Xác suất thống kê

Mục lục

[1. Ôn tập tập hợp và giải tích tổ hợp 4](#_Toc92979141)

[1.1. Tập hợp (set) 4](#_Toc92979142)

[1.2. Giải tích tổ hợp 4](#_Toc92979143)

[1.2.1. Chỉnh hợp (permutation) 4](#_Toc92979144)

[1.2.2. Hoán vị (factorial) 4](#_Toc92979145)

[1.2.3. Tổ hợp (combination) 4](#_Toc92979146)

[1.2.4. Nhị thức newton 4](#_Toc92979147)

[2. Khái niệm cơ bản về xác suất 4](#_Toc92979148)

[2.1. Biến cố và xác suất 4](#_Toc92979149)

[2.1.1. Biến cố ngẫu nhiên 4](#_Toc92979150)

[2.1.2. Quan hệ giữa các biến cố 5](#_Toc92979151)

[2.1.3. Các phép toán trên biến cố 5](#_Toc92979152)

[2.2. Công thức tính xác suất cơ bản 5](#_Toc92979153)

[2.2.1. Cộng xác suất (A+B) 5](#_Toc92979154)

[2.2.2. Xác suất giao (AB) 5](#_Toc92979155)

[2.2.3. Xác suất với điều kiện (A|B) 5](#_Toc92979156)

[2.2.4. Hai biến cố độc lập (hid) 5](#_Toc92979157)

[2.2.5. Công thức đầy đủ xác suất 6](#_Toc92979158)

[2.2.6. Công thức bayes 6](#_Toc92979159)

[3. Biến ngẫu nhiên 6](#_Toc92979160)

[3.1. Khái niệm biến ngẫu nhiên 6](#_Toc92979161)

[3.2. Phân phối xác suất 6](#_Toc92979162)

[3.2.1. Bảng phân phối xác suất 6](#_Toc92979163)

[3.2.2. Hàm mật độ xác suất 6](#_Toc92979164)

[3.2.3. Hàm phân phối xác suất 6](#_Toc92979165)

[3.3. Phân phối xác suất liên tục 8](#_Toc92979166)

[3.3.1. Hàm mật độ xác suất 8](#_Toc92979167)

[3.3.2. Hàm phân phối xác suất 8](#_Toc92979168)

[3.4. Các đặc trưng số của biến ngẫu nhiên 8](#_Toc92979169)

[3.4.1. Kỳ vọng - E(X) - μ - trung bình - mean 8](#_Toc92979170)

[3.4.2. Phương sai - Var(X) 8](#_Toc92979171)

[3.4.3. Mod 9](#_Toc92979172)

[3.4.4. Trung vị - Med 9](#_Toc92979173)

[4. Các phân phối xác suất thường gặp 9](#_Toc92979174)

[4.1. Biến ngẫu nhiên Bernouli (rr) 9](#_Toc92979175)

[4.2. Phân phối nhị thức (rr) (Binomial Dis) 9](#_Toc92979176)

[4.3. Phân phối siêu bội (rr) (Hypergeometric Dis) 10](#_Toc92979177)

[4.4. Phân phối Poisson (Poisson Dis) 10](#_Toc92979178)

[4.5. Phân phối đều (Uniform Dis) 11](#_Toc92979179)

[4.6. Phân phối chuẩn (liên tục) (Normal Distribution) 11](#_Toc92979180)

[4.7. Các phân phối không học (khó) 14](#_Toc92979181)

[5. Biến ngẫu nhiên (bỏ) 14](#_Toc92979182)

[5.1. Rời rạc 2 chiều 14](#_Toc92979183)

[5.1.1. Hàm xác suất đồng thời 14](#_Toc92979184)

[5.1.2. Hàm xác suất có điều kiện 15](#_Toc92979185)

[5.1.3. Sự độc lập 16](#_Toc92979186)

[5.2. Liên tục 2 chiều 16](#_Toc92979187)

[5.2.1. Hàm xác suất đồng thời 16](#_Toc92979188)

[5.2.2. Hàm xác suất có điều kiện 17](#_Toc92979189)

[5.2.3. Sự độc lập 17](#_Toc92979190)

[5.3. Hiệp phương sai và hệ số tương quan 18](#_Toc92979191)

[5.3.1. Hiệp phương sai 18](#_Toc92979192)

[5.3.2. Hệ số tương quan 18](#_Toc92979193)

[5.4. Sửa bài tập tổng quát 19](#_Toc92979194)

[6. Thống kê mô tả 19](#_Toc92979195)

[6.1. Giới thiệu về thống kê 19](#_Toc92979196)

[6.2. Mô tả dữ liệu bằng đồ thị 19](#_Toc92979197)

[6.3. Mô tả dữ liệu bằng phương pháp số 19](#_Toc92979198)

[7. Phân phối mẫu 20](#_Toc92979199)

[7.1. Khái niệm 20](#_Toc92979200)

[7.2. Các kết quả quan trọng (đã ẩn) 20](#_Toc92979201)

[8. Khoảng tin cậy 20](#_Toc92979205)

[8.1. Giới thiệu 20](#_Toc92979206)

[8.2. Khoảng tin cậy trung bình (đã ẩn) 20](#_Toc92979207)

[8.2.4. Tóm tắt 20](#_Toc92979211)

[8.2.5. Ví dụ 20](#_Toc92979212)

[8.3. Khoảng tin cậy tỉ lệ 21](#_Toc92979213)

[9. Kiểm định giả thuyết thống kê 22](#_Toc92979214)

[9.1. Bài toán kiểm định 22](#_Toc92979215)

[9.2. Kiểm định giả thuyết trường hợp 1 mẫu 22](#_Toc92979216)

[9.2.1. Kiểm định μ, TH1: Biết 22](#_Toc92979217)

[9.2.1. Kiểm định μ, TH2: không biết , mẫu lớn 24](#_Toc92979218)

[9.2.2. Kiểm định μ, TH3: Không biết , mẫu nhỏ < 30 24](#_Toc92979219)

[9.2.3. Kiểm định giả thuyết cho tỷ lệ 24](#_Toc92979220)

[9.3. Kiểm định giả thuyết trường hợp 2 mẫu độc lập 25](#_Toc92979221)

[9.3.1. So sánh hai μ, biết σ 25](#_Toc92979222)

[9.3.2. So sánh hai μ, không biết σ, mẫu lớn > 30 26](#_Toc92979223)

