Chương 6: Kiểm định giả thuyết thống kê

Phan Quang Sáng

Bộ môn Toán-Khoa CNTT-VNUA

Hà Nội, Ngày 3 tháng 11 năm 2018
http://fita.vnua.edu.vn/vi/pqsang

Nội dung chính

1. Các khái niệm về kiểm định

2. Kiểm định giả thuyết tham số
   - 2.1. Kiểm định kỳ vọng của phân phối chuẩn
   - 2.2. Kiểm định so sánh hai kỳ vọng của hai PP chuẩn
   - 2.3. Kiểm định tỷ lệ (hay XS)
   - 2.4. Kiểm định so sánh hai tỷ lệ (hay XS)

3. Kiểm định giả thuyết phi tham số
   - 3.1. Kiểm định một phân phối xác suất
   - 3.2. Kiểm định tính độc lập của hai đặc tính định tính

Giả thuyết và đối thuyết (thống kê) là gì?

Giả thuyết là một mệnh đề (hay khẳng định) nào đó, ký hiệu, H_0

Đối thuyết là một mệnh đề trái với giả thuyết, ký hiệu, H_1

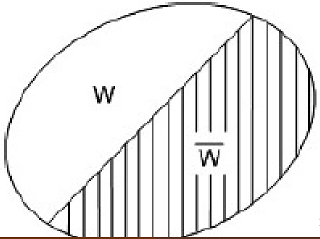
Ví dụ:

Một số dạng kiểm định thống kê:

- Kiểm định phi tham số: liên quan đến quy luật pp XS một đặc tính nào đó ở tổng thể
- Kiểm định tham số có mặt trong quy luật phân phối xs ở tổng thể (Vd các số đặc trưng, tỷ lệ...)

Bài toán Kiểm định giả thuyết thống kê?
Xây dựng một quy tắc sao cho từ mẫu ngẫu nhiên ta có thể ra quyết định chấp nhận giả thuyết H_0 hoặc bác bỏ H_0, cũng tương ứng là bác bỏ hoặc chấp nhận H_1.

Một quy tắc như thế tương đương với việc phân chia không gian mẫu (kích thước n) thành hai phần W và $\overline{W}$ sao cho:
- Nếu mẫu $(X_1, X_2, \cdots, X_n) \in W$ thì quyết định bác bỏ H_0;
- Nếu mẫu $(X_1, X_2, \cdots, X_n) \in \overline{W}$ thì quyết định chấp nhận H_0



Các loại sai lầm

(1) Sai lầm loại 1: H_0 đúng nhưng Qđ bác bỏ H_0 (chọn H_1) do mẫu $(X_1, X_2, \cdots, X_n) \in W$.
XS mắc phải sai lầm loại 1 là:

$$\alpha := P(W/H_0).$$

(2) Sai lầm loại 2: H_0 sai (tức là coi H_1 đúng) nhưng Qđ chấp nhận H_0 (bác H_1) do mẫu $(X_1, X_2, \cdots, X_n) \in \overline{W}$.
XS mắc phải sai lầm loại 2 là:

$$\beta := P(\overline{W}/H_1).$$

$\Rightarrow$ Mong muốn α và β bé nhất có thể

Tuy nhiên khi α giảm thì β tăng và ngược lại

Tuy nhiên khi α giảm thì β tăng và ngược lại

$\Rightarrow$ Với kích thước mẫu cố định, người ta giải quyết bài toán theo hướng: cố định XS sai lầm loại 1 α và xây dựng quy tắc sao cho β là nhỏ nhất hoặc ở mức chấp nhận được. (hoặc theo hướng ngược lại)

XS sai lầm loại 1 α **được gọi là mức ý nghĩa** của bài toán kiểm định

Một số dạng bài toán

1) Giả thuyết đơn- Đối thuyết đơn:

Với các số θ_0, $\theta_1 \neq \theta_0$ cho trước:

$$\begin{cases} H_0: & \theta = \theta_0 \\ H_1: & \theta = \theta_1 \end{cases}$$

Một số dạng bài toán

1) Giả thuyết đơn- Đối thuyết đơn:

Với các số θ_0, $\theta_1 \neq \theta_0$ cho trước:

$$\begin{cases} H_0 : & \theta = \theta_0 \\ H_1 : & \theta = \theta_1 \end{cases}$$

2) Giả thuyết đơn- Đối thuyết hợp:

