Chương 7
KIỂM ĐỊNH GIẢ THIẾT THỐNG KÊ

Dùng các thống kê từ một mẫu để khẳng định hay bác bỏ một giả thiết nào đó nói về tổng thể gọi là kiểm định giả thiết thống kê.

Giả sử cần kiểm định một giả thiết H. Khi kiểm định có thể xảy ra một trong hai loại sai lầm sau đây:

- Loại 1: bác bỏ H trong lúc H đúng
- Loại 2: chấp nhận H trong lúc H sai.

Phương pháp chung để kiểm định là cho phép xác suất xảy ra sai lầm loại 1 không quá α, số α gọi là mức ý nghĩa. Với mức ý nghĩa đã cho, ta chấp nhận H nếu xác suất xảy ra sai lầm loại 2 nhỏ nhất.

7.1. KIỂM ĐỊNH TỶ LỆ

1- Kiểm định giả thiết về tỷ lệ tổng thể

Bài toán. Giả sử tổng thể có tỷ lệ p. Mẫu có kích thước n, tỷ lệ mẫu f. Hãy kiểm định giả thiết H: $p = p_o$ với mức ý nghĩa α.

Phương pháp giải

Giả thiết thêm $n \geq 30$. Nếu H đúng thì:

$$\frac{F - p_o}{\sqrt{p_o(1 - p_o)}}\sqrt{n} \sim N(0,1)$$

$$P\left(\frac{|F - p_o|}{\sqrt{p_o(1 - p_o)}}\sqrt{n} \leq Z_\alpha\right) = 1 - \alpha$$

Từ đó ta có quy tắc kiểm định:

- Tìm Z_α từ hệ thức $2\Phi(Z_\alpha) = 1 - \alpha$

- Tính thống kê

$$Z_o = \frac{|f - p_o|}{\sqrt{p_o(1 - p_o)}}.\sqrt{n}$$

Nếu $Z_o \leq Z_\alpha$ thì chấp nhận H

Nếu $Z_o > Z_\alpha$ thì bác bỏ H.

Ví dụ 7.1. Theo báo cáo, tỷ lệ hàng phế phẩm trong kho là 12%. Kiểm tra ngẫu nhiên 100 sản phẩm, thấy có tám phế phẩm. Với mức ý nghĩa 5%, báo cáo trên có đáng tin không?

Giải. Gọi p là tỷ lệ phế phẩm trong kho, $p_o = 0,12$ là tỷ lệ phế phẩm giả thiết. Mẫu có kích thước n = 100, tỷ lệ mẫu f = 0,08.

Giả thiết H: $p = p_o = 0,12$ (báo cáo đáng tin).

$$2\Phi(Z_\alpha) = 1 - 0,05 \Rightarrow Z_\alpha = 1,96$$

$$Z_o = \frac{|0,08 - 0,12|}{\sqrt{0,12.0,88}}\sqrt{100} = 1,23$$

Vì $Z_o < Z_\alpha$ nên chấp nhận H. Vậy báo cáo đáng tin với mức ý nghĩa $\alpha = 5\%$.

2- Kiểm định so sánh hai tỷ lệ

Bài toán. Giả sử tổng thể I có tỷ lệ p_1; tổng thể II có tỷ lệ p_2. Từ tổng thể I có mẫu kích thước n_1, tỷ lệ mẫu f_1. Từ tổng thể II có mẫu kích thước n_2, tỷ lệ mẫu f_2. Hãy kiểm định giả thiết H: $p_1 = p_2$ với mức ý nghĩa α.

Phương pháp giải

Giả thiết thêm $n_1, n_2 \geq 30$. Ta có quy tắc kiểm định như sau:

- Tìm Z_α từ hệ thức $2\Phi(Z_\alpha) = 1 - \alpha$

- Tính thống kê: $Z_o = \dfrac{|f_1 - f_2|}{\sqrt{p_o(1-p_o)\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}}$; $p_o = \dfrac{n_1 f_1 + n_2 f_2}{n_1 + n_2}$

Nếu $Z_o \le Z_\alpha$ thì chấp nhận H

Nếu $Z_o > Z_\alpha$ thì bác bỏ H.

Ví dụ 7.2. Kiểm tra 100 sản phẩm ở kho I, thấy có sáu phế phẩm. Kiểm tra 200 sản phẩm ở kho II, thấy có 24 phế phẩm. Chất lượng hàng ở hai kho có khác nhau không với mức ý nghĩa 5%.

Giải. Gọi p_1 là tỷ lệ phế phẩm ở kho I, p_2 là tỷ lệ phế phẩm ở kho II. Ta có $n_1 = 100$; $f_1 = 0,06$; $n_2 = 200$; $f_2 = 0,012$.

