ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC KỸ THUẬT CÔNG NGHIỆP

BÁO CÁO TỔNG KẾT
ĐỀ TÀI KHOA HỌC VÀ CÔNG NGHỆ CẤP TRƯỜNG

XÂY DỰNG BÀI GIẢNG TRỰC TUYẾN MÔN XÁC SUẤT THỐNG KÊ

Mã số: T2022-VD18

Chủ nhiệm đề tài: ThS. Nguyễn Thị Xuân Mai

Thái Nguyên – 9/2023

ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC KỸ THUẬT CÔNG NGHIỆP

BÁO CÁO TỔNG KẾT
ĐỀ TÀI KHOA HỌC VÀ CÔNG NGHỆ CẤP TRƯỜNG

XÂY DỰNG BÀI GIẢNG TRỰC TUYẾN MÔN XÁC SUẤT THỐNG KÊ

Mã số: T2022-VD18

Xác nhận của tổ chức chủ trì	Chủ nhiệm đề tài
KT. HIỆU TRƯỞNG PHÓ HIỆU TRƯỞNG	
PGS. TS Vũ Ngọc Pi	Nguyễn Thị Xuân Mai

Thái Nguyên – 9/2023

MỤC LỤC

TRƯỜNG ĐẠI HỌC KỸ THUẬT CÔNG NGHIỆP

Đơn vị: Khoa Khoa học Cơ bản và Ứng dụng

THÔNG TIN KẾT QUẢ NGHIÊN CỨU

1. Thông tin chung:

- Tên đề tài: Xây dựng bài giảng trực tuyến môn Xác suất Thống kê
- Mã số: T2022-VD18
- Chủ nhiệm: Nguyễn Thị Xuân Mai
- Cơ quan chủ trì: Trường Đại học Kỹ thuật Công nghiệp
- Thời gian thực hiện: 04/2022 – 08/2023

2. Mục tiêu:

Xây dựng kho học liệu số môn Xác suất Thống kê trong trường Đại học Kỹ thuật Công nghiệp, đồng thời nâng cao kỹ năng ứng dụng công nghệ thông tin trong hoạt động dạy và học.

3. Kết quả nghiên cứu:

Xây dựng được 17 video bài giảng môn Xác suất Thống kê.

4. Sản phẩm:

Sản phẩm ứng dụng: 17 video bài giảng môn Xác suất Thống kê.

5. Hiệu quả:

Góp phần nâng cao hiệu quả việc học tập cho sinh viên, thay đổi cách tiếp cận trong việc giảng dạy cũng như cách học của người học đối với học phần Xác suất Thống kê.

6. Khả năng áp dụng và phương thức chuyển giao kết quả nghiên cứu:

Có khả năng áp dụng trong giảng dạy và học tập môn Xác suất Thống kê tại trường Đại học kỹ Thuật Công nghiệp – Đại học Thái Nguyên.

Ngày tháng năm 2023

Cơ quan chủ trì	**Chủ nhiệm đề tài**
KT.HIỆU TRƯỞNG	
PHÓ HIỆU TRƯỞNG	
PGS.TS. Vũ Ngọc Pi	**Nguyễn Thị Xuân Mai**

INFORMATION ON RESEARCH RESULTS

1. General information:

Project title: Developing online lectures on Probability Statistics

Code number: T2022-VD18

Coordinator: Nguyen Thi Xuan Mai

Implementing institution: Thai Nguyen University of Technology

Duration: from 4/2022 to 4/2023

2. Objective(s):

Building a digital repository of Probability Statistics subjects at the University of Technology and Industry, and at the same time improving skills in applying information technology in teaching and learning activities.

3. Research results:

Built 17 videos of lectures on Probability Statistics.

4. Products:

Application products: Built 17 videos of lectures on Probability Statistics.

5. Effects:

To contribute to improving students' learning efficiency, change the approach in teaching as well as the learners' learning style for the Probability Statistics.

6. Transfer alternatives of reserach results andapplic ability:

Able to apply in teaching and learning Probability Statistics at Thai Nguyen University of Technology

CHƯƠNG 1
MỞ ĐẦU

1. Tính cấp thiết của đề tài

Xu thế toàn cầu hóa và hội nhập quốc tế đặt ra những yêu cầu mới cho giáo dục đại học. Cách thức tổ chức và phương pháp giảng dạy tại các trường đại học cần thay đổi. Công nghệ phát triển với chi phí rẻ là điều kiện thuận lợi để các trường đại học đầu tư cơ sở vật chất, các công cụ và phương tiện giảng dạy hiện đại. Bên cạnh hình thức giảng dạy trực tiếp cho người học, các trường cần sử dụng nhiều hơn các hình thức khác như đào tạo online, thiết kế môi trường ảo để người học và người dạy có thể tương tác lẫn nhau và truyền đạt thông tin, tổ chức thực hành tại các phòng thí nghiệm hay phòng mô phỏng ảo…

Việc đưa hệ thống E-learning vào hoạt động tại trường Đại học, tạo ra một kênh học tập khác góp phần nâng cao chất lượng đào tạo. Hiện nay việc sử dụng hệ thống E-learning đã trở thành tự giác đối với hầu hết giảng viên và sinh viên trong Trường vì những lợi ích thiết thực mà hệ thống mang lại. Video là một phương tiện truyền thông phong phú và mạnh mẽ được sử dụng trong elearning. Nó có thể trình bày thông tin một cách hấp dẫn và nhất quán.

Để nâng cao chất lượng đào tạo trong tình hình mới, đáp ứng nhu cầu về nguồn lao động chất lượng cao cho đất nước, Trường Đại học Kỹ thuật Công nghiệp đã triển khai xây dựng các bài giảng điện tử. Từ đó, tác giả đăng ký đề tài "Xây dựng bài giảng trực tuyến môn Xác suất Thống kê" để cung cấp thêm nguồn tài liệu cho công tác giảng dạy và học tập của Nhà trường.

2. Mục tiêu nghiên cứu

Xây dựng kho học liệu số môn Xác suất Thống kê trong trường Đại học Kỹ thuật Công nghiệp, đồng thời nâng cao kỹ năng ứng dụng công nghệ thông tin trong hoạt động dạy và học.

3. Đối tượng và phạm vi nghiên cứu

3.1. Đối tượng nghiên cứu: Xây dựng video bài giảng các nội dung cơ bản của học phần Xác suất Thống kê.

3.2. Phạm vi nghiên cứu: Các nội dung dạy và học học phần Xác suất Thống kê tại trường Đại học Kỹ thuật Công nghiệp

4. Cách tiếp cận, phương pháp nghiên cứu

4.1. Cách tiếp cận: Ứng dụng công nghệ thông tin và các phương pháp dạy học tích cực để xây dựng các video bài giảng.

4.2. Phương pháp nghiên cứu:

- Phương pháp nghiên cứu lý thuyết.

- Phương pháp thực nghiệm sư phạm

5. Kết quả nghiên cứu

Ngoài phần mở đầu, mục lục, danh mục tài liệu tham khảo, phụ lục, đề tài nghiên cứu đã xây dựng được 17 video bài giảng môn Xác suất Thống kê. Thời lượng mỗi video từ 15-20 phút. Cụ thể như sau:

Chương 1: 04 video

Chương 2: 02 video

Chương 3: 02 video

Chương 4: 03 video

Chương 5: 04 video

Chương 6: 02 video

CHƯƠNG 2
NỘI DUNG VÀ KẾT QUẢ NGHIÊN CỨU

Sau đây là nội dung của 17 video bài giảng môn Xác suất Thống kê được trích từ các slide PowerPoint đã ghi âm.

CHƯƠNG 1

BIẾN CỐ NGẪU NHIÊN VÀ CÁC PHÉP TÍNH VỀ XÁC SUẤT

+ Định nghĩa về xác suất
+ Phân loại các loại biến cố
+ Các công thức tính xác suất

1.1. Phép thử và các loại biến cố :

1.1.1. Định nghĩa phép thử, biến cố

+ Khi thực hiện 1 số các điều kiện nào đó, ta gọi đó là 1 *phép thử*. Còn những hiện tượng được xét trong phép thử sẽ gọi là các *biến cố*.

+ Các biến cố thường được ký hiệu là các chữ cái in hoa như A, B, C,...,X, Y,...., A_1, A_2,....A_n, B_1, B_2,.....,B_n,....

Ví dụ :

+ Phép thử: Tung một con xúc xắc

+ Biến cố: Xúc xắc xuất hiện mặt có i chấm (i = 1..6)

Gọi A_i là biến cố xúc xắc xuất hiện **mặt có i chấm**

1.1.2. Phân loại các biến cố

+ **Biến cố không thể có**: Là biến cố nhất định không xảy ra sau phép thử, ký hiệu là $\emptyset$.

+ **Biến cố chắc chắn**: là biến cố nhất định xảy ra sau phép thử, ký hiệu là Ω.

+ **Biến cố ngẫu nhiên**: là biến cố có thể xảy ra hoặc không xảy ra sau phép thử.

Ví dụ : Tung một con xúc xắc

+ Gọi A là biến cố xúc xắc xuất hiện mặt có số chấm lớn hơn 6

→ A là biến cố không thể có (A = $\emptyset$).

+ Gọi B là biến cố xúc xắc xuất hiện mặt có số chấm nhỏ hơn hoặc bằng 6

→ B là biến cố chắc chắn (B = Ω).

+ A_2 : biến cố xúc xắc xuất hiện **mặt có 2 chấm**

→ A_2 là các biến cố ngẫu nhiên.

1.1.3. Mối quan hệ giữa các biến cố

+ **Biến cố kéo theo**:

Biến cố A được gọi là *kéo theo* biến cố B nếu sự xảy ra của A kéo theo sự xảy ra của B.

Ví dụ : Tung một con xúc xắc

Gọi:

+ **A** là biến cố xúc xắc xuất hiện mặt có 2 chấm.

+ **B** là biến cố xúc xắc xuất hiện mặt có số chấm chẵn.

Khi đó biến cố A kéo theo biến cố B.

+ **Hợp (tổng) của 2 biến cố**

Biến cố A là hợp (tổng) của hai biến cố B và C, kí hiệu là A = B + C, nếu A xảy ra khi ít nhất một trong hai biến cố B hoặc C phải xảy ra.

Ví dụ : Một người đi phỏng vấn xin việc vào hai công ty

Gọi:

+ A_1 là biến cố người đó trúng tuyển vào công ty thứ nhất.

+ A_2 là biến cố người đó trúng tuyển vào công ty thứ hai.

+ A là biến cố người đó xin được việc.

Khi đó : $A = A_1 + A_2$

+ Giao (tích) của 2 biến cố

Biến cố A gọi là biến cố giao (tích) của hai biến cố B và C, kí hiệu là A = B.C, nếu A xảy ra khi đồng thời cả hai biến cố B và C cùng xảy ra.

Ví dụ

Hai người cùng đến phỏng vấn xin việc vào một công ty

Gọi:

+ A_1 là biến cố người thứ nhất trúng tuyển.

+ A_2 là biến cố người thứ hai trúng tuyển.

+ A là biến cố cả hai người cùng trúng tuyển.

Khi đó : $A = A_1 . A_2$

+ Hợp (tổng) của 2 biến cố

Biến cố A là hợp (tổng) của hai biến cố B và C, kí hiệu là **A = B + C**, nếu A xảy ra khi ít nhất một trong hai biến cố B hoặc C phải xảy ra.

+ Giao (tích) của 2 biến cố

Biến cố A gọi là biến cố giao (tích) của hai biến cố B và C, kí hiệu là **A = B.C**, nếu A xảy ra khi đồng thời **cả hai** biến cố B **và** C cùng xảy ra.

+ Biến cố xung khắc:

Hai biến cố A và B được gọi là *xung khắc* với nhau nếu biến cố A xảy ra thì biến cố B không xảy ra và ngược lại.

→ A và B là hai biến cố xung khắc thì A.B = ∅

Ví dụ : Hai người cùng đến phỏng vấn xin việc vào một vị trí của một công ty (công ty chỉ tuyển một người cho vị trí đó)

+ A_1 là biến cố người thứ nhất trúng tuyển.

+ A_2 là biến cố người thứ hai trúng tuyển.

Khi đó : A_1 và A_2 là hai biến cố xung khắc.

+ Biến cố đối lập

Biến cố không xảy ra biến cố A được gọi là biến cố *đối lập* với A, ký hiệu là $\bar{A}$.

Ví dụ 1 : Một người đi câu cá.

Nếu gọi A là biến cố người đó câu được cá thì $\bar{A}$ là biến cố người đó không câu được cá.

Ví dụ 2 : Lấy 1 sản phẩm bất kỳ ra kiểm tra.

Nếu gọi A là biến cố sản phẩm đó không đạt chuẩn thì $\bar{A}$ là biến cố sản phẩm đó đạt chuẩn .

+ Biến cố độc lập

A và B được gọi là 2 biến cố độc lập nếu biến cố này xảy ra hay không cũng không ảnh hưởng đến xác suất xảy ra biến cố kia và ngược lại.

Ví dụ : Hai máy hoạt động độc lập, cùng tham gia sản xuất trong một dây chuyền.

Nếu gọi: A_1 là biến cố máy thứ nhất bị hỏng

A_2 là biến cố máy thứ hai bị hỏng

thì A_1 và A_2 là hai biến cố độc lập.

+ Hệ đầy đủ các biến cố

Một hệ gồm n biến cố $\{A_1, A_2, ..., A_n\}$ được gọi là một *hệ đầy đủ các biến cố* nếu trong kết quả của phép thử sẽ xảy ra một và chỉ một trong n biến cố này.

→ $\{A_1, A_2, ..., A_n\}$ – hệ đầy đủ các biến cố nếu các A_i đôi một xung khắc và tổng các A_i là biến cố chắc chắn.

Ví dụ : Tung một con xúc xắc

Gọi A_i là biến cố xúc xắc xuất hiện **mặt có i chấm** (i = 1..6)

→ $\{A_1, A_2, ..., A_6\}$ - hệ đầy đủ các biến cố.

Ví dụ : Có 3 hộp đựng bi. Lấy ra một hộp bất kì trong 3 hộp đó

Gọi A_1, A_2, A_3 lần lượt là biến cố lấy ra hộp thứ nhất, hộp thứ hai, hộp thứ ba.

→ $\{A_1, A_2, A_3\}$ - hệ đầy đủ các biến cố.

1.1.3. Mối quan hệ giữa các biến cố

+ Biến cố kéo theo.

+ Hợp (tổng) của hai biến cố.

+ Giao (tích) của hai biến cố.

+ Biến cố xung khắc.

+ Biến cố đối lập.

+ Biến cố độc lập.

+ Hệ đầy đủ các biến cố.

1.2. Định nghĩa xác suất

Xác suất của biến cố A là một con số đặc trưng cho khả năng xuất hiện của biến cố A trong phép thử tương ứng, ký hiệu là p(A).

* **<u>Định nghĩa:</u>** Nếu trong một phép thử có tất cả n trường hợp đồng khả năng, trong đó có m trường hợp thuận lợi cho biến cố A thì **p(A) =** $\frac{m}{n}$

→ Nếu A là biến cố chắc chắn thì p(A) = 1.

→ Nếu A là biến cố không thể xảy ra thì p(A) = 0.

→ Nếu A là biến cố ngẫu nhiên thì 0 < p(A) < 1.

<u>Ví dụ 1:</u> Tung một con xúc xắc cân đối đồng chất.

Tính xác suất để :

a) Xúc xắc xuất hiện mặt có một chấm.

b) Xúc xắc xuất hiện mặt có số chấm là chẵn.

a/ Gọi A là biến cố xúc xắc xuất hiện mặt có một chấm

$$p(A) = \frac{1}{6}$$

b/ Gọi B là biến cố xúc xắc xuất hiện mặt có số chấm là chẵn

$$p(B) = \frac{3}{6} = \frac{1}{2}$$

<u>Ví dụ 2:</u> Một lô hàng có <u>12 sản phẩm</u> trong đó có <u>8 chính phẩm</u> và <u>4 phế phẩm</u>. Lấy ngẫu nhiên cùng một lúc ra 3 sản phẩm.

a) Tính xác suất để cả ba sản phẩm lấy ra đều là chính phẩm.

b) Tính xác suất để trong ba sản phẩm lấy ra có đúng hai chính phẩm.

a/ Gọi A là biến cố cả 3 sản phẩm lấy ra đều là chính phẩm.

$$p(A) = \frac{C_8^3}{C_{12}^3} = \frac{14}{55}$$

b/ Gọi B là biến cố trong 3 sản phẩm lấy ra có đúng 2 chính phẩm.

$$p(B) = \frac{C_8^2 . C_4^1}{C_{12}^3} = \frac{28}{55}$$

Tính xác suất bằng định nghĩa: **p(A) =** $\frac{m}{n}$

1.3. Các phép tính về xác suất:

1.3.1. Định lý cộng xác suất:

Xác suất của một tổng hai biến cố xung khắc bằng tổng xác suất của hai biến cố đó.

Tức là: Nếu A và B là hai biến cố <u>xung khắc</u> thì:

p(A + B) = p(A) + p(B)

→ Tổng quát: Cho $A_1, A_2, \ldots, A_n$ là một hệ các biến cố xung khắc từng đôi. Khi đó:

$$p\left(\sum_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} p(A_i)$$

→ **Hệ quả 1**: Nếu $A_1, A_2, \ldots, A_n$ là một hệ đầy đủ các biến cố thì tổng xác suất của chúng bằng 1.

$$\sum_{i=1}^{n} p(A_i) = 1$$

→ **Hệ quả 2**: Với 2 biến cố đối lập A và $\overline{A}$, ta luôn có:

$p(A) = 1 - p(\overline{A})$

Ví dụ 1: Một xạ thủ bắn một viên đạn vào một bia được chia làm 3 phần. Giả sử xác suất để xạ thủ đó bắn trúng phần 1, phần 2, phần 3 của bia lần lượt là 0,3; 0,2 ; 0,4. Tính xác suất để :

a) Xạ thủ đó bắn trúng phần 1 **hoặc** phần 3 của bia.

b) Xạ thủ đó bắn trúng bia.

c) Xạ thủ đó bắn không trúng bia.

Gọi A_i là biến cố xạ thủ bắn trúng phần i của bia (i = 1;2;3)

→ $p(A_1) = 0{,}3$; $p(A_2) = 0{,}2$; $p(A_3) = 0{,}4$

Gọi A là biến cố xạ thủ bắn trúng phần 1 **hoặc** phần 3 của bia.

Khi đó: $A = A_1 + A_3$

→ $p(A) = p(A_1 + A_3) \xrightarrow{A_1 \text{ và } A_3 \text{ xung khắc}} p(A_1) + p(A_3)$

$= 0{,}3 + 0{,}4 = \mathbf{0{,}7}$

Ví dụ 1: Một xạ thủ bắn một viên đạn vào một bia được chia làm 3 phần. Giả sử xác suất để xạ thủ đó bắn trúng phần 1, phần 2, phần 3 của bia lần lượt là 0,3; 0,2 ; 0,4. Tính xác suất để :
a) Xạ thủ đó bắn trúng phần 1 hoặc phần 3 của bia.

