b. Tính $P(Y = 3|X = 2)$.

$$P(Y = 3|X = 2) = \frac{P(Y = 3, X = 2)}{P(X = 2)} = \frac{0,2}{0,35} = 0,5714$$

Ví dụ 4.5 Một chương trình bao gồm hai mô-đun. Đặt X là số lỗi trong mô-đun 1 và Y là số lỗi trong mô-đun 2 có xác suất đồng thời như sau $P(0,0) = P(0,1) = P(1,0) = 0,2; P(1,1) = P(1,2) = P(1,3) = 0,1; P(0,2) = P(0,3) = 0,05$.

a. Tìm phân phối xác suất thành phần của X.

b. Tìm phân phối của tổng số lỗi trong chương trình.

c. Các lỗi trong hai mô-đun có xảy ra độc lập hay không?

d. Giả sử chương trình có lỗi. Tính xác suất mô-đun 1 có lỗi.

e. Giả sử mô-đun 1 có lỗi. Tính xác suất mô-đun 2 có lỗi.

Giải. Bảng phân phối xác suất đồng thời của như sau

X \ Y	0	1	2	3
0	0,2	0,2	0,05	0,05
1	0,2	0,1	0,1	0,1

4.2 Vectơ ngẫu nhiên liên tục

Định nghĩa 4.6 Cho X, Y là các biến ngẫu nhiên liên tục.

1. **Hàm mật độ xác suất đồng thời** (joint probability density function) của hai biến ngẫu nhiên là một hàm $f(x,y) \geq 0$ thỏa mãn

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y) \mathrm{d}x \mathrm{d}y = 1.$$

2. Hàm mật độ xác suất thành phần (marginal probability density function) của X và Y được lần lượt xác định như sau

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

3. **Hàm phân phối xác suất đồng thời** (joint probability density function) của hai biến ngẫu nhiên

$$F(x,y) = P(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u,v) \mathrm{d}v \mathrm{d}u.$$

Định lý 4.7 Cho X, Y là các BNN liên tục. Khi đó

1. $P(a \leq X \leq b, c \leq Y \leq d) = \int_c^d \int_a^b f(x,y)\mathrm{d}x\mathrm{d}y$
2. $F(a,b) = P(X \leq a, Y \leq b) = \int_{-\infty}^b \int_{-\infty}^a f(x,y)\mathrm{d}x\mathrm{d}y$
3. $f(x,y) = \dfrac{\partial^2}{\partial x \partial y} F(x,y)$.

Định lý 4.8 Cho X, Y là các BNN liên tục. Các điều sau là tương đương

1. X, Y là độc lập
2. $f(x,y) = f_X(x).f_Y(y), \forall x, y$
3. $F(x,y) = F_X(x).F_Y(y), \forall x, y$.

Ví dụ 4.9 Cho hàm mật độ xác suất đồng thời của các BNN X, Y như sau

$$f(x,y) = \begin{cases} cx(x-y), & 0 < x < 2, -x < y < x \\ 0, & \text{các trường hợp khác} \end{cases}$$

a. Tìm c.
b. Tìm hàm mật độ thành phần của X và Y.

Giải. a. Ta có

$$\begin{aligned} 1 &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y)\mathrm{d}x\mathrm{d}y \\ &= \int_0^2 \int_{-x}^{x} cx(x-y)\mathrm{d}y\mathrm{d}x \\ &= \int_0^2 2cx^3\mathrm{d}x \\ &= 8c \end{aligned}$$

Suy ra $c = \dfrac{1}{8}$.
b. Tìm hàm mật độ thành phần của X.

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{+\infty} f(x,y)\mathrm{d}y \\ &= \int_{-x}^{x} \frac{1}{8}x(x-y)\mathrm{d}y = \frac{x^3}{4} \end{aligned}$$

Như vậy,

$$f_X(x) = \begin{cases} \dfrac{x^3}{4}, & 0 < x < 2 \\ 0, & \text{các trường hợp khác} \end{cases}$$

Tìm hàm mật độ thành phần của Y.

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{+\infty} f(x,y)\mathrm{d}x \\ &= \int_0^2 \frac{1}{8}x(x-y)\mathrm{d}x = \frac{1}{3} - \frac{y}{4} \end{aligned}$$

Như vậy,

$$f_Y(y) = \begin{cases} \dfrac{1}{3} - \dfrac{y}{4}, & -x < y < x \\ 0, & \text{các trường hợp khác} \end{cases}$$

Ví dụ 4.10 Cho X, Y là các biến ngẫu nhiên liên tục có hàm mật độ xác suất đồng thời như sau

$$f(x,y) = \begin{cases} 6e^{-2x-3y}, & x > 0, y > 0 \\ 0, & \text{các trường hợp còn lại} \end{cases}$$

a. Tính $P(2 < X < 3, 1 < Y < 2)$.
b. Tìm hàm phân phối xác suất đồng thời của X và Y.
c. Tìm hàm mật độ thành phần của X và Y.
d. X và Y có độc lập không?
Giải. a. Tính $P(1 < X < 2, 2 < Y < 3)$.

$$P(1 < X < 2, 2 < Y < 3) = \int_2^3 \int_1^2 6e^{-2x-3y}\mathrm{d}x\mathrm{d}y = \ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

b. Tìm hàm phân phối xác suất đồng thời của X và Y.

$$F(x,y) = P(X \le x, Y \le y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u,v)\mathrm{d}v\mathrm{d}u$$

Với $x > 0, y > 0$, ta có

$$\begin{aligned} F(x,y) &= \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\mathrm{d}v\mathrm{d}u = \int_{0}^{x}\int_{0}^{y} 6e^{-2u-3v}\mathrm{d}v\mathrm{d}u \\ &= (1-e^{-2x})(1-e^{-3y}) \end{aligned}$$

Như vậy,

$$F(x,y) = \begin{cases} (1-e^{-2x})(1-e^{-3y}), & x > 0, y > 0 \\ 0, & \text{các trường hợp còn lại} \end{cases}$$

c. Tìm hàm mật độ thành phần của X. Với $x > 0$,

$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)\mathrm{d}y = \int_{0}^{+\infty} 6e^{-2x-3y}\mathrm{d}y = \dots\dots\dots\dots\dots\dots$$

Suy ra

$$f_X(x) = \begin{cases} \dots\dots\dots\dots\dots\dots, & x > 0 \\ 0, & \text{các trường hợp còn lại} \end{cases}$$

Tìm hàm mật độ thành phần của Y. Với $y > 0$,

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x,y)\mathrm{d}x = \int_{0}^{+\infty} 6e^{-2x-3y}\mathrm{d}x = \dots\dots\dots\dots\dots\dots$$

Suy ra

$$f_Y(y) = \begin{cases} \dots\dots\dots\dots\dots\dots, & y > 0 \\ 0, & \text{các trường hợp còn lại} \end{cases}$$

d. Ta thấy

$$f_X(x)f_Y(y) = \begin{cases} \dots\dots\dots\dots\dots\dots, & x > 0, y > 0 \\ 0, & \text{các trường hợp còn lại} \end{cases}$$

Suy ra

$$f_X(x)f_Y(y)\dots\dots\dots f(x,y)$$

Như vậy X, Y ....

Ví dụ 4.11 Cho X, Y là các biến ngẫu nhiên liên tục có hàm mật độ đồng thời

$$f(x,y) = \begin{cases} 4xy & ,0 < x < 1, 0 < y < 1 \\ 0, & \text{trường hợp khác} \end{cases}$$

Tính $P(X < Y)$.

Giải.

$$\begin{aligned} P(X < Y) &= \int_0^1 \int_0^y 4xy \mathrm{d}x\mathrm{d}y \\ &= \int_0^1 \left(2x^2y\Big|_0^y \right) \mathrm{d}y \\ &= \int_0^1 2y^3 \mathrm{d}y = \frac{1}{2} \end{aligned}$$

Ví dụ 4.12 Cho hàm mật độ xác suất đồng thời của các BNN X, Y như sau

$$f(x,y) = \begin{cases} Ce^{-x}e^{-2y}, & x > 0, y > 0 \\ 0, & \text{các trường hợp khác} \end{cases}$$

a. Tìm C.
b. Tính $P(X > 1, Y < 1)$.
c. Tính $P(X < Y)$.
d. Tính $P(X < a)$.

Giải. ....
....
....
....
....
....
....
....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

Định nghĩa 4.13 Cho X, Y là các biến ngẫu nhiên liên tục.

1. **Hàm mật độ xác suất có điều kiện** của X khi đã biết $Y = y$

$$f_X(x|y) = \begin{cases} \dfrac{f(x,y)}{f_Y(y)}, & f_Y(y) > 0 \\ 0, & \text{các trường hợp khác} \end{cases}$$

2. **Hàm mật độ xác suất có điều kiện** của Y khi đã biết $X = x$

$$f_Y(y|x) = \begin{cases} \dfrac{f(x,y)}{f_X(x)}, & f_X(x) > 0 \\ 0, & \text{các trường hợp khác} \end{cases}$$

3. Trung bình thành phần của X, Y lần lượt là

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

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

Ví dụ 4.14 Đặt biến ngẫu nhiên X biểu thị thời gian cho đến khi máy chủ kết nối với máy của bạn (tính bằng mili giây) và đặt Y biểu thị thời gian cho đến khi máy chủ ủy quyền cho bạn với tư cách là người dùng hợp lệ (tính bằng mili giây). Giả sử X, Y là các biến ngẫu nhiên liên tục có hàm mật độ xác suất đồng thời như sau

$$f(x,y) = \begin{cases} 6.10^{-6}e^{-0{,}001x-0{,}002y}, & 0 < x < y \\ 0, & \text{các trường hợp khác} \end{cases}$$

a. Tìm $f_Y(y|x)$.
b. Tìm $P(Y > 2000|X = 1500)$ và $E(Y|X = 1500)$.
Giải. a. Tìm hàm mật độ thành phần của X. Với $x > 0$, ta có

$$\begin{aligned} f_X(x) &= \int_x^{+\infty} f(x,y)\mathrm{d}y \\ &= \int_x^{+\infty} 6.10^{-6}e^{-0{,}001x-0{,}002y}\mathrm{d}y \\ &= 6.10^{-6}.e^{-0{,}001x}\left(\frac{e^{-0{,}002y}}{-0,002}\Big|_x^{+\infty}\right) \\ &= 0,003.e^{-0{,}003x} \end{aligned}$$

