XÁC SUẤT THỐNG KÊ

PHAN I: VECTO NGẪU NHIÊN

Độc lập khi

$$\boxed{f(x,y) = f_X(x).f_Y(y)}$$

VECTOR NN LIÊN TỤC

Sharing is learning

- **Tính chất:** $\displaystyle\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)dxdy = 1$
- **Hàm mật độ thành phần:**
  - **Hàm mật độ của X**

    $$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)dy$$

  - **Hàm mật độ của Y**

    $$f_Y(y) = \int_{-\infty}^{+\infty} f(x,y)dx$$

  - **Trung bình thành phần của X, Y:**

    $$E(X) = \int_{-\infty}^{+\infty} xf_X(x)dx = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xf(x,y)dxdy$$

    $$E(Y) = \int_{-\infty}^{+\infty} yf_Y(x)dy = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} yf(x,y)dxdy$$

- **Hàm mật độ xác suất có điều kiện:**
  - **Của X khi đã biết Y = y**

    $$f_X\left(x\middle|y\right) = \frac{f(x,y)}{f_Y(y)}.$$

  - **Của Y khi đã biết X = x**

    $$f_Y\left(y\middle|x\right) = \frac{f(x,y)}{f_X(x)}.$$

Vector ngẫu nhiên rời rạc

Bảng phân phối xác suất đồng thời của X và Y

X \ Y	0	1	2
0	0.10	0.05	0.15
1	0.05	0.15	0.10
2	0.20	0.10	0.10

Tính chất:

$$\sum_{\substack{x\in\Omega(X)\\ y\in\Omega(Y)}} P_{(X,Y)}(x,y) = 1$$

Có những yêu cầu điển hình như sau:

3. Hỏi X, Y có độc lập nhau không? Vì sao?

$$X, Y \text{ độc lập} \iff P_{(X,Y)}(x,y) = P_X(x)\cdot P_Y(y) \quad \forall x,y$$

Chứng minh X,Y phụ thuộc nhau: Chỉ ra một cặp (x,y) sao cho

$$P_{(X,Y)}(x,y) \neq P_X(x)\cdot P_Y(y).$$

PHAN II: UOC LUONG

ƯỚC LƯỢNG KHOẢNG TRUNG BÌNH

Một số kí hiệu cần nhớ :

α : Mức ý nghĩa

$1 - \alpha$: Độ tin cậy

$\bar{x}$: Trung bình mẫu

σ : Độ lệnh chuẩn của tổng thể

ε : Độ chính xác của ước lượng

s : Độ lệnh chuẩn của mẫu đã hiệu chỉnh

ƯỚC LƯỢNG KHOẢNG TRUNG BÌNH

Một số kí hiệu cần nhớ :

s : Độ lệnh chuẩn của mẫu đã hiệu chỉnh

- Nếu đề cho bảng số liệu:

$$s = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n-1}}$$

- Nếu đề cho độ lệnh chuẩn của mẫu chưa hiệu chỉnh($\hat{s}$)

$$s = \sqrt{\frac{n}{n-1}}\hat{s}$$

- n: kích thước của mẫu (cỡ mẫu).
- s: độ lệch chuẩn đã hiệu chỉnh.
- $\hat{s}$: độ lệch chuẩn chưa hiệu chỉnh.

a) ***Trường hợp* 1.** Kích thước mẫu $n \geq 30$ và phương sai tổng thể σ^2 đã biết.

- Từ mẫu ta tính $\overline{x}$ (trung bình mẫu).
- Từ $1-\alpha \Rightarrow \dfrac{1-\alpha}{2} = \varphi(t_\alpha) \xrightarrow{\textit{tra bảng B}} t_\alpha$.
- Khoảng ước lượng là:

$$\boxed{\left(\overline{x}-\varepsilon;\, \overline{x}+\varepsilon\right),\ \varepsilon = t_\alpha . \frac{\sigma}{\sqrt{n}}.}$$

b) ***Trường hợp* 2.** Kích thước mẫu $n \geq 30$ và phương sai tổng thể σ^2 chưa biết.