Gọi A_i là biến cố xạ thủ bắn trúng phần i của bia (i = 1;2;3)

→ $p(A_1) = 0{,}3$; $p(A_2) = 0{,}2$; $p(A_3) = 0{,}4$

Gọi B là biến cố xạ thủ bắn trúng bia.

(xạ thủ bắn trúng phần 1 hoặc phần 2 hoặc phần 3 của bia)

Khi đó: $B = A_1 + A_2 + A_3$

→ $p(B) = p(A_1 + A_2 + A_3) \xrightarrow[\text{đôi một xung khắc}]{A_1, A_2, A_3} p(A_1) + p(A_2) + p(A_3)$

$= 0{,}3 + 0{,}2 + 0{,}4$

$= \mathbf{0{,}9}$

Ví dụ 1: Một xạ thủ bắn một viên đạn vào một bia được chia làm 3 phần. Giả sử xác suất để xạ thủ đó bắn trúng phần 1, phần 2, phần 3 của bia lần lượt là 0,3; 0,2 ; 0,4. Tính xác suất để :
b) Xạ thủ đó bắn trúng bia.

Gọi A_i là biến cố xạ thủ bắn trúng phần i của bia (i = 1;2;3)

→ $p(A_1) = 0{,}3$; $p(A_2) = 0{,}2$; $p(A_3) = 0{,}4$

Gọi C là biến cố xạ thủ bắn không trúng bia.

Khi đó: $C = \overline{B}$

$\Rightarrow p(C) = p(\overline{B}) = 1 - p(B) = 1 - 0{,}9 = 0{,}1$

Ví dụ 1: Một xạ thủ bắn một viên đạn vào một bia được chia làm 3 phần. Giả sử xác suất để xạ thủ đó bắn trúng phần 1, phần 2, phần 3 của bia lần lượt là 0,3; 0,2 ; 0,4. Tính xác suất để :
b) Xạ thủ đó bắn trúng bia.
c) Xạ thủ đó bắn không trúng bia

Ví dụ 2:

Một hộp đựng **10 quả** cầu, trong đó có 6 quả cầu đỏ và 4 quả cầu xanh. Lấy ngẫu nhiên cùng một lúc ra 4 quả cầu.

Tính xác suất để trong 4 quả cầu lấy ra **có ít nhất** 2 quả màu đỏ.

(Trong 4 quả, lấy ra được 2 quả hoặc 3 quả hoặc 4 quả màu đỏ)

Gọi A_2, A_3, A_4 lần lượt là biến cố trong 4 quả lấy ra có 2 quả, 3 quả, 4 quả màu đỏ.

Gọi A là biến cố trong 4 quả lấy ra có **ít nhất** 2 quả màu đỏ

Khi đó: $A = A_2 + A_3 + A_4$

→ $p(A) = p(A_2 + A_3 + A_4) \xrightarrow[\text{đôi một xung khắc}]{A_2, A_3, A_4} p(A_2) + p(A_3) + p(A_4)$

Gọi A_2, A_3, A_4 lần lượt là biến cố trong 4 quả lấy ra có 2 quả, 3 quả, 4 quả màu đỏ.

$$p(A_2)=\frac{C_6^2.C_4^2}{C_{10}^4}=\frac{3}{7}$$

$$p(A_3)=\frac{C_6^3.C_4^1}{C_{10}^4}=\frac{8}{21}$$

$$p(A_4)=\frac{C_6^4}{C_{10}^4}=\frac{1}{14}$$

Vậy: $p(A)=p(A_2)+p(A_3)+p(A_4)$

$$=\frac{3}{7}+\frac{8}{21}+\frac{1}{14}\approx 0,88$$

> Ví dụ 2:
> Một hộp đựng 10 quả cầu, trong đó có 6 quả cầu đỏ và 4 quả cầu xanh. Lấy ngẫu nhiên cùng một lúc ra 4 quả cầu. Tính xác suất để trong 4 quả cầu lấy ra có ít nhất 2 quả màu đỏ.

1.3.2. Định lý nhân xác suất:

- **Định nghĩa**: Xác suất của biến cố A được tính với điều kiện biến cố B đã xảy ra gọi là xác suất có điều kiện của A, ký hiệu là p(A/B).

Gọi A là biến cố lần thứ nhất lấy được bi đỏ (A đã xảy ra)

Khi đó trong hộp chỉ còn 9 viên bi (gồm 6 viên bi đỏ và 3 viên bi xanh)

Gọi B là biến cố lần thứ hai lấy được bi đỏ

$$p(B)=p(B/A)$$

$$=\frac{C_6^1}{C_9^1}=\frac{2}{3}$$

> Ví dụ :
> Một hộp đựng 10 viên bi, trong đó có 7 viên bi đỏ và 3 viên bi xanh. Lấy ngẫu nhiên lần lượt 2 viên bi. Giả sử lần thứ nhất đã lấy được viên bi đỏ. Tính xác suất để lần thứ hai cũng lấy được bi đỏ?

*** Định lý nhân xác suất :**

Xác suất của tích hai biến cố A và B bằng tích xác suất của một trong hai biến cố đó với xác suất có điều kiện của biến cố còn lại.

p(A.B) = p(A).p(B/A)

Hoặc

p(A.B) = p(B).p(A/B)

Ví dụ 1: Một hộp đựng 10 viên bi xanh và 4 viên bi trắng. Lấy ngẫu nhiên lần lượt hai lần, mỗi lần một viên. Tính xác suất để cả hai lần đều lấy được bi xanh.

Gọi A là biến cố cả hai lần đều lấy được bi xanh

Gọi A_1, A_2 lần lượt là biến cố lần thứ nhất, lần thứ hai lấy được bi xanh

Khi đó: $A = A_1 . A_2$

→ $p(A) = p(A_1 .A_2) = p(A_1). p(A_2 /A_1)$

Trong đó: $p(A_1)=\frac{C_{10}^1}{C_{14}^1}=\frac{5}{7}$ $\qquad p(A_2/A_1)=\frac{C_9^1}{C_{13}^1}=\frac{9}{13}$

$$\Rightarrow p(A)=\frac{5}{7}.\frac{9}{13}=\frac{45}{91}$$

→ *Hệ quả* : Cho $A_1, A_2, ..., A_n$ là n biến cố độc lập thì:

$p(A_1.A_2.....A_n) = p(A_1).p(A_2).....p(A_n)$

Ví dụ 2: Một xí nghiệp có 3 máy nổ hoạt động độc lập nhau. Xác suất để trong một ngày máy nổ thứ nhất, thứ hai, thứ ba bị hỏng tương ứng là 0,2; 0,1; 0,15. Tính xác suất để trong một ngày:

a. Có đúng một máy nổ bị hỏng. b. Có đúng hai máy bị hỏng.

c. Cả ba máy đều không bị hỏng

Gọi A_1, A_2, A_3 lần lượt là biến cố trong một ngày máy nổ thứ nhất, thứ hai, thứ ba bị hỏng

→ $p(A_1) = 0,2$; $\qquad p(\overline{A_1}) = 1 - 0,2 = 0,8.$

$p(A_2) = 0,1$; $\qquad p(\overline{A_2}) = 1 - 0,1 = 0,9.$

$p(A_3) = 0,15$ $\qquad p(\overline{A_3}) = 1 - 0,15 = 0,85.$

Gọi A là biến cố trong một ngày có đúng một máy nổ bị hỏng

- Máy 1 hỏng và máy 2, máy 3 vẫn chạy $(A_1.\overline{A_2}.\overline{A_3})$
- Máy 2 hỏng và máy 1, máy 3 vẫn chạy $(\overline{A_1}.A_2.\overline{A_3})$
- Máy 3 hỏng và máy 1, máy 2 vẫn chạy $(\overline{A_1}.\overline{A_2}.A_3)$

Khi đó: $A=A_1.\overline{A_2}.\overline{A_3}+\overline{A_1}.A_2.\overline{A_3}+\overline{A_1}.\overline{A_2}.A_3$

$$\Rightarrow p(A)=p\left(A_1.\overline{A_2}.\overline{A_3}+\overline{A_1}.A_2.\overline{A_3}+\overline{A_1}.\overline{A_2}.A_3\right)$$

$$=p\left(A_1.\overline{A_2}.\overline{A_3}\right)+p\left(\overline{A_1}.A_2.\overline{A_3}\right)+p\left(\overline{A_1}.\overline{A_2}.A_3\right)$$

$$=p(A_1).p\left(\overline{A_2}\right).p\left(\overline{A_3}\right)+p\left(\overline{A_1}\right).p(A_2).p\left(\overline{A_3}\right)+p\left(\overline{A_1}\right).p\left(\overline{A_2}\right).p(A_3)$$

= 0,329

> Ví dụ 2: Một xí nghiệp có 3 máy nổ hoạt động độc lập nhau. Xác suất để trong một ngày máy nổ thứ nhất, thứ hai, thứ ba bị hỏng tương ứng là 0,2; 0,1; 0,15. Tính xác suất để trong một ngày:
> a. Có đúng một máy nổ bị hỏng.

Gọi B là biến cố trong một ngày có đúng hai máy nổ bị hỏng

- Máy 1 và máy 2 hỏng, máy 3 vẫn chạy $\left(A_1.A_2.\overline{A_3}\right)$
- Máy 1 và máy 3 hỏng, máy 2 vẫn chạy $\left(A_1.\overline{A_2}.A_3\right)$
- Máy 2 và máy 3 hỏng, máy 1 vẫn chạy $\left(\overline{A_1}.A_2.A_3\right)$

Khi đó: $B = A_1.A_2.\overline{A_3} + A_1.\overline{A_2}.A_3 + \overline{A_1}.A_2.A_3$

$$\Rightarrow p(B) = p(A_1.A_2.\overline{A_3} + A_1.\overline{A_2}.A_3 + \overline{A_1}.A_2.A_3)$$
$$= p(A_1.A_2.\overline{A_3}) + p(A_1.\overline{A_2}.A_3) + p(\overline{A_1}.A_2.A_3)$$
$$= p(A_1).p(A_2).p(\overline{A_3}) + p(A_1).p(\overline{A_2}).p(A_3) + p(\overline{A_1}).p(A_2).p(A_3)$$
$$= 0,2.0,1.0,85 + 0,2.0,9.0,15 + 0,8.0,1.0,15$$
$$= 0,056$$

Ví dụ 2: Một xí nghiệp có 3 máy nổ hoạt động độc lập nhau. Xác suất để trong một ngày máy nổ thứ nhất, thứ hai, thứ ba bị hỏng tương ứng là 0,2; 0,1; 0,15. Tính xác suất để trong một ngày:
b. Có đúng hai máy nổ bị hỏng.

Gọi C là biến cố trong một ngày cả ba máy nổ đều không bị hỏng

(Máy 1 chạy, máy 2 chạy và máy 3 chạy)

Khi đó: $C = \overline{A_1}.\overline{A_2}.\overline{A_3}$

$$\Rightarrow p(C) = p(\overline{A_1}.\overline{A_2}.\overline{A_3})$$
$$= p(\overline{A_1}).p(\overline{A_2}).p(\overline{A_3})$$
$$= 0,8.0,9.0,85$$
$$= 0,612$$

Ví dụ 2: Một xí nghiệp có 3 máy nổ hoạt động độc lập nhau. Xác suất để trong một ngày máy nổ thứ nhất, thứ hai, thứ ba bị hỏng tương ứng là 0,2; 0,1; 0,15. Tính xác suất để trong một ngày:
c. Cả ba máy nổ đều không bị hỏng.

1.3.3. Hệ quả của ĐL cộng và ĐL nhân xác suất

Xác suất của tổng hai biến cố không xung khắc bằng tổng xác suất các biến cố đó trừ đi xác suất của tích các biến cố đó.

$$\mathbf{p(A + B) = p(A) + p(B) - p(A.B)}$$

Ví dụ 1: Hai người cùng bắn vào một mục tiêu, mỗi người bắn một viên đạn với xác suất trúng mục tiêu của người thứ nhất là 0,7 và người thứ hai là 0,8. Tính xác suất để mục tiêu bị trúng đạn.

Gọi A_1, A_2 lần lượt là biến cố người thứ nhất, người thứ hai bắn trúng mục tiêu.

→ $p(A_1) = 0,7$; $p(A_2) = 0,8$

Gọi A là biến cố mục tiêu bị trúng đạn.

Khi đó $A = A_1 + A_2$

$$\rightarrow p(A) = p(A_1 + A_2) \xrightarrow{A_1 \text{ và } A_2 \text{ không xung khắc}} p(A_1) + p(A_2) - p(A_1.A_2)$$
$$\xrightarrow{A_1 \text{ và } A_2 \text{ độc lập}} p(A_1) + p(A_2) - p(A_1).p(A_2)$$
$$= 0,7 + 0,8 - 0,7.0,8 = 0,94$$

Ví dụ 1: Hai người cùng bắn vào một mục tiêu, mỗi người bắn một viên đạn với xác suất trúng (mục tiêu) của người thứ nhất là 0,7 và người thứ hai là 0,8. Tính xác suất để mục tiêu bị trúng đạn.

Ví dụ 2: Xác suất để động cơ thứ nhất của máy bay bị trúng đạn là 0,2; để động cơ thứ hai của máy bay bị trúng đạn là 0,3. Xác suất trúng đạn của phi công là 0,1. Tính xác suất để máy bay bị rơi, biết rằng máy bay rơi khi phi công bị trúng đạn **hoặc** cả hai động cơ bị trúng đạn.

(A_3) **($A_1.A_2$)**

Gọi A_1, A_2, A_3 lần lượt là biến cố động cơ thứ nhất, động cơ thứ hai, phi công bị trúng đạn.

→ $p(A_1) = 0,2$; $p(A_2) = 0,3$; $p(A_3) = 0,1$

Gọi A là biến cố máy bay bị rơi.

Khi đó: **$A = A_3 + A_1.A_2$**

$$\rightarrow p(A) = p(A_3 + A_1.A_2)$$
$$\xrightarrow{A_3 \text{ và } A_1.A_2 \text{ không xung khắc}} p(A_3) + p(A_1.A_2) - p(A_3.A_1.A_2)$$
$$\xrightarrow{A_1,\, A_2 \text{ và } A_3 \text{ độc lập}} p(A_3) + p(A_1).p(A_2) - p(A_3).p(A_1).p(A_2)$$
$$= 0,1 + 0,2.0,3 - 0,1.0,2.0,3 = \mathbf{0,154}$$

Ví dụ 2: Xác suất để động cơ thứ nhất của máy bay bị trúng đạn là 0,2; để động cơ thứ hai của máy bay bị trúng đạn là 0,3. Xác suất trúng đạn của phi công là 0,1. Tính xác suất để máy bay bị rơi, biết rằng máy bay rơi khi phi công bị trúng đạn hoặc cả hai động cơ bị trúng đạn.

1.4. Công thức xác suất đầy đủ, công thức Bayes

1.4.1. Công thức xác suất đầy đủ (toàn phần):

Giả sử $A_1, A_2, \ldots, A_n$ là một **hệ đầy đủ** các biến cố.

Biến cố A có thể xảy ra đồng thời với một trong các biến cố trên.

Khi đó:

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

Hay:

$$p(A) = p(A_1).p(A/A_1) + p(A_2).p(A/A_2) + \ldots + p(A_n).p(A/A_n)$$

1.4.2. Công thức Bayes:

Giả sử $A_1, A_2, \ldots, A_n$ là một **hệ đầy đủ** các biến cố. Biến cố A có thể xảy ra đồng thời với một trong các biến cố trên.

Giả sử **A đã xảy ra**. Khi đó:

$$p(A_i/A) = \frac{P(A_i).p(A/A_i)}{p(A)}$$

Ví dụ :

Có ba hộp. Hộp một có 7 chính phẩm và 3 phế phẩm.

Hộp hai có 8 chính phẩm và 2 phế phẩm. Hộp ba có 10 chính phẩm và 4 phế phẩm.

Lấy ngẫu nhiên một hộp, từ đó lấy ngẫu nhiên một sản phẩm.

a) Tính xác suất để sản phẩm lấy ra là chính phẩm.

b) Giả sử sản phẩm lấy ra là chính phẩm. Vậy sản phẩm đó có khả năng thuộc hộp nào nhất?

Gọi A là biến cố sản phẩm lấy ra là chính phẩm.

A_i là biến cố lấy ra hộp thứ i (i = 1, 2, 3).

Khi đó A_1, A_2, A_3 tạo thành hệ đầy đủ các biến cố và A xảy ra cùng một trong ba biến cố này.

$$\Rightarrow p(A) = p(A_1).p(A/A_1) + p(A_2).p(A/A_2) + p(A_3).p(A/A_3)$$

Trong đó: $p(A_1) = p(A_2) = p(A_3) = \frac{1}{3}$

$$p(A/A_1) = \frac{C_7^1}{C_{10}^1} = \frac{7}{10}$$

$$p(A/A_2) = \frac{C_8^1}{C_{10}^1} = \frac{8}{10} \qquad p(A/A_3) = \frac{C_{10}^1}{C_{14}^1} = \frac{10}{14}$$

$$\Rightarrow p(A) = \frac{1}{3}\left(\frac{7}{10} + \frac{8}{10} + \frac{10}{14}\right) \approx 0,74$$

Ví dụ : Có ba hộp. Hộp một có 7 chính phẩm và 3 phế phẩm. Hộp hai có 8 chính phẩm và 2 phế phẩm. Hộp ba có 10 chính phẩm và 4 phế phẩm. Lấy ngẫu nhiên một hộp, từ đó lấy ngẫu nhiên một sản phẩm.

a. Tính xác suất để sản phẩm lấy ra là chính phẩm.

Biến cố A đã xảy ra → Áp dụng công thức Bayes

$$p(A_1/A) = \frac{p(A_1).p(A/A_1)}{p(A)} = \frac{\frac{1}{3}.\frac{7}{10}}{0,74} \approx 0,315$$

$$p(A_2/A) = \frac{p(A_2).p(A/A_2)}{p(A)} = \frac{\frac{1}{3}.\frac{8}{10}}{0,74} \approx 0,36$$

$$p(A_3/A) = \frac{p(A_3).p(A/A_3)}{p(A)} = \frac{\frac{1}{3}.\frac{10}{14}}{0,74} \approx 0,322.$$

→ Vậy chính phẩm lấy ra có khả năng thuộc hộp hai nhất.

Ví dụ : Có ba hộp. Hộp một có 7 chính phẩm và 3 phế phẩm. Hộp hai có 8 chính phẩm và 2 phế phẩm. Hộp ba có 10 chính phẩm và 4 phế phẩm. Lấy ngẫu nhiên một hộp, từ đó lấy ngẫu nhiên một sản phẩm.

b) Giả sử sản phẩm lấy ra là chính phẩm. Vậy sản phẩm đó có khả năng thuộc hộp nào nhất?