Với số θ_0 cho trước và D là một miền trong $\mathbb{R}$ không chứa θ_0:

$$\begin{cases} H_0 : & \theta = \theta_0 \\ H_1 : & \theta \in D \end{cases}$$

Một số dạng bài toán

1) Giả thuyết đơn- Đối thuyết đơn:

Với các số θ_0, $\theta_1 \neq \theta_0$ cho trước:

$$\begin{cases} H_0: & \theta = \theta_0 \\ H_1: & \theta = \theta_1 \end{cases}$$

2) Giả thuyết đơn- Đối thuyết hợp:

Với số θ_0 cho trước và D là một miền trong $\mathbb{R}$ không chứa θ_0:

$$\begin{cases} H_0: & \theta = \theta_0 \\ H_1: & \theta \in D \end{cases}$$

Đặc biệt:

- Nếu $H_1: \theta \neq \theta_0$ thì gọi là đối thuyết 2 phía của H_0 (BT 1)
- Nếu $H_1: \theta > \theta_0$ thì gọi là đối thuyết phía phải của H_0 (BT 2)
- Nếu $H_1: \theta < \theta_0$ thì gọi là đối thuyết phía trái của H_0 (BT 3)

Giả sử đặc trưng X ở tổng thể có PP chuẩn $X \sim N(\mu, \sigma^2)$. Với mức ý nghĩa α, từ mẫu ngẫu nhiên $(X_1, X_2, \cdots, X_n)$ xây dựng quy tắc kiểm định cặp:

$$\begin{cases} H_0 : & \mu = \mu_0 \\ H_1 : & \mu \, \square \, \mu_0 \end{cases} \tag{1}$$

Giả sử đặc trưng X ở tổng thể có PP chuẩn $X \sim N(\mu, \sigma^2)$. Với mức ý nghĩa α, từ mẫu ngẫu nhiên $(X_1, X_2, \cdots, X_n)$ xây dựng quy tắc kiểm định cặp:

$$\begin{cases} H_0: & \mu = \mu_0 \\ H_1: & \mu \,\square\, \mu_0 \end{cases} \tag{1}$$

Trường hợp 1: Phương sai σ^2 đã biết
Trường hợp 2: Phương sai σ^2 chưa biết

Giả sử đặc trưng X ở tổng thể có PP chuẩn $X \sim N(\mu, \sigma^2)$. Với mức ý nghĩa α, từ mẫu ngẫu nhiên $(X_1, X_2, \cdots, X_n)$ xây dựng quy tắc kiểm định cặp:

$$\begin{cases} H_0 : & \mu = \mu_0 \\ H_1 : & \mu \square \mu_0 \end{cases} \qquad (1)$$

Trường hợp 1: Phương sai σ^2 đã biết
Trường hợp 2: Phương sai σ^2 chưa biết

Các bài toán tương ứng:

Bài toán 1: $H_0 : \mu = \mu_0$, $H_1 : \mu \neq \mu_0$,
Bài toán 2: $H_0 : \mu = \mu_0$, $H_1 : \mu > \mu_0$,
Bài toán 3: $H_0 : \mu = \mu_0$, $H_1 : \mu < \mu_0$,

Bài toán 1 trong **Trường hợp 1**: σ^2 đã biết,

$$\begin{cases} H_0: & \mu = \mu_0 \\ H_1: & \mu \neq \mu_0 \end{cases}$$

Bài toán 1 trong **Trường hợp 1**: σ^2 đã biết,

$$\begin{cases} H_0: & \mu = \mu_0 \\ H_1: & \mu \neq \mu_0 \end{cases}$$

Quy tắc: chấp nhận H_0 và bác H_1 nếu $|\overline{X} - \mu_0| \leq \varepsilon$ đủ nhỏ; bác H_0 và chấp nhận H_1 nếu $|\overline{X} - \mu_0| > \varepsilon$ với một số $\varepsilon > 0$.
Đặt

$$Z = \frac{\overline{X} - \mu_0}{\sigma}\sqrt{n} \text{ (chú ý rằng Z có PP chuẩn)}$$

Quy tắc trên cũng tương đương với việc so sánh $|Z| \square t$, với $t > 0$ là một mức giới hạn đủ nhỏ (sẽ được chọn sao cho sai lầm loại 1 là α)