- Tính $\overline{x}$ và s (độ lệch chuẩn mẫu đã hiệu chỉnh).
- Từ $1-\alpha \Rightarrow \dfrac{1-\alpha}{2} = \varphi(t_\alpha) \xrightarrow{\textit{tra bảng B}} t_\alpha$.
- Khoảng ước lượng là:

$$\boxed{\left(\overline{x}-\varepsilon;\, \overline{x}+\varepsilon\right),\ \varepsilon = t_\alpha \frac{s}{\sqrt{n}}.}$$

<u>*Chú ý*</u>

Mối liên hệ giữa độ lệch chuẩn mẫu đã hiệu chỉnh s và chưa hiệu chỉnh $\hat{s}$ là:

$$\boxed{s^2 = \frac{n}{n-1}\hat{s}^2 \Rightarrow s = \sqrt{\frac{n}{n-1}\hat{s}^2}.}$$

c) ***Trường hợp*** 3. Kích thước mẫu $n < 30$, σ^2 đã biết và ***X có phân phối chuẩn*** thì ta làm như trường hợp 1.

Chương 6. Mẫu thống kê & Ước lượng tham số

d) ***Trường hợp*** 4. Kích thước mẫu $n < 30$, σ^2 chưa biết và *X có phân phối chuẩn*.

- Từ mẫu ta tính $\overline{x}$, s.
- Từ $1-\alpha \Rightarrow \alpha \xrightarrow{\text{tra bảng } C} t_{\alpha}^{n-1}$

  (nhớ giảm bậc thành $n-1$ rồi mới tra bảng!)
- Khoảng ước lượng là:

$$\boxed{\left(\overline{x}-\varepsilon; \overline{x}+\varepsilon\right), \varepsilon = t_{\alpha}^{n-1}.\frac{s}{\sqrt{n}}.}$$

ƯỚC LƯỢNG KHOẢNG CHO TỈ LỆ TỔNG THỂ

Một TH duy nhất !

- *B1 : Tìm $f = \dfrac{m}{n}$ trong đó:*
  - *+ f : Tỉ lệ mẫu.*
  - *+ m : Số phần tử ta quan tâm (số con cá được đánh dấu) !*
  - *+ n : Cỡ mẫu (số con cá bắt ra) !*
- B2 : Từ $1-\alpha \Rightarrow (1-\alpha)/2 = \Phi(t_\alpha) \xrightarrow{\text{Tra bảng Laplace}} t_\alpha$
- B3 : Tính $\varepsilon = t_\alpha \sqrt{\dfrac{f(1-f)}{n}}$
- B4 : Khoảng ước lượng là: $[f-\varepsilon; f+\varepsilon]$

PHAN III: KIỂM ĐỊNH

KIỂM ĐỊNH VỀ TRUNG BÌNH

TH1: Đã biết độ lệch chuẩn tổng thể

- Bài toán 1: $H_0: \mu = \mu_0, \quad H_1: \mu \neq \mu_0$
  Bác bỏ H_0 nếu $|z| > z_{\alpha/2}$
  Chấp nhận H_0 nếu $|z| \le z_{\alpha/2}$
- Bài toán 2: $H_0: \mu = \mu_0, \quad H_1: \mu > \mu_0$
  Bác bỏ H_0 nếu $z > z_\alpha$
  Chấp nhận H_0 nếu $z \le z_\alpha$
- Bài toán 3: $H_0: \mu = \mu_0, \quad H_1: \mu < \mu_0$
  Bác bỏ H_0 nếu $z < -z_\alpha$
  Chấp nhận H_0 nếu $z \ge -z_\alpha$

Đặt $z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$

VD 1. Sở Điện lực *A* báo cáo rằng: trung bình một hộ hàng tháng phải trả 250 ngàn đồng tiền điện, với độ lệch chuẩn là 20 ngàn. Người ta khảo sát ngẫu nhiên 500 hộ thì tính được trung bình hàng tháng một hộ trả 252 ngàn đồng tiền điện.
Trong kiểm định giả thuyết H: "**trung bình một hộ phải trả hàng tháng là 250 ngàn đồng tiền điện**" với mức ý nghĩa 1%, hãy cho biết giá trị thống kê *t* và kết luận ?

