Chương 6: Kiểm định giả thuyết thống kê

Việc xác định qui luật xác suất của các biến có mặt trong tổng thể là một điều cần thiết trong xử lí số liệu. Bài toán ước lượng tham số mới giải quyết việc ước lượng tham tham số có mặt trong phân phối xác suất của tổng thể. Trong chương này chúng ta sẽ xây dựng các qui tắc đánh giá giả thuyết về các tham số, giả thuyết về các qui luật xác suất dựa trên mẫu ngẫu nhiên Qua các qui tắc kiểm định, người học có thể biết được cách xây dựng các giả thuyết và đối thuyết trong từng trường hợp cụ thể. Bài toán kiểm định giả thuyết thống kê là một bài toán lớn và quan trọng của thống kê toán học. Vì thời lượng chương trình có hạn, giáo trình chỉ đề cập tới một số qui tắc kiểm định thông dụng nhất. Một số qui tắc phi tham số giới thiệu trong giáo trình được đơn giản hóa bằng cách thay các thống kê dùng để kiểm định các qui tắc này bởi các qui luật xấp xỉ tương ứng.

I. Giả thuyết - Đối thuyết

1. Giả thuyết: Một mệnh đề (một câu khẳng định) về một vấn đề chưa biết nào đó được gọi là một giả thuyết. Các mệnh đề sau đều được gọi là các giả thuyết:

Vào năm 2010 con người sẽ có mặt trên sao hoả

Tham số $\theta=\theta_0$

Tham số $\theta\in[\theta_1;\theta_2]$

$X \sim N(\mu;\sigma^2)$

Sự kiện A độc lập với sự kiện B

Ta thường dùng H_0 để chỉ một giả thuyết. Giả thuyết là một mệnh đề nên có thể đúng hoặc không đúng. Tuy nhiên để kiểm tra tính đúng của một mệnh đề ta phải dựa trên tiêu chí thế nào là một mệnh đề đúng. Để khẳng định tính đúng sai của một mệnh đề ta thường kiểm tra mệnh đề này có thoả một số yêu cầu nào đó hay không hoặc đưa ra một mệnh đề khác trái với mệnh đề đã cho, trên cơ sở thực tế ta đưa ra quyết định coi mệnh đề ban đầu là đúng hoặc mệnh đề mới đưa ra là đúng. Trong thống kê ta sẽ theo hướng thứ hai.

2. Đối thuyết: Một mệnh đề trái với giả thuyết được gọi là một đối thuyết. Ta thường dùng H_1 để chỉ đối thuyết.

Ví dụ 1: H_0: Vào năm 2010 con người sẽ có mặt trên sao hoả.

Các mệnh đề sau là đối thuyết của giả thuyết H_0

H_1: Vào năm 2020 con người mới có mặt trên sao hoả

H_1: Vào năm 2010 con người chưa thể có mặt trên sao hoả

Ví dụ 2: H_0: $X \sim N(\mu;\sigma^2)$

Các đối thuyết của giả thuyết trên có thể là

H_1: $X \sim B(n, p)$ hoặc H_1: X không có phân phối chuẩn $N(\mu;\sigma^2)$

Nhận xét:

* Giả thuyết H_0 có thể đứng độc lập
* Đối thuyết phải đi kèm với mệnh đề trước đó được gọi là giả thuyết
* Mỗi giả thuyết có thể có nhiều đối thuyết khác nhau

* Một mệnh đề là giả thuyết trong trường hợp này có thể là đối thuyết trong trường hợp khác.

3. Giả thuyết thống kê và đối thuyết thống kê

Những giả thuyết và đối thuyết nói tới tham số có mặt trong qui luật xác suất của các đặc trưng có mặt trong tổng thể hoặc đề cập đến qui luật phân phối xác suất của những đặc trưng này được gọi là các giả thuyết và đối thuyết thống kê.

Ví dụ: H_0 : $X \sim N(\mu;\sigma^2)$

H_1 : $X \sim B(n, p)$

Hoặc H_0 : $\theta = \theta_0$

H_1 : $\theta \neq \theta_1$

là các giả thuyết và đối thuyết thống kê

4. Giả thuyết và đối thuyết tham số

Các giả thuyết và đối thuyết nói về tham số có mặt trong qui luật phân phối xác suất của tổng thể được gọi là các giả thuyết và đối thuyết tham số.

Ví dụ: Biết đặc trưng X ở tổng thể có phân phối chuẩn $N(\mu;\sigma^2)$

H_0 : $\mu=\mu_0$

H_1 : $\mu=\mu_1$ $(\mu_1 \neq \mu_2)$ hoặc: H_1 : $\mu \neq \mu_0$

là các giả thuyết và đối thuyết tham số

4.1 Giả thuyết đơn - Đối thuyết đơn

Giả thuyết đơn là giả thuyết trong đó tham số nhận một giá trị cụ thể nào đó.

Đối thuyết đơn là đối thuyết trong đó tham số nhận một giá trị cụ thể nào đó.

Ví dụ: Biết X~ B(n, p)

H_0 : $p = p_0$ là giả thuyết đơn

H_1 : $p = p_1$; ($p_1 \neq p_0$) là đối thuyết đơn của giả thuyết vừa nêu

4.2 Giả thuyết hợp - Đối thuyết hợp

Các giả thuyết hoặc đối thuyết trong đó tham số nhận hơn một giá trị gọi là giả thuyết hợp và đối thuyết hợp.

Ví dụ: Biết: $X \sim N(\mu;\sigma^2)$

H_0 : $\mu \in [\mu_1;\mu_2]$ là giả thuyết hợp

H_1 : $\mu < \mu_1$ hoặc $\mu > \mu_2$ là các đối thuyết hợp tương ứng với giả thuyết H_0

5. Giả thuyết và đối thuyết phi tham số: Những giả thuyết và đối thuyết thống kê không phải là các giả thuyết và đối thuyết tham số được gọi là các giả thuyết và đối thuyết phi tham số.

Ví dụ: H_0 : $X \sim N(\mu;\sigma^2)$

H_1 : $X \sim B(n, p)$

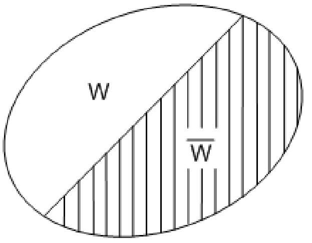
Hoặc: H_0 : A độc lập với B

H_1 : A không độc lập với B

là các giả thuyết và đối thuyết phi tham số

6. Kiểm định giả thuyết thống kê

Từ mẫu đã cho ta xây dựng một qui tắc chấp nhận giả thuyết H_0 (tương ứng với việc bác bỏ đối thuyết H_1) hoặc bác bỏ giả thuyết H_0 (tương ứng với việc chấp nhận đối thuyết H_1) được gọi là bài toán kiểm định một giả thuyết thống kê. Việc đưa ra một qui tắc chấp nhận hoặc bác bỏ giả thuyết H_0 dựa trên mẫu đã cho tương đương với việc xây dựng một qui tắc chia không gian mẫu V ra làm hai phần W và $\overline{W}$



Hình 1

Nếu mẫu ($X_1, X_2, \ldots, X_n$) $\in$ W ta quyết định bác bỏ giả thuyết H_0

Nếu mẫu ($X_1, X_2, \ldots, X_n$) $\in \overline{W}$ ta quyết định chấp nhận giả thuyết H_0

Kiểm định một giả thuyết thống kê không phải là một phép chứng minh về tính đúng hoặc không đúng của giả thuyết. Kiểm định một giả thuyết thống kê thực chất là xây dựng một qui tắc hành động dựa vào mẫu đã có đưa ra quyết định lựa chọn giả thuyết H_0 hoặc đối thuyết H_1

7. Các loại sai lầm: Với một qui tắc hành động chấp nhận hay bác bỏ H_0 ta có thể mắc phải các loại sai lầm sau:

7.1. Sai lầm loại 1: Bác bỏ giả thuyết H_0 khi H_0 đúng

Điều này có nghĩa là giả thuyết H_0 đúng nhưng mẫu ($X_1, X_2, \ldots, X_n$) $\in$ W nên ta bác bỏ H_0. Tương ứng với sai lầm loại 1 là *xác suất sai lầm loại 1:* $P(W/H_0) = \alpha$

7.2. Sai lầm loại 2: Chấp nhận giả thuyết H_0 khi H_0 sai. Điều này cũng có nghĩa là:

Khi H_0 sai (tức là coi H_1 đúng) nhưng mẫu ngẫu nhiên ($X_1, X_2, \ldots, X_n$) $\in \overline{W}$ nên ta chấp nhận H_0. Tương ứng với sai lầm loại 2 là *xác suất sai lầm loại 2:* $P(\overline{W}/H_1) = \beta$.

Ta có nhận xét:

Xác suất sai lầm loại 1: $P(W/H_0) = \alpha$ là xác suất bác bỏ H_0 mà thực ra H_0 đúng được gọi là mức ý nghĩa của bài toán kiểm định. Khi đó xác suất để chấp nhận H_0 khi H_0 đúng là $1 - \alpha$.

Xác suất sai lầm loại 2: $P(\overline{W}/H_1) = \beta$ là xác suất chấp nhận H_0 khi H_0 sai. Vậy xác suất bác bỏ H_0 khi H_0 sai là $1 - \beta$. Giá trị $1 - \beta$ được gọi là lực lượng của phép kiểm định. Mong muốn của người làm thống kê là xây một qui tắc chấp nhận hoặc bác bỏ một giả thuyết sao cho xác suất cả hai loại sai lầm càng nhỏ càng tốt. Tuy nhiên ta có

$$P(W/H_0) + P(\overline{W}/H_0) = 1 \quad ; \quad P(W/H_1) + P(\overline{W}/H_1) = 1$$

Từ đây suy ra khi α giảm thì β tăng và ngược lại. Với mẫu có kích thước cố định, để xây dựng một qui tắc hành động chấp nhận hoặc bác bỏ giả thuyết ta có thể đi theo một trong hai hướng sau:

Hướng thứ nhất: Cố định xác suất sai lầm loại 1 xây dựng một qui tắc sao cho xác suất sai lầm loại 2 là nhỏ nhất hoặc có thể chấp nhận được.

Hướng thứ hai ngược lại với hướng thứ nhất

Do đối thuyết H_1 thường là mệnh đề hợp (là hợp của các mệnh đề) nên việc cố định xác suất sai lầm loại hai là phức tạp và khó khả thi. Trong giáo trình này chúng ta sẽ đi theo hướng thứ nhất để xây dựng qui tắc kiểm định giả thuyết. Với mỗi cặp giả thuyết và đối thuyết đã cho, không phải lúc nào cũng tồn tại hoặc tìm được một qui tắc sao cho lực lượng của phép kiểm định $1 - \beta$ là lớn nhất. Những qui tắc đưa ra trong giáo trình này là những qui tắc thông dụng.

II. Kiểm định các giả thuyết tham số

1. Giả thuyết đơn - Đối thuyết đơn

Cặp giả thuyết: H_0: $\theta = \theta_0$

Đối thuyết H_1: $\theta = \theta_1 \quad (\theta_1 \neq \theta_0)$

là cặp giả thuyết và đối thuyết đơn

* Một qui tắc kiểm định với cặp giả thuyết - đối thuyết đơn được gọi là mạnh nhất nếu nó có lực lượng của phép kiểm định là lớn nhất

* Định lý Neyman - Pearson đã chỉ ra rằng: Nếu đặc trưng X ở tổng thể có hàm mật độ $f(x, \theta)$ thì tồn tại qui tắc mạnh nhất kiểm định cặp giả thuyết - đối thuyết đơn vừa nêu. Việc phát biểu và chứng minh định lý Neyman - Pearson không được nêu ra trong giáo trình này. Người đọc muốn biết có thể tham khảo ở các sách đã dẫn ra ở cuối giáo trình.

2. Giả thuyết đơn - Đối thuyết hợp.

Giả thuyết H_0 : $\theta=\theta_0$

Với đối thuyết: H_1 : $\theta \in D$; (D là một miền không chứa θ_0)

được gọi là cặp giả thuyết đơn với đối thuyết hợp.

Nhận thấy rằng: Với mỗi $\theta \in D$, sai lầm loại 2: $\beta = \beta(\theta)$ là một hàm số xác định trên D. Qui tắc kiểm định cặp giả thuyết, đối thuyết trên sao cho $\beta(\theta)$ cực tiểu $\forall\ \theta \in D$ được gọi là qui tắc kiểm định mạnh đều nhất.

Ví dụ : Biết biến X ở tổng thể có phân phối chuẩn $N(\mu;\sigma^2)$ với σ^2 đã biết. Xét cặp giả thuyết, đối thuyết đơn

H_0 : $\mu=\mu_0$

H_1 : $\mu=\mu_1 \quad (\mu_1 > \mu_0)$

Với mức ý nghĩa α, qui tắc mạnh nhất để kiểm định cặp giả thuyết, đối thuyết trên là:

* *Bác bỏ H_0 nếu:* $$\frac{\overline{X}-\mu_0}{\sigma}\sqrt{n} > U_\alpha$$

* *Chấp nhận H_0 nếu:* $$\frac{\overline{X}-\mu_0}{\sigma}\sqrt{n} \leq U_\alpha$$

Bởi qui tắc kiểm định vừa nêu không phụ thuộc vào μ_1 mà chỉ cần yêu cầu $\mu_1 > \mu_0$ nên qui tắc trên cũng là qui tắc mạnh đều nhất kiểm định cặp giả thuyết đơn, đối thuyết hợp

H_0 : $\mu=\mu_0$

H_1 : $\mu>\mu_0$ với mức ý nghĩa α.

3. Đối thuyết một phía và hai phía

Xét giả thuyết đơn : H_0 : $\theta = \theta_0$

* Mệnh đề: H_1 : $\theta \neq \theta_0$ gọi là đối thuyết hai phía của H_0
* Mệnh đề: H_1 : $\theta < \theta_0$ gọi là đối thuyết phía trái của H_0
* Mệnh đề: H_1 : $\theta > \theta_0$ gọi là đối thuyết phía phải của H_0.

Không phải lúc nào cũng tồn tại qui tắc mạnh đều nhất để kiểm định giả thuyết H_0 với một trong ba đối thuyết vừa nêu. Các qui tắc kiểm định được giới thiệu trong giáo trình này hoặc là qui tắc mạnh đều nhất hoặc là qui tắc tốt và thông dụng trong thống kê . Qui tắc "tốt" ở đây có thể hiểu theo nghĩa: Qui tắc mạnh đều nhất là tối ưu toàn cục thì qui tắc " tốt" là tối ưu bộ phận

4. Kiểm định kì vọng của phân phối chuẩn khi phương sai đã biết.

Giả sử đặc trưng X ở tổng thể có phân phối chuẩn $N(\mu;\sigma^2)$ với σ^2 đã biết.

Từ mẫu ngẫu nhiên $(X_1, X_2,....X_n)$ ta xây dựng qui tắc kiểm định giả thuyết

H_0 : $\mu = \mu_0$ trong các trường hợp sau:

4.1. Trường hợp 1: Đối thuyết H_1 : $\mu \neq \mu_0$

Thống kê $Z = \dfrac{\overline{X} - \mu}{\sigma}\sqrt{n}$ có phân phối chuẩn tắc.

Với $t \in (0, 1)$ xác suất $P[|\dfrac{\overline{X} - \mu}{\sigma}\sqrt{n}| > U_{\frac{t}{2}}] = t$. Nếu giả thuyết H_0 đúng thì xác suất

$P[|\dfrac{\overline{X} - \mu_0}{\sigma}\sqrt{n}| > U_{\frac{t}{2}}] = t$. Nếu lấy t = α là mức ý nghĩa của bài toán kiểm định thì

$P[|\dfrac{\overline{X} - \mu_0}{\sigma}\sqrt{n}| > U_{\frac{\alpha}{2}}] = \alpha$.

Bất đẳng thức $\left|\dfrac{\overline{X} - \mu_0}{\sigma}\sqrt{n}\right| > U_{\frac{\alpha}{2}} \Leftrightarrow$ Mẫu ngẫu nhiên $(X_1, X_2,....X_n) \in W$.

Điều này có nghĩa là: $P[\left|\dfrac{\overline{X} - \mu_0}{\sigma}\sqrt{n}\right| > U_{\frac{\alpha}{2}}] = P(W/ H_0) = \alpha$

Vậy qui tắc kiểm định: Giả thuyết H_0: $\mu = \mu_0$

Đối thuyết H_1: $\mu \neq \mu_0$.

với mức ý nghĩa α là:

Qui tắc 1: Nếu: $Z_T = \left|\dfrac{\overline{X} - \mu_0}{\sigma}\sqrt{n}\right| > U_{\frac{\alpha}{2}}$ quyết định bác bỏ H_0.

Nếu: $Z_T = \left|\dfrac{\overline{X} - \mu_0}{\sigma}\sqrt{n}\right| \leq U_{\frac{\alpha}{2}}$ quyết định chấp nhận H_0.

