Bài 1: **KHÁI NỆM CƠ BẢN VỀ XÁC SUẤT**

I. PHÉP THỬ, KHÔNG GIAN MẪU.

Phép thử ngẫu nhiên là một quá trình sinh ra một tập dữ liệu mà không thể dự báo trước được sự xảy ra của mỗi dữ liệu, mặc dù có thể biết toàn bộ tập dữ liệu đó.

Tập hợp tất cả kết quả có thể của một phép thử được gọi là **không gian mẫu,** ký hiệu S hoặc Ω.

Mỗi kết quả trong không gian mẫu gọi là một **phần tử** của không gian mẫu, hoặc là một **điểm mẫu**.

Cách mô tả không gian mẫu bằng

+ **liệt kê** các phần tử

+ **Biểu diễn** bằng một **mệnh đề, quy tắc.**

Có thể dùng **sơ đồ cây** liệt kê những phần tử của không gian mẫu để có nhiều thông tin hơn.

Ví dụ

a) Tìm không gian mẫu Ω của phép thử tung một đồng xu.

b)Tìm không gian mẫu của phép thử là tung một xúc xắc.

c)Tìm không gian mẫu của một phép thử là tung một đồng xu. Nếu xuất hiện mặt sấp thì tung nó lần thứ hai, còn xuất hiện mặt ngửa, thì tung một con xúc xắc lên.

d)Một lô hàng có 6 sản phẩm, trong đó có 2 phế phẩm. Lấy ngẫu nhiên lần lượt 3 sản phẩm. Tìm không gian mẫu của phép thử.

II. BIẾN CỐ

Biến cố là một tập con chứa điểm mẫu của không gian mẫu.

Dùng $A, B, C, A_1, A_2, \ldots$ để ký hiệu cho biến cố.

Một tập con của S không chứa bất kỳ một phần tử nào gọi là **biến cố không thể,** ký hiệu ∅.

Tập hợp là toàn bộ không gian mẫu S gọi là **biến cố chắc chắc**.

III. PHÉP TOÁN VÀ QUAN HỆ CÁC BIẾN CỐ

Giả sử A, B là hai biến cố trong một phép thử, hay là tập con của không gian mẫu S.

1. **Phần bù** của một biến cố A trong S là tập con gồm tất cả những phần tử của S mà không nằm trong A.

   Ký hiệu là $\overline{A}$ hoặc A'.

2. **Giao của hai biến cố *A* và *B***, ký hiệu $A \cap B$ hoặc AB, là biến cố chứa tất cả những phần tử chung của A và B.

3. **Hợp của hai biến cố *A* và *B***, ký hiệu $A \cup B$ hoặc $A + B$, là biến cố chứa tất cả những phần tử mà thuộc A hoặc thuộc B hoặc thuộc cả hai. (hay **ít nhất** trong A **hoặc** B).

4. **Định lý De Morgan:** $\overline{AB} = \overline{A} + \overline{B}$, $\overline{A+B} = \overline{A}\ \overline{B}$

Ví dụ

a. Cho $A = \{2,4,6\}$ và $B = \{4,5,6\}$ là các tập con trong không gian mẫu $S = \{1,2,3,4,5,6\}$. Tìm $\overline{A}, \overline{B}$, $A + B$, AB, $\overline{A}B, \overline{A} + B, \overline{A}\,\overline{B}, \overline{A} + \overline{B}$.

b. Hai xạ thủ cùng bắn vào một mục tiêu. Gọi biến cố A = "Xạ thủ 1 bắn trúng", B = "Xạ thủ 2 bắn trúng". Hãy biểu diễn các biến cố sau qua biến cố A và B.

   M = "mục tiêu bắn trúng",

   N = "mục tiêu bắn trượt",

   H = "chỉ một xạ thủ bắn trúng".

c. Lấy 4 sản phẩm trong kho có nhiều phế phẩm. Hãy biểu diễn các biến cố A = " có nhiều nhất 1 phế phẩm", B = " có ít nhất 1 phế phẩm".

5. **Quan hệ xung khắc:** Hai biến cố A và B gọi là xung khắc nếu $AB = \varnothing$.

IV. TÍNH CHẤT CÁC PHÉP TOÁN BIẾN CỐ.

1. **Hằng đẳng thức:**

$$A \cup A = A, \quad A \cup S = S, \quad A \cup \varnothing = A,$$

$$AA = A, \quad AS = A, \quad A\varnothing = \varnothing,$$

$$\overline{\overline{A}} = A, \quad A \cup \overline{A} = S, \quad A\overline{A} = \varnothing$$

2. **Tính chất :**

   Giao hoán: $A \cup B = B \cup A, \quad AB = BA$, Kết hợp: $(A\cup B)\cup C = A\cup(B\cup C)$, $(AB)C = A(BC)$

Phân phối: $A(B\cup C) = (AB)\cup(AC)$, $A\cup(BC) = (A\cup B)(A\cup C)$

V. ĐẾM CÁC ĐIỂM MẪU

1. Quy tắc nhân:

Nếu một công việc chia ra k **giai đoạn**. Giai đoạn 1 có n_1 cách thực hiện. Giai đoạn 2 có n_2 cách ...

Giai đoạn k có n_k cách thực hiện $\Rightarrow N = n_1 n_2 \ldots n_k$ cách thực hiện xong công việc.

2. Qui tắc cộng:

Nếu một công việc được chia ra k **trường hợp** để thực hiện. Trường hợp 1 có n_1 cách,trường hợp 2 có n_2 cách ,...,trường hợp k có n_k cách thực hiện, không có bất kỳ một cách thực hiện nào ở trường hợp này lại trùng với cách thực hiện ở trường hợp khác thì có $N = n_1 + n_2 + \ldots + n_k$ cách thực hiện xong công việc.

Ví dụ

a. Từ A đến B có 4 con đường, từ B đến C có 3 con đường. Hỏi có bao nhiêu cách đi từ A đến C đi qua B.

b. Có bao nhiêu số có 3 chữ số có tổng bằng 4.

3. Chỉnh hợp.

Cho tập hợp n phần tử. Một nhóm *có **tính thứ tự*** gồm k $(0 \le k \le n)$ *phần tử khác nhau* lấy từ n phần tử đã cho gọi là một chỉnh hợp chập k của n.

Số các chỉnh hợp chập k của n phần tử là $A_n^k = \dfrac{n!}{(n-k)!}$

4. Hoán vị

Một nhóm *có **tính thứ tự*** gồm *n phần tử khác nhau* lấy từ n phần tử đã cho gọi là một hoán vị.

Số các hoán vị của n, kí hiệu là P_n, cũng là số các chỉnh hợp chập n của n.

$$\boxed{P_n = A_n^n = n!}$$

Số những hoán vị của n phần tử phân biệt được sắp xếp theo một vòng tròn là $(n-1)!$

5. Tổ hợp.

Cho tập hợp n phần tử. Nhóm ***không tính thứ tự*** gồm k $(0 \le k \le n)$ *phần tử khác nhau* lấy từ n phần tử đã cho gọi là một tổ hợp chập k của n.

Số các tổ hợp chập k của n là $C_n^k = \dfrac{A_n^k}{k!} = \dfrac{n!}{k!(n-k)!}$

6. Phân hoạch

Phân hoạch một tập gồm n phần tử thành r tập con được gọi là các **ngăn**.

Số cách phân hoạch một tập gồm n phần tử thành r ngăn mà có n_1 phần tử trong ngăn thứ nhất, n_2 phần tử trong ngăn thứ hai,..., là

$$\boxed{C_n^{n_1, n_2, \ldots, n_r} = \frac{n!}{n_1! n_2! \cdots n_r!} \quad \text{trong đó} \quad n_1 + n_2 + \ldots + n_r = n}$$

Ví dụ

a. Chọn trong 5 người làm 2 việc A, B. Hỏi có bao nhiêu cách xếp mỗi người một việc.

b. Có bao nhiêu cách sắp 4 người vào một bàn ăn có 4 chỗ ngồi?

c. Một giải bóng đá có 12 đội tham gia, thi đấu đôi một với nhau. Hỏi có bao nhiêu trận đấu của giải.

d. Có bao nhiêu cách sắp khác nhau để tạo thành một xâu đèn của cây thông Noel có 3 bóng đèn đỏ, 4 bóng đèn vàng, và 2 bóng đèn xanh vào 9 đui đèn?

Chú ý:

Có 4 cách lấy k phần tử từ tập hợp n phần tử là:

- Cách lấy theo nghĩa tổ hợp.
- Cách lấy theo nghĩa chỉnh hợp.
- Cách lấy ra từng phần tử một không hoàn lại
- Cách lấy ra từng phần tử một có hoàn lại.

VI. XÁC SUẤT CỦA MỘT BIẾN CỐ.

Xác suất của một biến cố là số đặc trưng cho khả năng xuất hiện của biến cố A, kí hiệu là P(A).

$$\text{P}(A) = \frac{\textbf{Số trường hợp có lợi cho A}}{\textbf{Tổng số trường hợp}} = \frac{n_A}{N}$$

Tính chất: $0 \le P(A) \le 1$; $P(\varnothing) = 0$; $P(S) = 1$

Ví dụ

a. Lấy ngẫu nhiên 5 cây bài trong bộ tú lơ khơ 52 quân. Tìm xác suất để có 2 cây Át, 3 cây từ 2 đến 6.

b. Lấy ngẫu nhiên 3 viên bi trong một hộp có 6 bi đỏ, 4 bi xanh. Tính xác suất của biến cố:

A = "Không có quá 1 viên bi xanh",

B = "Có ít nhất 1 xanh, 1 đỏ".

c. Một con xúc sắc được đổ chì sao cho khả năng xuất hiện chấm chẵn gấp đôi khả năng xuất hiện chấm lẻ. Tính xác suất để số chấm nhỏ hơn 4 xuất hiện trong 1 lần tung xúc sắc.

BÀI 2: CÁC ĐỊNH LÝ VỀ XÁC SUẤT

I. QUY TẮC CỘNG.

Định lý Nếu A và B là hai biến cố tùy ý thì $P(A+B) = P(A) + P(B) - P(AB)$

Hệ quả 1 Nếu A và B là 2 biến cố xung khắc thì $P(A+B) = P(A) + P(B)$

Hệ qủa 2 Nếu $A_1, ..., A_n$ xung khắc với nhau thì

$$P(A_1 + A_2 + ... + A_n) = P(A_1) + P(A_2) + ... + P(A_n)$$

Hệ qủa 3 Nếu A và $\overline{A}$ là hai biến cố phần bù của nhau thì $P(A) + P(\overline{A}) = 1$.

Ví dụ Trong kỳ thi 2 môn Toán và tiếng Anh, xác suất để Paula thi đỗ môn toán là 0,7 và xác suất thi đỗ môn tiếng Anh là 0,6, xác suất để thi đỗ cả 2 môn là 0,4. Tính xác suất để Paula

(a) thi đỗ ít nhất một môn.

(b) không đỗ môn nào.

(c) thi trượt ít nhất một môn.

(d) thi đỗ đúng một môn.

Chú ý: $S = \{\text{ít nhất một}\} \cup \{\text{không có}\}$

II. XÁC SUẤT CÓ ĐIỀU KIỆN.

Định nghĩa Xác suất có điều kiện của biến cố B với điều kiện A, ký hiệu $P(B|A)$, được xác định như sau:

$$\boxed{P(B|A) = \frac{P(AB)}{P(A)} \quad \text{nếu } P(A) > 0}$$

Ý nghĩa: $P(B|A)$ là số đo (tỷ lệ) khả năng xảy ra của biến cố B trong điều kiện biến cố A đã xảy ra.

$P(B)$ là số đo khả năng xảy ra của B trong S.

Hai biến cố A và B được gọi là **độc lập với nhau** khi và chỉ khi

$$P(B|A) = P(B) \text{ hoặc } P(A|B) = P(A).$$

Trường hợp ngược lại ta nói A và B **phụ thuộc nhau**.

Ví dụ Xác suất một chuyến bay khởi hành đúng giờ là $P(A) = 0,83$, xác suất đến đúng giờ là $P(B) = 0,82$, xác suất khởi hành và đến đều đúng giờ là $P(AB) = 0,78$. Tính xác suất để một chuyến bay:

(a) đến đúng giờ, biết nó đã khởi hành đúng giờ.

(b) đến đúng giờ, biết nó khởi hành không đúng giờ.

(c) khởi hành đúng giờ biết nó đã đến đúng giờ.