Giả thiết H: $p_1 = p_2$ (chất lượng hàng hai kho giống nhau).

$$2\Phi(Z_\alpha) = 1 - \alpha \Rightarrow Z_\alpha = 1,96$$

$$p_o = \frac{6+24}{100+200} = 0,1 \Rightarrow Z_o = \frac{|0,06 - 0,12|}{\sqrt{0,1.0,9\left(\dfrac{1}{100}+\dfrac{1}{200}\right)}} = 1,63$$

Vì $Z_o < Z_\alpha$ nên chấp nhận H. Vậy chất lượng hàng ở hai kho là giống nhau ở mức ý nghĩa 5%.

7.2. KIỂM ĐỊNH GIÁ TRỊ TRUNG BÌNH

1. Kiểm định giả thiết về trung bình của tổng thể

Bài toán. Giả sử tổng thể có trung bình (kỳ vọng) a. Mẫu có kích thước n, trung bình mẫu $\overline{x}$, phương sai mẫu hiệu chỉnh s^2. Hãy kiểm định giả thiết H: $a = a_o$ với mức ý nghĩa α.

1- Trường hợp n ≥ 30

Nếu H đúng thì: $$\frac{\overline{X} - a_o}{s}\sqrt{n} \sim N(0,1)$$

$$P\left(\frac{\left|\overline{X} - a_o\right|}{s}\sqrt{n} \le Z_\alpha\right) = 1 - \alpha$$

Từ đó ta có quy tắc kiểm định sau:

- Tìm Z_α từ hệ thức $2\Phi(Z_\alpha) = 1 - \alpha$

- Tính thống kê $$Z_o = \frac{\left|\overline{x} - a_o\right|}{s}\sqrt{n}$$

Nếu $Z_o \le Z_\alpha$ thì chấp nhận H

Nếu $Z_o > Z_\alpha$ thì bác bỏ H.

Ví dụ 7.3. Điểm trung bình môn toán năm học trước là 5,72. Năm nay theo dõi 100 sinh viên được số liệu:

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

Với mức ý nghĩa 1%, phải chăng điểm trung bình môn toán năm nay cao hơn năm trước?

Giải. Gọi a là điểm trung bình môn toán của năm nay. Từ mẫu đã cho ta có n = 100; = 5,90; s = 1,21.

Giả thiết H: $a = a_o = 5,72$ (điểm năm nay bằng năm trước)

$$2\Phi(Z_\alpha) = 1 - 0,01 \Rightarrow Z_\alpha = 2,58$$

$$Z_o = \frac{|5,90 - 5,72|}{1,21}\sqrt{100} = 1,49$$

Vì $Z_o < Z_\alpha$ nên chấp nhận H. Vậy điểm môn toán năm nay không cao hơn năm trước với mức ý nghĩa 1%.

2- Trường hợp tổng thể có phân phối chuẩn đã biết phương sai σ^2

Trường hợp này được kiểm định như trường hợp 1 với:

$$Z_o = \frac{|\bar{x} - a_o|}{\sigma}\sqrt{n}$$

Chú ý rằng trường hợp này không phân biệt n.

3- Trường hợp n < 30, tổng thể có phân phối chuẩn, chưa biết phương sai

Nếu H đúng thì: $\dfrac{\overline{X}-a_o}{s}\sqrt{n} \sim T(n-1)$

$$P\left(\dfrac{\left|\overline{X}-a_o\right|}{s}\sqrt{n} \le T_\alpha(n-1)\right) = 1-\alpha$$

Từ đó ta có quy tắc kiểm định:

- Tìm $T_\alpha = T_\alpha(n-1)$ từ bảng phân phối Student
- Tính thống kê $T_o = \dfrac{\left|\overline{x}-a_o\right|}{s}\sqrt{n}$

Nếu $T_o \le T_\alpha$ thì chấp nhận H

Nếu $T_o > T_\alpha$ thì bác bỏ H.

Ví dụ 7.4. Một vườn ươm cây giống, theo quy định khi nào cây cao trung bình trên 1 m thì đem ra trồng. Đo ngẫu nhiên 25 cây, được số liệu:

Chiều cao	0,8	0,9	1,0	1,1	1,2	1,3
Số cây	1	2	9	7	4	2

Với mức ý nghĩa 5%, có thể đem cây ra trồng được chưa? (giả thiết chiều cao của cây theo luật phân phối chuẩn).

Giải. Gọi a là chiều cao trung bình của cây trong vườn ươm. Từ mẫu ta có:

$$\overline{x} = 1{,}068; \quad s = 0{,}122$$

Giả thiết H: $a = a_o = 1$ (chưa nên đem cây ra trồng).

- $T_\alpha = T0{,}05(24) = 2{,}064$

- $T_o = \dfrac{|1{,}068 - 1|}{0{,}122}\sqrt{25} = 2{,}787.$

Vì $T_o > T_\alpha$ nên bác bỏ H. $D_o > a_o$, nên ta kết luận nên đem cây ra trồng.