Với mẫu cụ thể $(x_1, x_2,....x_n)$ đã cho ta thực hiện bài toán kiểm định theo các bước sau:

Bước 1: Tính đại lượng $Z_T = \dfrac{|\overline{x} - \mu_0|}{\sigma}\sqrt{n}$

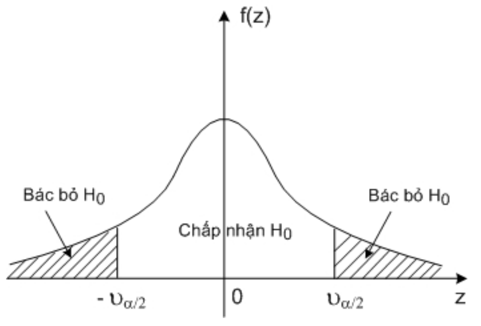
Bước 2: Tìm $U_{\frac{\alpha}{2}}$

Bước 3: So sánh hai giá trị trên rồi đưa ra quyết định:

Nếu: $Z_T > U_{\frac{\alpha}{2}}$ quyết định bác bỏ H_0.

Nếu: $Z_T \le U_{\frac{\alpha}{2}}$ quyết định chấp nhận H_0.

Hình vẽ sau mô tả miền chấp nhận và miền bác bỏ giả thuyết H_0



Hình 2

Các giá trị $-U_{\frac{\alpha}{2}}$ và $U_{\frac{\alpha}{2}}$ là các ngưỡng so sánh khi quyết định chấp nhận hay bác bỏ H_0.

Ví dụ 1: Từ tổng thể có phân phối chuẩn N(μ, 4) ta lấy mẫu có kích thước n = 9 và tìm được $\bar{x}$ = 21,20.Với mức ý nghĩa α = 0,05 hãy kiểm định giả thuyết :

H_0: $\mu = 20$

H_1: $\mu \neq 20$.

Tính $Z_T = \dfrac{|\overline{x} - \mu_0|}{\sigma}\sqrt{n} = \dfrac{|21,2 - 20,0|}{2}3 = 1,8$; Tìm $U_{0,025}$ = 1,96.

Vì Z_T =1,80 < 1,96 = $U_{0,025}$ ta quyết định chấp nhận H_0.

4.2 Trường hợp 2: Đối thuyết $\mu > \mu_0$

Tương tự như trường hợp trên thống kê $Z = \dfrac{\overline{X} - \mu}{\sigma}\sqrt{n}$ ta có phân phối chuẩn tắc. Dựa vào thống kê này ta có qui tắc kiểm định giả thuyết :

H_0 : $\mu = \mu_0$.

H_1 : $\mu > \mu_0$ với mức ý nghĩa α là:

Qui tắc 2: Nếu : $\frac{\overline{X} - \mu_0}{\sigma}\sqrt{n} > U_\alpha$ quyết định bác bỏ H_0.

Nếu: $\frac{\overline{X} - \mu_0}{\sigma}\sqrt{n} \leq U_\alpha$ quyết định chấp nhận H_0.

Với mẫu cụ thể ($x_1, x_2, \ldots, x_n$) ta cũng thực hiện bài toán kiểm định thao các bước:

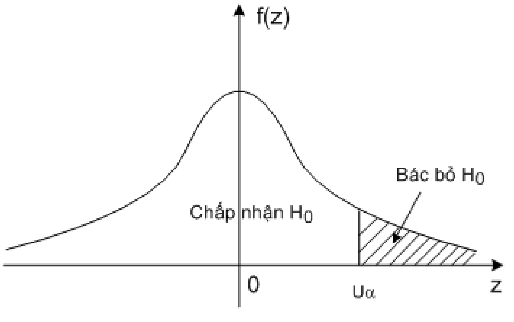
Bước 1: Tính $Z_T = \frac{\bar{x} - \mu_0}{\sigma}\sqrt{n}$

Bước 2: Tìm U_α

Bước 3: Nếu: $Z_T > U_\alpha$ quyết định bác bỏ H_0.

Nếu: $Z_T \leq U_\alpha$ quyết định chấp nhận H_0.

Miền chấp nhận và miền bác bỏ được mô tả bởi hình vẽ sau:



Hình 3

.

4.3 Trường hợp 3: Đối thuyết $\mu < \mu_0$

Tương tự như trường hợp trên thống kê $Z = \frac{\overline{X} - \mu}{\sigma}\sqrt{n}$ ta có phân phối chuẩn tắc. Dựa vào thống kê này ta có qui tắc kiểm định giả thuyết :

H_0 : $\mu = \mu_0$

H_1 : $\mu < \mu_0$ với mức ý nghĩa α là:

Qui tắc 3: Nếu: $Z_T = \dfrac{\overline{X}-\mu_0}{\sigma}\sqrt{n} < -U_\alpha$ quyết định bác bỏ H_0.

Nếu: $Z_T = \dfrac{\overline{X}-\mu_0}{\sigma}\sqrt{n} \geq -U_\alpha$ quyết định chấp nhận H_0.

Với mẫu cụ thể ($x_1, x_2,....x_n$) ta cũng thực hiện bài toán kiểm định thao các bước:

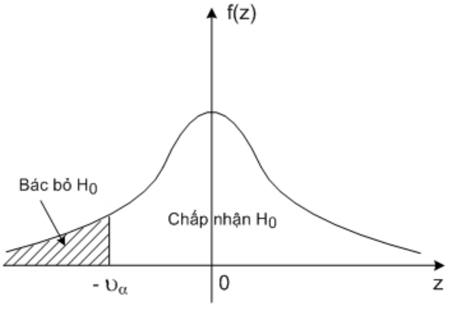
Bước 1: Tính $Z_T = \dfrac{\overline{x}-\mu_0}{\sigma}\sqrt{n}$

Bước 2: Tìm U_α .

Bước 3: Nếu $Z_T < -U_\alpha$ quyết định bác bỏ H_0.

Nếu $Z_T \geq -U_\alpha$ quyết định chấp nhận H_0.

Miền chấp nhận và miền bác bỏ được mô tả bởi hình vẽ sau:



Hình 4

.

Ví dụ 3: Trọng lượng X gói mì ăn liền tuân theo qui luât chuẩn N(μ, 25).Từ mẫu 25 gói mì ăn liền ta tìm được $\overline{x} = 82$ gam. Với mức ý nghĩa $\alpha = 0{,}05$ hãy kiểm định giả thuyết

H_0 : $\mu = 80$

H_1 : $\mu > 80$.

Ta có $Z_T = \dfrac{\overline{x}-\mu_0}{\sigma}\sqrt{n} = \dfrac{82-80}{5}.5 = 2{,}0$; $U_{0,05} = 1{,}68$.

Từ $Z_T > U_{0,05}$, áp dụng qui tắc 2 ta quyết định bác bỏ H_0

5. Kiểm định kỳ vọng của phân phối chuẩn khi không biết phương sai

Giả sử biến X ở tổng thể có phân phối chuẩn N($\mu;\sigma^2$).

Từ mẫu ngẫu nhiên ($X_1, X_2, ..., X_n$) xây dựng quy tắc kiểm định giả thuyết

H_0: $\mu = \mu_0$ trong các trường hợp đối thuyết hai phía và đối thuyết một phía

5.1 *Kiểm định giả thuyết:* H_0: $\mu = \mu_0$

H_1: $\mu \neq \mu_0$

Xét thống kê $Z = \dfrac{\overline{X} - \mu}{S}\sqrt{n}$, với $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$

Do Z ~ T_{n-1} nên $\forall$ $t \in (0,1)$ ta có

$$P\left[\left|\frac{\overline{X} - \mu}{S}\sqrt{n}\right| > t_{\frac{t}{2},n-1}\right] = t \qquad (1)$$

Nếu H_0 đúng thì:

$$P\left[\left|\frac{\overline{X} - \mu_0}{S}\sqrt{n}\right| > t_{\frac{t}{2},n-1}\right] = t \qquad (2)$$

Lấy t = α là mức ý nghĩa của bài toán kiểm định ta có:

$$P\left[\left|\frac{\overline{X} - \mu_0}{S}\sqrt{n}\right| > t_{\frac{\alpha}{2},n-1}\right] = \alpha \qquad (3)$$

Bất đẳng thức:

$$P\left[\left|\frac{\overline{X} - \mu_0}{S}\right| > t_{\frac{\alpha}{2},n-1}\right] = \alpha \qquad \Leftrightarrow (X_1, X_2, ..., X_n) \in W \Rightarrow P(W/H_0) = \alpha.$$

Từ đây có quy tắc kiểm định

H_0: $\mu = \mu_0$

H_1: $\mu \neq \mu_0$ với mức ý nghĩa α là

Qui tắc 4: Nếu $\left|\dfrac{\overline{X} - \mu_0}{S}\sqrt{n}\right| > t_{\frac{\alpha}{2},n-1}$ quyết định bác bỏ H_0

Nếu $\left|\frac{\overline{X}-\mu_0}{S}\sqrt{n}\right| \le t_{\frac{\alpha}{2},n-1}$ ta chấp nhận H_0

Cũng như qui tắc 1: Với mẫu cụ thể ($x_1, x_2,....x_n$) ta tiến hành bài toán kiểm định theo các bước sau:

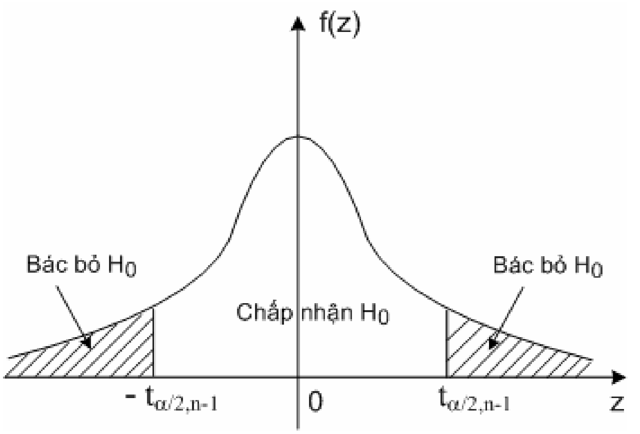
Bước 1: Tính $Z_T = \left|\frac{\overline{x}-\mu_0}{s}\sqrt{n}\right|$

Bước 2: Tìm $t_{\frac{\alpha}{2},n-1}$

Bước 3: Nếu $Z_T > t_{n-1,\frac{\alpha}{2}}$ quyết định bác bỏ H_0

Nếu $Z_T \le t_{\frac{\alpha}{2},n-1}$ quyết định chấp nhận H_0

Miền chấp nhận và miền bác bỏ H_0 cho bởi hình sau:



Hình 6

Ví dụ 1: Năng suất lúa là một đại lượng ngẫu nhiên X có phân phối chuẩn N($\mu;\sigma^2$). Điều tra năng suất giống lúa trên ở 200 ruộng ta được bảng các số liệu sau:

Năng suất tạ/ha	46	48	49	50	51	53	54	58
Số thửa ruộng	17	18	35	45	42	23	10	10

Với mức ý nghĩa α = 0,05. Hãy kiểm định cặp giả thuyết đối thuyết:

H_0: $\mu = 52$

H_1: $\mu \ne 52$

Để tính $\overline{x}$ và s ta lập bảng sau:

x_i	n_i	$x_i n_i$	$n_i x_i^2$
46	17	782	35972
48	18	804	41472
49	35	1715	84035
50	45	2250	112500
51	42	2142	109242
53	23	1219	64607
54	10	540	29160
58	10	580	33640
$\sum$	200	. 10032	. 510628

$$\overline{x} = 50{,}16,\ \hat{s}^2 = \overline{x^2} - \overline{x}^2 = 37{,}1144, s^2 = \frac{200\hat{s}^2}{199} = 37{,}3009;\ s = 6{,}11.$$

$$Z_T = \frac{|50{,}16 - 52{,}00|}{6{,}11}\sqrt{200} = 4{,}26,\ t_{0,025\ ,\ 199} = U_{0,025} = 1{,}96$$

$Z_T = 4{,}26 > 1{,}96 \Rightarrow$ bác bỏ H_0

5.2 *Trường hợp 2.* Đối thuyết H_1: $\mu < \mu_0$

Tương tự như trường hợp 1. Từ thống kê $Z = \frac{\overline{X} - \mu}{S}\sqrt{n}$ có phân phối T_{n-1} nếu H_0 đúng ta có qui tắc kiểm định cặp giả thuyết đối thuyết ở mức ý nghĩa α là

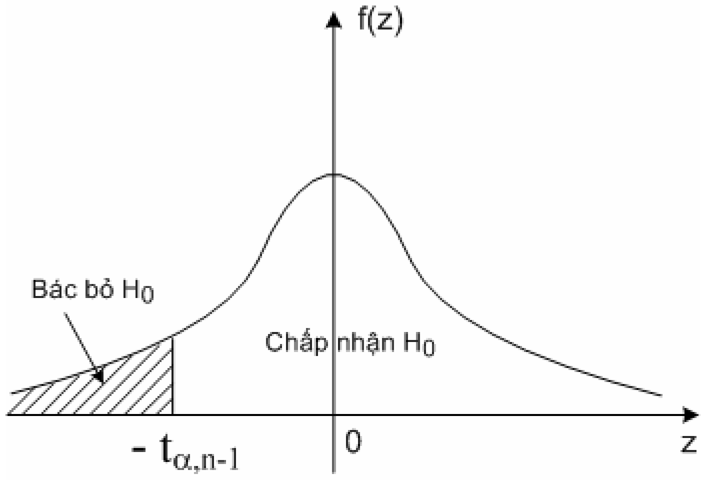
H_0: $\mu = \mu_0$

H_1: $\mu < \mu_0$

Qui tắc 5: Nếu $Z = \frac{\overline{X} - \mu_0}{S}\sqrt{n} < -t_{\alpha, n-1}$ bác bỏ H_0

Nếu $Z = \frac{\overline{X}-\mu_0}{S}\sqrt{n} \geq -\ t_{\alpha,n-1}$ chấp nhận H_0

Hình vẽ sau cho miền chấp nhận và bác bỏ H_0



Hình 7

Dựa vào qui tắc 5, các bước thực hiện bài toán kiểm định như trường hợp 1

5.3 *Đối thuyết H_1:* $\mu > \mu_0$

Tương tự như trường hợp 2, qui tắc kiểm định cặp giả thuyết đối thuyết

H_0: $\mu = \mu_0$

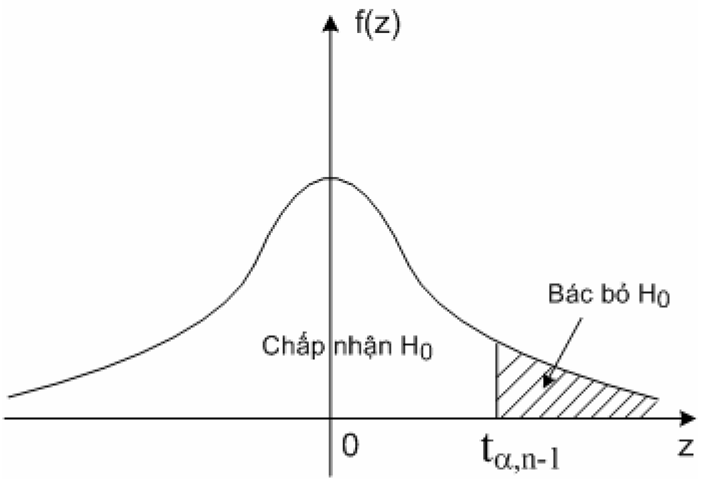
H_1: $\mu > \mu_0$

ở mức ý nghĩa α

Qui tắc 6: Nếu: $Z_T = \frac{\overline{X}-\mu_0}{S}\sqrt{n} > t_{\alpha,n-1}$, bác bỏ H_0

Nếu: $Z = \frac{\overline{X}-\mu_0}{S}\sqrt{n} \leq t_{\alpha,n-1}$ chấp nhận H_0

Hình vẽ sau cho miền chấp nhận và bác bỏ H_0



Hình 8

6.Kiểm định phương sai của phân phối chuẩn

Từ mẫu ngẫu nhiên (X_1, X_2, ... ,X_n) của tổng thể có phân phối chuẩn N($\mu;\sigma^2$), với mức ý nghĩa α xây dựng qui tắc kiểm định cặp giả thuyết , đối thuyết :

$$H_0: \quad \sigma^2 = \sigma_0^2$$

$$H_1: \quad \sigma^2 > \sigma_0^2$$

Thống kê $Z = \dfrac{(n-1)S^2}{\sigma^2}$ có phân phối χ^2_{n-1}, với $t \in (0,1)$ ta có

$$P[\frac{(n-1)S^2}{\sigma^2} > \chi^2_{t,n-1}] = t \qquad (1)$$

Nếu H_0 đúng thì $P[\dfrac{(n-1)S^2}{\sigma_0{}^2} > \chi^2_{t,n-1}] = t$ (2)

Cho $t = \alpha$ là mức ý nghĩa của bài toán kiểm định có:

$$P[\frac{(n-1)S^2}{\sigma_0{}^2} > \chi^2_{\alpha,n-1}] = \alpha \qquad (3)$$

Bất đẳng thức:

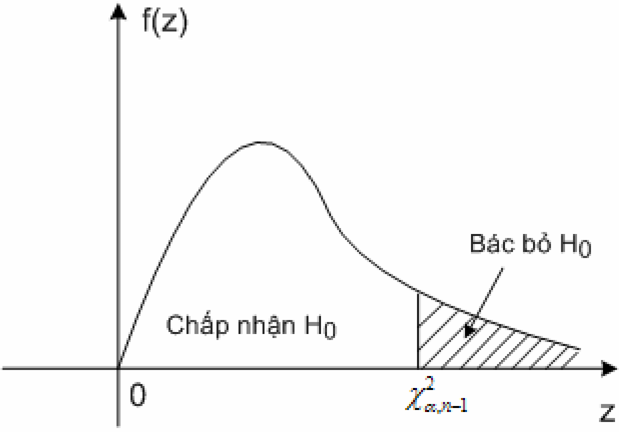
$\dfrac{(n-1)S^2}{\sigma_0{}^2} > \chi^2_{\alpha,n-1}$ tương đương với (X_1, X_2, ... ,X_n) $\in W \Leftrightarrow P(W/H_0) = \alpha$

Qui tắc kiểm định H_0 là:

Qui tắc 7: Nếu $\dfrac{(n-1)S^2}{{\sigma_0}^2} > \chi^2_{\alpha,n-1}$ bác bỏ H_0

Nếu $\dfrac{(n-1)S^2}{{\sigma_0}^2} \le \chi^2_{\alpha,n-1}$ chấp nhận H_0

Hình vẽ sau chỉ miền chấp nhận và miền bác bỏ H_0



Hình 9

Từ mẫu cụ thể ($x_1, x_2, \ldots x_n$) theo qui tắc ta thực hiện bài toán kiểm định cặp giả thuyết , đối thuyết: H_0: $\sigma^2 = \sigma_0^2$

H_1: $\sigma^2 > \sigma_0^2$ ở mức ý nghĩa α theo các bước sau:

Bước 1: Tính $Z_T = \dfrac{(n-1)s^2}{{\sigma_0}^2}$

Bước 2: Tìm $\chi^2_{\alpha,n-1}$

Bước 3: Nếu $Z_T > \chi^2_{\alpha,n-1}$ ta bác bỏ H_0

Nếu $Z_T \le \chi^2_{\alpha,n-1}$ ta chấp nhận H_0

Ví dụ: Một máy sản xuất các tấm chất dẻo được thường xuyên theo dõi về độ dày của sản phẩm Biết độ dày X của các tấm chất dẻo tuân theo qui luật chuẩn $N(\mu;\sigma^2)$. Nếu độ lệch chuẩn vượt quá 0,3mm thì chất lượng sản phẩm không được đảm bảo về kĩ thuật. Người ta chọn ngẫu nhiên 10 tấm chất dẻo rồi đo độ dày của mỗi tấm và được kết quả sau (đơn vị đo mm)

22,0 ; 22,6 ; 23,2 ; 22,7 ; 22,5 ; 22,8 ; 22,5 ; 22,8 ; 22,9 ; 23,0.

