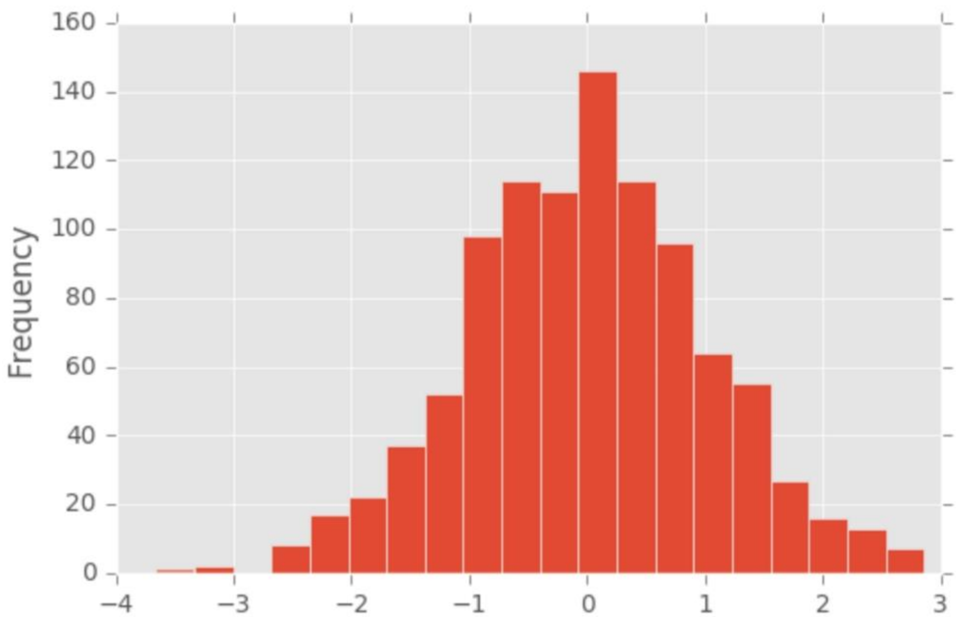
****

BÀI GIẢNG

XÁC SUẤT VÀ THỐNG KÊ



► Yêu cầu của môn học: Máy tính bỏ túi Fx 570, Fx 580

Sinh viên: ……………………………………...

Lớp :……….……………………………………

Gv: Võ Thanh Hải

06/ 2023

Chương 1. XÁC SUẤT CƠ BẢN

MỤC TIÊU

Nội dung chương này giúp người học có khả năng:

- Hiểu được khái niệm phép thử, không gian mẫu, biến cố và các định nghĩa xác suất của biến cố .
- Nhận biết được các quan hệ xung khắc, độc lập và họ đầy đủ các biến cố.
- Vận dụng được công thức cộng và công thức nhân để tính xác suất.
- Hiểu được khái niệm xác suất có điều kiện và tính được xác suất bằng công thức xác suất đầy đủ, công thức Bayes.

1. MỘT SỐ KHÁI NIỆM VỀ GIẢI TÍCH TỔ HỢP

1.1 Qui tắc đếm

Qui tắc cộng

Để hoàn thành một công việc có thể thực hiện theo **2 trường hợp** khác nhau:

+ nếu thực hiện theo trường hợp 1: có ***n*** cách để hoàn thành.

+ nếu thực hiện theo trường hợp 2: có ***m*** cách để hoàn thành.

Khi đó số cách để hoàn thành công việc sẽ là: ***n + m***.

Qui tắc nhân

Để hoàn thành một công việc phải thực hiện qua **2 giai đoạn** khác nhau:

+ nếu thực hiện giai đoạn 1: có ***n*** cách để hoàn thành

+ nếu thực hiện giai đoạn 2: có ***m*** cách để hoàn thành.

Khi đó số cách để hoàn thành công việc sẽ là: ***n.m***.

Chú ý: *Trong thực tế khi diễn đạt làm công việc nếu dùng chữ* ***HOẶC*** *thì ta dùng quy tắc* ***CỘNG****, còn nếu dùng chữ* ***VÀ*** *thì ta dùng quy tắc* ***NHÂN****,*

Ví dụ 1. Một người tham gia một trò chơi trên truyền hình bằng cách chọn một câu hỏi để trả lời. Có 8 câu hỏi về thể thao và 7 câu hỏi về lịch sử. Hỏi người đó có bao nhiêu lựa chọn.
Giải.

……………………………………………………………………………………………
……………………………………………………………………………………………
…………………………………………………………………………………….

Ví dụ 2. Một khách du lịch đang dự định đi tham quan một đỉnh núi. Khi đi lên có thể đi bộ hoặc đi cáp treo, còn đi xuống có thể đi bộ, đi cáp treo hoặc đi máng trượt. Hỏi người này có bao nhiêu cách lựa chọn để đi?
Giải:

……………………………………………………………………………………………
……………………………………………………………………………………………
…………………………………………………………………………………….

1.2 Chỉnh hợp

Một cách chọn lần lượt không lặp lại (không hoàn lại), có thứ tự, *r* phần tử từ một tập hợp có *n* phần tử khác nhau được gọi là một chỉnh hợp chập r của n phần tử $(1 \le r \le n)$. Ký hiệu A_n^r là số các chỉnh hợp chập *r* của *n* phần tử thì ta có: $A_n^r = \dfrac{n!}{(n-r)!}$.

1.3 Tổ hợp

Một cách chọn đồng thời không phân biệt thứ tự, *r* phần tử từ một tập hợp có *n* phần tử khác nhau được gọi là một tổ hợp chập *r* của *n* phần tử $(1 \le r \le n)$. Ký hiệu C_n^r là số các tổ hợp chập *r* của *n* phần tử ta có: $C_n^r = \dfrac{n!}{r!(n-r)!}$.

Chú ý:

Khái niệm chỉnh hợp và tổ hợp khác nhau cơ bản ở chỗ chỉnh hợp có phân biệt *thứ tự r* phần tử lấy ra, còn tổ hợp thì không phân biệt *thứ tự*.

Ví dụ 3. Chọn 10 bạn sinh viên bất kỳ trong một lớp học có 100 sinh viên để làm một bài kiểm tra nhanh về Tiếng Anh. Hỏi có mấy cách chọn ?

Giải

Chọn 10 bạn từ 100 bạn để làm bài kiểm tra là không có thứ tự. Vậy có $C_{100}^{10} = 17310309456440$ cách chọn.

2. CÁC KHÁI NIỆM CƠ BẢN CỦA XÁC SUẤT

2.1 Phép thử

Thực hiện một công việc và quan tâm đến kết quả của công việc này, thì việc làm này gọi chung là **một phép thử**. Ký hiệu một phép thử là *T*.

2.2 Không gian mẫu (Sample Space)

Tập hợp tất cả các kết quả có thể xảy ra của một phép thử được gọi là ***không gian mẫu*** và được ký hiệu là Ω.

2.3 Biến cố (Events)

- Mỗi một tập hợp chứa một số các kết quả của một phép thử gọi là một *biến cố ngẫu nhiên* hay ***biến cố.*** Ký hiệu các biến cố là: $A, B, C, \ldots$
- Nếu một phép thử mà mỗi lần chỉ cho một kết quả đơn (không phân tách được dưới dạng các kết quả khác) thì những kết quả đó gọi là *các* ***biến cố sơ cấp***.
- Nếu kết quả của phép thử thuộc vào biến cố A thì ta nói ***biến cố A xảy ra***, ngược lại ta nói ***biến cố A không xảy ra***.

Biến cố chắc chắn xảy ra khi thực hiện phép thử gọi là "***biến cố chắc chắn***", ký hiệu là Ω.

Biến cố không bao giờ xảy ra khi thực hiện phép thử gọi là "***biến cố không thể***", ký hiệu là $\varnothing$.

Ví dụ 4. Tung một con súc sắc đồng chất cân đối một lần xem mặt mấy chấm xuất hiện.

Con súc sắc.

Ta có: - Tung con súc sắc là một phép thử.

- Có 6 biến cố sơ cấp là: $\{1\},\{2\},\{3\},\{4\},\{5\},\{6\}$.

- Không gian mẫu là: $\Omega=\{1,2,3,4,5,6\}$.

- Biến cố A = "số chấm xuất hiện lớn hơn 2".

- Biến cố B = "số chấm xuất hiện nhỏ hơn 5".

- Nếu tung được "6 chấm" ta nói biến cố A xảy ra, còn biến cố B không xảy ra.

Ví dụ 5. Hai người cùng bắn vào một mục tiêu, mỗi người bắn một viên đạn. Hãy tìm không gian mẫu.

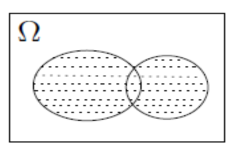
Giải:

………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………

2.4 Các phép toán của biến cố

Tổng của hai biến cố

Biến cố tổng $A+B$ (hay $A\cup B$) xảy ra khi và chỉ khi A xảy ra hoặc B xảy ra (có ít nhất một trong hai biến cố *A*, *B* xảy ra).

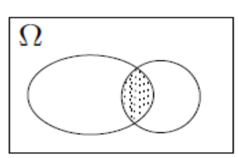


A+B

Tổng $A\cup B$.

Tích của hai biến cố

Biến cố tích AB (hay $A\cap B$) xảy ra khi và chỉ khi cả hai biến cố *A*, *B* cùng xảy ra.



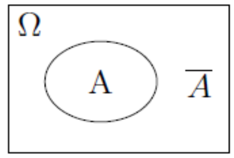
AB

Tích $A\cap B$.

Phần bù của biến cố

Phần bù của biến cố *A* được ký hiệu $\overline{A}$ và $\overline{A}=\Omega\setminus A$.

Luật đối ngẫu (De-Morgan): $\overline{A+B}=\overline{A}.\overline{B}$ và $\overline{A.B}=\overline{A}+\overline{B}$.



Phần bù.

Ví dụ 6. Tung một con súc sắc 6 mặt, xét các biến cố sau:

A: Xuất hiện mặt có số chấm chẵn.

B: Xuất hiện mặt có số chấm nhỏ hơn 4.

Ta có: $A+B=\{1,2,3,4,6\}$, $AB=\{2\}$, $\overline{A}=\{1,3,5\}$.

Ví dụ 7. Tung 1 con súc sắc 6 mặt, xét các biến cố sau:
A: Xuất hiện mặt có số chấm chẵn.
B: Xuất hiện mặt có số chấm lớn hơn hay bằng 3.
C: Xuất hiện mặt có số chấm không quá 4.
Ta có: A+B = B.C = B+C = $\overline{C}$ =

3. CÁC ĐỊNH NGHĨA CỦA XÁC SUẤT

3.1 Khái niệm xác suất (Probability)

Xác suất của biến cố *A* ký hiệu $P(A)$ là một số thực nói lên khả năng xảy ra của *A* khi phép thử được thực hiện,

3.2 Định nghĩa xác suất theo cổ điển

Giả sử phép thử *T* thỏa 2 điều kiện:

i. Không gian mẫu có **hữu hạn** phần tử.

ii. Các khả năng xảy ra của các phần tử không gian mẫu là như nhau (**đồng khả năng**).

Khi đó nếu *A* là một biến cố bất kỳ của phép thử *T*, ta có: $P(A)=\dfrac{m_A}{n}$.

Trong đó: *n* là số khả năng có thể xảy ra của phép thử (số phần tử của không gian mẫu). m_A là số khả năng để biến cố *A* xảy ra.

Ví du 8. Từ một lô hàng có 13 chính phẩm và 7 phế phẩm, lấy ngẫu nhiên 1 sản phẩm.
Gọi A là biến cố *lấy được chính phẩm*, ta có: $P(A)=.....$

Ví du 9. Xếp ngẫu nhiên 5 sinh viên A, B, C, D, E vào một ghế dài có 5 chỗ ngồi. Tính xác suất B và C ngồi cạnh nhau.
Giải:

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
…………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
…………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………

Ví dụ 10 – Bài toán cơ bản 3C
Một lô hàng có 15 sản phẩm, trong đó có 10 chính phẩm và 5 phế phẩm. Lấy ngẫu nhiên từ lô sản phẩm đó 8 sản phẩm. Tìm xác suất để trong 8 sản phẩm lấy ra có 6 chính phẩm và 2 phế phẩm.
Giải:
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
…………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
…………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………

Ví dụ 11. Bốn người A, B, C, D chơi bài tiến lên.Tính xác suất để người A được 4 lá Hai.

Giải:
……………………………………………………………………………………………
……………………………………………………………………………………………
………………………………………………………………………………………
……………………………………………………………………………………………
…………………………………………………………………………………………

Ví dụ 12. Ba sinh viên cùng thi môn Xác suất thống kê (điểm thi là các số nguyên từ 0 đến 10). Tính xác suất để điểm thi của cả 3 bạn đều khác nhau.
Giải:

……………………………………………………………………………………………
……………………………………………………………………………………………
…………………………………………………………………………………….
……………………………………………………………………………………………
………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
…………………………………………………………………………………….

3.3 Định nghĩa xác suất theo thống kê

Ta nhận thấy rằng tính xác suất theo định nghĩa cổ điển sẽ không thực hiện được khi không thỏa cả hai điều kiện đã nêu trong định nghĩa. Tức là xác suất sẽ không tính được khi số phần tử của không gian mẫu là vô hạn hoặc hữu hạn nhưng không đồng khả năng. Vì vậy người ta đưa ra định nghĩa về xác suất theo thống kê như sau:

Định nghĩa

Giả sử phép thử *T* có thể được thực hiện lặp lại nhiều lần trong những điều kiện giống hệt nhau. Nếu trong *n* lần thực hiện phép thử *T*, biến cố *A* xuất hiện m_A lần thì tỷ số $f_n(A)=\frac{m_A}{n}$ được gọi là tần suất xuất hiện của biến cố *A* trong *n* phép thử. Khi *n* tăng lên vô hạn mà $f_n(A)$ tiến đến một giới hạn xác định $P(A)$ thì giới hạn này được gọi là xác suất của biến cố *A*, tức là $P(A)=\lim\limits_{n\to+\infty} f_n\left(A\right)$. Trong thực tế khi *n* đủ lớn ta lấy xác suất $P(A)$ xấp xỉ tần suất $f_n(A)$.

Ví dụ 13. Để nghiên cứu khả năng xuất hiện mặt sấp (biến cố *A*) khi tung một đồng tiền, hai nhà khoa học Buffon và Pearson tiến hành tung đồng tiền nhiều lần và thu được kết quả sau:

Người tiến hành thử	Số lần tung (n)	Số lần được mặt sấp (m_A)	Tần suất $f_n(A)$
Buffon	4040	2048	0,5080
Pearson	12000	6019	0,5016
Pearson	24000	12012	0,5005

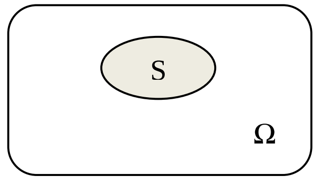
Từ các kết quả trên ta nhận thấy khi số phép thử *n* tăng lên, tần suất xuất hiện mặt sấp tiến dần đến 0,5. Vậy ta nói xác suất xuất hiện mặt sấp khi tung đồng tiền là 0,5.

3.4 Định nghĩa xác suất theo hình học

Khi số phép thử *n* là vô hạn và khả năng xảy ra của các phần tử là như nhau, ta không thể áp dụng định nghĩa cổ điển để tính xác suất. Trong nhiều trường hợp, ta có thể sử dụng định nghĩa xác suất theo quan điểm hình học như sau:

Định nghĩa: Giả sử một phép thử *T* được xem tương tự một điểm *M* rơi ngẫu nhiên vào một miền Ω có diện tích hữu hạn. Khả năng điểm *M* rơi vào miền *S* là một miền con của Ω tỷ lệ với diện tích của miền này. Khi đó xác suất của biến cố *A* " điểm *M* rơi ngẫu nhiên vào miền *S*" được xác định bởi công thức: $P(A)=\frac{sd\left(S\right)}{sd\left(\Omega\right)}$.

Trong đó $sd(S)$, $sd(\Omega)$ là số đo diện tích của các miền S, Ω.



4. CÁC TÍNH CHẤT VÀ ĐỊNH LÝ VỀ XÁC SUẤT

4.1 Tính chất

Cho *A* là một biến cố bất kỳ, ta luôn có: $P(A) = 1 - P(\overline{A})$ và $0 \leq P(A) \leq 1$.

4.2 Quan hệ xung khắc

Hai biến cố *A*, *B* được gọi là ***xung khắc*** nếu $A.B = \varnothing$. Nghĩa là 2 biến cố này không thể đồng thời cùng xảy ra.

Các biến cố $A_1, A_2, ..., A_n$ gọi là họ xung khắc hoặc họ xung khắc từng đôi nếu bất kỳ 2 biến cố trong họ đều xung khắc nhau.

Chú ý: ***Xung khắc là biến cố này xảy ra thì biến cố kia không xảy ra***

Ví dụ 14. Tung 1 con súc sắc 6 mặt, xét các biến cố sau:

A: Xuất hiện mặt có số chấm chẵn.
B: Xuất hiện mặt 1 chấm.
C: Xuất hiện mặt có số chấm lớn hơn 2.
D: Xuất hiện mặt có số chấm lẻ.

Các cặp biến cố xung khắc là: ……………………………………………………..

4.3 Công thức cộng xác suất

Nếu *A*, *B* là hai biến cố **bất kỳ** thì:

$$P(A+B) = P(A) + P(B) - P(AB) \quad \text{hay} \quad P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Nếu *A*, *B* là hai biến cố **xung khắc** thì:

$$P(A+B) = P(A) + P(B) \quad \text{hay} \quad P(A \cup B) = P(A) + P(B)$$

Nếu $A_1, A_2, ..., A_n$ là một **họ xung khắc** thì: $P(A_1 + A_2 + \ldots + A_n) = P(A_1) + P(A_2) + \ldots + P(A_n)$

Ví dụ 15. Lớp có 120 học sinh trong đó có 65 học sinh giỏi Anh văn, 50 học sinh giỏi Pháp văn và 40 học sinh giỏi cả 2 ngoại ngữ. Chọn ngẫu nhiên 1 học sinh trong lớp. Tính xác suất học sinh này giỏi ít nhất một ngoại ngữ.

Giải:

………………………………………………………………………………………
………………………………………………………………………………………
…………………………………………………………………………………….
………………………………………………………………………………………
……………………………………………………………………………………

Ví dụ 16. Trong một lớp học, tỉ lệ sinh viên học tiếng Anh là 70%, tỉ lệ sinh viên học tiếng Nhật là 9% và tỉ lệ sinh viên học cả hai thứ tiếng Anh và Nhật là 3%. Chọn ngẫu nhiên một sinh viên trong lớp. Tính xác suất để sinh viên đó học ít nhất một môn ngoại ngữ Anh hoặc Nhật?
Giải:
……………………………………………………………………………………………
……………………………………………………………………………………………
…………………………………………………………………………………….

Câu 17. Một cửa hàng chỉ bán hai loại điện thoại Iphone và Samsung. Một cuộc khảo sát 200 khách hàng đầu tiên đã mua hàng, thấy có 60% mua điện thoại Iphon, 70% mua điện thoại Samsung. Chọn một khách hàng bất kì trong 200 khách đó, hỏi xác suất chọn được người mua cả 2 loại là bao nhiêu?
Giải:
……………………………………………………………………………………………
……………………………………………………………………………………………
…………………………………………………………………………………….
……………………………………………………………………………………………
…………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………

Ví dụ 18. Một lô hàng có 25% sản phẩm loại I, 55% sản phẩm loại II và 20% sản phẩm loại III. Sản phẩm được cho là đạt chất lượng nếu thuộc loại I hoặc loại II. Chọn ngẫu nhiên 1 sản phẩm tìm xác suất để sản phẩm này đạt tiêu chuẩn chất lượng.
Giải:
……………………………………………………………………………………………
……………………………………………………………………………………………
…………………………………………………………………………………….
……………………………………………………………………………………………
…………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………

Ví dụ 19. Một lô hàng chứa 15 sản phẩm gồm 10 sản phẩm tốt và 5 sản phẩm xấu. Chọn ngẫu nhiên từ lô hàng 4 sản phẩm.Tính xác suất để trong 4 sản phẩm được chọn ra có:
a. Nhiều hơn 2 sản phẩm tốt.
b. Ít nhất 1 sản phẩm xấu.
Giải:
…………………………………………………………………………………………
……………………………………………………………………………………………
………………………………………………………………………………………
……………………………………………………………………………………………
…………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
…………………………………………………………………………………….
……………………………………………………………………………………………

5. XÁC SUẤT CÓ ĐIỀU KIỆN

5.1 Định nghĩa

Xác suất có điều kiện $P(A|B)$ là xác suất của biến cố *A* được tính trong khi biến cố ***B* đã xảy ra** rồi.

Công thức xác suất có điều kiện: $P(A|B)=\dfrac{P(AB)}{P(B)}$

Ví dụ 20. Tung 1 con súc sắc đồng chất 6 mặt. Xét các biến cố sau:

A là biến cố xuất hiện mặt có số chấm chẵn.
B là biến cố xuất hiện mặt có số bé hơn 5.
C là biến cố xuất hiện mặt có số chấm nhỏ hơn hay bằng 2.
D là biến cố xuất hiện mặt có số chấm lớn hơn hay bằng 4.
Khi đó

- P(A|B)=……………………………………………………………………………………
  ……………………………………………………………………………………………
  ……………………………………………………………………………………………
  …………………………………………………………………………………………

- P(D|A) = ; P(C|B) = ; P(C|D) = ; P(C|A) =

Ví dụ 21. Một nhóm 100 người trong đó có 40 nữ. Biết rằng nhóm 100 người này có 30 người hút thuốc trong đó có 5 người hút thuốc là nữ. Chọn ngẫu nhiên một người trong nhóm 100 người này. Tính xác suất:

a. Người này hút thuốc biết rằng người này là nữ.

b. Người này là nữ biết rằng người này hút thuốc.

Giải.

……………………………………………………………………………………………
……………………………………………………………………………………………
………………………………………………………………………………………
……………………………………………………………………………………………
…………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………

5.2 Quan hệ độc lập

Hai biến cố *A*, *B* được gọi là ***độc lập*** nếu A xảy ra hay không xảy ra không ảnh hưởng đến khả năng xảy ra của B, và ngược lại.

Các biến cố $A_1, A_2, ..., A_n$ gọi là độc lập nếu việc xảy ra hay không xảy ra của một nhóm bất kỳ *k* biến cố $(1 \le k \le n)$ trong đó, đều không làm ảnh hưởng đến việc xảy ra hay không xảy ra của một nhóm các biến cố còn lại.

Chú ý: ***Độc lập là biến cố này xảy ra không ảnh hưởng (xác suất) đến biến cố kia xảy ra***

Hệ quả

1) Nếu hai biến cố *A*, *B* là độc lập thì các cặp biến cố $A, \overline{B}$; $\overline{A}, B$; $\overline{A}, \overline{B}$ cũng độc lập.

2) Hai biến cố *A*, *B* là độc lập khi và chỉ khi $P(A|B) = P(A)$ và $P(B|A) = P(B)$.

5.3 Công thức nhân xác suất

Nếu $A_1, A_2, ..., A_n$ là các biến cố **bất kỳ** thì

$$P(A_1A_2A_3...A_n) = P(A_1).P(A_2|A_1).P(A_3|A_1A_2)....P(A_n|A_1A_2A_3...A_{n-1}).$$

Hay

$$P(A_1 \cap A_2 \cap ... \cap A_n) = P(A_1).P(A_2|A_1)....P(A_n|A_1A_2A_3...A_{n-1})$$

Nếu các biến cố $A_1, A_2, ..., A_n$ là **độc lập** thì

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

Hay

$$P(A_1 \cap A_2 \cap ... \cap A_n) = P(A_1).P(A_2).....P(A_n)$$

Ví dụ 22. Một hộp đựng 4 bi trắng và 6 bi đen. Thực hiện hai lần lấy bi, mỗi lần lấy 1 bi (lấy không hoàn lại). Gọi các biến cố:
A: “Lần 1 lấy được bi đen”
B: “Lần 2 lấy được bi trắng”
Hai biến cố A và B có độc lập không?
Giải

……………………………………………………………………………………………
………………………………………………………………………………………………
…………………………………………………………………………………………
………………………………………………………………………………………………
…………………………………………………………………………………………

Ví dụ 23. Hai xạ thủ mỗi người bắn một viên đạn vào cùng một mục tiêu, xác suất bắn trúng mục tiêu lần lượt của mỗi người là 0,7 và 0,8. Tính xác suất:
a. Cả 2 người bắn trúng mục tiêu.
b. Có đúng một người bắn trúng mục tiêu.
Giải:

……………………………………………………………………………………………
………………………………………………………………………………………………
…………………………………………………………………………………………

…………………………………………………………………………………………………
……………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………………
……………………………………………………………………………………………

Ví dụ 24. Sản phẩm trước khi xuất khẩu phải qua hai lần kiểm tra. Bình quân 80% sản phẩm làm ra qua được lần kiểm tra thứ nhất. Nếu sản phẩm đã qua lần kiểm tra thứ nhất thì 90% sẽ qua được lần kiểm tra thứ hai. Tính tỷ lệ sản phẩm được xuất khẩu.
Giải:
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
……………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………………
……………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………………
……………………………………………………………………………………………

Ví dụ 25. Một người bắn liên tiếp vào một mục tiêu cho đến khi có một phát đạn trúng mục tiêu thì ngưng bắn. Biết rằng xác suất trúng mục tiêu của mỗi lần bắn là như nhau và bằng 0,6. Tính xác suất sao cho khi bắn đến phát thứ ba thì ngưng bắn.
Giải:
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………………
……………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………………
……………………………………………………………………………………………

Ví dụ 26. Bạn đang cầm một chùm 3 chìa khóa khác nhau của 3 phòng và mở các phòng bằng chìa khóa đó một cách ngẫu nhiên.Tính xác suất để bạn mở được 3 phòng mà mỗi phòng chỉ đúng một lần mở khóa.
Giải:
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………

……………………………………………………………………………………………
…………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
………………………………………………………………………………………
……………………………………………………………………………………………

Ví dụ 27. Một lô hàng gồm 10 sản phẩm trong đó có 2 phế phẩm. Người ta chia ngẫu nhiên lô hàng thành 2 gói, mỗi gói có số sản phẩm bằng nhau. Tính xác suất để trong mỗi gói đều có đúng một phế phẩm.

Giải:

……………………………………………………………………………………………
……………………………………………………………………………………………
………………………………………………………………………………………
……………………………………………………………………………………………
…………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
………………………………………………………………………………………
…………………………………………………………………………………………
……………………………………………………………………………………………
…………………………………………………………………………………………
……………………………………………………………………………………………

Ví dụ 28. Một công ty tham gia đấu thầu hai dự án A và B. Đợt đầu tham gia đấu thầu dự án A, khả năng trúng thầu là 80%. Nếu đợt đầu công ty trúng thầu thì khả năng trúng thầu dự án B là 90%, ngược lại khả năng trúng thầu dự án B chỉ còn 60%. Tính xác suất để công ty chỉ trúng thầu một dự án.

Giải:

……………………………………………………………………………………………
……………………………………………………………………………………………
………………………………………………………………………………………
……………………………………………………………………………………………
…………………………………………………………………………………………
……………………………………………………………………………………………

Ví dụ 29. Một phân xưởng có 3 máy làm việc. Trong một ca máy thứ I có thể hỏng với xác suất 0,15; máy thứ II có thể hỏng với xác suất 0,1; máy thứ III có thể hỏng với xác suất 0,12. Tính xác suất để trong 1 ca có ít nhất 1 máy hỏng.

Giải:

……………………………………………………………………………………………
……………………………………………………………………………………………
…………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………

5.4 Họ đầy đủ

Các biến cố $A_1, A_2, ..., A_n$ được gọi là **họ đầy đủ** nếu thỏa 2 điều kiện sau:

i. $A_1, A_2, ..., A_n$ là họ **xung khắc** từng đôi.

ii. $A_1 + A_2 + ... + A_n = \Omega$ tức là **chắc chắn có một biến cố trong họ xảy ra**.

5.5 Công thức xác suất đầy đủ và công thức Bayes

Cho $A_1, A_2, ..., A_n$ là một họ đầy đủ các biến cố và *B* là một biến cố nào đó $(P(B) \neq 0)$ của phép thử mà khi *B* xảy ra thì có A_i xảy ra. Khi đó:

i. Công thức xác suất đầy đủ: $P(B) = P(A_1).P(B \mid A_1) + ... + P(A_n).P(B \mid A_n) = \sum P(A_i B)$.

ii. Công thức Bayes: $P(A_i \mid B) = \dfrac{P(A_i).P(B \mid A_i)}{P(B)}$.

Ví dụ 30. Một kho hàng có 2 kiện hàng loại I mỗi kiện có 8 sản phẩm tốt và 2 sản phẩm xấu; và 3 kiện hàng loại II mỗi kiện có 6 sản phẩm tốt và 4 sản phẩm xấu. Chọn ngẫu nhiên từ kho hàng một kiện hàng và từ kiện hàng đó chọn ngẫu nhiên 1 sản phẩm.
a. Tính xác suất để sản phẩm chọn được là sản phẩm tốt.
b. Biết sản phẩm chọn được là sản phẩm tốt. Khả năng sản phẩm này thuộc kiện hàng nào?
Giải:

……………………………………………………………………………………………
……………………………………………………………………………………………
…………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
…………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
…………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
…………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………

Ví dụ 31. Tỷ lệ người dân nghiện thuốc lá ở 1 vùng là 60%. Biết rằng tỷ lệ người bị viêm họng trong số người nghiện thuốc lá là 80%, còn tỷ lệ người viêm họng trong số người không nghiện thuốc lá là 20%. Chọn ngẫu nhiên 1 người trong vùng đó:

a. Tính xác suất để người đó bị viêm họng.

b. Biết người đó bị viêm họng. Tính xác suất để người đó không nghiện thuốc lá.

Giải

……………………………………………………………………………………………

……………………………………………………………………………………………

…………………………………………………………………………………….

……………………………………………………………………………………………

………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

…………………………………………………………………………………….

……………………………………………………………………………………………

………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

…………………………………………………………………………………….

……………………………………………………………………………………………

………………………………………………………………………………………

CHÚ Ý CÁC DẠNG TOÁN THƯỜNG ÁP DỤNG CÔNG THỨC XS ĐẦY ĐỦ

……………………………………………………………………………………………

……………………………………………………………………………………………

…………………………………………………………………………………….

……………………………………………………………………………………………

………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

…………………………………………………………………………………….

……………………………………………………………………………………………

………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

…………………………………………………………………………………….

……………………………………………………………………………………………

………………………………………………………………………………………

Ví dụ 32. Một trại chăn nuôi nhận 50 con giống từ ba cơ sở, trong đó 15 con thuộc cơ sở I; 10 con thuộc cơ sở II và 25 con thuộc cơ sở III. Tỉ lệ con giống không đạt tiêu chuẩn của mỗi cơ sở tương ứng là 16%, 15% và 12%. Kiểm tra ngẫu nhiên 1 con từ trại chăn nuôi này thấy đạt tiêu chuẩn. Hãy xét xem khả năng con giống này thuộc về cơ sở nào là lớn hơn?

Giải:

……………………………………………………………………………………

……………………………………………………………………………………

…………………………………………………………………………….

……………………………………………………………………………………

………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

…………………………………………………………………………….

……………………………………………………………………………………

………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

Ví dụ 33. Một lớp có 40% là sinh viên nữ dự thi môn tiếng Anh. Tỷ lệ đậu của nữ là 80%, của nam là 50%. Chọn ngẫu nhiên 1 sinh viên dự thi, tính xác suất đậu của sinh viên này.
Giải:

……………………………………………………………………………………

……………………………………………………………………………………

…………………………………………………………………………….

……………………………………………………………………………………

………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

…………………………………………………………………………….

