BỘ GIÁO DỤC VÀ ĐÀO TẠO

TRƯỜNG ĐẠI HỌC DÂN LẬP HẢI PHÒNG

ISO 9001 : 2008

ĐỀ TÀI

NGHIÊN CỨU KHOA HỌC

NGHIÊN CỨU, BIÊN SOẠN TẬP BÀI GIẢNG MÔN XÁC SUẤT THỐNG KÊ DÙNG CHO TRƯỜNG ĐẠI HỌC DÂN LẬP HẢI PHÒNG

Chủ nhiệm đề tài : VŨ VĂN ÁNH

Thành viên : HOÀNG HẢI VÂN

HẢI PHÒNG, 2012

BỘ GIÁO DỤC VÀ ĐÀO TẠO

TRƯỜNG ĐẠI HỌC DÂN LẬP HÀI PHÒNG

ISO 9001 : 2008

NGHIÊN CỨU, BIÊN SOẠN TẬP BÀI GIẢNG MÔN XÁC SUẤT THỐNG KÊ DÙNG CHO TRƯỜNG ĐẠI HỌC DÂN LẬP HẢI PHÒNG

Chủ nhiệm đề tài : VŨ VĂN ÁNH

Thành viên : HOÀNG HẢI VÂN

HẢI PHÒNG, 2012

MỤC LỤC

MỞ ĐẦU

1.Lý do chọn đề tài.

Ra đời từ nửa cuối thế kỷ 17 ở nước Pháp, xác suất là một bộ phận của toán học nghiên cứu các hiện tượng ngẫu nhiên. Nói một cách đại khái thì hiện tượng ngẫu nhiên là hiên tượng ta không thể nói trước nó xảy ra hay không xảy ra khi thực hiện một lần quan sát. Tuy nhiên nếu tiến hành quan sát khá nhiều lần một hiện tượng ngẫu nhiên trong những hoàn cảnh như nhau, thì trong nhiều trường hợp ta có thể rút ra được những kết luận khoa học về hiện tượng này. Dựa vào các thành tựu của lý thuyết xác suất, thống kê toán xây dựng các phương pháp ra quyết định trong điều kiện thông tin không đầy đủ. Hơn 300 năm phát triển, đến nay nội dung và các phương pháp xác suất thống kê rất phong phú, được ứng dụng rộng rãi trong nhiều lĩnh vực tự nhiên và xã hội khác nhau, từ âm nhạc tới vật lý, từ cơ học tới thị trường chứng khoán, từ dự báo thời tiết tới kinh tế, từ nông học tới y học…

Môn học Xác suất thống kê là một môn học quan trọng ở bậc đại học.Ở nước ta, số sách về xác suất thống kê đã được xuất bản khá nhiều. Tuy nhiên đề tài này chỉ trình bày những vấn đề tương ứng với nội dung giảng dạy môn học Xác suất thống kê cho sinh viên trường Đại học Dân lập Hải Phòng. Vì vậy, đề tài này được viết hoàn toàn theo quan điểm thực hành, chỉ chú trọng việc áp dụng các phương pháp của xác suất và thống kê toán trong quản lý kinh tế và quản trị kinh doanh mà bỏ qua cơ sở toán học của các kết quả đó. Mỗi khái niệm, vấn đề hay phương pháp đều được minh họa bằng các ví dụ trong lĩnh vực thực tế, giúp giói thiệu cho sinh viên khả năng ứng dụng rộng rãi của các phương pháp đó trong việc giải quyết các vấn đề thực tiễn.

Vì vậy, nhóm tác giả đã lựa chọn đề tài: ***"Nghiên cứu biên soạn tập bài giảng môn xác suất thống kê dùng cho Trường Đại học Dân lập Hải Phòng".***

Cho đến nay, có thể khẳng định đây là 1 đề tài hoàn toàn mới.

2.Mục tiêu đề tài.

Đề tài tập trung xây dựng một bài giảng vừa đáp ứng yêu cầu chuẩn mực của sách giáo khoa, vừa có giá trị thực tiễn và phù hợp với phương thức tự học, tự nghiên cứu của sinh viên.

3.Các phương pháp nghiên cứu.

Đề tài sử dụng tổng hợp nhiều phương pháp như: tổng hợp, thống kê, phân tích, …

4.Bố cục của đề tài

Ngoài phần mở đầu và kết luận, đề tài gồm 2 phần có 7 chương:

Phần I: XÁC SUẤT

Chương 1: Giải tích tổ hợp

Chương 2: Biến cố và xác suất của biến cố

Chương 3: Đại lượng ngẫu nhiên và quy luật phân phối xác suất

Phần II: THỐNG KÊ

Chương 4: Lý thuyết mẫu

Chương 5: Ước lượng tham số

Chương 6: Kiểm định giả thuyết thống kê

Chương 7: Lý thuyết tương quan và hồi quy

PHẦN I: XÁC SUẤT

CHƯƠNG I: GIẢI TÍCH TỔ HỢP

1.1. Quy tắc cộng.

Định nghĩa: Nếu có m_1 cách chọn đối tượng x_1, m_2 cách chọn đối tượng x_2, …, m_k cách chọn đối tượng x_k, và cách chọn x_i không trùng với cách chọn x_j ($i \neq j$) thì có: $m_1 + m_2 + \ldots + m_k$ cách chọn một trong các đối tượng $x_1, x_2, \ldots, x_k$

Ví dụ: Trong hộp bút của sinh viên Tuấn có 5 bút màu xanh, 3 bút màu đen, 6 bút màu đỏ. Hỏi Tuấn có bao nhiêu cách chọn một bút để viết.

Giải: Tuấn có: - 5 cách chọn 1 bút màu xanh

- 3 cách chọn 1 bút màu đen

- 6 cách chọn 1 bút màu đỏ

Nếu Tuấn chọn bút màu xanh thì không chọn bút các màu khác và ngược lại. Do vậy Tuấn có: 5 + 3 + 6 = 14 cách chọn 1 bút để viết.

1.2. Quy tắc nhân.

Định nghĩa: Nếu có m_1 cách chọn đối tượng x_1, sau đó với mỗi cách chọn x_1 có m_2 cách chọn đối tượng x_2, sau đó với mỗi cách chọn x_1 và x_2 như thế có m_3 cách chọn đối tượng x_3, …, cuối cùng với mỗi cách chọn $x_1\, x_2 \ldots x_{k-1}$ có m_k cách chọn đối tượng x_k. Khi đó có tất cả: $m_1.m_2 \ldots m_k$ cách chọn dãy: $x_1\, x_2 \ldots x_k$

Ví dụ: Sinh viên Tuấn có 5 bút màu xanh, 3 bút màu đen, 6 bút màu đỏ. Hỏi Tuấn có bao nhiêu cách chọn mỗi loại một bút để mang tới lớp.

*Giải:*Tuấn có 5 cách chọn 1 bút màu xanh.

Sau khi chọn được bút màu xanh, Tuấn có 3 cách chọn 1 bút màu đen.

Sau khi chọn được bút màu xanh, bút màu đen, Tuấn có 6 cách chọn 1 bút màu đỏ

Do vậy Tuấn có: 5.3.6 = 90 cách chọn mỗi loại 1 bút để tới lớp.

1.3. Hoán vị.

Định nghĩa: Cho một tập X có n phần tử. Hoán vị của n phần tử là một nhóm có thứ tự gồm đủ mặt n phần tử đã cho.

Kí hiệu: Số các hoán vị của tập hợp n phần tử là: P_n

Công thức tính: $P_n = n!$

Ví dụ : Có 3 bộ sách: bộ 1 có 5 tập, bộ 2 có 3 tập, bộ 3 có 6 tập. Tất cả được đặt lên một giá sách, có bao nhiêu cách xếp nếu:

a. Sắp tùy ý

b. Các tập được đặt theo từng bộ

c. 3 tập được chỉ định phải sắp cùng nhau

d. 2 tập được chỉ định phải sắp cuối cùng.

Giải: a) Sắp tùy ý

Mỗi cách sắp là một hoán vị của 14 phần tử => Số cách sắp tùy ý là $P_{14} = 14!$

b) Sắp theo bộ:

- Mỗi bộ sách là một phần tử lớn => Có $P_3 = 3!$ Cách xếp 3 phần tử này.
- Các tập sách trong mỗi có thể hoán vị với nhau => có $P_5.P_3.P_6 = 5!3!6!$ cách xếp
- Vây có tất cả: 3! 5!3!6! cách xếp.

c) 3 tập được chỉ định phải xếp cùng nhau

- 3 tập được chỉ định xếp cùng nhau coi như một phần tử cùng xếp với 11 tập còn lại => có $P_{12} = 12!$ Cách xếp
- Cách xếp của 3 tập chỉ định với nhau có: $P_3 = 3!$ Cách xếp
- Vậy có tất cả: $P_{12}P_3 = 12!3!$ cách xếp.

d) 2 tập được chỉ định phải xếp cuối cùng.

- 2 tập được chỉ định xếp cuối cùng có $P_2 = 2!$ Cách xếp
- Xếp 12 tập còn lại có $P_{12} = 12!$ cách xếp.
- Vậy có tất cả: $P_2P_{12} = 2!12!$ cách xếp.

1.4.Chỉnh hợp (chỉnh hợp không lặp).

Định nghĩa: Cho một tập X có n phần tử. Chỉnh hợp chập k từ n phần tử ($k \leq n$) là một nhóm có thứ tự gồm k phần tử khác nhau được chọn ra từ n phần tử đã cho.

Kí hiệu: Số các chỉnh hợp chập k từ tập hợp n phần tử là: A_n^k

Công thức tính: $A_n^k = \dfrac{n!}{(n-k)!}$

Ví dụ1 : Có bao nhiêu số tự nhiên có 3 chữ số khác nhau được tạo ra từ các chữ số {1,2,3,4,5}.

Giải: Mỗi số tự nhiên thỏa mãn yêu cầu bài toán là một bộ 3 chữ số khác nhau (có kể thứ tự) chọn ra từ 5 chữ số {1,2,3,4,5}, nên mỗi số này là một chỉnh hợp chập 3 của 5 phần tử. Vậy số các số tự nhiên thỏa mãn là:

$$A_5^3 = \frac{5!}{(5-3)!} = 5.4.3 = 60$$

Ví dụ2 : Một lớp học có 30 sinh viên, có bao nhiêu cách chọn ra 3 sinh viên để làm Lớp trưởng, Lớp phó và Bí thư. Biết rằng khả năng được chọn của mỗi sinh viên là như nhau và mỗi sinh viên chỉ nhận một chức.

Giải: Mỗi cách chọn ra 3 sinh viên từ 30 sinh viên thỏa mãn yêu cầu bài toán là một chỉnh hợp chập 3 của 30 phần tử. Vậy số cách chọn là:

$$A_{30}^{3} = \frac{30!}{(30-3)!} = 30.29.28 = 24360$$

1.5.Tổ hợp.

Định nghĩa: Cho một tập X có n phần tử. Tổ hợp chập k từ n phần tử (k ≤ n) là một nhóm không phân biệt thứ tự gồm k phần tử khác nhau được chọn ra từ n phần tử đã cho.

Kí hiệu: Số các tổ hợp chập k từ tập hợp n phần tử là: C_n^k

Công thức tính: $C_n^k = \dfrac{n!}{(n-k)!k!}$

Ví dụ : Một lớp học có 30 sinh viên, có bao nhiêu cách chọn ra 3 sinh viên để tham gia vào đội sinh viên tình nguyện của, biết rằng khả năng được chọn của mỗi sinh viên là như nhau.

Giải: Mỗi cách chọn ra 3 sinh viên từ 30 sinh viên thỏa mãn yêu cầu bài toán là một tổ hợp chập 3 của 30 phần tử. Vậy số cách chọn là:

$$C_{30}^{3} = \frac{30!}{(30-3)!3!} = 4060$$

Tính chất thường gặp về tổ hợp

$$C_n^k = C_n^{n-k}$$

$$C_n^k = C_{n-1}^k + C_{n-1}^{k-1}$$

Bài tập chương 1

Bài 1: Một lớp học gồm 30 sinh viên nam và 20 sinh viên nữ. Hỏi có bao nhiêu cách chọn từ lớp đó :

a. 4 sinh viên vào đội sinh viên tình nguyện của trường.
b. 4 sinh viên vào ban cán sự lớp, trong đó có 1 lớp trưởng, 1 lớp phó phụ trách học tập, 1 lớp phó phụ trách đời sống và 1 bí thư đoàn.
c. 4 sinh viên trong đó có 2 sinh viên nam và 2 sinh viên nữ đi tập văn nghệ.
d. 4 sinh viên đi dự Đại hội, trong đó có ít nhất 1 sinh viên nam.
e. 4 sinh viên đi tập erobic, trong đó có sinh viên Tuấn (Tuấn là một trong 30 sinh viên nam)

Bài 2: Có bao nhiêu cách chọn 7 quân bài từ 1 bộ bài tú lơ khơ gồm 52 quân bài trong các trường hợp sau:

a. Bất kỳ
b. Có 2 quân K, 3 quân Q
c. Có không quá 1 quân A
d. Có ít nhất 1 J
e. Có quân J cơ

Bài 3: Trong một buổi liên hoan có 10 cặp nam nữ, trong đó có 7 cặp là vợ chồng. Hỏi có bao nhiêu cách chọn ra 3 người trong các trường hợp sau:

a. 3 người bất kỳ.
b. 3 người, trong đó có ít nhất 1 nữ.
c. 3 người trong đó có 1 cặp là vợ chồng.
d. 3 người, trong đó không có cặp vợ chồng nào.

CHƯƠNG 2: BIẾN CỐ NGẪU NHIÊN VÀ XÁC SUẤT CỦA NÓ

2.1. Phép thử, biến cố và mối quan hệ giữa các biến cố.

2.1. 1.Phép thử và biến cố

Định nghĩa: Việc thực một nhóm các điều kiện cơ bản để quan sát một hiện tượng nào đó được gọi là một phép thử. Các kết quả có thể xảy ra của phép thử được gọi là biến cố (sự kiện).

Một phép thử được thực hiện nhiều lần trong cùng điều kiện, nếu phép thử cho ra cùng kết quả thì ta nói phép thử đó là phép thử tất nhiên. Ngược lại nếu phép thử cho ta các kết quả khác nhau gọi là phép thử ngẫu nhiên, các kết quả của phép thử ngẫu nhiên gọi là các biến cố ngẫu nhiên. Chú ý trong xác suất ta chỉ quan tâm tới phép thử ngẫu nhiên.

Ví dụ:

i)Tung một đồng xu lên một mặt bàn cứng là một phép thử. Các kết quả có thể có là xuất hiện mặt sấp hoặc ngửa là các biến cố.

ii) Bắn một viên đạn vào một cái bia là một phép thử. Các kết quả có thể có là viên đạn trúng hoặc trượt bia các biến cố.

2.1.2. Các loại biến cố

Biến cố chắc chắn: Là biến cố nhất định sẽ xảy ra khi thực hiện phép thử, kí hiệu là:U hoặc Ω.

Biến cố không thể có: Là biến cố không bao giờ xảy ra khi thực hiện phép thử, kí hiệu là:V hoặc ∅.

Biến cố ngẫu nhiên: Là biến cố có thể xảy ra cũng có thể không xảy ra khi thực hiện phép thử, kí hiệu là:A, B, C, … hoặc là: A_1, A_2, A_3,…

Ví dụ: Tung một con xúc xắc.

- Gọi V là biến cố xuất hiên mặt 7 chấm, V là biến cố không thể có.

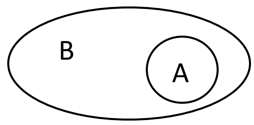
- Gọi A_i là biến cố xuất hiện mặt i chấm, i = {1, 2,…,6}, A_i là biến cố ngẫu nhiên.

- Gọi U là biến cố xuất hiện mặt có số chấm nhỏ hơn 7, U là biến cố chắc chắn.

2.1.3. Quan hệ giữa các biến cố

2.1.3.1. Quan hệ kéo theo: Biến cố A được gọi là kéo theo biến cố B, kí hiệu $A \subset B$ hoặc A=> B, nếu A xảy ra thì B xảy ra.

Biểu đồ Ven



Ví dụ:

i) Một người mua một tờ vé số

- Gọi A là biến cố người đó trúng số độc đắc
- Gọi B là biến cố người đó trúng số

Khi đó $A \subset B$

ii) Gieo một con xúc xắc

- Gọi A là biến cố xuất hiện mặt 2 chấm
- Gọi B là biến cố xuất hiện mặt có số chấm chẵn

Khi đó $A \subset B$

2.1.3.2. Quan hệ tương đương: Hai biến cố A và B được gọi là tương đương nhau nếu $A \subset B$ và $B \subset A$, kí hiệu là A= B

Ví dụ:

i) Gieo một con xúc xắc

- Gọi A là biến cố xuất hiện mặt 2, 4, 6 chấm
- Gọi B là biến cố xuất hiện mặt có số chấm chẵn

Khi đó A= B

i) Một hộp có 10 bi, trong đó có 7 bi trắng, 3 bi xanh. Lấy ngẫu nhiên ra 2 bi.

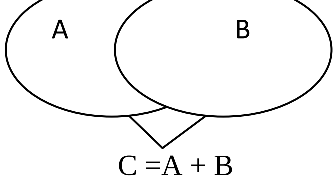
- Gọi A là biến cố lấy được 1 bi trắng
- Gọi B là biến cố lấy được 1 bi xanh

Khi đó A= B

2.1.3.3. Biến cố tổng (hợp).

Định nghĩa: Biến cố C được gọi là tổng của 2 biến cố A và B, kí hiệu: C =A + B hoặc C = A U B, nếu C xảy ra khi và chỉ khi có ít nhất 1 trong 2 biến cố A hoặc B xảy ra khi thực hiện phép thử.

Biểu đồ Ven



C =A + B

Ví dụ:

i) Gieo một con xúc xắc

- Gọi A là biến cố xuất hiện mặt 2, 4 chấm
- Gọi B là biến cố xuất hiện mặt 4, 6 chấm
- Gọi C là biến cố xuất hiện mặt có số chấm chẵn

Khi đó C = A + B

ii) Hai người cùng bắn vào một mục tiêu

- Gọi A là biến cố người thứ nhất bắn trúng

- Gọi B là biến cố người thứ hai bắn trúng

- Gọi C là biến cố mục tiêu trúng đạn

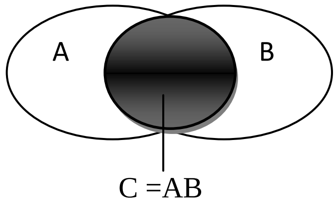
Khi đó C = A + B

Mở rộng: Biến cố A gọi là tổng của n biến cốA_1, A_2,…, A_n , nếu A xảy ra khi có ít nhất một trong n biến cố đó xảy ra khi thực hiện phép thử.

kí hiệu là: $A = A_1 + A_2 + \ldots + A_n$,

2.1.3.4. Biến cố tích (giao)

Định nghĩa: Biến cố C được gọi là tích của 2 biến cố A và B, Kí hiệu: C = AB hoặc C = A ∩ B, nếu C xảy ra khi và chỉ khi cả 2 biến cố A và B cùng xảy ra, khi thực hiện phép thử.



Biểu đồ Ven

Ví dụ:

i) Gieo một con xúc xắc

- Gọi A là biến cố xuất hiện mặt 2, 4 chấm

- Gọi B là biến cố xuất hiện mặt 4, 6 chấm

- Gọi C là biến cố xuất hiện mặt 4 chấm

Khi đó C = AB

ii) Hai người cùng bắn vào một mục tiêu

- Gọi A là biến cố người thứ nhất bắn trượt

- Gọi B là biến cố người thứ hai bắn trượt

- Gọi C là biến cố mục tiêu không trúng đạn

Khi đó C = AB

Mở rộng: Biến cố A gọi là tổng của n biến cốA_1, A_2,…, A_n , nếu A xảy ra khi và chỉ khi đồng thời n biến cố đó cùng xảy ra khi thực hiện phép thử.

kí hiệu là: $A = A_1 + A_2 + \ldots + A_n$

2.1.3.5. Biến cố sơ cấp

Định nghĩa: Biến cố sơ cấp là biến cố không thể phân tích được nữa.

Ví dụ: Gieo một con xúc xắc

- Gọi A_i là biến cố xuất hiện mặt i chấm. $i = \{1,2,\ldots,6\}$

- Gọi B là biến cố xuất hiện mặt có số chấm chẵn

Khi đó: A_i là biến cố sơ cấp

B không là biến cố

2.1.3.6. Biến cố xung khắc

Định nghĩa: Hai biến cố A và B được gọi là xung khắc với nhau khi và chỉ khi chúng không đồng thời xảy ra khi thực hiện phép thử. Trong trường hợp ngược lại, nếu hai biến cố đó có thể xảy ra trong cùng một phép thử thì được gọi là không xung khắc.

*Ví dụ:*Gieo một con xúc xắc

- Gọi A là biến cố xuất hiện mặt 2 chấm
- Gọi B là biến cố xuất hiện mặt 1, 3 chấm
- Gọi C là biến cố xuất hiện mặt có số chấm chẵn
- Gọi D là biến cố xuất hiện mặt có số chấm lẻ

Khi đó: A và B xung khắc nhau

A và D xung khắc nhau

C và D xung khắc nhau

A và C không xung khắc nhau

B và D không xung khắc nhau

Mở rộng: Nhóm n biến cố $A_1, A_2, \ldots, A_n$ được gọi là xung khắc từng đôi nếu bất kỳ hai biến cố nào trong nhóm này cũng xung khắc nhau.

2.1.3.7. Nhóm biến cố đầy đủ

Định nghĩa: Nhóm các biến cố $A_1, A_2, \ldots.., A_n$ được gọi là 1 nhóm biến cố đầy đủ nếu trong kết quả của phép thử chỉ xảy ra 1 và chỉ 1 trong các biến cố ấy.

Nói cách khác, các biến cố này được gọi là 1 nhóm biến cố đầy đủ nếu chúng xung khắc từng đôi với nhau và tổng của chúng là 1 biến cố chắc chắn.

Ví dụ: Gieo một con xúc xắc

- Gọi A_i là biến cố xuất hiện mặt i chấm. $i = \{1,2,\ldots,6\}$
- Gọi B là biến cố xuất hiện mặt có số chấm chẵn
- Gọi C là biến cố xuất hiện mặt có số chấm < 3

Khi đó: $\{A_1, A_2, A_3, A_4, A_5, A_6\}$ là nhóm biến cố đầy đủ

$\{B, A_1, A_3, A_5\}$ là nhóm biến cố đầy đủ không là biến cố

$\{C, A_3, A_4, A_5, A_6\}$ là nhóm biến cố đầy đủ

$\{A_1, A_2, A_3, A_4, A_5, A_6, B, C\}$ không là nhóm biến cố đầy đủ

2.1.3.8. Biến cố đối lập

Định nghĩa: Hai biến cố A và B được gọi là 2 biến cố đối lập nếu A và B không đồng thời xảy ra, và 1 trong 2 biến cố A hoặc B phải xảy ra, khi thực hiện phép thử. Nói cách khác, 2 biến cố được gọi là đối lập với nhau nếu chúng tạo thành 1 nhóm biến cố đầy đủ.

Biến cố đối lập của A được kí hiệu là $\overline{A}$ hoặc A*

Ví dụ: Gieo một con xúc xắc

Nếu gọi A là biến cố xuất hiện mặt có số chấm chẵn

Khi đó $\overline{A}$ biến cố xuất hiện mặt có số chấm lẻ

Nếu gọi B là biến cố xuất hiện mặt có số chấm < 3

Khi đó $\overline{B}$ biến cố xuất hiện mặt có số chấm ≥ 3

2.1.3.9. Biến cố độc lập.

Định nghĩa: Biến cố A được gọi là độc lập với biến cố B nếu sự xảy ra hay không xảy ra của B không ảnh hưởng gì đến sự xảy ra hay không xảy ra của A, và ngược lại.

2.2. Xác suất

- Qua việc quan sát các sự kiện ngẫu nhiên ta thấy rằng khả năng xuất hiện của các biến cố ngẫu nhiên nói chung không đồng đều, một số thường hay xảy ra, một số khác thường ít xảy ra.Từ đó nảy sinh vấn đề tìm cách đo lường "độ chắc" của một biến cố. Muốn vậy người ta tìm cách gán cho mỗi biến cố một số p không âm, số này được gọi là xác suất của biến cố đó.Ký hiệu xác suất của biến cố A là: P(A)

- Để phù hợp với nội dung thước đo "độ chắc" của biến cố trong phép thử, xác suất phải được xây dựng sao cho thỏa mãn các đòi hỏi sau:

+ Xác suất của biến cố chắc chắn U bằng 1: P(U) = 1

+ Xác suất của biến cố không thể có V bằng 0: P(V) = 0

+ Xác suất của biến cố ngẫu nhiên A nằm giữa 0 và 1: $0 \le P(A) \le 1$

2.2.1.Định nghĩa xác suất cổ điển

Định nghĩa: Giả sử trong một phép thử có tất cả n kết cục đồng khả năng, trong đó có m kết cục thuận lợi cho biến cố A. Khi đó xác suất của biến cố A là:

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

Ví dụ1: Trong 1 bình kín có 5 cầu trắng, 3 cầu đen. Lấy ngẫu nhiên 2 quả. Tìm xác suất để:

a. Lấy được 2 cầu trắng

b. Lấy được 2 cầu đen

c. Lấy được một cầu trắng và một cầu đen

Giải: Do lấy ngẫu nhiên 2 cầu từ bình 8 cầu nên số kết cục đồng khả năng là: $n = C_8^2$

Gọi A là biến cố "Lấy được 2 cầu trắng"

Số kết cục thuận lợi cho A là: $m_A = C_5^2$

$$=> \ \mathrm{P(A)} = \frac{m_A}{n} = \frac{C_5^2}{C_8^2}$$

b. Gọi B là biến cố "Lấy được 2 cầu đen"

Số kết cục thuận lợi cho B là: $m_B = C_3^2$

$$=> \mathrm{P(B)} = \frac{m_B}{n} = \frac{C_3^2}{C_8^2}$$

c. Gọi C là biến cố "Lấy được 1 cầu đen, 1 cầu trắng"

Số kết cục thuận lợi cho C là: $m_C = C_3^1 C_5^1$

$$=> \ \mathrm{P(C)} = \frac{m_C}{n} = \frac{C_3^1 C_5^1}{C_8^2}$$

2.2.2.Định nghĩa xác suất theo tần xuất

Định nghĩa: Nếu ta làm đi làm lại một phép thử nào đó n lần mà có m lần biến cố A xuất hiện thì tỷ số: m/n gọi là tần suất xuất hiện của biến cố A , ký hiệu: f(A) = m/n.