Từ yêu cầu của thực tế với mức ý nghĩa α = 0,05 hãy lập cặp giả thuyết và đối thuyết thích hợp đánh giá tình trạng làm việc của máy sản xuất các tấm chất dẻo trên

Độ lệch chuẩn ở mức cho phép không vượt quá 0,3 mm tương ứng với phương sai σ^2 không vượt quá 0,09 mm^2. Ta có cặp giả thuyết đối thuyết sau:

H_0: $\sigma^2 = 0{,}09$

H_1: $\sigma^2 > 0{,}09$

Với mẫu đã cho ta có: $\bar{x} = 2{,}27$, $s^2 = 0{,}1089$, $Z_T = \dfrac{(n-1)s^2}{\sigma_0^2} = \dfrac{9.0{,}1089}{0{,}09} = 10{,}89$

ở mức ý nghĩa 0,05 ta có:

$\chi^2_{9,0,05} = 16{,}92$; $Z_T = 10{,}89 < 16{,}92 = \chi^2_{0,05,9}$ quyết định chấp nhận H_0, điều này có nghĩa là máy sản xuất các tấm dẻo vẫn hoạt động bình thường.

Với cặp giả thuyết H_0: $\sigma^2 = \sigma_0^2$

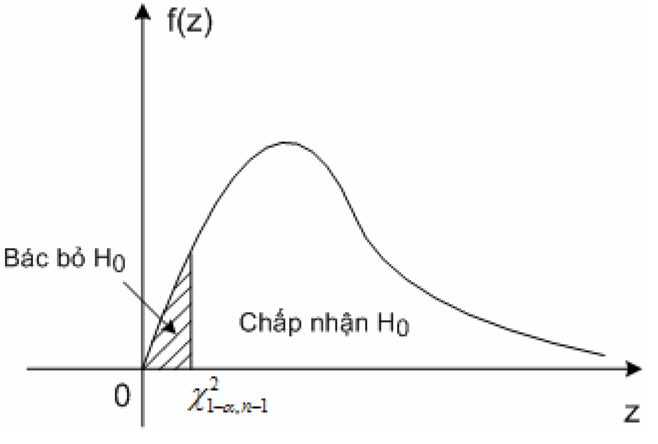
H_1: $\sigma^2 < \sigma_0^2$

Ta có qui tắc kiểm định là

Qui tắc 8: Nếu $\dfrac{(n-1)s^2}{\sigma_0^2} < \chi^2_{1-\alpha,n-1}$ ta bác bỏ H_0

Nếu $\dfrac{(n-1)s^2}{\sigma_0^2} \geq \chi^2_{1-\alpha,n-1}$ ta chấp nhận H_0

Miền bác bỏ và miền chấp nhận H_0 cho bởi hình



H ình 10

Với cặp giả thuyết : H_0: $\sigma^2 = \sigma_0^2$

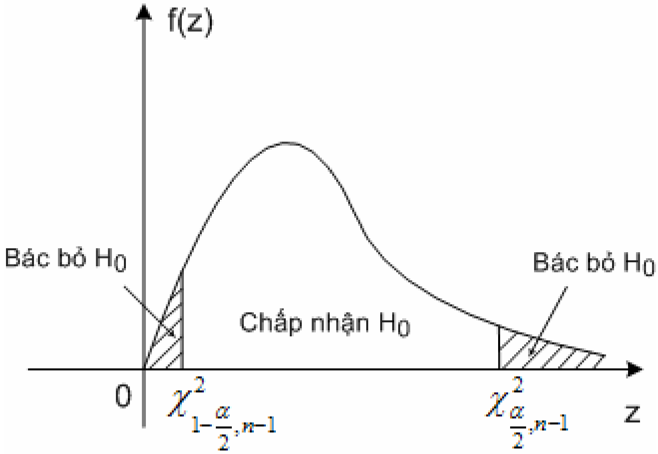
H_1: $\sigma^2 \neq \sigma_0^2$

Ta có qui tăc kiểm định là:

Qui tắc 9: Nếu $\dfrac{(n-1)s^2}{\sigma_0^2} < \chi^2_{1-\frac{\alpha}{2},n-1}$ hoặc $\dfrac{(n-1)s^2}{\sigma_0^2} > \chi^2_{\frac{\alpha}{2},n-1}$ bác bỏ H_0

Nếu $\chi^2_{1-\frac{\alpha}{2},n-1} \leq \dfrac{(n-1)s^2}{\sigma_0^2} \leq \chi^2_{\frac{\alpha}{2},n-1}$ chấp nhận H_0.

Miền bác bỏ và miền chấp nhận H_0 cho bởi hình



Hình 11

Nhận xét: Do phương sai là số do đặc trưng cho sai số nên các bài toán kiểm định về phương sai người ta thường kiểm định với đối thuyết một phía.

7. Kiểm định xác suất .

Xác suất xuất hiện sự kiện A ở mỗi phép thử P(A) = p. Tiến hành n phép thử độc lập có n_A lần xuất hiện A. Với mức ý nghĩa α ta xây dựng qui tắc kiểm định cặp giả thuyết, đối thuyết :

H_0: $p = p_0$

H_1: $p \neq p_0$

Xét thống kê $Z = \frac{f-p}{\sqrt{pq}}\sqrt{n}$, $f = \frac{n_A}{n}$, q =1 - p, thống kê này theo định lý giới hạn trung tâm có phân phối xấp xỉ chuẩn tắc. Nếu H_0 đúng thì

$Z = \frac{f-p_0}{\sqrt{p_0q_0}}\sqrt{n}$ có phân phối xấp xỉ chuẩn tắc. Giống như bài toán kiểm định kỳ vọng của phân phối chuẩn, ta có qui tắc kiểm định cặp giả thuyết đối thuyết trên là:

Qui tắc 10: Nếu: $\left|\frac{f-p_0}{\sqrt{p_0q_0}}\sqrt{n}\right| > U_{\frac{\alpha}{2}}$ ta bác bỏ H_0

Nếu: $\left|\frac{f-p_0}{\sqrt{p_0q_0}}\sqrt{n}\right| \leq U_{\frac{\alpha}{2}}$ ta chấp nhận H_0.

Ví dụ: Xét nghiệm 1000 mẫu máu của những người dân ở vùng Tây Nguyên ta thấy có 232 mẫu máu có ký sinh trùng sốt rét. Hãy kiểm định :

H_0: $p = 0,2$

H_1: $p \neq 0,2$, mức ý nghĩa $\alpha = 0,05$.

Ta có $f = \frac{232}{1000} = 0,0232$, $Z_T = \left|\frac{f-p_0}{\sqrt{p_0q_0}}\sqrt{n}\right| = \left|\frac{f-0,2}{\sqrt{0,2.0,8}}\sqrt{1000}\right| = 2,53$.

$U_{0,025}$= 1,96 ; Z_T = 2,53 > 1,96 vậy giả thuyết H_0 bị bác bỏ.

Chú ý 1: Với cặp giả thuyết đối thuyết

H_0: $p = p_0$

H_1: $p > p_0$

Có qui tắc kiểm định sau:

Qui tắc 11: Nếu $\frac{f - p_0}{\sqrt{p_0 q_0}}\sqrt{n} > U_\alpha$ bác bỏ H_0

. Nếu $\frac{f - p_0}{\sqrt{p_0 q_0}}\sqrt{n} \le U_\alpha$ chấp nhận H_0.

Chú ý 2: Với cặp giả thuyết đối thuyết

H_0: $p = p_0$

H_1: $p < p_0$

Ta có qui tắc kiểm định sau:

Qui tắc 12: Nếu $\frac{f - p_0}{\sqrt{p_0 q_0}}\sqrt{n} <- U_\alpha$ bác bỏ H_0

. Nếu $\frac{f - p_0}{\sqrt{p_0 q_0}}\sqrt{n} \ge U_\alpha$ chấp nhận H_0.

Ví dụ: Một kho bảo quản hạt giống được xem là chưa đảm bảo kỹ thuật nếu tỷ lệ hạt nảy mầm dưới 70%. Người ta lấy 500 hạt trong kho đem gieo và thấy có 340 hạt nảy mầm.Với mức ý nghĩa $\alpha = 0{,}05$ hãy lập giả thuyết và đối thuyết thích hợp để xét xem kho bảo quản hạt giống trên đã đảm bảo kỹ thuật hay chưa.

Yêu cầu trên tương đương với việc kiểm định cặp giả thuyết đối thuyết :

H_0: $p = 0{,}7$

H_1: $p < 0{,}7$

Ta có $f = 0{,}68$, $1 - p_0 = 0{,}3$, $Z_T = \frac{f - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n} = \frac{0{,}68 - 0{,}7}{\sqrt{0{,}7.0{,}3}}\sqrt{500} = -1{,}79$

$U_{0,025} = 1{,}96$; $Z_T = -1{,}79 > -U_{0,025} = -1{,}96$

Với kết quả trên ta chưa đủ cơ sở để kết luận tiêu chuẩn kĩ thuật của kho bảo quản hạt giống là có vấn đề.

8. Kiểm định sự bằng nhau của hai kỳ vọng của hai phân phối chuẩn.

Từ hai mẫu ngẫu nhiên độc lập lấy từ hai tổng thể có phân phối chuẩn $N(\mu_X ; \sigma_X^2)$ $N(\mu_Y ; \sigma_Y^2)$) ta xây dựng qui tắc kiểm định cặp giả thuyết, đối thuyết:

H_0: $\mu_X = \mu_Y$

H_1: $\mu_X \ne \mu_Y$

8.1: Trường hợp σ_X^2 , σ_Y^2 đã biết.

Thống kê $Z = \frac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}$ có phân phối chuẩn tắc

Nếu H_0 đúng thì $Z = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}$ có phân phối chuẩn tắc, Khi đó: $P\left[\dfrac{|\overline{X}-\overline{Y}|}{\sqrt{\dfrac{\sigma_X^2}{n}+\dfrac{\sigma_Y^2}{m}}} > U_{\frac{\alpha}{2}}\right] = \alpha$

Từ đây có qui tắc bác bỏ H_0 là:

Qui tắc 13: Nếu $\dfrac{|\overline{X}-\overline{Y}|}{\sqrt{\dfrac{\sigma_X^2}{n}+\dfrac{\sigma_Y^2}{m}}} > U_{\frac{\alpha}{2}}$ quyết định bác bỏ H_0

Nếu $\dfrac{|\overline{X}-\overline{Y}|}{\sqrt{\dfrac{\sigma_X^2}{n}+\dfrac{\sigma_Y^2}{m}}} \le U_{\frac{\alpha}{2}}$ quyết định chấp nhận H_0.

Với hai mẫu cụ thể $(x_1, x_2, ..., x_n)$ và $(y_1, y_2, ... y_m)$ thực hiện bài toán kiểm định cặp giả thuyết , đối thuyết trên theo các bước sau:

Bước 1: Tính $Z_T = \dfrac{|\overline{x}-\overline{y}|}{\sqrt{\dfrac{\sigma_x^2}{n}+\dfrac{\sigma_y^2}{m}}}$

Bước 2: Tìm $U_{\frac{\alpha}{2}}$.

Bước 3: Nếu $Z_T > U_{\frac{\alpha}{2}}$ bác bỏ H_0

Nếu $Z_T \le U_{\frac{\alpha}{2}}$ chấp nhận H_0

8.2. Trường hợp σ_X^2; σ_Y^2 chưa biết nhưng n, m ≥ 30.(gọi là mẫu lớn)

Do thống kê $Z = \dfrac{\overline{X}-\overline{Y}-(\mu_X-\mu_Y)}{\sqrt{\dfrac{\hat{S}_X^2}{n}+\dfrac{\hat{S}_Y^2}{m}}}$ có phân phối xấp xỉ phân phối chuẩn tắc. Nếu H_0 đúng thì

$Z = \dfrac{\overline{X}-\overline{Y}}{\sqrt{\dfrac{\hat{S}_X^2}{n}+\dfrac{\hat{S}_Y^2}{m}}}$ có phân phối xấp xỉ chuẩn tắc

Cũng như ở trường hợp trên ta có qui tắc chấp nhận hoặc bác bỏ giả thuyết H_0 với đối thuyết hai phía là:

Qui tắc 14: Nếu: $\dfrac{|\overline{X}-\overline{Y}|}{\sqrt{\dfrac{\hat{S}_x^2}{n}+\dfrac{\hat{S}_y^2}{m}}} > U_{\frac{\alpha}{2}}$ quyết định bác bỏ H_0

Nếu: $\dfrac{|\overline{X}-\overline{Y}|}{\sqrt{\dfrac{\hat{S}_x^2}{n}+\dfrac{\hat{S}_y^2}{m}}} \le U_{\frac{\alpha}{2}}$ quyết định chấp nhận H_0.

8.3. Trường hợp σ_X^2, σ_Y^2 chưa biết nhưng biết $\sigma_X^2 = \sigma_Y^2$.

Thống kê $Z = \dfrac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{S\sqrt{\dfrac{1}{n} + \dfrac{1}{m}}}$ có phân phối T_{n+m-2}

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2 + \sum_{j=1}^{m}(Y_j - \overline{Y})^2}{n+m-2} = \frac{(n-1)S_x^2 + (m-1)S_y^2}{n+m-2}$$

Nếu H_0 đúng thì $Z = \dfrac{\overline{X} - \overline{Y}}{S\sqrt{\dfrac{1}{n} + \dfrac{1}{m}}}$ có phân phối T_{n+m-2} và có qui tắc kiểm định cặp giả thuyết, đối thuyết ở mức ý nghĩa α là:

H_0 : $\mu_X = \mu_Y$

H_1 : $\mu_X \neq \mu_Y$

Qui tắc 15: Nếu $\dfrac{\left|\overline{X} - \overline{Y}\right|}{S\sqrt{\dfrac{1}{n} + \dfrac{1}{m}}} > t_{\frac{\alpha}{2}n+m-2}$ quyết định bác bỏ H_0

Nếu $\dfrac{\left|\overline{X} - \overline{Y}\right|}{S\sqrt{\dfrac{1}{n} + \dfrac{1}{m}}} \le t_{\frac{\alpha}{2},n+m-2}$ quyết định chấp nhận H_0

Với hai mẫu cụ thể (x_1, x_2, ..., x_n) và (y_1, y_2,...y_m) thực hiện bài toán kiểm định cặp giả thuyết , đối thuyết trên theo các bước sau:

Bước 1: Tính $Z_T = \dfrac{\left|\overline{x} - \overline{y}\right|}{s\sqrt{\dfrac{1}{n} + \dfrac{1}{m}}}$ với $s^2 = \dfrac{(n-1)s_X^2 + (m-1)s_Y^2}{n+m-2}$

Bước 2: Tìm $t_{\frac{\alpha}{2},n+m-2}$

Bước 3: Nếu $Z_T > t_{\frac{\alpha}{2},n+m-2}$ quyết định bác bỏ H_0

Nếu $Z_T \le t_{\frac{\alpha}{2},n+m-2}$ quyết định chấp nhận H_0

Ví dụ: Điều tra năng suất của 8 thửa ruộng trồng giống lúa A và năng suất của 10 thửa ruộng trồng giống lúa B với cùng một điều kiện canh tác ta có kết quả sau:

X	40	38	40	42	44	41	36	39		
Y	41	44	38	42	40	45	39	37	43	41

Biết rằng X~N(μ_X ;σ^2), Y~N(μ_Y ;σ^2). Với mức ý nghĩa $\alpha = 0{,}05$ hãy kiểm định:

H_0: $\mu_X = \mu_Y$

H_1: $\mu_X \neq \mu_Y$

Ta có $\overline{x} = 40$, $\overline{y} = 41$, $s_X^2 = \dfrac{42}{7} = 6$, $s_Y^2 = 6{,}66$, $s^2 = 6{,}375$, $s = 2{,}52$.

$$Z_T = \frac{|\overline{x}-\overline{y}|}{s\sqrt{\frac{1}{n}+\frac{1}{m}}} = \frac{1}{2{,}52\sqrt{\frac{1}{8}+\frac{1}{10}}} = 0{,}835.$$

$t_{0,025\,,\,16} = 2{,}12$; $Z_T = 0{,}835 < t_{0,025\,,\,16} = 2{,}12$. Chấp nhận H_0.

