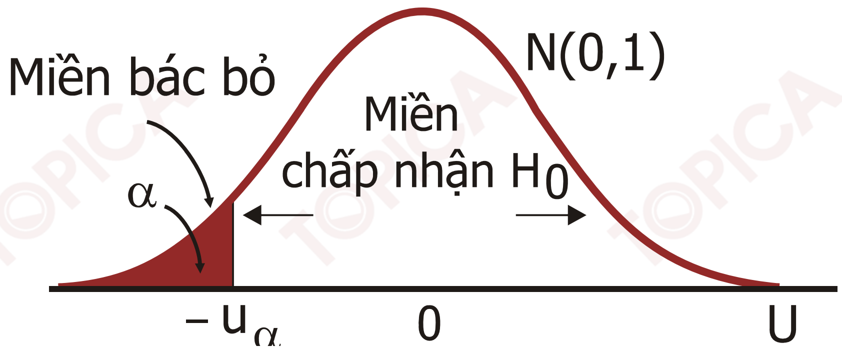
2.1. KIỂM ĐỊNH GIẢ THUYẾT KỲ VỌNG (tiếp theo)

Bài toán 3:

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu < \mu_0 \end{cases}$$

Để bác bỏ giả thuyết H_0 thì giá trị của thống kê U phải đủ nhỏ, tồn tại giá trị u_α sao cho: $P\{U < -u_\alpha\} = \alpha$

Miền bác bỏ: $W = (-\infty; -u_\alpha)$



2.1. KIỂM ĐỊNH GIẢ THUYẾT KỲ VỌNG (tiếp theo)

Ví dụ:

Theo tiêu chuẩn thì trọng lượng các bao gạo do một máy tự động đóng là 50kg. Sau một thời gian hoạt động người ta nghi ngờ máy hoạt động không bình thường làm cho trọng lượng các bao gạo giảm đi.

Lấy ngẫu nhiên 90 bao và cân thử thu được trọng lượng trung bình là 48,5kg.

Với mức ý nghĩa 5%, có thể kết luận về điều nghi ngờ trên?

Biết rằng trong lượng bao gạo là biến ngẫu nhiên có phân phối chuẩn với độ lệch chuẩn là 2kg.

2.1. KIỂM ĐỊNH GIẢ THUYẾT KỲ VỌNG (tiếp theo)

Giải: Gọi X là trọng lượng bao gạo X ~ $N(\mu,\sigma^2)$ và $\sigma = 2$

Ta cần kiểm định $\begin{cases} H_0: \mu = 50 \\ H_1: \mu < 50 \end{cases}$

mức ý nghĩa: $\alpha = 0,05 \Rightarrow \Phi_0(u_{0,05}) = 1 - 0,05 = 0,95.$

Tra bảng phân phối chuẩn ta có $u_{0,05} = 1,65$

Vậy miền bác bỏ là: $W = (-\infty; -1,65).$

Với mẫu đã cho ta có: $\overline{x} = 48,5$

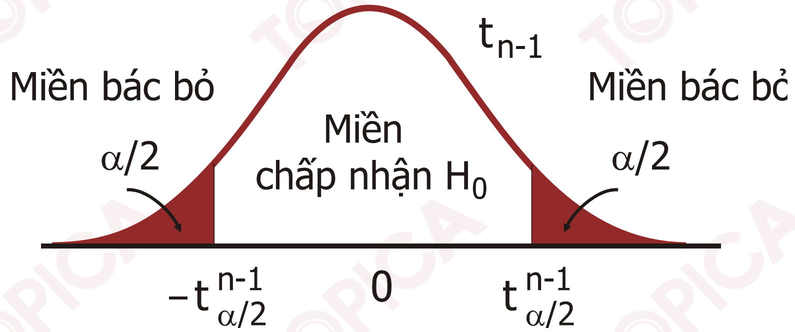
Giá trị của tiêu chuẩn thống kê là: $u_{qs} = \dfrac{\overline{x} - \mu_0}{\sigma}\sqrt{n} = \dfrac{48,5 - 50}{2}\sqrt{90} = -7,11$

Kiểm tra ta thấy $u_{qs} \notin W$ vậy ta bác bỏ giả thuyết H_0.

1. KIỂM ĐỊNH GIẢ THUYẾT KỲ VỌNG (tiếp theo)

Trường hợp σ chưa biết

Xét thống kê $T = \dfrac{\overline{X} - \mu_0}{S'}\sqrt{n}$, nếu giả thuyết H_0 là đúng thì T có quy luật phân phối student với n-1 bậc tự do.



Bài toán 1 $\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$

Ta bác bỏ giả thuyết H_0 nếu giá trị tuyệt đối của thống kê T đủ lớn,

tức là: $P\left\{|T| > t^{n-1}_{\alpha/2}\right\} = \alpha$

trong đó phân vị $t^{n-1}_{\alpha/2}$ tìm từ bảng phân phối student.

miền bác bỏ là: $W = (-\infty;\ -t^{n-1}_{\alpha/2}) \cup (t^{n-1}_{\alpha/2};\ +\infty)$

với mẫu cụ thể giá trị của tiêu chuẩn thống kê

$$t_{qs} = \frac{\overline{x} - \mu_0}{s'}\sqrt{n}$$

2.1. KIỂM ĐỊNH GIẢ THUYẾT KỲ VỌNG (tiếp theo)

Ví dụ:

Trong các năm trước thu nhập trung bình của công nhân là 15 (triệu/năm), năm nay điều tra thu nhập của 25 công nhân ta có số liệu sau:

Thu nhập	10-12	12-14	14-16	16-18	18-20
Số công nhân	2	4	10	6	3

Với mức ý nghĩa 5% hãy kiểm định xem thu nhập trung bình của công nhân năm nay có khác so với năm trước hay không? Biết rằng thu nhập của công nhân là biến ngẫu nhiên có phân phối chuẩn.

2.1. KIỂM ĐỊNH GIẢ THUYẾT KỲ VỌNG (tiếp theo)

Giải: Gọi X là thu nhập của công nhân $X \sim N(\mu,\sigma^2)$

Ta kiểm định $\begin{cases} H_0 : \mu = 15 \\ H_1 : \mu \neq 15 \end{cases}$

Ta có:

$$\overline{x} = 15,32,\ s'^2 = 4,893,\ s' = 2,212$$

Vậy ta có miền bác bỏ: $W = (-\infty; -2,06) \cup (2,06; +\infty)$.

Với mẫu đã cho ta có: $t^{24}_{0,025} = 2,06$.

