

CHƯƠNG V: PHÂN TÍCH HỒI QUY VÀ TƯƠNG QUAN

I

SƠ LƯỢC HỒI
QUY VÀ
TƯƠNG QUAN

II

LIÊN HỆ
TƯƠNG QUAN
TUYẾN TÍNH
GIỮA HAI TIÊU
THỨC SỐ
LƯỢNG

III

LIÊN HỆ
TƯƠNG QUAN
PHI TUYẾN
GIỮA HAI TIÊU
THỨC SỐ
LƯỢNG

IV

LIÊN HỆ
TƯƠNG QUAN
TUYẾN TÍNH
GIỮA NHIỀU
TIÊU THỨC SỐ
LƯỢNG

I. SƠ LƯỢC HỒI QUY VÀ TƯƠNG QUAN

1

Thế nào là phân tích hồi quy tương quan?

2

Mối liên hệ của các hiện tượng kinh tế xã hội

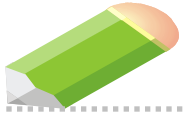
3

Nhiệm vụ phân tích hồi quy tương quan

4

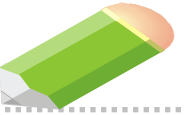
Các bước tiến hành xây dựng mô hình hồi quy

1. Thế nào là phân tích hồi quy tương quan?



Phân tích hồi quy

Nghiên cứu quan hệ giữa một biến phụ thuộc với một hoặc nhiều biến độc lập khác.



Phân tích tương quan

Nhằm đo mức độ tuyến tính giữa hai biến và các biến này có tính chất đối xứng.

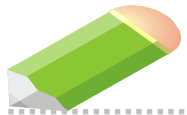


Phân tích hồi quy tương quan

Hồi quy tương quan thực chất là sử dụng phương pháp toán để phân tích mối liên hệ tương quan giữa các hiện tượng.

2. Mối liên hệ của các hiện tượng kinh tế xã hội

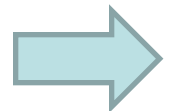
*Các hiện tượng KTXH luôn có mối liên hệ với nhau, các mối liên hệ này có thể diễn ra trong **thời gian** hoặc trong **không gian**.*



Liên hệ **hàm số**

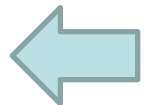


Liên hệ **tương quan**



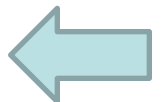
Liên hệ hàm số

- **Khái niệm:** mối liên hệ hoàn toàn chặt chẽ giữa tiêu thức nguyên nhân và tiêu thức kết quả. Sự thay đổi của tiêu thức nguyên nhân sẽ quyết định đến sự thay đổi của tiêu thức kết quả theo một quan hệ tỷ lệ.
- **Đặc điểm:** Liên hệ được biểu hiện trên từng đơn vị cá biệt



Liên hệ tương quan

- **Khái niệm:** mối liên hệ không hoàn toàn chặt chẽ giữa tiêu thức nguyên nhân và tiêu thức kết quả. Sự thay đổi của tiêu thức nguyên nhân có thể ảnh hưởng đến sự thay đổi của tiêu thức kết quả nhưng không có tính quyết định.
- **Đặc điểm:** Liên hệ không được biểu hiện trên từng đơn vị cá biệt mà phải quan sát số lớn.

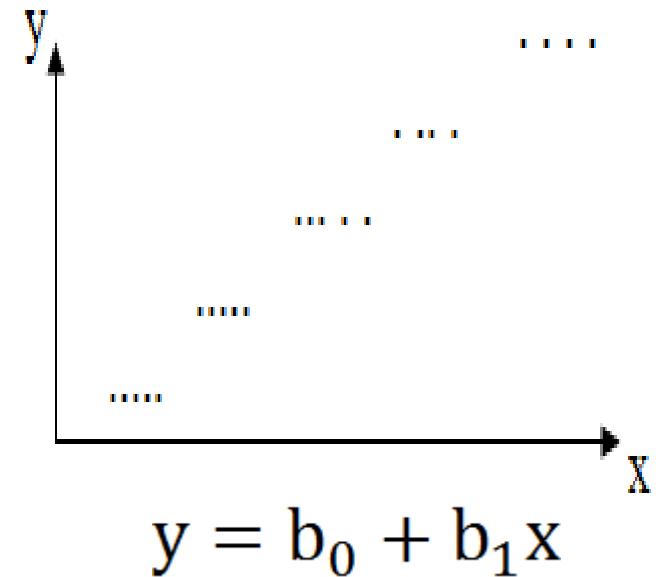


3. Nhiệm vụ phân tích hồi quy tương quan

- ✓ Xây dựng phương trình hồi quy và xác định tính chất (thuận, nghịch), hình thức của mối liên hệ.
- ✓ Đánh giá trình độ chặt chẽ của mối liên hệ tương quan và sự phù hợp của mô hình thông qua hai chỉ tiêu: hệ số t- ơng quan và tỷ số t- ơng quan.