Bài toán 1 trong **Trường hợp 1**: σ^2 đã biết,

$$\begin{cases} H_0 : & \mu = \mu_0 \\ H_1 : & \mu \neq \mu_0 \end{cases}$$

Quy tắc: chấp nhận H_0 và bác H_1 nếu $|\overline{X} - \mu_0| \leq \varepsilon$ đủ nhỏ; bác H_0 và chấp nhận H_1 nếu $|\overline{X} - \mu_0| > \varepsilon$ với một số $\varepsilon > 0$.
Đặt

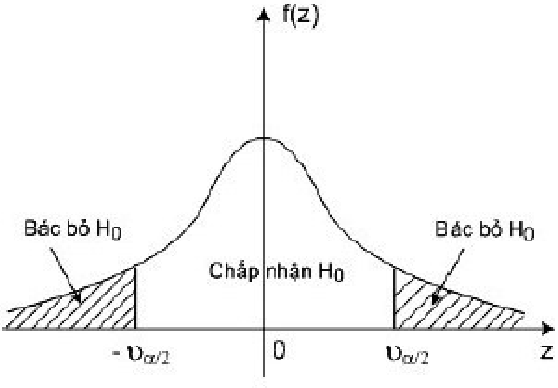
$$Z = \frac{\overline{X} - \mu_0}{\sigma}\sqrt{n} \text{ (chú ý rằng Z có PP chuẩn)}$$

Quy tắc trên cũng tương đương với việc so sánh $|Z| \;\square\; t$, với $t > 0$ là một mức giới hạn đủ nhỏ (sẽ được chọn sao cho sai lầm loại 1 là α)

Cố định α?

$$\alpha = P(\text{bác } H_0/\text{khi } H_0 \text{ đúng }) = P(|Z| > t/H_0)$$

Khi H_0 đúng: $\mu_0 = \mu$ và do đó Z có phân phối chuẩn tắc



$\Rightarrow P(|Z| > t/H_0) = \alpha$ khi $t = U_{\frac{\alpha}{2}}$.

Tóm lại quy tắc cho Bài toán 1 như sau:

1. Tính $Z = \frac{\overline{X} - \mu_0}{\sigma}\sqrt{n}$
2. Mức giới hạn $t = U_{\frac{\alpha}{2}}$
3. So sánh $|Z| \quad \square \quad U_{\frac{\alpha}{2}}$:
   - Nếu dấu $>$ thì KL bác H_0 (chấp nhận H_1).
   - Nếu dấu $\leq$ thì KL chấp nhận H_0 (bác H_1)

Tóm lại quy tắc cho Bài toán 1 như sau:

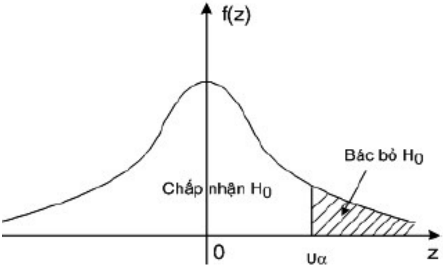
1. Tính $Z = \frac{\overline{X} - \mu_0}{\sigma} \sqrt{n}$
2. Mức giới hạn $t = U_{\frac{\alpha}{2}}$
3. So sánh $|Z| \quad \square \quad U_{\frac{\alpha}{2}}$:
   - Nếu dấu $>$ thì KL bác H_0 (chấp nhận H_1).
   - Nếu dấu $\leq$ thì KL chấp nhận H_0 (bác H_1)

Ví dụ: giả sử năng suất X của một giống lúa trong vùng có PP chuẩn với phương sai là 0,8. Khi điều tra người ta được bảng số liệu sau

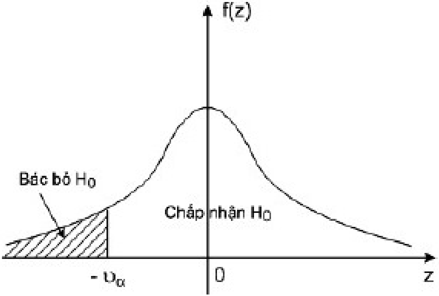
Năng suất	5	5,2	5,4	5,6	5,8	6
Số thửa	3	5	10	9	6	3

Với mức ý nghĩa 0,05 có thể cho rằng năng suất lúa trung bình của vùng là 5,5 tấn/ha không? Cho $U_{0,05} = 1,65$. $U_{0,025} = 1,96$.

