

Chương 7. Lý thuyết kiểm định

§1: Khái niệm chung về kiểm định

Việc dùng kết quả của mẫu để khẳng định hay bác bỏ một giả thiết H nào đó được gọi là kiểm định giả thiết H . Khi kiểm định ta có thể mắc 1 trong 2 loại sai lầm sau:

cuu duong than cong . com

1. **Sai lầm loại 1:** Là sai lầm mắc phải nếu ta bác bỏ H trong khi H đúng. Ta ký hiệu xác suất để mắc sai lầm loại này là α và gọi α là mức ý nghĩa.
2. **Sai lầm loại 2:** Là sai lầm mắc phải nếu ta công nhận H trong khi H sai. Ta ký hiệu xác suất để mắc sai lầm loại này là β và gọi $1-\beta$ là lực kiểm định.

Trong các bài toán kiểm định ta sẽ xét sau này mức ý nghĩa α là cho trước.

Giả thiết

$$H: P = P_0$$

Giả thiết đối lập: $\bar{H}$ $\begin{cases} P < P_0 \text{ (thiếu)} \\ P > P_0 \text{ (thừa)} \\ P \neq P_0 \text{ (đối xứng-ta chỉ xét bài này)} \end{cases}$

§2: Kiểm định giả thiết về tỉ lệ

1. Bài toán 1 mẫu:

Bài toán: Ký hiệu tỉ lệ của 1 tổng thể là P (chưa biết). Từ tổng thể lấy 1 mẫu kích thước n , có tỉ lệ mẫu f . Với mức ý nghĩa α hãy kiểm định giả thiết:

$$H: P = P_0$$

Giải:

Bước 1: Tra ngưỡng Z_α

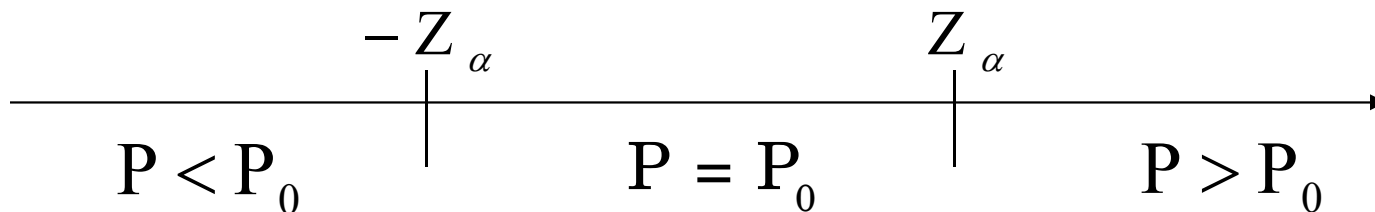
Bước 2: Tính giá trị quan sát:
$$U_{qs} = \frac{(f - P_0)\sqrt{n}}{\sqrt{P_0(1 - P_0)}}$$

Bước 3: Kết luận:

$$|U_{qs}| \leq Z_\alpha \Rightarrow H \text{ đúng} \Rightarrow P = P_0$$

$$|U_{qs}| > Z_\alpha \Rightarrow H \text{ sai} \Rightarrow P \neq P_0$$

$$P \neq P_0 \begin{cases} U_{qs} < -Z_\alpha \Rightarrow P < P_0 \\ U_{qs} > Z_\alpha \Rightarrow P > P_0 \end{cases}$$



2. Bài toán 2 mẫu

Bài toán: kí hiệu tỉ lệ của tổng thể 1, 2 là P_1, P_2 (cả 2 chưa biết). Từ các tổng thể lấy các mẫu kích thước n_1, n_2 , có tỉ lệ mẫu $f_1 = \frac{m_1}{n_1}, f_2 = \frac{m_2}{n_2}$. Với mức ý nghĩa α , hãy kiểm

định giả thiết:

$$H: P_1 = P_2$$

Bước 1: Tra ngưỡng Z_α

Bước 2:

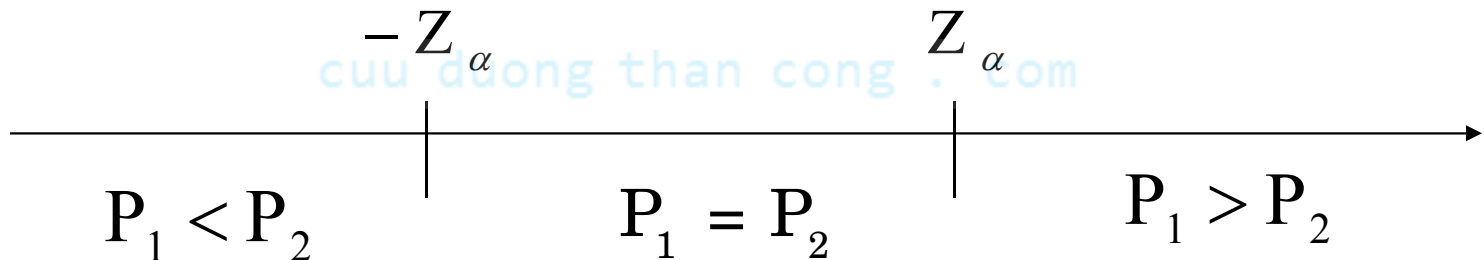
$$U_{qs} = \frac{\frac{m_1}{n_1} - \frac{m_2}{n_2}}{\sqrt{\frac{m_1 + m_2}{n_1 \cdot n_2} \left(1 - \frac{m_1 + m_2}{n_1 + n_2} \right)}}$$

