BỘ Y TẾ

XÁC SUẤT THỐNG KÊ

(DÙNG CHO ĐÀO TẠO BÁC SỸ ĐA KHOA)

Chủ biên : TS. ĐẶNG ĐỨC HẬU

NHÀ XUẤT BẢN GIÁO DỤC

BỘ Y TẾ

XÁC SUẤT THỐNG KÊ

(DÙNG CHO ĐÀO TẠO BÁC SĨ ĐA KHOA)

Mã số: Đ.01.X.02

NHÀ XUẤT BẢN GIÁO DỤC
HÀ NỘI – 2008

Chỉ đạo biên soạn:

VỤ KHOA HỌC VÀ ĐÀO TẠO – BỘ Y TẾ

Chủ biên:

TS. ĐẶNG ĐỨC HẬU

Tham gia biên soạn:

TS. ĐẶNG ĐỨC HẬU

TS. HOÀNG MINH HẰNG

Thư kí biên soạn:

TS. HOÀNG MINH HẰNG

Tham gia tổ chức bản thảo:

ThS. PHÍ VĂN THÂM

04 - 2009/CXB/262-2117/GD

Mã số: 7B725y9 - DAI

LỜI GIỚI THIỆU

Thực hiện một số điều của Luật Giáo dục, Bộ Giáo dục & Đào tạo và Bộ Y tế đã ban hành chương trình khung đào tạo ***Bác sĩ đa khoa****. Bộ Y tế tổ chức biên soạn tài liệu dạy – học các môn cơ sở và chuyên môn theo chương trình trên nhằm từng bước xây dựng bộ sách đạt chuẩn chuyên môn trong công tác đào tạo nhân lực y tế.*

Sách XÁC SUẤT THỐNG KÊ được biên soạn dựa vào chương trình giáo dục của Trường Đại học Y Hà Nội trên cơ sở chương trình khung đã được phê duyệt. Sách được TS. Đặng Đức Hậu (Chủ biên), TS. Hoàng Minh Hằng biên soạn theo phương châm: kiến thức cơ bản, hệ thống; nội dung chính xác, khoa học; cập nhật các tiến bộ khoa học, kỹ thuật hiện đại và thực tiễn Việt Nam.

Sách XÁC SUẤT THỐNG KÊ đã được Hội đồng chuyên môn thẩm định sách và tài liệu dạy – học chuyên ngành ***Bác sĩ đa khoa*** *của Bộ Y tế thẩm định năm 2008. Bộ Y tế quyết định ban hành tài liệu dạy – học đạt chuẩn chuyên môn của ngành trong giai đoạn hiện nay. Trong thời gian từ 3 đến 5 năm, sách phải được chỉnh lý, bổ sung và cập nhật.*

Bộ Y tế chân thành cảm ơn các tác giả và Hội đồng chuyên môn thẩm định đã giúp hoàn thành cuốn sách; cảm ơn PGS.TS. Đỗ Văn Dũng, ThS. Nguyễn Phan Dũng đã đọc và phản biện để cuốn sách sớm hoàn thành, kịp thời phục vụ cho công tác đào tạo nhân lực y tế.

Lần đầu xuất bản, chúng tôi mong nhận được ý kiến đóng góp của đồng nghiệp, các bạn sinh viên và các độc giả để lần xuất bản sau sách được hoàn thiện hơn.

VỤ KHOA HỌC VÀ ĐÀO TẠO – BỘ Y TẾ

LỜI NÓI ĐẦU

Lý thuyết xác suất và thống kê phát triển mạnh mẽ trong thế kỷ XX. Vào những năm của nửa cuối thế kỷ XX, xác suất thống kê được áp dụng rộng rãi trong nhiều lĩnh vực, trong đó có kinh tế, xã hội, điều khiển học và sinh, y học. Ngày nay không một công trình nghiên cứu nào mà không sử dụng các phương pháp thống kê khi xử lí số liệu.

Từ những năm 60 của thế kỷ trước, bộ môn Toán đã giảng dạy xác suất thống kê cho các sinh viên y và hướng dẫn xử lý số liệu thu được trong các nghiên cứu. Sau nhiều năm giảng dạy và ứng dụng, nội dung của cuốn sách dần hình thành và được chọn lọc, nó cũng chính là nội dung cho lần xuất bản này.

Bài giảng xác suất và thống kê được viết lần này theo chương trình Đại học đại cương có mở rộng và nâng cao. Cuốn sách không những cung cấp các kiến thức cơ bản về xác suất thống kê mà còn đưa ra một số ví dụ ứng dụng gần gũi và thiết thực về xác suất thống kê trong y học. Nội dung của cuốn sách là tài liệu học tập cho sinh viên hệ bác sĩ đa khoa và đồng thời cũng có thể là tài liệu tham khảo cho học viên sau đại học, cho các cán bộ giảng dạy xác suất thống kê trong ngành y và cho những người cần xử lý số liệu trong các nghiên cứu y học.

Với thời lượng 45 tiết, bài giảng xác suất và thống kê bao gồm hai phần chính là xác suất và thống kê. Xác suất làm cho ta hiểu rõ hơn về khả năng xuất hiện của các hiện tượng ngẫu nhiên cũng như các quy luật xác suất của chúng và nhờ đó giúp ta đánh giá đúng, phán đoán đúng hơn về các hiện tượng ngẫu nhiên. Thống kê giúp xử lí số liệu từ đó có thể so sánh đánh giá đúng về hiệu quả chẩn đoán và điều trị của các phương pháp, góp phần đưa ra các khuyến cáo về chẩn đoán và điều trị.

Khi đọc tài liệu này cần có các kiến thức cơ bản về giải tích, các kiến thức đó được trình bày trong các sách toán cao cấp phần giải tích.

Ứng dụng xác suất thống kê vào thực tiễn, đặc biệt là trong y học, là việc làm rất quan trọng và cần thiết. Viết tài liệu này cũng là một phần mong mỏi đáp ứng yêu cầu trên. Tuy vậy đây cũng là việc làm có nhiều khó khăn, khi đưa các lý thuyết toán học rất chặt chẽ và chính xác vào ứng dụng trong một ngành khoa học mang nhiều tính chủ quan, cá biệt và không đồng nhất. Với thời gian và khả năng có hạn, chắc chắn giáo trình khó tránh khỏi những hạn chế và thiếu sót. Bộ môn Toán và các tác giả rất mong nhận được sự đóng góp ý kiến của bạn đọc.

CÁC TÁC GIẢ

MỤC LỤC

PHỤ LỤC

Chương 1

XÁC SUẤT

Bài 1

TẦN SUẤT

MỤC TIÊU

1. Thực hiện được ba phép toán tập hợp (phép hợp, phép giao, phép trừ).

2. Tính được số lượng mẫu chỉnh hợp lặp, chỉnh hợp không lặp, tổ hợp không lặp và tổ hợp lặp.

3. Tính được tần suất của hiện tượng và nêu được ý nghĩa.

1. TẬP HỢP

1.1. Khái niệm tập hợp

▪ Mọi người thường nói tập hợp bàn ghế, tập hợp số, tập hợp thầy thuốc, tập hợp bệnh nhân v.v...

Tập hợp là khái niệm chưa xác định vì vậy để hiểu và thực hiện các phép toán với tập hợp thường thông qua cách cho một tập hợp. Khi đó tập hợp được xác định.

Có hai cách cho tập hợp: Hoặc cho danh sách các phần tử của tập hợp hoặc cho các đặc tính, tính chất để xác định một phần tử thuộc tập hợp.

Thường ký hiệu các chữ A, B, C, ... để chỉ tập hợp, các chữ x, y, z,... để chỉ phần tử của tập hợp.

A_1 = {Danh sách (tổ viên) tổ 1},

A_2 = {Danh sách lớp Y_1},

A = {x thực : thoả mãn tính chất Q(x)}.

Phần tử x thuộc A viết là $x \in A$. Phần tử x không thuộc B viết là $x \notin B$ hoặc $x \overline{\in} B$.

▪ Tập hợp trống là tập hợp không chứa một phần tử nào. Thường ký hiệu tập hợp trống là ϕ.

Ví dụ: $A = \{x \text{ thực} : x^2 + 1 = 0\}$,

B = {Bác sỹ chuyên mổ tim ở bệnh viện huyện},

C = {Bệnh nhân "Đao" trên 50 tuổi}.

A, B, C là các tập hợp trống.

▪ Tập hợp con

A là tập hợp con của B nếu mọi phần tử $x \in A$ đều là các phần tử $x \in B$.

Ký hiệu: $A \subseteq B$, đọc là A bao hàm trong B hoặc $B \supseteq A$, đọc là B bao hàm A hoặc B chứa A.

Tổ là tập hợp con của lớp. lớp là tập hợp con của khối.

Tập hợp bệnh nhân trong khoa bao hàm trong tập hợp bệnh nhân toàn viện.

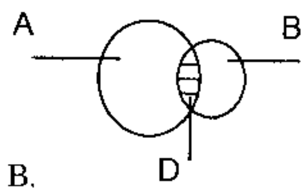
▪ Tập hợp bằng nhau.

Cho hai tập hợp A và B. Nếu mọi phần tử của A là những phần tử của B và ngược lại mọi phần tử của B cũng là những phần tử của A thì A = B.

Để chứng tỏ điều này cần chứng minh $A \subseteq B$ và $B \subseteq A$.

1.2. Phép toán tập hợp

▪ Phép giao



Cho A, B, C. Ký hiệu dấu $\cap$ đọc là giao.

Giao của hai tập hợp $A \cap B = D$

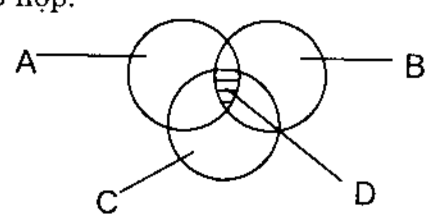
D là tập hợp có các phần tử vừa thuộc A vừa thuộc B.

Giao của ba tập hợp $A \cap B \cap C = D$

D là tập hợp có các phần tử vừa thuộc A vừa thuộc B vừa thuộc C.

Chú ý: Phép giao có thể mở rộng cho nhiều tập hợp.

Thường viết $A \cap B$ hoặc viết tắt là AB.



▪ Phép hợp

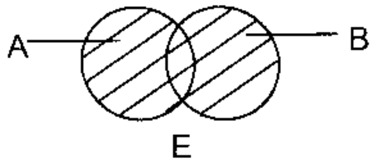
Cho A, B, C. Ký hiệu dấu $\cup$ đọc là hợp.

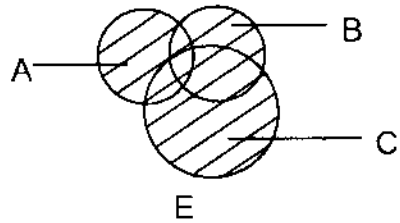
Hợp của hai tập hợp $A \cup B = E$

E là tập hợp có các phần tử hoặc thuộc A hoặc thuộc B hoặc thuộc A và B hay E là tập hợp có các phần tử thuộc ít nhất một trong hai tập hợp A, B.

Hợp của ba tập hợp $A \cup B \cup C = E$

E là tập hợp có các phần tử thuộc ít nhất một trong ba tập hợp A, B, C.

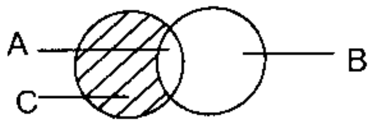




- Phép trừ

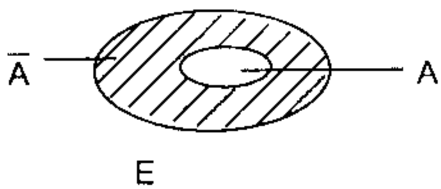
Cho A, B. Ký hiệu $A \setminus B$ đọc là A trừ B hay hiệu của A và B.

$A \setminus B = C$. C là tập hợp có các phần tử chỉ thuộc A mà không thuộc B



Cho $A \subset E$. $E \setminus A = C_E A = \overline{A}$

$C_E A$ được gọi là phần bù của A trong E hay $\overline{A}$



- Một số tính chất

$A \cap B = B \cap A$, $A \cap A = A$, $A \cap \phi = \phi$ vì $\phi \subset A$

$A \cup B = B \cup A$, $A \cup A = A$, $A \cup \phi = A$

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

$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.

1.3. Các khái niệm khác

- Tích *Đecart (R. Đecart)*

Cho A = (x, y, z), B = (1, 2, 3).

Tích *Đecart* của A và B viết là $A \times B$.

$A \times B = \{ (x, 1), (x, 2), ..., (z, 3) \}$.

Tích *Đecart* của A và B là một tập hợp mà mỗi phần tử là một cặp sắp thứ tự, phần tử thứ nhất thuộc A, phần tử thứ hai thuộc B.

Như vậy, một điểm trong mặt phẳng 0xy là một phần tử của tập hợp tích $R \times R$. $M(x, y) \in R \times R = R^2$. Một điểm trong không gian ba chiều 0xyz là một phần tử thuộc tập hợp tích *Đecart* $R \times R \times R$

$M(x, y, z) \in R \times R \times R = R^3$

- Sự phân hoạch một tập hợp

Cho E. Chia E thành $E_1, E_2, ..., E_n$ sao cho thoả mãn các tính chất:

$$E_i \neq \phi \ \forall_i = \overline{1, n}, \ E_i \cap E_j = \phi \ \forall i \neq j = \overline{1, n}, \ \bigcup_{i=1}^{n} E_i = E$$

được gọi là phân hoạch tập hợp E.

Thực chất sự phân hoạch là việc chia sao cho mỗi phần tử của E chỉ thuộc về duy nhất một tập hợp E_i mà thôi.

Chia một lớp thành 4 tổ hoặc chia bệnh nhân về các khoa là phân hoạch tập hợp.

2. CÔNG THỨC ĐẾM CÁC MẪU (GIẢI TÍCH TỔ HỢP)

Cho $A = (x_1, x_2,.., x_n)$

Có bao nhiêu cách lấy k phần tử từ A ? Số cách lấy hay số mẫu phụ thuộc vào tính chất của mẫu.

Mẫu lặp là mẫu có phần tử xuất hiện trong mẫu trên một lần, mẫu không lặp là mẫu có mỗi phần tử trong mẫu chỉ xuất hiện một lần.

Khi thay đổi thứ tự các phần tử trong mẫu mà được mẫu mới thì đó là mẫu có thứ tự, nếu vẫn là mẫu cũ thì đó là mẫu không thứ tự. Hay nói cách khác, mẫu có thứ tự là mẫu phụ thuộc thứ tự các phần tử trong mẫu, ngược lại là mẫu không thứ tự.

2.1. Chỉnh hợp lặp

- Định nghĩa

Cho $A = (x_1, x_2,.., x_n)$. *Chỉnh hợp lặp là mẫu k phần tử có lặp, có thứ tự lấy từ n phần tử của A.*

- Công thức đếm

Gọi số cách lấy mẫu hay số lượng mẫu chỉnh hợp lặp là F_n^k

Công thức tính: $F_n^k = n^k$. Công thức vẫn đúng khi k > n.

Một số tự nhiên có 3 chữ số là một mẫu có lặp, có thứ tự xây dựng từ các chữ số 0, 1, ..., 9.

Số mẫu = $9.F_{10}^2 = 9 \times 10^2 = 900$

Xếp tuỳ ý 5 bệnh nhân vào 3 khoa là một mẫu có lặp, có thứ tự xây dựng từ 3 khoa. Số mẫu = $F_3^5 = 3^5 = 243$.

2.2. Chỉnh hợp không lặp

- Định nghĩa

Cho A = (x_1, x_2,.., x_n). Chỉnh hợp không lặp là mẫu k phần tử không lặp, có thứ tự lấy từ n phần tử của A.

- Công thức đếm

Gọi số cách lấy mẫu chỉnh hợp không lặp là A_n^k

Công thức tính : $A_n^k = n(n-1)...(n-k+1)$.

Ký hiệu: n! = 1. 2. 3... n và quy ước 1! = 1, 0! = 1.

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

Công thức đúng khi $k \leq n$.

Một số tự nhiên có 3 chữ số khác nhau là một mẫu không lặp, có thứ tự xây dựng từ 10 số 0, 1, ..., 9. Số mẫu = $9 \times A_9^2 = 9 \times 9 \times 8 = 648$.

Xếp 3 bệnh nhân vào 5 khoa sao cho có nhiều nhất một người trong khoa là mẫu gồm 3 khoa không lặp, có thứ tự xây dựng từ 5 khoa. Số mẫu = $A_5^3 = 5 \times 4 \times 3 = 60$.

- Hoán vị: cho A = (x_1, x_2,.., x_k), mỗi cách sắp xếp k phần tử là một hoán vị.

$x_1\ x_2\ x_3 \ldots x_k$ và $x_2\ x_1\ x_3 \ldots x_k$ là hai hoán vị khác nhau.

Vậy hoán vị là mẫu k phần tử không lặp, có thứ tự lấy từ k phần tử.

Gọi số hoán vị là P_k ta có công thức tính: $P_k = k!$

Nhận xét : Chỉnh hợp lặp và chỉnh hợp không lặp là những mẫu có thứ tự.

2.3. Tổ hợp không lặp

- Định nghĩa

Cho A = (x_1, x_2,..., x_n). Tổ hợp không lặp là mẫu k phần tử không lặp, không thứ tự lấy từ n phần tử của A.

- Công thức đếm

Gọi số cách lấy mẫu tổ hợp không lặp là C_n^k. Do tổ hợp không lặp là mẫu không thứ tự của k phân tử lấy ra cho nên nhân số tổ hợp không lặp với k! sẽ được số chỉnh hợp không lặp.

Công thức: $$C_n^k = \frac{A_n^k}{k!} = \frac{n!}{(n-k)!\,k!}, \qquad (k \leq n)$$

Nhận xét : $C_n^k = C_n^{n-k}$

– Chọn 5 chấp hành chi đoàn trong số 8 ứng cử và đề cử là lấy mẫu không lặp, không thứ tự

Số cách chọn : $$C_8^5 = \frac{8!}{(8-5)!\,5!} = 56.$$

– Gia đình 3 con trong đó có 2 gái là mẫu không lặp, không thứ tự, lấy 2 gái trong số 3 gái. Số loại gia đình: $C_3^2 = 3$.

Lập luận tương tự theo số con trai cũng được kết quả trên.

2.4. Tổ hợp lặp

- Định nghĩa

Cho A = (x_1, x_2,..., x_n). Tổ hợp lặp là mẫu k phần tử có lặp, không thứ tự lấy từ n phần tử của A.

- Công thức đếm

Nếu mẫu lặp k phần tử thì chỉ thêm k –1 phần tử lặp vào A dẫn đến cách lấy mẫu k phần tử không lặp, không thứ tự từ n + k – 1 phần tử.

Công thức tính: $$C_{n+k-1}^k = \frac{(n+k-1)\,!}{(n-1)!\,k!}$$

Khi k > n công thức cũng đúng.

– Đơn thức bậc 5 lặp từ a và b là mẫu có lặp, không thứ tự.

Số đơn thức là: $$C_{2+5-1}^5 = \frac{6!}{1!\,5!} = 6$$

– Gia đình 4 con là mẫu có lặp, không thứ tự lặp từ hai phần tử T (trai), G (gái).

Số mẫu là: $$C_{2+4-1}^4 = \frac{5!}{1!\,4!} = 5$$

Nhận xét: Mẫu tổ hợp không lặp và mẫu tổ hợp lặp là những mẫu không thứ tự.

Sau đây xét một ví dụ tổng quát các loại mẫu.

Ví dụ: Cho A = (1, 2, 3, 4).

a) Có bao nhiêu số tự nhiên có 3 chữ số lập từ 4 số đã cho ?

b) Có bao nhiêu số tự nhiên có 3 chữ số khác nhau lập từ 4 số đã cho?

c) Có bao nhiêu nhóm có 3 chữ số khác nhau lập từ 4 số đã cho ?

d) Có bao nhiêu nhóm có 3 chữ số lập từ 4 số đã cho ?

Giải:

a) Số tự nhiên có 3 chữ số là mẫu có lặp, có thứ tự lập từ 4 số.

Số mẫu bằng F_4^3: $F_4^3 = 4^3 = 64$

b) Số tự nhiên có 3 chữ số khác nhau là mẫu không lặp, có thứ tự lập từ 4 số.

Số mẫu bằng A_4^3: $A_4^3 = \dfrac{4!}{(4-3)!} = 24$

c) Nhóm có 3 chữ số khác nhau là mẫu không lặp, không thứ tự lập từ 4 số.

Số mẫu bằng C_4^3: $C_4^3 = \dfrac{4!}{1!\ 3!} = 4$

d) Nhóm có 3 chữ số là mẫu có lặp, không thứ tự lập từ 4 số.

Số mẫu bằng C_{4+3-1}^3: $C_{4+3-1}^3 = \dfrac{6!}{3!\ 3!} = 20$.

Nhận xét: $F_4^3 - A_4^3 = 40$. Đó là các mẫu có lặp thật sự và có thứ tự.

2.5. Khai triển nhị thức Newton

$$\begin{aligned}(a+b)^n &= C_n^0 a^0 b^n + C_n^1 a^1 b^{n-1} + \ldots + C_n^n a^n b^0 \\ &= \sum_{k=0}^{n} C_n^k a^k b^{n-k}\end{aligned}$$

Đổi vai trò a cho b công thức cũng đúng.

Lấy a = b = 1, có công thức

$$2^n = C_n^0 + C_n^1 + \ldots + C_n^n$$

Cho p + q = 1, có công thức :

$$1 = (p+q)^n = \sum_{k=0}^{n} C_n^k p^k q^{n-k}$$

3. TẦN SUẤT

3.1. Các khái niệm

Để hiểu và thực hiện các phép toán đối với tần suất cũng như xác suất sau này, cần xây dựng một số khái niệm.

- Phép thử là nhóm điều kiện có thể lặp đi lặp lại nhiều lần trong cùng điều kiện. Thường ký hiệu phép thử bởi các chữ ε, α, β .... Khi nghiên cứu, đo chiều cao, làm xét nghiệm, chẩn đoán bệnh, điều trị bệnh ... là các phép thử.

▪ Hiện tượng hay biến cố là kết quả của một phép thử. Các hiện tượng được ký hiệu bởi các chữ A, B, C ... Xét nghiệm dương tính: A, chẩn đoán có bệnh: B, điều trị khỏi: K, là các hiện tượng hay gặp trong y.

Khi thực hiện các phép thử nhiều lần, số lần xuất hiện của một hiện tượng được gọi là tần số xuất hiện. Tần số ký hiệu bởi m.

▪ Khi nghiên cứu một đối tượng, không nghiên cứu mọi mặt mà chỉ nghiên cứu một số đặc tính hay tính chất nào đó. Dấu hiệu nghiên cứu là đặc tính hay tính chất cần nghiên cứu. Có thể chia dấu hiệu nghiên cứu ra làm hai loại: dấu hiệu về chất và dấu hiệu về lượng. Dấu hiệu về chất được nghiên cứu khả năng xuất hiện, còn dấu hiệu về lượng được hướng tới việc tính các tham số mẫu. Dựa vào khả năng xuất hiện chia các hiện tượng thành 3 loại.

▪ Hiện tượng ngẫu nhiên là hiện tượng không biết trước có xảy ra hay không khi thực hiện phép thử. Sự xuất hiện của hiện tượng ngẫu nhiên phụ thuộc vào nhiều yếu tố mà không có yếu tố chủ yếu quyết định sự xuất hiện đó. Ký hiệu các chữ A, B, C ... để chỉ các hiện tượng ngẫu nhiên.

▪ Hiện tượng chắc chắn xuất hiện là hiện tượng luôn luôn xảy ra khi thực hiện phép thử. Ký hiệu Ω để chỉ hiện tượng chắc chắn xảy ra.

▪ Hiện tượng trống, ký hiệu là ϕ, là hiện tượng nhất định không xảy ra khi thực hiện phép thử.

Khám bệnh cho một người có khi người đó bị bệnh, có khi không bị bệnh ; chữa bệnh có khi chắc chắn khỏi, có khi không bao giờ khỏi.

Giữa các hiện tượng có thể phụ thuộc nhau hay không phụ thuộc nhau.

▪ Hiện tượng A xung khắc với hiện tượng B nếu như A và B không đồng thời xuất hiện.

Khi đó $A \cap B = \phi$ tương đương với A và B xung khắc với nhau.

▪ $E_1, E_2, \ldots, E_n$ được gọi là nhóm đầy đủ các hiện tượng nếu: $E_i \neq \phi \ \forall i = \overline{1,n}$,

$E_i \cap E_j = \phi \ \forall i \neq j = \overline{1,n}$, $\bigcup\limits_{i=1}^{n} E_i = W$.

Như vậy khi phân hoạch Ω thành $E_1, E_2, \ldots, E_n$ sẽ được nhóm đầy đủ các hiện tượng.

Khi A, B lập thành nhóm đầy đủ hai hiện tượng thì A, B được gọi là 2 hiện tượng đối lập nhau. Khi đó B được ký hiệu là $\overline{A}$ và viết là A, $\overline{A}$.

▪ Hai hiện tượng A và B được gọi là độc lập với nhau nếu A xuất hiện hay không xuất hiện cũng không ảnh hưởng đến B xuất hiện hay không xuất hiện và ngược lại.

Hai hiện tượng xung khắc với nhau thì không độc lập với nhau. Cũng như vậy hai hiện tượng độc lập với nhau thì không xung khắc với nhau.

Chữa bệnh khỏi hoặc không khỏi, chẩn đoán có bệnh hoặc không có bệnh, sinh con trai hoặc sinh con gái là các cặp hiện tượng đối lập nhau. Ngày nay không thể dựa vào lần này sinh con trai thì suy ra lần sau sẽ sinh con trai hoặc gái. Như vậy sinh con trai hay gái giữa các lần sinh khác nhau độc lập với nhau.

3.2. Tần suất

- Định nghĩa

Thực hiện phép thử ε n lần độc lập, hiện tượng A xuất hiện m lần. Ký hiệu ω(A) là tần suất xuất hiện A.

$\omega(A) = \frac{m}{n}$. Tần suất là tỷ lệ giữa số lần xuất hiện A và số lần thực hiện phép thử.

ω là đại lượng không có đơn vị, được viết dưới dạng % hay $^{0}/_{00}$.

$0 \le \omega(A) \le 1$, ω(A) cho biết khả năng xuất hiện của A khi thực hiện phép thử một lần

$\omega(\phi) = 0$. Khi $\omega(A) = 0$ chưa chắc $A = \phi$,

$\omega(\Omega) = 1$. Khi $\omega(B) = 1$ chưa chắc $B = \Omega$.

- Tính chất

Khi n thay đổi, m thay đổi thì ω thay đổi. Khi n đủ lớn, ω thay đổi ít. Tính thay đổi ít của ω khi n lớn được gọi là tính ổn định của ω.

Buffon tung đồng xu 4040 lần thấy ω(s) = 50,79%,

Pearson tung đồng xu 12000 lần thấy ω(s) = 50,16%,

Pearson tung đồng xu 24.000 lần thấy ω(s) = 50,05%,

trong đó s ký hiệu là hiện tượng mặt sấp đồng xu xuất hiện.

$\omega(A) \ge 0,95$: A hầu như chắc chắn xuất hiện khi thực hiện phép thử

$\omega(B) \le 0,05$: B hầu như chắc chắn không xuất hiện khi thực hiện phép thử.

Đó là các quyết định dựa vào mong muốn càng đúng nhiều càng tốt và càng sai ít càng tốt mà không phải là các nguyên lý hay định lý luôn luôn đúng.

Bệnh nhân đến khám sớm (khi chưa có triệu chứng đặc hữu) được chữa theo bệnh hay gặp nhất ở thời gian đó.

Bệnh nhân bị bỏng trên 70% diện tích da, từ độ II trở lên có tỷ lệ tử vong cao song vẫn được cứu chữa tích cực với hy vọng cứu được một người trong số rất nhiều người không cứu được.

- Các phản ví dụ

Nồng độ pha loãng của dịch ($^0/_{00}$) không là tần suất.

$\frac{\text{số trẻ chết}}{\text{1000 trẻ sống}}$: không là tần suất.

Tỷ lệ tiêm chủng mở rộng:

Tỉnh A đạt 99,8% : là tần suất.

Tỉnh B đạt 101% : không là tần suất.

Tỉnh C đạt 102% : không là tần suất.

$\frac{\text{Chiều cao ngồi}}{\text{Chiều cao đứng}}$: không là tần suất.

CÂU HỎI TỰ LƯỢNG GIÁ

Mỗi bài lượng giá gồm 4 câu. Làm bài trong 30 phút.

Mỗi câu chỉ chọn một kết quả đúng.

Đúng 4 câu: Giỏi (10 điểm), Đúng 3 câu: Khá (7 điểm),

Đúng 2 câu: Đạt (5 điểm), Đúng 1 câu: Không đạt (3 điểm).

Không đúng câu nào: Kém (0 điểm).

Hãy chọn một kết quả đúng:

1. Khoa nội có 6 bác sỹ nữ, 4 bác sỹ nam. Khoa ngoại có 8 bác sỹ nam. Lập tổ công tác 3 người cần có nam, có nữ, có nội khoa, có ngoại khoa. Hỏi có bao nhiêu cách?

Kết quả:

A. 576 B. 480 C. 816 D. 360 E. số khác.

2. Một tổ sinh viên có 8 nam, 7 nữ. Chia thành 3 nhóm trực đồng thời tại 3 bệnh viện A, B, C. Hỏi có bao nhiêu cách phân công nếu: bệnh viện A cần 3 nam 2 nữ, bệnh viện B cần 5 người trong đó có ít nhất 4 nam, số còn lại đến bệnh viện C ?

Kết quả:

A. 30576 B. 61152 C. 29400 D. 1176 E. số khác.

3. Có 4 thuốc loại I và 3 thuốc loại II. Hỏi có bao nhiêu cách điều trị cho 5 người bị bệnh A, nếu mỗi người bị bệnh A cần 2 thuốc loại I và 1 thuốc loại II ?

Kết quả:

A. 45 B. 59.049 C. 90 D. 1.889.568 E. số khác.

4. Cho ngẫu nhiên đồng thời 6 kháng thể vào 6 kháng nguyên (khi chưa ghi nhãn) để tìm các kháng thể, kháng nguyên cùng cặp. Giả sử không có ngưng kết chéo, hỏi có bao nhiêu trường hợp xảy ra nếu chỉ có 1 cặp ngưng kết ?

Kết quả:

A. 135 B. 265 C. 264 D. 455 E. số khác.

Bài 2

XÁC SUẤT

MỤC TIÊU

1. Trình bày được định nghĩa đồng khả năng và định nghĩa thống kê của xác suất.

2. Trình bày được các công thức nhân xác suất, cộng xác suất, xác suất toàn phần và xác suất Bayes.

3. Giải được một số bài toán xác suất trong y dựa vào các công thức xác suất nêu trên.

Trước khi thực hiện phép thử, đoán xem một hiện tượng ngẫu nhiên nào đó có xảy ra hay không là một việc rất khó khăn. Khi thực hiện phép thử nhiều lần, biết khả năng xuất hiện của hiện tượng, từ đó đoán sự xuất hiện của hiện tượng dễ dàng hơn.

Khả năng xuất hiện hiện tượng A là xác suất xuất hiện A, ký hiệu là P(A), là hằng số p nằm giữa 0 và 1, tồn tại một cách khách quan, không phụ thuộc vào ý muốn chủ quan của con người.

1. ĐỊNH NGHĨA

1.1. Định nghĩa đồng khả năng

Giả sử có một bình cầu chứa n quả cầu hoàn toàn giống nhau. Trong n quả cầu có m quả có dấu. Xáo trộn đều các quả cầu trong bình và lấy ngẫu nhiên một quả. Gọi A là hiện tượng lấy được quả có dấu.

Xác suất xuất hiện hiện tượng A là tỷ lệ giữa số trường hợp thuận lợi cho A và tổng số các trường hợp có thể xảy ra

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

Xác suất đúng khi các quả cầu có cùng khả năng được lấy. Vì vậy định nghĩa trên được gọi là định nghĩa đồng khả năng.

Cần chú ý là các công thức tính xác suất được xây dựng trên cơ sở đồng khả năng. Xác suất tính được sẽ đúng đắn, chính xác chỉ khi điều kiện trên thoả mãn.

1.2. Định nghĩa thống kê

Thực hiện phép thử ε n lần độc lập, hiện tượng A xuất hiện m lần

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

Khi n đủ lớn, ω (A) ổn định, xác suất chính là giá trị ổn định của tần suất. Lấy tần suất gán cho xác suất được gọi là ước lượng điểm của xác suất. Ước lượng xác suất bằng tần suất giúp cho việc sử dụng rất thuận tiện nhưng có thể sai sót.

Giữa xác suất, hằng số xác định và tần suất có sự khác biệt, đó chính là sai số δ_1.

$$|P(A) - \omega(A)| \le \delta_1 \qquad \text{với } \delta_1 = t(\alpha/2)\sqrt{\frac{\omega(1-\omega)}{n}},$$

trong đó $t(\alpha/2)$ phụ thuộc vào α được tra trong bảng chuẩn tắc (bảng 1), n là số lần thực hiện phép thử, $t(0{,}05/2) = 1{,}96$.

Dẫn đến: $\omega - \delta_1 \le P(A) \le \omega + \delta_1$, $\omega \pm \delta_1$ được gọi là khoảng tin cậy mức $1 - \alpha$ của P(A). Khi α bé, mức tin cậy cao song khoảng ước lượng lớn không thuận tiện cho việc sử dụng. Nên chọn α phù hợp với bài toán thực tiễn.

Ví dụ:

1. Khám 7534 trẻ từ 5 – 15 tuổi thấy 19 trẻ bị thấp tim. Hãy đánh giá tỷ lệ thấp tim.

Gọi A là hiện tượng thấp tim

Ước lượng điểm: $P(A) \approx \omega(A) = \dfrac{19}{7534} = 0{,}0025.$

$$\delta_1 = 1{,}96\sqrt{\frac{0{,}0025 \times 0{,}9975}{7534}} = 0{,}0011, \text{ lấy } \alpha = 0{,}05.$$

Ước lượng khoảng:

$$P(A) \in \omega \pm \delta_1 \Rightarrow 0{,}0014 \le P(A) \le 0{,}0036$$

Như vậy tỷ lệ thấp tim ít nhất là $1{,}4\,^0/_{00}$, nhiều nhất là $3{,}6\,^0/_{00}$.

2. Điều tra năm 1989 tại một địa phương thấy 48,53% trẻ bị sâu răng. Điều trị và súc họng bằng Fluo 0,2% trong 8 năm, điều tra lại 1250 trẻ ban đầu thấy 181 trẻ sâu răng.

Hãy đánh giá tỷ lệ trẻ sâu răng sau 8 năm điều trị và súc họng.

Gọi A là hiện tượng trẻ sâu răng

Ước lượng điểm: $P(A) \approx \omega(A) = \frac{181}{1250} = 0,1448.$

$$\delta_1 = 1,96\sqrt{\frac{0,1448 \times 0,8552}{1250}} = 0,0195 \text{, lấy } \alpha = 0,05.$$

Ước lượng khoảng:

$$P(A) \in \omega \pm \delta_1 \Rightarrow 0,1253 \le P(A) \le 0,1643.$$

Sau 8 năm điều trị và phòng bệnh, tỷ lệ sâu răng ít nhất là 12,53%, nhiều nhất là 16,43%.

2. CÔNG THỨC TÍNH XÁC SUẤT

2.1. $P(\Omega) = 1, \quad P(\phi) = 0$

2.2. Công thức nhân xác suất

- Xác suất có điều kiện

Trong các công thức tính xác suất, thường gặp cách viết :

P (A/B), P(B/A), P(A/BC).

P (A/B) là xác suất xuất hiện hiện tượng A với điều kiện hiện tượng B đã xảy ra.

P (B/A) là xác suất xuất hiện hiện tượng B với điều kiện hiện tượng A đã xảy ra.

P (A/BC) là xác suất xuất hiện hiện tượng A với điều kiện hiện tượng B và C đã xảy ra.

Các xác suất trên được gọi là các xác suất có điều kiện.

Trong đám đông thường cho tỷ lệ bị bệnh nói chung của cả nam và nữ, đó là xác suất không điều kiện, còn tỷ lệ bị bệnh của riêng nam, tỷ lệ bị bệnh của riêng nữ là các xác suất có điều kiện.

Làm xét nghiệm chẩn đoán bệnh sẽ thu được tỷ lệ dương tính của nhóm bị bệnh và tỷ lệ âm tính của nhóm không bị bệnh. Đó là các xác suất có điều kiện. Nếu không phân biệt bị bệnh hay không bị bệnh ta có các xác suất dương tính của cả bị bệnh và không bị bệnh, xác suất âm tính của cả bị bệnh và không bị bệnh của xét nghiệm. Chúng là các xác suất không điều kiện.

- A, B, C là các hiện tượng không độc lập

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

$$P(A \cap B \cap C) = P(ABC) = P(A)\,P(B/A)\,P(C/AB) = \ldots$$

$$= P(ACB) = P(A)\,P(C/A)\,P(B/AC)$$

Có thể mở rộng công thức cho nhiều hiện tượng.

Thật vậy, từ một nghiên cứu với 2 phép thử α và β, thu được kết quả sau:

$\beta \backslash \alpha$	A	$\overline{A}$	m_{io}
B	m_{11}	m_{12}	m_{10}
$\overline{B}$	m_{21}	m_{22}	m_{20}
m_{oj}	m_{01}	m_{02}	n

$$P(AB) = \frac{m_{11}}{n}$$

$$P(A)P(B/A) = \frac{m_{01}}{n} \cdot \frac{m_{11}}{m_{01}} = \frac{m_{11}}{n}$$

$$P(B)P\left(A/B\right) = \frac{m_{10}}{n} \cdot \frac{m_{11}}{m_{10}} = \frac{m_{11}}{n}.$$

điều đó chứng tỏ $P(AB) = P(A)P(B/A) = P(B)P(A/B)$

- A, B, C là các hiện tượng độc lập

$$P(A \cap B) = P(AB) = P(A)P(B).$$

$$P(A \cap B \cap C) = P(ABC) = P(A)P(B)P(C).$$

Do các hiện tượng độc lập dẫn đến:

$$P\left(A/B\right) = P(A), \quad P\left(B/A\right) = P(B), \quad P\left(A/BC\right) = P(A).$$

Có thể nói khi các hiện tượng độc lập thì xác suất của giao các hiện tượng bằng tích các xác suất của từng hiện tượng.

2.3. Công thức cộng xác suất

- A, B, C là các hiện tượng ngẫu nhiên

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

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)$$

Nhận xét: Số hiện tượng lẻ thì xác suất có dấu +,

Số hiện tượng chẵn thì xác suất có dấu –.

Dựa vào nhận xét, có thể mở rộng công thức cho n hiện tượng.

- A, B, C xung khắc từng đôi

$$P(A \cup B) = P(A+B) = P(A) + P(B),$$

$$P(A \cup B \cup C) = P(A+B + C) = P(A) + P(B) + P(C).$$

Do các hiện tượng xung khắc từng đôi nên:

$$P(AB) = P(AC) = P(BC) = P(\phi) = 0 \qquad P(ABC) = P(\phi.C) = P(\phi) = 0.$$

Có thể nói khi các hiện tượng xung khắc từng đôi thì xác suất của tổng các hiện tượng bằng tổng các xác suất của từng hiện tượng.

- A, $\overline{A}$ hai hiện tượng đối lập

$$P(\Omega) = P(A + \overline{A}) = P(A) + P(\overline{A}) = 1 \Leftrightarrow P(\overline{A}) = 1 - P(A).$$

Ví dụ:

1. Tại một địa phương có 5000 người, điều tra thấy 510 người bị sốt rét. Trong số sốt rét có 15 người sốt rét ác tính. Trong số sốt rét ác tính có 5 người chết.

a) Tìm tỷ lệ sốt rét thường.

b) Tìm tỷ lệ chết của sốt rét ác tính.

Giải:

Gọi T là sốt rét thường.

A là sốt rét ác tính

C là chết

a) $P(T) = \dfrac{510-15}{5000} = 0,099$

b) $P(C/A) = \dfrac{5}{15} = 0,333$.

Cần phân biệt với $P(C) = \dfrac{5}{5000} = 0,001$.

$$P(C/S) = \frac{5}{510} = 0,0098$$

trong đó S là sốt rét nói chung.

2. Xác suất sinh con trai bằng 0,514.

a) Tìm xác suất sinh bằng được con trai ở lần sinh thứ 4.

b) Tìm xác suất sinh được 3 con đều là gái.

c) Tìm xác suất sinh được 3 con có ít nhất một gái.

Giải:

Gọi T_i là sinh con trai ở lần i.

G_i là sinh con gái ở lần i.

A_4 là sinh bằng được trai ở lần 4.

B là sinh được 3 con gái.

C là sinh được 3 con có ít nhất một gái.

a) $P(A_4) = P(G_1\, G_2\, G_3\, T_4) = P(G_1)\, P(G_2)\, P(G_3)\, P(T_4)$

$= 0{,}486^3 \text{ x } 0{,}514 = 0{,}059.$

b) $P(B) = P(G_1\, G_2\, G_3) = P(G_1)\, P(G_2)\, P(G_3)$

$= 0{,}486^3 = 0{,}115.$

c) $P(C) = P(G_1 \cup G_2 \cup G_3) = p_1 + p_2 + p_3$

$= 1 - p_0 = 1 - P(T_1\, T_2\, T_3)$

$= 1 - P(T_1)\, P(T_2)\, P(T_3) = 1 - 0{,}514^3 = 0{,}864,$

trong đó p_i là xác suất sinh 3 con có i là gái.

3. Trong một hộp thuốc cấp cứu có 100 ống thuốc tiêm, trong đó có 10 ống Atropin. Lấy ngẫu nhiên lần lượt 3 ống thuốc.Tìm xác suất sao cho lấy được:

a) 3 ống Atropin.

b) 2 ống Atropin.

Giải:

Gọi A_i là lấy được ống Atropin ở lần i.

A là lấy được 3 ống Atropin.

B là lấy 3 ống được 2 ống Atropin.

a) $P(A) = P(A_1\, A_2\, A_3) = P(A_1)\; P(A_2 / A_1)\; P(A_3 / A_1\, A_2)$

$$= \frac{10}{100} \times \frac{9}{99} \times \frac{8}{98} = 0{,}0007.$$

b) $P(B) = P(A_1\, A_2\, \overline{A}_3 + A_1\, \overline{A}_2\, A_3 + \overline{A}_1\, A_2\, A_3)$

$= P(A_1\, A_2\, \overline{A}_3) + P\,(A_1\, \overline{A}_2\, A_3) + P\,(\overline{A}_1\, A_2\, A_3)$

$= P(A_1) \times P(A_2 / A_1) \times P(\overline{A}_3 / A_1 A_2) +$
$P(A_1) \times P(\overline{A}_2 / A_1) \times P(A_3 / A_1 \overline{A}_2) + \ldots$

$$= \frac{10}{100} \times \frac{9}{99} \times \frac{90}{98} + \frac{10}{100} \times \frac{90}{99} \times \frac{9}{98} + \frac{90}{100} \times \frac{10}{99} \times \frac{9}{98} = 0{,}025.$$

Có thể tính cách khác. Lấy mẫu không lặp, không thứ tự là tổ hợp không lặp

$$P(A) = \frac{C_{10}^3}{C_{100}^3} = 0,0007, \qquad P(B) = \frac{C_{10}^2 \times C_{90}^1}{C_{100}^3} = 0,025$$

Nhận xét : P(A), P(B) rất nhỏ cho nên không được lấy thuốc ngẫu nhiên.

4. Ba bác sĩ độc lập nhau khám bệnh. Xác suất chẩn đoán sai của các bác sĩ tương ứng bằng 0,05, 0,1 và 0,15. Ba người đã khám cho một bệnh nhân. Tìm xác suất sao cho

a) Không ai chẩn đoán sai.

b) Không ai chẩn đoán đúng.

c) Ít nhất một người chẩn đoán đúng.

Giải:

Gọi A_i là bác sĩ thứ i chẩn đoán đúng.

A là không ai chẩn đoán sai ; B là không ai chẩn đoán đúng ; C là ít nhất một người chẩn đoán đúng.

a) $P(A) = P(A_1 A_2 A_3) = P(A_1)\ P(A_2)\ P(A_3) \quad = 0,95 \times 0,9 \times 0,85 = 0,72675.$

b) $P(B) = P(\bar{A}_1\ \bar{A}_2\ \bar{A}_3) = P(\bar{A}_1)\ P(\bar{A}_2)\ P(\bar{A}_3) \ = 0,05 \times 0,1 \times 0,15 = 0,00075.$

c) $P(C) = P(A_1 \cup A_2 \cup A_3) = p_1 + p_2 + p_3,$

trong đó p_i là xác suất có i người đúng.

$P(C) = 1 - P(\bar{A}_1\ \bar{A}_2\ \bar{A}_3) = 1 - 0,00075 = 0,99925.$

Nhận xét: Sau hội chẩn thường điều trị theo chẩn đoán của số quá bán các bác sĩ nếu trình độ các bác sĩ đồng đều. Ngược lại, sẽ điều trị theo chẩn đoán của người giỏi nhất.

5. Một bác sĩ có khả năng xác định đúng triệu chứng với xác suất 0,9. Khả năng chẩn đoán đúng bệnh với điều kiện đã xác định đúng triệu chứng bằng 0,8. Khi điều trị, mặc dù đã xác định đúng triệu chứng và chẩn đoán đúng bệnh, khả năng khỏi bằng 0,95.

Tìm xác suất không khỏi của người bệnh khi khám và điều trị bác sĩ trên.

Giải:

Gọi T là xác định đúng triệu chứng.

B là chẩn đoán đúng bệnh.

K là điều trị khỏi.

$P(T) = 0,9$ $\quad\quad$ $P(B/T) = 0,8$ $\quad\quad$ $P(K/TB) = 0,95$

$P(K) = P(TBK) = P(T)\,P(B/T)\,P(K/TB)$

$= 0,9 \times 0,8 \times 0,95 = 0,684$

$P(\overline{K}) = 1 - P(K) = 1 - 0,684 = 0,316.$

Chú ý: Trong thực tế lâm sàng có trường hợp chẩn đoán sai bệnh hoặc chẩn đoán không ra bệnh mà điều trị khỏi. Điều này nên quan niệm là rất hiếm gặp.

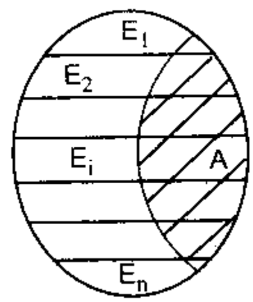
Có bác sĩ cho rằng chỉ có khả năng chẩn đoán đúng bệnh 95% các trường hợp nhưng đảm bảo rằng khả năng chữa khỏi các bệnh nhân đến khám và điều trị 99% các trường hợp. Điều này có đúng không ?