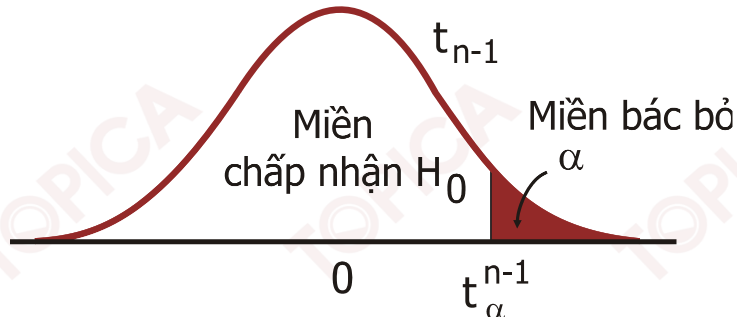
Giá trị tiêu chuẩn thống kê: $t_{qs} = \dfrac{15,32 - 15}{2,212}\sqrt{25} = 0,723$.

Ta thấy $t_{qs} \notin W$, do đó chưa bác bỏ được giả thuyết H_0.

2.1. KIỂM ĐỊNH GIẢ THUYẾT KỲ VỌNG (tiếp theo)

Bài toán 2: $\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu > \mu_0 \end{cases}$

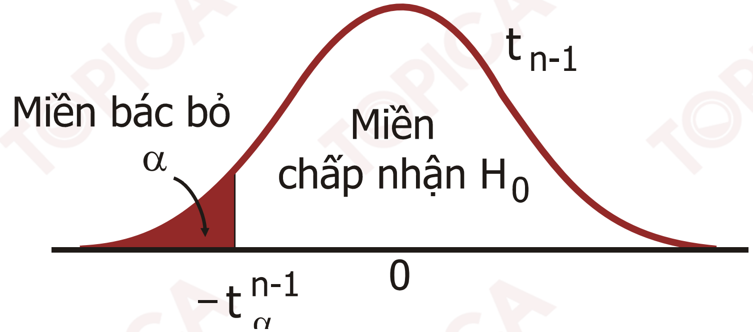
Miền bác bỏ: $W = (t_\alpha^{n-1}; +\infty)$



Bài toán 3: $\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu < \mu_0 \end{cases}$

Tương tự ta có miền bác bỏ

$W = (-\infty;\ -t_\alpha^{n-1})$



phân vị t_α^{n-1} tìm từ bảng phân phối student.

2.1. KIỂM ĐỊNH GIẢ THUYẾT KỲ VỌNG (tiếp theo)

Ví dụ:

Mức xăng hao phí cho một xe ôtô chạy trên đoạn đường AB ở năm trước là 50 lít. Năm nay do đoạn đường bị xuống cấp và mức xăng hao phí tăng lên. Điều tra 30 chuyến xe chạy trên đoạn đường AB ta có số liệu sau:

Mức xăng hao phí	49-49,5	49,5-50	50-50,5	50,5-51	51-51,5
Số chuyến	5	7	10	6	2

Với mức ý nghĩa 1% hãy kết luận về điều nghi ngờ trên. Biết rằng mức xăng hao phí là biến ngẫu nhiên có phân phối chuẩn.

Giải: Gọi X là mức xăng hao phí, X ~ N(μ,σ²)

Ta kiểm định
$$\begin{cases} H_0: \mu = 50 \\ H_1: \mu > 50 \end{cases}$$

tra bảng phân phối student ta có $t_{0,01}^{29} = 2,46.$

Với mẫu cụ thể đã cho, tính toán ta được:

$\bar{x} = 50,133;\ s'^2 = 0,339;\ s' = 0,583.$

Giá trị của tiêu chuẩn thống kê: $t_{qs} = \dfrac{50,133 - 50}{0,538}\sqrt{30} = 1,254$

Ta thấy $t_{qs} \notin W$ do đó chưa có cơ sở để bác bỏ giả thuyết H_0.

2.2. KIỂM ĐỊNH GIẢ THUYẾT PHƯƠNG SAI

Cho biến ngẫu nhiên X ~ N(μ,σ²). Giả thuyết $H_0 : \sigma^2 = \sigma_0^2$.

Trường hợp kỳ vọng μ đã biết

Xét thông kê $\chi^2 = \dfrac{nS^{*2}}{\sigma_0^2}$, trong đó $S^{*2} = \dfrac{1}{n}\sum(X_i - \mu)^2$

Khi đó χ^2 có phân phối khi bình phương với n bậc tự do.

Với mẫu cụ thể ta có : $\chi^2_{qs} = \dfrac{ns^{*2}}{\sigma_0^2}$.

Bài toán 1: $\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 \neq \sigma_0^2 \end{cases}$

Mức ý nghĩa cho trước ta có miền bác bỏ: $W = (0;\ \chi^2_{1-\alpha/2,n}) \cup (\chi^2_{\alpha/2,n};+\infty)$

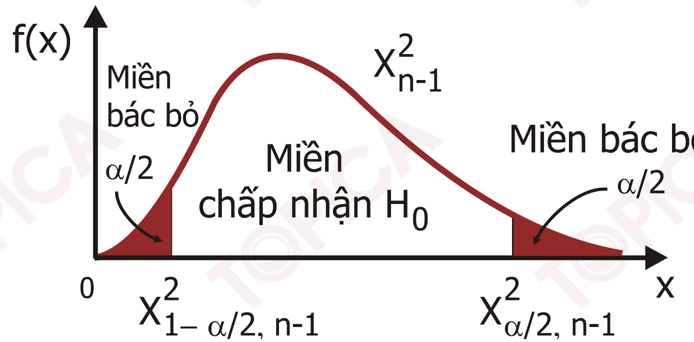
Bài toán 2: $\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 > \sigma_0^2 \end{cases}$ Miền bác bỏ: $W = (\chi^2_{\alpha,n};+\infty)$

Bài toán 3: $\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 < \sigma_0^2 \end{cases}$ Miền bác bỏ: $W = (0;\ \chi^2_{1-\alpha,n})$

2.2. KIỂM ĐỊNH GIẢ THUYẾT PHƯƠNG SAI (tiếp theo)

Trường hợp kỳ vọng μ chưa biết

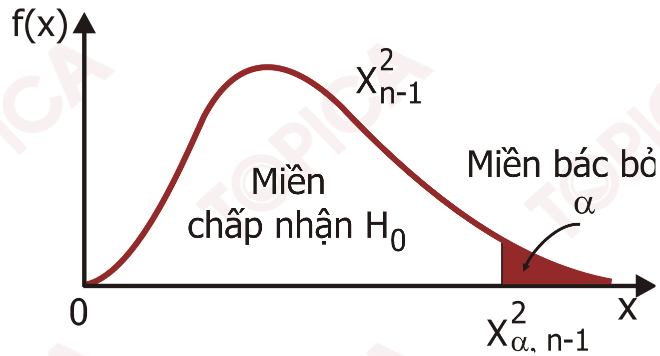
Ta có thống kê: $$\chi^2 = \frac{(n-1)S'^2}{\sigma_0^2} \sim \chi^2(n-1)$$



