TRƯỜNG CAO ĐẲNG CNTT TPHCM **BỘ MÔN TOÁN**

GVC ThS NGUYỄN THỊ MINH THƯ Chủ biên

ThS DƯƠNG THỊ XUÂN AN; ThS NGUYỄN THỊ THU THỦY

GIÁO TRÌNH

LÝ THUYẾT XÁC SUẤT VÀ THỐNG KÊ TOÁN

(LƯU HÀNH NỘI BỘ)

TP HỒ CHÍ MINH 2013

Hoan nghênh bạn đọc góp ý phê bình
Chân thành cảm ơn

LỜI NÓI ĐẦU

Nhằm đáp ứng nhu cầu học tập và giảng dạy môn Xác suất và thống kê toán, Bộ môn Toán Trường Cao Đẳng Công Nghệ Thông Tin TPHCM đã tổ chức biên soạn giáo trình "Lý thuyết xác suất và thống kê toán".

Giáo trình biên soạn trên cơ sở đề cương môn học theo tín chỉ đã được Hội Đồng Khoa học Trường Cao Đẳng Công Nghệ Thông Tin TPHCM phê duyệt.

Nội dung cuốn sách gồm 2 phần, phần 1: Lý thuyết xác suất, phần 2: Thống kê toán. Cuốn sách giải quyết các vấn đề trọng yếu của môn học, giúp sinh viên có nền tảng kiến thức để tiếp cận các môn học khác trong chương trình đào tạo hệ cao đẳng. Phần lý thuyết được trình bày logic, ngắn gọn, dễ hiểu, với nhiều ví dụ mẫu phù hợp với đối tượng là sinh viên hệ cao đẳng. Ngoài ra, sau mỗi chương đều có bài tập để sinh viên tự rèn luyện và nghiên cứu.

Đây là tài liệu được sử dụng chính thức trong trường, giúp sinh viên học tập và thi kết thúc học phần có hiệu quả tốt theo chương trình đào tạo tín chỉ. Trong quá trình giảng dạy, giáo trình sẽ được cập nhật, chỉnh lý để ngày càng hoàn thiện và đầy đủ hơn.

Do khả năng có hạn và cũng là lần đầu biên soạn theo hướng đào tạo tín chỉ nên giáo trình không tránh khỏi sai sót. Tập thể giáo viên bộ môn Toán rất mong nhận được các ý kiến góp ý, phê bình của bạn đọc trong và ngoài trường. Các ý kiến góp ý, phê bình của bạn đọc xin gửi về chủ biên: NGUYỄN THỊ MINH THƯ - Trưởng bộ môn TOÁN Trường Cao đẳng Công nghệ Thông tin TP HCM.
Địa chỉ: minhthu15916@gmail.com.

Xin chân thành cảm ơn.

BỘ MÔN TOÁN

MỤC LỤC

PHẦN I LÝ THUYẾT XÁC SUẤT

CHƯƠNG I
BIẾN CỐ NGẪU NHIÊN VÀ XÁC SUẤT

1.1. BỔ TÚC VỀ GIẢI TÍCH TỔ HỢP

I. Giai thừa

Kí hiệu n! là một tích của n số nguyên dương liên tiếp từ 1 đến n.

$$n! = 1.2.3...(n-1).n$$

Qui ước: 0! = 1

II. Qui tắc nhân và qui tắc cộng

1. Qui tắc nhân

Nếu một hiện tượng nào đó có thể chia làm k giai đoạn. Giai đoạn 1 xảy ra trong n_1 cách khác nhau và sau đó giai đoạn thứ 2 xảy ra trong n_2 cách khác nhau, tiếp theo giai đoạn thứ 3 xảy ra trong n_3 cách khác nhau... và tiếp theo giai đoạn thứ k lại xảy ra trong n_k cách khác nhau thì hiện tượng theo thứ tự nói trên đã xảy ra trong $(n_1.n_2.n_3...n_k)$ cách.

Ví dụ 1 Với các chữ số 1, 2, 3, 4, 5

a) Có thể lập ra bao nhiêu số gồm 3 chữ số?

b) Có bao nhiêu số gồm 3 chữ số khác nhau?

c) Có bao nhiêu số chẵn gồm 3 chữ số khác nhau?

d) Có bao nhiêu số gồm 5 chữ số viết không lặp lại? Trong tập này có bao nhiêu số chia hết cho 5?

BÀI GIẢI

a) Ta có thể chia thành 3 giai đoạn: giai đoạn 1 chọn 1 số trong 5 số đã cho để làm chữ số hàng đơn vị, nghĩa là có 5 cách chọn chữ số hàng đơn vị. Giai đoạn 2 chọn 1 trong 5 số đã cho để làm chữ số hàng chục, cũng có 5 cách chọn chữ số hàng chục. Giai đoạn 3 chọn 1 trong 5 số đã cho để làm chữ số hàng trăm, cũng có 5 cách chọn chữ số hàng trăm. Do đó từ 5 chữ số đã cho, ta có thể lập được 5.5.5 = 125 số gồm 3 chữ số.

b) Ta có thể chia thành 3 giai đoạn: giai đoạn 1 chọn 1 số trong 5 số đã cho để làm chữ số hàng đơn vị, nghĩa là có 5 cách chọn chữ số hàng đơn vị. Giai đoạn 2 chọn 1 trong 4 số đã cho còn lại để làm chữ số hàng chục, cũng có 4 cách chọn chữ số hàng chục khác chữ số hàng đơn vị. Giai đoạn 3 chọn 1 trong 3 số đã cho còn lại để làm chữ số hàng trăm, cũng có 3 cách chọn chữ số hàng trăm khác chữ số hàng đơn vị và chữ số hàng chục. Do đó từ 5 chữ số đã cho, ta có thể lập được 5.4.3= 60 số gồm 3 chữ số.

c) Số chẵn là số có chữ số ở hàng đơn vị là số chẵn.Trong 5 chữ số đã cho có 2 chữ số chẵn là số 2 và số 4. Do đó có 2 cách chọn chữ số chẵn cho hàng đơn vị. Có 4 cách chọn chữ số hàng chục khác với chữ số hàng đơn vị. Có 3 cách chọn chữ số hàng trăm khác với chữ số hàng đơn vị và hàng chục. Vậy trong tập hợp các số gồm 5 chữ số đã cho có 2.4.3 = 24 số chẵn gồm 3 chữ số khác nhau.

d) Có 5 cách chọn chữ số hàng đơn vị.Có 4 cách chọn chữ số hàng chục khác với chữ số hàng đơn vị.Có 3 cách chọn chữ số hàng trăm khác với các chữ số ở hai hàng kia. Có 2 cách chọn chữ số hàng ngàn khác với các chữ số đã chọn trước. Cuối cùng chỉ có 1 cách chọn chữ số hàng chục ngàn khác với 4 chữ số kia.

Vậy có 1.2.3.4.5 = 120 = 5! số gồm 5 chữ số khác nhau.

Một số chia hết cho 5 khi chữ số hàng đơn vị là chữ số 0 hoặc chữ số 5. Vậy chỉ có thể chọn trong bài toán này là chữ số 5 làm chữ số đứng ở hàng đơn vị mà thôi. Lí luận như trên ta có 4! = 24 cách chọn 4 chữ số khác nhau cho 4 vị trí còn lại.

Vậy trong tập hợp các số gồm 5 chữ số khác nhau được viết từ 5 chữ số đã cho có 1.4! = 24 số chia hết cho 5.

2. Qui tắc cộng

Nếu một hiện tượng nào đó có thể chia làm k trường hợp (sao cho 2 trường hợp bất kỳ không có cách chung): trường hợp 1 xảy ra trong n_1 cách khác nhau, trường hợp 2 xảy ra trong n_2 cách khác nhau, trường hợp thứ 3 xảy ra trong n_3 cách khác nhau..., trường hợp thứ k lại xảy ra trong n_k cách khác nhau thì hiện tượng nói trên đã xảy ra trong $(n_1 + n_2 + n_3 + \ldots + n_k)$ cách.

Ví dụ 2 Có 3 lớp sinh viên: ngân hàng 1 có 20 sinh viên nam và 30 sinh viên nữ, ngân hàng 2 có 25 sinh viên nam và 31 sinh viên nữ, ngân hàng 3 có 19 sinh viên nam và 35 sinh viên nữ. Tổng số cách chọn một sinh viên nữ của 3 lớp này là: 30+31+35=96.

III. Hoán vị

Người ta gọi hoán vị n phần tử không lặp lại là số cách sắp xếp n phần tử khác nhau vào n vị trí đã cho. Kí hiệu: $P_n = n!$

Ví dụ 3 Hỏi có bao nhiêu cách sắp xếp chỗ ngồi cho 3 người vào một cái bàn dài có 3 chỗ ngồi?

BÀI GIẢI Có 3! = 6 cách sắp xếp chỗ ngồi

Ví dụ 4 Một hội nghị bàn tròn có phái đoàn của các nước: 3 người Việt Nam, 5 người Mỹ, 2 người Nhật, 3 người Singapore và 4 người Hongkong. Hỏi có bao nhiêu cách sắp xếp chỗ ngồi cho mọi thành viên sao cho người có cùng quốc tịch thì ngồi cạnh nhau.

BÀI GIẢI

Có thể mời phái đoàn của một nước nào đó ngồi vào chỗ trước và sắp xếp 4 phái đoàn còn lại. Do đó có 4! = 24 cách sắp xếp các phái đoàn ngồi theo quốc gia của mình, trong đó có:

3! = 6 cách sắp xếp cho 3 người Việt Nam.

5! = 120 cách sắp xếp cho 5 người Mỹ.

2! = 2 cách sắp xếp cho 2 người Nhật.

3! = 6 cách sắp xếp cho 3 người Singapore.

4! = 24 cách sắp xếp cho 4 người Hongkong.

Vậy có tất cả là: 4!3!5!2!3!4! = 4976640 cách.

IV. Chỉnh hợp

Chỉnh hợp chập k của n phần tử ($k \le n$) là một nhóm có thứ tự gồm k phần tử khác nhau chọn từ n phần tử đã cho.

Số chỉnh hợp chập k của n phần tử kí hiệu là

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

Ví dụ 5 Một lớp học có 50 người. Chọn Ban Cán Sự lớp gồm 1 lớp trưởng, 1 lớp phó, 1 ủy viên. Hỏi có bao nhiêu cách chọn?

BÀI GIẢI

Số cách chọn ban cán sự lớp bằng số cách chọn có thứ tự 3 người từ 50 người là

$$A_{50}^3 = \frac{50!}{(50-3)!} = 117600$$

V. Chỉnh hợp lặp

Chỉnh hợp lặp chập k của n phần tử là một nhóm có thứ tự gồm k phần tử chọn từ n phần tử đã cho. Trong đó, mỗi phần tử có thể có mặt 1, 2,..., k lần trong nhóm đó.

Số chỉnh hợp lặp chập k của n phần tử là $B_n^k = n^k$.

Ví dụ 6 Xếp ngẫu nhiên 10 người lên 8 toa tàu một cách tùy ý.

Hỏi có bao nhiêu cách?

BÀI GIẢI

Xếp ngẫu nhiên 10 người lên 8 toa tàu một cách tùy ý ta có thể chia thành 10 giai đoạn (mỗi giai đoạn xếp 1 người). Mỗi giai đoạn đều có 8 cách. Vậy tổng số cách là $B_8^{10} = 8^{10}$.

VI. Tổ hợp

Tổ hợp chập k của n phần tử ($k \le n$) là một nhóm không phân biệt thứ tự, gồm k phần tử khác nhau chọn từ n phần tử đã cho.

Số tổ hợp chập k của n phần tử là

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

Chú ý: $C_n^0 = C_n^n = 1;\ C_n^1 = n$

Phân biệt Chỉnh hợp khác tổ hợp ở

- Hai chỉnh hợp khác nhau :

+ Hoặc có ít nhất một phần tử khác nhau.

+ Hoặc chỉ khác nhau về thứ tự sắp xếp của các phần tử.

- Hai tổ hợp chỉ khác nhau khi có ít nhất một phần tử khác nhau.

Ví dụ 7 Một hộp đựng 7 quả cầu trắng và 3 quả cầu đỏ. Ta lấy ngẫu nhiên ra 4 quả cầu:

a) Hỏi có bao nhiêu cách ?

b) Trong đó có bao nhiêu cách lấy được 2 quả cầu đỏ ?

c) Có bao nhiêu cách lấy nhiều nhất 2 quả cầu màu đỏ ?

d) Ít nhất là 2 quả cầu màu đỏ ?

e) Ít nhất là 1 quả cầu màu đỏ ?

BÀI GIẢI

a) Có 10 quả cầu, lấy ra 4 quả thì có $C_{10}^4 = 210$ cách.

b) Có 3 quả cầu đỏ, lấy ra 2 quả thì có C_3^2 cách.

Có 7 quả cầu trắng, lấy ra 2 quả thì có C_7^2 cách.

Suy ra có $C_3^2.C_7^2 = 3.21 = 63$ cách.

c) Có thể chọn: (2 đỏ + 2 trắng), (1 đỏ + 3 trắng), (4 trắng).

Do đó có: $C_3^2C_7^2 + C_3^1C_7^3 + C_7^4 = 63 + 105 + 35 = 203$ cách

d) Có thể chọn: (2 đỏ + 2 trắng), (3 đỏ + 1 trắng).

Do đó có: $C_3^2C_7^2 + C_3^3C_7^1 = 63 + 7 = 70$ cách.

e) Có thể chọn: (1 đỏ + 3 trắng), (2 đỏ + 2 trắng), (3 đỏ + 1 trắng).

Do đó có: $C_3^1C_7^3 + C_3^2C_7^2 + C_3^3C_7^1 = 105 + 63 + 7 = 175$ cách.

Cách khác: - Không có quả cầu màu đỏ, có: $C_3^0C_7^4 = 35$ cách.

- Lấy 4 quả cầu một cách tùy ý, có: $C_{10}^4 = 210$ cách.

Lấy được ít nhất là 1 quả cầu màu đỏ, có:

C_{10}^{4} - $C_3^0 C_7^4$ =210- 35 =175 cách.

VII. Nhị thức Newton

$$(a+b)^n = \sum_{k=0}^{n} C_n^k a^{n-k} b^k$$

Đặc biệt

Khi $n = 2$ ta có

$\left(a+b\right)^2 = C_2^0 a^2 + C_2^1 ab + C_2^2 b^2 = a^2 + 2ab + b^2$.

Khi $n = 3$ ta có

$\left(a+b\right)^3 = C_3^0 a^3 + C_3^1 a^2 b + C_3^2 ab^2 + C_3^3 b^3 = a^3 + 3a^2 b + 3ab^2 + b^3$

Tổng quát khi tính hệ số C_n^k trong khai triển nhị thức Newton người ta thường dùng tam giác Pascal.

Trong khi áp dụng giải tích tổ hợp vào lý thuyết xác suất và thống kê toán thông thường ta gọi tập W là tập hợp chính mà từ đó ta rút ra một số phần tử nào đó, chẳng hạn k các phần tử. Tập hợp lập nên bởi các phần tử được lấy ra gọi là mẫu, số phần tử của mẫu được gọi là cỡ của mẫu.

Thông thường ta hay xét hai cách lấy mẫu: lấy mẫu có hoàn lại và lấy mẫu không hoàn lại.

a) Lấy mẫu có hoàn lại

Trong cách lấy mẫu này sau khi đã chọn một phần tử ở tập chính ra, ta lại trả phần tử đó về tập chính trước khi chọn tiếp phần tử khác. Như vậy, số mẫu có cỡ k từ tập hợp có n phần tử có thể có là n^k.

b) Lấy mẫu không hoàn lại

Trong cách lấy mẫu này, khi đã chọn một phần tử nào đó ta bỏ phần tử đó khỏi tập hợp chính, sau đó mới lấy tiếp phần tử khác. Như vậy trong mẫu, mỗi phần tử chỉ có thể gặp không quá một lần và nếu k là cỡ mẫu thì k ≤ n. Số cỡ mẫu k từ tập chính gồm n phần tử bằng số chỉnh hợp chập k của tập hợp gồm n phần tử.

1.2 CÁC KHÁI NIỆM VỀ XÁC SUẤT

I. Đối tượng nghiên cứu của lý thuyết xác suất – thống kê

Trong tự nhiên có 2 loại hiện tượng:

- Hiện tượng tất nhiên: có thể dự đoán được kết quả của nó.
- Hiện tượng ngẫu nhiên: không thể dự đoán được kết quả.

Đối tượng nghiên cứu của lý thuyết xác suất – thống kê là các hiện tượng ngẫu nhiên. Chẳng hạn, khi ta tung một đồng xu có 2 mặt sấp – ngửa vài lần thì không thể biết mặt nào sẽ xuất hiện. Nhưng khi số lần tung khá lớn thì số lần xuất hiện mặt sấp xấp xỉ số lần xuất hiện mặt ngửa.

Mục đích: tìm ra các qui luật của các hiện tượng ngẫu nhiên.
Qui luật của hiện tượng ngẫu nhiên chỉ biểu hiện ra ngoài khi nó được lặp lại nhiều lần.
Lý thuyết xác suất: Tìm ra mô hình xác suất của các hiện tượng ngẫu nhiên.
Lý thuyết thống kê: Dựa vào dữ liệu thống kê (lấy từ thực tế) để chính xác hóa mô hình xác suất, đưa ra các quyết định hoặc dự báo.

II. Sự kiện (biến cố)

1. Phép thử

Định nghĩa xác suất được xây dựng trên cơ sở khái niệm phép thử. Đó là việc quan sát hoặc làm 1 thí nghiệm để ta nghiên cứu 1 đối tượng hay 1 hiện tượng ngẫu nhiên nào đó. Các phép thử thường do một nhóm điều kiện xác định. Khi các điều kiện này được thỏa mãn, ta gọi là đã thực hiện một phép thử. Kết quả của phép thử có thể được đặc trưng theo chất lượng hoặc đặc trưng theo số lượng. Kết quả chất lượng của phép thử được gọi là một sự kiện hoặc một biến cố. Kết quả số lượng của phép thử được gọi là đại lượng ngẫu nhiên đến chương II ta sẽ xét.

Ví dụ:

-Phép thử là tung 1 con xúc xắc cân đối và đồng chất xem mặt có mấy chấm xuất hiện.

-Phép thử là kiểm tra chất lượng của một lô hàng.

-Phép thử là nghiên cứu tác dụng phụ của một loại thuốc kháng sinh đối với trẻ em.

-Phép thử là bắn một viên đạn vào một cái bia xem viên đạn trúng bia ở vòng có điểm là bao nhiêu?

2. Phân loại biến cố

Ta thường gặp 3 loại biến cố

a. Biến cố chắc chắn là biến cố nhất định xảy ra sau khi thực hiện phép thử. Kí hiệu: Ω

b. Biến cố không thể có là biến cố nhất định không xảy ra sau khi thực hiện phép thử. Kí hiệu: $\varnothing$

c. Biến cố ngẫu nhiên là biến cố sau khi thực hiện phép thử nó có thể xảy ra mà cũng có thể không xảy ra.

Kí hiệu: A, B, C, A_1, A_2, ..., A_n,…

Ví dụ: Phép thử: thả hòn bi từ độ cao 1m

Biến cố: "hòn bi rơi xuống", đây là biến cố chắc chắn.

Ví dụ: Phép thử: sinh viên thi môn XSTK

Biến cố: "Sinh viên thi đạt", "Sinh viên thi không đạt", đây là các biến cố ngẫu nhiên.

Ví dụ: Phép thử là tung 1 con xúc xắc cân đối và đồng chất.

A là biến cố ra mặt chẵn có {2, 4, 6}

B là biến cố ra mặt lẻ có {1, 3, 5}

A_j là biến cố ra mặt có j chấm j=1,2,3,4,5,6

Biến cố ra mặt có số chấm lớn hơn 6 là $\varnothing$.

III. Mối quan hệ giữa các biến cố

1. Định nghĩa 1

Biến cố A và B được gọi là hai biến cố tương đương nếu A xảy ra thì B cũng xảy ra và ngược lại. Ký hiệu A = B

2. ***Định nghĩa 2***

Biến cố A được gọi là thuận lợi cho biến cố B nếu biến cố A xảy ra thì biến cố B cũng phải xảy ra. Kí hiệu: $A \subset B$

3. Định nghĩa 3

Biến cố C được gọi là tổng của 2 biến cố A và B.Biến cố C xảy ra khi và chỉ khi có ít nhất một trong hai biến cố A hoặc B xảy ra.

Ký hiệu là C = A + B

Ví dụ 1 Một hộp có 10 sản phẩm trong đó có 4 sản phẩm hỏng. Lấy ngẫu nhiên 3 sản phẩm

Gọi A_1 là biến cố 3 sản phẩm lấy ra có đúng 1 sản phẩm hỏng.

A_2 là biến cố 3 sản phẩm lấy ra có đúng 2 sản phẩm hỏng.

A là biến cố có 1 hoặc 2 sản phẩm hỏng thì $A = A_1 + A_2$

4. Định nghĩa 4

Biến cố A gọi là tổng của n biến cố: $A_1, A_2, \dots, A_n$. Biến cố A xảy ra khi và chỉ khi có ít nhất một trong n biến cố đó xảy ra.

Ký hiệu là $A = A_1 + A_2 + \dots + A_n$

5. ***Định nghĩa 5***

Hiệu của 2 biến cố A và B là một biến cố, xảy ra khi và chỉ khi biến cố A xảy ra nhưng biến cố B không xảy ra.Kí hiệu A\B

6. Định nghĩa 6

Biến cố C được gọi là tích của hai biến cố A và B. Biến cố C xảy ra khi và chỉ khi cả A và B đồng thời xảy ra. Ký hiệu C = A.B

Ví dụ 2 A là biến cố bạn Hà thi đậu môn Toán, B là biến cố bạn Hà thi đậu môn Anh văn thì A+B là biến cố bạn Hà thi đậu ít nhất 1 môn Toán hoặc Anh văn; A.B là biến cố bạn Hà thi đậu 2 môn Toán và Anh văn.

7. Định nghĩa 7

Biến cố A được gọi là tích của n biến cố: $A_1, A_2, \dots, A_n$ nếu A xảy ra khi và chỉ khi cả n biến cố ấy đồng thời xảy ra.

Ký hiệu $A = A_1A_2 \dots A_n$.

8. Định nghĩa 8

Hai biến cố A và B được gọi là xung khắc nhau nếu chúng không thể đồng thời xảy ra trong một phép thử. Nghĩa là $AB = \varnothing$

Ví dụ 3 Ở ví dụ 1 ta thấy ngay A_1 và A_2 là xung khắc vì đã có đúng 1 sản phẩm hỏng thì không thể có 2 sản phẩm hỏng.

9. Định nghĩa 9

Nhóm n biến cố $A_1, A_2, \dots, A_n$ được gọi là xung khắc từng đôi nếu bất kỳ hai trong n biến cố này xung khắc với nhau.

Nghĩa là $A_iA_j = \varnothing$ với $\forall i \neq j$

10. Định nghĩa 10

Các biến cố $A_1, A_2, \dots, A_n$ được gọi là nhóm biến cố đầy đủ và xung khắc nếu chúng xung khắc từng đôi và tổng của chúng là biến cố chắn chắn.

Nghĩa là $A_iA_j = \varnothing$ với $\forall i \neq j$ và $A_1 + A_2 + ... + A_n = \Omega$

Ví dụ 4 Có 2 hộp thuốc

Gọi A_1 là biến cố lấy hộp 1, A_2 là biến cố lấy hộp 2.

Các biến cố $\{A_1, A_2\}$ là nhóm biến cố đầy đủ và xung khắc từng đôi.

11. Định nghĩa 11

Biến cố A và $\overline{A}$ gọi là hai biến cố đối lập nhau nếu chúng tạo nên một nhóm biến cố đầy đủ và xung khắc: $A+\overline{A} = \Omega$; $A\overline{A} = \varnothing$

Như vậy $\overline{A}$ gọi là biến cố đối lập của biến cố A, nếu nó xảy ra khi và chỉ khi biến cố A không xảy ra.

Ví dụ 5 Một hộp có 10 sản phẩm trong đó có 4 sản phẩm hỏng. Lấy ngẫu nhiên 3 sản phẩm

Gọi A là biến cố trong 3 sản phẩm lấy ra có ít nhất 1 sản phẩm hỏng. Ta có $\overline{A}$ là biến cố trong 3 sản phẩm lấy ra không có sản phẩm nào hỏng.

12. Định nghĩa 12

Hai biến cố A và B được gọi là độc lập với nhau nếu việc xảy ra hay không xảy ra của biến cố này không làm thay đổi sự xuất hiện của biến cố kia và ngược lại.

Chú ý: Nếu A và B độc lập thì A và $\overline{B}$; $\overline{A}$ và B ; $\overline{A}$ và $\overline{B}$ cũng độc lập với nhau và $\overline{A+B} = \overline{A}.\overline{B}$; $\overline{A.B} = \overline{A} + \overline{B}$

Ví dụ 6 Hộp thứ nhất đựng 5 lọ thuốc tốt và 3 lọ kém phẩm chất. Hộp thứ hai có 3 lọ thuốc tốt và 2 lọ kém phẩm chất. Lấy ngẫu nhiên từ mỗi hộp ra một lọ.

Gọi A là biến cố lấy được 2 lọ thuốc tốt.

A_1 là biến cố lấy được 1 lọ thuốc tốt ở hộp 1.

A_2 là biến cố lấy được 1 lọ thuốc tốt ở hộp 2.

Ta có A_1 , A_2 là 2 biến cố độc lập và $A = A_1A_2$

13. Định nghĩa 13

Các biến cố $A_1, A_2,...,A_n$ được gọi là độc lập từng đôi nếu mỗi cặp hai biến cố bất kỳ trong n biến cố ấy độc lập với nhau.

14. Định nghĩa 14

Các biến cố $A_1, A_2, \ldots, A_n$ được gọi là độc lập toàn phần nếu mỗi biến cố độc lập với tích của một tổ hợp bất kỳ trong các biến cố còn lại.

Chú ý: Các biến cố độc lập từng đôi thì chưa chắc độc lập toàn phần. Điều kiện độc lập toàn phần mạnh hơn độc lập từng đôi.

15. Định nghĩa 15

Nhóm biến cố đồng khả năng là nhóm biến cố có khả năng xuất hiện như nhau

16. Định nghĩa 16 Không gian các sự kiện sơ cấp

Tập hợp các sự kiện $A_1, A_2, \ldots, A_n$ được gọi là không gian các sự kiện sơ cấp (hay không gian mẫu *)* nếu chúng là một hệ đầy đủ không thể tách nhỏ hơn.Ký hiệu là S.

Ví dụ 7 Phép thử là tung 1 con xúc xắc

A_j là biến cố ra mặt có j chấm j=1,2,3,4,5,6.

Không gian mẫu của phép thử này là: $S=\{ A_1, A_2, A_3, A_4, A_5, A_6 \}$

Ví dụ 8 Cho ba biến cố A, B, C. Viết biểu thức chỉ biến cố:

a) Cả ba biến cố cùng xảy ra.

b) Không có biến cố nào trong các biến cố đó xảy ra.

c) A và B xảy ra, nhưng C không xảy ra.

d) Có ít nhất một trong các biến cố A, B, C xảy ra.

e) Chỉ có A xảy ra.

f) Có một và chỉ một trong các biến cố đó xảy ra.

g) Chỉ có hai trong các biến cố đó xảy ra.

h) Có ít nhất hai biến cố cùng xảy ra.

i) Có không quá 2 biến cố trong các biến cố đó xảy ra.

BÀI GIẢI

a) ABC ; b) $\overline{A}.\overline{B}.\overline{C}$; c) $AB\overline{C}$; d) A + B +C; e) $A\overline{B}.\overline{C}$

f) $A.\overline{B}.\overline{C} + \overline{A}.B.\overline{C} + \overline{A}.\overline{B}.C$

g) $AB\overline{C} + A\overline{B}C + \overline{A}BC$

h) $AB\overline{C} + A\overline{B}C + \overline{A}BC + ABC$

i) $\overline{A}.\overline{B}.\overline{C} + A.\overline{B}.\overline{C} + \overline{A}.B.\overline{C} + \overline{A}.\overline{B}.C + A.B.\overline{C} + A.\overline{B}.C + \overline{A}.B.C$

Ví dụ 9 Một dụng cụ điện tử gồm có 3 bóng đèn loại 1 và 4 bóng đèn loại 2. Dụng cụ tiếp tục làm việc được nếu có ít nhất một bóng đèn loại 1 tốt và không ít hơn ba bóng đèn loại 2 tốt.

Hãy viết biểu thức chỉ biến cố dụng cụ tiếp tục làm việc.

BÀI GIẢI

Gọi A_k (k = 1, 2, 3) là biến cố chỉ bóng đèn loại 1 thứ k tốt.

B_j (j = 1, 2, 3,4) là bóng đèn loại 2 thứ j tốt.

C là biến cố chỉ dụng cụ tiếp tục làm việc được:

$$C = (A_1 + A_2 + A_3)\,[\,\overline{B}_1B_2B_3B_4 + B_1\overline{B}_2B_3B_4 + B_1B_2\overline{B}_3B_4 + B_1B_2B_3\overline{B}_4 + B_1B_2B_3B_4]$$

Ví dụ 10 Chọn ngẫu nhiên không hoàn lại từ kiện thứ nhất ra 4 sản phẩm và chọn ngẫu nhiên từ kiện thứ hai ra 5 sản phẩm để kiểm tra. Gọi C_i (i = 0, 1, 2, 3, 4) là biến cố có i sản phẩm đạt tiêu chuẩn trong 4 sản phẩm chọn ra từ kiện thứ nhất.
D_j (j = 0, 1, 2, 3, 4, 5) là biến cố có j sản phẩm đạt tiêu chuẩn trong 5 sản phẩm chọn ra từ kiện thứ hai.
Hãy viết các biến cố sau theo C_i và D_j

a) Có 5 sản phẩm đạt tiêu chuẩn trong 9 sản phẩm lấy ra từ hai kiện.

b) Có ít nhất 7 sản phẩm đạt tiêu chuẩn trong 9 sản phẩm lấy ra từ 2 kiện.

BÀI GIẢI

a) Gọi A là biến cố có 5 sản phẩm đạt tiêu chuẩn trong 9 sản phẩm lấy ra từ hai kiện.

$$A = C_1D_4 + C_2D_3 + C_3D_2 + C_4D_1 + C_0D_5$$

b) Gọi B là biến cố có ít nhất 7 sản phẩm đạt tiêu chuẩn trong 9 sản phẩm lấy ra từ 2 kiện.

$$B = C_2D_5 + C_3D_4 + C_4D_3 + C_3D_5 + C_4D_4 + C_4D_5$$

1.3 CÁC ĐỊNH NGHĨA VỀ XÁC SUẤT

Khi quan sát các biến cố ngẫu nhiên, ta thấy một số biến cố thường hay xảy ra, một số biến cố khác thường ít xảy ra. Từ đó người ta muốn đo lường khả năng xuất hiện của một biến cố.

I. Định nghĩa xác suất theo cách cổ điển

Xét một phép thử, giả sử không gian mẫu S có hữu hạn biến cố sơ cấp và các biến cố đồng khả năng

Xác suất của biến cố A chính là số đo khả năng xảy ra của biến cố A. Xác suất của biến cố A là một số, ký hiệu và định nghĩa là

$$P(A) = \frac{m}{n}$$

Trong đó : m là số trường hợp thuận lợi cho A.

n là số trường hợp đồng khả năng có thể xảy ra trong phép thử.

Xác suất phải thỏa các tiên đề sau

1) $P(A) \geq 0$
2) $P(\Omega) = 1$
3) $P(\varnothing) = 0$
4) $0 \leq P(A) \leq 1$

Ví dụ 1. Một lô hàng gồm 10 sản phẩm, trong đó có 3 sản phẩm kém chất lượng. Lấy ngẫu nhiên 1 sản phẩm từ lô hàng này. Tìm xác suất để sản phẩm lấy ra là sản phẩm tốt?

BÀI GIẢI

Gọi *A* là biến cố sản phẩm lấy ra là sản phẩm tốt

Có 10 cách chọn 1 sản phẩm từ 10 sản phẩm nên số trường hợp đồng khả năng là n=10.

Có 7 cách chọn 1 sản phẩm tốt từ 7 sản phẩm tốt nên số trường hợp thuận lợi cho *A* là m=7.

Vậy $P(A) = \frac{7}{10} = 0,7$.

Ví dụ 2 Xếp ngẫu nhiên 8 người lên 10 toa tàu. Tìm xác suất để

a) 8 người lên cùng một toa số 1?

b) 8 người lên cùng một toa?

c) 8 người lên 8 toa đầu?

d) 8 người lên 8 toa khác nhau?

BÀI GIẢI

Xếp ngẫu nhiên 8 người lên 10 toa tàu ta có thể chia thành 8 giai đoạn (mỗi giai đoạn xếp 1 người). Mỗi giai đoạn đều có 10 cách.

Vậy số trường hợp đồng khả năng có thể xảy ra là $n = B_{10}^{8} = 10^{8}$.

a) 8 người lên cùng toa số 1.

Đặt A là biến cố 8 người lên cùng toa số 1 thì chỉ có 1 trường hợp thuận lợi cho A.

Vậy m = 1. Do đó $P(A) = \frac{m}{n} = \frac{1}{10^8}$.

b) 8 người lên cùng 1 toa: Đặt B là biến cố 8 người lên cùng 1 toa.

Có 10 toa nên có 10 trường hợp thuận lợi cho B. Vậy m = 10.

Do đó $P(B) = \frac{m}{n} = \frac{10}{10^8} = \frac{1}{10^7}$

c) 8 người lên 8 toa đầu.

Đặt C là biến cố 8 người lên 8 toa đầu, xếp 8 người lên 8 toa là 1 hoán vị của 8 phần tử. Vậy số trường hợp thuận lợi là: m = 8!.

Do đó $P(C) = \frac{m}{n} = \frac{8!}{10^8} = 0,0004032$

d) 8 người lên 8 toa khác nhau.

Đặt D là biến cố 8 người lên 8 toa khác nhau. Ở đây ta lấy 8 toa từ 10 toa (có xếp thứ tự) nên số trường hợp thuận lợi cho D là $m = A_{10}^{8}$.

Do đó: $P(D) = \frac{m}{n} = \frac{A_{10}^{8}}{10^8} = 0,018144$

Ví dụ 3 Một lớp có 30 học sinh. Trong kỳ thi môn toán có 6 học sinh đạt điểm giỏi, 10 học sinh đạt điểm khá, 9 học sinh đạt điểm trung bình và 5 học sinh không đạt yêu cầu. Gọi tên ngẫu nhiên 3 học sinh của lớp.Tìm các xác suất sau

a) Gọi được 3 học sinh đều không đạt yêu cầu?

b) Gọi được 3 học sinh đạt điểm khá?

c) Gọi được 2 học sinh đạt điểm trung bình và 1 học sinh đạt điểm giỏi?

BÀI GIẢI

Gọi 3 học sinh từ 30 học sinh nên số trường hợp đồng khả năng là n = C_{30}^3.

a) Gọi A là biến cố gọi được 3 học sinh điểm không đạt yêu cầu. Ở đây ta lấy 3 học sinh từ 5 học sinh có điểm không đạt yêu cầu. Vậy số trường hợp thuận lợi cho biến cố A là m = C_5^3.

Do đó $P(A)=\dfrac{C_5^3}{C_{30}^3}=\dfrac{1}{406}$

b) Gọi B là biến cố gọi 3 học sinh đạt điểm khá.

Ở đây ta lấy 3 học sinh từ 10 học sinh đạt điểm khá. Vậy m = C_{10}^3

Do đó $P(B)=\dfrac{C_{10}^3}{C_{30}^3}=\dfrac{6}{203}$

c) Gọi C là biến cố gọi được 2 học sinh đạt điểm trung bình và 1 học sinh đạt điểm giỏi.

Ở đây ta lấy 2 học sinh từ 9 học sinh đạt điểm trung bình và lấy 1 học sinh từ 6 học sinh đạt điểm giỏi. Vậy m = $C_9^2.C_6^1$.

Do đó $P(C)=\dfrac{C_9^2.C_6^1}{C_{30}^3}=\dfrac{54}{1015}$

Hạn chế của định nghĩa xác suất theo cổ điển

Do đòi hỏi phải có hữu hạn các biến cố sơ cấp và tính đồng khả năng của chúng mà trong thực tế lại có nhiều phép thử không có tính chất đó. Để khắc phục những hạn chế trên người ta đưa ra một số định nghĩa xác suất khác như sau.

II. Định nghĩa xác suất theo thống kê

1. Định nghĩa tần suất

Nếu lặp lại n lần một phép thử trong đó có k lần xuất hiện biến cố A thì tần suất của A trong dãy n phép thử được kí hiệu và định nghĩa: $f_n(A) = \frac{k}{n}$.

2. Định nghĩa xác suất theo thống kê

Xác suất của biến cố A là $P(A) = \lim_{n \to +\infty} f_n(A)$.

Trong thực hành với n đủ lớn ta lấy $P(A) \approx f_n(A)$.

III. Định nghĩa xác suất theo hình học

Cho phép thử có không gian Ω là vô hạn và đồng khả năng. Khi đó nếu 2 biến cố A và Ω có thể biểu diễn bằng các miền hình học thì $P(A) = \frac{mesA}{mes\Omega}$

với mes A : độ lớn của hình biểu diễn cho biến cố A

mes Ω : độ lớn của hình biểu diễn cho biến cố Ω

Ví dụ 4 Tung 1 hạt cát vào 1 hình vuông cạnh a. Tính xác suất để hạt cát nằm trong hình tròn nội tiếp hình vuông

BÀI GIẢI

Gọi A là biến cố hạt cát nằm trong hình tròn nội tiếp hình vuông

Ta có: Ω là hình vuông; A là hình tròn nội tiếp

$$\Rightarrow \quad P(A) = \frac{mesA}{mes\Omega} = \frac{\pi\frac{a^2}{4}}{a^2} = \frac{\pi}{4}$$

Ví dụ 5 Hai người hẹn gặp nhau trong khoảng thời gian [0 , T] với giao hẹn nếu người đến trước đợi một khoảng thời gian t mà không gặp thì đi về. Tính xác suất để 2 người đó không gặp nhau.

BÀI GIẢI

Phép thử: thời điểm đến của 2 người đó

Gọi x: thời điểm đến của người thứ nhất;

y: thời điểm đến của người thứ hai

A là biến cố 2 người đó không gặp nhau.

Ta có: $\Omega = \{(x,y) \,/\, 0 \leq x, y \leq T\}$

$A = \{(x,y) \,/\, |x-y| > t \,;\, 0 \leq x, y \leq T\}$

$$\Rightarrow \quad P(A) = \frac{mesA}{mes\Omega} = \frac{(T-t)^2}{T^2}$$

IV. Nguyên lí xác suất lớn và nguyên lí xác suất nhỏ

Trong nhiều bài toán thực tế, ta thường gặp các biến cố có xác suất rất nhỏ, tức gần bằng 0. Qua nhiều lần quan sát, người ta thấy rằng: các biến cố có xác suất nhỏ gần như sẽ không xảy ra khi thực hiện phép thử. Trên cơ sở đó có thể đưa ra "Nguyên lý thực tế không thể có của các biến cố có xác suất nhỏ" sau đây: Nếu một biến cố có xác suất rất nhỏ thì thực tế có thể cho rằng trong một phép thử, biến cố đó sẽ không xảy ra.

Việc quy định một mức xác suất được coi là "rất nhỏ" tùy thuộc vào từng bài toán cụ thể. Chẳng hạn: Nếu xác suất để một loại dù không mở khi nhảy dù là 0,01 thì xác suất đó chưa thể coi là nhỏ và ta không nên sử dụng loại dù đó. Song nếu xác suất để một chuyến xe lửa đến ga chậm 10 phút là 0,01 thì ta có thể coi mức xác suất đó là nhỏ tức có thể cho rằng xe lửa đến ga đúng giờ. Một xác suất khá nhỏ mà với nó ta có thể cho rằng: biến cố đang xét không xảy ra trong một phép thử được gọi là mức ý nghĩa. Tùy theo từng bài toán cụ thể, mức ý nghĩa thường được lấy trong khoảng 0,01 đến 0,05.

Tương tự như vậy ta có thể nêu ra "nguyên lý thực tế chắc chắn xảy ra của các biến cố có xác suất lớn" như sau: Nếu một biến cố có xác suất gần bằng 1 thì thực tế có thể cho rằng biến cố đó sẽ xảy ra trong một phép thử. Thông thường người ta lấy trong khoảng từ 0,95 đến 0,99.

1.4. MỘT SỐ CÔNG THỨC TÍNH XÁC SUẤT

I. Công thức cộng xác suất

1. Công thức cộng xác suất cho 2 biến cố A và B bất kì

Nếu hai biến cố A và B bất kì thì

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

Chứng minh

Giả sử phép thử có n kết cục đồng khả năng. Trong đó có m_1 kết cục thuận lợi cho biến cố A và m_2 kết cục thuận lợi cho biến cố B. Vì A và B là 2 biến cố bất kì (không xung khắc) nên sẽ có k kết cục thuận lợi cho biến cố A và B.

Khi đó số kết cục thuận lợi cho biến cố A + B sẽ là: $m_1 + m_2 - k$.

Theo định nghĩa xác suất cổ điển ta có

$$P(A+B) = \frac{m_1 + m_2 - k}{n} = \frac{m_1}{n} + \frac{m_2}{n} - \frac{k}{n} = P(A) + P(B) - P(AB)$$

Ví dụ 1 Một lớp có 50 sinh viên. Trong đó có 20 người học giỏi toán, 30 người học giỏi ngoại ngữ, 10 người giỏi cả toán và ngoại. Chọn ngẫu nhiên một người trong lớp. Tìm xác suất để chọn được một người học giỏi ít nhất một môn trong hai môn toán và ngoại ngữ.

BÀI GIẢI

Gọi A là biến cố chọn được người học giỏi toán.

B là biến cố chọn được người học giỏi ngoại ngữ.

C là biến cố chọn được người học giỏi ít nhất một trong hai môn.

Ta thấy $C = A + B$ (A và B là 2 biến cố không xung khắc).

Do đó: $P(C) = P(A) + P(B) - P(A.B)$

$$= \frac{C_{20}^1}{C_{50}^1} + \frac{C_{30}^1}{C_{50}^1} - \frac{C_{10}^1}{C_{50}^1} = \frac{20}{50} + \frac{30}{50} - \frac{10}{50} = 0,8$$

Mở rộng cho 3 biến cố A, B và C bất kì

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

Chứng minh

P(A+B+C) = P(A + B) + P(C) – P((A + B)C)

= P(A) + P(B) – P(AB) + P(C) – P(AC + BC)

= P(A) + P(B) – P(AB) + P(C) – P(AC) – P(BC) + P(AC.BC)

= P(A) + P(B) + P(C) – P(AB) – P(BC) – P(CA) + P(ABC)

Tổng quát cho n biến cố $A_1, A_2, \dots, A_n$ bất kì

Tương tự như trên,ta có

$$P\left(\sum_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{1\le i<j\le n} P(A_iA_j) +$$

$$\sum_{1\le i<j<k\le n} P(A_iA_jA_k) + \dots + (-1)^{n-1}P(A_1A_2\dots A_n)$$

2. Công thức cộng xác suất cho 2 biến cố A, B xung khắc

Nếu hai biến cố A, B xung khắc với nhau thì

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

Ta nhắc lại định nghĩa 8

Hai biến cố A và B được gọi là xung khắc nhau nếu chúng không thể đồng thời xảy ra trong một phép thử. ($AB = \varnothing$)

Chứng minh

Giả sử phép thử có n kết cục đồng khả năng. Trong đó có m_1 kết cục thuận lợi cho biến cố A và m_2 kết cục thuận lợi cho biến cố B. Vì A và B là 2 biến cố xung khắc nên sẽ không có kết cục thuận lợi cho biến cố A và B.

