Chương 5:

KIỂM ĐỊNH GIẢ THIẾT

I. Bài toán kiểm định

Cho một Tổng thể có tham số θ chưa biết. Người ta đưa ra một giá thiết *H* cho θ là:

$$H:\theta=\theta_0$$
$$H:\theta\le\theta_0$$
$$H:\theta\ge\theta_0$$

Khi đó, ta sẽ đưa ra một **giải thiết đối** để kiểm định *H* với độ tin cậy $1-\alpha \Leftrightarrow$ mức ý nghĩa α.

$$\overline{H}:\theta\ne\theta_0 \qquad \text{(Dạng 1)}$$
$$\overline{H}:\theta>\theta_0 \qquad \text{(Dạng 2)}$$
$$\overline{H}:\theta<\theta_0 \qquad \text{(Dạng 3)}$$

* **Có 2 loại sai lầm**:

- Sai lầm loại 1: Bác bỏ giả thiết *H* nhưng thực tế giả thiết *H* đúng (*bác bỏ giả thuyết đúng*).

- Sai lầm loại 2: Chấp nhận giả thiết *H* nhưng thực tế giả thiết *H* sai (*Chấp nhận giả thiết sai*).

II. KIểm định tỷ lệ

Giả sử một tổng thể có hai loại phần tử bao gồm loại phần tử có tính chất *A* và loại phần tử không có tính chất *A*. Khi đó:

Gọi *p* là tỷ lệ loại phần tử có tính chất *A* của tổng thể đã cho.

Giả sử *p* chưa biết. Ta cần kiểm định *p* với mức ý nghĩa α.

Bước 1: Xác định giả thiết và giả thiết đối

$$\begin{cases} H: p=p_0 \\ \overline{H}: p\ne p_0 \end{cases} \qquad \begin{cases} H: p\le p_0 \\ \overline{H}: p> p_0 \end{cases} \qquad \begin{cases} H: p\ge p_0 \\ \overline{H}: p< p_0 \end{cases}$$

(Dạng 1) (Dạng 2) (Dạng 3)

Bước 2: Với α cho trước, tìm miền giá trị bác bỏ giả thiết *H* là W_α

- Nếu ở **Bước 1** chọn dạng 1 thì

$$W_\alpha=\left(-\infty;-Z_{1-\frac{\alpha}{2}}\right)\cup\left(Z_{1-\frac{\alpha}{2}};+\infty\right)$$

- Nếu ở **Bước 1** chọn dạng 2 thì

$$W_\alpha=\left(Z_{1-\alpha};+\infty\right)$$

- Nếu ở **Bước 1** chọn dạng 3 thì

$$W_\alpha=\left(-\infty;-Z_{1-\alpha}\right)$$

Bước 3: Chọn 1 mẫu có kích thước *n*, tính *f*

(So sánh f và p_0)

- Tính giá trị quan sát:

$$u_{qs} = \frac{(f - p_0)}{\sqrt{p_0(1-p_0)}}\sqrt{n}$$

Bước 4: So sánh u_{qs} và W_α

- $u_{qs} \in W_\alpha \Rightarrow$ Bác bỏ H và chấp nhận $\overline{H}$.

- $u_{qs} \notin W_\alpha \Rightarrow$ Chấp nhận H và bác bỏ $\overline{H}$.

Ví dụ 1: Một bệnh gây tử vong 20% bệnh nhân. Người ta muốn dùng 1 loại thuốc mới điều trị bệnh này. Khi điều trị thử nghiệm trên 200 người bị bệnh, vẫn còn 30 người tử vong. Dựa vào số liệu đã cho có đủ cơ sở để kết luận loại thuốc điều trị mới có hiệu quả hay chưa, $\alpha = 5\%$.

<u>Giải</u>

*** Phân tích**:

- Ban đầu tỷ lệ tử vong $p_0 = 20\%$

- Gọi *p* là tỷ lệ người bị tử vong sau khi sử dụng thuốc mới điều trị.

+ *p*: chưa biết

- Ta cần kiểm định *p* với mức ý nghĩa $\alpha = 5\%$.

- Xác định giả thiết và giả thiết đối.