Bước 3: Kết luận:

$$|U_{qs}| \leq Z_{\alpha} \Rightarrow H \text{ đúng} \Rightarrow P_1 = P_2$$

$$|U_{qs}| > Z_{\alpha} \Rightarrow H \text{ sai} \Rightarrow P_1 \neq P_2$$

$$P_1 \neq P_2 \begin{cases} U_{qs} < -Z_{\alpha} \Rightarrow P_1 < P_2 \\ U_{qs} > Z_{\alpha} \Rightarrow P_1 > P_2 \end{cases}$$



Ví dụ 2.1: Nếu áp dụng phương pháp I thì tỉ lệ phế phẩm là 6%, còn nếu áp dụng phương pháp II thì trong 100 sản phẩm có 5 phế phẩm. Vậy có thể kết luận áp dụng phương pháp thứ II thì tỉ lệ phế phẩm ít hơn phương pháp thứ I không? Hãy kết luận với mức ý nghĩa 0,05.

Giải: Ký hiệu $P_0 = 0,06$ là tỉ lệ phế phẩm của phương pháp I ;
 P là tỉ lệ phế phẩm của phương pháp II (chưa biết)

$$H : P = P_0 = 0,06$$

Bước 1: $Z_{\alpha} = 1,96, f = 0,05$

Bước 2:
$$U_{qs} = \frac{(f - P_0) \sqrt{n}}{\sqrt{P_0 (1 - P_0)}} = \frac{(0,05 - 0,06) \cdot 10}{\sqrt{0,06 \cdot 0,94}} = -0,42$$

Bước 3: $|U_{qs}| < Z_{0,05} = 1,96 \Rightarrow P = P_0$.Vậy tỉ lệ phế phẩm của phương pháp II bằng với tỉ lệ của phương pháp I **hoặc** Chưa đủ cơ sở để kết luận áp dụng phương pháp thứ II thì tỉ lệ phế phẩm ít hơn phương pháp thứ I

- **Ví dụ 2.2.** Thống kê số phế phẩm của 2 nhà máy cùng sản xuất một loại sản phẩm có bảng số liệu :

Nhà máy	Số sản phẩm	Số phế phẩm
I	1200	20
II	1400	60

Với mức ý nghĩa 0.05 ,hãy xét xem tỷ lệ phế phẩm ở 2 nhà máy trên có như nhau hay không ?

P_1 -tỷ lệ phế phẩm của nhà máy I

P_2 -tỷ lệ phế phẩm của nhà máy II

$$H : P_1 = P_2$$

Bước 1

$$\alpha = 0,05 \Rightarrow Z_{\alpha} = 1,96$$

Bước 2

$$U_{qs} = \frac{\frac{20}{1200} - \frac{60}{1400}}{\sqrt{\frac{20+60}{1200 \cdot 1400} \left(1 - \frac{80}{2600}\right)}} = -3,855$$

Bước 3

$$U_{qs} < -Z_{\alpha} = -1,96 \Rightarrow P_1 < P_2$$

Vậy tỷ lệ phế phẩm của nhà máy 1 thấp hơn nhà máy 2

§ 3. Kiểm định giả thiết về giá trị trung bình

1. Bài toán 1 mẫu:

Ký hiệu trung bình của 1 tổng thể là a (chưa biết). Từ tổng thể lấy 1 mẫu kích thước n có trung bình mẫu $\bar{x}$ và phương sai điều chỉnh mẫu S^2 . Với mức ý nghĩa α , hãy kiểm định giả thiết:

$$H : a = a_0$$

Giải:

Trường hợp 1: Đã biết phương sai tổng thể σ^2

B1: Tra ngưỡng Z_α

B2:

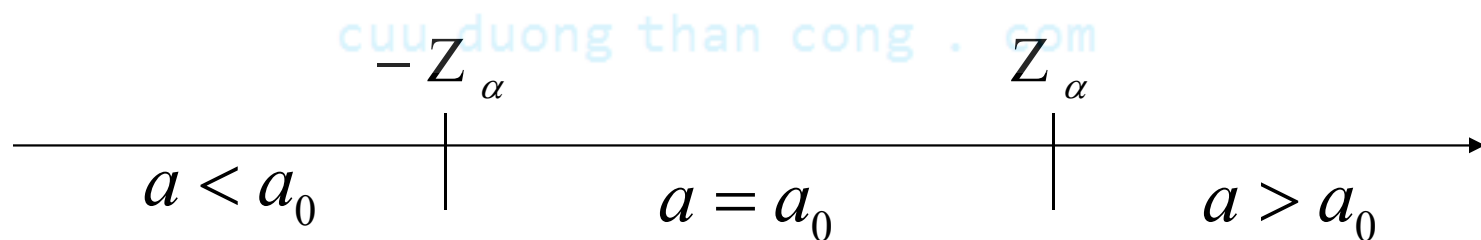
$$U_{qs} = \frac{(\bar{x} - a_0) \sqrt{n}}{\sigma}$$

B3:

$$|U_{qs}| \leq Z_\alpha \Rightarrow H \text{ đúng} \Rightarrow a = a_0$$

$$|U_{qs}| > Z_\alpha \Rightarrow H \text{ sai} \Rightarrow a \neq a_0$$

$$a \neq a_0 : \begin{cases} U_{qs} < -Z_\alpha \Rightarrow a < a_0 \\ U_{qs} > Z_\alpha \Rightarrow a > a_0 \end{cases}$$