Bài toán 2



Bài toán 3



Tương tự quy tắc cho Bài toán 2 và Bài toán 3: với mức ý nghĩa α,

$$\begin{cases} H_0: & \mu = \mu_0 \\ H_1: & \mu \square \mu_0 \end{cases} \tag{2}$$

Tính

$$Z = \frac{\overline{X} - \mu_0}{\sigma}\sqrt{n}$$

Bài toán	H_1	Mức GH	So sánh	KL
Bài toán 1	$\neq$	$U_{\frac{\alpha}{2}}$	$\lvert Z\rvert \quad \square \quad U_{\frac{\alpha}{2}}$	$>$ bác H_0
Bài toán 2	$>$	U_α	$Z \quad \square \quad U_\alpha$	$>$ bác H_0
Bài toán 3	$<$	$-U_\alpha$	$Z \quad \square - U_\alpha$	$<$ bác H_0

Tương tự quy tắc cho Bài toán 2 và Bài toán 3: với mức ý nghĩa α,

$$\begin{cases} H_0: & \mu = \mu_0 \\ H_1: & \mu \;\square\; \mu_0 \end{cases} \qquad (2)$$

Tính

$$Z = \frac{\overline{X} - \mu_0}{\sigma}\sqrt{n}$$

Bài toán	H_1	Mức GH	So sánh	KL
Bài toán 1	$\neq$	$U_{\frac{\alpha}{2}}$	$\lvert Z\rvert \;\square\; U_{\frac{\alpha}{2}}$	$>$ bác H_0
Bài toán 2	$>$	U_α	$Z \;\square\; U_\alpha$	$>$ bác H_0
Bài toán 3	$<$	$-U_\alpha$	$Z \;\square\; -U_\alpha$	$<$ bác H_0

Trường hợp 2: Phương sai σ^2 chưa biết

Quy tắc tương tự như trên nhưng thay

$$\sigma \text{ bởi } S,$$

$$U_{\frac{\alpha}{2}} \text{ bởi } t_{\frac{\alpha}{2},n-1}, \text{ hoặc } U_\alpha \text{ bởi } t_{\alpha,n-1}$$

Ví dụ...

Giả sử $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$. Với mức ý nghĩa α, xây dựng quy tắc kiểm định từ các mẫu ngẫu nhiên kích thước n của X, kích thước m của Y cặp:

$$\begin{cases} H_0: & \mu_X = \mu_Y \\ H_1: & \mu_X \,\square\, \mu_Y \end{cases}$$

Hai bài toán tương ứng với dấu trong H_1: $\neq, >$.

Ví dụ...

Giả sử $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$. Với mức ý nghĩa α, xây dựng quy tắc kiểm định từ các mẫu ngẫu nhiên kích thước n của X, kích thước m của Y cặp:

$$\begin{cases} H_0 : & \mu_X = \mu_Y \\ H_1 : & \mu_X \square \mu_Y \end{cases}$$

Hai bài toán tương ứng với dấu trong H_1: $\neq, >$.

Trường hợp 1: σ_X^2, σ_Y^2 đã biết

Trường hợp 2: σ_X^2, σ_Y^2 chưa biết nhưng $n, m \geq 30$ (mẫu lớn)

Trường hợp 3: σ_X^2, σ_Y^2 chưa biết nhưng biết $\sigma_X^2 = \sigma_Y^2$

Trường hợp 1: σ_X^2, σ_Y^2 đã biết

$$\begin{cases} H_0: & \mu_X = \mu_Y \\ H_1: & \mu_X \,\square\, \mu_Y \end{cases}$$

Tính

$$Z = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}$$

Bài toán	H_1	Mức GH	So sánh	KL
Bài toán 1	$\neq$	$U_{\frac{\alpha}{2}}$	$\lvert Z\rvert \;\square\; U_{\frac{\alpha}{2}}$	$>$ bác H_0
Bài toán 2	$>$	U_α	$Z \;\square\; U_\alpha$	$>$ bác H_0

Trường hợp 2: σ_X^2, σ_Y^2 chưa biết nhưng $n, m \geq 30$ (mẫu lớn)

$$\begin{cases} H_0: & \mu_X = \mu_Y \\ H_1: & \mu_X \square \mu_Y \end{cases}$$

Tính

$$Z = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}}$$

Bài toán	H_1	Mức GH	So sánh	KL
Bài toán 1	$\neq$	$U_{\frac{\alpha}{2}}$	$\lvert Z\rvert \square U_{\frac{\alpha}{2}}$	$>$ bác H_0
Bài toán 2	$>$	U_α	$Z \square U_\alpha$	$>$ bác H_0

