BAN HỌC TẬP ĐOÀN KHOA CNPM

Chuỗi Training cuối kì II năm học 2021 – 2022

ĐOÀN KHOA CÔNG NGHỆ PHẦN MỀM

BAN HỌC TẬP

Sharing is learning

Ban học tập

Khoa Công Nghệ Phần Mềm
Trường ĐH Công Nghệ Thông Tin

Email / Group

Bht.cnpm.uit@gmail.com
fb.com/groups/bht.cnpm.uit/

Training cuối kì

Xác suất thống kê

10h ngày 16/6/2022

Giảng đường 1 (A1)

Trainer: Nguyễn Bích Phượng - 21522884 – CNCL 2021.2
Huỳnh Tiến Phát – 21520388 – MTIO2021

CẤU TRÚC ĐỀ THI

CẤU TRÚC ĐỀ THI NĂM TRƯỚC

- Dạng 1 : Xác suất. Xấp xỉ xác suất.
- Dạng 2 : Vector ngẫu nhiên 2 chiều.
- Dạng 3 : Ước lượng khoảng, ước lượng tỉ lệ, kiểm định.
- Dạng 4: Phương trình hồi quy tuyến tính.

NỘI DUNG TRAINING

NỘI DUNG TRAINING

- Phân phối nhị thức, phân phối chuẩn, xấp xỉ các phân phối thành PP chuẩn
- Vector ngẫu nhiên 2 chiều
  - Rời rạc
  - Liên tục
- Ước lượng:
  - Ước lượng khoảng trung bình
  - Ước lượng tỉ lệ
- Kiểm định:
  - Kiểm định trung bình
  - Kiểm định tỉ lệ

NỘI DUNG TRAINING

- Bài toán tương quan & Hồi quy
  - Hệ số tương quan mẫu r
  - Phương trình hồi quy tuyến tính

PHẦN 1: ÔN TẬP PHÂN PHỐI

PHÂN PHỐI NHỊ THỨC

Ôn lại kiến thức Phân phối Nhị Thức ~ B(n; p)

VD: Một người chơi lô tô n lần với xác suất thắng là p (và q = 1 – p).

a) Tính xác suất để người đó thắng k lần.

$$\boxed{P(X = k) = C_n^k p^k q^{n-k}}$$

b) Tính xác suất để người đó thắng từ a đến b lần, $(a \leq b)$

$$\boxed{P(a \leq X \leq b) = \sum_{i=a}^{b} C_n^i p^i q^{n-i}}$$

HÀM MẬT ĐỘ PP CHUẨN

- X ~ N(0;1)   (Chuẩn đơn giản – PP Gauss)

  Hàm mật độ XS có dạng: $f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}, t \in \mathbb{R}$

- X ~ N(μ; σ^2) $\boxed{P(a \le X \le b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}$

ModX = EX = μ (Trung bình)
VarX = σ^2 (Phương sai)

Chú ý tính chất của hàm Laplace:

- Φ(-x) = -Φ(x) vì hàm Φ(x) lẻ
- Φ(-∞) = -0,5 và Φ(+∞) = +0,5
- Nếu x ≥ 4 thì Φ(x) ≈ 0,5

CÁCH SỬ DỤNG BẢNG PPXS

Bảng B: tích phân Laplace

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706

VD: $\Phi(1.24) = \Phi(1.2 + 0.04)$

Tra hàng   Tra cột

$\Rightarrow \Phi(1.24) = \boxed{0.3925}$

XÁC SUẤT CỦA PHÂN PHỐI CHUẨN

Ví dụ 1 **(Đề CK 2017–2018):** Đường kính (ĐK) của một loại trục máy tiện làm ra là một đại lượng ngẫu nhiên có phân phối chuẩn với giá trị trung bình là 25mm và phương sai 44,1mm^2. Trục máy được xem là đạt tiêu chuẩn trong khoảng từ 23,44mm đến khoảng 26,56mm.

a) Tìm tỉ lệ trục máy đạt tiêu chuẩn kĩ thuật.

b) Phải sản xuất ra ít nhất bao nhiêu trục để khả năng có ít nhất 1 trục đạt tiêu chuẩn kĩ thuật không dưới 99,73%.

XÁC SUẤT CỦA PHÂN PHỐI CHUẨN

Sharing is learning

Ví dụ 1 **(Đề CK 2017–2018):** Đường kính (ĐK) của một loại trục máy tiện làm ra là một đại lượng ngẫu nhiên có phân phối chuẩn với giá trị trung bình là 25mm và phương sai 44,1mm². Trục máy được xem là đạt tiêu chuẩn trong khoảng từ 23,44mm đến khoảng 26,56mm.

a) Tìm tỉ lệ trục máy đạt tiêu chuẩn kĩ thuật.

b) Phải sản xuất ra ít nhất bao nhiêu trục để khả năng có ít nhất 1 trục đạt tiêu chuẩn kĩ thuật không dưới 99,73%.

Theo đề bài ta có:

+ ĐK có phân phối chuẩn $\begin{cases} \mu = 25 \\ \sigma^2 = 44{,}1 \end{cases}$

+ ĐK đạt tiêu chuẩn $\in$ [23,44; 26,56]

XÁC SUẤT CỦA PHÂN PHỐI CHUẨN

Theo đề bài ta có:

+ ĐK có phân phối chuẩn $\begin{cases} \mu = 25 \\ \sigma^2 = 44{,}1 \end{cases}$

+ ĐK đạt tiêu chuẩn ∈ [23,44; 26,56]

Công thức:

$$P(a \le X \le b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

a) Tìm tỉ lệ trục máy đạt tiêu chuẩn kĩ thuật.

- Gọi X là đường kính của trục máy tiện → $X \sim N(25; 44{,}1)$
- Tỉ lệ trục máy đạt tiêu chuẩn, tức là tìm $P(23{,}44 \le X \le 26{,}56)$

$$P(23{,}44 \le X \le 26{,}56) = \Phi\left(\frac{26{,}56 - 25}{\sqrt{44{,}1}}\right) - \Phi\left(\frac{23{,}44 - 25}{\sqrt{44{,}1}}\right)$$

$$= \Phi(0{,}23) - \Phi(-0{,}23) = 0{,}5910 - 0{,}4090 = 0{,}182$$

XÁC SUẤT CỦA PHÂN PHỐI CHUẨN

b) Phải sản xuất ra ít nhất bao nhiêu trục để khả năng có ít nhất 1 trục đạt tiêu chuẩn kĩ thuật không dưới 99,73%.