Chú ý 1: Bài toán kiểm định cặp giả thuyết và đối thuyết trên còn được gọi là bài toán so sánh sự bằng nhau và khác nhau hai kỳ vọng của hai phân phối chuẩn.

Chú ý 2: Trong thực tế nhiều khi ta cần so sánh kỳ vọng của biến ngẫu nhiên X với kỳ vọng của biến ngẫu nhiên Y cộng thêm một hằng số μ_0 khi đó ta có bài toán kiểm định cặp giả thuyết đối thuyết sau:

H_0 : $\mu_X = \mu_Y + \mu_0$

H_1 : $\mu_X \neq \mu_Y + \mu_0$

Khi gặp bài toán này có các qui tắc kiểm định sau:

* Nếu biết σ_x^2, σ_y^2 qui tắc kiểm định cặp giả thuyết, đối thuyết trên là:

Qui tắc 16: Nếu: $\dfrac{|\overline{X}-\overline{Y}-\mu_0|}{\sqrt{\dfrac{\sigma_x^2}{n}+\dfrac{\sigma_y^2}{m}}} > U_{\frac{\alpha}{2}}$ quyết định bác bỏ H_0

Nếu: $\dfrac{|\overline{X}-\overline{Y}-\mu_0|}{\sqrt{\dfrac{\sigma_x^2}{n}+\dfrac{\sigma_y^2}{m}}} \le U_{\frac{\alpha}{2}}$ quyết định chấp nhận H_0.

* Nếu chưa biết σ_x^2, σ_y^2 nhưng mẫu n, m $\ge$ 30 có qui tắc sau:

Qui tắc 17: Nếu: $\dfrac{|\overline{X}-\overline{Y}-\mu_0|}{\sqrt{\dfrac{\hat{S}_x^2}{n}+\dfrac{\hat{S}_y^2}{m}}} > U_{\frac{\alpha}{2}}$ quyết định bác bỏ H_0

Nếu: $\dfrac{|\overline{X}-\overline{Y}-\mu_0|}{\sqrt{\dfrac{\hat{S}_x^2}{n}+\dfrac{\hat{S}_y^2}{m}}} \le U_{\frac{\alpha}{2}}$ quyết định chấp nhận H_0.

* Chưa biết σ_X^2, σ_Y^2 nhưng biết $\sigma_x^2 = \sigma_y^2$. Ta có qui tắc sau:

Qui tắc 18: Nếu: $\dfrac{|\overline{X}-\overline{Y}-\mu_0|}{S\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}} > t_{\frac{\alpha}{2},n+m-2}$ quyết định bác bỏ H_0

Nếu: $\dfrac{|\overline{X}-\overline{Y}-\mu_0|}{S\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}} \le t_{\frac{\alpha}{2},n+m-2}$ quyết định chấp nhận H_0.

Chú ý 3: Nếu không có thông tin gì về σ_x^2 và σ_y^2 mà kích thước của mẫu n, m < 30 ta tiến hành theo một trong hai cách sau:

* *Cách 1:* Thu thập thêm dữ liệu mẫu để n, m ≥ 30 sau đó sử dụng qui tắc kiểm định như đã nêu trong mục 4.2.
* *Cách 2:* Xây dựng cặp giả thuyết đối thuyết phụ:

H'_0: $\sigma_x^2 = \sigma_y^2$

H'_1: $\sigma_x^2 \neq \sigma_y^2$ với cùng mức ý nghĩa α.

Qui tắc kiểm định cặp giả thuyết đối thuyết này sẽ trình bày ở mục sau.
Nếu H'_0 được chấp nhận ta sử dụng qui tắc như đã nêu ở mục 4.3.
Nếu H'_0 bị bác bỏ thì về mặt lý thuyết việc so sánh hai kỳ vọng trên chưa giải quyết được. (Bài toán Behrens-Fisher).

9. Kiểm định sự bằng nhau của hai kỳ vọng với đối thuyết một phía.

Từ hai mẫu ngẫu nhiên độc lập:

$(X_1, X_2,....X_i,....X_n)$; $X_i \sim N(\mu_x, \sigma_x^2)$,

$(Y_1, Y_2,\ldots Y_j,\ldots Y_n)$; $Y_j \sim N(\mu_y, \sigma_y^2)$.

Ta xây dựng qui tắc kiểm định cặp giả thuyết đối thuyết :

H_0: $\mu_X = \mu_Y$

H_1: $\mu_X > \mu_Y$

9.1 Trường hợp 1: Khi σ_x^2, σ_y^2 đã biết.

Thống kê $Z = \dfrac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{\sqrt{\dfrac{\sigma_X^2}{n} + \dfrac{\sigma_Y^2}{m}}}$ có phân phối chuẩn tắc.

Nếu H_0 đúng thì $Z = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{\sigma_X^2}{n} + \dfrac{\sigma_Y^2}{m}}}$ có phân phối chuẩn tắc.

Từ đây ta có qui tắc kiểm định cặp giả thuyết đối thuyết trên là:

Qui tắc 19: Nếu: $\dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{\sigma_X^2}{n} + \dfrac{\sigma_Y^2}{m}}} > U_\alpha$ quyết định bác bỏ H_0

Nếu: $\dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{\sigma_X^2}{n} + \dfrac{\sigma_Y^2}{m}}} \leq U_\alpha$ quyết định chấp nhận H_0

9.2. Trường hợp 2: Khi σ_x^2, σ_y^2 chưa biết nhưng m, n ≥ 30.

Tương tự như trên ta có qui tăc kiểm định là:

Qui tắc 20: Nếu: $\dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{\hat{S}_X^2}{n} + \dfrac{\hat{S}_Y^2}{m}}} > U_\alpha$ quyết định bác bỏ H_0

Nếu: $\dfrac{\overline{X}-\overline{Y}}{\sqrt{\dfrac{\hat{S}_X^2}{n}+\dfrac{\hat{S}_Y^2}{m}}} \le U_\alpha$ quyết định chấp nhận H_0

9.3 Trường hợp 3: Khi σ_x^2, σ_y^2 chưa biết nhưng biết $\sigma_x^2 = \sigma_y^2$

Do biến $Z = \dfrac{\overline{X}-\overline{Y}-(\mu_X-\mu_Y)}{S\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}}$ có phân phối T_{n+m-2}

Nếu H_0 đúng thì $Z = \dfrac{\overline{X}-\overline{Y}}{S\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}}$ có phân phối T_{n+m-2}

Từ đây ta có qui tắc chấp nhận và bác bỏ H_0 là:

Qui tắc 21: Nếu $\dfrac{\overline{X}-\overline{Y}}{S\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}} > t_{\alpha,n+m-2}$ quyết định bác bỏ H_0

Nếu $\dfrac{\overline{X}-\overline{Y}}{S\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}} \le t_{\alpha,n+m-2}$ quyết định chấp nhận H_0.

Chú ý: Để kiểm định cặp giả thuyết đối thuyết

H_0 : $\mu_X = \mu_Y + \mu_0$

H_1 : $\mu_X > \mu_Y + \mu_0$

* Trường hợp σ_x^2, σ_y^2 đã biết qui tắc kiểm định là:

Qui tắc 22 : Nếu $\dfrac{\overline{X}-\overline{Y}-\mu_0}{\sqrt{\dfrac{\sigma_X^2}{n}+\dfrac{\sigma_Y^2}{m}}} > U_\alpha$ quyết định bác bỏ H_0

Nếu: $\dfrac{\overline{X}-\overline{Y}-\mu_0}{\sqrt{\dfrac{\sigma_X^2}{n}+\dfrac{\sigma_Y^2}{m}}} \le U_\alpha$ quyết định chấp nhận H_0

Trường hợp chưa biết σ_x^2, σ_y^2 nhưng n, m ≥ 30 ta có qui tắc kiểm định sau:

Qui tắc 23: Nếu: $\dfrac{\overline{X}-\overline{Y}-\mu_0}{\sqrt{\dfrac{\hat{S}_X^2}{n}+\dfrac{\hat{S}_Y^2}{m}}} > U_\alpha$ quyết định bác bỏ H_0

Nếu: $\dfrac{\overline{X}-\overline{Y}-\mu_0}{\sqrt{\dfrac{\hat{S}_X^2}{n}+\dfrac{\hat{S}_Y^2}{m}}} \le U_\alpha$ quyết định chấp nhận H_0

Trường hợp chưa biết σ_x^2, σ_y^2 nhưng biết $\sigma_x^2 = \sigma_y^2$ ta có qui tắc kiểm định sau:

Qui tắc 24: Nếu : $\dfrac{\overline{X}-\overline{Y}-\mu_0}{S\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}} > t_{\alpha,n+m-2}$ quyết định bác bỏ H_0

Nếu: $\dfrac{\overline{X}-\overline{Y}-\mu_0}{S\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}} \le t_{\alpha,n+m-2}$ quyết định chấp nhận H_0.

Ví dụ: Theo dõi 10 thửa ruộng trồng giống lúa A với năng suất X ta được:

$\overline{x} = 6{,}2$ tấn/ ha, $s_x^2 = 0{,}16$.

Theo dõi 8 thửa ruộng trồng giống lúa B với năng suất Y ta được:

$\overline{y} = 5{,}4$ tấn/ ha, $s_y^2 = 0{,}20$.

Biết chi phí canh tác giống lúa A tốn kém hơn giống lúa B qui ra thóc là 0,4 tấn/ ha. Với mức ý nghĩa $\alpha = 0{,}05$, xây dựng giả thuyết và đối thuyết thích hợp để khẳng định giống lúa A có hiệu quả kinh tế hơn giống lúa B không. Biết rằng năng suất của hai giống lúa này đều có phân phối chuẩn với phương sai bằng nhau và giá thành của hai loại lúa này là như nhau.

Yêu cầu trên tương ứng với việc kiểm định cặp giả thuyết đối thuyết sau:

H_0 : $\mu_X = \mu_Y + 0{,}4$

H_1: $\mu_X > \mu_Y + 0{,}4$

Ta có $s^2 = \dfrac{(n-1)s_X^2 + (m-1)s_Y^2}{n+m-2} = \dfrac{9.0{,}16 + 7.0{,}20}{16} = 0{,}18;\ s = 0{,}42$

$$Z_T = \frac{\overline{x}-\overline{y}-\mu_0}{s\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}} = \frac{6{,}2-5{,}4-0{,}4}{0{,}42\sqrt{\dfrac{1}{10}+\dfrac{1}{8}}} = 2{,}01 \ ; \ t_{0{,}05{,}16} = 1{,}75$$

Do $Z_T = 2{,}01 > 1{,}75 = t_{0{,}05{,}16}$ quyết định bác bỏ H_0. Điều này có nghĩa là giống lúa A có hiệu quả kinh tế hơn giống lúa B.

10.Phương pháp so sánh cặp đôi

Xét n cặp mẫu ngẫu nhiên $(X_1,Y_1),(X_2,Y_2),....,(X_n,Y_n)$

X_i có cùng phân phối với X có phân phối chuẩn $N(\mu_X,\sigma_X^2)$. Y_i có cùng phân phối với Y có phân phối chuẩn $N(\mu_Y,\sigma_Y^2)$, X độc lập với Y.

Đặt $D = X - Y$, $D_i = X_i - Y_i$, $\mu_V = \mu_X - \mu_Y$, $\sigma_V^2 = \sigma_X^2 + \sigma_Y^2$

Khi đó $D \sim N(\mu_V,\sigma_V^2)$; $\overline{D} = \dfrac{1}{n}\sum_{i=1}^{n} D_i = \overline{X} - \overline{Y}$; $S_D^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(D_i - \overline{D})$

Cặp giả thuyết đối thuyết:

H_0 : $\mu_X = \mu_Y + \mu_0$

H_1 : $\mu_X \ne \mu_Y + \mu_0$

tương đương với cặp giả thuyết đối thuyết:

H_0': $\mu_V = \mu_0$

H_1': $\mu_V \ne \mu_0$

Sử dụng quy tắc 4 ta có:

Nếu $Z_T = \dfrac{\left|\overline{D}-\mu_0\right|}{S_D}\sqrt{n} > t_{\frac{\alpha}{2},n-1}$ ta quyết định bác bỏ H_0.

Nếu $Z_T \le t_{\frac{\alpha}{2},n-1}$ ta quyết định chấp nhận H_0.

Cặp giả thuyết đối thuyết:

$$H_0: \quad \mu_X = \mu_Y + \mu_0$$
$$H_1: \quad \mu_X > \mu_Y + \mu_0$$

tương đương với cặp giả thuyết đối thuyết:

$$H_0': \quad \mu_V = \mu_0$$
$$H_1': \quad \mu_V > \mu_0$$

Sử dụng *quy tắc 6* ta có:

Nếu $Z_T = \dfrac{\overline{D}-\mu_0}{S_V}\sqrt{n} > t_{\alpha,n-1}$ ta quyết định bác bỏ H_0.

Nếu $Z_T \le t_{\alpha,n-1}$ ta quyết định chấp nhận H_0.

Ví dụ: Tại một câu lạc bộ thẩm mỹ người ta quảng cáo rằng sau một khóa tập luyện giảm béo người tham gia tập luyện có thể giảm trọng lượng hơn 5 kg. Một mẫu ngẫu nhiên gồm 10 người được đo trọng lượng trước và sau khi tập luyện cho bởi bảng sau:

X	85	90	96	93	86	89	82	84	100	102
Y	80	86	90	86	81	81	78	81	91	93

Hãy lập cặp giả thuyết đối thuyết thích hợp để kiểm tra tính đúng đắn của việc quảng cáo nói trên. Biết rằng $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$.

Xét cặp giả thuyết đối thuyết:

$$H_0: \quad \mu_X = \mu_Y + 5$$
$$H_1: \quad \mu_X > \mu_Y + 5$$

Đặt D = X – Y , $D_i = X_i - Y_i$, D_i nhận các giá trị sau:

V_i: 5 4 6 7 5 8 4 3 9 9

Ta có : $\overline{d} = 6,\ s_v^2 = \dfrac{40}{9} \Rightarrow s_d = 2{,}11$, $\quad t_{0,05,9} = 1{,}87$

$$Z_T = \frac{\overline{d}-5}{s_v}\sqrt{n} = \frac{6-5}{2{,}11}\sqrt{10} = 1{,}50$$

$Z_T = 1{,}50 < 1{,}87$ quyết định chấp nhận giả thuyết H_0. Điều này có nghĩa với mức ý nghĩa $\alpha = 0{,}05$ trọng lượng của những người tập luyện giảm béo chỉ ở mức giảm đi nhỏ hơn hoặc bằng 5 kg.

Phương pháp so sánh cặp đôi thường được sử dụng trong các thí nghiệm khoa học. Để xem một phương pháp canh tác, một chế độ cho gia súc ăn, một loại thuốc mới … có tốt hơn loại cũ hay không, người ta sẽ bố trí n cặp thí nghiệm, một thực hiện theo phương pháp cũ một thực hiện theo phương pháp mới, sau thời gian thí nghiệm thu được n cặp dữ liệu từ đó đưa ra kế luận.

11.Phương pháp loại bỏ các sai số thô

Khi thu thập và kiểm tra các số liệu mẫu, do nhiều nguyên nhân chủ quan và khách quan, việc gặp phải các sai số là điều không thể tránh khỏi. Các loại sai số thường gặp trong việc điều tra và thu thập các số liệu là :

11.1. Sai số thô: Là sai số xuất hiện do vi phạm các nguyên tắc cơ bản của việc đo đạc hoặc do những sơ suất mà người thu thập số liệu gây ra một cách cố ý hoặc vô ý.

11.2. Sai số hệ thống: Là những sai số do các dụng cụ đo gây ra hoặc do kĩ thuật viên không nắm được qui tắc vận hành dụng cụ đo. Sai số loại này dễ phát hiện để loại bỏ.

11.3. Sai số ngẫu nhiên: Là sai số chịu tác động của nhiều nguyên nhân, các sai số này thường nhỏ và không chịu sự tác động của người thu thập số liệu.

Trước khi tiến hành phân tích và xử lý số liệu, việc loại bỏ các số liệu dị thường (các sai số thô) ra khỏi tập các số liệu cần xử lý là điều cần chú ý, có như vậy các thông tin thu được sau xử lý mới đảm bảo tính chính xác với độ tin cậy cao.

11.4. Phương pháp loại bỏ sai số thô:

Khi tiến hành loại bỏ sai số thô (số liệu lạ) ta cần chú ý:

- Trước tiên cần kiểm tra xem có sơ suất hoặc có vi phạm các nguyên tắc cơ bản khi thu thập số liệu không?
- Thử loại bỏ x_0 là số liệu bị nghi ngờ rồi tiến hành xử lý số liệu xem kết luận có khác so với khi giữ lại x_0 hay không? Nếu không có sai khác đáng kể thì nên giữ lại số liệu x_0.
- Nên tham khảo các tài liệu chuyên môn liên quan có thể giải thích cho việc xuất hiện số liệu lạ này sau đó mới quyết định nên giữ hay nên bỏ.