Trường hợp 3: σ_X^2, σ_Y^2 chưa biết nhưng biết $\sigma_X^2 = \sigma_Y^2$

Tính

$$S^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}$$

Trường hợp 3: σ_X^2, σ_Y^2 chưa biết nhưng biết $\sigma_X^2 = \sigma_Y^2$

Tính

$$S^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2} = \frac{n(\overline{X^2} - \overline{X}^2) + m(\overline{Y^2} - \overline{Y}^2)}{n+m-2},$$

Trường hợp 3: σ_X^2, σ_Y^2 chưa biết nhưng biết $\sigma_X^2 = \sigma_Y^2$

Tính

$$S^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2} = \frac{n(\overline{X^2} - \overline{X}^2) + m(\overline{Y^2} - \overline{Y}^2)}{n+m-2},$$

$$Z = \frac{\overline{X} - \overline{Y}}{S\sqrt{\frac{1}{n} + \frac{1}{m}}}$$

Bài toán	H_1	Mức GH	So sánh	KL
Bài toán 1	$\neq$	$t_{\frac{\alpha}{2}, n+m-2}$	$\|Z\| \quad \square \quad t_{\frac{\alpha}{2}, n+m-2}$	$>$ bác H_0
Bài toán 2	$>$	$t_{\alpha, n+m-2}$	$Z \quad \square \quad t_{\alpha, n+m-2}$	$>$ bác H_0

Phương pháp so sánh cặp đôi

Ví dụ...

Giả sử X, Y là hai Bnn ĐL tuân theo PP chuẩn. Từ các cặp MNN $(X_1, Y_1), (X_2, Y_2) \cdots, (X_n, Y_n)$ của (X, Y), giả sử cần kiểm định bài toán

$$\begin{cases} H_0: & \mu_X = \mu_Y + \mu_0 \\ H_1: & \mu_X \square \mu_Y + \mu_0 \end{cases}$$

với mức ý nghĩa α, trong đó μ_0 là một hằng số.

Khi đó đặt $U = X - Y$ thì U có pp chuẩn với

$$\mu_U = \mu_X - \mu_Y, \ D(U) = D(X) + D(Y).$$

Bài toán KĐ trên tương đương với bài toán KĐ (đã biết cách giải quyết):

$$\begin{cases} H_0: & \mu_U = \mu_0 \\ H_1: & \mu_U \square \mu_0 \end{cases}$$

với MNN $U_i = X_i - Y_i, i = 1, 2, \cdots, n$.

Ví dụ: có phải tỷ lệ cá chép trong hồ là 25%?

Ví dụ: có phải tỷ lệ cá chép trong hồ là 25%?

Giả sử ở mỗi phép thử $P(A) = p$, và p_0 là một hằng số cho trước. Với mức ý nghĩa α, ta kiểm định cặp GT-ĐT sau:

$$\begin{cases} H_0 : & p = p_0 \\ H_1 : & p \square p_0 \end{cases}$$

Ví dụ: có phải tỷ lệ cá chép trong hồ là 25%?

Giả sử ở mỗi phép thử $P(A) = p$, và p_0 là một hằng số cho trước. Với mức ý nghĩa α, ta kiểm định cặp GT-ĐT sau:

$$\begin{cases} H_0: & p = p_0 \\ H_1: & p \square p_0 \end{cases}$$

Tiến hành n phép thử thấy A xuất hiện n_A lần.

Đặt $f = \dfrac{n_A}{n}$ và $Z = \dfrac{f - p_0}{\sqrt{p_0 q_0}}\sqrt{n}$, với $q_0 = 1 - p_0$.

Bài toán	H_1	Mức GH	So sánh	KL
Bài toán 1	$\neq$	$U_{\frac{\alpha}{2}}$	$\lvert Z\rvert \quad \square \quad U_{\frac{\alpha}{2}}$	$>$ bác H_0
Bài toán 2	$>$	U_α	$Z \quad \square \quad U_\alpha$	$>$ bác H_0
Bài toán 3	$<$	$-U_\alpha$	$Z \quad \square - U_\alpha$	$<$ bác H_0

Ví dụ: so sánh tỷ lệ cá chép ở trong hai hồ khác nhau?