[9.3.3. So sánh hai μ, không biết σ, mẫu nhỏ 26](#_Toc92979224)

[9.3.4. So sánh hai tỷ lệ 28](#_Toc92979225)

[10. Hồi quy và tương quan 29](#_Toc92979226)

[10.1. Giới thiệu 29](#_Toc92979227)

[10.2. Mô hình hồi quy tuyến tính đơn 29](#_Toc92979228)

[10.2.1. Mô hình 29](#_Toc92979229)

[10.2.2. Phương pháp bình phương bé nhất 30](#_Toc92979230)

1. Ôn tập tập hợp và giải tích tổ hợp
   1. Tập hợp (set)

- Các phép toán trên tập hợp:

* A Giao B:
* A hợp/hội B:
* A hiệu B:
* Phần bù của A:
  1. Giải tích tổ hợp

- Quy tắc nhân, quy tắc cộng

- Các phương pháp nhóm: nhóm có thứ tự, nhóm không có thứ tự, nhóm có lặp và nhóm không có lặp

* + 1. Chỉnh hợp (permutation)

- Chỉnh hợp chập k của n là số cách chọn một nhóm có thứ tự gồm k phần tử khác nhau từ n phần tử, kí hiệu hay, tính chất

- Chỉnh hợp lặp:

* + 1. Hoán vị (factorial)

- Hoán vị là trường hợp đặc biệt của chỉnh hợp:

* + 1. Tổ hợp (combination)

- Tổ hợp chập k của n phần tử là số cách chọn một nhóm k phần tử không phân biệt thứ tự khác nhau chọn từ n phần tử đã cho, kí hiệu , tính chất

- Tính chất:

- Tổ hợp lặp (học ở môn TRR)

* + 1. Nhị thức newton

- Với mọi n ∈ N và mọi cặp số a,b:

1. Khái niệm cơ bản về xác suất
   1. Biến cố và xác suất
      1. Biến cố ngẫu nhiên

- Các khái niệm:

* ~ Không gian mẫu ~ không gian biến cố sơ cấp: tập hợp tất cả kết quả có thể xảy ra
* ω ~ Biến cố sơ cấp: là một kết quả của phép thử ngẫu nhiên
* Biến cố ngẫu nhiên (biến cố): là tập con của và là tập hợp của các biến cố nhất định
* ∅ ~ Biến cố bất khả: là biến cố không thể xảy ra

- Ví dụ: Tung 2 con xúc xắc. Gọi "tổng số chấm ở mặt trên của 2 con xúc xắc = i" . Không gian các biến cố sơ cấp:

🡪 A, B, D gọi là các biến cố ngẫu nhiên

* + 1. Quan hệ giữa các biến cố

- A kéo theo B: , A xảy ra thì B phải xảy ra (A là biến cố thuận lợi cho B)

- A tương đương B: A = B , nếu A xảy ra thì B xảy ra và ngược lại

* + 1. Các phép toán trên biến cố

- Các phép toán trên biến cố:

* Tích ~ giao ~ và ~ AB ~ : là biến cố khi A và B cùng xảy ra, khi AB = ta nói A xung khắc B
* Hiệu ~ : là biến cố khi A xảy ra nhưng B không xảy ra
* Biến cố đối lập của A, kí hiệu là biến cố xảy ra khi A không xảy ra và ngược lại, hay

- Tính chất:

* (Tôi không mạnh **hay** không yếu : tôi không đồng thời mạnh và yếu : không xảy ra 1 trong 2 hoặc không xảy ra cả 2).
* = (Tôi không mạnh **và** không yếu : tôi đồng thời không mạnh và không yếu)

- Hệ/họ đầy đủ các biến cố là tập hợp các biến cố mà

* 1. Công thức tính xác suất cơ bản
     1. Cộng xác suất (A+B)

- Công xác xuất ta trừ phần giao, nếu xung khắc thì bỏ qua

- Cộng 2:

- Cộng 3:

- Cộng n biến cố (hided)

* + 1. Xác suất giao (AB)

- Cho 9 khóa, 2 mở được, tìm xác suất mở dc đúng lần thứ 2: là giao giữa lần 1 không trúng và lần 2 trúng:

🡪 Tổng quát (hid):

* + 1. Xác suất với điều kiện (A|B)

- Xác suất với điều kiện hay xác suất xảy ra biến cố A với điều kiện biến cố B đã xảy ra là:

(Từ công thức giao biến cố)

- Tính chất:

* + 1. Hai biến cố độc lập (hid)

- A và B độc lập khi 🡪 Khi đó

(Lưu ý: độc lập ≠ xung khắc, A, B độc lập nhưng )

🡪 Tổng quát:

* + 1. Công thức đầy đủ xác suất

- Biến cố của A khi đã xảy ra B với từng trường hợp của B:

C/M: A = = = =

🡪 Cho 2 biến cố A và B bất kì:

- Tổng quát: Cho Bi là một họ đầy đủ thì:

* + 1. Công thức bayes

- Là phiên bản ngược của công thức đầy đủ xác suất

- Xét là một họ đầy đủ các biến cố, biến cố A với . Với i = 1,2,..,n

- Lưu ý:

* P(AB̅) + P(AB) = P(A)
* P(A|B) + P(A̅|B) = 1

1. Biến ngẫu nhiên
   1. Khái niệm biến ngẫu nhiên

- Biến ngẫu nhiên X là 1 ánh xạ từ không gian biến cố sơ cấp (Ω) vào R. Các biến ngẫu nhiên thường dùng các chữ cái in hoa X, Y, Z để kí hiệu, chữ thường x, y, z,... để chỉ các giá trị của biến ngẫu nhiên

- Biến ngẫu nhiên gồm 2 loại: rời rạc (hữu hạn và đếm được) và liên tục

* 1. Phân phối xác suất
     1. Bảng phân phối xác suất

- Để mô tả xác suất ngẫu nhiên X nhận giá trị xác suất tương ứng, người ta dùng bảng phân phối xác suất

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  |  |  |  |

* + 1. Hàm mật độ xác suất

Chart

Description automatically generated- Xét biến ngẫu nhiên rời rạc X nhận các giá trị , một hàm mật độ xác suất của X là hàm thỏa:

VD: Khi tung 1 súc sắc, X là số nút nhận được, hàm mật độ của X:

* + 1. Hàm phân phối xác suất

Rời rạc

- Hàm phân-phối/tích-lũy xác suất của 1 biến X (VD: tỉ lệ học sinh (X) dưới hoặc bằng x điểm khi x thuộc từ 7-9 điểm) được xác định là:

=

Chart, line chart

Description automatically generatedVD: tiếp VD 3.2.2 - Tung 1 xúc sắc: hàm phân phối của X:

🡪 Nhận xét hàm phân phối xác suất:

* Là một hàm không giảm
* Là một hàm liên tục phải hay
* F(-∞) = 0, F(+∞) = 1
* vì:
* VD: Viết hàm phân phối xác suất, suy ra biến ngẫu nhiên khác

A picture containing calendar

Description automatically generated

Calendar

Description automatically generated

Table, Excel

Description automatically generated

* 1. Phân phối xác suất liên tục
     1. Hàm mật độ xác suất

- Mọi hàm f(x) không âm trên R và thì là một hàm mật độ của một biến ngẫu nhiên X nào đó

* + 1. Hàm phân phối xác suất

- Hàm phân phối xác suất của biến ngẫu nhiên X có hàm mật độ f(x) là

- Nếu X là biến ngẫu nhiên liên tục, với x1 và x2­ bất kỳ: = ... (ko xét dấu bằng với hàm liên tục)

* VD: Viết hàm phân phối xác suất

**Đề:**

là hàm mật độ của biến ngẫu nhiên vì   
Khi   
Khi   
Khi

* 1. Các đặc trưng số của biến ngẫu nhiên
     1. Kỳ vọng - E(X) - μ - trung bình - mean

- Lấy tất cả giá trị x nhân với xác suất của nó, là giá trị trung bình theo xác suất, kỳ vọng phản ánh giá trị trung tân (điểm trung bình) của một phần phối

- Kỳ vọng của X, kí hiệu E(X), là một định nghĩa:

- Tính chất của kỳ vọng: Cho là 2 biến ngẫu nhiên bất kỳ và thì

* Nếu và độc lập thì
  + 1. Phương sai - Var(X)

- Là bình phương của trung bình chênh lệch của giá trị thực x so với kỳ vọng, nó phản ánh mức độ phân tán các giá trị của biến ngẫu nhiên quanh giá trị cân bằng. Trong công nghiệp, phương sai biểu thị độ chính xác trong sản suất. Trong canh tác, phương sai biểu thị mức độ ổn định của năng suất, trong đo lường phương sai thể hiện trình độ “ổn định” của phép đo

- Phương sai của , ký hiệu là , được định nghĩa:

* Rời rạc:
* Liên tục:

- Độ lệch chuẩn:

- Tính chất của phương sai:

* Nếu X và Y độc lập:
  + 1. Mod

- Mod là giá trị x của X mà xảy ra nhiều nhất, theo đó:

* Rời rạc: Mod(X) là x0 mà có xác xuất xảy ra cao nhất
* Liên tục
  + 1. Trung vị - Med

- Nếu liên tục thì Med(X) là điểm x0 mà cắt đồ thị ra làm đôi:

- Trung vị - Med(X) = m với m là một phần tử của X mà

- VD: Biến ngẫu nhiên X có Med(X) = 2 và 3

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 |
|  | 0,1 | 0,4 | 0,3 | 0,2 |

1. Các phân phối xác suất thường gặp
   1. Biến ngẫu nhiên Bernouli (rr)

- Là 1 phép thử biến cố A với xác suất là p, khi A xảy ra thì biến ngẫu nhiên X nhận giá trị là 1, ngược lại nhận giá trị 0, ký hiệu X ~ B(1,p). Khi đó:

Bảng phân phối:

|  |  |  |
| --- | --- | --- |
|  | 1 | 0 |
|  |  |  |

Kỳ vọng E(X) = p

Phương sai Var(X) = pq

* 1. Phân phối nhị thức (rr) (Binomial Dis)

- Thực hiện n phép thử Bernouli có xác suất p với số lần thành công là , ký hiệu .

VD: Tung đồng xu 6 lần độc lập, gọi X là số lần xuất hiện mặt sấp, Xi là xuất hiện mặt tung mặt sấp lần tung i ()

Tổng quát:

Kỳ vọng:

Phương sai:

Giá trị tin chắc: Mod(X) ∈

* Ví dụ: phân phối nhị thức (hid - khá dễ)

Trong 1 gia đình có 6 người con. Tính xác xuất

Giải

Quan sát sinh con trai 6 lần: P(sinh con trai) = 0.5

Gọi X là số con trai trong 6 lần sinh 🡪

Ta có bảng phân phối

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|  | 0,016 | 0,094 | 0,234 | 0,312 | 0,234 | 0,094 | 0,016 |

(1) Xác suất để gia đình này có đúng 3 con trai

(2) Xác suất để gia đình này có nhiều nhất 2 con trai

(3) Xác suất để gia đình này có ít nhất 3 con trai

- Chứng minh: (hide)

Text

Description automatically generated

Text, letter

Description automatically generated

* 1. Phân phối siêu bội (rr) (Hypergeometric Dis)

- Ví dụ: Có 3 bi đỏ, 7 bi đen, lấy ngẫu nhiên không hoàn lại 5 bi cùng lúc. Xác suất trong số bi lấy ra có 1 bị đỏ là P(X = 1) = , khi đó ta kí hiệu:

- Tổng quát: Cho biến ngẫu nhiên rời rạc nhận các giá trị . Biến ngẫu nhiên có phân phối siêu bội, ký hiệu , hàm xác suất của X có dạng

Đặt thì:

- Lưu ý khi xét 1 số quá lớn và lấy ra 1 số nhỏ không hoàn lại cũng không ảnh hưởng nhiều xác xuất nhận được thì ta có thể sử dụng luôn pp nhị phân VD: Chart, scatter chart

Description automatically generated

* 1. Phân phối Poisson (Poisson Dis)

- Cho biến ngẫu nhiên rời rạc có phân phối Poisson với tham số , ký hiệu , khi đó hàm xác suất có dạng:

Kỳ vọng .

Phương sai .

- Ví dụ: Tại 1 đại lý bưu điện, các cuộc gọi đến xuất hiện ngẫu nhiên, độc lập với nhau và có tốc độ trung bình 1 cuộc gọi trong 0.5 phút. Biết rằng số cuộc gọi trong 1 khoảng thời gian cố định có phân phối Poisson. Tìm xác suất để:  
(1) có đúng 5 cuộc gọi trong 2 phút.  
(2) không có cuộc gọi nào trong khoảng thời gian 30 giây.  
(3) có ít nhất một cuộc gọi trong khoảng thời gian 10 giây.