KIỂM ĐỊNH VỀ TRUNG BÌNH

Theo đề bài ta có:
- $+ \mu_0 = 250\,000$đ
- $+ \sigma = 20\,000$đ
- + n = 500
- $+ \bar{x} = 252\,000$
- + α = 0,01

Ta kiểm định:

Giả thuyết H_0: $\mu = 250\,000$

Đối thuyết H_1: $\mu \neq 250\,000$

- Bài toán 1: $H_0: \mu = \mu_0, \quad H_1: \mu \neq \mu_0$
  Bác bỏ H_0 nếu $|z| > z_{\alpha/2}$
  Chấp nhận H_0 nếu $|z| \le z_{\alpha/2}$

- B1 : B1 : Tính $\bar{x}$ và s (đề đã cho)
- B2 : Từ α = 0,01. Ta có $\Phi(z_{\alpha/2}) = 1 - \alpha/2 = 0,995$
  Sau đó dò Bảng Phân Phối Chuẩn $\Rightarrow z_{\alpha/2} = 2,58$
- B3 : Tính $z = \dfrac{\bar{x} - \mu_0}{\sigma}\sqrt{n} = \dfrac{252000 - 250000}{20000}\sqrt{500} = 2,236$
- B4 : Có $|z| \le z_{\alpha/2}$ => Chấp nhận H_0

KIỂM ĐỊNH VỀ TRUNG BÌNH

TH2: Chưa biết độ lệch chuẩn tổng thể và n ≥ 30

Đặt $z = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$

- Bài toán 1: $H_0: \ \mu = \mu_0, \quad H_1: \ \mu \neq \mu_0$
  Bác bỏ H_0 nếu $|z| > z_{\alpha/2}$
  Chấp nhận H_0 nếu $|z| \leq z_{\alpha/2}$
- Bài toán 2: $H_0: \ \mu = \mu_0, \quad H_1: \ \mu > \mu_0$
  Bác bỏ H_0 nếu $z > z_\alpha$
  Chấp nhận H_0 nếu $z \leq z_\alpha$
- Bài toán 3: $H_0: \ \mu = \mu_0, \quad H_1: \ \mu < \mu_0$
  Bác bỏ H_0 nếu $z < -z_\alpha$
  Chấp nhận H_0 nếu $z \geq -z_\alpha$

Sharing is learning

KIỂM ĐỊNH VỀ TRUNG BÌNH

TH3: Chưa biết độ lệch chuẩn tổng thể và n < 30

Đặt $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$

+Bài toán 1 $\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$ Giả thuyết H_0 sẽ bị bác bỏ khi : $t > t_{n-1,\alpha}$

+Bài toán 2 $\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu < \mu_0 \end{cases}$ Giả thuyết H_0 sẽ bị bác bỏ khi : $t > -t_{n-1,\alpha}$

+Bài toán 3 $\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}$ Giả thuyết H_0 sẽ bị bác bỏ khi : $|t| > t_{n-1,\alpha/2}$

***Chú ý**: Giá trị $t_{n-1,\alpha}$ sẽ tra từ bảng phân phối Student*

Sharing is learning

KIỂM ĐỊNH VỀ TRUNG BÌNH

Bài toán: Một trại chăn nuôi gà đã nuôi thí nghiệm bằng khẩu phần thức ăn có bổ sung kháng sinh. Kiểm tra 81 con gà ta có số liệu:

Trọng lượng (kg)	3,8	3,9	4,0	4,1	4,2	4,3	4,4	4,5	4,6	4,7
Số gà	5	7	9	12	15	10	9	6	5	3

a)Trại chăn nuôi báo cáo trọng lượng trung bình của những con gà nuôi thí nghiệm sau 8 tuần nuôi là 4,3kg thì có đúng không với độ tin cậy 95%?

b) Giả sử những con gà có trọng lượng lớn hơn 4,3 kg được xếp loại I và trọng lượng của nó có phân phối chuẩn. với mức ý nghĩa 5%, chúng ta có thể kết luận trọng lượng trung bình của những con gà loại I lớn hơn 4,5 kg được không?

a. Giả thuyết $H_0: \mu = \mu_0 = 4{,}3$

Đối thuyết $H_1: \mu \neq 4{,}3$

Bấm máy ta được $\bar{x}$ = 4,212 và s=0,2358

$$t = \frac{\bar{x} - \mu_0}{s}\sqrt{n} = \frac{4.212 - 4{,}3}{0{,}2358}\sqrt{81} = -3{,}3588$$

$$\Rightarrow t_{\alpha/2} = 1.96$$

Ta thấy | t | > $t_{\alpha/2}$. Ta bác bỏ giả thuyết.