Ví dụ: so sánh tỷ lệ cá chép ở trong hai hồ khác nhau?

Giả sử XS của Sk A và B lần lượt là $P(A) = p_1$ và $P(B) = p_2$. Với mức ý nghĩa α, ta kiểm định cặp GT-ĐT sau:

$$\begin{cases} H_0: & p_1 = p_2 \\ H_1: & p_1 \ \square \ p_2 \end{cases}$$

Tiến hành n phép thử ĐL thấy A xuất hiện n_A lần.

Tiến hành m phép thử thấy B xuất hiện m_B lần.

Đặt

$$f_1 = \frac{n_A}{n}, f_2 = \frac{m_B}{m}, f = \frac{n_A + m_B}{n + m};$$

$$Z = \frac{f_1 - f_2}{\sqrt{f(1-f)(\frac{1}{n} + \frac{1}{m})}}.$$

Bài toán	H_1	Mức GH	So sánh	KL
Bài toán 1	$\neq$	$U_{\frac{\alpha}{2}}$	$\lvert Z\rvert \quad \square \quad U_{\frac{\alpha}{2}}$	$>$ bác H_0
Bài toán 2	$>$	U_α	$Z \quad \square \quad U_\alpha$	$>$ bác H_0

Ví dụ:

Ví dụ 1: có đúng thế hệ cà chua F_2 tuân theo quy luật phân ly tính trạng 1 : 2 : 1?

Ví dụ 2: có phải tỷ lệ cá chép, cá trôi, cá mè trong hồ là 25%, 40%, 35% không?

Ví dụ 1: có đúng thế hệ cà chua F_2 tuân theo quy luật phân ly tính trạng 1 : 2 : 1?

Ví dụ 2: có phải tỷ lệ cá chép, cá trôi, cá mè trong hồ là 25%, 40%, 35% không?

Cho $A_1, A_2, \ldots, A_k$ là một hệ đầy đủ. Với mức ý nghĩa α, giả sử cần kiểm định bài toán:

$$\begin{cases} H_0: & P(A_1) = p_1,\ P(A_2) = p_2, \cdots, P(A_k) = p_k \\ H_1: & \exists\, j \text{ sao cho } P(A_j) \neq p_j \end{cases}$$

Ví dụ 1: có đúng thế hệ cà chua F_2 tuân theo quy luật phân ly tính trạng 1 : 2 : 1?

Ví dụ 2: có phải tỷ lệ cá chép, cá trôi, cá mè trong hồ là 25%, 40%, 35% không?

Cho $A_1, A_2, \ldots, A_k$ là một hệ đầy đủ. Với mức ý nghĩa α, giả sử cần kiểm định bài toán:

$$\begin{cases} H_0: & P(A_1) = p_1,\ P(A_2) = p_2, \cdots, P(A_k) = p_k \\ H_1: & \exists\, j \text{ sao cho } P(A_j) \neq p_j \end{cases}$$

Tiến hành n phép thử ĐL ta thu được bảng số liệu sau:

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

Quy tắc kiểm định

Đặt

$$Z = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i}, \text{ hay}$$

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

Quy tắc kiểm định

Đặt

$$Z = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i}, \text{ hay}$$

$$Z = \sum_{i=1}^{k} \frac{n_i^2}{np_i} - n \, \square \, \chi^2_{k-1;\alpha}.$$

Nếu dấu ">" thì bác H_0 (chấp nhận H_1)...

Ví dụ: Câu IV- đề số 3, 12/06/2018, HK2 17-18

Câu IV (3,5 điểm) Phân khúc thị trường sữa tươi của các công ty TH True milk, Vinamilk, Mộc châu ở một vùng là 40%, 35% và 25%. Để tăng thị phần của mình Vinamilk tiến hành một chiến dịch quảng cáo với quy mô lớn. Sau đợt quảng cáo người ta khảo sát ngẫu nhiên 200 khách hàng dùng sữa tươi ở vùng đó thu được kết quả:

Công ty	TH True milk	Vinamilk	Mộc châu
Số người dùng	70	85	45

1. (2,0đ) Với mức ý nghĩa 5%, có thể cho rằng phân khúc thị trường sữa của ba công ty trên đã thay đổi sau chiến dịch quảng cáo của Vinamilk hay không? (gợi ý: kiểm định luật phân phối xác suất)
2. (1,5đ) Ước lượng tỷ lệ khách hàng dùng sữa tươi Vinamilk trong số khách hàng dùng sữa tươi ở vùng trên sau chiến dịch quảng cáo với độ tin cậy 95%.

