Trường Đại Học Khoa Học Tự Nhiên TP.HCM

# XÁC SUẤT THỐNG KÊ

Lưu hành nội bộ
2020





## Định nghĩa 1

**Giả thuyết thống kê** là những giả thuyết nói về các tham số, dạng quy luật phân phối, hoặc tính độc lập của các đại lượng ngẫu nhiên.

Việc tìm ra kết luận bác bỏ hay chấp nhận một giả thuyết gọi là **kiểm định giả thuyết thống kê**





## Ví dụ 1

Trong một báo cáo nói rằng: thu nhập bình quân của những người làm trong ngành thư viện ở Việt Nam là 7 triệu đồng một tháng thì ta có thể coi đó là một giả thuyết thống kê, giả thuyết này nói về một tham số (kỳ vọng) của biến ngẫu nhiên $X$ biểu thị mức lương của những người làm trong ngành thư viện. Dựa vào số liệu của một mẫu điều tra về thu nhập và quy tắc kiểm định để đưa một kết luận là bác bỏ hay chấp nhận giả thuyết nói trên.





## Cách đặt giả thuyết.

1. Giả thuyết được đặt ra với ý đồ bác bỏ nó, nghĩa là giả thuyết đặt ra ngược lại với điều ta muốn chứng minh, muốn thuyết phục.
2. Giả thuyết được đặt ra sao cho khi chấp nhận hay bác bỏ nó sẽ có tác dụng trả lời mà bài toán thực tế đặt ra.
3. Giả thuyết được đặt ra nếu nó đúng thì ta sẽ xác định được qui luật phân phối xác suất của đại lượng ngẫu nhiên được chọn làm tiêu chuẩn kiểm định.
4. Khi đặt giả thuyết ta thường so sánh cái chưa biết với cái đã biết. Cái chưa biết là điều mà ta cần kiểm định, kiểm tra, làm rõ. "Cái đã biết" mà ta nói ở đây thường là những thông tin quá khứ, các định mức kinh tế, kỹ thuật.
5. Giả thuyết đặt ra thường mang ý nghĩa: "không khác nhau", hoặc "khác mà không có ý nghĩa" hoặc "bằng nhau".

## Giả thuyết không và đối thuyết

Giả thuyết cần kiểm định được gọi là Giả thuyết không ký hiệu $H_0$. Một mệnh đề đối lập với $H_0$ được gọi là Giả thuyết đối (đối thuyết) và được ký hiệu là $H_1$.

**Ví dụ 2**

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

Nếu ta kiểm định giả thuyết với đối thuyết dạng như trên thì kiểm định được gọi là kiểm định giả thuyết hai phía. Nếu kiểm định giả thuyết với đối thuyết có dạng $H_1 : \theta > \theta_0$ hoặc $H_1 : \theta < \theta_0$ thì kiểm định được gọi là kiểm định giả thuyết một phía.

XÁC SUẤT THỐNG KÊ
N.T.M.Ngọc
Kiểm định giả thuyết thống kê
Kiểm định giả thuyết về so sánh kì vọng
Kiểm định giả thuyết về so sánh kì vọng với 1 số
Kiểm định giả thuyết về so sánh hai kì vọng
Kiểm định giả thuyết về so sánh tỷ lệ
Kiểm định giả thuyết về so sánh tỷ lệ với một số
Kiểm định giả thuyết về so sánh hai tỷ lệ

## Tiêu chuẩn kiểm định.

Xuất phát từ yêu cầu của bài toán thực tế , ta nêu ra giả thuyết $H_0$ và đối thuyết của nó.
Giả sử rằng $H_0$ đúng, từ đó tìm một biến cố có xác suất đủ bé để có thể tin rằng biến cố đó hầu như không thể xảy ra trong một phép thử. Muốn vậy từ mẫu ngẫu nhiên $(X_1, X_2, ..., X_n)$ ta chọn $Z = f(X_1, ..., X_n, \theta_0)$ sao cho:
Nếu $H_0$ đúng thì ta sẽ xác định được quy luật phân phối xác suất của $Z$ và với mẫu cụ thể ta có thể tính được giá trị của $Z$. Đại lượng ngẫu nhiên $Z$ được gọi là tiêu chuẩn kiểm định giả thuyết $H_0$.

XÁC SUẤT THỐNG KÊ
N.T.M.Ngọc
Kiểm định giả thuyết thống kê
Kiểm định giả thuyết về so sánh kì vọng
Kiểm định giả thuyết về so sánh kì vọng với 1 số
Kiểm định giả thuyết về so sánh hai kì vọng
Kiểm định giả thuyết về so sánh tỷ lệ
Kiểm định giả thuyết về so sánh tỷ lệ với một số
Kiểm định giả thuyết về so sánh hai tỷ lệ

## Miền bác bỏ, mức ý nghĩa.

Do quy luật phân phối xác suất của $Z$ đã biết nên với $\alpha$ bé tùy ý ta có thể tìm được miền $W_\alpha$ sao cho $\mathbb{P}(Z \in W_\alpha) = \alpha$. Miền $W_\alpha$ được gọi là miền bác bỏ giả thuyết $H_0$ và $\alpha$ được gọi là mức ý nghĩa của kiểm định.
Thực hiện một phép thử đối với mẫu ngẫu nhiên $(X_1, ..., X_n)$ ta thu được mẫu cụ thể $(x_1, ..., x_n)$. Từ mẫu cụ thể này ta tính được giá trị của $Z$ (ký hiệu là $z$) và gọi là giá trị thực nghiệm $z = f(x_1, ..., x_n, \theta_0)$.