Ví dụ 34. Một hộp có 15 quả bóng trong đó có 9 quả mới và 6 quả cũ. Lần đầu chọn ra 2 quả để sử dụng, lần hai chọn ra 3 quả. Tính xác suất để 3 quả lấy ra ở lần hai đều là bóng mới.
Giải:

……………………………………………………………………………………

……………………………………………………………………………………

…………………………………………………………………………….

……………………………………………………………………………………

………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

…………………………………………………………………………….

……………………………………………………………………………………

………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

…………………………………………………………………………….

……………………………………………………………………………………

………………………………………………………………………………

……………………………………………………………………………………

Ví dụ 35. Một kho hàng chứa sản phẩm cùng loại của 3 nhà máy với tỷ lệ sản phẩm tương ứng là 30%, 50%, 20%. Tỷ lệ phế phẩm của 3 nhà máy tương ứng là 1%, 3%, 5%. Một khách hàng mua một sản phẩm của kho hàng là phế phẩm. Theo Anh, Chị khả năng phế phẩm đó là của nhà máy nào?

Giải:
……………………………………………………………………………………………
……………………………………………………………………………………………
………………………………………………………………………………………
……………………………………………………………………………………………
…………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
………………………………………………………………………………………
……………………………………………………………………………………………
…………………………………………………………………………………………

BÀI TẬP CHƯƠNG 1

Bài 1: Một nhà máy sản xuất sản phẩm với 40% sản phẩm loại I, 50% sản phẩm loại II, còn lại là phế phẩm. Lấy ngẫu nhiên 1 sp của nhà máy. Tính xác suất sản phẩm lấy ra không phải là phế phẩm.
A. 0,1 B. 0,9 C. 0,4 D. 0,5
Bài 2: Một lô hàng có 9 sản phẩm. Môt lần kiểm tra chất lượng lấy ngẫu nhiên 3 sp. Sau khi kiểm tra xong lại trả vào lô hàng. Tính xác suất để sau 3 lần kiểm tra lô hàng tất cả sản phẩm đều được kiểm tra.
A. 0,03 B. 0,001 C. 0,0028 D. 0,15
Bài 3: Hai máy cùng sản xuất một loại sản phẩm. Tỷ lệ phế phẩm cuả máy I là 3%, của máy II là 2%. Từ một kho gồm 2/3 sản phẩm của máy I và 1/3 sản phẩm của máy II ta lấy ra 1 sản phẩm. Tính xác suất để sản phẩm đó là tốt.
A. 0,027 B. 0,94 C. 0,828 D. 0,973
Bài 4: Có 2 xạ thủ loại I và 8 xạ thủ loại II, xác suất bắn trúng đích của các xạ thủ theo thứ tự là 0,9 và 0,8. Chọn ngẫu nhiên 1 xạ thủ và xạ thủ đó bắn 1 viên đạn. Tìm xác suất để viên đạn đó trúng đích.
A. 0,82 B. 0,84 C. 0,18 D. 0,72
Bài 5: Bắn 2 phát đạn vào 1 máy bay với xác suất trúng tương ứng là 0,4 và 0,7. Nếu trúng 1 phát thì xác suất rơi máy bay là 0,2; nếu trúng 2 phát thì xác suất rơi máy bay là 0,6;. Tìm xác suất để máy bay rơi.
A. B. C. D.
Bài 6: Có ba khẩu súng I, II và III bắn độc lập vào một mục tiêu. Mỗi khẩu bắn 1 viên. Xác suất bắn trúng mục tiêu của 3 khẩu lần lượt là 0,7; 0,8 và 0,5. Tính xác suất để ít nhất 1 khẩu bắn trúng.
A. 0,03 B. 0,97 C. 0,28 D. 0,72

Bài 7: Có 2 xạ thủ loại I và 8 xạ thủ loại II, xác suất bắn trúng đích cuả các xạ thủ theo thứ tự là 0,9 và 0,8. Chọn ngẫu nhiên 2 xạ thủ và mỗi xạ thủ bắn 1 viên đạn. Tìm xác suất để cả 2 viên đều trúng đích.
A. 0,52 B. 0,67 C. 0,18 D. 0,72
Bài 8: Có hai hộp I và II mỗi hộp chứa 10 bi, trong đó hộp I gồm 9 bi đỏ, 1bi trắng; hộp II gồm 6 bi đỏ, 4 bi trắng. Lấy ngẫu nhiên từ mỗi hộp 1 bi. Tính xác suất để được 1 bi đỏ và 1 bi trắng.
A. 0,42 B. 0,64 C. 0,28 D. 0,52

Bài 9: Một lô hàng chứa 10 sản phẩm gồm 6 sản phẩm tốt và 4 sản phẩm xấu. Khách hàng kiểm tra bằng cách lấy ra từng sản phẩm (không hoàn lại) cho đến khi nào được 2 sản phẩm tốt thì dừng lại. Biết khách hàng đã dừng lại ở lần kiểm tra thứ 3. Tính xác suất để ở lần kiểm tra thứ 2 khách hàng gặp sản phẩm xấu.

A. B. C. D.

Bài 10 : Một hộp bi gồm 5 bi đỏ, 4 bi trắng và 3 bi xanh có cùng cỡ. Từ hộp ta rút ngẫu nhiên không hoàn lại từng bi một cho đến khi được bi đỏ thì dừng lại. Tính xác suất để được 2 bi trắng, 1 bi xanh và 1 bi đỏ.

A. 0,045 B. 0,014 C. 0,032 D. 0,12

Bài 11: Sản phẩm X bán ra ở thị trường do một nhà máy gồm hai phân xưởng I và II sản xuất, trong đó phân xưởng I chiếm 60%; phân xưởng II chiếm 40%.Tỷ lệ phế phẩm do hai phân xưởng I và II sản xuất lần lượt là 0,5% và 0,7%. Chọn mua ngẫu nhiên 1 sản phẩm X ở thị trường.Tính xác suất để mua nhầm phế phẩm.

A. 0,025 B. 0,0014 C. 0,002 D. 0,0058

Bài 12: Bạn quên số cuối cùng trong 8 số của số điện thoại và quay nó một cách ngẫu nhiên. Hãy tính xác suất để bạn quay đúng số mà không quá 3 lần quay.

A. 0,3 B. 0,44 C. 0,02 D. 0,28

Bài 13: Có 2 kiện hàng. Kiện thứ nhất có 12 sp loại I và 8 sp loại II, kiện thứ 2 có 6 sp loại I và 4 sp loại II. Chọn ngẫu nhiên 1 kiện rồi từ kiện đó lấy ngẫu nhiên 1 sp. Tính xác suất để chọn được sp loại I.

A. 0,5 B. 0,6 C. 0,4 D. 0,45

Bài 14.Một thủ kho có 10 chiếc chìa khóa bên ngoài giống hệt nhau nhưng chỉ có hai chiếc mở được cửa nhà kho. Anh ta thử mở cửa từng chiếc một, chiếc nào không mở được thì bỏ ra ngoài. Tính xác suất để người thủ kho mở được cửa ở lần mở thứ 4.

A. 0,12 B. 0,13 C. 0,1 D. 0,14

Bài 15.Sản phẩm X trên thị trường là do ba cơ sở sản xuất cung cấp: Cơ sở I chiếm 25% số lượng hàng với tỉ lệ phế phẩm là 1%; Cơ sở II chiếm 25% số lượng hàng với tỉ lệ phế phẩm là 5%; Cơ sở III chiếm 50% số lượng hàng với tỉ lệ phế phẩm là 10%. Một người mua một sản phẩm X trên thị trường và mua được sản phẩm phế phẩm. Tính xác suất để sản phẩm này là sản phẩm của cơ sở II cung cấp.

A. 0,232 B. 0,192 C. 0,645 D. 0,483

Bài 16.Có hai thùng đựng cùng một loại sản phẩm giống nhau. Thùng thứ nhất chứa 12 sản phẩm trong đó có 8 sản phẩm loại A và 4 sản phẩm loại B; thùng thứ hai chứa 12 sản phẩm trong đó có 7 sản phẩm loại A và 5 sản phẩm loại B. Người ta lấy ngẫu nhiên từ thùng thứ nhất 1 sản phẩm bỏ vào thùng thứ hai sau đó lấy ngẫu nhiên từ thùng thứ hai 4 sản phẩm ra ngoài. Tính xác suất để 4 sản phẩm chọn ra sau cùng đều là những sản phẩm loại A.

A. 0,071 B. 0,082 C. 0,91 D. 0,054

Bài 17. Có 3 nhóm xạ thủ tập bắn. Nhóm thứ nhất có 5 người, nhóm thứ hai có 7 người và nhóm thứ ba có 4 người. Xác suất bắn trúng của mỗi người trong một nhóm là như nhau và xác suất bắn trúng của mỗi người thuộc 3 nhóm tương ứng là 0,8; 0,7; 0,6. Chọn ngẫu nhiên 1 xạ thủ và biết xạ thủ này bắn trượt. Hãy xét xem khả năng xạ thủ này thuộc vào nhóm nào?

Bài 18. Có 2 lô hàng, lô I chứa 8 sản phẩm tốt và 6 sp xấu; lô II chứa 5 sp tốt và 10 sp xấu. Từ mỗi lô chọn 1 sp. Sau đó, từ 2 sp thu được chọn ra 1sp. Tính xác suất để sp chọn được sau cùng là sp tốt.

Bài 19. Một lô hàng gồm 60 sản phẩm trong đó có 5 phế phẩm. Lô hàng được chấp nhận nếu chọn ngẫu nhiên 30 sản phẩm để kiểm tra thì số phế phẩm không quá 2. Tính xác suất để lô hàng được chấp nhận.

Bài 20. Một hộp đựng 5 bi trắng và 10 bi đen. Lấy lần lượt không hoàn lại 3 viên bi (mỗi lần lấy 1 viên). Tính xác suất để viên bi lấy lần thứ nhất và thứ 2 màu trắng, viên bi lấy lần thứ 3 là đen.

Bài 22. Chọn lần lượt không hoàn lại mỗi lần một lá bài từ một bộ bài 52 lá đến khi đủ 2 lá Át thì dừng. Tính xác suất để việc chọn dừng lại ở lần chọn thứ 3.

Bài 22. Một đám đông có số đàn ông bằng một nửa số phụ nữ. Xác suất để đàn ông bị bệnh tim là 0,06 và phụ nữ là 0,036. Chọn ngẫu nhiên 1 người từ đám đông, tính xác suất để người này bị bệnh tim.

Bài 23. Trong một trường đại học, tỷ lệ sinh viên học tiếng Nhật là 9%, học tiếng Anh là 12%, học cả hai thứ tiếng Anh và Nhật là 7%. Chọn ngẫu nhiên một sinh viên của trường đó, biết sinh viên đó học tiếng Nhật, tính xác suất để sinh viên đó học tiếng Anh.

Bài 24. Có hai lô hàng, mỗi lô chứa 15 sản phẩm, trong đó lô I gồm 13 sp tốt, 2 sp xấu; lô II gồm 8 sp tốt và 7 sp xấu. Chọn ngẫu nhiên từ lô I ra 2 sp bỏ sang lô II, sau đó từ lô II lấy ra 1 sp. Tính xác suất để trong sản phẩm chọn ra từ lô II là sản phẩm tốt.

Bài 25. Một người buôn bán bất động sản đang cố gắng bán một mảnh đất lớn. Ông ta tin rằng nếu nền kinh tế tiếp tục phát triển, khả năng mảnh đất được mua là 80%; ngược lại nếu nền kinh tế ngừng phát triển, ông ta chỉ có thể bán được mảnh đất đó với xác suất 40%. Theo dự báo của một chuyên gia kinh tế, xác suất nền kinh tế tiếp tục tăng trưởng là 65%. Tính xác suất để bán được mảnh đất.

Bài 26. Một thùng kín đựng 2 loại thuốc: Số lượng lọ thuốc loại A bằng 2/3 thuốc số lượng lọ thuốc loại B. Tỉ lệ lọ thuốc A, B đã hết hạn sử dụng lần lượt là 10% và 8%. Từ thùng lấy ngẫu nhiên một lọ thuốc và được lọ thuốc còn hạn sữ dụng, tính xác suất lọ này là lọ thuốc B.

Bài 27. Có 3 bình đựng bi: bình I có 4 bi trắng và 6 bi đen; bình II có 7 bi trắng và 3 bi đen; bình III có 6 bi trắng và 8 bi đen. Từ bình I và bình II, mỗi bình lấy 1 bi và bỏ sang bình III. Tiếp theo, từ bình III lấy ra tiếp 3 bi. Tính xác suất 3 bi lấy ra từ bình III có hai bi trắng.

Bài 28. Có ba lô hàng mỗi lô có 20 sản phẩm, số sản phẩm loại A có trong lô I, II, III lần lượt là: 12; 14; 16. Bên mua chọn ngẫu nhiên từ mỗi lô hàng 3 sản phẩm, nếu lô nào cả 3 sản phẩm đều loại A thì bên mua nhận mua lô hàng đó. Tính xác suất có ít nhất một lô được mua

Chương 2. BIẾN NGẪU NHIÊN VÀ CÁC LUẬT PHÂN PHỐI XÁC SUẤT THÔNG DỤNG

MỤC TIÊU CHƯƠNG

Nội dung chương này giúp người học có khả năng:

- Hiểu được khái niệm biến ngẫu nhiên rời rạc, liên tục và các tham số đặc trưng của biến ngẫu nhiên .
- Nhận biết được một số quy luật phân phối xác suất thông dụng: Nhị thức, Poisson, Siêu bội và phân phối chuẩn.
- Vận dụng được các quy luật phân phối xác suất vào các bài toán tính xác suất.

1. BIẾN NGẪU NHIÊN (Random Variables)

1.1 Định nghĩa

Một *biến ngẫu nhiên X* là một hàm giá trị thực xác định trên một không gian mẫu Ω:

$$X:\Omega \to R$$
$$\omega \mapsto X(\omega) = x.$$

Người ta thường dùng các chữ cái in *X, Y, Z, …* để ký hiệu cho một biến ngẫu nhiên và chữ thường $x_1, x_2, ..., x_n$ để ký hiệu các giá trị của biến ngẫu nhiên *X* nào đó.

Không gian mẫu Ω	X
ω_1 →	x_1
ω_2 →	x_2
ω_3 →	x_3
…	…

Ví dụ 1. Bắn liên tiếp 2 viên đạn vào một mục tiêu. Có 4 khả năng xảy ra là: TT, TO, OT và OO. Nếu gọi Y là số viên đạn bắn trúng mục tiêu thì Y là một biến ngẫu nhiên nhận các giá trị là 0, 1, 2. Y có 3 giá trị nên Y có hữu hạn phần tử.

Ví dụ 2. Tung một đồng tiền đến khi xuất hiện mặt sấp thì dừng. Gọi *X* là số lần tung thì *X* là một biến ngẫu nhiên. Tập các giá trị của biến ngẫu nhiên *X* gồm $\{1,2,3,....,n,...\}$, là tập vô hạn đếm được các phần tử.

Ví dụ 3. Gọi *Z* là nhiệt độ tại một thời điểm trong một ngày quan sát thì *Z* là một biến ngẫu nhiên. Có vô hạn các giá trị của *Z* và ta không liệt kê được, chỉ có thể mô tả các giá trị của *Z* là đoạn nào đó của tập số thực. Do đó ta nói *Z* là tập vô hạn không đếm được các phần tử.

Ta có thể hiểu khái niệm ***đếm được*** như sau: Tập *X* gọi là đếm được nếu các giá trị của *X* là các **số nguyên**.

1.2 Phân loại biến ngẫu nhiên

Có hai loại biến ngẫu nhiên

- Biến ngẫu nhiên X được gọi là **rời rạc** nếu X chỉ nhận hữu hạn hoặc vô hạn **đếm được** các giá trị.
- Biến ngẫu nhiên X được gọi là **liên tục** nếu X nhận các giá trị là một khoảng **số thực**.

Ví dụ 4. Biến ngẫu nhiên X trong ví dụ 1 và 2 ở trên là rời rạc, biến ngẫu nhiên X trong ví dụ 3 là liên tục.

1.3 Luật phân phối của biến ngẫu nhiên

Để xác định một biến ngẫu nhiên, trước hết ta phải biết biến ngẫu nhiên đó có thể nhận những giá trị nào. Nhưng mặt khác ta phải biết nó nhận các giá trị đó với xác suất tương ứng là bao nhiêu.

Một hình thức biểu diễn mối quan hệ giữa các giá trị có thể có của biến ngẫu nhiên và các xác suất tương ứng được gọi là qui luật phân phối xác suất của biến ngẫu nhiên đó.

Để thiết lập qui luật phân phối xác suất của một biến ngẫu nhiên người ta có thể dùng: *bảng phân phối xác suất, hàm phân phối xác suất và hàm mật độ xác suất.*

1.4 Luật phân phối xác suất (Probability Distributions)

Mối liên hệ giữa các giá trị của một biến ngẫu nhiên với xác suất tương ứng của chúng được gọi là luật phân phối của biến ngẫu nhiên đó.

1.5 Bảng phân phối xác suất

Nếu X là một biến ngẫu nhiên rời rạc nhận các giá trị được sắp xếp theo thứ tự tăng dần $x_1, x_2, ..., x_n$ thì bảng phân phối xác suất của X gồm 2 dòng như sau:

X	x_1	x_2	…..	x_n
P	p_1	p_2	……	p_n

Dòng thứ nhất ghi các giá trị của X theo thứ tự tăng dần, dòng thứ hai ghi xác suất tương ứng để X nhận giá trị đó. Tức là: $p_i = P(X = x_i)$; $p_1 + p_2 + ... + p_n = 1$ hay **tổng dòng hai luôn bằng 1.**
Bảng phân phối xác suất của X còn được gọi là luật phân phối của X.

Ví dụ 5. Đo thị lực (X) của một số sinh viên trường đại học A, kết quả như sau:

X	5/10	6/10	7/10	8/10	9/10	10/10
Số sv	10	16	50	60	50	14

Hãy lập bảng phân phối xác suất của X.

Giải

Số sinh viên được khảo sát là: $n = 10 + 16 + 50 + 60 + 50 + 14 = 200$.

Xác suất để một sinh viên có thị lực 5/10 là: $p = \frac{10}{200} = 0,05$.

Tương tự cho các trường hợp còn lại. Ta được bảng phân phối xác suất của X là:

X	5/10	6/10	7/10	8/10	9/10	10/10
P	0,05	0,08	0,25	0,3	0,25	0,07

Chú ý: Các bước giải bài toán lập bảng ppxs
1) Gọi biến ngẫu nhiên X (đề bài yêu cầu lập bảng ppxs của đại lượng nào thì gọi đó là X)
2) Liệt kê tất cả những trường hợp có thể xảy ra
3) Mỗi trường hợp xét xem X bằng bao nhiêu và tính xác suất tương ứng.
4) Dựa vào các kết quả trên để lập bảng ppxs.
5) Cộng dòng xác suất kiểm tra xem phải bằng 1 mới đúng.

Ví dụ 6. Một xạ thủ có 4 viên đạn, bắn lần lượt từng viên vào một mục tiêu một cách độc lập. Xác suất trúng mục tiêu ở mỗi lần bắn là 0,7. Nếu có một viên trúng mục tiêu hoặc hết đạn thì dừng. Gọi X là số viên đạn đã bắn, lập bảng phân phối xác suất của X.
Giải:
……………………………………………………………………………………………
……………………………………………………………………………………………
………………………………………………………………………………………
……………………………………………………………………………………………
…………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………

Ví dụ 7. Một kiện hàng có 6 sản phẩm tốt và 4 phế phẩm. Chọn ngẫu nhiên từ kiện hàng đó ra 2 sản phẩm. Gọi X là số phế phẩm trong 2 sp chọn ra. Hãy lập bảng phân phối xác suất của X .
Giải:
……………………………………………………………………………………………
……………………………………………………………………………………………
………………………………………………………………………………………
……………………………………………………………………………………………
…………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
………………………………………………………………………………………
……………………………………………………………………………………………
…………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………

Ví du 8. Một khoa điều trị có 3 bác sỹ và 9 y tá. Chọn không oàn lại ngẫu nhiên từng người một cho đến khi gặp y tá thì dừng chọn. Hãy lập bảng phân phối xác suất của số bác sỹ được chọn.
Giải:
……………………………………………………………………………………………
……………………………………………………………………………………………
………………………………………………………………………………………
……………………………………………………………………………………………
…………………………………………………………………………………………

…………………………………………………………………………………………
…………………………………………………………………………………………
……………………………………………………………………………………
…………………………………………………………………………………………
……………………………………………………………………………………

Ví dụ 9. Cho hai biến ngẫu nhiên rời rạc X và Y có bảng phân phối xác suất:

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

Y	1	2
P	0,4	0,6

Hãy lập bảng phân phối xác suất của X + Y.
Giải:
…………………………………………………………………………………………
…………………………………………………………………………………………
……………………………………………………………………………………
…………………………………………………………………………………………
……………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
……………………………………………………………………………………
…………………………………………………………………………………………

Ví dụ 10. Cho biến ngẫu nhiên rời rạc X có bảng phân phối xác suất:

X	2	3	4	5	6
P	0,05	0,15	0,25	0,2	0,35

a. Tính xác suất p = P(X ≤ 4).
b. Tính xác suất p = P(3 ≤ X < 5).
Giải:
Vẽ biểu đồ
…………………………………………………………………………………………
…………………………………………………………………………………………
……………………………………………………………………………………
…………………………………………………………………………………………
……………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
……………………………………………………………………………………
…………………………………………………………………………………………

1.7 Hàm phân phối xác suất

Định nghĩa

Cho biến ngẫu nhiên X. Với mỗi số thực x, xác định duy nhất một biến cố (X ≤ x) và do đó có tương ứng một và chỉ một xác suất P(X ≤ x). Quan hệ tương ứng này cho ta một hàm số xác định trên R, hàm số này được ký hiệu là F(x).

Hàm phân phối xác suất *của một biến ngẫu nhiên X còn gọi là* ***hàm phân phối tích lũy****, ký hiệu* $F(x)$*, và được xác định bởi biểu thức:* $F(x)=P(X\le x)$.

a. Biến ngẫu nhiên rời rạc

Nếu *X* là BNN rời rạc nhận các giá trị là: $x_1<x_2<x_3<....<x_n$ thì $F(x)=\sum_{x_i\le x}P(X=x_i)$

tức là *X* có bảng phân phối xác suất:

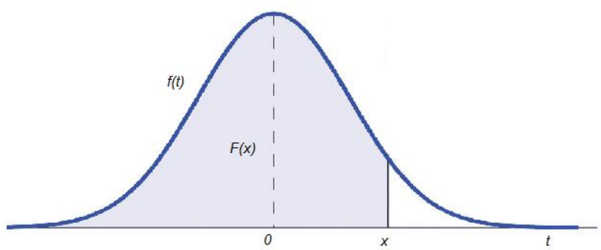
X	x_1	x_2	…..	x_n
P	p_1	p_2	……	p_n

và hàm phân phối xác suất của *X* có dạng:

$$F(x)=\begin{cases}0 & khi\ x<x_1\\ p_1 & khi\ x_1\le x<x_2\\ p_1+p_2 & khi\ x_2\le x<x_3\\ \text{........} & \\ p_1+p_2+...+p_{n-1} & khi\ x_{n-1}\le x<x_n\\ 1 & khi\ x\ge x_n.\end{cases}$$

b. Biến ngẫu nhiên liên tục

BNN liên tục X có hàm phân phối xác suất F(x). Nếu tồn tại hàm số *f(x)* sao cho: $F'(x)=f(x)$ thì hàm số *f(x)* được gọi là ***hàm mật độ xác suất*** của X.



Giá trị hàm $F(x)$.

Hàm mật độ $f(x)$ thỏa các tính chất sau:

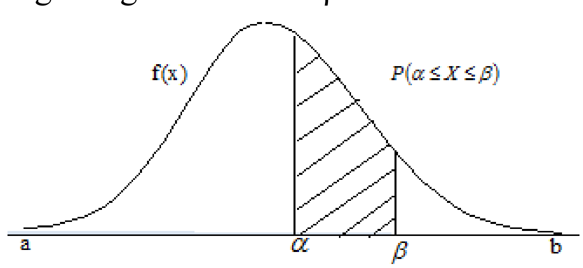
i) $f(x)\ge 0,\ \forall x\in R$

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

iii) $P(\alpha\le X\le\beta)=\int_{\alpha}^{\beta}f(x)dx,\ \ \alpha,\beta\in R$.

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

Về mặt hình học thì kết quả trên có thể minh họa như sau: Xác suất để biến ngẫu nhiên liên tục X nhận giá trị trong khoảng (α; β) bằng diện tích của hình thang cong giới hạn bởi trục Ox, đường cong f(x) và các đường thẳng x = α và x = β



Diện tích miền gạch chéo là xác suất để $\alpha\le X\le\beta$.

2. CÁC THAM SỐ ĐẶC TRƯNG CỦA BIẾN NGẪU NHIÊN

2.1 Kỳ vọng (Expectation) (trung bình)

Kỳ vọng (**trung bình**) của *X*, ký hiệu $E(X)$, là một số thực được xác định như sau:

- Nếu *X* là biến ngẫu nhiên rời rạc có bảng phân phối xác suất

X	x_1	x_2	…..	x_n
P	p_1	p_2	……	p_n

thì: $E(X) = x_1p_1 + x_2p_2 + ... + x_np_n$.

- Nếu *X* là biến ngẫu nhiên liên tục với hàm mật độ $f(x)$ có miền xác định $[a;b]$ thì: $E(X) = \int_a^b x.f(x)dx$.

Tính chất của kỳ vọng

Cho *X, Y* là 2 biến ngẫu nhiên, ta có:

i) $E(C) = C$, với *C* là hằng số.

ii) $E(X+Y) = E(X) + E(Y)$.

iii) $E(\lambda X) = \lambda.E(X)$, λ là hằng số.

Ví dụ 11. Cho X là biến ngẫu nhiên rời rạc có bảng phân phối xác suất

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

thì $E(X) = 1.0{,}5 + 2.0{,}2 + 3.0{,}3 = 1{,}8$.

2.2 Phương sai (Variance)

Cho *X* là biến ngẫu nhiên, ***phương sai*** của *X*, ký hiệu $V(X)$, là một số được xác định như sau:

- Nếu *X* là biến ngẫu nhiên rời rạc có bảng phân phối xác suất

X	x_1	x_2	…..	x_n
P	p_1	p_2	……	p_n

thì: $V(X) = \sum_{i=1}^{n} [x_i - E(X)]^2 p_i$.

- Nếu *X* là biến ngẫu nhiên liên tục với hàm mật độ $f(x)$ có miền xác định $[a;b]$ thì

$$V(X) = \int_a^b [X - E(x)]^2 f(x)dx = \int_a^b x^2 f(x)dx - [E(x)]^2.$$

Phương sai là giá trị trung bình của bình phương độ lệch giữa các giá trị của X so với trung bình của nó. Phương sai lớn có nghĩa là các giá trị của X cách xa giá trị trung bình, còn phương sai nhỏ thì các giá trị của X tập trung gần giá trị trung bình của nó.

Trong kinh doanh, phương sai đặc trưng cho độ rủi ro khi đầu tư. Trong kỹ thuật, phương sai đặc trưng cho sai số của thiết bị

Tính chất

Cho X, Y là 2 biến ngẫu nhiên, ta có:

i) $V(C) = 0$, với C là hằng số. ii) $V(X) \geq 0$.

iii) $V(\lambda X) = \lambda^2.V(X)$, với λ là hằng số. vi) $V(\lambda + X) = V(X)$, với λ là hằng số.

2.3 Độ lệch chuẩn

Cho X là biến ngẫu nhiên, khi đó ***độ lệch chuẩn*** của X ký hiệu là $\sigma = \sqrt{V(X)}$ (đọc là "xích ma").

Ý NGHĨA VÀ ỨNG DỤNG CỦA KỲ VỌNG TOÁN

• Kỳ vọng toán là giá trị trung bình theo xác suất của các giá trị của biến ngẫu nhiên. Nó thể hiện xu thế trung tâm của biến ngẫu nhiên ấy.

• Trong kinh tế, kỳ vọng đặc trưng cho năng suất trung bình của một phương án sản suất, lợi nhuận trung bình của một danh mục đầu tư, trọng lượng trung bình hoặc tuổi thọ trung bình của một loại sản phẩm,... Do đó trong kinh tế, kỳ vọng là một tiêu chuẩn để ra quyết định khi có nhiều phương án lựa chọn khác nhau.

Ý NGHĨA CỦA PHƯƠNG SAI

• Phương sai phản ánh mức độ phân tán của các giá trị của biến ngẫu nhiên xung quanh giá trị trung bình của nó là kỳ vọng toán.

• Tùy từng nội dung của biến ngẫu nhiên, phương sai đặc trưng cho: độ phân tán, độ biến động, độ rủi ro, độ ổn định, độ đồng đều, độ chính xác,… Do đó trong kinh tế, phương sai là một tiêu chuẩn để ra quyết định trong trường hợp có nhiều phương án lựa chọn khác nhau.

• Lưu ý:

- Phương sai càng lớn thì ta nói BNN càng biến động, càng dao động, càng phân tán…
- Phương sai càng nhỏ thì ta nói BNN càng ổn định, càng tập trung, càng đồng đều…
- Đơn vị của phương sai là bình phương đơn vị của BNN, chẳng hạn nếu X đơn vị là mét thì V(X) có đơn vị là mét^2.

Ví dụ 12. Một trường học thi môn Tiếng Anh với kết quả như sau:

Điểm (X)	4	5	7	8
Xác suất p	0,15	0,3	0,35	0,2

Tính trung bình, phương sai và độ lệch chuẩn của *X*.

Giải

……………………………………………………………………………………………
……………………………………………………………………………………………
………………………………………………………………………………….
……………………………………………………………………………………………
……………………………………………………………………………………

………………………………………………………………………………………
………………………………………………………………………………………
……………………………………………………………………………….
………………………………………………………………………………………
…………………………………………………………………………………
………………………………………………………………………………………
………………………………………………………………………………………
……………………………………………………………………………….

Ví dụ 13. Cho biến ngẫu nhiên rời rạc X có bảng phân phối xác suất:

X	3	5	6	7
P	0,1	0,3	0,35	0,25

Tính trung bình, phương sai và độ lệch chuẩn của X.
Giải:

……………………………………………………………………………………
………………………………………………………………………………………
…………………………………………………………………………………….

Ví dụ 14. Cho biến ngẫu nhiên rời rạc X có bảng phân phối xác suất:

X	2	4	5	7
P	0,2	0,4	0,1	0,3

Tính $E(3X + 2)$
Giải

………………………………………………………………………………………
………………………………………………………………………………………
……………………………………………………………………………….
………………………………………………………………………………………
…………………………………………………………………………………
………………………………………………………………………………………

Ví dụ 15. Một cửa hàng điện máy bán 1 cái máy lạnh A thì lời 0,8 triệu đồng nhưng nếu máy lạnh đó phải bảo hành thì lỗ 1 triệu đồng. Biết xác suất máy lạnh A phải bảo hành của cửa hàng là 10%, tính mức lời trung bình khi bán 1 máy lạnh A ?
Giải:

………………………………………………………………………………………
………………………………………………………………………………………
……………………………………………………………………………….
………………………………………………………………………………………
…………………………………………………………………………………
………………………………………………………………………………………
………………………………………………………………………………………
……………………………………………………………………………….
………………………………………………………………………………………
…………………………………………………………………………………

Ví dụ 16. Một nhân viên bảo hiểm sẽ được thưởng hàng tháng 3 triệu đồng nếu anh ta kí được 1 hợp đồng. Biết số hợp đồng mà các nhân viên kí được và tiền thưởng có bảng phân phối xác suất sau.