1.5. Công thức Bernoulli

Tiến hành *n* phép thử độc lập. Trong mỗi phép thử chỉ có hai trường hợp : biến cố *A* xảy ra hoặc biến cố *A* không xảy ra ($\overline{A}$)

Giả sử p(A)= ***p*** và $p(\overline{A})$ = ***q = 1-p***

Khi đó: xác suất để trong *n* phép thử độc lập nói trên biến cố *A* xuất hiện đúng *k* lần là:

$$P_n(k) = C^k_n.p^k.q^{n-k}$$

Và xác suất để trong *n* phép thử độc lập nói trên biến cố *A* xuất hiện từ k_1 đến k_2 lần là:

$$P(k_1, k_2) = \sum_{k=k_1}^{k_2}(C^k_n.p^k.q^{n-k})$$

__Ví dụ:__ Một phân xưởng có 5 máy hoạt động độc lập nhau.

Xs để trong mỗi ca làm việc mỗi máy bị hỏng đều bằng 0,1.

a. Tính xác suất để trong một ca làm việc có đúng hai máy bị hỏng.

b. Tính xác suất để trong một ca làm việc có ít nhất hai máy bị hỏng.

Bài toán thỏa mãn lược đồ Bernoulli nên

+ Xác suất để trong một ca làm việc có đúng hai máy bị hỏng là:

$$P_5(2) = C_5^2(0,1)^2(0,9)^3 = 0,0729$$

+ Xác suất để trong một ca làm việc có ít nhất hai máy bị hỏng là:

$$P(2;5) = \sum_{k=2}^{5} C_5^k (0,1)^k (0,9)^{5-k}$$

$$= C_5^2(0,1)^2(0,9)^3 + C_5^3(0,1)^3(0,9)^2 + C_5^4(0,1)^4(0,9) + C_5^5(0,1)^5$$

$$= 0,0729 + 0,0081 + 0,00045 + 0,00001$$

$$= 0,08146.$$

- Tính xác suất bằng định nghĩa cổ điển.
- Các công thức tính xác suất:
  - Định lí cộng xác suất
  - Định lí nhân xác suất
  - Hệ quả của ĐL cộng và ĐL nhân xác suất
  - Công thức xác suất đầy đủ và công thức Bayes
  - Công thức Becnuli

Chương 2

ĐẠI LƯỢNG NGẪU NHIÊN VÀ QUY LUẬT PHÂN PHỐI XÁC SUẤT.

1. Đại lượng ngẫu nhiên một chiều

2. Quy luật phân phối xác suất của đại lượng ngẫu nhiên

3. Các tham số đặc trưng của đại lượng ngẫu nhiên

2.1. Đại lượng ngẫu nhiên một chiều

2.1.1. Định nghĩa

Đại lượng ngẫu nhiên (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.

Các ĐLNN thường ký hiệu là: X, Y, Z....,X_1, Y_1, ...Y_n,...Các giá trị của chúng ký hiệu là: x, y, ..., x_1,.., x_n, y_1,..., y_n....

Ví dụ 1: Tung một con xúc xắc.

Gọi X là "số chấm xuất hiện" thì X là đại lượng ngẫu nhiên

Ví dụ 2: Gọi Y là " kích thước của chi tiết do một máy sản xuất ra" thì Y là đại lượng ngẫu nhiên.

2.1.2. Phân loại ĐLNN

ĐLNN được chia làm 2 loại:

+ **ĐLNN rời rạc** nếu các giá trị có thể có của nó lập nên một tập hợp hữu hạn hay đếm được.

+ **ĐLNN liên tục** nếu các giá trị có thể có của nó lấp đầy một khoảng trên trục số.

Ví dụ 1: Tung một con xúc xắc.

Gọi X là "số chấm xuất hiện" thì X là đại lượng ngẫu nhiên rời rạc

Ví dụ 2: Gọi Y là " kích thước của chi tiết do một máy sản xuất ra" thì Y là đại lượng ngẫu nhiên liên tục.

2.2. Quy luật phân phối xác suất của ĐLNN

Sự biểu diễn mối quan hệ giữa **các giá trị** có thể có của ĐLNN và **các xác suất** tương ứng của nó được gọi là quy luật phân phối xác suất của ĐLNN đó.

Có 3 phương pháp để mô tả quy luật phân phối xác suất của ĐLNN là:

+ Bảng phân phối xác suất.

+ Hàm phân phối xác suất.

+ Hàm mật độ xác suất.

a. Bảng phân phối xác suất (chỉ dùng cho ĐLNN rời rạc)

Giả sử ĐLNN rời rạc X nhận các giá trị x_1, x_2,..., x_n với các xác suất tương ứng là p_1, p_2, ... , p_n. Khi đó bảng phân phối xác suất của ĐLNN rời rạc X có dạng

X	x_1	x_2	. . .	x_n
P	p_1	p_2	. . .	p_n

Trong đó $\begin{cases} 0 \le p_i \le 1 \\ \sum_{i=1}^{n} p_i = 1 \end{cases}$ với $\forall i = 1...n$ (2.1)

Ví dụ 1: Tung một con xúc xắc. Gọi X là "số chấm xuất hiện"

Hãy xây dựng quy luật phân phối xác suất của X.

X là ĐLNN rời rạc với các giá trị có thể có là { 1, 2, 3, 4, 5, 6} với các xác suất tương ứng đều bằng 1/6

→ Bảng phân phối xác suất của X là:

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

Ví dụ 2: Trong một lô hàng có 10 sản phẩm trong đó có 6 chính phẩm. Lấy ngẫu nhiên ra 2 sản phẩm.

Xây dựng quy luật phân phối xác suất của số chính phẩm được lấy ra.

Gọi X là: "Số chính phẩm được lấy ra trong 2 sản phẩm" thì X là ĐLNN rời rạc với các giá trị có thể có là { 0, 1, 2} .

$P(X=0)=\frac{C_4^2}{C_{10}^2}=\frac{6}{45}=\frac{2}{15}$ $\quad$ $P(X=1)=\frac{C_6^1C_4^1}{C_{10}^2}=\frac{24}{45}=\frac{8}{15}$

$P(X=2)=\frac{C_6^2}{C_{10}^2}=\frac{15}{45}=\frac{5}{15}$ (Hoặc $P(X=2)=1-\left(\frac{2}{15}+\frac{8}{15}\right)=\frac{5}{15}$)

→ Bảng phân phối xác suất của X là:

X	0	1	2
p	2/15	8/15	5/15

b. Hàm phân phối xác suất

(dùng cho cả ĐLNN rời rạc và ĐLNN liên tục)

Hàm phân phối xác suất của ĐLNN X, ký hiệu là F(x), là **xác suất** để ĐLNN X nhận giá trị nhỏ hơn x, với x là một số thực bất kỳ.

$$\mathbf{F(x) = p(X < x)}$$

Và nếu X là ĐLNN rời rạc thì hàm phân phối xác suất được tính bằng công thức:

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

Ví dụ: Cho ĐLNN rời rạc X có bảng phân phối xác suất sau. Hãy xây dựng hàm phân phối xác suất của X?

X	1	3	4
p	0,1	0,5	0,4

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

\+ Nếu $x\le 1$ thì biến cố (X < x) là biến cố không thể có

→ F(x) = P(X < x) = 0

x 1 3 4

\+ Nếu $1<x\le 3$ thì F(x) = P(X < x) = P(X = 1) = 0,1

1 x 3 4

1 3 x 4

\+ Nếu $3<x\le 4$ thì

→F(x) = P(X < x) = P(X = 1) + P(X = 3) = 0,1 + 0,5 = 0,6

Nếu $x>4$ thì F(x) = P(X < x)

= P(X = 1) + P(X = 3) + P(x = 4)

= 0,1 + 0,5 +0,4 = 1.

1 3 4 x

Vậy hàm phân phối xác suất là:

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

*** Các tính chất của hàm phân phối xác suất:**

Tính chất 1: $0 \le F(x) \le 1$.

Và nếu X là ĐLNN liên tục thì hàm phân phối xs của nó liên tục

Tính chất 2: Xác suất để ĐLNN X nhận giá trị trong [a, b) bằng hiệu số của hàm phân phối xác suất tại hai đầu khoảng đó:

P(a ≤ X < b) = F(b) – F(a)

-> ***Hệ quả 1*** : Xác suất để ĐLNN liên tục X nhận một giá trị xác định bằng không, tức là:

P(X = x) = 0

-> ***Hệ quả 2:*** Nếu X là ĐLNN liên tục thì:

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

Ví dụ : ĐLNN liên tục X có hàm phân phối xác suất:

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

a) Tìm hệ số a và b.

b) Tính xác suất để trong kết quả của phép thử X nhận giá trị trong khoảng (0 ; 1/3).

Vì X là ĐLNN liên tục nên F(x) là hàm số liên tục

→ Ta cần tìm a, b để F(x) là hàm số liên tục tại x = -1 và x = 1/3

$$\lim_{x\to -1^-} F(x) = \lim_{x\to -1^+} F(x) = F(-1) \Leftrightarrow -a+b=0$$

$$\lim_{x\to \frac{1}{3}^-} F(x) = \lim_{x\to \frac{1}{3}^+} F(x) = F(\tfrac{1}{3}) \Leftrightarrow \tfrac{1}{3}a+b=0$$

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

Giải hệ ta được $a=\frac{3}{4}$, $b=\frac{3}{4}$.

Vậy hàm phân phối xác suất là

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

c. ***Hàm mật độ xác suất*** (chỉ có với ĐLNN liên tục)

Hàm mật độ xác suất của ĐLNN liên tục X, ký hiệu là f(x), là **đạo hàm bậc nhất** của hàm phân phối xác suất của ĐLNN đó.

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

*** Các tính chất của hàm mật độ xác suất:**

1. $f(x)\ge 0 \quad \forall x$
2. $P(a<X<b)=\int_a^b f(x)dx$
3. $F(x)=\int_{-\infty}^{x} f(x)dx$
4. $\int_{-\infty}^{+\infty} f(x)dx=1$

__Ví dụ 1:__ Cho ĐLNN liên tục X có hàm mật độ xác suất là:

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

a) Tìm hệ số a.

b) Tìm hàm phân phối xác suất F(x) .

c) Tính xác suất để đại lượng ngẫu nhiên X nhận giá trị trong $(0;\frac{\pi}{4})$.

Theo tính chất của hàm mật độ xác xuất thì f (x) phải thỏa mãn:

$$\begin{cases} f(x)\ge 0 \quad \forall x & (1) \\ \int_{-\infty}^{+\infty} f(x)dx=1 & (2) \end{cases}$$

Điều kiện (1) suy ra $a\ge 0$

Từ (2) ta có

$$1=\int_{-\infty}^{+\infty} f(x)dx=\int_{-\infty}^{-\frac{\pi}{2}} 0.dx+\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} a\cos x dx+\int_{\frac{\pi}{2}}^{+\infty} 0.dx \;=\; 2a \Leftrightarrow a=\frac{1}{2}$$

$$\begin{cases} f(x)\ge 0 \quad \forall x & (1) \\ \int_{-\infty}^{+\infty} f(x)dx=1 & (2) \end{cases} \qquad f(x)=\begin{cases} a\cos x & x\in\left[-\frac{\pi}{2};\frac{\pi}{2}\right] \\ 0 & x\notin\left[-\frac{\pi}{2};\frac{\pi}{2}\right] \end{cases}$$

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

Nếu $x<-\frac{\pi}{2}$ thì $F(x)=\int_{-\infty}^{x} 0.dx=0$

$$f(x)=\begin{cases} \frac{1}{2}\cos x & x\in\left[-\frac{\pi}{2};\frac{\pi}{2}\right] \\ 0 & x\notin\left[-\frac{\pi}{2};\frac{\pi}{2}\right] \end{cases}$$

Nếu $-\frac{\pi}{2}\le x\le\frac{\pi}{2}$ thì

$$F(x)=\int_{-\infty}^{x} f(x)dx=\int_{-\infty}^{-\frac{\pi}{2}} 0.dx+\int_{-\frac{\pi}{2}}^{x}\frac{1}{2}\cos x.dx=\frac{1}{2}\sin x\Big|_{-\frac{\pi}{2}}^{x}=\frac{1}{2}(\sin x+1)$$

Nếu $x>\frac{\pi}{2}$ thì

$$F(x)=\int_{-\infty}^{x} f(x)dx=\int_{-\infty}^{-\frac{\pi}{2}} 0.dx+\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\frac{1}{2}\cos x.dx+\int_{\frac{\pi}{2}}^{x} 0.dx=\frac{1}{2}\sin x\Big|_{-\frac{\pi}{2}}^{\frac{\pi}{2}}=1$$

Vậy hàm phân phối xác suất của X là

$$F(x)=\begin{cases} 0 & x<-\frac{\pi}{2} \\ \frac{1}{2}(\sin x+1) & -\frac{\pi}{2}\le x\le\frac{\pi}{2} \\ 1 & x>\frac{\pi}{2} \end{cases}$$

c) Tính $P(0<X<\frac{\pi}{4})$

$$P(0<X<\tfrac{\pi}{4})=F(\tfrac{\pi}{4})-F(0)=\frac{1}{2}(\sin\frac{\pi}{4}+1)-\frac{1}{2}(\sin 0+1)=\frac{\sqrt{2}}{4}$$

Hoặc cũng có thể tính như sau:

$$P(0<X<\tfrac{\pi}{4})=\int_0^{\frac{\pi}{4}} f(x)dx=\int_0^{\frac{\pi}{4}}\frac{1}{2}\cos x dx=\frac{1}{2}\sin x\Big|_0^{\frac{\pi}{4}}=\frac{\sqrt{2}}{4}$$

2.3. Các tham số đặc trưng của ĐLNN

2.3.1. Kỳ vọng toán (Kỳ vọng)

Kỳ vọng của ĐLNN X, ký hiệu là E(X), là <u>một số</u> được xác định theo X như sau:

+ Nếu X là ĐLNN rời rạc có bảng phân phối xác suất

X	x_1	x_2	.....	x_n
P	p_1	p_2	.....	p_n

$$\rightarrow E(X)=\sum_{i=1}^{n} x_i p_i$$

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

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

Ví dụ 1: Xác định kỳ vọng của ĐLNN rời rạc X có bảng phân phối xác suất sau:

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

Theo định nghĩa kỳ vọng của ĐLNN rời rạc ta có:

$$E(X)=\sum_{i=1}^{3} x_i p_i$$

$$=1.0,2+2.0,5+3.0,3=2,1$$

Ví dụ 2: Cho ĐLNN liên tục X có hàm phân phối xác suất là:

$$F(x)=\begin{cases} 0 & x\le 2 \\ \frac{1}{4}(x-2)^2 & 2<x\le 4 \\ 1 & x>4 \end{cases}$$

Tìm kỳ vọng của X.

Theo định nghĩa kỳ vọng của ĐLNN liên tục ta có: $E(X)=\int_{-\infty}^{+\infty} x.f(x)dx$

Trong đó:

$$f(x)=F'(x)\Rightarrow f(x)=\begin{cases} \frac{1}{2}(x-2) & x\in[2;4] \\ 0 & x\notin[2;4] \end{cases}$$

$$\Rightarrow E(X)=\int_{-\infty}^{2} 0.dx+\int_{2}^{4}\frac{x}{2}(x-2)dx+\int_{4}^{+\infty} 0.dx=\frac{10}{3}$$

- **<u>Chú ý</u>**: Kỳ vọng phản ánh <u>giá trị trung tâm</u> của phân phối xác suất của ĐLNN.

<u>Ví dụ:</u>

Thời gian để sản xuất ra một sản phẩm A là đại lượng ngẫu nhiên liên tục X có hàm mật độ xác suất là (đơn vị: phút)

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

Tìm **<u>thời gian trung bình</u>** sản xuất một sản phẩm

2.3.2. Phương sai – Độ lệch chuẩn

+ Phương sai của ĐLNN X, ký hiệu là D(X), là kỳ vọng toán của bình phương độ lệch của ĐLNN với kỳ vọng của nó: **$D(X)=E[X-E(X)]^2$**

+ Độ lệch tiêu chuẩn của đại lượng ngẫu nhiên X là: $\sigma(X)=\sqrt{D(X)}$

* **<u>Chú ý:</u>** trong thực tế, ta dùng công thức dạng tương đương để việc tính toán dễ dàng hơn $D(X)=E(X^2)-[E(X)]^2$

+ Nếu X là ĐLNN rời rạc thì: $D(X)=\sum_{i=1}^{n} x_i^2 p_i-\left(\sum_{i=1}^{n} x_i p_i\right)^2$

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

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

Ví dụ 1: Cho ĐLNN rời rạc X có bảng phân phối xác suất là:

X	0	1	2	3
P	0,064	0,288	0,432	0,216

Hãy tính D(X).

Áp dụng công thức $D(X)=E(X^2)-[E(X)]^2$

$$E(X)=\sum_{i=1}^{4} x_i p_i=0.0,064+1.0,288+2.0,432+3.0,216=1,8$$

$$E(X^2)=\sum_{i=1}^{4} x_i^2 p_i=0^2.0,064+1^2.0,288+2^2.0,432+3^2.0,216=3,96$$

Vậy: $D(X)=E(X^2)-[E(X)]^2=3,96-(1,8)^2=0,72.$

Ví dụ 2: Cho ĐLNN liên tục X có hàm mật độ xác suất là:

$$f(x)=\begin{cases}3x^2 & x\in(0;1)\\ 0 & x\notin(0;1)\end{cases}$$

Tìm phương sai D(X) và độ lệch chuẩn của X.

Áp dụng công thức $D(X)=E(X^2)-\left[E(X)\right]^2$

$$E(X)=\int_{-\infty}^{+\infty}x.f(x)dx=\int_0^1 3x^3dx=\frac{3}{4}x^4\Big|_0^1=\frac{3}{4}$$

$$E(X^2)=\int_{-\infty}^{+\infty}x^2.f(x)dx=\int_0^1 3x^4dx=\frac{3}{5}x^5\Big|_0^1=\frac{3}{5}$$

$$\Rightarrow D(X)=E(X^2)-\left[E(X)\right]^2=\frac{3}{5}-\left(\frac{3}{4}\right)^2=\frac{3}{80}$$

$$\sigma(X)=\sqrt{D(X)}=\sqrt{\frac{3}{80}}\approx 0,194$$

Chương 2

ĐLNN VÀ QUY LUẬT PHÂN PHỐI XÁC SUẤT.