Giải

Gọi là số cuộc gọi xuất hiện trong khoảng thời gian 2 phút. Ta có và do đó

Gọi là số cuộc gọi xuất hiện trong khoảng thời gian 30 giây. Ta có và do đó

Gọi là số cuộc gọi xuất hiện trong khoảng thời gian 10 giây. Ta có và do đó

* Xấp xỉ pp nhị thức bằng pp poisson

- Vấn đề: Khi tiêm truyền một loại huyết thanh, trung bình có một trường hợp phản ứng trên 1000 trường hợp. Dùng loại huyết thanh này tiêm cho 2000 người. Tính xác suất để:

Có 3 trường hợp phản ứng:

Có 100 trường hợp phản ứng: = (tràn số!)

🡪 Ta không thể tính bằng cách thông thường. Trong thực tế, khi , trong đó và thì ta có thể dùng xấp xỉ .

* 1. Phân phối đều (Uniform Dis)

- Một biến ngẫu nhiên liên tục X có phân phối đều trên [a;b], ký hiệu nếu hàm hàm mật độ xác suất X có dạng:

Hàm phân phối của X:

Kỳ vọng: =

Phương sai:

- VD: Một xe buýt xuất hiện tại bến đợi cứ 15 phút một chuyến. Một hành khách tới biến vào một thời điểm ngẫu nhiên. Gọi X là thời gian chờ xe của hành khách đó. Khi đó X có phân bố đều trên khoảng [0 ; 15]

Giải

* 1. Phân phối chuẩn (liên tục) (Normal Distribution)
* Chuẩn tắc

- Biến ngẫu nhiên được gọi là có phân phối chuẩn hóa (chuẩn tắc hay phân phối Gauss), ký hiệu , nếu hàm mật độ có dạng:

Biến ngẫu nhiên có kì vọng và phương sai

Đồ thị:

Chart, line chart, histogram

Description automatically generated

Hàm phân phối:

Tính chất hàm phân phối:

* (vì: nhìn hình ta có)
* Chuẩn thường

- Cho biến ngẫu nhiên liên tục, với kì vọng và phương sai là 2 tham số, có phân phối chuẩn, ký hiệu , khi đó hàm mật độ X có dạng:

Đồ thị hàm mật độ X:

Chart, line chart

Description automatically generated

Đồ thị hàm phân phối:

Chart, line chart

Description automatically generated

Nhận xét:

* Đồ thị dạng hình chuông với: E(X) (trung bình) Trung vị Median Mode
* Là phân phối đối xứng qua trục μ
* Độ phân tán (độ dạt ra 2 bên) được xác định bởi độ lệch chuẩn .

- Chuẩn hóa phân phối chuẩn thành phân phối chuẩn tắc:

🡪 Phương pháp biến đổi:

Nếu thì

- Phân vị chuẩn hóa: Cho biến ngẫu nhiên , phân vị chuẩn hóa mức , ký hiệu , là giá trị của biến ngẫu nhiên thỏa mãn điều kiện

Diagram

Description automatically generated

- Giá trị được cho nhiều nhất, nên nhớ: P(X ≤ 1.96) = 0.975

- Bấm máy tính: Mode 🡪 3 🡪 AC 🡪 Shift + 1 🡪 5 🡪 1

* Quy tắc k-sigma (hid)

- Quy tắc k-sigma: cho , xét

VD: Với k = 3, ta có quy tắc 3-sigma:

- Ý nghĩa quy tắc sigma: số k sigma chia đồ thị với tỉ lệ định trước 68; 96; 99,7

Chart, histogram

Description automatically generated

* Ví dụ

Chỉ số thông minh (IQ), được đo băng bài IQ Standford-Binet, có phân phối chuẩn trong một tổng thể nào đó. IQ trung bình là 100 điểm và độ lệch chuẩn là 16 điểm

Giải

🡪

Phần trăm số người trong tổng trên có   
(1) từ 140 trở lên:

(với Z = )

🡪

- Ví dụ:

Text, letter

Description automatically generated

* Định lý giới hạn trung tâm

- Nếu là các biến ngẫu nhiên độc lập có cùng dạng phân phối có kỳ vọng μ và phương sai σ2 . Xét . Khi đó:

* và phương sai

Khi (lớn hơn 30) , xét

* Xấp xỉ phân phối nhị thức thành phân phối chuẩn:

- Điều kiện:

* Xác xuất

- Hiệu chỉnh liên tục: vì phần phối nhị thức là phân phối rời rạc nên ta cần chỉnh nó về phân phối liên tục như sau:

- Cho , ta có và . Khi lớn, theo định lý giới hạn trung tâm phân phối của biến ngẫu nhiên được xấp xỉ bằng phân phối chuẩn , ký hiệu . Xác suất

* 1. Các phân phối không học (khó)

- Xem rõ hơn trong slide

* Phân phối gamma:
* Phân phối Chi bình phương:
* Phân phối student:
* Phân phối Fisher:
* Phân phối mũ

Graphical user interface, text, application, email

Description automatically generated

1. Biến ngẫu nhiên (bỏ)
   1. Rời rạc 2 chiều
      1. Hàm xác suất đồng thời

- Hàm xác suất đồng thời của Vector ngẫu nhiên (X,Y) ký hiệu là hàm thỏa:

- Bảng phân phối xác suất đồng thời:

Table

Description automatically generated

🡪 Hàm xác suất lề: (tổng khi có xi)

- Hàm xác suất lề:

- Bảng phân phối lề của biến ngẫu nhiên X:

A picture containing letter

Description automatically generated

Với:

- Cho X có hàm xác suất lề thì:

* + 1. Hàm xác suất có điều kiện

- Xét vecto ngẫu nhiên rời rạc , khi biết trước thì hàm xác suất có điều kiện của được xác định bởi công thức xác suất có điều kiện:

trong đó và .