Với mẫu cụ thể giá trị của tiêu chuẩn thống kê là: $$\chi^2_{qs} = \frac{(n-1)s'^2}{\sigma_0^2}$$

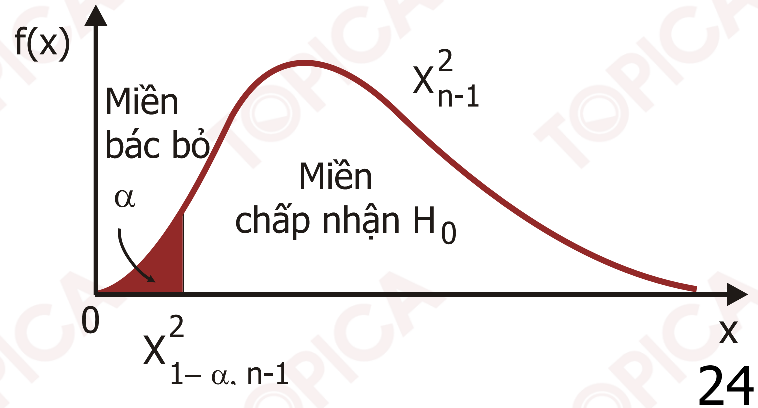
Bài toán 1 $\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 \neq \sigma_0^2 \end{cases}$

Miền bác bỏ: $W = (0;\ \chi^2_{1-\alpha/2,n-1}) \cup (\chi^2_{\alpha/2,n-1};\ +\infty)$



Bài toán 2 $\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 > \sigma_0^2 \end{cases}$

Miền bác bỏ: $W = (\chi^2_{\alpha,n-1};\ +\infty)$



Bài toán 3 $\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 < \sigma_0^2 \end{cases}$

Miền bác bỏ: $W = (0;\ \chi^2_{1-\alpha,n-1})$

2.2. KIỂM ĐỊNH GIẢ THUYẾT PHƯƠNG SAI (tiếp theo)

Ví dụ:

Lấy ngẫu nhiên 20 chai nước do một máy đóng chai tự động đóng ta thu được độ lệch chuẩn mẫu là $s'^2 = 0{,}0153$ (l^2). Máy được gọi là đạt chuẩn nếu độ phân tán không sai khác 0,01(l^2).

Với mức ý nghĩa 5% hãy kiểm định xem máy đóng chai có đạt chuẩn hay không? Biết rằng thể tích nước trong chai là biến ngẫu nhiên có phân phối chuẩn.

2.2. KIỂM ĐỊNH GIẢ THUYẾT PHƯƠNG SAI (tiếp theo)

Giải: Gọi X là thể tích nước trong chai, ta có X ~ N(μ,σ²)

Ta cần kiểm định $\begin{cases} H_0 : \sigma^2 = 0,01 \\ H_1 : \sigma^2 \neq 0,01 \end{cases}$

Mức ý nghĩa α =0,05

Tra bảng phân phối khi bình phương ta có: $\chi^2_{0,025,19} = 32,9,\ \chi^2_{0,975,19} = 8,91.$

Miền bác bỏ: $W = (0;\ \ 8,91) \cup (32,9;\ \ +\infty).$

Ta có: $s'^2 = 0,0153(l^2).$
do đó giá trị tiêu chuẩn thống kê: $\chi_{qs} = \dfrac{(19-1)0,0153}{0,01} = 29,07.$

Kiểm tra ta thấy $\chi_{qs} \notin W$. Vậy ta chưa có cơ sở bác bỏ giả thuyết H_0 tức là máy đóng chai vẫn đạt chuẩn.

2.3. KIỂM ĐỊNH GIẢ THUYẾT CHO XÁC SUẤT (TỶ LỆ)

Cho biến ngẫu nhiên X~A(p). Với p chưa biết.

Lấy mẫu cỡ n từ biến ngẫu nhiên y, gọi m là số lần mẫu nhận giá trị 1.

Ta có $f = \frac{m}{n}$ là tần suất xuất hiện biến cố 1. Ta đưa ra giả thuyết H_0: $p = p_0$.

Xét thống kê $U = \frac{(f - p_0)}{\sqrt{p_0(1 - p_0)}}\sqrt{n}$, khi đó U ~ N(0;1).

Với mẫu cụ thể ta có giá trị của thống kê U là: u_{ps}

Bài toán 1 $\begin{cases} H_0 : p = p_0 \\ H_1 : p \neq p_0 \end{cases}$ Miền bác bỏ: $W = (-\infty; -u_{\alpha/2}) \cup (u_{\alpha/2}; +\infty)$

với $\Phi_0(u_{\alpha/2}) = 1 - \alpha/2$

Bài toán 2 $\begin{cases} H_0 : p = p_0 \\ H_1 : p > p_0 \end{cases}$ Miền bác bỏ: $W = (u_\alpha; +\infty)$, với $\Phi_0(u_\alpha) = 1 - \alpha$

Bài toán 3 $\begin{cases} H_0 : p = p_0 \\ H_1 : p < p_0 \end{cases}$ Miền bác bỏ: $W = (-\infty; u_\alpha)$.

2.3. KIỂM ĐỊNH GIẢ THUYẾT CHO XÁC SUẤT (TỶ LỆ) (tiếp theo)

Ví dụ:

Những năm trước nhà máy áp dụng công nghệ A sản xuất cho tỷ lệ phế phẩm là 6%. Năm nay người ta nhập công nghệ B để sản xuất, lấy ngẫu nhiên 100 sản phẩm thấy có 5 phế phẩm.

Có thể cho rằng tỷ lệ phế phẩm của công nghệ B nhỏ hơn công nghệ A hay không? Hãy kết luận với mức ý nghĩa 5%.

2.3. KIỂM ĐỊNH GIẢ THUYẾT CHO XÁC SUẤT (TỶ LỆ) (tiếp theo)

Giải:

Gọi p là tỷ lệ phế phẩm của công nghệ B.

Ta cần kiểm định $\begin{cases} H_0 : p = 0,06 \\ H_1 : p < 0,06 \end{cases}$

mức ý nghĩa α = 0,05, tra bảng phân phối chuẩn ta có $u_{0,05}$ = 1,65

Ta có miền bác bỏ: $W = (-\infty; \ -1,65)$.

Ta có n =100; m = 5, do đó f = m/n = 0,05.

Tính giá trị tiêu chuẩn thống kê ta thu được:

$$u_{qs} = \frac{(0,05 - 0,06)}{\sqrt{0,06(1-0,06)}}\sqrt{100} = -0,42.$$

Kiểm tra ta thấy $u_{qs} \notin W$. Vậy chấp nhận giả thuyết H_0.