Khi n thay đổi, tần suất m/n cũng thay đổi nhưng nó luôn dao động quanh một số cố định nào đó, nếu n càng lớn thì m/n càng gần số cố định đó. Số cố định ấy được gọi là xác suất của biến cố A theo nghĩa thống kê. Ký hiệu: P(A)

Trên thực tế thì : $\mathrm{f(A)} \xrightarrow{n\to\infty} \mathrm{P(A)}$

Ví dụ: Để nghiên cứu khả năng xuất hiện mặt sấp khi tung một đồng xu, người ta tiến hành tung 1 đồng xu nhiều lần và thu được kết quả sau:

Người làm thí nghiệm	Số lần tung:n	Số lần được mặt sấp: m	Tần suất $f(A) = m/n$
Buffon	4040	2048	0.5069
Pearson	12000	6019	0.5016
Pearson	24000	12012	0.5005

Nhận xét: Qua ví dụ trên ta thấy, khi số phép thử tăng lên thì tần suất xuất hiện mặt sấp sẽ dao động càng ít hơn xung quanh giá trị không đổi là 0.5. Điều đó hy vọng rằng khi số phép thử tăng lên vô hạn, tần suất sẽ hội tụ về xác suất, tức là P(A) = f(A) = 0.5

2.2.3.Định nghĩa xác suất theo hình học

Định nghĩa: Xét một phép thử có vô hạn kết cục đồng khả năng. Giả sử ta có thể biểu diễn tập hợp mọi kết cục này bởi một miền hình học G nào đó(1 đoạn thẳng, 1 miền phẳng, một khối không gian,…) và những kết cục thích hợp cho biến cố A bởi các điểm thuộc miền con g của G. Với các giả thiết trên, xác suất của biến cố A được tính như sau.

$$P(A) = \frac{mes(g)}{mes(G)} = \frac{\text{kích thuoc mien g}}{\text{kích thuoc mien G}}$$

Ví dụ: Đường dây điện thoại ngầm nối hai trạm A và B bỗng nhiên bị đứt. Dây dài 800m chôn trong lòng đất đồng chất. Hãy tính xác suất của biến cố: chỗ đứt cách A không quá 100m.

Giải: Ta thấy dây có thể đứt tại bất kỳ 1 điểm nào trên đoạn AB, do đó có thể biểu diễn tập hợp mọi kết cục đồng khả năng của phép thử bởi đoạn AB.

Gọi H là biến cố “ chỗ đứt cách A không quá 100 m”

Các kết cục thích hợp cho biến cố H biểu thị bởi đoạn AC

Do đó:

$$P(H) = \frac{\text{do dai AC}}{\text{do dai AB}} = \frac{100}{800} = \frac{1}{8}$$

2.3. Các định lí cơ bản của xác suất

2.3.1. Định lí nhân xác suất.

2.3.1.1.Xác suất có điều kiện: Xác suất của biến cố A được tính trong điều kiện biết rằng biến cố B đã xảy ra được gọi là xác suất của A với điều kiện B.

Kí hiệu: P(A/B).

Biến cố B được gọi là biến cố điều kiện

Ví dụ: Trong một bình có 5 cầu trắng, 3 cầu đen. Lấy ngẫu nhiên lần lượt hai cầu theo phương thức không hoàn lại. Tìm xác suất để lần hai lấy được cầu trắng. Biết rằng lần thứ nhất đã lấy được cầu trắng.

Giải: Gọi A là biến cố "lần 1 lấy được cầu trắng"

Gọi B là biến cố "lần 2 lấy được cầu trắng"

=> P(A) = 5/8;

P(B/A) = 4/7

2.3.1.2. Định lý: Xác suất của tích 2 biến cố được xác định là tích xác suất của 1 trong 2 biến cố ấy với xác suất có điều kiện của biến cố còn lại.

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

Ví dụ: Trong 1 hộp kín có 10 sản phẩm tốt, 3 sản phẩm xấu. Lấy ngẫu nhiên lần lượt ra 2 sản phẩm. Tìm xác suất để 2 sản phẩm lấy ra đều là tốt.(theo phương thức không hoàn lại)

Giải: Gọi A là biến cố "sản phẩm lấy ra lần 1 là tốt"

B là biến cố "sản phẩm lấy ra lần 2 là tốt"

C là biến cố "cả 2 sản phẩm lấy ra đều tốt"

=> C = AB

$$\Rightarrow P(C) = P(AB) = P(A).P(B/A) = \frac{10}{13}.\frac{9}{12} = \frac{15}{26}$$

2.3.1.3. Hệ quả:

* Nếu $P(A) \neq 0$, $P(B) \neq 0$ thì

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

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

* Nếu A, B là 2 biến cố độc lập thì

$$P(AB) = P(A)\,P(B)$$

2.3.1.4. Tổng quát:

* Nếu $A_1, A_2, \ldots, A_n$ không độc lập từng đôi với nhau thì

$$P(A_1 A_2 \ldots A_n) = P(A_1)P(A_2/A_1)P(A_3/A_1A_2)\ldots P(A_n/A_1A_2\ldots A_{n-1})$$

* Nếu $A_1, A_2, \ldots, A_n$ độc lập từng đôi với nhau thì

$$P(A_1 A_2 \ldots A_n) = P(A_1)P(A_2)P(A_3)\ldots P(A_n)$$

* Nếu $A \subset B$ thì $P(AB) = P(A)$

Ví dụ1: Trong 1 hộp kín có 5 bi trắng, 3 bi đen. Lấy ngẫy nhiên từ hộp này lần lượt từng viên bi cho đến khi lấy được bi màu đen thì dừng lại. Tìm xác suất để lấy ra ngoài đúng 4 viên, biết cách thức lấy là không hoàn lại.

Giải: Gọi A_i là biến cố " lấy được bi trắng ở lần lấy thứ i"

$\bar{A}_i$ là biến cố " lấy được bi đen ở lần lấy thứ i"

B là biến cố " lấy được đúng 4 viên bi"

$\Rightarrow B = A_1A_2A_3\bar{A}_4$

$\Rightarrow P(B) = P(A_1A_2A_3\bar{A}_4)$

$= P(A_1).P(A_2/A_1).P(A_3/A_1A_2).P(\bar{A}_4/A_1A_2A_3)$

$$= \frac{5}{8}.\frac{4}{7}.\frac{3}{6}.\frac{3}{5} = \frac{3}{28}$$

Ví dụ2: Hai xạ thủ cùng bắn vào một mục tiêu một cách độc lập, mỗi người bắn 1 phát, xác suất trúng đích 2 người lần lượt là 0,7 và 0,8. Tìm xác suất để cả 2 người cùng bắn trúng mục tiêu.

Giải: Gọi A_i là biến cố "người thứ i bắn trúng mục tiêu i", i=1,2

B là biến cố " cả hai người cùng bắn trúng mục tiêu"

$\Rightarrow B = A_1A_2$

$\Rightarrow P(B) = P(A_1A_2)$

$= P(A_1).P(A_2) = 0,7.0,8 = 0,56$

(do A_1 và A_2 độc lập)

Ví dụ3: Một nhóm sinh viên có 7 nam và 3 nữ, chọn ngẫu nhiên ra 3 người.

a. Tìm xác suất để có 3 nam được chọn.

b. Tìm xác suất để có 3 nam được chọn, biết rằng có ít nhất 1 nam đã được chọn.

c. Giả sử Tuấn là 1 trong 7 nam. Tìm xác suất để Tuấn được chọn nếu biết rằng có ít nhất 1 nam đã được chọn.

Giải: Gọi A là biến cố "3 nam được chọn"

B là biến cố "có ít nhất 1 nam được chọn"

C là biến cố " Tuấn được chọn"

Số kết cục duy nhất đồng khả năng là: $n = C_{10}^3$

a. Số kết cục thuận lợi cho A là : $m_A = C_7^3$

$$P(A) = \frac{m_A}{n} = \frac{C_7^3}{C_{10}^3}$$

b. $$P(A/B) = \frac{P(AB)}{P(B)} = \frac{P(A)}{P(B)} = \frac{C_7^3}{C_{10}^3} : \left[1 - \frac{C_3^3}{C_{10}^3}\right]$$

c. $$P(C/B) = \frac{P(CB)}{P(B)} = \frac{P(C)}{P(B)} = \frac{C_1^1 C_9^2}{C_{10}^3} : \left[1 - \frac{C_3^3}{C_{10}^3}\right]$$

2.3.2. Công thức cộng

2.3.2.1. Định lý

* Nếu 2 biến cố A, B không xung khắc thì:

P(A + B) = P(A) + P(B) – P(AB)

* Nếu 2 biến cố A, B xung khắc thì:

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

Ví dụ1: Hai xạ thủ mỗi người bắn 1 phát vào 1 bia, xác suất trúng bia của mỗi người lần lượt là 0,7 và 0,8. Tìm xác suất để bia trúng đạn.

Giải: Gọi A_i là biến cố "người thứ i bắn trúng bia" i = 1, 2

Gọi B là biến cố "bia trúng đạn" P(B) =?

$\Rightarrow$ B = $A_1 + A_2$

$\Rightarrow$ P(B) = P($A_1 + A_2$) = P(A_1) + P(A_2) - P(A_1 A_2)

= 0,7 + 0,8 – 0,7.0,8 = 0,94

Ví dụ2: Trong 1 hộp kín có 10 sản phẩm tốt, 3 sản phẩm xấu. Lấy ngẫu nhiên ra 6 sản phẩm. Tìm xác suất để 6 sản phẩm lấy ra có không quá 1 sản phẩm xấu.

Giải: Gọi A_0 là biến cố "6 sản phẩm lấy ra không có sản phẩm xấu "

A_1 là biến cố "6 sản phẩm lấy ra có 1 sản phẩm xấu"

B là biến cố "6 sản phẩm có không quá 1 sản phẩm xấu" P(B) =?

$\Rightarrow$ B = $A_0 + A_1$. Do A_0 và A_1 là xung khắc nhau nên

=> P(B) = P($A_0 + A_1$) = P(A_0) + P(A_1)

$$\left.\begin{aligned} P(A_0) &= \frac{C_{10}^6}{C_{13}^6} \\ P(A_1) &= \frac{C_{10}^5 C_3^1}{C_{13}^6} \end{aligned}\right\} \Rightarrow P(B) = \frac{C_{10}^6}{C_{13}^6} + \frac{C_{10}^5 C_3^1}{C_{13}^6}$$

2.3.2.2. Hệ quả:

i) Nếu $A_1, A_2, \ldots, A_n$ là các biến cố bất kỳ thì:

$$P\left(\sum_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i A_j) + \sum_{i<j<k} P(A_i A_j A_k) + \ldots + (-1)^{n-1} P(A_1 A_2 \ldots A_n)$$

ii) Nếu $A_1, A_2, \ldots, A_n$ độc lập toàn phần thì:

$$P(A_1 + A_2 + \ldots + A_n) = 1 - \prod_{i=1}^{n} P(\overline{A_i})$$

iii) Nếu $A_1, A_2, \ldots, A_n$ xung khắc từng đôi thì

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

iv) Nếu $A_1, A_2, \ldots, A_n$ là nhóm đầy đủ các biến cố thì:

$$P(A_1 + A_2 + \ldots + A_n) = P(A_1) + P(A_2) + \ldots + P(A_n) = 1$$

v) Nếu A và $\bar{A}$ Là hai biến cố đối lập thì:

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

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

Chú ý: Nếu A, $\bar{A}$ đối lập và B là biến cố bất kỳ thì:

A/B và $\bar{A}/B$ cũng đối lập.

Ví dụ3: Hai sinh viên cùng thi 1 môn tại hai phòng khác nhau, xác suất để 2 sinh viên thi đỗ lần lượt là 0,7 và 0,8. Tìm xác suất để :

a. Có ít nhất 1 sinh viên thi đỗ

b. Có đúng 1 sinh viên thi đỗ.

c. Biết có đúng 1 sinh viên thi đỗ, tìm xác suất để sinh viên thứ nhất thi đỗ.

Giải: Gọi A_i là biến cố "sinh viên thứ i thi đỗ"

$\bar{A}_i$ là biến cố "sinh viên thứ i thi trượt" $i = 1, 2$

a. Gọi B là biến cố "Có ít nhất 1 sinh viên thi đỗ"

là biến cố "không có sinh viên nào thi đỗ"

b. Gọi C là biến cố "Có đúng 1 sinh viên thi đỗ "

$$\Rightarrow P(C) = P(A_1\overline{A_2} + \overline{A_1}A_2) = P(A_1\overline{A_2}) + P(\overline{A_1}.A_2)$$

$$= P(A_1)P(\overline{A_2}) + P(\overline{A_1}).P(A_2) = 0,7.0,2 + 0,3.0,8 = 0,38$$

$$\Rightarrow P(B) = 1 - P(\overline{B}) = 1 - P(\overline{A_1}.\overline{A_2}) = 1 - 0,3.0,2 = 0,94$$

c. Xác suất để sính viên thứ nhất thi đỗ biết rằng có đúng 1 sinh viên thi đỗ chính là:

$$P(A_1/C)=\frac{P(A_1C)}{P(C)}=\frac{P(A_1\overline{A_2})}{P(C)}=\frac{P(A_1)P(\overline{A_2})}{P(C)}=\frac{0,7.0,2}{0,38}=\frac{7}{19}$$

Ví dụ4: Hai sinh viên cùng vào phòng thi 1 môn, xác suất để 2 sinh viên thi đỗ lần lượt là 0,7 và 0,8. Tuy nhiên xác suất để cả 2 sinh viên cùng đỗ là 0,6. Tìm xác suất để có đúng 1 sinh viên thi đỗ.

Giải: Gọi A_i là biến cố "sinh viên thứ i thi đỗ"

$\bar{A}_i$ là biến cố "sinh viên thứ i thi trượt" i = 1, 2

Gọi B là biến cố "Có đúng 1 sinh viên thi đỗ "

$$\Rightarrow P(B)=P(A_1\overline{A}_2+\overline{A}_1A_2)=P(A_1\overline{A}_2)+P(\overline{A}_1.A_2)$$

$$= P(A_1)P(\overline{A}_2/A_1) + P(A_2)P(\overline{A}_1/A_2)$$

$$=P(A_1).\left[1-P\left(A_2/A_1\right)\right]+P(A_2)\left[1-P\left(A_1/A_2\right)\right]$$

$$=P(A_1)-P(A_1A_2)+P(A_2)-P(A_1A_2)=0,6$$

2.2.3.Công thức xác suất đầy đủ

Định lý: Nếu $A_1,A_2,\ldots,A_n$ là nhóm đầy đủ các biến cố, B là một biến cố bất kỳ có thể xảy ra đồng thời với một trong các biến cố đó. Khi đó:

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

Ví dụ: Có 2 hộp đựng sản phẩm, hộp 1 có 8 chính phẩm và 2 phế phẩm, hộp 2 có 12 chính phẩm và 3 phế phẩm. Lấy ngẫu nhiên 1 hộp và từ hộp đó lấy ra 2 sản phẩm. Tìm xác suất để 2 sản phẩm lấy ra là chính phẩm.

Giải: Gọi A_i là biến cố "lấy được hộp thứ i" i = 1, 2

A_1, A_2 là nhóm đầy đủ các biến cố và $P(A_1) = P(A_2) =0,5$

Gọi B là biến cố "lấy được 2 chính phẩm"

B xảy ra đồng thời với một trong 2 biến cố A_1, A_2

$$P(B) = P(A_1)P(B/A_1) + P(A_2)P(B/A_2)$$

$$= 0,5.\frac{C_8^2}{C_{10}^2}+0,5.\frac{C_{12}^2}{C_{15}^2}$$

2.2.4.Công thức Bayes

Định lý: Nếu $A_1,A_2,\ldots,A_n$ là nhóm đầy đủ các biến cố, B là một biến cố bất kỳ xảy ra đồng thời với một trong các biến cố đó. Khi đó:

$$P(A_k/B)=\frac{P(A_k)P(B/A_k)}{P(B)}=\frac{P(A_k)P(B/A_k)}{\sum_{i=1}^{n}P(A_i)P(B/A_i)};\quad k=1\div n$$

Ví dụ: Có 2 hộp đựng sản phẩm, hộp 1 có 8 chính phẩm và 2 phế phẩm, hộp 2 có 12 chính phẩm và 3 phế phẩm. Lấy ngẫu nhiên 1 hộp và từ hộp đó lấy ra 2 sản phẩm. Tìm xác suất để 2 sản phẩm lấy ra là của hộp 1 biết rằng 2 sản phẩm lấy ra là chính phẩm.

Giải: Gọi A_i là biến cố "lấy được hộp thứ i" i = 1, 2

A_1, A_2 là nhóm đầy đủ các biến cố và $P(A_1) = P(A_2) = 0,5$

Gọi B là biến cố "lấy được 2 chính phẩm"

B xảy ra đồng thời với một trong 2 biến cố A_1, A_2

$$P(B) = P(A_1)P(B/A_1) + P(A_2)P(B/A_2) = 0,5.\frac{C_8^2}{C_{10}^2}+0,5.\frac{C_{12}^2}{C_{15}^2}$$

Xác suất để lấy được 2 chính phẩm từ hộp 1 là:

$$P(A_1/B)=\frac{P(A_1)P(B/A_1)}{P(B)}=0,497$$

2.3.5. Công thức Bernoulli.

2.3.5.1. Dãy phép thử độc lập: Các phép thử được gọi là độc lập với nhau nếu việc xảy ra hay không xảy ra của một biến cố ở phép thử này không làm ảnh hưởng đến việc nó xảy ra hay không xảy ra ở phép thử khác.

Ví dụ: - Tung một đồng xu nhiều lần sẽ tạo nên các phép thử độc lập

- Lấy nhiều lần 1 đồ vật từ 1 thùng đồ vật theo phương thức hoàn lại tạo nên các phép thử độc lập.
- Lấy nhiều lần 1 sản phẩm từ 1 kho sản phẩm (rất lớn) tạo nên các phép thử độc lập.

……………

2.3.5.2. Công thức Bernoulli: Giả sử tiến hành n phép thử:

- Độc lập.

- Trong kết quả của mỗi phép thử chỉ có 2 khả năng xảy ra hoặc A hoặc Ā xuất hiện.

- Xác suất xảy ra biến cố A trong mỗi phép thử đều bằng p và xác suất xảy ra biến cố Ā trong mỗi phép thử đều bằng q = 1- p.

Những bài toán thỏa mãn 3 điều kiện trên được gọi là những bài toán tuân theo lược đồ Bernoulli. Khi đó xác suất để trong n phép thử nói trên biến cố A xuất hiện đúng k lần, kí hiệu: $P_n(k)$ và được tính bằng công thức:

$$P_n(k) = C_n^k p^k q^{n-k} \qquad k = 0 \div n$$

Ví dụ1: Một bài thi trắc nghiệm gồm 20 câu hỏi, mỗi câu hỏi có 5 câu trả lời, trong đó chỉ có một câu trả lời đúng. Giả sử rằng mỗi một câu trả lời đúng được 5 điểm và mỗi câu trả lời sai bị trừ đi 1 điểm, một người trả lời ngẫu nhiên. Tìm xác suất để:

a. Người thi đó được đúng 46 điểm?

b. Người thi đó bị điểm âm?

Giải: Gọi A là biến cố "trả lời đúng 1 câu" => P(A) = 0,2

Gọi k là số câu trả lời đúng => số điểm anh ta đạt được là

$$5k + (20 - k)(-1) = 6k - 20$$

a. Gọi B là biến cố "anh ta được 46 điểm"

$$P(B) = P_{n=20}(6k-20=46) = P_{n=20}(k=11)$$
$$= C_{20}^{11}(0,2)^{11}(0,8)^{9} = 4,6.10^{-4}$$

C: " anh ta bị điểm âm"

$$\Rightarrow P(C) = P_{n=20}(6k-20<0) = P_{n=20}(k<3,3)$$
$$= P_{n=20}(k=0) + P_{n=20}(k=1) + P_{n=20}(k=2) + P_{n=20}(k=3)$$

$$= C_{20}^{0}(0,2)^{0}(0,8)^{20} + C_{20}^{1}(0,2)^{1}(0,8)^{19} + C_{20}^{2}(0,2)^{2}(0,8)^{18} + C_{20}^{3}(0,2)^{3}(0,8)^{17}$$
$$= 0,41$$

Bài tập chương 2

Câu 1 :Trứng gà nở với xác suất là 0,8. Nếu trứng gà nở thì khả năng nở ra gà mái và gà trống là như nhau.

a.Cho ấp 1 quả. Tìm xác suất nở ra gà mái?

b.Cho ấp 2 quả. Tìm xác suất nở ra 2 mái? 1 mái và 1 trống?

Câu 2 : Hai người cùng bắn vào một mục tiêu. Xác suất để trúng đích của từng người là: 0,7 và 0,8. Tìm xác suất để:

a.Chỉ có một người bắn trúng.

b.Biết rằng chỉ có 1 người bắn trúng, tìm xác suất để đó là người thứ nhất.

Câu 3 : Một công ty cần tuyển 2 nhân viên. Có 6 người nộp đơn, trong đó có 2 nam và 4 nữ.Biết rằng khả năng được tuyển của mỗi người là như nhau.

a.Tìm xác suất để 2 nữ được chọn, nếu biết rằng có ít nhất 1 nữ đã được chọn

b.Giả sử rằng Hoa là 1 trong 4 nữ. Tìm xác suất để Hoa được chọn nếu biết rằng có ít nhất 1 nữ được chọn.

Câu 4 :Hai công ty A và B cùng kinh doanh 1 mặt hàng. Xác suất công ty A thua lỗ là 0,2. Xác suất để công ty B thua lỗ là 0,4. Tuy nhiên trên thực tế, khả năng cả 2 công ty thua lỗ là 0,1. Tìm xác suất:

a.Để chỉ có 1 công ty thua lỗ?

b.Để có ít nhất 1 công ty làm ăn không thua lỗ?

Câu 5 : Xếp ngẫu nhiên 10 người đi tàu lên 3 toa. Tìm xác suất để:

a.Toa đầu có 3 khách?

b. 1 toa có 3 khách và 1 toa có 4 khách?

Câu 6 :Trong 1 buổi liên hoan có 6 cặp nam nữ, trong đó có 3 cặp là vợ chồng.Chọn ngẫu nhiên ra 3 người. Tìm xác suất để:

a.Chọn được đúng 1 nam?

b.Không chọn được cặp vợ chồng nào?

Câu 7: Có 5 người vào mua hàng ở một cửa hàng có 3 quầy hàng một cách ngẫu nhiên.Tìm xác suất để:

a.Có 3 người cùng vào quầy 1?

b.5 người vào 2 quầy?

Câu 8 : Có 2 hộp bi. Hộp 1 có 6 bi trắng, 4 bi đỏ. Hộp 2 có 5 bi trắng và 5 bi đỏ. Từ hộp 1 lấy ngẫu nhiên ra 1 bi và từ hộp 2 lấy ngẫu nhiên ra 2 bi. Tìm xác suất để:

a.Lấy được đúng 2 bi đỏ?

b.Lấy được bi đỏ?

Câu 9 : Có 2 anh em lái xe ở hai đội xe gồm 5 và 7 người tương ứng. Trong 1 đợt công tác, mỗi đội xe phải chọn ngẫu nhiên ra 3 người chở hàng xuống 3 kho khác nhau và phân công mỗi người chở hàng xuống 1 kho một cách ngẫu nhiên. Tìm xác suất để:

a.Cả anh và em đều đi chở hàng?

b.Anh và em cùng chở hàng xuống kho thứ nhất?

Câu 10 :Một bài thi trắc nghiệm gồm 12 câu hỏi, mỗi câu hỏi cho 5 câu trả lời, trong đó chỉ có một câu trả lời đúng. Giả sử rằng mỗi một câu trả lời đúng được 4 điểm và mỗi câu trả lời sai bị trừ đi 1 điểm. Tìm xác suất để:

a.Người thi được 13 điểm?

b.Người thi bị điểm âm?

Câu 11 : Một người say rượu bước 8 bước, mỗi bước anh ta tiến lên phía trước 1 mét hoặc lui lại phía sau 1mét với xác suất như nhau. Tìm xác suất để sau 8 bước:

a.Anh ta trở lại vạch xuất phát?

b.Anh ta cách điểm xuất phát hơn 1 mét?

Câu 12 : Một sọt cam rất lớn được phân loại theo cách sau: chọn ngẫu nhiên 20 quả cam làm mẫu. Nếu mẫu này không chứa quả cam hỏng nào thì sọt cam được xếp loại 1. Nếu mẫu có 1 hoặc 2 quả hỏng thì xếp loại 2. Trong trường hợp còn lại thì sọt cam được xếp loại 3. Trên thực tế 3% số cam trong sọt bị hỏng. Tìm xác suất để sọt cam được xếp loại:

a.Xếp loại 1

b.Xếp loại 3

Câu 13: Một nhân viên bán hàng mỗi ngày đi bán hàng ở 10 nơi với xác suất bán được hàng ở mỗi nơi đều là 0,2. Tìm xác suất để:

a.Bán được hàng ở ít nhất 8 nơi?

b.Bán được hàng trong 1 ngày.

Câu 14 : Một sinh viên phải thi 3 môn một cách độc lập với nhau. Xác suất để nhận được cùng một điểm số nào đó ở cả 3 môn đều như nhau. Xác suất để thi 1 môn được 8 là 0,18, dưới 8 là 0,65. Xác suất để cả 3 môn đều được điểm 10 là 0,343.10-3. Tính xác suất để sinh viên thi 3 môn được ít nhất 28 điểm. Biết rằng điểm thi cho theo thang điểm 10 và không có điểm lẻ.

Câu 15 :Trong 1 kho rượu, số lượng rượu loại A bằng 2 lần loại B . Chọn ngẫu nhiên 1 chai rượu trong kho và đưa cho 5 người sành rượu nếm thử. Giả sử mỗi người có xác suất chẩn đoán đúng là 0,75. Có 4 người kết luận chai rượu loại A và 1 người kết luận chai rượu loại B. Hỏi xác suất chai rượu thuộc loại A là bao nhiêu?

Câu 16 : Trong trò chơi hái hoa có thưởng có 10 phiếu hoa, trong đó có 5 phiếu hoa có thưởng. Ba người đầu tiên tham gia trò chơi, mỗi người hái 1 phiếu hoa (hoa nào được hái thì sẽ không còn trên cây nữa). Hỏi khả năng hái được hoa có thưởng của mỗi người là bao nhiêu?

Câu 17 : Có 30 xạ thủ được chia thành 4 nhóm. Nhóm 1 gồm 6 người, xác suất bắn trúng đích của mỗi người là 0,8. Nhóm 2 gồm 9 người, xác suất bán trúng đích của mỗi người là 0,7. Nhóm 3 gồm 12 người, xác suất bắn trúng đích của mỗi người là 0,6. Nhóm 4 gồm 3 người, xác suất bắn trúng đích của mỗi người là 0,4. Chọn ngẫu nhiên 1 xạ thủ từ 30 xạ thủ này:

a.Tìm xác suất bắn trúng đích của xạ thủ được chọn ra.

b.Xạ thủ được chọn ra bắn 1 viên đạn và bị trượt. Hỏi xạ thủ này có khả năng thuộc nhóm nào nhất

Câu 18: Một hộp có 10 quả bóng bàn, trong đó có 6 quả mới. Hôm qua nhóm tập đã lấy ra 3 quả để chơi, sau đó trả lại vào hộp. Hôm nay nhóm tập đã lấy ra 3 quả. Tìm xác suất để 3 quả bóng bàn lấy ra hôm nay đều là bóng mới.

Câu 19 : Một công ty tư vấn về địa điểm đặt đại lý bán bếp gas ở một thành phố đã xếp loại các vị trí theo 3 hạng: Tốt, trung bình và kém. Công ty cũng phân loại các đại lý thành 2 loại là tiêu thụ nhanh và tiêu thụ chậm. Theo số liệu, đối với các đại lý tiêu thụ nhanh thì 70% có vị trí tốt, 20% có vị trí trung bình và 10% có vị trí kém. Đối với các đại lý tiêu thụ chậm thì 20% có vị trí tốt, 30% có vị trí trung bình và 50% có vị trí kém. Công ty cũng biết rằng trong tất cả các đại lý có 60% đại lý tiêu thụ nhanh và 40% đại lý tiêu thụ chậm.

a.Chọn ngẫu nhiên 1 đại lý. Tìm xác suất để nó được đánh giá là có vị trí tốt?

b.Chọn ngẫu nhiên 3 đại lý. Tìm xác suất để có ít nhất 1 đại lý tiêu thụ nhanh?