Khi đó số kết cục thuận lợi cho biến cố A + B sẽ là $m_1 + m_1$

Theo định nghĩa xác suất cổ điển ta có

$$P(A+B) = \frac{m_1 + m_2}{n} = \frac{m_1}{n} + \frac{m_2}{n} = P(A) + P(B)$$

Ví dụ 2 Một hộp có 10 sản phẩm trong đó có 4 sản phẩm hỏng. Lấy ngẫu nhiên 3 sản phẩm. Tính xác suất để có một hoặc hai sản phẩm hỏng.

BÀI GIẢI

Gọi A_1 là biến cố 3 sản phẩm lấy ra có đúng 1 sản phẩm hỏng.

A_2 là biến cố 3 sản phẩm lấy ra có đúng 2 sản phẩm hỏng.

Ta phải tính xác suất biến cố A có 1 hoặc 2 sản phẩm hỏng

Ta có $A = A_1 + A_2$

Ở đây ta thấy ngay A_1 và A_2 là xung khắc vì đã có đúng 1 sản phẩm hỏng thì không thể có 2 sản phẩm hỏng.

$$P(A) = P(A_1 + A_2) = P(A_1) + P(A_2) = \frac{C_4^1 C_6^2}{C_{10}^3} + \frac{C_4^2 C_6^1}{C_{10}^3} = \frac{1}{2} + \frac{3}{10} = \frac{4}{5}$$

Mở rộng cho 3 biến cố A_1, A_2, A_3 xung khắc với nhau

$$P(A) = P(A_1 + A_2 + A_3) = P(A_1) + P(A_2) + P(A_3)$$

Ví dụ 3 Một hộp có 10 sản phẩm trong đó có 4 sản phẩm hỏng. Lấy ngẫu nhiên 3 sản phẩm. Tính xác suất để có ít nhất 1 sản phẩm hỏng.

BÀI GIẢI

Gọi A là biến cố trong 3 sản phẩm lấy ra có ít nhất 1 sản phẩm hỏng.

Gọi A_j là biến cố trong 3 sản phẩm lấy ra có đúng j sản phẩm hỏng j=1,2,3. ta thấy ngay $\{A_1, A_2, A_3\}$ là xung khắc

$$P(A) = P(A_1 + A_2 + A_3) = P(A_1) + P(A_2) + P(A_3)$$

$$= \frac{C_4^1 C_6^2}{C_{10}^3} + \frac{C_4^2 C_6^1}{C_{10}^3} + \frac{C_4^3 C_6^0}{C_{10}^3} = \frac{1}{2} + \frac{3}{10} + \frac{1}{30} = \frac{5}{6}$$

Tổng quát $A_1, A_2, \ldots, A_n$ xung khắc với nhau từng đôi một

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

3. Xác suất của biến cố đối lập

Nếu A và $\overline{A}$ là hai biến cố đối lập với nhau thì $P(A) = 1 - P(\overline{A})$

Ta nhắc lại định nghĩa 11: Biến cố A và $\overline{A}$ gọi là hai biến cố đối lập nhau nếu chúng tạo nên một nhóm biến cố đầy đủ và xung khắc:

$$A + \overline{A} = \Omega ;\quad A\overline{A} = \varnothing$$

Từ định nghĩa ta thấy $\overline{A}$ là biến cố "biến cố A không xảy ra"

Chứng minh

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

suy ra $P(A) = 1 - P(\overline{A})$ ta có điều phải chứng minh

Ví dụ 4 Một hộp có 10 sản phẩm trong đó có 4 sản phẩm hỏng. Lấy ngẫu nhiên 3 sản phẩm. Tính xác suất để có ít nhất 1 sản phẩm hỏng.

BÀI GIẢI

Đây là ví dụ 3, ở đây ta giải bằng cách dùng công thức xác suất của biến cố đối lập sẽ ngắn gọn hơn.

Gọi A là biến cố trong 3 sản phẩm lấy ra có ít nhất 1 sản phẩm hỏng => $\overline{A}$ là biến cố trong 3 sản phẩm lấy ra không có sản phẩm hỏng.

Ta có $P(A) = 1 - P(\overline{A}) = 1 - \frac{C_4^0 C_6^3}{C_{10}^3} = 1 - \frac{20}{120} = \frac{5}{6}$

II. Công thức nhân xác suất

1.Xác suất có điều kiện

Cho hai biến cố A và B. Xác suất của biến cố A với điều kiện biến cố B đã xảy ra được gọi là xác suất có điều kiện.

Kí hiệu $P(A/B)$

Xác suất $P(A/B)$ được tính theo công thức

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

Chứng minh

$$P(A/B) = \frac{[AB]}{[B]} = \frac{[AB]/[\Omega]}{[B]/[\Omega]} = \frac{P(AB)}{P(B)}$$

Tương tự ta có $$P(B/A) = \frac{P(AB)}{P(A)}$$

Ví dụ 5 Tung hai con xúc xắc.Gọi A là biến cố con xúc xắc thứ nhất xuất hiện mặt 1 chấm.B là biến cố con xúc xắc thứ hai xuất hiện mặt có số chấm lớn hơn con xúc xắc thứ nhất. Tính P(A/B).

BÀI GIẢI

Ta giả sử biến cố B đã xảy ra:

B = {12, 13, 14, 15, 16, 23, 24, 25, 26, 34, 35, 36, 45, 46, 56}

Ta có 15 phần tử (trong số 15 khả năng xảy ra biến cố B ở trên có 5 khả năng xảy ra biến cố A)

P(A/B) chính là khả năng xảy ra biến cố A khi biến cố B đã xảy ra

$$\Rightarrow \quad P(A/B) = \frac{5}{15} = \frac{1}{3}$$

Ví dụ 6 Một công ty cần tuyển 2 nhân viên. Có 6 người nộp đơn, trong đó có 4 nữ và 2 nam. Khả năng được tuyển của mỗi người là như nhau.

a) Tính xác suất để có ít nhất một người nữ được chọn.

b) Tính xác suất để cả hai người nữ được chọn, biết rằng có ít nhất một người nữ được chọn.

BÀI GIẢI

a) Gọi A là biến cố có ít nhất một người nữ được chọn.

=> $\overline{A}$ là biến cố không có người nữ nào được chọn

Ta có $$P(A) = 1 - P\left(\overline{A}\right) = 1 - \frac{C_2^2}{C_6^2} = 1 - \frac{1}{15} = \frac{14}{15}$$

b) Gọi *B* là biến cố cả 2 người nữ được chọn

$\Rightarrow$ Ta cần tính: $P(B \mid A) = ?$

Ta có $$P(B \mid A) = \frac{P(BA)}{P(A)} = \frac{\dfrac{C_4^2}{C_6^2}}{\dfrac{14}{15}} = \frac{\dfrac{2}{5}}{\dfrac{14}{15}} = \frac{3}{7}$$

2. Công thức nhân xác suất

a) Công thức nhân xác suất cho hai biến cố A và B bất kì

Nếu hai biến cố A và B là bất kì thì

$$P(AB) = P(B)P(A/B) = P(A)P(B/A)$$

Chứng minh

Thật vậy từ định nghĩa xác suất có điều kiện

$$P(A/B)=\frac{P(AB)}{P(B)} \Rightarrow P(AB)=P(B)P(A/B)$$

Tương tự ta có $P(B/A)=\dfrac{P(AB)}{P(A)} \Rightarrow P(AB)=P(A)P(B/A)$

Ví dụ 7 Một hộp có 10 sản phẩm trong đó có 2 sản phẩm hỏng. Lấy ngẫu nhiên thứ tự không hoàn lại từ trong hộp ra 2 sản phẩm. Tính xác suất để 2 sản phẩm lấy ra là 2 sản phẩm hỏng.

BÀI GIẢI

Gọi A là biến cố sản phẩm lấy lần thứ nhất là hỏng

B là biến cố sản phẩm lấy lần thứ hai là hỏng

$$\Rightarrow P(AB)=P(A)P(B/A)=\frac{C_2^1}{C_{10}^1}.\frac{C_1^1}{C_9^1}=\frac{2}{10}\times\frac{1}{9}=\frac{1}{45}$$

Mở rộng cho 3 biến cố A_1, A_2, A_3 bất kì

$$P(A_1A_2A_3)=P(A_1)P(A_2/A_1)P(A_3/A_1A_2)$$

Chứng minh

$$P(A_1A_2A_3)=P(A_1A_2)P(A_3/A_1A_2)=P(A_1)P(A_2/A_1)P(A_3/A_1A_2)$$

Ví dụ 8 Một hộp có 10 sản phẩm trong đó có 3 sản phẩm hỏng. Lấy ngẫu nhiên có thứ tự không hoàn lại từ trong hộp ra 3 sản phẩm để kiểm tra. Tính xác suất để 3 sản phẩm lấy ra đều là sản phẩm hỏng.

BÀI GIẢI

Gọi A là biến cố 3 sản phẩm lấy ra là sản phẩm hỏng

A_i là biến cố sản phẩm lấy lần thứ i là sản phẩm hỏng (i=1,2,3)

$$P(A)=P(A_1A_2A_3)=P(A_1)P(A_2/A_1)P(A_3/A_1A_2)$$

$$=\frac{C_3^1}{C_{10}^1}.\frac{C_2^1}{C_9^1}.\frac{C_1^1}{C_8^1}=\frac{3}{10}\times\frac{2}{9}\times\frac{1}{8}=\frac{1}{120}$$

Tổng quát cho n biến cố $A_1, A_2, \ldots, A_n$ bất kì.

$$P(A_1A_2\ldots A_{n-1}A_n)=P(A_1)P(A_2/A_1)\ldots P(A_n/A_1A_2\ldots A_{n-1})$$

Chứng minh tương tự như trên

b) Công thức nhân xác suất cho hai biến cố A và B độc lập

Nếu A, B là hai biến cố độc lập thì $P(AB)=P(A).P(B)$

Ta nhắc lại định nghĩa 12: Hai biến cố A và B được gọi là độc lập với nhau nếu việc xảy ra hay không xảy ra của biến cố này không làm thay đổi sự xuất hiện của biến cố kia và ngược lại.

Chú ý: * Nếu A và B độc lập, thì A và $\overline{B}$; $\overline{A}$ và B ; $\overline{A}$ và $\overline{B}$ cũng độc lập với nhau.

* Ta định nghĩa hai biến cố độc lập theo xác suất như sau:

A và B là hai biến cố độc lập nếu $P(A) = P(A/B)$

hay $P(B) = P(B/A)$

Mở rộng: Nếu A_1, A_2 , A_3 là các biến cố độc lập toàn phần thì:

$$P(A_1A_2\,A_3) = P(A_1).P(A_2).P(A_3)$$

Ta nhắc lại định nghĩa 14: Các biến cố $A_1, A_2, \ldots, A_n$ được gọi là độc lập toàn phần nếu mỗi biến cố độc lập với tích của một tổ hợp bất kỳ trong các biến cố còn lại.

Ví dụ 9 Một công nhân đứng 3 máy. Xác suất để trong 1 ca làm việc máy I không hư hỏng là 0,7, máy II không hư hỏng là 0,8 và máy III không hư hỏng là 0,9. Tìm xác suất để trong ca làm việc

a) Cả 3 máy không hư hỏng.

b) Cả 3 máy đều hư hỏng.

c) Ít nhất 1 máy hư hỏng.

d) Ít nhất 1 máy không hỏng.

BÀI GIẢI

Gọi A_1, A_2, A_3 là các biến cố tương ứng với máy I, II, III không hư hỏng. Ta có $\{A_1, A_2, A_3\}$ độc lập toàn phần.

a) Gọi A là biến cố 3 máy không bị hỏng. Suy ra $A = A_1 .A_2 .A_3$

Do tính độc lập toàn phần của các biến cố nên:

$P(A) = P(A_1\,A_2\,A_3) = P(A_1)P(A_2)P(A_3) = 0{,}7 \,.\, 0{,}8 \,.\, 0{,}9 = 0{,}504$

b) Gọi B là biến cố cả 3 máy bị hỏng. Suy ra $B = \overline{A_1}\;\overline{A_2}\;\overline{A_3}$

$$P(B) = P(\overline{A_1})P(\overline{A_2})P(\overline{A_3}) = 0{,}3 \times 0{,}2 \times 0{,}1 = 0{,}006$$

c) Gọi C là biến cố có ít nhất 1 máy hư hỏng.

$P(C) = 1-P(\overline{C})=1 - P(A_1\ A_2\ A_3) = 1 - 0{,}504 = 0{,}496$

d) Gọi D là biến cố có ít nhất 1 máy không hỏng.

$P(D) = 1-P(\overline{D})=1 - P(\overline{A_1}.\overline{A_2}.\overline{A_3}) = 1 - 0{,}006 = 0{,}994$

Tổng quát: Nếu $A_1, A_2, \ldots, A_n$ là các biến cố độc lập toàn phần thì

$$P(A_1A_2\ldots A_n) = P(A_1).P(A_2)\ldots P(A_n)$$

Ví dụ 10 Tung 10 lần một đồng xu, tính xác suất để cả 10 lần đều ra mặt sấp.

BÀI GIẢI

Gọi A_i là biến cố lần tung thứ i ra mặt sấp, i = 1, … ,10

Vì $A_1, A_2, A_3, \ldots, A_{10}$ độc lập toàn phần nên:

$P(A_1\ A_2\ A_3 \ldots A_{10}) = P(A_1)P(A_2)P(A_3) \ldots P(A_{10})$

$$= \frac{1}{2}\times\frac{1}{2}\times\ldots\times\frac{1}{2} = \frac{1}{2^{10}}$$

ĐỘ TIN CẬY CỦA MỘT HỆ THỐNG (Phần đọc thêm)

Ta xét một hệ thống gồm *n* bộ phận $\beta_1, \beta_2, \ldots, \beta_n$ hoạt động độc lập với nhau. Xác suất để bộ phận β_i hoạt động tốt là $p_i\ (i=1,2,...,n)$.

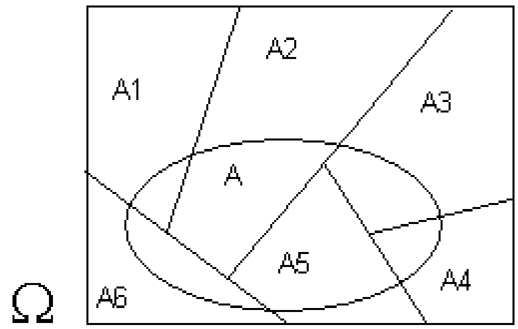
- Trường hợp *n* bộ phận này được mắc nối tiếp, có nghĩa là hệ thống hoạt động tốt khi mọi bộ phận đều hoạt động tốt. Xác suất để hệ thống hoạt động tốt là $p = p_1p_2...p_n$ .Vì $p_i \in (0,1)$ $(i=1,2,...,n)$ nên nếu *n* khá lớn thì *p* sẽ rất nhỏ mặc dù mỗi bộ phận có độ tin cậy cao (p_i gần bằng 1)

- Trường hợp *n* bộ phận này được mắc song song, nghĩa l hệ thống hoạt động tốt khi có ít nhất một bộ phận hoạt động tốt. Gọi *A* là biến cố để bộ phận hoạt động tốt, ta có: $P(A) = 1 - P(\overline{A}) = 1 - q_1q_2...q_n$

$(q_i = 1 - p_i)$ Xác suất để hệ thống hoạt động tốt khá gần 1 khi n khá lớn.Nguyên tắc này không chỉ áp dụng trong kỹ thuật mà còn được áp dụng trong quản lý, đầu tư, kinh doanh…

III. Công thức xác suất đầy đủ (Công thức xác suất toàn phần)

Cho một biến cố A và các biến cố $A_1, A_2, \dots, A_n$ là một nhóm đầy đủ và xung khắc.



Khi đó, ta có công thức xác suất toàn phần

$$P(A)=P(A_1)P(A/A_1)+P(A_2)P(A/A_2)+\dots+P(A_n)P(A/A_n)$$

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

Chứng minh

Do $A_1, A_2, \dots, A_n$ là hệ đầy đủ

Ta có $A=\Omega A=(A_1+A_2+\dots+A_n)A$

$A=A_1A+A_2A+\dots+A_nA$. Suy ra

$$P(A)=P(A_1A+A_2A+\dots+A_nA)$$
$$=P(A_1A)+P(A_2A)+\dots+P(A_nA)$$

(do $A_1, A_2, \dots, A_n$ xung khắc từng đôi nên $A_1A, A_2A, \dots, A_nA$ cũng xung khắc từng đôi).

Theo công thức nhân cho 2 biến cố bất kỳ ta có

$$P(A)=P(A_1)P(A/A_1)+P(A_2)P(A/A_2)+\dots+P(A_n)P(A/A_n)$$
$$=\sum_{i=1}^{n} P(A_i)P(A/A_i)$$

Ví dụ 11 Có 2 lô sản phẩm:
Lô 1 có 20 sản phẩm, trong đó có 15 sản phẩm tốt;
Lô 2 có 20 sản phẩm, trong đó có 10 sản phẩm tốt.
Lấy ngẫu nhiên ra một lô và từ lô này, chọn ngẫu nhiên ra 1 sản phẩm. Tính xác suất để sản phẩm lấy ra là sản phẩm tốt.

BÀI GIẢI

Gọi A là biến cố sản phẩm lấy ra sau cùng là sản phẩm tốt.

A_i là biến cố chọn được lô thứ i, $i = 1,2$

Ta cần tính: $P(A) = ?$

Khi đó, $\{A_1, A_2\}$ là hệ đầy đủ và xung khắc; $P(A_1) = P(A_2) = \dfrac{1}{2}$

Áp dụng công thức xác suất toàn phần, ta có

$$P(A) = P(A_1).P(A \mid A_1) + P(A_2).P(A \mid A_2)$$

$$= \frac{1}{2}\frac{C_{15}^1}{C_{20}^1} + \frac{1}{2}\frac{C_{10}^1}{C_{20}^1} = \frac{1}{2}\cdot\frac{15}{20} + \frac{1}{2}\cdot\frac{10}{20} = \frac{5}{8}$$

Ví dụ 12 Cho một bình có 10 viên bi trong đó có 2 bi đỏ. Cho hai người lần lượt bốc ngẫu nhiên mỗi người 1 bi (không hoàn lại), nếu bốc trúng bi đỏ thì có thưởng.

Hỏi người bốc trước hay bốc sau có lợi hơn?

BÀI GIẢI

Gọi A là biến cố người bốc trước trúng thưởng. Ta có $P(A) = \dfrac{2}{10}$

B là biến cố người bốc sau trúng thưởng

Do $\{A, \overline{A}\}$ là một nhóm đầy đủ và xung khắc nên theo công thức xác suất toàn phần

$$P(B) = P(A)P(B/A) + P(\overline{A})P(B/\overline{A}) = \frac{C_2^1}{C_{10}^1}\frac{C_1^1}{C_9^1} + \frac{C_8^1}{C_{10}^1}\frac{C_2^1}{C_9^1}$$

$$= \frac{2}{10}\times\frac{1}{9} + \frac{8}{10}\times\frac{2}{9} = \frac{2}{10} \Rightarrow P(A) = P(B) = \frac{2}{10}$$

Vậy bốc trước hay bốc sau là như nhau.

Qua ví dụ trên ta có thể lí giải vì sao người ta hay đưa ra đề nghị nên bốc thăm là để đảm bảo sự công bằng.

V. Công thức Bayes

Cho biến cố A và một nhóm đầy đủ và xung khắc $\{A_1, A_2, \dots, A_n\}$. Giả sử biến cố A đã xảy ra. Khi đó, ta có công thức Bayes

$$P(A_i \mid A) = \frac{P(A_i)P(A \mid A_i)}{\sum_{k=1}^{n} P(A_k)P(A \mid A_k)}$$

Chứng minh

$$P(A_i/A) = \frac{P(A_iA)}{P(A)} = \frac{P(A_i)P(A/A_i)}{P(A)} = \frac{P(A_i)P(A/A_i)}{\sum_{k=1}^{n} P(A_k)P(A/A_k)}$$

Chú ý: $P(A/A_i)$ gọi là xác suất tiên nghiệm.

$P(A_i/A)$ gọi là xác suất hậu nghiệm.

Ví dụ 13 Một nhà máy có 3 máy cùng sản xuất một loại sản phẩm. Máy I sản xuất 40% tổng sản lượng của nhà máy.

Máy II sản xuất 35% tổng sản lượng của nhà máy.

Máy III sản xuất 25% tổng sản lượng của nhà máy.

Tỉ lệ phế phẩm của 3 máy lần lượt là 0,04; 0,03; 0,02.

Lấy ngẫu nhiên một sản phẩm từ trong kho của nhà máy.

a) Tính xác suất để đó là phế phẩm.

b) Giả sử sản phẩm lấy được là phế phẩm, tính xác suất để sản phẩm đó là của máy I.

BÀI GIẢI

a) Gọi A là biến cố sản phẩm lấy được là phế phẩm

Gọi A_i là biến cố lấy được 1 sản phẩm của máy thứ i (i=1,2,3)

Khi đó $\{A_1, A_2, A_3\}$ là một hệ đầy đủ.

Theo công thức xác suất toàn phần

$$P(A) = P(A_1)P(A/A_1) + P(A_2)P(A/A_2) + P(A_3)P(A/A_3)$$

$= 0.4\times 0.04 + 0.35\times 0.03 + 0.25\times 0.02 = 0.0315$

b) Ta có: $P(A_1/A) = \dfrac{P(A_1A)}{P(A)} = \dfrac{P(A_1)P(A/A_1)}{P(A)}$

$$= \frac{0.4\times 0.04}{0.0315} = \frac{32}{63} \approx 0,5079$$

VI. Công thức Bernoulli

Giả sử tiến hành n phép thử độc lập. Trong mỗi phép thử chỉ có thể xảy ra một trong hai trường hợp: hoặc biến cố A xảy ra hoặc A không xảy ra. Xác suất xảy ra biến cố A trong mỗi phép thử đều bằng p và xác suất A không xảy ra bằng 1 - p = q.

Khi đó xác suất để trong n phép thử độc lập nói trên biến cố A xảy ra đúng k lần được tính theo công thức Bernoulli như sau

$$P_n(k) = C_n^k p^k q^{n-k}; (k = 0, 1, 2, ..., n)$$

Ví dụ 14 Một lô hàng có tỉ lệ phế phẩm là 0,05. Lấy ngẫu nhiên có thứ tự nhưng hoàn lại từ lô hàng đó ra 5 sản phẩm để kiểm tra.

Tìm xác suất để có 2 phế phẩm trong 5 sản phẩm lấy ra kiểm tra.

BÀI GIẢI

Nếu ta coi việc kiểm tra một sản phẩm là một phép thử, theo giả thiết ta có 5 phép thử độc lập.

Gọi A là biến cố "sản phẩm lấy ra kiểm tra là phế phẩm".

Ta thấy trong mỗi phép thử chỉ có thể xảy ra một trong hai trường hợp:

-hoặc sản phẩm kiểm tra là phế phẩm (tức A xảy ra)

-hoặc sản phẩm kiểm tra là sản phẩm tốt (tức A không xảy ra).

Xác suất để A xảy ra trong mỗi lần kiểm tra đều bằng 0,05.

Xác suất để A không xảy ra trong mỗi lần kiểm tra đều bằng 0,95

Vậy bài toán thỏa mãn lược đồ Bernoulli.

Do đó xác suất để có 2 phế phẩm trong 5 sản phẩm lấy ra kiểm tra là:

$$P_5(2) = C_5^2 (0,05)^2 (0,95)^3 \approx 0,0214$$

Ví dụ 15 Trong một hộp có 20 bóng đèn, trong đó có 5 bóng hỏng. Lấy ngẫu nhiên có thứ tự (hoàn lại) 3 bóng để dùng.

Tìm xác suất để

a) Có 3 bóng đều hỏng.

b) Có 3 bóng đều không hỏng.

c) Có ít nhất 1 bóng không hỏng.

BÀI GIẢI

Ta thấy trong mỗi lần lấy chỉ có thể xảy ra một trong hai trường hợp:

- hoặc bóng hỏng (tức A xảy ra)
- hoặc bóng tốt (tức A không xảy ra).

Xác suất để A xảy ra trong mỗi lần lấy đều bằng 5/20=0,25.

Vậy bài toán thỏa mãn lược đồ Bernoulli với n=5; p=0,25 ;q=0,75

a) B là biến cố có 3 bóng đều hỏng

$$P(B) = C_3^3 (0,25)^3 (0,75)^0 = \frac{1}{64} \approx 0,0156$$

b) C là biến cố có 3 bóng đều không hỏng

$$P(C) = C_3^0 (0,25)^0 (0,75)^3 = \frac{27}{64} = 0,421875$$

c) D là biến cố có ít nhất 1 bóng không hỏng

$$P(D) = 1-P(\overline{D}) = 1- C_3^3 (0,25)^3 (0,75)^0 = 1 - \frac{1}{64} = \frac{63}{64}$$

BÀI TẬP MẪU CHƯƠNG I

1. Cho $P(A)=\frac{1}{3}$, $P(B)=\frac{1}{2}$ và $P(A+B)=\frac{3}{4}$.

Tính $P(AB)$, $P(\bar{A}\bar{B})$, $P(\bar{A}+\bar{B})$, $P(A\bar{B})$ và $P(\bar{A}B)$.

BÀI GIẢI

Từ lý thuyết tập hợp ta có: $\bar{A}\bar{B}=\overline{A+B}$; $\bar{A}+\bar{B}=\overline{AB}$
Do hai biến cố A và B bất kì: $P(A+B)=P(A)+P(B)-P(AB)$,

ta suy ra : $P(AB)=P(A)+P(B)-P(A+B)=\frac{1}{12}$.

Do $\bar{A}\bar{B}=\overline{A+B}$, nên $P\left(\bar{A}\bar{B}\right)=P\left(\overline{A+B}\right)=1-P\left(A+B\right)=\frac{1}{4}$.

Tương tự, vì $\bar{A}+\bar{B}=\overline{AB}$ ta suy ra $P\left(\bar{A}+\bar{B}\right)=1-P\left(AB\right)=\frac{11}{12}$.

Xuất phát từ đẳng thức $B+\bar{B}=\Omega\Rightarrow A=A(B+\bar{B})=AB+A\bar{B}$ và vì AB, $A\bar{B}$ là các biến cố xung khắc, ta được $P(A)=P\left(AB\right)+P\left(A\bar{B}\right)$ và do đó

$$P\left(A\bar{B}\right)=P(A)-P\left(AB\right)=\frac{1}{4}.$$

Tương tự, ta có $P\left(\bar{A}B\right)=P(B)-P\left(AB\right)=\frac{5}{12}$.

2. Hộp thứ nhất có 5 lọ thuốc tốt và 3 lọ kém phẩm chất.

Hộp thứ hai có 3 lọ thuốc tốt và 2 lọ kém phẩm chất.

Lấy ngẫu nhiên từ mỗi hộp ra một lọ. Tìm các xác suất:

a) Lấy được 2 lọ thuốc tốt?

b) Lấy được 1 lọ tốt và 1 lọ kém phẩm chất?

c) Nếu lấy được 1 lọ tốt và 1 lọ kém phẩm chất. Tìm xác suất để lọ kém phẩm chất là của hộp thứ nhất?

BÀI GIẢI

a) Gọi A là biến cố lấy được 2 lọ thuốc tốt.

A_1 là biến cố lấy được 1 lọ thuốc tốt ở hộp 1.

A_2 là biến cố lấy được 1 lọ thuốc tốt ở hộp 2.

Vì A_1 và A_2 là 2 biến cố độc lập nên: $A = A_1A_2$

Suy ra $P(A) = P(A_1).P(A_2) = \frac{C_5^1}{C_8^1}.\frac{C_3^1}{C_5^1} = \frac{3}{8}$

b) Gọi B là biến cố lấy được 1 lọ tốt và 1 lọ kém phẩm chất

$\overline{A_1}$ là biến cố lấy được 1 lọ thuốc kém phẩm chất ở hộp 1

$\overline{A_2}$ là biến cố lấy được 1 lọ thuốc kém phẩm chất ở hộp 2

Vì $B = A_1\overline{A_2} + A_2\overline{A_1}$ nên $P(B) = P(A_1)P(\overline{A_2}) + P(A_2)P(\overline{A_1})$

$$= \frac{C_5^1}{C_8^1}.\frac{C_2^1}{C_5^1} + \frac{C_3^1}{C_5^1}.\frac{C_3^1}{C_8^1} = \frac{19}{40}$$

c) Ta có

$$P(\overline{A}_1 / B) = \frac{P(\overline{A}_1.B)}{P(B)} = \frac{P(\overline{A}_1)P(B / \overline{A}_1)}{P(B)} = \frac{\left(\frac{C_3^1}{C_8^1}.\frac{C_3^1}{C_5^1}\right)}{\frac{19}{40}} = \frac{9}{19}$$

3. Một hộp đựng 3 bi đỏ và 7 bi trắng. Rút ngẫu nhiên từ hộp ra một bi rồi sau đó bỏ vào hộp một bi khác màu với viên bi rút ra, sau đó rút tiếp một bi từ hộp.

a) Tìm xác suất để bi rút ra lần sau từ hộp là viên bi đỏ?

b) Nếu hai bi rút ra cùng màu, tìm xác suất để hai bi đó màu trắng?

BÀI GIẢI

a) Gọi A là biến cố bi rút ra lần sau từ hộp là viên bi đỏ.

A_i là biến cố lấy ra lần i là 1 viên bi đỏ, i=1,2

B_i là biến cố lần i lấy ra 1 viên bi màu trắng, i=1,2

Ta có: $A = A_1.A_2 + B_1.A_2$ thì

$P(A)=P(A_1).P(A_2/A_1)+P(B_1).P(A_2/B_1)$

Do đó $P(A)=\frac{C_3^1}{C_{10}^1}\frac{C_2^1}{C_{10}^1}+\frac{C_7^1}{C_{10}^1}\frac{C_4^1}{C_{10}^1}=\frac{3}{10}.\frac{2}{10}+\frac{7}{10}.\frac{4}{10}=\frac{17}{50}$

b) Gọi B là biến cố hai viên bi rút ra cùng màu.

C là biến cố hai viên bi rút ra có màu trắng.

Ta có: $B = A_1A_2 + B_1B_2$

$P(B) = P(A_1A_2) + P(B_1B_2) = P(A_1).P(A_2/A_1)+P(B_1).P(B_2/B_1)$

$$=\frac{C_3^1}{C_{10}^1}\frac{C_2^1}{C_{10}^1}+\frac{C_7^1}{C_{10}^1}\frac{C_6^1}{C_{10}^1}=\frac{3}{10}.\frac{2}{10}+\frac{7}{10}.\frac{6}{10}=\frac{12}{25}$$

Do đó $P(C/B)=\frac{P(CB)}{P(B)}=\frac{P(B_1B_2)}{P(B)}=\frac{7}{8}$

4. Hộp thứ nhất có 5 lọ thuốc tốt và 3 lọ kém phẩm chất.

Hộp thứ hai có 3 lọ thuốc tốt và 2 lọ kém phẩm chất.

Chọn ngẫu nhiên 1 hộp, từ hộp đã chọn lấy ngẫu nhiên ra 2 lọ.

Tìm các xác suất :

a) Lấy được 2 lọ thuốc tốt?

b) Lấy được 1 lọ tốt và 1 lọ kém phẩm chất?

c) Nếu lấy được 1 lọ tốt và 1 lọ kém phẩm chất. Tìm xác suất để lọ kém phẩm chất là của hộp thứ nhất?

BÀI GIẢI

D_1 là biến cố lấy hộp 1.

D_2 là biến cố lấy hộp 2.

$P(D_1)= P(D_2)= \frac{1}{2}$ và $\{A_1, A_2\}$ là một nhóm đầy đủ và xung khắc

a) Gọi D là biến cố lấy được 2 lọ thuốc tốt.

Áp dụng công thức xác suất toàn phần ta có

$$P(D) = P(D_1)P(D/D_1) + P(D_2)P(D/D_2) = \frac{1}{2}.\frac{C_5^2}{C_8^2} + \frac{1}{2}.\frac{C_3^2}{C_5^2} = \frac{23}{70}$$

b) Gọi E là biến cố lấy được 1 lọ thuốc tốt và 1 lọ thuốc kém phẩm chất.

$P(E) = P(D_1)P(E/D_1) + P(D_2)P(E/D_2)$

$$= \frac{1}{2}.\left(\frac{C_5^1C_3^1}{C_8^2} + \frac{C_3^1C_2^1}{C_5^2}\right) = \frac{159}{280}$$

c) Gọi G là biến cố để lọ thuốc kém phẩm chất là ở hộp 1

$$P(G) = P(D_1 / E) = \frac{P(D_1)P(E / D_1)}{P(E)} = \frac{\frac{1}{2}.\frac{C_5^1C_3^1}{C_8^2}}{P(E)} = \frac{25}{53}$$

5. Cho 2 hộp , hộp I có 3 quả bóng màu đỏ và 4 quả bóng màu trắng, hộp II có 5 quả bóng màu đỏ và 3 quả bóng màu trắng. Bốc ngẫu nhiên 1 quả bóng từ hộp I bỏ vào hộp II, rồi lấy ngẫu nhiên 1 quả bóng từ hộp II bỏ ra ngoài.

Tính xác suất để quả bóng lấy ra từ hộp II là quả bóng màu trắng

BÀI GIẢI

Gọi Ti là biến cố quả bóng lấy ra lần thứ i là màu trắng i=1,2

Ta có $\left\{T_1, \overline{T}_1\right\}$ là 1 nhóm đầy đủ và xung khắc.

Do đó $P(T_2) = P(T_1)P(T_2 / T_1) + P(\overline{T}_1)P(T_2 / \overline{T}_1)$

$$= \frac{C_4^1}{C_7^1}\frac{C_4^1}{C_9^1} + \frac{C_3^1}{C_7^1}\frac{C_3^1}{C_9^1} = \frac{4}{7}.\frac{4}{9} + \frac{3}{7}.\frac{3}{9} = \frac{25}{63} \approx 0,3968$$

6. Hai hộp phấn giống nhau, mỗi hộp có 5 viên phấn màu trắng và 7 viên phấn màu đỏ. Lấy ngẫu nhiên mỗi hộp 1 viên phấn. Từ 2 viên phấn này chọn ngẫu nhiên 1 viên phấn. Tính xác suất để viên phấn chọn ra sau cùng là viên phấn màu trắng

BÀI GIẢI

Gọi T_i là biến cố viên phấn lấy ra từ hộp i là viên phấn màu trắng, i = 1, 2

A_i là biến cố 2 viên phấn lấy ra có đúng i viên phấn màu trắng, i = 0, 1, 2

F là biến cố viên phấn lấy ra sau cùng là viên phấn màu trắng.

$\left\{T_i,\overline{T}_i\right\},\left\{A_0,A_1,A_2\right\}$ là những nhóm đầy đủ và xung khắc.

$\left\{T_1,T_2\right\}$ là nhóm độc lập

$A_0=\overline{T}_1.\overline{T}_2$ nên

$$P\left(A_0\right)=P\left(\overline{T}_1\right).P\left(\overline{T}_2\right)=\frac{7}{12}.\frac{7}{12}=\frac{49}{144}$$

$A_1=T_1.\overline{T}_2+\overline{T}_1.T_2$ là tổng của 2 biến cố xung khắc nhau

$$P\left(A_1\right)=P\left(T_1.\overline{T}_2\right)+P\left(\overline{T}_1.T_2\right)=P\left(T_1\right).P\left(\overline{T}_2\right)+P\left(\overline{T}_1\right).P\left(T_2\right)$$

$$=\frac{C_5^1}{C_{12}^1}\frac{C_7^1}{C_{12}^1}+\frac{C_7^1}{C_{12}^1}\frac{C_5^1}{C_{12}^1}=\frac{5}{12}.\frac{7}{12}+\frac{7}{12}.\frac{5}{12}=\frac{70}{144}=\frac{35}{72}$$

$$P\left(A_2\right)=P\left(T_1\right).P\left(T_2\right)=\frac{C_5^1}{C_{12}^1}\frac{C_5^1}{C_{12}^1}=\frac{5}{12}.\frac{5}{12}=\frac{25}{144}$$

$$P(F)=P\left(A_0\right)P\left(F/A_0\right)+P\left(A_1\right)P\left(F/A_1\right)+P\left(A_2\right)P\left(F/A_2\right)$$

$$==\frac{49}{144}.\frac{0}{2}+\frac{70}{144}.\frac{1}{2}+\frac{25}{144}.\frac{2}{2}=\frac{5}{12}$$

BÀI TẬP CHƯƠNG I

1.1. Cô dâu và chú rể mời 4 người bạn đứng thành 1 hàng để chụp ảnh cùng với mình. Có bao nhiêu cách xếp hàng nếu:

a) Cô dâu đứng cạnh chú rể

b) Cô dâu không đứng cạnh chú rể

c) Cô dâu đứng bên trái chú rể

1.2. Một tổ bộ môn của một trường đại học có 16 giảng viên.

a) Có bao nhiêu cách chọn một hội đồng gồm 5 thành viên của tổ sao cho mỗi thành viên được phân công ở một vị trí đã định.

b) Có bao nhiêu cách chọn một hội đồng gồm 5 thành viên của tổ.

c) Trong 16 giảng viên có 7 nữ và 9 nam. Có bao nhiêu cách chọn một hội đồng gồm 5 thành viên của tổ nếu trong hội đồng có ít nhất 1 nữ và ít nhất 1 nam.

1.3. Kiểm tra 3 sản phẩm. Gọi A_k là biến cố sản phẩm thứ k tốt. Hãy trình bày cách biểu diễn qua A_k các biến cố sau đây

a) A là biến cố cả 3 sản phẩm đều xấu.

b) B là biến cố cả 3 sản phẩm đều tốt.

c) C là biến cố có đúng 1 sản phẩm xấu.

d) D là biến cố có ít nhất một sản phẩm xấu.

c) C là biến cố không phải tất cả sản phẩm đều tốt.

f) F là biến cố có ít nhất 2 sản phẩm tốt.

g) G là biến cố có ít nhất một sản phẩm tốt.

1.4. Có 3 hộp, mỗi hộp đựng 5 quả cầu, trong đó hộp thứ i có i quả cầu màu trắng (i=1, 2, 3). Lấy ngẫu nhiên từ mỗi hộp ra một quả cầu.

a) Tìm xác suất để lấy được 3 quả cầu màu trắng?

b) Tìm xác suất để trong 3 quả cầu lấy ra có 1 quả cầu màu trắng?

c) Nếu trong 3 quả cầu lấy ra có một quả cầu màu trắng, tìm xác suất để quả cầu màu trắng đó là của hộp thứ nhất?

1.5. Có 3 hộp, mỗi hộp đựng 5 viên phấn, trong đó hộp thứ i có i viên phấn trắng (i=1, 2, 3). Chọn ngẫu nhiên một hộp, rồi từ hộp đã chọn lấy ngẫu nhiên một lần ra 3 viên phấn

a) Tìm xác suất để lấy được 3 viên phấn trắng?

b) Nếu trong 3 viên phấn lấy ra có 1 viên phấn trắng, tìm xác suất để viên phấn trắng đó là của hộp thứ nhất?

1.6. Trong một hộp có 15 bóng đèn, trong đó có 3 bóng hỏng. Lấy ngẫu nhiên có thứ tự không hoàn lại 3 bóng để dùng. Tìm xác suất:

a) Cả 3 bóng đều hỏng.

b) Cả 3 bóng đều không hỏng.

c) Có ít nhất 1 bóng không hỏng.

d) Chỉ có bóng thứ 2 hỏng.

1.7. Một nhà máy có 3 phân xưởng cùng sản xuất một loại sản phẩm. Phân xưởng I sản xuất 25%; phân xưởng II sản xuất 25% và phân xưởng III sản xuất 50% tổng số sản phẩm của toàn nhà máy. Tỉ lệ phế phẩm của từng phân xưởng tương ứng là 1%; 5%; 10%. Lấy ngẫu nhiên một sản phẩm từ lô hàng do 3 phân xuởng sản xuất.

a) Tìm xác suất để lấy sản phẩm tốt ? Ý nghĩa thực tế của kết quả đó là gì?

b) Nếu lấy được một sản phẩm xấu, theo bạn thì sản phẩm đó do phân xưởng nào sản xuất? Tại sao?

1.8. Một mạch điện mắc song song sẽ hoạt động được nếu có ít nhất một thành phần của nó hoạt động. Xét mạch điện mắc song song có 3 thành phần hoạt động độc lập, với xác suất hoạt động mỗi thành phần là 0,5. Tính xác suất có 1 thành phần hoạt động, biết mạch đó hoạt động bình thường.

1.9. Tỷ lệ phế phẩm của một lô hàng (lớn) là 1%. Từ lô hàng này, lấy ra n sản phẩm. Hỏi n ít nhất phải là bao nhiêu để xác suất nhận được ít nhất một phế phẩm lớn hơn 0,95

CHƯƠNG II

ĐẠI LƯỢNG NGẪU NHIÊN VÀ CÁC QUY LUẬT PHÂN PHỐI XÁC SUẤT ĐẶC BIỆT

2.1. ĐẠI LƯỢNG NGẪU NHIÊN (BIẾN NGẪU NHIÊN)

Trong thực tế kết quả của phép thử được biểu diễn bởi một đại lượng Trong chương I ta đã nghiên cứu các biến cố ngẫu nhiên, trong chương II ta quan tâm đến các đại lượng mà chúng nhận giá trị nào đó một cách ngẫu nhiên ta không biết chắc được

I.Định nghĩa

Đại lượng ngẫu nhiên là đại lượng nhận giá trị có thể có với xác suất tương ứng.

Kí hiệu đại lượng ngẫu nhiên: X, Y, Z, … hoặc X_1, X_2,…

Ví dụ Một hộp có 6 viên phấn màu trắng và 4 viên phấn màu đỏ. Lấy ngẫu nhiên 3 viên phấn. Gọi X là số viên phấn màu trắng trong 3 viên phấn lấy ra. X có thể có các giá trị: 0,1,2,3

Ví dụ Một xạ thủ bắn vào một tấm bia, các viên đạn được bắn lần lượt và xạ thủ dừng bắn khi có một viên đạn trúng bia. Gọi X là số lần bắn. X có thể có các giá trị: 1,2,3…n…

II. Phân loại đại lượng ngẫu nhiên

1. Đại lượng ngẫu nhiên rời rạc

Là đại lượng ngẫu nhiên có tập hợp các giá trị là hữu hạn hoặc vô hạn đếm được

Ví dụ Số trẻ em sơ sinh, số người bị một loại bệnh nào đó, số cơn bão, …

2. Đại lượng ngẫu nhiên liên tục

Là đại lượng ngẫu nhiên có tập hợp các giá trị là một đoạn [a,b] nào đó.

Ví dụ Độ dài , trọng lượng của một loại sản phẩm do một nhà máy sản xuất…

Nhận xét

Đại lượng ngẫu nhiên là đại diện cho một phép thử. Thông thường người ta quan tâm đến giá trị của đại lượng ngẫu nhiên và xác suất của nó hơn là phép thử.

III. Bảng phân phối xác suất

Giả sử đại lượng ngẫu nhiên X có thể nhận một trong các giá trị:

$x_1, x_2, \dots, x_n$ với các xác suất tương ứng là $p_1, p_2, \dots, p_n$.
Bảng phân phối xác suất là một bảng gồm hai dòng:
- Dòng thứ nhất gồm các giá trị có thể có của đại lượng ngẫu nhiên
- Dòng thứ hai là các xác suất tương ứng.

Bảng phân phối xác suất của X có dạng

X	x_1	x_2	…	x_n
P	P_1	p_2	…	P_n

Với: $\sum_{i=1}^{n} p_i = 1$

Bảng này cho biết quy luật phân phối xác suất của X

Ví dụ 1 Một hộp phấn có 6 viên màu trắng và 4 viên màu đỏ. Lấy ngẫu nhiên 3 viên. Gọi X là số viên màu trắng trong 3 viên lấy ra.