2. Kiểm định so sánh hai trung bình

Bài toán. Giả sử tổng thể I có trung bình a_1; tổng thể II có trung bình a_2. Từ tổng thể I có mẫu kích thước n_1, trung bình mẫu $\overline{x}_1$, phương sai mẫu hiệu chỉnh s_1^2. Từ tổng thể II có mẫu kích thước n_2, trung bình mẫu $\overline{x}_2$, phương sai mẫu hiệu chỉnh s_2^2. Hãy kiểm định giả thiết H: $a_1 = a_2$ với mức ý nghĩa α.

1- Trường hợp $n_1, n_2 \geq 30$

Ta có quy tắc kiểm định như sau:

- Tìm Z_α từ hệ thức $2\Phi(Z_\alpha) = 1 - \alpha$

- Tính thống kê $Z_o = \dfrac{\left|\overline{x_1} - \overline{x}\right|}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$

Nếu Zo ≤ Zα thì chấp nhận H

Nếu Zo > Zα thì bác bỏ H.

Ví dụ 7.5. Điểm trung bình học tập của 50 học sinh lớp 4 trường A là 6,72; phương sai mẫu hiệu chỉnh là (0,72)2. Điểm trung bình học tập của 80 học sinh lớp 4 trường B là 6,46; phương sai mẫu hiệu chỉnh là (0,91)2. Phải chăng điểm trung bình học tập của học sinh của hai trường là khác nhau với mức ý nghĩa 5%?

Giải. Gọi a1 là điểm trung bình học tập của học sinh lớp 4 trường A, a2 là điểm trung bình học tập của học sinh lớp 4 trườngB. Từ số liệu ta có n1 = 50;$\overline{x}_1$= 6,27;s_1^2= (0,72)2; n2 = 80;$\overline{x}_2$= 6,64;s_2^2 = (0,91)2 .

Giả thiết H: a1 = a2 (điểm trung bình hai trường bằng nhau).

$$2\Phi(Z_\alpha) = 1 - 0{,}05 \Rightarrow Z_\alpha = 1{,}96$$

- $$Z_o = \frac{|6{,}72 - 6{,}46|}{\sqrt{\frac{(0{,}72)^2}{50} + \frac{(0{,}91)^2}{80}}} = 1{,}32$$

Vì $Z_0 < Z_\alpha$ nên chấp nhận H. Vậy chưa kết luận được điểm trung bình học tập của học sinh lớp 4 hai trường là khác nhau với mức ý nghĩa 5%.

2- Trường hợp n1 < 30 hoặc n2 < 30

Giả thiết thêm các tổng thể có phân phối chuẩn. Ta có quy tắc kiểm định sau:

- Tìm $T_\alpha = T_\alpha(n_1 + n_2 - 2)$ từ bảng phân phối Student

- Tính thống kê $$T_o = \frac{|\overline{x_1} - \overline{x_2}|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Nếu To ≤ Tα thì chấp nhận H

Nếu To > Tα thì bác bỏ H.

Ví dụ 7.6. Trọng lượng của một loại sản phẩm theo luật phân phối chuẩn. Quan sát một số sản phẩm do máy I và máy II sản xuất ta được số liệu tương ứng như sau:

Máy I	Trọng lượng (g)	9,0	9,5	10,0	10,5
	Số sản phẩm	2	4	7	2
Máy II	Trọng lượng (g)	9,0	9,5	10,0	10,5
	Số sản phẩm	1	4	6	3

Với mức ý nghĩa 5%, phải chăng trọng lượng trung bình của sản phẩm do hai máy sản xuất là khác nhau?

Giải. Gọi a1, a2 là trọng lượng trung bình của sản phẩm tương ứng do máy I, máy II sản xuất. Từ số liệu ta có:

$$n_1 = 15; \quad \overline{x}_1 = 9{,}80; \quad s_1^2 = 0{,}207$$

$$n_2 = 14; \quad \overline{x}_2 = 9{,}89; \quad s_2^2 = 0{,}199$$

Giả thiết H: a1 = a2 (trọng lượng sản phẩm hai nhà máy bằng nhau).

$$T_\alpha = T_{0,05}(27) = 2{,}052$$

$$T_o = \frac{|9{,}80 - 9{,}89|}{\sqrt{\frac{0{,}207}{15} + \frac{0{,}199}{14}}} = 0{,}538$$

Vì $T_o < T_\alpha$ nên chấp nhận H. Vậy trọng lượng trung bình sản phẩm do hai máy sản xuất là không khác nhau với α = 5%.

7.3. KIỂM ĐỊNH PHƯƠNG SAI

Bài toán. Giả sử tổng thể có phân phối chuẩn, phương sai σ2. Mẫu có kích thước n, phương sai mẫu hiệu chỉnh s2. Hãy kiểm định giả thiết H: σ2 = với mức ý nghĩa α.