2.4. Công thức xác suất toàn phần

Giả sử A là một hiện tượng ngẫu nhiên nào đấy, khi tính P(A) theo phương pháp đồng khả năng nhưng không tính được. Cần xây dựng công thức tính.

Giả sử $E_1, E_2, \ldots, E_n$ là nhóm đầy đủ các hiện tượng, nghĩa là:

$$E_i \neq \phi \quad \forall i = \overline{1,n} \;,\quad E_i \cap E_j = \phi \quad \forall i \neq j = \overline{1,n} \;,\quad \bigcup_{i=1}^{n} E_i = \Omega .$$



Khi đó: $\quad A = A \cap \Omega = A \cap \left(\bigcup_{i=1}^{n} E_i\right) = \bigcup_{i=1}^{n} (A \cap E_i)$

Do đó: $\quad P(A) = P\left\{\bigcup_{i=1}^{n} (A \cap E_i)\right\} = \sum_{i=1}^{n} P(A \cap E_i)$

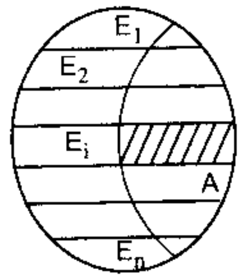
Vậy $\quad P(A) = \sum_{i=1}^{n} P(E_i)\,P(A / E_i)$

Công thức trên được gọi là công thức xác suất toàn phần.

Muốn tìm xác suất P(A) cần lấy tổng các xác suất từng phần của $A \cap E_i$, $i = \overline{1,n}$.

Công thức trên cũng được hiểu là xác suất đồng khả năng hoặc là xác suất trung bình có trọng lượng của các xác suất $P(A/E_i)$ với $i = \overline{1,n}$.

2.5. Công thức xác suất Bayes



$$P(A \cap E_i) = P(A).P(E_i / A) = P(E_i)\,P(A / E_i)$$

Nếu $P(A) \neq 0$, dẫn đến

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

Vậy $$P(E_i / A) = \frac{P(E_i)P(A / E_i)}{\sum_{i=1}^{n} P(E_i)P(A / E_i)}, \ i = \overline{1, n}.$$

Công thức trên do Bayes lập ra nên mang tên ông. Ngoài ra, do dạng của công thức nên cũng được gọi là công thức xác suất các giả thiết.

Từ công thức trên suy ra $\sum_{i=1}^{n} P(E_i / A) = 1$

Dẫn đến $P\left(\overline{A} / B\right) = 1 - P\left(A / B\right)$

$$P\left(\overline{B} / A\right) = 1 - P\left(B / A\right)$$

Chú ý: Do $\sum_{i=1}^{n} P(E_i / A) = 1$ nên:

$$P(A) = \sum_{i=1}^{n} P(A \cap E_i) = \sum_{i=1}^{n} P(A) \times P(E_i / A) = P(A)$$

Vậy không tính được P(A) theo phương pháp này.

Ví dụ:

6. Điều trị tương ứng phương pháp1, phương pháp 2, phương pháp 3 cho 5000, 3000 và 2000 bệnh nhân. Xác suất khỏi của các phương pháp tương ứng bằng 0,85; 0,9 và 0,95.

a) Tìm xác suất khỏi của ba phương pháp khi điều trị riêng rẽ từng phương pháp cho bệnh nhân.

b) Điều trị một trong ba phương pháp cho bệnh nhân đã khỏi, tìm tỷ lệ điều trị của từng phương pháp.

c) Tìm xác suất khỏi khi điều trị phối hợp ba phương pháp cho bệnh nhân.

Giải:

Gọi E_i là điều trị phương pháp thứ i cho bệnh nhân. $\overline{i = 1, 3}$.

A là điều trị khỏi.

Tổng số bệnh nhân điều trị ba phương pháp bằng 10.000 người.

$$P(E_1) = \frac{5000}{10.000} = 0,5 \qquad P(E_2) = \frac{3000}{10.000} = 0,3 \qquad P(E_3) = \frac{2000}{10.000} = 0,2$$

$$P(A/E_1) = 0,85 \qquad P(A/E_2) = 0,9 \qquad P(A/E_3) = 0,95.$$

a) $$P(A) = \sum_{i=1}^{3} P(E_i)\ P(A/E_i)$$

$$= 0,5 \times 0,85 + 0,3 \times 0,9 + 0,2 \times 0,95 = 0,885.$$

Có thể hiểu P(A) là xác suất đồng khả năng, là tỷ lệ giữa số người khỏi khi điều trị bởi ba phương pháp và tổng số người điều trị của ba phương pháp. Cũng có thể hiểu P(A) là xác suất trung bình có trọng lượng của các xác suất khỏi của từng phương pháp.

b) $$P(E_1/A) = \frac{P(E_1)P(A/E_1)}{P(A)} = \frac{0,5 \times 0,85}{0,885} = 0,48$$

$$P(E_2/A) = \frac{P(E_2)P(A/E_2)}{P(A)} = \frac{0,3 \times 0,9}{0,885} = 0,305$$

$$P(E_3/A) = \frac{P(E_3)P(A/E_3)}{P(A)} = \frac{0,2 \times 0,95}{0,885} = 0,215$$

Nhận xét: $\sum_{i=1}^{3} P\left(E_i/A\right) = 0,48 + 0,305 + 0,215 = 1$.

c) Đổi tên gọi các hiện tượng để tính toán thuận tiện hơn.

Gọi A_i là hiện tượng khỏi của phương pháp điều trị thứ i, $\overline{i = 1,3}$.

Điều trị phối hợp ba phương pháp thì một phương pháp điều trị khỏi hay hai phương pháp điều trị khỏi hay cả ba phương pháp điều trị khỏi, bệnh nhân sẽ khỏi. Hay nói cách khác bệnh nhân sẽ khỏi khi ít nhất một trong ba phương pháp điều trị khỏi.

Gọi F là hiện tượng khỏi khi điều trị phối hợp ba phương pháp.

$$P(F) = P(A_1 \cup A_2 \cup A_3) = p_1 + p_2 + p_3,$$

trong đó p_i là xác suất khỏi khi điều trị 3 phương pháp có i phương pháp khỏi

$$P(F) = 1 - P(\overline{A}_1 \overline{A}_2 \overline{A}_3) = 1 - P(\overline{A}_1)P(\overline{A}_2)P(\overline{A}_3)$$

$$= 1 - 0,15 \times 0,1 \times 0,05 = 0,99925.$$

7. Tỷ lệ bệnh B tại một địa phương bằng 0,02. Dùng một phản ứng giúp chẩn đoán, nếu người bị bệnh thì phản ứng dương tính 95%; nếu người không bị bệnh thì phản ứng dương tính 10%.

a) Tìm xác suất dương tính của phản ứng.

b) Một người làm phản ứng thấy dương tính, tìm xác suất sao cho đó là người bị bệnh.

c) Tìm xác suất chẩn đoán đúng của phản ứng.

Giải:

Gọi α là phép thử dương tính A hay âm tính $\overline{A}$

β là phép thử xác định có bệnh B hay không bệnh $\overline{B}$

ε là phép thử xác định đúng Đ hay sai S

Tổ chức y tế thế giới quy ước gọi:

$P(A/B)$ là độ nhạy.

$P(\overline{A}/\overline{B})$ là độ đặc hiệu.

$P(B/A)$ là giá trị của phản ứng dương tính.

$P(\overline{B}/\overline{A})$ là giá trị của phản ứng âm tính.

P(Đ) là giá trị của phản ứng.

$$P(Đ) = P(AB) + P(\overline{A}\,\overline{B})$$

$$= P(B)P(A/B) + P(\overline{B})P(\overline{A}/\overline{B})$$

$$= P(A)P(B/A) + P(\overline{A})P(\overline{B}/\overline{A})$$

Như vậy giá trị của phản ứng là giá trị trung bình của độ nhạy và độ đặc hiệu hoặc giá trị trung bình của giá trị dương tính và giá trị âm tính.

$P(B) = 0{,}02 \qquad P(A/B) = 0{,}95 \qquad P(A/\overline{B}) = 0{,}1$

a) $P(A) = P(B)P(A/B) + P(\overline{B})P(A/\overline{B})$

$= 0{,}02 \times 0{,}95 + 0{,}98 \times 0{,}1$

$= 0{,}117.$

b) $P(B/A) = \dfrac{P(B)P(A/B)}{P(A)} = \dfrac{0{,}02 \times 0{,}95}{0{,}117} = 0{,}162$

c) $P(Đ) = P(B)P(A/B) + P(\overline{B})P(\overline{A}/\overline{B})$

$= 0{,}02 \times 0{,}95 + 0{,}98 \times 0{,}9 = 0{,}901.$

8. Tại một địa phương tỷ lệ bị bệnh B bằng 0,05. Dùng một phản ứng giúp chẩn đoán, nếu phản ứng dương tính thì bị bệnh 20%; nếu phản ứng âm tính thì bị bệnh 1,25%.

a) Tìm xác suất dương tính của phản ứng.

b) Tìm độ nhạy, độ đặc hiệu của phản ứng.

c) Tìm xác suất sai của phản ứng.

Giải:

Ký hiệu các hiện tượng như ví dụ 7.

a) $P(B) = 0,05 \qquad P(B/A) = 0,2 \qquad P(B/\bar{A}) = 0,0125$

$$P(B) = P(A).P(B/A) + P(\bar{A})P(B/\bar{A})$$

$$= P(A).P(B/A) + [1 - P(A)].P(B/\bar{A})$$

$$0,05 = P(A) \times 0,2 + [1 - P(A)] \times 0,0125$$

$$P(A) = \frac{0,05 - 0,0125}{0,2 - 0,0125} = 0,2$$

b) $$P(A/B) = \frac{P(A)P(B/A)}{P(B)} = \frac{0,2 \times 0,2}{0,05} = 0,8$$

$$P(\bar{A}/\bar{B}) = \frac{P(\bar{A})P(\bar{B}/\bar{A})}{P(\bar{B})} = \frac{0,8 \times 0,9875}{0,95} = 0,832$$

c) $$P(S) = P(A\bar{B}) + P(\bar{A}B)$$

$$= P(A)P(\bar{B}/A) + P(\bar{A})P(B/\bar{A})$$

$$= 0,2 \times 0,8 + 0,8 \times 0,0125$$

$$= 0,17$$

Nhận xét:

- Từ công thức xác suất toàn phần của P(B), giải ngược lại sẽ tìm được P(A).
- Có thể tính P(S) dựa vào P(Đ).

Để giải các bài toán xác suất đỡ khó khăn, cần đọc kỹ đầu bài, đặt tên các hiện tượng và sử dụng công thức tính xác suất phù hợp với bài đã cho.

CÂU HỎI TỰ LƯỢNG GIÁ

Hãy chọn một kết quả đúng.

1. Điều trị 1 bệnh bởi phương pháp I, II, III, IV thấy tỷ lệ khỏi tương ứng bằng 0,6; 0,7; 0,8 và 0,85.

Điều trị cho 4 bệnh nhân, mỗi người một cách, tìm xác suất sao cho có từ 1 đến 3 người khỏi.

Kết quả:

A. 0,0486 B. 0,9964 C. 0,2892 D. 0,7108 E. số khác

2. Tỷ lệ điều trị phương pháp I, II, III, IV tương ứng bằng : 0,2; 0,25; 0,25; 0,3. Xác suất khỏi của các phương pháp tương ứng bằng : 0,75; 0,82; 0,84; 0,8. Một người điều trị một trong 4 phương pháp đã khỏi, tìm xác suất sao cho người đó được điều trị khỏi bởi phương pháp III.

Kết quả:

A. 0,18875 B. 0,8 C. 0,2625 D. 0,31125 E. số khác

3. Dùng một phản ứng chẩn đoán bệnh, phản ứng có độ nhạy bằng 0,84 và giá trị âm tính bằng 0,968. Biết giá trị của phản ứng bằng 0,852, tìm giá trị dương tính.

Kết quả:

A. 0,854.118 B. 0,504 C. 0,25 D. 0,852 E. số khác.

4. Kiểm tra lại những người chẩn đoán bị bệnh ở bệnh viện I, II tuyến dưới thấy tương ứng 90% và 96% bị bệnh. Xác suất khỏi trước kiểm tra của 2 bệnh viện tương ứng bằng 0,955 và 0,94.

Tìm xác suất khỏi của hai bệnh viện sau kiểm tra, biết rằng số người bị bệnh sau kiểm tra của bệnh viện I bằng 5/3 bệnh viện II.

Kết quả:

A. 0,945.3125 B. 0,875.5875 C. 0,953.0875 D. 0,949.375 E. số khác.

Bài 3

QUY LUẬT XÁC SUẤT CỦA ĐẠI LƯỢNG NGẪU NHIÊN LIÊN TỤC

MỤC TIÊU

1. *Trình bày được bốn quy luật xác suất của đại lượng ngẫu nhiên liên tục (Chuẩn, Khi bình phương χ^2, Student, Fisher-Snedecor).*
2. *Tra được các giá trị tới hạn.*

Các đại lượng ngẫu nhiên được chia thành hai loại : đại lượng ngẫu nhiên rời rạc và đại lượng ngẫu nhiên liên tục.

Đại lượng ngẫu nhiên X được gọi là rời rạc nếu X nhận các giá trị 0, 1, 2, ..., n.

Số con của 1 gia đình, số người bị bệnh trong n người đến khám, số bệnh nhân điều trị khỏi trong tháng hay năm, số hồng cầu, số bạch cầu của một người là những đại lượng ngẫu nhiên rời rạc.

Đại lượng ngẫu nhiên X được gọi là đại lượng ngẫu nhiên liên tục nếu X nhận giá trị tuỳ ý trong đoạn [a, b].

Một người có chiều cao 160 cm là người có chiều cao đo được từ trên 159,5 cm đến dưới 160,5 cm nếu chấp nhận sai lệch 0,5 cm. Như vậy chiều cao là đại lượng ngẫu nhiên liên tục. Tương tự như chiều cao, cân nặng, các kích thước đo được của cơ thể, của các cơ quan nội tạng ... là các đại lượng ngẫu nhiên liên tục.

1. HÀM MẬT ĐỘ XÁC SUẤT VÀ HÀM PHÂN PHỐI XÁC SUẤT

1.1. Hàm mật độ xác suất

Hàm f(x) được gọi là hàm mật độ xác suất của đại lượng ngẫu nhiên liên tục X nếu:

+ $f(x) \ge 0 \ \forall x \in R$

+ $\int_{-\infty}^{+\infty} f(x)dx = P\{-\infty < X < +\infty\} = 1.$

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

Nhận thấy: $f(x) \geq 0 \ \forall x \in R$

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 1 \qquad \text{(Tích phân Laplace)}$$

Vậy hàm đã cho là hàm mật độ xác suất của đại lượng ngẫu nhiên X nào đấy.

Tương tự $f(x) = \frac{1}{a\sqrt{2\pi}} e^{-\frac{(x-b)^2}{2a^2}}$, trong đó a > 0 và b là các tham số, cũng là hàm mật độ xác suất của một đại lượng ngẫu nhiên liên tục.

Chú ý: Người ta thường ký hiệu hàm $e^{-x} = \exp(-x)$

1.2. Hàm phân phối xác suất

Giả sử f(x) là hàm mật độ xác suất của đại lượng ngẫu nhiên X,

$F(x) = \int_{-\infty}^{x} f(t)dt$ *được gọi là hàm phân phối xác suất của đại lượng ngẫu nhiên X.*

Nhận thấy F(x) là tích phân phụ thuộc cận trên cho nên nó là nguyên hàm của f(x). Đó chính là mối liên hệ giữa hàm phân phối xác suất và hàm mật độ xác suất.

Hàm phân phối xác suất F(x) có một số tính chất sau :

$F(-\infty) = 0$; $F(+\infty) = 1$

F(x) là hàm tăng vì $f(x) \geq 0 \ \forall x \in R$

F(x) là hàm liên tục bên trái.

2. CÁC ĐẶC TRƯNG CỦA ĐẠI LƯỢNG NGẪU NHIÊN

2.1. Trung bình lý thuyết (Kỳ vọng toán học)

Trung bình lý thuyết của đại lượng ngẫu nhiên X ký hiệu là MX, giá trị của nó ký hiệu là μ, được xác định như sau:

$$MX = \begin{cases} \sum_{i=1}^{n} p_i x_i & \text{X là đại lượng ngẫu nhiên rời rạc với } p_i = P\{X = xi\}. \\ \int_{-\infty}^{+\infty} x.f(x)dx & \text{X là đại lượng ngẫu nhiên liên tục, f(x) là hàm mật độ xác suất.} \end{cases}$$

MX là hằng số xác định của đại lượng ngẫu nhiên. Nó cho biết tâm phân phối của đại lượng ngẫu nhiên.

- Ước lượng điểm của MX

Khi không biết MX, lấy $MX \approx \overline{x}$ và được gọi là ước lượng điểm của MX. Ước lượng điểm rất thuận lợi trong sử dụng.

- Ước lượng khoảng của MX

Ký hiệu sai số giữa MX và $\overline{x}$ là δ_2

$$|MX - \overline{x}| \le \delta_2,\ \delta_2 = \begin{cases} t(\alpha/2)\dfrac{\sigma}{\sqrt{n}} & \text{biết } DX = \sigma^2 \\ t(n-1;\alpha/2)\dfrac{s}{\sqrt{n}} & \text{không biết } DX \end{cases}$$

Trong biểu thức trên $t(\alpha/2)$ tra trong bảng chuẩn tắc (bảng 1), $t(n-1;\alpha/2)$ tra trong bảng Student (bảng 2), DX là phương sai chuẩn. Bỏ trị số tuyệt đối được ước lượng khoảng

$$\overline{x} - \delta_2 \le MX \le \overline{x} + \delta_2$$

$\overline{x} \pm \delta_2$ được gọi là khoảng tin cậy mức $1 - \alpha$ của MX.

Ví dụ:

Cân các vật có khối lượng từ 50g – 200g, một cân có sai số là :

$\delta = \sqrt{DX} = 0{,}045g$. Cân một vật ba lần được các kết quả : 87,32g; 87,27g; 87,39g. Giá trị trung bình của ba lần cân là:

$$\overline{x} = \frac{1}{3}(87{,}27 + 87{,}32 + 87{,}39) = 87{,}327g$$

$$\delta_2 = 1{,}96\ \frac{0{,}045}{\sqrt{3}} = 0{,}051$$

Khối lượng đúng của vật $MX \approx \overline{x} = 87{,}327g$

Ước lượng khoảng của vật đó:

$$MX \in [\overline{x} - \delta_2\ ;\ \overline{x} + \delta_2] = [87{,}276\ ;\ 87{,}378].$$

2.2. Phương sai và Độ lệch chuẩn

- Phương sai lý thuyết của đại lượng ngẫu nhiên X ký hiệu là DX, giá trị của nó ký hiệu là σ^2, được xác định như sau:

$$Dx = \begin{cases} \sum_{i=1}^{n} p_i (x_i - MX)^2 & \text{X là đại lượng ngẫu nhiên rời rạc, } p_i = P\{X = x_i\}. \\ \int_{-\infty}^{+\infty} f(x)(x - MX)^2\, dx & \text{X là đại lượng ngẫu nhiên liên tục, f(x) là hàm mật độ xác suất của X.} \end{cases}$$

Phương sai là hằng số đặc trưng cho độ tản mạn của các giá trị của đại lượng ngẫu nhiên.

Ước lượng của DX

$DX \approx s_x^2$ là ước lượng điểm của DX

$DX \in \left[\dfrac{(n-1)s_x^2}{q(n-1;\alpha/2)} ; \dfrac{(n-1)s_x^2}{q(n-1;1-\alpha/2)} \right]$ là ước lượng khoảng của Dx, trong đó s_x^2 là phương sai thực nghiệm, $q(n-1;\ \alpha/2)$ và $q(n-1;\ 1-\alpha/2)$ là giá trị tra bảng khi bình phương.

- Độ lệch chuẩn

$\sigma = \sqrt{DX}$ được gọi là độ lệch tiêu chuẩn của đại lượng ngẫu nhiên X.

Khi không biết DX thường lấy $\sigma \approx s_x$ là ước lượng điểm của độ lệch chuẩn,

$\sigma \in [s - \delta_3 ; s + \delta_3]$ là ước lượng khoảng của độ lệch chuẩn, trong đó

$\delta_3 = t(\alpha/2).\dfrac{s}{\sqrt{2n}}$. $s \pm \delta_3$ được gọi là khoảng tin cậy mức $1-\alpha$ của độ lệch chuẩn.

Ngoài các hằng số MX, DX đặc trưng cho đại lượng ngẫu nhiên, mômen bậc k của đại lượng ngẫu nhiên được xác định như sau:

$M(X - MX)^k = \sum_{i=1}^{n} p_i (x_i - MX)^k$ với X là đại lượng ngẫu nhiên rời rạc.

$k = 3$, mô men bậc 3 cho độ nhọn của đại lượng ngẫu nhiên.

$k = 4$, mô men bậc 4 cho hệ số đối xứng của đại lượng ngẫu nhiên.

3. QUY LUẬT CHUẨN (GAUSS - LAPLACE)

Quy luật chuẩn do Gauss, Laplace nghiên cứu đầu tiên.

3.1. Định nghĩa

Đại lượng ngẫu nhiên X liên tục, nhận giá trị trên R được gọi là có quy luật chuẩn, gọi tắt là đại lượng ngẫu nhiên chuẩn hay biến chuẩn với tham số μ và σ^2 nếu hàm mật độ xác suất của X có dạng:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$, trong đó MX = μ; DX = σ^2 là các tham số đã biết.

Khi $\mu = 0$, $\sigma^2 = 1$, đại lượng ngẫu nhiên X được gọi là chuẩn tắc. Như vậy chuẩn tắc là chuẩn có tham số đặc biệt 0 và 1.

Các đại lượng bình thường trong sinh, y học thường có quy luật chuẩn. Dẫn đến ký hiệu X chuẩn với μ và σ^2 như sau X : N (μ, σ^2) và chuẩn tắc là X : N (0 ; 1).

3.2. Các đặc trưng của quy luật chuẩn

Giả sử $X : N(\mu\,;\sigma^2)$ thì $MX = \mu$, $DX = \sigma^2$

Giả sử $X : N(0\,;1)$ thì $MX = 0$, $DX = 1$

Thật vậy $MX = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} x \exp(-\frac{x^2}{2}) dx = 0$

Do hàm lấy tích phân là hàm lẻ, miền tích phân đối xứng.

$$DX = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^2 \exp\left(-\frac{x^2}{2}\right) dx$$

Tích phân từng phần đặt u = x, dv = $x \exp(-\frac{x^2}{2})dx$

$$DX = \lim_{\substack{a \to -\infty \\ b \to +\infty}} \left[-\frac{x}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \Big|_a^b \right] + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left(-\frac{x^2}{2}\right) dx = 1.$$

3.3. Tra bảng

Ký hiệu $\Pi(x) = \int_{-\infty}^{x} f(t)dt = P\{-\infty < X \le x\}$

$\Pi(x)$ là hàm phân phối xác suất của đại lượng ngẫu nhiên chuẩn tắc

$$X : N\ (0\,;1) \text{ với } f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Tích phân $\pi(x)$ lấy các giá trị x từ 0 đến 4,5 lập được bảng $\pi(x)$ (bảng 1).

Trong các sách đôi khi lập bảng $\phi(x)$ với

$$\phi(x) = \int_{-x}^{x} f(t)dt = P\{-x \le X \le x\}$$

dẫn đến $\Pi(x) = 0,5 + \frac{1}{2}\phi(x)$

- $X : N(0\,;1)$

$$\Pi(-x) = 1 - \Pi(x)$$

$$P\{a \le x \le b\} = \Pi(b) - \Pi(a)$$

$$P\{x > b\} = 1 - P\{-\infty < x \le b\} = 1 - \Pi(b)$$

- $X : N\,(\mu, \sigma^2)$

$$P\{a \le x \le b\} = P\left\{\frac{a-\mu}{\sigma} \le \frac{x-\mu}{\sigma} \le \frac{b-\mu}{\sigma}\right\} =$$

$$= \Pi(\frac{b-\mu}{\sigma}) - \Pi(\frac{a-\mu}{\sigma})$$

Dựa vào bảng $\Pi(x)$ tra được

$\Pi(1) = 0,8413$ $\qquad$ $\Pi(-1) = 1 - \Pi(1) = 0,1587$

$\Pi(2) = 0,9772$ $\qquad$ $\Pi(-2) = 1 - \Pi(2) = 0,0228$

$\Pi(3) = 0,99865$ $\qquad$ $\Pi(-3) = 1 - \Pi(3) = 0,00135$

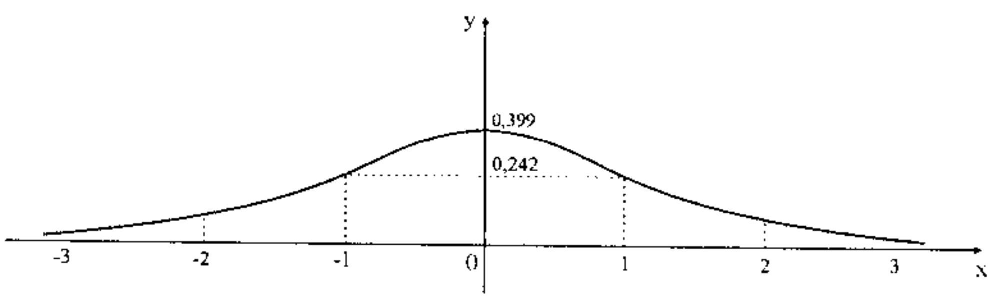
3.4. Thang phân loại

	$\mu - 3\sigma$	$\mu - 2\sigma$	$\mu - \sigma$	μ	$\mu + \sigma$	$\mu + 2\sigma$	$\mu + 3\sigma$
X:N(0; 1):	– 3	– 2	– 1	0	1	2	3
$\Pi(x)$	0,00135	0,0228	0,1587	0,5	0,8413	0,9772	0,99865

Rất thấp	thấp	hơi thấp	trung bình	hơi cao	cao	rất cao
0,00135	0,02145	0,1359	← 0,6826 →	0,1359	0,02145	0,00135

← Hơi thấp đến hơi cao →
0,9544

← Thấp đến cao →
0,9973

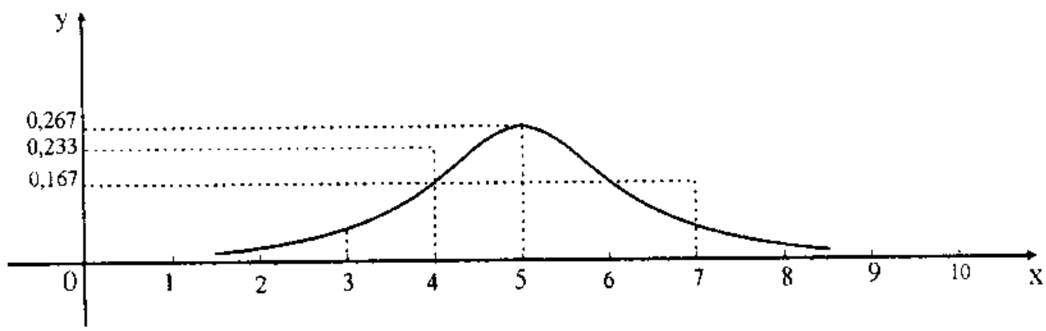


Thang phân loại của X : N(0, 1)

Ví dụ:

Gọi X là áp lực động mạch phổi thì tâm trương người bình thường (mm Hg). Nghiên cứu đã thu được số liệu với tần suất tương ứng sau:

X_i	2	3	4	5	6	7	8	9	Σ
m_i	1	4	7	8	2	5	2	1	30
ω_i	0,033	0,133	0,233	0,267	0,067	0,167	0,067	0,033	1



Đồ thị đường cong tần suất chuẩn làm trơn

Làm trơn đường gấp khúc nối các đỉnh tần suất tương ứng với các giá trị x_i được đường cong gần với đường cong của hàm mật độ xác suất chuẩn:

$$f(x) = \frac{1}{1,7\sqrt{2\pi}} \exp\left\{-\frac{(x-5,1)^2}{2\times 1,7^2}\right\},$$

trong đó $\mu \approx \overline{x} = 5,1,\ \sigma^2 \approx s^2 = 1,7^2$,

Dựa vào tính chuẩn để tính các sai số δ_1, δ_2, δ_3 và các khoảng tin cậy của P(A), MX và $\sqrt{DX}$.

Các bài toán so sánh trung bình lý thuyết hay so sánh phương sai của hai đại lượng ngẫu nhiên cũng được tiến hành trên cơ sở giả thiết các đại lượng ngẫu nhiên có quy luật chuẩn.

3.5. Quy luật log chuẩn

- Đại lượng ngẫu nhiên X (biến ngẫu nhiên X) được gọi là có quy luật ln chuẩn với tham số μ và σ^2 nếu Y = ln X có quy luật chuẩn : $N(\mu,\sigma^2)$ với $MY = \mu$, $DY = \sigma^2$

Khi đó $\quad MX = \exp(\mu + \frac{\sigma^2}{2}) \qquad DX = (\exp \sigma^2 - 1).\exp(2\mu + \sigma^2)$

- $P\{a \le X \le b\} = P(\ln a \le Y \le \ln b)$

$$= \Pi\left(\frac{\ln b - \mu}{\sigma}\right) - \Pi\left(\frac{\ln a - \mu}{\sigma}\right)$$

Ví dụ:

Gọi X là áp lực tâm thu động mạch phổi bệnh nhân hẹp hai lá, $\bar{x} = 72,4\text{mmHg}$ $s_x^2 = 30,49^2$. $Y = \ln X$, $Y : N(4,18; 0,46^2)$

$$MX = \exp(4,18 + \frac{0,2116}{2}) = 72,66$$

$$DX = [\exp(0,2116) - 1]\exp(2 \times 4,18 + 0,2116) = 35,27^2.$$

4. CÁC QUY LUẬT KHÁC

4.1. Hàm Gamma

Hàm Gamma, ký hiệu là $\Gamma(n)$, có biểu thức sau:

$$\Gamma(n) = \int_0^{\infty} e^{-x} x^{n-1} dx.$$

Ứng với các giá trị $n \le 2$, $\Gamma(n)$ được cho trong một bảng. Ví dụ $\Gamma(1) = 1$, $\Gamma(1,46) = 0,8856$, $\Gamma(2) = 1$.

Ứng với các giá trị $n > 2$, $\Gamma(n)$ được tính xấp xỉ theo công thức

$\Gamma(n) = (n-1)\Gamma(n-1)$.

Với n nguyên dương thì $\Gamma(n) = n!$

Với n đủ lớn, $\Gamma(n)$ được tính theo công thức sau:

$$\Gamma(n) = n^n e^{-n} \sqrt{2\pi} / \sqrt{n}.$$

4.2. Quy luật khi bình phương χ^2

- Định lý 1

Giả sử $X_1, X_2, \ldots, X_n$ là n đại lượng ngẫu nhiên chuẩn tắc độc lập thì

$Q_n = \sum_{i=1}^{n} X_i^2$ là đại lượng nhẫu nhiên χ^2 với n bậc tự do

X_1 là biến chuẩn tắc thì $Q_1 = X_1^2$ là một biến χ^2 với 1 bậc tự do.

Chia n giá trị nghiên cứu thành k hàng, có k –1 bậc tự do.

Chia n giá trị nghiên cứu thành k hàng và l cột, có (k – 1)(l – 1) bậc tự do.

- Q_n là đại lượng ngẫu nhiên có quy luật χ^2 với n bậc tự do, khi đó hàm mật độ xác suất của Q_n có biểu thức sau:

$$f(x) = \frac{1}{2^{n/2}\Gamma(n/2)} e^{-\frac{x}{2}} x^{\frac{n}{2}-1}, \ (x \geq 0).$$

4.3. Quy luật Student (Gosset W.S)

- Định lý 2

Giả sử X là đại lượng ngẫu nhiên chuẩn tắc độc lập với Q_n là đại lượng ngẫu nhiên χ^2 với n bậc tự do thì

$T_n = \dfrac{X}{\sqrt{\dfrac{Q_n}{n}}}$ là đại lượng ngẫu nhiên có quy luật Student với n bậc tự do.

Khi $n \to \infty$, $T_n \Rightarrow T : N(0, 1)$

- T_n là đại lượng ngẫu nhiên có quy luật Student với n bậc tự do. Khi đó hàm mật độ xác suất của T_n có biểu thức sau:

$$f(x) = \frac{\Gamma\left(\frac{n}{2}+\frac{1}{2}\right)}{n^{\frac{1}{2}}\,\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{1}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n}{2}-\frac{1}{2}}, \ (x \in R).$$

4.4. Quy luật Fisher – Snedecor

- Định lý 3

Giả sử Q_{n_1} là đại lượng ngẫu nhiên χ^2 với n_1 bậc tự do độc lập với Q_{n_2} là đại lượng ngẫu nhiên χ^2 với n_2 bậc tự do thì :

$$F_{n_1,n_2} = \frac{Q_{n_1}}{n_1} : \frac{Q_{n_2}}{n_2}$$

là đại lượng ngẫu nhiên có quy luật Fisher–Snedecor với n_1 và n_2 bậc tự do.

- F_{n_1,n_2} là đại lượng ngẫu nhiên có quy luật Fisher–Snedecor với n_1, n_2 bậc tự do. Khi đó hàm mật độ xác suất của F_{n_1,n_2} có biểu thức sau:

$$f(x) = \frac{\Gamma\left(\frac{n_1}{2}+\frac{n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)}\left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}}\frac{x^{\frac{n_1}{2}-1}}{\left(1+\frac{n_1}{n_2}\right)^{\frac{n_1}{2}+\frac{n_2}{2}}}, \ (x \geq 0).$$

5. GIÁ TRỊ TỚI HẠN

Hiện tượng A có $P(A) \geq 0{,}95$, khi thực hiện phép thử, A hầu như chắc chắn xuất hiện.

Hiện tượng B có $P(B) \leq 0{,}05$, khi thực hiện phép thử, B hầu như chắc chắn không xuất hiện.

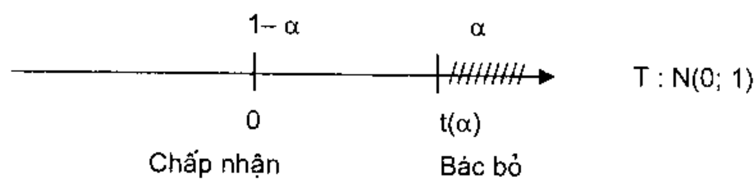
Những khẳng định trên không phải là các định lý hoàn toàn đúng mà chỉ là các quyết định đúng nhiều, sai ít.

Trong các bài toán kiểm định giả thiết thống kê cần quyết định chấp nhận giả thiết đưa ra hay bác bỏ giải thiết đó. Khi đó dựa vào mức ý nghĩa của xác suất để xác định một giá trị, căn cứ vào giá trị này mà quyết định chấp nhận hay bác bỏ giả thiết thống kê. Giá trị này được gọi là giá trị tới hạn. Chọn mức xác suất α là 5% hoặc 1% sẽ có giá trị tới hạn mức 95% hay 99%. Mỗi quy luật xác suất có các giá trị tới hạn riêng.

5.1. Quy luật chuẩn tắc T : N (0;1)

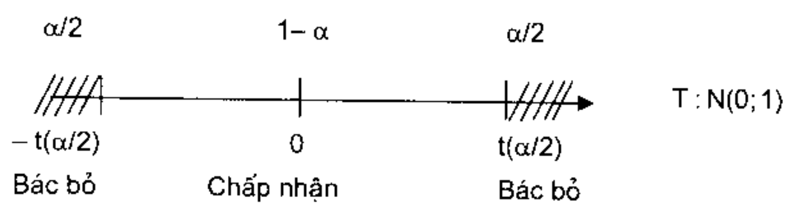
- $P\{-\infty < T < +\infty\} = P\{-\infty < T \leq t(\alpha)\} + P\{T > t(\alpha)\}$

Nếu xác định được $t(\alpha)$ sao cho $P\{T > t(\alpha)\} = \alpha$ thì $t(\alpha)$ được gọi là giá trị tới hạn một phía mức $1 - \alpha$



- $P\{-\infty < T < +\infty\} = P\{\,|T| \leq t(\alpha/2)\} + P\{\,|T| > t(\alpha/2)\}$

Nếu xác định được $t(\alpha/2)$ sao cho $P\{|T| > t(\alpha/2)\} = \alpha$ thì $t(\alpha/2)$ được gọi là giá trị tới hạn 2 phía mức $1 - \alpha$



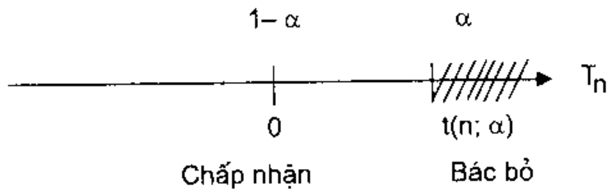
$t(\alpha)$, $t(\alpha/2)$ tra ở chỉ dẫn 2 bảng 1.

Ví dụ $t(0{,}05) = 1{,}6449$; $t(0{,}05/2) = 1{,}96$; $t(0{,}1/2) = t(0{,}05)$.

5.2. Quy luật Student T_n

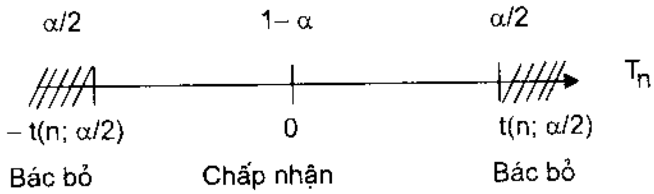
- $P\{-\infty < T_n < +\infty\} = P\{-\infty < T_n \leq t(n;\alpha)\} + P\{T_n > t(n;\alpha)\}$

Nếu xác định được t(n;α) sao cho $P\{T_n > t(n;\alpha)\} = \alpha$ thì t(n; α) được gọi là giá trị tới hạn một phía mức 1 – α



- $P\{-\infty < T_n < +\infty\} = P\{\,|T_n| \le t(n;\alpha/2)\} + P\{\,|T_n| > t(n;\alpha/2)\}$

Nếu xác định được t(n; α/2) sao cho $P\{\,|T_n| > t(n;\alpha/2)\} = \alpha$ thì t(n; α/2) được gọi là giá trị tới hạn hai phía mức 1 – α



t(n;α) và t(n; α/2) tra ở bảng 2.

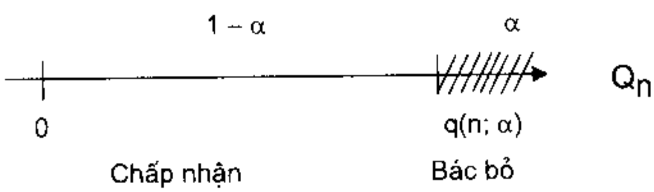
Ví dụ: t(1; 0,05) = 6,314

t(1;0,05/2) = 12,706; t(61; 0,1/2) = t(61; 0,05) = 1,645.

5.3. Quy luật khi bình phương Q_n

$$P\{0 \le Q_n < +\infty\} = P\{0 \le Q_n \le q(n;\alpha)\} + P\{Q_n > q(n;\alpha)\}$$

Nếu xác định được q(n;α) sao cho $P\{Q_n > q(n;\alpha)\} = \alpha$ thì q(n;α) được gọi là giá trị tới hạn mức 1 – α

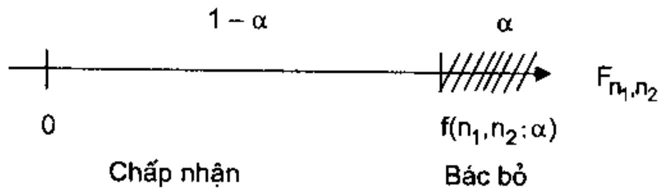


q(n;α) tra ở bảng 3. Ví dụ q(1; 0,05) = 3,841; q(2; 0,01) = 9,210.

5.4. Quy luật Fisher – Snedecor F_{n_1,n_2}

$$P\{0 \le F_{n_1,n_2} < +\infty\} = P\{0 \le F_{n_1,n_2} \le f(n_1,n_2;\alpha)\} + P\{F_{n_1,n_2} > f(n_1,n_2;\alpha)\}$$

Nếu xác định được f(n_1, n_2; α) sao cho $P\{F_{n_1,n_2} > f(n_1,n_2:\alpha)\} = \alpha$ thì f(n_1, n_2; α) được gọi là giá trị tới hạn mức 1 – α.



$f(n_1, n_2; \alpha)$ với $\alpha = 0{,}05$ tra ở bảng 4.

Ví dụ:

$f(10; 1; 0{,}05) = 242$, n_1 tra ở cột, n_2 tra ở hàng.

Các giá trị trên một cột không tuyến tính cho nên không nội suy mà lấy giá trị gần nhất: $f(1;\ \ 34;\ \ 0{,}05) \approx f(1;\ \ 30;\ \ 0{,}05) = 4{,}17$

$$f(1;\ \ 35;\ \ 0{,}05) \approx f(1;\ \ 40;\ \ 0{,}05) = 4{,}08$$

Các giá trị trên một hàng xem là tuyến tính cho nên nội suy các giá trị không có trong bảng. Ví dụ cần tra $f(12; 1; 0{,}05)$

$$\left.\begin{array}{l} f(15;\ \ 1;\ \ 0{,}05) = 246 \\ f(10;\ \ 1;\ \ 0{,}05) = 242 \end{array}\right\} \quad \Rightarrow n_1 : \text{tăng 5 bậc, giá trị f tăng 4}$$

vậy $f(12;1;0{,}05) = f(10;1;0{,}05) + 2 \times \dfrac{4}{5} = 243{,}6$

hoặc $f(12; 1; 0{,}05) = f(15; 1; 0{,}05) - 3 \times \dfrac{4}{5} = 243{,}6.$

5.5. Tra ngược

Giả sử cần tra $t(0{,}15)$. Giá trị này không có trong bảng giá trị tới hạn $t(\alpha)$. Có thể dựa vào bảng $\Pi(x)$ để tìm $t(\alpha)$ khi biết α.

Giả sử X là đại lượng ngẫu nhiên, $X : N(0;1)$.

$P\{-\infty < X \le x\} = \Pi(x) = p$. Như vậy cho x, tìm được p trong bảng $\Pi(x)$.

Giả sử T là đại lượng ngẫu nhiên, $T : N(0;1)$.

$$P\{T > t(\alpha)\} = \alpha$$

$$P\{T \le t(\alpha)\} = \Pi(t(\alpha)) = 1 - \alpha$$

Đặt $1 - \alpha = p \qquad p = \Pi(t(\alpha)) \Leftrightarrow t(\alpha) = x.$

Như vậy cho α, tìm được $t(\alpha)$ trong bảng $\Pi(x)$ nhờ tra ngược.

Khi $\alpha > 0{,}5 \qquad p = 1 - \alpha \ = \Pi(t(\alpha))$ không có trong bảng.

$$\Pi(x) = \alpha = 1 - \Pi(t(\alpha)) = \Pi(-t(\alpha)).$$

Vậy $\qquad t(\alpha) \ = -x.$

Tra $t(\alpha/2)$ như tra $t(\alpha)$. Khi đó α chính là giá trị $\alpha/2$.

Ví dụ:

- Tra $t(0,15)$.

  $\alpha = 0,15$. $\quad p = 1 - 0,15 = 0,85 = \Pi(1,04)$

  $t(0,15) = 1,04$.

- Tra $t(0,15/2)$

  $\alpha = 0,15/2 = 0,075$. $\quad p = 1 - 0,075 = 0,925 = \Pi(1,44)$

  $t(0,075) = t(0,15/2) = 1,44$.

- Tra $t(0,6)$

  $\alpha = 0,6$. $\quad 0,6 = \Pi(0,25)$

  $t(0,6) = -0,25$.

CÂU HỎI TỰ LƯỢNG GIÁ

Hãy chọn một kết quả đúng.

1. Gọi X là chiều cao nam thanh niên Việt Nam (đv:cm), X : $N(\mu,\sigma^2)$ với $\mu = 163,72$ và $\sigma^2 = 4,67^2$. Đo chiều cao một nam thanh niên, tìm xác suất sao cho chiều cao người đó nằm trong khoảng [156,248; 171,192].

Kết quả:

A. 0,1096 B. 0,9452 C. 0,8904 D. 0,0548 E. số khác.

2. Số lượng hồng cầu trong máu ngoại vi (đv: T/l) là đại lượng ngẫu nhiên chuẩn với $\mu = 5,05$ và $\sigma^2 = 0,38^2$. Đếm hồng cầu cho 4673 người, có bao nhiêu người có số hồng cầu trên 6T/l.

Kết quả:

A. 1 B. 29 C. 312 D. 6 E. số khác.

3. Giả sử T là đại lượng ngẫu nhiên chuẩn : N(0;1). Hãy cho biết $t(0,06/2)$.

Kết quả:

A. 0,5239 B. 0,5120 C. –1,88 D. 1,88 E. số khác.

4. Giả sử T là đại lượng ngẫu nhiên chuẩn tắc. Hãy cho biết $t(0,56)$.

Kết quả:

A. –0,15 B. 0,7123 C. 0,15 D. –0,7123 E. số khác.

Bài 4

QUY LUẬT XÁC SUẤT CỦA ĐẠI LƯỢNG NGẪU NHIÊN RỜI RẠC

MỤC TIÊU

1. *Trình bày được bốn quy luật xác suất (Nhị thức, Poisson, Siêu bội, Đa thức).*
2. *Trình bày được ý nghĩa của bốn quy luật.*
3. *Giải được bài toán xác suất có quy luật nhị thức.*

Đại lượng ngẫu nhiên liên tục thường gặp là đại lượng ngẫu nhiên có quy luật chuẩn hoặc Student hoặc khi bình phương. Các đại lượng ngẫu nhiên rời rạc có các quy luật xác suất khác nhau tuỳ thuộc các phép thử độc lập hay không độc lập. Trong bài này sẽ xét các quy luật xác suất của đại lượng ngẫu nhiên rời rạc hay gặp trong nghiên cứu sinh, y học.

1. QUY LUẬT NHỊ THỨC – BERNOULLI

1.1. Định nghĩa

Thực hiện phép thử ε n lần độc lập, hiện tượng A có xác suất P(A) = p. Gọi X là số lần xuất hiện A khi thực hiện phép thử n lần.

Đại lượng ngẫu nhiên X được gọi là đại lượng có quy luật nhị thức với các tham số n và p nếu:

+ *X nhận các giá trị 0, 1, 2, 3, ..., n*

+ $P(X = r) = C_n^r p^r (1-p)^{n-r} = B(r, n, p)$.