4. Các bước tiến hành xây dựng mô hình hồi quy

- Giải thích sự tồn tại thực tế và bản chất của mối liên hệ bằng phân tích lý luận (xác định tiêu thức nguyên nhân và tiêu thức kết quả)
- Thăm dò mối liên hệ bằng các phương pháp thống kê.
- Lập phương trình hồi quy biểu hiện mối liên hệ.
- Tính các tham số và giải thích ý nghĩa.



$$y = b_0 + b_1 x + b_2 x^2$$

$$y = b_0 + \frac{b_1}{x}$$

$$y = b_0 \cdot b_1^x$$

II. Liên hệ tương quan tuyến tính giữa hai tiêu thức số lượng

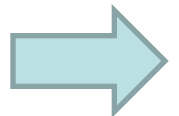
1

Xây dựng phương trình hồi quy

- a. Một số khái niệm cơ bản
- b. Cách xác định tham số: Phương pháp bình phương nhỏ nhất

2

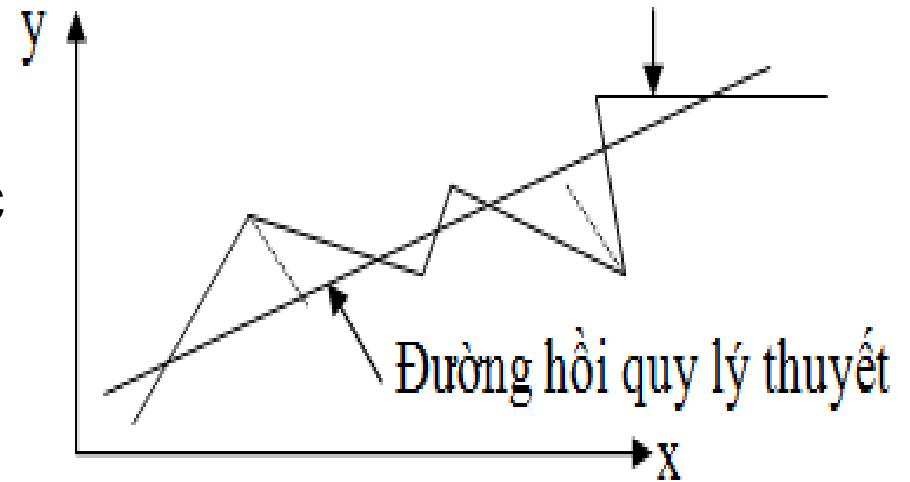
Hệ số tương quan tuyến tính (r)



1a. Một số khái niệm cơ bản

✓ *Đường hồi quy thực nghiệm*: đường được hình thành bởi các tài liệu thực tế.

✓ *Đường hồi quy lý thuyết*: đường điều chỉnh bù trừ các chênh lệch ngẫu nhiên, nêu ra xu hướng cơ bản của hiện tượng.



✓ *Phương trình hồi quy*: phương trình xác định vị trí của đường hồi quy lý thuyết sao cho mô tả gần đúng nhất mối liên hệ thực tế, gồm 2 loại: tuyến tính (đơn, bội) và phi tuyến.

MH

Phương trình hồi quy tuyến tính đơn

Tham số tự do (hệ số chặn)

Hệ số hồi quy (hệ số góc)

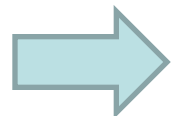
$$\hat{y}_x = b_0 + b_1 x$$

Giá trị lý
thuyết của

{
Biến phụ thuộc
Biến được giải thích
Dự báo
Kết quả

Biến độc lập
Biến giải thích
Công cụ dự báo
Nguyên nhân

Tại quan sát thứ i : $\hat{y}_i = b_0 + b_1 x_i$, $i = 1..N$



1b. Phương pháp bình phương nhỏ nhất (OLS)

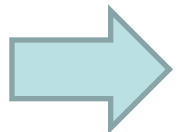
Ý tưởng: Tìm các tham số sao cho tổng bình phương các chênh lệch giữa giá trị thực tế và giá trị lý thuyết của tiêu thức kết quả là nhỏ nhất. Trong mô hình hồi quy tuyến tính đơn:

$$\begin{aligned} \text{Tìm } b_0, b_1 / S &= \sum_{i=1}^N (y_i - \hat{y}_i)^2 \min \Leftrightarrow S = \sum_{i=1}^N (y_i - b_0 - b_1 x_i)^2 \min \\ \Leftrightarrow \begin{cases} \frac{\partial S}{\partial b_0}(b_0) = \sum_{i=1}^N 2(y_i - b_0 - b_1 x_i)(-1) = 0 \\ \frac{\partial S}{\partial b_1}(b_1) = \sum_{i=1}^N 2(y_i - b_0 - b_1 x_i)(-x_i) = 0 \end{cases} &\Leftrightarrow \begin{cases} b_0 = \bar{Y} - b_1 \bar{X} \\ b_1 = \frac{\overline{XY} - \bar{Y}\bar{X}}{\overline{X^2} - \bar{X}^2} \end{cases} \\ &\quad \left(\bar{X} = \frac{\sum_{i=1}^N x_i}{N} \right) \end{aligned}$$