- Nếu $z \in W_\alpha$ thì ta bác bỏ giả thuyết $H_0$, thừa nhận $H_1$.
- Nếu $z \notin W_\alpha$ thì ta chấp nhận $H_0$.

XÁC SUẤT THỐNG KÊ
N.T.M.Ngọc
Kiểm định giả thuyết thống kê
Kiểm định giả thuyết về so sánh kì vọng
Kiểm định giả thuyết về so sánh kì vọng với 1 số
Kiểm định giả thuyết về so sánh hai kì vọng
Kiểm định giả thuyết về so sánh tỷ lệ
Kiểm định giả thuyết về so sánh tỷ lệ với một số
Kiểm định giả thuyết về so sánh hai tỷ lệ

## Sai lầm loại 1 và sai lầm loại 2.

Khi kiểm định giả thuyết thống kê, chúng ta có thể mắc phải một trong hai loại sai lầm sau đây:

a. Sai lầm loại 1: Là sai lầm mắc phải khi ta bác bỏ giả thuyết $H_0$ trong khi thực tế thì giả thuyết $H_0$ đúng.
b. Sai lầm loại 2: Là sai lầm mắc phải khi ta chấp nhận giả thuyết $H_0$ trong khi thực tế thì giả thuyết $H_0$ sai.

## Quy trình kiểm định.

Quá trình kiểm định giả thuyết thống kê được tiến hành theo các bước sau đây

1. Phát biểu giả thuyết không $H_0$ và đối thuyết $H_1$. Quyết định dữ liệu nào cần được thu thập và thu thập dưới các điều kiện nào. Chọn lựa một kiểm định thống kê (cùng với mô hình thống kê liên kết với nó) để kiểm định $H_0$.
2. Từ một số kiểm định có thể được dùng cho mô hình nghiên cứu, chọn ra kiểm định thích hợp nhất dựa trên cơ sở là các điều kiện của nghiên cứu và các giả định cơ sở của kiểm định.
3. Chọn mức ý nghĩa $\alpha$ và kích thước mẫu $n$.

XÁC SUẤT THỐNG KÊ
N.T.M.Ngọc
Kiểm định giả thuyết thống kê
Kiểm định giả thuyết về so sánh kì vọng
Kiểm định giả thuyết về so sánh kì vọng với 1 số
Kiểm định giả thuyết về so sánh hai kì vọng
Kiểm định giả thuyết về so sánh tỷ lệ
Kiểm định giả thuyết về so sánh tỷ lệ với một số
Kiểm định giả thuyết về so sánh hai tỷ lệ

4. Tìm phân phối mẫu của kiểm định thống kê dưới điều kiện $H_0$ đúng.
5. Trên cơ sở (2), (3) và (4) đã trình bày ở trên, xác định miền bác bỏ của kiểm định thống kê tương ứng.
6. Thu thập dữ liệu. Sử dụng dữ liệu thu được từ mẫu, tính giá trị của kiểm định. Nếu giá trị của thống kê nằm trong miền bác bỏ, ta bác giả thuyết $H_0$, nếu giá trị thu được nằm ngoài miền bác bỏ, kết luận không thể bác bỏ giả thuyết $H_0$ ở mức ý nghĩa đã chọn.

XÁC SUẤT THỐNG KÊ
N.T.M.Ngọc
Kiểm định giả thuyết thống kê
Kiểm định giả thuyết về so sánh kì vọng
Kiểm định giả thuyết về so sánh kì vọng với 1 số
Kiểm định giả thuyết về so sánh hai kì vọng
Kiểm định giả thuyết về so sánh tỷ lệ
Kiểm định giả thuyết về so sánh tỷ lệ với một số
Kiểm định giả thuyết về so sánh hai tỷ lệ

## Quy trình kiểm định trong bài làm

1. Từ mẫu cụ thể đã cho tính giá trị của các thống kê tương ứng với tiêu chuẩn kiểm định trong trường hợp tương ứng.
2. Với mức ý nghĩa $\alpha$ cho trước, xác định miền bác bỏ.
3. Kiểm tra giá trị của tiêu chuẩn kiểm định có nằm trong miền bác bỏ hay không và kết luận.

XÁC SUẤT THỐNG KÊ
N.T.M.Ngọc
Kiểm định giả thuyết thống kê
Kiểm định giả thuyết về so sánh kì vọng
Kiểm định giả thuyết về so sánh kì vọng với 1 số
Kiểm định giả thuyết về so sánh hai kì vọng
Kiểm định giả thuyết về so sánh tỷ lệ
Kiểm định giả thuyết về so sánh tỷ lệ với một số
Kiểm định giả thuyết về so sánh hai tỷ lệ

## Kiểm định giả thuyết về so sánh kì vọng với một số

**Bài toán**

Cho tổng thể với trung bình $\mu$ chưa biết với phương sai có thể đã biết hoặc chưa biết. Từ mẫu ngẫu nhiên $(X_1, X_2, ..., X_n)$ hãy kiểm định

a) $\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu < \mu_0 \end{cases}$ b) $\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu > \mu_0 \end{cases}$ c) $\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$

với mức ý nghĩa $\alpha$.