Công thức tính xác suất là số hạng tổng quát của khai triển nhị thức. Đôi khi ký hiệu q = 1 – p.

Ví dụ:

1. Tỷ lệ bị bệnh tại phòng khám đa khoa bằng 0,2. Khám bệnh cho 10 người, tìm xác suất sao cho có 2 người bị bệnh.

Khám bệnh cho 10 người, họ có bị bệnh hay không là các phép thử độc lập. Gọi X là số người bị bệnh khi khám cho 10 người. X có quy luật nhị thức với tham số n = 10, p = 0,2.

$$P(X = 2) = C_{10}^{2} \times 0,2^2 \times 0,8^8 = \frac{10!}{8!\ 2!} 0,2^2 \times 0,8^8 = 0,302.$$

2. Điều trị một bệnh có xác suất khỏi bằng 0,8. Điều trị cho 10 người bệnh trên, tìm xác suất sao cho có 8 người khỏi.

Chữa cho 10 người bị bệnh khỏi hay không là các phép thử độc lập. Gọi X là số khỏi khi chữa cho 10 người. X có quy luật nhị thức với tham số n = 10, p = 0,8.

$$P(X = 8) = C_{10}^{8} \times 0,8^8 \times 0,2^2 = \frac{10!}{2!8!} 0,8^8 \times 0,2^2 = 0,302.$$

3. Xác suất phản ứng thuốc khi điều trị kháng sinh cho bệnh nhân bằng 0,001. Điều trị cho 100 người, tìm xác suất sao cho có 1 người bị phản ứng thuốc.

Bệnh nhân có bị phản ứng thuốc hay không khi điều trị kháng sinh là do thuốc và cơ địa người bệnh. Điều trị thuốc cho 100 người bị bệnh là các phép thử độc lập. Gọi X là số người bệnh bị phản ứng thuốc khi điều trị. X có quy luật nhị thức với n = 100 và p = 0,001.

$$P(X=1) = C_{100}^{1} \times 0,001^1 \times 0,999^{99} = 0,091.$$

1.2. Tính chất

1.2.1. Đại lượng ngẫu nhiên X có quy luật nhị thức với tham số n và p thì:

$$P(0 \le X \le n) = 1.$$

Thật vậy:

$$P(0 \le X \le n) = \sum_{r=0}^{n} C_n^r p^r (1-p)^{n-r} = (p + 1 - p)^n = 1.$$

1.2.2. Đại lượng ngẫu nhiên X có quy luật nhị thức với tham số n và p thì:

$$MX = np,\ DX = npq.$$

Chứng minh

$$MX = \sum_{i=0}^{n} p_i x_i \qquad \text{với} \qquad p_i = P(X = i) = C_n^i p^i (1-p)^{n-i}, \quad x_i = i$$

$$= \sum_{i=0}^{n} \frac{n!}{(n-i)!\ i!} p^i (1-p)^{n-i}.i$$

$$= \sum_{i=1}^{n} \frac{n!}{(n-i)!\ (i-1)!} p^i (1-p)^{n-i}$$

$$= np\sum_{i=1}^{n}\frac{(n-1)!}{(n-i)!\,(i-1)!}p^{i-1}(1-p)^{n-i}$$ Đặt $i-1=s$

$$MX = np\sum_{s=0}^{n-1}C_{n-1}^{s}\,p^{s}(1-p)^{n-1-s} = np.$$

$$DX = \sum_{i=0}^{n}p_i(x_i-np)^2$$

$$= \sum_{i=0}^{n}C_n^i\,p^i(1-p)^{n-i}\left[i^2-2i.np+(np)^2\right]$$

$$= \sum_{i=0}^{n}\frac{n!}{(n-i)!\,i!}p^i(1-p)^{n-i}.i^2-(np)^2$$

$$= \sum_{i=0}^{n}\frac{n!}{(n-i)!i!}p^i(1-p)^{n-i}.\,i(i-1)+np-(np)^2$$

$$= \sum_{i=2}^{n}\frac{n!}{(n-i)!\,(i-2)!}\,p^i(1-p)^{n-i}+np-(np)^2$$

$$= n(n-1)p^2\sum_{i=2}^{n}\frac{(n-2)!}{(n-i)!\,(i-2)!}\,p^{i-2}(1-p)^{n-i}+np-(np)^2$$

$$= n(n-1)p^2\sum_{s=0}^{n-2}\frac{(n-2)!}{(n-2-s)!\,s!}\,p^{s}(1-p)^{n-2-s}+np-(np)^2$$, với $s=i-2$

$= (np)^2 - np^2 + np - (np)^2$

$= np(1-p) = npq.$

Xét ví dụ 1: $MX = 10 \times 0{,}2 = 2$, $DX = 10 \times 0{,}2 \times 0{,}8 = 1{,}6$.

ví dụ 2: $MX = 10 \times 0{,}8 = 8$, $DX = 10 \times 0{,}8 \times 0{,}2 = 1{,}6$.

1.2.3. Đại lượng ngẫu nhiên X có quy luật nhị thức với tham số n và p thì:

$P(X = r)$ lớn nhất khi $r = [(n+1)p]$, trong đó [...] là hàm phần nguyên.

Phần nguyên của một số là số nguyên lớn nhất nhỏ hơn hay bằng số đó

Biến X nhận giá trị từ 0 đến n, có n + 1 xác suất.

$$\frac{P(X=r)}{P(X=r-1)}=\frac{\dfrac{n!}{(n-r)!\,r!}\,p^r(1-p)^{n-r}}{\dfrac{n!}{(n-r+1)!\,(r-1)!}\,p^{r-1}(1-p)^{n-r+1}}=\frac{(n-r+1)}{r}\times\frac{p}{1-p}\geq 1.$$

Dẫn đến $r \le (n+1)p$. Khi $r \le (n+1)p$ thì $P(X = r) \ge P(X = r - 1)$

Khi $r > (n+1)p$ thì $P(X = r) < P(X = r - 1)$.

Vậy $r = [(n+1)p]$ thì $P(X = r)$ lớn nhất. Khi $(n+1)p$ là số nguyên thì $P(X=r) = P(X=r-1)$ là giá trị lớn nhất, r còn được ký hiệu là Mo (Mode).

Điều trị một bệnh có xác suất khỏi bằng 0,7. Điều trị cho 10 người.

Tìm xác suất có i người khỏi, qua đó tìm xác suất lớn nhất.

Chữa cho 10 người khỏi hay không là các phép thử độc lập. Gọi X là số khỏi khi chữa cho 10 người. X có quy luật nhị thức với n = 10 và p = 0,7.

$r = [(10 + 1)0,7] = [7,7] = 7$

$$P(X = 0) = C_{10}^{0}.0,7^{0}.0,3^{10} = 0,000006$$

$$P(X = 1) = C_{10}^{1}.0,7^{1}.0,3^{9} = 0,000138$$

....

$$P(X = 6) = C_{10}^{6}.0,7^{6}.0,3^{4} = 0,200121$$

$$P(X = 7) = C_{10}^{7}.0,7^{7}.0,3^{3} = 0,2668$$

$$P(X = 8) = C_{10}^{8}.0,7^{8}.0,3^{2} = 0,2335 \ \ldots.$$

Vậy P (X = 7) lớn nhất. X = 7 là giá trị hay gặp nhất.

Xét ví dụ 1 với n = 10, p = 0,2, r = [(10+1)0,2] = 2.

$P(X=2) = C_{10}^{2}0,2^{2}.0,8^{8} = 0,302$ Đó là giá trị xác suất lớn nhất.

1.2.4. Đại lượng ngẫu nhiên X có quy luật nhị thức với tham số n và p, khi $n \to \infty$, p không gần 0 hoặc 1 thì biến X có quy luật nhị thức sẽ tiến tới quy luật chuẩn với tham số $\mu = MX = np$ và $\sigma^2 = DX = npq$.

$$P\{m_1 \le X \le m_2\} = \sum_{k=m_1}^{m_2} C_n^k\, p^k(1-p)^{n-k}$$

$$\approx \Pi\left(\frac{m_2 - np}{\sqrt{npq}}\right) - \Pi\left(\frac{m_1 - np}{\sqrt{npq}}\right)$$

Ví dụ:

Điều trị kháng sinh C_0 cho trẻ bị viêm nhiễm đường hô hấp trên do vi khuẩn có tỷ lệ khỏi bằng 0,6.

Tìm xác suất sao cho điều trị cho 100 trẻ có:

a) Số trẻ khỏi từ 55 đến 70 trẻ.

b) Đúng 60 trẻ khỏi.

Giải:

Điều trị cho 100 trẻ là các phép thử độc lập.

Gọi X là số khỏi khi điều trị cho 100 trẻ. X có quy luật nhị thức với n = 100 và p = 0,6.

a) $P\{55 \le X \le 70\} = \sum_{r=55}^{70} C_{100}^{r} \times 0,6^{r}(1-0,6)^{100-r}$

$$\approx \Pi\left(\frac{70-100\times0,6}{\sqrt{100\times0,6\times0,4}}\right)-\Pi\left(\frac{55-100\times0,6}{\sqrt{100\times0,6\times0,4}}\right)$$

$$= \Pi(2)-\Pi(-1,02) = 0,9772+0,8461-1 = 0,8233.$$

b) $P\{X=60\} = C_{100}^{60}.0,6^{60}.0,4^{40}$

$$\approx P\{60-0,5 \le X \le 60+0,5\}$$

$$= \Pi\left(\frac{60,5-60}{\sqrt{24}}\right)-\Pi\left(\frac{59,5-60}{\sqrt{24}}\right)$$

$$\approx \Pi(0,10)-\Pi(-0,10) = 2\Pi(0,1)-1$$

$$= 2 \times 0,5398 - 1 = 0,0796.$$

2. QUY LUẬT POISSON

2.1. Định nghĩa :

Đại lượng ngẫu nhiên X được gọi là có quy luật Poisson với tham số $\lambda > 0$ nếu :

\+ *X nhận các giá trị 0, 1, 2,..., n.*

\+ $P(X=r) = \dfrac{e^{-\lambda}.\lambda^{r}}{r!} = P(r,\lambda)$,

trong đó: $e = \lim\limits_{x\to0}(1+x)^{\frac{1}{x}} = 2,7182818284590$.

2.2. Tính chất

2.2.1. Đại lượng ngẫu nhiên X có quy luật Poisson với tham số $\lambda > 0$ thì $MX = DX = \lambda$.

2.2.2. Đại lượng ngẫu nhiên X có quy luật nhị thức với tham số n và p, khi n tiến tới ∞ và p tiến tới 0 sao cho $np \to \lambda$ thì đại lượng ngẫu nhiên X có quy luật nhị thức sẽ tiến tới quy luật Poisson với tham số $\lambda > 0$.

Thực hiện n phép thử độc lập, hiện tượng A có xác suất P(A) = p. Gọi X là số lần xuất hiện A. X có quy luật nhị thức với tham số n và p.

$$P(X = r) = C_n^r p^r (1-p)^{n-r} = \frac{n!}{(n-r)!\,r!} p^r (1-p)^{n-r}.$$

Khi $n \to \infty$ và $p \to 0$ sao cho $np \to \lambda$ thì

$$P(X = r) = \frac{n!}{(n-r)!r!} p^r (1-p)^{n-r}$$

$$= \frac{(n-r+1)(n-r+2)\ldots n}{r!} \times \frac{n^r p^r}{n^r} (1-p)^{n-r}$$

$$= \frac{1}{r!}\left(1 - \frac{r-1}{n}\right)\left(1 - \frac{r-2}{n}\right)\ldots 1 \times (np)^r (1-p)^{n-r}$$

$$\lim_{\substack{n\to\infty \\ p\to 0}} (1-p)^{n-r} = \lim_{\substack{n\to\infty \\ p\to 0}} \left\{[1+(-p)]^{-\frac{1}{p}}\right\}^{-p(n-r)} = e^{-\lambda}.$$

Vậy $$P(X = r) \approx \frac{e^{-\lambda}\lambda^r}{r!}$$

P(A) = p rất gần 0 cho nên A là hiện tượng hiếm gặp.

Số lần xuất hiện hiện tượng hiếm gặp khi thực hiện số lớn phép thử độc lập sẽ có quy luật Poisson.

P(A) = p rất gần 0, q = 1 − p rất gần 1 cho nên MX = np, DX = npq dẫn đến DX ≈ MX = λ.

Ví dụ:

1. Xác suất mắc bệnh sau khi dùng vacxin bằng 0,001. Dùng vacxin cho 2000 trẻ. Tìm xác suất sao cho có 4 trẻ bị bệnh.

Giải:

Gọi X là số trẻ bị bệnh sau khi dùng vacxin.

X có quy luật nhị thức với n = 2000 và p = 0,001.

$$P(X = 4) = C_{2000}^4 \times 0{,}001^4 \times 0{,}999^{1996} = 0{,}090.$$

Nếu xem n = 2000 là đủ lớn, p = 0,001 gần 0.

MX = 2000 × 0,001 = 2, DX = 2000 × 0,001× 0,999 = 1,998

$\Rightarrow DX \approx MX = \lambda = 2$

Tính xác suất theo quy luật Poisson

$$P(X=4)=\frac{e^{-2}\times 2^4}{4!}=0,090$$

Sai số giữa hai cách tính rất không đáng kể.

2. Đếm hồng cầu trong 400 ô của kính hiển vi. Xác suất để một hồng cầu rơi vào một ô bằng 0,0025. Tìm xác suất sao cho trong số 1000 hồng cầu có 3 hồng cầu rơi vào một ô.

Giải:

Gọi X là số hồng cầu rơi vào một ô.

X có quy luật nhị thức với n =1000 và p = 0,0025 (phép thử độc lập)

$$P(X=3)=C_{1000}^{3}\times\ 0,0025^3\times 0,9975^{997}=0,214\,.$$

Khi n = 1000 xem là đủ lớn, p = 0,0025 là gần 0,

$MX = 1000 \times 0,0025 = 2,5$, $DX = 1000 \times 0,0025\times 0,9975 = 2,49375$

$\Rightarrow DX \approx MX = \lambda = 2,5$.

Tính xác suất theo quy luật Poisson

$$P(X=3)=\frac{e^{-2,5}.2,5^3}{3!}=0,2138\,.$$

Sai số rất nhỏ.

3. QUY LUẬT SIÊU BỘI

3.1. Định nghĩa

Nghiên cứu một đám đông có m phần tử, trong đó m_1 phần tử có đặc tính A, số còn lại m_2 không có đặc tính A. Lấy ngẫu nhiên, không hoàn lại một mẫu n phần tử từ m phần tử. Gọi X là số phần tử có đặc tính A trong n phần tử lấy ra.

Đại lượng ngẫu nhiên X được gọi là có quy luật siêu bội với các tham số m, m_1 và n nếu :

+ *X nhận các giá trị 0, 1, 2, ..., n*

+ $$P(X=r)=\frac{m_1!m_2!(m-n)!n!}{(m_1-r)!r!(m_2-n+r)!(n-r)!m!}=H(r,m,m_1,n).$$

Lấy mẫu không hoàn lại (không lặp), không thứ tự là mẫu tổ hợp không lặp.

Số thuận lợi cho x = r là $C_{m_1}^{r}\times C_{m_2}^{n-r}$.

Tổng số khả năng không phân biệt A hay là $\overline{A}$ là C_m^n.

Vậy $P(X=r)=\dfrac{C_{m_1}^{r}\times C_{m_2}^{n-r}}{C_m^n}$. Khai triển các tổ hợp được công thức trên.

3.2. Tính chất

3.2.1. Đại lượng ngẫu nhiên X có quy luật siêu bội với tham số m, m_1, n thì:

$$MX = n\times\frac{m_1}{m}$$

$$DX = n\times\frac{m_1}{m}\times\frac{m_2}{m}\left(1-\frac{n-1}{m-1}\right).$$

Phương sai tính theo quy luật siêu bội thường nhỏ hơn phương sai tính theo quy luật nhị thức.

3.2.2. Đại lượng ngẫu nhiêu X có quy luật siêu bội với tham số m, m_1, n khi $m\to\infty$ sao cho $\frac{m_1}{m}\to p$, n cố định thì X có quy luật siêu bội sẽ tiến tới X có quy luật nhị thức với tham số n và p.

$$P(X=r)=\frac{m_1!\,m_2!\,(m-n)!\,n!}{(m_1-r)!\,r!\,(m_2-n+r)!\,(n-r)!\,m!}\to C_n^r\,p^r(1-p)^{n-r}.$$

Ví dụ:

1. Xét nghiệm nhóm máu cho 100 người, trong đó 48% có nhóm máu O. Mỗi lần xét nghiệm cho 10 người, tìm xác suất sao cho có 5 người có nhóm máu O.

Giải:

Xét nghiệm cho 10 người được thực hiện theo cách lấy mẫu không hoàn lại.

Gọi X là số người có nhóm máu O trong 10 xét nghiệm.

X có quy luật siêu bội với m = 100, m_1 = 48, n = 10.

$$P(X=5)=\frac{48!\;52!\;90!\;10!}{(48-5)!\;5!\;(52-10+5)!\;(10-5)!\;100!}=H(5,100,48,10)$$

$$=0{,}257$$

Nếu m = 100 đủ lớn, $\frac{m_1}{m}=0{,}48$.

Tìm xác suất theo nhị thức

$$P(X=5)=C_{10}^5\times 0{,}48^5\times 0{,}52^5=0{,}244.$$

Sai số $\delta = 0{,}257-0{,}244=0{,}013\to\frac{0{,}013}{0{,}257}=0{,}05$. Khi m lớn hơn sai số sẽ bé đi.

2. Kiểm tra X quang cho 120 người trong đó 10% bị lao. Mỗi lần kiểm tra 10 người, tìm xác suất sao cho có một người bị lao.

Giải:

Kiểm tra X quang cho 10 người được thực hiện theo cách lấy mẫu không hoàn lại.

Gọi X là số người bị lao trong 10 người kiểm tra.

X có quy luật siêu bội với $m = 120$, $m_1 = 12$, $n = 10$

$$P(X = 1) = \frac{12!\ \ 108!\ \ 110!\ \ 10!}{11!\ \ 1!\ \ 99!\ \ 9!\ \ 120!} = H(1, 120, 12, 10) = 0,404$$

Nếu m =120 là đủ lớn, $\frac{m_1}{m} = 0,1$.

Tính xác suất theo nhị thức

$$P(X = 1) = C_{10}^{1} \times 0,1^{1} \times 0,9^{9} = 0,387.$$

Sai số $\delta = 0,404 - 0,387 = 0,017 \rightarrow \frac{0,017}{0,404} = 0,04.$

Có thể xem m = 120 là đủ lớn.

4. QUY LUẬT ĐA THỨC

4.1. Định nghĩa

Thực hiện n phép thử độc lập. Kết quả là nhóm đầy đủ các hiện tượng A_1, A_2, ..., A_k tương ứng với các xác suất p_1, p_2, ..., p_k.

Gọi X_1, X_2, ..., X_k là số lần xuất hiện tương ứng các hiện tượng A_1, A_2, ..., A_k.

Các đại lượng ngẫu nhiên X_1, X_2, ..., X_k được gọi là k biến có quy luật đa thức với k tham số n, p_1, p_2, ..., p_k nếu:

+ *X_i nhận giá trị từ 0, 1, 2,..., n; $i = \overline{1,k}$*

+ $P(X_1 = r_1, X_2 = r_2, ..., X_k = r_k) = \dfrac{n!}{r_1!\ r_2!\ ...\ r_k!} p_1^{r_1}.p_2^{r_2}....p_k^{r_k}$

Trong n kết quả có r_1 lần A_1, r_2 lần A_2,...,r_k lần A_k, số trường hợp là số mẫu không lặp, không thứ tự $C_n^{r_1} \times C_{n-r_1}^{r_2} \times ... \times C_{r_k}^{r_k}$.

Khai triển các tổ hợp trên có kết quả: $\dfrac{n!}{r_1!\ r_2!\ ...\ r_k!}$.

Do n phép thử độc lập, cho nên xác suất của một trường hợp là tích xác suất $p_1^{r_1}.p_2^{r_2}....p_k^{r_k}$. Dẫn đến công thức đã nêu.

Chú ý: $X_1 + X_2 + ... + X_k = r_1 + r_2 + ... + r_k = n$.

4.2. Tính chất

4.2.1. Giả sử $X_1, X_2, ..., X_k$ là k biến có quy luật đa thức với các tham số n, $p_1, p_2, ..., p_{k-1}$ thì $X_1, X_2, ..., X_h$, S ($S = X_{h+1} + ... + X_k$) gồm h +1 biến có quy luật đa thức với các tham số n, $p_1, p_2, ..., p_h$

Áp dụng tính chất trên, X_1 và $S = \sum_{i=2}^{k} X_i$ là hai biến có quy luật đa thức với các tham số n và p_1.

$$P\{X_1 = r_1, S = n - r_1\} = \frac{n!}{r_1!\ (n - r_1)!} p_1^{r_1} \times (1 - p_1)^{n-r_1}$$

Như vậy X_1 là đại lượng ngẫu nhiên có quy luật nhị thức với tham số n và p_1.

Dẫn đến: $MX_1 = np_1$, $DX_1 = np_1(1 - p_1)$.

Tương tự như trên X_i là đại lượng ngẫu nhiên có quy luật nhị thức với tham số n và p_i.

Dẫn đến: $MX_i = np_i$, $DX_i = np_i(1 - p_i)$ $i = \overline{1,k}$.

4.2.2. Giả sử $X_1, X_2, ..., X_k$ là k biến có quy luật đa thức với k tham số n, $p_1, p_2, ..., p_{k-1}$ thì

$$P(X_1 = n, X_i = 0) = \frac{n!}{n!\ 0!\ ...\ 0!} p_1^n.p_2^0....p_k^0 = p_1^n$$

$$P(X_2 = n, X_i = 0) = p_2^n$$

$$P(X_1 = n, X_2 = n, X_i = 0) = 0 \neq p_1^n . p_2^n .$$

Vậy X_1, X_2 là hai biến trong k biến có quy luật đa thức không độc lập với nhau.

Ví dụ:

1. Tỷ lệ nhóm máu AB bằng 0,04, nhóm máu A bằng 0,2, nhóm máu B bằng 0,28, nhóm máu O bằng 0,48. Xét nghiệm nhóm máu cho 10 người, tìm xác suất sao cho có một người nhóm máu AB, hai người nhóm máu A, hai người nhóm máu B và năm người nhóm máu O.

Giải:

Xét nghiệm nhóm máu cho 10 người là các phép thử độc lập.

Gọi X_1, X_2, X_3, X_4 là số người có nhóm máu tương ứng AB, A, B và O.

X_1, X_2, X_3, X_4 là 4 biến có quy luật đa thức với các tham số n =10 và $p_1 = 0{,}04$; $p_2 = 0{,}2$; $p_3 = 0{,}28$; $p_4 = 0{,}48$.

$$P(X_1 = 1, X_2 = 2, X_3 = 2, X_4 = 5) = \frac{10!}{1!\ 2!\ 2!\ 5!} 0{,}04^1 \times 0{,}2^2 \times 0{,}28^2 \times 0{,}48^5$$

$$= 0{,}0242.$$

2. Lai hai cây hoa màu hồng dị hợp tử, thế hệ sau sẽ có hoa trắng, hoa hồng và hoa đỏ với các xác suất tương ứng: 0,25; 0,5 và 0,25. Theo dõi hoa của 10 cây thế hệ sau, tìm xác suất sao cho có 3 cây hoa trắng, 4 cây hoa hồng và 3 cây hoa đỏ.

Giải:

Các cây thế hệ sau có màu hoa độc lập với nhau.

Gọi X_1, X_2, X_3 là số cây có hoa trắng, hoa hồng, hoa đỏ tương ứng.

X_1, X_2, X_3 là 3 biến có quy luật đa thức với các tham số n = 10; $p_1 = 0{,}25$; $p_2 = 0{,}5$; $p_3 = 0{,}25$.

$$P(X_1 = 3, X_2 = 4, X_3 = 3) = \frac{10!}{3!\ 4!\ 3!} 0{,}25^3 \times 0{,}5^4 \times 0{,}25^3 = 0{,}064.$$

3. Ba cô A, B, C khéo léo như nhau. Sau một thời gian làm thủ thuật có 5 thủ thuật không đạt, tìm xác suất sao cho A làm không đạt 1, B làm không đạt 2 và C làm không đạt 2 thủ thuật.

Giải:

Ba cô làm thủ thuật không đạt độc lập với nhau

Gọi X_1, X_2, X_3 là số thủ thuật không đạt tương ứng của A, B, C.

Ba cô khéo léo như nhau nên xác suất không đạt của mỗi cô bằng $\frac{1}{3}$.

X_1, X_2, X_3 là 3 biến có quy luật đa thức với các tham số n = 10; $p_i = \frac{1}{3}$; i = 1, 2, 3.

$$P(X_1 = 1, X_2 = 2, X_3 = 2) = \frac{5!}{1!\ 2!\ 2!} \left(\frac{1}{3}\right)^1 \times \left(\frac{1}{3}\right)^2 \times \left(\frac{1}{3}\right)^2 = 0{,}123.$$

4.4. Kiểm định χ^2

Giả sử X_1, X_2 là số lần xuất hiện A_1, A_2 khi thực hiện n phép thử độc lập. X_1, X_2 là hai biến có quy luật đa thức với n và p_1, p_2.

X_1 là biến có quy luật nhị thức với tham số n và p_1.

Khi đó: $MX_1 = np_1$, $DX_1 = np_1(1 - p_1) = np_1p_2$.

Tương tự với X_2 có $MX_2 = np_2$, $DX_2 = np_2(1 - p_2) = np_1p_2$

Khi n đủ lớn X_i có quy luật nhị thức tiến tới quy luật chuẩn với tham số

$$\mu_i = MX_i = np_i, \qquad \sigma_i^2 = DX_i = np_1p_2 \qquad i = \overline{1,2}$$

$\dfrac{X_1 - np_1}{\sqrt{np_1p_2}}$ là biến chuẩn tắc. Vậy $\left(\dfrac{X_1 - np_1}{\sqrt{np_1p_2}}\right)^2$ là biến χ^2 với 1 bậc tự do.

$$Q_1 = \frac{(X_1 - np_1)^2}{np_1p_2} = \frac{(X_1 - np_1)^2}{np_1} + \frac{(X_1 - np_1)^2}{np_2}.$$

$$(X_1 - np_1)^2 = (n - X_2 - np_1)^2 = (-X_2 + n(1 - p_1))^2 = (X_2 - np_2)^2.$$

$$Q_1 = \frac{(X_1 - np_1)^2}{np_1} + \frac{(X_2 - np_2)^2}{np_2}.$$

Thay X_i là số lần xuất hiện A_i bởi m_i và $np_i = M_i$ dẫn đến $Q_1 = \sum\limits_{i=1}^{2}\dfrac{(m_i - M_i)^2}{M_i}$ là biến χ^2 với 1 bậc tự do.

Mở rộng cho k biến có quy luật đa thức $Q_{k-1} = \sum\limits_{i=1}^{k}\dfrac{(m_i - M_i)^2}{M_i}$ là biến χ^2 với k – 1 bậc tự do.

CÂU HỎI TỰ LƯỢNG GIÁ

Hãy chọn một kết quả đúng.

1. Tỷ lệ mắc bệnh B tại cộng đồng bằng 0,0016. Khám bệnh cho 100 người. Tìm xác suất của số M_0.

Kết quả:

A. 0,852 B. 0,863.456 C. 0,136.544 D. 0,148 E. số khác.

2. Điều trị bệnh B bằng kháng sinh I có tỷ lệ khỏi bằng 0,82, bằng kháng sinh II có tỷ lệ khỏi bằng 0,89. Điều trị phối hợp 2 kháng sinh trên cho 100 người bệnh B, tìm xác suất sao cho có 95 người khỏi.

Kết quả:

A. 0,727 B. 0,273 C. 0,034 D. 0,966 E. số khác

3. Dùng một phản ứng chẩn đoán bệnh, phản ứng có tỷ lệ chẩn đoán đúng bằng 0,8. Biết xác suất chẩn đoán bệnh cho 100 người (bằng phản ứng trên) có ít nhất 100 – m người được chẩn đoán đúng không nhỏ hơn 0,96. Tìm m.

Kết quả:

A. 87 B. 13 C. 73 D. 27 E. số khác

4. Khi nâng cấp khoa xét nghiệm, bác sỹ A đề nghị chia khoa thành 2 bộ phận độc lập vì cho rằng ít nhất một bộ phận làm việc vẫn đảm bảo được xét nghiệm cho bệnh nhân; Bác sỹ B đề nghị chia khoa thành 4 bộ phận độc lập vì cho rằng ít nhất 2 bộ phận làm việc vẫn đảm bảo được xét nghiệm cho bệnh nhân. Xác suất để một bộ phận làm việc là p, không làm việc là q (q = 1 – p). Tìm q sao cho cách chia 2 có khả năng làm việc tốt hơn cách chia 4

Kết quả:

A. ≤ 1/3 B. > 1/3 C. 1/3 D. 2/3 E. số khác.

Bài 5

LUẬT SỐ LỚN

MỤC TIÊU

Trình bày được ước lượng $P(A) \approx \omega(A)$ *và* $MX \approx \overline{x}$.

Khi thực hiện một phép thử, không thể đoán trước một cách chắc chắn xem đại lượng ngẫu nhiên sẽ nhận giá trị nào trong các giá trị nó có thể nhận.

Khi thực hiện nhiều phép thử với số lượng lớn các đại lượng ngẫu nhiên, có thể dễ đoán nhận kết quả của phép thử. Khi đó tính ngẫu nhiên của phép thử sẽ "mất đi" và cho quy luật "tất nhiên" của hiện tượng.

1. BẤT ĐẲNG THỨC TREBUSEV

- Bất đẳng thức

Nếu đại lượng ngẫu nhiên X có trung bình lý thuyết (Kì vọng toán) và phương sai hữu hạn thì với mọi số dương ε bé tuỳ ý, ta luôn có:

$$P\{|X - MX| < \varepsilon\} \geq 1 - \frac{DX}{\varepsilon^2}. \quad (1)$$

- Chứng minh

Giả sử X là đại lượng ngẫu nhiên rời rạc. X nhận các giá trị $X_1, X_2, \ldots, X_n$ tương ứng với các xác suất $p_1, p_2, \ldots, p_n$.

Giả sử với k giá trị đầu $|X_i - MX| < \varepsilon$, $i = \overline{1,k}$ và (n – k) giá trị còn lại thì $|X_i - MX| \geq \varepsilon$, $i = \overline{k+1,n}$.

Hai hiện tượng trên là 2 hiện tượng đối lập, dẫn đến

$$P\{|X - MX| < \varepsilon\} = 1 - P\{|X - MX| \geq \varepsilon\} \quad (2)$$

Vì X là đại lượng ngẫu nhiên rời rạc nên

$$DX = \sum_{i=1}^{n} p_i (x_i - MX)^2$$

$$DX \geq \sum_{i=k+1}^{n} p_i (x_i - MX)^2 \quad \text{(vì các số hạng không âm)}$$

$$\geq \sum_{i=k+1}^{n} p_i \varepsilon^2$$

$$\geq \varepsilon^2 \sum_{i=k+1}^{n} p_i .$$

Theo công thức cộng xác suất, $\sum_{i=k+1}^{n} p_i$ là xác suất để đại lượng ngẫu nhiên X nhận một trong các giá trị x_{k+1}, x_{k+2}, ..., x_n mà mọi giá trị trên đều thoả mãn bất đẳng thức $|x_i - MX| \geq \varepsilon$.

$$\sum_{i=k+1}^{n} p_i = P\{|X - MX| \geq \varepsilon\}$$

$$DX \geq \varepsilon^2 P\{|X - MX| \geq \varepsilon\}$$

$$P\{|X - MX| \geq \varepsilon\} \leq \frac{DX}{\varepsilon^2} \qquad (3)$$

$$P\{|X - MX| < \varepsilon\} \geq 1 - \frac{DX}{\varepsilon^2}$$

- Nhận xét

*Có thể xem (3) là dạng khác của bất đẳng thức Trebưsev.

*Bất đẳng thức cũng đúng cho trường hợp X là đại lượng ngẫu nhiên liên tục.

* Nếu $DX \geq \varepsilon^2$ thì (1) là hiển nhiên, điều này không có giá trị trong thực tiễn.

Bất đẳng thức Trebưsev cho cận trên hoặc cận dưới xác suất để đại lượng ngẫu nhiên X nhận giá trị sai lệch so với trung bình lý thuyết của nó lớn hơn hay nhỏ hơn ε.

Bất đẳng thức Trebưsev được sử dụng để chứng minh các định lý về luật số lớn.

2. ĐỊNH LÝ TREBƯSEV

- Định lý:

Giả sử các đại lượng ngẫu nhiên X_1, X_2, ..., X_n độc lập, có các trung bình lý thuyết hữu hạn và các phương sai bị chặn trên bởi hằng số c ($DX_i \leq c, \forall i = \overline{1,n}$) thì với mọi số dương ε bé tuỳ ý ta luôn có:

$$\lim_{n \to \infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} MX_i\right| < \varepsilon\right\} = 1. \qquad (4)$$

- Chứng minh

Ký hiệu $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Tìm trung bình lý thuyết và phương sai của $\overline{X}$.

$$M\overline{X} = M\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} MX_i$$

$$D\overline{X} = D\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} DX_i .$$

Áp dụng bất đẳng thức Trebưsev

$$P\left\{ \left|\overline{X} - M\overline{X}\right| < \varepsilon \right\} \geq 1 - \frac{D\overline{X}}{\varepsilon^2} = 1 - \frac{\sum_{i=1}^{n} DX_i}{n^2\varepsilon^2}$$

$$\geq 1 - \frac{nc}{n^2\varepsilon^2} = 1 - \frac{c}{n\varepsilon^2} .$$

Lấy giới hạn 2 vế:

$$\lim_{n\to\infty} P\left\{ \left|\overline{X} - M\overline{X}\right| < \varepsilon \right\} \geq \lim_{n\to\infty}\left[1 - \frac{c}{n\varepsilon^2}\right] = 1 .$$

Do xác suất của một hiện tượng không vượt quá 1, vậy

$$\lim_{n\to\infty} P\left\{ \left|\overline{X} - M\overline{X}\right| < \varepsilon \right\} = 1 .$$

- Hệ quả

Giả sử các đại lượng ngẫu nhiên $X_1, X_2, \ldots, X_n$ độc lập, có cùng trung bình lý thuyết $(MX_i = \mu \quad \forall i = \overline{1,n})$ và các phương sai cùng bị chặn trên $(DX_i \leq c$, $\forall i = \overline{1,n})$ thì với mọi số dương ε bé tuỳ ý ta luôn có

$$\lim_{n\to\infty} P\left\{ \left|\overline{X} - \mu\right| < \varepsilon \right\} = 1 . \quad (5)$$

Về bản chất, định lý Trebưsev chứng tỏ sự hội tụ theo xác suất của trung bình số học của số lớn các đại lượng nhẫu nhiên về trung bình số học của các trung bình lý thuyết của chúng.

Như vậy trung bình số học của các đại lượng ngẫu nhiên ổn định xung quanh trung bình số học của các trung bình lý thuyết của các đại lượng ngẫu nhiên.

Từng đại lượng ngẫu nhiên có thể nhận giá trị khác biệt nhiều so với trung bình lý thuyết của chúng song trung bình số học của số lớn các đại lượng ngẫu nhiên lại nhận giá trị gần bằng trung bình số học của các trung bình lý thuyết của chúng với xác suất rất lớn. Do đó có thể dự đoán được trung bình số học của các đại lượng ngẫu nhiên.

Trong thực tiễn để có giá trị của một đại lượng thường đo đại lượng nhiều lần. Để có $M\overline{X}$ ta thống kê số lượng lớn các mẫu từ đó có được $\overline{X}$ và gán nó cho $M\overline{X}$.

3. ĐỊNH LÝ BERNOULLI

- Định lý

Giả sử W là tần suất xuất hiện hiện tượng A trong n phép thử độc lập và p là xác suất xuất hiện hiện tượng A trong mỗi phép thử đó thì với mọi ε dương bé tuỳ ý ta luôn có

$$\lim_{n\to\infty} P\{|W-P|<\varepsilon\}=1. \qquad (6)$$

- Chứng minh

Gọi X là số lần xuất hiện hiện tượng A trong n phép thử độc lập.

Xét đại lượng ngẫu nhiên $W=\frac{X}{n}$, là tần suất xuất hiện hiện tượng A trong n phép thử độc lập

Ta có: $MW=M(\frac{X}{n})=\frac{1}{n}MX$

$$DW=D(\frac{X}{n})=\frac{1}{n^2}DX.$$

X là đại lượng ngẫu nhiên có quy luật nhị thức với các tham số n và p do đó MX = np, DX = npq.

Vậy $MW=\frac{1}{n}.np=p$, $DW=\frac{1}{n^2}.npq=\frac{pq}{n}$.

Áp dụng bất đẳng thức Trebưsev đối với đại lượng ngẫu nhiên W

$$P\{|W-p|<\varepsilon\}\ge 1-\frac{pq}{n\varepsilon^2}.$$

Lấy giới hạn hai vế

$$\lim_{n\to\infty} P\{|W-p|<\varepsilon\}\ge \lim_{n\to\infty}\left[1-\frac{pq}{n\varepsilon^2}\right]=1.$$

Mặt khác do xác suất của một hiện tượng không vượt quá 1 dẫn đến

$$\lim_{n\to\infty} P\{|W-p|<\varepsilon\}=1.$$

- Nhận xét

Định lý chứng tỏ sự hội tụ theo xác suất của tần suất xuất hiện hiện tượng trong n phép thử độc lập về xác suất xuất hiện hiện tượng đó trong mỗi phép thử khi số phép thử tăng lên vô hạn. Như vậy tần suất ổn định xung quanh xác suất.

Định lý là cơ sở của định nghĩa thống kê của xác suất.

Ví dụ:

1. Đo chiều cao của 125 nam thanh niên. Tìm xác suất sao cho độ lệch giữa chiều cao trung bình và trung bình lý thuyết không vượt quá 2cm. Biết $DX = 4{,}7^2$.

Giải:

$$P\left\{ \left|\bar{X} - MX\right| < 2 \right\} \ge 1 - \frac{4{,}7^2}{125 \times 2^2} = 0{,}95582 .$$

2. Điều trị cho 500 người. Tìm xác suất sao cho độ lệch giữa tần suất khỏi và xác suất khỏi không vượt quá 0,05. Biết xác suất khỏi khi điều trị bằng 0,85

Giải:

$$P\left\{ \left|W - 0{,}85\right| < 0{,}05 \right\} \ge 1 - \frac{0{,}85 \times 0{,}15}{500 \times 0{,}05^2} = 0{,}898 .$$

Chương 2

THỐNG KÊ

Bài 1

THAM SỐ MẪU

MỤC TIÊU

1. Trình bày được công thức định nghĩa và công thức tính các tham số mẫu.

2. Tính được các tham số mẫu và nêu được ý nghĩa của chúng.

1. CÁC KHÁI NIỆM

- Khoảng số thực

khoảng đóng [a, b] = {x là số thực : $a \le x \le b$}

khoảng nửa đóng nửa mở

[a, b) = {x là số thực : $a \le x < b$}

hoặc (a, b] = {x là số thực : $a < x \le b$}

khoảng mở (a, b) = {x là số thực : $a < x < b$}.

- Ký hiệu tổng:

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \ldots + x_n$$

$$\sum_{i=1}^{n} (x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$$

$$\sum_{i=1}^{n} a x_i = a \sum_{i=1}^{n} x_i$$

$$\sum_{i=1}^{n} a = n\, a.$$

- Tập hợp tổng quát và tập hợp mẫu

Tập hợp tổng quát là tập hợp bao gồm tất cả các đối tượng cần nghiên cứu. Số phần tử của tập hợp tổng quát gọi là kích thước tập hợp tổng quát, ký hiệu là N.

Vì các điều kiện hạn chế, thường lấy ra một mẫu để nghiên cứu. Tập hợp mẫu là tập hợp gồm các đối tượng lấy ra để nghiên cứu. Số phần tử của tập hợp mẫu gọi là kích thước mẫu, ký hiệu n. Nói chung $N \geq n$.

Cần lấy mẫu ngẫu nhiên, khách quan sao cho tính chất của tập hợp mẫu phản ánh đúng tính chất tập hợp tổng quát.

Có hai cách lấy các phần tử ra để nghiên cứu. Lấy có hoàn lại là lấy ra một phần tử để nghiên cứu rồi trả lại tập hợp mẫu. Kết quả các lần nghiên cứu sau không phụ thuộc các kết quả nghiên cứu trước đó, phép thử độc lập. Lấy không hoàn lại là lấy ra một phần tử để nghiên cứu sau đó không trả lại tập hợp mẫu. Kết quả các nghiên cứu sau phụ thuộc kết quả các nghiên cứu trước, phép thử không độc lập.

- Dấu hiệu nghiên cứu

Khi nghiên cứu chỉ quan tâm xem xét một số mặt, một số tính chất của đối tượng nghiên cứu. Các đặc tính, tính chất cần nghiên cứu gọi là dấu hiệu nghiên cứu. Có dấu hiệu nghiên cứu về chất, có dấu hiệu nghiên cứu về lượng. Các dấu hiệu về chất được nghiên cứu khả năng xuất hiện của chúng, các dấu hiệu về lượng được tính các tham số mẫu.

2. SẮP XẾP SỐ LIỆU

Khi tiến hành nghiên cứu, số liệu thu được theo thứ tự thời gian. Như vậy số liệu chưa có thứ tự theo giá trị. Trước khi tính các tham số mẫu, số liệu được sắp xếp theo thứ tự giá trị.

Việc sắp xếp lại số liệu không làm thay đổi kết quả tính. Có những bài toán mà thuật toán đòi hỏi phải giữ nguyên thứ tự thu được theo thời gian thì không được sắp xếp lại số liệu.

Kí hiệu n giá trị thu được ban đầu là x_i, $i = \overline{1, n}$. Có những cách sắp xếp số liệu sau :

Sắp xếp số liệu thành dãy tăng hoặc bằng gọi là dãy không giảm

$$x_1 \leq x_2 \leq x_3 \leq \ldots \leq x_n \quad (1)$$

Sắp xếp số liệu thành dãy giảm hoặc bằng gọi là dãy không tăng

$$x_1 \geq x_2 \geq x_3 \geq \ldots \geq x_n \quad (2)$$

Có thể sắp xếp số liệu thành dãy các giá trị khác nhau tăng dần tương ứng với tần số xuất hiện của chúng.

$$\begin{array}{llll} x_1 & x_2 & \ldots & x_k \\ m_1 & m_2 & \ldots & m_k \end{array} \qquad \text{với} \quad \sum_{i=1}^{k} m_i = n \quad (3)$$

Với những nghiên cứu có kích thước mẫu n rất lớn, để tính các tham số mẫu thuận tiện mà sai số không đáng kể, có thể phân chia số liệu thành nhiều lớp.

Gọi k là số lớp cần phân chia : $k \geq 1 + 3{,}32 \lg n$.

Gọi R_x là miền biến thiên của x: $R_x = \max\limits_{i=\overline{1,n}} x_i - \min\limits_{i=\overline{1,n}} x_i$.

Gọi khoảng rộng của mỗi lớp là Δx

$$\Delta x \leq \frac{R_x}{k}$$

Như vậy sai số $\delta = \frac{\Delta x}{2}$. Với Δx đã biết, phân chia số liệu vào các lớp từ α_{i-1} đến α_i.

Kết quả thu được dãy giá trị giữa các lớp tương ứng với tần số xuất hiện của lớp:

$$x_1 \quad x_2 \quad \ldots\ldots \quad x_k$$

$$m_1 \quad m_2 \quad \ldots\ldots \quad m_k \qquad \text{với} \quad \sum_{i=1}^{k} m_i = n$$

Đôi khi từ số liệu thu được, chọn δ sao cho phù hợp với số liệu, từ đó có:

$\Delta x = 2\delta$, sau đó phân chia số liệu vào các lớp như trên.

Gọi x là áp lực động mạch phổi thì tâm thu bệnh nhân hẹp hai lá (mmHg).

Đo 153 bệnh nhân, $\max\limits_{\forall i} x_i = 157$, $\min\limits_{\forall i} x_i = 15$

$$R_x = 157 - 15 = 142$$

$$k \geq 1 + 3{,}32 \lg 153 = 8{,}2 .$$

Lấy k = 9 $\quad \Delta x \leq \frac{142}{9} = 15{,}77 \qquad \Rightarrow \quad \Delta x = 15$.

Sắp xếp số liệu vào 9 lớp được kết quả sau:

$(\alpha_{i-1} - \alpha_i)$	13 – 28	28 – 43	43 – 58	58 – 73	73 – 88	88 – 103	103 – 118	118 – 133	133 – 148	148 – 163
x_i	20,5	35,5	50,5	65,5	80,5	95,5	110,5	125,5	140,5	155,5
m_i	6	20	33	24	28	12	17	8	4	1

$$\sum_{i=1}^{10} m_i = 153$$

Chú ý : Từ số liệu chia k lớp sẽ thành k + 1 lớp.

Tính các tham số mẫu khi chia lớp sẽ có sai số.

3. CÁC THAM SỐ MẪU

Trong phần này chỉ nêu các tham số mẫu thường dùng. Đó là trung bình mẫu, phương sai và độ lệch mẫu.

3.1. Trung bình mẫu $\overline{x}$

- Định nghĩa và công thức tính

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \text{theo (1)} \qquad (4)$$

$$= \frac{1}{n}\sum_{i=1}^{k} m_i\, x_i \qquad \text{theo (3)} \qquad (5)$$

$$= x_0 + \Delta x \,.\, \frac{1}{n}\sum_{i=1}^{k} m_i\, u_i = x_0 + \Delta x\overline{u}\,. \qquad (6)$$

Trong (6) $u_i = \dfrac{x_i - x_0}{\Delta x}$ với x_0 và Δx tuỳ chọn.

Từ (5) suy ra (6) bằng cách thay $x_i = \Delta x.u_i + x_0$ vào (5)

$$\frac{1}{n}\sum_{i=1}^{k} m_i(\Delta x.u_i + x_0) = \frac{\Delta x}{n}\sum_{i=1}^{k} m_i u_i + \frac{x_0}{n}\sum_{i=1}^{k} m_i = x_0 + \Delta x\frac{1}{n}\sum_{i=1}^{k} m_i u_i$$

Trung bình cộng là trị số bình quân của các giá trị khác nhau, nhưng thuộc cùng một loại.