Hàm mật độ có điều kiện của Y. Với $0 < x < y$, ta có

$$\begin{aligned} f_Y(y|x) &= \frac{f(x,y)}{f_X(x)} \\ &= \frac{6.10^{-6}e^{-0{,}001x-0{,}002y}}{0,003.e^{-0{,}003x}} \\ &= 0,002.e^{0{,}002x-0{,}002y} \end{aligned}$$

b. Tìm $P(Y > 2000|X = 1500)$.

$$\begin{aligned} P(Y > 2000|X = 1500) &= \int_{2000}^{+\infty} f_Y(y|x)\mathrm{d}y \\ &= \int_{2000}^{+\infty} 0,002.e^{0{,}002.1500-0{,}002y}\mathrm{d}y \\ &= 0,002.e^{3}\left(\frac{e^{-0{,}002y}}{-0,002}\Big|_{2000}^{+\infty}\right) \\ &= 0,368 \end{aligned}$$

$$
\begin{aligned}
E(Y|X=1500) &= \int_{1500}^{+\infty} y f_Y(y|X=1500)\mathrm{d}y \\
&= \int_{1500}^{+\infty} y0,002.e^{0,002.1500-0,002y}\mathrm{d}y \\
&= 0,002.e^3 \int_{1500}^{+\infty} ye^{-0,002y}\mathrm{d}y \\
&= 0,002.e^3 \left(y\frac{e^{-0,002y}}{-0,002}\Big|_{1500}^{+\infty} - \int_{1500}^{+\infty} \frac{e^{-0,002y}}{-0,002}\mathrm{d}x \right) \\
&= 0,002.e^3 \left(\frac{1500.e^{-3}}{0,002} + \frac{e^{-3}}{0,002.0,002} \right) \\
&= 2000
\end{aligned}
$$

Ví dụ 4.15 Cho X, Y là các biến ngẫu nhiên liên tục có hàm mật độ xác suất đồng thời như sau

$$
f(x,y) = \begin{cases} \frac{2}{5}(2x+3y) & ,0 \le x \le 1, 0 \le y \le 1 \\ 0 & ,\ \text{các trường hợp còn lại} \end{cases}
$$

a. Tìm hàm mật độ có điều kiện $f_Y(y|x)$.

b. Tính $P(\frac{1}{4} < Y < 1|X = \frac{3}{4})$.

Giải. a. Hàm mật độ thành phần của X. Với $0 \le x \le 1, 0 \le y \le 1$

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

....

....

....

....

....

Tìm hàm mật độ điều kiện $f_Y(y|x)$.

$$f_Y(y|x) = \begin{cases} \dfrac{f(x,y)}{f_X(x)}, & 0 \leq y \leq 1 \\ 0, & \text{các trường hợp khác} \end{cases}$$

Định lý 4.16

Cho X, Y là các biến ngẫu nhiên và một hàm $h(X, Y)$. Kỳ vọng của hàm $h(X, Y)$, ký hiệu là $E(h(X, Y))$, được xác định như sau

1. Nếu X, Y là các biến ngẫu nhiên rời rạc thì

$$E(h(X,Y)) = \sum_x \sum_y h(x,y)P(x,y)$$

2. Nếu X, Y là các biến ngẫu nhiên liên tục có hàm mật độ đồng thời $f(x, y)$ thì

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

Ví dụ 4.17 Cho X, Y là các biến ngẫu nhiên liên tục có hàm mật độ đồng thời xác định như sau

$$f(x,y) = \begin{cases} 24xy, & 0 \leq x \leq 1; 0 \leq y \leq 1; x+y \leq 1 \\ 0, & \text{các trường hợp khác} \end{cases}$$

và hàm $h(X, Y) = 0,75 + 0,75X + 1,5Y$. Tính $E(h(X, Y))$.

Giải. Ta có

$$\begin{aligned} E(h(X,Y)) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h(x,y)f(x,y)dxdy \\ &= \int_0^1 \int_0^{1-y} (0,75 + 0,75x + 1,5y)24xydxdy \\ &= 1,65. \end{aligned}$$

Ví dụ 4.20 Cho X, Y là các biến ngẫu nhiên liên tục có hàm mật độ xác suất đồng thời như sau

$$f(x,y) = \begin{cases} x^2 + \dfrac{xy}{3} & , 0 \le x \le 1, 0 \le y \le 2 \\ 0 & , \text{ các trường hợp còn lại} \end{cases}$$

a. Tìm hàm mật độ có điều kiện $f_X(x|y)$.

b. Tính $E(X|Y = \dfrac{1}{2})$.

Giải. ....

BÀI TẬP

Bài 4.1 Cho hai biến ngẫu nhiên X, Y có phân phối xác suất đồng thời như sau

$X \backslash Y$	1	2	3	4
1	0	0,06	0,06	0,1
2	0,1	0,1	0,04	0,04
3	0,4	0,1	0	0

a. X và Y có độc lập không? Vì sao?
b. Tính xác suất $P(X + Y \leq 3)$
c. Tính $P(X > 1|Y = 2)$
Bài 4.2 Cho X, Y là các biến ngẫu nhiên rời rạc có phân phối xác suất đồng thời như sau

$X \backslash Y$	0	1	2
0	0,1	0,04	0.02
1	0,08	0,2	0,06
2	0,06	0,14	0,3

a. Tính $P(X \leq 1, Y \leq 1)$.
b. Tính $P(X > 0, Y > 0)$
c. Tìm bảng phân phối xác suất thành phần của X và Y.
d. X, Y có độc lập không? Tại sao?

Bài 4.3 Một hộp có 3 bi đỏ, 2 bi vàng và 3 bi xanh. Lấy ngẫu nhiên 2 bi từ hộp. Gọi X là số bi đỏ và Y là số bi vàng trong 2 bi lấy ra.
a. Lập bảng phân phối xác suất đồng thời của X và Y.
b. Tính $P(X + Y \leq 1)$. (ĐS: 9/14)
c. Tìm các phân phối xác suất thành phần của X và Y.
Bài 4.4 Cho X, Y là các biến ngẫu nhiên liên tục có hàm mật độ đồng thời như sau

$$f(x, y) = \begin{cases} kx^3y^2, & 0 \leq x \leq 2, 0 \leq y \leq 2 \\ 0, & \text{các trường hợp khác} \end{cases}$$

a. Tìm k.
b. Tìm hàm mật độ thành phần của X và Y.
c. Tìm hàm phân phối xác suất thành phần của X.
Bài 4.5 Cho X, Y là hai biến ngẫu nhiên có hàm mật độ xác suất đồng thời

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

a. Tìm C.
b. Tính $P(X \leq 1, Y > 2)$.
c. Tính $P(X \leq 1|Y > 2)$
Bài 4.6 Giả sử X, Y là tuổi thọ trung bình của 2 thiết bị có hàm mật độ đồng thời

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

a. Tìm hàm mật độ thành phần của X và Y.
b. Tìm $E(X + Y), V(X + Y)$.

Bài 4.7 Giả sử X, Y là hai biến ngẫu nhiên liên tục có hàm mật độ đồng thời

$$f(x,y) = \begin{cases} e^{-y(x+1)}, & 0 \le x; 0 \le y \\ 0, & \text{các trường hợp khác} \end{cases}$$

a. Tìm hàm mật độ thành phần của X và Y.
b. X, Y có độc lập không?
c. Tìm $P(0 < X < 1 | Y = 2)$.

Bài 4.8 Giả sử X, Y là hai biến ngẫu nhiên liên tục có hàm mật độ đồng thời

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

a. Tìm k.
b. Tìm hàm mật độ thành phần của X và Y.
c. X, Y có độc lập không?

Bài 4.9 Giả sử X, Y là hai biến ngẫu nhiên liên tục có hàm mật độ đồng thời

$$f(x,y) = \begin{cases} ke^{-x-y}, & 0 < y < x \\ 0, & \text{các trường hợp khác} \end{cases}$$

a. Tìm k.
b. Tìm hàm mật độ thành phần của X và Y.
c. X, Y có độc lập không?

Bài 4.10 Cho X, Y là các biến ngẫu nhiên liên tục có hàm mật độ đồng thời

$$f(x,y) = \begin{cases} C(x^2 + y), & -1 \le x < 1, 0 \le y \le 1. \\ 0, & \text{các trường hợp khác} \end{cases}$$

a. Tìm C.
b. Tìm các hàm mật độ thành phần của X và Y. Các biến ngẫu nhiên X, Y có độc lập không?
c. Tính $P(Y < 0,6)$ và $P(Y < 0,6 | X < 0,5)$.

Bài 4.11 Giả sử X, Y là hai biến ngẫu nhiên liên tục có hàm mật độ đồng thời

$$f(x,y) = \begin{cases} x + y, & 0 \le x \le 1; 0 \le y \le 1 \\ 0, & \text{các trường hợp khác} \end{cases}$$

a. Tìm $P(X < \frac{1}{2}, Y > \frac{1}{4})$.
b. Tìm $P(X + Y < 1)$.
c. Tìm hàm mật độ thành phần của X và Y.

Bài 4.12 Giả sử X, Y là hai biến ngẫu nhiên liên tục có hàm mật độ đồng thời

$$f(x,y) = \begin{cases} cxy, & 0 \le y \le x; 0 \le x \le 2 \\ 0, & \text{các trường hợp khác} \end{cases}$$

a. Tìm c.
b. Tìm hàm mật độ thành phần của X và Y. Hai biến X và Y có độc lập không?
c. Tìm $P(X < \frac{1}{2}, Y < \frac{3}{4})$.
d. Tìm $P(X \le \frac{1}{2} | Y < \frac{3}{4})$.

Bài 4.13 Giả sử X, Y là hai biến ngẫu nhiên liên tục có hàm mật độ đồng thời

$$f(x,y) = \begin{cases} cx^2y, & 0 < x^2 \le y < 1, x > 0 \\ 0, & \text{các trường hợp khác} \end{cases}$$

a. Tìm c.
b. Tìm $P(0 < X < \frac{3}{4}, \frac{1}{4} \le Y < 1)$.

