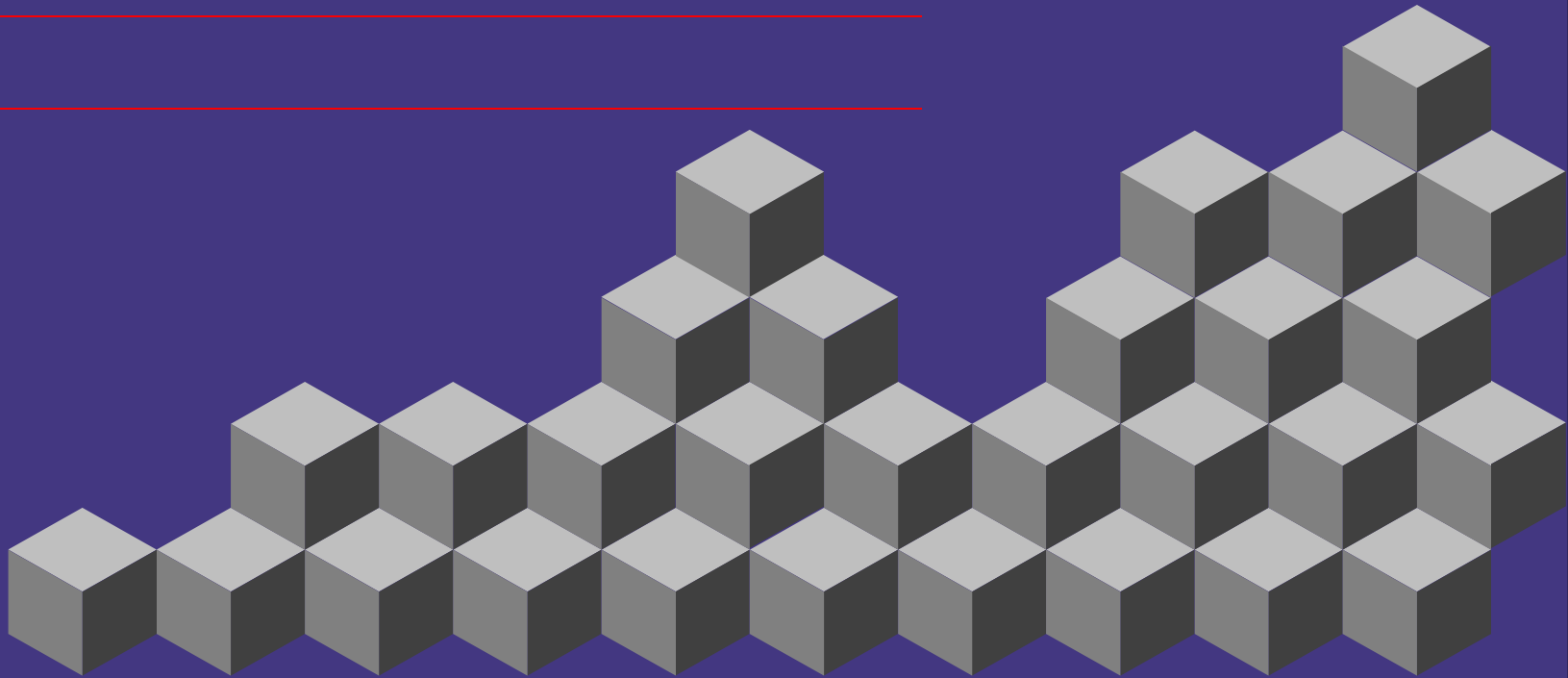


CHƯƠNG 5

BIẾN GIẢ TRONG PHÂN TÍCH HỒI QUY



BIẾN GIẢ

MỤC TIÊU

1. Biết cách đặt biến giả
2. Nắm phương pháp sử dụng biến giả trong phân tích hồi quy

NỘI DUNG

- 1 Khái niệm biến giả
- 2 Sử dụng biến giả trong mô hình hồi quy
- 3 Ứng dụng sử dụng biến giả