- Hàm phân phối xác suất có điều kiện: hàm phân phối của biến ngẫu nhiên X khi biết trước Y nhận giá trị y được định nghĩa:

- Kỳ vọng có điều kiện: cho trước X = x, thì kỳ vọng có điều kiện của biến ngẫu nhiên Y là:

* + 1. Sự độc lập

- Hai biến ngẫu nhiên rời rạc X và Y được gọi là độc lập với nhau khi thỏa một trong các điều sau:

- Ví dụ BT:

Text, application

Description automatically generated

* 1. Liên tục 2 chiều
     1. Hàm xác suất đồng thời

- Hàm mật độ xác suất đồng thời của vector ngẫu nhiên (X,Y) khi là liên tục thỏa dk sau:

- Hàm phân phối xác suất đồng thời:

- Hàm phân phối lề: tương tự rời rạc

- Cho vector ngẫu nhiên (X,Y):

- Nếu là hàm hai biến và vecto ngẫu nhiên có phân phối đồng thời thì:

- Ví dụ BT:

Text

Description automatically generated

* + 1. Hàm xác suất có điều kiện

- Tương tự rời rạc

- Xét vecto ngẫu nhiên liên tục có hàm mật độ xác suất đồng thời , hàm phân phối xác suất có điều kiện của cho trước giá trị được xác định bởi

Tương tự, phân phối có điều kiện của cho trước giá trị là

- Kỳ vọng có điều kiện:

* + 1. Sự độc lập

- Tương tự rời rạc

- Hai biến ngẫu nhiên liên tục X và Y gọi là độc lập nhau nếu thỏa 1 trong các tính chất:

- Ví dụ BT:

Graphical user interface, text, application

Description automatically generated

* 1. Hiệp phương sai và hệ số tương quan
     1. Hiệp phương sai

- Cho hai biến ngẫu nhiên, hiệp phương sai của X và Y được xác định là:

- Ý nghĩa (xem trong slide | không học)

- Nếu hai biến độc lập và có Var hữu hạn thì: Cov(X,Y) = 0 và phương sai của X + Y là:

- Tổng phương sai của các biến ngẫu nhiên:

(tổng quát trong slide không học)

* + 1. Hệ số tương quan

- Hệ số tương quan của hai biến ngẫu nhiên được xác định:

- Ví dụ bài tập:

Graphical user interface, text, application, chat or text message

Description automatically generated

* 1. Sửa bài tập tổng quát

Text, letter

Description automatically generated

Text, letter

Description automatically generated

1. Thống kê mô tả
   1. Giới thiệu về thống kê

- Là các khái niệm về thống kê, xem đầy đủ trong slide

* 1. Mô tả dữ liệu bằng đồ thị

- Không thi, xem slide

* 1. Mô tả dữ liệu bằng phương pháp số

- Các thông số quan tâm gồm:

* Mean - trung bình số học
* Median - trung vị: xếp tăng dần, trung vị là index , nếu index là số không nguyên là trung bình 2 ô gần index nhất

🡪 Nếu mean < median thì lệch trái, nếu bằng thì đối xứng

* Miền giá trị (khoảng cách giữa giá trị lớn nhất và bé nhất)
* Tứ phân vị chia đồ thị hàm mật độ thành 4 khoảng bằng nhau (theo trục tung) gồm , Median, . (Interquantile range) IQR = Q3 - Q1
* - Outlier là điểm nằm ngoài , Extreme outlier là điểm nằm ngoài
* Phương sai tổng thể (σ2 tính như cũ) và phương sai mẫu:

- Xem qua đồ thị hộp: boxplot (không thi nhưng ứng dụng)

- Định lý Chybeshev: cho k lớn hơn 1 và tập n số đo. Khi đó có ít nhất \*100% các số đó nằm trong khoảng k độ lệch chuẩn

- Hệ số biến thiên là độ lệch chuẩn theo tỉ lệ phần trăm trung bình:

1. Phân phối mẫu
   1. Khái niệm

- Lấy các giá trị có kích thước n từ thực tế mà Xi là độc lập và tuân theo phân phối nhất định. Ta có các thông số thống kê:

* Trung bình mẫu:
* Phương sai mẫu:

Bấm máy: Nhập bảng số liệu (chình hiện frequency bằng Shift + Mode + ↓ + STAT) 🡪 Shift + 1 + 4 + 4 bấm chọn bình phương

* Độ lệch chuẩn mẫu:
* Miền giá trị:
  1. Các kết quả quan trọng (đã ẩn)
     1. Các TH tương đương của phân phối chuẩn

- Giả sử là 1 mẫu ngẫu nhiên được lấy từ 1 tổng thể có phân phối chuẩn với trung bình và phương sai . Khi đó

* và độc lập nhau.

- Chứng minh (đã ẩn)

* + 1. TH xác suất chưa biếtText, letter

       Description automatically generated

- Xét mẫu ngẫu nhiên lấy từ 1 phân phối có trung bình hữu hạn và phương sai . Ta có biên ngẫu nhiên và đều có phân phối **xấp xỉ** với phân phối chuẩn tắc . (trong thực tế khi n ≫ thì xấp xỉ xảy ra)

* + 1. Phân phối mẫu tỉ lệ

- Giả sử cần khảo sát đặc trưng của tổng thể, khảo sát phần tử và đặt

thu được mẫu ngẫu nhiên với , với là 1 tỷ lệ phần tử thỏa đặc trưng .

Khi đó, được gọi là tỷ lệ mẫu. Đây là 1 ước lượng cho tỷ lệ tổng thể . Hơn nữa

🡪 Xét mẫu ngẫu nhiên lấy từ 1 phân phối Bernoulli . Ta có các biến ngẫu nhiên và có phân phối xấp xỉ với phân phối chuẩn .

1. Khoảng tin cậy
   1. Giới thiệu

- Khoảng tin cậy cho biến ngẫu nhiên θ với độ tin cậy nếu

* 1. Khoảng tin cậy trung bình (đã ẩn)
     1. Có kích thước mẫu n ≥ 30 (hoặc n < 30 nhưng tuân theo pp chuẩn), σ2 đã biết

- Khi đó:

- Với độ tin cậy , ta có

với là phân vị mức của phân phối chuẩn hóa .

