LÝ THYẾT XÁC SUẤT & THỐNG KÊ

10/2021

TÀI LIỆU LƯU HÀNH NỘI BỘ

Sưu tầm – Biên soạn

MỤC LỤC

PHẦN

I

XÁC SUẤT

CHƯƠNG 1 PHÉP ĐẾM

1. QUY TẮC NHÂN

Định nghĩa (Quy tắc nhân)
Giả sử để hoàn thành một công việc thì phải thực hiện k giai đoạn.

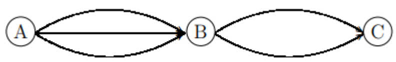
- ☑ Giai đoạn thứ nhất có n_1 cách thực hiện và
- ☑ Giai đoạn thứ hai có n_2 cách thực hiện và
- ☑ ...
- ☑ Giai đoạn thứ k có n_k cách thực hiện.

Khi đó, ta có

$$n = n_1 \cdot n_2 \cdots n_k$$

cách hoàn thành công việc.

VÍ DỤ 1.1. Giả sử để đi từ A đến C ta bắt buộc phải đi qua điểm B như hình bên dưới.



Nhận xét:

- ☑ Có 3 đường khác nhau đi từ A đến B và
- ☑ có 2 đường khác nhau đi từ B đến C.

Vậy có $n = 3 \cdot 2 = 6$ cách khác nhau để đi từ A đến C.

VÍ DỤ 1.2. Nếu một người có 6 đôi vớ khác nhau và 4 đôi giày khác nhau. Có bao nhiêu cách kết hợp giữa vớ và giày?
Nhận xét:

- ☑ Có 6 đôi vớ khác nhau và
- ☑ có 4 đôi giày khác nhau.

Vậy có $6 \cdot 4 = 24$ cách khác nhau để kết hợp giữa vớ và giày.

BÀI 1.1. Có bao nhiêu cách thiết kế cho một trang web gồm bốn màu, ba phông chữ và ba vị trí cho một hình ảnh?

4.3.3

BÀI 1.2. Quảng cáo trên web có thể được thiết kế từ bốn màu khác nhau, ba loại phông chữ, năm kích thước phông chữ, ba hình ảnh và năm cụm từ văn bản. Có thể thiết kế bao nhiêu mẫu khác nhau?

4 . 3 . 5 . 3 . 5

BÀI 1.3. Một ổ khóa có ba vòng khóa, mỗi vòng có 10 chữ số: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9. Hỏi có tất cả bao nhiêu mã khóa?

10 . 10 . 10

BÀI 1.4. Một thiết kế cho một máy tính có thể chỉ định bất kỳ một trong năm kích thước bộ nhớ, một trong ba loại màn hình, một trong bốn kích cỡ của một đĩa cứng và có thể bao gồm hoặc không bao gồm một cây bút điện tử. Có bao nhiêu hệ thống máy tính khác nhau có thể được thiết kế?

5 . 3 . 4 . 2

BÀI 1.5. Thiết kế mới cho một bể xử lý nước thải đã được đề xuất với ba hình dạng có thể, bốn kích thước có thể, ba vị trí cho van đầu vào và bốn vị trí cho van đầu ra. Có thể thiết kế bao nhiêu sản phẩm khác nhau?

3 . 4 . 3 . 4

2. HOÁN VỊ

Định nghĩa (Hoán vị): Số hoán vị của n phần tử khác nhau là

$$n! = n \cdot (n-1) \cdots 2 \cdot 1.$$

Hệ quả: Số hoán vị của $n = n_1 + n_2 + \cdots + n_k$ phần tử gồm n_1 phần tử loại 1, n_2 phần tử loại 2,... và n_k phần tử loại k được tính

$$\frac{n!}{n_1! n_2! \ldots n_k!}$$

Quy ước: $0! = 1$.

VÍ DỤ 2.1. Mỗi cách xếp 4 học sinh vào một bàn có 4 chỗ ngồi là một hoán vị của 4 phần tử. Số cách xếp 4 học sinh vào một bàn có 4 chỗ ngồi là $4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24$.

VÍ DỤ 2.2. Từ các chữ số 0, 1, 2, 3, 4 có thể lập được bao nhiêu số gồm 5 chữ số phân biệt?

✍ **LỜI GIẢI.**
Gọi $\overline{a_1a_2a_3a_4a_5}$ với $a_1 \neq 0$ và a_1, a_2, a_3, a_4, a_5 phân biệt là số cần lập.

- ☑ Bước 1: chữ số $a_1 \neq 0$ nên có 4 cách chọn a_1.
- ☑ Bước 2: Sắp 4 chữ số còn lại vào 4 vị trí có $4! = 24$ cách.

Vậy có $4 \cdot 24 = 96$ số. □

3. CHỈNH HỢP

Định nghĩa (Chỉnh hợp): Số hoán vị của các tập con gồm k phần tử được chọn từ một tập hợp n phần tử khác nhau là

$$A_n^k = n \cdot (n-1) \cdot (n-2) \cdots (n-k+1) = \frac{n!}{(n-k)!}$$

VÍ DỤ 3.1. Sắp xếp 5 người vào một băng ghế có 7 chỗ. Hỏi có bao nhiêu cách?

✍ **LỜI GIẢI.**
Mỗi cách chọn ra 5 chỗ ngồi từ băng ghế để sắp 5 người vào và có hoán vị là một chỉnh hợp chập 5 của 7. Vậy tộng cộng, có $A_7^5 = \dfrac{7!}{(7-5)!} = 2520$ cách sắp. □

VÍ DỤ 3.2. Từ các chữ số 0, 1, 2, 3, 4, 5 có thể lập được bao nhiêu số gồm 4 chữ số phân biệt?

✍ **LỜI GIẢI.**
Gọi $\overline{a_1a_2a_3a_4}$ với $a_1 \neq 0$ và a_1, a_2, a_3, a_4 phân biệt là số cần lập.

- ☑ Bước 1: chữ số $a_1 \neq 0$ nên có 5 cách chọn a_1.
- ☑ Bước 2: Chọn 3 trong 5 chữ số còn lại để sắp vào 3 vị trí chính là chỉnh hợp chập 3 của 5 phần tử: $A_5^3 = \dfrac{5!}{(5-3)!} = 60$.

Vậy có $5 \cdot A_5^3 = 300$ số. □

4. TỔ HỢP

4.Định nghĩa (Tổ hợp): Số tổ hợp, tập hợp con gồm k phần tử được chọn từ một tập hợp n phần tử được tính

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

VÍ DỤ 4.1. Có 10 cuốn sách toán khác nhau. Chọn ra 4 cuốn, hỏi có bao nhiêu cách?

✍ LỜI GIẢI.

Mỗi cách chọn ra 4 trong 10 cuốn sách là một tổ hợp chập 4 của 10. Vậy ta có

$$C_{10}^4 = \frac{10!}{4!(10-4)!} = 210 \text{ cách.}$$

□

BÀI 1.6. Trong một lớp gồm 30 sinh viên, cần chọn ra 3 sinh viên để làm lớp trưởng, lớp phó và thủ quỹ. Hỏi có bao nhiêu cách bầu chọn?

C_{30}^3

BÀI 1.7. Một hộp đựng 6 bi trắng và 4 bi đen.

a) Có tất cả bao nhiêu cách lấy ra 5 bi?

b) Có bao nhiêu cách lấy ra 5 bi trong đó có đúng 2 bi trắng?

a) C_{10}^5

b) $C_6^2 \cdot C_4^3$

BÀI 1.8. Một lô 140 chip bán dẫn được kiểm tra bằng cách chọn một mẫu 5 chip. Giả sử 10 trong số các chip này không phù hợp với yêu cầu của khách hàng.

a) Có bao nhiêu cách chọn mẫu khác nhau?

b) Có bao nhiêu mẫu trong số năm chip chứa chính xác một chip không phù hợp?

c) Có bao nhiêu mẫu trong số năm chip chứa ít nhất một chip không phù hợp?

a) C_{140}^5

b) $C_{10}^1 \cdot C_{130}^4$

c) 1 4 : $C_{10}^1 \cdot C_{130}^4$

2 3 : $C_{10}^2 \cdot C_{130}^3$ +

3 2 : $C_{10}^3 \cdot C_{130}^2$ +

4 1 : $C_{10}^4 \cdot C_{130}^1$ +

5 0 : C_{10}^5 +

BÀI 1.9. Mã vùng điện thoại là bộ gồm ba chữ số được sử dụng để đại diện cho một khu vực địa lý cụ thể.

a) Có bao nhiêu mã vùng có thể được tạo từ các chữ số từ 0 đến 9?

b) Như một phần (a), có bao nhiêu mã vùng có thể không bắt đầu bằng 0 hoặc 1, nhưng chứa 0 hoặc 1 làm chữ số giữa?

c) Có thể có bao nhiêu mã vùng trong đó không có chữ số nào xuất hiện nhiều hơn một lần trong đó?

a) $10 \cdot 10 \cdot 10$

b) 0 bắt đầu từ 0 và 1 : C_8^2
chữ giữa 0 or 1 : C_2^1 $\Rightarrow C_8^2 \cdot C_2^1 \cdot 10$
còn lại : 10

c) $10 \cdot 9 \cdot 8$

BÀI 1.10. Một thùng chứa 50 phần trong đó 5 phần bị lỗi. Một mẫu 10 phần được chọn ngẫu nhiên, không hoàn lại. Có bao nhiêu mẫu chứa ít nhất bốn bộ phận bị lỗi?

$C_5^4 \cdot C_{45}^6 + C_5^5 \cdot C_{45}^5$

BÀI 1.11. Trong một nhóm ứng viên gồm 7 nam và 3 nữ.

a) Có bao nhiêu cách thành lập một ủy ban bao gồm 3 người?

b) Có bao nhiêu cách thành lập một ủy ban bao gồm 3 người trong đó có đúng 1 nữ?

c) Có bao nhiêu cách thành lập một ủy ban bao gồm 3 người trong đó có ít nhất 1 nữ?

....

....

....

....

....

BÀI 1.12. Một hộp có 8 bi đỏ, 6 bi trắng, 4 bi vàng. Người ta chọn ra 6 bi từ hộp đó. Hỏi có bao nhiêu cách chọn nếu:

a) Không yêu cầu gì thêm.

b) Phải có 2 bi đỏ, 2 bi trắng, 2 bi vàng.

c) Có đúng 2 bi vàng.

....

....

....

....

....

....

BÀI 1.13. Một đồn cảnh sát khu vực có 9 người. Trong ngày cần cử 3 người làm nhiệm vụ ở địa điểm A, 2 người ở địa điểm B còn 4 người trực tại đồn. Hỏi có bao nhiêu cách phân công?

....

....

BÀI 1.14. Có 6 học sinh được sắp xếp ngồi vào 6 chỗ đã ghi số thứ tự trên một bàn dài. Tìm số cách xếp

a) 6 học sinh vào bàn.

b) 6 học sinh này vào bàn sao cho 2 học sinh A, B ngồi cạnh nhau.

c) 6 học sinh này ngồi vào bàn sao cho 2 học sinh A, B không ngồi cạnh nhau.

....

....

....

....

BÀI 1.15. Năm người A, B, C, D, E sẽ phát biểu trong một hội nghị. Có bao nhiêu cách sắp xếp để

a) B phát biểu sau A.

b) A phát biểu xong thì đến lượt B.

....

....

....

....

BÀI 1.16.

a) Có bao nhiêu cách xếp 3 nam và 3 nữ ngồi thành một hàng?

b) Có bao nhiêu cách xếp 3 nam và 3 nữ ngồi thành một hàng nếu mỗi nam và mỗi nữ ngồi cạnh nhau?

c) Có bao nhiêu cách xếp nếu 3 nam phải ngồi cạnh nhau?

d) Có bao nhiêu cách xếp nếu không có hai nam hoặc hai nữ nào được ngồi cạnh nhau?

....

....

....

....

....

BÀI 1.17. Từ 8 sinh viên nữ và 6 sinh viên nam, một nhóm làm việc gồm 3 nam và 3 nữ phải được lập ra. Có bao nhiêu cách lập nhóm nếu

a) 2 trong số các sinh viên nam không chịu làm việc cùng nhau?

b) 2 trong số các sinh viên nữ không chịu làm việc cùng nhau?

c) 1 nam và 1 nữ không chịu làm việc cùng nhau?

....

....

....

....

....

CHƯƠNG 2

XÁC SUẤT CƠ BẢN

1. MỘT SỐ KHÁI NIỆM

1.1. KHÔNG GIAN MẪU, BIẾN CỐ

- ☑ Một thí nghiệm có thể dẫn đến các kết quả khác nhau, mặc dù nó được lặp lại theo cùng một cách thức trong mỗi lần thực hiện, được gọi là một *thí nghiệm ngẫu nhiên*.
- ☑ *Không gian mẫu* Ω của một hiện tượng ngẫu nhiên là tập hợp tất cả các khả năng có thể xảy ra.
- ☑ *Biến cố* là một khả năng hay một tập hợp các khả năng của một hiện tượng ngẫu nhiên. Nói cách khác, biến cố là một tập con A của không gian mẫu Ω và kí hiệu là: $A \subset \Omega$.

VÍ DỤ 1.1. Gieo một đồng xu cân đối một lần. Gọi S là kết quả đồng xu "sấp" và N là kết quả đồng xu "ngửa". Không gian mẫu là

$$\Omega = \{S, N\}$$

Ta gọi tập con $A = \{S\} \subset \Omega$ là biến cố "đồng xu là sấp".

VÍ DỤ 1.2. Tung một con xúc xắc. Không gian mẫu là

$$\Omega = \{1; 2; 3; 4; 5; 6\}$$

Ta gọi tập con $B = \{1; 3; 5\} \subset \Omega$ là biến cố "kết quả tung xúc xắc là mặt lẻ".

VÍ DỤ 1.3. Hãy chỉ rõ không gian mẫu trong mỗi hiện tượng ngẫu nhiên sau:

a) Thảy một đồng xu.

....

b) Để một cây bút chì rơi tự do vào một tờ giấy có ghi những chữ số một cách ngẫu nhiên, sau đó ghi lại số có dấu chấm của đầu bút chì.

....

c) Thảy một đồng xu 4 lần rồi ghi lại chuỗi kết quả. Hãy liệt kê không gian mẫu.

....

....

....

....

....

....

Nếu ta chỉ quan tâm tới số lượng mặt ngửa trong chuỗi kết quả. thì cho biết không gian mẫu lúc này là

....

....

d) Bạn là nhà thiết kế trang web và bạn thiết lập một trang với 5 liên kết khác nhau. Người dùng có thể nhấp vào một trong các liên kết hoặc họ có thể rời khỏi trang web. Mô tả không gian mẫu cho kết quả của khách truy cập vào trang web của bạn.

....

....

VÍ DỤ 1.4. Thảy một đồng xu 4 lần rồi ghi lại chuỗi kết quả. Hãy mô tả biến cố A là chuỗi kết quả có chính xác 2 mặt ngửa xuất hiện.

....

1.2. CÁC QUY TẮC XÁC SUẤT

- ☑ Xác suất $\mathbb{P}(A)$ của biến cố A thỏa $0 \leq \mathbb{P}(A) \leq 1$.
- ☑ Nếu Ω là không gian mẫu của mô hình xác suất thì $\mathbb{P}(\Omega) = 1$.
- ☑ Hai *Biến cố rời nhau* khi chúng không có khả năng chung và không bao giờ xảy ra cùng nhau. Khi hai biến cố A, B rời nhau thì

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B).$$

Đây là *quy tắc cộng cho hai biến cố rời nhau*.

- ☑ *Biến cố đối* A^c của biến cố A là những khả năng biến cố A không xảy ra. Quy tắc đối là

$$\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$$

- ☑ Hai *biến cố độc lập* nhau khi biết rằng xác suất xảy ra biến cố này không làm thay đổi xác suất xảy ra biến cố kia, khi đó

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B).$$

Đây là *quy tắc nhân của hai biến cố độc lập*.

Định lý (Xác suất trong không gian hữu hạn)
Khi ta thực hiện phép gán xác suất cho mỗi khả năng riêng lẻ với một số nằm giữa 0 và 1, đồng

thời có tổng là 1. Xác suất của một biến cố bất kì là tổng các xác suất của mỗi khả năng xảy ra trong biến cố đó. Khi đó, ta có được một mô hình xác suất trong không gian mẫu hữu hạn.
Chú ý: khi ta gán xác suất cho mỗi khả năng bằng nhau thì ta gọi đó là mô hình xác suất đồng khả năng.

BÀI 2.1. Chọn ngẫu nhiên một số tự nhiên nhỏ hơn 10.

a) Hãy mô tả không gian mẫu của phép thử ngẫu nhiên trên.

b) Tính xác suất số được chọn là số không bé hơn 5.

c) Tính xác suất số được chọn là số 3.

....

....

....

....

....

....

....

....

BÀI 2.2. Nếu ta tung một con xúc xắc cân bằng thì xác suất để được mặt chẵn hoặc lớn hơn 4 chấm là bao nhiêu?

....

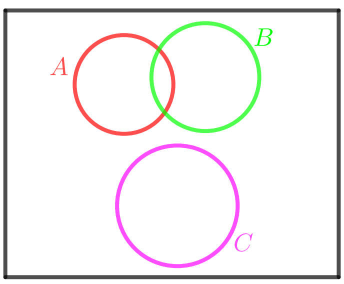
....

....

....

2. BIỂU ĐỒ VENN

Biểu đồ Venn (sơ đồ Venn hay sơ đồ tập hợp) là một sơ đồ cho thấy tất cả các mối quan hệ logic có thể có giữa một số lượng hữu hạn các tập hợp.
Trong biểu đồ Venn, người ta dùng những hình giới hạn bởi một đường khép kín (đường tròn, elip,...) để biểu diễn tập hợp.



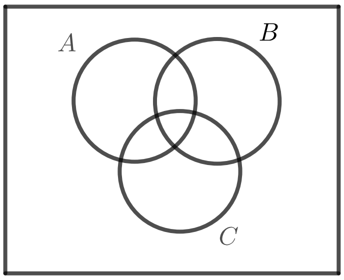
BÀI 2.3. Sử dụng biểu đồ Veen, mô tả các khái niệm: hai biến cố rời nhau và hai biến cố đối nhau.

....

....

....

BÀI 2.4. Cho ba biến cố được biểu diễn bởi biểu đồ Veen như sau:



Vẽ lại hình trên rồi tô đậm những vùng tương ứng với biến cố sau:

a) A^c b) $A \cup B$ c) $(A \cup B) \cap C$

d) $(B \cap C)^c$ e) $(A \cup B)^c \cap C$

....

....

....

....

....

....

....

....

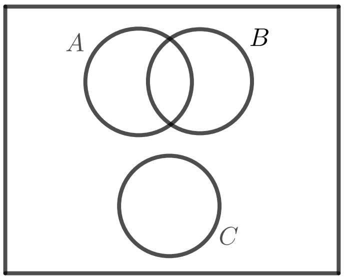
....

....

....

....

BÀI 2.5. Cho ba biến cố được biểu diễn bởi biểu đồ Veen như sau:



Vẽ lại hình trên rồi tô đậm những vùng tương ứng với biến cố sau:

a) A^c

b) $(A \cup B) \cap (A \cup B^c)$

c) $(A \cup B) \cap C$

d) $(B \cap C)^c$

e) $(A \cup B)^c \cap C$

....

....

....

....

....

....

....

....

....

....

....

....

BÀI 2.6. Xác suất xảy ra của 5 khả năng trong một thí nghiệm ngẫu nhiên là như nhau. Không gian mẫu $\Omega = \{a, b, c, d, e\}$. Đặt các biến cố $A = \{a, b\}$ và $B = \{c, d, e\}$. Tính:

a) $\mathbb{P}(A)$

b) $\mathbb{P}(B)$

c) $\mathbb{P}(A^c)$

d) $\mathbb{P}(A \cap B)$

e) $\mathbb{P}(A \cup B)$

....

....

....

....

....

....

BÀI 2.7. Không gian mẫu của một thí nghiệm ngẫu nhiên là $\Omega = \{a, b, c, d, e\}$ với xác suất tương ứng 0.1, 0.1, 0.2, 0.4 và 0.2. Đặt các biến cố $A = \{a, b, c\}$ và $B = \{c, d, e\}$. Tính:

a) $\mathbb{P}(A)$

b) $\mathbb{P}(B)$

c) $\mathbb{P}(A^c)$

d) $\mathbb{P}(A \cap B)$

e) $\mathbb{P}(A \cup B)$

3. CÔNG THỨC CỘNG HAI BIẾN CỐ

Công thức cộng hai biến cố A, B tổng quát là

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

Đặc biệt, nếu A, B **rời nhau**, tức là $\mathbb{P}(A \cap B) = 0$, thì

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B).$$

Định lý De Morgan
Cho A và B là hai tập hợp bất kỳ thì

1. $(A \cap B)^c = A^c \cup B^c$.
2. $(A \cup B)^c = A^c \cap B^c$.

BÀI 2.8. Cho $\mathbb{P}(A) = \dfrac{1}{3}$, $\mathbb{P}(B) = \dfrac{1}{2}$ và $\mathbb{P}(A \text{ hoặc } B) = \dfrac{3}{4}$

a) $\mathbb{P}(A \text{ và } B)$
b) $\mathbb{P}(A^c \text{ và } B^c)$
c) $\mathbb{P}(A^c \text{ hoặc } B^c)$
d) $\mathbb{P}(A^c \text{ và } B)$
e) $\mathbb{P}(A \text{ và } B^c)$

BÀI 2.9. Cho $\mathbb{P}(A) = 0.3$, $\mathbb{P}(B) = 0.2$ và $\mathbb{P}(A \cap B) = 0.1$. Tính các xác suất sau

a) $\mathbb{P}(A^c)$

b) $\mathbb{P}(A \cup B)$

c) $\mathbb{P}(A^c \cap B)$

d) $\mathbb{P}(A \cap B^c)$

e) $\mathbb{P}(A \cup B^c)$

f) $\mathbb{P}(A^c \cup B)$

....

....

....

....

....

....

....

....

....

....

BÀI 2.10. Ta xem như một hộ gia đình giàu có nếu thu nhập của họ vượt quá $100,000\$$ và xem hộ gia đình trí thức nếu chủ nhà hoàn thành bậc đại học. Chọn ngẫu nhiên một hộ gia đình người Mỹ và xét A là biến cố mà hộ gia đình được chọn là giàu có và B là biến cố gia đình trí thức. Theo khảo sát dân số hiện tại, người ta thấy xác suất $\mathbb{P}(A) = 0.138$; $\mathbb{P}(B) = 0.261$ và xác suất một gia đình vừa giàu có vừa trí thức là $\mathbb{P}(A \text{ và } B) = 0.082$.

a) Hãy tính xác suất chọn một gia đình hoặc là giàu có hoặc là trí thức.

b) Vẽ biểu đồ Venn biểu diễn mối quan hệ giữa hai biến cố A và B.

c) Biễu diễn qua biểu đồ Venn và tính những xác suất sau:

   i. $\{A \text{ và } B\}$

   ii. $\{A^c \text{ và } B\}$

   iii. $\{A \text{ và } B^c\}$

   iv. $\{A^c \text{ và } B^c\}$

....

....

....

....

....

....

....

....

....

....

BÀI 2.11. Tỷ lệ người mắc bệnh tim trong một vùng dân cư là 9%, mắc bệnh huyết áp là 12% và mắc cả hai bệnh là 7%. Chọn ngẫu nhiên một người trong vùng. Tính xác suất để người đó

a) Bị bệnh tim hay bị bệnh huyết áp.

b) Không bị bệnh tim cũng không bị bệnh huyết áp.

c) Không bị bệnh tim hay không bị bệnh huyết áp.

d) Bị bệnh tim nhưng không bị bệnh huyết áp.

e) Không bị bệnh tim nhưng bị bệnh huyết áp.

....

....

....

....

....

....

....

....

BÀI 2.12. Theo *Consumer Digest* (Tháng 7/8 1996), vị trí của các máy tính để bàn (PC) trong nhà là như sau

Phòng ngủ người lớn:	0.03
Phòng ngủ trẻ em:	0.15
Phòng ngủ khác:	0.14
Phòng làm việc:	0.40
Các phòng khác:	0.28

a) Hỏi xác suất để PC trong phòng ngủ là bao nhiêu?

b) Hỏi xác suất để PC không ở trong phòng ngủ là bao nhiêu?

c) Giả sử một căn hộ được chọn ngẫu nhiên từ các căn hộ có PC; hỏi bạn kỳ vọng sẽ thấy PC trong phòng nào?

BÀI 2.13. Một số tiểu bang đang xem xét luật sẽ cấm sử dụng điện thoại di động trong khi lái xe vì họ tin rằng lệnh cấm sẽ giảm tai nạn xe hơi liên quan đến điện thoại. Một nghiên cứu phân loại các loại tai nạn này vào các ngày trong tuần khi chúng xảy ra. Trong ví dụ này, ta sử dụng các giá trị từ nghiên cứu này làm mô hình xác suất. Dưới đây là xác suất:

Thứ	Hai	Ba	Tư	Năm	Sáu	Bảy	Chủ nhật
Xác suất	0.19	0.18	0.23	0.19	0.16	0.02	0.03

Hãy kiểm tra các quy tắc xác suất của mô hình trên. Sau đó, áp dụng các quy tắc xác suất, tính các trường hợp sau

a) Xác suất xảy ra tai nạn vào ngày nghỉ cuối tuần (thứ bảy và chủ nhật).

b) Xác suất xảy ra tai nạn vào ngày trong tuần.

....

....

....

BÀI 2.14 (Phân bố các loại máu). Máu người có thể là một trong các nhóm: O, A, B hoặc AB nhưng phân bố các loại khác nhau giữa các nhóm ở người. Bảng sau là sự phân bố các loại máu cho một người được chọn ngẫu nhiên tại Hoa Kỳ:

Loại máu	A	B	AB	O
Xác suất	0.40	0.11	0.04	?

a) Tính xác suất của nhóm máu O ở Hoa Kỳ.

b) Maria có máu loại B. Cô ấy có thể được truyền máu một cách an toàn từ những người có nhóm máu O và B. Xác suất khi chọn ngẫu nhiên một người Mỹ có thể hiến máu cho Maria?

....

....

....

....

BÀI 2.15. Số giả mạo trong bản khai thuế, hồ sơ thanh toán, hóa đơn, xác nhận quyền sở hữu tài khoản và nhiều loại giấy tờ khác thường có ở những mẫu không có trong hồ sơ hợp lệ. Một số mẫu qua mặt quản lí dễ dàng bởi một kẻ lừa đảo thông minh. Tuy nhiên, có một nghiên cứu chuyên sâu đã chỉ ra rằng các chữ số đầu tiên của các con số trong hồ sơ hợp pháp thường theo một phân phối được gọi là quy luật Benford.

Số đầu tiên	1	2	3	4	5	6	7	8	9
Xác suất	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046

Chú ý số 0 không thể đứng đầu tiên.

a) Xét những biến cố: $A = \{$ chữ số 1 đứng đầu $\}$ và $B = \{$ chữ số đầu là 6 hoặc lớn hơn $\}$. Hãy tính xác suất của mỗi biến cố.

b) Tính xác suất khi chữ số đầu tiên lớn hơn 1.

c) Sử dụng xác suất biến cố A và B, tính xác suất khi số đầu tiên là 1 hoặc 6 hoặc lớn hơn.

d) Tính xác suất biến cố C chữ số đầu tiên là số lẻ. Sau đó suy ra xác suất P(B hoặc C) và chứng minh nhỏ hơn tổng xác suất 2 biến cố B,C. Giải thích.

BÀI 2.16. Giả sử rằng xác suất để một thiết bị điện tử hoạt động trên 6000 giờ là 0.42. Giả sử rằng xác suất thiết bị hoạt động không quá 4000 giờ là 0.04.

a) Hỏi xác suất để tuổi thọ của thiết bị nhỏ hơn hoặc bằng 6000 giờ là bao nhiêu?

b) Hỏi xác suất để tuổi thọ lớn hơn 4000 giờ?

4. CÔNG THỨC NHÂN HAI BIẾN CỐ

Định nghĩa
Hai biến cố độc lập nhau khi biết rằng xác suất xảy ra biến cố này không làm thay đổi xác suất xảy ra của biến cố kia và ngược lại, khi đó

$$\mathbb{P}(B \text{ và } A) = \mathbb{P}(A).\mathbb{P}(B).$$

BÀI 2.17. Hai bố con bạn An đi hội chợ, chơi trò ném banh vào lon đặt trong ô trống, mỗi người ném 1 trái. Xác suất ném trúng lon của bố An và An lần lượt là là 0,85 và 0,43. Tính xác suất để cả hai bố con bạn An ném trúng lon.

....

....

....

BÀI 2.18. Ở hiệp phụ tại vòng chung kết giải mùa hè, hai xạ thủ A, B thi đấu với nhau, bắn mỗi người một mũi tên. Xác suất bắn trúng của A, B lần lượt là 0,86 và 0,83.

a) Tính xác suất để cả hai bắn trúng.

b) Tính xác suất để A bắn trượt và B bắn trúng.

....

....

....

....

....

BÀI 2.19. Một nhóm sinh viên thi SAT gồm 3 thí sinh A, B, C. Xác suất thi đậu của các thí sinh A, B, C lần lượt là 0,9; 0,85 ; 0,95. Tính xác suất nhóm này không thi đậu được cả ba thí sinh?

....

....

....

BÀI 2.20. 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 bị hỏng thì thiết bị ngừng hoạt động. Xác suất để cụm thứ nhất bị hỏng trong ngày là 0.1, cụm thứ hai là 0.05 và cụm thứ ba là 0.15. Tìm xác suất để thiết bị không ngừng hoạt động trong ngày. (Đs: 0.7267)

5. CÔNG THỨC XÁC SUẤT CÓ ĐIỀU KIỆN

Định nghĩa

Khi cần tính xác suất của biến cố B mà ta đã biết trước thông tin của biến cố A, ta sẽ sử dụng khái niệm *xác suất có điều kiện* $\mathbb{P}(B|A)$ theo công thức

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

với $\mathbb{P}(A) > 0$.

BÀI 2.21. Đĩa nhựa polycarbonate từ một nhà cung cấp được phân tích về khả năng chống trầy xước và sốc. Kết quả từ 100 đĩa được tóm tắt như sau

		Chống sốc	
		Cao	Thấp
Chống xước	Cao	70	9
	Thấp	16	5

Đặt A là biến cố một đĩa có khả năng chống sốc cao, và B là biến cố đĩa có khả năng chống xước cao. Xác định xác suất sau

a. $\mathbb{P}(A)$ b. $\mathbb{P}(B)$ c. $\mathbb{P}(A|B)$ d. $\mathbb{P}(B|A)$

....

....

....

....

BÀI 2.22. Bảng sau đây tóm tắt phân tích các mẫu thép mạ kẽm cho trọng lượng lớp phủ và độ nhám bề mặt

		Trọng lượng lớp phủ	
		Cao	Thấp
Độ nhám bề mặt	Cao	12	16
	Thấp	88	34

a. Nếu trọng lượng lớp phủ của mẫu cao, xác suất độ nhám bề mặt cao là bao nhiêu?

b. Nếu độ nhám bề mặt của mẫu cao, xác suất trọng lượng lớp phủ cao là bao nhiêu?

c. Nếu độ nhám bề mặt của mẫu thấp, xác suất trọng lượng lớp phủ thấp là bao nhiêu?

....

....

....

....

BÀI 2.23. Trong kì thi cuối kì của ĐH New Harmony có 10000 kết quả thi của 3 khoa chính: khoa nghệ thuật, khoa kỹ thuật và vật lí, khoa sức khỏe được thống kê trong bảng sau

Khoa	Điểm A	Điểm B	Điểm dưới B	Tổng
Nghệ thuật	2142	1890	2268	6300
Kỹ thuật và vật lí	368	432	800	1600
Sức khỏe	882	630	588	2100

Tính xác suất:

a) Lấy ngẫu nhiên được 1 điểm loại dưới B.

b) Lấy ngẫu nhiên được 1 điểm loại dưới B với thông tin điểm đó lấy từ khoa kĩ thuật.

c) Xác suất là bao nhiêu để điểm lấy ra từ khoa sức khỏe.

d) Lấy ngẫu nhiên được 1 điểm A.

e) Lấy ngẫu nhiên được 1 điểm A với thông tin điểm đó lấy từ khoa sức khỏe.

f) Hãy tính xác suất lấy được một điểm loại A từ trường ĐH New Harmony khoa nghệ thuật bằng 2 cách: số lượng trong bảng và công thức xác suất điều kiện. So sánh kết quả.
Hãy giải thích sự khác nhau ở đáp số câu a với câu b; câu d với câu e.