Cho: $U_{0,025} = 1,96$; $t_{0,025;10} = 2,228$; $t_{0,05;10} = 1,812$; $\chi^2_{0,05;2} = 5,991$; $\Phi(1,7391) = 0,959$; $t_{0,025;5} = 2,57$

Ta quan tâm đến 2 đặc tính định tính A và B của mỗi cá thể trong một đám đông.

Ví dụ: màu mắt và màu tóc, nhóm máu và huyết áp, thời gian học tập và kết quả học tập của Sv...

Ta quan tâm đến 2 đặc tính định tính A và B của mỗi cá thể trong một đám đông.

Ví dụ: màu mắt và màu tóc, nhóm máu và huyết áp, thời gian học tập và kết quả học tập của Sv...

Câu hỏi: hai đặc tính A và B có độc lập không?

Ta quan tâm đến 2 đặc tính định tính A và B của mỗi cá thể trong một đám đông.

Ví dụ: màu mắt và màu tóc, nhóm máu và huyết áp, thời gian học tập và kết quả học tập của Sv...

Câu hỏi: hai đặc tính A và B có độc lập không?

Kiểm định cặp

$$\begin{cases} H_0: & A, B \text{ độc lập} \\ H_1: & A, B \text{ không độc lập} \end{cases}$$

với mức ý nghĩa α.

Ví dụ: Để tìm hiểu mối quan hệ giữa việc nghiện thuốc lá (A) và huyết áp (B) người ta điều tra được bảng số liệu sau:

(A có $k = 3$ mức (cột) và B có $m = 2$ mức (hàng))

B \ A	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 5 % hãy kiểm định tính ĐL của A, B.

Giả sử A được chia thành k mức, B được chia thành m mức.

Lấy mẫu ngẫu nhiên có kích thước n được bảng số liệu của các $[n_{ij}]$.

Gọi n_i là tổng các hàng, m_j là tổng các cột,

$$n = \sum n_i = \sum m_j.$$

Khi đó:

$$f_i = \frac{n_i}{n}, f_j = \frac{m_i}{n}, f_{ij} = \frac{n_{ij}}{n},$$

lần lượt là ƯL xs của A_i, B_j, A_iB_j.

Khi H_0 đúng thì các A_i và B_j độc lập nên

$$P(A_iB_j) = P(A_i)P(B_j) \text{ với mọi } i, j.$$

Chấp nhận H_0 nếu

$$f_{ij} \approx f_i f_j \Leftrightarrow n_{ij} \approx \frac{n_i m_j}{n}, \text{ với mọi } i, j.$$

(nếu không thì bác bỏ)

Đặt

$$Z = \sum_{ij} \frac{(n_{ij} - \frac{n_i m_j}{n})^2}{\frac{n_i m_j}{n}}.$$

Hơn nữa,

$$Z = n[\sum_{ij} \frac{n_{ij}^2}{n_i m_j} - 1].$$

Quy tắc kiểm định: so sánh

$$Z = n[\sum_{ij} \frac{n_{ij}^2}{n_i m_j} - 1] \square \chi^2_{(k-1)(m-1),\alpha}.$$

- Nếu dấu $>$ thì KL bác H_0 (chấp nhận H_1).
- Nếu dấu $\leq$ thì KL chấp nhận H_0 (chấp nhận H_1)

Lời giải:
Ta kiểm định cặp Giả thuyết- Đối thuyết sau:

$$\begin{cases} H_0: & A, B \text{ độc lập} \\ H_1: & A, B \text{ không độc lập} \end{cases}$$

với mức ý nghĩa 5 %.
Tổng các hàng, các cột là:

$$n_1 = 50 + 25 + 28 = 103, \;\; n_2 = 30 + 35 + 32 = 97$$

$$m_1 = 80, m_2 = 60, m_3 = 60.$$

$$n = 103 + 97 = 200.$$

$$Z = 200[... - 1] =$$

Mức giới hạn $\chi^2_{\alpha,(k-1)(m-1)} = \chi^2_{0,05;2} = 5,991.$
Vì $Z \,\Box\, 5,991$ nên
KL:...

Ví dụ: Để tìm hiểu mối quan hệ giữa màu tóc và màu mắt người ta điều tra được bảng số liệu sau:

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