ĐẠI HỌC ĐÀ NẴNG
TRƯỜNG ĐẠI HỌC SƯ PHẠM

Lê Văn Dũng

Giáo trình
XÁC SUẤT THỐNG KÊ

NHÀ XUẤT BẢN THÔNG TIN VÀ TRUYỀN THÔNG

LỜI NÓI ĐẦU

Xác suất thống kê là một bộ phận của toán học, nghiên cứu các hiện tượng ngẫu nhiên và có phạm vi ứng dụng rộng rãi trong khoa học cũng như thực tiễn. Hiện nay, Xác suất thống kê là môn học thuộc khối khoa học cơ bản được giảng dạy hầu hết tại các trường đại học, cao đẳng trên toàn quốc.

Học phần Xác suất thống kê bao gồm hai nội dung chính là lý thuyết xác suất và thống kê toán. Mục đích của lý thuyết xác suất là nghiên cứu quy luật của các hiện tượng ngẫu nhiên và phân tích để rút ra các quy luật và khả năng xuất hiện các hiện tượng đó. Nhờ vào ứng dụng của lý thuyết xác suất, thống kê toán nghiên cứu các phương pháp thu thập và phân tích dữ liệu để khám phá ra các tri thức và thông tin còn ẩn náu. Thống kê toán đã được ứng dụng rộng rãi trong các lĩnh vực như: Kinh tế, Sinh học, Xã hội học,...

"Giáo trình Xác suất thống kê" được biên soạn theo chương trình đào tạo Cử nhân Sư phạm Sinh học của Trường Đại học Sư phạm - Đại học Đà Nẵng với thời lượng 3 tín chỉ (45 tiết). Ngoài ra, giáo trình cũng có thể được sử dụng để giảng dạy các học phần Xác suất thống kê 2 tín chỉ của các ngành đào tạo khác của trường.

Nội dung giáo trình gồm 8 chương. Chương 1 giới thiệu các kiến thức về lý thuyết xác suất. Chương 2 giới thiệu về khái niệm biến ngẫu nhiên và các định lý giới hạn, trong đó Luật số lớn và Định lý giới hạn trung tâm là các định lý quan trọng trong ứng dụng thống kê. Chương 3 trình bày các kiến thức về vectơ ngẫu nhiên. Nội dung chương 4 đến chương 8 trình bày các kiến thức về thống kê toán. Bên cạnh đó, cuối mỗi chương còn có phần bài tập giúp các bạn sinh viên hệ thống lại các kiến thức lý thuyết đã học, rèn luyện kỹ năng tính toán phát triển tư duy xác suất và thống kê toán. Ngoài ra, giáo trình còn có phần phụ lục cung cấp giá trị của các hàm phân phối chuẩn tắc, giá trị tới hạn của phân bố Student (t - phân bố), giá trị tới hạn của phân bố khi bình phương và giá trị tới hạn của phân bố F.

Mặc dù đã có nhiều cố gắng trong công tác biên soạn, tham khảo nhiều tài liệu và trình bày một cách có hệ thống để giúp các bạn đọc dễ dàng tiếp cận hơn, song giáo trình được xuất bản lần đầu sẽ khó tránh khỏi các sai sót. Tác giả rất mong nhận được sự đóng góp ý kiến của các bạn đọc để giáo trình được hoàn thiện hơn trong lần tái bản sau. Mọi góp ý xin được gửi về địa chỉ: Khoa Toán, Trường Đại học Sư phạm - Đại học Đà Nẵng.

Qua đây tác giả xin bày tỏ lời cảm ơn tới các thầy, cô trong khoa Toán, Trường Đại học Sư phạm - Đại học Đà Nẵng; Nhà xuất bản Thông tin và Truyền thông đã ủng hộ cho việc biên soạn và xuất bản giáo trình này.

Đà Nẵng, tháng 6 năm 2016

Tác giả

MỤC LỤC

Chương 1
XÁC SUẤT

1.1. Không gian mẫu và biến cố

1.1.1. Phép thử

Trong thực tế có nhiều thí nghiệm có thể lặp đi lặp lại nhiều lần trong cùng một điều kiện như nhau nhưng chúng ta không thể biết chắc chắn kết quả nào sẽ xảy ra khi thực hiện thí nghiệm đó. Những thí nghiệm đó ta gọi là *phép thử ngẫu nhiên* (hay gọi tắt là *phép thử*).

Ví dụ 1.1.
- Gieo một con xúc xắc.
- Hỏi tháng sinh của một sinh viên được chọn ngẫu nhiên.
- Đo chiều cao của một sinh viên được chọn ngẫu nhiên.

Định nghĩa 1.1. *Phép thử* là những thí nghiệm mà khi thực hiện sẽ xảy ra kết quả hoàn toàn ngẫu nhiên ngay cả khi thí nghiệm đó được lặp lại nhiều lần trong cùng một điều kiện giống nhau.

1.1.2. Không gian mẫu

Định nghĩa 1.2. Tập tất cả các kết quả có thể xảy ra của một phép thử được gọi là *không gian mẫu*. Kí hiệu không gian mẫu là Ω.

Ví dụ 1.2. Khi tung một đồng xu, có hai kết quả có thể xảy ra: xuất hiện mặt sấp (S) hoặc xuất hiện mặt ngửa (N). Không gian mẫu trong trường hợp này là $\Omega = \{S; N\}$.

Ví dụ 1.3. Hỏi tháng sinh của một sinh viên được chọn ngẫu nhiên trong lớp học. Ta có không gian mẫu:

$$\Omega = \{1,2,3,4,5,6,7,8,9,10,11,12\}$$

Ví dụ 1.4. Gieo đồng thời hai con xúc xắc. Nếu ta quan tâm đến số chấm xuất hiện trên hai mặt của hai xúc xắc thì không gian mẫu sẽ là:

$$\Omega = \{(i; j) : i, j = 1; 2; 3; 4; 5; 6\}$$

Ví dụ 1.5. Đo chiều cao của một sinh viên được chọn ngẫu nhiên trong lớp học (đơn vị: mét). Ta có không gian mẫu:

$$\Omega = \{x \in \mathbb{R} : 0 < x < 2\}$$

1.1.3. Biến cố

Định nghĩa 1.3. Mỗi tập con của không gian mẫu được gọi là *biến cố*. Biến cố chỉ có 1 phần tử được gọi là *biến cố sơ cấp*, biến cố rỗng ($\emptyset$) gọi là *biến cố không thể*, không gian mẫu (Ω) gọi là *biến cố chắc chắn*.

Một biến cố xảy ra khi thực hiện phép thử nếu kết quả của thực hiện phép thử rơi vào biến cố đó.

Ví dụ 1.6. Cho không gian mẫu tuổi thọ (năm) của một thiết bị điện tử là $\Omega = \{x \in \mathbb{R} : x \geq 0\}$. Biến cố thiết bị điện tử bị hỏng trước 5 năm là $A = \{x \in \mathbb{R} : 0 \leq x < 5\}$.

Ví dụ 1.7. Hỏi tháng sinh của một sinh viên được chọn ngẫu nhiên trong lớp học.

- Biến cố sinh viên sinh vào tháng chẵn là $A = \{2,4,6,8,10,12\}$.
- Biến cố sinh viên có tháng sinh 32 ngày là $\emptyset$.
- Biến cố sinh viên có tháng sinh bé hơn 32 ngày là Ω.

1.1.4. Các phép toán trên biến cố

Cho A và B là hai biến cố của không gian mẫu Ω.

a) Phép giao

$A \cap B$ (hoặc kí hiệu là: $A.B$ hay đơn giản là AB), là biến cố xảy ra khi và chỉ khi đồng thời hai biến cố A và B cùng xảy ra.

Nếu hai biến cố A và B không thể đồng thời xảy ra ($A \cap B = \emptyset$) thì ta nói A và B xung khắc.

$$A \cap B = \{\omega \in \Omega : \omega \in A \text{ và } \omega \in B\}$$

b) Phép hợp

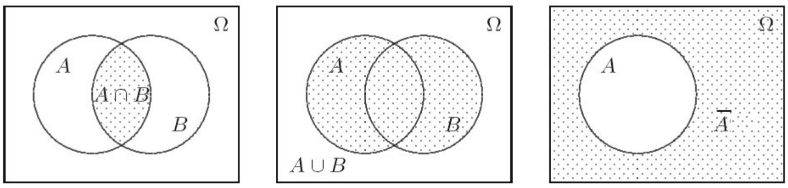
$A \cup B$ là biến cố xảy ra khi và chỉ khi có ít nhất một trong hai biến cố A, B xảy ra.

$$A \cup B = \{\omega \in \Omega : \omega \in A \text{ hoặc } \omega \in B\}$$

c) Phép lấy phần bù

Biến cố $\overline{A} = \Omega \backslash A$ được gọi là biến cố đối của A. Nếu A xảy ra thì $\overline{A}$ không xảy ra và ngược lại.

$$\overline{A} = \{\omega \in \Omega : \omega \notin A\}$$



Hình 1.1: *Biểu đồ Ven minh họa biến cố giao, biến cố hợp, biến cố đối*

Ví dụ 1.8. Tung một con xúc xắc cân đối đồng chất, khi đó có thể xuất hiện mặt 1 chấm, 2 chấm, 3 chấm,..., 6 chấm.

\+ Không gian mẫu $\Omega = \{1; 2; 3; 4; 5; 6\}$.

\+ Biến cố sơ cấp $\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}$.

\+ Biến cố $A = \{$số chấm của mặt xuất hiện bé hơn 4$\} = \{1; 2; 3\}$.

\+ Biến cố $B = \{$xuất hiện mặt chẵn$\} = \{2; 4; 6\}$.

Tìm các biến cố $A \cup B$, $A \cap B$, $\overline{A}$.

Ví dụ 1.9. Đo chiều cao một sinh viên được chọn ngẫu nhiên trong lớp học (đơn vị: mét). Không gian mẫu là:

$$\Omega = \mathbb{R}^+ = \{x \in \mathbb{R} | x > 0\}$$

Với:

$$A = \{x | 1,5 \leq x < 1,7\} \text{ và } B = \{x | 1,6 < x < 1,8\}$$

Tìm $A \cup B$, $A \cap B$, $\overline{A}$, $\overline{A} \cap B$.

Ví dụ 1.10. Hai xạ thủ cùng bắn vào một mục tiêu, kí hiệu A là biến cố xạ thủ 1 bắn trúng mục tiêu, B là biến cố xạ thủ 2 bắn trúng mục tiêu. Hãy biểu diễn qua A và B các biến cố sau.

a) Xạ thủ 1 không bắn trúng mục tiêu.

b) Cả hai xạ thủ bắn trúng mục tiêu.

c) Có ít nhất một xạ thủ bắn trúng mục tiêu.

d) Có đúng một xạ thủ bắn trúng mục tiêu.

e) Không có xạ thủ nào bắn trúng mục tiêu.

1.2. Xác suất của biến cố

1.2.1. Hệ tiên đề xác suất

Cho một phép thử và Ω là không gian mẫu của phép thử đó. Để đo lường khả năng xảy ra một biến cố ta sẽ đặt tương ứng mỗi biến cố A của Ω với một thực $P(A)$ thỏa mãn 3 tiên đề sau:

Tiên đề 1: $0 \leq P(A) \leq 1$ với mọi biến cố A.

Tiên đề 2: $P(\Omega) = 1$.

Tiên đề 3: Nếu $A_1, A_2, ..., A_n, ...$ là một dãy các biến cố đôi một xung khắc thì:

$$P(A_1 \cup A_2 \cup A_3 \cup ...) = \sum_{n=1}^{\infty} P(A_n)$$

Khi đó $P(A)$ được gọi là *xác suất* của biến cố A.

Ví dụ 1.11. Tung một đồng xu. Giả sử khả năng xuất hiện mặt sấp (S) và mặt ngửa (N) là như nhau trong mỗi lần tung, tức là:

$$P(\{S\}) = P(\{N\})$$

Mặt khác, do không gian mẫu $\Omega = \{S, N\} = \{S\} \cup \{N\}$ nên:

$$1 = P(\Omega) = P(\{S\}) + P(\{S\})$$

Từ đó ta có:

$$P(\{S\}) = P(\{S\}) = \frac{1}{2}$$

Ví dụ 1.12. Gieo một con xúc xắc. Giả sử rằng 6 mặt của xúc xắc có khả năng xuất hiện như nhau trong mỗi lần gieo. Khi đó ta có:

$$P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = P(\{5\}) = P(\{6\}) = \frac{1}{6}$$

Vì vậy xác suất xuất hiện mặt chẵn sẽ là:

$$P(\{2,4,6\}) = P(\{2\}) + P(\{4\}) + P(\{6\}) = \frac{3}{6} = 0,5$$

1.2.2. Một số tính chất cơ bản của xác suất

Tính chất 1.1.

$$P(\emptyset) = 0$$

Chứng minh. Từ Tiên đề 3 lấy $A_1 = \Omega$, $A_n = \emptyset$ với mọi $n \geq 2$ ta được:

$$P(\Omega) = \sum_{n=1}^{\infty} P(A_n) = P(\Omega) + \sum_{n=2}^{\infty} P(\emptyset).$$

Suy ra:

$$P(\emptyset) = 0$$

■

Tính chất 1.2.

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

Chứng minh. Vì $\Omega = A \cup \overline{A}$ và $A \cap \overline{A} = \emptyset$ nên

$$1 = P(\Omega) = P(A \cup \overline{A}) = P(A) + P(\overline{A})$$

■

Tính chất 1.3. *Nếu* $A \subset B$ *thì* $P(A) \leq P(B)$.

Chứng minh. Vì $A \subset B$ nên $B = A \cup (\overline{A}B)$. Do đó:

$$P(B) = P(A \cup \overline{A}B) = P(A) + P(\overline{A}B) \geq P(A)$$

■

Tính chất 1.4. *Với* A *và* B *là hai biến cố bất kì,*

$$P(A \cup B) = P(A) + P(B) - P(AB).$$

Chứng minh. Áp dụng Tiên đề 3 ta có các đẳng thức sau:

$$P(A \cup B) = P(A\overline{B}) + P(AB) + P(\overline{A}B) \quad (1.1)$$

$$P(A \cup B) = P(A) + P(\overline{A}B) \quad (1.2)$$

$$P(B) + P(\overline{B}A) = P(A \cup B) \quad (1.3)$$

Cộng vế với vế ba đẳng thức (1.1), (1.2) và (1.3) ta được điều phải chứng minh. ■

Sử dụng Tính chất 1.4 ta dễ dàng chứng minh được tính chất sau.

Tính chất 1.5. *Với* A, B *và* C *là hai biến cố bất kì,*

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(BC) - P(AC) + P(ABC)$$

1.2.3. Không gian mẫu gồm các biến cố sơ cấp đồng khả năng

Cho không gian mẫu Ω gồm N biến cố sơ cấp có khả năng xảy ra bằng nhau, tức là:

$$\Omega = \{\omega_1, \omega_2, ..., \omega_N\}$$

và

$$P(\{\omega_1\}) = P(\{\omega_2\}) = ... = P(\{\omega_N\})$$

Khi đó, theo Tiên đề 2 ta có:

$$P(\{\omega_1\}) = P(\{\omega_2\}) = ... = P(\{\omega_N\}) = \frac{1}{N}$$

Kết hợp Tiên đề 3 ta có: Với A là một biến cố bất kì của Ω

$$P(A) = \frac{|A|}{|\Omega|}$$

trong đó $|A|$ là số phần tử của A.

Ví dụ 1.13. Một hộp đựng 4 viên bi xanh, 5 viên bi đỏ và 6 viên bi vàng. Các viên bi đồng chất, giống nhau hoàn toàn về kích thước và khối lượng. Lấy ngẫu nhiên 5 viên bi. Tính xác suất các biến cố sau:

a) A: lấy được 1 bi xanh, 2 bi đỏ và 2 bi vàng.
b) B: lấy được 3 bi xanh.
c) C: lấy được ít nhất 4 bi đỏ.
d) D: lấy được ít nhất 1 bi vàng.

Giải. $|\Omega| = C_{15}^5$.

a) $|A| = C_4^1 C_5^2 C_6^2$ suy ra $P(A) = \dfrac{200}{1001} \approx 0,2$.

b) $|B| = C_4^3 C_1^2$ suy ra $P(B) = \dfrac{20}{273} \approx 0,073$.

c) $|C| = C_5^4 C_{10}^1 + C_5^5$ suy ra $P(C) = \dfrac{226}{3003} \approx 0,075$.

d) $|\overline{D}| = C_9^5$ suy ra $P(D) = 1 - P(\overline{D}) = 1 - \dfrac{6}{143} \approx 0,985$. ■

1.3. Đại số tổ hợp

1.3.1. Quy tắc nhân

Nếu một công việc được thực hiện qua k bước.

Bước 1 có n_1 cách thực hiện,

Bước 2 có n_2 cách thực hiện,

...

Bước k có n_k cách thực hiện.
Khi đó, có $n_1 \times n_2 \times ... \times n_k$ cách thực hiện công việc đó.

1.3.2. Hoán vị

Số cách sắp xếp n phần tử vào n vị trí sao cho mỗi vị trí có đúng 1 phần tử là $n!$.

1.3.3. Tổ hợp

Số tập con k phần tử của một tập n phần tử là:

$$C_n^k = \frac{n!}{k!(n-k)!}\ (0 \le k \le n)$$

1.3.4. Chỉnh hợp

Số cách lấy ra k phần tử từ tập n phần tử rồi sắp xếp theo một thứ tự nào đó là:

$$A_n^k = \frac{n!}{(n-k)!} = C_n^k.k!\ (1 \le k \le n)$$

1.4. Xác suất có điều kiện

Chúng ta xét ví dụ sau: Ở một lớp học phần môn Triết học gồm 17 sinh viên nam và 13 sinh viên nữ. Trong số đó có 12 sinh viên nam và 11 sinh viên nữ thi qua môn Triết học.

Chọn ngẫu nhiên một sinh viên, xác suất sinh viên đó thi qua môn Triết học là 23/30.

Nhưng nếu chọn ngẫu nhiên một sinh viên nam thì xác suất sinh viên đó thi qua môn Triết học sẽ là 12/17.

Rõ ràng 2 xác suất trên không bằng nhau. Để phân biệt 2 xác suất trên ta kí hiệu A là biến cố sinh viên đó thi qua môn Triết học, B là điều kiện sinh viên được chọn là sinh viên nam. Khi đó P(A|B)=12/17 được gọi là xác suất của biến cố A với điều kiện B.

Chú ý rằng:

$$P(A|B) = \frac{|A \cap B|}{|B|} = \frac{|A \cap B|/|\Omega|}{|B|/|\Omega|} = \frac{P(A \cap B)}{P(B)}$$

Định nghĩa 1.4. Cho hai biến cố A và B với $P(B) \neq 0$, *xác suất của A với điều kiện B* đã xảy ra, kí hiệu $P(A|B)$, xác định bởi

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Ví dụ 1.14. Một hộp đựng 20 bóng đèn tốt, 7 bóng đèn sẽ hỏng sau 1 giờ sử dụng và 3 bóng đèn hỏng. Lấy ngẫu nhiên một chiếc sử dụng thấy rằng nó không phải là bóng đèn hỏng. Tính xác suất đó là chiếc bóng đèn tốt.

Giải. Gọi A là biến cố lấy được bóng đèn tốt, B là biến cố lấy được bóng đèn không phải là bóng đèn hỏng.

$$P(A|B) = 20/27 \approx 0,74$$

■

Ví dụ 1.15. Trong một vùng dân cư tỉ lệ người hút thuốc là 60%, tỉ lệ người vừa hút thuốc vừa bị viêm phổi là 35%. Chọn ngẫu nhiên một người của vùng dân cư đó thấy người này hút thuốc. Tìm xác suất người này bị viêm phổi.

Giải. Gọi A là biến cố người được chọn hút thuốc, B là biến cố người được chọn bị viêm phổi. Xác suất để người này bị viêm phổi là:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{0,35}{0,6} \approx 0,583$$

■

Tính chất 1.6.

1) $P(\emptyset|B) = 0,\ P(B|B) = 1,\ P(\Omega|B) = 1.$

2) $P(A|B) + P(\overline{A}|B) = 1.$

3) Nếu A_1 và A_2 xung khắc thì:

$$P(A_1 \cup A_2|B) = P(A_1|B) + P(A_2|B)$$

4) Nếu $P(B) \neq 0$ thì $P(A \cap B) = P(B)P(A|B)$.

Nếu $P(A) \neq 0$ thì $P(A \cap B) = P(A)P(B|A)$.

5) $P(B|A) = \dfrac{P(B)P(A|B)}{P(A)}.$

1.5. Công thức nhân xác suất

Định lý 1.1. *Cho* $A_1, A_2, ..., A_n$ *là các biến có của không gian mẫu* Ω *thỏa mãn* $P(A_1A_2...A_{n-1}) \neq 0$. *Khi đó:*

$$P(A_1A_2...A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2)....P(A_n|A_1A_2...A_{n-1})$$

Chứng minh.

$$\begin{aligned} P(A_1A_2...A_n) &= P(A_1A_2...A_{n-1})P(A_n|A_1A_2...A_{n-1}) \\ &= P(A_1...A_{n-2})P(A_{n-1}|A_1...A_{n-2})P(A_n|A_1...A_{n-1}) \\ &= ... \\ &= P(A_1)P(A_2|A_1)P(A_3|A_1A_2)....P(A_n|A_1A_2...A_{n-1}) \end{aligned}$$

■

Ví dụ 1.16. Một hộp đựng 4 chiếc bút mới và 6 chiếc bút cũ. Mỗi ngày lấy ngẫu nhiên một chiếc ra sử dụng, cuối ngày trả bút đó lại hộp. Tính xác suất:

a) Sau 3 ngày sử dụng hộp còn đúng 1 bút mới.

b) Sau 2 ngày sử dụng hộp còn đúng 3 bút mới.

Giải. Kí hiệu A_k là biến cố ngày thứ k lấy được bút mới.

a) $P(A_1A_2A_3) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2) = \frac{4}{10}.\frac{3}{10}.\frac{2}{10} = 0,24.$

b) Biến cố sau 2 ngày sử dụng hộp còn đúng 3 bút mới là $(A_1\overline{A}_2) \cup (\overline{A}_1A_2)$, nên:

$$\begin{aligned} P((A_1\overline{A}_2) \cup (\overline{A}_1A_2)) &= P(A_1\overline{A}_2) + P(\overline{A}_1A_2) \\ &= P(A_1)P(\overline{A}_2|A_1) + P(\overline{A}_1)P(A_2|\overline{A}_1) \\ &= \frac{4}{10}.\frac{7}{10} + \frac{6}{10}.\frac{4}{10} = 0,52 \end{aligned}$$

■

Ví dụ 1.17. Trong một trường đại học có 40% sinh viên học tiếng Anh, 30% sinh viên học tiếng Pháp, trong số sinh viên học tiếng Anh có 55% sinh viên học tiếng Pháp. Chọn ngẫu nhiên một sinh viên, biết sinh viên đó học tiếng Pháp. Tính xác suất để sinh viên đó học tiếng Anh.

Giải. Gọi A là biến cố chọn được sinh viên biết tiếng Anh, B là biến cố chọn được sinh viên biết tiếng Pháp.

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(A).P(B|A)}{P(B)} = \frac{0,4.0,55}{0,3} \approx 0,733$$

■

1.6. Các biến cố độc lập

Hai biến cố A và B độc lập nếu việc xảy ra hay không xảy ra biến cố này không làm thay đổi xác suất xảy ra của biến cố kia.

Tức là:

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

Khi đó ta có:

$$P(A \cap B) = P(A)P(B)$$

Từ đó ta định nghĩa hai biến cố độc lập như sau.

Định nghĩa 1.5. Hai biến cố A và B được gọi là *độc lập* nếu

$$P(A \cap B) = P(A)P(B)$$

Trong trường hợp tổng quát ta có định nghĩa sau.

Định nghĩa 1.6. Một tập hữu hạn các biến cố $\{A_1; A_2; ..., A_n\}$ $(n \geq 2)$ được gọi là *độc lập* nếu với mọi k $(2 \leq k \leq n)$ biến cố bất kì A_{n_1}, A_{n_2},..., A_{n_k} ta có:

$$P(A_{n_1}.A_{n_2}...A_{n_k}) = P(A_{n_1})P(A_{n_2})...P(A_{n_k})$$

Trường hợp $n = 3$, ba biến cố A, B, C độc lập khi và chỉ khi thỏa mãn 4 đẳng thức sau:

$$P(AB) = P(A)P(B)$$
$$P(BC) = P(B)P(C)$$
$$P(CA) = P(C)P(A)$$
$$P(ABC) = P(A)P(B)P(C)$$

Định lý 1.2. *Nếu A và B độc lập thì A và $\overline{B}$, $\overline{A}$ và B, $\overline{A}$ và $\overline{B}$ là những cặp biến cố độc lập.*

Ví dụ 1.18. Hộp I có 3 bi đỏ và 7 bi xanh, hộp II có 6 bi đỏ và 4 bi xanh. Lấy ngẫu nhiên từ mỗi hộp ra 1 viên bi. Tìm xác suất:
a) Lấy được hai viên bi cùng màu đỏ.
b) Lấy được 1 bi xanh và 1 bi đỏ.

Giải. Gọi A là biến cố lấy từ hộp I được viên bi màu đỏ, B là biến cố lấy từ hộp II được viên bi màu đỏ. A và B là 2 biến cố độc lập.
a) $P(AB) = P(A).P(B) = \dfrac{3}{10}.\dfrac{6}{10} = 0,18.$
b) $P(A\overline{B} \cup \overline{A}B) = P(A)P(\overline{B}) + P(\overline{A})P(B) = 0,54.$ ■

1.7. Công thức xác suất toàn phần và công thức Bayes

1.7.1. Hệ đầy đủ

Định nghĩa 1.7. Một hệ gồm n biến cố $E_1, E_2, \ldots, E_n$ được gọi là *hệ đầy đủ* nếu thỏa mãn hai điều kiện:

(i) $E_i \cap E_j = \emptyset$ nếu $i \neq j$ (các biến cố đôi một xung khắc);

(ii) $E_1 \cup E_2 \cup \ldots \cup E_n = \Omega$ (chắc chắn có 1 biến cố xảy ra).

Từ định nghĩa hệ đầy đủ ta suy ra: nếu $E_1, E_2, \ldots, E_n$ là hệ đầy đủ thì:

$$P(E_1) + P(E_2) + ... + P(E_n) = 1$$

Ví dụ 1.19. Hỏi tháng sinh của một sinh viên được chọn ngẫu nhiên. Kí hiệu:
E_1 là biến cố sinh viên được hỏi sinh vào quý 1;
E_2 là biến cố sinh viên được hỏi sinh vào quý 2;
E_3 là biến cố sinh viên được hỏi sinh vào quý 3;
E_4 là biến cố sinh viên được hỏi sinh vào quý 4.

Khi đó E_1, E_2, E_3, E_4 là hệ đầy đủ.

Ví dụ 1.20. Một hộp đựng 5 bi xanh, 6 bi đỏ và 7 bi vàng. Lấy ngẫu nhiên 2 viên bi. Hãy chỉ ra một số hệ đầy đủ.

1.7.2. Công thức xác suất toàn phần và công thức Bayes

Định lý 1.3. *Giả sử* $\{E_i; 1 \le i \le n\}$ *là một hệ đầy đủ sao cho* $P(E_i) > 0$, *A là biến cố bất kì. Khi đó:*

1) $P(A) = P(E_1)P(A|E_1) + P(E_2)P(A|E_2) + ... + P(E_n)P(A|E_n)$.

2) Nếu thêm điều kiện $P(A) > 0$ *thì*

$$P(E_i|A) = \frac{P(E_i)P(A|E_i)}{P(A)}$$
$$= \frac{P(E_i)P(A|E_i)}{P(E_1)P(A|E_1) + P(E_2)P(B|E_2) + ... + P(E_n)P(A|E_n)}$$

Ví dụ 1.21. Hộp I đựng 4 bi xanh và 3 bi đỏ và 2 bi vàng, hộp II đựng 5 bi xanh 2 bi đỏ và 3 bi vàng. Từ hộp I lấy ngẫu nhiên ra một viên bi bỏ vào hộp II, sau đó từ hộp II lấy ngẫu nhiên ra hai viên bi. Tính xác suất hai viên bi lấy ra ở lần thứ hai là 2 bi xanh.

Giải. Gọi E là biến cố viên bi lấy từ hộp I bỏ vào hộp II là bi xanh, A là biến cố 2 viên bi lấy lần 2 là 2 viên bi xanh.

$$P(A) = P(E)P(A|E) + P(\overline{E})P(A|\overline{E}) = \frac{4}{9}.\frac{C_6^2}{C_{11}^2} + \frac{5}{9}.\frac{C_5^2}{C_{11}^2} = \frac{2}{9} \approx 0,22$$

■

Ví dụ 1.22. Một nhà máy có 3 phân xưởng sản xuất. Phân xưởng I sản xuất 50% sản phẩm, phân xưởng II sản xuất 30% sản phẩm, phân xưởng III sản xuất 20% sản phẩm. Biết rằng tỉ lệ phế phẩm do phân xưởng I, phân xưởng II, phân xưởng III sản xuất ra tương ứng là 2%, 1% và 3%. Lấy ngẫu nhiên 1 sản phẩm của nhà máy.

a) Tìm xác suất để sản phẩm lấy ra là phế phẩm.

b) Giả sử sản phẩm lấy ra là chính phẩm. Tính xác suất để sản phẩm đó do phân xưởng I sản xuất.

Giải. Gọi E_1, E_2, E_3 lần lượt là các biến cố sản phẩm lấy ra là của phân xưởng I, II và III. Khi đó: $\{E_1, E_2, E_3\}$ là hệ đầy đủ.

a) Gọi A là biến cố sản phẩm lấy ra là phế phẩm. Theo công thức xác suất toàn phần:

$$\begin{aligned} P(A) &= P(E_1).P(A|E_1) + P(E_2)P(A|E_2) + P(E_3)P(A|E_3) \\ &= 0,5.0,02 + 0,3.0,01 + 0,20,03 = 0,019 \end{aligned}$$

b) $P(E_1|\overline{A}) = \dfrac{P(\overline{A}|E_1).P(E_1))}{P(\overline{A})} = \dfrac{0,98.0,5}{1-0,019} \approx 0,4995.$ ■

Ví dụ 1.23. Một công ty sử dụng hai máy cùng sản xuất 1 loại sản phẩm. Tỉ lệ phế phẩm của máy I là 3% và của máy II là 2%. Số lượng sản phẩm do máy I sản xuất là 2/3 và máy II sản xuất là 1/3 tổng sản phẩm của công ty. Tính tỉ lệ phế phẩm của công ty đó.

Giải. Chọn ngẫu nhiên 1 sản phẩm. Gọi E là biến cố chọn được sản phẩm của nhà máy I, A là biến cố chọn được phế phẩm.

$$\begin{aligned} P(A) &= P(E)P(A|E) + P(\overline{E})P(A|\overline{E}) \\ &= \frac{2}{3}.0,03 + \frac{1}{3}.0,02 \approx 0,027 \end{aligned}$$

Vậy tỉ lệ phế phẩm của công ty là 2,7%. ■

1.8. Công thức Bernoulli

Định lý 1.4. *Cho Ω là không gian mẫu của một phép thử và A là một biến cố thỏa mãn $P(A) = p \in (0; 1)$.*

Thực hiện phép thử n lần độc lập, xác suất có đúng k lần xuất hiện biến cố A là:

$$p_n(k) = C_n^k p^k (1-p)^{n-k}$$

Ví dụ 1.24. Tung 10 lần một con xúc xắc cân đối đồng chất.
a) Tính xác suất có đúng 6 lần xuất hiện mặt một chấm.
b) Tính xác suất có ít nhất 9 lần xuất hiện mặt một chấm.
c) Tính xác suất có ít nhất 1 lần xuất hiện mặt một chấm.

Giải. Gọi A là biến cố xuất hiện mặt một chấm ở mỗi lần tung xúc xắc, $p = P(A) = 1/6$.

a) $p_{10}(6) = C_{10}^6(\frac{1}{6})^6(\frac{5}{6})^4 \approx 0,0022.$

b) $p_{10}(k \geq 9) = C_{10}^9(\frac{1}{6})^9(\frac{5}{6})^1 + (\frac{1}{6})^{10} \approx 8.10^{-7}.$

c) $p_{10}(k \geq 1) = 1 - p_{10}(k = 0) = 1 - (\frac{5}{6})^{10} \approx 0,84.$ ■

Định lý 1.5. *Cho $n \in \mathbb{Z}$, $n \geq 1$ và $p \in (0;1)$. Hàm số*

$$p_n(k) = C_n^k p^k (1-p)^{n-1} \text{ với } k \in \{0, 1, 2..., n\}$$

đạt giá trị lớn nhất tại

$$k = \begin{cases} [(n+1)p] & \text{nếu } (n+1)p \notin \mathbb{Z} \\ (n+1)p - 1 \text{ và } (n+1)p & \text{nếu } (n+1)p \in \mathbb{Z} \end{cases}$$

Ví dụ 1.25. Xác suất bắn trúng mục tiêu của một xạ thủ là $0,6$. Cho xạ thủ này bắn độc lập 20 phát vào mục tiêu. Tìm số lần bắn trúng mục tiêu có xác suất lớn nhất.

Giải. $(n+1)p = 21.0,6 = 12,6 \notin \mathbb{Z}$ nên số lần bắn trúng mục tiêu có xác suất lớn nhất là $k = 12$. ■

BÀI TẬP CHƯƠNG 1

▷ **1.1.** Gieo đồng thời 2 con xúc xắc. Tính xác suất:
a) Tổng số chấm xuất hiện trên 2 con là 7.
b) Số chấm xuất hiện trên 2 con hơn kém nhau 2.

▷ **1.2.** Một nhà khách có 6 phòng đơn. Có 10 khách đến thuê phòng, trong đó có 6 nam và 4 nữ. Người quản lí chọn 6 người. Tính xác suất:
a) Cả 6 người đều là nam.
b) Có 4 nam và 2 nữ.
c) Có ít nhất 2 nữ.
d) Có ít nhất 1 nữ.

▷ **1.3.** Một hộp đựng 6 quả cầu trắng, 4 quả cầu đỏ và 2 quả cầu đen. Chọn ngẫu nhiên 6 quả cầu. Tìm xác suất để chọn được 3 quả trắng, 2 đỏ và 1 đen.

▷ **1.4.** Có 30 tấm thẻ đánh số từ 1 đến 30. Chọn ngẫu nhiên ra 10 tấm thẻ. Tìm xác suất:

a) Tất cả 10 tấm đều mang số chẵn.

b) Có đúng 5 tấm mang số chia hết cho 3.

▷ **1.5.** Ở một nước có 50 tỉnh, mỗi tỉnh có 2 đại biểu Quốc hội. Người ta chọn ngẫu nhiên 50 đại biểu trong số 100 đại biểu để thành lập một ủy ban. Tính xác suất:

a) Trong ủy ban có ít nhất 1 đại biểu của thủ đô.

b) Mỗi tỉnh đều có đúng 1 đại biểu của ủy ban.

▷ **1.6.** Viết các chữ số: 1, 2, 3, 4, 5, 6, 7, 8 và 9 lên các tấm phiếu, sau đó sắp thứ tự ngẫu nhiên thành một hàng.

a) Tính xác suất để được một số chẵn.

b) Cũng từ 9 tấm phiếu trên chọn ngẫu nhiên 4 tấm rồi xếp thứ tự thành hàng, tính xác suất để được 1 số chẵn.

▷ **1.7.** Bộ bài có 52 lá, trong đó có 4 lá Át. Lấy ngẫu nhiên 3 lá. Tính xác suất:

a) Có 1 lá Át.

b) Có 2 lá Át.

c) Có ít nhất 1 lá Át.

▷ **1.8.** Một bình có 10 bi, trong đó có 3 bi đỏ, 4 bi xanh, 3 bi đen. Lấy ngẫu nhiên 4 viên. Tính xác suất:

a) Có đúng 2 bi xanh.

b) Có 1 bi xanh, 1 bi đỏ và 2 bi đen.

▷ **1.9.** Có 15 sản phẩm, trong đó có 3 phế phẩm, được bỏ ngẫu nhiên vào 3 cái hộp I, II, III, mỗi hộp 5 sản phẩm. Tính xác suất:

a) Ở hộp thứ I chỉ có 1 phế phẩm.

b) Các hộp đều có phế phẩm.

c) Các phế phẩm đều ở hộp thứ III.

▷ **1.10.** Một cửa hàng đồ điện nhập một lô bóng đèn điện đóng thành từng hộp, mỗi hộp 12 chiếc. Chủ cửa hàng kiểm tra chất lượng bằng cách lấy ngẫu nhiên 3 bóng để thử và nếu cả 3 bóng cùng tốt thì hộp bóng điện đó được chấp nhận. Tìm xác suất một hộp bóng điện được chấp nhận nếu trong hộp có 4 bóng bị hỏng.

▷ **1.11.** Trong đề cương ôn tập môn học gồm 10 câu hỏi lý thuyết và 30 bài tập. Mỗi đề thi gồm có 1 câu hỏi lý thuyết và 3 bài tập được lấy ngẫu nhiên trong đề cương. Một học sinh A chỉ học 4 câu lí thuyết và 12 câu bài tập trong đề cương. Khi thi học sinh A chọn ngẫu nhiên 1 đề thi trong các đề thi được tạo thành từ đề cương. Biết rằng học sinh A chỉ trả lời được câu lí thuyết và bài tập đã học. Tính xác suất:

a) Học sinh A không trả lời được lí thuyết.

b) Học sinh A chỉ trả lời được 2 câu bài tập.

c) Học sinh A đạt yêu cầu, biết rằng muốn đạt yêu cầu thì phải trả lời được câu hỏi lý thuyết và ít nhất 2 bài tập.

▷ **1.12.** Chọn ngẫu nhiên một vé xổ số có 5 chữ số từ 0 đến 9. Tính xác suất:

a) Số trên vé không có chữ số 1.

b) Số trên vé không có chữ số 2.

c) Số trên vé không có chữ số 1 hoặc không có chữ số 2.

▷ **1.13.** Xếp ngẫu nhiên 5 người A, B, C, D và E vào một cái bàn dài có 5 chỗ ngồi, tính xác suất:

a) A và B đầu bàn.

b) A và B cạnh nhau.

▷ **1.14.** Một máy bay có 3 bộ phận A, B, C có tầm quan trọng khác nhau. Máy bay sẽ rơi khi có một viên đạn trúng vào A hoặc hai viên đạn trúng vào B hoặc ba viên trúng vào C. Giả sử các bộ phận A, B, C lần lượt chiếm 15%, 30% và 55% diện tích máy bay. Bắn 3 phát vào máy bay. Tính xác suất máy bay rơi nếu:

a) máy bay bị trúng 2 viên đạn.

b) máy bay bị trúng 3 viên đạn.

▷ **1.15.** Một công ty sử dụng hai hình thức quảng cáo là quảng cáo trên đài phát thanh và quảng cáo trên tivi. Giả sử có 25% khách hàng biết được thông tin quảng cáo qua tivi và 34% khách hàng biết được thông tin quảng cáo qua đài phát thanh và 10% khách hàng biết được thông tin quảng cáo qua cả hai hình thức quảng cáo. Chọn ngẫu nhiên một khách hàng, tìm xác suất khách hàng đó biết được thông tin quảng cáo của công ty.

▷ **1.16.** Một lớp sinh viên có 50% học tiếng Anh, 40% học tiếng Pháp, 30% học tiếng Đức, 20% học tiếng Anh và tiếng Pháp, 15% học tiếng Pháp và tiếng Đức, 10% học tiếng Anh và tiếng Đức, 5% học cả ba thứ tiếng Anh, Pháp và Đức. Chọn ngẫu nhiên ra một sinh viên. Tìm xác suất:

a) Sinh viên đó học ít nhất 1 trong 3 ngoại ngữ kể trên.

b) Sinh viên đó chỉ học tiếng Anh và tiếng Đức.

c) Sinh viên đó học tiếng Pháp, biết sinh viên đó học tiếng Anh.

▷ **1.17.** Một công ty đầu tư hai dự án A và B. Xác suất công ty bị thua lỗ dự án A là 0,1, bị thua lỗ dự án B là 0,2 và thua lỗ cả 2 dự án là 0,05. Tính xác suất công ty có đúng 1 dự án bị thua lỗ.

▷ **1.18.** Một sinh viên phải thi liên tiếp 2 môn là triết học và toán. Xác suất qua môn triết học là 0,6 và qua môn toán là 0,7. Nếu trước đó đã qua môn triết thì xác suất qua môn toán là 0,8. Tính xác suất:

a) Sinh viên đó thi qua cả hai môn.

b) Sinh viên đó thi qua ít nhất 1 môn.

c) Sinh viên đó thi qua đúng 1 môn.

d) Sinh viên đó thi qua môn toán biết rằng đã không qua môn triết học.

▷ **1.19.** Một hộp bút có 10 cây bút, trong đó có 7 cây đã sử dụng. Ngày thứ 1 người ta lấy ngẫu nhiên từ hộp bút 1 cây để sử dụng, cuối ngày trả cây bút vào hộp, ngày thứ 2 và ngày thứ 3 cũng thực hiện như thế. Tính xác suất:

a) Sau ngày thứ 3 trong hộp không còn cây bút mới nào.

b) 3 cây bút lấy ra ở 3 ngày đều là bút đã sử dụng.

c) 2 ngày đầu lấy bút mới, ngày thứ 3 lấy bút đã sử dụng.

▷ **1.20.** Có hai lô hàng. Lô I có 90 chính phẩm và 10 phế phẩm, lô II có 80 chính phẩm và 20 phế phẩm. Lấy ngẫu nhiên từ mỗi lô 1 sản phẩm. Tính xác suất:

a) Lấy được 1 chính phẩm.

b) Lấy được ít nhất 1 chính phẩm.

▷ **1.21.** Một thiết bị có 2 bộ phận hoạt động độc lập. Cho biết trong thời gian hoạt động xác suất chỉ 1 bộ phận hỏng là 0,38 và xác suất bộ phận thứ 2 hỏng là 0,8. Tính xác suất bộ phận thứ nhất bị hỏng trong thời gian hoạt động.

▷ **1.22.** Ba khẩu súng được bắn độc lập vào một mục tiêu, xác suất để 3 khẩu bắn trúng lần lượt bằng 0,7; 0,8; 0,5. Mỗi khẩu bắn 1 viên, tính xác suất:

a) Một khẩu bắn trúng.

b) Hai khẩu bắn trúng.

c) Cả ba khẩu bắn trật.

d) Ít nhất một khẩu trúng.

e) Khẩu thứ nhất bắn trúng biết rằng có 2 viên trúng.

▷ **1.23.** Một thiết bị gồm 3 cụm chi tiết, mỗi cụm bị hỏng không ảnh hưởng gì đến các cụm khác và chỉ cần một cụm hỏng là thiết bị ngừng hoạt động. Xác suất để cụm thứ nhất bị hỏng trong ngày làm việc là $0,1$; tương tự cho hai cụm còn lại là $0,5$ và $0,15$. Tính xác suất để thiết bị không bị ngừng hoạt động trong ngày.

▷ **1.24.** Một nhà máy sản xuất linh kiện điện tử có 4 phân xưởng. Phân xưởng 1 sản xuất 40%; phân xưởng 2 sản xuất 30%; phân xưởng 3 sản xuất 20% và phân xưởng 4 sản xuất 10% sản phẩm của toàn xí nghiệp. Tỉ lệ phế phẩm của các phân xưởng 1, 2, 3, 4 tương ứng là 1%, 2%, 3%, 4%. Kiểm tra ngẫu nhiên một sản phẩm do nhà máy sản xuất.

a) Tìm xác suất để sản phẩm lấy ra là sản phẩm tốt.

b) Cho biết sản phẩm lấy ra kiểm tra là phế phẩm. Tính xác suất để phế phẩm đó do phân xưởng 1 sản xuất.

▷ **1.25.** Một dây chuyền lắp ráp nhận các chi tiết từ hai nhà máy khác nhau, tỷ lệ chi tiết do nhà máy thứ nhất cung cấp là 60%, còn lại của nhà máy thứ 2. Tỷ lệ chính phẩm của nhà máy thứ nhất là 90% và của nhà máy thứ 2 là 85%. Lấy ngẫu nhiên một chi tiết trên dây chuyền và thấy rằng nó là chính phẩm, tìm xác suất chi tiết đó do nhà máy thứ nhất sản xuất.

▷ **1.26.** Một cửa hàng máy tính chuyên kinh doanh 3 loại nhãn hiệu là IBM, Dell và Toshiba. Trong cơ cấu hàng bán, máy IBM chiếm 50%; Dell 30% và còn lại là máy Toshiba. Tất cả máy bán ra có thời hạn bảo hành là 12 tháng. Kinh nghiệm kinh doanh của chủ cửa hàng cho thấy 10% máy IBM phải sửa chữa trong thời hạn bảo hành; tỷ lệ sản phẩm cần sửa chữa của hai hiệu còn lại lần lượt là 20% và 25%.

a) Nếu có khách hàng mua một máy tính, tìm khả năng để máy tính của khách hàng đó phải đem lại sửa chữa trong thời hạn bảo hành.

b) Có một khách hàng mua máy tính mới 9 tháng đã phải đem lại vì có trục trặc, tính xác suất mà máy của khách này hiệu Toshiba.

▷ **1.27.** Hai máy cùng sản xuất 1 loại sản phẩm. Tỉ lệ phế phẩm của máy I là 3% và của máy II là 2%. Từ một kho gồm 2/3 sản phẩm của máy I và 1/3 sản phẩm của máy II lấy ngẫu nhiên ra 1 sản phẩm.

a) Tính xác suất để lấy được chính phẩm.

b) Biết sản phẩm lấy ra là phế phẩm. Tính xác suất để sản phẩm đó do máy I sản suất.

▷ **1.28.** Tỉ lệ người dân nghiện thuốc lá ở một vùng là 30%. Biết rằng người bị viêm họng trong số người nghiện thuốc lá là 60%, còn tỉ lệ người bị viêm họng trong số người không hút thuốc lá là 40%. Lấy ngẫu nhiên 1 người.

a) Biết người đó viêm họng, tính xác suất người đó nghiện thuốc.

b) Nếu người đó không bị viêm họng, tính xác suất người đó nghiện thuốc.

▷ **1.29.** Trong một trường đại học có 40% sinh viên học tiếng Anh, 30% sinh viên học tiếng Pháp, trong số sinh viên không học tiếng Anh có 45% sinh viên học tiếng Pháp. Chọn ngẫu nhiên 1 sinh viên, biết sinh viên đó học tiếng Pháp. Tính xác suất sinh viên đó học tiếng Anh.

▷ **1.30.** Có ba hộp bi bên ngoài giống hệt nhau. Hộp I có 6 trắng, 1 đen, 2 vàng; hộp II có 5 trắng, 2 đen, 3 vàng; hộp III có 4 trắng, 3 đen, 1 vàng. Lấy ngẫu nhiên 1 hộp rồi từ hộp đó lấy ngẫu nhiên 4 viên bi.

a) Tính xác suất 4 bi lấy ra có ít nhất 2 màu.

b) Giả sử 4 bi lấy ra cùng màu. Tính xác suất chọn được hộp I.