Tìm bảng phân phối xác suất của X

BÀI GIẢI

X có thể có các giá trị: 0,1,2,3

P(X=0) = P{3 viên màu đỏ} = $\dfrac{C_4^3}{C_{10}^3} = \dfrac{4}{120}$

P(X=1) = P{1 viên màu trắng} = $\dfrac{C_6^1 C_4^2}{C_{10}^3} = \dfrac{36}{120}$

P(X=2) = P{2 viên màu trắng} = $\dfrac{C_6^2 C_4^1}{C_{10}^3} = \dfrac{60}{120}$

P(X=3) = P{3 viên màu trắng} = $\dfrac{C_6^2 C_4^1}{C_{10}^3} = \dfrac{60}{120}$ $\dfrac{C_6^3}{C_{10}^3} = \dfrac{20}{120}$

Bảng phân phối xác suất

X	0	1	2	3
P	$\frac{4}{120}$	$\frac{36}{120}$	$\frac{36}{120}$	$\frac{20}{120}$

Ví dụ 2 Thùng A có 20 lọ thuốc trong đó có 2 lọ hỏng.

Thùng B có 20 lọ thuốc trong đó có 3 lọ hỏng.

Lấy ở mỗi thùng 1 lọ thuốc. Gọi X là số lọ hỏng trong hai lọ lấy ra.

Tìm bảng phân phối xác suất của X.

BÀI GIẢI

X có thể có các giá trị: 0,1,2

A là biến cố nhận được 1 lọ hỏng từ thùng A

B là biến cố nhận được 1 lọ hỏng từ thùng B

Chú ý : A, B là các biến cố độc lập. Ta có

$$P(X=0)=P(\bar{A}\bar{B})=P(\bar{A})P(\bar{B})=\frac{18}{20}\cdot\frac{17}{20}=\frac{306}{400}=0.765$$

$$P(X=1)=P(A\bar{B}+\bar{A}B)=P(A)P(\bar{B})+P(\bar{A})P(B)$$
$$=\frac{2}{20}\cdot\frac{17}{20}+\frac{18}{20}\cdot\frac{3}{20}=\frac{88}{400}=0.22$$

$$P(X=2)=P(AB)=P(A)P(B)=\frac{2}{20}\cdot\frac{3}{20}=\frac{6}{400}=0.015$$

Từ đó ta được bảng phân phối xác suất

X	0	1	2
P	0.765	0.22	0.015

V. Hàm phân phối xác suất

Ngoài bảng phân phối xác suất, quy luật phân phối xác suất của đại lượng ngẫu nhiên còn được xác định bằng hàm phân phối xác suất

1. Định nghĩa

X là đại lượng ngẫu nhiên thì hàm phân phối xác suất của X kí hiệu là F(x) và định nghĩa F(x)= P(X < x).

Nếu X là đại lượng ngẫu nhiên rời rạc thì hàm F(x) có dạng

$$F(x) = \sum_{x_i < x} P(X = x_i) = \sum_{x_i < x} p_i$$

nghĩa là với mỗi x thì F(x) là xác suất để X lấy giá trị trong nửa đường thẳng (-∞, x]

2. Tính chất

a) $0 \le F(x) \le 1$

b) F(x) là hàm không giảm nghĩa là nếu $x_2 \ge x_1$ thì $F(x_2) \ge F(x_1)$

Chứng minh

Ta có $\{X \le x_2\} = \{X \le x_1\} + \{x_1 < X \le x_2\}$ là hai sự kiện xung khắc

Suy ra

$$P\{X \le x_2\} = P\{X \le x_1\} + P\{x_1 < X \le x_2\}$$
$$F(x_2) = F(x_1) + P\{x_1 < X \le x_2\} \ \ge F(x_1)$$

c) $P(x_1 < X \le x_2) = P\{X \in (x_1, x_2)\} = F(x_2) - F(x_1)$

d) $\lim\limits_{x \to -\infty} F(x) = 0; \lim\limits_{x \to +\infty} F(x) = 1$

Ví dụ 3 Cho đại lượng ngẫu nhiên có bảng phân phối xác suất

X	-1	1	2
P	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$

Tìm hàm phân phối xác suất của X.

BÀI GIẢI

* $x \le -1$: F(x) = P(X<x) = 0

* $-1 < x \le 1$: F(x) = P(X<1) = P(X= -1) =1/2

* $1 < x \le 2$:

F(x) = P(X<2) = P(X = -1) + P(X = 1) =1/2+1/4 = 3/4

* $x > 2$:

F(x) = P(X < +∞) = P(X = -1) + P(X = 1) + P(X = 2)

=1/2 +1/4 + 1/4 =1

Vậy $F(x)=\begin{cases} 0 \text{ khi } x \le -1 \\ \dfrac{1}{2} \text{ khi } -1 < x \le 1 \\ \dfrac{3}{4} \text{ khi } 1 < x \le 2 \\ 1 \text{ khi } x > 2 \end{cases}$

VI. Hàm mật độ xác suất

Để mô tả quy luật phân phối xác suất của đại lượng ngẫu nhiên rời rạc ta dùng bảng phân phối xác suất, trong trường hợp đại lượng ngẫu nhiên liên tục ta dùng hàm mật độ xác suất

1. Định nghĩa

X là đại lượng ngẫu nhiên liên tục thì hàm mật độ xác suất của X kí hiệu là f(x) và định nghĩa f(x) = F'(x) nghĩa là

$$f(x) = \lim_{\Delta x \to o} \frac{F(x+\Delta x) - F(x)}{\Delta x} = \lim_{\Delta x \to o} \frac{P(x < X \le x + \Delta x)}{\Delta x}$$

2. Tính chất

a) $f(x) \ge 0$ với $\forall x \in R$

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

c) $P(x_1 < X \le x_2) = \int_{x_1}^{x_2} f(x)dx$

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

Ví dụ 4 Cho đại lượng ngẫu nhiên X và hàm số

$$f(x) = \begin{cases} \dfrac{x^2}{9} \text{ khi } x \in [0,3] \\ 0 \text{ khi } x \notin [0,3] \end{cases}$$

a) Chứng minh f(x) có phải hàm mật độ của X

b) Tính $P(0 \le X \le 2)$

c) Tìm hàm phân phối F(x).

BÀI GIẢI

a) Kiểm tra hàm mật độ xác suất

$$\int_{-\infty}^{+\infty} f(x)\,dx = \int_{-\infty}^{3} 0dx + \int_{0}^{3} \frac{x^2}{9}dx + \int_{3}^{+\infty} 0dx = \int_{0}^{3} \frac{x^2}{9}dx = \left.\frac{x^3}{27}\right|_0^3 = 1$$

b) Ta có

$$P(0 \le X \le 2) = \int_{0}^{2} f(x)\,dx = \int_{0}^{2} \frac{x^2}{9}dx = \left.\frac{x^3}{27}\right|_0^2 = \frac{8}{27}$$

c) Tìm hàm phân phối xác suất $\quad F(x) = \int_{-\infty}^{x} f(x)dx$

$$x \le 0 \text{ thì } F(x) = \int_{-\infty}^{x} 0dx = 0$$

$$0 < x \le 3 \text{ thì } F(x) = \int_{-\infty}^{x} f(x)dx = \int_{-\infty}^{0} 0dx + \int_{0}^{x} \frac{x^2}{9}dx = \frac{1}{27}x^3$$

$$3 < x < +\infty \text{ thì } F(x) = \int_{-\infty}^{0} 0dx + \int_{0}^{3} \frac{x^2}{9}dx + \int_{3}^{x} 0dx = ... = 1$$

$$\text{Vậy } F(x)= \begin{cases} 0 \text{ khi } x \le 0 \\ \dfrac{x^3}{27} \text{ khi } 0 < x \le 3 \\ 1 \text{ khi } x > 3 \end{cases}$$

3. Ý nghĩa của hàm mật độ f(x)

Hàm mật độ f(x) của đại lượng ngẫu nhiên liên tục X cho biết mức độ tập trung xác suất tại điểm x.

Khác với đại lượng ngẫu nhiên rời rạc, đối với đại lượng ngẫu nhiên liên tục thì $P(X = x_0) = 0$ với mọi $x_0 \in R$.

2.2. CÁC ĐẶC TRƯNG BẰNG SỐ CỦA ĐẠI LƯỢNG NGẪU NHIÊN

Trong thực tế, đối với đại lượng ngẫu nhiên ta không chỉ cần biết quy luật phân phối xác suất của đại lượng ngẫu nhiên (bằng bảng phân phối xác suất, hàm phân phối xác suất, hàm mật độ xác suất) mà còn quan tâm đến những thông tin phản ánh tổng hợp những đặc trưng quan trọng của đại lượng ngẫu nhiên

I. Kì vọng (Expectation)

1. Định nghĩa

Kì vọng của đại lượng ngẫu nhiên X là một số.

Kí hiệu là E(X) hay M(X)

a) Nếu đại lượng ngẫu nhiên X rời rạc có bảng phân phối xác suất

X	x_1	x_2	…	x_{n-1}	x_n
P	P_1	p_2	…	P_{n-1}	P_n

thì $$E(X)=\sum_{i=1}^{n} x_i p_i$$

b) Nếu đại lượng ngẫu nhiên X liên tục có hàm mật độ f(x) thì

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

Ví dụ 1 Từ bảng phân phối xác suất

X	1	2	3	4	5
P	0.6	0.24	0.096	0.0384	0.0256

Tính E(X)=?

BÀI GIẢI

$$E(X)=\sum_i x_i p_i = 1\times 0.6 + 2\times 0.24 + \ldots + 5\times 0.0256 = 1.6496,$$

Ví dụ 2 Cho đại lượng ngẫu nhiên X có hàm mật độ xác suất

$$f(x)=\begin{cases} \dfrac{x^2}{9} \text{khi } x\in[0,3] \\ 0 \text{ khi } x\notin[0,3] \end{cases}$$

Tính E(X).

BÀI GIẢI

$$E(X)=\int_{-\infty}^{+\infty} x.f(x)dx=\int_{-\infty}^{0} x.0dx+\int_{0}^{3}\frac{x^3}{9}dx+\int_{3}^{+\infty} x.0dx=...=\frac{9}{4}$$

2. Tính chất của kì vọng

a) Với C là hằng số thì E(C)= C

b) E(CX) = CE(X)

c) E(X+Y) = E(X) + E(Y)

d) Nếu X,Y độc lập thì E(X.Y) = E(X) . E(Y)

(Hai đại lượng X,Y được gọi là độc lập nếu việc X nhận giá trị nào không ảnh hưởng gì đến xác suất nhận giá trị của Y)

Ví dụ 3 Từ bảng phân phối xác suất

X	1	2	3	4	5
P	0.6	0.24	0.096	0.0384	0.0256

Tính E(Y) với Y=3X +5

BÀI GIẢI

$$E(X)=\sum_{i=1}^{5} x_i p_i = 1\times 0.6+2\times 0.24+...+5\times 0.0256=1.6496,$$

E(Y)=E(3X+5)=E(3X)+E(5)=3E(X)+5=3.1,6496+5=9,9488

Ví dụ 4 Một công ty có 600 nhân viên, bảng sau đây cho biết thu nhập một tháng của nhân viên trong công ty.

Thu nhập (triệu đồng/tháng)	3	3,5	4	5,5	6	10
Số người	46	100	152	200	62	40

Chọn ngẫu nhiên một nhân viên của công ty, gọi X là thu nhập một tháng của nhân viên này thì X là đại lượng ngẫu nhiên có bảng phân phối xác suất như sau

X	3	3,5	4	5,5	6	10
P	$\frac{46}{600}$	$\frac{100}{600}$	$\frac{152}{600}$	$\frac{200}{600}$	$\frac{62}{600}$	$\frac{40}{600}$

$$E(X)=3.\frac{46}{600}+3,5.\frac{100}{600}+4.\frac{152}{600}+5,5.\frac{200}{600}+6.\frac{62}{600}+10.\frac{40}{600}$$

$$=4,95 \text{ (triệu đồng/tháng)}$$

Trong ví dụ này, từ biểu thức tính EX ta thấy EX chính là thu nhập trung bình (triệu đồng/tháng/người) của nhân viên công ty này.
Mặt khác, ta thấy có nhiều nhân viên có thu nhập gần thu nhập trung bình.

2. Ý nghĩa của kỳ vọng

Kỳ vọng của đại lượng ngẫu nhiên X là giá trị trung bình của X

II. Phương sai (Variance)

Trong thực tế, người ta không chỉ quan tâm đến giá trị trung bình của đại lượng ngẫu nhiên mà còn quan tâm đến sự phân tán của các giá trị của đại lượng ngẫu nhiên quanh giá trị trung bình. Người ta dùng phương sai để đo sự phân tán này.

1. Định nghĩa phương sai

Phương sai của đại lượng ngẫu nhiên X là một số, kí hiệu D(X) hay Var(X) định nghĩa là

$$\mathrm{Var}(X) = E\left[(X-E(X))^2\right]$$

Công thức tính phương sai:

$$\mathrm{Var}X = E(X^2)-(EX)^2$$

Ta sử dụng tính chất của kỳ vọng chứng minh
Thật vậy

$$\begin{aligned}\mathrm{Var}(X) &= E\left[(X-E(X))^2\right] = E\left[(X^2) - 2X\,E(X) + (EX)^2\right]\\ &= E(X^2) - 2E(X).E(X) + E(EX)^2\\ &= E(X^2) - 2(EX)^2 + (EX)^2 = E(X^2) - (EX)^2\end{aligned}$$

Suy ra công thức để vận dụng tính toán như sau

a) Nếu X rời rạc thì $\mathrm{Var}(X) = \sum_{i=1}^{n} x_i^2 p_i - [EX]^2$

Ví dụ 5 Từ bảng phân phối xác suất

X	1	2	3	4	5
P	0.6	0.24	0.096	0.0384	0.0256

Tính Var(X)=?

BÀI GIẢI

$$E(X) = \sum_{i=1}^{5} x_i p_i = 1\times 0.6 + 2\times 0.24 + \ldots + 5\times 0.0256 = 1.6496,$$

và phương sai Var(X)= $\sum_{i=1}^{5} x_i^2 p_i$ - $[EX]^2$

$= 1^2 \times 0.6 + 2^2 \times 0.24 + ... + 5^2 \times 0.0256 - (1.6496)^2 \approx 0.95722.$

b) Nếu X liên tục

$$\text{Var(X)} = \int_{-\infty}^{+\infty} x^2 f(x)dx - \left[EX\right]^2$$

Ví dụ 6 Cho đại lượng ngẫu nhiên X có hàm mật độ xác suất

$$f(x) = \begin{cases} \dfrac{x^2}{9} \text{ khi } x \in [0,3] \\ 0 \text{ khi } x \notin [0,3] \end{cases}$$

Tính Var(X).

BÀI GIẢI

$$\text{E(X)} = \int_{-\infty}^{+\infty} \text{x.f(x)dx} = \int_{-\infty}^{0} \text{x.0dx} + \int_{0}^{3} \frac{x^3}{9}\text{dx} + \int_{3}^{+\infty} \text{x.0dx} = ... = \frac{9}{4}$$

$$\text{Var(X)}=\int_{-\infty}^{+\infty} x^2 f(x)dx - \left[EX\right]^2$$

$$= \int_{-\infty}^{0} x^2\text{.0dx} + \int_{0}^{3} \frac{x^4}{9}\text{dx} + \int_{3}^{+\infty} x^2\text{.0dx} - \left(\frac{9}{4}\right)^2 = ... = 27/80$$

2. Tính chất của phương sai

a) Var(C) = 0

b) Var(CX) = C^2 Var(X)

c) Nếu X, Y độc lập thì Var(X+Y) = Var(X) + Var(Y)

3. Ý nghĩa của phương sai

Phương sai là trung bình bình phương các sai lệch của các giá trị của đại lượng ngẫu nhiên đối với trung bình của đại lượng ngẫu nhiên.

Như vậy phương sai cho ta thấy mức độ phân tán các giá trị của X xung quanh giá trị trung bình.

Phương sai càng nhỏ thì độ phân tán càng tập trung. Do đó phương sai thường dùng để đo độ chính xác của sản xuất.

CÁC VÍ DỤ TỔNG QUÁT

Ví dụ 7 Ba xạ thủ độc lập bắn vào bia, mỗi xạ thủ có một viên đạn. Xác suất để từng xạ thủ bắn trúng bia lần lượt là 0,8 ; 0,7; 0,6.

Gọi X là số viên đạn bắn trúng bia.

a) Lập bảng phân phối xác suất cho X.

b) Tính kỳ vọng E(X) và phương sai Var(X).

c) Tính xác suất $P(2 \le X \le 7)$.

d) Tìm hàm phân phối xác suất của X.

BÀI GIẢI

a) Gọi A_i là biến cố người thứ i bắn trúng bia (i=1, 2, 3)

$\Rightarrow \overline{A}_i$ là biến cố người thứ i bắn trượt bia (i = 1, 2, 3)

$P(A_1) = 0,8;$ $P(A_2) = 0,7;$ $P(A_3) = 0,6$

$P(\overline{A}_1) = 0,2;$ $P(\overline{A}_2) = 0,3;$ $P(\overline{A}_3) = 0,4$

Gọi X là số viên đạn trúng bia. X có thể nhận 4 giá trị là 0,1,2,3

$P_0 = P(X = 0) = P(\overline{A}_1 \overline{A}_2 \overline{A}_3) = P(\overline{A}_1)P(\overline{A}_2)P(\overline{A}_3) = 0,024$

$P_1 = P(X = 1) = P\left(A_1.\overline{A_2}.\overline{A_3} + \overline{A_1}.A_2.\overline{A_3} + \overline{A_1}.\overline{A_2}.A_3\right)$

$= 0,8 . 0,3 . 0,4 + 0,2 . 0,7 . 0,4 + 0,2 . 0,3 . 0,6 = 0,188$

$P_2 = P(X = 2) = P\left(A_1.A_2.\overline{A_3} + A_1.\overline{A_2}.A_3 + \overline{A_1}.A_2.A_3\right)$

$= 0,8 . 0,7 . 0,4 + 0,8 . 0,3 . 0,6 + 0,2 . 0,7 . 0,6 = 0,452$

$P_3 = P(X = 3) = P(A_1 A_2 A_3) = 0,8 . 0,7 . 0,6 = 0,336$

Ta có bảng phân phối xác suất

X	0	1	2	3
P	0,024	0,188	0,452	0,336

b) $E(X) = \sum_{i=0}^{3} x_i p_i = 2,1; Var(X) = \sum_{i=0}^{3} x_i{}^2 p_i - [E(X)]^2 = 5,02 - 4,41 = 0,61$

c) P (2 ≤ X ≤ 7) = P(X=2) + P(X=3) = 0,788

d) Tìm hàm phân phối xác suất của X.

* $x \le 0$: $F(x) = P(X<0) = 0$
* $0 < x \le 1$: $F(x) = P(X<1) = P(X=0) = 0{,}024$
* $1 < x \le 2$: $F(x) = P(X<2) = P(X = 0) + P(X = 1)$ $=0{,}024+0{,}188 = 0{,}212$
* $2 < x \le 3$: $F(x) = P(X<3) = P(X = 0) + P(X = 1) + P(X = 2)$ $= 0{,}664$
* $x > 3$:

$F(x) = P(X < +\infty) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) =1$

$$\text{Vậy } F(x)= \begin{cases} 0 \text{ khi } x \le 0 \\ 0{,}024 \text{ khi } 0 < x \le 1 \\ 0{,}212 \text{ khi } 1 < x \le 2 \\ 0{,}664 \text{ khi } 2<x \le 3 \\ 1 \text{ khi } x > 3 \end{cases}$$

Ví dụ 8 Một hộp đựng 3 sản phẩm trong đó có 1 sản phẩm giả. Người ta lần lượt kiểm tra từng sản phẩm không hoàn lại cho đến khi phát hiện được sản phẩm giả thì thôi (giả thiết là các sản phẩm phải qua kiểm tra mới xác định được là sản phẩm giả hay thật).

Gọi X là số sản phẩm được kiểm tra. Tìm E(X);Var(X).

BÀI GIẢI

X là đại lượng ngẫu nhiên rời rạc có thể nhận các giá trị: 1, 2, 3.

Ta cần tìm các xác suất tương ứng.

Gọi A_i (i=1, 2, 3) là biến cố lần thứ i kiểm tra được sản phẩm giả.

Các biến cố phụ thuộc nhau.

Ta có $P(X = 1) = P(A_1) = \frac{1}{3}$

$$P(X = 2) = P(\overline{A}_1 A_2) = P(\overline{A}_1)\, P(A_2 / \overline{A}_1) = \frac{2}{3}.\frac{1}{2} = \frac{1}{3}$$

Tương tự ta tính được $P(X = 3) = \frac{1}{3}$

Vậy bảng phân phối xác suất của X là

X	1	2	3
P	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

$$E(X)=\sum_{i=1}^{3}x_ip_i=2;\quad Var(X)=\sum_{i=1}^{3}x_i^{\,2}p_i-[E(X)]^2=\frac{2}{3}$$

Ví dụ 9 Hàm mật độ xác suất f(x) của đại lượng ngẫu nhiên X có dạng

$$f\left(x\right)=\begin{cases}\dfrac{1}{2L} & khi\ \ |x-a|\ \le\ L\\ 0\ \ khi\ \ |x-a|\ >\ L\end{cases}$$

Trong đó a, L là hằng số.

a) Xác định kỳ vọng và phương sai của X.

b) Tìm hàm phân phối xác suất F(x).

BÀI GIẢI

a) Xác định kỳ vọng và phương sai của X.

$$E(X)=\int_{-\infty}^{a-L}x.0dx+\int_{a-L}^{a+L}\frac{1}{2L}.xdx+\int_{a+L}^{+\infty}x.0dx=a$$

$$Var(X)=\int_{-\infty}^{+\infty}x^2f(x)dx-[E(X)]^2$$

$$=\int_{-\infty}^{a-L}x^2.0dx+\int_{a-L}^{a+L}\frac{1}{2L}.x^2dx+\int_{a+L}^{+\infty}x.^20dx-a^2=\frac{L^2}{3}$$

b) Tìm hàm phân phối xác suất F(x): $F(x)=\int_{-\infty}^{x}f(x)dx$

$x\le a-L$ thì $F(x)=0$

$a-L<x\le a+L$ thì $F(x)=\int_{-\infty}^{x}f(x)dx=\int_{-\infty}^{a-L}0dx+\int_{a-L}^{x}\frac{1}{2L}dx$

$$= \frac{1}{2L}x - \frac{1}{2L}(a-L)$$

$a + L < x < +\infty$ thì $F(x) = \int_{-\infty}^{a-L} 0dx + \int_{a-L}^{a+L} \frac{1}{2L}dx + \int_{a+L}^{x} 0dx = 1$

Vậy $F(x)= \begin{cases} 0 \text{ khi } x \le a-L \\ \frac{x}{2L} - \frac{1}{2L}(a\text{-}L) \text{ khi } a\text{-}L<x \le a+L \\ 1 \text{ khi } x > a+L \end{cases}$

III. Một số đặc trưng khác

1. Mode

Định nghĩa: mode là giá trị của đại lượng ngẫu nhiên X có khả năng xuất hiện lớn nhất trong một lân cận nào đó của nó.

Kí hiệu: Mod X

a) Đại lượng ngẫu nhiên rời rạc: Mode là giá trị của X ứng với xác suất lớn nhất trong bảng phân phối xác suất của nó.

b) Đại lượng ngẫu nhiên X liên tục: có hàm mật độ f(x) thì Mode là giá trị làm cho hàm mật độ đạt cực đại.

Một đại lượng ngẫu nhiên có thể có nhiều Mode.

2. Median (trung vị)

Định nghĩa: Median là giá trị của đại lượng ngẫu nhiên X.

Kí hiệu Med X, thỏa

$$P\left(X \le m\right) \ge \frac{1}{2} \text{ và } P\left(X \ge m\right) \ge \frac{1}{2}$$

Nếu *X* là đại lượng ngẫu nhiên liên tục thì

$$P(X \le MedX) = P(X > MedX) = \frac{1}{2} \Leftrightarrow F(medX) = \frac{1}{2}$$

Ví dụ 10 Cho đại lượng ngẫu nhiên X có bảng phân phối xác suất

X	1	2	3
P	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$

Thì ModX = 3; MedX=2

Ví dụ 11 Cho đại lượng ngẫu nhiên X có hàm mật độ

$$f(x) = \begin{cases} 0 \;\; khi \;\; x \le 0 \\ \frac{x}{2} e^{-\frac{x^2}{4}} \;\; khi \;\; x > 0 \end{cases}$$

a) Tìm Mod X

b) Tìm MedX

BÀI GIẢI

a) Mod X là nghiệm của phương trình: f ' (x)=0

$$f'(x) = \frac{1}{2} e^{-\frac{x^2}{4}} - \frac{x^2}{4} e^{-\frac{x^2}{4}} \Rightarrow f'(x) = 0 \Leftrightarrow e^{-\frac{x^2}{4}} \left(\frac{1}{2} - \frac{x^2}{4}\right) = 0$$

$$\Leftrightarrow x^2 = 2 \Leftrightarrow x = \sqrt{2} \; (x > 0)$$

$$f''(x) = e^{-\frac{x^2}{4}} \left(\frac{x^3}{8} - \frac{3x}{4}\right) \Rightarrow f''(\sqrt{2}) < 0$$

Vậy f(x) đạt cực đại tại $x = \sqrt{2} \Rightarrow \text{ModX} = \sqrt{2}$

b) MedX là nghiệm của phương trình $F(x) = \frac{1}{2}$

$$\Leftrightarrow \int_0^x \frac{t}{2} e^{-\frac{t^2}{4}} dt = \frac{1}{2} \Leftrightarrow \int_0^x e^{-\left(\frac{t}{2}\right)^2} d\left(\frac{t^2}{4}\right) = \frac{1}{2}$$

$$\Leftrightarrow 1 - e^{-\frac{x^2}{4}} = \frac{1}{2} \Leftrightarrow e^{-\frac{x^2}{4}} = \frac{1}{2} \Leftrightarrow -\frac{x^2}{4} = -\ln 2 \Leftrightarrow x = 2\sqrt{\ln 2}$$

$\Rightarrow$ MedX= $2\sqrt{\ln 2}$

3. Độ lệch chuẩn*:* $\sigma = \sqrt{\text{Var}(X)}$

Khi phân tích đại lượng ngẫu nhiên X người ta thường dùng độ lệch chuẩn vì độ lệch chuẩn có cùng đơn vị đo với đại lượng ngẫu nhiên X, còn phương sai có đơn vị là đo là bình phương đơn vị đo của X.

Ví dụ X đo bằng cm thì độ lệch chuẩn có cùng đơn vị đo là cm, còn phương sai có đơn vị đo là $(cm)^2$.

Ngoài ra còn các đặc trưng bằng số khác như là: moment, hệ số bất đối xứng, hệ số nhọn.

2.3. CÁC QUY LUẬT PHÂN PHỐI XÁC SUẤT ĐẶC BIỆT

Một quy tắc phân phối xác suất của đại lượng ngẫu nhiên được gọi là quy luật phân phối của đại lượng ngẫu nhiên đó.

I. Quy luật siêu bội

1. Định nghĩa

Xét tổng thể có N phần tử, trong đó có M phần tử mang một tính chất T nào đó. Từ tổng thể này, ta lấy ngẫu nhiên đồng thời n phần tử. Gọi X là số phần tử mang tính chất T trong n phần tử lấy ra, thì đại lượng ngẫu nhiên rời rạc X gọi là có phân phối siêu bội

Ký hiệu $X \sim H(N, M, n)$.

Công thức xác suất

$$P(X = k) = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n} \;;\; k = 0, 1, 2, \ldots, n$$

2. ***Các đặc trưng bằng số***

Khi $X \sim H(N, M, n)$, ta có

Trung bình $E(X) = np$, với $p = \frac{M}{N}$

Phương sai $Var(X) = npq\dfrac{N-n}{N-1}$, với $p = \frac{M}{N}$ và $q = 1 - p$.

Ví dụ 1 Trong bình có 10 viên bi trong đó có 6 viên đỏ. Lấy ngẫu nhiên 4 viên. Tính xác suất để trong số đó có 3 viên đỏ.

BÀI GIẢI

X là số bi đỏ trong 4 bi lấy ra. $X \sim H(N = 10, M = 6, n = 4)$

$$P(X = 3) = \frac{C_6^3 C_4^1}{C_{10}^4} = \frac{80}{210} \approx 0,381$$

Ví dụ 2 Thùng A có 20 lọ thuốc gồm 3 lọ hỏng và 17 lọ tốt.

Lấy ra 3 lọ. Gọi X là số lọ hỏng trong 3 lọ lấy ra.

a) Tìm bảng phân phối xác suất của X.

b) Tìm $E(X); Var(X)$

BÀI GIẢI

a) Ta có $X \sim H(N=20; M=3; n=3)$,

$$P(X=k) = \frac{C_3^k C_{17}^{3-k}}{C_{20}^3} \text{ với } k = 0,1,2,3$$

Ta có bảng phân phối xác suất

X	0	1	2	3
P	$\frac{34}{57}$	$\frac{34}{95}$	$\frac{17}{380}$	$\frac{1}{1140}$

b) Trung bình $E(X) = np = 3\frac{3}{20}$=0,45

Phương sai $Var(X) = npq\frac{N-n}{N-1} = 3.\frac{3}{20}.\frac{17}{20}.\frac{20-3}{20-1}$=0,3422

Hoặc dựa vào bảng phân phối xác suất

$$E(X) = \sum_{i=0}^{3} x_i p_i = 0,45; \;\; Var(X) = \sum_{i=0}^{3} x_i^2 p_i - [E(X)]^2 = 0,3422$$

II. Quy luật phân phối nhị thức

1. Định nghĩa

Xét một phép thử, trong phép thử chỉ có thể xảy ra một trong hai trường hợp:

- hoặc biến cố A xảy ra P(A) = p
- hoặc A không xảy ra với $P(\overline{A})$ = 1- p = q .

Lặp lại phép thử này n lần một cách độc lập (độc lập có nghĩa là kết quả của phép thử này không ảnh hưởng đến kết quả của phép thử kia và ngược lại)

Gọi X là số lần xuất hiện biến cố A trong n phép thử đó.

Khi đó, ta nói X là đại lượng ngẫu nhiên có luật phân phối nhị thức.

Kí hiệu $X \sim B(n,p)$.

Công thức tính xác suất

$$P(X=k) = C_n^k p^k q^{n-k} \quad ; k = 0,1,\ldots,n$$

2.Các đặc trưng bằng số

Khi $X \sim B(n;p)$ ta có

Trung bình $EX = np$

Phương sai $DX = np(1-p) \equiv npq$

Độ lệch chuẩn $\sigma_X = \sqrt{npq}$

Mod X : $np\text{-}q \le ModX \le np+p$

Chú ý: Khi ta lấy mẫu có hoàn lại n lần từ một tổng thể có N phần tử, trong đó có M phần tử mang một tính chất T nào đó.

Gọi X là số phần tử mang tính chất T trong n phần tử lấy ra, thì $X \sim B(n;p)$ với $p = \frac{M}{N}$.

Ví dụ 3 Trong một lô thuốc (rất nhiều) với xác suất nhận được thuốc hỏng là $p = 0,1$. Lấy ngẫu nhiên 3 lọ để kiểm tra. Gọi X là số lọ hỏng trong 3 lọ lấy ra để kiểm tra.

a) Lập bảng phân phối xác suất cho X.

b) Tìm kỳ vọng E(X) và phương sai Var(X).

BÀI GIẢI

a) X là số lọ hỏng trong 3 lọ lấy ra để kiểm tra, suy ra X=0,1,2,3.

Ta có $X \sim B(n = 3; p = 0,1)$. Do đó xác suất để

$$P(X = k) = C_3^k (0.1)^k (1-0.1)^{3-k} \text{ với } k = 0,1,2,3$$

Ta có bảng phân phối xác suất

X	0	1	2	3
P	0.729	0.243	0.027	0.001

b) Tìm kỳ vọng E(X) và phương sai Var(X).

Kỳ vọng $E(X) = np = 3.0,1 = 0,3$

Phương sai $Var(X) = np(1-p) \equiv npq = 3.0,1.0,9 = 0,27$

Hoặc dựa vào bảng phân phối xác suất

$$E(X) = \sum_{i=0}^{3} x_i p_i = 0,3 \quad ; \quad Var(X) = \sum_{i=0}^{3} x_i^2 p_i - [E(X)]^2 = 0,27$$

Ví dụ 4 Tung 5 lần một con xúc xắc.
Tính xác suất để có 4 lần mặt 6 chấm xuất hiện

BÀI GIẢI

Gọi X là số lần xuất hiện mặt 6 chấm trong 5 lần tung
A là biến cố xuất hiện mặt 6 chấm trong một lần tung

P(A) = 1/6 = p, q = 1-p = 5/6; Ta có $X \sim B\left(n = 5; p = \frac{1}{6}\right)$.

$$P(X = 4) = C_5^k\left(\frac{1}{6}\right)^k\left(\frac{5}{6}\right)^{5-k} = C_5^4\left(\frac{1}{6}\right)^4\left(\frac{5}{6}\right)^1 \approx 0{,}003$$

III. Quy luật phân phối Poisson

1. Định nghĩa

Gọi X là số lần A xuất hiện trong một khoảng thời gian hoặc trên một vùng một miền nào đó, thì X gọi là đại lượng ngẫu nhiên có phân phối Poisson với tham số $\lambda > 0$ là số trung bình của số lần A xảy ra.

Ký hiệu $X \sim P(\lambda)$

Công thức xác suất

$$P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}\ ;\ \text{với } k = 0, 1, ..., n, ... ; e \approx 2,718...$$

2. Các đặc trưng bằng số

Khi $X \sim P(\lambda)$, ta có

Trung bình $E(X) = \lambda$

Phương sai $D(X) = \lambda$

Độ lệch chuẩn $\sigma_X = \sqrt{\lambda}$;

Mod X: $\lambda - 1 \le ModX \le \lambda$

Trong thực tế: dòng người đi đến siêu thị, nhà ga, bến tàu, tàu xe về bến bãi… đều tuân theo quy luật phân phối Poisson .

Ví dụ 5 Ở một tổng đài điện thoại, các cú điện thoại gọi đến xuất hiện ngẫu nhiên, độc lập với nhau và tốc độ trung bình 2 cuộc gọi trong 1 phút. Tìm xác suất để có đúng 5 cú điện thoại trong 1 phút.

BÀI GIẢI

Gọi X là số cú điện thoại gọi đến một tổng đài điện thoại trong 1 phút thì $X \sim P(\lambda = 2)$. Xác suất để có đúng 5 cú điện thoại trong 1 phút là

$$P(X = 5) = e^{-2}\frac{2^5}{5!}$$

IV. Quy luật phân phối chuẩn

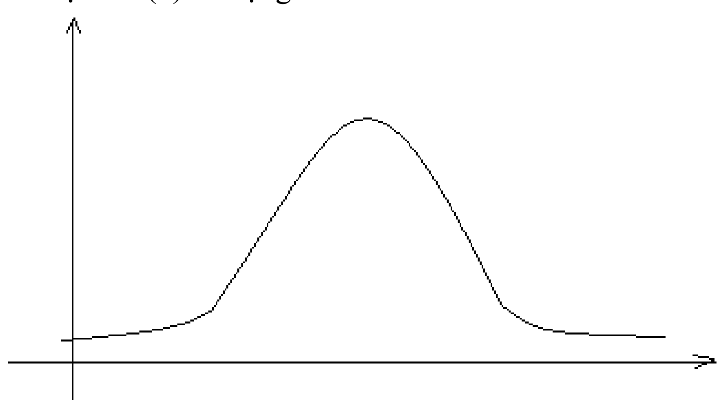
1.Định nghĩa

Đại lượng ngẫu nhiên X liên tục được gọi là có phân phối chuẩn với kỳ vọng μ và phương sai σ^2 nếu X có hàm mật độ f(x) dạng

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

Kí hiệu $X \sim N(\mu;\sigma^2)$

Đồ thị của f(x) có dạng



Trường hợp đặc biệt: $X \sim N(0;1)$. Khi đó ta nói X tuân theo phân phối chuẩn tắc với

Hàm mật độ xác suất $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ (hàm Gauss)

và hàm phân phối xác suất $F(x) = \int\limits_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$

2. Công thức tính xác suất của phân phối chuẩn

Khi $X \sim N\left(\mu; \sigma^2\right)$ thì xác suất

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

Với $\Phi(x) = \int\limits_{0}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$ (hàm tích phân Laplace)

Giá trị của tích phân này được tính sẵn ở bảng tra phần phục lục 1 (Bảng 1)

***Chú ý**:*

* $\Phi(-x) = -\Phi(x)$
* Khi $x > 4$ thì $\Phi(x) \approx \Phi(4) \approx 0,5$
* Khi $X \sim N\left(\mu; \sigma^2\right)$, mà $a = -\infty$ hay $b = +\infty$ thì

$$P\left(-\infty < X \le b\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi(-\infty) = \Phi\left(\frac{b-\mu}{\sigma}\right) + 0.5$$

$$P\left(a \le X < +\infty\right) = \Phi(+\infty) - \Phi\left(\frac{a-\mu}{\sigma}\right) = 0.5 - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

* Hơn nữa, do tính chất của tích phân, ta có

$$P(a \le X \le b) = P(a < X \le b) = P(a \le X < b) = P(a < X < b)$$

* Quy tắc k- xích ma: $P(|X-\mu| < k.\sigma) = 2\Phi(k)$ thật vậy

$$P(|X-\mu| < k.\sigma) = \Phi\left(\frac{k.\sigma+\mu-\mu}{\sigma}\right) - \Phi\left(\frac{-k.\sigma+\mu-\mu}{\sigma}\right) = 2\Phi(k)$$

3. Các đặc trưng bằng số

Khi $X \sim N\left(\mu;\sigma^2\right)$ ta có

Trung bình $EX = \mu$

Phương sai $DX = \sigma^2$

Độ lệch chuẩn $\sqrt{DX} = \sigma$ và Mod X$=\mu$

Ví dụ 6 Đường kính của một chi tiết máy do một máy tiện tự động sản xuất có phân phối chuẩn với trung bình $\mu = 50$ mm và độ lệch chuẩn $\sigma = 0,05$ mm. Chi tiết máy được xem là đạt yêu cầu nếu đường kính không sai quá 0,1mm so với đường kính thiết kế.

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.

BÀI GIẢI

Gọi X là đường kính của chi tiết máy thì $X \sim N(\mu;\sigma^2)$, với $\mu = 50$ mm và $\sigma = 0,05$ mm.

a) Xét biến cố A: "nhận được sản phẩm đạt yêu cầu"

$$P(A) = P\left(\left|X - 50\right| \le 0,1\right) = P\left(50 - 0,1 \le X \le 50 + 0,1\right)$$

$$P\left(49,9 \le X \le 50,1\right) = \Phi\left(\frac{50,1 - 50}{0,05}\right) - \Phi\left(\frac{49,9 - 50}{0,05}\right)$$

$$= \Phi\left(2\right) - \Phi\left(-2\right) = 2\Phi\left(2\right) = 2.0,4772 = 0,9544$$

Vậy xác suất để nhận được sản phẩm đạt yêu cầu là 95,44%.

b) Gọi X là số sản phẩm đạt yêu cầu trong 3 sản phẩm lấy ra thì $X \sim B(3;0,9544)$.

Suy ra xác suất để có ít nhất một sản phẩm đạt yêu cầu là

$$P\left(X \ge 1\right) = 1 - P\left(X = 0\right) = 1 - C_3^0\left(0,9544\right)^0\left(0,0456\right)^3$$

$$=0,9999$$

Chú ý: Sự liên hệ giữa các quy luật phân phối nhị thức, siêu bội, Poisson và phân phối chuẩn

i. Khi $X \sim H(N, M, n)$ mà n quá nhỏ so với N thì ta xấp xỉ $X \sim B(n,p)$ với $p = \frac{M}{N}$.

Ví dụ 7 Một lô bóng đèn có 10.000 bóng, trong số đó có 4000 bóng 110V. Lấy ngẫu nhiên 10 bóng. Tính xác suất có

a) 3 bóng 110V.

b) Ít nhất 1 bóng 110V.

BÀI GIẢI

Gọi X là số bóng 110V trong 10 bóng lấy ra.

Ta có: $X \sim H(N = 10000, M = 4000; n = 10)$.

Vì N khá lớn và n nhỏ nên có thể xem $X \sim B(10; 0{,}4)$.

a) $P(X = 3) = \dfrac{C_{4000}^3 C_{6000}^7}{C_{10000}^{10}} \approx C_{10}^3 (0{,}4)^3 (0{,}6)^7 = 0{,}21499$

b) $P(X \ge 1) = 1 - P(X = 0) = 1 - \dfrac{C_{6000}^{10}}{C_{10000}^{10}} \approx 1 - C_{10}^0 (0{,}4)^0 (0{,}6)^{10} = 0{,}9939$

ii. Khi $X \sim B(n, p)$ mà n lớn và p nhỏ và $np \le 5$ ta có thể xấp xỉ phân phối nhị thức $B(n, p)$ bằng phân phối Poisson $X \sim P(\lambda = np)$

Ví dụ 8 Một thiết bị có 1000 linh kiện, xác suất hỏng của mỗi linh kiện là 0,001. Tính xác suất để thiết bị đó có 10 linh kiện bị hỏng.

BÀI GIẢI

Phép thử là kiểm tra 1 linh kiện**;** n = 1000, k = 10

Gọi A là biến cố linh kiện hỏng trong 1 phép thử, P(A) = p = 0,001

X là số linh kiện bị hỏng. Vậy $X \sim B(n=1000; p= 0{,}001)$

$\Rightarrow P(X = 10) = C_{1000}^{10} 0.001^{10} 0.999^{990}$

Vì n lớn và p nhỏ và $np = \lambda = 1000.0{,}001=1$ nên ta xấp xỉ

$X \sim P(\lambda) \Rightarrow \quad P(X = 10) \approx \dfrac{e^{-1}1^{10}}{10!} = \dfrac{1}{e.10!} \approx 0{,}1014.10^{-6}$

iii. Khi $X \sim B(n,p)$ mà n lớn và p nhỏ và tích np > 5 ta có thể xấp xỉ phân phối nhị thức $B(n;p)$ bằng phân phối chuẩn $N(\mu;\sigma^2)$ với $\mu = np$, $\sigma^2 = npq$. Khi đó xác suất

$$P(X=k) = \frac{1}{\sigma}\varphi(u) \quad \text{Trong đó } u = \frac{k-\mu}{\sigma};$$

$$\varphi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$$ Tra bảng 2 (phần phụ lục 1)

Ví dụ 9 Một nhà máy sản xuất ra sản phẩm loại A, với xác suất là 0,8. Hãy tính xác suất để trong 400 sản phẩm do nhà máy sản xuất ra có

a) 336 sản phẩm loại A

b) Từ 304 đến 328 sản phẩm loại A

BÀI GIẢI

Gọi X là số sản phẩm loại A trong 400 sản phẩm sản xuất ra $X \sim B(n=400; p=0,8)$ mà n lớn và p nhỏ và np =320 > 5 ta xấp xỉ bằng phân phối chuẩn $N(\mu;\sigma^2)$ với $\mu = np = 400.0,8 = 320$ và $\sigma^2 = npq = 400.0,8.0,2 = 64$.

a) Tính xác suất để có 336 sản phẩm loại A.

Áp dụng công thức

$$P(X=k) = \frac{1}{\sigma}\varphi(u) \quad \text{Trong đó } u = \frac{k-\mu}{\sigma};$$

$$P(X=336) = \frac{1}{8}\varphi(u); u = \frac{336-320}{8} = 2;$$

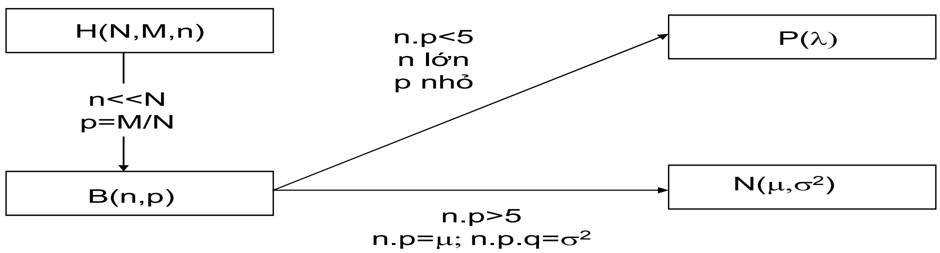
Tra bảng 2 (phần phụ lục 1) thì $\varphi(2) = 0,054$

$\Rightarrow P(X=336) = 0,00675$

b) $$P(304 \le X \le 328) = \Phi\left(\frac{328-320}{8}\right) - \Phi\left(\frac{304-320}{8}\right)$$
$$= \Phi(1) - \Phi(-2) = 0,34 - (-0,4772) = 0,8185.$$

iv. Nếu $X \sim N(\mu;\sigma^2)$ thì $\alpha X + \beta \sim N(\alpha\mu + \beta;(\alpha\sigma)^2)$ với α, β là hằng số

Sự liên hệ giữa các quy luật cho bởi hình vẽ



Ngoài ra, ta còn có một số phân phối dùng trong thống kê như sau

V. Quy luật phân phối χ^2 (chi- bình phương)

Định nghĩa

Cho các đại lượng ngẫu nhiên $X_1, X_2, \ldots, X_n$ độc lập, có cùng luật phân phối chuẩn N(0,1).