- $\overline{x}$ có cùng đơn vị x_i. Số thập phân của $\overline{x}$ hơn số thập phân của x_i một chữ số.
- $\overline{x}$ là tâm quần tụ của tập hợp mẫu.
- Tính chất

$$y_i = x_i + x_0 \;\Rightarrow\; \overline{y} = \overline{x} + x_0 \;\Leftrightarrow \overline{x} = \overline{y}\text{-}x_0$$

$$y_i = \frac{x_i}{\Delta x}\quad (\Delta x \neq 0) \;\Rightarrow\; \overline{y} = \frac{\overline{x}}{\Delta x} \;\Leftrightarrow\; \overline{x} = \Delta x\overline{y}$$

$$z_i = y_i + x_i \;\Rightarrow\; \overline{z} = \overline{y} + \overline{x}\,.$$

3.2. Phương sai s^2, độ lệch mẫu s

- Định nghĩa và công thức tính

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 \;\text{ theo (1)} \qquad (7)$$

$$= \frac{1}{n-1}\sum_{i=1}^{k} m_i(x_i - \overline{x})^2 \qquad \text{theo (3)} \qquad (8)$$

$$= \frac{1}{n(n-1)}\left[n\sum_{i=1}^{k} m_i x_i^2 - \left(\sum_{i=1}^{k} m_i x_i\right)^2\right] \qquad (9)$$

$$= \frac{\Delta x^2}{n(n-1)}\left[n\sum_{i=1}^{k} m_i u_i^2 - \left(\sum_{i=1}^{k} m_i u_i\right)^2\right], \qquad (10)$$

trong đó $u_i = \dfrac{x_i - x_0}{\Delta x}$ với Δx, x_0 tuỳ chọn, $\Delta x \neq 0$.

Từ (8), sau khi bình phương và thay $\overline{x} = \dfrac{1}{n}\sum_{i=1}^{k} m_i x_i$ suy ra (9).

Trong (9) thay $x_i = \Delta x.u_i + x_0$ dẫn đến

$$s^2 = \frac{1}{n(n-1)}\left[n\sum_{i=1}^{k} m_i(\Delta x u_i + x_0)^2 - \left(\sum_{i=1}^{k} m_i(\Delta x u_i + x_0)\right)^2\right]$$

$$= \frac{1}{n(n-1)}\left[n\Delta x^2\sum_{i=1}^{k} m_i u_i^2 + 2nx_0\Delta x\sum_{i=1}^{k} m_i u_i + (nx_0)^2 - \Delta x^2\left(\sum_{i=1}^{k} m_i u_i\right)^2 - \right.$$
$$\left. -2nx_0\Delta x\sum_{i=1}^{k} m_i u_i - (nx_0)^2\right]$$

$$= \frac{\Delta x^2}{n(n-1)}\left[n\sum_{i=1}^{k} m_i u_i^2 - \left(\sum_{i=1}^{k} m_i u_i\right)^2\right].$$ (10) được chứng minh.

- s^2 không cùng đơn vị với x_i.
- $s = \sqrt{s^2}$ được gọi là độ lệch mẫu.

s có cùng đơn vị và số thập phân với $\overline{x}$. Như vậy s^2 có số thập phân gấp hai số thập phân của s.

- s^2 là trung bình của bình phương khoảng lệch giữa x_i và $\overline{x}$ cho nên gọi tắt là phương sai. s^2 hay s cho biết mức độ tản mạn của x_i so với tâm của mẫu là $\overline{x}$, như vậy cũng cho biết độ đại diện của $\overline{x}$ cho các x_i tốt hay không. Khi đo một đại lượng nhiều lần, s^2 và s cho biết độ chính xác của các giá trị đo được, s^2 hay s được xem là sai số của cách đo.

s và $\overline{x}$ cùng đơn vị, có cùng số thập phân. Người ta thường viết $\overline{x} \pm s$ đại diện cho mẫu thu được.

Công thức (6) và (10) được sử dụng khi các x_i lớn hoặc có số thập phân hoặc cách đều.

- Tính chất

$y_i = x_i + x_0 \Rightarrow s_y^2 = s_x^2$

$y_i = \dfrac{x_i}{\Delta x}$ $\quad (\Delta x \neq 0) \qquad s_y^2 = \dfrac{s_x^2}{\Delta x^2} \Leftrightarrow s_x^2 = \Delta x^2 s_y^2$

$z_i = x_i + y_i \Rightarrow s_z^2 = s_x^2 + s_y^2$ khi X và Y là hai đại lượng độc lập.

- Các công thức khác

Trong một số trường hợp, phương sai được cho dưới dạng sau:

$$s^{*2} = \frac{1}{n}\sum_{i=1}^{k} m_i (x_i - MX)^2 \text{ với MX đã biết.} \qquad (11)$$

s^{*2} được xem là phương sai lý thuyết DX của đại lượng ngẫu nhiên khi n đủ lớn.

Hoặc là $s_*^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{k} m_i (x_i - \overline{x})^2$

$$= \overline{x^2} - \overline{x}^2. \qquad (12)$$

s_*^2 là phương sai chệch của phương sai lý thuyết của đại lượng ngẫu nhiên. Cách viết (12) thường gặp trong các công thức tính tham số của đường cong hồi quy và hệ số tương quan tuyến tính.

3.3. Phương sai của k dãy giá trị

Trong các nghiên cứu đồng thời k đại lượng, số liệu được cho dưới dạng sau:

X_1	X_2	...	X_j	...	X_k
x_{11}	x_{12}	...	x_{1j}	...	x_{1k}
x_{21}	x_{22}	...	x_{2j}	...	x_{2k}
⋮	⋮	...	⋮	...	⋮
x_{i1}	x_{i2}	...	x_{ij}	...	x_{ik}
⋮	⋮	...	⋮	...	⋮
$x_{n_1 1}$	$x_{n_2 2}$	...	$x_{n_j j}$	...	$x_{n_k k}$

trong đó $\sum_{j=1}^{k} n_j = N$

Gọi $\overline{x}$ là trung bình chung của k dãy, $\overline{x}_j$ là trung bình của dãy thứ j

$$\overline{x} = \frac{1}{N}\sum_{j,i=1}^{k,n_j} x_{ij} \tag{13}$$

$$\overline{x}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} x_{ij}\ , \qquad j = \overline{1,k} \tag{14}$$

Tuỳ thuộc k dãy giá trị của cùng một đại lượng hay của k đại lượng khác nhau sẽ có tương ứng hai phương sai.

- Phương sai của k dãy giá trị của cùng một đại lượng $\tilde{S}^2$

$$\tilde{S}^2 = \frac{1}{k-1}\sum_{j=1}^{k} n_j(\overline{x}_j - \overline{x})^2 \tag{15}$$

$$= \frac{1}{k-1}\left[\sum_{j=1}^{k}\frac{1}{n_j}\left(\sum_{i=1}^{n_j} x_{ij}\right)^2 - \frac{1}{N}\left(\sum_{j,i=1}^{k,n_j} x_{ij}\right)^2\right] \tag{16}$$

$$= \frac{B-C}{k-1} \text{ với } B = \sum_{j=1}^{k}\frac{1}{n_j}\left(\sum_{i=1}^{n_j} x_{ij}\right)^2,\ C = \frac{1}{N}\left(\sum_{j,i=1}^{k,n_j} x_{ij}\right)^2.$$

$\tilde{S}^2$ là trung bình của bình phương khoảng lệch giữa trung bình của từng dãy và trung bình chung của k dãy.

Thực hiện bình phương công thức (15)

$$\tilde{S}^2 = \frac{1}{k-1}\sum_{j=1}^{k} n_j(\ \overline{x}_j^2 - 2\overline{x}_j\overline{x} + \overline{x}^2\ )$$

$$= \frac{1}{k-1}\left[\sum_{j=1}^{k}\frac{1}{n_j}\left(\sum_{i=1}^{n_j} x_{ij}\right)^2 - 2\overline{x}\sum_{j,i=1}^{k,n_j} x_{ij} + N\overline{x}^2\right]$$

$$= \frac{1}{k-1}\left[\sum_{j=1}^{k}\frac{1}{n_j}\left(\sum_{i=1}^{n_j} x_{ij}\right)^2 - \frac{1}{N}\left(\sum_{j,i=1}^{k,n_j} x_{ij}\right)^2\right].$$

Thu được công thức (16)

- Phương sai của k dãy giá trị của k đại lượng khác nhau thuộc cùng một loại S^2

$$S^2 = \frac{1}{N-k}\sum_{j,\,i=1}^{k,\,n_j}(x_{ij}-\overline{x}_j)^2 \tag{17}$$

$$= \frac{1}{N-k}\left[\sum_{j,i=1}^{k,n_j} x_{ij}^2 - \sum_{j=1}^{k}\frac{1}{n_j}\left(\sum_{i=1}^{n_j} x_{ij}\right)^2\right] \tag{18}$$

$$S^2 = \frac{A-B}{N-k}, \qquad \text{với} \quad A = \sum_{j,i=1}^{k,n_j} x_{ij}^2 \qquad \text{và B đã biết.}$$

S^2 là trung bình của bình phương các khoảng lệch giữa các giá trị trong dãy và trung bình của dãy.

Thực hiện bình phương công thức (17)

$$S^2 = \frac{1}{N-k}\sum_{j,\,i=1}^{k,\,n_j}(x_{ij}-\overline{x}_j)^2$$

$$= \frac{1}{N-k}\sum_{j,i=1}^{k.n_j}(x_{ij}^2 - 2\overline{x}_j x_{ij} + \overline{x}_j^{\,2})$$

$$= \frac{1}{N-k}\left[\sum_{j,i=1}^{k,n_j} x_{ij}^2 - 2\sum_{j=1}^{k}\overline{x}_j\sum_{i=1}^{n_j} x_{ij} + \sum_{j=1}^{k} n_j\left(\frac{1}{n_j}\sum_{i=1}^{n_j} x_{ij}\right)^2\right]$$

$$= \frac{1}{N-k}\left[\sum_{j,i=1}^{k,n_j} x_{ij}^2 - \sum_{j=1}^{k}\frac{1}{n_j}\left(\sum_{i=1}^{n_j} x_{ij}\right)^2\right]$$

Công thức (18) được chứng minh.

3.4. Các tham số khác

- Hệ số biến thiên C_v

$$C_v = \frac{s}{\overline{x}}\,(^0/_{00})$$

C_v cho biết độ chính xác tương đối giữa s so với $\overline{x}$. C_v là tỷ số, viết dưới dạng % hay $^0/_{00}$, cho phép so sánh độ chính xác tương đối giữa các đại lượng không cùng đơn vị.

▪ Số trung vị M_e: $M_e = \begin{cases} x_{\frac{n+1}{2}} & \text{n lẻ} \\ x_{\frac{n}{2}} \quad \text{hoặc} \quad x_{\frac{n}{2}+1} & \text{n chẵn} \end{cases}$

M_e là giá trị giữa của n giá trị đã sắp xếp.

▪ Số mốt M_0

$M_0 = x_i$ mà m_i lớn nhất trong các $m_1, m_2, \ldots, m_k$

M_0 là giá trị hay gặp nhất trong k giá trị $x_1, x_2, \ldots, x_k$.

Với số liệu chuẩn theo một nghĩa nào đấy thì $M_e = M_0 = \overline{x}$

Vậy M_e, M_0 là các giá trị cũng cho biết tâm của tập hợp mẫu.

▪ Trung bình nhân, Trung bình điều hoà.

Khi nghiên cứu thu được dãy số liệu $x_1\, x_2 \ldots x_n$.

Đôi khi sử dụng trung bình nhân $\overline{g}$ hoặc trung bình điều hoà $\overline{h}$ trong xử lý số liệu. Công thức tính có dạng sau:

$$\overline{g} = \sqrt[n]{x_1 x_2 \ldots x_n}$$

$$\overline{h} = \frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_n}.$$

Ví dụ:

1. Gọi X là áp lực động mạch phổi thời tâm trương người bình thường

Đo 30 người được kết quả sau:

Giá trị x_i (mm Hg)	2	3	4	5	6	7	8	9
Số người m_i	1	4	7	8	2	5	2	1

Tính các tham số của mẫu trên.

Giải:

Cách 1. Lập bảng tính $\overline{x}$ theo (5) và s_x^2 theo (9)

i	x_i	m_i	$m_i\, x_i$	$m_i\, x_i^2$
1	2	1	2	4
2	3	4	12	36
3	4	7	28	112
4	5	8	40	200
5	6	2	12	72
6	7	5	35	245
7	8	2	16	128
k = 8	9	1	9	81
Σ		**30**	**154**	**878**

$$\overline{x} = \frac{154}{30} = 5,133 \approx 5,1.$$

$$s_x^2 = \frac{1}{30\times 29}[30\times 878 - 154^2] = \frac{2624}{870} = 3,0161 \approx 1,74^2$$

$$C_v = \frac{1,74}{5,13} = 0,339$$

$$M_e = x_{\frac{30}{2}} = 5 \quad , \quad M_0 = 5.$$

Cách 2. Lập bảng kiểm tra, tính $\overline{x}$ theo (6) và s_x^2 theo (10).

Chọn $x_0 = 5$ và $\Delta x = 1$ dẫn đến $u_i = \frac{x_i - 5}{1} = x_i - 5$

i	x_i	m_i	u_i	$m_i u_i$	$m_i u_i^2$
1	2	1	–3	– 3	9
2	3	4	–2	– 8	16
3	4	7	–1	– 7	7
4	5	8	0	0	0
5	6	2	1	2	2
6	7	5	2	10	20
7	8	2	3	6	18
k = 8	9	1	4	4	16
Σ		**30**		**4**	**88**

$$\overline{x} = 5 + \frac{1}{30}\times 4 = 5,133 \approx 5,1.$$

$$s_x^2 = \frac{1^2}{30\times 29}[30\times 88 - 4^2] = \frac{2624}{870} = 3,0161 \approx 1,74^2$$

Các giá trị của $\overline{x}$ và s_x^2 trùng với các kết quả trên.

2. Gọi X là lượng Protein huyết thanh người bình thường (g/l). Điện di 17 mẫu của 17 người thu được kết quả sau:

Giá trị x_i (g/l)	6,9	7,2	7,6	8,2	8,5
Số người m_i	2	3	5	6	1

Tính các tham số của mẫu trên

Giải:

Lập bảng tính $\overline{x}$ theo (6) và s_x^2 theo (10) với $u_i = \dfrac{x_i - 7,5}{0,1}$ và $v_i = \dfrac{x_i - 8}{0,1}$

i	x_i	m_i	u_i	m_iu_i	$m_iu_i^2$	v_i	m_iv_i	$m_iv_i^2$
1	6,9	2	– 6	–12	72	–11	–22	242
2	7,2	3	– 3	– 9	27	–8	–24	192
3	7,6	5	1	5	5	–4	–20	80
4	8,2	6	7	42	294	2	12	24
k = 5	8,5	1	10	10	100	5	5	25
Σ		**17**		**36**	**498**		**– 49**	**563**

$$\overline{x} = 7,5 + \frac{0,1}{17} \times 36 = 8 + \frac{0,1}{17} \times (-49) = 7,71$$

$$s_x^2 = \frac{0,1^2}{17 \times 16}\left[17 \times 498 - 36^2\right] = \frac{0,1^2}{17 \times 16}\left[17 \times 563 - 49^2\right] = \frac{0,1^2 \times 7170}{272} = 0,2636 = 0,51^2$$

$$\overline{x} \pm s = 7,71 \pm 0,51 (g/l) \quad C_v = \frac{0,51}{7,71} = 0,066, \qquad M_e = 7,6, \qquad M_0 = 8,2\,.$$

3. Gọi X_1, X_2, X_3, X_4 là thời gian hết ký sinh trùng sốt rét trong máu (giờ) của bốn nhóm bệnh nhân điều trị theo bốn cách khác nhau. Kết quả nghiên cứu thu được số liệu sau:

X_1	18	37	46	46	46	51	62	78	85	90
X_2	38	41	41	42	43	44	45	50	50	52
X_3	36	48	50	52	58	60	60	68	74	74
X_4	36	38	40	42	48	60	62	70	72	72

Tính các tham số $\overline{x}$, s, $\tilde{S}^2, S^2$ của bốn dãy số liệu.

i	X_1	X_1^2	X_2	X_2^2	X_3	X_3^2	X_4	X_4^2
1	18	324	38	1444	36	1296	36	1296
2	37	1369	41	1681	48	2304	38	1444
3	46	2116	41	1681	50	2500	40	1600
4	46	2116	42	1764	52	2704	42	1764
5	46	2116	43	1849	58	3364	48	2304
6	51	2601	44	1936	60	3600	60	3600
7	62	3844	45	2025	60	3600	62	3844
8	78	6084	50	2500	68	4624	70	4900
9	85	7225	50	2500	74	5476	72	5184
10	90	8100	52	2704	74	5476	72	5184
Σ	**559**	**35.895**	**446**	**20.084**	**580**	**34.944**	**540**	**31.120**

$$\sum_{i=1}^{4} n_j = 10 + 10 + 10 + 10 = 40$$

$$\bar{x}_1 = \frac{559}{10} = 55,9 \qquad s_{x_1}^2 = \frac{1}{10 \times 9}\ [10 \text{ x } 35.895 - 559^2] = 516,3222 = 22,72^2$$

$$\bar{x}_2 = \frac{446}{10} = 44,6 \qquad s_{x_2}^2 = \frac{1}{10 \times 9}\left[10 \times 20.084 - 446^2\right] = 21,3778 = 4,62^2$$

$$\bar{x}_3 = \frac{580}{10} = 58 \qquad s_{x_3}^2 = \frac{1}{10 \times 9}\left[10 \times 34.944 - 580^2\right] = 144,8889 = 12,04^2$$

$$\bar{x}_4 = \frac{540}{10} = 54 \qquad s_{x_4}^2 = \frac{1}{10 \times 9}\left[10 \times 31.120 - 540^2\right] = 217,7778 = 14,76^2$$

$$\bar{x} = \frac{1}{40}(559 + 446 + 580 + 540) = 53,125.$$

$$A = 35.895 + 20.084 + 34.944 + 31.120 = 122.043$$

$$B = \frac{559^2}{10} + \frac{446^2}{10} + \frac{580^2}{10} + \frac{540^2}{10} = 113.939,7$$

$$C = \frac{1}{40}[559 + 446 + 580 + 540]^2 = 112.890,625$$

$$\tilde{S}^2 = \frac{1}{4-1}[113.939,7 - 112.890,625] = 349,6917$$

$$\tilde{S}^2 = \frac{1}{40-4}[122.043 - 113.939,7] = 225,0917.$$

Chú ý: Nếu k dãy số liệu của cùng một đại lượng, có thể đổi biến

$u_j = \dfrac{X_j - x_0}{\Delta x}$ với x_0 và Δx tuỳ chọn $j = \overline{1,k}$, tính toán sẽ thuận lợi hơn.

Khi đó $\tilde{S}^2 = \dfrac{\Delta x^2(B-C)}{k-1}$, B và C tính theo u_j.

Chú ý: Đôi khi giá trị trung bình không phản ánh đúng kết quả nghiên cứu như ở các ví dụ dưới đây.

4. Đánh giá một phương pháp điều trị ngoại khoa mới kéo dài 10 năm nhận thấy:

Năm 1, 2, 3 điều trị cho 47 bệnh nhân, kết quả tốt: 31 người

Năm 4, 5, 6, 7 điều trị cho 96 bệnh nhân, kết quả tốt: 71 người.

Năm 8, 9, 10 điều trị cho 64 bệnh nhân, kết quả tốt: 58 người.

Tỷ lệ tốt trung bình của phương pháp điều trị bằng $\frac{160}{207} = 0,773$. Từ năm 11 trở đi tỷ lệ điều trị tốt lớn hơn $\frac{58}{64}(90,6\%)$. Vậy giá trị trung bình không phản ánh đúng kết quả nghiên cứu.

5. Chỉ tiêu tuyển sinh vào khoa I (ĐK) năm 2000 của ĐH X là 260.

Số thí sinh đăng ký thi : 3267; Trung bình 13 thí sinh lấy 1 người.

Chỉ tiêu tuyển sinh vào khoa II (KTYH) của ĐH X là 50.

Số thí sinh đăng ký thi : 641; Trung bình 13 thí sinh lấy 1 người.

Chỉ tiêu tuyển sinh vào khoa III (YTCC) của ĐH X là 30.

Số thí sinh đăng ký thi : 1134; Trung bình 38 thí sinh lấy 1 người.

Thí sinh thi vào khoa III có nên chuyển sang thi vào khoa I không?

Để đỗ vào khoa I, mỗi thí sinh phải hơn ít nhất 3007 thí sinh khác.

Để đỗ vào khoa III, mỗi thí sinh chỉ phải hơn ít nhất 1104 thí sinh khác. Thí sinh thi vào khoa II không nên đổi nguyện vọng sang khoa khác vì khó hơn.

CÂU HỎI TỰ LƯỢNG GIÁ

Hãy chọn một kết quả đúng.

1. Định lượng Protein dịch não tủy người bình thường (X, đv mg%) thu được số liệu sau:

11	17	19	12	17	19	14	18	19	16	18	20
16	18	20	16	18	20	16	19	20	16	19	20
16	19	21	17	19	21	17	19	21	17	19	22

Tính $\overline{x} \pm s$ của số liệu trên theo công thức tính.

Kết quả:

A. 17,94±2,37; B. 17,94±2,40; C. 18,48±2,40; D. 18,48±2,37; E. số khác

2. Gọi X là áp lực trung bình của động mạch phổi bệnh nhân hẹp hai lá đơn thuần (đv: mmHg), nghiên cứu thu được số liệu sau:

x_i	13	23	33	43	53	63	73	83	93	103
m_i	5	20	27	24	25	23	15	10	4	2

Tính $\overline{x} \pm s$ của số liệu trên.

Kết quả:

A. 50,162±20,690; B. 49,839±20,690; C. 50,162±20,757; D. 49,839±20,757; E. số khác

3. Đếm nhịp tim (tần số tim) của trẻ nam 3 lứa tuổi thu được kết quả sau:

Nhóm I	9 tuổi	$n_1 = 30$	$\overline{x}_1 \pm s_1 = 72{,}77 \pm 4{,}60$
Nhóm II	10 tuổi	$n_2 = 45$	$\overline{x}_2 \pm s_2 = 72{,}47 \pm 5{,}06$
Nhóm III	11 tuổi	$n_3 = 32$	$\overline{x}_3 \pm s_3 = 73{,}63 \pm 5{,}42$.

Tính phương sai chung S^2 của 3 nhóm số liệu trên.

Kết quả:

A. 25,3800; B. 25,2674; C. 25,4891; D. 12,9012; E. số khác.

4. Theo dõi số chuột chết khi cho các lô chuột thí nghiệm sử dụng các liều thuốc có độc tăng dần thu được số liệu sau:

x_i(liều, mg/kg)	0,015	0,02	0,025	0,03	0,035	0,04
Số chuột mỗi lô	20	69	95	78	44	20
Số chết	0	11	50	61	37	20

Tính liều chết trung bình của số liệu trên (Số liệu Finney).

Kết quả:

A. 0,02846; B. 0,0247; C. 0,0253; D. 0,0255; E. số khác.

Bài 2

KIỂM ĐỊNH GIẢ THIẾT THỐNG KÊ

MỤC TIÊU

Trình bày được các bước của bài toán kiểm định.

Điều trị một bệnh bằng nhiều phương pháp, mỗi phương pháp có một tỷ lệ khỏi nhất định. Các tỷ lệ khỏi của các phương pháp có như nhau không ?

Định lượng Protein toàn phần trong máu trẻ suy dinh dưỡng trước và sau điều trị. Phương pháp điều trị có hiệu quả không ? Nói cách khác, lượng Protein toàn phần trung bình sau điều trị có cao hơn hẳn lượng Protein toàn phần trung bình trước điều trị không ?

Điều tra n đối tượng nghiên cứu thấy m đối tượng có đặc tính A. Khả năng xuất hiện hiện tượng A là p_o có đúng không ?

Trên đây là những bài toán kiểm định giả thiết thống kê.

1. GIẢ THIẾT VÀ ĐỐI GIẢ THIẾT

Trong bài toán kiểm định giả thiết thống kê, giả thiết cần kiểm định ký hiệu H_0, được nêu ra dưới dạng: các tỷ lệ như nhau, các trung bình như nhau... Các giả thiết đối lập với giả thiết H_0 gọi tắt là đối thiết, ký hiệu H_1. Đối giả thiết không như nhau hay khác nhau được gọi là đối giả thiết hai phía. Đối giả thiết lớn hơn hay nhỏ hơn là các đối giả thiết một phía. Tuỳ theo giá trị thu được trong nghiên cứu để đưa ra đối giả thiết một phía hay hai phía.

2. ĐIỀU KIỆN

Các bài toán khác nhau có những điều kiện khác nhau, song để đảm bảo tính đúng đắn và chính xác của kiểm định có một số điều kiện sau:

+ Điều kiện chuẩn.

+ Điều kiện n đủ lớn.

+ Điều kiện đám đông thuần nhất.

3. TÍNH GIÁ TRỊ CỦA ĐẠI LƯỢNG NGẪU NHIÊN

Đó là các giá trị của đại lượng ngẫu nhiên chuẩn T hoặc Student T_n hoặc đại lượng ngẫu nhiên χ^2... Các công thức tính được nêu trong từng bài toán cụ thể.

4. TRA GIÁ TRỊ TỚI HẠN

Trước hết cần chọn mức α, sau đó tra giá trị tới hạn tương ứng mức α đó. Mức α thường chọn là 0,05, cũng có khi chọn tới mức 0,01 hay 0,001.

Giá trị tới hạn chia miền giá trị của đại lượng ngẫu nhiên thành hai miền: miền có giá trị ứng với xác suất lớn $1 - \alpha$ là miền giữ giả thiết H_0, miền có giá trị ứng với xác suất bé α là miền bác giả thiết H_0. Tuỳ theo giá trị tính được của đại lượng ngẫu nhiên thuộc miền nào mà quyết định kết luận bài toán kiểm định.

5. CÁC XÁC SUẤT CỦA BÀI TOÁN KIỂM ĐỊNH

- H_0 đúng

Giá trị của đại lượng ngẫu nhiên thuộc miền giữ giả thiết. Xác suất giữ giả thiết khi giả thiết đúng gọi là độ tin cậy.

Giá trị của đại lượng ngẫu nhiên thuộc miền bác giả thiết. Xác suất bác giả thiết khi giả thiết đúng gọi là nguy hiểm loại I hay sai lầm loại I. Do H_0 đúng, sai lầm loại I chính là α, còn độ tin cậy là $1 - \alpha$.

- H_0 sai

Giá trị của đại lượng ngẫu nhiên thuộc miền giữ giả thiết. Xác suất giữ giả thiết khi giả thiết sai gọi là nguy hiểm loại II hay sai lầm loại II. Hàm sai lầm loại II ký hiệu là β.

Giá trị của đại lượng ngẫu nhiên thuộc miền bác giả thiết. Xác suất bác giả thiết khi giả thiết sai gọi là lực của kiểm định. Lực của kiểm định bằng $1 - \beta$.

- Khi α bé, $1 - \alpha$ lớn thì β sẽ lớn.

Nếu n đủ lớn thì α và β sẽ có giá trị nhỏ.

Khi n lớn, kinh phí nghiên cứu lớn vì vậy cần chọn n, α và β phù hợp với nhau; khuyến cáo nên chọn α mức 0,05.

Bài 3

SO SÁNH PHƯƠNG SAI, SO SÁNH TRUNG BÌNH CỦA HAI BIẾN CHUẨN

MỤC TIÊU

1. Giải được bài toán so sánh 2 phương sai, 2 trung bình.

2. Nêu được ý nghĩa bài toán.

1. SO SÁNH PHƯƠNG SAI

Nghiên cứu đại lượng ngẫu nhiên X thu được dãy giá trị $x_1, x_2 \ldots x_n$ (1)

Nghiên cứu đại lượng ngẫu nhiên Y thu được dãy giá trị $y_1, y_2 \ldots y_m$ (2)

Độ chính xác của các số liệu của hai đại lượng hoặc độ tản mạn của hai dãy số liệu của hai đại lượng có như nhau không ?

Giải bài toán trên cần so sánh phương sai của hai đại lượng ngẫu nhiên X và Y.

1.1. Tính tham số mẫu

- Tính tham số mẫu của dãy (1) : $\overline{x} \pm s_x$ với n đã biết.
- Tính tham số mẫu của dãy (2) : $\overline{y} \pm s_y$ với m đã biết.

1.2. Các bước của bài toán

- Đưa ra giả thiết H_0: DX = DY và H_1: DX ≠ DY.
- Kiểm tra điều kiện: Đại lượng ngẫu nhiên X chuẩn; Đại lượng ngẫu nhiên Y chuẩn.
- Tính giá trị F.

$$\text{Giả sử } s_x^2 > s_y^2, \quad F_{n-1,\, m-1} = \frac{s_x^2}{s_y^2}. \tag{3}$$

- Tra bảng

$F_{n-1,\, m-1}$ trong (3) là giá trị của đại lượng ngẫu nhiên tuân theo quy luật Fisher – Snedecor với n – 1 và m – 1 bậc tự do.

Tra f(n – 1; m – 1; 0,05) trong bảng quy luật Fisher–Snedecor, n – 1 tra ở cột n_1 và có thể nội suy, m – 1 tra ở hàng n_2 và lấy giá trị gần nhất.

▪ Kết luận

$F_{n-1,\, m-1} \le f(n-1, m-1; 0,05)$: chấp nhận giả thiết H_0.

$F_{n-1,\, m-1} > f(n-1, m-1; 0,05)$: bác bỏ giả thiết H_0, chấp nhận đối giả thiết H_1.

Từ kết luận trên suy ra ý nghĩa của bài toán.

Ví dụ

Đo đường kính của viên thuốc (mm) do hai máy thuộc hai loại dập ra, thu được số liệu sau:

Máy I	X:	5,54	5,69	5,62	5,80	5,67	5,52	5,77	5,65		
Máy II	Y:	5,64	5,42	5,58	5,52	5,29	5,50	5,67	5,48	5,32	5,44

Độ chính xác của hai máy có như nhau không ?

Giải

1. Tham số mẫu của hai dãy số liệu

$n = 8, \quad \bar{x} \pm s_x = 5,658 \pm \sqrt{0,0098}$.

$m = 10, \quad \bar{y} \pm s_y = 5,486 \pm \sqrt{0,0156}$.

2. So sánh hai phương sai

▪ $H_0: DX = DY \qquad H_1: DX \ne DY$

▪ Điều kiện

Giả sử X tuân theo quy luật chuẩn.

Giả sử Y tuân theo quy luật chuẩn.

▪ Tính F

$$F = \frac{0,0156}{0,0098} = 1,59.$$

▪ Kết luận

Tra bảng quy luật Fisher – Snedecor f(10 – 1; 8 – 1; 0,05)

$$f(9; 7; 0,05) = \frac{1}{2}[f(8; 7; 0,05) + f(10; 7; 0,05)]$$

$$= \frac{1}{2}[3,73 + 3,63] = 3,68.$$

Kết luận: 1,59 < 3,68 : chấp nhận giả thiết H_0 nghĩa là hai phương sai như nhau. Hai máy có độ chính xác như nhau.

2. SO SÁNH HAI TRUNG BÌNH LÝ THUYẾT

Khi nghiên cứu thường gặp bài toán: giá trị trung bình của nhóm nam X có bằng giá trị trung bình của nhóm nữ Y không hoặc giá trị trung bình của nhóm điều trị cách một X có bằng giá trị trung bình của nhóm điều trị cách hai Y không.

Giải bài toán, cần so sánh giá trị trung bình lý thuyết của hai nhóm.

2.1. Tính tham số mẫu

Từ hai dãy số liệu thu được n giá trị của đại lượng ngẫu nhiên X và m giá trị của đại lượng ngẫu nhiên Y, cần tính $\overline{x} \pm s_x$ và $\overline{y} \pm s_y$.

2.2. So sánh hai trung bình lý thuyết

- Đặt giả thiết H_0: MX = MY.

Đặt giả thiết đối lập H_1: MX > MY (Trường hợp 1)

hoặc MX ≠ MY (Trường hợp 2)

- Kiểm tra điều kiện:

Đại lượng ngẫu nhiên X chuẩn.

Đại lượng ngẫu nhiên Y chuẩn.

- Tính giá trị T.

Công thức tính T phụ thuộc vào giá trị DX, DY của hai đại lượng X và Y có biết không.

2.2.1. Biết DX, DY: $DX = \sigma_x^2$, $DY = \sigma_y^2$

$$T = \frac{|\overline{x} - \overline{y}|}{\sqrt{\dfrac{\sigma_x^2}{n} + \dfrac{\sigma_y^2}{m}}} \qquad (1.1)$$

T là giá trị của đại lượng ngẫu nhiên chuẩn tắc.

- Kết luận:

Tra giá trị tới hạn t(α) ứng với (Trường hợp 1) hoặc t(α/2) ứng với (Trường hợp 2) trong bảng chuẩn, lấy α = 0,05.

Khi T ≤ t(α) hoặc t(α/2): chấp nhận giả thiết H_0.

Ngược lại T > t(α) hoặc t(α/2): bác bỏ giả thiết H_0, chấp nhận đối thiết H_1.

2.2.2. Không biết DX, DY, nhưng giả thiết rằng DX = DY

$$T = \frac{|\bar{x} - \bar{y}|}{s\sqrt{\frac{1}{n} + \frac{1}{m}}}. \quad (1.2)$$

Trong (1.2) s^2 là phương sai mẫu chung của hai dãy số liệu.

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 + \sum_{i=1}^{m}(y_i - \bar{y})^2}{n+m-2} = \frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2} \quad (2)$$

- Kết luận:

T là giá trị của đại lượng ngẫu nhiên có quy luật Student với n + m – 2 bậc tự do. Tra giá trị t(n + m –2; α) hoặc t(n + m –2; α/2) trong bảng Student.

Khi T ≤ t(n + m – 2; α) hoặc t(n + m – 2; α/2): chấp nhận giả thiết H_0.

Ngược lại T > t(n + m – 2; α) hoặc t(n + m – 2; α/2): bác bỏ giả thiết H_0, chấp nhận đối thiết H_1.

2.2.3. Không biết DX, DY

$$T = \frac{|\bar{x} - \bar{y}|}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}}. \quad (1.3)$$

- Kết luận:

T là giá trị của đại lượng ngẫu nhiên xấp xỉ Student, khi đó giá trị gần đúng τ (α) được tính theo công thức:

$$\tau(\alpha) = \frac{s_x^2 t(n-1;\alpha) + s_y^2 t(m-1;\alpha)}{s_x^2 + s_y^2} \quad (3)$$

τ(α/2) tính tương tự (3).

Khi T ≤ τ (α) hoặc τ (α/2): chấp nhận giả thiết H_0.

Ngược lại T > τ (α) hoặc τ (α/2): bác bỏ giả thiết H_0, chấp nhận đối thiết H_1.

Ví dụ

1. Gọi X là đường kính các viên thuốc do máy I dập có kết quả:

$$n = 8;\ \bar{x} \pm s_x = 5{,}658 \pm \sqrt{0{,}0098}.$$

Gọi Y là đường kính các viên thuốc do máy II dập có kết quả: m = 10;

$$\overline{y} \pm s_y = 5{,}486 \pm \sqrt{0{,}0156}\,.$$

Đường kính trung bình của các viên thuốc do hai máy dập ra có như nhau không?

Giải:

- $H_0: MX = MY$, $H_1: MX \neq MY$.
- Điều kiện

Đại lượng ngẫu nhiên X tuân theo quy luật chuẩn.

Đại lượng ngẫu nhiên Y tuân theo quy luật chuẩn.

- Tính T

Theo kết quả so sánh phương sai ở trên, ta có phương sai của biến X và biến Y là như nhau, nên cần tính phương sai chung của hai biến.

$$s^2 = \frac{(8-1)0{,}0098 + (10-1)0{,}0156}{8+10-2} = 0{,}1143^2,$$

$$T = \frac{5{,}658 - 5{,}486}{0{,}1143\sqrt{\frac{1}{8} + \frac{1}{10}}} = 3{,}173\,.$$

- Kết luận

Tra giá trị tới hạn $t(8 + 10 - 2; 0{,}05/2) = 2{,}120$,

$t(8 + 10 - 2; 0{,}01/2) = 2{,}921$.

Do $T = 3{,}173 > 2{,}921$: bác bỏ giả thiết H_0. Trung bình hai dãy số liệu khác nhau mức 99%.

Đường kính trung bình của các viên thuốc do hai máy dập ra là khác biệt có ý nghĩa thống kê. Không nên dùng hai máy để dập các viên thuốc. Nếu cần dùng cả hai máy thì phải chỉnh máy.

2. Định lượng Protein toàn phần trong huyết thanh bệnh nhi suy dinh dưỡng trước điều trị X và sau điều trị Y, thu được số liệu sau:

X(g/l) 55,8 53,3 30,1 51,0 37,8 68,6 57,7 59,1 49,4 35,4 53,4 42,7 21,2 28,3 57,3 42,4 61,4

Y(g/l) 60,4 58,7 28,9 48,0 39,7 68,8 57,5 70,4 56,8 40,6 57,3 44,3 32,2 47,7 77,0 55,1 66,1

Phương pháp điều trị có hiệu quả không ?

Giải

- Tính tham số mẫu

Trước điều trị n = 17, $\overline{x} \pm s_x = 47,35 \pm \sqrt{173,6564}$.

Sau điều trị m = 17, $\overline{y} \pm s_y = 53,5 \pm \sqrt{182,0925}$.

- H_0: MX = MY, H_1: MX < MY.
- Điều kiện :

Giả sử đại lượng ngẫu nhiên X tuân theo quy luật chuẩn.

Giả sử đại lượng ngẫu nhiên Y tuân theo quy luật chuẩn.

- Tính T

Giả sử hai phương sai như nhau, cần tính s^2.

$$s^2 = \frac{(17-1)173,6564 + (17-1)182,0925}{17+17-2} = 177,8745 = 13,34^2.$$

$$T = \frac{53,5 - 47,35}{13,34\sqrt{\frac{1}{17} + \frac{1}{17}}} = 1,344.$$

- Kết luận

Tra bảng Student t(17 + 17 –2; 0,05) ≈ t(30; 0,05) = 1,697.

T = 1,344 < 1,697. Giữ giả thiết mức 95%.

Lượng Protein toàn phần trong huyết thanh bệnh nhi trước và sau điều trị như nhau. Phương pháp điều trị chưa thật sự hiệu quả.

Chú ý: Khi quan niệm xác suất là giá trị trung bình của các tần suất thì có thể áp dụng thuật toán so sánh hai trung bình để so sánh hai tỷ lệ.

3. Điều trị phương pháp I cho 405 bệnh nhân có 328 người khỏi.

Điều trị phương pháp II cho 155 bệnh nhân có 122 người khỏi.

Tỷ lệ khỏi của hai phương pháp có như nhau không?

Giải

- Tính các xác suất

Gọi xác suất khỏi của phương pháp I là p_1 $P(\overline{A}) = 0,5$

Gọi xác suất khỏi của phương pháp II là p_2 $p_2 = \frac{122}{155} = 0,781$.

Ký hiệu $n_1 = 405$; $m_1 = 328$

$n_2 = 155$; $m_2 = 122$

- $H_0 : p_1 = p_2$, $H_1 : p_1 \neq p_2$

- Điều kiện n_1, n_2 đủ lớn.
- Tính T

$$T = \frac{\frac{m_1}{n_1} - \frac{m_2}{n_2}}{\sqrt{\frac{m_1+m_2}{n_1+n_2}(1-\frac{m_1+m_2}{n_1+n_2})(\frac{1}{n_1}+\frac{1}{n_2})}} = \frac{\frac{328}{405} - \frac{122}{155}}{\sqrt{\frac{328+122}{405+155}(1-\frac{328+122}{405+155})(\frac{1}{405}+\frac{1}{155})}}$$
$$= 0,607$$

với n_1, n_2 đủ lớn, T là giá trị của đại lượng ngẫu nhiên chuẩn tắc.

- Kết luận

Tra bảng chuẩn tắc t(0,05/2) = 1,96

Do T = 0,607 < 1,96 dẫn đến chấp nhận H_0. Tỷ lệ khỏi của 2 phương pháp điều trị như nhau.

- **Chú ý:** Đặt $H_1 : p_1 > p_2$ thì giá trị tới hạn tra $t(\alpha)$.

3. SO SÁNH TỪNG CẶP

Trên một đối tượng nghiên cứu có khi thu được hai giá trị của cùng một đại lượng. Cân nặng của trẻ suy dinh dưỡng trước và sau điều trị, đường huyết của bệnh nhân đái tháo đường trước và sau điều trị... là một cặp giá trị của cùng một đại lượng. Số liệu của n đối tượng nghiên cứu là n cặp giá trị (x_i, y_i) $i = \overline{1,n}$.

Phần trên đưa ra phương pháp so sánh hai trung bình của hai đại lượng, phần này tiến hành so sánh cặp hay còn gọi là so sánh hiệu.

3.1. Tính tham số mẫu

Giả sử hầu hết các giá trị y_i lớn hơn x_i, khi đó đặt Z = Y – X, như vậy có n giá trị z_i, $z_i = y_i - x_i$, $i = \overline{1,n}$.

Từ các giá trị z_i tính các tham số mẫu $\bar{z}$ và s_z^2.

3.2. Các bước

- $H_0: MZ = 0$, $H_1: MZ > 0$ (1)

  hoặc $MZ \neq 0$. (2)

- Điều kiện:

Giả sử đại lượng ngẫu nhiên Z có quy luật chuẩn.

- Tính giá trị T.

$$T = \frac{|\bar{z}|}{\sqrt{\frac{s_z^2}{n}}} = \frac{|\bar{z}|\sqrt{n}}{s_z} \qquad (3)$$

T là giá trị của đại lượng ngẫu nhiên có quy luật Student với n – 1 bậc tự do.

- Kết luận:

Tra bảng Student $t(n-1; \alpha)$ ứng với (1)

hoặc $t(n-1; \alpha/2)$ ứng với (2).

Khi $T \le t(n-1; \alpha)$ hoặc $t(n-1; \alpha/2)$: chấp nhận giả thiết H_0.

Ngược lại $T > t(n-1; \alpha)$ hoặc $t(n-1; \alpha/2)$: bác bỏ giả thiết H_0, chấp nhận đối thiết H_1.

Từ kết luận của bài toán kiểm định suy ra ý nghĩa y học.

Ví dụ

1. Gọi X là lượng Protein toàn phần trong huyết thanh bệnh nhi suy dinh dưỡng trước điều trị. Gọi Y là lượng Protein toàn phần trong huyết thanh bệnh nhi suy dinh dưỡng sau điều trị. Z là lượng Protein toàn phần trong huyết thanh bệnh nhi suy dinh dưỡng tăng lên sau điều trị. Như vậy Z = Y – X.

Từ 17 cặp giá trị trước và sau điều trị (Ví dụ 2.2 ở phần trước), suy ra Z nhận các giá trị sau:

Z(g/l)	4,6	5,4	–1,2	–3,0	1,9	0,2	–0,2	11,3	7,4
	5,2	3,9	1,6	11,0	19,4	19,7	12,7	4,7	

Sau điều trị lượng Protein toàn phần có thật sự tăng lên không ?

Giải

- Từ dãy số liệu tính được các tham số mẫu:

$$n = 17;\ \bar{z} \pm s_z = 6{,}153 \pm 6{,}694\ (g/l).$$

- $H_0: MZ = 0,\ H_1: MZ > 0$.

- Điều kiện:

Giả sử đại lượng ngẫu nhiên Z có quy luật chuẩn.

- Tính giá trị T.

$$T = \frac{6{,}153\sqrt{17}}{6{,}694} = 3{,}790.$$

- Kết luận:

Tra bảng $t(17-1; 0{,}05) = 1{,}746$; $t(17-1; 0{,}01) = 2{,}583$.

$T > t(16\ ;\ 0{,}01)$: bác giả thiết H_0 mức 99%, chấp nhận H_1. Vậy lượng Protein toàn phần thật sự có tăng lên sau điều trị. Phương pháp điều trị mang lại hiệu quả cao.

2. Đo giá trị p đồng thời tại hai điểm trên cơ thể 12 người bệnh. Gọi Z là hiệu giữa điểm I và điểm II, thu được kết quả sau:

Z	0,012	0,002	0,006	0,027	0,005	0,015
	0,001	0,003	– 0,016	– 0,007	0,003	– 0,015

Giá trị đo được tại hai điểm có như nhau không ?

Giải

- Tính tham số mẫu.

$$n = 12 \qquad \sum z_i = 0{,}036 \qquad \sum z_i^2 = 0{,}001712$$

$$\overline{z} \pm s_z = 0{,}003 \pm 0{,}012.$$

- $H_0: MZ = 0,\ H_1: MZ \neq 0.$

- Điều kiện:

Z là đại lượng ngẫu nhiên có quy luật chuẩn.

- Tính giá trị T.

$$T = \frac{0{,}003\sqrt{12}}{0{,}012} = 0{,}866.$$

T là giá trị của đại lượng ngẫu nhiên tuân theo quy luật Student với 11 bậc tự do.

- Kết luận

Tra bảng Student, $t(11;\ 0{,}05/2) = 2{,}201$

$T < t(11\ ;\ 0{,}05/2)$: giữ giả thiết H_0. Không có sự khác biệt giữa hai giá trị đo được tại hai nơi.

Bài 4

SO SÁNH CÁC TRUNG BÌNH CÁC BIẾN CHUẨN, KIỂM ĐỊNH GIÁ TRỊ TRUNG BÌNH LÝ THUYẾT

MỤC TIÊU

1. *Giải được bài toán so sánh các trung bình và kiểm định $MX = \mu_0$.*
2. *Tính được sai lầm loại II.*

1. SO SÁNH CÁC TRUNG BÌNH CÁC BIẾN CHUẨN (Phân tích phương sai)

Nghiên cứu k nhóm tương ứng k đại lượng, thu được bảng giá trị sau:

$$\begin{array}{cccc}
X_1 & X_2 \ldots\ldots & X_j \ldots\ldots & X_K \\
x_{11} & x_{12} \ldots\ldots & x_{1j} \ldots\ldots & x_{1k} \\
x_{21} & x_{22} \ldots\ldots & x_{2j} \ldots\ldots & x_{2k} \\
\cdot & \cdot & \cdot & \cdot \\
x_{i1} & x_{i2} \ldots\ldots & x_{ij} \ldots\ldots & x_{ik} \\
\cdot & \cdot & \cdot & \cdot \\
x_{n_1 1} & x_{n_2 2} \ldots\ldots & x_{n_j j} \ldots\ldots & x_{n_k k}
\end{array}$$

Các giá trị trung bình của k dãy có như nhau không ?

