Sự độc lập của 2 BNN rời rạc

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

Hai BNN X, Y độc lập nghĩa là các biến cố $\{X = x\}$ và $\{Y = y\}$ độc lập nhau với mọi giá trị (x, y).

Chứng minh X, Y phụ thuộc nhau: chỉ ra 1 cặp giá trị (x, y) sao cho

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

Vector ngẫu nhiên liên tục

Cho BNN liên tục X, Y, **hàm phân phối xác suất đồng thời** của X, Y là hàm

$$F_{(X,Y)}(x, y) = P(X \leq x \cap Y \leq y)$$

Hàm mật độ đồng thời:

$$f(x, y) = \frac{\partial^2}{\partial x \partial y} F_{(X,Y)}(x, y)$$

Tính chất:

- $P((X, Y) \in D) = \iint_D f(x, y) dxdy$
- $\iint_{\mathbb{R}^2} f(x, y) dxdy = 1$
- $F(x, y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(u, v)\, du\, dv$

1. Hàm mật độ thành phần (mật độ biên)
    - **Hàm mật độ của X**

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

    - **Hàm mật độ của Y**

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

2. Trung bình thành phần
    - Trung bình của X: $E(X) = \int_{-\infty}^{\infty} x\, f_X(x)dx$
    - Trung bình của Y: $E(Y) = \int_{-\infty}^{\infty} y\, f_Y(y)dy$

Sự độc lập của 2 BNN liên tục

$$\begin{aligned} X, Y \text{ độc lập} &\iff F_{(X,Y)}(x,y) = F_X(x) \cdot F_Y(y) \quad \forall\, x, y \\ &\iff f(x,y) = f_X(x) \cdot f_Y(y) \quad \forall\, x, y \end{aligned}$$

Để chứng minh X, Y phụ thuộc nhau, chỉ cần chỉ ra 1 cặp giá trị (x, y) sao cho

$$F_{(X,Y)}(x,y) \neq F_X(x) \cdot F_Y(y)$$

hoặc

$$f(x,y) \neq f_X(x) \cdot f_Y(y).$$

- Hàm mật độ của X trong điều kiện $Y = y, \quad f_Y(y) > 0$

$$f_X(x|y) = \frac{f(x, y)}{f_Y(y)}$$

(Với mỗi giá trị y, đây là hàm 1 biến theo x.)

- Hàm mật độ của Y trong điều kiện $X = x, \quad f_X(x) > 0$

$$f_Y(y|x) = \frac{f(x, y)}{f_X(x)}$$

(Với mỗi giá trị x, đây là hàm 1 biến theo y)

Kiểm định giả thiết về kì vọng

Giả sử X có phân phối chuẩn $N(\mu, \sigma^2)$.
Gọi μ là giá trị trung bình của tổng thể. Ta sẽ kiểm định giả thiết
$H_0 : \ \mu = \mu_0$
Giả sử H_0 đúng, nghĩa là $\mu = \mu_0$, thì

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$$

Kí hiệu $z_{\alpha/2}$ là giá trị thỏa

$$P(|Z| > z_{\alpha/2}) = \alpha \iff P(Z \leq z_{\alpha/2}) = 1 - \frac{\alpha}{2}$$

Kí hiệu z_α là giá trị thỏa

$$P(Z > z_\alpha) = \alpha \iff P(Z \leq z_\alpha) = 1 - \alpha$$

TH: σ đã biết (X có phân phối chuẩn)

Đặt

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

- 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 $z \geq -z_\alpha$

TH: σ chưa biết

a) $n \geq 30$

Đặt

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

với s là độ lệch chuẩn có hiệu chỉnh của BNN X tính từ mẫu dữ liệu.

