KIỂM ĐỊNH GIẢ THIẾT THỐNG KÊ

CHƯƠNG :

KIỂM ĐỊNH GIẢ THIẾT THỐNG KÊ

1.KHÁI NIỆM

Dựa trên các số liệu thu được từ các mẫu, chúng ta thiết lập một giả thiết thống kê mà chúng ta muốn nghiên cứu,

- Giả thiết thống kê là một giả thiết được đặc trưng bởi các tham số của tổng thể (ví dụ : trung bình tổng thể, tỷ lệ tổng thể, phương sai tổng thể)

Giả sử θ là tham số của tổng thể mà chúng ta cần kiểm định và θ_0 là giá trị của tham số biết được dựa trên một nguồn tin nào đó.

Lập giả thiết

$H_0 : \theta = \theta_0$ (thường được gọi là giả thiết không).

- Vấn đề đặt ra ở đây là trên cơ sở các dữ liệu có được, chúng ta sẽ bác bỏ hay chấp nhận giả thiết trên, công việc mà chúng ta tiến hành để đi đến kết luận chấp nhận hay bác bỏ giả thiết H_0 đó được gọi là kiểm định giả thiết thống kê.

VD:

Theo một nguồn tin tại một địa phương A, năng suất trung bình về lúa của vụ mùa Đông Xuân là 7 tấn/ mẫu.

Gọi θ là năng suất lúa trung bình của vụ mùa Đông xuân tại địa phương trên (chưa biết)

θ_0 là năng suất lúa trung bình của vụ mùa Đông xuân theo nguồn tin = 7 tấn/mẫu.

Lập giả thiết

$H_0 : \theta = \theta_0$ = 7 tấn/mẫu

Vấn đề đặt ra là chấp nhận hay bác bỏ giả thiết H_0.

VD:

Trước ngày bầu cử Tổng thống .Tại một dịa phương, theo nguồn tin của một hãng thông tấn có 65% cử tri sẽ bầu cho ứng cử viên B.

Gọi θ là tỷ lệ cử tri sẽ bầu cho ứng cử viên B (chưa biết).

θ_0 là tỷ lệ cử tri bầu cho ứng cử viên B theo nguồn tin của hãng thông tấn = 65%.

Lập giả thiết

$H_0 : \theta = \theta_0$ = 65%

Chấp nhận hay bác bỏ giả thiết H_0.

- Trong quá trình kiểm định có thể mắc phải hai sai lầm:
- Sai lầm loại 1 : thực sự giả thiết H_0 đúng nhưng qua kiểm định ta kết luận giả thiết H_0 sai.
- Sai lầm loại 2: giả thiết H_0 sai nhưng qua kiểm định ta kết luận giả thiết H_0 đúng.

NHẬN XÉT:

Hai sai lầm này có tính đối kháng nhau , nghĩa là muốn hạn chế sai lầm loại 1,thì phải mở rộng miền chấp nhận, như vậy làm tăng khả năng mắc sai lầm loại 2. Ngược lại muốn hạn chế sai lầm loại 2, thì phải thu hẹp miền chấp nhận, như vậy sẽ dể mắc phải sai lầm loại 1.

- Nếu :P(mắc sai lầm loại 1)=P(bác bỏ H_0 / H_0 đúng) mà giảm ,

  thì : P(mắc sailầm loại 2)=P(chấpnhận H_0 / H_0 sai) tăng.

- Nếu P(mắc sailầm loại 2)=P(chấpnhận H_0 / H_0 sai)mà giảm,

  thì :P(mắc sai lầm loại 1)=P(bác bỏ H_0 / H_0 đúng)tăng.

Ngoài giả thiết H_0 , người ta còn lập ra một giả thiết trái ngược với giả thiết H_0 , gọi là giả thiết đối, ký hiệu là H_1

Ví dụ

$$H_0 : \theta = \theta_0 \quad ; \quad H_1 : \theta \neq \theta_0$$

$$H_0 : \theta \geq \theta_0 \quad ; \quad H_1 : \theta < \theta_0$$

$$H_0 : \theta \leq \theta_0 \quad ; \quad H_1 : \theta > \theta_0$$

Vấn đề đặt ra ở đây là dựa vào đâu để chấp nhận hay bác bỏ một giả thiết thống kê.

Người ta dựa trên nguyên lý sau:

"Nếu một biến cố có xác suất rất nhỏ, thì trong một hay vài phép thử, biến cố đó sẽ không xãy ra "

- Vậy giả thiết H_0 sẽ bị bị bác bỏ.

  Nếu P(bác bỏ H_0 / H_0 đúng) rất nhỏ.

Một kiểm định giả thiết thống kê lý tưởng nếu cả hai loại sai lầm loại 1 và sai lầm loại 2 đạt cực tiểu, tuy nhiên điều này không thực hiện được.

Giửa hai loại sai lầm, thì sai lầm loại 1 đáng quan tâm hơn. Vì thế người ta ấn định trước xác suất mắc sai lầm loại 1, xác suất này được gọi là **mức ý nghĩa**, ký hiệu là α .

Thông thường α được cho với các giá trị : 1%, 2% , 3%, 4% , 5%.

Ví dụ P(bác bỏ H_0 / H_0 đúng) $= \alpha = 2\%$.

Xác suất mắc sai lầm loại 2 được ký hiệu là β ,

- Nếu P(chấp nhận H_0 / H_0 sai) $= \beta$

Thì P(bác bỏ H_0 / H_0 sai) $= 1-\beta$

$(1-\beta)$ được gọi là năng lực kiểm định.

- Nếu P(bác bỏ H_0 / H_0 đúng) $= \alpha$

  Thì P(chấp nhận H_0 / H_0 đúng) $= 1-\alpha$

CHÚ Ý:

Trong quá trình kiểm định giả thiết thống kê nếu dẫn đến việc chấp nhận giả thiết H_0, chúng ta không nên hiểu giả thiết H_0 đúng, mà chỉ nên hiểu với những dữ liệu đã có chưa đủ cơ sở để bác bỏ giả thiết H_0, cần phải nghiên cứu tiếp.

Công việc kiểm định một giả thiết thống kê được tiến hành qua các bước sau:

i) Lập giả thiết không $H_0 : \theta = \theta_0$
(hay $H_0 : \theta \geq \theta_0$ hay $H_0 : \theta \leq \theta_0$)

ii) Lập giả thiết đối $H_1 : \theta \neq \theta_0$
(hay $H_1 : \theta < \theta_0$ hay $H_1 : \theta > \theta_0$)

iii) Xác định mức ý nghĩa α.

iv) Tùy trường hợp chọn thống kê kiểm định thích hợp.

v) Trên cơ sở giả thiết H_0 đúng, chọn miền bác bỏ giả thiết H_0.

vi) Từ mẫu quan sát được, tính thống kê kiểm định.

vii) Dựa vào giá trị thống kê kiểm định có rơi vào miền bác bỏ H_0 hay không để kết luận chấp nhận hay bác bỏ giả thiết H_0.

Trong chương này chúng ta sẽ kiểm định giả thiết về trung bình tổng thể, tỷ lệ tổng thể, phương sai tổng thể, so sánh hai trung bình, so sánh hai tỷ lệ, kiểm định tính độc lập giửa hai dấu hiệu của tổng thể.

2. KIỂM ĐỊNH TRUNG BÌNH CỦA TỔNG THỂ

2.1. TỔNG THỂ CÓ PHƯƠNG SAI ĐÃ BIẾT

2.1.1 KIỂM ĐỊNH MỘT PHÍA

i) TRƯỜNG HỢP 1:

Giả sử chúng ta có một mẫu ngẫu nhiên của n quan sát từ một tổng thể có phân phối chuẩn trung bình là μ chưa biết và phương sai σ^2 đã biết.

μ_0 là trung bình của tổng thể được biết dựa theo một nguồn tin nào đó.

Nếu trung bình của mẫu cụ thể là $\overline{x}$.

Với mức ý nghĩa α.

Chúng ta lập các giả thiết sau:

Giả thiết không

$$H_0 : \mu = \mu_0 \qquad \text{hay} \qquad H_0 : \mu \le \mu_0$$

Giả thiết đối

$$H_1 : \mu \succ \mu_0$$

Chọn thống kê kiểm định là

$$Z = \frac{(\overline{X} - \mu)\sqrt{n}}{\sigma}$$

Thì ĐLNN Z có phân phối chuẩn tắc. $Z \sim N(0,1)$

Vì giả sử giả thiết H_0 đúng nghĩa là:

$$\mu = \mu_0$$

Nên ĐLNN $Z = \dfrac{(\overline{X} - \mu_0)\sqrt{n}}{\sigma} \sim N(0,1)$

Tính giá trị thống kê $z = \dfrac{(\overline{x} - \mu_0)\sqrt{n}}{\sigma}$

Với mức ý nghĩa α ,

Sử dụng phân phối chuẩn

Từ $P(Z > z_\alpha) = \alpha \Rightarrow z_\alpha$

- z_α được gọi là giá trị tới hạn

Gọi W là miền bác bỏ giả thiết H_0

Chúng ta bác bỏ giả thiết H_0 nếu $z \in W$

Ta có

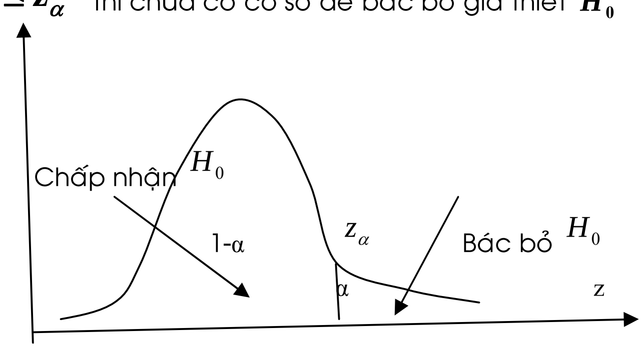
$$P(z \in W) = \alpha \Leftrightarrow P(z > z_\alpha) = \alpha$$

$$W = \{\ z / z > z_\alpha\ \}$$

Quy tắc kiểm định

- Nếu $z > z_\alpha$ thì bác bỏ giả thiết H_0
- Nếu $z \le z_\alpha$ thì chưa có cơ sở để bác bỏ giả thiết H_0

HÌNH VẼ



VD:

Tại một địa phương chiều cao trung bình của thanh niên vào năm 1998 là 165 cm. Một nhóm nhà nghiên cứu về nhân trắc học muốn tìm hiểu xem chiều cao trung bình của thanh niên hiện tại có cao hơn so với trước đây hay không, Họ chọn ngẫu nhiên 169 thanh niên thì thấy chiều cao trung bình của các thanh niên này là 168cm.

Với mức ý nghĩa 3% có thể đưa ra một kết luận gì về chiều cao trung bình của thanh niên. Cho biết chiều cao của thanh niên có phân phối chuẩn với độ lệch chuẩn là 2cm.

GIẢI:

Gọi X(cm) là chiều cao của thanh niên hiện nay.

μ là chiều cao trung bình của thanh niên hiện nay.

μ_0 là chiều cao trung bình của thanh niên vào năm 1998.

Lập các giả thiết

$$H_0 : \mu = \mu_0 = 165$$

$$H_1 : \mu > \mu_0$$

$\alpha = 3\%$

$n = 169$

$\sigma = 2$

Chọn thống kê $Z = \dfrac{(\overline{X} - \mu)\sqrt{n}}{\sigma}$

Ta có ĐLNN X có phân phối chuẩn, suy ra Z~N(0,1)

Từ $P(Z > z_\alpha) = \alpha = 0{,}03 \Rightarrow z_\alpha = 2{,}17$

Tính $z = \dfrac{(\overline{x} - \mu_0)\sqrt{n}}{\sigma} = \dfrac{(168 - 165)\sqrt{169}}{2} = 19{,}5$

Suy ra $z > z_\alpha$

Vậy bác bỏ giả thiết H_0, nghĩa là chiều cao trung bình của thanh niên hiện nay cao hơn chiều cao trung bình của thanh niên vào năm 1998.

ii) TRƯỜNG HỢP 2:

Giả sử chúng ta có một mẫu ngẫu nhiên của n quan sát từ một tổng thể có phân phối chuẩn trung bình là μ chưa biết và phương sai là σ^2 đã biết.

μ_0 là trung bình của tổng thể dựa trên một nguồn tin nào đó.

Nếu mẫu cụ thể có trung bình là $\overline{x}$.