🡪 Vậy:

* Với mẫu ngẫu nhiên (mẫu thực nghiệm thì viết thành ), khoảng tin cậy cho tham số với độ tin cậy là
* Đại lượng được gọi là dung sai (sai số giới hạn) của khoảng tin cậy.
  + 1. Có kích thước mẫu n ≥ 30, σ2 chưa biết

- Khi đó:

* Ta ước lượng Var(X) là S2 để thay thế σ2, còn lại hoàn toàn tương tự trường hợp 1, công thức chỉ sửa chữ σ thành S
  + 1. n < 30, σ2 chưa biết, X tuân theo phân phối chuẩn

- Khi đó:

(có phân phối Student với n-1 bậc tự do)

- Với độ tin cậy 1 - σ, thì:

(với là phân vị mức của phân phối Student với bậc tự do)

* + 1. Tóm tắt

- Các bước tìm khoảng tin cậy cho trung bình:

* B1: tìm x̅ và phương sai mẫu s2
* B2: xác định trường hợp và tính phân vị và dung sai với
* n ≥ 30 (hoặc x < 30 nhưng tuân theo pp chuẩn) và σ2 đã biết, khoảng tin cậy:
* n ≥ 30, σ2 chưa biết, khoảng tin cậy giống (TH1) những thay σ bằng độ lệch chuẩn mẫu
* n < 30, σ2 chưa biết, X tuân theo phân phối mẫu, khoảng tin cậy giống (TH2) nhưng với
* B3: Kết luận: khoảng tin cậy cho trung bình của tổng thể là
  + 1. Ví dụ

- Ví dụ:

Graphical user interface, application, Word

Description automatically generated

Giải

Text

Description automatically generated

Text, letter

Description automatically generated

Nhận xét thêm:

* Tin chắc rằng 96% mức lương trung bình thuộc khoảng
  1. Khoảng tin cậy tỉ lệ

- Vấn đề đặt ra: cho tổng thể X, trong đó tỉ lệ các cá thể mang đặc tính A là p. Từ mẫu ngẫu nhiên (X1, X2, ..., Xn) hãy tìm khoảng tin cậy cho p với độ tin cậy 1 - σ

- Các bước thực hiện:

Lưu ý: Các bài so về thống kê tỷ lệ phải đặt biên

* B1: Tìm tỉ lệ mẫu: (Y là số phần tử thỏa tính chất A trong n phần tử khảo sát nên Y ∼ B(n,p))
* B2:
* B4:
* B5: Kết luận: khoảng tin cậy 100( 1 - α )% cho tỉ lệ tổng thể là

Graphical user interface

Description automatically generated

- Ví dụ:

Table

Description automatically generated

Giải

Text, letter

Description automatically generated

1. Kiểm định giả thuyết thống kê
   1. Bài toán kiểm định

- Trong bài toán kiểm định giả thuyết, giả thuyết cần được kiểm định là giả thuyết không () và mệnh đề đối với gọi là đối thuyết ()

- Khi nào đề cho thẳng chữ 'σ' thì mới biết là có σ nếu nói độ lệch chuẩn hay độ lệch tiêu chuẩn thì tức là chưa biết σ

- Các loại sai lầm:

Table

Description automatically generated

* 1. Kiểm định giả thuyết trường hợp 1 mẫu
     1. Kiểm định μ, TH1: Biết

- Các bước kiểm định cho trung bình:

* B1: Phát biểu giả thuyết và đối thuyết của bài toán
* B2: Tính:
* B3: Tính miền bác bỏ:

Text

Description automatically generated

* B4: Kết luận: Bác bỏ / Chưa đủ cơ sở để bác bỏ
* B3': Sử dụng p-value thay vì miền bác bỏ: Bác bỏ nếu . Xác định p-value như sau:

Text, table

Description automatically generated

- Ví dụ: Dây chuyền sản xuất kem đánh răng P/S được thiết kế để đóng những tuýt kem có **trọng lượng trung bình là 6 oz** (1 ounce = 28g). Một mẫu gồm **30 tuýt** kem được chọn ngẫu nhiên để kiểm tra định kỳ. Bộ phận điều khiển dây chuyền phải đảm bảo để trọng lượng trung bình mỗi tuýt kem là 6oz; **nếu nhiều hơn hay ít hơn, dây chuyền phải được điều chỉnh lại**. Giả sử **trung bình mẫu của 30 tuýt kem là 6, 1 oz** và độ lệch **tiêu** chuẩn của tổng thể là **σ = 0, 2 oz** Thực hiện kiểm định giả thuyết với **mức ý nghĩa 3%** để xác định xem dây chuyền sản xuất có vận hành tốt hay không?

Giải

Giả thuyết:

Giá trị thống kê kiểm định:

Miền bác bỏ của khi Với . Vậy ta bác bỏ khi:

Do nên ta bác bỏ .

Vậy với độ tin cậy 97% độ tin cậy, trọng lượng trung bình mỗi tuyết kem không bằng 6 oz

- Ví dụ 2: Metro EMS: Một bệnh viện tại trung tâm thành phố cung cấp dịch vụ cấp cứu tại nhà. Với khoảng **2**0 xe cấp cứu, **mục tiêu** của trung tâm là cung cấp dịch vụ cấp cứu trong khoảng **thời gian trung bình là 12 phút sau khi nhận được điện thoại** yêu cầu. Một mẫu ngẫu nhiên gồm thời gian đáp ứng khi có yêu cầu của **40 ca cấp cứu được chọn**. **Trung bình mẫu là 13, 25 phút.** Biết rằng độ lệch **tiêu chuẩn của tổng thể là σ = 3, 2 phút**. Giám đốc EMS muốn thực hiện 1 kiểm định, với mức **ý nghĩa 5%,** để xác định xem liệu thời gian 1 ca cấp cứu có **bé hơn** hoặc bằng 12 phút hay không?

Giải

Giả thuyết:

* Thời gian đáp ứng của dịch vụ cấp cứu đạt yêu cầu
* Thời gian đáp ứng của dịch vụ cấp cứu đạt không yêu cầu

Giá trị thống kê kiểm định:

Miền bác bỏ :

Vậy với độ tin cậy 95%, Metro EMS không đáp ứng được mục tiêu thời gian phục vụ khách hàng từ 12p trở xuống

* + 1. Kiểm định μ, TH2: không biết , mẫu lớn

- Tương tự TH1, chỉ thay S cho σ

- Ví dụ: Trạm cảnh sát giao thông trên đường cao tốc sẽ thực hiện việc bắn tốc độ định kỳ tại các địa điểm khác nhau để kiểm tra tốc độ các phương tiện giao thông. Một mẫu về tốc độ của các loại xe được chọn để thực hiện **kiểm định giả thuyết sau** . Những vị trí mà bác bỏ là những vị trí tốt nhất được chọn để đặt radar kiểm soát tốc độ. Tại địa điểm F, một **mẫu gồm tốc độ của 64 phương tiện** được bắn tốc độ ngẫu nhiên có trung **bình là 66, 2 mph và độ lệch tiêu chuẩn 4, 2 mph**. Sử dụng **α = 5%** để kiểm định giả thuyết.