Chú ý: $P(\overline{A}B) = P(B) - P(AB)$

Ví dụ Trong một thí nghiệm nghiên cứu mối liên hệ giữa bệnh cao huyết áp với thói quen hút thuốc lá, số liệu thu thập được từ 180 người như sau:

	Không hút	Hút thuốc	Nghiện thuốc
Huyết áp cao	25	38	27
Bình thường	48	24	18

Chọn ngẫu nhiên một trong những người này. Tìm xác suất

(a) để người đó bị cao huyết áp, biết người đó nghiện thuốc.

(b) để người đó nghiện thuốc, biết người đó huyết áp bình thường.

III. QUY TẮC NHÂN

1. Định lý Nếu trong một phép thử, các biến cố A và B có thể cùng xảy ra thì

$$\boxed{P(AB) = P(A).P(B \mid A) = P(B).P(A \mid B)}$$

Hệ quả: Hai biến cố A và B là độc lập với nhau khi và chỉ khi

$$\boxed{P(AB) = P(A).P(B).}$$

2. Định lý Nếu trong một phép thử, các biến cố $A_1, A_2, ..., A_k$ có thể xảy ra thì

$$\boxed{P(A_1A_2...A_k) = P(A_1).P(A_2 \mid A_1)...P(A_k \mid A_1A_2...A_{k-1})}$$

Hệ quả: Nếu $A_1, A_2, ..., A_k$ độc lập thì $P(A_1A_2...A_k) = P(A_1).P(A_2)...P(A_k)$

Ví dụ Trong một dây chuyền sản xuất có 2 máy kiểm tra liên tiếp các sản phẩm. Qua máy kiểm tra thứ nhất có 2% số sản phẩm bị loại. Qua máy kiểm tra thứ hai có 0,5% sản phẩm bị loại. Tìm xác suất để một sản phẩm trong dây chuyền đó không bị loại.

Ví dụ Trong lô hàng có 100 sản phẩm, trong đó có 80 sản phẩm tốt. Lấy ngẫu nhiên 3 sản phẩm lần lượt từng chiếc không hoàn lại. Tính xác suất để 3 sản phẩm tốt.

Ví dụ Một hộp chứa 3 bi trắng, 7 bi đỏ và 10 bi xanh. Hộp 2 chứa 5 bi trắng, 2 bi đỏ và 3 bi xanh. Ta lấy ngẫu nhiên từ mỗi hộp 1viên bi

a. Tìm xác suất để 2 viên lấy ra cùng màu.

b. Khi hai viên cùng màu, tìm xs để hai viên bi đó màu xanh.

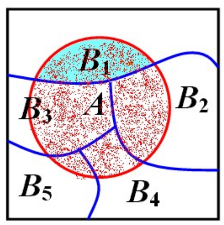
IV. CÔNG THỨC XÁC SUẤT ĐẦY ĐỦ VÀ BAYES

Định lý: Nếu các biến cố $B_1, B_2, ..., B_k$ là một phân hoạch của không gian mẫu S, thì với biến cố A bất kì của S, ta có:

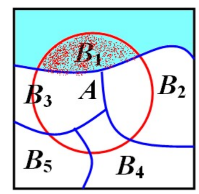
Công thức xác suất đầy đủ $P(A)=\sum_{i=1}^{k}P(B_iA)\boxed{=\sum_{i=1}^{k}P(B_i)P(A\mid B_i)}$

Công thức Bayes $P(B_i\mid A)=\dfrac{P(B_i).P(A\mid B_i)}{P(A)}\boxed{,\ \text{với}\ \ i=1,2,\ldots,k}$

Chú ý: Phân biệt $P(B_1\mid A)$ và $P(A\mid B_1)$



Tỉ lệ B_1 có mặt trong A.



Tỉ lệ A có mặt trong B_1

Ví dụ Trong một dây chuyền có ba máy 1,2,3 sản xuất tương ứng 30%, 45% và 25% sản phẩm. Qua kiểm tra cho biết tỷ lệ phế phẩm do mỗi máy tương ứng là 3%, 4% và 2%. Lấy ngẫu nhiên một sản phẩm.

(a) Tính xác suất để nó là phế phẩm.

(b) Biết lấy ra là phế phẩm, tìm xác suất để nó do máy 1 sản xuất.

Ví dụ Trong túi 1 có 4 bóng trắng, 3 bóng đen. Túi 2 có 3 bóng trắng, 5 bóng đen. Lấy ngẫu nhiên một quả bóng từ túi 1 bỏ sang túi 2. Tính xác suất để lấy ra từ túi 2 được một quả bóng màu đen.

BÀI 3: BIẾN NGẪU NHIÊN MỘT CHIỀU

I. BIẾN NGẪU NHIÊN.

1. **Biến ngẫu nhiên** là một hàm số đặt tương ứng mỗi phần tử trong không gian mẫu với một số thực. Tên biến ngẫu nhiên viết chữ in hoa, ví dụ *X*.

Các giá trị của nó viết chữ thường, ví dụ *x*.

Ví dụ

Ba quả bóng được lấy lần lượt theo cách không hoàn lại từ một bình chứa 4 quả bóng đỏ và 2 quả bóng trắng. Hãy mô tả không gian mẫu qua

a) Biến cố Đ = “ lấy bóng đỏ”, T = “ lấy bóng trắng.

b) Biến ngẫu nhiên X là số bóng đỏ.

2. **Biến ngẫu nhiên được gọi là rời rạc** nếu tập các kết cục có thể xảy ra của nó là đếm được.

Biến ngẫu nhiên được gọi là liên tục nếu tập giá trị của nó là một khoảng thực nào đó.

II. PHÂN PHỐI XÁC SUẤT RỜI RẠC.

1. **Hàm phân phối xác suất** của biến ngẫu nhiên rời rạc X :

là hàm số tương ứng với mỗi điểm mẫu của không gian mẫu S với một số thực $f(x)$ thỏa mãn các điều kiện sau:

1) $f(x) \geq 0$

2) $\sum_{x} f(x) = 1$

3) $P(X = x) = f(x)$

Ví dụ Một kiện hàng có 8 máy vi tính, trong đó 3 máy bị lỗi. Một trường học mua ngẫu nhiên 2 máy. Gọi biến ngẫu nhiên X là số máy mua bị lỗi, tìm hàm phân phối xác suất $f(x)$ của X.

Ví dụ Một người bắn ba viên đạn độc lập nhau vào một mục tiêu với xác suất bắn một viên trúng mục tiêu là 0,6.

(a) Gọi X là số viên đạn bắn trúng mục tiêu. Tìm hàm phân phối xác suất $f(x)$ của X.

(b) Gọi Y là số viên đạn đã bắn cho đến khi trúng mục tiêu hoặc hết đạn thì dừng lại. Tìm phân phối xác suất $g(Y)$ của Y.

2. **Hàm phân phối tích lũy xác suất** $\mathrm{F(x)}$ của biến ngẫu nhiên rời rạc X với phân phối xác suất $\mathrm{f(x)}$ là:

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t), \text{ với } -\infty < x < +\infty .$$

Hệ quả: $P(a < X \le b) = F(b) - F(a)$.

$$P(a < X < b) = F(b) - F(a) - P(X = b).$$

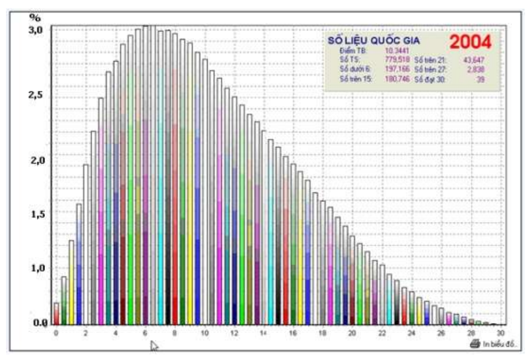
Ý nghĩa Từ tổng các giá trị xác suất trong miền $(-\infty, x]$ sẽ tính được xác suất trong miền $[a,b]$ bất kỳ trong khoảng đó.

Ví dụ Tìm hàm phân phối tích lũy $F(x)$ của biến ngẫu nhiên X và tính $P(0,5 < X \le 2,5)$

x	0	1	2	3
$f(x)$	0,4	0,2	0,1	0,3

3. Biểu đồ xác suất.

Vẽ các điểm $(x, f(x))$ và nối các điểm này đến trục Ox bởi một đường nét đứt (hoặc liền nét), ta được một biểu đồ hình cây.



Ta cũng có thể vẽ các hình chữ nhật có bề rộng đáy bằng nhau và mỗi giá trị x được đặt chính giữa đáy, còn chiều cao của chúng bằng xác suất tương ứng được cho bởi $f(x)$.được gọi là một **biểu đồ xác suất**.

III. PHÂN PHỐI XÁC SUẤT LIÊN TỤC.

1. **Hàm mật độ xác suất** $f(x)$ của biến ngẫu nhiên liên tục X, xác định trên tập các số thực $\mathbb{R}$, nếu:

    1) $f(x) \ge 0, \quad \forall x \in \mathbb{R}$.

    2) $\int_{-\infty}^{+\infty} f(x)dx = 1$.

3) $P(a < X < b) = \int_a^b f(x)dx$.

Ví dụ Giả sử sai số của nhiệt độ phản ứng (đơn vị ^{0}C) trong một thí nghiệm là biến ngẫu nhiên liên tục X có hàm mật độ xác suất $f(x) = \begin{cases} kx^2, & x \in (-1,2) \\ 0 \quad , & x \notin (-1,2) \end{cases}$

(a) Tìm hằng số k.

(b) Tìm $P(0 < X < 1)$.

2. **Hàm phân phối tích lũy** $F(x)$ của biến ngẫu nhiên liên tục X với hàm mật độ $f(x)$ là: $F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)dt$, với $-\infty < x < \infty$

Hệ quả: $P(a < x < b) = F(b) - F(a)$

$$f(x) = \frac{dF(x)}{dx} \text{ (nếu đạo hàm tồn tại).}$$

Ví dụ Thời hạn sử dụng (đơn vị: ngày) của một loại thuốc nào đó là biến ngẫu nhiên X có hàm mật độ sau:

$$f(x) = \begin{cases} \dfrac{20000}{(x+100)^3}, & x > 0 \\ 0 \quad , & x \le 0 \end{cases}$$

Tìm xác suất để một chai thuốc nói trên sẽ có thời hạn sử dụng

(a) ít nhất là 200 ngày.

(b) từ 80 đến 120 ngày.

BÀI 4: BIẾN NGẪU NHIÊN HAI CHIỀU

I. PHÂN PHỐI XÁC SUẤT ĐỒNG THỜI.

Định nghĩa:

Hàm $f(x,y)$ là **phân phối xác suất đồng thời** của các biến ngẫu nhiên rời rạc X và Y nếu:

1) $f(x,y) \ge 0, \quad \forall (x,y)$

2) $\sum_x \sum_y f(x,y) = 1$.

3) $P(X = x, Y = y) = f(x,y)$.

Công thức tính xác suất trên miền A trong mặt phẳng $0xy$: $P[(X,Y) \in A] = \sum\sum_A f(x,y)$.

Ví dụ Hai chiếc bút được chọn ngẫu nhiên từ 1 hộp gồm 3 bút xanh, 2 bút đỏ, 3 bút vàng. Gọi X là số bút xanh, Y là số bút đỏ được chọn, tìm

(a) Phân phối xác suất đồng thời $f(x,y)$.

(b) $P[(X,Y) \in A]$ với $A = \{(x,y) | x + y \le 1\}$.

Ví dụ Giả sử X là số lần xuất hiện mặt ngửa và Y là số lần xuất hiện mặt ngửa trừ đi số lần xuất hiện mặt sấp khi tung 3 đồng xu. Tìm phân phối xác suất đồng thời của X và Y.

II. HÀM MẬT ĐỘ ĐỒNG THỜI.

Định nghĩa: Hàm $f(x,y)$ được gọi là **hàm mật độ đồng thời** của các biến ngẫu nhiên liên tục X và Y nếu:

1) $f(x,y) \ge 0, \quad \forall (x,y)$.

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

3) $P[(X,Y) \in A] = \iint_A f(x,y)dxdy$

Ví dụ Cho biến ngẫu nhiên X, Y có hàm mật độ đồng thời là:

$$f(x,y) = \begin{cases} \frac{2}{5}(2x+3y), (x,y) \in D \\ 0 \qquad\qquad , (x,y) \notin D \end{cases}$$

Với $D: \{0 \le x \le 1, 0 \le y \le 1\}$.

Tìm $P[(X,Y) \in A]$ với $A = \left\{ (x,y) \middle| 0 < x < \frac{1}{2}; \frac{1}{4} < y < \frac{1}{2} \right\}$.