2.4. KIỂM ĐỊNH GIẢ THUYẾT CHO HAI KỲ VỌNG

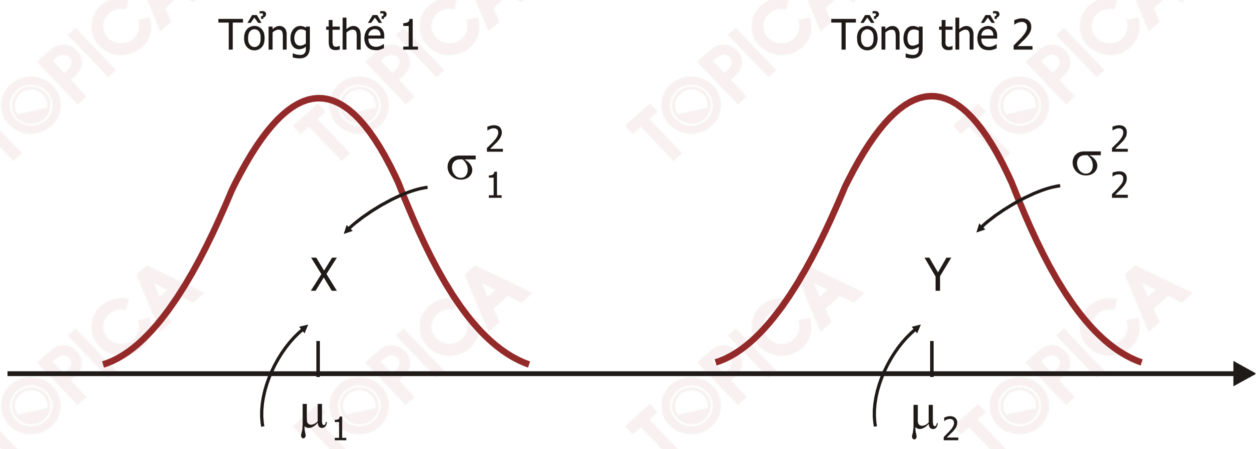
Cho hai biến ngẫu nhiên độc lập

$X \sim N(\mu_1,\sigma_1^2)$; $Y \sim N(\mu_2,\sigma_2^2)$

Mẫu ngẫu nhiên $(X_1,X_2,\ldots X_n)$ nhận giá trị $(x_1,x_2,\ldots,x_n)$

Mẫu ngẫu nhiên $(Y_1,Y_2,\ldots Y_m)$ nhận giá trị $(y_1,y_2,\ldots,y_m)$.

Ta cần kiểm định giả thuyết: $H_0 : \mu_1 = \mu_2$.



Mẫu 1:
$(x_1, x_2, \ldots, x_n)$

Mẫu 2:
$(y_1, y_2, \ldots, y_m)$

2.4. KIỂM ĐỊNH GIẢ THUYẾT CHO HAI KỲ VỌNG (tiếp theo)

Trường hợp σ_1, σ_2 đã biết:

Ta có thống kê $U = \dfrac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{m}}}$ có phân phối chuẩn N(0,1)

- Nếu giả thuyết H_0 là đúng khi đó $U = \dfrac{(\overline{X} - \overline{Y})}{\sqrt{\dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{m}}}$.
- Với mẫu cụ thể giá trị của tiêu chuẩn thống kê là $u_{qs} = \dfrac{(\overline{x} - \overline{y})}{\sqrt{\dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{m}}}$.

Bài toán 1: $\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases}$ Miền bác bỏ: $W = \left(-\infty; -u_{\alpha/2}\right) \cup \left(u_{\alpha/2}; +\infty\right)$	**Bài toán 2:** $\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 > \mu_2 \end{cases}$ Miền bác bỏ: $W = \left(u_\alpha; +\infty\right)$	**Bài toán 3:** $\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 < \mu_2 \end{cases}$ Miền bác bỏ: $W = \left(-\infty; -u_\alpha\right)$

2.4. KIỂM ĐỊNH GIẢ THUYẾT CHO HAI KỲ VỌNG (tiếp theo)

Ví dụ:

Hai trường A và B cùng học môn toán, khảo sát kết quả thi hết môn ta thu được kết quả sau:

Trường A: n = 64;

Trường B: m = 68;

Biết rằng điểm thi của hai trường là biến ngẫu nhiên có phân phối chuẩn với độ lệch chuẩn tương ứng là σ_1=1,09 và σ_2=1,12.

Với mức ý nghĩa 1% có thể cho rằng kết quả thi của trường B cao hơn trường A hay không?

2.4. KIỂM ĐỊNH GIẢ THUYẾT CHO HAI KỲ VỌNG (tiếp theo)

Ví dụ:

Hai trường A và B cùng học môn toán, khảo sát kết quả thi hết môn ta thu được kết quả sau:

Trường A: n = 64;

Trường B: m = 68;

Biết rằng điểm thi của hai trường là biến ngẫu nhiên có phân phối chuẩn với độ lệch chuẩn tương ứng là σ_1=1,09 và σ_2=1,12.

Với mức ý nghĩa 1% có thể cho rằng kết quả thi của trường B cao hơn trường A hay không?

Giải: Gọi X và Y là kết quả thi của hai trường A và B,

$X \sim N(\mu_1;\sigma_1^2);\ Y \sim N(\mu_2;\sigma_2^2).$

Ta cần kiểm định $\begin{cases} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 < \mu_2 \end{cases}$

Với mức ý nghĩa đã cho, tra bảng phân phối chuẩn, ta có

$u_{0,01} = 2,33.$

Miền bác bỏ: $W = (-\infty; -2,33)$

Tính giá trị của tiêu chuẩn thống kê ta được

$$u_{qs} = \frac{7,32 - 7,66}{\sqrt{\frac{1,09^2}{64} + \frac{1,12^2}{68}}} = -31,43$$

$u_{qs} \in W$, vậy ta bác bỏ giả thuyết H_0, kết quả thi ở trường B cao hơn trường A.

Trường hợp phương sai σ_1, σ_2 chưa biết

Giả sử rằng $\sigma_1^2 = \sigma_2^2 = \sigma^2$. Xét thống kê: $T = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{nS_x^2 + mS_y^2}{n+m-2}}\sqrt{\dfrac{n+m}{nm}}}$

khi đó $T \sim T(n + m - 2)$.