MH

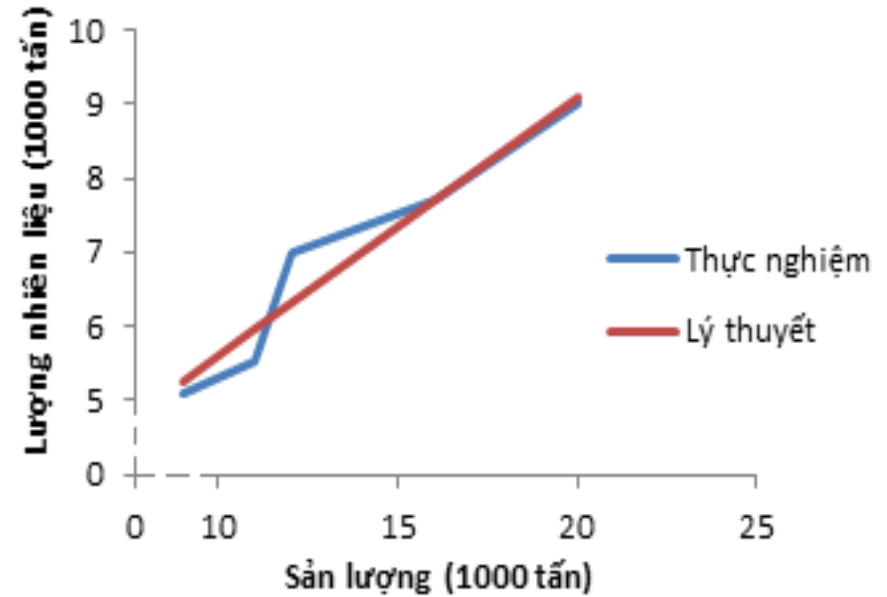
Ý nghĩa các tham số

- b_0 : phản ánh ảnh hưởng của các nguyên nhân khác (ngoài nguyên nhân x) tới kết quả y .
- b_1 : phản ánh ảnh hưởng trực tiếp của nguyên nhân x tới kết quả y . Cụ thể, khi x tăng thêm 1 đơn vị thì y thay đổi bình quân là b_1 đơn vị.
 - $b_1 > 0$: x và y có mối liên hệ thuận (cùng chiều)
 - $b_1 < 0$: x và y có mối liên hệ nghịch (ngược chiều)



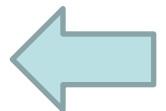
VD1: Lập phương trình hồi quy biểu diễn mối liên hệ giữa sản lượng và lượng nhiên liệu tiêu thụ (1000 tấn) cho tài liệu sau:

XN	SL x_i	NL y_i	x_i^2	$x_i y_i$
A	9	5,1	81	45,9
B	11	5,5	121	60,5
C	12	7,0	144	84
D	16	7,7	256	123,2
E	20	9,0	400	180
Σ	68	34,3	1002	493,6
TB	13,6	6,86	200,4	98,72



$$\text{PTHQ: } \widehat{NL} = b_0 + b_1 * SL = 2,082 + 0,3513 * SL$$

$$\begin{cases} b_0 = \bar{Y} - b_1 \bar{X} = 6,86 - 0,3513 * 13,6 = 2,082 \\ b_1 = \frac{\overline{XY} - \bar{Y}\bar{X}}{\overline{X^2} - \bar{X}^2} = \frac{98,72 - 13,6 * 6,86}{200,4 - 13,6^2} = 0,3513 \end{cases}$$



Sử dụng phương trình hồi quy vào dự báo

$$\widehat{NL} = 2,082 + 0,3513 * SL$$

☞ Muốn đạt được sản lượng 15000 tấn, lượng nhiên liệu tiêu thụ khoảng bao nhiêu?

$$\begin{aligned} SL = 15 \text{ (nghtấn)} &\Rightarrow \widehat{NL} = 2,082 + 0,3513 * 15 \\ &= 7,3515 \text{ (nghtấn)} \end{aligned}$$

☞ Giả sử một xí nghiệp còn lại 8000 tấn nhiên liệu, ước tính sản lượng mà xí nghiệp có thể đạt được?

$$\widehat{NL} = 8 \text{ (nghtấn)} \Rightarrow 2,082 + 0,3513 * SL = 8$$

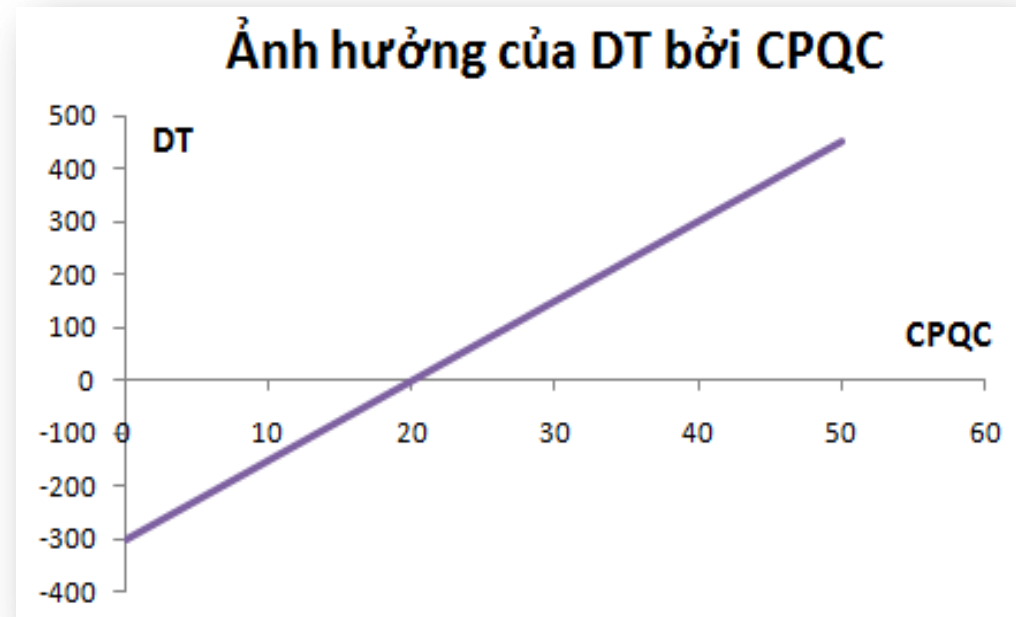
$$\Leftrightarrow SL = (8 - 2,082)/0,3513 = 16,846 \text{ (nghtấn)}$$