1. Phân biệt ĐLNN rời rạc và ĐLNN liên tục
2. Quy luật phân phối xác suất của đại lượng ngẫu nhiên
   - + Bảng phân phối xác suất
   - + Hàm phân phối xác suất
   - + Hàm mật độ xác suất
3. Các tham số đặc trưng của đại lượng ngẫu nhiên
   - + Kỳ vọng
   - + Phương sai, độ lệch chuẩn

Chương 3
LÝ THUYẾT MẪU

3.1. Khái niệm về phương pháp mẫu, mẫu ngẫu nhiên

3.2. Một số đặc trưng của mẫu

3.3. Thực hành tính các giá trị của đặc trưng mẫu thông qua các mẫu cụ thể

3.1. Khái niệm về phương pháp mẫu, mẫu ngẫu nhiên

3.1.1. Khái niệm về PP mẫu:

Phương pháp chọn mẫu là phương pháp (PP) từ tập hợp cần nghiên cứu chọn ra một số phần tử, phân tích các phần tử này và dựa vào đó để kết luận về dấu hiệu cần nghiên cứu của tổng thể.

Tập hợp gồm n phần tử lấy ra từ tổng thể gọi là mẫu, n được gọi là kích thước mẫu (hay cỡ mẫu).

Ví dụ: Để xác định chiều cao, cân nặng trung bình của trẻ 5-6 tuổi ở một trường mầm non , người ta tiến hành đo cân nặng, chiều cao của 50 trẻ bất kì

→ Kích thước mẫu n = 50

3.1.2. Mẫu ngẫu nhiên:

Mẫu ngẫu nhiên cỡ n là tập hợp của n biến ngẫu nhiên độc lập $X_1, X_2,...,X_n$ được thành lập từ ĐLNN X trong tổng thể nghiên cứu và có cùng quy luật phân phối xác suất với X.

Ký hiệu mẫu ngẫu nhiên cỡ n : $W_x = (X_1, X_2,..,X_n)$

Thực hiện phép thử với mẫu ngẫu nhiên W_x , giả sử khi đó X_i nhận giá trị x_i (i = 1..n) . Khi đó $w = (x_1, x_2,..,x_n)$ gọi là mẫu cụ thể.

3.1.3. Các phương pháp chọn mẫu:

a) Mẫu ngẫu nhiên hoàn lại

b) Mẫu ngẫu nhiên không hoàn lại

c) Mẫu phân tổ

d) Mẫu" điển hình"

e) Mẫu nhiều cấp

3.2 Một số đặc trưng của mẫu

Giả sử từ đại lượng ngẫu nhiên gốc X trong tổng thể, lập mẫu ngẫu nhiên cỡ n: $W_x = (X_1, X_2, ..., X_n)$, trong đó X_i có tần số là n_i.

1. Trung bình mẫu

Trung bình mẫu là một thống kê, kí hiệu là $\overline{X}$, được xác định bởi công thức: $$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} n_i X_i$$

Trong đó tần số của X_i là n_i (i = 1...n)

2. Phương sai mẫu, độ lệch tiêu chuẩn

Phương sai mẫu ngẫu nhiên là một thống kê, được ký hiệu là S^{*2} và xác định bởi công thức:

$$S^{*2} = \frac{1}{n}\sum_{i=1}^{k} n_i (X_i - \overline{X})^2$$

Phương sai mẫu hiệu chỉnh có dạng:

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

Độ lệch tiêu chuẩn là căn bậc hai của phương sai mẫu.

$$S^* = \sqrt{S^{*2}}$$

Độ lệch tiêu chuẩn hiệu chỉnh là căn bậc hai của phương sai mẫu hiệu chỉnh.

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

3. Tần suất mẫu

Giả sử từ tổng thể, lấy ra mẫu ngẫu nhiên cỡ n, trong đó có m phần tử mang dấu hiệu nghiên cứu.

Khi đó tần suất mẫu là một thống kê được xác định bởi công thức

$$f = \frac{m}{n}$$

3.3. Ví dụ thực hành tính các giá trị của đặc trưng mẫu thông qua các mẫu cụ thể:

Giả sử cho mẫu cụ thể $w_X = (x_1, x_2, ..., x_n)$.

Khi đó ta có các đặc trưng mẫu được tính bởi công thức:

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

$$S^{*2} = \frac{1}{n}\sum_{i=1}^{n} n_i x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} n_i x_i\right)^2$$

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

$$S^* = \sqrt{S^{*2}} \; ; \; S = \sqrt{S^2}$$

Ví dụ 1: Đo độ dài của 30 chi tiết máy được chọn ngẫu nhiên của một loại sản phẩm ta thu được bảng số liệu sau:

Độ dài (x_i) (cm)	39	40	41	42	43	44
Số chi tiết (n_i)	4	5	9	7	4	1

Tính các đặc trưng của mẫu $\overline{X}, S^{*2}, S^2, S$.

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{k} n_i X_i$$

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

Độ dài (x_i) (cm)	39	40	41	42	43	44
Số chi tiết (n_i)	4	5	9	7	4	1

Từ bảng tính, ta có:

$$\overline{X} = \frac{1}{30}\sum_{i=1}^{6} n_i x_i = \frac{1}{30}.1235 \approx 41{,}17$$

$$S^{*2} = \frac{1}{30}\sum_{i=1}^{6} n_i x_i^2 - \left(\frac{1}{30}\sum_{i=1}^{6} n_i x_i\right)^2$$

$$= \frac{1}{30}.50893 - (41{,}17)^2 \approx 1{,}46.$$

$$S^2 = \frac{n}{n-1} S^{*2} = \frac{30}{29}.1{,}46 \approx 1{,}5$$

$$S = \sqrt{S^2} \approx 1{,}23.$$

x_i	n_i	$n_i x_i$	$n_i x_i^2$
39	4	156	6.084
40	5	200	8.000
41	9	369	15.129
42	7	294	12.348
43	4	172	7.396
44	1	44	1.936
$\sum$	30	1.235	50.893

*** Chú ý 1:**

Khi các **x_i gần nhau** và **n_i khá lớn** thì ta nên AD phép đổi biến để thuận tiện cho việc tính toán

Chọn $x_0 = x_i$ (với x_i có n_i lớn nhất). Khi đó:

$$\overline{X} = x_0 + \frac{1}{n}\sum_{i=1}^{k} n_i (x_i - x_0)$$

$$S^{*2} = \frac{1}{n}\sum_{i=1}^{k} n_i (x_i - x_0)^2 - \left[\frac{1}{n}\sum_{i=1}^{k} n_i (x_i - x_0)\right]^2$$

$$S^* = \sqrt{S^{*2}}$$

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

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

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{k} n_i X_i$$

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

Ví dụ 2: Đo đường kính của 100 chi tiết máy ta thu được bảng số liệu sau:

Độ dài đường kính (X – mm)	80	82	84	86	88	90
Số chi tiết (n_i)	7	15	30	20	15	13

Tính độ dài trung bình của đường kính và độ lệch tiêu chuẩn hiệu chỉnh. (Tính $\overline{X}, S$)

x_i	n_i	$x_i - x_0$	$n_i(x_i - x_0)$	$n_i(x_i - x_0)^2$
80	7	– 4	– 28	112
82	15	– 2	– 30	60
84	30	0	0	0
86	20	2	40	80
88	15	4	60	240
90	13	6	78	468
$\sum$	100		120	960

$$S^2 = \frac{n}{n-1}.S^{*2} = \frac{100}{99}.8,16 \approx 8,24$$

$$\Rightarrow S = \sqrt{S^2} \approx 2,87.$$

Chọn $x_0 = 84$

$$\overline{X} = x_0 + \frac{1}{n}\sum_{i=1}^{k} n_i(x_i - x_0) = 84 + \frac{1}{100}.120 = 85,2$$

$$S^{*2} = \frac{1}{n}\sum_{i=1}^{k} n_i(x_i - x_0)^2 - \left[\frac{1}{n}\sum_{i=1}^{k} n_i(x_i - x_0)\right]^2$$

$$= \frac{1}{100}.960 - \left(\frac{1}{100}.120\right)^2 = 9,6 - 1,44 = 8,16.$$

- **Chú ý 2:**

Nếu X là ĐLNN liên tục, ta sẽ phân chia các giá trị có thể có của X thành từng khoảng bằng nhau có độ dài là h.

+ Gọi x_i^* là trung điểm mỗi khoảng

+ Chọn $x_0 = x_i^*$ (x_i^* ứng với n_i lớn nhất)

+ Đặt $y_i = \dfrac{x_i^* - x_0}{h}$

Khi đó:

$$\overline{X} = x_0 + \frac{h}{n}\sum_{i=1}^{k} n_i y_i$$

$$S^{*2} = \frac{h^2}{n}\sum_{i=1}^{k} n_i y_i^2 - \left(\frac{h}{n}\sum_{i=1}^{k} n_i y_i\right)^2$$

Ví dụ 3 : Điều tra doanh số hàng tháng của 100 hộ kinh doanh một loại hàng hoá, người ta thu được bảng số liệu sau:

Doanh số (X– triệu đồng)	40 – 42	42 – 44	44 – 46	46 – 48	48 – 50
Số hộ kinh doanh (n_i)	10	15	35	30	10

Tính doanh số trung bình và độ lệch tiêu chuẩn hiệu chỉnh

x_i	n_i	x_i^*	y_i	$n_i y_i$	$n_i y_i^2$
40 – 42	10	41	– 2	– 20	40
42 – 44	15	43	– 1	– 15	15
44 – 46	35	45	0	0	0
46 – 48	30	47	1	30	30
48 – 50	10	49	2	20	40
$\sum$	100			15	125

Doanh số (X– triệu đồng)	40 – 42	42 – 44	44 – 46	46 – 48	48 – 50
Số hộ kinh doanh (n_i)	10	15	35	30	10

Chọn $x_0 = 45$;

Độ dài mỗi khoảng h = 2; n = 100

Khi đó:

$$\overline{X} = x_0 + \frac{h}{n}\sum_{i=1}^{k} n_i y_i = 45 + \frac{2}{100}.15 = 45 + 0,3 = 45,3.$$

$$S^{*2} = \frac{h^2}{n}\sum_{i=1}^{k} n_i y_i^2 - \left(\frac{h}{n}\sum_{i=1}^{k} n_i y_i\right)^2 = \frac{4}{100}.125 - \left(\frac{2}{100}.15\right)^2 = 5 - 0,09 = 4,91.$$

$$\Rightarrow S^2 = \frac{100}{99}.4,91 \approx 4,96. \quad \Rightarrow S = \sqrt{S^2} \approx 2,23.$$

Chương 4

ƯỚC LƯỢNG THAM SỐ CỦA ĐẠI LƯỢNG NGẪU NHIÊN

4.1. Phương pháp ước lượng điểm

4.2. Phương pháp ước lượng khoảng

Bài toán:

Cần phải nghiên cứu dấu hiệu γ nào đó của tổng thể.

Giả sử ĐLNN X là mô hình hóa của dấu hiệu γ. Giả thiết rằng bằng phân tích lý thuyết đã xác định dạng phân phối xác suất của nó, nhưng chưa biết tham số θ nào đó của nó.

Vấn đề đặt ra là cần ước lượng (xác định một cách gần đúng) tham số θ đó.

4.1. Phương pháp ước lượng điểm

Giả sử ĐLNN X có tham số θ chưa biết

+ Từ X lập mẫu ngẫu nhiên cỡ n là $W_x = (X_1, X_2, ..., X_n)$.

+ Chọn thống kê $G = G(X_1, X_2, ..., X_n)$ thay cho θ.

+ Tiến hành lập mẫu cụ thể $w_x = (x_1, x_2, ..., x_n)$.

+ Tính giá trị cụ thể của G ứng với mẫu nói trên, giả sử là $g = g(x_1, x_2, ..., x_n)$.

→ Khi đó ước lượng điểm của θ là $g = g(x_1, x_2, ..., x_n)$ vừa tìm được.

4.2. Phương pháp ước lượng khoảng

4.2.1. Phương pháp ước lượng bằng khoảng tin cậy

<u>Bài toán</u>: Ước lượng tham số θ của ĐLNN X

+ Lập mẫu ngẫu nhiên cỡ n : $W_x = (X_1, X_2, .., X_n)$

+ Chọn thống kê $G = G(X, \theta)$.

+ Xác định khoảng $(G_1; G_2)$ của G sao cho với một xác suất cho trước thì $\theta \in (G_1; G_2)$.

→ Khi đó khoảng $(G_1; G_2)$ được gọi là <u>khoảng tin cậy</u> của θ nếu với xác suất α khá bé, ta được $P(G_1 < \theta < G_2) = 1 - \alpha$

$\gamma = 1 - \alpha$ gọi là <u>độ tin cậy</u> của ước lượng.

Độ dài $l = G_2 - G_1$ được gọi là <u>độ dài của khoảng tin cậy</u>.

4.2. Phương pháp ước lượng khoảng

4.2.1. Phương pháp ước lượng bằng khoảng tin cậy

<u>Bài toán</u>: Ước lượng tham số θ của ĐLNN X

+ Lập mẫu ngẫu nhiên cỡ n : $W_x = (X_1, X_2, .., X_n)$

+ Chọn thống kê $G = G(X, \theta)$.

+ Xác định khoảng $(G_1; G_2)$ của G sao cho với một xác suất cho trước thì $\theta \in (G_1; G_2)$.

→ Khi đó khoảng $(G_1; G_2)$ được gọi là <u>khoảng tin cậy</u> của θ nếu với xác suất α khá bé, ta được $P(G_1 < \theta < G_2) = 1 - \alpha$

$\gamma = 1 - \alpha$ gọi là <u>độ tin cậy</u> của ước lượng.

Độ dài $l = G_2 - G_1$ được gọi là <u>độ dài của khoảng tin cậy</u>.

4.2.2. Ước lượng khoảng tin cậy cho kỳ vọng của ĐLNN X có phân phối chuẩn

Giả sử trong tổng thể, ĐLNN X có phân phối chuẩn $N(a,\sigma^2)$ với kỳ vọng $E(X) = a$ chưa biết.

Tức là: hàm mật độ xác suất của X có dạng

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

Trong đó $a = E(X)$; $\sigma^2 = D(X)$.

Ta cần ước lượng $E(X) = a$.

Từ X, lập mẫu ngẫu nhiên cỡ n : $W_x = (X_1, X_2, .., X_n)$

A- Trường hợp 1: Đã biết phương sai $D(X) = \sigma^2$

+ Chọn lập thống kê $G = \frac{\overline{X}-a}{\sigma}\sqrt{n}$ với $\overline{X}$ là trung bình mẫu

+ Khi đó khoảng tin cậy đối xứng của E(X) = a có dạng

$$\left(\overline{X}-t_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}};\ \overline{X}+t_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$$

$t_{\frac{\alpha}{2}}$ được xác định từ hệ thức $\Phi\left(t_{\frac{\alpha}{2}}\right)=\frac{1-\alpha}{2}$

+ Khoảng tin cậy bên phải là: $\left(\overline{X}-t_{\alpha}\frac{\sigma}{\sqrt{n}};+\infty\right)$ với $\Phi(t_\alpha)=\frac{1-2\alpha}{2}$.

+ Khoảng tin cậy bên trái là: $\left(-\infty;\overline{X}+t_{\alpha}\frac{\sigma}{\sqrt{n}}\right)$

- Nhận xét: Từ khoảng tin cậy đối xứng của E(X) = a có dạng

$$\left(\overline{X}-t_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}};\ \overline{X}+t_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$$

+ Đặt : $\varepsilon = t_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$

→ ε gọi là độ chính xác của ước lượng

+ Độ dài của khoảng tin cậy $l = 2\varepsilon = t_{\frac{\alpha}{2}}\frac{2\sigma}{\sqrt{n}}$

→ Công thức xác định kích thước mẫu tối thiểu sao cho với độ tin cậy $\gamma = 1-\alpha$ cho trước, độ dài khoảng tin cậy không vượt quá giá trị l_0 cho trước :

$$t_{\frac{\alpha}{2}}\frac{2\sigma}{\sqrt{n}} \le l_0 \Rightarrow n \ge \left(t_{\frac{\alpha}{2}}\frac{2\sigma}{l_0}\right)^2 = t_{\frac{\alpha}{2}}^2\frac{\sigma^2}{\varepsilon_0^2};\ n\in Z^+$$

Ví dụ 1 : Chiều dài của một loại sản phẩm (X) là ĐLNN có phân phối theo quy luật chuẩn với độ lệch chuẩn là 0,38. Đo thử 30 sản phẩm loại này, ta thu được kết quả sau:

Chiều dài (cm)	36	38	40	42
Số sản phẩm tương ứng	6	11	8	5

a) Với độ tin cậy 0,95 hãy ước lượng chiều dài trung bình của loại sản phẩm nói trên bằng khoảng tin cậy đối xứng.

b) Nếu yêu cầu độ chính xác của ước lượng là 0,05 và giữ nguyên độ tin cậy thì phải điều tra một mẫu kích thước bao nhiêu?

Ta có $\sigma(X)$ = 0,38, E(X) = a chưa biết. Cần ước lượng E(X).

Khoảng tin cậy đối xứng là: $\left(\overline{X}-t_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}};\ \overline{X}+t_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$

Trong đó: n = 30 ; σ = 0,38

Độ tin cậy $\gamma = 1-\alpha = 0{,}95 \Rightarrow \Phi\left(t_{\frac{\alpha}{2}}\right)=\frac{1-\alpha}{2}=\frac{0{,}95}{2}=0{,}475$

Bảng 2: HÀM LAPLACE $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_0^x e^{\frac{-t^2}{2}}dt$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0039	0079	0119	0159	0199	0239	0279	0318	0358
0,1	03983	0438	0477	0517	0556	0596	0635	0674	0714	0753
0,2	07926	0831	0870	0909	0948	0987	1025	1064	1102	1140
......	......	......	......	......	......	......	......	......	......	......
1,6	44520	4463	4473	4481	4495	4505	4515	4525	4535	4544
1,7	45543	4563	4572	4581	4590	4599	4608	4616	4624	4632
1.8	46407	4648	4656	4663	4671	4678	4685	4692	4699	4706
1,9	47128	4719	4725	4732	4738	4744	4750	4755	4761	4767
2,0	47725	4777	4783	4788	4793	4798	4803	4807	4812	4816
2,1	48214	4825	4830	4834	4838	4842	4846	4850	4853	4857

$\Rightarrow t_{\frac{\alpha}{2}} = 1{,}96$

Trung bình mẫu:

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{4}n_i.X_i=\frac{1}{30}(6.36+11.38+8.40+5.42)=38{,}8$$

Vậy khoảng tin cậy đối xứng cần tìm là:

$$\left(\overline{X}-t_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}};\ \overline{X}+t_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)=\left(38{,}8-1{,}96.\frac{0{,}38}{\sqrt{30}};\ 38{,}8+1{,}96.\frac{0{,}38}{\sqrt{30}}\right)$$
$$=(38{,}66\ ;\ 38{,}94)$$

Chiều dài (cm)	36	38	40	42
Số sản phẩm tương ứng	6	11	8	5

Ví dụ 1 : Chiều dài của một loại sản phẩm (X) là ĐLNN có phân phối theo quy luật chuẩn với độ lệch chuẩn là 0,38. Đo thử 30 sản phẩm loại này, ta thu được kết quả sau:

Chiều dài (cm)	36	38	40	42
Số sản phẩm tương ứng	6	11	8	5

b) Nếu yêu cầu độ chính xác của ước lượng là 0,05 và giữ nguyên độ tin cậy thì phải điều tra một mẫu kích thước bao nhiêu?