5.1 KHÁI NIỆM

- ❖ **Biến định lượng**: các giá trị quan sát được thể hệ bằng con số
- ❖ **Biến định tính**: thể hiện một số tính chất nào đó
- ❖ Để đưa những thuộc tính của biến định tính vào mô hình hồi quy, cần lượng hóa chúng => sử dụng biến giả (dummy variables)

5.2 Sử dụng biến giả trong mô hình hồi quy

Ví dụ 5.1: Xét mô hình $Y_i = \beta_1 + \beta_2 X_i + \beta_3 D_i + U_i$

với

Y Tiền lương (triệu đồng/tháng)

X Bậc thợ

D=1 nếu công nhân làm trong khu vực tư nhân

D=0 nếu công nhân làm trong khu vực nhà nước

D được gọi là biến giả trong mô hình

5.2 Sử dụng biến giả trong mô hình hồi quy

$$E(Y/X, D) = \beta_1 + \beta_2 X_i + \beta_3 D_i \quad (5.1)$$

$$E(Y/X, D=0) = \beta_1 + \beta_2 X_i \quad (5.2)$$

$$E(Y/X, D=1) = \beta_1 + \beta_2 X_i + \beta_3 \quad (5.3)$$

(5.2): tiền lương trung bình của công nhân làm việc trong khu vực quốc doanh với bậc thợ là X

(5.3): tiền lương trung bình của công nhân làm việc trong khu vực tư nhân với bậc thợ là X

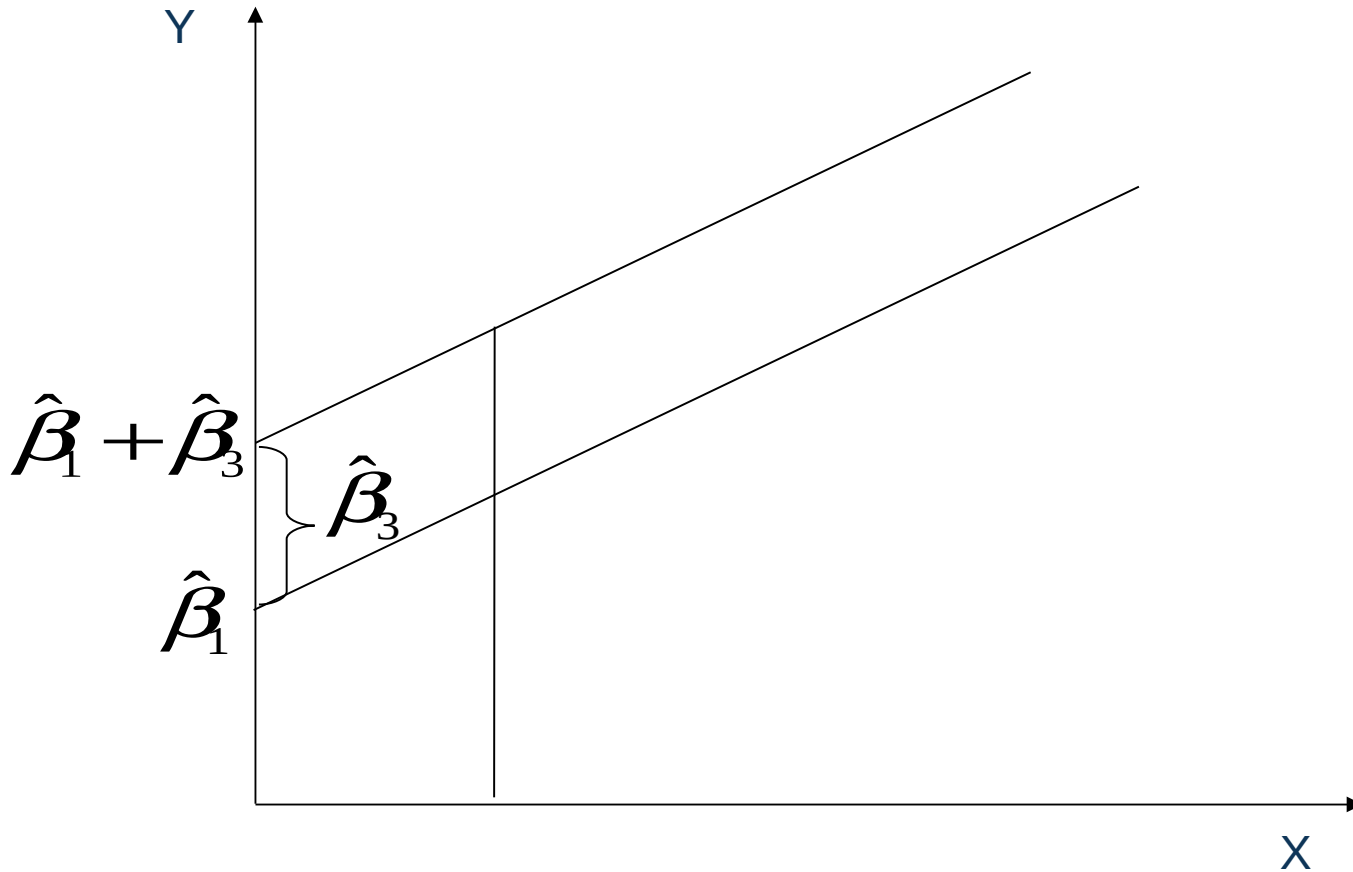
5.2 Sử dụng biến giả trong mô hình hồi quy