Vậy báo cáo của trại chăn nuôi là không đúng.

b. Giả thuyết $H_0: \mu = \mu_0 = 4{,}5$

Đối thuyết $H_1: \mu > 4{,}5$

Bấm máy ta được $\bar{x}$ = 4,5087 và s=0,1083

$$t = \frac{\bar{x} - \mu_0}{s}\sqrt{n} = \frac{4.5087 - 4{,}5}{0{,}1083}\sqrt{23} = 0.3853$$

$$n - 1 = 22 \Rightarrow t_{n-1,\alpha} = 1.717$$

Ta thấy $t < t_{22{,}0.05}$. Ta chấp nhận H_0.

KL : Với mức ý nghĩa 5%, báo cáo của trại chăn nuôi là không đúng.

KIỂM ĐỊNH VỀ TỈ LỆ

Một TH duy nhất !

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

- Bài toán 1: $H_0 : \ p = p_0, \quad H_1 : \ p \neq p_0$
  Bác bỏ H_0 nếu $|z| > z_{\alpha/2}$
  Chấp nhận H_0 nếu $|z| \leq z_{\alpha/2}$
- Bài toán 2: $H_0 : \ p = p_0, \quad H_1 : \ p > p_0$
  Bác bỏ H_0 nếu $z > z_\alpha$
  Chấp nhận H_0 nếu $z \leq z_\alpha$
- Bài toán 3: $H_0 : \ p = p_0, \quad H_1 : \ p < p_0$
  Bác bỏ H_0 nếu $z < -z_\alpha$
  Chấp nhận H_0 nếu $z \geq -z_\alpha$

Bài toán: Kiểm tra mức độ nghiêm túc của sinh viên trong giờ học.
Khảo sát ngẫu nhiên 400 sinh viên về mức độ nghiêm túc trong giờ học thì thấy 13 sinh viên thừa nhận có ngủ trong giờ học. Trong kiểm định giả thiết H: "Có 2% sinh viên ngủ trong giờ học", cho biết kết luận với mức ý nghĩa 0,05.

IỂM ĐỊNH VỀ TỈ LỆ

Theo đề bài ta có:
+ n = 400 sinh viên
+ m = 13 sinh viên
+ p_0 = 2%
+ α = 0,05

- B1 : Tìm $f = \frac{m}{n} = \frac{13}{400} = 0{,}0325$
- B2 : Từ 1 – α = 95% ⇒ α = 0,05. Ta có $\Phi(z_{\alpha/2}) = 1 - \alpha/2 = 0{,}975$
  Sau đó dò Bảng Phân Phối Chuẩn ⇒ $z_{\alpha/2} = 1{,}96$
- B3 : Tính giá trị thống kê $z = \frac{|f - p_0|}{\sqrt{p_0 q_0}}\sqrt{n} = \frac{|0{,}0325 - 0{,}02|}{\sqrt{0{,}02 . 0{,}98}}\sqrt{400} = 1{,}786$
- B4 : $z \leq z_{\alpha/2}$ => Giả thuyết đúng

PHẦN IV: HỒI QUY TƯƠNG QUAN

HỆ SỐ TƯƠNG QUAN MẪU

Công thức : $r = \dfrac{\overline{xy} - \bar{x}\bar{y}}{\hat{s}_x \hat{s}_y}$; $\overline{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i$

Tính chất : $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$

- $-1 \le r \le 1$
- *Nếu r = 0 thì X, Y không có quan hệ tuyến tính (nhưng không có nghĩa là độc lập) !*
- *Nếu r = ±1 thì X,Y có quan hệ tt tuyệt đối.*
- *r < 0 : Quan hệ giữa X,Y giảm biến*
- *r > 0 : Quan hệ giữa X,Y đồng biến*

PHƯƠNG TRÌNH HỒI QUY TUYẾN TÍNH

Đường hồi quy tuyến tính của Y theo X là :

$$Y = a + bX$$

Với $b = \dfrac{\overline{xy} - \bar{x}\bar{y}}{\hat{s}_x^2}$

Với $a = \bar{y} - b\,\bar{x}$

Đường hồi quy tuyến tính của X theo Y là :

$$X = a + bY$$

Với $b = \dfrac{\overline{xy} - \bar{x}\bar{y}}{\hat{s}_y^2}$

Với $a = \bar{x} - b\bar{y}$