**Trường hợp kích thước mẫu $n \geq 30$ (hoặc $n < 30$ nhưng $X$ có phân phối chuẩn), $\sigma^2$ đã biết**

Chọn thống kê

$$Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma}$$

làm tiêu chuẩn kiểm định. Nếu giả thuyết $H_0$ đúng thì $Z \sim N(0,1)$. Từ đây ta suy ra miền bác bỏ tương ứng với từng loại đối thuyết.

XÁC SUẤT THỐNG KÊ
N.T.M.Ngọc
Kiểm định giả thuyết thống kê
Kiểm định giả thuyết về so sánh kì vọng
Kiểm định giả thuyết về so sánh kì vọng với 1 số
Kiểm định giả thuyết về so sánh hai kì vọng
Kiểm định giả thuyết về so sánh tỷ lệ
Kiểm định giả thuyết về so sánh tỷ lệ với một số
Kiểm định giả thuyết về so sánh hai tỷ lệ

| Giả thuyết | Miền bác bỏ $H_0$ |
|---|---|
| $H_0: \mu = \mu_0$<br>$H_1: \mu < \mu_0$ | $W_\alpha = \left\{ z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} : z < z_\alpha \right\}$ |
| $H_0: \mu = \mu_0$<br>$H_1: \mu > \mu_0$ | $W_\alpha = \left\{ z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} : z > z_{1-\alpha} \right\}$ |
| $H_0: \mu = \mu_0$<br>$H_1: \mu \neq \mu_0$ | $W_\alpha = \left\{ z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} : \lvert z \rvert > z_{1-\alpha/2} \right\}$ |

Chú ý : $z_\alpha = -z_{1-\alpha} = \mathbb{P}(Z \leq z_\alpha) = \alpha$

XÁC SUẤT THỐNG KÊ
N.T.M.Ngọc
Kiểm định giả thuyết thống kê
Kiểm định giả thuyết về so sánh kì vọng
Kiểm định giả thuyết về so sánh kì vọng với 1 số
Kiểm định giả thuyết về so sánh hai kì vọng
Kiểm định giả thuyết về so sánh tỷ lệ
Kiểm định giả thuyết về so sánh tỷ lệ với một số
Kiểm định giả thuyết về so sánh hai tỷ lệ

**Ví dụ :** Trong năm trước trọng lượng trung bình trước khi xuất chuồng của bò ở một trại chăn nuôi là 380 kg. Năm nay người ta áp dụng thử một chế độ chăn nuôi mới với hi vọng là bò sẽ tăng trọng nhanh hơn. Sau một thời gian áp dụng thử người ta lấy ngẫu nhiên 50 con bò trước khi xuất chuồng đem cân và tính được trọng lượng trung bình của chúng là 390 kg. Với mức ý nghĩa $\alpha = 0.01$ có thể cho rằng trọng lượng trung bình của bò trước khi xuất chuồng đã tăng lên hay không?
Giả thiết trọng lượng của bò là BNN có phân phối chuẩn với độ lệch chuẩn là 35.2 kg.

XÁC SUẤT THỐNG KÊ
N.T.M.Ngọc
Kiểm định giả thuyết thống kê
Kiểm định giả thuyết về so sánh kì vọng
Kiểm định giả thuyết về so sánh kì vọng với 1 số
Kiểm định giả thuyết về so sánh hai kì vọng
Kiểm định giả thuyết về so sánh tỷ lệ
Kiểm định giả thuyết về so sánh tỷ lệ với một số
Kiểm định giả thuyết về so sánh hai tỷ lệ

**Trường hợp kích thước mẫu $n \geq 30$, $\sigma^2$ chưa biết**

Ta có thể dùng ước lượng của $Var(X)$ là $S^2$ để thay thế cho $\sigma^2$.
Chọn thống kê

$$Z = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} = \frac{\sqrt{n}(\bar{X} - \mu_0)}{S}$$

làm tiêu chuẩn kiểm định. Nếu giả thuyết $H_0$ đúng thì $Z \sim N(0,1)$. Từ đây ta suy ra miền bác bỏ tương ứng với từng loại đối thuyết.

| Giả thuyết | Miền bác bỏ $H_0$ |
|---|---|
| $H_0 : \mu = \mu_0$ <br> $H_1 : \mu < \mu_0$ | $W_\alpha = \left\{ z = \frac{\bar{x}-\mu_0}{s/\sqrt{n}} : z < z_\alpha \right\}$ |
| $H_0 : \mu = \mu_0$ <br> $H_1 : \mu > \mu_0$ | $W_\alpha = \left\{ z = \frac{\bar{x}-\mu_0}{s/\sqrt{n}} : z > z_{1-\alpha} \right\}$ |
| $H_0 : \mu = \mu_0$ <br> $H_1 : \mu \neq \mu_0$ | $W_\alpha = \left\{ z = \frac{\bar{x}-\mu_0}{s/\sqrt{n}} : \lvert z \rvert > z_{1-\alpha/2} \right\}$ |



Ví dụ :

Đo đường kính của 36 chi tiết máy ta được bảng số liệu sau:

| Độ dài đường kính | 10.10 | 10.12 | 10.20 | 10.25 | 10.30 |
|---|---|---|---|---|---|
| Số chi tiết | 3 | 15 | 14 | 2 | 2 |

Với mức ý nghĩa $\alpha = 0.05$ hãy cho kết luận về ý kiến: "Trung bình đường kính là 10.20"



Trường hợp kích thước mẫu $n < 30$, $\sigma^2$ chưa biết, $X$ tuân theo quy luật phân phối chuẩn.