β_2 tốc độ tăng lương theo bậc thợ

β_3 chênh lệch tiền lương trung bình của công nhân làm việc ở hai khu vực và cùng bậc thợ

(Giả thiết của mô hình: tốc độ tăng lương theo bậc thợ ở hai khu vực giống nhau)

$$E(Y/X, Z) = \beta_1 + \beta_2 X_i + \beta_3 D_i$$



Hình 5.1 mức thu nhập bình quân tháng của người lao động tại KVQD và KVTN khi có bậc thợ là X

5.2 Sử dụng biến giả trong mô hình hồi quy

Ví dụ 5.2: Xét sự phụ thuộc của thu nhập (Y) (triệu đồng/tháng) vào thời gian công tác (X) (năm) và nơi làm việc của người lao động (DNNN, DNTN và DNLD)

Dùng 2 biến giả Z_1 và Z_2 với

- $Z_{1i} = 1$ nơi làm việc tại DNNN
- $Z_{1i} = 0$ nơi làm việc tại nơi khác
- $Z_{2i} = 1$ nơi làm việc tại DNTN
- $Z_{2i} = 0$ nơi làm việc tại nơi khác

- $Z_{1i} = 0$ và $Z_{2i} = 0$ phạm trù cơ sở

5.2 Sử dụng biến giả trong mô hình hồi quy

$$E(Y/X, Z1, Z2) = \beta_1 + \beta_2 X_i + \beta_3 Z_{1i} + \beta_4 Z_{2i}$$

$$E(Y/X, Z1=0, Z2=0) = \beta_1 + \beta_2 X_i$$

$$E(Y/X, Z1=1, Z2=0) = \beta_1 + \beta_2 X_i + \beta_3$$

$$E(Y/X, Z1=0, Z2=1) = \beta_1 + \beta_2 X_i + \beta_4$$

- β_3 chênh lệch thu nhập trung bình của nhân viên làm việc tại DNNN và DNLD khi có cùng thời gian làm việc X năm

- β_4 chênh lệch thu nhập trung bình của nhân viên làm việc tại DNTN và DNLD khi có cùng thời gian làm việc X năm

5.2 Sử dụng biến giả trong mô hình hồi quy

Ví dụ 5.3. thu nhập còn phụ thuộc vào trình độ người lao động (từ đại học trở lên, cao đẳng và khác)

$$D_{1i} = \begin{cases} 1: \text{nếu trình độ từ đại học trở lên} \\ 0: \text{trường hợp khác} \end{cases}$$
$$D_{2i} = \begin{cases} 1: \text{nếu trình độ cao đẳng} \\ 0: \text{trường hợp khác} \end{cases}$$