Phương pháp giải

Nếu H đúng thì:

$$\frac{(n-1)S^2}{\sigma_1^2} \sim X\ 2(n-1)$$

$$P(X^2_{1-\frac{\alpha}{2}}(n-1) \leq \frac{(n-1)S^2}{\sigma_o^2} \leq X^2_{\frac{\alpha}{2}}(n-1)) = 1-\alpha$$

Từ đó ta có quy tắc kiểm định:

- Tìm $X^2_{\frac{\alpha}{2}}(n-1)$ và $X^2_{1-\frac{\alpha}{2}}(n-1)$ từ bảng phân phối X 2

- Tính thống kê $X_o^2 = \dfrac{(n-1)s^2}{\sigma_o^2}$.

Nếu $X^2_{1-\frac{\alpha}{2}} \le X_o^2 \le X^2_{\frac{\alpha}{2}}$ thì chấp nhận H

Nếu $X_0^2 < X^2_{1-\frac{\alpha}{2}}$ hoặc $X_0^2 > X^2_{\frac{\alpha}{2}}$ thì bác bỏ H.

Ví vụ 7.7. Nếu máy móc hoạt động bình thường thì chiều dài của một loại sản phẩm là đại lượng ngẫu nhiên có phân phối chuẩn với phương sai 3 cm. Nghi ngờ máy hoạt động không bình thường người ta đo thử một số sản phẩm thì được số liệu:

Chiều dài (cm)	105	107	109	111
Số sản phẩm	2	4	5	2

Giải. Gọi σ2 là phương sai của sản phẩm hiện nay. Từ mẫu ta có n = 13, s2 = 3,74.

Giả thiết H: $\sigma^2 = \sigma_o^2 = 3$ (máy vẫn hoạt động bình thường).

- $X^2_{1-\frac{\alpha}{2}} = X^2_{0,975}(12) = 4,4$

$X^2_{\frac{\alpha}{2}} = X^2_{0,025}(12) = 23,3$

$$X_o^2 = \frac{12.3,74}{3} = 14,96$$

Vì $X^2_{1-\frac{\alpha}{2}} < X_0^2 < X^2_{\frac{\alpha}{2}}$ nên chấp nhận H. Vậy máy móc vẫn hoạt động bình thường.

7.4. KIỂM ĐỊNH QUY LUẬT PHÂN PHỐI

Bài toán. Giả sử đại lượng ngẫu nhiên X chưa rõ phân phối. Cho một mẫu kích thước n. Hãy kiểm định giả thiết:

H : X có phân phối F(x)

với mức ý nghĩa α.

1- Trường hợp X là đại lượng ngẫu nhiên rời rạc

Xét mẫu:

Xét mẫu:

x_i	x_1	x_2	...	x_k
n_i	n_1	n_2	...	n_k

$$\Sigma n_i = n$$

Từ mẫu đã cho ta tính được các ước lượng hợp lý cực đại của các tham số của F(x) nếu các tham số đó chưa biết. Từ các tham số đó ta sẽ tính được các xác suất: $p_i = P(X = x_i), i = \overline{1,k}$

Ta có quy tắc kiểm định như sau:

- Tìm $X_\alpha^2(k - r - 1)$ từ bảng phân phối *X* 2, ở đây r là số tham số của F(x)

- Tính thống kê $X_o^2 = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i}$

Nếu $X_0^2 \leq X_\alpha^2$ *thì chấp nhận H*

Nếu $X_0^2 > X_\alpha^2$ *thì bác bỏ H.*

Ví dụ 7.8. Số lần gọi X đến một trạm điện thoại trong một phút được cho ở bảng sau:

Số lần gọi	0	1	2	3	4	5	6
Số phút	17	22	26	20	11	2	2

Với mức ý nghĩa 1%, có thể coi X có phân phối Poisson được hay không?

Giải. Giả sử X ~ P(a). Khi đó ước lượng hợp lý cực đại của a là $\bar{x} = 2$.

Từ đó:

$$p_i = P(X = x_i) = \frac{e^{-2} 2^{x_i}}{x_i!}$$

Tra bảng ta được: $p_1 = 0{,}1353;\ p_2 = 0{,}2707;\ p_3 = 0{,}2707;\ p_4 = 0{,}1804$

$p_5 = 0{,}0902;\ p_6 = 0{,}0361;\ p_7 = 0{,}0120$

Số tham số của phân phối Poisson là r = 1, do đó:

$$X^2_\alpha = X^2_{0,01}(7-1-1) = X^2_{0,01}(5) = 15{,}1$$

$$X^2_o = \frac{(17-1353)^2}{13{,}53} + \frac{(22-27{,}04)^2}{27{,}04} + \frac{(26-27{,}07)^2}{27{,}07} + \frac{(20-18{,}04)^2}{18{,}04} +$$

$$+ \frac{(11-9{,}02)^2}{9{,}02} + \frac{(2-3{,}61)^2}{3{,}61} + \frac{(2-1{,}20)^2}{1{,}20} = 3{,}78$$

Vì $X_0^2 < X_\alpha^2$ nên chấp nhận giả thiết X có phân phối Poisson.