Áp dụng công thức: $n \ge \left(t_{\frac{\alpha}{2}}\frac{2\sigma}{l_0}\right)^2 = t_{\frac{\alpha}{2}}^2\frac{\sigma^2}{\varepsilon_0^2}$

Trong đó : $t_{\frac{\alpha}{2}} = 1{,}96; \sigma = 0{,}38; \varepsilon_0 = 0{,}05 \Rightarrow n \ge 221{,}9$

Mà n là số nguyên dương → n = 222.

→ Vậy cần phải điều tra mẫu kích thước là 222 để đạt được độ chính xác 0,05 đã đề ra.

4.2. Phương pháp ước lượng khoảng

4.2.2. Ước lượng khoảng tin cậy cho kỳ vọng của ĐLNN X có phân phối chuẩn

A- Trường hợp đã biết phương sai $D(X) = \sigma^2$

+ Khoảng tin cậy đối xứng của E(X) = a có dạng

$$\left(\overline{X} - t_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}};\ \overline{X} + t_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$$

$t_{\frac{\alpha}{2}}$ được xác định từ hệ thức $\Phi\left(t_{\frac{\alpha}{2}}\right) = \frac{1-\alpha}{2}$

+ Khoảng tin cậy bên phải là: $\left(\overline{X} - t_\alpha\frac{\sigma}{\sqrt{n}};+\infty\right)$ với $\Phi(t_\alpha) = \frac{1-2\alpha}{2}$.

+ Khoảng tin cậy bên trái là: $\left(-\infty;\overline{X} + t_\alpha\frac{\sigma}{\sqrt{n}}\right)$

Ví dụ 2: Với độ tin cậy 87,9%, hãy ước lượng lượng xăng hao phí tối đa cho 1 oto chạy từ A đến B. Biết rằng nếu chạy thử 30 lần trên đoạn đường này, người ta ghi nhận được lượng xăng hao phí như sau:

Xăng hao phí (lít)	9,6-9,8	9,8-10	10-10,2	10,2-10,4	10,4-10,6
Số lần chạy thử	3	5	10	8	4

Giả thiết lượng xăng hao phí là ĐLNN tuân theo quy luật phân phối chuẩn, có độ lệch chuẩn là 0,21.

Ycbt ↔ Xác định khoảng tin cậy bên trái.

Khi đó lượng xăng hao phí tối đa là : $\overline{X} + t_\alpha\frac{\sigma}{\sqrt{n}}$

Trong đó: n = 30; $\sigma = 0,21$

Độ tin cậy $\gamma = 1 - \alpha = 87,9\% = 0,879 \Rightarrow \alpha = 0,121$

$\Rightarrow \Phi(t_\alpha) = \frac{1-2\alpha}{2} = 0,379$

Bảng 2: HÀM LAPLACE $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_0^x e^{\frac{-t^2}{2}}dt$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0039	0079	0119	0159	0199	0239	0279	0318	0358
0,1	03983	0438	0477	0517	0556	0596	0635	0674	0714	0753
0,2	07926	0831	0870	0909	0948	0987	1025	1064	1102	1140
0,3	11791	1217	1255	1293	1330	1368	1405	1443	1480	1517
0,4	15542	1591	1627	1664	1700	1736	1772	1808	1843	1879
0,5	19146	1949	1984	2019	2019	2088	2122	2156	2190	2224
0,6	22575	2290	2323	2356	2389	2421	2453	2485	2517	2549
0,7	25804	2611	2642	2673	2703	2733	2763	2793	2823	2852
0,8	28814	2910	2938	2967	2995	3023	3051	3078	3105	3132
0,9	31594	3185	3212	3288	3263	3289	3314	3339	3364	3389
1,0	34134	3437	3461	3485	3508	3531	3554	3576	3599	3621
1,1	36433	3665	3686	3707	3728	3749	3769	3790	3810	3829
1,2	38493	3868	3887	3906	3925	3943	3961	3979	3997	4014

$\Phi(t_\alpha) = 0,379 \Rightarrow t_\alpha = 1,17$

Lập bảng tính trung bình mẫu:

x_i	n_i	x_i^*	y_i	$n_i y_i$
9,6-9,8	3	9,7	-2	-6
9,8-10	5	9,9	-1	-5
10-10,2	10	10,1	0	0
10,2-10,4	8	10,3	1	8
10,4-10,6	4	10,5	2	8
				$\sum = 5$

Độ dài mỗi khoảng h = 0,2;

Kích thước mẫu n = 30

Chọn x_0 = 10,1

$$\overline{X} = x_0 + \frac{h}{n}\sum_{i=1}^{4} n_i.y_i = 10,1 + \frac{0,2}{30}.5 = 10,133$$

Vậy lượng xăng hao phí tối đa là :

$$\overline{X} + t_\alpha\frac{\sigma}{\sqrt{n}} = 10,133 + 1,17.\frac{0,21}{\sqrt{30}} = 10,178\ (l)$$

A- TH1: đã biết phương sai $D(X) = \sigma^2$

B - TH2: chưa biết phương sai D(X) và mẫu cỡ n < 30

Khoảng tin cậy đối xứng của E(X) = a là:

$$\overline{X} - t_{\frac{\alpha}{2}}^{(n-1)}\frac{S}{\sqrt{n}} < a < \overline{X} + t_{\frac{\alpha}{2}}^{(n-1)}\frac{S}{\sqrt{n}}$$

Trong đó:

$\overline{X}, S$ lần lượt là trung bình mẫu và độ lệch tiêu chuẩn hiệu chỉnh

$t_{\frac{\alpha}{2}}^{(n-1)}$ được tra từ bảng phân vị Student

$$\left(\overline{X} - t_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}};\ \overline{X} + t_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$$

→ Các khoảng tin cậy một phía của E(X) = a là:

$\overline{X} - t_\alpha^{(n-1)}\frac{S}{\sqrt{n}} < a < +\infty$ -- khoảng tin cậy bên phải

$-\infty < a < \overline{X} + t_\alpha^{(n-1)}\frac{S}{\sqrt{n}}$ -- khoảng tin cậy bên trái

$$\overline{X} - t_{\frac{\alpha}{2}}^{(n-1)}\frac{S}{\sqrt{n}} < a < \overline{X} + t_{\frac{\alpha}{2}}^{(n-1)}\frac{S}{\sqrt{n}}$$

Ví dụ: Đo đường kính của 25 chi tiết do một máy tiện sản xuất, người ta thu được bảng số liệu sau

Độ dài (mm)	48 - 50	50 - 52	52 - 54	54 - 56
Số chi tiết	4	8	7	6

Với độ tin cậy 95%, hãy ước lượng độ dài tối thiểu của đường kính chi tiết máy, biết đường kính chi tiết máy là đại lượng ngẫu nhiên X có phân phối chuẩn.

Ycbt ↔ Xác định khoảng tin cậy bên phải.

Khi đó độ dài tối thiểu của đường kính chi tiết máy là :

$$\overline{X} - t_{\alpha}^{(n-1)} \frac{S}{\sqrt{n}}$$

Trong đó: n = 25.

Độ tin cậy $\gamma = 1 - \alpha = 95\% = 0,95 \Rightarrow \alpha = 0,05$

$\Rightarrow t_{\alpha}^{(n-1)} = t_{0,05}^{(24)} = 1,711$

$$\overline{X} - t_{\alpha}^{(n-1)} \frac{S}{\sqrt{n}}$$

Bảng 3: PHÂN VỊ STUDENT $P\{X > t_{\alpha}^{(n)}\} = \alpha$ với $X \sim t(n)$

n \ α	0,1	0,05	0,025	0,01	0,005	0,0005
1	3,078	6,314	12,706	31,820	63,626	363,6
2	1,886	2,920	4,303	6,965	9,925	31,600
3	1,638	2,353	3,182	4,541	5,841	12,922
4	1,533	2,132	2,776	3,747	4,604	8,610
5	1,476	2,015	2,571	3,365	4,032	6,869
....	....	....	....	....	....	....
24	1,318	1,711	2,064	2,492	2,797	3,745
25	1,316	1,708	2,060	2,485	2,787	3,725

$\Rightarrow t_{0,005}^{(25)} = 2,787$

Lập bảng tính $\overline{X}, S$

Chọn $x_0 = 51$,

h = 2

x_i	n_i	x'_i	y_i	$n_i y_i$	$n_i y_i^2$
48 – 50	4	49	-1	-4	4
50 – 52	8	51	0	0	0
52 – 54	7	53	1	7	7
54 – 56	6	55	2	12	24
	n = 25			$\sum = 15$	$\sum = 35$

$$\overline{X} = x_0 + \frac{h}{n}\sum_{i=1}^{4} n_i y_i = 51 + \frac{2}{25}.15 = 52,2$$

$$S'^2 = \frac{h^2}{n}\sum n_i y_i^2 - \left(\frac{h}{n}\sum n_i y_i\right)^2 = \frac{4}{25}.35 - (\frac{2}{25}.15)^2 = 4,16$$

$$\Rightarrow S^2 = \frac{n}{n-1} S'^2 = \frac{25}{24}.4,16 = 4,33 \Rightarrow S = \sqrt{S^2} \approx 2,08$$

Vậy độ dài tối thiểu của đường kính chi tiết máy là

$$52,2 - 1,711.\frac{2,08}{\sqrt{25}} = 51,49(mm)$$

C – Trường hợp chưa biết phương sai D(X) và mẫu cỡ n ≥ 30

+ Khoảng tin cậy đối xứng của E(X) = a là:

$$\overline{X} - t_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}} < a < \overline{X} + t_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}$$

Trong đó:

$\overline{X}, S$ lần lượt là trung bình mẫu và độ lệch tiêu chuẩn hiệu chỉnh

$t_{\frac{\alpha}{2}}$ được xác định từ hệ thức $\Phi\left(t_{\frac{\alpha}{2}}\right) = \frac{1-\alpha}{2}$

Đã biết phương sai D(X) = σ^2: $\overline{X} - t_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < a < \overline{X} + t_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$

Chưa biết phương sai D(X) = $\sigma^2, n < 30$: $\overline{X} - t_{\frac{\alpha}{2}}^{(n-1)} \frac{S}{\sqrt{n}} < a < \overline{X} + t_{\frac{\alpha}{2}}^{(n-1)} \frac{S}{\sqrt{n}}$

→ Các khoảng tin cậy một phía của E(X) = a là:

$\overline{X} - t_{\alpha} \frac{S}{\sqrt{n}} < a < +\infty$ -- khoảng tin cậy bên phải

$-\infty < a < \overline{X} + t_{\alpha} \frac{S}{\sqrt{n}}$ -- khoảng tin cậy bên trái

Trong đó t_α được xác định từ hệ thức: $\Phi(t_\alpha) = \frac{1-2\alpha}{2}$

Ví dụ: Hao phí nguyên liệu cho một đơn vị sản phẩm là một ĐLNN tuân theo quy luật phân phối chuẩn.

Người ta sản xuất thử 50 sản phẩm và thu được bảng số liệu sau:

Mức hao phí ng.liệu (gam)	**19,5 – 20,0**	**20,0 – 20,5**	**20,5 – 21,0**	**21,0 – 21,5**
Số sp tương ứng	**15**	**20**	**10**	**5**

Với độ tin cậy 96% hãy ước lượng **mức hao phí nguyên liệu trung bình** cho một đơn vị sản phẩm bằng khoảng tin cậy đối xứng.

Gọi X là "mức hao phí nguyên liệu cho một đơn vị sản phẩm" thì X là ĐLNN có phân phối chuẩn.

Khoảng tin cậy đối xứng cần tìm là: $\overline{X} - t_{\frac{\alpha}{2}}\dfrac{S}{\sqrt{n}} < a < \overline{X} + t_{\frac{\alpha}{2}}\dfrac{S}{\sqrt{n}}$

Trong đó: Kích thước mẫu n = 50

Độ tin cậy $\gamma = 1 - \alpha = 96\% = 0,96 \Rightarrow \Phi\left(t_{\frac{\alpha}{2}}\right) = \dfrac{1-\alpha}{2} = 0,48$

$\Rightarrow t_{\frac{\alpha}{2}} = 2,06$

Bảng 2: HÀM LAPLACE $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_0^x e^{\frac{-t^2}{2}}dt$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0039	0079	0119	0159	0199	0239	0279	0318	0358
0,1	03983	0438	0477	0517	0556	0596	0635	0674	0714	0753
0,2	07926	0831	0870	0909	0948	0987	1025	1064	1102	1140
......	......	......	......	......	......	......	......	......	......	......
1,6	44520	4463	4473	4481	4495	4505	4515	4525	4535	4544
1,7	45543	4563	4572	4581	4590	4599	4608	4616	4624	4632
1,8	46407	4648	4656	4663	4671	4678	4685	4692	4699	4706
1,9	47128	4719	4725	4732	4738	4744	4750	4755	4761	4767
2,0	47725	4777	4783	4788	4793	4798	4803	4807	4812	4816
2,1	48214	4825	4830	4834	4838	4842	4846	4850	4853	[illegible]

Lập bảng tính $\overline{X}, S$

Chọn $x_0 = 20,25$

h = 0,5

x_i	n_i	x^*_i	y_i	$n_i y_i$	$n_i y_i^2$
19,5 – 20,0	15	19,75	-1	-15	15
20,0 – 20,5	20	20,25	0	0	0
20,5 – 21,0	10	20,75	1	10	10
21,0 – 21,5	5	21,25	2	10	20
	n = 50			$\sum = 5$	$\sum = 45$

$$\overline{X} = x_0 + \frac{h}{n}\sum_{i=1}^{4} n_i y_i = 20,25 + \frac{0,5}{50}.5 = 20,3$$

$$S^{*2} = \frac{h^2}{n}\sum n_i y_i^2 - \left(\frac{h}{n}\sum n_i y_i\right)^2 = \frac{0,5^2}{50}.45 - \left(\frac{0,5}{50}.5\right)^2 \approx 0,2225$$

$$\Rightarrow S^2 = \frac{n}{n-1}S^{*2} = \frac{50}{49}.0,2225 = 0,227 \Rightarrow S = \sqrt{S^2} \approx 0,476$$

Vậy khoảng tin cậy đối xứng cần tìm là

$$\left(20,3 - 2,06.\frac{0,476}{\sqrt{50}};\ 20,3 + 2,06.\frac{0,476}{\sqrt{50}}\right) = (20,16\ ;\ 20,44)$$

4.2. Phương pháp ước lượng khoảng

4.2.2. Ước lượng khoảng tin cậy cho kỳ vọng của ĐLNN X có phân phối chuẩn

A – Trường hợp đã biết phương sai D(X) = σ^2

B – Trường hợp chưa biết phương sai D(X) và cỡ mẫu n < 30

C – Trường hợp chưa biết phương sai D(X) và mẫu cỡ n ≥ 30

4.2. Phương pháp ước lượng khoảng

4.2.3. Ước lượng khoảng tin cậy cho xác suất p của ĐLNN

Giả sử trong tổng thể kích thước n, có m phần tử mang dấu hiệu nghiên cứu. Nếu lấy ngẫu nhiên 1 phần tử và gọi X là số phần tử mang dấu hiệu nghiên cứu được lấy ra thì X có bảng phân phối xác suất là:

X	0	1
P	1 - p	p

với p là xác suất để lấy ngẫu nhiên một phần tử thì được phần tử mang dấu hiệu nghiên cứu → p = $\frac{m}{n}$.

Khi đó: E(X) = p và D(X) = $\dfrac{p(1-p)}{n}$

Ở đây ta chỉ xét cỡ mẫu đủ lớn $\Rightarrow f \approx$ p

+ Khoảng tin cậy đối xứng của p là:

$$\left(f - t_{\frac{\alpha}{2}}\frac{\sqrt{f(1-f)}}{\sqrt{n}};\ \ f + t_{\frac{\alpha}{2}}\frac{\sqrt{f(1-f)}}{\sqrt{n}}\right)$$

+ Khoảng tin cậy bên phải của p là:

$$\left(f + t_{\alpha}\frac{\sqrt{f(1-f)}}{\sqrt{n}};\ \ +\infty\right)$$

+ Khoảng tin cậy bên trái của p là:

$$\left(-\infty;\ \ f + t_{\alpha}\frac{\sqrt{f(1-f)}}{\sqrt{n}}\right)$$

<u>**Ví dụ :**</u> Kiểm tra ngẫu nhiên 300 sản phẩm do một máy sản xuất thấy có 15 sản phẩm là phế phẩm. Với độ tin cậy 0,95 hãy **ước lượng tỉ lệ** phế phẩm <u>tối đa</u> của máy đó.

Gọi p là tỉ lệ phế phẩm của máy đó.

Vậy khoảng tin cậy bên trái của p là $\left(-\infty;\ f + t_{\alpha}\dfrac{\sqrt{f(1-f)}}{\sqrt{n}}\right)$

Độ tin cậy $\gamma = 1 - \alpha = 95\% = 0,95 \Rightarrow \Phi(t_{\alpha}) = \dfrac{1-2\alpha}{2} = 0,45$

$\Rightarrow t_{\alpha} = 1,65$

Bảng 2: HÀM LAPLACE $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_0^x e^{\frac{-t^2}{2}}dt$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0039	0079	0119	0159	0199	0239	0279	0318	0358
0,1	03983	0438	0477	0517	0556	0596	0635	0674	0714	0753
0,2	07926	0831	0870	0909	0948	0987	1025	1064	1102	1140
......	......	......	......	......	......	......	......	......	......	......
1,6	44520	4463	4473	4481	4495	4505	4515	4525	4535	4544
1,7	45543	4563	4572	4581	4590	4599	4608	4616	4624	4632
1,8	46407	4648	4656	4663	4671	4678	4685	4692	4699	[illegible]
1,9	47128	4719	4725	4732	4738	4744	4750	4755	4761	4767

<u>Ví dụ :</u> Kiểm tra ngẫu nhiên 300 sản phẩm do một máy sản xuất thấy có 15 sản phẩm là phế phẩm. Với độ tin cậy 0,95 hãy **ước lượng tỉ lệ** phế phẩm <u>tối đa</u> của máy đó.

$f = \frac{15}{300} = 0,05$

→ Khoảng tin cậy bên trái của p là

$$\left(-\infty;\ f + t_\alpha \frac{\sqrt{f(1-f)}}{\sqrt{n}}\right) = (-\infty;\ \ 0,07)$$

Vậy tỉ lệ phế phẩm tối đa của máy đó là 0,07 = 7%.