Một chỉ tiêu chất lượng có n phạm trù (thuộc tính) khác nhau thì dùng $n-1$ biến giả

5.2 Sử dụng biến giả trong mô hình hồi quy

VD 5.4: Khảo sát lương của giáo viên theo số năm giảng dạy

$$\text{Mô hình: } Y_i = \beta_1 + \beta_3 X_i$$

Trong đó

Y lương giáo viên

X số năm giảng dạy

và xem xét yếu tố giới tính có tác động đến thu nhập không

Z giới tính với $Z=1$: nam; $Z=0$: nữ

5.2 Sử dụng biến giả trong mô hình hồi quy

- ❑ TH1: Lương khởi điểm của gv nam và nữ khác nhau nhưng tốc độ tăng lương theo số năm giảng dạy như nhau
- ❑ TH2: Lương khởi điểm như nhau nhưng tốc độ tăng lương khác nhau
- ❑ TH3: Lương khởi điểm và tốc độ tăng lương khác nhau

5.2 Sử dụng biến giả trong mô hình hồi quy

TH1: Dịch chuyển số hạng tung độ gốc

Hàm PRF: $Y = \beta_1 + \beta_2 Z + \beta_3 X + U$

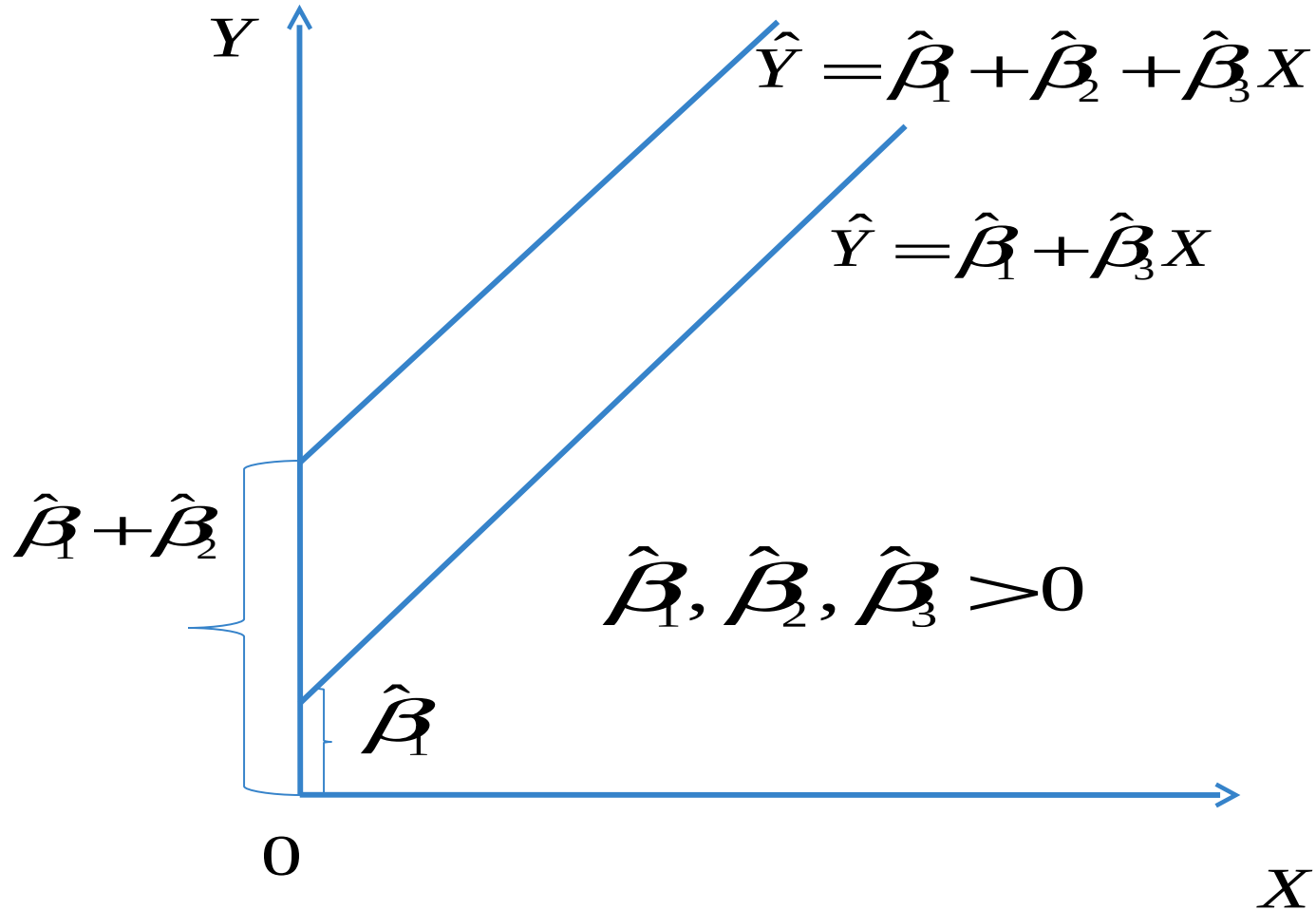
Hàm SRF ứng với nữ ($Z=0$) :