- 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 $z \geq -z_\alpha$

TH: σ chưa biết

b) $n < 30$, X có phân phối chuẩn

$$T = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \sim \text{PP Student bậc tự do } n-1$$

với s là độ lệch chuẩn có hiệu chỉnh của BNN X tính từ mẫu dữ liệu.
Kí hiệu t_α và $t_{\alpha/2}$ là các số thực thỏa

$$P(T > t_\alpha) = \alpha, \quad P(T > t_{\alpha/2}) = \frac{\alpha}{2}$$

với T tuân theo phân phối Student bậc tự do $n-1$ (tra bảng A5)

TH: σ chưa biết, X có phân phối chuẩn

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

b) $n < 30$

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

Kiểm định giả thiết về tỷ lệ

Giả sử ta quan tâm đến những phần tử có đặc trưng A nào đó trong tổng thể.
Gọi p là tỷ lệ phần tử có đặc trưng A trong tổng thể.
Ta sẽ kiểm định giả thiết $H_0: \ p = p_0$.
Xét thống kê (với n lớn)

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

Kiểm định giả thiết về tỷ lệ

Dựa trên mẫu dữ liệu, tính

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

với f là tỷ lệ phần tử có tính chất A trong mẫu cụ thể.

- 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 $z \geq -z_\alpha$

Các khái niệm và kí hiệu chung cho bài toán ước lượng

Xét ĐLNN X trong tổng thể.

Xét một mẫu ngẫu nhiên gồm n phần tử.

μ: giá trị trung bình của ĐLNN X trong tổng thể.

$$\boxed{\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}}$$

trung bình mẫu cụ thể

σ: độ lệch chuẩn của X trong tổng thể.

s: độ lệch chuẩn có hiệu chỉnh của mẫu cụ thể

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

Ước lượng trung bình tổng thể μ

- **Mục tiêu:** ước lượng μ với độ tin cậy 1- α
  - Độ tin cậy 1- α là khả năng μ thuộc vào khoảng tin cậy mà ta đưa ra.
  - Dùng giá trị trung bình của mẫu $(\overline{x})$ để ước lượng trung bình của tổng thể μ
- Ta cần tìm khoảng tin cậy $[\overline{x}-\varepsilon;\overline{x}+\varepsilon]$ sao cho:

$$P(\overline{x}-\varepsilon \leq \mu \leq \overline{x}+\varepsilon)=1-\alpha$$

- Ví dụ: Khoảng tin cậy là [156; 173] với độ tin cậy là 95% nghĩa là có 95% khả năng μ thuộc khoảng [156; 173] và có 5% khả năng μ không thuộc khoảng [156; 173].

Ước lượng μ với độ tin cậy 1-α

1) TH: X có phân phối chuẩn, σ đã biết

$$Z = \frac{\overline{X} - \mu}{(\sigma / \sqrt{n})} \sim N(0,1)$$

Tìm $z_{\alpha/2}$ sao cho $P(-z_{\alpha/2} \le Z \le z_{\alpha/2}) = 1 - \alpha$,
nghĩa là tìm $z_{\alpha/2}$ (tra bảng A4) thỏa

$$P(Z \le z_{\alpha/2}) = 1 - (\alpha/2)$$

Tính độ sai số của ước lượng $\boxed{\varepsilon = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}}$

Khoảng tin cậy cho μ với độ tin cậy 1-α:

$$[\overline{x} - \varepsilon, \overline{x} + \varepsilon]$$

Ước lượng μ với độ tin cậy 1-α

2) TH: σ chưa biết

a) n ≥ 30 : thay σ bởi s (X có phân phối bất kì)

$$Z = \frac{\bar{X} - \mu}{(s/\sqrt{n})} \sim N(0,1)$$

Tìm $z_{\alpha/2}$ (tra bảng A4) thỏa

$P(Z \le z_{\alpha/2}) = 1 - (\alpha/2)$

Tính độ sai số của ước lượng $\boxed{\varepsilon = z_{\alpha/2} \frac{s}{\sqrt{n}}}$

Khoảng tin cậy cho μ với độ tin cậy 1-α:

$[\overline{x} - \varepsilon, \overline{x} + \varepsilon]$

Ước lượng μ với độ tin cậy 1-α

2) TH: σ chưa biết

b) n < 30, X có phân phối chuẩn :

