BÀI GIẢNG
TOÁN XÁC SUẤT VÀ THỐNG KÊ

GIỚI THIỆU HỌC PHẦN:

TOÁN XÁC SUẤT THỐNG KÊ

Số tín chỉ: 2TC
Thời lượng: 34 tiết
Lý thuyết: 26 tiết
Bài tập: 8 tiết

ĐÁNH GIÁ KẾT QUẢ HỌC TẬP

Cách đánh giá:

- Điểm quá trình: 30%;

- Điểm thi kết thúc HP: 70%

- Điểm quá trình bao gồm:

điểm kiểm tra giữa kỳ, điểm thảo luận, sửa bài tập trên lớp, . . .

GIỚI THIỆU HỌC PHẦN

Lý thuyết xác suất & thống kê
Tống Đình Quỳ (thư viện)

Giáo trình lý thuyết xác suất & thống kê toán học

Lý Hoàng Tú- Trần Tuấn Điệp

Và các tài liệu khác
Google ....

PHẦN 1

XÁC SUẤT

- Đến đầu thế kỷ 20, Andrei Nikovlaevich Kolmogorov đã thành công trong việc xây dựng cơ sở Toán học để trình bày khái niệm của lý thuyết xác suất, bằng việc xây dựng tiên đề cho lý thuyết xác suất.

Andrey Nikovlaevich Kolmogorov
(1903-1987)

CHƯƠNG 1: Các khái niệm về xác suất

1.1 – Biến cố- Không gian các biến cố

- **1.1.1 – Phép thử và biến cố**
- **a) Phép thử**
- Một thí nghiệm dùng để nghiên cứu một đại lượng hay một hiện tượng nào đó được gọi là phép thử. Ký hiệu một phép thử là $\mathcal{G}, \alpha, \beta$ ...
- **Ví dụ**: Mua ngẫu nhiên một vé số.
- **Ví dụ:** Gieo một con xúc sắc.
- **Ví dụ:** Tung một đồng xu.
- **b) Biến cố**

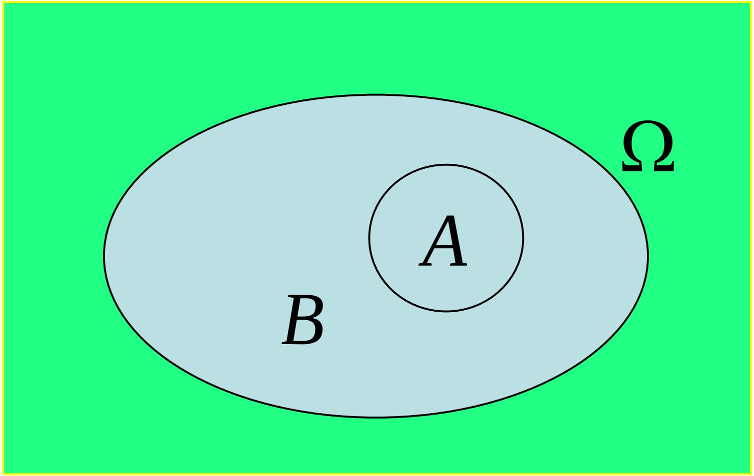
- – **Biến cố sơ cấp:** Kết quả đơn giản nhất có thể xảy ra khi thực hiện phép thử.
- **Ví dụ:**
  - Tung một đồng xu đồng chất cân đối ta thấy BCSC của phép thử này là: N, S.
  - Gieo một con xúc sắc đồng chất cân đối ta thấy các BCSC của phép thử này là: A_1, A_2, A_3, A_4, A_5, A_6
- Nếu $A_{i,}$ i=1,2, …6 là mặt i chấm xuất hiện

- – **Biến cố chắc chắn (tất yếu):** Là biến cố nhất định phải xảy ra khi phép thử được thực hiện. Ký hiệu là Ω (hoặc **U**).
- Ví dụ: Gieo một con xúc sắc đồng chất cân đối, biến cố số chấm xuất hiện nhỏ hơn 7 là biến cố chắc chắn.
- – **Biến cố không thể (bất khả):** Là biến cố không thể xảy ra khi phép thử được thực hiện. Ký hiệu là Ø (hoặc **V**).
- - **Biến cố ngẫu nhiên:** là biến cố có thể xảy ra hoặc không khi thực hiện phép thử.
- Tập hợp các biến cố sơ cấp của một phép thử còn gọi là không gian các biến cố sơ cấp và ký hiệu là Ω

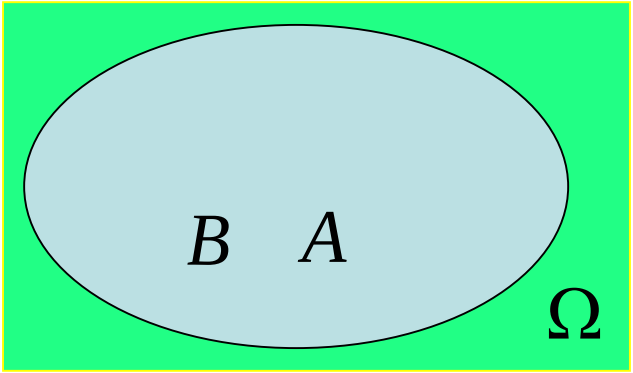
1.2 – Quan hệ giữa các biến cố

- Giả sử phép thử $\mathcal{G}$ có các biến cố A, B, C…

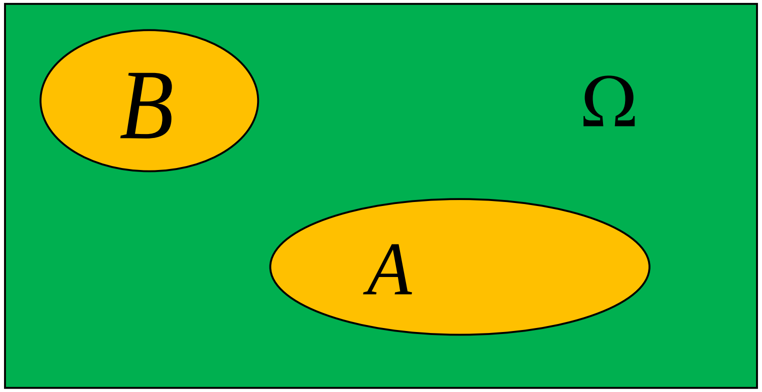
a) Kéo theo: Biến cố A được gọi là kéo theo biến cố B, ký hiệu là $A \subset B$ hoặc $A \Rightarrow B$ nếu A xảy ra thì B cũng xảy ra.



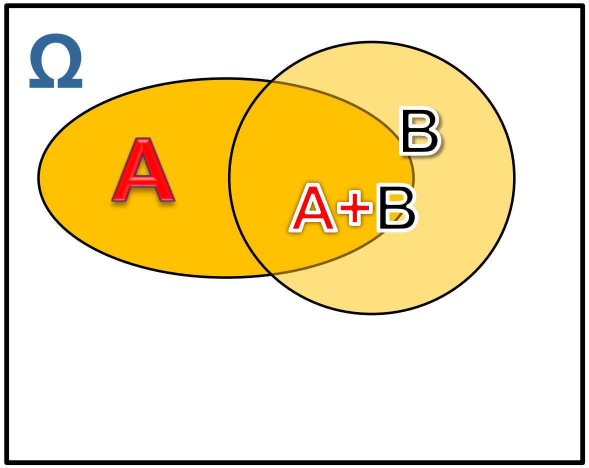
b) Tương đương: Biến cố A và B được gọi là hai *biến cố tương đương*, ký hiệu là $A = B$ nếu $A \subset B$ và $B \subset A$.



- **c) Xung khắc:** A, B là 2 biến cố xung khắc nếu chúng không thể đồng thời xảy ra khi thực hiện phép thử.



- **d) Tổng 2 biến cố:** Tổng của 2 biến cố A và B là một biến cố, ký hiệu là A+B (hoặc $A \cup B$). Biến cố này xảy ra khi và chỉ khi có ít nhất một trong hai biến cố A, B xảy ra.
- **Thí dụ:** Quan sát 2 xạ thủ cùng bắn vào một bia. Mỗi xạ thủ bắn một viên. Gọi A là biến cố "*xạ thủ thứ nhất bắn trúng bia*", B là biến cố "*xạ thủ thứ hai bắn trúng bia*", C là biến cố "*bia trúng đạn*" thì C = A + B



- Tổng của n biến cố $A_1, A_2, \ldots, A_n$ trong cùng một phép thử là một biến cố C ký hiệu là
- $C = A_1 + A_2 + \ldots + A_n$. Biến cố này xảy ra khi và chỉ khi có ít nhất một biến cố A_i nào đó xảy ra khi phép thử được thực hiện.

- **f) Tích của 2 biến cố:** Tích của hai biến cố A và B là một biến cố, ký hiệu là AB (hoặc $A \cap B$), biến cố này xảy ra khi và chỉ khi cả A và B xảy ra.

Ví dụ: Xét phép thử quan sát hai xạ thủ cùng bắn vào một bia (mỗi người bắn một viên). Gọi A là biến cố "*xạ thủ thứ nhất bắn trượt*", B là biến cố "*xạ thủ thứ hai bắn trượt*" và C là biến cố " *bia không trúng đạn*".

- Thì: C = AB.

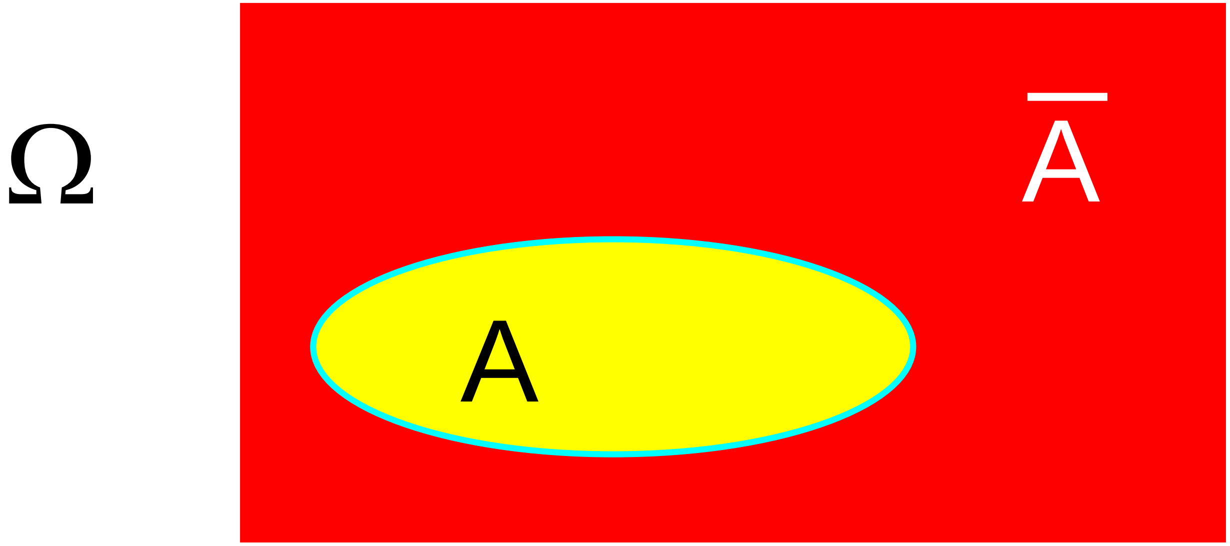
- Tích của n biến cố $A_1, A_2, \ldots, A_n$ trong cùng một phép thử là một biến cố C ký hiệu là

$$C = A_1 A_2 \ldots A_n$$

Biến cố này xảy ra khi và chỉ khi tất cả các biến cố A_i đều xảy ra khi phép thử được thực hiện.

g) Biến cố đối lập: Biến cố đối lập với biến cố A, ký hiệu là $\bar{A}$, nếu A, $\bar{A}$ xung khắc và $A \cup \bar{A} = \Omega$.

Biến cố đối



Thí dụ: Kiểm tra 5 sản phẩm. Gọi A là biến cố "có ít nhất 3 sản phẩm tốt", $\bar{A}$ là biến cố "số sản phẩm tốt không quá 2".

1.3- Định nghĩa xác suất cổ điển.

a)Định nghĩa

Giả sử phép thử $\mathcal{G}$ có n biến cố sơ cấp đồng khả năng; A là biến cố trong cùng phép thử và có m kết quả thuận lợi cho A (nghĩa là số khả năng xảy ra biến cố A)

- Ta gọi tỉ số $\frac{m}{n}$ là xác suất của biến cố A và ký hiệu là P(A)

$$\boxed{P(A) = \frac{\text{số khả năng thuận lợi cho } A}{\text{Tổng số khả năng}} = \frac{m}{n}}$$

Thí dụ: Tung một con xúc sắc cân đối và đồng chất, tính xác suất mặt xuất hiện có số chấm chia hết cho 3?