$$\hat{Y} = \hat{\beta}_1 + \hat{\beta}_3 X$$

Hàm SRF ứng với nam ($Z=1$) :

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

5.2 Sử dụng biến giả trong mô hình hồi quy



Hình 5.2 Lương khởi điểm của gv nam và nữ khác nhau

5.2 Sử dụng biến giả trong mô hình hồi quy

TH2: Dịch chuyển số hạng độ dốc

Hàm PRF:

$$Y = \beta_1 + \beta_2 X + \beta_3 (ZX) + U$$

Với ZX gọi là biến tương tác

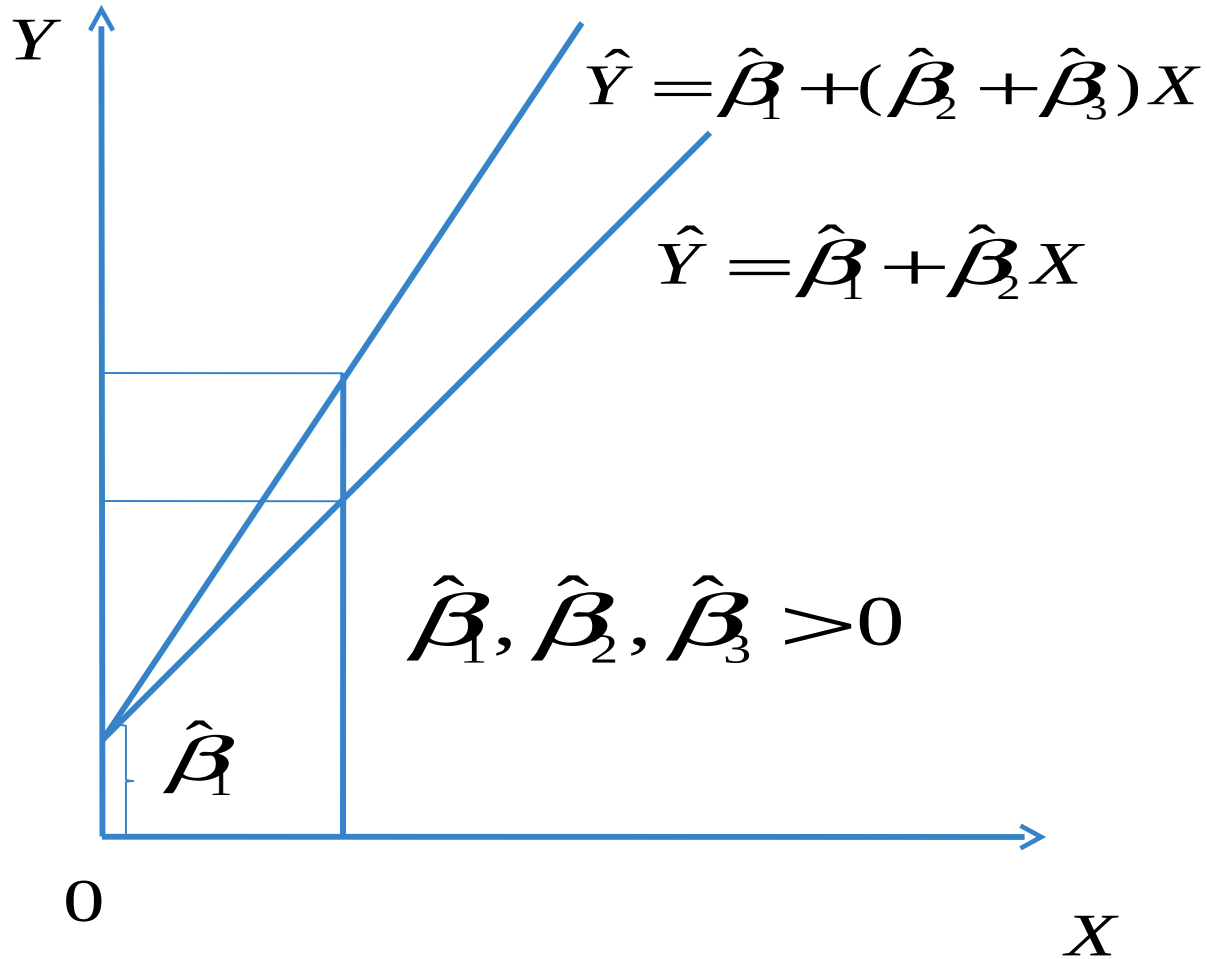
Hàm SRF ứng với nữ (Z=0) :

