

BÀI 8

TƯƠNG QUAN VÀ HỒI QUY

TS. Nguyễn Mạnh Thế

TÌNH HUỐNG KHỞI ĐỘNG BÀI

Tình huống:

- Siêu thị ABC muốn mở 01 siêu thị tại khu dân cư Vạn Phúc. Để xác định được quy mô siêu thị, doanh nghiệp cần biết được chi phí nhu yếu phẩm của người dân trong vùng. Biết chi phí nhu yếu phẩm của 01 cá nhân phụ thuộc chính vào mức thu nhập của cá nhân đó. Siêu thị tiến hành điều tra mức thu nhập (X) và chi tiêu (Y) cho những nhu yếu phẩm của cá nhân.
- Kết quả cho bảng số liệu sau (Đơn vị: Triệu đồng):

X\Y	0,5	0,8	1,0
1,5	4	3	0
2,0	6	2	1
2,5	2	5	2
3,0	1	1	4

Câu hỏi gợi mở:

Câu 1: Tính hệ số tương quan mẫu.

Câu 2: Viết phương trình hồi quy tuyến tính mẫu.

Câu 3: Ước lượng sai số hồi quy.

Câu 4: Dự báo giá trị của Y khi mức thu nhập X là 4,0 triệu đồng.

TÌNH HUỐNG KHỞI ĐỘNG BÀI (tiếp theo)

Kết luận:

Với mẫu ngẫu nhiên $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ ta xây dựng hàm hồi quy mẫu (hàm hồi quy thực nghiệm) có dạng:

$$\hat{y} = \hat{a}x + \hat{b}$$

Với:

$$\hat{b} = \bar{y} - \hat{a}\bar{x}$$

$$\hat{a} = \frac{\sum_{f=1}^n x_f y_f - (\bar{x})(\bar{y})}{\frac{1}{n} \sum_{f=1}^n x_f^2 - (\bar{x})^2} = r \frac{s_y}{s_x}$$

Trong đó:

$$r = \frac{\overline{xy} - (\bar{x})(\bar{y})}{s_x s_y}$$

Phương trình hồi quy mẫu (đường hồi quy trung bình tuyến tính thực nghiệm) có dạng:

$$\hat{y} = r \frac{s_y}{s_x} x + \bar{y} - r \frac{s_y}{s_x} \bar{x} = \hat{a}x + \hat{b}$$

Trong đó:

$$r = \frac{\overline{xy} - (\bar{x})(\bar{y})}{s_x s_y}$$

NỘI DUNG

- Hệ số tương quan mẫu;
- Đường hồi quy bình phương trung bình tuyến tính thực nghiệm của hai biến ngẫu nhiên.



1. HỆ SỐ TƯƠNG QUAN MẪU

- Hệ số tương quan của hai biến ngẫu nhiên X, Y , được xác định bởi công thức:

$$\rho = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E(XY) - (EX)E(Y)}{\sigma_X \sigma_Y}$$

- Hệ số tương quan đo mức độ phụ thuộc tuyến tính giữa hai biến ngẫu nhiên X và Y .

Giả sử ta có mẫu ngẫu nhiên $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ rút ra từ véc tơ ngẫu nhiên (X, Y) với giá trị mẫu $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$.

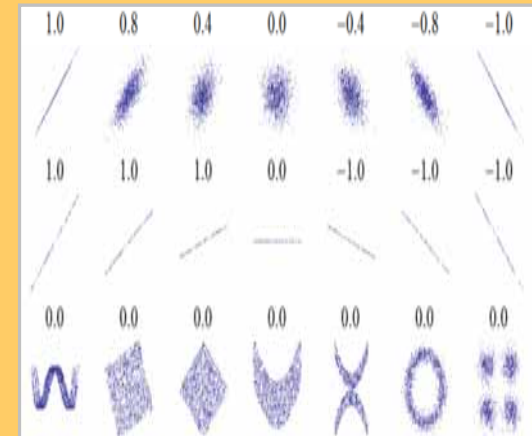
Hệ số tương quan mẫu:

$$R = \frac{\overline{XY} - (\bar{X})(\bar{Y})}{S_x S_y}$$

Trong đó thống kê:

$$\overline{XY} = \frac{1}{n} \sum_{k=1}^n x_k y_k$$

R là ước lượng của hệ số tương quan lý thuyết ρ



2. ĐƯỜNG HỒI QUY TRUNG BÌNH TUYẾN TÍNH THỰC NGHIỆM

Với mẫu cụ thể giá trị của R là: $r = \frac{\overline{xy} - (\bar{x})(\bar{y})}{s_x s_y}$

$|\rho|$ càng gần 1 thì mức độ phụ thuộc tuyến tính giữa X và Y càng chặt chẽ, tức là ta có thể xấp xỉ Y theo X dưới dạng $f(x) = ax + b$;

Thông thường khi $|\rho| > 0,8$ thì xấp xỉ được gọi là chặt chẽ, và quan hệ giữa X và Y được thể hiện: $y = ax + b + \varepsilon$

ε là sai số ngẫu nhiên.