TH 2: Chưa biết phương sai tổng thể σ^2 , $n \geq 30$

B1: Tra ngưỡng Z_α

B2:

$$U_{qs} = \frac{(\bar{x} - a_0) \sqrt{n}}{S}$$

B3: Kết luận

$$|U_{qs}| \leq Z_\alpha \Rightarrow H \text{ đúng} \Rightarrow a = a_0$$

$$|U_{qs}| > Z_\alpha \Rightarrow H \text{ sai} \Rightarrow a \neq a_0$$

$$a \neq a_0 \begin{cases} U_{qs} < -Z_\alpha \Rightarrow a < a_0 \\ U_{qs} > Z_\alpha \Rightarrow a > a_0 \end{cases}$$

TH3: Chưa biết phương sai tổng thể $\sigma^2, n < 30$

B1. Tra ngưỡng $T_{\alpha}^{(n-1)}$

B2:

$$T_{qs} = \frac{(\bar{x} - a_0) \sqrt{n}}{S}$$

B3: Kết luận

$$|T_{qs}| \leq T_{\alpha}^{(n-1)} \Rightarrow H \text{ đúng : } a = a_0$$

$$|T_{qs}| > T_{\alpha}^{(n-1)} \Rightarrow H \text{ sai : } a \neq a_0$$

$$a \neq a_0 \begin{cases} T_{qs} < -T_{\alpha}^{(n-1)} \Rightarrow a < a_0 \\ T_{qs} > T_{\alpha}^{(n-1)} \Rightarrow a > a_0 \end{cases}$$

Ví dụ 3.1. Trọng lượng (X) của một loại sản phẩm do nhà máy sản xuất ra là đại lượng ngẫu nhiên có phân phối chuẩn với độ lệch chuẩn là $\sigma = 1kg$, trọng lượng trung bình là 50kg. Nghi ngờ máy hoạt động không bình thường làm thay đổi trọng lượng trung bình của sản phẩm, người ta cân thử 100 sản phẩm và thu được kết quả sau:

Trọng lượng sản phẩm(kg)	48	49	50	51	52
Số lượng sản phẩm	10	60	20	5	5

Với mức ý nghĩa 0.05, hãy kết luận về nghi ngờ nói trên.

. **Giải.** Ký hiệu a là trọng lượng trung bình của sản phẩm.

Ta kiểm định giả thiết : $H : a = a_0 = 50$

Vì $\sigma = 1$ nên đây là trường hợp 1

$$\bar{x} = 49,35$$

$$U_{qs} = \frac{(\bar{x} - a_0) \sqrt{n}}{\sigma} = \frac{(49,35 - 50) \sqrt{100}}{1} = -6,5 < -Z_{0,05} = -1,96$$

$$\Rightarrow a < a_0 = 50$$

Vậy máy đã hoạt động không bình thường làm giảm trọng lượng trung bình của sản phẩm.

Ví dụ 3.2.

Mức hao phí xăng(X) cho một loại xe ô tô chạy trên đoạn đường AB là một đại lượng ngẫu nhiên có phân phối chuẩn có kỳ vọng là 50 lít. Nay do đường được tu sửa lại, người ta cho rằng hao phí trung bình đã giảm xuống. Quan sát 36 chuyến xe chạy trên đường AB ta thu được bảng số liệu sau :

Mức hao phí(lít)	48,5-49,0	49,0-49,5	49,5-50,0	50,0-50,5	50,5-51,0
Số chuyến xe n_i	10	11	10	3	2

Với mức ý nghĩa $\alpha = 0,05$ hãy cho kết luận về ý kiến trên.

a mức hao phí xăng sau khi sửa lại đường
 a_0 mức hao phí xăng khi chưa sửa lại đường

$$H : a = a_0 = 50$$

Vì $n=36 > 30$ nên đây là trường hợp 2

$$Z_{0,05} = 1,96$$

$$\bar{x} = 49,4167; S = 0,573$$

$$U_{qs} = \frac{(\bar{x} - a_0) \sqrt{n}}{S} = \frac{(49,4167 - 50) \sqrt{36}}{0,573}$$

$$= -6,1 < -Z_{\alpha} = -1,96$$

$$\Rightarrow a < a_0$$

Vậy mức hao phí xăng trung bình đã giảm .

Ví dụ 3.3. Định mức để hoàn thành 1 sản phẩm là 14,5 phút. Có nên thay đổi định mức không, nếu theo dõi thời gian hoàn thành của 25 công nhân, ta có bảng số liệu sau:

Thời gian sản xuất một sản phẩm(phút)	10-12	12-14	14-16	16-18	18-20
Số công nhân tương ứng (n_i)	2	6	10	4	3

Hãy kết luận với mức ý nghĩa 0.05 biết rằng thời gian hoàn thành một sản phẩm (X) là một đại lượng ngẫu nhiên có phân phối chuẩn.

. Giải

$$H : a = a_0 = 14,5$$

$a_0 = 14,5$ là định mức cũ ,a là năng suất trung bình mới

$$n = 25 < 30 \Rightarrow TH3 \Rightarrow T_{0.05}^{(24)} = t_{24;0,025} = 2,064;$$

$$\bar{x} = 15; S = 2,236 \Rightarrow$$

$$T_{qs} = \frac{(\bar{x} - a_0) \sqrt{n}}{S} = \frac{(15 - 14,5) \sqrt{25}}{2,236} = 1,118$$

$$|T_{qs}| = 1,118 < 2.064 \Rightarrow a = a_0$$

Vậy không nên thay đổi định mức.