$$\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X$$

Hàm SRF ứng với nam (Z=1) :

$$\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X + \hat{\beta}_3 X = \hat{\beta}_1 + (\hat{\beta}_2 + \hat{\beta}_3) X$$

5.2 Sử dụng biến giả trong mô hình hồi quy



Hình 5.3 Mức tăng lương theo số năm giảng dạy của gv nam và nữ khác nhau

5.2 Sử dụng biến giả trong mô hình hồi quy

TH3: Dịch chuyển số hạng tung độ gốc và số hạng độ dốc

Hàm PRF: $Y = \beta_1 + \beta_2 Z + \beta_3 X + \beta_4 (ZX) + U$

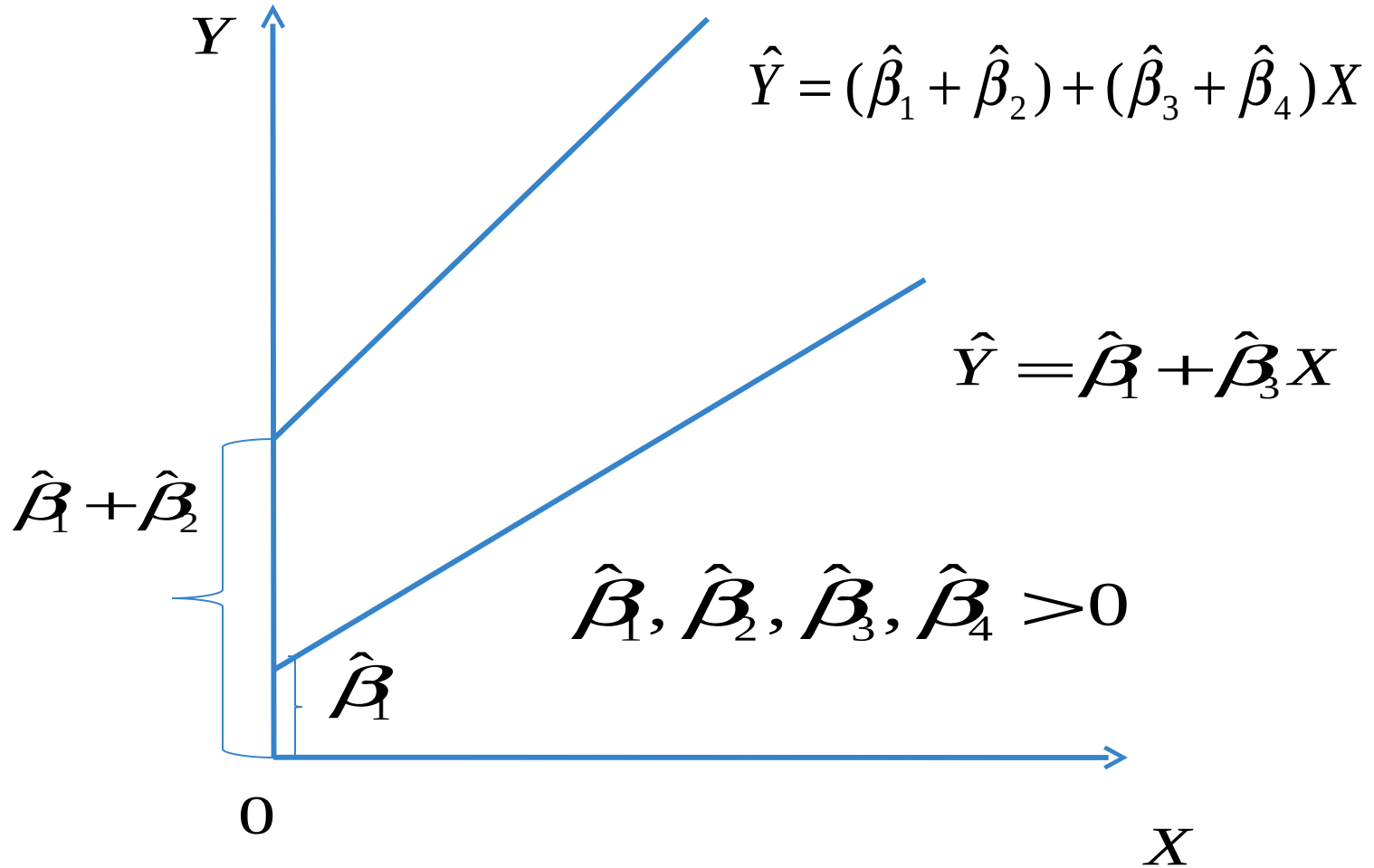
Hàm SRF ứng với nữ ($Z=0$) :

$$\hat{Y} = \hat{\beta}_1 + \hat{\beta}_3 X$$

Hàm SRF ứng với nam ($Z=1$) :

$$\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3 X + \hat{\beta}_4 X = (\hat{\beta}_1 + \hat{\beta}_2) + (\hat{\beta}_3 + \hat{\beta}_4) X$$

5.2 Sử dụng biến giả trong mô hình hồi quy



Hình 5.4 Lương khởi điểm và mức tăng lương của gv nam và nữ khác nhau

5.3 Ứng dụng sử dụng biến giả

5.3.1 Sử dụng biến giả trong phân tích mùa

Y chi tiêu cho tiêu dùng

X thu nhập