Với mẫu cụ thể, ta tính giá trị của tiêu chuẩn thống kê T:

$$t_{qs} = \frac{\overline{x} - \overline{y}}{\sqrt{\dfrac{ns_x^2 + ms_y^2}{n+m-2}}\sqrt{\dfrac{n+m}{nm}}}.$$

Bài toán 1: $\begin{cases} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{cases}$	**Bài toán 2:** $\begin{cases} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 > \mu_2 \end{cases}$	**Bài toán 3:** $\begin{cases} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 < \mu_2 \end{cases}$
Miền bác bỏ: $W = \left(-\infty; -t_{\alpha/2}^{n+m-2}\right) \cup \left(t_{\alpha/2}^{n+m-2}; +\infty\right)$	Miền bác bỏ: $W = \left(t_{\alpha/2}^{n+m-2}; +\infty\right)$	Miền bác bỏ: $W = \left(-\infty; -t_{\alpha/2}^{n+m-2}\right)$

Ví dụ:

Điều tra thu nhập (tính theo $) trong một tháng của công nhân ở hai nhà máy sản xuất thiết bị điện tử A và B ta thu được số liệu sau:

Nhà máy A: 91,5 94,18 92,18 95,39 91,79 89,07 94,72 89,21.

Nhà máy B: 89,19 90,95 90,46 93,21 97,19 97,04 91,07 92,75.

Với mức ý nghĩa 5% có thể cho rằng thu nhập trung bình của công nhân trong hai nhà máy A và B là như nhau hay không? Biết rằng thu nhập của công nhân trong hai nhà máy có phân phối chuẩn.

2.4. KIỂM ĐỊNH GIẢ THUYẾT CHO HAI KỲ VỌNG (tiếp theo)

Giải: Gọi X và Y là thu nhập của công nhân trong hai nhà máy A và B

$X \sim N(\mu_1; \sigma_1^2); N(\mu_1; \sigma_2^2)$. Ta cần kiểm định: $\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_2 : \mu_1 \neq \mu_2 \end{cases}$

Ta có n = 8; m = 8, $\overline{x} = 92,255$; $s_x^2 = 4,998$; $\overline{y} = 92,733$; $s_y^2 = 7,77$.

Với mức ý nghĩa α=0,05 tra bảng phân phối Student ta được:

$$t_{0,025}^{8+8-2} = t_{0,025}^{14} = 2,14.$$

Ta có miền bác bỏ: $W = (-\infty; \ -2,14) \cup (2,14; \ +\infty)$.

Với mẫu đã cho, tính toán ta được giá trị tiêu chuẩn thống kê:

$$t_{qs} \notin W,\ t_{qs} = \frac{92,255 - 92,733}{\sqrt{\frac{8.4,998 + 8.7,77}{8+8-2}}\sqrt{\frac{8+8}{8.8}}} = -0,353.$$

Vậy chấp nhận H_0, tức là công nhân hai nhà máy có thu nhập như nhau.

2.5. KIỂM ĐỊNH GIẢ THUYẾT CHO 2 XÁC SUẤT

Cho hai biến ngẫu nhiên X ~ A(p_1), Y ~ A(p_2)

Xét hai mẫu ngẫu nhiên rút từ X và Y tương ứng là: $(X_1, X_2, ..., X_n)$ và $(Y_1, Y_2, ..., Y_m)$

Gọi k_1 số lần mẫu ngẫu nhiên của X nhận giá trị 1

Gọi k_2 số lần mẫu ngẫu nhiên của Y nhận giá trị 1

Đặt $f_1 = \frac{k_1}{n}$, $f_2 = \frac{k_2}{m}$, $f = \frac{k_1 + k_2}{n + m}$ là các tần suất mẫu.

Ta cần kiểm định giả thuyết H_0: $p_1 = p_2$ và các đối thuyết

$H_1 : p_1 \neq p_2$; $H_1 : p_1 > p_2$; $H_1 : p_1 < p_2$

2.5. KIỂM ĐỊNH GIẢ THUYẾT CHO 2 XÁC SUẤT

Xét thống kê $U = \dfrac{f_1 - f_2}{\sqrt{f(1-f)\left(\dfrac{1}{n} + \dfrac{1}{m}\right)}}$. Thống kê U có phân phối N(0,1)

$\begin{cases} H_0: p_1 = p_2 \\ H_1: p_1 \neq p_2 \end{cases}$ $W = (-\infty; -u_{\alpha/2}) \cup (u_{\alpha/2}; +\infty)$

$\begin{cases} H_0: p_1 = p_2 \\ H_1: p_1 > p_2 \end{cases}$ $W = (u_\alpha; +\infty)$

$\begin{cases} H_0: p_1 = p_2 \\ H_1: p_1 < p_2 \end{cases}$ $W = (-\infty; -u_\alpha)$

Giải:

Gọi p_1 và p_2 là tỷ lệ học sinh bỏ học ở vùng nông thôn A và B.

Ta cần kiểm định: $\begin{cases} H_0: p_1 = p_2 \\ H_1: p_1 < p_2 \end{cases}$

Ta có n = 1900; k_1 = 175, tần suất f_1 = 0,092.

m = 2600; k_2 = 325, tần suất f_2 = 0,125.

f = 0,111.

Vậy giá trị tiêu chuẩn thống kê:

$$u_{qs} = \frac{0,092 - 0,125}{\sqrt{0,111(1-0,111)\left(\frac{1}{1900}+\frac{1}{2600}\right)}} = -3,48.$$

Với mức ý nghĩa 0,05, tra bảng ta được $u_{0,05}$ = 1,65,

$u_{qs} \in W$, vậy miền bác bỏ là: $W = (-\infty; -1,65)$.

Ta bác bỏ giả thuyết H_0, tức là tình trạng bỏ học ở vùng A là ít nghiêm trọng hơn vùng B.

6. KIỂM ĐỊNH GIẢ THUYẾT CHO HAI PHƯƠNG SAI

Cho

Mẫu

Mẫu

Ta c

Các

Xét

S_y / σ_2

và m-1 bậc tự do.

Nếu

Với

Bài toán 1: $\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1 \neq \sigma_2 \end{cases}$

Miền bác bỏ:

Bài toán 2 $\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1 > \sigma_2 \end{cases}$

Miền bác bỏ:

Bài toán 3: $\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1 < \sigma_2 \end{cases}$

Miền bác bỏ:

2.6. KIỂM ĐỊNH GIẢ THUYẾT CHO HAI PHƯƠNG SAI (tiếp theo)

Ví dụ:

Hai máy A và B cùng gia công một loại chi tiết. Người ta muốn kiểm tra xem hai máy có độ chính xác của hai máy và để làm điều đó người ta tiến hành lấy mẫu và thu được kết quả sau:

Máy A:	135	138	136	140	138	135	139
Máy B:	140	135	140	138	135	138	140

Với mức ý nghĩa 5% hãy kiểm tra xem hai máy có độ chính xác như nhau hay không? Biết kích thước chi tiết là biến ngẫu nhiên có phân phối chuẩn.

Giải: Gọi X là kích thước chi tiết của máy A, $X \sim N(\mu_1; \sigma_1^2)$

Y là kích thước chi tiết của máy B, $Y \sim N(\mu_1; \sigma_1^2)$.