- Phải sản xuất ra ít nhất bao nhiêu trục (tức là tìm n)
- Tỉ lệ trục máy đạt tiêu chuẩn kĩ thuật là: 0,182 (tức là p)

$\}$ **B(n; p)**

- Gọi Y là số trục đạt tiêu chuẩn → **Y~B(n; 0,182)**
- Ta có : $P(Y \geq 1) \geq 99{,}73\ \%$

$$\rightarrow 1 - P(Y = 0) \geq 99{,}73\%$$

$$\boxed{P(X = k) = C_n^k p^k q^{n-k}}$$

$$\rightarrow 1 - C_n^0 p^0 q^n \geq 99{,}73\ \%$$

$$(C_n^0 p^0 = 1)$$

$$\rightarrow 1.q^n \leq 0{,}0027 \rightarrow (1 - 0{,}182)^{\mathbf{n}} \leq 0{,}0027 \rightarrow n \leq 29{,}4$$

Vậy số trục cần sản xuất để thoả mãn đề bài là **30 trục**

PHẦN 2: XẤP XỈ PHÂN PHỐI

SIÊU BỘI → NHỊ THỨC

- **Ứng dụng tốt khi n rất nhỏ so với N (n < 5%.N)**

$H(N; N_A; n)$

↓

$B(n; p)$

VD: Trong hộp có N phần tử, với N_A phần tử có thuộc tính A. Lấy n phần tử ra. Tính XS lấy được k phần tử có thuộc tính A.

Đặt $p = \dfrac{N_A}{N}$

Khi đó : $\dfrac{C_{N_A}^{k} C_{N-N_A}^{n-k}}{C_N^n} \xrightarrow{d} C_n^k p^k q^{n-k}$

NHỊ THỨC → CHUẨN (QUAN TRỌNG)

- Ứng dụng tốt khi **np ≥ 5 và n(1-p) ≥ 5**

$B(n; p)$

↓

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

VD: Một người chơi lô tô với xác suất thắng là p. Người đó chơi n lần.

a) Tính xác suất để người đó thắng từ k1 đến k2 lần.

b) Tính xác suất để người đó thắng k lần.

Đặt $\mu = np$, $\sigma^2 = npq$ (với $q = 1-p$)

$$\boxed{P(k_1 \le X \le k_2) = \varphi\left(\frac{k_2-\mu}{\sigma}\right) - \varphi\left(\frac{k_1-\mu}{\sigma}\right).}$$

$$\boxed{P(X = k) \approx P(k - 0,5 \le X \le k + 0,5).}$$

NHỊ THỨC → CHUẨN (QUAN TRỌNG)

Ví dụ 2 (Đề thi cuối kì 2018 – 2019): Xác suất virus máy tính V có thể gây hại cho một tập tin bất kì là 35%. Giả sử virus V xâm nhập vào một thư mục gồm 2400 tập tin. Tính xác suất có từ 800 đến 850 tập tin bị nhiễm virus.

NHỊ THỨC → CHUẨN (QUAN TRỌNG)

Ví dụ 2 **(Đề thi cuối kì 2018 – 2019):** Xác suất virus máy tính V có thể gây hại cho một tập tin bất kì là 35%. Giả sử virus V xâm nhập vào một thư mục gồm 2400 tập tin. Tính xác suất có từ 800 đến 850 tập tin bị nhiễm virus.

Theo đề bài ta có:

+ $p = 35\%$
+ $n = 2400$ (ta thấy n rất lớn)
+ Tính xs từ 800 đến 850 **tập tin bị nhiễm virus**

NHỊ THỨC → CHUẨN (QUAN TRỌNG)

Theo đề bài ta có:

+ $p = 35\% \rightarrow q = 1 - p = 65\%$
+ $n = 2400$ (ta thấy n rất lớn)
+ Tính xs từ 800 đến 850 **tập tin bị nhiễm virus**

B(n; p) → N(**μ**; **σ**²)

$$\mu = np$$
$$\sigma^2 = npq$$

$$P(a \le X \le b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

- Gọi X là số tập tin bị nhiễm virus → X ~ B(**2400**; 0,35)
- Nhận xét **np = 2400 . 0,35 > 5 và nq = 2400 . 0,65 > 5**

Ta dùng xấp xỉ phân phối từ Nhị Thức sang Chuẩn với: $\begin{cases} \mu = np = \mathbf{840} \\ \sigma^2 = npq = \mathbf{546} \end{cases}$

X ~ N(**840**; **546**)

$$P(800 \le X \le 850) = \Phi\left(\frac{850-840}{\sqrt{546}}\right) - \Phi\left(\frac{800-840}{\sqrt{546}}\right)$$

$$= \Phi(0,43) - \Phi(-1,71) = 0,6664 - 0,0436 = 0,6228$$

PHẦN 3: VECTOR NN 2 CHIỀU

VECTOR NN RỜI RẠC

Sharing is learning

❖ **Bảng phân phối xác suất đồng thời của (X,Y)**

$X \backslash Y$	y_1	y_2	$\dots$	y_j	$\dots$	y_n	Tổng dòng
x_1	p_{11}	p_{12}	$\dots$	p_{1j}	$\dots$	p_{1n}	$p_{1\bullet}$
x_2	p_{21}	p_{22}	$\dots$	p_{2j}	$\dots$	p_{2n}	$p_{2\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_i	p_{i1}	p_{i2}	$\dots$	p_{ij}	$\dots$	p_{in}	$p_{i\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_m	p_{m1}	p_{m2}	$\dots$	p_{mj}	$\dots$	p_{mn}	$p_{m\bullet}$
Tổng cột	$p_{\bullet 1}$	$p_{\bullet 2}$	$\dots$	$p_{\bullet j}$	$\dots$	$p_{\bullet n}$	1

$$p_{i,j} = P(X = x_i, Y = y_j)$$

❑ **Bảng phân phối xác suất của X (tổng dòng ngang)**

X	x_1	x_2	$\dots$	x_m
P	$p_{1\bullet}$	$p_{2\bullet}$	$\dots$	$p_{m\bullet}$

- Kì vọng X: $\boxed{EX = x_1 p_{1\bullet} + x_2 p_{2\bullet} + \dots + x_m p_{m\bullet}.}$