Bài 4.14 Giả sử X, Y là hai biến ngẫu nhiên liên tục có hàm mật độ đồng thời

$$f(x,y) = \begin{cases} x+y, & 0 \le x \le 1, 0 \le y \le 1 \\ 0, & \text{các trường hợp khác} \end{cases}$$

a. Tìm hàm mật độ thành phần của X, Y
b. X và Y có độc lập không?
c. Tìm hiệp phương sai và hệ số tương quan của X và Y.

Bài 4.15 Giả sử X, Y là hai biến ngẫu nhiên liên tục có hàm mật độ đồng thời

$$f(x,y) = \begin{cases} k(x+y), & 0 \le x \le 2, 0 \le y \le 2 \\ 0, & \text{các trường hợp khác} \end{cases}$$

a. Tìm k.
b. Tìm các hàm mật độ có điều kiện $f_X(x|y), f_Y(y|x)$
c. Tìm $P(0 \le Y \le 0,5|X = 1)$.

Chương 5. Lý thuyết mẫu

Nguyễn Minh Trí

Trường Đại học Công nghệ Thông tin

Ngày 16 tháng 4 năm 2023

Ví dụ 6.9 Thu nhập trung bình hàng tháng của 30 hộ dân trong một thành phố được cho như sau (đơn vị triệu đồng)

12.23	16.56	4.39	2.89	1.24	2.17	13.19	9.16	1.42	73.25
1.91	14.64	11.59	6.69	1.06	8.74	3.17	18.13	7.92	4.78
16.85	40.22	2.42	21.58	5.01	1.47	12.24	2.27	12.77	2.76

Tìm khoảng tin cậy 90% của thu nhập trung bình hàng tháng của toàn thành phố. Biết thu nhập trung bình hàng tháng có phân phối chuẩn và có độ lệch chuẩn 14.405.

Giải. ....

....

....

....

....

....

....

....

....

6.2.2 Ước lượng khoảng cho trung bình tổng thể khi chưa biết σ

Bài toán 2. Thu nhập trung bình hàng tháng của **30** hộ dân trong một thành phố được cho như sau (đơn vị triệu đồng)

12.23	16.56	4.39	2.89	1.24	2.17	13.19	9.16	1.42	73.25
1.91	14.64	11.59	6.69	1.06	8.74	3.17	18.13	7.92	4.78
16.85	40.22	2.42	21.58	5.01	1.47	12.24	2.27	12.77	2.76

Tìm khoảng thu nhập trung bình hàng tháng của toàn thành phố với độ tin cậy 90%.

Bài toán 3. Theo một thống kê cho thấy số thu nhập của 7 công nhân trong năm 2021 của một công ty được cho như sau (đơn vị triệu đồng)

54,6	59	60,9	63,1	71,6	84,4	99,3

Giả sử thu nhập trong năm 2021 của công ty có phân phối chuẩn. Tính khoảng thu nhập trung bình của công ty này với độ tin cậy 99%.

- Cho một tổng thể có phân phối chuẩn với trung bình μ.
- Cho $\overline{X}$ là công thức trung bình của mẫu ngẫu nhiên và $\overline{x}, s$ lần lượt là giá trị của trung bình và độ lệch chuẩn của một mẫu có kích thước n.
- Biến ngẫu nhiên

$$T = \frac{\overline{X} - \mu}{s/\sqrt{n}}$$

  có phân phối Student với bậc tự do $n - 1$.
- khoảng tin cậy của μ với độ tin cậy $(1 - \alpha)$ là

$$\left[\overline{x} - t_{\alpha/2}\frac{s}{\sqrt{n}}; \overline{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}\right].$$

Trường hợp 1: Kích thước mẫu $n \geq 30$.

1. $\overline{x}, s$ là trung bình và độ lệch chuẩn của một mẫu cụ thể
2. Đổi biến $Z = \dfrac{\overline{X} - \mu}{s/\sqrt{n}}$, khi đó $Z \sim N(0; 1)$.
3. Tra bảng $A4$, tìm $z_{\alpha/2}$.
4. Khoảng tin cậy của μ với độ tin cậy $1 - \alpha$ là

$$[\overline{x} - z_{\alpha/2}\frac{s}{\sqrt{n}}; \overline{x} + z_{\alpha/2}\frac{s}{\sqrt{n}}].$$

Trường hợp 2: Kích thước mẫu $n < 30$ và tổng thể có phân phối chuẩn

1. Đổi biến $T = \dfrac{\overline{X} - \mu}{s/\sqrt{n}}$, khi đó $T \sim St(n)$.
2. Tra bảng $A5$ dòng $n - 1$, tìm $t_{\alpha/2}$ thỏa mãn $P(T > t_{\alpha/2}) = \dfrac{\alpha}{2}$.
3. Khoảng tin cậy của μ với độ tin cậy $1 - \alpha$ là

$$[\overline{x} - t_{\alpha/2}\frac{s}{\sqrt{n}}; \overline{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}].$$

Cách tìm t_β trong bảng A5: bậc tự do của phân phối Student là $n - 1$. Cột đầu tiên bên trái của bảng A5 là cột bậc tự do, hai hàng đầu tiên bên trên là giá trị của β. Số nằm ở vị trí của giao của hàng tương ứng với bậc tự do $n - 1$ và cột tương ứng với β là giá trị của t_β.

Ví dụ 6.10 Tìm giá trị $t_{0,005}$ với bậc tự do 17. Theo bảng A5, ta có $t_{0,005} = 2,898$.

ν (d.f.)	α, the right-tail probability									
	.10	.05	.025	.02	.01	.005	.0025	.001	.0005	.0001
1	3.078	6.314	12.706	15.89	31.82	63.66	127.3	318.3	636.6	3185
2	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60	70.71
3	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92	22.20
4	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610	13.04
5	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.894	6.869	9.676
6	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959	8.023
7	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408	7.064
8	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041	6.442
9	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781	6.009
10	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587	5.694
11	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437	5.453
12	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318	5.263
13	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221	5.111
14	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140	4.985
15	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073	4.880
16	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015	4.790
17	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965	4.715
18	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.610	3.922	4.648
19	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883	4.590
20	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850	4.539

Ví dụ 6.11 Theo một thống kê cho thấy số thu nhập của 7 công nhân trong năm 2021 của một công ty được cho như sau (đơn vị triệu đồng)

54,6 59 60,9 63,1 71,6 84,4 99,3

Giả sử thu nhập trong năm 2021 của công ty có phân phối chuẩn. Tính khoảng thu nhập trung bình của công ty này với độ tin cậy 99%.

Giải.

- Trung bình mẫu: $\overline{x} =$ ....
- Độ lệch chuẩn mẫu: $s =$ ....
- Tìm $t_{\alpha/2}$ với độ tin cậy $1 - \alpha = 0,99$ và bậc tự do 6. Ta có $t_{\alpha/2} =$ ....
- Khoảng tin cậy cần tìm

$$\overline{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} \leq \mu \leq \overline{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}$$

$$70,414 - 3,707\frac{16,103}{\sqrt{7}} \leq \mu \leq 70,414 + 3,707\frac{16,103}{\sqrt{7}}$$

$$47,852 \leq \mu \leq 92,976$$

Ví dụ 6.12 Kiểm tra tuổi thọ (tính bằng giờ) của 50 bóng đèn do nhà máy A sản xuất, người ta được bảng số liệu sau

Tuổi thọ	3300	3500	3600	4000
Số bóng đèn	10	20	12	8

a. Ước tính tuổi thọ trung bình của các bóng đèn do nhà máy A sản xuất với độ tin cậy 97%.
b. Dựa vào mẫu trên để ước lượng tuổi thọ trung bình của các bóng đèn do nhà máy A sản xuất có độ chính xác 59,02 giờ thì phải đảm bảo độ tin cậy là bao nhiêu?
c. Dựa vào mẫu trên, nếu muốn ước lượng tuổi thọ trung bình của các bóng đèn do nhà máy A sản suất có độ chính xác nhỏ hơn 40 giờ với độ tin cậy 98% thì cần phải kiểm tra tối thiểu bao nhiêu bóng đèn?
Giải. a. (Kích thước mẫu bàng 50 và chưa biết độ lệch chuẩn tổng thể)

- Trung bình mẫu: $\overline{x} = \ldots\ldots\ldots\ldots$
- Độ lệch chuẩn mẫu: $s = \ldots\ldots\ldots\ldots$
- Độ tin cậy $1-\alpha = 0,97$. Suy ra $\Phi(z_{\alpha/2}) = 1-\alpha/2 = 0,985$. Do đó $z_{\alpha/2} = \ldots\ldots\ldots\ldots$
- Độ chính xác:

$$z_{\alpha/2}\frac{s}{\sqrt{n}} = 2,17\frac{217,3683}{\sqrt{50}} = \ldots\ldots\ldots\ldots$$

- Khoảng tin cậy của tuổi thọ trung của bóng đèn với độ tin cậy 97% là

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

b. Ta có độ chính xác bằng .... giờ, tức là

$$z_{\alpha/2}\frac{s}{\sqrt{n}} = \ldots\ldots\ldots\ldots$$

Suy ra

$$z_{\alpha/2} = 59,02\frac{\sqrt{n}}{s} = \ldots\ldots\ldots\ldots$$

Do đó

$$\Phi(1,92) = \Phi(z_{\alpha/2}) = 1-\frac{\alpha}{2}.$$

Trang bảng A4, ta có $\Phi(\ldots\ldots\ldots\ldots) = \ldots\ldots\ldots\ldots$ và do đó $\alpha = \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$
Như vậy, độ tin cậy là ....
c. Ta có độ chính xác nhỏ hơn 40 giờ với độ tin cậy 98%, tức là

$$z_{\alpha/2}\frac{s}{\sqrt{n}} < 40.$$

Suy ra

$$\sqrt{n} > z_{\alpha/2}\frac{s}{40}$$

Vì $1-\alpha = 0,98$ nên $\Phi(z_{\alpha/2}) = 1 - \dfrac{\alpha}{2} = 0,99$. Suy ra $z_{\alpha/2} = \ldots\ldots\ldots\ldots$Do đó

$$\sqrt{n} > z_{\alpha/2}\frac{s}{40} = 2,33\frac{217,3683}{40} = \ldots\ldots\ldots\ldots$$

Như vậy, $n >$ .... và do đó cần khảo sát ít nhất................ bóng đèn.