Chọn thống kê

$$T = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} = \frac{\sqrt{n}(\bar{X} - \mu_0)}{S}$$

làm tiêu chuẩn kiểm định. Nếu giả thuyết $H_0$ đúng thì $T \sim T(n-1)$ (Phân phối Student với $n-1$ bậc tự do). Từ đây ta suy ra miền bác bỏ tương ứng với từng loại đối thuyết.



| Giả thuyết | Miền bác bỏ $H_0$ |
|---|---|
| $H_0 : \mu = \mu_0$ <br> $H_1 : \mu < \mu_0$ | $W_\alpha = \left\{ t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}} : t < t_\alpha^{n-1} \right\}$ |
| $H_0 : \mu = \mu_0$ <br> $H_1 : \mu > \mu_0$ | $W_\alpha = \left\{ t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}} : t > t_{1-\alpha}^{n-1} \right\}$ |
| $H_0 : \mu = \mu_0$ <br> $H_1 : \mu \neq \mu_0$ | $W_\alpha = \left\{ t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}} : \lvert t \rvert > t_{1-\alpha/2}^{n-1} \right\}$ |

Ví dụ :

Cho 8 kết quả đo đạc về một đại lượng bởi cùng một máy đo không có sai lầm hệ thống:

$$369, 378, 315, 420, 385, 401, 372, 383$$

Với mức ý nghĩa $\alpha = 0.05$, hãy cho kết luận về ý kiến: "Giá trị trung bình là 380". Biết rằng đại lượng được đo có phân phối chuẩn.

XÁC SUẤT THỐNG KÊ
N.T.M.Ngọc
Kiểm định giả thuyết thống kê
Kiểm định giả thuyết về so sánh kì vọng
Kiểm định giả thuyết về so sánh kì vọng với 1 số
Kiểm định giả thuyết về so sánh hai kì vọng
Kiểm định giả thuyết về so sánh tỷ lệ
Kiểm định giả thuyết về so sánh tỷ lệ với một số
Kiểm định giả thuyết về so sánh hai tỷ lệ

## Kiểm định giả thuyết về so sánh hai kì vọng

**Bài toán**

Quan sát $X$ trên 2 mẫu lấy từ hai tổng thể $A$ và $B$.

- Trên tổng thể $A$: $X \sim N(\mu_1, \sigma_1^2)$, mẫu cỡ $n_1$, trung bình mẫu $\overline{X_1}$, phương sai mẫu $S_1^2$.
- Trên tổng thể $B$: $X \sim N(\mu_2, \sigma_2^2)$, mẫu cỡ $n_2$, trung bình mẫu $\overline{X_2}$, phương sai mẫu $S_2^2$.

Hãy kiểm định

a) $\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 < \mu_2 \end{cases}$ b) $\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 > \mu_2 \end{cases}$ c) $\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases}$

với mức ý nghĩa $\alpha$.

XÁC SUẤT THỐNG KÊ
N.T.M.Ngọc
Kiểm định giả thuyết thống kê
Kiểm định giả thuyết về so sánh kì vọng
Kiểm định giả thuyết về so sánh kì vọng với 1 số
Kiểm định giả thuyết về so sánh hai kì vọng
Kiểm định giả thuyết về so sánh tỷ lệ
Kiểm định giả thuyết về so sánh tỷ lệ với một số
Kiểm định giả thuyết về so sánh hai tỷ lệ

**Trường hợp $\sigma_1^2$, $\sigma_2^2$ đã biết.**

Chọn thống kê

$$Z = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

làm tiêu chuẩn kiểm định. Nếu giả thuyết $H_0$ đúng thì $Z \sim N(0, 1)$. Từ đây ta suy ra miền bác bỏ tương ứng với từng loại đối thuyết.

XÁC SUẤT THỐNG KÊ
N.T.M.Ngọc
Kiểm định giả thuyết thống kê
Kiểm định giả thuyết về so sánh kì vọng
Kiểm định giả thuyết về so sánh kì vọng với 1 số
Kiểm định giả thuyết về so sánh hai kì vọng
Kiểm định giả thuyết về so sánh tỷ lệ
Kiểm định giả thuyết về so sánh tỷ lệ với một số
Kiểm định giả thuyết về so sánh hai tỷ lệ

| Giả thuyết | Miền bác bỏ $H_0$ |
|---|---|
| $H_0 : \mu_1 = \mu_2$<br>$H_1 : \mu_1 < \mu_2$ | $W_\alpha = \left\{ z = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} : z < z_\alpha \right\}$ |
| $H_0 : \mu_1 = \mu_2$<br>$H_1 : \mu_1 > \mu_2$ | $W_\alpha = \left\{ z = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} : z > z_{1-\alpha} \right\}$ |
| $H_0 : \mu_1 = \mu_2$<br>$H_1 : \mu_1 \neq \mu_2$ | $W_\alpha = \left\{ z = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} : \lvert z \rvert > z_{1-\alpha/2} \right\}$ |