❑ **Bảng phân phối xác suất của Y (tổng cột dọc)**

Y	y_1	y_2	$\dots$	y_n
P	$p_{\bullet 1}$	$p_{\bullet 2}$	$\dots$	$p_{\bullet n}$

- Kỳ vọng Y: $\boxed{EY = y_1 p_{\bullet 1} + y_2 p_{\bullet 2} + \dots + y_n p_{\bullet n}.}$

VECTOR NN RỜI RẠC

❑ Xác suất có điều kiện của X :

$$P\left(X=x_i \middle| Y=y_j\right)=\frac{P(X=x_i,\,Y=y_j)}{P(Y=y_j)}=\frac{p_{ij}}{p_{\bullet j}}$$

Kỳ vọng của X với điều kiện $Y = y_j$

$$E(X|Y=y_j)=x_1\frac{p_{1j}}{p_{*j}}+x_2\frac{p_{2j}}{p_{*j}}+\cdots+x_n\frac{p_{nj}}{p_{*j}}.$$

❑ Xác suất có điều kiện của Y :

$$P\left(Y=y_j \middle| X=x_i\right)=\frac{P(X=x_i,\,Y=y_j)}{P(X=x_i)}=\frac{p_{ij}}{p_{i\bullet}}$$

Kỳ vọng của Y với điều kiện $X = x_i$

$$E(Y\,|\,X=x_i)=y_1\frac{p_{i1}}{p_{i*}}+y_2\frac{p_{i2}}{p_{i*}}+\ldots+y_n\frac{p_{in}}{p_{i*}}.$$

VECTOR NN RỜI RẠC

Ví dụ 3 **(Đề thi cuối kì 2018– 2019):** Cho X và Y là số lần phần cứng bị hỏng trong 2 phòng A và B trong một tháng. Phân phối đồng thời của X và Y được cho bởi bảng sau:

P(x, y)		x		
		0	1	2
y	0	0.52	0.20	0.04
	1	0.14	0.02	0.01
	2	0.06	0.01	0

a) Tính xác suất $P(X + Y \geq 1)$.
b) X và Y có độc lập không vì sao ?
c) Giả sử phòng A bị hỏng phần cứng trong tháng 1, tính xác suất phòng B cũng bị hỏng phần cứng trong tháng 1.

VECTOR NN RỜI RẠC

a) Tính xác suất P(X + Y ≥ 1)

P(x, y)		x		
		0	1	2
y	0	0.52	0.20	0.04
	1	0.14	0.02	0.01
	2	0.06	0.01	0

$$
\begin{aligned}
&P(X+Y\ge 1)\\
&=P(X=0,Y=1)+P(X=0,Y=2)\\
&+P(X=1,Y=0)+P(X=1,Y=1)\\
&+P(X=1,Y=2)+P(X=2,Y=0)\\
&+P(X=2,Y=1)+P(X=2,Y=2)\\
&=0{,}2+0{,}04+0{,}14+0{,}02+0{,}01+0{,}06+0{,}01=0{,}48
\end{aligned}
$$

$$
\begin{aligned}
&P(X+Y\ge 1)\\
&=1-P(X+Y<1)\\
&=1-P(X=0,Y=0)\\
&=1-0{,}52=0{,}48
\end{aligned}
$$

VECTOR NN RỜI RẠC

b) X và Y có độc lập không ? Vì sao ?

P(x, y)		x		
		0	1	2
y	0	0.52	0.20	0.04
	1	0.14	0.02	0.01
	2	0.06	0.01	0

P(X = x) . P(Y = y) = P(X = x, Y = y)

Nếu chứng minh **không độc lập**, ta chỉ cần chỉ rõ trường hợp **đẳng thức trên là sai**.

$$\mathbf{P(X=0)} = P(X=0, Y=0) + P(X=0, Y=1) + P(X=0, Y=2) = 0,72$$

$$\mathbf{P(Y=0)} = P(X=0, Y=0) + P(X=1, Y=0) + P(X=2, Y=0) = 0,76$$

Ta có **P(X = 0) . P(Y = 0) ≠ P(X = 0, Y = 0)** (0,5472 ≠ 0,52)

Vậy X và Y không độc lập nhau.

VECTOR NN RỜI RẠC

c) Giả sử phòng A bị hỏng phần cứng trong tháng 1, tính xác suất phòng B cũng bị hỏng phần cứng trong tháng 1.

P(x, y)		x		
		0	1	2
y	0	0.52	0.20	0.04
	1	0.14	0.02	0.01
	2	0.06	0.01	0

Phòng A bị lỗi phần cứng $\boxed{X \geq 1}$

Phòng B bị lỗi phần cứng $\boxed{Y \geq 1}$

$$P\left(Y=y_j \middle| X=x_i\right) = \frac{P(X=x_i, Y=y_j)}{P(X=x_i)} = \frac{p_{ij}}{p_{i\bullet}}$$

P(Y ≥ 1 [redacted] ≥ 1) $= \frac{P(X \geq 1, Y \geq 1)}{P(X \geq 1)} = \frac{0{,}02 + 0{,}01 + 0{,}01 + 0}{0{,}28} = \frac{1}{7}$

VECTOR NN LIÊN TỤC

❑ Tính chất: $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)dxdy = 1$

❑ 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:**

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

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

❑ 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 NN LIÊN TỤC

Ví dụ 4 (Đề thi học kì 2018 – 2019): Cho hàm mật độ xác suất đồng thời của X, Y:

$$f(x,y)=\begin{cases} C(x-2)e^{-y} & \text{nếu } 0 \le x \le 2;\ y \ge 0 \\ 0 & \text{các trường hợp khác} \end{cases}$$

a) Tìm C.

b) Tìm hàm mật độ thành phần của X. Tính $P(X < 1)$.

c) Tính xác suất $P(X \le 1, Y > 2)$.