Ví dụ 6.13 Một thống kê cho thấy chi phí (tính bằng triệu) của các mẫu quảng cáo 30-giây trên một số đài truyền hình được cho như sau

14	55	165	9	15	66	23	30	150	22	12	13	54	73	55	41	78

Giả sử chi phí cho một video quảng cáo 30-giây có phân phối chuẩn. Ước tính chi phí trung bình cho một quảng cáo 30-giây trên truyền hình với độ tin cậy 90%.

Giải. . . . .
. . . .
. . . .
. . . .
. . . .
. . . .
. . . .
. . . .
. . . .
. . . .

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

6.2.3 Ước lượng tỉ lệ của tổng thể

Bài toán 4. Thăm dò ý kiến của 100 cử tri được chọn ngẫu nhiên tại một thành phố cho thấy có 80% trong số cử tri này ủng hộ ứng viên A. Với độ tin cậy 98%, hãy ước lượng tỉ lệ của tất cả các cử tri ủng hộ ứng viên A tại thành phố này.

- p : tỉ lệ tổng thể (tỉ lệ phần tử có tính chất $\mathcal{P}$ trong tổng thể)
- f : tỉ lệ mẫu cụ thể (tỉ lệ phần tử có tính chất $\mathcal{P}$ trong mẫu)
- Khi $nf > 5$ và $n(1-f) > 5$ thì tỉ lệ mẫu ngẫu nhiên sẽ xấp xỉ phân phối chuẩn $N(f; \dfrac{f(1-f)}{n})$.
- Với độ tin cậy $1-\alpha$, khoảng tin cậy chứa tỉ lệ tổng thể là

$$[f - z_{\alpha/2}\sqrt{\frac{f(1-f)}{n}}; f + z_{\alpha/2}\sqrt{\frac{f(1-f)}{n}}]$$

  trong đó $\Phi(z_{\alpha/2}) = 1 - \dfrac{\alpha}{2}$ (xem bảng A4).

- Độ chính xác (sai số) là $z_{\alpha/2}\sqrt{\dfrac{f(1-f)}{n}}$.

Theo bất đẳng thức Cauchy, ta có độ chính xác (sai số)

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

Do đó, sai số tối đa trong ước lượng tỉ lệ tổng thể là $\dfrac{z_{\alpha/2}}{2\sqrt{n}}$.

Giải Bài toán 4. Theo đề bài

- Tỉ lệ mẫu cụ thể $f = ...$
- Kích thước mẫu $n = ...$
- Độ tin cậy $1-\alpha = ...$ suy ra $1-\dfrac{\alpha}{2} = ...$ Do đó $z_{\alpha/2} = ...$
- Độ chính xác (sai số) là

$$z_{\alpha/2}\sqrt{\frac{f(1-f)}{n}} = 2,33\frac{0,4}{10} = 0,0932.$$

- Khoảng tin cậy $[0,7068; 0,8932]$.

Như vậy có từ 70,68% đến 89,32% cử tri ủng hộ ứng viên A.

Ví dụ 6.14 Lấy ngẫu nhiên 200 sản phẩm trong một kho hàng để kiểm tra thì thấy có 21 sản phẩm có lỗi.
a. Với độ tin cậy 95%, hãy ước lượng tỉ lệ sản phẩm lỗi của cả kho hàng.
b. Dựa vào mẫu trên, để ước tính tỉ lệ sản phẩm bị lỗi có độ chính xác là 0,035 thì độ tin cậy bằng bao nhiêu?
c. Dựa vào mẫu trên, nếu muốn ước lượng tỉ lệ sản phẩm bị lỗi với độ chính xác nhỏ hơn 0,01 với độ tin cậy 93% thì cần kiểm tra ít nhất bao nhiêu sản phẩm.

Giải. ....

....

....

....

....

....

....

....

....

. . . .
. . . .
. . . .
. . . .
. . . .
. . . .
. . . .
. . . .
. . . .
. . . .
. . . .
. . . .
. . . .
. . . .
. . . .

BÀI TẬP

Bài 6.1 Trọng lượng của những vĩ thuốc do một công ty sản xuất có phân phối chuẩn với độ lệch chuẩn 0,038mg. Một mẫu ngẫu nhiên gồm 10 vĩ thuốc có trọng lượng trung bình 4,87mg. Hãy ước lượng trọng lượng trung bình của các vĩ thuốc do công ty sản xuất với độ tin cậy 95%.
Bài 6.2 Đo đường kính trung bình của một mẫu ngẫu nhiên gồm 100 vòng bi do một máy sản suất trong một tuần có đường kính trung bình 0,824 cm và độ lệch chuẩn mẫu hiệu chỉnh là 0,042 cm. Hãy tìm khoảng tin cậy của tất cả các vòng bi với độ tin cậy 96%.
Bài 6.3 Để nghiên cứu khối lượng rác sinh hoạt thải ra trong một ngày tại một thành phố, người ta khảo sát ngẫu nhiên 400 gia đình. Kết quả khảo sát như sau

Khối lượng (kg/ngày)	0,5	1,5	2,5	3,5	4,5	5,5	6,5	7,5
Số gia đình	10	35	86	132	78	31	18	10

a. Hãy ước tính khối lượng rác trung bình thải ra của toàn bộ các hộ gia đình trong 1 ngày với độ tin cậy 99%. Biết rằng khối lượng rác thải ra có phân phối chuẩn và có độ lệch chuẩn là 0,75 kg.
b. Với mẫu khảo sát trên, nếu ước lượng lượng rác thải hàng ngày này với độ chính xác nhỏ hơn 0,8kg/ngày và độ tin cậy 95% thì cần khảo sát tối thiểu bao nhiêu gia đình?
Bài 6.4 Khảo sát giá (triệu đồng) của 10 loại laptop có RAM 8G tại một số cửa hàng kinh doanh online ta được bảng số liệu sau

18,5   27,5   26,4   17,9   28   27   14,5   22   24   28

Hãy ước tính giá trung bình của các loại laptop có RAM 8G với độ tin cậy 95%.

Bài 6.5 Một tỉnh nọ có 1 triệu thanh niên trên 18 tuổi. Người ta khảo sát ngẫu nhiên 20 000 thanh niên của tỉnh này về trình độ học vấn thì thấy có 12575 thanh niên đã tốt nghiệp THPT. Hãy ước tính tỉ lệ thanh niên tốt nghiệp THPT của tỉnh này với độ tin cậy 95%.

Bài 6.6 Lấy ngẫu nhiên 200 sản phẩm trong một kho hàng để kiểm tra thì thấy có 21 sản phẩm có lỗi.

a. Dựa vào mẫu trên, để ước tính tỉ lệ sản phẩm bị lỗi có độ chính xác là 0,035 thì độ tin cậy bằng bao nhiêu?

b. Dựa vào mẫu trên, nếu muốn ước lượng tỉ lệ sản phẩm bị lỗi với độ chính xác nhỏ hơn 0,01 với độ tin cậy 93% thì cần kiểm tra ít nhất bao nhiêu sản phẩm.

Bài 6.7 Một nhà sản xuất bóng đèn tuyên bố rằng tuổi thọ của các bóng đèn của họ được phân phối chuẩn với giá trị trung bình là 60.000 giờ và độ lệch chuẩn là 4.000 giờ. Một mẫu ngẫu nhiên gồm 16 bóng đèn có tuổi thọ trung bình là 58.500 giờ. Nếu tuyên bố của nhà sản xuất là đúng thì xác suất giá trị trung bình mẫu là 58.500 hoặc thấp hơn là bao nhiêu? (0,0668)

Bài 6.8 Người ta ước tính rằng 43% sinh viên tốt nghiệp ngành công nghệ thông tin tin rằng một khóa học về lập trình Python là rất quan trọng để có thể tìm được việc làm tại các công ty lớn. Tìm xác suất để hơn một nửa mẫu ngẫu nhiên gồm 80 sinh viên tốt nghiệp ngành công nghệ thông tin có niềm tin này. (0,102)

Bài 6.9 Một mẫu ngẫu nhiên gồm 270 laptop được lấy từ một lượng lớn các laptop cũ để ước tính tỷ lệ laptop có ổ cứng bị lỗi. Nếu trên thực tế, 20% laptop có ổ cứng bị lỗi thì xác suất để tỷ lệ mẫu nằm trong khoảng từ 16% đến 24% là bao nhiêu? (0.905)

Bài 6.10 Thăm dò 500 người dân tại thành phố Hồ Chí Minh về việc xây lại Dinh Độc lập, có 380 người không đồng ý.

a. Hãy ước lượng tỉ lệ người dân TPHCM không đồng ý xây lại Dinh Độc Lập với độ tin cậy 95%.

b. Nếu muốn độ chính xác của ước lượng này là 3% thì độ tin cậy bằng bao nhiêu?

c. Nếu muốn độ chính xác của ước lượng này nhỏ hơn 3% với độ tin cậy 99% thì cần khảo sát ít nhất bao nhiêu người?

Chương 7. Kiểm định giả thuyết

Nguyễn Minh Trí

Trường Đại học Công nghệ Thông tin

Ngày 16 tháng 4 năm 2023

7.1 Các khái niệm

Bài toán. Một hiệu trưởng của một trường THPT tại TPHCM đọc báo thấy rằng điểm trung bình của bài thi Đánh giá năng lực đợt 1 của Đại học Quốc gia TPHCM năm 2022 là 646,1 điểm. Hiệu trưởng nói rằng điểm trung bình của tất cả các học sinh của trường thi đáng giá năng lực lớn hơn 646,1. Sau đó, một phóng viên chọn ngẫu nhiên 50 học sinh của trường đã thi Đánh giá năng lực và thấy rằng điểm trung bình của nhóm học sinh này là 665. Như vậy, có đủ căn cứ để chấp nhận phát biểu của hiệu trưởng không?

Định nghĩa 7.1 Giả thuyết thống kê là một dự đoán về một tham số của tổng thể.