Số hợp đồng bảo hiểm	0	1	2	3
Tiền thưởng (triệu đồng)	0	3	6	9
P	0,1	0,3	x	0,2

Kỳ vọng số tiền thưởng hàng tháng của anh ta là:
Giải:
…………………………………………………………………………………………
…………………………………………………………………………………………
………………………………………………………………………………….
…………………………………………………………………………………………

Ví dụ 17. Theo thống kê, một người Mỹ 25 tuổi sẽ sống thêm trên 1 năm có xác suất là 0,992 và người đó chết trong vòng 1 năm tới là 0,008. Một công ty bảo hiểm A đề nghị người đó bảo hiểm sinh mạng cho 1 năm với số tiền chi trả là 10000 USD, phí bảo hiểm là 100 USD. Hỏi trung bình công ty A lãi bao nhiêu khi bán bảo hiểm cho người đó? (20USD)

Giải.
………………………………………………………………………………….
…………………………………………………………………………………………
…………………………………………………………………………………………
………………………………………………………………………………….

Bài thảo luận: Nhu cầu hằng ngày của 1 khu phố về 1 loại thực phẩm tươi sống có bảng phân phối xác suất:

Nhu cầu (kg)	31	32	33	34
P	0.15	0,25	0,45	0,15

Một cửa hàng trong khu phố nhập về mỗi ngày 34 kg loại thực phẩm này với giá 25.000 đồng/kg và bán ra với giá 40.000 đồng/kg. Nếu bị ế, cuối ngày cửa hàng phải bán hạ giá còn 15.000 đồng/kg mới bán hết hàng. Tính tiền lời trung bình của cửa hàng này về loại thực phẩm trên trong 1 ngày.
Giải.
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………

4. Dùng máy tính bỏ túi fx- 500ES, fx- 570ES, Plus tính trung bình và độ lệch chuẩn:

1. Mở chế độ bảng thống kê có **2 cột**: ***Ấn lần lượt 5 phím*** (những lần sau không cần bấm lại).

SHIFT **MODE** **RELAY**(nút mũi tên xuống ↓)

4(start) **1(on)**

Vào tính thống kê bấm lần lượt 3 phím sau: **MODE** **3(Start)** **1(1-Var)**

Ví dụ:

X	2	4	5	7
P	0,2	0,4	0,1	0,3

2. Nhập số liệu:

Nhập hàng thứ nhất X:

Bấm: **2** =
4 =
5 =
7 =

Bấm nút: **Relay** (->) để qua cột thứ hai, rồi dời dấu nháy lên đầu cột nhập tiếp hàng 2.

Nhập hàng thứ hai P:

Bấm: **0.2** =
0.4 =
0.1 =
0.3 =

Nhập toàn bộ dữ liệu xong bấm phím : **AC**

a.Tính trung bình : ***Ấn lần lượt 5 phím***

Bấm: **SHIFT** **1** **4(Var)** **2**$(\overline{X})$ =

Kết quả: $\overline{X}$ = 4,6

b.Tính độ lệch chuẩn**:** ***Ấn lần lượt 5 phím***

Bấm: **SHIFT** **1** **4(Var)** **3**$(x\sigma n)$ =

Kết quả: σ = 1,8

c. Tính phương sai: bấm bình phương độ lệch chuẩn: $\sigma^2 = (1,8)^2 = 3,24$

3**.** Thoát khỏi chế độ tính thống kê ấn**:** **MODE** **1**

5. Loại máy fx 580 VN

Ví dụ:

X	2	4	5	7
P	0,2	0,4	0,1	0,3

Bấm: SHIFT - MENU – ▼ - 3(Thống kê) – 1(Mở)

Bấm: MENU – 6 - 1(Tính Tkê 1 – biến) - Nhập dữ liệu- AC

Bấm: OPTN – 2(Tính 1-biến)

Kết quả: Trung bình: $\overline{X} = 4,6$

Phương sai: $\sigma^2 x = (1,8)^2 = 3,24$

Độ lệch chuẩn: $\sigma x = 1,8$

3. QUI LUẬT PHÂN PHỐI XÁC SUẤT CỦA BIẾN NGẪU NHIÊN RỜI RẠC

3.1 Luật phân phối nhị thức

Một biến ngẫu nhiên rời rạc có luật phân phối xác suất mà ta thường gặp trong các tình huống nghiên cứu kinh tế, xã hội, y học … là phân phối nhị thức. Đó là những hiện tượng mà kết quả của phép thử luôn là 2 biến cố xung khắc $A, \overline{A}$ và xác suất biến cố *A* xảy ra **không thay đổi** trong các lần thử độc lập.

Định nghĩa

Một phép thử được thực hiện độc lập *n* lần trong những điều kiện như nhau, ở mỗi lần biến cố *A* xảy ra với **xác suất bằng** p **không đổi** và không xảy ra với xác suất $(1-p)$. Gọi *X* là số lần biến cố *A* xảy ra trong *n* lần thử, thì *X* được gọi là biến ngẫu nhiên có ***phân phối nhị thức***, ký hiệu $X \sim B(n;p)$.

Xác suất để trong *n* lần thử có *k* lần biến cố *A* xảy ra là: $P(X=k) = C_n^k p^k (1-p)^{n-k}$.

Tính chất. Nếu $X \sim B(n;p)$ thì $E(X) = np$, $\sigma^2 = V(X) = np(1-p)$.

Chú ý: Khi giải bài toán có phân phối nhị thức

1) Dấu hiệu của pp nhị thức là xác suất không đổi.

2) Gọi X là số lần biến cố A xảy ra trong n lần

3) X là số tự nhiên nhỏ nhất bằng 0 và lớn nhất bằng n

4) Nếu đề bài cho trước tỷ lệ %, hoặc cho trước xác suất thì bài toán là pp nhị thức.

5) Bấm máy tính cho công thức tính xác suất của phân phối nhị thức như sau:

$$P(X=k) = \sum_{X=k}^{k} \left(nCX.p^X.(1-p)^{n-X} \right) =$$

$$P\left(X \geq k\right) = \sum_{X=k}^{n}\left(nCX.p^{X}.\left(1-p\right)^{n\text{-}X}\right) =$$

$$P\left(X \leq k\right) = \sum_{X=0}^{k}\left(nCX.p^{X}.\left(1-p\right)^{n\text{-}X}\right) =$$

Ví dụ 18. Cho $X \sim B(12;0,4)$, tính $P(X<8)$ và $P(X \geq 6)$

Giải

…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………

Ví du 19. Tỉ lệ cử tri ủng hộ ứng cử viên A trong một cuộc bầu cử là 60%. Người ta hỏi ý kiến 20 cử tri được chọn một cách ngẫu nhiên. Gọi X là số người ủng hộ cho ông A trong số 20 người đó.

a. Tìm $P(X \leq 10)$

b. Tìm $P(X > 12)$

c. Tìm $P(X = 11)$

d. Tìm $P(X < 15)$

e. Tìm $P(14 \leq X \leq 18)$

f. Tìm $P(X \geq 17)$

Ví dụ 20. Một phân xưởng có 15 máy hoạt động độc lập, xác suất để trong một ngày mỗi máy bị hỏng đều bằng 0,08. Tìm xác suất để trong một ngày có không quá 4 máy hỏng.
Giải:

…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………

Ví dụ 21. Một máy có xác suất sản xuất ra sản phẩm loại tốt là 80%. Cho máy sản suất 40 sản phẩm. Tính xác suất để trong 40 sản phẩm do máy sản xuất ra có ít hơn 10 sản phẩm xấu.
Giải.

…………………………………………………………………………………………

……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………

Ví dụ 22. Một nhà máy kiểm tra chất lượng hàng theo quy trình sau: Từ mỗi lô hàng lấy ngẫu nhiên ra lần lượt 15 sản phẩm để kiểm tra, nếu tất cả đều tốt thì lô hàng được chấp nhận, nếu ngược lại thì trả lại để kiểm tra toàn bộ. Biết rằng xác suất để được một sản phẩm tốt là 0,9. Hãy tìm xác suất để lô hàng được nhận.

Giải:

……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
………………………………………………………………………………………

Ví dụ 23. Một bài thi trắc nghiệm có 30 câu, mỗi câu có 4 đáp án, trong đó chỉ có 1 đáp án đúng và điểm bài thi làm tròn số nguyên tối đa là 10 điểm. Một sinh viên làm hết bài thi bằng cách chọn mỗi câu ngẫu nhiên một đáp án, tính xác suất để sinh viên này được từ 5 điểm trở lên.

Giải:

……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………

Ví dụ 24. Một máy sản xuất một loại sản phẩm có tỉ lệ sản phẩm loại A của máy là 80% . Cho máy đó sản xuất ra 3 sản phẩm. Lập bảng phân phối xác suất của số sản phẩm loại A có trong 3 sản phẩm sản suất ra.

Giải:

……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
…………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
…………………………………………………………………………………………
……………………………………………………………………………………………

Ví dụ 25. Một lô hàng có tỷ lệ sản phẩm loại A là 65%. Hỏi phải lấy từ lô hàng ra ít nhất bao nhiêu sản phẩm để cho xác suất của biến cố: “có ít nhất một sản phẩm loại A trong số các sản phẩm lấy ra” không bé hơn 90%?

Giải:

……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………

Ví dụ 26. Một lô hàng chứa 20 sản phẩm trong đó có 4 phế phẩm. Chọn liên tiếp 3 lần (có hoàn lại) từ lô hàng, mỗi lần chọn ra 4 sản phẩm. Tính xác suất để trong 3 lần chọn có đúng 1 lần chọn được 4 phế phẩm.

Giải:

……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………

3.2 Luật phân phối POISSON

Những hiện tượng xảy ra trong thực tế mà ta quan tâm đến số lần một biến cố cụ thể xảy ra trong một đơn vị thời gian hay không gian xác định là phân phối Poisson. Như số lỗi trên 1 trang đánh máy, số tai nạn giao thông trong 1 giờ…

Định nghĩa

Giả sử một phép thử mà biến cố *A* xảy ra trong một **khoảng thời gian (hoặc trên một vùng, miền).** Biết **trung bình** số lần biến cố *A* xuất hiện trong khoảng thời gian đó là λ. Gọi *X* là số lần xuất hiện của biến cố *A* trong **khoảng thời gian** đó thì *X* gọi là biến ngẫu nhiên có ***phân phối Poisson***, ký hiệu là $X \sim P(\lambda)$.

Xác suất biến cố *A* xuất hiện k lần là*:* $P(X=k)=\dfrac{e^{-\lambda}\lambda^k}{k!}$.

Tính chất: Nếu *X* có phân phối Poisson $X \sim P(\lambda)$ thì $E(X)=V(X)=\lambda$.

Chú ý: Khi giải bài toán pp Poisson

1) Dấu hiệu của pp Poisson là biến cố A xảy ra trong một khoảng thời gian (hoặc trên một vùng, miền).

2) ***Nếu X có pp Poisson thì X là số tự nhiên nhỏ nhất bằng 0 và lớn nhất là vô hạn.***

3) Khoảng thời gian hay vùng miền là trong câu hỏi của đề bài.

4) λ ***là trung bình số lần biến cố A xảy ra trong khoảng thời gian đề bài hỏi***

5) Bấm máy tính cho công thức tính xác suất của phân phối Poisson như sau:

$$P\left(X=k\right)=\sum_{X=k}^{k}\frac{e^{-\lambda}.\lambda^X}{X!}=$$

$$P\left(X \le k\right) = \sum_{X=0}^{k} \frac{e^{-\lambda}.\lambda^{X}}{X!} =$$

$$P\left(X > k\right) = 1 - P\left(X \le k\right) = 1 - \sum_{X=0}^{k} \frac{e^{-\lambda}.\lambda^{X}}{X!} =$$

Ví dụ 26. Cho $X \sim P(\lambda = 9)$. Hãy tính $P(X \le 6)$ và $P(X \ge 5)$

Giải

…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………

Ví dụ 27. Tin nhắn gởi đến tổng đài điện thoại tuân theo qui luật Poisson có trung bình là 9 (tin/30 phút). Tìm xác suất cho các trường hợp sau:

a. Có 10 tin nhắn đến trong 1 giờ.

b. Không có tin nhắn nào đến trong 10 phút.

c. Có không quá 5 tin nhắn trong 20 phút.

d. Có trên 30 tin nhắn trong 2 giờ.

Giải

…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
……………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………

Ví dụ 28. Ở một tổng đài điện thoại, các cuộc điện thoại gọi đến xuất hiện ngẫu nhiên, độc lập với nhau và trung bình 70 cuộc gọi trong 10 phút. Tính xác suất có hơn 6 cuộc gọi trong 1 phút.

Giải.

…………………………………………………………………………………………………
…………………………………………………………………………………………………
……………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………

Ví dụ 29. Tại một cửa hàng, trong 1 giờ có trung bình 20 khách hàng đến mua hàng. Tính xác suất trong 30 phút tại cửa hàng đó không quá 8 người đến mua hàng.
Giải.

…………………………………………………………………………………………………
…………………………………………………………………………………………………
………………………………………………………………………………………………
…………………………………………………………………………………………………

Ví dụ 30. Một trường Đại Học có quẹt thẻ xe điện tử ở cửa ra vào. Thông thường cứ quẹt 20 xe thì sẽ có 2 xe có vấn đề khi lấy thẻ. Tính xác suất buổi sáng có đúng 40 xe có vấn đề khi lấy thẻ biết rằng có 300 xe vào trường.
Giải.

…………………………………………………………………………………………………
…………………………………………………………………………………………………
………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………

Ví dụ 31. Tại một máy rút tiền ATM của 1 ngân hàng trung bình 10 phút có 1 khách hàng đến giao dịch. Tìm xác suất để có ít nhất 1 người đến giao dịch trong khoảng thời gian 5 phút.
Giải:

…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
……………………………………………………………………………………………

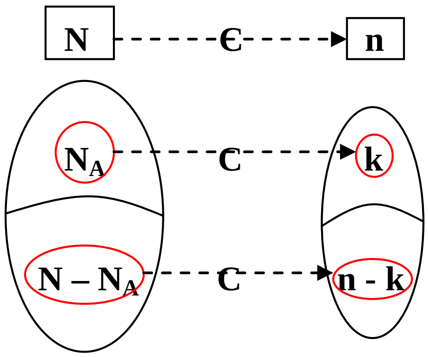
Ví dụ 32. Một ga ra cho thuê ôtô thấy rằng số người đến thuê ôtô vào ngày thứ 7 cuối tuần là một biến ngẫu nhiên X có phân phối Poisson với trung bình 4 người. Giả sử ga ra có 6 ôtô và mỗi người đến chỉ được thuê một chiếc. Hãy tìm xác suất để vào ngày thứ 7 Ga ra không đáp ứng được yêu cầu.
Giải:

…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………

3.3 Luật phân phối siêu bội (Hypergeometric Distribution)

Một tập hợp có *N* phần tử gồm **2 loại:** phần tử loại A và phần tử không phải loại A, trong đó số phần tử loại *A* là N_A. Lấy từ tập hợp này ra *n* phần tử (không hoàn lại). Gọi *X* là số phần tử loại *A* có trong *n* phần tử lấy ra. Khi đó *X* gọi là có *phân phối siêu bội*, ký hiệu là $X \sim H(N; N_A; n)$.

Xác suất có k phần tử loại A trong n phần tử là: $P(X=k)=\dfrac{C_{N_A}^{k}C_{N-N_A}^{n-k}}{C_N^n}$.



Tính chất. Nếu $X\sim H(N;N_A;n)$ thì: $E(X)=np,\ p=\dfrac{N_A}{N}$, $V(X)=np\left(1-p\right)\dfrac{N-n}{N-1}$.

Chú ý:

1) X là số tự nhiên nhỏ nhất bằng 0 và lớn nhất bằng n

2) Bấm máy tính xác suất :

$$P\left(X\le k\right)=\sum_{X=0}^{k}\frac{C_{N_A}^{X}.C_{N-N_A}^{n-X}}{C_N^n}$$

Ví dụ 33. Sản phẩm của một nhà máy sản xuất xong được đóng thành từng kiện. Mỗi kiện chứa 20 sản phẩm trong đó có 12 sản phẩm loại A và 8 sản phẩm loại B. Chọn ngẫu nhiên một kiện rồi từ kiện đó chọn ngẫu nhiên 6 sản phẩm. Tính xác suất để trong 6 sản phẩm chọn ra có hơn 4 sản phẩm loại A.

Giải.

……………………………………………………………………………………………

…………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

…………………………………………………………………………………

Ví dụ 34. Một kiện hàng gồm 100 sản phẩm, biết rằng tỉ lệ sản phẩm lỗi trong kiện hàng là 5%. Lấy ngẫu nhiên 20 sản phẩm trong kiện hàng này. Tính xác suất để có 2 sản phẩm bị lỗi.

Giải

……………………………………………………………………………………………

…………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

…………………………………………………………………………………

Ví dụ 35. Một nhà máy sản xuất một loại sản phẩm. Sản phẩm được đóng thành các lô hàng, mỗi lô gồm 40 sản phẩm, theo qui định gồm 37 sản phẩm loại A và 3 sản phẩm loại B. Bộ phận kiểm tra sẽ chọn ngẫu nhiên từ lô hàng 10 sản phẩm, nếu có quá 1 sản phẩm loại B thì lô hàng bị trả lại. Tính xác suất một lô hàng được chấp nhận.

Giải

…………………………………………………………………………………………
…………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
……………………………………………………………………………………

Ví dụ 36. Giả sử 18 công ty máy tính lớn hoạt động tại Hoa Kỳ, trong đó có 12 công ty được đặt tại thung lũng Silicon California. Nếu 3 công ty máy tính được chọn ngẫu nhiên trong toàn bộ danh sách 18 công ty trên, thì xác suất có nhiều hơn 1 công ty được chọn nằm ở thung lũng Silicon là bao nhiêu?

Giải

…………………………………………………………………………………………
…………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
……………………………………………………………………………………

4 MỘT SỐ QUI LUẬT PHÂN PHỐI CỦA BIẾN NGẪU NHIÊN LIÊN TỤC

4.1 Phân phối chuẩn (Normal distribution)

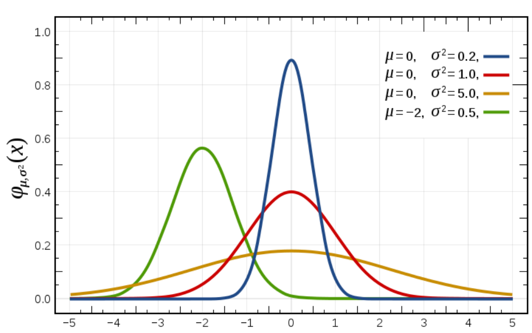
Phân phối chuẩn được Carl Friedrich Gauss, một nhà toán học và nhà khoa học người Đức tìm ra năm 1809 nên nó còn được gọi là phân phối Gauss. Trong thực tế người ta nhận thấy rằng nhiều biến số có thể mô tả bằng phân phối chuẩn hoặc gần như phân phối chuẩn (Định lý giới hạn trung tâm) như: các đại lượng vật lý, thiên văn, điểm thi, năng suất, mức lãi suất, huyết áp, thân nhiệt, nồng độ hemoglobin…

Định nghĩa

Biến ngẫu nhiên liên tục X có *phân phối chuẩn* với tham số $\mu;\sigma^2$, được ký hiệu là $X \sim N\left(\mu;\sigma^2\right)$,

nếu hàm mật độ xác suất có dạng: $f(x) = \dfrac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}; \qquad \forall x \in R, \sigma \neq 0$,

trong đó μ là giá trị trung bình đọc là "muy", σ là độ lệch chuẩn, σ^2 là phương sai của X.



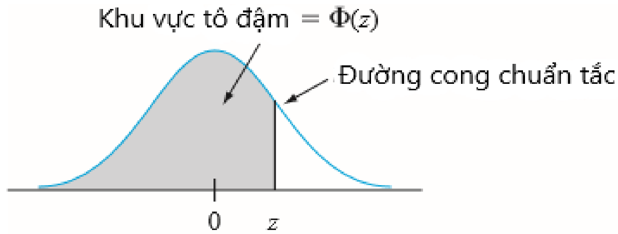
Đồ thị hàm mật độ phân phối chuẩn là hình chuông.

Tính chất : Đồ thị hàm mật độ $f(x)$ là một hình chuông cân đối, có hoành độ đỉnh là trung bình μ và có diện tích bằng 1.

Định nghĩa phân phối chuẩn tắc

Biến ngẫu nhiên Z có phân phối chuẩn tắc nếu Z là phân phối chuẩn với $\mu=0,\sigma=1$, ký hiệu: $Z \sim N(0;1)$. Hàm mật độ của phân phối chuẩn tắc: $f(z)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}$, $-\infty<z<+\infty$.

Hàm phân phối xác suất của phân phối chuẩn tắc: $F(z)=\Phi(z)=P(Z\le z)=\int_{-\infty}^{z} f(t)dt=\int_{-\infty}^{z}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}t^2}dt$.



Giá trị hàm $\Phi(z)=P(Z\le z)$.

Hình chuông của phân phối chuẩn tắc đối xứng qua trục tung và có tổng diện tích bằng 1 (đơn vị diện tích). Để đơn giản trong thực hành người ta tính sẵn các giá trị của hàm $\Phi(z)$ thành 2 bảng có tên là **Bảng phân phối chuẩn tắc của giá trị dương** và **Bảng phân phối chuẩn tắc của giá trị âm.**

$$\Phi(z)=P(Z\le z)=\int_{-\infty}^{z} f(t)dt=\int_{-\infty}^{z}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}t^2}dt.$$

Tính chất

1) $P(a<z<b)=\Phi(b)-\Phi(a)$

2) $P(z>a)=1-\Phi(a)$

3) $P(z<a)=\Phi(a)$

4) Nếu $z>3,5$ thì $\Phi(z)=1$ và nếu $z<-3,5$ thì $\Phi(z)=0$

Chú ý: *Đối với phân phối chuẩn tắc thì xác suất khi z nhận một giá trị cụ thể là bằng 0, tức là* $P(z=k)=0$. *Do đó các công thức trên vẫn đúng nếu các dấu bất đẳng thức là:* $\le;\ge$.

Ví dụ 37. Cho Z là BNN có phân phối chuẩn tắc, tính: $P(Z\le 1,25)$, $P(Z>1,05)$, $P(Z\le -2,14)$

Giải.

....
....
....
....
....
....
....
....
....
....

Ví dụ 38. Cho Z là BNN có phân phối chuẩn tắc, tính: $P(-0,56 \le Z \le 0,85)$.

Giải.

....

....

....

....

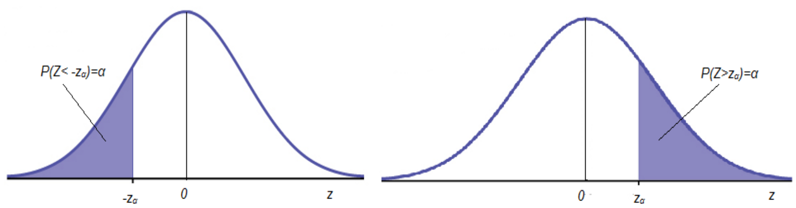
....

....

Điểm tới hạn mức α (phân vị mức α)

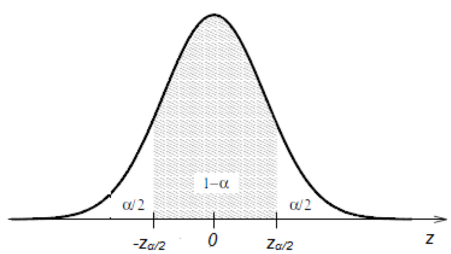
Cho biến ngẫu nhiên $Z \sim N(0;1)$, điểm $z_\alpha \ge 0$ gọi là điểm tới hạn *mức α* nếu thỏa: $P(z > z_\alpha) = \alpha$ hoặc $P(z < -z_\alpha) = \alpha$.

Để tính điểm tới hạn mức α là z_α ta dùng công thức $\Phi(z_\alpha) = 1 - \alpha$. **(**tra ngược bảng hàm $\Phi(z)$**).**



Điểm tới hạn mức α.

Tương tự, tính điểm tới hạn mức α/2 là $z_{\alpha/2}$ ta dùng công thức $\Phi(z_{\alpha/2}) = 1 - \frac{\alpha}{2}$ (tra ngược bảng hàm $\Phi(z)$) **.**



Điểm tới hạn mức α/2.

Ví dụ 39. Cho $\alpha = 0,06$, hãy tính mức phân vị $z_{\alpha/2}$ và z_α.

Giải

....

....

....

....

....

....

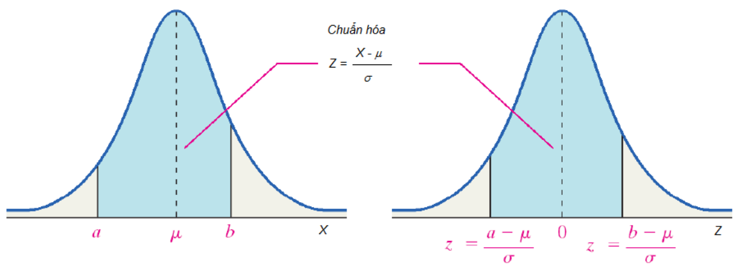
Ví dụ 40. Với $\alpha = 0,03$, tính $z_{\alpha/2}$

………………………………………………………………………………………………
………………………………………………………………………………………………
……………………………………………………………………………………
………………………………………………………………………………………………

Ví dụ 41. Với $\alpha = 0,02$, tính z_{α}

………………………………………………………………………………………………
………………………………………………………………………………………………

Chuẩn hóa số liệu Nếu $X \sim N(\mu;\sigma^2)$ thì $Z = \dfrac{X-\mu}{\sigma} \sim N(0,1^2)$.



Biến ngẫu nhiên chuẩn hóa.

Ví dụ 42. Cho $X \sim N(50;100)$. Tính $P(X \le 45)$.

Giải

Ta có $\mu = 50,\ \sigma = \sqrt{100} = 10$

$$P(X \le 45) = P\left(\frac{X-\mu}{\sigma} \le \frac{45-\mu}{\sigma}\right) = P\left(\frac{X-\mu}{\sigma} \le \frac{45-50}{10}\right) = P(z \le -0,5) = \Phi(-0,5) = 0,3085$$

Công thức tính xác suất của pp chuẩn:

Nếu $X \sim N(\mu;\sigma^2)$ thì μ: là trung bình, σ: độ lệch chuẩn, σ^2: phương sai

$$1).\ P(X \le a) = \Phi\left(\frac{a-\mu}{\sigma}\right)$$

$$2).\ P(X \ge b) = 1 - \Phi\left(\frac{b-\mu}{\sigma}\right)$$

$$3).\ P(a \le X \le b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

$$4).\ P(|X-\mu| \le a) = 2\Phi\left(\frac{a}{\sigma}\right) - 1$$

Chú ý.

1) *Đối với phân phối chuẩn thì xác suất khi X nhận một giá trị cụ thể là bằng 0, tức là* $P(X = k) = 0$. *Do đó các công thức trên vẫn đúng nếu các dấu bất đẳng thức là:* $<;>$.

2) Đối với bảng **Dương** thì: $\Phi(z) \ge 0,5$.

Đối với bảng **Âm** thì: $\Phi(z) \le 0,5$

Ví dụ 43. Cho X là BNN có phân phối chuẩn N(6; 2). Tính P(X < 7,5) và P(X > 5,2).
Giải:

…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
………………………………………………………………………………………………
…………………………………………………………………………………………………

Ví dụ 44. Trọng lượng X của một sản phẩm có phân phối chuẩn với trung bình 500gam và phương sai 30gam^2. Chọn ngẫu nhiên một sản phẩm, tính xác suất để chọn được sản phẩm có trọng lượng lớn hơn 510 gam.
Giải:
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
………………………………………………………………………………………………
…………………………………………………………………………………………………

Ví dụ 45. Trọng lượng của một loại vật nuôi trong nông trại là một đại lượng ngẫu nhiên có phân phối chuẩn với trung bình 50 kg và độ lệch chuẩn là 7 kg . Quy định rằng những con vật có trọng lượng từ 48 kg trở lên là những con đạt chuẩn. Hãy tính tỉ lệ những con vật đạt chuẩn của nông trại đó.
Giải:
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
………………………………………………………………………………………………
…………………………………………………………………………………………………

Ví dụ 46. Chiều cao của nam giới đã trưởng thành là biến ngẫu nhiên X(cm) có phân phối chuẩn N(163; 25). Chọn ngẫu nhiên 1 nam giới đã trưởng thành, tính xác suất người này cao từ 1,62m đến 1,7m.
Giải:
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
………………………………………………………………………………………………

Ví dụ 47. Cho $X \sim N\left(10; 0,6^2\right)$. Biết $P\left(X < b\right) = 0,0179$. Hãy tính b.
Giải:
…………………………………………………………………………………………………
…………………………………………………………………………………………………

……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………

Ví dụ 48. Cho $X \sim N(150;49)$. Biết $P(X > a) = 0,32$. Hãy tính a.
Giải:
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………

Ví dụ 49. Trọng lượng một loại sản phẩm là biến ngẫu nhiên có phân phối chuẩn với độ lệch chuẩn 50 gam. Chọn ngẫu nhiên 1 sản phẩm, tính xác suất để sản phẩm này có trọng lượng lệch không quá 25 gam so với trọng lượng trung bình.
Giải:
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………

Ví dụ 50. Tuổi thọ của một loại bóng đèn có phân phối chuẩn với trung bình là 10 ngàn giờ và độ lệch tiêu chuẩn 900 giờ. Nhà sản xuất đưa ra thời hạn bảo hành cho bóng đèn là 9 ngàn giờ. Tính tỷ lệ bóng đèn cần bảo hành.
Giải:
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………

Ví dụ 51. Thời gian X (tháng) từ lúc vay đến lúc trả tiền của một khách hàng tại ngân hàng A là biến ngẫu nhiên có phân phối chuẩn N(10; 9). Chậm nhất là bao lâu để 95% khách hàng trả tiền cho ngân hàng.

Giải:

………………………………………………………………………………………

………………………………………………………………………………………

………………………………………………………………………………………

………………………………………………………………………………………

………………………………………………………………………………………

………………………………………………………………………………………

………………………………………………………………………………………

………………………………………………………………………………………

………………………………………………………………………………………

………………………………………………………………………………………

Ví dụ 52. Khi thâm nhập một thị trường mới doanh nghiệp chỉ dự kiến doanh số hàng tháng là biến ngẫu nhiên có phân phối chuẩn với trung bình là 50 tỷ/tháng. Khả năng đạt doanh số trên 60 tỷ là 0,3. Tính độ lệch chuẩn của doanh số.

Giải:

………………………………………………………………………………………

………………………………………………………………………………………

………………………………………………………………………………………

………………………………………………………………………………………

………………………………………………………………………………………

………………………………………………………………………………………

…………………………………………………………………………….

………………………………………………………………………………………

………………………………………………………………………………………

……………………………………………………………………………….

……………………………………………………………………………………

Ví dụ 53. Điểm Toeic của sinh viên sắp tốt nghiệp ở trường đại học có phân phối chuẩn với điểm trung bình 550 và độ lệch chuẩn 60 điểm. Giả sử nhà trường muốn xác định điểm Toeic tối thiểu để sinh viên có thể ra trường với tỉ lệ 90%. Tính điểm Toeic tối thiểu (lấy phần nguyên).

Giải:

………………………………………………………………………………………

………………………………………………………………………………………

………………………………………………………………………………………

………………………………………………………………………………………

………………………………………………………………………………………

………………………………………………………………………………………

…………………………………………………………………………….

………………………………………………………………………………………

………………………………………………………………………………………

……………………………………………………………………………….

……………………………………………………………………………………

Ví dụ 54. Tuổi thọ của máy cắt cỏ là biến ngẫu nhiên có phân phối chuẩn với trung bình là 6 năm. Nhà sản xuất bảo hành sản phẩm khi bán ra là 24 tháng. Giả sử 3% sản phẩm bị trả lại (lỗi) trong thời gian bảo hành. Tính độ lệch chuẩn của tuổi thọ sản phẩm này.