VECTOR NN LIÊN TỤC

$$f(x,y) = \begin{cases} C(x-2)e^{-y} & \text{nếu } 0 \le x \le 2,\ y \ge 0 \\ 0 & \text{nếu } (x,y) \text{ khác} \end{cases}$$

a) Tìm C

$$\boxed{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\,dxdy = 1} \Leftrightarrow \int_0^2 \left[\int_0^{+\infty} C(x-2)e^{-y}dy \right] dx = 1$$

$$\Leftrightarrow C\int_0^2 (x-2)dx \int_0^{+\infty} e^{-y}dy = 1$$

$$\Leftrightarrow C\int_0^2 (x-2)dx = 1 \Rightarrow C = -\frac{1}{2}$$

VECTOR NN LIÊN TỤC

$$f(x,y) = \begin{cases} C(x-2)e^{-y} & \text{nếu } 0 \le x \le 2,\ y \ge 0 \\ 0 & \text{nếu } (x, y) \text{ khác} \end{cases}$$

b) Tìm mật độ thành phần của X. Tính P(X < 1)

$$f_X(x) = \int_0^{+\infty} -\frac{1}{2}(x-2)e^{-y}dy = -\frac{1}{2}(x-2)$$

$$P(X<1) = \int_0^1 f_X(x)dx = \int_0^1 -\frac{1}{2}(x-2)dx = \frac{3}{4}$$

VECTOR NN LIÊN TỤC

$$f(x,y) = \begin{cases} C(x-2)e^{-y} & \text{nếu } 0 \le x \le 2,\ y \ge 0 \\ 0 & \text{nếu } (x, y) \text{ khác} \end{cases}$$

c) Tính xác suất P(X ≤ 1, Y > 2)

$$P(X \le 1, Y > 2) = \int_0^1 \int_2^{+\infty} f(x,y)\,dy\,dx$$

$$= \int_0^1 \left[\int_2^{+\infty} -\frac{1}{2}(x-2)e^{-y}\,dy \right] dx$$

$$= -\frac{1}{2}\int_0^1 (x-2)e^{-2}\,dx = \frac{3}{4}e^{-2}$$

VECTOR NN LIÊN TỤC

Ví dụ 5 Cho hàm mật độ xác suất đồng thời của X, Y:

$$f(x,y) = \begin{cases} 6x & \text{nếu } 0<x<1;\ 0<y<1-x \\ 0 & \text{các trường hợp khác} \end{cases}$$

a) Tính trung bình thành phần của Y.

b) Tính P(X > 0,3 | Y = 0,5).

c) X, Y có độc lập không?

d) Tính P(X + Y ≤ 0,5).

e) Tính P(Y ≥ 0,5).

VECTOR NN LIÊN TỤC

Ví dụ 5 Cho hàm mật độ xác suất đồng thời của X, Y:

$$f(x,y)=\begin{cases}6x & \text{nếu } 0<x<1;\ 0<y<1-x\\ 0 & \text{các trường hợp khác}\end{cases}$$

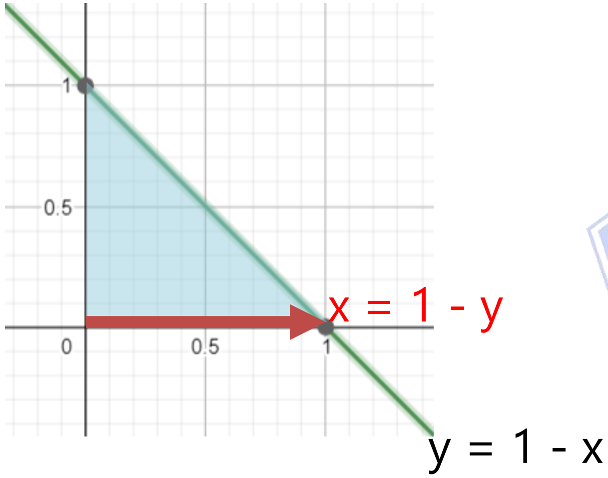
a) Tính trung bình thành phần của Y.

$$E(Y)=\int_{-\infty}^{+\infty} y.f_Y(y)dy=\int_{-\infty}^{+\infty} y.\int_{-\infty}^{+\infty} f(x,y)dxdy$$

$$=\int_0^1 y.\int_0^{1-y} 6xdxdy$$

$$=\int_0^1 y.3x^2\Big|_0^{1-y}dy$$

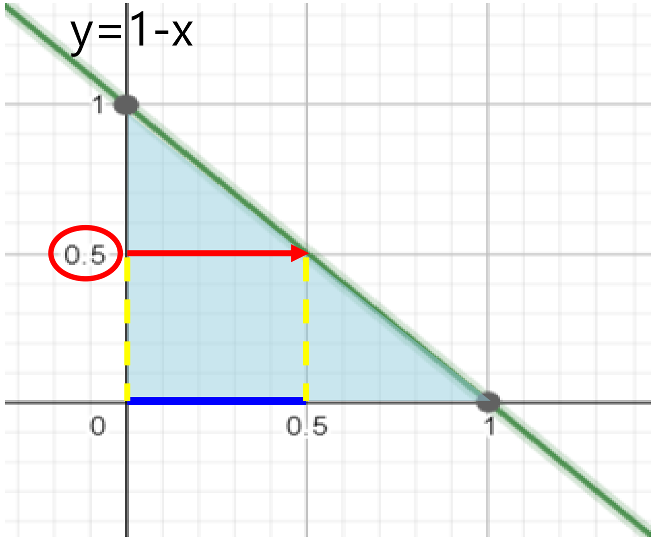
$$=\int_0^1 y.3(1-y)^2dy=\frac{1}{4}$$



VECTOR NN LIÊN TỤC

Ví dụ 5 Cho hàm mật độ xác suất đồng thời của X, Y:

$$f(x,y)=\begin{cases}6x & \text{nếu } 0<x<1;\ 0<y<1-x\\ 0 & \text{các trường hợp khác}\end{cases}$$



b) Tính P(X > 0,3 | Y = 0,5).

$$f_Y(y)=\int_{-\infty}^{+\infty} f(x,y)dx=\begin{cases}\int_0^{1-y}6xdx, & 0<y<1\\ 0, & y\notin(0;1)\end{cases}=\begin{cases}3(1-y)^2, & 0<y<1\\ 0, & y\notin(0;1)\end{cases}$$

$$f_X(x|y)=\begin{cases}\dfrac{f(x,y)}{f_Y(y)}, & f_Y(y)>0\\ 0, & \text{các TH khác}\end{cases}=\begin{cases}\dfrac{6x}{3(1-y)^2}, & 0<x<1;\ 0<y<1-x\\ 0, & \text{các TH khác}\end{cases}$$

$$P(X>0,3|Y=0,5)=\int_{0,3}^{+\infty} f_X(x|y=0,5)dx=\int_{0,3}^{0,5}\frac{6x}{3(1-0,5)^2}dx=0,64$$