▷ **1.31.** Một thùng có 20 sản phẩm, trong đó có 3 sản phẩm loại I và 17 sản phẩm loại II. Trong quá trình vận chuyển bị mất 1 sản phẩm không rõ chất lượng. Lấy ngẫu nhiên 1 sản phẩm trong 19 sản phẩm còn lại.

a) Tính xác suất lấy được sản phẩm loại I.

b) Giả sử lấy được sản phẩm loại I. Tính xác suất lấy tiếp 2 sản phẩm nữa được 1 sản phẩm loại I và 1 sản phẩm loại II.

▷ **1.32.** Trong số 10 xạ thủ có 5 người bắn trúng bia với xác suất 0,9 (nhóm I); có 3 người bắn trúng bia với xác suất 0,8 (nhóm II) và 2 người bắn trúng bia với xác suất 0,7 (nhóm III). Chọn ngẫu nhiên một xạ thủ và cho anh ta bắn một viên đạn nhưng kết quả không trúng bia. Tính xác suất xạ thủ đó thuộc nhóm I?

▷ **1.33.** Một chiếc máy bay có thể xuất hiện ở vị trí A với xác suất 2/3 và ở vị trí B với xác suất 1/3. Có 3 phương án bố trí 4 khẩu pháo bắn máy bay như sau:

Phương án 1: 3 khẩu đặt tại A, 1 khẩu đặt tại B.

Phương án 2: 2 khẩu đặt tại A, 2 khẩu đặt tại B.

Phương án 3: 1 khẩu đặt tại A, 3 khẩu đặt tại B.

Biết rằng xác suất bắn trúng máy bay của mỗi khẩu pháo là 0,7 và các khẩu pháo hoạt động độc lập với nhau, hãy chọn phương án tốt nhất.

▷ **1.34.** Tỉ lệ sản phẩm bị lỗi của công ty A là 0,9%. Công ty sử dụng một bộ phận kiểm soát chất lượng sản phẩm, bộ phận đó xác định chính xác một sản phẩm bị lỗi với xác suất 99% và xác định sai một sản phẩm không bị lỗi với xác suất 0,5%. Sản phẩm được bộ phận kiểm soát chất lượng xác nhận không bị lỗi mới được bán ra thị trường.

a) Chọn ngẫu nhiên 1 sản phẩm sau khi kiểm soát chất lượng xác nhận không bị lỗi, tính xác suất sản phẩm đó không bị lỗi.

b) Chọn ngẫu nhiên 1 sản phẩm sau khi kiểm soát chất lượng xác nhận bị lỗi, tính xác suất sản phẩm đó không bị lỗi.

▷ **1.35.** Một phương pháp phân tích mới nhằm phát hiện chất gây ô nhiễm trong nước đang được một nhà sản xuất tiến hành thử nghiệm. Nếu thành công, phương pháp phân tích này sẽ phát hiện cùng một lúc 3 loại chất gây ô nhiễm có thể có trong nước: chất hữu cơ, các dung môi dễ bay hơi, các hợp chất Clo. Các kỹ sư cho rằng thí nghiệm mới này có thể phát hiện chính xác nguồn nước bị ô nhiễm bởi các chất hữu cơ với xác suất 99,7%, bởi các dung môi dễ bay hơi với xác suất 99,95% và bởi các hợp chất Clo với xác suất 89,7%. Còn nếu nguồn nước không bị ô nhiễm bởi ba loại trên thì phương pháp phân tích cho kết quả chính xác 100%. Các mẫu nước được chuẩn bị cho tiến hành thử nghiệm có 60% mẫu bị nhiễm các chất hữu cơ, 27% mẫu bị nhiễm các dung môi dễ bay hơi và 13% mẫu bị nhiễm các hợp chất Clo. Chọn ngẫu nhiên một mẫu để áp dụng.

a) Tính xác suất phương pháp phân tích cho kết quả mẫu bị ô nhiễm.

b) Giả sử phương pháp phân tích cho kết quả mẫu bị ô nhiễm. Tính xác suất phương pháp phân tích cho kết quả mẫu bị ô nhiễm bởi hợp chất Clo.

▷ **1.36.** Công ty A thường thăm dò ý kiến khách hàng trước khi đưa sản phẩm mới ra thị trường. Thông tin quá khứ cho thấy một sản phẩm rất thành công có 95% ý kiến thăm dò đánh giá tốt, một sản phẩm thành công vừa phải có 60% ý kiến thăm dò đánh giá tốt và một sản phẩm không thành công có 10% ý kiến thăm dò đánh giá tốt. Ngoài ra, công ty đã có 40% sản phẩm rất thành công, 35% sản phẩm thành công vừa phải và 25% sản phẩm không thành công. Tìm xác suất một sản phẩm có ý kiến thăm dò đánh giá tốt.

▷ **1.37.** Một thiết bị điện tử bao gồm 40 vi mạch độc lập. Xác suất một vi mạch bị lỗi là 0,01. Thiết bị này chỉ hoạt động khi không có vi mạch nào bị lỗi. Tính xác suất thiết bị này không hoạt động.

▷ **1.38.** Bắn ba viên đạn độc lập vào 1 mục tiêu. Xác suất trúng mục tiêu của mỗi viên lần lượt là 0,7; 0,8; 0,9. Biết rằng nếu chỉ 1 viên trúng hoặc 2 viên trúng thì mục tiêu bị phá hủy với xác suất lần lượt là 0,4 và 0,6. Còn nếu 3 viên trúng thì mục tiêu bị phá hủy. Tìm xác suất để mục tiêu bị phá hủy.

▷ **1.39.** Hai người cùng bắn vào một mục tiêu. Khả năng bắn trúng của từng người là 0,8 và 0,9. Tính xác suất:

a) chỉ có 1 người bắn trúng.
b) có người bắn trúng mục tiêu.
c) cả hai người bắn trượt.

▷ **1.40.** Một nồi hơi có 3 van bảo hiểm. Xác suất hỏng của van 1, 2 và 3 trong thời gian làm việc là 0,05; 0,05 và 0,06. Các van hoạt động độc lập. Nồi hơi gặp nguy hiểm nếu có 2 van bị hỏng. Tính xác suất nồi hơi hoạt động bình thường trong thời gian làm việc.

▷ **1.41.** Bắn liên tiếp vào mục tiêu đến khi nào có viên đạn trúng thì ngừng bắn. Tìm xác suất sao cho phải bắn đến viên đạn thứ 4, biết xác suất viện đạn trúng mục tiêu là 0,6 và các lần bắn độc lập với nhau.

▷ **1.42.** Tỉ lệ phế phẩm của một nhà máy là 5%. Tìm xác suất để trong 12 sản phẩm do nhà máy đó sản xuất ra:

a) Có đúng 2 phế phẩm.
b) Có không quá 2 phế phẩm.

▷ **1.43.** Đề thi trắc nghiệm có 10 câu hỏi, mỗi câu có 5 cách trả lời, trong đó chỉ có 1 cách trả lời đúng. Một thí sinh chọn cách trả lời một cách hoàn toàn hú họa. Tìm xác suất để thí sinh đó thi đỗ, biết rằng để thi đỗ phải trả lời đúng ít nhất 8 câu.

▷ **1.44.** Một người bắn bia với xác suất bắn trúng là 0,7.

a) Bắn liên tiếp 3 viên, tính xác suất để có ít nhất một lần trúng bia.

b) Hỏi phải bắn ít nhất mấy lần để có xác suất ít nhất 1 lần trúng bia không bé hơn 0,9?

▷ **1.45.** Một lô hàng có tỷ lệ phế phẩm là 5%, cần phải lấy mẫu cỡ bao nhiêu sao cho xác suất để có ít nhất một phế phẩm không bé hơn 0,95.

▷ **1.46.** Tín hiệu thông tin được phát 3 lần với xác suất thu được mỗi lần là 0,4.

a) Tìm xác suất để nguồn thu nhận được thông tin đó.

b) Cần phải phát tín hiệu ít nhất bao nhiêu lần để suất thu được lớn hơn 0,9.

Chương 2
BIẾN NGẪU NHIÊN

2.1. Biến ngẫu nhiên

Định nghĩa 2.1. Cho không gian mẫu Ω. *Biến ngẫu nhiên* là quy tắc X đặt tương ứng mỗi biến cố sơ cấp $\omega \in \Omega$ với một số thực duy nhất kí hiệu là $X(\omega)$. Tập tất cả các giá trị của X được gọi là *miền giá trị* của X và kí hiệu là $X(\Omega)$.

Sở dĩ ta gọi quy tắc X như trên là biến ngẫu nhiên là vì khi thực hiện phép thử ta sẽ được kết quả là một biến cố sơ cấp ω ngẫu nhiên và do đó $X(\omega)$ là một số thực ngẫu nhiên.

Ví dụ 2.1. Tung đồng thời 2 con xúc xắc. Gọi X là tổng số chấm trên mặt xuất hiện của hai con xúc xắc.

Ta có không gian mẫu $\Omega = \{(m; n) : m = 1, 2, ..., 6; n = 1, ..., 6\}$.

Khi đó X xác định bởi $X((m, n)) = m + n$.

Miền giá trị của X là $X(\Omega) = \{2, 3, ..., 12\}$.

Ví dụ 2.2. Tung một đồng xu cho đến khi nào xuất hiện mặt sấp thì dừng lại. Gọi X là số lần tung.

Kí hiệu hai mặt sấp và ngửa của đồng xu là S và N. Ta có không gian mẫu: $\Omega = \{S, NS, NNS, ...\}$.

Biến ngẫu nhiên X xác định như sau:

$$X(\{\underbrace{N...N}_{k-1 \text{ lần}} S\} = k, \quad k \geq 1$$

Ví dụ 2.3. Chọn ngẫu nhiên một sinh viên của trường đại học A, gọi X là chiều cao của sinh viên đó.

Ta có không gian mẫu $\Omega = \{$toàn bộ sinh viên của đại học A$\}$.

Khi đó với mỗi $sv \in \Omega$, $X(sv) =$ chiều cao của sv.

Chú ý 2.1. Để cho gọn trong trình bày, với $A \subset \mathbb{R}$, ta kí hiệu:

$$(X \in A) := \{\omega \in \Omega : X(\omega) \in A\}$$

Chẳng hạn:

$$(a < X \leq b) := \{\omega \in \Omega : a < X(\omega) \leq b\}$$
$$(X = a) := \{\omega \in \Omega : X(\omega) = a\}$$

2.2. Hai loại biến ngẫu nhiên

2.2.1. Biến ngẫu nhiên rời rạc

Định nghĩa 2.2. Nếu biến ngẫu nhiên X có miền giá trị hữu hạn hoặc vô hạn đếm được thì X được gọi là *biến ngẫu nhiên rời rạc*.

Giả sử biến ngẫu nhiên rời rạc X có miền giá trị $X(\Omega) = \{x_1, x_2, ...\}$, hàm số $p : \mathbb{R} \to \mathbb{R}$ xác định bởi:

$$p(x) = \begin{cases} P(X = x) & \text{nếu } x \in X(\Omega) \\ 0 & \text{nếu } x \notin X(\Omega) \end{cases}$$

được gọi là hàm xác suất (the probability mass function) của biến ngẫu nhiên X. Trong trường hợp $X(\Omega)$ hữu hạn thì ta có thể lập bảng các giá trị của $p(x)$ như sau:

x	x_1	x_2	...	x_n
$p(x)$	$P(X = x_1)$	$P(X = x_2)$	...	$P(X = x_n)$

Bảng trên được gọi là *bảng phân bố xác suất* của biến ngẫu nhiên X.

Ví dụ 2.4. Một hộp đựng 3 viên bi xanh và 4 viên bi đỏ, các viên bi giống nhau hoàn toàn về kích thước và khối lượng. Lấy ngẫu nhiên ra 3 viên bi, gọi X là số bi xanh có trong 3 viên bi lấy ra.

a) Lập bảng phân bố xác suất của X.
b) Tính xác suất $P(X \leq 1)$.

Giải.
a) $P(X = 0) = P(\text{lấy được 3 bi đỏ}) = \dfrac{C_4^3}{C_7^3} = \dfrac{4}{35}$.
Tương tự ta có:
$$P(X = 1) = \frac{18}{35},\ P(X = 2) = \frac{12}{35},\ P(X = 3) = \frac{1}{35}$$
Vì vậy, bảng phân bố xác suất của biến ngẫu nhiên X là:

x	0	1	2	3
$p(x)$	$\frac{4}{35}$	$\frac{18}{35}$	$\frac{12}{35}$	$\frac{1}{35}$

b) Hai biến cố $(X = 0)$ và $(X = 1)$ xung khắc nên:
$$P(X \leq 1) = P((X = 0) \cup (X = 1)) = P(X = 0) + P(X = 1) = 22/35.$$
■

Ví dụ 2.5. Tung một con xúc xắc cho đến khi xuất hiện mặt một chấm thì dừng lại. Gọi X là số lần tung.
a) Tìm hàm xác suất của biến ngẫu nhiên X.
b) Tính xác suất $P(2 \leq X \leq 5)$.

Giải.
a) Miền giá trị của X là $X(\Omega) = \{1, 2, ...\} = \mathbb{N}^*$.
Hàm xác suất của X là
$$p(k) = \frac{1}{6}\left(\frac{5}{6}\right)^{k-1}, \quad k \in X(\Omega)$$
b) Áp dụng **Tiên đề 3** ta có:
$$P(2 \leq X \leq 5) = \sum_{k=2}^{5} \frac{1}{6}\left(\frac{5}{6}\right)^{k-1} = \frac{3355}{7776}$$
■

Định lý 2.1. *Cho biến ngẫu nhiên X có miền giá trị $X(\Omega) = \{x_1, x_2, ..., x_n, ...\}$ và hàm xác suất là $p(x)$. Khi đó:*

1) $p(x) \geq 0$ với mọi x.

2) $P(a \leq X \leq b) = \sum\limits_{a \leq x_i \leq b} p(x_i)$.

3) $\sum\limits_{x_i \in X(\Omega)} p(x_i) = 1$.

2.2.2. Biến ngẫu nhiên liên tục

Định nghĩa 2.3. Nếu biến ngẫu nhiên X có miền giá trị là hợp một số khoảng trên trục số thì X được gọi là *biến ngẫu nhiên liên tục*.

Nếu tồn tại hàm số $y = f(x)$ thỏa mãn $f(x) \geq 0\ \forall x$ sao cho với mọi $a \leq b$ ta có:

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

thì $f(x)$ được gọi là *hàm mật độ xác suất* của X.

Định lý 2.2. *Cho biến ngẫu nhiên liên tục X có hàm mật độ xác suất $f(x)$. Khi đó:*

1) $P(X = a) = 0\ \forall\ a \in \mathbb{R}$.

2) $P(X < b) = P(X \leq b) = \int_{-\infty}^b f(x)dx$.

3) $P(X > a) = P(X \geq a) = \int_a^\infty f(x)dx$.

4) $\int\limits_{-\infty}^{+\infty} f(x)dx = 1$.

Ví dụ 2.6. Tuổi thọ (năm) của một loại thiết bị điện là biến ngẫu nhiên liên tục X có hàm mật độ xác suất:

$$f(x) = \begin{cases} 0,5.e^{-0,5x} & \text{nếu } x > 0 \\ 0 & \text{nếu } x \leq 0 \end{cases}$$

Chọn ngẫu nhiên một thiết bị điện loại trên. Tính xác suất:

a) Thiết bị đó có tuổi thọ thấp hơn 1 năm.
b) Thiết bị đó có tuổi thọ cao hơn 2 năm.

Giải.
a) $P(X < 1) = \int_{-\infty}^{0} 0dx + \int_{0}^{1} 0,5e^{-0,5x}dx = 1 - e^{-0,5}$.

b) $P(X > 2) = \int_{2}^{\infty} 0,5e^{-0,5x}dx = e^{-1}$. ■

Ví dụ 2.7. Cho biến ngẫu nhiên liên tục X có hàm mật độ xác suất:

$$f(x) = \begin{cases} kx^2 & \text{nếu } x \in [0;3] \\ 0 & \text{nếu } x \notin [0;3] \end{cases}$$

a) Tìm hằng số k.
b) Tính xác suất $P(|X| \le 1)$.

Giải.

a) Ta có $\int_{-\infty}^{\infty} f(x)dx = 1 \Leftrightarrow \int_{0}^{3} kx^2dx = 1 \Leftrightarrow k = 1/9$.

b)$P(|X| \le 1) = P(-1 \le X \le 1) = \int_{0}^{1} \dfrac{x^2}{9}dx = 1/27$. ■

2.3. Hàm phân phối xác suất

Định nghĩa 2.4. Cho biến ngẫu nhiên X, hàm số:

$$F(x) = P(X < x), x \in \mathbb{R}$$

được gọi là *hàm phân phối xác suất* của X.

1. Nếu biến ngẫu nhiên X có miền giá trị $\{x_1, x_2, ..., x_n, ...\}$ thì:

$$F(x) = \sum_{x_k < x} P(X = x_k) = \sum_{x_k < x} p(x_k)$$

2. Nếu biến ngẫu nhiên X có hàm mật độ xác suất $f(x)$ thì:

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

Ví dụ 2.8. Cho biến ngẫu nhiên X có bảng phân bố xác suất:

x	0	1	2
$p(x)$	$0,1$	$0,6$	$0,3$

Tìm hàm phân phối xác suất $F(x)$ của biến ngẫu nhiên X.

Giải. Hàm phân phối xác suất của biến ngẫu nhiên X là:

$$F(x) = \begin{cases} 0 & \text{nếu } x \leq 0 \\ 0,1 & \text{nếu } 0 < x \leq 1 \\ 0,7 & \text{nếu } 1 < x \leq 2 \\ 1 & \text{nếu } x > 2 \end{cases}$$

■

Ví dụ 2.9. Cho biến ngẫu nhiên liên tục X có hàm mật độ xác suất:

$$f(x) = \begin{cases} e^{-x} & \text{nếu } x \geq 0 \\ 0 & \text{nếu } x < 0 \end{cases}$$

Tìm hàm phân phối xác suất $F(x)$ của biến ngẫu nhiên X.

Giải. Hàm phân phối xác suất của biến ngẫu nhiên X là:

$$F(x) = \begin{cases} 0 & \text{nếu } x < 0 \\ 1 - e^{-x} & \text{nếu } x \geq 0 \end{cases}$$

■

Tính chất 2.1. *Hàm phân phối xác suất $F(x)$ của biến ngẫu nhiên X có một số tính chất sau:*

1) $0 \leq F(x) \leq 1 \, \forall \, x \in \mathbb{R}$, $\lim\limits_{x \to -\infty} F(x) = 0$, $\lim\limits_{x \to +\infty} F(x) = 1$.

2) Không giảm: nếu $x_1 \leq x_2$ *thì* $F(x_1) \leq F(x_2)$.

3) Nếu X là biến ngẫu nhiên liên tục có hàm mật độ xác suất $f(x)$ thì:

$$F'(x) = f(x)$$

2.4. Kì vọng

Định nghĩa 2.5. Cho biến ngẫu nhiên X xác định trên không gian mẫu Ω. *Kì vọng* của biến ngẫu nhiên X, kí hiệu là $E(X)$, được xác định như sau:

1. Nếu biến ngẫu nhiên rời rạc X có hàm xác suất $p(x)$ thì:

$$E(X) = \sum_{x_k \in X(\Omega)} x_k p(x_k)$$

2. Nếu biến ngẫu nhiên liên tục X có hàm mật độ xác suất $f(x)$ thì:

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

Tính chất 2.2.

1) Nếu $X = C$ là hằng số thì $E(C) = C$.

2) Nếu $a, b \in \mathbb{R}$ và X, Y là hai biến ngẫu nhiên cùng xác định trên không gian mẫu Ω thì:

$$E(aX + b) = aE(X) + b \text{ và } E(X \pm Y) = E(X) \pm E(Y)$$

Ví dụ 2.10. Tính kì vọng của các biến ngẫu nhiên trong Ví dụ 2.4 và Ví dụ 2.7.

2.5. Phương sai và độ lệch chuẩn

Định nghĩa 2.6. Cho biến ngẫu nhiên X. Khi đó, đại lượng:

$$V(X) = E(X - E(X))^2$$

được gọi là *phương sai* của X, $SD(X) = \sqrt{V(X)}$ được gọi là *độ lệch chuẩn* của X.

Tính chất 2.3.

1) $V(X) \geq 0$, $V(X) = 0$ khi và chỉ khi $X = C$ (hằng số).

2) $V(X) = E(X^2) - (E(X))^2$.

3) $V(aX + b) = a^2 V(X)$ với mọi $a, b \in \mathbb{R}$.

Định lý 2.3.

1) Nếu biến ngẫu nhiên rời rạc X có hàm xác suất p(x) thì:

$$V(X) = \sum_{x_k \in X(\Omega)} x_k^2 p(x_k) - \left(\sum_{x_k \in X(\Omega)} x_k p(x_k) \right)^2$$

2) Nếu biến ngẫu nhiên liên tục X có hàm mật độ xác suất f(x) thì:

$$V(X) = \int_{-\infty}^{+\infty} x^2 f(x)dx - \left(\int_{-\infty}^{+\infty} x f(x)dx \right)^2$$

Ý nghĩa. Phương sai dùng để đo độ phân tán các giá trị của biến ngẫu nhiên quanh kỳ vọng của nó. Phương sai càng lớn thì độ phân tán càng rộng.

Ví dụ 2.11. Tính phương sai và độ lệch chuẩn của các biến ngẫu nhiên trong Ví dụ 2.4 và Ví dụ 2.7.

2.6. Trung vị

Định nghĩa 2.7. Số thực m được gọi là *trung vị* của biến ngẫu nhiên X nếu:

$$P(X < m) \leq 0,5 \text{ và } P(X > m) \leq 0,5$$

Kí hiệu $med(X) = m$.

Ví dụ 2.12. Tìm trung vị của biến ngẫu nhiên X có bảng phân bố xác suất:

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

Giải. Vì $P(X < 2) = 0,4 < 0,5$ và $P(X > 2) = 0,2 < 0,5$ nên $Med(X) = 2$. ■

Ví dụ 2.13. Tìm trung vị của biến ngẫu nhiên X có bảng phân bố xác suất:

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

Giải.
$Med(X) = m \in [1; 2]$ vì $P(X < m) = 0,5$ và $P(X > m) = 0,5$ với mọi $m \in [1; 2]$. ■

Định lý 2.4. *Nếu biến ngẫu nhiên X có hàm phân phối xác suất $F(x)$ liên tục trên $\mathbb{R}$ thì trung vị là nghiệm phương trình $F(x) = 0,5$.*

Chứng minh. Thật vậy, $P(X < m) \leq 0,5$ tương đương với

$$F(m) \leq 0,5 \tag{2.1}$$

Mặt khác:

$$P(X > m) \leq 0,5 \Leftrightarrow P(X \leq m) \geq 0,5$$

Do X là biến ngẫu nhiên có hàm phân phối xác suất liên tục nên bất đẳng thức trên tương đương với:

$$P(X < m) \geq 0,5$$

Hay:

$$F(m) \geq 0,5 \tag{2.2}$$

Kết hợp (2.1) và (2.2) ta được $F(m) = 0,5$. ■

Ví dụ 2.14. Tìm trung vị của biến ngẫu nhiên X có hàm mật độ xác suất

$$f(x) = \begin{cases} e^{-x} & \text{nếu } x \geq 0 \\ 0 & \text{nếu } x < 0 \end{cases}$$

Giải. Hàm phân phối xác suất của biến ngẫu nhiên X là:

$$F(x) = \begin{cases} 0 & \text{nếu } x < 0 \\ 1 - e^{-x} & \text{nếu } x \geq 0 \end{cases}$$

Ta có:

$$F(x) = 0,5 \Leftrightarrow \begin{cases} 1 - e^{-x} = 0,5 \\ x \geq 0 \end{cases} \Leftrightarrow x = \ln 2$$

Vậy $Med(X) = \ln 2$. ■

2.7. Biến ngẫu nhiên độc lập

Định nghĩa 2.8. Các biến ngẫu nhiên $X_1, X_2, ..., X_n$ $(n \geq 2)$ được gọi là *độc lập* nếu với mọi $x_1, x_2, ..., x_n \in \mathbb{R}$ ta có:

$$P(\bigcap_{k=1}^{n}\{X_k < x_k\}) = P(\{X_1 < x_1\})P(\{X_2 < x_2\})...P(\{X_n < x_n\})$$

Định lý 2.5. *Nếu hai biến ngẫu nhiên X và Y độc lập thì:*

1) $E(XY) = E(X)E(Y)$.

2) $V(X \pm Y) = V(X) + V(Y)$.

2.8. Một số phân bố xác suất quan trọng

2.8.1. Phân bố Bernoulli

Định nghĩa 2.9. Biến ngẫu nhiên rời rạc X có *phân bố Bernoulli* với tham số p $(0 < p < 1)$ nếu X có miền giá trị $X(\Omega) = \{0, 1\}$ và hàm xác suất:

$$p(k) = P(X = k) = \begin{cases} 1 - p & \text{nếu } k = 0 \\ p & \text{nếu } k = 1 \\ 0 & \text{nếu } k \notin \{0, 1\} \end{cases}$$

Kí hiệu: $X \sim Ber(p)$.

Tính chất 2.4. *Nếu $X \sim Ber(p)$ thì $E(X) = p$ và $V(X) = p(1 - p)$.*

2.8.2. Phân bố nhị thức

Định nghĩa 2.10. Biến ngẫu nhiên rời rạc X có *phân bố nhị thức* với tham số n và p ($n \in \mathbb{N}^*$ và $0 < p < 1$) nếu X có miền giá trị $X(\Omega) = \{0, 1, ..., n\}$ và hàm xác suất:

$$p(k) = C_n^k p^k (1 - p)^{n-k}, k \in X(\Omega).$$

Kí hiệu: $X \sim B(n, p)$.

Tính chất 2.5.

1) Nếu $X \sim B(n,p)$ *thì* $E(X) = np$ *và* $V(X) = np(1-p)$.

2) Nếu $X_1, X_2, ..., X_n$ *là n biến ngẫu nhiên độc lập, cùng phân bố xác suất với* $X \sim Ber(p)$ *thì biến ngẫu nhiên* $T = X_1 + X_2 + ... + X_n$ *có phân bố nhị thức* $B(n,p)$.

Ví dụ 2.15. Tỉ lệ phế phẩm của một nhà máy là 12%. Các sản phẩm của nhà máy được đóng gói thành từng hộp, mỗi hộp 20 sản phẩm.

a) Trung bình mỗi hộp chứa bao nhiêu phế phẩm? Tính độ lệch chuẩn số phế phẩm trong mỗi hộp.

b) Một khách hàng mua ngẫu nhiên một hộp sản phẩm. Tính xác suất hộp đó có chứa phế phẩm.

c) Tìm số phế phẩm trong hộp có xác suất lớn nhất.

Giải. Gọi X là số phế phẩm trong mỗi hộp. Khi đó, $X \sim B(20; 0,12)$.

a) $E(X) = np = 2,4$; $SD(X) = \sqrt{np(1-p)} \approx 1,45$.

b) $P(X \geq 1) = 1 - P(X = 0) = 1 - 0,88^{20}$.

c) $(n+1)p = 2,52 \notin \mathbb{Z}$ nên $P(X = k) = C_{20}^{k}.0,12^{k}.0,88^{20-k}$ đạt giá trị lớn nhất tại $k = 2$. ■

2.8.3. Phân bố Poisson

Định nghĩa 2.11. Biến ngẫu nhiên rời rạc X có *phân bố Poisson* với tham số λ ($\lambda > 0$) nếu X có miền giá trị $\mathbb{N} = \{0, 1, 2, ...\}$ và hàm xác suất:

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

Kí hiệu: $X \sim Poi(\lambda)$.

Phân bố Poisson thường gặp thể hiện phân bố số lần xuất hiện 1 biến cố nào đó trong một khoảng thời gian T.

Tính chất 2.6.

1) Nếu $X \sim Poi(\lambda)$ *thì* $E(X) = \lambda$, $V(X) = \lambda$.

2) Nếu $X_1, X_2, ..., X_n$ là n biến ngẫu nhiên độc lập, cùng phân bố với $X \sim Poi(\lambda)$ thì biến ngẫu nhiên $T = X_1 + X_2 + ... + X_n$ có phân bố Poisson $Poi(n\lambda)$.

Ví dụ 2.16. Một gara cho thuê xe ôtô có 2 ôtô loại A. Số đơn đặt hàng ôtô loại này vào ngày cuối tuần có phân bố Poisson với số đơn trung bình 2 đơn/ngày. Tính xác suất trong ngày cuối tuần:

a) có một ôtô loại A được thuê.

b) có 2 ôtô loại A được thuê.

c) gara không đáp ứng nhu cầu thuê ôtô loại này.

Giải. Gọi X là số đơn đặt hàng thuê ô tô ngày cuối tuần của gara. Ta có $X \sim Poi(2)$ (do $E(X) = \lambda = 2$).

a) $P(X = 1) = e^{-2}\dfrac{2^1}{1!} \approx 0,27$.

b) $P(X \geq 2) = 1 - P(X < 2) = 1 - P(X = 0) - P(X = 1) = 1 - e^{-2} - e^{-2}\dfrac{2^1}{1!} \approx 0,59$.

c) $P(X > 2) = 1 - P(X \leq 2) \approx 0,32$. ■

Ví dụ 2.17. Ở một tổng đài bưu điện, số cuộc điện thoại gọi đến xuất hiện là biến ngẫu nhiên có phân bố Poisson với số cuộc điện thoại trung bình là 2 cuộc gọi trong 1 phút. Tính xác suất có đúng 5 cuộc trong khoảng thời gian 1 phút.

Giải. Gọi X là số cuộc điện thoại gọi đến trong khoảng thời gian 1 phút, theo giả thiết, X có phân bố Poisson. Vì $E(X) = 2$ nên $\lambda = 2$. Do đó:

$$P(X = 5) = e^{-2}.\frac{2^5}{5!} \approx 0,036$$

■

Định lý 2.6. *(Luật biến cố hiếm) Cho $\{X_n; n \geq 1\}$ là dãy biến ngẫu nhiên có phân bố nhị thức $X_n \sim B(n; p_n)$. Nếu tồn tại giới hạn $\lim\limits_{n\to\infty} np_n = \lambda$ thì:*

$$\lim_{n\to\infty} P(X_n = k) = e^{-\lambda}\frac{\lambda^k}{k!},\ k = 0, 1, 2, ...$$

Chứng minh. Ta có:

$$\begin{aligned}P(X_n = k) &= C_n^k p_n^k (1-p_n)^{n-k}\\ &= \frac{n(n-1)...(n-k+1)}{k!} p_n^k (1-p_n)^{n-k}\\ &= \frac{(np_n)^k}{k!}\left(1-\frac{1}{n}\right)...\left(1-\frac{k-1}{n}\right)(1-p_n)^n(1-p_n)^{-k}\end{aligned}$$

Ta lại có:

$$\lim_{n\to\infty} \frac{(np_n)^k}{k!} = \frac{\lambda^k}{k!}$$

Đặt $\lambda_n = np_n$. Khi đó:

$$\lim_{n\to\infty} (1-p_n)^n = \lim_{n\to\infty}\left[\left(1-\frac{\lambda_n}{n}\right)^{\frac{n}{\lambda_n}}\right]^{-\lambda_n} = e^{-\lambda}$$

Các thừa số khác có giới hạn bằng 1. Từ đó ta có điều phải chứng minh. ■

Ứng dụng: Nếu $X \sim B(n;p)$ với n khá lớn và p khá bé thì X có xấp xỉ phân bố Poisson với $\lambda = np$, tức là:

$$P(X = k) = C_n^k p^k (1-p)^{n-k} \approx \frac{\lambda^k}{k!} e^{-\lambda}$$

Ví dụ 2.18. Tỉ lệ phế phẩm của một nhà máy là 0,006. Lấy ngẫu nhiên 1.000 sản phẩm của nhà máy, tính xác suất có đúng 9 phế phẩm.

Giải. Gọi X là số phế phẩm trong 1.000 sản phẩm, khi đó $X \sim B(1.000; 0,006)$. Vì $n = 1.000$ khá lớn và $p = 0,006$ khá bé nên ta có thể tính bằng xấp xỉ phân bố Poisson với $\lambda = np = 6$:

$$P(X = 9) \approx e^{-6}.\frac{6^9}{9!} \approx 0,069$$

■

Ví dụ 2.19. Một xưởng in sách thấy rằng trung bình một cuốn sách 500 trang có chứa 300 lỗi. Tìm xác suất trong một trang:

a) Có đúng 2 lỗi.

b) Có ít nhất 2 lỗi.

Giải. Gọi p là xác suất một chữ bị lỗi, X là số lỗi trong 1 trang có n chữ. Khi đó $X \sim B(n;p)$ và $E(X) = np = 300/500 = 0,6$. Vì xác suất 1 chữ bị lỗi rất nhỏ và số chữ trong 1 trang rất lớn nên có thể xấp xỉ X bởi phân bố Poisson với $\lambda = 0,6$. Do đó:

a) $P(X = 2) = \dfrac{0,6^2}{2!}e^{-0,6} \approx 0,099$.

b) $PX \geq 2) = 1 - P(X < 2) \approx 1 - 0,549 - 0,359 = 0,122$. ■

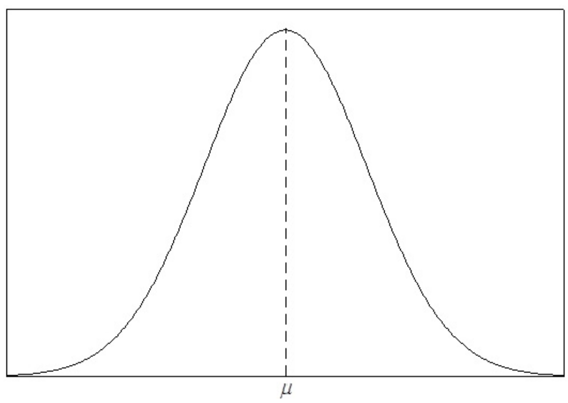
2.8.4. Phân bố chuẩn

Định nghĩa 2.12. Biến ngẫu nhiên liên tục X có *phân bố chuẩn* với tham số μ và σ ($-\infty < \mu < +\infty$, $\sigma > 0$) nếu có hàm mật độ xác suất:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}},\ x \in \mathbb{R}$$

Kí hiệu $X \sim N(\mu, \sigma^2)$.

Dưới đây là hình dáng đồ thị của hàm mật độ xác suất $f(x)$:



Hình 2.1

Phân bố chuẩn tắc

Biến ngẫu nhiên phân bố chuẩn với $\mu = 0$ và $\sigma = 1$ được gọi là *phân bố chuẩn tắc* và kí hiệu là Z. Khi đó, hàm mật độ xác suất được kí hiệu là $\varphi(x)$,

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

Hàm phân bố xác suất được kí hiệu là $\Phi(x)$,

$$\Phi(x) = \int\limits_{-\infty}^{x} \varphi(t)dt = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$$

Chú ý rằng $\Phi(-x) = 1 - \Phi(x), \forall\, x \in \mathbb{R}$.

Tính $\Phi(x)$ bằng máy tính Casio

1) CASIO FX570MS:
- Vào Mode tìm SD: Mode $\rightarrow$ Mode $\rightarrow$ 1 (SD);
- Shift $\rightarrow$ 3 (Distr) $\rightarrow$ 1;
- Nhập x.

2) CASIO FX570ES, FX570ES - PLUS, FX570VN - PLUS:
- Vào Mode tìm 1-Var: Mode $\rightarrow$ 3 (Stat) $\rightarrow$ 1 (1-Var)$\rightarrow$ AC;
- Shift $\rightarrow$ 1(Stat) $\rightarrow$ 5 (Distr) $\rightarrow$ 1;
- Nhập x.

Tính hàm ngược $\Phi^{-1}(y)$ bằng máy tính CASIO FX570VN PLUS

- Mode $\rightarrow$ 3 (DIST) $\rightarrow$ 3;
- Nhập $y \rightarrow = \rightarrow \sigma = 1 = \rightarrow \mu = 0 =$.

Ví dụ 2.20. Tính $\Phi(1,96), \Phi(-1,65)$

Tính chất 2.7. *Cho biến ngẫu nhiên $X \sim N(\mu, \sigma^2)$. Khi đó:*

1) $E(X) = \mu$, $V(X) = \sigma^2$.

2) $Z = \dfrac{X - \mu}{\sigma} \sim N(0; 1)$.

3) Nếu $X_1, X_2, ..., X_n$ là các biến ngẫu nhiên độc lập, cùng phân bố xác suất với $X \sim N(\mu; \sigma^2)$ thì:

$$T = X_1 + X_2 + ... + X_n \sim N(n\mu; n\sigma^2)$$

và

$$\overline{X} = \frac{X_1 + X_2 + ... + X_n}{n} \sim N(\mu; \sigma^2/n)$$

(4) $P(X < a) = P(X \leq a) = \Phi(\frac{a-\mu}{\sigma})$.

(5) Với $\alpha < \beta$ *ta có:*

$$P(\alpha < X < \beta) = P(\alpha \leq X \leq \beta) = \Phi(\frac{\beta-\mu}{\sigma}) - \Phi(\frac{\alpha-\mu}{\sigma})$$

Ví dụ 2.21. Cho biến ngẫu nhiên liên tục $X \sim N(1; 4)$. Tính $P(X < 3,5)$, $P(X > 0)$, $P(0,5 < X \leq 2,5)$.

Giải.

$P(X < 3,5) = \Phi(\frac{3,5-1}{2}) = \Phi(1,25) = 0,8944$;

$P(X > 0) = 1 - P(X \leq 0) = 1 - \Phi(-0,5) = \Phi(0,5) = 0,6915$;

$P(0,5 < X \leq 2,5) = \Phi(0,75) - \Phi(-0,25) = 0,3721$. ■

Ví dụ 2.22. Giả sử số đo chiều dài của một sợi dây kim loại do một máy tự động cắt ra là một biến ngẫu nhiên chuẩn với $\mu = 10$mm, $\sigma^2 = 4\text{mm}^2$.

a) Tính xác suất lấy ra được một sợi dây có chiều dài lớn hơn 13mm.

b) Tìm tỉ lệ sợi dây do máy cắt ra có chiều dài từ $8,5$mm đến $12,5$mm.

Giải.

a) $P(X > 13) = 1 - P(X \leq 13) = 1 - \Phi(1,5) = 0,067$.

b) $P(8,5 \leq X \leq 12,5) = \Phi(1,25) - \Phi(-0,75) = 0,668$. ■

Ví dụ 2.23. Đường kính của một trục trong ổ đĩa quang là một biến ngẫu nhiên chuẩn với đường kính trung bình là $0,2508$inch và độ lệch chuẩn $0,0005$inch. Thông số kỹ thuật ghi trên trục là $0,25 \pm 0,0015$inch. Tìm tỉ lệ trục có đường kính phù hợp với thông số kỹ thuật.

Giải. Gọi X là đường kính của trục ổ đĩa quang, ta có:

$$X \sim N(0,2508; 0,0005^2)$$

$$\begin{aligned} P(0,25 - 0,0015 \leq X \leq 0,25 + 0,0015) &= \Phi(1,4) - \Phi(-4,6) \\ &= 0,919 \end{aligned}$$

■

Ví dụ 2.24. Tỉ lệ lợi nhuận X (%) của một dự án đầu tư được xem là một biến ngẫu nhiên có phân bố chuẩn. Theo đánh giá của ủy ban đầu tư thì với xác suất 0,1587 cho tỉ lệ lợi nhuận cao hơn 20% và với xác suất 0,0228 cho tỉ lệ lợi nhuận lớn hơn 25%. Tìm xác suất $P(X > 0)$.

Giải. Gọi X là lãi suất đầu tư vào 1 dự án trong 1 năm, khi đó $X \sim N(\mu; \sigma^2)$. Từ giả thiết ta có:

$$\begin{cases} P(X > 0,2) = 0,1587 \\ P(X > 0,25) = 0,0228 \end{cases} \Leftrightarrow \begin{cases} 1 - \Phi(\dfrac{0,2 - \mu}{\sigma}) = 0,1587 \\ \\ 1 - \Phi(\dfrac{0,25 - \mu}{\sigma}) = 0,0228 \end{cases}$$

$$\Leftrightarrow \begin{cases} \Phi(\dfrac{0,2 - \mu}{\sigma}) = 0,8413 = \Phi(1) \\ \\ \Phi(\dfrac{0,25 - \mu}{\sigma}) = 0,9772 = \Phi(2) \end{cases} \Leftrightarrow \begin{cases} \mu = 0,15 \\ \sigma = 0,05 \end{cases}$$

Vì vậy:

$$P(X > 0) = 1 - \Phi(\frac{0 - 0,15}{0,05}) = \Phi(3) = 0,9987$$

■

Ví dụ 2.25. Chiều cao X (mét) của nam thanh niên trưởng thành ở quốc gia A tuân theo quy luật phân bố chuẩn $N(\mu; 0,1^2)$. Chọn ngẫu nhiên 100 nam thanh niên của quốc gia A. Tính xác suất sai số tuyệt đối giữa chiều cao trung bình của 100 nam thanh niên được chọn với μ không vượt quá $0,03$.

Giải.

Gọi X_k là chiều cao của nam thanh niên thứ k ($k = 1, 2, ..., 100$). Khi đó:

$$\overline{X} = \frac{X_1 + X_2 + ... + X_{100}}{100}$$

là chiều cao trung bình của 100 nam thanh niên được chọn. Vì $\overline{X} \sim N(\mu; 0,01^2)$ nên ta có:

$$P(|\overline{X} - \mu| \leq 0,03) = 2\Phi(3) - 1 = 0,9974$$

Như vậy, khi chọn ngẫu nhiên 100 nam thanh niên thì hầu như chắc chắn rằng chiều cao trung bình của 100 nam thanh niên đó rơi vào đoạn $[\mu - 0,03; \mu + 0,03]$. ■

2.8.5. Phân bố đều

Định nghĩa 2.13. Biến ngẫu nhiên liên tục X có *phân bố đều* trên đoạn $[a; b]$ ($a < b$) nếu có hàm mật độ xác suất:

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{nếu} \quad x \in [a; b] \\ 0 & \text{nếu} \quad x \notin [a; b] \end{cases}$$

Kí hiệu: $X \sim U([a; b])$.

Tính chất 2.8. *Nếu* $X \sim U([a; b])$ *thì* $E(X) = \dfrac{a+b}{2}$, $V(X) = \dfrac{(b-a)^2}{12}$.

2.8.6. Phân bố mũ

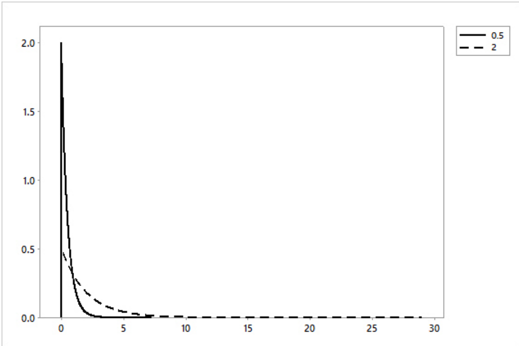
Định nghĩa 2.14. Biến ngẫu nhiên liên tục X có *phân bố mũ* với tham số λ ($\lambda > 0$) nếu có hàm mật độ

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{nếu} \quad x \geq 0 \\ 0 & \text{nếu} \quad x < 0 \end{cases}$$

Kí hiệu: $X \sim Exp(\lambda)$.

Trong cuộc sống, phân bố mũ thể hiện phân bố thời gian sống của các đối tượng,...

Tính chất 2.9. *Nếu* $X \sim Exp(\lambda)$ *thì* $E(X) = 1/\lambda$, $V(X) = 1/\lambda^2$.



Hình 2.2: *Phân bố mũ* ($\lambda = 0,5$ *và* $\lambda = 2$))

Ví dụ 2.26. Giả sử tuổi thọ (X) của một chiếc quạt trong máy tính là một biến ngẫu nhiên phân bố mũ với tuổi thọ trung bình là 3.300 giờ. Tính xác suất:

a) Chiếc quạt hỏng trước 10.000 giờ.

b) Chiếc quạt có tuổi thọ lớn hơn 7.000 giờ.

Giải. Theo giả thiết $E(X) = \dfrac{1}{\lambda} = 3.300$ nên:

a) $P(X < 10.000) = \int\limits_0^{10.000} \dfrac{1}{3.300} e^{-x/3.300} dx \approx 0,952.$

b) $P(X > 7.000) = 1 - P(X \leq 7.000) = \int\limits_0^{7.000} \dfrac{1}{3.300} e^{-x/3.300} dx \approx$ $0,88.$ ■

2.8.7. Phân bố Student (t-distribution)

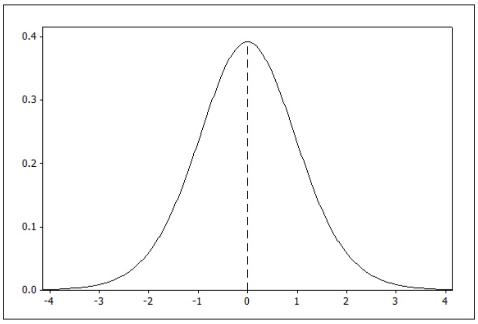
Định nghĩa 2.15. Biến ngẫu nhiên liên tục X có *phân bố Student* (t-phân bố) với n bậc tự do nếu có hàm mật độ xác suất:

$$f_n(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\,\Gamma(\frac{n}{2})} \left[1 + \frac{x^2}{n}\right]^{-\frac{n+1}{2}} \qquad \forall x \in \mathbb{R}$$

Trong đó $\Gamma(.)$ là hàm Gamma được xác định bởi: $\Gamma(x) = \int_0^\infty u^{x-1} e^{-u} du$.

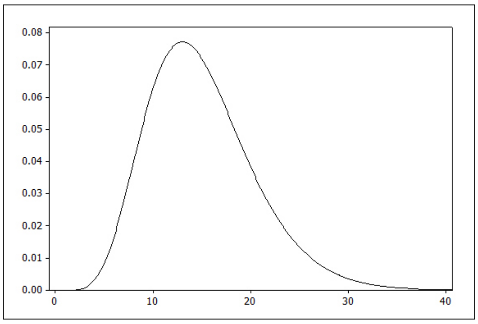
Kí hiệu $X \sim T_n$.

Đồ thị hàm mật độ $f_n(x)$ của phân bố T_n có dạng như Hình 2.3 ($f_n(x)$ là hàm số chẵn).



Hình 2.3: *Phân bố Student* $(n = 15)$*)*

2.8.8. Phân bố khi bình phương



Hình 2.4: *Phân bố khi bình phương* $(n = 15)$*)*

Định nghĩa 2.16. Biến ngẫu nhiên liên tục X có phân bố *khi bình phương* n bậc tự do nếu có hàm mật độ xác suất:

$$f_n(x) = \begin{cases} \dfrac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} & \text{nếu } x > 0 \\ 0 & \text{nếu } x \leq 0 \end{cases}$$

Kí hiệu $X \sim \chi_n^2$.

Đồ thị hàm mật độ $f_n(x)$ của phân bố χ_n^2 có dạng như Hình 2.4.