Chú ý

VD2: Cho phương trình hồi quy biểu hiện mối liên hệ giữa doanh thu và chi phí quảng cáo như sau:

$$\widehat{DT} = -300 + 15 * CPQC$$

→ chỉ khi $CPQC \geq 20$
thì CPQC mới có ảnh hưởng đến DT.



2. Hệ số tương quan tuyến tính (r)

Hệ số tương quan tuyến tính là chỉ tiêu được sử dụng để *xác định chiều hướng và đánh giá trình độ chặt chẽ* (cường độ) của mối liên hệ tương quan tuyến tính giữa hai tiêu thức số lượng.

$$r = \frac{\overline{XY} - \bar{X}\bar{Y}}{\sigma_X \sigma_Y} = b_1 \frac{\sigma_X}{\sigma_Y}$$

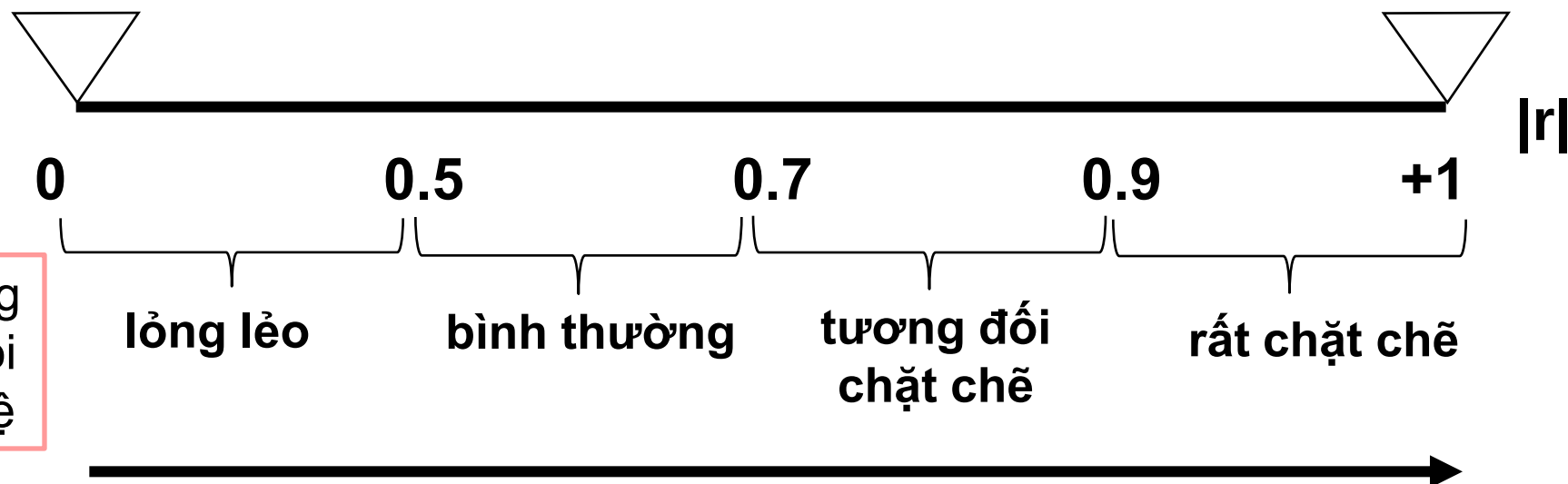
$$\sigma_X = \sqrt{\sigma_X^2} = \sqrt{\frac{\sum_{i=1}^N x_i^2}{N} - \left(\frac{\sum_{i=1}^N x_i}{N}\right)^2} = \sqrt{\overline{X^2} - \bar{X}^2};$$

$$\sigma_Y = \sqrt{\sigma_Y^2} = \sqrt{\frac{\sum_{i=1}^N y_i^2}{N} - \left(\frac{\sum_{i=1}^N y_i}{N}\right)^2} = \sqrt{\overline{Y^2} - \bar{Y}^2}$$