2. Bài toán 2 mẫu:

Kí hiệu trung bình của tổng thể 1,2 là μ_1, μ_2 (cả hai chưa biết). Từ các tổng thể lấy các mẫu kích thước n_1, n_2 có trung bình mẫu $\bar{x}_1, \bar{x}_2$ và phương sai hiệu chỉnh mẫu s_1^2, s_2^2 . Với mức ý nghĩa α , hãy kiểm định giả thiết:

$$H_0: \mu_1 = \mu_2$$

Trường hợp 1. Đã biết phương sai tổng thể σ_1^2, σ_2^2

B1: Tra ngưỡng Z_α

B2:

$$U_{qs} = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

B3. Kết luận

$$|U_{qs}| \leq Z_\alpha \Rightarrow H \text{ đúng} \Rightarrow a_1 = a_2$$

$$|U_{qs}| > Z_\alpha \Rightarrow H \text{ sai} \Rightarrow a_1 \neq a_2$$

$$a_1 \neq a_2 \begin{cases} U_{qs} < -Z_\alpha \Rightarrow a_1 < a_2 \\ U_{qs} > Z_\alpha \Rightarrow a_1 > a_2 \end{cases}$$

TH2: Chưa biết σ_1^2, σ_2^2 , n_1 và $n_2 \geq 30$

B1: : Tra ngưỡng Z_α

B2:

$$U_{qs} = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

B3: Kết luận

$$|U_{qs}| \leq Z_\alpha \Rightarrow H \text{ đúng} \Rightarrow a_1 = a_2$$

$$|U_{qs}| > Z_\alpha \Rightarrow H \text{ sai} \Rightarrow a_1 \neq a_2$$

$$a_1 \neq a_2 \begin{cases} U_{qs} < -Z_\alpha \Rightarrow a_1 < a_2 \\ U_{qs} > Z_\alpha \Rightarrow a_1 > a_2 \end{cases}$$

TH3: Chưa biết $\sigma_1^2, \sigma_2^2, n_1$ hoặc $n_2 < 30$

B1. Tra ngưỡng $T_\alpha^{(n_1+n_2-2)}$

B2.

$$T_{qs} = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

B3. $|T_{qs}| \leq T_\alpha^{(n_1+n_2-2)} \Rightarrow$ H đúng : $a_1 = a_2$

$|T_{qs}| > T_\alpha^{(n_1+n_2-2)} \Rightarrow$ H sai : $a_1 \neq a_2$

$$a_1 \neq a_2 \begin{cases} T_{qs} < -T_\alpha^{(n_1+n_2-2)} \Rightarrow a_1 < a_2 \\ T_{qs} > T_\alpha^{(n_1+n_2-2)} \Rightarrow a_1 > a_2 \end{cases}$$

Ví dụ 3.4: Người ta thí nghiệm 2 phương pháp chăn nuôi gà khác nhau, sau 1 tháng kết quả tăng trọng như sau:

Phương pháp	Số gà được theo dõi	Mức tăng trọng trung bình (kg)	Độ lệch tiêu chuẩn
I	100	1,2	0,2
II	150	1,3	0.3

Với mức ý nghĩa 0.05 có thể kết luận phương pháp II hiệu quả hơn phương pháp I không?

Giải: $n_1 = 100, n_2 = 150, \sigma_1 = 0,2, \sigma_2 = 0,3, \bar{x}_1 = 1,2, \bar{x}_2 = 1,3$

a_1 - Mức tăng trong trung bình của phương pháp I

a_2 - Mức tăng trọng trung bình của phương pháp II

$$H: a_1 = a_2$$

$$Z_\alpha = 1,96$$

$$U_{qs} = \frac{1,2 - 1,3}{\sqrt{\frac{0,04}{100} + \frac{0,09}{150}}} = -3,16 < -Z_\alpha \Rightarrow a_1 < a_2$$

Vậy phương pháp 2 hiệu quả hơn phương pháp 1

Ví dụ 3.5: Tương tự ví dụ trên nhưng thay bằng số liệu sau

$$n_1 = 10; n_2 = 15; S_1 = 0,2; S_2 = 0,3$$

$$\overline{x}_1 = 1,2 \quad ; \quad \overline{x}_2 = 1,3$$

$$T_{qs} = \frac{1,2 - 1,3}{\sqrt{\frac{0,2^2}{10} + \frac{0,3^2}{15}}} = -1$$

$$|T_{qs}| < T_{0,05}^{(23)} = 2,069 \Rightarrow a_1 = a_2$$

Vậy hai phương pháp hiệu quả như nhau.

§4. Kiểm định giả thiết về phương sai

Bài toán: Kí hiệu phương sai của tổng thể là σ^2 , từ tổng thể lấy 1 mẫu kích thước n có phương sai hiệu chỉnh mẫu S^2 , với mức ý nghĩa α , hãy kiểm định giả thiết: $H: \sigma^2 = \sigma_0^2$

B1: Tra bảng $\chi^2_{1-\frac{\alpha}{2}}(n-1) < \chi^2_{\frac{\alpha}{2}}(n-1)$

B2: $\chi^2_{qs} = \frac{(n-1).S^2}{\sigma_0^2}$

B3: Kết luận: $\chi^2_{1-\frac{\alpha}{2}}(n-1) \leq \chi^2_{qs} \leq \chi^2_{\frac{\alpha}{2}}(n-1) \Rightarrow \sigma^2 = \sigma_0^2$

$$\left\{ \begin{array}{l} \chi^2_{qs} < \chi^2_{1-\frac{\alpha}{2}}(n-1) \Rightarrow \sigma^2 < \sigma_0^2 \\ \chi^2_{qs} > \chi^2_{\frac{\alpha}{2}}(n-1) \Rightarrow \sigma^2 > \sigma_0^2 \end{array} \right.$$

Ví dụ 4.1.