2.8.9. Phân bố F

Định nghĩa 2.17. Cho W và Y là hai biến ngẫu nhiên độc lập có phân bố khi bình phương với bậc tự do lần lượt là m và n. Khi đó tỉ số:

$$F = \frac{W/m}{Y/n}$$

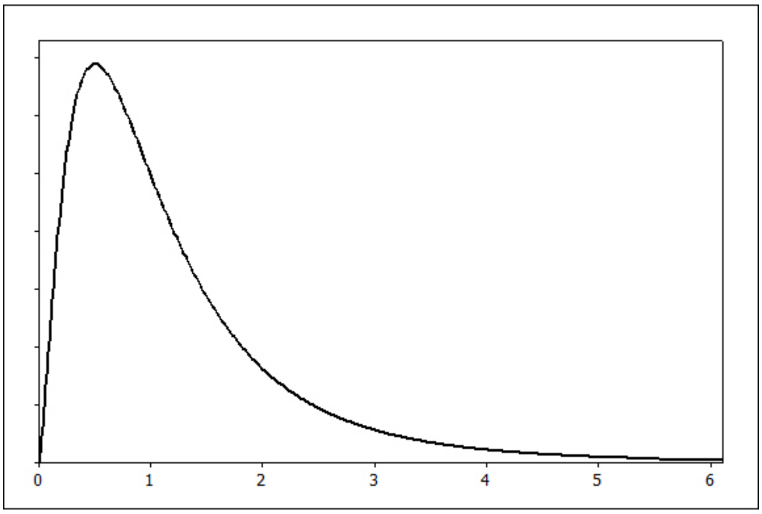
có hàm mật độ xác suất là:

$$f_{m,n}(x) = \begin{cases} \dfrac{\Gamma(\dfrac{m+n}{2}) \left(\dfrac{m}{n}\right)^{m/2} x^{(m/2)-1}}{\Gamma(\dfrac{m}{2})\Gamma(\dfrac{n}{2}) \left[\dfrac{m}{n}x + 1\right]^{(m+n)/2}} & \text{nếu } x > 0 \\ 0 & \text{nếu } x \leq 0 \end{cases}$$

và được gọi là *phân bố F* với hai bậc tự do m và n.

Kí hiệu $F \sim F_{m,n}$.

Đồ thị hàm mật độ $f_{m,n}(x)$ của phân bố F có dạng như Hình 2.5.



Hình 2.5: *Phân bố F ($m = 5$, $n = 10$))*

2.9. Các định lí giới hạn

2.9.1. Luật số lớn

Định lý 2.7 (Bất đẳng thức Markov)**.** *Nếu X là biến ngẫu nhiên không âm thì với số thực dương bất kì $a > 0$ ta có:*

$$P(X \geq a) \leq \frac{E(X)}{a}$$

Chứng minh. Với $a > 0$, xét biến ngẫu nhiên I có bảng phân bố xác suất như sau:

x	0	1
$p(x)$	$P(X < a)$	$P(X \geq a)$

Khi đó: $I \leq \dfrac{X}{a}$.

Lấy kì vọng hai vế ta được: $E(I) \leq \dfrac{E(X)}{a}$.

Vì $E(I) = P(X \geq a)$ nên:

$$P(X \geq a) \leq \frac{E(X)}{a}$$

■

Định lý 2.8 (Bất đẳng thức Chebyshev)**.** *Cho X là biến ngẫu nhiên. Khi đó, với mọi $\varepsilon > 0$ ta có:*

$$P(|X - E(X)| > \varepsilon) \leq \frac{V(X)}{\varepsilon^2}$$

Chứng minh. Với $\varepsilon > 0$ tùy ý, đặt $a = \varepsilon^2$, $Y = |X - E(X)|^2$. Áp dụng bất đẳng thức Markov ta được điều phải chứng minh. ■

Định lý 2.9 (Luật yếu số lớn)**.** *Nếu $\{X_n, n \geq 1\}$ là dãy các biến ngẫu nhiên độc lập, cùng phân bố xác suất với biến ngẫu nhiên X có kì vọng $E(X) = \mu$ hữu hạn và phương sai $V(X) = \sigma^2$ hữu hạn thì với mọi $\varepsilon > 0$,*

$$\lim_{n \to \infty} P(|\overline{X} - \mu| < \varepsilon) = 1$$

trong đó $\overline{X} = \dfrac{1}{n} \sum_{k=1}^{n} X_k$.

Chứng minh. Đặt $T = X_1 + X_2 + ... + X_n$. Do các biến ngẫu nhiên X_k độc lập, cùng phân bố xác suất nên $E(T) = n\mu$ và $V(T) = n\sigma^2$. Áp dụng bất đẳng thức Chebyshev, ta có:

$$P(|\frac{T}{n} - \mu| \geq \varepsilon) = P(|T - E(T)| \geq n\varepsilon) \leq \frac{V(T)}{(n\varepsilon)^2} = \frac{\sigma^2}{n\varepsilon^2} \to 0$$

Suy ra $\lim\limits_{n\to\infty} P(|\overline{X} - \mu| < \varepsilon) = 1$. ■

Ý nghĩa của luật số lớn: Nếu $X_1, X_2, ..., X_n$ là các biến ngẫu nhiên độc lập, cùng phân bố xác suất với biến ngẫu nhiên X thì với n đủ lớn ta có:

$$\overline{X} = \frac{X_1 + X_2 + ... + X_n}{n} \approx E(X)$$

Hệ quả 2.1. *Dãy $\{X_n, n \geq 1\}$ các biến ngẫu nhiên độc lập, cùng phân bố xác suất với biến ngẫu nhiên X có kì vọng $E(X) = \mu$ hữu hạn và phương sai $V(X) = \sigma^2$ hữu hạn thì với mọi $\varepsilon > 0$,*

$$\lim_{n\to\infty} P(|S^2 - \sigma^2| < \varepsilon) = 1$$

trong đó $S^2 = \frac{1}{n}\sum_{k=1}^{n}(X_k - \overline{X})^2$.

2.9.2. Định lí giới hạn trung tâm

Định lý 2.10. *Nếu $\{X_n, n \geq 1\}$ là dãy các biến ngẫu nhiên độc lập, cùng phân bố xác suất với biến ngẫu nhiên X có kì vọng $E(X) = \mu$ hữu hạn và phương sai $V(X) = \sigma^2$ hữu hạn thì:*

$$\lim_{n\to\infty} P(\frac{(X_1 + ... + X_n) - n\mu}{\sqrt{n}\sigma} < x) = \Phi(x) \quad x \in \mathbb{R}$$

Ý nghĩa Định lí giới hạn trung tâm: Nếu $X_1, X_2, ..., X_n$ là các biến ngẫu nhiên độc lập, cùng phân bố xác suất (không cần thiết có phân bố chuẩn) thì với n đủ lớn ta có:

$$T = X_1 + X_2 + ... + X_n$$

có phân bố xấp xỉ phân bố chuẩn $N(n\mu; n\sigma^2)$.

Ví dụ 2.27. Tuổi thọ làm việc của một linh kiện điện tử là một biến ngẫu nhiên X có kì vọng 250 giờ và độ lệch chuẩn là 250 giờ. Tính xác suất 100 linh kiện được chọn ngẫu nhiên có tổng tuổi thọ ít nhất 1 năm (365 ngày).

Giải. Gọi X_k là tuổi thọ của linh kiện thứ k ($1 \leq k \leq 100$), khi đó các biến ngẫu nhiên $X_1, X_2, ..., X_{100}$ độc lập, cùng phân bố xác suất với X. Theo Định lí giới hạn trung tâm ta có:

$$T = X_1 + X_2 + ... + X_{100}$$

có phân bố xấp xỉ phân bố chuẩn $N(100 \times 250; 100 \times 250^2)$. Do đó:

$$P(T \geq 365 \times 24) = 1 - P(T < 8.760) \approx 1 - \Phi(-6,496) \approx 1$$

■

Hệ quả 2.2. *(Định lý giới hạn tích phân Moivre-Laplace) Giả sử X_n là biến ngẫu nhiên có phân bố nhị thức $B(n;p)$. Đặt:*

$$Z_n = \frac{X_n - np}{\sqrt{np(1-p)}}$$

Khi đó với mọi $x \in \mathbb{R}$,

$$\lim_{n\to\infty} P(Z_n < x) = \Phi(x)$$

Nói cách khác, với n đủ lớn, $B(n;p)$ có phân bố xấp xỉ phân bố chuẩn $N(np; np(1-p))$.

Xấp xỉ trên tốt nhất khi $np > 5$ và $n(1-p) > 5$.

Ví dụ 2.28. Xác suất bắn trúng mục tiêu của một xạ thủ là $0,7$. Cho xạ thủ bắn 100 phát độc lập vào mục tiêu, tính xác suất có ít nhất 75 phát trúng mục tiêu.

Giải. Gọi X là số phát trúng trong 100 phát đã bắn. Khi đó, $X \sim B(100; 0,7)$. Áp dụng hệ quả trên, X có phân bố xấp xỉ phân bố chuẩn $N(70; 21)$. Do đó,

$$P(X \geq 75) = 1 - P(X < 75) = 1 - \Phi(\frac{75-70}{\sqrt{21}}) \approx 0,14$$

■

Ví dụ 2.29. Có 10.000 xe máy mua bảo hiểm của một công ty trong 1 năm. Mỗi chủ xe phải nộp phí 100.000 đồng/1 năm và trung bình nhận lại 5×10^6 đồng nếu xe máy bị tai nạn giao thông. Qua thống kê cho biết tỉ lệ xe máy bị tai nạn giao thông trong 1 năm là 0,006. Tính xác suất:

a) trong một năm hoạt động công ty bị lỗ.

b) trong một năm hoạt động công ty lãi ít nhất 800 triệu đồng.

Giải. Gọi X là số xe máy mua bảo hiểm của công ty bị tai nạn trong một năm, khi đó $X \sim B(10^4; 0,006)$. Vì $np = 60$ và $np(1-p) = 59,64$ nên ta có thể xấp xỉ X bởi phân bố chuẩn $N(60; 59,64)$.

a) Xác suất sau một năm hoạt động công ty bị lỗ là:

$$\begin{aligned} P(10^9 - 5 \times 10^6 X < 0) &= P(X > 200) = 1 - P(X \leq 200) \\ &= 1 - \Phi(18,13) = 0 \end{aligned}$$

b) Xác suất sau một năm hoạt động công ty lãi ít nhất 800 triệu đồng là:

$$P(10^9 - 5 \times 10^6 X \geq 8.10^8) = P(X \leq 40) = \Phi(-2,59) \approx 0,005$$

■

BÀI TẬP CHƯƠNG 2

▷ **2.1.** Một thiết bị gồm 3 bộ phận hoạt động độc lập với nhau. Xác suất trong thời gian T các bộ phận bị hỏng tương ứng là 0,4; 0,2 và 0,1. Gọi X là số bộ phận bị hỏng trong thời gian T.

a) Lập bảng phân bố xác suất của X.

b) Tính xác suất trong thời gian T có không quá 2 bộ phận bị hỏng.

▷ **2.2.** Ba xạ thủ độc lập bắn vào một mục tiêu. Xác suất bắn trúng tương ứng là 0,7; 0,8; 0,5; mỗi xạ thủ bắn một viên. Gọi X là số viên trúng.

a) Lập bảng phân phối của X.

b) Tìm kì vọng, phương sai và trung vị.

c) Tính xác suất có ít nhất 2 viên trúng.

▷ **2.3.** Có hai lô sản phẩm. Lô 1 có 8 chính phẩm và 2 phế phẩm, lô 2 có 7 chính phẩm và 3 phế phẩm. Từ lô 1 lấy ngẫu nhiên 2 sản phẩm bỏ vào lô 2, sau đó từ lô 2 lấy ngẫu nhiên ra 2 sản phẩm. Lập bảng phân phối xác suất của số chính phẩm được lấy ra ở lần 2.

▷ **2.4.** Một thiết bị có 3 bộ phận hoạt động độc lập. Gọi X là số bộ phận hỏng trong thời gian T, X có bảng phân bố xác suất sau:

x	0	1	2	3
$p(x)$	0,024	0,188	0,452	0,336

a) Tính kì vọng và phương sai của biến ngẫu nhiên X.

b) Biết xác suất bộ phân 1 hỏng trong thời gian T là 0,8. Tìm xác suất hỏng trong thời gian T của mỗi bộ phận còn tại.

▷ **2.5.** Một công ty khai thác dầu đang có hai dự án khai thác dầu, một ở châu Á và một ở châu Âu. Gọi X là số dự án thành công, X có bảng phân bố xác suất sau:

x	0	1	2
$p(x)$	0,02	0,26	0,72

Giả sử xác suất thành công mỗi dự án là độc lập nhau. Tìm xác suất thành công của mỗi dự án.

▷ **2.6.** Xác suất để một người bắn trúng bia là 0,8. Người ấy được phát từng viên đạn để bắn cho đến khi trúng bia. Gọi X là số viên đạn bắn trượt, tìm hàm xác suất của biến ngẫu nhiên X.

▷ **2.7.** Một xạ thủ được cung cấp 4 viên đạn và 80.000 đồng. Xạ thủ đó bắn độc lập từng viên cho tới khi một viên trúng đích hoặc hết đạn thì dừng lại. Xác suất bắn trúng đích của xạ thủ là 0,7. Nếu bắn trúng 1 viên thì được nhận 50.000 đồng còn nếu bắn trật 1 viên thì mất 20.000 đồng. Gọi X là số tiền có được của xạ thủ sau khi bắn xong. Lập bảng phân bố xác suất của X và tính E(X).

▷ **2.8.** Một người tham gia 1 trò chơi gieo đồng thời 3 đồng tiền cân đối đồng chất. Mỗi đồng tiền có 2 mặt kí hiệu là S và N. Người đó bỏ ra x đồng cho 1 lần gieo. Nếu kết quả gieo cả 3 mặt giống nhau thì người đó không thu về đồng nào còn nếu kết quả gieo cả 3 mặt không giống nhau thì được 3x đồng. Người này có nên thường xuyên tham gia trò chơi này không? Vì sao?

▷ **2.9.** Trong kì thi hết môn học A thầy giáo cho đề cương ôn tập gồm 10 câu lý thuyết và 15 câu bài tập. Thầy giáo tạo đề thi gồm 1 câu lý thuyết và 2 câu bài tập được lấy ngẫu nhiên trong đề cương. Sinh viên B chỉ học và trả lời được 7 câu lý thuyết và chỉ làm được 10 câu bài tập trong đề cương. Nếu trả lời đúng câu lý thuyết thì được 4 điểm và làm đúng mỗi câu bài tập thì được 3 điểm, không có điểm từng phần trong từng câu. Gọi X là số điểm môn học A của sinh viên B sau khi thi. Lập bảng phân bố xác suất của X và tính kì vọng $E(X)$.

▷ **2.10.** Một sinh viên thi vấn đáp trả lời 5 câu hỏi độc lập với nhau. Khả năng trả lời đúng mỗi câu hỏi đều bằng 65%. Nếu trả lời đúng thì sinh viên được 4 điểm, nếu sai thì bị trừ 2 điểm.

a) Tìm xác suất để sinh viên đó trả lời đúng 3 câu hỏi.

b) Tìm số điểm trung bình mà sinh viên đó đạt được.

▷ **2.11.** Trong một kênh truyền hình kĩ thuật số, xác suất thiết bị đầu cuối nhận được 1 bit (đơn vị dùng để đo lượng thông tin) bị lỗi là 0,001 và các bit bị lỗi độc lập nhau. Kí hiệu X là số bit thiết bị đầu cuối nhận được bị lỗi trong 1.000 bit được truyền đi.

a) Tính các xác suất $P(X = 1)$, $P(X \geq 1)$, $P(X \leq 2)$.

b) Tính kì vọng và phương sai của biến ngẫu nhiên X.

▷ **2.12.** Tuổi thọ (năm) của một loại bóng đèn là biến ngẫu nhiên X có hàm mật độ xác suất như sau:

$$f(x) = \begin{cases} kx^2(4-x) & \text{nếu } x \in [0;4] \\ 0 & \text{nếu } x \notin [0;4] \end{cases}$$

a) Tìm k và tính xác suất tuổi thọ bóng đèn bé hơn 1 năm tuổi.

b) Tìm kì vọng và phương sai của biến ngẫu nhiên X.

▷ **2.13.** Biến ngẫu nhiên X có hàm mật độ xác suất:

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

a) Tìm a và xác định hàm phân phối xác suất $F(x)$ của X.
b) Tính xác suất để X nhận giá trị trong khoảng $(\frac{\pi}{4}; \pi)$.

▷ **2.14.** Biến ngẫu nhiên X có hàm phân phối xác suất:

$$F(x) = a + \frac{1}{\pi}\arctan\frac{x}{2},\ x \in \mathbb{R}$$

a) Tìm a.
b) Tìm m sao cho $P(X > m) = 0,25$.

▷ **2.15.** Biến ngẫu nhiên X có hàm phân phối xác suất:

$$F(x) = \begin{cases} 0 & \text{nếu } x \le 0 \\ a\sin 2x & \text{nếu } 0 < x \le \pi/4 \\ 1 & \text{nếu } x > \pi/4 \end{cases}$$

a) Tìm a và hàm mật độ xác suất của X.
b) Tính kì vọng của X.

▷ **2.16.** Biến ngẫu nhiên liên tục X có hàm phân phối xác suất:

$$F(x) = \begin{cases} 0 & \text{nếu } x < -\pi/2, \\ a + b\sin x & \text{nếu } -\pi/2 \le x \le \pi/2 \\ 1 & \text{nếu } x > \pi/2 \end{cases}$$

Tìm các hằng số a và b. Tính $P(X > \pi/4)$.

▷ **2.17.** Theo thống kê, tỉ lệ tai nạn giao thông nặng là 1/1000 và tỉ lệ tai nạn giao thông nhẹ là 4/1000. Một công ty bảo hiểm bán bảo hiểm với mức phí 80.000 đồng. Hỏi tiền lãi trung bình của công ty bán bảo hiểm là bao nhiêu? Biết rằng thuế và các chi phí khác chiếm 30% phí bảo hiểm; ngoài ra nếu bị tai nạn giao thông nặng, nhẹ thì được công ty bảo hiểm bồi thường số tiền tương ứng là 10 triệu đồng và 5 triệu đồng.

$\triangleright$ **2.18.** Trọng lượng của một gói đường đóng bằng máy tự động có phân bố chuẩn. Trong 1.000 gói đường có 70 gói có trọng lượng lớn hơn 1.015g, trọng lượng trung bình của 1.000 gói đường là 1.012g. Hãy ước lượng xem trong 1.000 gói đường có bao nhiêu gói đường có trong lượng ít hơn 1.008g.

$\triangleright$ **2.19.** Một chi tiết máy được xem là đạt tiêu chuẩn nếu sai số tuyệt đối giữa chiều dài của nó so với chiều dài quy định không vượt quá 10mm. Biến ngẫu nhiên X chỉ độ lệch của chiều dài chi tiết so với chiều dài quy định có phân phối chuẩn $N(\mu, \sigma^2)$, với $\mu = 0$mm và $\sigma = 5$mm.

a) Chọn ngẫu nhiên một chi tiết, tính xác suất chi tiết đó đạt tiêu chuẩn.

b) Tìm số trung bình các chi tiết đạt tiêu chuẩn khi lấy ra 100 chi tiết.

$\triangleright$ **2.20.** Tuổi thọ (năm) của một thiết bị điện tử là một biến ngẫu nhiên X có hàm mật độ xác suất:

$$f(x) = \begin{cases} 0,25e^{-0,25x} & \text{nếu } x \geq 0 \\ 0 & \text{nếu } x < 0 \end{cases}$$

Bán được một thiết bị nếu không phải bảo hành thì lãi 15.000 đồng nhưng nếu phải bảo hành thì lỗ 5.000 đồng.

a) Tính $P(X > 2)$.

b) Để trung bình mỗi thiết bị lãi 10.000 thì nên quy định thời gian bảo hành bao nhiêu năm?

c) Với thời gian quy định bảo hành 6 tháng, cửa hàng A nhập về 10.000 thiết bị để bán. Tính xác suất với 10.000 thiết bị được bán hết cửa hàng A lãi ít nhất 125 triệu đồng.

$\triangleright$ **2.21.** Chiều cao của nam giới khi trưởng thành ở một vùng dân cư là biến ngẫu nhiên có phân phối chuẩn với $\mu = 160$cm và $\sigma = 6$cm. Một thanh niên bị coi là lùn nếu có chiều cao nhỏ hơn 155cm.

a) Tìm tỉ lệ thanh niên lùn ở vùng đó.

b) Tìm xác suất để lấy ngẫu nhiên 4 người thì có ít nhất một người không lùn.

▷ **2.22.** Tuổi thọ của một bóng đèn là biến ngẫu nhiên X có $E(X) = 250$ giờ và $\sigma(X) = 250$ giờ.

a) Một cửa hàng mua 30 bóng đèn để khi hỏng có thể thay thế ngay. Dùng định lí giới hạn trung tâm để tính gần đúng xác suất cửa hàng duy trì được ánh sáng liên tục trong ít nhất 8.750 giờ ($\approx$ 1 năm).

b) Dùng định lí giới hạn trung tâm để tính: chủ cửa hàng phải mua bao nhiêu bóng đèn để duy trì ánh sáng liên tục ít nhất 8.750 giờ với xác suất lớn hơn 0,9772.

▷ **2.23.** Một nhà nghỉ có 1.000 người. Nhà ăn phục vụ ăn trưa trong hai đợt liên tiếp. Mỗi người chọn ăn trưa một trong hai đợt này với xác suất như nhau. Dùng định lý giới hạn trung tâm tính: nhà ăn cần tối thiểu bao nhiêu chỗ để đảm bảo đủ chỗ cho khách vào ăn trưa với xác suất không bé hơn 0,99?

▷ **2.24.** Một hộp đựng 50 viên pin loại A và 50 viên pin loại B. Pin loại A có tuổi thọ trung bình $\mu_1 = 500$ giờ và độ lệch chuẩn $\sigma_1 = 15$ giờ, pin loại B có tuổi thọ trung bình $\mu_2 = 400$ giờ và độ lệch chuẩn $\sigma_2 = 6$ giờ. Tính gần đúng xác suất tổng tuổi thọ của 100 viên pin trên lớn hơn 45.200 giờ.

Chương 3
VECTƠ NGẪU NHIÊN HAI CHIỀU

3.1. Định nghĩa

Định nghĩa 3.1. Cho $X : \Omega \to \mathbb{R}$ và $Y : \Omega \to \mathbb{R}$ là hai biến ngẫu nhiên. Ánh xạ $Z : \Omega \to \mathbb{R} \times \mathbb{R}$ xác định bởi $Z(\omega) = (X(\omega), Y(\omega))$ được gọi là *vectơ ngẫu nhiên 2 chiều* và kí hiệu là $Z = (X, Y)$. Miền giá trị của $Z = (X, Y)$ được kí hiệu $Z(\Omega)$.

3.2. Phân bố xác suất của vectơ ngẫu nhiên

3.2.1. Vectơ ngẫu nhiên 2 chiều rời rạc

Định nghĩa 3.2. Cho vectơ ngẫu nhiên $Z = (X, Y)$ có miền giá trị:

$$Z(\Omega) = \{(x_i, y_j) : i, j \geq 1\}$$

Hàm xác suất đồng thời của $Z = (X, Y)$ là hàm số $p : \mathbb{R}^2 \to \mathbb{R}$ xác định bởi:

$$p(x, y) = \begin{cases} P(X = x, Y = y) & \text{nếu } (x, y) \in Z(\Omega) \\ 0 & \text{nếu } (x, y) \notin Z(\Omega) \end{cases}$$

Trong trường hợp X và Y có miền giá trị lần lượt là $\{x_1, x_2, ..., x_m\}$, $\{y_1, y_2, ..., y_n\}$, đặt:

$$p_{ij} = P(X = x_i; Y = y_j)$$

Bảng chữ nhật sau được gọi là *bảng phân bố xác suất đồng thời* của vectơ ngẫu nhiên (X, Y).

$x \backslash y$	y_1	y_2	...	y_n
x_1	p_{11}	p_{12}	...	p_{1n}
x_2	p_{21}	p_{22}	...	p_{2n}
...	...	...	...	...
x_m	p_{m1}	p_{m2}	...	p_{mn}

Ví dụ 3.1. Gieo đồng thời 1 đồng xu và 1 con xúc xắc cân đối đồng chất. Gọi X là số mặt sấp xuất hiện của đồng xu, Y là số chấm xuất hiện trên mặt con xúc xắc. Lập bảng phân phối xác suất đồng thời của (X, Y).

Giải. Gọi X là số mặt sấp xuất hiện của đồng xu, Y là số chấm xuất hiện trên mặt con xúc xắc. Ta có $X(\Omega) = \{0, 1\}$, $Y(\Omega) = \{1, 2, 3, 4, 5, 6\}$.

$$\begin{aligned}
P(X = 0, Y = 1) &= \frac{1}{2}.\frac{1}{6} = \frac{1}{12} \\
P(X = 0, Y = 2) &= \frac{1}{2}.\frac{1}{6} = \frac{1}{12} \\
&\dots \\
P(X = 1, Y = 1) &= \frac{1}{12} \\
&\dots
\end{aligned}$$

Bảng phân bố xác suất đồng thời của X và Y là:

$x \backslash y$	1	2	3	4	5	6
0	1/12	1/12	1/12	1/12	1/12	1/12
1	1/12	1/12	1/12	1/12	1/12	1/12

■

Ví dụ 3.2. Trong một hộp có 5 quả bóng bàn, trong đó có 3 quả chưa sử dụng (mới) và 2 quả đã sử dụng (cũ). Lần 1 lấy ngẫu nhiên 2 quả ra sử dụng sau đó trả lại hộp. Lần thứ 2 lấy ra 2 quả để sử dụng. Gọi X là số bóng mới lấy ra ở lần thứ nhất, Y là số bóng mới lấy ra ở lần thứ 2. Lập bảng phân phối xác suất đồng thời của (X, Y).

Giải.

$$P(X = 0, Y = 0) = P(X = 0)P(Y = 0|X = 0) = \frac{C_2^2}{C_5^2}\frac{C_2^2}{C_5^2} = 0,01$$

$$P(X = 0, Y = 1) = P(X = 0)P(Y = 1|X = 0) = \frac{C_2^2}{C_5^2}\frac{C_3^1 C_2^1}{C_5^2} = 0,06$$

$$P(X = 0, Y = 2) = P(X = 0)P(Y = 2|X = 0) = \frac{C_2^2}{C_5^2}\frac{C_3^2}{C_5^2} = 0,03$$

$$P(X = 1, Y = 0) = P(X = 1)P(Y = 0|X = 1) = \frac{C_3^1 C_2^1}{C_5^2}\frac{C_3^2}{C_5^2} = 0,18$$

$$P(X = 1, Y = 1) = P(X = 1)P(Y = 1|X = 1) = \frac{C_3^1 C_2^1}{C_5^2}\frac{C_3^1 C_2^1}{C_5^2} = 0,36$$

$$P(X = 1, Y = 2) = P(X = 1)P(Y = 2|X = 1) = \frac{C_3^1 C_2^1}{C_5^2}\frac{C_2^2}{C_5^2} = 0,06$$

$$P(X = 2, Y = 0) = P(X = 2)P(Y = 0|X = 2) = \frac{C_3^2}{C_5^2}\frac{C_4^2}{C_5^2} = 0,18$$

$$P(X = 2, Y = 2) = P(X = 2)P(Y = 2|X = 2) = 0$$

$$P(X=2,Y=1)=P(X=2)P(Y=1|X=2)=\frac{C_3^2}{C_5^2}\frac{C_4^3}{C_5^2}=0,12$$

Bảng phân bố xác suất đồng thời của (X,Y) là:

$x \backslash y$	0	1	2
0	$0,01$	$0,06$	$0,03$
1	$0,18$	$0,36$	$0,06$
2	$0,18$	$0,12$	0

■

Định lý 3.1. *Cho vectơ ngẫu nhiên $Z=(X,Y)$ có miền giá trị*

$$Z(\Omega)=\{(x_i,y_j): x_i\in X(\Omega), y_j\in Y(\Omega)\}$$

Khi đó:

1) $\sum\limits_{(x_i,y_j)\in Z(\Omega)} p(x_i;y_j)=1.$

2) Hàm xác suất của biến ngẫu nhiên X:

$$p_X(x)=\sum_{y_j\in Y(\Omega)} P(X=x;Y=y_j)$$

3) Hàm xác suất của biến ngẫu nhiên Y:

$$p_Y(y)=\sum_{x_i\in X(\Omega)} P(X=x_i;Y=y)$$

3.2.2. Vectơ ngẫu nhiên 2 chiều liên tục

Định nghĩa 3.3. Cho X và Y là hai biến ngẫu nhiên liên tục. Nếu tồn tại hàm số $f(x,y)\geq 0\ \forall\ x,y$ sao cho với mọi $a<b$ và $c<d$ ta có:

$$P(a\leq X\leq b, c\leq Y\leq d)=\int_a^b\int_c^d f(x,y)dxdy$$

thì $f(x,y)$ được gọi là *hàm mật độ xác suất đồng thời* của (X,Y).

Định lý 3.2. *Cho $f(x,y)$ là hàm mật độ đồng thời của (X,Y), $f_X(x)$ và $f_Y(x)$ lần lượt là hàm mật độ xác suất của X và Y. Khi đó:*

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

2) $f_X(x) = \int\limits_{-\infty}^{+\infty} f(x,y)dy.$

3) $f_Y(y) = \int\limits_{-\infty}^{+\infty} f(x,y)dx.$

Ví dụ 3.3. Cho vectơ ngẫu nhiên (X,Y) có hàm mật độ xác suất đồng thời:

$$f(x,y) = \begin{cases} ce^{-x-y} & \text{nếu } x \geq 0 \text{ và } y \geq 0 \\ 0 & \text{nếu trái lại} \end{cases}$$

a) Tìm hằng số c.
b) Tìm hàm mật độ xác suất của X và của Y.
c) Tìm xác suất để (X,Y) nhận giá trị trong miền chữ nhật:

$$\{(x,y) : 1 \leq x \leq 2, 0 < y < 2\}$$

Giải.
a) Ta có:

$$\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} f(x,y)dxdy = 1 \Leftrightarrow c\int\limits_{0}^{+\infty}\int\limits_{0}^{+\infty} e^{-x-y}dxdy = 1 \Leftrightarrow c = 1.$$

Vì vậy:

$$f(x,y) = \begin{cases} e^{-x-y} & \text{nếu } x \geq 0, y \geq 0 \\ 0 & \text{nếu trái lại} \end{cases}$$

b) $f_X(x) = \begin{cases} 0 & \text{nếu } x < 0 \\ e^{-x} & \text{nếu } x \geq 0 \end{cases}$ $\quad f_Y(y) = \begin{cases} 0 & \text{nếu } y < 0 \\ e^{-y} & \text{nếu } y \geq 0 \end{cases}$

c) $P(1 \leq x \leq 2, 0 < y < 2) = \int_1^2\int_0^2 e^{-x-y}dxdy$
$= e^{-4} - e^{-2} - e^{-3} + e^{-1}$ ■

3.2.3. Hàm phân phối xác suất đồng thời

Định nghĩa 3.4. Cho 2 biến ngẫu nhiên X và Y. Hàm số

$$F_{X,Y}(x,y) = P(X < x, Y < y)$$

được gọi là *hàm phân phối xác suất đồng thời* của vectơ ngẫu nhiên (X,Y).

1. Nếu vectơ ngẫu nhiên $(X;Y)$ có hàm xác suất đồng thời $p(x;y)$ thì:

$$F_{X,Y}(x,y) = \sum_{x_i<x;y_j<y} p(x_i,y_j)$$

2. Nếu vectơ ngẫu nhiên $(X;Y)$ có hàm mật độ xác suất đồng thời $f(x,y)$ thì:

$$F_{X,Y}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)dudv$$

Ví dụ 3.4. Tìm hàm phân phối xác suất đồng thời ở Ví dụ 3.3.

Giải. Với $x \geq 0$ và $y \geq 0$ ta có:

$$\begin{aligned} F_{X,Y}(x,y) &= \int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y} f(u,v)dudv \\ &= \int\limits_{0}^{x}\int\limits_{0}^{y} e^{-u-v}dudv \\ &= e^{-x-y} - e^{-x} - e^{-y} + 1 \end{aligned}$$

$F_{X,Y}(x,y) = 0$ nếu $x < 0$ hoặc $y < 0$.
Vì vậy:

$$F_{X,Y}(x,y) = \begin{cases} e^{-x-y} - e^{-x} - e^{-y} + 1 & \text{nếu } x \geq 0, y \geq 0 \\ 0 & \text{nếu trái lại} \end{cases}$$

■

Định lý 3.3. *Hai biến ngẫu nhiên X và Y độc lập khi và chỉ khi:*

$$f(x,y) = f_X(x).f_Y(y)$$

Trong đó $f(x,y)$, $f_X(x)$, $f_Y(y)$ lần lượt là hàm mật độ xác suất (hoặc hàm xác suất trong trường hợp biến ngẫu nhiên rời rạc) của (X,Y), X, Y.

3.3. Phân bố xác suất có điều kiện và kì vọng có điều kiện

3.3.1. Trường hợp rời rạc

Cho (X,Y) có bảng phân bố xác suất đồng thời:

$x \backslash y$	y_1	...	y_j	...	y_n
x_1	p_{11}	...	p_{1j}	...	p_{1n}
x_2	p_{21}	...	p_{2j}	...	p_{2n}
...	...	...	...	...	...
x_m	p_{m1}	...	p_{mj}	...	p_{mn}

Khi đó ta có:

$$\begin{aligned} P(X = x_i|Y = y_j) &= \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} \\ &= \frac{P(X = x_i, Y = y_j)}{\sum_{k=1}^{m} P(X = x_k)P(Y = y_j|X = x_k)} \\ &= \frac{p_{ij}}{p_{1j} + p_{2j} + ... + p_{mj}} \end{aligned}$$

Từ đó ta có hàm xác suất của X với điều kiện $Y = y_j$ như sau:

$$p_{y_j}(x) = \begin{cases} \dfrac{p_{ij}}{p_{1j} + p_{2j} + ... + p_{mj}} & \text{nếu } x = x_i \in X(\Omega) \\ 0 & \text{nếu } x \notin X(\Omega) \end{cases}$$

Kì vọng của X với điều kiện $Y = y_j$:

$$E(X|Y = y_j) = \frac{x_1p_{1j} + x_2p_{2j} + ... + x_mp_{mj}}{p_{1j} + p_{2j} + ... + p_{mj}}$$

3.3.2. Trường hợp liên tục

Định nghĩa 3.5. Cho X và Y là hai biến ngẫu nhiên liên tục và $y \in \mathbb{R}$. Nếu tồn tại hàm số $f_y(x) \geq 0$ $\forall x$ sao cho với mọi $a < b$ ta có:

$$P(a \leq X \leq b|Y = y) = \int_a^b f_y(x)dx$$

thì $f_y(x)$ được gọi là *hàm mật độ xác suất* của biến ngẫu nhiên X với điều kiện $Y = y$. Hàm mật độ xác suất $f_y(x)$ còn được kí hiệu bởi $f(x|y)$. Khi đó:

$$E(X|Y = y) = \int_{-\infty}^{+\infty} xf(x|y)dx$$

Định lý 3.4. *Cho (X, Y) có hàm mật độ xác suất đồng thời $f(x, y)$. Giả sử $f(x|y)$ là hàm mật độ xác suất của X với điều kiện $Y = y$ và và $f(y|x)$ hàm mật độ xác suất của Y với điều kiện $X = x$. Khi đó:*

1) $f(x|y) = \dfrac{f(x, y)}{f_Y(y)}$ *nếu* $f_Y(y) \neq 0$.

2) $f(y|x) = \dfrac{f(x, y)}{f_X(x)}$ *nếu* $f_X(x) \neq 0$.

Chứng minh. 1) Với $a < b$ ta có:

$$\begin{aligned} P(a < X < b|Y = y) &= \lim_{t \to 0} P(a < X < |y \leq Y \leq y + t) \\ &= \lim_{t \to 0} \frac{P(a < X < b, y \leq Y \leq t)}{P(y \leq Y \leq y + t)} \\ &= \lim_{t \to 0} \frac{\int_y^{y+t} \int_a^b f(x, v)dxdv}{\int_y^{y+t} f_Y(v)dv} \end{aligned}$$

Áp dụng quy tắc l'Hôpital ta được:

$$P(a < X < b|Y = y) = \lim_{t\to 0} \frac{\frac{d}{dt}\int_y^{y+t}\int_a^b f(x,v)dxdv}{\frac{d}{dt}\int_y^{y+t} f_Y(v)dv}$$

Áp dụng Định lý giá trị trung bình

$$\frac{d}{dt}\int_y^{y+t} g(x)dx = g(y+t)$$

Ta được:

$$\begin{aligned} P(a < X < b|Y = y) &= \lim_{t\to 0} \frac{\int_a^b f(x,y+t)dx}{f_Y(y+t)} \\ &= \frac{\int_a^b \lim_{t\to 0} f(u,y+t)du}{\lim_{t\to 0} f_Y(y+t)} = \int_a^b \frac{f(x,y)}{f_Y(y)}dx \end{aligned}$$

Vì vậy,

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

Chứng minh tương tự cho trường hợp (2). ■

Ví dụ 3.5. Cho vectơ ngẫu nhiên (X, Y) có hàm mật độ xác suất đồng thời:

$$f(x,y) = \begin{cases} 2e^{-x-y} & \text{nếu } y > x > 0 \\ 0 & \text{nếu trái lại} \end{cases}$$

a) Cho $x > 0$, tìm $f(y|x)$.
b) Cho $y > 0$, tìm $f(x|y)$.

Giải. a) Với $x > 0$ ta có:

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

$$= 0 + \int_x^{+\infty} 2e^{-x-y}dy = 2e^{-2x}$$

Do đó:

$$f(y|x) = \begin{cases} e^{x-y} & \text{nếu } y > x \\ 0 & \text{nếu } y \le x \end{cases}$$

b) Với $y > 0$ ta có:

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x,y)dx = \int_0^y 2e^{-x-y}dx = 2e^{-y}(1 - e^{-y})$$

Do đó:

$$f(x|y) = \begin{cases} \dfrac{e^{-x}}{1 - e^{-y}} & \text{nếu } 0 < x < y \\ 0 & \text{nếu trái lại} \end{cases}$$

■

3.4. Hiệp phương sai, hệ số tương quan

Định nghĩa 3.6. Cho vectơ ngẫu nhiên (X, Y). *Hiệp phương sai* của X và Y là một số xác định bởi công thức:

$$\mathrm{Cov}(X,Y) = E\left[(X - E(X)(Y - E(Y)))\right]$$

Định lý 3.5.
1) $\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y)$.
2) $\mathrm{Cov}(aX + a', bY + b') = ab\,\mathrm{Cov}(X,Y)$.
3) Nếu X và Y độc lập thì $\mathrm{Cov}(X,Y) = 0$.

Định lý 3.6. *Cho vectơ ngẫu nhiên $Z = (X, Y)$.*

$$E(XY) = \begin{cases} \sum_{i,j} x_i y_j p(x_i, y_j) & \textit{nếu } (X,Y) \textit{ rời rạc} \\ \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)dxdy & \textit{nếu } (X,Y) \textit{ liên tục} \end{cases}$$

trong đó $p(x,y)$ là hàm xác suất nếu (X,Y) là vectơ ngẫu nhiên rời rạc và $f(x,y)$ là hàm mật độ xác suất nếu (X,Y) là vectơ ngẫu nhiên liên tục.

Định lý 3.7.

1) Nếu X và Y là các biến ngẫu nhiên rời rạc có miền giá trị lần lượt là $\{x_1, x_2, ..., x_m\}$, $\{y_1, y_2, ..., y_n\}$ thì:

$$\mathrm{Cov}(X,Y) = \sum_{i=1}^{m}\sum_{j=1}^{n} x_i y_j p(x_i, y_j) - E(X)E(Y)$$

Trong đó $p_{ij} = P(X = x_i, Y = y_j)$.

2) Nếu (X,Y) có hàm mật độ xác suất đồng thời $f(x,y)$ thì:

$$\mathrm{Cov}(X,Y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyf(x,y)dxdy - E(X)E(Y)$$

Chứng minh. Do $\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y)$, áp dụng Định lí 3.6 ta có ngay các kết quả trên. ∎

Định nghĩa 3.7. *Hệ số tương quan* của hai biến ngẫu nhiên X và Y, kí hiệu là $\rho(X,Y)$, được xác định bởi công thức:

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}}$$

Định lý 3.8.

1) $|\rho(X,Y)| \leq 1$.

2) Nếu X và Y độc lập thì $\rho(X,Y) = 0$.

3) Nếu $Y = aX + b$ thì

$$\rho(X,Y) = \begin{cases} -1 & \text{nếu } a < 0 \\ 1 & \text{nếu } a > 0 \end{cases}$$

Chứng minh.

1) Đặt $X' = X - EX$, $Y' = Y - EY$. $\forall t \in R$, ta có:

$$\begin{aligned} E(tX' + Y')^2 \geq 0 &\Leftrightarrow EX'^2 t^2 + 2EtX'Y' + EY'^2 \geq 0 \\ &\Leftrightarrow E(X - EX)^2 t^2 + 2tE(X - EX)(X - EY) \\ &\quad + E(Y - EY)^2 \geq 0 \\ &\Leftrightarrow V(X)t^2 + 2t\,\mathrm{Cov}(X,Y) + V(Y) \geq 0 \end{aligned}$$

Đây là tam thức bậc hai theo t, do đó:

$$\begin{aligned}
\triangle' &= \operatorname{Cov}^2(X,Y) - V(X)V(Y) \le 0 \\
&\Leftrightarrow \operatorname{Cov}^2(X,Y) \le V(X)V(Y) \\
&\Leftrightarrow |\operatorname{Cov}(X,Y)| \le \sqrt{V(X)}\sqrt{V(Y)} \\
&\Leftrightarrow \frac{|\operatorname{Cov}(X,Y)|}{\sqrt{V(X)}\sqrt{V(Y)}} \le 1
\end{aligned}$$

Vậy:

$$|\rho(X,Y)| = \frac{|\operatorname{Cov}(X,Y)|}{\sqrt{V(X)}\sqrt{V(Y)}} \le 1$$

2) Nếu X, Y độc lập thì theo Định lý 3.6, ta có $\operatorname{Cov}(X,Y) = 0$, do đó:

$$|\rho(X,Y)| = \frac{|\operatorname{Cov}(X,Y)|}{\sqrt{V(X)}\sqrt{V(Y)}} = 0$$

3) $Y = aX + b \Rightarrow V(Y) = a^2V(X) \Rightarrow \sigma_Y = |a|SD(X)$.
Mặt khác do:

$$\begin{aligned}
\operatorname{Cov}(aX+b, X) &= E[(aX+b)X] - E(aX+b)EX \\
&= aEX^2 - a(EX)^2 \\
&= aV(X)
\end{aligned}$$

Suy ra:

$$\rho(aX+b, X) = \frac{aV(X)}{|a|V(X)} = \begin{cases} -1 & \text{nếu } a < 0 \\ 1 & \text{nếu } a > 0 \end{cases}$$

■

BÀI TẬP CHƯƠNG 3

▷ **3.1.** Số khách hàng mua máy ảnh kĩ thuật số hiệu Canon trong một tuần ở một cửa hàng là biến ngẫu nhiên X có bảng phân bố xác suất:

x	0	1	2	3	4
$p(x)$	$0,1$	$0,2$	$0,3$	$0,25$	$0,15$

Biết rằng 60% khách hàng mua máy ảnh kĩ thuật số hiệu Canon ở cửa hàng trên mua gói bảo hành mở rộng. Gọi Y là số khách hàng mua gói bảo hành mở rộng.

a) Tính xác suất $P(X = 4, Y = 2)$.

b) Tính xác suất $P(X = Y)$.

c) Lập bảng phân bố xác suất đồng thời của $(X; Y)$.

▷ **3.2.** Số trẻ em sinh ra trong 1 tuần ở làng A là 1 biến ngẫu nhiên X có bảng phân bố xác suất:

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

Số người chết trong 1 tuần ở làng đó là biến ngẫu nhiên Y có bảng phân bố xác suất:

y	0	1	2	3	4
$p(y)$	$0,1$	$0,3$	$0,4$	$0,15$	$0,05$

Giả sử X và Y độc lập.

a) Tìm bảng phân bố xác suất của vectơ ngẫu nhiên (X, Y).

b) Tính $P(X > Y)$.

▷ **3.3.** Cho (X, Y) có bảng phân bố xác suất đồng thời:

$X \backslash Y$	-1	1
-1	$1/6$	$1/4$
0	$1/6$	$1/8$
1	$1/6$	$1/8$

Tính $E(X)$, $E(Y)$, $\mathrm{Cov}(X, Y)$, $\rho(X, Y)$.

▷ **3.4.** Cho (X, Y) có bảng phân bố xác suất đồng thời:

X \ Y	1	2	3
2	$1/12$	$1/6$	$1/12$
3	$1/6$	0	$1/6$
4	0	$1/3$	0

Chứng minh X và Y không độc lập.

▷ **3.5.** Cho (X, Y) có hàm mật độ xác suất đồng thời:

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

a) Tìm hằng số c.

b) Tìm các hàm mật độ xác suất của X và Y.

▷ **3.6.** Cho (X, Y) có hàm mật độ

$$f(x, y) = \begin{cases} \dfrac{1}{6\pi} & \text{nếu } \dfrac{x^2}{9} + \dfrac{y^2}{4} < 1 \\ 0 & \text{nếu trái lại} \end{cases}$$

Tìm các hàm mật độ xác suất của X và Y.

▷ **3.7.** Cho (X, Y) có hàm mật độ xác suất đồng thời:

$$f(x, y) = \frac{c}{(1 + x^2)(1 + y^2)}$$

a) Tìm hằng số c.

b) Tìm hàm phân phối xác suất đồng thời của (X, Y).

c) Tìm xác suất (X, Y) nhận giá trị trong miền hình chữ nhật có các đỉnh $A(1; 1)$, $B(\sqrt{3}; 1)$, $C(1; 0)$ và $D(\sqrt{3}; 0)$.

▷ **3.8.** Cho (X, Y) có hàm mật độ xác suất đồng thời:

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

a) Với $0 \leq x \leq 1$, tìm $f(y|x)$.

b) Tính xác suất $P(X^2 + Y^2 \leq 1)$.

▷ **3.9.** Cho (X, Y) có hàm mật độ xác suất đồng thời:

$$f(x, y) = \begin{cases} c(1 - xy^3) & \text{nếu } |x| \leq 1, |y| \leq 1 \\ 0 & \text{nếu trái lại} \end{cases}$$

a) Tìm hằng số c.

b) Tìm $\rho(X, Y)$.

▷ **3.10.** Cho (X, Y) có hàm mật độ xác suất đồng thời:

$$f(x, y) = \begin{cases} e^{-y} & \text{nếu } y > x > 0 \\ 0 & \text{nếu trái lại} \end{cases}$$

a) Với $x > 0$, tính $E(Y|X = x)$.

b) Với $y > 0$, tính $E(X|Y = y)$.