Giải:

Các trường hợp đồng khả năng là: xúc sắc ra mặt 1, xúc sắc ra mặt 2, . . . , xúc sắc ra mặt 6, vậy n = 6.

Gọi A là biến cố mặt xuất hiện có số chấm chia hết cho 3, số khả năng A xảy ra là 2 khả năng

Vậy xác suất của A là $P(A) = \frac{2}{6}$.

b- Các tính chất của xác suất:

1. Nếu A là biến cố ngẫu nhiên thì:

$$0 \leq P(A) \leq 1$$

2. Nếu Ω là biến cố chắc chắn thì:

$$P(\Omega) = 1$$

3. Nếu ∅ là biến cố không thể thì:

$$P(\varnothing) = 0$$

Với B là biến cố bất kỳ, ta luôn có:

$$0 \leq P(B) \leq 1$$

4. Nếu A, B xung khắc thì

$$P(A+B) = P(A) + P(B).$$

5. $\overline{A}$ là biến cố đối của A thì

$$P(\overline{A}) = 1 - P(A)$$

6. Nếu $A \Rightarrow B$ thì $P(A) \leq P(B)$

1.2.2- Định nghĩa thống kê

Xét phép thử $\mathcal{G}$ và A là một biến cố.

Giả sử ta có thể thực hiện phép thử $\mathcal{G}$ n lần khi đó biến cố A xuất hiện m lần, và tỷ số $\frac{m}{n}$ gọi là tần suất xuất hiện của biến cố A. Cho số phép thử tăng lên vô hạn khi đó tần suất xuất hiện của biến cố A sẽ tiến tới một giá trị xác định gọi là xác suất của A, ký hiệu P(A)

$$\boxed{P(A) = \lim_{n\to\infty} \frac{m}{n}}$$

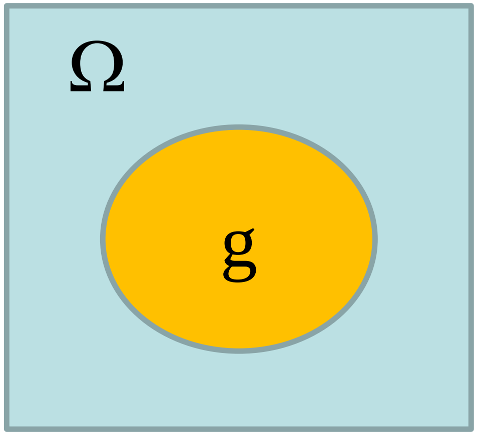
Ví dụ:

Người thí nghiệm	Số lần tung	Số lần sấp	Tần suất
Buffon	4040	2048	0.5080
Pearson	12000	6019	0.5016
Pearson	24000	12012	0.5005

1.2.3- Định nghĩa hình học

Xét một phép thử đồng khả năng, không gian mẫu có vô hạn phần tử và được biểu diễn thành một miền hình học Ω có độ đo xác định (độ dài, diện tích, thể tích…). Biến cố $A \subset \Omega$ được biểu diễn bởi miền hình học g. Khi đó, xác suất xảy ra A

$$\boxed{P(A) = \frac{mes(g)}{mes(\Omega)} = \frac{\text{Độ đo miền A}}{\text{Độ đo miền } \Omega}}$$



Ví dụ: Hai người X,Y hẹn gặp nhau ở một địa điểm trong khoảng thời gian 1 giờ (8h-9h sáng) và quy ước mỗi người đến điểm hẹn chờ người kia không quá 20 phút, nếu không thấy người kia sẽ đi về. Tìm xác suất để 2 người gặp được nhau.

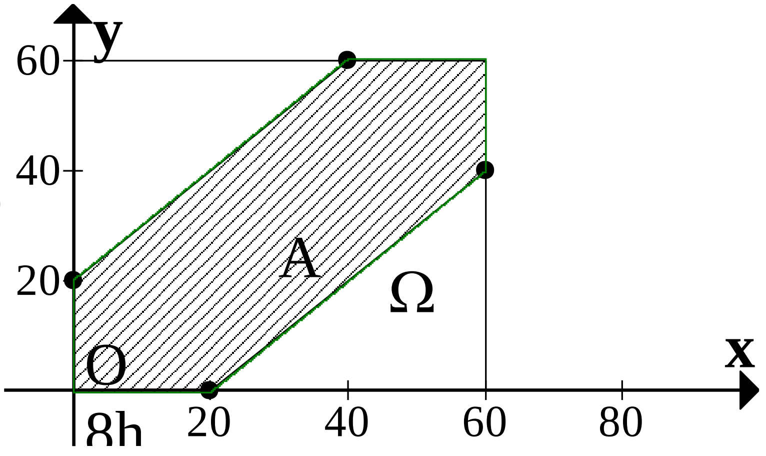
Giải: Gọi x, y là thời gian người X, Y đến địa điểm hẹn tính từ gốc 8h, (phút), ta có $0 \le x \le 60; 0 \le y \le 60$

$$\Omega = \{(x; y) \in \mathbb{R}^2 | 0 \le x, y \le 60\}$$

Gọi A là biến cố 2 người gặp nhau. Ta có

$$A = \{(x; y) \in \mathbb{R}^2 | 0 \le |x - y| \le 20\}$$

$$P(A) = \frac{S(A)}{S(\Omega)} = \frac{60^2 - 40^2}{60^2} = \frac{5}{9}$$



1.3- Các công thức tính xác suất

1.3.1- Công thức cộng xác suất:

❶ Nếu A và B là hai biến cố xung khắc thì:

$$P(A + B) = P(A) + P(B)$$

Tổng quát:

Nếu $A_1, A_2, \ldots, A_n$ là n biến cố xung khắc từng đôi, thì:

$$P(A_1 + A_2 + \ldots + A_n) = P(A_1) + P(A_2) + \ldots + P(A_n)$$

***Hệ quả*:** Nếu A và $\overline{A}$ là hai biến cố đối lập nhau thì:

$$\boxed{P(A) = 1 - P(\overline{A})}$$

Biến cố A độc lập với biến có B khi biến cố này không ảnh hưởng đến biến cố kia và ngược lại. Nếu A, B độc lập thì: $A, \bar{B}$; $\bar{A}$, B và $\bar{A}, \bar{B}$ cũng độc lập.

Nếu 2 biến cố A, B độc lập thì

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

❷ Công thức cộng

Nếu A và B là hai biến cố bất kỳ thì:

$$P(A + B) = P(A) + P(B) - P(AB)$$

Trường hợp n = 3: Nếu A_1, A_2, A_3 là các biến cố không xung khắc, thì:

$$\boxed{\begin{aligned} P(A_1 + A_2 + A_3) &= P(A_1) + P(A_2) + P(A_3) \\ &\quad - P(A_1A_2) - P(A_1A_3) - P(A_2A_3) + P(A_1A_2A_3) \end{aligned}}$$

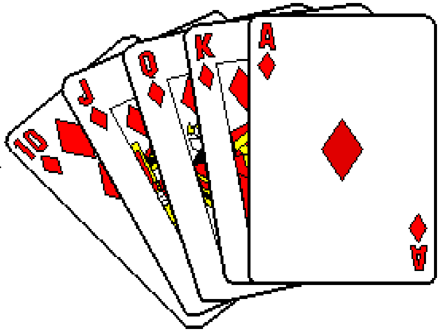
Ví dụ:

Hai xạ thủ mỗi người bắn một phát vào 1 tấm bia, xác suất bắn trúng của người thứ nhất và người thứ 2 lần lượt là 0,8 và 0,7.

a) Tính xác suất để cả 2 người bắn trúng.
b) Tính xác suất để có ít nhất một người bắn trúng.
c) Tính xác suất để có đúng một phát trúng đích.

1.3.2 Xác suất có điều kiện –Định lý nhân xác suất

a) Xác suất có điều kiện



Ví dụ. Một bộ bài tây có 52 lá, rút ngẫu nhiên 1 quân lần 1, sau đó lại rút ngẫu nhiên quân, gọi A là biến cố rút được 1 quân át, B là biến cố rút được 1 quân át ở lần thứ 2, ta có $P(B/A) = \frac{3}{51}; P(B/$

a) Định nghĩa:

Xác suất của biến cố A được tính với điều kiện biến cố B đã xảy ra gọi là xác suất có điều kiện của A khi biết chắc chắn **B đã xảy ra**, ký hiệu là P(A/B).

b) Định lý nhân xác suất.

Định lý 1: Nếu A, B là 2 biến cố bất kỳ ta có:

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

Ta có nếu P(A/B)=P(A) thì A, B độc lập, nếu không thì A,B phụ thuộc.

Định lý 2: Nếu A, B là 2 biến cố độc lập khi và chỉ khi

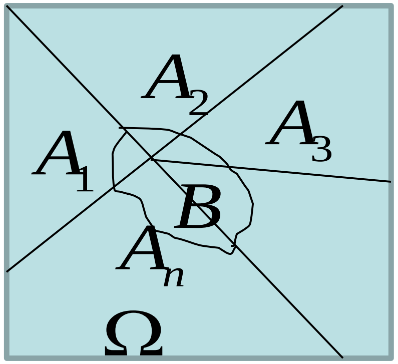
$$\boxed{P(AB) = P(A).P(B)}$$

Ví dụ: Một ngăn đựng 3 trục loại I và 7 trục loại II, người lắp máy chọn ngẫu nhiên 1 chiếc, sau đó lại rút tiếp chiếc thứ 2, tính xác suất để chiếc thứ nhất thuộc loại I và chiếc thứ 2 thuộc loại II.

1.3.3- Công thức xác suất toàn phần (đầy đủ)- Công thức Bayes

a) Hệ đầy đủ các biến cố (nhóm đầy đủ)

Cho không gian mẫu Ω và $A_1, A_2, \ldots, A_n$. Các biến cố $A_1, A_2, \ldots, A_n$ là hệ biến cố đầy đủ nếu chúng thỏa mãn 2 điều kiện sau:



$$\begin{cases} A_1 + A_2 + \ldots + A_n = \Omega \\ A_i A_j = \varnothing \quad \forall 1 \le i \ne j \le n \end{cases}$$

b) **Công thức xác suất toàn phần**

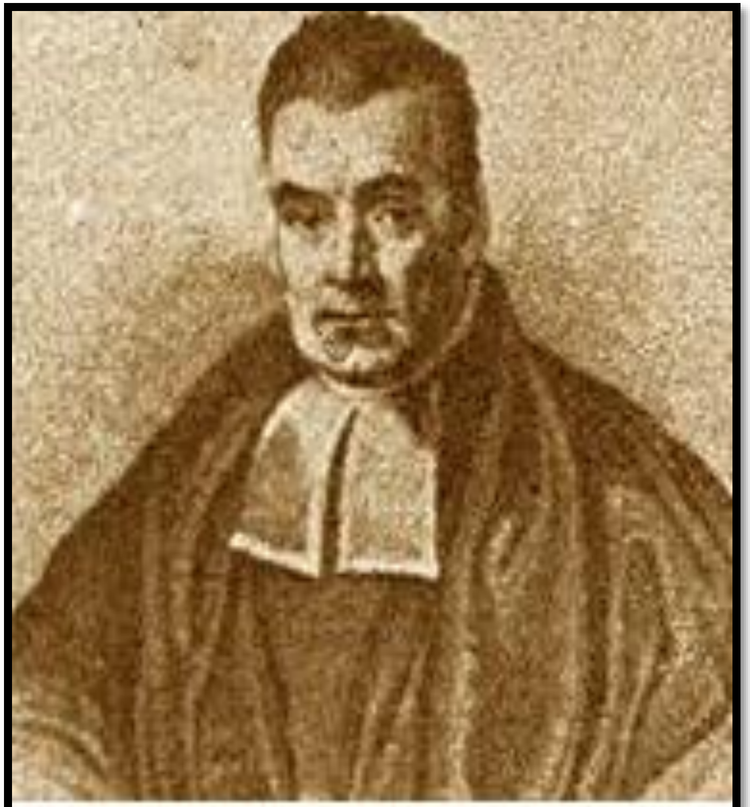
Giả sử ta có hệ biến cố đầy đủ $A_1, A_2, \ldots, A_n$, B là biến cố bất kỳ, ta có

$$P(B) = \sum_{i=1}^{n} P(A_i) P(B/A_i)$$

Ví dụ:

Một nhà máy có 4 phân xưởng sản xuất các chi tiết máy. Phân xưởng I sản xuất 40% tổng sản phẩm và có tỉ lệ phế phẩm là 5%, phân xưởng II sản xuất 20% tổng sản phẩm và có tỉ lệ phế phẩm là 2%, phân xưởng III sản xuất 30% tổng sản phẩm và có tỉ lệ phế phẩm là 3%, phân xưởng IV sản xuất 10% tổng sản phẩm và có tỉ lệ phế phẩm là 4%. Hãy xác định tỉ lệ phế phẩm chung của nhà máy.