Với mức ý nghĩa là α

Chúng ta lập các giả thiết sau:

Giả thiết không

$$H_0 : \mu = \mu_0 \quad \text{hay} \quad H_0 : \mu \geq \mu_0$$

Giả thiết đối

$$H_1 : \mu < \mu_0$$

Chọn thống kê kiểm định là

$$Z = \frac{(\overline{X} - \mu)\sqrt{n}}{\sigma}$$

Thì ĐLNN Z có phân phối chuẩn tắc. $Z \sim N(0,1)$

Vì giả sử giả thiết H_0 đúng, nghĩa là $\mu = \mu_0$

Do đó ĐLNN $Z = \dfrac{(\overline{X} - \mu_0)\sqrt{n}}{\sigma} \sim N(0,1)$

Từ $$P(Z < -z_\alpha) = \alpha \Rightarrow -z_\alpha$$

Tính $$z = \frac{(\overline{x} - \mu_0)\sqrt{n}}{\sigma}$$

Gọi W là miền bác bỏ giả thiết H_0

Chúng ta bác bỏ giả thiết H_0 nếu $z \in W$

Ta có

$$P(z \in W) = \alpha \Leftrightarrow P(z < -z_\alpha) = \alpha$$

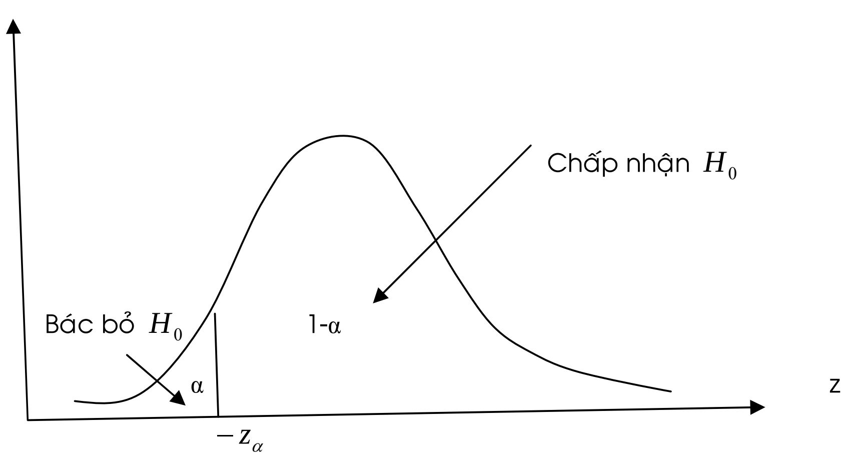
$$W = \{\ z / z < -z_\alpha\ \}$$

Quy tắc kiểm định

- Nếu $z < -z_\alpha$ thì bác bỏ giả thiết H_0
- Nếu $z \geq -z_\alpha$ thì chưa đủ cơ sở để bác bỏ giả thiết H_0

HÌNH VẼ

.



VD:

Một hãng điện tử quãng cáo đèn hình TV của hãng có tuổi thọ là 9000 giờ. Kiểm tra 15 đèn hình TV của hãng thấy tuổi thọ trung bình là 8800 giờ.

Với mức ý nghĩa 5% hãy xét xem quãng cáo của hãng có đáng tin cậy không ? Cho biết tuổi thọ của đèn hình TV có phân phối chuẩn với độ lệch chuẩn là 500 giờ.

GIẢI:

Gọi X(giờ) là tuổi thọ của đèn hình.

μ là tuổi thọ trung bình của đèn hình của hãng điện tử trên.

μ_0 = 9000 là tuổi thọ trung bình của đèn hình theo quãng cáo.

Lập các giả thiết

$$H_0 : \mu = \mu_0 = 9000$$

$$H_1 : \mu < \mu_0$$

$\overline{x}$ = 8800 là tuổi thọ trung bình của đèn hình theo mẫu kiểm tra.

σ = 500 là độ lệch chuẩn của tổng thể

n = 15 là kích thước mẫu.

α = 5% là mức ý nghĩa

Chọn thống kê $Z = \dfrac{(\overline{X} - \mu)\sqrt{n}}{\sigma}$

Ta có ĐLNN Z có phân phối chuẩn tắc. Z~N(0,1)

Từ $P(Z < -z_\alpha) = \alpha = 0{,}05 \Rightarrow -z_\alpha = -1{,}65$

Tính $z = \dfrac{(\overline{x} - \mu_0)\sqrt{n}}{\sigma} = \dfrac{(8800 - 9000)\sqrt{15}}{500} = -1{,}55$

Suy ra $z > -z_\alpha$

Vậy chưa đủ cơ sở để bác bỏ giả thiết H_0,

nghĩa là quãng cáo đáng tin cậy với mức ý nghĩa 5%.

2.1.2.KIỂM ĐỊNH HAI PHÍA

Giả sử chúng ta có một mẫu ngẫu nhiên của n quan sát từ một tổng thể có phân phối chuẩn với trung bình μ chưa biết và phương sai σ^2 đã biết.

Nếu mẫu cụ thể có trung bình $\overline{x}$

Với mức ý nghĩa α

Chúng ta lập các giả thiết sau:

Giả thiết không

$$H_0 : \mu = \mu_0$$

Giả thiết đối

$$H_1 : \mu \neq \mu_0$$

Chọn thống kê kiểm định là : $Z = \dfrac{(\overline{X} - \mu)\sqrt{n}}{\sigma}$

Thì ĐLNN Z có phân phối chuẩn tắc. $Z \sim N(0,1)$

Từ $\quad P(|Z| > z_\alpha) = \alpha \Rightarrow z_{\frac{\alpha}{2}}$

Tính $\quad z = \dfrac{(\overline{x} - \mu_0)\sqrt{n}}{\sigma}$

Gọi W là miền bác bỏ giả thiết H_0

Chúng ta sẽ bác bỏ giả thiết H_0 nếu $z \in W$

Ta có

$$P(z \in W) = \alpha \Leftrightarrow P(|z| > z_{\frac{\alpha}{2}}) = \alpha \Leftrightarrow$$

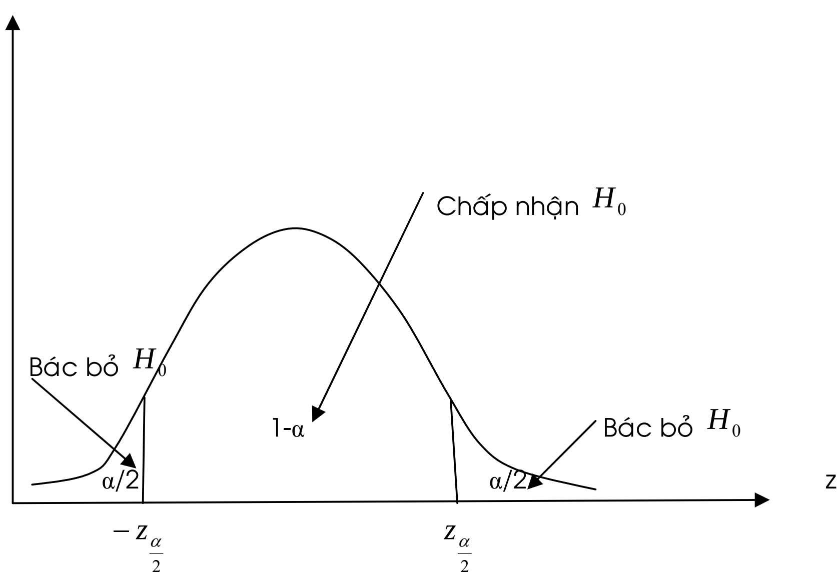
$$\Leftrightarrow P(z > z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}; P(z < -z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$$

$$W = \left\{\ z / |z| > z_{\frac{\alpha}{2}}\ \right\}$$

Quy tắc kiểm định

- Nếu $|z| > z_{\frac{\alpha}{2}}$ thì bác bỏ giả thiết H_0
- Nếu $|z| \leq z_{\frac{\alpha}{2}}$ chưa đủ cơ sở để bác bỏ giả thiết H_0

HÌNH VẼ



VD:

Khảo sát 121 công nhân của công ty may về thu nhập, nhận thấy thu nhập trung bình của một công nhân là 1,6 triệu đồng/tháng.

Theo một nguồn tin từ ban giám đốc thì thu nhập trung bình của công nhân công ty là 1,65 triệu đồng/tháng. Với mức ý nghĩa 5% hãy xét xem nguồn tin từ ban giám đốc có đáng tin cậy không? Cho biết thu nhập của công nhân có phân phối chuẩn với độ lệch chuẩn là 100 ngàn đồng.

GIẢI:

Gọi X(triệu đồng) là thu nhập của công nhân

μ là thu nhập trung bình của công nhân công ty

μ_0 là thu nhập trung bình của công nhân theo nguồn tin từ ban giám đốc = 1,65

$\overline{x}$ là thu nhập trung bình của công nhân theo mẫu = 1,6

σ độ lệch chuẩn = 0,100

α mức ý nghĩa = 0,05

n kích thước mẫu = 121

Lập các giả thiết

$$H_0 : \mu = \mu_0 = 1,65$$

$$H_1 : \mu \neq \mu_0$$

Chọn thống kê kiểm định $Z = \dfrac{(\overline{X} - \mu)\sqrt{n}}{\sigma}$

Ta có ĐLNN Z có phân phối chuẩn tắc. Z~N(0,1)

Tính $z = \dfrac{(\overline{x} - \mu_0)\sqrt{n}}{\sigma} = \dfrac{(1,6 - 1,65)\sqrt{121}}{0,1} = -5,5$

Từ $P(|Z| > z_{\frac{\alpha}{2}}) = \alpha = 0,05 \Rightarrow z_{\frac{\alpha}{2}} = 1,96$

Suy ra $|z| > z_{\frac{\alpha}{2}}$

Vậy bác bỏ giả thiết H_0, nghĩa là nguồn tin từ ban giám đốc không đáng tin cậy.

2.2 TỔNG THỂ CÓ PHƯƠNG SAI CHƯA BIẾT

2.2.1 KÍCH THƯỚC MẪU LỚN

Giả sử chúng ta có một mẫu ngẫu nhiên của n quan sát từ một tổng thể có trung bình là μ chưa biết và phương sai là σ^2 chưa biết.

Nếu kích thước mẫu lớn (n≥30).

Trong trường hợp này chúng ta vẫn dùng thống kê kiểm định như trên nhưng thay thế σ^2 bởi phương sai mẫu hiệu chỉnh $\hat{s}^2$

Ta có: $$Z = \frac{(\overline{X} - \mu)\sqrt{n}}{\hat{s}}$$

<u>CHÚ Ý:</u>

Trong trường hợp kích thước mẫu lớn ($n \geq 30$).

Tổng thể không có phân phối chuẩn, theo định lý giới hạn trung tâm thống kê kiểm định

$$Z = \frac{(\overline{X} - \mu)\sqrt{n}}{\hat{s}}$$

có phân phối xấp xỉ gần chuẩn. $Z \sim N(0,1)$

VD:

Theo nguồn tin từ ban giám đốc một siêu thị, số tiền trung bình một khách hàng sử dụng để mua hàng vào các ngày của tháng 10, 11, 12 là 250 ngàn đồng.Trong tuần lể đầu tiên của tháng 1, khảo sát 64 khách hàng thấy trung bình một khách hàng sử dụng số tiền là 280 ngàn đồng để mua hàng và độ lệch chuẩn hiệu chỉnh của mẫu là 80 ngàn đồng.

Với mức ý nghĩa là 2%, hãy xét xem phải chăng sức mua của khách hàng vào những ngày cuối năm (Âm lịch) có khuynh hướng tăng.

GIẢI:

Gọi X(ngàn đồng) là số tiền một khách hàng sử dụng để mua hàng vào các ngày cuối năm (Âm lịch)

μ là số tiền trung bình một khách hàng sử dụng để mua hàng vào các ngày cuối năm (Âm lịch)

μ_0 là số tiền trung bình một khách hàng sử dụng để mua hàng trong tháng 10, 11, 12

Lập các giả thiết:

Giả thiết không

$$H_0 : \mu = \mu_0 = 250$$

Giả thiết đối

$$H_1 : \mu > \mu_0$$

$\overline{x}$ số tiền trung bình một khách hàng sử dụng để mua hàng vào các ngày cuối năm theo mẫu = 280

$\hat{s}$ độ lệch chuẩn hiệu chỉnh của mẫu = 80

α mức ý nghĩa = 0,02

Chọn thống kê kiểm định là $Z = \frac{(\overline{X} - \mu)\sqrt{n}}{\hat{s}}$

Thì ĐLNN Z có phân phối chuẩn tắc. $Z \sim N(0,1)$

Từ $P(Z > z_\alpha) = \alpha \Rightarrow z_\alpha = 2{,}33$

Tính $$z = \frac{(\overline{x} - \mu_0)\sqrt{n}}{\hat{s}} = \frac{(280 - 250)\sqrt{64}}{80} = 3$$

Suy ra $z > z_\alpha$

Vậy bác bỏ giả thiết H_0, nghĩa là sức mua của khách hàng vào các ngày cuối năm Âm lịch có khuynh hướng tăng.

2.2.2 KÍCH THƯỚC MẪU NHỎ (n < 30)

Giả sử chúng ta có một mẫu ngẫu nhiên của n quan sát từ một tổng thể có phân phối chuẩn với trung bình là μ chưa biết và phương sai là σ^2 chưa biết

Nếu mẫu cụ thể có trung bình là $\overline{x}$ và phương sai hiệu chỉnh của mẫu là $\hat{s}^2$

Với mức ý nghĩa α.