+ Sử dụng thuốc mới - Có hiệu quả: $p < 20\%$

- Chưa có hiệu quả: $p \geq 20\%$

$$\Rightarrow \begin{cases} H : p \geq 20\% \\ \overline{H} : p < 20\% \end{cases} \text{(Dạng 3)}$$

* Với $\alpha = 5\%$, thì $W_\alpha = (-\infty; -Z_{1-\alpha}) = (-\infty; -Z_{0.95}) = (-\infty; -1.645)$

* Phân tích mẫu: có *n=200* người + 30 tử vong

+ 170 còn lại

- Ta có: $f = \frac{30}{200} = 0.15$

- Giá trị quan sát: $u_{qs} = \frac{(f - p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}} = \frac{(0.15-0.2)\sqrt{200}}{\sqrt{0.2(1-0.2)}} = -1.768$

* So sánh u_{qs} và W_α

- Ta có: $u_{qs} \in W_\alpha \Rightarrow$ Bác bỏ H và chấp nhận $\overline{H}$.

Vậy có đủ cơ sở kết luận loại thuốc mới điều trị bệnh có hiệu quả.

Chú ý: u_{qs} cùng dấu với W_α

Ví dụ 2: Một trường Đại học cho biết tỷ lệ sinh viên mới tốt nghiệp có việc làm của Trường mình là 80%. Để kiểm tra người ta chọn ra ngẫu nhiên 300 sinh viên mới tốt nghiệp Trường đó, thì có 210 người có việc làm. Với mức ý nghĩa 5% có thể cho rằng báo cáo của Trường đó là quá cao hay không?

Giải

*** Phân tích**:

- Ban đầu tỷ lệ có việc làm: $p_0 = 80\%$.

- Gọi *p* là tỷ lệ sinh viên mới tốt nghiệp có việc làm khi điều tra thực tế

+ *p*: chưa biết

- Ta cần kiểm định *p* với mức ý nghĩa $\alpha = 5\%$

- Xác định giả thiết và gia thiết đối

+ Báo cáo + Quá cao: nếu tỷ lệ SV mới tốt nghiệp có việc làm thực tế $< 80\%$

+ Không quá cao: nếu tỷ lệ SV mới TN có việc làm thực tế $\geq 80\%$

$\Rightarrow \begin{cases} H: p \geq 80\% \\ \overline{H}: p < 80\% \end{cases}$ (Dạng 3)

III. Kiểm định trung bình

Cho m là trung bình dấu hiệu của Tổng thể.

Giả sử m chưa biết. Ta cần kiểm định m với mức ý nghĩa α.

Bước 1: Xác định giả thiết và giả thiết đối

$$\begin{cases} H: m = m_0 \\ \overline{H}: m \neq m_0 \end{cases} \qquad \begin{cases} H: m \leq m_0 \\ \overline{H}: m > m_0 \end{cases} \qquad \begin{cases} H: m \geq m_0 \\ \overline{H}: m < m_0 \end{cases}$$

Trường hợp 1: Biết phương sai $V(X) = \sigma^2$ và $n \geq 30$ hoặc $n < 30$

Bước 2: Với α cho trước, tìm miền giá trị bác bỏ giả thiết *H* là W_α

- Nếu $\overline{H}: m \neq m_0$ thì $W_\alpha = \left(-\infty; -Z_{1-\frac{\alpha}{2}}\right) \cup \left(Z_{1-\frac{\alpha}{2}}; +\infty\right)$

- Nếu $\overline{H}: m > m_0$ thì $W_\alpha = \left(Z_{1-\alpha}; +\infty\right)$

- Nếu $\overline{H}: m < m_0$ thì $W_\alpha = \left(-\infty; -Z_{1-\alpha}\right)$

Bước 3: Tính giá trị quan sát:

$$u_{qs} = \frac{\left(\overline{x} - m_0\right)}{\sigma}\sqrt{n}$$

Bước 4: So sánh u_{qs} và W_α

- $u_{qs} \in W_\alpha \Rightarrow$ Bác bỏ H và chấp nhận $\overline{H}$.

- $u_{qs} \notin W_\alpha \Rightarrow$ Chấp nhận H và bác bỏ $\overline{H}$.

Trường hợp 2: Chưa biết phương sai $V(X)=\sigma^2$ và $n\geq 30$

Bước 2: Với α cho trước, tìm miền giá trị bác bỏ giả thiết H là W_α

- Nếu $\overline{H}: m\neq m_0$ thì $W_\alpha=\left(-\infty;-Z_{1-\frac{\alpha}{2}}\right)\cup\left(Z_{1-\frac{\alpha}{2}};+\infty\right)$

- Nếu $\overline{H}: m> m_0$ thì $W_\alpha=\left(Z_{1-\alpha};+\infty\right)$

- Nếu $\overline{H}: m< m_0$ thì $W_\alpha=\left(-\infty;-Z_{1-\alpha}\right)$

Bước 3: Tính giá trị quan sát:

$$u_{qs}=\frac{\left(\overline{x}-m_0\right)}{s}\sqrt{n}$$

Bước 4: So sánh u_{qs} và W_α

- $u_{qs}\in W_\alpha \Rightarrow$ Bác bỏ H và chấp nhận $\overline{H}$.