Giải:

……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
………………………………………………………………………………….
……………………………………………………………………………………………
……………………………………………………………………………………………
…………………………………………………………………………………….
…………………………………………………………………………………………

SỬ DỤNG MÁY TÍNH BỎ TÚI fx- 570VN PLUS TÍNH XÁC SUẤT CỦA BNN

1. Máy tính fx- 570VN PLUS tính giá trị phân phối chuẩn

$X \sim N\left(\mu;\sigma^2\right)$**. Tìm b sao cho:** $P\left(X < b\right) = p$**.**

Ví dụ: Cho $X \sim N\left(\mu = 10; \sigma^2 = 0,5^2\right)$**.** Tìm b sao cho: $P\left(X < b\right) = 0,0179$.

- Bấm 5 phím: **MODE** ▼ **3(DIST** **3(Inverse Normal)**
- Bấm số:

0.0179 =

0.5 =

10 =

Kết quả được: **b = 8,95**

6. Máy Casio fx 580VNx

Cho $X \sim N\left(\mu = 10; \sigma^2 = 0,5^2\right)$**.** Tìm b sao cho: $P\left(X < b\right) = 0,0179$.

Bấm: Menu - 7 – 3

Vùng: 0,0179

σ : 0,5

μ : 10

4.2 Xấp xỉ phân phối nhị thức bằng phân phối chuẩn

Cho $X \sim B(n;\ p)$ thỏa mãn điều kiện n đủ lớn và p không quá gần 0 và 1 thì xấp xỉ

$X \sim N(\mu = np; \sigma^2 = np(1-p))$,

1) $P(X \le a) \approx P(X \le a + 0,5)$,
2) $P(X < a) \approx P(X < a - 0,5)$,
3) $P(X \ge a) \approx P(X \ge a - 0,5)$,
4) $P(X > a) \approx P(X > a + 0,5)$,
5) $P(X = a) \approx P(a - 0,5 \le X \le a + 0,5)$,
6) $P(a \le X \le b) \approx P(a - 0,5 \le X \le b + 0,5)$.

Ví dụ 55. Một cuộc khảo sát cho thấy 60% số dân ở một địa phương là nam giới. Chọn ngẫu nhiên 300 người thuộc địa phương này.Tính xác suất để có từ 200 đến 230 nam giới .

Giải:

………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
…………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………

Ví dụ 56. Tỷ lệ phế phẩm do một máy tự động sản xuất ra là 20%. Lấy 1000 sản phẩm do máy sản xuất để kiểm tra. Tính xác suất có 400 phế phẩm.

Giải.

………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………

Ví dụ 57. Tỉ lệ sản phẩm loại A của một nhà máy là 75%. Người ta chọn ngẫu nhiên 200 sản phẩm (trong số rất nhiều sản phẩm) để kiểm tra. Tính xác suất để 200 sản phẩm chọn ra có nhiều nhất 150 sản phẩm loại A.

Giải.

………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………

BÀI TẬP CHƯƠNG 2

Bài 1. Người thợ chép tranh mỗi tuần chép hai bức tranh độc lập A và B với xác suất hỏng tương ứng là 0,03 và 0,05. Biết rằng nếu thành công thì người thợ sẽ kiếm lời từ bức tranh A là 1,3 triệu đồng và B là 0,9 triệu đồng, nhưng nếu hỏng thì bị lỗ do bức tranh A là 0,8 triệu đồng và do B là 0,6 triệu đồng. Hỏi trung bình người thợ kiếm được bao nhiêu tiền chép tranh mỗi tuần?

Bài 2 Nhà máy X sản xuất một loại linh kiện điện tử với tỷ lệ sản phẩm tốt là 80%. Sản phẩm được đóng thành từng kiện, mỗi kiện 4 sản phẩm. Kiện thuộc loại A nếu trong đó số tốt nhiều hơn xấu. Khách hàng mua một kiện hàng.Tính xác suất khách hàng đó mua được kiện loại A.

A. 0,82 B. 0,41 C. 0,72 D. 0,48

Bài 3: Một nhân viên bán hàng mỗi ngày đi chào hàng ở 10 nơi với xác suất bán được hàng ở mỗi nơi là 0,2.Tìm xác suất để người đó bán được hàng ở 2 nơi.

A. 0,302 B. 0,484 C. 0,185 D. 0,458

Bài 4: Ba sinh viên làm bài thi, xác suất đạt yêu cầu của mỗi em lần lượt là 0,8; 0,7; 0,6. Gọi X là số Sv đạt yêu cầu. Tính trung bình, phương sai của X.

A. E(X)= 2,1 ; Var(X)= 0,76 B. E(X)= 2,1 ; Var(X)= 0,61

C. E(X)= 1,8 ; Var(X)= 0,73 D. E(X)= 1,8 ; Var(X)= 0,69

Bài 5: Bắn 2 phát vào bia. Bia gồm có 2 phần A, B không giao nhau với xác suất trúng lần lượt là: 0,3; 0,7. Nếu trúng A được 10 điểm, trúng B được 5 điểm. Gọi X là tổng số điểm đạt được sau 2 lần bắn. Lập bảng phân phối xác suất của X.

A.

X	5	10	15	20
P	0,49	0,21	0,21	0,09

B.

X	5	10	20
P	0,49	0,42	0,09

C.

X	10	15	20
P	0,49	0,21	0,09

D.

X	10	15	20
P	0,49	0,42	0,09

Bài 6. Trọng lượng của 1 bao gạo là một ĐLNN theo luật phân phối chuẩn với trung bình là 50kg và phương sai là $0,25kg^2$. Quy định những bao gạo có trọng lượng từ 49,5kg trở lên được gọi là loại 1. Tính tỷ lệ bao loại 1

A. 82,5% B. 84,13% C. 56% D. 48%

Bài 7.: Đề thi trắc nghiệm có 10 câu hỏi , mỗi câu có 5 đáp án trả lời, trong đó chỉ có một đáp án trả lời đúng.Một thí sinh chọn trả lời một cách hú hoạ. Tìm xác suất để người đó trả lời đúng ít nhất 8 câu.

A. 0,000074 B. 0,0000078 C. 0,000078 D. 0,0000074

Bài 8: Bắn 5 viên đạn vào mục tiêu. Xác suất trúng đích của mỗi lần bắn như nhau và bằng 0,2. Mục tiêu bị phá huỷ khi có ít nhất 3 viên trúng mục tiêu. Tìm xác suất để mục tiêu bị phá hủy.

A. 0,0512 B. 0,0064 C. 0,0579 D. 0,0067

Bài 9: Một nghiên cứu cho thấy 70% công chức cho rằng việc nghỉ làm 2 ngày 1 tuần sẽ nâng cao được hiệu suất công tác. Nếu chọn ngẫu nhiên 15 công chức ở một bộ để phỏng vấn thì xác suất để có ít nhất 10 người đồng ý với ý kiến trên là bao nhiêu.
A. 0,72 B. 0,206 C. 0,515 D. 0,869

Bài 10: Xác suất để khỏi bệnh khi dùng loại thuốc A là 3/4. Có 12 người mắc bệnh dùng thuốc A. Tìm xác suất để có 3 người khỏi bệnh.
A. 0,00035 B. 0,999 C. 0,00039 D. 0,0028

Bài 11: Một gia đình có 10 người con. Giả sử xác suất sinh con trai và con gái là như nhau. Tính xác suất để trong 10 người con đó có số trai ít hơn số gái.
A. 0,205 B. 0,828 C. 0,623 D. 0,377

Bài 12: Bưu điện dùng 1 máy tự động đọc địa chỉ trên bì thư để phân loại theo từng khu vực gởi đi. Trung bình khả năng đọc sai 1 địa chỉ trên bì thư trong 1 phút là 2. Tính xác suất để trong 1 phút máy đọc sai 3 bì thư.
A. 0,86 B. 0,18 C. 0,14 D. 0,27

Bài 13: Có 2 lô hàng, mỗi lô có nhiều sản phẩm. Tỷ lệ sp loại I của từng lô tương ứng là: 0,8; 0,7. Chọn mỗi lô ra 10 sp để kiểm tra. Nếu trong 10 sp đó có từ 8 sp loại 1 trở lên thì mua lô hàng đó. Tìm xác suất để chỉ có 1 lô hàng được mua.
A. 0,66 B. 0,38 C. 0,54 D. 0,67

Bài 14: Trong số 20 công nhân của một công ty có 12 người có tay nghề khá. Tìm xác suất để kiểm tra ngẫu nhiên tay nghề của 5 công nhân thì có 3 người có tay nghề khá.
A. 0,6 B. 0,3 C. 0,5 D. 0,4

Bài 15: Trong 20 giấy thông báo thuế thu nhập có 3 giấy mắc sai sót. Nhân viên kiểm tra ngẫu nhiên 6 giấy thông báo. Tính xác suất trong 6 giấy kiểm tra không có giấy nào bị sai sót.
A. 0,68 B. 0,32 C. 0,54 D. 0,42

Bài 16: Một xe tải vận chuyển 1000 chai rượu vào kho. Xác suất để khi vận chuyển 1 chai bị vỡ là 0,004. Tìm xác suất để sau khi vận chuyển có 5 chai bị vỡ.
A. 0,79 B. 0,16 C. 0,37 D. 0,75

Bài 17: 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 chỉ bị hạ nếu có ít nhất 2 viên trúng. Tính xác suất để máy bay bị hạ.
A. 0,95 B. 0,12 C. 0,08 D. 0,88

Bài 18: Trọng lượng sản phẩm X do 1 máy tự động sản xuất là biến ngẫu nhiên theo qui luật chuẩn với trung bình 100 gam và độ lệch chuẩn là 1 gam. Sản phẩm được coi là đạt tiêu chuẩn kỷ thuật nếu trọng lượng của nó đạt từ 98 đến 102 gam. Tìm tỷ lệ sp đạt tiêu chuẩn kỷ thuật của máy.
A. 97% B. 23% C. 95,54% D. 88%

Bài 19: Trong hệ thống tỷ giá hối đoái thả nổi, sự biến động của tỷ giá hối đoái chịu sự tác động của rất nhiều nhân tố và có thể xem như biến ngẫu nhiên có phân phối chuẩn. Giả sử ở một giai đọan nào đó tỷ giá của USD với VND có trung bình là 15000 đ và độ lệch chuẩn là 500đ. Tìm xác suất để trong 1 ngày nào đó. Tỷ giá sẽ cao hơn 16000đ.
A. 0,97 B. 0,013 C. 0,023 D. 0,82

Bài 20: Trọng lượng một loại trái cây có qui luật phân phối chuẩn với trọng lượng trung bình là 250g, đô lệch chuẩn là 5g. Một người lấy 1 trái từ trong sọt trái cây ra. Tính xác suất người này lấy được trái có trọng lượng hơn 260g.
A. 0,097 B. 0,023 C. 0,0228 D. 0,088

Bài 21: Nghiên cứu chiều cao của những người trưởng thành, người ta nhận thấy rằng chiều cao đó tuân theo qui luật phân phối chuẩn với trung bình là 175cm và độ lệch chuẩn 4cm. Hãy xác định giá trị h nếu biết rằng 30% người trưởng thành có tầm vóc dưới mức h.

A. 177 B. 179,5   C.166,61 D. 172,92

Bài 22. Một công ty bảo hiểm bán thẻ bảo hiểm với giá 100 nghìn đồng/1 người/1 năm. Nếu người tham gia bảo hiểm gặp rủi ro trong năm đó thì nhận được số tiền bồi thường là 1 triệu đồng. Theo thống kê biết rằng tỷ lệ người tham gia bảo hiểm bị rủi ro trong năm là 0,05, hãy tính tiền lãi trung bình khi bán mỗi thẻ bảo hiểm.

A. 45.000

Bài 23. Cho phân phối xác suất của số máy hỏng X trong một ca làm việc trong bảng

X	0	1	2
P	0,9	0,09	0,01

a. Tìm số máy hỏng trung bình trong một ca làm việc

b. Mỗi máy hỏng phải sửa hết 2 triệu đồng, tính tiền sửa máy trung bình trong một ca làm việc.

Bài 24. Cho X và Y tương ứng là các biến ngẫu nhiên độc lập chỉ lợi nhuận (tính theo %) hàng năm khi đầu tư vào hai ngành A và B nào đó. Giả sử E(X) = 12, Var(X) = 25, E(Y) = 14, Var(Y) = 36. Một người đầu tư vào cả hai ngành A và B thì cần lựa chọn tỷ lệ đầu tư như thế nào để ít rủi ro nhất.

Bài 25. Một người hằng ngày từ nhà đến cơ quan phải qua 4 ngã tư. Xác suất gặp đèn đỏ ở mỗi ngã tư là 25%. Lập bảng phân phối xác suất số lần gặp đèn đỏ của người đó.

Bài 26. Một người có thể lựa chọn giữa hai vị trí làm việc. Vị trí thứ nhất là tại một văn phòng và nhận một mức lương tháng cố định là 6 triệu đồng. Vị trí thứ hai là tại một đơn vị kinh doanh và nhận lương tháng theo số hợp đồng ký được. Mỗi hợp đồng ký được sẽ được nhận 5 triệu đồng. Biết rằng, số hợp đồng ký được trong 1 tháng có thể là 0, 1, 2 hoặc 3 hợp đồng với khả năng tương ứng là 10%, 30%, 40% và 20%. Làm thế nào để có thể so sánh, đánh giá về mức lương trong hai vị trí trên để từ đó đưa ra lựa chọn?

Bài 27: Trọng lượng các sản phẩm là 1 đại lượng ngẫu nhiên có phân phối chuẩn với phương sai $100g^2$. Tính tỷ lệ sản phẩm lệch không quá 20 so với trung bình.

A. 95,44%   B. %   C.7%   D. %

Bài 28. Thời gian X (tính bằng phút) của một khách hàng chờ để được phục vụ tại 1 cửa hàng là biến ngẫn nhiên có phân phối chuẩn N(4,5; 1,21). Thời gian t phải chờ tối thiểu là bao nhiêu để có không quá 7% số khách phải chờ có thời gian vượt quá t.

Bài 29. Giả sử X là biến ngẫn nhiên có phân phối chuẩn với trung bình 2100 và độ lệch chuẩn 200. Tìm b để $P(1800 < X < b) = 0{,}85$.

Bài 30. Để khảo sát trọng lượng trung bình của một loại trái cây của một vườn cây, người ta chọn ngẫu nhiên 500 quả từ vườn cây thì thấy có 150 quả nặng hơn 800 gam và 200 quả có trọng lượng nhỏ hơn 600 gam. Biết trọng lượng trái cây của vườn có phân phối chuẩn, tính trong lượng trung bình và độ lệch chuẩn của loại trái cây trên.

Bài 31. Lãi suất X (%) của 1 doanh nghiệp đầu tư vào một dự án trong năm 2011 theo luật phân phối chuẩn. Theo dự báo của công ty đầu tư thì lãi suất lớn hơn 20% có xác suất là 0,1585 và lãi suất lớn hơn 25% có xác suất là 0,0225. Hãy xác định kỳ vọng và phương sai của lãi suất X.

Bài 32. Vận chuyển 1000 chai rượu vào kho, xác suất 1 chai bị vỡ sau khi vận chuyển là 0,04. Tính xác suất sau khi vận chuyển có số chai vỡ từ 35 đến 50 chai

Bài 33. Thời gian mang thai của sản phụ là biến ngẫu nhiên có phân phối chuẩn với trung bình 280 ngày. Cho biết tỷ lệ một sản phụ mang thai trên 290 ngày là 25,14%, tính độ lệch chuẩn của thời gian mang thai.

Bài 34. Chiều dài của loại linh kiện điện tử A tại cửa hàng B là biến ngẫu nhiên X (mm) có phân phối chuẩn N(12; 2, 5). Một công ty cần mua loại linh kiện này với chiều dài từ 11,98mm đến 13mm và họ chọn lần lượt 7 chiếc từ cửa hàng B. Tính xác suất để trong 7 chiếc được chọn có 5 chiếc sử dụng được.

Chương 3. THỐNG KÊ VÀ ƯỚC LƯỢNG

MỤC TIÊU CHƯƠNG

Nội dung chương này giúp người học có khả năng:

• Hiểu biết về thống kê.

• Tính toán được các bài toán ước lượng trung bình và tỷ lệ tổng thể.

3.1. TỔNG QUAN

3.1.1 THỐNG KÊ NGHĨA LÀ GÌ?

Thống kê (statistics)

Thống kê, cũng giống như nhiều lĩnh vực nghiên cứu, có ngôn ngữ riêng. Chúng ta thường xuyên thấy thống kê trong ngôn ngữ hàng ngày. Cách sử dụng phổ biến là thống kê đề cập đến các thông tin bằng số. Ví dụ như mức lương khởi điểm trung bình của sinh viên tốt nghiệp đại học, số người chết vì ung thư vào năm ngoái, và tỉ lệ thất nghiệp năm 2019. Trong các ví dụ này, số liệu thống kê là một giá trị hoặc tỷ lệ phần trăm. Một tập hợp các thông tin số được gọi là số liệu thống kê.

Thống kê là khoa học thu thập, tổ chức, trình bày, phân tích và giải thích dữ liệu để hỗ trợ đưa ra quyết định hiệu quả hơn.

Trước hết chúng ta tìm hiểu một số khái niệm cơ bản về thống kê.

Tổng thể (Population)

Tổng thể là tập hợp toàn bộ những người, đồ vật hoặc đối tượng ta quan tâm hay các phép đo thu được từ tất cả các cá nhân hoặc đối tượng quan tâm.

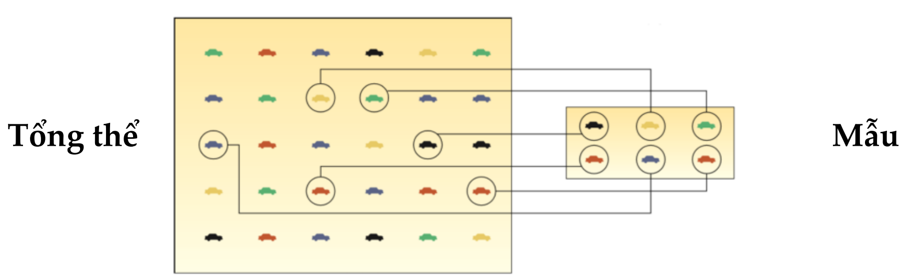
Tổng thể là một tập hợp các đối tượng, chẳng hạn như "tất cả công nhân đang làm việc cho Microsoft", hoặc " tất cả các xe máy Vision được sản xuất vào ngày 3 tháng 2 năm 20017, bởi Công ty Hoda tại nhà máy Vĩnh Phúc".

Mẫu (Sample)

Để suy ra điều gì đó về tổng thể, chúng ta thường lấy một mẫu từ tổng thể.

Mẫu là một phần, hoặc một bộ phận, của tổng thể quan tâm.

Ví dụ 1. Chúng ta muốn ước tính trung bình số km đi được cho 4,5 lít xăng (MPG) của xe Toyota Vios. Một mẫu 6 chiếc Vios được lựa chọn từ tổng thể để thử nghiệm. MPG trung bình của sáu xe được sử dụng để ước tính MPG cho toàn bộ xe Toyota Vios.



Tham số của tổng thể (Population Parameter)

Một thước đo mô tả về tổng thể được gọi là một *tham số*. Các tham số thường được biểu thị bằng các chữ cái Hy Lạp. Ví dụ về các tham số là *trung bình tổng thể* (μ), *phương sai tổng thể* (σ^2), *độ lệch chuẩn tổng thể* (σ) và tỷ lệ tổng thể (p).

Trị thống kê của mẫu (Sample Statistic)

Một thước đo mô tả của một mẫu được gọi là một *trị thống kê*. Các trị thống kê thường được biểu thị bằng chữ La Mã. Ví dụ về trị thống kê là *trung bình mẫu* ($\overline{x}$), *phương sai mẫu* (s^2), *độ lệch chuẩn mẫu* (s) và tỷ lệ mẫu (f).

3.1.3 CÁC LOẠI THỐNG KÊ

Thống kê thường được chia thành hai loại: thống kê mô tả và thống kê suy luận.

Thống kê mô tả (Descriptive Statistics)

Thống kê đề cập đến tổ chức, trình bày, vẽ biểu đồ, nhận xét. . . dữ liệu. Khía cạnh thống kê này thường được gọi là *thống kê mô tả*.

Thống kê mô tả là phương pháp tổ chức, tóm tắt và trình bày dữ liệu theo cách để làm rõ thông tin.

Chẳng hạn, dân số Việt Nam là 87.860.000 vào năm 2011; 89.759.000 vào năm 2013; và 91.713.000 vào năm 2015, thông tin này là thống kê mô tả. Tuy nhiên, nếu chúng ta sử dụng chúng để ước tính dân số Việt Nam trong năm 2025 hoặc tăng trưởng phần trăm từ năm 2015 đến 2020 thì đó không phải là thống kê mô tả. Vì những thống kê này không được sử dụng để tóm tắt các giá trị trong quá khứ mà để ước tính các giá trị trong tương lai.

Thống kê suy luận (Inferential Statistics)

Loại thống kê thứ hai là thống kê suy luận, còn được gọi là suy luận thống kê. Mối quan tâm chính liên quan đến thống kê suy luận là tìm ra điều gì đó về tổng thể từ một mẫu được lấy từ tổng thể đó. Ví dụ, một cuộc khảo sát cho thấy chỉ 46% học sinh trung học có thể giải quyết các vấn đề liên quan đến phân số, số thập phân và tỷ lệ phần trăm. Vì đây là những suy luận về tổng thể (tất cả học sinh trung học) dựa trên dữ liệu mẫu khảo sát, nên ta gọi là thống kê suy luận. Như vậy thống kê suy luận như một dự đoán tốt nhất về giá trị tổng thể dựa trên thông tin mẫu.

Thống kê suy luận là các phương pháp được sử dụng để ước tính một đặc tính của tổng thể trên cơ sở mẫu.

Nếu một nhà nghiên cứu thu thập dữ liệu từ một mẫu và sử dụng các số liệu đó để đạt được kết luận về tổng thể mà từ đó mẫu được lấy, thống kê đó là thống kê suy luận.

Nội dung của môn học tập trung giới thiệu về thống kê suy luận.

3.2. THỐNG KÊ MÔ TẢ

3.2.1 XÂY DỰNG BẢNG TẦN SỐ (Frequency Table)

Đầu tiên chúng ta tìm hiểu về bảng tần số.

Bảng tần số là một tập dữ liệu được chia thành các nhóm loại trừ lẫn nhau và có số lượng quan sát trong mỗi nhóm.

Tần số tương đối (Relative Frequency)

Tần số tương đối (tần suất) là tỷ lệ của tần số trong một nhóm và tổng tần số. Tức là để tính tần số tương đối của một nhóm ta lấy tần số của nhóm đó chia cho tổng số quan sát.

Tần số tích lũy (Cumulative Frequency)

Tần số tích lũy của nhóm đầu tiên bằng chính là tần số của nhóm này. Tần số tích lũy của nhóm thứ hai bằng tần số của nhóm đó cộng với tần số tích lũy của nhóm thứ nhất. Quá trình này tiếp tục cho đến nhóm cuối, khi đó tần số tích lũy của nhóm cuối đúng bằng số giá trị quan sát của dữ liệu.

Bảng tần số cho dữ liệu định lượng có ít biểu hiện (Bảng dạng điểm)

Ví dụ 2. Khảo sát 200 người đã đi làm về số tờ báo họ đọc trong một tuần. Người ta thu được kết quả:

Số tờ báo đọc	Tần số	Tần suất	Tần số tích lũy
0	44	0.22	44
1	24	0.12	68
2	18	0.09	86
3	16	0.08	102
4	20	0.10	122
5	22	0.11	144
6	26	0.13	170
7	30	0.15	**200**
Tổng số	**200**	**1**	

Bảng tần số cho dữ liệu định lượng có nhiều biểu hiện (Bảng dạng nhóm, khoảng)

Khi xây dựng bảng phân phối tần số cho dữ liệu định lượng có nhiều biểu hiện, nhà nghiên cứu thường xác định *phạm vi* (**range**) của dữ liệu. Phạm vi thường được định nghĩa là sự khác biệt giữa giá trị lớn nhất và nhỏ nhất.

Ví dụ 3. Xem xét lợi nhuận (triệu đồng) của 180 xe máy được bán tại đại lý của Hoda:

Lợi nhuận				**%**	**Tích lũy**	
Cận dưới Cận trên	**Điểm giữa**	**Độ rộng**	**Tần số**	**Tần suất**	**Tần số**	**Tần suất**
0.2 – 0.6	0.4	0.4	8	4	8	4
0.6 – 1.0	0.8	0.4	11	6	19	10
1.0 – 1.4	1.2	0.4	23	13	42	23
1.4 – 1.8	1.6	0.4	38	21	80	44
1.8 – 2.2	2.0	0.4	45	25	125	69
2.2 – 2.6	2.4	0.4	32	18	157	87
2.6 – 3.0	2.8	0.4	19	11	176	98
3.0 – 3.4	3.2	0.4	4	2	180	100
Tổng số			**180**	**100**		

3.3 CÁC THAM SỐ TỔNG THỂ VÀ TRỊ THỐNG KÊ MẪU

3.3.1 TRUNG BÌNH

TRUNG BÌNH TỔNG THỂ

Đo lường vị trí trung tâm người ta thường dùng giá trị trung bình của dữ liêu. Trung bình là thước đo vị trí hiển thị giá trị trung tâm của dữ liệu. Trung bình xuất hiện hàng ngày trên TV, trên các trang web khác nhau, trên báo và trong các tạp chí khác.

Đối với dữ liệu thô, nghĩa là dữ liệu chưa được nhóm trong phân phối tần số, *trung bình tổng thể* bằng tổng của tất cả các giá trị tổng thể chia cho số lượng giá trị trong tổng thể N.

TRUNG BÌNH TỔNG THỂ $\mu = \dfrac{\sum x}{N}$

Ví dụ 4. Khảo sát số công nhân của tất cả các tổ sản xuất trong nhà máy Z15 ta thu được số liệu sau:

11	4	10	4	9	3	8	10	3	14	8	10	3	5	6
2	2	5	6	4	2	2	3	7	4	3	7	8	10	7
7	4	7	5	2	2	5	5	6	3	3	2	2	9	8

Tại sao thông tin này là một tổng thể ? Tính trung bình số công nhân mỗi tổ?

Lời giải

Đây là một tổng thể vì chúng ta đang xem xét tất cả các tổ sản xuất có trong nhà máy Z15. Trong nhà máy có tất cả 45 tổ nên trung bình số công nhân mỗi tổ là:

$$\mu = \frac{\sum x}{N} = \frac{11+4+10+...+2+9+8}{45} = \frac{232}{45} = 5.2\,.$$

Giá trị của 5.2 là số tượng trưng cho số công nhân một tổ.

TRUNG BÌNH MẪU

Chúng ta thường chọn một mẫu từ tổng thể để ước tính một đặc tính cụ thể của tổng thể. *Giá trị trung bình mẫu* bằng tổng của tất cả các giá trị mẫu chia cho tổng số giá trị của mẫu.

TRUNG BÌNH MẪU $\bar{x} = \dfrac{\sum x}{n}$; $\bar{x} = \dfrac{\sum fM}{n}$.

Ví dụ 5. Mobifone đang nghiên cứu số phút được khách hàng sử dụng trong một gói cước điện thoại di động cụ thể. Một mẫu ngẫu nhiên gồm 30 khách hàng cho thấy số phút sau được sử dụng vào tháng trước như sau:

111	94	90	74	89	93	108	110	93	91	88	80	99	95	96
82	92	85	96	74	82	112	113	97	94	83	97	88	90	97

Tính trung bình số phút sử dụng của khách hàng?

Giải

Trung bình số phút sử dụng của khách hàng là:

$$\bar{x} = \frac{\sum x}{n} = \frac{111+94+90+...+88+90+97}{30} = \frac{2793}{30} = 93.1\,.$$

Số phút trung bình được sử dụng trong tháng trước của mẫu người dùng mạng điện thoại di động Mobifone là 93.1 phút.

Ví dụ 6. Xem xét báo cáo lợi nhuận (triệu đồng) của 180 xe máy được bán trong tháng trước tại bốn đại lý của Hoda tại Đà Nẵng. Như sau:

Lợi nhuận	**Tần số**
0.2 – 0.6	8
0.6 – 1.0	11
1.0 – 1.4	23
1.4 – 1.8	38
1.8 – 2.2	45
2.2 – 2.6	32
2.6 – 3.0	19
3.0 – 3.4	4
Tổng số	**180**

Xác định lợi nhuận trung bình trên mỗi chiếc xe.

Giải

Lợi nhuận	**Điểm giữa (*M*)**	**Tần số (*f*)**
0.2 – 0.6	0.4	8
0.6 – 1.0	0.8	11
1.0 – 1.4	1.2	23
1.4 – 1.8	1.6	38
1.8 – 2.2	2.0	45
2.2 – 2.6	2.4	32
2.6 – 3.0	2.8	19
3.0 – 3.4	3.2	4
Tổng số		**180**

Lợi nhuận trung bình trên mỗi chiếc xe là:

$$\overline{x} = \frac{\sum fM}{n} = \frac{0,4.8 + 0,8.11 + ... + 3,2.4}{180} = \frac{333,2}{180} = 1,85.$$

3.3.2 PHƯƠNG SAI VÀ ĐỘ LỆCH CHUẨN

Phương sai và *độ lệch chuẩn* mô tả độ lệch giữa các giá trị so với giá trị trung bình.

PHƯƠNG SAI là trung bình số học của độ lệch bình phương giữa các giá trị so với trung bình của chúng.

ĐỘ LỆCH CHUẨN là căn bậc hai của phương sai.

PHƯƠNG SAI TỔNG THỂ $\sigma^2 = \dfrac{\sum(x-\mu)^2}{N}$.

trong đó: σ là chữ Hy Lạp viết thường, được đọc là "sigma "

ĐỘ LỆCH CHUẨN TỔNG THỂ $\sigma = \sqrt{\dfrac{\sum(x-\mu)^2}{N}}$.

Ví dụ 7. Số lượng xe vi phạm giao thông trên đường cao tốc Long Thành năm 2018 theo tháng, được báo cáo dưới đây:

Tháng	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Số xe	19	17	22	18	28	34	45	39	38	44	34	10

Xác định phương sai, độ lệch chuẩn của số lượng xe vi phạm giao thông trên đường cao tốc Long Thành năm 2018 theo tháng.

Giải

Ta có số lượng xe vi phạm giao thông trên đường cao tốc Long Thành năm 2018 theo tháng là một tổng thể.

Trung bình tổng thể: $\mu = \frac{19+17+...+34+10}{12} = \frac{348}{12} = 29$ (xe)

Phương sai tổng thể: $\sigma^2 = \frac{(19-29)^2 + (17-29)^2 + ... + (34-29)^2 + (10-29)^2}{12} = \frac{1488}{12} = 124$ (xe)^2.

Độ lệch chuẩn tổng thể: $\sigma = \sqrt{24} = 4,9.$

Các công thức tính cho phương sai mẫu là.

PHƯƠNG SAI MẪU $\qquad s^2 = \frac{\sum(x-\bar{x})^2}{n-1}$.

ĐỘ LỆCH CHUẨN MẪU $\qquad s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$.

Việc sử dụng chính của phương sai mẫu và độ lệch chuẩn mẫu là ước tính cho phương sai tổng thể và độ lệch chuẩn tổng thể. Trong đó phương sai mẫu và độ lệch chuẩn mẫu sử dụng $n-1$ trong mẫu số thay vì *n* vì sử dụng *n* trong mẫu số của phương sai mẫu có xu hướng đánh giá thấp phương sai tổng thể. Việc sử dụng $n-1$ cho phép nó trở thành một công cụ ước lượng tốt hơn. Việc đánh giá về các tính chất của công cụ ước tính tốt nằm ngoài phạm vi của tài liệu này.