Câu 20 : Một lô hàng gồm 25 sản phẩm loại 1 và 35 sản phẩm loại 2 được đóng gói để gửi cho khách hàng. Nơi nhận thấy thất lạc 1 sản phẩm. Lấy ngẫu nhiên từ lô hàng nhận được ra 1 sản phẩm thì thấy nó là sản phẩm loại 1. Tìm xác suất để sản phẩm bị thất lạc cũng là sản phẩm loại 1.

Câu 21: Trong đám đông, số nam bằng nửa số nữ. Xác suất bị bạch tạng đối với nam là 0,0006; với nữ là 0,000036. Tìm xác suất để:

a.Gặp người bị bạch tạng trong đám đông?

b.Gặp người nam biết rằng người đó bị bạch tạng?

Câu 22: Ở 1 trạm cấp cứu bỏng, có 65% bệnh nhân bỏng do nóng, 35% bênh nhân bỏng do hoá chất. Bị bỏng do nóng có 25% bị biến chứng, bị bỏng do hoá chất có 40% bị biến chứng.

a.Tìm xác suất gặp bệnh nhân không bị biến chứng?

b.Gặp ngẫu nhiên 1 bệnh nhân bị biến chứng, khả năng bệnh nhân đó bị bỏng do nguyên nhân nào nhiều hơn?

Câu 23: Có 2 lô sản phẩm giống hết nhau. Lô 1 có 90% chính phẩm và lô 2 có 80% chính phẩm. Người ta lấy ngẫu nhiên 1 lô và từ đó lấy ngẫu nhiên 1 sản phẩm khác. Tìm xác suất để sản phẩm lần 2 lấy ra là phế phẩm?

Câu 24: Có 2 chuồng thỏ, chuồng 1 có 3 thỏ trắng và 3 thỏ nâu, chuồng thứ 2 có 6 thỏ trắng và 4 thỏ nâu.Bắt ngẫu nhiên 4 con thỏ ở chuồng 1 bỏ sang chuông 2 rồi sau đó lại bắt ngẫu nhiên 1 con thỏ ở chuồng 2 bỏ ra ngoài. Tìm xác suất để bắt được thỏ nâu từ chuồng 2?

Câu 25: Để thi nâng bậc, một công nhân phải chọn ngẫu nhiên 1 trong 3 loại sản phẩm để gia công và phải gia công 5 sản phẩm.Giả sử 1 công nhân có mức độ thành thạo gia công 3 loại sản phẩm trên khác nhau. Cụ thể là xác suất để công nhân đó gia công được sản phẩm đạt tiêu chuẩn tương ứng với 3 loại trên là 0,8; 0,9 và 0,95. Biết rằng sau khi thi người đó đã đỗ. Tìm xác suất người đó chọn đúng sản phẩm mà mình gia công thạo nhất? Biết rằng để đỗ thì trong 5 sản phẩm phải gia công không có sản phẩm nào không đạt tiêu chuẩn.

Câu 26: Hai người cùng bắn vào 1 mục tiêu một cách độc lập với nhau. Xác suất trúng đích của 2 người lần lượt là 0,8 và 0,7. Tìm xác suất để

a.Người thứ nhất bán trúng đích ngay trong 3 phát đầu

b.Ít nhất có một người bắn trúng đích khi mỗi người bắn 1 phát.

CHƯƠNG 3: ĐẠI LƯỢNG NGẪU NHIÊN

3.1.Định nghĩa và phân loại đại lượng ngẫu nhiên.

3.1.1. Định nghĩa:

- ĐLNN là đại lượng mà trong kết quả của phép thử sẽ nhận một và chỉ một trong các giá trị có thể có của nó với một xác suất tương ứng xác định.

-ĐLNN được ký hiệu là: X, Y, Z,… hoặc X1, X2, … còn giá trị có thể có của nó là x, y, z, … hoặc x1, x2, …

-Ta gọi đại lượng X là ĐLNN vì trước khi thực hiện phép thử ta chưa thể biết chắc chắn nó sẽ nhận giá trị bằng bao nhiêu mà chỉ có thể dự đoán điều này với một xác suất nhất định mà thôi.

3.1.2. Ví dụ:

1/. Tung một con xúc xắc.

Gọi X là "số chấm xuất hiện trên mặt con xúc xắc"

X là đại lượng ngẫu nhiên vì trong kết quả của phép thử nó sẽ nhận 1 trong 6 giá trị {1; 2; 3; 4; 5; 6}. Khi đó ta viết:X = {1; 2; 3; 4; 5; 6}.

2/. Có 1 kho đạn, cho một người tập bắn súng lấy đạn từ kho bắn vào bia theo phương thức bắn đến khi nào trúng thì dừng lại.

Gọi X là "số đạn đã bắn".

Khi đó X là đại lượng ngẫu nhiên và X = {1; 2; 3;…; n; …}.

3/. Nghiên cứu tuổi thọ của bóng đèn (giờ)

Gọi X là "tuổi thọ của bóng đèn" thì X nhận giá trị trong khoảng (a; b). X là đại lượng ngẫu nhiên và viết $X \in (a; b)$.

3.1.3. Phân loại ĐLNN

Có 2 loại đại lượng ngẫu nhiên

*3.1.3.1.ĐLNN rời rạc:*Là đại lượng mà nó chỉ nhận một số hữu hạn hoặc vô hạn đếm được các giá trị. (Ta có thể liệt kê được các giá trị của ĐLNN)

Ví dụ: Trong ví dụ1 và 2 ở trên.

3.1.3.2.ĐLNN liên tục: Đại lượng ngẫu nhiên X được gọi là liên tục nếu các giá trị có thể có của nó lấp đầy một khoảng trên trục số , $X \in (a; b)$

Ví dụ: Trong ví dụ 3 ở trên.

3.2. Quy luật phân phối xác suất của ĐLNN.

3.2.1. Định nghĩa: - Bất kỳ một phương thức nào cho phép biểu diễn mối quan hệ giữa các giá trị có thể có của ĐLNN và các xác suất tương ứng của nó đều được gọi là quy luật phân phối xác suất của ĐLNN ấy.

-Trong thực tế người ta thường sử dụng 3 phương pháp để thiết lập quy luật phân phối xác suất của ĐLNN:

+ Bảng phân phối xác suất.

+ Hàm phân phối xác suất.

+ Hàm mật độ xác suất.

3.2.2. Bảng phân phối xác suất:

- Bảng phân phối xác suất chỉ dùng để thiết lập quy luật phân phối xác suất của các ĐLNN rời rạc.

-Giả sử ĐLNN rời rạc X nhận 1 trong các giá trị có thể có là: $x_1, x_2, \dots x_n$ với các xác suất tương ứng là: $p_1, p_2, \dots p_n$. Thì quy luật phân phối xác suất của ĐLNN được mô tả bằng bảng phân phối như sau:

X	x_1	x_2	…	x_n
P	P_1	p_2	…	p_n

Trong đó $0 \le p_i \le 1 \quad \forall i$ và $\sum_{i=1}^{n} p_i = 1$

Ví dụ:

1/. Tung 1 con xúc xắc. Lập bảng phân bố xác suất cho số chấm xuất hiện trên mặt con xúc xắc.

Giải: Gọi X là "số chấm xuất hiện trên mặt con xúc xắc"

=>X chỉ nhận một trong các giá trị có thể có là {1;2;3;4;5;6} với các xác suất tương ứg đều bằng 1/6.

=> X là một ĐLNN rời rạc => X có bảng phân bố xs như sau:

X	1	2	3	4	5	6
P	1/6	1/6	1/6	1/6	1/6	1/6

2/. Tung 1 đồng xu. Lập bảng phân bố xác suất cho số lần xuất hiện mặt sấp.

Giải: Gọi X là "số lần xuất hiện mặt sấp"

=>X chỉ nhận một trong các giá trị có thể có là {0;1} với các xác suất tương ứng đều bằng 1/2.

=> X là một ĐLNN rời rạc => X có bảng phân bố xs như sau:

X	0	1
P	1/2	1/2

3/ Có hai hộp đựng bi, hộp 1 có 4 bi trắng 6 bi đen, hộp 2 có 7 bi trắng 8 bi đen. Lấy ngẫu nhiên ra ngoài 2 bi từ hộp 1, 3 bi từ hộp 2. Lập bảng phân phối xác suất cho số bi trắng được lấy ra

Giải: Gọi X là "số bi trắng được lấy ra từ hộp 1 và hộp 2"

=>X = {0;1;2;3}

Gọi A_i là biến cố "lấy được i bi trắng từ hộp 1" i = {0;1;2}

Gọi B_i là biến cố "lấy được i bi trắng từ hộp 2" i = {0;1}

Ta có:

$$P(X=0)=P(A_0B_0)=P(A_0)P(B_0)=\frac{C_6^2}{C_{10}^2}.\frac{C_8^1}{C_{15}^1}=\frac{8}{45}$$

$$P(X=1)=P(A_1B_0+A_0B_1)=P(A_1)P(B_0)+P(A_0)P(B_1)$$
$$=\frac{C_4^1C_6^1}{C_{10}^2}.\frac{C_8^1}{C_{15}^1}+\frac{C_6^2}{C_{10}^2}.\frac{C_7^1}{C_{15}^1}=\frac{11}{25}$$

$$P(X=2)=P(A_1B_1+A_2B_0)=P(A_1)P(B_1)+P(A_2)P(B_0)$$
$$=\frac{C_4^1C_6^1}{C_{10}^2}.\frac{C_7^1}{C_{15}^1}+\frac{C_4^2}{C_{10}^2}.\frac{C_8^1}{C_{15}^1}=\frac{8}{25}$$

$$P(X=3)=P(A_2B_1)=P(A_2)P(B_1)=\frac{C_4^2}{C_{10}^2}.\frac{C_7^1}{C_{15}^1}=\frac{14}{225}$$

=> X có bảng phân bố xs như sau:

X	0	1	2	3
P	$\frac{8}{45}$	$\frac{11}{25}$	$\frac{8}{25}$	$\frac{18}{225}$

3.2.3. Hàm phân phối xác suất

3.2.3.1.Nhận xét:

Như đã nêu ở trên, ĐLNN rời rạc hoàn toàn xác định khi biết x_i và p_i (i=1÷n). Song đối với ĐLNN liên tục thì ta không thể chỉ ra được hết các giá trị x_i, hơn nữa xác suất để một ĐLNN liên tục nhận một giá trị cho trước = 0. Vì vậy để có thể đặc trưng cho ĐLNN bất kỳ ta dùng khái niệm hàm phân phối xác suất.

Cho X là ĐLNN bất kỳ và xét biến cố (X < x) với x là số thực nào đó. Khi x thay đổi thì P(X < x) cũng thay đổi. Nghĩa là xác suất là một hàm số của x.

3.2.3.2.Định nghĩa:

Cho X là ĐLNN và x là một số thực tùy ý ($x\in X$) thì xác suất của biến cố (X < x) gọi là hàm phân phối xác suất của ĐLNN X.

Ký hiệu: F(x) = P(X < x).

Chú ý: Định nghĩa trên chỉ là tổng quát của hàm phân phối F(x). Đối với từng loại ĐLNN hàm phân phối xác suất được tính theo những công thức riêng biệt khác nhau.

Đặc biệt nếu X là đlnn rời rạc có bảng phân phối xác suất:

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

Ở đây ký hiệu $x_i < x$ dưới dấu $\sum$ có nghĩa là tổng này được lấy theo trị số x_i của đlnn bé hơn x. Cụ thể:

$$F(x) = \begin{cases} 0 & \text{nếu } x \le x_1 \\ p_1 & \text{nếu } x_1 < x \le x_2 \\ p_1 + p_2 & \text{nếu } x_2 < x \le x_3 \\ \cdots\cdots\cdots & \\ 1 & \text{nếu } x > x_n \end{cases}$$

3.2.3.3.Ví dụ: Cho X là đlnn có bảng phân phối xác suất như sau:

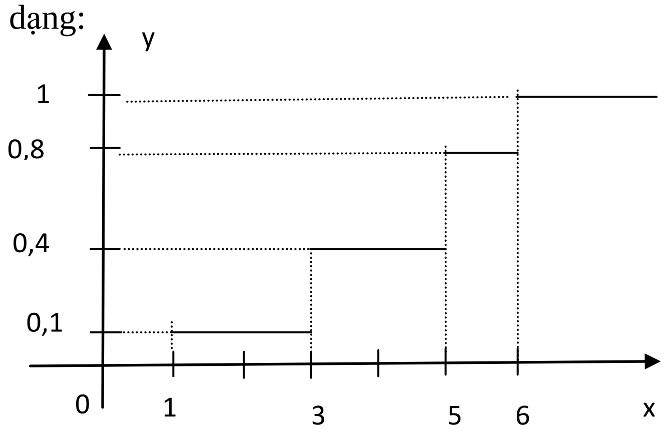
X	1	3	5	6
P	0,1	0,3	0,4	0,2

Tìm hàm phân phối F(x) và vẽ đồ thị của nó.

- Hàm phân phối xs có dạng:

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

- Đồ thị hàm phân phối xác suất có dạng:



=> Đồ thị của hàm phân phối xác suất của ĐLNN rời rạc có dạng bậc thang có số điểm gián đoạn bằng số giá trị có thể có của X.

3.2.3.4.Tính chất:

1/. Hàm phân phối xác suất luôn nhận giá trị trong [0;1]

2/. $F(-\infty) = P(X< -\infty) = P(V) = 0$; $F(+\infty) = P(X< +\infty) = P(U) = 1$

3/. Hàm phân phối xác suất xác suất là hàm không giảm nghĩa là với $x_1 \le x_2$ thì $F(x_1) \le F(x_2)$.

4/. Tính xs của một biến cố khi biết hàm phân phối xs:

+) Nếu X là đlnn rời rạc thì: $P(a \le X < b) = F(b) – F(a)$

+) Nếu X là đlnn liên tục thì:

*) $P(X = x_0) = 0$

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

$= P(a < X < b) = F(b) – F(a)$

5/. Nếu X chỉ nhận giá trị trong [a;b] thì

+) với $x < a$ => $F(x) = 0$; với $x>b$ => $F(x) = 1$

6/. Nếu X là đlnn liên tục thì F(x) liên tục.

Ví dụ: Hàm phân phối xác suất của ĐLNN liên tục X được cho bởi:

$$F(x)=\begin{cases} 0 & \text{nếu} \quad x \le -1 \\ ax + b & \text{nếu} \quad -1 < x \le 3 \\ 1 & \text{nếu} \quad x > 3 \end{cases}$$

a.Tìm a, b và vẽ đồ thị hàm số.

b.Tìm P(0 < X < 1).

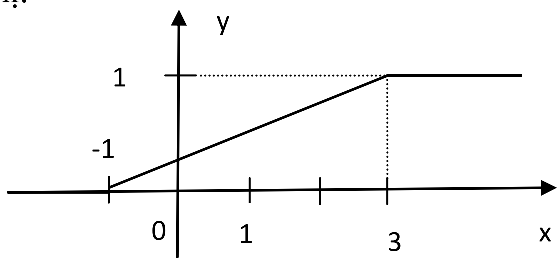
c.Tìm xs để trong 3 phép thử độc lập có đúng 1 lần X nhận giá trị trong (0;1)

Giải:

a.Do F(x) liên tục nên: $\begin{cases} F(-1^-) = F(-1^+) \\ F(3^-) = F(3^+) \end{cases}$ <=> $\begin{cases} -a + b = 0 \\ 3a + b = 1 \end{cases}$ <=> a = b = $\frac{1}{4}$

$$F(x) = \begin{cases} 0 & \text{nếu} \quad x \le -1 \\ \frac{1}{4}x + \frac{1}{4} & \text{nếu} \quad -1 < x \le 3 \\ 1 & \text{nếu} \quad x > 3 \end{cases}$$

Vẽ đồ thị:



b. Tìm: P(0 < x < 1) = F(1) – F(0) = ¼ =0,25

c. Gọi A là biến cố “ Trong 3 phép thử độc lập có đúng 1 lần X nhận giá trị (0;1)”

$$P(A) = P_{n=3}(k=1) = C_3^1(0,25)^1(0,75)^2 = 0,422$$

3.2.4. Hàm mật độ xác suất

3.2.4.1. Nhận xét:

Như đã thấy ở trên hàm phân phối xác suất F(x) hoàn toàn có thể đặc trưng cho đlnn X, nhưng nó lại có nhược điểm là dựa vào nó ta chưa thấy rõ tính chất phân phối xác suất ở lân cận điểm này hay điểm khác trên trục số. Vì vậy đối với ĐLNN liên tục người ta dùng một đặc trưng khác là mật độ xác suất để đặc trưng cho quy luật phân phối xác suất của nó**.**

3.2.4.2. Định nghĩa:

Cho đlnn liên tục X có hàm phân phối xác suất F(x). Nếu hàm phân phối có đạo hàm bậc nhất F’(x) thì đạo hàm này được gọi là hàm mật độ xác suất của ĐLNN X.

Ký hiệu: f(x) = F’(x).

Ví dụ: Tìm hàm mật độ f(x) của ĐLNN X biết hàm phân phối F(x):

$$F(x) = \begin{cases} 0 & \text{nếu } x \le -1 \\ \frac{1}{4}x + \frac{1}{4} & \text{nếu } -1 < x \le 3 \\ 1 & \text{nếu } x > 3 \end{cases}$$

Giải:

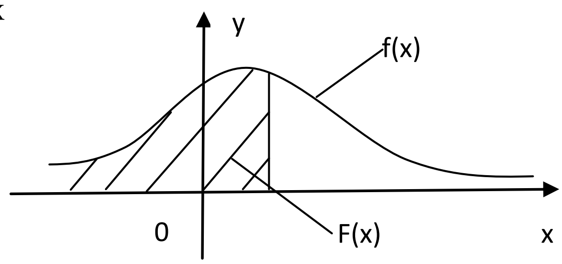
Ta có $f(x) = F'(x) = \begin{cases} 0 & \text{nếu } x \notin (-1;3] \\ \frac{1}{4} & \text{nếu } x \in (-1;3] \end{cases}$

3.2.4.3.Xác định hàm phân phối F(x) biết hàm mật độ f(x):

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

3.2.4.4.Tính xác suất của đlnn X nhận giá trị trong (a;b) khi biết f(x):

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



3.2.4.5.Tính chất hàm mật độ:

+/ $f(x) \ge 0 \quad \forall x$

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

=> *Ý nghĩa:* Toàn bộ diện tích hình phẳng giới hạn bởi đường f(x) và trục Ox bằng 1 đơn vị diện tích.

Ví dụ: Cho đlnn liên tục X có hàm mât độ:

$$f(x) = \begin{cases} 0 & \text{nếu } x \notin [1;3] \\ kx & \text{nếu } x \in [1;3] \end{cases}$$

a.Tìm k

b.Tìm F(x)

c.Tìm xác suất để trong 5 phép thử độc lập X nhận giá trị trong (1;2) ít nhất 1 lần.

Giải:

a. Ta có $\int_{-\infty}^{+\infty} f(x)dx = 1 <=> \int_1^3 kxdx = 1 <=> k = \frac{1}{4}$

$$f(x) = \begin{cases} 0 & \text{nếu } x \notin [1;3] \\ \frac{1}{4}x & \text{nếu } x \in [1;3] \end{cases}$$

b.Tìm F(x)

+) Với x < 1

$$F(x) = \int_{-\infty}^{x} f(x)dx = \int_{-\infty}^{x} 0dx = 0$$

+) Với 1 ≤ x ≤ 3

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

+) Với 1 ≤ x ≤ 3 $\quad = \int_{-\infty}^{1} 0dx + \int_{1}^{x} \frac{1}{4}xdx = \frac{1}{8}x^2 - \frac{1}{8}$

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

$$= \int_{-\infty}^{1} 0dx + \int_{1}^{3} \frac{1}{4}xdx + \int_{3}^{x} 0dx = 1$$

Vậy $F(x) = \begin{cases} 0 & \text{nếu } x \le 1 \\ \frac{1}{8}x^2 - \frac{1}{8} & \text{nếu } 1 < x \le 3 \\ 1 & \text{nếu } x > 3 \end{cases}$

c.Tìm xác suất để trong 5 phép thử độc lập X nhận giá trị trong (1;2) ít nhất 1 lần.

- Xác suất để X nhận giá trị trong (1;2) là:

P(1<X<2) = F(2) – F(1)

$$= \int_{1}^{2} f(x)dx = \frac{3}{8}$$

- Xác suất để trong 5 phép thử độc lập X nhận giá trị trong (1;2) ít nhất 1 lần là:

$$P_{n=5}(k \ge 1) = 1 - P_{n=5}(k = 0) = 1 - C_5^0 \left(\frac{3}{8}\right)^0 \left(\frac{5}{8}\right)^5$$

3.3.Các tham số đặc trưng của ĐLNN

3.3.1. Kỳ Vọng

3.3.1.1.Định nghĩa: Kỳ vọng của đlnn X, ký hiệu: E(X) được định nghĩa riêng cho từng loại đlnn.

+)Nếu X là đlnn rời rạc có bảng phân phối xác suất như sau:

X	x_1	x_2	…	x_n
P	P_1	p_2	…	p_n

Trong đó $0 \le p_i \le 1 \quad \forall i$ và $\sum_{i=1}^{n} p_i = 1$

Thì: $E(X) = x_1p_1 + x_2p_2 + \ldots + x_np_n$

+) Nếu X là đlnn liên tục có hàm mật độ xác suất là f(x) thì:

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

3.3.1.2.Ví dụ

1/ Cho X là đlnn rời rạc có bảng phân phối xác suất như sau:

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

Tìm E(X)

E(X) = -1. 0,1 + 2. 0,3 + 3. 0,2 + 5. 0,4 = 3,1

2/ Cho X là đlnn liên tục có hàm mật độ xác suất là:

$$f(x) = \begin{cases} 0 & \text{nếu } x \notin [1;3] \\ \frac{1}{4}x & \text{nếu } x \in [1;3] \end{cases}$$

Tìm E(X)

Giải:

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx = \int_{1}^{3} x\frac{1}{4}xdx = \frac{1}{12}x^3 \Big|_1^3 = \frac{13}{6}$$

3/ Lớp học có 10 sinh viên. Điểm số môn xác suất thống kê của lớp như sau:

Điểm số	3	5	6	8	9
Số SV	1	2	3	3	1

a. Tìm điểm trung bình của lớp

b. Chọn ngẫu nhiên 1 sinh viên trong lớp ra xem điểm thi. Gọi X là điểm số của sinh viên này. Lập bảng phân phối xác suất cho X và tìm E(X)?

Giải:

a. Điểm trung bình của lớp là:

$$\bar{x} = \frac{1}{10}(1.3 + 2.5 + 3.6 + 3.8 + 1.9) = 6.4 \text{ điểm}$$

b. Ta có bảng phân phối xác suất của X là:

X	3	5	6	8	9
p	0,1	0,2	0,3	0,3	0,1

E(X) = 3.0,1 + 5.0,2 + 6.0,3 + 8.0,3 + 9.0,1 = 6,4

Vậy E(X) chính là điểm số trung bình.

Ý nghĩa: Kỳ vọng của đlnn X chính là giá trị trung bình theo xác suất của đlnn X.

4/ Có hai hộp đựng bi, hộp 1 có 4 bi trắng 6 bi đen, hộp 2 có 7 bi trắng 8 bi đen. Lấy ngẫu nhiên ra ngoài 2 bi từ hộp 1, 3 bi từ hộp 2. Tìm số bi trắng trung bình lấy được.

Giải: Gọi X là "số bi trắng được lấy ra từ hộp 1 và hộp 2"

=>X = {0;1;2;3}

Gọi A_i là biến cố "lấy được i bi trắng từ hộp 1" i = {0;1;2}

Gọi B_i là biến cố "lấy được i bi trắng từ hộp 2" i = {0;1}

Ta có:

$$P(X=0) = P(A_0B_0) = P(A_0)P(B_0) = \frac{C_6^2}{C_{10}^2}.\frac{C_8^1}{C_{15}^1} = \frac{8}{45}$$

$$P(X=1) = P(A_1B_0 + A_0B_1) = P(A_1)P(B_0) + P(A_0)P(B_1)$$
$$= \frac{C_4^1C_6^1}{C_{10}^2}.\frac{C_8^1}{C_{15}^1} + \frac{C_6^2}{C_{10}^2}.\frac{C_7^1}{C_{15}^1} = \frac{11}{25}$$

$$P(X=2) = P(A_1B_1 + A_2B_0) = P(A_1)P(B_1) + P(A_2)P(B_0)$$
$$= \frac{C_4^1C_6^1}{C_{10}^2}.\frac{C_7^1}{C_{15}^1} + \frac{C_4^2}{C_{10}^2}.\frac{C_8^1}{C_{15}^1} = \frac{8}{25}$$

$$P(X=3)=P(A_2B_1)=P(A_2)P(B_1)=\frac{C_4^2}{C_{10}^2}\cdot\frac{C_7^1}{C_{15}^1}=\frac{14}{225}$$

=> X có bảng phân bố xs như sau:

X	0	1	2	3
P	$\frac{8}{45}$	$\frac{11}{25}$	$\frac{8}{25}$	$\frac{14}{225}$

⇨ Số bi trắng trung bình lấy được chính là E(X)

$$E(X) = 0.\frac{8}{45} + 1.\frac{11}{25} + 2\frac{8}{25} + 3.\frac{14}{225} = \frac{59}{45}$$

4/Thời gian xếp chờ mua hàng của khác hàng tại một cửa hàng là đlnn liên tục T có hàm mật độ xác suất là:

$$f(t) = \begin{cases} 0 & \text{nếu } x \notin (0;3) \\ \frac{4}{81}t^3 & \text{nếu } x \in (0;3) \end{cases} \qquad \text{t (phút)}$$

Tìm thời gian xếp hàng trung bình của khách hàng.

Giải:

Thời gian xếp hàng trung bình của khách hàng chính là kỳ vọng toán của T – thời gian xếp chờ mua hàng. Theo định nghĩa :

$$E(T) = \int_{-\infty}^{+\infty} tf(t)dt = \int_0^3 t\frac{4}{81}t^3dt = 2,4phút$$

3.3.1.3. Tính chất:

+) E(c) = c

+) E(c.X) = c.E(X) ; c _ hằng số

+) E(X ± Y) = E(X) ± E(Y)

$$=> E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i)$$

+) Nếu X, Y độc lập: E(X.Y) = E(X).E(Y)

3.3.2. Phương sai

3.3.2.1.Định nghĩa:

Phương sai của đlnn X, ký hiệu D(X) hay V(X), là một số không âm được xác định là kỳ vọng của bình phương sai lệch giữa đlnn với giá trị trung bình của nó, nghĩa là: Nếu đlnn X có E(X) = a thì: $D(X) = E\{(X-a)^2\} = E(X^2) - (E(X))^2$

Cụ thể:

+ Nếu X là đlnn rời rạc thì :

$$D(X) = \sum_{i=1}^{n}(x_i - a)^2 p_i = \sum_{i=1}^{n} x_i^2 p_i - \left(\sum_{i=1}^{n} x_i p_i\right)^2$$

+ Nếu X là đlnn liên tục thì:

$$D(X) = \int_{-\infty}^{+\infty} (x-a)^2 f(x)dx \quad = \int_{-\infty}^{+\infty} x^2 f(x)dx - \left(\int_{-\infty}^{+\infty} xf(x)dx \right)^2$$

3.3.2.2.Tính chất: Phương sai có tính chất sau:

Nếu X, Y độc lập và c là hằng số thì:

+) D(c) = 0

+) $D(X) \geq 0; \forall X$; D(X) = 0 $\Leftrightarrow$ X = c .