Chúng ta xét các trường hợp sau:

i) **KIỂM ĐỊNH MỘT PHÍA**

- **TRƯỜNG HỢP 1**

Lập các giả thiết:

Giả thiết không

$$H_0 : \mu = \mu_0 \quad \text{hay} \quad H_0 : \mu \le \mu_0$$

Giả thiết đối

$$H_1 : \mu > \mu_0$$

Chọn thống kê kiểm định

$$T = \frac{(\overline{X} - \mu)\sqrt{n}}{\hat{s}}$$

Thì ĐLNN T có phân phối Student bậc tự do là k=n-1. $T \sim T(n-1)$

Từ $P(T > t_\alpha) = \alpha \Rightarrow t_\alpha$

Tính $$t = \frac{(\overline{x} - \mu_0)\sqrt{n}}{\hat{s}}$$

Gọi W là miền bác bỏ giả thiết H_0

Chúng ta sẽ bác bỏ giả thiết H_0 nếu $t \in W$ Ta có

$$P(t \in W) = \alpha \Leftrightarrow P(t > t_\alpha) = \alpha$$

$$W = \{\ t / t > t_\alpha\ \}$$

Quy tắc kiểm định

- Nếu $t > t_\alpha$ thì bác bỏ giả thiết H_0
- Nếu $t \le t_\alpha$ thì không đủ cơ sở để bác bỏ giả thiết H_0

  - **TRƯỜNG HỢP 2**

Lập các giả thiết:

Giả thiết không

$$H_0 : \mu = \mu_0 \quad \text{hay} \quad H_0 : \mu \ge \mu_0$$

Giả thiết đối

$$H_1 : \mu < \mu_0$$

Chọn thống kê kiểm định

$$T = \frac{(\overline{X} - \mu)\sqrt{n}}{\hat{s}}$$

Thì ĐLNN T có phân phối Student bậc tự do là k=n-1. $T \sim T(n-1)$

Vì giả sử giả thiết H_0 đúng, nghĩa là $\mu = \mu_0$

Nên ĐLNN $T = \dfrac{(\overline{X} - \mu_0)\sqrt{n}}{\hat{s}} \sim T(n-1)$

Từ $\quad P(T < -t_\alpha) = \alpha \Rightarrow -t_\alpha$

Tính $\quad t = \dfrac{(\overline{x} - \mu_0)\sqrt{n}}{\hat{s}}$

Gọi W là miền bác bỏ giả thiết H_0

Chúng ta bác bỏ giả thiết H_0 nếu $t \in W$

Ta có

$$P(t \in W) = \alpha \Leftrightarrow P(t < -t_\alpha) = \alpha$$

$$W = \{\ t / t < -t_\alpha\ \}$$

Quy tắc kiểm định

- Nếu $t < -t_\alpha$ thì bác bỏ giả thiết H_0
- Nếu $t \ge -t_\alpha$ thì chưa đủ cơ sở để bác bỏ giả thiết H_0

VD:

Tại một trại chăn nuôi gà trong năm 2008, trọng lượng một con gà khi đưa ra thị trường là 2,9kg. Năm 2009 trại sử dụng một loại thức ăn mới cho gà, chọn ngẫu nhiên 16 con gà trong số gà đưa ra thị trường thấy trọng lượng trung bình là 3 kg và độ lệch chuẩn hiệu chỉnh của mẫu là 0,1 kg.

Với mức ý nghĩa là 2%, hãy xét xem có phải thức ăn mới làm tăng trọng lượng gà.

GIẢI:

Gọi X(kg) là trọng lượng gà khi đưa ra thị trường trong năm 2009

μ trọng lượng trung bình của một con gà khi đưa ra thị trường trong năm 2009

μ_0 trọng lượng trung bình của một con gà khi đưa ra thị trường trong năm 2008

Lập các giả thiết

Giả thiết không

$$H_0 : \mu = \mu_0 = 2{,}9$$

Giả thiết đối

$$H_1 : \mu > \mu_0$$

$\overline{x}$ trọng lượng trung bình một con gà khi đưa ra thị trường trong năm 2009 theo mẫu = 3

$\hat{s}$ độ lệch chuẩn hiệu chỉnh của mẫu = 0,1

n kích thước mẫu = 16

α mức ý nghĩa = 0,01

Chọn thống kê kiểm định $T = \dfrac{(\overline{X} - \mu)\sqrt{n}}{\hat{s}}$

Sử dụng phân phối Student với bậc tự do là k=n-1=16-1=15

Từ $P(T > t_\alpha) = 0{,}01 \Rightarrow t_\alpha = 2{,}602$

Tính $t = \dfrac{(\overline{x} - \mu_0)\sqrt{n}}{\hat{s}} = \dfrac{(3 - 2{,}9)\sqrt{16}}{0{,}1} = 4$

Suy ra $t > t_\alpha$

Vậy bác bỏ giả thiết H_0, nghĩa là thức ăn mới làm tăng trọng lượng gà.

VD:

Môt loại mì ăn liền có trọng lượng trung bình ghi trên bao bì là 60g, kiểm tra 20 gói thấy trọng lượng trung bình một gói là 59g và độ lệch chuẩn hiệu chỉnh của mẫu là 2g.

Với mức ý nghĩa là 5%, hãy xét xem phải chăng trọng lượng trung bình ghi trên bao bì cao hơn trọng lượng trung bình thực sự của một gói mì ăn liền.

GIẢI:

Gọi X(g) là trọng lượng một bao mì ăn liền

μ trọng lượng trung bình của một bao mì ăn liền

μ_0 trọng lượng trung bình ghi trên bao bì

Lập các giả thiết:

Gỉa thiết không

$$H_0 : \mu = \mu_0 = 60$$

Giả thiết đối

$$H_1 : \mu < \mu_0$$

$\overline{x}$ trọng lượng trung bình một gói mì theo mẫu = 59

$\hat{s}$ độ lệch chuẩn hiệu chỉnh = 2g

n kích thước mẫu = 20

α mức ý nghĩa = 0,05

Chọn thống kê kiểm định $T = \dfrac{(\overline{X} - \mu)\sqrt{n}}{\hat{s}}$

$$T \sim T(n-1) \Rightarrow T \sim T(19)$$

Sử dụng phân phối Student với bậc tự do là k=n-1=19

Từ $P(T < -t_\alpha) = \alpha \Rightarrow -t_\alpha =$ -1,729

Tính $t = \dfrac{(\overline{x} - \mu)\sqrt{n}}{\hat{s}} = \dfrac{(59 - 60)\sqrt{20}}{2} = -2,24$

Suy ra

$$t < -t_\alpha$$

Vậy bác bỏ giả thiết H_0, nghĩa là trọng lượng trung bình của một gói mì ghi trên bao bì cao hơn trọng lượng trung bình thực sự.

ii) KIỂM ĐỊNH HAI PHÍA

Lập các giả thiết:

Giả thiết không

$$H_0 : \mu = \mu_0$$

Giả thiết đối

$$H_1 : \mu \neq \mu_0$$

Chọn thống kê kiểm định

$$T = \frac{(\overline{X} - \mu)\sqrt{n}}{\hat{s}}$$

Thì ĐLNN T có phân phối Student với bậc tự do là k=n-1. $T \sim T(n-1)$

Vì giả sử giả thiết H_0 đúng, nghĩa là $\mu = \mu_0$

Nên ĐLNN $T=\dfrac{(\overline{X}-\mu_0)\sqrt{n}}{\hat{s}} \sim T(n-1)$

Từ $\boldsymbol{P(|T|>t_{\frac{\alpha}{2}})=\alpha \Rightarrow t_{\frac{\alpha}{2}}}$

Tính $\boldsymbol{t=\dfrac{(\overline{x}-\mu_0)\sqrt{n}}{\hat{s}}}$

Gọi W là miền bác bỏ giả thiết $\boldsymbol{H_0}$

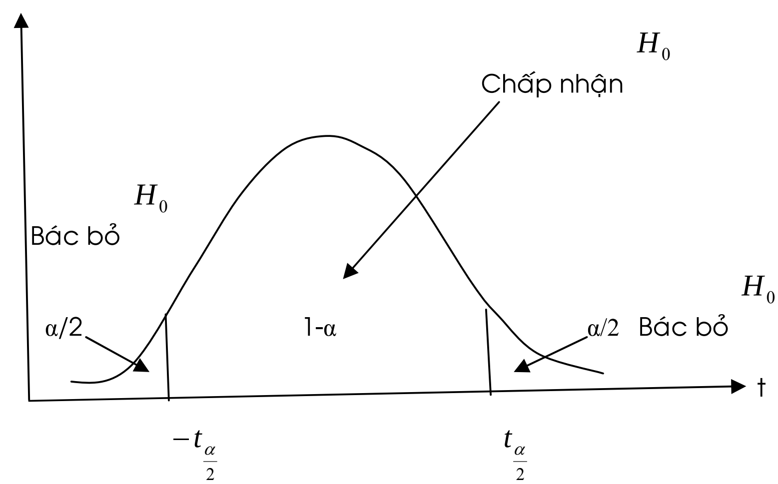
Chúng ta bác bỏ giả thiết $\boldsymbol{H_0}$ nếu $\boldsymbol{t \in W}$

Ta có $P(t \in W)=\alpha \Leftrightarrow P(|t|>t_{\frac{\alpha}{2}})=\alpha$

$$\boldsymbol{W=\left\{\ t\,/\,|t|>t_{\frac{\alpha}{2}}\ \right\}}$$

Quy tắc kiểm định

- ❖ Nếu $|t|>t_{\frac{\alpha}{2}}$ thì bác bỏ giả thiết H_0
- ❖ Nếu $|t|\le t_{\frac{\alpha}{2}}$ thì chưa đủ cơ sở để bác bỏ giả thiết H_0



VD

Chuỗi cửa hàng bán lẻ của một tập đoàn bán lẻ qua thống kê hằng năm đã đưa ra nhận định: trung bình lượng hàng hóa bán ra của các cửa hàng trong tháng 12 cao hơn tháng 11 là 20%. Trong năm nay vào cùng thời điểm như trên người ta chọn ngẫu nhiên 6 cửa hàng và được số liệu như sau (lượng hàng bán ra của tháng 12 cao hơn tháng 11)

19,2%   18,4%   19,8%   20,2%   20,4%   19%

Giả sử rằng tổng thể có phân phối chuẩn.

Với mức ý nghĩa 10%, hãy xét xem nhận định trên có còn đúng cho năm nay không?

GIẢI:

Gọi X(%) là(phần trăm) lượng hàng bán ra trong tháng 12 cao hơn so với tháng 11

μ là (phần trăm) trung bình lượng hàng bán ra trong tháng 12 cao hơn tháng 11 trong năm nay.

μ_0 là (phần trăm) trung bình lượng hàng bán ra trong tháng 12 cao hơn tháng 11 trong các năm trước.

Lập giả thiết

Giả thiết không

$$H_0 : \mu = \mu_0 = 20$$

Giả thiết đối

$$H_1 : \mu \neq \mu_0$$

$\overline{x}$ =19,5

$\hat{s}$ = 0,767

Chọn thống kê kiểm định

$$T = \frac{(\overline{X} - \mu)\sqrt{n}}{\hat{s}}$$

$$T \sim (n-1) \Rightarrow T \sim T(5)$$

Sử dụng phân phối Student với bậc tự do là k=n-1=5

Từ $P(|T| > t_{\frac{\alpha}{2}}) = \alpha = 0,10 \Rightarrow t_{\frac{\alpha}{2}} = 2,015$

Tính $t = \frac{(\overline{x} - \mu_0)\sqrt{n}}{\hat{s}} = \frac{(19,5 - 20)\sqrt{6}}{0,767} = -1,597$

Suy ra $|t| < t_{\frac{\alpha}{2}}$

Vậy chưa đủcơ sở để bác bỏ giả thiết H_0 ,nghĩa là nhận định trên vẫn còn đúng với mức ý nghĩa 10%.

3. KIỂM ĐỊNH TỶ LỆ CỦA TỔNG THỂ

TRƯỜNG HỢP KÍCH THƯỚC MẪU LỚN

Giả sử chúng ta có một mẫu của n quan sát từ một tổng thể có tỷ lệ thành công là p. Nếu kích thước mẫu lớn (n ≥ 30) và tỷ lệ thành công của mẫu ngẫu nhiên là f.

Với mức ý nghĩa α.

Chúng ta xét các trường hợp sau:

3.1. KIỂM ĐỊNH MỘT PHÍA

3.1.1. TRƯỜNG HỢP 1

Lập các giả thiết sau:

Giả thiết không

$$H_0 : p = p_0 \quad \text{hay} \quad H_0 : p \le p_0$$

Gỉa thiết đối

$$H_1 : p > p_0$$

Chọn thống kê kiểm định

$$Z = \frac{(f-p)\sqrt{n}}{\sqrt{p(1-p)}}$$

Vì kích thước mẫu lớn nên theo định lý giới hạn trung tâm ĐLNN Z có phân phối xấp xỉ chuẩn . Z~N(0,1).