Định nghĩa 7.2

1. **Giả thuyết** (null hypothesis), kí hiệu H_0, là một giả thuyết thống kê nói rằng **không có sự khác biệt** giữa một tham số và một giá trị cụ thể hoặc không có sự khác biệt giữa hai tham số.
2. **Đối thuyết** (alternative hypothesis), ký hiệu H_1, là một giả thuyết thống kê cho biết **có sự khác biệt** giữa một tham số và một giá trị cụ thể, hoặc có sự khác biệt giữa hai tham số.

Kết quả của mỗi kiểm định là *chấp nhận* H_0 **hoặc** *bác bỏ* H_0 *và chấp nhận* H_1.

Các dạng toán kiểm định thường gặp:

Kiểm định 2 phía: Giả thuyết $H_0 : \theta = \theta_0$ và đối thuyết $H_1 : \theta \neq \theta_0$.

Kiểm định 1 phía trái: Giả thuyết $H_0 : \theta = \theta_0$ và đối thuyết $H_1 : \theta < \theta_0$.

Kiểm định 1 phía phải: Giả thuyết $H_0 : \theta = \theta_0$ và đối thuyết $H_1 : \theta > \theta_0$.

Ví dụ 7.3

1. Một nhà nghiên cứu nói rằng những trẻ em uống ít nhất 1 ly sữa mỗi ngày khi trưởng thành sẽ có chiều cao lớn hơn 170cm.
   Ta kiểm định: Giả thuyết $H_0 : \mu = 170$ và đối thuyết $H_1 : \mu > 170$.
2. Một giám đốc của một doanh nghiệp thấy rằng sau dịch Covid-19 mức lương trung bình của công nhân toàn công ty có thay đổi. Mức lương trung bình trước dịch Covid-19 là 8,2 triệu đồng/tháng.
   Ta kiểm định: Giả thuyết $H_0 : \mu = 8,2$ và đối thuyết $H_1 : \mu \neq 8,2$.
3. Một công nhân sản xuất gạch thấy rằng số lượng gạch làm ra trong 1 giờ giảm khi áp dụng quy trình sản xuất mới. Trước đây, trung bình công nhân làm được 35 viên gạch trong một giờ.
   Ta kiểm định: Giả thuyết $H_0 : \mu = 35$ và đối thuyết $H_1 : \mu < 35$.

4 Một nhân viên của một nhà hàng nói rằng thời gian trung bình khách phải chờ để được phục vụ của nhà hàng họ là không quá 10 phút.
Ta kiểm định: Giả thuyết $H_0 : \mu = 10$ và đối thuyết $H_1 : \mu > 10$.

Các sai lầm trong kiểm định giả thuyết

- Sai lầm loại 1: H_0 đúng nhưng bác bỏ H_0
- Sai lầm loại 2: H_0 sai nhưng chấp nhận H_0

Trong thực tế, sai lầm loại 1 là nguy hiểm hơn, do đó ta thiết kế mô hình kiểm định sao cho xác suất sai lầm loại 1 bị chặn bởi một số rất nhỏ α.
Định nghĩa 7.4 Số α được gọi là **mức ý nghĩa** của kiểm định nếu α là xác suất ta bác bỏ H_0 khi H_0 đúng.

7.2 Kiểm định giả thuyết về trung bình

Giả sử biến ngẫu nhiên X có phân phối chuẩn $N(\mu; \sigma^2)$ trong đó μ là trung bình của tổng thể.
Bài toán 1. Ta kiểm định

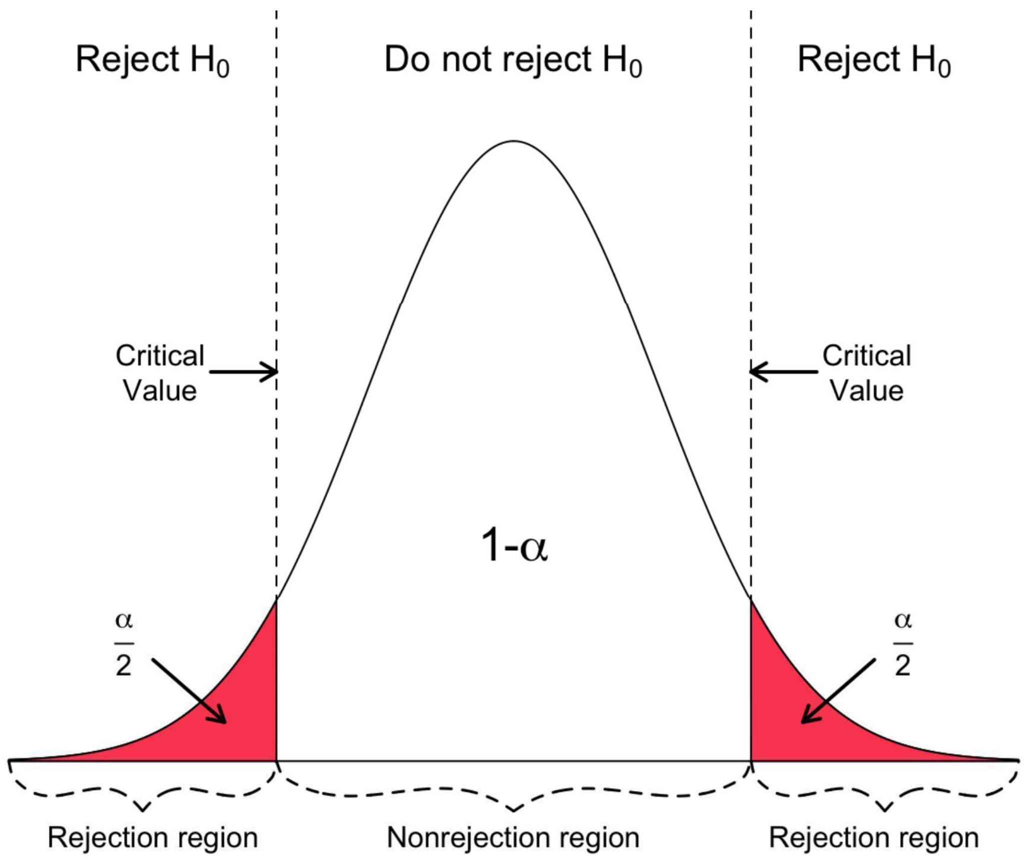
$$H_0 : \mu = \mu_0 \text{ và } H_1 : \mu \neq \mu_0.$$

Đặt $Z = \dfrac{\overline{X} - \mu_0}{\sigma}$ là biến ngẫu nhiên có phân phối chuẩn $N(0; 1)$.
Với mức ý nghĩa α, đặt $z_{\alpha/2}$ (giá trị tới hạn) là giá trị thỏa mãn

$$P(|Z| > z_{\alpha/2}) = \alpha$$

hay

$$\boxed{\Phi(z_{\alpha/2}) = P(Z \leq z_{\alpha/2}) = 1 - \frac{\alpha}{2}}$$



Bài toán 2. Với mức ý nghĩa α, ta kiểm định:

$$H_0 : \mu = \mu_0 \text{ và } H_1 : \mu > \mu_0.$$

Đặt z_α là giá trị thỏa mãn

$$P(Z > z_\alpha) = \alpha$$

hay

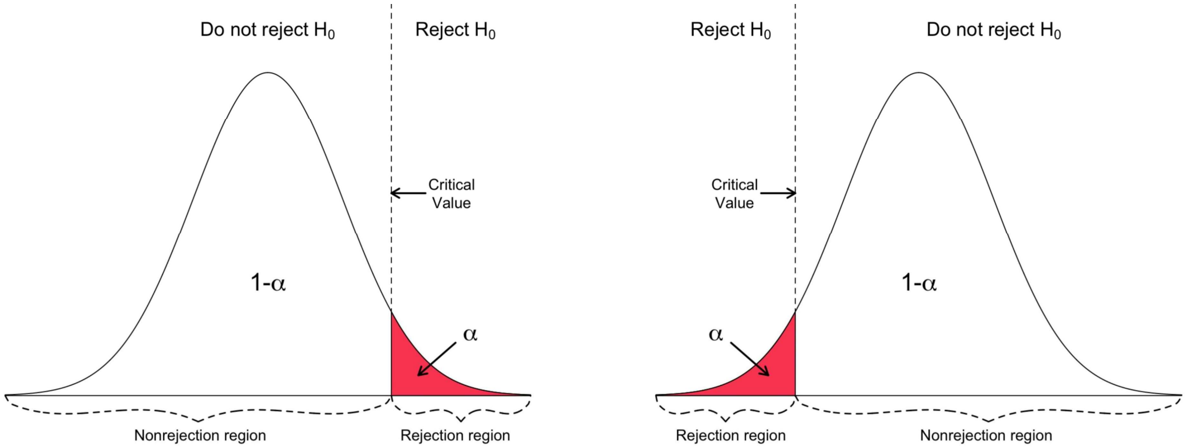
$$\boxed{\Phi(z_\alpha) = P(Z \le z_\alpha) = 1 - \alpha}$$

Bài toán 3. Với mức ý nghĩa α, ta kiểm định:

$$H_0 : \mu = \mu_0 \text{ và } H_1 : \mu < \mu_0.$$

Đặt z_α là giá trị thỏa mãn

$$\boxed{\Phi(z_\alpha) = P(Z < z_\alpha) = \alpha}$$



Cho $\overline{x}$ là trung bình mẫu, n là kích thước mẫu, s là độ lệch chuẩn mẫu.

Trường hợp 1. σ đã biết.	**Trường hợp 2.** σ chưa biết và kích thước mẫu $n \geq 30$.
Tính $z = \dfrac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}$	Tính $z = \dfrac{\overline{x} - \mu_0}{s/\sqrt{n}}$

Kiểm định	Bác bỏ H_0	Chấp nhận H_0
$H_0 : \mu = \mu_0; H_1 : \mu \neq \mu_0.$	$\lvert z \rvert \geq z_{\alpha/2}$	$\lvert z \rvert < z_{\alpha/2}$
$H_0 : \mu = \mu_0; H_1 : \mu > \mu_0.$	$z \geq z_\alpha$	$z < z_\alpha$
$H_0 : \mu = \mu_0; H_1 : \mu < \mu_0.$	$z \leq z_\alpha$	$z > z_\alpha$

Trường hợp 3: Với σ chưa biết và $n < 30$, tổng thể có phân phối chuẩn.
Đặt

$$T = \frac{\overline{X} - \mu_0}{s/\sqrt{n}} \sim St(n)$$

trong đó $\overline{X}$ là trung bình mẫu ngẫu nhiên, n là cỡ mẫu và s là độ lệch chuẩn mẫu.
Với mức ý nghĩa α, đặt $t_{\alpha/2}$ và t_α là các số thực thỏa mãn (xem bảng A5)

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

Đặt

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

trong đó $\overline{x}$ là trung bình mẫu cụ thể.