2- Trường hợp X là đại lượng ngẫu nhiên liên tục

Viết lại mẫu đã cho dưới dạng:

x_i	n_i
$(-\infty, a_1)$ (a_1, a_2) (a_2, a_3) $\vdots$ (a_{k-1}, ∞)	n_1 n_2 n_3 $\vdots$ n_k
	$\Sigma = n$

Từ mẫu ta tìm được các ước lượng hợp lý cực đại của các tham số của F(x) nếu các tham số này chưa biết. Từ đó ta tính được:

p1 = P(X < a1), p2 = P(a1 < X < a2), ..., pk = P(X > ak-1)

Ta có quy tắc kiểm định như sau:

- Tìm $X_\alpha^2 = X_\alpha^2(k-r-1)$ từ bảng phân phối X 2, ở đây là số tham số của F(x)

- Tính thống kê $X_o^2 = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i}$

Nếu $X_0^2 \leq X_\alpha^2$ thì chấp nhận H

Nếu $X_o^2 > X_\alpha^2$ thì bác bỏ H.

Ví dụ 7.9. Điểm trung bình học tập của 100 sinh viên, có số liệu:

x_i	n_i
(0,3)	8
(3,5)	11
(5,7)	50
(7,8)	22
(8,10)	9

Với mức ý nghĩa 5%, kiểm định giả thiết nói rằng điểm trung bình học tập của sinh viên theo luật phân phối chuẩn.

Giải. Từ mẫu ta có , các giá trị này là ước lượng hợp lý cực đại của a và σ2. Nếu X có phân phối chuẩn thì X ~ N(6,0; (1,83)2), từ đó:

$$p_1 = \Phi\left(\frac{3-6}{1,83}\right) + 0,5 = 0,0505$$

$$p_2 = \Phi\left(\frac{5-6}{1,83}\right) - \Phi\left(\frac{3-6}{1,83}\right) = 0,2407$$

$$p_3 = \Phi\left(\frac{7-6}{1,83}\right) - \Phi\left(\frac{5-6}{1,83}\right) = 0,4176$$

$$p_4 = \Phi\left(\frac{8-6}{1,83}\right) - \Phi\left(\frac{7-6}{1,83}\right) = 0,1523$$

$$p_5 = 0,5 - \Phi\left(\frac{8-6}{1,83}\right) = 0,1379$$

Số tham số của phân phối chuẩn là r = 2, nên:

$$X_\alpha^2 = X_{0,05}^2(5-2-1) = 6,0$$

$$X_o^2 = \frac{(8-5,05)^2}{5,05} + \frac{(11-24,07)^2}{24,07} + \frac{(50-41,76)^2}{41,76} +$$

$$+\frac{(22-15,23)^2}{15,23}+\frac{(9-13,79)^2}{13,79}=15,12$$

Vì $X_o^2 > X_\alpha^2$ nên không thể coi điểm trung bình học tập của sinh viên là có phân phối chuẩn với mức ý nghĩa 5%.

7.5. KIỂM ĐỊNH TÍNH ĐỘC LẬP

Bài toán. Cho X và Y là hai dấu hiệu trên cùng một tổng thể. Từ một mẫu kích thước n ta có số liệu:

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

trong đó: xi ($i=\overline{1,k}$) - là các dấu hiệu mà X nhận

y_j $(j = \overline{1,h})$ - là các dấu hiệu mà Y nhận

n_i $(i = \overline{1,k})$ - là số lần X nhận x_i

m_j $(j = \overline{1,h})$ - là số lần Y nhận y_j

$n_{i,j}$ $(i = \overline{1,k},\ j = \overline{1,h})$ - là số lần đồng thời X nhận x_i và Y nhận y_j.

Hãy kiểm định giả thiết:

H: X và Y độc lập

với mức ý nghĩa α.

Phương pháp giải

- Tìm $X_\alpha^2 = X_\alpha^2\left[(k-1)(h-1)\right]$ từ bảng phân vị *X* 2

- Tính thống kê

$$X_o^2 = \sum_{i=1}^{k}\sum_{j=1}^{h}\frac{\left(n_{ij}-\gamma_{ij}\right)^2}{\gamma_{ij}}, \qquad \gamma_{ij} = \frac{n_i m_j}{n}$$

Nếu $X_o^2 \le X_\alpha^2$ thì chấp nhận H

Nếu $X_o^2 > X_\alpha^2$ thì bác bỏ H.