Ví dụ 8. Một nhà nghiên cứu khảo sát một mẫu các luật sư trong một thành phố về tổng số năm kết án mà mỗi luật sư giành được cho các bị cáo trong tháng qua, như sau:

Luật sư	**1**	**2**	**3**	**4**	**5**
Số năm	55	40	25	30	50

Tính phương sai và độ lệch chuẩn cho mẫu này.

Giải

Trung bình mẫu $\bar{x} = \frac{55+40+25+30+50}{5} = \frac{200}{5} = 40$.

Phương sai mẫu $s^2 = \frac{(55-40)^2 + (40-40)^2 + ... + (50-40)^2}{5-1} = \frac{650}{4} = 162,5$.

Độ lệch chuẩn mẫu $s = \sqrt{162,5} = 12,75$.

4. Dùng máy tính bỏ túi fx- 570ES, Plus tính kích, trung bình và độ lệch chuẩn mẫu:
Ví dụ:

X	2	4	5	7
F	12	14	21	13

1. Mở chế độ tính thống kê có 2 cột: ***Ấn lần lượt 5 phím***

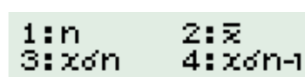
SHIFT **MODE** **RELAY(nút mũi tên xuống ↓)**

4(start) **1(on)**

Vào chế độ tính thống kê bấm lần lượt 3 phím sau:

MODE **3(Start)** **1(1-Var)**

2. Nhập số liệu:
<u>Nhập hàng thứ nhất X:</u>
Bấm: **2** =
4 =
5 =
7 =
Bấm nút: **Relay** (->) để qua cột thứ hai, rồi dời dấu nháy lên ứng với hàng bằng 2 để nhập tiếp dòng xác suất P.
<u>Nhập hàng thứ hai P:</u>
Bấm: **12** =
14 =
21 =
13 =
Nhập toàn bộ dữ liệu xong bấm phím : **AC**



a.Tính kích thước mẫu : ***Ấn lần lượt 5 phím***
Bấm: **SHIFT** **1** **4(Var)** **1**(n) =
Kết quả: n = 60

b.Tính trung bình mẫu : ***Ấn lần lượt 5 phím***
Bấm: **SHIFT** **1** **4(Var)** **2**$(\overline{X})$ =

Kết quả: $\overline{X}$ = 4,6

c.Tính độ lệch chuẩn mẫu**:** ***Ấn lần lượt 5 phím***
Bấm: **SHIFT** **1** **4(Var)** **4**(sx) =
Kết quả: $s = sx = x\sigma n - 1$ = 1,669
Máy 570 không có phương sai: nên phương sai là $s^2 = (1,669)^2 = 2,786$

3**.** Thoát khỏi chế độ tính thống kê ấn**:** **MODE** **1**

5. Máy fx 580 VN

X	2	4	5	7
F	12	14	21	13

Bấm: SHIFT - MENU – 3(Thống kê) – 1(Mở)

Bấm: MENU – 6 - 1(Tính Tkê 1 – biến) - Nhập dữ liệu- AC

Bấm: OPTN – 2(Tính 1-biến)

Kết quả:
Kích thước mẫu n = 60
Trung bình mẫu : $\overline{x}$ = 4,6
Độ lệch chuẩn mẫu**:** $s = sx$ = 1,669
Phương sai mẫu**:** $s_x^2 =$ = 2,786

Ví dụ 7. Tính kích thước mẫu, trung bình, độ lệch chuẩn và phương sai của dữ liệu mẫu sau:

Giá trị	Tần số
12	6
14	12
15	15
17	19
19	26
22	18
23	9
25	5

Giải
………………………………………………………………………………………………
……………………………………………………………………………………………

Ví dụ 8. Tính kích thước mẫu, trung bình, độ lệch chuẩn và phương sai của dữ liệu mẫu sau:

Nhóm	Tần số
10 – 15	6
15 – 20	22
20 – 25	35
25 – 30	29
30 – 35	16
35 – 40	8
40 – 45	4
45 - 50	2

Giải
………………………………………………………………………………………………
……………………………………………………………………………………………

Ví dụ 9. Tính kích thước mẫu, trung bình, độ lệch chuẩn và phương sai của dữ liệu mẫu sau:
70; 80; 65; 85; 72; 75; 68; 67; 82; 89; 79; 86; 90; 78

Giải
………………………………………………………………………………………………
………………………………………………………………………………………………

Ví dụ 10. Tính kích thước mẫu, trung bình, độ lệch chuẩn và phương sai của dữ liệu mẫu sau:

	Nhóm	Tần số
	11-14	10
	14 -17	30
	17 - 20	100
	20 - 23	90
	>23	40

Giải

………………………………………………………………………………………

3.3.3 ĐỊNH LÝ CHEBYSHEV

Chúng ta biết rằng độ lệch chuẩn nhỏ cho một tập hợp các giá trị chỉ ra rằng các giá trị này được tập trung gần với giá trị trung bình. Ngược lại, độ lệch chuẩn lớn cho thấy các quan sát được phân tán xa giá trị trung bình. Nhà toán học người Nga P. L. Chebyshev (1821 - 1894) đã phát triển một định lý cho phép chúng ta xác định tỷ lệ tối thiểu của các giá trị nằm trong một số độ lệch chuẩn tính từ giá trị trung bình.

ĐỊNH LÝ CHEBYSHEV Đối với bất kỳ tập hợp quan sát nào, có ít nhất là $1-\frac{1}{k^2}$ các giá trị nằm trong *k* độ lệch chuẩn tính từ giá trị trung bình, trong đó *k* là hằng số lớn hơn 1.
Cụ thể:

a. Với $k=3$: ít nhất $1-\frac{1}{k^2}=1-\frac{1}{3^2}=\frac{8}{9}\approx 89\%$ các giá trị nằm trong khoảng từ $\bar{x}-3s$ đến $\bar{x}+3s$.

b. Với $k=2$: ít nhất $1-\frac{1}{k^2}=1-\frac{1}{2^2}=\frac{3}{4}=75\%$ các giá trị nằm trong khoảng từ $\bar{x}-2s$ đến $\bar{x}+2s$.

3.4 CHỌN MẪU

3.4.1 Phương pháp chọn mẫu

Khi chọn phần tử để điều tra mẫu, người ta có thể chọn theo những quy tắc khác nhau, ngẫu nhiên hoặc không ngẫu nhiên. Có hai phương pháp chọn mẫu cơ bản được sử dụng phổ biến trong các cuộc điều tra là chọn mẫu ngẫu nhiên và chọn mẫu phi ngẫu nhiên.

• **Chọn ngẫu nhiên**: là phương pháp chọn hoàn toàn ngẫu nhiên không phụ thuộc vào ý muốn chủ quan của con người.

• **Chọn mẫu phi ngẫu nhiên:** là phương pháp chọn đơn vị phụ thuộc vào ý muốn chủ quan của người chọn.

Trong phạm vi và nội dung chương trình chỉ đề cập đến các vấn đề thuộc chọn mẫu ngẫu nhiên.

3.4.2 Phương pháp chọn mẫu ngẫu nhiên đơn giản

• **Khái niệm**: Là phương pháp tổ chức chọn mẫu một cách hoàn toàn ngẫu nhiên không qua một sự sắp xếp nào và có thể dùng phương pháp chọn một lần hoặc chọn nhiều lần.

° **Chọn hoàn lại (chọn lặp, chọn nhiều lần):** Từ tổng thể ta rút ngẫu nhiên một phần tử và ghi lại các đặc trưng cần quan tâm, sau đó trả lại phần tử đó về tổng thể và làm tương tự ở các lần tiếp theo cho tới khi ta được một mẫu cỡ n. Xác suất được chọn của mỗi phần tử là như nhau (đều bằng 1/N). Khi đó, số mẫu có thể hình thành là: $k = N^n$.

° **Chọn không hoàn lại (chọn một lần):** Làm tương tự như trên, chỉ khác là sau mỗi lần rút các phần tử ta loại phần tử đó ra khỏi tổng thể. Xác suất được chọn của các phần tử là hoàn toàn khác nhau, xác suất này tăng dần trong quá trình chọn. Khi đó, số mẫu có thể hình thành: $k = C_N^k$.

3.4.2 Sai số chọn mẫu (Sampling error)

Sai số chọn mẫu là sự khác biệt giữa một thống kê mẫu và tham số tổng thể tương ứng của nó, tức là: $\theta = \theta' \pm \varepsilon_M \pm \varepsilon_{0M}$.

Trong đó:

θ : Tham số tổng thể.

θ' : Thống kê của mẫu.

ε_M : Sai số chọn mẫu (sai số do tính đại diện)

ε_{0M} : Sai số phi chọn mẫu (sai số do ghi chép)

Các loại sai số trong điều tra chọn mẫu

Trong một cuộc điều tra chọn mẫu, thường xuất hiện hai loại sai số chủ yếu sau:

• **Sai số phi chọn mẫu:** Sai số này xảy ra ở tất cả các cuộc điều tra. Nguyên nhân là do điều tra viên cân đong đo đếm sai, ghi chép sai, đơn vị điều tra cung cấp sai sự thật. Sai số này không thể xoá bỏ được mà chỉ giảm bằng cách huấn luyện nhân viên điều tra cẩn thận hơn.

• **Sai số chọn mẫu:** Sai số này chỉ xảy ra trong điều tra chọn mẫu, do chỉ điều tra một số ít đơn vị nhưng kết quả lại ước lượng cho cả tổng thể. Sai số chọn mẫu được biểu hiện dưới hai hình thức:

o **Sai số hệ thống**: xảy ra do vi phạm nguyên tắc chọn, chọn một số đơn vị không đủ lớn để đảm bảo tính chất đại biểu, chọn mẫu không khách quan.

o **Sai số ngẫu nhiên**: chỉ xuất hiện trong trường hợp các phần tử của tổng thể được chọn theo nguyên tắc ngẫu nhiên, không lường trước được lệch về hướng nào, nhiều hơn hay ít hơn so với thực tế. Sai số này được giảm dần khi điều tra một số đủ lớn các phần tử.

Sai số chọn mẫu không phải là một trị số cố định. Với cùng một hiện tượng nhưng nếu tiến hành điều tra nhiều lần với các cách chọn mẫu khác nhau, kết cấu của mẫu khác nhau thì sẽ có các sai số chọn mẫu khác nhau. Do đó, để tính sai số nhằm đánh giá mức độ chính xác của ước lượng thì phải tính sai số bình quân chọn mẫu.

3.4.3 Sai số bình quân chọn mẫu

• Khái niệm

Sai số bình quân chọn mẫu là một số đại diện cho các giá trị của sai số chọn mẫu, nói cách khác, đó là bình quân của tất cả các sai số chọn mẫu do việc lựa chọn mẫu có kết cấu thay đổi. Trong thống kê, cùng một hiện tượng nghiên cứu nếu áp dụng các phương pháp tổ chức chọn mẫu khác nhau sẽ có sai số bình quân chọn mẫu khác nhau. Thống kê toán đã xác định công thức tính sai số bình quân chọn mẫu từ một tổng thể hữu hạn trong các trường hợp như sau:

Suy rộng \ Cách chọn	Chọn hoàn lại	Chọn không hoàn lại
Trung bình	$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}}$ hoặc $\sigma_{\bar{x}} = \sqrt{\frac{s^2}{n-1}}$	$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)}$ hoặc $\sigma_{\bar{x}} = \sqrt{\frac{s^2}{n-1}\left(\frac{N-n}{N-1}\right)}$
Tỷ lệ	$\sigma_p = \sqrt{\frac{p(1-p)}{n}}$ hoặc $\sigma_p = \sqrt{\frac{f(1-f)}{n-1}}$	$\sigma_p = \sqrt{\frac{p(1-p)}{n}\left(\frac{N-n}{N-1}\right)}$ hoặc $\sigma_p = \sqrt{\frac{f(1-f)}{n-1}\left(\frac{N-n}{N-1}\right)}$

Trong đó:

N : Số phần tử tổng thể. n : Số phần tử mẫu.

$\sigma_{\bar{x}}$: Sai số trung bình chọn mẫu khi ước lượng số trung bình.

σ_p : Sai số trung bình chọn mẫu khi ước lượng tỷ lệ.

σ^2 : Phương sai tổng thể. s^2 : Phương sai mẫu.

p : Tỷ lệ một đặc tính nào đó của tổng thể. f : Tỷ lệ một đặc tính nào đó của mẫu.

Từ cách tính sai số bình quân chọn mẫu ở trên, có thể thấy có rất nhiều nhân tố ảnh hưởng đến độ lớn của sai số chọn mẫu. Trong đó, ba nhân tố chủ yếu nhất, đó là:

o Số phần tử mẫu n: Khi số phần tử điều tra tăng lên, tổng thể mẫu sẽ gần với tổng thể chung, sai số chọn mẫu sẽ giảm. Khi n → N thì $|\mu - \bar{x}| \to 0, |p - f| \to 0.$.

o Phương pháp tổ chức chọn mẫu: Các phương pháp chọn mẫu khác nhau, tính đại diện của mẫu chọn ra khác nhau cũng sẽ dẫn đến những sai số chọn mẫu khác nhau.

o Độ đồng đều của tổng thể chung: Nếu tổng thể có độ đồng đều cao tức phương sai tổng thể σ^2 tương đối nhỏ thì sai số chọn mẫu sẽ nhỏ.

3.5 PHÂN PHỐI MẪU

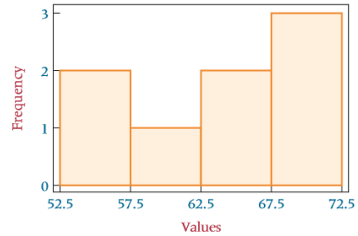
3.5.1 Phân phối trung bình mẫu (Sampling Distribution of the Sample Mean)

Phân phối trung bình mẫu là phân phối xác suất của tất cả giá trị trung bình của các mẫu có cùng kích thước và được thu thập từ một tổng thể.

Ví dụ 11. Giả sử một dữ liệu chỉ bao gồm N = 8 số:

54 55 59 63 64 68 69 70

Sử dụng biểu đồ, chúng ta có thể thấy hình dạng phân phối của tổng thể dữ liệu này.



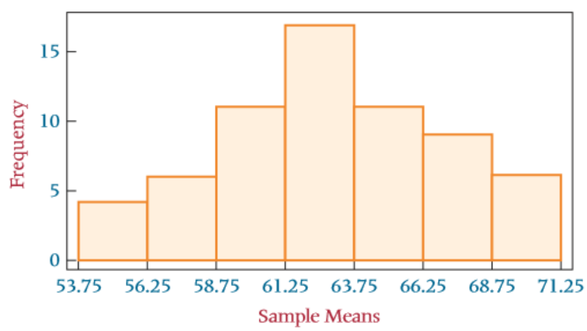
Giả sử chúng ta chọn (có hoàn lại) tất cả các mẫu có kích thước n = 2 từ tổng thể này. Kết quả là có $8^2 = 64$ mẫu sau đây:

(54,54)	(55,54)	(59,54)	(63,54)	(64,54)	(68,54)	(69,54)	(70,54)
(54,55)	(55,55)	(59,55)	(63,55)	(64,55)	(68,55)	(69,55)	(70,55)
(54,59)	(55,59)	(59,59)	(63,59)	(64,59)	(68,59)	(69,59)	(70,59)
(54,63)	(55,63)	(59,63)	(63,63)	(64,63)	(68,63)	(69,63)	(70,63)
(54,64)	(55,64)	(59,64)	(63,64)	(64,64)	(68,64)	(69,64)	(70,64)
(54,68)	(55,68)	(59,68)	(63,68)	(64,68)	(68,68)	(69,68)	(70,68)
(54,69)	(55,69)	(59,69)	(63,69)	(64,69)	(68,69)	(69,69)	(70,69)
(54,70)	(55,70)	(59,70)	(63,70)	(64,70)	(68,70)	(69,70)	(70,70)

Các trung bình của từng mẫu này như sau.

54 54.5 56.5 58.5 59 61 61.5 62 54.5 55 57 59 59.5 61.5 62 62.5 56.5 57 59 61 61.5 63.5 64 64.5 58.5 59 61 63 63.5 65.5 66 66.5 59 59.5 61.5 63.5 64 66 66.5 67 60 61.5 63.5 65.5 66 68 68.5 69 61.5 62 64 66 66.5 68.5 69 69.5 62 62.5 64.5 66.5 67 69 69.5 70.

Phân phối của các trung bình mẫu này là:



Tóm lại, các mối quan hệ quan trọng giữa phân phối tổng thể và phân phối trung bình mẫu:

1. Giá trị trung bình của trung bình các mẫu bằng trung bình tổng thể.

2. Sự phân tán của phân phối trung bình mẫu hẹp hơn phân phối tổng thể.

3. Phân phối trung bình mẫu có xu hướng trở thành hình chuông và gần với phân phối chuẩn.

3.5.2 Định lý giới hạn trung tâm

Trong phần này, chúng ta xem xét định lý giới hạn trung tâm ứng dụng vào phân phối trung bình mẫu, cho phép chúng ta sử dụng phân phối xác suất bình thường để tạo khoảng tin cậy cho trung bình tổng thể và thực hiện các kiểm định về giả thuyết .

Định lý giới hạn trung tâm

Nếu các mẫu có kích thước n được lấy ngẫu nhiên từ một tổng thể thể có giá trị trung bình μ và độ lệch chuẩn σ, thì trung bình mẫu $\bar{x}$, có phân phối chuẩn cho các cỡ mẫu lớn (n ≥30) bất kể hình dạng phân phối tổng thể. Nếu tổng thể có phân phối chuẩn thì trung bình cũng có phân phối chuẩn cho bất kỳ mẫu kích thước nào.Giá trị trung bình của các trung bình mẫu là trung bình tổng thể: $\mu_{\bar{x}} = \mu$ *,và độ lệch chuẩn của trung bình mẫu (sai số chuẩn của trung bình) là:* $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$.

3.5.3 Các dạng phân phối trung bình mẫu

3.5.1 Chọn mẫu từ tổng thể hữu hạn

Trong trường hợp tổng thể hữu hạn, có thể thực hiện điều chỉnh thống kê cho công thức z của trung bình mẫu. Điều chỉnh được gọi là hệ số hiệu chỉnh: $\sqrt{\dfrac{N-n}{N-1}}$. Công thức z cho trung bình mẫu khi các mẫu được rút ra từ các tổng thể hữu hạn: $z = \dfrac{\bar{x}-\mu}{\dfrac{\sigma}{\sqrt{n}}\sqrt{\dfrac{N-n}{N-1}}} \sim N(0;1)$.

3.5.2 Chọn mẫu từ tổng thể vô hạn (hoặc rất lớn)

Trong trường hợp tổng thể vô hạn hoặc tổng thể có phân phối chuẩn, hoặc tổng thể hữu hạn và n / N <0,05 thì công thức z cho trung bình mẫu khi các mẫu là: $z = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0;1)$

3.5.3 Phân phối của tỷ lệ mẫu

Làm thế nào để một nhà nghiên cứu sử dụng tỷ lệ mẫu trong phân tích? Định lý giới hạn trung tâm áp dụng cho tỷ lệ mẫu trong đó phân phối chuẩn xấp xỉ hình dạng phân phối tỷ lệ mẫu nếu n.p> 5 và n.(1-p)> 5 (p là tỷ lệ tổng thể). Giá trị trung bình của tỷ lệ mẫu đối với tất cả các mẫu có kích thước n được chọn ngẫu nhiên từ tổng thể là p và độ lệch chuẩn của tỷ lệ mẫu $\sqrt{\dfrac{p(1-p)}{n}}$ đôi khi được gọi là sai số chuẩn của tỷ lệ.

Tỷ lệ mẫu cũng có công thức z : $z = \dfrac{f-p}{\sqrt{\dfrac{p(1-p)}{n}}} \sim N(0;1)$

Trong đó:

f : là tỷ lệ mẫu

n : là kích thước mẫu

p : là tỷ lệ tổng thể.

3.6. CÁC KHÁI NIỆM CƠ BẢN ƯỚC LƯỢNG

Khái niệm: Ước lượng tham số là tính toán một cách gần đúng nhất giá trị của một tham số chưa biết của tổng thể dựa trên thông tin từ một mẫu thống kê $(x_1, x_2, ..., x_n)$.

Có ba tham số của tổng thể thông dụng ,vì vậy ta có ba bài toán:

• Ước lượng trung bình tổng thể μ.

• Ước lượng tỷ lệ tổng thể p.

• Ước lượng phương sai tổng thể σ^2.

Các tham số của tổng thể μ, σ^2, p sẽ dùng ký hiệu θ làm đại diện chung để trình bày trong các phần sau. Khi ước lượng cho tham số của tổng thể θ dựa trên thông tin từ mẫu thì có hai loại ước lượng là ước lượng điểm và ước lượng khoảng.

3.6.1 Ước lượng điểm

Khái niệm ước lượng điểm

Ước lượng một tham số θ nào đó của tổng thể mà khi tính toán trên mẫu $(x_1, x_2, ..., x_n)$ ta chỉ ra được một giá trị gần đúng $\hat{\theta}$ của θ thì $\hat{\theta}$ gọi là *ước lượng điểm* của θ.

Rõ ràng $\hat{\theta} = f(x_1, x_2, ..., x_n)$ là một hàm của giá trị mẫu hay là một thống kê. Vì có nhiều hàm của mẫu, do đó có nhiều ước lượng điểm cho θ, vấn đề là chọn cái nào. Vì vậy cần xây dựng các tiêu chuẩn để đánh giá các ước lượng điểm.

Thống kê đã chỉ ra được các tiêu chuẩn để có ước lượng điểm tốt là tính không chệch, tính vững và tính hiệu quả.

Tính không chệch

Ước lượng $\hat{\theta} = f(x_1, x_2, ..., x_n)$ được gọi là *ước lượng không chệch* của θ nếu: $E(\hat{\theta}) = \theta$.

Nghĩa là đòi hỏi trung bình của $\hat{\theta}$ bằng giá trị cần tìm.

Ví dụ 13. Cân 100 sản phẩm của 1 xí nghiệp ta có bảng sau:

X	498	502	506	510
n	40	20	20	20

$$\overline{x} = \frac{498.40 + 502.20 + 506.20 + 510.20}{100} = 502,8.$$

Dự đoán trọng lượng trung bình của các sản phâm trong xí nghiệp là $\mu = 502,8$.

Vậy ước lượng cho kỳ vọng μ là $\overline{x}$. Đó là ước lượng không chệch vì $E(\overline{x}) = \mu$.

Tính vững

Ước lượng $\hat{\theta} = f(x_1, x_2, ..., x_n)$ được gọi là *ước lượng vững* của θ nếu:

$$E(\hat{\theta}) \xrightarrow{xs} \theta \text{ hay } \hat{\theta} \to \theta \text{ khi } n \to N.$$

Tính hiệu quả

Ước lượng $\hat{\theta} = f(x_1, x_2, ..., x_n)$ được gọi là *ước lượng hiệu quả* của θ nếu nó là ước lượng không chệch và có phương sai nhỏ nhất.

Nhận xét: Các tham số trung bình mẫu, phương sai mẫu, tỷ lệ mẫu là các ước lượng không chệch, hiệu quả và vững của trung bình tổng thể, phương sai tổng thể và tỷ lệ tổng thể.

3.6.2 Ước lượng khoảng

Giả sử θ là một tham số của tổng thể chưa biết. Khi đó với mẫu cụ thể $(x_1, x_2, ..., x_n)$ ta tìm được một khoảng (a,b)chứa θ sao cho $P[\theta \in (a,b)] = \gamma$ cho trước, thì:

- Khoảng (a,b) gọi là **khoảng tin cậy** hay ước lượng khoảng cho θ.
- γ gọi là **độ tin cậy** của ước lượng.

Ước lượng điểm dù tốt nhất cũng chỉ cho ta một giá trị trong một tập vô hạn, nên ta không biết được độ tin cậy của ước lượng. Vì không đánh giá được mức độ sai lầm khi dùng $\hat{\theta}$ thay cho θ, do đó trong thực tế người ta dùng ước lượng khoảng cho các tham số của tổng thể.

3.6.3 Các bài toán ước lượng khoảng cho trung bình

Gọi μ là trung bình của biến ngẫu nhiên X chưa biết, với độ tin cậy γ cho trước, ta tìm khoảng $(a;\, b)$ chứa μ theo công thức: $(a;\, b) = (\overline{x} - \varepsilon;\, \overline{x} + \varepsilon)$, với ε gọi là **sai số** hay **độ chính xác** của ước lượng.

1. Bài toán 1: Tìm độ chính xác hoặc tìm khoảng ước lượng cho trung bình

Ta thực hiện như sau:

- Tính $z_{\alpha/2}$.
- Tính độ chính xác (sai số ước lượng): $\varepsilon = \dfrac{z_{\alpha/2}.s}{\sqrt{n}}$ (hoặc $\varepsilon = \dfrac{z_{\alpha/2}.\sigma}{\sqrt{n}}$).
- Ước lượng khoảng cho trung bình tổng thể μ là: $\mu \in (\overline{x} - \varepsilon; \overline{x} + \varepsilon)$

2. **Bài toán 2: Ước lượng trung bình biết γ và ε, tìm kích thước mẫu n?**

- Tính s.
- Tính $z_{\alpha/2}$.
- Kích thước mẫu là: $n = \left(\dfrac{z_{\alpha/2}.s}{\varepsilon}\right)^2$

Khi tính kích thước mẫu, kết quả cuối cùng là số tự nhiên và phải làm tròn lên 1 đơn vị.

3. **Bài toán 3: Ước lượng trung bình biết n và ε, tìm độ tin cậy γ?**

- Tính n, s.
- Tính $z_{\alpha/2} = \dfrac{\varepsilon\sqrt{n}}{s}$.
- Độ tin cậy đạt được là: $\gamma = 2.\Phi(z_{\alpha/2}) - 1$

Bảng một số giá trị tới hạn thường gặp

Độ tin cậy γ	**0,90**	**0,91**	**0,92**	**0,93**	**0,94**	**0,95**	**0,96**	**0,97**	**0,98**	**0,99**
Giá trị tới hạn $z_{\alpha/2}$	1,65	1,70	1,75	1,81	1,88	1,96	2,06	2,17	2,33	2,58

Ví dụ 1. Một cuộc điều tra muốn ước lượng chiều cao trung bình của sinh viên trường A. Giả sử chiều cao tuân theo luật phân phối chuẩn. Điều tra 150 sinh viên của trường chiều cao trung bình tính được là 162 cm và độ lệch chuẩn 4 cm. Ước lượng chiều cao trung bình của sinh viên trường A, với độ tin cậy 95%.

Giải:

....

....

....

....

....

....

....

....

Ví dụ 2. Năng suất lúa trong 1 vùng là biến ngẫu nhiên. Gặt ngẫu nhiên 100ha của vùng này, người ta thu được bảng số liệu:

Năng suất (tạ/ha)	41	44	45	46	48	52	54
Diện tích (ha)	10	20	30	15	10	10	5

Ước lượng năng suất lúa trung bình của vùng trên với độ tin cậy 96% thì sai số của ước lượng là bao nhiêu?

Giải:

....

....

....

....

....

....

....

....

Ví dụ 3. Khảo sát thu nhập tại một doanh nghiệp, có số liệu sau:

Thu nhập (trđ/tháng)	6 - 8	8 - 10	10 - 12	12 - 14	14 - 16
Số lao động	15	20	30	25	5

Hãy ước lượng thu nhập trung bình 1 lao động của doanh nghiệp, với độ tin cậy 97%.

Giải:

....

....

....

....

....

....

....

Ví dụ 4. Người ta chọn một số sản phẩm của kho hàng và thu được kết quả sau:

Chiều dài sản phẩm (cm)	4	4,5	4,8	5,1	5,6	5,9	6,2
Số sản phẩm	7	11	20	18	25	9	10

Những sản phẩm loại I là sản phẩm có chiều từ 5cm trở lên. Hãy ước lượng chiều dài trung bình của sản phẩm loại I trong kho, với độ tin cậy 93%.
Giải:

....

....

....

....

....

....

....

....

Ví dụ 5. Quan sát tuổi thọ X (giờ) của một số bóng đèn do công ty A sản suất, ta có kết quả

X	1000	1100	1200	1300	1400
Số bóng	10	20	30	20	25

Nếu muốn ước lượng trung bình tuổi thọ bóng đèn với độ tin cậy 99% thì sai số là bao nhiêu ?
Giải:

....

....

....

....

....

....

....

....

Ví dụ 6. Điều tra trọng lượng của một loại trái cây, người ta cân thử một số trái cây và được kết quả cho trong bảng sau:

Trọng lượng (g)	50-55	55-60	60-65	65-70	70-75	75-80	80-85	85-90
Số trái cây	21	35	44	87	43	16	26	19

Muốn ước lượng trọng lượng trung bình của loại trái cây ở nông trường với độ tin cậy 96% và độ chính xác 2 g thì cần điều tra tối thiểu bao nhiêu trái?
Giải:

....

....

....

....

....

....

....

....

....

Ví dụ 7. Để khảo sát chỉ tiêu X của một lọai sản phẩm, người ta quan sát 1 mẫu và có kết quả như sau:

X (cm)	15-19	19-23	23-27	27-31	31-35	35-39
Số sp	9	20	16	16	12	18

Nếu muốn ước lượng giá trị trung bình của chỉ tiêu X của loại sp trên với độ chính xác 1,5 cm thì sẽ đạt độ tin cậy là bao nhiêu ?

Giải

....
....
....
....
....
....
....

Ví dụ 8. Khảo sát thu nhập (tr đ/năm) của một số người trong 1 công ty người ta thu được bảng số liệu sau:

Thu nhập	60-64	64-68	68-72	72-76	76-80	80-84	84-88
Số người	10	18	25	28	19	18	14

Nếu muốn ước lượng thu nhập trung bình của 1 nhân viên trong công ty với độ chính xác 1,2 triệu đồng/năm và độ tin cậy 99% thì phải khảo sát thêm ít nhất bao nhiêu người nữa ?

Giải:

....
....
....
....
....
....
....
....
....
....
....

Ví dụ 9. Một nhà máy muốn ước lượng thời gian trung bình để sản xuất một sản phẩm. Mẫu ngẫu nhiên 50 sản phẩm cho thấy thời gian sản xuất trung bình là 20 phút, độ lệch chuẩn 5 phút. Ước lượng thời gian trung bình để sản xuất một sản phẩm của nhà máy, với độ tin cậy 96%.

Giải:

....
....
....
....
....
....
....
....
....

3.6.5 Các bài toán ước lượng khoảng cho tỷ lệ

Giả sử tỷ lệ p các phần tử có tính chất A của tổng thể chưa biết. Với độ tin cậy γ, ước lượng khoảng cho p là $(p_1; p_2)$ sao cho: $P(p_1 < p < p_2) = \gamma$.

4. Bài toán 4: Tìm độ chính xác hoặc tìm khoảng ước lượng cho tỷ lệ

- Tính kích thước mẫu n, tính số lần biến cố A xuất hiện trong mẫu đó là *m*, từ đó tính được tỷ lệ mẫu là $f = \dfrac{m}{n}$.
- Tính $z_{\alpha/2}$.
- Tính độ chính xác (sai số ước lượng): $\varepsilon = z_{\alpha/2}.\sqrt{\dfrac{f(1-f)}{n}}$.
- Ước lượng khoảng cho tỷ lệ tổng thể p là: $p \in (f - \varepsilon;\ f + \varepsilon)$

5. Bài toán 5: Ước lượng tỷ lệ biết γ và ε, tìm kích thước mẫu n?

- Tính f.
- Tính $z_{\alpha/2}$.
- Kích thước mẫu là: $n = \dfrac{(z_{\alpha/2})^2 . f(1-f)}{\varepsilon^2}$

6. Bài toán 6: Ước lượng tỷ lệ biết n và ε, tìm độ tin cậy γ?

- Tính n , m , $f = \dfrac{m}{n}$.
- Tính $z_{\alpha/2} = \varepsilon\sqrt{\dfrac{n}{f(1-f)}}$.
- Độ tin cậy đạt được là: $\gamma = 2.\Phi(z_{\alpha/2}) - 1$

7. Bài toán 7: Tìm kích thước mẫu nhỏ nhất để độ chích xác ước lượng tỷ lệ không vướt quá ε và độ tin cậy là γ, khi không biết tỷ lệ mẫu f .