Kiểm định	Bác bỏ H_0	Chấp nhận H_0
$H_0 : \mu = \mu_0; H_1 : \mu \neq \mu_0.$	$\lvert t \rvert \geq t_{\alpha/2}$	$\lvert t \rvert < t_{\alpha/2}$
$H_0 : \mu = \mu_0; H_1 : \mu > \mu_0.$	$z \geq t_\alpha$	$t < t_\alpha$
$H_0 : \mu = \mu_0; H_1 : \mu < \mu_0.$	$t \leq -t_\alpha$	$t > -t_\alpha$

Ví dụ 7.5 Theo báo cáo "Thị trường IT Việt Nam - Developers Recruitment State 2021" do TopDev công bố cho biết, tính đến quý II/2021, kỹ sư trí tuệ nhân tạo (AI) và máy học (Machine Learning) là vị trí có mức lương trung bình hàng tháng cao nhất trong các kỹ sư IT, đạt 3054 USD (khoảng 70 triệu đồng). Một cuộc khảo sát 30 kỹ sư trí tuệ nhân tạo tốt nghiệp từ một trường đại học X cho thấy họ có mức lương trung bình là 3105 USD/tháng. Hãy kiểm tra kết luận nói rằng các kỹ sư trí tuệ nhân tạo của trường X có mức thu nhập trung bình lớn hơn 3054 USD/tháng với mức ý nghĩa 0,05. Giả sử thu nhập của các kỹ sư trí tuệ nhân tạo có phân phối chuẩn với độ lệch chuẩn tổng thể là 120 USD.

Giải.

- Gọi μ thu nhập trung bình của các kỹ sư trí tuệ nhân tạo
- Ta kiểm định: Giả thuyết $H_0 : \mu = 3054$ và đối thuyết $H_1 : \mu > 3054$
- Theo đề bài, trung bình mẫu là $\overline{x} = 3105$, cỡ mẫu $n = 30$ và độ lệch chuẩn tổng thể $\sigma = 120$
- Vì mức ý nghĩa $\alpha = 0,05$ nên $z_\alpha = 1,65$.
- Đặt
$$z = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{3105 - 3054}{120/\sqrt{30}} = 2,34.$$
- Vì $z = 2,34 > 1,65$ nên bác bỏ H_0.
- Ta đồng ý với tuyên bố lương trung bình của các kỹ sư trí tuệ nhân tạo nhiều hơn 3054 USD/tháng.

Ví dụ 7.6 Một nhà nghiên cứu nói rằng trung bình giá tiền của một đôi giày thể thao nam là ít hơn 80 USD. Chọn ngẫu nhiên 36 đôi giày thể thao nam để khảo sát giá, ta được kết quả sau (USD/đôi)

60	70	75	55	80	55	50	40	80
70	50	95	120	90	75	85	80	60
110	65	80	85	85	45	75	60	90
90	60	95	110	85	45	90	70	70

Giả sử giá giày có phân phối chuẩn với độ lệch chuẩn là 19,2 USD. Tuyên bố của nhà nghiên cứu có chấp nhận được không với mức ý nghĩa 10%?

Giải.

- Gọi μ giá trung bình của một đôi giày thể thao nam.
- Ta điểm định: Giả thuyết $H_0 : \mu = 80$ và đối thuyết $H_1 : \mu < 80$
- Theo đề bài, trung bình mẫu là $\overline{x} = 75$, cỡ mẫu $n = 36$ và độ lệch chuẩn tổng thể $\sigma = 19,2$
- Vì mức ý nghĩa $\alpha = 0,1$ nên $z_\alpha = -1,28$.

- Đặt
$$z = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{75 - 80}{19,2/\sqrt{36}} = -1,56.$$
- Vì $z = -1,56 < -1,28$ nên bác bỏ H_0.
- Ta đồng ý với nhận xét giá tiền trung bình của một đôi giày thể thao nam ít hơn 80 USD.

Ví dụ 7.7 Một nhà nghiên cứu nói rằng trung bình giá tiền của một đôi giày thể thao nam là ít hơn 80 USD. Chọn ngẫu nhiên 16 đôi giày thể thao nam để khảo sát giá, ta được kết quả sau (USD/đôi)

60	70	75	55	80	55	50	40
70	50	95	120	90	75	85	80

Giả sử giá giày có phân phối chuẩn. Tuyên bố của nhà nghiên cứu có chấp nhận được không với mức ý nghĩa 10%?

Giải.

- Gọi μ giá trung bình của một đôi giày thể thao nam.
- Ta điểm định: Giả thuyết $H_0 : \mu = 80$ và đối thuyết $H_1 : \mu < 80$
- Theo đề bài, trung bình mẫu là $\overline{x} = 71,875$, cỡ mẫu $n = 16$ và độ lệch chuẩn mẫu $s =$ ....
- Vì mức ý nghĩa $\alpha = 0,1$ nên $t_\alpha =$ .... (bậc tự do 15).
- Đặt
$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \frac{.........}{......../\sqrt{.....}} = ...$$
- Vì .... nên ....
- ....

7.3 Kiểm định giả thuyết về tỉ lệ

Bài toán 4. Một người ăn kiêng nói rằng có 60% số người không ăn bánh ngọt. Một cuộc khảo sát 200 người, ta thấy có 128 người nói rằng họ không ăn bánh ngọt. Với mức ý nghĩa 5%, ta có thể bác bỏ tuyên bố của người ăn kiêng này không?

- Gọi p (f, F) là tỉ lệ các phần tử có tính chất $\mathcal{P}$ trong tổng thể (trong mẫu cụ thể, mẫu ngẫu nhiên).
- Kiểm định giả thuyết
$$H_0 : p = p_0$$
- Chọn một mẫu ngẫu nhiên có kích thước n.
- Với n đủ lớn, biến ngẫu nhiên
$$Z = \frac{F - p_0}{\sqrt{p_0(1-p_0)/n}} \sim N(0;1)$$

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

Với mức ý nghĩa α.

Kiểm định	Bác bỏ H_0	Chấp nhận H_0
$H_0 : p = p_0; H_1 : p \neq p_0.$	$\lvert z\rvert \geq z_{\alpha/2}$	$\lvert z\rvert < z_{\alpha/2}$
$H_0 : p = p_0; H_1 : p > p_0.$	$z \geq z_\alpha$	$z < z_\alpha$
$H_0 : p = p_0; H_1 : p < p_0.$	$z \leq z_\alpha$	$z > z_\alpha$

Giải Bài toán 4

- Gọi p là tỉ lệ người không ăn bánh ngọt.
- Ta kiểm định: Giả thuyết $H_0 : p = 60\%$ và đối thuyết $H_1 : p \neq 60\%$
- Tỉ lệ mẫu là $f = \dfrac{128}{200} = 0,64$ và cỡ mẫu $n = 200.$
- Vì mức ý nghĩa $\alpha = 0,05$ nên $z_{\alpha/2} = 1,96.$
- Đặt
$$z = \frac{f - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{0,64 - 0,6}{\sqrt{0,6(1-0,6)/200}} = 1,15.$$
- Vì $\lvert z\rvert = 1,15 < 1,96$ nên chấp nhận H_0.
- Ta đồng ý với phát biểu rằng có 60% người không ăn bánh ngọt.

Ví dụ 7.8 Một giáo viên nói rằng lương trung bình của giáo viên tại TPHCM ít hơn 16 triệu đồng/tháng trong năm 2021. Chọn ngẫu nhiên 8 giáo viên thì thấy lương hàng tháng (đơn vị là triệu đồng) của họ trong năm 2021

16 15,6 16 15,5 17 15,5 16 15,5

Giả sử lương của giáo viên có phân phối chuẩn. Với mức ý nghĩa 10%, tuyên bố của giáo viên đó có chấp nhận được không?

Giải.

- Gọi μ tiền lương trung bình hàng tháng của giáo viên trong năm 2021.
- Kiểm định: Giả thuyết $H_0 : \mu =$ ..... và đối thuyết $H_1 : \mu$.........
- Trung bình mẫu là $\overline{x} =$ .... cỡ mẫu $n = 8$ và độ lệch chuẩn mẫu là $s =$ ....
- Mức ý nghĩa $\alpha = 0,1$ suy ra .... (xem bảng A5)
- Đặt

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

- Vì $t =$ .... nên .... H_0.
- Ta .... với tuyên bố tiền lương trung bình trong một tháng của giáo viên ít hơn 16 triệu đồng.

Ví dụ 7.9 Một lập trình viên nói rằng có hơn 25% các lập trình viên đã học ngôn ngữ lập trình Python. Một cuộc khảo sát 200 lập trình viên tại một thành phố nọ, người ta thấy có 63 lập trình viên đã học Python. Với mức ý nghĩa 5%, hãy kết luận về nhận định của lập trình viên trên.

Giải. ....

....

....

....

....

....

....

....

....

....

....

....

BÀI TẬP

Bài 7.1 Giám đốc một công ty phát biểu rằng thu nhập trung bình của công nhân trong công ty của ông là hơn 6,7 triệu đồng/tháng. Khảo sát ngẫu nhiên 40 công nhân, người ta thấy rằng thu nhập trung bình của họ là 7,25 triệu đồng/tháng và độ lệch chuẩn mẫu là 1,02 triệu đồng. Với mức ý nghĩa 1%, phát biểu của giám đốc có chấp nhận được không?

Bài 7.2 Một nhà báo nói rằng học phí trung bình của 4 năm học đại học của một sinh viên hơn 11,4 triệu đồng. Cô ấy chọn ngẫu nhiên 36 ngành học 4 năm tại các trường đại học và nhận thấy mức học phí trung bình của 36 ngành này là 11,9 triệu đồng. Biết độ lệch chuẩn của tổng thể là 1,318 triệu đồng. Với mức ý nghĩa 5%, phát biểu của nhà báo đó có chấp nhận được không?