VECTOR NN LIÊN TỤC

Ví dụ 5 Cho hàm mật độ xác suất đồng thời của X, Y:

$$f(x,y)=\begin{cases}6x & \text{nếu } 0<x<1;\ 0<y<1-x\\ 0 & \text{các trường hợp khác}\end{cases}$$

Độc lập khi

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

c) X, Y có độc lập không?

$$f_X(x)=\int_{-\infty}^{+\infty} f(x,y)dy=\begin{cases}\int_0^{1-x}6xdy, & 0<x<1\\ 0, & x\notin(0;1)\end{cases}=\begin{cases}6x(1-x), & 0<x<1\\ 0, & x\notin(0;1)\end{cases}$$

$$f_Y(y)=\int_{-\infty}^{+\infty} f(x,y)dx=\begin{cases}\int_0^{1-y}6xdx, & 0<y<1\\ 0, & y\notin(0;1)\end{cases}=\begin{cases}3(1-y)^2, & 0<y<1\\ 0, & y\notin(0;1)\end{cases}$$

Ta có: $f_X(x).f_Y(y)\neq f(x,y)\Leftrightarrow 18x(1-x)(1-y)^2\neq 6x$

=> Vậy X và Y không độc lập

VECTOR NN LIÊN TỤC

Ví dụ 5 Cho hàm mật độ xác suất đồng thời của X, Y:

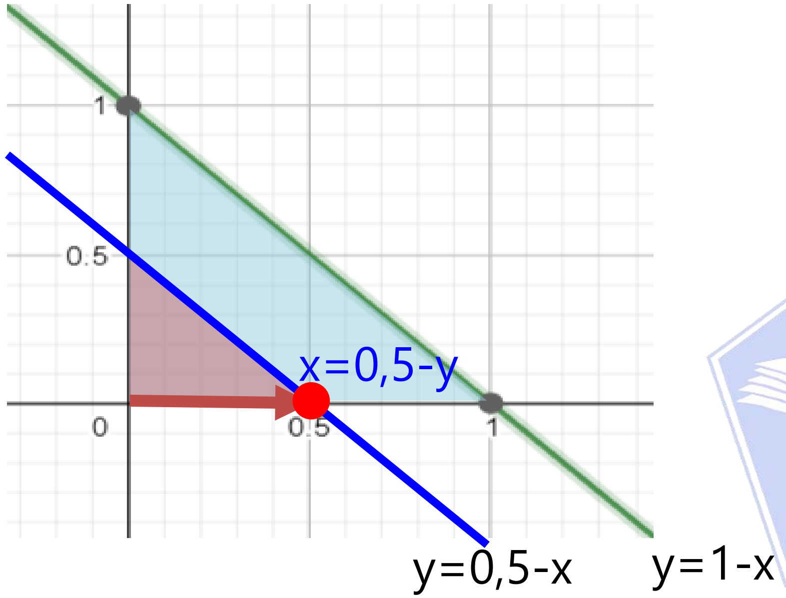
$$f(x,y) = \begin{cases} 6x & \text{nếu } 0<x<1;\ 0<y<1-x \\ 0 & \text{các trường hợp khác} \end{cases}$$

d) Tính P(X + Y ≤ 0,5).

$$P(X+Y \le 0,5) = \int_0^{0,5}\int_0^{0,5-y} 6x\,dx\,dy$$

$$= \int_0^{0,5} 3x^2 \Big|_0^{0,5-y} dy$$

$$= \int_0^{0,5} 3(0,5-y)^2 dy = \frac{1}{8}$$



PHẦN 4: ƯỚC LƯỢNG

1. ƯỚC LƯỢNG KHOẢNG TRUNG BÌNH
2. ƯỚC LƯỢNG TỶ LỆ

CÁCH SỬ DỤNG BẢNG PPXS

Bảng C: xác suất Student

$n\backslash\alpha$	0.95	0.90	0.80	0.70	0.60	0.50	0.40	0.30	0.20	0.10	0.05
1	0.079	0.158	0.325	0.510	0.727	1.000	1.376	1.963	3.078	6.314	12.706
2	0.071	0.142	0.289	0.445	0.617	0.816	1.061	1.386	1.886	2.920	4.303
3	0.068	0.137	0.277	0.424	0.584	0.765	0.978	1.250	1.638	2.353	3.182
4	0.067	0.134	0.271	0.414	0.569	0.741	0.941	1.190	1.533	2.132	2.776
5	0.066	0.132	0.267	0.408	0.559	0.727	0.920	1.156	1.476	2.015	2.571
6	0.065	0.131	0.265	0.404	0.553	0.718	0.906	1.134	1.440	1.943	2.447
7	0.065	0.130	0.263	0.402	0.549	0.711	0.896	1.119	1.415	1.895	2.365
8	0.065	0.130	0.262	0.399	0.546	0.706	0.889	1.108	1.397	1.860	2.306
9	0.064	0.129	0.261	0.398	0.543	0.703	0.883	1.100	1.383	1.833	2.262
10	0.064	0.129	0.260	0.397	0.542	0.700	0.879	1.093	1.372	1.812	2.228
11	0.064	0.129	0.260	0.396	0.540	0.697	0.876	1.088	1.363	1.796	2.201
12	0.064	0.128	0.259	0.395	0.539	0.695	0.873	1.083	1.356	1.782	2.179
13	0.064	0.128	0.259	0.394	0.538	0.694	0.870	1.079	1.350	1.771	2.160

Hàm mật độ XS có dạng:

t_α^n = K

Tra hàng

VD: $t_\alpha^n = t_{0.6}^9$

Tra cột

$=> t_{0.6}^9 = \boxed{0.543}$

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:

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

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

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

ƯỚC LƯỢNG KHOẢNG TRUNG BÌNH

Bài toán ước lượng trung bình có 4TH :