Chương 5

KIỂM ĐỊNH GIẢ THIẾT THỐNG KÊ

5.1. Khái niệm chung

5.2. Quy tắc kiểm định giả thiết thống kê

5.3. Kiểm định tham số

5.1. Khái niệm chung:

Định nghĩa: Giả thiết thống kê là giả thiết về dạng phân phối xác suất của đại lượng ngẫu nhiên, về các tham số đặc trưng của đại lượng ngẫu nhiên hoặc về tính độc lập của các đại lượng ngẫu nhiên.

+ Giả thiết thống kê đưa ra được gọi là giả thiết gốc và được ký hiệu là H_0

+ Khi đưa ra một giả thiết thống kê, người ta còn nghiên cứu kèm theo nó mệnh đề mâu thuẫn với nó, gọi là giả thiết đối và ký hiệu là H_1, để khi giả thiết H_0 bị bác bỏ thì thừa nhận giả thiết H_1.

H_0 và H_1 tạo nên cặp giả thiết thống kê

Ví dụ : Trọng lượng sản phẩm (X) do nhà máy sản xuất ra là ĐLNN phân phối chuẩn với độ lệch tiêu chuẩn $\sigma_x = 2$ kg và trọng lượng trung bình là 20 kg. Nghi ngờ máy hoạt động không bình thường làm thay đổi trọng lượng trung bình của sản phẩm, người ta cân thử 100 sản phẩm và thu được kết qủa sau:

Trọng lượng sản phẩm	19	20	21	22	23
Số sản phẩm tương ứng	10	60	20	5	5

Với mức ý nghĩa $\alpha = 0,05$ hãy kết luận về điều nghi ngờ nói trên.

+ Giả thiết gốc (H_0) : Trọng lượng trung bình của sản phẩm là 20kg

+ Giả thiết đối (H_1) : Trọng lượng trung bình của sản phẩm khác 20kg

Vì các giả thiết thống kê có thể đúng/sai nên cần phải kiểm định.

Việc kiểm định này gọi là kiểm định giả thiết thống kê.

Phương pháp chung để kiểm định giả thiết thống kê là: giả sử mệnh đề H_0 đúng → tìm được một biến cố A nào đó sao cho xác suất xảy ra biến cố A bé đến mức có thể coi như biến cố A không xảy ra → bằng một mẫu cụ thể ta thực hiện một phép thử đối với biến cố A.

+ Nếu A xảy ra thì chứng tỏ giả thiết H_0 sai và bác bỏ nó.

+ Nếu A không xảy ra thì ta chưa có cơ sở để bác bỏ H_0.

5.2. Quy tắc kiểm định giả thiết thống kê:

5.2.1. Tiêu chuẩn kiểm định

5.2.2. Mức ý nghĩa, miền bác bỏ

5.2.3. Quy tắc kiểm định

5.2.4. Sai lầm loại 1 và sai lầm loại 2

5.2. Quy tắc kiểm định giả thiết thống kê:

5.2.1. Tiêu chuẩn kiểm định

+ Từ đại lượng ngẫu nhiên X trong tổng thể, lập mẫu ngẫu nhiên cỡ n $W_x = (X_1, X_2, \ldots, X_n)$.

+ Chọn lập thống kê $G = G(X_1, X_2, \ldots, X_n, \theta_0)$

θ_0 là tham số liên quan đến giả thiết cần kiểm định.

Nếu giả thiết H_0 đúng thì quy luật phân phối xác suất của G hoàn toàn xác định.

→ Thống kê G như vậy gọi là tiêu chuẩn kiểm định.

5.2.2. Mức ý nghĩa, miền bác bỏ

Sau khi đã chọn được tiêu chuẩn kiểm định G, khi đó với một xác suất α khá bé ta có thể tìm được miền W_α sao cho với điều kiện giả thiết H_0 đúng thì $p(G \in W_\alpha) = \alpha$

+ Giá trị α gọi là mức ý nghĩa của kiểm định.

+ W_α gọi là miền bác bỏ giả thiết H_0 với mức ý nghĩa α.

5.2.3. Quy tắc kiểm định

+ Thực hiện một phép thử với mẫu ngẫu nhiên $W_x = (X_1, X_2, ..., X_n)$ thu được mẫu cụ thể $w = (x_1, x_2, ..., x_n)$.

+ Tính giá trị cụ thể của tiêu chuẩn kiểm định G, giả sử là

$$G_{qs} = f(x_1, x_2, ..., x_n, \theta_0)$$

G_{qs} gọi là giá trị quan sát của tiêu chuẩn kiểm định.

Khi đó:

+ Nếu $G_{qs} \in W_\alpha$ thì ta bác bỏ giả thiết H_0, thừa nhận giả thiết H_1.

+ Nếu $G_{qs} \notin W_\alpha$ thì chưa có cơ sở để bác bỏ giả thiết H_0 (trên thực tế là vẫn thừa nhận H_0).

5.2.4. Sai lầm loại 1 và sai lầm loại 2

+ Sai lầm loại 1: Bác bỏ giả thiết H_0 , trong khi H_0 đúng.

Xác suất mắc phải sai lầm loại này đúng bằng α.

+ Sai lầm loại 2: Thừa nhận H_0 đúng , trong khi H_0 sai.

Xác suất mắc phải sai lầm loại này bằng

$$\beta = p(G \notin W_\alpha / H_1) = 1 - p(G \in W_\alpha / H_1)$$

$\Rightarrow$ Xác suất bác bỏ giả thiết H_0 nếu nó sai là $p(G \in W_\alpha | H_1) = 1 - \beta$.

Xác suất $1 - \beta$ gọi là lực kiểm định.

5.3. Kiểm định tham số

5.3.1. Kiểm định giả thiết về kỳ vọng

Vấn đề: Giả sử ĐLNN X tuân theo quy luật phân phối chuẩn với kỳ vọng E(X) = a chưa biết. Nhưng có cơ sở nào đó nêu lên giả thiết

H_0: E(X) = a = a_0.

Cần kiểm định giả thiết này.

Xét 3 trường hợp sau:

+ Đã biết phương sai.

+ Chưa biết phương sai và kích thước mẫu n < 30.

+ Chưa biết phương sai và kích thước mẫu n ≥ 30.

a. TH1: Đã biết phương sai σ^2

+ Lập mẫu ngẫu nhiên cỡ n $W_x = (X_1, X_2, ..., X_n)$.

+ Chọn tiêu chuẩn kiểm định $U_{qs} = \dfrac{(\overline{X} - a_0)\sqrt{n}}{\sigma}$

+ Với mức ý nghĩa α cho trước, tùy thuộc vào giả thiết đối H_1 để xác định miền bác bỏ H_0 (W_α)

- Nếu H_0: a = a_0, H_1: a > a_0 thì $W_\alpha = (t_\alpha; +\infty)$
- Nếu H_0: a = a_0, H_1: a < a_0 thì $W_\alpha = (-\infty; -t_\alpha)$
- Nếu H_0: a = a_0, H_1: a ≠ a_0 thì

$$W_\alpha = \left(-\infty; -t_{\frac{\alpha}{2}}\right) \cup \left(t_{\frac{\alpha}{2}}; +\infty\right)$$

+ So sánh U_{qs} với W_α để kết luận

- Nếu $U_{qs} \in W_\alpha$ thì bác bỏ H_0, thừa nhận H_1
- Nếu $U_{qs} \notin W_\alpha$ thì chưa có cơ sở để bác bỏ H_0 (thực tế là thừa nhận H_0)

__Ví dụ 1:__ Định mức thời gian để SX ra một loại sản phẩm là __1 giờ__. Sau khi cải tiến công nghệ, người ta SX thử __150 sản phẩm__ thì thấy thời gian trung bình để được một sản phẩm là __50 phút__ với độ lệch tiêu chuẩn là __4 phút__. Hỏi với mức ý nghĩa 0,05 , có thể kết luận rằng ***công nghệ mới thực sự làm giảm chi phí thời gian SX*** hay không? Biết rằng định mức thời gian để sản xuất ra một loại sản phẩm là ĐLNN có phân phối chuẩn.

Ta chọn cặp giả thiết thống kê là H_0: a = 60; H_1: a < 60

Mẫu cỡ n = 150

Chọn tiêu chuẩn kiểm định $U_{qs} = \dfrac{(\overline{X}-a_0)\sqrt{n}}{\sigma} = \dfrac{(50-60)\sqrt{150}}{4} \approx -30,6$

Với mức ý nghĩa 0,05 và căn cứ vào cặp giả thiết thống kê, ta có miền bác bỏ H_0 là $W_\alpha = (-\infty;\ -t_\alpha)$

Trong đó:

$$\alpha = 0,05 \Rightarrow \Phi(t_\alpha) = \frac{1-2\alpha}{2} = 0,45 \Rightarrow t_\alpha = 1,65 \Rightarrow W_\alpha = (-\infty;\ -1,65)$$

Bảng 2: HÀM LAPLACE $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{t^2}{2}}dt$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0039	0079	0119	0159	0199	0239	0279	0318	0358
0,1	03983	0438	0477	0517	0556	0596	0635	0674	0714	0753
0,2	07926	0831	0870	0909	0948	0987	1025	1064	1102	1140
......	......	......	......	......	......	......	......	......	......	......
1,6	44520	4463	4473	4481	4495	4505	4515	4525	4535	4544
1,7	45543	4563	4572	4581	4590	4599	4608	4616	4624	4632
1,8	46407	4648	4656	4663	4671	4678	4685	4692	4699	4706
1,9	47128	4719	4725	4732	4738	4744	4750	4755	4761	4767
2,0	47725	4777	4783	4788	4793	4798	4803	4807	4812	4816
2,1	48214	4825	4830	4834	4838	4842	4846	4850	4853	4857

__Ví dụ 1:__ Định mức thời gian để SX ra một loại sản phẩm là __1 giờ__. Sau khi cải tiến công nghệ, người ta SX thử __150 sản phẩm__ thì thấy thời gian trung bình để được một sản phẩm là __50 phút__ với độ lệch tiêu chuẩn là __4 phút__. Hỏi với mức ý nghĩa 0,05 , có thể kết luận rằng ***công nghệ mới thực sự làm giảm chi phí thời gian SX*** hay không? Biết rằng định mức thời gian để sản xuất ra một loại sản phẩm là ĐLNN có phân phối chuẩn.

Vậy: U_{qs} = - 30,6 ; $W_\alpha = (-\infty;\ -1,65)$

Như vậy $U_{qs} \in W_\alpha$ chứng tỏ bác bỏ H_0, thừa nhận H_1

Hay có thể nói rằng việc áp dụng công nghệ mới đã làm giảm thời gian sản xuất.

5.3. Kiểm định tham số

5.3.1. Kiểm định giả thiết về kỳ vọng

__Vấn đề__: Giả sử ĐLNN X tuân theo quy luật phân phối chuẩn với kỳ vọng E(X) = a chưa biết. Nhưng có cơ sở nào đó nêu lên giả thiết

H_0: E(X) = a = a_0.

Cần kiểm định giả thiết này.

Xét 3 trường hợp sau:

+ Đã biết phương sai.

+ Chưa biết phương sai và kích thước mẫu n < 30.

+ Chưa biết phương sai và kích thước mẫu n ≥ 30.

Quy trình kiểm định giả thiết thống kê:

+ Xác định cặp giả thiết thống kê: H_0 và H_1

+ Từ mẫu cụ thể, tính giá trị quan sát U_{qs}

+ Xác định miền bác bỏ H_0 (miền W_α)

+ Kiểm tra giá trị U_{qs} có thuộc miền W_α hay không để kết luận về giả thiết thống kê nêu ra.

b. TH2: Chưa biết phương sai và kích thước mẫu n < 30

+ Xác định cặp giả thiết thống kê: H_0 và H_1

+ Lập mẫu ngẫu nhiên cỡ n $W_x = (X_1, X_2, ..., X_n)$.

Chọn tiêu chuẩn kiểm định $U_{qs} = \dfrac{(\overline{X}-a_0)\sqrt{n}}{S}$

+ Với mức ý nghĩa α cho trước, tùy thuộc vào giả thiết đối H_1 để xác định miền bác bỏ H_0 (W_α)

- Nếu H_0: a = a_0, H_1: a > a_0 thì $W_\alpha = \left(t_\alpha^{(n-1)}; +\infty\right)$
- Nếu H_0: a = a_0, H_1: a < a_0 thì $W_\alpha = \left(-\infty;\ -t_\alpha^{(n-1)}\right)$
- Nếu H_0: a = a_0, H_1: a ≠ a_0 thì

$$W_\alpha = \left(-\infty;\ -t_{\frac{\alpha}{2}}^{(n-1)}\right) \cup \left(t_{\frac{\alpha}{2}}^{(n-1)};\ +\infty\right)$$

+ So sánh U_{qs} với W_α để kết luận

- Nếu $U_{qs} \in W_\alpha$ thì bác bỏ H_0, thừa nhận H_1
- Nếu $U_{qs} \notin W_\alpha$ thì chưa có cơ sở để bác bỏ H_0 (thực tế là thừa nhận H_0)

<u>Ví dụ 1:</u> Mức xăng hao phí (X) cho một loại xe ôtô chạy từ A đến B là một đại lượng ngẫu nhiên có phân phối theo quy luật chuẩn, với <u>mức xăng hao phí trung bình là 50 lít</u>. Do đường được tu sửa lại, người ta ***cho rằng mức xăng hao phí trung bình đã <u>giảm xuống</u>***. Quan sát 25 chuyến xe trên đường AB, người ta thu được bảng số liệu sau:

Mức xăng hao phí	48,5 – 49,0	49,0 – 49,5	49,5 – 50,0	50,0 – 50,5	50,5 – 51,0
Số chuyến xe	3	8	9	3	2

Với mức ý nghĩa 0,05 hãy kết luận về ý kiến nêu trên.

Ta chọn cặp giả thiết thống kê là H_0: a = 50; H_1: a < 50

Mẫu cỡ n = 25 < 30

Chọn tiêu chuẩn kiểm định $U_{qs} = \dfrac{(\overline{X} - a_0)\sqrt{n}}{S}$ → Lập bảng tính $\overline{X}, S$

Chọn x_0 = 49,75

h = 0,5

x_i	n_i	x_i^*	y_i	$n_i y_i$	$n_i y_i^2$
48,5 – 49,0	3	48,75	-2	-6	12
49,0 – 49,5	8	49,25	-1	-8	8
49,5 – 50,0	9	49,75	0	0	0
50,0 – 50,5	3	50,25	1	3	3
50,5 – 51,0	2	50,75	2	4	8
	n = 25			$\sum = -7$	$\sum = 31$

<u>Ví dụ 1:</u> Mức xăng hao phí (X) cho một loại xe ôtô chạy từ A đến B là một đại lượng ngẫu nhiên có phân phối theo quy luật chuẩn, với <u>mức xăng hao phí trung bình là 50 lít</u>. Do đường được tu sửa lại, người ta ***cho rằng mức xăng hao phí trung bình đã <u>giảm xuống</u>***.

$$\overline{X} = x_0 + \frac{h}{n}\sum n_i y_i = 49{,}75 + \frac{0{,}5}{25}.(-7) = 49{,}61$$

$$S^{*2} = \frac{h^2}{n}\sum n_i y_i^2 - \left(\frac{h}{n}\sum n_i y_i\right)^2 = \frac{0{,}5^2}{25}.31 - \left[\frac{0{,}5}{25}.(-7)\right]^2 = 0{,}17$$

$$\Rightarrow S^2 = \frac{n}{n-1}S^{*2} = \frac{25}{24}.0{,}17 \approx 0{,}177 \Rightarrow S = \sqrt{S^2} \approx 0{,}42$$

$$\Rightarrow U_{qs} = \frac{(\overline{X} - a_0)\sqrt{n}}{S} = \frac{(49{,}61 - 50)\sqrt{25}}{0{,}42} \approx -4{,}643$$

Với mức ý nghĩa 0,05 và căn cứ vào cặp giả thiết thống kê, ta có miền bác bỏ H_0 là $W_\alpha = \left(-\infty; -t_\alpha^{(n-1)}\right)$

Trong đó: $t_\alpha^{(n-1)} = t_{0.05}^{(24)} = 1{,}711 \Rightarrow W_\alpha = (-\infty;\ -1{,}711)$

Như vậy U_{qs} = - 4,643 $\in W_\alpha$ → Bác bỏ H_0, thừa nhận H_1

→ Mức xăng hao phí trung bình đã giảm khi đường được tu sửa lại.

Bảng 3: PHÂN VỊ STUDENT $P(X > t_\alpha^{(n)}) = \alpha$ với $X \sim t(n)$

n \ α	0,1	0,05	0,025	0,01	0,005	0,0005
1	3,078	6,314	12,706	31,820	63,526	363,6
2	1,886	2,920	4,303	6,965	9,925	31,600
3	1,638	2,353	3,182	4,541	5,841	12,922
....	....	....	....	....	....	....
23	1,319	1,714	2,069	2,500	2,807	3,767
24	1,318	1,711	2,064	2,492	2,797	3,745
25	1,316	1,708	2,060	2,485	2,787	3,72[illegible]

<u>Ví dụ 2:</u> Thời gian trung bình để gia công một sp là 14 phút. Nghi ngờ trong thực tế sx, ***thời gian trung bình để hoàn thành một sp bị tăng lên***, người ta theo dõi quá trình gia công 28 sp của công nhân và thu được bảng số liệu

Thời gian gia công (phút)	10-12	12-14	14-16	16-18	18-20
Số sp tương ứng	3	6	10	4	5

Với mức ý nghĩa 0,05 hãy kết luận về điều nghi ngờ trên, biết rằng thời gian gia công sp là ĐLNN tuân theo quy luật phân phối chuẩn.