Thay σ bởi độ lệch chuẩn mẫu có điều chỉnh s

$$T = \frac{\bar{X} - \mu}{(s/\sqrt{n})} \sim$$ Phân phối Student bậc tự do n-1

Tìm $t_{\alpha/2}$ (tra bảng A5, dòng n-1) thỏa

$$P(T > t_{\alpha/2}) = \frac{\alpha}{2}$$

Tính độ sai số của ước lượng $\boxed{\varepsilon = t_{\alpha/2}\frac{s}{\sqrt{n}}}$

Khoảng tin cậy cho μ với độ tin cậy 1-α:

$$[\bar{x} - \varepsilon, \bar{x} + \varepsilon]$$

Ước lượng tỉ lệ tổng thể

Kí hiệu:

p: tỷ lệ phần tử có tính chất $\mathscr{P}$ trong tổng thể,

f: tỷ lệ phần tử có tính chất $\mathscr{P}$ trong mẫu.

Lưu ý: bài toán ước lượng tỷ lệ chỉ xét trong trường hợp mẫu lớn, do đó ta dùng phân phối chuẩn.

Ước lượng tỷ lệ tổng thể p với độ tin cậy 1- α

Khoảng tin cậy: $[f-\varepsilon\ ;\ f+\varepsilon]$

Độ chính xác (sai số)

$$\varepsilon = z_{\alpha/2}\sqrt{\frac{f(1-f)}{n}}$$

Trong đó $z_{\alpha/2}$ thỏa

$P(Z \leq z_{\alpha/2}) = 1 - (\alpha/2)$, $Z \sim N(0,1)$

(tra bảng A4 để xác định $z_{\alpha/2}$)

Bài toán xác định cỡ mẫu

Bài toán: Xác định cỡ mẫu cần khảo sát để đạt được ước lượng cho tỷ lệ tổng thể với độ tin cậy và độ sai số cho trước.

$$\varepsilon = z_{\alpha/2}\sqrt{\frac{f(1-f)}{n}} \le z_{\alpha/2}\frac{1}{2\sqrt{n}} \qquad \forall \ f \in [0,1]$$

$$\Rightarrow \qquad n \ge \left(\frac{z_{\alpha/2}}{2\varepsilon}\right)^2$$

Dùng khoảng tin cậy của p để ước lượng N

Để điều tra số cá trong hồ, cơ quan quản lý đánh bắt 300 con, làm dấu rồi thả xuống hồ. Lần hai bắt ngẫu nhiên 400 con thấy có 60 con được đánh dấu. Hãy xác định số cá trong hồ với độ tin cậy 95%.

- $p = \frac{M}{N} = \frac{300}{N}$ tỷ lệ cá ***được đánh dấu*** trong hồ.
- $f = \frac{m}{n} = \frac{60}{400} = 15\%$: tỷ lệ cá được đánh dấu trong mẫu ngẫu nhiên gồm $n = 400$con cá.
- Với độ tin cậy 95%, sử dụng phân phối chuẩn để ước lượng p, sau đó ước lượng N.

Đánh giá sai số của bài toán ước lượng tỉ lệ

$$\boxed{\varepsilon = z_{\alpha/2}\sqrt{\frac{f(1-f)}{n}} \le \frac{z_{\alpha/2}}{2\sqrt{n}} = \varepsilon_{\max}}$$

Vì lí do tài chính, một tờ báo quyết định chỉ khảo sát 2000 cử tri về dự định của họ trong cuộc bầu cử sắp tới. Với độ tin cậy 95%, sai số tối đa khi ước lượng tỷ lệ cử tri bầu cho ứng viên X là bao nhiêu ?

Hệ số tương quan mẫu

- Công thức

$$r = \frac{\overline{xy} - \overline{x}.\overline{y}}{\sqrt{\left(\overline{x^2} - (\overline{x})^2\right)\left(\overline{y^2} - (\overline{y})^2\right)}}$$

- $|r| > 0{,}9$: có sự tương quan tuyến tính mạnh giữa X, Y
  - $r > 0$: tương quan tuyến tính **thuận** (X, Y đồng biến)
  - $r < 0$: tương quan tuyến tính **nghịch** (X, Y nghịch biến)
- $|r| < 0{,}9$: tương quan tuyến tính giữa X và Y yếu.
- |r| càng gần 1 thì sự tương quan tuyến tính càng mạnh