- **ƯỚC LƯỢNG KHOẢNG TRUNG BÌNH**
  - **BIẾT ĐỘ LỆCH CHUẨN TỔNG THỂ (đề cho σ)**
    - **TH1: Kích thước mẫu n ≥ 30**
    - **TH2: Kích thước mẫu n < 30 và X có pp Chuẩn**
  - **CHƯA BIẾT ĐỘ LỆCH CHUẨN TỔNG THỂ (đề không cho σ)**
    - **TH3: Kích thước mẫu n ≥ 30**
    - **TH4: Kích thước mẫu n < 30 và X có pp Chuẩn**

(n là kích thước mẫu, σ là phương sai tổng thể)

ƯỚC LƯỢNG KHOẢNG TRUNG BÌNH

TH1 : $n \geq 30$ và σ đã biết (đề cho σ)

TH2 : $n \geq 30$ và σ chưa biết (đề không cho σ)

TH3 : $n < 30$, σ đã biết và X có pp Chuẩn

- B1 : Tính $\bar{x}$ sau đó tính s (nếu đề không cho σ) bằng cách bấm máy tính.
- B2 : Từ 1-α ⇒ $\varphi(t_\alpha) = (1-\alpha)/2$ → tra bảng Laplace → t_α
- B3 : Tính $\varepsilon = t_\alpha \dfrac{\sigma}{\sqrt{n}}$ (thay σ bằng s nếu không có σ)
- B4 : Khoảng ước lượng là : $[\overline{x} - \varepsilon; \overline{x} + \varepsilon]$

ƯỚC LƯỢNG KHOẢNG TRUNG BÌNH

TH4 : n < 30, σ chưa biết và X có pp Chuẩn

(Lúc này ta dùng bảng **pp Student**)

- B1 : Tính $\bar{x}$ và s bằng cách bấm máy tính.
- B2 : Từ $1-\alpha \Rightarrow \alpha \xrightarrow[\text{pp Student}]{\text{Tra bảng}} t_\alpha$

  (nhớ giảm bậc thành n-1 rồi mới tra bảng !!!)
- B3 : Tính $\varepsilon = t_\alpha \dfrac{s}{\sqrt{n}}$
- B4 : Khoảng ước lượng là : $[\overline{x}-\varepsilon;\overline{x}+\varepsilon]$

ƯỚC LƯỢNG KHOẢNG TRUNG BÌNH

Bài toán: Ước lượng chiều cao trung bình của dân số. Chúng ta cần tìm ra một khoảng (a, b) với độ tin cậy 1-α cho trước.

VD: Hãy ước lượng chiều cao trung bình của người Việt Nam, biết trung bình chiều cao của 250 người VN được khảo sát là 1.7m, độ lệnh chuẩn 0.5m, với độ tin cậy 95% (tức là xác suất 95% ước lượng là chính xác với thông số đã cho).

ƯỚC LƯỢNG KHOẢNG TRUNG BÌNH

Theo đề bài ta có:
* n = 250 người
* $\bar{x} = 1{,}7\text{m}$
* s = 0,5m
* 1 – α = 95%

→ **TH2 : n ≥ 30 và σ chưa biết**

- B1 : Tính $\bar{x}$ và s (đề đã cho)
- B2 : Từ 1 – α = 0,95 ⇒ Ta có $\varphi(t_\alpha)$ = (1-α)/ 2 = 0.475

  Sau đó dò Bảng Laplace $\Rightarrow t_\alpha = 1{,}96$
- B3 : Tính $\varepsilon = t_\alpha \frac{\sigma}{\sqrt{n}} = 1{,}96 \frac{0{,}5}{\sqrt{250}} \approx 0{,}062$
- B4 : Khoảng ước lượng là: $[\bar{x} - \varepsilon; \bar{x} + \varepsilon]$

[1,7 – 0,062; 1,7 + 0,062]

ƯỚC LƯỢNG KHOẢNG TRUNG BÌNH

Bài toán: Ước lượng chiều cao của cây.

VD: Chiều cao của loại cây A là biến ngẫu nhiên có phân phối chuẩn. Người ta đo ngẫu nhiên 20 cây A thì thấy chiều cao trung bình là 23,12m và độ lệch chuẩn của mẫu chưa hiệu chỉnh là 1,25m Tìm khoảng ước lượng chiều cao trung bình của loại cây A với độ tin cậy 95%.

ƯỚC LƯỢNG KHOẢNG TRUNG BÌNH

Theo đề bài ta có:
+ n = 20 cây
+ $\overline{x}$ = 23,12m
+ $\hat{s}$ = 1,25m
+ 1 – α = 95%

→ **TH4 : n < 30 và σ chưa biết**

- B1: Tính $\overline{x}$ và $\hat{s} \Rightarrow s = 1,25.\sqrt{\dfrac{20}{20-1}} = 1,2825m$
- B2: Từ $1-\alpha = 0,95 \Rightarrow \alpha = 0,05$
  Giảm bậc n-1 =19. Tra bảng Student ⇒ $t_\alpha = 2,093$
- B3: $\varepsilon = t_\alpha \dfrac{s}{\sqrt{n}} = 2,093\dfrac{1,2825}{\sqrt{20}} \approx 0,6$
- B4 : Khoảng ước lượng là: [23,12 – 0,6; 23,12 + 0,6]

ĐOÀN KHOA CÔNG NGHỆ PHẦN MỀM
BAN HỌC TẬP
Sharing is learning

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 $1-\alpha \Rightarrow (1-\alpha)/2 = \Phi(t_\alpha) \xrightarrow[\textbf{Laplace}]{\text{Tra bảng}} t_\alpha$

- B3 : Tính $\varepsilon = t_\alpha \sqrt{\dfrac{f(1-f)}{n}}$

- B4 : Khoảng ước lượng là: $[f-\varepsilon; f+\varepsilon]$

ƯỚC LƯỢNG KHOẢNG CHO TỈ LỆ TỔNG THỂ

Bài toán: Khảo sát 500 websites mới đăng ký trên Internet người ta phát hiện có 24 Website vô danh. Xây dựng khoảng ước lượng cho tỷ lệ website vô danh trong số những website mới với độ tin cậy là 95%. (Đề thi 2018 – 2019)

ƯỚC LƯỢNG KHOẢNG CHO TỈ LỆ TỔNG THỂ

Theo đề bài ta có:

+ n = 500 Websites
+ m = 24 Websites
+ $1 - \alpha = 95\%$

- B1 : Tìm $f = \frac{m}{n} = \frac{24}{500} = 0{,}048$
- B2 : Từ $1 - \alpha = 0.95 \Rightarrow \Phi(t_\alpha) = (1-\alpha)/2 = 0{,}475$