Đặt $X = X_1^2 + X_2^2 + \ldots + X_n^2$.

Khi đó ta nói đại lượng ngẫu nhiên X có luật phân phối chi bình phương với n bậc tự do.

Kí hiệu: $X \sim \chi^2_n$

VI. Quy luật phân phối Student

Định nghĩa

Cho 2 đại lượng ngẫu nhiên $Y \sim N(0,1)$ và $Z \sim \chi^2_n$ độc lập với nhau. Đặt $X = \dfrac{Y\sqrt{n}}{\sqrt{Z}}$.

Khi đó ta nói X có luật phân phối Student với n bậc tự do.

Kí hiệu: $T \sim St(n)$.

Chú ý

Phân phối Student với bậc tự do lớn, $n \geq 30$, được xấp xỉ bằng phân phối Gauss, nghĩa là:

Nếu $X \sim St(n)$, với $n \geq 30$, thì $X \sim N(0;1)$.

VII. Phân phối Fisher

Định nghĩa

Xuất phát từ hai đại lượng ngẫu nhiên độc lập có phân phối chi-bình phương, người ta xây dựng được phân phối Fisher.

Cụ thể, với hai đại lượng ngẫu nhiên độc lập X, Y trong đó $X \sim \chi^2(n)$ và $Y \sim \chi^2(m)$.

Ta đặt $F = \dfrac{X/n}{Y/m}$

Khi đó ta nói F có luật phân phối Fisher.
Kí hiệu: $F \sim F(n,m)$.

Chú ý: Bảng giá trị của một số xác suất liên quan đến đại lượng ngẫu nhiên liên tục có phân phối Gauss $N(0;1)$, Chi-Bình phương $\chi^2(n)$, Student $St(n)$ và Fisher $F(n,m)$ được lập thành bảng để tiện dùng ở phần phụ lục **1**

BÀI TẬP MẪU CHƯƠNG II

1. Cho đại lượng ngẫu nhiên X có luật phân phối như sau

X	0	1	4	6
P	$\frac{1}{8}$	$\frac{4}{8}$	$\frac{1}{8}$	$\frac{2}{8}$

a) Tìm kỳ vọng và phương sai của Y=5X + Var(X).

b) Tìm xác suất P(1 ≤ X ≤ 3) = ?

BÀI GIẢI

a) $E(X) = \sum_{i=0}^{3} x_i p_i = \frac{5}{2};$

$$Var(X) = \sum_{i=0}^{3} x_i^2 p_i - [E(X)]^2 = \frac{21}{4}$$

Dùng tính chất của kỳ vọng và phương sai:

$E(Y) = E[5X + Var(X)] = 5E(X) + Var(X) = 5.\frac{5}{2} + \frac{21}{4} = \frac{71}{4}$

$Var(Y) = Var[5X + Var(X)] = 25Var(X) + 0 = 5^2.\frac{21}{4} = \frac{525}{4}$

b) $P(1 \le X \le 3) = P(X = 1) = \frac{4}{8}$

2. Gieo đồng thời 2 con xúc xắc. Gọi X là tổng số chấm xuất hiện

a) Lập bảng phân phối xác suất cho X.

b) Tìm kỳ vọng E(X) và phương sai Var(X).

BÀI GIẢI

Gọi X_1 là số chấm xuất hiện của con xúc xắc thứ 1.

X_2 là số chấm xuất hiện của con xúc xắc thứ 2.

X là tổng số chấm xuất hiện khi gieo đồng thời 2 con xúc xắc

Ta có $X = X_1 + X_2$

Ta dùng bảng tính sau đây

X_1 / X_2	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

a) Bảng phân phối xác suất

X	2	3	4	5	6	7	8	9	10	11	12
P	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

b) Tìm kỳ vọng E(X) và phương sai Var(X).

Cách 1: Từ bảng phân phối xác suất của X theo định nghĩa kỳ vọng và phương sai ta có

$$E(X)=\sum_{i=2}^{12} x_i p_i = 7\ ;$$

$$\mathrm{Var(X)}=\sum_{i=2}^{12} x_i^2 p_i - \left[E(X)\right]^2 = \frac{1974}{36}-7^2=\frac{210}{36}$$

***Cách 2**:* Sử dụng tính chất của kỳ vọng và phương sai.

Bảng phân phối xác suất của X_1

X_1	1	2	3	4	5	6
P	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Bảng phân phối xác suất của X_2

X_2	1	2	3	4	5	6
P	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$\Rightarrow$ $E(X) = E(X_1 + X_2) = E(X_1) + E(X_2) = \frac{21}{6} + \frac{21}{6} = 7$

$Var(X) = Var(X_1) + Var(X_2) = \frac{210}{36}$

3. Một công ty cần trang bị một số lượng lớn máy cho khu vực sản xuất mới. Có hai loại máy được xem xét là máy do công ty A sản xuất và máy do công ty B sản xuất với số liệu thống kê như sau

Mức độ hỏng		1	2	3
Máy của công ty A	Tỷ lệ hỏng (%)	4	4	2
	Chi phí sửa chữa (triệu đồng/năm)	7,5	11	15,5
Máy của công ty B	Tỷ lệ hỏng (%)	2	5	3
	Chi phí sửa chữa (triệu đồng/năm)	6,5	10,5	14

Giả sử các yếu tố khác không có sự khác biệt đáng kể và công ty này chỉ quan tâm đến chi phí sản xuất hàng năm. Hỏi nên chọn mua máy do công ty nào sản xuất?

BÀI GIẢI

Gọi X là chi phí sửa chữa của một máy của công ty A (triệu đồng/năm). Ta xem X là đại lượng ngẫu nhiên có bảng phân phối xác suất như sau

X	0	7,5	11	15,5
P	0,9	0,04	0,04	0,02

Khi đó $EX = 0.0,9 + 7,5.0,04 + 11.0,04 + 15,5.0,02 = 1,05$

(triệu đồng/năm)

Gọi Y là chi phí sửa chữa của một máy của công ty B (triệu đồng/năm). Ta xem Y là đại lượng ngẫu nhiên có bảng phân phối xác suất như sau

Y	0	6,5	10,5	14
P	0,9	0,02	0,05	0,03

Khi đó $EY = 0.0,9 + 6,5.0,02 + 10,5.0,05 + 14.0,03 = 1,075$

(triệu đồng/năm).

Vì $EX < EY$ nên ta chọn mua máy của công ty A.

4. Một công ty kinh doanh thực phẩm có 3 cửa hàng. Gọi X_1, X_2, X_3 lần lượt là doanh thu trong một ngày (đơn vị tính: triệu đồng) của cửa hàng 1, 2, 3. Khi đó, X_1, X_2, X_3 là các đại lượng ngẫu nhiên có bảng phân phối xác suất như sau

X_1	15	19,5	29
P	$\frac{2}{7}$	$\frac{3}{7}$	$\frac{2}{7}$

X_2	7	17,5	21,5
P	$\frac{1}{7}$	$\frac{4}{7}$	$\frac{2}{7}$

X_3	10,5	15	26,5
P	$\frac{3}{7}$	$\frac{3}{7}$	$\frac{1}{7}$

Tính doanh thu trung bình trong một ngày của công ty.

BÀI GIẢI

Ta có

$$EX_1 = 15.\frac{2}{7} + 19,5.\frac{3}{7} + 29.\frac{2}{7} = \frac{146,5}{7}$$

$$EX_2 = 7.\frac{1}{7} + 17,5.\frac{4}{7} + 21,5.\frac{2}{7} = \frac{120}{7}$$

$$EX_3 = 10,5.\frac{3}{7} + 15.\frac{3}{7} + 26,5.\frac{1}{7} = \frac{103}{7}$$

Gọi X là doanh thu (triệu đồng) của công ty này trong một ngày.

$$X = X_1 + X_2 + X_3$$

$$EX = E\left(X_1 + X_2 + X_3\right)$$

$$= EX_1 + EX_2 + EX_3 = \frac{369,5}{7} = 52,786 \text{ (triệu đồng)}$$

5. Một nhà đầu tư có 3 dự án. Gọi X_i ($i = 1,2,3$) là lợi nhuận khi thực hiện dự án thứ *i*, còn giá trị âm chỉ số tiền bị thua lỗ. Qua nghiên cứu và bằng kinh nghiệm, nhà đầu tư có ước lượng như sau

X_1	-2	-1	10
P	0,4	0,2	0,4

X_2	-1	4	5
P	0,3	0,2	0,5

X_1	-3	-2,5	8
P	0,3	0,2	0,5

Đơn vị tính: tỷ đồng.
Nếu chọn một trong 3 dự án trên, theo bạn nên chọn dự án nào?
BÀI GIẢI.
Ta tính được: $EX_1 = 3$; $EX_2 = 3$; $EX_3 = 2,6$

$$Var(X_1) = 32,8\,;\ Var(X_2) = 7\,;\ Var(X_3) = 29,19$$

(Chú ý: Var(X) và EX không cùng đơn vị)
Nếu chọn một trong 3 dự án trên, ta nên chọn dự án nào có lợi nhuận kỳ vọng cao nhất và dự án nào có mức độ rủi ro nhỏ nhất.
Như vậy ta nên chọn dự án thứ 2.

6. Ở một bàn hỏi thi vấn đáp có 50 phiếu hỏi thi. Mỗi phiếu có 4 cách trả lời (trong đó có một cách trả lời đúng). Sinh viên A bốc ngẫu nhiên 6 phiếu thi và mỗi phiếu chọn ngẫu nhiên 1 trong 4 cách trả lời.

Tìm xác suất để sinh viên A trả lời đúng ít nhất 4 trong 10 phiếu hỏi thi đã bốc được?

BÀI GIẢI

Ta nhận thấy rằng 50 phiếu hỏi thi không ảnh hưởng gì đến việc bốc 6 phiếu thi của sinh viên A.Gọi X là số câu trả lời đúng trong 6 phiếu hỏi thi.

Ta coi việc trả lời từng câu hỏi của sinh viên A là 1 phép thử thì các phép thử này độc lập với nhau. Mỗi phép thử chỉ xảy ra trường hợp:Hoặc là sinh viên A trả lời đúng hoặc là sinh viên A trả lời sai.

p = 1/4 là xác suất để sinh viên A trả lời đúng trong từng phép thử.

q = 1 - p = 3/4 là xác suất để sinh viên A trả lời không đúng trong từng phép thử

Vậy X ~ B(n,p). Ta có: P(X = k) = $C_n^k p^k q^{n-k}$, k = 0,1,…,n

Ta có: P(X ≥ 4) = $P_6(4) + P_6(5) + P_6(6)$

$$= C_6^4.p^4.q^2 + C_6^5.p^5.q + C_6^6.p^6 = \frac{73}{729}$$

7. Mỗi chuyến xe người ta chở được 1250 chai dược phẩm. Xác suất để một chai bị vỡ khi vận chuyển là 0,0035. Tính xác suất để có ít nhất 2 chai bị vỡ khi vận chuyển.

BÀI GIẢI

Gọi X là số chai dược phẩm bị vỡ khi vận chuyển.

Ta có $X \sim B(1250;0,0035)$

Vì $n = 1250$ khá lớn, $p = 0,0035$ kh nhỏ, $np = 4,375$

Ta có thể xấp xỉ $X \sim P(4,375)$

Xác suất cần tìm là:

$$P(X \geq 2) = 1 - P(X < 2) = 1 - P(X = 0) - P(X = 1)$$

$$= 1 - \frac{e^{-4,375}.4,375^0}{0!} - \frac{e^{-4,375}.4,375^1}{1!} = 1 - 5,375e^{-4,375}$$

8. Một viên đạn súng trường bắn trúng máy bay với xác suất 0,001. Có 5000 khẩu bắn lên một lượt. Người ta biết rằng máy bay chắc chắn bị hạ nếu có ít nhất 2 viên trúng. Nếu có 1 viên trúng thì xác suất bị hạ chỉ là 80%. Tính xác suất để máy bay bị hạ.

BÀI GIẢI

Phép thử là bắn 1 viên đạn, ta có n = 5000 phép thử độc lập.

Đặt A là biến cố viên đạn trúng máy bay. P(A) = p = 0,001.

Gọi X là tổng số viên đạn trúng thì X~ B(5000 ; 0,001).

Do n lớn và p nhỏ và np=5 nên có xấp xỉ Poisson tức X~ P(5).

Đặt A_0 là biến cố có 0 viên trúng: $P(A_o) = e^{-5}$

A_1 là biến cố có 1 viên trúng :$P(A_1) = 5e^{-5}$

A_2 là biến cố có ít nhất 2 viên trúng :

$$P(A_2) = 1 - P(A_0) - P(A_1) = 1 - 6e^{-5}$$

$\{A_o , A_1 , A_2\}$ tạo thành hệ đầy đủ và xung khắc từng đôi.

Đặt B là biến cố máy bay bị hạ.

Theo công thức xác suất toàn phần:

$$\begin{aligned} P(B) &= P(A_o) P(B / A_o) + P(A_1) P(B / A_1) + P(A_2) P(B / A_2) \\ &= e^{-5} . 0 + 5e^{-5} . 0,8 + (1 - 6e^{-5}).1 \\ &= 1 - 2e^{-5} = 1 - 2 . 0,0067 = 0,9856 \end{aligned}$$

9. Trọng lượng của một loại sản phẩm được biết là có phân phối chuẩn với kỳ vọng là 500g và phương sai là 4.

a) Tính xác suất để sản phẩm có trọng lượng nằm trong khoảng là 494g đến 506g.

b) Sản phẩm được chia làm 3 loại: loại A nếu trọng lượng lớn hơn 504g, loại B nếu trọng lượng từ 498g đến 504g và loại C là loại còn lại. Tính tỷ lệ 3 loại A, B, C

c) Tính xác suất để sản phẩm có trọng lượng là 494g

BÀI GIẢI

a) $P\left(494 \le X \le 506\right) = \Phi\left(\dfrac{506-500}{2}\right) - \Phi\left(\dfrac{494-500}{2}\right)$

$$= \Phi(3) - \Phi(-3) = 2\Phi(3) = 2.0,4987 = 0,9974$$

b) Tỷ lệ loại C là:

$$P(C) = P(0 \le X \le 498) = \Phi\left(\frac{498-500}{2}\right) - \Phi\left(\frac{0-500}{2}\right)$$

$$= \Phi(-1) - \Phi(-250) = \Phi(250) - \Phi(1) = 0,5 - 0,3413 = 0,1587$$

Tỷ lệ loại B là:

$$P(B) = P(498 \le X \le 504) = \Phi\left(\frac{504-500}{2}\right) - \Phi\left(\frac{498-500}{2}\right)$$

$$= \Phi(2) - \Phi(-1) = \Phi(1) + \Phi(2) = 0,3413 + 0,4772 = 0,8185$$

Tỷ lệ loại A là: P(A) = 1 - (0,1587 + 0,8185) = 2,28%

c) $P(X = k) = \dfrac{1}{\sigma}\varphi(u); \quad u = \dfrac{k-\mu}{\sigma}$;

$$P(X = 494) = \frac{1}{2}\varphi(u); u = \frac{494-500}{2} = -3;$$

Tra bảng 2 (phần phụ lục 1) ta có $\varphi(-3) = \varphi(3) = 0,0044$ nên

$$P(X = 494) = \frac{0,0044}{2} = 0,0022$$

10. Lợi nhuận của một nhà đầu tư là đại lượng có phân phối chuẩn với trung bình $\mu = 550$ và độ lệch chuẩn $\sigma = 15$ (đơn vị tính: triệu đồng). Tính xác suất để lợi nhuận của nhà đầu tư này ở trong khoảng $(530;585)$

BÀI GIẢI

Gọi X là lợi nhuận của nhà đầu tư này, ta có $X \sim N(550;15^2)$.

Xác suất cần tìm là

$$P(530 < X < 585) = \Phi\left(\frac{585-550}{15}\right) - \Phi\left(\frac{530-550}{15}\right)$$

$$= \Phi(2,33) - \Phi(-1,33) = \Phi(2,33) + \Phi(1,33)$$

$$= 0,4901 + 0,4082 = 0,8983$$

11. Tuổi thọ của một loại sản phẩm là đại lượng ngẫu nhiên có phân phối chuẩn với $\mu = 2000$ giờ và $\sigma = 50$ giờ. Thời gian bảo hành cho mỗi sản phẩm là 1910 giờ.

a) Tính tỷ lệ sản phẩm phải bảo hành.

b) Muốn tỷ lệ sản phẩm phải bảo hành là 2,28% thì quy định thời gian bảo hành cho mỗi sản phẩm là bao nhiêu?

BÀI GIẢI

a) Đầu tiên ta tính xác suất để một sản phẩm được chọn ngẫu nhiên là sản phẩm phải bảo hành.

Ta có $X \sim N\left(2000, 50^2\right)$

$$P\left(0 \le X < 1910\right) = \Phi\left(\frac{1910-2000}{50}\right) - \Phi\left(\frac{0-2000}{50}\right)$$

$$= \Phi(-1,8) - \Phi(-40) = -\Phi(1,8) + \Phi(40)$$

$$= 0,4641 + 0,5 = 0,0359 = 3,59\%$$

b) Gọi t là thời gian bảo hành. Ta phải chọn t sao cho

$$P\left(0 \le X < t\right) = 0,0228$$

$$\Leftrightarrow \Phi\left(\frac{t-2000}{50}\right) - \Phi\left(\frac{0-2000}{50}\right) = 0,0228$$

$$\Leftrightarrow -\Phi\left(\frac{2000-t}{50}\right) = \Phi(-40) + 0,0228 = -\Phi(40) + 0,0228$$

$$\Leftrightarrow \Phi\left(\frac{2000-t}{50}\right) = 0,5 - 0,0228 = 0,4772$$

Tra bảng giá trị hàm Laplace, (bảng 1) ta có: $\dfrac{2000-t}{50} = 2$

Vậy thời gian bảo hành là $t = 1900$ giờ.

BÀI TẬP CHƯƠNG II

2.1. Có 3 kiện hàng. Kiện 1 chứa 10 sản phẩm tốt và 2 phế phẩm. Kiện thứ 2 chứa 9 sản phẩm tốt và 3 phế phẩm. Kiện thứ 3 chứa 7 sản phẩm tốt và 5 phế phẩm. Các sản phẩm giống hệt nhau về kích thước và trọng lượng. Lấy ngẫu nhiên mỗi kiện hàng 1 sản phẩm.Gọi X là số phế phẩm trong 3 sản phẩm lấy ra.

a) Lập bảng phân phối xác suất cho X

b) Tính kỳ vọng của X

c) Tính phương sai của X

2.2. Có 3 kiện hàng. Kiện 1 chứa 10 sản phẩm tốt và 2 phế phẩm. Kiện thứ 2 chứa 9 sản phẩm tốt và 3 phế phẩm. Kiện thứ 3 chứa 7 sản phẩm tốt và 5 phế phẩm. Các sản phẩm giống hệt nhau về kích thước và trọng lượng. Lấy ra ngẫu nhiên một kiện hàng. Từ đó lấy ra 3 sản phẩm.
Gọi Y là số sản phẩm tốt trong 3 sản phẩm lấy ra.

a) Lập bảng phân phối xác suất cho Y

b) Tính kỳ vọng E(Y)

c) Tính phương sai Var(Y)

2.3. Một kiện hàng có 6 sản phẩm loại I và 4 sản phẩm loại II.Tiền lời khi bán 1 sản phẩm loại I là 5.000đ, loại II là 3.000đ. Lấy ngẫu nhiên 3 sản phẩm để bán. Gọi X là số sản phẩm loại I trong 3 sản phẩm lấy ra để bán.

a) Lập bảng phân phối xác suất của X

b) Tính số tiền lời trung bình khi bán 3 sản phẩm trên

2.4. Có 2 chiếc bình, mỗi bình chứa 10 bi, trong đó có 4 bi đỏ. Lấy ngẫu nhiên từ bình I ra 2 bi bỏ sang bình II. Sau đó lấy ngẫu nhiên từ bình II ra 3 bi .

Gọi X là số bi đỏ lấy được từ bình II.

a) Lập bảng phân phối xác suất của X

b) Tính kỳ vọng E(X)

c) Tính phương sai Var(X)

HƯỚNG DẪN:

Gọi A_i là biến cố 2 bi lấy từ bình 1 có i bi đỏ (i=0,1,2)

$$P(A_i) = \frac{C_4^i C_6^{2-i}}{C_{10}^2} \qquad (i = \overline{0,2})$$

$$P(X = k) = \sum_{i=0}^{2} P(A_i)P(X = k / A_i) = \sum_{i=0}^{2} \frac{C_4^i C_6^{2-i}}{C_{10}^2}.\frac{C_{4+i}^k C_{8-i}^{3-k}}{C_{12}^3}; (k = \overline{0,3})$$

2.5. Người quản lý của một công ty xem xét tình trạng hoạt động của các máy trong công ty này. Dựa vào các thông tin trong quá khứ và bằng kinh nghiệm, người ta có bảng phân phối xác suất của số máy bị hỏng trong tuần như sau:

Số máy bị hỏng trong 1 tuần	0	1	2	3	4
P	0,1	0,24	0,42	0,16	0,08

a) Tính số máy bị hỏng trung bình trong một tuần và độ lệch chuẩn của số máy bị hỏng trong một tuần.

b) Chi phí thiệt hại do mỗi máy bị hỏng trong một tuần là 3 triệu đồng. Tính chi phí thiệt hại cho các máy bị hỏng trung bình của công ty này trong một tuần.

2.6. Một người có một lô hàng và đang cân nhắc xem bán như thế nào

- Phương án 1: bán ngay với giá 20 triệu đồng.

- Phương án 2: chờ đến mùa mua sắm đặc biệt, khi đó có thể bán được với giá cao hơn, tuy nhiên cũng có thể giảm giá, hoặc chỉ bán được một phần. Sau khi ước lượng các yếu tố cần quan tâm (kể cả yếu tố lãi suất theo thời gian), người này có đánh giá như sau: X (đơn vị tính: triệu đồng) là số tiền bán được lô hàng vào mùa mua sắm.

X	10	15	30
P	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

a) Tính EX.

b) Người này nên bán ngay hay chờ đến mùa mua sắm để bán.

2.7. Hàm mật độ xác suất f(x) của đại lượng ngẫu nhiên X có dạng:

$$f(x) = \begin{cases} a\cos x & khi \quad |x| \leq \frac{\pi}{2} \\ 0 & khi \quad |x| > \frac{\pi}{2} \end{cases}$$

a) Tìm hằng số a
b) Tính kỳ vọng E(X)
c) Tính phương sai Var (X)
d) Tìm hàm phân phối xác suất F(x)

2.8. Gọi X là số lần mặt nhất xuất hiện sau 3 lần tung một con xúc xắc.

a) Lập bảng phân phối xác suất của X.

b) Tính xác suất có ít nhất 1 lần được mặt nhất.

c) Tính xác suất có tối đa 2 lần được mặt nhất.

d) Tính E(X), Var(X).

2.9. Một gia đình có 10 người con. Giả sử xác suất sinh con trai, con gái như nhau. Tính xác suất

a) Có 5 người con trai

b) Số con trai ít hơn số con gái

2.10. Một người nuôi 100 con gà đẻ. Xác suất để một con gà đẻ trứng trong ngày là 60%.

a) Tính số trứng trung bình có được trong một ngày

b) Nếu muốn trung bình mỗi ngày thu được 120 trứng thì phải nuôi bao nhiêu con gà

2.11. Ở một tổng đài điện thoại, các cú điện thoại gọi đến xuất hiện ngẫu nhiên, độc lập với nhau và tốc độ trung bình 2 cuộc gọi trong 1 phút. Tìm xác suất để:

a) Có đúng 5 cú điện thoại trong 2 phút

b) Không có cú nào trong khoảng thời gian 30 giây

c) Có ít nhất một cú trong khoảng thời gian 10 giây

2.12. Một loại chi tiết máy được coi là đạt tiêu chuẩn nếu đường kính của nó sai lệch so với đường kính thiết kế không quá 0,33 mm về giá trị tuyệt đối. Cho biết đường kính của loại chi tiết máy là đại lượng ngẫu nhiên phân phối theo qui luật chuẩn của độ lệch tiêu chuẩn là 0,3 mm.

a) Tính xác suất sản xuất ra được một chi tiết đạt tiêu chuẩn

b) Tính số chi tiết máy đạt tiêu chuẩn trung bình khi sản xuất 100 chi tiết

HƯỚNG DẪN

a) Gọi X là đường kính của chi tiết máy. Theo giả thiết X là đại lượng ngẫu nhiên phân phối theo quy luật chuẩn có $\sigma = 0,3$.

Vậy $p = P\left(\left|X - \mu\right| < 0,33\right) = ... = 0,73$

b) Gọi Y là số chi tiết đạt tiêu chuẩn có trong 100 chi tiết sản xuất ra thì $Y \sim B(100\ ;\ p)$ trong đó p là xác suất để sản xuất được chi tiết đạt tiêu chuẩn.

Số chi tiết đạt tiêu chuẩn trung bình khi sản xuất 100 chi tiết chính là E(Y).
Vì $Y \sim B(100\ ;\ 0,73)$ nên $E(Y) = 100 \,.\, 0,73 = 73$

2.13.

a) Cho $X \sim B(5; 0,8)$, tính $P(X=2)$.

b) Cho $X \sim P(2)$, tính $P(X < 2)$.

c) Cho $X \sim H(1000; 600; 5)$, tính $P(X < 1)$.

d) Cho $X \sim N(5; 0,16)$, tính $P(X > 4,6)$.

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

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

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

2.15. Một nhà máy sản xuất ra sản phẩm được đóng thành các kiện hàng. Giả sử khối lượng của một kiện hàng có phân phối chuẩn với khối lượng trung bình là 1000 gam và độ lệch chuẩn về khối lượng là 30 gam. Một người lấy ra một kiện hàng từ trong lô hàng của nhà máy.

a) Tính xác suất để người này lấy được kiện hàng có khối lượng lớn hơn 1030 gam.

b) Kiện hàng được gọi là đạt tiêu chuẩn nếu nó có khối lượng trong khoảng [991 gam; 1015 gam]. Tính xác suất để người này lấy được kiện hàng đạt chuẩn.

c) Nếu lấy được kiện hàng đạt tiêu chuẩn thì sẽ mua lô hàng đó. Kiểm tra 10 lô hàng. Tính xác suất để người này mua được 4 lô hàng.

PHẦN II THỐNG KÊ TOÁN

CHƯƠNG III

MẪU NGẪU NHIÊN

3.1 TỔNG THỂ VÀ MẪU

I. Tổng thể

Khi ta nghiên cứu các vấn đề kinh tế – xã hội, cũng như các vấn đề thuộc các lĩnh vực khác, người ta phải khảo sát một hoặc một số dấu hiệu nào đó. Các dấu hiệu này thể hiện trên nhiều phần tử. Tập hợp tất cả các phần tử đó theo mục đích và phạm vi vấn đề đang nghiên cứu được gọi là tập hợp chính hay tổng thể hoặc một đám đông.

Ví dụ: Tổng thể là tập hợp các sinh viên của một lớp, dấu hiệu ta khảo sát là điểm thi môn Xác suất thống kê.

Giả sử tổng thể có N phần tử. Gọi X^* là dấu hiệu mà ta khảo sát.

x_i ($i=\overline{1,k}$) là giá trị của X^* đo được trên phần tử của tổng thể.

n_i ($i=\overline{1,k}$) là tần số của x_i ; p_i ($i=\overline{1,k}$) là tần suất của x_i.

Ta có thể lập bảng cơ cấu của tổng thể theo dấu hiệu X^*

Giá trị của X^*	x_1	x_2	…	x_i	…	x_k
Tần suất (p_i)	p_1	p_2	…	p_i	…	p_k

Để phân tích dấu hiệu X^* người ta tóm tắt bảng trên bằng các số đặc trưng sau đây

1. Trung bình của tổng thể ký hiệu là m: $m=\sum_{i=1}^{k} p_i x_i$

2. Phương sai của tổng thể ký hiệu là σ^2 : $\sigma^2=\sum_{i=1}^{k}(x_i-m)^2 p_i$

3. Độ lệch tiêu chuẩn của tổng thể ký hiệu là $\sigma = \sqrt{\sigma^2}$

4. Tỷ lệ tổng thể ký hiệu là p

Giả sử tổng thể có N phần tử, trong đó có M phần tử có tính chất A. Gọi $p = \dfrac{M}{N}$ là tỷ lệ các phần tử có tính chất A của tổng thể.

Ví dụ Một công ty chỉ có 40 công nhân, khảo sát dấu hiệu X là năng suất lao động (số sản phẩm/ đơn vị thời gian) ta được bảng số liệu như sau

Năng suất	50	55	60	65	70	75
Số công nhân	3	5	10	12	7	3

Nếu ta gọi những người có năng suất lớn hơn hoặc bằng 65 là những người có năng suất cao thì tỷ lệ những người có năng suất cao là $p = \dfrac{22}{40} = 55\%$

II. Mẫu

Giả sử tổng thể có N phần tử. Khi nghiên cứu vì một số lý do, ta không thể nghiên cứu tất cả mọi phần tử của tập hợp mà chỉ lấy ra một tập hợp gồm n phần tử để nghiên cứu. Phương pháp này gọi là phương pháp mẫu.Các lý do đó là:

1. Phải chịu chi phí rất lớn về tiền, về thời gian, nhân lực và phương tiện, …

2. Có những trường hợp điều tra sẽ làm phá hủy các phần tử được điều tra.

Ví dụ: Kiểm tra chất lượng đồ hộp.

3. Có trường hợp không xác định được hết toàn bộ N phần tử của tổng thể.

Ví dụ Số sinh viên nghiện thuốc lá trong một ký túc xá.

Từ tổng thể ta lấy ra n phần tử để nghiên cứu ta gọi là một mẫu cỡ n. Từ phương pháp toán học, phân tích nghiên cứu kỹ n phần tử ta đưa ra kết luận chung cho toàn bộ tổng thể. Do đó mẫu phải đảm bảo tính ngẫu nhiên phản ánh đúng bản chất của tổng thể.

Có nhiều cách lấy mẫu như sau

1. Lấy mẫu ngẫu nhiên

Ta đánh số các phần tử của tổng thể từ 1 đến N. Để lấy mẫu n phần tử, ta dùng bảng số ngẫu nhiên hoặc dùng cách bốc thăm lấy đủ n phần tử.

2. Chọn mẫu cơ giới

Là các phần tử của tổng thể đưa vào mẫu cách nhau một khoảng xác định.

Ví dụ: Trên dây chuyền sản xuất, sau một khoảng thời gian t lại lấy một phần tử cho vào mẫu.

3. Chọn mẫu bằng cách phân lớp (hoặc phân nhóm)

Là chia tổng thể thành một số lớp theo 1 tiêu chí nào đó. Sau đó lấy ngẫu nhiên mỗi lớp một số phần tử đưa vào mẫu.

Lấy mẫu tiến hành theo hai phương thức:

- Lấy mẫu có hoàn lại là từ tổng thể lấy một phần tử ra nghiên cứu, sau đó lại trả lại phần tử đó vào tập chính rồi mới lấy phần tử tiếp theo ra nhiên cứu.Như vậy phần tử lấy ở lần sau có thể trùng với phần tử ở lần lấy trước đó.Cứ như vậy cho đến khi được mẫu cỡ n.

- Lấy mẫu không hoàn lại là từ tổng thể lấy một phần tử ra nghiên cứu, sau đó lại không trả lại phần tử đó vào tập chính rồi mới lấy phần tử tiếp theo ra nhiên cứu.Như vậy phần tử lấy ở lần sau không thể trùng với phần tử ở lần lấy trước đó. Cứ như vậy cho đến khi được mẫu cỡ n.

Theo định lý giới hạn của xác suất, người ta chứng minh được rằng: khi số phần tử của tổng thể đủ lớn thì coi hai cách lấy mẫu theo hai phương thức trên là như nhau.

3.2. MÔ HÌNH XÁC SUẤT CỦA TỔNG THỂ VÀ MẪU

I. Đại lượng ngẫu nhiên gốc

Ta có thể mô hình hóa dấu hiệu X^* bằng một đại lượng ngẫu nhiên. Thật vậy, lấy ngẫu nhiên từ tổng thể ra một phần tử và gọi X là giá trị của dấu hiệu X^* đo được trên phần tử lấy ra. Do giá trị của X thay đổi từ phần tử này qua phần tử khác của tổng thể nên X là đại lượng ngẫu nhiên hay còn gọi là đại lượng ngẫu nhiên đám đông, có phân phối xác suất như sau

X	x_1	x_2	…	x_i	…	x_k
P	p_1	P_2	…	p_i	…	p_k

Đại lượng X gọi là đại lượng ngẫu nhiên gốc. Quy luật phân phối xác suất của X gọi là quy luật phân phối gốc.

Do trên một đám đông mỗi lần ta chỉ xét một dấu hiệu X, nên các đặc trưng của đám đông theo dấu hiệu X mang các thông tin tổng hợp về đám đông.

Ta có các tham số của đại lượng ngẫu nhiên gốc

Kì vọng toán: $E(X)=\sum_{i=1}^{k} p_i x_i$

Như vậy trung bình của tổng thể chính là kì vọng toán của đại lượng ngẫu nhiên X.

Phương sai của tổng thể: $Var(X)=\sum_{i=1}^{k}\left(x_i - E(X)\right)^2 p_i$

Như vậy phương sai Var(X) chính là phương sai của tổng thể Var(X)= σ^2.

II. Mẫu ngẫu nhiên

Từ tổng thể lấy ra n phần tử. Gọi X_i là giá trị của dấu hiệu X^* đo được trên phần tử thứ i ($i=\overline{1,n}$). Các đại lượng X_i là những đại lượng ngẫu nhiên độc lập có cùng quy luật phân phối xác suất với đại lượng X. Một mẫu có kích thước n được thành lập từ đại lượng ngẫu nhiên X là n đại lượng ngẫu nhiên độc lập có cùng phân phối xác suất với X, gọi là một mẫu ngẫu nhiên.

Kí hiệu $W_X=(X_1, X_2, \ldots, X_n)$.

Khi các đại lượng X_i nhận giá trị cụ thể x_i, ta có mẫu cụ thể kí hiệu $w_x=(x_1, x_2,\ldots, x_n)$.

Ví dụ Một lớp có 100 sinh viên là một tổng thể. Để nghiên cứu kết quả điểm thi môn Xác suất thống kê ta lấy một mẫu cỡ n=8 vì chưa có tên sinh viên cụ thể thì $W_X=(X_1, X_2, \ldots, X_8)$ là mẫu ngẫu nhiên.Khi lấy ngẫu nhiên tên cụ thể của 8 sinh viên trong lớp ta có mẫu cụ thể. Giả sử có điểm là $w_x=(5, 8,3,7,9,6,8,10)$.

Khi nghiên cứu lý thuyết ta xét mẫu tổng quát còn làm toán ta xét với mẫu cụ thể.

Ta có thể nói rằng: xác suất nghiên cứu đám đông và nhờ nó ta hiểu về mẫu, còn thống kê thì nghiên cứu mẫu và nhờ nó mà ta hiểu về đám đông.

Phân phối thực nhiệm là luật phân phối mẫu xét cho 1 mẫu cụ thể.

III. Sai số quan sát

Trong việc lấy mẫu, do nhiều nguyên nhân khác nhau, sẽ không tránh khỏi các sai số trong các số liệu mẫu. Vì vậy trước khi dùng các thống kê để phân tích, xử lý ta cần loại bỏ các sai số không đáng có trong mẫu đã cho.

Giả sử X là kết quả quan sát và a là giá trị đúng của đại lượng đang quan sát.

Khi đó $Z = x - a$ là sai số. Vì a chưa biết nên Z cũng chưa biết.

Ta phân loại các sai số như sau

1. Sai số thô

Là sai số do vi phạm các điều kiện cơ bản của việc lấy mẫu hoặc do sơ suất của người thực hiện, chẳng hạn người kiểm tra cố ý chọn ra các sản phẩm tốt để kiểm tra khi đánh giá chất lượng, hoặc kỹ thuật viên ghi nhầm kết quả thu được.

2. Sai số hệ thống

Là sai số do không điều chỉnh chính xác dụng cụ hoặc không thống nhất với nhau về cách xác định một đại lượng nào đó, dẫn đến một loạt kết quả quan sát lệch đi một tỷ lệ nhất định nào đó.

3. Sai số ngẫu nhiên

Là sai số phát sinh do một số lớn các nguyên nhân mà tác dụng của chúng nhỏ đến mức không thể tách riêng và tính riêng biệt cho từng nguyên nhân được.

Trong 3 loại sai số trên, sai số thô và sai số hệ thống cần phát hiện sớm và khử bỏ ngay, còn sai số ngẫu nhiên không thể khử bỏ được trong mỗi lần quan sát.

Việc khử bỏ sai số thô và sai số hệ thống được thực hiện tốt nhất khi ta phát hiện chúng ngay trong quá trình thu thập mẫu. Người thu nhập mẫu cảm thấy sự khác thường ở các số liệu đó, tự họ kiểm tra và câu trả lời chính xác dễ được tìm ra hơn. Còn khi nhà thống kê phát hiện ra những sự nghi ngờ thì câu trả lời khó được tìm ra, nhất là đối với sai số hệ thống.

4. Luật phân phối của sai số ngẫu nhiên

Sau khi bỏ sai số thô và sai số hệ thống chỉ còn sai số ngẫu nhiên Z=X- a thì thông thường $Z \sim N(0,\sigma^2)$ với σ là độ chính xác của dụng cụ đo đạc.

Ta có

$$P(a < Z < b) = \Phi(\frac{b}{\sigma}) - \Phi(\frac{a}{\sigma})$$

$$P(|Z| < k\sigma) = \Phi(k) - \Phi(-k) = 2\Phi(k)$$

Khi k=1 thì $P(|Z| < \sigma) = 2\Phi(1) \approx 0,68$

Khi k=2 thì $P(|Z| < 2\sigma) = 2\Phi(2) \approx 0,95$

Khi k=3 thì $P(|Z| < 3\sigma) = 2\Phi(3) \approx 0,9974$ suy ra

$P(|Z| > 3\sigma) = 1 - 0,9974 \approx 0,0026$.

Xác suất 0,0026 là quá bé, cho nên trong thực tế ta xem như sai số ngẫu nhiên không vượt quá giới hạn $\pm 3\sigma$

3.3. CÁC THAM SỐ ĐẶC TRƯNG CỦA MẪU

I. Các tham số đặc trưng của mẫu

Xét mẫu ngẫu nhiên kí hiệu $W_X=(X_1, X_2, \ldots, X_n)$.

1. Trung bình mẫu ngẫu nhiên $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

Các đại lượng X_i là những đại lượng ngẫu nhiên độc lập có cùng quy luật phân phối xác suất với đại lượng X nên $\overline{X}$ là đại lượng ngẫu nhiên.

Giả sử $E(X) = a$ và $Var(X) = \sigma^2$ thì $E(X_i) = a$; $Var(X_i) = \sigma^2$

Ta cũng chứng minh được rằng

a) Kỳ vọng của trung bình mẫu ngẫu nhiên: $E(\overline{X}) = a$

b) Phương sai của trung bình mẫu ngẫu nhiên: $Var(\overline{X}) = \frac{\sigma^2}{n}$

Từ mẫu cụ thể $w_x=(x_1, x_2,\ldots, x_n)$ ta tính được giá trị của $\overline{X}$ ký hiệu là $\overline{x}$.

2. Phương sai mẫu ngẫu nhiên $\hat{S}_X^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\overline{X}\right)^2$

Các đại lượng X_i là những đại lượng ngẫu nhiên độc lập có cùng quy luật phân phối xác suất với đại lượng X nên $\hat{S}_X^2$ là đại lượng ngẫu nhiên.

Kỳ vọng của phương sai mẫu ngẫu nhiên $E(\hat{S}^2) = \frac{n-1}{n}\sigma^2$.

Từ mẫu cụ thể $w_x=(x_1, x_2,\ldots, x_n)$ ta tính được giá trị của $\hat{S}_X^2$ ký hiệu là $\hat{s}^2$.

3. Phương sai mẫu ngẫu nhiên có hiệu chỉnh $S^2 = \frac{n}{n-1}\hat{S}^2$

Cách khác $S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - n(\overline{X})^2\right]$

Kỳ vọng của phương sai mẫu có hiệu chỉnh $E(S^2) = \sigma^2$

Độ lệch chuẩn mẫu ký hiệu là $S = \sqrt{S^2}$

Từ mẫu cụ thể $w_x=(x_1, x_2,\dots, x_n)$ ta tính được giá trị của s

4. Hệ số tỷ lệ mẫu ngẫu nhiên ký hiệu là F_n hoặc F

Giả sử đám đông chia thành 2 loại phần tử: phần tử có tính chất A và phần tử không có tính chất A. Tỉ lệ phần tử có tính chất A của đám đông chưa biết. Lấy ra mẫu cỡ n.

Gọi V là số phần tử được đánh dấu trên mẫu W. Do mẫu là ngẫu nhiên nên V là một đại lượng ngẫu nhiên. Khi đó $F = \frac{V}{n}$ được gọi là hệ số tỷ lệ mẫu ngẫu nhiên. Trên đám đông xét đại lượng ngẫu nhiên X bằng 1 nếu phần tử đám đông được đánh dấu và bằng 0 nếu phần tử đám đông không được đánh dấu. Khi đó nếu p là tỷ lệ các phần tử được đánh dấu trên đám đông thì X có bảng phân phối sau đây

X	0	1
P	q	p

và $V = X_1 + \cdots + X_n$ thì $F = \frac{V}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$ suy ra hệ số tỷ lệ mẫu ngẫu nhiên F chính là trung bình mẫu ngẫu nhiên $\overline{X}$.

Do đó có thể nói, hệ số tỷ lệ mẫu ngẫu nhiên F là trường hợp riêng của trung bình mẫu ngẫu nhiên $\overline{X}$.

Kỳ vọng của hệ số tỷ lệ mẫu ngẫu nhiên F: $E(F)= p$.

Phương sai của hệ số tỷ lệ mẫu ngẫu nhiên F: $D(F) = \frac{pq}{n}$

Từ mẫu cỡ n cụ thể ta tính được giá trị của F ký hiệu là $f = \frac{n_A}{n}$

Với n_A là tổng số phần tử có tính chất A trong mẫu cỡ n cụ thể

Chú ý:

Từ mẫu cụ thể $w_x=(x_1, x_2, \ldots, x_n)$

Tính các giá trị của các thống kê mẫu.

* Khi X nhận giá trị x_i với số lần $n_i = 1$ Ta có

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \ ; \qquad \hat{s}_x^{\,2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\overline{x}\right)^2 ;$$

Phương sai mẫu hiệu chỉnh

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

* Khi X nhận giá trị x_i với số lần n_i ta có $\sum_{i=1}^{k} n_i = n$; Thì

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

Phương sai mẫu hiệu chỉnh

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

II. Cách tính các đặc trưng mẫu

1.Trường hợp 1 X nhận giá trị x_i với số lần n_i và mẫu cỡ n nhỏ

Ví dụ 1 Điều tra năng suất lúa trên diện tích 100ha trồng lúa của 1 vùng, ta thu được bảng số liệu sau

Năng suất(tạ/ha)	41	44	45	46	48	52	54
Diện tích (ha)	10	20	30	15	10	10	5

Tính giá trị trung bình và phương sai mẫu hiệu chỉnh năng suất lúa.