BÀI 2.24. Trong một nhóm sinh viên đại học, người ta phân loại theo giới tính và mức độ thường xuyên uống rượu bia hay không. Dưới đây là xác suất

	Nam	Nữ
Thường xuyên	0.11	0.12
Không thường xuyên	0.32	0.45

Kiểm tra xem bảng xác suất trên có tổng là 1 không? Hãy tính xác suất chọn ngẫu nhiên

a) Một người không thường uống rượu bia.

b) Một người nam sinh viên không thường uống rượu bia. So sánh với kết quả trên.

c) Một người nam sinh viên thường uống rượu bia; một người nữ sinh viên thường uống rượu bia.

d) Xác suất một sinh viên thường uống rươu bia với điều kiện phải là sinh viên nam. Xác suất chọn một sinh viên thường uống rươu bia với điều kiện phải là sinh viên nữ.

....

....

....

....

....

....

BÀI 2.25. Theo dõi dự báo thời tiết trên đài truyền hình (nắng, sương mù, mưa) và so sánh với thời tiết thực tế xảy ra, ta có bảng thống kê sau

Thực tế	Dự báo		
	nắng	sương mù	mưa
nắng	30	5	5
sương mù	4	20	2
mưa	10	4	20

nghĩa là có 30 lần dự báo nắng, trời nắng, 4 lần dự báo nắng, trời sương mù; 10 lần dự báo nắng, trời mưa,...

a) Tính xác suất dự báo trời nắng của đài truyền hình.

b) Tính xác suất dự báo của đài truyền hình là đúng thực tế.

c) Được tin dự báo là trời nắng, tính xác suất để thực tế thì trời mưa? trời sương mù? trời nắng?

BÀI 2.26. Xác suất để một chiếc ôtô đang được đổ xăng cũng cần thay dầu là 0,25; xác suất nó cần một bộ lọc dầu mới là 0,40; xác suất cần thay cả dầu và bộ lọc là 0,14

a) Nếu dầu đã được thay, thì xác suất để cần thay một bộ lọc dầu mới là bao nhiêu?

b) Nếu một lọc dầu mới đã thay, thì xác suất để dầu phải được thay là bao nhiêu?

BÀI 2.27. Xác suất để một người đàn ông có vợ xem một chương trình ti vi T là 0,4, xác suất để một người phụ nữ có chồng xem chương trình T là 0,5. Xác suất để một người đàn ông xem chương trình T, biết rằng vợ ông ta cũng xem là 0,7. Tìm xác suất để

a) một cặp vợ chồng xem chương trình T;

b) người vợ xem chương trình T, biết rằng chồng cô ấy cũng xem;

c) ít nhất một trong hai vợ chồng sẽ xem chương trình T.

6. CÔNG THỨC XÁC SUẤT TOÀN PHẦN

Định nghĩa (Hệ đầy đủ các biến cố).
Hệ các biến cố A_i, $(i = 1, ..., n)$ gọi là **hệ đầy đủ các biến cố** nếu thỏa mãn hai điều kiện sau

① A_1, A_2,...,A_n đôi một rời nhau, tức là $A_i \cap A_j = \emptyset$ với mọi $i \neq j$.

② $A_1 \cup A_2 \cup ... \cup A_n = \Omega$.

Định nghĩa (Công thức xác suất toàn phần).
Cho A_i, $(i = 1, ..., n)$ là hệ đầy đủ các biến cố với $\mathbb{P}(A_i) > 0$, $(i = 1, ..., n)$, B là một biến cố nào đó thì

$$\mathbb{P}(B) = \mathbb{P}(A_1)\mathbb{P}(B|A_1) + \cdots + \mathbb{P}(A_n)\mathbb{P}(B|A_n).$$

Công thức trên gọi là công thức xác suất toàn phần (đầy đủ).

....

....

....

....

....

....

....

....

....

....

....

....

7. CÔNG THỨC BAYES

Định lý (Công thức Bayes).
Cho A_i, $(i = 1, ..., n)$ là hệ đầy đủ các biến cố với $\mathbb{P}(A_i) > 0$, $(i = 1, ..., n)$, B là một biến cố nào đó sao cho $\mathbb{P}(B) > 0$. Khi đó với mọi i, $(i = 1, ..., n)$

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(A_i)\mathbb{P}(B|A_i)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A_i)\mathbb{P}(B|A_i)}{\mathbb{P}(A_1)\mathbb{P}(B|A_1) + \cdots + \mathbb{P}(A_n)\mathbb{P}(B|A_n)}.$$

BÀI 2.28 (Sử dụng biểu đồ cây để tính xác suất toàn phần). Một nhà máy có ba phân xưởng A; B; C tương ứng làm ra 25%; 35% và 40% tổng sản phẩm của nhà máy. Giả sử xác suất làm ra một sản phẩm hỏng của các phân xưởng A; B và C lần lượt là 0.01, 0.02 và 0.025. Hãy tính xác suất nhận được một sản phẩm hỏng.

....

....

....

....

....

....

....

....

BÀI 2.29. 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%, của nhà máy thứ hai là 40%. Tỷ lệ chính phẩm của nhà máy thứ nhất là 90%, của nhà máy thứ hai là 85%. Lấy ngẫu nhiên một chi tiết trên dây chuyền và thấy rằng nó tốt. Tìm xác suất để chi tiết đó do nhà máy thứ nhất sản xuất.

....

....

....

....

....

....

....

....

BÀI 2.30. Trong một vùng dân cư, cứ 100 người thì có 30 người hút thuốc lá. Biết tỷ lệ người bị viêm họng trong số người hút thuốc lá là 60%, trong số người không hút thuốc lá là 30%. Khám ngẫu nhiên một người và thấy người đó bị viêm họng. Tìm xác suất để người đó hút thuốc lá. Nếu người đó không bị viêm họng thì xác suất để người đó hút thuốc lá là bao nhiêu?

....

....

....

....

....

BÀI 2.31. Cảnh sát có kế hoạch thực thi việc giới hạn tốc độ bằng cách sử dụng các trạm radar tại bốn địa điểm khác nhau trong thành phố. Các trạm radar tại mỗi địa điểm D_1, D_2, D_3, D_4 sẽ được vận hành 40%, 30%, 20% và 30%. Nếu một người đang chạy quá tốc độ trên đường đi làm có các xác suất tương ứng 0,2; 0,1; 0,5 và 0,2 vượt qua các vị trí này.

a) Tính xác suất để người này sẽ nhận một vé phạt.

b) Nếu người này đã nhận một vé phạt tốc độ trên đường đi làm, hỏi xác suất để người này đi qua trạm radar đặt ở D_2 là bao nhiêu?

BÀI 2.32. Một chuỗi cửa hàng sơn sản xuất và bán sơn mủ và sơn bán bóng. Dựa trên doanh số bán hàng trong thời gian dài, xác suất để một khách hàng sẽ mua sơn mủ là 0,75. Trong số những người mua sơn mủ, 60% cũng mua con lăn. Nhưng chỉ 30% người mua sơn bán bóng mua con lăn. Một người mua được chọn ngẫu nhiên, người này mua con lăn và một hộp sơn, tính xác suất để sơn là sơn mủ.

BÀI 2.33. Một lớp học về vật lý gồm 10 sinh viên năm hai, 30 sinh viên năm ba và 10 sinh viên năm cuối. Điểm cuối kỳ cho thấy 3 sinh viên năm hai, 10 sinh viên năm ba và 5 sinh viên năm cuối nhận điểm A cho môn học. Nếu một sinh viên được chọn ngẫu nhiên từ lớp này và thấy rằng đã nhận điểm A, hỏi xác suất người này là sinh viên năm ba?

....

....

....

....

....

....

....

....

CHƯƠNG 3 BIẾN NGẪU NHIÊN RỜI RẠC

1. KHÁI NIỆM VÀ PHÂN LOẠI BIẾN NGẪU NHIÊN

Xét phép thử τ với không gian mẫu Ω. Giả sử, ứng với mỗi biến cố sơ cấp $\omega \in \Omega$, ta liên kết với một số thực $X(\omega) \in \mathbb{R}$, thì X được gọi là một **biến ngẫu nhiên**.

VÍ DỤ 1.1. Với trò chơi sấp ngửa bằng cách thảy đồng xu, giả sử nếu xuất hiện mặt sấp, ta được 1 đồng, nếu xuất hiện mặt ngửa, ta mất 1 đồng. Khi đó, ta có

- Phép thử τ: "*thảy đồng xu*",
- Không gian mẫu $\Omega = \{S, N\}$,
- Biến ngẫu nhiên X với $X(S) = 1$, $X(N) = -1$.

VÍ DỤ 1.2. Lấy ngẫu nhiên một sinh viên khoa Vật lý K18 và đo chiều cao. Khi đó, ta có

- Phép thử τ: "*lấy một sinh viên khoa Vật lý K18*",
- Không gian mẫu Ω là tập hợp tất cả các sinh viên khoa Vật lý K18,
- Biến ngẫu nhiên Y với $\omega \in \Omega$ thì $Y(\omega)$ là chiều cao của ω.

Tổng quát, biến ngẫu nhiên X của một phép thử τ với không gian mẫu Ω là một ánh xạ

$$\begin{array}{rccc} X: & \Omega & \to & \mathbb{R} \\ & \omega & \mapsto & X(\omega) \end{array}$$

- **Biến ngẫu nhiên** được gọi là **rời rạc** nếu tập hợp các giá trị mà nó có thể nhận $X(\Omega)$ là một tập hữu hạn hoặc vô hạn đếm được.
- **Biến ngẫu nhiên** được gọi là **liên tục** nếu tập hợp các giá trị mà nó có thể nhận $X(\Omega)$ là một khoảng dạng (a, b) hoặc toàn bộ $\mathbb{R}$.

Ở 2 ví dụ trên, ta thấy $X(\Omega) = \{-1; 1\}$, $Y(\Omega) = (145; 181)$, do đó X là biến ngẫu nhiên rời rạc còn Y là biến ngẫu nhiên liên tục.

BÀI 3.1. Trong mỗi tình huống dưới đây, biến ngẫu nhiên là liên tục hay rời rạc. Hãy cho biết lí do:

a) Trang web của bạn có năm liên kết khác nhau và người dùng có thể nhấp vào một trong các liên kết hoặc có thể rời khỏi trang. Bạn ghi lại khoảng thời gian người dùng bỏ ra trên trang web trước khi nhấp vào một trong các liên kết hoặc rời khỏi trang.

b) Số lần truy cập trên trang web của bạn.

c) Lượng khách truy cập hằng năm của trang web.

....

....

....

....

....

2. QUY LUẬT PHÂN PHỐI XÁC SUẤT

2.1. BẢNG PHÂN PHỐI XÁC SUẤT

Phân phối của biến ngẫu nhiên rời rạc X là một bảng gồm các giá trị và xác suất tương ứng của chúng. Cụ thể, cho biến ngẫu nhiên rời rạc X. Khi đó bảng phân phối xác suất cho X có dạng

X	x_1	x_2	...	x_n	...
$\mathbb{P}$	p_1	p_2	...	p_n	...

trong đó $p_i = \mathbb{P}(X = x_i)$ với mọi $i = 1, 2, \ldots$

2.2. HÀM XÁC SUẤT

Định nghĩa (Hàm xác suất): Cho biến ngẫu nhiên rời rạc X nhận hữu hạn giá trị x_1, ..., x_n, **hàm xác suất** (probability mass function) f thỏa mãn

i) $f(x_i) \geq 0$

ii) $\sum_{i=1}^{n} f(x_i) = 1$

iii) $f(x_i) = \mathbb{P}(X = x_i)$

2.3. HÀM PHÂN PHỐI XÁC SUẤT

Định nghĩa (Hàm phân phối xác suất): Hàm phân phối xác suất của biến ngẫu nhiên rời rạc X là:

$$F(a) = \mathbb{P}(X \leq a) = \sum_{x_i \leq a} f(x_i) \quad \text{với mọi số thực } a$$

Cụ thể hơn, trường hợp biến ngẫu nhiên rời rạc X

$$F(a) = \begin{cases} 0 & \text{nếu} \quad a < x_1 \\ p_1 & \text{nếu } x_1 \le a < x_2 \\ p_1 + p_2 & \text{nếu } x_2 \le a < x_3 \\ \dots & \dots \\ p_1 + \dots + p_k & \text{nếu } x_k \le a < x_{k+1} \\ \dots & \dots \end{cases}$$

Tính chất:
Với mọi $a < b$

$$\begin{aligned} \mathbb{P}(a < X \le b) &= \mathbb{P}(X \le b) - \mathbb{P}(X \le a) \\ &= F(b) - F(a). \end{aligned}$$

VÍ DỤ 2.1. Cho biến ngẫu nhiên X có bảng phân phối xác suất

X	1	2	3	4	5
$\mathbb{P}$	0,5	0,1	0,2	0,1	0,1

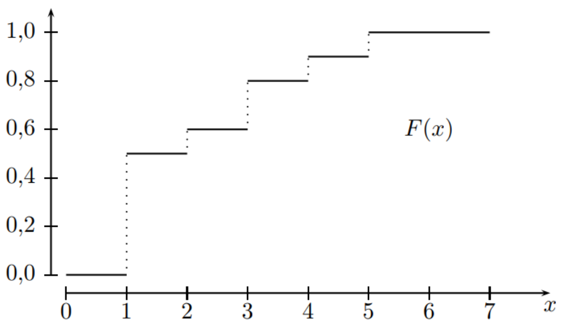
✍ **LỜI GIẢI.**

Hàm phân phối xác suất của biến ngẫu nhiên X là

$$F(x) = \begin{cases} 0 & \text{nếu} \quad x < 1 \\ 0{,}5 & \text{nếu } 1 \le x < 2 \\ 0{,}6 & \text{nếu } 2 \le x < 3 \\ 0{,}8 & \text{nếu } 3 \le x < 4 \\ 0{,}9 & \text{nếu } 4 \le x < 5 \\ 1 & \text{nếu } 5 \le x \end{cases}$$

Theo tính chất hàm phân phối xác suất, ta tính được:

$$\mathbb{P}(0{,}7 < X \le 3{,}27) = F(3{,}27) - F(0{,}7) = 0{,}8 - 0 = 0{,}8.$$



Hình 1.Hàm phân phối xác suất của biến ngẫu nhiên X.

□

VÍ DỤ 2.2. Cho biến ngẫu nhiên X có bảng phân phối xác suất

X	0	1	3	4
$\mathbb{P}$	0,2	0,1	0,4	0,3

Hãy xác định hàm phân phối xác suất $F(x)$ của X và vẽ đồ thị $F(x)$.

....

....

....

....

....

....

3. CÁC ĐẶC TRƯNG CỦA BIẾN NGẪU NHIÊN

3.1. TRUNG BÌNH (KỲ VỌNG)

Định nghĩa:
Trung trình hay kỳ vọng (mean, expected value) của biến ngẫu nhiên rời rạc X nhận các giá trị $\{x_1, ..., x_n\}$ tương ứng với xác suất $\{p_1, ..., p_n\}$, được ký hiệu là $\mathbb{E}(X)$ (hoặc μ_X) và xác định theo công thức:

$$\mathbb{E}(X) \equiv \mu_X = \bar{x} = \sum_{i=1}^{n} x_i p_i.$$

Tính chất: Cho X, Y là hai biến ngẫu nhiên bất kỳ và hằng số $a, b \in \mathbb{R}$ thì kỳ vọng của biến ngẫu nhiên có các tính chất sau

i) $\mathbb{E}(a) = c$.

ii) $\mathbb{E}(a + bX) = a + b\mathbb{E}(X)$.

iii) $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$.

iv) Nếu X và Y độc lập thì $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.

VÍ DỤ 3.1. Biến ngẫu nhiên rời rạc X nhận hai giá trị 0 và 1 có bảng phân phối xác suất như sau

X	0	1
$\mathbb{P}$	$\frac{1}{2}$	$\frac{1}{2}$

thì kỳ vọng của biến ngẫu nhiên X

$$\mathbb{E}(X) = 0.\frac{1}{2} + 1.\frac{1}{2} = \frac{1}{2}.$$

VÍ DỤ 3.2. Tung một con xúc sắc cân đối, gọi X là số chấm trên mặt xuất hiện thì phân phối xác suất của X là

X	1	2	3	4	5	6
$\mathbb{P}$	1/6	1/6	1/6	1/6	1/6	1/6

....

....

3.2. KỲ VỌNG HÀM CỦA BIẾN NGẪU NHIÊN

Mệnh đề: Cho g là hàm số thực bất kỳ, kỳ vọng của hàm g của biến ngẫu nhiên rời rạc X, ký hiệu là $\mathbb{E}(g(X))$ xác định theo công thức:

$$\mathbb{E}(g(X)) = \sum_i g(x_i)p_i.$$

Đặc biệt,

- Nếu $g(x) = x^r$, ta gọi $\mathbb{E}(g(X))$ là *moment bậc r*.
- Nếu $g(x) = e^{tx}$ thì ta gọi $\mathbb{E}(g(X))$ là *hàm sinh moment*.
- Nếu $g(x) = e^{-itx}$ thì ta gọi $\mathbb{E}(g(X))$ là *hàm đặc trưng* của biến ngẫu nhiên X.

VÍ DỤ 3.3. Biến ngẫu nhiên rời rạc X nhận hai giá trị 0 và 1 có bảng phân phối xác suất như sau

X	0	1
$\mathbb{P}$	$\frac{1}{2}$	$\frac{1}{2}$

Ta có kỳ vọng của biến ngẫu nhiên X^2 là

$$\mathbb{E}(X^2) = 0^2.\frac{1}{2} + 1^2.\frac{1}{2} = \frac{1}{2}.$$

VÍ DỤ 3.4. Tung một con xúc sắc cân đối, gọi X là số chấm trên mặt xuất hiện thì phân phối xác suất của X là

X	1	2	3	4	5	6
$\mathbb{P}$	1/6	1/6	1/6	1/6	1/6	1/6

Tìm $\mathbb{E}(X^2)$.

....

....

....

3.3. PHƯƠNG SAI

Định nghĩa:
Phương sai của biến ngẫu nhiên X được ký hiệu là $Var(X)$ (hoặc σ^2) và xác định theo công thức

$$Var(X) = \mathbb{E}\left(X - \mathbb{E}(X)\right)^2 = \mathbb{E}(X^2) - \left(\mathbb{E}(X)\right)^2.$$

Cụ thể: Nếu X nhận các giá trị $\{x_1, ..., x_n\}$ tương ứng với xác suất $\{p_1, ..., p_n\}$:

$$Var(X) \equiv \sigma_X^2 = \sum_{i=1}^{n}(x_i - \mu_X)^2 p_i = \sum_{i=1}^{n} x_i^2 p_i - \mu_X^2.$$

Tính chất: Cho X, Y là hai biến ngẫu nhiên bất kỳ và hằng số $c \in \mathbb{R}$ thì phương sai của biến ngẫu nhiên có các tính chất sau

i) $Var(c) = c$.

ii) $Var(cX) = c^2 Var(X)$.

iii) Nếu X và Y độc lập thì $Var(X \pm Y) = Var(X) + Var(Y)$.

iv) Nếu X và Y là hai biến ngẫu nhiên có hệ số tương quan ρ thì $Var(X \pm Y) = Var(X) + Var(Y) \pm \rho Var(X)Var(Y)$.

3.4. ĐỘ LỆCH CHUẨN

Độ lệch chuẩn của biến ngẫu nhiên X, ký hiệu $\sigma(X)$, là căn bậc hai của $Var(X)$.

$$\sigma(X) = \sqrt{Var(X)}.$$

VÍ DỤ 3.5. Biến ngẫu nhiên rời rạc X nhận hai giá trị 0 và 1 có bảng phân phối xác suất như sau

X	0	1
$\mathbb{P}$	$\frac{1}{2}$	$\frac{1}{2}$

Ta có $\mathbb{E}(X) = \frac{1}{2}$ và $\mathbb{E}(X^2) = \frac{1}{2}$. Do đó

Phương sai của X là $Var(X) \equiv \sigma_X^2 = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \frac{1}{2} - \left(\frac{1}{2}\right)^2 = \frac{1}{4}$.

Độ lệch tiêu chuẩn của X là $\sigma_X = \sqrt{Var(X)} = \sqrt{\frac{1}{4}} = \frac{1}{2}$.

VÍ DỤ 3.6. Tung một con xúc sắc cân đối, gọi X là số chấm trên mặt xuất hiện thì phân phối xác suất của X là

X	1	2	3	4	5	6
$\mathbb{P}$	1/6	1/6	1/6	1/6	1/6	1/6

Tìm phương sai và độ lệch tiêu chuẩn của X.

....

....

....

....

3.5. MODE

Mode của biến ngẫu nhiên X, ký hiệu là $Mode(X)$, là giá trị mà biến ngẫu nhiên X nhận được với xác suất lớn nhất. Cụ thể hơn, trường hợp rời rạc X nhận các giá trị $\{x_1, ..., x_n\}$ tương ứng với xác suất $\{p_1, ..., p_n\}$:

$$Mode(X) = x_i \Leftrightarrow p_i = \mathbb{P}(X = x_i) = \max\{p_1, p_2, ..., p_n\}.$$

VÍ DỤ 3.7. Tìm $Mode$ của biến ngẫu nhiên rời rạc X với bảng phân phối xác suất

X	1	2	3	4	5
$\mathbb{P}$	0,3	0,25	0,18	0,14	0,13

Dễ dàng nhận thấy, $Mode(X) = 1$.

VÍ DỤ 3.8. Tìm $Mode$ của biến ngẫu nhiên rời rạc X với bảng phân phối xác suất

X	1	2	3	4	5
$\mathbb{P}$	0,1	0,15	0,4	0,35	0,13

....

3.6. TRUNG VỊ

Trung vị của biến ngẫu nhiên X là số x thỏa điều kiện $\mathbb{P}(X \leq x) = \mathbb{P}(X \geq x) = \dfrac{1}{2}$. Trong trường hợp xác suất đồng khả năng, để tìm trung vị, người ta xếp tập $X(\Omega)$ tăng dần, sau đó nếu số phần tử của $X(\Omega)$ là số lẻ thì ta lấy số chính giữa. Ngược lại, ta sẽ lấy trung bình của hai số chính giữa nếu số phần từ của $X(\Omega)$ là số chẵn.

VÍ DỤ 3.9. Cho biến ngẫu nhiên X có bảng phân phối xác suất

X	1	2	3	4	5
$\mathbb{P}$	0,1	0,3	0,1	0,2	0,2

Tìm trung vị của X.

....

VÍ DỤ 3.10. Cho biến ngẫu nhiên X có bảng phân phối xác suất

X	1	2	3	4	5
$\mathbb{P}$	0,2	0,3	0,1	0,2	0,2

Tìm trung vị của X.

....

....

BÀI 3.2. Trò chơi đánh bài Texas bắt đầu với việc mỗi người chơi nhận được 2 là bài trên tay. Sau đây là bảng phân phối số lượng con bài át trong hai lá bài đó

Số lượng át	0	1	2
Xác suất	0.559	0.382	0.059

a) Hãy kiểm tra xem mô hình trên có phải là một biến ngẫu nhiên rời rạc?

b) Xác suất mà trong hai lá chứa ít nhất một át? Tính toán bằng hai cách khác nhau.

....

....

....

....

....

....

BÀI 3.3. Phần mềm kiểm tra chính tả bắt lỗi "lỗi không phải từ", là lỗi tạo thành bởi một chuỗi các chữ cái sắp xếp không tạo thành một từ, ví dụ chữ "the" được nhập là "teh". Khi sinh viên đại học được yêu cầu viết một bài luận 250 từ (không được kiểm tra lại lỗi chính tả), và X số lượng từ bị lỗi có phân phối sau:

Giá trị của X	0	1	2	3	4
Xác suất	0.1	0.3	0.3	0.2	0.1

a) Viết biến cố "có ít nhất 1 lỗi" theo biến ngẫu nhiên X. Xác suất của biến cố này là bao nhiêu?

b) Phát biểu thành lời biến cố $X \leq 2$. Tính xác suất biến cố đó và xác suất của biến cố $X < 2$.

....

....

....

....

....

....

BÀI 3.4. Không gian mẫu của một biến ngẫu nhiên là $\{a; b; c; d; e; f\}$ và các sự kiện sơ cấp có cùng khả năng xảy ra. Biến ngẫu nhiên đó được xác định như sau

Sự kiện sơ cấp	a	b	c	d	e	f
X	0	0	1,5	2,5	2	3

Viết hàm xác suất của X, sau đó sử dụng hàm xác suất vừa tìm được để tính các xác suất sau

a) $\mathbb{P}(X > 3)$.

b) $\mathbb{P}(0{,}5 < X < 2{,}7)$.

c) $\mathbb{P}(X = 0 \text{ hoặc } X = 2)$.

d) $\mathbb{P}(0 \leq X < 3)$.

BÀI 3.5. Cho hàm số $f(x) = \dfrac{8}{7}\left(\dfrac{1}{2}\right)^x$, $x = 1, 2, 3$.

a) Hàm số trên có phải là hàm xác suất không?

b) Tính các xác suất sau

i $\mathbb{P}(X > 1)$

ii $\mathbb{P}(2 < X < 6)$

iii $\mathbb{P}(X \leq 1 \text{ hoặc } X = 3)$

....

....

....

....

BÀI 3.6. Cho hàm số $f(x) = \dfrac{2x+1}{25}$, $x = 0, 1, 2, 3, 4$.

a) Hàm số trên có phải là hàm xác suất không?

b) Tính các xác suất sau

i $\mathbb{P}(X > 1)$

ii $\mathbb{P}(2 < X < 6)$

iii $\mathbb{P}(X \leq 1$ hoặc $X = 3)$

iv $\mathbb{P}(2 \leq X \leq 4.5)$

BÀI 3.7. Thả hai con súc sắc cân bằng với xác suất của mỗi mặt chấm xuất hiện là như nhau. Gọi biến ngẫu nhiên X là tổng số chấm xuất hiện khi thả hai con súc sắc.

a) Hãy viết không gian mẫu những khả năng khi tung 2 súc sắc.

b) Tính xác suất của mỗi khả năng.

c) Sử dụng kết quả câu b để lập bảng phân phối xác suất.

d) Người chơi sẽ thắng cược khi tổng súc sắc là 7 hoặc 11. Tính xác suất để người chơi thắng khi thả súc sắc.

....

....

....

....

....

....

....

....

BÀI 3.8. Đại học bang North Carolina đăng bản phân phối cấp lớp cho các khóa học trực tuyến. Học sinh trong học phần tiếng Anh 210 của học kỳ mùa xuân 2006 đã nhận được 31% A, 40% B, 20% C, 4% D và 5% F. Chọn ngẫu nhiên một học sinh Anh ngữ 210. "Chọn ngẫu nhiên" có nghĩa là mọi học sinh cùng một cơ hội được chọn. Điểm của học sinh theo thang điểm bốn điểm (với $A = 4$) là một biến ngẫu nhiên X.

a) Hãy lập bảng phân phối xác suất và tính xác suất học sinh được chọn B hoặc tốt hơn.

b) Giả sử khi đạt điểm D và F trong học phần tiếng Anh 210, học sinh đó bị coi như chưa hoàn thành chuyên ngành ngôn ngữ học. Hãy tính xác suất để học sinh chọn không thỏa yêu cầu của chuyên ngành.

BÀI 3.9. Sử dụng số liệu từ Bài tập 2.15, ta xét biến ngẫu nhiên V chính là giá trị của chữ số đầu tiên. Hãy tính trung bình và độ lệch tiêu chuẩn của V.

BÀI 3.10.

a) Cho biến ngẫu nhiên X có kì vọng $\mu_X = 10$ thì biến ngẫu nhiên $Y = 15 + 8X$ có kì vọng là bao nhiêu?

b) Cho biến ngẫu nhiên U có trung bình $\mu_U = 20$ và biến ngẫu nhiên V có trung bình $\mu_V = 20$ thì biến ngẫu nhiên $Z = 0.5U + 0.5V$ có trung bình là bao nhiêu?

BÀI 3.11. Linda là một nhân viên bán hàng tại một đại lý ô tô lớn. Với mức hoa hồng 25% lợi nhuận tính trên mỗi chiếc xe cô bán, Linda dự kiến sẽ kiếm được 350$ cho mỗi chiếc xe hơi bán được và 400$ cho mỗi chiếc xe tải hoặc SUV bán được. Linda thúc đẩy bản thân bằng cách sử dụng ước lượng xác suất cho doanh thu của mình. Vào một ngày Thứ Bảy của tháng Tư, cô ước tính doanh số bán xe hơi của mình như sau:

Số xe bán	0	1	2	3
Xác suất	0.3	0.4	0.2	0.1

và xe tải hoặc SUV như sau

Số xe bán	0	1	2
Xác suất	0.4	0.5	0.1

Đặt X là số xe hơi mà Linda bán và Y là số xe tải hoặc SUV mà Linda bán.

a. Hãy tính trung bình lượng xe mỗi loại cô ta bán từ đó suy ra doanh thu trung bình cô ta thu được.

b. Tính phương sai và độ lệch chuẩn cho biến X.

....

....

....

....

....

....

BÀI 3.12. Một nhà sản xuất ổ đĩa bán ra các thiết bị lưu trữ với dung lượng 1 terabyte, 500 gigabyte và 100 gigabyte với tỷ lệ tương ứng là 50%, 30% và 20%. Doanh thu ước tính trong năm tương ứng với từng thiết bị bán ra là 50 triệu \$, 25 triệu \$ và 10 triệu \$. Xác định hàm xác suất, hàm phân phối cho doanh thu của thiết bị lưu trữ bán ra trong năm đó. Tính doanh thu trung bình của thiết bị lưu trữ trong năm đó.

BÀI 3.13. Một kết cấu bao gồm 3 thành phần cơ khí hoạt động độc lập với nhau. Giả sử rằng xác suất mà thành phần thứ 1, thứ 2 và thứ 3 đáp ứng được các chi tiết kỹ thuật lần lượt là 0,95, 0,98 và 0,99. Tìm hàm xác suất, hàm phân phối cho số lượng các thành phần đáp ứng chi tiết kỹ thuật trong kết cấu.

BÀI 3.14. Miền giá trị của biến ngẫu nhiên Y là $\{0; 1; 2; 3; y\}$, trong đó y chưa biết và $\mathbb{P}(Y = a) = 0{,}2$, với $a = 0, 1, 2, 3, y$. Tìm y để giá trị trung bình của biến ngẫu nhiên Y là 6.

BÀI 3.15. Trong một hộp có hai đồng xu 10 cent và một đồng xu 50 cent. Chọn ngẫu nhiên hai đồng xu từ hộp. Gọi X là tổng giá trị của hai đồng xu.

a) X có thể nhận những giá trị nào?

b) Tìm hàm phân phối của X.

BÀI 3.16. Khi Sở y tế kiểm tra hai tạp chất thường thấy trong nước uống ở những giếng tư nhân trong một tỉnh. Kết quả kiểm tra cho thấy, có 20% các giếng có cả hai tạp chất, 40% có tạp chất A và 50% có tạp chất B. Nếu chọn ngẫu nhiên một giếng trong tỉnh, tìm phân phối xác suất cho Y, với Y là số tạp chất được tìm thấy trong giếng.

BÀI 3.17. Trong một bài kiểm tra đưa ra cho các đứa trẻ, có ba tấm hình về các động vật, yêu cầu các đứa trẻ nối một tấm hình đến một từ để nhật biết động vật đó. Nếu một đứa trẻ gán ba từ một cách ngẫu nhiên đến ba hình ảnh, tìm phân phối xác suất cho Y - số nhận diện đúng.