Ví dụ 7.10. Trong một tổng công ty, theo dõi 1.000 công nhân trong một năm thấy ngày nghỉ việc của họ như sau:

Số ngày nghỉ việc trong năm	**Nữ**	**Nam**
0 – 10	300	500
10 – 20	80	70
20 trở lên	20	30

Với mức ý nghĩa 5%, có phải số ngày nghỉ việc phụ thuộc vào giới tính?

Giải. Từ số liệu của bài toán, ta có bảng sau:

X \ Y	y_1	y_2	n_i
x_1	300	500	800
x_2	80	70	150
x_3	20	30	50
m_j	400	600	n = 1000

trong đó: X - là các khoảng ngày nghỉ việc

Y - là các giới tính.

Ta kiểm định giả thiết:

H: X và Y độc lập (số ngày nghỉ việc không phụ thuộc giới tính)

- $X_{\alpha}^{2} = X_{0,05}^{2}\left[(3-1)(2-1)\right] = X_{0,05}^{2}(2) = 6,0$

- Các giá trị của γij:

γ_{ij}	1	2
1	320	480
2	60	90
3	20	30

từ đó:

$$X_{o}^{2} = \frac{(3000-320)^2}{320} + \frac{(500-480)^2}{480} + \frac{(80-60)^2}{60} + \frac{(70-90)^2}{90} + \frac{(20-20)^2}{30} + \frac{(30-30)^2}{30} = 13,19$$

Vì $X_{o}^{2} > X_{\alpha}^{2}$ nên bác bỏ H. Vậy số ngày nghỉ việc của công nhân có phụ thuộc vào giới tính.

BÀI TẬP

7.1. Tỷ lệ phế phẩm cho một máy tự động sản xuất là 5%. Kiểm tra ngẫu nhiên 360 sản phẩm thấy có 24 sản phẩm là phế phẩm. Từ đó có ý kiến tỷ lệ phế phẩm do máy đó sản xuất đã tăng lên. Hãy kiểm tra ý kiến trên với mức ý nghĩa $\alpha = 0{,}05$.

7.2. Nếu áp dụng phương pháp I thì tỷ lệ phế phẩm là 6%; còn nếu áp dụng áp dụng phương pháp II thì trong 100 phế phẩm có năm phế phẩm. Vậy có thể kết luận áp dụng phương pháp thứ II thì tỷ lệ phế phẩm ít hơn phương pháp thứ I không? Hãy kết luận với mức ý nghĩa $\alpha = 0{,}05$.

7.3. Tỷ lệ bệnh nhân bị bệnh T khi điều trị bằng thuốc A là 85%. Thí nghiệm dùng thuốc B điều trị thì trong 900 người mắc bệnh T thì có 800 người khỏi bệnh. Vậy với mức ý nghĩa $\alpha = 0{,}05$ có thể kết luận thuốc B có hiệu quả hơn thuốc A không?

7.4. Người ta khẳng định rằng tỷ lệ thanh niên tốt nghiệp phổ thông trung học ở một thành phố nọ là 0,5. Để kiểm chứng điều này người ta phỏng vấn ngẫu nhiên 625 thanh niên thấy có 300 người đã tốt nghiệp trung học phổ thông. Với mức ý nghĩa $\alpha = 0{,}05$, hãy cho kết luận về khẳng định trên.

7.5. Thống kê số phế phẩm của hai 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 $\alpha = 0{,}05$, hãy xét xem tỷ lệ phế phẩm của hai nhà máy đó có như nhau không?

7.6. Ở một tỉnh nọ, đánh giá xem tỉ lệ mắc bệnh lao ở thành thị và nông thôn có như nhau không, người ta chọn ngẫu nhiên 1.200 người ở thành phố kiểm tra thấy có 40 người mắc bệnh lao, chọn ngẫu nhiên 1.500 người ở nông thôn kiểm tra thấy có 40 người mắc bệnh lao. Với mức ý nghĩa $\alpha = 0{,}01$, hãy kiểm định giả thuyết: tỷ lệ người bị bệnh lao ở thành thị và nông thôn là như nhau.

7.7. Kiểm tra chất lượng hai kho sản phẩm, người ta kiểm tra ngẫu nhiên 500 sản phẩm ở kho I thì thấy có 50 phế phẩm, kiểm tra ngẫu nhiên 400 sản phẩm ở kho II thì thấy có 60 phế phẩm. Với mức ý nghĩa $\alpha = 0{,}05$ ta có thể kết luận gì về hai kho sản phẩm trên?

7.8. 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 lệnh chuẩn $\sigma = 1$ kg, trọng lượng trung bình là 50 kg.

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 bảng kết quả sau:

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

Với mức ý nghĩa $\alpha = 0{,}05$, hãy kết luận về nghi ngờ nói trên.

7.9. Mì chính được đóng gói theo quy định trên một máy tự động là 453 g/gói. Có thể coi trọng lượng gói mì chính tuân theo quy luật chuẩn với độ lệch chuẩn $\sigma = 16$ gam. Kiểm tra ngẫu nhiên 81 gói thấy trọng lượng trung bình là 448 gam. Với mức ý nghĩa $\alpha = 0{,}05$ có thể kết luận trong lượng trung bình gói mì chính có xu hướng giảm hay không?