BÀI GIẢI

Ta lập bảng tính như sau:

x_i	n_i	$x_i.n_i$	$x_i^2.n_i$
41	10	410	16.810
44	20	880	38.720
45	30	1.350	60.750
46	15	690	31.740
48	10	480	23.040
52	10	520	27.040
54	5	270	14.580
Tổng	n = 100	4.600	212.680

Từ kết quả tính ở bảng trên ta có:

Năng suất trung bình $\overline{x} = \frac{1}{n}\sum_{i=1}^{k} x_i.n_i = \frac{4600}{100} = 46$ (tạ/ha)

Phương sai của năng suất

$$\hat{s}_x^{\,2} = \frac{1}{n}\sum_{i=1}^{k} x_i^2.n_i - \left(\overline{x}\right)^2 = \frac{212680}{100} - \left(46\right)^2 = 10,8$$

Phương sai mẫu hiệu chỉnh $S^2 = \frac{n}{n-1}\hat{S}^2 = \frac{100}{100-1}10,8 = 10,9$

Cách khác

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{k} n_i x_i^2 - n(\overline{x})^2\right] = \frac{1}{99}\left[212680 - 100.(46)^2\right] \approx 10,9$$

2.***Trường hợp 2*** X nhận giá trị x_i với số lần n_i và cỡ mẫu n lớn.

Khi đó ta chia thành k khoảng. Ta đếm các giá trị rơi vào khoảng thứ i là n_i

Ví dụ 2 Cho bảng số liệu sau

$x_i - x_{i+1}$	n_i	$x_i - x_{i+1}$	n_i
4 – 12	143	44 – 52	9
12 – 20	75	52 – 60	5
20 – 28	53	60 – 68	4
28 – 36	27	68 – 76	3
36 – 44	14	76 – 80	3

Tính giá trị trung bình mẫu và phương sai mẫu.

BÀI GIẢI

Đặt $x_i^o = \dfrac{x_{i+1} + x_i}{2}$, ta lập bảng tính toán như sau

$x_i - x_{i+1}$	n_i	x_i^0	$n_i \cdot x_i^0$	$n_i \cdot (x_i^0)^2$
4 – 12	143	8	1144	9152
12 – 20	75	16	1200	19.200
20 – 28	53	24	1272	30.528
28 – 36	27	32	864	27.648
36 – 44	14	40	560	22.400
44 – 52	9	48	432	20.736
52 – 60	5	56	280	15.680
60 – 68	4	64	256	16.384
68 – 76	3	72	216	15.552
76 – 80	3	78	234	18.252
Tổng	n = 336		6458	195.532

Trung bình mẫu $\overline{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i^0 = 19,22$

Phương sai mẫu $\hat{s}_x^2 = \frac{1}{n}\sum_{i=1}^{k}\left(x_i^0\right)^2 n_i - \left(\overline{x}\right)^2 = 211,93$

Phương sai mẫu hiệu chỉnh $S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{k} n_i \left(x_i^0\right)^2 - n(\overline{x})^2\right] = 213$

Cách khác

Nếu số liệu lớn thì sử dụng công thức đổi biến cho đơn giản

Chọn x_0 là x_i^0 ứng với max n_i và h là độ rộng của khoảng.

$$x_i^o = \frac{x_{i+1} + x_i}{2} \;;\; u_i = \frac{x_i^o - x_o}{h};$$

Ta tính lại ví dụ 2 ở trên bằng phương pháp đổi biến.

Ta chọn $x_o = 8$ (ứng với max $n_i = 143$); $h = 8$

$u_i = \frac{x_i^o - x_o}{h} = \frac{x_i^o - 8}{8}$ Lập bảng tính như sau

$x_i - x_{i+1}$	x_i^o	n_i	u_i	$n_i u_i$	$n_i . u_i^2$
4 – 12	8	143	0	0	0
12 – 20	16	75	1	75	75
20 – 28	24	53	2	106	212
28 – 36	32	27	3	81	243
36 – 44	40	14	4	56	224
44 – 52	48	9	5	45	225
52 – 60	56	5	6	30	180
60 – 68	64	4	7	28	196
68 – 76	72	3	8	24	192
76 – 80	78	3	8,75	26,25	229,6875
Tổng		n=336		471,25	1776,6875

với $\overline{u} = \frac{1}{n}\sum_{i=1}^{k} n_i u_i$; $\overline{x} = h.\overline{u} + x_o$;

$\hat{S}^2 = h^2\left[\frac{1}{n}\sum_{i=1}^{k} n_i u_i^2 - \left(\overline{u}\right)^2\right]$; $S^2 = \frac{n}{n-1}\hat{S}^2$

Ta có $\overline{u} = \frac{1}{n}\sum_{i=1}^{k} n_i u_i = \frac{471,25}{336}$ =1,4025, $\overline{x} = h.\overline{u} + x_o$ = 19,22

$\hat{s}^2 = h^2\left[\frac{1}{n}\sum_{i=1}^{k} n_i u_i^2 - \left(\overline{u}\right)^2\right]$ = 211,93 ; S^2 = 213

III. Luật số lớn và định lý giới hạn trung tâm

1. Hội tụ theo xác suất

Cho dãy đại lượng ngẫu nhiên (X_n) và đại lượng ngẫu nhiên X. Chúng ta nói rằng X_n hội tụ về X theo xác suất, ký hiệu là $X_n \xrightarrow{P} X$, nếu với mọi $\varepsilon > 0$, xác suất của biến cố $\left(|X_n - X| \geq \varepsilon\right)$ hội tụ về 0 khi $n \to \infty$. Nghĩa là $\lim_{n\to\infty} P\left(|X_n - X| \geq \varepsilon\right) = 0$.

2. Hội tụ với xác suất bằng 1

Cho dãy đại lượng ngẫu nhiên (X_n) và đại lượng ngẫu nhiên X. Chúng ta nói rằng X_n hội tụ về X với xác suất bằng 1 (còn gọi là hội tụ hầu chắc chắn), ký hiệu là $P(X_n \to X) = 1$, nếu biến cố $(X_n \to X)$ là biến cố chắc chắn.

Nếu X_n hội tụ về X hầu chắc chắn thì X_n hội tụ theo xác suất về X.

3. Định lý Bernoulli

Xét phép thử T, gọi p là xác suất biến cố A xuất hiện sau mỗi lần thực hiện phép thử T.

Thực hiện phép thử n lần độc lập, gọi X_n là số lần biến cố A xuất hiện. Khi đó tần suất $v_n = \frac{X_n}{n}$ của A (là dãy các đại lượng ngẫu nhiên) hội tụ theo xác suất về xác suất p của biến cố A, khi $n \to \infty$, nghĩa là: $\lim_{n\to\infty} P\left(|v_n - p| \geq \varepsilon\right) = 0$.

4. Định lý Lindeberg - Levy

Giả sử $X_1, X_2, \ldots$ là dãy các đại lượng ngẫu nhiên độc lập có cùng phân phối với đại lượng ngẫu nhiên X, với kỳ vọng là μ và phương sai là σ^2.

Gọi $\overline{X}_n$ là trung bình cộng của n đại lượng ngẫu nhiên đầu tiên của dãy nói trên, nghĩa là

$$\overline{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

Ta có $\overline{X}_n$ là dãy đại lượng ngẫu nhiên.

Ký hiệu $\overline{X}_n^* = \dfrac{\overline{X}_n - E(\overline{X}_n)}{\sqrt{Var(\overline{X}_n)}}$ là đại lượng ngẫu nhiên trung tâm chuẩn hóa của $\overline{X}_n$.

Khi đó với mọi x cố định, $P\left(\overline{X}_n^* < x\right) \to \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{x} e^{-t^2/2}\, dt$ khi $n \to \infty$.

Từ đó suy ra, với n khá lớn, có thể xem $\overline{X}_n^*$ có phân phối chuẩn tắc N(0,1) và $\overline{X}_n$ có phân phối chuẩn $N\left(\mu, \dfrac{\sigma^2}{n}\right)$.

Định lý Lindeberg – Levy là cơ sở trong các lý luận tiếp theo sau đây.

Người ta cũng đã chứng minh được rằng

1) $\overline{X}_n \to \mu$ khi $n \to \infty$

2) $\hat{S}_n^2 \to \sigma^2$, $S_n^2 \to \sigma^2$ khi $n \to \infty$

3) $F_n \to p$ khi $n \to \infty$

Do đó đến chương ước lượng người ta có thể dùng:

- $\overline{X}$ để ước lượng μ
- S^2 để ước lượng σ^2 và F để ước lượng p.

5. Luật phân phối xác suất của các đặc trưng mẫu ngẫu nhiên

Bổ đề

Cho các đại lượng ngẫu nhiên X_1, X_2, … độc lập, có cùng phân phối với đại lượng ngẫu nhiên X. Giả sử $E(X) = \mu$ và $Var(X)=\sigma^2$.

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

Chúng ta có $\overline{X}$ và $\hat{S}^2$ độc lập.
Khi đó

$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ và $\frac{n\hat{S}^2}{\sigma^2} \sim \chi^2_{n-1}$ (phân phối χ^2 $(n-1)$ bậc tự do)

Từ đó suy ra $\frac{\left(\overline{X} - \mu\right)\sqrt{n}}{S} \sim T_{n-1}$ (phân phối Student (*n-1*) bậc tự do)

Từ bổ đề này, định lý Lindeberg – Levy ta suy ra

a) Nếu X có phân phối chuẩn N(μ, σ^2) thì

$$\overline{X} \sim N\left(a, \frac{\sigma^2}{n}\right); \quad \frac{\left(\overline{X} - \mu\right)\sqrt{n}}{\sigma} \sim N(0,1); \quad \frac{\left(\overline{X} - \mu\right)\sqrt{n}}{S} \sim T_{n-1}$$

b) Nếu X có phân phối bất kỳ và n khá lớn thì

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right); \quad \frac{\left(\overline{X} - \mu\right)\sqrt{n}}{\sigma} \sim N(0,1)$$

c) Hệ số tỷ lệ F ~ B(n,p) , do đó nếu n khá lớn thì

$$F \sim N(p, \frac{pq}{n}); \quad \frac{(F-p)\sqrt{n}}{\sqrt{pq}} \sim N(0,1)$$

Có thể xem F là trường hợp riêng của $\overline{X}$ khi X có phân phối 0-1, Do đó khi n khá lớn F có phân phối chuẩn hơn nữa $F \sim N(p, \frac{pq}{n})$.

Trong thực tế, khi n > 30 chúng ta có thể xem là n khá lớn nên áp dụng được các quy tắc ở trên. Trong trường hợp không biết σ^2 thì có thể thay σ^2 bằng s^2, trong đó s^2 là phương sai mẫu có hiệu chỉnh lấy trên một mẫu có kích thước n đủ lớn nào đó.

BÀI TẬP CHƯƠNG III

3.1. 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_i	12,00	12,05	12,10	12,15	12,20	12,25	12,30	12,35	12,40
n_i	2	3	7	9	10	8	6	5	3

Với n_i chỉ số trường hợp tính theo từng giá trị của X (mm).

Tính trung bình mẫu $\overline{x}$ và độ lệch chuẩn s_x của mẫu.

Đáp số: $n = 53$; trung bình $\overline{x} = 12,2066$; độ lệch $s_x = 0,1029$.

3.2. Đem cân một số trái xoài vừa thu hoạch, ta được kết quả sau

x_i (gam)	200-210	210-220	220-230	230-240	240-250
Số trái:n_i	12	17	20	18	15

Tính trung bình mẫu $\overline{x}$ và độ lệch chuẩn s_x của mẫu.

Đáp số: $n = 82; \overline{x} = 225,8537; s_x = 13,2592$.

3.3. 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;

Tính trung bình mẫu $\overline{x}$ và độ lệch chuẩn s_x của mẫu.

Đáp số: $n = 12; \overline{x} = 137,8333; s_x = 4,4073; s_x^2 = 19,4242$

3.4. Ở một cửa hàng bán xăng dầu, theo dõi nhu cầu của mặt hàng xăng trong một số ngày, ta có kết quả ở bảng sau

Số bán ra (lít)	Số ngày	Số bán ra (lít)	Số ngày
20 – 30	3	70 – 80	25
30 – 40	8	80 – 90	17
40 – 50	30	90 – 100	9
50 – 60	45	>100	4
60 – 70	20		

Tính trung bình mẫu $\overline{x}$ và độ lệch chuẩn s_x của mẫu.

Đáp số: $n = 161; \overline{x} = 62,5776; s_x = 17,9156$

3.5. Khảo sát về thu nhập X (triệu đồng/tháng) của một số người, ta có bảng số liệu sau:

Thu nhập	0 – 4	4 – 8	8 – 12	12 – 16	16 – 20	20 – 24
Số người	8	12	20	30	16	10

Tính trung bình mẫu $\overline{x}$ và độ lệch chuẩn s_x của mẫu.

Đáp số: $n = 96; \overline{x} = 12,667; s_x = 5,587$

3.6. Cân thử 100 trái cây của một nông trường, ta có kết quả như sau

Trọng lượng (g)	Số trái	Trọng lượng (g)	Số trái
35 – 55	3	115 – 135	20
55 – 75	10	135 – 155	6
75 – 95	25	155 – 175	1
95 – 115	35		

Tính trung bình mẫu $\overline{x}$ và độ lệch chuẩn s_x của mẫu.

Đáp số: $n = 100; \overline{x} = 101,2; s_x = 22,9013$

3.7. Đo lượng cholesterol (đơn vị mg%) cho một số người, ta được

X(mg%)	150-160	160-170	170-180	180-190	190-200	200-210
Số người	2	4	5	6	4	3

a) Tính trung bình mẫu và độ lệch chuẩn mẫu.

b) Một mẫu thứ nhì Y có 30 người cho $\overline{y}$ =180 mg% , s_y =16 mg%.

Nhập hai mẫu lại, tính trung bình và độ lệch chuẩn của mẫu nhập.

CHƯƠNG IV
ƯỚC LƯỢNG CÁC THAM SỐ ĐẶC TRƯNG CỦA TỔNG THỂ

Các đặc trưng bằng số của đại lượng ngẫu nhiên X đặc trưng cho tổng thể như: kì vọng, phương sai, tỷ lệ … được sử dụng rất nhiều trong phân tích kinh tế-xã hội và cả trong lĩnh vực khoa học kỹ thuật. Nhưng các số đặc trưng này thường chưa biết. Vì vậy vấn đề đặt ra là ta đi ước lượng chúng bằng phương pháp mẫu.

Cho đại lượng ngẫu nhiên X có thể đã biết hoặc chưa biết quy luật phân phối xác suất và gọi θ là một tham số nào đó của đại lượng ngẫu nhiên X. Hãy ước lượng θ bằng phương pháp mẫu.

Có 2 phương pháp ước lượng θ:

- Dùng một con số nào đó để ước lượng tham số θ ta gọi là ước lượng điểm.

- Chỉ ra θ rơi vào khoảng (g_1, g_2) ta gọi đó là phương pháp ước lượng khoảng.

Ta sẽ xét cả 2 phương pháp, các phương pháp này xuất phát từ cơ sở hợp lý nào đó để tìm ước lượng của θ chứ không phải là sự chứng minh chặt chẽ.

4.1. CÁC PHƯƠNG PHÁP TÌM ƯỚC LƯỢNG ĐIỂM

Các đặc trưng bằng số của đại lượng ngẫu nhiên X đặc trưng cho tổng thể thông thường chưa biết, ta đi ước lượng chúng bằng phương pháp mẫu. Từ mẫu ta tính ra được một số nào đó dùng số này để ước lượng tham số cần tìm. Ta gọi là ước lượng điểm.

I. Phương pháp hàm ước lượng

1. Mô tả phương pháp

Gọi θ là một tham số nào đó của đại lượng ngẫu nhiên X. Từ X lập mẫu ngẫu nhiên kích thước n :

$W_X=(X_1, X_2, \ldots, X_n)$. Chọn $G=f(X_1, X_2, \ldots, X_n)$.

G là hàm của các đại lượng ngẫu nhiên $X_1, X_2, \ldots, X_n$ nên G là một đại lượng ngẫu nhiên. G gọi là hàm ước lượng của θ . Trong thực tế ta thường chọn hàm ước lượng như sau:

- **Ước lượng điểm cho trung bình đám đông**

Người ta thường chọn $$G = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- **Ước lượng điểm cho phương sai đám đông**

Người ta thường chọn $$G = S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - n(\overline{X})^2\right]$$

- **Ước lượng điểm cho tỷ lệ đám đông** chọn tỷ lệ mẫu f

2. Các tiêu chuẩn lựa chọn hàm ước lượng

Ta thấy có vô số cách chọn dạng hàm f làm hàm ước lượng của θ. Vì vậy cần có một tiêu chuẩn để đánh giá chất lượng của ước lượng. Từ đó lựa chọn được hàm ước lượng tốt hơn theo nghĩa nào đó. Sau đây là một số tiêu chuẩn

a) Ước lượng không chệch

G là ước lượng không chệch của θ nếu E(G)= θ

Ý nghĩa: (G - θ) là sai số của ước lượng, mà theo tính chất của kỳ vọng ta có: E (G-θ)= E(G) - E(θ)= θ–θ =0.

Như vậy ước lượng không chệch là ước lượng có sai số trung bình bằng không. Tức là giá trị của G không bị lệch về một phía, nếu dùng G để ước lượng θ thì không mắc phải sai số hệ thống.

Chú ý: G là ước lượng không chệch của θ không có nghĩa là mọi giá trị của G đều trùng khít với θ mà chỉ có nghĩa là: trung bình các giá trị của G bằng θ, một giá trị của G có thể sai khác nhiều so với θ.

Trong thưc tế người ta thường chọn như sau

*** Ước lượng điểm cho trung bình đám đông**

$$G = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{vì} \quad E(\overline{X}) = a$$

Nên dùng trung bình mẫu $\overline{X}$ là ước lượng không chệch cho trung bình đám đông.

***Ước lượng điểm cho phương sai đám đông**

$$G = S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - n(\overline{X})^2\right] \text{ vì } E(S^2) = \sigma^2$$

Nên dùng phương sai mẫu có hiệu chỉnh s^2 là ước lượng không chệch cho phương sai đám đông

*** Ước lượng điểm cho tỷ lệ đám đông** vì E(f)=p nên dùng tỷ lệ mẫu f là ước lượng không chệch cho tỷ lệ đám đông

Ví dụ 1 Cân thử 100 trái cây của một nông trường, ta có kết quả như sau

Trọng lượng (g)	Số trái	Trọng lượng (g)	Số trái
35 – 55	3	115 – 135	20
55 – 75	10	135 – 155	6
75 – 95	25	155 – 175	1
95 – 115	35		

a) Tìm ước lượng không chệch cho trọng lượng trung bình của 1 trái cây trong nông trường.

b) Tìm ước lượng không chệch cho đại lượng biểu thị độ đồng đều của các trái cây trong nông trường.

c) Nếu xem các trái có trọng lượng không quá 95g là trái loại II. Tìm ước lượng không chệch cho tỷ lệ trái loại II trong nông trường.

BÀI GIẢI

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i = \frac{10120}{100} = 101{,}2$$

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{k} x_i^2 n_i - n\left(\overline{x}\right)^2\right) = \frac{1}{99}\left(1080700 - 100(101{,}2)^2\right)$$

$$s^2 = 571{,}2727 \Rightarrow \sqrt{s^2} \approx 23{,}9013$$

a) Theo đề bài đây là bài toán ước lượng điểm cho trung bình đám đông. Ta biết trung bình mẫu là ước lượng không chệch cho trung bình đám đông. Theo bảng tính ở trên thì: $\overline{x}$ = 101,2 và ta dự đoán ước lượng không chệch cho trọng lượng trung bình của trái cây trong nông trường là 101,2 g.

b) Đây là bài toán ước lượng điểm cho phương sai đám đông. Ta biết phương sai mẫu có hiệu chỉnh là ước lượng không chệch cho phương sai đám đông. Theo trên ta tính được: s^2 = 571,2727. Vậy dự đoán độ đồng đều của trái cây trong nông trường vào khoảng 571,227.

c) Đây là bài toán ước lượng điểm cho tỷ lệ đám đông. Ta biết tỷ lệ mẫu là ước lượng không chệch cho tỷ lệ đám đông. Theo đề bài ta tính được tỷ lệ mẫu f = (3 + 10 + 25)/100 = 0,38. Vậy dự đoán tỷ lệ trái cây loại II trong nông trường vào khoảng 38%.

b) Ước lượng hiệu quả

$G=f(X_1, X_2, \ldots, X_n)$, là ước lượng không chệch của θ. Áp dụng bất đẳng thức Chebyshev cho G ta có $P(|G-\theta|<\varepsilon)\geq 1-\dfrac{Var(G)}{\varepsilon^2}$.

Như vậy nếu phương sai Var(G) càng nhỏ thì xác suất để G nhận giá trị càng gần θ bao nhiêu cũng được, càng lớn. Do đó phương sai của G là một chỉ tiêu quan trọng phản ánh chất lượng của hàm ước lượng $G=f(X_1, X_2, \ldots, X_n)$.

Định nghĩa

$G=f(X_1, X_2, \ldots, X_n)$ là ước lượng không chệch của θ và phương sai Var(G) bằng cận dưới các phương sai của các hàm ước lượng được xây dựng từ mẫu ngẫu nhiên $W_X=(X_1, X_2, \ldots, X_n)$ thì G được gọi là ước lượng hiệu quả của **θ.**

c) Ước lượng vững

Định nghĩa Cho mẫu ngẫu nhiên $W_X=(X_1, X_2, \ldots, X_n)$.

Hàm ước lượng $G=f(X_1, X_2, \ldots, X_n)$ của θ được gọi là ước lượng vững nếu với mọi $\varepsilon > 0$ bé tùy ý cho trước ta đều có

$\lim\limits_{n\to\infty} P[|f(X_1, X_2, ...X_n) - \theta| < \varepsilon] = 1$.

Định lý G=f(X_1, X_2, …, X_n) là ước lượng không chệch của θ và $\lim\limits_{n\to\infty} Var(G) = 0$ thì G là ước lượng vững của θ.

II. Phương pháp ước lượng hợp lý cực đại

Giả sử ta cần ước lượng tham số đặc trưng θ của đại lượng ngẫu nhiên *X*. $(X_1, X_2, ..., X_n)$ là mẫu ngẫu nhiên kích thước *n* được lập thành từ *X*.

Hàm mật độ của *X* có dạng $f(x,\theta)$

Ta lập hàm số $L(\theta) = f(x_1,\theta).f(x_2,\theta)...f(x_n,\theta)$

(Nếu *X* rời rạc thì $L(\theta) = P(x_1,\theta).P(x_2,\theta)...P(x_n,\theta)$)

Hàm $L(\theta)$ được gọi là hàm hợp lí của θ.

Trong biểu thức của $L(\theta)$ có x_1, x_2, …, x_n và θ nhưng ta xem x_1, x_2, …, x_n (giá trị cụ thể bất kỳ của mẫu) cố định và θ là biến số thuộc khoảng (a,b) nào đó.

$G = G(x_1, x_2, ..., x_n)$ được gọi là ước lượng hợp lý cực đại của θ nếu $L(G) \ge L(\theta)$ với mọi $\theta \in (a,b)$

Hàm $G = G(X_1, X_2, ..., X_n)$ được gọi là hàm ước lượng hợp lý cực đại.

Vì hàm L và hàm $\ln L$ đạt cực đại tại cùng một giá trị của θ nên ta có thể tìm giá trị của θ để $\ln L$ đạt cực đại theo các bước sau:

Bước 1: Tìm đạo hàm bậc nhất của $\ln L$ theo θ.

Bước 2: Lập phương trình: $\dfrac{\partial \ln L}{\partial \theta} = 0$ Phương trình này gọi là phương trình hợp lý. Giả sử nó có nghiệm là $\theta_0 = \varphi(x_1, x_2, ..., x_n)$.

Bước 3: Tìm đạo hàm bậc hai $\dfrac{\partial^2 \ln L}{\partial \theta^2}$

Nếu tại điểm $\theta_0 = \varphi\left(x_1, x_2, ..., x_n\right)$, đạo hàm bậc hai âm thì tại điểm này hàm $\ln L$ đạt cực đại. Do đó, $\theta_0 = \varphi\left(x_1, x_2, ..., x_n\right)$ là ước lượng hợp lý tối đa của θ.

Ví dụ 2 Giả sử $X \sim N\left(\mu, \sigma^2\right)$, trong đó σ^2 đã biết, hãy tìm ước lượng hợp lý cực đại của μ.

BÀI GIẢI

Ta lập hàm $L(\theta) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x_i-\mu)^2}{2\sigma^2}} = \frac{1}{\left(\sigma\sqrt{2\pi}\right)^n} e^{\frac{-1}{2\sigma^2}\sum(x_i-\mu)^2}$

Với $\theta = \mu \ \ (\mu \in R)$ thì

- $\ln L(\theta) = \ln L(\mu) = n\ln\frac{1}{\sigma\sqrt{2\pi}} - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i - \mu\right)^2$

- $\frac{d\ln L(\mu)}{d\mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(x_i - \mu\right)$

- Giải phương trình $\frac{d\ln L(\mu)}{d\mu} = 0$ ta được:

$$G\left(x_1, x_2, ..., x_n\right) = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- Hơn nữa $\frac{d^2\ln L(\mu)}{d\mu^2} = -\frac{n}{\sigma^2} < 0$

Như vậy $\overline{X}$ là hàm ước lượng hợp lý cực đại của μ.

4.2. ƯỚC LƯỢNG KHOẢNG

Gọi θ là một tham số nào đó của đại lượng ngẫu nhiên X cần ước lượng. θ rơi vào khoảng (g_1, g_2) ta gọi đó là phương pháp ước lượng khoảng.

I. Mô tả phương pháp ước lượng khoảng.

Gọi θ là một tham số nào đó của đại lượng ngẫu nhiên X cần ước lượng

Từ X lập mẫu ngẫu nhiên $W_X=(X_1, X_2, \dots, X_n)$.

Chọn thống kê $G= f(X_1, X_2, \dots, X_n)$ chứa tham số θ sao cho: mặc dù chưa biết giá trị của θ nhưng qui luật phân phối xác suất của G vẫn hoàn toàn xác định. Do đó với xác suất α_1 khá bé ta tìm được phân vị g_{α_1} của thống kê G, tức là g_{α_1} thỏa mãn: $P(G < g_{\alpha_1})= \alpha_1$

Với xác suất α_1 mà $\alpha_1 + \alpha_2 = \alpha$ khá bé (trong thực tế ta lấy $\alpha \le 0,05$) ta tìm phân vị $g_{1-\alpha_2}$ tức $P(G < g_{1-\alpha_2})= 1-\alpha_2$.

Suy ra $P(g_{\alpha_1} \le G \le g_{1-\alpha_2}) = P(G < g_{1-\alpha_2}) - P(G < g_{\alpha_1}) = 1-\alpha_2-\alpha_1 = 1-\alpha$

Từ $g_{\alpha_1} \le G \le g_{1-\alpha_2}$ giải ra được θ : $P(G_1 \le \theta \le G_2) = 1-\alpha$

+) Khoảng (G_1, G_2) được gọi là khoảng ước lượng của θ.

+) $1-\alpha$ độ tin cậy của ước lượng.

Phương pháp ước lượng khoảng có ưu điểm là không những chỉ ra được khoảng (G_1, G_2) để ước lượng θ mà còn cho biết độ tin cậy của ước lượng. Tuy nhiên nó cũng chứa đựng khả năng mắc phải sai lầm, xác suất mắc phải sai lầm là α.

Chú ý: - Khoảng tin cậy càng lớn thì độ tin cậy càng cao. Với độ tin cậy bằng 1 thì khoảng tin cậy lớn nhất và bằng $(-\infty, +\infty)$. Khi đó ước lượng không có ý nghĩa, nên chỉ xét với độ tin cậy nhỏ hơn 1.

Trong thực hành thường lấy độ tin cậy là một trong các xác suất 90%, 95%, 99%.

- Trong tất cả các khoảng tin cậy (G_1, G_2) ứng với độ tin cậy $1-\alpha$ cho trước thì khoảng tin cậy có dạng đối xứng (-G,G) có độ dài ngắn nhất.

II. Ước lượng trung bình của tổng thể là m

(hay ước lượng kỳ vọng E(X))

Giả sử tổng thể có m =E(X) chưa biết. Dựa vào mẫu ngẫu nhiên $(X_1, X_2,\ldots,X_n)$, cụ thể là dựa vào $\overline{X}$ hãy tìm 2 số G_1, G_2 sao cho $P(G_1 < m < G_2) = 1-\alpha$ khá lớn cho trước, xảy ra các trường hợp sau:

1. Trường hợp 1: $X \sim N(m;\sigma^2)$ với σ^2 đã biết.

Dựa vào định lí giới hạn trung tâm $\overline{X} = \frac{1}{n}\left(X_1 + X_2 + \cdots + X_n\right)$ tuân theo quy luật phân phối xấp xỉ chuẩn $N\left(m, \frac{\sigma^2}{n}\right)$.

Khi đó $\frac{\left(\overline{X}-m\right)\sqrt{n}}{\sigma} \sim N(0,1)$

Theo quy tắc k-σ ta có

$$P\left(\left|\overline{X}-m\right| < \varepsilon\right) = 1-\alpha \Leftrightarrow P\left(\frac{\left|\overline{X}-m\right|\sqrt{n}}{\sigma} < \frac{\varepsilon\sqrt{n}}{\sigma}\right) = 1-\alpha.$$

Đặt $t_\alpha = \frac{\varepsilon\sqrt{n}}{\sigma}$ thì $2\Phi\left(t_\alpha\right) = 1-\alpha$.

Với $1-\alpha$ cho trước từ bảng tích phân Laplace (bảng 1 hoặc bảng 3) ta tìm được t_α và đặt $\varepsilon = t_\alpha \frac{\sigma}{\sqrt{n}}$.

ε gọi là độ chính xác hoặc sai số của ước lượng.Vậy $\overline{X}-\varepsilon < m < \overline{X}+\varepsilon$. Ta có: $P\left(\overline{X}-\varepsilon < m < \overline{X}+\varepsilon\right) = 1-\alpha$.

Qui tắc thực hành: Từ mẫu cỡ n cụ thể *$(x_1, x_2, \ldots, x_n)$* tính $\overline{x}$

Khi đó khoảng ước lượng của m là: $\left(\overline{x} - \varepsilon < m < \overline{x} + \varepsilon\right)$

với $\varepsilon = t_\alpha \dfrac{\sigma}{\sqrt{n}}$;

Từ độ tin cậy $1 - \alpha$ cho trước tra bảng phân vị chuẩn (tra bảng 1 hoặc bảng 3) ta có t_α.

Ví dụ 1 Cân thử trọng lượng 15 con gà một trại chăn nuôi khi xuất chuồng, ta được các kết quả sau (đơn vị đo là kg)

3,25	2,5	4	3,75	3,8	3,9	4,02	3,8
3,2	3,82	3,4	3,6	3,75	4	3,5	

Giả sử trọng lượng các con gà tuân theo luật phân phối chuẩn với phương sai 0,01. Hãy ước lượng trọng lượng trung bình của một con gà với độ tin cậy 99%.

BÀI GIẢI

Đây là bài toán ước lượng khoảng cho trung bình đám đông.

Gọi X là trọng lượng của một con gà trong một trại chăn nuôi

m là trọng lượng trung bình của một con gà. Ta ước lượng m:

Trường hợp này kích thước mẫu n = 15 và đã biết $\sigma^2 = 0{,}01$

Vậy khoảng ước lượng của m là:

$\left(\overline{x} - \varepsilon < m < \overline{x} + \varepsilon\right)$ với $\varepsilon = t_\alpha \dfrac{\sigma}{\sqrt{n}}$

Từ độ tin cậy $1 - \alpha = 99\%$ thì tra bảng phân vị chuẩn (bảng 3) có $t_\alpha = 2{,}5758$.

Từ bảng số liệu ta tính được: $\overline{x} = \dfrac{1}{n}\sum_{i=1}^{k} n_i x_i = 3{,}6193$ (kg) .

Từ đó ta có: $\varepsilon = t_\alpha \dfrac{\sigma}{\sqrt{n}} = 2,5758.\dfrac{0,1}{\sqrt{15}} = 0,0665$

Vậy với độ tin cậy 99% khoảng ước lượng của m là:

(3,6193-0,0665 < m < 3,6193+0,0665) hay (3,5528kg; 3,6858kg)

2. Trường hợp 2: Cỡ mẫu $n \geq 30$ và $X \sim N(m;\sigma^2)$ với σ^2 chưa biết.

Vì $\overline{X} \sim N\left(m, \dfrac{\sigma^2}{n}\right)$.

Lập luận tương tự như trường hợp 1 nhưng thay $\sigma^2 = s^2$.

Qui tắc thực hành: Từ mẫu cỡ n cụ thể: $(x_1, x_2, \ldots, x_n)$ tính $\overline{x}$ và s.

Khi đó khoảng ước lượng của m là

$$\left(\overline{x} - \varepsilon < m < \overline{x} + \varepsilon\right) \qquad \text{với } \varepsilon = t_\alpha \frac{s}{\sqrt{n}}$$

Từ độ tin cậy 1 - α cho trước tra bảng phân vị chuẩn (tra bảng 3) có t_α.

Ví dụ 2 Ở một cửa hàng bán xăng dầu, theo dõi nhu cầu của mặt hàng xăng trong một số ngày, ta có kết quả ở bảng sau

Số bán ra (lít)	Số ngày	Số bán ra (lít)	Số ngày
20 – 30	3	70 – 80	25
30 – 40	8	80 – 90	17
40 – 50	30	90 – 100	9
50 – 60	45	>100	4
60 – 70	20		

a) Hãy ước lượng lượng xăng trung bình bán ra trong một ngày với độ tin cậy 99%. Giả sử lượng xăng bán ra trong ngày tuân theo phân phối chuẩn.

b) Giả sử giá một lít xăng là 20.000 đồng. Cửa hàng này phải dự trù một món tiền trung bình bằng bao nhiêu để cung cấp xăng cho khách hàng trong một tháng.

BÀI GIẢI

Từ bảng số liệu ta tính được: $\overline{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i = \frac{10075}{161} \approx 62,5776;$

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{k} x_i^2 n_i - n\left(\overline{x}\right)^2\right) = \frac{1}{160}\left(681825 - 161(62,5776)^2\right)$$

$$s^2 = 320,9705 \quad \Rightarrow \quad s \approx 17,9156$$

a) Đây là bài toán ước lượng khoảng cho trung bình đám đông.

Gọi X là lượng xăng bán ra trong ngày

m là lượng xăng trung bình bán ra trong ngày, ta tìm ước lượng của m

Trường hợp này kích thước mẫu n = 161 và chưa biết σ^2

Vậy khoảng ước lượng của m là:

$(\overline{x} - \varepsilon < m < \overline{x} + \varepsilon)$ với $\varepsilon = t_\alpha \frac{s}{\sqrt{n}}$

Với độ tin cậy 1 - α = 99% thì tra bảng phân vị chuẩn (bảng 3) có t_α = 2,5758. Từ số liệu đã tính toán ở bảng trên ta có : $\overline{x}$ = 62,5776 (lít) và s = 17,9156.

Từ đó $\varepsilon = t_\alpha \frac{s}{\sqrt{n}} = 2,5758\frac{17,9156}{\sqrt{161}} = 3,6362$

Vậy với độ tin cậy 99% khoảng ước lượng của m là:

(62,5776 - 3,6362< m < 62,5776 + 3,6362)

hay (58,9414 lít; 66,2138 lít)

b) Giả sử giá một lít xăng là 20.000 đồng và cửa hàng nhập xăng hàng ngày thì phải chuẩn bị một số vốn trung bình cho một tháng là từ:

(30 x 58,9414 x 20.000)đồng đến (30x 66,2138 x 20.000) đồng

3. Trường hợp 3: Cỡ mẫu $n < 30$ và $X \sim N(m;\sigma^2)$ với σ^2 chưa biết.

Khi $n < 30$, σ^2 chưa biết vì $\overline{X} \sim N\left(m, \dfrac{\sigma^2}{n}\right)$. Thay $\sigma^2 = s^2$, khi đó

$T = \dfrac{\left(\overline{X} - m\right)\sqrt{n}}{s}$ ~ ***T(n-1).***

Ta có : $P\left(\left|\overline{X} - m\right| < \varepsilon\right) = 1 - \alpha \Leftrightarrow P\left(\dfrac{\left|\overline{X} - m\right|\sqrt{n}}{s} < \dfrac{\varepsilon\sqrt{n}}{s}\right) = 1 - \alpha$

Tra bảng phân phối Student bậc (n-1) là bảng 4 tìm được t_{α}^{n-1} sao cho $P\left(|T| < t_{\alpha}^{n-1}\right) = 1 - \alpha$. Khi đó:

$$P\left(\frac{\left|\overline{X} - m\right|\sqrt{n}}{s} < t_{\alpha}^{(n-1)}\right) = 1 - \alpha$$

$$\Leftrightarrow P\left(\overline{X} - t_{\alpha}^{(n-1)}\frac{s}{\sqrt{n}} < m < \overline{X} + t_{\alpha}^{(n-1)}\frac{s}{\sqrt{n}}\right) = 1 - \alpha$$

Ta có khoảng ước lượng của m là

$$\overline{X} - t_{\alpha}^{(n-1)}\frac{s}{\sqrt{n}} < m < \overline{X} + t_{\alpha}^{(n-1)}\frac{s}{\sqrt{n}}$$

Đặt $\varepsilon = t_{\alpha}^{(n-1)}\dfrac{s}{\sqrt{n}}$ thì $\overline{X} - \varepsilon < m < \overline{X} + \varepsilon$

Qui tắc thực hành

Từ mẫu cỡ n cụ thể: $(x_1, x_2, \ldots, x_n)$ tính $\overline{x}$ và s.

Khoảng ước lượng của m là: $(\overline{x} - \varepsilon < m < \overline{x} + \varepsilon)$

với $\varepsilon = t_{\alpha}^{(n-1)}\dfrac{s}{\sqrt{n}}$

Từ độ tin cậy $1 - \alpha$ tra bảng phân vị Student với (n - 1) bậc tự do. (bảng 4) ta có t_{α}^{n-1}.

Ví dụ 4 Đo đường kính của 20 trục máy do một máy tiện tự động sản xuất ra, ta được kết quả sau:

250	249	251	253	248	250	250	252	257	245
248	247	249	250	280	250	247	253	256	249

Giả sử đường kính của các trục máy là đại lượng ngẫu nhiên có phân phối chuẩn. Hãy ước lượng đường kính trung bình của các trục máy do máy tiện ra với độ tin cậy 95%.

BÀI GIẢI

Gọi X là đường kính của các trục máy.

m là đường kính trung bình của các trục máy.

Ta tìm khoảng ước lượng của m.

Trường hợp này kích thước mẫu n = 20 và chưa biết σ^2.

Vậy khoảng ước lượng của m là:

$$\left(\overline{x} - \varepsilon < m < \overline{x} + \varepsilon\right) \text{ với } \varepsilon = t_{\alpha}^{(n-1)} \frac{s}{\sqrt{n}}$$

Với độ tin cậy 1 - α = 95% cho trước tra bảng phân vị student với (n - 1) bậc tự do (bảng 4): t_{α}^{n-1} = 2,093

Từ bảng số liệu, ta có n = 20; $\sum_{i=1}^{k} n_i x_i = 5034$; $\sum_{i=1}^{k} n_i x^2{}_i = 1268062$

Ta tính được: $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} n_i x_i$ = 251,7 (mm) và

$s^2 = \frac{1}{n}\sum_{i=1}^{n} n_i x_i^2 - \left(\overline{x}\right)^2$ = 52,85; s=7,27

Từ đó ta có: $\varepsilon = t_{\alpha}^{(n-1)} \frac{s}{\sqrt{n}} = 2{,}093\frac{7{,}27}{\sqrt{20}} = 3{,}4$.

Vậy với độ tin cậy 95% khoảng ước lượng của m là:

(251,7 – 3,4 < m < 251,7+3,4) hay (248,3mm ; 255,1mm)

III. Ước lượng tỷ lệ tổng thể P

Giả sử tổng thể có 2 loại phần tử, tỷ lệ những phần tử có tính chất A là p (p chưa biết). Dựa vào mẫu ngẫu nhiên *($X_1, X_2, \ldots, X_n$)*, cụ thể là dựa vào tỷ lệ mẫu f, hãy tìm 2 số p_1, p_2 sao cho $P(p_1 < p < p_2) = 1 - \alpha$ khá lớn cho trước.

Khi n khá lớn (n ≥ 30). Ta có:

$$P\left(|F - p| < \varepsilon\right) = P\left(\frac{|F - p|}{\sqrt{F(1-F)}}\sqrt{n} < \varepsilon\frac{\sqrt{n}}{\sqrt{F(1-F)}}\right) = 1 - \alpha .$$

Đặt $t = \dfrac{\varepsilon\sqrt{n}}{\sqrt{F(1-F)}}$, theo quy tắc k-$\sigma$:

$$P\left(\frac{|F - p|}{\sqrt{F(1-F)}}\sqrt{n} < t\right) = 2\Phi(t) = 1 - \alpha$$

Với độ tin cậy 1 - α cho trước, tra bảng 3 ta tìm được t_α .

Từ đó ta có $\varepsilon = t_\alpha \dfrac{\sqrt{F(1-F)}}{\sqrt{n}}$.

Vậy khoảng ước lượng của p là $f - \varepsilon < p < f + \varepsilon$.

Qui tắc thực hành: Từ mẫu cụ thể tính f là tỷ lệ phần tử có tính chất A trong mẫu cỡ n. Ta có khoảng ước lượng của p là:

$(f - \varepsilon < p < f + \varepsilon)$ với $\varepsilon = t_\alpha\sqrt{\dfrac{f(1-f)}{n}}$

Ứng với độ tin cậy 1 - α tra bảng phân vị chuẩn (tra bảng 3) tìm t_α

Ví dụ 5 Một nhà máy sản xuất hàng loạt, người ta kiểm tra một số mẫu. Kết quả cho ở bảng sau

Kích thước mẫu	Số sản phẩm loại A	Kích thước mẫu	Số sản phẩm loại A
60	14	80	18
50	03	80	20
50	05	40	17
60	09	300	32
80	10		

Tìm ước lượng cho tỷ lệ sản phẩm loại A với độ tin cậy 95%.

BÀI GIẢI

Đây là bài toán ước lượng khoảng cho tỷ lệ đám đông.

Gọi tỷ lệ sản phẩm loại A là p. Ta ước lượng p với độ tin cậy 95%.

Ta có khoảng ước lượng của p là: $(f - \varepsilon < p < f + \varepsilon)$

với $\varepsilon = t_\alpha \sqrt{\dfrac{f(1-f)}{n}}$

Từ bảng số liệu đã cho, ta có tỷ lệ sản phẩm loại A của mẫu là:

$$f = \frac{128}{800} = 0{,}16$$

Ứng với độ tin cậy $1 - \alpha = 95\%$ thì tra bảng phân vị chuẩn (bảng 3) có $t_\alpha = 1{,}96$.

Vậy $\varepsilon = 1{,}96\sqrt{\dfrac{0{,}16(1-0{,}16)}{800}} = 0{,}0254$

Khoảng ước lượng của p là:

$(0{,}16 - 0{,}0254 < p < 0{,}16 + 0{,}0254)$ hay $(0{,}1346 < p < 0{,}1854)$

IV. Các bài toán kéo theo

1. Tìm độ tin cậy $1-\alpha$.

Giả sử có mẫu cụ thể kích thước mẫu là n và yêu cầu đảm bảo độ chính xác ε thì độ tin cậy đạt được là bao nhiêu %?

* Trong bài toán ước lượng trung bình m= E(X).

a) Trường hợp σ^2 đã biết:

Ta có $\varepsilon = t_\alpha \dfrac{\sigma}{\sqrt{n}}$ suy ra $t_\alpha = \dfrac{\varepsilon\sqrt{n}}{\sigma}$

Tra bảng 1 ta có $\Phi(t_\alpha) = \Phi\left(\dfrac{\varepsilon\sqrt{n}}{\sigma}\right)$. Từ đó suy ra $1-\alpha = 2\Phi(t_\alpha)$.