+) $D(c.X) = c^2.D(X)$

+) D(X ± c) = D(X)

+) D(X ± Y) = D(X) + D(Y),

Ví dụ: Khảo sát điểm số môn xác suất thống kê của 2 lớp được kết quả như sau:

Gọi X là điểm số của lớp A:

Điểm số (X)	3	5	6	8	9
Số SV	1	2	3	3	1

Gọi Y là điểm số của lớp B:

Điểm số (Y)	5	6	7	8
Số SV	3	2	3	2

Tìm kỳ vọng và phương sai của điểm số của từng lớp

Giải:

- Ta có bảng phân phối xác suất của lớp A là:

X	3	5	6	8	9
p	0,1	0,2	0,3	0,3	0,1

E(X) = 3.0,1 + 5.0,2 + 6.0,3 + 8.0,3 + 9.0,1 = 6,4

$E(X^2) = 3^2.0,1 + 5^2.0,2 + 6^2.0,3 + 8^2.0,3 + 9^2.0,1 = 44$

$D(X) = E(X^2) – [E(X)]^2 = 3,04$

- Ta có bảng phân phối xác suất của lớp B là:

Y	5	6	7	8
p	0,3	0,2	0,3	0,2

E(X) = 5.0,3 + 6.0,2 + 7.0,3 + 8.0,2 = 6,4

$E(X^2) = 5^2.0,3 + 6^2.0,2 + 7^2.0,3 + 8^2.0,2 = 42,2$

$D(X) = E(X^2) – [E(X)]^2 = 1,24$

Nhận xét: Hai lớp trên có điểm số trung bình như nhau nhưng phương sai khác nhau

3.3.2.3.Ý nghĩa của phương sai:

- Giả sử đại lượng ngẫu nhiên X có E(X) = a, ta thấy: X – a là độ lệch khỏi giá trị trung bình nên $D(X) = E\{(X – a)^2\}$ là độ lệch bình phương trung bình. Do đó phương sai phản ánh mức độ phân tán các giá trị của đlnn chung quanh giá trị TB.

- Trở lại ví dụ xét điểm số môn xác suất thống kê của hai lớp ở trên. Ta muốn xem lớp học có "đều" không, nghĩa là các điểm số x_i có tập trung gần điểm trung bình E(X) = a không, ta xét: $\sum(x_i - a)^2 p_i = D(X)$ và mong muốn nó càng nhỏ càng tốt. Nếu D(X) nhỏ thì ta nói các x_i tập trung quanh E(X), D(X) lớn ta nói các x_i phân tán ra xa E(X), do vậy trong ví dụ trên D(X) > D(Y) nên lớp B học đều hơn lớp A.

- Chú ý: Đơn vị đo của phương sai bằng đơn vị đo của X bình phương. Ta hay gặp ký hiệu cho giá trị của phương sai: σ^2.

3.3.3. Độ lệch chuẩn

3.3.3.1.Định nghĩa:

Độ lệch chuẩn của đlnn X, ký hiệu σ(X), được tính là căn bậc hai của phương sai: $\sigma(X) = \sqrt{D(X)}$

=> σ(X) có cùng đơn vị đo với X.

3.3.3.2..Ý nghĩa:

Độ lệch chuẩn có ý nghĩa giống như phương sai, nghĩa là nó phản ánh mức độ tập trung hay phân tán các giá trị của đlnn xung quanh giá trị trung bình E(X).

3.3.4.Mode (giá trị tin cậy nhất) của X

Định nghĩa: Mode của đlnn X, ký hiệu mod(X), là giá trị x_i của X tương ứng với:

Xác suất p_i lớn nhất nếu X là đlnn rời rạc.

Giá trị cực đại của hàm mật độ xác suất nếu X là đlnn liên tục.

Chú ý: Đlnn X có thể có nhiều mod, cũng có thể không có mod

Ví dụ: Tìm mod(X) trong ví dụ sau:

X	5	6	7	8
p	0,1	0,2	0,5	0,2

Ta thấy mod(X) = 7 vì P(X=7) = 0,5 lớn nhất.

3.3.5. Median (Trung vị) của X

Định nghĩa: Median của đlnn X, ký hiệu med(X), là giá trị x_i của X chia phân phối của đlnn X thành hai phần bằng nhau.

Cụ thể:

+) Nếu X là đlnn rời rạc thì med(X) = x_i nếu nó thỏa mãn: $F(x_i) \le 0,5 \le F(x_{i+1})$

+) Nếu X liên tục thì med(X) thỏa mãn:

$$\int_{-\infty}^{med(X)} f(x)dx = 0,5$$

Ví dụ: Cho X là đlnn có bảng phân phối xác suất như sau:

X	1	3	5	6
P	0,1	0,3	0,4	0,2

Tìm mod(X) và med(X)

Giải:- Ta thấy mod(X) = 5 vì P(X=5) = 0,4 lớn nhất.

- Hàm phân phối xác suất có dạng:

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

Med(X) = 5 vì F(5) = 0,4 < 0,5 < F(6) = 0,8

3.4. Một số quy luật phân phối thường gặp

3.4.1. Quy luật phân phối siêu bội

Ví dụ: 1 hộp có 10 bi, trong đó có 4 bi trắng, chọn ngẫu nhiên ra 3 bi từ hộp. Tính xác suất lấy được 2 bi trắng

Giải: Gọi X là số bi trắng lấy được trong 3 bi lấy ra X ={0;1;2;3}

Ta cần tính P(X = 2)

$$P(X = 2) = \frac{C_4^2 . C_6^1}{C_{10}^3}$$

Tổng quát: Giả sử có tập hợp có N phần tử, trong đó có M phần tử có tính chất A nào đó. Lấy ngẫu nhiên ra n phần tử từ tập hợp. Tính xác suất trong n phần tử lấy ra có đúng k phần tử có tính chất A

Giải: Gọi X là số phần tử có tính chất A trong n phần tử lấy ra.

$$P(X=k)=\frac{C_M^k.C_{N-M}^{n-k}}{C_N^M} \qquad (1)$$

Định nghĩa: Đại lượng ngẫu nhiên rời rạc X nhận 1 trong các giá trị có thể có là: 0; 1; …; n với xác suất tương ứng được tính bởi công thức (1) được gọi là tuân theo quy luật phân phối siêu bội. Ký hiệu: X ~ H(N,M,n).

Tính chất: Cho X ~ H(N,M,n) , khi đó ta có:

+) E(X) = np, với $p=\frac{M}{N}$

$+) D(X)=npq\frac{N-n}{N-1}$

$+) \sigma(X)=\sqrt{D(X)}$

Trong đó: $\frac{N-n}{N-1}$ gọi là hệ số hiệu chỉnh

Ví dụ: 1 hộp có 10 bi, trong đó có 4 bi trắng, chọn ngẫu nhiên ra 3 bi từ hộp.

a.Tính số bi trắng lấy được trung bình

b. Tính phương sai của số bi trắng lấy được.

Giải: Gọi X là số bi trắng lấy được trong 3 bi lấy ra.

X={0;1;2;3}. X ~ H (N,M,n) với N = 10; M = 4; n = 3

a. Số bi trắng lấy được trung bình chính là E(X)

E(X) = np =3.4/10 = 12/10 = 1,2

b. Phương sai của số bi trắng lấy được D(X)

$$D(X)=npq\frac{N-n}{N-1}=3\frac{4}{10}\left(1-\frac{4}{10}\right)\frac{10-3}{10-1}=0,56$$

3.4.2. Quy luật phân phối nhị thức

3.3.2.1.Bài toán: Giả sử có n phép thử:

+) Độc lập

+) Trong kết quả của mỗi phép thử chỉ xảy ra 1 trong 2 khả năng: hoặc A xuất hiện hoặc A không xuất hiện.

+) Xác suất xuất hiện biến cố A trong mỗi phép thử đều bằng p, không xuất hiện A đều bằng q = 1-p.

Gọi X là số lần xuất hiện biến cố A trong n phép thử. X = { 0; 1; … ; n }

Bài toán tuân theo lược đồ Bernoulli, xác suất để A xuất hiện đúng k lần trong n phép thử là:

$$P(X=k)=C_n^k p^k q^{n-k} \qquad k=0\div n \qquad (2)$$

3.3.2.2.Định nghĩa: Đại lượng ngẫu nhiên rời rạc X nhận 1 trong các giá trị có thể có là: {0; 1; …; n} với xác suất tương ứng được tính bởi công thức (2) được gọi là tuân theo quy luật phân phối nhị thức với các tham số n và p. Ký hiệu: X ~ B(n,p).

3.3.2.3.Tính chất: Cho X ~ B(n,p) , khi đó ta có:

+) E(X) = np

+) D(X) = npq

+) $\sigma(X) = \sqrt{D(X)}$

+) np – q ≤ mod(X) ≤ np + p; mod(X) $\in$ Z

+) P(k ≤ X ≤ k + h) = $p_k + p_{k+1} + \ldots + p_{k+h}$

3.3.2.3.Ví dụ:

1/ Một nhân viên bán hàng mỗi ngày đi bán ở 10 nơi với xác suất bán được hàng ở mỗi nơi đều bằng 0,2. Vậy nếu 1 năm người đó đi bán hàng 300 ngày thì có trung bình bao nhiêu ngày người đó bán được hàng.

Giải:

Gọi X là số nơi nhân viên bán được hàng trong mỗi ngày, X = {0; 1; …;10}

Việc bán được hàng trong 10 nơi đi bán là độc lập nhau, xác suất bán được mỗi nơi đều bằng 0,2. => X~B(n=10; p=0,2)

Xác suất để nhân viên đó bán được ít nhất 1 nơi là:

P(X ≥ 1) = 1 – P(X = 0) $= 1 - C_{10}^{0}.(0,2)^{0}(0,8)^{10} = 0,8926$

Gọi Y là số ngày nhân viên đó bán được hàng trong 1 năm.

Y = {0; 1; …; 300}. Lập luận tương tự ta có Y~B(300; 0,8926)

=>Số ngày trung bình bán được trong 1 năm chính là E(Y)

E(Y) = np = 300. 0,8926 = 267,78 ngày

2/Một máy sản xuất được 200 sản phẩm một ngày. Xác suất để máy sản xuất ra phế phẩm là 0,05. Tìm số phế phẩm trung bình và số phế phẩm có khả năng tin chắc của máy đó trong 1 ngày.

Giải:

Gọi X là số phế phẩm của máy trong 1 ngày.

=> X ~ B(n=200; p=0,05).

+) Số phế phẩm trung bình của máy trong 1 ngày là:

E(X) = np = 200×0,05 = 10

+) Số phế phẩm tin chắc trong ngày là mod(X). Ta có

$np - q = 200 \times 0,05 - 0,95 = 9,05$

$np + p = 200 \times 0,05 + 0,05 = 10,05$

$\Rightarrow 9,05 \le mod(X) \le 10,05$

Vì X~B(n=200; p=0,05) nên mod(X) $\in$ Z. Do đó mod(X) = 10

3.4.3. Quy luật phân phối Poisson

3.4.3.1.Công thức Poisson:

Giả sử X là đại lượng nhẫu nhiên có phân phối nhị thức với các tham số (n,p) => X ~ B(n, p)

và λ = np trong đó n khá lớn và p khá bé.

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

Khi đó ta có thể thay công thức Bernoulli bởi công thức Poisson:

$$P_n(k) = P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} \qquad (3)$$

3.4.3.2. Định nghĩa:

Đại lượng ngẫu rời rạc X nhận 1 trong các giá trị có thể có là: {0; 1; …; n} với xác suất tương ứng được tính bởi công thức (3) được gọi là tuân theo quy luật phân phối Poisson với tham số λ. Ký hiệu: X ~ P(λ).

3.4.3.3.Tính chất: Cho X ~ P(λ), khi đó ta có:

+) E(X) = λ

+) D(X) = λ

+) $\sigma(X) = \sqrt{\lambda}$

+) $\lambda - 1 \le mod(X) \le \lambda$; mod(X) $\in$ Z

+) $P(k \le X \le k + h) = p_k + p_{k+1} + \ldots + p_{k+h}$

Ví dụ: Một máy sản xuất được 5000 sản phẩm một ngày. Xác suất để máy sản xuất ra phế phẩm là 0,001. Tìm xác suất trong 1 ngày máy sản xuất được không quá 1 phế phẩm

Giải:

Gọi X là số phế phẩm của máy trong 1 ngày.

Do n = 5000 đủ lớn, p = 0,001 đủ nhỏ, $\lambda = np = 5000 \times 0,001 = 5$ nên X ~ P(λ).

Do đó xác suất để có không quá 1 phế phẩm là:

$$P(0 \le X \le 1) = P(X = 0) + P(X = 1)$$

$$= e^{-5}\frac{5^0}{0!} + e^{-5}\frac{5^1}{1!} = 0,04043$$

3.4.4. Quy luật phân phối mũ

3.4.4.1.Định nghĩa: Đại lượng ngẫu nhiên liên tục X được gọi là có phân phối mũ với tham số λ nếu nó có hàm mật độ xác suất là:

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

Ký hiệu: X ~ E(λ).

⇒ Hàm phân phối xác suất của nó là:

$$F(x) = \int_{-\infty}^{x} f(x)dx = \begin{cases} 1 - e^{-\lambda x} & \text{nếu} \quad x > 0 \\ 0 & \text{nếu} \quad x \le 0 \end{cases}$$

3.4.4.2.Các tham số đặc trưng:

Cho X là đlnn có phân phối mũ với tham số λ , X ~ E(λ), khi đó:

+) Kỳ vọng của X là:

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx = \frac{1}{\lambda}$$

+) Phương sai của X là:

$$D(X) = E\left(X^2\right) - \left(E(X)\right)^2 = \frac{1}{\lambda^2}$$

+) Xác suất để X nhận giá trị trong khoảng (a;b) là:

P(a < X < b) = F(b) – F(a)

Ví dụ: Giả sử tuổi thọ (tính bằng năm) của một mạch điện tử trong máy tính là một đlnn có phân phối mũ với kỳ vọng là 6,25. Thời gian bảo hành của mạch điện tử này là 5 năm.

a.Hỏi có bao nhiêu phần trăm mạch điện tử bán ra phải thay thế trong thời gian bảo hành?

b.Muốn tỷ lệ bảo hành chỉ còn 10% thì cần quy định thời gian bảo hành là bao nhiêu?

Giải:

a.Gọi X là tuổi thọ của mạch.

=> X ~ E(λ) với $\lambda = \dfrac{1}{E(X)} = \dfrac{1}{6,25}$

Xác suất để lấy ra 1 mạch điện tử thì phải bảo hành là:

$$P(X<5)=F(5)=1-e^{-\lambda.5}=1-e^{-\frac{5}{6,25}}=0,5506$$

Vậy có khoảng 55% số mạch điện tử bán ra phải thay thế trong thời gian bảo hành.

b.Gọi a là thời gian quy định bảo hành, ta có:

$$P(X<a)=0,1 <=> F(a)=0,1 <=> 1-e^{-\lambda.a}=0,1$$

$$<=> 1-e^{-\frac{a}{6,25}}=0,1 <=> a=-6,25.\ln 0,9$$

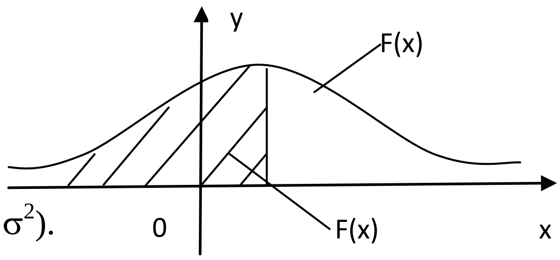
Vậy muốn tỷ lệ bảo hành chỉ còn 10% thì cần quy định thời gian bảo hành là:

a = -6,25ln0,9

3.4.4. Quy luật phân phối chuẩn

3.4.4.1.Định nghĩa: Đlnn liên tục X nhận giá trị trong (- ∞; + ∞) được gọi là có phân phối chuẩn với tham số μ và σ^2 nếu hàm mật độ xác suất của nó có dạng:

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



Trong đó μ, σ là những hằng số. Ký hiệu: X ~ N(μ, σ^2).

=> Hàm phân phối của X ~ N(μ, σ^2) là:

$$F(x)=\int_{-\infty}^{x} f(x)dx=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}dx$$

3.4.4.2.Phân phối chuẩn hóa: Đlnn X gọi là có phân phối chuẩn hóa nếu nó có phân phối chuẩn với tham số μ = 0 và σ^2 = 1. Ký hiệu:X ~ N(0, 1). Hàm mật độ xác suất của nó có dạng: $\varphi(x)=\frac{1}{\sqrt{2\pi}}.e^{-\frac{1}{2}x^2}$

=> Hàm phân phối của X ~ N(μ, σ^2) là:

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

3.4.4.3.Các tham số đặc trưng: Cho X ~ N(μ,σ^2), khi đó:

+) Kỳ vọng của X là: E(X) = μ

+) Phương sai của X là: D(X) = σ^2

+) Độ lệch chuẩn của X là: σ(X) = σ

+) mod(X) = med(X) = μ

3.4.4.4.Tính xác suất: Cho X ~ $N(\mu,\sigma^2)$, khi đó:

1) $P(a < X < b) = \phi\left(\frac{b - \mu}{\sigma}\right) - \phi\left(\frac{a - \mu}{\sigma}\right)$

2) $P(X < b) = \frac{1}{2} + \phi\left(\frac{b - \mu}{\sigma}\right)$

3) $P(a < X) = \frac{1}{2} - \phi\left(\frac{a - \mu}{\sigma}\right)$

4) $P\left(|X-\mu| < \alpha\right) = 2\phi\left(\frac{\alpha}{\sigma}\right)$

5) $P(|X| < \alpha) = \phi\left(\frac{\alpha - \mu}{\sigma}\right) + \phi\left(\frac{\alpha + \mu}{\sigma}\right)$

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

Chú ý: $\phi(x)$ có các tính chất sau:

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

+) Với $x \geq 5$ thì $\phi(x) = 0{,}5$

+) Các giá trị của $\phi(x)$ được tính sẵn bằng bảng phụ lục.

3.4.4.5.Ví dụ:

1/ Thời gian làm xong bài thi của học sinh là một đại lượng ngẫu nhiên có phân phối chuẩn với thời gian trung bình là 110 phút và độ lệch tiêu chuẩn là 20 phút.

a. Xác định tỷ lệ học sinh làm bài xong trong thời gian quy định là 120 phút.

b. Nếu muốn 90% học sinh làm xong bài thi thì phải quy định thời gian là bao nhiêu**?**

Giải: Gọi X là thời gian làm xong bài thi; X ~ $N(110;20^2)$

a.Xác suất để gặp học sinh làm xong bài thi là:

$$P(X < 120) = \frac{1}{2} + \phi\left(\frac{120-\mu}{\sigma}\right) = 0{,}5 + \phi(0{,}5) = 0{,}6915$$

- Vậy tỷ lệ học sinh làm xong bài thi là 69,15%

b. Gọi thời gian làm bài cần quy định là a

$$P(X < a) = 0{,}9 \Leftrightarrow \frac{1}{2} + \phi\left(\frac{a-\mu}{\sigma}\right) = 0{,}9$$

$$\Leftrightarrow \phi\left(\frac{a-110}{20}\right) = 0{,}4 = \phi(1{,}28)$$

$$\Leftrightarrow \frac{a-110}{20} = 1,28 \Leftrightarrow a = 135,6$$

2/Chiều cao của một loại cây là một đại lượng ngẫu nhiên có phân phối chuẩn. Trong 1 mẫu gồm 640 cây có 25 cây thấp hơn 18m và 110 cây cao hơn 24m.

a.Tính chiều cao trung bình của cây và độ lệch tiêu chuẩn?

b.Tìm tỷ lệ cây có chiều cao từ 16m đến 20m?

Giải

Gọi T là chiều cao của cây. T ~ N(μ;σ^2)

$$a. \quad \begin{cases} P\ T<18\ = \dfrac{25}{640} = 0,039 \\ P\ T>24\ = \dfrac{110}{640} = 0,1718 \end{cases} \Leftrightarrow \begin{cases} 0,5+\phi\left(\dfrac{18-\mu}{\sigma}\right) = 0,039 \\ 0,5-\phi\left(\dfrac{24-\mu}{\sigma}\right) = 0,1718 \end{cases}$$

$$\Leftrightarrow \begin{cases} \phi\left(\dfrac{18-\mu}{\sigma}\right) = \phi\ \ -1,76 \\ \phi\left(\dfrac{24-\mu}{\sigma}\right) = \phi\ \ 0,95 \end{cases} \Leftrightarrow \begin{cases} \mu = 21,9 \\ \sigma = 2,22 \end{cases}$$

$$b.\ P\ \ 16<T<20\ = \phi\left(\frac{20-21,9}{2,22}\right) - \phi\left(\frac{16-21,9}{2,22}\right) = 0,191$$

Vậy tỷ lệ cây có chiều cao từ 16m đến 20m là: 19,1%

Bài tập chương 3

<u>Câu 1</u>: Một xạ thủ đem 6 viên đạn để bắn kiểm tra trước ngày thi. Xạ thủ bắn từng viên vào bia với xác suất trúng vòng 10 là 0,85. Nếu bắn được 3 viên liên tiếp trúng vòng 10 thì thôi không bắn nữa. Gọi X là số viên đạn xạ thủ này đã bắn.

a.Lập bảng phân bố xác suất của X

b.Tìm số viên đạn trung bình mà xạ thủ này bắn và số viên đạn có khả năng bắn lớn nhất?

<u>Câu 2</u>: Một túi có 4 quả cầu trắng và 3 quả cầu đen. Hai người A và B lần lượt rút ra 1 quả cầu trong túi (rút xong không trả lại túi). Trò chơi kết thúc khi có người rút được quả cầu đen, người đó xem như thua cuộc và phải trả cho người kia số tiền là số quả cầu đã rút ra nhân với 5USD. Giả sử A là người rút trước và X là số tiền A thu được:

a.Lập bảng phân bố xác suất của X?

b.Tính EX, DX và mod X?

<u>Câu 3</u>: Hai đấu thủ A và B thi đấu cờ. Xác suất thắng của A là 0,4 trong mỗi ván chơi (không có hoà). Ở mỗi ván người thắng được 1 điểm, người thua không được điểm nào. Trận đấu sẽ kết thúc khi hoặc A giành được 3 điểm trước (khi đó A là người thắng), hoặc khi B giành được 5 điểm trước (khi đó B là người thắng).

a.Tìm xác suất thắng của A?

b.Gọi X là số ván cần thiết của toàn bộ trận đấu. Lập bảng phân bố xác suất của X.

<u>Câu 4</u>: Hai xạ thủ A và B tập bắn. Mỗi người bắn hai phát, xác suất bắn trúng đích của A trong mỗi lần bắn là 0,4, còn của B là 0,5. Gọi X là số phát bắn trúng của A trừ đi số phát bắn trúng của B.

a.Tìm phân bố xác suất của X?

b.Tìm phân bố xác suất của $|X|$?

<u>Câu 5</u>: Một thiết bị gồm 3 bộ phận hoạt động độc lập với nhau. Xác suất trong khoảng thời gian T, các bộ phận bị hỏng tưởng ứng là 0,2; 0,3 và 0,25. Gọi X là số bộ phận bị hỏng trong khoảng thời gian T.

a.Tìm quy luật phân phối xác suất của X

b.Tìm $P(0 < X \le 4)$

<u>Câu 6</u>: Một trạm cho thuê xe taxi có 3 chiếc xe, hàng ngày trạm phải nộp thuế 8USD cho 1 chiếc xe (dù xe đó có được thuê hay không), mỗi chiếc xe được cho thuê với giá 20USD. Giả sử số yêu cầu thuê xe của trạm là một đại lượng ngẫu nhiên X có phân phối pat xông với tham số □ =2,8. Gọi Y là số tiền thu được trong ngày của trạm.

a.Lập bảng phân bố xác suất của Y?

b.Tìm số tiền trung bình trạm thu được trong 1 ngày?

<u>Câu 7</u>: Theo thống kê 10000 người Mỹ ở tuổi 25 thì có 8 người chết trong 1năm tới. Một công ty bảo hiểm đề nghị họ mua bảo hiểm sinh mạng cho 1 năm với giá là 10USD và trong trường hợp người đóng bảo hiểm bị chết thì số tiền bồi thường là 1000USD. Hỏi lợi nhuận trung bình của công ty khi bán mỗi thẻ bảo hiểm loại này là bao nhiêu?

<u>Câu 8</u>: Một nhà đầu tư đang cân nhắc việc đầu tư vào 2 dự án A và B trong 2 lĩnh vực độc lập nhau. Khả năng thu hồi vốn sau 2 năm(tính bằng %) của 2 dự án là các đại lượng ngẫu nhiên có phân phối xác suất:

Dự án A:

X	65	67	68	69	70	71	73
P	0,04	0,12	0,16	0,28	0,24	0,08	0,08

Dự án B:

Y	66	68	69	70	71
P	0,12	0,28	0,32	0,2	0,08

Hãy chỉ ra phương án nào có tỷ lệ thu hồi vốn kỳ vọng cao hơn? Phương án nào có độ rủi ro của thu hồi vốn thấp hơn?

<u>Câu 9</u>: Cho X là đại lượng ngẫu nhiên có phân bố mũ với tham số bằng 1. Xét đại lượng ngẫu nhiên Y=2X2. Tìm:

a.P(2 < Y < 18)?

b.P(Y < 4)?

<u>Câu 10</u>: Thời gian đi từ nhà đến trường của sinh viên An là một đại lượng ngẫu nhiên T (đơn vị: phút) có phân phối chuẩn. Biết rằng 65% số ngày An đến trường mất hơn 20 phút và 8% số ngày mất hơn 30 phút.

a.Tính thời gian đến trường trung bình của An và độ lệch tiêu chuẩn?

b.Giả sử An xuất phát từ nhà trước giờ vào học 25 phút. Tìm xác suất để An bị muộn học?

<u>Câu 11</u>: Trong một cuộc thi, có 2 hình thức thi. Hình thức thứ nhất là: mỗi người phải trả lời 2 Câu hỏi, mỗi Câu trả đúng được 5 điểm, hình thức thứ 2 là nếu trả lời đúng Câu thứ nhất mới được trả lời Câu thứ 2, Câu thứ nhất đúng được 5 điểm, Câu thứ 2 đúng được 10 điểm. Trong cả hai hình thức thi, Câu trả lời sai đều không được điểm. Giả sử xác suất trả lời đúng mỗi Câu đều là 0,75 và việc trả lời đúng mỗi Câu là độc lập với nhau. Hãy cho biết nên chọn hình thức thi nào để số điểm trung bình đạt được nhiều hơn?

Câu 12: Chiều cao của một loại cây là một đại lượng ngẫu nhiên có phân phối chuẩn. Trong 1 mẫu gồm 640 cây có 25 cây thấp hơn 18m và 110 cây cao hơn 24m.

a.Tính chiều cao trung bình của cây và độ lệch tiêu chuẩn?

b.Tìm tỷ lệ cây có chiều cao trung bình từ 16m đến 20m?