Tính chất của hệ số tương quan tuyến tính

Không có mối liên hệ
tương quan tuyến tính

Liên hệ hàm số,
hoàn toàn chặt chẽ

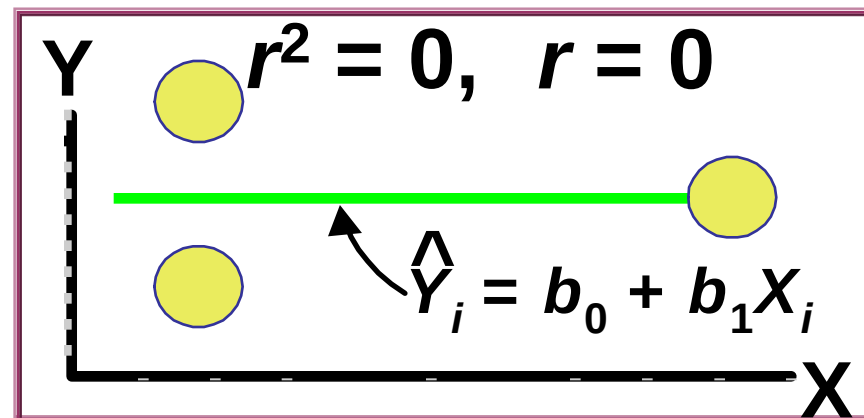
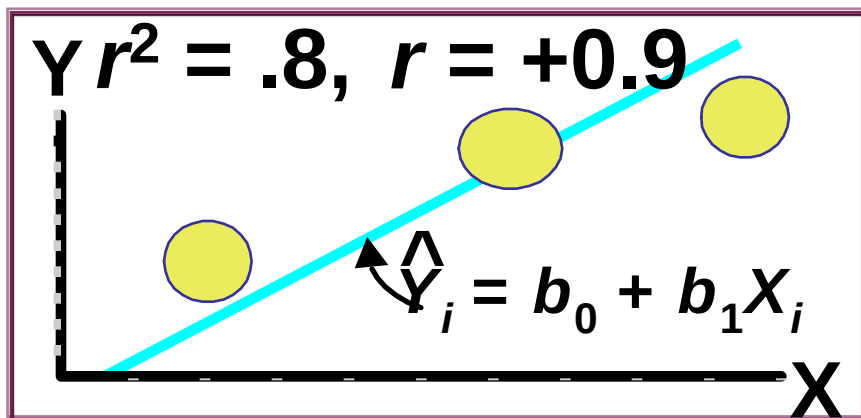
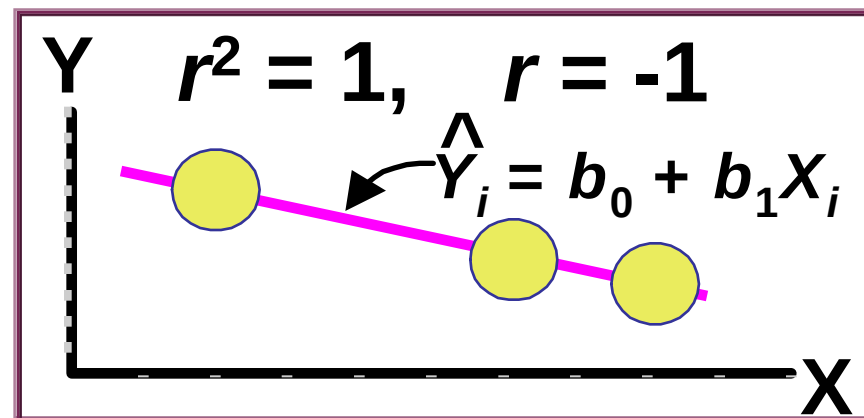
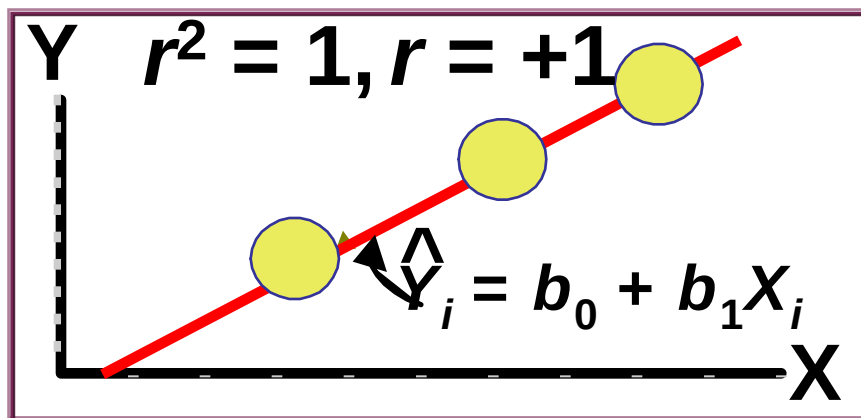


$r > 0$: mối liên hệ tương quan tuyến tính thuận

$r < 0$: mối liên hệ tương quan tuyến tính nghịch

Chiều hướng
mối liên hệ

Tính chất của hệ số tương quan tuyến tính



VD1: Đánh giá mức độ chặt chẽ của mối liên hệ giữa sản lượng và lượng nhiên liệu tiêu thụ (1000 tấn) cho tài liệu sau:

XN	SL x_i	NL y_i	x_i^2	$x_i y_i$	y_i^2
A	9	5,1	81	45,9	26,01
B	11	5,5	121	60,5	30,25
C	12	7,0	144	84	49
D	16	7,7	256	123,2	59,29
E	20	9,0	400	180	81
Σ	68	34,3	1002	493,6	245,55
TB	13,6	6,86	200,4	98,72	49,11

$$\begin{aligned}
 r &= \frac{\overline{XY} - \bar{X}\bar{Y}}{\sigma_X \sigma_Y} = b_1 \frac{\sigma_X}{\sigma_Y} \\
 &= \frac{98,72 - 13,6 * 6,86}{3,93 * 1,43} \\
 &= 0,3513 * \frac{3,93}{1,43} \\
 &= 0.965
 \end{aligned}$$

$$\sigma_X = \sqrt{\overline{X^2} - \bar{X}^2} = \sqrt{200,4 - (13,6)^2} = 3,93$$

$$\sigma_Y = \sqrt{\overline{Y^2} - \bar{Y}^2} = \sqrt{49,11 - (6,86)^2} = 1,43$$

Mối liên hệ
tuyến tính thuận
rất chặt chẽ.