1.1. Tính $\tilde{S}^2$ và S^2

$$N = \sum_{j=1}^{k} n_j \qquad \overline{x}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} x_{ij} \quad j = \overline{1,k} \qquad \overline{x} = \frac{1}{N}\sum_{j,i=1}^{k,n_j} x_{ij}$$

$$A = \sum_{j,i=1}^{k,\,n_j} x_{ij}^2 \qquad B = \sum_{j=1}^{k} \frac{1}{n_j}\left(\sum_{i=1}^{n_j} x_{ij}\right)^2 \qquad C = \frac{1}{N}\left(\sum_{j,i=1}^{k,n_j} x_{ij}\right)^2$$

$$\tilde{S}^2 = \frac{B-C}{k-1} \qquad S^2 = \frac{A-B}{N-k}.$$

1.2. Các bước kiểm định

- Giả thiết và đối thiết

$H_0 : MX_1 = MX_2 = \ldots = MX_K$

H_1 : Các MX_i không đồng thời bằng nhau

- Điều kiện

$X_1, X_2, ..., X_K$ là k đại lượng ngẫu nhiên chuẩn.

- Tính F

$$F_{k-1,N-k} = \frac{\tilde{s}^2}{s^2}$$

F là giá trị của đại lượng ngẫu nhiên có quy luật Fisher – Snedecor với k – 1 và N – k bậc tự do.

- Kết luận

Tra bảng Fisher – Snedecor giá trị f(k – 1; N – k; 0,05)

Khi $F \le f(k-1; N-k; \alpha)$: chấp nhận giả thiết H_0.

Ngược lại $F > f(k-1; N-k; \alpha)$: bác bỏ giả thiết H_0, chấp nhận đối thiết H_1.

Ví dụ

Theo dõi thời gian khỏi (ngày) của ba nhóm bệnh nhân điều trị bằng ba cách thu được bảng số liệu sau:

X_1	10	12	14	11	13	12		
X_2	20	18	19	12	14	16	15	18
X_3	4	6	7	5	8	6	7	

Thời gian khỏi trung bình của ba cách điều trị có như nhau không ?

Giải

1. Tính các tham số $\tilde{s}^2$ và s^2

i	X_1	X_1^2	X_2	X_2^2	X_3	X_3^2
1	10	100	20	400	4	16
2	12	144	18	324	6	36
3	14	196	19	361	7	49
4	11	121	12	144	5	25
5	13	169	14	196	8	64
6	12	144	16	256	6	36
7			15	225	7	49
8			18	324		
Σ	72	874	132	230	43	275

$\bar{x}_j$	12	16,5	6,14
s_j^2	$1,4142^2$	$2,7255^2$	$1,3452^2$

$$N = 6+8+7 = 21 \qquad \bar{x} = \frac{1}{21}(72+132+43) = 11,76.$$

$$A = 874 + 2230 + 275 = 3379$$

$$B = \frac{72^2}{6} + \frac{132^2}{8} + \frac{43^2}{7} = 3306,1429$$

$$C = (72+132+43)^2 \times \frac{1}{21} = 2905,1905$$

$$\bar{s} = \frac{1}{3-1}[3306,1429 - 2905,1905] = 200,4762,$$

$$s^2 = \frac{1}{21-3}[3379 - 3306,1429] = 4,0476.$$

2. So sánh các trung bình

- Giả thiết và đối giả thiết

$H_0 : MX_1 = MX_2 = MX_3$

H_1 : Các MX_i không đồng thời bằng nhau

- Điều kiện

X_1, X_2, X_3 là các đại lượng ngẫu nhiên chuẩn.

- Tính F

$$F = \frac{200,4762}{4,0476} = 49,53$$

F là giá trị của đại lượng ngẫu nhiên có quy luật Fisher – Snedecor với 3 – 1 và 21 – 3 bậc tự do.

- Kết luận

Tra bảng Fisher – Snedecor f(2 ; 18 ; 0,05) = 3,55.

F > f(2 ; 18 ; 0,05): bác bỏ giả thiết H_0, chấp nhận đối thiết H_1.

Thời gian khỏi trung bình của ba cách điều trị không như nhau. Có cách điều trị khỏi nhanh, có cách điều trị lâu khỏi.

2. KIỂM ĐỊNH GIÁ TRỊ TRUNG BÌNH LÝ THUYẾT

Giả sử đại lượng ngẫu nhiên X có quy luật chuẩn với tham số $MX = \mu_0$.

Đo n giá trị của đại lượng ngẫu nhiên X thu được $x_1, x_2,..., x_n$. Từ n giá trị của X tính được tham số mẫu $\overline{x} \pm s_x$.

Với dãy số liệu thu được MX = μ_0 có đúng không ?

2.1. Các bước kiểm định

- Đặt giả thiết và đối thiết

H_0: MX = μ_0, H_1: MX > μ_0 (Trường hợp 1)

hoặc MX ≠ μ_0 (Trường hợp 2)

- Điều kiện của kiểm định

Đại lượng ngẫu nhiên X tuân theo quy luật chuẩn $N(\mu_0, \sigma^2)$.

- Tính T.

2.1.1. Biết DX = σ^2

$$T = \frac{|\overline{x} - \mu_0|}{\frac{\sigma}{\sqrt{n}}}$$

T là giá trị của đại lượng ngẫu nhiên tuân theo quy luật chuẩn tắc.

- Kết luận

Tra giá trị tới hạn t(α) ứng với (Trường hợp 1) hoặc t(α/2) ứng với (Trường hợp 2).

Khi T ≤ t(α) hoặc t(α/2) : chấp nhận giả thiết H_0.

Ngược lại T > t(α) hoặc t(α/2): bác bỏ giả thiết H_0, chấp nhận đối thiết H_1.

2.1.2. Không biết DX

$$T = \frac{|\overline{x} - \mu_0|}{\frac{s}{\sqrt{n}}}$$

T là giá trị của đại lượng ngẫu nhiên tuân theo quy luật Student với n –1 bậc tự do.

- Kết luận

Tra t(n – 1; α) ứng với (Trường hợp 1) hoặc t(n – 1; α/2) ứng với (Trường hợp 2).

Khi T ≤ t(n – 1; α) hoặc t(n – 1; α/2): chấp nhận giả thiết H_0.

Ngược lại T > t(n – 1; α) hoặc t(n – 1; α/2): bác bỏ giả thiết H_0, chấp nhận đối thiết H_1.

2.2. Các xác suất của bài toán kiểm định

Khi tiến hành bài toán kiểm định giả thiết thống kê, kết luận của bài toán kiểm định đúng hay sai phụ thuộc vào H_0 đúng hay sai. Trong phần này cần xét các xác suất liên quan tới kiểm định. Bài toán được giải với giả thiết:

$H_0: MX = \mu_0$; $H_1: MX \neq \mu_0$ và biết $DX = \sigma^2$.

2.2.1. Giả thiết H_0: MX = μ_0 đúng

- $|T| \leq t(\alpha/2)$: giữ giả thiết H_0.

Xác suất giữ giả thiết H_0 khi H_0 đúng gọi là độ tin cậy.

- $|T| > t(\alpha/2)$ bác giả thiết H_0.

Xác suất bác giả thiết H_0 khi H_0 đúng gọi là nguy hiểm loại I hay sai lầm loại I.

Do H_0 đúng cho nên sai lầm loại I là α và độ tin cậy là $1-\alpha$. Như vậy chọn α trong bài toán kiểm định chính là ấn định sai lầm loại I.

2.2.2. Giả thiết H_0: MX = μ_0 sai. Khi đó giả sử MX = μ đúng

- $|T| \leq t(\alpha/2)$: giữ giả thiết H_0.

Xác suất giữ giả thiết H_0 khi H_0 sai gọi là nguy hiểm loại II hay sai lầm loại II. Sai lầm loại II ký hiệu là β và phụ thuộc vào μ cho nên viết là $\beta(\mu)$.

- $|T| > t(\alpha/2)$ bác giả thiết H_0.

Xác suất bác giả thiết H_0 khi H_0 sai được gọi là lực của kiểm định.

2.2.3. Tính $\beta(\mu)$

$$\beta(\mu) = P\{|T| \leq t(\alpha/2)\} = P\{-t(\alpha/2) \leq \frac{\overline{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}} \leq t(\alpha/2)\}$$

$$= P\{\mu_0 - t(\alpha/2)\frac{\sigma}{\sqrt{n}} \leq \overline{x} \leq \mu_0 + t(\alpha/2)\frac{\sigma}{\sqrt{n}}\}.$$

Để tính xác suất trên, ta thừa nhận định lý sau:

Giả sử $X_1, X_2, ..., X_n$ là n biến chuẩn độc lập có cùng $MX_i = \mu$ và $DX_i = \sigma^2$ $i = 1...n$, thì $\overline{X}$ là đại lượng ngẫu nhiên chuẩn với tham số $M\overline{X} = \mu$ và $D\overline{X} = \frac{\sigma^2}{n}$, trong đó $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

$$\text{Vậy } \beta(\mu) = \Pi\left(\frac{\mu_0 + t(\alpha/2)\frac{\sigma}{\sqrt{n}} - \mu}{\sigma/\sqrt{n}}\right) - \Pi\left(\frac{\mu_0 - t(\alpha/2)\frac{\sigma}{\sqrt{n}} - \mu}{\sigma/\sqrt{n}}\right)$$

$$= \Pi\left(\frac{\sqrt{n}}{\sigma}(\mu_0 - \mu) + t(\alpha/2)\right) - \Pi\left(\frac{\sqrt{n}}{\sigma}(\mu_0 - \mu) - t(\alpha/2)\right).$$

2.2.4. Nhận xét

- Khi $\alpha\downarrow \Leftrightarrow (1-\alpha)\uparrow \Leftrightarrow \beta(\mu)\uparrow$.

Chỉ khi n đủ lớn $\Leftrightarrow$ α và $\beta(\mu)$ sẽ cùng nhỏ. Vì vậy khuyến cáo lấy $\alpha = 0{,}05$ để n và $\beta(\mu)$ vừa phải.

- μ_0 và μ chênh lệch ít thì $\beta(\mu)$ sẽ lớn ; ngược lại μ_0 và μ chênh lệch nhiều thì $\beta(\mu)$ sẽ nhỏ.

- Khi kiểm định một phía với $H_1 : MX > \mu_0$ thì sai lầm loại II được tính như sau:

$$\beta(\mu) = P\{-\infty < T \le t(\alpha)\} = \Pi\left(\frac{\sqrt{n}}{\sigma}(\mu_0 - \mu) + t(\alpha)\right).$$

- Khi DX không biết, tính gần đúng $\beta(\mu)$ cần thay σ bằng s, $t(\alpha)$ bằng $t(n-1;\alpha)$ và $t(\alpha/2)$ bằng $t(n-1;\alpha/2)$.

Ví dụ

1. Gọi X là chiều cao nam thanh niên (cm). Giả sử X: $N(158{,}5; 5^2)$. Sau 10 năm, đo chiều cao 270 nam thanh niên được kết quả $\overline{x} \pm s_x = 162{,}3 \pm 5{,}5$.

a) Thế hệ sau có thừa nhận MX = 158,5 không? Lấy $\alpha = 0{,}05$.

b) Nếu MX = 160 đúng, hãy tính $\beta(160)$.

Giải

a) Kiểm định giả thiết.

- H_0: MX = 158,5 ; H_1: MX > 158,5.

- Điều kiện

Đại lượng ngẫu nhiên X tuân theo quy luật chuẩn.

- Tính T

$$T = \frac{(162{,}3 - 158{,}5)}{5} \times \sqrt{270} = 12{,}488.$$

- Kết luận

Tra bảng chuẩn tắc t(0,05) = 1,6449.

T > t(0,05): bác bỏ giả thiết H_0, chấp nhận đối thiết H_1.

Thế hệ sau cao hơn, không thừa nhận MX = 158,5.

b) Tính $\beta(\mu)$

$$\beta(160) = \Pi\left(\frac{\sqrt{270}}{5}(158,5-160)+1,6449\right) = \Pi(-3,28) = 1-0,99952 = 0,00048.$$

2. Điều tra 1600 gia đình 4 con thu được kết quả sau:

$$\bar{x} \pm s_x = 2,0469 \pm 1,0333,$$

trong đó X là số con trai gia đình 4 con. Giả sử MX = 2 và DX = 1.

a) Số liệu trên có thừa nhận MX = 2 không ? Lấy $\alpha = 0,05$.

b) Giả sử MX = 2,056 là giá trị đúng, hãy tính $\beta(2,056)$.

Giải

a) Kiểm định giả thiết

- H_0: MX = 2 ; H_1: MX $\neq$ 2.
- Điều kiện: n = 4 quá nhỏ.
- Tính T

$$T = \frac{(2,0469-2)}{1}.\sqrt{1600} = 1,876.$$

- Kết luận

Tra bảng chuẩn tắc t(0,05/2) = 1,96.

T < t(0,05/2) : giữ giả thiết H_0

Số liệu trên thừa nhận MX = 2.

b) Tính $\beta(\mu)$

$$\beta(2,056) = \Pi(\frac{\sqrt{1600}}{1}(2-2,056)+1,96) - \Pi(\frac{\sqrt{1600}}{1}(2-2,056)-1,96)$$

$$= \Pi(-0,28) - \Pi(-4,2) = \Pi(4,2) - \Pi(0,28)$$

$$= 0,999968 - 0,6103 \quad = 0,389668$$

$$\approx 0,39.$$

CÂU HỎI TỰ LƯỢNG GIÁ

(Bài 3 & 4)

Hãy chọn một kết quả đúng.

1. Định lượng Protein toàn phần trong máu hai nhóm trẻ bị bệnh thu được kết quả sau:

Nhóm I	n = 26	$\overline{x}_1 \pm s_x = 47,3 \pm 11,25$ mg%.
Nhóm II	m = 31	$\overline{y}_1 \pm s_y = 53,5 \pm 10,49$ mg%.

Giá trị trung bình của hai nhóm có như nhau không ? Kiểm định 1 phía, lấy $\alpha = 0,05$.

Kết quả:

A. 2,1448 B. 2,1370 C. 2,1435 D. 2,150 E. số khác

2. Điều tra 53680 gia đình 8 con, gọi X là số con trai của gia đình, thu được kết quả sau: $\overline{x} \pm s_x = 4,11742 \pm 1,43786$. Biết DX=2, hãy cho biết MX=4 có đúng không ?

Kết quả:

A. 18,920 B. 19,237 C. 13,602 D. 13,162 E. số khác

3. Gọi X là lượng Protein trong máu trẻ bị bệnh nhẹ (mg%), X : N(60; 40,96).

Định lượng Protein trong máu cho 69 trẻ bị bệnh trên được kết quả $\overline{x} \pm s_x = 61,2 \pm 6,6$, nếu MX = 60 là sai, hãy tính $\beta(62)$. Kiểm định 1 phía, lấy $\alpha = 0,05$.

Kết quả:

A. 0,1711 B. 0,8289 C. 0,999.968 D. 0,2611 E. số khác

4. Xét nghiệm Cholesterol toàn phần (X:mmol/l) cho 2 nhóm trẻ thu được kết quả:

Nhóm I X_1	n=48	$\overline{x}_1 \pm s_x = 3,40 \pm 0,65$
Nhóm II X_2	m=52	$\overline{x}_2 \pm s_x = 3,82 \pm 0,72$

Giả sử X:N(3,50; 0,4624). Lượng Cholesterol toàn phần trung bình chung của hai nhóm có khác biệt với hằng số đã cho không? Kiểm định 1 phía, lấy $\alpha = 0,05$.

Kết quả:

A. 1,6058 B. 1,6176 C. 1,7412 D. 1,7285 E. số khác.

Bài 5

SO SÁNH CÁC TỶ LỆ VÀ KIỂM ĐỊNH TÍNH ĐỘC LẬP

MỤC TIÊU

1. *Giải được bài toán so sánh các tỷ lệ, kiểm định tính độc lập bằng thuật toán χ^2 thường gặp.*

2. *Nêu được ý nghĩa của các bài toán.*

Trong nhiều nghiên cứu thường gặp các câu hỏi như tỷ lệ khỏi của các phương pháp có như nhau không hay tỷ lệ mắc bệnh của các địa phương có như nhau không hoặc tỷ lệ mắc bệnh B có phụ thuộc vào sự nghiện thuốc lá, phụ thuộc vào giới hay phụ thuộc vào nghề nghiệp không ...

Nếu tỷ lệ khỏi của các phương pháp như nhau nghĩa là tỷ lệ khỏi không phụ thuộc vào phương pháp hay tỷ lệ khỏi "độc lập" với các phương pháp. Như vậy từ bài toán so sánh các tỷ lệ cũng có thể dẫn đến bài toán kiểm định tính độc lập giữa các đặc tính.

Giả sử $X_1, X_2, ..., X_k$ là k biến có quy luật đa thức với các tham số n và $p_1, p_2, ..., p_k$ thì

$$Q = \sum_{i=1}^{k}\left(\frac{X_i - np_i}{\sqrt{np_i}}\right)^2$$ là đại lượng ngẫu nhiên có quy luật χ^2 với k − 1 bậc tự do.

Khi đó $Q = \sum_{i=1}^{k}\frac{(m_i - M_i)^2}{M_i}$ là giá trị của đại lượng ngẫu nhiên trên.

Các bài toán nêu trên có thể giải bằng kiểm định χ^2.

1. CÁC BƯỚC

- Giả thiết H_0

Hoặc các tỷ lệ như nhau

Hoặc hai đặc tính độc lập.

- Đối giả thiết H_1

Hoặc các tỷ lệ không như nhau

Hoặc hai đặc tính không độc lập.

- Điều kiện

Tần số xuất hiện của các hiện tượng : m_i hay $m_{ij} \geq 5$ hay lớn hơn 10 càng tốt.

Các đối tượng nghiên cứu phải thuần nhất.

- Tính Q

Các giá trị phân phối thành k hàng : $Q_1 = \sum_{i=1}^{k} \frac{(m_i - M_i)^2}{M_i}$,

trong đó m_i là tần số thực nghiệm, M_i tương ứng là tần số lý thuyết.

Q_1 là giá trị của biến χ^2 với k – 1 bậc tự do.

Các giá trị phân phối thành k hàng, l cột : $Q_2 = \sum_{i,j=1}^{k,l} \frac{(m_{ij} - M_{ij})^2}{M_{ij}}$,

trong đó m_{ij} là tần số thực nghiệm và M_{ij} tương ứng là tần số lý thuyết.

Q_2 là giá trị của biến χ^2 với (k – 1)(l – 1) bậc tự do.

- Kết luận

Tra giá trị tới hạn q(k – 1; 0,05) hoặc q((k – 1)(l – 1); 0,05) trong bảng χ^2 với bậc tự do n = k – 1 hoặc n = (k – 1)(l – 1) và p hay α = 0,05.

Giả sử Q_1 < q (k – 1; 0,05) : Chấp nhận giả thiết H_0.

Ngược lại Q_1 > q (k – 1; 0,05) : Bác bỏ giả thiết H_0, chấp nhận đối thiết H_1.

Tương tự Q_2 < q ((k – 1)(l – 1); 0,05) Chấp nhận giả thiết H_0.

Q_2 > q ((k – 1)(l–1); 0,05) : Bác bỏ giả thiết H_0, chấp nhận đối thiết H_1.

Trên cơ sở kết luận của bài toán kiểm định, cần suy ra ý nghĩa y học.

2. CÁC BÀI TOÁN

2.1. Bài toán 1

Theo dõi 50 trẻ sơ sinh thấy 20 trẻ gái ra đời. Số còn lại là trẻ trai. Tỷ lệ sinh con gái bằng 0,5 có đúng không?

Giải

Nếu tỷ lệ sinh con gái bằng 0,5 thì tỷ lệ sinh con trai cũng bằng 0,5. Ký hiệu T là trai, G là gái và HT: Hiện tượng.

- $H_0 : P(T) = P(G) = 0,5$.

  $H_1 : P(T) \neq P(G)$

- Điều kiện:

  $m_1 = m(G) = 20$, $m_2 = m(T) = 30$. $m_i > 10$, $i = 1,2$

- Lập bảng tính Q

  $M_1 = np_1 = 50 \times 0,5 = 25$

  $M_2 = np_2 = 50 \times 0,5 = 25$

HT	m_i	M_i	$(m_i - M_i)^2/M_i$
G	20	25	1
T	30	25	1
Σ	50	50	2

- Kết luận

Bậc tự do : 2 – 1 = 1 ; q(1;0,05) = 3,841.

Q = 2 < 3,841 : giữ giả thiết H_0. Vậy P(G) = 0,5.

Nhận xét : $\omega(G) = \dfrac{20}{50} = 0,4$, $\omega(T) = \dfrac{30}{50} = 0,6$

Kết luận của bài toán kiểm định là tỷ lệ 0,4 và tỷ lệ 0,6 được "xem là" như nhau. Điều này chưa được thoả đáng. Khi n tăng lên kết luận của bài toán sẽ thay đổi.

2.2. Bài toán 2

Điều trị một bệnh bằng hai phương pháp (FF) thu được kết quả (KQ) sau:

(K : khỏi, Đ : đỡ, TB : thất bại)

FF \ KQ	K	Đ	TB	m_{i0}
I	230	40	50	320
II	76	12	7	95
m_{0j}	306	52	57	415

Hai phương pháp điều trị có hiệu quả như nhau không?

Giải

- H_0 : Hai phương pháp điều trị hiệu quả như nhau.

  H_1 : Hai phương pháp điều trị hiệu quả không như nhau.

- Điều kiện :

  $m_{ij} > 5 \ \forall i,j$.

Đám đông bệnh nhân điều trị bằng hai cách thuần nhất.

- Tính Q

Dựa vào H_0 dẫn đến $M_{ij} = \frac{m_{i0} \times m_{0j}}{n}$

$M_{11} = 235{,}952$ $\quad M_{12} = 40{,}096$ $\quad M_{13} = 43{,}952$

$M_{21} = 70{,}048$ $\quad M_{22} = 11{,}904$ $\quad M_{23} = 13{,}048$

$$Q = \sum_{i,j=1}^{2,3} \frac{(m_{ij} - M_{ij})^2}{M_{ij}} = 4{,}292.487.$$

- Kết luận

Bậc tự do là (2 –1)(3 – 1) = 2. q(2; 0,05) = 5,991.

Q < p(2 ; 0,05) : giữ giả thiết H_0.

Hai phương pháp điều trị hiệu quả như nhau.

Tính theo công thức tính nhanh, Q = 4,292.645. Sự khác biệt giữa hai kết quả tính là không đáng kể.

2.3. Bài toán 3

Điều tra tình hình mắc ba bệnh (B) B_1, B_2, B_3 tại hai phân xưởng (FX) I và II của nhà máy X thu được kết quả sau:

FX \ B	B_1	B_2	B_3	m_{i0}
I	588	369	89	1046
II	304	171	50	525
m_{0j}	892	540	139	1571

Tỷ lệ ba bệnh tại hai phân xưởng có như nhau không?

Giải

- H_0 : Tỷ lệ ba bệnh tại hai phân xưởng như nhau.

  H_1 : Tỷ lệ ba bệnh tại hai phân xưởng khác nhau.

- Điều kiện :

  $m_{ij} > 10 \ \forall i,j$.

Điều tra tất cả công nhân của hai phân xưởng

- Tính Q

Dựa vào H_0 suy ra $M_{ij} = \frac{m_{i0} \times m_{0j}}{n}$

$M_{11} = 593.91$ $\quad M_{12} = 359{,}54$ $\quad M_{13} = 92{,}55$

$M_{21} = 298{,}09$ $\qquad$ $M_{22} = 180{,}46$ $\qquad$ $M_{23} = 46{,}45$.

$$Q = \sum_{i,j=1}^{2,3} \frac{(m_{ij} - M_{ij})^2}{M_{ij}} = 1{,}328$$

- Kết luận

Bậc tự do bằng (2 –1)(3 – 1) = 2 ; q(2; 0,05) = 5,991.

$Q < q(2\,;\,0{,}05)$: giữ giả thiết H_0.

Tỷ lệ các bệnh B_1, B_2, B_3 ở hai phân xưởng như nhau.

Để trả lời câu hỏi tỷ lệ mắc các bệnh tại hai phân xưởng có như nhau không, bài toán chỉ được giải trong trường hợp đặc biệt.

2.4. Bài toán 4

Theo dõi phương pháp điều trị ngoại khoa cải tiến trong 10 năm, thu được bảng số liệu sau:

KQ \ Năm	1 2 3	4 5 6 7	8 9 10	Σ
T	31	71	58	160
$\bar{T}$	16	25	6	47
Σ	47	96	64	207

Tỷ lệ tốt của 3 nhóm có như nhau không?

Giải

- H_0 : Tỷ lệ tốt của ba nhóm như nhau.

  H_1 : Tỷ lệ tốt của ba nhóm khác nhau.

- Điều kiện:

$m_{ij} > 5 \ \forall\, i,j$. Các nhóm bệnh nhân thuần nhất.

- Tính Q

$$Q = \sum_{i,j=1}^{2,3} \frac{(m_{ij} - M_{ij})^2}{M_{ij}} = 10{,}531$$

- Kết luận: Bậc tự do bằng 2. q (2; 0,05) = 5,991.

  $Q > q(2;\,0{,}05)$: bác H_0, chấp nhận H_1.

Tỷ lệ tốt của 3 nhóm tương ứng với 3 thời kỳ không như nhau.

Chú ý: Từ năm 11 trở đi phương pháp cải tiến có tỷ lệ tốt trên 90,625% (58/64).

2.5. Bài toán 5

Chẩn đoán bệnh cho 1000 người tại cộng đồng, tỷ lệ mắc bệnh bằng 2,5%. Dùng một phản ứng chẩn đoán, phản ứng có độ nhạy bằng 0,8 và xác suất dương tính của nhóm không bệnh bằng 0,08. Độ nhạy và độ đặc hiệu của phản ứng có như nhau không?

Giải

- $H_0 : P(A/B) = P(\bar{A}/\bar{B})$ hay $P(Đ/B) = P(Đ/\bar{B})$

  $H_1 : P(A/B) \neq P(\bar{A}/\bar{B})$.

ε / β	Đ	S	Σ
B	20	5	25
$\bar{B}$	897	78	975
Σ	917	83	1000

- Điều kiện:

  $m_{ij} \geq 5 \; \forall$ i,j.

- Tính Q

  Q = 4,612.

- Kết luận: Bậc tự do bằng 1.

q (1; 0,05) = 3,841. Q > q(1; 0,05).

Độ đặc hiệu cao hơn hẳn độ nhạy. Nếu phản ứng không đắt, nên dùng để chẩn đoán phân biệt ở cộng đồng.

2.6. Bài toán 6

Điều tra một đám đông người nước ngoài với hai đặc tính màu tóc (MT) và màu mắt (MM) thu được số liệu sau:

MT / MM	Đen	Hung	Nâu	Bạch kim	m_{i0}
Xanh	35	19	36	25	115
Đen	14	14	16	10	54
m_{0j}	49	33	52	35	169

Hai đặc tính di truyền có độc lập với nhau không?

Giải

- H_0 : Hai đặc tính di truyền độc lập nhau.

  H_1 : Hai đặc tính di truyền không độc lập nhau.

- Điều kiện: $m_{ij} \geq 10 \quad \forall\, i, j$.
- Tính Q

Từ giả thiết H_0 suy ra $M_{ij} = \dfrac{m_{i0} \times m_{0j}}{n}$

$M_{11} = 33{,}343$ $\quad M_{12} = 22{,}456$ $\quad M_{13} = 35{,}385$ $\quad M_{14} = 23{,}817$

$M_{21} = 15{,}657$ $\quad M_{22} = 10{,}544$ $\quad M_{23} = 16{,}615$ $\quad M_{24} = 11{,}183$.

$$Q = \sum_{i,j=1}^{2,4} \frac{(m_{ij} - M_{ij})^2}{M_{ij}} = 2{,}140.$$

- Kết luận

Bậc tự do bằng $(2 - 1)(4 - 1) = 3$. $\quad q\,(3; 0{,}05) = 7{,}815$

$Q < q\,(3; 0{,}05)$: giữ giả thiết H_0.

Hai đặc tính di truyền độc lập nhau.

3. CÔNG THỨC TÍNH NHANH

3.1. k, l lớn hơn hoặc bằng 2

$$Q = \sum_{i,j=1}^{k,l} \frac{(m_{ij} - M_{ij})^2}{M_{ij}} = n\left[\sum_{i,j=1}^{k,l} \frac{m_{ij}^2}{m_{i0} \times m_{0j}} - 1\right]$$

3.2. k = 2, l ≥ 2

$$Q = \sum_{i,j=1}^{2,l} \frac{(m_{ij} - M_{ij})^2}{M_{ij}} = \sum_{j=1}^{l} \frac{(m_{1j} \times m_{20} - m_{2j} \times m_{10})^2}{m_{0j} \times m_{10} \times m_{20}}$$

Áp dụng với số liệu bài toán 6

$$Q = \frac{(35 \times 54 - 14 \times 115)^2}{49 \times 115 \times 54} + \ldots + \frac{(25 \times 54 - 115 \times 10)^2}{35 \times 115 \times 54} = 2{,}139 \approx 2{,}140.$$

3.3. k = l = 2

$$Q = \sum_{i,j=1}^{2,2} \frac{(m_{ij} - M_{ij})^2}{M_{ij}} = \frac{(m_{11} \times m_{22} - m_{12} \times m_{21})^2.n}{m_{10} \times m_{20} \times m_{01} \times m_{02}}$$

Ví dụ

Quan sát các cây với hai đặc tính màu hoa (H) và dạng lá (L) thu được số liệu sau:

H \ L	Đỏ	Hồng	m_{io}
Phẳng	328	122	450
Nhăn	77	33	110
m_{0j}	405	155	560

Hai đặc tính di truyền có độc lập không?

Giải

- H_0 : Hai đặc tính di truyền độc lập nhau.

  H_1 : Hai đặc tính di truyền không độc lập.

- Điều kiện : $m_{ij} > 10\ \forall i,j$.

Không có loại cây nào trong nhóm nghiên cứu chết non.

- Tính Q

$$Q = \frac{(328\times 33 - 122\times 77)^2 \times 560}{450\times 110\times 405\times 155} = 0,3685$$

Chú ý : $Q = \sum_{i,j=1}^{2.2} \frac{(m_{ij} - M_{ij})^2}{M_{ij}} = 0,3685.$

- Kết luận

Bậc tự do bằng (2 –1)(2 – 1) = 1. $q(1; 0,05) = 3,841.$

$Q < q(1\ ;\ 0,05)$: Hai đặc tính di truyền độc lập với nhau.

CÂU HỎI TỰ LƯỢNG GIÁ

Hãy chọn một kết quả đúng.

1. Năm 1998 có 8110 người ngộ độc cấp, trong đó tỷ lệ tử vong bằng 3,76%.

Năm 1999 có 8339 người ngộ độc cấp, trong đó tỷ lệ tử vong bằng 3,02%.

Tỷ lệ tử vong của bệnh nhân ngộ độc cấp trong 2 năm có như nhau không ?

Kết quả:

A. 6,8603 B. 3,8049 C. 5,0431 D. 6,9413 E. số khác

2. Tỷ lệ bị bệnh tại 1 bệnh viện bằng 0,24. Chẩn đoán bệnh cho 1000 người bởi phương pháp I, phương pháp I có độ nhạy bằng 0,875. Chẩn đoán bệnh cho 1000 người bởi phương pháp II, phương pháp II có độ nhạy bằng 0,775.

Độ nhạy của 2 phương pháp có như nhau không ?

Kết quả:

A. 1,4545 B. 34,6320 C. 6,0606 D. 8,3117 E. số khác

3. Một phản ứng có xác suất dương tính bằng 0,89, xác suất sai bằng 0,1 và xác suất bị bệnh của nhóm đúng bằng 0,9. Dùng phản ứng chẩn đoán cho 400 người. Giá trị dương tính của phản ứng có bằng giá trị âm tính của phản ứng không ?

Kết quả:

A. 125,0177 B. 3,6772 C. 147,2072 D. 5,9800 E. số khác

4. Tỷ lệ chết chu sinh của 3 địa phương A, B, C tương ứng bằng 18/1000 20/1000 10/1000. Xác suất chết chu sinh của 3 địa phương có như nhau không ?

Kết quả:

A. 3,500 B. 3,5569 C. 3,4537 D. 0,0569 E. số khác.

Bài 6

KIỂM ĐỊNH QUY LUẬT XÁC SUẤT CỦA ĐẠI LƯỢNG NGẪU NHIÊN

MỤC TIÊU

1. Giải được bài toán kiểm định quy luật chuẩn.

2. Giải được bài toán kiểm định quy luật nhị thức.

Trong bài này trình bày cách tiến hành kiểm định χ^2 xem đại lượng ngẫu nhiên X có quy luật chuẩn, có quy luật nhị thức hay có quy luật xác suất nào đấy không. Ngoài kiểm định quy luật chuẩn theo kiểm định χ^2 cũng có thể kiểm định quy luật chuẩn theo Kolmogorov hoặc các kiểm định gần đúng khác.

1. KIỂM ĐỊNH QUY LUẬT NHỊ THỨC CỦA ĐẠI LƯỢNG NGẪU NHIÊN X

1.1. Các bước

- H_0 : X tuân theo quy luật nhị thức với tham số n và p.

  p có thể cho trước hoặc ước lượng bởi ω.

  H_1 : X không tuân theo quy luật nhị thức với tham số n và p.

- Điều kiện :

$m_i \geq 5; \quad i = \overline{0, n}$. Nếu lớp i có tần số m_i dưới 5 thì ghép lớp này với lớp kề bên sao cho tần số các lớp không dưới 5 và sàn sàn nhau.

- Lập bảng tính Q

Trước hết tính p_i. $p_i = P\{X = i\} = C_n^i p^i (1-p)^{n-i} \qquad i = \overline{0, n}$

Tiếp theo tính M_i. $M_i = np_i \qquad i = \overline{0, n}$

Sau đó tính Q: $$Q = \sum_{i=1}^{K} \frac{(m_i - M_i)^2}{M_i}.$$

- Bậc tự do

Nếu p cho trước, bậc tự do bằng (k – 1)

Nếu ước lượng p bằng tần suất tính từ số liệu, bậc tự do bằng (k–2).

Tra bảng χ^2 với bậc tự do đã tính, lấy $\alpha = 0,05$.

- Kết luận

$Q \leq q(...; 0,05)$: giữ giả thiết H_0.

X tuân theo quy luật nhị thức với tham số n và p.

$Q > q(...; 0,05)$: bác giả thiết H_0, chấp nhận đối thiết H_1.

X không tuân theo quy luật nhị thức với tham số n và p.

Ví dụ

Gọi X là số con trai trong gia đình 4 con. Điều tra 1600 gia đình 4 con thu được số liệu sau:

x_i(số con trai)	0	1	2	3	4	Σ
m_i(số gia đình)	111	367	576	428	118	1600

Đại lượng ngẫu nhiên X có tuân theo quy luật nhị thức với n = 4 và p = 0,5 không?

Giải

- H_0 : X nhị thức với n = 4 và p = 0,5.

  H_1 : X không nhị thức với n = 4 và p = 0,5.

- Điều kiện

  $m_i > 100;\ i = \overline{0,4}$.

Điều tra hết các gia đình 4 con.

- Lập bảng tính Q

  $p_i = P(X = i) = C_4^i \times 0,5^i\, 0,5^{4-i} = C_4^i \,.\, 0,0625$.

  $M_i = np_i = 1600 \times p_i$

i	x_i	m_i	p_i	M_i	$(m_i-M_i)^2/M_i$
0	0	111	0,0625	100	1,210
1	1	367	0,250	400	2,7225
2	2	576	0,375	600	0,96
3	3	428	0,250	400	1,96
4	4	118	0,0625	100	3,24
Σ		**1600**	**1,**	**1600**	**10,0925**

$$Q = \sum_{i=0}^{4} \frac{(m_i - M_i)^2}{M_i} = 10,0925 \approx 10,093$$

- Bậc tự do: p = 0,5 cho trước nên bậc tự do bằng (5 – 1) = 4.

  q(4; 0,05) = 9,488.

- Kết luận

Q = 10,093 > 9,488. Bác giả thiết H_0, chấp nhận đối thiết H_1.

X không tuân theo quy luật nhị thức với n = 4 và p = 0,5.

Nếu lấy p là xác suất sinh con trai từ số liệu, đại lượng ngẫu nhiên X sẽ có quy luật nhị thức với n = 4 và $p \approx \omega(T)$.

Nhận xét : Kiểm định đại lượng ngẫu nhiên X có quy luật poisson hoặc quy luật siêu bội tiến hành tương tự trên.

Kiểm định sẽ càng đúng khi n càng lớn.

2. KIỂM ĐỊNH QUY LUẬT CHUẨN CỦA ĐẠI LƯỢNG NGẪU NHIÊN X

2.1. Các bước

- H_0 : X : $N(\mu, \sigma^2)$

μ và σ^2 có thể được cho trước hoặc ước lượng bởi tham số mẫu.

H_1 : X không chuẩn với tham μ và σ^2.

- Điều kiện :

Phải chia miền giá trị của đại lượng ngẫu nhiên X thành k lớp sao cho $k \geq 4$, tần số thực nghiệm của lớp thứ i từ α_{i-1} đến α_i phải từ 5 trở lên và tần số các lớp sàn sàn nhau.

- Lập bảng tính Q

Ứng với lớp thứ i từ α_{i-1} đến α_i, tính xác suất theo giả thiết H_0.

$$P\{\alpha_{i-1} < X \leq \alpha_i\} = \Pi\left(\frac{\alpha_i - \mu}{\sigma}\right) - \Pi\left(\frac{\alpha_{i-1} - \mu}{\sigma}\right) = p_i \qquad i = \overline{1,k}.$$

Tính tần số lý thuyết tương ứng với xác suất p_i :

$$M_i = np_i \quad i = \overline{1,k}.$$

Sau đó tính Q: $Q = \sum_{i=1}^{K} \frac{(m_i - M_i)^2}{M_i}$.

- Bậc tự do

Nếu μ và σ^2 cho trước, bậc tự do bằng (k – 1)

Nếu ước lượng μ và σ^2 bằng tham số mẫu tương ứng là $\overline{x}$ và s^2, bậc tự do bằng (k–1–2).

Tra bảng χ^2 với bậc tự do đã tính, lấy $\alpha = 0,05$.

- Kết luận

$Q \leq q(...; 0,05)$: giữ giả thiết H_0.

Đại lượng ngẫu nhiên X có quy luật chuẩn với tham số μ và σ^2.

$Q > q(...; 0,05)$: bác giả thiết H_0; chấp nhận H_1.

Ví dụ

Gọi X là áp lực động mạch phổi thì tâm trương người bình thường. Đo 30 người, thu được số liệu sau:

x_i (mm Hg)	2	3	4	5	6	7	8	9	Σ
m_i (số người)	1	4	7	8	2	5	2	1	30

Đại lượng ngẫu nhiên X có quy luật chuẩn với tham số $\mu \approx \overline{x}$ và $\sigma^2 \approx s^2$ không?

Giải

a) Tính tham số mẫu.

Kết quả trong bài tham số mẫu là $\overline{x} = 5,1$; $s^2 = 1,7^2$.

b) Kiểm định quy luật chuẩn.

- $H_0 : X : N(\mu, \sigma^2)$ với $\mu \approx \overline{x}$ và $\sigma^2 \approx s^2$.

  H_1: X không chuẩn với μ và σ^2.

- Điều kiện

Chia miền giá trị X thành 4 lớp sao cho tần số lớp không nhỏ hơn 5 và sàn sàn nhau.

- Lập bảng tính Q.

$$t_i = \frac{\alpha_i - \mu}{\sigma} \qquad p_i = \Pi(t_i) - \Pi(t_{i-1}) \qquad i = \overline{1,4}$$

i	$[\alpha_{i-1} \div \alpha_i)$	m_i	t_i	$\Pi(t_i)$	p_i	M_i	$(m_i - M_i)^2/M_i$
1	$-\infty \div 3,5$	5	– 0,94	0,1736	0,1736	5,2	0,0076
2	$3,5 \div 4,5$	7	– 0,35	0,3632	0,1896	5,7	0,2964
3	$4,5 \div 6,5$	10	0,82	0,7939	0,4307	12,9	0,6488
4	$6,5 \div \infty$	8	∞	1	0,2061	6,2	0,5225
Σ		**30**			**1**	**30,**	**1,4753**

- Bậc tự do

Bậc tự do bằng (4 –1 – 2) = 1 ; q(1; 0,05) = 3,841.

- Kết luận

Q = 1,475 < 3,841 : giữ giả thiết H_0 : X : N(5,1; $1,7^2$).

Nhận xét :

Chia miền giá trị thành 4 lớp với tần số mỗi lớp không nhỏ hơn 5 vì vậy n ≥ 30.

Mỗi cách chia lớp miền giá trị sẽ dẫn đến kết quả Q tính được khác nhau, như vậy có thể làm thay đổi kết luận kiểm định.

Có thể tiến hành ngược lại với cách làm trên nghĩa là chia lớp các giá trị của đại lượng ngẫu nhiên sao cho tần số lý thuyết M_i của các lớp bằng nhau và không nhỏ hơn 5 trên cơ sở đó có xác xuất của các lớp bằng nhau.

Gọi số lớp là k : $k \le \frac{n}{M_i} = \frac{n}{5}$ và $p_i = \frac{1}{k}$ $\quad i = \overline{1,k}$.

Xét lại ví dụ trên:

$$k \le \frac{30}{5} = 6 \qquad p_i = \frac{1}{6} \qquad i = \overline{1,6}$$

- H_0 : X : N(5,1 ; $1,7^2$)

  H_1: X không chuẩn với 5,1 và $1,7^2$

- Điều kiện : $M_i \ge 5$ $\quad i = \overline{1,6}$

- Tính Q

Có bảng tính sau, với $x_i = 5,1 + 1,7.t_i$

i	p_i	$\Pi(t_i)$	t_i	x_i	m_i	M_i	$(m_i - M_i)^2/M_i$
1	1/6	0,1667	– 0,97	3,451	5	5	0
2	1/6	0,3333	– 0,43	4,369	7	5	0,8
3	1/6	0,5	0	5,1	8	5	1,8
4	1/6	0,6667	0,43	5,831	0	5	5,
5	1/6	0,8333	0,97	6,749	2	5	1,8
6	1/6	1	∞	∞	8	5	1,8
Σ	**1**				**30**	**30**	**11,2**

- Bậc tự do bằng (6 – 1 – 2) = 3 q(3 ; 0,05) = 7,815.
- Kết luận Q > q(3 ; 0,05) : bác giả thiết X : N (5,1 ; $1,7^2$).

Ngoài kiểm định tính chuẩn của đại lượng ngẫu nhiên theo kiểm định χ^2 có thể kiểm định theo phương pháp khác.

2.3. Kiểm định Kolmogorov

Giả sử đại lượng ngẫu nhiên X nhận các giá trị $x_1, x_2, \ldots, x_n$.

Xét giả thiết H_0 X : $N(\mu, \sigma^2)$ với μ và σ^2 cho trước.

Gọi $F_n(x)$ là hàm phân phối thực nghiệm. Khi đó:

$$F_n(x) = \frac{k}{n}$$

trong đó k là số giá trị x_i thực sự nhỏ hơn x. Như vậy $F_n(x)$ chính là tần suất cộng dồn của những giá trị nhỏ hơn x.

F(x) là hàm phân phối chuẩn của đại lượng ngẫu nhiên X.

Đặt $D_n = \max_x |F_n(x) - F(x)|$.

Theo Kolmogorov

$$P\left\{\sqrt{n}\, D_n > Z\right\} = P\left\{D_n > \frac{Z}{\sqrt{n}}\right\} \approx 1 - \theta(Z).$$

Với các giá trị Z = 1,3581 và Z = 1,6276 thì θ(1,3581) = 0,95 và θ(1,6276) = 0,99 do đó:

$$P\left\{D_n > \frac{1,3581}{\sqrt{n}}\right\} \approx 0,05 \text{ và } P\left\{D_n > \frac{1,6276}{\sqrt{n}}\right\} \approx 0,01.$$

Bài toán kiểm định quy luật chuẩn của đại lượng ngẫu nhiên X với tham số μ và σ^2 sẽ được chấp nhận khi $D_n \le \frac{1,3581}{\sqrt{n}}$ và bác bỏ khi $D_n > \frac{1,3581}{\sqrt{n}}$ với độ tin cậy 95%.

Bài toán kiểm định sẽ rất tốt nếu $n \ge 30$ và có thể chấp nhận được khi n = 10.

Khi giả thiết H_0 X : $N(\mu, \sigma^2)$ với $\mu \approx \overline{x}$ và $\sigma^2 \approx s_x^2$ thì giá trị Z được thay bởi 0,878 và 0,989 tương ứng với các xác suất 0,95 và 0,99.

Khi đó $D_n \le \frac{0,878}{\sqrt{n}}$ chấp nhận giả thiết H_0

$D_n > \frac{0,878}{\sqrt{n}}$ bác bỏ giả thiết H_0 với độ tin cậy 95%.

Xét một ví dụ để minh hoạ phương pháp này.

Gọi x là nhịp tim nam bình thường 9 tuổi. Đếm nhịp tim 30 trẻ thu được số liệu sau:

X_i (l/ph)	64	66	68	70	72	74	75	76	77	78	80	Σ
m_i	1	2	6	1	6	2	1	4	2	2	3	30

Tính tham số mẫu có kết quả $\overline{x} \pm s_x = 72,77 \pm 4,60$.

X có chuẩn với $\mu \approx \overline{x}$ và $\sigma^2 \approx s^2$ không?

Giải

- H_0 : X : N(72,8 ; $4,6^2$), H_1: X không chuẩn với các tham số 72,8 và $4,6^2$.
- Điều kiện $n \geq 30$.
- Lập bảng tính D_n

X_i	64	66	68	70	72	74	75	76	77	78	80	∞
K	0	1	3	9	10	16	18	19	23	25	27	30
$F_N(X)$	0	0,033	0,1	0,3	0,3333	0,5333	0,6	0,6333	0,7667	0,8333	0,9	1
F(X)	0,0281	0,0694	0,1492	0,271	0,4325	0,6026	0,6879	0,7580	0,8186	0,8810	0,9418	1
D_N								0,1247				

- Tính giá trị tới hạn $\dfrac{0,878}{\sqrt{30}} = 0,1603$.
- Kết luận $D_n = 0,1247 < 0,1603$: chấp nhận H_0.

Đại lượng ngẫu nhiên X chuẩn với $\mu \approx 72,8$ và $\sigma^2 \approx 4,6^2$.

Nhận xét:

Khi kiểm định giả thiết đại lượng ngẫu nhiên X chuẩn với μ và σ^2 cho trước thì kiểm định Kolmogorov mạnh hơn kiểm định χ^2 . Nếu $\mu \approx \overline{x}$ và $\sigma^2 \approx s^2$ thì kiểm định χ^2 mạnh hơn.

Nếu số liệu đủ điều kiện cho phép kiểm định tính chuẩn thì nên kiểm định trước, sau đó mới thực hiện các bài toán khác.