Trường hợp $\sigma_1^2$, $\sigma_2^2$ chưa biết nhưng $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

Chọn thống kê

$$T = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

làm tiêu chuẩn kiểm định. Nếu giả thuyết $H_0$ đúng thì $T \sim t(n_1 + n_2 - 2)$.
Trong đó,

$$S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

được gọi là phương sai mẫu gộp.
Từ đây ta suy ra miền bác bỏ tương ứng với từng loại đối thuyết.
*Chú ý: Khi n đủ lớn ($n \geq 30$) thì $t(n) \approx N(0,1)$.*

XÁC SUẤT THỐNG KÊ
N.T.M.Ngọc
Kiểm định giả thuyết thống kê
Kiểm định giả thuyết về so sánh kì vọng
Kiểm định giả thuyết về so sánh kì vọng với 1 số
Kiểm định giả thuyết về so sánh hai kì vọng
Kiểm định giả thuyết về so sánh tỷ lệ
Kiểm định giả thuyết về so sánh tỷ lệ với một số
Kiểm định giả thuyết về so sánh hai tỷ lệ

| Giả thuyết | Miền bác bỏ $H_0$ |
|---|---|
| $H_0 : \mu_1 = \mu_2$ <br> $H_1 : \mu_1 < \mu_2$ | $W_\alpha = \left\{ t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} : t < t_\alpha^{n_1+n_2-2} \right\}$ |
| $H_0 : \mu_1 = \mu_2$ <br> $H_1 : \mu_1 > \mu_2$ | $W_\alpha = \left\{ t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} : t > t_{1-\alpha}^{n_1+n_2-2} \right\}$ |
| $H_0 : \mu_1 = \mu_2$ <br> $H_1 : \mu_1 \neq \mu_2$ | $W_\alpha = \left\{ t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} : \lvert t\rvert > t_{1-\alpha/2}^{n_1+n_2-2} \right\}$ |

XÁC SUẤT THỐNG KÊ
N.T.M.Ngọc
Kiểm định giả thuyết thống kê
Kiểm định giả thuyết về so sánh kì vọng
Kiểm định giả thuyết về so sánh kì vọng với 1 số
Kiểm định giả thuyết về so sánh hai kì vọng
Kiểm định giả thuyết về so sánh tỷ lệ
Kiểm định giả thuyết về so sánh tỷ lệ với một số
Kiểm định giả thuyết về so sánh hai tỷ lệ

Ví dụ :

Dùng hai phương pháp để làm cùng một loại sản phẩm. Phương pháp A được một nhóm 12 người thực hiện có năng suất trung bình là 45 sản phẩm trong một ca làm việc, với độ lệch mẫu $s_A = 5$ sản phẩm. Phương pháp B được nhóm 15 người khác thực hiện, có năng suất trung bình 53 sản phẩm với độ lệch mẫu $s_B = 6$ sản phẩm. Với mức ý nghĩa $\alpha = 0.05$, hãy so sánh hiệu quả của hai phương pháp.

XÁC SUẤT THỐNG KÊ
N.T.M.Ngọc
Kiểm định giả thuyết thống kê
Kiểm định giả thuyết về so sánh kì vọng
Kiểm định giả thuyết về so sánh kì vọng với 1 số
Kiểm định giả thuyết về so sánh hai kì vọng
Kiểm định giả thuyết về so sánh tỷ lệ
Kiểm định giả thuyết về so sánh tỷ lệ với một số
Kiểm định giả thuyết về so sánh hai tỷ lệ

## Kiểm định giả thuyết về so sánh tỷ lệ với một số

**Bài toán**

Quan sát tỷ lệ các phần tử loại $A$ trên một mẫu lấy ra từ tổng thể. Giả sử tỉ lệ phần tử loại $A$ trên tổng thể là $p$ (chưa biết), cỡ mẫu $n$, tần suất $\hat{p}$.
Hãy kiểm định

a) $\begin{cases} H_0 : p = p_0 \\ H_1 : p < p_0 \end{cases}$ b) $\begin{cases} H_0 : p = p_0 \\ H_1 : p > p_0 \end{cases}$ c) $\begin{cases} H_0 : p = p_0 \\ H_1 : p \neq p_0 \end{cases}$

với mức ý nghĩa $\alpha$.

- Gọi $Y$ là số phần tử loại $A$ trên một mẫu lấy ra từ tổng thể thì $Y \sim B(n,p)$. Đặt
$$\hat{P} = \frac{Y}{n}$$
- Khi $n$ lớn và $p$ không quá gần 0 hoặc 1 (đk: $n\hat{P} \geq 5$ và $n(1-\hat{P}) \geq 5$), ta chọn thống kê
$$Z = \frac{\hat{P} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{\sqrt{n}(\hat{P} - p_0)}{\sqrt{p_0 q_0}}$$
làm tiêu chuẩn kiểm định. Nếu giả thuyết $H_0$ đúng thì $Z \sim N(0,1)$. Từ đây ta suy ra miền bác bỏ tương ứng với từng loại đối thuyết.