- $u_{qs}\notin W_\alpha \Rightarrow$ Chấp nhận H và bác bỏ $\overline{H}$.

Trường hợp 3: Chưa biết phương sai $V(X)=\sigma^2$ và $n<30$

Bước 2: Với α cho trước, tìm miền giá trị bác bỏ giả thiết H là W_α

- Nếu $\overline{H}: m\neq m_0$ thì $W_\alpha=\left(-\infty;-t_{\frac{\alpha}{2}}(n-1)\right)\cup\left(t_{\frac{\alpha}{2}}(n-1);+\infty\right)$

- Nếu $\overline{H}: m> m_0$ thì $W_\alpha=\left(t_\alpha(n-1);+\infty\right)$

- Nếu $\overline{H}: m< m_0$ thì $W_\alpha=\left(-\infty;-t_\alpha(n-1)\right)$

Bước 3: Tính giá trị quan sát:

$$u_{qs}=\frac{\left(\overline{x}-m_0\right)}{s}\sqrt{n}$$

Bước 4: So sánh u_{qs} và W_α

- $u_{qs}\in W_\alpha \Rightarrow$ Bác bỏ H và chấp nhận $\overline{H}$.

- $u_{qs}\notin W_\alpha \Rightarrow$ Chấp nhận H và bác bỏ $\overline{H}$.

Ví dụ 3: Biết chiều cao của nam thanh niên Việt Nam là đại lượng ngẫu nhiên *X (cm)* có phân phối chuẩn *N(m ; 100)*. Có tài liệu cho rằng chiều cao trung bình của nam thanh niên Việt Nam là 164cm. Để kiểm tra, người ta chọn ngẫu nhiên một mẫu gồm 80 nam thanh niên để đo chiều cao và tính được chiều cao trung bình là 162cm. Với mức ý nghĩa 5%, giá trị cho biết có phù hợp với mẫu quan sát hay không?

Giải

*** Phân tích**:

- Ban đầu chiều cao nam thanh niên Việt Nam: $m_0 = 164$.

- Gọi *m* là chiều cao trung bình của nam thanh niên Việt Nam

+ m chưa biết

- Xác định giá thiết và giả thiết đối

+ Chiều cao nam thanh niên Việt Nam: + Theo tài liệu: $m=164cm$

+ Thực tế: $m \neq 164cm$

$$\Rightarrow \begin{cases} H: m = 164 \\ \overline{H}: m \neq 164 \end{cases}$$

- Với $\alpha = 5\%$, tìm W_α

$$W_\alpha = \left(-\infty; -Z_{1-\frac{\alpha}{2}}\right) \cup \left(Z_{1-\frac{\alpha}{2}}; +\infty\right) = \left(-\infty; -Z_{0.975}\right) \cup \left(Z_{0.975}; +\infty\right) = \left(-\infty; -1.96\right) \cup \left(1.96; +\infty\right)$$

- Giá trị quan sát:

$$u_{qs} = \frac{\left(\overline{x} - m_0\right)\sqrt{n}}{\sigma} = \frac{(162-164)\sqrt{80}}{10} = -1.789$$

- So sánh u_{qs} và W_α

+ Ta có: $u_{qs} \notin W_\alpha \Rightarrow$ Chấp nhận H và bác bỏ $\overline{H}$.

Vậy giá trị cho biết có phù hợp với mẫu quan sát.

Ví dụ 4: Trọng lượng của các bao gạo là đại lượng ngẫu nhiên có phân phối chuẩn với trọng lượng trung bình là 50kg. Sau một thời gian hoạt động người ta nghi ngờ trọng lượng các bao gạo bị thiếu. Cân thử 25 bao gạo thu được kết quả sau:

Trọng lượng (kg)	48 – 48.5	48.5 – 49	49 – 49.5	49.5 – 50	50 – 50.5
Số bao gạo	1	4	9	8	3

Với độ tin cậy 99% hãy kết luận về điều nghi ngờ nói trên

Giải

*** Phân tích**:

- Trọng lượng trung bình là $m_0 = 50kg$

- Gọi *m* là trọng lượng trung bình của các bao gạo

+ *m* chưa biết?