- **c) Công thức Bayes.**



Thomas Bayes (1702–1761)

Giả sử ta có nhóm đầy đủ A_1, A_2, . . . , A_n, B là biến cố bất kỳ, phép thử được thực hiện, ta có

$$\boxed{P(A_i / B) = \frac{P(A_i)P(B / A_i)}{P(B)}}, \quad i = 1, 2, 3 \ldots n$$

Ví dụ: Quay lại ví dụ trên, giả sử một người mua ngẫu nhiên được một phế phẩm, tính xác suất để phế phẩm là của phân xưởng I sản xuất.

Các xác suất $P(A_i/B)$ được xác định sau khi đã biết kết quả của phép thử là B đã xảy ra nên thường được gọi là các *xác suất hậu nghiệm*.

Công thức Bayes xác định lại các *xác suất tiên nghiệm* $P(A_i)$ khi biết thông tin là B xảy ra.

1.3.4 Dãy các phép thử độc lập-Công thức Bernoulli

a)Dãy các phép thử độc lập

Một dãy các phép thử gọi là độc lập với nhau nếu mọi biến cố trong phép thử này độc lập với mọi biến cố trong phép thử khác.

b) Phép thử đơn (phép thử Bernoulli).

Phép thử G gọi là phép thử đơn nếu trong phép thử chỉ xét 2 biến cố A và $\bar{A}$.

c) Phép thử lặp (dãy phép thử Bernoulli)

Lặp đi lặp lại n lần phép thử đơn (phép thử Bernoulli) ta có phép thử lặp.

d) Công thức Bernoulli

Bernoulli (1700-1782)

d) Công thức Bernoulli

Cho một dãy n phép thử Bernoulli với xác suất biến cố A xuất hiện trong mỗi phép thử bằng nhau và bằng p.

Khi đó xác suất để biến cố A trong phép thử đó xuất hiện đúng k lần được tính theo công thức:

$$\boxed{P_n(k) = C_n^k p^k q^{n-k} \quad (0 \le k \le n)}; (1) \quad q = 1 - p$$

(1) được gọi là công thức Bernoulli

Hệ quả: Xác suất để biến cố A xuất hiện đúng k lần, $0 \le k_1 \le k \le k_2 \le n$:là

$$\boxed{P_n(k_1; k_2) = \sum_{k=k_1}^{k_2} C_n^k p^k q^{n-k}}; (2) \quad q = 1 - p$$

Ví dụ:

Một xạ thủ bắn 5 viên đạn vào một mục tiêu, xác suất bắn trúng mục tiêu một viên là 0,8.

a) Tính xác suất để có đúng 2 phát trúng đích

b) Tính xác suất để số phát trúng đích từ 2 đến 4.

1.3.5 Số lần xuất hiện chắc chắn nhất.

Bài toán: Tìm k_0 sao cho $P_n(k_0) = \max\limits_{0 \le k \le n} P_n(k)$

- Nếu np-q ≥ 0

❖ Nếu np-q $\in \mathbb{Z}$ $thì\ k_0 = np - q\ hoặc\ k_0$=np-q+1.

❖ Nếu np-q $\notin \mathbb{Z}$ $thì\ k_0 = [np - q] + 1$

- Nếu np-q<0 thì $k_0 = 0$.

1.3.6. Công thức Moivre-Laplace

Định lý 1. Nếu trong mỗi phép thử độc lập biến cố A xuất hiện với xác suất p (0<p<1) thì:

$$\boxed{\lim_{n\to\infty}\left(P_n(k)-\frac{1}{\sqrt{2\pi npq}}e^{-\frac{(k-np)^2}{2npq}}\right)=0}$$

Vậy với n đủ lớn $P_n(k)\approx\frac{1}{\sqrt{npq}}\varphi(x_0),\ x_0=\frac{k-np}{\sqrt{npq}}$

Với $\varphi(x_0)$ là hàm Gauss được xác định bởi $\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$

Chú ý Hàm Gauss là hàm chẵn, bảng tra từ 0 đến 3,99 nên ta có

$+\varphi(-x)=\varphi(x)$

+ x>3,99 $\varphi(x)\approx\varphi(3{,}99)$

Ví dụ:

Một xạ thủ bắn 200 viên đạn vào một mục tiêu, xác suất bắn trúng mục tiêu một viên là 0,8.

Tính xác suất để xạ thủ bắn trúng 80 phát.

Định lý 2. Nếu trong mỗi phép thử độc lập biến cố A xuất hiện với xác suất p (0<p<1) thì:

$$\boxed{\lim_{n\to\infty}\left(P_n(k_1;k_2)-\frac{1}{\sqrt{2\pi}}\int_{x_1}^{x_2} e^{-\frac{t^2}{2}}dt\right)=0}$$

Vậy với n đủ lớn $\boxed{P_n(k_1;k_2)=\Phi(x_2)-\Phi(x_1)}$

Với $x_1=\frac{k_1-np}{\sqrt{npq}}$; $x_2=\frac{k_2-np}{\sqrt{npq}}$; $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{t^2}{2}}$dt.

Φ(x) gọi là hàm Laplace.

Chú ý +Bảng 2: x<0, Φ(-x) = −Φ(x)

x>5,0; Thì Φ(x) = Φ(5,0)

Ví dụ:

Một xạ thủ bắn 200 viên đạn vào một mục tiêu, xác suất bắn trúng mục tiêu một viên là 0,8.

Tính xác suất để xạ thủ bắn trúng 80 đến 160 phát.

TÓM TẮT CHƯƠNG 1

Phép thử

Biến cố

Xác suất của biến cố.

- ĐN cổ điển
- ĐN thống kê
- ĐN hình học
- Các công thức cơ bản

- Các loại biến cố
- Mối quan hệ

Điều kiện áp dụng

Chương 2

BIẾN NGẪU NHIÊN (BNN) MỘT CHIỀU

2.1 – Biến ngẫu nhiên một chiều

2.2 – Luật phân phối xác suất của biến ngẫu nhiên

2.3 – Các đặc trưng của biến ngẫu nhiên

2.4 – Một số quy luật phân phối thường gặp

2.1 - BIẾN NGẪU NHIÊN MỘT CHIỀU

2.1.1 Định nghĩa.

- Cho phép thử có không gian mẫu Ω. Ánh xạ X từ Ω vào tập số thực R được gọi là một biến ngẫu nhiên.

$$X: \Omega \longrightarrow R, \qquad (X(\omega) \in R, \omega \in \Omega)$$

Tập $X(\Omega) = \{X(\omega) \in R: \omega \in \Omega\}$ là tập giá trị của X.

Các biến ngẫu nhiên ký hiệu bằng các chữ in hoa như X, Y, Z,…, các giá trị của biến ngẫu nhiên ký hiệu bằng chữ in thường như x, y, z, …

Các thí dụ:

❶ Kiểm tra 3 sản phẩm và quan tâm đến *số sản phẩm đạt tiêu chuẩn* có trong 3 sản phẩm kiểm tra.

❷ Khảo sát điểm thi môn toán xác suất thống kê của một sinh viên hệ chính qui và quan tâm đến *điểm thi* của sinh viên này.

❸ Khảo sát doanh thu của một siêu thị trong một ngày và quan tâm đến *doanh thu* (triệu đồng) của siêu thị.

2.1.2 – PHÂN LOẠI

a) Biến ngẫu nhiên rời rạc:

Biến ngẫu nhiên X gọi là BNN rời rạc nếu tập giá trị của X là hữu hạn hoặc vô hạn đếm được.

Ví dụ 1: Gieo một con xúc sắc, gọi X là số chấm xuất hiện của con xúc sắc.

Ví dụ 2: X là số con trai trong một gia đình có hai con.

b) Biến ngẫu nhiên liên tục

Biến ngẫu nhiên X gọi là BNN liên tục nếu tập giá trị của nó lấp đầy một khoảng trên trục số.

Ví dụ : X là trọng lượng của một trẻ sơ sinh.

2.2 Luật phân phối xác suất của biến ngẫu nhiên

2.2.1 Bảng phân phối xác suất của BNN rời rạc

Bảng phân phối xác suất biến ngẫu nhiên rời rạc X có dạng

X	x_1	x_2	…	x_n
P_i	p_1	p_2	…	p_n

Trong đó biến ngẫu nhiên X có tập giá trị $X(\Omega) = \{x_1, x_2, \ldots, x_n\}$ và xác suất tại các điểm là $P(X = x_i) = p_i, i = 1,2, \ldots n.$

Ví dụ:

Một xạ thủ có 3 viên đạn, anh ta bắn từng phát một cho tới khi trúng đích hoặc hết đạn thì thôi, Gọi X là số viên đạn anh ta bắn, cho biết xác suất bắn trúng 1 viên là 0,6. Lập bảng phân phối xác suất của X.

Tính chất:

1) $p_i \in [0;1]$

2) $\sum_{i=1}^{n} p_i = 1$

2.2.2 Hàm mật độ (Dùng cho biến ngẫu nhiên liên tục)

Định nghĩa:

Hàm $f(x)$ gọi là hàm mật độ nếu thỏa mãn:

1, Tập xác định R. $(D_f = R)$

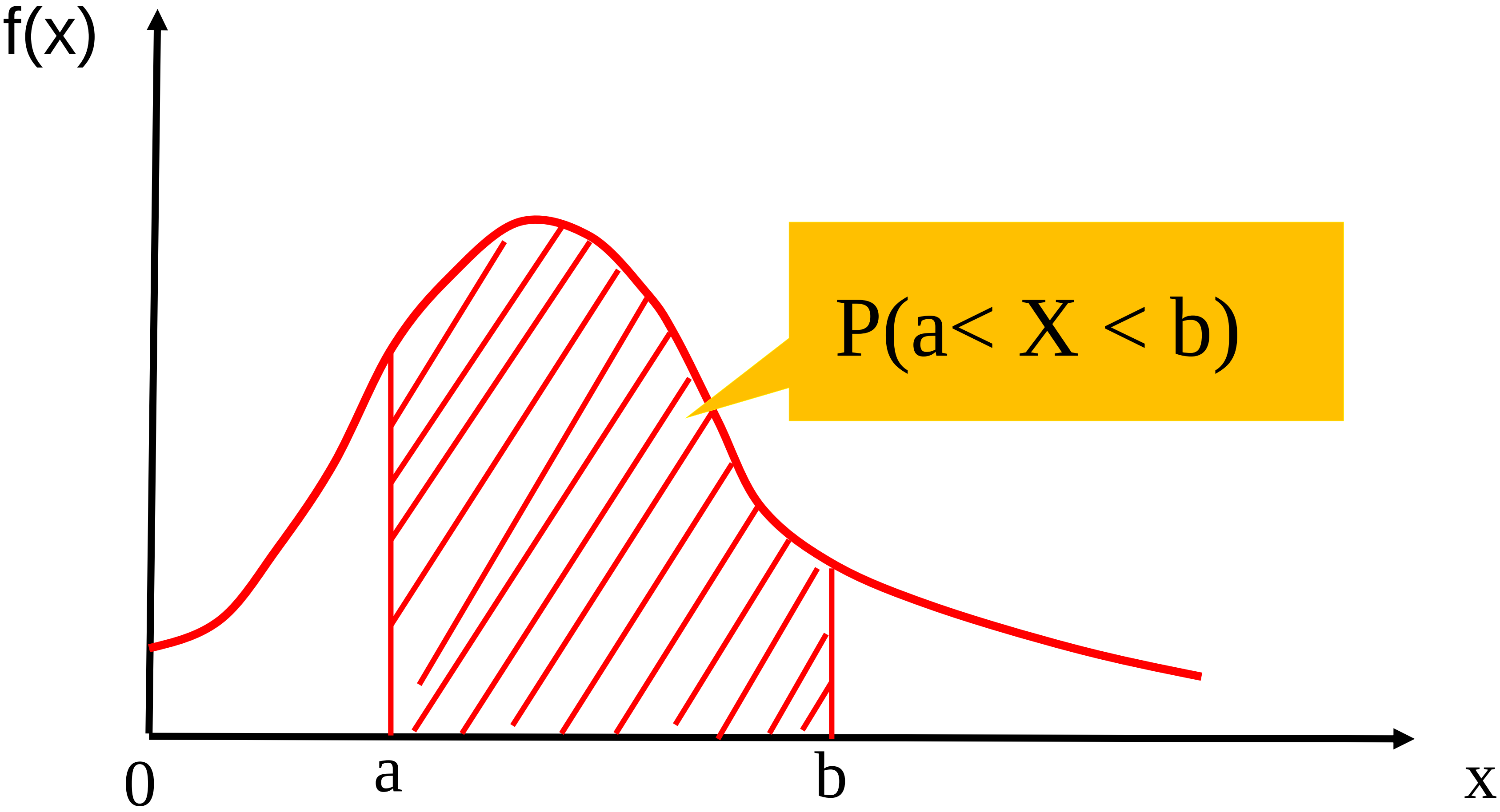
2, $f(x) \geq 0, \forall x$

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

Tính chất $P(a < X < b) = P(a < X \leq b) =$
$P(a \leq X < b) = P(a \leq X \leq b) = \int_a^b f(x)dx$

Chú ý

Đối với biến ngẫu nhiên liên tục X thì P(X = a) = 0.


f(x)
P(a< X < b)
0
a
b
x

Ví dụ

Cho biến ngẫu nhiên liên tục X có hàm mật độ

$$f(x) = \begin{cases} 0 & khi \quad x \notin [0;3] \\ ax^2 & khi \quad x \in [0;3] \end{cases}$$

a, Tìm giá trị của a.

b. Tính P(1<X<5).

c. Tính P(X>2).