Giả sử ta có dãy số liệu: $x_0, x_1, ..., x_n$ ở đó x_0 bị nghi ngờ là số dị thường (giá trị nhỏ nhất hoặc lớn nhất) trong dãy số trên. Khi đó ta xét đại lượng: $Z_T = \dfrac{|x_0 - \overline{x}|}{s}$. Nếu

$Z_T > t_{\frac{\alpha}{2}, n-1}$ ta quyết định loại bỏ giá trị x_0 ra khỏi dãy các số liệu trên.

Nếu $Z_T \le t_{\frac{\alpha}{2}, n-1}$ ta kết luận dãy số liệu trên không có số dị thường. Trong thực tế tùy yêu cầu chính xác của việc xử lý số liệu người ta thường lấy α ở các mức 0,05 hoặc 0,01.

Việc đưa ra tiêu chuẩn loại bỏ sai số thô nói trên dựa trên giả thiết các số liệu mẫu $x_0, x_1, ..., x_n$ lấy từ tổng thể có phân phối chuẩn $N(\mu, \sigma^2)$.

Ví dụ: Người ta đo 10 trục thép do một dây chuyền cơ khí sản xuất tính được $\overline{x} = 1{,}98$, $s = 0{,}04$ trong đó có trục có đường kính lớn nhất là 2,03. Với mức $\alpha = 0{,}05$ hỏi trục có đường kính nêu trên có phải là trục dị thường không?

Ta có: $Z_T = \dfrac{|x_0 - \overline{x}|}{s}\sqrt{n} = \dfrac{|1{,}98 - 2{,}03|}{0{,}04}\sqrt{10} \approx 3{,}64$

$t_{0,025,9} = 2{,}26,\ Z_T = 3{,}64 > t_{0,025,9} = 2{,}26$ trục có đường kính 2,03 nên loại khỏi mẫu.

Chú ý: Để kiểm tra giá trị nhỏ nhất và giá trị lớn nhất của mẫu có phải là sai số thô hay không trong trường hợp kích thước mẫu không lớn ta có thể thực hiện theo qui tắc sau:

*Tính: $Z_{TM} = \dfrac{x_{max} - \overline{x}}{\hat{s}}$ hoặc $Z_{Tm} = \dfrac{\overline{x} - x_{min}}{\hat{s}}$

* Từ mức α ta tìm C_α từ bảng 7

* So sánh Z_{TM} và Z_{Tm} với C_α. Nếu $Z_{TM} > C_\alpha$ ta loại x_{max}. Nếu $Z_{Tm} > C_\alpha$ ta loại bỏ x_{min}

12.Kiểm định sự bằng nhau của hai xác suất

Xác suất xuất hiện sự kiện A trong một dãy n phép thử độc lập $P(A) = p_1$
Xác suất xuất hiện sự kiện B trong một dãy n phép thử độc lập $P(B) = p_2$
Ta xây dựng qui tắc kiểm định giả thuyết

$$H_0: \quad p_1 = p_2$$

12.1 Trường hợp đối thuyết hai phía $H_1: p_1 \neq p_2$

Giả sử sau dãy n phép thử độc lập có n_A lần sự kiện A xuất hiện
Trong m phép thử khác có m_B lần sự kiện B xuất hiện

Đặt $f_1 = \dfrac{n_A}{n}, f_2 = \dfrac{m_B}{m}, f = \dfrac{n_A + m_B}{n+m}$

Thống kê $Z = \dfrac{f_1 - f_2 - (p_1 - p_2)}{\sqrt{f(1-f)(\frac{1}{n}+\frac{1}{m})}}$ có phân phối xấp xỉ chuẩn tắc

Nếu H_0 đúng thì $Z = \dfrac{f_1 - f_2}{\sqrt{f(1-f)(\frac{1}{n}+\frac{1}{m})}}$ có phân phối xấp xỉ chuẩn tắc nên

$$P\left[\frac{|f_1 - f_2|}{\sqrt{f(1-f)(\frac{1}{n}+\frac{1}{m})}} > U_{\frac{\alpha}{2}}\right] = \alpha.$$

Từ đây ta có qui tắc kiểm định đối với cặp giả thuyết đối thuyết:

$$H_0: \quad p_1 = p_2$$
$$H_1: \quad p_1 \neq p_2$$ ở mức ý nghĩa α là

Qui tắc 25: Nếu $\dfrac{|f_1 - f_2|}{\sqrt{f(1-f)(\frac{1}{n}+\frac{1}{m})}} > U_{\frac{\alpha}{2}}$ bác bỏ H_0

Nếu $\dfrac{|f_1 - f_2|}{\sqrt{f(1-f)(\frac{1}{n}+\frac{1}{m})}} \le U_{\frac{\alpha}{2}}$ chấp nhận H_0

Ví dụ: Điều tra một loại bệnh ở hai trại gà ta có kết quả sau:
Trại thứ nhất: Kiểm tra 500 con có 60 con mắc bệnh
Trại thứ hai: Kiểm tra 400 con có 50 con mắc bệnh
Gọi p_1, p_2 là xác suất để mỗi con gà ở trại thứ nhất và trại thứ hai mắc bệnh. Hãy kiểm định giả thuyết , đối thuyết

$$H_0: \quad p_1 = p_2$$
$$H_1: \quad p_1 \neq p_2$$ ở mức ý nghĩa $\alpha = 0{,}05$

Ta có: $f_1 = 0{,}12$, $f_2 = 0{,}125$, $f = 0{,}122$

$$Z_T = \frac{|f_1 - f_2|}{\sqrt{f(1-f)(\frac{1}{n}+\frac{1}{m})}} = \frac{0{,}005}{\sqrt{0{,}122.0{,}878(\frac{1}{500}+\frac{1}{400})}} \approx 0{,}23 \text{ ; } U_{0,025} = 1{,}96$$

$Z_T = 0{,}23 < U_{0,025} = 1{,}96$; giả thuyết được chấp nhận.

12.2 Trường hợp đối thuyết một phía: H_1: $p_1 > p_2$

Tương tự như trường hợp trên ta có qui tắc kiểm định là

Qui tắc 26: Nếu $\frac{f_1 - f_2}{\sqrt{f(1-f)(\frac{1}{n}+\frac{1}{m})}} > U_\alpha$ bác bỏ H_0

Nếu $\frac{f_1 - f_2}{\sqrt{f(1-f)(\frac{1}{n}+\frac{1}{m})}} \le U_\alpha$ chấp nhận H_0

13.Kiểm định sự bằng nhau của hai phương sai

Giả sử X~ N(μ_X ;σ_X^2)), (X_1, X_2, ..., X_n) là một mẫu ngẫu nhiên tương ứng.

Y ~ N(μ_Y ;σ_Y^2)), (Y_1, Y_2, ..., Y_n) là một mẫu ngẫu nhiên tương ứng.

Ta xây dựng qui tắc kiểm định giả thuyết H_0: $\sigma_X^2 = \sigma_Y^2$ ở mức ý nghĩa α

13.1 Trường hợp đối thuyết một phía: H_1: $\sigma_X^2 > \sigma_Y^2$

Thống kê $Z = \frac{S_X^2}{\sigma_X^2}.\frac{\sigma_Y^2}{S_Y^2}$ có phân phối $F_{n-1,\, m-1}$. Nếu H_0 đúng thì $Z = \frac{S_X^2}{S_Y^2}$ có phân phối $F_{n-1,\, m-1}$. $\Rightarrow P(\frac{S_X^2}{S_Y^2} > F_{\alpha, n-1, m-1}) = \alpha$

Từ đây ta có qui tắc kiểm định cặp giả thuyết đối thuyết

H_0: $\sigma_X^2 = \sigma_Y^2$

H_1: $\sigma_X^2 > \sigma_Y^2$ ở mức ý nghĩa α là

Qui tắc 27: Nếu $\frac{S_X^2}{S_Y^2} > F_{\alpha, n-1, m-1}$ ta bác bỏ H_0

Nếu $\frac{S_X^2}{S_Y^2} \le F_{\alpha, n-1, m-1}$ ta chấp nhận H_0

13.2 Trường hợp đối thuyết 2 phía H_1: $\sigma_X^2 \ne \sigma_Y^2$

Với cặp giả thuyết đối thuyết

. H_0: $\sigma_X^2 = \sigma_Y^2$

H_1: $\sigma_X^2 \ne \sigma_Y^2$

ở mức ý nghĩa α ta có qui tắc kiểm định sau:

Qui tắc 28: Nếu: $\dfrac{S_X^2}{S_Y^2} \notin \left[F_{1-\frac{\alpha}{2},n-1,\ m-1} ; F_{\frac{\alpha}{2},n-1,\ m-1}\right]$ ta bác bỏ H_0

Nếu: $\dfrac{S_X^2}{S_Y^2} \in \left[F_{1-\frac{\alpha}{2},n-1,\ m-1} ; F_{\frac{\alpha}{2},n-1,\ m-1}\right]$ ta chấp nhận H_0

Ví dụ: Một mẫu gồm 17 phần tử lấy từ tổng thể có phân phối chuẩn N($\mu_X ;\sigma_X^2$) ta được: $s_X^2 = 123{,}5$.

Một mẫu ngẫu nhiên khác gồm 15 phần tử cũng lấy từ tổng thể có phân phối chuẩn N($\mu_Y ;\sigma_Y^2$) ta được s_Y^2 = 60,4. Với mức ý nghĩa α = 0,05 hãy kiểm định cặp giả thuyết đối thuyết

H_0: $\sigma_X^2 = \sigma_Y^2$

H_1: $\sigma_X^2 > \sigma_Y^2$

Ta có $Z_T = \dfrac{s_X^2}{s_Y^2} = \dfrac{123{,}5}{60{,}4} = 2{,}04$, $F_{0,05\ ,16,14} = 2{,}48$, ta chấp nhận giả thuyết :H_0: $\sigma_X^2 = \sigma_Y^2$

14. Kiểm định sự bằng nhau của nhiều kì vọng

Việc so sánh sự bằng nhau của nhiều kì vọng là một yêu cầu khá phổ biến trong nông học, sinh học cũng như trong lâm học. Chẳng hạn như so sánh như so sánh năng suất của k giống lúa khác nhau hay so sánh chiều cao trung bình của k chủng người khác nhau...Những vấn đề vừa nêu sẽ được đề cập kĩ hơn trong giáo trình thống kê nâng cao. Trong mục này ta xây dựng phương pháp kiểm định giả thuyết đối thuyết

H_0: $\mu_1 = ... = \mu_i = ... = \mu_k$

H_1: $\exists\, i \neq j$ để $\mu_i \neq \mu_j$

Giả sử :

($X_{11}, X_{12}, ..., X_{1n_1}$) là mẫu ngẫu nhiên lấy từ tổng thể có phân phối chuẩn N($\mu_1 ;\sigma^2$)

($X_{i1}, X_{i2}, ..., X_{in_i}$) là mẫu ngẫu nhiên lấy từ tổng thể có phân phối chuẩn N($\mu_i ;\sigma^2$)

($X_{k1}, X_{k2}, ..., X_{kn_k}$) là mẫu ngẫu nhiên lấy từ tổng thể có phân phối chuẩn N($\mu_k ;\sigma^2$)

$$\overline{X}_1 = \frac{1}{n_1}\sum_{j=1}^{n_1} X_{1j}\ ,...,\quad \overline{X}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}\ , ...,\quad \overline{X}_k = \frac{1}{n_k}\sum_{j=1}^{n_k} X_{kj}$$

$\overline{X}_i$ là trung bình của mẫu ngẫu nhiên thứ i

$SS_i = \sum_{j=1}^{n_i}(X_{ij} - \overline{X}_i)^2$ tổng bình phương độ lệch của nhóm thứ i

$SSW = \sum_{i=1}^{k} SS_i$ là tổng bình phương độ lệch bên trong các nhóm

$\overline{X} = \dfrac{1}{n}\sum_{i=1}^{k} n_i \overline{X}_i$ là trung bình chung, $n = \sum_{i=1}^{k} n_i$ là kích thước mẫu

$SSB = \sum n_i (\overline{X}_i - \overline{X})^2$ là tổng bình phương độ lệch giữa các nhóm

$SST = \sum_{i,j=1}^{k,n_i}(X_{ij} - \overline{X})^2$ là tổng bình phương độ lệch toàn phần

Ta có SST = SSW + SSG

$MSW = \dfrac{SSW}{n-k}$ gọi là bình phương trung bình bên trong các nhóm. Đây là một ước lượng không chệch của phương sai σ^2.

$MSB = \dfrac{SSB}{k-1}$ gọi là bình phương trung bình giữa các nhóm. Đây cũng là một ước lượng không chệch của phương sai σ^2.

Khi giả thuyết H_0 đúng người ta thường chứng minh được rằng biến $Z = \dfrac{MSB}{MSW}$ có phân phối $F_{k-1,n-k}$

Từ khẳng định trên ta có qui tắc kiểm định cặp giả thuyết đối thuyết

H_0: $\mu_1 = ... = \mu_i = ... = \mu_k$

H_1: $\exists i \neq j$ để $\mu_i \neq \mu_j$

ở mức ý nghĩa α là

Qui tắc 29: Nếu $\dfrac{MSG}{MSW} > F_{k-1;n-k;\alpha}$ ta bác bỏ H_0

Nếu $\dfrac{MSB}{MSW} \leq F_{\alpha,k-1;n-k}$ ta chấp nhận H_0

Với mẫu cụ thể

X_1 x_{11} x_{12} x_{1n_1}	X_2 x_{21} x_{22} x_{2n_2}		X_i x_{i1} x_{i2} x_{in_i}		X_k x_{k1} x_{k2} x_{kn_k}

Ta tiến hành thực hiện bài toán kiểm định cặp giả thuyết , đối thuyết

H_0: $\mu_1 = ... = \mu_i = ... = \mu_k$

H_1: $\exists i \neq j$ để $\mu_i \neq \mu_j$

ở mức ý nghĩa α theo các bước sau:

Bước 1: Tính các số liệu ứng với mẫu đã cho:

$\overline{x}_i = \dfrac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}$, $\overline{x} = \dfrac{1}{n}\sum_{i=1}^{k} n_i\overline{x}_i$, $ss_i = \sum_{j=1}^{n_i}(x_{ij} - \overline{x}_i)^2$, $ssw = \sum_{i=1}^{k} ss_i$, $ssb = \sum_{i=1}^{k} n_i(\overline{x}_i - \overline{x})^2$

$msb = \dfrac{ssg}{k-1}$, $msw = \dfrac{ssw}{n-k}$, $Z_T = \dfrac{ssb}{ssw}$

Bước 2: Tìm $F_{\alpha,k-1,\,n-k}$

Bước 3: Nếu $Z_T > \quad F_{\alpha,k-1,\,n-k}$ quyết định bác bỏ H_0

Nếu $Z_T \le F_{\alpha,k-1,n-k}$ quyết định chấp nhận H_0

Ví dụ: Để so sánh năng suất của 3 giống lúa A, B, C người ta thực hiện thí nghiệm trên 20 thửa ruộng, sau khi thu hoạch ta có kết quả sau:

Giống lúa A trồng trên 7 thửa ruộng năng suất X_1 của từng thửa ruộng là:

54 57 55 58 61 54 52.

Giống lúa B trồng trên 6 thửa ruộng năng suất X_2 của từng thửa ruộng là:

50 55 58 54 61 52

Giống lúa C trồng trên 7 thửa ruộng năng suất X_3 của từng thửa ruộng là

58 60 54 56 61 60 57

ở mức ý nghĩa α = 0,05, hãy kiểm định giả thuyết

H_0: Năng suất của 3 giống lúa trên là như nhau

Biết rằng X_1, X_2, X_3 là các biến chuẩn có cùng phương sai

Ta có $\overline{x}_1 = \frac{1}{7}(\,54 + 57 + 55 + 58 + 61 + 54 + 52) = 56.$

$$\overline{x}_2 = \frac{1}{6}(\,50 + 55 + 58 + 54 + 61 + 52) = 55.$$

$$\overline{x}_3 = \frac{1}{7}(\,58 + 60 + 54 + 56 + 61 + 60 + 57) = 58.$$

$$\overline{x} = \frac{1}{20}(\,7.56 + 6.55 + 7.58\,) = 56{,}4.$$

$ss_1 = (54 - 56)^2+(57 - 56)^2 +(55 - 56)^2 +(58 - 56)^2 +(61 - 56)^2 +(54 - 56)^2 +(52 - 56)^2 = 55$

$ss_2 = (50 - 55)^2 + (55 - 55)^2 + (58 - 55)^2 + (54 - 55)^2 + (61 - 55)^2 + (52 -55)^2 = 80$

$ss_3 = (58 - 58)^2+(60 - 58)^2+(54 - 58)^2 +(56 - 58)^2 +(61 - 58)^2 +(60 - 58)^2 +(57 - 58)^2 = 38$

$ssw = ss_1 + ss_2 + ss_3 = 173$

$ssb = 7.\,(56 - 56{,}4)^2 + 6.\,(55 - 56{,}4)^2 +7.\,(58 - 56{,}4)^2 = 30{,}8$

$$msb = \frac{ssg}{k-1} = \frac{30{,}8}{2} = 15{,}40\,,\ \ msw = \frac{ssw}{n-k} = \frac{173}{17} = 10{,}18$$

$$Z_T = \frac{msb}{msw} = \frac{15{,}40}{10{,}18} = 1{,}51 \quad F_{k-1,n-k,0,05} = F_{2,17,0,05} = 3{,}59$$

$$Z_T = 1{,}51 < 3{,}59 = F_{0,05\,,\,2\,,17}$$

Giả thuyết được chấp nhận, không có sự khác nhau về năng suất của ba giống lúa trên.