XÁC SUẤT THỐNG KÊ
N.T.M.Ngọc
Kiểm định giả thuyết thống kê
Kiểm định giả thuyết về so sánh kì vọng
Kiểm định giả thuyết về so sánh kì vọng với 1 số
Kiểm định giả thuyết về so sánh hai kì vọng
Kiểm định giả thuyết về so sánh tỷ lệ
Kiểm định giả thuyết về so sánh tỷ lệ với một số
Kiểm định giả thuyết về so sánh hai tỷ lệ

| Giả thuyết | Miền bác bỏ $H_0$ |
|---|---|
| $H_0: p = p_0$<br>$H_1: p < p_0$ | $W_\alpha = \left\{ z = \frac{(\hat{p}-p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}} : z < z_\alpha \right\}$ |
| $H_0: p = p_0$<br>$H_1: p > p_0$ | $W_\alpha = \left\{ z = \frac{(\hat{p}-p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}} : z > z_{1-\alpha} \right\}$ |
| $H_0: p = p_0$<br>$H_1: p \neq p_0$ | $W_\alpha = \left\{ z = \frac{(\hat{p}-p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}} : \lvert z \rvert > z_{1-\alpha/2} \right\}$ |

XÁC SUẤT THỐNG KÊ
N.T.M.Ngọc
Kiểm định giả thuyết thống kê
Kiểm định giả thuyết về so sánh kì vọng
Kiểm định giả thuyết về so sánh kì vọng với 1 số
Kiểm định giả thuyết về so sánh hai kì vọng
Kiểm định giả thuyết về so sánh tỷ lệ
Kiểm định giả thuyết về so sánh tỷ lệ với một số
Kiểm định giả thuyết về so sánh hai tỷ lệ

Ví dụ :

Tỉ lệ người mắc bệnh $A$ ở một địa phương là 5%. Trong một lần kiểm tra sức khỏe ngẫu nhiên 300 người thấy có 25 người mắc bệnh $A$. Với mức ý nghĩa $\alpha = 0.05$ có thể cho rằng tỉ lệ người bị bệnh $A$ có xu hướng tăng lên hay không?

XÁC SUẤT THỐNG KÊ
N.T.M.Ngọc
Kiểm định giả thuyết thống kê
Kiểm định giả thuyết về so sánh kì vọng
Kiểm định giả thuyết về so sánh kì vọng với 1 số
Kiểm định giả thuyết về so sánh hai kì vọng
Kiểm định giả thuyết về so sánh tỷ lệ
Kiểm định giả thuyết về so sánh tỷ lệ với một số
Kiểm định giả thuyết về so sánh hai tỷ lệ

## Kiểm định giả thuyết về so sánh hai tỉ lệ

**Bài toán**

Xét cỡ mẫu lớn: $n_1 \geq 30$, $n_2 \geq 30$.
Quan sát tỉ lệ các phần tử loại $A$ trên hai mẫu lấy ra từ hai tổng thể.

- Trên tổng thể 1: tỉ lệ các phần tử loại $A$ là $p_1$, mẫu cỡ $n_1$, tần suất $\hat{p}_1$.
- Trên tổng thể 2: tỉ lệ các phần tử loại $A$ là $p_2$, mẫu cỡ $n_2$, tần suất $\hat{p}_2$.

Hãy kiểm định

a) $\begin{cases} H_0 : p_1 = p_2 \\ H_1 : p_1 < p_2 \end{cases}$ b) $\begin{cases} H_0 : p_1 = p_2 \\ H_1 : p_1 > p_2 \end{cases}$ c) $\begin{cases} H_0 : p_1 = p_2 \\ H_1 : p_1 \neq p_2 \end{cases}$

với mức ý nghĩa $\alpha$.

- Gọi $X_1$ và $X_2$ là số phần tử loại $A$ trong mẫu 1 và mẫu 2. Khi đó, $X_1 \sim B(n_1, p_1)$ và $X_2 \sim B(n_2, p_2)$. Đặt

$$\hat{P}_1 = \frac{X_1}{n_1}, \hat{P}_2 = \frac{X_2}{n_2}, \hat{P} = \frac{X_1 + X_2}{n_1 + n_2}$$

- Ta chọn thống kê

$$Z = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P}(1-\hat{P})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

làm tiêu chuẩn kiểm định. Nếu giả thuyết $H_0$ đúng thì $Z \sim N(0,1)$. Từ đây ta suy ra miền bác bỏ tương ứng với từng loại đối thuyết.