BÀI 3.18. Các thư đến trung tâm dịch vụ của một nhà sản xuất hệ thống thông tin đã được phân loại dựa trên số lượng từ khóa và loại tin nhắn: email hoặc voice. Giả sử, 70% tin nhắn đến qua email và phần còn lại là voice.

Số từ khóa	0	1	2	3	4
email	0,1	0,1	0,2	0,4	0,2
voice	0,3	0,4	0,2	0,1	0

Xác định hàm xác suất của số từ khóa trong một bài viết?

BÀI 3.19. Xác định hàm xác suất của biến ngẫu nhiên có hàm phân phối như sau

$$F(x) = \begin{cases} 0; & x < 2 \\ 0{,}2; & 2 \le x < 5{,}7 \\ 0{,}5; & 5{,}7 \le x < 6{,}5 \\ 0{,}8; & 6{,}5 \le x < 8{,}5 \\ 1; & 8{,}5 \le x. \end{cases}$$

4. MỘT SỐ PHÂN PHỐI THƯỜNG GẶP

4.1. PHÂN PHỐI NHỊ THỨC

- **Phép thử Bernoulli** là thí nghiệm (hay phép thử) chỉ có hai kết quả và thường được sử dụng để xây dựng một chuỗi các phép thử ngẫu nhiên.
- **Phép thử n Bernoulli** là một nhóm các phép thử Beronoulli thỏa 3 điều kiện: các phép thử trong chuỗi là độc lập; mỗi phép thử chỉ có 2 kết quả là “thành công” và “thất bại”; xác suất thành công của mỗi phép thử là p và không đổi.
- **Biến ngẫu nhiên nhị thức** X chính là số lần thành công của phép thử n Bernoulli tham số $0 < p < 1$, ký hiệu $X \sim B(n; p)$, với xác suất tương ứng

$$\mathbb{P}(X = x) = \mathrm{C}_n^x p^x (1-p)^{n-x}, \quad x = \overline{0, n}.$$

- **Tính chất:** Nếu X là biến ngẫu nhiên có phân phối nhị thức $B(n; p)$ thì

  i) $\mathbb{E}(X) = np$.

  ii) $Var(X) = npq$, trong đó $q = 1 - p$.

  iii) $np - q \le Mod(X) \le np - q + 1$, người ta còn gọi $Mod(X)$ là số lần xuất hiện chắc chắn nhất.

  iv) Với x, h là hai số nguyên dương thì

$$\mathbb{P}(x \le X \le x + h) = \mathbb{P}(X = x) + \mathbb{P}(X = x + 1) + ... + \mathbb{P}(X = x + h).$$

VÍ DỤ 4.1. Một bài thi trắc nghiệm có 10 câu hỏi, mỗi câu hỏi có bốn lựa chọn trong đó chỉ một lựa chọn đúng, giả sử ở mỗi câu hỏi thí sinh chọn câu trả lời một cách ngẫu nhiên. Đặt

$$X_i = \begin{cases} 1 & \text{nếu trả lời đúng câu } i \\ 0 & \text{nếu trả lời sai câu } i \end{cases}$$

thì X_i ($i = 1, ..., 10$) là các biến ngẫu nhiên Bernoulli với tham số $p = 1/4$. Gọi X là số câu thí sinh trả lời đúng thì

$$X = X_1 + ... + X_{10}$$

Khi đó $X \sim B(10; 1/4)$.

a) Tính xác suất để thí sinh trả lời đúng 5 câu hỏi.

b) Giả sử để đỗ kỳ thi này thí sinh phải trả lời đúng ít nhất 6 câu. Xác thí sinh thi đỗ là?

✍ **LỜI GIẢI.**

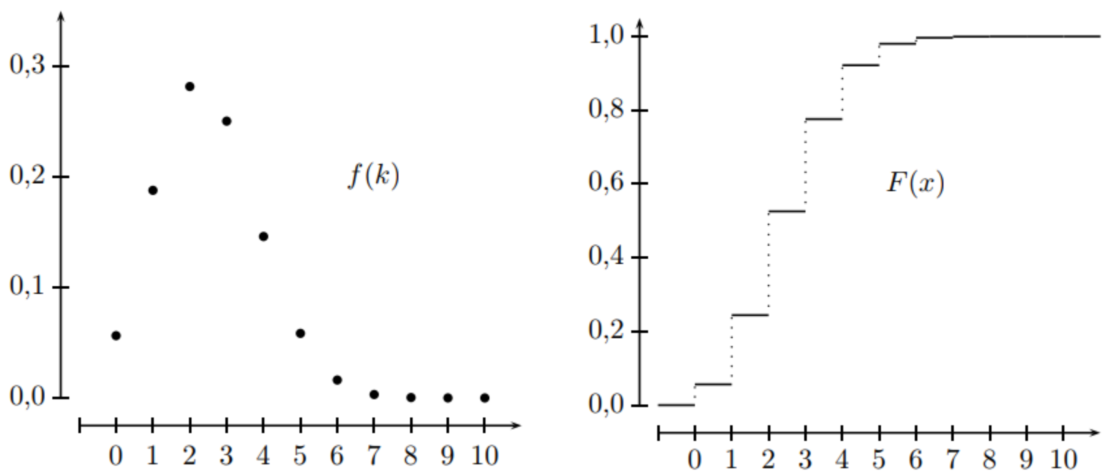
a) Xác suất để thí sinh trả lời đúng 5 câu hỏi là

$$\mathbb{P}(X = 5) = \mathrm{C}_{10}^{5} \left(\frac{1}{4}\right)^5 \left(1 - \frac{1}{4}\right)^{10-5} \approx 0.0584.$$

b) Giả sử để đỗ kỳ thi này thí sinh phải trả lời đúng ít nhất 6 câu. Xác suất thí sinh thi đỗ là

$$\begin{aligned}
\mathbb{P}(X \geq 6) &= \mathbb{P}(X = 6) + \cdots + \mathbb{P}(X = 10) \\
&= \mathrm{C}_{10}^{6} \left(\frac{1}{4}\right)^6 \left(1 - \frac{1}{4}\right)^{10-6} + \cdots + \mathrm{C}_{10}^{10} \left(\frac{1}{4}\right)^{10} \left(1 - \frac{1}{4}\right)^{10-10} \\
&= \sum_{x=6}^{10} \mathrm{C}_{10}^{x} \left(\frac{1}{4}\right)^x \left(1 - \frac{1}{4}\right)^{10-x} \\
&\approx 0.0197.
\end{aligned}$$

□



Hình 2.Hàm giá trị xác suất và hàm phân phối xác suất của biến ngẫu nhiên $X \sim B(10; 1/4)$.

VÍ DỤ 4.2. Cho biến ngẫu nhiên $X \sim B(9; 1/3)$, ta tìm kỳ vọng của biến ngẫu nhiên X theo công thức

$$\mathbb{E}(X) = np = 9 \cdot \frac{1}{3} = 3$$

và phương sai

$$Var(X) = npq = np(1-p) = 9 \cdot \frac{1}{3} \cdot \left(1 - \frac{1}{3}\right) = 9 \cdot \frac{1}{3} \cdot \frac{2}{3} = 2.$$

Số lần xuất hiện chắc chắn ($Mod(X)$) thỏa điều kiện

$$np - q \leq Mod(X) \leq np - q + 1,$$

thay số vào ta được

$$3 - \frac{2}{3} \leq Mod(X) \leq 3 + \frac{1}{3}$$

Mod của biến ngẫu nhiên $X \sim B(9; 1/3)$ là một số nguyên dương, vậy ta tìm được $Mod(X) = 3$.

BÀI 3.20. Trong mỗi tình huống dưới đây, có hợp lý khi sử dụng phân phối nhị thức cho biến ngẫu nhiên X không ? Đưa ra lý do cho câu trả lời của bạn trong mỗi trường hợp. Nếu là phân phối nhị thức, hãy cho các giá trị của n và p.

a) Một cuộc thăm dò ý kiến của 200 sinh viên đại học hỏi bạn có thường hay cáu kỉnh vào buổi sáng hay không. X là số người trả lời rằng họ thường dễ cáu kỉnh buổi sáng.

b) Bạn ném một đồng xu cân bằng cho đến khi mặt ngửa xuất hiện. X là số lần tung mà bạn thực hiện.

c) Hầu hết các cuộc gọi điện thoại khảo sát được thực hiện ngẫu nhiên và mẫu được coi là không thành công khi không nói chuyện trực tiếp với một người. Trong số các cuộc gọi đến thành phố New York, chỉ $1/12$ thành công. Cuộc khảo sát cuộc gọi 500 số được chọn ngẫu nhiên ở thành phố New York. X là số tiếp cận một người trực tiếp.

d) Một quy trình sản xuất hàng nghìn đầu dò nhiệt độ. Cho X biểu thị số đầu dò không phù hợp trong một mẫu có kích thước 30 được chọn ngẫu nhiên từ quá trình sản xuất trên.

e) Xét X biểu thị số vụ tai nạn xảy ra trên đường cao tốc liên bang ở Arizona trong thời gian một tháng.

....

....

....

....

....

BÀI 3.21. Biến ngẫu nhiên X có phân phối nhị thức với $n = 10$ và $p = 0.5$. Xác định xác suất:

a) $\mathbb{P}(X = 5)$

b) $\mathbb{P}(X \leq 2)$

c) $\mathbb{P}(X \geq 9)$

d) $\mathbb{P}(3 \leq X < 5)$

e) Hãy vẽ hàm phân phối tích lũy của mô hình trên.

....

....

....

....

....

....

....

....

BÀI 3.22. Biến ngẫu nhiên X có phân phối nhị thức với $n = 10$ và $p = 0.01$. Xác định xác suất:

a) $\mathbb{P}(X = 5)$

b) $\mathbb{P}(X \leq 2)$

c) $\mathbb{P}(X \geq 9)$

d) $\mathbb{P}(3 \leq X < 5)$

e) Hãy vẽ hàm phân phối tích lũy của mô hình trên.

BÀI 3.23. Một phân xưởng có 5 máy. Xác suất để trong một ca, mỗi máy bị hỏng là 0.1. Tìm xác suất để trong một ca, có đúng 2 máy bị hỏng. (Đs: 0.0729)

....

....

....

BÀI 3.24. Tính xác suất để gieo con súc sắc 10 lần, mặt một chấm xuất hiện không quá 3 lần. (Đs: 0.93)

BÀI 3.25. Giả sử tỷ lệ sinh con trai và con gái là bằng nhau và bằng 1/2. Một gia đình có 4 người con. Tính xác suất để 4 đứa con đó gồm

a) 2 trai và 2 gái. b) 1 trai và 3 gái. c) 4 trai.

....

....

....

....

....

BÀI 3.26. Tỷ lệ một loại bệnh bẩm sinh trong dân số là $p = 0.01$. Bệnh này cần sự chăm sóc đặc biệt lúc mới sinh. Một nhà bảo sinh thường có 20 ca sinh trong một tuần. Tính xác suất để

a) không có trường hợp nào cần chăm sóc đặc biệt;

b) có đúng một trường hợp cần chăm sóc đặc biệt;

c) có nhiều hơn một trường hợp cần chăm sóc đặc biệt.
(Đs: 0.8179; 0.1652; 0.0168)

BÀI 3.27. Các đường dây (lines) điện thoại đến hệ thống đặt vé máy bay bận chiếm 40% số lần gọi. Giả sử rằng việc các đường dây bận khi cuộc gọi đến là độc lập. Giả sử rằng 10 cuộc gọi được đặt cho hãng hàng không.

a) Xác suất để có đúng 3 cuộc gọi tới bị bận đường dây.

b) Xác suất để có ít nhất 1 cuộc gọi tới không bị bận.

c) Tính trung bình số cuộc gọi tới bị bận bằng hai cách.
(Đs: 0.215; 0.994; 4)

....

....

....

....

....

BÀI 3.28. Một đèn giao thông trên tuyến đường một người đi làm vào buổi sáng có màu xanh chiếm 20% số lần người ta tới ngã tư đó. Giả sử mỗi buổi sáng đi làm để thu thập số liệu là độc lập

a) Với 5 buổi sáng đi làm, hãy tính xác suất gặp đèn xanh đúng 1 ngày.

b) Với 20 buổi sáng đi làm, hãy tính xác suất gặp đèn xanh đúng 4 ngày.

c) Với 20 buổi sáng đi làm, hãy tính xác suất gặp đèn xanh nhiều hơn 4 ngày.
(Đs: 0.410; 0.218; 0.37)

BÀI 3.29. Bài kiểm tra trắc nghiệm chứa 25 câu hỏi, mỗi câu hỏi có bốn câu trả lời. Giả sử một học sinh chọn câu trả lời ở mỗi câu hỏi một cách ngẫu nhiên.

a) Xác suất để học sinh đó có nhiều hơn 20 câu trả lời đúng.

b) Xác suất để học sinh đó có ít hơn 5 câu trả lời đúng.

BÀI 3.30. Một bài thi trắc nghiệm gồm 12 câu hỏi, mỗi câu có 5 câu trả lời, trong đó chỉ có một câu đúng. Giả sử mỗi câu trả lời đúng, thí sinh được 4 điểm, mỗi câu trả lời sai, thí sinh bị trừ 1 điểm. Một thí sinh làm bài bằng cách chọn ngẫu nhiên các câu trả lời. Tìm xác suất để

a) Thí sinh được 13 điểm.

b) Thí sinh bị điểm âm.

(Đs: 0.0532; 0.558)

BÀI 3.31. Một người bắn bia với xác suất bắn trúng là $p = 0.7$

a) Bắn liên tiếp 3 phát. Tính xác suất có ít nhất 1 lần trúng bia.

b) Hỏi phải bắn ít nhất mấy lần để có xác suất ít nhất một lần trúng bia ≥ 0.9.

(Đs: 0.973; ít nhất 2)

BÀI 3.32. Một nhà máy sản xuất với tỷ lệ phế phẩm là 7%.

a) Quan sát ngẫu nhiên 10 sản phẩm. Tính xác suất để có

   i) đúng một phế phẩm   ii) ít nhất một phế phẩm   iii) nhiều nhất một phế phẩm

   (Đs: 0.3643; 0.516; 0.8483)

b) Hỏi phải quan sát ít nhất bao nhiêu sản phẩm để xác suất nhận được ít nhất một phế phẩm ≥ 0.9.
   (Đs: ít nhất 32)

4.2. PHÂN PHỐI SIÊU BỘI

☑ **Định nghĩa:** Ta xét tập hợp N đối tượng bao gồm: K đối tượng có tính chất A và $N - K$ đối tượng không có tính chất A, lấy một mẫu ngẫu nhiên gồm n đối tượng (không hoàn lại) từ N đối tượng này ($K < N$ và $n < N$). Khi đó, **biến ngẫu nhiên siêu bội** là số đối tượng có tính chất A trong mẫu n đối tượng với xác suất tương ứng

$$\mathbb{P}(X = x) = \frac{C_K^x C_{N-K}^{n-x}}{C_N^n}, \quad \max\{0, n + K - N\} \leq x \leq \min\{K, n\},$$

ký hiệu $X \sim H(N; K; n)$.

☑ **Tính chất:** Nếu X là biến ngẫu nhiên siêu bội $H(N; K; n)$ thì

i) $\mathbb{E}(X) = np$ với $p = \dfrac{K}{N}$.

ii) $Var(X) = npq\dfrac{N-n}{N-1}$ với $q = 1 - p$.

Chú ý: đại lượng $\dfrac{N-n}{N-1}$ được gọi là hệ số hiệu chỉnh. Khi n nhỏ hơn N thì hệ số hiệu chỉnh nhỏ và phân phối siêu bội $X \sim H(N; K; n)$ có thể được xấp xỉ bởi phân phối nhị thức $X \sim B(n, K/N)$.

VÍ DỤ 4.3. Một lớp có 50 sinh viên trong đó có 30 sinh viên nữ. Cần chọn 10 bạn vào một đội văn nghệ, giả sử khả năng được chọn của các sinh viên là như nhau. Gọi X là số sinh viên nữ được chọn, khi đó $X \sim H(50; 30; 10)$ và các giá trị có thể của X là $k \in \{0; ..., 10\}$.

a) Tính xác suất để có đúng 2 sinh viên nữ trong đôi văn nghệ.

b) Tính xác suất để số sinh viên nữ được chọn không quá 3.

c) Tính xác suất để chọn được ít nhất một sinh viên nữ.

✍ **LỜI GIẢI.**

a) Xác suất để có đúng 2 sinh viên nữ trong đôi văn nghệ là

$$\mathbb{P}(X=2)=\frac{C_{30}^2C_{20}^8}{C_{50}^{10}}\approx 0.0053.$$

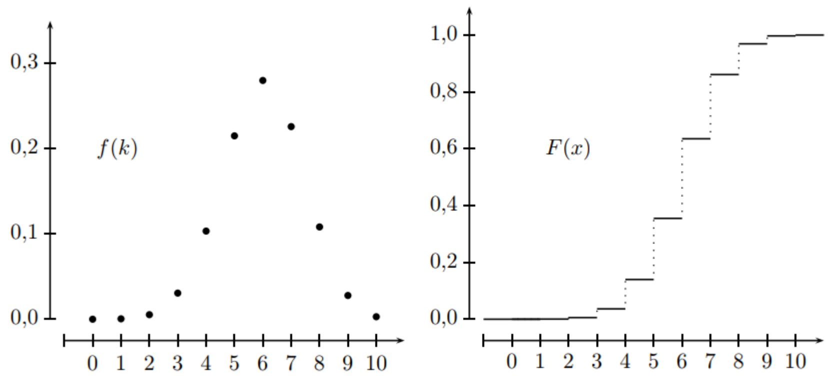
b) Xác suất để số sinh viên nữ được chọn không quá 3 là

$$\begin{aligned}\mathbb{P}(X\le 3)&=\mathbb{P}(X=0)+\mathbb{P}(X=1)+\mathbb{P}(X=2)+\mathbb{P}(X=3)\\&=\frac{C_{30}^0C_{20}^{10}}{C_{50}^{10}}+\frac{C_{30}^1C_{20}^9}{C_{50}^{10}}+\frac{C_{30}^2C_{20}^8}{C_{50}^{10}}+\frac{C_{30}^3C_{20}^7}{C_{50}^{10}}\approx 0.03648.\end{aligned}$$

c) Xác suất để chọn được ít nhất một sinh viên nữ

$$\begin{aligned}\mathbb{P}(X>1)&=1-\mathbb{P}(X<1)\\&=1-\mathbb{P}(X=0)\\&=1-\frac{C_{30}^0C_{20}^{10}}{C_{50}^{10}}\approx 0.99998.\end{aligned}$$

□



Hình 3.Hàm giá trị xác suất và hàm phân phối xác suất của biến ngẫu nhiên $X\sim H(50;30;10)$.

BÀI 3.33. Có một cái hộp chứa 8 quả cầu trắng và 3 quả cầu đen. Lấy ngẫu nhiên không hoàn lại 4 quả cầu. Gọi X là số quả cầu trắng lấy được. Tính xác suất

a) Lấy được ít nhất 1 quả cầu trắng.

b) Lấy được 2 quả cầu trắng.

c) Tính $\mathbb{E}(X)$ và $Var(X)$.

....

....

....

....

....

....

BÀI 3.34. Một lô chứa 36 tế bào vi khuẩn và 12 tế bào trong đó không có khả năng sao chép (sinh sản) tế bào. Giả sử bạn kiểm tra 3 tế bào vi khuẩn được chọn ngẫu nhiên, không cần hoàn lại.

a) Hãy xác định hàm phân phối ứng với biến ngẫu nhiên X là số tế bào có thể sao chép trong mẫu lấy ra.

b) Tính trung bình và phương sai của X.

c) Tính xác suất để có ít nhất 1 tế bào trong mẫu không thể sao chép.

....

....

....

....

....

....

....

....

BÀI 3.35. Một lô gồm 100 bóng đèn trong đó 8 bóng hư, còn lại là bóng tốt. Lấy ngẫu nhiên 10 bóng đèn trong lô hàng trên, tính xác suất để trong 10 bóng được lấy trên có đúng 2 bóng hư.

BÀI 3.36. Một cửa hàng có 20 sản phẩm trong đó có 15 sản phẩm loại A. Chọn ngẫu nhiên từ cửa hàng 4 sản phẩm. Tính xác suất có 2 sản phẩm loại A trong 4 sản phẩm lấy ra?

BÀI 3.37 (Sử dụng phân phối Nhị thức xấp xỉ phân phối siêu bội). Một công ty có 800 người đàn ông dưới 55 tuổi. Giả sử 30% mang dấu hiệu trên nhiễm sắc thể nam biểu thị nguy cơ cao huyết áp.

a) Nếu 10 người đàn ông trong công ty được xét nghiệm dấu hiệu của nhiễm sắc thể này, xác suất có chính xác một người đàn ông mang dấu hiệu đó với nhận xét số người được xét nghiệm $n = 10$ nhỏ hơn rất nhiều so với $N = 800$?

b) Nếu 10 người đàn ông trong công ty được xét nghiệm dấu hiệu của nhiễm sắc thể này, xác suất có nhiều hơn một người đàn ông mang dấu hiệu đó?

(Đs: 0.1201; 0.8523)

4.3. PHÂN PHỐI POISSON

☑ **Định nghĩa: Biến ngẫu nhiên Poisson**: một biến ngẫu nhiên rời rạc được mô hình bằng phân phối Poisson nếu thỏa các điều kiện:

i) Các biến cố xảy ra một cách ngẫu nhiên và độc lập giữa những khoảng thời gian hoặc không gian nhất định.

ii) Số lượng trung bình λ các biến cố xảy ra trong những khoảng là đồng nhất và hữu hạn.

Khi đó, số lượng biến cố xảy ra trong những khoảng thời gian tương ứng với xác suất

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

ký hiệu $X \sim P(\lambda)$.

☑ **Tính chất:** Nếu biến ngẫu nhiên X có phân phối Poisson với tham số λ, $X \sim P(\lambda)$ thì

i) $\mathbb{E}(X) = \lambda$.

ii) $Var(X) = \lambda$.

Chú ý: Khi $X \sim B(n, p)$ với p nhỏ và n lớn thì ta có thể xấp xỉ bởi phân phối Poisson $X \sim P(np)$.

VÍ DỤ 4.4. Một nhà máy dệt có số ống sợi bị đứt trong một giờ tuân theo phân phối Poisson với tham số $\lambda = 4$.

a) Tìm xác suất để trong 1 giờ máy hoạt động có đúng 2 ống sợi bị đứt.

b) Tìm xác suất để trong 1 giờ máy hoạt động không có quá 2 ống sợi bị đứt.

✍ **LỜI GIẢI.**

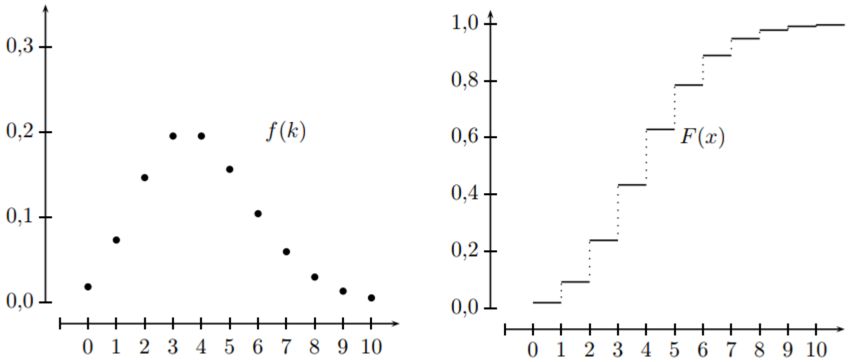
Gọi X là số ống sợi bị đứt trong 1 giờ, khi đó $X \sim P(4)$.

a) Xác suất để trong 1 giờ máy hoạt động có đúng 2 ống sợi bị đứt là

$$\mathbb{P}(X = 2) = \frac{e^{-4} \cdot 4^2}{2!} \approx 0.1465.$$

b) Xác suất để trong 1 giờ máy hoạt động không có quá 2 ống sợi bị đứt là

$$\begin{aligned}\mathbb{P}(X \leq 2) &= \mathbb{P}(X = 0) + \mathbb{P}(X = 1) + \mathbb{P}(X = 2)\\ &= \frac{e^{-4} \cdot 4^0}{0!} + \frac{e^{-4} \cdot 4^1}{1!} + \frac{e^{-4} \cdot 4^2}{2!} = 13e^{-4}.\end{aligned}$$



Hình 4.Hàm giá trị xác suất và hàm phân phối xác suất của biến ngẫu nhiên $X \sim P(4)$.

□

BÀI 3.38. Giả sử X có phân phối Poisson với trung bình là 4. Tính xác suất

a) $\mathbb{P}(X = 0)$ b) $\mathbb{P}(X \leq 2)$ c) $\mathbb{P}(X = 4)$ d) $\mathbb{P}(X = 8)$

BÀI 3.39. Số lượng các cuộc gọi điện thoại đến trao đổi với một trạm giao dịch thường được mô hình hóa bằng biến ngẫu nhiên Poisson. Giả sử trung bình có 10 cuộc gọi mỗi giờ. Xác suất có chính xác

a) 5 cuộc gọi trong một giờ.
b) 3 cuộc gọi hoặc ít hơn trong một giờ.
c) 15 cuộc gọi trong hai giờ.
d) 5 cuộc gọi trong 30 phút.

BÀI 3.40. Số lỗ hỏng trên một tấm vải của nhà máy dệt được giả định là phân phối Poisson với trung bình 0,1 lỗ trên một mét vuông vải. Xác suất để có

a) 2 lỗ hỏng trên 1 m^2 vải.
b) 1 lỗ hỏng trên 10 m^2 vải.
c) 0 lỗ hỏng trên 20 m^2 vải.
d) ít nhất 2 lỗ hỏng trên 10 m^2 vải.

....

....

....

....

....

....

BÀI 3.41. Số lượng danh mục của một trang Web thay đổi tuân theo phân phối Poisson với trung bình là 0.25 trên 1 ngày. Xác suất có

a) nhiều hơn hay bằng 2 thay đổi trong một ngày.

b) không có thay đổi trong năm ngày.

c) ít hơn hay bằng 2 thay đổi trong năm ngày.

(Đs: 0.026; 0.287; 0.868)

BÀI 3.42. Một trung tâm bưu điện nhận được trung bình 3 cuộc điện thoại trong mỗi phút. Tính xác suất để trung tâm này nhận được 1 cuộc, 2 cuộc, 3 cuộc gọi trong 1 phút, biết rằng số cuộc gọi trong một phút có phân phối Poisson. (Đs: 0.1494; 0.224; 0.224)

BÀI 3.43. Khi nhà sản xuất đĩa máy tính kiểm tra đĩa, họ ghi vào đĩa và sau đó kiểm tra nó bằng cách sử dụng một chương trình xác nhận. Trình xác nhận đến số xung hoặc lỗi bị thiếu. Số lỗi trên vùng thử của đĩa có phân phối Poisson với $\lambda = 0.2$.

a) Xác định số lỗi trung bình trên phần diện tích được kiểm tra.

b) Tỉ lệ phần trăm diện tích kiểm tra để có hai lỗi hay ít hơn. (Đs: 99.89%)

BÀI 3.44. Số lượng các lỗ hổng bề mặt trong các tấm nhựa được sử dụng làm nội thất của xe ô tô có phân bố Poisson với giá trị trung bình 0.05 lỗ trên một mét vuông của bảng nhựa. Giả sử một bộ phận nội thất ô tô chứa 10 mét vuông của bảng nhựa.

a) Xác suất để không có lỗi nào trên bộ phận nội thất đó.

b) Nếu 10 chiếc xe được bán cho một công ty cho thuê, xác suất để không có bất kì chiếc trong số 10 chiếc xe có bất kỳ sai sót bề mặt nào?

c) Nếu 10 chiếc xe được bán cho một công ty cho thuê, xác suất để nhiều nhất một chiếc trong số 10 chiếc xe có bất kỳ sai sót bề mặt nào?

(Đs: 0.6065; 0.0067; 0.0504)

BÀI 3.45. Tỷ lệ một loại bệnh bẩm sinh trong dân số là $p = 0.01$. Bệnh này cần sự chăm sóc đặc biệt lúc mới sinh. Một nhà bảo sinh thường có 20 ca sinh trong một tuần. Tính xác suất để

a) Không có trường hợp nào cần chăm sóc đặc biệt.

b) Có đúng một trường hợp cần chăm sóc đặc biệt.

c) Có nhiều hơn một trường hợp cần chăm sóc đặc biệt.

Tính bằng quy luật nhị thức rồi dùng quy luật Poisson để so sánh kết quả khi ta xấp xỉ phân phối nhị thức bằng phân phối Poisson .

BÀI 3.46. Giả sử rằng số lượng khách hàng bước vào ngân hàng trong một giờ là một biến ngẫu nhiên Poisson và giả sử rằng $\mathbb{P}(X = 0) = 0.05$. Xác định trung bình và phương sai của X.

BÀI 3.47. Các nhà thiên văn học đếm số lượng các ngôi sao trong một thể tích không gian cho trước được xem như là biến ngẫu nhiên Poisson. Mật độ trong thiên hà Milky Way trong vùng lân cận với hệ mặt trời của chúng ta là một ngôi sao trên 16 năm ánh sáng

a) Xác suất để có từ 2 ngôi sao trở lên trong 16 năm ánh sáng.

b) Cần bao nhiêu năm ánh sáng để xác suất có 1 hoặc nhiều hơn ngôi sao lớn hơn 0.95.

(Đs: 0.264; 48)

CHƯƠNG 4

BIẾN NGẪU NHIÊN LIÊN TỤC

1. MỘT SỐ KHÁI NIỆM

1.1. ĐỊNH NGHĨA

Biến ngẫu nhiên X của một phép thử τ với không gian mẫu Ω là một ánh xạ

$$\begin{array}{rccc} X: & \Omega & \to & \mathbb{R} \\ & \omega & \mapsto & X(\omega) \end{array}$$

Biến ngẫu nhiên được gọi là **liên tục** nếu tập hợp các giá trị mà nó có thể nhận $X(\Omega)$ là một khoảng dạng (a, b) hoặc toàn bộ $\mathbb{R}$.

1.2. HÀM MẬT ĐỘ XÁC SUẤT

Hàm mật độ xác suất (probability density function) của một biến ngẫu nhiên liên tục X là hàm số thỏa

i) $f(x) \geq 0$

ii) $\displaystyle\int_{-\infty}^{\infty} f(x)dx = 1$

iii) $\mathbb{P}(a \leq X \leq b) = \displaystyle\int_a^b f(x)dx.$

VÍ DỤ 1.1. Tìm k để hàm sau là hàm mật độ.

$$f(x) = \begin{cases} k(1+2x) & \text{khi } 0 \leq x \leq 2, \\ 0 & \text{chỗ khác.} \end{cases}$$

✍ **LỜI GIẢI.**

Với $0 \leq x \leq 2$, ta có $1 + 2x > 0$. Do đó, để $f(x) \geq 0, \forall x$ thì $k \geq 0$.

Ta có

$$\begin{aligned}\int_{-\infty}^{+\infty} f(x)\mathrm{d}x &= \int_{-\infty}^{0} f(x)\mathrm{d}x + \int_{0}^{2} f(x)\mathrm{d}x + \int_{2}^{+\infty} f(x)\mathrm{d}x \\ &= \int_{-\infty}^{0} 0\mathrm{d}x + \int_{0}^{2} k(1+2x)\mathrm{d}x + \int_{2}^{+\infty} 0\mathrm{d}x \\ &= 0 + \frac{k}{2}(2x+1)^2\Big|_0^2 + 0 \\ &= 6k.\end{aligned}$$

Mặt khác, để f là hàm mật độ thì $\int_{-\infty}^{+\infty} f(x)\mathrm{d}x = 1$, do đó $6k = 1 \Rightarrow k = \frac{1}{6}$.
Vậy hàm mật độ được xác định là

$$f(x) = \begin{cases} \frac{1}{6}(1+2x) & \text{khi } 0 \le x \le 2, \\ 0 & \text{chỗ khác.} \end{cases}$$

□

1.3. HÀM PHÂN PHỐI TÍCH LŨY

Hàm phân phối tích lũy (cumulative distribution function) của biến ngẫu nhiên liên tục X là:

$$F(a) = \mathbb{P}(X \le a) = \int_{-\infty}^{a} f(x)dx, \quad x \in \mathbb{R}.$$

VÍ DỤ 1.2. Cho X là biến ngẫu nhiên liên tục có hàm mật độ

$$f(x) = \begin{cases} \frac{1}{6}(1+2x) & \text{khi } 0 \le x \le 2, \\ 0 & \text{chỗ khác.} \end{cases}$$

Tìm hàm phân phối tích lũy của X.

