

# MÔ HÌNH HỒI QUY NHIỀU BIẾN

I

MÔ HÌNH HỒI  
QUY NHIỀU  
BIẾN

II

PHƯƠNG PHÁP  
BÌNH PHƯƠNG  
NHỎ NHẤT

III

KHOẢNG TIN CẬY  
VÀ KIỂM ĐỊNH  
GIẢ THUYẾT

IV

PHÂN TÍCH PHƯƠNG  
SAI VÀ KIỂM ĐỊNH GIẢ  
THUYẾT ĐỒNG THỜI

V

PHÂN TÍCH HỒI  
QUY VÀ DỰ BÁO

# I. MÔ HÌNH HỒI QUY NHIỀU BIẾN

## I.1 *Mô hình hồi quy tổng thể*

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + U_i$$

Trong đó:

$Y_i$  : giá trị của biến phụ thuộc Y ( $i = 1..n$ );

$\beta_1$  : hệ số chặn (hệ số tự do);

$\beta_j$ : hệ số góc (hệ số hồi quy riêng) của biến giải thích thứ j ( $j = 2..k$ );

$U_i$  : sai số ngẫu nhiên

## I.2

## Mô hình hồi quy mẫu

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \dots + \hat{\beta}_k X_{ki}$$

Được xây dựng trên mẫu ngẫu nhiên gồm n quan sát:

$$\{(Y_i, X_{2i}, X_{3i}, \dots, X_{ki}), i = 1..n\}$$

Trong đó:

$\hat{Y}_i$  : ước lượng của  $Y_i$  hoặc  $E(Y/X_i)$  ( $i = 1..n$ )

$\hat{\beta}_j$  : ước lượng của các hệ số hồi quy tổng thể ( $j=2..k$ )

### I.3

## Dạng ma trận của mô hình hồi quy nhiều biến

Ta kí  
hiệu  
các  
ma  
trận:

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_n \end{pmatrix} \quad \hat{Y} = \begin{pmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \dots \\ \hat{Y}_n \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_k \end{pmatrix} \quad \hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \dots \\ \hat{\beta}_k \end{pmatrix}$$

$$U = \begin{pmatrix} U_1 \\ U_2 \\ \dots \\ U_n \end{pmatrix} \quad X = \begin{pmatrix} 1 & X_{21} & X_{31} & \dots & X_{k1} \\ 1 & X_{22} & X_{32} & \dots & X_{k2} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & X_{2n} & X_{3n} & \dots & X_{kn} \end{pmatrix}$$

Hàm hồi quy tổng thể có dạng:  $Y = X\beta + U$

Hàm hồi quy mẫu có dạng:  $\hat{Y} = X\hat{\beta}$

## I.4

### *Các giả thiết của mô hình hồi quy nhiều biến*

**Giả thiết 1.** Các biến giải thích  $X_j$  ( $j=2..k$ ) là phi ngẫu nhiên, giá trị của nó là xác định, đã biết.

**Giả thiết 2.** Trung bình ảnh hưởng các yếu tố ngẫu nhiên đến biến phụ thuộc bằng 0:  $E(U_i) = E(U|X_{2i}, X_{3i}, \dots, X_{ki}) = 0 \forall i$

**Giả thiết 3.** Các yếu tố ngẫu nhiên có phương sai bằng nhau (phương sai thuần nhất):  $Var(U_i) = Var(U|X_{2i}, X_{3i}, \dots, X_{ki}) = \sigma^2 \forall i$

**Giả thiết 4.** Các  $U_i$  không tương quan :  $Cov(U_i, U_j) = 0 \forall i \neq j$

**Giả thiết 5.** Các  $X_j$  ( $j=2..k$ ) độc lập tuyến tính với nhau  $\leftrightarrow$   
 $\text{rank}(X)=k$

**Giả thiết 6.**  $U_i \sim N(0, \sigma^2) \forall i$

## II. PHƯƠNG PHÁP BÌNH PHƯƠNG NHỎ NHẤT

### II.1 Nội dung phương pháp

Xác định các  $\hat{\beta}_j$  ( $j = 1..k$ ) hay ma trận  $\hat{\beta}$  sao cho:

$$S = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \min \Leftrightarrow S = \sum_{i=1}^n e_i^2 = e^T e \min \Leftrightarrow \frac{\partial(e^T e)}{\partial \hat{\beta}} = 0$$

$$\Leftrightarrow \hat{\beta} = (X^T X)^{-1} \cdot X^T Y$$

$$X^T X = \begin{pmatrix} n & \sum X_{2i} & \sum X_{3i} & \dots & \sum X_{ki} \\ \sum X_{2i} & \sum X_{2i}^2 & \sum X_{2i}X_{3i} & \dots & \sum X_{2i}X_{ki} \\ \dots & \dots & \dots & \dots & \dots \\ \sum X_{ki} & \sum X_{ki}X_{2i} & \sum X_{ki}X_{3i} & \dots & \sum X_{ki}^2 \end{pmatrix} \quad X^T Y = \begin{pmatrix} \sum Y_i \\ \sum Y_i X_{2i} \\ \dots \\ \sum Y_i X_{ki} \end{pmatrix}$$