2.2.3. Hàm phân phối(phân bố) xác suất

a) Định nghĩa

Hàm F(x) xác định trên R gọi là hàm phân phối của biến ngẫu nhiên X nếu **F(x) = P(X<x).**

Ví dụ

Cho biến ngẫu nhiên X có bảng phân phối xác suất

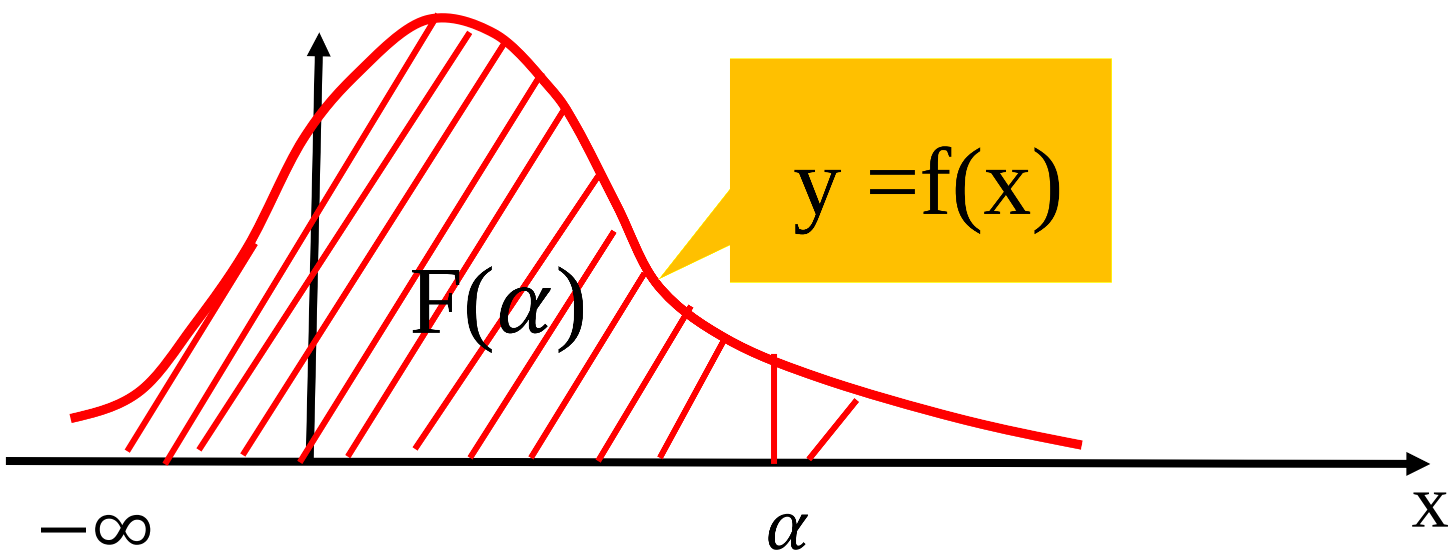
X	0	1	2
P	0,5	0,3	0,2

Tìm hàm phân phối xác suất của X.

b) Liên hệ giữa hàm mật độ và hàm phân phối

$$F(x) = P(X < x) = P(-\infty < X < x) = \int_{-\infty}^{x} f(x)dx$$

$$P\ (a < X < b) = \int_a^b f(x)dx = \mathrm{F(b)} - \mathrm{F(a)}$$



c) Tính chất hàm phân phối

i, $F(x) \in [0;1], F(-\infty) = 0, F(+\infty) = 1$

ii, F(x) là hàm không giảm nghĩa là $x_1 < x_2$ thì $F(x_1) \leq F(x_2)$.

iii, $P(a \leq X < b) = F(b) - F(a)$

iv, Nếu X là biến ngẫu nhiên liên tục thì F(x) liên tục trên R

F’(x) = f(x)

v, F(x) liên tục trái tại mọi $\alpha \in \mathbb{R}$, tức là $\lim\limits_{x \to \alpha^-} F(x) = F(\alpha)$

d- *Ý nghĩa của hàm phân phối xác suất:*

Hàm F(x) phản ánh mức độ tập trung xác suất về phía bên trái của điểm x. Giá trị của hàm F(x) cho biết có bao nhiêu phần của một đơn vị xác suất phân phối trong khoảng ($-\infty$, x).

Ví dụ:

Cho biến ngẫu nhiên liên tục X có hàm mật độ

$$f(x) = \begin{cases} \frac{1}{2} cosx & nếu\ x \in [-\frac{\pi}{2}; \frac{\pi}{2}] \\ 0 & nếu\ x \notin [-\frac{\pi}{2}; \frac{\pi}{2}] \end{cases}$$

a) Tìm xác suất để X nhận giá trị thuộc $(\boldsymbol{\frac{\pi}{4}}; \boldsymbol{\pi})$.

b) Tìm hàm phân phối xác suất của X.

2.3 CÁC ĐẶC TRƯNG CỦA BIẾN NGẪU NHIÊN

2.3.1. Kỳ vọng

Định nghĩa 1.

Cho biến ngẫu nhiên rời rạc X có bảng phân phối xác suất

X	x_1	x_2	…	x_n
P_i	p_1	p_2	…	p_n

Kỳ vọng của X ký hiệu là EX được xác định bởi

$$EX = x_1p_1 + x_2p_2 + \cdots + x_np_n = \sum_{i=1}^{n} x_ip_i$$

Định nghĩa 2.

Cho X là biến ngẫu nhiên liên tục X có hàm mật độ f(x),

tích phân $\int_{-\infty}^{+\infty} xf(x)$ Hội tụ tuyệt đối", kỳ vọng của X

ký hiệu là EX và được xác định bởi:

$$\boxed{EX = \int_{-\infty}^{+\infty} xf(x)dx}$$

Ví dụ

Cho X là biến ngẫu nhiên có bảng phân phối

X	0	1	2	3
P	1/8	3/8	3/8	1/8

Tính EX?

Ví dụ

Cho X có hàm mật độ

$$f(x) = \begin{cases} 0 & khi \quad x \notin [0;3] \\ \frac{1}{9}x^2 & khi \quad x \in [0;3] \end{cases}$$

Tính EX.

Tính chất

i, $E(C) = C$, $C=$ const

ii, $E(CX) = CEX$

iii, $E(X \pm Y) = EX \pm EY$

iv, X, Y là 2 biến ngẫu nhiên độc lập thì $E(XY) = EX.EY$

v, Y = g(X) thì

$$EY = Eg(X) = \begin{cases} \displaystyle\sum_{i=1}^{n} g(x_i)p_i & khi\ X\ rời\ rạc \\ \displaystyle\int_{-\infty}^{+\infty} g(x)f(x)dx & khi\ X\ liên\ tục \end{cases}$$

2.3.2. Phương sai

Định nghĩa

Phương sai của biến ngẫu nhiên X ký hiệu là VX (VarX) hoặc DX,

được xác định bởi: $\mathrm{V}(X) = E(X - EX)^2$

$$VX = E(X - EX)^2 = \begin{cases} \displaystyle\sum_{i=1}^{n} (x_i - EX)^2 p_i \quad khi\ X\ rời\ rạc \\ \displaystyle\int_{-\infty}^{+\infty} (x - EX)^2 f(x)dx \quad khi\ X\ liên\ tục \end{cases}$$

Trong thực tế người ta còn tính VX theo công thức:

$$\boxed{VX = EX^2 - (EX)^2}$$

Tương tự như trên ta có:

$$VX = EX^2 - (EX)^2$$

$$= \begin{cases} \sum_{i=1}^{n} {x_i}^2 p_i - (EX)^2 \quad khi\ X\ rời\ rạc \\ \int_{-\infty}^{+\infty} x^2 f(x)dx - (EX)^2 \quad khi\ X\ liên\ tục \end{cases}$$

Tính chất

i, $V(C) = 0$, $C=$ const; $VX \geq 0, \forall X$

ii, $V(CX) = C^2\, VX$

iii, X, Y là 2 biến ngẫu nhiên độc lập $V(X \pm Y) = VX + VY$

2.4. Một số quy luật phân phối thường gặp

a) Quy luật phân phối chuẩn $N(a; \sigma^2)$

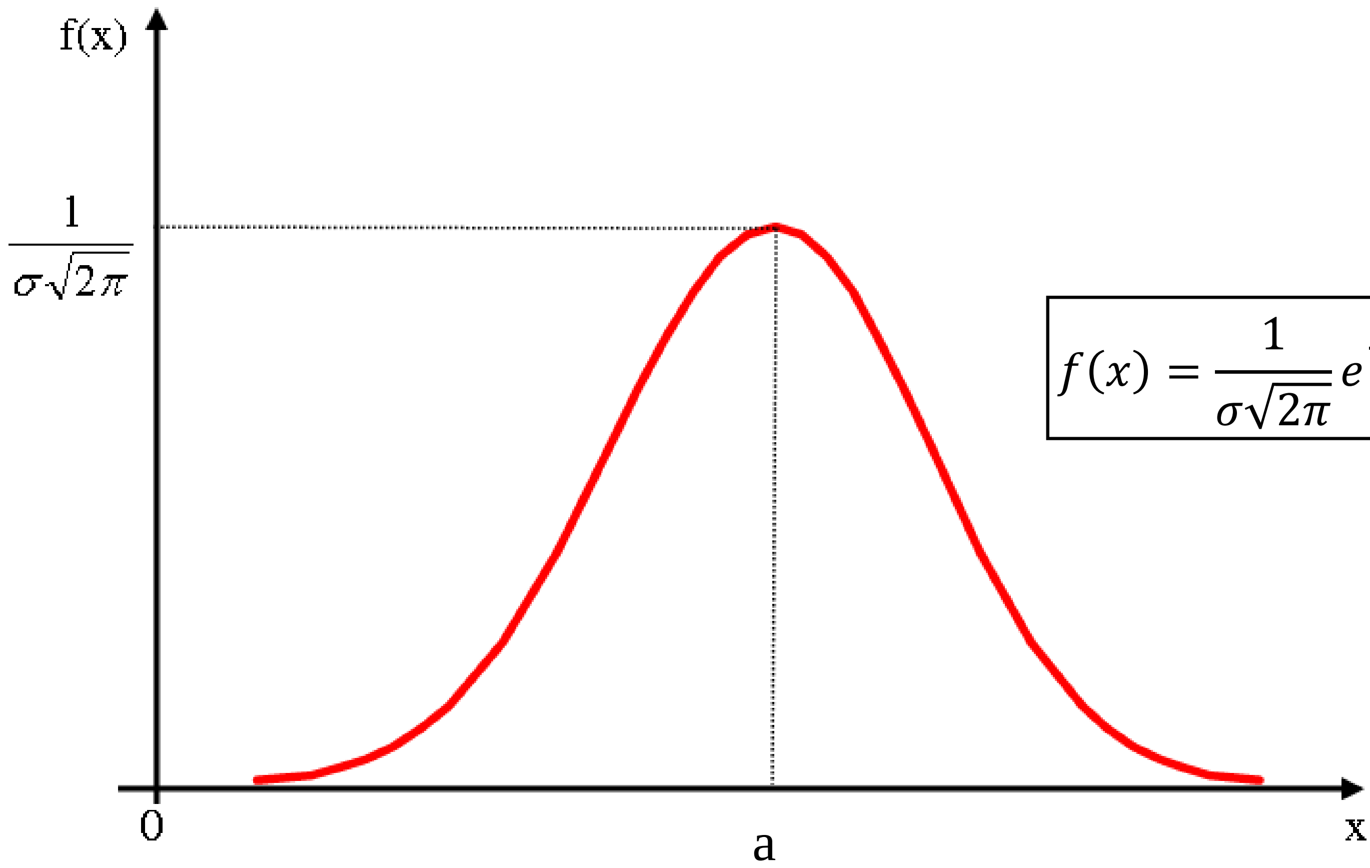
Định nghĩa

Biến ngẫu nhiên X gọi là có quy luật chuẩn với tham số a và σ^2 nếu hàm mật độ có dạng

$$\boxed{f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}}$$

Ký hiệu $X \sim N(a; \sigma^2)$

Khi đó i, $EX = \text{a}$ ii, $VX = DX = \sigma^2$.