- Xác định giả thiết và giả thiết đối:

+ Trọng lượng của các bao gạo: + Ban đầu: $m \geq 50$

+ Nghi ngờ: $m < 50$

$$\Rightarrow \begin{cases} H: m \geq 50 \\ \overline{H}: m < 50 \end{cases}$$

- Với $1-\alpha=99\%$, tìm W_α

$$W_\alpha=\left(-\infty;-t_\alpha\left(n-1\right)\right)=\left(-\infty;-t_{0.01}\left(24\right)\right)=\left(-\infty;-2.492\right)$$

- Giá trị quan sát:

+ Từ bảng dữ liệu ta có: $n=25,\overline{x}=49.41,s=0.515$

$$\Rightarrow u_{qs}=\frac{\left(\overline{x}-m_0\right)\sqrt{n}}{s}=\frac{\left(49.41-50\right)\sqrt{25}}{0.515}=-5.728$$

- So sánh u_{qs} và W_α

+ Ta có: $u_{qs}\in W_\alpha\Rightarrow$ Bác bỏ H và chấp nhận $\overline{H}$.

Vậy việc nghi ngờ của khách hàng đúng.

IV. So sánh hai tỷ lệ

Giả sử hai tổng thể 1 và 2 cùng chứa hai loại phần tử bao gồm loại phần tử có tính chất loại *A* và loại phần tử không có tính chất *A*.

Gọi P_1,P_2 là tỷ lệ của loại phần tử có tính chất *A* của tổng thể 1 và 2.

Giả sử P_1,P_2 chưa biết. Ta cần so sánh P_1 và P_2 với mức ý nghĩa α.

Bước 1: Xác định giả thiết và giá thiết đối

$$\begin{cases}H:p_1=p_2\\ \overline{H}:p_1\neq p_2\end{cases}\qquad \begin{cases}H:p_1\le p_2\\ \overline{H}:p_1> p_2\end{cases}\qquad \begin{cases}H:p_1\ge p_2\\ \overline{H}:p_1< p_2\end{cases}$$

(Dạng 1) (Dạng 2) (Dạng 3)

Bước 2: Với α cho trước, tìm W_α

- Nếu ở **Bước 1** chọn dạng 1 thì

$$W_\alpha=\left(-\infty;-Z_{1-\frac{\alpha}{2}}\right)\cup\left(Z_{1-\frac{\alpha}{2}};+\infty\right)$$

- Nếu ở **Bước 2** chọn dạng 2 thì

$$W_\alpha=\left(Z_{1-\alpha};+\infty\right)$$

- Nếu ở **Bước 3** chọn dạng 3 thì

$$W_\alpha=\left(-\infty;-Z_{1-\alpha}\right)$$

Bước 3: Chọn 2 mẫu có kích thước n_1 và n_2, tính f_1 và f_2

- Mẫu 1: f_1

- Mẫu 2: f_2

$$\Rightarrow f=\frac{n_1f_1+n_2f_2}{n_1+n_2}$$

- Tính giá trị quan sát:

$$u_{qs}=\frac{f_1-f_2}{\sqrt{f\left(1-f\right)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

Bước 4: So sánh u_{qs} và W_α

- $u_{qs}\in W_\alpha \Rightarrow$ Bác bỏ H và chấp nhận $\overline{H}$.

- $u_{qs}\notin W_\alpha \Rightarrow$ Chấp nhận H và bác bỏ $\overline{H}$.

Ví dụ 5: Người ta cần so sánh sự đánh giá người xem đối với 2 bộ phim đang chiếu trên thị trường. Khi khảo sát 500 người đã từng xem bộ phim A thì có 420 thích bộ phim này. Trong khi đó, khảo sát 400 người đã từng xem bộ phim B thì có 320 người trả lời thích bộ phim này. Với mức nghĩa 5%, hãy kiểm định bộ phim A được yêu thích nhiều hơn bộ phim B hay không?

<u>Giải</u>

Gọi p_1, p_2 là tỷ lệ người thích xem bộ phim A và B

+ p_1, p_2 chưa biết.

Ta cần so sánh p_1 và p_2 với mức ý nghĩa $\alpha=5\%$.