Ta cần kiểm định: $\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \end{cases}$

Mức ý nghĩa là 5% và cỡ mẫu n = 7, m = 7, tra bảng phân phối F được

$f_{0,975,6,6} = 0,2;\ f_{0,025,6,6} = \dfrac{1}{0,2} = 4,995.$

Vậy miền bác bỏ $W = (0;\ 0,2) \cup (4,995; +\infty)$.

Với mẫu tính toán ta được: $s_x^{'2} = 3,905;\ s_y^{'2} = 5$

Giá trị tiêu chuẩn thống kê: $f_{qs} = 3,905/5 = 0,781$

Ta thấy $f_{qs} \notin W$. Vậy chấp nhận giả thuyết H_0, tức là độ chính xác hai máy là như nhau.

3. MỘT SỐ TIÊU CHUẨN KIỂM ĐỊNH PHI THAM SỐ

Kiểm định giả thuyết về phân phối của biến ngẫu nhiên:

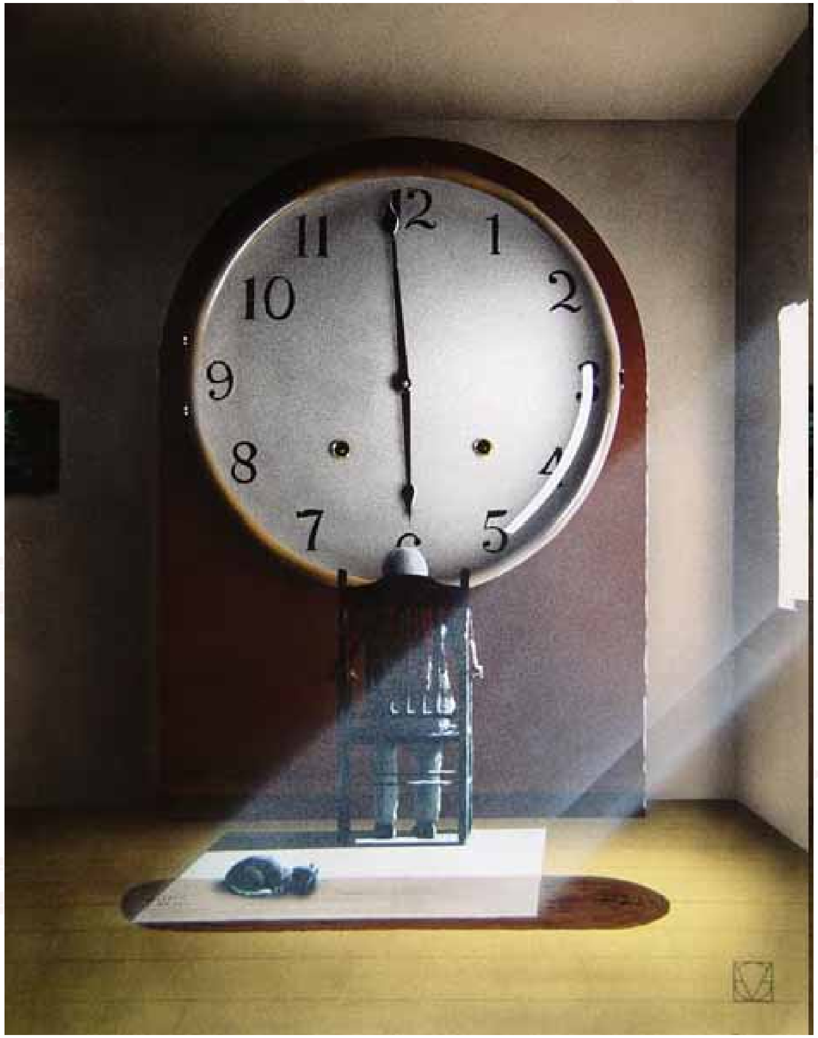
$$\begin{cases} H_0: X \sim F_0(x, \theta_1, \theta_2, \ldots, \theta_k) \\ H_1: X \nsim F_0(x, \theta_1, \theta_2, \ldots, \theta_k) \end{cases}$$

So sánh nhiều tỷ lệ:

$$\begin{cases} H_0: p_1 = p_2 = \ldots = p_k \\ H_1: \exists p_i \neq p_j, i \neq j \end{cases}$$

Kiểm tra tính độc lập

$$\begin{cases} H_0: \text{X độc lập với Y} \\ H_1: \text{X không độc lập với Y} \end{cases}$$



3.1. KIỂM ĐỊNH GIẢ THUYẾT PHÂN PHỐI CỦA BIẾN NGẪU NHIÊN

Cho biến ngẫu nhiên X có hàm phân phối F(x)

Ta có bài toán kiểm định sau:

H_0 : X có phân phối $F_0(x, \theta_1, \theta_2, ..., \theta_r)$

H_1 : X không có phân phối $F_0(x, \theta_1, \theta_2, ..., \theta_r)$

Trong đó $\theta_1, \theta_2, ..., \theta_r$các là các tham số của phân phối .

Xét mẫu ngẫu nhiên $(X_1, X_2, \ldots X_n)$ với giá trị mẫu $(x_1, x_2, \ldots x_n)$.

Chia miền giá trị của X thành k miền không giao nhau S_i; $i = (1...k)$.

Ký hiệu n_i là số giá trị mẫu rơi vào khoảng $(S_1, S_2 \ldots S_k)$, n_i được gọi là tần số thực nghiệm.

Nếu giả thuyết H_0 là đúng thì khi đó X có phân phối xác định là F_0 do đó ta tính được các xác suất: $p_i = P\{X \in S_i\}$, $i = 1, 2, .., k$.

Đặt $E_i = n_i.p_i, (i = 1, 2...k)$, E_i được gọi là tần số lý thuyết.

3.1. KIỂM ĐỊNH GIẢ THUYẾT PHÂN PHỐI CỦA BIẾN NGẪU NHIÊN (tiếp theo)

Trường hợp tham số $\theta_1, \theta_2, \ldots, \theta_r$ đã biết:

Xét thống kê $\chi^2 = \sum_{i=1}^{k} \frac{(n_i - E_i)^2}{E_i}$ có phân phối khi bình phương với k-1 bậc tự do.

Ta có miền bác bỏ: $W = (\chi^2_{\alpha,k-1};\ +\infty)$,

Với mẫu cụ thể ta tính giá trị của tiêu chuẩn thống kê χ^2_{qs} và so sánh với miền bác bỏ.

Ví dụ: Quan sát biến ngẫu nhiên X ta thu được số liệu mẫu sau:

Giá trị X	0-1	1-3	3-6	6-7	7-10
Số lần	3	4	2	5	4

Với mức ý nghĩa 5% hãy kiểm định xem biến ngẫu nhiên X có phân phối đều trên [0; 10] hay không?