Chương 4
THỐNG KÊ MÔ TẢ

4.1. Khái niệm mẫu và tổng thể

Giả sử ta cần nghiên cứu tính chất X nào đó của các phần tử trong tập hợp Ω mà Ω có số phần tử khá lớn ($|\Omega|$ có thể bằng vô cùng). Khi đó ta khó có thể nghiên cứu được tính chất X trên tất cả các phần tử. Phương pháp thống kê là chọn ngẫu nhiên một số lượng hữu hạn n phần tử để nghiên cứu. Trên cơ sở kết quả nghiên cứu của n phần tử này sẽ đưa ra kết luận cho toàn bộ tổng thể. Ta đưa ra các khái niệm sau:

1) *Tổng thể* là tập hợp tất cả các phần tử của Ω mà ta cần nghiên cứu tính chất X.
2) *Mẫu* là một tập con n phần tử của tổng thể được chọn ngẫu nhiên đề nghiên cứu. n được gọi là *kích thước mẫu* (hoặc *cỡ mẫu*).
3) Nếu mỗi phần tử của tổng thể có tính chất X là một số thực thì với phương pháp chọn mẫu ngẫu nhiên, ta có X là biến ngẫu nhiên, tập các giá trị của X trong mẫu được gọi là *mẫu số liệu*.

Ví dụ 4.1. X là chiều cao của thanh niên Việt Nam 22 tuổi hiện nay. Khi đó:

- Tổng thể là tập hợp toàn bộ thanh niên Việt Nam 22 tuổi.
- Vì số lượng thanh niên 22 tuổi trên cả nước rất lớn nên ta không thể điều tra hết được mà chỉ chọn ra 1 tập hợp con để điều

tra. Tập hợp con được chọn ra đó được gọi là một mẫu, số phần tử của mẫu là kích thước mẫu, tập tất cả các giá trị chiều cao của các cá thể trong mẫu là mẫu số liệu.

4.2. Các số đặc trưng của một mẫu số liệu

4.2.1. Trung bình mẫu, phương sai mẫu và độ lệch chuẩn mẫu

Cho $\{x_1, x_2, ..., x_n\}$ là mẫu số liệu của biến ngẫu nhiên X.

1) Trung bình mẫu, kí hiệu là $\overline{x}$, được tính theo công thức:

$$\overline{x} = \frac{x_1 + x_2 + ... + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

2) Phương sai mẫu, kí hiệu là s^2, được tính theo công thức:

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

3) Độ lệch chuẩn mẫu.

$$s = \sqrt{s^2} = \sqrt{\frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - n(\overline{x})^2\right]}$$

Ví dụ 4.2. Giả sử ta có mẫu số liệu về chiều cao (mét) của 10 sinh viên một trường đại học như sau:

1,75	1,69	1,73	1,77	1,68
1,73	1,77	1,70	1,74	1,71

Tìm trung bình mẫu, phương sai mẫu và độ lệch chuẩn mẫu.

Chú ý 4.1.

1) Mẫu số liệu cho dạng bảng phân bố tần số rời rạc:

X	x_1	x_2	...	x_m
n_i	n_1	n_2	...	n_m

- Kích thước mẫu: $n = n_1 + n_2 + ... + n_m$.
- Trung bình mẫu: $\overline{x} = \frac{1}{n}\sum\limits_{i=1}^{m} n_i x_i$.
- Phương sai mẫu: $s^2 = \frac{1}{n-1}\left[\sum\limits_{i=1}^{m} n_i x_i^2 - n\overline{x}^2\right]$.

2) Mẫu số liệu cho dạng bảng phân bố tần số liên tục:

X	$a_0 - a_1$	$a_1 - a_2$	...	$a_{m-1} - a_m$
n_i	n_1	n_2	...	n_m

Trong đó $a_{k-1} - a_k = [a_{k-1}; a_k)$. Đặt $x_k = \frac{a_{k-1} + a_k}{2}$ ta được:

X	x_1	x_2	...	x_m
n_i	n_1	n_2	...	n_m

3) Tính $\overline{x}$ và s bằng máy tính CASIO FX570VN PLUS.
- Mode → 3 → 1;
- Bật/tắt tần số: Shift → SETUP → REPLAY → 4(Stat);
- Nhập số liệu, kết thúc nhập: bấm AC;
- Lấy $\overline{x}$: Shift → 1 → 4 → 2 → =;
- Lấy s: Shift → 1 → 4 → 4 → =.

Ví dụ 4.3. Doanh thu X (triệu đồng) trong 100 ngày được chọn ngẫu nhiên của 1 cửa hàng cho bởi bảng sau:

19,0 - 19,4	19,4 - 19,8	19,8 - 20,2	20,2 - 20,6	20,6 - 21,0
15	25	30	20	10

Tìm trung bình mẫu và độ lệch chuẩn mẫu.

Giải. Đưa về bảng tần số rời rạc:

X	19,2	19,6	20,0	20,4	20,8
n_i	15	25	30	20	10

Áp dụng Chú ý 1 ta tính được $\overline{x} = 19,94$ và $s \approx 0,48$. ■

4.2.2. Trung vị mẫu

Sắp xếp mẫu số liệu theo thứ tự tăng dần, giả sử $x_1 \le x_2 \le ... \le x_n$. *Trung vị mẫu*, kí hiệu $\tilde{x}$, xác định bởi:

$$\tilde{x} = \begin{cases} x_{\frac{n+1}{2}} & \text{nếu } n \text{ lẻ,} \\ \dfrac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} & \text{nếu } n \text{ chẵn} \end{cases}$$

4.2.3. Hệ số tương quan mẫu

Cho $\{(x_1, y_1); (x_2, y_2); ...; (x_n, y_n)\}$ là mẫu hai chiều của vectơ ngẫu nhiên (X, Y). *Hệ số tương quan* mẫu được xác định bởi:

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

4.3. Biểu đồ

4.3.1. Biểu đồ phân bố tần số (Histogram)

Cho $(x_1, x_2, ..., x_n)$ là mẫu số liệu của biến ngẫu nhiên X.

Trường hợp 1: X là biến ngẫu nhiên rời rạc.

Lập bảng phân bố tần số rời rạc của số liệu đã cho như sau:

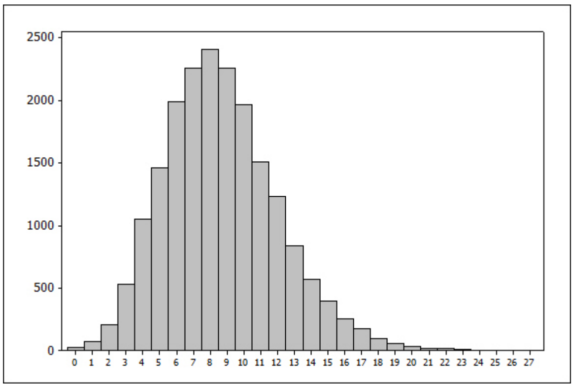
X	x_1^*	x_2^*	...	x_m^*
n_i	n_1	n_2	...	n_m

Sử dụng hệ trục tọa độ Descartes vuông góc để vẽ biểu đồ với trục hoành là các giá trị $x_1^*, x_2^*, ..., x_m^*$, trục tung là tần số (hoặc tần số tương đối).

Ví dụ 4.4. Trong một cuộc thi game online có 27 màn được tổ chức có 19.383 game thủ tham gia. Kết quả cho bởi bảng sau:

Vượt qua màn	Số game thủ	Tần số tương đối	Vượt qua màn	Số game thủ	Tần số tương đối
0	20	0,0010	14	569	0,0294

1	72	0,0037	15	393	0,0203
2	209	0,0108	16	253	0,0131
3	527	0,0272	17	171	0,0088
4	1048	0,0541	18	97	0,0050
5	1457	0,0752	19	53	0,0027
6	1988	0,1026	20	31	0,0016
7	2256	0,1164	21	19	0,0010
8	2403	0,1240	22	13	0,0007
9	2256	0,1164	23	5	0,0003
10	1967	0,1015	24	1	0,0001
11	1509	0,0779	25	0	0,0000
12	1230	0,0635	26	1	0,0001
13	834	0,0430	27	1	0,0001



Hình 4.1: *Biểu đồ tần số của cuộc thi game online*

Trường hợp 2: X **là biến ngẫu nhiên liên tục**.
Lập bảng phân bố tần số liên tục:

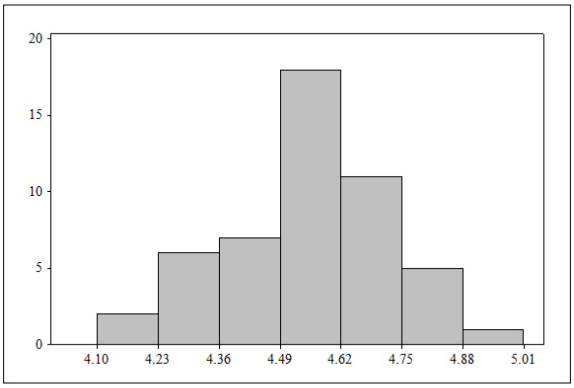
X	$[a_0; a_1)$	$[a_1; a_2)$	...	$[a_{m-1}; a_m)$
n_i	n_1	n_2	...	n_m

Trong đó, số khoảng cần chia tốt nhất là từ 5 đến 20 khoảng, có thể chọn xấp xỉ bằng $\sqrt{n}$ (hoặc $1+\log_2(n)$). Nếu ta chia dữ liệu thành m khoảng thì độ dài mỗi khoảng xấp xỉ $(\max\{x_k\} - \min\{x_k\})/m$.

Sử dụng hệ trục tọa độ Descartes vuông góc với trục hoành là các khoảng giá trị $[a_{k-1}; a_k)$, trục tung là tần số (hoặc tần số tương đối).

Ví dụ 4.5. Đo chiều dài (mm) của 50 con bọ cánh cứng ở một khu vực, nhà sinh học thu được kết quả sau:

4,1	4,35	4,45	4,5	4,55	4,55	4,6	4,65	4,7	4,75
4,2	4,35	4,45	4,5	4,55	4,6	4,6	4,65	4,7	4,75
4,3	4,35	4,45	4,5	4,55	4,6	4,6	4,65	4,7	4,75
4,3	4,4	4,45	4,5	4,55	4,6	4,65	4,65	4,725	4,78
4,35	4,4	4,45	4,525	4,55	4,6	4,65	4,7	4,75	4,95



Hình 4.2: *Biểu đồ tần số chiều dài bọ cánh cứng*

Vì $n = 50$, $\sqrt{50} \approx 7$ nên ta sẽ chia số liệu thành 7 khoảng, mỗi khoảng có độ dài $d = (\max\{x_i\} - \min\{x_i\})/7 \approx 0,13$. Từ đó ta có bảng phân bố tần số liên tục:

4,1-4,23	4,23-4,36	4,36-4,49	4,49-4,62	4,62-4,75	4,75-4,88	4,88-5,01
2	6	7	18	11	5	1

4.3.2. Biểu đồ thân - lá

Biểu đồ này tương tự histogram, chỉ khác ở chỗ chúng trình bày giá trị dữ liệu thay vì dùng các cột. Biểu đồ thân - lá gồm 3 thành phần là: phần thân (gồm một hoặc 2 chữ số đầu của một số liệu), phần lá (gồm những chữ số còn lại), tần số. Biểu đồ này thường chỉ dùng cho các nhóm dữ liệu nhỏ. Để tạo biểu đồ thân - lá ta làm như sau:

(1) Chia mỗi số liệu x_k thành 2 phần: phần thân gồm một hoặc 2 chữ số đầu, phần lá là những chữ số còn lại;

(2) Ghi phần thân thành một cột;

(3) Mỗi số liệu x_k ghi lại phần lá ứng với phần thân trên cùng một hàng;

(4) Với mỗi x_k ghi lại phần lá trên hàng của cột 2 ứng với phần thân;

(5) Ghi tần số trên cột thứ 3 (số phần lá ứng với phần thân).

(Tốt nhất chia số liệu từ 5 đến 20 thân)

Ví dụ 4.6. Vẽ biểu đồ thân - lá trong Ví dụ 4.5.

Lấy phần thân là các số $7, 8, 9, ..., 24$, khi đó ta được biểu đồ thân - lá như sau:

Tần số	Thân	Lá
1	41	0
2	42	0
8	43	005555
15	44	0555555
(11)	45	00002555555
24	46	0000000555555
11	47	0000255557
1	48	
1	49	5

4.3.3. Biểu đồ xác suất chuẩn

Giả sử mẫu số liệu của biến ngẫu nhiên X đã sắp thứ tự tăng dần:

$$x_1 \leq x_2 \leq x_3 \leq ... \leq x_n$$

Hàm phân phối tần số thực nghiệm của X được xác định như sau:

$$F(x) = \frac{\text{số phần tử của mẫu số liệu} < x}{n}$$

Do có đúng $j-1$ phần tử của mẫu số liệu bé hơn x_j và có đúng j phần tử của mẫu bé hơn hoặc bằng x_j nên:

$$F(x_j) = \frac{j-1}{n} \approx \frac{j-0,5}{n}$$

Nếu biến ngẫu nhiên X có phân bố chuẩn $N(\mu;\sigma^2)$ thì:

$$P(X < x_j) = \Phi(\frac{x_j - \mu}{\sigma}) \approx \frac{j-0,5}{n}$$

$$\Leftrightarrow \frac{x_j - \mu}{\sigma} \approx \Phi^{-1}(\frac{j-0,5}{n}) = z_j$$

Từ đó ta có $x_j \approx \sigma z_j + \mu$. Tức là các điểm $(z_j; x_j)$ với $j = 1, 2, ..., n$ nằm xấp xỉ trên đường thẳng

$$x = \sigma z + \mu$$

Do đó nếu $(z_j; x_j)$ với $j = 1, 2, ..., n$ nằm xấp xỉ trên 1 đường thẳng thì có thể xem biến ngẫu nhiên X có phân bố chuẩn.

Định nghĩa 4.1. *Biểu đồ xác suất chuẩn* là tập hợp các điểm có tọa độ $(z_i; x_i)$ với $i = 1, 2, ..., n$ trên hệ trục tọa độ Descartes vuông góc Ozx, trong đó:

$$\Phi(z_j) = \frac{j-0,5}{n}$$

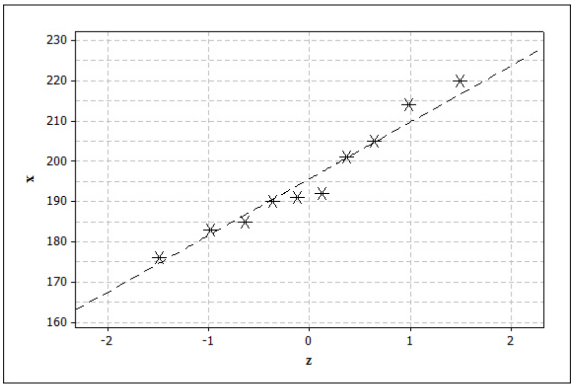
Ví dụ 4.7. Xây dựng biểu đồ xác suất chuẩn của số liệu sau.

176	183	185	190	191	192	201	205	214	220

Giải.

j	x_j	$\frac{j-0,5}{n}$	$z_j = \Phi^{-1}(\frac{j-0,5}{n})$
1	176	0,05	-1,64
2	183	0,15	-1,04
3	185	0,25	-0,67
4	190	0,35	-0,39
5	191	0,45	-0,13
6	192	0,55	0,13
7	201	0,65	0,39
8	205	0,75	0,67
9	214	0,85	1,04
10	220	0,95	1,64

Biểu đồ xác suất chuẩn là:



Hình 4.3: *Biểu đồ xác suất chuẩn của Ví dụ 4.7*

■

4.4. Mẫu ngẫu nhiên

Cho X là biến ngẫu nhiên của một tổng thể cần nghiên cứu. Để có thể áp dụng lý thuyết xác suất vào thống kê toán ta đưa ra định nghĩa mẫu ngẫu nhiên như sau.

Định nghĩa 4.2. Cho biến ngẫu nhiên X. Các biến ngẫu nhiên $(X_1, X_2, ..., X_n)$ được gọi là mẫu ngẫu nhiên của X nếu thỏa mãn 2 điều kiện sau:

(i) $X_1, X_2, ..., X_n$ là các biến ngẫu nhiên độc lập.

(ii) $X_1, X_2, ..., X_n$ có cùng phân bố xác suất với X.

Như vậy, mẫu số liệu $(x_1, x_2, ..., x_n)$ có thể xem là 1 giá trị của mẫu ngẫu nhiên $(X_1, X_2, ..., X_n)$.

Ví dụ 4.8. Tung 1 con xúc xắc cân đối đồng chất, gọi X là số chấm xuất hiện trên xúc xắc. Biến ngẫu nhiên X có hàm xác suất:

$$p(k) = \frac{1}{6} \text{ với } k \in \{1, 2, 3, 4, 5, 6\}$$

Bây giờ tiến hành tung xúc xắc 10 lần, gọi X_i là số chấm xuất sẽ hiện ở lần tung thứ i. Khi đó X_1, X_2,..., X_{10} là các biến ngẫu nhiên độc lập và có phân bố xác suất giống X. Vì vậy $(X_1, ..., X_{10})$ là một mẫu ngẫu nhiên của X.

4.5. Chọn mẫu ngẫu nhiên đơn giản

4.5.1. Chọn mẫu từ tổng thể hữu hạn

Giả sử tổng thể cần nghiên cứu có kích thước $|\Omega| = N$, ta cần chọn ra 1 mẫu có kích thước n.

Định nghĩa 4.3. Một mẫu ngẫu nhiên đơn giản kích thước n được chọn ra từ 1 tổng thể kích thước N là mẫu được chọn sao cho các phần tử được chọn vào mẫu có xác suất bằng nhau.

Có hai phương pháp chọn mẫu ngẫu nhiên đơn giản cơ bản: Chọn mẫu ngẫu nhiên có hoàn lại và chọn mẫu ngẫu nhiên không hoàn lại.

a) *Chọn mẫu ngẫu nhiên có hoàn lại*: Chọn ngẫu nhiên 1 phần tử, ghi lại thông tin cần nghiên cứu, trả phần tử đó về tổng thể và chọn ngẫu nhiên phần tử tiếp theo. Lặp lại quá trình chọn mẫu n lần. Như vậy mỗi phần tử có thể được chọn nhiều hơn 1 lần vào mẫu. Các phần tử được chọn vào mẫu là độc lập.

Ví dụ 4.9. Tổng thể $\Omega = \{a, b, c\}$, ta cần chọn 1 mẫu có kích thước $n = 2$ theo phương pháp chọn mẫu có hoàn lại.

Mẫu	1	2	3	4	5	6	7	8	9
X_1	a	a	a	b	b	b	c	c	c
X_2	a	b	c	a	b	c	a	b	c

Nhận xét: Nếu $(X_1, X2, ..., X_n)$ là mẫu được chọn theo phương pháp có hoàn lại thì $X_1, X_2, ..., X_n$ là các biến ngẫu nhiên độc lập và có cùng phân bố với tổng thể.

b) *Chọn mẫu ngẫu nhiên không hoàn lại*: Chọn ngẫu nhiên 1 phần tử, ghi lại thông tin cần nghiên cứu, không trả phần tử đó về tổng thể và chọn ngẫu nhiên phần tử tiếp theo. Lặp lại quá trình chọn mẫu n lần. Như vậy mỗi phần tử được chọn không quá 1 lần vào mẫu. Với phương pháp chọn mẫu này, mỗi phần tử được chọn vào mẫu có xác suất bằng nhau và bằng n/N.

Ví dụ 4.10. Tổng thể $\Omega = \{a, b, c\}$, ta cần chọn 1 mẫu có kích thước $n = 2$ theo phương pháp chọn mẫu không hoàn lại.

Mẫu	1	2	3	4	5	6
X_1	a	a	b	b	c	c
X_2	b	c	a	c	a	b

Nếu $(X_1, X2, ..., X_n)$ là mẫu được chọn theo phương pháp không hoàn lại thì $X_1, X_2, ..., X_n$ là các biến ngẫu nhiên không độc lập và có cùng phân bố với tổng thể. Tuy nhiên trong trường hợp tổng thể có kích thước tổng thể N lớn hơn rất nhiều so với kích thước mẫu n, thường được giả thiết $n/N \leq 0,05$, thì $X_1, X_2, ..., X_n$ gần như độc lập.

Nói chung, khi tổng thể có kích thước tổng thể N rất lớn thì không có khác biệt đáng kể giữa hai phương pháp chọn mẫu trên. Và trên thực tế phương pháp chọn mẫu không hoàn lại được áp dụng nhiều hơn. Trong phạm vi giáo trình này ta luôn giả thiết $n/N \leq 0,05$.

Để áp dụng phương pháp chọn mẫu ngẫu nhiên không hoàn lại ta có thể sử dụng bảng số ngẫu nhiên hoặc sử dụng các phần mềm thống kê. Chẳng hạn sử dụng phần mềm Minitab để chọn ngẫu nhiên n phần tử từ 1 danh sách được đánh số thứ tự từ 1 đến N ta thực hiện như sau:

- Tạo một danh sách số thứ tự từ 1 đến N: Calc → Make Patterned Data → Simple Set of Numbers.

◇ Store Patterned Data in: chọn C1

◇ From first value: chọn 1

◇ To last value: chọn N

◇ In steps of: chọn 1

◇ Number of times to list each value: chọn 1

◇ Number of times to list the sequence: chọn 1

- Chọn ngẫu nhiên n phần tử từ danh sách: Calc → Random Data → Sample from Columns.

◇ Number of rows to sample: chọn n

◇ From Colmns: chọn C1

◇ Store sample in: chọn C2

Nếu chọn mẫu có hoàn lại thì chọn: Sample with replacement.

4.5.2. Chọn mẫu từ tổng thể vô hạn

Trong một số trường hợp ta cần chọn mẫu kích thước n từ 1 tổng thể có vô hạn phần tử ($N = \infty$). Chẳng hạn chọn một mẫu các sản phẩm được sản xuất bởi một nhà máy; chọn một mẫu là khách hàng vào một cửa hàng;... Đối với trường hợp tổng thể có vô hạn phần tử, một mẫu ngẫu nhiên $(X_1, X_2, ..., X_n)$ được chọn phải thỏa mãn các điều kiện sau:

1) Mỗi phần tử được chọn vào mẫu là ngẫu nhiên.
2) Các phần tử phải được chọn độc lập nhau.

4.6. Phân bố của trung bình mẫu

Trong mục này ta sẽ nghiên cứu phân bố của trung bình mẫu. Trước hết ta xét ví dụ sau:

Ví dụ 4.11. Cho tổng thể $\Omega = \{1, 2, 3, 4, 5, 6\}$. Tìm phân bố của trung bình mẫu có kích thước $n = 2$ chọn theo phương pháp có hoàn lại.

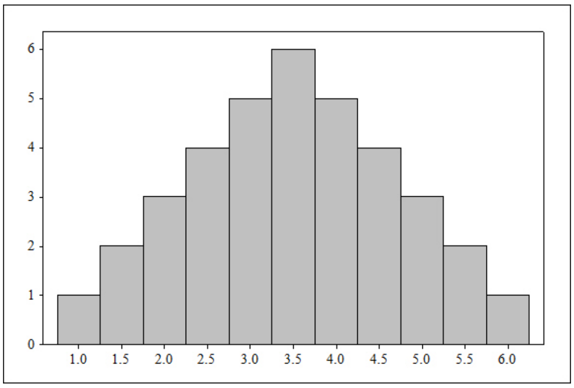
Kí hiệu (X_1, X_2) là mẫu ngẫu nhiên. Ta có bảng các giá trị của (X_1, X_2):

X_1 \ X_2	1	2	3	4	5	6
1	(1;1)	(1;2)	(1;3)	(1;4)	(1;5)	(1;6)
2	(2;1)	(2;2)	(2;3)	(2;4)	(2;5)	(2;6)
3	(3;1)	(3;2)	(3;3)	(3;4)	(3;5)	(3;6)
4	(4;1)	(4;2)	(4;3)	(4;4)	(4;5)	(4;6)
5	(5;1)	(5;2)	(5;3)	(5;4)	(5;5)	(5;6)
6	(6;1)	(6;2)	(6;3)	(6;4)	(6;5)	(6;6)

Như vậy ta có 36 giá trị của mẫu ngẫu nhiên (X_1, X_2) nên sẽ có 36 giá trị trung bình mẫu. Bảng phân bố tần số của trung bình mẫu:

$\overline{x}$	1	1,5	2	2,5	3	3,5	4	4,5	5	5,5	6
n_i	1	2	3	4	5	6	5	4	3	2	1

Biểu đồ tần số của trung bình mẫu:



Hình 4.4

Nhận xét: Phân bố của trung bình mẫu $\overline{x}$ có hình dáng của phân bố chuẩn.

Định lý 4.1. *Nếu* $(X_1, X_2, ..., X_n)$ *là mẫu ngẫu nhiên của biến ngẫu nhiên* X *có phân bố chuẩn* $N(\mu;\sigma^2)$ *thì:*

$$\overline{X} = \frac{1}{n}(X_1 + X_2 + ... + X_n)$$

cũng có phân bố chuẩn với kì vọng $E(\overline{X}) = \mu$ *và độ lệch chuẩn* $SD(\overline{X}) = \dfrac{\sigma}{\sqrt{n}}$.

Trong trường hợp tổng thể không có phân bố chuẩn thì ta chọn mẫu có kích thước $n > 30$, khi đó áp dụng Định lí giới hạn trung tâm ta có $\overline{X}$ có xấp xỉ phân bố chuẩn với kì vọng μ và độ lệch chuẩn $\sigma/\sqrt{n}$.

Định lý 4.2. *Nếu* $(X_1, X_2, ..., X_n)$ *là mẫu ngẫu nhiên của biến ngẫu nhiên* X *có phân bố chuẩn* $N(\mu;\sigma^2)$ *thì biến ngẫu nhiên*

$$\frac{\overline{X} - \mu}{S/\sqrt{n}}$$

có phân phối Student $n-1$ *bậc tự do* (T_{n-1})*. Trong đó:*

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}.$$

BÀI TẬP CHƯƠNG 4

▷ **4.1.** Chỉ số khối cơ thể (BMI) được tính bằng tỉ số giữa cân nặng (kg) và bình phương chiều cao (m^2) được dùng để đo mức độ béo hay gầy của một người. Giả sử rằng phân bố chỉ số BMI của nam giới có phân bố chuẩn với độ lệch chuẩn $\sigma = 3$ (kg/m^2). Một mẫu ngẫu nhiên gồm 49 nam giới được chọn để ước lượng chỉ số BMI trung bình. Tính xác suất sai số tuyệt đối giữa chỉ số BMI trung bình và trung bình mẫu không vượt quá 1 (kg/m^2).

▷ **4.2.** Một sinh viên thực hiện thí nghiệm gieo 8 hạt đậu trong 8 cái cốc riêng biệt và được đặt dưới ánh sáng của đèn huỳnh quang chiếu sáng liên tục. Sau 14 ngày, sinh viên này tiến hành đo chiều cao (cm) của các cây đậu được gieo ở trên và thu được kết quả sau:

Cây	1	2	3	4	5	6	7	8
Chiều cao (cm)	7,5	10,1	8,3	9,8	5,7	10,3	9,2	8,7

Tìm trung bình mẫu, độ lệch chuẩn mẫu và trung vị mẫu.

▷ **4.3.** Đo nồng độ cholesterol trong huyết thanh của 1.067 đàn ông Mỹ tuổi từ 25 đến 34 được chọn ngẫu nhiên, nhà nghiên cứu thu được kết quả như sau.

Nồng độ cholesterol (mg/100 ml)	Số người
80–119	13
120–159	150
160–199	442
200–239	299
240–279	115
280–319	34
320–399	14

a) Vẽ biểu đồ tần số.
b) Tìm trung bình mẫu, độ lệch chuẩn mẫu và trung vị mẫu.

▷ **4.4.** Nồng độ hemoglobin trong máu của 60 người được làm xét nghiệm ở một địa phương nghi có dịch sốt xuất huyết cho bởi:

43	63	63	75	95	75	80	48	62	71	76	90
51	61	74	103	93	82	74	65	63	53	64	67
80	77	60	69	73	76	91	55	65	69	84	78
50	68	72	89	75	57	66	79	85	70	59	71
87	67	72	52	35	67	99	81	97	74	61	62

a) Vẽ biểu đồ tần số, biểu đồ xác suất chuẩn.
b) Tìm trung bình mẫu, độ lệch chuẩn mẫu và trung vị mẫu.

▷ **4.5.** Đo chiều cao (đơn vị là cm) 10 nam thanh niên trưởng thành được chọn ngẫu nhiên ở vùng A và 10 nam thanh niên trưởng thành được chọn ngẫu nhiên ở vùng B. Số đo chiều cao của hai nhóm người này được cho như sau.

X	165	167	174	172	165	167	168	172	170	173
Y	177	175	172	174	176	172	178	170	168	179

Hãy vẽ biểu đồ xác suất chuẩn của hai mẫu số liệu của X và Y trên cùng 1 hệ trục tọa độ. Nhận xét về độ lệch chuẩn của X và Y.

▷ **4.6.** Hàm lượng asen trong 20 mẫu nước ngầm được lấy ngẫu nhiên ở hai vùng dân cư A và B được cho như sau:

X (10^{-3} mg/l)	3	7	25	10	15	6	12	25	15	7
Y (10^{-3} mg/l)	48	44	40	38	33	21	20	12	1	18

Hãy vẽ biểu đồ xác suất chuẩn của hai mẫu số liệu của X và Y trên cùng 1 hệ trục tọa độ. Nhận xét về độ lệch chuẩn của X và Y.

▷ **4.7.** Đo kích thước chiều dài (mm) và chiều rộng (mm) của một mẫu gồm 50 con bọ cánh cứng khác nhau ở một khu vực gần close Bristol, Vương quốc Anh, các nhà khoa học (Peacock, L., Carter, P., Powers, S. and Karp, A. 2003) thu được như sau:

TT	Dài	Rộng	TT	Dài	Rộng	TT	Dài	Rộng
1	4,60	1,50	18	4,55	1,50	35	4,60	1,60
2	4,70	1,65	19	4,60	1,70	36	4,55	1,65
3	4,50	1,55	20	4,55	1,60	37	4,775	1,55
4	4,55	1,65	21	4,35	1,60	38	4,60	1,60
5	4,75	1,65	22	4,45	1,60	39	4,45	1,475
6	4,40	1,50	23	4,55	1,55	40	4,60	1,60
7	4,20	1,70	24	4,35	1,55	41	4,65	1,625
8	4,70	1,55	25	4,65	1,65	42	4,725	1,65
9	4,55	1,60	26	4,50	1,55	43	4,95	1,725
10	4,70	1,65	27	4,45	1,50	44	4,65	1,65
11	4,65	1,55	28	4,45	1,60	45	4,60	1,625
12	4,50	1,55	29	4,10	1,40	46	4,45	1,55
13	4,30	1,50	30	4,50	1,50	47	4,30	1,525
14	4,65	1,65	31	4,60	1,60	48	4,75	1,60
15	4,75	1,65	32	4,75	1,575	49	4,525	1,55
16	4,65	1,60	33	4,70	1,65	50	4,35	1,50
17	4,45	1,60	34	4,35	1,55			

a) Vẽ biểu đồ tần số và biểu đồ xác suất chuẩn của các số liệu về chiều dài và chiều rộng.

b) Tìm hệ số tương quan mẫu.

Chương 5
ƯỚC LƯỢNG THAM SỐ

5.1. Ước lượng điểm

5.1.1. Ước lượng điểm và hàm ước lượng

Một biến ngẫu nhiên X của một tổng thể có các số đặc trưng của nó như kì vọng, phương sai,... Các số đặc trưng này sẽ được gọi chung là tham số. Các số đặc trưng của một biến ngẫu nhiên X thường khó biết được chính xác giá trị của nó là bao nhiêu nếu không biết được phân bố xác suất của nó. Chẳng hạn, gọi μ là chiều cao trung bình của nam thanh niên trưởng thành ở một địa phương nào đó mà ta chưa biết. Một mẫu gồm 10 nam thanh niên trưởng thành được chọn ngẫu nhiên ở địa phương trên, gọi x_1, x_2, ..., x_{10} là số đo chiều cao của 10 nam thanh niên này. Khi đó ta có thể sử dụng trung bình mẫu $\overline{x} = (x_1 + x_2 + ... + x_{10})/10$ để ước lượng. Tuy nhiên ta cũng có thể sử dụng một hàm khác để ước lượng. Chẳng hạn, ta có thể sử dụng hàm trung vị $\tilde{x}$ để ước lượng μ. Hàm của mẫu số liệu dùng để ước lượng tham số được gọi là hàm ước lượng.

> *Ước lượng điểm* của tham số θ dựa trên mẫu số liệu $(x_1, x_2, ..., x_n)$ là một hàm n biến:
> $$\hat{\theta} = \hat{\theta}(x_1, x_2, ..., x_n).$$

5.1.2. Ước lượng không chệch

Cho biến ngẫu nhiên X có tham số θ, gọi $(X_1, X_2, ..., X_n)$ là mẫu ngẫu nhiên của X. Khi đó hàm:

$$\hat{\theta} = \hat{\theta}(X_1, X_2, ..., X_n)$$

cũng được gọi là hàm ước lượng. Hàm ước lượng $\hat{\theta}$ được gọi là ước lượng không chệch đối với tham số θ nếu $E(\hat{\theta}) = \theta$. Ngược lại, ta gọi $\hat{\theta}$ là ước lượng chệch và $E(\hat{\theta}) - \theta$ gọi là độ chệch của ước lượng.

5.1.3. Ước lượng không chệch của kì vọng và phương sai

Định lý 5.1. *Cho biến ngẫu nhiên X của 1 tổng thể có $E(X) = \mu$, $V(X) = \sigma^2$. Với $(X_1, X_2, ..., X_n)$ là mẫu ngẫu nhiên của X. Khi đó:*

1) $\overline{X} = \dfrac{X_1 + X_2 + ... + X_n}{n}$ là ước lượng không chệch của μ.

2) $S^2 = \dfrac{1}{n-1}\sum\limits_{i=1}^{n}(X_i - \overline{X})^2$ là ước lượng không chệch của σ^2.

Chứng minh. Thật vậy, ta có:

$$E(\overline{X}) = E\left(\frac{X_1 + X_2 + ... + X_n}{n}\right)$$
$$= \frac{E(X_1) + E(X_2 + ... + E(X_n)}{n} = \frac{n\mu}{n} = \mu$$

Đối với S^2 ta có:

$$S^2 = \frac{1}{n-1}\left(X_1^2 + ... + X_n^2 - \frac{(X_1 + ... + X_n)^2}{n}\right)$$

Do đó:

$$E(S^2) = \frac{1}{n-1}\left(E(X_1^2) + ... + E(X_n^2) - \frac{E(X_1 + ... + X_n)^2}{n}\right)$$
$$= \frac{1}{n-1}\Big(n(\sigma^2 + \mu^2) -$$
$$- \frac{V(X_1 + ... + X_n) + (E(X_1 + ... + X_n))^2}{n}\Big)$$

$$= \frac{1}{n-1}\left(n(\sigma^2+\mu^2) - \frac{n\sigma^2+(n\mu)^2}{n}\right)$$
$$= \sigma^2$$

■

5.1.4. Ước lượng không chệch tỉ lệ

Giả sử p là tỉ lệ phần tử có tính chất A nào đó trong 1 tổng thể (chẳng hạn tỉ lệ phế phẩm do 1 dây chuyền sản xuất, tỉ lệ nam giới ở một địa phương,...). Ta sẽ sử dụng phân bố Bernoulli để mô tả bằng cách mỗi phần tử của tổng thể được gán bởi 1 nếu có tính chất A và được gán bởi 0 nếu không có tính chất A. Chọn ngẫu nhiên 1 phần tử của tổng thể đó, ta đặt:

$$X = \begin{cases} 1 & \text{nếu phần tử đó có tính chất A} \\ 0 & \text{nếu phần tử đó không có tính chất A} \end{cases}$$

Khi đó X có phân bố Bernoulli với tham số p.

Định lý 5.2. *Cho biến ngẫu nhiên X có phân bố Bernoulli với tham số p. Gọi $(X_1, X_2, ..., X_n)$ là mẫu ngẫu nhiên của X. Khi đó:*

$$\hat{P} = \frac{X_1 + X_2 + ... + X_n}{n}$$

là một ước lượng không chệch của tham số p.

5.2. Nguyên lí xác suất nhỏ và nguyên lí xác suất lớn

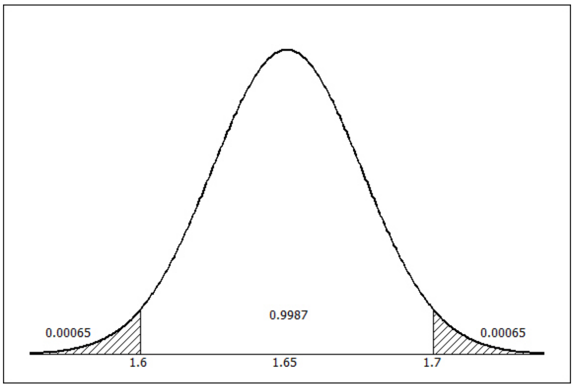
Một biến cố không thể xảy ra có xác suất bằng 0 nhưng một biến cố có xác suất bằng 0 vẫn có thể xảy ra khi ta thực hiện môt số lượng lớn phép thử. Qua thực tế người ta nhận thấy rằng một biến cố có xác suất bé sẽ hầu như không xảy ra khi ta thực hiện một hoặc hai lần phép thử. Vì vậy các nhà thống kê thừa nhận một nguyên lý sau đây và gọi là "nguyên lí xác suất nhỏ": Một biến cố có xác suất rất nhỏ gần bằng 0 thì biến cố đó hầu như chắc chắn không xảy ra khi thực hiện phép thử một lần. Chẳng hạn khi mua một vé xổ số thì xác suất trúng giải đặc biệt rất nhỏ nên có thể xem biến cố trúng giải đặc biệt sẽ không xảy ra khi mua 1 vé xổ số.

Tương tự như vậy ta có nguyên lí xác suất lớn: Một biến cố có xác suất gần bằng 1 thì biến cố đó hầu như chắc chắn sẽ xảy ra khi thực hiện phép thử.

Ví dụ 5.1. Một hộp đựng 9.999 viên bi xanh và 1 viên bi đỏ, các viên bi giống nhau về kích thước và khối lượng. Lấy ngẫu nhiên 1 viên bi. Kí hiệu: A là biến cố lấy được viên bi xanh, B là biến cố lấy được viên bi đỏ.

Khi đó ta có $P(A) = 0,999$ và $P(B) = 0,001$. Theo nguyên lí xác suất lớn và nguyên lí xác suất nhỏ thì khi ta lấy ngẫu nhiên 1 viên bi (thực hiện 1 lần) thì hầu như chắc chắn rằng ta sẽ lấy được viên bi xanh còn biến cố lấy được viên bi đỏ hầu như chắc chắn rằng sẽ không xảy ra.

Ví dụ 5.2. Đại học A có khoảng 5.000 nam sinh viên. Chiều cao X của nam sinh viên của trường đại học A có phân bố chuẩn với chiều cao trung bình $\mu = 1,65$m và độ lệch chuẩn $\sigma = 0,1$m.



Hình 5.1

Gọi $(X_1, X_2, ..., X_{36})$ là mẫu ngẫu nhiên của X. Khi đó:

$$\overline{X} = \frac{X_1 + X_2 + ... + X_{36}}{36}$$

có phân bố chuẩn $N(1,65; \frac{0,1^2}{36})$. Vì vậy,

$$P(1,6 < \overline{X} < 1,7) = 2\Phi(3) - 1 = 0,9987$$

và

$$P(\overline{X} \notin (1,6; 1,7)) = 0,0013$$

Đại học A có 5.000 nam sinh viên nên sẽ có $C_{5000}^{36} = 344774 \times 10^{86}$ mẫu kích thước $n = 36$ khác nhau, trong đó có 99,87% mẫu có trung bình mẫu nằm trong khoảng $(1,6; 1,7)$ và 0,13% mẫu có trung bình mẫu nằm ngoài khoảng $(1,6; 1,7)$. Theo nguyên lí xác suất lớn và nguyên lí xác suất nhỏ thì biến cố chọn ngẫu nhiên được một mẫu 36 nam sinh viên có trung bình mẫu $\overline{X} \in (1,6; 1,7)$ sẽ xảy ra, còn biến cố chọn ngẫu nhiên được một mẫu 36 nam sinh viên có trung bình mẫu $\overline{X} \notin (1,6; 1,7)$ sẽ không xảy ra.

5.3. Khoảng tin cậy cho kì vọng

Cho biến ngẫu nhiên X của một tổng thể có $E(X) = \mu$ chưa biết, khoảng tin cậy cho kì vọng μ có dạng $l < \mu < u$. Để tìm l và u ta tiến hành các bước như sau:

(1) Cho trước một số $\alpha \in (0; 1)$ khá bé gọi là mức ý nghĩa;

(2) Với mẫu ngẫu nhiên $(X_1, X_2, ..., X_n)$, tìm hai hàm ngẫu nhiên:

$$L = L(X_1, X_2, ..., X_n)$$

$$U = U(X_1, X_2, ..., X_n)$$

sao cho:

$$P(L < \mu < U) = 1 - \alpha$$

(3) Nếu kết quả chọn mẫu ngẫu nhiên thu được mẫu số liệu $(x_1, x_2, ..., x_n)$ thì thay $X_1 = x_1, X_2 = x_2, ..., X_n = x_n$ vào L và U ta được $L(x_1, x_2, ..., x_n)$, $U(x_1, x_2, ..., x_n)$. Khi đó khoảng tin cậy cho kì vọng μ với mức ý nghĩa α là $L(x_1, x_2, ..., x_n) < \mu < U(x_1, x_2, ..., x_n)$.

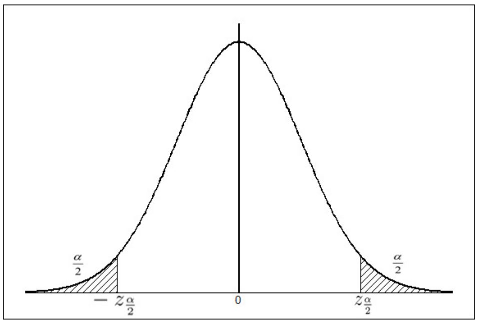
Giá trị $1 - \alpha$ gọi là độ tin cậy của ước lượng.

Cơ sở của phương áp ước lượng trên là dựa vào nguyên lý xác suất lớn. Do α nhỏ nên $1-\alpha$ lớn, vì vậy khi chọn một mẫu ngẫu nhiên $(X_1, X_2, ..., X_n)$, biến cố $L < \mu < U$ hầu như chắc chắn rằng sẽ xảy ra. Do đó, với kết quả chọn mẫu ngẫu nhiên $X_1 = x_1$, $X_2 = x_2$,..., $X_n = x_n$ ta được khoảng tin cậy cho kì vọng:

$$L(x_1, x_2, ..., x_n) < \mu < U(x_1, x_2, ..., x_n)$$

5.3.1. $X \sim N(\mu;\sigma^2)$ với σ^2 đã biết

Bài toán. Cho biến ngẫu nhiên X của một tổng thể có phân bố chuẩn $N(\mu;\sigma^2)$ với μ chưa biết và σ^2 đã biết. Tìm khoảng tin cậy cho μ với mức ý nghĩa α.



Hình 5.2

Nếu $(X_1, X_2, ..., X_n)$ là mẫu ngẫu nhiên của X thì

$$\overline{X} \sim N(\mu;\frac{\sigma^2}{n}).$$

Do đó:

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

Với $\alpha \in (0;1)$, gọi $z_{\alpha/2}$ thỏa mãn:

$$\frac{1}{\sqrt{2\pi}}\int_{z_{\alpha/2}}^{\infty} e^{-\frac{x^2}{2}}dx = \frac{\alpha}{2} \Leftrightarrow z_{\alpha/2} = \Phi^{-1}(1-\frac{\alpha}{2}).$$

Khi đó ta có:

$$P(-z_{\alpha/2} < \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2}) = 1 - \alpha$$

$$\Leftrightarrow P\left(\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Từ đó ta có định nghĩa:

> Cho biến ngẫu nhiên X có phân bố chuẩn $N(\mu;\sigma^2)$ với μ chưa biết và σ^2 đã biết. Nếu $(x_1, x_2, ..., x_n)$ là mẫu số liệu của X thì với độ tin cậy $1-\alpha$, khoảng tin cậy cho μ là:
>
> $$\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$
>
> Trong đó $z_{\alpha/2} = \Phi^{-1}(1 - \frac{\alpha}{2})$.

Ví dụ 5.3. Trọng lượng (kg) sản phẩm của công ty A có phân phối chuẩn $N(\mu;\sigma^2)$ với $\sigma = 1$ (kg). Chọn ngẫu nhiên 25 sản phẩm người ta tính được trung bình mẫu $\overline{x} = 50,1$ (kg). Với độ tin cậy 95% hãy tìm khoảng tin cậy cho trọng lượng trung bình của sản phẩm công ty A.

Giải. $\alpha = 0,05$ suy ra $z_{\alpha/2} = z_{0,025} = 1,96$;

$$z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 1,96\frac{1}{\sqrt{25}} = 0,4.$$

Với độ tin cậy 95%, khoảng tin cậy cho trọng lượng trung bình của sản phẩm: $49,7 < \mu < 50,5$. ■

Chọn cỡ mẫu

Từ công thức khoảng tin cậy cho μ ta thấy rằng sai số của ước lượng $|\overline{x} - \mu|$ bé hơn hoặc bằng $z_{\alpha/2}\frac{\sigma}{n}$. Do đó với độ tin cậy $1-\alpha$, nếu muốn có ước lượng μ có sai số không vượt quá Δ cho trước thì ta cần chọn cỡ mẫu n thỏa mãn:

$$z_{\alpha/2}\frac{\sigma}{n} < \Delta$$

Tương đương với:

$$n > \left(\frac{z_{\alpha/2}\sigma}{\Delta}\right)^2$$

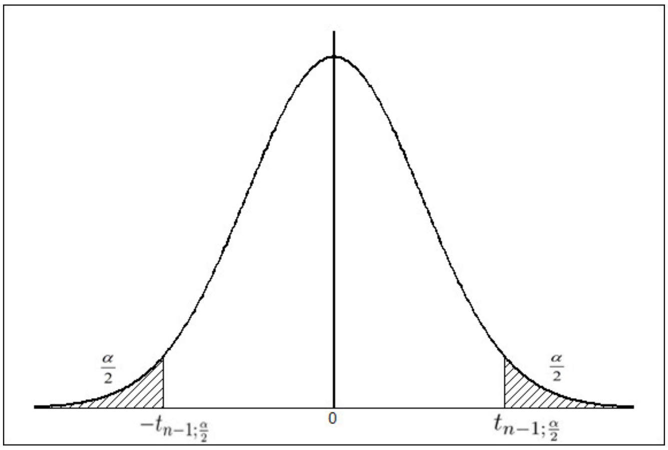
Khoảng tin cậy một phía

Khoảng tin cậy trong trường hợp $l = \infty$ hoặc $u = \infty$, thay $z_{\alpha/2}$ bởi z_α ta thu được khoảng tin cậy một phía như sau.