$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$$

Trường hợp đặc biệt khi $X \sim N(0;1)$ thì X gọi là có quy luật chuẩn tắc, khi đó hàm mật độ gọi là hàm Gauss có dạng

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Khi đó hàm phân phối của quy luật chuẩn tắc có dạng

$$\boxed{\mathrm{F}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt = \frac{1}{2} + \Phi(x)}$$

- **Định lý**

Cho X là ĐLNN có phân phối chuẩn $X \in N(a; \sigma^2)$

- Ta có công thức:

$$P(\alpha < X < \beta) = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right)$$

$$\text{Trong đó}: \Phi(x) = \int_0^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

Ví dụ 1:

- Cho X có phân phối chuẩn N(2100; 200^2)
- Tính P(1700 < X < 2200)

Ví dụ 2:

Người ta sản xuất một chi tiết máy có độ dài theo quy định là a= 20 cm, cho biết độ dài là BNN X $\sim N(a;\ \sigma^2)$. Tính xác suất để độ dài chi tiết được sản xuất ra lệch so với độ dài quy định không quá 0,3; biết σ =0,2

b) Phân phối Bernoulli

Định nghĩa.

Biến ngẫu nhiên rời rạc X gọi là phân phối Bernoulli với tham số p nếu có bảng phân phối xác suất Ký hiệu quy luật không-một là X~B(1;p).

X	0	1
P	1-p	p

Khi đó

i, $EX = p$

ii, $VX = p(1 - p)$

c) Quy luật nhị thức B(n;p)

Định nghĩa

Biến ngẫu nhiên X gọi là có quy luật nhị thức với tham số n ∈ N và p∈[0;1] nếu nó có tập giá trị {0, 1, 2, …, n} và:

$$P(X = k) = C_n^k p^k q^{n-k}, k = 0,1, \dots, n \text{ và } q = 1 - p$$

Biến ngẫu nhiên X có quy luật nhị thức ký hiệu là X~B(n;p).

Khi đó

i, $EX = np$

ii, $VX = npq$

d) Quy luật Poisson P(λ)

Định nghĩa

Biến ngẫu nhiên X gọi là có quy luật Poisson với tham số $\lambda > 0$ nếu nó có tập giá trị là tập N và

$$P(X = k) = e^{-\lambda}\frac{\lambda^{k}}{k!}, k \in N$$

Khi đó ký hiệu X~P(λ). Khi đó

i, $EX = \lambda$

ii, $VX = \lambda$

Chương 3

BIẾN NGẪU NHIÊN NHIỀU CHIỀU

3.1 – Biến ngẫu nhiên hai chiều

3.2 – Phân phối xác suất của biến ngẫu nhiên 2 chiều rời rạc

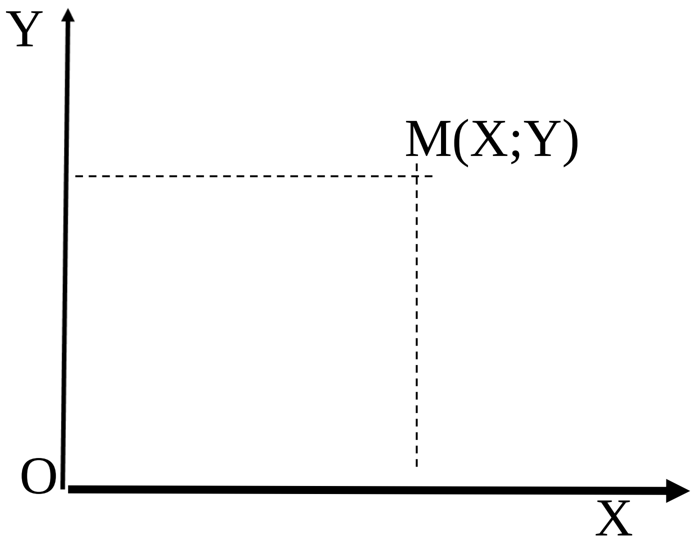
3.3 – Phân phối xác suất của biến ngẫu nhiên liên tục

3.1 - BIẾN NGẪU NHIÊN HAI CHIỀU

a) ĐN: Ta gọi BNN $\xi(X;Y)$ một véc tơ biến ngẫu nhiên 2 chiều có các thành phần X, Y là các biến ngẫu nhiên một chiều tương ứng

Chú ý : * X, Y là biến ngẫu nhiên rời rạc thì $\xi(X;Y)$ là BNN rời rạc.

* X, Y là biến ngẫu nhiên Liên tục thì $\xi(X;Y)$ là BNN liên tục.



b) Hàm phân phối

❶ ĐN: Ta gọi hàm phân phối xác suất của BNN 2 chiều là

$$\boxed{F(x;y) = P(X < x; Y < y),\ \forall x, y \in R}$$

Trong đó (X<x; Y<y) là biến cố đồng thời X< x; Y<y.

❷Tính chất hàm phân phối

a) $0 \le F(x;y) \le 1$.

b) $F(x;y)$ là hàm không giảm đối với từng biến số.

c) $F(x;y)$ liên tục trái đối với từng biến số

$$\lim_{x \to x_1^-} F(x;y) = F(x_1;y)$$

d) $F(-\infty;y) = F(x;-\infty) = 0; F(+\infty;+\infty) = 1$

e) $F_1(x) = F(x;\ +\infty); F_2(y) = F(+\infty;y)$.

3.2 – PHÂN PHỐI XÁC SUẤT CỦA BNN 2 CHIỀU RỜI RẠC

a) Bảng phân phối xác suất của BNN 2chiều rời rạc

X \ Y	y_1	y_2	…………	y_j	…………	y_m
x_1	p_{11}	p_{12}	…………	p_{1j}	…………	p_{1m}
x_2	p_{21}	p_{22}	…………	p_{2j}	…………	p_{2m}
………	………	………	………	………	………	
x_i	p_{i1}	p_{i2}	…………	p_{ij}	…………	p_{im}
………						
x_n	p_{n1}	p_{n2}	…………	p_{nj}	…………	p_{nm}

trong đó $p_{ij} = P\left(X = x_i; Y = y_j\right)$

b) Tính chất + $p_{ij} \geq 0, \forall i = \overline{1, n}\,; \forall j = \overline{1, \mathrm{m}}$

$$+ \sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} = 1.$$

c) Chú ý:

$$\boxed{F(x,y) = P(X < x; Y < y) = \sum_{\substack{i,j \\ x_i < x; y_j < y}} p_{ij}}$$

Ví dụ:

Bảng phân phối xác suất của BNN 2 chiều rời rạc

X \ Y	**1**	**2**
1	0,15	0,35
2	0,20	0,05
3	0,10	0,15

a) Tìm luật phân phối của BNN X, BNN Y?

b) Tìm F(2; 3).

c) Tìm F(x;y)?

Thí dụ: Cho biết bảng phân phối xác suất của BNN 2 chiều (X, Y), trong đó X là doanh thu và Y là chi phí quảng cáo của các công ty tư nhân kinh doanh cùng một mặt hàng như sau: (đơn vị tính của X và Y đều là triệu đồng/tháng). Tìm bảng phân phối xác suất của từng biến ngẫu nhiên, tính kỳ vọng và phương sai của từng biến ngẫu nhiên X, Y?

Y \ X	100	150	200
0	0,1	0,05	0,05
1	0,05	0,2	0,15
2	0	0,1	0,3

3.3 Phân phối xác suất của biến ngẫu nhiên liên tục

+) Định nghĩa: Giả sử rằng véc tơ NN $\xi(X;Y)$ liên tục (các BNN thành phần liên tục) có hàm phân phối xác suất là F(x;y) liên tục khắp nơi và có đạo hàm cấp 2 $\frac{\partial^2 F}{\partial x \partial y}$; $\frac{\partial^2 F}{\partial y \partial x}$ cũng là hàm liên tục khắp nơi. Khi đó hàm mật độ đồng thời của BNN $\xi(X;Y)$là

$$\boxed{f(x;y) = \frac{\partial^2 F(x;y)}{\partial x \partial y}}$$

c)Tính chất hàm mật độ

+)Tính chất

$$1)\ f(x;y) \geq 0$$

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

+)Mối liên hệ giữa hàm phân phối và hàm mật độ xác suất của BNN 2 chiều

$$F(x;y) = P(X < x; Y < y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(x;y)dxdy$$

Hệ Quả:

$$P\left(\left(X;Y\right)\in D\right)=\iint\limits_{D} f(x;y)dxdy$$

Liên hệ giữa hàm mật độ và hàm phân bố xác suất của BNN thành phần

$$F_1(x) = P(X < x; Y < +\infty) = \int_{-\infty}^{x}\int_{-\infty}^{+\infty} f(x;y)dxdy$$

$$f_1(x) = F'_1(x) = \int_{-\infty}^{+\infty} f(x,y)dy$$

Ví dụ:

Hàm mật độ của BNN 2chiều liên tục

$$f(x,y)=\begin{cases} a(x+y) \text{ nếu } (x,y)\in[0,1]\times[0,1] \\ 0 \quad\quad \text{nếu trái lại} \end{cases}$$

a) Tìm a?

b) Tìm $f_1(x)$; $f_2(y)$.

c) Tìm $P(0<X<1/2;0<Y<1/2)$?

Ví dụ:

Hàm mật độ của BNN 2chiều liên tục

$$f(x,y) = \begin{cases} a\,xy & \text{nếu } 0 < x < y < 1 \\ 0 & \text{nếu trái lại} \end{cases}$$

a) Tìm a?

b) Tìm $f_1(x)$; $f_2(y)$.

Ví dụ:

Hàm mật độ của BNN 2chiều liên tục

$$f(x,y)=\begin{cases} a(x+y)\ nếu\ 0<x<y<1 \\ 0 \quad\quad nếu\ trái\ lại \end{cases}$$

a) Tìm a?

b) Tìm $f_1(x)$; $f_2(y)$.

c) Tìm EX, EY?

Phần 2: Thống kê
Chương 4
MẪU THỐNG KÊ

4.1 – Một số khái niệm về mẫu.

4.2 – Mẫu

4.3 – Các phương pháp chọn mẫu

4.4– Các phương pháp sắp xếp số liệu

4.5 Các đặc trưng của mẫu

Chương 4: Mở đầu về Thống kê

4.1. Một số khái niệm về mẫu.

1 .Tổng thể: Khái niệm: Tập hợp tất cả các phần tử để nghiên cứu theo 1 dấu hiệu nghiên cứu nào đó gọi là tổng thể. Số phần tử của tổng thể được gọi là kích thước N của nó. Đại lượng ngẫu nhiên đặc trưng cho dấu hiệu nghiên cứu gọi là đại lượng ngẫu nhiên gốc X.

Dấu hiệu nghiên cứu được chia ra làm 2 loại: Định lượng và định tính.

-Định lượng: $E(X) = a, V(X) = \sigma^2$

-Định tính: $E(X) = p, V(X) = p.q$

Gọi a là trung bình tổng thể , p là tỉ lệ tổng thể

σ^2 gọi là phương sai tổng thể

σ gọi là độ lệch tổng thể

Chú ý: Định tính là trường hợp riêng của định lượng với hai lượng là 0 và 1. Cho nên p là trường hợp riêng của a, còn p.q là trường hợp riêng của σ^2

4.2-Mẫu:

Từ tổng thể lấy ngẫu nhiên ra n phân tử để nghiên cứu được gọi là lấy một mẫu kích thước n.

Định nghĩa: Từ đại lượng ngẫu nhiên gốc X, xét n đại lượng ngẫu nhiên độc lập có cùng phân phối với X.Véc tơ ngẫu nhiên n chiều $W=(X_1, X_2...X_n)$ được gọi là một mẫu kích thước n. Thực hiện phép thử ta nhận được $w=(x_{1,} x_2 ... x_n)$ là giá trị cụ thể hay giá trị thực hành mẫu W.

4.3 – Các phương pháp chọn mẫu

- 1) Mẫu cơ học: Chia tập nền thành n tập nhỏ sau đó chọn từ mỗi tập một phần tử làm đại diện
- 2) Mẫu điển hình: Chia tập nền thành n tập nhỏ có tính điển hình. Sau đó chọn từ mỗi tập một phần tử làm đại diện
- 3) Mẫu dãy: Chia tập nền thành n dãy sau đó chọn từ mỗi dãy một phần tử làm đại diện.

- 4) Chọn mẫu ngẫu nhiên:
- Các phần tử được chọn một cách ngẫu nhiên làm đại diện
- Chọn mẫu có suy luận:đự trên ý kiến các chuyên gia hoặc các nhà phân tích về vấn đề đó.