III . Kiểm định giả thuyết phi tham số

1. Kiểm định một phân phối xác suất

Theo dõi một dãy n phép thử độc lập ta có kết quả sau:

Sự kiện	A_1	A_2	...	A_i	...	A_k
Tần số	n_1	n_2	...	n_i	...	n_k

Ta tiến hành kiểm định cặp giả thuyết đối thuyết ở mức ý nghĩa α

H$_0$: $P(A_1) = p_1, \ \ldots \ , P(A_i) = p_i, \ \ldots \ , P(A_k) = p_k$

H$_1$: $\exists j$ để $P(A_j) \neq p_j$

Nhận thấy rằng: H_0 đúng thì $n_i \approx np_i$, để kiểm định cặp giả thuyết đối thuyết trên cần đưa ra một thống kê thích hợp để khi thống kê này vượt quá ngưỡng cho phép nào đó thì ta nói có sự khác biệt giữa n_i và np_i. Khi đó ta sẽ quyết định bác bỏ giả thuyết còn nếu ngược lại ta chấp nhận giả thuyết hay nói đúng hơn là mẫu đã cho phù hợp với giả thuyết.

1.1.Trường hợp các p_i đã biết

Người ta chứng minh được rằng nếu H_0 đúng thì thống kê $Z = \sum_{i=1}^{k}\frac{(n_i - np_i)^2}{np_i}$ có phân phối giới hạn χ^2_{k-1}. Khi thống kê này vượt qua ngưỡng $\chi^2_{k-1,\alpha}$ ta quyết định bác bỏ giả thuyết.

Qui tắc kiểm định cặp giả thuyết đối thuyết

H$_0$: $P(A_1) = p_1, \ \ldots \ , P(A_i) = p_i, \ \ldots \ , P(A_k) = p_k$

H$_1$: $\exists j$ để $P(A_j) \neq p_j$

ở mức ý nghĩa α là

Qui tắc 1: Nếu $\sum_{i=1}^{k}\frac{(n_i - np_i)^2}{np_i} > \chi^2_{\alpha,k-1}$ ta bác bỏ H_0

Nếu $\sum_{i=1}^{k}\frac{(n_i - np_i)^2}{np_i} \leq \chi^2_{\alpha,k-1}$ ta chấp nhận H_0

Ví dụ 1: Hai cá thể ở thế hệ F_1 (cùng mang kiểu gen Aa) đem lai với nhau. Các cá thể ở thế hệ F_2 có một trong 3 kiểu gen AA, Aa, aa. Điều tra 200 cá thể ở thế hệ F_2 có:

Kiểu gen	AA	Aa	aa
Số cá thể	40	105	55

ở mức ý nghĩa $\alpha = 0{,}05$ kiểm định cặp giả thuyết , đối thuyết

H$_0$: Kiểu gen ở thế hệ F_2 tuân theo luật Mendel

H$_1$: Kiểu gen ở thế hệ F_2 không tuân theo luật Mendel

Giả thuyết H_0 tương ứng với $P(AA) = \frac{1}{4}$, $P(Aa) = \frac{2}{4}$, $P(aa) = \frac{1}{4}$

Sử dụng qui tắc 1 ta thực hiện bài toán kiểm định theo các bước sau:

Bước 1: Tính $Z_T = \sum_{i=1}^{3}\frac{(n_i - np_i)^2}{np_i} = \frac{(40-50)^2}{50} + \frac{(105-100)^2}{100} + + \frac{(55-50)^2}{50} = 2{,}75$

Bước 2: Tìm $\chi^2_{0,05,2} = 5{,}99$

Bước 3: $Z_T = 2{,}75 < 5{,}99 = \chi^2_{0,05,2}$. Theo qui tắc1 giả thuyết H_0 được chấp nhận điều này có nghĩa là mẫu đã cho phù hợp với qui luật Mendel

Chú ý: Xét $Z_T = \sum_{i=1}^{k}\frac{(n_i - np_i)^2}{np_i} = \sum_{i=1}^{k}\frac{n_i^2 - 2n_i np_i + n^2p_i^2}{np_i}$

$$= \sum_{i=1}^{k}\frac{n_i^2}{np_i} - 2\sum_{i=1}^{k} n_i + \sum_{i=1}^{k} np_i = \sum_{i=1}^{k}\frac{n_i^2}{np_i} - n \ \ (1)$$

Khi sử dụng qui tắc1 ta có thể tính Z_T theo công thức (1)

1.2 Khi các p_i phụ thuộc vào r tham số chưa biết (r < k-1)

Giả sử $p_i = p_i(\theta_1, \theta_2, ..., \theta_r)$ là các hàm phụ thuộc vào r tham số $\theta_1, \theta_2, ..., \theta_r$. Để thực hiện bài toán kiểm định trong trường hợp này trước hết ta cần tìm các ước lượng điểm của θ_i theo phương pháp hợp lý nhất. Nếu $\hat{\theta}_i$ là một ước lượng điểm của θ_i thì

$\hat{p}_i = p_i(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_r)$ là ước lượng điểm của p_i. Tương tự như qui tắc 1, thống kê $Z = \sum_{i=1}^{k} \frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i}$ có phân phối giới hạn χ^2_{k-r-1} nếu H_0 đúng. Từ đây ta có qui tắc kiểm định cặp giả thuyết, đối thuyết:

H_0: $P(A_1) = p_1, \dots, P(A_i) = p_i, \dots, P(A_k) = p_k$

H_1: $\exists j$ để $P(A_j) \neq p_j$

ở mức ý nghĩa α là

Qui tắc 2: Nếu $\sum_{i=1}^{k} \frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i} > \chi^2_{\alpha, k-r-1}$ ta bác bỏ H_0

Nếu $\sum_{i=1}^{k} \frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i} \le \chi^2_{\alpha, k-r-1}$ ta chấp nhận H_0

Ví dụ 2: Để xem có sự lây lan của " bệnh nấm mầm" từ cây này sang cây khác ở các cây cọ dầu hay không người ta trồng 500 cặp cây cọ dầu vào 500 hốc tại một vườn ươm cây. Sau một thời gian kiểm tra ta thu được kết quả sau:

Cả 2 cây bị bệnh	1 cây bị bệnh	0 cây bị bệnh
73	185	242

Với mức ý nghĩa $\alpha = 0{,}05$ kiểm định cặp giả thuyết đối thuyết

H_0: Không có sự lây bệnh từ cây này sang cây khác

H_1: Có sự lây bệnh từ cây này sang cây khác

Gọi p là xác suất để mỗi cây cọ dầu bị " bệnh nấm mầm "

Nếu giả thuyết H_0 đúng thì:

Xác suất để cả hai cây bị mắc bệnh $p_2 = p^2$.

Xác suất để một trong hai cây bị mắc bệnh $p_1 = 2p(1 - p)$.

Xác suất để không cây nào mắc bệnh $p_2 = (1 - p)^2$

Giả thuyết H_0 tương ứng với giả thuyết $p_2 = p^2$, $p_1 = 2p(1 - p)$, $p_0 = (1 - p)^2$. Bài toán kiểm định thực hiện theo các bước:

Bước 1: Ước lượng xác suất p bởi tần suất $f = \frac{2.73 + 185}{1000} = 0{,}331$

$\hat{p}_0 = (1 - 0{,}331)^2 = 0{,}44754$; $\hat{p}_1 = 2.0{,}331\,(1 - 0{,}331) = 0{,}4429$; $\hat{p}_2 = 0{,}331^2 = 0{,}10956$

$$\frac{(73 - 500\hat{p}_0)^2}{500\hat{p}_0} + \frac{(185 - 500\hat{p}_1)^2}{500\hat{p}_1} + \frac{(242 - 500\hat{p}_2)^2}{500\hat{p}_2} = \frac{18{,}22^2}{54{,}78} + \frac{36{,}45^2}{22{,}15} + \frac{18{,}23^2}{223{,}77} = 12{,}55$$

Bước 2: Tìm $\chi^2_{0.05,1} = 3{,}84$

Bước 3: $Z_T = 12{,}55 > 3{,}84 = \chi^2_{0.05,1}$

Giả thuyết H_0 bị bác bỏ, có sự lây lan " bệnh nấm mầm" từ cây này sang cây khác.

Chú ý 1: Khi sử dụng hai qui tắc vừa nêu ta phải thực hiện yêu cầu np_i hoặc $n\hat{p}_i$ ít nhất phải bằng 5

Chú ý 2: Để kiểm định giả thuyết

H_0: $X \sim F(x, \theta_1, \theta_2, ..., \theta_r)$

Khi mẫu đã cho được phân chia thành k lớp

Lớp	$<x_1$	$x_1 - x_2$		$x_{i-1} - x_i$		$\geq x_{k-1}$
Tần số	n_1	n_2		n_i		n_k

Trường hợp 1: Nếu các tham số $\theta_1, \theta_2, ..., \theta_r$ đã biết.

Ta có: $p_1 = F(x_1, \theta_1, \theta_2, ..., \theta_r)$ $p_2 = F(x_2, \theta_1, \theta_2, ..., \theta_r) - F(x_1, \theta_1, \theta_2, ..., \theta_r)$

$p_i = F(x_i, \theta_1, \theta_2, ..., \theta_r) - F(x_{i-1}, \theta_1, \theta_2, ..., \theta_r)$; $p_k = 1 - (p_1 + p_2 + ... + p_{k-1})$

Nếu $np_i \geq 5$ với mọi i = 1, k ta sử dụng qui tắc 1 để thực hiện bài toán kiểm định.

Nếu có những lớp mà $np_i < 5$ ta phải thực hiện ghép lớp này vào các lớp liền kề để các $np_i \geq 5$ sau đó sử dụng qui tắc 1.

Trường hợp 2: Nếu các tham số $\theta_1, \theta_2, ..., \theta_r$ chưa biết. Ta phải tìm các ước lượng điểm của $\theta_1, \theta_2, ..., \theta_r$ theo phương pháp hợp lý nhất. Giả sử $\hat{\theta}_i$ là ước lượng điểm của θ_i.

Ta có $\hat{p}_1 = F(x_1, \hat{\theta}_1, ...\hat{\theta}_r)$, $\hat{p}_2 = F(x_2, \hat{\theta}_1, ...\hat{\theta}_r) - F(x_1, \hat{\theta}_1, ...\hat{\theta}_r)$

$\hat{p}_i = F(x_i, \hat{\theta}_1, ...\hat{\theta}_r) - F(x_{i-1}, \hat{\theta}_1, ...\hat{\theta}_r)$, $\hat{p}_k = 1 - (\hat{p}_1 + \hat{p}_2 + ... + \hat{p}_{k-1})$

Nếu với mọi i =1, k mà $n\hat{p}_i \geq 5$ ta sử dụng qui tắc 2 để thực hiện bài toán kiểm định.

Nếu có những lớp mà $np_i < 5$ ta phải thực hiện ghép lớp này vào các lớp liền kề để các $np_i \geq 5$ sau đó mới sử dụng qui tắc 2.

Chú ý 3: Việc phân các số liệu mẫu vào các lớp nếu thoả mãn các yêu cầu sau thì lực lượng của phép kiểm định sẽ lớn (xác suất sai lầm loại 2 nhỏ)

*Xác suất để X nhận giá trị trong các lớp xấp xỉ nhau.

*Nếu kích thước mẫu nhỏ hơn 100 thì số lớp k phải lớn nhất thoả mãn

$np_i \geq 5$; $n\hat{p}_i \geq 5$

*Nếu kích mẫu lớn hơn 100 thì số lớp k phải xấp xỉ $3{,}2n^{2/5}$ và yêu cầu np_i hoặc $n\hat{p}_i \geq 5$ vẫn được bảo đảm.

Ví dụ: Sản lượng của loại đậu xám(tạ/ ha) trên 36 mảnh đất gần nhau được cho bởi bảng sau:

19,2	17,7	22,0	21,1	18,5	21,0	19,3	19,0	18,2
17,1	19,2	19,1	20,1	14,3	19,5	17,3	16,3	19,6
17,5	19,1	19,7	16,0	16,7	16,4	20,0	18,8	20,8
19,3	16,0	17,4	17,2	17,6	11,4	16,3	11,5	16,1

ở mức ý nghĩa $\alpha = 0{,}05$ có thể xem mẫu đã cho có phù hợp với giả thuyết :

H_0 : Sản lượng loại đậu xám có phân phối chuẩn $N(\mu; \sigma^2)$

Ta ước lượng μ bởi $\overline{x} = 17{,}95$, ước lượng σ^2 bởi $s^2 = 5{,}617$.

Lấy các $\hat{p}_i$ bằng nhau ta có số lớp k ở đây là 7. Vậy các $\hat{p}_i$ đều bằng $\frac{1}{7}$

x_1 là số thoả mãn $\phi(\frac{x_1-\overline{x}}{s})=\frac{1}{7} \Rightarrow x_1 = 15{,}42$

x_2 là số thoả mãn $\phi(\frac{x_2-\overline{x}}{s})=\frac{2}{7} \Rightarrow x_2 = 16{,}61$

x_3 là số thoả mãn $\phi(\frac{x_3-\overline{x}}{s})=\frac{3}{7} \Rightarrow x_3 = 17{,}53$

x_4 là số thoả mãn $\phi(\frac{x_4-\overline{x}}{s})=\frac{4}{7} \Rightarrow x_4 = 18{,}37$

x_5 là số thoả mãn $\phi(\frac{x_5-\overline{x}}{s})=\frac{5}{7} \Rightarrow x_5 = 19{,}29$

x_6 là số thoả mãn $\phi(\frac{x_6-\overline{x}}{s})=\frac{6}{7} \Rightarrow x_6 = 20{,}48$

Các số liệu mẫu được xếp vào các lớp theo bảng sau:

Lớp	$< x_1$	$[x_1;x_2)$	$[x_2;x_3)$	$[x_3;x_4)$	$[x_4;x_5)$	$[x_5;x_6)$	$\geq x_6$
n_i	3	6	6	3	7	7	4

Sử dụng *qui tắc 2* tiến hành kiểm định theo các bước sau:

Bước 1: Tính $Z_T = \sum_{i=1}^{k}\frac{(n_i-n\hat{p}_i)^2}{n\hat{p}_i} =$

$$\frac{(3-\frac{36}{7})^2}{\frac{36}{7}}+\frac{(6-\frac{36}{7})^2}{\frac{36}{7}}+\frac{(6-\frac{36}{7})^2}{\frac{36}{7}}+\frac{(3-\frac{36}{7})^2}{\frac{36}{7}}+\frac{(7-\frac{36}{7})^2}{\frac{36}{7}}+\frac{(7-\frac{36}{7})^2}{\frac{36}{7}}+\frac{(4-\frac{36}{7})^2}{\frac{36}{7}} = 3{,}666$$

Bước 2: Do phải ước lượng hai tham số μ và σ^2 nên số bậc tự do là 4, $\chi^2_{0,05,4} = 9{,}448$.

Bước 3: $Z_T = 3{,}666 < 9{,}488 = \chi^2_{0,05,4}$ mẫu đã cho phù hợp với giả thuyết.

Chú ý: $Z_T = \sum_{i=1}^{k}\frac{(n_i-n\hat{p}_i)^2}{n\hat{p}_i} = \sum_{i=1}^{k}\frac{n_i^2-2n_in\hat{p}_i+n^2\hat{p}_i^2}{n\hat{p}_i}$

$$= \sum_{i=1}^{k}\frac{n_i^2}{n\hat{p}_i} - 2\sum_{i-1}^{k}n_i + \sum_{i=1}^{k}n_i\hat{p}_i = \sum_{i=1}^{k}\frac{n_i^2}{n\hat{p}_i} - n \quad (2)$$

Khi sử dụng qui tắc 2 ta có thể tính Z_T theo công thức cho bởi (2)

2. Kiểm định tính độc lập của hai đặc tính định tính.

Xét một đám đông mỗi cá thể ta để ý tới hai đặc tính định tính A và B. Giả sử đặc tính A được chia thành k mức A_1, A_2,... A_k, đặc tính B được chia thành m mức B_1, B_2...B_m.Từ đám đông lấy ra một mẫu ngẫu nhiên có kích thước n ta có kết quả sau:

B / A	B_1	B_2		B_j		B_m
A_1	n_{11}	n_{12}		n_{1j}		n_{1m}
A_2	n_{21}	n_{22}		n_{2i}		n_{2m}
A_i	n_{i1}	n_{i2}		n_{ij}		n_{im}
A_k	n_{k1}	n_{k2}		n_{kj}		n_{km}

n_{ij} là số cá thể có đặc tính $A = A_i$ và đặc tính $B = B_j$ trong mẫu.

Từ mẫu trên xây dựng qui tăc kiểm định cặp giả thuyết, đối thuyết

H_0 : A độc lập với B.

H_1 : A không độc lập với B ở mức ý nghĩa α.