✍ **LỜI GIẢI.**

∗ Với $x < 0$, khi đó

$$F(x) = \mathbb{P}(X \le x) = \int_{-\infty}^{x} f(t)\mathrm{d}t = \int_{-\infty}^{x} 0\mathrm{d}t = 0.$$

∗ Với $0 \le x < 2$, khi đó

$$\begin{aligned}F(x) = \mathbb{P}(X \le x) &= \int_{-\infty}^{x} f(t)\mathrm{d}t \\ &= \int_{-\infty}^{0} f(t)\mathrm{d}t + \int_{0}^{x} f(t)\mathrm{d}t \\ &= \int_{-\infty}^{0} 0\mathrm{d}t + \int_{0}^{x} \frac{1}{6}(1+2t)\mathrm{d}t \\ &= 0 + \frac{1}{24}(2t+1)^2\Big|_0^x\end{aligned}$$

$$= \frac{1}{24}\left[(2x+1)^2 - 1\right].$$

$*$ Với $x \geq 2$, khi đó

$$\begin{aligned}
F(x) = \mathbb{P}(X \leq x) &= \int_{-\infty}^{x} f(t)\mathrm{d}t \\
&= \int_{-\infty}^{0} f(t)\mathrm{d}t + \int_{0}^{2} f(t)\mathrm{d}t + \int_{2}^{x} f(t)\mathrm{d}t \\
&= \int_{-\infty}^{0} 0\mathrm{d}t + \int_{0}^{2} \frac{1}{6}(1+2t)\mathrm{d}t + \int_{2}^{x} 0\mathrm{d}t \\
&= 0 + \frac{1}{24}(2t+1)^2\Big|_0^2 + 0 \\
&= 1.
\end{aligned}$$

Vậy hàm phân phối tích tũy của X là

$$F(x) = \begin{cases} 0 & \text{nếu } x < 0 \\ \dfrac{1}{24}\left[(2x+1)^2 - 1\right] & \text{nếu } 0 < x \leq 2 \\ 1 & \text{nếu } 2 \leq x. \end{cases}$$

$\square$

2. MỘT SỐ ĐẶC TRƯNG CỦA BIẾN NGẪU NHIÊN

1) **Trung trình** (**kỳ vọng**) của biến ngẫu nhiên liên tục X có hàm mật độ $f(x)$ được định nghĩa là

$$\mathbb{E}(X) \equiv \mu_X = \int_{-\infty}^{+\infty} x f(x)\mathrm{d}x.$$

2) **Kỳ vọng của hàm của biến ngẫu nhiên**
Cho g là hàm số thực bất kỳ, kỳ vọng của hàm g của biến ngẫu nhiên rời rạc X, ký hiệu là $\mathbb{E}(g(X))$ xác định theo công thức:

$$\mathbb{E}(g(X)) = \int_{-\infty}^{+\infty} g(x) f(x)\mathrm{d}x.$$

Đặc biệt,

- ☑ Nếu $g(x) = x^r$, ta gọi $\mathbb{E}(g(X))$ là *moment bậc r*.
- ☑ Nếu $g(x) = e^{tx}$ thì ta gọi $\mathbb{E}(g(X))$ là *hàm sinh moment*.
- ☑ Nếu $g(x) = e^{-itx}$ thì ta gọi $\mathbb{E}(g(X))$ là *hàm đặc trưng* của biến ngẫu nhiên X.

3) **Phương sai** của biến ngẫu nhiên liên tục X có hàm mật độ $f(x)$ được định nghĩa là

$$Var(X) \equiv \sigma_X^2 = \int_{-\infty}^{+\infty} (x-\mu_X)^2 f(x)\mathrm{d}x = \int_{-\infty}^{+\infty} x^2 f(x)\mathrm{d}x - \mu_X^2.$$

4) **Độ lệch tiêu chuẩn** của biến ngẫu nhiên liên tục X được định nghĩa là căn bậc hai của phương sai
$$\sigma_X = \sqrt{Var(X)}.$$

VÍ DỤ 2.1. Cho X là biến ngẫu nhiên liên tục có hàm mật độ
$$f(x) = \begin{cases} \frac{1}{6}(1+2x) & \text{khi } 0 \leq x \leq 2, \\ 0 & \text{chỗ khác.} \end{cases}$$

a) Tính $\mathbb{E}(X)$ và $\mathbb{E}(X^2)$.

b) Tính phương sai và độ lệch tiêu chuẩn của X.

✍ **LỜI GIẢI.**

a) Ta có $\mathbb{E}(X) \equiv \mu_X = \int_{-\infty}^{+\infty} xf(x)\mathrm{d}x = \int_0^2 x \cdot \frac{1}{6}(1+2x)\mathrm{d}x = \frac{11}{9}$.
$\mathbb{E}(X^2) = \int_{-\infty}^{+\infty} x^2 f(x)\mathrm{d}x = \int_0^2 x^2 \cdot \frac{1}{6}(1+2x)\mathrm{d}x = \frac{16}{9}$.

b) Phương sai của X là $Var(X) \equiv \sigma_X^2 = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = \frac{16}{9} - \left(\frac{11}{9}\right)^2 = \frac{23}{81}$.
Lưu ý: Ta cũng có thể tính phương sai của X như sau
$$Var(X) \equiv \sigma_X^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)\mathrm{d}x = \int_0^2 \left(x - \frac{11}{9}\right)^2 \cdot \frac{1}{6}(1+2x)\mathrm{d}x = \frac{23}{81}.$$
Độ lệch tiêu chuẩn của X là
$$\sigma_X = \sqrt{Var(X)} = \sqrt{\frac{23}{81}} = \frac{23}{9}.$$

□

BÀI 4.1. Dòng điện trong một mạch nhất định được đo bằng một ampe kế là biến ngẫu nhiên liên tục X với hàm mật độ sau
$$f(x) = \begin{cases} 0{,}075x + 0{,}2 & \text{khi } 3 \leq x \leq 5, \\ 0 & \text{chỗ khác.} \end{cases}$$

a) Hãy vẽ hàm mật độ của phân phối và kiểm tra phần diện tích bên dưới đường cong của hàm mật độ là 1.

b) Tính $\mathbb{P}(X \leq 4)$ và so sánh với $\mathbb{P}(X > 4)$.

c) Tính $\mathbb{P}(3{,}5 \leq X \leq 4{,}5)$ và $\mathbb{P}(X > 4{,}5)$.

....

....

....

....

....

....

....

....

BÀI 4.2. Lỗi liên quan đến việc thực hiện một phép đo nhất định là một biến ngẫu nhiên X liên tục với hàm mật độ

$$f(x) = \begin{cases} 0{,}09375(4 - x^2) & \text{khi } -2 \leq x \leq 2, \\ 0 & \text{chỗ khác.} \end{cases}$$

a) Hãy vẽ hàm mật độ của phân phối.

b) Tính $\mathbb{P}(-1 < X < 1)$.

c) Tính $\mathbb{P}(X > 0)$.

d) Tính $\mathbb{P}(X < -0{,}5 \text{ hoặc } X > 0{,}5)$.

....

....

....

....

....

....

....

....

3. MỘT SỐ PHÂN PHỐI THƯỜNG GẶP

3.1. PHÂN PHỐI ĐỀU

Biến ngẫu nhiên liên tục X được gọi là có **phân phối đều** trên đoạn $[a;b]$, ký hiệu $X \sim U[a;b]$, nếu hàm mật độ xác suất của X có dạng

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

Từ định nghĩa trên ta có được hàm phân phối xác suất của $X \sim U[a;b]$

$$F(x) = \begin{cases} 0 & \text{khi } x < a \\ \dfrac{x-a}{b-a} & \text{khi } x \in [a;b] \\ 1 & \text{khi } x > b \end{cases}$$

BÀI 4.3. Giả sử nhiệt độ phản ứng X (tính theo °C) trong một quá trình phản ứng hóa học nhất định có phân phối đều với $a = -5$ và $b = 5$.

a) Tính $\mathbb{P}(X < 0)$.

b) Tính $\mathbb{P}(-2{,}5 < X < 2{,}5)$.

c) Tính $\mathbb{P}(-2 \le X \le 3)$.

d) Với k thỏa $-5 < k < k+4 < 5$, hay tính $\mathbb{P}(k < X < k+4)$.

....

....

....

....

....

....

BÀI 4.4. Một bài báo: "Second Moment Reliability Evaluation vs. Monte Carlo Simulations for Weld Fatigue Strength (Quality and Reliability Engr. Intl., 2012: 887-896)" xem xét việc sử dụng phân phối đều với $a = 0{,}20$ và $b = 4{,}25$ cho đường kính X của những mối hàn (mm).

a) Tìm và vẽ hàm mật độ của X.

b) Xác suất đường kính vượt quá 3 mm là bao nhiêu?

c) Xác suất đường kính trong vòng 1 mm của đường kính trung bình là bao nhiêu?

d) Với a thỏa điều kiện $0{,}2 < a < a+1 < 4{,}5$ thì xác suất $\mathbb{P}(a < X < a+1)$ bao nhiêu?

BÀI 4.5. Khi đi làm, một giáo sư trước tiên phải lên xe buýt gần nhà cô và sau đó chuyển sang tuyến xe buýt thứ hai. Nếu thời gian chờ đợi (tính bằng phút) tại mỗi điểm dừng có phân bố đều với $a = 0$ và $b = 5$, khi đó có thể thấy rằng tổng thời gian chờ đợi Y có hàm mật độ

$$f(y) = \begin{cases} \dfrac{1}{25}y & \text{khi } 0 \le y \le 5, \\ \dfrac{2}{5} - \dfrac{1}{25}y & \text{khi } 5 \le y \le 10, \\ 0 & \text{chỗ khác.} \end{cases}$$

a) Vẽ hàm mật độ của Y.

b) Kiểm tra tính chất $\int_{-\infty}^{+\infty} f(y)dy = 1$.

c) Xác suất tổng thời gian chờ tối đa 3 phút là bao nhiêu?

d) Xác suất tổng thời gian chờ tối đa 8 phút là bao nhiêu?

e) Xác suất tổng thời gian chờ từ 3 đến 8 phút là bao nhiêu?

f) Xác suất mà tổng thời gian chờ đợi là ít hơn 2 phút hoặc lớn hơn 6 phút?

BÀI 4.6. Gọi X biểu thị ứng suất rung (với đơn vị psi) trên lưỡi tuabin gió ở tốc độ gió cụ thể trong đường hầm gió. Bài báo "Blade Fatigue Life Assessment with Application to VAWTS" (J. of Solar Energy Engr., 1982: 107–111) đề xuất phân phối Rayleigh, với hàm mật độ

$$f(x;\theta) = \begin{cases} \dfrac{x}{\theta^2} e^{-\frac{x^2}{2\theta^2}} & \text{khi } x > 0, \\ 0 & \text{chỗ khác.} \end{cases}$$

là mô hình cho phân phối của X.

a) Kiểm tra tính hợp lí của hàm $f(x;\theta)$.

b) Giả sử $\theta = 100$ (một giá trị được đề xuất bằng một đồ thị có trong bài báo). Tính xác suất X nhiều nhất là 200? Ít hơn 200? Nhiều hơn 200?

c) Tính xác suất X nằm giữa 100 và 200 (vẫn với giả thiết về θ như trên).

d) Tính hàm $\mathbb{P}(X \leq x)$.

BÀI 4.7. Dựa trên cơ sở phân tích dữ liệu, một bài báo trên: "Pedestrians' Crossing Behaviors and Safety at Unmarked Roadways in China (Accident Analysis and Prevention, 2011: 1927-1936)" đã đề xuất hàm mật độ $f(x) = 0{,}15e^{-0{,}15(x-1)}$ với $x \geq 1$ làm mô hình cho phân phối của X = thời gian (giây) được sử dụng ở dòng trung bình.

a) Xác suất mà thời gian chờ tối đa là 5 giây là bao nhiêu? Hơn 5 giây?

b) Xác suất mà thời gian chờ đợi là từ 2 đến 5 giây là bao nhiêu?

BÀI 4.8. Gọi X là tuổi thọ của con người. Một công trình nghiên cứu cho biết hàm mật độ của X là

$$f(x) = \begin{cases} cx^2(100-x)^2 & \text{khi } 0 \leq x \leq 100, \\ 0 & \text{khi } x < 0 \text{ hay } x > 100. \end{cases}$$

a) Xác định hằng số c.

b) Tính trung bình và phương sai của X.

c) Tính xác suất của một người có tuổi thọ ≥ 60.

d) Tính xác suất của một người có tuổi thọ ≥ 60, biết rằng người đó hiện nay đã 50 tuổi.

(Đs: $3{,}10^{-9}$; 50, 2500/7; 0,31744; 0,63548)

BÀI 4.9. Một giáo sư đại học không bao giờ kết thúc bài giảng của mình trước khi hết giờ và luôn hoàn thành bài giảng của mình trong vòng 2 phút sau giờ học. Cho X là thời gian trôi qua giữa thời điểm hết tiết học và kết thúc bài giảng của giáo sư. Giả sử hàm mật độ của X là

$$f(x) = \begin{cases} kx^2 & \text{khi } 0 \le x \le 2, \\ 0 & \text{chỗ khác.} \end{cases}$$

a) Tìm k và vẽ hàm mật độ tương ứng.

b) Hãy tính xác suất bài giảng kết thúc trong vòng 1 phút sau khi giờ học kết thúc.

c) Hãy tính xác suất bài giảng tiếp tục diễn ra sau khi giờ học kết thúc từ 60 s tới 90 s.

d) Xác suất mà bài giảng tiếp tục trong ít nhất 90 s ngoài giờ kết thúc là bao nhiêu?

BÀI 4.10. Tỷ lệ thời gian Y mà một rô bốt công nghiệp hoạt động trong suốt một tuần 40 giờ là một biến ngẫu nhiên có hàm mật độ xác suất

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

a) Tìm $\mathbb{E}(Y)$ và $Var(Y)$.

b) Đối với các rô bốt đang được nghiên cứu, lợi nhuận X mỗi tuần được cho bởi $X = 200Y - 60$. Tìm $\mathbb{E}(X)$ và $Var(X)$.

BÀI 4.11. Cho X là biến ngẫu nhiên liên tục với hàm phân phối tích lũy

$$F(x) = \begin{cases} 0 & \text{khi } x \le 0, \\ \dfrac{x}{4}\left(1 + \ln\dfrac{4}{x}\right) & \text{khi } 0 < x \le 4, \\ 1 & \text{khi } x \ge 4. \end{cases}$$

(Hàm phân phối tích lũy này được đề xuất trong bài báo "Variability in Measured Bedload Transport Rates (Water Resources Bull., 1985: 39–48)" được xem như mô hình cho biến ngẫu nhiên về thủy văn. Hãy tính

a) $\mathbb{P}(X \le 1)$.

b) $\mathbb{P}(1 \le X \le 3)$.

c) Xác định hàm mật độ của X.

....

....

....

....

....

....

....

....

BÀI 4.12. Cho X là biến ngẫu nhiên liên tục với hàm phân phối tích lũy

$$F(x) = \begin{cases} 0 & \text{khi } x < -2, \\ \dfrac{1}{2} + \dfrac{3}{32}\left(4x - \dfrac{x^3}{3}\right) & \text{khi } -2 \le x < 2, \\ 1 & \text{khi } x \ge 2. \end{cases}$$

a) Tính $\mathbb{P}(X < 0)$.

b) Tính $\mathbb{P}(-1 < X < 1)$.

c) Tính $\mathbb{P}(X > 0{,}5)$.

d) Kiểm tra lại hàm mật độ $f(x)$ là đạo hàm của $F'(x)$.

....

....

....

....

....

....

....

....

BÀI 4.13. Cho X là một biến ngẫu nhiên có hàm phân phối tích lũy như sau

$$F(x) = \begin{cases} 0 & \text{khi } x < 0, \\ \dfrac{x^2}{4} & \text{khi } 0 \le x < 2, \\ 1 & \text{khi } x \ge 2. \end{cases}$$

a) Tính $\mathbb{P}(X \le 1)$.

b) Tính $\mathbb{P}(0{,}5 \le X \le 1)$.

c) Tính $\mathbb{P}(X \ge 1{,}5)$.

d) Tìm hàm mật độ $f(x)$.

e) Tính $\mathbb{E}(X)$.

f) Tính $Var(X)$ và σ_X.

BÀI 4.14. Trong bài báo: "Modeling Sediment and Water Column Interactions for Hydrophobic Pollutant (Water Research, 1984: 1169 - 1174)"đề xuất phân bố đồng đều trên khoảng $(7{,}5, 20)$ làm mô hình cho độ sâu (cm) của lớp sinh học trong trầm tích ở một khu vực nhất định.

a) Tính trung bình và phương sai của độ sâu.

b) Tính hàm phân phối tích lũy của độ sâu.

c) Xác suất quan sát độ sâu tối đa là 10? Từ 10 đến 15?

d) Xác suất mà độ sâu quan sát được trong phạm vi 1 lần độ lệch chuẩn của giá trị trung bình là bao nhiêu? Trong vòng 2 lần độ lệch chuẩn?

BÀI 4.15. Gọi X là lượng không gian bị chiếm bởi một văn kiện được đặt trong một thùng container loại 1 ft^3. Hàm mật độ xác suất của X là

$$f(x) = \begin{cases} 90x^8(1-x) & \text{khi } 0 < x < 1, \\ 0 & \text{chỗ khác.} \end{cases}$$

a) Vẽ hàm mật độ của X. Tìm hàm phân phối tích lũy của X và vẽ hàm này.

b) Tìm $\mathbb{P}(X \leq 0.5)$? So sánh với $F(0{,}5)$.

c) Sử dụng kết quả từ câu a tính xác suất $\mathbb{P}(0{,}25 < X \leq 0{,}5)$. Kết quả có khác với $\mathbb{P}(0{,}25 \leq X \leq 0{,}5)$.

d) Hãy tìm vị trí x khi xác suất đạt 75%.

e) Tính $\mathbb{E}(X)$ và σ_X.

3.2. PHÂN PHỐI CHUẨN

① **Định nghĩa:** Biến ngẫu nhiên liên tục X được gọi là có **phân phối chuẩn** nếu hàm mật độ xác suất có dạng

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

trong đó $\mu = \mathbb{E}(X)$ và $\sigma^2 = Var(X)$, ký hiệu $X \sim N(\mu, \sigma^2)$.

Đặc biệt, nếu $\mu = 0$ và $\sigma^2 = 1$ thì ta gọi Z là biến ngẫu nhiên có **phân phối chuẩn tắc**, ký hiệu $Z \sim N(0; 1)$ và đặt Φ là hàm phân phối tích lũy có dạng

$$\Phi(z) = \mathbb{P}(Z \leq z).$$

② **Tính chất:**

i) Hàm phân phối xác suất của $Z \sim N(0; 1)$ là $\Phi(a) = \mathbb{P}(Z \leq a) = \displaystyle\int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$. Giá trị của hàm này được tra trong bảng.

ii) $\Phi(z) = 1 - \Phi(-z)$. Công thức này dùng để tìm giá trị của hàm Φ tại một z **âm**.

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

iv) Nếu $X \sim N(\mu; \sigma^2)$ thì $\mathbb{P}(X \leq x) = \mathbb{P}\left(\dfrac{X-\mu}{\sigma} \leq \dfrac{x-\mu}{\sigma}\right) = \mathbb{P}(Z \leq z) = \Phi(z)$.

v) Nếu $X \sim N(\mu; \sigma^2)$ thì

$$\mathbb{P}(a \leq X \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

BÀI 4.16. Sử dụng bảng phân phối chuẩn tắc hoặc máy tính, tính các xác suất sau

a) $\mathbb{P}(Z < 1{,}32)$ b) $\mathbb{P}(Z < 3)$ c) $\mathbb{P}(Z > 1{,}45)$

d) $\mathbb{P}(Z > -2{,}15)$ e) $\mathbb{P}(-2{,}34 < Z < 1{,}76)$

✍ **LỜI GIẢI.**

Ta có $Z \sim N(0;1) \Rightarrow \mu = 0$; $\sigma^2 = 1$.
Hàm mật độ của Z là $f(z) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$.

a) $\mathbb{P}(Z < 1{,}32) = \displaystyle\int_{-\infty}^{1{,}32} f(z)\mathrm{d}z = \int_{-\infty}^{1{,}32} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \mathrm{d}z$.

Tiến hành bấm máy hoặc tra bảng, ta tính được $\mathbb{P}(Z < 1{,}32) = \displaystyle\int_{-\infty}^{1{,}32} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \mathrm{d}z \approx 0{,}90658$.

□

....

....

....

....

....

....

BÀI 4.17. Sử dụng bảng phân phối chuẩn tắc hoặc máy tính, tính các xác suất sau

a) $\mathbb{P}(-1 < Z < 1)$ b) $\mathbb{P}(-2 < Z < 2)$ c) $\mathbb{P}(-3 < Z < 3)$

d) $\mathbb{P}(Z > 3)$ e) $\mathbb{P}(0 < Z < 1)$

....

....

....

....

....

....

....

....

BÀI 4.18 (Bài toán $z-$value). Sử dụng bảng phân phối chuẩn tắc hoặc máy tính, vẽ và tính các $z-$value nếu biết

a) $\mathbb{P}(Z < z) = 0{,}9$

b) $\mathbb{P}(Z < z) = 0{,}5$

c) $\mathbb{P}(Z > z) = 0{,}1$

d) $\mathbb{P}(Z > z) = 0{,}9$

e) $\mathbb{P}(-1{,}24 < Z < z) = 0{,}8$

✍ **LỜI GIẢI.**

Ta có $Z \sim N(0; 1) \Rightarrow \mu = 0; \sigma^2 = 1$.
Hàm mật độ của Z là $f(z) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$.

a) Ta có

$$\begin{aligned} \mathbb{P}(Z < z) &= 0{,}9 \\ \Leftrightarrow \int_{-\infty}^{z} f(t)\mathrm{d}t &= 0{,}9 \end{aligned}$$

Cách 1: Dùng máy tính cầm tay có chức năng tra cứu DIST.

- ☑ Bước 1: Vào môi trường DIST/ Phân Phối
  - Ở máy 570VN PLUS/VINACAL: MODE $\to$ ▼ $\to$ 3:DIST.
  - Ở máy 580VN: MENU $\to$ ▼ $\to$ 7:Distribution.
- ☑ Bước 2: Truy cập tra ngược phân phối chuẩn: 3: Inverse Normal
- ☑ Bước 3: Area = xác suất cần tra (ở câu a) này là $= 0{,}9$).
- ☑ Bước 4: Nhập giá trị cho μ và σ. Lưu ý: $Z \sim N(0; 1)$ nên $\mu = 0$ và $\sigma = 1$.
- ☑ Bước 5: Ghi kết quả cần tìm.

Ta được $z \approx 1{,}28155$.
Cách 2: Tra bảng phân phối

- ☑ Bước 1: Đưa xác suất cần tính về hàm $\Phi(a) = \mathbb{P}(Z \leq a) = \mathbb{P}(Z < a)$.

$$\begin{aligned} \mathbb{P}(Z < z) &= 0{,}9 \\ \Leftrightarrow \int_{-\infty}^{z} f(t)\mathrm{d}t &= 0{,}9 \\ \Leftrightarrow \Phi(z) &= 0{,}9. \end{aligned}$$

- ☑ Bước 2: Tìm xác suất bên trong bảng gần với 0,9 nhưng không lớn hơn.
- ☑ Bước 3: Từ vị trí xác suất xác định ở Bước 2, ta gióng theo hàng, theo cột sẽ được kết quả $z \approx 1{,}28$.

$$\Phi(z) = P(Z \le z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}\, du$$

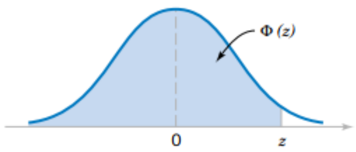


Table II Cumulative Standard Normal Distribution (*continued*)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.500000	0.503989	0.507978	0.511967	0.515953	0.519939	0.532922	0.527903	0.53[illegible]881	0.535856
0.1	0.539828	0.543795	0.547758	0.551717	0.555760	0.559618	0.563559	0.567495	0.57[illegible]424	0.575345
0.2	0.579260	0.583166	0.587064	0.590954	0.594835	0.598706	0.602568	0.606420	0.61[illegible]261	0.614092
0.3	0.617911	0.621719	0.625516	0.629300	0.633072	0.636831	0.640576	0.644309	0.64[illegible]027	0.651732
0.4	0.655422	0.659097	0.662757	0.666402	0.670031	0.673645	0.677242	0.680822	0.68[illegible]386	0.687933
0.5	0.691462	0.694974	0.698468	0.701944	0.705401	0.708840	0.712260	0.715661	0.71[illegible]043	0.722405
0.6	0.725747	0.729069	0.732371	0.735653	0.738914	0.742154	0.745373	0.748571	0.75[illegible]748	0.754903
0.7	0.758036	0.761148	0.764238	0.767305	0.770350	0.773373	0.776373	0.779350	0.78[illegible]305	0.785236
0.8	0.788145	0.791030	0.793892	0.796731	0.799546	0.802338	0.805106	0.807850	0.81[illegible]570	0.813267
0.9	0.815940	0.818589	0.821214	0.823815	0.826391	0.828944	0.831472	0.833977	0.83[illegible]457	0.838913
1.0	0.841345	0.843752	0.846136	0.848495	0.850830	0.853141	0.855428	0.857690	0.85[illegible]929	0.862143
1.1	0.864334	0.866500	0.868643	0.870762	0.872857	0.874928	0.876976	0.878999	0.88[illegible]000	0.882977
1.2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	$\boxed{0.899727}$	0.901475
1.3	0.903199	0.904902	0.906582	0.908241	0.909877	0.911492	0.913085	0.914657	0.916207	0.917736
1.4	0.919243	0.920730	0.922196	0.923641	0.925066	0.926471	0.927855	0.929219	0.930563	0.931888

Lưu ý: Dù dùng cách nào, thì việc tra ngược chỉ tra được cho trường hợp tích phân có dạng $\Phi(a) = \mathbb{P}(Z \le a) = \mathbb{P}(Z < a) = \displaystyle\int_{-\infty}^{z} f(t)\mathrm{d}t$.

c) Hướng dẫn

$$\begin{aligned} \mathbb{P}(Z > z) &= 0{,}1 \\ \Leftrightarrow 1 - \mathbb{P}(Z \le z) &= 0{,}1 \\ \Leftrightarrow \mathbb{P}(Z \le z) &= 0{,}9 \end{aligned}$$

e) Hướng dẫn

$$\begin{aligned} \mathbb{P}(-1{,}24 < Z < z) &= 0{,}8 \\ \Leftrightarrow \int_{-1{,}24}^{z} f(t)\mathrm{d}t &= 0{,}8 \\ \Leftrightarrow \int_{-\infty}^{z} f(t)\mathrm{d}t - \int_{-\infty}^{-1{,}24} f(t)\mathrm{d}t &= 0{,}8 \\ \Leftrightarrow \int_{-\infty}^{z} f(t)\mathrm{d}t - 0{,}10749 &= 0{,}8 \\ \Leftrightarrow \int_{-\infty}^{z} f(t)\mathrm{d}t &= 0{,}90749 \\ \Rightarrow z &\approx \ldots \end{aligned}$$

□

....

....

BÀI 4.19. Sử dụng bảng phân phối chuẩn tắc hoặc máy tính, tính các $z-$value nếu biết

a) $\mathbb{P}(-z < Z < z) = 0{,}95$

b) $\mathbb{P}(-z < Z < z) = 0{,}99$

c) $\mathbb{P}(-z < Z < z) = 0{,}68$

d) $\mathbb{P}(-z < Z < z) = 0{,}9973$

✍ LỜI GIẢI.

Ta có $Z \sim N(0;1) \Rightarrow \mu = 0; \sigma^2 = 1$.
Lưu ý: Vì $Z \sim N(0;1)$ nên có hàm mật độ $f(z)$ đối xứng qua Oy, do đó

$$\mathbb{P}(-z < Z < z) = \int_{-z}^{z} f(t)\mathrm{d}t = 2\int_{0}^{z} f(t)\mathrm{d}t$$

a) Ta có

$$\begin{aligned} \mathbb{P}(-z < Z < z) &= 0{,}95 \\ \Leftrightarrow \int_{-z}^{z} f(t)\mathrm{d}t &= 0{,}95 \\ \Leftrightarrow 2\int_{0}^{z} f(t)\mathrm{d}t &= 0{,}95 \\ \Leftrightarrow \int_{0}^{z} f(t)\mathrm{d}t &= 0{,}475 \end{aligned}$$

Tới đây dùng câu e) Bài 4.18.

□

BÀI 4.20. Cho X là biến ngẫu nhiên phân phối chuẩn với trung bình 10 và phương sai 2. Xác định

a) $\mathbb{P}(X < 13)$ b) $\mathbb{P}(X > 9)$ c) $\mathbb{P}(6 < X < 14)$

d) $\mathbb{P}(2 < X < 4)$ e) $\mathbb{P}(-2 < X < 8)$

✍ **LỜI GIẢI.**

Ta có $X \sim N(10; 2) \Rightarrow \mu = 10$ và $\sigma = \sqrt{2}$.

Lưu ý: Nếu $X \sim N(\mu; \sigma^2)$ **thì** $Z = \dfrac{X-\mu}{\sigma} \sim N(0; 1)$.

a) $\mathbb{P}(X < 13) = \mathbb{P}\left(\dfrac{X-10}{\sqrt{2}} < \dfrac{13-10}{\sqrt{2}}\right) = \mathbb{P}\left(Z < \dfrac{3\sqrt{2}}{2}\right) = \displaystyle\int_{-\infty}^{\frac{3\sqrt{2}}{2}} f(z)\mathrm{d}z \approx 0{,}98305.$

□

BÀI 4.21. Cho X là biến ngẫu nhiên phân phối chuẩn với trung bình 10 và phương sai 2. Xác định giá trị của x thỏa

a) $\mathbb{P}(X > x) = 0{,}5$ b) $\mathbb{P}(X > x) = 0{,}95$

c) $\mathbb{P}(x < X < 10) = 0{,}2$ d) $\mathbb{P}(-x < X - 10 < x) = 0{,}95$

e) $\mathbb{P}(-x < X - 10 < x) = 0{,}99$

✍ **LỜI GIẢI.**

Ta có $X \sim N(10; 2) \Rightarrow \mu = 10$ và $\sigma = \sqrt{2}$.
Ta sẽ chuyển X sang Z rồi làm như Bài 4.18.

b) Ta có

$$\begin{aligned}
\mathbb{P}(X > x) &= 0{,}95\\
\Leftrightarrow 1 - \mathbb{P}(X \le x) &= 0{,}95\\
\Leftrightarrow \mathbb{P}(X \le x) &= 0{,}05\\
\Leftrightarrow \mathbb{P}\left(Z \le \frac{x-10}{\sqrt{2}}\right) &= 0{,}05\\
\Rightarrow \frac{x-10}{\sqrt{2}} &\approx -1{,}645\\
\Rightarrow x &\approx 7{,}674.
\end{aligned}$$

$\square$

BÀI 4.22. Trọng lượng X (tính bằng gam) một loại trái cây có phân phối chuẩn với trung bình 500 gam và phương sai 16 gam^2. Trái cây thu hoạch được phân loại theo trọng lượng như sau:

- Loại 1: trên 505 gam.
- Loại 2: từ 495 – 505 gam.
- Loại 3: dưới 495 gam.

Tính tỷ lệ mỗi loại. (Đs: 0,10565; 0,7887; 0,10565)

BÀI 4.23. Cường độ nén của các mẫu xi măng có thể được mô hình hóa bởi một phân bố chuẩn với giá trị trung bình 6000 kg/cm^2 và độ lệch chuẩn là 100 kg/cm^2.

a) Xác suất để cường độ nén của mẫu nhỏ hơn 6250 kg/cm^2.

b) Xác suất để cường độ nén của mẫu trong khoảng 5800 đến 5900 kg/cm^2.

c) Độ nén là bao nhiêu để có thể chiếm ít nhất 95% mẫu.