7.10. Mức hao phí xăng (X) cho một loại xe ôtô chạy trên đoạn đường AB là một đại lượng ngẫu nhiên có phân phối chuẩn có kỳ vọng là 50 lít. Nay do đường mới được tu sửa lại, người ta cho rằng mức hao phí xăng 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 - 5,10
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.

7.11. Định mức cũ để sản xuất một sản phẩm là 20,5 phút. Nay do cải tiến kỹ thuật, người ta sản xuất thử 100 sản phẩm thu được bảng số liệu sau:

Thời gian để sản xuất một sản phẩm (phút).	16 - 17	17 - 18	18 - 19	19 - 20	20 - 21	21 - 22
Số sản phẩm tương ứng (n_i)	6	10	24	30	18	12

Thời gian để sản xuất một sản phẩm (phút).16 - 1717 - 1818 - 1919 – 2020 - 2121 - 22Số sản phẩm tương ứng (ni)61024301812 Với mức ý nghĩa $\alpha = 0,05$, có thể nói việc cải tiến kỹ thuật đã làm tăng năng suất hay không?

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

Thửa thứ nhất lấy mẫu 1000 bông lúa thấy số hạt trung bình là $\overline{x}_1 = 70$ hạt và độ lệch chuẩn hiệu chỉnh S1 = 10

Thửa thứ hai lấy mẫu 500 bông lúa thấy số hạt trung bình là $\overline{x}_2 = 72$ hạt và độ lệch chuẩn hiệu chỉnh S2 = 20. Hỏi với mức ý nghĩa $\alpha = 0{,}05$ thì sự khác nhau của $\overline{x}_1$ và $\overline{x}_2$ là ngẫu nhiên hay bản chất?

7.13. Định mức để hoàn thành một sản phẩm là 14,5 phút. Có cần thay đổi định mức không, nếu theo dõi thời gian hoàn thành sản phẩm 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 $\alpha = 0{,}05$, biết rằng thời gian hoàn thành một sản phẩm (X) là một đại lượng ngẫu nhiên có phân phối chuẩn.

7.14. Người ta thí nghiệm hai phương pháp chăn nuôi gà khác nhau, sau một 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	1,2	0,3

Với mức ý nghĩa $\alpha = 0,05$, có thể kết luận phương pháp II hiệu quả hơn phương pháp I không?

7.15. Điều tra về trọng lượng trung bình của trẻ sơ sinh ở hai khu vực thành thị và nông thôn có kết quả:

Khu vực	**Số trẻ được cân**	**Trọng lượng trung bình**	**Phương sai hiệu chỉnh**
Nông thôn	$n_1=2500$	$\overline{x}_1 = 3,0$ kg	$S_1^2 = 0,04$
Thành phố	$n_2=500$	$\overline{x}_2 = 3,1$ kg	$S_2^2 = 0,09$

Với mức ý nghĩa $\alpha = 0,05$, có thể coi trọng lượng trung bình của trẻ sơ sinh ở hai khu vực thành thị và nông thôn bằng nhau được không?

7.16. Chọn ngẫu nhiên 47 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 $\alpha = 0,05$, hãy cho kết luận?

7.17. Xét đại lượng ngẫu nhiên X có phân phối chuẩn N(a,σ2). Xét mẫu:

x_i	2,7	2,8	2,9	3,0	3,1	3,2	3,3
n_i	1	2	5	6	5	4	2

Với mức ý nghĩa α = 0,05, hãy kiểm định giả thiết $H_o : \sigma^2 = 0{,}03$ với đối giả thiết $H_1 : \sigma^2 \neq 0{,}03$.

7.18. Trọng lượng gà con lúc mới nở là một đại lượng ngẫu nhiên có phân phối chuẩn. Nghi ngờ độ đồng đều về trọng lượng gà con giảm sút, người ta cân thử 12 con và tìm được phương sai hiệu chỉnh là 11,41 (gam)2. Với mức ý nghĩa α = 0,05, hãy kết luận về điều nghi ngờ trên. Biết rằng bình thường độ phân tán của trọng lượng gà con là σ2 = 10 (gam)2.

7.19. Nếu máy móc hoạt động bình thường thì trọng lượng sản phẩm tuân theo quy luật chuẩn với độ lệch chuẩn 1 kg. Có thể coi máy móc hoạt động tốt hay không nếu cân thử 30 sản phẩm ta thấy độ lệch chuẩn hiệu chỉnh mẫu là 1,1 kg. Yêu cầu kết luận với mức ý nghĩa α = 0,01.