  Sau đó dò Bảng Laplace $\Rightarrow t_\alpha = 1{,}96$
- B3 : Tính $\varepsilon = t_\alpha\sqrt{\frac{f(1-f)}{n}} = 1{,}96\sqrt{\frac{0{,}048(1-0{,}048)}{500}} \approx 0{,}0187$
- B4 : Khoảng ước lượng là: [0,048 – 0,0187; 0,048 + 0,0187]

PHẦN 5: KIỂM ĐỊNH GIẢ THIẾT THỐNG KÊ

1. KIỂM ĐỊNH GIẢ THUYẾT VỀ TRUNG BÌNH
2. KIỂM ĐỊNH GIẢ THUYẾT VỀ TỈ LỆ

ĐOÀN KHOA CÔNG NGHỆ PHẦN MỀM
BAN HỌC TẬP
Sharing is learning

KIỂM ĐỊNH GIẢ THIẾT THỐNG KÊ

Quy tắc chung :

Ta nêu lên 2 mệnh đề trái ngược nhau :

- Một mệnh đề là giả thiết H_0
- Một mệnh đề là đối thiết H_1

Dựa vào mẫu mà ta quan sát được (hay khảo sát được), ta chấp nhận hoặc bác bỏ giả thiết H (khi giá trị thống kê rơi vào miền bác bỏ).

KIỂM ĐỊNH VỀ TRUNG BÌNH

- **CÓ 3 TRƯỜNG HỢP:**
  - Đã biết độ lệch chuẩn tổng thể (TH1)
  - Chưa biết độ lệch chuẩn tổng thể, **có $n \geq 30$ (TH2)**
  - Chưa biết độ lệch chuẩn tổng thể, **có $n < 30$ (TH3)**

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 $|z| \leq z_{\alpha/2}$
- 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 $z \geq -z_\alpha$

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

KIỂM ĐỊNH VỀ TRUNG BÌNH

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:

+ μ_0 = 250 000đ
+ $\sigma = 20\,000$đ
+ n = 500
+ $\overline{x} = 252\,000$
+ α = 0,01

Ta kiểm định:

Giả thuyết H_0: μ = 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 $|z| \leq z_{\alpha/2}$

- 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 $z = \dfrac{\overline{x} - \mu_0}{\sigma}\sqrt{n} = \dfrac{252000 - 250000}{20000}\sqrt{500} = 2,236$
- B4 : Có $|z| \leq z_{\alpha/2}$ => Chấp nhận H_0

KIỂM ĐỊNH VỀ TRUNG BÌNH

Bài toán: Sở cứu hỏa Scoottsdale đặt mục tiêu là phản hồi những cuộc gọi cứu hỏa trong thời gian trung bình 4 phút. Thới gian phản hồi có phân phối chuẩn với độ lệch chuẩn là 1 phút. Một mẫu khảo sát gồm 18 cuộc gọi cứu hỏa với thời gian phản hồi trung bình là 4 phút 30 giây có chỉ ra rằng sở cứu hỏa đó không đạt được mục tiêu ở mức ý nghĩa 0.01 không ?

(Đề thi 2018-2019)

KIỂM ĐỊNH VỀ TRUNG BÌNH

Theo đề bài ta có:

+ μ_0 = 4 phút
+ $\sigma = 1$ phút
+ n = 18 cuộc
+ $\overline{x}$ = 4,5
+ α = 0,01

Ta kiểm định:

Giả thuyết H_0: $\mu = 4$

Đối thuyết H_1: $\mu > 4$

- 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 $z \leq z_\alpha$

- B1 : B1 : Tính $\bar{x}$ và s (đề đã cho)
- B2 : Từ α = 0,01. Ta có $\Phi(z_\alpha) = 1 - \alpha = 0,99$

  Sau đó dò Bảng Phân Phối Chuẩn $\Rightarrow z_\alpha = 2,33$

- B3 : Tính $z = \dfrac{\overline{x} - \mu_0}{\sigma}\sqrt{n} = \dfrac{4,5-4}{1}\sqrt{18} = 2,1213$
- B4 : Có $z \leq z_\alpha$ => Chấp nhận H_0

KIỂM ĐỊNH VỀ TRUNG BÌNH

TH2: Chưa biết độ lệch chuẩn tổng thể và n $\geq$ 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 $z \geq -z_\alpha$

KIỂM ĐỊNH VỀ TRUNG BÌNH

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

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

+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*

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?

KIỂM ĐỊNH VỀ TRUNG BÌNH

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.

KIỂM ĐỊNH VỀ TRUNG BÌNH

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 !

- 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 $z \geq -z_{\alpha}$

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

KIỂM ĐỊNH VỀ TỈ LỆ

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.

KIỂM 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 $\Rightarrow z_{\alpha/2} = 1{,}96$
- 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 : $z \le z_{\alpha/2}$ => Giả thuyết đúng

PHẦN 6 : BÀI TOÁN TƯƠNG QUAN & HỒI QUY

HỆ SỐ TƯƠNG QUAN MẪU

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

Tính chất :

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

- $-1 \le r \le 1$
- 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 $b = \frac{\overline{xy} - \bar{x}\bar{y}}{\hat{s}_x^2}$

Với $a = \bar{y} - b\,\bar{x}$

Đường hồi quy tuyến tính của X theo Y là :

$$X = a + bY$$

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

Với $a = \bar{x} - b\bar{y}$

PHƯƠNG TRÌNH HỒI QUY TUYẾN TÍNH

HƯỚNG DẪN CÁCH BẤM MÁY

ĐOÀN KHOA CÔNG NGHỆ PHẦN MỀM
BAN HỌC TẬP
Sharing is learning

HƯỚNG DẪN CÁCH BẤM MÁY

<u>VD 1.</u> Kết quả đo lường độ cholesterol (Y) có trong máu của 10 đối tượng nam ở độ tuổi (X) như sau:

X	20	52	30	57	28	43	57	63	40	49
Y	1,9	4,0	2,6	4,5	2,9	3,8	4,1	4,6	3,2	4,0

Tính hệ số tương quan mẫu giữa X và Y.