(Đs: 0,99379; 0,13591; 5835)

BÀI 4.24. Thời gian cho đến khi cần sạc lại pin cho một máy tính xách tay trong điều kiện bình thường là phân phối chuẩn với trung bình 260 phút và độ lệch chuẩn là 50 phút.

a) Xác suất pin sử dụng kéo dài hơn bốn giờ là bao nhiêu?

b) Xác định thời gian sử dụng pin tại những *giá trị phân vị*, là giá trị của z sao cho xác suất $\mathbb{P}(Z < z)$ đạt 25% và 75%.

c) Xác định thời gian sử dụng pin tương ứng với xác suất ít nhất 0,95.

....

....

....

....

....

....

....

....

BÀI 4.25. Đường kính của một chi tiết máy do một máy tiện tự động sản xuất có phân phối chuẩn với trung bình 0,001 mm và độ lệch chuẩn 0,05 mm. Chi tiết máy được xem là đạt yêu cầu nếu đường kính không sai quá 0,1 mm.

a) Tính tỷ lệ sản phẩm đạt yêu cầu.

b) Lấy ngẫu nhiên 3 sản phẩm. Tính xác suất có ít nhất một sản phẩm đạt yêu cầu.
(*sử dụng kết quả câu a và phân phối nhị thức*)

BÀI 4.26. Trọng lượng mỗi trái bí ngòi xanh trong vườn là một đại lượng ngẫu nhiên có phân phối chuẩn với trung bình 1,2 kg và độ lệch chuẩn là 0,2 kg. Biết rằng những trái có trọng lượng từ 0,8 kg trở lên là những trái có thể thu hoạch. Tính xác suất những trái bí ngòi xanh có thể thu hoạch trong vườn.

BÀI 4.27. Nuôi ốc bươu đen (hay còn gọi là ốc nhồi) là một trong những nghề giúp nông dân miền Tây làm giàu. Kích thước mỗi con ốc bươu đen trong hầm nuôi là một đại lượng ngẫu nhiên có phân phối chuẩn với trung bình 4,2 cm và độ lệch chuẩn là 0,3 cm. Biết rằng những con có kích thước từ 3,8 cm trở lên là những con có thể bán thương phẩm. Tính tỉ lệ những con ốc bươu đen có thể bán thương phẩm trong hầm.

4. ĐỊNH LÝ GIỚI HẠN TRUNG TÂM

4.1. ĐỊNH LÝ GIỚI HẠN TRUNG TÂM

Cho X_1, X_2, ..., X_n là các biến ngẫu nhiên độc lập và có cùng phân phối với trung bình μ và phương sai σ^2. Khi đó $\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ có phân phối xấp xỉ chuẩn tắc $N(0;1)$ khi $n \to +\infty$. Trong đó $\bar{X} = \dfrac{X_1 + \cdots + X_n}{n}$.

4.2. XẤP XỈ PHÂN PHỐI NHỊ THỨC BẰNG PHÂN PHỐI CHUẨN

Định lý 1: Cho $X \sim B(n;p)$ thì biến ngẫu nhiên $Z = \dfrac{X - np}{\sqrt{np(1-p)}}$ được xấp xỉ bởi phân phối chuẩn tắc $N(0;1)$ đủ tốt khi $np > 5$ và $n(1-p) > 5$. Khi đó, ta sử dụng hiệu chỉnh liên tục như sau

- ☑ $\mathbb{P}(X \leq x) = \mathbb{P}(X \leq x + 0{,}5) \approx \mathbb{P}\left(Z \leq \dfrac{x + 0{,}5 - np}{\sqrt{np(1-p)}}\right)$,
- ☑ $\mathbb{P}(X \geq x) = \mathbb{P}(X \geq x - 0{,}5) \approx \mathbb{P}\left(Z \geq \dfrac{x - 0{,}5 - np}{\sqrt{np(1-p)}}\right)$.

BÀI 4.28. Giả sử X có phân phối nhị thức với $n = 200$ và $p = 0{,}4$. Hãy xấp xỉ xác suất

a) $\mathbb{P}(X \leq 70)$ b) $\mathbb{P}(70 \leq X \leq 90)$ c) $\mathbb{P}(X = 80)$

(ĐS: 0,0853; 0,8293; 0,0575)

....

....

....

....

....

....

....

....

BÀI 4.29. Biết rằng tỷ lệ phế phẩm của phân xưởng A là 3%. Kiểm tra lô hàng gồm 500 sản phẩm từ phân xưởng A, hãy tính xác suất để

a) có nhiều nhất 60 phế phẩm, b) số phế phẩm từ 55 đến 75.

....

....

....

....

....

....

BÀI 4.30. Một trại nuôi cá sấu tiến hành thử nghiệm ấp 500 trứng cá sấu bằng máy ấp của nhà sản suất A, theo quảng cáo, khi sử dụng máy ấp và thực hiện đúng quy trình, tỷ lệ trứng nở là 78%. Tính tỉ lệ để trong 500 trứng có ít nhất 380 trứng nở.

....

....

....

....

....

BÀI 4.31. Một shiper giao một ngày 500 đơn hàng. Mỗi đơn có xác suất không liên lạc được khách hàng là 0,25. Tính xác suất có ít nhất 130 đơn không liên lạc được khách hàng?

....

....

....

....

....

4.3. XẤP XỈ PHÂN PHỐI POISSON BẰNG PHÂN PHỐI CHUẨN

Định lý 2: Cho $X \sim P(\lambda)$ thì biến ngẫu nhiên $Z = \dfrac{X-\lambda}{\sqrt{\lambda}}$ được xấp xỉ bởi phân phối chuẩn tắc $N(0;1)$ đủ tốt khi $\lambda > 5$. Khi đó, ta sử dụng công thức hiệu chỉnh liên tục như sau

- ☑ $\mathbb{P}(X \le x) = \mathbb{P}(X \le x + 0{,}5) \approx \mathbb{P}\left(Z \le \dfrac{x + 0{,}5 - \lambda}{\sqrt{\lambda}}\right)$,
- ☑ $\mathbb{P}(X \ge x) = \mathbb{P}(X \ge x - 0{,}5) \approx \mathbb{P}\left(Z \ge \dfrac{x - 0{,}5 - \lambda}{\sqrt{\lambda}}\right)$.

BÀI 4.32. Giả sử X có phân phối Poisson với $\lambda = 6$.

a) Tính chính xác xác suất X nhỏ hơn 4.

b) Sử dụng cách xấp xỉ để tính xác suất trên và so sánh kết quả.

c) Xấp xỉ xác suất để $8 < X < 12$. Tính bằng hai cách.

....

....

....

....

....

....

BÀI 4.33. Giả sử rằng số lượng các hạt amiăng trong một mét vuông trên bề mặt tuân theo phân bố Poisson với giá trị trung bình là 1000. Nếu phân tích một mét bụi bình phương, xác suất nào có thể tìm thấy ít hơn hoặc bằng 950 hạt. (Đs: 0,058)

....

....

....

....

....

....

PHẦN

II

THỐNG KÊ

CHƯƠNG 5 ƯỚC LƯỢNG

1. ƯỚC LƯỢNG ĐIỂM

1.1. THAM SỐ ĐẶC TRƯNG THỐNG KÊ MẪU

Giả sử ta có một bộ dữ liệu $x_1, \ldots, x_n$ của một tổng thể X mà ta cần quan tâm nghiên cứu, khi đó, ta có các đặc trưng thống kê mẫu sau

① Trung bình mẫu

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

② Phương sai mẫu

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

③ Độ lệch chuẩn mẫu

$$s = \sqrt{s^2}$$

1.2. TÍNH CÁC ĐẶC TRƯNG TỪ DỮ LIỆU TRÊN MÁY TÍNH CẦM TAY

1.2.1. Dữ liệu không có tần số

Fx 570	Fx 580
$\boxed{\text{Mode}}$ → $\boxed{\text{3:STAT}}$ → $\boxed{\text{1: 1-VAR}}$ → Nhập dữ liệu → $\boxed{\text{AC}}$	$\boxed{\text{MENU}}$ → $\boxed{\text{6 : Stat}}$ → $\boxed{\text{1 :1-VAR}}$ → Nhập dữ liệu → $\boxed{\text{AC}}$
$\boxed{\text{SHIFT}}$ → $\boxed{1}$ → $\boxed{\text{4 :Var}}$ → 1 : n; 2 : $\bar{x}$; 4 : s	$\boxed{\text{OPTN}}$ → $\boxed{\downarrow}$ → $\boxed{\text{2 : Variable}}$ → 6 n; 1 : $\bar{x}$; 5 : s

VÍ DỤ 1.1. Ta có bộ dữ liệu: 129, 132, 140, 141, 138, 143, 133, 137, 140, 143, 138, 140, thực hiện các bước trên, ta sẽ tính được: $n = 12$, $\bar{x} = 137{,}83333$ và $s = 4{,}40729$.

1.2.2. Dữ liệu có tần số

Fx 570	Fx 580
$\boxed{\text{SHIFT}}$ → $\boxed{\text{Mode}}$ → $\boxed{\downarrow}$ → $\boxed{\text{4:STAT}}$ → $\boxed{\text{1:On}}$	$\boxed{\text{SHIFT}}$ → $\boxed{\text{MENU}}$ → $\boxed{\downarrow}$ → $\boxed{\text{3:STAT}}$ → $\boxed{\text{1:On}}$
$\boxed{\text{Mode}}$ → $\boxed{\text{3:STAT}}$ → $\boxed{\text{1: 1-VAR}}$ → Nhập dữ liệu → $\boxed{\text{AC}}$	$\boxed{\text{MENU}}$ → $\boxed{\text{6 : Stat}}$ → $\boxed{\text{1 :1-VAR}}$ → Nhập dữ liệu → $\boxed{\text{AC}}$
$\boxed{\text{SHIFT}}$ → $\boxed{\text{1}}$ → $\boxed{\text{4 :Var}}$ 1 : n 2 : $\bar{x}$ 4 : s	$\boxed{\text{OPTN}}$ → $\boxed{\downarrow}$ → $\boxed{\text{2 : Variable}}$ 6 n 1 : $\bar{x}$ 5 : s

VÍ DỤ 1.2. Ta có bảng dữ liệu

X	0	1	2	3	4
Số gia đình	5	19	28	7	3

thực hiện các bước trên, ta tính được: $n = 62$, $\bar{x} = 1{,}7419$, $s = 0{,}9398$.

1.2.3. Dữ liệu tần số dạng khoảng

X	$(a_1; b_1]$	$(a_2; b_2]$	...	$(a_k; b_k]$
N	n_1	n_2	...	n_k

X	$\dfrac{a_1+b_1}{2}$	$\dfrac{a_2+b_2}{2}$	...	$\dfrac{a_k+b_k}{2}$
N	n_1	n_2	...	n_k

VÍ DỤ 1.3. Ta có bảng dữ liệu

X	140 – 145	145 – 150	150 – 155	155 – 160	160 – 165	165 – 170
Số người	1	3	7	9	5	2

Thực hiện biến đổi, ta được

X	142,5	147,5	152,5	157,5	162,5	167,5
Số người	1	3	7	9	5	2

Từ đây, ta thực hiện các bước ở dạng dữ liệu 2 (dữ liệu tần số) để tính các đặc trưng.

2. ƯỚC LƯỢNG KHOẢNG

2.1. KHOẢNG TIN CẬY CHO TRUNG BÌNH

2.1.1. Khoảng tin cậy cho trung bình khi phương sai σ^2 đã biết

Định nghĩa 1. Nếu $\bar{x}$ là trung bình mẫu được tính từ mẫu ngẫu nhiên với cỡ mẫu n được lấy từ một tổng thể có phân phối chuẩn với phương sai σ^2 đã biết, thì khoảng tin cậy $100(1-\alpha)\%$ cho μ được xác định

$$\bar{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$$

trong đó $z_{1-\alpha/2}$ thỏa $\mathbb{P}(Z < z_{1-\frac{\alpha}{2}}) = 1 - \dfrac{\alpha}{2}$ với $Z \sim N(0;1)$.

Độ chính xác và cỡ mẫu:

- $\epsilon = z_{1-\alpha/2}\dfrac{\sigma}{\sqrt{n}}$ gọi là độ chính xác (hay sai số) của ước lượng.
- Chiều dài khoảng tin cậy: 2ϵ.
- Cho trước sai số ϵ và độ tin cậy $100(1-\alpha)\%$, công thức tính cỡ mẫu

$$n = \left(\frac{z_{1-\alpha/2}\sigma}{\epsilon}\right)^2$$

BÀI 5.1. Cho tổng thể có phân phối chuẩn với tham số σ^2 đã biết. Xác định

a) Độ tin cậy của ước lượng khoảng

$$\bar{x} - 2.14\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + 2.14\frac{\sigma}{\sqrt{n}}$$

b) Độ tin cậy của ước lượng khoảng

$$\bar{x} - 2.49\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + 2.49\frac{\sigma}{\sqrt{n}}$$

c) Độ tin cậy của ước lượng khoảng

$$\bar{x} - 1.85\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + 1.85\frac{\sigma}{\sqrt{n}}$$

(Đs: 96.76%; 98.72%; 93.56%)

✍ **LỜI GIẢI.**

Nhắc lại: Ước lượng khoảng cho trung bình khi phương sai σ^2 đã biết có dạng

$$\bar{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$$

trong đó $z_{1-\alpha/2}$ thỏa $\mathbb{P}(Z < z_{1-\frac{\alpha}{2}}) = 1 - \dfrac{\alpha}{2}$ với $Z \sim N(0;1)$.

a) Ta có $z_{1-\alpha/2} = 2.14$.

$$\begin{aligned} \text{mà} \quad \mathbb{P}(Z < z_{1-\frac{\alpha}{2}}) &= 1 - \frac{\alpha}{2} \\ \Leftrightarrow \mathbb{P}(Z < 2.14) &= 1 - \frac{\alpha}{2} \\ \Leftrightarrow 0.9838 &= 1 - \frac{\alpha}{2} \\ \Rightarrow \alpha &= 0.0324. \end{aligned}$$

$\Rightarrow$ Độ tin cậy: $(1-\alpha) \cdot 100\% = 96.76\%$.

....

....

....

....

....

....

□

BÀI 5.2. Cho tổng thể có phân phối chuẩn với tham số σ^2 đã biết. Xác định

a) Giá trị của $z_{1-\frac{\alpha}{2}}$ ứng với độ tin cậy 98% trong công thức xây dựng khoảng tin cậy.

b) Giá trị của $z_{1-\frac{\alpha}{2}}$ ứng với độ tin cậy 80% trong công thức xây dựng khoảng tin cậy.

c) Giá trị của $z_{1-\frac{\alpha}{2}}$ ứng với độ tin cậy 75% trong công thức xây dựng khoảng tin cậy.

(Đs: 2.33; 1.285; 1.151)

✍ **LỜI GIẢI.**

Phương sai σ^2: đã biết.

a) Độ tin cậy: $98\% \Rightarrow (1-\alpha) \cdot 100\% = 98\%$

$$\Rightarrow \alpha = 0.02.$$

$$\begin{aligned} \text{Ta có} \quad \mathbb{P}(Z < z_{1-\frac{\alpha}{2}}) &= 1 - \frac{\alpha}{2} \\ \Leftrightarrow \mathbb{P}(Z < z_{1-\frac{\alpha}{2}}) &= 1 - \frac{0.02}{2} \\ \Leftrightarrow \mathbb{P}(Z < z_{1-\frac{\alpha}{2}}) &= 0.99 \\ \Rightarrow z_{1-\frac{\alpha}{2}} &= 2.3263. \end{aligned}$$

....

....

....

....

....

....

□

BÀI 5.3. Ước lượng khoảng tin cậy cho độ hoàn thiện mạch của một thiết bị bán dẫn. Giả sử mức độ hoàn thiện tuân theo phân phối chuẩn với độ lệch chuẩn $\sigma = 20$. Tính

a) Khoảng tin cậy 95% cho μ khi $n = 10$ và $\bar{x} = 1000$.

b) Khoảng tin cậy 95% cho μ khi $n = 25$ và $\bar{x} = 1000$.

c) Khoảng tin cậy 99% cho μ khi $n = 10$ và $\bar{x} = 1000$.

d) Khoảng tin cậy 99% cho μ khi $n = 25$ và $\bar{x} = 1000$.

e) Nhận xét độ rộng của khoảng tin cậy khi thay đổi cỡ mẫu và độ tin cậy.

(Đs: $[987.6; 1012.4]$; $[992.16; 1007.84]$; $[983.71; 1016.28]$; $[989.7; 1010.3]$)

✍ **LỜI GIẢI.**

$\sigma = 20$: đã biết.

a) • **Độ tin cậy:** $95\% \Rightarrow (1-\alpha)\cdot 100\% = 95\%$

$$\Rightarrow \alpha = 0.05.$$

$\Rightarrow z_{1-\frac{\alpha}{2}} = z_{0.975} = 1.96.$

• **Sai số:**
$$\begin{aligned} \epsilon &= z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \\ &= 1.96 \cdot \frac{20}{\sqrt{10}} \\ &\approx 12.40. \end{aligned}$$

• **Khoảng tin cậy** 95% **cho** μ **là**
$$\begin{aligned} \bar{x} - \epsilon \leq \quad & \mu \quad \leq \bar{x} + \epsilon \\ \Leftrightarrow 1000 - 12.40 \leq \quad & \mu \quad \leq 1000 + 12.40 \\ \Leftrightarrow 987.60 \leq \quad & \mu \quad \leq 1012.40 \end{aligned}$$
$\Rightarrow \mu \in [987.60; 1012.40].$

....

....

....

□

BÀI 5.4. Đường kính của những cái lỗ trên dây nịt có phân phối chuẩn với độ lệch chuẩn 0.01 inch. Một mẫu ngẫu nhiên có cỡ là 10 với đường kính trung bình là 1.5054 inch. Tìm khoảng tin cậy 99% cho ước lượng trung bình đường kính lỗ. (Đs:$[1.49725; 1.51355]$)

BÀI 5.5. Một kỹ sư xây dựng phân tích cường độ nén của bê tông. Cường độ nén thường có phân phối chuẩn với phương sai là 1000 $(\text{psi})^2$. Một mẫu ngẫu nhiên gồm 12 mẫu có cường độ nén trung bình $\bar{x} = 3250$ psi.

a) Xây dựng khoảng tin cậy 95% cho trung bình cường độ nén của bê tông. (Đs:$[3232.11; 3267.89]$)

b) Xây dựng khoảng tin cậy 99% cho trung bình cường độ nén của bê tông và so sánh độ rộng với câu trên. (Đs:$[3226.4; 3273.6]$)

c) Để có được ước lượng cho trung bình độ nén với sai số không vượt quá 15 psi ở dộ tin cậy 99% thì cỡ mẫu phải là bao nhiêu? (Đs: 30)

....

....

....

....

....

....

....

....

....

BÀI 5.6. Một mẫu ngẫu nhiên được chọn từ tổng thể có phân phối chuẩn và các khoảng tin cậy sau đây được xây dựng từ cùng một bộ dữ liệu $(38{,}02; 61{,}98)$ và $(39{,}95; 60{,}05)$

a) Tính giá trị trung bình mẫu.
Hướng dẫn: Với khoảng tin cậy $[a; b]$ hoặc $(a; b)$ thì độ dài khoảng tin cậy là $2\epsilon = b - a \Rightarrow \epsilon = ...$
Mà cận bên trái khoảng tin cậy là $\bar{x} - \epsilon = 38{,}02 \Rightarrow \bar{x} = ...$

b) Một trong hai khoảng tin cậy trên được xây dựng với độ tin cậy 95% và cái còn lại là 90%. Khoảng tin cậy nào tương ứng với 95% và giải thích?
Hướng dẫn: Dựa vào nhận xét của bài 6.4 đề xác định.

BÀI 5.7. Một mẫu ngẫu nhiên được chọn từ tổng thể có phân phối chuẩn và các khoảng tin cậy sau đây được xây dựng từ cùng một bộ dữ liệu $(37{,}53; 49{,}87)$ và $(35{,}59; 51{,}81)$

a) Tính giá trị trung bình mẫu.

b) Một trong hai khoảng tin cậy trên được xây dựng với độ tin cậy 95% và cái còn lại là 99%. Khoảng tin cậy nào tương ứng với 95% và giải thích?

BÀI 5.8. Một nhà sản xuất ống piston cho động cơ tự động. Được biết, đường kính ống có phân phối chuẩn với độ lệch chuẩn là 0,01 mm. Một mẫu ngẫu nhiên gồm 15 ống piston có đường kính trung bình là $\bar{x} = 74{,}036$ mm. Tìm khoảng tin cậy 99% cho đường kính trung bình của ống piston.

BÀI 5.9. Giả sử tuổi thọ của bóng đèn do một công ty sản xuất xấp xỉ phân phối chuẩn với độ lệch chuẩn là 40 giờ. Một mẫu 30 bóng đèn cho thấy tuổi thọ trung bình là 780 giờ.

a) Hãy tìm khoảng tin cậy 96% cho tuổi thọ trung bình của tất cả các bóng đèn do công ty sản xuất.

b) Nếu muốn sai số ước lượng không quá 10 giờ, thì phải quan sát ít nhất bao nhiêu bóng đèn?

Phân phối Student - t

Định nghĩa 2. Xét $X_1, ..., X_n \sim N(\mu;\sigma)$ với μ và σ^2 không biết. Biến ngẫu nhiên

$$T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$$

có phân phối Student t với $n-1$ bậc tự do.
Hàm mật độ của T có dạng

$$f(t) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\sqrt{\pi k}\left(\frac{t^2}{k}+1\right)^{(k+1)/2}} \quad -\infty < t < +\infty$$

2.1.2. Khoảng tin cậy cho trung bình khi phương sai σ^2 chưa biết

Định nghĩa 3. Nếu $\bar{x}$ và s lần lượt là trung bình mẫu và độ lệch tiêu chuẩn của một mẫu ngẫu nhiên cỡ n được chọn từ tổng thể có phân phối chuẩn với **trung bình μ và phương sai σ^2 không biết**, khoảng tin cậy $100(1-\alpha)\%$ cho μ được xác định như sau

$$\bar{x} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} \le \mu \le \bar{x} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$$

với $t_{\alpha/2,n-1}$ thỏa $\mathbb{P}(|T| > t_{\alpha/2,n-1}) = \alpha$ với $T \sim t(n-1)$.

BÀI 5.10 (Bài toán $t-$value). Xác định giá trị t trong xây dựng khoảng tin cậy tương ứng với

a) Độ tin cậy 95% với bậc tự do 12.

b) Độ tin cậy 95% với bậc tự do 24.

c) Độ tin cậy 99% với bậc tự do 13.

d) Độ tin cậy 99.9% với bậc tự do 15.

(Đs: 2.179; 2.0644; 3.012; 4.073)

✍ **LỜI GIẢI.**

a) • **Độ tin cậy:** $95\% \Rightarrow (1-\alpha)\cdot 100\% = 95\%$

$$\Rightarrow \alpha = 0.05.$$

$\Rightarrow t_{\frac{\alpha}{2};n-1} = t_{0.025;12}$.
Tiến hành tra bảng phân phối Student, ta được $t_{0.025;12} = 2.179$.

....

....

....

....

□

BÀI 5.11. Kỹ sư nghiên cứu cho một nhà sản xuất lốp xe đang nghiên cứu tuổi thọ lốp làm từ một hợp chất cao su mới và đã chế tạo 16 lốp xe, thử nghiệm tuổi thọ khi chúng chạy trên đường. Giá trị trung bình mẫu và độ lệch chuẩn là 60139.7 km và 3645.94 km. Tìm khoảng tin cậy 95% cho tuổi thọ trung bình của lốp. (Đs: $[58197.33; 62082.07]$)

✍ **LỜI GIẢI.**

....

....

....

....

....

....

....

....

....

□

BÀI 5.12. Độ sáng của ống hình ảnh của tivi có thể được đánh giá bằng cách đo lượng dòng điện cần thiết để đạt được một mức độ sáng cụ thể. Một mẫu của 10 ống cho kết quả độ lệch chuẩn là 15.7 và trung bình là 317.2. Tìm khoảng tin cậy 99% cho trung bình thực tế yêu cầu.(Đs: [301.06; 333.34])

BÀI 5.13. Cục Khí tượng của Chính phủ Úc đã cung cấp lượng mưa trung bình hàng năm (milimet) ở Úc 1983-2002 như sau (http://www.bom.gov.au/climate/change/rain03.txt):

499.2, 555.2, 398.8, 391.9, 453.4, 459.84, 483.7, 417.6, 469.2, 452.4,
499.3, 340.6, 522.8, 469.9, 527.2, 565.5, 584.1, 727.3, 558.6, 338.6

Xây dựng khoảng tin cậy 95% cho lượng mưa trung bình hàng năm.
(Đs: $\bar{x} = 485.755$; $s^2 = 1149.229$; $[469.89; 501.62]$)

✍ **LỜI GIẢI.**

....

....

....

....

....

....

....

....

....

□

BÀI 5.14. Người ta đo ion Na^+ trên một số người và ghi nhận lại được kết quả như sau

$$129,\ 132,\ 140,\ 141,\ 138,\ 143,\ 133,\ 137,\ 140,\ 143,\ 138,\ 140$$

a) Tính trung bình mẫu và phương sai mẫu. (Đs: 137.83; 19.42)

b) Ước lượng trung bình của tổng thể ở độ tin cậy 0.95. (Đs: $[135.01; 140.63]$; $[9.76; 56.1]$)

c) Nếu muốn sai số ước lượng trung bình không quá $\epsilon = 1$ với độ tin cậy 0.95 thì phải quan sát mẫu gồm ít nhất mấy người? (Đs: 75)

✍ **LỜI GIẢI.**

....

....

....

....

....

....

....

....

....

....

....

....

□

BÀI 5.15. Một bài báo về Kỹ thuật hạt nhân quốc tế (tháng 2 năm 1988, trang 33) mô tả một số đặc tính của các thanh nhiên liệu được sử dụng trong lò phản ứng thuộc sở hữu của một công ty điện ở Na Uy. Các phép đo về tỷ lệ làm giàu của 12 thanh đã được báo cáo như sau:

$$2.94;\ 3.00;\ 2.90;\ 2.75;\ 3.00;\ 2.95;\ 2.90;\ 2.75;\ 2.95;\ 2.82;\ 2.81;\ 3.05$$

Tìm khoảng tin cậy 99% cho tỷ lệ phần trăm trung bình của làm giàu. Bạn có đồng ý với tuyên bố rằng tỷ lệ phần trăm trung bình của làm giàu là 2.95? Tại sao? (Đs: [2.813;2.991])

BÀI 5.16. Quan sát tuổi thọ X (giờ) của một số bóng đèn do xí nghiệp A sản xuất, ta ghi nhận

X	1000	1100	1200	1300	1400	1500	1600	1700	1800
N	10	14	16	17	18	16	16	12	9

a) Tính trung bình mẫu và độ lệch tiêu chuẩn mẫu.
(Đs: 1391.41; 234.45)

b) Ước lượng tuổi thọ trung bình của bóng đèn ở độ tin cậy 0.95. (Đs:[1350.79; 1432.03])

c) Nếu muốn sai số ước lượng trung bình không quá $\epsilon = 30$ với độ tin cậy 0.95 thì phải quan sát mẫu gồm ít nhất mấy bóng đèn?(Đs: 235)

✍ LỜI GIẢI.

....

....

....

....

....

....

....

....

....

....

....

....

□

BÀI 5.17. Đo đường kính của một chi tiết máy do một máy tiện tự động sản xuất, ta ghi nhận được số liệu như sau

X	12.00	12.05	12.10	12.15	12.20	12.25	12.30	12.35	12.40
N	2	3	7	9	10	8	6	5	3

với N chỉ số trường hợp tính theo từng giá trị của X (mm).

a) Tính trung bình mẫu và độ lệch chuẩn của mẫu. (Đs: 12.21, 0.103)

b) Ước lượng đường kính trung bình μ ở độ tin cậy 0.95. (Đs: $[12.18; 12.24]$)

c) Nếu muốn sai số ước lượng không quá $\epsilon = 0.02$ mm ở độ tin cậy 0.95 thì phải quan sát ít nhất mấy trường hợp. (Đs: 102)

BÀI 5.18. Quan sát chiều cao X (cm) của một số người, ta ghi nhận

X	$140-145$	$145-150$	$150-155$	$155-160$	$160-165$	$165-170$
Số người	1	3	7	9	5	2

a) Tính trung bình mẫu và phương sai mẫu.(Đs: 156.2; 37.68)

b) Ước lượng trung bình của tổng thể ở độ tin cậy 0.95.(Đs:$[153.77; 158.63]$)

BÀI 5.19. Đem cân một số trái cây vừa thu hoạch, ta được kết quả sau

X (g)	$200-210$	$210-220$	$220-230$	$230-240$	$240-250$
Số trái	12	17	20	18	15

a) Tìm khoảng ước lượng của trọng lượng trung bình của trái cây với độ tin cậy 0.95 và 0.99. (Đs: $[222.98; 228.72]$, $[222.08; 229.63]$)

b) Nếu muốn sai số ước lượng không quá $\epsilon = 2$ g ở độ tin cậy 99% thì phải quan sát ít nhất bao nhiêu trái? (Đs: 293)

2.2. KHOẢNG TIN CẬY CHO TỶ LỆ

Định nghĩa 4. Nếu $\hat{p}$ là tỷ lệ mẫu các phần tử thỏa tính chất A quan tâm của một mẫu ngẫu nhiên cỡ n, khoảng tin cậy với độ tin cậy $100(1-\alpha)\%$ cho tỷ lệ p các phần tử thỏa tính chất A của tổng thể là

$$\hat{p} - z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

trong đó $z_{1-\alpha/2}$ thỏa $\mathbb{P}(Z < z_{1-\frac{\alpha}{2}}) = 1 - \dfrac{\alpha}{2}$ với $Z \sim N(0;1)$.

☑ Độ chính xác (sai số) của ước lượng

$$\epsilon = z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

☑ Với độ chính xác ϵ, tỷ lệ p và độ tin cậy $100(1-\alpha)\%$ cho trước, công thức xác định cỡ mẫu

$$n = \left(\frac{z_{1-\alpha/2}}{\epsilon}\right)^2 p(1-p)$$

☑ Nếu muốn **ít nhất** $100(1-\alpha)\%$ độ tin cậy rằng độ chính xác trong ước lượng p bởi $\hat{p}$ bé hơn ϵ thì cỡ mẫu là

$$n = \left(\frac{z_{1-\alpha/2}}{\epsilon}\right)^2 (0.25)$$

BÀI 5.20. Phần nhỏ của các mạch tích hợp khiếm khuyết được tạo ra trong quá trình quang khắc được nghiên cứu. Một mẫu ngẫu nhiên của 300 mạch được kiểm tra đã phát hiện được 13 khiếm khuyết. Tính khoảng tin cậy 95% trên phần mạch bị lỗi do quá trình trên tạo ra. (Đs: $[0.02029; 0.06637]$)

✍ **LỜI GIẢI.**

Ta có $n = 300$; $y = 13 \Rightarrow \hat{p} = \dfrac{y}{n} = \dfrac{13}{300}$.

• **Độ tin cậy:** $95\% \Rightarrow (1-\alpha)\cdot 100\% = 95\%$

$$\Rightarrow \alpha = 0.05.$$

$\Rightarrow z_{1-\frac{\alpha}{2}} = z_{0.975} = 1.96.$

• **Sai số:** $\epsilon = z_{1-\alpha/2}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

$$= 1.96 \cdot \sqrt{\frac{\frac{13}{300}\left(1-\frac{13}{300}\right)}{300}}$$

$\approx 0.023.$

• **Khoảng tin cậy** 95% **cho tỷ lệ mạch bị khiếm khuyết là**

$$\begin{aligned} \hat{p}-\epsilon \le\ & p \le \hat{p}+\epsilon \\ \Leftrightarrow \frac{13}{300} - 0.023 \le\ & p \le \frac{13}{300} + 0.023 \\ \Leftrightarrow 0.0203 \le\ & p \le 0.0663 \end{aligned}$$

$\Rightarrow p \in [0.0203; 0.0663]$. □

BÀI 5.21. Trong số liệu từ cuộc bầu cử tổng thống năm 2004, một bang quan trọng là bang Ohio đã cho kết quả sau đây: đã có 2020 người trả lời trong các cuộc thăm dò xuất cảnh và 768 là sinh viên tốt nghiệp đại học. Trong số các sinh viên tốt nghiệp đại học có 412 bầu cho George Bush. Xây dựng khoảng tin cậy 95% cho tỉ lệ sinh viên tốt nghiệp đại học bầu cho George Bush.(Đs: $[0.501; 0.571]$)

✍ **LỜI GIẢI.**

....