Z = 1 nếu quan sát trong mùa (tháng 1-6)

Z = 0 nếu quan sát không nằm trong mùa (tháng 7-12)

TH1: Nếu yếu tố mùa chỉ ảnh hưởng đến hệ số chặn

TH2: Nếu yếu tố mùa có ảnh hưởng đến hệ số góc

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{\beta}_3 Z_i \quad \hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{\beta}_3 Z_i + \hat{\beta}_4 X_i Z_i (*)$$

Mô hình * có tính tổng quát hơn. Qua việc kiểm định giả thiết để biết được hệ số góc nào có ý nghĩa.

5.3 Ứng dụng sử dụng biến giả

5.3.2 Kiểm định tính ổn định cấu trúc của các mô hình hồi quy
Ví dụ 5.5. Số liệu tiết kiệm và thu nhập cá nhân ở nước Anh từ 1946-63 (triệu pounds)

TK I	Tiết kiệm	Thu nhập	TK II	Tiết kiệm	Thu nhập
1946	0.36	8.8	1955	0.59	15.5
1947	0.21	9.4	1956	0.9	16.7
1948	0.08	10	1957	0.95	17.7
1949	0.2	10.6	1958	0.82	18.6
1950	0.1	11	1959	1.04	19.7
1951	0.12	11.9	1960	1.53	21.1
1952	0.41	12.7	1961	1.94	22.8
1953	0.5	13.5	1962	1.75	23.9
1954	0.43	14.3	1963	1.99	25.2

5.3 Ứng dụng sử dụng biến giả

Mục tiêu: Kiểm tra hàm tiết kiệm có thay đổi cấu trúc giữa 2 thời kỳ hay không.