Câu 13: Tuổi thọ của 1 loại bóng đèn là đại lượng ngẫu nhiên tuân theo phân phối chuẩn với trung bình là 4,2 năm và độ lệch chuẩn là 1,5 năm. Khi bán 1 bóng đèn thì được lãi 100 nghìn đồng, nhưng nếu đèn phải bảo hành thì lỗ 300 nghìn đồng. Vậy để tiền lãi trung bình khi bán mỗi bóng đèn là 30 nghìn đồng thì phải quy định thời gian bảo hành là bao nhiêu?

Câu 14: Một xí nghiệp may quần áo bằng phương pháp thủ công. Giá bán mỗi bộ quần áo là 100.000 đồng. Số bộ quần áo bán được trong tháng là một đại lượng ngẫu nhiên có phân phối chuẩn với kỳ vọng là 530 bộ. Độ lệch tiêu chuẩn là 30. Để sản xuất, hàng tháng xí nghiệp này phải bỏ ra một khoản chi phí cố định là 1.000.000 đồng, các chi phí khác là 85.000 đồng cho một bộ quần áo.

a.Tìm trung bình và độ lệch tiêu chuẩn của lợi nhuận trong 1 tháng

b.Tìm xác suất để xí nghiệp có số lãi ít nhất 6275 nghìn đồng/tháng

Câu 15: Thời gian làm xong bài thi của học sinh là một đại lượng ngẫu nhiên có phân phối chuẩn với thời gian trung bình là 110 phút và độ lệch tiêu chuẩn là 20 phút.

a.Xác định tỷ lệ học sinh là bài xong trong thời gian quy định là 120 phút.

b.Nếu muốn 90% học sinh làm xong bài thi thì phải quy định thời gian là bao nhiêu

Câu 16: Tuổi thọ của một loại bóng đèn của một nhà máy là một đại lượng ngẫu nhiên có phân phối chuẩn vời kỳ vọng là 5000 giờ và độ lệch tiêu chuẩn là 250 giờ. Bóng đèn được gọi là không đạt tiêu chuẩn kỹ nếu tuổi thọ của nó nhỏ hơn 4500 giờ. Lấy ngẫu nhiên 5 bóng đèn loại trên. Tìm xác suất để trong 5 bóng lấy ra có ít nhất 1 bóng không đạt tiêu chuẩn.

Câu 17: chiều dài của chi tiết được gia công trên máy tự động là một đại lượng ngẫu nhiên tuân theo phân phối chuẩn với độ lệch tiêu chuẩn là 0,01 mm. Chi tiết được coi là đạt tiêu chuẩn nếu kích thước thực tế của nó sai lệch so với kích thước trung bình không vượt quá 0,02 mm.

a.Tìm tỷ lệ chi tiết không đạt tiêu chuẩn.

b.Xác định độ đồng đều cần thiết của sản phẩm để tỷ lệ chi tiết không đạt tiêu chuẩn chỉ còn 1%.

Câu 18: Độ dài của chi tiết (tính bằng cm) do một máy tự động sản xuất ra là một đại lượng ngẫu nhiên có phân phối chuẩn với độ lệch tiêu chuẩn là 9cm. Nếu được biết 84,13% chi tiết do máy sản xuất có độ dài không vượt quá 84cm thì xác suất để lấy ngẫu nhiên 3 chi tiết được ít nhất 1 chi tiết có độ dài không vượt quá 80cm là bao nhiêu?

Câu 19: Thời gian hoạt động tốt (không phải sửa chữa) của một loại ti vi là một đại lượng ngẫu nhiên có phân phối chuẩn với $\mu = 4150$ giờ và $\sigma = 250$ giờ. Giả sử mỗi ngày người ta dùng trung bình là 10giờ và thời hạn bảo hành miễn phí là 1năm (365 ngày).

a.Hãy tính tỷ lệ sản phẩm phải bảo hành.

b.Phải nâng chất lượng sản phẩm bằng cách tăng thời gian hoạt động tốt trung bình của sản phẩm lên bao nhiêu để tỷ lệ bảo hành chỉ còn 1%? Giả thiết thời gian bảo hành và σ^2 không thay đổi.

Câu 20: Kết quả thi môn Anh Văn là một đại lượng ngẫu nhiên có phân phối chuẩn với phương sai bằng 100. Xác suất của 1 học sinh có kết quả thi này đạt tối thiểu 85 điểm bằng 0,0668. Hãy tính.

a.Xác suất kết quả thi môn trên của học sinh đạt ít hơn 80 điểm

b.Xác suất để có ít nhất 1 trong 4 sinh viên được chọn ngẫu nhiên có kết quả thi đạt tối thiểu 80 điểm.

Câu 21: Một luật sư kinh tế nhận cãi một vụ kiện cho một doanh nghiệp. Doanh nghiệp này đề ra 2 phương án trả thù lao cho luật sư như sau:

a.Hoặc luật sư nhận trọn gói ngay từ đầu 5 triệu đồng, bất kể kết quả vụ kiện này ra sao.

b.Hoặc luật sư sẽ nhận 15 triệu đồng nếu doanh nghiệp thắng kiên, còn nếu thua kiện thì sẽ chỉ trả 100 nghìn đồng gọi là tiền giấy bút.

Sau khi nghiên cứu hồ sơ của vụ kiện, luật sư đã chấp nhận phương án thứ 2. Nếu vậy luật sư đã nhận định xác suất p để doanh nghiệp thắng kiện tối thiểu là bao nhiêu?

Câu 22: Chiều cao của một loại cây lấy gỗ là một đại lượng ngẫu nhiên có phân phối chuẩn với kỳ vọng là 20 m và độ lệch tiêu chuẩn là 2,5 m. Cây đạt tiêu chuẩn khai thác là cây có chiều cao tối thiểu 15 m. Nếu cây đạt tiêu chuẩn sẽ lãi 10 nghìn đồng, ngược lại sẽ lỗ 50 nghìn đồng. Tìm tiền lãi trung bình khi khai thác cây?

Câu 23: Cho X và Y là hai đại lượng ngẫu nhiên độc lập có phân phối xác suất tương ứng là:

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

Y	-1	0	1
P	0,3	0,4	0,3

Tính EZ và DZ với Z = min(X,Y)

PHẦN II: THỐNG KÊ

CHƯƠNG 4: LÝ THUYẾT MẪU

4.1. Tổng thể, mẫu và phương pháp lấy mẫu

4.1.1. Khái niệm.

Giả sử cần nghiên cứu một tập hợp có rất nhiều phần tử, vì một số lý do mà ta không thể khảo sát toàn bộ tập hợp lớn này (khảo sát tất cả các phần tử) nhưng ta lại muốn có kết quả trên tập hợp lớn. Ta có thể giải quyết như sau:

Từ tập hợp lớn lấy ra một tập hợp nhỏ hơn để nghiên cứu, ta thu được kết quả trên tập hợp nhỏ, từ kết quả tập nhỏ ta suy ra cho kết quả tập hợp lớn.

Phương pháp làm việc như trên gọi là phương pháp mẫu. Tập hợp lớn cần nghiên cứu gọi là tổng thể hay đám đông, số phần tử của tập lớn gọi là kích thước của tổng thể (đám đông) ký hiệu là: N. Tập hợp nhỏ gọi là mẫu, số phân tử của mẫu gọi là kích thước của mẫu hay cỡ mẫu, ký hiệu: n

Với mỗi tổng thể, ta không nghiên cứu trực tiếp tổng thể mà thông qua một số đặc trưng nào đó của nó (các đặc trưng này có thể là định tính, có thể là định lượng). Các đặc trưng cần nghiên cứu được gọi là dấu hiệu nghiên cứu.

*Ví dụ:*Khi nghiên cứu về người dân Việt Nam

- Toàn bộ người dân Việt Nam là một tổng thể.

- Giả sử hiện nay dân số Việt Nam là 87 trệu dân thì kích thước tổng thể là N = 87 triệu.

- Ta không nghiên cứu toàn bộ người dân Việt Nam mà chỉ nghiên cứu ngẫu nhiên 1 triệu người thì 1 triệu người đó được gọi là 1 mẫu nghiên cứu và có kích thước n = 1 triệu.

- Khi nghiên cứu về người dân Việt Nam có rất nhiều đặc trưng cần nghiên cứu như: chiều cao, sức khỏe, thu nhập, …. Nếu ta nghiên cứu về chiều cao thì chiều cao là một dấu hiệu nghiên cứu.

4.1.2. Các lý do không thể nghiên cứu toàn bộ tổng thể.

Để nghiên cứu một tổng thể nào đó người ta ít khi mang toàn bộ các phần tử của tổng thể ra để nghiên cứu vì một số lý do sau:

-Giới hạn về thời gian, tài chính.

Ví dụ: Muốn khảo sát xem chiều cao trung bình của thanh niên Việt Nam hiện nay có tăng lên so với trước đây hay không, ta phải đo chiều cao của toàn bộ thanh niên Việt Nam(Giả sử xấp xỉ N = 45 triệu người), điều này tuy làm được nhưng rõ ràng tốn nhiều thời gian, tiền bạc, công sức….Ta có thể khảo sát khoảng 1 triệu thanh niên và từ chiều cao trung bình của n = 1 triệu người này, ta suy ra chiều cao trung bình của toàn bộ thanh niên Việt Nam.

-Phá vỡ tổng thể nghiên cứu.

Ví dụ: Khi ta cất vào kho 10000 hộp sữa, muốn biết tỷ lệ hộp hỏng trong kho sau 1 thời gian bảo quản thì phải kiểm tra từng hộp để xác định số hộp hỏng M = 500 (tỷ lệ hộp hỏng trong kho khi đó là M/N = 500/10000). Rõ ràng 1 sản phẩm sau khi bị kiểm tra thì bị mát phẩm chất, do đó, khi ta kiểm tra xong cả cái kho thì cũng hỏng luôn cái kho… Ta có thể lấy ngẫu nhiên n = 100 hộp ra kiểm tra, giả sử có 10 hộp hỏng, suy ra tỷ lệ hộp hỏng của cả kho là 10%.

-Không xác định được chính xác tổng thể.

Ví dụ : Như muốn khảo sát xem tỷ lệ những người bị nhiễm HIV qua đường tiêm chích ma túy là bao nhiêu phần trăm. Trong tình huống này thì tổng thể chính là những người bị nhiễm HIV, nhưng ta không thể xác định chính xác những người bị nhiễm HIV vì chỉ có những người tự nguyện đến xét nghiệm thì mới biết được, còn những người không đi xét nghiệm thì không biết được. Do vậy, ta chỉ biết được 1 phần của tổng thể là những người đã đi xét nghiệm. Ngoài ra số người bị nhiễm mới HIV và bị chết bởi HIV có thể thay đổi từng giây nên số phần tử của tổng thể cũng thay đổi từng giây.

4.1.3. Nguyên tắc chọn mẫu

Muốn từ kết quả của mẫu suy ra kết quả cho tổng thể tốt thì mẫu phải đại diện được cho tổng thể, muốn vậy thì mẫu phải được lấy một cách ngẫu nhiên. Mẫu có ngẫu nhiên không thì nó phụ thuộc vào kỹ thuật lấy mẫu, phương pháp lấy mẫu.

a) Kỹ thuật lấy mẫu

- Mẫu giản đơn.
- Mẫu hệ thống.
- Mẫu chùm.
- Mẫu phân tổ.
- Mẫu nhiều cấp.

b) Phương pháp lấy mẫu

- Mẫu có hoàn lại.

- Mẫu không hoàn lại.

4.1.4. Mẫu ngẫu nhiên và mẫu cụ thể.

-Tổng thể được đặc trưng bởi dấu hiệu nghiên cứu X, là 1 ĐLNN. Do vậy, khi nói về X tức là nói về tổng thể.

-Mẫu ngẫu nhiên (có cỡ mẫu là n) được ký hiệu: $W_X = (X_1, X_2, \ldots, X_n)$ là 1 véc tơ có n thành phần, mỗi thành phần X_i là 1 ĐLNN. Các ĐLNN này độc lập với nhau và có cùng quy luật phân phối giống với X.

-Mẫu cụ thể (có cỡ mẫu n) được ký hiệu: $w_X = (x_1, x_2, \ldots, x_n)$ là 1 véc tơ có n thành phần, mỗi thành phần x_i là 1 con số cụ thể.

-Ứng với 1 mẫu ngẫu nhiên thì có nhiều mẫu cụ thể tương ứng với kết quả của các phép thử ngẫu nhiên khác nhau.

Ví dụ: Trong 1 cửa hàng băng đĩa, trên 1 kệ chứa các đĩa nhạc có các loại đĩa tương ứng với số tiền như sau:

Giá(ngàn đồng)	20	25	30	34	40
Số đĩa	35	10	25	17	13

Xét tổng thể về mặt định lượng:

Lấy ngẫu nhiên 1 đĩa nhạc trong giá.

Gọi X là giá tiền của đĩa nhạc này. X = { 20; 25; 30; 34; 40}

Ta thấy X có quy luật phân phối xác suất như sau:

X	20	25	30	34	40
p	0,35	0,1	0,25	0,17	0,13

X đặc trưng cho dấu hiệu nghiên cứu là giá tiền của tổng thể, tổng thể là số đĩa trên giá.

Lấy ngẫu nhiên (có hoàn lại) 4 đĩa nhạc từ kệ.

Gọi X_i là giá tiền của đĩa nhạc thứ i lấy được, $i=1\div4$

Ta thấy các X_i độc lập và có cùng quy luật phân phối xác suất giống như X.

Lập $W_X =(X_1,X_2,X_3,X_4)$, gọi là mẫu ngẫu nhiên.

Bây giờ ta xét xem giá cụ thể của từng đĩa lấy ra thấy như sau:

Đĩa1: giá 20 ngàn đồng

Đĩa2: giá 30 ngàn đồng

Đĩa3: giá 20 ngàn đồng

Đĩa4: giá 40 ngàn đồng

Lập $w_x = (x_1,x_2,x_3,x_4) = (20,30,20,40)$ gọi là mẫu cụ thể

Bây giờ ta xét tổng thể về mặt định tính:

Đĩa có giá dưới 25 ngàn đồng là đĩa "lậu".

Lấy ngẫu nhiên 1 đĩa từ kệ.

Gọi Y là số đĩa lậu lấy được. Y = {0; 1}

Ta thấy Y có quy luật phân phối xác suất như sau:

Y	0	1
p	0,65	0,35

Y đặc trưng cho dấu hiệu nghiên cứu là đĩa lậu hay không lậu của tổng thể là số đĩa trên giá.

Lấy ngẫu nhiên (có hoàn lại) 4 đĩa nhạc từ kệ.

Gọi Y_i là số đĩa lậu lấy được khi lấy 1 đĩa ở lần lấy thứ i, $i=1\div4$

Ta thấy các Y_i độc lập và có cùng quy luật phân phối xác suất giống như Y.

Lập $W_Y =(Y_1,Y_2,Y_3,Y_4)$, gọi là mẫu ngẫu nhiên.

Bây giờ ta xét xem giá cụ thể của từng đĩa lấy ra thấy như sau:

Đĩa1: giá 20 ngàn đồng => $y_1 = 1$

Đĩa2: giá 30 ngàn đồng => $y_2 = 0$

Đĩa3: giá 20 ngàn đồng => $y_3 = 1$

Đĩa4: giá 40 ngàn đồng => $y_4 = 0$

Lập $w_y = (y_1,y_2,y_3,y_4) = (1, 0, 1, 0)$ gọi là mẫu cụ thể.

4.2.Các tham số đặc trưng.

4.2.1.Các tham số đặc trưng của tổng thể.

- Ta xét tổng thể về mặt định lượng: tổng thể được đặc trưng bởi dấu hiệu nghiên cứu X, X là ĐLNN.

+) $E(X) = \mu$ là trung bình tổng thể.

+) $D(X) = \sigma^2$ là phương sai tổng thể.

+) σ là độ lệch chuẩn của tổng thể.

- Ta xét tổng thể về mặt định tính: tổng thể có kích thước N, trong đó có M phần tử có tính chất A quan tâm. Khi đó, p = M/N gọi là tỷ lệ tổng thể.

Chú ý: Các tham số trên gọi là các tham số lý thuyết của tổng thể. Vì thực chất ta không thể biết được chính xác mà ta chỉ biết được nó thông qua các tham số của mẫu (ước lượng).

4.2.2. Các tham số đặc trưng của mẫu ngẫu nhiên.

Cho mẫu ngẫu nhiên $W_X = (X_1, X_2, \ldots, X_n)$. Khi đó

-Trung bình mẫu ngẫu nhiên được xác định là

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

-Phương sai mẫu ngẫu nhiên được xác định là

$$S^2 = \frac{(X_1 - \overline{X})^2 + (X_2 - \overline{X})^2 + \ldots + (X_n - \overline{X})^2}{n}$$

-Phương sai hiệu chỉnh của mẫu ngẫu nhiên được xác định là:

$$S'^2 = \frac{n}{n-1} S^2$$

-Độ lệch tiêu chuẩn của mẫu ngẫu nhiên là: $S = \sqrt{S^2}$

-Độ lệch tiêu chuẩn hiệu chỉnh của mẫu ngẫu nhiên là: $S' = \sqrt{S'^2}$

-Tỷ lệ mẫu: f = m/n

trong đó: m số lần xuất hiện A trong mẫu, n – kích thước mẫu

4.2.3. Các tham số đặc trưng của mẫu cụ thể.

4.2.3.1.Nếu mẫu cụ thể có dạng w = ($x_1, \ldots, x_n$). Khi đó:

- Trung bình mẫu: $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

- Phương sai mẫu: $s^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \overline{x})^2$

- Phương sai hiệu chỉnh: $s'^2 = \frac{n}{n-1} s^2$

- Độ lệch chuẩn: $s = \sqrt{s^2}$

- Độ lệch chuẩn hiệu chỉnh: $s' = \sqrt{s'^2} = \sqrt{\frac{n}{n-1} s^2}$

-Tỷ lệ mẫu: f = m/n

trong đó: m số lần xuất hiện A trong mẫu, n – kích thước mẫu.

Ví dụ: Nghiên cứu về mức tiêu thụ điện năng (kw) của một khu dân cư, ta nghiên cứu mức tiêu thụ điện năng của 5 hộ gia đình trong một khu dân cư đó thì thu được kết quả sau: 75; 123; 97; 145; 215

a. Tính trung bình số điện tiêu thụ của các hộ gia đình. Tìm phương sai, độ lệch tiêu chuẩn hiệu chỉnh của số điện tiêu thụ.

b. Nếu quy ước những hộ có mức tiêu thụ ít hơn hoặc bằng 123 kw thì được coi là tiêu thụ điện bình thường. Tìm tỷ lệ các hộ tiêu thụ điện bình thường.

Giải

Gọi X là mức tiêu thụ điện của các hộ gia đình

a. - Trung bình số điện tiêu thụ của các hộ gia đình là:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{5}(75+123+97+145+215) = 131$$

- Phương sai:

$$s^2 = \frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}} \left(x_{\mathrm{i}} - \overline{x}\right)^2 = \frac{1}{5}\left[(-56)^2 + (-8)^2 + (-34)^2 + (14)^2 + (84)^2\right] = 2321,6$$

- Phương sai hiệu chỉnh:

$$\mathrm{s'}^2 = \frac{\mathrm{n}}{\mathrm{n\text{-}1}} s^2 = 2902$$

- Độ lệch tiêu chuẩn:

$$\mathrm{s} = \sqrt{s^2} = 40,183$$

- Độ lệch tiêu chuẩn hiệu chỉnh:

$$\mathrm{s'} = \sqrt{\mathrm{s'}^2} = \sqrt{\frac{n}{n-1} s^2} = 53,87$$

b. Tỷ lệ hộ tiêu thụ điện bình thường:

$$\mathrm{f} = \frac{\mathrm{m}}{\mathrm{n}} = \frac{3}{5}$$

4.2.3.1.Trường hợp mẫu cho dưới dạng bảng:

x_i	x_1	x_2	….	x_k
n_i	n_1	n_2	….	n_k

với $n = \sum_{i=1}^{k} n_i$ Khi đó:

+) Giá trị trung bình của mẫu cụ thể là:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i$$

+) Phương sai của mẫu cụ thể là:

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

+) Độ lệch tiêu chuẩn của mẫu cụ thể là:

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

+) Phương sai hiệu chỉnh của mẫu cụ thể là:

$$s'^2 = \frac{n}{n-1}s^2$$

+) Độ lệch tiêu chuẩn hiệu chỉnh của mẫu cụ thể là:

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

+) Tỷ lệ mẫu: $f = \frac{m}{n}$

trong đó, n là cỡ mẫu; m là số phần tử trong mẫu có tính chất A quan tâm.

Ví dụ: Kiểm tra thể lực của một nhóm sinh viên, ta có kết quả về cân nặng như sau:

Kg	45	50	55	60	65
Số sv	8	14	28	18	12

a. Tìm cân nặng trung bình của nhóm sinh viên trên. Phương sai, phương sai hiệu chỉnh, độ lệch tiêu chuẩn và độ lệch tiêu chuẩn hiệu chỉnh.

b. Tìm tỷ lệ sinh viên có cân nặng lớn hơn hoặc bằng 55 kg.

Giải: Gọi X là cân nặng sinh viên

a. - Cân nặng trung bình của nhóm sinh viên trên: $\overline{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i = 55,75$

- Phương sai: $s^2 = \left(\frac{1}{n}\sum_{i=1}^{5} n_i x_i^2\right) - \left(\overline{x}\right)^2 = 34,46$

- Phương sai hiệu chỉnh: $s'^2 = \frac{n}{n-1}s^2 = 34,93$

- Độ lệch tiêu chuẩn: $s = \sqrt{s^2} = 5,87$

- Độ lệch tiêu chuẩn hiệu chỉnh: $s' = \sqrt{s'^2} = \sqrt{\frac{n}{n-1}s^2} = 5,91$

b. Tỷ lệ sinh viên có cân nặng lớn hơn hoặc bằng 55 kg:

$$f = \frac{m}{n} = \frac{58}{80} = 0,725$$

4.2.3.1.Trường hợp mẫu cho dưới dạng khoảng

Để tính các tham số đặc trưng cho mẫu dưới dạng này, ta lập bảng sau:

$(x_i;x_{i+1})$	$(x_1;x_2)$	$(x_2;x_3)$	…	$(x_k;x_{k+1})$
n_i	n_1	n_2	….	n_k

Để tính các tham số đặc trưng cho mẫu dưới dạng này, ta lập bảng sau:

$(x_i ; x_{i+1})$	n_i	$x_i^o = \frac{x_i + x_{i+1}}{2}$	$n_i x_i^o$	$n_i x_i^{02}$
$(x_1 ; x_2)$	n_1	x_1^0	$n_1 x_1^o$	$n_1 x_1^{02}$
$(x_2 ; x_3)$	n_2	x_2^0	$n_2 x_2^o$	$n_2 x_2^{02}$
…	…	…	….	…
$(x_k; x_{k+1})$	n_k	x_k^0	$n_k x_k^o$	$n_k x_k^{02}$

Khi đó:

với $n = \sum_{i=1}^{k} n_i$ Khi đó:

+) Giá trị trung bình của mẫu cụ thể là: $\bar{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i^0$

+) Phương sai của mẫu cụ thể là:

$$s^2 = \frac{1}{n}\sum_{i=1}^{k} n_i \; x_i^0 - \bar{x} \;{}^2 = \frac{1}{n}\sum_{i=1}^{k} n_i x_i^{02} - \left(\frac{1}{n}\sum_{i=1}^{k} n_i x_i^0\right)^2$$

+) Độ lệch tiêu chuẩn của mẫu cụ thể là:

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

+) Phương sai hiệu chỉnh của mẫu cụ thể là: $s'^2 = \frac{n}{n-1} s^2$

+) Độ lệch tiêu chuẩn hiệu chỉnh của mẫu cụ thể là: $s' = \sqrt{s'^2}$

+) Tỷ lệ mẫu: $f = \frac{m}{n}$

trong đó, n là cỡ mẫu; m là số phần tử trong mẫu có tính chất A quan tâm.

Bài tập chương 4:

Bài 1:Theo dõi sự phát triển chiều cao của một loại cây sau một năm tuổi ta có:

X (cm)	250-300	300-350	350-400	400-450	450-500	500-550	550-600
Số cây	5	20	25	30	28	23	14

Tìm chiều cao trung bình, độ lệch tiêu chuẩn hiệu chỉnh của cây sau một năm tuổi.

Quy ước những cây cao không quá 3,5 m là cây chậm lớn, tìm tỷ lệ cây chậm lớn.

Bài 2: Điều tra chiều dài của một loại sản phẩm (đơn vị tính bằng cm) thì thu được kết quả sau:

Xi	80-95	95-110	110-125	125-140	140-155	155-170	170-185
Ni	13	19	23	30	28	23	14

Hãy tìm chiều dài trung bình, độ lệch tiêu chuẩn hiệu chỉnh của sản phẩm.

Nếu quy ước những sản phẩm có chỉ tiêu từ 155 cm trở lên là sản phẩm ngoại cỡ, tìm tỷ lệ sản phẩm ngoại cỡ.

Bài3: Kết quả quan sát về hàm lượng vitamin C của một loại trái cây cho trong bảng sau:

Hàm lượng (%)	6-7	7-8	8-9	9-10	10-11	11-12
số trái	5	10	20	35	25	5

Hãy tìm hàm lượng vitamin C trung bình, độ lệch tiêu chuẩn hiệu chỉnh.

Nếu quy ước những trái loại A là những trái có hàm lượng vitamin C trên 10%, tìm tỷ lệ trái loại A

Bài 4: Kiểm tra lượng xăng hao phí trên đoạn đường từ A đến B, người ta theo dõi một số ôtô chạy trên đoạn đường này và thu được kết quả:

X(Lit)	9,5-9,7	9,7-9,9	9,9-10,1	10,1-10,3	10,3-10,5
Ni	7	9	15	11	8

Hãy tìm lượng xăng hao phí trung bình, độ lệch tiêu chuẩn hiệu chỉnh.

Nếu quy ước những xe có lượng xăng hao phí từ 10,1 lít trở lên là xe tốn xăng, tìm tỷ lệ xe tốn xăng.

Bài 5:Khi kiểm tra thể lực của một nhóm sinh viên, ta có kết quả về cân nặng như sau:

X(kg)	42,5-47,5	47,5-52,5	52,5-57,5	57,5-62,5	62,5-67,5
Số sinh viên	8	14	28	18	12

Hãy tìm trọng lượng trung bình, độ lệch tiêu chuẩn hiệu chỉnh.

Tìm tỷ lệ sinh viên có trọng lượng ít hơn hoặc bằng 52.5 kg.

Bài 6: Nghiên cứu chiều cao X(cm) của học sinh tiểu học, ta có kết quả sau:

X(cm)	120-125	125-130	130-135	135-140	140-145
Ni	9	16	34	30	19

Hãy tìm chiều cao trung bình, độ lệch tiêu chuẩn hiệu chỉnh.

Tìm tỷ lệ các cháu có chiều cao trung bình lớn hơn hoặc bằng 140 cm.

Bài 7: Để nghiên cứu sự phát triển của một loại giống cây trồng, người ta tiến hành đo đường kính X(cm) và thu được kết quả:

X(cm)	20-22	22-24	24-26	26-28	28-30
Ni	7	14	33	27	19

Hãy tìm đường kính trung bình, độ lệch tiêu chuẩn hiệu chỉnh.