7.20. Số con của 2000 phụ nữ ở một vùng dân cư ở độ tuổi 26 cho ở bảng sau:

x_i (số con)	0	1	2	3	4
n_i (số phụ nữ)	1090	650	220	30	10

Với mức ý nghĩa α = 0,05, có thể xét xem mẫu trên phù hợp với phân phối Poisson được không?

7.21. 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 α = 0,01, có thể coi mẫu trên phù hợp với phân phối Poisson hay không?

7.22. Kết quả đo kích thước của 1000 chi tiết cho trong bảng sau

Kích thước mm	97,75-98,25	98,25-98,75	98,75-99,25	99,25-99,75	99,75 -100,25	100,25-100,75	100,75-101,25	101,25 - 101,75	101,75-102,25	102,25-102,75
Số chi tiết tương ứng	21	47	87	158	181	201	142	97	41	25

Với mức ý nghĩa α = 0,05, có thể coi kích thước chi tiết sản xuất ra tuân theo quy luật chuẩn hay không?

7.23. Từ bộ sản phẩm của một máy tiện người ta chọn ra 200 chiếc. Bán kính sản phẩm được đo đạc và cho như sau:

Bán kính x_i	3,2	3,4	3,6	3,8	4,0	4,2	4,4	4,6	4,8	5,0
Số lượng n_i	1	5	4	18	86	62	14	6	3	1

Với mức ý nghĩa $\alpha = 0{,}05$ có thể coi bán kính sản phẩm của máy tiện đó tuân theo quy luật chuẩn?

7.24. Cho mẫu sau:

x_i	5 - 15	15 - 25	25 - 35	35 - 45	45 - 55	55 - 65	65 - 75
n_i	45	197	308	202	198	22	18

Với mức ý nghĩa $\alpha = 0{,}05$ có thể coi mẫu trên phù hợp với phân phối chuẩn hay không?

7.25. Thống kê về chiều cao của một loại cây sau hai tháng tuổi cho kết quả sau:

Độ cao (cm)	5	7	9	11	13	15	17	19	21
Số lượng	11	26	27	32	25	22	24	20	13

Với mức ý nghĩa $\alpha = 0,01$ hãy kiểm định xem mẫu trên có phù hợp với phân phối chuẩn không?

7.26. Ở một trường đại học, để nghiên cứu xem khả năng toán học của sinh viên có tương quan gì với sự yêu thích môn thống kê hay không, người ta chọn ngẫu nhiên 200 sinh viên điều tra có kết quả:

Thái độ đối với môn thống kê	**Khả năng toán học**		
	Thấp	Trung bình	Cao
Ít thích	60	15	15
Thích vừa	15	45	10
Rất thích	5	10	25

Với mức ý nghĩa $\alpha = 0,05$, kiểm định xem sự yêu thích môn thống kê có phụ thuộc vào khả năng toán học của sinh viên trường này hay không?

7.27. Trong một thí nghiệm khoa học người ta nghiên cứu độ dày của lớp mạ kền khi dùng ba loại bể mạ khác nhau. Sau một thời gian mạ, người ta đo độ dày của lớp mạ nhận được ở các bể:

Độ dày lớp mạ kền tính bằng μm	**Số lần đo ở bể mạ**		
	A	B	C
4 - 8	32	51	68
8 - 12	123	108	80
12 - 16	10	26	26
16 - 20	41	24	28
20 - 24	19	20	28

với mức ý nghĩa $\alpha = 0{,}05$, hãy kiểm định giả thiết: độ dày lớp mạ sau khoảng thời gian nói trên không phụ thuộc loại bể mạ được dùng.

7.28. Người ta nghiên cứu xem có sự phụ thuộc giữa huyết áp của cha và huyết áp của con hay không bằng cách chọn ngẫu nhiên 97 học sinh lớp 11 ở một trường phổ thông, đo và có kết quả:

Huyết áp cha	**Huyết áp con**		
	Thấp	Vừa phải	Cao
Thấp	14	11	8
Vừa phải	11	11	9
Cao	6	10	12

Với mức ý nghĩa $\alpha = 0,05$. Hãy xét xem huyết áp của con có phụ thuộc vào huyết áp của cha hay không?

7.29. Nghiên cứu sự ảnh hưởng của thành phần thức ăn của bố mẹ (X) đối với giới tính (Y) của con có kết quả:

Giới tính Y của con	**Thành phần thức ăn X**	
	Không có vitamin	Có vitamin
Nam	123	145
Nữ	153	150

Với mức ý nghĩa $\alpha = 0,01$, có thể xem X và Y độc lập với nhau được không?

7.30. Nghiên cứu ảnh hưởng của hoàn cảnh gia đình đối với tình trạng phạm tội của trẻ em vị thành niên có kết quả:

Tình trạng phạm tội	**Hoàn cảnh gia đình**		
	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 $\alpha = 0,05$ có thể kết luận hoàn cảnh gia đình độc lập với tình trạng phạm tội của trẻ em hay không?