....

....

....

....

....

....

....

....

....

....

....

□

BÀI 5.22. Trong số 1000 trường hợp ung thư phổi được chọn ngẫu nhiên, 823 kết quả tử vong trong vòng 10 năm

a) Tính toán khoảng tin cậy 95% về tỷ lệ tử vong do ung thư phổi.

b) Nếu muốn sai số $\epsilon \leq 0.03$ thì kích thước mẫu khảo sát cần thiết là bao nhiêu? (Đs: 622)

✍ **LỜI GIẢI.**

....

....

....

....

....

....

....

....

....

....

....

....

□

BÀI 5.23. Một loại thuốc mới đem điều trị cho 50 người bị bệnh B, kết quả có 40 người khỏi bệnh.

a) Ước lượng tỷ lệ khỏi bệnh p nếu dùng thuốc đó điều trị với độ tin cậy 0.95 và 0.99.(Đs: $[0.689; 0.911]$; $[0.654; 0.946]$)

b) Nếu muốn sai số ước lượng không quá 0.02 ở độ tin cậy 0.95 thì phải quan sát ít nhất mấy trường hợp ?(Đs: 1537)

BÀI 5.24. Một loại bệnh có tỷ lệ tử vong là 0.01. Muốn chứng tỏ một loại thuốc có hiệu nghiệm (nghĩa là hạ thấp được tỷ lệ tử vong nhỏ hơn 0.005) ở độ tin cậy 0.95 thì phải thử thuốc đó trên ít nhất bao nhiêu người?(Đs: 1522)

....

....

....

....

BÀI 5.25. Biết lương tháng của công nhân (Đv: triệu đồng) trong một nhà máy có phân phối chuẩn. Chọn ngẫu nhiên 16 công nhân khảo sát

Lương tháng	0.8	1.0	1.2	1.3	1.5	1.7	2.0	2.3	2.5
Số công nhân	1	1	2	2	2	3	2	2	1

a) Công nhân gọi là có thu nhập cao nếu lương tháng từ 2 triệu đồng trở lên. Hãy lập khoảng tin cậy 95% cho tỷ lệ công nhân có thu nhập cao.

b) Nếu muốn sai số $\epsilon \leq 0.08$ thì cần khảo sát thêm bao nhiêu công nhân nữa?

CHƯƠNG 6 KIỂM ĐỊNH GIẢ THUYẾT MỘT MẪU

1. BÀI TOÁN KIỂM ĐỊNH GIẢ THUYẾT THỐNG KÊ

1.1. ĐỊNH NGHĨA

Định nghĩa 1. **Giả thuyết thống kê** là một phát biểu về những tham số của một hay nhiều tổng thể. Việc tìm ra kết luận để bác bỏ hoặc chấp nhận một giả thuyết gọi là **kiểm định giả thuyết thống kê**.

VÍ DỤ 1.1. Giám đốc một nhà máy sản xuất bo mạch chủ máy vi tính tuyên bố rằng tuổi thọ của một bo mạch chủ do nhà máy sản xuất ra là 5 năm; đây là một giả thuyết về kỳ vọng của biến ngẫu nhiên $X =$ tuổi thọ của một bo mạch chủ. Để đưa ra kết luận là chấp nhận hay bác bỏ giả thuyết trên, ta cần dựa vào mẫu điều tra và quy tắc kiểm định thống kê.

1.2. GIẢ THUYẾT KHÔNG VÀ ĐỐI THUYẾT

Định nghĩa 2. Trong bài toán kiểm định giả thuyết, giả thuyết cần được kiểm định gọi là **giả thuyết không**, ký hiệu là H_0. Mệnh đề đối lập với H_0 gọi là **đối thuyết**, ký hiệu là H_1.

VÍ DỤ 1.2.

a) Gọi μ là độ thay đổi trung bình trong huyết áp của một bệnh nhân sau khi dùng thuốc, bác sĩ điều trị cần quan tâm đến giả thuyết sau

$$\begin{cases} H_0 : \mu = 0 \text{ Không có ảnh hưởng của thuốc lên huyết áp của bệnh nhân} \\ H_1 : \mu \neq 0 \text{ Có ảnh hưởng của thuốc lên huyết áp của bệnh nhân} \end{cases}$$

b) Một khách hàng quan tâm đến tỷ lệ sản phẩm kém chất lượng trong một lô hàng mua của một nhà cung cấp. Giả sử tỷ lệ sản phẩm kém tối đa được phép là 5%. Khách hàng cần quan tâm đến giả thuyết sau

$$\begin{cases} H_0 : p \geq 0.05 \text{ Tỷ lệ sản phẩm kém cao hơn mức cho phép} \\ H_1 : p < 0.05 \text{ Tỷ lệ sản phẩm kém ở mức cho phép} \end{cases}$$

1.3. CÁCH ĐẶT GIẢ THUYẾT

① Giả thuyết được đặt ra với ý đồ bác bỏ nó, nghĩa là giả thuyết đặt ra ngược lại với điều ta muốn chứng minh, muốn thuyết phục.

② Giả thuyết được đặt ra sao cho khi chấp nhận hay bác bỏ nó sẽ có tác dụng trả lời bài toán thực tế đặt ra.

③ Giả thuyết được đặt ra sao cho nếu nó đúng thì ta sẽ xác định được quy luật phân phối xác suất của đại lượng ngẫu nhiên được chọn làm tiêu chuẩn kiểm định.

④ Khi đặt giả thuyết, ta thường so sánh cái chưa biết với cái đã biết. Cái chưa biết là điều mà ta cần kiểm định, kiểm tra, làm rõ. "Cái đã biết" là những thông tin trong quá khứ, các định mức kinh tế, kỹ thuật.

⑤ Giả thuyết đặt ra thường mang ý nghĩa: "Không khác nhau" hoặc "khác nhau không có ý nghĩa" hoặc "bằng nhau".

Tổng quát, một bài toán kiểm định giả thuyết cho tham số θ sẽ có một trong 3 dạng dưới đây (θ_0 là giá trị kiểm định đã biết):
Hai phía:

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{cases}$$

Một phía bên trái:

$$\begin{cases} H_0 : \theta \geq \theta_0 \\ H_1 : \theta < \theta_0 \end{cases}$$

Một phía bên phải:

$$\begin{cases} H_0 : \theta \leq \theta_0 \\ H_1 : \theta > \theta_0 \end{cases}$$

1.4. CÁC LOẠI SAI LẦM VÀ ĐỘ MẠNH KIỂM ĐỊNH

① **Sai lầm kiểm định**

i) **Sai lầm loại I**: là sai lầm nếu **bác bỏ** giả thuyết H_0 khi H_0 **đúng**.

ii) **Sai lầm loại II**: là sai lầm nếu **chấp nhận** giả thuyết H_0 khi H_0 **sai**.

② **Xác suất sai lầm loại I**

$$\alpha = \mathbb{P}(\text{sai lầm loại I}) = \mathbb{P}(\text{bác bỏ } H_0 \text{ khi } H_0 \text{ đúng}).$$

Giá trị α gọi là **mức ý nghĩa** của bài toán kiểm định.

③ **Xác suất sai lầm loại II**

$$\beta = \mathbb{P}(\text{sai lầm loại II}) = \mathbb{P}(\text{chấp nhận } H_0 \text{ khi } H_0 \text{ sai}).$$

④ **Độ mạnh (power)** của kiểm định thống kê là xác suất mà ta bác bỏ giả thuyết H_0 khi đối thuyết H_1 đúng. Độ mạnh được tính là $1-\beta$. Một tiêu chuẩn kiểm định tốt sẽ có độ mạnh cao.

1.5. $P-$GIÁ TRỊ

Giá trị $p-$giá trị là mức ý nghĩa nhỏ nhất dẫn tới quyết định bác bỏ giả thuyết H_0 ứng với dữ liệu đã cho.
Định lý 1. Nếu $p-$giá trị $< \alpha$ thì bác bỏ giả thuyết H_0.

2. KIỂM ĐỊNH TRUNG BÌNH

2.1. KIỂM ĐỊNH CHO TRUNG BÌNH KHI BIẾT PHƯƠNG SAI

2.1.1. Bài toán

① **Các giả định:**

- ☑ Mẫu ngẫu nhiên $X_1,..., X_n$ được chọn từ tổng thể có phân phối chuẩn $N(\mu;\sigma^2)$ với kỳ vọng μ chưa biết.
- ☑ Phương sai σ^2 đã biết.
- ☑ Cho trước giá trị μ_0, cần so sánh kỳ vọng μ với μ_0.

② **Bài toán có 3 trường hợp**

$$(1)\begin{cases}H_0: \mu=\mu_0\\ H_1: \mu\neq\mu_0\end{cases}\qquad (2)\begin{cases}H_0: \mu=\mu_0\\ H_1: \mu<\mu_0\end{cases}\qquad (3)\begin{cases}H_0: \mu=\mu_0\\ H_1: \mu>\mu_0\end{cases}$$

với mức ý nghĩa α cho trước.

2.1.2. Các bước thực hiện

① Phát biểu giả thuyết không và đối thuyết.

② Xác định mức ý nghĩa α.

③ Tính thống kê kiểm định

$$Z_0=\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$$

④ Xác định bác bỏ H_0 theo bảng dưới đây

Đối thuyết	**Miền bác bỏ H_0**
$H_1 : \mu \neq \mu_0$	$\|z_0\| > z_{1-\alpha/2}$
$H_1 : \mu < \mu_0$	$z_0 < -z_{1-\alpha}$
$H_1 : \mu > \mu_0$	$z_0 > z_{1-\alpha}$

Ngược lại, chưa đủ cơ sở bác bỏ H_0.

⑤ Kết luận.

Ở bước 4, ta có thể sử dụng $p-$giá trị thay thế bằng cách tính $p-$giá trị theo bảng dưới đây

Đối thuyết	**$p-$giá trị**
$H_1 : \mu \neq \mu_0$	$p-$giá trị $= 2\left[1 - \Phi(\|z_0\|)\right]$
$H_1 : \mu < \mu_0$	$p-$giá trị $= \Phi(z_0)$
$H_1 : \mu > \mu_0$	$p-$giá trị $= 1 - \Phi(z_0)$

Bác bỏ H_0 khi $p-$giá trị $< \alpha$. Ngược lại, chưa đủ cơ sở bác bỏ H_0.

BÀI 6.1. Hãy phát biểu giả thuyết H_0 và đối thuyết H_1 trong các trường hợp sau:

a) Kiểm định cung cấp bằng chứng cho thấy trung bình của tổng thể lớn hơn 10.

b) Kiểm định cung cấp bằng chứng cho thấy trung bình của tổng thể không bằng 7.

c) Kiểm định cung cấp bằng chứng cho thấy trung bình của tổng thể nhỏ hơn 10.

....

....

....

BÀI 6.2. Một giả thuyết sẽ được kiểm định rằng trung bình tổng thể bằng 7 có đối thuyết là trung bình khác 7 với phương sai đã biết. Hãy tìm giá trị để đưa ra quyết định ứng với mức ý nghĩa:

a) 0.01 b) 0.05 c) 0.1

....

....

....

....

....

....

BÀI 6.3. Một giả thuyết sẽ được kiểm định rằng trung bình tổng thể bằng 5 có đối thuyết là trung bình nhỏ hơn 5 với phương sai đã biết. Hãy tìm giá trị để đưa ra quyết định ứng với mức ý nghĩa:

a) 0.01 b) 0.05 c) 0.1

BÀI 6.4. Cho giả thuyết kiểm định $H_0 : \mu = 7$ với đối thuyết $H_1 : \mu \neq 7$ và phương sai đã biết, tính $p-$giá trị trong các trường hợp sau:

a) 2.05 b) -1.84 c) 0.4

BÀI 6.5. Cho giả thuyết kiểm định $H_0 : \mu = 5$ với đối thuyết $H_1 : \mu < 5$ và phương sai đã biết, tính $p-$giá trị trong các trường hợp sau:

a) 2.05 b) -1.84 c) 0.4

(Đs: 0.98; 0.03; 0.65)

VÍ DỤ 2.1 (Kiểm định 2 phía). Dây chuyền sản xuất kem đánh răng P/S được thiết kế để đóng hộp những tuýt kem có trọng lượng trung bình là 6 oz (1 oz = 28 g). Một mẫu gồm 30 tuýt kem được chọn ngẫu nhiên để kiểm tra định kỳ. Bộ phận điều khiển dây chuyền phải đảm bảo để trọng lượng trung bình mỗi tuýt kem là 6 oz, nếu nhiều hoặc ít hơn, dây chuyền phải được điều chỉnh lại.
Giả sử trung bình mẫu của 30 tuýt kem là 6.1 oz và độ lệch tiêu chuẩn của tổng thể là $\sigma = 0.2$ oz.
Thực hiện kiểm định giả thuyết với mức ý nghĩa 3% để xác định xem dây chuyền sản xuất có vận hành tốt hay không?

✍ **LỜI GIẢI.**

Gọi X là trọng lượng của một tuýt kem đánh răng (đv: oz), giả sử $X \sim N(\mu; \sigma^2)$.
$n = 30$; $\bar{x} = 6.1$.
$\sigma = 0.2$: đã biết.
Các bước thực hiện như sau:

1) Phát biểu giả thuyết: $\begin{cases} H_0 : \mu = 6 \\ H_1 : \mu \neq 6 \end{cases}$; Kiểm định 2 phía.

2) Xác định mức ý nghĩa: $\alpha = 3\%$.

3) Tính thống kê kiểm định
$$z_0 = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{6.1 - 6}{0.2/\sqrt{30}} = 2.74.$$

4) Xác định miền bác bỏ: Bác bỏ H_0 khi $|z_0| > z_{1-\alpha/2}$.
$\alpha = 3\%$ nên $z_{1-\alpha/2} = z_{0.985} = 2.17$.

5) Kết luận: Ta có:
$$|z_0| > z_{1-\alpha/2}$$
$$\Leftrightarrow 2.74 > 2.17 \text{ (đúng)}$$
$\Rightarrow$ nên bác bỏ H_0.
Ta kết luận: Với 97% độ tin cậy rằng trọng lượng trung bình mỗi tuýt kem không bằng 6, nghĩa là dây chuyền phải được điều chỉnh lại.

Tính $p-$giá trị:
$$p\text{-giá trị} = 2\left[1 - \Phi(|z_0|)\right] = 2\left[1 - \Phi(|2.74|)\right] = 2\left[1 - 0.9969\right] = 0.0062$$
.

5) Kết luận: Vì $p-$giá trị $= 0.0062 < 0.03 = \alpha$ nên bác bỏ H_0.
Ta kết luận: Với 97% độ tin cậy rằng trọng lượng trung bình mỗi tuýt kem không bằng 6, nghĩa là dây chuyền phải được điều chỉnh lại.

□

BÀI 6.6. Một hệ thống tên lửa phản lực sử dụng động cơ đẩy nhiên liệu rắn. Tốc độ cháy của nhiên liệu rắn là một đặc trưng quan trọng của động cơ. Thông số kĩ thuật yêu cầu tốc độ cháy trung bình của thanh nhiên liệu là 50 cm/s. Các kĩ sư biết rằng độ lệch chuẩn của tốc độ cháy là 2 cm/s. Những kĩ sư kiểm nghiệm xác định xác suất của sai lầm loại I hoặc mức ý nghĩa $\alpha = 0.05$ và chọn cỡ mẫu là $n = 25$ với trung bình mẫu tốc độ cháy là $\bar{x} = 51.3$ cm/s. Kết luận rút ra như thế nào? (Đs: bác bỏ)

....

....

....

....

....

....

....

....

....

....

....

....

VÍ DỤ 2.2 (Kiểm định 1 phía). **Metro EMS:** là một bệnh biện tại trung tâm thành phố cung cấp dịch vụ cấp cứu tại nhà. Với khoảng 20 xe cấp cứu, mục tiêu của bệnh viện là cung cấp dịch vụ cấp cứu trong khoảng thời gian trung bình là 12 phút sau khi nhận được điện thoại yêu cầu. Một mẫu ngẫu nhiên gồm thời gian đáp ứng khi có yêu cầu của 40 ca cấp cứu được chọn. Trung bình mẫu là 13.25 phút. Biết rằng độ lệnh chuẩn của tổng thể là $\sigma = 3.2$ phút. Giám đốc EMS muốn thực hiện một kiểm định, với mức ý nghĩa 5%, để xác định xem liệu thời gian một ca cấp cứu có bé hơn hoặc bằng 12 phút hay không?

✍ **LỜI GIẢI.**

1) Phát biểu giả thuyết
   $H_0 : \mu = 12$: Thời gian đáp ứng của dịch vụ cấp cứu đạt yêu cầu.
   $H_1 : \mu > 12$: Thời gian đáp ứng của dịch vụ cấp cứu không đạt yêu cầu.

2) Xác định mức ý nghĩa: $\alpha = 0.05$

3) Tính thống kê kiểm định
$$z_0 = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{13.25 - 12}{3.2/\sqrt{40}} = 2.47$$

4) Xác định miền bác bỏ: Bác bỏ H_0 nếu $z_0 > z_{1-\alpha}$.
   $\alpha = 0.05 \Rightarrow z_{1-\alpha} = z_{0.95} = 1.645$.

5) Kết luận: Ta có

$$z_0 > z_{1-\alpha}$$
$$\Leftrightarrow 2.47 > 1.645 \text{ (đúng)}$$

$\Rightarrow$ bác bỏ H_0.
Ta kết luận: Với 95% độ tin cậy, Metro EMS không đáp ứng được mục tiêu thời gian phục vụ khách hàng từ 12 phút trở xuống.

Sử dụng p−giá trị:
Tính p−giá trị:

$$p\text{-giá trị} = 1 - \Phi(z_0) = 1 - \Phi(2.47) = 1 - 0.9932 = 0.0068$$

.

5) Kết luận: Vì p−giá trị $= 0.0068 < 0.05 = \alpha$ nên bác bỏ H_0.
Ta kết luận: Với 95% độ tin cậy, Metro EMS không đáp ứng được mục tiêu thời gian phục vụ khách hàng từ 12 phút trở xuống.

□

BÀI 6.7. Nhiệt độ nước trung bình hạ lưu từ ống tháp xả giải nhiệt của nhà máy điện không được lớn hơn 100°F. Kinh nghiệm quá khứ đã chỉ ra rằng độ lệch chuẩn của nhiệt độ là 2°F. Nhiệt độ nước được đo trên chín ngày được lựa chọn ngẫu nhiên, và nhiệt độ trung bình được tìm thấy là 98°F.

a) Có bằng chứng gì cho ta thấy nhiệt độ nước có thể chấp nhận được hay không với mức ý nghĩa 0.05?

b) Tính p−giá trị của kiểm định.

....

....

....

....

....

....

....

....

....

....

....

....

BÀI 6.8. Các kĩ sư nghiên cứu độ bền sức kéo của một hợp kim được sử dụng làm trục của gậy đánh golf biết rằng độ bền xấp xỉ phân phối chuẩn với $\sigma = 60$ psi. Một mẫu ngẫu nhiên gồm 12 mẫu vật có trung bình độ bền là 3450 psi.

a) Kiểm định giả thuyết rằng trung bình độ bền là 3500 psi với $\alpha = 0.01$.

b) Giải thích thêm cho kết luận ở phần trên bằng cách sử dụng khoảng tin cậy của μ.

BÀI 6.9. Tuổi thọ của pin được xem như có phân phối xấp xỉ phân phối chuẩn với độ lệch chuẩn $\sigma = 1.25$ h. Một mẫu ngẫu nhiên gồm 10 viên pin có tuổi thọ trung bình là $x = 40.5$ h, $\alpha = 0.05$.

a) Có thêm bằng chứng gì để hỗ trợ cho tuyên bố rằng tuổi thọ của pin không vượt quá 40h?

b) Tính $p-$giá trị cho phép kiểm định ở câu trên.

2.2. KIỂM ĐỊNH CHO TRUNG BÌNH KHI KHÔNG BIẾT PHƯƠNG SAI

2.2.1. Bài toán

① **Các giả định:**

- Mẫu ngẫu nhiên $X_1,..., X_n$ được chọn từ tổng thể có phân phối chuẩn $N(\mu;\sigma^2)$ với kỳ vọng μ và phương sai σ^2 chưa biết.
- Cho trước giá trị μ_0, cần so sánh kỳ vọng μ với μ_0.

② **Bài toán có 3 trường hợp**

$$(1)\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases} \qquad (2)\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu < \mu_0 \end{cases} \qquad (3)\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$$

với mức ý nghĩa α cho trước.

2.2.2. Các bước thực hiện

① Phát biểu giả thuyết không và đối thuyết.

② Xác định mức ý nghĩa α.

③ Tính thông kê kiểm định

$$T_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}.$$

Khi H_0 đúng thì $T_0 \sim t(n-1)$.

④ Xác định bác bỏ H_0 theo bảng dưới đây

Đối thuyết	Miền bác bỏ H_0
$H_1 : \mu \neq \mu_0$	$\lvert t_0 \rvert > t_{\alpha/2;n-1}$
$H_1 : \mu < \mu_0$	$t_0 < -t_{\alpha;n-1}$
$H_1 : \mu > \mu_0$	$t_0 > t_{\alpha;n-1}$

Ngược lại, chưa đủ cơ sở bác bỏ H_0.

5 Kết luận.

Ở bước 4, ta có thể sử dụng $p-$giá trị thay thế bằng cách tính $p-$giá trị theo bảng dưới đây

Đối thuyết	$p-$giá trị
$H_1 : \mu \neq \mu_0$	$p-$giá trị $= 2\mathbb{P}(T \geq \lvert t_0 \rvert)$
$H_1 : \mu < \mu_0$	$p-$giá trị $= \mathbb{P}(T \leq t_0)$
$H_1 : \mu > \mu_0$	$p-$giá trị $= \mathbb{P}(T \geq t_0)$

Bác bỏ H_0 khi $p-$giá trị $< \alpha$. Ngược lại, chưa đủ cơ sở bác bỏ H_0.

BÀI 6.10. Một bài viết về Tăng trưởng: tạp chí dành cho các vấn đề về tăng trưởng bình thường và bất thường: "So sánh tỷ lệ béo và chất béo ước tính được đo lường, chất béo, kali và nitơ của lợn" (Vol. 46, No. 4, 1982, pp. 306–321)] báo cáo kết quả của một nghiên cứu đo trọng lượng cơ thể (tính bằng gam) đối với lợn guinea khi sinh.

421.0	452.6	456.1	494.6	373.8	90.5
110.7	96.4	81.7	102.4	241.0	296.0
317.0	290.9	256.5	447.8	687.6	705.7
879.0	88.8	296.0	273.0	268.0	227.5
279.3	258.5	296.0			

a) Kiểm tra giả thuyết trọng lượng trung bình là 300 gram với $\alpha = 0.05$. Tính giá trị $p-$giá trị.

b) Giải thích thêm cho kết luận ở phần trên bằng cách sử dụng khoảng tin cậy của μ.

....

....

....

....

....

....

....

....

....

BÀI 6.11. Một bài báo năm 1992 trên Tạp chí Hiệp hội Y khoa Hoa Kỳ ("Thẩm định quan trọng 98.6° F, giới hạn trên của nhiệt độ cơ thể bình thường và các di sản khác của Carl Reinhold August Wunderlich") đã báo cáo nhiệt độ cơ thể, giới tính và nhịp tim cho một số đối tượng. Nhiệt độ cơ thể cho 25 đối tượng nữ theo sau:

97.8	97.2	97.4	97.6	97.8	97.9	98.0	98.0	98.0
98.1	98.2	98.3	98.3	98.4	98.4	98.4	98.5	98.6
98.6	98.7	98.8	98.8	98.9	98.9	99.0		

a) Kiểm tra giả thuyết $H_0 : \mu = 98.6$ có đối thuyết $H_1 : \mu \neq 98.6$ với $\alpha = 0.05$.

b) Giải thích thêm cho kết luận ở phần trên bằng cách sử dụng khoảng tin cậy của μ.

BÀI 6.12. Hàm lượng natri của hai mươi hộp bắp hữu cơ 300 gram được xác định. Dữ liệu (tính bằng miligam) như sau:

131.15	130.69	130.91	129.54	129.64	128.77	130.72	128.33	128.24	129.65
130.14	129.29	128.71	129.00	129.39	130.42	129.53	130.12	129.78	130.92

a) Bạn hãy kiểm định giá trị trung bình có khác 130 milligram với $\alpha = 0.05$.

b) Giải thích thêm cho kết luận ở phần trên bằng cách sử dụng khoảng tin cậy của μ.

BÀI 6.13. Đo cholesterol (đơn vị mg%) cho một nhóm người, ta ghi nhận lại được

Chol.	150 – 160	160 – 170	170 – 180	180 – 190	190 – 200	200 – 210
Số người	3	9	11	3	2	1

a) Tính trung bình và phương sai mẫu.

b) Tìm khoảng ước lượng cho trung bình cholesterol trong dân số với độ tin cậy 0.95.

c) Có tài liệu cho biết lượng cholesterol trung bình là 175 mg%. Giá trị này có phù hợp với mẫu quan sát không ? (kết luận với $\alpha = 0.05$).

BÀI 6.14. Một máy đóng gói các sản phẩm có khối lượng 1kg. Nghi ngờ máy hoạt động không bình thường, người ta chọn ra một mẫu ngẫu nhiên gồm 100 sản phẩm thì thấy như sau:

Khối lượng	0.95	0.97	0.99	1.01	1.03	1.05
Số gói	9	31	40	15	3	2

Với mức ý nghĩa 0.05, hãy kết luận về nghi ngờ trên.

BÀI 6.15. Một xí nghiệp đúc một số rất lớn các sản phẩm bằng thép với số khuyết tật trung bình ở mỗi sản phẩm là 3. Người ta cải tiến cách sản xuất và kiểm tra 36 sản phẩm. Kết quả như sau:

Số khuyết tật trên sản phẩm	0	1	2	3	4	5	6
Số sản phẩm tương ứng	7	4	5	7	6	6	1

Giả sử số khuyết tật của các sản phẩm có phân phối chuẩn.

a) Hãy ước lượng số khuyết tật trung bình ở mỗi sản phẩm sau khi cải tiến, với độ tin cậy 90%.

b) Hãy cho kết luận về hiệu quả của việc cải tiến sản xuất với mức ý nghĩa 0.05.

3. KIỂM ĐỊNH CHO TỶ LỆ

3.1. BÀI TOÁN

Cho tổng thể X, trong đó tỷ lệ phần tử mang đặc tính A nào đó trong tổng thể là p chưa biết. Một mẫu dữ liệu $x_1,...,x_n$ đã được thu thập. Hãy kiểm định

$$(1)\begin{cases}H_0: p = p_0\\ H_1: p \neq p_0\end{cases}\qquad (2)\begin{cases}H_0: p = p_0\\ H_1: p < p_0\end{cases}\qquad (3)\begin{cases}H_0: p = p_0\\ H_1: p > p_0\end{cases}$$

với mức ý nghĩa α cho trước. Trong đó, p_0 đã biết.

3.2. CÁC BƯỚC THỰC HIỆN

1. Phát biểu giả thuyết không và đối thuyết.
2. Xác định mức ý nghĩa α.
3. Tính thống kê kiểm định

$$Z_0 = \frac{\hat{P} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$$

4. Xác định bác bỏ H_0 theo bảng dưới đây

Đối thuyết	**Miền bác bỏ H_0**
$H_1: p \neq p_0$	$\lvert z_0\rvert > z_{1-\alpha/2}$
$H_1: p < p_0$	$z_0 < -z_{1-\alpha}$
$H_1: p > p_0$	$z_0 > z_{1-\alpha}$

Ngược lại, chưa đủ cơ sở bác bỏ H_0.

5. Kết luận.

Ở bước 4, ta có thể sử dụng $p-$giá trị thay thế bằng cách tính $p-$giá trị theo bảng dưới đây

Đối thuyết	$p-$**giá trị**
$H_1 : p \neq p_0$	$p-$giá trị $= 2\left[1 - \Phi(\lvert z_0\rvert)\right]$
$H_1 : p < p_0$	$p-$giá trị $= \Phi(z_0)$
$H_1 : p > p_0$	$p-$giá trị $= 1 - \Phi(z_0)$

Bác bỏ H_0 khi $p-$giá trị $< \alpha$. Ngược lại, chưa đủ cơ sở bác bỏ H_0.

VÍ DỤ 3.1. Trong kỳ nghỉ giáng sinh và đầu năm mới, Cục An toàn giao thông đã thống kê được rằng có 500 người chết và 25000 người bị thương do các vụ tai nạn giao thông trên toàn quốc. Theo thông báo của Cục ATGT thì khoảng 50% số vụ tai nạn có liên quan đến rượu bia.
Khảo sát ngẫu nhiên 120 vụ tai nạn thấy có 67 vụ do ảnh hưởng của rượu bia. Sử dụng số liệu trên để kiểm định lời khẳng định của Cục An toàn giao thông với mức ý nghĩa $\alpha = 5\%$.

....

....

....

....

....

....

....

....

....

....

....

....

BÀI 6.16. Một nhà sản xuất chất bán dẫn sản xuất bộ điều khiển sử dụng trong công nghệ động cơ ô tô. Khách hàng yêu cầu phần khiếm khuyết ở các bước sản xuất quan trọng không vượt quá 0.05 và nhà sản xuất chứng minh khả năng xử lý ở mức chất lượng này bằng cách sử dụng $\alpha = 0.05$. Nhà sản xuất chất bán dẫn lấy mẫu ngẫu nhiên gồm 200 thuyết bị và thấy rằng bốn thuyết bị này bị lỗi. Nhà sản xuất có thể chứng minh khả năng xử lý cho khách hàng không? (Đs: kết luận rằng quá trình này là có khả năng)

BÀI 6.17. Giả sử rằng 1000 khách hàng được khảo sát và 850 người hài lòng hoặc rất hài lòng với các sản phẩm và dịch vụ của công ty

a) Kiểm định với giả thuyết $H_0 : p = 0.9$ với đối thuyết $H_1 : p \neq 0.9$ với mức $\alpha = 0.05$. Tìm $p-$giá trị.

b) Giải thích thêm cho kết luận trên bằng cách sử dụng khoảng tin cậy cho p.

BÀI 6.18. Giả sử người ta kiểm tra 500 thành phần máy móc do một nhà máy sản xuất và thấy có 10 thành phần bị loại bỏ. Kiểm định giả thuyết $H_0 : p = 0.03$ với đối thuyết $H_1 : p < 0.03$ và $\alpha = 0.05$. Tìm $p-$giá trị.

BÀI 6.19. Một bài báo trên tạp chí y khoa Anh "So sánh điều trị sỏi thận bằng phẫu thuật phẫu thuật, cắt bỏ sỏi thận, và Lithotrips sóng bổ sung," (1986, Vol. 292, pp. 879–882)] thấy rằng tác động qua da (PN) có tỷ lệ thành công trong việc loại bỏ sỏi thận của 289 trong số 350 bệnh nhân. Phương pháp truyền thống đạt hiệu quả 78%. Có bằng chứng gì cho thấy tỉ lệ thành công của PN lớn hơn so với truyền thống với mức ý nghĩa $\alpha = 1\%$? Tìm $p-$giá trị.

BÀI 6.20. Một nhà nghiên cứu tuyên bố rằng ít nhất 10% của tất cả các mũ bảo hiểm bóng chày có lỗi sản xuất có khả năng có thể gây thương tích cho người đội. Một mẫu 200 mũ bảo hiểm cho thấy 16 mũ bảo hiểm chứa các khuyết tật như vậy.

a) Phát hiện này có ủng hộ tuyên bố của nhà nghiên cứu không? Tính $p-$ giá trị của kiểm định.

b) Giải thích thêm cho kết luận trên bằng cách sử dụng khoảng tin cậy cho p.