Ví dụ Cho biến ngẫu nhiên X, Y có hàm mật độ đồng thời là:

$$f(x,y)=\begin{cases} kx, & (x,y)\in D \\ 0, & (x,y)\notin D \end{cases}$$. Với $D=\{0\le x\le 1, 0\le y\le 1-x\}$. Tìm *k*.

III. PHÂN PHỐI BIÊN DUYÊN.

Định nghĩa:

Phân phối biên duyên của X và Y, với biến ngẫu nhiên rời rạc là

$$g(x)=\sum_y f(x,y) \text{ và } h(y)=\sum_x f(x,y)$$

với biến ngẫu nhiên liên tục là $g(x)=\int_{-\infty}^{\infty} f(x,y)dy$ và $h(y)=\int_{-\infty}^{\infty} f(x,y)dx$

Ví dụ Hai chiếc bút được chọn ngẫu nhiên từ 1 hộp gồm 3 bút xanh, 2 bút đỏ, 3 bút vàng. Gọi X là số bút xanh, Y là số bút đỏ được chọn. Tìm phân phối biên duyên của X và Y.

Ví dụ Giả sử hàm mật độ đồng thời của X,Y là:

$$f(x,y)=\begin{cases} \frac{2}{5}(2x+3y), & (x,y)\in D \\ 0 & ,(x,y)\notin D \end{cases}$$

Với $D:\{0\le x\le 1, 0\le y\le 1\}$. Tìm $g(x), h(y)$.

Ví dụ Cho biến ngẫu nhiên X,Y, với X là sự thay đổi nhiệt độ, Y là tỷ lệ thay đổi quang phổ mà một nguyên tử tạo ra, có hàm mật độ đồng thời là:

$$f(x,y)=\begin{cases} 10xy^2, & (x,y)\in D \\ 0 & ,(x,y)\notin D \end{cases}$$

với $D=\{(x,y)\,|\,0<x<y<1\}$. Tìm hàm mật độ biên duyên $g(x)$, $h(y)$.

IV. PHÂN PHỐI XÁC SUẤT CÓ ĐIỀU KIỆN VÀ ĐỘC LẬP THỐNG KÊ.

Nhắc lại: Xác suất có điều kiện của biến cố B với điều kiện biến cố A xảy ra:

$$P(B\,|\,A)=\frac{P(AB)}{P(A)}\ ,\quad P(A)>0$$

Định nghĩa: Giả sử X và Y là hai biến ngẫu nhiên rời rạc hoặc liên tục.

Phân phối xác suất có điều kiện của biến ngẫu nhiên Y với điều kiện $X=x$ là

$$f(y\,|\,x)=\frac{f(x,y)}{g(x)},\quad g(x)>0$$

Phân phối có điều kiện của biến ngẫu nhiên X với điều kiện $Y=y$ là

$$f(x\,|\,y)=\frac{f(x,y)}{h(y)},\quad h(y)>0$$

Hệ quả: Xác suất để biến ngẫu nhiên rời rạc X lấy giá trị trong khoảng (a,b) khi đã biết biến ngẫu nhiên rời rạc $Y=y$ là:

$$P(a<X<b\,|\,Y=y)=\sum_x f(x\,|\,y),$$

trong đó tổng được lấy trên tất cả các giá trị của X nằm giữa *a* và *b*.

Khi X và Y liên tục thì $P(a<X<b\,|\,Y=y)=\int_a^b f(x\,|\,y)dx$

Định nghĩa: Các biến ngẫu nhiên X và Y được gọi là **độc lập thống kê** khi và chỉ khi $f(x,y)=g(x).h(y)$, với $\forall(x,y)$ nằm trong miền giá trị của ...

Chú ý: Nếu $f(x\,|\,y)$ không phụ thuộc vào *y* thì $f(x\,|\,y)=g(x)$

Nếu $f(x,y)\neq f(x)g(y)$ thì biến ngẫu nhiên *X* và Y không độc lập thống kê.

Liên hệ với 2 biến cố *A*, *B* độc lập nhau $\Leftrightarrow P(A|B)=P(A)\Leftrightarrow P(AB)=P(A).P(B)$

Ví dụ Cho hàm mật độ đồng thời

$$f(x,y)=\begin{cases}\dfrac{x(1+3y^2)}{4}, & 0<x<2,\, 0<y<1\\ \quad 0 \quad, & (x,y)\notin(0,2)\times(0,1)\end{cases}$$

Tìm $g(x),h(y),f(x\,|\,y)$ và $P(\frac{1}{4}<X<\frac{1}{2}\,|\,Y=\frac{1}{3})$. Hỏi *X*, *Y* có độc lập tuyến tính không?

Ví dụ Cho phân phối xác suất đồng thời

Y \ X	0	1	2
– 1	0,12	0,18	0,3
3	0,08	0,12	0,2

(a) Tính $f(X=x\,|\,Y=3)$ với $X=0,1,2$

(b) Hỏi X, Y có độc lập thống kê không?

$5: CÁC SỐ ĐẶC TRƯNG CỦA BNN 1 CHIỀU

Ví dụ mở đầu: Kiểm tra ngẫu nhiên 100 hộp sữa có khối lượng như sau.

x (gam)	490	495	500	505	510
m (số hộp)	15	19	28	22	16

Tính khối lượng trung bình của các hộp sữa đó.

Tổng quát: $\overline{x}=\dfrac{\sum_{i=1}^{n}x_i m_i}{n}=\sum_{i=1}^{n}x_i\dfrac{m_i}{n}$. Với n đủ lớn thì tần suất $f(x_i)=\dfrac{m_i}{n}$ dần đến hàm phân phối xác suất $f(x)$ nên $\overline{x}\to\sum_x x.f(x)$.

I. TRUNG BÌNH (KỲ VỌNG - Expected).

1. Định nghĩa: Cho X là một biến ngẫu nhiên với phân phối xác suất là $f(x)$.

Trung bình (kỳ vọng) của X là $\mu=E(X)=\sum_x xf(x)$ nếu X là bnn rời rạc.

Và $\mu=E(X)=\int_{-\infty}^{+\infty}xf(x)dx$ nếu X là bnn liên tục.

2. Ý nghĩa: Trong thực tiễn thì các giá trị bnn X luôn biến đổi nhưng trung bình $\mu=E(X)$ thường ổn định nên $\mu=E(X)$ là giá trị đại diện cho X.

$E(X)$ còn gọi là tâm của phân phối xác suất.

3. Định lí: Cho X là bnn có phân phối xác suất $f(x)$ thì giá trị trung bình của biến ngẫu nhiên $g(X)$ là

$\boxed{\mu_{g(X)}=E(g(X))=\sum_x g(x).f(x)}$ nếu X rời rạc

$\mu_{g(X)}=E(g(X))=\int_{-\infty}^{+\infty}g(x).f(x)dx$ nếu X liên tục.

Ví dụ: Lấy ngẫu nhiên một mẫu gồm 3 sản phẩm từ lô hàng có 4 sản phẩm loại A và 3 sản phẩm loại *B*. Gọi X là số sản phẩm loại A trong 3 sản phẩm lấy ra.

(a) Tìm giá trị và phân phối xác suất của *X*.

(b) Tìm số sản phẩm trung bình loại *A* lấy ra.

(c) Biết giá tiền của 1 SP loại *A* là 3 triệu và 1 SP loại *B* là 0,2 triệu. Tính số tiền trung bình của mẫu đó.

Ví dụ Tuổi thọ X (giờ) của một thiết bị điện tử là một biến ngẫu nhiên có hàm mật độ xác suất là

$$f(x) = \begin{cases} \dfrac{20000}{x^3}, & x > 100 \\ 0 \quad , & x \le 100 \end{cases}.$$

Tính tuổi thọ trung bình của thiết bị điện tử.

II. PHƯƠNG SAI.

1. Định nghĩa :

Cho bnn X với ppxs $f(x)$ và trung bình là μ. Phương sai của X là trung bình của bình phương độ lệch giữa X và $\mu = E(X)$, tức là:

nếu X là rời rạc thì $\boxed{\sigma^2 = E(X-\mu)^2 = \sum_x (x-\mu)^2 f(x)}$

nếu X liên tục thì $\boxed{\sigma^2 = E(X-\mu)^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 . f(x)dx}$

σ gọi là **độ lệch chuẩn** của X.

Giá trị $x-\mu$ được gọi là **độ lệch của giá trị quan sát** so với giá trị trung bình.

2. Ý nghĩa: Phương sai biểu thị độ phân tán của X xung quanh giá trị trung bình μ.

3. Định lí: Phương sai của biến ngẫu nhiên X là

$$\boxed{\sigma^2 = E(X^2) - \mu^2 = \begin{cases} \sum_x x^2 f(x) - \mu^2, & X \text{ roi rac} \\ \int_{-\infty}^{+\infty} x^2 f(x)dx - \mu^2, & X \text{ lien tuc} \end{cases}}.$$

4. **Định lí:** Phương sai của biến ngẫu nhiên $g(X)$:

nếu X rời rạc là: $\boxed{\sigma^2_{g(X)} = E(g(X)-\mu_{g(X)})^2 = \sum_x (g(x)-\mu_{g(X)})^2 f(x).}$

nếu X liên tục là: $\boxed{\sigma^2_{g(X)} = E(g(X)-\mu_{g(X)})^2 = \int_{-\infty}^{+\infty} (g(x)-\mu_{g(X)})^2 f(x)dx}$

Công thức thường dùng để tính phương sai:

$\sigma^2[g(X)] = E[g^2(X)] - [E(g(X))]^2$

Đặc biệt: Nếu a và b là các hằng số, thì $\sigma^2(aX+b) = a^2\sigma^2(X)$.

Ví dụ Gọi X là bnn biểu thị số xe ô tô được sử dụng cho mục đích kinh doanh chính thức trong một ngày làm việc. Phân phối xác suất của X tại công ty A là

x	1	2	3
$f(x)$	0,3	0,4	0,3

và tại công ty B là

x	0	1	2	3	4
$f(x)$	0,2	0,1	0,3	0,3	0,1

Hãy so sánh trung bình và phương sai của biến ngẫu nhiên X tại công ty A và B.

Ví dụ: Lượng Pepsi (đv 1000 lít) bán được ở một địa phương là một bnn liên tục X với hàm mật độ xác suất như sau

$$f(x) = \begin{cases} 2(x-1), & x \in (1,2) \\ 0, & x \notin (1,2) \end{cases}$$

Tìm kỳ vọng và phương sai của $g(X) = 6X+2$.

§6: CÁC SỐ ĐẶC TRƯNG CỦA BNN 2 CHIỀU

1. Kỳ vọng, phương sai của hàm 2 bnn.

Cho X, Y là các bnn với ppxs đồng thời là $f(x,y)$ Kỳ vọng của biến ngẫu nhiên $\mathrm{g(X,Y)}$ là nếu X, Y là rời rạc $\mu_{g(X,Y)} = E[g(X,Y)] = \sum_x \sum_y g(x,y) f(x,y)$

nếu X, Y là bnn liên tục, $\mu_{g(X,Y)} = E[g(X,Y)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x,y) f(x,y) dxdy$

Ví dụ: Cho X, Y là các bnn với ppxs như sau:

$Y \backslash X$	0	1	2
0	3/28	9/28	3/28
1	6/28	6/28	0
2	1/28	0	0

Tính kỳ vọng của biến ngẫu nhiên $g(X, Y) = XY$.

Ví dụ: Hãy tính $E\left(\frac{Y}{X}\right)$ biết rằng hàm mật độ đồng thời là

$$f(x,y) = \begin{cases} \frac{x(1+3y^2)}{4}, & (x;y) \in (0;2) \times (0;1) \\ 0, & (x;y) \notin (0;2) \times (0;1) \end{cases}$$

Định lí

- $E[g(X,Y) \pm h(X,Y)] = E[g(X,Y)] \pm E[h(X,Y)]$.
- $E(\mathrm{a}X + bY) = aE(X) + bE(Y)$
- Nếu X và Y độc lập thì $E(XY) = E(X)E(Y)$

2. Covariance.

Định nghĩa: Cho X và Y là các bnn với ppxs đồng thời $f(x,y)$.

Covariance của X và Y là:

Với X, Y rời rạc

$$\sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] = \sum_x \sum_y (x - \mu_X)(y - \mu_Y) f(x,y)$$

Với X, Y liên tục

$$\sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x-\mu_X)(y-\mu_Y)f(x,y)dxdy$$

Định lí:

Covariance của các bnn X và Y với các kì vọng tương ứng là μ_X, μ_Y, được xác định bởi:.. hay

Với X, Y rời rạc $\sigma_{XY} = \sum_x \sum_y xyf(x,y) - \mu_X\mu_Y$

Với X, Y liên tục $\sigma_{XY} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyf(x,y)dxdy - \mu_X\mu_Y$

3. Hệ số tương quan.

Định nghĩa: Cho X và Y là các bnn với covariance σ_{XY} và các độ lệch chuẩn tương ứng là σ_X và σ_Y.