* Xác định giả thiết và giả thiết đối

+ Bộ phim A được yêu thích nhiều hơn bộ phim B: $p_1>p_2$

+ Ngược lại: $p_1\le p_2$

$$\Rightarrow\begin{cases}H: p_1\le p_2\\ \overline{H}: p_1>p_2\end{cases}\text{ (Dạng 2)}$$

* $\alpha=5\%$, tìm W_α

Dạng 2 $\Rightarrow W_\alpha=\left(Z_{1-\alpha};+\infty\right)=\left(1.645;+\infty\right)$

* Chọn 2 mẫu có kích thước n_1 và n_2, tính f_1 và f_2

$$f_1=\frac{420}{500}=0.84\quad f_2=\frac{320}{400}=0.8\quad\Rightarrow f=\frac{740}{900}=0.822$$

- Giá trị quan sát:

$$u_{qs}=\frac{f_1-f_2}{\sqrt{f\left(1-f\right)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}=\frac{0.84-0.8}{\sqrt{0.822*0.178\left(\frac{1}{500}+\frac{1}{400}\right)}}=1.559$$

* So sánh u_{qs} và W_α.

$u_{qs}\notin W_\alpha$ chấp nhận H và bác bỏ $\overline{H}$

Vậy không thể cho rằng bộ phim A được thích nhiều hơn bộ phim B.

V. So sánh hai trung bình

Cho 2 tổng thể. Gọi m_1, m_2 là trung bình dấu hiệu của tổng thể 1 và 2.

Giả sử m_1, m_2 chưa biết.

Ta cần so sánh m_1 và m_2 với mức ý nghĩa α.

Bước 1: Xác định giả thiết và giả thiết đối

$$\begin{cases} H: m_1 = m_2 \\ \overline{H}: m_1 \neq m_2 \end{cases} \qquad \begin{cases} H: m_1 \leq m_2 \\ \overline{H}: m_1 > m_2 \end{cases} \qquad \begin{cases} H: m_1 \geq m_2 \\ \overline{H}: m_1 < m_2 \end{cases}$$

(Dạng 1) (Dạng 2) (Dạng 3)

*** Trường hợp 1**: Khi phương sai chưa biết, kích thước mẫu $n_1, n_2 \geq 30$ (Giống với trường hợp khi phương sai đã biết $V(X_1) = \sigma_1^2, V(X_1) = \sigma_2^2$))

Bước 2: Với α cho trước, tìm W_α

- Nếu ở **Bước 1** chọn dạng 1 thì

$$W_\alpha = \left(-\infty; -Z_{1-\frac{\alpha}{2}}\right) \cup \left(Z_{1-\frac{\alpha}{2}}; +\infty\right)$$

- Nếu ở **Bước 1** chọn dạng 2 thì

$$W_\alpha = \left(Z_{1-\alpha}; +\infty\right)$$

- Nếu ở **Bước 1** chọn dạng 3 thì

$$W_\alpha = \left(-\infty; -Z_{1-\alpha}\right)$$

Bước 3: Chọn 2 mẫu có kích thước n_1 và n_2 khi đó:

- Mẫu 1: tính $n_1, \overline{x_1}, s_1$

- Mẫu 2: tính $n_2, \overline{x_2}, s_2$

- Tính giá trị quan sát:

$$u_{qs} = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Bước 4: So sánh u_{qs} và W_α

- $u_{qs} \in W_\alpha \Rightarrow$ Bác bỏ H và chấp nhận $\overline{H}$.

- $u_{qs} \notin W_\alpha \Rightarrow$ Chấp nhận H và bác bỏ $\overline{H}$.

*** Trường hợp 2**: Khi phương sai chưa biết, kích thước mẫu $n_1, n_2 < 30$

Bước 2: Với α cho trước, tìm W_α

- Nếu ở **Bước 1** chọn dạng 1 thì

$$W_\alpha = \left(-\infty; -t_{\frac{\alpha}{2}}(n_1 + n_2 - 2)\right) \cup \left(t_{\frac{\alpha}{2}}(n_1 + n_2 - 2); +\infty\right)$$

- Nếu ở **Bước 1** chọn dạng 2 thì

$$W_\alpha = \left(t_\alpha\left(n_1 + n_2 - 2\right); +\infty\right)$$

- Nếu ở **Bước 1** chọn dạng 3 thì

$$W_\alpha = \left(-\infty; -t_\alpha\left(n_1 + n_2 - 2\right)\right)$$

Bước 3: Chọn 2 mẫu có kích thước n_1 và n_2 khi đó:

- Mẫu 1: tính $n_1, \overline{x_1}, s_1$

- Mẫu 2: tính $n_2, \overline{x_2}, s_2$

- Tính giá trị quan sát: $$u_{qs} = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}}$$

Trong đó: $s = \dfrac{\left(n_1 - 1\right)s_1^2 + \left(n_2 - 1\right)s_2^2}{n_1 + n_2 - 2}$

Bước 4: So sánh u_{qs} và W_α

- $u_{qs} \in W_\alpha \Rightarrow$ Bác bỏ H và chấp nhận $\overline{H}$.