4.4– Các phương pháp sắp xếp số liệu

- 1) Sắp xếp theo giá trị quan sát:

X^*	x_1	x_2	…	x_k
n_i	n_1	n_2	…	n_k

Ta gọi bảng này là bảng tần số của mẫu

- 2) Sắp xếp số liệu dạng khoảng:

X^*	$x_1' - x_1''$	$x_2' - x_2''$	…	$x_k' - x_k''$
n_i	n_1	n_2	…	n_k

Khi đó để lập bảng tần số ta chọn $x_i = (x_i' + x_i'')/2$

- 3) Bảng tần suất

X^*	x_1	x_2	…	x_k
f_i	f_1	f_2	…	f_k

Ta gọi bảng này là bảng tần suất của mẫu $f_i = n_i/n$

- 4)Mẫu đơn giản $(x_1; x_2; x_3; \ldots .x_n)$

4.5 Các đặc trưng của mẫu (số liệu cụ thể dung ký hiệu chữ nhỏ)

Cho mẫu ngẫu nhiên ($X_1;X_2;....X_n$), mẫu cụ thể có kích thước n ($x_1;x_2;....x_n$),

Trung bình mẫu: $\overline{X} = \dfrac{\sum_{i=1}^{n} X_i}{n}$ với mẫu cụ thể $\overline{x} = \dfrac{\sum_{i=1}^{k} n_i x_i}{n} = \dfrac{\sum_{i=1}^{n} x_i}{n}$

Phương sai mẫu: $\hat{S}^2 = \dfrac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \dfrac{1}{n}\sum_{i=1}^{k} n_i\left(X_i - \overline{X}\right)^2$

Mẫu cụ thể ta có:

$$\hat{s}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 = \frac{1}{n}\sum_{i=1}^{k} n_i\left(x_i - \overline{x}\right)^2$$

- Phương sai mẫu hiệu chỉnh với mẫu ngẫu nhiên:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \frac{1}{n-1}\sum_{i=1}^{k} n_i\left(X_i - \overline{X}\right)^2$$

Với mẫu cụ thể ta có $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 = \frac{1}{n-1}\sum_{i=1}^{k} n_i\left(x_i - \overline{x}\right)^2$

Khi đó ta có : $S^2 = \frac{n}{n-1}\mathrm{S}^2; s^2 = \frac{n}{n-1}s^2$

- Tỉ lệ mẫu: $f_n = f_n(A) = \frac{m}{n}$

- Độ lệch mẫu ngẫu nhiên và mẫu cụ thể là $\hat{S} = \sqrt{\hat{S}^2}; \hat{s} = \sqrt{\hat{s}^2}$

- Độ lệch mẫu hiệu chỉnh ngẫu nhiên và mẫu cụ thể $S = \sqrt{S^2}; s = \sqrt{s^2}$

- **Ví dụ:**
- Cho bảng số liệu:

x_i	4	5	7	9
n_i	10	15	13	12

Tính trung bình mẫu và phương sai mẫu, phương sai mẫu hiệu chỉnh, độ lệch mẫu, độ lệch mẫu hiệu chỉnh.

Phương pháp sử dụng máy tính bỏ túi F_x570 MS

- **Bước 1:** Vào chương trình thống kê (SD)
- Mod – SD – xóa dữ liệu có trong máy – vào lại SD
- **Bước 2:** Nhập số liệu
- x_i ; n_i – M^+ (i = 1…n)
- **Bước 3:** Đọc kết quả
- Gọi n; $\sum x_i$; $\sum x_i^2$: Bấm shift – sum
- Gọi $\bar{x};\ S\left(x\sigma n-1\right);\ \hat{S}\ \left(x\sigma n\right)$ Bấm shift - var

- **Bước 2:**
- Nhập số liệu theo cột – nhập xong bấm AC
- **Bước 3:**
- Đọc kết quả
- Gọi $\sum x_i$; $\sum x_i^2$: Bấm Sift – 1 – chọn 4
- Gọi n; xσn; xσn-1: Bấm Sift – 1 – chọn 5

Cách dùng máy tính bỏ túi ES

- Mở tần số(1 lần): Shift Mode ↓ Stat On(Off)
- Nhập: Mode Stat 1-var

x_i	n_i
48	20
49	15
50	25

AC: báo kết thúc nhập dữ liệu

Cách đọc kết quả: Shift Stat Var

$\overline{x} = 49,0833$

$x\sigma n = 0,8620$

$x\sigma n - 1 = 0,8693$

CHƯƠNG 5. ƯỚC LƯỢNG

5.1. Khái niệm chung về ước lượng.

5.2. Ước lượng khoảng của tỷ lệ tổng thể p.

5.3. Ước lượng khoảng của trung bình (kỳ vọng)

5.1 Khái niệm chung về ước lượng.

-Việc dùng kết quả của mẫu để đánh giá 1 tham số θ nào đó của tổng thể được gọi là ước lượng.

1.Ước lượng điểm:

Chọn G=G(W(x_1,x_2,…x_n) gọi là ước lượng điểm của tham số θ sau đó lấy $\theta \approx G$

Khi đo G goi là ước lượng

1.Không chệch nếu $E(G) = \theta$

2.Vững: $\lim\limits_{n\to\infty} G = \theta$

3.Hiệu quả: $V\left(G\right) = D(G) \to \min$

Kết quả: $a \approx \overline{x}$: có đủ bốn tính chất trên.

$p \approx f$: có đủ bốn tính chất trên.

$\sigma^2 \approx S^2$: là ước lượng không chệch.

2.Ước lượng khoảng:

Định nghĩa: khoảng(θ_1, θ_2) được gọi là khoảng ước lượng của tham số θ với độ tin cậy $\gamma = 1 - \alpha$ nếu:

$I = \theta_2 - \theta_1$ -độ dài khoảng ước lượng hay khoảng tin cậy.

$$\mathrm{P}\left(\theta_1 < \theta < \theta_2\right) = 1 - \alpha$$

Giải: Chọn $G(\mathrm{W},\theta)$ sao cho G có quy luật phân phối xác suất đã biết, tìm hai số g_1, g_2 sao cho

$$\mathrm{P}\left(g_1 < G < g_2\right) = 1 - \alpha$$

$$\Rightarrow g_1 < g\left(w,\theta\right) < g_2$$

$$\Leftrightarrow \theta_1 < \theta < \theta_2$$

5.2 Ước lượng khoảng của tỷ lệ tổng thể p.

Bài toán: Từ tổng thể lấy một mẫu kích thước n có tỷ lệ mẫu f. Với độ tin cậy γ ,hãy tìm khoảng tin cậy của p.

Giải: Chọn $$G = U = \frac{(f-p)\sqrt{n}}{\sqrt{f(1-f)}} \approx \mathrm{N}(0,1)$$

Xét $$\Rightarrow \mathrm{P}\left(u_{\alpha_1} < U < u_{1-\alpha_2}\right) = 1-\alpha$$

$$\Rightarrow u_{\alpha_1} < \frac{(f-p)\sqrt{n}}{\sqrt{f(1-f)}} < u_{1-\alpha_2}$$

$$\Leftrightarrow f - \frac{\sqrt{f(1-f)}}{\sqrt{n}}.\mathrm{u}_{\alpha_2} < p < f + \frac{\sqrt{f(1-f)}}{\sqrt{n}}.\mathrm{u}_{\alpha_1}$$

Ta xét ba trường hợp riêng

1. Ước lượng tối đa $p < f + \dfrac{\sqrt{f(1-f)}}{\sqrt{n}}.\mathrm{u}_{\alpha}$

2. Ước lượng tối thiểu $f - \dfrac{\sqrt{f(1-f)}}{\sqrt{n}}.\mathrm{u}_{\alpha} < p$

3. Ước lượng đối xứng $f - \varepsilon < p < f + \varepsilon$

với $\varepsilon = \dfrac{\sqrt{f(1-f)}}{\sqrt{n}}.\mathrm{u}_{\alpha/2}$ là độ chính xác (của ước lượng)

Độ dài khoảng tin cậy $I = 2\varepsilon$

Kích thước mẫu tối thiểu $n = \left[\dfrac{f(1-f)}{\varepsilon^2}.\mathrm{u}_{\alpha/2}{}^2\right] + 1$

$$\Phi.(\mathrm{u}_{\alpha}) = 0,5 - \alpha;\ \Phi.(\mathrm{u}_{\alpha/2}) = \frac{1-\alpha}{2} = \frac{\gamma}{2}$$

Quy ước: Nếu đề bài không nói rõ thì ta xét ước lượng đối xứng.

Ví dụ : Để điều tra số cá trong hồ lớn, một cơ quan quản lý đánh bắt 3000 con, làm dấu rồi thả xuống hồ, lần hai bắt ngẫu nhiên 4000 con thấy 540 con có dấu. Hãy xác định số cá trong hồ với đô tin cậy bằng 0,95.

Gọi N là số cá trong hồ

p là tỷ lệ cá bị đánh dấu trong hồ $p = \dfrac{3000}{N}$

$$n = 4000, m = 540 \to f = 0,135; \alpha = 1 - 0,95 = 0,05$$

$$\varepsilon = \frac{\sqrt{0,135.0,865}}{\sqrt{4000}}.\mathrm{u}_{0,025} = \frac{\sqrt{0,135.0,865}}{\sqrt{4000}}.1,96 = 0,0105$$

$$\to f - \varepsilon < p = \frac{3000}{N} < f + \varepsilon \Rightarrow ? < N < ?$$

$$0,1245 < p = \frac{3000}{N} < 0,1455; \quad 20618 < N < 24096$$

Ví dụ : Cần lập một mẫu ngẫu nhiên với kích thước bao nhiêu để tỷ lệ phế phẩm của mẫu là 0,2; độ dài khoảng tin cậy đối xứng là 0,02 và độ tin cậy là 0,95.

5.3 . Ước lượng khoảng của trung bình (kỳ vọng)

Bài toán: Từ tổng thể lấy một mẫu kích thước n có trung bình $\overline{x}$ mẫu và phương sai điều chỉnh mẫu s^2. Với độ tin cậy γ hãy tìm khoảng ước lượng của trung bình tổng thể.

Bài giải.Ta xét 3 trường hợp:

TH1. Đã biết phương sai tổng thể σ^2

Chọn $G = U = \dfrac{\left(\overline{X} - a\right)\sqrt{n}}{\sigma} \sim N(0,1)$

Xét $\alpha_{1,2} \geq 0; \alpha_1 + \alpha_2 = \alpha \Rightarrow -u_{2\alpha_1} < \dfrac{\left(\overline{x} - a\right)\sqrt{n}}{\sigma} < u_{2\alpha_2}$

1. Ước lượng trung bình tối đa $\boxed{a < \overline{x} + \frac{\sigma}{\sqrt{n}}.\mathrm{u}_{\alpha}}$

2. Ước lượng tối thiểu $\boxed{\overline{x} - \frac{\sigma}{\sqrt{n}}.\mathrm{u}_{\alpha} < a}$

3. Ước lượng Đối xứng $\boxed{\overline{x} - \varepsilon < a < \overline{x} + \varepsilon}$

với $\varepsilon = \frac{\sigma}{\sqrt{n}}.\mathrm{u}_{\alpha/2}$ là độ chính xác của khoảng tin cậy

Độ rộng tin cậy $I = 2\varepsilon$

Kích thước mẫu tối thiểu $\boxed{n = \left[\left(\frac{\sigma}{\varepsilon}.\mathrm{u}_{\alpha/2}\right)^2\right] + 1,}$

$$\Phi.(\mathrm{u}_{\alpha}) = 0,5 - \alpha;\ \Phi.(\mathrm{u}_{\alpha/2}) = \frac{1-\alpha}{2} = \frac{\gamma}{2}$$

TH2. Chưa biết phương sai σ^2 tổng thể mẫu lớn $n \geq 30$

1. Ước lượng trung bình tối đa $\boxed{a < \overline{x} + \frac{s}{\sqrt{n}}.\mathrm{u}_{\alpha}}$

2. Ước lượng tối thiểu $\boxed{\overline{x} - \frac{s}{\sqrt{n}}.\mathrm{u}_{\alpha} < a}$

3. Ước lượng Đối xứng $\boxed{\overline{x} - \varepsilon < a < \overline{x} + \varepsilon}$

với $\varepsilon = \frac{s}{\sqrt{n}}.\mathrm{u}_{\alpha/2}$ là độ chính xác của khoảng tin cậy

Độ rộng tin cậy $I = 2\varepsilon$

Kích thước mẫu tối thiểu $\boxed{n = \left[\left(\frac{s}{\varepsilon}.\mathrm{u}_{\alpha/2}\right)^2\right] + 1,}$

$$\Phi.(\mathrm{u}_{\alpha}) = 0,5 - \alpha;\ \Phi.(\mathrm{u}_{\alpha/2}) = \frac{1-\alpha}{2} = \frac{\gamma}{2}$$

TH3.Chưa biết phương sai tổng thể σ^2 **,** $n < 30$

Xét
$$G = T = \frac{\left(\overline{X} - a\right)\sqrt{n}}{S} \sim T_{n-1}$$

1. Ước lượng trung bình tối đa $\boxed{a < \overline{x} + \frac{s}{\sqrt{n}}.t_{n-1;\alpha}}$

2. Ước lượng tối thiểu $\boxed{\overline{x} - \frac{s}{\sqrt{n}}.t_{n-1;\alpha} < a}$

3. Ước lượng Đối xứng $\boxed{\overline{x} - \varepsilon < a < \overline{x} + \varepsilon}$

với $\varepsilon = \frac{s}{\sqrt{n}}.t_{n-1;\alpha/2}$ là độ chính xác của khoảng tin cậy

Độ rộng tin cậy $I = 2\varepsilon$

Kích thước mẫu tối thiểu $n = \left[\left(\frac{s}{\varepsilon}.t_{n-1;\alpha/2}\right)^2\right] + 1,$

$$\Phi.(\mathrm{u}_\alpha) = 0,5 - \alpha;\ \Phi.(\mathrm{u}_{\alpha/2}) = \frac{1-\alpha}{2} = \frac{\gamma}{2}$$

Ví dụ: Để ước lượng xăng hao phí trung bình cho 1 loại xe ô tô chạy trên đoạn đương từ A đến B ,chạy thử 49 lần trên đoạn đường này ta có bảng số liệu:

Lượng xăng hao phí(lit)	9,6-9,8	9,8-10,0	10,0-10,2	10,2-10,4	10,4-10,6
Số lần	4	8	25	8	4

Với độ tin cậy 0,95, hãy tìm khoảng tin cậy cho mức hao phí xăng trung bình của loại xe nói trên.