Vì giả sử giả thiết H_0 đúng, nghĩa là $p = p_0$

Do đó ĐLNN

$$Z = \frac{(f-p_0)\sqrt{n}}{\sqrt{p_o(1-p_0)}} \sim N(0,1)$$

Ta có $P(Z > z_\alpha) = \alpha \Rightarrow z_\alpha$: giá trị tới hạn

Từ mẫu cụ thể tính giá trị $z = \dfrac{(f-p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}}$

Gọi W là miền bác bỏ giả thiết H_0

Chúng ta bác bỏ giả thiết H_0 nếu $z \in W$

Ta có $P(z \in W) = \alpha \Leftrightarrow P(z > z_\alpha) = \alpha$

Suy ra $W = \{\ z / z > z_\alpha\ \}$

Quy tắc kiểm định

- Nếu $z > z_\alpha$ thì bác bỏ giả thiết H_0
- Nếu $z \le z_\alpha$ thì chưa đũ cơ sở để bác bỏ giả thiết H_0

3.1.2 TRƯỜNG HỢP 2

Lập các giả thiết sau:

Giả thiết không

$$H_0 : p = p_0 \text{ hay } \quad H_0 : p \geq p_0$$

Giả thiết đối

$$H_1 : p < p_0$$

Chọn thống kê kiểm định

$$Z = \frac{(f - p)\sqrt{n}}{\sqrt{p(1-p)}}$$

Thì ĐLNN Z có phân phối chuẩn tắc. Z~N(0,1)

Vì giả sử giả thiết H_0 đúng, nghĩa là $p = p_0$

Do đó ĐLNN $Z = \dfrac{(f - p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}} \sim N(0,1)$

Ta có $\quad P(Z < -z_\alpha) = \alpha \Rightarrow -z_\alpha$

Từ mẫu cụ thể tính $z = \dfrac{(f - p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}}$

Miền bác bỏ $W = \{ \; z / z < -z_\alpha \; \}$

Quy tắc kiểm định

- Nếu $z < -z_\alpha$ thì bác bỏ giả thiết H_0
- Nếu $z \geq -z_\alpha$ thì chưa đủ cơ sở để bác bỏ giả thiết H_0

3.2. KIỂM ĐỊNH HAI PHÍA

Lập các giả thiết sau:

Giả thiết không

$$H_0 : p = p_0$$

Giả thiết đối

$$H_1 : p \neq p_0$$

Chọn thống kê kiểm định

$$Z = \frac{(f - p)\sqrt{n}}{\sqrt{p(1-p)}}$$

Thì ĐLNN Z có phân phối chuẩn tắc. Z~N(0,1)

Vì giả sử giả thiết H_0 đúng, nghĩa là $p = p_0$

Do đó ĐLNN $Z=\frac{(f-p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}}\sim N(0,1)$

Ta có $P(|Z|>z_{\frac{\alpha}{2}})=\alpha\Rightarrow z_{\frac{\alpha}{2}}$

Từ mẫu cụ thể tính $z=\frac{(f-p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}}$

Miền bác bỏ $W=\left\{\ z/|z|>z_{\frac{\alpha}{2}}\ \right\}$

Quy tắc kiểm định

- Nếu $|z|>z_{\frac{\alpha}{2}}$ thì bác bỏ giả thiết H_0
- Nếu $|z|\le z_{\frac{\alpha}{2}}$ thì chưa có cơ sở để bác bỏ giả thiết H_0

VD:

Giám đốc một công ty kiểm toán cho biết tỷ lệ nhân viên kiểm toán của công ty có văn bằng kiểm toán quốc tế là 50%. Chọn ngẫu nhiên 400 nhân viên kiểm toán của công ty thấy có 180 nhân viên có văn bằng kiểm toán quốc tế. Với mức ý nghĩa 5%, xét xem giám đốc công ty có nói hơi quá không?

GIẢI:

Gọi p là tỷ lệ nhân viên kiểm toán có văn bằng kiểm toán quốc tế của công ty.

p_0 là tỷ lệ nhân viên kiểm toán có văn bằng kiểm toán quốc tế theo giám đốc công ty.

Lập giả thiết

Giả thiết không

$$H_0: p=p_0=0{,}50$$

Giả thiết đối

$$H_1: p<p_0$$

f là tỷ lệ nhân viên kiểm toán có văn bằng kiểm toán quốc tế theo mẫu

n = 400 kích thước mẫu

α là mức ý nghĩa = 0,05

Chọn thống kê kiểm định $Z=\frac{(f-p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}}$

Z~N(0,1)

Ta có: $P(Z<-z_\alpha)=\alpha=0{,}05\Rightarrow -z_\alpha=-1{,}65$

Từ mẫu cụ thể tính: $z = \dfrac{(f - p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}} = -2$

Suy ra $z < -z_\alpha$

Vậy bác bỏ giả thiết H_0, nghĩa là giám đốc công ty nói hơi quá với mức ý nghĩa là 5%

VD:

Một hãng điện tử quảng cáo có 40% người tiêu dùng sử dụng sản phẩm của hãng.Khảo sát 169 người thấy có 40 người sử dụng sản phẩm của hãng. Với mức ý nghĩa 2%, xét xem quảng cáo của hãng có đáng tin cậy không?

GIẢI:

Gọi p là tỷ lệ người tiêu dùng sử dụng sản phẩm của hãng.

p_0 tỷ lệ người tiêu dùng sử dụng sản phẩm của hãng theo quảng cáo.

Lập các giả thiết:

Giả thiết không

$$H_0 : p = p_0 = 0{,}40$$

Gỉa thiết đối

$$H_1 : p < p_0$$

f là tỷ lệ người tiêu dùng sử dụng sản phẩm của hãng theo mẫu

n = 169

α mức ý nghĩa = 0,02

Chọn thống kê kiểm định $Z = \dfrac{(f - p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}}$

Thì Z~N(0,1)

Ta có: $P(Z < -z_\alpha) = \alpha = 0{,}02 \Rightarrow -z_\alpha = -2{,}05$

Từ mẫu cụ thể tính: $z = \dfrac{(f - p_0)\sqrt{n}}{\sqrt{p_o(1-p_0)}}$ = - 0,4111

Suy ra $z > -z_\alpha$

Vậy chưa đủ cơ sở để bác bỏ giả thiết H_0, nghĩa là quãng cáo đáng tin cậy với mức ý nghĩa 2%.

VD:

Tại một vườn trồng phong lan thời gian trước đây nhà vườn sử dụng loại thuốc A để diệt trừ sâu rầy, tỷ lệ thành công là 80%. Hiện nay nhà vườn sử dụng một loại thuốc mới B để diệt trừ sâu rầy, phun thuốc trên 1000 cây phong lan có 900 cây hết sâu rầy. Với mức ý nghĩa 1% xét xem tỷ lệ thành công của hai loại thuốc có khác nhau không?

GIẢI:

Gọi p là tỷ lệ thành công của loại thuốc B

p_0 là tỷ lệ thành công của loại thuốc A

Lập các giả thiết:

Giả thiết không

$$H_0 : p = p_0 = 0{,}80$$

Giả thiết đối

$$H_1 : p \neq p_0$$

f tỷ lệ thành công của loại thuốc B theo mẫu

n = 1000

α mức ý nghĩa = 0,01

Chọn thống kê kiểm định $Z = \dfrac{(f - p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}}$

Thì Z~N(0,1)

Ta có: $P(|Z| > z_{\frac{\alpha}{2}}) = \alpha = 0{,}01 \Rightarrow z_{\frac{\alpha}{2}} = 2{,}58$

Từ mẫu cụ thể tính: $z = \dfrac{(f - p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}} = 7{,}906$

Suy ra $|z| > z_{\frac{\alpha}{2}}$

Vậy bác bỏ giả thiết H_0, nghĩa là có sự khác biệt giửa tỷ lệ thành công của hai loại thuốc.

4. KIỂM ĐỊNH PHƯƠNG SAI CỦA TỔNG THỂ CÓ PHÂN PHỐI CHUẨN

Giả sử chúng ta có một mẫu ngẫu nhiên của n quan sát từ một tổng thể có phân phối chuẩn với phương sai là σ^2 (chưa biết).

Phương sai hiệu chỉnh của mẫu là $\hat{s}^2$.

Với mức ý nghĩa là α .

Chúng ta xét các trường hợp sau:

4.1. KIỂM ĐỊNH MỘT PHÍA

4.1.1 TRƯỜNG HỢP 1

Lập các giả thiết:

Giả thiết không

$H_0 : \sigma^2 = \sigma_0^2$ hay $H_0 : \sigma^2 \le \sigma_0^2$

Giả thiết đối

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

σ_0^2 là phương sai của tổng thể được biết dựa theo một nguồn tin nào đó.

n là kích thước mẫu

Chọn thống kê kiểm định là: $\chi^2 = \dfrac{(n-1)\hat{s}^2}{\sigma^2}$

Thì ĐLNN χ^2 có phân phối chi bình phương với bậc tự do là k=n-1. $\chi^2 \sim \chi^2(n-1)$

Vì giả sử giả thiết H_0 đúng, nghĩa là $\sigma^2 = \sigma_0^2$

Do đó ĐLNN $\chi^2 = \dfrac{(n-1)\hat{s}^2}{\sigma_0^2} \sim \chi^2(n-1)$

Ta có: $P(\chi^2 > \chi_\alpha^2) = \alpha \Rightarrow \chi_\alpha^2$

Từ mẫu cụ thể tính: $\chi^2 = \dfrac{(n-1)\hat{s}^2}{\sigma_0^2}$

Miền bác bỏ: $W = \left\{ \chi^2 / \chi^2 > \chi_\alpha^2 \right\}$

Quy tắc kiểm định

- Nếu $\chi^2 > \chi_\alpha^2$ thì bác bỏ giả thiết H_0
- Nếu $\chi^2 \le \chi_\alpha^2$ thì chưa đủ cơ sở để bác bỏ giả thiết H_0

4.1.2 TRƯỜNG HỢP 2

Lập các giả thiết:

Giả thiết không

$H_0 : \sigma^2 = \sigma_0^2$ hay $H_0 : \sigma^2 \ge \sigma_0^2$

Giả thiết đối

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

Chọn thống kê kiểm định là $\chi^2 = \dfrac{(n-1)\hat{s}}{\sigma^2}$

Thì ĐLNN χ^2 có phân phối chi bình phương với bậc tự do là k=n-1. $\chi^2 \sim \chi^2(n-1)$

Vì giả sử giả thiết H_0 đúng, nghĩa là $\sigma^2 = \sigma_0^2$

Do đó ĐLNN $\chi^2 = \dfrac{(n-1)\hat{s}^2}{\sigma_0^2} \sim \chi^2(n-1)$

Ta có: $P(\chi^2 < \chi^2_{1-\alpha}) = \alpha \Rightarrow \chi^2_{1-\alpha}$

Từ mẫu cụ thể tính: $\chi^2 = \dfrac{(n-1)\hat{s}^2}{\sigma_0^2}$

Miền bác bỏ $W = \{\ \chi^2 / \chi^2 < \chi^2_{1-\alpha}\ \}$

Quy tắc kiểm định

- ❖ Nếu $\chi^2 < \chi^2_{1-\alpha}$ thì bác bỏ giả thiết H_0
- ❖ Nếu $\chi^2 \geq \chi^2_{1-\alpha}$ thì chưa đủ cơ sở để bác bỏ giả thiết H_0

4.2. KIỂM ĐỊNH HAI PHÍA

Lập các giả thiết:

Giả thiết không

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

Giả thiết đối

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

Chọn thống kê kiểm định $\chi^2 = \dfrac{(n-1)\hat{s}^2}{\sigma^2}$

Thì ĐLNN χ^2 có phân phối chi bình phương với bậc tự do là k=n-1. $\chi^2 \sim \chi^2(n-1)$

Vì giả sử giả thiết H_0 đúng, nghĩa là $\sigma^2 = \sigma_0^2$

Do đó ĐLNN $\chi^2 = \dfrac{(n-1)\hat{s}^2}{\sigma_0^2} \sim \chi^2(n-1)$

Ta có: $P(\chi^2 > \chi^2_{\frac{\alpha}{2}}) = \dfrac{\alpha}{2} \Rightarrow \chi^2_{\frac{\alpha}{2}}$; $P(\chi^2 < \chi^2_{1-\frac{\alpha}{2}}) = \dfrac{\alpha}{2} \Rightarrow \chi^2_{1-\frac{\alpha}{2}}$

Từ mẫu cụ thể tính: $\chi^2 = \dfrac{(n-1)\hat{s}^2}{\sigma_0^2}$

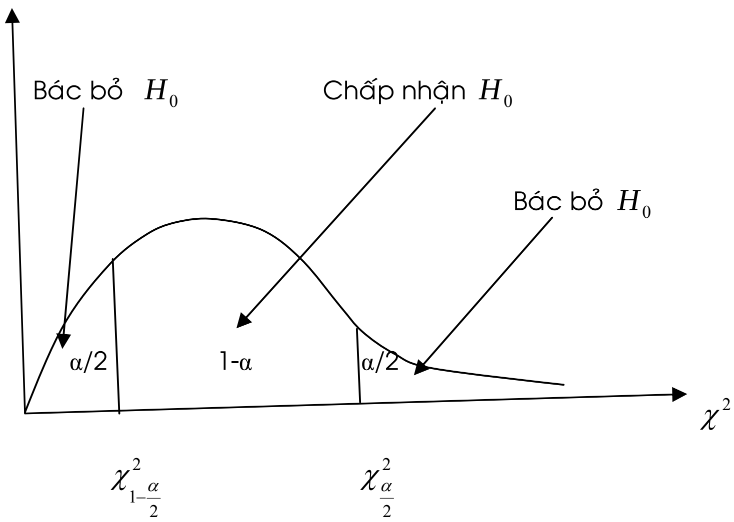
Miền bác bỏ $W = \{\ \chi^2 / \chi^2 < \chi^2_{1-\frac{\alpha}{2}}$ hay $\chi^2 > \chi^2_{\frac{\alpha}{2}}\ \}$

Quy tắc kiểm định

- ❖ Nếu $\chi^2 < \chi^2_{1-\frac{\alpha}{2}}$ hay $\chi^2 > \chi^2_{\frac{\alpha}{2}}$ thì bác bỏ giả thiết H_0

- Nếu $\chi^2_{1-\frac{\alpha}{2}} < \chi^2 < \chi^2_{\frac{\alpha}{2}}$ thì chưa đủ cơ sở để bác bỏ giả thiết H_0

HÌNH VẼ



VD:

Tại một đại lý bán vé máy bay thời gian một khách hàng chờ đợi mua vé là một đại lượng ngẫu nhiên có phân phối chuẩn. Được biết trước đây thời gian chờ đợi trung bình là 5 phút và độ lệch chuẩn là 3 phút.Sau một thời gian cải tiến cách phục vụ nhằm làm giảm độ lệch chuẩn, theo dõi 25 khách hàng thấy thời gian chờ đợi là 4,5 phút với độ lệch chuẩn có hiệu chỉnh là 2,5 phút. Với mức ý nghĩa 5%, hãy xét xem độ lệch chuẩn của thời gian phuc vụ đã giảm được chưa ?