Sự phù hợp của mô hình hồi quy tuyến tính đơn

- r^2 càng lớn, mô hình càng phù hợp
- r^2 cho biết tỷ lệ % thay đổi của biến phụ thuộc được giải thích bởi mô hình.

Ở VD1: $r^2 = 0.965^2 = 0.931 \rightarrow$ mô hình phù hợp.

III. Liên hệ tương quan phi tuyến giữa hai tiêu thức số lượng

1

Mối liên hệ tương quan phi tuyến

Là mối liên hệ tương quan giữa các tiêu thức không được biểu hiện bằng một đường thẳng mà bằng các đường cong có hình dáng khác nhau.

2

Một số dạng phương trình hồi quy thường gặp

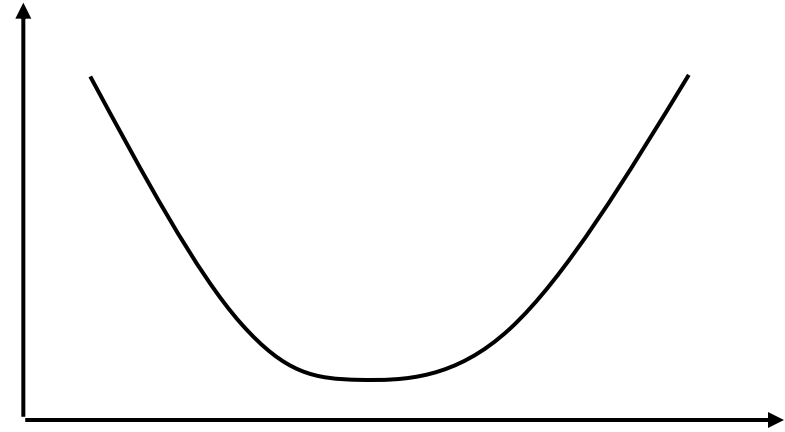
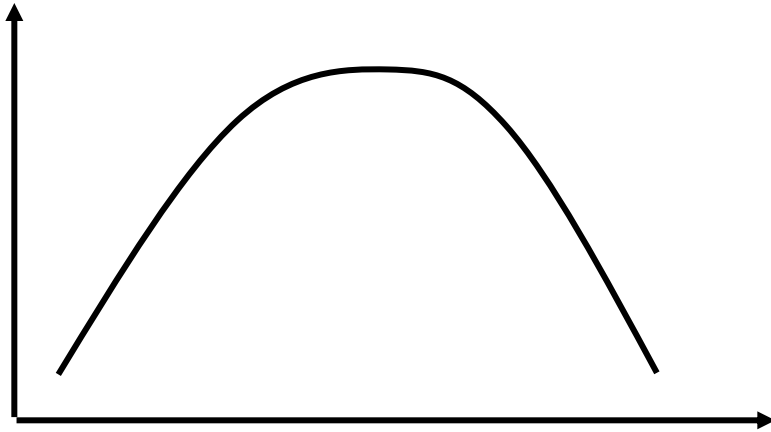
3

Tỷ số tương quan

2. Một số dạng phương trình hồi quy thường gặp

- Phương trình Parabol:

$$\hat{y}_x = b_0 + b_1x + b_2x^2$$



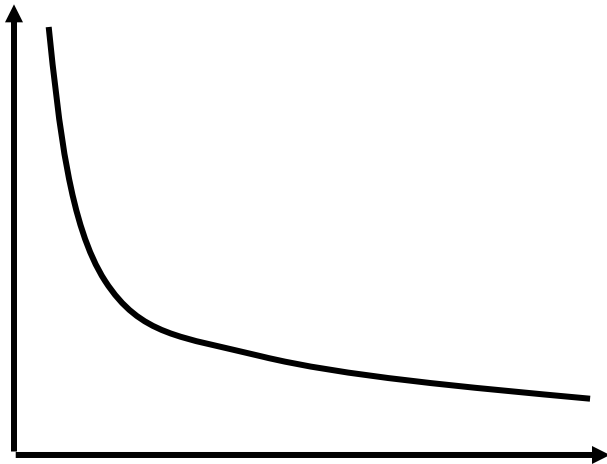
Các tham số là nghiệm của hệ:

$$\begin{cases} \sum y_i = Nb_0 + b_1 \sum x_i + b_2 \sum x_i^2 \\ \sum x_i y_i = b_0 \sum x_i + b_1 \sum x_i^2 + b_2 \sum x_i^3 \\ \sum x_i^2 y_i = b_0 \sum x_i^2 + b_1 \sum x_i^3 + b_2 \sum x_i^4 \end{cases}$$