> Với độ tin cậy $1-\alpha$, khoảng tin cậy tối đa của μ là:
> $$\mu < \overline{x} + z_\alpha \frac{\sigma}{\sqrt{n}}$$
> Với độ tin cậy $1-\alpha$, khoảng tin cậy tối thiểu của μ là:
> $$\mu > \overline{x} - z_\alpha \frac{\sigma}{\sqrt{n}}$$

5.3.2. $X \sim N(\mu; \sigma^2)$ với σ^2 chưa biết

Bài toán: Cho biến ngẫu nhiên X của một tổng thể có phân bố chuẩn $N(\mu; \sigma^2)$ với μ chưa biết và σ^2 chưa biết. Tìm khoảng tin cậy cho μ với độ tin cậy $1-\alpha$.



Hình 5.3

Nếu $(X_1, X_2, ..., X_n)$ là mẫu ngẫu nhiên của X thì theo Định lí 4.2 ta có biến ngẫu nhiên:

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

có phân bố Student $n-1$ bậc tự do.

Với mức ý nghĩa α lấy giá trị $t_{n-1;\alpha/2}$ sao cho:

$$P(T_{n-1} \geq t_{n-1;\alpha/2}) = \alpha$$

Trong đó T_{n-1} là phân bố Student $n-1$ bậc tự do. Khi đó ta có:

$$P\left(-t_{n-1;\alpha/2} < \frac{\overline{X} - \mu}{S/\sqrt{n}} < t_{n-1;\alpha/2}\right) = 1 - \alpha$$

$$\Leftrightarrow P\left(\overline{X} - t_{n-1;\alpha/2}\frac{S}{\sqrt{n}} < \mu < \overline{X} - t_{n-1;\alpha/2}\frac{S}{\sqrt{n}}\right) = 1 - \alpha$$

Vì vậy, ta định nghĩa khoảng tin cậy cho μ cho trường hợp chưa biết phương sai như sau.

Cho biến ngẫu nhiên X có phân bố chuẩn $N(\mu; \sigma^2)$ với μ chưa biết và σ^2 chưa biết. Nếu $(x_1, x_2, ..., x_n)$ là mẫu số liệu của X thì với độ tin cậy $1-\alpha$, khoảng tin cậy cho μ là:

$$\overline{x} - t_{n-1;\alpha/2}\frac{s}{\sqrt{n}} < \mu < \overline{x} + t_{n-1;\alpha/2}\frac{s}{\sqrt{n}}$$

Trong đó $t_{n-1;\alpha/2}$ tra ở Bảng II, với $n > 30$: $t_{n-1;\alpha/2} \approx z_{\alpha/2}$.

Ví dụ 5.4. Một mẫu 16 pin dùng cho smartphone được chọn ngẫu nhiên của công ty A có tuổi thọ trung bình mẫu $\overline{x} = 24.308$ (giờ) và độ lệch chuẩn mẫu $s = 727$ (giờ). Giả sử rằng tuổi thọ pin smartphone có phân bố chuẩn. Với độ tin cậy 95%, hãy tìm khoảng tin cậy tuổi thọ trung bình smartphone được sản xuất bởi công ty A.

Giải. $t_{n-1;\alpha/2} = t_{15;0,025} = 2,1314;$

$$t_{n-1;\alpha/2}\frac{s}{\sqrt{n}} = 2,1314\frac{727}{\sqrt{16}} \approx 387.$$

Khoảng tin cậy tuổi thọ trung bình của pin smartphone công ty A với độ tin cậy 95% là:

$$23921 < \mu < 24695$$

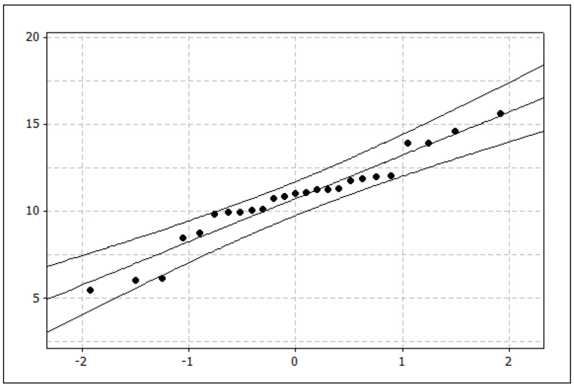
■

Ví dụ 5.5. Kết quả khảo sát hàm lượng asen trong nước máy sinh hoạt của 25 mẫu được chọn ngẫu nhiên ở thành phố A thu được như sau (đơn vị 10^{-3} mg/l).

6,06	11,07	10,77	15,67	10,15
5,51	10,87	12,02	12,08	6,17
9,96	9,95	13,95	10,06	11,25
8,79	8,52	11,31	13,92	11,81
11,27	11,12	11,89	14,64	9,83

Với độ tin cậy 95%, hãy tìm khoảng tin cậy cho hàm lượng asen trung bình trong nước máy sinh hoạt.

Giải. Để kiểm tra điều kiện phân bố chuẩn của tổng thể ta vẽ biểu đồ xác suất chuẩn.



Hình 5.4

Từ biểu đồ xác suất chuẩn (Hình 5.4) có thể kết luận hàm lượng asen trong nước máy sinh hoạt có phân bố chuẩn.

$\overline{x} = 10,75$; $s = 2,49$; $t_{n-1;\alpha/2} = t_{24;0,025} = 2,0639$;

$t_{n-1;\alpha/2}\dfrac{s}{\sqrt{n}} = 2,0639.\dfrac{2,49}{\sqrt{25}} \approx 1.$

Với độ tin cậy 95%, khoảng tin cậy cho hàm lượng asen trung bình trong nước máy sinh hoạt là:

$$9,75 < \mu < 11,75$$

■

Ví dụ 5.6. Một bài báo trong năm 1993 của Hiệp hội Thủy sản Mỹ báo cáo kết quả của một nghiên cứu để điều tra về ô nhiễm thủy ngân trong loài cá vược miệng rộng. Một mẫu cá đã được lựa chọn từ 53 hồ ở Florida, kết quả nồng độ thủy ngân được như sau (đơn vị: $10^{-4}\%$).

1,23	0,49	0,49	1,08	0,59	0,28	0,18	0,10	0,94
1,33	0,19	1,16	0,98	0,34	0,34	0,19	0,21	0,40
0,04	0,83	0,05	0,63	0,34	0,75	0,04	0,86	0,43
0,04	0,81	0,15	0,56	0,84	0,87	0,49	0,52	0,25
1,20	0,71	0,19	0,41	0,50	0,56	1,10	0,65	0,27
0,27	0,50	0,77	0,73	0,34	0,17	0,16	0,27	

Với độ tin cậy 95% hãy tìm khoảng tin cậy cho nồng độ thủy ngân trung bình có trong loài cá trên.

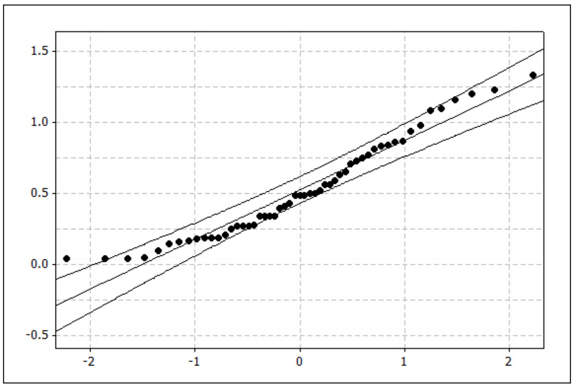
Giải. Biểu đồ xác suất chuẩn (Hình 5.5) cho thấy nồng độ thủy ngân có trong cá không có phân bố chuẩn, tuy nhiên do cỡ mẫu $n = 53 > 30$ nên áp dụng Định lí giới hạn trung tâm ta vẫn có thể tiến hành ước lượng.

$\overline{x} = 0,52$; $s = 0,35$; $t_{n-1;\alpha/2} \approx z_{0,025} = 1,96$;

$t_{n-1;\alpha/2}\dfrac{s}{\sqrt{n}} = 0,09.$

Với độ tin cậy 95%, khoảng tin cậy cho nồng độ thủy ngân trung bình trong cá vược là:

$$0,43 < \mu < 0,61$$



Hình 5.5

■

Khoảng tin cậy một phía

Với độ tin cậy $1-\alpha$, khoảng tin cậy tối đa cho kì vọng μ là:

$$\mu < \overline{x} + t_{n-1;\alpha}\frac{s}{\sqrt{n}}$$

Với độ tin cậy $1-\alpha$, khoảng tin cậy tối thiểu cho kì vọng μ là:

$$\overline{x} - t_{n-1;\alpha}\frac{s}{\sqrt{n}} < \mu$$

Ví dụ 5.7. Với độ tin cậy 95%, hãy tìm khoảng tin cậy tối đa và khoảng tin cậy tối thiểu hàm lượng asen trung bình trong Ví dụ 5.5.

Giải. $t_{n-1;\alpha} = t_{24;0,05} = 1,7109$;

$$t_{n-1;\alpha}\frac{s}{\sqrt{n}} = 1,7109.\frac{2,49}{\sqrt{49}} \approx 1.$$

Với độ tin cậy 95%, khoảng tin cậy tối đa hàm lượng asen trung bình trong nước máy sinh hoạt là: $\mu < 11,36$.

Với độ tin cậy 95%, khoảng tin cậy tối thiểu hàm lượng asen trung bình trong nước máy sinh hoạt là: $\mu > 10,314$. ■

5.4. Khoảng tin cậy cho tỷ lệ

Giả sử biến ngẫu nhiên X có phân bố Bernoulli với tham số p là tỷ lệ phần tử trong tổng thể có tính chất A nào đó. Trong phần này ta sẽ xây dựng công thức khoảng tin cậy cho p với mức ý nghĩa α.

Gọi $(X_1, X_2, ..., X_n)$ mẫu ngẫu nhiên của X. Đặt:

$$\hat{P} = \frac{X_1 + X_2 + ... + X_n}{n}$$

Theo Định lí giới hạn trung tâm, với n đủ lớn ta có:

$$\frac{\hat{P} - p}{\sqrt{p(1-p)}}\sqrt{n}$$

có phân bố xấp xỉ phân phối chuẩn tắc $N(0; 1)$. Với $\alpha \in (0; 1)$ cho trước, lấy $z_{\alpha/2} = \Phi^{-1}(1 - \frac{\alpha}{2})$, ta có:

$$P\left(-z_{\alpha/2} < \frac{\hat{P} - p}{\sqrt{p(1-p)}}\sqrt{n} < z_{\alpha/2}\right) \approx 1 - \alpha$$

Tương đương với:

$$P\left(\hat{P} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} < p < \hat{P} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right) \approx 1 - \alpha$$

Mặt khác, do $\hat{P}$ là một ước lượng của p nên ta có:

$$P\left(\hat{P} - z_{\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}} < p < \overline{X} + z_{\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}\right) \approx 1 - \alpha$$

Vì vậy, ta định nghĩa khoảng tin cậy cho p như sau.

> Nếu $\hat{p} = k/n$ là một ước lượng của tỷ lệ p từ 1 mẫu ngẫu nhiên kích thước n với $k \geq 10$ và $n - k \geq 10$ thì với độ tin cậy $1 - \alpha$, khoảng tin cậy cho p là:
>
> $$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Ví dụ 5.8. Với độ tin cậy 95% hãy tìm khoảng tin cậy cho tỷ lệ phế phẩm của một nhà máy biết rằng kiểm tra 100 sản phẩm của nhà máy thì thấy có 10 phế phẩm.

Giải.

$$\begin{cases} k = 10 \\ n = 100 \\ \alpha = 0,05 \end{cases} \Rightarrow \begin{cases} \hat{p} = 0,1 \\ z_{\alpha/2} = z_{0,025} = 1,96 \end{cases}$$

$$\epsilon = z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1,96\sqrt{\frac{0,1.0,9}{100}} = 0,059.$$

Với độ tin cậy 95%, khoảng tin cậy cho tỷ lệ phế phẩm của nhà máy là:

$$0,041 < p < 0,159$$

■

Khoảng tin cậy một phía

Cho $\hat{p} = k/n$ là một ước lượng của tỷ lệ p từ một mẫu ngẫu nhiên kích thước n.

Với độ tin cậy $1-\alpha$, khoảng tin cậy tối đa cho p là:

$$p < \hat{p} + z_\alpha\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Với độ tin cậy $1-\alpha$, khoảng tin cậy tối thiểu cho p là:

$$p > \hat{p} - z_\alpha\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

BÀI TẬP CHƯƠNG 5

▷ **5.1.** Công ty bao bì HP mới nhập về 1 lô hàng gồm 20.000 bao hạt nhựa của một nhà cung cấp quen. Dữ liệu quá khứ cho thấy khối lượng của các bao hạt nhựa này có phân bố chuẩn với phương sai $36(kg^2)$. Chọn ngẫu nhiên 25 bao hạt nhựa từ lô hàng trên để cân và thu được giá trị trung bình mẫu là 96 kg/bao. Với độ tin cậy 95% hãy ước lượng khoảng tin cậy khối lượng trung bình của 20.000 bao hạt nhựa này.

▷ **5.2.** Cho một ô tô chạy thử 32 lần trên đoạn đường từ A đến B người ta ghi nhận được lượng xăng hao phí như sau:

Lượng xăng hao phí (lít)	Tần số
$[9,6; 9,8)$	3
$[9,8; 10,0)$	7
$[10,0; 10,2)$	10
$[10,2; 10,4)$	8
$[10,4; 10,6)$	4

a) Vẽ biểu đồ tần số kiểm tra phân bố chuẩn.

b) Với độ tin cậy 95% hãy ước lượng khoảng tin cậy lượng xăng hao phí trung bình của xe trên chạy từ A đến B.

▷ **5.3.** Để định mức thời gian gia công một chi tiết máy, người ta theo dõi ngẫu nhiên quá trình gia công 25 chi tiết và thu được số liệu sau:

Thời gian gia công (phút)	Tần số
$[15; 17)$	1
$[17; 19)$	3
$[19; 21)$	4
$[21; 23)$	12
$[23; 25)$	3
$[25; 27)$	2

a) Vẽ biểu đồ tần số kiểm tra phân bố chuẩn.

b) Hãy ước lượng khoảng tin cậy thời gian gia công trung bình một chi tiếu máy với độ tin cậy $1 - \alpha = 0,95$.

▷ **5.4.** Để ước lượng khoảng tin cậy cho kích thước trung bình của chi tiết máy được gia công bởi của một máy gia công, người ta lấy ngẫu nhiên 25 chi tiết do máy đó gia công, đem đo và thu được các kích thước như sau:

24,1	27,2	26,7	23,6	26,4
25,8	27,3	23,2	26,9	27,1
22,7	26,9	24,8	24,0	23,4
24,5	26,1	25,9	25,4	22,9
26,4	25,4	23,3	23,0	24,3

a) Vẽ biểu đồ tần số, biểu đồ xác suất chuẩn để kiểm tra phân bố chuẩn.

b) Với độ tin cậy 95% hãy ước lượng khoảng tin cậy kích thước trung bình các chi tiết do máy đó gia công.

▷ **5.5.** Một nghiên cứu lương kĩ sư tốt nghiệp trường Đại học Bách khoa sau 3 năm, 32 kĩ sư được chọn ngẫu nhiên có kết quả như sau (đơn vị: triệu đồng):

12,3	5,1	9,2	14,3	10,2	5,9	14,7	12,8
14,7	11,5	15,6	11,8	12,0	11,9	13,7	9,7
11,6	14,1	12,8	9,6	12,1	12,6	4,0	11,2
10,7	11,3	13,9	6,8	13,4	10,1	10,5	9,6

a) Vẽ biểu đồ tần số, biểu đồ xác suất chuẩn.

b) Với độ tin cậy 95% hãy ước lượng khoảng tin cậy lương trung bình của kĩ sư tốt nghiệp sau 3 năm.

▷ **5.6.** Hãy ước lượng khoảng tin cậy tỷ lệ chính phẩm của một nhà máy với độ tin cậy 0,95 biết rằng kiểm tra 100 sản phẩm của nhà máy thì thấy có 10 phế phẩm.

▷ **5.7.** Mở 200 hộp của một kho đồ hộp, người ta thấy có 28 hộp bị biến chất. Với độ tin cậy 95%, hãy ước lượng khoảng tin cậy tỷ lệ đồ hộp biến chất ở trong kho.

▷ **5.8.** Trong đợt vận động bầu cử tổng thống người ta phỏng vấn ngẫu nhiên 1.600 cử tri thì được biết 960 người trong số đó sẽ bỏ phiếu cho ứng cử viên A. Với độ tin cậy 90%, hãy tìm khoảng tin cậy tỷ lệ phiếu bầu cho ứng cử viên A.

▷ **5.9.** Nhà máy A sản xuất 1 loại sản phẩm. Để ước lượng tỷ lệ thành phẩm người ta chọn ngẫu nhiên 400 sản phẩm và chia thành 40 nhóm để kiểm tra.

Số thành phẩm trong nhóm	1	2	3	4	5	6	7	8	9	10
Số nhóm	2	1	3	6	8	10	4	5	1	0

Với độ tin cậy 90% hãy ước lượng khoảng tin cậy tỷ lệ thành phẩm của nhà máy.

Chương 6
KIỂM ĐỊNH GIẢ THUYẾT

6.1. Khái niệm chung

6.1.1. Giả thuyết thống kê và kiểm định giả thuyết thống kê

Cho biến ngẫu nhiên X của một tổng thể có tham số θ.

Giả thuyết thống kê về θ là các khẳng định về giá trị của θ.

Ví dụ 6.1.

(1) μ là tuổi thọ trung bình của người Việt Nam. Giả thuyết thống kê có thể là: $\mu = 60$ (tuổi) hoặc $\mu > 60$, hoặc $\mu \neq 60$,...

(2) X là biến ngẫu nhiên đo trọng lượng sản phẩm của một nhà máy và σ^2 là phương sai của X. Giả thuyết thống kê có thể là $\sigma^2 = 1$, hoặc $\sigma^2 \neq 0,5$;...

(3) p là tỉ lệ phế phẩm của nhà máy A. Giả thuyết thống kê có thể là: $p < 0,1$ hoặc $p = 0,1$ hoặc $p \neq 0,1$;...

Ngoài giả thuyết thống kê về tham số, trong thống kê còn đưa ra nhiều giả thuyết thống kê khác như giả thuyết về phân bố, giả thuyết về độc lập,...

Bài toán kiểm định giả thuyết là một cặp giả thuyết thống kê mâu thuẫn nhau được đưa ra xem xét để chọn một giả thuyết đúng. Chẳng hạn, giả thuyết tỉ lệ phế phẩm của nhà máy: $p < 0,1$ và $p \geq 0,1$, giả thuyết tuổi thọ trung bình: $\mu = 60$ và $\mu \neq 60$. Một trong hai giả thuyết đó được giả định ban đầu là giả thuyết đúng, gọi là **giả thuyết gốc** và được kí hiệu là H_0. Giả thuyết còn lại gọi là **đối thuyết**, được kí hiệu là H_1.

Kiểm định giả thuyết là phương pháp sử dụng mẫu số liệu được chọn ngẫu nhiên để đưa ra quyết định bác bỏ H_0 hay chấp nhận H_0. Nếu từ mẫu số liệu có đủ cơ sở chứng tỏ H_0 sai thì bác bỏ H_0 và chấp nhận H_1 là giả thuyết đúng. Ngược lại, nếu từ mẫu số liệu chưa đủ cơ sở để cho rằng H_0 sai thì phải chấp nhận H_0 là giả thuyết đúng. Việc công nhận H_0 đúng ở đây cần hiểu là từ mẫu số liệu thu thập được chưa thể cho rằng H_0 sai, cần phải nghiên cứu tiếp.

Phương pháp kiểm định giả thuyết là dựa vào nguyên lý xác suất nhỏ. Ta sẽ tìm một biến cố A sao cho nếu H_0 đúng thì $P(A) = \alpha \approx 0$, tức là A hầu như chắc chắn rằng sẽ không xảy ra khi thực hiện phép thử trong điều kiện H_0 đúng. Nếu kết quả phép thử chỉ ra rằng biến cố A xảy ra thì bác bỏ H_0, còn nếu A không xảy ra thì chấp nhận H_0. Số thực α được gọi là *mức ý nghĩa* của kiểm định.

Ví dụ 6.2. Cách đây 10 năm về trước, chiều cao trung bình của nam thanh niên từ 22 tuổi là $1,55$ (m). Cho rằng chiều cao trung bình hiện nay (kí hiệu là μ) cao hơn, lúc đó ta lập bài toán kiểm định giả thuyết:

$$\begin{cases} H_0: & \mu = 1,55 \text{ (m)} \\ H_1: & \mu > 1,55 \text{ (m)} \end{cases}$$

Ở bài toán kiểm định giả thuyết trên H_0 được giả định ban đầu là giả thuyết đúng. Để bác bỏ H_0 ta tiến hành chọn một mẫu ngẫu nhiên kích thước n về chiều cao của nam thanh niên từ 22 tuổi là $(x_1, x_2, ..., x_n)$ và tính trung bình mẫu $\overline{x}$. Giá trị $\overline{x}$ càng lớn thì càng có cơ sở bác bỏ H_0.

Ví dụ 6.3. Tỉ lệ phế phẩm của một dây chuyền sản xuất là 15%. Sau khi tiến hành cải tiến kĩ thuật hy vọng rằng tỉ lệ phế phẩm sẽ thấp hơn 15%. Lúc đó ta lập bài toán kiểm định giả thuyết:

$$\begin{cases} H_0: & p = 0,15 \\ H_1: & p < 0,15 \end{cases}$$

Trong đó p là tỉ lệ phế phẩm sau cải tiến kĩ thuật.

H_0 được giả định ban đầu là giả thuyết đúng. Để bác bỏ H_0 ta tiến hành kiểm tra ngẫu nhiên n sản phẩm, gọi số sản phẩm loại I trong n sản phẩm là k. Khi đó $\hat{p} = k/n$ càng nhỏ thì càng có cơ sở bác bỏ H_0.

Thống kê kiểm định là một hàm của mẫu ngẫu nhiên được sử dụng để đưa ra quyết định bác bỏ hay chấp nhận H_0.

6.1.2. Sai lầm loại I và sai lầm loại II

Khi tiến hành kiểm định giả thuyết thống kê ta phải đưa ra quyết định hoặc là bác bỏ H_0 hoặc là chấp nhận H_0. Có 1 trong 4 trường hợp sau xảy ra:

- Bác bỏ H_0 và trên thực tế H_0 cũng là giả thuyết sai. Khi đó quyết định bác bỏ H_0 là quyết định đúng.

- Bác bỏ H_0 trong khi thực tế là H_0 là giả thuyết đúng. Khi đó quyết định bác bỏ H_0 là quyết định sai, sai lầm này gọi là sai lầm loại I.

- Chấp nhận H_0 và trên thực tế H_0 cũng là giả thuyết đúng. Khi đó quyết định chấp nhận H_0 là quyết định đúng.

- Chấp nhận H_0 trong khi thực tế là H_0 là giả thuyết sai. Khi đó quyết định chấp nhận H_0 là quyết định sai, sai lầm này gọi là sai lầm loại II.

Như vậy, khi quyết định bác bỏ H_0 thì có thể mắc sai lầm loại I, còn khi quyết định chấp nhận H_0 thì có thể phạm phải sai lầm loại II.

	H_0 đúng	H_0 sai
Bác bỏ H_0	sai lầm loại I	quyết định đúng
Chấp nhận H_0	quyết định đúng	sai lầm loại II

6.2. Kiểm định kì vọng của phân phối chuẩn

6.2.1. Đã biết phương sai

Cho biến ngẫu nhiên X của một tổng thể có phân bố chuẩn $N(\mu;\sigma^2)$ với kì vọng μ chưa biết và phương sai σ^2 đã biết. Xét bài toán kiểm định giả thuyết:

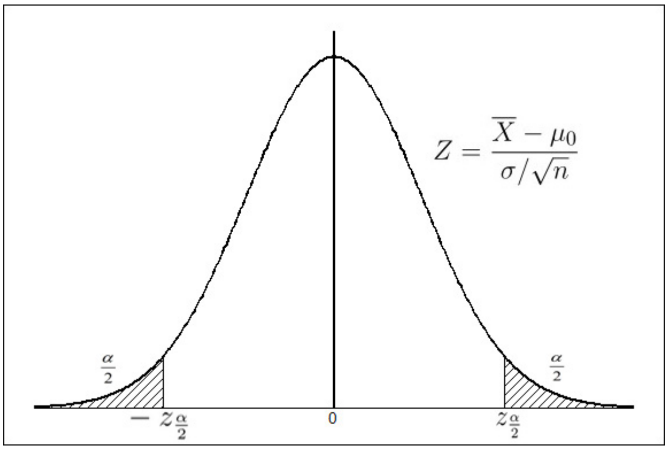
$$\begin{cases} H_0\colon \mu = \mu_0 \\ H_1\colon \mu \neq \mu_0 \end{cases}$$

Trong đó μ_0 là một số thực đã cho.

Giả sử rằng H_0 đúng, tức là $\mu = \mu_0$.
Gọi $(X_1, X_2, ..., X_n)$ là mẫu ngẫu nhiên của X. Khi đó:

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$$

có phân bố chuẩn tắc.



Hình 6.1

Vì vậy, với $\alpha \in (0;1)$ cho trước, lấy $z_{\alpha/2} = \Phi^{-1}(1 - \frac{\alpha}{2})$ ta có:

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

Suy ra:

$$P(Z \notin (-z_{\alpha/2}; z_{\alpha/2})) = \alpha$$

Với α là một số khá bé (gần bằng 0) thì biến cố chọn được một mẫu kích thước n có $Z \notin (-z_{\alpha/2}; z_{\alpha/2})$ hầu như chắc chắn sẽ không xảy ra. Vì vậy, nếu kết quả chọn ngẫu nhiên được 1 mẫu $(x_1, x_2, ..., x_n)$ thỏa mãn:

$$z = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} \in (-\infty; -z_{\alpha/2}] \cup [z_{\alpha/2}; +\infty)$$

thì bác bỏ H_0.

$W_\alpha = (-\infty; -z_{\alpha/2}] \cup [z_{\alpha/2}; +\infty)$ được gọi là *miền bác bỏ* H_0.

Lý luận tương tự ta có:

- Đối với bài toán kiểm định giả thuyết:

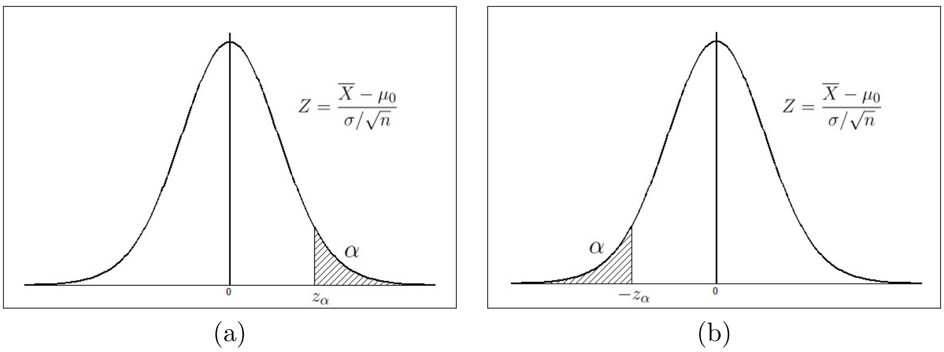
$$\begin{cases} H_0\text{: } \mu = \mu_0 \\ H_1\text{: } \mu > \mu_0 \end{cases}$$

Miền bác bỏ H_0 là $W_\alpha = [z_\alpha; +\infty)$ (xem Hình 6.2 (a)).

- Đối với bài toán kiểm định giả thuyết:

$$\begin{cases} H_0\text{: } \mu = \mu_0 \\ H_1\text{: } \mu < \mu_0 \end{cases}$$

Miền bác bỏ H_0 là $W_\alpha = (-\infty; -z_\alpha]$ (xem Hình 6.2 (b)).



Hình 6.2

Kết luận:

$X \sim N(\mu; \sigma^2)$ với σ^2 đã biết.
Giả thuyết gốc H_0: $\mu = \mu_0$
Giá trị thống kê kiểm định: $z = \dfrac{(\overline{x} - \mu_0)}{\sigma}\sqrt{n}$

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

Nếu $z \in W_\alpha$ thì bác bỏ H_0, nếu $z \notin W_\alpha$ thì chấp nhận H_0.

Ví dụ 6.4. Một nhà sản xuất máy tính xách tay quan tâm đến nguồn cấp điện cho máy tính, nguồn cấp đạt tiêu chuẩn đối với máy tính là 19 volt. Đo nguồn cấp điện của một mẫu 25 sạc pin được chọn ngẫu nhiên của hãng sản xuất A người ta tính được trung bình mẫu $\overline{x} = 19,25$. Giả sử guồn cấp điện của xạc pin trên có phân bố chuẩn với độ lệch chuẩn $\sigma = 0,5$ volt. Với mức ý nghĩa $\alpha = 0,05$ hãy kiểm định giả thuyết gốc H_0: $\mu = 19$ (volt) với đối thuyết H_1: $\mu \neq 19$ (volt) với μ là nguồn cấp điện trung bình của loại xạc pin trên.

Giải. Miền bác bỏ H_0 là $W = (-\infty; -1,96] \cup [1,96; +\infty)$.
$z = \dfrac{\overline{x} - \mu_0}{\sigma}\sqrt{n} = 2,5 \in W$. Do đó, có cơ sở bác bỏ H_0. ■

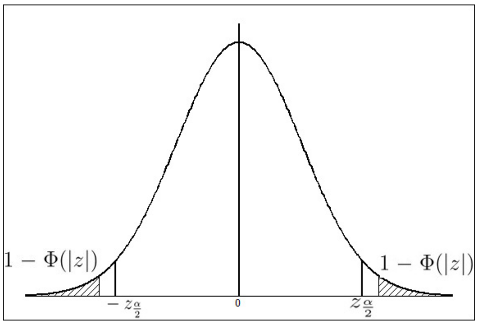
Ví dụ 6.5. Trong năm trước trọng lượng trung bình của bò xuất chuồng ở một trang trại là 380kg. Năm nay người ta áp dụng thử một chế độ ăn mới với hy vọng là bò sẽ tăng trọng nhanh hơn. Sau thời gian áp dụng thử người ta lấy ngẫu nhiên 50 con bò xuất chuồng đem cân và tính được trọng lượng trung bình của chúng là $\overline{x} = 390$kg. Với mức ý nghĩa $\alpha = 0,05$ có thể cho rằng trọng lượng trung bình của bò xuất chuồng đã tăng lên không? Giả sử rằng trọng lượng của bò có phân phối chuẩn với độ lệch chuẩn $\sigma = 25,2$kg.

Giải. Gọi μ là trọng lượng trung bình của bò áp dụng chế độ ăn mới. Bài toán kiểm định giả thuyết H_0: $\mu = 380$, H_1: $\mu > 380$.

Miền bác bỏ H_0 là $W = [z_{0,05}; +\infty) = [1,645; +\infty)$.

$z = \dfrac{\overline{x} - \mu_0}{\sigma}\sqrt{n} \approx 2,8 \in W$. Do đó, có cơ sở bác bỏ H_0. Tức là với mức ý nghĩa $\alpha = 0,05$ có thể cho rằng trọng lượng trung bình của bò xuất chuồng đã tăng lên. ■

p-giá trị



Hình 6.3

Trở lại bài toán kiểm định giả thuyết:

$$\begin{cases} H_0\colon \mu = \mu_0 \\ H_1\colon \mu \neq \mu_0 \end{cases}$$

Với mức ý nghĩa α, giả thuyết gốc H_0 bị bác bỏ khi và chỉ khi:

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

Tương đương với:

$$\Phi(z) \leq \frac{\alpha}{2} \text{ hoặc } 1 - \Phi(z) \leq \frac{\alpha}{2}$$

Tương đương với:

$$2(1 - \Phi(-z)) \leq \alpha \text{ hoặc } 2(1 - \Phi(z)) \leq \alpha$$

Tương đương với:

$$2(1 - \Phi(|z|)) \leq \alpha$$

Giá trị xác suất $2(1-\Phi(|z|))$ được gọi là p-giá trị. Tương tự như vậy ta có:

- Đối với bài toán kiểm định giả thuyết: $\begin{cases} H_0\colon \mu = \mu_0 \\ H_1\colon \mu > \mu_0 \end{cases}$

p-giá trị $= 1 - \Phi(z)$.

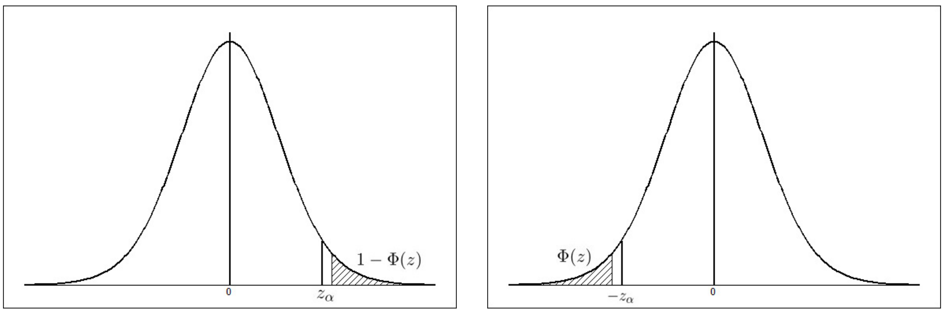
- Đối với bài toán kiểm định giả thuyết: $\begin{cases} H_0\colon \mu = \mu_0 \\ H_1\colon \mu < \mu_0 \end{cases}$

p-giá trị $= \Phi(z)$.

Khi đó, thủ tục kiểm định được thực hiện như sau:

1. Tính p-giá trị.

2. Nếu p-giá trị $\leq \alpha$ thì bác bỏ H_0, nếu p-giá trị $> \alpha$ thì chấp nhận H_0.



(a) p-giá trị $= 1 - \Phi(z)$ (b) p-giá trị $= \Phi(z)$

Hình 6.4

6.2.2. Chưa biết phương sai

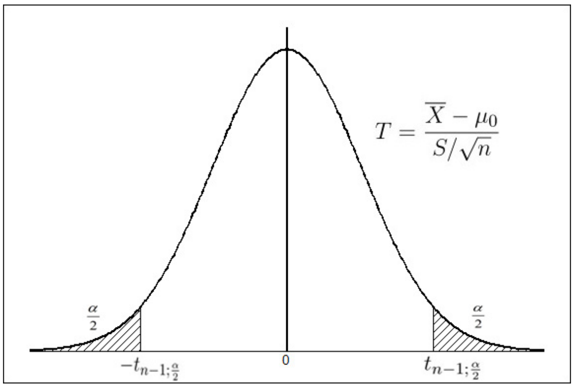
Cho biến ngẫu nhiên X của một tổng thể có phân bố chuẩn $N(\mu; \sigma^2)$ với kì vọng μ chưa biết và phương sai σ^2 chưa biết. Xét bài toán kiểm định giả thuyết:

$$\begin{cases} H_0\colon \mu = \mu_0 \\ H_1\colon \mu \neq \mu_0 \end{cases}$$

Giả sử H_0 đúng, khi đó $\mu = \mu_0$. Với $(X_1, X_2, ..., X_n)$ là mẫu ngẫu nhiên của X, khi đó:

$$T = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}$$

có phân bố student $n-1$ bậc tự do.



Hình 6.5

Với một số $\alpha \in (0;1)$ cho trước, lấy $t_{n-1;\alpha/2}$ thỏa mãn:

$$P(T_{n-1} > t_{n-1;\alpha/2}) = \frac{\alpha}{2}$$

Trong đó T_{n-1} là phân bố student $n-1$ bậc tự do. Khi đó ta có:

$$P(T \in (-\infty; -t_{n-1;\alpha/2}] \cup [t_{n-1;\alpha/2}; +\infty)) = \alpha$$

Với α là một số khá bé thì biến cố chọn ngẫu nhiên được một mẫu kích thước n có $T \in (-\infty; -t_{n-1;\alpha/2}] \cup [t_{n-1;\alpha/2}; +\infty)$ hầu như chắc chắn không xảy ra. Vì vậy, nếu chọn ngẫu nhiên được một mẫu số liệu $(x_1, x_2, ..., x_n)$ thỏa mãn:

$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} \in (-\infty; -t_{n-1;\alpha/2}] \cup [t_{n-1;\alpha/2}; +\infty)$$

thì bác bỏ H_0.

Như vậy, miền bác bỏ H_0 là:

$$W_\alpha = (-\infty; -t_{n-1;\alpha/2}] \cup [t_{n-1;\alpha/2}; +\infty)$$

Lý luận tương tự như trên ta có:

- Bài toán kiểm định giả thuyết:

$$\begin{cases} H_0\colon \mu = \mu_0 \\ H_1\colon \mu > \mu_0 \end{cases}$$

Miền bác bỏ H_0 là $W_\alpha = [t_{n-1;\alpha}; +\infty)$.

- Bài toán kiểm định giả thuyết:

$$\begin{cases} H_0\colon \mu = \mu_0 \\ H_1\colon \mu < \mu_0 \end{cases}$$

Miền bác bỏ H_0 là $W_\alpha = (-\infty; -t_{n-1;\alpha}]$.

Kết luận:

$X \sim N(\mu; \sigma^2)$ với σ^2 chưa biết.
Giả thuyết gốc H_0: $\mu = \mu_0$
Giá trị thống kê kiểm định: $t = \dfrac{\overline{x} - \mu_0}{s}\sqrt{n}$

Đối thuyết	Miền bác bỏ H_0	p-giá trị
H_1: $\mu \neq \mu_0$	$(-\infty; -t_{n-1;\alpha/2}] \cup [t_{n-1;\alpha/2}; +\infty)$	$2P(T_{n-1} > \lvert t\rvert)$
H_1: $\mu > \mu_0$	$[t_{n-1;\alpha}; +\infty)$	$P(T_{n-1} > t)$
H_1: $\mu < \mu_0$	$(-\infty; -t_{n-1;\alpha}]$	$P(T_{n-1} < t)$

Trong trường hợp $n > 30$: $t_{n-1;\alpha} \approx z_\alpha$.

Ví dụ 6.6. Tuổi thọ trung bình của một loại bóng đèn do nhà máy A sản xuất khi chưa cải tiến kĩ thuật là 2.000 giờ. Sau thời gian cải tiến kĩ thuật người ta chọn ngẫu nhiên 25 bóng đèn cho lắp thử nghiệm, kết quả thực nghiệm thu được tuổi thọ trung bình mẫu $\overline{x} = 2.010$ giờ và độ lệch chuẩn mẫu $s = 15$ giờ. Với mức ý nghĩa $0,025$ có thể kết luận "sau khi cải tiến kĩ thuật, tuổi thọ bóng đèn có tăng lên" không? Biết tuổi thọ bóng đèn có phân phối chuẩn.

Giải. Gọi μ là tuổi thọ trung bình của bóng đèn sau cải tiến kĩ thuật. Bài toán kiểm định giả thuyết H_0: $\mu = 2.000$, H_1: $\mu > 2.000$.

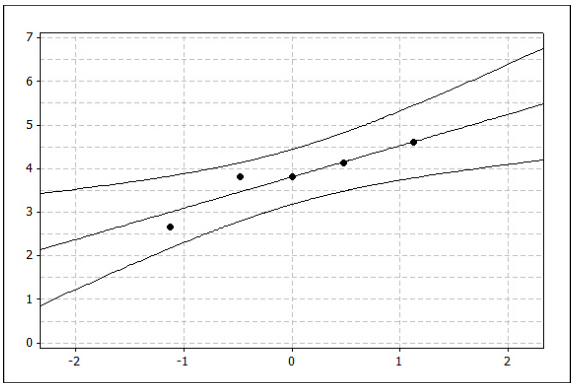
$t_{n-1;\alpha} = t_{24;0,025} = 2,0639$.

Miền bác bỏ H_0 là $W = [2,0639; +\infty)$.

$t = \dfrac{\overline{x} - \mu_0}{s}\sqrt{n} = 3,33 \in W$. Do đó, có cơ sở bác bỏ H_0, tức là có cơ sở để kết luận "sau khi cải tiến kĩ thuật, tuổi thọ bóng đèn có tăng lên". ■

Ví dụ 6.7. Một mẫu số liệu về nồng độ glycerol (mg/ml) trong rượu vang trắng của công ty A như sau: $2,67; 4,62; 4,14; 3,81; 3,83$. Giả sử nồng độ glycerol trung bình trong rượu vang trắng đạt tiêu chuẩn là 4 (mg/ml). Với mức ý nghĩa 5% có thể cho rằng nồng độ glycerol trung bình trong rượu vang trắng của công ty A đạt tiêu chuẩn không?

Giải. Từ biểu đồ xác suất chuẩn (Hình 6.6) ta có thể chấp nhận tổng thể có phân bố chuẩn.



Hình 6.6

Gọi μ là nồng độ glycerol trung bình trong rượu vang trắng của công ty A. Bài toán kiểm định giả thuyết:

$$\begin{cases} H_0\colon \mu = 4 \\ H_1\colon \mu \neq 4 \end{cases}$$

Miền bác bỏ H_0: $W = (-\infty; -2,7764] \cup [2,7764; +\infty)$.

Từ mẫu số liệu ta có $n = 5$; $\overline{x} = 3,814$; $s = 0,718$.

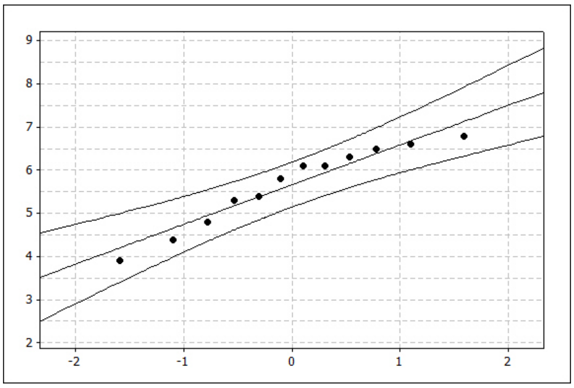
$t = \dfrac{3,814 - 4}{0,718}.\sqrt{5} \approx -0,6 \notin W$. Do đó, chưa có cơ sở bác bỏ H_0. ■

Ví dụ 6.8. Nguyên nhân gây ra hiện tượng mưa axít là do lượng khí thải SO_2 và NO_x tạo ra từ các quá trình phát triển sản xuất con người tiêu thụ nhiều than đá, dầu mỏ, từ khí thải sản xuất công nghiệp,... Nếu nước mưa có độ pH dưới 6,0 thì được xem là mưa axit. Với 12 mẫu nước mưa được lấy ngẫu nhiên ở một thành phố, phân tích cho kết quả độ pH như sau:

$$6,1 \quad 5,4 \quad 4,8 \quad 5,8 \quad 6,6 \quad 5,3 \quad 6,1 \quad 4,4 \quad 3,9 \quad 6,8 \quad 6,5 \quad 6,3$$

Với mức ý nghĩa $\alpha = 0,05$ có thể cho rằng độ pH trung bình trong nước mưa ở thành phố trên thấp hơn 6,0 không?

Giải. Từ biểu đồ xác suất chuẩn (Hình 6.7) có thể chấp nhận nồng độ pH trong nước mưa có phân bố chuẩn.



Hình 6.7

Gọi μ là độ pH trung bình trong nước mưa. Bài toán kiểm định giả thuyết:

$$\begin{cases} H_0\colon \mu = 6 \\ H_1\colon \mu < 6 \end{cases}$$

Miền bác bỏ H_0 là $W = (-\infty; -1,7959]$.

$t = \dfrac{\overline{x} - \mu_0}{s}\sqrt{n} = -1,25 \notin W$. Do đó, chưa thể bác bỏ H_0. ∎

Ví dụ 6.9. Tiến hành kiểm tra ngẫu nhiên 121 sản phẩm do xí nghiệp đó sản xuất và tính được trung bình mẫu $\overline{x} = 5,8$kg và độ lệch chuẩn mẫu $s = 1,4$kg. Với mức ý nghĩa $\alpha = 5\%$ có thể cho rằng trọng lượng trung bình sản phẩm của xí nghiệp là 6kg không? Biết rằng trọng lượng sản phẩm của xí nghiệp có phân phối chuẩn.

Giải. Gọi μ là trọng lượng sản phẩm trung bình của sản phẩm. Bài toán kiểm định giả thuyết:

$$\begin{cases} H_0\colon \mu = 6 \\ H_1\colon \mu \neq 6 \end{cases}$$

$t_{n-1;\alpha/2} = t_{120;0,025} \approx z_{0,025} = 1,96$.

Miền bác bỏ H_0 là $W = (-\infty; -1,96] \cup [1,96; +\infty)$.

$t = \dfrac{\overline{x} - \mu_0}{s}\sqrt{n} = -1,571 \notin W$. Do đó, chưa có cơ sở bác bỏ H_0. ■

6.3. So sánh 2 kì vọng của 2 phân phối chuẩn chưa biết 2 phương sai

Cho X và Y biến số ngẫu nhiên của hai tổng thể độc lập nhau và lần lượt có phân bố chuẩn $N(\mu_x; \sigma_x^2)$ và $N(\mu_y; \sigma_y^2)$. Trong mục này ta xét bài toán so sánh hai kì vọng μ_x và μ_y. Giả thiết quan trọng cho bài toán này là:

(i) $(X_1, X_2, ..., X_m)$ là một mẫu ngẫu nhiên của biến ngẫu nhiên $X \sim N(\mu_x; \sigma_x^2)$.
(ii) $(Y_1, Y_2, ..., Y_n)$ là một mẫu ngẫu nhiên của biến ngẫu nhiên $Y \sim N(\mu_y; \sigma_y^2)$.
(iii) Hai mẫu ngẫu nhiên trên độc lập với nhau.

Từ các giả thiết trên ta có kì vọng và phương sai của $\overline{X} - \overline{Y}$ lần lượt là:

$$E(\overline{X} - \overline{Y}) = \mu_x - \mu_y,\ V(\overline{X} - \overline{Y}) = \frac{\sigma_x^2}{m} + \frac{\sigma_y^2}{n}$$

Vì vậy, biến ngẫu nhiên:

$$Z = \frac{(\overline{X} - \overline{Y}) - (\mu_x - \mu_y)}{\sqrt{\dfrac{\sigma_x^2}{m} + \dfrac{\sigma_y^2}{n}}}$$

có phân phối chuẩn tắc $N(0; 1)$.

6.3.1. Cỡ mẫu lớn

Khi chưa biết hai phương sai σ_x^2 và σ_y^2 nhưng cỡ mẫu lớn thì ta sử dụng S_x và S_y để ước lượng σ_x và σ_y tương ứng và áp dụng Định lí giới hạn trung tâm ta có:

$$Z = \frac{(\overline{X} - \overline{Y}) - (\mu_x - \mu_y)}{\sqrt{\dfrac{S_x^2}{m} + \dfrac{S_y^2}{n}}}$$

có xấp xỉ phân phối chuẩn tắc $N(0; 1)$.

Từ đó ta có kết quả sau.

$X \sim N(\mu_x; \sigma_x^2)$ và $Y \sim N(\mu_y; \sigma_y^2)$ trong đó σ_x^2 và σ_y^2 đều chưa biết; $m > 30$ và $n > 30$.