Bài 7

KIỂM ĐỊNH GIÁ TRỊ CỦA XÁC SUẤT

MỤC TIÊU

1. *Giải được bài toán kiểm định xác suất.*
2. *Trình bày được định nghĩa và tính được sai lầm loại II.*

Xác suất khỏi khi điều trị một bệnh bằng bao nhiêu? Tỷ lệ bị bệnh tại một vùng đang có dịch là bao nhiêu? Độ nhạy của một phương pháp chẩn đoán bằng bao nhiêu?

Có thể tiến hành phương pháp kiểm định giá trị của xác suất để trả lời các câu hỏi trên.

1. ƯỚC LƯỢNG XÁC SUẤT

- Thực hiện phép thử ε n lần độc lập, hiện tượng A xuất hiện m lần.

$$\omega(A) = \frac{m}{n}.$$

- Khi không biết P(A), có ước lượng điểm sau :

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

- Ước lượng khoảng của P(A).

Chọn $\alpha = 0{,}05$, có khoảng ước lượng 95%.

$$\omega - t(\alpha/2)\sqrt{\frac{\omega(1-\omega)}{n}} \le P(A) \le \omega + t(\alpha/2)\sqrt{\frac{\omega(1-\omega)}{n}}.$$

Như vậy có 5% trường hợp P(A) có thể nằm ngoài khoảng trên.

2. KIỂM ĐỊNH HAI PHÍA

Thực hiện phép thử n lần độc lập, hiện tượng A xuất hiện m lần, P(A) có bằng P_0 không?

Biến ngẫu nhiên S là số lần xuất hiện A có quy luật nhị thức với tham số n và P_0.

Khi n đủ lớn có thể thay quy luật nhị thức bằng quy luật chuẩn $N(\mu, \sigma^2)$ với $\mu = MS = np_0$ và $\sigma^2 = DS = np_0(1-p_0)$.

2.1. Các bước

- H_0 : $P(A) = p_0$

  H_1 : $P(A) \neq p_0$.

- Điều kiện

n đủ lớn khi $np_0 \geq 10$ và $n(1 - p_0) \geq 10$.

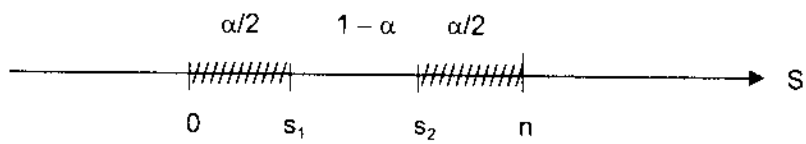
- Tính giá trị $s_1(p_0)$ và $s_2(p_0)$

Dựa vào ước lượng khoảng của xác suất suy ra :

$$S_1(p_0) = np_0 - t(\alpha/2)\sqrt{np_0(1-p_0)}$$

$$S_2(p_0) = np_0 + t(\alpha/2)\sqrt{np_0(1-p_0)}$$

Biểu diễn các giá trị trên một trục số :



- Kết luận

$S = m \in [s_1, s_2]$: giữ giả thiết H_0. Chấp nhận $P(A) = p_0$.

$S = m \notin [s_1, s_2]$: Bác giả thiết H_0, chấp nhận đối giả thiết H_1. Như vậy $P(A) \neq p_0$.

2.2. Các xác suất

- Trường hợp H_0 đúng.

  $P(A) = p_0$, $S = m \in [s_1, s_2]$: Kết luận giữ giả thiết H_0 khi H_0 đúng.

Xác suất giữ giả thiết khi giả thiết đúng gọi là độ tin cậy của kiểm định.

$S = m \notin [s_1, s_2]$: Kết luận bác giả thiết H_0 khi H_0 đúng.

Xác suất bác giả thiết khi giả thiết đúng gọi là nguy hiểm loại I hay sai lầm loại I.

Do H_0 đúng nên sai lầm loại I chính là α còn độ tin cậy là $1 - \alpha$.

- Trường hợp H_0 sai

  $P(A) \neq p_0$, $P(A) = p$ với p là giá trị đúng.

  $S = m \in [s_1, s_2]$: kết luận giữ giả thiết H_0 khi H_0 sai.

Xác xuất giữ giả thiết khi giả thiết sai gọi là nguy hiểm loại II hay sai lầm loại II. Hàm sai lầm loại II phụ thuộc p đúng, ký hiệu $\beta(p)$.

$S = m \overline{\in} [s_1, s_2]$: kết luận bác giả thiết H_0 khi H_0 sai.

Xác suất bác giả thiết H_0 khi H_0 sai gọi là lực của kiểm định.

- Tính sai lầm loại II.

$$\beta(p) = P\{s_1(p_0) \leq S \leq s_2(p_0)\} = \sum_{r=s_1}^{s_2} C_n^r p^r (1-p)^{n-r}$$

$$\approx \Pi\left(\frac{s_2 - np}{\sqrt{npq}}\right) - \Pi\left(\frac{s_1 - np}{\sqrt{npq}}\right)$$

$$= \Pi\left(\frac{\sqrt{n}(p_0 - p) + t(\alpha/2)\sqrt{p_0(1-p_0)}}{\sqrt{pq}}\right) - \Pi\left(\frac{\sqrt{n}(p_0 - p) - t(\alpha/2)\sqrt{p_0(1-p_0)}}{\sqrt{pq}}\right)$$

- Khi $\alpha \searrow \;\Leftrightarrow (1-\alpha) \nearrow \;\Leftrightarrow \beta(p) \nearrow$

Khi $n \nearrow \;\Leftrightarrow \alpha \searrow$ và $\beta(p) \searrow$

Khuyến cáo lấy $\alpha = 0{,}05$ để n và $\beta(p)$ không quá lớn.

Ví dụ

Theo dõi 900 trẻ sơ sinh thấy 427 gái ra đời. P(G) = 0,5 có đúng không? Nếu P(G) = 0,486 là giá trị đúng, hãy tính $\beta(0{,}486)$ lấy $\alpha = 0{,}05$.

Giải

- $H_0 : P(G) = 0{,}5$, $\quad H_1 : P(G) \neq 0{,}5$.
- Điều kiện:

  $np_0 = 900 \times 0{,}5 = 450$, $n(1 - p_0) = 900 \times 0{,}5 = 450$.
- Tính s_1 và s_2.

$$S_1(p_0) = 900 \times 0{,}5 - 1{,}96\sqrt{900 \times 0{,}5 \times 0{,}5}$$
$$= 450 - 29{,}4 = 420{,}6$$
$$= 421.$$

$$S_2(p_0) = 450 + 29{,}4 = 479{,}4$$
$$\approx 479.$$

- Kết luận:

  $S = 427 \in [421; 479]$: giữ giả thiết H_0.

Số liệu trên thừa nhận P(G) = 0,5.

- $\beta(0{,}486) = P\{420{,}6 \le S \le 479{,}4\}$

$$= \Pi\left(\frac{479{,}4 - 900\times 0{,}486}{\sqrt{900\times 0{,}486\times 0{,}514}}\right) - \Pi\left(\frac{420{,}6 - 900\times 0{,}486}{\sqrt{900\times 0{,}486\times 0{,}514}}\right)$$

$$= \Pi(2{,}8) - \Pi(-1{,}12) = 0{,}9974 - 1 + 0{,}8686$$

$$= 0{,}866.$$

Hàng trăm lần tiến hành kiểm định P(G) = 0,5, khi giữ giả thiết H_0 mà H_0 sai thì bị sai khoảng 87 lần. Sai lầm quá lớn.

3. KIỂM ĐỊNH MỘT PHÍA

3.1. Các bước

- $H_0 : P(A) = p_0$

  H_1 : Hoặc $P(A) > p_0$ (Trường hợp 1)

  Hoặc $P(A) < p_0$ (Trường hợp 2)

- Điều kiện:

  $np_0 \ge 10$, $n(1 - p_0) \ge 10$.

- Tính giá trị tới hạn.

Khi $P(A) > p_0$ tính $s'_2(p_0)$.

$$s'_2(p_0) = np_0 + t(\alpha)\sqrt{np_0(1-p_0)}$$

- Kết luận:

  $S = m \in [0,\ s'_2]$: Giữ giả thiết H_0.

  $S = m \notin [0,\ s'_2]$: Bác giả thiết H_0. Chấp nhận đối thiết H_1

Như vậy $P(A) > p_0$.

Khi $P(A) < p_0$ tính $s'_1(p_0)$.

$$s'_1(p_0) = np_0 - t(\alpha)\sqrt{np_0(1-p_0)}$$

- Kết luận:

  $S = m \in [\ s_1'\ ;\ n]$: giữ giả thiết H_0.

  $S = m \notin [s_1'\ ;\ n]$: bác giả thiết H_0, chấp nhận đối thiết H_1

Như vậy $P(A) < p_0$.

3.2. Tính sai lầm loại II

Tương tự như kiểm định hai phía, xác suất giữ giả thiết H_0 khi H_0 sai là sai lầm loại II, ký hiệu là $\beta(p)$.

Trường hợp (1) : $P(A) > p_0$

$$\beta(p) = \Pi\left(\frac{s_2'(p_0) - np}{\sqrt{npq}}\right) = \Pi\left(\frac{\sqrt{n}(p_0 - p) + t(\alpha)\sqrt{p_0(1-p_0)}}{\sqrt{pq}}\right).$$

Trường hợp (2) : $P(A) < p_0$

$$\beta(p) = 1 - \Pi\left(\frac{s_1'(p_0) - np}{\sqrt{npq}}\right) = 1 - \Pi\left(\frac{\sqrt{n}(p_0 - p) - t(\alpha)\sqrt{p_0(1-p_0)}}{\sqrt{pq}}\right),$$

trong đó $\Pi\left(\frac{s_1'(p_0) - np}{\sqrt{npq}}\right)$ chính là lực của kiểm định.

Ví dụ

1. Xác định nhóm máu cho 2000 người thấy 1002 người có nhóm máu 0. Tỷ lệ người có nhóm máu 0 bằng 0,48 có đúng không? Lấy $\alpha = 0,05$.

Giải

- $H_0 : P(0) = 0,48$

  $H_1 : P(0) > 0,48$.
- Điều kiện:

  $np_0 = 2000 \times 0,48 = 960$; $n(1 - p_0) = 2000 \times 0,52 = 1040$.
- Tính $s_2'(p_0)$

  $$s_2'(p_0) = 2000 \times 0,48 + 1,6449\sqrt{2000 \times 0,48 \times 0,52} = 996,75.$$
  $$\simeq 997.$$
- Kết luận:

  $S = 1002 \notin [0; 997]$: bác giả thiết H_0. Chấp nhận đối thiết H_1.

Với số liệu trên, $P(0) > 0,48$.

2. Khi sử dụng một loại thuốc, do tai biến, có ý kiến cho là nên cấm, có ý kiến cho là nên tiếp tục sử dụng.

Phải hỏi bao nhiêu người? Bao nhiêu người đồng ý cấm thì ra lệnh cấm?

Nếu cho rằng : $P(C) = 0,5$; lấy $\alpha = 0,05$ sao cho $\beta(0,6) \le 0,2$, trong đó C ký hiệu là cấm dùng.

Giải

- $H_0 : P(C) = 0,5$

  $H_1 : P(C) > 0,5.$

- $$\beta(0,6) = \Pi\left(\frac{\sqrt{n}(p_0 - p) + t(\alpha)\sqrt{p_0(1-p_0)}}{\sqrt{pq}}\right) \le 0,2 = \Pi(-0,84)$$

$\Pi(x)$ là hàm đồng biến dẫn đến :

$$\frac{\sqrt{n}(0,5-0,6) + 1,6449\sqrt{0,5\times 0,5}}{\sqrt{0,6\times 0,4}} \le -0,84\,.$$

$$n = (\sqrt{n})^2 \ge \left(\frac{0,84\times\sqrt{0,24} + 1,6449\times 0,5}{0,1}\right)^2$$

$$\ge 152,26 \approx 153.$$

$$s_2' = 153\times 0,5 + 1,6449\sqrt{153\times 0,5\times 0,5} = 86,67$$
$$\approx 87.$$

Để thoả mãn các điều kiện của bài toán, cần hỏi 153 người, số người đồng ý cấm trên 87 người thì ra lệnh cấm.

3. Hỏi n người dùng thuốc A hoặc B chữa một bệnh có m người cho là A tốt hơn B, số còn lại cho là B tốt hơn A. Lấy $\alpha = 0,05$, có thật sự A tốt hơn B không?

Giải bài toán với n = 100 và m = 55.

Giải

Quy ước rằng nếu A tốt hơn B ghi dấu (+) hay C

B tốt hơn A ghi dấu (–).

Như vậy số dấu (+) bằng số dấu (–) khi hai thuốc như nhau.

Dẫn đến bài toán kiểm định xem P(C) có bằng 0,5 không.

- $H_0 : P(C) = 0,5$

  $H_1 : P(C) > 0,5.$

- Điều kiện:

$np_0 = 100 \times 0,5 = 50$, $n(1 - p_0) = 100 \times 0,5 = 50$. Điều kiện bài toán thoả mãn.

- Tính $s_2'(p_0)$

$$s_2' = 100\times 0,5 + 1,6449\sqrt{100\times 0,5\times 0,5} = 58,2245$$
$$\approx 58.$$

- Kết luận:

$S = 55 \in [0; 58]$: giữ giả thiết H_0.

Hai thuốc tốt như nhau.

Nếu cho rằng có 60% bác sĩ cho là A thật sự tốt hơn B thì sai lầm loại II là bao nhiêu?

$$\beta(0,6) = \Pi\left(\frac{58,2245 - 100 \times 0,6}{\sqrt{100 \times 0,6 \times 0,4}}\right) = \Pi(-0,36)$$

$$= 1 - 0,6406 = 0,3594.$$

Chú ý : Những người cho là A và B như nhau không kể trong n.

CÂU HỎI TỰ LƯỢNG GIÁ

Hãy chọn một kết quả đúng.

1. Khám lao cho 100.000 người thấy 89 người mắc lao. Tỷ lệ mắc lao bằng 0,001 có đúng không? Kiểm định 1 phía, lấy $\alpha = 0,05$.

Kết quả:

A. 120 B. 116 C. 84 D. 80 E. số khác

2. Hỏi 114 người dùng thuốc A hoặc B điều trị một bệnh, có 9 người cho là A, B như nhau, 55 người cho là A tốt hơn B, số còn lại cho là B tốt hơn A. Hãy cho biết có thật sự A tốt hơn B không? Kiểm định 1 phía, với giả thiết thường quy.

Kết quả:

A. 42 B. 44 C. 64 D. 61 E. số khác

3. Người ta cho rằng xác suất chẩn đoán sai bằng 0,125 khi khám kiểm tra cho 120 người. Giả sử xác suất chẩn đoán sai có giá trị đúng bằng 0,08 thì sai lầm loại II bằng bao nhiêu? Kiểm định 2 phía, lấy $\alpha = 0,05$.

Kết quả:

A. 0,715.668 B. 0,5753 C. 0,999.928 D. 0,4247 E. số khác

4. Khi sử dụng thuốc A vì tai biến, có ý kiến đề nghị cấm dùng. Bao nhiêu người đề nghị cấm (C) thì ra lệnh cấm? Nếu quan niệm rằng: $P(c) = 0,5$ với $\alpha = 0,05$ thì $\beta(0,4) \leq 0,15$.

Kết quả:

A. 80 B. số khác C. 129 D. 99 E. 103.

Bài 8

ĐỘ KHÔNG XÁC ĐỊNH (ENTRÔPI)

MỤC TIÊU

1. *Trình bày được khái niệm độ không xác định và lượng tin.*
2. *Giải được bài toán ứng dụng và nêu được ý nghĩa.*

1. KHÁI NIỆM

Đoán kết quả của một phép thử dễ hay khó trước khi thực hiện phép thử, đoán phép thử nào dễ hơn giữa các phép thử... cần phải dựa vào khái niệm độ không xác định của phép thử để trả lời các câu hỏi trên.

1.1. Định nghĩa

Thực hiện phép thử ε, kết quả lập thành nhóm đầy đủ k hiện tượng $A_1, A_2, \ldots, A_k$ tương ứng với các xác suất $p_1, p_2, \ldots, p_k$.

Độ không xác định của phép thử ε ký hiệu H(ε), là giá trị được xác định bởi biểu thức:

$$H(e) = -\sum_{i=1}^{k} p_i \log_a p_i \qquad (1)$$

Đơn vị của H(ε) là thập phân khi a = 10

Đơn vị của H(ε) là nhị phân, gọi tắt là bit, khi a = 2.

Có thể chuyển đổi đơn vị cho nhau.

$\log_{10} p = 0{,}30103 \log_2 p$.

$\log_2 p = 3{,}3219 \log_{10} p$.

1.2. Bảng tra

Bảng cho giá trị $-p\log_2 p$ với $0{,}001 \le p \le 0{,}999$ tại bảng 5.

Khi p = 1, $-p\log_2 p = 0$

Khi $p = \left(\frac{1}{10}\right)^n$, $-\left(\frac{1}{10}\right)^n \log_2\left(\frac{1}{10}\right)^n = 3{,}3219 \times \frac{n}{10^n}$. Như vậy $p \to 0$ thì $-p\log_2 p$ xem là bằng 0.

1.3. Tính chất

- $-\sum_{i=1}^{k} p_i \log_2 p_i \geq 0$ (2)

- $-\sum_{i=1}^{k} p_i \log_2 p_i \leq \log_2 k$ (3)

Chứng minh (3)

$$H(\varepsilon) = -\sum_{i=1}^{k} p_i \log_2 p_i = \sum_{i=1}^{k} \log_2 \left(\frac{1}{p_i}\right)^{p_i}$$

$$= \log_2 \left[\left(\frac{1}{p_1}\right)^{p_1} \cdot \left(\frac{1}{p_2}\right)^{p_2} \cdots \left(\frac{1}{p_k}\right)^{p_k}\right]$$

$$H(\varepsilon) \leq \log_2 \left\{\frac{1}{k}\left[\left(\frac{1}{p_1}\right)^{p_1} + \left(\frac{1}{p_2}\right)^{p_2} + \ldots + \left(\frac{1}{p_k}\right)^{p_k}\right]\right\}^k \quad \text{(theo Cauchy)}$$

$$\leq \log_2 k.$$

1.4. Ý nghĩa

Độ không xác định cho biết đoán nhận kết quả của phép thử khó hay dễ.

Độ không xác định nhỏ, gần 0 dễ đoán nhận kết quả phép thử. Phép thử gần xác định.

Độ không xác định lớn, gần $\log_2 k$, khó đoán nhận kết quả phép thử. Phép thử khó xác định.

Ví dụ

– Bác sỹ thứ nhất chữa một bệnh có xác suất khỏi bằng 1, xác suất không khỏi bằng 0; $H(\varepsilon_1) = 0$. Phép thử xác định, dễ đoán nhận kết quả phép thử.

– Bác sỹ thứ hai chữa bệnh đó có xác suất khỏi bằng 0,6, xác suất không khỏi bằng 0,4; $H(\varepsilon_2) = -0{,}6 \log_2 0{,}6 - 0{,}4 \log_2 0{,}4 = 0{,}971 \approx \log_2 2 = 1$.

Độ không xác định lớn, khó đoán nhận kết quả của phép thử.

Đoán bác sỹ thứ hai chữa khỏi hay không, khó hơn đoán bác sỹ thứ nhất chữa khỏi hay không.

– Bác sỹ thứ ba chữa bệnh trên có xác suất khỏi bằng 0,4, xác suất không khỏi bằng 0,6.

$H(\varepsilon_3) = H(\varepsilon_2) = 0{,}971.$

Độ không xác định không đề cập tới nội dung hay bản chất các hiện tượng, chỉ quan tâm tới xác suất của chúng và đoán nhận kết quả phép thử khó hay dễ. Bản chất các hiện tượng cần được đề cập ở dạng nghiên cứu khác.

2. ĐỘ KHÔNG XÁC ĐỊNH CỦA HAI PHÉP THỬ

Giả sử α là phép thử có k kết quả lập thành nhóm đầy đủ các hiện tượng A_1, A_2, ..., A_k tương ứng với các xác suất $P(A_i)$, $i = \overline{1,k}$.

Giả sử β là phép thử có n kết quả lập thành nhóm đầy đủ các hiện tượng B_1, B_2, ..., B_n tương ứng với các xác suất $P(B_j)$, $j = \overline{1,n}$.

Thực hiện đồng thời 2 phép thử α và β. Kết quả thu được là một bảng gồm kxn hiện tượng. Dựa vào bảng kết quả, xây dựng khái niệm độ không xác định của hai phép thử α và β.

α \ β	B_1	...	B_j	...	B_n
A_1	A_1B_1		A_1B_j		A_1B_n
...	...		...		...
A_i	A_iB_1	...	A_iB_j	...	A_iB_n
...	...		...		...
A_k	A_kB_1	...	A_kB_j	...	A_kB_n

2.1. Định nghĩa

Độ không xác định của hai phép thử α và β, ký hiệu $H(\alpha.\beta)$, là giá trị được xác định bởi biểu thức

$$H(a.b) = -\sum_{i,j=1}^{k,n} P(A_iB_j)\log_2 P(A_iB_j) \qquad (4.1)$$

Chú ý: Từ đây trở đi sẽ không viết cơ số 2 của loga và đơn vị độ không xác định là bit hay nhị phân.

2.2. Công thức tính

2.2.1. α và β là hai phép thử không độc lập.

- Khai triển hàng 1 với ký hiệu $H_1(\alpha.\beta)$.

$$\begin{aligned} H_1(\alpha.\beta) &= -P(A_1B_1)\log P(A_1B_1) - \ldots - P(A_1B_n)\log P(A_1B_n) \\ &= -P(A_1).P(B_1/A_1)[\log P(A_1) + \log P(B_1/A_1)]\ldots \\ &\quad -P(A_1).P(B_n/A_1)[\log P(A_1) + \log P(B_n/A_1)] \end{aligned}$$

$$= -\ P(A_1)\log P(A_1)\left[P(B_1/A_1)+\ \dots\ +P(B_n/A_1)\right]+$$
$$+\ P(A_1)\left[-P(B_1/A_1)\log P(B_1/A_1)-\dots-P(B_n/A_1)\log P(B_n/A_1)\right]$$
$$= -\ P(A_1)\log P(A_1)+P(A_1).\left(-\sum_{j=1}^{n}P(B_j/A_1)\log P(B_j/A_1)\right)$$
$$= -\ P(A_1)\log P(A_1)+P(A_1).H_{A_1}(\beta),\ \text{với}$$

$$H_{A_1}(\beta)\ = -\ \sum_{j=1}^{n}P(B_j/A_1)\log P(B_j/A_1)$$

Tương tự trên

$$H_2(\alpha.\beta)\ = -\ P(A_2)\log P(A_2)+P(A_2)H_{A_2}(\beta)$$

. . . .

$$H_k(\alpha.\beta)\ = -\ P(A_k)\log P(A_k)+P(A_k)H_{A_k}(\beta).$$

Cộng các hàng từ 1 đến k dẫn đến:

$$H(\alpha.\beta)\ = -\sum_{i=1}^{k}P(A_i)\log P(A_i)+\sum_{i=1}^{k}P(A_i)H_{Ai}(\beta)$$

$$H(\alpha.\beta)\ = H(\alpha)+H_\alpha(\beta) \qquad (4.2)$$

trong đó: $H_\alpha(\beta)\ = \sum_{i=1}^{k}P(A_i)H_{A_i}(\beta)$

Khai triển tương tự theo cột có kết quả:

$$H(\alpha.\beta)=H(\beta)+H_\beta(\alpha)=H(\beta.\alpha) \qquad (4.3)$$

trong đó: $H_\beta(\alpha)\ = \sum_{j=1}^{n}P(B_j).H_{B_j}(\alpha)$.

Trong công thức (4.2) và (4.3) $H_\alpha(\beta)$ và $H_\beta(\alpha)$ được gọi là các độ không xác định có điều kiện.

2.2.2. α và β là hai phép thử độc lập

$$H(\alpha.\beta)=H(\alpha)+H(\beta) \qquad (4.4)$$

$$H_\alpha(\beta)=H(\beta),\qquad H_\beta(\alpha)=H(\alpha)$$

2.3. Tính chất của $H_\alpha(\beta)$

- $H_\alpha(\beta)\ =0$ khi β là phép thử đã xác định.

- $H_\alpha(\beta) = H(\beta)$ khi α và β là hai phép thử độc lập.

- Khi α không độc lập với β nghĩa là α và β phụ thuộc nhau, vì vậy α xảy ra trước làm giảm độ không xác định của β.

$$H_\alpha(\beta) \le H(\beta).$$

- Giả sử $H(\beta) > H(\alpha)$.

Từ (4.2) và (4.3) suy ra

$$H_\alpha(\beta) = H(\beta) - H(\alpha) + H_\beta(\alpha).$$

$H_\alpha(\beta) \ge H(\beta) - H(\alpha)$. Dấu "=" đạt được khi $H_\beta(\alpha) = 0$.

Khi đó β là phép thử đã xảy ra sẽ xác định các kết quả của α

Tóm lại luôn luôn có $0 \le H(\beta) - H(\alpha) \le H_\alpha(\beta) \le H(\beta)$.

Ví dụ

1. Tỷ lệ mắc một bệnh tại một cộng đồng bằng 2%. Dùng một xét nghiệm kiểm tra, nếu người bị bệnh, xét nghiệm luôn luôn dương tính; nếu người không bị bệnh, xét nghiệm 50% dương tính và 50% âm tính. Ký hiệu α là phép thử xét nghiệm dương tính A hay âm tính $\bar{A}$, β là phép thử xác định bị bệnh B hay không bị bệnh $\bar{B}$. Tính H(α β).

Giải

Từ số liệu bài cho

$P(B) = 0{,}02$ $\qquad$ $P(\bar{B}) = 0{,}98$

$P(A/B) = 1$ $\qquad$ $P(\bar{A}/B) = 0$

$P(A/\bar{B}) = 0{,}5$ $\qquad$ $P(\bar{A}/\bar{B}) = 0{,}5$

Tính H(α β)

$$H(\alpha.\beta) = H(\beta) + H_\beta(\alpha)$$

$$\begin{aligned} H(\beta) &= -P(B)\log P(B) - P(\bar{B})\log P(\bar{B}) \\ &= -0{,}02\log 0{,}02 - 0{,}98\log 0{,}98 = 0{,}1415. \end{aligned}$$

$$\begin{aligned} H_\beta(\alpha) &= P(B).H_B(\alpha) + P(\bar{B})H_{\bar{B}}(\alpha) \\ &= 0{,}02\left[-1\log 1 - 0\log 0\right] + 0{,}98\left[-0{,}5\log 0{,}5 - 0{,}5\log 0{,}5\right] \\ &= 0{,}98. \end{aligned}$$

$H(\alpha\,\beta) = 0{,}1415 + 0{,}98 = 1{,}1215.$

Tính H(α β) theo công thức (4.1)

$P(AB) = 0,02 \times 1 = 0,02$ $\qquad$ $P(\bar{A}B) = 0,02 \times 0 = 0$

$P(A\bar{B}) = 0,98 \times 0,5 = 0,49$ $\qquad$ $P(\bar{A}\,\bar{B}) = 0,98 \times 0,5 = 0,49$

$$H(\alpha.\beta) = -0,02\ \log 0,02 - 0 - 0,49\ \log 0,49 - 0,49\ \log 0,49$$
$$= 1,1215$$

2. Tỷ lệ sinh con trai bằng 0,514. Đoán hai phụ nữ cùng sinh mỗi người một con là trai hay gái dễ hay khó?

Giải

Ký hiệu

$E_1 = T_1T_2$	$P(E_1) = 0,264$	$E_2 = T_1G_2$	$P(E_2) = 0,250$
$E_3 = G_1T_2$	$P(E_3) = 0,250$	$E_4 = G_1G_2$	$P(E_4) = 0,236$.

α là phép thử phụ nữ thứ nhất sinh con trai hoặc gái, β là phép thử phụ nữ thứ hai sinh con trai hoặc gái.

$$H(\alpha.\beta) = -\ 0,264\ \log 0,264 - 2 \times 0,25\ \log 0,25 - 0,236\ \log 0,236$$

$$= \quad 1,9988 \approx \log_2 4 = 2.$$ Vậy rất khó đoán.

Vì α và β là hai phép thử độc lập, dẫn đến:

$$H(\alpha.\beta) = H(\alpha) + H(\beta) = 2 \times 0,9994 = 1,9988.$$

3. KHÁI NIỆM VỀ LƯỢNG TIN

3.1. Định nghĩa

$H(\beta)$ đặc trưng cho độ không xác định của phép thử β.

$H(\beta)$ lớn hay nhỏ thì sự đoán nhận kết quả của β khó hoặc dễ.

$H(\beta) = 0$ chứng tỏ kết quả của phép thử β đã biết hay β là phép thử xác định.

Nếu một phép thử α nào đó xảy ra trước khi thực hiện phép thử β, có thể xảy ra các trường hợp sau:

- $H_\alpha(\beta) = H(\beta)$ nếu α độc lập với β. Như vậy việc thực hiện trước α không làm giảm độ không xác định của β.

- Nếu kết quả của α hoàn toàn xác định kết quả của β thì độ không xác định của β với điều kiện α đã xảy ra sẽ giảm tới 0: $H_\alpha(\beta) = 0$.

- $H_\alpha(\beta) < H(\beta)$: α xảy ra trước β đã làm giảm độ không xác định của β. Vì vậy có thể xem α là phép thử phụ.

Hiệu $H(\beta) - H_\alpha(\beta)$ chỉ ra rằng việc thực hiện α trước sẽ làm giảm độ không xác định của β đến mức nào, nghĩa là việc thực hiện α làm cho biết thêm một chút gì đó về β.

Lượng thông tin về phép thử β chứa trong phép thử α, ký hiệu là $I(\alpha, \beta)$, được xác định bởi biểu thức:

$$I(\alpha, \beta) = H(\beta) - H_\alpha(\beta) \qquad (5)$$

Thường gọi vắn tắt lượng thông tin là lượng tin

Với quan niệm lượng tin như trên, cũng có thể nói $H(\beta)$ cũng là lượng tin.

Thật vậy $I(\beta, \beta) = H(\beta) - H_\beta(\beta)$.

Việc thực hiện β rồi tất nhiên hoàn toàn xác định kết quả của nó, do đó $H_\beta(\beta) = 0$

Vậy $I(\beta, \beta) = H(\beta)$.

$H(\beta)$ chính là lượng tin về β chứa trong bản thân phép thử đó.

Cũng có thể hiểu $H(\beta)$ là lượng thông tin lớn nhất về β mà nó có thể có, hay là lượng thông tin đầy đủ về β.

Hoặc nói cách khác $H(\beta)$ là lượng tin nhận được sau khi thực hiện phép thử β. Cũng có thể nói rằng $H(\beta)$ là lượng tin trung bình chứa trong các kết quả của phép thử β.

3.2. Ví dụ

Tại β_1 xác suất bị bệnh B của phòng khám bằng 0,4.

Tại β_2 xác suất bị bệnh B của phòng khám bằng 0,8.

Sử dụng phương pháp chẩn đoán mới tại β_1, với khẳng định có bệnh thì đúng 3/5 trường hợp; với khẳng định không bệnh thì đúng 4/5 trường hợp.

Sử dụng phương pháp chẩn đoán trên tại β_2, với khẳng định có bệnh thì đúng 9/10 trường hợp; với khẳng định không bệnh thì đúng 5/10 trường hợp.

a) Tại nơi nào phương pháp chẩn đoán mới cho một lượng tin về bệnh lớn hơn?

b) Tìm xác suất chẩn đoán đúng tại hai nơi.

Giải

Gọi β_1 là phép thử xác định một người có bệnh B hay không tại β_1.

β_2 là phép thử xác định một người có bệnh B hay không tại β_2.

Gọi α là phép thử xác định phương pháp chẩn đoán mới dương tính A hay âm tính $\overline{A}$

- Xét tại β_1

Các xác suất đã cho:

$$P(B) = 0,4 \qquad P(\overline{B}) = 0,6$$

$$P(B/A) = 0,6 \qquad P(\overline{B}/A) = 0,4$$

$$P(\overline{B}/\overline{A}) = 0,8 \qquad\qquad P(B/\overline{A}) = 0,2.$$

Tính $I(\alpha, \beta_1)$

$$I(\alpha, \beta_1) = H(\beta_1) - H_\alpha(\beta_1)$$

Cần tính P(A).

$$P(B) = P(A)P(B/A) + P(\overline{A}).P(B/\overline{A})$$

$$0,4 = P(A) \times 0,6 + [1 - P(A)] \times 0,2$$

$P(A) = 0,5$, $P(\overline{A}) = 0,5$

Tính $H(\beta_1)$ và $H_\alpha(\beta_1)$

$$H(\beta_1) = -0,4 \log 0,4 - 0,6 \log 0,6 = 0,971.$$

$$H_\alpha(\beta_1) = P(A)H_A(\beta_1) + P(\overline{A})H_{\overline{A}}(\beta_1)$$

$$= 0,5[-0,6\log 0,6 - 0,4\log 0,4] + 0,5[-0,8\log 0,8 - 0,2\log 0,2]$$

$$= 0,84645.$$

$$I(\alpha, \beta_1) = 0,971 - 0,84645 = 0,12455.$$

Tính P(Đ)

$$P(Đ) = P(A)\, P(B/A) + P(\overline{A})\, P(\overline{B}/\overline{A})$$

$$= 0,5 \times 0,6 + 0,5 \times 0,8 = 0,7.$$

- Xét tại β_2

Các xác suất đã cho

$$P(B) = 0,8 \qquad\qquad P(\overline{B}) = 0,2$$

$$P(B/A) = 0,9 \qquad\qquad P(\overline{B}/A) = 0,1$$

$$P(\overline{B}/\overline{A}) = 0,5 \qquad\qquad P(B/\overline{A}) = 0,5.$$

Tính $I(\alpha, \beta_2)$

$$I(\alpha, \beta_2) = H(\beta_2) - H_\alpha(\beta_2)$$

Tính P(A)

$$P(B) = P(A)P(B/A) + P(\overline{A}).P(B/\overline{A})$$

$$0,8 = P(A) \times 0,9 + [1 - P(A)] \times 0,5$$

$$P(A) = 0,75\ ,\ P(\overline{A}) = 0,25.$$

Tính $H(\beta_2)$ và $H_\alpha(\beta_2)$

$$H(\beta_2) = -P(B) \log P(B) - P(\overline{B}) \log(\overline{B})$$

$= -0,8\log 0,8 - 0,2\log 0,2 = 0,7219$

$H_\alpha(\beta_2) = P(A)H_A(\beta_2) + P(\bar{A})H_{\bar{A}}(\beta_2)$

$= 0,75[-0,9\log 0,9 - 0,1\log 0,1] + 0,25[-0,5\log 0,5 - 0,5\log 0,5]$

$= 0,75 \times 0,469 + 0,25 \times 1 = 0,60175.$

$I(\alpha, \beta_2) = 0,7219 - 0,60175 = 0,12015.$

Tính P(Đ)

P(Đ) $= P(A)\,P(B/A) + P(\bar{A})\,P(\bar{B}/\bar{A})$

$= 0,75 \times 0,9 + 0,25 \times 0,5 = 0,8.$

- Nhận xét

$I(\alpha, \beta_1) > I(\alpha, \beta_2)$

Lượng tin về bệnh chứa trong phương pháp chẩn đoán mới tại β_1 lớn hơn tại β_2 nghĩa là dùng phương pháp chẩn đoán mới giúp chẩn đoán bệnh B tại β_2 có giá trị hơn tại β_1. Điều này cũng phù hợp với xác suất chẩn đoán đúng tại β_2 lớn hơn xác suất chẩn đoán đúng tại β_1.

3.3. Tính chất

- $0 \le I(\alpha, \beta) = H(\beta) - H_\alpha(\beta) \le H(\beta)$.
- $I(\alpha, \beta) = 0$

Hoặc $H(\beta) = H_\alpha(\beta)$: phép thử α độc lập với phép thử β.

Hoặc $H(\beta) = 0$: β là phép thử đã xác định.

- $I(\alpha, \beta) = H(\beta)$. Khi đó $H_\alpha(\beta) = 0$: phép thử α xác định các kết quả của phép thử β.
- $I(\alpha, \beta)$ không đề cập tới nội dung hay bản chất của các hiện tượng.
- $I(\alpha, \beta) = I(\beta, \alpha)$.

Thật vậy: $H(\alpha.\beta) = H(\alpha) + H_\alpha(\beta) = H(\beta) + H_\beta(\alpha) \Leftrightarrow$

$$I(\beta, \alpha) = H(\alpha) - H_\beta(\alpha) = H(\beta) - H_\alpha(\beta) = I(\alpha, \beta) \quad (6)$$

$$= H(\alpha) + H(\beta) - H(\alpha.\beta) \quad (7)$$

Lượng tin về phép thử β chứa trong phép thử α bằng lượng tin về phép thử α chứa trong phép thử β. Như vậy $I(\alpha, \beta)$ là lượng tin tương hỗ giữa hai phép thử α và β.

3.4. Công thức khác

Gọi ε là phép thử xác định chẩn đoán đúng hay sai. Tương tự công thức (6), (7) có các công thức tính $I(\alpha, \varepsilon)$ và $I(\beta, \varepsilon)$.

$I(\alpha, \varepsilon) = H(\alpha) - H_\varepsilon(\alpha) = H(\varepsilon) - H_\alpha(\varepsilon) = H(\alpha) + H(\varepsilon) - H(\alpha.\varepsilon)$

$I(\beta, \varepsilon) = H(\beta) - H_\varepsilon(\beta) = H(\varepsilon) - H_\beta(\varepsilon) = H(\beta) + H(\varepsilon) - H(\beta.\varepsilon)$

Mặt khác từ định nghĩa dẫn đến:

$H(\alpha.\beta) = H(\alpha.\varepsilon) = H(\beta.\varepsilon)$

$H_\alpha(\beta) = H_\alpha(\varepsilon) \quad H_\beta(\alpha) = H_\beta(\varepsilon) \quad H_\varepsilon(\alpha) = H_\varepsilon(\beta).$

CÂU HỎI TỰ LƯỢNG GIÁ

Hãy chọn một kết quả đúng.

1. Điều trị một bệnh có xác suất khỏi bằng 0,8. Điều trị cho 2 người, đoán ai khỏi ai không khó hay dễ?

Kết quả:

A. 1,124 B. 1,444 C. 1,021 D. 0,722 E. số khác.

2. Một phòng cấp cứu điều trị cho 3 bệnh nhân nặng A, B, C. Xác suất cấp cứu của A, B, C trong 1 giờ tương ứng bằng 0,4 0,3 0,2. Đoán xem mấy người cấp cứu mấy người không trong 1 giờ khó hay dễ?

Kết quả:

A. 0,971 B. 2,6042 C. 1,6289 D. 1,1002 E. số khác.

3. Dùng một phản ứng chẩn đoán bệnh, phản ứng có độ nhạy bằng 0,9 và độ đặc hiệu bằng 0,75. Xác suất chẩn đoán đúng bằng 0,81. Kí hiệu α: phản ứng dương tính hay âm tính; ε: đúng hay sai. Tính $H_\alpha(\varepsilon)$.

Kết quả:

A. 1,617 B. 0,674 C. 0,646 D. 0,701 E. số khác.

4. Một xét nghiệm có xác suất đúng bằng 0,95. Tỷ lệ bị bệnh tại cộng đồng bằng 0,2. Độ nhạy của xét nghiệm bằng 0,8. Dùng xét nghiệm chẩn đoán bệnh. Gọi α là xét nghiệm dương tính hay âm tính; ε: đúng hay sai. Tính $I(\alpha, \varepsilon)$.

Kết quả:

A. 0,065 B. 0,436 C. 0,944 D. 0,000 E. số khác.

Bài 9

PHƯƠNG PHÁP BÌNH PHƯƠNG BÉ NHẤT VÀ ỨNG DỤNG

MỤC TIÊU

1. *Lập được hàm tương quan tuyến tính, hàm tương quan bậc 2.*
2. *Lập được hàm mật độ xác suất chuẩn nhờ tuyến tính hóa.*

1. BÀI TOÁN

Giả sử trên mỗi đối tượng nghiên cứu thu được hai giá trị x và y của hai đại lượng X và Y. Kết quả của n đối tượng nghiên cứu được cho trong bảng sau:

X	x_1	x_2	...	x_n
Y	y_1	y_2	...	y_n

Giả sử giữa Y và X có mối tương quan hàm số $y = ax + b$ hay $y = ax^2 + bx + c$ ..., từ n cặp giá trị hãy lập hàm số $y = f(x)$.

2. LẬP HÀM BẬC NHẤT

2.1 . Giải bài toán

Biểu diễn điểm $M(x_i, y_i)$ trên mặt phẳng toạ độ 0xy.

Giả sử hàm số $y = ax + b$ đã lập được

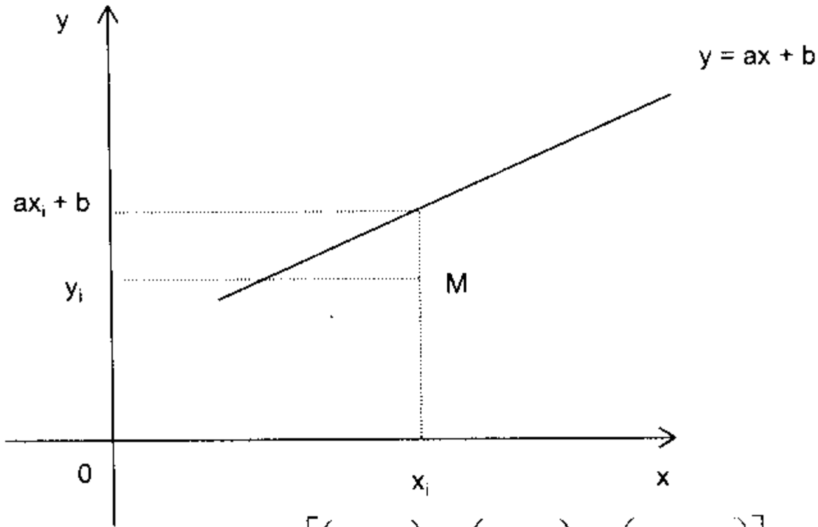
Gọi δ_i là bình phương khoảng lệch thứ i: $\delta_i = (ax_i + b - y_i)^2$.

Với n điểm $\sum_{i=1}^{n}\delta_i = \sum_{i=1}^{n}(ax_i + b - y_i)^2$

Hàm số $y = ax + b$ được lập với điều kiện

$$f(a,b) = \sum_{i=1}^{n}(ax_i + b - y_i)^2 \text{ bé nhất.} \qquad (1)$$

Khi đó cần tính các đạo hàm f'_a và f'_b.



$$f'_a = 2\sum_{i=1}^{n}(ax_i + b - y_i)x_i = 2\left[\left(\sum_{i=1}^{n} x_i^2\right)a + \left(\sum_{i=1}^{n} x_i\right)b - \left(\sum_{i=1}^{n} x_i y_i\right)\right]$$

$$f'_b = 2\sum_{i=1}^{n}(ax_i + b - y_i)1 = 2\left[\left(\sum_{i=1}^{n} x_i\right)a + n.b - \sum_{i=1}^{n} y_i\right]$$

Giải hệ 2 phương trình bậc nhất đối với a và b.

$$\begin{cases} f'_a = 0 \\ f'_b = 0 \end{cases} \Leftrightarrow \begin{cases} \left(\sum_{i=1}^{n} x_i^2\right)a + \left(\sum_{i=1}^{n} x_i\right)b = \sum_{i=1}^{n} x_i y_i \\ \left(\sum_{i=1}^{n} x_i\right)a + \quad n.b \qquad = \sum_{i=1}^{n} y_i \end{cases} \tag{2}$$

Giải hệ (2) $\det D = \begin{vmatrix} \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & n \end{vmatrix} = n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2$

$$\det D_a = \begin{vmatrix} \sum_{i=1}^{n} x_i y_i & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} y_i & n \end{vmatrix} = n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \times \sum_{i=1}^{n} y_i$$

$$\det D_b = \begin{vmatrix} \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i y_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} y_i \end{vmatrix} = \sum_{i=1}^{n} x_i^2 \times \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i . \sum_{i=1}^{n} x_i y_i$$

Khi det D ≠ 0, hệ xác định

$$a = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i . \sum_{i=1}^{n} y_i}{n.\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \qquad (3.1)$$

$$b = \frac{\sum_{i=1}^{n} x_i^2 . \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i}{n.\sum x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \qquad (4.1)$$

Tính đạo hàm f''_a, f''_b và f''_{ab}

$f''_a = 2\left(\sum_{i=1}^{n} x_i^2\right)$; $f''_a > 0$ ∀a. Hàm f(a, b) đạt cực tiểu tại a tìm được.

$f''_b = 2n$; $f''_b > 0$ ∀b. Hàm f(a, b) đạt cực tiểu tại b tìm được.

$f''_{ab} = 2\left(\sum_{i=1}^{n} x_i\right)$. Với điều kiện $\left(f''_{ab}\right)^2 - f''_a.f''_b < 0$ dẫn đến a và b tính theo (3.1) và (4.1) là các giá trị thoả mãn điều kiện (1).

Hàm số y = ax + b đã tìm được.

Hàm số $x = a'y + b'$ được lập tương tự trên sao cho $f(a', b') = \sum_{i=1}^{n}\left(a'y_i + b' - x_i\right)^2$ bé nhất.

2.2. Công thức tính

Từ (3.1), các công thức tính khác phụ thuộc vào số liệu thu được:

- Chia tử số và mẫu cho n^2 ta được $a = \frac{\overline{xy} - \overline{x}\,.\,\overline{y}}{\overline{x^2} - \overline{x}^2}$ (3.2)
- x_i và y_i quá lớn hoặc là số thập phân hoặc các số cách đều nhau:

$$a = \frac{\overline{u.v} - \overline{u}\ .\ \overline{v}}{\overline{u^2} - \overline{u}^2} \times \frac{\Delta y}{\Delta x}, \quad (3.3)$$

trong đó $u_i = \dfrac{x_i - x_0}{\Delta x}$, $v_i = \dfrac{y_i - y_0}{\Delta y}$ với x_0, y_0, Δx, Δy tuỳ chọn.

Trong tính toán, không tính b theo (4.1). Từ phương trình (2) với a đã biết dẫn đến:

$$b = \overline{y} - a.\overline{x} \quad (4.2)$$

Nhận xét: Hàm số y = ax + b luôn luôn đi qua điểm $M(\overline{x}, \overline{y})$.

Ví dụ

1. Cho 2 dãy số liệu

X	1	2	3	4	5
Y	3	5	7	9	11

Lập hàm số y = ax + b thoả mãn điều kiện (1).