Hồi quy

- Mục đích: **dự đoán đại lượng này** (*đại lượngphụ thuộc*) **từ các đại lượng khác** (*các đại lượngđộc lập*).
  - Vd: biết chiều dài của một con cá mập → cân nặng, tuổi, chiều rộng hàm cá mập?
- Tiến trình ước lượng này được gọi là tiến trình hồi quy.
- Nếu Y được ước lượng từ Xbằng một biểu thức nào đó thì ta gọi **biểu thức này** là phương trình hồi quy *của* Y *theo* X.
- **Đường cong tương ứng**được gọi là đường cong hồi quy*của Y theo X.*
  - **Đường thẳng tương ứng** được gọi là đường thẳng hồi quy*của Y theo X.*

Sử dụng máy tính

Ví dụ Bài toán cho dạng cặp (x_i, y_i) như sau:

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

Tìm hệ số tương quan r_{xy}, đường hồi qui mẫu $y = a + bx$.

a. Máy fx-570VN Plus

– Bước 1: SHIFT; MODE; ↓; chọn 4 (Stat); chọn 2 (Off)

– Bước 2: MODE; chọn 3 (Stat); chọn 2 (A+Bx)

* Nhập giá trị của X 20= 52= ⋯
* Nhập giá trị của Y 1.9= 4= ⋯

– Bước 3: Xuất kết quả On; SHIFT; chọn 1 (Stat); chọn 5 (Reg); 1(A =a); 2(B=b); 3(r=r_{xy}).

Kết quả $r_{xy} = 0,9729$; $y = 0,9311 + 0,0599x$.

ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN
BỘ MÔN TOÁN – LÝ

ĐỀ THI CUỐI KỲ
MÔN XÁC SUẤT THỐNG KÊ
Học kỳ 2, năm học 2018-2019.
Ngày thi: / / 2019.
Thời gian làm bài: 90 phút.
Được sử dụng tài liệu giấy.

(Lưu ý: Đề thi gồm có 2 trang)

Câu 1.(1.5 điểm) Tuổi thọ của một loại chip máy tính là đại lượng ngẫu nhiên X (đơn vị tính là giờ) có phân phối chuẩn, trong đó tuổi thọ trung bình là 120000 giờ và độ lệch chuẩn là $\sigma = 30000$ giờ.

a) Tính tỷ lệ chip có tuổi thọ trên hơn 140000 giờ.

b) Hãy tính xác suất trong 100 chip loại này có ít nhất 20 chip có tuổi thọ trên 140000 giờ.

Câu 2.(1.5 điểm) Cho biến ngẫu nhiên X và Y có phân phối xác suất đồng thời như sau :

$P(x,y)$		x			
		1	2	3	4
y	1	0	0.06	0.06	0.10
	2	0.10	0.10	0.04	0.04
	3	0.40	0.10	0	0

a) X và Y có độc lập nhau hay không ? Vì sao ?

b) Tính xác suất $P(X+Y \le 3), P(X > 1|Y = 2)$.

Câu 3.(1.5 điểm) Cho 2 biến ngẫu nhiên X, Y có hàm mật độ đồng thời

$$f(x,y) = \begin{cases} Cxy & \text{nếu } x \in [0,2], y \in [1,3], \\ 0 & \text{nếu } (x,y) \text{ khác.} \end{cases}$$

a) Tìm C.

b) Tính xác suất $P(X \le 1 \cap Y > 2)$.

c) Tính xác suất $P(X \le 1|Y > 2)$.

Câu 4.(2 đ)

a) 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%.

b) Nếu muốn ước lượng tỷ lệ người trưởng thành Việt Nam biết nhóm máu của mình với độ tin cậy 95% và sai số tối đa 0.02 thì cần khảo sát ít nhất bao nhiêu người ?