Giải

Giả thuyết:

Giá trị kiểm định thống kê:

Bác bỏ khi:

Do

Vậy với 95% độ tin cậy. Địa điểm F là địa điểm tốt để đặt radar kiểm soát

* + 1. Kiểm định μ, TH3: Không biết , mẫu nhỏ < 30

- Tương tự TH1, thay độ lệch tiêu chuển bằng độ lệch chuẩn S, thay bằng

( tuân theo phân phối student với n-1 bậc tự do)

miền bác bỏ tương tự chỉ thay từ Norm sang Student

trong trường hợp cần xác dịnh p.value:

Text

Description automatically generated

* + 1. Kiểm định giả thuyết cho tỷ lệ

- Một số khái niệm:

* Để giải bài toán, ta cần kiểm tra phân phối chuẩn xấp xỉ phân phối nhị thức tốt cần có
* Ước lượng không chệch: (Y là số lần thực hiện biến cố)

- Các bước kiểm định:

Lưu ý: Các bài so về thống kê tỷ lệ phải đặt biên

* B0: Tính các thống số
* B1: Phát biểu giả thuyết không và đối thuyết
* B2:
* B3: Tính giá trị thống kê kiểm định:
* B4: Xác định miền bác bỏ: (giống TH1)
* B5: Kết luận: Bác bỏ / Chưa đủ cơ sở để bác bỏ

- Ví dụ: Trong kỳ nghỉ giáng sinh vào đầu năm mới, Cục An toàn giao thông đã thống kê được rằng có 500 người chết và 25000 người bị thương do các vụ tai nạn giao thông trên toàn quốc. Theo thông cáo của Cục ATGT thì khoảng **50% số vụ tai nạn có liên quan đến rượu bia.** **Khảo sát ngẫu nhiên 120 vụ** tai nạn thấy có **67 vụ do ảnh hưởng của rượu bia**. Sử dụng số liệu trên để kiểm định lời khẳng định của Cục ATGT với mức **ý nghĩa α = 5%**

Giải

Giả thuyết:

Giá trị thống kê kiểm định:

Ta bác bỏ khi

Do nên ta kết luận chưa đủ cơ sở để bác bỏ giả thuyết

Vậy với độ độ tin cậy 95% khoảng 50% số vụ tai nạn có liên quan đến rượu bia

* 1. Kiểm định giả thuyết trường hợp 2 mẫu độc lập
     1. So sánh hai μ, biết σ

- Các bước kiểm định:

* B1: Phát biểu giả thuyết và đối thuyết
* B2: Tính thống kê kiểm định
* B3: Xác định miền bác bỏ: (giống với TH1)
* B4: Kết luận: Nếu bác bỏ , ta kết luân đúng với Nếu không ta kết luận chưa đủ cơ sở để bác bỏ với α cho trước

- Ví dụ: Một công ty sản xuất sơn nghiên cứu về 1 loại phụ gia làm giảm thời gian khô của sơn. Thực hiện **thí nghiệm trên 2 mẫu**: **mẫu thứ nhất gồm 10 mẫu vật được sơn bằng loại sơn bình thường; mẫu thứ hai gồm 10 mẫu vật được sơn với sơn có chất phụ gia mới**. Trong những nghiên cứu trước, biết rằng **độ lệch tiêu chuẩn của thời gian khô sau khi quét sơn là 8 phút và không thay đổi khi thêm phụ gia vào**. Trung bình của mẫu 1 và mẫu 2 lần lượt là = 121 phút và = 112 phút. Với mức ý nghĩa 5%, hãy cho kết luận về loại sơn với chất phụ gia mới.

Giải

Giả thuyết

Giá trị thống kê kiểm định

Do

Vậy với độ tin cậy 95%. Chất phụ giá có hiệu quả làm giảm bớt thời gian khô sau khi sơn

* + 1. So sánh hai μ, không biết σ, mẫu lớn > 30

- Các bước kiểm định giống với khi biết σ ta chỉ thay σ bằng S

- Ví dụ: Khảo sát về chiều cao của sinh viên khoa Toán và CNTT: chọn ngẫu nhiên 50 sinh viên khoa Toán, tính được chiều cao trung bình là 163 cm và độ lệch chuẩn 5 cm. Đo chiều cao 50 sinh viên khoa CNTT, có trung bình mẫu là 166 cm và độ lệch chuẩn 8 cm. Với mức ý nghĩa α = 1%, hãy cho kết luận về chiều cao của sinh viên hai khoa.

Giải

Gọi X, Y lần lượt là mẫu ngẫu nhiên 50 sinh viên khoa Toán và CNTT

Cần kiểm định

Giá trị thống kê kiểm định:

Do

Vậy với độ tin cậy 99% ta kết luận chiều cao sinh viên 2 khoa là bằng nhau

* + 1. So sánh hai μ, không biết σ, mẫu nhỏ
* Trường hợp

- Các bước kiểm định:

* B1: Nêu giả thuyết và đối thuyết
* B2: Xác định phương sai mẫu chung
* B3: Đặt
* B4: Miền bác bỏ của :

|  |  |  |
| --- | --- | --- |
| Đối thuyết |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

* B5: Kết luận: bác bỏ / Chưa đủ cơ sở để bác bỏ
* Trường hợp

- Các bước kiểm định giống TH đầu nhưng giá trị thống kê:

và miền bác bỏ cũng có bậc tự do với:

(làm tròn như thông thường nếu số lẻ)

* Ví dụ

Tại một thành phố, ở **khu vực A**, người ta **chọn ngẫu nhiên 17 sinh viên** và cho làm 1 bài kiểm tra để cho chỉ số IQs, thu được **trung bình mẫu là 106 và độ lệch tiêu chuẩn bằng 10**; tại **khu vực B**, chỉ số IQs **trung bình của 1 mẫu gồm 14 sinh viên bằng 109 với độ lệch tiêu chuẩn là 7**. Giả sử **2 phương sai bằng** nhau. Có sự khác biệt về chỉ số IQs của sinh viên ở hai khu vực A và B hay không (nếu cho khác biệt bằng 7 hay không thì ta cho ) ? với α = 0, 02.