Giải

- Lập bảng tính với $u_i = x_i - 3$, $v_i = (y_i - 7)/2$.

i	x_i	y_i	u_i	v_i	u_i^2	u_iv_i
1	1	3	–2	–2	4	4
2	2	5	–1	–1	1	1
3	3	7	0	0	0	0
4	4	9	1	1	1	1
5	5	11	2	2	4	4
Σ			0	0	10	10
TB			0	0	2	2

- Tính các tham số

$$a = \frac{\overline{u.v} - \overline{u}\ .\ \overline{v}}{\overline{u^2} - (\overline{u})^2} \times \frac{\Delta y}{\Delta x} = \frac{2 - 0\times 0}{2 - 0^2} \times \frac{2}{1} = 2$$

$$\overline{x} = x_0 + \Delta x\ .\ \overline{u} \quad = 3 + 1\times 0 \quad = 3$$

$$\overline{y} = y_0 + \Delta y\ .\ \overline{v} \quad = 7 + 2\times 0 \quad = 7$$

$$b = \overline{y} - a\overline{x} \quad = 7\text{-}2\times 3 \quad = 1$$

Hàm số bậc nhất cần lập có dạng y = 2x + 1.

2. Đo một đại lượng tại hai điểm khác nhau trên cơ thể. Tại điểm I bằng phương pháp I, ký hiệu là X; tại điểm II bằng phương pháp II, ký hiệu là Y, thu được số liệu sau:

X	32,6	34,5	39,0	39,1	39,1	39,3	39,7	42,3	45,4	53,3	59,4	71,9
Y	32,3	37,6	39,2	37,4	39,6	40,9	39,0	42,8	46,1	55,6	55,1	71,3

Lập hàm số y = ax + b thoả mãn điều kiện (1).

Giải

- Lập bảng tính với $u_i = (x_i - 39{,}0)/0{,}1$ và $v_i = (y_i - 40{,}0)/0{,}1$.

i	x_i	y_i	u_i	v_i	u_i^2	v_i^2	u_iv_i
1	32,6	32,3	–64	–77	4.096	5.929	4.928
2	34,5	37,6	–45	–24	2.025	576	1.080
3	39,0	39,2	0	–8	0	64	0
4	39,1	37,4	1	–26	1	676	–26
5	39,1	39,6	1	–4	1	16	–4
6	39,3	40,9	3	9	9	81	27
7	39,7	39,0	7	–10	49	100	–70
8	42,3	42,8	33	28	1.089	784	924
9	45,4	46,1	64	61	4.096	3.721	3.904
10	53,3	55,6	143	156	20.449	24.336	22.308
11	59,4	55,1	204	151	41.616	22.801	30.804
12	71,9	71,3	329	313	108.241	97.969	102.977
Σ			**676**	**569**	**181.672**	**157.053**	**166.852**

- Tính các tham số.

$$a = \frac{12\times 166.852 - 676\times 569}{12\times 181.672 - 676\times 676}\times\frac{0.1}{0.1} = 0.938768072$$

$$b = \frac{1}{12}\left[12\times 40 + 0{,}1\times 569 - 0{,}938.768.072\times(12\times 39 + 0{,}1\times 676)\right]$$

$$= 2{,}841.318.386$$

Hàm số cần lập có phương trình:

$y = 0{,}938.768.072x + 2{,}841.318.386.$

3. Theo dõi số dân (X, đv: 1000người) và tỷ lệ sinh (Y, đv: $^0/_{00}$) của cả nước thu được số liệu sau:

X	64.412	66.233	67.744	69.405	71.026	72.510	73.959
Y	31,3	29,9	30,4	30,0	28,5	28,3	25,3

Lập phương trình hàm số y = ax + b.

Giải

- Tính các kết quả trung gian

$\sum x$	$\sum x^2$	$\sum y$	$\sum y^2$	$\sum xy$
485.289	$3{,}371.434.605\times 10^{10}$	203,7	5.951,09	14.085.466,6

- Tính các tham số

  a = $-5{,}153.505.987 \times 10^{-4} = -0{,}000.515.350.5987$

  b = 64,827.710.96.

Phương trình cần lập có dạng:

y = – 0,000.515.350.5987x + 64,827.710.96.

3. LẬP HÀM BẬC HAI

3.1. Giải bài toán

Từ n cặp giá trị (x_i, y_i), $i = \overline{1,n}$ lập hàm bậc hai $y = ax^2 + bx + c$.

Làm tương tự như lập hàm bậc nhất.

Gọi δ_i là bình phương khoảng lệch thứ i: $\delta_i = (a x_i^2 + bx_i + c - y_i)^2$

Với n điểm:

$$\sum_{i=1}^{n} \delta_i = \sum_{i=1}^{n} (ax_i^2 + bx_i + c - y_i)^2 \text{, ký hiệu là } f(a, b, c).$$

Tìm các tham số a, b, c sao cho:

$$f(a,b,c) = \sum_{i=1}^{n} (ax_i^2 + bx_i + c - y_i)^2 \text{ bé nhất.} \qquad (1')$$

Tính các đạo hàm, cho các đạo hàm bằng 0 dẫn đến hệ phương trình:

$$\begin{cases} \left(\sum_{i=1}^{n} x_i^4\right)a + \left(\sum_{i=1}^{n} x_i^3\right)b + \left(\sum_{i=1}^{n} x_i^2\right)c = \sum_{i=1}^{n} x_i^2 y_i \\ \left(\sum_{i=1}^{n} x_i^3\right)a + \left(\sum_{i=1}^{n} x_i^2\right)b + \left(\sum_{i=1}^{n} x_i\right)c = \sum_{i=1}^{n} x_i y_i \\ \left(\sum_{i=1}^{n} x_i^2\right)a + \left(\sum_{i=1}^{n} x_i\right)b + n.c = \sum_{i=1}^{n} y_i \end{cases} \qquad (2')$$

Giải hệ (2′) theo phương pháp Gauss hoặc theo phương pháp Cramer sẽ tìm được a, b, c.

Với các điều kiện phức tạp thường không xét a, b, c tìm được là các tham số thoả mãn điều kiện (1′).

Ví dụ

Cho 2 dãy số liệu

X	1	2	3	4
Y	–2	0	4	10

Lập hàm số bậc hai $y = ax^2 + bx + c$ từ số liệu trên.

Giải

- Lập bảng tính các hệ số theo (2′)

i	x_i	y_i	x_i^2	x_i^3	x_i^4	$x_i^2y_i$	x_iy_i
1	1	–2	1	1	1	–2	–2
2	2	0	4	8	16	0	0
3	3	4	9	27	81	36	12
4	4	10	16	64	256	160	40
Σ	10	12	30	100	354	194	50

- Dựa vào hệ (2′) lập được ma trận mở rộng A′

$$A' = \begin{bmatrix} 354 & 100 & 30 & 194 \\ 100 & 30 & 10 & 50 \\ 30 & 10 & 4 & 12 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 0,3 & 0,1 & 0,5 \\ 0 & -6,2 & -5,4 & 17 \\ 0 & 1 & 1 & -3 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 1 & 0,3 & 0,1 & 0,5 \\ 0 & 1 & 54/62 & -170/62 \\ 0 & 0 & 1 & -2 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -2 \end{bmatrix}$$

$a = 1,\ b = -1,\ c = -2$

Phương trình bậc 2 có biểu thức $y = x^2 - x - 2$.

4. PHƯƠNG PHÁP TUYẾN TÍNH HOÁ

4.1. Các hàm cần lập

- Hàm mũ

$$y = Ae^{Bx}$$

Lấy loga cơ số e hai vế:

$$\ln y = \ln A + Bx$$

Đặt $Y = \ln y$, $a = \ln A$. Phương trình cần lập có dạng:

$$Y = a + Bx.$$

Từ phương trình lập được suy ra $y = Ae^{Bx}$ với $y = e^{Y}$ và $A = e^{a}$.

- Hàm loga

$$y = A + B\ln x.$$

Đặt $X = \ln x$, phương trình cần lập có dạng $y = A + BX$.

- Hàm luỹ thừa

$$y = A.x^{B}$$

Lấy loga cơ số e hai vế:

$$\ln y = \ln A + B\ln x.$$

Đặt $Y = \ln y$, $X = \ln x$ và $a = \ln A$, phương trình cần lập có dạng:

$$Y = a + BX.$$

Lấy "e mũ" hai vế ta được hàm luỹ thừa cần lập.

- Hàm nghịch đảo

$$y = A + \frac{B}{x}.$$

Đặt $X = \frac{1}{x}$, phương trình cần lập có dạng $y = A + BX$.

Ví dụ

1. Đo áp lực động mạch phổi thì tâm trương người bình thường (X, mmHg) thu được số liệu sau:

x_i	2	3	4	5	6	7	8	9
m_i	1	4	7	8	2	5	2	1

Giả sử X là đại lượng ngẫu nhiên chuẩn, hãy lập hàm mật độ xác suất của X từ đó cho biết MX và DX.

Giải

- Theo giả thiết $X : N(\mu, \sigma^2)$. Khi đó hàm mật độ xác suất của X có dạng:

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

Dẫn đến $p_i = P\{\alpha_{i-1} < X \le \alpha_i\} = \int_{\alpha_{i-1}}^{\alpha_i} f(t)dt = f(\xi)(\alpha_i - \alpha_{i-1}) \approx f(x_i)$,

với $\alpha_i - \alpha_{i-1}$ thường xấp xỉ bằng 1.

Lấy loga cơ số e hai vế của hàm mật độ xác suất f(x)

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

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

Phương trình cần lập có dạng:

$$Y = ax^2 + bx + c,$$

trong đó $Y = \ln f(x)$, $a = -\frac{1}{2\sigma^2}$, $b = \frac{\mu}{\sigma^2}$, $c = -\left[\frac{\mu^2}{2\sigma^2} + \ln(\sigma\sqrt{2\pi})\right]$

Từ đó suy ra μ và σ^2.

- Từ dãy số liệu đã cho (x_i, m_i), $i = \overline{1,8}$, suy ra dãy (x_i, y_i) $i = \overline{1,8}$.

x_i	2	3	4	5
y_i	–3,401.197.382	–2,014.903.021	–1,455.287.233	–1,321.755.84
x_i(tiếp)	6	7	8	9
y_i(tiếp)	–2,708.050.201	–1,791.759.469	–2,708.050.201	–3,401.197.382

Phương trình cần lập có biểu thức sau:

$$Y = -0,133.628.251x^2 + 1,400.131.636x - 5,307.196.161.$$

Dẫn đến $\sigma^2 = 3,741.723.747 = 1,934.353.573^2 \approx 1,934^2$

$\mu = 5,238.905.791 \approx 5,239.$

Hàm mật độ xác suất của đại lượng ngẫu nhiên X có dạng:

$$f(x) = \frac{1}{1,934\sqrt{2\pi}} e^{-(x-5,239)^2/2\times 3,741.724}$$

- **Chú ý:** Từ dãy số liệu đã cho, tính được tham số mẫu: $\overline{x} \pm s = 5,133 \pm 1,737$.

Kiểm định quy luật chuẩn đại lượng ngẫu nhiên X thì X : N (MX, DX) với $MX \approx \overline{x}$ và $DX \approx s_X^2$. Tuy nhiên hàm mật độ xác suất của X lập theo bình phương bé nhất sẽ có tương quan hàm số chặt chẽ hơn.

2. Một địa phương có 908 người. Theo dõi nhiều ngày khi có dịch, thu được số liệu sau: t (thời gian, ngày), x (số người, đơn vị: người).

t_i	1	2	3	4	5	6
x_i	47	56	68	79	96	112

Lập hàm phát triển dịch x = f(t), từ đó cho biết số người bị dịch ngày thứ 10. Dịch không chữa được cho nên cách ly hoàn toàn với xung quanh.

Giải

- Hàm cần lập có dạng:

$$x = f(t) = \frac{a+b}{1+\frac{b}{a}e^{-\alpha(a+b)t}}.$$

- Bằng phương pháp bình phương bé nhất tuyến tính hoá thu được kết quả:

$$x = \frac{908}{1+22.14718508e^{-0,190.245.796t}}$$

$x(10) = 210,9 \approx 211$ người.

3. Cho các lô chuột nhắt trắng dùng Cocain Chlohydrat với liều tăng dần, Theo dõi số chuột chết, Finney thu được kết quả sau:

Liều (mg/kg)	0,015	0,02	0,025	0,030	0,035	0,04
Số chuột của lô	20	69	95	78	44	20
Số chuột chết	0	11	50	61	37	20
Tỷ lệ chết (%)	0	15,94	52,63	78,21	84,09	100

Hãy tính liều chết 50% (LD 50).

Giải

- Trevan nhận thấy:

– Tỷ lệ chết của chuột là hàm số của liều lượng. Đồ thị biểu diễn hàm số có dạng hình chữ s "nằm ngang".

– Liều chết ứng với 50%, gọi là LD 50, nằm ở chỗ dốc nhất của đường biểu diễn. Vì vậy khi thay đổi rất ít liều lượng cũng gây nên thay đổi lớn về tỷ lệ chết. Đó chính là vùng dễ có sai số lớn.

- Gọi y là tỷ lệ chết (đv: %) và x là liều lượng (đv: % mg/kg). Hàm số có dạng sau:

$$y = \frac{A}{1+B\,e^{-cx}}$$

Bằng phương pháp bình phương bé nhất tuyến tính hoá, qua 3 cặp giá trị ứng với các liều lượng 0,02 0,025 và 0,03. Lập được hàm số sau:

$$y = \frac{100}{1 + 1.710,88438e^{-2.940.645.524x}}$$

Từ hàm số trên suy ra:

LD50 = 2,531.677.36% mg/kg

≈ 0,0253 mg/kg

Bằng các phương pháp tính khác đối với số liệu trên, các tác giả cũng thu được các kết quả tương tự.

Bài 10

HỆ SỐ TƯƠNG QUAN TUYẾN TÍNH

MỤC TIÊU

1. Tính được hệ số tương quan tuyến tính của hàm bậc 1 và trình bày được ý nghĩa.

2. Tính được hệ số tương quan tuyến tính của một số hàm tuyến tính hoá.

Trong bài trước đã giới thiệu cách lập hàm số y = f(x) từ hai dãy số liệu.

Bài này giới thiệu một hệ số mà giá trị của nó cho biết lập hàm số y = ax + b có phù hợp với số liệu không. Đó là hệ số tương quan tuyến tính.

1. HIỆP PHƯƠNG SAI

1.1. Định nghĩa

Cho hai đại lượng X và Y

Hiệp phương sai của hai đại lượng X, Y ký hiệu C_{0v} (X, Y) là hằng số được xác định như sau:

$$C_{0v}(X, Y) = M\{(X - MX)(Y - MY)\} \tag{1}$$

Khi không biết MX và MY, hiệp phương sai được ước lượng bởi hiệp phương sai mẫu:

$$C_{0v}(X,Y) \approx \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) \tag{2}$$

Từ (2) dẫn đến công thức tính gần đúng của hiệp phương sai:

$$C_{0v}(X,Y) \approx \frac{1}{n}\sum_{i=1}^{n}x_i y_i - \frac{1}{n}\sum_{i=1}^{n}x_i \frac{1}{n}\sum_{i=1}^{n}y_i$$

$$\approx \overline{xy} - \overline{x}.\overline{y}.$$

Khi x_i, y_i nhận giá trị lớn hoặc có số thập phân hoặc cách đều ta có công thức tính sau:

$$C_{0v}(X,Y) \approx \Delta x.\Delta y\left(\overline{u.v} - \overline{u}.\overline{v}\right), \tag{3}$$

trong đó $u_i = \frac{x_i - x_0}{\Delta x}$, $v_i = \frac{y_i - y_0}{\Delta y}$ với x_0, y_0, $\Delta x \neq 0$ và $\Delta y \neq 0$ tuỳ chọn.

1.2. Tính chất

$C_{0v}(X, Y) = C_{0v}(Y, X)$.

$C_{0v}(X, X) = DX = \sigma^2$.

$C_{0v}(aX, bY) = ab\, C_{0v}(X, Y)$, a và b là các tham số thực.

$C_{0v}(X, Y) = 0$ khi X và Y độc lập với nhau.

2. HỆ SỐ TƯƠNG QUAN TUYẾN TÍNH

2.1. Định nghĩa

Cho hai đại lượng X và Y.

Hệ số tương quan tuyến tính của hai đại lượng X và Y là một số xác định, ký hiệu là $R_{x,y}$ gọi tắt là hệ số tương quan:

$$R_{xy} = \frac{C_{0v}(X,Y)}{\sqrt{DX}\sqrt{DY}}. \tag{4}$$

Gọi r_{xy} là hệ số tương quan mẫu

Khi không biết MX, MY hệ số tương quan được ước lượng bởi hệ số tương quan mẫu.

$R_{xy} \approx r_{xy}$

$$r_{xy} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2}} \tag{5}$$

2.2. Công thức tính hệ số tương quan mẫu

$$r_{xy} = \frac{n\sum_{i=1}^{n}x_i y_i - \sum_{i=1}^{n}x_i \sum_{i=1}^{n}y_i}{\sqrt{n\sum_{i=1}^{n}x_i^2 - \left(\sum_{i=1}^{n}x_i\right)^2} \times \sqrt{n\sum_{i=1}^{n}y_i^2 - \left(\sum_{i=1}^{n}y_i\right)^2}} \tag{6}$$

$$= \frac{\overline{xy} - \overline{x}.\overline{y}}{\sqrt{\overline{x^2} - (\overline{x})^2} \times \sqrt{\overline{y^2} - (\overline{y})^2}} = a\frac{\sqrt{\overline{x^2} - (\overline{x})^2}}{\sqrt{\overline{y^2} - (\overline{y})^2}} \tag{7}$$

$$= \frac{\overline{uv} - \bar{u}.\bar{v}}{\sqrt{\overline{u^2} - (\bar{u})^2} \times \sqrt{\overline{v^2} - (\bar{v})^2}}, \text{ trong đó u, v theo (3)} \qquad (8)$$

2.3. Tính chất

2.3.1. $R_{x,y}$ là hệ số không có đơn vị, thường viết đến phần nghìn

2.3.2. $R_{xy} = R_{yx}$ viết tắt là R.

2.3.3. Giả sử a, b là các số thực dương và $X' = aX$, $Y' = bY$.

$$R_{x'y'} = \frac{C_{0v}(X', Y')}{\sqrt{DX'\,DY'}} = \frac{C_{0v}(X, Y)}{\sqrt{DX\,DY}} = R_{xy}.$$

2.3.4. Hai đại lượng X, Y độc lập thì $R_{xy} = 0$.

2.3.5. Giả sử Y = aX + b

$$R_{xy} = \frac{C_{0v}(X, Y)}{\sqrt{DX\,DY}} = \frac{C_{0v}(X, aX+b)}{\sqrt{DXD\,(aX+b)}}$$

$$= \frac{aC_{0v}(X, X)}{\sqrt{DX}\,.\,\sqrt{a^2DX}} = \frac{a}{|a|} = \pm 1$$

Như vậy khi y là hàm bậc nhất của x thì hệ số tương quan tuyến tính bằng ± 1.

Từ (3.4) và (3.5) dẫn đến quy ước:

$0 \le |r| < 0,3$: x và y không tương quan tuyến tính.

$0,3 \le |r| \le 0,6$: x và y có tương quan tuyến tính.

$0,6 < |r| \le 1$: x và y có tương quan tuyến tính chặt chẽ.

Từ (7) suy ra r và a luôn cùng dấu:

$r < 0 \Leftrightarrow a < 0$ hàm bậc nhất nghịch biến

$r > 0 \Leftrightarrow a > 0$ hàm bậc nhất đồng biến.

Ví dụ

1. Gọi X, Y là giá trị đo được của một đại lượng tại hai điểm trên cơ thể bằng hai cách. Đo 12 người thu được kết quả sau:

X	32,6	34,5	39,0	39,1	39,1	39,3	39,7	42,3	45,4	53,3	59,4	71,9
Y	32,3	37,6	39,2	37,4	39,6	40,9	39,0	42,8	46,1	55,6	55,1	71,3

Hai dãy số liệu trên có tương quan tuyến tính không?

Giải

- Tính các kết quả trung gian

$\sum x$	$\sum x^2$	$\sum y$	$\sum y^2$	$\sum xy$
535,6	25.341,52	536,9	25.322,53	25.311,62.

▪ Tính r

$$r = \frac{12\times 25.311,62 - 535,6\times 536,9}{\sqrt{12\times 25.341,52 - 535,6^2}\ \sqrt{12\times 25.322,53 - 536,9^2}} = 0,986.343.023$$

$$\approx 0,986.$$

Hai dãy số liệu trên tương quan tuyến tính đồng biến rất chặt chẽ.

2. Theo dõi số dân (X, đv: triệu) và chỉ tiêu phát triển dân số (Y, đv: $^0/_{00}$) của cả nước thu được số liệu sau:

X	64,412	66,233	67,744	69,405	71,026	72,510	73,959	75,355	76,710
Y	21,0	21,9	22,9	23,00	21,8	21,6	18,6	18,8	18

Hai dãy số liệu trên có tương quan tuyến tính không?

Giải

▪ Tính các kết quả trung gian

$\sum x$	$\sum x^2$	$\sum y$	$\sum y^2$	$\sum xy$
637,354	45.277,146.18	187,6	3.939,22	13.238,4815.

▪ Tính r

$$r = \frac{9\times 13.238,4815 - 637,354\times 187,6}{\sqrt{9\times 45.277,146.18 - 637,354^2}\ \sqrt{9\times 3.939,22 - 187,6^2}} = -\ 0,733.019.569$$

$$\approx -\ 0,733.$$

Hai dãy số liệu trên tương quan tuyến tính nghịch biến rất chặt chẽ.

Chú ý

▪ Dùng các máy tính bấm tay đủ mạnh có thể gọi được kết quả hệ số tương quan tuyến tính trên máy.

▪ Trong ví dụ lập hàm số $Y = ax^2 + bx + c$, tuyến tính hoá cũng tính được hệ số tương quan tuyến tính.

CÂU HỎI TỰ LƯỢNG GIÁ

(Bài 9 & 10)

Hãy chọn một kết quả đúng.

1. Theo dõi số dân (x, đv: 1 người) và chỉ tiêu phát triển dân số ($y = s - c$, đv: $^0/_{00}$) thu được 2 dãy số liệu sau:

x_i	73000	74000	75000	77000
y_i	21,6	18,6	18,8	18,0

Lập phương trình $y = ax + b$ trong đó a, b tính theo công thức.

Kết quả:

A. – 0,000.74857x + 75,20785 B. – 0,000.74857x + 75,2057

C. – 0,000.7486x + 75,2057 D. – 0,0007486x + 75,20785

E. Biểu thức khác.

2. Theo dõi 2 đại lượng x và y thu được kết quả sau:

x_i	54.927	56.713	57.442	58.669	59.872
y_i	30,02	29,8	29,3	28,4	28,44

Lập phương trình bậc hai $y = ax^2 + bx + c$.

Kết quả:

A. $-1{,}503.918.96 \times 10^{-8} x^2 + 1{,}352.295.553 \times 10^{-3}x + 1{,}210.425.323$

B. $-1.162{,}951.089x^2 + 65.475{,}23069x - 862.267{,}3944$.

C. $-1.162{,}951.08x^2 + 1{,}352.295.553.10^{-3}x + 1{,}210.425.323$.

D. $-1{,}503.918.96.10^{-8}x^2 + 65.475{,}23069x - 862267{,}3944$.

E. Phương trình khác.

3. Theo dõi số dân (x, đv: 1000 người) và chỉ tiêu phát triển dân số ($y = s - c$, ‰) thu được số liệu sau:

x_i	54900	56700	57600	58500	59400
y_i	23,04	22,7	22,36	21,68	21,34

Từ phương trình đã lập $x = ay + b$ hãy cho biết x_c.

Kết quả:

A. 113,989,567 B. 110.091,792

C. 131.989,567 D. 101.091,792 E. số khác.

4. Gọi x là lứa tuổi và y là nhịp tim trung bình, nghiên cứu thu được kết quả sau:

x_i	9	10	11	12	13	14	15
y_i	72,8	72,5	73,6	69,8	69,2	68,6	70,2

Từ hai dãy số liệu trên đã lập được hai phương trình:

$y = -0{,}714x + 78{,}529$ (1)

$y = -0{,}714x + 80{,}029$ (2)

Hãy tính một giá trị để biết phương trình nào tốt hơn.

Kết quả:

A. 15,978 B. –0,783 C. 9,031 D. 10,809 E. số khác.

Bài tập

RÚT MẪU

1. Cho 10 chữ số: 0, 1, 2, ..., 9.

a) Có bao nhiêu số chẵn có 5 chữ số khác nhau lập từ 10 chữ số ?

b) Có bao nhiêu số có 5 chữ số khác nhau, nhỏ hơn 30.000, không bắt đầu từ 123 lập từ 10 chữ số.

2. Bệnh viện N 10 tầng có thang máy, 5 bệnh nhân cùng vào thang máy tầng 1 để lên các tầng trên. Hỏi có bao nhiêu cách nếu:

a) 5 bệnh nhân tuỳ ý ra các tầng?

b) ít nhất 2 bệnh nhân cùng ra 1 tầng, số còn lại mỗi người một tầng?

3. Một lớp gồm 40 học viên trong đó có 20 nam, 20 nữ. Chia lớp làm 10 nhóm bằng nhau, hỏi có bao nhiêu cách chia nếu:

a) Số nam, nữ trong nhóm tuỳ ý ?

b) Mỗi nhóm nhiều nhất 2 nam ?

4. Đơn vị A có 15 nam và 5 nữ. Lập ban chỉ huy có 4 người. Hỏi có bao nhiêu cách lập nếu:

a) Số nam, nữ trong ban tuỳ ý ?

b) Ban chỉ huy phải có nữ ?

5. Có n bệnh nhân ngồi thành hàng chờ khám bệnh. Có bao nhiêu trường hợp để 2 người chọn trước luôn luôn cách nhau r người ($r < n$) ?

6. Trong phòng nghiên cứu có n cặp kháng nguyên kháng thể khác nhau. Đặt các kháng nguyên lên giá trước mà quên ghi nhãn. Phải làm bao nhiêu lần để tìm được đúng n cặp?

7. Một bác sĩ có 15 bệnh án. Hỏi có bao nhiêu cách lấy bệnh án nghiên cứu nếu:

a) Lấy tuỳ ý 10 bệnh án ?

b) Lấy 10 bệnh án có chọn lọc ?

8. Một Khoa có 20 bác sĩ. Lập quy hoạch bồi dưỡng liên tục. Hỏi có bao nhiêu cách sắp xếp nếu lấy một người đi thi NCS và 3 người đi thi CKI, trong đó có 1 chỉ tiêu dự bị ?

9. Có 3 thuốc cùng loại điều trị cho 4 bệnh nhân. Hỏi có bao nhiêu cách điều trị nếu:

a) Mỗi bệnh nhân dùng không quá 2 thuốc ?

b) Số thuốc dùng tuỳ ý cho mỗi bệnh nhân ?

10. Ba kỹ thuật viên thay nhau làm thủ thuật. Sau một thời gian có 4 thủ thuật không đạt.

a) Có bao nhiêu trường hợp xảy ra khi 1 trong 3 người làm hỏng 3 thủ thuật?

b) Một trong 3 người làm hỏng 3 thủ thuật, người đó có vụng không ?

11. Trên một bảng đồng có 3 cột kim loại. Trên 1 cột đã xếp 64 vòng vàng thành tháp Phạn: vòng to ở dưới, vòng nhỏ ở trên. Cần chuyển 64 vòng sang 1 cột khác. Chỉ có 1 người làm việc, nếu mệt thay người khác, mỗi lần chỉ chuyển 1 vòng, phải đặt vòng nhỏ lên trên vòng lớn.

a) Cần tất cả bao nhiêu lần để chuyển cho xong ?

b) Cần bao nhiêu thời gian nếu mỗi vòng chuyển trung bình mất 1 giây ?

XÁC SUẤT

1. Một phòng điều trị cho 3 bệnh nhân nặng A, B, C. Trong 1 giờ, xác suất để bệnh nhân A, B, C cấp cứu tương ứng bằng 0,6; 0,7; 0,8. Tìm xác suất sao cho trong 1 giờ:

a) Cả 3 bệnh nhân cấp cứu.

b) Có ít nhất 1 bệnh nhân cấp cứu.

2. Tỷ lệ mổ của bệnh K bằng 15%. Trong số những người mổ K có 10% mổ sớm. Biết tỷ lệ mổ sớm sống trên 5 năm của những người bệnh K bằng 0,00375.

a) Tìm tỷ lệ mổ sớm của bệnh K.

b) Tìm tỷ lệ sống trên 5 năm của những người mổ K.

3. Ba bệnh nhân nặng A, B, C cùng điều trị tại bệnh viện. Trong một giờ, xác suất cấp cứu tương ứng với 3 bệnh nhân bằng 0,8; 0,7 và 0,6. Trong một giờ có 2 bệnh nhân cấp cứu, tìm xác suất gặp 2 bệnh nhân cấp cứu là A và B.

4. Trong 20 kháng sinh (KS) có 1 KS chữa được xoắn trùng. Lấy ngẫu nhiên 1KS điều trị xoắn trùng, nếu không khỏi lấy ngẫu nhiên một KS khác. Tìm xác suất điều trị xoắn trùng khỏi ở lần thứ 5.

5. Xác suất sinh con trai bằng 0,514. Ai có khả năng thực hiện mong muốn của mình hơn?

a) Phụ nữ A mong muốn sinh bằng được con gái.

b) Phụ nữ B mong muốn sinh bằng được con trai.

6. Tại một khoa, xếp 12 bệnh nhân vào 3 phòng bệnh (các phòng có thể xếp tối đa 12 bệnh nhân).

a) Tìm xác suất sao cho có 3, 4, 5 bệnh nhân vào 3 phòng tương ứng.

b) Tìm xác suất sao cho phòng 1 có 3 bệnh nhân, còn lại xếp tuỳ ý.

7. Trong một khay đựng n cặp bơm tiêm khác nhau. Lấy ngẫu nhiên ra 2r chiếc: $2r \le n$.

a) Tìm xác suất sao cho không lấy được cặp bơm tiêm nào cùng đôi.

b) Tìm xác suất sao cho được đúng một cặp bơm tiêm cùng đôi.

8. Xác suất bị bạch tạng của đàn ông bằng 0,0006; của đàn bà bằng 0,000.036. Trong đám đông số đàn ông bằng 0,5 số đàn bà. Tìm xác suất gặp một người đàn ông trong đám đông bị bạch tạng

9. Gọi E_1 là hiện tượng sinh đôi thật. Hai trẻ luôn luôn cùng giới.

Gọi E_2 là hiện tượng sinh đôi giả. Hai trẻ cùng giới hoặc khác giới; xác suất cùng giới bằng 0,5. Xác suất sinh đôi thật bằng p.

a) Tìm xác suất sinh đôi thật của nhóm cùng giới.

b) Nếu sinh đôi khác giới thì xác suất sinh đôi giả bằng bao nhiêu?

10. Xác suất dương tính của Xquang bằng 0,2. Giá trị của Xquang dương tính bằng 0,2. Biết tỷ lệ bị bệnh trong nhóm Xquang âm tính bằng 0,0125. Dùng Xquang chẩn đoán bệnh.

Tìm độ nhạy, độ đặc hiệu của Xquang.

11. Xét nghiệm HS có xác suất sai bằng 0,197. Tỷ lệ bị bệnh tại cộng đồng bằng 0,02. Biết độ đặc hiệu bằng 0,8. Dùng xét nghiệm chẩn đoán bệnh. Tìm giá trị của xét nghiệm dương tính.

12. Ba người cùng đến khám bệnh. Người thứ i nghi bệnh B_i, i = 1,2,3. Xác suất bị bệnh B_1 bằng 0,01; B_2 bằng 0,02. Biết xác suất sao cho có người bị bệnh bằng 0,058906. Ba người được khám thấy 1 người bị bệnh, tìm xác suất sao cho 2 người không bị bệnh là người thứ 1 và thứ 3.

13. Tại một khoa nội, tỷ lệ 3 nhóm bệnh tim mạch, huyết học, tiêu hoá là 1: 1: 2. Xác suất gặp bệnh nhân nặng của nhóm tim mạch bằng 0,4 và của huyết học bằng 0,5. Xác suất gặp bệnh nhân nặng của 3 nhóm bằng 0,375.

Khám tất cả bệnh nhân nặng, tìm tỷ lệ gặp bệnh nhân nhóm tiêu hoá.

14. Điều trị riêng rẽ 2 kháng sinh (KS) cho bệnh nhân, xác suất phản ứng của KSI bằng 0,002, KSII bằng 0,001. Biết xác suất phản ứng của 2KS khi điều trị riêng rẽ bằng 0,0014.

a) Một người dùng KS bị phản ứng, tìm xác suất sao cho người đó dùng KSII.

b) Tìm xác suất sao cho 2 người dùng KS thì cả 2 không bị phản ứng.

15. Xác suất mắc bệnh B tại phòng khám bằng 0,4. Khi sử dụng phương pháp chẩn đoán mới, với khẳng định là có bệnh thì đúng 3/5 trường hợp; với khẳng định là không bệnh thì đúng 4/5 trường hợp.

a) Tìm xác suất chẩn đoán có bệnh của phương pháp trên.

b) Tìm xác suất chẩn đoán sai.

c) Khi xác suất mắc bệnh B thay đổi, bài toán đúng với xác suất mắc bệnh B là bao nhiêu?

16. Tại một bệnh viện, tỷ lệ mắc bệnh B bằng 0,1. Để chẩn đoán xác định, người ta làm phản ứng MD, nếu khẳng định có bệnh thì đúng 50%, nếu người không bị bệnh thì sai 10%.

a) Tìm xác suất phản ứng dương tính của nhóm bị bệnh.

b) Tìm giá trị của chẩn đoán MD.

17. Khi chẩn đoán bệnh B, một phản ứng có xác suất dương tính bằng 0,75. Nếu phản ứng dương tính thì đúng 9 /10 trường hợp. Giá trị của phản ứng âm tính bằng 0,5. Một người được chẩn đoán đúng, tìm xác suất sao cho đó là người bị bệnh; đó là người có phản ứng âm tính.

18. Khám bệnh ngoài da cho các cháu tại một nhà trẻ, các bác sỹ thấy 70% trẻ mắc bệnh A, 50% trẻ mắc bệnh B.

Dùng thuốc T_1 chữa bệnh, xác suất khỏi khi chữa bệnh A bằng 0,8; bệnh B bằng 0,5; cả 2 bệnh bằng 0,35.

Dùng thuốc T_2 chữa bệnh, xác suất khỏi khi chữa bệnh A bằng 0,6; bệnh B bằng 0,8; cả 2 bệnh bằng 0,3.

Biết rằng giá thuốc, khối lượng thuốc 2 loại như nhau. Nên dùng thuốc nào để chữa bệnh?

19. Tỷ lệ dùng thuốc A bằng 0,2, thuốc B bằng 0,4 và thuốc C bằng 0,4. Xác suất khỏi của thuốc A bằng 0,9, thuốc B bằng 0,85. Biết xác suất dùng thuốc C trong số khỏi bằng 0,35.

a) Tìm xác suất khỏi của 3 thuốc khi dùng riêng rẽ từng thuốc.

b) Tìm xác suất khỏi khi dùng phối hợp 3 thuốc.

20. Dùng 3 thuốc A, B, C điều trị một bệnh thấy xác suất kháng thuốc A bằng 0,15, thuốc B bằng 0,3. Tỷ lệ dùng thuốc A bằng 0,4. Biết xác suất dùng thuốc A, B trong số kháng thuốc tương ứng bằng 0,3 và 0,6. Tìm xác suất kháng thuốc của C.

Bài tự làm

21. Một phản ứng có xác suất chẩn đoán đúng bằng 0,74. Dùng phản ứng chẩn đoán bệnh, giá trị của phản ứng dương tính bằng 0,7 và độ đặc hiệu bằng 0,64.

Hãy cho biết độ nhạy của phản ứng.

22. Dùng một xét nghiệm chẩn đoán bệnh, xét nghiệm có độ nhạy bằng 0,9 và độ đặc hiệu bằng 0,7. Biết xác suất dương tính của nhóm đúng bằng 0,5625. Tìm giá trị của xét nghiệm dương tính?

23. Kiểm tra những người chẩn đoán bị bệnh ở bệnh viện I, II tuyến dưới thấy tương ứng 90% và 96% bị bệnh. Xác suất khỏi sau kiểm tra của hai bệnh viện tương ứng bằng 0,95 và 0,9375. Tìm xác suất khỏi của hai bệnh viện trước kiểm tra, biết rằng số kiểm tra của bệnh viện I bằng 5/3 bệnh viện II.

24. Người có nhóm máu AB có thể nhận bất kỳ nhóm máu nào. Người có nhóm máu còn lại có thể nhận máu của người cùng nhóm máu với mình hoặc của người có nhóm máu O. Tỷ lệ các nhóm máu O, A, B, AB tương ứng bằng 0,24 0,29 0,32 0,15 (người ÊĐÊ). Chọn ngẫu nhiên một người nhận máu và một người cho máu dân tộc trên, tìm xác suất để sự truyền máu được thực hiện.

NHỊ THỨC

1. Xác suất sinh con trai bằng 0,514. Hiện tượng sinh được hai con trai trong 4 con có xác suất lớn hay bé hơn hiện tượng sinh được 2 con gái trong 4 con?

2. Tỷ lệ sinh viên bị cận thị tại một trường đại học bằng 1%. Cần lấy một mẫu n bằng bao nhiêu(có hoàn lại), sao cho với xác suất không nhỏ hơn 0,95, trong mẫu đó có ít nhất một sinh viên bị cận thị?

3. Tỷ lệ bị lao trong dân cư bằng 0,1%. Khám lao cho n người. Gọi A là hiện tượng có 1 người bị lao. Gọi B là hiện tượng có 1 người không bị lao. Hiện tượng A hay hiện tượng B có khả năng xuất hiện lớn hơn ? Nêu ý nghĩa.

4. Xác suất sinh bằng được con gái ở lần thứ 3 bằng 0,127449. Xác suất sinh được một con trai trong 3 lần sinh bằng 0,367353.

Tìm xác suất sinh được con trai trong 1 lần sinh.

5. Một bác sỹ chữa bệnh B có xác suất khỏi bằng 0,8. Có người cho rằng cứ 5 người bệnh B đến chữa, chắc chắn có 4 người khỏi. Người khác cho rằng: cứ 10 người bệnh B đến chữa, chắc chắn có 8 người khỏi.

Ai nói đúng ? Tính 2 xác suất trên.

6. Dùng thuốc A, B điều trị một bệnh thấy xác suất khỏi của A bằng 0,8, xác suất khỏi của B bằng 0,6. Điều trị phối hợp hai thuốc cho n người bệnh, xác suất có 10 người khỏi là lớn nhất. Tìm xác suất trên.

7. Một bác sỹ chữa bệnh B có xác suất khỏi bằng 0,8. Chữa cho đến người thứ bao nhiêu để xác suất không nhỏ hơn 0,9 có thể tin rằng số người không khỏi từ dưới 75 người?

8. Dùng thuốc mới chữa thử bệnh B có tỷ lệ khỏi bằng p. Trước khi đưa ra sử dụng chính thức, người ta điều trị thử cho 100 người bệnh.

Thuốc được chấp nhận đưa ra sử dụng với xác suất 1; 0,8 và 0 với số người khỏi tương ứng là trên 80 người, từ 60 đến 80 người và dưới 60 người.

Tìm xác suất thuốc được chấp nhận sử dụng với:

a) p = 0,8.

b) p = 0,6.

9. Tỷ lệ bị bệnh sau dùng vác xin bằng 0,0001. Dùng vác xin cho 100.000 trẻ, tìm xác suất sao cho:

a) Có r trẻ bị bệnh, biết xác suất có r trẻ bị bệnh là lớn nhất.

b) Có nhiều nhất 10 trẻ bị bệnh.

10. Xét nghiệm nhóm máu AB cho 100 người. Tỷ lệ nhóm máu AB bằng 0,05. Tìm xác suất sao cho có từ 3 đến 7 người có nhóm máu AB.

THAM SỐ MẪU, SO SÁNH PHƯƠNG SAI SO SÁNH TRUNG BÌNH

1. Điều tra 1600 gia đình có 4 con thu được kết quả sau:

x_i(số con trai)	0	1	2	3	4
m_i(số gia đình)	111	367	576	428	118

Tính các tham số $\overline{x} \pm s$ và C_v của số liệu trên.

2. Đo áp lực động mạch phổi thời tâm thu bệnh nhân hẹp 2 lá thu được kết quả sau:

Áp lực (mmHg)	20,5	35,5	50,5	65,5	80,5	95,5	110,5	125,5	140,5	155,5
Số người	6	20	33	24	28	12	17	8	4	1

Tính các tham số $\overline{x} \pm s$ và C_v của số liệu trên.

3. Theo dõi số chuột chết khi cho các nhóm chuột sử dụng các liều thuốc có độc (X: mg/kg) thu được kết quả sau:

x_i(mg/kg)	0,015	0,02	0,025	0,03	0,035	0,04
Số chuột của nhóm	10	10	10	10	10	10
Số chết của nhóm	0	2	5	8	9	10

Tính $\overline{x} \pm s$ của số liệu trên.

4. Theo dõi dấu hiệu viêm khớp khi điều trị (ĐT) bệnh nhân viêm đa khớp thu được số liệu sau:

Trước ĐT	3	2	6	4	7	12	5	4	8	15	18	15
Sau 1 tháng ĐT	3	2	4	4	6	10	5	4	8	14	18	13
Sau 2 tháng ĐT	2	0	4	2	4	7	3	2	4	10	15	7

Trước ĐT (tiếp)	20	16	8	15	17	16	18	15	9	13
Sau 1 tháng ĐT(tiếp)	18	15	9	14	15	14	20	15	8	12
Sau 2 tháng ĐT(tiếp)	15	13	7	8	10	12	17	13	7	10

a) Tính các tham số $\overline{x} \pm s$ và C_v của 3 dãy số liệu: trước điều trị, sau 1 tháng ĐT, sau 2 tháng ĐT.

b) Tính các tham số $\overline{z} \pm s$ và C_v của chênh lệch trước ĐT và sau 1 tháng ĐT, của trước ĐT và sau 2 tháng ĐT, của sau 1 tháng ĐT và sau 2 tháng ĐT.

c) Tính các phương sai $\tilde{S}^2$ và S^2 của 3 dãy số liệu.

d) Hãy so sánh trung bình của 2 trong 3 dãy.

e) Hãy so sánh từng cặp của các chênh lệch.

f) Hãy so sánh 3 trung bình của 3 dãy.

5. Điều trị sốt rét bằng 4 cách. Theo dõi thời gian hết ký sinh trùng sốt rét trong máu (giờ) của từng bệnh nhân thu được số liệu sau:

Cách 1	18	37	46	46	46	50,5	61,5	78	84,5	90
Cách 2	38	41	41,1	42	43,1	44,1	45,2	50	50	52
Cách 3	36	48	50	52	58	60	60	68	74	74
Cách 4	36	38	40	42	48	60	62	70	72	72

a) Tính các tham số $\overline{x} \pm s$ và C_v của từng cách điều trị.

b) Tính các phương sai $\tilde{S}^2$ và S^2 của 4 dãy số liệu.

c) Hãy so sánh trung bình của 2 trong 4 dãy số liệu.

d) Hãy so sánh 4 giá trị trung bình của 4 dãy số liệu.

6. Đo chiều cao (CC: cm); Vòng ngực (VN: cm); Cân nặng(CN: kg); Vòng bụng (VB: cm); Sải tay (ST: cm) của 2 nhóm trẻ nữ 9 tuổi được 10 dãy số liệu sau:

Nhóm I					Nhóm II				
CC	**VN**	**CN**	**VB**	**ST**	**CC**	**VN**	**CN**	**VB**	**ST**
115	54	16	49	115	115	54	17	50	110
112	51	19	51	106	118	52	18	46	118
103	49	14	47	98	126	56	21	50	121
117	56	20,5	51	115	115	59	20	49	112
115	53	17	50	110	118	55	20	51	112
112	54	17	50	110	121	56	20,5	50	120
117	59	20	51	110	122	56	22,5	49	120
130	67	25	51	125	129	57	24	51	128
114	58	18,5	48	110	120	53	18	49	118
115	55	19	50	110	110	50	16	48	102
126	57	22,5	54	122	120	55	18,5	50	120
117	51	17	49	112	125	56	21	50	120
113	60	20	48	113	115	54	17,5	49	113
112	59	20	51	110	132	55	26	50	132
110	53	15	49	110	130	57	23	51	128
125	61	23	51	120	122	55	20	50	122
125	60	21,5	50	120	122	53	18	49	120
130	60	26	51	126	112	51	16	49	108
120	55	20	49	115	109	54	15	49	104
120	55	20	52	115	121	55	20	51	121
121	57	21	49	121	117	54	19	48	111
116	55	19	49	115	134	62	25	49	130

a) Tính các tham số $\overline{x} \pm s$ và C_v của từng dãy số liệu.

b) Hãy so sánh từng cặp giữa Chiều cao – Sải tay, giữa Vòng bụng – Vòng ngực của nhóm I. Làm tương tự với nhóm II.

c) Hãy so sánh trung bình Chiều cao I và Chiều cao II,..., Sải tay I và Sải tay II.

d) Gọi x là chiều cao nữ 9 tuổi, X : N(121,14; 5,08^2). Chiều cao nhóm I; nhóm II tương ứng có thừa nhận MX = 121,14 không ?

e) Gọi y là cân nặng nữ 9 tuổi. Giả sử MY = 19,5 là sai. Hãy tính $\beta(20,55)$ với n = 44, biết DY = 2,44^2. Kiểm định 1phía với α = 0,05.

KIỂM ĐỊNH χ^2

1. Điều tra 1600 bà mẹ sinh 4 con, người ta thấy:

111 bà mẹ không có con trai — 428 bà mẹ có 1 con gái.

367 bà mẹ có 1 con trai — 118 bà mẹ không có con gái

576 bà mẹ có 2 con gái.

Tỷ lệ sinh con trai của các bà mẹ 4 con có bằng 0,5 không?

2. Điều trị kháng sinh (KS) I, KSII và KSIII, mỗi loại cho 80 người, 120 người và 200 người. Xác suất khỏi của mỗi loại KS tương ứng bằng 0,9; 0,85 và 0,8.

Tỷ lệ khỏi của 3 loại KS trên có như nhau không?

3. Dùng Xquang và siêu âm, mỗi loại kiểm tra 100 người bị bệnh. Xquang và siêu âm xác định đúng tương ứng bằng 0,8 và 0,9. Độ nhạy của 2 phương pháp có như nhau không?