XÁC SUẤT THỐNG KÊ
N.T.M.Ngọc
Kiểm định giả thuyết thống kê
Kiểm định giả thuyết về so sánh kì vọng
Kiểm định giả thuyết về so sánh kì vọng với 1 số
Kiểm định giả thuyết về so sánh hai kì vọng
Kiểm định giả thuyết về so sánh tỷ lệ
Kiểm định giả thuyết về so sánh tỷ lệ với một số
Kiểm định giả thuyết về so sánh hai tỷ lệ

| Giả thuyết | Miền bác bỏ $H_0$ |
|---|---|
| $H_0: p_1 = p_2$<br>$H_1: p_1 < p_2$ | $W_\alpha = \left\{ z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} : z < z_\alpha \right\}$ |
| $H_0: p_1 = p_2$<br>$H_1: p_1 > p_2$ | $W_\alpha = \left\{ z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} : z > z_{1-\alpha} \right\}$ |
| $H_0: p_1 = p_2$<br>$H_1: p_1 \neq p_2$ | $W_\alpha = \left\{ z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} : \lvert z \rvert > z_{1-\alpha/2} \right\}$ |

XÁC SUẤT THỐNG KÊ
N.T.M.Ngọc
Kiểm định giả thuyết thống kê
Kiểm định giả thuyết về so sánh kì vọng
Kiểm định giả thuyết về so sánh kì vọng với 1 số
Kiểm định giả thuyết về so sánh hai kì vọng
Kiểm định giả thuyết về so sánh tỷ lệ
Kiểm định giả thuyết về so sánh tỷ lệ với một số
Kiểm định giả thuyết về so sánh hai tỷ lệ

Ví dụ :

Kiểm tra ngẫu nhiên sản phẩm sản xuất từ hai cơ sở ta có số liệu

- Cơ sở 1: Có 20 phế phẩm trong 1000 sản phẩm kiểm tra.
- Cơ sở 2: Có 30 phế phẩm trong 900 sản phẩm kiểm tra.

Với mức ý nghĩa $\alpha = 0.05$ có thể coi rằng tỉ lệ phế phẩm của hai cơ sở sản xuất trên như nhau hay không?

XÁC SUẤT THỐNG KÊ
N.T.M.Ngọc
Kiểm định giả thuyết thống kê
Kiểm định giả thuyết về so sánh kì vọng
Kiểm định giả thuyết về so sánh kì vọng với 1 số
Kiểm định giả thuyết về so sánh hai kì vọng
Kiểm định giả thuyết về so sánh tỷ lệ
Kiểm định giả thuyết về so sánh tỷ lệ với một số
Kiểm định giả thuyết về so sánh hai tỷ lệ

## Kiểm định giả thuyết về tính độc lập

- **Bài toán:**
  - Giả sử mỗi phần tử trong một tổng thể có thể được phân loại theo hai đặc tính khác nhau, gọi là đặc tính $X$ và đặc tính $Y$. $X$ có $r$ giá trị và $Y$ có $s$ giá trị. Gọi

$$P_{ij} = \mathbb{P}(X = x_i, Y = y_j)$$

  với $i = 1, \ldots, r$ và $j = 1, \ldots, s$. $P_{ij}$ là xác suất chọn được một phần tử trong tổng thể có đặc tính $X$ bằng $i$ và đặc tính $Y$ bằng $j$.
  - Gọi

$$p_i = \mathbb{P}(X = x_i) = \sum_{j=1}^{s} P_{ij}, \quad i = 1, \ldots, r$$

và

$$q_j = \mathbb{P}(Y = y_j) = \sum_{i=1}^{r} P_{ij}, \quad j = 1, \ldots, s$$

## Kiểm định giả thuyết về tính độc lập

$p_i$ là xác suất chọn được một phần tử của tổng thể có đặc tính $X$ bằng $x_i$, $q_j$ là xác suất chọn được môt phần tử của tổng thể có đặc tính $Y$ bằng $y_j$.

- Ta cần kiểm định xem $X$ có độc lập với $Y$ hay không? Phát biểu giả thuyết

$$H_0 : P_{ij} = p_i q_j \quad \forall i = 1, \ldots, r;\ j = 1, \ldots, s$$

và đối thuyết

$$H_1 : \exists\ (i, j) \text{ sao cho } P_{ij} \neq p_i q_j$$



## Kiểm định giả thuyết về tính độc lập

- Khảo sát $N$ phần tử, ta được bảng kết quả, trong bài toán này gọi là bảng ngẫu nhiên (contingency table):

| $X$ \ $Y$ | $y_1$ | $y_2$ | $\cdots$ | $y_s$ | Tổng hàng |
|---|---|---|---|---|---|
| $x_1$ | $n_{11}$ | $n_{12}$ | $\cdots$ | $n_{1s}$ | $n_1$ |
| $x_2$ | $n_{21}$ | $n_{22}$ | $\cdots$ | $n_{2s}$ | $n_2$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $x_r$ | $n_{r1}$ | $n_{r2}$ | $\cdots$ | $n_{rs}$ | $n_r$ |
| Tổng cột | $m_1$ | $m_2$ | $\cdots$ | $m_s$ | $N$ |

Bảng:

trong đó, các $n_{ij}$ gọi là tần số thực nghiệm.



## Kiểm định giả thuyết về tính độc lập

- Ước lượng của $p_i$ và $q_j$ lần lượt bằng

$$\hat{p}_i = \frac{n_i}{N}, \quad i = 1, \ldots, r$$
$$\hat{q}_j = \frac{m_j}{N}, \quad j = 1, \ldots, s$$

- Gọi $N_{ij}$ là số phần tử có đặc tính $(x_i, y_j)$ trong $N$ phần tử khảo sát, thì $N_{ij} \sim B(N, P_{ij})$. Khi đó,

$$\mathbb{E}(N_{ij}) = NP_{ij} = Np_i q_j \ \text{ khi } H_0 \text{ đúng}$$

Đặt

$$e_{ij} = N\hat{p}_i \hat{q}_j = \frac{n_i m_j}{N}$$

$e_{ij}$ gọi là tần số lý thuyết.