GIẢI:

Gọi X(phút) là thời gian chờ đợi mua vé của một khách hàng

σ là độ lệch chuẩn của thời gian chờ đợi sau khi cải tiến cách làm việc

σ_0 là độ lệch chuẩn của thời gian chờ đợi trước khi cải tiến cách làm việc

Lập các giả thiết:

Giả thiết không

$$H_0 : \sigma^2 = \sigma_0^2 = (3)^2$$

Giả thiết đối

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

$\hat{s}$ là độ lệch chuẩn hiệu chỉnh của mẫu = 2,5

α là mức ý nghĩa = 0,05

Chọn thống kê kiểm định là $\chi^2 = \dfrac{(n-1)\hat{s}}{\sigma_0^2}$

Thì χ^2 có phân phối chi bình phương với bậc tự do là k=n-1=24

Ta có:

$$P(\chi^2 < \chi^2_{1-\alpha}) = \alpha \Leftrightarrow 1 - P(\chi^2 > \chi^2_{1-\alpha}) = \alpha \Leftrightarrow$$
$$\Leftrightarrow P(\chi^2 > \chi^2_{1-\alpha}) = 1-\alpha \Rightarrow P(\chi^2 > \chi^2_{0,95}) = 0{,}95 \Rightarrow \chi^2_{0,95} = 13{,}85$$

Từ mẫu cụ thể tính: $\chi^2 = \dfrac{(n-1)\hat{s}^2}{\sigma_0^2} = \dfrac{24(2{,}5)^2}{3^2} = 16{,}67$

Suy ra $\chi^2 > \chi^2_{1-\alpha}$

Vậy chưa đủ cơ sở để bác bỏ giả thiết H_0, nghĩa là sau thời gian cải tiến cách làm việc, độ lệch chuẩn của thời gian chờ đợi chưa giảm với mức ý nghĩa 5%.

VD:

Một hãng điện tử sử dụng công nghệ mới trong sản xuất đèn hình TV.Kiểm tra 20 bóng đèn hình TV loại mới thấy tuổi thọ trung bình là 10 ngàn giờ và độ lệch chuẩn hiệu chỉnh là 100 giờ.Tuổi thọ trung bình của đèn hình TV loại cũ là 8 ngàn giờ và độ lệch chuẩn là 160 giờ. Với mức ý nghĩa 1% xét xem độ lệch chuẩn của tuổi thọ đèn hình loại mới có giảm so với đèn hình loại cũ không? Cho biết tuổi thọ của đèn hình có phân phối chuẩn.

GIẢI:

Gọi X(giờ) là tuổi thọ của đèn hình TV

σ là độ lệch chuẩn của tuổi thọ đèn hình TV lọai mới

σ_0 là độ lệch chuẩn của tuổi thọ đèn hình TV loại cũ

Lập các giả thiết:

Giả thiết không

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

Giả thiết đối

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

Chọn thống kê kiểm định $\chi^2 = \dfrac{(n-1)\hat{s}^2}{\sigma_0^2}$

Ta có: χ^2 có phân phối chi bình phương với bậc tự do là k=n-1= 20-1=19

$$P(\chi^2 < \chi^2_{1-\alpha}) = \alpha \Leftrightarrow P(\chi^2 > \chi^2_{1-\alpha}) = 1-\alpha = 0{,}99 \Rightarrow \chi^2_{1-\alpha} = 7{,}63$$

Từ mẫu cụ thể tính: $\chi^2 = \dfrac{(n-1)\hat{s}^2}{\sigma^2} = \dfrac{19.100^2}{160^2} = 7{,}42$

Suy ra $\chi^2 < \chi^2_{1-\alpha}$

Vậy bác bỏ giả thiết H_0, nghĩa là độ lệch chuẩn của tuổi thọ đèn hình loại mới có giảm so với loại cũ.

5. PHƯƠNG PHÁP P_GIÁ TRỊ

5.1. ĐỊNH NGHĨA:

Mức ý nghĩa nhỏ nhất mà chúng ta có thể bác bỏ giả thiết H_0 được gọi là P_giá trị.

5.2. QUY TẮC KIỂM ĐỊNH BẰNG PHƯƠNG PHÁP P_GIÁ TRỊ

i) TRƯỜNG HỢP ĐÃ BIẾT MỨC Ý NGHĨA

- Nếu P_giá trị $\leq \alpha$ thì bác bỏ giả thiết H_0
- Nếu P_giá trị $> \alpha$ thì chưa có cơ sở để bác bỏ giả thiết H_0

ii) TRƯỜNG HỢP CHƯA BIẾT MỨC Ý NGHĨA

- Nếu P_giá trị > 0,05 thì không đủ cơ sở để bác bỏ giả thiết H_0
- Nếu 0,01< P_giá trị < 0,05 thì có đủ cơ sở để bác bỏ giả thiết H_0
- Nếu P_giá trị < 0,01 thì có cơ sở rất mạnh để bác bỏ giả thiết H_0

5.3 XÁC ĐỊNH P_GIÁ TRỊ

Xét bài toán kiểm định giả thiết trung bình.

Chúng ta xét các trường hợp sau:

i) TRƯỜNG HỢP 1

Lập các giả thiết:

Giả thiết không

$$H_0 : \mu = \mu_0$$

Giả thiết đối

$$H_1 : \mu > \mu_0$$

$\overline{X}$ trung bình mẫu ngẫu nhiên

$\overline{x}$ trung bình mẫu cụ thể

Trong trường hợp giả sử giả thiết H_0 đúng,

Ta có P_giá trị = P$(\overline{X} \geq \overline{x})$

- Nếu n $\geq$ 30 sử dụng phân phối chuẩn tính P_giá trị.

$$P(\overline{X} \geq \overline{x}) = P\left(\frac{(\overline{X} - \mu_0)\sqrt{n}}{\sigma} \geq \frac{(\overline{x} - \mu_0)\sqrt{n}}{\sigma}\right) = P\left(Z \geq \frac{(\overline{x} - \mu_0)\sqrt{n}}{\sigma}\right)$$

- Nếu n< 30, X có phân phối chuẩn và chưa biết σ thì sử dụng phân phối Student để tính P_giá trị.

ii) TRƯỜNG HỢP 2

Lập các giả thiết:

Giả thiết không

$$H_0 : \mu = \mu_0$$

Giả thiết đối

$$H_1 : \mu < \mu_0$$

Ta có P_giá trị = P($\overline{X} \leq \overline{x}$)

iii) TRƯỜNG HỢP 3

Lập các giả thiết:

Giả thiết không

$$H_0 : \mu = \mu_0$$

Giả thiết đối

$$H_1 : \mu \neq \mu_0$$

Ta có P_giá trị = 2.P($\overline{X} \leq \overline{x}$)

CHÚ Ý:

Bài toán kiểm định tỷ lê lập luận tương tự.

VD:

Một hãng sản xuất điện thoại di động quảng cáo thời gian trung bình sử dụng giữa hai lần sạc pin của loại điện thoại A là 72 giờ. Kiểm tra 36 điện thoại thấy thời gian sử dụng là 70 giờ và độ lệch chuẩn hiệu chỉnh là 6 giờ. Với mức ý nghĩa 1%, sử dụng phương pháp P_giá trị kiểm định xem quảng cáo có đáng tin cậy không?

GIẢI:

Lập các giả thiết:

Giả thiết không

$$H_0 : \mu = \mu_0 = 72$$

Giả thiết đối

$$H_1 : \mu < \mu_0$$

Ta có

P_giá trị = P$(\overline{X} \leq \overline{x})$

CHÚ Ý: từ $n \geq 30$ suy ra $Z = \dfrac{(\overline{X} - \mu)\sqrt{n}}{\sigma} \sim N(0,1)$

Vì σ chưa biết thay bởi $\hat{s}$ và từ giả sử giả thiết H_0 đúng thay $\mu = \mu_0$,

Suy ra $Z = \dfrac{(\overline{X} - \mu_0)\sqrt{n}}{\hat{s}} \sim N(0,1)$

Do đó $P(\overline{X} \le \overline{x}) = P(\overline{X} \le 70) = P(\dfrac{(\overline{X} - \mu_0)\sqrt{n}}{\hat{s}} \le \dfrac{(70-72)\sqrt{36}}{6}) = P(Z \le -2) = 0{,}0228$

Vậy P_giá trị = 0,0228 ; mà α= 0,01

Suy ra P_giá trị > α, vậy chưa có cơ sở để bác bỏ giả thiết H_0, nghĩa là quảng cáo đáng tin cậy với mức ý nghĩa $\alpha = 1\%$.

VD:

Theo ban giám đốc công ty A thu nhập trung bình của công nhân công ty là 2 triệu đồng/tháng. Khảo sát thu nhập của 25 công nhân của công ty thấy thu nhập trung bình là 1,8 triệu đồng/tháng và độ lệch chuẩn có hiệu chỉnh là 0,2 triệu đồng.Với mức ý nghĩa 2%, hãy sử dụng phương pháp P_giá trị xét xem nguồn tin từ ban giám đốc có đáng tin cậy không?

Cho biết thu nhập của công nhân có phân phối chuẩn.

GIẢI:

Gọi X(triệu đồng) là thu nhập của công nhân

μ_0 thu nhập trung bình của công nhân công ty A theo ban giám đốc

μ thu nhập trung bình của công nhân công ty A

Lập các giả thiết

$$H_0 : \mu = \mu$$

$$H_1 : \mu < \mu_0$$

Ta có P_giá trị = P($\overline{X} < \overline{x}$)

CHÚ Ý: n=25 < 30 .

X có phân phối chuẩn và σ chưa biết, do đó sử dụng phân phối Student với bậc tự do là k=n-1

$$T = \frac{(\overline{X} - \mu)\sqrt{n}}{\hat{s}} \sim T(n-1)$$

Vì giả sử giả thiết H_0 đúng, nên thay $\mu = \mu_0$

Suy ra $T = \dfrac{(\overline{X} - \mu_0)\sqrt{n}}{\hat{s}}$ ~T(n-1)

Sử dụng phân phối Student với bậc tự do là k=n-1=24

Ta có: P_giá trị=P$(T < \frac{(\overline{x}-\mu_0)\sqrt{n}}{\hat{s}}) = P(T < -5)$

$\Rightarrow$P_giá trị = 0,00002 < α =0,02

Vậy bác bỏ giả thiết H_0, nghĩa là nguồn tin từ ban giám đốc không đáng tin cậy với mức ý nghĩa 2%.

VD:

Sở giáo dục một địa phương cho biết tỷ lệ học sinh bỏ học là 15%, khảo sát 121 học sinh thì có 25 học sinh bỏ học. Với mức ý nghĩa 2%, sử dụng phương pháp P_giá trị kiểm định xem báo cáo của sở giáo dục có chấp nhận được không?

GIẢI:

p là tỷ lệ học sinh bỏ học tại địa phương.

p_0 là tỷ lệ học sinh bỏ học theo báo cáo của sở giáo dục.

Lập các giả thiết

$$H_0 : p = p_0$$

$$H_1 : p > p_0$$

Ta có

P_giá trị = P($(f \geq \frac{25}{121}) = P(\frac{(f-p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}} \geq \frac{(0{,}207-0{,}15)\sqrt{121}}{\sqrt{0{,}15.0{,}85}})$ = P(Z $\geq$1,85) = 0,0322

(CHÚ Ý: $Z = \frac{(f-p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}} \sim N(0,1)$)

Mà α = 0,02

Suy ra P_giá trị > α

Vậy bác bỏ giả thiết H_0, nghĩa là nguồn tin từ sở giáo dục không đáng tin cậy với mức ý nghĩa 2%.

6. KIỂM ĐỊNH HIỆU CỦA HAI TRUNG BÌNH

6.1. MẪU CỦA n CẶP QUAN SÁT

Giả sử chúng ta có một mẫu ngẫu nhiên gồm n cặp quan sát từ tổng thể có trung bình là $\mu_X ; \mu_Y$.

Gọi d và s_d^2 lần lượt là trung bình mẫu và phương sai hiệu chỉnh của n giá trị $(x_i - y_i)$, Nếu tổng thể của các hiệu số trên có phân phối chuẩn.

Với mức ý nghĩa α.