BÀI 6.21. Một thiết bị radar mới đang được xem xét cho một hệ thống phòng thủ tên lửa. Hệ thống được kiểm tra bằng cách thử nghiệm với phi cơ trong đó việc tiêu diệt hay không tiêu diệt được mô phỏng. Nếu trong 300 phép thử có 250 lần bị tiêu diệt thì chấp nhận hay bác bỏ, tại mức ý nghĩa 0.04, công bố rằng xác suất tiêu diệt với hệ thống mới không vượt quá xác suất 0.8 của thiết bị đang dùng.

BÀI 6.22. Người ta tin rằng ít nhất 60% cư dân trong một khu vực ủng hộ việc sáp nhập của một thành phố lân cận. Kết luận nào được rút ra nếu chỉ 110 người ủng hộ trong một mẫu 200 cử tri? Sử dụng mức ý nghĩa 0.05.

BÀI 6.23. Một công ty dầu lửa công bố rằng một phần năm căn hộ trong một thành phố nào đó được sưởi ấm bằng dầu lửa. Hỏi ta có lý do để tin rằng có ít hơn một phần năm được sưởi ấm bằng dầu lửa không nếu trong một mẫu ngẫu nhiên 1000 căn hộ trong thành phố này, có 136 căn được sưởi ấm bằng dầu lửa? Sử dụng $p-$giá trị trong kết luận của bạn. Mức ý nghĩa 0.05.

BÀI 6.24. Ở một trường cao đẳng nào đó, người ta ước tính rằng không quá 25% sinh viên chạy xe đạp đi học. Đây có phải là một ước tính hợp lệ nếu trong một mẫu ngẫu nhiên 90 sinh viên cao đẳng có 28 sinh viên được thấy đi xe đạp đi học? Sử dụng mức ý nghĩa 0.05.

BÀI 6.25. Quan sát số hoa hồng bán ra trong một ngày của một cửa hàng bán hoa sau một thời gian, người ta ghi được số liệu sau:

Số hoa hồng (đóa)	12	13	15	16	17	18	19
Số ngày	3	2	7	7	3	2	1

a) Tìm ước lượng điểm của số hoa hồng trung bình bán được trong một ngày với độ tin cậy 0.95.

b) Sau khi tính toán, ông chủ cửa hàng nói rằng nếu trung bình một ngày không bán được 15 đoá hoa thì chẳng thà đóng cửa còn hơn. Dựa vào số liệu trên, anh (chị) hãy kết luận giúp ông chủ cửa hàng xem có nên tiếp tục bán hay không ở mức ý nghĩa 0.05.

c) Giả sử những ngày bán được từ 13 đến 17 đoá hồng là những ngày "bình thường". Hãy ước lượng tỉ lệ của những ngày bình thường của cửa hàng ở độ tin cậy 90%.(Giả thiết rằng số hoa bán ra trong ngày có phân phối chuẩn).

CHƯƠNG 7 HỒI QUY - TƯƠNG QUAN

1. GIỚI THIỆU

Bài toán: trong các hoạt động về khoa học - kỹ thuật, kinh tế, xã hội, . . . ta có nhu cầu xác định mối liên giữa hai hay nhiều biến ngẫu nhiên với nhau. Ví dụ

- ☑ Mối liên hệ giữa chiều cao và cỡ giầy của một người, từ đó một cửa hàng bán giầy dép có thể xác định chính xác cỡ giầy của một khách hàng khi biết chiều cao;
- ☑ Độ giãn nở của một loại vật liệu theo nhiệt độ môi trường;
- ☑ Doanh thu khi bán 1 loại sản phẩm và số tiền chi cho quảng cáo và khuyến mãi;
- ☑ . . .

Để giải quyết các vấn đề trên, ta sử dụng kỹ thuật **phân tích hồi quy** (Regression Analysis).

Phân tích hồi quy được sử dụng để xác định mối liên hệ giữa

- ☑ một biến phụ thuộc Y (biến đáp ứng) và một hay nhiều biến độc lập X_1, X_2, . . . , X_p, các biến này còn được gọi là biến giải thích.
  - Biến phụ thuộc Y phải là biến liên tục;
  - Các biến độc lập X_1, X_2, . . . , X_p, có thể là biến liên tục, rời rạc hoặc phân loại.
- ☑ Mối liên hệ giữa X_1, X_2, . . . , X_p và Y được biểu diễn bởi một hàm tuyến tính.
- ☑ Sự thay đổi trong Y được giả sử do những thay đổi trong X_1, X_2, . . . , X_p gây ra.

Trên cơ sở xác định mối liên hệ giữa biến phụ thuộc Y và các biến giải thích X_1, X_2, . . . , X_p, ta có thể

- ☑ dự đoán, dự báo giá trị của Y,
- ☑ giải thích tác động của sự thay đổi trong các biến giải thích lên biến phụ thuộc.

2. MÔ HÌNH HỒI QUY TUYẾN TÍNH ĐƠN

2.1. ĐỊNH NGHĨA

Định nghĩa 1. Một **mô hình hồi quy tuyến tính đơn** liên quan đến một biến ngẫu nhiên Y và một biến giải thích x là phương trình có dạng

$$Y = \beta_0 + \beta_1 x + \epsilon$$

với

- ☑ β_0 là hệ số chặn của đường thẳng hồi quy.
- ☑ β_1 là hệ số góc của đường thẳng hồi quy.
- ☑ ϵ là thành phần sai số, ϵ được giả sử có phân phối chuẩn với $\mathbb{E}(\epsilon) = 0$ và $Var(\epsilon) = \sigma^2$.

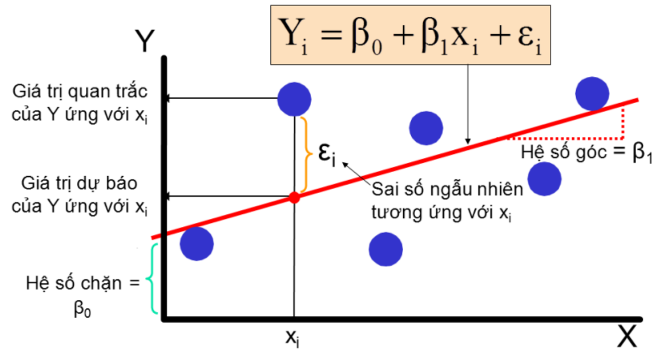
Đường thẳng hồi quy thực sự

$$Y = \beta_0 + \beta_1 x$$

Với $(x_1, y_1), \ldots, (x_n, y_n)$ là n cặp giá trị quan trắc của một mẫu ngẫu nhiên cỡ n, ta có

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad i = 1, \ldots, n$$

Sử dụng đồ thị phân tán để biểu diễn các cặp giá trị quan trắc (x_i, y_i) trên hệ trục tọa độ Oxy



2.2. ƯỚC LƯỢNG CÁC HỆ SỐ HỒI QUY

Gọi $\hat{\beta}_0$ và $\hat{\beta}_1$ là các ước lượng của β_0 và β_1. Khi đó, đường thẳng hồi quy với các hệ số ước lượng

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x.$$

Một đường thẳng ước lượng tốt phải "gần với các điểm dữ liệu".
Để tìm $\hat{\beta}_0$ và $\hat{\beta}_1$, ta dùng phương pháp bình phương bé nhất (PPBPBN).
Với dữ liệu (x_i, y_i), $i = 1, \ldots, n$, ta có

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_i.$$

Độ lệch giữa giá trị quan trắc y_i và giá trị dự đoán $\hat{y}_i$ gọi là **giá trị thặng dư** thứ i, xác định như sau

$$e_i = y_i - \hat{y}_i = y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)$$

Định nghĩa 2. **Tổng bình phương sai số** (Sum of Squares for Errors - SSE) hay tổng bình phương thặng dư cho n điểm dữ liệu được định nghĩa như sau

$$SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left[y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)\right]^2$$

Nội dung của PPBPBN là tìm các ước lượng $\hat{\beta}_0$ và $\hat{\beta}_1$ sao cho SSE đạt giá trị bé nhất.

Lấy đạo hàm theo β_0 và β_1

$$\frac{\partial SSE}{\partial \beta_0} = -2\sum_{i=1}^{n}\left[y_i - (\beta_0 + \beta_1 x_i)\right] = 0$$
$$\frac{\partial SSE}{\partial \beta_1} = -2\sum_{i=1}^{n}\left[y_i - (\beta_0 + \beta_1 x_i)\right] x_i = 0$$

ta thu được hệ phương trình

$$n\beta_0 + \beta_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$
$$\beta_0 \sum_{i=1}^{n} x_i + \beta_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i$$

Giải hệ trên, ta tìm được các ước lượng BPBN của β_0 và β_1 là

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - \dfrac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n}}{\sum_{i=1}^{n} x_i^2 - \dfrac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}} = \frac{S_{xy}}{S_{xx}}$$
$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

với S_{xy} và S_{xx} xác định bởi

$$S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n}$$
$$S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}$$

trong đó $\bar{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$ và $\bar{y} = \dfrac{1}{n}\sum_{i=1}^{n} y_i$.

- ☑ Các ước lượng $\hat{\beta}_0$ và $\hat{\beta}_1$ tìm được gọi là các **ước lượng BPBN**.
- ☑ Đường thẳng $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ gọi là đường thẳng BPBN, thỏa các tính chất sau

  i) $SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$ đạt giá trị bé nhất.

  ii) $SE = \sum_{i=1}^{n} (y_i - \hat{y}_i) = \sum_{i=1}^{n} e_i = 0$ với SE là tổng các thặng dư (Sum of Errors).

VÍ DỤ 2.1 (Số năm sử dụng và giá xe). Một công ty sản xuất ô tô muốn điều tra giá của một dòng xe của họ giảm giá thế nào theo số năm sử dụng. Bộ phận nghiên cứu của công ty đã lấy mẫu gồm 8 chiếc của dòng xe này và thu thập thông tin về số năm sử dụng (X - năm) và giá của xe (Y - triệu \$) được mô tả như sau

$$\sum_{i=1}^{n} x_i = 41; \quad \sum_{i=1}^{n} x_i^2 = 259; \quad \sum_{i=1}^{n} y_i = 1144;$$

$$\sum_{i=1}^{n} y_i^2 = 221772; \quad \sum_{i=1}^{n} x_i y_i = 4197.$$

a) Tìm đường thẳng hồi quy biểu diễn mối liên hệ giữa giá của xe Y theo số năm sử dụng X.

Hướng dẫn: Tính S_{xy}, $S_{xx} \Rightarrow \hat{\beta}_1$; $\bar{x}$, $\bar{y} \Rightarrow \hat{\beta}_0 \Rightarrow \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$

....

....

....

....

....

....

....

....

....

2.3. DỰ ĐOÁN GIÁ TRỊ QUAN TRẮC MỚI

Giả sử với giá trị x_0, ta cần dự đoán giá trị quan trắc Y_0 trong tương lai tương ứng với x_0 bằng bao nhiêu. Từ mô hình hồi quy, ta có

$$\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$$

$\hat{Y}_0$ là một ước lượng điểm của giá trị quan trắc mới Y.

2.4. Ý NGHĨA CỦA CÁC HỆ SỐ HỒI QUY

Từ mô hình hồi quy, ta có

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x.$$

Ý nghĩa của các hệ số hồi quy

- ☑ $\hat{\beta}_0$: Khi các yếu tố bị loại khỏi mô hình ($x = 0$), thì trung bình của Y là $\hat{\beta}_0$.

☑ $\hat{\beta}_1$: khi x thay đổi 1 đơn vị, các yếu tố khác không đổi thì Y thay đổi $\hat{\beta}_1$ đơn vị.

Tiếp tục Ví dụ 2.1 (Số năm sử dụng và giá xe) Từ câu a) ta có $\hat{y} = 317.69567 - 34.08696x$

b) Dựa vào đường thẳng hồi quy vừa ước lượng, hãy cho biết giá của chiếc xe sau 3 năm sử dụng?

c) Dựa vào đường thẳng hồi quy vừa ước lượng, giả sử các yếu tố khác không đổi thì giá của xe sẽ thay đổi như thế nào?

....

....

....

....

....

3. HỆ SỐ XÁC ĐỊNH VÀ HỆ SỐ TƯƠNG QUAN

3.1. ĐO SỰ BIẾN THIÊN CỦA DỮ LIỆU

Gọi

☑ SST: Tổng bình phương toàn phần (Total Sum of Squares)

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

☑ SSR: Tổng bình phương hồi quy (Regression Sum of Squares)

$$SSR = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

☑ SSE: Tổng bình phương sai số (Error Sum of Squares)

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

trong đó

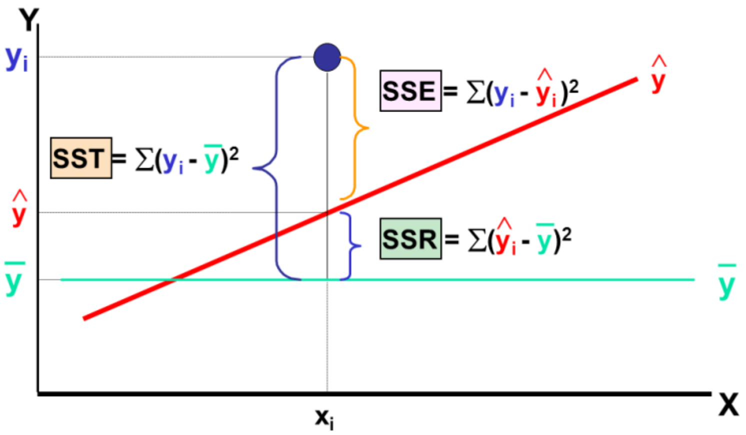
☑ SST: đo sự biến thiên của các giá trị y_i xung quanh giá trị trung tâm của dữ liệu $\bar{y}$.

☑ SSR: giải thích sự biến thiên liên quan đến mối quan hệ tuyến tính của X và Y.

☑ SSE: giải thích sự biến thiên của các nhân tố khác (không liên quan đến mối quan hệ tuyến tính của X và Y).

Ta có

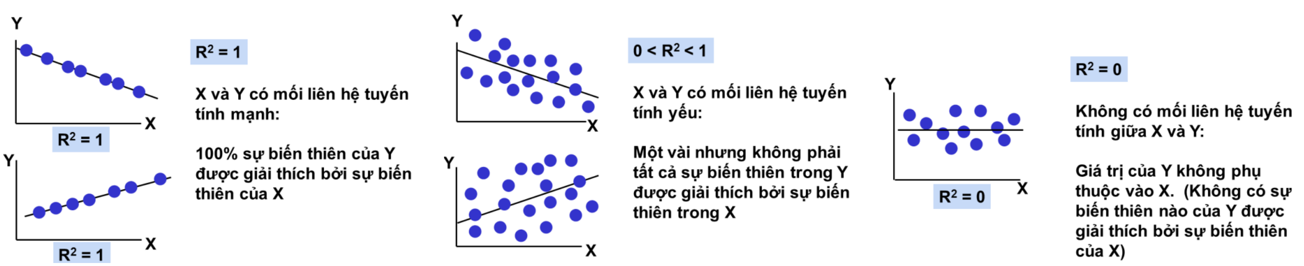
$$\begin{aligned}\sum_{i=1}^{n}(y_i-\bar{y})^2 &= \sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2+\sum_{i=1}^{n}(y_i-\hat{y}_i)^2\\ SST &= \quad SSR \quad + \quad SSE.\end{aligned}$$



3.2. HỆ SỐ XÁC ĐỊNH

> **Định nghĩa 3. Hệ số xác định** (Coefficient of Determination) là tỷ lệ của tổng sự biến thiên trong biến phụ thuộc gây ra bởi sự biến thiên của các biến độc lập (biến giải thích) so với tổng sự biến thiên toàn phần. Hệ số xác định thường được gọi là $R-$bình phương ($R-$squared), ký hiệu là R^2. Công thức tính:
> $$R^2 = \frac{SSR}{SST}.$$

Chú ý: $0 \leq R^2 \leq 1$.
Hệ số xác định của một mô hình hồi quy cho phép ta đánh giá mô hình tìm được có giải thích tốt cho mối liên hệ giữa biến phụ thuộc Y và biến phụ thuộc X hay không.



3.3. HỆ SỐ TƯƠNG QUAN MẪU

Định nghĩa 4. Với mẫu ngẫu nhiên cỡ n: (x_i, y_i), $i = 1, \ldots, n$. **Hệ số tượng quan mẫu**, ký hiệu r_{xy}, được xác định như sau

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} = \frac{S_{xy}}{\sqrt{S_{xx} SST}}$$

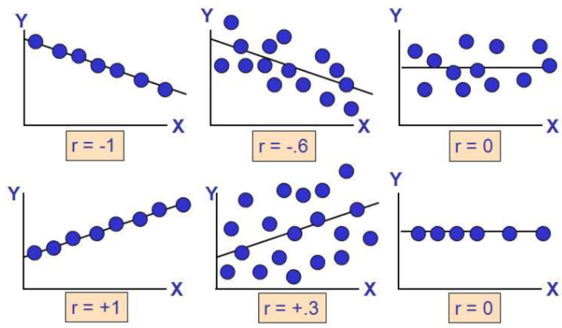
trong đó $SST = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$.

Hệ số xác định R^2 của mô hình hồi quy tuyến tính đơn bằng với bình phương của hệ số tương quan mẫu

$$R^2 = r_{xy}^2$$

Đánh giá hệ số tương quan:

- Miền giá trị: $-1 \leq r_{xy} \leq 1$.
- $-1 \leq r_{xy} < 0$: tương quan âm. r_{xy} càng gần -1 biểu thị mối liên hệ tuyến tính nghịch giữa X và Y càng mạnh.
- $0 < r_{xy} \leq 1$: tương quan dương. r_{xy} càng gần 1 biểu thị mối liên hệ tuyến tính thuận giữa X và Y càng mạnh.
- r_{xy} càng gần 0 biểu thị mối liên hệ tuyến tính yếu. $r_{xy} = 0$: không có mối liên hệ tuyến tính giữa X và Y.



Tiếp tục Ví dụ 2.1 (Số năm sử dụng và giá xe) Ta có

$$n = 8; S_{xy} = -1666; S_{xx} = 48.875; \hat{\beta}_1 = -34.08696; \sum_{i=1}^{n} y_i = 1144; \sum_{i=1}^{n} y_i^2 = 221772$$

d) Tính hệ số xác định R^2 và hệ số tương quan r_{xy}^2.

- **Tính tổng các sai số**

- $SST = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} = 221772 - \frac{1144^2}{8} = 58180.$

- ☑ $SSE = SST - \hat{\beta}_1 Sxy = 58180 - \left(-\dfrac{-1666}{48.875}\right)\cdot(-1666) \approx 1391.13.$

- ☑ $SSR = SST - SSE = 58180 - 1391.130435 \approx 56788.87.$

$\Rightarrow$ Hệ số xác định $R^2 = \dfrac{SSR}{SST} \approx \dfrac{56788.87}{58180} \approx 0.98.$
Như vậy, mô hình tuyến tính giải thích được 98% dữ liệu quan trắc.
$\Rightarrow$ Hệ số tương quan mẫu $r_{xy} = \dfrac{S_{xy}}{\sqrt{S_{xx}\cdot SST}} = \dfrac{-1666}{\sqrt{48.875\cdot 58180}} \approx -0.99.$
Như vậy mối quan hệ tuyến tính nghịch giữa số năm sử dụng và giá của xe tương ứng là rất mạnh.
BÀI 7.1. Dữ liệu sau đây cung cấp thông tin về kinh nghiệm (X - năm) và lương hàng tháng (Y - trăm \$) của 9 thư ký được chọn ngẫu nhiên.

$$\sum_{i=1}^{n} x_i = 80; \quad \sum_{i=1}^{n} x_i^2 = 968; \quad \sum_{i=1}^{n} y_i = 425;$$

$$\sum_{i=1}^{n} y_i^2 = 21841; \quad \sum_{i=1}^{n} x_i y_i = 4404.$$

a) Tính những ước lượng bình phương tối thiểu cho hệ số góc và tung độ góc.

b) Sử dụng đường thẳng hồi quy vừa ước lượng, hãy dự đoán mức lương hàng tháng của một thư ký có 4 năm kinh nghiệm?

c) Dựa vào đường thẳng hồi quy vừa ước lượng, giả sử các yếu tố khác không đổi thì mức lương hàng tháng của một thư ký sẽ thay đổi như thế nào?

d) Tính hệ số xác định và hệ số tương quan mẫu.

....

....

....

....

....

....

....

....

....

....

....

....

VÍ DỤ 3.1. Một công ty sản xuất ô tô muốn điều tra giá của một dòng xe của họ giảm giá thế nào theo số năm sử dụng. Bộ phận nghiên cứu của công ty đã lấy mẫu gồm 8 chiếc của dòng xe này và thu thập thông tin về số năm sử dụng (X - năm) và giá của xe (Y - triệu \$) được mô tả ở bảng dưới đây

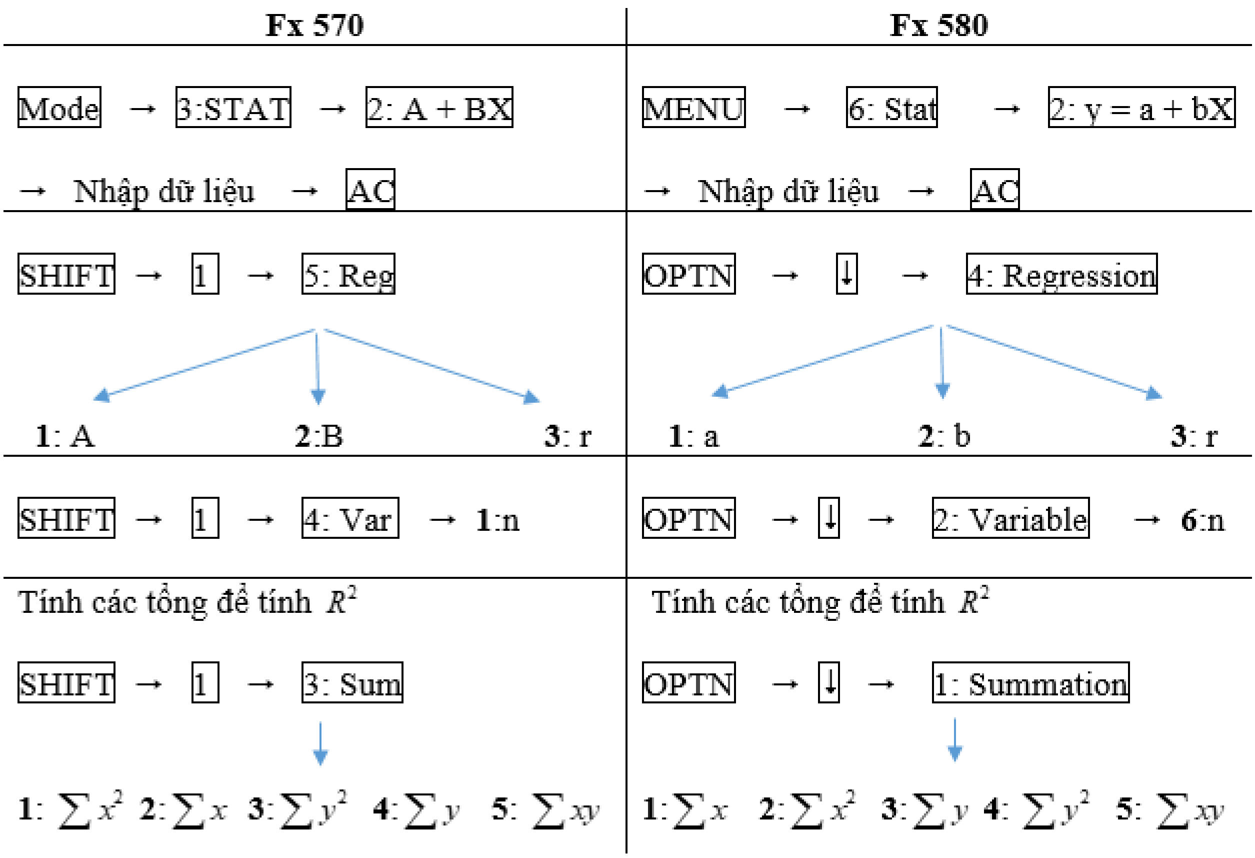
Số năm sử dụng	6	3	8	9	2	5	6	2
Giá xe	112	220	38	33	267	134	95	245

a) Tìm đường thẳng hồi quy biểu diễn mối liên hệ giữa giá của xe Y theo số năm sử dụng X.

b) Dựa vào đường thẳng hồi quy vừa ước lượng, hãy cho biết giá của chiếc xe sau 2 năm sử dụng?

c) Dựa vào đường thẳng hồi quy vừa ước lượng, giả sử các yếu tố khác không đổi thì giá của xe sẽ thay đổi như thế nào?

d) Tính hệ số xác định và hệ số tương quan mẫu.



....

....

....

....

....

BÀI 7.2. Để thực hiện một công trình nghiên cứu về mối quan hệ giữa chiều cao (Y - mét) và đường kính (X - cm) của một loại cây, người ta quan sát trên một mẫu ngẫu nhiên và có kết quả sau:

X	26	28	23	25	31	37	29	42
Y	4	5	4	4	7	8	6	9

a) Tìm đường thẳng hồi quy biểu diễn mối liên hệ giữa chiều cao Y theo đường kính X.

b) Dựa vào đường thẳng hồi quy vừa ước lượng, hãy cho biết chiều cao của cây nếu có đường kính 30 cm?

c) Dựa vào đường thẳng hồi quy vừa ước lượng, giả sử các yếu tố khác không đổi thì chiều cao của cây sẽ thay đổi như thế nào?

....

....

....

....

....

....

....

....

BÀI 7.3. Thu thập về kinh nghiệm lái xe (X - năm) và phí bảo hiểm ô tô hàng tháng (Y - \$) của 8 tài xế ô tô được chọn ngẫu nhiên từ một thị trấn nhỏ có tham gia bảo hiểm ô tô, ta được dữ liệu như sau

$$\sum_{i=1}^{8} x_i = 90; \quad \sum_{i=1}^{8} x_i^2 = 1396; \quad \sum_{i=1}^{8} y_i = 474;$$

$$\sum_{i=1}^{8} y_i^2 = 29642; \quad \sum_{i=1}^{8} x_i y_i = 4739$$

a) Tìm phương trình hồi quy tuyến tính ước lượng phí bảo hiểm ô tô hàng tháng Y theo kinh nghiệm lái xe X.

b) Dựa vào đường thẳng hồi quy ước lượng ở trên, hãy cho biết phí bảo hiểm ô tô hàng tháng của một tài xế có 8 năm kinh nghiệm.

c) Dựa vào đường thẳng hồi quy ước lượng ở trên, nếu các yếu tố khác không đổi thì phí bảo hiểm ô tô hàng tháng của một tài xế Y sẽ thay đổi như thế nào?

d) Tìm hệ số xác định và hệ số tương quan mẫu.

BÀI 7.4. Với một mẫu gồm 20 lần quan sát hàng tháng, một nhà phân tích tài chính muốn thực hiện hồi quy tỷ lệ phần trăm lợi nhuận (Y - %) của cổ phiếu phổ thông của một công ty theo tỷ lệ phần trăm lợi nhuận (X - %) của chỉ số S&P 500 (Standard & Poor's 500). Kết quả được tổng hợp bên dưới

$$\sum_{i=1}^{20} x_i = 25.4; \quad \sum_{i=1}^{20} x_i^2 = 145.7; \quad \sum_{i=1}^{20} y_i = 22.6;$$

$$\sum_{i=1}^{20} y_i^2 = 196.2; \quad \sum_{i=1}^{20} x_i y_i = 150.5$$

a) Tính những ước lượng bình phương tối tiểu cho hệ số góc và tung độ gốc.

b) Dựa vào đường thẳng hồi quy ước lượng ở trên, hãy cho biết giá trị của y khi $x = 1.46\%$.

c) Dựa vào đường thẳng hồi quy ước lượng ở trên, nếu các yếu tố khác không đổi thì tỷ lệ phần trăm lợi nhuận (Y) sẽ thay đổi như thế nào?

d) Tìm hệ số xác định và hệ số tương quan mẫu.

BÀI 7.5. Một công ty tổ chức một bài kiểm tra năng lực cho tất cả các đại diện bán hàng mới vào công ty. Ban quản lý công ty cho rằng những người đại diện bán hàng có điểm năng lực càng cao sẽ có khả năng bán hàng càng tốt. Bảng dưới đây ghi lại doanh số bán hàng trung bình hàng tuần (Y - nghìn đô la) và điểm kiểm tra năng lực (X - điểm) với một mẫu ngẫu nhiên gồm 8 đại diện bán hàng mới.

Doanh số hàng tuần	10	12	28	24	18	16	15	12
Điểm năng lực	55	60	85	75	80	85	65	60

a) Ước lượng đường thẳng hồi quy tuyến tính của doanh số bán hàng hàng tuần (Y) theo điểm năng lực (X).

b) Dựa vào đường thẳng hồi quy ước lượng ở trên, hãy cho biết doanh số bán hàng hàng tuần của một đại diện bán hàng mới có điểm năng lực là 87 điểm.

c) Dựa vào đường thẳng hồi quy ước lượng ở trên, nếu các yếu tố khác không đổi thì doanh số bán hàng hàng tuần (Y) sẽ thay đổi như thế nào?

d) Tìm hệ số xác định và hệ số tương quan mẫu.

BÀI 7.6. Chủ một xưởng sản xuất găng tay nhỏ quan tâm đến chi phí điều hòa sẽ cao vào mùa hè nhưng cũng sợ nhiệt độ trong xưởng quá cao sẽ làm giảm năng suất. Trong suốt mùa hè, người chủ đã thử nghiệm với các cài đặt nhiệt độ từ 68°F đến 81°F và đo năng suất mỗi ngày. Bảng sau đây cho biết nhiệt độ trong xưởng (X - °F) và số lượng đôi găng tay (Y - tính theo đơn vị trăm đôi) được sản xuất vào mỗi ngày trong 8 ngày được chọn ngẫu nhiên.

Nhiệt độ (°F)	72	71	78	75	81	77	68	76
Số đôi găng tay (trăm đôi)	37	37	32	36	33	35	39	34

a) Tìm phương trình hồi quy tuyến tính ước lượng số đôi găng tay Y theo nhiệt độ trong xưởng X.

b) Dựa vào đường thẳng hồi quy ước lượng ở trên, hãy cho biết số đôi găng tay được sản xuất khi nhiệt độ trong xưởng là 74°F.

c) Dựa vào đường thẳng hồi quy ước lượng ở trên, nếu các yếu tố khác không đổi thì số đôi găng tay được sản xuất Y sẽ thay đổi như thế nào?

d) Tìm hệ số xác định và hệ số tương quan mẫu.

BÀI 7.7. Bệnh tiểu đường và béo phì là những vấn đề sức khỏe nghiêm trọng ở Hoa Kỳ và phần lớn các nước phát triển. Đo lượng mỡ cơ thể của một người là một cách để theo dõi tiến độ kiểm soát cân nặng, nhưng đo chính xác nó phải sử dụng đến thiết bị X-quang đắt tiền hoặc nhúng cơ thể xuống một hồ bơi. Thay vào đó, chỉ số khối cơ thể (BMI) thường được sử dụng làm đại diện cho mỡ cơ thể vì nó dễ đo: BMI = khối lượng (kg)/(chiều cao (m))^2 = 703 khối lượng (lb)/(chiều cao (in))^2. Trong một nghiên cứu của 250 người đàn ông tại Đại học Bingham Young, cả BMI (X) và mỡ cơ thể (Y) được đo lường. Các nhà nghiên cứu đã tìm thấy các thống kê tóm tắt sau:

$$\sum_{i=1}^{n} x_i = 6322.28; \quad \sum_{i=1}^{n} x_i^2 = 162674.18; \quad \sum_{i=1}^{n} y_i = 4757.90;$$

$$\sum_{i=1}^{n} y_i^2 = 107679.27; \quad \sum_{i=1}^{n} x_i y_i = 125471.10$$

a) Tính những ước lượng bình phương tối thiểu cho hệ số góc và tung độ góc.

b) Sử dụng đường thẳng hồi quy, hãy tiên đoán lượng mỡ cơ thể của một người đàn ông sẽ được quan trắc nếu có chỉ số BMI là 30?

c) Xác định hệ số tương quan mẫu r_{xy}.

d) Tính hệ số xác định R^2.