Tìm tỷ lệ cây đường kính trung bình lớn hơn hoặc bằng 26 cm

Bài 8:Theo dõi số kẹo bán được (kg) trong 1 tuần lễ tại một xí nghiệp ta có kết quả:

X(kg)	0-50	50-100	10-150	150-200	200-250	250-300	300-350
số tuần	9	23	27	30	25	20	5

Hãy tìm số kẹo trung bình bán được trong 1 tuần, độ lệch tiêu chuẩn hiệu chỉnh.

Tìm tỷ lệ những tuần bán được lớn hơn hoặc bằng 250 kg

Bài 9: Để nghiên cứu nhu cầu X của một loại hàng hoá trong 1 tháng, người ta tiến hành điều tra ở 100 gia đình và thu được kết quả sau:

X(kg)	1-2	2-3	3-4	4-5	5-6	6-7	7-8
ni	10	13	22	25	18	15	7

a.Tìm nhu cầu trung bình về mặt hàng này và độ lệch tiêu chuẩn hiệu chỉnh

b.Hãy tìm tỷ lệ gia đình có nhu cầu ít nhất 5 kg về mặt hàng này.

CHƯƠNG 5: ƯỚC LƯỢNG THAM SỐ

5.1. Đặt vấn đề

Giả sử cần nghiên cứu dấu hiệu nào đó của tổng thể và dấu hiệu đó được mô hình hóa bằng đlnn X. Như đã biết một trong những mục tiêu cơ bản của việc nghiên cứu là xác định các tham số đặc trưng của tổng thể, đó là:

- Trung bình tổng thể: $E(X) = \mu$
- Phương sai tổng thể: $D(X) = \sigma 2$
- Độ lệch chuẩn tổng thể: $\sigma(X) = \sigma$
- Tỷ lệ tổng thể: $p = M/N$

Ta gọi chung các tham số này của tổng thể là θ, θ là một số cố định nhưng chưa biết của tổng thể, ta cần phải ước lượng hay chính là xác định một cách gần đúng giá trị của θ dựa trên các thông tin thu được từ mẫu quan sát của X. Quá trình này gọi là quá trình ước lượng tham số. Có hai dạng ước lượng cơ bản là ước lượng điểm và ước lượng khoảng.

5.2. Ước lượng điểm

5.2.1. Định nghĩa:

Quá trình ước lượng tham số θ được gọi là ước lượng điểm nếu ta dùng một số θ* nào đó để ước lượng cho θ.

Chú ý: θ* thường là các tham số mẫu

5.2.2. Một số tính chất:

5.2.2.1.Ước lượng chệch, không chệch

Tham số θ* của một mẫu được gọi là ước lượng chệch của tham số θ của đlnn X nếu $E(\theta^*) \neq \theta$ và là ước lượng không chệch của θ nếu $E(\theta^*) = \theta$.

Ta dễ dàng chứng minh được rằng:

+ là ước lượng không chệch của $E(X) = \mu$.

+ S’2 là ước lượng không chệch của $D(X) = \sigma 2$.

+ f là ước lượng không chệch của p.

+ S2 là ước lượng chệch của $D(X) = \sigma 2$.

5.2.2.2.Ước lượng hiệu quả

Tham số θ* của một mẫu được gọi là ước lượng hiệu quả nhất của tham số θ của tổng thể nếu nó là ước lượng không chệch của tham số θ và là ước lượng không chệch có phương sai nhỏ nhất trong số các ước lượng không chệch của θ.

Tức là: $\begin{cases} E(\theta^*) = \theta \\ D(\theta^*) \longrightarrow \min \end{cases}$

Ngoài hai tính chất trên ra còn một số tính chất như: ước lượng vững, ước lượng hợp lý tối đa, …

5.3.Ước lượng khoảng

5.3.1. Định nghĩa:

Quá trình ước lượng tham số θ được gọi là ước lượng khoảng nếu như ta dùng một khoảng (θ* - ε; θ* + ε) nào đó để ước lượng cho θ.

- (θ* - ε; θ* + ε) gọi là khoảng tin cậy đối xứng của θ.
- ε gọi là sai số ước lượng (hay còn gọi là độ chính xác).
- I = 2ε gọi là độ dài khoảng tin cậy.
- Số γ thỏa mãn P (θ* - ε < θ < θ* + ε) = γ thì γ gọi là độ tin cậy.

Chú ý: Ta có các dạng ước lượng cơ bản sau:

+ Ước lượng giá trị trung bình.

+ Ước lượng tỷ lệ.

5.3.2.Ước lượng giá trị trung bình E(X) = μ

Giả sử dấu hiệu nghiên cứu của tổng thể được mô hình hóa bằng đlnn X và X ~ N(μ; σ^2). Với độ tin cậy γ cho trước thì khoảng tin cậy đối xứng cho μ là:

$$\mu \in \left(\overline{x} - \varepsilon; \overline{x} + \varepsilon\right)$$

Trong đó: $\overline{x}$: là trung bình mẫu.

ε : là sai số ước lượng.

I = 2ε: là độ dài khoảng tin cậy.

5.3.2.1. Nếu biết σ (độ lệch chuẩn tổng thể) khi đó

-Với độ tin cậy γ cho trước thì sai số ước lượng: $\varepsilon = \dfrac{\sigma}{\sqrt{n}} u(\dfrac{\alpha}{2})$

Trong đó $\alpha = 1 - \gamma$ và $u\left(\frac{\alpha}{2}\right)$ được xác định bằng $\phi\left(u(\frac{\alpha}{2})\right) = \frac{\gamma}{2}$

-Xác định độ tin cậy γ nếu biết sai số ước lượng là ε_0

$$\gamma = 2\Phi\left(u(\frac{\alpha}{2})\right) \text{ và } \quad u(\frac{\alpha}{2}) = \frac{\varepsilon_0\sqrt{n}}{\sigma}$$

Cỡ mẫu tối thiểu sao cho độ tin cậy γ cho trước, độ dài khoảng tin cậy không vượt quá giá trị I_0 (hoặc sai số ước lượng không vượt quá giá trị ε_0 cho trước).

$$n_{tt} = \left[\left(\frac{2\sigma}{I_0}u(\frac{\alpha}{2})\right)^2\right] + 1 \quad \text{hoặc} \quad n_{tt} = \left[\left(\frac{\sigma}{\varepsilon_0}u(\frac{\alpha}{2})\right)^2\right] + 1$$

Trong đó [a] là lấy phần nguyên của a

Ví dụ: Trọng lượng của một loại sản phẩm là đlnn tuân theo quy luật phân phối chuẩn với độ lệch tiêu chuẩn là 3 kg. Cân thử ngẫu nhiên các sản phẩm này thì thu được kết quả là:

X(kg)	42,5-47,5	47,5-52,5	52,5-57,5	57,5-62,5	62,5-67,5
Số sản phẩm	8	14	28	18	12

a. Hãy ước lượng cho trọng lượng trung bình của số sản phẩm trên với độ tin cậy 96%.

b. Nếu muốn sai số ước lượng trên là 0,5 kg thì cần đảm bảo độ tin cậy là bao nhiêu.

c. Muốn ước lượng trọng lượng trung bình của số sản phẩm với độ chính xác là 0,5 kg và độ tin cậy 98% thì cần theo dõi thêm nữa không?

Giải: Gọi X là trọng lượng của sản phẩm, $X \sim N(\mu; \sigma^2)$ với $\sigma = 3$

a. Ước lượng cho trọng lượng trung bình của số sản phẩm với độ tin cậy 96%.

Tức là tìm $\bar{x}$ và ε để $\mu \in (\bar{x} - \varepsilon; \bar{x} + \varepsilon)$

Tìm $\bar{x}$:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i^0 = 55,75$$

Ta có

$$\varepsilon = \frac{\sigma}{\sqrt{n}}u(\frac{\alpha}{2}) = \frac{3}{\sqrt{80}}.2,05 = 0,688$$

- Tìm $u(\alpha/2)$:

Ta có $\gamma = 0,96 \Rightarrow \alpha = 1 - \gamma = 1 – 0,96 = 0,04$

$\Rightarrow$ $u(\alpha/2) = u(0,02)$ => $\phi(u(0,02)) = 0,48$ => $u(0,02) = 2,05$

- Tìm sai số ước lượng:

- Vậy khoảng tin cậy đối xứng trọng lượng trung bình của số sản phẩm trên là :

$\bar{x} - \varepsilon; \bar{x} + \varepsilon$ =(55,062; 56,438)

b. Nếu muốn sai số ước lượng là 0,5 kg thì cần đảm bảo độ tin cậy là bao nhiêu?

- Tức là tìm γ biết $\varepsilon = 0,5$

- Ta có: $\varepsilon = \dfrac{\sigma}{\sqrt{n}} u(\alpha/2) = 0,5$

$$\Rightarrow u(\alpha/2) = \frac{0,5.\sqrt{n}}{\sigma} = \frac{0,5.\sqrt{80}}{3} = 1,491$$

$\gamma = 2.\phi(u(\alpha/2)) = \phi(1,491) = 2.0,4319$

$\Rightarrow \gamma = 0,8638$

Vậy muốn sai số ước lượng là 0,5 kg thì cần đảm bảo độ tin cậy là 86,38%

c. Muốn ước lượng trọng lượng trung bình của sản phẩm với độ chính xác là 0,5 kg và độ tin cậy 98% thì cần theo dõi thêm nữa không?

Ta có cỡ mẫu tối thiểu n sao cho với độ tin cậy γ và độ chính xác ε_0 là:

$$n_{tt} = \left[\left(\frac{\sigma}{\varepsilon_0} u(\frac{\alpha}{2})\right)^2\right] + 1$$

Trong đó: $\sigma = 3$

$\gamma = 0,98 = 1 - \alpha \Rightarrow \alpha = 0,02 \Rightarrow u(\alpha/2) = u(0,01) = 2,32$

$\varepsilon_0 = 0,5$

Vậy: $$n_{tt} = \left[\left(\frac{\sigma}{\varepsilon_0} u(\frac{\alpha}{2})\right)^2\right] + 1 = \left[\left(\frac{3}{0,5} 2,32\right)^2\right] + 1 = 194$$

=> Có phải điều tra thêm 194 – 80 = 114 sản phẩm nữa

5.3.2.2. Nếu chưa biết σ và kích thước mẫu $n \geq 30$:

-Với độ tin cậy γ cho trước thì sai số ước lượng: $\varepsilon = \dfrac{s'}{\sqrt{n}} u(\dfrac{\alpha}{2})$

Trong đó: s' là độ lệch chuẩn hiệu chỉnh của mẫu

$\alpha = 1 - \gamma$ và $u\left(\dfrac{\alpha}{2}\right)$ được xác định bằng $\phi\left(u(\dfrac{\alpha}{2})\right) = \dfrac{\gamma}{2}$

-Xác định độ tin cậy γ nếu biết sai số ước lượng là ε_0

$$\gamma = 2\Phi\left(u(\frac{\alpha}{2})\right) \quad \text{và} \quad u(\frac{\alpha}{2}) = \frac{\varepsilon_0\sqrt{n}}{s'}$$

-Cỡ mẫu tối thiểu sao cho độ tin cậy γ cho trước, độ dài khoảng tin cậy không vượt quá giá trị I_0 (hoặc sai số ước lượng không vượt quá giá trị ε_0 cho trước).

$$n_{tt} = \left[\left(\frac{2s'}{I_0}u(\frac{\alpha}{2})\right)^2\right] + 1 \quad \text{hoặc} \quad n_{tt} = \left[\left(\frac{s'}{\varepsilon_0}u(\frac{\alpha}{2})\right)^2\right] + 1$$

5.3.2.3. Nếu chưa biết σ và kích thước mẫu n < 30:

-Với độ tin cậy γ cho trước thì sai số ước lượng: $\varepsilon = \frac{s'}{\sqrt{n}}.t^{\,n-1}(\alpha/2)$

Chú ý: $t^{(n-1)}(\alpha/2)$ là giá trị tới hạn của phân phối Student và nó được xác định bởi bảng phụ lục cuối sgk.

-Cỡ mẫu tối thiểu sao cho độ tin cậy γ cho trước, độ dài khoảng tin cậy không vượt quá giá trị I_0 (hoặc sai số ước lượng không vượt quá giá trị ε_0 cho trước).

$$n_{tt} = \left[\left(\frac{2s'}{I_0}t^{n-1}(\frac{\alpha}{2})\right)^2\right] + 1 \quad \text{hoặc} \quad n_{tt} = \left[\left(\frac{s'}{\varepsilon_0}t^{n-1}(\frac{\alpha}{2})\right)^2\right] + 1$$

5.3.3.Ước lượng tỷ lệ (Ước lượng cho p)

Gọi p là xác suất xuất hiện biến cố A của tổng thể.Khi đó

- Với độ tin cậy γ cho trước thì khoảng tin cậy cho p là:

$p \in (f - \varepsilon; f + \varepsilon)$

Trong đó:

+) f = m/n là tần suất xuất hiện biến cố A trong mẫu.

+) $\varepsilon_0 = \sqrt{\frac{f(1-f)}{n}}.u(\alpha/2)$ là sai số ước lượng

- Kích thước mẫu tối thiểu là:

$$n_{tt} = \left[\left(\frac{2.\sqrt{f(1-f)}}{I_0}.u(\alpha/2)\right)^2\right] + 1 \quad \text{hoặc} \quad n_{tt} = \left[\left(\frac{\sqrt{f(1-f)}}{\varepsilon_0}.u(\alpha/2)\right)^2\right] + 1$$

Bài tập chương 5

Câu 1: Theo dõi sự phát triển chiều cao của bạch đàn sau một năm tuổi ta có:

X (cm)	250-300	300-350	350-400	400-450	450-500	500-550	550-600
Số cây	5	20	25	30	28	23	14

a.Hãy ước lượng chiều cao trung bình của những cây với độ tin cậy 95%

b.Để ước lượng chiều cao trung bình cho bạch đàn sau 1 năm tuổi với độ chính xác là 0,2m thì đảm bảo độ tin cậy là bao nhiêu?

c. Muốn ước chiều cao trung bình cho bạch đàn sau 1 năm tuổi với độ chính xác là 30cm độ tin cậy 98% thì cần theo dõi thêm nữa không?

d. Hãy ước lượng chiều cao trung bình của những cây chậm lớn với độ tin cậy 98%

e. Hãy ước lượng tỷ lệ cây chậm lớn với độ tin cậy 95%

(Quy ước những cây cao không quá 3,5 m là cây chậm lớn).

Câu 2: Điều tra chỉ tiêu X của một loại sản phẩm (đơn vị tính bằng cm) thì thu được kết quả sau (X tuân theo phân phối chuẩn)

X_i	80-95	95-110	110-125	125-140	140-155	155-170	170-185
N_i	13	19	23	30	28	23	14

Quy ước những sản phẩm có chỉ tiêu từ 155 cm trở lên là sản phẩm ngoại cỡ.

a.Để ước lượng tỷ lệ sản phẩm ngoại cỡ với độ chính xác là 2% thì đảm bảo độ tin cậy là bao nhiêu?

b.Hãy ước lượng trung bình chỉ tiêu của những sản phẩm ngoại cỡ với độ tin cậy là 99%?

Câu 3: Kết quả quan sát về hàm lượng vitamin C của một loại trái cây cho trong bảng sau:

Hàm lượng (%)	6-7	7-8	8-9	9-10	10-11	11-12
số trái	5	10	20	35	25	5

Hàm lượng vitamin C của trái tuân theo phân phối chuẩn. Những trái loại A có hàm lượng vitamin C trên 10%.

a.Muốn độ chính xác khi ước lượng hàm lượng vitamin C trung bình của trái loại A là 0,1% với độ tin cậy 95% thì cần quan sát thêm bao nhiêu trái nữa.

b.Hãy ước lượng hàm lượng vitamin C trung bình trong 1 trái với độ tin cậy 98%.

c.Hãy ước lượng tỷ lệ trái loại A với độ tin cậy 98%.

d.Để ước lượng hàm lượng vitamin C trung bình cho mỗi trái với độ chính xác là 0,1% thì đảm bảo độ tin cậy là bao nhiêu?

Câu 4: Kiểm tra lượng xăng hao phí trên đoạn đường từ A đến B, người ta theo dõi một số ôtô chạy trên đoạn đường này và thu được kết quả:

X(Lit)	9,5-9,7	9,7-9,9	9,9-10,1	10,1-10,3	10,3-10,5
N_i	7	9	15	11	8

(X tuân theo quy luật chuẩn)

a. Những xe có lượng xăng hao phí từ 10,1 lít trở lên là xe tốn xăng. Hãy ước lượng tỷ lệ xe tốn xăng với độ tin cậy 95%.

b. Muốn ước lượng lượng xăng hao phí trung bình với độ chính xác là 0,1 lít và độ tin cậy 98% thì cần theo dõi thêm nữa không?

Câu 5:Khi kiểm tra thể lực của một nhóm sinh viên, ta có kết quả về cân nặng như sau:

X(kg)	42,5-47,5	47,5-52,5	52,5-57,5	57,5-62,5	62,5-67,5
Số sinh viên	8	14	28	18	12

Giả sử cân nặng của sinh viên tuân theo quy luật chuẩn

a.Để ước lượng tỷ lệ sinh viên có cân nặng lớn hơn hoặc bằng 52,5 kg với độ tin cậy 99% và độ chính xác không quá 2% thì tối thiểu cần điều tra cân nặng của bao nhiêu sinh viên?

b.Hãy ước lượng cho cân nặng trung bình của nhóm sinh viên trên với độ tin cậy 96%.

Câu 6:Nghiên cứu chiều cao X(cm) của học sinh tiểu học, ta có kết quả sau (X tuân theo phân phối chuẩn)

X(cm)	120-125	125-130	130-135	135-140	140-145
N_i	9	16	34	30	19

a. Để ước lượng chiều cao trung bình của các học sinh trên với độ chính xác 2 cm và độ tin cậy 94% thì cần điều tra thêm nữa không?

b. Hãy ước lượng tỷ lệ các học sinh có chiều cao lớn hơn hoặc bằng 135 cm với độ tin cậy 92%?

Câu 7: Để nghiên cứu sự phát triển của một loại giống cây trồng, người ta tiến hành đo đường kính X(cm) và thu được kết quả (X tuân theo phân phối chuẩn)

X(cm)	20-22	22-24	24-26	26-28	28-30
N_i	7	14	33	27	19

a. Để ước lượng đường kính trung bình lớn hơn hoặc bằng 26 cm với độ chính xác là 0,5 cm thì đảm bảo độ tin cây là bao nhiêu?

b.Để ước lượng đường kính trung bình của cây trên với độ chính xác 5 cm và độ tin cậy 95% thì cần điều tra thêm nữa không?

Câu 8:Theo dõi số kẹo bán được (kg) trong 1 tuần lễ tại một xí nghiệp ta có kết quả:

X(kg)	0-50	50-100	100-150	150-200	200-250	250-300	300-350
số tuần	9	23	27	30	25	20	5

Biết rằng X tuân theo quy luật chuẩn

a. Để ước lượng số kẹo trung bình bán được trong 1 tuần với độ chính xác 10 kg và độ tin cậy 97% thì cần điều tra thêm bao nhiêu tuần nữa?

b. Hãy ước lượng tỷ lệ những tuần bán được lớn hơn hoặc bằng 250 kg với độ tin cậy 88%?

Câu 9:Để nghiên cứu nhu cầu X của một loại hàng hoá trong 1 tháng, người ta tiến hành điều tra ở 100 gia đình và thu được kết quả sau:

X(kg)	1-2	2-3	3-4	4-5	5-6	6-7	7-8
n_i	10	13	22	25	18	15	7

Biết rằng X tuân theo quy luật chuẩn

a. Muốn ước lượng nhu cầu trung bình về mặt hàng này với độ chính xác không quá 0,5 kg và độ tin cậy 95% thì cần điều tra tối thiểu bao nhiêu gia đình?

b.Hãy ước lượng tỷ lệ gia đình có nhu cầu ít nhất 5 kg về mặt hàng này với độ tin cậy 90%?

Câu 10: Nghiên cứu năng suất X của một giống lúa A tại một vùng, người ta gặt ngẫu nhiên một số thửa ruộng của vùng đó và thu được kết quả:

X(tạ/ha)	40-42	42-44	44-46	46-48	48-50	50-52
n_i	7	13	25	35	15	5

Biết rằng X tuân theo quy luật chuẩn

a. Để ước lượng năng suất lúa trung bình của giống lúa trên với độ chính xác 0,4088 tạ/ha thì đảm bảo độ tin cậy là bao nhiêu?

b. Hãy ước lượng tỷ lệ thửa ruộng có năng suất lớn hơn hoặc bằng 48 tạ/ha với độ tin cậy 93%?

Câu 11:Khảo sát thu nhập X của một số người làm việc ở một công ty có kết quả:

X(ngàn đồng/tháng)	780-820	820-860	860-900	900-940	940-980
Số người	8	12	17	16	7

Biết rằng X tuân theo quy luật chuẩn

a. Nếu muốn độ tin cậy khi ước lượng thu nhập trung bình đạt 99% thì độ chính xác là bao nhiêu?

b. Hãy ước lượng thu nhập trung bình của những người có thu nhập tối thiểu 900 nghìn đồng/tháng với độ tin cậy 94%?

Câu 12: Điều tra lượng bia bán được X(lit) (X có phân phối chuẩn) trong 1 ngày tại các quán bia ở một quận có kết quả:

X(lít)	75-125	125-175	175-225	225-275	275-325	325-375	375-425
n_i	4	9	10	18	16	15	7

Các quán bán được 325 lít trở lên được gọi là đông khách

a. Muốn độ chính xác khi ước lượng tỷ lệ của các quán đông khách là 0,04 với độ tin cậy 89% thì cần phải điều tra thêm bao nhiêu quán nữa.

b. Hãy ước lượng bia bán được trung bình trong 1 ngày của các quán đông khách với độ tin cậy là 95%?

Câu 13: Tiến hành quan sát về độ bền X(kg/mm^2) của một loại sắt, ta có kết quả (X tuân theo phân phối chuẩn)

X	80-100	100-120	120-140	140-160	160-180	180-200
n_i	16	20	24	30	17	15

a. Để ước lượng độ bền trung bình của loại sắt trên với độ chính xác 38 kg/mm^2 thì đảm bảo độ tin cậy là bao nhiêu?

b.Hãy ước lượng tỷ lệ sắt có độ bền lớn hơn hoặc bằng 160 kg/mm^2 với độ tin cậy 90%

Câu 14: Đo đường kính X của một số chi tiết do 1 máy sản xuất ta có kết quả (X tuân theo phân phối chuẩn)

X(mm)	19,8-19,85	19,85-19,9	19,9-19,95	19,95-20	20-20,05	20,05-20,1
Số chi tiết	12	16	20	14	7	5

Quy ước những chi tiết có đường kính từ 19,9 mm đến 20,05 mm là những chi tiết đạt tiêu chuẩn.

a. Để ước lượng đường kính trung bình của các chi tiết đạt độ chính xác là 0,02 mm và ước lượng tỷ lệ chi tiết đạt tiêu chuẩn đạt độ chính xác là 5% với cùng độ tin cậy 89% thì cần đo thêm bao nhiêu chi tiết nữa?

b. Hãy ước lượng đường kính trung bình của các chi tiết đạt tiêu chuẩn với độ tin cậy 95%.

Câu 15: Đo độ nhạy X của một kênh truyền hình của một số máy thu hình, ta có kết quả (X tuân theo quy luật chuẩn)

khoảng (μv)	225-275	275-325	325-375	375-425	425-475	475-525
Số máy	9	18	28	16	15	7

a.Với độ tin cậy 98%, hãy ước lượng độ nhạy trung bình cho các máy trên.

b.Muốn độ chính xác của ước lượng tỷ lệ các máy có độ nhạy ≤ 325□ v là 2% với độ tin cậy 97% thì cần điều tra thêm nữa không?

Câu 16: Theo dõi chiều dài X của một loại sản phẩm ta có kết quả

X(m)	0,5-1,5	1,5-2,5	2,5-3,5	3,5-4,5	4,5-5,5	5,5-6,5
Số sp (n_i)	14	22	28	24	13	8

Biết rằng X tuân theo quy luật chuẩn

a.Hãy ước lượng chiều dài trung bình của sản phẩm với độ tin cậy là 85%.

b. Để ước lượng tỷ lệ các sản phẩm có chiều dài tối thiểu 4,5 m với độ chính xác là 5% thì đảm bảo độ tin cậy là bao nhiêu?

Câu 17:Điều tra lượng Glucoza X trong máu ở một số người thì thu được kết quả:

X(mg%)	65-70	70-75	75-80	80-85	85-90	90-95	95-100
Số người	3	5	8	16	18	17	9

Biết rằng X tuân theo quy luật chuẩn

a. Để ước lượng Glucoza trung bình của những người có lương Glucoza trong máu ≥ 90 mg% với độ chính xác là 0,3 mg% và độ tin cậy 94% thì cần điều tra thêm nữa không?

b. Hãy ước lượng Glucoza trung bình trong máu của số người nói trên với độ tin cậy 88%.

Câu 18: Điều tra doanh thu hàng năm (đơn vị: 1.000.000đồng) của các đại lý cho một công ty kinh doanh tại vùng A thu được số liệu:

Mức doanh thu	25-30	30-35	35-40	40-45	45-50	50-55
Số đại lý	20	30	100	80	15	10

Biết rằng doanh thu hàng năm của các đại lý ở vùng A có phân phối chuẩn

a.Hãy ước lượng tỷ lệ đại lý có doanh thu tối thiểu 40 triệu ở vùng A với độ tin cậy 95%.

b. Nếu muốn độ chính xác khi ước lượng doanh thu trung bình của các đại lý trên là 2 triệu và độ chính xác khi ước lượng tỷ lệ đại lý có doanh thu tối thiểu 40 triệu là 0,05 với độ tin cậy 90% thì cần điều tra bao nhiêu đại lý nữa.

Câu 19: Trọng lượng sản phẩm của một nhà máy là một đại lượng ngẫu nhiên tuân theo quy luật chuẩn. Cân ngẫu nhiên 1 số sản phẩm thì thu được kết quả:

Trọng lượng (kg)	39,75-39,85	39,85-39,95	39,95-40,05	40,05-40,15	40,15-40,25	40,25-40,35
Số sản phẩm	7	12	18	15	6	2

Quy ước những sản phẩm có trọng lượng lớn hơn hoặc bằng 40,05 kg là sản phẩm nặng.

a. Để ước lượng trọng lượng trung bình của sản phẩm nặng đạt được độ chính xác là 0,03 kg thì đảm bảo độ tin cậy là bao nhiêu?

b.Hãy ước lượng trọng lượng trung bình của các sản phẩm trên với độ tin cậy là 88%

Câu 20: Mức tăng trọng của chuột thí nghiệm là đại lượng ngẫu nhiên tuân theo quy luật chuẩn. Kết quả thí nghiệm trên 1 số con chuột cho ở bảng sau:

trọng lượng (g)	10-15	15-20	20-25	25-30	30-35	35-40
số con	2	5	7	10	6	2

a.Hãy ước lượng mức tăng trọng trung bình của chuột thí nghiệm với độ tin cậy là 99%?

b.Nếu muốn độ chính xác khi ước lượng tỷ lệ chuột có trọng lượng ≥ 30g là 3% thì đảm bảo độ tin cậy là bao nhiêu?