Hệ số tương quan của X và Y là $\rho_{XY} = \dfrac{\sigma_{XY}}{\sigma_X\sigma_Y}$

Ý nghĩa:

Covariance của hai biến ngẫu nhiên cung cấp thông tin về mối liên hệ giữa hai biến ngẫu nhiên, **nhưng độ lớn của σ_{XY} không cho ta biết về mức độ quan hệ** của hai biến ngẫu nhiên, bởi vì σ_{XY} còn phụ thuộc vào đơn vị đo của cả X và Y.

Hệ số tương quan ρ_{XY} là một phiên bản của covariance mà không phụ thuộc vào đơn vị đo, thường so sánh mức độ quan hệ giữa 2 cặp biến khác nhau và được sử dụng rộng rãi trong thống kê.

Hệ số tương quan thoả mãn bất đẳng thức $-1 \le \rho_{XY} \le 1$.

Khi $\sigma_{XY} = 0$ thì $\rho_{XY} = 0$.

Khi $\rho_{XY} = \pm 1$ thì X, Y có sự phụ thuộc tuyến tính, tức là $Y \equiv a + bX$,

Dấu của σ_{XY}, ρ_{XY} cho biết mối quan hệ X, Y phụ thuộc là thuận hay là nghịch.

Ví dụ Tỉ lệ X nam vận động viên và tỉ lệ Y nữ vận động viên điền kinh hoàn thành bài thi trong cuộc thi marathon được mô tả bằng hàm mật độ đồng thời sau

$f(x,y)=\begin{cases}8xy, & x\in D\\ 0, & x\notin D\end{cases}$. Với $D:\{0\le x\le 1, 0\le y\le x\}$. Hãy tìm σ_{XY}.

Ví dụ Hai anh *A*, *B* ném bóng rổ độc lập nhau, có xác suất trúng rổ lần lượt là 0.8 , 0.6. Anh *A* ném lần lượt 2 quả, anh *B* ném 1 quả. Gọi *X* và *Y* là số bóng trúng rổ của anh *A* và *B*. Hãy tìm phân phối xác suất đồng thời của *X*, *Y* và tính σ_{XY}.

Nhận xét giá trị σ_{XY} và giải thích.

Ví dụ Cho biến ngẫu nhiên hai chiều *X*, *Y* có phân phối xác suất đồng thời là :

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

Tính hệ số tương quan của *X*, *Y*

Bài 7. CÁC PHÂN PHỐI XÁC SUẤT

I. PHÂN PHỐI NHỊ THỨC

1. **Phép thử Bernoulli** có tính chất sau:

+ Gồm ***n phép thử cùng loại được lặp lại***.

+ Mỗi phép thử chỉ có 2 biến cố xảy ra gọi là biến cố thành công có xác suất p và biến cố thất bại có xác suất là $q=1-p$.

+ Các phép thử là độc lập.

2. Phân phối nhị thức:

Cho phép thử Bernoulli với biến cố thành công xuất hiện trong một lần thực hiện có xác suất là p và biến cố thất bại có xác suất là $q=1-p$. Phân phối xác suất của bnn X, biểu thị ***số lần thành công trong n phép thử độc lập***, là

$$\boxed{f(x)=P(X=x)=C_n^x p^x q^{n-x},\quad x=0,1,2,\dots,n}$$

3. Hệ quả:

Xác suất không quá k lần thành công là $P(X \le k) = \sum_{x=0}^{k} C_n^x p^x q^{n-x}$

Xác suất ít nhất k lần thành công là $P(X \ge k) = 1 - P(X < k) = 1 - \sum_{x=0}^{k-1} C_n^x p^x q^{n-x}$

Ý nghĩa: Ứng dụng vào việc kiểm tra sản phẩm của một dây chuyền sản xuất, mỗi lần kiểm tra cho biết là chính phẩm hay phế phẩm.

4. Kì vọng và phương sai của pp nhị thức.

Định lý: Giá trị trung bình và phương sai của phân phối nhị thức là

$$\boxed{\mu = np, \quad \sigma^2 = npq \quad \text{với } q = 1 - p \quad .}$$

Ví dụ: Xác suất để một bệnh nhân sống sót sau khi mắc một loại bệnh hiếm về máu là 0,4. Biết rằng đã có 15 người mắc loại bệnh này, tìm xác suất để

(a) có đúng 5 người sống sót.

(b) có ít nhất 10 người sống sót;

II. PHÂN PHỐI SIÊU BỘI

1. Phân phối siêu bội.

Phân phối xác suất của biến ngẫu nhiên X, biểu thị ***số phần tử thành công trong mẫu cỡ n*** được chọn ngẫu nhiên từ N phần tử trong đó có k phần tử được đặt tên là thành công và $N-k$ phần tử được đặt là thất bại, là

$$\boxed{f(x) = P(X = x) = \frac{C_k^x C_{N-k}^{n-x}}{C_N^n}, \quad x = 0,1,2,\dots,n}$$

2. Kì vọng và phương sai của pp siêu bội.

Định lý: Giá trị trung bình và phương sai của phân phối nhị thức là

$$\boxed{\mu = \frac{nk}{N}, \quad \sigma^2 = \frac{N-n}{N-1} n \frac{k}{N}\left(1 - \frac{k}{N}\right).}$$

Ví dụ: Một lô hàng có 40 sản phẩm, trong đó có 3 phế phẩm. Lô hàng sẽ bị loại nếu kiểm tra ngẫu nhiên một mẫu 5 sản phẩm mà thấy ít nhất 2 phế phẩm. Tính xác suất để lô hàng bị loại.

III. PHÂN PHỐI CHUẨN

1. Định nghĩa:
Biến ngẫu nhiên X với trung bình μ và độ lệch chuẩn σ gọi là phân phối chuẩn nếu hàm mật độ là

$$\boxed{f(x)=n(x;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}} \text{ với } -\infty<x<+\infty$$

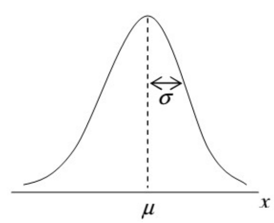
Biến ngẫu nhiên Z có $\mu=0,\sigma=1$ gọi là **phân phối tiêu chuẩn** $n(z,0,1)$.

Đồ thị **đường cong chuẩn** là đường đối xứng có dạng hình quả chuông.

Tính chất của đường cong chuẩn

1. Mode, là điểm trên trục hoành mà tại đó đường cong đạt giá trị lớn nhất $\max f(x)=\frac{1}{\sigma\sqrt{2\pi}}$ tại $x=\mu$.
2. Trục đối xứng là đường thẳng đứng đi qua μ.
3. Có hai điểm uốn tại $x=\mu\pm\sigma$.
4. Đường cong tiệm cận với trục hoành.
5. Tổng diện tích của phần bên dưới đường cong và bên trên trục hoành bằng 1.

2. Công thức xác suất của bnn chuẩn



Diện tích dưới đường cong chắn bởi hai đường $x=x_1,x=x_2$ chính bằng xác suất để bnn nhận giá trị nằm giữa $x=x_1,x=x_2$.

$$P(x_1<X<x_2)=\int_{x_1}^{x_2}n(x;\mu,\sigma)dx=\frac{1}{\sigma\sqrt{2\pi}}\int_{x_1}^{x_2}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx$$

3. Cách tra bảng A3:

Bảng tra A3 có các giá trị của tích phân với $\mu=0,\sigma=1$.

Bảng A3 chỉ ra diện tích phần bên dưới đường cong tiêu chuẩn ứng với $P(Z<z)$.

Ví dụ: Tính (tra thuận)

(a) $P(Z < 2,34)$:

(b) $P(Z > -1,26)$

(c) $P(1,21 < Z < 2,57)$.

Ví dụ (**Tra ngược**) Tìm z_0 nếu

(a) $P(Z < z_0) = 0,8203$

(b) $P(Z > z_0) = 0,2661$,

(c) $P(-z_0 < Z < z_0) = 0,2458$

Đặt $Z = \dfrac{X-\mu}{\sigma}$ đưa bnn chuẩn bất kỳ thành bnn tiêu chuẩn Z với $\mu = 0, \sigma = 1$.

$P(x_1 < X < x_2) = P(z_1 < Z < z_2)$ với $z_1 = \dfrac{x_1 - \mu}{\sigma}, z_2 = \dfrac{x_2 - \mu}{\sigma}$.

IV. ỨNG DỤNG CỦA PHÂN PHỐI CHUẨN

1. **Bài toán 1:** Cho X tuân theo pp chuẩn với $\mu,\ \sigma$. Tìm xác suất để X nằm trong (a,b) hoặc $X > b$ hoặc $X \le a$.

Bước 1: Gọi X là bnn chỉ …

Bước 2: Đặt $Z = \dfrac{X-\mu}{\sigma}$ đưa bnn chuẩn bất kỳ thành bnn tiêu chuẩn Z với $\mu = 0, \sigma = 1$.

Bước 3: Tại $x_1 = a \Rightarrow z = z_1$, $x_2 = b \Rightarrow z = z_2$.

Bước 4: $P(x_1 < X < x_2) = P(z_1 < Z < z_2)$

Ví dụ

a. Một loại pin có tuổi thọ trung bình là 3 năm và độ lệch chuẩn là 0,5 năm. Giả sử tuổi thọ của pin có phân phối chuẩn, tìm xác suất để pin có tuổi thọ ít hơn 2,3 năm.

b. Một loại bóng đèn có tuổi thọ tuân theo phân phối chuẩn với trung bình bằng 800 giờ và độ lệch chuẩn là 40 giờ. Tìm xác suất để bóng đèn có tuổi thọ từ 778 đến 834 giờ.

2. **Bài toán 2**: Cho X tuân theo pp chuẩn với $\mu,\ \sigma$. Tìm giá trị nhỏ nhất, lớn nhất của X hoặc X có giá trị trong (a,b) để xác suất (tỷ lệ) là α.

Ví dụ

c. Trong kỳ thi, điểm thi có phân phối chuẩn với điểm trung bình là 74, độ lệch chuẩn là 7. Nếu có 12% là điểm loại *A* thì một sinh viên cần điểm tối thiểu là bao nhiêu để được nhận điểm loại *A*? Một sinh viên muốn đạt được điểm loại *B* thì cần được tối đa bao nhiêu điểm?

d. Đường kính của một loại pít-tông có phân phối chuẩn với trung bình là 10 cm và độ lệch tiêu chuẩn là 0,03 cm.

a) Hỏi tỷ lệ pít-tông có đường kính vượt quá 10,075 cm là bao nhiêu?

b) Hỏi xác suất để một pít-tông có đường kính từ 9,97 đến 10,03 cm là bao nhiêu?

c) Tìm giá trị *d* của đường kính để 15% số pít-tông có đường kpính nhỏ hơn giá trị đó?

Bài 8: MỘT SỐ THỐNG KÊ MẪU QUAN TRỌNG

I. TỔNG THỂ, MẪU VÀ THỐNG KÊ.

1. **Tổng thể:** là tất cả các quan sát có thể có của một đối tượng mà chúng ta quan tâm.
2. **Một mẫu:** là một tập con của một tổng thể.
3. **Thống kê:** là hàm của các biến ngẫu nhiên tạo thành mẫu ngẫu nhiên.

Cho mẫu nn $X_1, X_2, ..., X_n$ thì $\hat{\Theta} = g(X_1, X_2, ..., X_n)$ gọi là thống kê.

Gọi $X_1, X_2, ..., X_n$ là **mẫu ngẫu nhiên cỡ *n*** từ tổng thể.

Từ những số liệu của mẫu ngẫu nhiên, rút ra được kết luận khách quan về tổng thể và xây dựng phương pháp để rút ra kết luận đó (gọi là **phương pháp mẫu**).

II. MỘT SỐ THỐNG KÊ QUAN TRỌNG

Cho một mẫu ngẫu nhiên $X_1, X_2, ..., X_n$ có cỡ n.

1. **Trung bình mẫu** $\overline{X} = \dfrac{\sum_{i=1}^{n} X_i}{n}$

Chú ý: Thống kê $\overline{X}$ có giá trị $\overline{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$ khi $X_1, X_2, ...$ nhận giá trị $x_1, x_2, ...$

Median mẫu $\quad X = \begin{cases} X_{(n+1)/2} & n = 2k+1 \\ \dfrac{X_{n/2} + X_{(n/2)+1}}{2} & n = 2k \end{cases}$

với $X_1, X_2, ..., X_n$ được sắp xếp theo thứ tự tăng dần.