Ta chọn cặp giả thiết thống kê là H_0: a = 14; H_1: a > 14

Mẫu cỡ n = 28 < 30

Chọn tiêu chuẩn kiểm định $U_{qs} = \dfrac{(\overline{X} - a_0)\sqrt{n}}{S}$

Chọn $x_0 = 15$; $h = 2$

x_i	n_i	x^*_i	y_i	n_iy_i	$n_iy_i^2$
10 – 12	3	11	-2	-6	12
12 – 14	6	13	-1	-6	6
14 – 16	10	15	0	0	0
16 – 18	4	17	1	4	4
18 – 20	5	19	2	10	20
	n = 28			$\sum = 2$	$\sum = 42$

$$\overline{X} = x_0 + \frac{h}{n}\sum n_i y_i = 15 + \frac{2}{28}.2 = 15,143$$

$$S'^2 = \frac{h^2}{n}\sum n_i y_i^2 - \left(\frac{h}{n}\sum n_i y_i\right)^2 = \frac{2^2}{28}.42 - \left(\frac{2}{28}.2\right)^2 = 5,980$$

$$\Rightarrow S^2 = \frac{n}{n-1}S'^2 = \frac{28}{27}.5,980 = 6,201 \Rightarrow S = \sqrt{S^2} = 2,490$$

$$\Rightarrow U_{qs} = \frac{(\overline{X} - a_0)\sqrt{n}}{S} = \frac{(15,143 - 14)\sqrt{28}}{2,490} \approx 2,429$$

Với mức ý nghĩa 0,05 và căn cứ vào cặp giả thiết thống kê, ta có miền bác bỏ H_0 là $W_\alpha = \left(t_\alpha^{(n-1)}; +\infty\right)$

Trong đó: $t_\alpha^{(n-1)} = t_{0,05}^{(27)} = 1,703 \Rightarrow W_\alpha = (1,703; +\infty)$

Như vậy $U_{qs} = 2,429 \in W_\alpha$ → Bác bỏ H_0, thừa nhận H_1

→ Thời gian trung bình để hoàn thành một sản phẩm là trên 14 phút

Bảng 3: PHÂN VỊ STUDENT $P(X > t_\alpha^{(n)}) = \alpha$ với $X \sim t(n)$

n \ α	0,1	0,05	0,025	0,01	0,005	0,0005
1	3,078	6,314	12,706	31,820	63,526	363,6
2	1,886	2,920	4,303	6,965	9,925	31,600
3	1,638	2,353	3,182	4,541	5,841	12,922
......	......	......	......	......	......	......
26	1,315	1,706	2,056	2,479	2,779	3,707
27	1,314	1,703	2,052	2,473	2,771	3,690
28	1,313	1,701	2,048	2,467	2,763	3,674
29	1,311	1,699	2,045	2,462	2,756	3,659

5.3.1. Kiểm định giả thiết về kỳ vọng

c. TH3: Chưa biết phương sai và kích thước mẫu $n \geq 30$

+ Xác định cặp giả thiết thống kê: H_0 và H_1

+ Lập mẫu ngẫu nhiên cỡ n $W_x = (X_1, X_2, \ldots, X_n)$.

+ Chọn tiêu chuẩn kiểm định $U_{qs} = \dfrac{(\overline{X} - a_0)\sqrt{n}}{S}$

+ Với mức ý nghĩa α cho trước, tùy thuộc vào giả thiết đối H_1 để xác định miền bác bỏ H_0 (W_α)

- Nếu H_0: $a = a_0$, H_1: $a > a_0$ thì $W_\alpha = (t_\alpha; +\infty)$
- Nếu H_0: $a = a_0$, H_1: $a < a_0$ thì $W_\alpha = (-\infty; -t_\alpha)$
- Nếu H_0: $a = a_0$, H_1: $a \neq a_0$ thì $W_\alpha = \left(-\infty; -t_{\frac{\alpha}{2}}\right) \cup \left(t_{\frac{\alpha}{2}}; +\infty\right)$

+ So sánh U_{qs} với W_α để kết luận

- Nếu $U_{qs} \in W_\alpha$ thì bác bỏ H_0, thừa nhận H_1
- Nếu $U_{qs} \notin W_\alpha$ thì chưa có cơ sở để bác bỏ H_0 (thực tế là thừa nhận H_0)

<u>Ví dụ</u>: Định mức thời gian hoàn thành một sản phẩm là 14 phút. Có cần <u>thay đổi định mức</u> không nếu theo dõi thời gian hoàn thành một sản phẩm ở 33 công nhân, ta thu được bảng số liệu sau:

Thời gian sản xuất một sản phẩm (phút)	10-12	12-14	14-16	16-18	18-20
Số công nhân tương ứng	5	6	7	10	5

Với mức ý nghĩa 0,01 hãy kết luận về điều nghi ngờ trên, biết rằng thời gian hoàn thành một sản phẩm là ĐLNN tuân theo quy luật phân phối chuẩn.

Ta chọn cặp giả thiết thống kê là H_0: $a = 14$; H_1: $a \neq 14$

Mẫu cỡ n = 33 > 30

Chọn tiêu chuẩn kiểm định $U_{qs} = \dfrac{(\overline{X} - a_0)\sqrt{n}}{S}$

x_i	n_i	x^*_i	y_i	n_iy_i	$n_iy_i^2$
10 – 12	5	11	-3	-15	45
12 – 14	6	13	-2	-12	24
14 – 16	7	15	-1	-7	7
16 – 18	10	17	0	0	0
18 – 20	5	19	1	5	5
	n = 33			$\sum = -29$	$\sum = 81$

Chọn $x_0 = 17$; $h = 2$

$$\overline{X} = x_0 + \frac{h}{n}\sum n_i y_i = 17 + \frac{2}{33}.(-29) \approx 15,24$$

$$S'^2 = \frac{h^2}{n}\sum n_i y_i^2 - \left(\frac{h}{n}\sum n_i y_i\right)^2 = \frac{4}{33}.81 - \left[\frac{2}{33}.(-29)\right]^2 \approx 6,729$$

$$\Rightarrow S^2 = \frac{33}{32}.6,729 \approx 6,94 \Rightarrow S \approx 2,63$$

$$\Rightarrow U_{qs} = \frac{(\overline{X} - a_0)\sqrt{n}}{S} = \frac{(15,24 - 14)\sqrt{33}}{2,63} \approx 2,704$$

Với mức ý nghĩa 0,01 và căn cứ vào cặp giả thiết thống kê, ta có miền bác bỏ H_0 là $W_\alpha = \left(-\infty;\ -t_{\frac{\alpha}{2}}\right) \cup \left(t_{\frac{\alpha}{2}};\ +\infty\right)$

$$\alpha = 0,01 \Rightarrow \Phi\left(t_{\frac{\alpha}{2}}\right) = \frac{1-\alpha}{2} = 0,495 \Rightarrow t_{\frac{\alpha}{2}} = 2,58$$

H_0: a = 14; H_1: a ≠ 14

$$\Rightarrow W_\alpha = (-\infty;\ -2,58) \cup (2,58; +\infty)$$

Như vậy U_{qs} = 2,704 $\in W_\alpha$ → Bác bỏ H_0, thừa nhận H_1

Bảng 2: HÀM LAPLACE $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{\frac{-t^2}{2}} dt$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0039	0079	0119	0159	0199	0239	0279	0318	0358
0,1	03983	0438	0477	0517	0556	0596	0635	0674	0714	0753
0,2	07926	0831	0870	0909	0948	0987	1025	1064	1102	1140
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
2,4	49180	4920	4922	4924	4926	4928	4930	4932	4934	4936
2,5	49379	4939	4941	4943	4944	4946	4947	4949	4950	4952
2,6	49534	4954	4956	4957	4958	4959	4960	4962	4963	4964

<u>Ví dụ 2:</u> Trọng lượng sản phẩm (X) do một máy sx ra là ĐLNN có quy luật pp chuẩn với trọng lượng trung bình 20kg. Nghi ngờ máy hoạt động không bình thường làm ***trọng lượng trung bình của sp <u>tăng lên</u>,*** người ta cân thử 100 sản phẩm và thu được kết quả sau:

Trọng lượng sản phẩm (kg)	19	20	21	22	23
Số sản phẩm tương ứng	10	60	20	5	5

Với mức ý nghĩa 0,05 hãy kết luận về điều nghi ngờ nói trên.

Ta chọn cặp giả thiết thống kê là H_0: a = 20; H_1: a > 20

Mẫu cỡ n = 100 > 30

Chọn tiêu chuẩn kiểm định $U_{qs} = \frac{(\overline{X} - a_0)\sqrt{n}}{S}$

x_i	n_i	$x_i - x_0$	$n_i(x_i - x_0)$	$n_i(x_i - x_0)^2$
19	10	-1	-10	10
20	60	0	0	0
21	20	1	20	20
22	5	2	10	20
23	5	3	15	45
	n = 100		$\sum = 35$	$\sum = 95$

Chọn x_0 = 20

$$\overline{X} = x_0 + \frac{1}{n}\sum n_i(x_i - x_0) = 20 + \frac{1}{100}.53 = 20,53$$

$$S'^2 = \frac{1}{n}\sum n_i(x_i - x_0)^2 - \left(\frac{1}{n}\sum n_i(x_i - x_0)\right)^2 = \frac{1}{100}.95 - \left(\frac{1}{100}.35\right)^2 = 0,828$$

$$\Rightarrow S^2 = \frac{n}{n-1}S'^2 = \frac{100}{99}.0,828 = 0,836 \Rightarrow S = \sqrt{S^2} = 0,914$$

$$\Rightarrow U_{qs} = \frac{(\overline{X} - a_0)\sqrt{n}}{S} = \frac{(20,53 - 20)\sqrt{100}}{0,914} \approx 5,799$$

Với mức ý nghĩa 0,05 và căn cứ vào cặp giả thiết thống kê, ta có miền bác bỏ H_0 là $W_\alpha = (t_\alpha;\ +\infty)$

$$\alpha = 0,05 \Rightarrow \Phi(t_\alpha) = \frac{1-2\alpha}{2} = 0,450 \quad \Rightarrow t_\alpha = 1,65 \Rightarrow W_\alpha = (1,65\ ;\ +\infty)$$

Như vậy U_{qs} = 5,799 $\in W_\alpha$ → Bác bỏ H_0, thừa nhận H_1

Bảng 2: HÀM LAPLACE $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{\frac{-t^2}{2}} dt$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0039	0079	0119	0159	0199	0239	0279	0318	0358
0,1	03983	0438	0477	0517	0556	0596	0635	0674	0714	0753
0,2	07926	0831	0870	0909	0948	0987	1025	1064	1102	1140
......	......	......	......	......	......	......	......	......	......	......
1,6	44520	4463	4473	4481	4495	4505	4515	4525	4535	4544
1,7	45543	4563	4572	4581	4590	4599	4608	4616	4624	4632
1,8	46407	4648	4656	4663	4671	4678	4685	4692	4699	4706
1,9	47128	4719	4725	4732	4738	4744	4750	4755	4761	4767
2,0	47725	4777	4783	4788	4793	4798	4803	4807	4812	4816

5.3. Kiểm định tham số

5.3.2. Kiểm định giả thiết về tỷ lệ (xác suất)

<u>Vấn đề</u>: Giả sử ĐLNN X tuân theo quy luật phân phối không – một với tham số p chưa biết. Nhưng có cơ sở nào đó nêu lên giả thiết

H_0 : p = p_0.

Ta cần kiểm định giả thiết này.

+ Xác định cặp giả thiết thống kê H_0 , H_1 của bài toán.

+ Lập mẫu ngẫu nhiên cỡ n $W_x = (X_1, X_2, ..., X_n)$.

Chọn tiêu chuẩn kiểm định $U_{qs} = \frac{(\overline{X} - p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}} \approx \frac{(f - p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}}$

+ Với mức ý nghĩa α cho trước, tùy thuộc vào giả thiết đối H_1 để xác định miền bác bỏ H_0 (W_α)

- Nếu H_0: p = p_0, H_1: p > p_0 thì $W_\alpha = (t_\alpha; +\infty)$
- Nếu H_0: p = p_0, H_1: p < p_0 thì $W_\alpha = (-\infty;\ -t_\alpha)$
- Nếu H_0: p = p_0, H_1: p ≠ p_0 thì

$$W_\alpha = \left(-\infty;\ -t_{\frac{\alpha}{2}}\right) \cup \left(t_{\frac{\alpha}{2}};\ +\infty\right)$$

+ So sánh U_{qs} với W_α để kết luận

- Nếu $U_{qs} \in W_\alpha$ thì bác bỏ H_0, thừa nhận H_1
- Nếu $U_{qs} \notin W_\alpha$ thì chưa có cơ sở để bác bỏ H_0 (thực tế là thừa nhận H_0)

Ví dụ: Tỉ lệ phế phẩm ở một nhà máy là 0,07. Trong một lần kiểm tra chất lượng sản phẩm, người ta kiểm tra ngẫu nhiên 350 sản phẩm thì thấy có 30 sản phẩm là phế phẩm. Với mức ý nghĩa 0,05 có thể khẳng định tỉ lệ phế phẩm tăng lên hay không?

Kích thước mẫu n = 350.

Ta chọn cặp giả thiết thống kê là H_0: p = 0,07 ; H_1: p > 0,07

Chọn tiêu chuẩn kiểm định $U_{qs} = \dfrac{(f - p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}}$

Trong đó $f = \dfrac{m}{n} = \dfrac{30}{350} \approx 0{,}086 \Rightarrow U_{qs} = \dfrac{(0{,}086 - 0{,}07)\sqrt{350}}{\sqrt{0{,}07.(1-0{,}07)}} \approx 1{,}17$

Với mức ý nghĩa 0,05 và căn cứ vào cặp giả thiết thống kê, ta có miền bác bỏ H_0 là $W_\alpha = (t_\alpha ;\ +\infty)$

$$\alpha = 0{,}05 \Rightarrow \Phi(t_\alpha) = \frac{1-2\alpha}{2} = 0{,}450 \quad \Rightarrow t_\alpha = 1{,}65 \Rightarrow W_\alpha = (1{,}65 ;\ +\infty)$$

Như vậy U_{qs} = 1,17 $\notin W_\alpha$ → Thừa nhận H_0

Bảng 2: HÀM LAPLACE $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{\frac{-t^2}{2}} dt$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0039	0079	0119	0159	0199	0239	0279	0318	0358
0,1	03983	0438	0477	0517	0556	0596	0635	0674	0714	0753
0,2	07926	0831	0870	0909	0948	0987	1025	1064	1102	1140
......	....	......	......	......	......	......	......	......	......	......
1,6	44520	4463	4473	4481	4495	4505	4515	4525	4535	4544
1,7	45543	4563	4572	4581	4590	4599	4608	4616	4624	4632
1,8	46407	4648	4656	4663	4671	4678	4685	4692	4699	4706
1,9	47128	4719	4725	4732	4738	4744	4750	4755	4761	4767
2,0	47725	4777	4783	4788	4793	4798	4803	4807	4812	4816

Chương 5

KIỂM ĐỊNH GIẢ THIẾT THỐNG KÊ

*** CÁC KIẾN THỨC CẦN NẮM CHẮC:**

- Nắm được các khái niệm về kiểm định giả thiết thống kê, cặp giả thiết thống kê, quy tắc kiểm định giả thiết thống kê.
- Nhận dạng bài toán kiểm định giả thiết thống kê về tham số nào, trường hợp nào, cặp giả thiết thống kê là gì.
- Biết áp dụng các công thức kiểm định tương ứng.

CHƯƠNG 6

LÝ THUYẾT TƯƠNG QUAN VÀ HÀM HỒI QUY

6.1. Mở đầu

6.2. Hệ số tương quan mẫu

6.3. Hàm hồi quy

6.4. Thực hành tính hệ số tương quan

6.1. Mở đầu

Trong nhiều bài toán thực tế người ta quan tâm đến quan hệ của hai hoặc nhiều biến ngẫu nhiên.

+ Nếu hai biến ngẫu nhiên độc lập, ta có thể nghiên cứu chúng riêng rẽ.

+ Nếu hai biến ngẫu nhiên không độc lập, ta cần xác định mức độ phụ thuộc và quan hệ hàm giữa các biến.

6.2. Hệ số tương quan mẫu

Xét cặp ĐLNN (X,Y) với bộ số liệu quan sát (x_i, y_i)

Nếu X và Y có sự phụ thuộc nhau (tương quan với nhau) thì sự phụ thuộc này được đặc trưng bởi **hệ số tương quan r** xác định bằng công thức:

$$r = \frac{\overline{XY} - \overline{X}.\overline{Y}}{S_X^*.S_Y^*}$$

*** Tính chất, ý nghĩa của hệ số tương quan mẫu**

1) $|r| \leq 1$

2) Nếu X, Y độc lập thì r = 0.

Như vậy:

+ Nếu r = 0 thì X và Y không có mối tương quan với nhau.

+ Nếu r khá gần 0 thì X và Y tương quan yếu.

+ Nếu $|r|$ khá gần 1 thì X và Y tương quan chặt (tuyến tính với nhau).

6.3. Hàm hồi quy

Khi X, Y có tương quan tuyến tính thì phương trình hồi quy tuyến tính có dạng:

+ Phương trình của Y theo X là:

$$Y - \overline{Y} = a(X - \overline{X}), \quad a = r.\frac{S^*_Y}{S^*_X}$$

+ Phương trình của X theo Y là:

$$X - \overline{X} = b(Y - \overline{Y}), \quad b = r.\frac{S^*_X}{S^*_Y}$$

Các hệ số a, b gọi là hệ số hồi quy.

6.4. Thực hành tính hệ số tương quan và viết phương trình hồi quy tuyến tính

Ví dụ 1: Nghiên cứu khả năng học Toán (X) và khả năng học Lý (Y) của 15 em học sinh trong một nhóm, ta có bảng số liệu sau:

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
4	1	[illegible]	[illegible]	[illegible]
5		3	2	4
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

a. Hãy ước lượng hệ số tương quan giữa khả năng học Toán và khả năng học Lý của nhóm học sinh trên.

b. Viết phương trình hồi quy tuyến tính của Y theo X. Từ đó dự đoán: một học sinh đạt 7 điểm môn Toán thì khả năng đạt mấy điểm môn Lý?

a/ Hệ số tương quan r : $r = \dfrac{\overline{XY}-\overline{X}.\overline{Y}}{S_X^{*}.S_Y^{*}}$

X_i	n_i	$n_i.X_i$	$n_i.X_i^2$
4	3	12	48
5	9	45	225
6	3	18	108
	15	**75**	**381**

$\overline{X}=\frac{1}{n}\sum n_i.X_i=\frac{1}{15}.75=5$

$S^{*}{}_X^{2}=\frac{1}{n}\sum n_i.x_i^2-\left(\overline{X}\right)^2$

$=\frac{1}{15}.381-(5)^2=0,4\Rightarrow S^{*}{}_X\approx 0,63$

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
5		3	2	4
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Y_j	m_j	$m_j.Y_j$	$m_j.Y_j^2$
2	1	2	4
3	5	15	45
4	3	12	48
5	6	30	150
	15	59	247

$\overline{Y}=\frac{1}{m}\sum m_j.Y_j=\frac{1}{15}.59\approx 3,93$

$S^{*}{}_Y^{2}=\frac{1}{m}\sum m_j.Y_j^2-\left(\overline{Y}\right)^2$

$=\frac{1}{15}.247-(3,93)^2\approx 1,02\Rightarrow S^{*}{}_Y\approx 1,01$

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	1	[illegible]	[illegible]	[illegible]
5		3	2	4
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

$\overline{XY}=\frac{1}{m}\sum m_{ij}.X_i.Y_j$

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
5		45	40	100
[illegible]	[illegible]	[illegible]	24	[illegible]
[illegible]	8	69	64	160

$\overline{XY}=\frac{1}{m}\sum m_{ij}.X_i.Y_j$

$=\frac{1}{15}.(8+69+64+160)\approx 20,07$

$\overline{X}=\frac{1}{n}\sum n_i.X_i=\frac{1}{15}75=5$

$\overline{Y}=\frac{1}{m}\sum m_j.Y_j=\frac{1}{15}59\approx 3,93$

$S_X^{*2}=\frac{1}{15}381-(5)^2=0,4\Rightarrow S_X^{*}=\sqrt{S_X^{*2}}\approx 0,63;$

$S_Y^{*2}=\frac{1}{15}247-(3,93)^2\approx 1,02\Rightarrow S_Y^{*}=\sqrt{S_Y^{*2}}\approx 1,01$

$\overline{XY}=\frac{1}{n}\sum n_{ij}.X_iY_j=\frac{1}{15}301\approx 20,07$

$\Rightarrow r=\dfrac{\overline{XY}-\overline{X}.\overline{Y}}{S_X^{*}S_Y^{*}}=\dfrac{20,07-5.3,93}{0,63.1,01}\approx 0,66$

b. Viết phương trình hồi quy tuyến tính của Y theo X.