DÙNG MÁY TÍNH BỎ TÚI TÌM ĐƯỜNG HỒI QUY

1. Số liệu không có tần số

a) Máy tính f_x500MS, f_x570MS

VD 1. Bài toán cho ở dạng cặp (x_i, y_i) như sau:

X	20	52	30	57	28	43	57	63	40	49
Y	1,9	4,0	2,6	4,5	2,9	3,8	4,1	4,6	3,2	4,0

Tìm hệ số r, đường hồi quy Y theo X: $y = a + bx$.

Nhập số liệu:

$$\textbf{MODE} \longrightarrow \textbf{REG} \longrightarrow \textbf{LIN}$$

$$X, \quad Y \longrightarrow \mathbf{M}^+$$

$$20,\ 1.9 \longrightarrow \mathbf{M}^+$$

$$52,\ 4.0 \longrightarrow \mathbf{M}^+$$

$$\ldots \quad \ldots\ \ldots \quad \ldots$$

$$49,\ 4.0 \longrightarrow \mathbf{M}^+$$

DÙNG MÁY TÍNH BỎ TÚI TÌM ĐƯỜNG HỒI QUY

Xuất kết quả:

SHIFT $\rightarrow$ 2 $\rightarrow$ (dịch chuyển mũi tên phải 2 lần)

$\rightarrow$ **1** (***A*** chính là ***a*** trong phương trình)

$\rightarrow$ **2** (***B*** chính là ***b*** trong phương trình)

$\rightarrow$ **3** (***r*** chính là r).

Đáp số: $r = 0,9729;\ y = 0,9311 + 0,0599x.$

DÙNG MÁY TÍNH BỎ TÚI TÌM ĐƯỜNG HỒI QUY

b) Máy tính f_x500ES, f_x570ES

Xét lại VD 1 ở trên.

Nhập số liệu:

SHIFT $\longrightarrow$ **MODE** $\longrightarrow$ dịch chuyển mũi tên tìm chọn mục **Stat** $\longrightarrow$ **2** (chế độ không tần số)

MODE $\longrightarrow$ **3** (stat) $\longrightarrow$ **2** ($A+Bx$) $\longrightarrow$ (nhập các giá trị của X, Y vào 2 cột)

X	Y
20	1.9
52	4.0
…	…
49	4.0

Xuất kết quả:

SHIFT $\rightarrow$**1** $\rightarrow$**7** $\rightarrow$ **1**(A chính là ***a*** trong phương trình)

SHIFT $\rightarrow$**1** $\rightarrow$**7** $\rightarrow$ **2**(B chính là ***b*** trong phương trình)

SHIFT $\rightarrow$**1** $\rightarrow$**7** $\rightarrow$ **3**(r chính là ***r*** trong phương trình).

ĐOÀN KHOA CÔNG NGHỆ PHẦN MỀM
BAN HỌC TẬP
Sharing is learning

Đề tham khảo năm (2019-2020)

Sau đây là dữ liệu của hai biến ngẫu nhiên X và Y:

X	100	230	320	500	550	1000	1300	1500	2000	2100	3200	3800
Y	11	34	25	35	40	45	55	65	80	75	125	115

a) Tính hệ số tương quan và nhận xét về tính tuyến tính của X và Y (Mạnh hay yếu? Nghịch biến hay đồng biến?)

b) Viết phương trình hồi qui tuyến tính của Y theo X. Dự đoán giá trị của Y khi X= 5000.

DÙNG MÁY TÍNH BỎ TÚI TÌM ĐƯỜNG HỒI QUY

Đáp án

a) r = 0,9756 => tính tuyến tính mạnh đồng biến.

b) Y = 19,3414 + 0,0285.X
Khi X = 5000 => Y = 161,84

2. Số liệu có tần số

a) Máy tính f_x500MS, f_x570MS

VD 2. Tìm hệ số r, đường hồi quy thực nghiệm Y theo X: $y = a + bx$ với bài toán cho ở dạng bảng như sau:

$Y \backslash X$	21	23	25
3	2		
4	5	3	
5		11	8

DÙNG MÁY TÍNH BỎ TÚI TÌM ĐƯỜNG HỒI QUY

Y \ X	21	23	25
3	2		
4	5	3	
5		11	8

Nhập số liệu:

$$\mathbf{MODE} \to \mathbf{REG} \to \mathbf{LIN}$$

$$X,\quad Y;\ n \to \mathbf{M}^{+}$$

$$21,\quad 3;\ 2 \to \mathbf{M}^{+}$$

$$21,\quad 4;\ 5 \to \mathbf{M}^{+}$$

$$\ldots \qquad\qquad \ldots$$

$$25,\ 5;\ 8 \to \mathbf{M}^{+}$$

***Xuất kết quả*: làm như 1a).**

Đáp số: $r = 0,7326;\ y = -2,6694 + 0,3145x.$

ĐOÀN KHOA CÔNG NGHỆ PHẦN MỀM
BAN HỌC TẬP
Sharing is learning

DÙNG MÁY TÍNH BỎ TÚI TÌM ĐƯỜNG HỒI QUY

b) Máy tính f_x500ES, f_x570ES

Xét lại VD 2 ở trên

***Nhập số liệu*:**

SHIFT $\rightarrow$ **MODE** $\rightarrow$ dịch chuyển mũi tên tìm chọn
Mục **Stat** $\rightarrow$ **1** (chế độ có tần số)

MODE $\rightarrow$ **3** (stat) $\rightarrow$ **2** ($A+Bx$) $\rightarrow$ (nhập các giá trị của X, Y, tần số vào 3 cột)

X	Y	FREQ
21	3	2
21	4	5
...	...	...
25	5	8

***Xuất kết quả*: làm như 1b).**

PHẦN 7 : TÀI LIỆU THAM KHẢO

BẢNG PHÂN PHỐI XÁC SUẤT

DOWNLOAD TẠI : http://bit.ly/2LBILm9

PHẦN ĐIỂM DANH

https://bom.so/oZWa2P

BAN HỌC TẬP ĐOÀN KHOA CNPM

Chuỗi Training cuối học kì II năm học 2021 - 2022

HẾT

Cảm ơn các bạn đã theo dõi
Chúc các bạn có kết quả thi thật tốt !

Ban học tập

Khoa Công Nghệ Phần Mềm
Trường ĐH Công Nghệ Thông Tin

Email / Group

Bht.cnpm.uit@gmail.com
fb.com/groups/bht.cnpm.uit/