2. **Mode** là giá trị của mẫu xảy ra thường xuyên nhất hoặc có tần số lớn nhất.

Mode có thể không tồn tại (khi tất cả các quan sát xảy ra cùng tần số).
Mode không nhất thiết là giá trị duy nhất (khi có nhiều giá trị xảy ra cùng với tần số lớn nhất)

3. **Phương sai mẫu** $S^2 = \dfrac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1}$

Định lý: $S^2 = \dfrac{n\sum_{i=1}^{n} X_i^2 - (\sum_{i=1}^{n} X_i)^2}{n(n-1)} = \dfrac{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2}{n-1}$

4. **Độ lệch chuẩn mẫu** ký hiệu bằng S là căn số bậc hai dương của phương sai mẫu.

5. **Tỉ lệ mẫu:** $\hat{P} = \dfrac{X}{n}$ với X là số lần thành công trong n phép thử.

III. TÍNH CÁC THỐNG KÊ BẰNG MÁY CASIO

Ví dụ Khoảng thời gian bằng phút của 10 bệnh nhân đợi ở phòng bác sĩ trước khi khám là

5 11 9 5 10 15 6 10 5 10

(a) Xác định cỡ mẫu

(b) Tính trung bình mẫu, median mẫu, mode

(c) Phương sai mẫu, độ lệch tiêu chuẩn mẫu.

(d) Thời gian lớn hơn 10 phút gọi là đợi lâu. Tính tỉ lệ bệnh nhân không phải đợi lâu.

Ví dụ Khối lượng hộp sữa X (gam) của một dây chuyền đóng hộp cho bởi số liệu sau:

X	996	998	1000	1002	1004	1006
m	2	10	24	35	21	8

(a) Xác định cỡ mẫu.

(b) Tính trung bình mẫu, median mẫu, mode.

(c) Phương sai mẫu, độ lệch chuẩn mẫu

(d) Gọi hộp sữa là chuẩn nếu khối lượng là 1000 ± 5 (gam). Tính tỉ lệ hộp sữa đạt chuẩn.

IV. CÁC HÀM PHÂN PHỐI CỦA THỐNG KÊ MẪU

1. Phân phối lấy mẫu của giá trị trung bình.

Định nghĩa: Phân phối xác suất của một thống kê được gọi là **phân phối lấy mẫu.**

Giả sử một mẫu ngẫu nhiên cỡ n được lấy từ một tổng thể tuân theo phân phối chuẩn có giá trị trung bình μ và phương sai σ^2.

$\overline{X}$ là phân phối chuẩn. Từ $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ ta có $\mu_{\overline{X}} = \frac{1}{n}\sum_{i=1}^{n}\mu_{X_i} = \mu$, $\sigma_{\overline{X}}^2 = \frac{1}{n^2}\sum_{i=1}^{n}\sigma_{X_i}^2 = \frac{\sigma^2}{n}$

Định lý giới hạn trung tâm:

Nếu $\overline{X}$ là giá trị trung bình của một mẫu ngẫu nhiên có kích thước n lấy từ tổng thể có giá trị trung bình μ và phương sai hữu hạn σ^2, khi đó giới hạn phân phối của thống kê $Z = \dfrac{\overline{X} - \mu}{\sigma / \sqrt{n}}$ **là phân phối tiêu chuẩn** $N(z\,;0\,,1)$ khi $n \to \infty$.

Chú ý:

Nếu $n \geq 30$: Phép xấp xỉ chuẩn đối với $\overline{X}$ nói chung tốt cho bất kỳ hình dạng của tổng thể.

Nếu $n < 30$: mức xấp xỉ sẽ chỉ tốt nếu tổng thể không quá khác so với phân phối chuẩn.

Nếu tổng thể được xác định là chuẩn, thì phân phối lấy mẫu của $\overline{X}$ sẽ cho phép một phân phối chuẩn chính xác, cho dù cỡ của mẫu nhỏ.

Ví dụ Một công ty điện sản xuất các loại bóng điện có tuổi thọ tuân theo phân phối chuẩn, với giá trị trung bình bằng 800 giờ và độ lệch tiêu chuẩn 40 giờ.

(a) Tìm xác suất của biến ngẫu nhiên tuổi thọ trung bình từ 775 đến 825 giờ của 16 bóng điện.

(b) Tìm xác suất của biến ngẫu nhiên tuổi thọ từ 775 giờ đến 825 giờ của bóng điện.

Ví dụ Biến ngẫu nhiên X, biểu diễn số trái anh đào trong một bánh kem, có phân phối xác suất như sau :

x	4	5	6	7
$P(X=x)$	0,2	0,4	0,3	0,1

a) Tìm giá trị trung bình μ và phương sai σ^2 của X.

b) Tìm giá trị trung bình $\mu_{\overline{X}}$ và phương sai $\sigma^2_{\overline{X}}$ của giá trị trung bình $\overline{X}$ cho các mẫu ngẫu nhiên của 36 bánh kem.

c) Tìm xác suất khi số lượng trung bình trái anh đào trong cỡ mẫu là 36 bánh kem nhỏ hơn 5,5.

Phân phối mẫu của sự sai khác giữa hai giá trị trung bình.

Nếu các mẫu độc lập cỡ n_1 và n_2 được lấy ngẫu nhiên từ hai tổng thể, có các giá trị trung bình μ_1 và μ_2, các phương sai σ_1^2 và σ_2^2 tương ứng, khi đó phân phối của các thống kê mẫu của hiệu hai giá trị trung bình $\overline{X}_1 - \overline{X}_2$, là phân phối xấp xỉ chuẩn có giá trị trung bình và phương sai bằng:

$$\mu_{\overline{X}_1-\overline{X}_2} = \mu_1 - \mu_2 \quad \text{và} \quad \sigma^2_{\overline{X}_1-\overline{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$

Khi đó thống kê: $Z = \dfrac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{(\sigma_1^2/n_1) + (\sigma_2^2/n_2)}}$ có phân phối xấp xỉ với **phân phối tiêu chuẩn** $n(z; 0,1)$.

Ví dụ Pp các chiều cao của một giống chó sục có chiều cao trung bình 72cm và độ lệch chuẩn 10cm, trong khi đó phân phối của các chiều cao một giống chó xù có chiều cao trung bình 28cm với độ lệch chuẩn 5cm. Tìm xác suất khi giá trị trung bình mẫu cho mẫu nn các chiều cao của 64 chó sục lớn hơn giá trị trung bình mẫu đối với một bnn có chiều cao của 100 chó xù tối đa là 44,2 cm.

2. Phân phối mẫu của phương sai mẫu S^2.

Định lý: Nếu S^2 là phương sai của mẫu ngẫu nhiên có kích thước n được rút ra từ một tổng thể có phân phối chuẩn có phương sai σ^2.

Khi đó thống kê $\chi^2 = \dfrac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n} \dfrac{(X_i - \overline{X})^2}{\sigma^2}$ có phân phối **Khi bình phương** với $v = n-1$ bậc tự do.

Ví dụ Tìm xác suất để một mẫu ngẫu nhiên của 25 quan sát, từ một tổng thể chuẩn có phương sai $\sigma^2 = 6$, sẽ có phương sai mẫu S^2

(a) lớn hơn 9,1.

(b) trong khoảng 3,462 và 10,745.

Bài 9: BÀI TOÁN ƯỚC LƯỢNG TRUNG BÌNH

Biến nn X được xác định dựa trên giá trị x và phân phối xác suất $f(x)$, nhưng các số đặc trưng μ, σ, p mới có tính ổn định và có thể ước lượng được thông qua những quan sát trong thực tế.

Sử dụng các **phương pháp truyền thống** để ước lượng **tham số tổng thể** (μ, p) bằng cách tính các thống kê $\overline{X}$, $\hat{P}$ của mẫu ngẫu nhiên.

I. ƯỚC LƯỢNG ĐIỂM.

Một **ước lượng điểm** về tham số tổng thể θ là một giá trị đơn $\hat{\theta}$ của một thống kê $\hat{\Theta}$.

II. ƯỚC LƯỢNG KHOẢNG

Cho X có dạng ppxs nào đó nhưng các tham số chung θ chưa biết.

Từ mẫu thực nghiệm kích thước n, ta muốn ước lượng θ với xác suất xảy ra $(1-\alpha)100\%$ khá lớn.

Cần xác định khoảng $\theta \in (\hat{\theta}_1, \hat{\theta}_2)$ sao cho $P(\hat{\theta}_L < \theta < \hat{\theta}_U) = 1 - \alpha$.

1. **Định nghĩa:** Khoảng $(\hat{\theta}_1, \hat{\theta}_2)$ gọi là khoảng tin cậy của θ với độ tin cậy $(1-\alpha)100\%$ nếu $\boxed{P(\hat{\theta}_1 < \theta < \hat{\theta}_2) = 1 - \alpha}$

2. **Chú ý:** Với một giá trị của $(1-\alpha)100\%$ có nhiều khoảng ước lượng của θ nên thường chọn khoảng $(\hat{\theta}_1, \hat{\theta}_2)$ sao cho $|\hat{\theta}_2 - \hat{\theta}_1|$ nhỏ nhất, giúp ta xác định chính xác hơn tham số cần tìm.

III. ƯỚC LƯỢNG CỦA MỘT TRUNG BÌNH μ

1. Tổng thể có phân phối chuẩn, khi biết σ

Khoảng tin cậy (1-α)100% của μ, khi biết σ là $\boxed{\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}$

với $z_{\alpha/2}$ là giá trị z tạo một diện tích $\alpha/2$ bên phía phải của nó.

Chú ý: Cách tra bảng tìm giá trị $z_{\alpha/2}$:

Đổi $P(Z > z_{\alpha/2}) = \alpha/2 \Leftrightarrow P(Z < z_{\alpha/2}) = 1 - \alpha/2$.

Ví dụ Tuổi thọ (giờ) bóng đèn do một công ty sản xuất có phân phối chuẩn với độ lệch chuẩn là 40 giờ. Kiểm tra một số bóng đèn cho số liệu sau:

Tuổi thọ	750	800	850	900
Số bóng đèn	3	11	9	5

Tìm khoảng tin cậy 95% của tuổi thọ trung bình các bóng đèn

Định lý: Nếu $\bar{x}$ được dùng để ước lượng μ, khi đó với độ tin cậy $(1-\alpha)100\%$ thì sai số $\boxed{|\bar{x} - \mu| < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}$

Định lý: Nếu $\bar{x}$ được dùng để ước lượng μ, khi đó với độ tin cậy $(1-\alpha)100\%$ thì sai số $|\bar{x} - \mu| < e$ khi cỡ mẫu là $n = \left(\frac{z_{\alpha/2}\sigma}{e}\right)^2$ (n được làm tròn đến số nguyên tiếp theo).

Ví dụ Hàm lượng kẽm (g/ml) đo tại một số điểm trên sông Theme cho số liệu sau

1.9	2.2	1.8	1.9	2.0	2.1	2.2

Giả thiết hàm lượng kẽm có phân phối chuẩn với độ lệch chuẩn là 0,3. Nếu muốn sai số của ước lượng μ nhỏ hơn 0,05 với độ tin cậy 98% thì cỡ mẫu đo là bao nhiêu?

2. Tổng thể có phân phối bất kỳ, khi chưa biết σ nhưng biết cỡ mẫu lớn.

Người ta chứng minh với $n \geq 30$ thì $\sigma \approx s$.

Khoảng tin cậy $(1-\alpha)100\%$ của μ, khi chưa biết σ nhưng $n \geq 30$ là

$$\boxed{\overline{x} - z_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \overline{x} + z_{\alpha/2}\frac{s}{\sqrt{n}}}$$

Ví dụ Để đo mức tiêu hao xăng X (lít) của một loại ôtô, người ta thử nghiệm và ghi được số liệu như sau. Tìm khoảng tin cậy 94% của mức tiêu hao xăng trung bình của loại ôtô đó.

X	9,7	9,8	10,0	10,1	10,3
Số xe	5	7	13	9	6

3. Tổng thể có phân phối chuẩn, khi chưa biết σ nhưng biết cỡ mẫu nhỏ.

Khoảng tin cậy $(1-\alpha)100\%$ của μ, khi chưa biết σ nhưng $n < 30$ là

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

trong đó $t_{\alpha/2}$ là giá trị t với $v = n-1$ bậc tự do, sinh ra một diện tích bằng $\alpha/2$ bên phía phải.