- $u_{qs} \notin W_\alpha \Rightarrow$ Chấp nhận H và bác bỏ $\overline{H}$.

Ví dụ 6: Khảo sát trọng lượng của một số trẻ sơ sinh ta có bảng số liệu sau đây.

Trọng lượng (Kg)	<2.8	2.8 – 3	3 – 3.2	3.2 – 3.4	>3.4
Số bé Trai	8	17	30	35	22
Số bé Gái	12	15	32	33	18

Với mức ý nghĩa 5%, hãy kiểm định định trong lượng trung bình của bé Trai và bé Gái sơ sinh có bằng nhau hay không?

Giải

Gọi m_1, m_2 lần lượt là trọng lượng trung bình của bé Trai và bé Gái sơ sinh.

+ Giả sử m_1, m_2 chưa biết.

Ta cần so sánh m_1 và m_2 với mức ý nghĩa $\alpha = 5\%$.

Xác định giả thiết và giả thiết đối

$$\begin{cases} H: m_1 = m_2 \\ \overline{H}: m_1 \neq m_2 \end{cases}$$

Với $\alpha = 5\%$ tìm W_α

$$W_\alpha = \left(-\infty; -Z_{1-\frac{\alpha}{2}}\right) \cup \left(Z_{1-\frac{\alpha}{2}}; +\infty\right) = \left(-\infty; -Z_{0.975}\right) \cup \left(Z_{0.975}; +\infty\right) = \left(-\infty; -1.96\right) \cup \left(1.96; +\infty\right)$$

Từ mẫu số liệu, ta có:

- Mẫu 1: $\begin{cases} n_1 = 112 \\ \overline{x_1} = 3.182 \\ s_1 = 0.235 \end{cases}$ Mẫu 2: $\begin{cases} n_2 = 110 \\ \overline{x_2} = 3.155 \\ s_2 = 0.242 \end{cases}$

- Giá trị quan sát:

$$u_{qs} = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{3.182 - 3.155}{\sqrt{\frac{0.235_1^2}{112} + \frac{0.242_1^2}{110}}} = 0.84$$

So sánh u_{qs} và W_α

- $u_{qs} \notin W_\alpha \Rightarrow$ Chấp nhận H và bác bỏ $\overline{H}$.

Vậy trọng lượng trung bình của bé Trai và bé Gái sơ sinh bằng nhau.

VI. Kiểm định nhiều hơn hai tỷ lệ

Giả sử một tổng thể có *n* loại phần tử bao gồm loại phần tử có tính chất loại I, loại phần tử có tính chất loại II, ... loại k (*k*>*2*).

- Tổng thể + Loại I
  + Loại II
  + Loại k (*k*>*2*)

Gọi $p_1, p_2, ..., p_k$ là tỷ lệ loại $1, 2, 3, ..., k$ của tổng thể.

Giả sử $p_1, p_2, ..., p_k$ chua biết.

Ta cần kiểm định $p_1, p_2, ..., p_k$ với mức ý nghĩa α.

Bước 1:

$$\begin{cases} H: p_1 = p_{01}; p_2 = p_{02}; ...; p_k = p_{0k} \\ \overline{H}: \quad \text{Có ít nhất 1 tỷ lệ sai} \end{cases}$$

Bước 2: Với α cho trước, tìm W_α

$$W_\alpha = \left(\chi_\alpha^2 (k-1); +\infty\right)$$

Bước 3: Chọn 1 mẫu có kích thước *n*

Loại phần tử	1	2	…	k
Tần số (mẫu)	n_1	n_2		n_k
Tần số (gt)	$\hat{n}_1 = n * p_{01}$	$\hat{n}_2 = n * p_{02}$		$\hat{n}_k = n * p_{0k}$

- Giá trị quan sát:

$$u_{qs} = \frac{(n_1 - \hat{n}_1)^2}{\hat{n}_1} + \frac{(n_2 - \hat{n}_2)^2}{\hat{n}_2} + ... + \frac{(n_k - \hat{n}_k)^2}{\hat{n}_2}$$

Bước 4: So sánh u_{qs} và W_α

Ví dụ 7: Trong một buổi hội thảo khoa học, có báo cáo cho rằng tỷ lệ người mang nhóm máu O, A, B, AB lần lượt là: 49%, 18%, 28%, 5%.