2. Một số dạng phương trình hồi quy thường gặp

- Phương trình Hypebol:

$$\hat{y}_x = b_0 + \frac{b_1}{x}$$



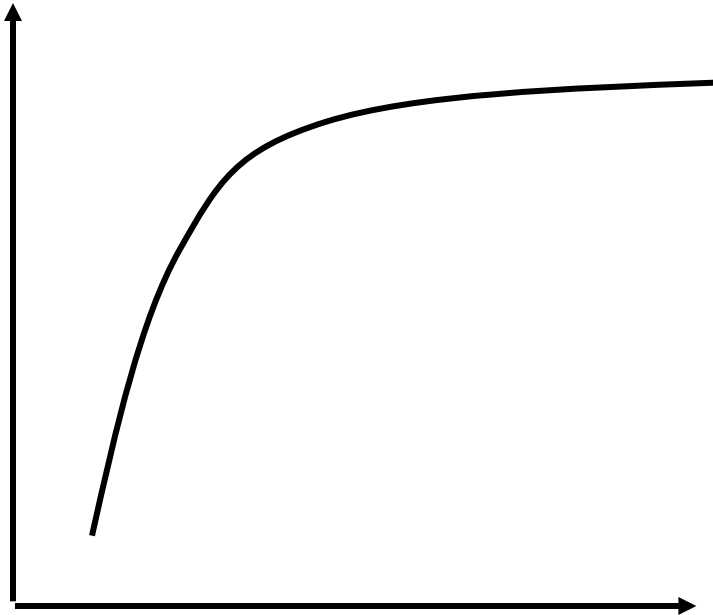
Các tham số là nghiệm của hệ:

$$\begin{cases} \Sigma y_i = Nb_0 + b_1 \Sigma \frac{1}{x_i} \\ \Sigma \frac{1}{x_i} y_i = b_0 \Sigma \frac{1}{x_i} + b_1 \Sigma \frac{1}{x_i^2} \end{cases}$$

2. Một số dạng phương trình hồi quy thường gặp

- Phương trình hàm mũ:

$$\hat{y}_x = b_0 \cdot b_1^x$$



Các tham số là nghiệm của hệ:

$$\begin{cases} \sum \ln y_i = N \ln b_0 + \ln b_1 \sum x_i \\ \sum x \ln y_i = \ln b_0 \sum x_i + \ln b_1 \sum x_i^2 \end{cases}$$

3. Tỷ số tương quan (η)

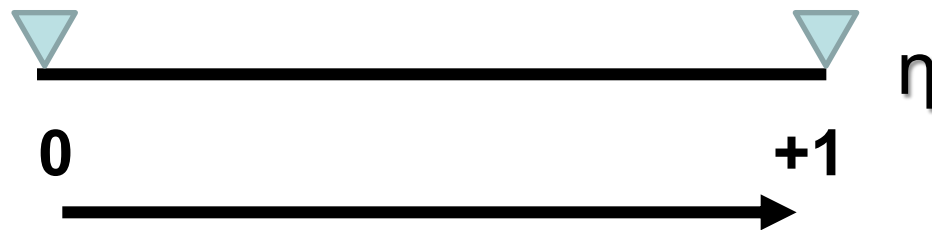
Tỷ số tương quan là chỉ tiêu đánh giá trình độ chặt chẽ của mối liên hệ tương quan phi tuyến.

$$\eta = \sqrt{1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}} = \sqrt{1 - \frac{RSS}{TSS}}$$

Tính chất:

Không có mối liên hệ

Mối liên hệ hoàn
toàn chặt chẽ



Mối liên hệ càng chặt chẽ

So sánh r và η

	r	η
Dạng mlhệ	Tuyến tính	Phi tuyến (tuyến tính)
Miền xác định	$[-1,1]$	$[0,1]$
Phạm vi nghiên cứu	Chiều hướng và cường độ mối liên hệ	Cường độ mối liên hệ
Khả năng thay thế	Không thay được cho η	Thay được cho r trong trường hợp $r \geq 0$