Phương trình hồi quy tuyến tính của Y theo X là:

$$Y-\overline{Y}=a(X-\overline{X}),\qquad a=r.\frac{S^{*}{}_Y}{S^{*}{}_X}$$

với $\overline{X}=5$; $\overline{Y}=3,93$

$a=r.\dfrac{S_Y^{*}}{S_X^{*}}=0,66.\dfrac{1,01}{0,63}\approx 1,06$

$\Rightarrow Y=1,06(X-5)+3,93=1,06X-1,37$

Thay X = 7 → Y = 6,05

Chứng tỏ: Một học sinh đạt 7 điểm môn Toán thì có khả năng đạt 6,05 điểm môn Lý.

Vậy hệ số tương quan r :

$$r = \frac{\overline{XY} - \overline{X}.\overline{Y}}{S_X^*.S_Y^*} = \frac{8,382 - 2,836.3}{0,173.1,044} \approx -0,698$$

Hệ số tương quan $r \neq 0$ chứng tỏ X và Y phụ thuộc tương quan với nhau.

b) Phương trình tương quan tuyến tính của Y đối với X :

$$Y - \overline{Y} = a(X - \overline{X}), \quad a = r.\frac{S^*_Y}{S^*_X}$$

với $\overline{X} = 2,836$; $\overline{Y} = 3$

$$a = r.\frac{S_Y^*}{S_X^*} = -0,698.\frac{1,044}{0,173} \approx 4,212$$

$$\Rightarrow Y - 3 = 4,212.(X - 2,836) \Rightarrow Y = 4,212.X - 21,58$$

Ví dụ 2: Điều tra độ mòn của lưỡi dao của một xưởng kim khí, người ta đo độ dày của lưỡi dao và thu được bảng số liệu sau:

Y / X	2	3	5
2,6		1	2
2,8	1	2	
3,0	3	2	

Trong đó :

X - Độ dày của lưỡi dao (tính bằng mm).

Y - Thời gian sử dụng (tính theo ngày).

a) Giữa X và Y có phụ thuộc tương quan không? Vì sao?

b) Viết phương trình tương quan tuyến tính của Y đối với X.

CHƯƠNG 6

LÝ THUYẾT TƯƠNG QUAN VÀ HÀM HỒI QUY

*** CÁC KIẾN THỨC CẦN NẮM CHẮC :**

+ Ý nghĩa của hệ số tương quan mẫu r.

+ Công thức xác định hệ số tương quan mẫu r.

+ Phương trình hồi quy tuyến tính của Y đối với X, của X đối với Y .

CHƯƠNG 3
KẾT LUẬN

Đề tài bài giảng trực tuyến môn Xác suất Thống kê có ý nghĩa rất quan trọng trong việc nâng cao chất lượng đào tạo sinh viên trường Đại học Kỹ thuật Công nghiệp. Bên cạnh đó, hệ thống video bài giảng là tài liệu tham khảo giúp sinh viên có thể ôn lại bài giảng khi học trực tiếp trên lớp. Đề tài nghiên cứu này đã xây dựng được bộ video bài giảng học phần Xác suất Thống kê, bao gồm 17 video có nội dung bám sát theo đề cương chi tiết, thời gian mỗi video phù hợp.

Đề tài có ý nghĩa quan trọng đối với quá trình chuyển đổi số của Nhà trường, là tài liệu quan trọng phục vụ chương trình đào tạo từ xa, giảng dạy và học tập môn Xác suất Thống kê trong trường Đại học kỹ thuật Công nghiệp.

DANH MỤC TÀI LIỆU THAM KHẢO

[1]. Nguyễn Cao Văn; Trần Thái Ninh – Giáo trình lý thuyết Xác suất và thống kê Toán – Nhà xuất bản Thống kê, 2005.

[2]. Tống Đình Quỳ – Giáo trình Xác suất thống kê – Nhà xuất bản Đại học Quốc gia Hà Nội, 2004.

[3]. Phạm Văn Kiều – Giáo trình Xác suất thống kê – Nhà xuất bản Giáo dục, 2007.

[4]. Nguyễn Cao Văn; Trần Thái Ninh; Nguyễn Thế Hệ – Bài tập Xác suất và thống kê Toán – Nhà xuất bản Đại học Kinh tế Quốc dân, 2006.

[5]. Bộ môn Toán, Khoa Khoa học Cơ bản, Đại học Kỹ thuật Công nghiệp Thái Nguyên – Bài giảng môn Toán 5.

ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC KỸ THUẬT CÔNG NGHIỆP

THUYẾT MINH
ĐỀ TÀI KHOA HỌC VÀ CÔNG NGHỆ CẤP TRƯỜNG
NĂM 2022

TÊN ĐỀ TÀI: XÂY DỰNG BÀI GIẢNG TRỰC TUYẾN MÔN XÁC SUẤT THỐNG KÊ
MÃ SỐ: T2022-VD18

Chủ nhiệm đề tài: ThS Nguyễn Thị Xuân Mai

THÁI NGUYÊN, NĂM 2022

THUYẾT MINH ĐỀ TÀI
KHOA HỌC VÀ CÔNG NGHỆ CẤP TRƯỜNG NĂM 2022

1. TÊN ĐỀ TÀI: Xây dựng bài giảng trực tuyến môn Xác suất Thống kê	2. MÃ SỐ: T2022-VD18

3. LĨNH VỰC NGHIÊN CỨU

Khoa học Tự nhiên	x	Khoa học Kỹ thuật và Công nghệ	
Khoa học Y, dược		Khoa học Nông nghiệp	
Khoa học Xã hội		Khoa học Nhân văn	

4. LOẠI HÌNH NGHIÊN CỨU

Cơ bản	Ứng dụng	Triển khai
		x

5. THỜI GIAN THỰC HIỆN DỰ KIẾN: 12 tháng

Từ tháng 04 năm 2022 đến tháng 04 năm 2023

6. CHỦ NHIỆM ĐỀ TÀI

Họ và tên: Nguyễn Thị Xuân Mai   Học vị: Thạc sỹ

Chức danh khoa học: Năm sinh: 1985

Địa chỉ cơ quan: ĐH KTCN   Điện thoại di động: 0983984802

Điện thoại cơ quan: Fax:

E-mail: nguyenthixuanmai@tnut.edu.vn

7. NHỮNG THÀNH VIÊN THAM GIA NGHIÊN CỨU ĐỀ TÀI

TT	Họ và tên	Đơn vị công tác và lĩnh vực chuyên môn	Nội dung nghiên cứu cụ thể được giao	Chữ ký
1	Phạm Thị Thu Hằng	Khoa KHCB – ĐH KTCN. Lĩnh vực chuyên môn: Toán học	Thiết kế bài giảng power point chương 1, 2,3.	
2	Ngô Văn Giang	Khoa KHCB – ĐH KTCN. Lĩnh vực chuyên môn: Toán học	Thiết kế bài giảng power point chương 4,5.	

8. ĐƠN VỊ PHỐI HỢP CHÍNH

Tên đơn vị trong và ngoài nước	Nội dung phối hợp nghiên cứu	Họ và tên người đại diện đơn vị

9. TỔNG QUAN TÌNH HÌNH NGHIÊN CỨU THUỘC LĨNH VỰC CỦA ĐỀ TÀI Ở TRONG VÀ NGOÀI NƯỚC

9.1. Trong nước *(phân tích, đánh giá tình hình nghiên cứu thuộc lĩnh vực của đề tài ở Việt Nam, liệt kê danh mục các công trình nghiên cứu, tài liệu có liên quan đến đề tài được trích dẫn khi đánh giá tổng quan)*

- Nhờ sự thành công rực rỡ của cuộc cách mạng khoa học công nghệ 4.0 nên hiện nay ngoài phương thức học trực tiếp, người học còn có xu thế tham gia học trực tuyến. Theo đó, người học có thể dễ dàng tìm kiếm được trên Internet nhiều video bài giảng ở nhiều lĩnh vực khác nhau của Toán học, trong đó có Xác suất thống kê để học tập và tham khảo.

- Các video bài giảng về các nội dung kiến thức Xác suất thống kê cũng đang được quan tâm nhiều trong giai đoạn hiện nay, tuy nhiên số lượng các video bài giảng về lĩnh vực này chưa được nhiều và phong phú như các lĩnh vực khác của Toán học.

9.2. Ngoài nước *(phân tích, đánh giá tình hình nghiên cứu thuộc lĩnh vực của đề tài trên thế giới, liệt kê danh mục các công trình nghiên cứu, tài liệu có liên quan đến đề tài được trích dẫn khi đánh giá tổng quan)*

9.3. Danh mục các công trình đã công bố thuộc lĩnh vực của đề tài của chủ nhiệm và những thành viên tham gia nghiên cứu (*họ và tên tác giả; bài báo; ấn phẩm; các yếu tố về xuất bản*)

a) Của chủ nhiệm đề tài

b) Của các thành viên tham gia nghiên cứu

(Những công trình được công bố trong 5 năm gần nhất)

10. TÍNH CẤP THIẾT CỦA ĐỀ TÀI

- Hiện nay, trước tình hình dịch bệnh Covid đang diễn biến phức tạp, khó lường, việc học tập được linh hoạt chuyển từ hình thức học trực tiếp sang hình thức học trực tuyến là xu thế tất yếu không chỉ ở Việt Nam mà tại nhiều quốc gia trên thế giới. Học tập trực tuyến đem lại rất nhiều lợi ích hơn so với cách học truyền. Theo đó, người học giờ đây đóng vai trò trung tâm và chủ động của quá trình đào tạo, họ có thể học mọi lúc, mọi nơi, học theo thời gian biểu cá nhân,xem đi xem lại nhiều lần với nhịp độ tuỳ theo khả năng và mục đích của mình, có thể chọn các nội dung học,...

- Tuy nhiên, hình thức học tập này chưa thực sự đạt được chất lượng như mong muốn do nhiều yếu tố chủ quan và khách quan như điều kiện về cơ sở vật chất, ý thức học tập của sinh viên, khó khăn trong việc quản lý sinh viên trong giờ học của giáo viên, ...

- Học phần Xác suất thống kê là học phần được giảng dạy cho sinh viên một số Khoa của trường ĐH KTCN như Khoa Kinh tế Công nghiệp, Khoa Xây dựng và Môi trường . Nội dung kiến thức của học phần này được sử dụng nhiều trong các môn học chuyên ngành của sinh viên trong những năm học tiếp theo

- Với mong muốn nâng cao chất lượng học tập cho sinh viên trường ĐH KTCN trong giai đoạn học online cũng như sau này. Đồng thời giúp các sinh viên không được học học phần Xác suất thống kê trong khung chương trình của mình vẫn có thể tiếp cận với môn học này, nhóm nghiên cứu đề xuất đề tài: *"Xây dựng bài giảng trực tuyến môn Xác suất thống kê"* để sinh viên có thể tiếp thu kiến thức tốt hơn khi xem các video giảng dạy trước và sau mỗi buổi học cũng như ôn tập cuối kỳ.

11. MỤC TIÊU ĐỀ TÀI

- Nghiên cứu tìm hiểu cách xây dựng và biên tập video bài giảng.
- Xây dựng các video tóm tắt được các nội dung cơ bản của học phần Xác suất thống kê giúp cho sinh viên học tập tốt hơn học phần này.

12. ĐỐI TƯỢNG, PHẠM VI NGHIÊN CỨU

12.1. Đối tượng nghiên cứu: Xây dựng video bài giảng các nội dung cơ bản của học phần Xác suất thống kê.

12.2. Phạm vi nghiên cứu: Các nội dung dạy và học học phần Xác suất thống kê tại trường ĐH Kỹ thuật Công nghiệp.

13. CÁCH TIẾP CẬN, PHƯƠNG PHÁP NGHIÊN CỨU

13.1. Cách tiếp cận: Thu thập thông tin – Luận cứ lý thuyết, thực tiễn – Phân tích, thảo luận – Kết luận, đề nghị.

13.2. Phương pháp nghiên cứu: Đề tài sử dụng các phương pháp nghiên cứu:

- Phương pháp nghiên cứu lý thuyết.
- Phương pháp thực nghiệm sư phạm.

14. NỘI DUNG NGHIÊN CỨU VÀ TIẾN ĐỘ THỰC HIỆN

14.1. Nội dung nghiên cứu (***Mô tả chi tiết những nội dung nghiên cứu của đề tài***)

14.2. Tiến độ thực hiện

STT	Các nội dung, công việc thực hiện	Sản phẩm	Thời gian (bắt đầu-kết thúc)	Người thực hiện
1	Xây dựng thuyết minh đề tài, xây dựng đề cương cho các video bài giảng, nghiên cứu tìm hiểu các phần mềm thiết kế xây dựng video bài giảng.	Báo cáo	04/2022-06/2022	Nguyễn Thị Xuân Mai
2	Xây dựng video bài giảng chương 1, 2, 3 của học phần Xác suất thống kê.	Các video bài giảng	07/2022-12/2022	Nguyễn Thị Xuân Mai; Phạm Thị Thu Hằng; Ngô Văn Giang.
3	Xây dựng video bài giảng chương 4,5 của học phần Xác suất thống	Các video bài giảng	01/2023-04/2023	Nguyễn Thị Xuân Mai;

	kê.			Phạm Thị Thu Hằng; Ngô Văn Giang.

15. SẢN PHẨM

Stt	Tên sản phẩm	Số lượng	Yêu cầu chất lượng sản phẩm *(mô tả chi tiết chất lượng sản phẩm đạt được như nội dung, hình thức, các chỉ tiêu, thông số kỹ thuật,...)*
I	Sản phẩm khoa học (Các công trình khoa học sẽ được công bố: sách, bài báo khoa học, ..)		
1.1			
1.2			
...			
II	Sản phẩm đào tạo (cử nhân, thạc sĩ, tiến sĩ,...)		
2.1			
2.2			
...			
III	Sản phẩm ứng dụng		
3.1	Video bài giảng học phần Xác suất thống kê	Các video bài giảng	
3.2			
...			

16. PHƯƠNG THỨC CHUYỂN GIAO KẾT QUẢ NGHIÊN CỨU VÀ ĐỊA CHỈ ỨNG DỤNG

16.1. Phương thức chuyển giao

16.2. Địa chỉ ứng dụng

17. TÁC ĐỘNG VÀ LỢI ÍCH MANG LẠI CỦA KẾT QUẢ NGHIÊN CỨU

17.1. Đối với lĩnh vực giáo dục và đào tạo: tạo ra sản phẩm học thuật có chất lượng và có ý nghĩa thực tiễn trong dạy và học các học phần Toán cho sinh viên khối trường đại học chuyên ngành kỹ thuật.

17.2. Đối với lĩnh vực khoa học và công nghệ có liên quan: Phát triển hướng nghiên cứu đổi mới phương pháp dạy và học theo định hướng phát triển năng lực người học và đổi mới giáo dục trong thời đại 4.0.

17.3. Đối với phát triển kinh tế-xã hội: Tạo ra cơ sở khoa học cho việc xây dựng và phát triển thương hiệu một trường đại học, cơ sở giáo dục nghề nghiệp cho đất nước trong thời đại công nghiệp hóa ngày nay.

17.4. Đối với tổ chức chủ trì và các cơ sở ứng dụng kết quả nghiên cứu: nâng cao chất lượng dạy và học trong quá trình đào tạo của trường ĐH Kỹ Thuật Công Nghiệp, giúp định hướng

vai trò truyền thông ngày nay trong công cuộc xây dựng niềm tin của người học vào cơ sở đào tạo nghề nghiệp.
18. KINH PHÍ THỰC HIỆN ĐỀ TÀI **Tổng kinh phí: 3.600.000đ** *Bằng chữ: Ba triệu sáu trăm nghìn đồng chẵn.* *(Dự toán chi tiết các mục chi đính kèm có xác nhận của các đơn vị liên quan.)*
Ngày 15 tháng 4 năm 2022 **Chủ nhiệm đề tài** — **PHÒNG KHCN&HTQT** **Th.S Nguyễn Thị Xuân Mai**
HỘI ĐỒNG KHOA KHCB — **KT. HIỆU TRƯỞNG** **PHÓ HIỆU TRƯỞNG** Nguyễn Văn Thắng — **PGS.TS. Vũ Ngọc Pi**

ĐVT: VNĐ

DỰ TOÁN KINH PHÍ ĐỀ TÀI KH&CN CẤP TRƯỜNG NĂM 2022

Tên đề tài: Xây dựng bài giảng trực tuyến môn Xác suất Thống kê

Chủ nhiệm đề tài: ThS. Nguyễn Thị Xuân Mai

Thành viên chính: ThS.GVC Phạm Thị Thu Hằng, ThS. Ngô Văn Giang

Thành viên:

ĐVT: VNĐ

STT	Nội dung	Dự toán			
		Người thực hiện	Số ngày công	Hệ số tiền công theo ngày (2)*	Thành tiền
1	*Mục chi tiền công lao động tham gia trực tiếp (1)*				
1.1	Xây dựng thuyết minh đề tài. Nghiên cứu tìm hiểu các phần mềm thiết kế xây dựng video bài giảng.	Nguyễn Thị Xuân Mai	0,5	0,45	335.250
1.2	Xây dựng hệ thống video bài giảng chương 1, 2, 3	Phạm Thị Thu Hằng	1	0,3	447.000
1.3		Ngô Văn Giang	1	0,3	447.000
1.4		Nguyễn Thị Xuân Mai	1	0,45	670.500
1.5	Xây dựng hệ thống video bài giảng chương 4, 5	Ngô Văn Giang	1	0,45	447.000
1.6		Nguyễn Thị Xuân Mai	1	0,45	670.500
1.7	Viết báo cáo nghiệm thu	Nguyễn Thị Xuân Mai	0,5	0,45	335.250
	Tổng 1		6		**3.352.500**
2	*Mục chi phí khác*				
2.1	Phô tô, in ấn				**247.500**
	Tổng 2				**247.500**
	Tổng 1+2				**3.600.000**

Cơ quan chủ trì KT. HIỆU TRƯỞNG PHÓ HIỆU TRƯỞNG	TRƯỞNG PHÒNG KHCN&HTQT	CHỦ NHIỆM ĐỀ TÀI
PGS.TS. Vũ Ngọc Pi		ThS. Nguyễn Thị Xuân Mai

KHTC