Giả thuyết thống kê H_0: $\mu_x - \mu_y = \Delta_0$

Giá trị thống kê kiểm định: $z = \dfrac{\overline{x} - \overline{y} - \Delta_0}{\sqrt{\dfrac{s_x^2}{m} + \dfrac{s_y^2}{n}}}$

Đối thuyết	Miền bác bỏ H_0	p-giá trị
H_1: $\mu_x - \mu_y \neq \Delta_0$	$(\infty; -z_{\alpha/2}] \cup [z_{\alpha/2}; +\infty)$	$2(1 - \Phi(\lvert z \rvert))$
H_1: $\mu_x - \mu_y > \Delta_0$	$[z_\alpha; +\infty)$	$1 - \Phi(z)$
H_1: $\mu_x - \mu_y < \Delta_0$	$(\infty; -z_\alpha]$	$\Phi(z)$

Ví dụ 6.10. Người ta cân trẻ sơ sinh ở hai khu vực thành thị và nông thôn, kết quả thu được như sau:

Khu vực	Số trẻ	Trung bình mẫu	Phương sai mẫu
Nông thôn	$m = 60$	$\overline{x} = 3,0$ kg	$s_x^2 = 0,4$ kg^2
Thành thị	$n = 50$	$\overline{y} = 3,1$ kg	$s_y^2 = 0,5$ kg^2

Với mức ý nghĩa $0,05$ có thể coi trọng lượng trung bình của trẻ sơ sinh ở hai khu vực khác nhau không? Biết trọng lượng trẻ sơ sinh ở hai khu vực có phân phối chuẩn.

Giải. Gọi trọng lượng trung bình của trẻ sơ sinh ở nông thôn và thành thị lần lượt là μ_x (kg) và μ_y (kg). Bài toán kiểm định giả thuyết: H_0: $\mu_x = \mu_y$, H_1: $\mu_x \neq \mu_y$.

Miền bác bỏ H_0: $W = (-\infty; -1,96] \cup [1,96; +\infty)$.

$z = \dfrac{\overline{x} - \overline{y}}{\sqrt{\dfrac{s_x^2}{m} + \dfrac{s_y^2}{n}}} = -0,77 \notin W$. Do đó, chấp nhận H_0. ■

6.3.2. Cỡ mẫu nhỏ và hai phương sai bằng nhau

Giả sử rằng chưa biết hai phương sai σ_x^2 và σ_y^2 nhưng $\sigma_x^2 = \sigma_y^2 = \sigma^2$. Khi đó:

$$V(\overline{X} - \overline{Y}) = \frac{\sigma_x^2}{m} + \frac{\sigma_y^2}{n} = \sigma^2(\frac{1}{m} + \frac{1}{n})$$

Để ước lượng σ^2 ta sử dụng hàm ước lượng:

$$S_p^2 = \frac{(m-1)S_x^2 + (n-1)S_y^2}{m+n-2}$$

Ta có định lí sau.

Định lý 6.1. *Biến ngẫu nhiên*

$$\frac{(\overline{X} - \overline{Y}) - (\mu_x - \mu_y)}{S_p\sqrt{1/m + 1/n}}$$

có phân phối Student $m + n - 2$ bậc tự do.

Áp dụng định lí trên ta có kết quả sau.

$X \sim N(\mu_x; \sigma_x^2)$, $Y \sim N(\mu_y; \sigma_y^2)$ với $\sigma_x = \sigma_y$ chưa biết.
Giả thuyết thống kê H_0: $\mu_x - \mu_y = \Delta_0$
Giá trị thống kê kiểm định:

$$t = \frac{(\overline{x} - \overline{y}) - \Delta_0}{s_p\sqrt{\dfrac{1}{m} + \dfrac{1}{n}}}$$

Với $s_p^2 = \dfrac{(m-1)s_x^2 + (n-1)s_y^2}{m+n-2}$.

Đối thuyết	Miền bác bỏ H_0	p-giá trị
H_1: $\mu_x - \mu_y \neq \Delta_0$	$(\infty; -t_{m+n-2;\frac{\alpha}{2}}] \cup [t_{m+n-2;\frac{\alpha}{2}}; +\infty)$	$2P(T_{m+n-2} > \|t\|)$
H_1: $\mu_x - \mu_y > \Delta_0$	$[t_{m+n-2;\alpha}; +\infty)$	$P(T_{m+n-2} > t)$
H_1: $\mu_x - \mu_y < \Delta_0$	$(\infty; -t_{t_{m+n-2;\alpha}}]$	$P(T_{m+n-2} < t)$

Ví dụ 6.11. Một nghiên cứu được thực hiện đối với 20 người ở phường A và 19 người ở phường B trong một thành phố để xem thu nhập trung bình hàng năm của dân cư hai phường đó thực sự khác nhau hay không. Các mẫu số liệu thu được như sau:

Phường A: m=20, $\overline{x} = 18,27$, $s_x^2 = 8,74$.

Phường B: m=19, $\overline{y} = 16,78$, $s_x^2 = 6,58$.

Với mức ý nghĩa $0,05$ có thể cho rằng thu nhập trung bình của dân cư hai phường đó khác nhau hay không? Giả sử thu nhập hàng năm của dân cư hai phường đó có phân phối chuẩn và hai phương sai bằng nhau.

Giải. Gọi μ_x và μ_y tương ứng là thu nhập trung bình hàng năm của dân cư hai phường A và B. Bài toán kiểm định giả thuyết:

$$\begin{cases} H_0\colon \mu_x = \mu_y \\ H_1\colon \mu_x \neq \mu_y \end{cases}$$

Miền bác bỏ H_0 là $W = (-\infty; -1,96] \cup [1,96; +\infty)$.

Từ giả thiết bài toán ta tính được:

$$s_p^2 = \frac{(m-1)s_x^2 + (n-1)s_y^2}{m+n-2} = 2,773$$

$t = \dfrac{\overline{x} - \overline{y}}{s_p\sqrt{1/m + 1/n}} = 1,667 \notin W$. Do đó, chấp nhận H_0. Tức là chưa có cơ sở cho rằng thu nhập trung bình của dân cư hai phường đó khác nhau.

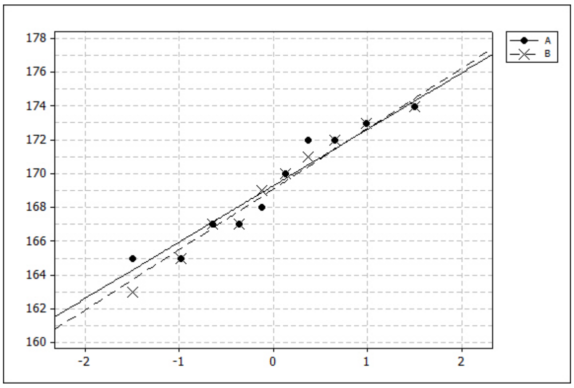
Hoặc tính p-giá trị $= 2P(T_{37} \geq 1,667) = 0,102 \geq \alpha$. ■

Ví dụ 6.12. Để so sánh chiều cao trung bình của nam thanh niên trưởng thành ở hai vùng dân cư A và B người ta chọn ngẫu nhiên 10 nam thanh niên trưởng thành ở vùng A và 10 nam thanh niên trưởng thành ở vùng B. Số đo chiều cao của hai nhóm người này được cho như sau (đơn vị: cm).

A	165	167	174	172	165	167	168	172	170	173
B	172	170	167	169	171	167	173	165	163	174

Với mức ý nghĩa 5% hãy so sánh chiều cao trung bình của nam thanh niên trưởng thành ở hai vùng dân cư trên.

Giải. Từ biểu đồ xác suất chuẩn (Hình 6.8) ta có thể chấp nhận hai tổng thể có phân phối chuẩn và hai phương sai bằng nhau.



Hình 6.8

Gọi μ_x và μ_y lần lượt là chiều cao trung bình của nam thanh niên trưởng thành ở vùng dân cư A và B.

Bài toán so sánh:

$$\begin{cases} H_0\colon \mu_x = \mu_y \\ H_1\colon \mu_x \neq \mu_y \end{cases}$$

Miền bác bỏ H_0: $W = (-\infty; -2,1] \cup [2,1; +\infty)$.

$\overline{x} = 169,3$; $s_x = 3,33$; $\overline{y} = 169,1$; $s_y = 3,57$; $s_p = 3,64$.

$t = 0,1 \notin W$. Do đó, chưa có cơ sở bác bỏ H_0.

Hoặc tính p-giá trị $= 0,898 > 5\%$. ■

6.3.3. Cỡ mẫu nhỏ và hai phương sai không bằng nhau

Giả sử rằng chưa biết hai phương sai σ_x^2 và σ_x^2 và $\sigma_x^2 \neq \sigma_y^2$. Nếu giả thuyết thống kê $H_0 : \mu_x - \mu_y = \Delta_0$ đúng thì:

$$T = \frac{\overline{X} - \overline{Y} - \Delta_0}{\sqrt{\dfrac{S_x^2}{m} + \dfrac{S_y^2}{n}}}$$

có phân phối Student ϑ bậc tự do (T_ϑ), trong đó ϑ là phần nguyên của:

$$\frac{\left(\dfrac{S_x^2}{m} + \dfrac{S_y^2}{n}\right)^2}{\dfrac{(S_x^2/m)^2}{m-1} + \dfrac{(S_y^2/n)^2}{n-1}}$$

Kết luận:

$X \sim N(\mu_1; \sigma_1^2)$, $Y \sim N(\mu_2; \sigma_2^2)$, chưa biết σ_x^2 và σ_y^2; $\sigma_1^2 \neq \sigma_2^2$.
Giả thuyết thống kê $H_0 : \mu_x - \mu_y = \Delta_0$
Giá trị thống kê kiểm định:

$$t = \frac{(\overline{x} - \overline{y}) - \Delta_0}{\sqrt{\dfrac{s_x^2}{m} + \dfrac{s_y^2}{n}}}$$

Đối thuyết	Miền bác bỏ H_0	p-giá trị
$H_1 : \mu_x - \mu_y \neq \Delta_0$	$(\infty; -t_{v;\frac{\alpha}{2}}] \cup [t_{v;\frac{\alpha}{2}}; +\infty)$	$2P(T_v > \|t\|)$
$H_1 : \mu_x - \mu_y > \Delta_0$	$[t_{v;\alpha}; +\infty)$	$P(T_v > t)$
$H_1 : \mu_x - \mu_y < \Delta_0$	$(\infty; -t_{t_{v;\alpha}}]$	$P(T_v < t)$

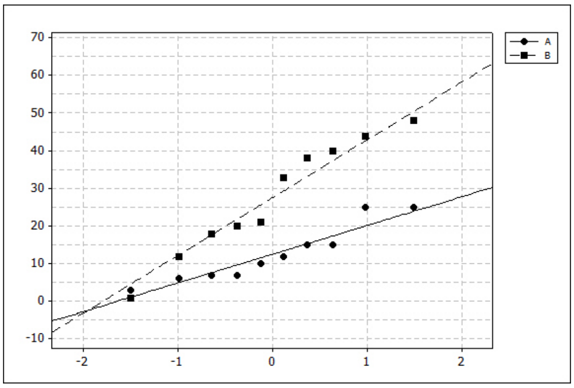
Trong đó υ là phần nguyên của:

$$\frac{\left(\dfrac{s_x^2}{m}+\dfrac{s_y^2}{n}\right)^2}{\dfrac{(s_x^2/m)^2}{m-1}+\dfrac{(s_y^2/n)^2}{n-1}}$$

Ví dụ 6.13. Hàm lượng asen trong 20 mẫu nước ngầm được lấy ngẫu nhiên ở hai vùng dân cư A và B được cho như sau (đơn vị: 10^{-3} mg/l):

A	3	7	25	10	15	6	12	25	15	7
B	48	44	40	38	33	21	20	12	1	18

Với mức ý nghĩa 5% có thể cho rằng hàm lượng asen ở vùng B cao hơn hàm lượng asen ở vùng A không?



Hình 6.9

Giải. Từ biểu đồ xác suất chuẩn (Hình 6.9) ta có thể khẳng định hai tổng thể có phân phối chuẩn và hai phương sai khác nhau.

Từ mẫu số liệu ta tính được $t=-2,8$, $\upsilon=13$.

Xét bài toán so sánh H_0: $\mu_x=\mu_y$, H_1: $\mu_x<\mu_y$.

Miền bác bỏ H_0: $W=(-\infty;-2,16]$.

Giá trị thống kê kiểm định $t=-2,8\in W$ nên bác bỏ H_0.

Hoặc tính p-giá trị $=P(T_{13}<-2,8)=0,008<0,05$. ■

6.4. So sánh cặp

Giả sử $(X;Y)$ là một cặp gồm 2 biến ngẫu nhiên (nói chung phụ thuộc nhau) với $E(X) = \mu_x$, $E(Y) = \mu_y$. Chúng ta muốn so sánh μ_x và μ_y.

Giả sử $(x_1;y_1),(x_2;y_2),...,(x_n;y_n)$ là 1 mẫu số liệu về $(X;Y)$. Do X và Y phụ thuộc nhau nên ta không thể dùng phương pháp so sánh ở mục trước. Để giải quyết bài toán này ta xét hiệu:

$$D = X - Y.$$

Khi đó ta được mẫu số liệu của D là $d_i = x_i - y_i$ $(i = 1,2,...,n)$ và $E(D) = \mu_d = \mu_x - \mu_y$. Giả thuyết gốc $H_0 : \mu_x - \mu_y = \Delta_0$ được đưa về $H_0 : \mu_d = \Delta_0$.

Như vậy ta đưa bài toán về kiểm định giả thuyết về kì vọng của biến ngẫu nhiên D.

Giả thuyết gốc: $H_0 : \mu_D = \Delta_0$

(trong đó $D = X - Y$ và $\mu_D = \mu_x - \mu_y$)

Giá trị thống kê kiểm định: $t = \dfrac{\overline{d} - \Delta_0}{s_D}\sqrt{n}$

(trong đó $d_1 = x_1 - y_1$, $d_2 = x_2 - y_2$,...,$d_n = x_n - y_n$)

Đối thuyết	Miền bác bỏ H_0	p-giá trị
$H_1 : \mu_D \neq \Delta_0$	$W = (\infty; -t_{\alpha/2,n-1}]$ $\cup[t_{\alpha/2,n-1}; +\infty)$	$2P(T_{n-1} > \|t\|)$
$H_1 : \mu_D > \Delta_0$	$W = [t_{\alpha,n-1}; +\infty)$	$P(T_{n-1} > t)$
$H_1 : \mu_D < \Delta_0$	$W = (\infty; -t_{\alpha,n-1}]$	$P(T_{n-1} < t)$

Ví dụ 6.14. Để khảo sát tác dụng của việc bón thêm 1 loại phân bón mới (A) người ta chia mỗi thửa ruộng thí nghiệm làm 2 mảnh. Một mảnh đối chứng không có phân bón A (X) và mảnh kia có phân bón A (Y). Sản lượng của 17 thửa ruộng được ghi lại như sau:

Thửa	X	Y	Thửa	X	Y
1	55,8	60,4	9	53,3	58,7
2	30,1	28,9	10	51,0	48,0
3	37,8	39,7	11	68,6	68,8
4	57,7	57,5	12	59,1	70,4
5	49,4	56,8	13	35,4	40,6
6	53,4	57,3	14	42,7	44,3
7	21,2	32,2	15	28,3	47,7
8	57,3	77,0	16	42,4	55,1
			17	61,4	66,1

Với mức ý nghĩa 5% hãy nhận xét việc bón phân có tác dụng không?

Giải.

Thửa	X	Y	D	Thửa	X	Y	D
1	55,8	60,4	-4,6	9	53,3	58,7	-4,7
2	30,1	28,9	1,2	10	51,0	48,0	-5,4
3	37,8	39,7	-1,9	11	68,6	68,8	3
4	57,7	57,5	0,2	12	59,1	70,4	-0,2
5	49,4	56,8	-7,4	13	35,4	40,6	-11,3
6	53,4	57,3	-3,9	14	42,7	44,3	-5,2
7	21,2	32,2	-11	15	28,3	47,7	-1,6
8	57,3	77,0	-19,7	16	42,4	55,1	-19,4
				17	61,4	66,1	-12,7

Bài toán kiểm định giả thuyết:

$$\begin{cases} H_0\colon \mu_x = \mu_y \\ H_1\colon \mu_x < \mu_y \end{cases}$$

Tương đương với toán kiểm định giả thuyết:

$$\begin{cases} H_0\colon \mu_D = 0 \\ H_1\colon \mu_D < 0 \end{cases}$$

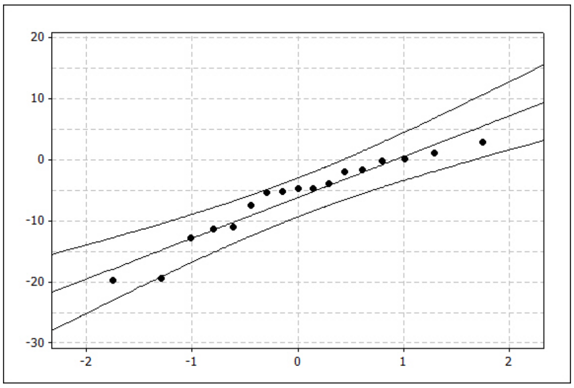
Từ biểu đồ xác suất chuẩn (Hình 6.10) của số liệu $d_i = x_i - y_i$ ta có thể chấp nhận $D = X - Y$ có phân phối chuẩn.

Miền bác bỏ H_0 là $W = (-\infty; t_{16;0,05}] = (-\infty; -1,753]$.

$\overline{d} = -6,15$; $s_D = 6,694$.

$t = \dfrac{\overline{d}}{s_D}\sqrt{n} = -3,79 \in W$. Do đó, có cơ sở bác bỏ H_0.

Hoặc tính p-giá trị $= P(T_{16} < -3,79) \approx 0$.



Hình 6.10

■

6.5. Kiểm định giả thuyết về tỷ lệ

Giả sử biến số ngẫu nhiên X của một tổng thể có phân bố Bernoulli với tham số p là tỷ lệ phần tử trong tổng thể có tính chất A nào đó.

Xét bài toán kiểm định giả thuyết

$$\begin{cases} H_0\colon p = p_0 \\ H_1\colon p \neq p_0 \end{cases}$$

Giả sử H_0 đúng, khi đó $p = p_0$.

Gọi $(X_1, X_2, ..., X_n)$ là mẫu ngẫu nhiên của X. Đặt
$$\hat{P} = \frac{X_1 + X_2 + ... + X_n}{n}$$
Với n đủ lớn, theo Định lí giới hạn trung tâm ta có:
$$Z = \frac{\hat{P} - p_0}{\sqrt{p_0(1 - p_0)/n}}$$
có phân bố xấp xỉ phân bố chuẩn tắc $N(0; 1)$.

Với một số $\alpha \in (0; 1)$ cho trước, lấy $z_{\alpha/2} = \Phi^{-1}(1 - \frac{\alpha}{2})$, ta có:
$$P(Z \in (-z_{\alpha/2}; z_{\alpha/2})) = 1 - \alpha$$
Suy ra:
$$P(Z \notin (-z_{\alpha/2}; z_{\alpha/2})) = \alpha$$

Với α là một số khá bé thì biến cố chọn được một mẫu kích thước n có $Z \notin (-z_{\alpha/2}; z_{\alpha/2})$ hầu như chắc chắn không xảy ra khi thực hiện 1 lần. Vì vậy nếu chọn ngẫu nhiên được 1 mẫu số liệu $(x_1, ..., x_n)$ có:
$$\hat{p} = \frac{x_1 + x_2 + ... + x_n}{n} = \frac{k}{n}$$
thỏa mãn:
$$z = \frac{k/n - \mu_0}{\sqrt{p_0(1 - p_0)/n}} \in (-\infty; -z_{\alpha/2}] \cup [z_{\alpha/2}; +\infty)$$
thì bác bỏ H_0.

Vì vậy, miền bác bỏ H_0 là $W_\alpha = (-\infty; -z_{\alpha/2}] \cup [z_{\alpha/2}; +\infty)$.

Lý luận tương tự ta như trên ta có:

- Đối với bài toán kiểm định giả thuyết:
$$\begin{cases} H_0\text{: } p = p_0 \\ H_1\text{: } p > p_0 \end{cases}$$
Miền bác bỏ H_0 là $W_\alpha = [z_\alpha; +\infty)$.

- Đối với bài toán kiểm định giả thuyết:

$$\begin{cases} H_0\colon p = p_0 \\ H_1\colon p < p_0 \end{cases}$$

Miền bác bỏ H_0 là $W_\alpha = (-\infty; -z_\alpha]$.

Kết luận:

Cho $\hat{p} = k/n$ là một ước lượng của tỷ lệ p từ 1 mẫu ngẫu nhiên kích thước n.

Giả thuyết gốc H_0: $p = p_0$

Giá trị thống kê kiểm định: $z = \dfrac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n}$

Đối thuyết	Miền bác bỏ H_0	p-giá trị
$H_1 : p \neq p_0$	$(-\infty; -z_{\alpha/2}] \cup [z_{\alpha/2}; +\infty)$	$2(1 - \Phi(\lvert z\rvert))$
H_1: $p > p_0$	$[z_\alpha; +\infty)$	$1 - \Phi(z)$
H_1: $p < p_0$	$(-\infty; -z_\alpha]$	$\Phi(z)$

Ví dụ 6.15. Giám đốc một công ty tuyên bố 90% sản phẩm của công ty đạt tiêu chuẩn quốc gia. Một công ty kiểm định độc lập đã tiến hành kiểm tra 200 sản phẩm của công ty đó thì thấy có 168 sản phẩm đạt yêu cầu. Với mức ý nghĩa $\alpha = 0,05$ có thể cho rằng tỷ lệ sản phẩm đạt tiêu chuẩn quốc gia thấp hơn 90% không?

Giải.

Gọi p là tỷ lệ sản phẩm của công ty đạt chuẩn quốc gia.

Bài toán kiểm định giả thuyết:

$$\begin{cases} H_0\colon p = 0,9 \\ H_1\colon p < 0,9 \end{cases}$$

$\alpha = 0,05 \Rightarrow z_\alpha = z_{0,05} = 1,645$.

Miền bác bỏ H_0 là $W = (-\infty; -1,645]$.

$z = \dfrac{k/n - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n} \approx -2,83 \in W$. Do đó, có thể bác bỏ H_0, tức là không có cơ sở để tin vào tuyên bố của vị giám đốc trên. ■

6.6. So sánh hai tỷ lệ

$X \sim Ber(p_1)$, $Y \sim Ber(p_2)$, X và Y độc lập.

Cho $\hat{p}_1 = k/m$ và $\hat{p}_2 = l/n$ lần lượt là ước lượng của p_1 và p_2 từ hai mẫu ngẫu nhiên độc lập.

Giả thuyết gốc H_0: $p_1 = p_2$

Giá trị thống kê kiểm định:

$$z = \frac{\dfrac{k}{m} - \dfrac{l}{n}}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{m} + \dfrac{1}{n}\right)}} \text{ với } \hat{p} = \frac{k+l}{m+n}$$

Đối thuyết	Miền bác bỏ H_0	p-giá trị
$H_1 : p_1 \neq p_2$	$(-\infty; -z_{\alpha/2}] \cup [z_{\alpha/2}; +\infty)$	$2(1 - \Phi(\lvert z \rvert))$
$H_1 : p_1 > p_2$	$[z_\alpha; +\infty)$	$1 - \Phi(z)$
$H_1 : p_1 < p_2$	$(-\infty; -z_\alpha]$	$\Phi(z)$

Ví dụ 6.16. Kiểm tra ngẫu nhiên các sản phẩm cùng loại do hai nhà máy sản xuất thu được số liệu sau:

Nhà máy	Số sản phẩm được kiểm tra	số phế phẩm
A	m=1000	k=20
B	n=900	l=30

Với mức ý nghĩa $\alpha = 0,05$ có thể coi tỷ lệ phế phẩm của hai nhà máy trên bằng nhau không?

Giải.

Gọi p_1 và p_2 lần lượt là tỉ lệ phế phẩm của nhà máy A và B.

Bài toán kiểm định giả thuyết:

$$\begin{cases} H_0\colon p_1 = p_2 \\ H_1\colon p_1 \neq p_2 \end{cases}$$

Miền bác bỏ H_0 là $W = (-\infty; -1,96] \cup [1,96; +\infty)$.

$\hat{p} = \dfrac{k+l}{m+n} = 0,0263.$

$z = \dfrac{\dfrac{k}{m} - \dfrac{l}{n}}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{m}+\dfrac{1}{n}\right)}} = -1,81 \notin W$. Do đó, chưa có cơ sở bác bỏ H_0. ■

BÀI TẬP CHƯƠNG 6

▷ **6.1.** Trung tâm hỗ trợ người tiêu dùng nhận được khá nhiều lời phàn nàn về sản phẩm bột giặt loại 4kg của công ty SC. Để hỗ trợ người tiêu dùng, Trung tâm tiến hành chọn ngẫu nhiên 36 gói bột giặt của công ty để cân và thu được kết quả trung bình mẫu 3,95kg. Giả sử trọng lượng bột giặt sản xuất của công ty tuân theo quy luật phân phối chuẩn với độ lệch chuẩn là 0,15kg.

a) Với mức ý nghĩa 5%, trung tâm có thể cho rằng trọng lượng trung bình của sản phẩm thấp hơn 4kg không?

b) Với mức ý nghĩa 2%, trung tâm có thể cho rằng trọng lượng trung bình của sản phẩm thấp hơn 4kg không?

▷ **6.2.** Trọng lượng (X) sản phẩm do nhà máy sản xuất ra là biến ngẫu nhiên phân phối chuẩn với độ lệch chuẩn $\sigma = 2$kg và trọng lượng trung bình là 20kg. Nghi ngờ máy hoạt động không bình thường làm thay đổi trọng lượng trung bình của sản phẩm người ta cân thử 100 sản phẩm và thu được kết quả sau:

Trọng lượng sản phẩm	19	20	21	22	23
Số sản phẩm	10	50	20	15	5

Với mức ý nghĩa 0,05 hãy kết luận về điều nghi ngờ trên.

▷ **6.3.** Mỳ chính được đóng gói 453g một gói trên máy tự động. Có thể coi trọng lượng các gói mỳ chính tuân theo quy luật chuẩn với độ lệch chuẩn 36g. Kiểm tra ngẫu nhiên 81 gói thấy trọng lượng trung bình là 448g. Với mức ý nghĩa 0,05 có thể kết luận trọng lượng các gói mỳ chính có xu hướng bị đóng thiếu không?

▷ **6.4.** Một nhà máy sản xuất bánh ngọt tuyên bố rằng mỗi chiếc bánh của họ trung bình có 88 calo. Một mẫu ngẫu nhiên với 46 chiếc bánh được kiểm tra cho thấy lượng calo trung bình trong mỗi chiếc bánh là 90 calo với độ lệch tiêu chuẩn là 4 calo. Với mức ý nghĩa 5%, kiểm định xem có phải trên thực tế mỗi chiếc bánh về trung bình chứa nhiều hơn 88 calo hơn hay không?

▷ **6.5.** Năng suất lúa trung bình của giống lúa A được công bố là 43 tạ/ha. Một nhóm gồm 60 thửa ruộng thí nghiệm được kiểm tra cho thấy năng suất lúa trung bình của nhóm là 46,2 tạ/ha với độ lệch chuẩn 12 tạ/ha. Với mức ý nghĩa 5%, nhận định xem có phải công bố là thấp hơn so với sự thật không?

▷ **6.6.** Điều tra giá của một loại hàng hóa A tại 100 cửa hàng được chọn ngẫu nhiên thu được số liệu sau:

Giá (nghìn đồng)	95	95,5	96	96,5	97	97,5	98	98,5	99
Số cửa hàng	7	12	15	17	20	13	8	6	2

a) Với độ tin cậy 95% hãy ước lượng khoảng tin cậy giá trung bình của loại hàng A trên thị trường.

b) Giá niêm yết loại hàng hóa A của công ty là 97 (nghìn đồng). Với mức ý nghĩa 5% có thể cho rằng giá trung bình loại hàng hóa A trên thị trường thấp hơn giá niêm yết của công ty không?

▷ **6.7.** Khảo sát lượng nước tiêu thụ trong 1 tháng của 36 hộ gia đình 4 người được chọn ngẫu nhiên trên địa bàn DN có mẫu số liệu sau (đơn vị: m^3).

26	27	32	26	26	27	26	19	26
30	30	24	24	23	30	28	32	24
24	24	28	31	24	29	29	27	28
30	26	27	24	30	25	26	28	28

a) Vẽ biểu đồ xác suất chuẩn.

b) Với mức ý nghĩa 5% có thể cho rằng lượng nước tiêu thụ trung bình trong một tháng của hộ gia đình 4 người cao hơn $26m^3$ không?

c) Cần chọn ngẫu nhiên ít nhất bao nhiêu hộ gia đình 4 người để lượng nước tiêu thụ trung bình trong 1 tháng của các hộ gia đình này cao hơn 26m^3 có xác suất lớn hơn 95%.

▷ **6.8.** Kết quả khảo sát hàm lượng sắt trong nước biển ở bãi tắm TH của 49 mẫu được chọn ngẫu nhiên thu được như sau (đơn vị 10^{-1} mg/l).

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

Theo tiêu chuẩn của Bộ Y tế nước máy sinh hoạt có hàm lượng sắt tối đa cho phép là 0,5mg/l. Với mức ý nghĩa 0,03 có thể cho rằng hàm lượng sắt trung bình cao hơn 0,5mg/l không?

▷ **6.9.** Lô hàng đủ tiêu chuẩn xuất khẩu nếu tỷ lệ phế phẩm không vượt quá 3%. Kiểm tra ngẫu nhiên 400 sản phẩm của lô hàng thấy có 14 phế phẩm. Với mức ý nghĩa $\alpha = 0,05$ có cho phép lô hàng xuất khẩu được hay không?

▷ **6.10.** Tỷ lệ phế phẩm do một nhà máy tự động sản xuất là 5%. Kiểm tra ngẫu nhiên 300 sản phẩm thấy có 24 phế phẩm. Nên có ý kiến cho rằng tỷ lệ phế phẩm do nhà máy sản xuất có chiều hướng tăng lên. Hãy kết luận ý kiến trên với mức ý nghĩa $\alpha = 0,05$.

▷ **6.11.** Một tỉnh báo cáo tỷ lệ học sinh tốt nghiệp của họ là 88%. Một mẫu ngẫu nhiên 100 học sinh được chọn thì chỉ có 82 em đỗ. Với mức ý nghĩa 5% kiểm định xem báo cáo của tỉnh có cao hơn sự thật.

▷ **6.12.** Tại thành phố M, mỗi hộ dùng không quá một điện thoại bàn và các điện thoại bàn chỉ sử dụng dịch vụ của một trong 2 công ty viễn thông A và B. Điều tra ngẫu nhiên 3.600 hộ tại thành phố M thấy có 2.500 hộ dùng điện thoại bàn, trong đó có 1.300 hộ dùng điện thoại bàn sử dụng dịch vụ viễn thông của công ty A.

a) Với độ tin cậy 95% hãy ước lượng khoảng tin cậy tỷ lệ hộ dùng điện thoại bàn tại thành phố M.

b) Với mức ý nghĩa 1% có thể cho rằng số điện thoại bàn sử dụng dịch vụ viễn thông của công ty A nhiều công ty B không?

▷ **6.13.** Công ty truyền hình cáp SV đã lắp đặt truyền hình cáp cho 8.000 hộ ở địa phương F. Để mở rộng kinh doanh và dự định nâng cấp chương trình truyền hình cáp tốt hơn, công ty SV điều tra 10.000 hộ ở địa phương F và thấy có 3.600 hộ lắp đặt truyền hình cáp. Trong số 3.600 hộ lắp đặt truyền hình cáp đó có 720 hộ lắp đặt truyền hình cáp của công ty SV.

a) Với độ tin cậy 95% hãy ước lượng khoảng tin cậy tỷ lệ hộ lắp đặt truyền hình cáp tại địa phương F.

b) Với độ tin cậy 95% hãy ước lượng khoảng tin cậy số hộ lắp đặt truyền hình cáp tại địa phương F.

c) Trong số 720 hộ lắp đặt truyền hình cáp SV đó, có 400 hộ đồng ý nâng cấp chương trình truyền hình. Biết rằng nếu có trên 50% khách hàng đồng ý nâng cấp chương trình thì công ty SV sẽ nâng cấp. Với mức ý nghĩa $\alpha = 0,025$, hỏi công ty SV có nâng cấp chương trình không?

Biết rằng mỗi hộ chỉ lắp đặt truyền hình cáp của 1 công ty.

▷ **6.14.** Doanh số (triệu đồng) bán ra của một nhà hàng A có phân phối chuẩn. Theo dõi doanh số bán ra của nhà hàng A trong 100 ngày có số liệu như sau:

Doanh số (triệu đồng)	118	123	127	135	140
Số ngày	5	26	40	20	9

Chủ nhà hàng báo cáo với nhân viên thu thuế là doanh số bán ra trung bình của nhà hàng trong 1 ngày là 127 triệu đồng. Nhân viên thu thuế nghi ngờ doanh số bán ra trung bình của nhà hàng A lớn hơn 127 triệu đồng. Dựa vào kết quả của mẫu ở trên, hãy tìm $\alpha \in (0; 0,5]$ sao cho với mức ý nghĩa α đó chưa có cơ sở để bác bỏ báo cáo của chủ nhà hàng.

▷ **6.15.** Để so sánh trọng lượng trung bình của trẻ sơ sinh ở thành thị và nông thôn người ta cân thử 1.000 trẻ ở hai khu vực và thu được số liệu:

Vùng	Số trẻ	Trung bình mẫu	Độ lệch chuẩn mẫu
Nông thôn	800	3,0kg	0,3kg
Thành thị	200	3,2kg	0,3kg

Với mức ý nghĩa $\alpha = 0,05$ có thể coi trọng lượng trung bình của trẻ sơ sinh ở thành thị cao hơn ở nông thôn hay không? Giả thiết trọng lượng trẻ sơ sinh có phân phối chuẩn.

▷ **6.16.** Người ta nghiên cứu năng suất lúa mỳ ở hai vùng chế độ canh tác khác nhau, kết quả thu được như sau:

Vùng	Số thửa ruộng	Trung bình mẫu	Phương sai mẫu
A	39	24,6 (tạ/ha)	0,24 (tạ/ha)^2
B	46	25,8 (tạ/ha)	0,16 (tạ/ha)^2

Với mức ý nghĩa $\alpha = 0,05$ hỏi có sự khác nhau đáng kể về năng suất lúa trung bình giữa hai vùng đất canh tác không?

▷ **6.17.** Kiểm tra chất lượng của hai lô sản phẩm người ta thấy ở lô thứ nhất trong 500 sản phẩm được kiểm tra có 50 phế phẩm, ở lô thứ hai trong 400 sản phẩm được kiểm tra có 60 phế phẩm. Với mức ý nghĩa $\alpha = 0,05$ có thể xem tỷ lệ phế phẩm của hai lô hàng bằng nhau không?

▷ **6.18.** Độ tinh khiết của một chất xúc tác rất quan trọng trong nghiên cứu hóa học. Người ta thử nghiệm hai phương pháp khác nhau: bằng phương pháp I (hữu cơ) làm 32 mẫu và bằng phương pháp II (vô cơ) làm 36 mẫu. Kết quả thu được như sau (lượng chất bẩn trên một đơn vị chất):

Phương pháp I

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

Phương pháp II

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

Với mức ý nghĩa $\alpha = 0,05$ có thể coi lượng chất bẩn trung bình của hai phương pháp trên là khác nhau được không?

▷ **6.19.** Một nghiên cứu viên muốn thực hiện một chương trình nhằm phát triển khả năng tập đọc của trẻ mẫu giáo. Để nghiên cứu chương trình này, nghiên cứu viên chọn ngẫu nhiên 6 cặp trẻ sinh đôi, một em trong mỗi cặp được chọn vào nhóm thực nghiệm, em kia được đưa vào nhóm đối chứng. Trong đó nhóm thực nghiệm sinh hoạt mỗi ngày 1 giờ trong một căn phòng được trang bị toàn những đồ chơi có tính cách giáo dục, nhóm đối chứng được chơi mỗi ngày 1 giờ trong một phòng có toàn đồ chơi không mang tính giáo dục. Thành tích học tập của mỗi em được đo lường bằng tuổi (tính theo số tháng) vào lúc mỗi em bắt đầu đọc được chữ trên một cuốn sách vỡ lòng. Kết quả nghiên cứu như sau:

Cặp sinh đôi	Nhóm thực nghiệm (X)	Nhóm đối chứng (Y)
Cặp 1	60	64
Cặp 2	54	57
Cặp 3	62	68
Cặp 4	71	73
Cặp 5	57	63
Cặp 6	63	66

a) Vẽ biểu đồ xác suất chuẩn của hiệu $X - Y$.

b) Với mức ý nghĩa 5% hãy kiểm định giả thuyết số tháng trung bình trẻ bắt đầu đọc được chữ trên một cuốn sách vỡ lòng của nhóm thực nghiệm thấp hơn số tháng trung bình trẻ bắt đầu đọc được chữ trên một cuốn sách vỡ lòng của nhóm đối chứng.

▷ **6.20.** Khảo sát 395 giáo viên tiểu học và 266 giáo viên THCS được chọn ngẫu nhiên cho kết quả có 224 giáo viên tiểu học và

126 giáo viên THCS hài lòng với công việc của họ. Với mức ý nghĩa 5% có thể cho rằng tỷ lệ giáo viên hài lòng với công việc ở bậc tiểu học và bậc THCS là như nhau không?

▷ **6.21.** Chọn ngẫu nhiên 200 thí sinh thi THPT Quốc gia 2015 ở hai hội đồng thi DND và DHU cho kết quả điểm thi môn văn như sau:

	Cỡ mẫu	Trung bình mẫu	Độ lệch chuẩn mẫu
DND	$m = 100$	$\overline{x} = 4,8$	$s_x = 1,3$
DHU	$n = 100$	$\overline{y} = 5,1$	$s_y = 1,9$

Với mức ý nghĩa 5% có thể cho rằng mức điểm trung bình môn Văn ở DND bằng điểm trung bình môn Văn ở DHU không? Biết điểm thi môn Văn ở hai hội đồng thi trên có phân bố chuẩn.

▷ **6.22.** Người ta làm một bài kiểm tra trên 15 sinh viên để tìm hiểu hiệu quả của việc giảng dạy một vấn đề theo phương pháp mới. Trước khi học phương pháp này sinh viên được cho làm một bài kiểm tra và sau khi học làm bài kiểm tra thứ 2 (điểm chấm theo thang điểm 100). Kết quả thu được như sau:

Số thứ tự	1	2	3	4	5	6	7	8
Trước học	54	79	91	75	68	43	33	85
Sau học	66	85	83	88	93	40	78	91

Số thứ tự	9	10	11	12	13	14	15	
Trước học	22	56	73	63	29	75	87	
Sau học	44	82	59	81	64	83	81	

Với mức ý nghĩa $\alpha = 0,05$ có thể nói rằng học theo phương pháp mới hiệu quả hơn không?

Chương 7
KIỂM ĐỊNH KHI BÌNH PHƯƠNG

7.1. Kiểm định giả thuyết về quy luật phân phối xác suất

Giả sử $\{x_1; x_2, ..., x_n\}$ là mẫu số liệu các giá trị của biến ngẫu nhiên X mà ta chưa biết quy luật phân phối xác suất của X. Từ mẫu số liệu của X là cần kiểm định giả thuyết X tuân theo quy luật phân bố xác suất A nào đó. Ta có bài toán kiểm định giả thuyết:

$$\begin{cases} H_0\text{: } X \text{ có phân bố xác suất A} \\ H_0\text{: } X \text{ không có phân bố xác suất A} \end{cases}$$

7.1.1. Biến ngẫu nhiên rời rạc

Từ mẫu số liệu lập bảng phân bố tần số rời rạc:

X	x_1	x_2	...	x_m
n_i	n_1	n_2	...	n_m

$n_1 + n_2 + ... + n_m = n.$

Giả sử H_0 đúng, tính các xác suất $p_k = P(X = x_k/H_0$ đúng) và đặt $n_k^{'} = np_k$ ta được bảng phân bố tần số của X với điều kiện H_0 đúng:

X	x_1	x_2	...	x_m
$n_i^{'}$	$n_1^{'}$	$n_2^{'}$	...	$n_m^{'}$

$n_1^{'} + n_2^{'} + ... + n_m^{'} = n.$

Người ta đã chứng minh được nếu tất các các giá trị $n'_k \geq 5$ thì:

$$v = \sum_{k=1}^{m} \frac{(n_k - n_k^{'})^2}{n_k^{'}}$$

có phân bố khi bình phương $m - r - 1$ bậc tự do với r là số tham số cần ước lượng của quy luật cần kiểm định (chẳng hạn phân phối Bernoulli và Poisson có $r = 1$). Giả thuyết H_0 bị bác bỏ nếu v lớn một cách bất thường.

Vì vậy, với mức ý nghĩa α, miền bác bỏ H_0 là $W = [\chi^2_{m-r-1}(\alpha); +\infty)$.

p-giá trị $= P(\chi^2_{m-r-1} > v)$ với $v = \sum\limits_{k=1}^{m} \dfrac{(n_k - n_k^{'})^2}{n_k^{'}}$.

Ví dụ 7.1. Số cuộc gọi đến (X) ở một trạm điện thoại trong một phút được cho bởi bảng sau:

X	0	1	2	3	4	5	≥ 6
n_i	17	22	26	20	11	2	2

Với mức ý nghĩa $\alpha = 0,05$ có thể xem X có phân bố Poisson không?

Giải. Từ mẫu số liệu trên ta tính được $\overline{x} = 2$.

Ta cần kiểm định giả thuyết $H_0 : X$ có phân bố Poisson với đối thiết $H_1 : X$ không có phân bố Poisson.

Giả sử H_0 đúng, ta lấy $\lambda = \overline{x} = 2$. Gộp 2 lớp cuối cùng lại, ta có bảng sau:

x_k	n_k	$p_k = e^{-2}\dfrac{2^{x_k}}{x_k!}$	$n_k^{'} = np_k$	$\dfrac{(n_k - n_k^{'})^2}{n_k^{'}}$
0	17	0,1353	13,53	0,89
1	22	0,2707	27,07	0,95
2	26	0,2707	27,07	0,04
3	20	0,1804	18,04	0,21
4	11	0,0902	9,02	0,43

≥ 5	4	0,0527	5,27	0,31
	100	1		$\upsilon = 2,83$

Từ bảng trên ta thấy rằng $n'_k \geq 5$. Với mức ý nghĩa $\alpha = 5\%$, $m = 6$ và $r = 1$ ta có $\chi^2_4(0,05) = 9,49$ nên miền bác bỏ H_0 là $W = [9,49; +\infty)$.

Ta có $\upsilon \notin W$ nên chưa có cơ sở bác bỏ H_0. Vì vậy, với mức ý nghĩa 5% có thể kết luận: X có phân phối Poisson với $\lambda = 2$.

p-giá trị $= P(\chi^2_5 \geq 2,83) = 0,73 > \alpha$. ■

7.1.2. Biến ngẫu nhiên liên tục

Chia miền giá trị của X thành m khoảng rời nhau: $S_1, S_2, ..., S_m$. Từ mẫu số liệu lập bảng phân phối tần số liên tục:

X	S_1	S_2	...	S_m
n_i	n_1	n_2	...	n_m

Giả sử H_0 đúng, tính các xác suất $p_k = P(X \in S_k|H_0$ đúng) và đặt $n'_k = np_k$ ta được bảng phân phối của X với điều kiện H_0 đúng:

X	S_1	S_2	...	S_m
n'_i	n'_1	n'_2	...	n'_m

Nếu $n'_k \geq 5$ với mọi k thì

$$\upsilon = \sum_{k=1}^{m} \frac{(n_k - n'_k)^2}{n'_k}$$

có phân bố khi bình phương $m - r - 1$ bậc tự do nên với mức ý nghĩa α, miền bác bỏ H_0 là $W = [\chi^2_{m-r-1}(\alpha); +\infty)$.

Ví dụ 7.2. Tiến hành đo ngẫu nhiên chiều cao (X) của 100 cây bạch đàn trong khu rừng trồng bạch đàn của một lâm trường ta được kết quả sau:

X (mét)	n_k	X (mét)	n_k
[8,275;8,325)	1	[8,625-8,675)	17
[8,325;8,375)	2	[8,675;8,725)	12
[8,375;8,425)	4	[8,725;8,775)	9
[8,425;8,475)	5	[8,775;8,825)	7
[8,475;8,525)	8	[8,725;8,775)	6
[8,525;8,575)	10	[8,775;8,925)	0
[8,575;8,625)	18	[8,925;8,975)	1

Với mức ý nghĩa $\alpha = 0,05$ có thể xem chiều cao cây bạch đàn ở khu rừng trên có phân phối chuẩn không?

Giải. Tiến hành ghép lớp để đảm bảo các $n'_k \geq 5$ ta được:

X (mét)	n_k	X (mét)	n_k
(-∞;8,425)	7	[8,625;8,675)	17
[8,425-8,475)	5	[8,675;8,725)	12
[8,475-8,525)	8	[8,725;8,775)	9
[8,525-8,575)	10	[8,775;8,825)	7
[8,575-8,625)	18	[8,825;+∞)	7

Từ mẫu trên ta tính được $\overline{x} = 8,63$, $s = 0,128$.

Giả sử X có phân phối chuẩn $N(\mu;\sigma^2)$, khi đó ta xấp xỉ $\mu = 8,63$ và $\sigma = 0,128$ để tính các p_k theo công thức:

$$p_k = P(a_{k-1} \leq X < a_k) = \Phi(\frac{a_k - 8,63}{0,128}) - \Phi(\frac{a_{k-1} - 8,63}{0,128})$$

Từ đó ta có:

X (mét)	n_k	p_k	$n'_k = np_k$	$\dfrac{(n_k - n'_k)^2}{n'_k}$
(-∞;8,425)	7	0,0548	5,48	0,4216
[8,425-8,475)	5	0,0583	5,83	0,1182

[8,475-8,525)	8	0,0930	9,3	0,1817
[8,525-8,575)	10	0,1295	12,95	0,6720
[8,575-8,625)	18	0,1484	14,84	0,6729
[8,625;8,675)	17	0,1528	15,28	0,1936
[8,675;8,725)	12	0,1735	17,35	0,1365
[8,725;8,775)	9	0,1004	10,04	0,1077
[8,775;8,825)	7	0,0650	6,5	0,0385
[8,825;+∞)	7	0,0643	6,43	0,0505
	100			$v = 2,5923$

Từ bảng trên ta thấy rằng $n'_k \geq 5$. Với mức ý nghĩa $\alpha = 5\%$, $m = 10$ và $r = 2$ ta có $\chi^2_7(0,05) = 14,07$ nên miền bác bỏ H_0 là $W = [14,07; +\infty)$.