Chương 6: KIỂM ĐỊNH GIẢ THUYẾT

6.1.Khái niệm mở đầu:

Ví dụ: Một nhà máy sản xuất bóng đèn điện tử có đặt chỉ tiêu kỹ thuật là tuổi thọ của đèn 10 ngàn giờ. Bài toán đặt ra hãy đánh giá các bóng đèn sản xuất ra của nhà máy có đảm bảo chỉ tiêu kỹ thuật không?

Để giải quyết bài toán này thì ta không thể kiểm tra từng bóng đèn được và dù việc kiểm tra được tiến hành cũng không có kết luận đúng. Vậy làm thế nào?

Trong thực tế cuộc sống và xã hội có rất nhiều bài toán tương tự như vậy được đặt ra, để giải quyết các bài toán dạng này ta phải sử dụng các phương pháp được xây dựng trên cơ sở toán học.

Một trong các phương pháp quan trọng đó là lý thuyết kiểm định giả thuyết thống kê nghĩa là nêu lên một giả thuyết cần kiểm tra và dùng một chỉ tiêu chuẩn đã được xây dựng với một số điều kiện cho trước để đánh giá giả thuyết này đúng hay sai. Vấn đề đặt ra là tiêu chuẩn đó được xây dựng như thế nào? Vận dụng ra sao?

Giả sử cần nghiên cứu một dấu hiệu nào đó của tổng thể và dấu hiệu đó được mô hình hóa bằng đlnn X có phân phối xác suất xác định song còn một số tham số nào đó chưa biết. Khi đó ta nêu:

- Giả thuyết: H_0 (Tham số chưa biết nhận giá trị xác định).

- Đối thuyết: H_1 (Tham số chưa biết không nhận giá trị xác định đó).

Cần xác định xem H_0 đúng hay H_1 đúng dựa trên các quy tắc kiểm định sau:

- Từ tổng thể lấy một mẫu kích thước n: $(X_1, X_2, \ldots, X_n)$, gọi $M = (X_1, X_2, \ldots, X_n)$ là điểm mẫu, ký hiệu R^n là không gian mẫu n chiều.(Tập hợp tất cả những điểm mẫu kích thước n).

- Giả sử giả thuyết H_0 đúng, vận dụng các kết quả của lý thuyết xác suất ta sẽ tìm một miền $S \subset R^n$ sao cho khi mẫu $(X_1, X_2, \ldots, X_n) \in S$ thì giả thuyết H_0 sai (bác bỏ giả thuyết H_0) còn mẫu $(X_1, X_2, \ldots, X_n) \notin S$ thì ta tạm thời chấp nhận giả thuyết H_0 cho đến khi có thông tin mới.(bác bỏ H_1).

- Miền S nói trên được gọi là miền tiêu chuẩn hay miền bác bỏ giả thuyết H_0. làm theo phương pháp này ta sẽ phạm phải một trong hai sai lầm sau:

+ Sai lầm loại 1 là bác bỏ H_0 nhưng thực tế H_0 đúng.

+ Sai lầm loại 2 là chấp nhận H_0 nhưng thực tế H_0 sai.

Bài toán kiểm định được gọi là tối ưu nếu ta chọn được một giả thuyết nào đó mà cả hai khả năng mắc sai lầm đều đạt cực tiểu.

Tuy nhiên trong xác suất thống kê thì bài này không có lời giải vì khi ta cực tiểu được sai lầm loại này thì sai lầm loại kia sẽ tăng lên và ngược lại.

Do đó để giải quyết bài toán này người ta cho trước một giới hạn trên, ký hiệu: α - là xác suất phạm sai lầm loại 1 và bài toán dẫn đến tìm miền tiêu chuẩn S sao cho xác suất phạm sai lầm loại 1 không vượt quá α. Còn xác suất phạm sai lầm loại 2 sẽ đạt min. Số α nói trên gọi là mức ý nghĩa của kiểm định.

6.2. Môt số bài toán kiểm định giả thuyết

6.2.1. Bài toán KĐGT về GTTB của đlnn $X \sim N(\mu; \sigma^2)$

6.2.1.1. Đã biết σ

i) Với bài toán kiểm định:

- Giả thuyết H_0 : $\mu = \mu_0$
- Đối thuyết H_1 : $\mu > \mu_0$

Với mức ý nghĩa α cho trước.

Chọn thống kê: $u = \dfrac{\overline{x} - \mu_0}{\sigma}\sqrt{n}$

thì miền bác bỏ giả thuyết H_0 là: $S=\{u / u > u(\alpha)\}$

Trong đó $u(\alpha)$ là giá trị tới hạn chuẩn được xác định: $\phi(u(\alpha)) = 0,5 - \alpha$

Ví dụ: Kích thước một loại sản phẩm là đlnn có phân phối chuẩn với độ lệch chuẩn là 9 mm và độ dài trung bình 21 mm. Nghi ngờ kích thước của sản phẩm có xu hướng tăng lên, lấy ngẫu nhiên một mẫu gồm 64 sản phẩm để kiểm tra thì ta được độ dài trung bình là 23,7 mm.

Hãy xét xem điều nghi ngờ trên đúng không? Với mức ý nghĩa là 5%.

Giải: Gọi X là kích thước sản phẩm. $X \sim N(\mu; \sigma^2)$

Bài toán dẫn tới kiểm định

- Giả thuyết H_0 : $\mu = 21$
- Đối thuyết H_1 : $\mu > 21$

Với mức ý nghĩa $\alpha = 5\%$

Chọn thống kê: $u = \dfrac{\overline{x} - \mu_0}{\sigma}\sqrt{n}$

Khi đó miền bác bỏ giả thuyết H_0 là: $S=\{u / u > u(\alpha)\}$

Ta có: $u = \frac{\overline{x} - \mu_0}{\sigma}\sqrt{n} = \frac{23,7 - 21}{9}\sqrt{64} = 2,4$

$u(\alpha) = u(0,05) = 1,64$

=> $u > u(\alpha)$ thỏa mãn miền S

Vậy bác bỏ H_0 , chấp nhận H_1 => Điều nghi ngờ là đúng.

ii) Với bài toán kiểm định:

- Giả thuyết H_0 : $\mu = \mu_0$
- Đối thuyết H_1 : $\mu < \mu_0$

Với mức ý nghĩa α cho trước.

Chọn thống kê: $u = \frac{\overline{x} - \mu_0}{\sigma}\sqrt{n}$

Khi đó miền bác bỏ giả thuyết H_0 là: $S = \{u / u < - u(\alpha)\}$

Trong đó $u(\alpha)$ là giá trị tới hạn chuẩn được xác định

$\phi(u(\alpha)) = 0,5 - \alpha$

Ví dụ: Kích thước một loại sản phẩm là đlnn có phân phối chuẩn với độ lệch chuẩn là 9 mm và độ dài trung bình 21 mm. Nghi ngờ kích thước của sản phẩm có xu hướng giảm đi, lấy ngẫu nhiên một mẫu gồm 64 sản phẩm để kiểm tra thì ta được độ dài trung bình là 20,7 mm. Hãy xét xem điều nghi ngờ trên đúng không? Với mức ý nghĩa là 5%.

*Giải:*Gọi X là kích thước sản phẩm. $X \sim N(\mu; \sigma^2)$

Bài toán dẫn tới kiểm định

- Giả thuyết H_0 : $\mu = 21$
- Đối thuyết H_1 : $\mu < 21$

Với mức ý nghĩa $\alpha = 5\%$

Chọn thống kê: $u = \frac{\overline{x} - \mu_0}{\sigma}\sqrt{n}$

Khi đó miền bác bỏ giả thuyết H_0 là: $S=\{u / u < - u(\alpha)\}$

Ta có: $u = \frac{\overline{x} - \mu_0}{\sigma}\sqrt{n} = \frac{20,7 - 21}{9}\sqrt{64} = -0,27$

$u(\alpha) = u(0,05) = 1,64$

=> $u > - u(\alpha)$ không thỏa mãn miền S

Vậy bác bỏ H_1 , chấp nhận H_0 => Điều nghi ngờ là sai.

iii) Với bài toán kiểm định:

- Giả thuyết H_0 : $\mu = \mu_0$
- Đối thuyết H_1 : $\mu \neq \mu_0$

Với mức ý nghĩa α cho trước.

Chọn thống kê: $u = \frac{\overline{x} - \mu_0}{\sigma}\sqrt{n}$

Khi đó miền bác bỏ giả thuyết H_0 là: $S = \{u \,/\, |u| > u(\alpha/2)\}$

Trong đó $u(\alpha/2)$ là giá trị tới hạn chuẩn được xác định $\phi(u(\alpha)) = 0,5 - \alpha/2$

Ví dụ: Kích thước một loại sản phẩm là đlnn có phân phối chuẩn với độ lệch chuẩn là 9 mm và độ dài trung bình 21 mm. Nghi ngờ kích thước của sản phẩm đã bị thay đổi, lấy ngẫu nhiên một mẫu gồm 64 sản phẩm để kiểm tra thì ta được độ dài trung bình là 20 mm. Hãy xét xem điều nghi ngờ trên đúng không? Với mức ý nghĩa là 5%.

Giải: Gọi X là kích thước sản phẩm. $X \sim N(\mu; \sigma^2)$

Bài toán dẫn tới kiểm định

- Giả thuyết H_0 : $\mu = 21$
- Đối thuyết H_1 : $\mu \neq 21$

Với mức ý nghĩa $\alpha = 5\%$

Chọn thống kê: $u = \frac{\overline{x} - \mu_0}{\sigma}\sqrt{n}$

Khi đó miền bác bỏ giả thuyết H_0 là: $S = \{u \,/\, |u| > u(\alpha/2)\}$

Ta có: $u = \frac{\overline{x} - \mu_0}{\sigma}\sqrt{n} = \frac{20 - 21}{9}\sqrt{64} = -0,89$

$u(\alpha) = u(0,025) = 1,96$

$\Rightarrow |u| < u(\alpha/2)$ không thỏa mãn miền S

Vậy bác bỏ H_1 , chấp nhận H_0 => Điều nghi ngờ là sai.

6.2.1.2.chưa biết σ và $n \geq 30$

Với bài toán kiểm định

- Giả thuyết H_0 : $\mu = \mu_0$
- Đối thuyết H_1 : $\mu > \mu_0$
  $\mu < \mu_0$
  $\mu \neq \mu_0$

Với mức ý nghĩa α cho trước.

Chọn thống kê: $u = \dfrac{\overline{x} - \mu_0}{s'}\sqrt{n}$

Khi đó miền bác bỏ giả thuyết H_0 ứng với các đối thuyết là:

$S = \{ u \ / \ \ u > u(\alpha) \}$

$S = \{ u \ / \ \ u < - u(\alpha) \}$

$S = \{ u \ / \ \ |u| > u(\alpha/2) \}$

Ví dụ1: Kích thước một loại sản phẩm là đlnn có độ dài trung bình 21 mm. Nghi ngờ kích thước của sản phẩm không đúng , lấy ngẫu nhiên một mẫu gồm 64 sản phẩm để kiểm tra thì ta được độ dài trung bình là 20 mm và độ lệch tiêu chuẩn là 9 mm. Hãy xét xem điều nghi ngờ trên đúng không? Với mức ý nghĩa là 5%.

Giải: Gọi μ là kích thước trung bình thực tế của sản phẩm

Bài toán dẫn tới kiểm định

- Giả thuyết H_0 : $\mu = 21$
- Đối thuyết H_1 : $\mu \neq 21$

Với mức ý nghĩa $\alpha = 5\%$

Chọn thống kê: $u = \dfrac{\overline{x} - \mu_0}{s'}\sqrt{n}$

Khi đó miền bác bỏ giả thuyết H_0 là: $S = \{u \ / \ |u| > u(\alpha/2)\}$

Ta có: $s' = \sqrt{\dfrac{n}{n-1}s^2} = \sqrt{\dfrac{64}{63}9^2} = 9{,}07$

$$=> u = \frac{\overline{x} - \mu_0}{s'}\sqrt{n} = \frac{20 - 21}{9{,}07}\sqrt{64} = -0{,}88$$

$u(\alpha) = u(0{,}025) = 1{,}96$

=> $|u| < u(\alpha/2)$ không thỏa mãn miền S

Vậy bác bỏ H_1 , chấp nhận H_0 => Điều nghi ngờ là sai.

Ví dụ 2: Một công ty có 1 hệ thống máy tính có thể xử lý 1200 hoá đơn trong 1 giờ. Công ty mới nhập một hệ thống máy tính mới. Hệ thống này khi chạy kiểm tra trong 40 giờ cho thấy số hoá đơn được xử lý trung bình trong một giờ là 1260 với độ lệch tiêu chuẩn là 215. Với mức ý nghĩa 5%, hãy nhận định xem hệ thống mới có tốt hơn hệ thống cũ không?

Giải: Gọi μ Số hoá đơn được xử lý TB trong 1 giờ của hệ thống mới.

Bài toán dẫn tới kiểm định

- Giả thuyết H_0 : $\mu = 1200$

- Đối thuyết H_1 : $\mu > 1200$

Với mức ý nghĩa $\alpha = 5\%$

Chọn thống kê: $u = \dfrac{\overline{x} - \mu_0}{s'}\sqrt{n}$

Khi đó miền bác bỏ giả thuyết H_0 là: $S = \{ u \ / \ u > u(\alpha) \}$

Ta có: $s' = \sqrt{\dfrac{n}{n-1}s^2} = \sqrt{\dfrac{40}{39}215^2} = 217,7$

$\Rightarrow u = \dfrac{\overline{x} - \mu_0}{s'}\sqrt{n} = \dfrac{1260 - 1200}{217,7}\sqrt{40} = 1,74$

$u(\alpha) = u(0,05) = 1,64$

=> $u > u(\alpha)$ thỏa mãn miền S

Vậy bác bỏ H_0 , chấp nhận H_1 => hệ thống mới có tốt hơn hệ thống cũ

6.2.1.3. Chưa biết σ và n < 30

Với bài toán kiểm định

- Giả thuyết H_0 : $\mu = \mu_0$

- Đối thuyết H_1 : $\mu > \mu_0$

  $\mu < \mu_0$

  $\mu \neq \mu_0$

Với mức ý nghĩa α cho trước.

Chọn thống kê: $u = \dfrac{\overline{x} - \mu_0}{s'}\sqrt{n}$

Khi đó miền bác bỏ giả thuyết H_0 ứng với các đối thuyết là:

$S = \{ u \ / \ u > t^{(n-1)}(\alpha) \}$

$S = \{ u \ / \ u < - t^{(n-1)}(\alpha) \}$

$S = \{ u \ / \ |u| > t^{(n-1)}(\alpha/2) \}$

Ví dụ: Thời gian sống trung bình sau mổ của một bệnh nhân mắc bệnh ung thư A là 5 năm. Một loại thuốc mới tìm thấy, khi dùng thử trên 20 bệnh nhân loại này sau khi mổ thì thấy thời gian sống trung bình của họ là 5,7 năm với độ lệch tiêu chuẩn là 1,2 năm. Với mức ý nghĩa là 1% hãy cho biết loại thuốc mới này có kéo dài được thời gian sống sau mổ của bệnh nhân hay không?

*Giải:*Gọi μ là thời gian sống trung bình sau mổ của bệnh nhân khi được dùng thuốc.

Bài toán dẫn tới kiểm định:

- Giả thuyết H_0: $\mu = 5$
- Đối thuyết H_1: $\mu > 5$

Với mức ý nghĩa 1%

Chọn thống kê: $u = \dfrac{\overline{x} - \mu_0}{s'}\sqrt{n}$

Miền bác bỏ H_0 : $S = \{ u \ : \ u > t^{(n-1)}(\alpha) \}$

Có: $s' = \sqrt{\dfrac{n}{n-1}s^2} = \sqrt{\dfrac{20}{19}1{,}2^2} = 1{,}23$

$$\Rightarrow \ u = \frac{\overline{x} - \mu_0}{s'}\sqrt{n} = \frac{5{,}7 - 5}{1{,}23}\sqrt{20} = 2{,}55$$

=> $t^{(n-1)}(\alpha) = t^{(19)}(0{,}01) = 2{,}539$ => $u > t^{(n-1)}(\alpha)$

=> Chấp nhận H_1, bác bỏ H_0 => Loại thuốc mới này có kéo dài được thời gian sống sau mổ của bệnh nhân.

6.2.2. KĐGT về sự bằng nhau của 2 GTTB.

Giả sử có: $X \sim N(\mu_1 ; \sigma_1^2)$ và $Y \sim N(\mu_2 ; \sigma_2^2)$

6.2.2.1. Đã biết σ_1 và σ_2

Với bài toán kiểm định

- Giả thuyết H_0 : $\mu_1 = \mu_2$
- Đối thuyết H_1 : $\mu_1 > \mu_2$

  $\mu_1 < \mu_2$

  $\mu_1 \neq \mu_2$

Với mức ý nghĩa α cho trước.

Chọn thống kê: $u = \dfrac{\overline{x_1} - \overline{x_1}}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$

Khi đó miền bác bỏ giả thuyết H_0 ứng với các đối thuyết là:

$S = \{ u \ : \ u > u(\alpha) \}$

$S = \{ u \ : \ u < -u(\alpha) \}$

$S = \{ u \ : |u| > u(\alpha/2) \}$

Ví dụ: Theo dõi tuổi thọ của 36 chiếc bóng đèn nhãn hiệu T, thấy tuổi thọ trung bình của chúng là 1250 giờ. Theo dõi 36 chiếc bóng đèn nhãn hiệu E thấy tuổi thọ trung bình là 1260 giờ. Với mức ý nghĩa 5%, có thể xem tuổi thọ trung bình của 2 loại bóng đèn là như nhau không? Biết rằng tuổi thọ của 2 loại bóng đèn này tuân theo quy luật chuẩn với độ lệch tiêu chuẩn của bóng nhãn hiệu T là 20 giờ, của bóng nhãn hiệu E là 35 giờ.

Giải: Gọi μ_1 là tuổi thọ trung bình của bóng đèn T

μ_2 là tuổi thọ trung bình của bóng đèn E

Bài toán dẫn tới kiểm định:

- Giả thuyết H_0: $\mu_1 = \mu_2$
- Đối thuyết H_1: $\mu_1 \neq \mu_2$

Với mức ý nghĩa 5%

Chọn thống kê: $u = \dfrac{\overline{x}_1 - \overline{x}_1}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$

Miền bác bỏ H_0 : $S = \{ u \ : |u| > u(\alpha/2) \}$

Có: $u = \dfrac{\overline{x}_1 - \overline{x}_1}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} = \dfrac{1250 - 1260}{\sqrt{\dfrac{20^2}{36} + \dfrac{35^2}{36}}} = -1,49$

$u(\alpha/2) = u(0,025) = 1,96 \quad \Rightarrow |u| < u(\alpha/2)$

=> Chấp nhận H_0, bác bỏ H_1

=> Tuổi thọ trung bình của 2 loại bóng đèn là khác nhau.

6.2.2.1.Chưa biết σ_1 và σ_2

Với bài toán kiểm định

- Giả thuyết H_0 : $\mu_1 = \mu_2$
- Đối thuyết H_1 : $\mu_1 > \mu_2$

$\mu_1 < \mu_2$

$\mu_1 \neq \mu_2$

Với mức ý nghĩa α cho trước.

Chọn thống kê: $u = \dfrac{\overline{x}_1 - \overline{x}_1}{\sqrt{\dfrac{s_1'^2}{n_1} + \dfrac{s_2'^2}{n_2}}}$

Khi đó miền bác bỏ giả thuyết H_0 ứng với các đối thuyết là:

$S = \{ u : u > u(\alpha) \}$

$S = \{ u : u < - u(\alpha) \}$

$S = \{ u : |u| > u(\alpha/2) \}$

Ví dụ: Theo dõi tuổi thọ của 36 chiếc bóng đèn nhãn hiệu T, thấy tuổi thọ trung bình của chúng là 1250 giờ với độ lệch tiêu chuẩn là 20 giờ. Theo dõi 36 chiếc bóng đèn nhãn hiệu E thấy tuổi thọ trung bình là 1260 giờ với độ lệch tiêu chuẩn là 35 giờ. Với mức ý nghĩa 5%, có thể xem tuổi thọ trung bình của 2 loại bóng đèn là như nhau không? Biết rằng tuổi thọ của 2 loại bóng đèn này tuân theo quy luật chuẩn.

Giải: Gọi μ_1 là tuổi thọ trung bình của bóng đèn T

μ_2 là tuổi thọ trung bình của bóng đèn E

Bài toán dẫn tới kiểm định:

- Giả thuyết H_0: $\mu_1 = \mu_2$
- Đối thuyết H_1: $\mu_1 \neq \mu_2$

Với mức ý nghĩa 5%

Chọn thống kê: $u = \dfrac{\overline{x}_1 - \overline{x}_1}{\sqrt{\dfrac{s_1'^2}{n_1} + \dfrac{s_2'^2}{n_2}}}$

Miền bác bỏ H_0 : $S = \{ u : |u| > u(\alpha/2) \}$

Có: $s_1'^2 = \sqrt{\dfrac{n_1}{n_1-1} s_1^2} = \sqrt{\dfrac{36}{35} 20^2} = 20,28$

$s_2'^2 = \sqrt{\dfrac{n_2}{n_2-1} s_2^2} = \sqrt{\dfrac{36}{35} 35^2} = 35,5$

$$u = \frac{\overline{x}_1 - \overline{x}_1}{\sqrt{\dfrac{s'^2_1}{n_1} + \dfrac{s'^2_2}{n_2}}} = \frac{1250 - 1260}{\sqrt{\dfrac{20,28^2}{36} + \dfrac{35,5^2}{36}}} = -1,468$$

$u(\alpha/2) = u(0,025) = 1,96 \quad \Rightarrow |u| < u(\alpha/2)$

=> Chấp nhận H_0, bác bỏ H_1

=> Tuổi thọ trung bình của 2 loại bóng đèn là khác nhau.

6.2.3. Bài toán KĐGT về tỷ lệ (xác suất)

Gọi p là xác suất xuất hiện biến cố A trong tổng thể.

Khi đó với bài toán kiểm định:

- Giả thuyết H_0 : $p = p_0$

- Đối thuyết H_1 : $p > p_0$

$p < p_0$

$p \neq p_0$

Với mức ý nghĩa α cho trước.

Chọn thống kê: $u = \dfrac{f - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n}$

Khi đó miền bác bỏ giả thuyết H_0 ứng với các đối thuyết là:

$S = \{ u : u > u(\alpha) \}$

$S = \{ u : u < - u(\alpha) \}$

$S = \{ u : |u| > u(\alpha/2) \}$

và: $f = m/n$ là tần suất xuất hiện biến cố A trong mẫu.

Ví dụ: Cơ quan cảnh sát giao thông cho rằng 62% số người lái xe trên đường là có bằng lái. Kiểm tra ngẫu nhiên 130 người lái xe thấy chỉ có 68 người có bằng lái xe. Với mức ý nghĩa 5%, hãy kết luận về ý kiến trên.

Giải: Gọi p là tỷ lệ người lái xe trên đường là có bằng lái.

Bài toán dẫn tới kiểm định:

- Giả thuyết H_0: $p = 0,62$
- Đối thuyết H_1: $p \neq 0,62$

Với mức ý nghĩa 5%.

Chọn thống kê: $u = \dfrac{f - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n}$

Khi đó miền bác bỏ giả thuyết H_0 là:

$S = \{ u : |u| > u(\alpha/2) \}$

Có: $f = \dfrac{m}{n} = \dfrac{68}{130}$

$$u = \frac{f - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n} = \frac{0,523 - 0.62}{\sqrt{0,62(1-0,62)}}\sqrt{130} = -2,28$$

$u(\alpha/2) = u(0,025) = 1,96 \Rightarrow |u| > u(\alpha/2)$

=> Bác bỏ H_0, chấp nhận H_1

=> Cơ quan cảnh sát giao thông cho rằng 62% số người lái xe trên đường là có bằng lái là không đúng.

6.2.4. Bài toán KĐGT về sự bằng nhau của hai tỷ lệ (xác suất)

- Giả thuyết H_0 : $p_1 = p_2$
- Đối thuyết H_1 : $p_1 > p_2$

  $p_1 < p_2$

  $p_1 \neq p_2$

Với mức ý nghĩa α cho trước.

Chọn thống kê: $$u = \frac{\dfrac{m_1}{n_1} - \dfrac{m_2}{n_2}}{\sqrt{\dfrac{m_1+m_2}{n_1+n_2}\left(1-\dfrac{m_1+m_2}{n_1+n_2}\right)\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}}$$

Khi đó miền bác bỏ giả thuyết H_0 ứng với các đối thuyết là:

$S = \{ u : u > u(\alpha) \}$

$S = \{ u : u < -u(\alpha) \}$

$S = \{ u : |u| > u(\alpha/2) \}$

Ví dụ: Công ty nước giải khát Coca - Cola đang nghiên cứu việc đưa vào một công thức mới để cải tiến sản phẩm của mình. Với công thức cũ, khi cho 500 người dùng thử thì có 120 người ưa thích nó. Với công thức mới, khi đưa cho 1000 người khác dùng thử thì có 300 người tỏ ra ưa thích nó. Với mức ý nghĩa là 2% hãy cho biết có phải công thức mới đưa vào đã làm tăng tỷ lệ người ưa thích Coca - Cola không?

Giải: Gọi P_1 làTỷ lệ người thích Coca - Cola với công thức cũ

P_2 là Tỷ lệ người thích Coca - Cola với công thức mới

Bài toán dẫn tới kiểm định:

- Giả thuyết H_0: $p_1 = p_2$
- Đối thuyết H_1: $p_1 < p_2$

Với mức ý nghĩa 2%

Chọn thống kê: $$u = \frac{\dfrac{m_1}{n_1} - \dfrac{m_2}{n_2}}{\sqrt{\dfrac{m_1+m_2}{n_1+n_2}\left(1-\dfrac{m_1+m_2}{n_1+n_2}\right)\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}}$$

Miền bác bỏ H_0 : $S = \{ u : u < -u(\alpha) \}$

Có: $u = \dfrac{\dfrac{m_1}{n_1} - \dfrac{m_2}{n_2}}{\sqrt{\dfrac{m_1+m_2}{n_1+n_2}\left(1-\dfrac{m_1+m_2}{n_1+n_2}\right)\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}}$

$$= \frac{\dfrac{120}{500} - \dfrac{300}{1000}}{\sqrt{\dfrac{120+300}{500+1000}\left(1-\dfrac{120+300}{500+1000}\right)\left(\dfrac{1}{500}+\dfrac{1}{1000}\right)}} = -2,5$$

$u(\alpha) = u(0,02) = 2,05$ => $u < -u(\alpha)$

=> Bác bỏ H_0, chấp nhận H_1

Công thức mới đưa vào đã làm tăng tỷ lệ người ưa thích Coca - Cola

Bài tập chương 6

Câu 1: Thống kê 10650 trẻ sơ sinh ở một địa phương, người ta thấy có 5410 con trai. Hỏi tỷ lệ sinh con trai có thực sự cao hơn tỷ lệ sinh con gái không? Cho biết mức ý nghĩa là 1%.

Câu 2: Cơ quan cảnh sát giao thông cho rằng 62% số người lái xe trên đường là có bằng lái. Kiểm tra ngẫu nhiên 130 người lái xe thấy chỉ có 68 người có bằng lái xe. Với mức ý nghĩa 5%, hãy kết luận về ý kiến trên.