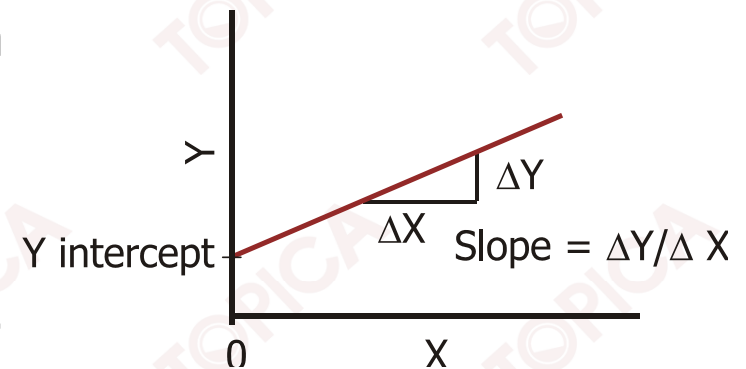


2. ĐƯỜNG HỒI QUY TRUNG BÌNH TUYẾN TÍNH THỰC NGHIỆM (tiếp theo)

Phương trình $y = ax + b$ được gọi là phương trình hồi quy tuyến tính của Y theo X.

Trong đó:

- Y được gọi là biến được giải thích hay biến phụ thuộc, được gọi là biến giải thích;
- X được gọi là biến giải thích;
 ε là sai số hồi quy;
- a được gọi là độ dốc (slope), cho biết khi biến X tăng một đơn vị thì giá trị của biến Y sẽ tăng hay giảm bao nhiêu đơn vị;
- b được gọi là hệ số chặn hoặc hệ số cắt (intercept), cho biết phương trình hồi quy có đi qua gốc tọa độ hay không và điểm xuất phát của Y khi X bằng 0 sẽ là bao nhiêu;
- a và b cũng được gọi là hệ số hồi quy.



2. ĐƯỜNG HỒI QUY TRUNG BÌNH TUYẾN TÍNH THỰC NGHIỆM (tiếp theo)

Với mẫu ngẫu nhiên $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ ta xây dựng hàm hồi quy mẫu (hàm hồi quy thực nghiệm) có dạng:

$$\hat{y} = \hat{a}x + \hat{b}$$

Tại quan sát thứ i ta có:

$$\hat{y}_i = \hat{a}x_i + \hat{b}$$

$$y_i = \hat{y}_i + \varepsilon_i = \hat{a}.x_i + \hat{b} + \varepsilon_i$$

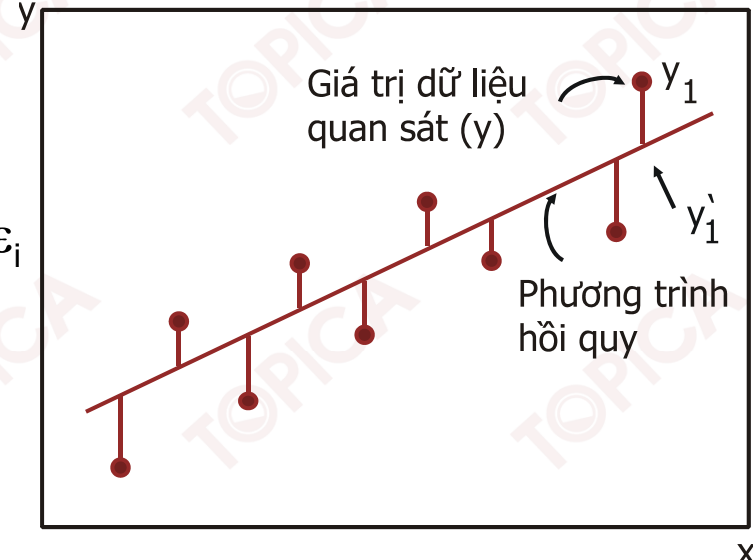
ε_i giá trị của ε tại quan sát thứ i ;