Chương 6. Lý thuyết kiểm định

6.1- Giả thuyết - Đối thuyết

6.2- Kiểm định tham số

6.1 Giả thuyết thống kê

Định nghĩa

• Giả thuyết là một mệnh đề (một câu khẳng định) về một vấn đề nào đó.

• Giả thuyết thống kê là giả thuyết về quy luật phân phối xác suất của biến ngẫu nhiên, về tham số đặc trưng của biến ngẫu nhiên hoặc về tính độc lập của biến ngẫu nhiên.

• Giả thuyết thống kê đưa ra thường được ký hiệu là H_0 (H).

Một mệnh đề trái với giả thuyết thống kê được gọi là đối thuyết thống kê, ký hiệu là H_1 (K).

Bài toán kiểm định một giả thuyết thống kê

Định nghĩa: Bài toán kiểm định giả thuyết thống kê là bài toán dựa trên cơ sở mẫu đã có, ta xây dựng một quy tắc chấp nhận hoặc bác bỏ một giả thuyết thống kê đã cho.

- Để thực hiện bài toán này, ta lập không gian mẫu. Trên không gian mẫu, ta xác định miền W là miền bác bỏ giả thuyết H_0 và phần bù của W là W là miền chấp nhận H_0 .
- Như vậy, nếu mẫu $(x_1,\ldots,x_n) \in W$ thì ta bác bỏ giả thuyết H_0. Nếu mẫu $(x_1,\ldots,x_n) \in W$ thì ta chấp nhận giả thuyết H_0.

Các loại sai lầm

Sai lầm loại 1 •

Là sai lầm mắc phải khi bác bỏ giả thuyết H_0 nhưng H_0 đúng.

- Xác suất sai lầm loại 1: P(bác bỏ H_0 / H_0 đúng) = $P(W/ H_0) = \alpha$

Sai lầm loại 2

- Là sai lầm mắc phải khi chấp nhận giả thuyết H_0 nhưng H_0 sai.
- Xác suất sai lầm loại 2: P(Chấp nhận H_0 / H_0 sai) = $P(\overline{W}/H_1) = \beta$

Nhận xét

- Cực tiểu 2 sai lầm. Nhưng với mẫu cụ thể thì không làm được.
- β phức tạp hơn.$\Rightarrow$ Cố định α, làm cực tiểu β
- α được gọi là mức ý nghĩa bài toán kiểm định.
- $1-\beta$ được gọi là lực lượng của bài toán kiểm định.

6.2 Kiểm định tham số

1. Kiểm định kỳ vọng của phân phối chuẩn.
2. Kiểm định xác suất (tỷ lệ).
3. Kiểm định sự bằng nhau của hai kỳ vọng.
4. Kiểm định sự bằng nhau của hai xác suất (hai tỷ lệ).

1.Kiểm định kỳ vọng của phân phối chuẩn

Bài toán : Giả sử, trong tổng thể, biến ngẫu nhiên X ~N(a;σ²).

Với mức ý nghĩa α cho trước, hãy kiểm định cặ giả thuyết, đối thuyết sau:

H_0: $a = a_0$ (a_0 đã biết)

H_1: $a \neq a_0$ (Đối thuyết hai phía)

H_1: $a > a_0$ (Đối thuyết một phía)

H_1: $a < a_0$ (Đối thuyết một phía)

- Từ tổng thể, ta chọn mẫu ngẫu nhiên, kích thước n:

$X_1,...,X_n$. $\Rightarrow X_i \sim N(a;\sigma^2)$.

- Chia ba trường hợp:
- Trường hợp: σ^2 đã biết.
- Trường hợp: σ^2 chưa biết nhưng cỡ mẫu $n \geq 30$.
- Trường hợp: σ^2 chưa biết nhưng cỡ mẫu $n<30$.

a)Trường hợp σ² đã biết:

- Xét thống kê $$Z=\frac{\left(\overline{X}-a_0\right)\sqrt{n}}{\sigma}$$ ~N(0;1) nếu H_0 đúng
- Với mẫu cụ thể, ta thực hiện bài toán kiểm định như sau:

Các bước thực hiện

B1: Tính $$Z_t=\frac{\left(\overline{x}-a_0\right)\sqrt{n}}{\sigma}$$

B2: Quyết định dựa vào bảng sau:

Quyết định H_1	$\boldsymbol{a \neq a_0}$	$\boldsymbol{a > a_0}$	$\boldsymbol{a < a_0}$
Bác bỏ H_0	$\lvert Z_t\rvert > u_{\frac{\alpha}{2}}$	$Z_t > u_\alpha$	$Z_t < -u_\alpha$
Chấp nhận H_0	$\lvert Z_t\rvert \leq u_{\frac{\alpha}{2}}$	$Z_t \leq u_\alpha$	$Z_t \geq -u_\alpha$

Ví dụ: Trọng lượng X các bao gạo là biến ngẫu nhiên tuân theo phân phối chuẩn N(50;0,01). Có ý kiến khách hàng phản ánh là trọng lượng bị thiếu. Một tổ thanh tra đã cân ngẫu nhiên 25 bao gạo từ lô bị nghi ngờ. Kết quả thu được $\bar{x}$= 49,27. Với mức ý nghĩa α = 0,05, có kết luận gì về ý kiến của khách hàng?

Ví dụ: Trọng lượng X gói mì ăn liền tuân theo luật phân phối chuẩn N(80;64). Lấy một mẫu ngẫu nhiên 25 gói mì ăn liền ta tìm được $\bar{x} = 82$. Với mức ý nghĩa 0,05 có thể cho rằng trọng lượng trung bình gói mì ăn liền là 80 gam được không?

b)Trường hợp σ^2 chưa đã biết: mẫu lớn n ≥ 30

- Xét thống kê $$Z = \frac{\left(\overline{X} - a_0\right)\sqrt{n}}{S}$$ ~N(0;1) nếu H_0 đúng
- Với mẫu cụ thể, ta thực hiện bài toán kiểm định như sau:

Các bước thực hiện

B1: Tính $$Z_t = \frac{\left(\overline{x} - a_0\right)\sqrt{n}}{s}$$

B2: Quyết định dựa vào bảng sau:

Quyết định H_1	$\boldsymbol{a \neq a_0}$	$\boldsymbol{a > a_0}$	$\boldsymbol{a < a_0}$
Bác bỏ H_0	$\lvert Z_t \rvert > u_{\frac{\alpha}{2}}$	$Z_t > u_\alpha$	$Z_t < -u_\alpha$
Chấp nhận H_0	$\lvert Z_t \rvert \leq u_{\frac{\alpha}{2}}$	$Z_t \leq u_\alpha$	$Z_t \geq -u_\alpha$

Trường hợp σ^2 chưa biết cỡ và cỡ mẫu n<30.

- Xét thống kê $Z = \dfrac{\left(\overline{X} - a_0\right)\sqrt{n}}{S}$ $\sim t_{n-1}$ nếu H_0 đúng
- Với mẫu cụ thể, ta thực hiện bài toán kiểm định như sau:

Các bước thực hiện

B1: Tính $Z_t = \dfrac{\left(\overline{x} - a_0\right)\sqrt{n}}{s}$

B2: Quyết định dựa vào bảng sau:

Quyết định H_1	$a \neq a_0$	$a > a_0$	$a < a_0$
Bác bỏ H_0	$\lvert Z_t \rvert > t_{n-1;\frac{\alpha}{2}}$	$Z_t > t_{n-1;\alpha}$	$Z_t < -t_{n-1;\alpha}$
Chấp nhận H_0	$\lvert Z_t \rvert \leq t_{n-1;\frac{\alpha}{2}}$	$Z_t \leq t_{n-1;\alpha}$	$Z_t \geq -t_{n-1;\alpha}$

Chú ý : Trong trường hợp cỡ mẫu n > 30, ta có thể sử dụng u_α thay vì $t_{n-1;\alpha}$ (hai giá trị này xấp xỉ nhau khi n > 30).

Ví dụ: Lương X (triệu/tháng) của người làm công nghệ thông tin (CNTT) Việt Nam tuân theo phân phối chuẩn $N(14;\sigma^2)$. Có ý kiến cho rằng thực tế thì lương của người làm CNTT cao hơn. Người ta điều tra ngẫu nhiên 40 người, thu được kết quả $\bar{x} = 18$ và độ lệch chuẩn s = 2,5. Với mức ý nghĩa 0,1 có kết luận gì về ý kiến trên?

2.Kiểm định tỉ lệ

Bài toán : Giả sử, trong tổng thể, gọi p là xác suất để một cá thể có tính chất A. Với mức ý nghĩa α cho trước, hãy kiểm định cặp giả thuyết, đối thuyết sau:

H_0: $p = p_0$ (p_0 đã biết)

H_1: $p \neq p_0$ (Đối thuyết hai phía)

H_1: $p > p_0$ (Đối thuyết một phía)

H_1: $p < p_0$ (Đối thuyết một phía)

- **Lấy mẫu ngẫu nhiên kích thước n.**
- Gọi f là tần suất cá thể đặc tính A trong mẫu: $f = \frac{n_A}{n}$

Xét thống kê $Z_t = \frac{f - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$, khi n đủ lớn thì $Z_t \sim N(0;1)$ nếu H_0 đúng.

Ta có các bước thực hiện như sau:

B1: Tính $Z_t = \frac{f - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$

B2: Quyết định dựa vào bảng sau:

Quyết định H_1	**p≠ p_0**	**p> p_0**	**p< p_0**
Bác bỏ H_0	$\|Z_t\| > u_{\frac{\alpha}{2}}$	$Z_t > u_\alpha$	$Z_t < -u_\alpha$
Chấp nhận H_0	$\|Z_t\| \leq u_{\frac{\alpha}{2}}$	$Z_t \leq u_\alpha$	$Z_t \geq -u_\alpha$

Ví dụ: Một máy sản xuất tự động với tỷ lệ chính phẩm là 0,98. Sau một thời gian làm việc, người ta kiểm tra ngẫu nhiên 500 sản phẩm do máy chế tạo thì thấy có 30 phế phẩm. Với mức ý nghĩa 0,05 hãy xem:

a. Chất lượng máy có thay đổi không?

b. Nếu chất lượng có thay đổi thì thay đổi như thế nào?

a. Ta có cặp giả thuyết đối thuyết :

H_0: $p = 0,98$

H_1: $p \neq 0,98$

• Tần suất chính phẩm trong mẫu: $f = \dfrac{500 - 30}{500} = 0,94.$

• Tính $Z_t = \dfrac{f - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} = \dfrac{0,94 - 0,98}{\sqrt{\dfrac{0,98.0,02}{500}}} = -6,3888$

• $\alpha = 0,05$, $\Phi(u_{0,025}) = 0,5 - 0,025 = 0,475 = \Phi(1,96) \Rightarrow$
$u_{0,025} = 1,96.$

• So sánh: $|Z_t| > u_{0,025} \Rightarrow$ bác bỏ H_0, hay chất lượng của máy đã thay đổi so với trước.

b. Ta có cặp giả thuyết đối thuyết :

H_0: $p = 0,98$

H_1: $p < 0,98$

- $u_{0,05} = 1,65$
- So sánh: $Z_t < -u_{0,05} \Rightarrow$ bác bỏ H_0, hay chất lượng của máy đã giảm so với trước.