Chọn ngẫu nhiên 27 vòng bi cùng loại thì thấy độ lệch trung bình $S=0.003$. Theo số liệu quy định thì độ lệch chuẩn cho phép không vượt quá 0.0025. Với mức ý nghĩa 0.05, hãy cho kết luận?

.Giải :

$$H : \sigma^2 = \sigma_0^2 = 0,0025^2$$

$$n = 27, \chi_{0.975}^2(26) = 13,84 ; \chi_{0.025}^2(26) = 41,92,$$

$$\chi_{qs}^2 = \frac{26 \cdot 0,003^2}{0,0025^2} = 37,44$$

$$\Rightarrow 13,84 < \chi_{qs}^2 < 41,92 \Rightarrow \sigma^2 = \sigma_0^2$$

Vậy lô vòng bi này chưa vượt mức cho phép về độ phân tán

§5. Kiểm định giả thiết về quy luật phân phối.

Bài toán: Giả sử đại lượng ngẫu nhiên gốc X của tổng thể chưa rõ phân phối. Từ tổng thể lấy một mẫu kích thước n . Với mức ý nghĩa α hãy kiểm định giả thiết :

H : X có phân phối $F(x)$

I. $F(x)$ là phân phối rời rạc

Giả sử bảng phân phối tần số mẫu có dạng

X	x_1	x_2	$\dots$	x_k
n_i	n_1	n_2	$\dots$	n_k

B1: Ký hiệu r là số tham số chưa biết của phân phối $F(x)$, ta thay các tham số đó bằng các ước lượng hợp lý tối đa.

B2: Tra $\chi^2_{\alpha}(k - r - 1)$

B3: Tính $p_i = P(X = x_i)$, nếu $X \sim F(x), i = 1, 2, \dots, k$

B4: Tính giá trị quan sát

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

B5: Kết luận:

$$\chi^2_{qs} \leq \chi^2_{\alpha}(k - r - 1) \Rightarrow \text{H đúng: } X \text{ có phân phối } F(X)$$

$$\chi^2_{qs} > \chi^2_{\alpha}(k - r - 1) \Rightarrow \text{H sai: } X \text{ không có phân phối } F(X)$$

1. Kiểm định giả thiết về phân phối đều rời rạc

$H: X$ có phân phối đều rời rạc

B1. $r = 0$ (do phân phối đều không có tham số chưa biết)

B2. $\chi^2_{\alpha}(k - 1)$

B3. $p_i = \frac{1}{k}, i = \overline{1, k}$

B4. $\chi^2_{qs} = \sum_{i=1}^k \frac{\left(n_i - \frac{n}{k}\right)^2}{\left(\frac{n}{k}\right)} = \frac{1}{nk} \sum_{i=1}^k (n_i \cdot k - n)^2$

B5. Kết luận : Theo bài toán chung như trên

Ví dụ 5.1. Tung 1 con xúc xắc ta được bảng điểm sau đây:

Số điểm	1	2	3	4	5	6
Số lần	3	7	6	5	6	4

Với mức ý nghĩa 0.05 ,hãy kết luận con xúc sắc trên có đều hay không?

Giải:

$$\chi_{qs}^2 = \frac{1}{31.6} \left[(3.6-31)^2 + (7.6-31)^2 + (6.6-31)^2 + (5.6-31)^2 + (4.6-31)^2 \right]$$

$$= 2.1 < \chi_{0.05}^2(5) = 11.07$$

Vậy con xúc xắc đều

2. Kiểm định giả thiết về phân phối Poisson.

$$H : X \sim P(a)$$

X	0	1	2	...	$k-1$
n_i	n_0	n_1	n_2	...	n_{k-1}

B1. $r=1$ (có 1 tham số chưa biết là a), ta thay $a = \bar{x}$

B2. $\chi^2_{\alpha}(k-2)$

B3. $p_i = p(X = i) = e^{-a} \frac{a^i}{i!}, i = \overline{0, k-1}$

B4. $\chi^2_{qs} = \sum_{i=0}^{k-1} \frac{(n_i - np_i)^2}{np_i}$

B5. Kết luận : Như b5 ở bài trên

Ví dụ 5.2:

Để kiểm tra công việc của 200 công nhân, người ta chọn ngẫu nhiên 1000 sản phẩm của mỗi người đem đi thử nghiệm để tìm ra phế phẩm. Kết quả như sau:

Số phế phẩm trên 1000 sản phẩm	0	1	2	3	4
Số công nhân	109	65	22	3	1

Với mức ý nghĩa 0.01, có thể coi mẫu trên phù hợp với phân phối Poisson hay không ?