Chúng ta xét các trường hợp sau:

i) TRƯỜNG HỢP 1:

Lập các giả thiết :

Giả thiết không

$$H_0 : \mu_X - \mu_Y = 0$$

Giả thiết đối

$$H_1 : \mu_X - \mu_Y > 0$$

Chọn thống kê kiểm định : $T = \dfrac{d\sqrt{n}}{s_d}$

Thì ĐLNN T có phân phối Student với bậc tự do là k=n-1

Ta có: $P(T > t_\alpha) = \alpha \Rightarrow t_\alpha$

Từ mẫu cụ thể tính: $t = \dfrac{d\sqrt{n}}{s_d}$

Quy tắc kiểm định

- ❖ Nếu $t > t_\alpha$ thì bác bỏ giả thiết H_0
- ❖ Nếu $t \le t_\alpha$ thì chưa đủ cơ sở để bác bỏ giả thiết H_0

ii) TRƯỜNG HỢP 2:

Lập các giả thiết:

Giả thiết không

$$H_0 : \mu_X - \mu_Y = 0$$

Giả thiết đối

$$H_1 : \mu_X - \mu_Y < 0$$

Chọn thống kê kiểm định: $T = \dfrac{d\sqrt{n}}{s_d}$

Thì ĐLNN T có phân phối Student với bậc tự do là k=n-1. T~T(n-1)

Ta có $P(T < -t_\alpha) = \alpha \Rightarrow -t_\alpha$

Từ mẫu cụ thể tính: $t = \dfrac{d\sqrt{n}}{s_d}$

Quy tắc kiểm định:

- ❖ Nếu $t < -t_\alpha$ thì bác bỏ giả thiết H_0
- ❖ Nếu $t \ge -t_\alpha$ thì chưa đủ cơ sở để bác bỏ giả thiết H_0

iii) TRƯỜNG HỢP 3:

Lập các giả thiết:

Giả thiết không

$$H_0 : \mu_X - \mu_Y = 0$$

Giả thiết đối

$$H_1 : \mu_X - \mu_Y \neq 0$$

Chọn thống kê kiểm định: $T = \dfrac{d\sqrt{n}}{s_d}$

Thì ĐLNN T có phân phối Student với bậc tự do là k=n-1. T~T(n-1)

Ta có: $P(T < -t_{\frac{\alpha}{2}}) = \dfrac{\alpha}{2} \Rightarrow -t_{\frac{\alpha}{2}}$; $P(T > t_{\frac{\alpha}{2}}) = \dfrac{\alpha}{2} \Rightarrow t_{\frac{\alpha}{2}}$

Từ mẫu cụ thể tính: $t = \dfrac{d\sqrt{n}}{s_d}$

Quy tắc kiểm định:

- Nếu $t < -t_{\frac{\alpha}{2}}$ hay $t > t_{\frac{\alpha}{2}}$ thì bác bỏ giả thiết H_0.
- Nếu $-t_{\frac{\alpha}{2}} \leq t \leq t_{\frac{\alpha}{2}}$ thì chưa đủ cơ sở để bác bỏ giả thiết H_0.

VD:

Doanh số của 6 cửa hàng của một tập đoàn bán lẻ trước và sau khi quãng cáo có số liệu như sau:

Trước khi quãng cáo (đơn vị triệu đồng)

620 600 640 630 570 600

Sau khi quãng cáo

660 620 670 620 580 630

Với mức ý nghĩa 1%, hãy kiểm định xem chiến dịch quãng cáo có hiệu quả không?

GIẢI

X(triệu đồng) là doanh thu của mỗi cửa hàng trước khi quãng cáo

Y là doanh thu của mỗi cửa hàng sau khi quãng cáo

μ_X là doanh thu trung bình của các cửa hàng trước khi quãng cáo

μ_Y là doanh thu trung bình của các cửa hàng sau khi quãng cáo

Lập các giả thiết

$$H_0 : \mu_X - \mu_Y = 0$$

$$H_1 : \mu_X - \mu_Y < 0$$

Chọn thống kê kiểm định : $T = \dfrac{d\sqrt{n}}{s_d}$

Thì ĐLNN T có phân phối Student với bậc tự do là k=n-1=6-1=5

Ta có

$$d = \frac{\sum_{i=1}^{6}(x_i - y_i)}{n} = -12$$

$$d_i = x_i - y_i$$

$$s_d^2 = \frac{\sum_{i=1}^{6}(d_i - d)^2}{n-1} = 396{,}8 \Rightarrow s_d = 19{,}92$$

Tính $t = \dfrac{(-12)\sqrt{6}}{19{,}92} = -1{,}4756$

Sử dụng phân phối Student với bậc tự do là k=n-1=6-1=5

Ta có $P(T < -t_\alpha) = \alpha = 0{,}01 \Rightarrow -t_\alpha = -3{,}365$

Suy ra $t > -t_\alpha$

Vậy chưa đủ cơ sở để bác bỏ giả thiết H_0, nghĩa là chiến dịch quãng cáo chưa mang lại hiệu quả với mức ý nghĩa 1%.

6.2. HAI MẪU ĐỘC LẬP

PHƯƠNG SAi σ^2 ĐÃ BIẾT HAY KÍCH THƯỚC MẪU LỚN

Giả sử chúng ta có hai mẫu ngẫu nhiên độc lập của $n_X; n_Y$ quan sát từ các tổng thể có phân phối chuẩn với trung bình $\mu_X; \mu_Y$ và phương sai $\sigma_X^2; \sigma_Y^2$.

Nếu mẫu cụ thể có trung bình là $\overline{x}; \overline{y}$

Với mức ý nghĩa α

Chúng ta xét các trường hợp sau:

6.2.1 KIỂM ĐỊNH MỘT PHÍA

i) TRƯỜNG HỢP 1

Lập các giả thiết:

Giả thiết không

$$H_0 : \mu_X - \mu_Y = 0$$

Giả thiết đối

$$H_1 : \mu_X - \mu_Y > 0$$

Chọn thống kê kiểm định: $Z = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{\sigma_X^2}{n_X} + \dfrac{\sigma_Y^2}{n_Y}}}$

Thì ĐLNN Z có phân phối chuẩn tắc. Z~N(0,1)

Ta có $P(Z > z_\alpha) = \alpha \Rightarrow z_\alpha$

Từ mẫu cụ thể tính: $z = \dfrac{\overline{x} - \overline{y}}{\sqrt{\dfrac{\sigma_X^2}{n_X} + \dfrac{\sigma_Y^2}{n_Y}}}$

Quy tắc kiểm định

- Nếu $z > z_\alpha$ thì bác bỏ giả thiết H_0
- Nếu $z \le z_\alpha$ thì chưa đủ cơ sở để bác bỏ giả thiết H_0

ii)TRƯỜNG HỢP 2

Lập các giả thiết:

Giả thiết không

$$H_0 : \mu_X - \mu_Y = 0$$

Giả thiết đối

$$H_1 : \mu_X - \mu_Y < 0$$

Chọn thống kê kiểm định: $Z = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{\sigma_X^2}{n_X} + \dfrac{\sigma_Y^2}{n_Y}}}$

Thì ĐLNN Z có phân phối chuẩn tắc. Z~N(0,1)

Ta có: $P(Z < -z_\alpha) = \alpha \Rightarrow -z_\alpha$

Từ mẫu cụ thể tính: $z = \dfrac{\overline{x} - \overline{y}}{\sqrt{\dfrac{\sigma_X^2}{n_X} + \dfrac{\sigma_Y^2}{n_Y}}}$

Quy tắc kiểm định

- Nếu $z < -z_\alpha$ thì bác bỏ giả thiết H_0
- Nếu $z \ge -z_\alpha$ thì chưa đủ cơ sở để bác bỏ giả thiết H_0

6.2.2 KIỂM ĐỊNH HAI PHÍA

Lập các giả thiết:

Giả thiết không

$$H_0 : \mu_X - \mu_Y = 0$$

Giả thiết đối

$$H_1 : \mu_X - \mu_Y \neq 0$$

Chọn thống kê kiểm định: $Z = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{\sigma_X^2}{n_X} + \dfrac{\sigma_Y^2}{n_Y}}}$

Thì ĐLNN Z có phân phối chuẩn tắc. Z~N(0,1)

Ta có: $P(|Z| > z_{\frac{\alpha}{2}}) = \alpha \Rightarrow z_{\frac{\alpha}{2}}$

Từ mẫu cụ thể tính: $z = \dfrac{\overline{x} - \overline{y}}{\sqrt{\dfrac{\sigma_X^2}{n_X} + \dfrac{\sigma_Y^2}{n_Y}}}$

Quy tắc kiểm định

- Nếu $|z| > z_{\frac{\alpha}{2}}$ thì bác bỏ giả thiết H_0
- Nếu $|z| \leq z_{\frac{\alpha}{2}}$ thì chưa có cơ sở để bác bỏ giả thiết H_0

CHÚ Ý:

Nếu kích thước mẫu $n_X ; n_Y$ lớn (≥ 30) thì có thể thay thế $\sigma_X^2 ; \sigma_Y^2$ bởi các phương sai mẫu hiệu chỉnh $s_X^2 ; s_Y^2$.

Trong trường hợp kích thước mẫu lớn, chúng ta cũng có thể xấp xỉ phân phối chuẩn nếu tổng thể không có phân phối chuẩn.

VD:

Trong tháng 11 và 12 theo dõi giá cổ phiếu của hai ngân hàng A và B trong 31 ngày nhận thấy như sau:

Giá cổ phiếu trung bình của ngân hàng A là 27,58 (ngàn đồng) và độ lệch tiêu chuẩn của mẫu là 1,50 (ngàn đồng)

Giá cổ phiếu trung bình của ngân hàng B là 28,24 (ngàn đồng) và độ lệch chuẩn của mẫu là 2,20 (ngàn đồng)

Với mức ý nghĩa 5%, kiểm định xem giá cổ phiếu trung bình của hai ngân hàng có thực sự khác nhau không?

Cho biết giá cổ phiếu của hai ngân hàng trên có phân phối chuẩn.

GIẢI:

μ_X là giá cổ phiếu trung bình của ngân hàng A

μ_Y là giá cổ phiếu trung bình của ngân hàng B

Lập các giả thiết:

Giả thiết không

$$H_0 : \mu_X - \mu_Y = 0$$

Giả thiết đối

$$H_1 : \mu_X - \mu_Y \neq 0$$

Chọn thống kê kiểm định: $Z = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{\hat{s}_X^2}{n_X} + \dfrac{\hat{s}_Y^2}{n_Y}}}$

Thì Z~N(0,1)

Ta có: $P(|Z| > z_{\frac{\alpha}{2}}) = \alpha = 0{,}05 \Rightarrow z_{\frac{\alpha}{2}} = z_{0,025} = 1{,}96$

Từ mẫu cụ thể tính: $z = \dfrac{\overline{x} - \overline{y}}{\sqrt{\dfrac{\hat{s}_X^2}{n_X} + \dfrac{\hat{s}_Y^2}{n_Y}}} = -1{,}38$

Suy ra $|z| < z_{\frac{\alpha}{2}}$

Vậy chưa đủ cơ sở để bác bỏ giả thiết H_0,

nghĩa là giá cổ phiếu trung bình của hai ngân hàng trên không thực sự khác nhau với mức ý nghĩa 5%.

6.3 HAI MẪU ĐỘC LẬP CỦA HAI TỔNG THỂ CÙNG CÓ PHÂN PHỐI CHUẨN VẢ CÓ PHƯƠNG SAI BẰNG NHAU.

Giả sử chúng ta có hai mẫu ngẫu nhiên độc lập của $n_X ; n_Y$ quan sát từ hai tổng thể có phân phối chuẩn với trung bình $\mu_X ; \mu_Y$ và có cùng phương sai.

Nếu phương sai hiệu chỉnh của hai mẫu lần lượt là $\hat{s}_X^2 ; \hat{s}_Y^2$ thì một ước lượng của phương sai chung của hai mẫu là

$$s^2 = \frac{(n_X - 1)\hat{s}_X^2 + (n_Y - 1)\hat{s}_Y^2}{n_X + n_Y - 2}$$

Hai mẫu cụ thể có trung bình lần lượt là $\overline{x}; \overline{y}$

α là mức ý nghĩa.

Chúng ta xét các trường hợp sau:

6.3.1 KIỂM ĐỊNH MỘT PHÍA

i) TRƯỜNG HỢP 1

Lập các giả thiết:

Giả thiết không

$$H_0 : \mu_X - \mu_Y = 0$$

Giả thiết đối

$$H_1 : \mu_X - \mu_Y > 0$$

Chọn thống kê kiểm định: $T = \dfrac{\overline{X} - \overline{Y}}{s\sqrt{\dfrac{n_X + n_Y}{n_X n_Y}}}$

Thì ĐLNN T có phân phối Student với bậc tự do là $k = n_X + n_Y - 2$

Ta có: $P(T > t_\alpha) = \alpha \Rightarrow t_\alpha$

Từ mẫu cụ thể tính: $t = \dfrac{\overline{x} - \overline{y}}{s\sqrt{\dfrac{n_X + n_Y}{n_X n_Y}}}$

Quy tắc kiểm định

- ❖ Nếu $t > t_\alpha$ thì bác bỏ giả thiết H_0
- ❖ Nếu $t \le t_\alpha$ thì chưa đủ cơ sở để bác bỏ giả thiết H_0

ii)TRƯỜNG HỢP 2

Lập các giả thiết:

Giả thiết không

$$H_0 : \mu_X - \mu_Y = 0$$

Giả thiết đối

$$H_1 : \mu_X - \mu_Y < 0$$

Chọn thống kê kiểm định: $T = \dfrac{\overline{X} - \overline{Y}}{s\sqrt{\dfrac{n_X + n_Y}{n_X n_Y}}}$

Thì ĐLNN T có phân phối Student với bậc tự do là $k = n_X + n_Y - 2$

Ta có: $P(T < -t_\alpha) = \alpha \Rightarrow -t_\alpha$

Từ mẫu cụ thể tính: $t = \dfrac{\overline{x} - \overline{y}}{s\sqrt{\dfrac{n_X + n_Y}{n_X n_Y}}}$

Quy tắc kiểm định

- ❖ Nếu $t < -t_\alpha$ thì bác bỏ giả thiết H_0

- Nếu $t \geq -t_\alpha$ thì chưa đủ cơ sở để bác bỏ giả thiết H_0

6.3.2. KIỂM ĐỊNH HAI PHÍA

Lập các giả thiết:

Giả thiết không

$$H_0 : \mu_X - \mu_Y = 0$$

Giả thiết đối

$$H_1 : \mu_X - \mu_Y \neq 0$$

Chọn thống kê kiểm định: $T = \dfrac{\overline{X} - \overline{Y}}{s\sqrt{\dfrac{n_X + n_Y}{n_X n_Y}}}$

Thì ĐLNN T có phân phối Student với bậc tự do là $k = n_X + n_Y - 2$

Ta có: $P(|T| > t_{\frac{\alpha}{2}}) = \alpha \Rightarrow t_{\frac{\alpha}{2}}$

Từ mẫu cụ thể tính: $t = \dfrac{\overline{x} - \overline{y}}{s\sqrt{\dfrac{n_X + n_Y}{n_X n_Y}}}$

Quy tắc kiểm định

- Nếu $|t| > t_\alpha$ thì bác bỏ giả thiết H_0
- Nếu $|t| \leq t_\alpha$ thì chưa đủ cơ sở để bác bỏ giả thiết H_0

VD:

Một công ty có hai phân xưởng cùng sản xuất một loại linh kiện điện tử.