Ví dụ Kiểm tra ngẫu nhiên 12 sinh viên đào tạo về đánh máy cho biết số chữ đánh trong một phút như sau.

81 78 83 79 80 82 80 78 77 83 75 80

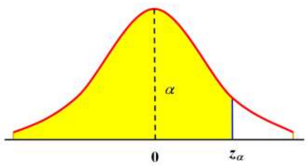
Tìm khoảng tin cậy 90% cho số chữ trung bình của tất cả sinh viên đánh máy. Giả thiết số chữ đánh máy là phân phối chuẩn.

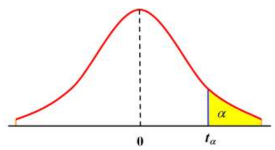
Ví dụ Hàm lượng axit sulfuric (lít) của 7 container là

9,8   10,2   10,4   9,8   10,0   10,2   9,6

Tìm khoảng tin cậy 95% của hàm lượng axit sulfuric trung bình các container, giả sử phân phối xấp xỉ chuẩn.

Chú ý: Miền có diện tích α trong A.3 nằm bên trái Z_α, còn trong bảng A.4 nằm bên phải t_α.





$10 ƯỚC LƯỢNG HIỆU 2 TRUNG BÌNH VÀ TỈ LỆ

I. ƯỚC LƯỢNG HIỆU HAI TRUNG BÌNH $\mu_1 - \mu_2$.

1. Tổng thể có phân phối chuẩn, khi biết σ_1, σ_2.

Khoảng tin cậy $(1-\alpha)100\%$ của $\mu_1 - \mu_2$, khi biết σ_1, σ_2 là

$$\bar{x}_1 - \bar{x}_2 - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < \bar{x}_1 - \bar{x}_2 + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Ví dụ

Tiến hành thí nghiệm hai loại động cơ *A* và *B* để so sánh số dặm đi được trên mỗi gallon xăng. Trong 50 thí nghiệm cho *A* và 75 thí nghiệm cho *B*, thì số dặm đi trung bình lần lượt là 36 và 42. Tìm khoảng tin cậy 96% của $\mu_B - \mu_A$, trong đó μ_A, μ_B là trung bình tổng thể của tổng số dặm của động cơ *A*, *B*. Giả thiết là phân phối chuẩn có $\sigma_A = 6, \sigma_B = 8$.

2. Tổng thể phân phối bất kỳ, khi chưa biết σ_1, σ_2 nhưng cỡ mẫu lớn.

Khoảng tin cậy $(1-\alpha)100\%$ của $\mu_1 - \mu_2$, khi chưa biết σ_1, σ_2 nhưng biết cỡ mẫu lớn là:

Với $n \geq 30$ chưa biết σ_1, σ_2 hoặc không biết phân phối của tổng thể thì vẫn sử dụng công thức trên bằng cách sử dụng s_1, s_2 thay cho σ_1, σ_2.

$$\bar{x}_1 - \bar{x}_2 - z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < \bar{x}_1 - \bar{x}_2 + z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$$

Ví dụ Một viện nghiên cứu về khối lượng một loại quả cam theo hai chế độ chăm sóc. Khi thu hoạch, họ ghi lại số liệu sau. Tìm khoảng tin cậy 98% của hiệu khối lượng trung bình quả cam theo 2 chế độ.

Khối lượng (g)	38	39	40	41	42
Số quả CĐ1	8	12	26	33	17
Số quả CĐ2	15	21	32	11	5

3. Tổng thể phân phối chuẩn, chưa biết σ_1, σ_2 cỡ mẫu nhỏ và $\sigma_1 = \sigma_2$.

Khoảng tin cậy $(1-\alpha)100\%$ của $\mu_1 - \mu_2$, chưa biết σ_1, σ_2 cỡ mẫu nhỏ và $\sigma_1 = \sigma_2$ là

$$\bar{x}_1 - \bar{x}_2 - t_{\alpha/2}s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} < \mu_1 - \mu_2 < \bar{x}_1 - \bar{x}_2 + t_{\alpha/2}s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$s_p = \sqrt{\dfrac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$ là độ lệch chuẩn chung của 2 mẫu. $t_{\alpha/2}$ là giá trị phân phối-*t* với $v = n_1 + n_2 - 2$ bậc tự do

Ví dụ Một phòng thị trường đã nghiên cứu mối quan hệ giữa chi phí quảng cáo X ($) và lợi nhuận Y ($) hàng tuần và ghi lại như sau:

X	40	20	25	20	30	40	50	20
Y	385	395	400	365	475	440	490	425

Tìm khoảng tin cậy 97% của hiệu hai trung bình tổng thể giữa lợi nhuận và chi phí quảng cáo. Giả thiết quan sát lấy từ các tổng thể chuẩn có phương sai bằng nhau.

II. ƯỚC LƯỢNG MỘT TỶ LỆ VỚI CỠ MẪU LỚN

Khoảng tin cậy $(1-\alpha)100\%$ của tỷ lệ $\boxed{\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}}$

với $\hat{p}$ là tỷ lệ của các thành công trong một mẫu ngẫu nhiên có kích thước n và $\hat{q} = 1 - \hat{p}$.

Định lý: Với độ tin cậy $(1-\alpha)100\%$ thì sai số $|\hat{p}-p|$ sẽ không vượt quá $z_{\alpha/2}\sqrt{\dfrac{\hat{p}\hat{q}}{n}}$.

Định lý: Với độ tin cậy $(1-\alpha)100\%$ thì sai số sẽ nhỏ hơn một lượng cụ thể e khi cỡ mẫu là $\boxed{n=\dfrac{z_{\alpha/2}^2\hat{p}\hat{q}}{e^2}}$ (n được làm tròn đến số nguyên tiếp theo)

Nếu không có mẫu sơ bộ và $\hat{p}$ được sử dụng là một ước lượng của p, với độ tin cậy $(1-\alpha)100\%$ thì sai số sẽ không vượt quá một lượng cụ thể e khi cỡ mẫu là

$$\boxed{n=\frac{z_{\alpha/2}^2}{4e^2}}$$

Ví dụ Trong một mẫu ngẫu nhiên 500 gia đình có ti vi tại Hà Nội, có 340 gia đình đăng ký HBO.

a) Tìm khoảng tin cậy 95% cho tỷ lệ thực các gia đình trong thành phố này đăng ký sử dụng HBO. Từ đó ước lượng số gia đình đăng ký HBO trong 2000 gia đình có ti vi tại thành phố này.

b) Nếu độ tin cậy là 95% thì cần điều tra bao nhiêu gia đình để sai số ước lượng p nhỏ hơn 0,02

(c) Nếu không có mẫu sơ bộ trên thì cần lấy mẫu cỡ bao nhiêu để độ tin cậy 95% của sai số ước lượng p nhỏ hơn 0,02

III. ƯỚC LƯỢNG HIỆU HAI TỶ LỆ VỚI CỠ MẪU LỚN

Khoảng tin cậy $(1-\alpha)100\%$ hiệu hai tỷ lệ p_1-p_2 là

$$\boxed{\hat{p}_1-\hat{p}_2-z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2}}<p_1-p_2<\hat{p}_1-\hat{p}_2+z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2}}}$$

Ví dụ Mười trường kỹ thuật ở Mỹ được khảo sát cho kết quả là: 250 kỹ sư điện có 80 người là nữ; 175 kỹ sư hóa học có 40 người là nữ. Tìm khoảng tin cậy 90% cho sai số của tỷ lệ nữ trong hai ngành kỹ thuật này.

Ví dụ Trong một nghiên cứu của Đại học Bách khoa Viginia xác định rằng ở 5^0C trong 200 hạt giống có 102 hạt nảy mầm, ở 15^0C trong số 250 hạt giống có 131 hạt nảy mầm. Tính khoảng tin cậy 98% sai số giữa tỷ lệ nảy mầm tại 2 nhiệt độ đó và xác định xem tỷ lệ đó có khác biệt nhau hay không?

Bài 11 KIỂM ĐỊNH GIẢ THUYẾT VỀ TRUNG BÌNH

I. CÁC KHÁI NIỆM CHUNG.

1. Giả thuyết thống kê

Một sự nhận định liên quan tới số liệu của một hay nhiều tổng thể gọi là một **giả thuyết thống kê.**

2. Kiểm định một giả thuyết thống kê.

Lấy một mẫu ngẫu nhiên từ tập hợp được quan tâm và sử dụng dữ liệu có trong mẫu để đưa ra bằng chứng mà theo đó, ta chấp nhận hoặc không chấp nhận giả thuyết.

3. Giả thuyết thống kê và đối thuyết

Một **giả thuyết thống kê** là một sự xác nhận hay phỏng đoán liên quan đến một hay nhiều tổng thể, ký hiệu là H_0.

Sự bác bỏ H_0 dẫn đến sự chấp nhận một đối thuyết, ký hiệu là H_1.

Giả thuyết $H_0 : \theta = \theta_0$, đối thuyết $H_1 : \theta \neq \theta_0$, gọi là kiểm định hai phía.

$H_0 : \theta = \theta_0$, $H_1 : \theta > \theta_0$ hoặc $H_0 : \theta = \theta_0$ $H_1 : \theta < \theta_0$, gọi là kiểm định một phía.

4. Các khả năng có thể xảy ra trong kiểm định giả thuyết.

	H_0 đúng	**H_0 sai**
Chấp nhận H_0	Quyết định đúng	Sai lầm loại II
Bác bỏ H_0	Sai lầm loại I	Quyết định đúng

Bác bỏ giả thuyết đúng gọi là **sai lầm loại I.** Chấp nhận giả thuyết sai gọi là **sai lầm loại II.**

Xác suất mắc phải sai lầm loại I, cũng được gọi là **mức ý nghĩa**, được ký hiệu là α.

5. Miền bác bỏ và miền chấp nhận

Chấp nhận một giả thuyết chỉ đơn thuần là dữ liệu không cho đủ thông tin để bác bỏ nó. Do sai lầm loại I dễ kiểm soát hơn nên chọn trước số α rất nhỏ như là một ngưỡng để xác định sai lầm loại I .

6. Các bước bài toán kiểm định giả thuyết

1. Mô tả giả thuyết $H_0 : \theta = \theta_0$ và đối thuyết phù hợp $H_1 : \theta \neq \theta_0$, hoặc $\theta > \theta_0$ hoặc $\theta < \theta_0$.
2. Chọn chỉ tiêu kiểm định phù hợp và tính giá trị của chỉ tiêu kiểm định từ dữ liệu mẫu.
3. Chọn mức ý nghĩa α và tìm miền bác bỏ H_0.
4. Kiểm tra và quyết định bác bỏ hay chấp nhận H_0.
5. Kết luận

II. KIỂM ĐỊNH VỀ MỘT TRUNG BÌNH.

1. Kiểm định về một trung bình, khi biết σ.

Giả thuyết H_0: $\mu = \mu_0$

Chỉ tiêu kiểm định: $Z = \dfrac{\overline{X} - \mu_0}{\dfrac{\sigma}{\sqrt{n}}}$

Đối thuyết H_1	Miền bác bỏ H_0
$\mu \neq \mu_0$	$D = (-\infty; -z_{\alpha/2}) \cup (z_{\alpha/2}; +\infty)$
$\mu > \mu_0$	$D = (z_\alpha; +\infty)$
$\mu < \mu_0$	$D = (-\infty; -z_\alpha)$

Ví dụ Một mẫu ngẫu nhiên gồm 100 giấy báo tử ở Mỹ cho thấy tuổi thọ trung bình là 71,8 năm. Giả sử độ lệch chuẩn tổng thể là 8,9 năm. Với mức ý nghĩa là 0,05, kiểm định giả thuyết

a. Tuổi thọ trung bình hiện nay lớn hơn 70 năm.

b. Tuổi thọ trung bình hiện nay nhỏ hơn 73 năm.

Ví dụ Một nhà sản xuất tuyên bố một loại dây câu cá mới có thể chịu được trọng lượng trung bình là 8 kg, với $\sigma = 0,5$ kg. Lấy ngẫu nhiên 50 dây kiểm tra thấy trọng lượng trung bình có thể chịu được là 7,8 kg. Hãy kiểm định lời tuyên bố của nhà sản xuất, với $\alpha = 0,01$

Ví dụ Khối lượng trứng quả gà của một trang trại cho số liệu sau

Khối lượng (g)	120	130	140	150
Số quả	8	24	31	19

Kiểm định giả thuyết khối lượng trung bình quả trứng gà lớn hơn 135 g. Biết rằng khối lượng quả trứng gà có phân phối chuẩn với độ lệch chuẩn 5g. Cho mức ý nghĩa 0,02.