BÀI 7.8. Một bài báo trong Nghiên cứu Bê tông "Đặc tính bề mặt gần bê tông: Tính thấm nội tại" (1989, Tập 41) trình bày dữ liệu về cường độ nén x và độ thấm nội tại y của các hỗn hợp bê tông và phương pháp xử lý khác nhau. Số liệu được tóm tắt như sau:

$$n = 14; \quad \sum_{i=1}^{n} y_i = 572; \quad \sum_{i=1}^{n} y_i^2 = 23530; \quad \sum_{i=1}^{n} x_i = 43;$$

$$\sum_{i=1}^{n} x_i^2 = 157.42; \quad \sum_{i=1}^{n} x_i y_i = 1697.80$$

a) Tính những ước lượng bình phương tối thiểu cho hệ số góc và tung độ góc.

b) Sử dụng đường thẳng hồi quy, hãy tiên đoán lượng độ thấm nội sẽ quan trắc được khi cường độ nén là $x = 4.3$?

c) Xác định hệ số tương quan mẫu r_{xy}.

d) Tính hệ số xác định R^2.

BÀI 7.9. Các phương pháp hồi quy đã được sử dụng để phân tích dữ liệu từ một nghiên cứu điều tra mối quan hệ giữa nhiệt độ bề mặt đường (x) và độ lún mặt đường (y). Số liệu được tóm tắt như sau

$$n = 20; \quad \sum_{i=1}^{n} y_i = 12.75; \quad \sum_{i=1}^{n} y_i^2 = 8.86; \quad \sum_{i=1}^{n} x_i = 1478;$$

$$\sum_{i=1}^{n} x_i^2 = 143215.8; \quad \sum_{i=1}^{n} x_i y_i = 1083.67$$

a) Tính những ước lượng bình phương tối thiểu cho hệ số góc và tung độ góc.

b) Sử dụng đường thẳng hồi quy, hãy tiên đoán lượng độ lún mặt đường sẽ quan trắc được khi nhiệt độ bề mặt đường là 85°F?

c) Xác định hệ số tương quan mẫu r_{xy}.

d) Tính hệ số xác định R^2.

BÀI 7.10. Điểm thi giữa kỳ (x) và cuối kỳ (y) của một lớp có 9 sinh viên là như sau

x	77	50	71	72	81	94	96	99	67
y	82	66	78	34	47	85	99	99	68

a) Ước lượng đường hồi quy tuyến tính.

b) Ước lượng điểm bài thi cuối kỳ của một sinh viên có điểm giữa kỳ là 85.

c) Xác định hệ số tương quan mẫu r_{xy}.

d) Tính hệ số xác định R^2.

BÀI 7.11. Một nghiên cứu về khối lượng đường bị biến đổi trong một quá trình nào đó ở các nhiệt độ khác nhau. Dữ liệu được mã hóa và ghi lại như sau

Nhiệt độ (x)	Đường bị biến đổi (y)
1.0	8.1
1.1	7.8
1.2	8.5
1.3	9.8
1.4	9.5
1.5	8.9
1.6	8.6
1.7	10.2
1.8	9.3
1.9	9.2
2.0	10.5

a) Ước lượng đường hồi quy tuyến tính.

b) Ước lượng khối lượng trung bình của đường bị biến đổi được tạo ra khi nhiệt độ được mã hóa là 1.75.

c) Xác định hệ số tương quan mẫu r_{xy}.

d) Tính hệ số xác định R^2.

BÀI 7.12. Khối lượng của một hợp chất hóa học y hòa tan trong 100 gram nước ở các nhiệt độ khác nhau x được ghi lại như sau

x (°C)	y (gram)		
0	8	6	8
15	12	10	14
30	25	21	24
45	31	33	28
60	44	39	42
75	48	51	44

a) Tìm phương trình của đường thẳng hồi quy.

b) Ước lượng khối lượng của hợp chất hóa học sẽ hòa tan trong 100 gram nước ở 50°.

c) Xác định hệ số tương quan mẫu r_{xy}.

d) Tính hệ số xác định R^2.

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} \, du$$

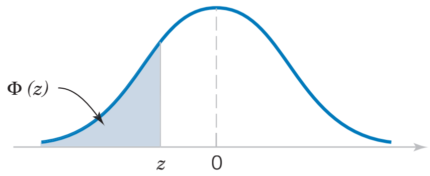


Table II Cumulative Standard Normal Distribution

z	−0.09	−0.08	−0.07	−0.06	−0.05	−0.04	−0.03	−0.02	−0.01	−0.00
−3.9	0.000033	0.000034	0.000036	0.000037	0.000039	0.000041	0.000042	0.000044	0.000046	0.000048
−3.8	0.000050	0.000052	0.000054	0.000057	0.000059	0.000062	0.000064	0.000067	0.000069	0.000072
−3.7	0.000075	0.000078	0.000082	0.000085	0.000088	0.000092	0.000096	0.000100	0.000104	0.000108
−3.6	0.000112	0.000117	0.000121	0.000126	0.000131	0.000136	0.000142	0.000147	0.000153	0.000159
−3.5	0.000165	0.000172	0.000179	0.000185	0.000193	0.000200	0.000208	0.000216	0.000224	0.000233
−3.4	0.000242	0.000251	0.000260	0.000270	0.000280	0.000291	0.000302	0.000313	0.000325	0.000337
−3.3	0.000350	0.000362	0.000376	0.000390	0.000404	0.000419	0.000434	0.000450	0.000467	0.000483
−3.2	0.000501	0.000519	0.000538	0.000557	0.000577	0.000598	0.000619	0.000641	0.000664	0.000687
−3.1	0.000711	0.000736	0.000762	0.000789	0.000816	0.000845	0.000874	0.000904	0.000935	0.000968
−3.0	0.001001	0.001035	0.001070	0.001107	0.001144	0.001183	0.001223	0.001264	0.001306	0.001350
−2.9	0.001395	0.001441	0.001489	0.001538	0.001589	0.001641	0.001695	0.001750	0.001807	0.001866
−2.8	0.001926	0.001988	0.002052	0.002118	0.002186	0.002256	0.002327	0.002401	0.002477	0.002555
−2.7	0.002635	0.002718	0.002803	0.002890	0.002980	0.003072	0.003167	0.003264	0.003364	0.003467
−2.6	0.003573	0.003681	0.003793	0.003907	0.004025	0.004145	0.004269	0.004396	0.004527	0.004661
−2.5	0.004799	0.004940	0.005085	0.005234	0.005386	0.005543	0.005703	0.005868	0.006037	0.006210
−2.4	0.006387	0.006569	0.006756	0.006947	0.007143	0.007344	0.007549	0.007760	0.007976	0.008198
−2.3	0.008424	0.008656	0.008894	0.009137	0.009387	0.009642	0.009903	0.010170	0.010444	0.010724
−2.2	0.011011	0.011304	0.011604	0.011911	0.012224	0.012545	0.012874	0.013209	0.013553	0.013903
−2.1	0.014262	0.014629	0.015003	0.015386	0.015778	0.016177	0.016586	0.017003	0.017429	0.017864
−2.0	0.018309	0.018763	0.019226	0.019699	0.020182	0.020675	0.021178	0.021692	0.022216	0.022750
−1.9	0.023295	0.023852	0.024419	0.024998	0.025588	0.026190	0.026803	0.027429	0.028067	0.028717
−1.8	0.029379	0.030054	0.030742	0.031443	0.032157	0.032884	0.033625	0.034379	0.035148	0.035930
−1.7	0.036727	0.037538	0.038364	0.039204	0.040059	0.040929	0.041815	0.042716	0.043633	0.044565
−1.6	0.045514	0.046479	0.047460	0.048457	0.049471	0.050503	0.051551	0.052616	0.053699	0.054799
−1.5	0.055917	0.057053	0.058208	0.059380	0.060571	0.061780	0.063008	0.064256	0.065522	0.066807
−1.4	0.068112	0.069437	0.070781	0.072145	0.073529	0.074934	0.076359	0.077804	0.079270	0.080757
−1.3	0.082264	0.083793	0.085343	0.086915	0.088508	0.090123	0.091759	0.093418	0.095098	0.096801
−1.2	0.098525	0.100273	0.102042	0.103835	0.105650	0.107488	0.109349	0.111233	0.113140	0.115070
−1.1	0.117023	0.119000	0.121001	0.123024	0.125072	0.127143	0.129238	0.131357	0.133500	0.135666
−1.0	0.137857	0.140071	0.142310	0.144572	0.146859	0.149170	0.151505	0.153864	0.156248	0.158655
−0.9	0.161087	0.163543	0.166023	0.168528	0.171056	0.173609	0.176185	0.178786	0.181411	0.184060
−0.8	0.186733	0.189430	0.192150	0.194894	0.197662	0.200454	0.203269	0.206108	0.208970	0.211855
−0.7	0.214764	0.217695	0.220650	0.223627	0.226627	0.229650	0.232695	0.235762	0.238852	0.241964
−0.6	0.245097	0.248252	0.251429	0.254627	0.257846	0.261086	0.264347	0.267629	0.270931	0.274253
−0.5	0.277595	0.280957	0.284339	0.287740	0.291160	0.294599	0.298056	0.301532	0.305026	0.308538
−0.4	0.312067	0.315614	0.319178	0.322758	0.326355	0.329969	0.333598	0.337243	0.340903	0.344578
−0.3	0.348268	0.351973	0.355691	0.359424	0.363169	0.366928	0.370700	0.374484	0.378281	0.382089
−0.2	0.385908	0.389739	0.393580	0.397432	0.401294	0.405165	0.409046	0.412936	0.416834	0.420740
−0.1	0.424655	0.428576	0.432505	0.436441	0.440382	0.444330	0.448283	0.452242	0.456205	0.460172
0.0	0.464144	0.468119	0.472097	0.476078	0.480061	0.484047	0.488033	0.492022	0.496011	0.500000

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} \, du$$

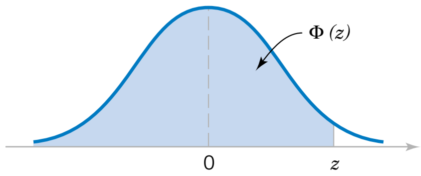


Table II Cumulative Standard Normal Distribution (*continued*)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.500000	0.503989	0.507978	0.511967	0.515953	0.519939	0.532922	0.527903	0.531881	0.535856
0.1	0.539828	0.543795	0.547758	0.551717	0.555760	0.559618	0.563559	0.567495	0.571424	0.575345
0.2	0.579260	0.583166	0.587064	0.590954	0.594835	0.598706	0.602568	0.606420	0.610261	0.614092
0.3	0.617911	0.621719	0.625516	0.629300	0.633072	0.636831	0.640576	0.644309	0.648027	0.651732
0.4	0.655422	0.659097	0.662757	0.666402	0.670031	0.673645	0.677242	0.680822	0.684386	0.687933
0.5	0.691462	0.694974	0.698468	0.701944	0.705401	0.708840	0.712260	0.715661	0.719043	0.722405
0.6	0.725747	0.729069	0.732371	0.735653	0.738914	0.742154	0.745373	0.748571	0.751748	0.754903
0.7	0.758036	0.761148	0.764238	0.767305	0.770350	0.773373	0.776373	0.779350	0.782305	0.785236
0.8	0.788145	0.791030	0.793892	0.796731	0.799546	0.802338	0.805106	0.807850	0.810570	0.813267
0.9	0.815940	0.818589	0.821214	0.823815	0.826391	0.828944	0.831472	0.833977	0.836457	0.838913
1.0	0.841345	0.843752	0.846136	0.848495	0.850830	0.853141	0.855428	0.857690	0.859929	0.862143
1.1	0.864334	0.866500	0.868643	0.870762	0.872857	0.874928	0.876976	0.878999	0.881000	0.882977
1.2	0.884930	0.886860	0.888767	0.890651	0.892512	0.894350	0.896165	0.897958	0.899727	0.901475
1.3	0.903199	0.904902	0.906582	0.908241	0.909877	0.911492	0.913085	0.914657	0.916207	0.917736
1.4	0.919243	0.920730	0.922196	0.923641	0.925066	0.926471	0.927855	0.929219	0.930563	0.931888
1.5	0.933193	0.934478	0.935744	0.936992	0.938220	0.939429	0.940620	0.941792	0.942947	0.944083
1.6	0.945201	0.946301	0.947384	0.948449	0.949497	0.950529	0.951543	0.952540	0.953521	0.954486
1.7	0.955435	0.956367	0.957284	0.958185	0.959071	0.959941	0.960796	0.961636	0.962462	0.963273
1.8	0.964070	0.964852	0.965621	0.966375	0.967116	0.967843	0.968557	0.969258	0.969946	0.970621
1.9	0.971283	0.971933	0.972571	0.973197	0.973810	0.974412	0.975002	0.975581	0.976148	0.976705
2.0	0.977250	0.977784	0.978308	0.978822	0.979325	0.979818	0.980301	0.980774	0.981237	0.981691
2.1	0.982136	0.982571	0.982997	0.983414	0.983823	0.984222	0.984614	0.984997	0.985371	0.985738
2.2	0.986097	0.986447	0.986791	0.987126	0.987455	0.987776	0.988089	0.988396	0.988696	0.988989
2.3	0.989276	0.989556	0.989830	0.990097	0.990358	0.990613	0.990863	0.991106	0.991344	0.991576
2.4	0.991802	0.992024	0.992240	0.992451	0.992656	0.992857	0.993053	0.993244	0.993431	0.993613
2.5	0.993790	0.993963	0.994132	0.994297	0.994457	0.994614	0.994766	0.994915	0.995060	0.995201
2.6	0.995339	0.995473	0.995604	0.995731	0.995855	0.995975	0.996093	0.996207	0.996319	0.996427
2.7	0.996533	0.996636	0.996736	0.996833	0.996928	0.997020	0.997110	0.997197	0.997282	0.997365
2.8	0.997445	0.997523	0.997599	0.997673	0.997744	0.997814	0.997882	0.997948	0.998012	0.998074
2.9	0.998134	0.998193	0.998250	0.998305	0.998359	0.998411	0.998462	0.998511	0.998559	0.998605
3.0	0.998650	0.998694	0.998736	0.998777	0.998817	0.998856	0.998893	0.998930	0.998965	0.998999
3.1	0.999032	0.999065	0.999096	0.999126	0.999155	0.999184	0.999211	0.999238	0.999264	0.999289
3.2	0.999313	0.999336	0.999359	0.999381	0.999402	0.999423	0.999443	0.999462	0.999481	0.999499
3.3	0.999517	0.999533	0.999550	0.999566	0.999581	0.999596	0.999610	0.999624	0.999638	0.999650
3.4	0.999663	0.999675	0.999687	0.999698	0.999709	0.999720	0.999730	0.999740	0.999749	0.999758
3.5	0.999767	0.999776	0.999784	0.999792	0.999800	0.999807	0.999815	0.999821	0.999828	0.999835
3.6	0.999841	0.999847	0.999853	0.999858	0.999864	0.999869	0.999874	0.999879	0.999883	0.999888
3.7	0.999892	0.999896	0.999900	0.999904	0.999908	0.999912	0.999915	0.999918	0.999922	0.999925
3.8	0.999928	0.999931	0.999933	0.999936	0.999938	0.999941	0.999943	0.999946	0.999948	0.999950
3.9	0.999952	0.999954	0.999956	0.999958	0.999959	0.999961	0.999963	0.999964	0.999966	0.999967

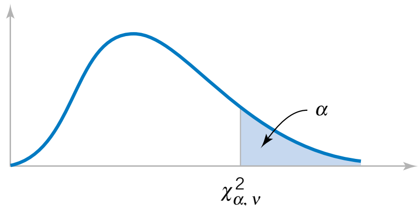


Table III Percentage Points $\chi^2_{\alpha,\nu}$ of the Chi-Squared Distribution

ν \ α	.995	.990	.975	.950	.900	.500	.100	.050	.025	.010	.005
1	.00+	.00+	.00+	.00+	.02	.45	2.71	3.84	5.02	6.63	7.88
2	.01	.02	.05	.10	.21	1.39	4.61	5.99	7.38	9.21	10.60
3	.07	.11	.22	.35	.58	2.37	6.25	7.81	9.35	11.34	12.84
4	.21	.30	.48	.71	1.06	3.36	7.78	9.49	11.14	13.28	14.86
5	.41	.55	.83	1.15	1.61	4.35	9.24	11.07	12.83	15.09	16.75
6	.68	.87	1.24	1.64	2.20	5.35	10.65	12.59	14.45	16.81	18.55
7	.99	1.24	1.69	2.17	2.83	6.35	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	7.34	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	8.34	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	9.34	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	10.34	17.28	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	6.30	11.34	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	12.34	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	13.34	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.27	7.26	8.55	14.34	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	15.34	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	16.34	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.87	17.34	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	18.34	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	19.34	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	20.34	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	21.34	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	22.34	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	23.34	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	24.34	34.28	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	25.34	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	26.34	36.74	40.11	43.19	46.96	49.65
28	12.46	13.57	15.31	16.93	18.94	27.34	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	28.34	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	29.34	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	39.34	51.81	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	37.69	49.33	63.17	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	46.46	59.33	74.40	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	55.33	69.33	85.53	90.53	95.02	100.42	104.22
80	51.17	53.54	57.15	60.39	64.28	79.33	96.58	101.88	106.63	112.33	116.32
90	59.20	61.75	65.65	69.13	73.29	89.33	107.57	113.14	118.14	124.12	128.30
100	67.33	70.06	74.22	77.93	82.36	99.33	118.50	124.34	129.56	135.81	140.17

ν = degrees of freedom.

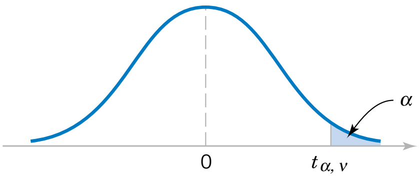


Table IV Percentage Points $t_{\alpha,\nu}$ of the t-Distribution

ν \ α	.40	.25	.10	.05	.025	.01	.005	.0025	.001	.0005
1	.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	.289	.816	1.886	2.920	4.303	6.965	9.925	14.089	23.326	31.598
3	.277	.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924
4	.271	.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	.267	.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	.265	.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	.263	.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	.262	.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	.261	.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	.260	.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	.260	.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	.259	.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	.259	.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	.258	.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	.258	.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	.258	.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	.257	.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	.257	.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	.257	.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	.257	.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	.257	.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	.256	.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	.256	.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	.256	.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	.256	.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	.256	.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	.256	.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	.256	.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	.256	.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	.256	.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	.255	.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
60	.254	.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
120	.254	.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	.253	.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

ν = degrees of freedom.

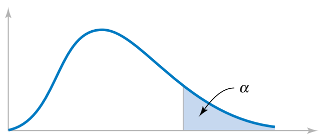


Table V Percentage Points f_{α,ν_1,ν_2} of the F-Distribution

$f_{0.25,\nu_1,\nu_2}$

Degrees of freedom for the numerator (ν_1); degrees of freedom for the denominator (ν_2)

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	5.83	7.50	8.20	8.58	8.82	8.98	9.10	9.19	9.26	9.32	9.41	9.49	9.58	9.63	9.67	9.71	9.76	9.80	9.85
2	2.57	3.00	3.15	3.23	3.28	3.31	3.34	3.35	3.37	3.38	3.39	3.41	3.43	3.43	3.44	3.45	3.46	3.47	3.48
3	2.02	2.28	2.36	2.39	2.41	2.42	2.43	2.44	2.44	2.44	2.45	2.46	2.46	2.46	2.47	2.47	2.47	2.47	2.47
4	1.81	2.00	2.05	2.06	2.07	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08
5	1.69	1.85	1.88	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.88	1.88	1.88	1.88	1.87	1.87	1.87
6	1.62	1.76	1.78	1.79	1.79	1.78	1.78	1.78	1.77	1.77	1.77	1.76	1.76	1.75	1.75	1.75	1.74	1.74	1.74
7	1.57	1.70	1.72	1.72	1.71	1.71	1.70	1.70	1.70	1.69	1.68	1.68	1.67	1.67	1.66	1.66	1.65	1.65	1.65
8	1.54	1.66	1.67	1.66	1.66	1.65	1.64	1.64	1.63	1.63	1.62	1.62	1.61	1.60	1.60	1.59	1.59	1.58	1.58
9	1.51	1.62	1.63	1.63	1.62	1.61	1.60	1.60	1.59	1.59	1.58	1.57	1.56	1.56	1.55	1.54	1.54	1.53	1.53
10	1.49	1.60	1.60	1.59	1.59	1.58	1.57	1.56	1.56	1.55	1.54	1.53	1.52	1.52	1.51	1.51	1.50	1.49	1.48
11	1.47	1.58	1.58	1.57	1.56	1.55	1.54	1.53	1.53	1.52	1.51	1.50	1.49	1.49	1.48	1.47	1.47	1.46	1.45
12	1.46	1.56	1.56	1.55	1.54	1.53	1.52	1.51	1.51	1.50	1.49	1.48	1.47	1.46	1.45	1.45	1.44	1.43	1.42
13	1.45	1.55	1.55	1.53	1.52	1.51	1.50	1.49	1.49	1.48	1.47	1.46	1.45	1.44	1.43	1.42	1.42	1.41	1.40
14	1.44	1.53	1.53	1.52	1.51	1.50	1.49	1.48	1.47	1.46	1.45	1.44	1.43	1.42	1.41	1.41	1.40	1.39	1.38
15	1.43	1.52	1.52	1.51	1.49	1.48	1.47	1.46	1.46	1.45	1.44	1.43	1.41	1.41	1.40	1.39	1.38	1.37	1.36
16	1.42	1.51	1.51	1.50	1.48	1.47	1.46	1.45	1.44	1.44	1.43	1.41	1.40	1.39	1.38	1.37	1.36	1.35	1.34
17	1.42	1.51	1.50	1.49	1.47	1.46	1.45	1.44	1.43	1.43	1.41	1.40	1.39	1.38	1.37	1.36	1.35	1.34	1.33
18	1.41	1.50	1.49	1.48	1.46	1.45	1.44	1.43	1.42	1.42	1.40	1.39	1.38	1.37	1.36	1.35	1.34	1.33	1.32
19	1.41	1.49	1.49	1.47	1.46	1.44	1.43	1.42	1.41	1.41	1.40	1.38	1.37	1.36	1.35	1.34	1.33	1.32	1.30
20	1.40	1.49	1.48	1.47	1.45	1.44	1.43	1.42	1.41	1.40	1.39	1.37	1.36	1.35	1.34	1.33	1.32	1.31	1.29
21	1.40	1.48	1.48	1.46	1.44	1.43	1.42	1.41	1.40	1.39	1.38	1.37	1.35	1.34	1.33	1.32	1.31	1.30	1.28
22	1.40	1.48	1.47	1.45	1.44	1.42	1.41	1.40	1.39	1.39	1.37	1.36	1.34	1.33	1.32	1.31	1.30	1.29	1.28
23	1.39	1.47	1.47	1.45	1.43	1.42	1.41	1.40	1.39	1.38	1.37	1.35	1.34	1.33	1.32	1.31	1.30	1.28	1.27
24	1.39	1.47	1.46	1.44	1.43	1.41	1.40	1.39	1.38	1.38	1.36	1.35	1.33	1.32	1.31	1.30	1.29	1.28	1.26
25	1.39	1.47	1.46	1.44	1.42	1.41	1.40	1.39	1.38	1.37	1.36	1.34	1.33	1.32	1.31	1.29	1.28	1.27	1.25
26	1.38	1.46	1.45	1.44	1.42	1.41	1.39	1.38	1.37	1.37	1.35	1.34	1.32	1.31	1.30	1.29	1.28	1.26	1.25
27	1.38	1.46	1.45	1.43	1.42	1.40	1.39	1.38	1.37	1.36	1.35	1.33	1.32	1.31	1.30	1.28	1.27	1.26	1.24
28	1.38	1.46	1.45	1.43	1.41	1.40	1.39	1.38	1.37	1.36	1.34	1.33	1.31	1.30	1.29	1.28	1.27	1.25	1.24
29	1.38	1.45	1.45	1.43	1.41	1.40	1.38	1.37	1.36	1.35	1.34	1.32	1.31	1.30	1.29	1.27	1.26	1.25	1.23
30	1.38	1.45	1.44	1.42	1.41	1.39	1.38	1.37	1.36	1.35	1.34	1.32	1.30	1.29	1.28	1.27	1.26	1.24	1.23
40	1.36	1.44	1.42	1.40	1.39	1.37	1.36	1.35	1.34	1.33	1.31	1.30	1.28	1.26	1.25	1.24	1.22	1.21	1.19
60	1.35	1.42	1.41	1.38	1.37	1.35	1.33	1.32	1.31	1.30	1.29	1.27	1.25	1.24	1.22	1.21	1.19	1.17	1.15
120	1.34	1.40	1.39	1.37	1.35	1.33	1.31	1.30	1.29	1.28	1.26	1.24	1.22	1.21	1.19	1.18	1.16	1.13	1.10
∞	1.32	1.39	1.37	1.35	1.33	1.31	1.29	1.28	1.27	1.25	1.24	1.22	1.19	1.18	1.16	1.14	1.12	1.08	1.00

Table V Percentage Points of the *F*-Distribution (*continued*)

$f_{0.10,v_1,v_2}$

v_2 \ v_1	Degrees of freedom for the numerator (v_1)																		
Degrees of freedom for the denominator (v_2)	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19	60.71	61.22	61.74	62.00	62.26	62.53	62.79	63.06	63.33
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.20	5.18	5.18	5.17	5.16	5.15	5.14	5.13
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79	3.78	3.76
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.12	3.10
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	2.47
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32	2.29
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	2.16
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.28	2.24	2.20	2.18	2.16	2.13	2.11	2.08	2.06
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.03	2.00	1.97
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.15	2.10	2.06	2.04	2.01	1.99	1.96	1.93	1.90
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.90	1.88	1.85
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	2.05	2.01	1.96	1.94	1.91	1.89	1.86	1.83	1.80
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.02	1.97	1.92	1.90	1.87	1.85	1.82	1.79	1.76
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	1.99	1.94	1.89	1.87	1.84	1.81	1.78	1.75	1.72
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.96	1.91	1.86	1.84	1.81	1.78	1.75	1.72	1.69
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.93	1.89	1.84	1.81	1.78	1.75	1.72	1.69	1.66
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96	1.91	1.86	1.81	1.79	1.76	1.73	1.70	1.67	1.63
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.64	1.61
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92	1.87	1.83	1.78	1.75	1.72	1.69	1.66	1.62	1.59
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.64	1.60	1.57
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89	1.84	1.80	1.74	1.72	1.69	1.66	1.62	1.59	1.55
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.61	1.57	1.53
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87	1.82	1.77	1.72	1.69	1.66	1.63	1.59	1.56	1.52
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.81	1.76	1.71	1.68	1.65	1.61	1.58	1.54	1.50
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85	1.80	1.75	1.70	1.67	1.64	1.60	1.57	1.53	1.49
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.79	1.74	1.69	1.66	1.63	1.59	1.56	1.52	1.48
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83	1.78	1.73	1.68	1.65	1.62	1.58	1.55	1.51	1.47
30	2.88	2.49	2.28	2.14	2.03	1.98	1.93	1.88	1.85	1.82	1.77	1.72	1.67	1.64	1.61	1.57	1.54	1.50	1.46
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.47	1.42	1.38
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.66	1.60	1.54	1.51	1.48	1.44	1.40	1.35	1.29
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.60	1.55	1.48	1.45	1.41	1.37	1.32	1.26	1.19
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.55	1.49	1.42	1.38	1.34	1.30	1.24	1.17	1.00

Table V Percentage Points of the F-Distribution (*continued*)

$f_{0.05,v_1,v_2}$

v_2 \ v_1	Degrees of freedom for the numerator (v_1)																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.55	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

Degrees of freedom for the denominator (v_2)

Table V Percentage Points of the *F*-Distribution (*continued*)

$f_{0.025,v_1,v_2}$

v_2 \ v_1	Degrees of freedom for the numerator (v_1)																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.7	963.3	968.6	976.7	984.9	993.1	997.2	1001	1006	1010	1014	1018
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.41	39.43	39.45	39.46	39.46	39.47	39.48	39.49	39.50
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.34	14.25	14.17	14.12	14.08	14.04	13.99	13.95	13.90
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.75	8.66	8.56	8.51	8.46	8.41	8.36	8.31	8.26
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52	6.43	6.33	6.28	6.23	6.18	6.12	6.07	6.02
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.37	5.27	5.17	5.12	5.07	5.01	4.96	4.90	4.85
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.67	4.57	4.47	4.42	4.36	4.31	4.25	4.20	4.14
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.20	4.10	4.00	3.95	3.89	3.84	3.78	3.73	3.67
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.87	3.77	3.67	3.61	3.56	3.51	3.45	3.39	3.33
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62	3.52	3.42	3.37	3.31	3.26	3.20	3.14	3.08
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.43	3.33	3.23	3.17	3.12	3.06	3.00	2.94	2.88
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28	3.18	3.07	3.02	2.96	2.91	2.85	2.79	2.72
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25	3.15	3.05	2.95	2.89	2.84	2.78	2.72	2.66	2.60
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15	3.05	2.95	2.84	2.79	2.73	2.67	2.61	2.55	2.49
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.96	2.86	2.76	2.70	2.64	2.59	2.52	2.46	2.40
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99	2.89	2.79	2.68	2.63	2.57	2.51	2.45	2.38	2.32
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	2.92	2.82	2.72	2.62	2.56	2.50	2.44	2.38	2.32	2.25
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87	2.77	2.67	2.56	2.50	2.44	2.38	2.32	2.26	2.19
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88	2.82	2.72	2.62	2.51	2.45	2.39	2.33	2.27	2.20	2.13
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68	2.57	2.46	2.41	2.35	2.29	2.22	2.16	2.09
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80	2.73	2.64	2.53	2.42	2.37	2.31	2.25	2.18	2.11	2.04
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76	2.70	2.60	2.50	2.39	2.33	2.27	2.21	2.14	2.08	2.00
23	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73	2.67	2.57	2.47	2.36	2.30	2.24	2.18	2.11	2.04	1.97
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64	2.54	2.44	2.33	2.27	2.21	2.15	2.08	2.01	1.94
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	2.61	2.51	2.41	2.30	2.24	2.18	2.12	2.05	1.98	1.91
26	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65	2.59	2.49	2.39	2.28	2.22	2.16	2.09	2.03	1.95	1.88
27	5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63	2.57	2.47	2.36	2.25	2.19	2.13	2.07	2.00	1.93	1.85
28	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61	2.55	2.45	2.34	2.23	2.17	2.11	2.05	1.98	1.91	1.83
29	5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.59	2.53	2.43	2.32	2.21	2.15	2.09	2.03	1.96	1.89	1.81
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.41	2.31	2.20	2.14	2.07	2.01	1.94	1.87	1.79
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.29	2.18	2.07	2.01	1.94	1.88	1.80	1.72	1.64
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.17	2.06	1.94	1.88	1.82	1.74	1.67	1.58	1.48
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	2.05	1.94	1.82	1.76	1.69	1.61	1.53	1.43	1.31
∞	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.94	1.83	1.71	1.64	1.57	1.48	1.39	1.27	1.00

Degrees of freedom for the denominator (v_2)

Table V Percentage Points of the *F*-Distribution (*continued*)

$f_{0.01,\nu_1,\nu_2}$

ν_2 \ ν_1	Degrees of freedom for the numerator (ν_1)																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
Degrees of freedom for the denominator (ν_2) 1	4052	4999.5	5403	5625	5764	5859	5928	5982	6022	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87	26.69	26.00	26.50	26.41	26.32	26.22	26.13
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.46
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	8.68	6.36	5.42	4.89	4.36	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.59
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00