Giải: $a = \bar{x} = 0,61$

$$n p_i = 200 . e^{-\bar{x}} . \frac{\bar{x}^i}{i!}, i = \overline{0, 4}$$

i	n_i	np_i
0	109	A=108,67
1	65	B=66,29
2	22	C=20,21
3	3	D=4,111
4	1	E=0,627

$$n.e^{-\bar{x}} = np_0, np_i = np_{i-1}.\bar{x} : i, \quad i = 1, 2, \dots, k-1$$

$$n.e^{-\bar{x}} = np_0 \rightarrow A, \times \bar{x} = np_1 \rightarrow B,$$

$$\times \bar{x} : 2 \rightarrow C, \times \bar{x} : 3 \rightarrow D, \times \bar{x} : 4 \rightarrow E$$

$$\Rightarrow \chi_{qs}^2 = \frac{(109-A)^2}{A} + \frac{(65-B)^2}{B} + \frac{(22-C)^2}{C} + \frac{(3-D)^2}{D} + \frac{(1-E)^2}{E} = 0,705 < \chi_{0.01}^2(3) = 11,34$$

Vậy mẫu trên phù hợp với phân phối Poisson.

$$n.e^{-\bar{x}} = np_0 \rightarrow A \quad (109-A)^2 : A \text{ SH STOB}$$

$$A \times \bar{x} = np_1 \rightarrow A \quad B+(65-A)^2 : A \text{ SH STOB}$$

$$A \times \bar{x} : 2 = np_2 \rightarrow A \quad B+(22-A)^2 : A \text{ SH STOB}$$

$$A \times \bar{x} : 3 = np_3 \rightarrow A \quad B+(3-A)^2 : A \text{ SH STOB}$$

$$A \times \bar{x} : 4 = np_4 \rightarrow A \quad B+(1-A)^2 : A \text{ SH STOB}$$

$$\Rightarrow \chi_{qs}^2 = B = 0,705 < \chi_{0.01}^2(3) = 11,34$$

Vậy mẫu trên phù hợp với phân phối Poisson.

II. Trường hợp $F(x)$ liên tục: $H : X \sim F(x)$

Giả sử bảng phân phối tần số mẫu có dạng:

X	(a_0, a_1)	(a_1, a_2)	$\dots$	(a_{k-1}, a_k)
n_i	n_1	n_2	$\dots$	n_k

B1. Ký hiệu r là số tham số chưa biết. Thay các tham số đó bằng các ước lượng hợp lý tối đa của chúng.

B2. Tra $\chi^2_\alpha(k-r-1)$

B3. Tính

$$p_1 = P(-\infty < X < a_1),$$
$$p_2 = P(a_1 < X < a_2), \dots, \quad \text{nếu } X \sim F(x)$$
$$p_{k-1} = P(a_{k-2} < X < a_{k-1})$$
$$p_k = P(a_{k-1} < X < +\infty)$$

Chú ý:

$$\sum_i p_i = 1$$

B4.

$$\chi^2_{qs} = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i} = \left(\sum_{i=1}^k \frac{n_i^2}{p_i} \right) \cdot \frac{1}{n} - n$$

B5. Kết luận : Giống trường hợp F(x) rời rạc.

Kiểm định về phân phối chuẩn.

$$H: X \sim N(a, \sigma^2)$$

X	(a_0, a_1)	(a_1, a_2)	...	(a_{k-1}, a_k)
n_i	n_1	n_2	...	n_k

Giải

B1: $r = 2, a = \bar{x}, \sigma = \hat{S} = \frac{\sigma}{\sqrt{n}}$

B2. $\chi^2_{\alpha} (k - 3)$

B3.

$$p_1 = \Phi \left(\frac{a_1 - \bar{x}}{\sigma} \right) + 0,5$$

$$p_2 = \Phi \left(\frac{a_2 - \bar{x}}{\sigma} \right) - \Phi \left(\frac{a_1 - \bar{x}}{\sigma} \right), \dots,$$

$$p_{k-1} = \Phi \left(\frac{a_{k-1} - \bar{x}}{\sigma} \right) - \Phi \left(\frac{a_{k-2} - \bar{x}}{\sigma} \right)$$

$$p_k = 0,5 - \Phi \left(\frac{a_{k-1} - \bar{x}}{\sigma} \right)$$

B4.

$$\chi_{qs}^2 = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i} = \left(\sum_{i=1}^k \frac{n_i^2}{p_i} \right) \cdot \frac{1}{n} - n$$

B5. Kết luận như b5 bài toán chung

B3.

$$p_1 = \overline{\Phi} \left(\frac{a_1 - \bar{x}}{\sigma} \right)$$

$$p_2 = \overline{\Phi} \left(\frac{a_2 - \bar{x}}{\sigma} \right) - \overline{\Phi} \left(\frac{a_1 - \bar{x}}{\sigma} \right), \dots,$$

$$p_{k-1} = \overline{\Phi} \left(\frac{a_{k-1} - \bar{x}}{\sigma} \right) - \overline{\Phi} \left(\frac{a_{k-2} - \bar{x}}{\sigma} \right)$$

$$p_k = 1 - \overline{\Phi} \left(\frac{a_{k-1} - \bar{x}}{\sigma} \right)$$

B4.

$$\chi_{qs}^2 = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i} = \left(\sum_{i=1}^k \frac{n_i^2}{p_i} \right) \cdot \frac{1}{n} - n$$

B5. Kết luận như b5 bài toán chung

Chú ý : Nếu cho bảng số liệu

Số điểm	1	3	5	7	9
Số học sinh	6	24	43	16	11

hoặc

Số điểm	$-\infty-2$	2-4	4-6	6-8	8- ∞
Số học sinh	6	24	43	16	11

thì ta chia khoảng như sau

Số điểm	0-2	2-4	4-6	6-8	8-10
Số học sinh	6	24	43	16	11

Chú ý : Nếu bảng số liệu như sau

X	2	6	10	14
X	(0-4)	(4-8)	(8-12)	(12-16)
n_i	1	4	5	2

thì hàng đầu dùng tính bước 1, hàng 2 dùng tính bước 3.