Câu 5.(2.5 đ)

a) Tổ chức bảo vệ môi trường của Mỹ (EPA) quan tâm chất lượng nước uống phục vụ trên các chuyến bay. Họ kiểm tra ngẫu nhiên 112 chuyến bay và phát hiện có 14 chuyến bay phục vụ nước uống bị nhiễm khuẩn trên mức cho phép. Để biết có thể đưa ra kết luận rằng có trên 10% chuyến bay có nước bị nhiễm khuẩn hay không, họ có thể dùng bài toán kiểm định giả thiết. Hãy phát biểu giả thiết H_0, và đối thiết H_1 của bài toán đó ?

b) Sở cứu hỏa Scottsdale đặt mục tiêu là phản hồi những cuộc gọi cứu hỏa trong thời gian trung bình 4 phút. Thời 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 $\alpha = 0.01$ không ?

Câu 6.(1 điểm) Sau đây là dữ liệu về tốc độ vi xử lý (Microprocessor Speed) và công suất tiêu tán năng lượng (Power dissipation) của các loại chip

Chip	Microprocessor Speed (MHz)	Power Dissipation (watts)
1989 Intel 80486	20	3
1993 pentium	100	10
1997 Pentium II	233	35
1998 Intel Celeron	300	20
1999 Pentium III	600	42
1999 AMD Athlon	600	50
2000 Pentium 4	1300	51
2004 Celeron D	2100	73
2004 Pentium 4	3800	115
2005 Pentium D	3200	130
2007 AMD Phenom	2300	95
2008 Intel Core 2	3200	136
2009 Intel Core i7	2900	95
2009 AMD Phenom II	3200	125

Nguồn: http://en.wikipedia.org và New Scientist, 208, no. 2780 (October 2, 2010), p. 41.

a) Tính hệ số tương quan và đánh giá về tính tuyến tính (mạnh hay yếu) của mối liên hệ phụ thuộc giữa X và Y. Viết phương trình hồi qui tuyến tính của công suất tiêu tán năng lượng theo tốc độ vi xử lí.

b) Dự đoán về công suất tiêu tán năng lượng nếu tốc độ vi xử lí là 3500 MHz ?

Hết

Cán bộ coi thi không giải thích gì thêm.

Trưởng Bộ môn Toán-Lý

CAO THANH TÌNH

ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN
BỘ MÔN TOÁN – LÝ

ĐỀ THI CUỐI KỲ
MÔN XÁC SUẤT THỐNG KÊ
Học kỳ 2, năm học 2019-2020.
Ngày thi: / / 2020
Thời gian làm bài: 90 phút
Được sử dụng tài liệu giấy.

(Lưu ý: Đề thi gồm có 2 trang)

Câu 1.(1.5 điểm) Giả sử X và Y là hai biến ngẫu nhiên có phân phối đồng thời như sau:

$$P(X=0,Y=0)=0.1,\ P(X=0,Y=1)=0.3,\ P(X=0,Y=2)=0.2,$$

$$P(X=1,Y=0)=0.1,\ P(X=1,Y=1)=0.2,\ P(X=1,Y=2)=0.1.$$

a) Hỏi X, Y có độc lập nhau không? Vì sao?
b) Giả sử $Y > 0$, tính xác suất $X > 0$.
c) Tính $P(Y > X)$.

Câu 2.(2 điểm) Cho 2 biến ngẫu nhiên X, Y có hàm mật độ đồng thời

$$f(x,y)=\begin{cases}6xy & \text{nếu } 0\le x\le 1,\ 0\le y\le 2-2x,\\ 0 & \text{nếu } (x,y) \text{ khác.}\end{cases}$$

a) Tìm hàm mật độ thành phần của X.
b) Tính $P(Y < 0.5|X < 0.5)$.

Câu 3.(2.5 điểm) Sau nhiều phàn nàn về khối lượng ngũ cốc trong hộp ít hơn 16 ounces, công ty sản xuất ngũ cốc Captain Crisp quyết định tiến hành kiểm tra. Họ chọn ngẫu nhiên một mẫu gồm 400 hộp và thấy có 94 hộp có lượng ngũ cốc ít hơn 16 ounce.

a) Xây dựng khoảng tin cậy với độ tin cậy 95% cho tỷ lệ hộp ngũ cốc nhẹ hơn 16 ounce (xét trên toàn bộ hộp ngũ cốc do công ty đó sản xuất).

b) Với mức ý nghĩa 0.1, có thể kết luận rằng tỷ lệ hộp ngũ cốc nhẹ hơn 16 ounce của công ty đó có trên 20% không?