Câu 3: Thời gian sống trung bình sau mổ của một bệnh nhân mắc bệnh ung thư A là 5 năm. Một loại thuốc mới tìm thấy, khi dùng thử trên 20 bệnh nhân loại này thì thấy thời gian sống trung bình của họ là 5,7 năm với độ lệch tiêu chuẩn là 1,2 năm. Với mức ý nghĩa là 1% hãy cho biết loại thuốc mới này có kéo dài được thời gian sống sau mổ của bệnh nhân hay không?

Câu 4: Bệnh A có thể chữa bằng thuốc H. Công ty sản xuất thuốc H tuyên bố tỷ lệ bệnh nhân khỏi bệnh do dùng thuốc của họ là 85%. Người ta dùng thử thuốc H cho 250 bệnh nhân bị bệnh A thì thấy có 210 người khỏi bệnh. Với mức ý nghĩa 5% hãy cho biết hiệu quả chữa bệnh của thuốc H có đúng như quảng cáo của công ty hay không?

Câu 5: Có 36 vòng bi loại cũ và 25 vòng bi loại mới. Biết rằng ma sát của các loại vòng bi này là đại lượng ngẫu nhiên tuân theo quy luật chuẩn. Có ý kiến cho rằng loại vòng bi mới có ma sát nhỏ hơn so với loại cũ. Để kiểm tra ý kiến này, người ta cho các vòng bi nói trên hoạt động theo cùng một chế độ và đo được ma sát trung bình của loại cũ là 52 với độ lệch tiêu chuẩn là 8; còn ma sát trung bình của loại mới là 44 với độ lệch tiêu chuẩn là 12. Với mức ý nghĩa là 10%, hãy kết luận về ý kiến trên.

Câu 6: Ở một khu dân cư, các hộ gia đình chỉ có thể mua gas ở một trong 2 cửa hàng A và B. Điều tra ngẫu nhiên 500 hộ dùng gas, thấy có 265 hộ dùng gas của cửa hàng A. Số còn lại dùng gas của cửa hàng B. Với mức ý nghĩa 5% có thể kết luận cửa hàng A thu hút khách trên địa bàn hơn cửa hàng B được không?

Câu 7:Cân ngẫu nhiên một số sản phẩm cùng loại của nhà máy, được kết quả là:

Trọng lượng (kg)	39,8	39,9	40	40,2	40,3	40,6
Số sản phẩm	12	13	17	26	15	12

Biết rằng trọng lượng sản phẩm trên tuân theo quy luật chuẩn.

Với mức ý nghĩa là 5%, có thể chấp nhận ý kiến cho rằng trọng lượng trung bình của loại sản phẩm đó là 40 kg được không?

Câu 8: Phỏng vấn ngẫu nhiên một số người ở thành phố A về mức tiêu dùng gạo hàng tháng của họ thu được số liệu:

mức tiêu dùng (kg)	10	11	12	13	14	15	16	17
Số người	5	10	25	30	15	10	3	2

Ở thành phố B, phỏng vấn 200 người, tìm được mức tiêu dùng gạo trung bình hàng tháng của họ là 15kg và độ lệch tiêu chuẩn là 0,5kg. Với mức ý nghĩa là 5% có thể cho rằng mức tiều dùng gạo trung bình hàng tháng của dân 2 thành phố đó khác nhau hay không? Giả thiết rằng mức tiêu dùng gạo tuân theo quy luật chuẩn.

Câu 9: Trọng lượng đóng bao gạo theo quy định là 20 kg. Với mức ý nghĩa là 5% có thể cho rằng bao gạo bị đóng thiếu hay không nếu kiểm tra ngẫu nhiên một số bao thu được kết quả:

Trọng lượng (kg)	19	19,3	19,4	19,6	19,8	20	20,2
Số bao	2	3	5	8	4	2	1

Giả thiết trọng lượng đóng bao gạo có phân phối chuẩn.

Câu 10: Theo dõi tuổi thọ của 36 chiếc bóng đèn nhãn hiệu T, thấy tuổi thọ trung bình của chúng là 1250 giờ với độ lệch tiêu chuẩn hiệu chỉnh là 20. Theo dõi 36 chiếc bóng đèn nhãn hiệu E thấy tuổi thọ trung bình là 1260 giờ với độ lệch tiêu chuẩn hiệu chỉnh là 35. Với mức ý nghĩa 5%, có thể xem tuổi thọ trung bình của 2 loại bóng đèn là như nhau không? Giả thiết tuổi thọ của 2 loại bóng đèn này tuân theo quy luật chuẩn.

Câu 11: Công ty nước giải khát Coca - Cola đang nghiên cứu việc đưa vào một công thức mới để cải tiến sản phẩm của mình. Với công thức cũ, khi cho 500 người dùng thử thì có 120 người ưa thích nó. Với công thức mới, khi đưa cho 1000 người khác dùng thử thì có 300 người tỏ ra ưa thích nó. Với mức ý nghĩa là 2% hãy cho biết có phải công thức mới đưa vào đã làm tăng tỷ lệ người ưa thích Coca - Cola không?

Câu 12: Một công ty sản xuất pin tuyên bố rằng pin của họ có tuổi thọ trung bình là 21,5 giờ. Một cơ quan kiểm tra chất lượng đã kiểm tra 6 chiếc pin của công ty và thu được kết quả về tuổi thọ 6 chiếc pin này là 19; 18; 22; 20; 16; 25 giờ. Kết quả này có xác nhận rằng quảng cáo của công ty là đúng hay không? Biết mức ý nghĩa là 5%.

Câu 13: Một nhóm nghiên cứu công bố rằng trung bình 1 người vào siêu thị A tiêu hết 140 ngàn đồng. Chọn một mẫu ngẫu nhiên gồm 50 người ta tính được số tiền trung bình họ tiêu là 154 ngàn với độ lệch tiêu chuẩn là 62 ngàn. Với mức ý nghĩa là 2%, hãy cho biết ý kiến về công bố trên

Câu 14: Người ta tiến hành 1 cuộc nghiên cứu để so sánh mức lương trung bình của một phụ nữ với mức lương trung bình của một nam giới ở một công ty lớn. Một mẫu ngẫu nhiên gồm 100 phụ nữ có mức lương trung bình là 7,23 $/giờ với độ lệch tiêu chuẩn là 1,64 $/giờ. Một mẫu gồm 75 nam giới có mức lương trung bình là 8,06$/giờ với độ lệch tiêu chuẩn là 1,85 $/giờ. Số liệu đã cho có chứng minh được rằng, mức lương trung bình của phụ nữ trong công ty là thấp hơn nam giới hay không? Cho biết mức ý nghĩa là 1% và giả thiết mức lương có phân phối chuẩn.

Câu 15:Người ta tiến hành một cuộc nghiên cứu để so sánh điểm trung bình của các vận động viên thể dục năm 1970 và năm 1995. Một mẫu ngẫu nhiên gồm 35 vận động viên của năm 1970 có số điểm trung bình là 267 với độ lệch tiêu chuẩn là 27, một mẫu gồm 40 vận động viên của năm 1995 có số điểm trung bình là 225 với độ lệch tiêu chuẩn là 30. Kiểm định xem có sự khác nhau hay không giữa 2 thế hệ vận động viên năm 1970 và 1995? Cho biết mức ý nghĩa là 5% và giả thiết số điểm của vận động viên có phân phối chuẩn.

Câu 16: Một quán ăn nói với nhân viên thu thuế rằng trung bình một ngày họ có 32 khách. Kiểm tra ngẫu nhiên 22 ngày cho thấy số khách trung bình 1 ngày là 37,2 và độ lệch tiêu chuẩn là 7,4. Với mức ý nghĩa là 2% hãy cho biết chủ quán nói đúng không?

Câu 17: Hai giáo sư A và B cùng dạy một môn ở 2 trường đại học lớn. Trong số 400 sinh viên theo học giáo sư A có 80 sinh viên thi trượt. Trong số 500 sinh viên theo học giáo sư B có 125 sinh viên thi trượt. Với mức ý nghĩa 5%, hãy kiểm định xem có sự khác nhau về tỷ lệ trượt giữa sinh viên của giáo sư A và sinh viên của giáo sư B không?

Câu 18: Một công ty có 1 hệ thống máy tính có thể xử lý 1200 hoá đơn trong 1 giờ. Công ty mới nhập một hệ thống máy tính mới. Hệ thống này khi chạy kiểm tra trong 40 giờ cho thấy số hoá đơn được xử lý trung bình trong một giờ là 1260 với độ lệch tiêu chuẩn là 215. Với mức ý nghĩa 5%, hãy nhận định xem hệ thống mới có tốt hơn hệ thống cũ không?

Câu 19: Tại xã A, kiểm tra 500 người thì có 50 người bị đau mắt hột. Tại xã B, kiểm tra 400 người thì có 60 người bị đau mắt hột với mức ý nghĩa 5%, có thể kết luận gi về tình hình đau mắt hột của 2 xã trên.

Câu 20: Trước chiến dịch vận động của một ứng viên A, người ta thăm dò ngẫu nhiên 10000 người dân, thấy có 4000 người ủng hộ người A. Sau cuộc vận động, trong số 20000 người được hỏi ngẫu nhiên thấy có 8500 người ủng hộ người A. Với mức ý nghĩa 5%, hãy kết luận xem cuộc vận động có làm tăng sự ủng hộ đối với A hay không?

CHƯƠNG 7: TƯƠNG QUAN HỒI QUY

7.1.Khái niệm

Ở các phần trước, ta chỉ khảo sát 1 ĐLNN trên tổng thể. Tuy nhiên, trong thực tế, đôi khi chúng ta cần khảo sát đồng thời 2 ĐLNN trong mối quan hệ ràng buộc với nhau trên cùng 1 tổng thể.

Ví dụ:

i.Quan hệ giữa chiều cao X và cân nặng Y của 1 người.

ii.Độ dài quãng đường di chuyển X và thời gian di chuyển Y

Qua các ví dụ trên, chúng ta thấy sự phụ thuộc giữa các đặc tính có thể thể hiện dưới 2 dạng chính sau:

Sự phụ thuộc hàm

Ở ví dụ (ii), mối quan hệ giữa X và Y được thể hiện bởi hàm: X = VY

Sự phụ thuộc thống kê

Ở ví dụ (i) các ĐLNN X, Y không bị ràng buộc với nhau bằng 1 quy tắc chạt chẽ như ví dụ (ii), nhưng giữa các đặc tính đó vẫn có 1 sự tương quan nhất định.

Khi cho X nhận 1 giá trị x, thì Y có thể nhận nhiều giá trị khác nhau, các giá trị này phân phối quanh giá trị trung bình $\overline{Y_X}$ tương ứng với x.

Trong trường hợp đó, ta nói Y có sự phụ thuộc thống kê đối với X. Nói cách khác, sự phụ thuộc thống kê của Y đối với X biểu hiện ở chỗ khi X thay đổi thì phân phối các giá trị của Y thay đổi.

Trong trường hợp đặc biệt, khi sự phụ thuộc thống kê biểu diễn dưới dạng: Nếu 1 trong 2 đại lượng X và Y thay đổi thì giá trị trung bình của đại lượng kia cũng thay đổi như 1 hàm của đại lượng đó thì ta nói, 2 đại lượng đó có tương quan lẫn nhau.

Ví dụ: Y: sản lượng lúa

X: lượng phân bón

Vì cùng 1 diện tích đất, cùng 1 lượng phân bón, sản lượng lúa có thể khác nhau nên Y không phụ thuộc hàm đối với X.

Tuy nhiên thực nghiêm cho thấy, sản lượng trung bình của lúa lại là hàm của phân bón, tức là Y và X tương quan lẫn nhau.

Nhận xét trực quan sự phụ thuộc tương quan bằng đồ thị phân tán

Xét 2 ĐLNN X và Y

Gọi x_i , y_i (i = 1, 2,…,n) là các giá trị cụ thể của các đặc tính X và Y.

Ta biểu diễn mỗi cặp số (x_i , y_i) bởi điểm M_i trên mặt phẳng hệ tọa độ xOy. Các điểm M_i lập nên một đám mây thống kê thường được gọi là đồ thị phân tán.
Các điểm M_i nằm sát theo đường thẳng hay một đường parabol,..thì ta phải nghĩ rằng giữa X và Y có liên quan hàm.
Các điểm M_i nằm rải rác không theo một quy tắc nào thì ta nghĩ rằng X và Y độc lập.
Các điểm M_i nằm tập trung vào 1 vùng nhất định có dạng hình bầu dục

*)Nếu trục lớn của hình bầu dục nghiêng lên, ta kết luận X và Y đồng biến, có tương quan thuận.

**)Nếu trục lớn của hình bầu dục nghiêng xuống, ta kết luận X và Y nghịch biến, có tương quan nghịch

Trong phạm vi của phần này, chúng ta chỉ xét dạng tương quan tuyến tính.

7.2. Mạng tương quan, bảng tương quan, đường hồi quy thực nghiệm.

7.2.1. Mạng tương quan

Giả sử X và Y là 2 ĐLNN trên một tổng thể. Chọn ngẫu nhiên một mẫu kích thước n, trong đó (X, Y) nhận các giá trị (x_1 , y_1), (x_2 , y_2) , …., (x_n , y_n). Sắp xếp và đánh số lại, chúng ta có thể viết các cặp giá trị này của X và Y vào bảng sau:

Y \ X	x_1	x_2	…	x_i	…	x_r	n_y
y_1	n_{11}	n_{12}	…	n_{1i}	…	n_{1r}	n_{y1}
y_2	n_{21}	n_{22}	…	n_{2i}	…	n_{2r}	n_{y2}
…	…	…	…	…	…	…	…
y_j	n_{j1}	n_{j2}	…	n_{ji}	…	n_{jr}	n_{yj}
…	…	…	…	…	…	…	…
y_s	n_{s1}	n_{s2}	…	n_{si}	…	n_{sr}	n_{ys}
n_x	n_{x1}	n_{x2}	…	n_{xs}	…	n_{xr}	n

Bảng trên được gọi là mạng tương quan.

Ví dụ: Độ tuổi X và chiều cao Y(cm) của 100 em bé được cho bằng mạng tương quan sau:

Y \ X	4	6	8	n_y
80	15	8	2	25
90	10	32	13	55
100	0	5	15	20
n_x	25	45	30	n = 100

7.2.2. Bảng tương quan

Ứng với mỗi giá trị X = x_i ta có 1 dãy giá trị Y xuất hiện những tần số khác nhau, nên ta tìm cách thay thế phân phối các giá trị Y đó bằng 1 đại lượng có đủ tư cách đại diện cho chúng.

Người ta sử dụng $\overline{Y_X}$ = E(Y_X) làm giá trị tiêu biểu đại diện cho các giá trị Y_X ứng với X = x.

Đặt g(x) = $\overline{Y_X}$. Khi đó

*) Hàm số này được gọi là hàm hồi quy của Y theo X.

*) Phương trình g(x) = $\overline{Y_X}$ được gọi là phương trình hồi quy của Y theo X.

*) Đồ thị của hàm g được gọi là đường hồi quy của Y theo X.

*) Tương tự, ta có h(y) = $\overline{X_Y}$ được gọi là phương trình hồi quy của X theo Y.

Ví dụ: xét độ tuổi và chiều cao của 100 em bé ở ví dụ (a), ta có:

+) Chiều cao trung bình của 1 em bé 4 tuổi là:

$$\overline{Y}_4 = \frac{80\times15+90\times10+100\times0}{25} = 84cm$$

+) Chiều cao trung bình của 1 em bé 6 tuổi là:

$$\overline{Y}_6 = \frac{80\times8+90\times32+100\times5}{45} = 89.33cm$$

+) Chiều cao trung bình của 1 em bé 8 tuổi là:

$$\overline{Y}_6 = \frac{80\times2+90\times13+100\times15}{30} = 94.33cm$$

Từ đó, ta có bảng tương quan của Y theo X là:

X	4	6	8
$\overline{Y_X}$	84	89.33	94.33

Qua ví dụ trên, ta thấy các điểm của hàm g(x) = $\overline{Y_X}$ sắp xếp theo 1 đường gần thẳng. Vì vậy, ta có thể dự đoán rằng có thể xấp xỉ g(x) với 1 hàm tuyến tính dạng : $\hat{y} = \hat{a}\,x + \hat{b}$

Vì trong sự xấp xỉ này có sai số nên ta cần phải xác định các giá trị $\hat{a}$ và $\hat{b}$ sao cho sai số là bé nhất.

7.2.3. Cách xác định đường hồi quy tuyến tính.

Giả sử các đại lượng X và Y có hồi quy tuyến tính, khi đó dường hồi quy tuyến tính là đường thẳng.

Để xác định đường hồi quy, ta thực hiện n phép thử độc lập và thu được các cặp giá trị của (X, Y) là (x_1, y_1), (x_2, y_2), …., (x_n, y_n).

Vấn đề đặt ra là từ các số liệu đó hãy xác định đường hồi quy. Ở đây ta xét trường hợp đơn giản, mỗi giá trị x_i của X chỉ tương ứng 1 giá trị của Y là y.

Vì đường hồi quy là đường thẳng nên có dạng $\hat{y} = \hat{a}\,x + \hat{b}$.

Ta cần xác định đường thẳng thể hiện sự phụ thuộc của Y theo X dựa vào các cặp(x_i, y_i) . Có nhiều cách xác định đường thẳng đó. Ở đây ta xác định đường thẳng sao cho tổng bình phương của các khoảng cách (theo chiều song song với oy) từ các điểm $M_i(x_1, y_1)$ đến dường đó là bé nhất..

Bài toán dẫn đến việc xác định $\hat{a}$ và $\hat{b}$ sao cho tổng sau bé nhất:

$$F = \sum_{i=1}^{n} (\hat{y_i} - y_i)^2 = \sum_{i=1}^{n} \left(\hat{a}\ x_i + \hat{b} - y_i\right)^2$$

Muốn thế ta phải có:

$$\begin{cases} \dfrac{\partial F}{\partial \hat{a}} = 0 \\ \dfrac{\partial F}{\partial \hat{b}} = 0 \end{cases} \Leftrightarrow \begin{cases} 2\displaystyle\sum_{i=1}^{n} \left(\hat{a}\ x_i + \hat{b} - y_i\right)x_i = 0 \\ 2\displaystyle\sum_{i=1}^{n} \left(\hat{a}\ x_i + \hat{b} - y_i\right) = 0 \end{cases}$$

Giải hệ trên ta có

$$\hat{a} = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \times \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$\hat{b} = \frac{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i \times \sum_{i=1}^{n} x_i y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

Phương trình $\hat{y} = \hat{a}\,x + \hat{b}$ thường được viết lại là g(x) = $\overline{Y_X} = \hat{a}\,x + \hat{b}$. Đây được gọi là phương trình hồi quy tuyến tính mẫu.

Cách xác định $\hat{a}$ và $\hat{b}$ như trên được gọi là phương pháp bình phương bé nhất. $\hat{a}$ và $\hat{b}$ còn có thể được viết lại dưới dạng

$$\hat{a} = \frac{\overline{XY} - \bar{X}\bar{Y}}{\overline{X^2} - (\bar{X})^2}; \qquad \hat{b} = \bar{y} - \hat{a}\bar{X}$$

Ví dụ: Lập phương trình hồi quy tuyến tính giữa độ tuổi X và chiều cao Y(cm) của 100 em bé:

Y \ X	4	6	8	n_y
80	15	8	2	25
90	10	32	13	55
100	0	5	15	20
n_x	25	45	30	n = 100

*)Lập phương trình hồi quy tuyến tính của Y đối với X.

Ta có:

$$\bar{X} = 6.1;\ \bar{Y} = 89.5;\ \overline{XY} = 551.6;\ \overline{X^2} = 39.4;\ S^2(X) = 2.19$$

Suy ra $\hat{a}$ = 2.58; $\hat{b} = 73.76$

Do vậy g(x) = 2.58$x + 73.76$ là phương trình hồi quy tuyến tính của Y đối với X.

**)Ta cũng có thể tính phương trình hồi quy tuyến tính của X đối với Y như sau

Ta có:

$$\bar{X} = 6.1;\ \bar{Y} = 89.5;\ \overline{XY} = 551.6;\ \overline{Y^2} = 8055;\ S^2(Y) = 44.75$$

Suy ra $\hat{a}$ = 0.126; $\hat{b} = -5.177$

Do vậy h(y) = 0,126y – 5.177 là phương trình hồi quy tuyến tính của Y đối với X.

7.2.4.Hệ số tương quan mẫu.

Cho X, Y là 2 ĐLNN rời rạc. Khi đó hệ số tương quan giữa X và Y là:

$$r = \frac{\overline{XY} - \bar{X}\bar{Y}}{S(X)S(Y)} = \frac{\sum_{i=1}^{n}(x_i - \bar{X})(y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{X})^2 \sum_{i=1}^{n}(y_i - \bar{Y})^2}}$$

**Tính chất của hệ số tương quan:*

+) -1 ≤ r ≤ 1

+) Nếu r = ± 1 thì X và Y tương quan tuyến tính.

+) Nếu r = 0 thì X và Y không tương quan tuyến tính

+) r đặc trưng cho mối quan hệ tuyến tính giữa X và Y. Nếu $|r|$ càng lớn thì X và Y càng có quan hệ chặt chẽ.

**) Liên hệ giữa hồi quy và hệ số tương quan:* $\hat{a} = r \times \frac{s_Y}{s_X}$

Trong đó, S_X, S_Y là các độ lệch chuẩn mẫu.

Ví dụ: Tìm hệ số tương quan mẫu giữa độ tuổi và chiều cao của 100 em bé

Y \ X	4	6	8	n_y
80	15	8	2	25
90	10	32	13	55
100	0	5	15	20
n_x	25	45	30	n = 100

Ta có :

$$r = \frac{\overline{XY} - \bar{X}\bar{Y}}{S(X)S(Y)} = 0.5707$$

Bài tập chương 7:

Bài 1: Nghiên cứu mối liên hệ giữa hai chỉ tiêu X(cm) và Y (%) của một loại thép, người ta tiến hành kiểm tra trên một số sản phẩm và thu được kết quả ở bảng dưới đây.

X \ Y	1 – 3	3 – 5	5 - 7	7 - 9	9 – 11
50	1	2	3		
60		6	6	5	
70		3	6	1	
80			4	6	2

Giả sử giữa *X* và *Y* có quan hệ tuyến tính $Y = a + bX$, hãy ước lượng các hệ số a, b bằng phương pháp bình phương bé nhất. Hãy cho kết quả dự đoán của *Y* nếu *X* = 90(cm).

Bài 2 : Cho *Y* là sản lượng mùa màng và *X* là lượng phân bón, thống kê được bảng sau đây :

X	1	2	3	4	5
Y	12	14	13	15	16

Hãy xác định hệ số tương quan mẫu của *Y* đối với *X*, chứng tỏ rằng có thể coi đường hồi quy trung bình mẫu của *Y* đối với X là đường thẳng. Hãy tìm đường thẳng đó.

Bài 3 : Hãy đánh giá mức độ tương quan giữa tổng giá trị hàng hóa xuất khẩu (Y) và tiền trợ cấp hưu trí (X) trên cơ sở số liệu của 10 năm như sau:

X (tỷ/năm)	5,1	5,3	5,2	4,9	4,8	4,7	4,5	5,0	4,6	4,4
$Y(tỷ/năm$	22	30	35	29	27	36	40	39	42	45

KẾT LUẬN

Đào tạo theo học chế tín chỉ đòi hỏi người học phải dành rất nhiều thời gian cho quá trình tự học, tự nghiên cứu.Tập bài giảng này tồn tại với tư cách là tài liệu có hướng dẫn nhưng sẽ được sử dụng có hiệu quả trong các hình thức tự học của sinh viên : Tự học không có sự hướng dẫn của giáo viên – Tự học có sự hướng dẫn từ xa của giáo viên – Tự học trong hoạt động dạy học và cũng được sử dụng có hiệu quả trong hình thức lên lớp (lớp – bài) – hình thức tổ chức dạy học cơ bản trong nhà trường. Bài giảng đã sử dụng nhiều ví dụ minh họa để giúp sinh viên dễ dàng nắm bắt được các khái niệm, các vấn đề, nhằm phục vụ việc giảng dạy môn Xác suất thống kê tại trường Đại học Dân lập Hải Phòng.

Nhân đây tác giả cũng xin bày tỏ rằng đây là một đề tài hoàn toàn mới, từ trước tới nay chưa có một công trình nghiên cứu tương tự để tác giả có thể học tập và rút kinh nghiệm. Tuy nhiên, tìm hiểu một lĩnh vực nghiên cứu mới mẻ và có ý nghĩa thực tiễn đối với hoạt động giảng dạy là đóng góp mới của đề tài. Tuy chắc chắn không tránh khỏi còn có thiếu sót và hạn chế, đề tài là kết quả làm việc nỗ lực và nghiêm túc của các tác giả.

TÀI LIỆU THAM KHẢO

1. Nguyễn Cao Văn, *Giáo trình Lý thuyết xác suất & thống kê,* Đề cương giảng dạy môn toán cao cấp theo niên chế của trường ĐH DL HP
2. Đề cương giảng dạy môn toán cao cấp của trường Học viện kỹ thuật quân sự.
3. Đề cương giảng dạy môn toán cao cấp của trường ĐH Hải phòng
4. Đề cương giảng dạy môn toán cao cấp của trường ĐH Công nghiệp Hà nội
5. Hoàng Xuân Sính, Đại số đại cương, NXBGD, 1997
6. Nguyễn Đình Trí, *Toán học cao cấp tập 1,* NXB GD, 2006
7. Nguyễn Đình Trí, *Toán học cao cấp tập 2,* NXB GD, 2006
8. Nguyễn Đình Trí, *Toán học cao cấp tập 3,* NXB GD, 2006
9. Nguyễn Đình Trí, *Bài tập toán học cao cấp tập 1,* NXB GD, 2006
10. Nguyễn Đình Trí, *Bài tập toán học cao cấp tập 2,* NXB GD, 2006
11. Nguyễn Đình Trí, *Bài tập toán học cao cấp tập 3,* NXB GD, 2006
12. Lê Ngọc Lăng –*Ôn thi học kỳ và thi vào giai đoạn 2, Tập 1,* NXBGD, 1997
13. Lê Ngọc Lăng –*Ôn thi học kỳ và thi vào giai đoạn 2, Tập 2,* NXBGD, 1997
14. G.M.Fichtengon – *Cơ sở giải tích toán học tập 1*, NXBĐH & THCN, 1986
15. G.M.Fichtengon – *Cơ sở giải tích toán học tập 2,* NXBĐH & THCN, 1986
16. Y.Y.Liaskô, *Giải tích toán học tập1*, NXBĐH & THCN, 1978
17. Y.Y. Liaskô, *Giải tích toán học tập2*, NXBĐH & THCN, 1978
18. Y.Y. Liaskô, *Giải tích toán học tập3*, NXBĐH & THCN, 1978
19. P.E.Đancô, Bài tập toán cao cấp phần 1, NXBĐH & THCN, 1983
20. P.E.Đancô, Bài tập toán cao cấp phần 1, NXBĐH & THCN, 1983
21. Jean- Marie Monier, *Giải tích 1*, NXBGD, 2006
22. Jean- Marie Monier, *Giải tích 2*, NXBGD, 2006
23. Jean- Marie Monier, *Giải tích 3*, NXBGD, 2006
24. Jean- Marie Monier, *Giải tích 4*, NXBGD, 2006
25. Jean- Marie Monier, *Đại số 1*, NXBGD, 2006
26. Jean- Marie Monier, *Đại số 2*, NXBGD, 2006
27. Jean- Marie Monier, *Hình học,* NXBGD, 2006