Ta có $\sum_{j=1}^{m} n_{ij} = n_{i.}$ là số cá thể có đặc tính $A = A_i$

$\sum_{i=1}^{k} n_{ij} = n_{.j}$ là số cá thể có đặc tính $B = B_j$

$$\sum_{i,j=1}^{k,m} n_{ij} = \sum_{i=1}^{k} n_{i.} = \sum_{j=1}^{m} n_{.j} = \mathrm{n}$$

Để đơn giản ta qui ước: $\sum_{i=1}^{k}\sum_{j=1}^{m} n_{ij} = \sum_{i,j=1}^{k,m} n_{ij}$)

$f_{ij} = \dfrac{n_{ij}}{n}$ là ước lượng của xác suất $P(A_iB_j)$

$f_{i.} = \dfrac{n_{i.}}{n}$ là ước lượng của xác suất $P(A_i)$

$f_{.j} = \dfrac{n_{.j}}{n}$ là ước lượng của xác suất $p(B_j)$

Nếu giả thuyết đúng thì A_i độc lập với B_j vì vậy có $P(A_iB_j) = P(A_i)P(B_j)$

$$\Rightarrow f_{ij} \approx f_{i\bullet}f_{\bullet j} \Rightarrow n_{ij} \approx \frac{n_{i\bullet}n_{\bullet j}}{n}$$

Ta đưa ra thống kê Z thích hợp để khi thống kê này vượt qua một giá trị xác định nào đó thì ta khẳng định có sự khác biệt giữa n_{ij} và $\dfrac{n_{i\bullet}n_{\bullet j}}{n}$, từ đó đưa ra quyết định bác bỏ giả thuyết. Người ta đã chứng minh được thống kê

$Z = \sum_{i,j=1}^{k,m} \dfrac{\left(n_{ij} - \dfrac{n_{i\bullet}n_{\bullet j}}{n}\right)^2}{\dfrac{n_{i\bullet}n_{\bullet j}}{n}}$ có phân phối giới hạn là $\chi^2_{(k-1)(m-1)}$ nếu giả thuyết đúng. Từ đây ta có qui tắc bác bỏ giả thuyết ở mức ý nghĩa α là:

Qui tắc 3: Nếu $\sum_{i,j=1}^{k,m} \frac{(n_{ij} - \frac{n_{i\bullet} n_{\bullet j}}{n})^2}{\frac{n_{i\bullet} n_{\bullet j}}{n}} > \chi^2_{\alpha,(k-1)(m-1)}$. ta quyết định bác bỏ H_0

Nếu $\sum_{i,j=1}^{k,m} \frac{(n_{ij} - \frac{n_{i\bullet} n_{\bullet j}}{n})^2}{\frac{n_{i\bullet} n_{\bullet j}}{n}} \le \chi^2_{\alpha,(k-1)(m-1)}$ ta quyết định chấp nhận H_0 hay mẫu đã cho phù hợp với giả thuyết A độc lập với B.

Chú ý: Khi sử dụng qui tắc 3 để kiểm định tính độc lập của hai đặc tính A, B cần đáp ứng yêu cầu $\frac{n_{i\bullet} n_{\bullet j}}{n} \ge 5$.

Ví dụ: Xét một đàn ốc sên rừng, đặc tính A là màu vỏ gồm màu vàng (A_1) và màu hồng (A_2). Đặc tính B là số vạch trên vỏ gồm : 0 vạch(B_0), 1 hoặc 2 vạch (B_1), 3 hoặc 4 vạch (B_2) và 5 vạch (B_3). Bắt ngẫu nhiên 169 con ốc sên rừng thuộc đàn ốc sên nói trên ta có bảng sau:

Số vạch / Màu vỏ	0 (B_0)	1-2 (B_1)	3-4 (B_2)	5 (B_3)
Vàng(A_1)	35	19	36	25
Hồng(A_2)	14	14	16	10

Với mức ý nghĩa $\alpha = 0,05$ hãy kiểm định giả thuyết

H_0: Màu vỏ độc lập về di truyền với số vạch trên vỏ

H_1 : Màu vỏ không độc lập về di truyền với số vạch trên vỏ

Ta có: $n_{1\bullet} = 115$, $n_{2\bullet} = 54$, $n_{\bullet 1} = 49$, $n_{\bullet 2} = 33$, $n_{\bullet 3} = 52$, $n_{\bullet 4} = 35$, $n = 169$

$$Z_T = \sum_{i,j=1}^{2,3} \frac{(n_{ij} - \frac{n_{i\bullet} n_{\bullet j}}{n})^2}{\frac{n_{i\bullet} n_{\bullet j}}{n}} = \frac{(35 - \frac{115.49}{169})^2}{\frac{115.49}{169}} + \frac{(19 - \frac{115.33}{169})^2}{\frac{115.33}{169}} + \frac{(36 - \frac{115.52}{169})^2}{\frac{115.52}{169}}$$

$$+ \frac{(25 - \frac{115.35}{169})^2}{\frac{115.35}{169}} + \frac{(14 - \frac{54.49}{169})^2}{\frac{54.49}{169}} + \frac{(14 - \frac{54.33}{169})^2}{\frac{54.33}{169}} + \frac{(16 - \frac{54.52}{169})^2}{\frac{54.52}{169}} + \frac{(10 - \frac{54.35}{169})^2}{\frac{54.35}{169}} = 2,13$$

$$\chi^2_{\alpha,(k-1)(m-1)} = \chi^2_{0,05,3} = 7,81$$

$Z_T = 2,13 < 7,81 = \chi^2_{0,05,3}$. Ta quyết định chấp nhận H_0 màu vỏ và số vạch trên vỏ độc lập với nhau về di truyền.

Chú ý 1: $Z_T = \sum_{i,j=1}^{k,m} \frac{(n_{ij} - \frac{n_{i.} n_{.j}}{n})^2}{n_{i.} n_{.j}} = \left[\sum_{i,j=1}^{k,m} \frac{n n_{ij}^2}{n_{i.} n_{.j}} - 2 \sum_{i,j=1}^{k,m} n_{ij} + \frac{1}{n} \sum_{i,j=1}^{k,m} n_{i.} n_{.j} \right]$

$$= n\sum_{i,j=1}^{k,m}\frac{n_{ij}^2}{n_{i.}n_{.j}} - 2n + \frac{1}{n}\sum_{i=1}^{k}n_{i.}\sum_{j=1}^{m}n_{.j} = n\left[\sum_{i,j=1}^{k,m}\frac{n_{ij}^2}{n_{i.}n_{.j}} - 1\right] \qquad (3)$$

Khi sử dụng qui tắc 3 có thể tính Z_T bằng công thức cho bởi (3)

Chú ý 2: Việc xây dựng qui tắc kiểm định tính thuần nhất của đám đông cũng được trình bày như tiêu chuẩn vừa nêu. Tiêu chuẩn đưa ra cũng giống như tiêu chuẩn vừa nêu.

3.Quy tắc dấu

Xét n cặp mẫu ngẫu nhiên : $(X_1,Y_1),(X_2,Y_2),....,(X_n,Y_n)$

X_i có cùng phân phối với X có hàm mật độ f(x)

Y_i có cùng phân phối với Y có hàm mật độ g(x)

Nếu X, Y là các biến chuẩn thì việc so sánh kì vọng của X và Y đã được trình bày trong phương pháp so sánh cặp đôi. Bây giờ ta đưa ra quy tắc kiểm định trong trường hợp tổng quát cặp giả thuyết đối thuyết.

H_0: X có cùng phân phối với Y

H_1: X và Y có phân phối khác nhau.

Đặt D = X - Y , $D_i = X_i - Y_i$.

Nếu H_0 đúng người ta có thể chứng minh rằng P(D > 0) = P(D < 0) = 0,5.

Gọi M là số các giá trị mà $D_i > 0$ ta thấy M có phân phối nhị thức B(n, $\frac{1}{2}$). Cặp giả thuyết đối thuyết nêu trên tương đương với cặp giả thuyết đối thuyết.

H_0': M có phân phối nhị thức B(n, $\frac{1}{2}$)

H_1': M không có phân phối nhị thức B(n, $\frac{1}{2}$)

Sử dụng định lý giới hạn: Biến Z = $\frac{M-0,5n}{0,5\sqrt{n}}$ có phân phối giới hạn chuẩn tắc ta có quy tắc kiểm định cặp giả thuyết H_0 và H_1 là :

Qui tắc 5: Nếu $Z_T = \frac{|M-0,5n|}{0,5\sqrt{n}} > U_{\frac{\alpha}{2}}$ bác bỏ H_0

Nếu $Z_T \le U_{\frac{\alpha}{2}}$ chấp nhận H_0

Trong thực hành khi gặp các cặp số liệu (x_i, y_i) mà $x_i = y_i$ ta loại bỏ cặp số liệu này ra khỏi mẫu.

Ví dụ: Chiều cao X của người bố và chiều cao Y của con trai tương ứng từ mẫu gồm 20 cặp bố con được cho ở bảng sau:

X	1,72	1,70	1,62	1,58	1,64	1,68	1,67	1,73	1,57	1,63
Y	1,74	1,68	1,65	1,55	1,61	1,70	1,67	1,74	1,59	1,60
X	1,74	1,76	1,58	1,67	1,55	1,68	1,71	1,58	1,75	1,65
Y	1,72	1,73	1,60	1,64	1,62	1,66	1,65	1,62	1,77	1,61

Với mức ý nghĩa α = 0,05, hãy kiểm định cặp giả thuyết đối thuyết

H_0: X có cùng phân phối xác suất với Y

Ta loại bỏ mẫu thứ bảy do chiều cao của cặp cha con này như nhau.
Đặt: $D = X - Y$, $d_i = x_i - y_i$
Số mẫu có d_i dương $m = 11$, kích thước mẫu $n = 19$.

. $$Z_T = \frac{|m - 0{,}5n|}{0{,}5\sqrt{n}} = 0{,}96 \;\; ; U_{0,025} = 1{,}96 \Rightarrow$$ quyết định chấp nhận H_0.

4.Quy tắc Wilcoxon

4.1 Thứ tự của dãy số

Cho dãy số : $x_1, x_2, \ldots\ldots, x_n$
Gọi $u_i = \text{rank}(x_i)$ là thứ hạng của số x_i khi xếp dãy số trên theo thứ tự tăng dần.
Nếu trong dãy số $x_1, x_2, \ldots\ldots, x_n$ có các giá trị bằng nhau được xếp từ thứ tự thứ k đến thứ k +m-1 thì thứ hạng của các số giống nhau này cùng bằng $k + \frac{m}{2}$.

Ví dụ: Cho dãy số: 1,4; 1,1; 1,4; 1,1; 1,5; 1,4; 1,6; 1,8; 1,7; 1,8.
Xếp dãy số trên theo thứ tự tăng dần ta có:
1,1; 1,1; 1,4; 1,4; 1,4; 1,5; 1,6; 1,7; 1,8; 1,8.
Khi đó:
rank(1,1) = 1,5, rank(1,4) = 4, rank(1,5) = 6, rank(1,6) =7, rank(1,7)=8, rank(1,8) = 9,5.

4.2 Quy tắc Wilcoxon

Dựa vào thứ tự của dãy số mẫu, Wilcoxon đưa ra quy tắc kiểm định cặp giả thuyết đối thuyết:

H_0: X có cùng phân phối xác suất với Y
H_1: X và Y có phân phối khác nhau

Wilcoxon giải quyết bài toán trên trong trường hợp mẫu gồm n cặp:
(X_1, Y_1) ; (X_2 , Y_2) ;……;(X_n , Y_n)
Mann và Whitney giải quyết bài toán trên trong trường hợp tổng quát với hai mẫu
$(X_1, X_2, \ldots, X_n)$ và $(Y_1, Y_2, \ldots, Y_m)$.
Gọi V_i là thứ tự của X_i trong dãy gồm n + m số:
$$X_1, X_2, \ldots, X_n, Y_1, Y_2, \ldots, Y_m$$
Đặt $V = \sum_{i=1}^{n} V_i$,nếu H_0 đúng có thể chứng minh rằng
$$E(V) = \frac{n(n+m+1)}{2}; \;\; D(V) = \frac{nm(n+m+1)}{12}$$
Khi đó thống kê $Z = \dfrac{V - \frac{n(n+m+1)}{2}}{\sqrt{\frac{nm(n+m+1)}{12}}}$ có phân phối xấp xỉ chuẩn tắc .
Từ đây ta có quy tắc kiểm định cặp giả thuyết đối thuyết

H_0: X có cùng phân phối xác suất với Y
H_1: X và Y có phân phối khác nhau

ở mức ý nghĩa α là :

Quy tắc 6: Nếu $Z_T = \dfrac{\left|V - \dfrac{n(n+m+1)}{2}\right|}{\sqrt{\dfrac{nm(n+m+1)}{12}}} > U_{\frac{\alpha}{2}}$ ta bác bỏ H_0.

Nếu $Z_T \le U_{\frac{\alpha}{2}}$ ta chấp nhận H_0.

Ví dụ: Theo dõi doanh thu X của 10 cửa hàng thóc giống tại Hà Tây và doanh thu Y của 12 cửa hàng thóc giống tại Thái Bình ta có kết quả sau:

X(triệu đồng/tháng): 32, 36, 28, 24, 30, 25, 32, 33, 26, 27

Y(triệu đồng/tháng): 31, 35, 27, 31, 26, 28, 34, 32, 30, 31, 26, 29

Với mức ý nghĩa $\alpha = 0{,}05$ hãy kiểm định cặp giả thiết đối thuyết:

H_0: X có cùng phân phối xác suất với Y

H_1: X và Y có phân phối khác nhau

Ta có tổng các thứ hạng của các x_i là v = 107,5

$$\frac{n(n+m+1)}{2} = 115 \ ; \ \sqrt{\frac{nm(n+m+1)}{12}} = 15{,}165$$

$$Z_T = \frac{\left|v - \dfrac{n(n+m+1)}{2}\right|}{\sqrt{\dfrac{nm(n+m+1)}{12}}} = 0{,}45 \ ; \ U_{0,025} = 1{,}96$$

$Z_T = 0{,}45 < U_{0,025} = 1{,}96$ ta quyết định chấp nhận H_0.

4.3 Quy tắc Kruskal-Wallis

Các dữ liệu thu được từ các cuộc điều tra trong sinh học, nông học, lâm học và y học thường được thu thập từ nhiều vùng khác nhau. Ta cần kiểm tra xem các dữ liệu này có cùng xuất phát từ một tập cơ bản (cùng một tổng thể) hay không? Giả sử mẫu được thu thập từ k vùng ($k \ge 3$) và giả sử rằng dãy các giá trị mẫu:

$x_{11}, x_{12}, \ldots, x_{1n_1}$ lấy từ vùng I, có đặc tính X_1

$x_{21}, x_{22}, \ldots, x_{2n_2}$ lấy từ vùng II, có đặc tính X_2

....

$x_{k1}, x_{k2}, \ldots, x_{kn_k}$ lấy từ vùng K, có đặc tính X_k

Kích thước mẫu $n = \sum_{j=1}^{k} n_j$.

Ta gọi n_{ij} là thứ tự của số liệu x_{ij} trong n số liệu trên, $1 \le i \le k, 1 \le j \le n_i$. Đặt $R_i = \sum_{j=1}^{n_i} n_{ij}$

Xét thống kê: $Z = \dfrac{12}{n(n+1)} \sum_{i=1}^{k} \dfrac{R_i^2}{n_i} - 3(n+1)$

Nếu $k \ge 3, n_i \ge 6$ thì Z có phân phối xấp xỉ phân phối khi bình phương với k-1 bậc tự do. Dựa vào quy luật phân phối xấp xỉ của biến Z với mức ý nghĩa α ta có quy tắc kiểm định cặp giả thuyết đối thuyết

H_0: Dãy các số liệu trên thu thập từ một tập cơ bản

Quy tắc 7: Nếu $Z_T = \dfrac{12}{n(n+1)}\sum_{i=1}^{k}\dfrac{R_i^2}{n_i} - 3(n+1) > \chi^2_{\alpha,k-1}$ bác bỏ H_0

Nếu $Z_T \le \chi^2_{\alpha,k-1}$ chấp nhận H_0.

Quy tắc trên được gọi là quy tắc Kruskal - Wallis.

Ví dụ: Nghiên cứu tác động của 3 loại thức ăn gia súc khác nhau đối với sự tăng trọng của một loài lợn người ta tiến hành thử nghiệm trên 20 con lợn.

Gọi: X_1 là mức tăng trọng trong một tháng ở mỗi con trong nhóm 6 con lợn dùng thức ăn loại A là:

17,5 13,5 9,0 12,5 11,0 16,5

X_2 là mức tăng trọng trong một tháng ở mỗi con lợn trong nhóm 7 con lợn dùng thức ăn loại B là:

16,0 14,5 11,5 8,5 12,0 15,0 10,5

X_3 là mức tăng trọng trong một tháng ở mỗi con lơn trong nhóm 7 con lợn dùng thức ăn loại C là:

17,0 9,5 14,0 13,0 10,0 15,5 8,0

Với mức ý nghĩa α = 0,05 hãy kiểm định cặp giả thuyết đối thuyết

H_0: Ba loại thức ăn có tác dụng như nhau với sự tăng trọng của lợn

H_1: Ba loại thức ăn có tác dụng khác nhau với sự tăng trọng của lợn

Giả thuyết H_0 tương đương với các số liệu mẫu trên lấy từ một đám đông thuần nhất.