\hat{a}, \hat{b} là các ước lượng của a và b ;

Xác định các hệ số \hat{a}, \hat{b} sao cho:

$$L = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (\hat{y}_i - \hat{a}.x_i - \hat{b})^2 \rightarrow \min$$

$$\Rightarrow \begin{cases} \hat{b} = \bar{y} - \hat{a}.\bar{x} \\ \hat{a} = \frac{\sum_{i=1}^n x_i y_i - (\bar{x})(\bar{y})}{\frac{1}{n} \sum_{i=1}^n x_i^2 - (\bar{x})^2} = r \frac{s_y}{s_x} \end{cases}$$



2. ĐƯỜNG HỒI QUY TRUNG BÌNH TUYẾN TÍNH THỰC NGHIỆM (tiếp theo)

Tính chất:

- Hàm hồi quy mẫu đi qua điểm (\bar{x}, \bar{y}) ;
- Các ước lượng \hat{a} , \hat{b} được xác định duy nhất;
- Giá trị trung bình các sai số:
$$\bar{\varepsilon} = \frac{1}{n} \sum_{i=1}^n \varepsilon_i = 0$$
- Giá trị trung bình của \hat{y}_i bằng giá trị các quan sát y_i ; $\bar{\hat{y}} = \bar{y}$
- Sai số $\varepsilon_i = y_i - \hat{y}_i$ cũng được gọi là phần dư (residual) biểu thị sự sai khác giữa quan sát y_i và giá trị \hat{y}_i

Đặt $RSS = \sum_{i=1}^n \varepsilon_i^2$ được gọi là tổng bình phương các phần dư $RSS = ns_y^2(1 - r^2)$

Ký hiệu $\varepsilon_{y/x}^2$ là ước lượng của sai số ε

$$\varepsilon_{y/x}^2 = \frac{1}{n} \sum_{i=1}^n \varepsilon_i^2 - (\bar{\varepsilon})^2 = \frac{RSS}{n} = s_y^2(1 - r^2)$$

- Ứng dụng quan trọng của hồi quy là dự báo giá trị của biến Y khi biến X nhận giá trị mới. Nếu biến X nhận giá trị mới là x_0 khi đó ta có: $\hat{y}_0 = \hat{a}x_0 + \hat{b}$
 \hat{y}_0 là ước lượng điểm tương ứng cho giá trị y_0 của Y.

2. ĐƯỜNG HỒI QUY TRUNG BÌNH TUYẾN TÍNH THỰC NGHIỆM (tiếp theo)

Ví dụ 1: Theo dõi mức lãi suất (Y) và tỷ lệ lạm phát (X) ở một số nước ta có số liệu sau:

Y	17,5	15,6	9,8	5,3	7,9	10,0	19,2	13,1
X	14,2	11,7	6,4	2,1	4,8	8,1	15,4	9,8

Lời giải:

d) Nếu tỷ lệ lạm phát $x_0 = 22,5$ thì mức lãi suất ngân hàng sẽ là:

$$y_0 = 1,045.22,5 + 2,83 = 26,343$$

2. ĐƯỜNG HỒI QUY TRUNG BÌNH TUYẾN TÍNH THỰC NGHIỆM (tiếp theo)

Ví dụ 2: Điều tra mức thu nhập (X) và chi tiêu (Y) cho những nhu cầu yếu phẩm của cá nhân ta có bảng số liệu sau (Đơn vị: Triệu đồng):

Y \ X	0,5	0,8	1,0
1,5	4	3	0
2,0	6	2	1
2,5	2	5	2
3,0	1	1	4

Lời giải:

d. Dự báo giá trị của Y khi $X = x_0 = 4,0$ ta có: $y_0 = 0,33.4,0 - 0,016 = 1,3$

**Articulate Quizmaker Quiz
Placeholder -
XSTK_chuong
8_Quiz_3_12_08**

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Select a term:

Biến được giải thích

Biến giải thích

Độ dốc (slope)

Đường hồi quy trung bình tuy...

Hằng số hồi quy (intercept)

Hệ số hồi quy

Hệ số tương quan mẫu

Phần dư (residual)

RSS

Sai số hồi quy

Ước lượng của sai số

Biến được giải thích

Biến được giải thích: Biến Y được gọi là biến được giải thích hay biến phụ thuộc.

PROPERTIES

Allow user to leave interaction:

Anytime

Show 'Next Slide' Button:

Don't show

Completion Button Label:

Next Slide



Properties...



Edit in Engage