(Hoặc tra ngược từ bảng 3)

b) Trường hợp $n \geq 30$, σ^2 chưa biết:

Ta có $\varepsilon = t_\alpha \dfrac{s}{\sqrt{n}}$ Suy ra $t_\alpha = \dfrac{\varepsilon\sqrt{n}}{s}$

Tra bảng 1 ta có $\Phi(t_\alpha) = \Phi\left(\dfrac{\varepsilon\sqrt{n}}{s}\right)$. Từ đó suy ra $1-\alpha = 2\Phi(t_\alpha)$.

c) Trường hợp $n < 30$, σ^2 chưa biết:

Ta có $\varepsilon = t_\alpha^{(n-1)} \dfrac{s}{\sqrt{n}}$ suy ra $t_\alpha^{(n-1)} = \dfrac{\varepsilon\sqrt{n}}{s}$

Tra bảng 4 với (n-1) bậc tự do ta tìm được $1-\alpha$.

Ví dụ 6 Điểm trung bình môn Xác suất thống kê của 100 sinh viên trường Cao đẳng công nghệ thông tin là 5,13 với phương sai mẫu hiệu chỉnh là 1,21.

a) Ước lượng điểm trung bình môn Xác suất thống kê của các sinh viên với độ tin cậy 95%.

b) Với độ chính xác 0,25 điểm thì đảm bảo độ tin cậy là bao nhiêu?

BÀI GIẢI

a) Gọi m là điểm trung bình môn Xác suất thống kê của sinh viên trường Cao đẳng công nghệ thông tin.

$n = 100 > 30$, σ^2 chưa biết , $\overline{x} = 5,13$; $s = \sqrt{s^2} = 1,1$

$1-\alpha = 95\%$. Tra bảng 3 ta có t_α =1,96

Khoảng ước lượng của m là

$$\overline{x} - t_\alpha \frac{s}{\sqrt{n}} < m < \overline{x} + t_\alpha \frac{s}{\sqrt{n}}$$

$$\Leftrightarrow \quad 5,13 - 1,96.\frac{1,1}{\sqrt{100}} < m < 5,13 + 1,96.\frac{1,1}{\sqrt{100}}$$

$$\Leftrightarrow \quad 4,9144 < m < 5,3456$$

Vậy điểm trung bình môn Xác suất thống kê của sinh viên trường Cao đẳng công nghệ thông tin nằm trong khoảng:
(4,9144 điểm; 5,3456 điểm) với độ tin cậy 95%.

b) Do n > 30, σ^2 chưa biết nên $t_\alpha = \dfrac{\varepsilon\sqrt{n}}{s}$.

Theo đề bài $\varepsilon = 0,25$ nên $t_\alpha = \dfrac{0,25\sqrt{100}}{1,1} = 2,27$

Từ đó suy ra

$1-\alpha = 2\Phi(t_\alpha) = 2\Phi(2,27) = 2.0,4884 = 0,9768 = 97,68\%$.

Vậy với mẫu trên với độ chính xác 0,25 thì độ tin cậy đạt được là 97,68%.

* Trường hợp ước lượng tỷ lệ tổng thể.

Ta có: $\varepsilon = t_\alpha \dfrac{\sqrt{f(1-f)}}{\sqrt{n}}$ nên

$$t_\alpha = \frac{\varepsilon\sqrt{n}}{\sqrt{f(1-f)}} \Rightarrow \Phi(t_\alpha) = \Phi\left(\frac{\varepsilon\sqrt{n}}{\sqrt{f(1-f)}}\right) \Rightarrow 1-\alpha = 2\Phi(t_\alpha)$$

Ví dụ 7 Trong 500 sinh viên thi môn Xác suất thống kê thì thấy có 150 sinh viên đạt trên 7 điểm. Với độ chính xác 3%, với mẫu trên thì đảm bảo độ tin cậy là bao nhiêu?

BÀI GIẢI

Đây là bài toán xác định độ tin cậy trong bài toán ước lượng tỷ lệ, ở đây:

$$n = 500;\ f = 0,3;\ \varepsilon = 3\% = 0,03$$

$$t_\alpha = \frac{\varepsilon\sqrt{n}}{\sqrt{f(1-f)}} = \frac{0,03.\sqrt{500}}{\sqrt{0,3(1-0,3)}} = 1,46 \Rightarrow \Phi(t_\alpha) = \Phi(1,46) = 0,4279$$

$$\Rightarrow 1-\alpha = 2\Phi(t_\alpha) = 0,8558 = 85,58\%$$

Vậy với độ chính xác 3% thì đảm bảo độ tin cậy là 85,58%.

2. Xác định kích thước mẫu

Chất lượng của ước lượng được phản ánh qua độ tin cậy (1-α) và độ chính xác (sai số) ε, kích thước mẫu càng lớn thì độ tin cậy lớn và độ chính xác cao. Nhưng ta muốn độ tin cậy và độ chính xác đạt được ở một mức nào đó cho trước thì cần kích thước mẫu tối thiểu là bao nhiêu?

* Xác định kích thước mẫu (tối thiểu) trong trường hợp ước lượng trung bình của tổng thể.

$$\varepsilon \geq t_\alpha \frac{\sigma}{\sqrt{n}} \Rightarrow \sqrt{n} \geq t_\alpha \frac{\sigma}{\varepsilon} \Rightarrow n \geq t_\alpha^2 \frac{\sigma^2}{\varepsilon^2}$$

Vậy n tối thiểu sẽ là $n = \left[t_\alpha^2 \frac{\sigma^2}{\varepsilon^2} \right] + 1$

(trong đó $\left[t_\alpha^2 \frac{\sigma^2}{\varepsilon^2} \right]$ là phần nguyên của $t_\alpha^2 \frac{\sigma^2}{\varepsilon^2}$).

Trong trường hợp σ^2 chưa biết thì ta có thể lấy mẫu đã cho (nếu chưa có mẫu thì ta có thể lấy mẫu lần đầu kích thước $n_1 \geq 30$ để tính s^2, rồi thay $\sigma^2 = s^2$). Nên khi đó kích thước của mẫu (tối thiểu) sẽ là:

$$n = \left[t_\alpha^2 \frac{s^2}{\varepsilon^2} \right] + 1$$

* Xác định kích thước mẫu (tối thiểu) trong trường hợp ước lượng tỷ lệ của tổng thể.

Ta có

$$\varepsilon \geq t_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}} \Rightarrow \sqrt{n} \geq t_\alpha \frac{\sqrt{p(1-p)}}{\varepsilon} \Rightarrow n \geq t_\alpha^2 \frac{p(1-p)}{\varepsilon^2}$$

p chưa biết, căn cứ vào mẫu đã cho (nếu chưa có mẫu thì có thể lấy mẫu ban đầu để tính f).

Ta thay p(1-p)=f (1-f) , nên $n \geq t_\alpha^2 \frac{f(1-f)}{\varepsilon^2}$

Vậy n tối thiểu sẽ là $n = \left[t_\alpha^2 \frac{f(1-f)}{\varepsilon^2} \right] + 1$

Và nếu $n > n_1$ thì ta phải điều tra thêm $n_2 = n - n_1$ để kích thước mẫu $n = n_1 + n_2$ để đảm bảo độ tin cậy và độ chính xác cho trước (nếu $n \leq n_1$ thì ta thường không phải điều tra thêm).

Ví dụ 8 Đem cân thử 100 trái xoài vừa thu hoạch ở khu vườn, ta có trung bình mẫu 400g, phương sai mẫu 100 và thấy có 65 trái đạt tiêu chuẩn.

a) Ước lượng trọng lượng trung bình của một trái xoài ở khu vườn này với độ tin cậy 96%?

b) Với mẫu đã cho với độ tin cậy 96%, độ chính xác 2g thì phải điều tra tối thiểu bao nhiêu trái xoài?

c) Với mẫu đã cho với độ tin cậy 98%, độ chính xác 4% thì phải điều tra thêm ít nhất bao nhiêu trái xoài nữa?

BÀI GIẢI

a) Gọi m là trọng lượng trung bình của một trái xoài ở khu vườn này.

$n = 100 > 30$, σ^2 chưa biết , $\overline{X} = 400, s = \sqrt{s^2} = 10$

$1 - \alpha = 96\% \Rightarrow t_\alpha = 2{,}0537$ (tra bảng 1)

Khoảng ước lượng của m là

$$\overline{x} - t_\alpha \frac{s}{\sqrt{n}} < m < \overline{x} + t_\alpha \frac{s}{\sqrt{n}}$$

$$\Leftrightarrow 400 - 2{,}0537.\frac{10}{\sqrt{100}} < m < 400 + 2{,}0537.\frac{10}{\sqrt{100}}$$

$$\Leftrightarrow 397{,}9463 < m < 402{,}0537$$

Vậy trọng lượng trung bình của mỗi trái xoài của vườn nằm trong khoảng: (397,9463; 402,0537g) với độ tin cậy 96%.

b) Đây là bài toán xác định kích thước mẫu tối thiểu trong trường hợp ước lượng trung bình tổng thể.

$$n = \left[t_\alpha^2 \frac{s^2}{\varepsilon^2} \right] + 1 = \left[2{,}0537^2 \frac{100}{2^2} \right] + 1 = [105{,}4421] + 1 = 106$$

c) Đây là bài toán xác định kích thước mẫu tối thiểu trong trường hợp ước lượng tỷ lệ tổng thể.

$$n = \left[t_\alpha^2 \frac{f(1-f)}{\varepsilon^2} \right] + 1$$

Ở đây $f = \frac{65}{100} = 0,65; \varepsilon = 2\% = 0,02$;

$1 - \alpha = 98\% = 0,98 \Rightarrow t_\alpha = 2,3263$

$$n = \left[t_\alpha^2 \frac{f(1-f)}{\varepsilon^2} \right] + 1 = \left[2,3263^2 \frac{0,65.(1-0,65)}{(0,04)^2} \right] + 1$$

$$= \left[769,4721 \right] + 1 = 770$$

Ta đã điều tra 100 sản phẩm, nên phải điều tra thêm

$n_2 = n - n_1 = 770 - 100 = 670$ (sản phẩm nữa)

Ví dụ 9 Lô trái cây của một chủ hàng được đóng thành từng sọt, mỗi sọt 100 trái. Khi kiểm tra 50 sọt thấy có 450 trái không đạt tiêu chuẩn.

a) Ước lượng tỷ lệ trái không đạt tiêu chuẩn của lô hàng với độ tin cậy 95%.

b) Muốn ước lượng tỷ lệ trái không đạt tiêu chuẩn với độ chính xác 0,5% thì độ tin cậy đạt được là bao nhiêu?

c) Muốn ước lượng tỷ lệ trái không đạt tiêu chuẩn với độ tin cậy 99% và độ chính xác 1% thì cần kiểm tra bao nhiêu sọt?

d) Muốn ước lượng tỷ lệ trái không đạt tiêu chuẩn với độ tin cậy 99,73% thì độ chính xác đạt được là bao nhiêu?

BÀI GIẢI

a) Gọi p là tỷ lệ trái không đạt tiêu chuẩn, ta cần ước lượng p

Ứng với độ tin cậy $1 - \alpha = 95\%$ tra bảng 3 có $t_\alpha = 1,96$

Tỷ lệ trái không đạt tiêu chuẩn của mẫu là

$$f = \frac{450}{5000} = 0,09$$

Độ chính xác

$$\varepsilon = t_\alpha \sqrt{\frac{f(1-f)}{n}} = 1,96.\sqrt{\frac{0,09.(1-0,09)}{5000}} = 0,008$$

Khoảng ước lượng của p là $(f - \varepsilon < p < f + \varepsilon)$

(0,09 - 0,008 < p < 0,09 + 0,008) hay (8,2% ; 9,8%)

b) Ta cần xác định độ tin cậy (1 - α) với độ chính xác ε = 0,5% khi ước lượng p.

Muốn vậy, ta cần xác định t_α, từ đó suy ra 1 - α.

Từ công thức $\varepsilon = t_\alpha\sqrt{\dfrac{f(1-f)}{n}}$ ta suy ra

$$u_\gamma = \varepsilon.\sqrt{\frac{n}{f(1-f)}} = 0{,}005.\sqrt{\frac{5000}{0{,}09(1-0{,}09)}} = 1{,}2354$$

Tra bảng 3 (tra ngược) ta được: 1 - α ≈ 0,78 hay 78%.

c) Ta cần xác định kích thước mẫu n sao cho thỏa mãn độ chính xác là 1% và độ tin cậy 99% khi ước lượng p. Ta có:

$$n = \left[(t_\alpha)^2\frac{f(1-f)}{\varepsilon^2}\right] = \left[(2{,}58)^2\frac{0{,}09(1-0{,}09)}{(0{,}01)^2}\right] = 5452$$

Vì mỗi sọt có 100 trái nên ta cần kiểm tra 55 sọt.

d) Ta cần xác định độ chính xác với độ tin cậy 99,73% (ứng với t_α =3) và kích thước mẫu n = 5000. Ta có:

$$\varepsilon = t_\alpha\sqrt{\frac{f(1-f)}{n}} = 3.\sqrt{\frac{0{,}09.(1-0{,}09)}{5000}} = 0{,}012$$

Tức độ chính xác đạt được 1,2%.

V. Ước lượng phương sai σ^2 khi X tuân theo phân phối chuẩn

1. Trường hợp 1 Ước lượng phương sai D(X) = σ^2 với E(X) = m đã biết.

Ta lập bảng tính toán để tính $\Sigma n_i (x_i - m)^2$

Ứng với độ tin cậy 1 - α tra (bảng 5) $\chi^2_{\alpha_1}$, $\chi^2_{1-\alpha_2}$ là các phân vị χ^2 với n bậc tự do.Khoảng ước lượng của σ^2 là:

$$\frac{\sum_{i=1}^{n}(x_i-m)^2}{\chi^2_{1-\alpha 2}} < \sigma^2 < \frac{\sum_{i=1}^{n}(x_i-m)^2}{\chi^2_{\alpha 1}}$$

Ví dụ 10 Mức hao phí nhiên liệu cho 1 đơn vị sản phẩm là đại lượng ngẫu nhiên X phân phối theo quy luật chuẩn. Quan sát 28 sản phẩm ta thu được kết quả sau

Lượng nguyên liệu hao phí (g)	19	19,5	20,0	20,5
Số sản phẩm	5	6	14	3

Với độ tin cậy 90%, hãy ước lượng phương sai của X khi E(X) = 20g (cho $\alpha_1 = \alpha_2 = 0,05$)?

BÀI GIẢI

Ta ước lượng phương sai $D(X) = \sigma^2$ trong trường hợp đã biết E(X).

Để tính $\Sigma(x_i - m)^2$ ta lập bảng tính toán tìm được:

$\Sigma n_i(x_i - 20)^2 = 7.25$

Ứng với độ tin cậy $1 - \alpha = 90\%$ mà $\alpha_1 = \alpha_2 = 0,05$

Tra (bảng 5) phân vị $\chi 2$ với bậc tự do n = 28 ta được:

$\chi^2_{\alpha 1} = \chi^2_{0,05} = 16,9; \chi^2_{1-\alpha 2} = \chi^2_{0,95} = 41,3$

Vậy với độ tin cậy 90%, khoảng ước lượng của σ^2 là:

$$\left[\frac{7,25}{41,3} < \sigma^2 < \frac{7,25}{16,9}\right]$$

hay $(0,1755 < \sigma^2 < 0,429)$

2. Trường hợp 2 Ước lượng phương sai $D(X) = \sigma^2$ trong trường hợp chưa biết kỳ vọng E(X)= m.

Trường hợp này khoảng ước lượng của σ^2 sẽ là:

$$\left[\frac{(n-1)s^2}{\chi^2_{1-\alpha 2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{\alpha 1}}\right]$$

Trong đó $\chi^2_{\alpha_1}$, $\chi^2_{1-\alpha_2}$ là các phân vị χ^2 với n - 1 bậc tự do.

Ví dụ 11 Theo dõi số lượng gạo bán được trong 1 ngày ở cửa hàng, ta có kết quả ghi ở bảng sau

Số hàng bán được (kg/ngày)	Số ngày
1900 – 1950	2
1950 – 2000	10
2000 – 2050	8
2050 – 2100	5

Hãy ước lượng phương sai của lượng hàng bán được mỗi ngày với độ tin cậy 95% (cho $\alpha_1 = \alpha_2 = 0{,}05$)?

BÀI GIẢI

Đây là bài toán ước lượng phương sai chưa biết E(X).

Trường hợp này khoảng ước lượng của σ^2 sẽ là:

$$\left[\frac{(n-1)s^2}{\chi^2_{1-\alpha 2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{\alpha 1}}\right]$$

Trong đó $\chi^2_{\alpha_1}$, $\chi^2_{1-\alpha_2}$ là các phân vị χ^2 với n - 1 = 24 bậc tự do.

Tra bảng 5, ta được: $\chi^2_{\alpha_1} = \chi^2_{0,05} = 13{,}8;\ \chi^2_{1-\alpha_2} = \chi^2_{0,95} = 36{,}4$

Để tính s^2 ta lập bảng tính sau đây

x_i	x_{i+1}	n_i	x_{io}	$x_{io}n_i$	$n_i x_{io} x_{io}$
1900	1950	2	1925	3850	7411250
1950	2000	10	1975	19750	39006250
2000	2050	8	2025	16200	32805000
2050	2100	5	2075	10375	21528125
Tổng		n=25		50175	100750625

Từ kết quả tính toán ở bảng trên ta có: $s^2 = 2058{,}338$

Vậy khoảng ước lượng của σ^2 là: ($1413{,}6994 < \sigma^2 < 3728{,}8732$).

BÀI TẬP CHƯƠNG IV

4.1. 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 cho trọng lượng trung bình của trái cây với độ tin cậy 0,99. Giả sử trọng lượng của trái cây tuân theo phân phối chuẩn.

b) Nếu muốn sai số ước lượng cho trọng lượng trung bình của trái cây không quá ε =2 g ở độ tin cậy 99% thì phải quan sát ít nhất bao nhiêu trái

c) Trái cây có trọng lượng lớn hơn 230g được xếp vào loại A. Nếu muốn sai số ước lượng cho tỷ lệ p của trái cây loại A không quá 0,04 ở độ tin cậy 0.99 thì phải tỷ quan sát ít nhất mấy trường hợp ?

Đáp số:

a) Từ số liệu của mẫu, ta có: $n = 82; \overline{x} = 225,854; s_X = 13,259$.

Khoảng ước lượng cho trung bình tổng thể khi chưa biết phương sai tổng thể là $\left(222,0822; 229,626\right)$.

b) Cần phải quan sát ít nhất 292 trái.

c) Phải quan sát ít nhất 1001 trường hợp.

4.2. 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 $\overline{x}$ và phương sai mẫu hiệu chỉnh s_X^2.

b) Ước lượng trung bình của tổng thể ở độ tin cậy 0,95.

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

Đáp số: a) Từ các số liệu nhận được của mẫu, ta có:

$$n = 12 \ ; \overline{x} = 137,83 \ ; \ s_X^2 = 19,42 \text{, và } S_X = 4,41 .$$

b) Khoảng ước lượng cho trung bình là $(135,01;140,63)$.

c) Vậy phải quan sát ít nhất 75 người.

4.3. Một công ty tiến hành khảo sát nhu cầu tiêu dùng về 1 loại sản phẩm do công ty sản xuất. Khảo sát trên 500 hộ gia đình ở 1 thành phố ta được bảng số liệu:

Số lượng (kg/tháng)	0-2	2-3	3-4	4-5	5-6	6-7	7-8
Số hộ	150	33	52	127	73	35	30

a) Hãy ước lượng số sản phẩm của công ty được tiêu thụ tại thành phố trong 1 tháng, với độ tin cậy 94%. Cho biết tổng số hộ gia đình trong toàn thành phố là 500.000 hộ.

b) Hãy ước lượng mức tiêu thụ trung bình trên mỗi hộ ở các hộ có nhu cầu sử dụng, với độ tin cậy 95%.

c) Những hộ có mức tiêu thụ > 5 kg/tháng gọi là những hộ có nhu cầu sử dụng cao. Nếu muốn ước lượng tỷ lệ những hộ có nhu cầu sử dụng cao với độ chính xác 4% và độ tin cậy 98% thì phải điều tra thêm bao nhiêu hộ nữa.

4.4. Muốn khảo sát tỷ lệ học sinh khá giỏi trong một tổng thể, ta lấy mẫu 300 học sinh thì thấy có 50 học sinh không khá giỏi.

a) Hãy ước lượng khoảng cho tỷ lệ học sinh khá giỏi với độ tin cậy 90% và hãy nêu ý nghĩa thực tiễn của số liệu tìm được.

b) Nếu dùng mẫu này để ước lượng tỷ lệ học sinh khá giỏi có sai số không quá 5% thì độ tin cậy là bao nhiêu ?

4.5. Mức hao phí nhiên liệu cho 1 đơn vị sản phẩm là đại lượng ngẫu nhiên X phân phối theo quy luật chuẩn. Quan sát 28 sản phẩm ta thu được kết quả sau:

Lượng nguyên liệu hao phí (gr)	19	19,5	20,0	20,5
Số sản phẩm	5	6	14	3

Với độ tin cậy 90%, hãy ước lượng phương sai của X (cho $\alpha_1 = \alpha_2 = 0,05$)? **Đáp số**: $(0,143 < \sigma^2 < 0,354)$

CHƯƠNG V

KIỂM ĐỊNH GIẢ THIẾT THỐNG KÊ

5.1. CÁC KHÁI NIỆM

I. Đặt bài toán

Giả thiết thống kê là những giả thiết nói về các tham số, dạng quy luật phân phối, hoặc tính độc lập của các đại lượng ngẫu nhiên.

Đám đông có đặc trưng θ chưa biết. Ta cần kiểm định giả thiết

H_0: $\theta = \theta_0$. Hãy đưa ra một quy tắc căn cứ vào mẫu điều tra

$(X_1, \ldots, X_n)$ mà chấp nhận hay bác bỏ giả thiết trên.

Trong thống kê toán, việc bác bỏ một giả thiết dựa vào xác suất xảy ra biến cố có liên đến giả thiết đó. Một giả thiết chỉ có thể xảy ra với xác suất rất nhỏ thì trên thực tế giả thiết đó hầu như không đúng, nên ta bác bỏ giả thiết đó.

Dựa vào các lí luận trên, khi đặt giả thiết thống kê ta lưu ý một số vấn đề sau:

- Giả thiết đặt ra với ý đồ bác bỏ nó, nghĩa là giả thiết đặt ra ngược lại với điều ta muốn chứng minh, muốn thuyết phục. Vì vậy khi bác bỏ được giả thiết có nghĩa là ta đã chứng minh điều ngược lại.
- Giả thiết đặt ra sao cho khi chấp nhận hoặc bác bỏ nó sẽ có tác dụng trả lời được câu hỏi mà bài toán thực tế đặt ra.
- Giả thiết đặt ra 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.

Khi đặt giả thiết ta thường so sánh cái chưa biết với cái đã biết. "Cái đã biết" mà ta nói ở đây thường là những thông tin quá khứ, cái định mức kinh tế, kỹ thuật.

Giả thiết đặt ra như vậy gọi là giả thiết cần kiểm định. Giả thiết cần kiểm định còn được gọi là giả thiết không, ký hiệu là H_0 (hoặc H). Một mệnh đề đối lập với H_0 được gọi là giả thiết đối và được ký hiệu là H_1 (hoặc $\overline{H}$).

Chẳng hạn: H_0: $\theta = \theta_0$; $H_1 : \theta \neq \theta_0$

(θ là một tham số nào đó của đại lượng ngẫu nhiên ta đang nghiên cứu; θ_0 là giá trị đã biết).

Nếu kiểm định giả thiết với giả thiết đối có dạng này được gọi là kiểm định giả thiết hai phía (vì miền bác bỏ nằm về hai phía của miền chấp nhận).

Giả thiết đối dạng $\theta \neq \theta_0$ thường được áp dụng khi ta chưa biết rõ trong thực tế $\theta > \theta_0$ hay $\theta < \theta_0$.

- Nếu kiểm định giả thiết với giả thiết đối có dạng: H_1: $\theta > \theta_0$ hoặc H_1: $\theta < \theta_0$; thì được gọi là kiểm định giả thiết một phía (vì miền bác bỏ nằm về một phía của miền chấp nhận).

Nếu giả thiết đối có dạng H_1: $\theta > \theta_0$; thì được gọi là kiểm định giả thiết về phía bên phải (vì miền bác bỏ nằm về phía bên phải của miền chấp nhận).

- Nếu giả thiết đối có dạng H_1: $\theta < \theta_0$; thì được gọi là kiểm định giả thiết về phía bên trái (vì miền bác bỏ nằm về phía bên trái của miền chấp nhận).

Nhiệm vụ của lý thuyết kiểm định giả thiết thống kê là bằng thực nghiệm (thông qua mẫu cụ thể) kiểm tra tính đúng (sai) của giả thiết H_0.

II. Mức ý nghĩa α và miền bác bỏ

Có thể mô tả phương pháp kiểm định giả thiết thống kê như sau: Xuất phát từ yêu cầu của bài toán thực tế, ta nêu ra một giả thiết H_0 và giả thiết đối của nó.

Giả sử rằng H_0 đúng, từ đó tìm một biến cố có xác suất đủ bé để có thể tin rằng biến cố đó hầu như không thể xảy ra trong một phép thử. Muốn vậy, từ mẫu ngẫu nhiên:

$$W_x=(X_1,X_2,\ldots,X_n)$$

ta chọn: $Z=f(X_1,X_2,\ldots,X_n, \theta_0)$

Z được chọn sao cho: nếu H_0 đúng thì ta sẽ xác định được quy luật phân phối xác suất của Z và với mẫu cụ thể ta có thể tính được giá trị của Z. Đại lượng ngẫu nhiên Z được gọi là tiêu chuẩn kiểm định giả thiết H_0.

Do quy luật phân phối xác suất của Z đã biết, nên với α tùy ý ta có thể tìm được được miền W sao cho $P(Z \in W_\alpha) = \alpha$. Miền W được gọi là miền bác bỏ giả thiết H_0. Trong thực tế thường ta chọn α trong khoảng (1%;5%), được gọi là mức ý nghĩa của kiểm định.

Thực hiện một phép thử đối với mẫu ngẫu nhiên W_x, ta thu được mẫu cụ thể $W_x=(x_1,x_2,\ldots,x_n)$. Từ mẫu cụ thể này ta tính được giá trị của Z (kí hiệu z) và gọi là giá trị thực nghiệm;

$Z=f(x_1,x_2,\ldots,x_n, \theta_0)$

Nếu $z \in W_\alpha$ thì ta bác bỏ giả thiết H_0 thừa nhận H_1.

Nếu $z \notin W_\alpha$ thì ta chấp nhận H_0.

Chú ý: Khi nói "chấp nhận H_0" điều đó không có nghĩa là giả thiết H_0 là đúng mà chỉ có nghĩa là với số liệu của mẫu ta chưa đủ cơ sở (chưa đủ bằng chứng) để bác bỏ H_0. Trong thực hành tốt hơn là nên nói rằng: "có thể chấp nhận H_0" hoặc "chưa có cơ sở để bác bỏ H_0".

III. Sai lầm loại I và sai lầm loại II

Giả thiết liên quan đến toàn bộ đám đông. Nhưng ta lại chỉ căn cứ vào một mẫu để kết luận, nên có thể có 4 khả năng xảy ra:

Trường hợp 1: H_0 đúng($\theta = \theta_0$) nhưng qua kiểm tra ta kết luận giả thiết sai, tức $\theta \neq \theta_0$.

Trường hợp 2: H_0 đúng và qua kiểm tra cũng kết luận giả thiết đúng.

Trường hợp 3: H_0 sai và qua kiểm tra, ta kết luận H_0 đúng.

Trường hợp 4: H_0 sai và qua kiểm tra ta cũng kết luận H_0 sai.

Trường hợp 2 và 4 ta đã có quyết định đúng còn trường hợp 1 và 3 ta đã mắc sai lầm. Người ta quy ước gọi sai lầm ở trường hợp 1 là sai lầm loại I, còn sai lầm ở trường hợp 3 là sai lầm loại II.

Hai loại sai lầm này có tính đối kháng tức là muốn hạn chế khả năng phạm sai lầm loại I, ta có xu hướng làm tăng khả năng phạm sai lầm II và ngược lại. Vì muốn hạn chế sai lầm loại I, ta có xu hướng dè dặt trong việc bác bỏ và sẽ có khuynh hướng dễ dãi trong việc chấp nhận. Khi đó lại dễ phạm sai lầm loại II. Còn muốn giảm sai lầm loại II, ta dè dặt trong việc chấp nhận và lại dẫn đến dễ dãi trong bác bỏ. Điều này làm cho nguy cơ phạm sai lầm loại I tăng lên. Tức là:

$$P(\text{sai loại I}) = P(\text{loại } H_0/H_0 \text{ đúng})\downarrow$$
$$\Rightarrow P(\text{sai loại II})=P(\text{nhận } H_0/H_0 \text{ sai})\uparrow;$$

$$P(\text{sai loại II})\downarrow \Rightarrow P(\text{sai loại I})\uparrow$$

Có hai cách khống chế khả năng mắc phải sai lầm:

Cách thứ nhất: Ta ấn định trước mức mắc phải sai lầm loại I và sai lầm loại II rồi tính toán tìm một mẫu có kích thước nhỏ nhất ứng với hai mức sai lầm này.

Cách thứ hai: Ta ấn định trước xác suất sai lầm loại I (tức cho trước mức ý nghĩa α) chọn miền bác bỏ W_α có xác suất sai lầm loại II nhỏ nhất.

Trong giáo trình này ta chọn cách thứ 2.

5.2. KIỂM ĐỊNH GIẢ THIẾT CÓ THAM SỐ

I. Kiểm định giả thiết về tỷ lệ đám đông

1. Bài toán

Giả sử đám đông chia 2 loại phần tử: phần tử có tính chất A và phần tử không có tính chất A. Tỷ lệ phần tử có tính chất A là p chưa biết. Ta cần kiểm tra giả thiết

$H_0 : P = P_0$ và $H_1 : P \neq P_0$

Hãy đưa ra quy tắc căn cứ vào mẫu gồm n quan sát $(x_1, \ldots, x_n)$ mà chấp nhận hay bác bỏ giả thiết trên với mức ý nghĩa α.

Để kiểm định giả thiết trên, ta lấy mẫu cỡ n khá lớn khi đó nếu H_0 đúng thì đại lượng ngẫu nhiên

$Z = \dfrac{(F - P_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}}$ phân phối xấp xỉ N(0,1)

2. Quy tắc thực hành

Ta đặt giả thiết $H_0 : P = P_0$ và $H_1 : P \neq P_0$

+ Từ mẫu cụ thể $(x_1, \ldots, x_n)$ ta tính f là tỷ lệ mẫu và tính kiểm định

$$t = \frac{(f - p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}}$$

+ Từ α đã biết tra bảng phân vị chuẩn (tra bảng 3) tìm t_α

Kết luận:

+ Nếu $|t| \le t_\alpha$ ta chấp nhận giả thiết H_0, coi $p = p_0$.

+ Nếu $|t| > t_\alpha$ ta bác bỏ giả thiết H_0, coi $p \ne p_0$.

Và khi đó

+ Nếu $f > p_0$ ta xem $p > p_0$

+ Nếu $f < p_0$ ta xem $p < p_0$

Ví dụ 1 Một máy sản xuất hàng loạt sản phẩm, tỷ lệ sản phẩm loại A lúc đầu là 0,155. Sau khi áp dụng phương pháp sản xuất mới, kiểm tra 1000 sản phẩm ta thấy tỷ lệ loại A trong mẫu này là 0,2. Cho kết luận về phương pháp sản xuất mới này với $\alpha = 1\%$.

BÀI GIẢI

Tỷ lệ loại A lúc đầu là $p_0 = 0{,}155$

Tỷ lệ loại A sau khi áp dụng phương pháp mới là p chưa biết.

Ta đặt giả thiết H_0: $p = p_0 = 0{,}155$ và $H_1 : p \ne p_0$

Ta đi kiểm tra giả thiết này:

+ Nếu nó được chấp nhận thì phương pháp sản xuất mới không có tác dụng gì.

+ Nếu giả thiết bị bác bỏ thì phương pháp sản xuất mới đã làm thay đổi tỷ lệ loại A.

Do $f = 0{,}2 > p_0 = 0{,}155$ nên p cũng lớn hơn p_0 và tỷ lệ loại A tăng lên. Phương pháp sản xuất mới có hiệu quả tốt.

Kiểm tra giả thiết:

Tính kiểm định $t = \dfrac{(f - p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}} = \dfrac{(0{,}2 - 0{,}155)\sqrt{1000}}{\sqrt{0{,}155(1-0{,}155)}} \approx 4$

Với $\alpha = 1\%$ tra bảng phân vị chuẩn (bảng 3) ta có $t_\alpha = 2{,}5758$

Vì $|t| > t_\alpha$ nên ta bác bỏ giả thiết H_0. Theo phân tích ở trên ta thấy phương pháp sản xuất mới này có hiệu quả tốt.

II. Kiểm định giả thiết về trung bình đám đông

Bài toán

Giả sử đám đông có trung bình m chưa biết. Ta cần kiểm tra giả thiết H_0: $m = m_0$ và $H_1 : m \neq m_0$

Hãy đưa quy tắc căn cứ vào mẫu gồm n quan sát độc lập (x_1, .., x_n) mà chấp nhận hay bác bỏ giả thiết trên với mức ý nghĩa α. Chia 3 trường hợp:

1. Trường hợp 1: $X \sim N(m;\sigma^2)$ với σ^2 đã biết.

Ta chọn thống kê : $Z = \dfrac{(\overline{X} - m_0)\sqrt{n}}{\sigma}$ làm tiêu chuẩn kiểm định.

Nếu giả thuyết H_0 đúng thì $Z \sim N(0,1)$

Với mức ý nghĩa , chọn miền bác bỏ giả thiết H_0: $W_\alpha = \{z : |z| > z_\alpha\}$

Trong đó z là giá trị $Z \sim N(0,1)$ của thỏa mãn: $P(|Z| > z_\alpha) = \alpha$

Ta có: $P(Z \in W_\alpha) = P(|Z| > Z_\alpha) = P(Z < -z_\alpha) + P(Z > z_\alpha)$

$$= \alpha/2 + 1 - (1 - \alpha/2) = \alpha$$

Như vậy xác suất để giá trị của Z rơi vào miền bác bỏ là α , tức xác suất để Z rơi vào miền chấp nhận sẽ là $1-\alpha$. Vì α nhỏ, nên xác suất để Z rơi vào miền chấp nhận sẽ lớn.

Nghĩa là: nếu giả thiết H_0 đúng thì có thể coi rằng hầu hết các giá trị của Z sẽ rơi vào miền chấp nhận. Còn nếu giá trị của Z rơi vào miền bác bỏ có nghĩa là ta đã tìm được "bằng chứng" để chứng tỏ giả thiết H_0 là không đúng và vì thế ta bác bỏ giả thiết đó.

Quy tắc thực hành

Ta đặt giả thiết H_0: $m = m_0$ và $H_1 : m \neq m_0$.

+ Từ mẫu tính $\overline{x}$

+ Tính kiểm định $t = \dfrac{(\overline{x} - m_0)\sqrt{n}}{\sigma}$

+ Với mức ý nghĩa α tra bảng phân vị chuẩn (tra bảng 3) tìm t_α

Kết luận: + Nếu $|t| > t_\alpha$ ta loại giả thiết H_0 coi $m \neq m_0$

+ Nếu $|t| \leq t_\alpha$ ta chấp nhận giả thiết H_0 coi $m = m_0$

Ví dụ 2 Một máy tiện tự động quy định đường kính trung bình của trục máy là 225 cm, độ lệch chuẩn là 36 cm. Sau một thời gian sản xuất, kiểm tra 81 trục, ta có $\overline{x} = 210$ cm. Cho kết luận về tình hình sản xuất với mức ý nghĩa $\alpha = 1\%$.

BÀI GIẢI

Trung bình quy định $m_0 = 225$ cm.

Trung bình thực tế sản xuất là m chưa biết.

Đat giả thiết $H_0 : m = m_0 = 225$ cm và $H_1 : m \neq m_0$

Chú ý: + Nếu giả thiết được chấp nhận thì sản xuất bình thường.

+ Nếu giả thiết bị bác bỏ thì máy sản xuất ra sản phẩm không đúng quy định.

Kiểm tra giả thiết: Ta có $n = 81$, $\sigma = 36$ đã biết

Tính kiểm định $t = \dfrac{(\overline{x} - m_0)\sqrt{n}}{\sigma} = \dfrac{(210-225)\sqrt{81}}{36} = -3,75$

Với mức ý nghĩa $\alpha = 1\%$ tra bảng 3 ta có $t_\alpha = 2,5758$

Vì $|t| > t_\alpha$ nên ta loại giả thiết H_0.

Theo chú ý ở trên, máy sản xuất các trục máy có đường kính trung bình nhỏ hơn quy định.

2. Trường hợp 2: $n \geq 30$ và $X \sim N(m;\sigma^2)$ với σ^2 chưa biết

Ta làm như trường hợp 1 nhưng thay $\sigma = s$.

Quy tắc thực hành

Ta đặt giả thiết H_0**:** $m = m_0$ và $H_1 : m \neq m_0$

+ Tính kiểm định $t = \dfrac{(\overline{x} - m_0)\sqrt{n}}{s}$

+ Với mức ý nghĩa α tra bảng phân vị chuẩn (bảng 3) tìm t_α

Kết luận: + Nếu $|t| > t_\alpha$ ta loại giả thiết H_0

+ Nếu $|t| \leq t_\alpha$ ta chấp nhận giả thiết H_0.

Ví dụ 3 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 GIẢI

Khối lượng trung bình quy định $m_0 = 1$ kg

Khối lượng trung bình thực tế sản xuất là m chưa biết.

Đưa giả thiết $H_0 : m = m_0 =$ 1 kg và $H_1 : m \neq m_0$.

Từ số liệu của mẫu, ta có

Cỡ mẫu : $n = 100$; $\overline{X} = 0,9856$; $S_X^2 = 0,00433$; $S_X = 0,021$,

Với số liệu trên, ta được

$$t = \frac{\left(0,9856 - 1\right)\sqrt{100}}{0,021} = -6,86 .$$

Với mức ý nghĩa $\alpha = 0,05$, ta có t_α = 1,96. Vì $|t| > t_\alpha$ ta loại giả thiết H_0, nghĩa là máy hoạt động không bình thường

3. Trường hợp 3: cỡ mẫu $n < 30$ và $X \sim N(m;\sigma^2)$ với σ^2 chưa biết

Trường hợp này chọn: $T = \dfrac{(\overline{X} - m_0)\sqrt{n}}{S}$ làm tiêu chuẩn kiểm định.

Nếu H_0 đúng thì T phân phối theo qui luật Student với n-1 bậc tự do $P\left(|T| > t_\alpha^{(n-1)}\right) = \alpha$ được xác định bằng cách tra bảng phân phối Student với bậc tự do n-1 (bảng 4).

Quy tắc thực hành

Ta đặt giả thiết H_0**:** $m = m_0$ và $H_1 : m \neq m_0$

+ Tính kiểm định $t = \dfrac{(\overline{x} - m_0)\sqrt{n}}{s}$

+ Tìm $t_{\alpha}^{(n-1)}$: tra bảng phân vị student bậc n-1 (tra bảng 4)

Kết luận: + Nếu $|t| > t_{\alpha}^{(n-1)}$ ta loại giả thiết H_0

+ Nếu $|t| \le t_{\alpha}^{(n-1)}$ ta chấp nhận giả thiết H_0

Chú ý: Nếu giả thiết H_0 đã bị bác bỏ, tức là ta xem $m \ne m_0$.

Khi đó:

+ Nếu $\overline{x} > m_0$ ta kết luận $m > m_0$.

+ Nếu $\overline{x} < m_0$ ta kết luận $m < m_0$.

Ví dụ 4 Một máy đóng bao với trọng lượng trung bình quy định là 40g. Cân 15 bao, ta có $\overline{x} = 39{,}8$g, $s^2 = 0{,}144$. Giả thiết X là trọng lượng của các bao bột có phân phối chuẩn. Cho kết luận về tình hình làm việc của máy (với mức ý nghĩa $\alpha = 5\%$).

BÀI GIẢI

Trọng lượng trung bình quy định $m_0 = 40$g

Trọng lượng trung bình thực tế sản xuất là m chưa biết.

Đặt giả thiết $H_0 : m = m_0 = 40$ g và $H_1 : m \ne m_0$

Ta thấy n = 15 < 30, X có phân phối chuẩn, σ^2 chưa biết nên thuộc trường hợp 3

Tính kiểm định $t = \dfrac{(\overline{x} - m_0)\sqrt{n}}{S} = \dfrac{(39{,}8 - 40)\sqrt{15}}{0{,}379} = -2{,}04$

Từ bảng phân phối Student (bảng 4) với 14 bậc tự do ta có

$t_{\alpha}^{n-1} = t_{5\%}^{14} = 2{,}145$.

Vì $|t| = 2{,}04 < 2{,}145$ nên ta chấp nhận giả thiết H_0.

Vậy tình hình sản xuất bình thường.

III. Kiểm định giả thiết về phương sai đám đông có phân phối chuẩn.

1. Bài toán

Giả sử đám đông đặc trưng bởi đại lượng ngẫu nhiên X có phân phối chuẩn $X \sim N(m, \sigma^2)$ với σ^2 chưa biết.

Ta cần kiểm định giả thiết $H_0 : \sigma^2 = \sigma_0^2$ và $H_1 : \sigma^2 \neq \sigma_0^2$

Hãy đưa quy tắc căn cứ vào mẫu $(x_1, \ldots, x_n)$ mà chấp nhận hay bác bỏ giả thiết trên với mức ý nghĩa α.

2. Quy tắc thực hành

Ta cần kiểm định giả thiết $H_0 : \sigma^2 = \sigma_0^2$ và $H_1 : \sigma^2 \neq \sigma_0^2$

Giả sử có mẫu $(x_1, \ldots, x_n)$ ta tính kiểm định $\chi^2 = \dfrac{(n-1)s^2}{\sigma_0^2}$

Biết α, từ χ^2_{n-1} (tra bảng 5) ta tra được $\chi^2_{n-1}\left(\dfrac{\alpha}{2}\right)$ và $\chi^2_{n-1}\left(1-\dfrac{\alpha}{2}\right)$

Nếu $\chi^2 < \chi^2_{n-1}\left(\dfrac{\alpha}{2}\right)$ hoặc $\chi^2 > \chi^2_{n-1}\left(1-\dfrac{\alpha}{2}\right)$ ta bác bỏ giả thiết H_0.

Nếu $\chi^2_{n-1}\left(\dfrac{\alpha}{2}\right) < \chi^2 < \chi^2_{n-1}\left(1-\dfrac{\alpha}{2}\right)$ ta chấp nhận giả thiết H_0

Trong trường hợp giả thiết bị bác bỏ:

+ Nếu $s^2 > \sigma_0^2$ ta kết luận $\sigma^2 > \sigma_0^2$

+ Nếu $s^2 < \sigma_0^2$ ta kết luận $\sigma^2 < \sigma_0^2$

Ví dụ 5 Tiến hành 25 quan sát cụ thể về chỉ tiêu X của một loại sản phẩm, ta tính được s^2=416,666. Có tài liệu nói phương sai của chỉ tiêu X là $\sigma^2 = 400$.

Cho nhận xét về tài liệu này với mức ý nghĩa $\alpha = 5\%$.

BÀI GIẢI

Ta cần kiểm định giả thiết $H_0 : \sigma^2 = 400$ và $H_1 : \sigma^2 \neq \sigma_0^2$

Ta thấy $\chi^2 = \dfrac{(n-1)s^2}{\sigma_0^2} = \dfrac{24.216,666}{400} = 25$

Tra bảng phân phối χ^2 (bảng 5) ta có :

$\chi^2_{24}(0,975) = 39,4$; $\chi^2_{24}(0,025) = 12,4$

Ta thấy $\chi^2_{n-1}\left(\dfrac{\alpha}{2}\right) < \chi^2 < \chi^2_{n-1}\left(1-\dfrac{\alpha}{2}\right)$

Nên ta chấp nhận giả thiết Ho: $\sigma^2 = \sigma_0^2$ và xem tài liệu cho con số chấp nhận được.