3. Kiểm định sự bằng nhau của hai kỳ vọng

Bài toán : Giả sử, trong hai tổng thể có hai biến ngẫu nhiên tương ứng X ~$N(a_1;\sigma_1^2)$, Y ~$N(a_2;\sigma_2^2)$.;Với mức ý nghĩa α cho trước, hãy kiểm định cặp giả thuyết, đối thuyết sau:

H_0: $a_1 = a_2$

H_1: $a_1 \neq a_2$ (Đối thuyết hai phía)

H_1: $a_1 > a_2$ (Đối thuyết một phía)

H_1: $a_1 < a_2$ (Đối thuyết một phía)

Từ hai tổng thể lấy ra hai mẫu ngẫu nhiên có kích thước tương ứng n, m:

$X=(X_1,\ldots,X_n;) \Rightarrow Xi \sim N(a_1;\sigma_1^2)\ \forall i = 1,2,\ldots;n.$

$Y=(Y_1,\ldots,Y_m) \Rightarrow Y_j \sim N(a_2;\sigma_2^2)\ \forall j = 1;m.$

Xét TH1: đã biết $\sigma_1^2;\ \sigma_2^2$

- Xét thống kê $Z_t = \dfrac{\bar{X}-\bar{Y}}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}}$; $\bar{X}$; $\bar{Y}$ là trung bình của mẫu X và Y nếu H_0

đúng thì $Z_t \sim N(0;1)$

- Với 2 mẫu cụ thể $(x_1,\ldots,x_n)$, $(y_1,\ldots,y_m)$, ta thực hiện bài toán kiểm định

như sau:

B1: Tính $Z_t = \dfrac{\bar{x} - \bar{y}}{\sqrt{\dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{m}}}$

B2: Quyết định dựa vào bảng sau:

Quyết định H_1	$\mathbf{a_1 \neq a_2}$	$\mathbf{a_1 > a_2}$	$\mathbf{a_1 < a_2}$
Bác bỏ H_0	$\lvert Z_t \rvert > u_{\frac{\alpha}{2}}$	$Z_t > u_\alpha$	$Z_t < -u_\alpha$
Chấp nhận H_0	$\lvert Z_t \rvert \leq u_{\frac{\alpha}{2}}$	$Z_t \leq u_\alpha$	$Z_t \geq -u_\alpha$

Xét TH2: chưa biết σ_1^2; σ_2^2, m,n $\geq$30

- Xét thống kê $Z_t = \dfrac{\bar{X}-\bar{Y}}{\sqrt{\frac{S_1^2}{n}+\frac{S_2^2}{m}}}$; $\bar{X}$; $\bar{Y}$ là trung bình của mẫu X và Y, S_1^2; S_2^2

là phương sai mẫu hiệu chỉnh của X, Y, nếu H_0 đúng thì $Z_t \sim N(0;1)$

- Với 2 mẫu cụ thể $(x_1,\ldots,x_n)$, $(y_1,\ldots,y_m)$, ta thực hiện bài toán kiểm định như sau:

B1: Tính $Z_t = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{S_1^2}{n}+\frac{S_2^2}{m}}}$

B2: Quyết định dựa vào bảng sau:

Quyết định H_1	**$a_1 \neq a_2$**	**$a_{1>}a_2$**	**$a_{1<}a_2$**
Bác bỏ H_0	$\lvert Z_t \rvert > u_{\frac{\alpha}{2}}$	$Z_t > u_\alpha$	$Z_t < -u_\alpha$
Chấp nhận H_0	$\lvert Z_t \rvert \leq u_{\frac{\alpha}{2}}$	$Z_t \leq u_\alpha$	$Z_t \geq -u_\alpha$

Ví dụ: Để so sánh trọng lượng con rạ (sinh từ lần thứ hai trở đi) và trọng lượng con so (sinh lần đầu) qua thống kê ở một nhà hộ sinh ta được kết quả sau:

Trọng lượng(g)	1700-2000	2000-2300	2300-2600	2600-2900	2900-3200
Số con rạ	9	13	18	42	18
Số con so	5	10	22	40	45

Với mức ý nghĩa 0,05 có thể coi trọng lượng con so lớn hơn con rạ được không?

Xét TH3: chưa biết σ_1^2; σ_2^2, m,n <30, biết $\sigma_1^2 = \sigma_2^2$

- Xét thống kê $Z_t = \dfrac{\bar{X}-\bar{Y}}{\sqrt{\dfrac{(n-1)S_1^2+(m-1)S_2^2}{m+n-2}\left(\dfrac{1}{n}+\dfrac{1}{m}\right)}}$; $\bar{X}$; $\bar{Y}$ là trung bình của mẫu X và Y, S_1^2; S_2^2 là phương sai mẫu hiệu chỉnh của X, Y, nếu H_0 đúng thì $Z_t \sim t_{n+m-2}$
- Với 2 mẫu cụ thể $(x_1,\dots,x_n)$, $(y_1,\dots,y_m)$, ta thực hiện bài toán kiểm định như sau:

B1: Tính $Z_t = \dfrac{\bar{X} - \bar{Y}}{\sqrt{\dfrac{(n-1)S_1^2 + (m-1)S_2^2}{m+n-2}\left(\dfrac{1}{n} + \dfrac{1}{m}\right)}}$

B2: Quyết định dựa vào bảng sau:

Quyết định H_1	$\mathbf{a_1 \neq a_2}$	$\mathbf{a_1 > a_2}$	$\mathbf{a_1 < a_2}$
Bác bỏ H_0	$\lvert Z_t \rvert > t_{n+m-2;\frac{\alpha}{2}}$	$Z_t > t_{n+m-2;\alpha}$	$Z_t < -t_{n+m-2;\alpha}$
Chấp nhận H_0	$\lvert Z_t \rvert \leq t_{n+m-2;\frac{\alpha}{2}}$	$Z_t \leq t_{n+,m-2;\alpha}$	$Z_t \geq -t_{n+m-2;\alpha}$

Ví dụ: Điều tra thu nhập (Trăm Dollar/tháng) của 8 CEO công nghệ thông tin miền Bắc và 10 CEO miền Nam có kết quả sau:

X	40	38	40	42	44	41	36	39		
Y	41	44	38	42	40	45	39	37	43	41

Biết rằng X và Y tuân theo phân phối chuẩn với cùng phương sai. Với mức ý nghĩa 0,05 có thể cho rằng trung bình thu nhập của các CEO 2 miền là như nhau được không?

4. Kiểm định sự bằng nhau của hai xác suất (tỷ lệ)

Bài toán : • Xác suất xuất hiện biến cố A trong một dãy n phép thử độc lập là $P(A) = p_1$

• Xác suất xuất hiện biến cố B trong một dãy m phép thử độc lập là $P(B) = p_2$

• Với mức ý nghĩa α cho trước, hãy kiểm định cặp giả thuyết, đối thuyết sau:

H_0: $p_1 = p_2$

H_1: $p_1 \neq p_2$ (Đối thuyết hai phía)

H_1: $p_1 > p_2$ (Đối thuyết một phía)

H_1: $p_1 < p_2$ (Đối thuyết một phía)

- Chọn hai mẫu ngẫu nhiên từ hai tổng thể, kích thước n, m tương ứng.
- Gọi n_1 là số cá thể đặc tính A trong mẫu thứ nhất.
- Gọi m_1 là số cá thể đặc tính B trong mẫu thứ hai.

Đặt $f_1 = \frac{n_1}{n}, f_2 = \frac{m_1}{m}, f = \frac{n_1+m_1}{n+m}$

Xét biến ngẫu nhiên Z $= \frac{f_1-f_2-(p_1-p_2)}{\sqrt{f(1-f)(\frac{1}{n}+\frac{1}{m})}} \sim N(0; 1)$

Nếu H_0 đúng thì $Z_t = \frac{f_1-f_2}{\sqrt{f(1-f)(\frac{1}{n}+\frac{1}{m})}} \sim N(0; 1)$

- Ta tiến hành bài toán kiểm định như sau:

B1: Tính $Z_t = \dfrac{f_1 - f_2}{\sqrt{f(1-f)(\frac{1}{n}+\frac{1}{m})}}$

B2: Quyết định dựa vào bảng sau:

Quyết định H_1	$p_1 \neq p_2$	$p_1 > p_2$	$p_1 < p_2$
Bác bỏ H_0	$\lvert Z_t \rvert > u_{\frac{\alpha}{2}}$	$Z_t > u_\alpha$	$Z_t < -u_\alpha$
Chấp nhận H_0	$\lvert Z_t \rvert \leq u_{\frac{\alpha}{2}}$	$Z_t \leq u_\alpha$	$Z_t \geq -u_\alpha$

Ví dụ: Điều tra một loại bệnh ở hai vùng dân cư. Ở vùng I, người ta kiểm tra 500 người dân thì thấy có 60 người mắc bệnh. Ở vùng II, người ta kiểm tra 400 người thấy có 50 người mắc bệnh. Với mức ý nghĩa 0,05 có kết luận gì về tỷ lệ mắc bệnh ở hai vùng dân cư.

Chương 7: Tương quan và hồi quy mẫu

7.1 Hệ số tương quan mẫu.

7.2.Đường hồi quy

Chương 7: Tương quan và hồi quy mẫu

7.1 Hệ số tương quan mẫu.

Định nghĩa 1.1: Hệ số tương quan mẫu giữa X và Y là:

$$r_{XY} = \frac{\overline{xy} - \overline{x}.\overline{y}}{\hat{s}_X . \hat{s}_Y}$$

Hệ số tương quan mẫu là một ước lượng của hệ số tương quan giữa X và Y.

7.2.Đường hồi quy

1.Đường hồi quy mẫu.

Định nghĩa 2.1: Ký hiệu

$$\overline{Y}_{x_i} = (\overline{Y} / X = x_i) = \frac{1}{n_i}\sum_{j=1}^{h} y_j . n_{ij}, i = \overline{1,k} \Rightarrow \mathrm{M}_i\left(x_i, \overline{Y}_{x_i}\right), \overline{1,k}$$

Đường gấp khúc $M_1 M_2 \ldots M_k$ được gọi là đường hồi quy mẫu cua Y theo X.

2. Đường hồi quy tuyến tính mẫu.

Định nghĩa 2.2:Đường hồi quy tuyến tính mẫu của Y theo X là đương thẳng y=a+bx sao cho:

$$Q(a,b) = \sum_{i=1}^{k}\left[\overline{Y_{x_i}} - (a + bx_i)\right]^2 .n_i \to \min$$

Định lý:

$$\boxed{b = \frac{\overline{xy} - \overline{x}.\overline{y}}{\hat{s}_X^2}, a = \overline{y} - b\overline{x}}$$

Ý nghĩa: Đường hồi quy tuyến tính mẫu là đường thẳng xấp xỉ nội suy từ bảng số liệu của X và Y theo phương pháp bình phương tối tiểu.Nếu X và Y có tương quan xấp xỉ tuyến tính thì đường hồi quy tuyến tính mẫu cho ta một dự báo đơn giản:

$$\boxed{X = x_0 \Rightarrow Y \approx y_0 = a + bx_0}$$

3.Cách dùng máy tính bỏ túi: Nhập số liệu như để tính E(X),D(X),E(XY),… ở chương 3,$6 .Sau đó đọc kết quả theo cách sau:

a)Loại ES

SHIFT START VAR	$\overline{x} \Longrightarrow \overline{x}$
SHIFT START VAR	$\overline{y} \Longrightarrow \overline{y}$
SHIFT START REG	$r \Longrightarrow r_{xy}$
SHIFT START REG	$a \Longrightarrow a$
SHIFT START REG	$b \Longrightarrow b$
SHIFT START SUM	$\sum xy \Longrightarrow n.\overline{xy}$

b) Loại MS:

.SHIFT START S-VAR	$\overline{x} \Longrightarrow \overline{x}$
.SHIFT START S-VAR	$\triangleright\ \overline{y} \Longrightarrow \overline{y}$
.SHIFT START S-VAR	$\triangleright\triangleright\ r \Longrightarrow r_{xy}$
.SHIFT START S-VAR	$a \Longrightarrow a$
.SHIFT START S-VAR	$b \Longrightarrow b$
.SHIFT START S-SUM	$\sum xy \Longrightarrow n.\overline{xy}$

Ví dụ : Số vốn đầu tư X và lợi nhuận Y trong một đơn vị thời gian của 100 quan sát, được bảng số liệu:

X \ Y	0,3	0,7	1,0	
1	20	10		30
2		30	10	40
3		10	20	30
	20	50	30	N=100

Tìm hệ số tương quan thực nghiệm r, phương trình hồi quy tuyến tính của Y theo X.