Bài 7.3 Một nhân viên bán hàng tại một cửa hàng laptop nói rằng tuổi thọ trung bình của laptop hiệu Z là 30000 giờ. Một cuộc khảo sát 40 laptop hiệu Z cho thấy tuổi thọ trung bình của chúng là 30456 giờ. Biết rằng độ lệch chuẩn tổng thể là 1684 giờ. Với mức ý nghĩa 10%, phát biểu của nhân viên bán hàng có chấp nhận được không?

Bài 7.4 Một báo cáo cho biết rằng trung bình số lần mua hàng online của một phụ nữ trong một tháng là 5,8 lần. Một nhà nghiên cứu chọn ngẫu nhiên 20 phụ nữ và thu được bảng số liệu về số lần mua hàng online trong một tháng như sau

3	2	1	3	7	2	9	4	6	6
8	0	5	6	4	2	1	3	4	1

Giả sử số lần mua hàng online của phụ nữ có phân phối chuẩn. Với mức ý nghĩa 5%, hãy kết luận về báo cáo trên.

Bài 7.5 Một công ty cung cấp dịch vụ internet nói rằng có 40% khách hàng của họ gặp sự cố về đường truyền trong một năm. Một nhóm gồm 100 khách hàng được chọn và người ta thấy rằng có 37 khách hàng gặp sự cố về đường truyền. Với mức ý nghĩa 1%, hãy kết luận về tuyên bố của công ty cung cấp dịch vụ internet.

Bài 7.6 Một quy trình sản xuất các chai dầu gội đầu, khi vận hành chính xác, sẽ tạo ra các chai có trọng lượng trung bình là 200 gam. Một mẫu ngẫu nhiên gồm chín chai từ một lần sản xuất duy nhất mang lại các trọng lượng hàm lượng sau (tính bằng gam):

201,4 109,7 109,7 200,6 200,8 200,1 190,7 200,3 200,9

Giả sử rằng trọng lượng của các chai dầu gội này có phân phối chuẩn. Hãy kiểm tra ở mức 5% đối với giả thuyết rằng quy trình đang vận hành chính xác.

Bài 7.7 Một lập trình viên nói rằng có hơn 25% các lập trình viên đã học ngôn ngữ lập trình Python. Một cuộc khảo sát 200 lập trình viên tại một thành phố nọ, người ta thấy có 63 lập trình viên đã học Python. Với mức ý nghĩa 5%, hãy kết luận về nhận định của lập trình viên trên.

Bài 7.8 Một báo cáo cho thấy rằng có ít hơn 78% sinh viên sử dụng Google Translate khi đọc các trang web bằng tiếng Anh. Chọn ngẫu nhiên 143 sinh viên tại một trường đại học và người ta thấy có 100 sinh viên sử dụng Google Translate khi đọc các trang web tiếng Anh. Với mức ý nghĩa 5%, hãy kết luận về nhận định của báo cáo trên.

Bài 7.9 Market Research, Inc., muốn biết liệu người mua có nhạy cảm với giá của các mặt hàng được bán trong siêu thị hay không. Một mẫu ngẫu nhiên gồm 802 người mua sắm đã được thu thập và 378 người trong số những người mua sắm ở siêu thị đó có thể đưa ra giá chính xác của một mặt hàng ngay sau khi đặt nó vào giỏ hàng của họ. Kiểm định ở mức 7% giả thuyết cho rằng ít nhất một nửa số người mua sắm có thể đưa ra mức giá chính xác.

Ta cần một số đo để đo mức độ chặt chẽ trong quan hệ tuyến tính giữa hai biến ngẫu nhiên.

Định nghĩa 8.1 Hệ số tương quan mẫu của hai biến ngẫu nhiên X, Y

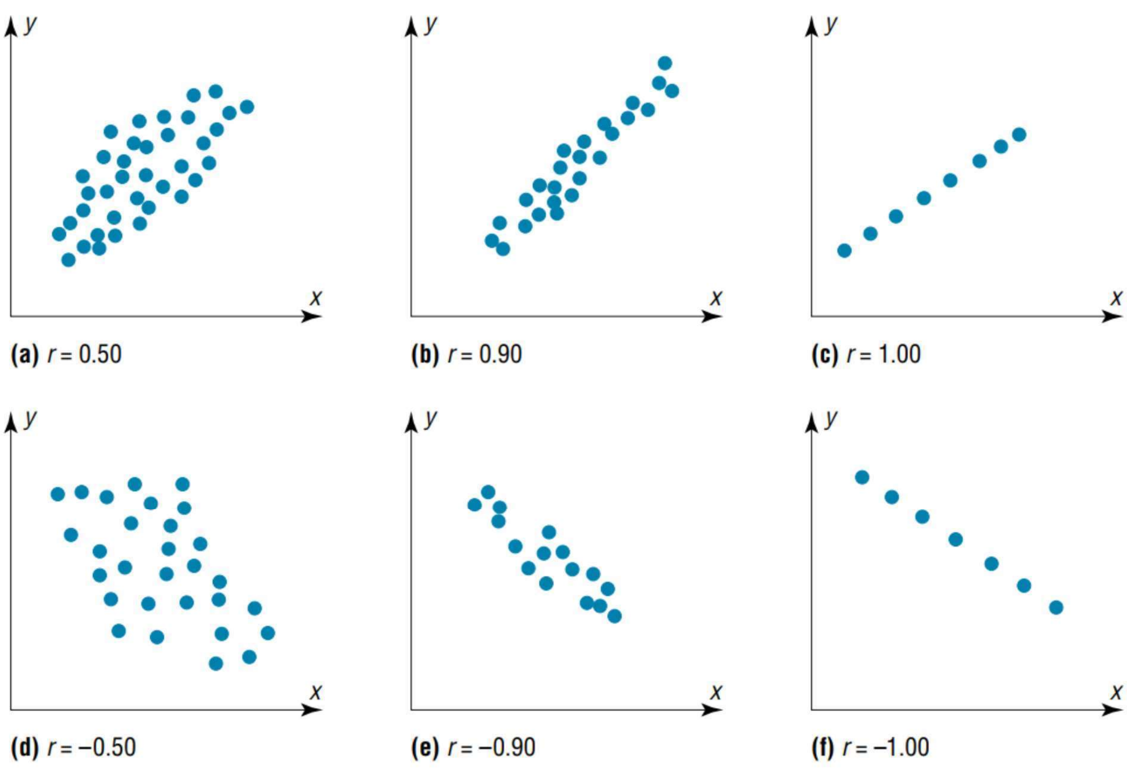
$$r = \frac{\overline{xy} - \overline{x}.\overline{y}}{\sqrt{(\overline{x^2} - \overline{x}^2)(\overline{y^2} - \overline{y}^2)}}$$

Hay

$$r = \frac{n(\sum_{i=1}^n x_i y_i) - (\sum_{i=1}^n x_i).(\sum_{i=1}^n y_i)}{\sqrt{(n(\sum_{i=1}^n x_i^2) - (\sum_{i=1}^n x_i)^2)(n(\sum_{i=1}^n y_i^2) - (\sum_{i=1}^n y_i)^2)}}$$

trong đó n là số cặp điểm dữ liệu thực nghiệm.

- Ta có $-1 \leq r \leq 1$.
- Nếu $0,8 \leq |r| \leq 1$ thì ta nói X, Y có tương quan tuyến tính mạnh.
- Nếu $|r| < 0,8$ thì ta nói X, Y có tương quan tuyến tính yếu.
- Nếu r gần bằng 1 thì ta nói có sự tương quan tuyến tính thuận giữa X và Y.
- Nếu r gần bằng -1 thì ta nói có sự tương quan tuyến tính nghịch giữa X và Y.



Ví dụ 8.2 Điểm số môn Xác suất thống kê và số buổi vắng của 7 sinh viên được cho bên dưới

Số buổi vắng (X)	6	2	15	9	12	5	8
Điểm (Y)	8,2	8,6	4,3	7,4	5,8	9,0	7,8

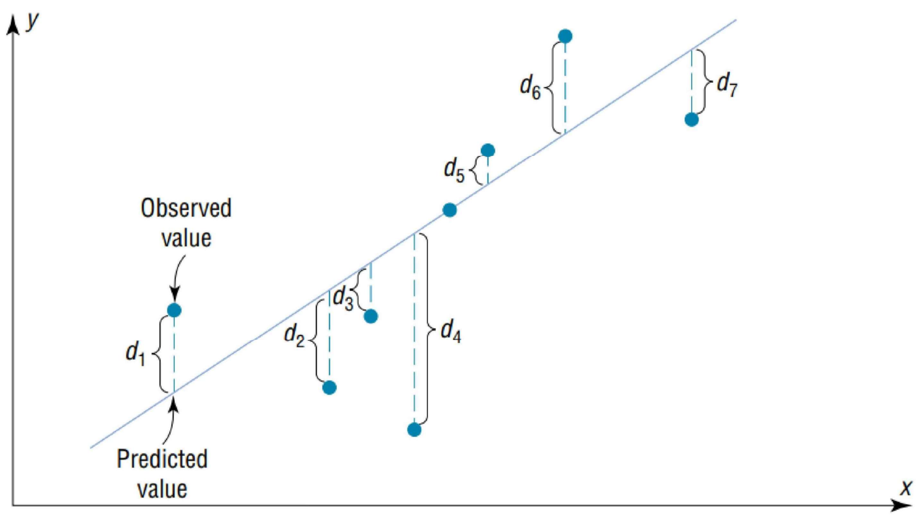
Tìm hệ số tương quan giữa số buổi nghỉ học và điểm môn Xác suất thống kê.
Ta có

$$\begin{aligned}
\overline{xy} &= \frac{6.8,2 + 2.8,6 + 15.4,3 + 9.7,4 + 12.5,8 + 5.9,0 + 8.7,8}{7} = 53,5 \\
\overline{x} &= \frac{6 + 2 + 15 + 9 + 12 + 5 + 8}{7} = 8,14 \\
\overline{y} &= \frac{8,2 + 8,6 + 4,3 + 7,4 + 5,8 + 9,0 + 7,8}{7} = 7,3 \\
\overline{x^2} &= \frac{6^2 + 2^2 + 15^2 + 9^2 + 12^2 + 5^2 + 8^2}{7} = 82,71 \\
\overline{y^2} &= \frac{8,2^2 + 8,6^2 + 4,3^2 + 7,4^2 + 5,8^2 + 9,0^2 + 7,8^2}{7} = 55,7
\end{aligned}$$

Do đó, hệ số tương quan là

$$r = \frac{53,5 - 8,14.7,3}{\sqrt{(82,71 - 8,14^2)(55,7 - 7,3^2)}} = \frac{-5,992}{6,296} = -0,9517.$$

Có một sự tương quan tuyến tính mạnh giữa số buổi vắng và số điểm. Nếu số buổi vắng càng nhiều thì số điểm càng thấp.