Giải

Gọi 17 sinh viên khu vực là mẫu 1 và là mẫu sinh viên khu vực là mẫu 2 và là mẫu   
Cần kiểm định:

Ta có phương sai mẫu chung:

Tính thống kê:

Ta có : .

Vậy với độ tin cậy 98% chưa đủ cơ sở để bác bỏ chỉ số IQs của sinh viên 2 khu vực là như nhau

* Ví dụ

Hàm lượng thạch tín (Asen) (Đv: ppb) trong nước càng cao càng có hại cho sức khỏe. Người ta kiểm tra hàm lượng thạch tín ở hai khu vực là trung tâm thành phố Biên Hòa và khu vực gần san bay Biên Hòa. Tại mỗi khu vực, người ta đo ngẫu nhiên hàm lượng thạch tín trong nước ứng với **10 địa điểm** khác nhau. Giả sử . Số liệu cho bởi bảng thống kê bên dưới

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Trung tâm TP | 3 | 7 | 25 | 10 | 15 | 6 | 12 | 25 | 15 | 7 |
| Khu vực gần sân bay | 48 | 44 | 40 | 38 | 33 | 21 | 20 | 12 | 1 | 18 |

Với α = 0, 05, hãy kiểm tra xem có sự khác biệt về hàm lượng thạch tín ở 2 khu vực này

Giải

Gọi mẫu ngẫu nhiên hàm lượng thạch tín trong 10 địa điểm ở TP và gần sân bay lần lượt là X và Y

Cần kiểm định

Do nên bác bỏ

Vậy với độ tin cậy 95% hàm lượng thạch tín trong nước ở hai khu vực là khác nhau

* + 1. So sánh hai tỷ lệ

- Các bước kiểm định:

Lưu ý: Các bài so về thống kê tỷ lệ phải đặt biên

* B0: Tính các thống số
* B1: Phát biểu giá thuyết và đối thuyết
* B2:
* B3: Tính thống kê

với

* B4: Xác định miền bác bỏ:
* B5: Nếu bác bỏ , ta thấy đúng với độ tin cậy, ngước lại kết luận chưa đủ cơ sở để bác bỏ
* Ví dụ:

Một công ty sản xuất thuốc cần kiểm tra 1 loại thuốc có tác dụng là giảm việc xuất hiện cơn đau ngực ở các bệnh nhân. Công ty thực hiện thí nghiệm trên 400 người, chia làm 2 nhóm: nhóm 1 gồm **200 được uống thuốc** và nhóm 2 gồm **200 người được uống giả dược**. Theo dõi thấy ở nhóm **1 có 8 người lên cơn đau ngực** và **nhóm 2 có 25 người lên cơn đau ngực**. Với **α = 0, 05**, hãy cho kết luận về hiệu quả của thuốc mới sản xuất.

Giải

Gọi X, Y là số người bị đau ngực ở nhóm 1 và nhóm 2

là tỷ lệ người bị đau ngực tổng thể nhóm 1 và nhóm 2

Cần kiểm định:

Thống kê:

Ta có . Vậy bác bỏ với mức ý nghĩa 95% thuốc mới sản xuất thật sự có hiệu quả

1. Hồi quy và tương quan
   1. Giới thiệu

- Phân tích hồi quy được sử dụng để xác định mối liên hệ giữa:

* Một hay nhiều biến độc lập (biến giải thích)
* Một biến phụ thuộc Y (biến đáp ứng)
* Sự thay đổi trong Y được giả sử do những thay đổi trong

🡪 Trên cơ sở sự thay đổi X ta dự báo được giá trị của Y

* 1. Mô hình hồi quy tuyến tính đơn
     1. Mô hình

- Mô hình thống kê tuyến tính đơn với biến ngẫu nhiên Y và biến giải thích x là pt có dạng:

trong đó:

* x là độc lập, giải thích cho y
* là thành phần sai số, được giả sử có phân phối chuẩn với và
* là các cặp quan trắc 1 mẫu ngẫu nhiên cỡ n được biểu diễn bởi các cặp giá trị quan trắc trong đồ thị phân tán (scatter plot)

Chart

Description automatically generated

🡪 Từ đồ thị: 🡪

🡪 Cần tìm đường thẳng hồi quy với các hệ số ước lượng . Đường thẳng ấy được xác định sao cho gần với các điểm dư liệu, ta dùng pp bình phương bé nhất

* + 1. Phương pháp bình phương bé nhất

- Từ định nghĩa, ta có: , Tổng bình phương sai số (**S**um of **S**quares for **E**rror - SSE) cho n điểm dữ liệu là

🡪 Ta cần tìm các ước lượng đạt giá trị bé nhất (ta đạo hàm theo để SSE' = 0 - slide 13 14 15 C7), Giải ra ta được:

Với

(tương tự ta có )

🡪 Khi đó đường thẳng gọi là đường thẳng BPBN, thỏa các tính chất sau

* đạt giá trị bé nhất.
* với là tổng các thặng dư (Sum of Error)

- Bấm máy tính: Mode 2 3 để nhập số liệu ; shift 1 3 để xem thông số ; shift 1 5 để kiểm tra lại phương trình hồi quy

- Đo độ biến thiên dữ liệu:

* : Tổng bình phương toàn phần (Total Sum of Squares): là độ biến thiên của xung quanh giá trị trung tâm dữ liệu
* : Tổng bình phương hồi quy (Regression Sum of Squares): là sự biến thiên liên quan đến mối liên hệ tuyến tính của X và Y
* SSE : Tổng bình phương sai số (Error Sum of Squares) là sự biến thiên của các nhân tố khác (không liên quan đến mối quan hệ tuyến tính)

🡪 Total = Regresion + Error

Diagram

Description automatically generated with medium confidence

- Hệ số xác định (coefficient of determination) là tỷ lệ sự biến thiên của các độc lập với tổng sự biến thiên:

nó cho ta đánh giá mô hình tìm được có giải thích tốt cho mối liên hệ phụ thuộc giữa Y và X hay không

Chart

Description automatically generated Chart, scatter chart

Description automatically generatedChart

Description automatically generated with medium confidence

- Hiệp tương quan (covariant)

🡪 Hệ số tương quan:

Hệ số tương quan mẫu r được xác định

(Cho R: res = resid(lm(y~x) = ))