## Kiểm định giả thuyết về tính độc lập

[Pearson] Với $N_{ij}$ và $E_{ij} = NP_{ij}$, biến ngẫu nhiên

$$\sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(N_{ij} - E_{ij})^2}{E_{ij}}$$

sẽ hội tụ theo phân phối về biến ngẫu nhiên Chi bình phương $\chi^2_{(r-1)(s-1)}$ bậc tự do.

## Kiểm định giả thuyết về tính độc lập

**Các bước kiểm định**

1. Phát biểu giả thuyết $H_0$: $X$ và $Y$ độc lập
2. Xác định tần số thực nghiệm $n_{ij}$ và tần số lý thuyết

$$e_{ij} = \frac{n_i m_j}{N}$$

với $n_i$ và $m_j$ là tổng hàng $i$ và tổng cột $j$ tương ứng,
Điều kiện: $e_{ij} \geq 5$.



## Kiểm định giả thuyết về tính độc lập

3. Tính thống kê kiểm định

$$Q^2 = \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{(n_{ij}-e_{ij})^2}{e_{ij}} = \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{n_{ij}^2}{e_{ij}} - N \quad (1)$$

Nếu $H_0$ đúng, thống kê $Q^2$ có phân phối Chi bình phương với $(r-1)(s-1)$ bậc tự do

4. Bác bỏ $H_0$ khi

$$Q^2 > \chi^2_{(r-1)(s-1)}(\alpha) \quad (2)$$

4b. Sử dụng $p$-giá trị:

$$p = \mathbb{P}\left(\chi^2_{(r-1)(s-1)} \geq Q^2\right) \quad (3)$$

Bác bỏ $H_0$ khi: $p \leq \alpha$.



## Kiểm định giả thuyết về tính độc lập

Một báo cáo khoa học trong y khoa tuyên bố rằng việc sở hữu một thú cưng trong nhà (chó hoặc mèo) sẽ làm tăng khả năng sống sót của những người chủ mà thường bị lên cơn đau tim. Một mẫu ngẫu nhiên gồm 95 người đã lên cơn đau tim được chọn để khảo sát. Dữ liệu của mỗi người khảo sát được chia làm 2 loại:

- Những người sống sót/tử vong 1 năm sau khi lên cơn đau tim.
- Người sống sót/tử vong có nuôi thú cưng trong nhà hay không.

Kết quả cho bởi bảng sau

| | Có nuôi thú cưng | Không nuôi thú cưng |
|---|---|---|
| Sống sót | 28 | 44 |
| Tử vong | 8 | 15 |



## Kiểm định giả thuyết về tính độc lập

1. Phát biểu giả thuyết, $H_0$: Bệnh lên cơn đau tim độc lập với việc nuôi thú cưng,
2. Tính tần số thực nghiệm: với $n_1 = 72$, $n_2 = 23$, $m_1 = 36$, $m_2 = 59$

$$e_{11} = \frac{n_1 m_1}{N} = \frac{72\times 36}{95} = 27.284; \qquad e_{12} = \frac{n_1 m_2}{N} = \frac{72\times 59}{95} = 44.716$$

$$e_{21} = \frac{n_2 m_1}{N} = \frac{23\times 36}{95} = 8.716; \qquad e_{22} = \frac{n_2 m_2}{N} = \frac{23\times 59}{95} = 14.284$$

3. Tính giá trị thống kê $Q^2$

$$Q^2 = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{n_{ij}^2}{e_{ij}} - n = \left(\frac{28^2}{27.284} + \frac{44^2}{44.716} + \frac{8^2}{8.716} + \frac{15^2}{15.284}\right) - 95 = 0.125$$

## Kiểm định giả thuyết về tính độc lập

4. Bác bỏ $H_0$ khi: $Q^2 > \chi^2_{(r-1)(s-1)}(\alpha) = \chi^2_1(0.05)$.
   Tra bảng Chi - bình phương, ta được $\chi^2_1(0.05) = 3.841$.
   $Q^2 = 0.125$, suy ra $Q^2 < 3.841$. Ta kết luận chưa đủ cơ sở để bác bỏ $H_0$ tức là bệnh lên cơn đau tim độc lập với việc nuôi thú cưng.



## Kiểm định giả thuyết về tính độc lập

Vé máy bay của hãng hàng không Việt Nam Airline được chia làm 3 loại: Hạng thường (C), hạng trung (B) và hạng doanh nhân (A). Hành khách đi máy bay của VN Airlines nằm trong 1 trong 2 dạng sau: bay nội địa hoặc quốc tế. Khảo sát 920 hành khách đã bay của hãng, cho kết quả sau:

| | **Loại chuyến bay** | |
|---|---|---|
| **Loại vé** | Nội địa | Quốc tế |
| Hạng thường | 29 | 22 |
| Hạng trung | 95 | 121 |
| Hạng doanh nhân | 518 | 135 |

Có ý kiến cho rằng hành khách mua loại vé nào (A, B, C) sẽ phụ thuộc vào việc người đó bay nội địa hay quốc tế. Với mức ý nghĩa 5%, hãy kiểm tra ý kiến trên.