Khi đó

$$B = \frac{\overline{xy} - \overline{x}.\overline{y}}{\overline{x^2} - \overline{x}^2} \text{ và } A = \overline{y} - B\overline{x}.$$

Ví dụ 8.3 Điểm số môn Xác suất thống kê và số buổi vắng của 7 sinh viên được cho bên dưới

Số buổi vắng (X)	6	2	15	9	12	5	8
Điểm (Y)	8,2	8,6	4,3	7,4	5,8	9,0	7,8

Tìm phương trình đường thẳng hồi quy tuyến tính và dự đoán điểm của sinh viên chỉ vắng 1 buổi học.

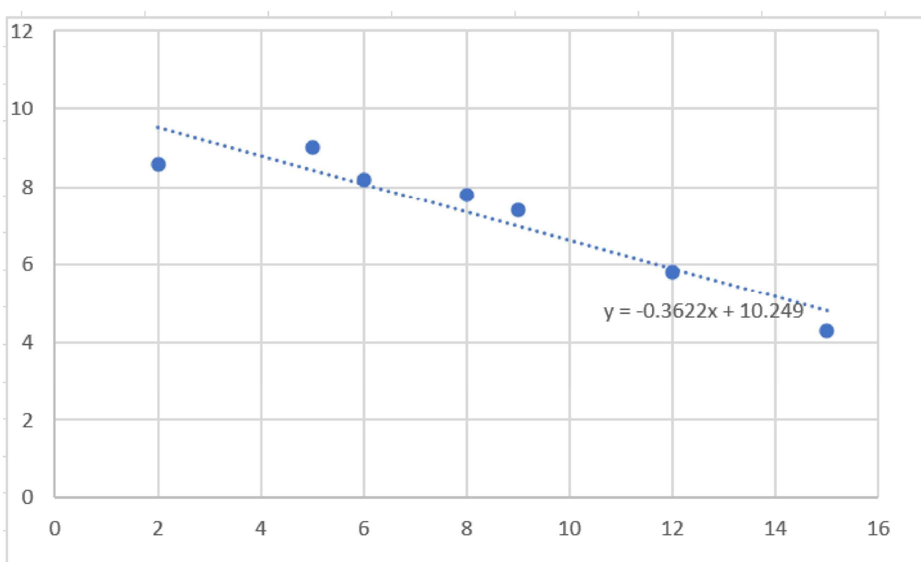
Giải.

- Phương trình hồi quy tuyến tính $Y = A + BX$.
- $B = \dfrac{\overline{xy} - \overline{x}.\overline{y}}{\overline{x^2} - \overline{x}^2} = \ldots\ldots\ldots$
- $A = \overline{y} - B\overline{x} = \ldots\ldots.$
- Phương trình đường thẳng hồi quy tuyến tính cần tìm là ....
- Khi $X = 1$ thì $Y = \ldots$

Giải. Dùng máy tính CASIO fx-570VN-PLUS

- $\boxed{SHIFT} \rightarrow \boxed{MODE} \rightarrow \triangledown \rightarrow$ chọn STAT (trên màn hình - phím 4)
- Màn hình xuất hiện **Frequency**, chọn **OFF**
- $\boxed{SHIFT} \rightarrow \boxed{MODE} \rightarrow$ **Data** (phím $\boxed{2}$)
- Nhập dữ liệu cột X : $\boxed{6}$ $\boxed{=}$ $\boxed{2}$ $\boxed{=}$...
- Nhập dữ liệu cột Y : $\boxed{8.2}$ $\boxed{=}$ $\boxed{8.6}$ $\boxed{=}$...
- $\boxed{ON}$
- $\boxed{SHIFT} \rightarrow \boxed{1} \rightarrow$ **Reg** (phím $\boxed{5}$)
- Chọn **A** (phím $\boxed{1}$) $\boxed{=}$
- $\boxed{ON}$
- $\boxed{SHIFT} \rightarrow \boxed{1} \rightarrow$ **Reg** (phím $\boxed{5}$)
- Chọn **B** (phím $\boxed{2}$) $\boxed{=}$

Khi đó $A = 10,2493$ và $B = -0,3722$. Đường thẳng hồi quy tuyến tính là $Y = 10,2493 - 0,3622X$.



Nếu $X = 1$ thì $Y = 9,8871$. Do đó nếu sinh viên vắng một buổi học thì điểm số của sinh viên có thể đạt được là 9,8871 điểm.

Dùng Microsoft Excel để tìm đường thẳng hồi quy

- Tạo bảng dữ liệu trong Microsoft Excel
- Tạo biểu đồ phân tán: Chọn bảng dữ liệu → **Insert** → **Charts** → **All Charts** → **X Y (Scatter)** → **OK**
- Tạo đường thẳng hồi quy: Nhấp vào ⊞ bên góc phải của Chart vừa hiện ra → **Chart Elements**, chọn **Trendline**
- Hiện phương trình đường thẳng hồi quy: Bên cạnh **Trendline** → ▶ **More Options**
- Trong bảng **Format Trendline**, chọn , kéo xuống bên dưới và chọn **Display Equation on chart**.

Một vài lưu ý

- Đường thẳng hồi quy tuyến tính theo phương pháp bình phương tối tiểu luôn đi qua điểm $(\overline{x}, \overline{y})$
- Khi tính toán cần xác định rõ biến độc lập và biến phụ thuộc
  - Phương tình hồi quy tuyến tính của Y theo X

$$Y = A + BX$$

  - Phương tình hồi quy tuyến tính của X theo Y

$$X = A + BY$$

Ví dụ 8.4 Bảng khảo sát doanh thu bán hàng online Y và chi phí quảng cáo online X (trong 15 phút) của 7 cửa hàng được cho như sau: Đơn vị tính là USD

Doanh số bán hàng	368	340	665	954	331	556	376
Chi phí quảng cáo	1,7	1,5	2,8	5	1,3	2,2	1,3

a. Tính hệ số tương quan và nhận xét về tính tuyến tính của X và Y (mạnh hay yếu).
b. Viết phương trình hồi quy tuyến tính của Y theo X. Dự đoán doanh số bán hàng khi chi phí quảng cáo online trong 15 phút là 4 USD.

Giải. ....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

BÀI TẬP

Bài 8.1 Lợi nhuận của 7 công ty cho thuê xe Y (tỉ USD) trong 1 năm và số lượng xe cho thuê X (nghìn chiếc) được cho như sau

Công ty	Số xe (X)	Lợi nhuận (Y) (tỉ USD)
A	630	7
B	290	3,9
C	208	2,1
D	191	2,8
E	134	1,4
F	85	1,5

a. Tính hệ số tương quan giữa số xe cho thuê và lợi nhuận hàng năm.
b. Viết phương trình hồi quy tuyến tính của Y theo X. Dự đoán lợi nhuận trong một năm của một công ty có 200 000 xe cho thuê.
Bài 8.2 PCWorld đã đánh giá bốn đặc điểm thành phần cho 10 máy tính xách tay siêu di động: tính năng, hiệu suất, thiết kế và giá cả. Mỗi đặc điểm được đánh giá bằng thang điểm 0-100. Xếp hạng tổng thể, được gọi là PCW World Rating, sau đó được phát triển cho từng máy tính xách tay. Bảng sau đây cho thấy xếp hạng tính năng và PCW World Rating cho 10 máy tính xách tay (trang web PC World, ngày 5 tháng 2 năm 2009).

Model	Hạng tính năng (X)	PCW World Rating (Y)
Thinkpad X200	87	83
VGN-Z598U	85	82
U6V	80	81
Elitebook 2530P	75	78
X360	80	78
Thinkpad X300	76	78
Ideapad U110	81	77
Micro Express JFT2500	73	75
Toughbook W7	79	73
HP Voodoo Envy133	68	72

a. Tính hệ số tương quan và nhận xét về tính tuyến tính của X và Y (mạnh hay yếu).
b. Viết phương trình hồi quy tuyến tính của Y theo X. Dự đoán PCW World Rating của một laptop mới mà nó có hạng tính năng là 70.

Bài 8.3 Một nghiên cứu đã được thực hiện về lượng đường được chuyển đổi trong một quy trình nhất định ở các nhiệt độ khác nhau. Dữ liệu được mã hóa và ghi lại như sau:

Nhiệt độ (X)	Đường được chuyển đổi (Y)		Nhiệt độ (X)	Đường được chuyển đổi (Y)
1,0	8,1		1,6	8,6
1,1	7.8		1,7	10,2
1,2	8,5		1,8	9.3
1,3	9,8		1,9	9,2
1,4	9,5		2,0	10,5
1,5	8,9			

a. Tìm đường thẳng hồi quy tuyến tính.
b. Ước tính lượng đường chuyển đổi được tạo ra khi nhiệt độ được mã hóa là 1,75.

Bài 8.4 Thời gian sử dụng liên tục của 8 loại điện thoại Y (giờ) và số mAh X (nghìn chiếc) ghi trên pin của điện thoại được khảo sát như sau

Điện thoai	Số mAh (X)	Thời gian sử dụng (Y) (giờ)
A	2800	3,8
B	3000	3,9
C	3700	4,2
D	4000	3,8
E	4300	4,1
F	5000	5
G	5000	4.8
H	6000	4,9

a. Tính hệ số tương quan giữa số mAh trên pin và thời gian sử dụng.
b. Viết phương trình hồi quy tuyến tính của Y theo X. Dự đoán thời gian sử dụng của một loại pin điện thoại có 6550 mAh.