Kiểm tra 15 sản phẩm của phân xưởng A thấy tuổi thọ trung bình là 4 ngàn giờ và độ lệch chuẩn hiệu chỉnh là 200 giờ.

Kiểm tra 15 sản phẩm của phân xưởng B thấy tuổi thọ trung bình là 4,1 ngàn giờ và độ lệch chuẩn hiệu chỉnh là 220 giờ.

Với mức ý nghĩa 1%, kiểm định xem có sự khác nhau về tuổi thọ trung bình của sản phẩm do hai phân xưởng sản xuất không?

Cho biết tuổi thọ của sản phẩm do hai phân xưởng sản xuất có phân phối chuẩn và phương sai bằng nhau.

GIẢI:

μ_X là tuổi thọ trung bình của sản phẩm do phân xưởng A sản xuất

μ_Y là tuổi thọ trung bình của sản phẩm do phân xưởng B sản xuất

Lập các giả thiết:

Giả thiết không

$$H_0 : \mu_X - \mu_Y = 0$$

Giả thiết đối

$$H_1 : \mu_X - \mu_Y \neq 0$$

Phương sai chung của hai mẫu là:

$$s^2 = \frac{(15-1)(0,2)^2 + (15-1)(0,22)^2}{15+15-2} = 0,0442 \Rightarrow s = 0,21$$

Tính thống kê kiểm định

$$t = \frac{\overline{x} - \overline{y}}{s\sqrt{\frac{n_X + n_Y}{n_X n_Y}}} = \frac{4 - 4,1}{0,21\sqrt{\frac{15+15}{15.15}}} = 1,304$$

Sử dung phân phối Student với bậc tự do là:

$$k = n_X + n_Y - 2 = 15 + 15 - 2 = 28$$

Ta có:

$$P(|T| > t_{\frac{\alpha}{2}}) = \alpha = 0,01 \Rightarrow t_{\frac{\alpha}{2}} = 2,763$$

Suy ra: $|t| < t_{\frac{\alpha}{2}}$

Vậy chưa đủ cơ sở để bác bỏ giả thiết H_0, nghĩa là tuổi thọ trung bình của sản phẩm do hai phân xưởng sản xuất như nhau với mức ý nghĩa 1%.

7. KIỂM ĐỊNH HIỆU CỦA HAI TỶ LỆ

TRƯỜNG HỢP KÍCH THƯỚC MẪU LỚN (n≥30)

Giả sử f_X là tỷ lệ thành công trong mẫu ngẫu nhiên của n_X quan sát từ một tổng thể với tỷ lệ thành công là p_X.

f_Y là tỷ lệ thành công trong mẫu ngẫu nhiên độc lập của n_Y quan sát từ một tổng thể với tỷ lệ thành công là p_Y

Tỷ lệ chung của hai tỷ lệ mẫu

$$f = \frac{n_X f_X + n_Y f_Y}{n_X + n_Y}$$

α là mức ý nghĩa.

Chúng ta xét các trường hợp sau:

7.1 KIỂM ĐỊNH MỘt PHÍA

7.1.1 TRƯỜNG HỢP 1

Lập các giả thiết:

Giả thiết không

$$H_0 : p_X - p_Y = 0$$

Giả thiết đối

$$H_1 : p_X - p_Y > 0$$

Chọn thống kê kiểm định: $$Z = \frac{(f_X - f_Y) - (p_X - p_Y)}{\sqrt{\frac{p_X(1-p_X)}{n_X} + \frac{p_Y(1-p_Y)}{n_Y}}}$$

Thì ĐLNN Z có phân phối chuẩn tắc. Z~N(0,1)

Từ giả sử giả thiết H_0 đúng, suy ra

$$p_X = p_Y = p_0$$

Suy ra $$Z = \frac{(f_X - f_Y)}{\sqrt{p_0(1-p_0)(\frac{n_X + n_Y}{n_X n_Y})}} \sim N(0,1)$$

Ta có: $P(Z > z_\alpha) = \alpha \Rightarrow z_\alpha$

Từ mẫu cụ thể tính: $$z = \frac{(f_x - f_y)}{\sqrt{f(1-f)(\frac{n_x + n_y}{n_x n_y})}}$$

Quy tắc kiểm định:

- Nếu $z > z_\alpha$ thì bác bỏ giả thiết H_0
- Nếu $z \le z_\alpha$ thì chưa đủ cơ sở để bác bỏ giả thiết H_0

7.1.2. TRƯỜNG HỢP 2

Lập các giả thiết:

Giả thiết không

$$H_0 : p_X - p_Y = 0$$

Giả thiết đối

$$H_1 : p_X - p_Y < 0$$

Chọn thống kê kiểm định: $$Z = \frac{f_X - f_Y}{\sqrt{f(1-f)(\frac{n_X + n_Y}{n_X n_Y})}}$$

Thì ĐLNN Z có phân phối chuẩn tắc. Z~N(0,1)

Ta có: $P(Z < -z_\alpha) = \alpha \Rightarrow -z_\alpha$

Từ mẫu cụ thể tính: $z = \dfrac{f_x - f_y}{\sqrt{f(1-f)(\dfrac{n_x + n_y}{n_x n_y})}}$

Quy tắc kiểm định:

- Nếu $z < -z_\alpha$ thì bác bỏ giả thiết H_0
- Nếu $z \ge -z_\alpha$ thì chưa đủ cơ sở để bác bỏ giả thiết H_0

7.2. KIỂM ĐỊNH HAI PHÍA

Lập các giả thiết:

Giả thiết không

$$H_0 : p_X - p_Y = 0$$

Giả thiết đối

$$H_1 : p_X - p_Y \neq 0$$

Chọn thống kê kiểm định: $Z = \dfrac{f_X - f_Y}{\sqrt{f(1-f)(\dfrac{n_X + n_Y}{n_X n_Y})}}$

Thì ĐLNN Z có phân phối chuẩn tắc. Z~N(0,1)

Ta có: $P(|Z| > z_{\frac{\alpha}{2}}) = \alpha \Rightarrow z_{\frac{\alpha}{2}}$

Từ mẫu cụ thể tính: $z = \dfrac{f_x - f_y}{\sqrt{f(1-f)(\dfrac{n_x + n_y}{n_x n_y})}}$

Quy tắc kiểm định

- Nếu $z > z_{\frac{\alpha}{2}}$ thì bác bỏ giả thiết H_0
- Nếu $|z| \le z_{\frac{\alpha}{2}}$ thì chưa đủ cơ sở để bác bỏ giả thiết H_0

VD:

Tại một địa phương trước ngày bầu cử, một hãng thông tấn thăm dò 160 cử tri có độ tuổi ≥ 50 có 80 cử tri sẽ bầu cho ứng cử viên B, thăm dò 220 cử tri có độ tuổi dưới 50 có 120 cử tri sẽ bầu cho ứng cử viên B. Với mức ý nghĩa 2% kiểm định xem tỷ lệ cử tri của hai độ tuổi bầu cho ứng cử viên B có khác biệt nhau không?

GIẢI:

p_X là tỷ lệ cử tri ≥ 50 sẽ bầu cho ứng cử viên B

p_Y là tỷ lệ cử tri dưới 50 sẽ bầu cho ứng cử viên B

Lập các giả thiết:

$$H_0 : p_X - p_Y = 0$$

$$H_1 : p_X - p_Y \neq 0$$

Chọn thống kê kiểm định

$$Z = \frac{f_X - f_Y}{\sqrt{f(1-f)(\frac{n_X + n_Y}{n_X n_Y})}} \sim N(0,1)$$

Tỷ lệ chung của hai mẫu

$$f = \frac{n_X f_X + n_Y f_Y}{n_X + n_Y} = \frac{10}{19} = 0,53$$

Từ mẫu cụ thể tính $z = \dfrac{f_x - f_y}{\sqrt{f(1-f)(\frac{n_x + n_y}{n_x n_y})}}$ = - 0,12

Ta có: $P(|Z| > z_{\frac{\alpha}{2}}) = 0,2 \Rightarrow z_{\frac{\alpha}{2}} = 2,58$

Suy ra $|z| < z_{\frac{\alpha}{2}}$

Vậy chưa đủ cơ sở để bác bỏ giả thiết H_0, nghĩa là không có sự khác biệt về tỷ lệ cử tri của hai độ tuổi sẽ bầu cho ứng cử viên B với mức ý nghĩa 2%.

VD:

Một công ty dược phẩm cho biết sẽ đưa ra thị trường mợt loại thuốc mới A, qua thử nghiệm trên 100 bệnh nhân bị bệnh K có 70 bệnh nhân được lành bệnh khi sử dụng loại thuốc A, trong khi đó cho 100 bệnh nhân bị bệnh K sử dụng loại thuốc B có 60 bệnh nhân được lành bệnh. Với mức ý nghĩa 5%, kiểm định xem có thực sự loại thuốc A hiệu quả hơn loại thuốc B không?

GIẢI:

p_X là tỷ lệ bệnh nhân lành bệnh khi sử dụng thuốc A

p_Y là tỷ lệ bệnh nhân lành bệnh khi sử dụng thuốc B

Lập các giả thiết:

$$H_0 : p_X - p_Y = 0$$

$$H_1 : p_X > p_Y$$

Tính tỷ lệ chung của hai mẫu

$$f = \frac{n_X f_X + n_Y f_Y}{n_X + n_Y} = 0{,}65$$

Tính $$z = \frac{f_x - f_y}{\sqrt{f(1-f)(\frac{n_x + n_y}{n_x n_y})}} = 1{,}84$$

Sử dụng phân phối chuẩn, ta có:

$$P(Z > z_\alpha) = \alpha = 0{,}05 \Rightarrow z_\alpha = 1{,}65$$

Suy ra $z > z_\alpha$

Vậy bác bỏ giả thiết H_0, nghĩa là tỷ lệ bệnh nhân lành bệnh khi sử dụng thuốc A cao hơn khi sử dụng thuốc B (thuốc A hiệu quả hơn).

VD:

Ban giám đốc chuỗi cửa hàng bán lẻ của một tập đoàn kim khí điện máy cho biết tỷ lệ khách hàng mua sản phẩm LCD tại địa phương A cao hơn địa phương B trên 10%. Để kiểm tra kết luận này, chọn ngẫu nhiên 198 khách hàng tại địa phương A có 189 khách hàng mua LCD, chọn ngẫu nhiên 210 khách hàng tại địa phương B có 158 khách hàng mua LCD.

Với mức ý nghĩa 1%, kiểm định xem nguồn tin từ ban giám đốc có đáng tin cậy không?

GIẢI:

p_X là tỷ lệ khách hàng tại địa phương A mua LCD

p_Y là tỷ lệ khách hàng tại địa phương B mua LCD

$$f_X = \frac{189}{198} = 0{,}9545$$

$$f_Y = \frac{158}{210} = 0{,}7524$$

Lập các giả thiết

$$H_0 : p_X - p_Y = \delta = 0{,}1$$

$$H_1 : p_X - p_Y > \delta$$

Chọn thống kê kiểm định

$$Z = \frac{f_X - f_Y - \delta}{\sqrt{\frac{f_X(1-f_X)}{n_X} + \frac{f_Y(1-f_Y)}{n_Y}}} \sim N(0,1)$$

Ta có: $P(Z > z_\alpha) = \alpha = 0{,}01 \Rightarrow z_\alpha = 2{,}33$

Tính $$z = \frac{\frac{189}{198} - \frac{158}{210} - 0,1}{\sqrt{\frac{\frac{189}{198}.\frac{9}{198}}{198} + \frac{\frac{158}{210}.\frac{52}{210}}{210}}} = 3,07$$

Suy ra $z > z_\alpha$

Vậy bác bỏ giả thiết H_0, nghĩa là nguồn tin từ ban giám đốc đáng tin cậy với mức ý nghĩa 1%.