Câu 4.(1.5 điểm) Các nhà nghiên cứu lo lắng có quá nhiều clorine trong nước uống. Họ thu thập 25 mẫu nước uống để đo lượng clorine. Lượng clorine trung bình của các mẫu là 4.2mg/l với độ lệch chuẩn có hiệu chỉnh của mẫu là 0.6 mg/l. Cục quản lý Thực phẩm và Dược phẩm Hoa Kỳ (FDA) khuyến nghị rằng lượng clorine trung bình không nên vượt quá 4 mg/l. Các nhà nghiên cứu muốn biết rằng lượng clorine trung bình trong nước uống có vượt quá giới hạn do FDA đề nghị không. Giả sử lượng clorine trong nước uống có phân phối chuẩn. Hãy thực hiện kiểm định thích hợp với mức ý nghĩa 0.05.

Câu 5.(1.5 điểm) Tỷ lệ khách hàng phản hồi tin nhắn của một công ty là 20%. Trong tuần vừa qua, công ty đó gởi tin nhắn cho 400 khách hàng. Hỏi xác suất có hơn 50 khách hàng trong số khách hàng đó phản hồi tin nhắn?

Câu 6.(1 điểm) 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.

Hết

Cán bộ coi thi không giải thích gì thêm.

Trưởng Bộ môn Toán-Lý

CAO THANH TÌNH

ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN
BỘ MÔN TOÁN – LÝ

ĐỀ THI CUỐI KỲ MÔN XÁC SUẤT THỐNG KÊ
Học kỳ 2, năm học 2021-2022.
Ngày thi: .../.../2022
Thời gian làm bài: 90 phút
Dược sử dụng tài liệu giấy.

Câu 1: (2 điểm) Giả sử X và Y là hai biến ngẫu nhiên có phân phối xác suất đồng thời như sau:

$P(X,Y)$		X = 1	2	3
Y	1	0.01	0.02	0.08
	2	0.01	0.02	0.08
	3	0.07	0.08	0.63

a. Tính $P(X \leq 2, Y \leq 3)$.

b. X, Y có độc lập không? Vì sao?

c. Tính $P(X > 1|Y = 3)$.

d. Tính giá trị trung bình của X.

Câu 2: (2 điểm) Cho (X, Y) là véctơ ngẫu nhiên liên tục có hàm mật độ đồng thời là:

$$f(x,y) = \begin{cases} \dfrac{k}{x^2y} & \text{nếu } 1 \leq x \leq 2 \text{ và } 1 \leq y \leq e, \\ 0 & \text{nếu } (x,y) \text{ khác.} \end{cases}$$

a. Tìm k.

b. Tìm hàm mật độ thành phần của X. Tính $P(X < \frac{3}{2})$.

c. Tìm $P(X < \frac{3}{2}, Y > \sqrt{e})$. Tính $P(Y > \sqrt{e}|X < \frac{3}{2})$.

Câu 3: (1 điểm) Dữ liệu dưới đây thể hiện các khoản đầu tư X (tính bằng 1000 đô la) vào việc phát triển phần mềm mới của một công ty máy tính theo năm Y.

Year, Y	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Investment, X	17	23	31	29	33	39	39	40	41	44	47

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) Tìm phương trình hồi qui tuyến tính của Y theo X. Dự đoán giá trị đầu tư của công ty vào năm 2023.

Câu 4: (1.5 điểm) Một công ty đưa ra tuyên bố rằng một loại sản phẩm mới mà họ đã phân phối trong toàn bộ hệ thống đại lý bán lẻ của họ sau một năm đã có hơn 90% sản phẩm được khách hàng quan tâm và chọn mua. Kiểm tra ngẫu nhiên 300 đại lý thấy có đến 40 đại lý không bán được một sản phẩm nào loại này, trong khi số đại lý còn lại gần như đã bán hết sản phẩm. Với mức ý nghĩa 5%, hãy kiểm định xem tuyên bố trên của công ty có quá sự thật không?