2. Kiểm định về một trung bình, khi chưa biết σ nhưng biết cỡ mẫu lớn.

Vẫn có thể sử dụng chỉ tiêu kiểm định trên bằng cách thay $\sigma \approx s$, ta có $z = \dfrac{\overline{x} - \mu_0}{\dfrac{s}{\sqrt{n}}}$

Ví dụ Chiều cao của giống cây cho số liệu sau

Chiều cao (m)	1,5	2	2,2	2,5
Số cây	13	24	31	17

Kiểm định giả thuyết chiều cao trung bình giống cây nhỏ hơn 2,2m. Cho mức ý nghĩa 0,05.

3. Kiểm định về một trung bình có pp chuẩn, khi chưa biết σ nhưng biết cỡ mẫu nhỏ.

Giả thuyết H_0: $\mu = \mu_0$

Chỉ tiêu kiểm định: $t = \dfrac{\overline{x} - \mu_0}{\dfrac{s}{\sqrt{n}}}$ với $v = n - 1$ bậc tự do

Đối thuyết H_1	Miền bác bỏ H_0
$\mu \neq \mu_0$	$D = (-\infty; -t_{\alpha/2}) \cup (t_{\alpha/2}; +\infty)$
$\mu > \mu_0$	$D = (t_\alpha; +\infty)$
$\mu < \mu_0$	$D = (-\infty; -t_\alpha)$

Ví dụ Cơ thể người cần khoảng 220 miligam natri một ngày. Khảo sát 20 loại suất ăn của một hãng, ta có số liệu sau:

natri (mg)	200	210	220	230	240	250	260
Số suất ăn	2	3	4	4	2	3	2

Với mức ý nghĩa 0,05, có thể nói lượng natri trung bình trong các suất ăn của hãng này là lớn hơn 220 miligam không? Giả sử phân phối của lượng natri là chuẩn.

Ví dụ 3 Khối lượng (kg) của một loại hộp dầu, bị bay hơi sau 1 năm, cho số liệu như sau

1,5 1,8 2,0 1,9 1,7 1,8 1,5 2,1

Kiểm định giả thuyết khối lượng trung bình hộp dầu sau 1 năm là 1,75 kg, với mức ý nghĩa 0,05.

Bài 11 KIỂM ĐỊNH HIỆU 2 TRUNG BÌNH VÀ TỶ LỆ

I. KIỂM ĐỊNH VỀ HIỆU HAI TRUNG BÌNH.

1. Kiểm định hiệu hai trung bình, biết σ_1, σ_2.

Giả thuyết H_0: $\mu_1 - \mu_2 = d_0$

Chỉ tiêu kiểm định: $$z = \frac{\overline{x}_1 - \overline{x}_2 - d_0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

Đối thuyết H_1	Miền bác bỏ H_0
$\mu_1 - \mu_2 \neq d_0$	$D = (-\infty; -z_{\alpha/2}) \cup (z_{\alpha/2}; +\infty)$
$\mu_1 - \mu_2 > d_0$	$D = (z_\alpha; +\infty)$
$\mu_1 - \mu_2 < d_0$	$D = (-\infty; -z_\alpha)$

Chú ý:

Ngôn ngữ	Biểu thức
1) Kiểm định trung bình của X_1 **lớn hơn** trung bình của X_2 **là (bằng) *a***.	$\mu_1 - \mu_2 = a$
2) Kiểm định trung bình của X_1 **lớn hơn** trung bình của X_2 **trên *a***.	$\mu_1 - \mu_2 > a$
3) Kiểm định trung bình của X_1 **nhỏ hơn** trung bình của X_2 ***dưới* a**.	$\mu_1 - \mu_2 < a$

Ví dụ Thí nghiệm nuôi 100 con gà theo phương pháp 1 cho $\overline{x}_1 = 1{,}1$ kg và 150 con theo phương pháp 2 cho $\overline{x}_2 = 1{,}2$ kg. Kiểm định giả thuyết khối lượng trung bình gà nuôi theo 2 phương pháp là khác nhau. Giả thuyết khối lượng gà nuôi theo luật chuẩn với phương sai của phương pháp 1 và phương pháp 2 lần lượt là 0,04 và 0,098. Cho $\alpha = 0{,}05$.

2. Kiểm định hiệu hai trung bình, khi chưa biết σ_1, σ_2 nhưng biết cỡ mẫu lớn.

Thay $\sigma_1 \approx s_1, \sigma_2 \approx s_2$ trong chỉ tiêu kiểm định $z = \dfrac{\overline{x}_1 - \overline{x}_2 - d_0}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$

Ví dụ Một hãng sản xuất xe hơi thực hiện thí nghiệm 2 loại lốp *A* hay *B* và ghi lại số km (đơn vị 1000) đi được đến khi phải thay lốp. Kiểm định giả thuyết không có sự khác biệt giữa hai loại lốp, với mức ý nghĩa 0,04.

Số km	37,5	38	38,5	39	39,5	40	40,5	41
Số lốp A	5	8	14	6	8	22	11	6
Số lốp B	2	12	9	15	25	7	3	2

Ví dụ Điểm thi môn Toán 5 của sinh viên chương trình chính quy và tiến tiến cho kết quả sau:

Điểm (hệ số 4)	0	1	2	3	4
Số SV CTCQ	15	27	31	12	8
Số SV CTTT	6	18	24	15	7

Kiểm định giả thuyết điểm thi trung bình của sinh viên CTTT lớn hơn CTCQ trên 0,2 điểm. Cho $\alpha = 0,04$.

3. Kiểm định hiệu hai trung bình, khi chưa biết σ_1, σ_2 nhưng biết cỡ mẫu nhỏ và $\sigma_1 = \sigma_2$.

Giả thuyết H_0: $\mu_1 - \mu_2 = d_0$

Chỉ tiêu kiểm định: $t = \dfrac{\overline{x}_1 - \overline{x}_2 - d_0}{s_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$ với $v = n_1 + n_2 - 2$ bậc tự do

$$s_p = \sqrt{\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}}$$

Đối thuyết H_1	Miền bác bỏ H_0
$\mu_1 - \mu_2 \neq d_0$	$D = (-\infty; -t_{\alpha/2}) \cup (t_{\alpha/2}; +\infty)$
$\mu_1 - \mu_2 > d_0$	$D = (t_\alpha; +\infty)$
$\mu_1 - \mu_2 < d_0$	$D = (-\infty; -t_\alpha)$

Ví dụ Cân nặng (kg) trẻ sơ sinh gốc Phi và gốc Á ở một bệnh viện quốc tế cho số liệu sau:

Cân nặng	2,5	3	3,5	4
Số trẻ gốc Phi	1	3	7	5
Số trẻ gốc Á	4	5	3	1

Kiểm định giả thuyết cân nặng trung bình của trẻ gốc Phi lớn hơn trẻ gốc Á là 0,3 kg, với $\alpha = 0,02$. Biết tổng thể phân phối chuẩn có độ lệch chuẩn bằng nhau.

Ví dụ Kiểm tra 12 miếng kim loại 1 và 10 miếng kim loại 2 lần lượt có độ mài mòn trung bình là 85 và 81 đơn vị, với độ lệch mẫu $s_1 = 4, s_2 = 5$. Kiểm định giả thuyết độ mài mòn trung bình của kim loại 1 lớn hơn kim loại 2 trên 2 đơn vị, với mức ý nghĩa 0,05. Giả sử các độ mài mòn xấp xỉ chuẩn, phương sai bằng nhau.

II. KIỂM ĐỊNH TỶ LỆ

1. Kiểm định về một tỷ lệ.

Giả thuyết H_0 : $p = p_0$

Chỉ tiêu kiểm định: $z = \dfrac{x - np_0}{\sqrt{np_0q_0}}$, $q_0 = 1 - p_0$, x là số phần tử thành công trong mẫu n.

Đối thuyết H_1	Miền bác bỏ H_0
$p \neq p_0$	$D = (-\infty; -z_{\alpha/2}) \cup (z_{\alpha/2}; +\infty)$
$p > p_0$	$D = (z_\alpha; +\infty)$
$p < p_0$	$D = (-\infty; -z_\alpha)$

Ví dụ Một loại thuốc an thần thường dùng có tác dụng tới 60% người sử dụng. Kết quả thử nghiệm loại thuốc mới, trong 100 người cho thấy tác dụng tới 70 người. Có thể tin được hay không tỷ lệ tác dụng loại thuốc mới khác với loại thường dùng, với mức ý nghĩa 0,05.

Ví dụ Thống kê khối lượng của một loại hộp sữa bột, cho số liệu sau:

Khối lượng (g)	398	399	400	401
Số hộp sữa	11	60	117	62

Các hộp sữa có khối lượng dưới 399 g được coi là sai qui cách. Với mức ý nghĩa 1% hãy kiểm định giả thuyết tỷ lệ hộp sữa sai qui cách nhỏ hơn 5%.

2. Kiểm định về hiệu hai tỷ lệ

Giả thuyết H_0 : $p_1 = p_2$

Chỉ tiêu kiểm định: $z = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$ với $\hat{p} = \dfrac{x_1 + x_2}{n_1 + n_2}$ (tỷ lệ chung 2 mẫu)

Đối thuyết H_1	Miền bác bỏ H_0
$p_1 \neq p_2$	$B = (-\infty; -z_{\alpha/2}) \cup (z_{\alpha/2}; +\infty)$
$p_1 > p_2$	$B = (z_\alpha; +\infty)$
$p_1 < p_2$	$B = (-\infty; -z_\alpha)$

Ví dụ Trong một cuộc thi đồng đội trên truyền hình. Có 120 trên 200 sinh viên trường A trả lời đúng và 240 trên 500 sinh viên trường B trả lời đúng. Liệu có thể cho rằng tỷ lệ sinh viên trả lời đúng của 2 trường như nhau hay không. Cho mức ý nghĩa 0,025.

Ví dụ Người ta hỏi ngẫu nhiên 2 nhóm phụ nữ có gia đình về kế hoạch sinh con đầu vào thời điểm này không. Nhóm 1 có 240 trong số 300 người có gia đình chưa quá 2 năm và nhóm 2 có 288 trong số 400 người có gia đình 5 năm dự định sinh con đầu. Kiểm định giả thuyết tỷ lệ phụ nữ có kế hoạch sinh con ở nhóm 1 là cao hơn tỷ lệ ở nhóm 2, với mức ý nghĩa 0,01.

Ví dụ Thống kê số học sinh thi trượt tốt nghiệp ở 2 trường THPT như sau:

Trường	Tổng số học sinh	Số trượt tốt nghiệp
A	1900	380
B	2600	637

Với mức ý nghĩa 0,05, kiểm định giả thuyết tỷ lệ học sinh thi trượt tốt nghiệp ở 2 trường trên là bằng nhau.

Bài 12 HỒI QUY TUYẾN TÍNH ĐƠN

I. MỞ ĐẦU VỀ HỒI QUY TUYẾN TÍNH

Ta bắt đầu từ số liệu X, Y về tuổi, chiều cao (cm) của nhóm trẻ như sau:

X / Y	4	5	6	7

100	8	7	3	2
105	4	6	2	1
110	3	4	7	5
115	1	3	5	9

Câu hỏi đặt ra là với trẻ 5 tuổi của nhóm này thì chiều cao trung bình là bao nhiêu? Khi trẻ 8 tuổi thì có chiều cao trung bình là bao nhiêu?

Bài toán dẫn đến khái niệm $Y|x$ là biến ngẫu nhiên Y tương ứng với một giá trị cố định x.

Ký hiệu trung bình, phương sai của nó tương ứng là $\mu_{Y|x}$, $\sigma^2_{Y|x}$.

Với $x = x_i$, ký hiệu $Y|x_i$ là biến ngẫu nhiên Y_i với trung bình $\mu_{Y|x_i}$ và phương sai $\sigma^2_{Y|x_i}$.

Khi nối các điểm $(\mu_{Y|x_i}, x_i)$, sẽ có một đường xuyên qua các điểm của tập dữ liệu đã cho. Đường này gọi là **đường hồi quy.**

II. HỒI QUY TUYẾN TÍNH ĐƠN VÀ PHƯƠNG PHÁP BÌNH PHƯƠNG TỐI THIỂU

Nếu ta giả thiết rằng tất cả các số trung bình $\mu_{Y|x_i}$ nằm trên một đường thẳng, mỗi biến ngẫu nhiên Y_i có thể được mô tả bằng **mô hình hồi quy tuyến tính đơn:**

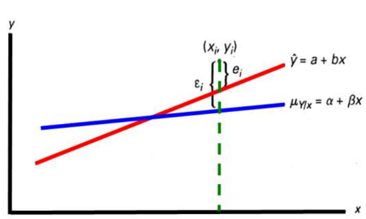
$$Y_i = \mu_{Y|x_i} + E_i = \alpha + \beta x_i + E_i$$

trong đó, sai số ngẫu nhiên E_i, sai số mô hình, cần phải có giá trị trung bình bằng 0.