Ta có $v \notin W$. Do đó, chưa có cơ sở bác bỏ H_0. ■

7.2. Kiểm định tính độc lập

Trong mục này ta sẽ xét bài toán kiểm tra tính độc lập của 2 dấu hiệu A và B trong một tổng thể. Ta chia dấu hiệu A làm m mức độ: $A_1, A_2, ..., A_m$ và dấu hiệu B ra n mức độ: B_1, B_2, ..., B_n. Xét một mẫu ngẫu nhiên có k_{ij} cá thể mang dấu hiệu A ở mức A_i và dấu hiệu B ở mức B_j. Khi đó ta có bảng sau:

A \ B	B_1	B_2	...	B_n	Tổng
A_1	k_{11}	k_{12}	...	k_{1n}	k_{1*}
A_2	k_{21}	k_{22}	...	k_{2n}	k_{2*}
...	...	...	...	...	...
A_m	k_{m1}	k_{m2}	...	k_{mn}	k_{m*}
Tổng	k_{*1}	k_{*2}	...	k_{*n}	N

Trong đó $N = \sum_{i=1}^{m}\sum_{j=1}^{n}$, $k_{i*} = \sum_{j=1}^{n} k_{ij}$, $k_{*j} = \sum_{i=1}^{m} k_{ij}$.

Kí hiệu p_{ij} là xác suất để một cá thể chọn ngẫu nhiên mang các dấu hiệu A_i và B_j, p_{i*} là xác suất để một cá thể chọn ngẫu nhiên mang dấu hiệu A_i, p_{*j} là xác suất để một cá thể chọn ngẫu nhiên mang dấu hiệu B_j.

Xét bài toán kiểm định giả thuyết $H_0 : A$ và B độc lập, $H_1 : A$ và B không độc lập (phụ thuộc).

Giả sử H_0 đúng, khi đó $P(AB) = P(A)P(B)$ nên ta có $p_{ij} = p_{i*}p_{*j}$.

Các xác suất p_{i*} và p_{*j} được ước lượng bởi:

$$p_{i*} \approx \frac{k_{i*}}{N},\ p_{*j} \approx \frac{k_{*j}}{N}$$

Do đó:

$$p_{ij} \approx \frac{k_{i*}k_{*j}}{N^2}$$

Số cá thể có đồng thời 2 dấu hiệu A_i và B_j khi chọn ngẫu nhiên N cá thể là $\hat{k}_{ij} = N.p_{ij} \approx \dfrac{k_{i*}k_{*j}}{N}$.

Các số $\hat{k}_{ij}$ được gọi là *tần số lý thuyết* còn các số k_{ij} được gọi là *tần số thực nghiệm*. Khoảng cách các tần số lý thuyết và tần số thực nghiệm được đo bởi đại lượng:

$$\upsilon = \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{(\hat{k}_{ij} - k_{ij})^2}{\hat{k}_{ij}}$$

Người ta đã chứng minh được rằng nếu N lớn và các tần số $\hat{k}_{ij} \geq 5$ thì υ có phân bố xấp xỉ phân bố $\chi^2_{(m-1)(n-1)}$. Giả thuyết H_0 bị bác bỏ nếu υ lớn một cách bất thường.

Vì vậy, với mức ý nghĩa α, miền bác bỏ H_0 là:

$$W = [\chi^2_{(m-1)(n-1)}(\alpha); +\infty)$$

p-giá trị $= P(\chi^2_{(m-1)(n-1)} \geq T)$.

Chú ý 7.1. Trong trường hợp có tần số lý thuyết bé hơn 5 thì ta tiến hành ghép cột hoặc ghép hàng để có tần số lý thuyết không bé hơn 5.

Ví dụ 7.3. Ở cây ngọc trâm lá có 2 dạng "phẳng" hoặc "nhăn", hoa có 2 dạng "bình thường" hoặc "hoàng hậu". Quan sát một mẫu gồm 560 cây ngọc trâm thu được kết quả:

Lá \ Hoa	Bình thường	Hoàng hậu	Tổng
Phẳng	328	122	450
Nhăn	77	33	110
Tổng	405	155	560

Với mức ý nghĩa 5% có thể cho rằng hai đặc tính của hoa và lá trên cây ngọc trâm là độc lập không?

Giải. Bài toán kiểm định giả thuyết:

H_0: hai đặc tính của hoa và lá trên cây ngọc trâm độc lập.

H_1: hai đặc tính của hoa và lá trên cây ngọc trâm không độc lập.

Miền bác bỏ H_0 là $W = [3,841; +\infty)$.

Bảng tần số lý thuyết:

Lá \ Hoa	Bình thường	Hoàng hậu
Phẳng	325,44	124,55
Nhăn	79,55	30,44

Ta thấy rằng tất cả các tần số lý thuyết đều lớn hơn 5.

$$
\begin{aligned}
v &= \frac{(325,44-328)^2}{235,44} + \frac{(124,55-122)^2}{124,55} \\
&\quad + \frac{(79,55-77)^2}{79,55} + \frac{(30,44-33)^2}{30,44} \\
&= 0,369 \notin W
\end{aligned}
$$

Do đó, chưa có cơ sở bác bỏ H_0. ■

Ví dụ 7.4. Một con ốc sên rừng có thể có màu vỏ là vàng hoặc hồng; số vạch trên vỏ có thể là 1, 2, 3, 4 hoặc 5. Một mẫu 169 con ốc sên có số liệu sau:

Màu vỏ / Số vạch	Vàng	Hồng	Tổng
A_1 (0 vạch)	35	14	49
A_2 (1 hoặc 2 vạch)	19	14	33
A_3 (3 hoặc 4 vạch)	36	16	52
A_4 (5 vạch)	25	10	35
Tổng	115	54	169

Với mức ý nghĩa 5% có thể cho rằng màu vỏ và số vạch trên vỏ của ốc sên độc lập không?

Giải. Bài toán kiểm định giả thuyết:
H_0: màu vỏ và số vạch trên vỏ của ốc sên độc lập.
H_1: màu vỏ và số vạch trên vỏ của ốc sên không độc lập.
Miền bác bỏ H_0 là $W = [7,81; +\infty)$.
Bảng tần số lý thuyết:

Màu vỏ / Số vạch	Vàng	Hồng
A_1 (0 vạch)	33,34	15,66
A_2 (1 hoặc 2 vạch)	22,46	10,54
A_3 (3 hoặc 4 vạch)	35,38	16,62
A_4 (5 vạch)	23,82	11,18

Như vậy các tần số lý thuyết đều lớn hơn 5.

$$\upsilon = \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{(\hat{k}_{ij} - k_{ij})^2}{\hat{k}_{ij}} = 2,139 \notin W$$

Do đó, chưa có cơ sở bác bỏ H_0. ■

7.3. Kiểm định sự phù hợp giữa lý thuyết và thực nghiệm

Trong khoa học thường sử dụng kiểm định khi bình phương để kiểm định sự phù hợp giữa lý thuyết và số liệu thực nghiệm. Ví dụ tung 200 lần một đồng xu. Nếu đồng xu cân đối và đồng chất thì theo lý thuyết ta sẽ có 100 lần xuất hiện mặt sấp và 100 lần xuất hiện mặt ngữa. Nếu kết quả thực hiện thu được 92 lần xuất hiện mặt sấp và 108 lần xuất hiện mặt ngữa thì liệu ta có thể chấp nhận giả thuyết đồng xu cân đối đồng chất không?

Một cách tổng quát ta xét một dấu hiện X có k trường hợp xảy ra là A_1, A_2,..., A_k. Ta lập bảng sau:

X	Tần số thực nghiệm (n_k)	Tần số lý thuyết $(n_k^{'})$	$(n_k - n_k^{'})^2/n_k^{'}$
A_1	n_1	$n_1^{'}$	$(n_1 - n_1^{'})^2/n_1^{'}$
A_2	n_2	$n_2^{'}$	$(n_2 - n_2^{'})^2/n_2^{'}$
...	...	...	...
A_k	n_k	$n_k^{'}$	$(n_k - n_k^{'})^2/n_k^{'}$
			$\upsilon = \sum_{i=1}^{k} \frac{(n_i - n_i^{'})^2}{n_i^{'}}$

Xét bài toán kiểm định giả thuyết:

$$\begin{cases} H_0\text{: Số liệu thu được phù hợp với lý thuyết} \\ H_1\text{: Số liệu thu được không phù hợp với lý thuyết} \end{cases}$$

Người ta đã chứng minh được rằng nếu H_0 đúng và $n_k' \geq 5$ với mọi k thì:

$$\upsilon = \sum_{i=1}^{k} \frac{(n_i - n_i^{'})^2}{n_i^{'}}$$

có phân bố khi bình phương $k-1$ bậc tự do.

Do đó, miền bác bỏ H_0 là $W_\alpha = [\chi^2_{k-1}(\alpha); +\infty)$.

Ví dụ 7.5. Một nhà di truyền học tiến hành phép lai giữa hai cá thể ruồi giấm F1 và thu được 176 cá thể F2 gồm 130 có kiểu hình hoang dại và 46 có kiểu hình đột biến. Với mức ý nghĩa 5% có thể cho rằng kết quả thu được có phù hợp với tỉ lệ 3 trội : 1 lặn theo định luật phân ly của Mendel không?

Giải. Bài toán kiểm định giả thuyết:

H_0: Kết quả thu được tuân theo Định luật phân ly của Mendel.

H_1: Kết quả thu được không tuân theo Định luật phân ly của Mendel.

Miền bác bỏ H_0: $W = [3,841; +\infty)$.

X	n_k	p_k	$n_k^{'} = np_k$	$\frac{(n_k - n_k^{'})^2}{n_k^{'}}$
Hoang dại	130	3/4	132	0,03
Đột biến	46	1/4	44	0,091
	176	1		$\upsilon = 0,121$

$\upsilon = 0,121 \notin W$. Do đó, chấp nhận H_0. ■

Ví dụ 7.6. Theo dõi sự di truyền của hai tính trạng chiều cao và dạng lá ở cà chua, người ta thực hiện một phép lai và thu được kết quả ở F2 như sau: Thân cao, lá chẻ 926; Thân cao, lá nguyên 288; Thân thấp, lá chẻ 293; Thân thấp, lá nguyên 104. Với mức ý nghĩa 5% có thể cho rằng kết quả thu được có phù hợp với tỉ lệ phân ly 9 : 3 : 3 : 1 theo định luật phân ly độc lập của Mendel không?

Giải. Bài toán kiểm định giả thuyết:

H_0: Kết quả thu được tuân theo Định luật phân ly độc lập của Mendel.

H_1: Kết quả thu được không tuân theo Định luật phân ly độc lập của Mendel.

Miền bác bỏ H_0: $W = [7,815; +\infty)$.

X	n_k	p_k	$n_k^{'} = np_k$	$\dfrac{(n_k - n_k^{'})^2}{n_k^{'}}$
Thân cao, lá chẻ	926	9/16	906,19	0,433
Thân cao, lá nguyên	288	3/16	302,06	0,654
Thân thấp, lá chẻ	293	3/16	302,06	0,272
Thân thấp, lá nguyên	104	1/16	100,69	0,109
	1611	1		$\upsilon = 1,468$

$\upsilon = 1.468 \notin W$. Do đó, chấp nhận H_0. ■

BÀI TẬP CHƯƠNG 7

▷ **7.1.** Một nhà sinh thái thực vật nghiên cứu một mẫu gồm 100 cây của một loài quí hiếm trên một vùng có diện tích 400 dặm. Ở mỗi cây ông ghi nhận đặc điểm của lá (lá có lông tơ hoặc không có) và đặc điểm của đất nơi cây mọc (có khoáng chất serpentine hoặc không có). Kết quả như sau:

Đất \ Lá cây	Có lông tơ	Không có lông tơ
Có serpentine	12	22
Không có serpentine	16	50

Với mức ý nghĩa 5% hãy kiểm tra xem đặc điểm của lá có phụ thuộc vào đặc điểm của loại đất nơi cây mọc hay không.

▷ **7.2.** Với mức ý nghĩa 5% hãy dùng kết quả ghi nhận trong bảng dưới đây để kiểm tra xem màu sắc của loài bọ Cicindela fulgida có thay đổi tùy theo mùa không?

Mùa \ Màu sắc bọ	Đỏ sáng	Không đỏ sáng
Đầu mùa xuân	29	11
Cuối mùa xuân	273	191

Đầu mùa hè	8	31
Cuối mùa hè	64	64

▷ **7.3.** Một nhà xã hội học muốn tìm hiểu mối quan hệ giữa các dạng tội phạm (A) đối với tuổi (B) của phạm nhân. Chọn ngẫu nhiên 100 phạm nhân trong hồ sơ của tòa án, ông ta thu được số liệu sau:

A \ B	Dưới 25	Từ 25 đến 49	Trên 49
Hình sự	15	30	10
Không hình sự	5	30	10

Với mức ý nghĩa 5% với số liệu trên có thể cho rằng tuổi và dạng tội phạm độc lập nhau không?

▷ **7.4.** Một cuộc thăm dò được nghiên cứu mối quan hệ giữa nghề nghiệp của một người với quan niệm của người đó về tiêu chuẩn đạo đức và tính trung thực. Kết quả khảo sát một mẫu ngẫu nhiên 380 người cho số liệu sau:

Nghề nghiệp \ Quan niệm	Cao	Trung bình	Thấp
Bác sĩ	53	35	10
Luật sư	24	43	27
Nhà kinh doanh	18	55	20
Nhà chính trị	14	43	38

Với mức ý nghĩa 5% hãy kiểm định xem có sự phụ thuộc giữa nghề nghiệp và quan niệm về tiêu chuẩn đạo đức không?

▷ **7.5.** Một nhà nghiên cứu muốn kiểm chứng giả thuyết cho rằng các dị tật bẩm sinh của trẻ có liên hệ đến tuổi của mẹ sinh ra chúng. Ông chọn 309 trường hợp dị tật bẩm sinh và phân loại theo 4 loại A, B, C và D, và mỗi loại dị tật được liên hệ với một trong ba lớp tuổi của người mẹ. Kết quả phân loại này được trình bày trong bảng dưới đây:

Lớp tuổi của người mẹ	Loại A	Loại B	Loại C	Loại D
Dưới 25 tuổi	51	46	25	15
Từ 25 đến dưới 40	33	17	49	20
Trên 40	4	11	35	3

Với mức ý nghĩa 5% hãy kiểm định xem có sự phụ thuộc giữa các dị tật bẩm sinh và tuổi của người mẹ không?

▹ **7.6.** Một người nghiên cứu soạn một thang thái độ, trong đó có một câu phát biểu yêu cầu sinh viên cho biết ý kiến về câu ấy theo 5 mức: 1. rất không đồng ý; 2. không đồng ý; 3. không có ý kiến; 4. đồng ý; 5. rất đồng ý. Người nghiên cứu chọn một mẫu ngẫu nhiên 400 sinh viên để khảo sát. Kết quả phản ứng của sinh viên về câu phát biểu ấy như sau:

Loại đáp ứng	1	2	3	4	5
Tần số SV trả lời	89	113	98	55	45

Có sự khác biệt hay không giữa các sinh viên về tần số đáp ứng với câu phát biểu trên thang đo thái độ ấy? Thực hiện kiểm định với mức ý nghĩa 5%.

▹ **7.7.** Một người nghiên cứu chọn một mẫu ngẫu nhiên gồm 956 sinh viên trong thành phố và hỏi họ thích theo học chương trình nào trong 3 chương trình P1, P2 và P3 ở đại học. Kết quả cho biết số sinh viên ưa thích theo học mỗi chương trình như sau:

P1	P2	P3
376	383	197

Có sự khác biệt hay không giữa các loại chương trình sinh viên ưa thích? Thực hiện kiểm định với mức ý nghĩa 5%.

▷ **7.8.** Trong mẫu 956 sinh viên trúng tuyển kì thi đại học, số sinh viên thuộc các khu vực (KV) được phân phối như sau:

KV1	KV2	KV3
225	419	312

Các sinh viên trúng tuyển có được phân bố đồng đều theo khu vực hay không? Thực hiện kiểm định với mức ý nghĩa 5%.

Chương 8
PHÂN TÍCH PHƯƠNG SAI

8.1. Phân tích phương sai một nhân tố

Trước hết ta xét ví dụ sau.

Ví dụ 8.1. Một thí nghiệm được tiến hành để so sánh trọng lượng (gam) của các con gà được nuôi bằng 4 khẩu phần ăn khác nhau. Hai mươi con gà có trọng lượng ban đầu như nhau được bố trí ngẫu nhiên vào từng nhóm, mỗi nhóm gồm 5 con. Kết quả thí nghiệm được ghi nhận trong bảng sau.

Khẩu phần 1	Khẩu phần 2	Khẩu phần 3	Khẩu phần 4
105	61	42	120
88	100	100	105
65	50	85	119
40	89	90	90
90	70	69	65

Câu hỏi đặt ra là khẩu phần ăn có ảnh hưởng tới trọng lượng gà không?

Kí hiệu X_{ij} là biến ngẫu nhiên trọng lượng của con gà thứ i ở khẩu phần j; μ là trọng lượng trung bình chung của gà; α_j là tác dụng của khẩu phần ăn thứ j, ϵ_{ij} là sai số ngẫu nhiên tác động lên con gà thứ i ở khẩu phần ăn thứ j.

Ta có mô hình toán học của ví dụ trên là:

$$X_{ij} = \mu + \alpha_j + \epsilon_{ij}$$

Bài toán kiểm định giả thuyết:

$$\begin{cases} H_0\text{: tác dụng của các khẩu phần ăn là như nhau} \\ H_1\text{: có ít nhất hai khẩu phần ăn có tác dụng khác nhau} \end{cases}$$

Hay tương đương với bài toán kiểm định giả thuyết

$$\begin{cases} H_0: \quad \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0 \\ H_1: \quad \alpha_1^2 + \alpha_2^2 + \alpha_3 + \alpha_k^2 \neq 0 \end{cases}$$

Mô hình tổng quát của tích phương sai một nhân tố như sau: Giả sử ta cần quan tâm tác động của nhân tố A lên biến số ngẫu nhiên X ở k mức A_1, A_2,...,A_k. Kí hiệu X_{ij} là kết quả của tác động mức A_j lên phần tử thứ i.

A_1	A_2	...	A_k
X_{11}	X_{12}	...	X_{1k}
X_{21}	X_{22}	...	X_{2k}
...	...	...	...
X_{m1}	X_{m2}	...	X_{mk}

Mô hình phân tích phương sai một nhân tố:

$$X_{ij} = \mu + \alpha_j + \epsilon_{ij} \quad i = 1, 2, .., m; j = 1, 2, ..., k$$

Trong đó ϵ_{ij} là các biến ngẫu nhiên độc lập, có cùng phân phối chuẩn $N(0; \sigma^2)$; α_j là tác dụng của mức nhân tố A_j lên các biến ngẫu nhiên X_{ij}, $i = 1, 2, ...m$.

Bài toán kiểm định giả thuyết:

$$\begin{cases} H_0: \alpha_1 = \alpha_2 = ... = \alpha_k = 0 \\ H_1: \alpha_1^2 + \alpha_2^2 + ... + \alpha_k^2 \neq 0 \end{cases} \qquad (*)$$

Từ giả thiết của mô hình ta có X_{ij} là các biến ngẫu nhiên có phân bố chuẩn với:

$$\mu_j = E(X_{ij}) = \mu + \alpha_j \text{ và } Var(X_{ij}) = \sigma^2$$

Bài toán kiểm định giả thuyết (*) cũng có thể phát biểu dưới dạng sau:

$$\begin{cases} H_0\text{: } \mu_1 = \mu_2 = ... = \mu_k = 0 \\ H_1\text{: có ít nhất 2 giá trị trung bình khác nhau} \end{cases}$$

Gọi $\{X_{ij} : 1 \leq i \leq m, 1 \leq j \leq k\}$ là mẫu ngẫu nhiên thu được từ thí nghiệm. Đặt:

$$T_j = \sum_{i=1}^{m} X_{ij}, \quad T = \sum_{i=1}^{m}\sum_{j=1}^{k} X_{ij}, \quad Q = \sum_{i=1}^{m}\sum_{j=1}^{k} X_{ij}^2$$

Ta đưa ra một số kí hiệu sau:

- Trung bình mẫu thứ j $(j = 1, 2, ..., k)$:

$$\overline{X}_j = \frac{1}{m}\sum_{i=1}^{m} X_{ij}$$

- Trung bình chung:

$$\overline{X} = \frac{1}{mk}\sum_{i=1}^{m}\sum_{j=1}^{k} X_{ij}$$

- Ước lượng giá trị X_{ij} từ mô hình:

$$\hat{X}_{ij} = \overline{X}_j$$

- Phần dư:

$$e_{ij} = X_{ij} - \overline{X_j}, i = 1, ..., m; j = 1, ..., k$$

- Tổng bình phương chung:

$$SST = \sum_{i=1}^{m}\sum_{j=1}^{k}(X_{ij} - \overline{X})^2 = Q - \frac{T^2}{mk}$$

- Tổng bình phương do nhân tố:

$$SSF = \sum_{i=1}^{m}\sum_{j=1}^{k}(\overline{X}_j - \overline{X})^2 = \frac{1}{m}\sum_{j=1}^{k} T_j^2 - \frac{T^2}{mk}$$

- Tổng bình phương do sai số:

$$SSE = \sum_{i=1}^{m}\sum_{j=1}^{k}(X_{ij} - \overline{X}_j)^2$$

Từ các công thức trên ta có $SST = SSF + SSE$.

- Trung bình bình phương của nhân tố:

$$MSF = \frac{SSF}{k-1}$$

- Trung bình bình phương của sai số:

$$MSE = \frac{SSE}{mk-k}$$

- Tỉ số F:

$$F = \frac{MSF}{MSE}$$

Các kết quả nói trên được trình bày trong bảng sau đây gọi là bảng ANOVA.

Nguồn	Bậc tự do	Tổng bình phương	Trung bình bình phương	Tỉ số F
Nhân tố	k-1	SSF	MSF	$\frac{MSF}{MSE}$
Sai số	mk-k	SSE	MSE	
Tổng	n-1	SST		

Người ta chứng minh được rằng nếu H_0 đúng thì:

$$F = \frac{MSF}{MSE}$$

có phân bố F với hai tham số $k-1$ và $n-k$. Vì vậy, miền bác bỏ H_0 với mức ý nghĩa α là: $W = [f_{k-1,mk-k}(\alpha); +\infty)$.

p-giá trị $= P(F_{k-1,mk-k} \geq F))$.

Ví dụ 8.2. Xét số liệu ở Ví dụ 8.1. Với mức ý nghĩa 5% có thể cho rằng khẩu phần ăn có ảnh hưởng tới trọng lượng gà không?

Giải. Bài toán kiểm định giả thuyết là:

$$\begin{cases} H_0\text{: Tác dụng của các khẩu phần ăn là như nhau} \\ H_1\text{: Có ít nhất hai khẩu phần ăn có tác dụng khác nhau} \end{cases}$$

Từ số liệu trên ta lập được bảng ANOVA:

Nguồn	Bậc tự do	Tổng bình phương	Trung bình bình phương	Tỉ số F
Nhân tố	3	2.115,75	705,25	1,34
Sai số	16	8.408,8	525,55	
Tổng	19	10.524,55		

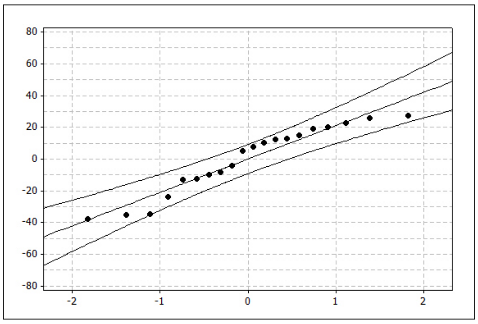
Miền bác bỏ H_0: $W = [3,239; \infty)$.

$F = 1,34 \notin W$. Do đó, chưa có cơ sở bác bỏ H_0.

Kiểm tra giả thiết của mô hình.

Để kiểm gia giả thiết của môn hình: sai số ngẫu nhiên có phân bố chuẩn $N(0; \sigma^2)$ ta vẽ biểu đồ xác suất chuẩn của phần dư. Ta có số liệu của phần dư như sau:

e_{i1}	27,4	10,4	-12,6	-37,6	12,4
e_{i2}	-13	26	-24	15	-4
e_{i3}	-35,2	22,8	7,8	12,8	-8,2
e_{i4}	20,2	5,2	19,2	-9,8	-34,8



Hình 8.1

Từ biểu đồ xác suất chuẩn (xem Hình 8.1) của phần dư nằm xấp xỉ trên đường thẳng đi qua gốc tọa độ ta có thể xem sai số ngẫu nhiên có phân phối chuẩn $N(0;\sigma^2)$. ■

8.2. Phân tích phương sai hai nhân tố

8.2.1. Phân tích phương sai hai nhân tố không lặp lại

Trên thực tế một biến lượng chịu tác động không chỉ 1 nhân tố mà có thể 2 hoặc nhiều nhân tố. Chẳng hạn năng suất cây trồng chịu ảnh hưởng của nhân tố giống và nhân tố đất.

Giả sử chúng ta quan tâm tới 2 nhân tố A và B. Nhân tố A được xem xét ở các mức $A_1, A_2, ..., A_m$ và nhân tố B được xem xét ở các mức $B_1, B_2, ..., B_n$.

Đặt X_{ij} là biến ngẫu nhiên đo lường hiệu quả của việc tác động của nhân tố A ở mức i và nhân tố B ở mức i lên cá thể, X_{ij} là một kết quả thu được từ biến ngẫu nhiên X_{ij} khi thực hiện thí nghiệm.

Ta có mô hình toán học cho phân tích phương sai hai nhân tố không lặp như sau.

Mô hình phân tích phương sai hai nhân tố không lặp:

$$X_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij} \quad i = 1, 2, .., m; j = 1, 2, ..., n.$$

Trong đó ϵ_{ij} là các biến ngẫu nhiên độc lập, có cùng phân phối chuẩn $N(0;\sigma^2)$.

1. Bài toán kiểm định giả thuyết 1:

$$\begin{cases} H_0^A\text{: } \alpha_1 = \alpha_2 = ... = \alpha_k = 0 \\ \text{(Các mức } A_1, ..., A_m \text{ có hiệu quả như nhau)} \\ H_1^A\text{: } \alpha_1^2 + \alpha_2^2 + ... + \alpha_k^2 \neq 0 \\ \text{(Có ít nhất 2 mức } A_i \text{ và } A_j \text{ có hiệu quả khác nhau)} \end{cases}$$

2. Bài toán kiểm định giả thuyết 2:

$$\begin{cases} H_0^B\text{: } \beta_1 = \beta_2 = ... = \beta_k = 0 \\ \text{(Các mức } B_1, ..., B_n \text{ có hiệu quả khác nhau)} \\ H_1^B\text{: } \beta_1^2 + \beta_2^2 + ... + \beta_k^2 \neq 0 \\ \text{(Có ít nhất 2 mức } B_i \text{ và } B_j \text{ có hiệu quả khác nhau)} \end{cases}$$

Từ giả thiết của mô hình ta có X_{ij} là các biến ngẫu nhiên có phân phối chuẩn với:

$$E(X_{ij}) = \mu_{ij} = \mu + \alpha_i + \beta_j \text{ và } V(X_{ij}) = \sigma^2$$

Giả sử $\{X_{ij}; 1 \leq i \leq m, 1 \leq j \leq n\}$ là mẫu ngẫu nhiên thu được khi thực hiện thí nghiệm. Khi đó ta lập được bảng sau:

A \ B	B_1	B_2	...	B_n	$T_{i*} = \sum\limits_{j=1}^{n} X_{ij}$
A_1	X_{11}	X_{12}	...	X_{1n}	T_{1*}
A_2	X_{21}	X_{22}	...	X_{2n}	T_{2*}
...	...	...	...	...	...
A_m	X_{m1}	X_{m2}	...	X_{mn}	T_{m*}
$T_{*j} = \sum\limits_{i=1}^{m} X_{ij}$	T_{*1}	T_{*2}	...	T_{*n}	$T = \sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n} X_{ij}$
$\sum\limits_{i=1}^{m} X_{ij}^2$	$\sum\limits_{i=1}^{m} X_{i1}^2$	$\sum\limits_{i=1}^{m} X_{i2}^2$	...	$\sum\limits_{i=1}^{m} X_{in}^2$	$Q = \sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n} X_{ij}^2$

- Trung bình mẫu:

$$\overline{X} = \frac{1}{mn}\sum_{i=1}^{m}\sum_{i=1}^{n} X_{ij}$$

- Trung bình mức A_i:

$$\overline{X}_{i*} = \frac{1}{n}\sum_{j=1} X_{ij}$$

- Trung bình mức B_j:

$$\overline{X}_{*j} = \frac{1}{m}\sum_{i=1} X_{ij}$$

- Ước lượng giá trị X_{ij} từ mô hình:

$$\hat{X}_{ij} = \overline{X}_{i*} + \overline{X}_{*j} - \overline{X}$$

- Phần dư:

$$e_{ij} = X_{ij} - \overline{X}_{i*} - \overline{X}_{*j} + \overline{X}$$

- Tổng bình phương chung:

$$SST = \sum_{i=1}^{m}\sum_{j=1}^{n}(X_{ij} - \overline{X})^2 = Q - \frac{T^2}{mn}$$

- Tổng bình phương cho nhân tố A:

$$SSA = \frac{\sum_{i=1}^{m} T_{i*}^2}{n} - \frac{T^2}{mn}$$

- Tổng bình phương cho nhân tố A:

$$SSB = \frac{\sum_{j=1}^{n} T_{*j}^2}{m} - \frac{T^2}{mn}$$

- Tổng bình phương do sai số:

$$SSE = SST - SSA - SSB$$

- Trung bình bình phương nhân tố A:

$$MSA = \frac{SSA}{m-1}$$

$m-1$ được gọi là bậc tự do của A.
- Trung bình bình phương nhân tố B

$$MSB = \frac{SSB}{n-1}$$

$n-1$ được gọi là bậc tự do của B.
- Trung bình bình phương sai số

$$MSE = \frac{SSE}{(m-1)(n-1)}$$

$(m-1)(n-1)$ được gọi là bậc tự do của sai số.
- Tỉ số F cho nhân tố A:

$$F_A = \frac{MSA}{MSE}$$

- Tỉ số F cho nhân tố B:

$$F_A = \frac{MSB}{MSE}$$

Các kết quả nói trên được trình bày trong bảng sau đây gọi là bảng ANOVA 2 nhân tố:

Nguồn	Bậc tự do	SS	MS	Tỉ số F
Nhân tố A	$m-1$	SSA	MSA	F_A
Nhân tố B	$n-1$	SSB	MSB	F_B
Sai số	$(m-1)(n-1)$	SSE	MSE	
Tổng	$mn-1$	SST		

Ta có các bài toán kiểm định giả thuyết sau.
1. Bài toán kiểm định giả thuyết 1:

$$\begin{cases} H_0^A\text{: Các mức } A_1, A_2, ..., A_m \text{ có hiệu quả như nhau} \\ H_1^A\text{: Có ít nhất 2 mức } A_i \text{ và } A_j \text{ có hiệu quả khác nhau} \end{cases}$$

Miền bác bỏ H_0^A là $W = [f_{m-1,(m-1)(n-1)}(\alpha); +\infty)$.
p-giá trị $= P(F_{m-1,(m-1)(n-1)} \geq F_A)$.
2. Bài toán kiểm định giả thuyết 2:

$$\begin{cases} H_0^B\text{: Các mức } B_1, B_2, ..., B_n \text{ có hiệu quả như nhau} \\ H_1^B\text{: Có ít nhất 2 mức } B_i \text{ và } B_j \text{ có hiệu quả khác nhau} \end{cases}$$

Miền bác bỏ H_0^B là $W = [f_{n-1,(m-1)(n-1)}(\alpha); +\infty)$.
p-giá trị $= P(F_{n-1,(m-1)(n-1)} \geq F_B)$.
Trong đó, $F_{x,y}$ là phân bố F, $f_{x,y}(\alpha)$ là giá trị tới hạn mức α của phân bố F.

Ví dụ 8.3. Chiết suất chất X từ 1 loại dược liệu bằng 3 phương pháp (PP) và 5 loại dung môi (DM), ta có kết quả:

PP chiết xuất / Dung môi	B_1	B_2	B_3
A_1	120	60	60
A_2	120	70	50
A_3	130	60	50
A_4	150	70	60
A_5	110	75	54

Với mức ý nghĩa $\alpha = 0,05$ hãy xét tác dụng của phương pháp chiết suất và dung môi đến kết quả chiết suất chất X.

Giải. Các bài toán kiểm định giả thuyết:

$$\begin{cases} H_0^A\text{: Tác dụng của 5 dung môi bằng nhau} \\ H_1^A\text{: Có ít nhất 2 dung môi có tác dụng khác nhau} \end{cases}$$

$$\begin{cases} H_0^B\text{: Tác dụng của 3 phương pháp bằng nhau} \\ H_1^B\text{: Có ít nhất 2 phương pháp có tác dụng khác nhau} \end{cases}$$

PP / DM	B_1	B_2	B_3	T_{i*}
A_1	120	60	60	240
A_2	120	70	50	240
A_3	130	60	50	240
A_4	150	70	60	280
A_5	110	75	54	239
T_{*j}	630	335	274	T=1239
$\sum_i X_{ij}^2$	80300	22625	15116	$Q = 118041$

Ta có bảng ANOVA sau:

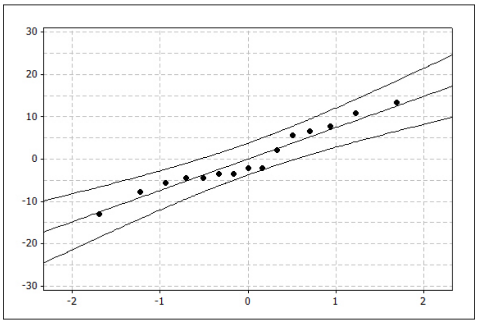
Nguồn	Bậc tự do	SS	MS	F
Nhân tố A	4	432,2667	108,0667	$F_A = 1,1249$
Nhân tố B	2	14498,8	7249,4	$F_B = 75,4622$
Sai số	8	768,5333	96,0667	
Tổng	14	15699,6		

Với mức ý nghĩa $\alpha = 0,05$, miền bác bỏ H_0^A là $W = [3,838; +\infty)$, miền bác bỏ H_0^B là $W = [4,459; +\infty)$. Vì vậy, với mức ý nghĩa $\alpha = 0,05$ có thể kết luận: dung môi không ảnh hưởng đến kết quả chiết suất và phương pháp ảnh hưởng đến kết quả chiết suất.

Kiểm tra giả thiết của mô hình.

Để kiểm gia giả thiết của môn hình: sai số ngẫu nhiên có phân bố chuẩn $N(0;\sigma^2)$ ta vẽ biểu đồ xác suất chuẩn của phần dư. Ta có số liệu của phần dư như sau:

e_{i1}	-3,4	-3,4	6,6	13,27	-13,07
e_{i2}	-4,4	5,6	-4,4	-7,73	10,93
e_{i3}	7,8	-2,2	-2,2	-5,53	2,13



Hình 8.2

Biểu đồ xác suất chuẩn (xem Hình 8.2) của phần dư nằm xấp xỉ trên đường thẳng đi qua gốc tọa độ nên giả thiết sai số ngẫu nhiên có phân bố chuẩn $N(0;\sigma^2)$ xem như được thỏa mãn. ■

8.2.2. Phân tích phương sai hai nhân tố có lặp

Tương tự như bài toán phân tích phương sai 2 nhân tố không lặp, chỉ khác mỗi mức $(A_i; B_j)$ đều có sự lặp lại r lần thí nghiệm và ta cần khảo sát thêm sự tương tác giữa 2 nhân tố A và B.

Gọi X_{ijk} là biến ngẫu nhiên đo lường hiệu quả của việc tác động của mức A_i và B_j lên cá thể. Ta có mô hình toán học cho phân tích phương sai hai nhân tố có lặp như sau:

> Mô hình phân tích phương sai hai nhân tố có lặp:
> $X_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk} \quad i = 1, 2, .., m; j = 1, 2, ..., n; 1 \leq k \leq r$
> Trong đó ϵ_{ijk} là các biến ngẫu nhiên độc lập, có cùng phân phối chuẩn $N(0; \sigma^2)$.

1. Bài toán kiểm định giả thuyết 1:

$$\begin{cases} H_0^A\text{: } \alpha_1 = \alpha_2 = ... = \alpha_k = 0 \\ \text{(Các mức nhân tố } A \text{ có hiệu quả trung bình như nhau)} \\ H_1^A\text{: } \alpha_1^2 + \alpha_2^2 + ... + \alpha_k^2 \neq 0 \\ \text{(Có ít nhất 2 mức nhân tố } A \text{ có hiệu quả khác nhau)} \end{cases}$$

2. Bài toán kiểm định giả thuyết 2:

$$\begin{cases} H_0^B\text{: } \beta_1 = \beta_2 = ... = \beta_k = 0 \\ \text{(Các mức nhân tố } B \text{ có hiệu quả như nhau)} \\ H_1^B\text{: } \beta_1^2 + \beta_2^2 + ... + \beta_k^2 \neq 0 \\ \text{(Có ít nhất 2 mức nhân tố } B \text{ có hiệu quả khác nhau)} \end{cases}$$

3. Bài toán kiểm định giả thuyết 3:

$$\begin{cases} H_0^{AB}\text{: } \gamma_{ij} = 0 \text{với mọi} i, j \\ \text{(Không có sự tương tác giữa hai nhân tố } A \text{ và } B\text{)} \\ H_1^{AB}\text{: Có ít nhất một} \gamma_{ij} \neq 0 \\ \text{(Có sự tương tác giữa hai nhân tố } A \text{ và } B\text{)} \end{cases}$$

Giả sử $\{X_{ijk} : 1 \leq k \leq r\}$ là mẫu kích thước r của biến ngẫu nhiên X_{ij} (được gọi là mẫu (i,j)). Ta đưa ra một số kí hiệu sau:

$$T_{ij} = \sum_{k=1}^{r} X_{ijk}, \quad T_{*j} = \sum_{i=1}^{n} T_{ij}; \quad T_{i*} = \sum_{j=1}^{m} T_{ij},$$

$$T = \sum_{j=1}^{n} T_{*j} = \sum_{i=1}^{m} T_{i*}, \quad Q = \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{r} X_{ijk}^2$$

- Trung bình mẫu (i, j):

$$\overline{X}_{ij} = \sum_{k=1}^{r} X_{ijk} = \frac{T_{ij}}{r}$$

- Trung bình của mức A_i:

$$\overline{X}_{i*} = \frac{\sum_{j=1}^{n}\sum_{k=1}^{r} X_{ijk}}{rn} = \frac{T_{i*}}{rn}$$

- Trung bình của mức B_j:

$$\overline{X}_{i*} = \frac{\sum_{i=1}^{m}\sum_{k=1}^{r} X_{ijk}}{rm} = \frac{T_{*j}}{rm}$$

- Ước lượng giá trị X_{ij} từ mô hình:

$$\hat{X}_{ijk} = \overline{X}_{ij}$$

- Phần dư:

$$e_{ijk} = X_{ijk} - \overline{X}_{ij}, 1 \leq i \leq m; 1 \leq j \leq n; 1 \leq k \leq r$$

- Trung bình chung:

$$\overline{X} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{r} X_{ijk}}{rmn} = \frac{T}{rmn}$$

- Tổng bình phương chung:

$$SST = Q - \frac{T^2}{rmn}$$

- Tổng bình phương cho nhân tố A:

$$SSA = \sum_{j=1}^{n} n_{j*}(\overline{X}_{j0} - \overline{X})^2 = \frac{1}{rn}\sum_{i=1}^{m} T_{i*}^2 - \frac{T^2}{rmn}$$

- Tổng bình phương cho nhân tố B:

$$SSB = \sum_{i=1}^{m} rm(\overline{X}_{*j} - \overline{X})^2 = \frac{1}{rm}\sum_{j=1}^{n} T_{*j}^2 - \frac{T^2}{rmn}$$

- Tổng bình phương do sai số:

$$SSE = \sum_{k=1}^{r}\sum_{i=1}^{m}\sum_{j=1}^{n}(X_{ijk} - \overline{X}_{ij})^2 = Q - \frac{1}{r}\sum_{i=1}^{m}\sum_{j=1}^{n} T_{ij}^2$$

- Tổng bình phương do tương tác:

$$SSI = \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{r}(\overline{X}_{ij} - \overline{X}_{j*} - \overline{X}_{*k} + \overline{X})^2 = SST - SSA - SSB - SSE.$$

- Trung bình bình phương nhân tố A:

$$MSA = \frac{SSA}{m-1}$$

$m-1$ được gọi là bậc tự do của A.
- Trung bình bình phương nhân tố B:

$$MSB = \frac{SSB}{n-1}$$

$n-1$ được gọi là bậc tự do của B.
- Trung bình bình phương sai số:

$$MSE = \frac{SSE}{rmn - mn}$$

$rmn - mn$ được gọi là bậc tự do của sai số.
- Trung bình bình phương của tương tác:

$$MSI = \frac{SSI}{(m-1)(n-1)}$$

$(m-1)(n-1)$ được gọi là bậc tự do của tương tác.

- Tỉ số F cho nhân tố A:

$$F_A = \frac{MSA}{MSE}$$

- Tỉ số F cho nhân tố B:

$$F_A = \frac{MSB}{MSE}$$

- Tỉ số F cho tương tác A và B:

$$F_{AB} = \frac{MSI}{MSE}$$

Các kết quả nói trên được trình bày trong bảng sau đây gọi là bảng ANOVA 2 nhân tố:

Nguồn	Bậc tự do	SS	MS	Tỉ số F
Nhân tố A	$m-1$	SSA	MSA	F_A
Nhân tố B	$n-1$	SSB	MSB	F_B
Tương tác	$(m-1)(n-1)$	SSI	MSI	F_{AB}
Sai số	$rmn-mn$	SSE	MSE	
Tổng	$rmn-1$	SST		

Ta có quy tắc kiểm định như sau.

1. Bài toán kiểm định giả thuyết 1:

$$\begin{cases} H_0^A\text{: Các mức nhân tố } A \text{ có hiệu quả như nhau} \\ H_1^A\text{: Có ít nhất hai mức nhân tố } A \text{ có hiệu quả khác nhau} \end{cases}$$

Miền bác bỏ H_0^A là $W = [f_{m-1,rmn-mn}(\alpha); +\infty)$.

p-giá trị $= P(F_{m-1,rmn-mn} > F_A)$.

2. Bài toán kiểm định giả thuyết 2:

$$\begin{cases} H_0^B\text{: Các mức nhân tố } B \text{ có hiệu quả như nhau} \\ H_1^B\text{: Có ít nhất hai mức nhân tố } B \text{ có hiệu quả khác nhau} \end{cases}$$

Miền bác bỏ H_0^B là $W = [f_{n-1,rmn-mn}(\alpha); +\infty)$.

p-giá trị $= P(F_{n-1,rmn-mn} > F_B)$.

3. Bài toán kiểm định giả thuyết 3:

$$\begin{cases} H_0^{AB}\text{: Không có sự tương tác giữa } A \text{ và } B \\ H_1^{AB}\text{: Có sự tương tác giữa } A \text{ và } B \end{cases}$$

Miền bác bỏ H_0^{AB} là $W = [f_{(m-1)(n-1),rmn-mn}(\alpha); +\infty)$.

p-giá trị $= P(F_{(m-1)(n-1),rmn-mn} > F_{AB})$.

Trong đó, $F_{x,y}$ là phân bố F, $f_{x,y}(\alpha)$ là giá trị tới hạn mức α của phân bố F.

Ví dụ 8.4. Hàm lượng saponin (mg) của cùng một loại dược liệu được thu hái trong 2 mùa (khô và mưa: trong mỗi mùa lấy mẫu 3 lần (đầu mùa, giữa mùa, cuối mùa) và từ 3 miền (Nam, Trung, Bắc) thu được kết quả sau:

Mùa	Thời điểm	Miền		
		Nam	Trung	Bắc
Khô	Đầu mùa	2,4	2,1	3,2
	Giữa mùa	2,4	2,2	3,2
	Cuối mùa	2,5	2,2	3,4
Mưa	Đầu mùa	2,5	2,2	3,4
	Giữa mùa	2,5	2,3	3,5
	Cuối mùa	2,6	2,3	3,5

Với mức ý nghĩa $\alpha = 0,05$ hãy cho biết hàm lượng saponin có khác nhau theo mùa hay miền không? Nếu có thì 2 yếu tố mùa và miền có sự tương tác với nhau hay không?

Giải. Ta có bảng các giá trị của T_{ij}:

$T_{11} = 7,3$	$T_{12} = 6,5$	$T_{13} = 9,8$	$T_{1*} = 23,6$
$T_{21} = 7,6$	$T_{22} = 6,8$	$T_{23} = 9,8$	$T_{2*} = 10,4$
$T_{*1} = 14,9$	$T_{*2} = 13,3$	$T_{*3} = 20,2$	$T = 48,4$

$Q = 134,64$.

Suy ra bảng ANOVA 2 nhân tố sau:

Nguồn	Bậc tự do	SS	MS	F
Nhân tố Mùa	1	$0,08$	$0,08$	$F_A = 16$
Nhân tố Miền	2	$4,3478$	$2,1739$	$F_B = 434,78$
Tương tác	2	$0,01$	$0,005$	$F_{AB} = 1$
Sai số	12	$0,06$	$0,005$	
Tổng	17	155699,6		

Miền bác bỏ H_0^A là $W = [4.7472; +\infty)$.

Miền bác bỏ H_0^B là $W = [3.8853; +\infty)$.

Miền bác bỏ H_0^{AB} là $W = [3.8853; +\infty)$.

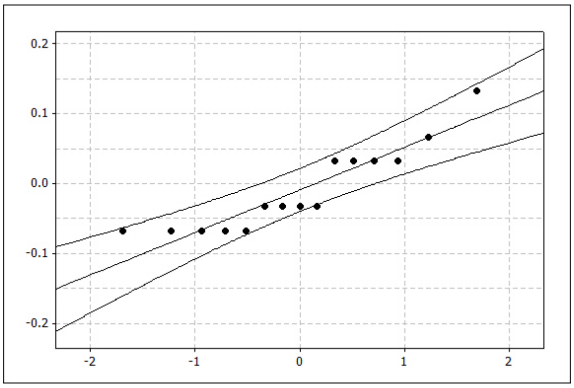
Vì vậy, có thể kết luận: Hàm lượng saponin trong dược liệu khác nhau theo mùa, theo miền và không có sự tương tác giữa mùa và miền trên hàm lượng saponin.

Kiểm tra giả thiết của mô hình.

Để kiểm gia giả thiết của môn hình: Sai số ngẫu nhiên có phân bố chuẩn $N(0; \sigma^2)$ ta vẽ biểu đồ xác suất chuẩn của phần dư. Ta có số liệu của phần dư như sau:

-0,033	-0,067	-0,067
-0,033	0,033	-0,067
0,067	0,033	0,133
-0,033	-0,067	-0,067
-0,033	0,033	0,033

Biểu đồ xác suất chuẩn (xem Hình 8.3) của phần dư nằm xấp xỉ trên đường thẳng đi qua gốc tọa độ nên giả thiết sai số ngẫu nhiên có phân bố chuẩn $N(0; \sigma^2)$ xem như được thỏa mãn.