Khi đó áp dụng BĐT Côsi ta luôn có: $f(1-f) \le \left(\dfrac{f+1-f}{2}\right)^2 = \dfrac{1}{4}$

- Tính $z_{\alpha/2}$.
- Kích thước mẫu là: $n \ge \dfrac{(z_{\alpha/2})^2}{4\varepsilon^2}$

Chú ý: ***Khi tính kích thước mẫu, kết quả cuối cùng là số tự nhiên và phải làm tròn lên 1 đơn vị.***

Độ tin cậy γ	**0,90**	**0,91**	**0,92**	**0,93**	**0,94**	**0,95**	**0,96**	**0,97**	**0,98**	**0,99**
Giá trị tới hạn $z_{\alpha/2}$	1,65	1,70	1,75	1,81	1,88	1,96	2,06	2,17	2,33	2,58

Ví dụ 10. Năng suất lúa trong 1 vùng là biến ngẫu nhiên. Gặt ngẫu nhiên một số ha của vùng này, người ta thu được bảng số liệu:

Năng suất (tạ/ha)	41	44	45	46	48	52	54
Diện tích (ha)	5	10	15	15	20	10	5

Những thửa ruộng trong vùng trên có năng suất từ 48 tạ/ha trở lên là những thửa có năng suất cao. Ước lượng tỉ lệ diện tích có năng suất cao với độ tin cậy 97%.
Giải:

....
....
....
....
....
....
....
....
....
....
....

Ví dụ 11. Từ một lô thuốc ta kiểm tra 200 ống thì thấy có 17 ống bị đục. Với đô tin cậy 94%, ước lượng khoảng của tỷ lệ thuốc không bị đục của toàn lô thuốc, thì độ chính xác (sai số) bằng bao nhiêu.
Giải:

....
....
....
....
....
....
....
....
....
....
....

Ví dụ 12. Lấy mẫu của 1 nhà máy sản xuất xi măng để kiểm tra về sức chịu lực R (kg/cm^2) ta được kết quả:

X (kg/cm^2)	10-11	11-12	12-13	13-14	14-15	15-16	16-17	17-18
Số sp	5	10	14	12	17	9	8	2

Xi măng được xếp vào loại II nếu sức chịu lực dưới 14 kg/cm^2. Hãy ước lượng tỷ lệ xi măng loại II của nhà máy, với độ tin cậy 96% .

Giải:

Ví dụ 13. Điều tra thu nhập (triệu đồng/tháng) một số người của công ty X người ta được kết quả sau:

12;	11;	12;	14;	14;	15;	13;	16;	16;	10;	13;	13;
14;	8;	10;	16;	11;	12;	14;	13;	17;	13;	16;	12;
10;	13;	14;	15;	14;	14;	13;	13;	12;	14;	11;	15;
11;	16;	14;	14;	12;	12;	11;	13;	14;	11;	16;	10.

Hãy ước lượng tỷ lệ người có thu nhập dưới 14 triệu đồng/tháng của công ty X, với độ tin cậy 98%.

Giải:

Ví dụ 14. Để khảo sát chỉ tiêu X của một loại sản phẩm, người ta quan sát 1 mẫu và có kết quả như sau:

X cm	11-15	15-19	19-23	23-27	27-31	31-35	35-39
Số sp	8	10	20	16	15	13	8

Những sản phẩm có chỉ tiêu X không quá 19 cm được xếp vào loại B. Nếu muốn ước lượng tỷ lệ các sp loại B với độ chính xác 8% thì sẽ đạt độ tin cậy là bao nhiêu ?

Giải:

Ví dụ 15. Để ước lượng tỷ lệ mắc bệnh gan đối với những người có dấu hiệu X với độ tin cậy 92% và sai số không vượt quá 3% thì cần phải khám ít nhất bao nhiêu người có dấu hiệu X này, biết rằng tỷ lệ mắc bệnh gan thực nghiệm là 0,8.
Giải:
....
....
....
....
....
....
....

Ví dụ 16. Muốn ước lượng tỷ lệ người thu nhập thấp tại thành phố A với sai số không quá 5% và độ tin cậy 95%. Hỏi phải chọn ít nhất bao nhiêu người.
Giải:
....
....
....
....
....
....
....

Ví dụ 17. Người ta muốn biết tỷ lệ học thêm của học sinh cấp tiểu học tại TP.HCM. Nếu muốn sai số ước lượng không quá 0,06 ở độ tin cậy 97% thì phải quan sát ít nhất mấy trường hợp?
....
....
....
....
....
....
....

Ví dụ 18. Người ta xếp 100 trái ổi vào 1 thùng, có rất nhiều thùng như thế. Kiểm tra ngẫu nhiên 50 thùng thấy có 200 trái ổi không đạt tiêu chuẩn. Ước lượng tỉ lệ trái ổi không đạt tiêu chuẩn với độ tin cậy 94%.
Giải:
....
....
....
....
....
....
....
....
....
....

Ví dụ 19. Khảo sát điểm thi môn Tiếng Anh của một lớp trong trường A, được kết quả sau:

X (điểm)	4	5	6	7	8	9	10
Số sinh viên	9	15	18	23	17	12	10

Những sinh viên đạt từ 9 điểm trở lên là những sinh viên đạt loại giỏi. Muốn ước lượng tỷ lệ sinh viên đạt điểm loại giỏi môn Tiếng Anh của trường A với độ tin cậy 93% thì sai số (độ chính xác) là bao nhiêu?

Giải.

....
....
....
....
....
....
....
....
....

Ví dụ 20. Từ một kho hàng gồm 5000 sản phẩm của hai nhà máy A và B. Người ta chọn ngẫu nhiên 200 sản phẩm trong kho thì thấy có 120 sản phẩm của nhà máy A. Hãy ước lượng số sản phẩm của nhà máy A trong kho hàng với độ tin cậy 95%.

Giải:

....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....

BÀI TẬP CHƯƠNG 3

Câu 1: Để khảo sát chỉ tiêu X của một loại sản phẩm trong nhà máy, người ta quan sát một mẫu và có kết quả sau:

X(cm)	11-15	15-19	19-23	23-27	27-31	31-35	35-39
Số sản phẩm	8	9	20	16	16	13	18

Những sản phẩm có chỉ tiêu X từ 23cm trở xuống gọi là những sản phẩm loại B. Nếu ước lượng tỷ lệ những sản phẩm loại B với độ chính xác 12% thì sẽ đạt độ tin cậy bao nhiêu?

Câu 2: Để khảo sát chiều cao của một giống cây trồng, người ta quan sát một mẫu và có kết quả sau:

X(cm)	95-105	105-115	115-125	125-135	135-145	145-155	155-165
Số cây	10	10	30	10	10	10	15

Những cây có chiều cao từ 135cm trở lên được gọi là những cây cao.Hãy ước lượng tỷ lệ cây cao , với độ tin cậy 95% .

Câu 3: Để khảo sát trọng lượng X của một loại vật nuôi trong nông trại, người ta quan sát một mẫu và có kết quả sau:

X(kg)	28-32	32-36	36-40	40-44	44-48	48-52	52-56
Số con	10	10	15	30	10	10	15

Ước lượng trọng lượng trung bình của loại vật nuôi trên với độ tin cậy 99%.

Câu 4: Điểm trung bình môn Toán của 100 thí sinh dự thi vào đại học là 5 với độ lệch mẫu hiệu chỉnh là 2,5. Muốn ước lượng điểm trung bình môn Toán với sai số 0,25 điểm thì độ tin cậy là bao nhiêu?

Câu 5: Để theo dõi sự phát triển chiều cao của một giống cây trồng trong một nông trại , người ta quan sát một mẫu và có kết quả sau:

X(cm)	70-74	74-78	78-82	82-86	86-90	90-94
Số cây	20	10	30	20	10	10

Những cây có chiều cao từ 82cm trở lên được gọi là những cây cao. Hãy ước lượng tỷ lệ cây cao, với độ tin cậy 97%.

Câu 6: Để theo dõi sự phát triển chiều cao của một giống cây trồng trong một nông trại , người ta quan sát một mẫu và có kết quả sau:

X(cm)	68-72	72-76	76-80	80-84	84-88	88-92
Số cây	20	10	30	20	10	10

Ước lượng chiều cao trung bình của giống cây trồng trên với độ tin cậy 96%.

Câu 7: Để ước lượng mức xăng tiêu hao trung bình cho 1 loại ôtô chạy từ A đến B, phòng kỷ thuật của công ty quan sát mức tiêu hao xăng X (lít) trong 30 chuyến xe, kết quả như sau:

X	$\leq 9{,}8$	9,8-10	10-10,2	10,2-10,4	>10,4
Số chuyến	3	5	10	8	4

Với độ tin cậy 95%, hãy ước lượng mức tiêu hao xăng trung bình.

Câu 8: Chọn ngẫu nhiên 36 công nhân của 1 xí nghiệp thì thấy lương trung bình là 600 ngàn đ /tháng và độ lệch chuẩn 30 ngàn đồng. Với độ tin cậy 95%, hãy ước lượng mức lương trung bình của công nhân toàn xí nghiệp.

Câu 9: Lô trái cây của 1 chủ hàng được đóng thành sọt 100 trái. Kiểm tra 50 sọt thấy có 450 trái không đạt tiêu chuẩn. Muốn ước lượng tỷ lệ trái cây không đạt tiêu chuẩn với độ tin cậy 99% thì độ chính xác là bao nhiêu?

Câu 10: Để khảo sát chỉ tiêu X của một loại sản phẩm trong nhà máy, người ta quan sát một mẫu và có kết quả sau:

X(kg)	15	20	25	30	35	40	45
Số sp	8	13	15	12	14	6	13

Nếu muốn ước lượng giá trị trung bình chỉ tiêu X của loại sản phẩm đã cho với độ tin cậy 99% và độ chính xác 2kg thì phải điều tra thêm bao nhiêu sản phẩm nữa?

Chương 4. KIỂM ĐỊNH GIẢ THUYẾT THỐNG KÊ

MỤC TIÊU CHƯƠNG

Nội dung chương này giúp người học có khả năng:

- Biết cách đặt giả thuyết kiểm định, hiểu được mức ý nghĩa của kiểm định và các loại sai lầm trong kiểm định.
- Trình bày được các bước giải bài toán kiểm định về trung bình, tỷ lệ và phương sai tổng thể.
- Hiểu được khái niệm *p-value* và áp dụng được vào bài toán kiểm định về trung bình và tỷ lệ tổng thể.

4.1 Kiểm định giả thuyết

4.1.1 Bài toán kiểm định giả thiết thống kê

Chương trước đã trình bày phương pháp dùng dữ liệu mẫu để ước lượng các đặc trưng chưa biết như: trung bình, tỷ lệ, phương sai của tổng thể. Bên cạnh việc ước lượng, các đặc trưng của tổng thể còn có thể được suy diễn từ dữ liệu mẫu thông qua phương pháp kiểm định giá thiết. Kiểm định giả thiết là dựa vào mẫu dữ liệu thu thập được để đưa ra kết luận –bác bỏ hay không thể bác bỏ – về các giả thiết của tổng thể.

Bài toán đặt ra là khi có một giả thiết về một vấn đề nào đó, yêu cầu chúng ta phải chấp nhận hay loại bỏ giả thuyết đó trên cơ sở thông tin từ một mẫu đại diện. Dĩ nhiên, yêu cầu là chọn tình huống nào để khả năng đúng cao hơn, khả năng sai thấp hơn. Tất nhiên không thể yêu cầu cho một câu trả lời đúng hoàn toàn, không mắc sai lầm nào, khi mà thông tin của chúng ta chỉ dựa trên một mẫu đại diện

4.1.2 Đặt giả thuyết thống kê

Có thể xác định nghiên cứu khoa học như là một qui trình thử nghiệm giả thuyết, theo các bước sau đây:

Bước 1: Nhà nghiên cứu cần phải định nghĩa một **giả thuyết không** (**null hypothesis**) ký hiệu là H_0, là một giả thuyết ngược lại với những gì mà nhà nghiên cứu tin là sự thật.

Ví dụ. Trong một nghiên cứu lâm sàng, gồm hai nhóm bệnh nhân: một nhóm được điều trị bằng thuốc mới A, và một nhóm được điều trị bằng thuốc B, nhà nghiên cứu có thể phát biểu giả thuyết H_0 rằng sự hiệu nghiệm thuốc mới A tương đương với sự hiệu nghiệm của thuốc B.

Bước 2: Nhà nghiên cứu cần phải định nghĩa một **giả thuyết đối** (**alternative hypothesis**) ký hiệu là H_1, là một giả thuyết mà nhà nghiên cứu nghĩ là sự thật, và điều cần được "chứng minh" bằng dữ kiện.

Chẳng hạn như trong ví dụ trên đây, nhà nghiên cứu có thể phát biểu giả thuyết thay thế H_1 rằng thuốc mới A có hiệu nghiệm cao thuốc B.

Bước 3: Sau khi đã thu thập đầy đủ những dữ kiện liên quan, nhà nghiên cứu dùng một hay nhiều phương pháp thống kê để kiểm tra xem trong hai giả thuyết trên, giả thuyết nào được xem là khả dĩ. Cách kiểm tra này được tiến hành để trả lời câu hỏi: nếu giả thuyết H_0 đúng, thì xác suất mà những dữ kiện thu thập được phù hợp với giả thuyết H_0 là bao nhiêu. Giá trị của xác suất này thường được đề cập đến trong các báo cáo bằng kí hiệu "*p- value*".

Bước 4: Quyết định chấp nhận hay loại bỏ giả thuyết H_0, bằng cách dựa vào giá trị xác suất trong bước thứ ba.

Chẳng hạn như theo truyền thống lựa chọn trong một nghiên cứu khoa học, nếu giá trị xác suất nhỏ hơn 5% thì nhà nghiên cứu sẵn sàng bác bỏ giả thuyết H_0: sự hiệu nghiệm của thuốc A khác với sự hiệu nghiệm của thuốc B. Tuy nhiên, nếu giá trị xác suất cao hơn 5%, thì nhà nghiên cứu chỉ có thể phát biểu rằng chưa có bằng chứng đầy đủ để bác bỏ giả thuyết H_0, và điều này không có nghĩa rằng giả thuyết H_0 là đúng, là sự thật. Nói một cách khác, thiếu bằng chứng không có nghĩa là không có bằng chứng.

Bước 5: Nếu giả thuyết H_0 bị bác bỏ, thì nhà nghiên cứu mặc nhiên thừa nhận giả thuyết H_1. Nhưng vấn đề khởi đi từ đây, bởi vì có nhiều giả thuyết thay thế khác nhau. Chẳng hạn như so sánh với giả thuyết H_0 ban đầu (A khác với B), nhà nghiên cứu có thể đặt ra nhiều giả thuyết H_1 khác nhau như sự hiệu nghiệm của thuốc A cao hơn B là 5%, 10% .... Nói tóm lại, một khi nhà nghiên cứu bác bỏ giả thuyết H_0, thì giả thuyết H_1 được mặc nhiên công nhận, nhưng nhà nghiên cứu không thể xác định giả thuyết H_0 nào là đúng với sự thật.

Việc kiểm tra xem chấp nhận được hay không chấp nhận giả thiết H_0 dựa vào thông tin của mẫu ngẫu nhiên (X_1, X_2, . . ., Xn) nên được gọi là kiểm định thống kê.

Ý cơ bản để giải bài toán này là:

Chia miền giá trị có thể của mẫu ngẫu nhiên (chia có lựa chọn) thành 2 phần:

- S là phần bác bỏ giả thiết H_0
- $\overline{S}$ là phần chấp nhận giả thiết H_0

và thực hiện qui tắc sau:

- Nếu mẫu $\left(X_1, X_2, ..., X_n\right) \in S$ thì ta bác bỏ H_0 (do đó tạm chấp nhận H_1).
- Nếu mẫu $\left(X_1, X_2, ..., X_n\right) \in \overline{S}$ thì ta chưa có cơ sở bác bỏ H_0 (do đó tạm chấp nhận H_0 và bác bỏ H_1).

4.1.3 Một số nguyên tắc liên quan đến việc đặt giả thiết

1. Giả thuyết H_0 và giả thuyết H_1 phải trái, ngược nhau.
2. Dấu trong cấu trúc của H_0 phải là: $=$; $\geq$; $\leq$.

   Dấu trong cấu trúc của H_1 phải là: $\neq$; $<$; $>$.
3. Giả thuyết H_1 sẽ quyết định trường hợp kiểm định, gồm 3 trường hợp: kiểm định hai phía $\left(H_1: \neq\right)$, kiểm định bên phải $\left(H_1: >\right)$ và kiểm định bên trái $\left(H_1: <\right)$.

Cần cận trọng khi thiết lập các giả thuyết một cách hợp lý sao cho kết luận thu được của kiểm định cung cấp thông tin mà người nghiên cứu hoặc người ra quyết định yêu cầu. Sau đây là ba trường hợp thường sử dụng đến quy trình kiểm định giả thuyết.

(1). Kiểm định giả thuyết nghiên cứu

Giả sử một mẫu xe hơi A đang đạt mức tiêu thụ nhiên liệu 100 km/7 lit. Nhóm nghiên cứu đang phát triển một hệ thống phun nhiên liệu mới để giảm mức tiêu thụ nhiên liệu. Để đánh giá hệ thống mới, họ thử nghiệm trên một mẫu xe A. Ở đây nhóm nghiên cứu đang tìm bằng chứng để kết luận rằng hệ thống mới có hiệu quả, tức làm giảm lượng xăng trung bình trên 100km. Trong trường hợp này giả thuyết nghiên cứu là lượng nhiên liệu tiêu thụ của loại xe A trên 100km ít hơn 7 lit. Theo thông lệ nên đặt giả thuyết nghiên cứu là giả thuyết đối, vậy gỉ thuyết cần đặt là: $H_0: \mu \ge 7$; $H_1: \mu < 7$.

Trong các nghiên cứu, giả thuyết không và giả thuyết đối nên thiết lập sao cho việc bác bỏ H_0 sẽ ủng hộ cho kết luận của nghiên cứu. Do vậy, giả thuyết nghiên cứu nên được đặt thành giả thuyết đối là H_1.

(2). Kiểm định hiệu lực của một tuyên bố

Một nhà sản xuất tuyên bố rằng chai nước tinh khiết loại 2 lít chứa trung bình ít nhất 1,96 lít nước tinh khiết. Một mẫu các chai nước loại này của nhà sản xuất sẽ được chọn ra đo lường lượng nước chứa bên trong để kiểm định phát biểu của nhà sản xuất. Đối với những trường hợp kiểm định giả thuyết dạng này, chúng ta thường giả định rằng tuyên bố của nhà sản xuất là đúng, trừ khi có bằng chứng từ mẫu chỉ ra sự mâu thuẫn. Vì vậy giả thuyết kiểm định cho bài toán này là $H_0: \mu \ge 1{,}96$; $H_1: \mu < 1{,}96$.

Trong kiểm định về hiệu lực của một tuyên bố, giả thuyết không H_0 thường dựa trên giả định tuyên bố là chính xác. Sau đó thiết lập giả thuyết đối sao cho khi bác bỏ H_0 sẽ cung cấp bằng chứng thống kê cho thấy tuyên bố trên là không chính xác.

(3). Kiểm định các tình huống ra quyết định

Một mẫu các linh kiện từ một lô hàng vừa gởi đến, người kiểm tra chất lượng sản phẩm phải quyết định xem có chấp nhận lô hàng từ nhà cung cấp hay phải trả lại lô hàng vì không đạt tiêu chuẩn. Giả sử rằng tiêu chuẩn cho một linh kiện là phải đạt độ dài trung bình 5 cm. Nếu độ dài trung bình lớn hơn hoặc nhỏ hơn 5 cm, linh kiện này sẽ khiến việc lắp ráp gặp vấn đề. Trong trường hợp này ta thiết lập giả thuyết như sau: $H_0: \mu = 5$; $H_1: \mu \ne 5$.

Nếu kết quả từ mẫu cho thấy không thể bác bỏ H_0, người kiểm tra sẽ không có lý do gì để nghi nghờ lô hàng này không đạt tiêu chuẩn, và lô hàng sẽ được chấp nhận. Tuy nhiên nếu kết quả từ mẫu cho thấy có thể bác bỏ H_0, thì kết luận các linh kiện không đạt tiêu chuẩn. Trong trường hợp này người kiểm tra có đủ bằng chứng để trả lại lô hàng cho nhà cung cấp.

4.1.4 Các loại sai lầm mắc phải

Làm theo qui tắc trên sẽ phạm phải hai sai lầm:

- **Sai lầm loại I:** Bác bỏ giả thiết H_0, nhưng thực tế H_0 đúng.
- **Sai lầm loại II:** Chấp nhận giả thiết H_0, nhưng thực tế H_0 sai.

Tất nhiên ta mong muốn chọn giả thiết nào để cực tiểu cả hai khả năng phạm sai lầm. Nhưng thống kê Toán học đã chứng minh bài toán không có lời giải. Người ta khống chế sai lầm loại I và cực tiểu khả năng phạm sai lầm loại II. Nghĩa là cho trước một giới hạn trên là α của xác suất phạm sai lầm loại I (số α thường nhỏ) và bài toán đưa đến là: hãy chọn giả thiết nào để sao cho:

1) Xác suất phạm sai lầm loại I không vượt quá α.

2) Khả năng phạm sai lầm loại II đạt cực tiểu.

Số α gọi là **mức ý nghĩa** của bài toán.

Ta có thể nghĩ tại sao không đặt $\alpha = 0$, vì khi đó khả năng phạm sai lầm loại II là bằng 1. Khi giảm khả năng mắc sai lầm loại I thì khả năng mắc sai lầm loại II tăng lên.

Như vậy, khi phát biểu bài toán kiểm định giả thiết thống kê ta phải chỉ rõ: H_0, H_1, α.

Người ta sử dụng xác suất phạm sai lầm loại I được ký hiệu bằng α để xác định trị tới hạn cho khu vực bác bỏ H_0 của phân phối mẫu. Việc lựa chọn giá trị α tùy thuộc vào mức độ tổn thất mà người kiểm định có thể "chịu đựng". Thường chọn giá trị 0,05 Vùng còn lại của phân phối mẫu có trị bằng (1- α), ký hiệu là γ được gọi là độ tin cậy, là vùng không bác bỏ H_0.

Cần hai yếu tố để xác định giá trị tới hạn đó là giả thuyết H_0 và xác suất chấp nhận sai số loại I là α (mức ý nghĩa). Khoảng giá trị thống kê tương ứng với xác suất chấp nhận sai lầm loại I được gọi là vùng *bác bỏ H_0*, khoảng giá trị thống kê còn lại được gọi là vùng *không bác bỏ H_0*.

4.1.5 Đặt vấn đề

Gọi θ là một tham số (giá trị trung bình, phương sai, hoặc tỷ lệ) chưa biết của tổng thể. Giả sử ta hình thành một giả thiết về θ, gọi là giả thiết H_0, và trong kiểm định giả thiết H_0 được xem là đúng cho đến khi đủ chứng cứ để kết luận khác hơn là giả thiết đối H_1.

Chẳng hạn, sản phẩm bột dinh dưỡng trẻ em của một công ty sản xuất được qui định đóng gói có trọng lượng của mỗi gói bột là 450g. Để kiểm tra qui định này, ta có thể đặt giả thiết:

$$\begin{cases} H_0 : \mu = 450 \\ H_1 : \mu \neq 450 \end{cases} \text{ hoặc } \begin{cases} H_0 : \mu = 450 \\ H_1 : \mu > 450 \end{cases} \text{ hoặc } \begin{cases} H_0 : \mu = 450 \\ H_1 : \mu < 450 \end{cases}$$

Như vậy, trong giả thiết H_0 thì θ có thể bằng một giá trị cụ thể θ_0 hoặc một khoảng giá trị nào đó $\theta \geq \theta_0$.

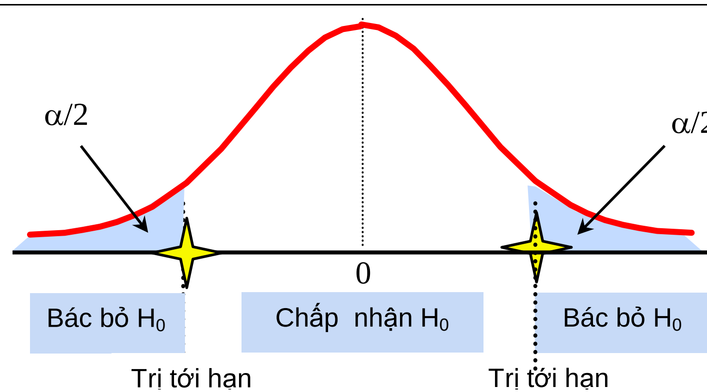
4.1.6 Qui tắc kiểm định

Dựa vào việc so sánh giá trị kiểm định với khoảng trị tới hạn được xác định. Kết luận được đưa ra tùy thuộc vùng mà giá trị kiểm định rơi vào. Nếu giá trị kiểm định nằm trong vùng "bác bỏ H_0" thì kết luận **bác bỏ giả thuyết H_0** và dùng giả thuyết H_1 làm kết luận của mình. Nếu giá trị kiểm định rơi vào vùng "**không bác bỏ H_0**" thì kết luận không có cơ sở để bác bỏ giả thuyết H_0.

Kiểm định 2 phía (two-sided)

Vùng bác bỏ H_0 sẽ bị chia làm đôi nên được gọi là kiểm định hai phía.

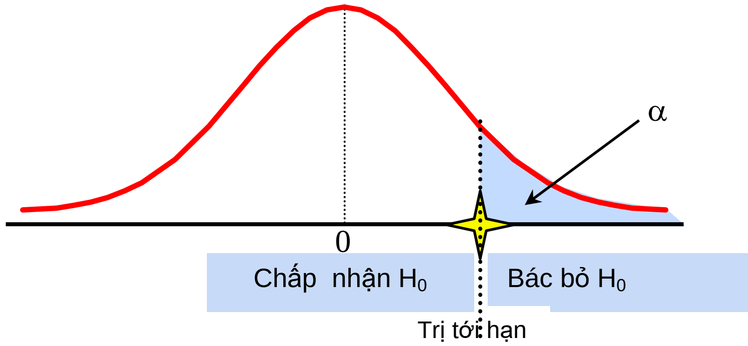
Giá trị tại dấu ✦ là trị tới hạn, dựa vào mức ý nghĩa và kiểu phân phối dữ liệu để tra tìm giá trị tới hạn.



Kiểm định hai phía.

Kiểm định bên phải (one-sided)

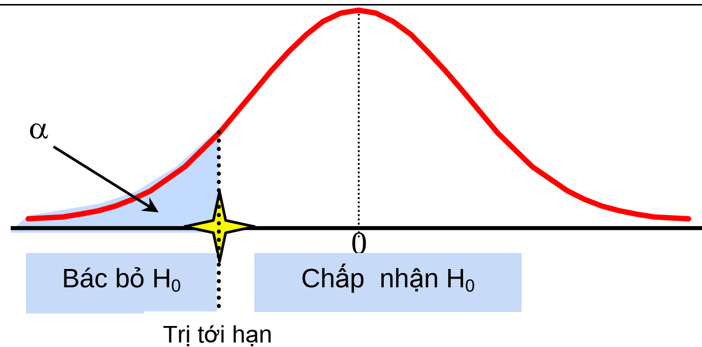
Vùng nhỏ hơn 0 của phân phối mẫu sẽ là vùng không bác bỏ H_0 . Vùng bác bỏ H_0 phải thuộc phía dương, tức nằm về bên phải điểm 0 nên được gọi là kiểm định bên phải.



Kiểm định bên phải.

Kiểm định bên trái (one-sided)

Phía dương không bác bỏ H_0; vùng bác H_0 thuộc về bên trái điểm 0 nên được gọi là kiểm định bên trái.



Kiểm định bên trái.

4.2 Kiểm định giả thiết về trung bình tổng thể

Giả sử ta có mẫu gồm n giá trị quan sát được chọn ngẫu nhiên từ một tổng thể nào đó có phân phối chuẩn $N\left(\mu,\sigma^2\right)$. Kiểm định giả thiết về trung bình tổng thể μ được thực hiện như sau (xét trường hợp phổ biến nhất):

Tính kích thước mẫu, trung bình và độ lệch chuẩn của mẫu: $n,\ \bar{x},\ s$

Giả thuyết	Giá trị kiểm định	Giá trị tới hạn	Quy tắc kiểm định
$I:\begin{cases}H_0:\mu=\mu_0\\H_1:\mu\neq\mu_0\end{cases}$		$z_{\alpha/2}$	Nếu $\lvert z\rvert > z_{\alpha/2}$ thì bác bỏ H_0
$II:\begin{cases}H_0:\mu\le\mu_0\\H_1:\mu>\mu_0\end{cases}$	$z=\dfrac{\left(\bar{x}-\mu_0\right)\sqrt{n}}{s}$	z_α	Nếu $z > z_\alpha$ thì bác bỏ H_0
$III:\begin{cases}H_0:\mu\ge\mu_0\\H_1:\mu<\mu_0\end{cases}$	(nếu biết σ thì thay s bằng σ)	z_α	Nếu $z < -z_\alpha$ thì bác bỏ H_0

CÁC BƯỚC SUY LUẬN KHI ĐẶT GIẢ THUYẾT:

1) *Tìm* ***câu văn*** *trong đề bài phát biểu về* ***"trung bình hay tỷ lệ so sánh với một số".***

2) *Nếu câu văn này có dạng câu là một:* ***Tuyên bố****;* ***Kết luận****;* ***Quy định****;* ***Báo cáo*** *thì câu văn này* ***là giả thuyết H_0****, và chuyển câu văn qua biểu thức Toán với các dấu****: = ; ≥ ; ≤***

3) *Nếu câu văn này có dạng câu là một:* ***nghi ngờ****,* ***điều muốn chứng minh****,* ***cần kiểm tra*** *thì câu văn này* ***là giả thuyết H_1,*** *và chuyển câu văn qua biểu thức Toán với các dấu****: ≠ ; > ; <***

4) *Nếu gặp trường hợp giả thuyết* ***H_0 là dấu " = "*** *thì* ***đọc tiếp trong câu hỏi đề bài*** *để xác định dấu của H_1 cho đúng.*

5) *Nếu đề bài nói đến một sự:* ***Cải tiến; Đổi mới; Nâng cấp; Thay đổi vv...*** *thì* ***cần phải kiểm tra xem có hiệu quả không là giả thuyết H_1.***

6) Giải bài toán kiểm định là dựa vào kết luận chấp nhận hay bác bỏ giả thuyết H_0 để trả lời câu hỏi của đề bài.

Bảng một số giá trị tới hạn thường gặp của phân phối chuẩn

α	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**	**0.1**
$z_{\alpha/2}$	2.576	2.326	2.17	2..054	1.96	1.881	1.812	1.751	1.695	1.645
z_{α}	2.326	2.054	1.881	1.751	1.645	1.555	1.476	1.405	1.341	1.282

Ví dụ 1. Một công ty điện thoại nói rằng thời gian lắp đặt điện thoại trung bình cho khách hàng trong thành phố chậm nhất là 2 tuần kể từ khi có yêu cầu. Kiểm tra ngẫu nhiên 50 khách hàng thấy thời gian trung bình chờ lắp điện thoại là 15 ngày và độ lệch chuẩn là 2 ngày. Với mức ý nghĩa 3%, có thể chấp nhận lời tuyên bố của công ty được không?