Với mức ý nghĩa 5%, hãy kiểm định xem báo cáo đã cho đúng hay sai.

Tiến hành khảo sát 500 người ta có bảng số liệu sau đây.

Nhóm	O	A	B	AB
Số người	260	100	130	10

<u>Giải</u>

Gọi p_1, p_2, p_3, p_4 lần lượt tỷ lệ nhóm máu O, A, B và AB.

- p_1, p_2, p_3, p_4 chưa biết. Ta cần so kiểm định các tỷ lệ này với mức ý nghĩa $\alpha = 5\%$.

- Xác định giả thiết và giả thiết đối

$$\begin{cases} H : p_1 = 0.49, p_2 = 0.18, p_3 = 0.28, p_4 = 0.05 \\ \bar{H} : \text{Có ít nhất 1 tỷ lệ bị sai} \end{cases}$$

Với α =5% và *k=4*, tìm W_α

$$W_\alpha = \left(\chi^2_\alpha(k-1); +\infty\right) = \left(\chi^2_{0.05}(3); +\infty\right) = (7.815; +\infty)$$

Nhóm	O	A	B	AB
Số người (mẫu)	260	100	130	10
Số người (Gt)	$n_1 = 500 * 0.49 = 245$	$n_2 = 500 * 0.18 = 90$	$n_3 = 500 * 0.28 = 140$	$n_4 = 500 * 0.05 = 25$

- Giá trị quan sát:

$$u_{qs} = \frac{(n_1 - \hat{n}_1)^2}{\hat{n}_1} + \frac{(n_2 - \hat{n}_2)^2}{\hat{n}_2} + \frac{(n_3 - \hat{n}_3)^2}{\hat{n}_3} + \frac{(n_4 - \hat{n}_4)^2}{\hat{n}_4}$$
$$= \frac{(260-245)^2}{245} + \frac{(100-90)^2}{90} + \frac{(130-140)^2}{140} + \frac{(10-25)^2}{25} = 11.744$$

- So sánh u_{qs} và W_α

$u_{qs} \in W_\alpha \Rightarrow$ bác bỏ H, chấp nhận $\bar{H}$

Vậy báo cáo khoa học đã cho là sai.

VII. Kiểm định sự độc lập

Tổng thể có 02 dấu hiệu X và Y

X	- Loại 1	Y	- Loại 1
	- Loại 2		- Loại 2
	...		...
	- Loại k		- Loại k

Ta cần kiểm định *X* và *Y* độc lập hay không với mức ý nghĩa α

Bước 1: Xác định giả thiết và giả thiết đối

H: *X* và *Y* độc lập

$\bar{H}$: *X* và *Y* phụ thuộc

Bước 2: Với α cho trước, tìm W_α.

$$W_\alpha = \left(\chi^2_\alpha\left[(k-1)(l-1)\right]; +\infty\right)$$

Bước 3: Chọn 1 mẫu có kích thước n. Lập bảng tần số:

X \ Y	Loại 1	Loại 2		Loại ℓ	
Loại 1	n_{11} $\hat{n}_{11} = \alpha_1\beta_1/n$	n_{12} $\hat{n}_{12} = \alpha_1\beta_2/n$		n_{1l} $\hat{n}_{1l} = \alpha_1\beta_l/n$	$\alpha_1 = n_{11} + n_{12} + \ldots + n_{1l}$
Loại 2	n_{21} $\hat{n}_{21} = \alpha_2\beta_1/n$	n_{22} $\hat{n}_{22} = \alpha_2\beta_2/n$		n_{2l} $\hat{n}_{2l} = \alpha_2\beta_l/n$	$\alpha_2 = n_{21} + n_{22} + \ldots + n_{2l}$
Loại k	n_{k1} $\hat{n}_{k1} = \alpha_k\beta_1/n$	n_{k2} $\hat{n}_{k2} = \alpha_k\beta_2/n$		n_{kl} $\hat{n}_{kl} = \alpha_k\beta_l/n$	$\alpha_k = n_{k1} + n_{k2} + \ldots + n_{kl}$
	$\beta_1 = n_{11} + n_{21} + \ldots + n_{k1}$	$\beta_2 = n_{12} + n_{22} + \ldots + n_{k2}$		$\beta_\ell = n_{1l} + n_{2l} + \ldots + n_{kl}$	