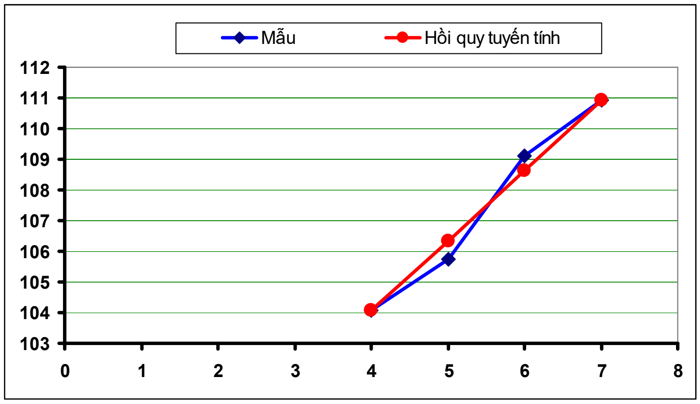
Mỗi cặp quan sát (x_i, y_i) trong mẫu thử thỏa mãn phương trình: $y_i = \alpha + \beta x_i + \varepsilon_i$.

trong đó ε_i là giá trị được tính toán bởi E_i khi Y_i nhận giá trị y_i.

Phương trình trên có thể được coi là mô hình cho quan sát đơn y_i.



So sánh ε_i với số dư e_i



Tương tự, sử dụng **đường hồi quy ước lượng** hoặc **đường hồi quy phù hợp** $\hat{y} = a + bx$, mỗi cặp quan sát thỏa mãn quan hệ: $y_i = a + bx_i + e_i$, trong đó $e_i = y_i - \hat{y}_i$ được gọi là **phần dư** và mô tả sai số trong mức độ phù hợp của mô hình tại điểm dữ liệu thứ *i*.

Phương pháp bình phương tối thiểu để ước lượng các hệ số hồi quy.

Cho mẫu (x_i, y_i), $i = 1,2,\ldots n$, các ước lượng bình phương tối thiểu a, b của hệ số hồi quy α, β là:

$$b = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} \text{ và } a = \frac{\sum_{i=1}^{n} y_i - b\sum_{i=1}^{n} x_i}{n} = \overline{y} - b\overline{x}$$

Chú ý:

Phương trình đường hồi quy tuyến tính thực nghiệm X theo Y là $\hat{x} = c + dy$.

Một số đường hồi quy phi tuyến được tuyến tính hóa qua phép đổi biến

$$\hat{y} = e^{a+bx} \Leftrightarrow y^* = a + bx, \quad y^* = \ln \hat{y}.$$

$$\hat{y} = \ln(a + bx) \Leftrightarrow y^* = a + bx, \quad y^* = e^{\hat{y}}.$$

$$\hat{y} = a + bx^m \Leftrightarrow \hat{y} = a + bz, \quad z = x^m.$$

Thống kê mẫu hai chiều

Mẫu cỡ n: cho dưới dạng cặp số: $(x_1, y_1), \ldots, (x_n, y_n)$ hoặc bảng tần số:

Y \ X	x_1	…	x_k

y_1	m_{11}	…	m_{k1}
…		…	
y_h	m_{1h}	…	m_{kh}

CASIO 570VN- Plus

+ Xoá bộ nhớ: $\boxed{SHIFT}\ \boxed{CLR}\ \boxed{3}\ \boxed{=}\ \boxed{=}$,

+ $\boxed{SHIFT}\ \boxed{MODE}\ \boxed{\vee}\ \boxed{4:STAT}\ \boxed{1:ON}$,

+ Vào chế độ REG: $\boxed{MODE}\ \boxed{3}\ \boxed{2:\text{a}X+b}$

+ Nhập số liệu: cột 1: x_i, cột 2: y_j, cột 3: tần số m_{ij}

+ Xem kết quả: $\boxed{SHIFT}\ \boxed{1}\ \boxed{3:SUM}$

$\boxed{SHIFT}\ \boxed{1}\ \boxed{4:Var} \Rightarrow n, \overline{x}, s_x, \overline{y}, s_y, A, B, r$

Ví dụ Số liệu về tuổi *X,* chiều cao *Y* của nhóm trẻ

Y \ X	4	5	6	7
100	8	7	3	2
105	4	6	2	1
110	3	4	7	5
115	1	3	5	9

a) Tính $n, \sum_{i=1}^{n} x_i,\ \sum_{i=1}^{n} y_i,\ \sum_{i=1}^{n} x_i^2,\ \sum_{i=1}^{n} x_i y_i$. Tìm phương trình đường hồi quy tuyến tính thực nghiệm của *Y* theo *X*.

b) Dự đoán chiều cao trung bình của trẻ 5 và 8 tuổi

Ví dụ Một cuộc nghiên cứu được tiến hành trên số lượng đường chuyển hóa theo một quy trình xác định tại các nhiệt độ khác nhau. Số liệu được mã hóa và ghi lại như sau:

Nhiệt độ x	1,0	1,1	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9	2,0
Đường chuyển hóa y	8,1	7,8	8,5	9,8	9,5	8,9	8,6	10,2	9,3	9,2	10,5

a) Tính các giá trị $n,\ \sum_{i=1}^{n} x_i, \sum_{i=1}^{n} y_i, \sum_{i=1}^{n} x_i^2, \sum_{i=1}^{n} y_i^2, \sum_{i=1}^{n} x_i y_i$

b) Tìm phương trình đường hồi quy tuyến tính thực nghiệm của Y theo X.

(c) Tìm phương trình đường hồi quy tuyến tính thực nghiệm của X theo Y.

(d) Tính số lượng đường chuyển hóa trung bình được sản xuất ra khi nhiệt độ là 1,75.

Tính chất của các ước lượng bình phương tối thiểu

Các giá trị *a* và *b*, dựa trên mẫu đã cho gồm *n* quan sát, chỉ là các ước lượng của thông số thực α, β.

Vì thế có thể coi là các giá trị được tính toán bởi các biến ngẫu nhiên A và B được xác định là:

$$B = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})Y_i}{\sum_{i=1}^{n}(x_i - \overline{x})^2} .$$

Kết quả: $\mu_B = \beta$, $\sigma_B^2 = \dfrac{\sigma^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$, $\mu_A = \alpha$, $\sigma_A^2 = \dfrac{\sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n}(x_i - \overline{x})^2}\sigma^2$

Các ước lượng bình phương nhỏ nhất đối với α, β đều là những ước lượng không chệch. ịnh lý: Ước lượng không chệch của σ^2 là $\boxed{s^2 = \sum_{i=1}^{n}\frac{(y_i - \hat{y})^2}{n-2} = \frac{S_{yy} - bS_{xy}}{n-2}}$

với $S_{xx} = \sum_{i=1}^{n}(x_i - \overline{x})^2$, $S_{yy} = \sum_{i=1}^{n}(y_i - \overline{y})^2$, $S_{xy} = \sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$

Chú ý: Thường tính toán các số trên ở dạng sau:

$$S_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{1}{n}(\sum_{i=1}^{n} x_i)^2,\ S_{yy} = \sum_{i=1}^{n} y_i^2 - \frac{1}{n}(\sum_{i=1}^{n} y_i)^2,\ S_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{1}{n}(\sum_{i=1}^{n} x_i).(\sum_{i=1}^{n} y_i),$$

$$b = \frac{S_{xy}}{S_{xx}}$$

Bài 13 KHOẢNG TIN CẬY VÀ KIỂM ĐỊNH GIẢ THIẾT CÁC HỆ SỐ ĐƯỜNG HỒI QUY

I. Khoảng tin cậy cho các hệ số đường hồi quy.

Khoảng tin cậy $(1-\alpha)100\%$ đối với tham số α,β trong đường hồi quy $\mu_{Y|x}=\alpha+\beta X$ là:

$$b-\frac{t_{\alpha/2}s}{\sqrt{S_{xx}}}<\beta<b+\frac{t_{\alpha/2}s}{\sqrt{S_{xx}}} \quad \text{và} \quad a-\frac{t_{\alpha/2}s\sqrt{\sum_{i=1}^{n}x_i^2}}{\sqrt{nS_{xx}}}<\alpha<a+\frac{t_{\alpha/2}s\sqrt{\sum_{i=1}^{n}x_i^2}}{\sqrt{nS_{xx}}}$$

trong đó $t_{\alpha/2}$ là giá trị của phân phối $-t$ với $n-2$ bậc tự do

Chú ý: Ký hiệu α được sử dụng ở đây theo cả hai cách hoàn toàn không liên quan: đó là mức ý nghĩa và điểm chặn của đường hồi quy.

Ví dụ Số liệu về tuổi *X*, chiều cao *Y* của nhóm trẻ

Y \ X	4	5	6	7
100	8	7	3	2
105	4	6	2	1
110	3	4	7	5
115	1	3	5	9

Từ bảng dữ liệu tìm 95% khoảng tin cậy đối với tham số

a) hệ số góc β

b) điểm chặn α trong đường hồi quy $\mu_{Y|x}=\alpha+\beta x$.

II. Kiểm định cho các hệ số đường hồi quy

Giả thiết $H_0:\beta=\beta_0$ Đối thiết: $H_1:\beta\neq\beta_0$.

Chỉ tiêu kiểm định: $T=\dfrac{B-\beta_0}{s/\sqrt{S_{xx}}} \Rightarrow t=\dfrac{b-\beta_0}{s/\sqrt{S_{xx}}}$

Giả thiết $H_0:\alpha=\alpha_0$ Đối thiết: $H_1:\alpha\neq\alpha_0$

Chỉ tiêu kiểm định: $T=\dfrac{A-\alpha}{S\sqrt{\sum_{i=1}^{n}x_i^2/nS_{xx}}} \Rightarrow t=\dfrac{a-\alpha_0}{s\sqrt{\sum_{i=1}^{n}x_i^2/nS_{xx}}}$

III. Khoảng tin cậy của trung bình có điều kiện

Khoảng tin cậy $(1-\alpha)100\%$ đối với trung bình có điều kiện $\mu_{Y|x_0}$ là:

$$\hat{y}_0 - t_{\alpha/2} s\sqrt{\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}} < \mu_{Y|x_0} < \hat{y}_0 + t_{\alpha/2} s\sqrt{\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}}$$

trong đó $t_{\alpha/2}$ là của giá trị của phân phối t với $n-2$ bậc tự do.

Ví dụ Sử dụng dữ liệu trong **Ví dụ trước**, hãy ước lượng khoảng tin cậy 95% cho kỳ vọng có điều kiện $\mu_{Y|x}$ khi $x_0 = 8$.

IV. Khoảng dự đoán cho giá trị biến y

Khoảng tin cậy $(1-\alpha)100\%$ đối với giá trị y_0 của biến đáp ứng (single repsonse) là:

$$\boxed{\hat{y}_0 - t_{\alpha/2} s\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}} < y_0 < \hat{y}_0 + t_{\alpha/2} s\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}}}$$

trong đó $t_{\alpha/2}$ là giá trị của phân phối t với $n-2$ bậc tự do

Ví dụ Sử dụng dữ liệu của bảng: Thước đo phần trăm chất rắn x (%) và nhu cầu oxy sinh hoá y (%)

x %	3	7	11	15	18	27	29	30	30	31
y %	5	11	21	16	16	28	27	25	35	30

x %	31	32	33	33	34	36	36	36	37	38
y %	40	32	34	32	34	37	38	34	36	38

x %	39	39	39	40	41	42	42	43	44	45	46	47	50
y %	37	36	45	39	41	40	44	37	44	46	46	49	51

Hãy tìm:

a. Phương trình đường hồi quy tuyến tính thực nghiệm và giá trị dự đoán y khi $x = 50$

b. Khoảng tin cậy 95% cho tham số β trong đường hồi quy $\mu_{Y|x} = \alpha + \beta x$

c. Khoảng tin cậy 95% cho α trong đường hồi quy $\mu_{Y|x} = \alpha + \beta x$

d Khoảng tin cậy 95% cho trung bình có điều kiện $\mu_{Y|x}$ tại $x_0 = 20\%$.

e. Khoảng dự đoán 95% cho y_0 khi $x_0 = 20\%$.