Ta có: $k = 3$, $n_1 = 6$, $n_2 = n_3 = 7$, $n = 20$

. $R_1 = 65,\ R_2 = 71,\ R_3 = 69$

$$Z_T = \frac{12}{n(n+1)}\sum_{i=1}^{k}\frac{R_i^2}{n_i} - 3(n+1) = 4{,}77\ ;\ \chi^2_{0,05,2} = 5{,}99$$

. $Z_T = 4{,}77 < 5{,}99 \Rightarrow$ giả thuyết H_0 được chấp nhận, điều nay có thể hiểu là 3 loại thức ăn trên có tác dụng như nhau với việc tăng trọng của lợn.

Chú ý: Các qui tắc kiểm định phi tham số có ưu điểm là không cần biết trước kiểu dạng phân phối xác suất của các đặc trưng ở tổng thể, nhưng do lượng lượng thông tin thu được từ tổng thể không nhiều nên lực lượng của phép kiểm định của các qui tắc này không cao.

Bài tập chương VI

1. Biết độ chịu lực X của các mẫu bê tông có phân phối chuẩn N($\mu;\sigma^2$). Đo độ chịu lực của 210 mẫu bê tông ta có kết quả sau:

Độ chịu lực X_i(kg/cm^2)	195	205	215	225	235	245
Số mẫu bê tông n_i	13	18	46	74	34	15

Với mức ý nghĩa α = 0,05, hãy kiểm định giả thuyết, đối thuyết:

H_0 : $\mu = 230$

H_1: $\mu \neq 230$ hoặc H_1: $\mu < 230$

2.Trọng lượng của mỗi gói mì ăn liền X (g/gói) do một nhà máy sản xuất là biến chuẩn với phương sai bằng 2,25. Lấy ngẫu nhiên 20 gói mì do nhà máy trên sản xuất đem cân ta có trọng lượng trung bình

$\bar{x}$ = 78,2. Với mức ý nghĩa α = 0,05 hãy kiểm định cặp giả thuyết, đối thuyết

H_0: $\mu = 80$; H_1: $\mu \neq 80$

3. Năng suất X của một giống lúa trong vùng là một biến chuẩn. Điều tra năng suất lúa trên 36 mảnh ruộng ta có kết quả sau:

X_i(tấn/ha)	5,0	5,2	5,4	5,6	5,8	6,0
Số mảnh n_i	3	5	10	9	6	3

Với mức ý nghĩa α = 0,05 hãy kiểm định cặp giả thuyết , đối thuyết

a. H_0: $\mu = 5{,}5$; H_1: $\mu \neq 5{,}5$

b. H_0: $\sigma^2 = 0{,}8$; H_1: $\sigma^2 > 0{,}8$

4. Một mẫu ngẫu nhiên gồm 600 học sinh lớp 12 các vùng nông thôn khu vực phía Bắc thấy có 122 nói sẽ nộp đơn thi vào trương Đại Học Nông nghiệp I.

Với mức ý nghĩa α = 0,05 hãy kiểm định cặp giả thuyết, đối thuyết

H_0: Tỉ lệ học sinh thi vào ĐHNNI p = 0,20

H_1: Tỉ lệ học sinh thi vào ĐHNNI p > 0,20 .

5. Để so sánh năng suất của hai giống lúa A (năng suất X), giống lúa B (năng suất Y), người ta trồng từng cặp trên các loại đất khác nhau sau thu hoạch ta được kết quả sau:

Giống A(năng suất X tấn / ha)	6	7	6,5	5,5	4,3	6,6	5,8	4,9	5,3	6,5
Giống B(năng suất Y tấn / ha)	5	4	7,5	5,5	5,5	5,6	6,8	4,2	6,3	4,5

Biết X và Y là các biến chuẩn. Với mức ý nghĩa 0,05 có thể coi năng suất hai giống lúa trên là khác nhau không? Sử dụng phương pháp so sánh cặp đôi . Hãy xét trong trường hợp lấy mẫu độc lập.

6. Để xét ảnh hưởng của hai loại phân bón A, B đối với một giống lúa người ta dùng phân A bón cho lúa trên 5 thửa ruộng. Dùng phân B bón cho lúa trên 6 thửa ruộng. Sau thu hoạch ta có kết quả:

X(tạ/ha)Năng suất lúa sử dụng phân A	45	47	43	44	46	
Y(tạ/ha)Năng suất lúa sử dụng phân B	46	49	43	46	50	44

Với mức ý nghĩa 0,05 có thể coi ảnh hưởng của hai loại phân trên đối với năng suất lúa là như nhau được không? Thực hiện như bài 5.

7.Để so sánh trọng lượng của con rạ (sinh từ lần thứ hai trở đi) và trọng lượng con so (sinh lần đầu) qua thống kê ở một nhà hộ sinh ta được kết quả sau:

Trọng lượng(g)	1700-2000	2000-2300	2300-2600	2600-2900	2900-3200
Số con rạ n_i	9	13	18	42	18
Số con so m_i	5	10	22	40	45

Với mức ý nghĩa 0,05 có thể coi trọng lượng con so lớn hơn trọng lượng con rạ không?

8. Theo dõi doanh thu X , Y hàng tháng của 8 cửa hàng bán giống cây trồng tại Nam Định và 10 cửa hàng bán giống cây trồng tại Thái Bình ta được kết quả sau:

X(triệu đồng/tháng)	32	36	28	24	30	25	32	33		
Y(triệu đồng/tháng)	31	35	27	36	31	26	28	34	32	30

Với mức ý nghĩa 0,05 có thể coi doanh thu của các cửa hàng bán giống cây trồng ở hai địa phương trên là khác nhau không?

9. Một nông trường bò sữa nhập ba giống bò A, B, C. Người ta thống kê sản lượng sữa của chúng theo ba mức: ít, trung bình và nhiều sữa. Từ bảng số liệu về sự phân bố ba giống bò trên theo ba mức:

Giống bò	A	B	C
Ít sữa	92	53	75
Trung bình	37	15	19
Nhiều sữa	46	19	12

Với mức ý nghĩa 0,05 hãy nhận định xem sản lượng sữa của 3 giống bò có khác nhau không?

10. Để điều tra mức độ xem phim của nhân dân một tỉnh người ta chia mức độ xem phim thành ba cấp (nhiều , vừa, ít). Kết quả điều tra 300 hộ như sau:

Mức độ / Vùng	Nhiều	Vừa	ít
Thành phố	48	26	26
Ven nội	38	34	28
Huyện	16	10	74

Có thể coi mức độ xem phim ở ba vùng là như nhau được không? Mức ý nghĩa 0,05.

11. Khảo sát màu mắt và màu tóc của 6800 người Pháp ta được kết quả sau:

Màu tóc / Màu mắt	Vàng	Nâu	Đen	Hung
Xanh	1768	807	189	47
Đen	946	1387	746	53
Nâu	115	438	288	16

Với mức ý nghĩa 0,05 hãy kiểm định giả thuyết:

H_0: Màu tóc độc lập với màu mắt.
H_1: Màu tóc không độc lập với màu mắt.

12. Để nghiên cứu mối liên hệ giữa việc nghiện thuốc lá (đặc tính A) và huyết áp (đặc tính B) người ta tiến hành điều tra 200 người kết quả cho bởi:

A / B	A_0(không nghiện)	A_1(nghiện nhẹ)	A_2(nghiện nặng)
B_0(huyết áp bt)	50	25	28
B_1(huyết áp cao)	30	35	32

Với mức ý nghĩa 0,05 hãy kiểm định giả thuyết :
H_0: A độc lập với B
H_1: A không độc lập với B

13. Một loài hoa có 3 giống A, B, C. Mỗi giống hoa có thể cho hoa đỏ hoặc hoa trắng. Từ số liệu thống kê:

Màu\ Loài	A	B	C
Hoa đỏ	58	102	65
Hoa trắng	102	118	75

Với mức ý nghĩa 0,05. Hay kiểm định các giả thuyết:
a. Màu hoa và giống hoa độc lập với nhau
b. Trong giống hoa B tỉ lệ giữa hoa đỏ và hoa trắng là 1 : 1

14. Điều tra 100 gia đình có hai con ta được kết quả sau:

Số con trai / Số gia đình	0	1	2
n_i	20	56	24

Với mức $\alpha=0,05$ hãy kiểm định giả thuyết:
a. H_0: Số con trai trong mỗi gia đình tuân theo phân phối nhị thức B(2 ; 0,5)
b. H_0: Số con trai trong mỗi gia đình tuân theo phân phối nhị thức B(2 ; p)

15. Một loại cây có gen A chỉ lá quăn, gen a chỉ lá phẳng, gen B hạt trắng, gen b chỉ hạt đỏ. Khi lai hai cây thuần chủng lá quăn hạt đỏ và lá thẳng hạt trắng ta được thế hệ F_1. Cho hai cá thể ở thế hệ F_1 lai với nhau ở thế hệ F_2 ta có kết quả sau:
1160 cây lá quăn hạt đỏ ; 380 cây lá quăn hạt trắng
350 cây lá thẳng hạt đỏ ; 110 cây lá thẳng hạt trắng
Với các số liệu trên ở mức ý nghĩa 0,05 hãy kiểm định cặp giả thuyết đối thuyết :
H_0: Kết quả phù hợp với qui luật phân li tính trạng 9 : 3 : 3 : 1
H_1: Trái với H_0.

16. Xét mối liên quan giữa vợ chồng và thể trạng ta có bảng số liệu sau:

Vợ / Chồng	Gầy	Béo	Trung bình
Gầy	24	12	12
Béo	10	40	15
Trung bình	20	12	115

Với mức ý nghĩa: $\alpha = 0{,}05$ hãy kiểm định cặp giả thuyết đối thuyết:

H_0: Thể trạng và mối quan hệ vợ chồng độc lập với nhau.

H_1: Thể trạng và mối quan hệ vợ chồng có liên quan với nhau.

17. Một gói mì ăn liền đạt yêu cầu về trọng lượng nếu có trọng lượng 80 gam. Kiểm tra mẫu gồm 20 gói mì được $\overline{x} = 78{,}5$, s = 2,5. Với mức ý nghĩa 0,05 hãy xây dựng giả thuyết và đối thuyết thích hợp về khâu đóng gói mì ăn liền của nhà máy đạt yêu cầu không?

18. Đo chỉ số mỡ sữa X của 130 con bò lai F_1 ta được kết quả sau

X	3,0- 3,6	3,6- 4,2	4,2– 4,8	4,8 –5,4	5,4 –6,0	6,0 – 6,6	6,6 –7,2
n_i	2	8	35	43	22	15	5

Biết chỉ số mỡ sữa trung bình của giống bò thuần chủng là 4,95. Với mức ý nghĩa 0,01. Hãy đưa ra kết luận về việc lai tạo giống biết rằng chỉ số mỡ sữa X có phân phối chuẩn.

19. Phân tích hàm lượng mùn trong một loại đất theo hai phương pháp ta có kết quả sau:

Phương pháp 1: 27,5 27,0 27,3 27,6 27,8 (đơn vị %)

Phương pháp 2: 27,9 27,2 26,5 26,3 27,0 27,4 27,3 26,8 (đơn vị %)

Với mức ý nghĩa 0,05 hãy xây dựng giả thuyết và đối thuyết thích hợp và đưa ra kết luận.

20. Người ta chiếu xạ liều 3000 Rơnghen vào một quần thể ruồi dấm thấy trong số 805 con ở thế hệ F_1 có 80 con bị đột biến. Trong khi đó cũng chiếu xạ vào một quần thể ruồi dấm khác có cho ăn kèm theo một loại đường thì trong số 2756 con ở thế hệ F_1 có 357 con bị đột biến . Với mức ý nghĩa 0,05 hãy xây dựng cặp giả thuyết đối thuyết thích hợp và đưa ra kết luận.

21. Để so sánh hai loại thức ăn đối với việc tăng trọng của lợn người ta đã tiến hành thí nghiệm trên hai mẫu :

Mẫu I cho 8 con lợn ăn loại thức ăn A sau 1 tháng được kết quả sau:

X : 12,3 13,4 14,6 11,0 16,1 11,3 12,9 10,7

Mẫu II cho 7 con lợn ăn loại thức ăn B sau 1 tháng được kết quả sau:

Y : 13,2 14,3 16,8 13,1 14,5 15,7 14,5

Với mức ý nghĩa 0,05 hãy đưa ra cặp giả thuyết đối thuyết thích hợp rồi đưa ra kết luận.

22. Để khảo sát tác dụng của việc bón phân cho ngô 70 đơn vị đạm/ha, người ta trồng liền nhau mảnh đối chứng (không bón đạm) và mảnh thực nghiệm trên 15 thửa ruộng sau khi thu hoạch được kết quả sau:

Trọng lượng mảnh đối chứng X				55,8	53,3	30,1	51,0	37,8	68,8
Trọng lượng mảnh thực nghiệm Y				60,4	58,7	28,9	48,0	39,7	68,8
X	57,7	59,1	49,4	35,4	42,7	21,2	28,3	57,3	42,4
Y	56,8	40,6	57,3	44,3	32,2	47,7	77,0	55,1	66,1

Biết X, Y là các biến chuẩn. Với mức ý nghĩa α = 0,05. Hãy xây dựng cặp giả thuyết đối thuyết thích hợp và đưa ra kết luận.

23. Điều tra 320 gia đình có 5 con ta có các số liệu sau:

Số con trai X	5	4	3	2	1	0
Số gia đình n_i	18	56	110	88	40	8

Với mức ý nghĩa α = 0,05 hãy kiểm định giả thuyết đối thuyết

H_0: Số con trai X ~ B(5, 0,5)

H_1: Trái với H_0

24. Số tai nạn giao thông xảy ra mỗi ngày X tại một thành phố được ghi trong bảng sau:

X	0	1	2	3	4	5	6	7	8
n_i	10	32	46	35	20	9	2	1	1

Với mức ý nghĩa α = 0,05 hãy kiểm định giả thuyết : Số tai nạn giao thông không xảy ra trong ngày tuân theo luật Poisson.

25. Chiều cao X của cây dầu sau 6 tháng tuổi quan sát được cho ở bảng sau:

X	24 - 30	30 - 36	36 - 42	42 - 48	48 - 54	54 - 60	60 - 66
n_i	12	24	35	47	43	32	7

Với mức ý nghĩa α = 0.05 hãy kiểm định giả thuyết X có phân phối chuẩn.

26. Một loài hoa hồng có 4 màu : đỏ, hồng, bạch và vàng. Với mẫu gồm 200 bông hoa hồng thuộc loài hoa trên ta có bảng số liệu sau:

Màu hoa	đỏ	hồng	bạch	vàng
Số hoa	27	65	75	33

Với mức ý nghĩa 0,05 hãy kiểm định giả thuyết H_0 : Các màu hoa đỏ, hồng, bạch, vàng theo tỉ lệ 1 : 2 : 2 : 1.

27. Chi phí về văn hoá X (Đơn vị 100000đ/năm) và chi phí về đi lại Y (Đơn vị 100000 đồng/năm) của 15 gia đình cho bởi bảng sau:

X	12	6,5	6,2	8,8	4,5	7,0	7,1	20	15	7,5	8,5	10,9	8,2	8	10,5
Y	5,9	6,7	4,5	4,8	10	5,5	5,2	15	7,0	4,0	5,5	8,2	5,4	8,4	7,0

Sử dụng tiêu chuẩn về dấu kiểm định giả thuyết: X và Y có cùng qui luật xác suất với mức ý nghĩa 0,05.

28. Mức tiêu thụ xăng của 3 loại xe A, B, C (lít/100km) lần lượt là X , Y, Z. Người ta cho chạy thử 7 xe A, 7 xe B và 8 xe C các số liệu thu được cho ở bảng sau:

Y : 9,4 7,5 6,9 8,9 9,4 10 8,1

Z : 7,1 8,4 7,0 9,8 8,7 10 7,9 8,2

Với mức ý nghĩa 0,05 sử dụng tiêu chuẩn Kruskal – Wallis hãy kiểm định giả thuyết: Mức tiêu thụ xăng của 3 loại xe nói trên có cùng qui luật xác suất

29. Một mẫu điều tra lương của công nhân một nhà máy may X_1, lương của công nhân nhà máy chế biến hải sản X_2, lương của công nhân nhà máy sản xuất dày da xuất khẩu X_3 và lương vủa công nhân nhà máy chế biến hàng nông sản X_4 tại một khu chế suất cho bởi bảng số liệu sau: (Đơn vị 100000 đồng/tháng)

X_1 : 8,5 8,8 7,9 8,5 9,2 9,5 8,3

X_2 : 9,0 9,1 8,7 8,6 9,4 9,2 8,5 9,1

X_3 : 10 9,4 9,2 8,6 8,7 8,1 9,9

X_4 : 8,1 8,8 8,6 9,0 9,2 7,8 8,7 8,9 9,1

Ở mức ý nghĩa 0,05 sử dụng tiêu chuẩn Kruskal – Wallis hãy kiểm định giả thuyết:

Mức lương của công nhân bốn nhà máy trên là như nhau.

30. Chiều cao X của một mẫu ngẫu nhiên của 12 sinh viên nam tại Hà nội và 14 sinh viên nam tại thành phố Hồ Chí Minh cho bởi bảng số liệu sau:

X: 1,65 1,72 1,60 1,68 1,59 1,75 1,77 1,66 1,78 1,80 1,56 1,70

Y: 1,59 1,61 1,64 1,70 1,68 1,57 1,55 1,78 1,72 1,77 1,60 1,64 1,62 1,77

Ở mức ý nghĩa 0,05 sử dụng tiêu chuẩn Mann – Whitney hãy kiểm định giả thuyết:

X và Y có cùng qui luật phân phối.