Giải:

....
....
....
....
....
....
....
....
....
....
....
....
....
....

Ví dụ 2. Trong năm trước doanh thu trung bình tại cửa hàng X là 80 triệu đồng/ngày. Đầu năm nay cửa hàng cải tạo lại mặt bằng kinh doanh. Để đánh giá xem việc làm này có hiệu quả hay không, người ta kiểm tra doanh thu của 30 ngày thì thấy doanh thu trung bình là 100 triệu đồng và độ lệch tiêu chuẩn hiệu chỉnh là 15 triệu đồng. Với mức ý nghĩa 2%, hãy cho biết kết luận .

Giải:

....
....
....
....
....
....
....
....
....
....
....
....
....

Ví dụ 3. Trọng lượng một loại sản phẩm do nhà máy A sản xuất có phân phối chuẩn và trọng lượng quy định là 500g. Nghi ngờ trọng lượng có xu hướng giảm sút, người ta cân ngẫu nhiên một số sản phẩm loại này và có bảng số liệu:

Trọng lượng (gam)	485	490	495	500	505	510
Số sản phẩm	13	18	15	13	24	8

Với mức ý nghĩa 0,03, hãy cho kết luận về điều nghi ngờ nói trên?

Giải:

....
....
....
....
....
....
....
....
....
....
....
....
....

Ví dụ 4. Trong kho có rất nhiều sản phẩm của xí nghiệp A, trọng lượng X (kg) của các sản phẩm này là biến ngẫu nhiên có phân phối chuẩn. Cân ngẫu nhiên một số sản phẩm loại này, có kết quả:

X	0,85	0,9	0,95	1	1,05	1,1
Số sp	10	20	10	15	15	25

Có thể nói rằng trọng lượng trung bình sản phẩm này đạt đến hơn 1 kg được không? Với mức ý nghĩa 4%.

Giải:

....

....

....

....

....

....

....

....

....

....

....

....

....

Ví dụ 5. Theo một báo cáo từ mười năm trước chiều cao trung bình của các em lứa tuổi lên 10 ở một vùng nông thôn không lớn hơn 140 cm. Để xem hiện nay chiều cao của các em lứa tuổi lên 10 ở vùng nông thôn này có tăng lên hay không. Người ta lấy ra một mẫu đại diện với các kết quả như sau:

Chiều cao (cm)	≤ 130	130-135	135-140	140-145	≥ 145
Số em	5	15	20	35	5

Giả sử chiều cao tuân theo luật phân phối chuẩn. Với mức ý nghĩa 0,05, hãy cho biết kết luận.

Giải:

....

....

....

....

....

....

....

....

....

....

....

Ví dụ 6. Quan sát doanh thu (tỷ đồng) trong năm qua của các cửa hàng trong một công ty ta được kết quả sau:

12	13	14	11	16	9	8	12	7	15	16	18	11	10
9	12	6	11	12	14	6	8	9	8	10	9	15	16
18	11	10.											

Người ta kết luận doanh thu trung bình của các cửa hàng là 12 tỷ đồng thì có chấp nhận được không? Với mức ý nghĩa 0,06.

Giải:

....
....
....
....
....
....
....
....
....
....

Ví dụ 7. Kiểm tra ngẫu nhiên số gạo bán ra hàng ngày ở một cửa hàng, có kết quả:

Số gạo bán ra (kg)	120	130	150	160	180	190	210	220
Số ngày bán	2	9	12	25	30	25	13	4

Chủ cửa hàng cho rằng nếu trung bình mỗi ngày bán ra dưới 180kg gạo thì tốt nhất là nghỉ bán. Từ số liệu trên, với mức ý nghĩa 3% cửa hàng nên quyết định thế nào?

Giải:

....
....
....
....
....
....
....
....
....
....
....
....
....

Ví dụ 8. Theo dõi sự phát triển chiều cao X(dm) của cây bạch đàn trồng trên đất phèn sau một năm tuổi, có kết quả:

X (dm)	25 – 30	30 – 35	35 – 40	40 – 45	45 – 50	50 – 55	55 - 60
Số cây	5	20	25	35	30	23	14

Biết chiều cao trung bình của bạch đàn sau một năm tuổi ở đất không có phèn là 45dm. Với mức ý nghĩa 4%, có cần tiến hành kháng phèn cho bạch đàn không?
Giải:

....

....

....

....

....

....

....

....

....

....

....

....

....

....

4.3 Kiểm định giả thiết về tỷ lệ

Giả sử tỷ lệ các phần tử mang dấu hiệu A nào đó của tổng thể là p chưa biết. Với mỗi số α khá bé, hãy dựa vào mẫu kích thước n (n khá lớn) để đưa ra qui tắc kiểm định.
Trong n phần tử của mẫu tìm xem có m phần tử mang dấu hiệu A, từ đó tính được tỷ lệ mẫu $f = m/n$.

Giả thuyết	Giá trị kiểm định	Giá trị tới hạn	Quy tắc kiểm định
$I:\begin{cases} H_0: p = p_0 \\ H_1: p \neq p_0 \end{cases}$		$z_{\alpha/2}$	Nếu $\lvert z \rvert > z_{\alpha/2}$ thì bác bỏ H_0
$II:\begin{cases} H_0: p \le p_0 \\ H_1: p > p_0 \end{cases}$	$z = \dfrac{(f - p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}}$	z_α	Nếu $z > z_\alpha$ thì bác bỏ H_0
$III:\begin{cases} H_0: p \ge p_0 \\ H_1: p < p_0 \end{cases}$		z_α	Nếu $z < -z_\alpha$ thì bác bỏ H_0

α	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**	**0.1**
$z_{\alpha/2}$	2.576	2.326	2.17	2..054	1.96	1.881	1.812	1.751	1.695	1.645
z_α	2.326	2.054	1.881	1.751	1.645	1.555	1.476	1.405	1.341	1.282

Ví dụ 9. Tỷ lệ phế phẩm của một nhà máy là 5%. Sau khi cải tiến kỷ thuật, người ta kiểm tra 400 sản phẩm thì thấy có 16 phế phẩm. Với mức ý nghĩa 0,01, hãy kết luận xem việc cải tiến kỷ thuật có hiệu quả hay không?

Giải:

....
....
....
....
....
....
....
....
....
....
....
....
....

Ví dụ 10. Thu nhập của công nhân X (triệu đồng) trong một năm của 1 phân xưởng của công ty A được thống kê như sau:

X	70 - 75	75 - 80	80 - 85	85 - 90	90 - 95
Số CN	6	18	28	40	16

Có thể cho rằng tỷ lệ công nhân có thu nhập thấp (là công nhân có thu nhập dưới 80 triệu đồng) của công ty A là trên 25% được không ? với mức ý nghĩa 6% .

Giải:

....
....
....
....
....
....
....
....
....
....
....
....
....
....

Ví dụ 11. Theo qui định tỷ lệ phế phẩm trong một lô hàng là không quá 0,02. Kiểm tra ngẫu nhiên 200 sản phẩm trong lô hàng thì thấy có 6 phế phẩm. Xét xem lô hàng có đúng qui định không? với mức ý nghĩa 0,03.

Giải:

....
....
....
....

....
....
....
....
....
....
....
....
....
....

Ví dụ 12. Một cơ sở bán hạt giống nói rằng tỷ lệ hạt đậu tương không nảy mầm ở đây dưới 10%. Gieo 300 hạt giống của cơ sở này thấy có 260 hạt nảy mầm. Với mức ý nghĩa 4% xem cơ sở nói có đúng không?

Giải:

....
....
....
....
....
....
....
....
....
....
....

Ví dụ 13. Một máy sản xuất tự động, lúc đầu tỷ lệ sản phẩm loại tốt là 30%. Sau khi áp dụng một phương pháp sản xuất mới, người ta lấy một số hộp (mỗi hộp gồm 10 sản phẩm) để kiểm tra. Kết quả kiểm tra cho ở bảng sau:

Số sp tốt/hộp	1	2	3	4	5	6	7	8	9
Số hộp	2	4	3	6	8	9	4	5	3

Với mức ý nghĩa 3 %. Hãy cho kết luận về phương pháp mới này.

Giải:

....
....
....
....
....
....
....
....

4.5 KIỂM ĐỊNH DÙNG P- VALUE

Phần trước đã trình bày phương pháp kiểm định truyền thống. Trong phần này sẽ xét một phương pháp khác hiện nay được các nhà thống kê sử dụng khá rộng rãi gọi là phương pháp *p – Value*.
p – value là mức ý nghĩa quan sát, là xác suất phạm sai lầm loại I tối đa khi bác bỏ giả thuyết H_0 với một mẫu dữ liệu đang quan sát.

1. Xét ví dụ: Giám đốc điều hành sản xuất của một nhà máy chế biến thực phẩm ăn liền đang quan tâm đến dây chuyền tự động đóng hộp ngũ cốc dinh dưỡng. Theo đúng quy định trọng lượng mỗi hộp ngũ cốc là 368 g và độ lệch chuẩn 15g. Ông ta nghi ngờ dây chuyền gặp trục trặc gì đó nên qui định trên không được đảm bảo. Ông ta chọn 25 hộp thì thấy trọng lượng trung bình là 372,5g. Với mức ý nghĩa 0,05, theo Anh (chị) ông ta kết luận như thế nào?
Giải

Giả thiết: H_0: $\mu = \mu_0 = 368$ H_1: $\mu \neq 368$

Giá trị kiểm định: $Z = \dfrac{(\bar{x} - \mu_0).\sqrt{n}}{\sigma} = \dfrac{(372,5 - 368).\sqrt{25}}{15} = 1,5$

$\alpha = 0,05 \Rightarrow z_{\alpha/2} = z_{0,025} = 1,96$

=> $z = 1,5 < z_{0,025} = 1,96$ => chấp nhận H_0 => dây chuyền hoạt động bình thường.

Ta thấy rằng khi α càng nhỏ thì giá trị tới hạn $z_{\alpha/2}$ càng lớn , do đó H_0 không thể bác bỏ ở mức ý nghĩa $\alpha = 0,05$, nên cũng không thể bác bỏ H_0 ở mức ý nghĩa nhỏ hơn 0,05. Vậy nếu mức ý nghĩa lớn hơn 0,05 thì sao? có bác bỏ được H_0 hay không?
Giá trị kiểm định tính được là 1,5 , nên H_0 sẽ bị bác bỏ ở bất kỳ mức ý nghĩa α nào mà $z_{\alpha/2} > 1,5$.(vì kiểm định 2 phía nên phải nhân hai khi tính α)

Vậy : $p - value = 2.P(z > 1,5) = 2.[1 - \Phi(1,5)] = 2(1 - 0,9332) = 0,1336$

<u>Mức ý nghĩa nhỏ nhất (ranh giới) mà ở đó giả thuyết H_0 sẽ bị bác bỏ gọi là P – Value.</u>

BÀI TOÁN TÍNH P-VALUE CỦA KIỂM ĐỊNH TRUNG BÌNH
Tính các tham số mẫu: $n;\ \bar{x};\ s$.

Giả thuyết kiểm định	Tính giá trị kiểm định	Tính giá trị p-value	Quy tắc kiểm định
$I: \begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}$		$p - value = 2\left[1 - \Phi(\lvert z \rvert)\right]$	Nếu $p - value \le \alpha$ thì bác bỏ H_0
$II: \begin{cases} H_0: \mu \le \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$	$z = \dfrac{(\bar{x} - \mu_0)\sqrt{n}}{s}$	$p - value = 1 - \Phi(z)$	
$III: \begin{cases} H_0: \mu \ge \mu_0 \\ H_1: \mu < \mu_0 \end{cases}$		$p - value = \Phi(z)$	

BÀI TOÁN TÍNH P-VALUE CỦA KIỂM ĐỊNH TỶ LỆ

Tính tỷ lệ mẫu: $f = \dfrac{m}{n}$.

Giả thuyết kiểm định	Tính giá trị kiểm định	Tính giá trị p-value	Quy tắc kiểm định
$I: \begin{cases} H_0: p = p_0 \\ H_1: p \neq p_0 \end{cases}$		$p-value = 2\left[1-\Phi\left(\lvert z\rvert\right)\right]$	Nếu $p-value \le \alpha$ thì bác bỏ H_0
$II: \begin{cases} H_0: p \le p_0 \\ H_1: p > p_0 \end{cases}$	$z = \dfrac{(f-p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}}$	$p-value = 1-\Phi(z)$	
$III: \begin{cases} H_0: p \ge p_0 \\ H_1: p < p_0 \end{cases}$		$p-value = \Phi(z)$	

2. Áp dụng

Từ một tổng thể có trung bình là μ chưa biết , lấy một mẫu kích thước n >30 ta tính được giá trị trung bình mẫu $\bar{x}$ và độ lệch chuẩn mẫu là s. Khi đó $\bar{x}$ là BNN có phân phối xấp xỉ chuẩn:

$$\bar{x} \sim N\left(\mu = \mu_0; \sigma^2_{\bar{x}} = \frac{s^2}{n}\right)$$

Ví dụ 1. Kiểm định giả thuyết $H_0: \mu = 368 = \mu_0$ **;** $H_1: \mu \neq 368$, với $\bar{x} = 372,5$; $\alpha = 0,05$ **;** $\sigma = 15$ **;** $\mu_0 = 368$ **;** $n = 25$ **.** Hãy tính P-value.

Giải:

....

....

....

....

Ví dụ 2. Từ một mẫu ta tính được $n = 36$; $\bar{x} = 5040$; $s = 780$, hãy dùng phương pháp p – value kiểm định giả thiết: H_0: $\mu = 4700$ H_1: $\mu > 4700$ với mức ý nghĩa 0,02.

Giải:

....

....

....

....

....

....

....

....

....

....

Ví dụ 3. Từ một mẫu ta tính được $n = 140$; $m = 54$, hãy sử dụng phương pháp p – value kiểm định giả thiết: H_0: $p = 0,4 = p_0$ H_1: $p < 0,4$ với mức ý nghĩa 0,05.

Giải:

....
....
....
....
....
....
....

Ví dụ 4. Một loại sản phẩm có chiều dài trung bình là 11 cm. Sau khi thực hiện một cải tiến kỷ thuật của máy sản xuất, người ta muốn biết cải tiến này có ảnh hưởng đến chiều dài sản phẩm hay không. Chọn ngẫu nhiên 50 sản phẩm do máy sản xuất thấy chiều dài trung bình là 10,3 cm và độ lệch chuẩn 2,1 cm. Hãy cho biết kết luận bằng cách sử dụng phương pháp kiểm định p – value giá trị, với mức ý nghĩa 4%.

Giải:

....
....
....
....
....
....
....
....
....
....

Ví dụ 5. Tốc độ xe trên đường cao tốc là biến ngẫu nhiên có phân phối chuẩn với trung bình là 80km/h và độ lệch chuẩn là 15 km. Người ta đã nâng cấp chất liệu mặt đường để xem có giúp xe chạy nhanh hơn không. Sau sau khi nâng cấp, người ta chọn ngẫu nhiên 100 xe để khảo sát và tính được tốc độ trung bình mẫu là 87 km/h. Hãy cho biết việc nâng cấp chất liệu mặt đường có giúp xe chạy nhanh hơn hay không? Yêu cầu xác định p-value của kiểm định và đưa ra kết luận với mức ý nghĩa 3%.

Giải.

....
....
....
....
....
....
....
....
....
....
....
....

Ví dụ 6. Tỷ lệ sản phẩm loại xấu của một nhà máy là 20%. Sau khi một phương pháp cải tiến kỷ thuật được áp dụng, người ta lấy 1000 sản phẩm kiểm tra thì thấy có 190 sản phẩm loại xấu. Với mức ý nghĩa 6%, hãy cho ý kiến về phương pháp cải tiến này bằng cách dùng kiểm định p-value.

Giải

....

....

....

....

....

....

....

....

....

....

....

BÀI TẬP CHƯƠNG 4

Câu 1: Tỷ lệ phế phẩm của 1 nhà máy trước đây là 5%. Năm nay nhà máy áp dụng một biện pháp kỷ thuật mới. Để nghiên cứu tác dụng của biện pháp kỷ thuật mới, người ta lấy 1 mẫu gồm 800 sản phẩm để kiểm tra và thấy có 24 phế phẩm. Với mức ý nghĩa 0,01. Hãy cho kết luận về biện pháp kỹ thuật mới này.

Câu 2: Để khảo sát chỉ tiêu X của một loại sản phẩm trong nhà máy, người ta quan sát một mẫu và có kết quả sau:

X(kg)	15	20	25	30	35	40	45
Số sp	8	13	15	12	14	6	13

Những sản phẩm có chỉ tiêu X từ 35kg trở lên được gọi là những sản phẩm loại A. Người ta nói tỷ lệ những sản phẩm loại A do nhà máy sản xuất ra là 48%, hãy cho kết luận với mức ý nghĩa 5%.

Câu 3: Để khảo sát chỉ tiêu X của một loại sản phẩm trong nhà máy, người ta quan sát một mẫu và có kết quả sau:

X(cm)	11-15	15-19	19-23	23-27	27-31	31-35	35-39
Số sản phẩm	8	9	20	16	16	13	18

Người ta kết luận giá trung bình của chỉ tiêu X của loại sản phẩm trên là 29cm thì có chấp nhận được không, với mức ý nghĩa 2%.

Câu 4: Giám đốc 1 xí nghiệp cho biết lương trung bình của 1 công nhân thuộc xí nghiệp hiện nay là 600 ngàn đồng / tháng. Chọn ngẫu nhiên 36 công nhân thấy lương trung bình là 520 ngàn đồng / tháng, và độ lệch chuẩn 40 ngàn đồng / tháng. Với mức ý nghĩa 5%, kết luận về khẳng định trên.

Câu 5: Một cửa hàng thực phẩm nhận thấy thời gian vừa qua trung bình 1 khách hàng mua 250 ngàn đồng thực phẩm trong ngày. Nay cửa hàng chọn ngẫu nhiên 55 khách hàng thấy trung bình 1 khách hàng mua 240 ngàn đồng trong ngày và phương sai là 40 . Với mức ý nghĩa 5%, hãy cho biết nhận xét ?

Câu 6: Một máy sản xuất tự động với tỷ lệ chính phẩm 98%. Sau 1 thời gian làm việc, người ta kiểm tra ngẫu nhiên 500 sản phẩm do nhà máy này chế tạo, thấy có 28 phế phẩm. Với mức ý nghĩa 0,05% hãy thử xem chất lượng làm việc của máy có còn như trước không?

Câu 7: Theo báo cáo thì tỷ lệ hộ dân thích xem thời sự trên Ti vi là 80%. Thăm dò 36 hộ dân thấy có 25 hộ thích xem thời sự. Với mức ý nghĩa 5%. Kiểm định xem nguồn tin này có đáng tin cậy không.

Câu 8: Tỷ lệ phế phẩm của 1 nhà máy trước đây là 5%. Năm nay nhà máy áp dụng một biện pháp kỷ thuật mới. Để nghiên cứu tác dụng của biện pháp kỷ thuật mới, người ta lấy 1 mẫu gồm 800 sản phẩm để kiểm tra và thấy có 24 phế phẩm. Nếu nhà máy báo cáo tỷ lệ phế phẩm sau khi áp dụng biện pháp kỹ thuật mới là 2% thì có chấp nhận được không, với mức ý nghĩa 0,05. Hãy dùng kiểm định p-value.

Câu 8: Để khảo sát chiều cao của một giống cây trồng, người ta quan sát một mẫu và có kết quả sau:

X(cm)	95-105	105-115	115-125	125-135	135-145	145-155	155-165
Số cây	10	10	30	10	10	10	15

Những cây có chiều cao từ 135cm trở lên được gọi là những cây cao. Người ta nói rằng tỷ lệ cây cao là 40%. hãy cho kết luận với mức ý nghĩa 5% . Hãy dùng kiểm định p-value.

Câu 9: Để khảo sát trọng lượng X của một loại vật nuôi trong nông trại, người ta quan sát một mẫu và có kết quả sau:

X(kg)	36	42	48	54	60	66	72
Số con	15	12	25	18	10	10	10

Những con vật có trọng lượng từ 60kg trở lên được gọi là những con đạt tiêu chuẩn. Người ta kết luận những con vật đạt tiêu chuẩn của loại vật nuôi trên là 38% thì có chấp nhận được không, với mức ý nghĩa 5%. Hãy dùng kiểm định p-value.

Câu 10: Trọng lượng X của trái quít có phân phối chuẩn. Cân thử 100 trái quít của một vườn quít , ta có bảng kết quả sau :

X(g)	40	50	60	70	80	90	100	110
Số trái	3	10	12	15	28	16	11	5

Những trái có trọng lượng lớn hơn 75g là trái loại I.Hãy cho kết luận về phát biểu"trọng lượng trung bình trái loại I trong vườn trên là 85g", với mức ý nghĩa 5%. Hãy dùng kiểm định p-value.

Câu 11: Để khảo sát trọng lượng X của một loại vật nuôi trong nông trại, người ta quan sát một mẫu và có kết quả sau:

X(kg)	28-32	32-36	36-40	40-44	44-48	48-52	52-56
Số con	10	10	15	30	10	10	15

Người ta nói rằng trọng lượng trung bình của loại vật nuôi trên là 45kg. Hãy cho kết luận, với mức ý nghĩa 1%. Hãy dùng kiểm định p-value.

Chương 5: **HỒI QUI TUYẾN TÍNH ĐƠN**

MỤC TIÊU

Sau khi học xong chương này sinh viên có thể:
- Thiết lập được mô hình hồi qui tuyến tính đơn giản của 2 biến định lượng.
- Nhận xét được mô hình có phù hợp hay không?
- Nêu được ý nghĩa của mô hình hồi qui đã thiết lập.
- Vận dụng được mô hình hồi qui để dự báo.

4.1. Khái niệm

Giả sử có hai biến biến định lượng *x* và *y*, trong đó biến *x* ảnh hưởng đến biến *y* hay *y* phụ thuộc vào *x*, ta cần tìm một mô hình toán học để mô hình hóa hiện tượng này. Ở đây ta chỉ xét mô hình tuyến tính đơn giản thể hiện mối liên hệ tuyến tính giữa 2 biến *x* và *y*. Mô hình này gọi là hàm hồi qui của *y* theo *x*. Hàm hồi qui được hiểu theo nghĩa thể hiện mối liên hệ phụ thuộc thống kê, chứ không phải phụ thuộc hàm số như trong toán học. Tức là trong hàm hồi qui biến *x* là biến giải thích sự phụ thuộc của biến *y* với yếu tố đang xem xét, ngoài ra còn nhiều yếu tố khác tác động đến *y* không xét đến trong hàm hồi qui.

4.2 Mô hình hồi qui và phương trình hồi quy

Giả sử có 2 biến trong đó *y* phụ thuộc *x* bởi mối phụ thuộc tuyến tính.
- *y* – biến phụ thuộc (dependent variable), là biến liên tục.
- *x* – biến độc lập ((independent variable), là biến liên tục hay không liên tục.

Mô hình hồi quy tuyến tính đơn: $y = \underbrace{\beta_0 + \beta_1 x}_{\text{Thành phân tuyên tính}} + \underbrace{\varepsilon}_{\text{Thành phân ngâu nhiên}}$

Công thức này gọi là mô hình hồi qui tuyến tính đơn biến tổng thể

Trong đó:
- *y* được gọi là biến phụ thuộc (hay biến được giải thích)
- *x* được gọi là biến hồi quy (hay biến độc lập, biến giải thích)
- β_0, β_1 là các tham số của mô hình
- Mô hình được gọi là tuyến tính vì nó tuyến tính với các tham số β_0, β_1 (có lũy thừa 1)
- Mô hình được gọi là **đơn** vì có một biến độc lập.
- ε (epsilon) là một biến ngẫu nhiên gọi là sai số, thể hiện biến thiên của *y* mà không thể giải thích bởi mối liên hệ tuyến tính giữa *x* và *y*.

Ta thấy rằng với mỗi giá trị *x* cho trước có thể xác định nhiều giá trị khác nhau của biến *y* vì còn nhiều yếu tố khác tác động đến biến *y* ngoài *x*. Do đó tại cùng một giá trị *x* sẽ có nhiều sai số khác nhau.

Phương trình hồi quy tuyến tính đơn: $E(y|x) = \beta_0 + \beta_1 x$

Đồ thị của phương trình hồi quy tuyến tính đơn (của tổng thể) là một đường thẳng cắt trục tung tại β_0 và có hệ số góc là β_1. $E(y|x)$ là giá trị trung bình của *y* với điều kiện *x*, nghĩa là giá trị đại diện cho nhiều giá trị *y* cùng xảy ra ở một giá trị *x* cụ thể.

Mô hình hồi qui được trình bày trên đây phải thỏa các giả thuyết (giả định) sau:

1. *Sai số* ε *là một biến ngẫu nhiên có trung bình bằng 0, tức là:* $E(\varepsilon) = 0$.

2. *Phương sai của* ε *là như nhau cho tất cả các giá trị của x.*
3. *Các giá trị của* ε *độc lập với nhau.*
4. *Sai số ngẫu nhiên* ε có *phân phối chuẩn.*

Ví dụ 2: Mức chi tiêu *y* của một người sẽ phụ thuộc vào mức thu nhập *x* của người đó. Tuy nhiên ta thấy rằng 2 người có cùng mức thu nhập thì chi tiêu vẫn khác nhau là do ảnh hưởng bởi các yếu tố khác như: hoàn cảnh gia đình, sở thích, thói quen tiêu dùng,..vv..

4.3 Phương trình hồi quy ước lượng (phương trình hồi quy mẫu)

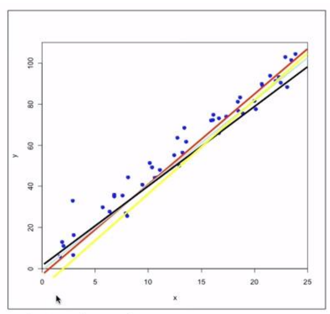
Trong thực tế ta không thể xác định được các tham số β_0, β_1 của phương trình hồi qui tuyến tính, mà chỉ có thể ước lượng chúng từ các giá trị quan sát được của mẫu.
Với dữ liệu là mẫu thu thập được, phương trình hồi qui ước lượng thể hiện mối liên hệ tuyến tính của biến *y* đối với biến *x* là:

$$\hat{y} = a + bx$$

Các hệ số *a, b* là các tham số mẫu, chúng lần lượt là ước lượng của tham số tổng thể β_0, β_1.

4.3.1 Phương pháp bình phương nhỏ nhất (OLS)

Chúng ta tìm hiểu phương pháp bình phương nhỏ nhất để tính toán các hệ số a, b trong phương trình hồi quy ước lượng như sau.
Giả sử ta có mẫu n cặp giá trị quan sát được $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ từ *x* và *y*.
Ta mong muốn là tìm được các giá trị a, b thích hợp nhất để ước lượng cho β_0, β_1.



Phương trình hồi quy ước lượng (đường thẳng) $\hat{y} = a + bx$ được xem là tốt nhất khi tổng bình phương các chênh lệch giữa giá trị thực tế *y* với giá trị ước lượng $\hat{y}$ là nhỏ nhất, tức là:

$$\sum d_i^2 = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - a - bx_i)^2 = \min \quad (1)$$

Phương pháp này gọi là phương pháp bình phương nhỏ nhất.
Điều kiện (1) là bài toán cực tiểu của hàm 2 biến, người ta chứng minh được kết quả sau:

➢ **phương trình hồi qui tuyến tính mẫu của *y* theo *x* :** $\hat{y} = a + bx$

trong đó:

$$b = \frac{n.\sum xy - \sum x.\sum y}{n.\sum x^2 - \left(\sum x\right)^2}, \qquad a = \frac{\sum y - b.\sum x}{n}$$

4.3.2 Ý nghĩa các hệ số trong hồi qui

Trong phương trình hồi qui: $\hat{y} = a + bx$ thì *a*, *b* gọi là các hệ số hồi qui, *a* gọi là hệ số chặn, hệ số tự do hay là tung độ gốc, *b* gọi là hệ số góc hay độ dốc.

Nếu *x* tăng lên 1 đơn vị (của *x*) thì trung bình của *y* sẽ tăng (nếu b > 0), giảm (nếu b < 0) |*b*| đơn vị (của *y*).

Ví dụ 3. Biến *y* doanh số bán hàng (triệu đ/tháng) phụ thuộc vào *x* chi phí quảng cáo (triệu đ) như sau: $\hat{y} = 8x + 7$ thì ý nghĩa của hệ số góc b = 8 là:
Nếu chi phí quảng cáo tăng lên 1 triệu đồng thì doanh số bán hàng trung bình tăng lên 8 triệu đồng / tháng.

4.4 Hệ số xác định

Giả sử phương trình hồi qui tuyến tính ước lượng là $\hat{y} = a + bx$ thì khi thay giá trị quan sát thứ i là x_i vào phương trình hồi qui ta được giá trị tương ứng $\hat{y}_i$ là **giá trị dự báo** hay **giá trị lý thuyết** của biến phụ thuộc.

Xét các tổng bình phương sau:

Sự khác biệt giữa giá trị thực tế y_i và giá trị lý thuyết $\hat{y}_i$, tức là $y_i - \hat{y}_i$, được gọi là phần dư thứ *i*. Tổng bình phương các phần dư (hay sai số) này được ký hiệu là SSE và gọi là ***tổng bình phương do sai số***: $SSE = \sum\left(y_i - \hat{y}_i\right)^2$. Đại lượng SSE thể hiện sự biến thiên của *y* do các yếu tố khác không nghiên cứu trong mô hình.

Sự khác biệt giữa giá trị thực tế y_i và giá trị trung bình $\overline{y}$, tức là $y_i - \overline{y}$, thể hiện sai số trong việc sử dụng $\overline{y}$ để ước tính y_i. Tổng bình phương các sai số này được ký hiệu là SST và gọi là ***tổng bình phương toàn bộ***: $SST = \sum\left(y_i - \overline{y}\right)^2$. Đại lượng SST thể hiện toàn bộ biến thiên của *y*.

Để đo lường các giá trị lý thuyết $\hat{y}_i$ khác biệt với giá trị trung bình $\overline{y}$, tức là $\hat{y}_i - \overline{y}$, ***tổng bình do phương trình hồi quy*** được ký hiệu SSR và tính như sau: $SSR = \sum\left(\hat{y}_i - \overline{y}\right)^2$. Đại lượng: SSR thể hiện sự biến thiên của *y* được giải thích bởi biến *x*.

Bằng các phép biến đổi toán học ta được: $SST = SSR + SSE \Leftrightarrow 1 = \dfrac{SSR}{SST} + \dfrac{SSE}{SST}$

Tỷ lệ $\dfrac{SSR}{SST}$, có giá trị giữa 0 và 1, được sử dụng để đánh giá **sự phù hợp của phương trình hồi quy**. Tỷ lệ này được gọi là ***hệ số xác định*** và ký hiệu là R^2.

$$R^2 = \frac{SSR}{SST}$$

Ý nghĩa của hệ số xác định: *Hệ số xác định R^2 nói lên tỷ lệ % biến thiên của biến y được giải thích bởi biến x .*

Ví dụ 4. Nếu hệ số xác định là $R^2 = 0{,}87$ thì trong phương trình hồi qui mẫu biến *x* giải thích được 87% sự thay đổi của biến phụ thuộc *y*, còn 13% sự thay đổi còn lại của *y* do các yếu tố ngẫu nhiên gây ra.