3.1. KIỂM ĐỊNH GIẢ THUYẾT PHÂN PHỐI CỦA BIẾN NGẪU NHIÊN (tiếp theo)

$$\theta_1, \theta_2, ..., \theta_r$$

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - E_i)^2}{E_i}$$

$$W = (\chi^2_{\alpha,k-1}; \quad +\infty),$$

$$\chi^2_{qs}$$

Giá trị X	0-1	1-3	3-6	6-7	7-10
Số lần	3	4	2	5	4

3.1. KIỂM ĐỊNH GIẢ THUYẾT PHÂN PHỐI CỦA BIẾN NGẪU NHIÊN (tiếp theo)

Giải: Ta cần kiểm định H_0: X có phân phối đều trên [0; 10],

H_1: X không có phân phối đều trên [0; 10].

Ta có các khoảng $S_1 = (0,1)$, $n_1 = 3$; $S_2 = (1; 3)$, $n_2 = 4$;

$S_3 = (3; 6)$, $n_3 = 2$; $S_4 = (6; 7)$, $n_4 = 5$; $S_5 = (7\text{-}10)$, $n_5 = 4$.

Cỡ mẫu n = 18.

Các xác suất tương ứng:

$$p_1 = P\{0 < X < 1\} = 1/10;\ p_2 = P\{1 < X < 3\} = 2/10 = 1/5.$$

$$p_3 = 3/10;\ p_4 = 1/10;\ p_5 = 3/10.$$

Tương tự ta tính được

Các tần số lý thuyết:

$$E_1 = 18.1/10 = 1,8;\ E_2 = 18.1/5 = 3,6;\ E_3 = 18.3/10 = 5,4;$$

$$E_4 = 18.1/10 = 1,8;\ E_5 = 18.3/10 = 5,4.$$

Tra bảng phân phối khi bình phương ta được: $\chi^2_{0,05,4} = 9,488$, miền bác bỏ là $W = (9,488;\ +\infty)$.

So sánh ta thấy $\chi^2_{qs} \notin W$, do đó ta chưa bác bỏ giả thuyết H_0.

3.1. KIỂM ĐỊNH GIẢ THUYẾT PHÂN PHỐI CỦA BIẾN NGẪU NHIÊN (tiếp theo)

Trường hợp các tham số $\theta_1, \theta_2, ..., \theta_r$ chưa biết

Ta cũng tiến hành các bước tương tự như trước nhưng các xác suất

$p_i = P(X \in S_i)$ sẽ phụ thuộc vào các tham số vì vậy ta thay các tham số bằng các ước lượng điểm tương ứng.

Ta cũng xét tiêu chuẩn thống kê như trong trường hợp trước nhưng lưu ý rằng tiêu chuẩn thông kê bây giờ có phân phối khi bình phương với k-r-1 bậc tự do.

Phân phối chuẩn: Tham số $\theta_1 = \mu; \theta_2 = \sigma^2, r = 2$, ta thay μ bằng $\overline{x}$, thay σ^2 bằng s'^2.

Phân phối Posson: Tham số λ (r = 1) được thay bằng $\overline{x}$.

Phân phối nhị thức: Tham số p (r = 1) được thay bằng tần suất f = m/n.

Phân phối mũ: Tham số λ (r = 1), thay tham số λ bằng $1/\overline{x}$.

Phân phối đều trên [a; b]: (r = 2). Tham số a được thay bằng $\overline{x} - \sqrt{3}s$,

Tham số b được thay bằng $\overline{x} + \sqrt{3}s$.

3.1. KIỂM ĐỊNH GIẢ THUYẾT PHÂN PHỐI CỦA BIẾN NGẪU NHIÊN (tiếp theo)

Ví dụ:

Quan sát biến ngẫu nhiên X ta thu được giá trị mẫu như sau:

Giá trị X	1-3	3-5	5 -7	7-9	9-11
n_i	3	6	4	7	2

Với mức ý nghĩa 5% có thể cho rằng biến ngẫu nhiên X có phân phối chuẩn hay không?

1. KIỂM ĐỊNH GIẢ THUYẾT PHÂN PHỐI CỦA BIẾN NGẪU NHIÊN (tiếp theo)

Giải: Ta cần kiểm định bài toán sau:

H_0: X có phân phối chuẩn $N(\mu,\sigma^2)$

H_1: X không có phân phối chuẩn $N(\mu,\sigma^2)$

Ta có miền giá trị của X là R do đó ta phải gộp các khoảng lại

$S_1 = (-\infty;\ 3),\ n_1 = 3;\ S_2 = (3;\ 5),\ n_2 = 6;\ S_3 = (5;\ 7),\ n_3 = 4;$

$S_4 = (7;\ 9),\ n_4 = 7;\ S_5 = (9;\ +\infty),\ n_5 = 2.$

Với mẫu đã cho ta tính được: $\overline{x} = 5,91;\ s'^2 = 6,25.$

Vậy $\mu = \overline{x} = 5,91;\ \sigma^2 = s'^2 = 6,25.$

Từ đó ta có các giá trị kì vọng sau:

$E_1 = 3,982;\quad E_2 = 6,05;$

$E_3 = 6,578;\quad E_4 = 3,894;$

$E_5 = 1,496$

$$\chi^2_{qs} = \frac{(3-3,982)^2}{3,982} + \frac{(6-6,05)^2}{6,05} + \frac{(4-6,578)^2}{6,578} + \frac{(7-3,894)^2}{3,894} + \frac{(2-1,496)^2}{1,496} = 3,9$$

Mức ý nghĩa 5% và số tham số là r = 2, tra bảng ta được: $\chi^2_{0,052} = 5,99,$

Vậy miền bác bỏ là W=(5,99; +∞). So sánh ta thấy chưa bác bỏ giả thuyết H_0.

2. SO SÁNH NHIỀU TỶ LỆ

Giả sử k biến ngẫu nhiên ($X_1,X_2...X_k$) độc lập cùng phân phối 0-1 các xác suất tương ứng là ($p_1,p_2...p_k$).

Ta kiểm định bài toán sau: $\begin{cases} H_0 : p_1 = p_2 = \ldots = p_k \\ H_1 : \exists p_i \neq p_j,\ i \neq j \end{cases}$

Xét mẫu n quan sát về biến ngẫu nhiên, ta có bảng số liệu mẫu:

Trong đó: $v_i = \sum_{s=1}^{2} n_{si};\ u_s = \sum_{i=1}^{k} n_{si}$

	X_1	...	X_i	...	X_k	
0	n_{11} E_{11}	...	n_{1i} E_{1i}	...	n_{1k} E_{1k}	u_1
1	n_{21} E_{21}	...	n_{2i} E_{2i}	...	n_{2k} E_{1k}	u_2
	v_1	...	v_i	...	V_k	n

$(v_1 + v_2 + \ldots + v_k = n)$; v_i là cỡ mẫu của biến ngẫu nhiên X_i

Đặt $E_{si} = \dfrac{u_s.v_i}{n};\ s = 1,2,\ i = 1,2,..,k \sim \chi^2(k-1)$

Ta viết giá trị E_{si} ngay cạnh giá trị của n_{si} trong bảng số liệu.