Giá trị quan sát:

$$u_{qs} = \frac{(n_{11} - \hat{n}_{11})^2}{\hat{n}_{11}} + \frac{(n_{12} - \hat{n}_{12})^2}{\hat{n}_{12}} + \ldots + \frac{(n_{kl} - \hat{n}_{kl})^2}{\hat{n}_{kl}}$$

Bước 4: So sánh u_{qs} và W_α

Ví dụ 8: Nghiên cứu 200 cặp vợ chồng có thời gian kết hôn trên 5 năm để tìm hiểu có mối quan hệ giữa thời gian tìm hiểu trước hôn nhân và tình trạng hiện tại của cuộc hôn nhân hay không. Có 3 mức độ thời gian tìm hiểu trước hôn nhân là ngắn, trung bình và dài. Cũng có 3 mức độ tình trạng hôn nhân hiện tại là hạnh phúc, không hạnh phúc và ly dị. Phỏng vấn 200 người ta có số liệu sau:

	Ngắn	Trung bình	Dài
Hạnh phúc	38	58	54
Không hạnh phúc	12	14	4
Ly dị	10	8	2

Với độ tin cậy 95%, đánh giá xem có mối quan hệ giữa thời gian tìm hiểu trước hôn nhân và tình trạng hiện tại của cuộc hôn nhân hay không?

Giải

Vợ chồng → Thời gian tìm hiểu (*X*)

Vợ chồng → Hạnh phúc sau cưới (*Y*)

Ta có:

X: Dài, Trung bình, Ngắn

Y: Hạnh phúc, Không hạnh phúc, Ly dị

Ta cần kiểm định *X* và *Y* có độc lập hay không với $\alpha = 5\%$

Xác định giả thiết và giả thiết đối

$$\begin{cases} H: & X \text{ và } Y \text{ độc lập} \\ \overline{H}: & X \text{ và } Y \text{ phụ thuộc} \end{cases}$$

Với $\alpha = 5\%$, tìm W_α:

$$W_\alpha = \left(\chi^2_\alpha\left[(k-1)(l-1)\right]; +\infty\right) = \left(\chi^2_{0.05}(4); +\infty\right) = (9.488; +\infty)$$

Phân tích mẫu có *n=200*

Lập bảng tần số:

Y / X	**Hạnh phúc**	**Không hạnh phúc**	**Ly dị**	
Dài	$n_{11} = 54$ $\hat{n}_{11} = \frac{60*150}{200} = 45$	$n_{12} = 4$ $\hat{n}_{12} = \frac{60*30}{200} = 9$	$n_{13} = 2$ $\hat{n}_{13} = \frac{60*20}{200} = 6$	**60** = 54+4+2
Trung bình	$n_{21} = 58$ $\hat{n}_{21} = \frac{80*150}{200} = 60$	$n_{22} = 14$ $\hat{n}_{22} = \frac{80*30}{200} = 12$	$n_{23} = 8$ $\hat{n}_{23} = \frac{80*20}{200} = 8$	**80** = 58+14+8
Ngắn	$n_{31} = 38$ $\hat{n}_{31} = \frac{60*150}{200} = 45$	$n_{32} = 12$ $\hat{n}_{32} = \frac{60*30}{200} = 9$	$n_{33} = 10$ $\hat{n}_{33} = \frac{60*20}{200} = 6$	**60** = 38+12+10
	150=54+58+38	**30**=4+14+12	**20**=2+8+10	

Giá trị quan sát:

$$u_{qs} = u_{qs} = \frac{(n_{11} - \hat{n}_{11})^2}{\hat{n}_{11}} + \frac{(n_{12} - \hat{n}_{12})^2}{\hat{n}_{12}} + \ldots + \frac{(n_{33} - \hat{n}_{33})^2}{\hat{n}_{33}}$$
$$= \frac{(54-45)^2}{45} + \frac{(4-9)^2}{9} + \ldots + \frac{(10-6)^2}{6} = 12.4$$

So sánh u_{qs} và W_α:

Ta có: $u_{qs} \in W_\alpha$ nên bác bỏ H, chấp nhận $\overline{H}$

Vậy là thời gian tìm hiểu ảnh hưởng đến hạnh phúc sau hôn nhân.