IV. Liên hệ tương quan tuyến tính giữa nhiều tiêu thức số lượng

1

Mô hình hồi quy tuyến tính bội

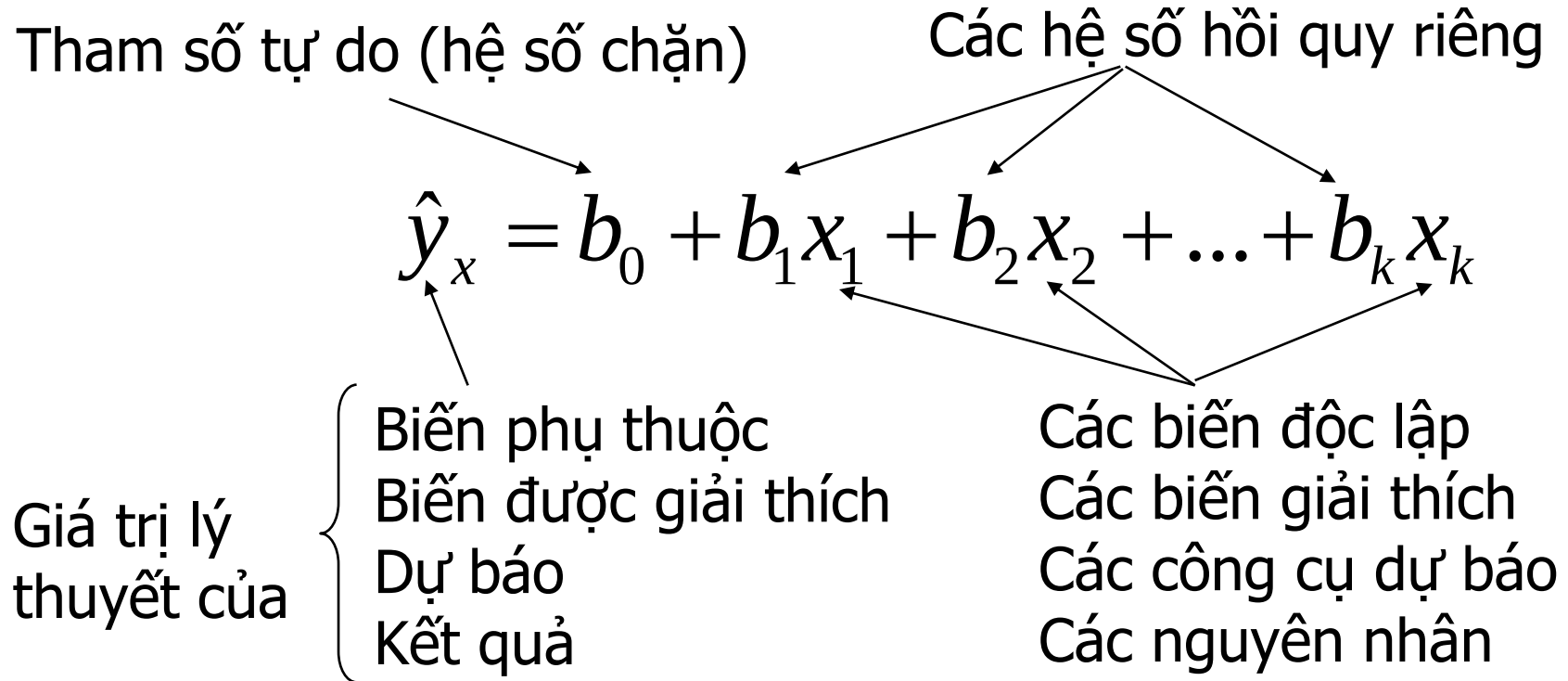
a. Mô hình

b. Cách xác định tham số: Phương pháp OLS

2

Hệ số tương quan tuyến tính bội (R)

1a. Mô hình hồi quy tuyến tính bội



Tại quan sát thứ i : $\hat{y}_i = b_0 + b_1x_{1i} + b_2x_{2i} + \dots + b_kx_{ki}$, $i = 1..N$

1b. Cách xác định tham số

$b_0, b_1, \dots, b_k$ là nghiệm của hệ phương trình:

$$\left\{ \begin{array}{l} \sum_{i=1}^N y_i = Nb_0 + b_1 \sum_{i=0}^N x_{1i} + b_2 \sum_{i=0}^N x_{2i} + \dots + b_k \sum_{i=0}^N x_{ki} \\ \sum_{i=0}^N x_{1i} y_i = b_0 \sum_{i=0}^N x_{1i} + b_1 \sum_{i=0}^N x_{1i}^2 + b_2 \sum_{i=0}^N x_{1i} x_{2i} + \dots + b_k \sum_{i=0}^N x_{1i} x_{ki} \\ \dots \\ \dots \\ \sum_{i=0}^N x_{ki} y_i = b_0 \sum_{i=0}^N x_{ki} + b_1 \sum_{i=0}^N x_{1i} x_{ki} + b_2 \sum_{i=0}^N x_{2i} x_{ki} + \dots + b_k \sum_{i=0}^N x_{ki}^2 \end{array} \right.$$

Dấu của b_i nói lên chiều hướng ảnh hưởng của tiêu thức nguyên nhân X_i tới tiêu thức kết quả Y .

2. Hệ số tương quan bội (R)

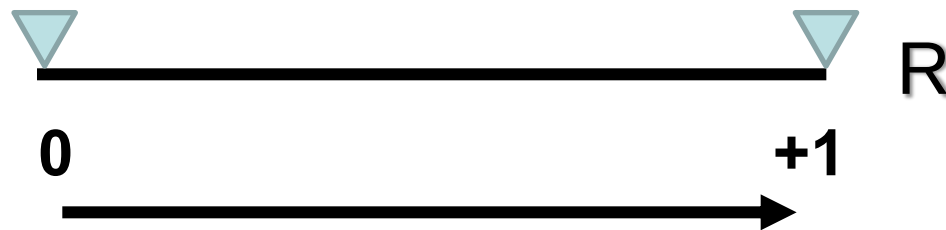
Hệ số tương quan bội là chỉ tiêu dùng để đánh giá mức độ chặt chẽ của mối liên hệ tương quan giữa tiêu thức kết quả y với tất cả các tiêu thức nguyên nhân được nghiên cứu.

$$R = \sqrt{1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}}$$

Tính chất:

Không có mối liên hệ

Liên hệ hàm số,
hoàn toàn chặt chẽ



Mối liên hệ càng chặt chẽ

Hệ số hồi quy chuẩn hóa (β)

Hệ số hồi quy chuẩn hóa là chỉ tiêu được sử dụng để đánh giá mức độ ảnh hưởng của từng tiêu thức nguyên nhân X_i đến tiêu thức kết quả Y . $|\beta_i|$ càng lớn, mức độ ảnh hưởng càng nhiều.

$$\beta_i = b_i \frac{\sigma_{X_i}}{\sigma_Y}$$