IV. So sánh hai tỉ lệ

Trong các vấn đề tiếp dưới đây, chúng ta chỉ nói về luật phân phối của thống kê T mà giá trị cụ thể t được dùng làm tiêu chuẩn thử nghiệm (test) theo nguyên lý biến cố xác suất nhỏ. Việc chứng minh luật phân phối đó các bạn có thể suy ra trên cơ sở các kiến thức trình bày trong cuốn sách này (đặc biệt là các phần: hàm các đại lượng ngẫu nhiên, phân phối các đặc trưng mẫu và thống kê…). Hơn nữa, cơ sở toán học của quy tắc thực hành các bạn cũng tự giải thích như các phần trước với nhiều tương tự.

1. Bài toán

Ta có hai đám đông: Gọi tỉ lệ các phần tử có tính chất A cần quan tâm trong đám đông thứ nhất là p_1 và tỉ lệ các phần tử có tính chất A cần quan tâm trong đám đông thứ hai là p_2.

Ta cần kiểm định giả thiết H_0: $p_1 = p_2$ và $H_1 : p_1 \neq p_2$

Hãy đưa quy tắc căn cứ vào mẫu điều tra với cỡ n_1 từ đám đông thứ nhất và cỡ n_2 từ đám đông thứ hai mà chấp nhận hay bác bỏ giả thiết trên với mức ý nghĩa α.

2. Quy tắc thực hành

Ta xét trường hợp n_1 và n_2 khá lớn. Giả sử cỡ mẫu thứ nhất là n_1 với tỷ lệ thực nghiệm f_1 và cỡ mẫu thứ hai là n_2 với tỷ lệ thực nghiệm f_2. Ta tính $f = \dfrac{n_1 f_1 + n_2 f_2}{n_1 + n_2}$ là tỷ lệ thực nghiệm chung trong 2 mẫu.

Sau đó tính kiểm định

$$t = \frac{f_1 - f_2}{\sqrt{f\,(1-f)\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$$

Biết mức ý nghĩa α ta tìm t_α từ bảng phân vị chuẩn (bảng 3).

+ Nếu $t \le t_\alpha$ ta chấp nhận giả thiết

+ Nếu $t > t_\alpha$ ta bác bỏ giả thiết.

Trong trường hợp này:

* Khi $f_1 > f_2$ thì ta kết luận $p_1 > p_2$.
* Khi $f_1 < f_2$ thì kết luận $p_1 < p_2$.

Ví dụ 6 Từ hai đám đông tiến hành hai mẫu với $n_1 = 100$ và $n_2 = 120$ quan sát và tính được $f_1 = 0{,}2$; $f_2 = 0{,}30$.

Hãy kiểm định giả thiết H_0: $p_1 = p_2$ với mức ý nghĩa $\alpha = 1\%$.

BÀI GIẢI

Ta tính $f = \dfrac{n_1 f_1 + n_2 f_2}{n_1 + n_2} = \dfrac{100.0,2 + 120.0,3}{100 + 120}$ =0,255

$$t = \frac{0,2 - 0,3}{\sqrt{0,255.0,745\left(\dfrac{1}{100} + \dfrac{1}{120}\right)}} = -1,695$$

Vì $|t| < t_\alpha$ =2,5758 nên ta chấp nhận giả thiết, tức là xem $p_1 = p_2$.

V. So sánh hai trung bình

1. Bài toán Xét hai đám đông

Gọi trung bình của đám đông thứ nhất là m_1

Trung bình của đám đông thứ hai là m_2.

Ta có giả thiết H_0: $m_1 = m_2$ và $H_1 : m_1 \neq m_2$

Lấy từ đám đông thứ nhất một mẫu cỡ n, lấy từ đám đông thứ hai một mẫu cỡ m.

Hãy đưa quy tắc căn cứ vào mẫu $(X_1, \ldots, X_n)$ và $(Y_1, \ldots, Y_m)$ mà chấp nhận hay bác bỏ giả thiết trên với mức ý nghĩa α.

2. Quy tắc thực hành

a) Trường hợp 1: n, m > 30 hoặc n, m < 30, σ_1^2, σ_2^2 đã biết.

Từ hai mẫu cụ thể $(x_1, \ldots, x_n)$ và $(y_1, \ldots, y_m)$

Ta tính
$$t = \frac{\overline{x} - \overline{y}}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}$$

Từ mức α ta tìm t_α (Tra bảng phân vị chuẩn là bảng 3).

+ Nếu $|t| \leq t_\alpha$ ta chấp nhận giả thiết.

+ Nếu $|t| > t_\alpha$ ta bác bỏ giả thiết và khi đó:

* Nếu $\overline{x} > \overline{y}$ ta kết luận $m_1 > m_2$.

* Nếu $\overline{x} < \overline{y}$ ta kết luận $m_1 < m_2$.

b) Trường hợp 2: n, m > 30, σ_1^2, σ_2^2 chưa biết.

Tương tự trường hợp 1 nhưng thay $\sigma_1^2 = S_1^2; \sigma_2^2 = S_2^2$

Ta tính
$$t = \frac{\overline{x} - \overline{y}}{\sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{m}}}$$

Từ mức α ta tìm t_α (Tra bảng phân vị chuẩn là bảng 3)

+ Nếu $|t| \le \tau_\alpha$ ta chấp nhận giả thiết.

+ Nếu $|t| > t_\alpha$ ta bác bỏ giả thiết và khi đó:

* Nếu $\overline{x} > \overline{y}$ ta kết luận $m_1 > m_2$.

* Nếu $\overline{x} < \overline{y}$ ta kết luận $m_1 < m_2$.

c) Trường hợp 3: n, m < 30, X, Y có phân phối chuẩn với σ_1^2, σ_2^2 chưa biết

Ta tính
$$t = \frac{\overline{x} - \overline{y}}{S\sqrt{\frac{1}{n} + \frac{1}{m}}}$$

Trong đó
$$s^2 = \frac{(n-1)s_1^2 + (m-1)s_2^2}{n-1+m-1}$$

Từ mức α tra bảng phân phối Student (bảng 4) với (n + m – 2) bậc tự do để tìm số t_α^{n+m-2},

+ Nếu $|t| \le t_\alpha^{n+m-2}$ ta chấp nhận giả thiết.

+ Nếu $|t| > t_\alpha^{n+m-2}$ ta bác bỏ giả thiết và khi đó:

* Nếu $\overline{x} > \overline{y}$ ta kết luận $m_1 > m_2$.

* Nếu $\overline{x} < \overline{y}$ ta kết luận $m_1 < m_2$.

Ví dụ 7 Theo một tài liệu của viện nghiên cứu phát triển gia cầm thì hai giống gà G_1 và G_2 có trọng lượng trung bình ở 3 tháng tuổi là như nhau. Người ta nuôi thử mỗi giống 100 con, và ở do 3 tháng tuổi cân lại người ta tính được kết quả tương ứng là

$\overline{x} = 1825g$, $s_1^2 = 1628$, $\overline{y} = 1973g$, $s_2^2 = 1876$.

Hãy căn cứ vào mẫu đó cho nhận xét về tài liệu trên với mức ý nghĩa α = 1%.

BÀI GIẢI

m_1 là trọng lượng trung bình giống gà G_1

m_2 là trọng lượng trung bình giống gà G_2

Ta cần kiểm tra giả thiết H_0: $m_1 = m_2$.

Ta thấy đây là trường hợp n, m > 30, σ_1^2, σ_2^2 chưa biết

Ta đi tính kiểm định

$$t = \frac{\overline{x} - \overline{y}}{\sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{m}}} = \frac{1825 - 1973}{\sqrt{\frac{1628 + 1876}{100}}} = -25$$

với mức ý nghĩa $\alpha = 1\%$ thì $t_\alpha = 2,58$.

Kết luận: $|t| > t_\alpha$ ta bác bỏ giả thiết và vì $\overline{y} > \overline{x}$ ta kết luận tài liệu không chính xác, trọng lượng trung bình giống gà G_2 lớn hơn trọng lượng trung bình giống gà G_1 (ở 3 tháng tuổi).

Ví dụ 8 Quan sát cân nặng của bé trai (X) và bé gái (Y) lúc sơ sinh (đơn vị g)

Trọnglượng(kg)	3-3,2	3,2-3,4	3,4-3,6	3,6-3,8	3,8-4
Số bé trai	1	3	8	10	3
Số bé gái	2	10	10	5	1

a) Tính $\overline{x}$, $\overline{y}$, s_X^2, s_Y^2.

b) So sánh các trung bình μ_X, μ_Y
(kết luận với mức ý nghĩa $\alpha = 5\%$).

c) Nhập hai mẫu lại. Tính trung bình và độ lệch chuẩn của mẫu nhập. Dùng mẫu nhập để ước lượng sức nặng trung bình của trẻ sơ sinh ở độ tin cậy 95%.

BÀI GIẢI Với n_1 là số bé trai quan sát, n_2 số bé gái quan sát.

Trọng lượng	3,1	3,3	3,5	3,7	3,9
n_1	1	3	8	10	3
n_2	2	10	10	5	1
Tổng số	3	13	18	15	4

a) Từ số liệu của mẫu, ta có $\overline{X} = 3,588$, $\overline{Y} = 3,450$,

$$S_X^2 = \frac{1}{24}\left(\sum_{i=1}^{5} n_{1i}X_i^2 - 25\overline{X}^2\right) = 40,26667,$$

$$S_Y^2 = \frac{1}{27}\left(\sum_{i=1}^{5} n_{2i}Y_i^2 - 28\overline{Y}^2\right) = 37,40741.$$

b) Ta có bài toán kiểm định $H: \mu_X = \mu_Y$

Đáp số: bác bỏ H, nghĩa là $\mu_X \neq \mu_Y$.

c) Nhập hai mẫu lại. Gọi Z là trong luong của trẻ sơ sinh. Từ bảng số liệu, ta có

$$\overline{Z} = \frac{1}{53}\left(3,1\cdot 3 + 3,3\cdot 13 + 3,5\cdot 18 + 3,7\cdot 15 + 3,9\cdot 4\right) = 3,5151$$

$$S_{XY}^2 = 42,8047\,;\ S_Z = 0,20689.$$

Để tìm khoảng ước lượng cho trung bình mẫu nhập Z,

ta dùng thống kê

$$T = \frac{\left(\overline{Z} - \mu\right)\sqrt{n}}{S_Z} = \frac{\left(3,5151 - \mu\right)\sqrt{53}}{0,20698} \sim N(0;1).$$

Với độ tin cậy $\gamma = 0.95$ thì $t_\alpha = 1,96$, $\varepsilon = t_\alpha \frac{s_z}{\sqrt{n}} = 0,0557$

ta suy ra khoảng ước lượng trung bình của mẫu nhập

(3,45938; 3,57082)

BÀI TẬP CHƯƠNG V

5.1. Trọng lượng trung bình khi xuất chuồng ở một trại chăn nuôi gà công nghiệp năm trước là 3,3kg/con. Năm nay người ta sử dụng một loại thức ăn mới. Cân thử 15 con khi xuất chuồng ta được số liệu sau (đơn vị tính là kg).

3,25	2,50	4,00	3,75	3,80	3,9	4,02	3,80
3,20	3,82	3,40	3,60	3,75	4,00	3,50	

Giả thiết trọng lượng gà là đại lượng ngẫu nhiên phân phối theo quy luật chuẩn với phương sai là 0,16.

a) Với mức ý nghĩa $\alpha = 0,05$, hãy cho kết luận về tác dụng của loại thức ăn này.

b) Nếu trại chăn nuôi báo cáo trọng lượng trung bình khi xuất chuồng là 3,5kg/con thì có chấp nhận được không?

(với mức ý nghĩa $\alpha = 0,05$).

5.2. Trọng lượng của một loại sản phẩm do một nhà máy sản xuất là đại lượng ngẫu nhiên phân phối theo quy luật chuẩn với trọng lượng trung bình là 500g. Nghi ngờ trọng lượng của loại sản phẩm này có xu hướng giảm sút, người ta cân thử 25 sản phẩm và thu được kết quả cho ở bảng sau:

Trọng lượng (g)	480	485	490	495	500	510
Số sản phẩm	2	3	8	5	3	4

Với mức ý nghĩa 0,05, hãy kết luận về điều nghi ngờ nói trên.

5.3. Khảo sát về thu nhập X (triệu đồng / tháng) của một số người, ta có bảng số liệu như sau:

Thu nhập	0 – 4	4 – 8	8 –12	12 – 16	16 –20	20 –24
Số người	8	12	20	30	16	10

a) Hãy ước lượng thu nhập trung bình của một người trong một tháng với độ tin cậy 90%.

b) Những người có thu nhập không quá 4 triệu đồng/ tháng là những người có mức thu nhập thấp. Hãy ước lượng tỷ lệ người thu nhập thấp trong một tháng với độ tin cậy 96%.

c) Giả sử có người báo cáo rằng "Mức thu nhập trung bình của một người là 13 triệu đồng / tháng" thì có chấp nhận được không (với mức ý nghĩa 5%) ?

d) Những người có thu nhập không quá 4 triệu đồng/ tháng là những người có mức thu nhập thấp. Giả sử có người báo cáo rằng "Tỷ lệ những người có mức thu nhập thấp là 10%" thì có chấp nhận được không với mức ý nghĩa 5%?

5.4. Để nghiên cứu nhu cầu về 1 loại hàng ở một khu vực, người ta khảo sát nhu cầu của mặt hàng này ở 400 gia đình. Kết quả cho ở bảng sau

Giả sử nhu cầu về loại hàng này tuân theo phân phối chuẩn

Nhu cầu (kg/tháng)	Số gia đình	Nhu cầu (kg/tháng)	Số gia đình
0 – 1	10	4 – 5	78
1 – 2	35	5 – 6	31
2 – 3	86	6 – 7	18
3 – 4	132	7 – 8	10

Giả sử khu vực đó có 4000 hộ gia đình.

a) Ước lượng nhu cầu trung bình về mặt hàng này của toàn khu vực trong 1 tháng với độ tin cậy 95%.

b) Những gia đình có nhu cầu của mặt hàng này lớn hợn 5kg/tháng là những gia đình có nhu cầu cao. Hãy ước lượng tỷ lệ gia đình có nhu cầu cao trong 1 tháng với đô tin cậy 99%

c) Nếu cho rằng nhu cầu trung bình về mặt hàng này của toàn khu vực là 14 tấn trong 1 tháng thì có chấp nhận được không (với mức ý nghĩa $\alpha = 0{,}01$)?

d) Nếu cho rằng tỷ lệ gia đình có nhu cầu cao trong 1 tháng là 16% thì có chấp nhận được không (với mức ý nghĩa $\alpha = 0{,}05$)?

5.5. Điều tra năng suất của một giống lúa trên 100 ha. Ta có bảng số liệu sau

Năng suất (tấn / ha)	8.5	9.5	10.5	11.5	12.5	13.5
Số ha	4	16	25	30	15	10

Giả sử năng suất của giống lúa có luật phân phối chuẩn.

a) Hãy ước lượng năng suất lúa trung bình với độ tin cậy 95%.

b) Những thửa ruộng có năng suất từ 10.5 (tấn/ha) trở lên được gọi là đạt tiêu chuẩn. Hãy ước lượng tỷ lệ các thửa ruộng đạt tiêu chuẩn với độ tin cậy 95%.

c) Muốn độ chính xác khi ước lượng năng suất trung bình không quá 0.2 với độ tin cậy là 95% thì cần quan sát thêm ít nhất bao nhiêu hecta nữa ?

d) Theo một tài liệu thống kê cho biết năng suất trung bình của giống lúa trên là 11 (tấn/ha). Hãy cho biết bảng số liệu trên có phù hợp với tài liệu này không (kết luận với mức ý nghĩa 5%)?

CHƯƠNG VI

LÝ THUYẾT TƯƠNG QUAN VÀ HÀM HỒI QUY

6.1. MỐI QUAN HỆ GIỮA HAI ĐẠI LƯỢNG NGẪU NHIÊN

Khi khảo sát hệ hai đại lượng ngẫu nhiên (X, Y) ta phân biệt các quan hệ sau đây giữa chúng:

I. X và Y độc lập với nhau: tức việc X (hay Y) nhận giá trị nào đó không làm thay đổi quy luật phân phối xác xuất của Y (hay X). Điều kiện cần và đủ để X và Y độc lập với nhau là:

$$f(x,y) = f_1(x)\, f_2(y)$$

Trong đó: fx,y) là ham mật độ xác suất của hệ (X, Y)'

$f_1(x)$, $f_2(y)$ là hàm mật độ xác suất của X và của Y.

II. X, Y có sự phụ thuộc hàm số: Tức các giá trị có thể có của X và Y có dạng $y=\varphi\ (x)$

III. X, Y có sự phụ thuộc tương quan và sự phụ thuộc không tương quan:

Vì các khái niệm độc lập, phụ thuộc, tương quan và không tương quan được định nghĩa xuất phát từ hàm mật độ và hệ số tương quan r nên có thể hệ thống hóa các loại quan hệ của X và Y dưới dạng bảng sau:

R	r=0	$o < \lvert r \rvert < 1$	$\lvert r \rvert = 1$
$F(x,y)=f_1(x)f_2(y)$	Độc lập không tương quan		
$F(x,y) \neq f_1(x)f_2(y)$	Phụ thuộc không tương quan	Phụ thuộc tương quan	Phụ thuộc hàm

Trong các loại quan hệ trên, thống kê toán đặc biệt chú ý nghiên cứu mối quan hệ phụ thuộc tương quan giữa hai đại lượng ngẫu nhiên X và Y, do đó trước hết phải trả lời câu hỏi: Có tồn tại mối quan hệ này hay không? Tức xét xem hệ số tương quan:

$$R_{xy} = \frac{E\{[X - E(X)].[Y - E(Y)]\}}{\sigma(X).\sigma(Y)}$$ có khác 0 hay không?

Thông thường R_{xy} chưa biết, nên phải ước lượng nhờ phương pháp

mẫu.
Nếu X, Y có sự phụ thuộc tương quan thì công cụ để nghiên cứu sự phụ thuộc này là hàm hồi quy của Y đối với X: f(x)=E(Y/x).
Hay hàm hồi quy của X đối với Y: g(y)=E(X/y)
E(Y/x) là kỳ vọng toán có điều kiện của Y đối với X.
E(X/y) là kỳ vọng toán có điều kiện của X đối với Y.
Chú ý:
+ Hàm f(x)=E(Y/x) phản ánh quan hệ hàm số giữa từng giá trị x của X với kỳ vọng (trung bình) có điều kiện của Y (điều kiện ở đây là X=x).
+ Tương tự như vậy, hàm g(y)=E(X/y) phản ánh quan hệ hàm số giữa từng giá trị y với kỳ vọng có điều kiện của X.
Trong thực tế các hàm hồi quy f(x), g(y) thường chưa biết, vì vậy cũng đặt ra bài toán ước lượng các hàm này bằng phương pháp mẫu.

6.2. BẢNG TƯƠNG QUAN THỰC NGHIỆM

Khi lập mẫu kích thước n, với sự quan sát đồng thời các giá trị của X và các giá trị của Y, ta thu được số liệu thức nghiệm và trình bày dưới dạng bảng sau:

X \ Y	Y_1	y_2	…	y_j	…	y_h	n_i
x_1	n_{11}	n_{12}	…	n_{1j}	…	N_{1h}	n_1
x_2	n_{21}	n_{22}	…	n_{2j}	…	N_{2h}	n_2
…	…	…	…	…	…	…	…
x_i	n_{i1}	n_{i2}	…	n_{ij}	…	n_{ih}	n_i
…	…	…	…	…	…	…	…
x_k	n_{k1}	n_{k2}	…	n_{kj}	…	n_{kh}	n_k
m_j	m_1	m_2	…	m_j	…	m_h	**N**

Trong đó
x_i $(i=\overline{l,k})$ là các giá trị có thể có của X
y_j $(j=\overline{l,h})$ là các giá trị có thể có của Y.
n_i $(i=\overline{l,k})$ là tần số của X (số lần X nhận giá trị x_i)
m_j $(j=\overline{l,h})$ là tần số của Y (số lần Y nhận giá trị y_j)

n_{ij} ($i=\overline{l,k}$; $j=\overline{l,h}$) là số lần X nhận giá trị x_i đồng thời Y nhận giá trị y_j.

Đối với bảng tương quan thực nghiệm ta luôn có:

$$\sum_{i=1}^{k} n_i = \sum_{j=1}^{h} m_j = \sum_{i=1}^{k}\sum_{j=1}^{h} n_{ij} = n$$

Từ bảng tương quan thực nghiệm ta có thể suy ra:

I. Phân phối thực nghiệm của X

X	x_1	x_2	…	x_i	…	x_k
Tần số	n_1	n_2	…	n_i	…	n_k

Từ phân phối thực nghiệm của X, ta tính được các đặc trưng mẫu theo đại lượng X

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

II. Phân phối thực nghiệm của Y

Y	y_1	y_2	…	y_j	…	y_h
Tần số	m_1	m_2	…	m_j	…	m_h

Từ đó ta tính được các đặc trưng mẫu của Y là:

$$\overline{y} = \frac{1}{n}\sum_{j=1}^{h} mj.y_j \text{ ; } \hat{s}_Y^2 = \frac{1}{n}\sum_{j=1}^{h} m_j y_j^2 - \left(\overline{y}\right)^2$$

III. Các phân phối có điều kiện của Y (điều kiện là X = x_i)

Y	y_1	y_2	…	y_j	…	y_h
Tần số	n_{i1}	n_{i2}	…	n_{ij}	…	n_{ih}

Từ các phân phối có điều kiện, ta có thể tính được các giá trị trung bình có điều kiện của Y: $\overline{y_{x_{ij}}} = \frac{1}{n_i}\sum_{j=1}^{h} n_{ij}.y_j$

IV. Đường hồi quy thực nghiệm

- Nếu đưa các điểm $M_1(x_i. \overline{y_{x_i}})$ lên hệ trục tọa độ và nối chúng lại, ta sẽ được một đường gấp khúc, gọi là đường hồi quy thực nghiệm của Y đối với X.

- Nếu kích thước mẫu (n) khá lớn thì:

$$f_{ij} = \frac{n_{ij}}{n} \approx P(Y = y_j / X = x_i)$$

Và khi đó đường hồi quy thực nghiệm sẽ xấp xỉ với đường hồi quy lý thuyết.

Vì vai trò của X và Y như nhau, nên ta cũng có thể tính các giá trị $\overline{x_{y_j}}$ và vẽ đường hồi quy thực nghiệm của X đối với Y một cách tương tự.

Ví dụ 1 Đo chiều cao và trọng lượng của 81 học sinh, ta được các số liệu cho bảng tương quan thực nghiệm sau

Trong đó: X là trọng lượng (đơn vị tính là kg)

Y là chiều cao (đơn vị tính là cm)

X \ Y	150	155	160	165	170
40	3	6			
45	5	10	9	4	
50		8	12	15	2
55					7

a) Hãy lập phân phối thực nghiệm của X

b) Hãy lập phân phối thực nghiệm của Y

c) Hãy tìm các trung bình có điều kiện của Y theo X

d) Hãy vẽ đường hồi quy thực nghiệm của Y đối với X

BÀI GIẢI

Ở bảng tương quan này, ta có

$$n = \sum n_i = \sum m_j = \sum n_{ij} = 81$$

a) Phân phối thực nghiệm của X:

x_i	40	45	50	55
Tần số	9	28	37	7

Từ bảng tính được:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} n_i.x_i = \frac{1}{81}3855 = 47,5926;$$

$$\hat{s}_x^2 = \frac{1}{n}\sum_{i=1}^{k} n_i.x_i^2 - \left(\overline{x}\right)^2 = \frac{1}{81}184775 - \left(47,5926\right)^2 = 16,1173$$

b) Phân phối thực nghiệm của Y:

y_j	150	155	160	165	170
Tần số	8	24	21	19	9

Từ bảng tính được:

$$\overline{y} = \frac{1}{n}\sum_{j=1}^{h} mj.y_j = \frac{1}{81}12945 = 159,815;$$

$$\hat{s}_y^2 = \frac{1}{n}\sum_{j=1}^{h} m_j.y_j^2 - \left(\overline{y}\right)^2 = \frac{1}{81}2071575 - \left(159,815\right)^2 = 34,1658$$

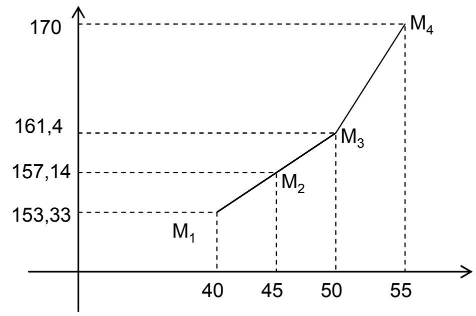
c) Các trung bình có điều kiện của Y:

$$\overline{y_{40}} = \frac{1}{9}\left(3.150 + 6.155\right) = 153,33$$

$$\overline{y_{45}} = \frac{1}{28}\left(5.150 + 10.155 + 9.160 + 4.165\right) = 157,14$$

$$\overline{y_{50}} = 161,49; \overline{y_{55}} = 170;$$

d) Từ đó ta có thể vẽ đường hồi quy thực nghiệm của Y đối với X.



6.3. ƯỚC LƯỢNG HỆ SỐ TƯƠNG QUAN VÀ HÀM HỒI QUY

I. Ước lượng hệ số tương quan

Chúng ta đã biết, hệ số tương quan của hai đại lượng ngẫu nhiên X và Y được xác định với công thức

$$R_{xy} = \frac{E\{[X - E(X)].[Y - E(Y)]\}}{\sigma(X).\sigma(Y)} = \frac{E(X.Y) - E(X).E(Y)}{\sigma(X).\sigma(Y)}$$

Để ước lượng R_{xy} ta dùng thống kê

$$R = \frac{\overline{XY} - \overline{X}.\overline{Y}}{\hat{S}_X.\hat{S}_Y} \quad \text{Với} \quad \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \ ;$$

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \ ; \qquad \overline{XY} = \frac{1}{n}\sum_{i=1}^{n} X_i Y_i$$

Trong đó

(Xi, Yi) là thành phần của mẫu ngẫu nhiên 2 chiều:

$W_{XY}=\{(X_1,Y_1), (X_2, Y_2),\ldots,(X_n,Y_n)$

Với mẫu cụ thể, ta sẽ tính được giá trị cụ thể của R. Ký hiệu giá trị này là r.

$$r = \frac{\overline{xy} - \overline{x}.\overline{y}}{\hat{s}_x.\hat{s}_y}$$

Trong đó: $\overline{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i$; $\overline{y} = \frac{1}{n}\sum_{j=1}^{h} m_j y_j$; $\overline{xy} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{h} n_{ij} x_i y_j$

Ví dụ 2 Từ số liệu cho ở ví dụ 1, hãy ước lượng hệ số tương quan của Y và X?

BÀI GIẢI

Ở ví dụ 1, ta đã tính được: $\overline{x}=47,5926$, $\hat{S}_x^2=16,3189$

$\overline{y}=159,815$, $\hat{S}_y^2=34,59375$

Tính $\overline{xy}=\frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{h}n_{ij}x_iy_j=\frac{1}{81}617400$

Từ kết quả tính toán ta có: $r=\dfrac{\overline{xy}-\overline{x}.\overline{y}}{\hat{s}_x.\hat{s}_y}$ =0,69

II. Phương pháp bình phương bé nhất để ước lượng hàm hồi quy

Giả sử hàm hồi quy của Y đối với X có dạng: E(Y/x)=f(x,A,B,C…)

Trong đó: A, B, C là tham số. Chẳng hạn: f(X,A,B)=Ax+B hoặc f(X,A,B,C)=Ax^2+Bx+C

Thông thường trong thực tế, hàm hồi quy của Y đối với X chưa biết, do đó cần phải ước lượng nó.

Trước hết, để xác định dạng của hàm f, người ta thường căn cứ vào đường hồi quy thực nghiệm, kết hợp với việc phân tích bản chất kinh tế, xã hội hoặc kỹ thuật của vấn đề đang nghiên cứu (Dạng của đường hồi quy thực nghiệm sẽ gợi ý cho ta dạng của đường hồi quy lý thuyết). Sau khi đã xác định dạng của hàm f, ta cần ước lượng cac tham số A, B, C … của hàm đó. Để ước lượng các tham số này, ta sử dụng phương pháp bình phương bé nhất.

Nội dung của phương pháp đó như sau:

Lựa chọn hàm f(X,a,b,c…) để ước lượng hàm f(X,A,B,C…) sao cho: đồ thị của hàm f(X,a,b,c…) “gần gũi” nhất so với đường hồi quy thực nghiệm. Khái niệm “gần gũi” ở đây được hiểu theo nghĩa:

$Q=\sum_{i=1}^{k}n_i\left[\overline{y_{x_i}}-f(x_i,a,b,c...)\right]^2$ đạt cực tiểu

Q là tổng bình phương các sai lệch giữa trung bình có điều kiện thực nghiệm $\overline{y_{x_i}}$ và giá trị của hàm f(x,a,b,c…) tại điểm x=x_i.

Điều kiện cần để Q đạt cực tiểu là: đạo hàm riêng bậc nhất theo các biến a, b, c … bằng 0. Tức là:

$$\begin{cases} \dfrac{\partial Q}{\partial a} = 0 \\ \dfrac{\partial Q}{\partial b} = 0 \\ \dfrac{\partial Q}{\partial c} = 0 \end{cases}$$

Hệ trên gọi là hệ phương trình chuẩn để xác định a,b,c…

Trong nhiều trường hợp, điều kiện cần cũng chính là điều kiện đủ. Giả sử a_0,b_0,c_0 … là nghiệm của hệ trên. Nếu a_0,b_0,c_0 … thỏa mãn điều kiện đủ để Q đạt cực tiểu thì hàm $f(x,a_0,b_0,c_0 \ldots)$ được gọi là hàm hồi quy mẫu được ước lượng bằng phương pháp bình phương bé nhất.

6.4 ƯỚC LƯỢNG MỘT SỐ HÀM HỒI QUY DẠNG ĐƠN GIẢN

I. Ước lượng hàm hồi quy tuyến tính một biến

Giả sử hàm hồi quy của Y đối với X có dạng tuyến tính.

Tức là: E(Y/x)=Ax+B (1)

Dùng phương pháp bình phương bé nhất và dựa vào số liệu của mẫu cụ thể, ta tìm hàm: $\overline{y_x} = ax + b$ (2) để ước lượng hàm (1).

Hàm (2) thường được gọi là hàm hồi quy tuyến tính mẫu.

Từ mẫu cụ thể, ta có các giá trị của X: $x_1,x_2,\ldots,x_k$ với các tần số $n_1, n_2,\ldots,n_k$ tương ứng. Từ các số liệu này ta tính được các giá trị của các trung bình có điều kiện của y: $\overline{y_{x_1}};\overline{y_{x_2}};\ldots;\overline{y_{x_k}}$

Theo phương pháp bình phương bé nhất. Ta phải xác định a,b sao cho:

$$Q = \sum_{i=1}^{k} n_i \left[\overline{y_{x_i}} - (ax_i + b)\right]^2$$ đạt cực tiểu.

Hệ phương trình có dạng:

$$\begin{cases} \dfrac{\partial Q}{\partial a} = 0 \\ \dfrac{\partial Q}{\partial b} = 0 \end{cases} \Leftrightarrow \begin{cases} \left(\sum n_i x_i\right) a + \left(\sum n_i x_i\right) b = \sum n_i x_i \overline{y_{x_i}} \\ \left(\sum n_i x_i\right) a + nb = \sum n_i \overline{y_{x_i}} \end{cases}$$

Giải hệ phương trình ta có công thức tính a và b như sau:

$$a = \frac{\overline{xy} - \overline{x}.\overline{y}}{\hat{s}_x^2}; \quad b = \overline{y} - a.\overline{x}$$

Ví dụ 3 Với bảng số liệu cho ở ví dụ 1.
Hãy ước lượng hàm hồi quy tuyến tính mẫu của Y theo X.
BÀI GIẢI
Theo kết quả tính toán ở các phần trên ta đã tính được
$\overline{x}$=47,5926, $\overline{y}$=159,815, $\sum_i \sum_j n_{ij} x_i y_j = 617400$

Vậy $a = \dfrac{\overline{xy} - \overline{x}.\overline{y}}{\hat{s}_x^2}$=1,006; $b = \overline{y} - a.\overline{x}$ =111,9464.

Nên hàm hồi quy tuyến tính mẫu của Y theo X là:
$\overline{y_x} = 1{,}0058x + 111{,}9464$

Chú ý:

1. Từ công thức tính $r = \dfrac{\overline{xy} - \overline{x}.\overline{y}}{\hat{s}_x.\hat{s}_y}$ ta có thể chứng minh được phương trình sau: $\dfrac{y - \overline{y}}{\hat{s}_y} = r\dfrac{x - \overline{x}}{\hat{s}_x}$.

Suy ra phương trình hồi quy tuyến tính mẫu: $\overline{y_x} = ax + b$

2. Tương tự, ta có hàm hồi quy tuyến tính mẫu của X theo Y là $\overline{x_y} = cy + d$ trong đó các hệ số c và d được tính theo công thức sau

$$c = \frac{\overline{xy} - \overline{x}.\overline{y}}{\hat{s}_y^2}; \qquad d = \overline{x} - c.\overline{y}$$

II. Ứng dụng của hàm hồi quy mẫu

1. Trong thực tế, hàm hồi quy mẫu thường được dùng làm công cụ dự báo.

Ví dụ Quay bảng số liệu ở ví dụ 1

Giả sử trọng lượng của một học sinh ở mức 48kg thì dự đoán chiều cao trung bình của một học sinh sẽ là bao nhiêu?

BÀI GIẢI

Từ hàm hồi quy tuyến tính mẫu của Y theo X là

$$\overline{y_x} = 1{,}0058x + 111{,}9464$$

Ta thay x=48 kg vào phương trình hồi quy ta dự đoán chiều cao trung bình của một học sinh sẽ là

$$\overline{y_{48}} = 1{,}006.48 + 111{,}94 = 160{,}23(cm)$$

2. Ngoài ra, ta có thể dùng hàm hồi quy để kiểm tra số liệu đã có. Chẳng hạn, ở ví dụ 1 nêu trên, trường hợp trong lượng (X) là 45kg. Theo hàm hồi quy mẫu thì trung bình của Y là

$$\overline{y_{45}} = 1{,}006.48 + 111{,}94 = 157{,}21(cm)$$

Nếu giá trị trung bình có điều kiện (với điều kiện X=45) theo số liệu đã cho quá sai lệch so với 157,21 thì ta cần kiểm tra, xem xét lại trường hợp này. Có thể số liệu khi thu thập có sai sót, hoặc có thể có mối quan hệ đặc biệt giữa X và Y trong trường hợp này khác với xu thế chung.

BÀI TẬP CHƯƠNG VI

6.1. Cho tương quan cặp của X, Y như sau

X	1	2	3	4	4	1	2	4	4	1	1
Y	3	5	1	3	3	3	6	2	6	1	3

Hãy chuyển từ tương quan cặp sang tương quan bảng.

6.2. Số liệu thống kê về tỷ lệ ngân sách chi cho giáo dục (X) và tỷ lệ tăng thu nhập quốc dân (Y) của một số nước được cho ở bảng sau

Xi (%)	8	10	12	9	14	15	10	11	12	13
Yi (%)	4	5	6	5	8	9	6	6	10	7

Tìm phương trình hồi quy tuyến tính mẫu của tỷ lệ tăng thu nhập quốc dân theo tỷ lệ ngân sách chi cho giáo dục.

6.3. Giả sử 2 đại lượng X và Y có bảng tương quan như sau

X \ Y	1	2
3	10	20
4	20	30

a) Tính các tổng.
b) Tính hệ số tương quan.
c) Tìm hàm hồi quy tuyến tính mẫu của Y theo X
d) Tìm hàm hồi quy tuyến tính mẫu của X theo Y

Hướng dẫn

a) Lập bảng sau

X \ Y	1	2	n_i	x_in_i	$x_i^2n_i$
3	10	20	30	90	270
	30	120			
4	20	30	50	200	800
	80	240			
m_j	30	50	n=80	$\sum x_i ni = 290$	$\sum x_i^2 n_i = 1070$
m_jy_j	30	100	$\sum m_j y_j = 130$		
$m_jy_j^2$	30	200	$\sum m_i y_j^2 = 230$		$\sum n_{ij} x_i y_j = 470$

b) r_{xy} = - 0,0667

c) Hàm hồi quy tuyến tính mẫu của Y theo X : $\overline{y_x} = ax + b$

Với $\overline{x}$ =3,625; $\overline{y}$ =1,625; $\hat{s}_X^2$ =0,2373; $\hat{s}_Y^2$ =0.2373

$a = \dfrac{\overline{xy} - \overline{x}.\overline{y}}{\hat{s}_X^2}$= - ,0667; $b = \overline{y} - a.\overline{x}$ =1,8668;

Hàm hồi quy tuyến tính mẫu của X theo Y: $\overline{x_y} = cy + d$,

$c = \dfrac{\overline{xy} - \overline{x}.\overline{y}}{\hat{s}_Y^2}$= - 0,0667 ; $d = \overline{x} + c.\overline{y}$ =3,7334

ĐỀ THI THAM KHẢO

ĐỀ THI KẾT THÚC HỌC PHẦN
Môn thi: Xác suất thống kê
Thời gian: 75 phút. ĐỀ THI SỐ 1
(Sinh viên chỉ được sử dụng các bảng số thống kê)
(ĐỀ THI CÁC LỚP - HỆ CĐ)

***Câu 1* (3 điểm)**

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

a) Lập bảng phân phối xác suất cho X.

b) Tính kỳ vọng cho Y=3X+D(X).

***Câu 2* (3 điểm)**

a) Việc tiêu dùng điện hàng tháng của các hộ gia đình ở TPHCM là biến ngẫu nhiên phân phối chuẩn với trung bình là 200 KWh và độ lệch chuẩn là 40KWh.
Tính xác suất để chọn ngẫu nhiên một hộ gia đình thì hộ đó có mức tiêu dùng điện hàng tháng dưới 180KWh.

b) Trong 100 bóng đèn có 40 bóng hỏng. Tìm xác suất để lấy được 3 bóng hỏng trong 5 bóng được kiểm tra ngẫu nhiên.

***Câu 3* (4 điểm)**

Theo dõi mức nguyên liệu hao phí để sản xuất ra 1 đơn vị sản phẩm ở một nhà máy, người ta thu được các số liệu quan sát sau:

Mức nguyên liệu hao phí(gam/sp)	28	29	30	31	32
Số sản phẩm	3	11	17	11	8

a) Tìm khoảng ước lượng về số tiền trung bình dùng để mua loại nguyên liệu này trong từng quý của nhà máy, với độ tin cậy 98%.
(biết giá loại nguyên liệu này là 600 ngàn đồng/gam và sản lượng của nhà máy trong 1 quý là 50.000 sản phẩm).

b) Trước đây, mức hao phí loại nguyên liệu này trung bình là 31 gam/sản phẩm. Số liệu của mẫu trên được thu thập sau khi nhà máy áp dụng một công nghệ sản xuất mới. Hãy cho nhận xét về công nghệ sản xuất mới (với mức ý nghĩa 4%) ?

ĐỀ THI KẾT THÚC HỌC PHẦN

Môn thi: Xác suất thống kê

Thời gian: 75 phút. ĐỀ THI SỐ 2

(Sinh viên chỉ được sử dụng các bảng số thống kê)

(ĐỀ THI CÁC LỚP - HỆ CĐ)

Câu 1 (3 điểm) Có 8 lô hàng trong đó có:

2 lô hàng loại A: mỗi lô hàng 6 sản phẩm loại 1 và 4 sản phẩm loại 2; 3 lô hàng loại B: mỗi lô hàng 7 sản phẩm loại 1 và 3 sản phẩm loại 2; 3 lô hàng loại C: mỗi lô hàng 8 sản phẩm loại 1 và 2 sản phẩm loại 2.

Lấy ngẫu nhiên một lô hàng và từ đó lấy ngẫu nhiên 2 sản phẩm
Gọi X là số sản phẩm loại 1 có trong 2 sản phẩm lấy ra .

a) Hãy lập bảng phân phối xác suất cho X?
b) Tính kì vọng của Y=5X+D(X).

Câu 2 (3 điểm)

a) Một trạm điện thoại có 1000 đầu dây. Xác suất để tại một thời điểm một nơi bất kỳ gọi đến trạm là 0,004. Tính xác suất để tại thời điểm đó có không quá 2 nơi gọi đến

b) Trong một lô hàng có 800 sản phẩm loại 1 và 200 sản phẩm loại 2. Lấy ngẫu nhiên ra 5 sản phẩm theo phương thức có hoàn lại. Tìm số sản phẩm loại 1 trung bình lấy được trong 5 sản phẩm lấy ra.

Câu 3 (4 điểm)

Một công ty tiến hành khảo sát nhu cầu tiêu dùng về 1 loại sản phẩm do công ty sản xuất. Khảo sát một số hộ gia đình ở 1 thành phố ta được bảng số liệu:

Sốlượng(kg/tháng)	1-2	2-3	3-4	4-5	5-6	6-7
Số hộ	20	30	80	60	40	10

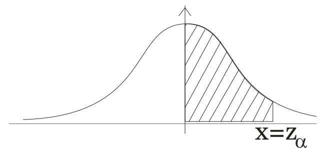
a) Hãy ước lượng số sản phẩm của công ty được tiêu thụ tại thành phố trong 1 tháng, với độ tin cậy 95%. Cho biết tổng số hộ gia đình trong toàn thành phố là 500.000 hộ.

b) Một tài liệu cho rằng: mức tiêu thụ trung bình của loại sản phẩm này ở thành phố là 1750 tấn/tháng thì có chấp nhận được không (với mức ý nghĩa 1%)?

PHỤ LỤC 1
CÁC BẢNG PHÂN PHỐI XÁC SUẤT

BẢNG 1 PHÂN PHỐI GAUSS

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_0^x e^{-t^2/2}dt = P\left(0 \le X \le x\right) \equiv \alpha$$



Với $X \sim N\left(0;1\right)$, $x \equiv z_\alpha$.

	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974

2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998
3.5	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998
3.6	0.4998	0.4998	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.7	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.8	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.9	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000
4.0	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000

Ví dụ: $\Phi(0,55) = 0.2088$; $\Phi(1.65) = 0.4505$

$\Phi(1.96) = 0.475$

$\Phi(-2,58) = -\Phi(2,58) = -0,4951$

Chú ý: Nếu $x > 4,09$ thì lấy $\Phi(x) = 0,5$

Ngược lại: Cho $\gamma = 0.95$, $\Phi(t_\alpha) = \frac{\gamma}{2} = 0.475$. Tra bảng, ta thấy giá trị 0.475 nằm ở hàng 1.9, cột 0.06, điều này có nghĩa là $\Phi(1.96) = 0.475$. Do đó $t_\alpha = 1.96$.