Ví dụ 5.3 : Bảng điểm của 1 lớp học như sau

Số điểm	0-2	2-4	4-6	6-8	8-10
Số học sinh	6	24	43	16	11

Với $\alpha = 0,05$ hãy kết luận bảng điểm này có phù hợp với phân phối chuẩn hay không?

$$\bar{x} = 5,04 \quad ; \quad \sigma = x\sigma n = 2,078$$

$$p_1 = \Phi\left(\frac{2-\bar{x}}{\sigma}\right) + 0,5 = \Phi(-1,46) + 0,5 = -0,42825 + 0,5 = 0,07175$$

$$p_2 = \Phi\left(\frac{4-\bar{x}}{\sigma}\right) - \Phi\left(\frac{2-\bar{x}}{\sigma}\right) = -\Phi(0,50) + 0,42825 = -0,19163 + 0,42825 = 0,23662$$

$$p_3 = \Phi\left(\frac{6-\bar{x}}{\sigma}\right) - \Phi\left(\frac{4-\bar{x}}{\sigma}\right) = \Phi(0,46) + 0,19163 = 0,17795 + 0,19163 = 0,36958$$

$$p_4 = \Phi\left(\frac{8-\bar{x}}{\sigma}\right) - \Phi\left(\frac{6-\bar{x}}{\sigma}\right) = \Phi(1,424) - 0,23565 = 0,42283 - 0,17795 = 0,24488$$

$$p_5 = 0,5 - \Phi\left(\frac{8-\bar{x}}{\sigma}\right) = 0,5 - 0,42283 = 0,07717$$

$$\chi_{qs}^2 = \left(\frac{6^2}{p_1} + \frac{24^2}{p_2} + \frac{43^2}{p_3} + \frac{16^2}{p_4} + \frac{11^2}{p_5} \right) : 100 - 100$$

$$= (10552,37815 : 100 - 100 = 5,5237 \leq \chi_{0,05}^2(2) = 6$$

Chú ý: Yêu cầu trình bày như sau $\bar{x} = 5,04$; $\sigma = x\sigma n = 2,078$

$$p_1 = \Phi \left(\frac{2 - \bar{x}}{\sigma} \right) + 0,5 = 0,07175$$

$$p_2 = \Phi \left(\frac{4 - \bar{x}}{\sigma} \right) - \Phi \left(\frac{2 - \bar{x}}{\sigma} \right) = 0,23662$$

$$p_3 = \Phi \left(\frac{6 - \bar{x}}{\sigma} \right) - \Phi \left(\frac{4 - \bar{x}}{\sigma} \right) = 0,36958$$

$$p_4 = \Phi \left(\frac{8 - \bar{x}}{\sigma} \right) - \Phi \left(\frac{6 - \bar{x}}{\sigma} \right) = 0,24488$$

$$p_5 = 0,5 - \Phi \left(\frac{8 - \bar{x}}{\sigma} \right) = 0,07717$$

$$\chi_{qs}^2 = \left(\frac{6^2}{p_1} + \frac{24^2}{p_2} + \frac{43^2}{p_3} + \frac{16^2}{p_4} + \frac{11^2}{p_5} \right) : 100 - 100 = 5,5237 \leq \chi_{0.05}^2(2) = 6$$

Chú ý:

$$Q(u) = |\Phi(u)|$$

$$(2 - \bar{x}) : x\sigma n = u_1 \quad (u_1 < 0 \Rightarrow \Phi(u_1) = -Q(u_1))$$

$$Q(ans) = \text{SH STO } X \quad -X + 0,5 = p_1 \quad 6^2 : ans = \text{SH STO } A$$

$$(4 - \bar{x}) : x\sigma n = u_2 \quad (u_2 < 0 \Rightarrow \Phi(u_2) = -Q(u_2))$$

$$Q(ans) = \text{SH STO } Y \quad -Y + X = p_2 \quad 24^2 : ans + A = \text{SH STO } A$$

$$(6 - \bar{x}) : x\sigma n = u_3 \quad (u_3 > 0 \Rightarrow \Phi(u_3) = Q(u_3))$$

$$Q(ans) = \text{SH STO } X \quad X + Y = p_3 \quad 43^2 : ans + A = \text{SH STO } A$$

$$(8 - \bar{x}) : x\sigma n = u_4 \quad (u_4 > 0 \Rightarrow \Phi(u_4) = Q(u_4))$$

$$Q(ans) = \text{SH STO } Y \quad Y - X = p_4 \quad 16^2 : ans + A = \text{SH STO } A$$

$$0,5 - Y = p_5 \quad 11^2 : ans + A = \text{SH STO } A$$

$$\chi_{qs}^2 = A : 100 - 100 = 5,5237 \leq \chi_{0.05}^2(2) = 6$$

Chú ý: Yêu cầu trình bày như sau $\bar{x} = 5,04$; $\sigma = x\sigma n = 2,078$

$$p_1 = \overline{\Phi} \left(\frac{2 - \bar{x}}{\sigma} \right) = 0,07175$$

$$p_2 = \overline{\Phi} \left(\frac{4 - \bar{x}}{\sigma} \right) - \overline{\Phi} \left(\frac{2 - \bar{x}}{\sigma} \right) = 0,23662$$

$$p_3 = \overline{\Phi} \left(\frac{6 - \bar{x}}{\sigma} \right) - \overline{\Phi} \left(\frac{4 - \bar{x}}{\sigma} \right) = 0,36958$$

$$p_4 = \overline{\Phi} \left(\frac{8 - \bar{x}}{\sigma} \right) - \overline{\Phi} \left(\frac{6 - \bar{x}}{\sigma} \right) = 0,24488$$

$$p_5 = 1 - \overline{\Phi} \left(\frac{8 - \bar{x}}{\sigma} \right) = 0,07717$$

$$\chi_{qs}^2 = \left(\frac{6^2}{p_1} + \frac{24^2}{p_2} + \frac{43^2}{p_3} + \frac{16^2}{p_4} + \frac{11^2}{p_5} \right) : 100 - 100 = 5,5237 \leq \chi_{0.05}^2(2) = 6$$