Chú ý: Ta có thể sử dụng các công thức tính các tổng bình phương như sau:

$$SST = \sum y_i^2 - \frac{\left(\sum y_i\right)^2}{n}, \qquad SSR = \mathrm{b}\left(\sum x_i y_i - \frac{\sum x_i . \sum y_i}{n}\right), \qquad SSE = \sum y_i^2 - b.\sum\left(x_i y_i\right) - a.\sum y_i$$

4.5 Sai số chuẩn của ước lượng – Standard Error

Dùng phương pháp bình phương bé nhất OLS ta được một đường thẳng phù hợp nhất với tập dữ liệu, nhưng không phải tất cả các điểm quan sát thực tế đều nằm trên đường thẳng này. Do

đó cần phải tính toán một đại lượng thống kê đo lường sự chênh lệch của giá trị thực tế *y* và giá trị lý thuyết $\hat{y}$ do đường hồi qui tính toán ra. Cũng giống như khi tính độ lệch chuẩn đo lường sự biến thiên của mỗi giá trị quan sát xung quanh trị trung bình của nó. Độ lệch chuẩn xung quanh đường hồi qui được gọi là ***sai số chuẩn*** của hồi qui. Vì trong phương trình hồi qui $\hat{y} = a + bx$ có 2 tham số được ước lượng là a và b nên trong phép tính sai số chuẩn s_e bậc tự do sẽ giảm đi 2 còn (n – 2).
Sai số chuẩn hồi qui là một đại lượng dùng để đánh giá mức độ chính xác của ước lượng. Nếu sai số chuẩn bé thì số liệu thực nghiệm tập trung gần đường hồi qui nên dự báo chính xác cao. Ngược lại, sai số chuẩn lớn, số liệu thực nghiệm phân tán xa đường hồi qui nên dự báo kém chính xác.

Công thức tính sai số chuẩn của ước lượng: $s_e = \sqrt{\dfrac{SSE}{n-2}}$

4.6 Dùng phương trình hồi qui để ước lượng (dự báo)

- Ước lượng điểm: Khi biết $x = x_0$ thì dự báo trung bình của *y* là: $\hat{y}_0 = a + bx_0$.

- Ước lượng khoảng:
  Khi biết $x = x_0$ thì dự báo khoảng **trung bình** của *y*, với độ tin cậy γ là:
  $$y = \hat{y}_0 \pm t_{n-2;\alpha/2}.s_e.\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}}, \quad \text{trong đó: } \hat{y}_0 = a + bx_0.$$
  Khi biết $x = x_0$ thì dự báo khoảng của *y*, với độ tin cậy γ là:
  $$y = \hat{y}_0 \pm t_{n-2;\alpha/2}.s_e.\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}}, \quad \text{trong đó: } \hat{y}_0 = a + bx_0.$$

Ví dụ 5. Ta có số liệu sau về GDP *x* (ngàn tỷ đồng) và tiêu dùng cá nhân *y* (ngàn tỷ đồng) ở Việt Nam trong 9 năm như sau:
$\sum x_i = 232; \sum x_i^2 = 6152$; $\sum y_i = 158$; $\sum y_i^2 = 2824$; $\sum x_i y_i = 4165$
Hãy lập phương trình hồi qui của *y* theo *x*.
Giải:
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………

Ví dụ 6. Lập phương trình hồi qui của lượng sản phẩm bán được (1000 sản phẩm) theo số phút quảng cáo sản phẩm của mẫu sau:

Số phút quảng cáo	36	45	37	28	30	34	33	32
lượng sản phẩm bán được	41	37	32	25	22	37	31	35

Dựa vào hệ số góc của phương trình hồi qui hãy cho nhận xét.
Giải:

……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………

Ví dụ 7. Khảo sát một mẫu 10 sinh viên ghi nhận số giờ ôn tập môn Tiếng Anh chuyên nghành và điểm thi môn này như sau:

Điểm	56	43	59	53	74	83	84	84	96	69
Số giờ	9	8	7	6	5	19	15	11	14	15

Hãy lập phương trình hồi quy của điểm thi theo số giờ ôn tập và dựa vào hệ số góc của phương trình hãy cho nhận xét.
Giải:

……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………

Ví dụ 8. Cho số liệu sau về mức chi tiêu tiêu dùng (Y-USD/tuần) và thu nhập hàng tuần (X-USD/tuần) của 15 gia đình ở thành phố A như sau:

$$\sum x_i = 1700; \sum x_i^2 = 322000\ ; \sum y_i = 1110; \sum x_i y_i = 205500$$

Hãy lập phương trình hồi qui của Y theo X và cho nhận xét về hệ số góc của phương trình tìm được.
Giải:

……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………

Ví dụ 9. Xét mẫu số liệu sau:

Thu nhập X (triệu đ/tháng)	8	9	10	11	12	15	15	16	17	20
Chi tiêu Y (triệu đ/tháng)	6	8	9	12	10	12	11	13	15	14

1. Lập phương trình hồi qui tuyến tính mô tả sự phụ thuộc của chi tiêu vào thu nhập.
2. Hãy dự báo chi tiêu khi thu nhập là 14 triệu đ/tháng.

Giải:

…………………………………………………………………………………………
…………………………………………………………………………………………
………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
………………………………………………………………………
……………………………………………………………………………………
…………………………………………………………………………………………

Ví dụ 10. Một người bán lẻ tin rằng có một mối liên hệ tuyến tính giữa giá của 1 kg Đường x (ngàn đồng/1kg) với lượng cầu hàng tuần y (kg). Người này thu thập dữ liệu và tổng kết lại như sau:

$\sum x = 101,4; \sum y = 869; \sum x^2 = 653,88; \sum y^2 = 48799; \sum xy = 5375,7\ ; n = 20$

1. Viết phương trình hồi quy tuyến tính ước lượng cầu theo giá đường, giải thích ý nghĩa hệ số góc của phương trình.
2. Hãy dự báo lượng cầu trung bình với giá 4,5 ngàn đồng /kg.

Giải:

…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………

Ví dụ 11. Giả sử có số liệu thống kê về lãi suất ngân hàng (% năm) và tổng vốn đầu tư (tỷ đồng) trên địa bàn tỉnh A qua 10 năm liên tiếp như sau:

Lãi suất ngân hàng	7	6,5	6,5	6	6	6	5,5	5,5	5	4,5
Tổng vốn đầu tư	38	32	30	34	34	35	40	42	48	50

1. Lập phương trình hồi qui tuyến tính của tổng vốn đầu tư theo lãi suất ngân hàng. Nêu ý nghĩa của hệ các số hồi qui tìm được.
2. Hãy ước lượng tổng vốn đầu tư trung bình khi lãi suất 6% năm.
Giải:

……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………

Ví dụ 12 Với một mẫu gồm 30 quan sát, một người phân tích tài chính muốn hồi qui tỷ suất sinh lợi của cổ phiếu thường (Y%) của một công ty theo tỷ suất sinh lợi của chỉ số Standar (X%). Có sẵn thông tin sau đây:

$$\sum_{i=1}^{20} y_i = 22,6; \sum_{i=1}^{20} x_i = 25,4; \sum_{i=1}^{20} y_i^2 = 196,2; \sum_{i=1}^{20} x_i^2 = 145,7; \sum_{i=1}^{20} x_i y_i = 150,5$$

1. Viết phương trình hồi qui tuyến tính ước lượng của Y theo X. Nêu ý nghĩa hệ số góc của phương trình hồi qui tuyến tính ước lượng.
2. Tìm tỷ suất sinh lợi kỳ vọng của cổ phiếu công ty khi tỷ suất sinh lợi của chỉ số Standar là 1%.
Giải:

……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………

Ví dụ 13. Câu lạc bộ thể thao Dáng Chuẩn chọn 8 người của câu lạc bộ ghi lại tổng số thời gian tập (giờ) aerobic trong suốt một tuần và số cân nặng giảm như sau:

Số cân nặng giảm	0,6	1,8	1,4	1,4	1,6	2,8	2,3	1,5
Tổng thời gian tập	4	9	7	11	8	10	9	8

1. Tìm phương trình hồi qui tuyến tính miêu tả mối phụ thuộc trọng lượng giảm theo lượng thời gian tập earobic của mẫu trên và nêu ý nghĩa hệ số góc của phương trình.
2. Sử dụng phương trình hồi qui này để dự báo số cân giảm được của một thành viên câu lạc bộ khi họ tập 7 giờ earobic mỗi tuần.
Giải:

……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………

Ví dụ 14. Mối liên hệ giữa độ tươi của những quả Đào và thời gian từ lúc thu hoạch chúng, được nghiên cứu bằng cách chọn ngẫu nhiên 12 quả Đào từ một vườn Đào. Cứ mỗi 2 ngày, 2 quả được chọn ngẫu nhiên và đo lường độ tươi dựa trên một thước đo 10 điểm. Quả nào đạt 10 điểm là tươi, quả nào 0 điểm là đã bị thối. Kết quả thu được:

Ngày thứ	0	2	4	6	8	10
Độ tươi	9,4 - 9,8	8,2 - 8,9	7,0 - 7,2	5,8 - 6,8	4,6 - 5,0	3,5 - 4,1

1. Hãy xác định phương trình hồi qui tuyến tính cho độ tươi của Đào phụ thuộc vào thời gian từ lúc thu hoạch. Nêu ý nghĩa hệ số góc của phương trình.
2. Dự đoán độ tươi của một quả Đào sau 3 ngày thu hoạch.
Giải:

……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………

***Dùng máy tính 570ES, 500ES tính hệ số A, B của phương* trình hồi quy:Y = A + BX**

Ví dụ: Lập phương trình hồi qui cho mẫu số liệu sau:

X	2	4	3	2	4
Y	24	34	28	27	37

* Bấm 3 phím: **MODE** **3(STAT)** **2(A+BX)**

*Nhập dữ liệu:

Nhập dữ liệu cho cột X:

2 =
4 =
3 =
2 =
4 =

Dùng phím **RELAY (▶)** chuyển dấu nháy sang cột Y và dùng phím **RELAY (▲)** di chuyển dấu nháy lên đầu hàng để nhập dữ liệu cho cột Y:

24 =
34 =
28 =
27 =
37 =

* Nhập xong toàn bộ dữ liệu bấm phím: **ON**

* Tính hệ số A: **SHIFT** 1 **7(Reg)** **1(A)** =

Kết quả: **A = 15**

* Tính hệ số góc B: **SHIFT** 1 **7(Reg)** **2(B)** =

Kết quả: **B =15**

* Thoát khỏi chế độ tính thống kê bấm: **MODE** **1**

Dùng máy tính 580:

Menu -> 6 -> 2 -> Nhập x, y -> Optn -> 4

Bài tập

Bài 1: Có số liệu về giá trị sản xuất (Y) và số công nhân (X) của các phân xưởng trong một tuần của nhà máy như sau:

X	44	47	48	48	43	46	42
Y	410	600	500	600	450	500	400

1.Viết phương trình hồi qui ước lượng của Y theo X, giải thích ý nghĩa hệ số góc của phương trình.
2. Dự đoán giá trị sản xuất của một phân xưởng 550 công nhân.

Bài 2: Xem xét mối liên quan giữa diện tích X và doanh số kinh doanh Y của cùng một mặt hàng tại 10 của hàng, ta có số liệu sau:

DT(m^2) X	7	10	8	5	11	3	7	11	12	6
DS (tr) Y	2	3	2,5	2	3,2	1,5	2,1	3,8	4	2,5

1. Tìm phương trình hồi qui.
2. Dự đoán Y trong trường hợp X=9.

Bài 3: Sau một tuần quảng cáo một loại sản phẩm, công ty A có số liệu sau:

Thời gian quảng cáo (giây)	25	18	32	21	35	28	30
Doanh số bán (tr)	16	11	20	15	26	32	20

1. Tìm phương trình hồi qui.
2. Nêu ý nghĩa hệ số góc của phương trình hồi quy.
3. Dự đoán doanh số trong trường hợp quảng cáo 20 giây.

Bài 4: Đo chiều cao X (cm) và đường kính Y(cm) của một loại cây, ta có số liệu sau:

X	28	28	30	60	30	32	42	43	49
Y	5	6	6	10	5	7	8	9	10

1. Lập phương trình hồi qui Y theo X.
2. Biết X = 50 hãy dự đoán Y.

Bài 5: Có số liệu cho biết tỷ lệ ngân sách chi cho ngành giáo dục X (%), tỷ lệ tăng thu nhập quốc dân Y (%) của một số nước như sau:

X	8	10	12	9	14	15	12	11	14	13
Y	4	5	6	5	8	9	7	6	8	7

1. Lập phương trình hồi qui Y theo X.
2. Nêu ý nghĩa hệ số góc của phương trình hồi quy.

Bảng phân phối chuẩn tắc của giá trị dương

z	Φ(z)	z	Φ(z)	z	Φ(z)	z	Φ(z)	z	Φ(z)	z	Φ(z)	z	Φ(z)
0.00	0.5000	**0.50**	0.6915	**1.00**	0.8413	**1.50**	0.9332	**2.00**	0.9772	**2.50**	0.9938	**3.00**	0.9987
0.01	0.5040	**0.51**	0.6950	**1.01**	0.8438	**1.51**	0.9345	**2.01**	0.9778	**2.51**	0.9940	**3.01**	0.9987
0.02	0.5080	**0.52**	0.6985	**1.02**	0.8461	**1.52**	0.9357	**2.02**	0.9783	**2.52**	0.9941	**3.02**	0.9987
0.03	0.5120	**0.53**	0.7019	**1.03**	0.8485	**1.53**	0.9370	**2.03**	0.9788	**2.53**	0.9943	**3.03**	0.9988
0.04	0.5160	**0.54**	0.7054	**1.04**	0.8508	**1.54**	0.9382	**2.04**	0.9793	**2.54**	0.9945	**3.04**	0.9988
0.05	0.5199	**0.55**	0.7088	**1.05**	0.8531	**1.55**	0.9394	**2.05**	0.9798	**2.55**	0.9946	**3.05**	0.9989
0.06	0.5239	**0.56**	0.7123	**1.06**	0.8554	**1.56**	0.9406	**2.06**	0.9803	**2.56**	0.9948	**3.06**	0.9989
0.07	0.5279	**0.57**	0.7157	**1.07**	0.8577	**1.57**	0.9418	**2.07**	0.9808	**2.57**	0.9949	**3.07**	0.9989
0.08	0.5319	**0.58**	0.7190	**1.08**	0.8599	**1.58**	0.9429	**2.08**	0.9812	**2.58**	0.9951	**3.08**	0.9990
0.09	0.5359	**0.59**	0.7224	**1.09**	0.8621	**1.59**	0.9441	**2.09**	0.9817	**2.59**	0.9952	**3.09**	0.9990
0.10	0.5398	**0.60**	0.7257	**1.10**	0.8643	**1.60**	0.9452	**2.10**	0.9821	**2.60**	0.9953	**3.10**	0.9990
0.11	0.5438	**0.61**	0.7291	**1.11**	0.8665	**1.61**	0.9463	**2.11**	0.9826	**2.61**	0.9955	**3.11**	0.9991
0.12	0.5478	**0.62**	0.7324	**1.12**	0.8686	**1.62**	0.9474	**2.12**	0.9830	**2.62**	0.9956	**3.12**	0.9991
0.13	0.5517	**0.63**	0.7357	**1.13**	0.8708	**1.63**	0.9484	**2.13**	0.9834	**2.63**	0.9957	**3.13**	0.9991
0.14	0.5557	**0.64**	0.7389	**1.14**	0.8729	**1.64**	0.9495	**2.14**	0.9838	**2.64**	0.9959	**3.14**	0.9992
0.15	0.5596	**0.65**	0.7422	**1.15**	0.8749	**1.65**	0.9505	**2.15**	0.9842	**2.65**	0.9960	**3.15**	0.9992
0.16	0.5636	**0.66**	0.7454	**1.16**	0.8770	**1.66**	0.9515	**2.16**	0.9846	**2.66**	0.9961	**3.16**	0.9992
0.17	0.5675	**0.67**	0.7486	**1.17**	0.8790	**1.67**	0.9525	**2.17**	0.9850	**2.67**	0.9962	**3.17**	0.9992
0.18	0.5714	**0.68**	0.7517	**1.18**	0.8810	**1.68**	0.9535	**2.18**	0.9854	**2.68**	0.9963	**3.18**	0.9993
0.19	0.5753	**0.69**	0.7549	**1.19**	0.8830	**1.69**	0.9545	**2.19**	0.9857	**2.69**	0.9964	**3.19**	0.9993
0.20	0.5793	**0.70**	0.7580	**1.20**	0.8849	**1.70**	0.9554	**2.20**	0.9861	**2.70**	0.9965	**3.20**	0.9993
0.21	0.5832	**0.71**	0.7611	**1.21**	0.8869	**1.71**	0.9564	**2.21**	0.9864	**2.71**	0.9966	**3.21**	0.9993
0.22	0.5871	**0.72**	0.7642	**1.22**	0.8888	**1.72**	0.9573	**2.22**	0.9868	**2.72**	0.9967	**3.22**	0.9994
0.23	0.5910	**0.73**	0.7673	**1.23**	0.8907	**1.73**	0.9582	**2.23**	0.9871	**2.73**	0.9968	**3.23**	0.9994
0.24	0.5948	**0.74**	0.7704	**1.24**	0.8925	**1.74**	0.9591	**2.24**	0.9875	**2.74**	0.9969	**3.24**	0.9994
0.25	0.5987	**0.75**	0.7734	**1.25**	0.8944	**1.75**	0.9599	**2.25**	0.9878	**2.75**	0.9970	**3.25**	0.9994
0.26	0.6026	**0.76**	0.7764	**1.26**	0.8962	**1.76**	0.9608	**2.26**	0.9881	**2.76**	0.9971	**3.26**	0.9994
0.27	0.6064	**0.77**	0.7794	**1.27**	0.8980	**1.77**	0.9616	**2.27**	0.9884	**2.77**	0.9972	**3.27**	0.9995
0.28	0.6103	**0.78**	0.7823	**1.28**	0.8997	**1.78**	0.9625	**2.28**	0.9887	**2.78**	0.9973	**3.28**	0.9995
0.29	0.6141	**0.79**	0.7852	**1.29**	0.9015	**1.79**	0.9633	**2.29**	0.9890	**2.79**	0.9974	**3.29**	0.9995
0.30	0.6179	**0.80**	0.7881	**1.30**	0.9032	**1.80**	0.9641	**2.30**	0.9893	**2.80**	0.9974	**3.30**	0.9995
0.31	0.6217	**0.81**	0.7910	**1.31**	0.9049	**1.81**	0.9649	**2.31**	0.9896	**2.81**	0.9975	**3.31**	0.9995
0.32	0.6255	**0.82**	0.7939	**1.32**	0.9066	**1.82**	0.9656	**2.32**	0.9898	**2.82**	0.9976	**3.32**	0.9995
0.33	0.6293	**0.83**	0.7967	**1.33**	0.9082	**1.83**	0.9664	**2.33**	0.9901	**2.83**	0.9977	**3.33**	0.9996
0.34	0.6331	**0.84**	0.7995	**1.34**	0.9099	**1.84**	0.9671	**2.34**	0.9904	**2.84**	0.9977	**3.34**	0.9996
0.35	0.6368	**0.85**	0.8023	**1.35**	0.9115	**1.85**	0.9678	**2.35**	0.9906	**2.85**	0.9978	**3.35**	0.9996
0.36	0.6406	**0.86**	0.8051	**1.36**	0.9131	**1.86**	0.9686	**2.36**	0.9909	**2.86**	0.9979	**3.36**	0.9996
0.37	0.6443	**0.87**	0.8078	**1.37**	0.9147	**1.87**	0.9693	**2.37**	0.9911	**2.87**	0.9979	**3.37**	0.9996
0.38	0.6480	**0.88**	0.8106	**1.38**	0.9162	**1.88**	0.9699	**2.38**	0.9913	**2.88**	0.9980	**3.38**	0.9996
0.39	0.6517	**0.89**	0.8133	**1.39**	0.9177	**1.89**	0.9706	**2.39**	0.9916	**2.89**	0.9981	**3.39**	0.9997
0.40	0.6554	**0.90**	0.8159	**1.4**	0.9192	**1.90**	0.9713	**2.40**	0.9918	**2.90**	0.9981	**3.40**	0.9997
0.41	0.6591	**0.91**	0.8186	**1.41**	0.9207	**1.91**	0.9719	**2.41**	0.9920	**2.91**	0.9982	**3.41**	0.9997
0.42	0.6628	**0.92**	0.8212	**1.42**	0.9222	**1.92**	0.9726	**2.42**	0.9922	**2.92**	0.9982	**3.42**	0.9997
0.43	0.6664	**0.93**	0.8238	**1.43**	0.9236	**1.93**	0.9732	**2.43**	0.9925	**2.93**	0.9983	**3.43**	0.9997
0.44	0.6700	**0.94**	0.8264	**1.44**	0.9251	**1.94**	0.9738	**2.44**	0.9927	**2.94**	0.9984	**3.44**	0.9997
0.45	0.6736	**0.95**	0.8289	**1.45**	0.9265	**1.95**	0.9744	**2.45**	0.9929	**2.95**	0.9984	**3.45**	0.9997
0.46	0.6772	**0.96**	0.8315	**1.46**	0.9279	**1.96**	0.9750	**2.46**	0.9931	**2.96**	0.9985	**3.46**	0.9997
0.47	0.6808	**0.97**	0.8340	**1.47**	0.9292	**1.97**	0.9756	**2.47**	0.9932	**2.97**	0.9985	**3.47**	0.9997
0.48	0.6844	**0.98**	0.8365	**1.48**	0.9306	**1.98**	0.9761	**2.48**	0.9934	**2.98**	0.9986	**3.48**	0.9997
0.49	0.6879	**0.99**	0.8389	**1.49**	0.9319	**1.99**	0.9767	**2.49**	0.9936	**2.99**	0.9986	**3.49**	0.9998

Bảng phân phối chuẩn tắc của giá trị âm

z	Φ(z)	z	Φ(z)	z	Φ(z)	z	Φ(z)	z	Φ(z)	z	Φ(z)	z	Φ(z)
0.00	0.5000	**-0.50**	0.3085	**-1.00**	0.1587	**-1.50**	0.0668	**-2.00**	0.0228	**-2.50**	0.0062	**-3.00**	0.0013
-0.01	0.4960	**-0.51**	0.3050	**-1.01**	0.1562	**-1.51**	0.0655	**-2.01**	0.0222	**-2.51**	0.0060	**-3.01**	0.0013
-0.02	0.4920	**-0.52**	0.3015	**-1.02**	0.1539	**-1.52**	0.0643	**-2.02**	0.0217	**-2.52**	0.0059	**-3.02**	0.0013
0.03	0.4880	**-0.53**	0.2981	**-1.03**	0.1515	**-1.53**	0.0630	**-2.03**	0.0212	**-2.53**	0.0057	**-3.03**	0.0012
-0.04	0.4840	**-0.54**	0.2946	**-1.04**	0.1492	**-1.54**	0.0618	**-2.04**	0.0207	**-2.54**	0.0055	**-3.04**	0.0012
-0.05	0.4801	**-0.55**	0.2912	**-1.05**	0.1469	**-1.55**	0.0606	**-2.05**	0.0202	**-2.55**	0.0054	**-3.05**	0.0011
-0.06	0.4761	**-0.56**	0.2877	**-1.06**	0.1446	**-1.56**	0.0594	**-2.06**	0.0197	**-2.56**	0.0052	**-3.06**	0.0011
-0.07	0.4721	**-0.57**	0.2843	**-1.07**	0.1423	**-1.57**	0.0582	**-2.07**	0.0192	**-2.57**	0.0051	**-3.07**	0.0011
-0.08	0.4681	**-0.58**	0.2810	**-1.08**	0.1401	**-1.58**	0.0571	**-2.08**	0.0188	**-2.58**	0.0049	**-3.08**	0.0010
-0.09	0.4641	**-0.59**	0.2776	**-1.09**	0.1379	**-1.59**	0.0559	**-2.09**	0.0183	**-2.59**	0.0048	**-3.09**	0.0010
-0.10	0.4602	**-0.60**	0.2743	**-1.10**	0.1357	**-1.60**	0.0548	**-2.10**	0.0179	**-2.60**	0.0047	**-3.10**	0.0010
-0.11	0.4562	**-0.61**	0.2709	**-1.11**	0.1335	**-1.61**	0.0537	**-2.11**	0.0174	**-2.61**	0.0045	**-3.11**	0.0009
-0.12	0.4522	**-0.62**	0.2676	**-1.12**	0.1314	**-1.62**	0.0526	**-2.12**	0.0170	**-2.62**	0.0044	**-3.12**	0.0009
-0.13	0.4483	**-0.63**	0.2643	**-1.13**	0.1292	**-1.63**	0.0516	**-2.13**	0.0166	**-2.63**	0.0043	**-3.13**	0.0009
-0.14	0.4443	**-0.64**	0.2611	**-1.14**	0.1271	**-1.64**	0.0505	**-2.14**	0.0162	**-2.64**	0.0041	**-3.14**	0.0008
-0.15	0.4404	**-0.65**	0.2578	**-1.15**	0.1251	**-1.65**	0.0495	**-2.15**	0.0158	**-2.65**	0.0040	**-3.15**	0.0008
-0.16	0.4364	**-0.66**	0.2546	**-1.16**	0.1230	**-1.66**	0.0485	**-2.16**	0.0154	**-2.66**	0.0039	**-3.16**	0.0008
-0.17	0.4325	**-0.67**	0.2514	**-1.17**	0.1210	**-1.67**	0.0475	**-2.17**	0.0150	**-2.67**	0.0038	**-3.17**	0.0008
-0.18	0.4286	**-0.68**	0.2483	**-1.18**	0.1190	**-1.68**	0.0465	**-2.18**	0.0146	**-2.68**	0.0037	**-3.18**	0.0007
-0.19	0.4247	**-0.69**	0.2451	**-1.19**	0.1170	**-1.69**	0.0455	**-2.19**	0.0143	**-2.69**	0.0036	**-3.19**	0.0007
-0.20	0.4207	**-0.70**	0.2420	**-1.20**	0.1151	**-1.70**	0.0446	**-2.20**	0.0139	**-2.70**	0.0035	**-3.20**	0.0007
-0.21	0.4168	**-0.71**	0.2389	**-1.21**	0.1131	**-1.71**	0.0436	**-2.21**	0.0136	**-2.71**	0.0034	**-3.21**	0.0007
-0.22	0.4129	**-0.72**	0.2358	**-1.22**	0.1112	**-1.72**	0.0427	**-2.22**	0.0132	**-2.72**	0.0033	**-3.22**	0.0006
-0.23	0.4090	**-0.73**	0.2327	**-1.23**	0.1093	**-1.73**	0.0418	**-2.23**	0.0129	**-2.73**	0.0032	**-3.23**	0.0006
-0.24	0.4052	**-0.74**	0.2296	**-1.24**	0.1075	**-1.74**	0.0409	**-2.24**	0.0125	**-2.74**	0.0031	**-3.24**	0.0006
-0.25	0.4013	**-0.75**	0.2266	**-1.25**	0.1056	**-1.75**	0.0401	**-2.25**	0.0122	**-2.75**	0.0030	**-3.25**	0.0006
-0.26	0.3974	**-0.76**	0.2236	**-1.26**	0.1038	**-1.76**	0.0392	**-2.26**	0.0119	**-2.76**	0.0029	**-3.26**	0.0006
-0.27	0.3936	**-0.77**	0.2206	**-1.27**	0.1020	**-1.77**	0.0384	**-2.27**	0.0116	**-2.77**	0.0028	**-3.27**	0.0005
-0.28	0.3897	**-0.78**	0.2177	**-1.28**	0.1003	**-1.78**	0.0375	**-2.28**	0.0113	**-2.78**	0.0027	**-3.28**	0.0005
-0.29	0.3859	**-0.79**	0.2148	**-1.29**	0.0985	**-1.79**	0.0367	**-2.29**	0.0110	**-2.79**	0.0026	**-3.29**	0.0005
-0.30	0.3821	**-0.80**	0.2119	**-1.30**	0.0968	**-1.80**	0.0359	**-2.30**	0.0107	**-2.80**	0.0026	**-3.30**	0.0005
-0.31	0.3783	**-0.81**	0.2090	**-1.31**	0.0951	**-1.81**	0.0351	**-2.31**	0.0104	**-2.81**	0.0025	**-3.31**	0.0005
-0.32	0.3745	**-0.82**	0.2061	**-1.32**	0.0934	**-1.82**	0.0344	**-2.32**	0.0102	**-2.82**	0.0024	**-3.32**	0.0005
-0.33	0.3707	**-0.83**	0.2033	**-1.33**	0.0918	**-1.83**	0.0336	**-2.33**	0.0099	**-2.83**	0.0023	**-3.33**	0.0004
-0.34	0.3669	**-0.84**	0.2005	**-1.34**	0.0901	**-1.84**	0.0329	**-2.34**	0.0096	**-2.84**	0.0023	**-3.34**	0.0004
-0.35	0.3632	**-0.85**	0.1977	**-1.35**	0.0885	**-1.85**	0.0322	**-2.35**	0.0094	**-2.85**	0.0022	**-3.35**	0.0004
-0.36	0.3594	**-0.86**	0.1949	**-1.36**	0.0869	**-1.86**	0.0314	**-2.36**	0.0091	**-2.86**	0.0021	**-3.36**	0.0004
-0.37	0.3557	**-0.87**	0.1922	**-1.37**	0.0853	**-1.87**	0.0307	**-2.37**	0.0089	**-2.87**	0.0021	**-3.37**	0.0004
-0.38	0.3520	**-0.88**	0.1894	**-1.38**	0.0838	**-1.88**	0.0301	**-2.38**	0.0087	**-2.88**	0.0020	**-3.38**	0.0004
-0.39	0.3483	**-0.89**	0.1867	**-1.39**	0.0823	**-1.89**	0.0294	**-2.39**	0.0084	**-2.89**	0.0019	**-3.39**	0.0003
-0.40	0.3446	**-0.90**	0.1841	**-1.40**	0.0808	**-1.90**	0.0287	**-2.40**	0.0082	**-2.90**	0.0019	**-3.4**	0.0003
-0.41	0.3409	**-0.91**	0.1814	**-1.41**	0.0793	**-1.91**	0.0281	**-2.41**	0.0080	**-2.91**	0.0018	**-3.41**	0.0003
-0.42	0.3372	**-0.92**	0.1788	**-1.42**	0.0778	**-1.92**	0.0274	**-2.42**	0.0078	**-2.92**	0.0018	**-3.42**	0.0003
-0.43	0.3336	**-0.93**	0.1762	**-1.43**	0.0764	**-1.93**	0.0268	**-2.43**	0.0075	**-2.93**	0.0017	**-3.43**	0.0003
-0.44	0.3300	**-0.94**	0.1736	**-1.44**	0.0749	**-1.94**	0.0262	**-2.44**	0.0073	**-2.94**	0.0016	**-3.44**	0.0003
-0.45	0.3264	**-0.95**	0.1711	**-1.45**	0.0735	**-1.95**	0.0256	**-2.45**	0.0071	**-2.95**	0.0016	**-3.45**	0.0003
-0.46	0.3228	**-0.96**	0.1685	**-1.46**	0.0721	**-1.96**	0.0250	**-2.46**	0.0069	**-2.96**	0.0015	**-3.46**	0.0003
-0.47	0.3192	**-0.97**	0.1660	**-1.47**	0.0708	**-1.97**	0.0244	**-2.47**	0.0068	**-2.97**	0.0015	**-3.47**	0.0003
-0.48	0.3156	**-0.98**	0.1635	**-1.48**	0.0694	**-1.98**	0.0239	**-2.48**	0.0066	**-2.98**	0.0014	**-3.48**	0.0003
-0.49	0.3121	**-0.99**	0.1611	**-1.49**	0.0681	**-1.99**	0.0233	**-2.49**	0.0064	**-2.99**	0.0014	**-3.49**	0.0002