4. Xác định bệnh cho 1000 người ở cộng đồng bằng một xét nghiệm, thấy 2,5% dương tính. Trong số dương tính có 80% bị bệnh; trong số âm tính có 8% bị bệnh.

Giá trị xét nghiệm dương tính có bằng giá trị xét nghiệm âm tính không ?

5. Điều tra số trẻ chết trước một tuổi tại xã A bị rải chất diệt cỏ và xã B không bị rải chất diệt cỏ, thu được kết quả sau:

Năm 71 – 75 xã A: số trẻ chết 52, số trẻ sống 1260

xã B: số trẻ chết 19, số trẻ sống 876.

Năm 81 – 86 xã A: số trẻ chết 61, số trẻ sống 1696

xã B: số trẻ chết 20, số trẻ sống 753.

Chất diệt cỏ có ảnh hưởng tới tỷ lệ chết của trẻ dưới 1 tuổi không?

Nếu có, ảnh hưởng như thế nào?

6. Khám lao cho 120.000; 100.000 và 90.000 người của 3 phường A, B, C tương ứng, người ta thấy tỷ lệ bị lao tương ứng mỗi phường bằng 0,001; 0,0015 và 0,0012.

a) Tỷ lệ bị lao của 3 phường có như nhau không?

b) Số người bị lao của 3 phường có như nhau không?

7. Điều trị một bệnh bằng 3 phương pháp A, B, C mỗi phương pháp cho 40 người, thu được kết quả sau:

Kết quả điều trị	Khỏi	Đỡ	Thất bại
Phương pháp A	14	18	8
Phương pháp B	22	16	2
Phương pháp C	32	8	0

Hiệu quả của ba phương pháp điều trị có như nhau không?

8. Tổng kết số liệu 10 năm của bệnh viện K người ta thấy trong số 2000 bệnh nhân đến khám và điều trị có 50% phải mổ. Trong số những người đã mổ có 12% số người mổ sớm và chạy tia, 8% số người mổ sớm và điều trị bằng hoá chất, số còn lại là mổ muộn. Trong số những người sống trên 5 năm sau mổ, có 50% là mổ sớm chạy tia, 40% mổ sớm và dùng hoá chất, số còn lại là mổ muộn. Có 90% số người được mổ sống dưới 5 năm. Hãy đánh giá hiệu quả các phương pháp điều trị và nêu ý nghĩa?

9. Xét nghiệm nhóm máu cho 100 người. Tỷ lệ gặp nhóm máu O bằng 0,48. Mỗi lần xét nghiệm cho 4 người. Gọi X là số người có nhóm máu O trong nhóm xét nghiệm. Số liệu thu được như sau:

x_i	0	1	2	3	4
m_i	1	7	10	6	1

a) X có quy luật siêu bội có đúng không ?

b) X có quy luật nhị thức với n =4, p = 0,48 có đúng không ?

10. Điều tra 53.680 gia đình 8 con. Gọi X là số con trai, thu được số liệu sau:

x_i (số con trai)	0	1	2	3	4	5	6	7	8
m_i (số gia đình)	215	1.485	5.331	10.649	14.959	11.929	6.678	2.092	342

X có quy luật nhị thức với n = 8 và p $\approx \omega$ không ?

11. Đếm số hồng cầu X rơi vào mỗi ô của máy đếm hồng cầu, thu được số liệu sau:

x_i (số HC)	0	1	2	3	4	5	6	7	8	9	10	11	12	13	>13
m_i (số ô)	2	21	48	57	92	67	47	34	15	8	4	2	2	1	0

X có quy luật Poisson với $\lambda = \overline{x}$ không ?

12. Định lượng Protein dịch não người tuỷ bình thường (x, đơn vị: mg%) thu được số liệu sau:

11,4	17,3	19,2	11,5	17,4	19,3	14,3	17,5	19,3	16,	17,6	19,5
16	17,7	19,6	16,1	18,2	19,7	16,3	18,5	20,	16,3	18,8	20,1
16,4	19	20,8	16,7	19,1	21,	17,1	19,1	21,1	17,2	19,2	22

a) Hãy kiểm định giả thiết: $X:N(\mu,\sigma^2)$ với $\mu \approx \overline{x}$, $\sigma^2 \approx s^2$ bằng kiểm định χ^2

b) Hãy kiểm định giả thiết: $X:N(\mu,\sigma^2)$ với $\mu \approx \overline{x}$, $\sigma^2 \approx s^2$ bằng kiểm định Kolmogorov.

13. Gọi X là áp lực trung bình của động mạch phổi bệnh nhân hẹp van 2 lá đơn thuần (mmHg). Nghiên cứu thu được số liệu sau:

x_i	13	23	33	43	53	63	73	83	93	103
m_i	5	20	27	24	25	23	15	10	4	2

trong đó x_i là giá trị giữa của lớp thứ i.

a) $X : N(\mu,\sigma^2)$ với $\mu \approx \overline{x}$, $\sigma^2 \approx s^2$ có đúng không ?

b) Giả sử $X:N(\mu,\sigma^2)$. Trong số bao nhiêu người bệnh có một người có áp lực trung bình của động mạch phổi từ trên 110 mmHg.

KIỂM ĐỊNH XÁC SUẤT

1. Nêu sự khác nhau giữa kiểm định "một phía" và "hai phía" trong bài toán kiểm định giá trị của xác suất.

2. Nêu sự giống nhau giữa kiểm định "một phía" và "hai phía" trong bài toán kiểm định giá trị của xác suất.

3. Với mức ý nghĩa $\alpha = 0,05$, kiểm định 2 phía bác bỏ giả thiết H_0 thì kiểm định 1 phía kết luận thế nào ? Giải thích?

Với mức ý nghĩa $\alpha = 0,05$, kiểm định 1 phía bác bỏ giả thiết H_0 thì kiểm định 2 phía kết luận thế nào ? Giải thích?

4. Sử dụng hai thuốc A, B cùng loại, có ý kiến cho là thuốc A tốt hơn thuốc B, có ý kiến ngược lại.

a) Hỏi 100 người, chỉ có 40 người cho là thuốc A tốt hơn B. Có thật sự thuốc A tốt hơn B không? Giả thiết là 2 thuốc như nhau, lấy $\alpha = 0,05$.

b) Hỏi 105 người, chỉ có 35 người cho là thuốc B tốt hơn A, 5 người cho là như nhau. Có thật sự thuốc A không tốt hơn B không? Giả thiết là 2 loại như nhau, lấy $\alpha = 0,05$.

5. Xét nghiệm nhóm máu cho 100 người, mỗi lần một nhóm 4 người, thu được số liệu sau (x: số người có nhóm máu O trong nhóm xét nghiệm):

x	0	1	2	3	4
Số nhóm	1	7	10	6	1

Tỷ lệ nhóm máu O bằng 0,48 có đúng không?

6. Điều tra 100.000 người ở tỉnh H thấy 32 người bị lao.

a) Tỷ lệ bị lao bằng 0,0005 có đúng không?

b) Nếu tỷ lệ bị lao đúng bằng 0,001, tính sai lầm loại 2 với $\alpha = 0,05$.

7. Tại một địa phương, điều tra 1.000 trẻ thấy 376 suy dinh dưỡng.

a) Tỷ lệ suy dinh dưỡng bằng 0,4 có đúng không?

Tính sai lầm loại 2, nếu tỷ lệ suy dinh dưỡng đúng bằng 0,37 với $\alpha = 0,05$.

b) Tỷ lệ suy dinh dưỡng bằng 0,35 có đúng không?

Tính sai lầm loại 2, nếu tỷ lệ suy dinh dưỡng đúng bằng 0,37 với $\alpha = 0,05$.

8. Khi sử dụng thuốc A vì tai biến có ý kiến cho là phải cấm dùng, có ý kiến ngược lại. Cần hỏi ý kiến bao nhiêu người? Trong đó bao nhiêu người đồng ý cấm thì ra lệnh cấm? Nếu quan niệm như sau:

a) $H_0 : p_0 = 0,5$	$H_1 : p_0 < 0,5$	$\beta(0,4) \leq 0,1$	$\alpha = 0,05$
b) $H_0 : p_0 = 0,5$	$H_1 : p_0 > 0,5$	$\beta(0,6) \leq 0,2$	$\alpha = 0,01$
c) $H_0 : p_0 = 0,5$	$H_1 : p_0 \neq 0,5$	$\beta(0,51) \leq 0,2$	$\alpha = 0,05$

ĐỘ KHÔNG XÁC ĐỊNH

1. Một phòng điều trị 3 bệnh nhân nặng A, B, C. Trong 1 giờ xác suất cấp cứu tương ứng của A, B, C bằng 0,6; 0,7; 0,8.

Trong 1 giờ đoán 3 người ai cấp cứu ai không khó hay dễ ?

2. Tỷ lệ mổ K bằng 0,15. Trong số những người mổ K có 10% mổ sớm. Biết tỷ lệ mổ sớm sống trên 5 năm bằng 0,00375.

Tìm độ không xác định lớn nhất trong các phép thử.

3. Trong 10 kháng sinh (KS) có một KS chữa được xoắn trùng. Lấy ngẫu nhiên một KS điều trị xoắn trùng, nếu không khỏi lấy ngẫu nhiên một KS khác.

Đoán xem bệnh nhân bị xoắn trùng điều trị khỏi ở lần nào khó hay dễ ?

4. Xác suất sinh con trai bằng 0,514

a) Đoán hai phụ nữ cùng sinh, mỗi người một con là trai hay gái, dễ hay khó ?

b) Đoán hai phụ nữ cùng sinh, mỗi người hai con là trai hay gái, dễ hay khó ?

5. Điều trị một bệnh có xác suất khỏi bằng 0,7

Điều trị cho 4 người, đoán mấy người khỏi mấy người không dễ hay khó ?

6. Tại một khoa nội thấy 15% bị bệnh A, 20% bị bệnh B, 30% bị bệnhC, số còn lại bị các bệnh khác.

Đoán một bệnh nhân vào khoa nội thuộc nhóm nào khó hay dễ ?

7. Tại một bệnh viện tổng kết thấy 30% người nghiện thuốc lá, trong đó 5% bị K. Những người không nghiện thuốc lá có 1% K.

Đoán một người nghiện thuốc lá và K không khó hay dễ?

8. Tỷ lệ ba nhóm bệnh A:B:C tại khoa nội bằng 2:1:2. Xác suất gặp bệnh nhân nặng tương ứng của mỗi nhóm bằng 0,35 0,5 và 0,4.

a) Đoán bệnh nhân trong khoa nặng hay không khó hay dễ ?

b) Gọi α : nặng hay không nặng; β: A, B, C. Tính I (α,β).

9. Một xét nghiệm có xác suất đúng bằng 0,95 và tỷ lệ bị bệnh tại bệnh viện bằng 0,2. Biết độ nhạy của xét nghiệm bằng 0,8. Dùng xét nghiệm chẩn đoán bệnh.

a) Tính độ không xác định của phép thử α với điều kiện β đã xảy ra, trong đó α: dương tính hay âm tính, β: bị bệnh hay không.

b) Lượng tin về bệnh chứa trong chẩn đoán xét nghiệm lớn hay bé?

10. Dùng một xét nghiệm để chẩn đoán bệnh. Xét nghiệm có xác suất đúng bằng 0,763. Giá trị dương tính của xét nghiệm bằng 0,1 và giá trị âm tính của xét nghiệm bằng 0,95.

a) Đoán một người xét nghiệm dương hay âm tính và có bệnh hay không khó hay dễ?

b) Gọi β: bị bệnh hay không, ε: đúng hay sai. Tính I (β,ε).

TƯƠNG QUAN

1. Lập phương trình $y = a_1x + b_1$, $x = a_2y + b_2$ và tính hệ số tương quan tuyến tính từ số liệu sau:

x_i	9	10	11	12	13	14	15
y_i	72,8	72,5	73,6	69,8	69,2	68,6	70,2

2. Lập phương trình $y = ax^2 + b$ từ số liệu sau:

x_i	1	2	3	4	5
y_i	0,1	3	8,1	14,9	23,9

3. Lập phương trình $y = ax^2 + bx + c$ từ các số liệu sau:

a)

x_i	1	2	3	4	5
y_i	2,9	8,9	19,1	33,2	50,8

b)

x_i	0,56	0,84	1,14	2,44	3,16
y_i	– 0,80	– 0,97	– 0,98	1,07	3,66

4. Tại một địa phương có 250 người. Khi thông báo dịch, có 25 người bị dịch. Sau 10 ngày thông báo có 33 người bị dịch. Sau 15 ngày thông báo có 37 người bị dịch. Lập hàm phát triển dịch x = f(t), từ đó cho biết số người bị dịch sau 30 ngày thông báo (Dịch không chữa được cho nên cách ly hoàn toàn với xung quanh).

5. Theo dõi phát triển dân số một quận thu được số liệu sau:

năm	x (số dân, người)	s (tỷ lệ sinh, đv: 1)	c (tỷ lệ chết, đv: 1)
1983	171.000	0,0240	0,00512
1984	175.300	0,0217	0,00499
1985	179.600	0,0194	0,00486
1986	183.900	0,0171	0,00473
1987	188.200	0,0148	0,00460

Tính hệ số tương quan tuyến tính và lập các phương trình $s=ax+b$, $c=a'x+b'$

6. Theo dõi phát triển dân số một xã thu được số liệu sau:

năm	x (số dân, người)	s (tỷ lệ sinh, đv: 1)	c (tỷ lệ chết, đv: 1)
1980	4.670	0,0411	0,0099
1981	4.860	0,0397	0,0074
1982	5.050	0,0352	0,0099
1983	5.170	0,0375	0,0064
1984	5.470	0,0336	0,0059

Tính hệ số tương quan tuyến tính và lập phương trình $y=ax+b$ với $y=s-c$.

7. Theo dõi phát triển dân số toàn quốc thu được số liệu sau:

năm	x(số dân, 1000ng)	s(tỷ lệ sinh, $^0/_{00}$)	c (tỷ lệ chết, $^0/_{00}$)
1981	54.927	30,02	6,98
1982	56.713	29,80	7,10
1983	57.442	29,30	7,08
1984	58.669	28,40	7,03
1985	59.872	28,44	6,94

Tính hệ số tương quan tuyến tính và lập các phương trình $s = ax + b$, $c = a'x + b'$ từ đó cho biết dân số ổn định cân bằng.

PHỤ LỤC

Bảng 1. HÀM PHÂN BỐ ∏ CỦA QUY LUẬT CHUẨN TẮC

Giá trị x nhỏ hơn 2

x	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	5398	5438	5478	5517	5557	5596	5636	5675	5714	5753
0,2	5793	5832	5871	5910	5948	5987	6026	6064	6103	6141
0,3	6179	6217	6255	6293	6331	6368	6406	6443	6480	6517
0,4	6554	6591	6628	6664	6700	6736	6772	6808	6844	6879
0,5	6915	6950	6985	7019	7054	7088	7123	7157	7190	7224
0,6	7257	7290	7324	7357	7389	7422	7454	7486	7517	7549
0,7	7580	7611	7642	7673	7704	7734	7764	7794	7823	7852
0,8	7881	7910	7939	7967	7995	8023	8051	8078	8106	8133
0,9	8159	8186	8212	8238	8264	8289	8315	8340	8365	8389
1,0	8413	8438	8461	8485	8508	8531	8554	8577	8599	8621
1,1	8643	8665	8686	8708	8729	8749	8770	8790	8810	8830
1,2	8849	8869	8888	8907	8925	8944	8962	8980	8997	9015
1,3	9032	9049	9066	9082	9099	9115	9131	9147	9162	9177
1,4	9192	9207	9222	9236	9251	9265	9279	9292	9306	9319
1,5	9332	9345	9357	9370	9382	9394	9406	9418	9429	9441
1,6	9452	9463	9474	9484	9495	9505	9515	9525	9535	9545
1,7	9554	9564	9573	9582	9591	9599	9608	9616	9625	9633
1,8	9641	9649	9656	9664	9671	9678	9686	9693	9699	9706
1,9	9713	9719	9726	9732	9738	9744	9750	9756	9761	9767

Giá trị x lớn hơn 2

x	2,0	2,1	2,2	2,3	2,4	2,5	2,6
π (x)	0,9772	0,9821	0,9861	0,9893	0,9918	0,9938	0,9953
x	2,7	2,8	2,9	3,0	3,1	3,2	3,3
π (x)	0,9965	0,9974	0,9981	0,99865	0,99904	0,99931	0,99952
x	3,4	3,5	3,6	3,8	4,0	4,5	
π (x)	0,99966	0,99976	0,999841	0,999928	0,999968	0,999997	

Chỉ dẫn

1. $\pi(x) = P\{-\infty < X \leq x\}$,

với số x nằm giữa 0 và 2, có hai số thập phân thì $\pi(x)$ nằm ở chỗ giao nhau của hàng và cột.

Ví dụ $x = 1,57$ cần tìm ở hàng 1,5 và cột 0,07. $\Pi(1,57) = 0,9418$

2. Trong các bài toán liên quan tới kiểm định giả thiết mức α, tra $t(\alpha)$ nếu kiểm định một phía, hoặc $t(\alpha/2)$ nếu kiểm định 2 phía trong bảng sau:

α	**0,2**	**0,1**	**0,05**	**0,01**	**0,001**
t(α)	0,8416	1,2816	1,6449	2,3263	3,0902
t(α/2)	1,2816	1,6449	1,9600	2,5758	3,2905

Bảng 2. QUY LUẬT STUDENT VỚI N BẬC TỰ DO

Giá trị của t(n, p)

n \ p (2p)	0,4 (0,8)	0,25 (0,50)	0,1 (0,2)	0,05 (0,10)	0,025 (0,050)	0,01 (0,02)	0,005 (0,01)	0,0005 (0,001)
1	0,325	1,000	3,078	6,314	12,706	31,821	63,657	636,62
2	0,289	0,816	1,886	2,920	4,303	6,965	9,925	31,598
3	0,277	0,765	1,638	2,353	3,182	4,541	5,841	12,924
4	0,271	0,741	1,533	2,132	2,776	3,747	4,604	8,610
5	0,267	0,727	1,476	2,015	2,571	3,365	4,032	6,869
6	0,265	0,718	1,440	1,943	2,447	3,143	3,707	5,959
7	0,263	0,711	1,415	1,895	2,365	2,988	3,499	5,408
8	0,262	0,706	1,397	1,860	2,306	2,896	3,355	5,041
9	0,261	0,703	1,383	1,833	2,262	2,821	3,250	4,781
10	0,260	0,700	1,372	1,812	2,228	2,764	3,169	4,587
11	0,260	0,697	1,363	1,796	2,201	2,718	3,106	4,437
12	0,259	0,695	1,356	1,782	2,179	2,681	3,055	4,318
13	0,259	0,694	1,350	1,771	2,160	2,650	3,012	4,221
14	0,258	0,692	1,345	1,761	2,145	2,624	2,977	4,140
15	0,258	0,691	1,341	1,753	2,131	2,602	2,947	4,073
16	0,258	0,690	1,337	1,746	2,120	2,583	2,921	4,015
17	0,257	0,689	1,333	1,740	2,110	2,567	2,898	3,965
18	0,257	0,688	1,330	1,734	2,101	2,552	2,878	3,022
19	0,257	0,688	1,328	1,729	2,093	2,539	2,861	3,883
20	0,257	0,687	1,325	1,725	2,086	2,528	2,845	3,850
21	0,257	0,686	1,323	1,721	2,080	2,518	2,831	3,819
22	0,256	0,686	1,321	1,717	2,074	2,508	2,819	3,792
23	0,256	0,685	1,319	1,714	2,069	2,500	2,807	3,767
24	0,256	0,685	1,318	1,711	2,064	2,492	2,797	3,745
25	0,256	0,684	1,316	1,708	2,060	2,485	2,787	3,725
26	0,256	0,684	1,315	1,706	2,056	2,479	2,779	3,707
27	0,256	0,684	1,314	1,703	2,052	2,473	2,771	3,690
28	0,256	0,683	1,313	1,701	2,048	2,467	2,763	3,674
29	0,256	0,683	1,311	1,699	2,045	2,462	2,756	3,659
30	0,256	0,683	1,310	1,697	2,042	2,457	2,750	3,646
40	0,255	0,681	1,303	1,684	2,021	2,423	2,704	3,551
60	0,254	0,679	1,296	1,671	2,000	2,390	2,660	3,460
∞	0,253	0,674	1,282	1,645	1,960	2,326	2,576	3,291

Chỉ dẫn

1. T_n có quy luật Student với n bậc tự do thì t(n,p) là giá trị được xác định thoả mãn điều kiện: $p\{T_n > t(n,p)\} = p$; $p\{|T_n| > t(n,p)\} = 2p$.

2. Số t(n,p) đọc ở hàng n cột p; ví dụ t(19; 0,05) = 1,729

3. Trong các bài toán liên quan tới kiểm định giả thiết mức α, tra t(n; α) nếu kiểm định một phía, hoặc t(n; α/2) nếu kiểm định 2 phía, khi đó tương ứng p = α hoặc 2p = α .

Bảng 3. QUY LUẬT χ^2 VỚI N BẬC TỰ DO

Giá trị của q(n, p)

n \ p	0,99	0,95	0,90	0,75	0,50	0,30	0,20	0,10	0,05	0,02	0,01
1	0,000	0,004	0,016	0,102	0,455	1,074	1,642	2,706	3,841	5,412	6,635
2	0,020	0,103	0,211	0,575	1,386	2,408	3,219	4,605	5,991	7,824	9,210
3	0,115	0,352	0,584	1,213	2,366	3,665	4,642	6,251	7,815	9,837	11,345
4	0,297	0,711	1,064	1,923	3,357	4,878	5,989	7,779	9,488	11,668	13,277
5	0,554	1,145	1,610	2,675	4,351	6,064	7,289	9,236	11,070	13,388	15,086
6	0,872	1,635	2,204	3,455	5,348	7,231	8,558	10,645	12,592	15,033	16,812
7	1,239	2,167	2,833	4,255	6,346	8,383	9,803	12,017	14,067	16,622	18,475
8	1,646	2,733	3,490	5,071	7,344	9,524	11,030	13,362	15,507	18,168	20,090
9	2,088	3,325	4,168	5,899	8,343	10,656	12,242	14,684	16,919	19,679	21,666
10	2,558	3,940	4,865	6,737	9,342	11,781	13,442	15,987	18,307	21,161	23,209
11	3,053	4,575	5,578	7,584	10,341	12,899	14,631	17,275	19,675	22,618	24,725
12	3,571	5,226	6,304	8,438	11,340	14,011	15,812	18,549	21,026	24,054	26,217
13	4,107	5,892	7,042	9,299	12,340	15,119	16,985	19,812	22,362	25,472	27,688
14	4,660	6,571	7,790	10,165	13,339	16,222	18,151	21,064	23,685	26,873	29,141
15	5,229	7,261	8,547	11,037	14,339	17,322	19,311	22,307	24,996	28,259	30,574
16	5,812	7,962	9,312	11,912	15,338	18,418	20,465	23,542	26,296	29,633	32,000
17	6,408	8,672	10,085	12,792	16,338	19,511	21,615	24,769	27,587	30,995	33,409
18	7,015	9,390	10,865	13,675	17,338	20,601	22,760	25,989	28,869	32,346	34,805
19	7,633	10,117	11,651	14,562	18,338	21,689	23,900	27,204	30,144	33,687	36,191
20	8,260	10,851	12,443	15,452	19,337	22,775	25,038	28,412	31,410	35,020	37,566
21	8,897	11,591	13,240	16,344	20,337	23,858	26,171	29,615	32,671	36,343	38,932
22	9,542	12,338	14,041	17,240	21,337	24,939	27,301	30,813	33,924	37,659	40,289
23	10,196	13,091	14,848	18,137	22,337	26,018	28,429	32,007	35,172	38,968	41,638
24	10,856	13,848	15,659	19,037	23,337	27,096	29,553	33,196	36,415	40,270	42,980
25	11,524	14,611	16,473	19,939	24,337	28,172	30,675	34,382	37,652	41,566	44,314
26	12,198	15,379	17,292	20,843	25,336	29,246	31,795	35,563	38,885	42,856	45,642
27	12,879	16,151	18,114	21,749	26,336	30,319	32,912	36,741	40,113	44,140	46,963
28	13,565	16,928	18,939	22,657	27,336	31,391	34,027	37,916	41,337	45,419	48,278
29	14,256	17,708	19,768	23,567	28,336	32,461	35,139	39,087	42,557	46,693	49,588
30	14,953	18,493	20,599	24,478	29,336	33,530	36,250	40,256	43,773	47,962	50,892

Chỉ dẫn

1. Q_n có quy luật χ^2 với n bậc tự do thì q(n,p) là giá trị được xác định thoả mãn điều kiện:

$$P\{Q_n > q(n,p)\} = p$$

2. Số q(n, p) đọc ở hàng n cột p, ví dụ: q(4; 0,05) = 9,488.

Bảng 4. QUY LUẬT FISHER – SNEDECOR

Giá trị của $f(n_1, n_2 ; 0{,}05)$

n_2 \ n_1	1	2	3	4	5	6	8	10	15	20	50	∞
1	161	199	216	225	230	234	239	242	246	248	252	254
2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4	19,5	19,5
3	10,1	9,55	9,28	9,12	9,01	8,94	8,84	8,78	8,70	8,66	8,58	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,96	5,85	5,80	5,70	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,74	4,62	4,56	4,44	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,06	3,94	3,87	3,75	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,63	3,51	3,44	3,32	3,21
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,34	3,22	3,15	3,03	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,13	3,01	2,93	2,80	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,97	2,84	2,77	2,64	2,54
11	4,84	3,98	3,59	3,36	3,20	3,10	2,95	2,86	2,72	2,65	2,50	2,40
12	4,75	3,89	3,49	3,26	3,11	3,00	2,85	2,76	2,62	2,54	2,40	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,77	2,67	2,53	2,46	2,32	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,60	2,46	2,39	2,24	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,55	2,40	2,33	2,18	2,07
16	4,50	3,63	3,24	3,01	2,85	2,74	2,59	2,49	2,35	2,28	2,13	2,01
17	4,45	3,59	3,19	2,96	2,81	2,70	2,55	2,45	2,31	2,23	2,08	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,41	2,27	2,19	2,04	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,38	2,23	2,15	2,00	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,35	2,20	2,12	1,96	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,32	2,18	2,09	1,93	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,30	2,15	2,07	1,91	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,28	2,13	2,04	1,88	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,35	2,26	2,11	2,02	1,86	1,73
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,24	2,09	2,00	1,84	1,71
26	4,22	3,37	2,97	2,74	2,59	2,47	2,32	2,22	2,07	1,99	1,82	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,31	2,20	2,06	1,97	1,80	1,67
28	4,20	3,34	2,95	2,71	2,56	2,45	2,29	2,19	2,04	1,96	1,78	1,65
29	4,18	3,33	2,93	2,70	2,55	2,43	2,28	2,18	2,03	1,94	1,77	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,16	2,01	1,93	1,76	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,07	1,92	1,84	1,66	1,51
60	4,00	3,15	2,76	2,53	2,37	2,25	2,10	1,99	1,84	1,75	1,56	1,39
100	3,94	3,09	2,70	2,46	2,30	2,19	2,03	1,92	1,77	1,68	1,48	1,28
200	3,89	3,04	2,65	2,41	2,26	2,14	1,98	1,87	1,71	1,62	1,42	1,19
∞	3,84	3,00	2,00	2,37	2,21	2,10	1,94	1,83	1,67	1,57	1,35	1,00

Chỉ dẫn

1. F_{n_1,n_2} có quy luật Fisơ – Snedecor với n_1 và n_2 bậc tự do thì $f(n_1,n_2,p)$ là giá trị được xác định thoả mãn điều kiện:

$$P\left\{F_{n_1,n_2} > f(n_1,n_2,p)\right\} = p$$

2. Số $f(n_1,n_2,\ 0{,}05)$ tìm trong cột n_1, hàng n_2 ; ví dụ: $f(15,1;\ 0{,}05) = 246$.

Bảng 5. GIÁ TRỊ CỦA HÀM SỐ $-P*LOG_2P$

P	0	1	2	3	4	5	6	7	8	9
0.00		0.0100	0.0179	0.0251	0.0319	0.0382	0.0443	0.0501	0.0557	0.0612
1	0.0664	0.0716	0.0766	0.0814	0.0862	0.0909	0.0955	0.0999	0.1043	0.1086
2	0.1129	0.1170	0.1211	0.1252	0.1291	0.1330	0.1369	0.1407	0.1444	0.1481
3	0.1518	0.1554	0.1589	0.1624	0.1659	0.1693	0.1727	0.1760	0.1793	0.1825
4	0.1858	0.1889	0.1921	0.1952	0.1983	0.2013	0.2043	0.2073	0.2103	0.2132
0.05	0.2161	0.2190	0.2218	0.2246	0.2274	0.2301	0.2329	0.2356	0.2383	0.2409
6	0.2435	0.2461	0.2487	0.2513	0.2538	0.2563	0.2588	0.2613	0.2637	0.2662
7	0.2686	0.2709	0.2733	0.2756	0.2780	0.2803	0.2826	0.2848	0.2871	0.2893
8	0.2915	0.2937	0.2959	0.2980	0.3002	0.3023	0.3044	0.3065	0.3086	0.3106
9	0.3127	0.3147	0.3167	0.3187	0.3207	0.3226	0.3246	0.3265	0.3284	0.3303
0.10	0.3322	0.3341	0.3359	0.3378	0.3396	0.3414	0.3432	0.3450	0.3468	0.3485
11	0.3503	0.3520	0.3537	0.3555	0.3571	0.3588	0.3605	0.3622	0.3638	0.3654
12	0.3671	0.3687	0.3703	0.3719	0.3734	0.3750	0.3766	0.3781	0.3796	0.3811
13	0.3826	0.3841	0.3856	0.3871	0.3886	0.3900	0.3915	0.3929	0.3943	0.3957
14	0.3971	0.3985	0.3999	0.4012	0.4026	0.4040	0.4053	0.4066	0.4079	0.4092
0.15	0.4105	0.4118	0.4131	0.4144	0.4156	0.4169	0.4181	0.4194	0.4206	0.4218
16	0.4230	0.4242	0.4254	0.4266	0.4278	0.4289	0.4301	0.4312	0.4323	0.4335
17	0.4346	0.4357	0.4368	0.4379	0.4390	0.4401	0.4411	0.4422	0.4432	0.4443
18	0.4453	0.4463	0.4474	0.4484	0.4494	0.4504	0.4514	0.4523	0.4533	0.4543
19	0.4552	0.4562	0.4571	0.4581	0.4590	0.4599	0.4608	0.4617	0.4626	0.4635
0.20	0.4644	0.4653	0.4661	0.4670	0.4678	0.4687	0.4695	0.4704	0.4712	0.4720
21	0.4728	0.4736	0.4744	0.4752	0.4760	0.4768	0.4776	0.4783	0.4791	0.4798
22	0.4806	0.4813	0.4820	0.4828	0.4835	0.4842	0.4849	0.4856	0.4863	0.4870
23	0.4877	0.4883	0.4890	0.4897	0.4903	0.4910	0.4916	0.4923	0.4929	0.4935
24	0.4941	0.4947	0.4954	0.4960	0.4966	0.4971	0.4977	0.4983	0.4989	0.4994
0.25	0.5000	0.5006	0.5011	0.5016	0.5022	0.5027	0.5032	0.5038	0.5043	0.5048
26	0.5053	0.5058	0.5063	0.5068	0.5072	0.5077	0.5082	0.5087	0.5091	0.5096
27	0.5100	0.5105	0.5109	0.5113	0.5119	0.5122	0.5126	0.5130	0.5134	0.5138
28	0.5142	0.5146	0.5150	0.5154	0.5158	0.5161	0.5165	0.5169	0.5172	0.5176
29	0.5179	0.5182	0.5186	0.5189	0.5192	0.5196	0.5199	0.5202	0.5205	0.5208
0.30	0.5211	0.5214	0.5217	0.5220	0.5222	0.5225	0.5228	0.5230	0.5233	0.5235
31	0.5238	0.5240	0.5243	0.5245	0.5247	0.5250	0.5252	0.5254	0.5256	0.5258
32	0.5260	0.5262	0.5264	0.5266	0.5268	0.5270	0.5272	0.5273	0.5275	0.5277
33	0.5278	0.5280	0.5281	0.5283	0.5284	0.5286	0.5287	0.5288	0.5289	0.5291
34	0.5292	0.5293	0.5294	0.5295	0.5296	0.5297	0.5298	0.5299	0.5299	0.5300
P	0	1	2	3	4	5	6	7	8	9

GIÁ TRỊ CỦA HÀM SỐ – P*LOG$_2$P (tiếp theo)

P	0	1	2	3	4	5	6	7	8	9
0.35	0.5301	0.5302	0.5302	0.5303	0.5304	0.5304	0.5305	0.5305	0.5305	0.5306
36	0.5306	0.5306	0.5307	0.5307	0.5307	0.5307	0.5307	0.5307	0.5307	0.5307
37	0.5307	0.5307	0.5307	0.5307	0.5307	0.5306	0.5306	0.5306	0.5305	0.5305
38	0.5305	0.5304	0.5304	0.5303	0.5302	0.5302	0.5301	0.5300	0.5300	0.5299
39	0.5298	0.5297	0.5296	0.5295	0.5294	0.5293	0.5292	0.5291	0.5290	0.5289
0.40	0.5288	0.5286	0.5285	0.5284	0.5283	0.5281	0.5280	0.5278	0.5277	0.5275
41	0.5274	0.5272	0.5271	0.5269	0.5267	0.5266	0.5264	0.5262	0.5260	0.5258
42	0.5256	0.5255	0.5253	0.5251	0.5249	0.5246	0.5244	0.5242	0.5240	0.5238
43	0.5236	0.5233	0.5231	0.5229	0.5226	0.5224	0.5222	0.5219	0.5217	0.5214
44	0.5211	0.5209	0.5206	0.5204	0.5201	0.5198	0.5195	0.5193	0.5190	0.5187
0.45	0.5184	0.5181	0.5178	0.5175	0.5172	0.5169	0.5166	0.5163	0.5160	0.5157
46	0.5153	0.5150	0.5147	0.5144	0.5140	0.5137	0.5133	0.5130	0.5127	0.5123
47	0.5120	0.5116	0.5112	0.5109	0.5105	0.5102	0.5098	0.5094	0.5090	0.5087
48	0.5083	0.5079	0.5075	0.5071	0.5067	0.5063	0.5059	0.5055	0.5051	0.5047
49	5043	0.5039	0.5034	0.5030	0.5026	0.5022	0.5017	0.5013	0.5009	0.5004
0.50	0.5000	0.4996	0.4991	0.4987	0.4982	0.4978	0.4973	0.4968	0.4964	0.4959
51	0.4954	0.4950	0.4945	0.4940	0.4935	0.4930	0.4926	0.4921	0.4916	0.4911
52	0.4906	0.4901	0.4896	0.4891	0.4886	0.4880	0.4875	0.4870	0.4865	0.4860
53	0.4854	0.4849	0.4844	0.4839	0.4833	0.4828	0.4822	0.4817	0.4811	0.4806
54	0.4800	0.4795	0.4789	0.4784	0.4778	0.4772	0.4767	0.4761	0.4755	0.4750
0.55	0.4744	0.4738	0.4732	0.4726	0.4720	0.4714	0.4708	0.4702	0.4696	0.4690
56	0.4684	0.4678	0.4672	0.4666	0.4660	0.4654	0.4648	0.4641	0.4635	0.4629
57	0.4623	0.4616	0.4610	0.4603	0.4597	0.4591	0.4584	0.4578	0.4571	0.4565
58	0.4558	0.4551	0.4545	0.4538	0.4532	0.4525	0.4518	0.4511	0.4505	0.4498
59	0.4491	0.4484	0.4477	0.4471	0.4464	0.4457	0.4450	0.4443	0.4436	0.4429
0.60	0.4422	0.4415	0.4408	0.4401	0.4393	0.4386	0.4379	0.4372	0.4365	0.4357
61	0.4350	0.4343	0.4335	0.4328	0.4321	0.4313	0.4306	0.4298	0.4291	0.4283
62	0.4276	0.4268	0.4261	0.4253	0.4246	0.4238	0.4230	0.4223	0.4215	0.4207
63	0.4199	0.4192	0.4184	0.4176	0.4168	0.4160	0.4152	0.4145	0.4137	0.4129
64	0.4121	0.4113	0.4105	0.4097	0.4089	0.4080	0.4072	0.4064	0.4056	0.4048
0.65	0.4040	0.4031	0.4023	0.4015	0.4007	0.3998	0.3990	0.3982	0.3973	0.3965
66	0.3956	0.3948	0.3940	0.3931	0.3923	0.3914	0.3905	0.3897	0.3888	0.3880
67	0.3871	0.3862	0.3854	0.3845	0.3836	0.3828	0.3819	0.3810	0.3801	0.3792
68	0.3783	0.3775	0.3766	0.3757	0.3748	0.3739	0.3730	0.3721	0.3712	0.3703
69	0.3694	0.3685	0.3676	0.3666	0.3657	0.3648	0.3639	0.3630	0.3621	0.3611
P	0	1	2	3	4	5	6	7	8	9

GIÁ TRỊ CỦA HÀM SỐ – P*LOG$_2$P (tiếp theo)

P	0	1	2	3	4	5	6	7	8	9
0.70	0.3602	0.3593	0.3583	0.3574	0.3565	0.3555	0.3546	0.3537	0.3527	0.3518
71	0.3508	0.3499	0.3489	0.3480	0.3470	0.3460	0.3451	0.3441	0.3432	0.3422
72	0.3412	0.3403	0.3393	0.3383	0.3373	0.3364	0.3354	0.3344	0.3334	0.3324
73	0.3314	0.3305	0.3295	0.3285	0.3275	0.3265	0.3255	0.3245	0.3235	0.3225
74	0.3215	0.3204	0.3194	0.3184	0.3174	0.3164	0.3154	0.3144	0.3133	0.3123
0.75	0.3113	0.3102	0.3092	0.3082	0.3072	0.3061	0.3051	0.3040	0.3030	0.3020
76	0.3009	0.2999	0.2988	0.2978	0.2967	0.2956	0.2946	0.2935	0.2925	0.2914
77	0.2903	0.2893	0.2882	0.2871	0.2861	0.2850	0.2839	0.2828	0.2818	0.2807
78	0.2796	0.2785	0.2774	0.2763	0.2752	0.2741	0.2731	0.2720	0.2709	0.2698
79	0.2687	0.2676	0.2665	0.2653	0.2642	0.2631	0.2620	0.2609	0.2598	0.2587
0.80	0.2575	0.2564	0.2553	0.2542	0.2530	0.2519	0.2508	0.2497	0.2485	0.2474
81	0.2462	0.2451	0.2440	0.2428	0.2417	0.2405	0.2394	0.2382	0.2371	0.2359
82	0.2348	0.2336	0.2325	0.2313	0.2301	0.2290	0.2278	0.2266	0.2255	0.2243
83	0.2231	0.2219	0.2208	0.2196	0.2184	0.2172	0.2160	0.2149	0.2137	0.2125
84	0.2113	0.2101	0.2089	0.2077	0.2065	0.2053	0.2041	0.2029	0.2017	0.2005
085	0.1993	0.1981	0.1969	0.1957	0.1944	0.1932	0.1920	0.1908	0.1896	0.1884
86	0.1871	0.1859	0.1847	0.1834	0.1822	0.1810	0.1797	0.1785	0.1773	0.1760
87	0.1748	0.1736	0.1723	0.1711	0.1698	0.1686	0.1673	0.1661	0.1648	0.1636
88	0.1623	0.1610	0.1598	0.1585	0.1572	0.1560	0.1547	0.1534	0.1522	0.1509
89	0.1496	0.1484	0.1471	0.1458	0.1445	0.1432	0.1420	0.1407	0.1394	0.1381
0.90	0.1368	0.1355	0.1342	0.1329	0.1316	0.1303	0.1290	0.1277	0.1264	0.1251
91	0.1238	0.1225	0.1212	0.1199	0.1186	0.1173	0.1159	0.1146	0.1133	0.1120
92	0.1107	0.1093	0.1080	0.1067	0.1054	0.1040	0.1027	0.1014	0.1000	0.0987
93	0.0974	0.0960	0.0947	0.0933	0.0920	0.0907	0.0893	0.0880	0.0866	0.0853
94	0.0839	0.0826	0.0812	0.0798	0.0785	0.0771	0.0758	0.0744	0.0730	0.0717
0.95	0.0703	0.0689	0.0676	0.0662	0.0648	0.0634	0.0621	0.0607	0.0593	0.0579
96	0.0565	0.0552	0.0538	0.0524	0.0510	0.0496	0.0482	0.0468	0.0454	0.0440
97	0.0426	0.0412	0.0398	0.0384	0.0370	0.0356	0.0342	0.0328	0.0314	0.0300
98	0.0286	0.0271	0.0257	0.0243	0.0229	0.0215	0.0201	0.0186	0.0172	0.0158
99	0144	0.0129	0.0115	0.0101	0.0086	0.0072	0.0058	0.0043	0.0029	0.0014
P	0	1	2	3	4	5	6	7	8	9

TÀI LIỆU THAM KHẢO

1. Guy LeFort, *Toán cao cấp tập 4 – Phép tính xác suất thống kê* (Lưu hành nội bộ), Bộ Đại học và Trung học chuyên nghiệp, 1971.
2. L.Z.Rumsixki, *Phương pháp toán học xử lý các kết quả thực nghiệm*, Nhà xuất bản Khoa học và kỹ thuật; 1972.
3. Nguyễn Đình Trí (Chủ biên), *Toán cao cấp tập III*, Nhà xuất bản Giáo dục, 1997.
4. Nguyễn Cao Văn (Chủ biên), *Lý thuyết xác suất và thống kê toán học*, Nhà xuất bản khoa học và kỹ thuật, 1996.
5. Đỗ Như Cương, *Giáo trình Toán cao cấp*, Trường Đại học Y Hà Nội, 1977.
6. Bộ Y tế, *Các giá trị sinh học người Việt nam bình thường thập kỷ 90 – thế kỷ XX*, Nhà xuất bản Y học, 2003.
7. James T.McClave, Frank H.Dietrich, *Statistic – Dellen publishing company*, 1979.
8. John Neter, William Wasserman, Michael H.Kutner, *Applied Linear Statistical Models*, 1990.

Chịu trách nhiệm xuất bản :
Chủ tịch HĐQT kiêm Tổng Giám đốc NGÔ TRẦN ÁI
Phó Tổng Giám đốc kiêm Tổng biên tập NGUYỄN QUÝ THAO

Chịu trách nhiệm nội dung :
Chủ tịch HĐQT kiêm Giám đốc Công ty CP Sách ĐH–DN TRẦN NHẬT TÂN

Biên tập nội dung và sửa bản in :
NGÔ THỊ THANH BÌNH

Biên tập mĩ thuật và trình bày bìa :
TẠ TRỌNG TRÍ

Thiết kế sách và chế bản :
BÌNH MINH

XÁC SUẤT THỐNG KÊ

Mã số: 7B725y9 - DAI

In 1.000 cuốn (QĐ04), khổ 19 x 27cm. In tại Công ti CP In Phúc Yên.
Địa chỉ: Đường Trần Phú, thị xã Phúc Yên, Vĩnh Phúc.
Số ĐKKH xuất bản: 04 - 2009/CXB/262-2117/GD.
In xong và nộp lưu chiểu tháng 1 năm 2009.

CÔNG TY CỔ PHẦN SÁCH ĐẠI HỌC - DẠY NGHỀ
HEVOBCO
25 HÀN THUYÊN - HÀ NỘI
Website : www.hevobco.com.vn

TÌM ĐỌC SÁCH THAM KHẢO Y HỌC
CỦA NHÀ XUẤT BẢN GIÁO DỤC

1.	**Sinh học phân tử**	*GS. TS. Nguyễn Văn Thanh (Chủ biên)*
2.	**Bào chế và sinh dược học – Tập 2**	*PGS. TS. Lê Quan Nghiệm – TS. Huỳnh Văn Hoá (Đồng chủ biên)*
3.	**Thực vật dược**	*TS. Trương Thị Đẹp (Chủ biên)*
4.	**Ký sinh trùng**	*PGS. TS. Phạm Văn Thân (Chủ biên)*
5.	**Hoá đại cương**	*PGS. TSKH. Phan An (Chủ biên)*
6.	**Điều dưỡng cơ bản 1**	*PGS. TS. Phạm Văn Linh – TS. Lê Văn An (Đồng Chủ biên)*
7.	**Điều dưỡng cơ bản 2**	*PGS. TS. Hoàng Ngọc Chương BSCKII. Trần Đức Thái (Đồng Chủ biên)*
8.	**Kiểm nghiệm thuốc**	*Trần Tích (Chủ biên)*
9.	**Nhãn khoa**	*PGS. TS. Hoàng Thị Phúc (Chủ biên)*
10.	**Sinh lý học**	*GS. TS. Phạm Thị Minh Đức (Chủ biên)*
11.	**Phẫu thuật miệng – Tập 1**	*TS. BS. Lê Đức Lánh (Chủ biên)*
12.	**Hoá phân tích – Tập 1**	*PGS. TS. Võ Thị Bạch Huệ (Chủ biên)*
13.	**Công nghệ bào chế dược phẩm**	*PGS. TS. Hoàng Minh Châu (Chủ biên)*
14.	**Dược lý học – Tập 1**	*GS. TS. Đào Văn Phan (Chủ biên)*
15.	**Vệ sinh phòng bệnh**	*PGS. TS. Trần Văn Dần (Chủ biên)*
16.	**Dinh dưỡng**	*TS. Phạm Thị Thuý Hoà (Chủ biên)*
17.	**Sức khoẻ sinh sản**	*TS. Bùi Thị Thu Hà (Chủ biên)*
18.	**Lý thuyết thiết bị hình ảnh y tế – Tập 1**	*KS. Trần Văn Sơn (Chủ biên)*
19.	**Lý thuyết thiết bị hình ảnh y tế – Tập 2**	*KS. Lê Tiến Khoan (Chủ biên)*

Bạn đọc có thể mua tại các Công ty Sách – Thiết bị trường học ở các địa phương hoặc các Cửa hàng sách của Nhà xuất bản Giáo dục :

Tại Hà Nội : 25 Hàn Thuyên ; 187B Giảng Võ ; 232 Tây Sơn ; 23 Tràng Tiền ;

Tại Đà Nẵng : Số 15 Nguyễn Chí Thanh ; Số 62 Nguyễn Chí Thanh ;

Tại Thành phố Hồ Chí Minh : Cửa hàng 451B – 453, Hai Bà Trưng, Quận 3 ;
240 Trần Bình Trọng – Quận 5.

Tại Thành phố Cần Thơ : Số 5/5, đường 30/4 ;

Website : www.nxbgd.com.vn

8934980936344

Giá: 32.000 đ