Câu 5: (1.5 điểm) Một giám đốc một công ty nói rằng mức lương trung bình của các kỹ sư IT của công ty ông là 24,6 triệu đồng/tháng. Giả sử mức lương hàng tháng của các kỹ sư IT có phân phối chuẩn với độ lệch chuẩn 2,52 triệu đồng. Chọn ngẫu nhiên 25 kỹ sư IT trong công ty, người ta thấy rằng lương trung bình của nhóm kỹ sư này là 25,33 triệu đồng/tháng. Hỏi có đủ căn cứ để chấp nhận tuyên bố của giám đốc này với mức ý nghĩa 1% không?

Câu 6: (2 điểm) Để ước lượng tỷ lệ phế phẩm của một dây chuyền sản xuất, người ta chọn ngẫu nhiên 300 sản phẩm và kiểm tra thấy có 20 phế phẩm.

a) Với độ tin cậy 99%, hãy tìm khoảng ước lượng cho tỷ lệ phế phẩm của dây chuyền trên.

b) Để sai số của việc ước lượng với độ tin cậy 95% không vượt quá 0.01 thì ta phải kiểm tra tối thiểu bao nhiêu sản phẩm?

Hết

Cán bộ coi thi không giải thích gì thêm.

ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN
BỘ MÔN TOÁN – LÝ

ĐỀ THI CUỐI KỲ MÔN XSTK

Học kỳ 1, năm học 2022-2023.
Ngày thi: .../.../2023
Thời gian làm bài: 90 phút.
Được sử dụng tài liệu giấy.

Câu 1.(2 điểm) Một thành phố A có 30% người theo đảng Cộng hòa, 50% theo đảng Dân chủ, và những người còn lại không theo đảng nào. Chọn ngẫu nhiên một người và xét 2 biến ngẫu nhiên X, Y được định nghĩa như sau:

$$X = \begin{cases} 1 & \text{người đó thuộc đảng Cộng hòa} \\ 0 & \text{người đó không thuộc đảng Cộng hòa} \end{cases}$$

$$Y = \begin{cases} 1 & \text{người đó thuộc đảng Dân chủ} \\ 0 & \text{người đó không thuộc đảng Dân chủ} \end{cases}$$

Tìm phân phối xác suất đồng thời của X và Y. Tính xác suất $P(X = 1|Y = 0)$.

Câu 2.(3 điểm) Cho (X, Y) là biến ngẫu nhiên có hàm mật độ xác suất đồng thời như sau:

$$f(x, y) = \begin{cases} 6x^2y & \forall\ 0 \le x, y \le 1, \\ 0 & \forall\ (x, y) \text{ khác.} \end{cases}$$

a. Tính $P(X + Y \ge 1)$.

b. Tính $P(X \ge \frac{1}{2}|X + Y \ge 1)$.

c. Tính giá trị trung bình của X trong điều kiện $Y = \frac{1}{2}$.

Câu 3.(2 điểm) Khảo sát ngẫu nhiên 178 phiên giao dịch của một loại cổ phiếu thì thấy có 112 phiên cổ phiếu này tăng giá. Có ý kiến cho rằng tỉ lệ phiên giao dịch tăng giá của loại cổ phiếu này trên 60%. Hãy tính giá trị kiểm định và cho nhận xét về ý kiến trên với mức ý nghĩa 5%.

Câu 4.(2 điểm) Để kiểm tra tuổi thọ của bóng đèn do nhà máy M sản xuất, người ta khảo sát ngẫu nhiên 256 bóng đèn thì thấy tuổi thọ trung bình là 1642,5 (giờ) với độ lệch chuẩn có hiệu chỉnh là 226,8281 (giờ). Hãy ước lượng tuổi thọ trung bình của bóng đèn do nhà máy M sản suất với độ tin cậy 95%.

Câu 5.(1 điểm) Xét 2 biến ngẫu nhiên Y, Z có số liệu như sau:

Y	16	19	22	21	19	19	26	21	22	25
Z	7	10	25	13	12	15	33	23	19	27

Tính hệ số tương quan và viết phương trình hồi qui tuyến tính của Y theo Z.

Hết

Cán bộ coi thi không giải thích gì thêm.