Hình 8.3

■

8.3. Đại cương về bố trí thí nghiệm

8.3.1. Một số khái niệm

Bố trí thí nghiệm: Là lập kế hoạch về các bước cần tiến hành để thu thập số liệu cho vấn đề đang nghiên cứu. Mục đích để có nhiều kết luận chính xác với chi phí thấp nhất.

Nghiệm thức: Là các mức nhân tố. Mỗi mức nhân tố được xem là một nghiệm thức.

Đơn vị thí nghiệm: Một đơn vị thí nghiệm là một đơn vị nghiên cứu trong thí nghiệm, hoặc cụ thể hơn đó là đơn vị nhỏ nhất mà một nghiệm thức được ứng dụng.

8.3.2. Hai nguyên tắc cơ bản về bố trí thí nghiệm

1. Lặp lại. Một nghiệm thức phải được lặp lại nhiều hơn 1 đơn vị thí nghiệm. Số nghiệm thức càng tăng thì sai số chuẩn càng nhỏ và độ chính xác của thí nghiệm càng cao.

2. Ngẫu nhiên hoá. Mẫu phải được chọn sao cho tất cả các đơn vị thí nghiệm được bố trí ngẫu nhiên vào các nghiệm thức.

Điều này giúp tránh được các thành kiến của người làm thí nghiệm cũng như các biến động sinh học, môi trường,...

8.3.3. Kỹ thuật ngẫu nhiên hoá

Để thực hiện việc ngẫu nhiên hoá, tránh những sai sót chủ quan của người thí nghiệm. Có hai phương pháp ngẫu nhiên thường dùng là dùng bảng số ngẫu nhiên và tạo các số ngẫu nhiên bằng phần mềm thống kê.

8.3.4. Các kiểu bố trí thí nghiệm phổ biến

1. Bố trí ngẫu nhiên hoàn toàn (CRD)

Đây là kiểu bố trí thí nghiệm đơn giản nhất, trong đó tất cả các đơn vị thí nghiệm được bố trí vào các nghiệm thức. Kiểu bố trí này được dùng khi các đơn vị thí nghiệm không có những sai khác mang tính hệ thống. Chẳng hạn tất cả các động vật thí nghiệm có cùng độ tuổi, tất cả các nông trại đều có kỹ thuật canh tác giống nhau...

Kiểu bố trí thí nghiệm này được sử dụng trong phân tích phương sai một nhân tố.

Ví dụ 8.5. Ta cần bố trí 20 con gà tương đương nhau vào 4 khẩu phần ăn, mỗi khẩu phần ăn có 5 con gà.

Sử dụng phần mềm thống kê Minitab ta thực hiện như sau: Kí hiệu các khẩu phần ăn là KP1, PH2, KP3 và KP4. Nhập vào cột C1 từ dòng 1 đến dòng 5 là KP1, từ dòng 6 đến dòng 10 là KP2, từ dòng 11 đến dòng 15 là KP3 và từ dòng 16 đến dòng 20 là KP4. Sau đó vào Calc → Random Data → Integer.

- Number of rows of data to generate: nhập 20;
- Store in column(s): nhập C2;
- Minimum value: nhập 1;
- Maximum value: nhập 20;
- Kết thúc chọn OK.

Trình bày số liệu của CRD

Các nghiệm thức / Lần lặp lại	1	2	...	k
1	X_{11}	X_{12}	...	X_{1k}
2	X_{21}	X_{22}	...	X_{2k}
...	...	...	...	...
m	X_{m1}	X_{m2}	...	X_{mk}

Sử dụng phân tích phương sai 1 nhân tố để so sánh hiệu quả của các nghiệm thức.

2. Bố trí khối hoàn toàn ngẫu nhiên (RCBD)

Là kiểu bố trí trong đó các đơn vị thí nghiệm được tập hợp thành từng khối (block) sao cho mỗi khối có đầy đủ tất cả các nghiệm thức, trong mỗi khối các đơn vị thí nghiệm có tính chất đồng đều. Giữa các khối có ảnh hưởng khác nhau tới kết quả thí nghiệm.

Trình bày số liệu của RCBD

Các nghiệm thức / Lần lặp lại khối	1	2	...	k
1	X_{11}	X_{12}	...	X_{1k}
2	X_{21}	X_{22}	...	X_{2k}
...	...	...	...	...
m	X_{m1}	X_{m2}	...	X_{mk}

Sử dụng phân tích phương sai 2 nhân tố (nhân tố hàng là các khối và nhân tố cột là các nghiệm thức) để so sánh hiệu quả của các nghiệm thức.

3. Bố trí ô vuông La tinh (LSD)

Bố trí thí nghiệm kiểu RCBD khắc phục được vấn đề của CRD đối với đơn vị thí nghiệm không đồng nhất theo một hướng. Tuy nhiên trong thực tế đôi khi ta gặp phải trường hợp các biến động xảy ra theo cả hai hướng. Trong trường hợp này nếu bố trí theo kiểu RCBD ta cũng sẽ gặp phải hai vấn đề là mất độ chính xác và kết luận sai. Vì vậy chúng ta cần bố trí khối theo hai hướng.

Trình bày số liệu của LSD

Hướng 2 (cột) / Hướng 1 (hàng)	1	2	...	m
1	X_{11}	X_{12}	...	X_{1m}
2	X_{21}	X_{22}	...	X_{2m}
...	...	...	...	...
m	X_{m1}	X_{m2}	...	X_{mm}

Các kết quả nói trên được trình bày trong bảng ANOVA sau đây.

Nguồn	SS	Bậc tự do	MS	Tỉ số F
Nghiệm thức	$SSTr$	$m-1$	MST	$MSTr/MSE$
Hàng	SSA	$m-1$	MSA	
Cột	SSB	$m-1$	MSB	
Sai số	SSE	$(m-1)(m-2)$	MSE	
Tổng	SST	m^2-1		

Trong đó SSA, SSB, SSE, SST, MSA, MSB tính tương tự trong phân tích phương sai 2 nhân tố không lặp.

$SSTr = \sum_{i=1}^{m} \dfrac{T_i^2}{m} - \dfrac{Q^2}{m^2}$ với T_i là tổng theo nghiệm thức.

$MSE = MST - MSA - MSB - MSTr.$

$MST = SSTr/(m-1)$, $MSE = SSE/(m-1)(m-2)$.

Bài toán kiểm định giả thuyết:
H_0: Các nghiệm thức có giá trị trung bình bằng nhau.
H_1: Có ít nhất 2 nghiệm thức có giá trị trung bình khác nhau.
Miền bác bỏ H_0 là $W = [f_{m-1,(m-1)(m-2)}(\alpha); +\infty)$.

BÀI TẬP CHƯƠNG 8

▷ **8.1.** Hàm lượng Oxygen trong nước là một chỉ tiêu để xem xét mức độ ô nhiễm môi trường. Trong một lần khảo sát người ta lấy ngẫu nhiên 24 mẫu ở 4 khu vực khác nhau, ký hiệu là KV1, KV2, KV3, KV4. Mẫu được đem phân tích và đo hàm lượng oxygen hòa tan (đơn vị tính theo phần triệu). Kết quả ghi nhận ở bảng sau:

Mẫu số	1	2	3	4	5	6
KV1	6,3	6,8	6,5	5,9	5,8	5,7
KV2	6,9	6,3	6,2	5,4	5,7	5,9
KV3	4,8	4,7	4,5	4,8	4,9	4,6
KV4	4,8	4,9	5,2	5,6	5,8	5,7

Với mức ý nghĩa 5%, bốn khu vực có thật sự khác nhau về hàm lượng oxygen hòa tan hay không?

▷ **8.2.** Một nhà nông học khảo sát hàm lượng phosphorus của lá cây từ 3 giống táo (1, 2 và 3). Mẫu được lấy ngẫu nhiên từ 5 lá của mỗi giống đem phân tích hàm lượng phosphorus. Dữ liệu được trình bày trong bảng sau:

Giống	Hàm lượng phosphorus				
1	0,35	0,40	0,58	0,50	0,47
2	0,65	0,70	0,90	0,84	0,79
3	0,60	0,80	0,75	0,73	0,66

Với mức ý nghĩa 5% hãy kiểm tra giả thuyết gốc "hàm lượng phosphorus trung bình của ba giống táo là giống nhau".

▷ **8.3.** Nghiên cứu năng suất của 5 giống bắp lai được thụ phấn ngẫu nhiên, thí nghiệm được lặp lại 5 lần trên mỗi giống, kết quả ghi nhận trong bảng sau:

Giống 1	Giống 2	Giống 3	Giống 4	Giống 5
4,5	3,8	4,4	4,7	3,5
4,6	4,2	6,3	4,3	4,2
4,8	3,8	4,5	4,6	4,3
4,6	3,9	5,3	4,6	4,2
4,1	4,1	5,9	3,8	3,2

Với mức ý nghĩa 5% hãy kiểm tra xem năng suất của 5 giống bắp nầy có phụ thuộc vào phẩm chất giống hay không?

▷ **8.4.** Một phòng thí nghiệm lớn có 4 loại thiết bị để đo độ pH của các mẫu đất. Họ muốn xác định liệu có sự sai khác về giá trị trung bình đọc được từ bốn thiết bị này hay không nên đã bố trí thí nghiệm như sau: 24 mẫu đất có độ pH đã biết được phân bố ngẫu nhiên thành 4 nhóm ứng với 4 thiết bị, mỗi nhóm gồm 6 mẫu đất. Các mẫu đất (MD) được đo pH bằng thiết bị (TB và sự sai khác giữa pH đọc từ thiết bị với pH đã biết được ghi nhận. Kết quả được trình bày trong bảng sau:

TB \ MD	1	2	3	4	5	6
A	-0,307	-0,294	0,079	0,019	-0,136	-0,324
B	-0,176	0,125	-0,013	0,082	0,091	0,459
C	0,137	-0,063	0,240	-0,050	0,318	0,154
D	-0,042	0,690	0,201	0,166	0,219	0,407

Với mức ý nghĩa 5% hãy xác định xem trung bình pH đọc được từ bốn thiết bị có sai khác nhau hay không?

▷ **8.5.** Các nhà nghiên cứu tiến hành thí nghiệm để so sánh hàm lượng tinh bột của khoai tây trồng trên đất cát được bổ sung chất dinh dưỡng A, B, C. Nhóm A là nhóm đối chứng (tưới bằng nước cất), nhóm B được cung cấp chất dinh dưỡng Hoagland với nồng độ thấp và nhóm C được cung cấp chất dinh dưỡng Hoagland với nồng độ cao. 18 cây khoai tây cùng một giống được bố trí ngẫu nhiên vào ba nhóm và hàm lượng tinh bột ở cuống lá (μg/mg) được ghi nhận sau 25 ngày trồng ở bảng sau:

Hàm lượng dinh dưỡng A	22	20	21	18	16	14
Hàm lượng dinh dưỡng B	12	14	15	10	9	6
Hàm lượng dinh dưỡng C	7	9	7	6	5	3

Với mức ý nghĩa 5% hãy kiểm tra sự khác biệt về hàm lượng tinh bột của cây thuộc ba nhóm.

▹ **8.6.** Ba kiểu gen khác nhau BB, Bb, bb của loài bọ cánh cứng Tribolium castaneum được nuôi ở mật độ 20 con/1 g bột. Trọng lượng khô (mg) của mỗi kiểu gen được ghi nhận ở 4 thời điểm khác nhau, mỗi thời điểm cách nhau vài tháng.

Thời điểm \ Kiểu gen	BB	Bb	bb
1	0,958	0,986	0,925
2	0,971	1,051	0,952
3	0,927	0,891	0,829
4	0,971	1,01	0,955

Với mức ý nghĩa 5% hãy kiểm tra xem nhân tố A (thời điểm) và nhân tố B (kiểu gen) có ảnh hưởng đến trọng lượng bọ hay không?

▹ **8.7.** Nhằm khảo sát sự sai khác trong tiêu thụ thức ăn khi mỡ tươi được thay bằng mỡ ôi trong khẩu phần ăn của chuột, người ta bố trí thí nghiệm trên 12 con chuột (6 đực, 6 cái) có độ tuổi từ 30 đến 34 ngày, mỗi khẩu phần ăn cho mỗi giới gồm 3 chuột. Lượng thức ăn tiêu thụ (g) trong suốt 73 ngày được ghi nhận trong bảng sau:

Giới tính \ Khẩu phần ăn	Mỡ tươi	Mỡ ôi
Đực	709	592
	679	538
	699	476
Cái	657	508
	594	505
	677	539

Với mức ý nghĩa 5% hãy kiểm tra xem:

a) Có sự khác nhau về tiêu thụ thức ăn giữa chuột đực và chuột cái không?

b) Có sự khác nhau về tiêu thụ thức ăn giữa mỡ tươi và mỡ ôi không?

▷ **8.8.** Một thầy giáo phân phối ngẫu nhiên 20 sinh viên trong có trình độ đồng đều ra thành 4 nhóm, mỗi nhóm 5 sinh viên. Nhóm A học 1 giờ/tuần, nhóm B học 2 giờ/tuần, nhóm C học 3 giờ/tuần và nhóm D học 4 giờ/tuần. Sau 4 tuần lễ, thầy giáo ra một bài khảo sát. Kết quả điểm số của 4 nhóm như sau:

Nhóm A	Nhóm B	Nhóm C	Nhóm D
19,2	18,4	22,0	24,5
21,0	19,0	23,1	27,0
23,3	17,6	19,8	26,7
20,4	19,5	17,0	27,5
22,5	20,1	20,4	27,8

Với mức ý nghĩa 5%, có thể cho rằng điểm trung bình của sinh viên ở 4 nhóm là bằng nhau không?

▷ **8.9.** Nhằm khảo sát ảnh hưởng của stress và cường độ sáng lên tăng trưởng của cây đậu nành, thí nghiệm được bố trí ngẫu nhiên thành 4 nghiệm thức, mỗi nghiệm thức gồm 13 hạt đậu trồng trong

13 chậu. Để gây stress người ta cho chậu vào máy lắc, mỗi ngày lắc 2 lần, mỗi lần 20 phút. Có thể tóm tắt bố trí thí nghiệm như sau:

Nghiệm thức 1: đối chứng, ánh sáng yếu;

Nghiệm thức 2: stress, ánh sáng yếu;

Nghiệm thức 3: đối chứng, ánh sáng trung bình;

Nghiệm thức 4: stress, ánh sáng trung bình.

Sau 16 ngày trồng, các cây được thu hoạch và đo tổng diện tích lá (cm^2) của mỗi cây. Kết quả ghi nhận trong bảng sau:

NT1	NT2	NT3	NT4
264	235	314	283
200	188	320	312
225	195	310	291
268	205	340	259
215	212	299	216
241	214	268	201
232	182	345	267
256	215	271	326
229	272	285	241
288	163	309	291
256	230	337	269
288	255	282	282
230	202	273	257

Với mức ý nghĩa 5% hãy kiểm tra xem:

a) Nhân tố ánh sáng có ảnh hưởng đến diện tích lá không?

b) Nhân tố stress có ảnh hưởng đến diện tích lá không?

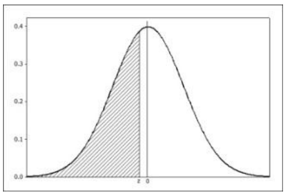
c) Hai yếu tố ánh sáng và stress có tương tác nhau không?

▷ **8.10.** Một thầy giáo phân ngẫu nhiên 24 học sinh đồng đều về trình độ trong lớp ra thành 4 nhóm và giảng dạy 4 nhóm ấy theo 4 phương pháp khác nhau. Sau 1 thời gian ông ra một bài trắc nghiệm thành quả học tập cho cả 4 nhóm. Kết quả điểm số như sau:

Nhóm 1	Nhóm 2	Nhóm 3	Nhóm 4
30	50	18	88
74	38	56	78
46	66	34	60
58	62	24	76
62	44	66	80
38	58	52	75

Với mức ý nghĩa 5%, có thể cho rằng điểm trung bình của sinh viên ở 4 nhóm là bằng nhau không?

BẢNG PHỤ LỤC



Bảng Ia.

Các giá trị của hàm phân phối chuẩn tắc

$$\Phi(z) = P(Z < z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{\frac{-t^2}{2}} dt$$

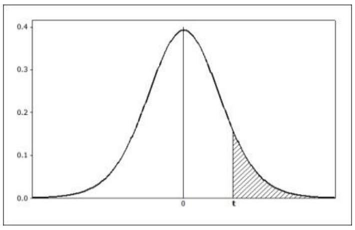
z	0	-0,01	-0,02	-0,03	-0,04	-0,05	-0,06	-0,07	-0,08	-0,09
0,0	0,500	0,496	0,492	0,488	0,484	0,480	0,476	0,472	0,468	0,464
-0,1	0,460	0,456	0,452	0,448	0,444	0,440	0,436	0,433	0,429	0,425
-0,2	0,421	0,417	0,413	0,409	0,405	0,401	0,397	0,394	0,390	0,386
-0,3	0,382	0,378	0,374	0,371	0,367	0,363	0,359	0,356	0,352	0,348
-0,4	0,345	0,341	0,337	0,334	0,330	0,326	0,323	0,319	0,316	0,312
-0,5	0,309	0,305	0,302	0,298	0,295	0,291	0,288	0,284	0,281	0,278
-0,6	0,274	0,271	0,268	0,264	0,261	0,258	0,255	0,251	0,248	0,245
-0,7	0,242	0,239	0,236	0,233	0,230	0,227	0,224	0,221	0,218	0,215
-0,8	0,212	0,209	0,206	0,203	0,200	0,198	0,195	0,192	0,189	0,187
-0,9	0,184	0,181	0,179	0,176	0,174	0,171	0,169	0,166	0,164	0,161
-1,0	0,159	0,156	0,154	0,152	0,149	0,147	0,145	0,142	0,140	0,138
-1,1	0,136	0,133	0,131	0,129	0,127	0,125	0,123	0,121	0,119	0,117
-1,2	0,115	0,113	0,111	0,109	0,107	0,106	0,104	0,102	0,100	0,099
-1,3	0,097	0,095	0,093	0,092	0,090	0,089	0,087	0,085	0,084	0,082
-1,4	0,081	0,079	0,078	0,076	0,075	0,074	0,072	0,071	0,069	0,068
-1,5	0,067	0,066	0,064	0,063	0,062	0,061	0,059	0,058	0,057	0,056
-1,6	0,055	0,054	0,053	0,052	0,051	0,049	0,048	0,047	0,046	0,046
-1,7	0,045	0,044	0,043	0,042	0,041	0,040	0,039	0,038	0,038	0,037
-1,8	0,036	0,035	0,034	0,034	0,033	0,032	0,031	0,031	0,030	0,029
-1,9	0,029	0,028	0,027	0,027	0,026	0,026	0,025	0,024	0,024	0,023
-2,0	0,023	0,022	0,022	0,021	0,021	0,020	0,020	0,019	0,019	0,018
-2,1	0,018	0,017	0,017	0,017	0,016	0,016	0,015	0,015	0,015	0,014
-2,2	0,014	0,014	0,013	0,013	0,013	0,012	0,012	0,012	0,011	0,011
-2,3	0,011	0,010	0,010	0,010	0,010	0,009	0,009	0,009	0,009	0,008
-2,4	0,008	0,008	0,008	0,008	0,007	0,007	0,007	0,007	0,007	0,006
-2,5	0,006	0,006	0,006	0,006	0,006	0,005	0,005	0,005	0,005	0,005
-2,6	0,005	0,005	0,004	0,004	0,004	0,004	0,004	0,004	0,004	0,004
-2,7	0,003	0,003	0,003	0,003	0,003	0,003	0,003	0,003	0,003	0,003
-2,8	0,003	0,002	0,002	0,002	0,002	0,002	0,002	0,002	0,002	0,002
-2,9	0,002	0,002	0,002	0,002	0,002	0,002	0,002	0,001	0,001	0,001
-3,0	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001
-3,1	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001
-3,2	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001
-3,3	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000

Bảng Ib.

Các giá trị của hàm phân phối chuẩn tắc

$$\Phi(z) = P(Z < z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{\frac{-t^2}{2}} dt$$

z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,500	0,504	0,508	0,512	0,516	0,520	0,524	0,528	0,532	0,536
0,1	0,540	0,544	0,548	0,552	0,556	0,560	0,564	0,567	0,571	0,575
0,2	0,579	0,583	0,587	0,591	0,595	0,599	0,603	0,606	0,610	0,614
0,3	0,618	0,622	0,626	0,629	0,633	0,637	0,641	0,644	0,648	0,652
0,4	0,655	0,659	0,663	0,666	0,670	0,674	0,677	0,681	0,684	0,688
0,5	0,691	0,695	0,698	0,702	0,705	0,709	0,712	0,716	0,719	0,722
0,6	0,726	0,729	0,732	0,736	0,739	0,742	0,745	0,749	0,752	0,755
0,7	0,758	0,761	0,764	0,767	0,770	0,773	0,776	0,779	0,782	0,785
0,8	0,788	0,791	0,794	0,797	0,800	0,802	0,805	0,808	0,811	0,813
0,9	0,816	0,819	0,821	0,824	0,826	0,829	0,831	0,834	0,836	0,839
1,0	0,841	0,844	0,846	0,848	0,851	0,853	0,855	0,858	0,860	0,862
1,1	0,864	0,867	0,869	0,871	0,873	0,875	0,877	0,879	0,881	0,883
1,2	0,885	0,887	0,889	0,891	0,893	0,894	0,896	0,898	0,900	0,901
1,3	0,903	0,905	0,907	0,908	0,910	0,911	0,913	0,915	0,916	0,918
1,4	0,919	0,921	0,922	0,924	0,925	0,926	0,928	0,929	0,931	0,932
1,5	0,933	0,934	0,936	0,937	0,938	0,939	0,941	0,942	0,943	0,944
1,6	0,945	0,946	0,947	0,948	0,949	0,951	0,952	0,953	0,954	0,954
1,7	0,955	0,956	0,957	0,958	0,959	0,960	0,961	0,962	0,962	0,963
1,8	0,964	0,965	0,966	0,966	0,967	0,968	0,969	0,969	0,970	0,971
1,9	0,971	0,972	0,973	0,973	0,974	0,974	0,975	0,976	0,976	0,977
2,0	0,977	0,978	0,978	0,979	0,979	0,980	0,980	0,981	0,981	0,982
2,1	0,982	0,983	0,983	0,983	0,984	0,984	0,985	0,985	0,985	0,986
2,2	0,986	0,986	0,987	0,987	0,987	0,988	0,988	0,988	0,989	0,989
2,3	0,989	0,990	0,990	0,990	0,990	0,991	0,991	0,991	0,991	0,992
2,4	0,992	0,992	0,992	0,992	0,993	0,993	0,993	0,993	0,993	0,994
2,5	0,994	0,994	0,994	0,994	0,994	0,995	0,995	0,995	0,995	0,995
2,6	0,995	0,995	0,996	0,996	0,996	0,996	0,996	0,996	0,996	0,996
2,7	0,997	0,997	0,997	0,997	0,997	0,997	0,997	0,997	0,997	0,997
2,8	0,997	0,998	0,998	0,998	0,998	0,998	0,998	0,998	0,998	0,998
2,9	0,998	0,998	0,998	0,998	0,998	0,998	0,998	0,999	0,999	0,999
3,0	0,999	0,999	0,999	0,999	0,999	0,999	0,999	0,999	0,999	0,999
3,1	0,999	0,999	0,999	0,999	0,999	0,999	0,999	0,999	0,999	0,999
3,2	0,999	0,999	0,999	0,999	0,999	0,999	0,999	0,999	0,999	0,999
3,3	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000



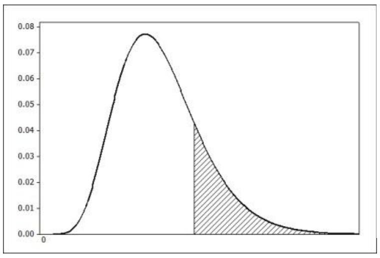
Bảng II.

Giá trị tới hạn của t-phân bố

$$P(T_n > t_{n,\alpha}) = \alpha$$

$n > 30 : t_{n;\alpha} \approx z_{\alpha}$

n	α					
	0,250	**0,100**	**0,050**	**0,025**	**0,010**	**0,005**
1	1,0000	3,0777	6,3138	12,7062	31,8205	63,6567
2	0,8165	1,8856	2,9200	4,3027	6,9646	9,9248
3	0,7649	1,6377	2,3534	3,1824	4,5407	5,8409
4	0,7407	1,5332	2,1318	2,7764	3,7469	4,6041
5	0,7267	1,4759	2,0150	2,5706	3,3649	4,0321
6	0,7176	1,4398	1,9432	2,4469	3,1427	3,7074
7	0,7111	1,4149	1,8946	2,3646	2,9980	3,4995
8	0,7064	1,3968	1,8595	2,3060	2,8965	3,3554
9	0,7027	1,3830	1,8331	2,2622	2,8214	3,2498
10	0,6998	1,3722	1,8125	2,2281	2,7638	3,1693
11	0,6974	1,3634	1,7959	2,2010	2,7181	3,1058
12	0,6955	1,3562	1,7823	2,1788	2,6810	3,0545
13	0,6938	1,3502	1,7709	2,1604	2,6503	3,0123
14	0,6924	1,3450	1,7613	2,1448	2,6245	2,9768
15	0,6912	1,3406	1,7531	2,1314	2,6025	2,9467
16	0,6901	1,3368	1,7459	2,1199	2,5835	2,9208
17	0,6892	1,3334	1,7396	2,1098	2,5669	2,8982
18	0,6884	1,3304	1,7341	2,1009	2,5524	2,8784
19	0,6876	1,3277	1,7291	2,0930	2,5395	2,8609
20	0,6870	1,3253	1,7247	2,0860	2,5280	2,8453
21	0,6864	1,3232	1,7207	2,0796	2,5176	2,8314
22	0,6858	1,3212	1,7171	2,0739	2,5083	2,8188
23	0,6853	1,3195	1,7139	2,0687	2,4999	2,8073
24	0,6848	1,3178	1,7109	2,0639	2,4922	2,7969
25	0,6844	1,3163	1,7081	2,0595	2,4851	2,7874
26	0,6840	1,3150	1,7056	2,0555	2,4786	2,7787
27	0,6837	1,3137	1,7033	2,0518	2,4727	2,7707
28	0,6834	1,3125	1,7011	2,0484	2,4671	2,7633
29	0,6830	1,3114	1,6991	2,0452	2,4620	2,7564
30	0,6828	1,3104	1,6973	2,0423	2,4573	2,7500

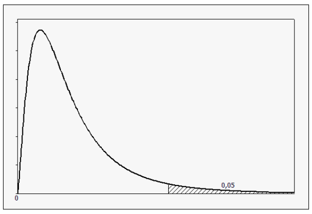


Bảng III.

Giá trị tới hạn của phân bố khi bình phương

$$P(\chi_n^2 > \chi_{n;\alpha}^2) = \alpha$$

	α									
n	**0,995**	**0,990**	**0,975**	**0,950**	**0,900**	**0,100**	**0,050**	**0,025**	**0,010**	**0,005**
1	0,000	0,000	0,001	0,004	0,016	2,706	3,841	5,024	6,635	7,879
2	0,010	0,020	0,051	0,103	0,211	4,605	5,991	7,378	9,210	10,597
3	0,072	0,115	0,216	0,352	0,584	6,251	7,815	9,348	11,345	12,838
4	0,207	0,297	0,484	0,711	1,064	7,779	9,488	11,143	13,277	14,860
5	0,412	0,554	0,831	1,145	1,610	9,236	11,070	12,833	15,086	16,750
6	0,676	0,872	1,237	1,635	2,204	10,645	12,592	14,449	16,812	18,548
7	0,989	1,239	1,690	2,167	2,833	12,017	14,067	16,013	18,475	20,278
8	1,344	1,646	2,180	2,733	3,490	13,362	15,507	17,535	20,090	21,955
9	1,735	2,088	2,700	3,325	4,168	14,684	16,919	19,023	21,666	23,589
10	2,156	2,558	3,247	3,940	4,865	15,987	18,307	20,483	23,209	25,188
11	2,603	3,053	3,816	4,575	5,578	17,275	19,675	21,920	24,725	26,757
12	3,074	3,571	4,404	5,226	6,304	18,549	21,026	23,337	26,217	28,300
13	3,565	4,107	5,009	5,892	7,042	19,812	22,362	24,736	27,688	29,819
14	4,075	4,660	5,629	6,571	7,790	21,064	23,685	26,119	29,141	31,319
15	4,601	5,229	6,262	7,261	8,547	22,307	24,996	27,488	30,578	32,801
16	5,142	5,812	6,908	7,962	9,312	23,542	26,296	28,845	32,000	34,267
17	5,697	6,408	7,564	8,672	10,085	24,769	27,587	30,191	33,409	35,718
18	6,265	7,015	8,231	9,390	10,865	25,989	28,869	31,526	34,805	37,156
19	6,844	7,633	8,907	10,117	11,651	27,204	30,144	32,852	36,191	38,582
20	7,434	8,260	9,591	10,851	12,443	28,412	31,410	34,170	37,566	39,997
21	8,034	8,897	10,283	11,591	13,240	29,615	32,671	35,479	38,932	41,401
22	8,643	9,542	10,982	12,338	14,041	30,813	33,924	36,781	40,289	42,796
23	9,260	10,196	11,689	13,091	14,848	32,007	35,172	38,076	41,638	44,181
24	9,886	10,856	12,401	13,848	15,659	33,196	36,415	39,364	42,980	45,559
25	10,520	11,524	13,120	14,611	16,473	34,382	37,652	40,646	44,314	46,928
26	11,160	12,198	13,844	15,379	17,292	35,563	38,885	41,923	45,642	48,290
27	11,808	12,879	14,573	16,151	18,114	36,741	40,113	43,195	46,963	49,645
28	12,461	13,565	15,308	16,928	18,939	37,916	41,337	44,461	48,278	50,993
29	13,121	14,256	16,047	17,708	19,768	39,087	42,557	45,722	49,588	52,336
30	13,787	14,953	16,791	18,493	20,599	40,256	43,773	46,979	50,892	53,672

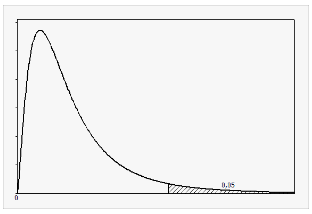


Bảng IVa.

Giá trị tới hạn của phân bố F

$$P(F_{m,n} > f_{m,n}(0,05)) = 0,05$$

m \ n	1	2	3	4	5	6	7	8	9	10	11
1	161,448	18,513	10,128	7,709	6,608	5,987	5,591	5,318	5,117	4,965	4,844
2	199,500	19,000	9,552	6,944	5,786	5,143	4,737	4,459	4,256	4,103	3,982
3	215,707	19,164	9,277	6,591	5,409	4,757	4,347	4,066	3,863	3,708	3,587
4	224,583	19,247	9,117	6,388	5,192	4,534	4,120	3,838	3,633	3,478	3,357
5	230,162	19,296	9,013	6,256	5,050	4,387	3,972	3,687	3,482	3,326	3,204
6	233,986	19,330	8,941	6,163	4,950	4,284	3,866	3,581	3,374	3,217	3,095
7	236,768	19,353	8,887	6,094	4,876	4,207	3,787	3,500	3,293	3,135	3,012
8	238,883	19,371	8,845	6,041	4,818	4,147	3,726	3,438	3,230	3,072	2,948
9	240,543	19,385	8,812	5,999	4,772	4,099	3,677	3,388	3,179	3,020	2,896
10	241,882	19,396	8,786	5,964	4,735	4,060	3,637	3,347	3,137	2,978	2,854
11	242,983	19,405	8,763	5,936	4,704	4,027	3,603	3,313	3,102	2,943	2,818
12	243,906	19,413	8,745	5,912	4,678	4,000	3,575	3,284	3,073	2,913	2,788
13	244,690	19,419	8,729	5,891	4,655	3,976	3,550	3,259	3,048	2,887	2,761
14	245,364	19,424	8,715	5,873	4,636	3,956	3,529	3,237	3,025	2,865	2,739
15	245,950	19,429	8,703	5,858	4,619	3,938	3,511	3,218	3,006	2,845	2,719
16	246,464	19,433	8,692	5,844	4,604	3,922	3,494	3,202	2,989	2,828	2,701
17	246,918	19,437	8,683	5,832	4,590	3,908	3,480	3,187	2,974	2,812	2,685
18	247,323	19,440	8,675	5,821	4,579	3,896	3,467	3,173	2,960	2,798	2,671
19	247,686	19,443	8,667	5,811	4,568	3,884	3,455	3,161	2,948	2,785	2,658
20	248,013	19,446	8,660	5,803	4,558	3,874	3,445	3,150	2,936	2,774	2,646
21	248,309	19,448	8,654	5,795	4,549	3,865	3,435	3,140	2,926	2,764	2,636
22	248,579	19,450	8,648	5,787	4,541	3,856	3,426	3,131	2,917	2,754	2,626
23	248,826	19,452	8,643	5,781	4,534	3,849	3,418	3,123	2,908	2,745	2,617
24	249,052	19,454	8,639	5,774	4,527	3,841	3,410	3,115	2,900	2,737	2,609
25	249,260	19,456	8,634	5,769	4,521	3,835	3,404	3,108	2,893	2,730	2,601
26	249,453	19,457	8,630	5,763	4,515	3,829	3,397	3,102	2,886	2,723	2,594
27	249,631	19,459	8,626	5,759	4,510	3,823	3,391	3,095	2,880	2,716	2,588
28	249,797	19,460	8,623	5,754	4,505	3,818	3,386	3,090	2,874	2,710	2,582
29	249,951	19,461	8,620	5,750	4,500	3,813	3,381	3,084	2,869	2,705	2,576
30	250,095	19,462	8,617	5,746	4,496	3,808	3,376	3,079	2,864	2,700	2,570
31	250,230	19,463	8,614	5,742	4,492	3,804	3,371	3,075	2,859	2,695	2,565

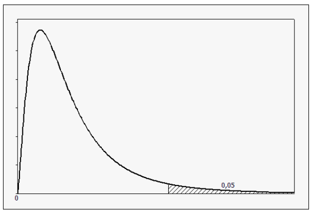


Bảng IVb.

Giá trị tới hạn của phân bố F

$$P(F_{m,n} > f_{m,n}(0,05)) = 0,05$$

m \ n	12	13	14	15	16	17	18	19	20	21	22
1	4,747	4,667	4,600	4,543	4,494	4,451	4,414	4,381	4,351	4,325	4,301
2	3,885	3,806	3,739	3,682	3,634	3,592	3,555	3,522	3,493	3,467	3,443
3	3,490	3,411	3,344	3,287	3,239	3,197	3,160	3,127	3,098	3,072	3,049
4	3,259	3,179	3,112	3,056	3,007	2,965	2,928	2,895	2,866	2,840	2,817
5	3,106	3,025	2,958	2,901	2,852	2,810	2,773	2,740	2,711	2,685	2,661
6	2,996	2,915	2,848	2,790	2,741	2,699	2,661	2,628	2,599	2,573	2,549
7	2,913	2,832	2,764	2,707	2,657	2,614	2,577	2,544	2,514	2,488	2,464
8	2,849	2,767	2,699	2,641	2,591	2,548	2,510	2,477	2,447	2,420	2,397
9	2,796	2,714	2,646	2,588	2,538	2,494	2,456	2,423	2,393	2,366	2,342
10	2,753	2,671	2,602	2,544	2,494	2,450	2,412	2,378	2,348	2,321	2,297
11	2,717	2,635	2,565	2,507	2,456	2,413	2,374	2,340	2,310	2,283	2,259
12	2,687	2,604	2,534	2,475	2,425	2,381	2,342	2,308	2,278	2,250	2,226
13	2,660	2,577	2,507	2,448	2,397	2,353	2,314	2,280	2,250	2,222	2,198
14	2,637	2,554	2,484	2,424	2,373	2,329	2,290	2,256	2,225	2,197	2,173
15	2,617	2,533	2,463	2,403	2,352	2,308	2,269	2,234	2,203	2,176	2,151
16	2,599	2,515	2,445	2,385	2,333	2,289	2,250	2,215	2,184	2,156	2,131
17	2,583	2,499	2,428	2,368	2,317	2,272	2,233	2,198	2,167	2,139	2,114
18	2,568	2,484	2,413	2,353	2,302	2,257	2,217	2,182	2,151	2,123	2,098
19	2,555	2,471	2,400	2,340	2,288	2,243	2,203	2,168	2,137	2,109	2,084
20	2,544	2,459	2,388	2,328	2,276	2,230	2,191	2,155	2,124	2,096	2,071
21	2,533	2,448	2,377	2,316	2,264	2,219	2,179	2,144	2,112	2,084	2,059
22	2,523	2,438	2,367	2,306	2,254	2,208	2,168	2,133	2,102	2,073	2,048
23	2,514	2,429	2,357	2,297	2,244	2,199	2,159	2,123	2,092	2,063	2,038
24	2,505	2,420	2,349	2,288	2,235	2,190	2,150	2,114	2,082	2,054	2,028
25	2,498	2,412	2,341	2,280	2,227	2,181	2,141	2,106	2,074	2,045	2,02
26	2,491	2,405	2,333	2,272	2,220	2,174	2,134	2,098	2,066	2,037	2,012
27	2,484	2,398	2,326	2,265	2,212	2,167	2,126	2,090	2,059	2,030	2,004
28	2,478	2,392	2,320	2,259	2,206	2,160	2,119	2,084	2,052	2,023	1,997
29	2,472	2,386	2,314	2,253	2,200	2,154	2,113	2,077	2,045	2,016	1,99
30	2,466	2,380	2,308	2,247	2,194	2,148	2,107	2,071	2,039	2,010	1,984
31	2,461	2,375	2,303	2,241	2,188	2,142	2,102	2,066	2,033	2,004	1,978



Bảng IVc.

Giá trị tới hạn của phân bố F

$$P(F_{m,n} > f_{m,n;0,05}) = 0,05$$

m \ n	23	24	25	26	27	28	29	30	31	32	33
1	4,279	4,260	4,242	4,225	4,210	4,196	4,183	4,171	4,160	4,149	4,139
2	3,422	3,403	3,385	3,369	3,354	3,340	3,328	3,316	3,305	3,295	3,285
3	3,028	3,009	2,991	2,975	2,960	2,947	2,934	2,922	2,911	2,901	2,892
4	2,796	2,776	2,759	2,743	2,728	2,714	2,701	2,690	2,679	2,668	2,659
5	2,640	2,621	2,603	2,587	2,572	2,558	2,545	2,534	2,523	2,512	2,503
6	2,528	2,508	2,490	2,474	2,459	2,445	2,432	2,421	2,409	2,399	2,389
7	2,442	2,423	2,405	2,388	2,373	2,359	2,346	2,334	2,323	2,313	2,303
8	2,375	2,355	2,337	2,321	2,305	2,291	2,278	2,266	2,255	2,244	2,235
9	2,320	2,300	2,282	2,265	2,250	2,236	2,223	2,211	2,199	2,189	2,179
10	2,275	2,255	2,236	2,220	2,204	2,190	2,177	2,165	2,153	2,142	2,133
11	2,236	2,216	2,198	2,181	2,166	2,151	2,138	2,126	2,114	2,103	2,093
12	2,204	2,183	2,165	2,148	2,132	2,118	2,104	2,092	2,080	2,07	2,06
13	2,175	2,155	2,136	2,119	2,103	2,089	2,075	2,063	2,051	2,04	2,03
14	2,150	2,130	2,111	2,094	2,078	2,064	2,050	2,037	2,026	2,015	2,004
15	2,128	2,108	2,089	2,072	2,056	2,041	2,027	2,015	2,003	1,992	1,982
16	2,109	2,088	2,069	2,052	2,036	2,021	2,007	1,995	1,983	1,972	1,961
17	2,091	2,070	2,051	2,034	2,018	2,003	1,989	1,976	1,965	1,953	1,943
18	2,075	2,054	2,035	2,018	2,002	1,987	1,973	1,960	1,948	1,937	1,926
19	2,061	2,040	2,021	2,003	1,987	1,972	1,958	1,945	1,933	1,922	1,911
20	2,048	2,027	2,007	1,990	1,974	1,959	1,945	1,932	1,920	1,908	1,898
21	2,036	2,015	1,995	1,978	1,961	1,946	1,932	1,919	1,907	1,896	1,885
22	2,025	2,003	1,984	1,966	1,950	1,935	1,921	1,908	1,896	1,884	1,873
23	2,014	1,993	1,974	1,956	1,940	1,924	1,910	1,897	1,885	1,873	1,863
24	2,005	1,984	1,964	1,946	1,930	1,915	1,901	1,887	1,875	1,864	1,853
25	1,996	1,975	1,955	1,938	1,921	1,906	1,891	1,878	1,866	1,854	1,844
26	1,988	1,967	1,947	1,929	1,913	1,897	1,883	1,870	1,857	1,846	1,835
27	1,981	1,959	1,939	1,921	1,905	1,889	1,875	1,862	1,849	1,838	1,827
28	1,973	1,952	1,932	1,914	1,898	1,882	1,868	1,854	1,842	1,83	1,819
29	1,967	1,945	1,926	1,907	1,891	1,875	1,861	1,847	1,835	1,823	1,812
30	1,961	1,939	1,919	1,901	1,884	1,869	1,854	1,841	1,828	1,817	1,806
31	1,955	1,933	1,913	1,895	1,878	1,863	1,848	1,835	1,822	1,81	1,799

Bảng V. Phân vị và tới hạn của Wilcoxon

	l_α	u_α	$P(W \le l_\alpha) = P(W \ge u_\alpha)$		l_α	u_α	$P(W \le l_\alpha) = P(W \ge u_\alpha)$
n=4	0	10	0,062		4	51	0,007
n=5	0	15	0,031		5	50	0,010
	1	14	0,062		6	49	0,014
	2	13	0,094		7	48	0,019
	3	12	0,156		8	47	0,024
n=6	0	21	0,016		9	46	0,032
	1	20	0,031		10	45	0,042
	2	19	0,047		11	44	0,053
	3	18	0,078		12	43	0,065
	4	17	0,109		13	42	0,080
	5	16	0,156		14	41	0,097
n=7	0	28	0,008		15	40	0,116
	1	27	0,016		16	39	0,138
	2	26	0,023	n=11	5	61	0,005
	3	25	0,039		6	60	0,007
	4	24	0,055		7	59	0,009
	5	23	0,078		8	58	0,012
	6	22	0,109		9	57	0,016
	7	21	0,148		10	56	0,021
n=8	0	36	0,004		11	55	0,027
	1	35	0,008		12	54	0,034
	2	34	0,012		13	53	0,042
	3	33	0,020		14	52	0,051
	4	32	0,027		15	51	0,062
	5	31	0,039		16	50	0,074
	6	30	0,055		17	49	0,087
	7	29	0,074		18	48	0,103
	8	28	0,098		19	47	0,120
	9	27	0,125		20	46	0,139
n=9	0	44	0,004	n=12	7	71	0,005
	1	43	0,006		8	70	0,006
	2	42	0,010		9	69	0,008
	3	41	0,014		10	68	0,010
	4	40	0,020		11	67	0,013
	5	39	0,027		12	66	0,017
	6	38	0,037		13	65	0,021
	7	37	0,049		14	64	0,026
	8	36	0,064		15	63	0,032
	9	35	0,082		16	62	0,039
	10	34	0,102		17	61	0,046
	11	33	0,125		18	60	0,055
n=10	3	52	0,005		19	59	0,065

TÀI LIỆU THAM KHẢO

[1] Dương Thiệu Tống (2005), *Thống kê ứng dụng trong nghiên cứu khoa học giáo dục*, Nhà Xuất bản Khoa học Xã hội.

[2] Nguyễn Duy Tiến và Vũ Viết Yên (2009), *Lý thuyết xác suất*, Nhà Xuất bản Giáo dục.

[3] Đặng Hùng Thắng (2009), *Mở đầu về lý thuyết xác suất và các ứng dụng*, Nhà Xuất bản Giáo dục.

[4] Đặng Hùng Thắng (2008), *Thống kê và ứng dụng*, Nhà Xuất bản Giáo dục.

[5] Đặng Hùng Thắng (2009), *Bài tập xác suất*, Nhà Xuất bản Giáo dục.

[6] Đặng Hùng Thắng (2008), *Bài tập thống kê*, Nhà Xuất bản Giáo dục.

[7] Jay L. Devore (2010), *Probability and Statistics for Engineering and the Sciences*, 8th Edition, Brooks/Cole.

[8] Shelldon M. Cross (2004), *Introduction to Probability and Statistics for engineers and scientists*, Elsevier Academic Press.

[9] Douglas C. Montgomery; George C. Runger (2007), *Applied Statistics and Probability for Engineers*(4th Edition), John Wiley and Sons, Inc.

[10] Welham,S.J., Gezan, S.A., Clark, S.J., Mead, A. (2013), *Statistical Methods in Biology: Design and Analysis of Experiments and Regression*, JChapman and Hall/CRC.

Giáo trình

XÁC SUẤT THỐNG KÊ

Chịu trách nhiệm xuất bản:
Giám đốc: TRẦN CHÍ ĐẠT
Chịu trách nhiệm nội dung:
Phó Giám đốc: NGÔ THỊ MỸ HẠNH

Biên tập : NGUYỄN TIẾN SỸ
LÊ HỒ DIỆU THẢO
Trình bày sách : LÊ HỒ DIỆU THẢO
Sửa bản in : LÊ HỒ DIỆU THẢO
Thiết kế bìa : TRẦN HỒNG MINH

NHÀ XUẤT BẢN THÔNG TIN VÀ TRUYỀN THÔNG
Website: www.nxbthongtintruyenthong.vn
Trụ sở chính: Tầng 6, 115 Trần Duy Hưng, Q. Cầu Giấy, TP. Hà Nội
Điện thoại: 04.35772138, 04.35772139 Fax: 04.35579858
E-mail: nxb.tttt@mic.gov.vn
Chi nhánh TP. Hồ Chí Minh: Số 8A đường D2, P. 25, Q. Bình Thạnh, TP. Hồ Chí Minh
Điện thoại: 08.35127750, 08.35127751 Fax: 08.35127751
E-mail: cnsg.nxbtttt@mic.gov.vn
Chi nhánh TP. Đà Nẵng: Lô C1 đường Trần Hưng Đạo, Q. Sơn Trà, TP. Đà Nẵng
Điện thoại: 0511.3897467 Fax: 0511.3843359
E-mail: cndn.nxbtttt@mic.gov.vn
Chi nhánh Tây Nguyên: Số 28B đường Y Bih Alêô, TP. Buôn Ma Thuột, tỉnh Đắk Lắk
Điện thoại: 0500.3808088 Email: cntn.nxbtttt@mic.gov.vn

Mã số: GD 20 ĐM 16 **Mã ISBN: 978-604-80-1930-3**

In 500 bản, khổ 16x24 cm tại Công ty TNHH In và Thương mại Hải Nam
Địa chỉ nơi in: Số 18 ngách 68/53/9 Quan Hoa, Q. Cầu Giấy, TP. Hà Nội
Số xác nhận đăng ký xuất bản: 2018 - 2016/CXBIPH/1 - 80/TTTT
Số quyết định xuất bản: 263/QĐ - NXB TTTT ngày 30 tháng 6 năm 2016.
In xong và nộp lưu chiểu quý IV năm 2016.