# Xác định phương trình hồi quy

- Bước 1: Tìm các ma trận:  $X^T X$ ,  $X^T Y$

$$X^T X = \begin{pmatrix} n & \sum X_{2i} & \sum X_{3i} & \dots & \sum X_{ki} \\ \sum X_{2i} & \sum X_{2i}^2 & \sum X_{2i}X_{3i} & \dots & \sum X_{2i}X_{ki} \\ \dots & \dots & \dots & \dots & \dots \\ \sum X_{ki} & \sum X_{ki}X_{2i} & \sum X_{ki}X_{3i} & \dots & \sum X_{ki}^2 \end{pmatrix} \quad X^T Y = \begin{pmatrix} \sum Y_i \\ \sum Y_i X_{2i} \\ \dots \\ \sum Y_i X_{ki} \end{pmatrix}$$

- Bước 2: Tính  $\det(X^T X)$ ,  $\widetilde{X^T X}$

- Bước 3: Tính  $\widetilde{X^T X} \cdot X^T Y$

- Bước 4: Tính  $\hat{\beta} = \frac{1}{\det(X^T X)} \widetilde{X^T X} \cdot X^T Y$

- Bước 5: Phương trình hồi quy:  $\hat{Y} = X\hat{\beta}$



Nghiên cứu mối quan hệ phụ thuộc giữa doanh số bán ra với chi phí dành cho quảng cáo và giá bán, người ta thu thập được các số liệu sau đây tại 10 cửa hàng cùng kinh doanh một loại mặt hàng:

|     |    |    |    |    |     |     |     |     |     |     |
|-----|----|----|----|----|-----|-----|-----|-----|-----|-----|
| Yi  | 84 | 90 | 92 | 96 | 100 | 108 | 120 | 126 | 130 | 136 |
| X2i | 8  | 9  | 10 | 9  | 10  | 12  | 12  | 14  | 14  | 15  |
| X3i | 9  | 8  | 8  | 7  | 7   | 8   | 7   | 7   | 6   | 6   |

Trong đó:

Yi: doanh số bán ra trong một tháng của cửa hàng thứ i (triệu đồng)

X2i: chi phí dành cho quảng cáo trong một tháng của cửa hàng thứ i (triệu đồng)

X3i: giá bán của cửa hàng thứ i (ngàn đồng/1 đv sp)

# Ý nghĩa của các hệ số hồi quy

- $\beta_1$ : ảnh hưởng của các nguyên nhân khác không được kể đến trong mô hình tới kết quả Y
- $\beta_j$ : ảnh hưởng của nguyên nhân  $X_j$  tới kết quả Y khi các nguyên nhân  $X_m$  còn lại không đổi ( $m \neq j, m = 2..k$ ).
  - $\beta_j > 0$ :  $X_j$  và Y có mối liên hệ thuận (cùng chiều)
  - $\beta_j < 0$ :  $X_j$  và Y có mối liên hệ nghịch (ngược chiều)

Khi  $X_j$  tăng lên 1 đơn vị thì Y tăng (giảm) trung bình  $\beta_j$  đơn vị.



## II.2 *Tính chất của ước lượng OLS*

i) Các  $\hat{\beta}_j$  ( $j = 1..k$ ) xác định duy nhất (với 1 mẫu thử cụ thể).  
Chúng là các ước lượng điểm, các đại lượng ngẫu nhiên.

ii) Hàm hồi quy tổng thể đi qua trung bình mẫu:

$$\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}_2 + \hat{\beta}_3 \bar{X}_3 + \dots + \hat{\beta}_k \bar{X}_k$$

iii) Giá trị trung bình của các  $\hat{Y}_i$  (được xác định thông qua hàm hồi quy mẫu) bằng trung bình các quan sát:  $\bar{\hat{Y}} = \bar{Y}$

iv) Tổng các phần dư bằng 0:  $\sum_{i=1}^n e_i = 0$



## II.2 *Tính chất của ước lượng OLS*

v) Các phần dư không tương quan với  $\hat{Y}_i$  :  $\sum_{i=1}^n \hat{Y}_i e_i = 0$

vi) Các phần dư không tương quan với  $X_{ji}$  ( $j = 2..k$ ):

$$\sum_{i=1}^n X_{ji} e_i = 0$$

vii) Với các giả thiết của mô hình hồi quy tuyến tính cổ điển thì các ước lượng bình phương nhỏ nhất  $\hat{\beta}_j$  ( $j = 1..k$ ) là các ước lượng hiệu quả của  $\beta_j$  ( $j = 1..k$ )

viii)  $\hat{\beta}_j \sim N(\beta_j, V(\hat{\beta}_j))$  ( $j = 1..k$ )