Thống kê: $$\chi^2 = \sum_{s=1}^{2} \sum_{i=1}^{k} \frac{(n_{si} - E_{si})^2}{E_{si}}$$

Với mức ý nghĩa α tra bảng phân phối khi bình phương ta tìm được giá trị phân vị $\chi^2_{\alpha,k-1}$, và tính được miền bác bỏ $W = (\chi^2_{\alpha,k-1};\ +\infty)$.

Giải:

Gọi p_1,p_2,p_3 là tỷ lệ phế phẩm của ba nhà máy A, B, C.

Ta cần kiểm định: $\begin{cases} H_0 : p_1 = p_2 = p_3 \\ H_1 : \exists p_i \neq p_j, i \neq j \end{cases}$

Với số liệu đã cho tính toán ta được các E_{si} được ghi trong bảng.

Ta có:

$$C_{qs}^2 = \frac{(12-13,14)^2}{13,14} + \frac{(88-86,86)^2}{86,86} + \frac{(16-15,77)^2}{15,77} + \frac{(104-104,23)^2}{104,23}$$

$$+ \frac{(18-17,09)^2}{17,09} + \frac{(112-112,91)^2}{112,91} = 0,174$$

Mức ý nghĩa 5% , k = 3, tra bảng ta thu được: $\chi^2_{0,05,2} = 5,99$,

và tính được miền bác bỏ W=(5,99; +∞)

So sánh ta thấy $\chi^2_{qs} \notin W$, do đó chấp nhận giả thuyết H_0.

3.3. KIỂM TRA TÍNH ĐỘC LẬP

Cho hai biến ngẫu nhiên X và Y với giả thiết:

H_0: X độc lập với Y.

H_1: X không độc lập với Y.

Xét một mẫu ngẫu nhiên hai chiều

$(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$

rút ra từ véc tơ ngẫu nhiên (X, Y),

giá trị mẫu $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$

Thu gọn giá trị mẫu ta có bảng

biểu diễn sau:

Trong đó n_{ij} là số lần cặp giá trị

(x_i, y_j) xuất hiện trong giá trị mẫu,

$$b_j = \sum_{i=1}^{r} n_{ij}, j = 1, 2, \ldots, s; a_i = \sum_{j=1}^{s} n_{ij}, i = 1, 2, \ldots, r$$

Đặt $E_{ij} = \dfrac{a_i . b_j}{n}$, ta viết giá trị của

E_{ij} vào trong ngoặc bên cạnh

của ô (i;j).

y_j / x_i	y_1	…	y_j	…	y_s	Σ
x_1	n_{11} (E_{11})	…	n_{1j} (E_{1j})	…	n_{1s} (E_{1s})	b_1
…						…
x_i	n_{i1} (E_{i1})	…	n_{ij} (E_{ij})	…	n_{is} (E_{is})	b_i
…						…
x_r	n_{r1} (E_{r1})	…	n_{rj} (E_{rj})	…	n_{rs} (E_{rs})	b_r
Σ	a_1	…	a_j	…	a_s	n

3.3. KIỂM TRA TÍNH ĐỘC LẬP (tiếp theo)

Thống kê $x^2 = \frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{s}\frac{(n_{ij}-E_{ij})^2}{E_{ij}} \sim x^2((r-1).(s-1))$

Miền bác bỏ là: $W = (x^2_{\alpha,(r-1)(s-1)};\ +\infty)$.

Ví dụ:

Nghiên cứu tình trạng hôn nhân trước ngày cưới của 542 cặp vợ chồng ta có bảng số liệu.

Với mức ý nghĩa 5% có thể cho rằng tình trạng hôn nhân của vợ và chồng là độc lập với nhau hay không?

Tình trạng hôn nhân vợ chồng	Chưa kết hôn lần nào	Ly hôn	Goá	$\sum$
Chưa kết hôn lần nào	180 (129,61)	34 (66,42)	36 (53,97)	250
Ly hôn	58 (97,47)	76 (49,95)	54 (40,58)	188
Goá	43 (53,92)	34 (27,63)	27 (22,45)	104
$\sum$	281	144	117	542

3.3. KIỂM TRA TÍNH ĐỘC LẬP (tiếp theo)

Giải: H_0: Tình trạng hôn nhân của chồng độc lập với vợ;

H_1: Tình trạng hôn nhân của chồng không độc lập với vợ.

Ta tính toán các E_{ij} và viết vào bảng số liệu.

Tiếp đó, ta tính giá trị tiêu chuẩn thống kê như sau:

$$\chi^2_{qs} = \frac{(180-129.61)^2}{129.61} + \frac{(58-97.47)^2}{97.47} + \frac{(43-53.92)^2}{53.92} + \frac{(34-66.42)^2}{66.42} + \frac{(76-49.95)^2}{49.95} + \frac{(34-27.63)^2}{27.63} + \frac{(36-53.97)^2}{53.97} + \frac{(54-40.58)^2}{40.58} + \frac{(27-22.45)^2}{22.45} = 80.$$

Ta có s = r = 3 và miền bác bỏ:

$$W = (9,48;\ +\infty)$$

So sánh, ta thấy $\chi^2_{qs} \in W$

Do đó ta bác bỏ giả thuyết H_0.

TÓM TẮT CUỐI BÀI

Các nội dung chính:

- Khái niệm kiểm định giả thuyết thống kê và phương pháp kiểm định giả thuyết thống kê.
- Kiểm định tham số và các bài toán kiểm định tham số.
- Kiểm định phi tham số và các bài toán kiểm định phi tham số.

Lưu ý:

- Đây là một kỹ thuật được dùng rất phổ biến trong thống kê cũng như trong các nghiên cứu xã hội học khác.
- Có thể kết hợp các bài toán về Ước lượng và các bài toán về kiểm định giả thuyết thống kê. Đây là hai bài rất quan trọng.

TỪ ĐIỂN THUẬT NGỮ

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Select a term:

- Đối thuyết
- Giả thuyết
- Giá trị của tiêu chuẩn thống kê
- Kiểm định phi tham số
- Miền bác bỏ
- Thuyết thống kê
- Tiêu chuẩn thống kê
- Xác suất ý nghĩa (p-value)

Đối thuyết

Đối thuyết:

Đi kèm với giả thuyết H_0 là mệnh đề đối lập được gọi là đối thuyết, ký hiệu là H_1.

PROPERTIES

Allow user to leave interaction: Anytime
Show 'Next Slide' Button: Don't show
Completion Button Label: Next Slide

Properties...

Edit in Engage