8. SO SÁNH NHIỀU TỶ LỆ

Giả sử chúng ta có k tổng thể

p_i là tỷ lệ phần tử loại A của tổng thể thứ i, i=1,2…,k.

Chúng ta muốn kiểm định giả thiết sau:

$$H_0 : p_1 = p_2 = ... = p_k$$

Từ mỗi tổng thể chọn ngẫu nhiên một mẫu ngẫu nhiên có kích thước n_i.

Trong mỗi mẫu có m_i phần tử loại A và $l_i = n_i - m_i$ phần tử không phải loại A.

$$\sum_{i=1}^{k} n_i = m + l = N \qquad \sum_{i=1}^{k} m_i = m$$

$$\sum_{i=1}^{k} l_i = l \qquad n_i = m_i + l_i$$

Giả sử giả thiết H_0 đúng

$$H_0 : p_1 = p_2 = ... = p_k = p$$

- p được ước lượng bởi $p_0 = \frac{m}{N}$

Chọn thống kê kiểm định :

$$\chi^2 = \frac{N^2}{m.l}\sum_{i=1}^{k}\frac{m_i^2}{n_i} - N.\frac{m}{l}$$

Thì ĐLNN χ^2 có phân phối chi bình phương với bậc tự do là : k-1

α là mức ý nghĩa.

Ta có: $P(\chi^2 > \chi^2_\alpha) = \alpha \Rightarrow \chi^2_\alpha$

Từ mẫu cụ thể tính: $\chi^2 = \frac{N^2}{ml}\sum_{i=1}^{k}\frac{m_i^2}{n_i} - N\frac{m}{l}$

Quy tắc kiểm định:

- Nếu $\chi^2 > \chi^2_\alpha$ thì bác bỏ giả thiết H_0
- Nếu $\chi^2 \le \chi^2_\alpha$ thì chưa đủ cơ sở để bác bỏ giả thiết H_0

VD:

Tại một địa phương người ta tiến hành một cuộc khảo để tìm hiểu phương tiện giao thông mà các công nhân viên chức sử dụng khi đi làm việc.Công việc khảo sát được tiến hành trên hai nhóm phụ nữ và nam giới. Kết quả thu được như sau:

Phương tiện	Xe máy	Xe buýt	Xe đạp
Nữ	125	100	25
Nam	205	120	75

Với mức ý nghĩa 1%, kiểm định xem có sự khác nhau về nhu cầu sử dụng phương tiện giao thông để đi làm giữa hai nhóm phụ nử và nam giới.

GIẢI:

Lập các giả thiết

H_0 :tỷ lệ sử dụng các phương tiện giao thông của hai nhóm là như nhau

H_1 :tỷ lệ sử dụng các phương tiện giao thông của hai nhóm là khác nhau

Chọn thống kê kiểm định là

$$\chi^2 = \frac{N^2}{ml}\sum_{i=1}^{3}\frac{m_i^2}{n_i} - N\frac{m}{l} \sim \chi^2(k-1=2)$$

Với: N=650 m=250 l=400

$m_i = 125;\ 100;\ 25$

$l_i = 205;\ 120;\ 75$

$n_i = 325;\ 220;\ 100$

Tính $\chi^2 = \dfrac{650^2}{250.400}\left[\dfrac{125^2}{325} + \dfrac{100^2}{220} + \dfrac{25^2}{100}\right] - \dfrac{650.250}{400} = 15{,}3267$

Sử dụng phân phối chi bình phương với bậc tự do là: 2

Từ $P(\chi^2 > \chi^2_\alpha) = 0{,}01 \Rightarrow \chi^2_\alpha = 9{,}21$

Suy ra $\chi^2 > \chi^2_\alpha$

Vậy bác bỏ giả thiết H_0 , nghĩa là có sự khác nhau về tỷ lệ sử dụng phương tiện giao thông giữa hai nhóm nam nữ.

VD:

Một công ty may sản xuất áo khoác với 4 màu: trắng, xanh, đỏ, vàng. Khảo sát 209 khách hàng nữ và 454 khách hàng nam về sở thích màu sắc của áo được số liệu sau:

Màu áo	trắng	xanh	đỏ	vàng
Nữ	42	34	62	71
Nam	54	223	125	52

Với mức ý nghĩa 1%, hãy so sánh tỷ lệ khách hàng nam và tỷ lệ khách hàng nữ về sở thích màu sắc của áo.

GIẢI:

Lập các giả thiết

H_0 :tỷ lệ khách hàng nam và nữ về sở thích màu sắc của áo là như nhau

H_1 :tỷ lệ khách hàng nam và nữ về sở thích màu sắc của áo là khác nhau

Chọn thống kê kiểm định là

$$\chi^2 = \frac{N^2}{m.l}\sum_{i=1}^{4}\frac{m_i^2}{n_i} - N\frac{m}{l} \sim \chi^2(k-1=3)$$

Với: N=663 m=209 l=454

m_i = 42; 34; 62; 71

l_i = 54; 223; 125; 52

n_i = 96; 257; 187; 123

Từ $P(\chi^2 > \chi^2_\alpha) = \alpha = 0{,}01 \Rightarrow \chi^2_\alpha = 11{,}34$

Tính $\chi^2 = \frac{663^2}{209.454}(\frac{42^2}{96} + \frac{34^2}{257} + \frac{62^2}{187} + \frac{71^2}{123}) - 663(\frac{209}{454}) = 85{,}84$ suy ra $\chi^2 > \chi^2_\alpha$

Vậy bác bỏ giả thiết H_0 , nghĩa là tỷ lệ khách hàng nam và nữ về sở thích màu sắc của áo là khác nhau.

9. KIỂM ĐỊNH TÍNH ĐỘC LẬP CỦA HAI DẤU HIỆU

Giả sử chúng ta cần nghiên cứu về tính độc lập của hai dấu hiệu định tính A và B, hay hai dấu hiệu định lượng A và B nào đó của tổng thể.

Dấu hiệu định tính, ví dụ: sự ưa chuộng một sản phẩm nào đó; chất lượng của sản phẩm, màu da, màu mắt của một người….

Dấu hiệu định lượng, ví dụ: giá tiền của sản phẩm, thu nhập của công nhân, chiều cao, trọng lượng của sinh viên….

Xét hai dấu hiệu A và B của tổng thể, dấu hiệu A có r thành phần $A_1; A_2; \ldots A_r$, dấu hiệu B có k thành phần $B_1; B_2; \ldots B_k$ .Giả sử chúng ta có một mẫu ngẫu nhiên gồm n phần tử, mỗi phần tử sẽ nhận đồng thời dấu hiệu A ở thành phần A_i và nhận dấu hiệu B ở thành phần B_j, chúng ta có bảng số liệu sau, được gọi là bảng liên hợp các dấu hiệu.

A \ B	B_1	B_2	...	B_k	Tổng
A_1	n_{11}	n_{12}	...	n_{1k}	n_1
A_2	n_{21}	n_{22}	...	n_{2k}	n_2
...	...	...	...	...	...
A_r	n_{r1}	n_{r2}	...	n_{rk}	n_r
Tổng	m_1	m_2	...	m_k	n

Với $n_i = \sum_{j=1}^{k} n_{ij}$; $m_j = \sum_{i=1}^{r} n_{ij}$; $n = \sum_{i=1}^{r} \sum_{j=1}^{k} n_{ij}$

$$p_{ij} = P(A_i B_j) = \frac{n_{ij}}{n} \quad ; \quad i = \overline{1,r}; j = \overline{1,k}$$

$$p_i = P(A_i) = \frac{n_i}{n}$$

$$q_j = P(B_j) = \frac{m_j}{n}$$

Lập các giả thiết:

Giả thiết không

H_0 : A và B độc lập

Giả thiết đối

H_1 : A và B phụ thuộc

Giả sử giả thiết H_0 đúng nghĩa là:

$$P(A_i B_j) = P(A_i).P(B_j); \forall i, j \Leftrightarrow p_{ij} = p_i q_j \text{ , } \forall i, j \Leftrightarrow \frac{n_{ij}}{n} = \frac{n_i}{n}.\frac{m_j}{n}; i = \overline{1,r}; j = \overline{1,k}$$

Chọn thống kê kiểm định

$$\chi^2 = n\left[\sum_{i=1}^{r}\sum_{j=1}^{k}\frac{n_{ij}^2}{n_i m_j} - 1\right]$$

CHÚ Ý:

Trường hợp n lớn thì

$\chi^2 = n\left[\sum_{i=1}^{r}\sum_{j=1}^{k}\frac{n_{ij}^2}{n_i m_j} - 1\right]$ có phân phối chi bình phương bậc tự do là: (r-1)(k-1).

$$\chi^2 = n\left[\sum_{i=1}^{r}\sum_{j=1}^{k}\frac{n_{ij}^2}{n_i m_j} - 1\right] \sim \chi^2\left[(r-1)(k-1)\right]$$

Với α là mức ý nghĩa

Ta có $P(\chi^2 > \chi_\alpha^2) = \alpha \Rightarrow \chi_\alpha^2$

Dựa vào số liệu của mẫu cụ thể chúng ta tính

$$\chi^2 = n\left[\sum_{i=1}^{r}\sum_{j=1}^{k}\frac{n_{ij}^2}{n_i m_j} - 1\right]$$

Quy tắc kiểm định

- Nếu $\chi^2 > \chi_\alpha^2$ thì bác bỏ giả thiết H_0
- Nếu $\chi^2 \le \chi_\alpha^2$ thì chưa đủ cơ sở để bác bỏ giả thiết H_0

VD:

Một nhật báo muốn thăm dò ý kiến khán giả về số trang dành cho tin thời sự quốc tế. Phiếu thăm dò đưa ra 4 mức:

A_1 : tăng số trang

A_2 : vẫn giữ như cũ

A_3 : giảm số trang

A_4 : không có ý kiến

Nhật báo thăm dò ý kiến ở hai nhóm độ tuổi: ≤ 40 và > 40. Kết quả thăm dò được số liệu như sau:

Tuổi / Ý kiến	≤ 40	> 40
Tăng	120	300
Như cũ	230	400
Giảm	55	80
Không ý kiến	35	70

Với mức ý nghĩa 5%, hãy nhận định xem độ tuổi và vấn đề tăng trang tin thời sự quốc tế có độc lập hay không?

GIẢI:

Lập các giả thiết:

H_0 : ý kiến tăng số trang tin thời sự quốc tế và độ tuổi độc lập với nhau.

H_1 : ý kiến tăng số trang tin thời sự quốc tế và độ tuổi không độc lập lập

Tuổi / Ý kiến	≤ 40	>40	Tổng số
Tăng	120	300	420
Như cũ	230	400	630
Giảm	55	80	135
Không YK	35	70	105
Tổng số	440	850	1290

Ta có: r=4 ; k=2

Tính:

$$\chi^2 = n\left[\sum_{i=1}^{4}\sum_{j=1}^{2}\frac{n_{ij}^2}{n_i m_j} - 1\right] =$$

$$= 1290\left[\frac{120^2}{440.420} + \frac{300^2}{850.420} + \ldots + \frac{35^2}{440.105} + \frac{70^2}{850.105}\right] = 10{,}059$$

Từ $\alpha = 0{,}05$ tra bảng phân phối chi bình phương với bậc tự do là:

(r-1)(k-1)=(4-1)(2-1)=3

Tìm được $\chi_\alpha^2 = \chi_{0,05}^2 = 7{,}815$

Suy ra $\chi^2 > \chi_\alpha^2$

Vậy bác bỏ giả thiết H_0 , nghĩa là ý kiến tăng số trang tin thời sự quốc tế và độ tuổi

không độc lập.

VD:

Tại một địa phương khảo sát thu nhập trong một tháng của công nhân thuộc hai nhóm tuổi : nhóm một ≤ 40 tuổi, nhóm hai từ 40-60 tuổi, được số liệu như sau: (đơn vị triệu đồng)

Thu nhập (triệu đ) / Tuổi	≤ 1	1 - 2	2 - 3	3 - 4	4 - 6	≥ 6
≤ 40	71	430	1072	1609	1178	158
40 - 60	54	324	894	1202	963	112

Với mức ý nghĩa 5%, hãy xét xem mức thu nhập và độ tuổi có độc lập hay không?

GIẢI:

Lập các giả thiết:

H_0 : mức thu nhập và độ tuổi độc lập.

H_1: mức thu nhập và độ tuổi không độc lập.

Thu nhập / Tuổi	≤ 1	1- 2	2 - 3	3- 4	4- 6	≥ 6	Tổng
≤ 40	71	430	1072	1609	1178	158	4518
40 - 60	54	324	894	1202	963	112	3549
Tổng	125	754	1966	2811	2141	270	8067

Ta có: r=2 ; k=6

Tính $\chi^2 = n\left[\sum_{i=1}^{2}\sum_{j=1}^{6}\frac{n_{ij}^2}{n_i m_j} - 1\right] = 4,27$

Từ mức ý nghĩa $\alpha = 0,05$, tra bảng phân phối chi bình phương bậc tự do là

(r-1)(k-1)=(2-1)(6-1)=5

Tìm được $\chi_\alpha^2 = \chi_{0,05}^2 = 11,07$

Suy ra $\chi^2 < \chi_\alpha^2$

Vậy chưa đủ cơ sở để bác bỏ giả thiết H_0 , nghĩa là mức thu nhập và độ tuổi độc lập với nhau.