Chú ý:

$$P(u) = \overline{\Phi}(u) = \Phi(u) + 0,5$$

$$P((2 - \bar{x}) : x\sigma n) = p_1 \quad \text{SH STO } X \quad 6^2 : \text{ans} = \text{SH STO } A$$

$$P((4 - \bar{x}) : x\sigma n) = \quad \text{SH STO } Y \quad Y - X = p_2$$

$$24^2 : \text{ans} + A = \text{SH STO } A$$

$$P((6 - \bar{x}) : x\sigma n) = \quad \text{SH STO } X$$

$$X - Y = p_3 \quad 43^2 : \text{ans} + A = \text{SH STO } A$$

$$P((8 - \bar{x}) : x\sigma n) = \quad \text{SH STO } Y$$

$$Y - X = p_4 \quad 16^2 : \text{ans} + A = \text{SH STO } A$$

$$1 - Y = p_5 \quad 11^2 : \text{ans} + A = \text{SH STO } A$$

$$\chi_{qs}^2 = A : 100 - 100 = 5,5237 \leq \chi_{0.05}^2(2) = 6$$

Vậy bảng điểm này có phù hợp với phân phối chuẩn

§6. Bảng phân phối tần số mẫu đồng thời hay bảng tương quan mẫu

Giả sử X, Y là 2 đại lượng ngẫu nhiên gốc của cùng 1 tổng thể. Bảng phân phối tần số mẫu đồng thời của X, Y là:

$X \backslash Y$	y_1	y_2	$\dots$	y_h
x_1	n_{11}	n_{12}	$\dots$	n_{1h}
x_2	n_{21}	n_{22}	$\dots$	n_{2h}
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
x_k	n_{k1}	n_{k2}	$\dots$	n_{kh}

Bảng phân phối tần số lẻ :

$X \backslash Y$	y_1	y_2	$\dots$	y_h	n_i
x_1	n_{11}	n_{12}	$\dots$	n_{1h}	n_1
x_2	n_{21}	n_{22}	$\dots$	n_{2h}	n_2
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
x_k	n_{k1}	n_{k2}	$\dots$	n_{kh}	n_k
m_j	m_1	m_2	$\dots$	m_h	n

x_i	y_j	n_{ij}
x_1	y_1	n_{11}
x_1	y_2	n_{12}
...	...	...
x_k	y_h	n_{kh}

§7. Kiểm định tính độc lập.

BÀI TOÁN. Giả sử X, Y là 2 đại lượng ngẫu nhiên gốc của cùng 1 tổng thể. Từ tổng thể lấy 1 mẫu kích thước n . Với mức ý nghĩa α hãy kiểm định giả thiết :

H: X, Y độc lập

B1. Tra bảng

$$\chi^2_{\alpha} \left[(k-1)(h-1) \right]$$

B2. Tính

$$\chi^2_{qs} = \sum_{i,j} \frac{(n_{ij} - \gamma_{ij})^2}{\gamma_{ij}}, \quad \text{với } \gamma_{ij} = \frac{n_i \cdot m_j}{n}$$

$$\text{hoặc } \chi^2_{qs} = \left[\left(\sum_{i,j} \frac{n_{ij}^2}{n_i \cdot m_j} \right) - 1 \right] \cdot n$$

B3.Kết luận:

$$\chi_{qs}^2 \leq \chi_{\alpha}^2 [(k-1)(h-1)] \Leftrightarrow X, Y \text{ độc lập}$$

$$\chi_{qs}^2 > \chi_{\alpha}^2 [(k-1)(h-1)] \Leftrightarrow X, Y \text{ phụ thuộc}$$

cuu duong than cong . com

Chú ý : người ta chứng minh được rằng chỉ khi

$$\gamma_{ij} \geq 5, \forall i, j$$

cuu duong than cong . com

thì tiêu chuẩn khi bình phương mới có thể cho một lời giải chính xác.

Ví dụ.7.1: Nghiên cứu ảnh hưởng của hoàn cảnh gia đình đối với tình trạng phạm tội của trẻ em có kết quả:

Tình trạng phạm tội	Bố mẹ đã mất	Bố mẹ ly hôn	Còn cả bố mẹ
Không phạm tội	20	25	13
Có phạm tội	29	43	18

Với mức ý nghĩa 0,05 có thể kết luận hoàn cảnh gia đình không ảnh hưởng tới tình trạng phạm tội hay không ?

Giải:

Tình trạng phạm tội	Bố mẹ đã mất	Bố mẹ ly hôn	Còn cả bố mẹ	n_i
Không phạm tội	20	25	13	58
Có phạm tội	29	43	18	90
m_j	49	68	31	148

$$\chi^2_{qs} = \left[\frac{20^2}{58 \cdot 49} + \frac{25^2}{58 \cdot 68} + \dots + \frac{18^2}{90 \cdot 31} - 1 \right] \cdot 148 = 0,32 < \chi^2_{0,05}(2) = 6$$

Vậy hoàn cảnh gia đình không ảnh hưởng tới tình trạng phạm tội