BẢNG 2 ***Hàm mật độ Gauss:*** $\varphi(x) = \dfrac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$

	0	1	2	3	4	5	6	7	8	9
0.0	0.3989	0.3989	0.3989	0.3988	0.3986	0.3984	0.3982	0.3980	0.3977	0.3973
0.1	0.3970	0.3965	0.3961	0.3956	0.3951	0.3945	0.3939	0.3932	0.3925	0.3918
0.2	0.3910	0.3902	0.3894	0.3885	0.3876	0.3867	0.3857	0.3847	0.3836	0.3825
0.3	0.3814	0.3802	0.3790	0.3778	0.3765	0.3752	0.3739	0.3725	0.3712	0.3697
0.4	0.3683	0.3668	0.3653	0.3637	0.3621	0.3605	0.3589	0.3572	0.3555	0.3538
0.5	0.3521	0.3503	0.3485	0.3467	0.3448	0.3429	0.3410	0.3391	0.3372	0.3352
0.6	0.3332	0.3312	0.3292	0.3271	0.3251	0.3230	0.3209	0.3187	0.3166	0.3144
0.7	0.3123	0.3101	0.3079	0.3056	0.3034	0.3011	0.2989	0.2966	0.2943	0.2920
0.8	0.2897	0.2874	0.2850	0.2827	0.2803	0.2780	0.2756	0.2732	0.2709	0.2685
0.9	0.2661	0.2637	0.2613	0.2589	0.2565	0.2541	0.2516	0.2492	0.2468	0.2444
1.0	0.2420	0.2396	0.2371	0.2347	0.2323	0.2299	0.2275	0.2251	0.2227	0.2203
1.1	0.2179	0.2155	0.2131	0.2107	0.2083	0.2059	0.2036	0.2012	0.1989	0.1965
1.2	0.1942	0.1919	0.1895	0.1872	0.1849	0.1826	0.1804	0.1781	0.1758	0.1736
1.3	0.1714	0.1691	0.1669	0.1647	0.1626	0.1604	0.1582	0.1561	0.1539	0.1518
1.4	0.1497	0.1476	0.1456	0.1435	0.1415	0.1394	0.1374	0.1354	0.1334	0.1315
1.5	0.1295	0.1276	0.1257	0.1238	0.1219	0.1200	0.1182	0.1163	0.1145	0.1127
1.6	0.1109	0.1092	0.1074	0.1057	0.1040	0.1023	0.1006	0.0989	0.0973	0.0957
1.7	0.0940	0.0925	0.0909	0.0893	0.0878	0.0863	0.0848	0.0833	0.0818	0.0804
1.8	0.0790	0.0775	0.0761	0.0748	0.0734	0.0721	0.0707	0.0694	0.0681	0.0669
1.9	0.0656	0.0644	0.0632	0.0620	0.0608	0.0596	0.0584	0.0573	0.0562	0.0551
2.0	0.0540	0.0529	0.0519	0.0508	0.0498	0.0488	0.0478	0.0468	0.0459	0.0449
2.1	0.0440	0.0431	0.0422	0.0413	0.0404	0.0396	0.0387	0.0379	0.0371	0.0363
2.2	0.0355	0.0347	0.0339	0.0332	0.0325	0.0317	0.0310	0.0303	0.0297	0.0290
2.3	0.0283	0.0277	0.0270	0.0264	0.0258	0.0252	0.0246	0.0241	0.0235	0.0229
2.4	0.0224	0.0219	0.0213	0.0208	0.0203	0.0198	0.0194	0.0189	0.0184	0.0180
2.5	0.0175	0.0171	0.0167	0.0163	0.0158	0.0154	0.0151	0.0147	0.0143	0.0139
2.6	0.0136	0.0332	0.0129	0.0126	0.0122	0.0119	0.0116	0.0113	0.0110	0.0107
2.7	0.0104	0.0101	0.0099	0.0096	0.0093	0.0091	0.0088	0.0086	0.0084	0.0081
2.8	0.0079	0.0077	0.0075	0.0073	0.0071	0.0069	0.0067	0.0065	0.0063	0.0061
2.9	0.0060	0.0058	0.0056	0.0055	0.0053	0.0051	0.0050	0.0048	0.0047	0.0046
3.0	0.0044	0.0043	0.0042	0.0040	0.0039	0.0038	0.0037	0.0036	0.0035	0.0034
3.1	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026	0.0025	0.0025
3.2	0.0024	0.0023	0.0022	0.0022	0.0021	0.0020	0.0020	0.0019	0.0018	0.0018
3.3	0.0017	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014	0.0013	0.0013
3.4	0.0012	0.0012	0.0012	0.0011	0.0011	0.0010	0.0010	0.0010	0.0009	0.0009
3.5	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007	0.0007	0.0007	0.0006
3.6	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005	0.0004
3.7	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003	0.0003	0.0003	0.0003
3.8	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002	0.0002	0.0002	0.0002	0.0002
3.9	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0001	0.0001
4.0	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001

Ví dụ: $\varphi(1,09) = 0,2203$

$\varphi(-2,80) = \varphi(2,80) = 0,0079$

Chú ý: Nếu $x > 4,09$ thì lấy $\varphi(x) = 0,0001$

BẢNG 3 ***Phân vị*** $|u|$***:*** $P(|U|<t_\alpha)=\gamma$ **với** $U \sim N(0,1)$**.**

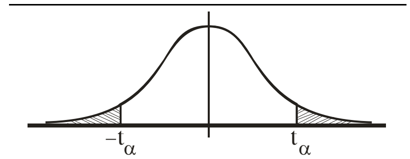
γ	t_α	γ	t_α	γ	t_α	γ	t_α
0.50	0.6745	0.710	1.0581	0.910	1.6954	0.979	2.3080
0.51	0.6903	0.720	1.0803	0.915	1.7224	0.980	2.3263
0.52	0.7063	0.730	1.1031	0.920	1.7507	0.981	2.3455
0.53	0.7225	0.740	1.1264	0.925	1.7805	0.982	2.3656
0.54	0.7388	0.750	1.1503	0.930	1.8119	0.983	2.3867
0.55	0.7554	0.760	1.1750	0.935	1.8453	0.984	2.4089
0.56	0.7722	0.770	1.2004	0.940	1.8808	0.985	2.4324
0.57	0.7892	0.780	1.2265	0.945	1.9189	0.986	2.4573
0.58	0.8064	0.790	1.2536	0.950	1.9600	0.987	2.4838
0.59	0.8239	0.800	1.2816	0.955	2.0047	0.988	2.5121
0.60	0.8416	0.810	1.3106	0.960	2.0537	0.989	2.5427
0.61	0.8596	0.820	1.3408	0.965	2.1084	0.990	2.5758
0.62	0.8779	0.830	1.3722	0.970	2.1701	0.991	2.6121
0.63	0.8965	0.840	1.4015	0.971	2.1835	0.992	2.6521
0.64	0.9154	0.850	1.4395	0.972	2.1973	0.993	2.6968
0.65	0.9346	0.860	1.4758	0.973	2.2115	0.994	2.7478
0.66	0.9542	0.870	1.5141	0.974	2.2262	0.995	2.8071
0.67	0.9741	0.880	1.5548	0.975	2.2414	0.996	2.8782
0.68	0.9945	0.890	1.5982	0.976	2.2571	0.997	2.9677
0.69	1.0152	0.900	1.6449	0.977	2.2734	0.998	3.0902
0.70	1.0364	0.905	1.6696	0.978	2.2904	0.999	3.2905

Ví dụ:

$$\gamma = 0,90 \Rightarrow t_\alpha = 1,6449 \approx 1,65$$
$$\gamma = 0,95 \Rightarrow t_\alpha = 1,9600$$
$$\gamma = 0,99 \Rightarrow t_\alpha = 2,5758 \approx 2,58$$

BẢNG 4 PHÂN PHỐI STUDENT

$$P\left(|T| \geq t_\alpha\right) = \alpha \text{ với } T \sim St(n)$$



Cột 1 : giá trị độ tự do n.

Hàng 1 : Giá trị nguy cơ sai lầm α

Nội dung bảng : Giá trị t_α tương ứng với n và α

	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2
1	63.66	31.82	21.21	15.89	12.71	10.58	9.058	7.916	7.026	6.314	4.165	3.078
2	9.925	6.965	5.643	4.849	4.303	3.896	3.578	3.320	3.104	2.920	2.282	1.886
3	5.841	4.541	3.896	3.482	3.182	2.951	2.763	2.605	2.471	2.353	1.924	1.638
4	4.604	3.747	3.298	2.999	2.776	2.601	2.456	2.333	2.226	2.132	1.778	1.533
5	4.032	3.365	3.003	2.757	2.571	2.422	2.297	2.191	2.098	2.015	1.699	1.476
6	3.707	3.143	2.829	2.612	2.447	2.313	2.201	2.104	2.019	1.943	1.650	1.440
7	3.499	2.998	2.715	2.517	2.365	2.241	2.136	2.046	1.966	1.895	1.617	1.415
8	3.355	2.896	2.634	2.449	2.306	2.189	2.090	2.004	1.928	1.860	1.592	1.397
9	3.250	2.821	2.574	2.398	2.262	2.150	2.055	1.973	1.899	1.833	1.574	1.383
10	3.169	2.764	2.527	2.359	2.228	2.120	2.028	1.948	1.877	1.812	1.559	1.372
11	3.106	2.718	2.491	2.328	2.201	2.096	2.007	1.928	1.859	1.796	1.548	1.363
12	3.055	2.681	2.461	2.303	2.179	2.076	1.989	1.912	1.844	1.782	1.538	1.356
13	3.012	2.650	2.436	2.282	2.160	2.060	1.974	1.899	1.832	1.771	1.530	1.350
14	2.977	2.624	2.415	2.264	2.145	2.046	1.962	1.887	1.821	1.761	1.523	1.345
15	2.947	2.602	2.397	2.249	2.131	2.034	1.951	1.878	1.812	1.753	1.517	1.341
16	2.921	2.583	2.382	2.235	2.120	2.024	1.942	1.869	1.805	1.746	1.512	1.337
17	2.898	2.567	2.368	2.224	2.110	2.015	1.934	1.862	1.798	1.740	1.508	1.333
18	2.878	2.552	2.356	2.214	2.101	2.007	1.926	1.855	1.792	1.734	1.504	1.330
19	2.861	2.539	2.346	2.205	2.093	2.000	1.920	1.850	1.786	1.729	1.500	1.328
20	2.845	2.528	2.336	2.197	2.086	1.994	1.914	1.844	1.782	1.725	1.497	1.325
21	2.831	2.518	2.328	2.189	2.080	1.988	1.909	1.840	1.777	1.721	1.494	1.323
22	2.819	2.508	2.320	2.183	2.074	1.983	1.905	1.835	1.773	1.717	1.492	1.321
23	2.807	2.500	2.313	2.177	2.069	1.978	1.900	1.832	1.770	1.714	1.489	1.319
24	2.797	2.492	2.307	2.172	2.064	1.974	1.896	1.828	1.767	1.711	1.487	1.318
25	2.787	2.485	2.301	2.167	2.060	1.970	1.893	1.825	1.764	1.708	1.485	1.316
26	2.779	2.479	2.296	2.162	2.056	1.967	1.890	1.822	1.761	1.706	1.483	1.315
27	2.771	2.473	2.291	2.158	2.052	1.963	1.887	1.819	1.758	1.703	1.482	1.314
28	2.763	2.467	2.286	2.154	2.048	1.960	1.884	1.817	1.756	1.701	1.480	1.313
29	2.756	2.462	2.282	2.150	2.045	1.957	1.881	1.814	1.754	1.699	1.479	1.311

Ví dụ: Với $T \sim St(10)$, để tìm t_α sao cho $P(|X| > t_\alpha) = 0.05$, nội dung bảng ứng với hàng 10, cột 0.05 cho giá trị $t_\alpha = 2.228$.

Cụ thể: Biết γ tìm t_α^{n-1}**:** $n = 11$**,** $\gamma = 0,95 \rightarrow t_{0,05}^{10} = 2,228$

Biết t_α^{n-1} tìm γ**:** $n = 20$**,** $t_\alpha^{n-1} = t_\alpha^{19} = 2,325 \rightarrow \gamma = 0,97$

Trường hợp tìm t_α sao cho $P(|T| \le t_\alpha) = \gamma$ được khảo sát giống như trường hợp phân phối Gauss

$$P(|T| > t_\alpha) = 1 - P(|T| \le t_\alpha) = 1 - \gamma.$$

Chú ý: Khi $T \sim St(n)$, với $n \ge 30$, ta có thể

– Dùng bảng phân phối Student với độ tự do $n = \infty$ (hàng cuối), hay

– Xấp xỉ phân phối Student bằng phân phối Gauss, nghĩa là $X \sim N(0,1)$.

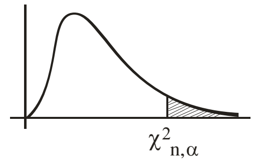
BẢNG 5: PHÂN PHỐI CHI – BÌNH PHƯƠNG

$P\left(X \geq \chi^2_{n,\alpha}\right) = \alpha$ khi $X \sim \chi^2(n)$

Hàng 1 : Giá trị của α.

Cột 1 : Giá trị độ tự do n.

Nội dung bảng : Giá trị $\chi^2_{n,\alpha}$.



	0.005	0.01	0.015	0.02	0.025	0.03	0.05	0.95	0.975	0.98	0.99	0.995
1	7.879	6.635	5.916	5.412	5.024	4.709	3.841	0.004	0.001	0.001	0.000	0.000
2	10.597	9.210	8.399	7.824	7.378	7.013	5.991	0.103	0.051	0.040	0.020	0.010
3	12.838	11.345	10.465	9.837	9.348	8.947	7.815	0.352	0.216	0.185	0.115	0.072
4	14.860	13.277	12.339	11.668	11.143	10.712	9.488	0.711	0.484	0.429	0.297	0.207
5	16.750	15.086	14.098	13.388	12.832	12.375	11.070	1.145	0.831	0.752	0.554	0.412
6	18.548	16.812	15.777	15.033	14.449	13.968	12.592	1.635	1.237	1.134	0.872	0.676
7	20.278	18.475	17.398	16.622	16.013	15.509	14.067	2.167	1.690	1.564	1.239	0.989
8	21.955	20.090	18.974	18.168	17.535	17.011	15.507	2.733	2.180	2.032	1.647	1.344
9	23.589	21.666	20.512	19.679	19.023	18.480	16.919	3.325	2.700	2.532	2.088	1.735
10	25.188	23.209	22.021	21.161	20.483	19.922	18.307	3.940	3.247	3.059	2.558	2.156
11	26.757	24.725	23.503	22.618	21.920	21.342	19.675	4.575	3.816	3.609	3.053	2.603
12	28.300	26.217	24.963	24.054	23.337	22.742	21.026	5.226	4.404	4.178	3.571	3.074
13	29.819	27.688	26.403	25.471	24.736	24.125	22.362	5.892	5.009	4.765	4.107	3.565
14	31.319	29.141	27.827	26.873	26.119	25.493	23.685	6.571	5.629	5.368	4.660	4.075
15	32.801	30.578	29.235	28.259	27.488	26.848	24.996	7.261	6.262	5.985	5.229	4.601
16	34.267	32.000	30.629	29.633	28.845	28.191	26.296	7.962	6.908	6.614	5.812	5.142
17	35.718	33.409	32.011	30.995	30.191	29.523	27.587	8.672	7.564	7.255	6.408	5.697
18	37.156	34.805	33.382	32.346	31.526	30.845	28.869	9.390	8.231	7.906	7.015	6.265
19	38.582	36.191	34.742	33.687	32.852	32.158	30.144	10.117	8.907	8.567	7.633	6.844
20	39.997	37.566	36.093	35.020	34.170	33.462	31.410	10.851	9.591	9.237	8.260	7.434
21	41.401	38.932	37.4B4	36.343	35.479	34.759	32.671	11.591	10.283	9.915	8.897	8.034
22	42.796	40.289	38.768	37.659	36.781	36.049	33.924	12.338	10.982	10.600	9.542	8.643
23	44.181	41.638	40.094	38.968	38.076	37.332	35.172	13.091	11.689	11.293	10.196	9.260
24	45.558	42.980	41.413	40.270	39.364	38.609	36.415	13.848	12.401	11.992	10.856	9.886
25	46.928	44.314	42.725	41.566	40.646	39.880	37.652	14.611	13.120	12.697	11.524	10.520
26	48.290	45.642	44.031	42.856	41.923	41.146	38.885	15.379	13.844	13.409	12.198	11.160
27	49.645	46.963	45.331	44.140	43.195	42.407	40.113	16.151	14.573	14.125	12.878	11.808
28	50.994	48.278	46.626	45.419	44.461	43.662	41.337	16.928	15.308	14.847	13.565	12.461
29	52.335	49.588	47.915	46.693	45.722	44.913	42.557	17.708	16.047	15.574	14.256	13.121
30	53.672	50.892	49.199	47.962	46.979	46.160	43.773	18.493	16.791	16.306	14.953	13.787
35	60.275	57.342	55.553	54.244	53.203	52.335	49.802	22.465	20.569	20.027	18.509	17.192
40	66.766	63.691	61.812	60.436	59.342	58.428	55.758	26.509	24.433	23.838	22.164	20.707
45	73.166	69.957	67.994	66.555	65.410	64.454	61.656	30.612	28.366	27.720	25.901	24.311

50	79.490	76.154	74.111	72.613	71.420	70.423	67.505	34.764	32.357	31.664	29.707	27.991
55	85.749	82.292	80.173	78.619	77.380	76.345	73.311	38.958	36.398	35.659	33.571	31.735
60	91.952	88.379	86.188	84.580	83.298	82.225	79.082	43.188	40.482	39.699	37.485	35.534
65	98.105	94.422	92.161	90.501	89.177	88.069	84.821	47.450	44.603	43.779	41.444	39.383
70	104.215	100.425	98.098	96.387	95.023	93.881	90.531	51.739	48.758	47.893	45.442	43.275
75	110.285	106.393	104.001	102.243	100.839	99.665	96.217	56.054	52.942	52.039	49.475	47.206
80	116.321	112.329	109.874	108.069	106.629	105.422	101.879	60.391	57.153	56.213	53.540	51.172
85	122.324	118.236	115.720	113.871	112.393	111.156	107.522	64.749	61.389	60.412	57.634	55.170
90	128.299	124.116	121.542	119.648	118.136	116.869	113.145	69.126	65.647	64.635	61.754	59.196
95	134.247	129.973	127.341	125.405	123.858	122.562	118.752	73.520	69.925	68.879	65.898	63.250
100	140.170	135.807	133.120	131.142	129.561	128.237	124.342	77.929	74.222	73.142	70.065	67.328

Ví dụ: Do $P(X < a) = 1 - P(X \ge a) = 1 - P(X > a)$ nên để giải các bài toán này, người ta tính sẵn một số giá trị x sao cho $P(X \ge x) = \alpha$, $X \sim \chi^2(n)$.

Với $X \sim \chi^2(10)$, để tìm x sao cho $P(X \ge x) = 0.025$, tra bảng ứng với hàng 10, cột 0.025, ta được $x = 20.483$.

$n = 14$, $\alpha = 5\%$

$$\chi^2_{\alpha/2}(n) = \chi^2_{0,025}(14) = 5,629\,;$$

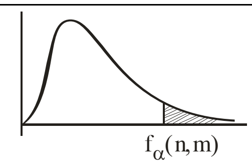
$$\chi^2_{1-\frac{\alpha}{2}}(n) = \chi^2_{0,975}(14) = 26,119$$

BẢNG 6: PHÂN PHỐI FISHER

$P\left(X \geq f_{\alpha}(n,m)\right) = \alpha$ khi $X \sim F(n,m)$

Hàng 1 : Giá trị của độ tự do (tử số) n. ***Cột 1 :*** Giá trị độ tự do (mẫu số) m.

Nội dung bảng : Giá trị $f_{\alpha}(n,m)$.



Trường hợp : $\alpha = 0.05$

	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161	200	216	225	230	234	237	239	241	242	244	246	248	249	250	251	252	253	254
2	18.51	19	19.16	19.25	19.3	19.33	19.35	19.37	19.38	19.4	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.5
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.7	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	5.96	5.91	5.86	5.8	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	4.5	4.46	4.43	4.4	4.37
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.1	4.06	4	3.94	3.87	3.84	3.81	3.77	3.74	3.7	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	3.3	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.5	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	2.9	2.86	2.83	2.79	2.75	2.71
10	4.96	4.1	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.7	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.2	3.09	3.01	2.95	2.9	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.4
12	4.75	3.89	3.49	3.26	3.11	3	2.91	2.85	2.8	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.3
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.6	2.53	2.46	2.42	2.38	2.34	2.3	2.25	2.21
14	4.6	3.74	3.34	3.11	2.96	2.85	2.76	2.7	2.65	2.6	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	2.9	2.79	2.71	2.64	2.59	2.54	2.48	2.4	2.33	2.29	2.25	2.2	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	3.2	2.96	2.81	2.7	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.1	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	2.9	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	3.1	2.87	2.71	2.6	2.51	2.45	2.39	2.35	2.28	2.2	2.12	2.08	2.04	1.99	1.95	1.9	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	2.1	2.05	2.01	1.96	1.92	1.87	1.81
22	4.3	3.44	3.05	2.82	2.66	2.55	2.46	2.4	2.34	2.3	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	2.8	2.64	2.53	2.44	2.37	2.32	2.27	2.2	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.4	3.01	2.78	2.62	2.51	2.42	2.36	2.3	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	2.6	2.49	2.4	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
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	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4	3.15	2.76	2.53	2.37	2.25	2.17	2.1	2.04	1.99	1.92	1.84	1.75	1.7	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.5	1.43	1.35	1.25
∞	3.84	3	2.6	2.37	2.21	2.1	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

Trường hợp: $\alpha = 0.01$

	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4052	4999	5404	5624	5764	5859	5928	5981	6022	6056	6107	6157	6209	6234	6260	6286	6313	6340	6366
2	98.5	99.0	99.1	99.2	99.3	99.3	99.3	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.5
3	34.1	30.8	29.4	28.7	28.2	27.9	27.6	27.4	27.3	27.2	27.0	26.8	26.6	26.6	26.5	26.4	26.3	26.2	26.1
4	21.2	18.0	16.6	15.9	15.5	15.2	14.9	14.8	14.6	14.5	14.3	14.2	14.0	13.9	13.8	13.7	13.6	13.5	13.5
5	16.2	13.2	12.0	11.3	10.9	10.6	10.4	10.2	10.1	10.0	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	13.7	10.9	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	12.2	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	11.2	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	4.86
9	10.5	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	10.0	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	3.6
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
15	8.68	6.36	5.42	4.89	4.56	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	5.19	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	2.49

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
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	1.8
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	1.6
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

Tương tự như phân phối chi-bình phương, với $X \sim F(n,m)$, các giá trị x sao cho $P(X \ge x) = \alpha$ được tính sẵn với một số α, m, n cho trước.

Ví dụ: nếu $X \sim F(5,10)$, để tìm x sao cho $P(X \ge x) = 0.05$, ta tra bảng (Trường hợp $\alpha = 0,05$), hàng 10, cột 5 và nhận được giá trị của x là 3.33

Để tìm x sao cho $P(X \ge x) = 0.01$, tra bảng (Trường hợp $\alpha = 0,01$), hàng 10, cột 5, ta được giá trị của x là 5.64

PHỤ LỤC 2
HƯỚNG DẪN DÙNG CÁC BẢNG PHÂN PHỐI XÁC SUẤT

Các bảng phân phối xác suất quan trọng gồm phân phối Gauss, Chi-Bình phương, Student và Fisher. Các giá trị xác suất đặc biệt của chúng được tính sẵn và liệt kê thành bảng như sau

BẢNG 1. ***Phân phối Gauss*** $N(0,1)$

Với $X \sim N(0,1)$, ta có hai bài toán xác suất quan trọng :

– Tìm $P(a \le X \le b)$, với $a, b \in \mathbb{R}$, $a \le b$ cho trước,

– Tìm giá trị t_α sao cho $P(-t_\alpha \le X \le t_\alpha) = P(|X| \le t_\alpha) = \gamma$, với γ cho trước.

I. Tìm $P(a \le X \le b)$

Do $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ là hàm mật độ của X nên từ tính chất của tích phân, ta có

$$P(a \le X \le b) = \int_a^b f(x)dx = \int_0^b f(x)dx - \int_0^a f(x)dx$$
$$\equiv \Phi(b) - \Phi(a),$$

trong đó $\Phi(x) = \int_0^x f(t)dt \equiv \frac{1}{\sqrt{2\pi}}\int_0^x e^{-t^2/2}dt$ được gọi là hàm Laplace.

Các giá trị của hàm Laplace được tính sẵn và liệt kê thành bảng gọi là bảng phân phối Gauss.

Ngoài ra, vì $\Phi(x)$ là hàm lẻ: $\Phi(-x) = -\Phi(x)$, $\forall x \in \mathbb{R}$, nên người ta chỉ cần liệt kê các giá trị của $\Phi(x)$ với $x > 0$.

Bảng phân phối Gauss gồm 400 giá trị của $\Phi(x)$, với x thay đổi từ 0.00, 0.01, 0.02, ..., 3.99 và được bố trí như sau

– Các hàng trong bảng, trừ hàng đầu, được đánh số từ 0.0, 0.1, đến 3.9.

– Các cột trong bảng, trừ cột đầu, được đánh số từ 0.00, 0.01 tới 0.09.

Khi đó, ứng với mỗi giá trị x trong khoảng từ 0.00 đến 3.99 với hai số lẻ thập phân dạng $x = a.bc$, giá trị $\Phi(x)$ nằm ở hàng đánh số a.b và cột đánh số 0.0c. Chẳng hạn, với $x = 1.52$, $\Phi(x) \equiv \Phi(1.52)$ nằm ở hàng 1.5, cột 0.02, nghĩa là $\Phi(1.52) = 0.4357$.

	...	...	0.01	***0.02***	0.03	...	...
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
1.4	...	...	0.4207	0.4222	0.4236	...	...
1.5	...	...	0.4345	**0.4357**	0.4370	...	...
1.6	...	...	0.4463	0.4474	0.4484	...	...
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...

Ngoài ra, vì $\Phi(+\infty) \equiv \dfrac{1}{\sqrt{2\pi}}\displaystyle\int_0^{+\infty} e^{-t^2/2}dt = 0.5$ nên $\Phi(x) = 0.5$, với $4 \le x \le +\infty$.

Tóm lại, khi $X \sim N(0,1)$ thì

$$P\left(a \le X \le b\right) = \Phi(b) - \Phi(a), \qquad (1)$$

với $\Phi(x)$ được tính như sau

– Khi $0 \le x \le 3.99$, giá trị $\Phi(x)$ được tìm thấy trong bảng,

– Khi $4 \le x \le +\infty$, $\Phi(x) = 0.5$,

– Khi $x < 0$, ta dùng công thức $\Phi(-x) = -\Phi(x)$.

Chú ý: Công thức (1) vẫn đúng cho trường hợp $a = -\infty$ và hay $b = +\infty$. Chẳng hạn

$$P\left(-\infty < X \le b\right) = \Phi(b) - \Phi(-\infty) = \Phi(b) + 0,5,$$

vì $\quad \Phi(-\infty) = -\Phi(+\infty) = -0,5,$

$$P\left(a \le X < +\infty\right) = \Phi(+\infty) - \Phi(a) = 0,5 - \Phi(a).$$

Hơn nữa, do tính chất của tích phân, ta có

$$P\left(a \le X \le b\right) = P\left(a < X \le b\right) = P\left(a \le X < b\right) = P\left(a < X < b\right).$$

II. Tìm t_α sao cho $P\left(-t_\alpha \le X \le t_\alpha\right) = P\left(\left|X\right| \le t_\alpha\right) = \gamma$, với γ cho trước

Xuất phát từ đẳng thức $\gamma = P\left(-t_\alpha \le X \le t_\alpha\right) = 2\Phi\left(t_\alpha\right)$, ta được $\Phi\left(t_\alpha\right) = \dfrac{\gamma}{2}$.

Do đó, ứng với giá trị γ cho trước, tính $\Phi\left(t_\alpha\right) = \dfrac{\gamma}{2}$ và tìm vị trí của số hạng này trong bảng. Bấy giờ, t_α chính là tổng của số chỉ hàng và số chỉ cột. Chẳng hạn, với $\gamma = 0.95$, $\Phi\left(t_\alpha\right) = \dfrac{\gamma}{2} = 0.475$. Tra bảng, ta thấy giá trị 0.475 nằm ở hàng 1.9, cột 0.06, điều này có nghĩa là $\Phi(1.96) = 0.475$. Do đó $t_\alpha = 1.96$.

	...	...	0.05	***0.06***	0.07	...	...
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
1.8	...	...	0.4678	0.4686	0.4693	...	...
1.9	...	...	0.4744	**0.4750**	0.4756	...	...
2.0	...	...	0.4798	0.4803	0.4808	...	...
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...

Chú ý: Ta còn gặp bài toán tìm t_α sao cho $P\left(\left|X\right| > t_\alpha\right) = \alpha$, với α cho trước. Khi đó

$$P\left(\left|X\right| \le t_\alpha\right) = 1 - P\left(\left|X\right| > t_\alpha\right) = 1 - \alpha = \gamma$$

và ta nhận trở lại bài toán vừa khảo sát. Thông thường, γ và α lần lượt được gọi là độ tin cậy và nguy cơ sai lầm.

Hoặc dùng BẢNG 3 CHO ĐƠN GIẢN.

BẢNG 4. Phân phối Student $St(n)$

Do phân phối Student thường chỉ dùng trong các bài toán thống kê nên với $T \sim St(n)$, người ta chỉ có các nhu cầu

- Tìm t_α sao cho $P(|T| \le t_\alpha) = \gamma$,
- Tìm t_α sao cho $P(|T| > t_\alpha) = \alpha$.

Để làm được điều này, người ta tính sẵn $P(|T| > t_\alpha) = \alpha$, $T \sim St(n)$, với một số giá trị của (nguy cơ sai lầm) α và (độ tự do) n và liệt kê trong bảng gọi là bảng phân phối Student.

Cụ thể, các hàng của bảng, trừ hàng 1, được đánh số theo độ tự do n, các cột của bảng, trừ cột 1, được đánh số theo (nguy cơ sai lầm) α. Khi đó, nội dung trong bảng ứng với hàng và cột nhận được chính là giá trị t_α cần tìm.

Ví dụ Với $T \sim St(10)$, để tìm t_α sao cho $P(|X| > t_\alpha) = 0.05$, nội dung bảng ứng với hàng 10, cột 0.05 cho giá trị $t_\alpha = 2.228$.

	...	...	0.04	***0.05***	0.06	...	...
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
9	...	...	2.398	2.262	2.150	...	...
10	...	...	2.359	**2.228**	2.120	...	...
11	...	...	2.328	2.201	2.096	...	...
...	...	...	...	...	...	...	...

Trường hợp tìm t_α sao cho $P(|T| \le t_\alpha) = \gamma$ được khảo sát giống như trường hợp phân phối Gauss

$$P\left(\left|T\right| > t_\alpha\right) = 1 - P\left(\left|T\right| \le t_\alpha\right) = 1 - \gamma.$$

Chú ý Khi $T \sim St(n)$, với $n \ge 30$, ta có thể

– Dùng bảng phân phối Student với độ tự do $n = \infty$ (hàng cuối), hay

– Xấp xỉ phân phối Student bằng phân phối Gauss, nghĩa là $X \sim N(0,1)$.

BẢNG 5. Phân phối Chi-Bình phương

Tương tự phân phối Student, phân phối Chi-Bình phương được dùng trong thống kê và ta gặp hai bài toán sau

– Tìm $a, b \in \mathbb{R}$ sao cho $P(X < a) = P(X > b) = \dfrac{\alpha}{2}$, với (nguy cơ sai lầm) α cho trước (bài toán ước lượng),

– Tìm $C \in \mathbb{R}$ sao cho $P(X > C) = \alpha$ (bài toán kiểm định).

Do $P(X < a) = 1 - P(X \ge a) = 1 - P(X > a)$ nên để giải các bài toán này, người ta tính sẵn một số giá trị x sao cho $P(X \ge x) = \alpha$, $X \sim \chi^2(n)$, tương ứng với các giá trị của α và n cho trước, và được liệt kê thành bảng. Các hàng, trừ hàng 1, được đánh số theo bậc tự do n, các cột, trừ cột 1, được đánh số theo các giá trị của α.

	...	...	0.02	***0.025***	0.03	...	...
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
9	...	...	19.679	19.023	18.480	...	...
10	...	...	21.161	**20.483**	19.922	...	...
11	...	...	22.618	21.920	21.342	...	...
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...

Giá trị trong bảng tương ứng với hàng và cột tìm được chính là giá trị x cần tìm. Chẳng hạn, với $X \sim \chi^2(10)$, để tìm x sao cho $P(X \geq x) = 0.025$, tra bảng ứng với hàng 10, cột 0.025, ta được $x = 20.483$.

BẢNG 6. Phân phối Fisher $F(n, m)$.

Tương tự như phân phối chi-bình phương, với $X \sim F(n, m)$, các giá trị x sao cho $P(X \geq x) = \alpha$ được tính sẵn với một số α, m, n cho trước.

Cụ thể, người ta chỉ xét hai giá trị của α là 0.05 và 0.01, và các giá trị của x được liệt kê thành hai bảng :

Bảng 1 ứng với $\alpha = 0.05$ và bảng 2 ứng với $\alpha = 0.01$.

Trong mỗi bảng, hàng 1 liệt kê các giá trị của n. Cột 1 liệt kê các giá trị của m và giá trị trong bảng là giá trị x cần tìm tương ứng.

Chẳng hạn, nếu $X \sim F(5, 10)$, để tìm x sao cho $P(X \geq x) = 0.05$, ta tra bảng 1, hàng 10, cột 5 và nhận được giá trị của x là 3.33

	...	...	4	***5***	6	...	...
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
9	...	...	3.63	3.48	3.37	...	...
10	...	...	3.48	**3.33**	3.22	...	...
11	...	...	3.36	3.2	3.09	...	...
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...

Để tìm x sao cho $P(X \geq x) = 0.01$, tra bảng 2, hàng 10, cột 5, ta được giá trị của x là 5.64

	...	...	4	***5***	6	...	...
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
9	...	...	6.42	6.06	5.80	...	...
10	...	...	5.99	**5.64**	5.39	...	...
11	...	...	5.67	5.32	5.07	...	...
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...

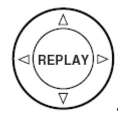
PHỤ LỤC 3

HƯỚNG DẪN SỬ DỤNG MÁY TÍNH

Trong phần này, ta chỉ khảo sát các phép tính thống kê trên ba loại máy: FX 500A, FX 500MS, 570MS và FX 500ES, 570ES.

I. Các ký hiệu, ghi chú

– Các phím bấm trên máy được ký hiệu bởi các biểu tượng được đóng khung, ví dụ : $\boxed{0}$, $\boxed{1}$, $\boxed{2}$, $\boxed{3}$, …, $\boxed{+}$, $\boxed{-}$, $\boxed{\times}$, $\boxed{\div}$, $\boxed{\text{SHIFT}}$, $\boxed{\text{MODE}}$, $\boxed{\sin}$, $\boxed{\cos}$, ….

– Các phím mũi tên như : ▶, ▲, ▼, ◀ là các phím trên nút Replay .

– Chuỗi các ký hiệu biểu tượng như " $\boxed{2} \to \boxed{3} \to \boxed{+} \to \boxed{1} \to \boxed{7} \to \boxed{=}$" nghĩa là bấm các phím $\boxed{2}$, $\boxed{3}$, $\boxed{+}$, $\boxed{1}$, $\boxed{7}$, $\boxed{=}$ trên máy theo thứ tự từ trái sang phải.

II. Các bước trong tính toán thống kê với các loại máy

FX–500A, FX–500MS, FX–570MS

Để nhận được các kết quả trong tính toán thống kê một biến, ta thực hiện tuần tự các bước sau

Bước 1: Vào chế độ thống kê (SD).

Bước 2: Xóa dữ liệu thống kê cũ.

Bước 3: Nhập số liệu thống kê mới.

Bước 4: Khai thác kết quả từ số liệu thống kê vừa nhập.

Bước 5: Thoát khỏi chế độ thống kê.

Cụ thể, ta có

Bước 1: Vào chế độ thống kê

Máy FX–500A: $\boxed{\text{MODE}} \to \boxed{\cdot}$

Máy FX–500MS: MODE → 2

Máy FX–570MS: MODE → 1

Bước 2: Xóa số liệu thống kê cũ

Trong chế độ SD, ta thực hiện như sau:

Máy FX–500A: MODE → 0 → MODE → •

Máy FX–500MS và FX–570MS:

SHIFT → MODE → 1 → = → AC

Bước 3: Nhập số liệu thống kê mới

Số liệu không có tần số: Chẳng hạn để nhập dãy số liệu của X

X	1	2	3	4	5

Máy FX–500A:

1 → M+ → 2 → M+ → 3 → M+ → 4 → M+

→ 5 → M+ → AC

Máy FX–500MS và FX–570MS:

1 → M+ → 2 → M+ → 3

→ M+ → 4 → M+ → 5 → M+ → AC

Số liệu có tần số: Chẳng hạn để nhập dãy số liệu của X

X	1	2	3	4	5
Tần số	3	2	4	5	2

Máy FX–500A: 1 → × → 3 → M+

→ 2 → × → 2 → M+

→ 3 → × → 4 → M+

→ 4 → × → 5 → M+

→ 5 → × → 2 → M+

$\rightarrow \boxed{\text{AC}}$

Máy FX–500MS và FX–570MS:

$\boxed{1} \rightarrow \boxed{\text{SHIFT}} \rightarrow \boxed{,} \rightarrow \boxed{3} \rightarrow \boxed{\text{M+}}$

$\rightarrow \boxed{2} \rightarrow \boxed{\text{SHIFT}} \rightarrow \boxed{,} \rightarrow \boxed{2} \rightarrow \boxed{\text{M+}}$

$\rightarrow \boxed{3} \rightarrow \boxed{\text{SHIFT}} \rightarrow \boxed{,} \rightarrow \boxed{4} \rightarrow \boxed{\text{M+}}$

$\rightarrow \boxed{4} \rightarrow \boxed{\text{SHIFT}} \rightarrow \boxed{,} \rightarrow \boxed{5} \rightarrow \boxed{\text{M+}}$

$\rightarrow \boxed{5} \rightarrow \boxed{\text{SHIFT}} \rightarrow \boxed{,} \rightarrow \boxed{2} \rightarrow \boxed{\text{M+}}$

$\rightarrow \boxed{\text{AC}}$

Bước 4: Khai thác kết quả

Với $\overline{X}$ chỉ trung bình mẫu;

σ_n chỉ phương sai mẫu chưa hiệu chỉnh;

σ_{n-1} chỉ phương sai mẫu có hiệu chỉnh;

$\sum x_i$ chỉ tổng số liệu mẫu;

$\sum x_i^2$ chỉ tổng bình phương các số liệu mẫu

n chỉ cỡ mẫu.

Ta có

Máy FX–500A:

$\overline{X}$: $\boxed{\text{SHIFT}} \rightarrow \boxed{7}$

σ_n: $\boxed{\text{SHIFT}} \rightarrow \boxed{8}$

σ_{n-1}: $\boxed{\text{SHIFT}} \rightarrow \boxed{9}$

$\sum x_i$: $\boxed{\text{SHIFT}} \rightarrow \boxed{5}$

$\sum x_i^2$: SHIFT → 4

n : SHIFT → 6

Máy FX–500MS và FX–570MS:

$\overline{X}$: SHIFT → 2 → 1 → =

σ_n : SHIFT → 2 → 2 → =

σ_{n-1} : SHIFT → 2 → 3 → =

$\sum x_i$: SHIFT → 1 → 2 → =

$\sum x_i^2$: SHIFT → 1 → 1 → =

n: SHIFT → 1 → 3 → =

Bước 5: Thoát khỏi chế độ thống kê

Máy FX–500A: MODE → 0

Máy FX–500MS: MODE → 1

Máy FX–570MS: MODE → 1

III. Các bước trong tính toán thống kê với các loại máy:

FX–500ES, FX–570ES

Để nhận được các kết quả trong tính toán thống kê một biến, ta thực hiện tuần tự các bước sau:

Bước 1: Vào chế độ thống kê (STAT).

Bước 2: Vào chế độ chỉnh sửa dữ liệu.

Bước 2.1: Xóa dữ liệu thống kê cũ.

Bước 2.2: Nhập số liệu thống kê mới.

Bước 3: Khai thác kết quả từ số liệu thống kê vừa nhập.

Bước 4: Thoát khỏi chế độ thống kê.

Chú ý: Chỉ có sự khác biệt giữa hai loại máy này trong bước 1.

Tất cả các bước còn lại là như nhau.

Cụ thể, ta có:

Bước 1: Vào chế độ thống kê

Máy FX–500ES: [MODE] → [2] → [1] → [AC]

Máy FX–570ES: [MODE] → [3] → [1] → [AC]

Bước 2: Vào chế độ chỉnh sửa dữ liệu

Trong chế độ STAT, ta thực hiện như sau:

[SHIFT] → [1] → [2]

Bước 2.1: Xóa số liệu thống kê cũ

[SHIFT] → [1] → [3] → [2]

Bước 2.2: Nhập số liệu thống kê mới

Số liệu không có tần số: Chẳng hạn để nhập dãy số liệu của X

X	1	2	3	4	5

[1] → [=] → [2] → [=] → [3] → [=] → [4]
→ [=] → [5] → [=] → [AC]

Số liệu có tần số: Chẳng hạn để nhập dãy số liệu của X

X	1	2	3	4	5
Tần số	3	2	4	5	2

Nếu trên màn hình không có cột Freq (cột để nhập tần số) thì bấm

[SHIFT] → [MODE] → [▼] → [4] → [1]

Nhập dữ liệu: [1] → [=] → [2] → [=] → [3] → [=] → [4]

$\rightarrow \boxed{=} \rightarrow \boxed{5} \rightarrow \boxed{=} \rightarrow$

Nhập tần số: ▼ $\rightarrow$ ▶ $\rightarrow \boxed{3} \rightarrow \boxed{=} \rightarrow \boxed{2} \rightarrow \boxed{=} \rightarrow \boxed{4}$

$\rightarrow \boxed{=} \rightarrow \boxed{5} \rightarrow \boxed{=} \rightarrow \boxed{2} \rightarrow \boxed{=} \rightarrow \boxed{AC}$

Bước 3: Khai thác kết quả

Với $\overline{X}$ chỉ trung bình mẫu;

σ_n chỉ phương sai mẫu chưa hiệu chỉnh;

σ_{n-1} chỉ phương sai mẫu có hiệu chỉnh;

$\sum x_i$ chỉ tổng số liệu mẫu;

$\sum x_i^2$ chỉ tổng bình phương các số liệu mẫu

n chỉ cỡ mẫu.

Ta có

$\overline{X}$: $\boxed{SHIFT} \rightarrow \boxed{1} \rightarrow \boxed{5} \rightarrow \boxed{2} \rightarrow \boxed{=}$

σ_n : $\boxed{SHIFT} \rightarrow \boxed{1} \rightarrow \boxed{5} \rightarrow \boxed{3} \rightarrow \boxed{=}$

σ_{n-1} : $\boxed{SHIFT} \rightarrow \boxed{1} \rightarrow \boxed{5} \rightarrow \boxed{4} \rightarrow \boxed{=}$

n: $\boxed{SHIFT} \rightarrow \boxed{1} \rightarrow \boxed{5} \rightarrow \boxed{1} \rightarrow \boxed{=}$

$\sum x_i$: $\boxed{SHIFT} \rightarrow \boxed{1} \rightarrow \boxed{4} \rightarrow \boxed{2} \rightarrow \boxed{=}$

$\sum x_i^2$: $\boxed{SHIFT} \rightarrow \boxed{1} \rightarrow \boxed{4} \rightarrow \boxed{1} \rightarrow \boxed{=}$

Bước 4: Thoát khỏi chế độ thống kê

TÀI LIỆU THAM KHẢO

1. Xác suất thống kê
GS. ĐẶNG HẤN – TRƯỜNG ĐẠI HỌC KINH TẾ TPHCM, 1991

2.Bài tập xác suất và thống kê toán
Chủ biên TS NGUYỄN CAO VĂN – TRƯỜNG ĐẠI HỌC KINH TẾ QUỐC DÂN HÀ NỘI, NXB GIÁO DỤC HÀ NỘI, 2002

3. Lý thuyết xác suất và thống kê toán
Thạc sĩ HOÀNG NGỌC NHẬM – TRƯỜNG ĐẠI HỌC KINH TẾ TPHCM, 2005

4. Bài tập xác suất và thống kê toán
Thạc sĩ LÊ KHÁNH LUẬN – TRƯỜNG ĐẠI HỌC KINH TẾ TPHCM, 2005

5. Lý thuyết xác suất và thống kê toán học
LÊ TRUNG TƯƠNG – TRƯỜNG ĐẠI HỌC BÁCH KHOA TPHCM