Cách 1 Lập hai mô hình tiết kiệm ở 2 thời kỳ

Thời kỳ tái thiết: 1946-54

$$Y_i = \alpha_1 + \alpha_2 X_i + U_{1i}$$

Thời kỳ hậu tái thiết: 1955-63

$$Y_i = \lambda_1 + \lambda_2 X_i + U_{2i}$$

Và kiểm định các trường hợp sau

$$\begin{cases} \alpha_1 = \lambda_1 \\ \alpha_2 = \lambda_2 \end{cases}$$

$$\begin{cases} \alpha_1 = \lambda_1 \\ \alpha_2 \neq \lambda_2 \end{cases}$$

$$\begin{cases} \alpha_1 \neq \lambda_1 \\ \alpha_2 = \lambda_2 \end{cases}$$

$$\begin{cases} \alpha_1 \neq \lambda_1 \\ \alpha_2 \neq \lambda_2 \end{cases}$$

5.3 Ứng dụng sử dụng biến giả

Cách 2 Sử dụng biến giả

B1. Lập hàm tiết kiệm tổng quát của cả 2 thời kỳ

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{\beta}_3 Z_i + \hat{\beta}_4 X_i Z_i + e_i$$

Với $n = n_1 + n_2$

$Z = 1$ quan sát thuộc thời kỳ tái thiết

$Z = 0$ quan sát thuộc thời kỳ hậu tái thiết

B2. Kiểm định giả thiết $H_0: \beta_3 = 0$

Nếu chấp nhận H_0 : loại bỏ Z ra khỏi mô hình

B3. Kiểm định giả thiết $H_0: \beta_4 = 0$

Nếu chấp nhận H_0 : loại bỏ $Z_i X_i$ ra khỏi mô hình

5.3 Ứng dụng sử dụng biến giả

Kết quả hồi quy theo mô hình như sau

$$Y_i = -1,75 + 0,15045X_i + 1,4839Z_i - 0,1034X_iZ_i + e_i$$

$$t = \quad (-5,27) \quad (9,238) \quad (3,155) \quad (-3,109)$$

$$p = \quad (0,000) \quad (0,000) \quad (0,007) \quad (0,008)$$

Nhận xét

- Tung độ gốc chênh lệch và hệ số góc chênh lệch có ý nghĩa thống kê
- Các hồi quy trong hai thời kỳ là khác nhau

5.3 Ứng dụng sử dụng biến giả

Thời kỳ tái thiết: $Z = 1$

$$\hat{Y}_i = -1,75 + 0,15045X_i + 1,4839 - 0,1034X_i$$

$$\hat{Y}_i = -0,2661 + 0,0475X_i$$

Thời kỳ hậu tái thiết: $Z = 0$

$$\hat{Y}_i = -1,75 + 0,15045X_i$$

5.3 Ứng dụng sử dụng biến giả

Tiết kiệm

Thời kỳ hậu tái thiết

$$\hat{Y}_i = -1,75 + 0,15045X_i$$

$$\hat{Y}_i = -0,2661 + 0,0475X_i$$

Thời kỳ tái thiết

Thu nhập

-0.27

-1.75

Hình 5.6 Mô hình hồi quy cho 2 thời kỳ

5.3 Ứng dụng sử dụng biến giả

5.3.3. Hàm tuyến tính từng khúc

Ví dụ 5.6: Doanh thu dưới X^* thì tiền hoa hồng sẽ khác với khi doanh thu trên X^* .

Hàm hồi quy có dạng

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 (X_i - X^*) Z_i + u_i$$

Y Tiền hoa hồng

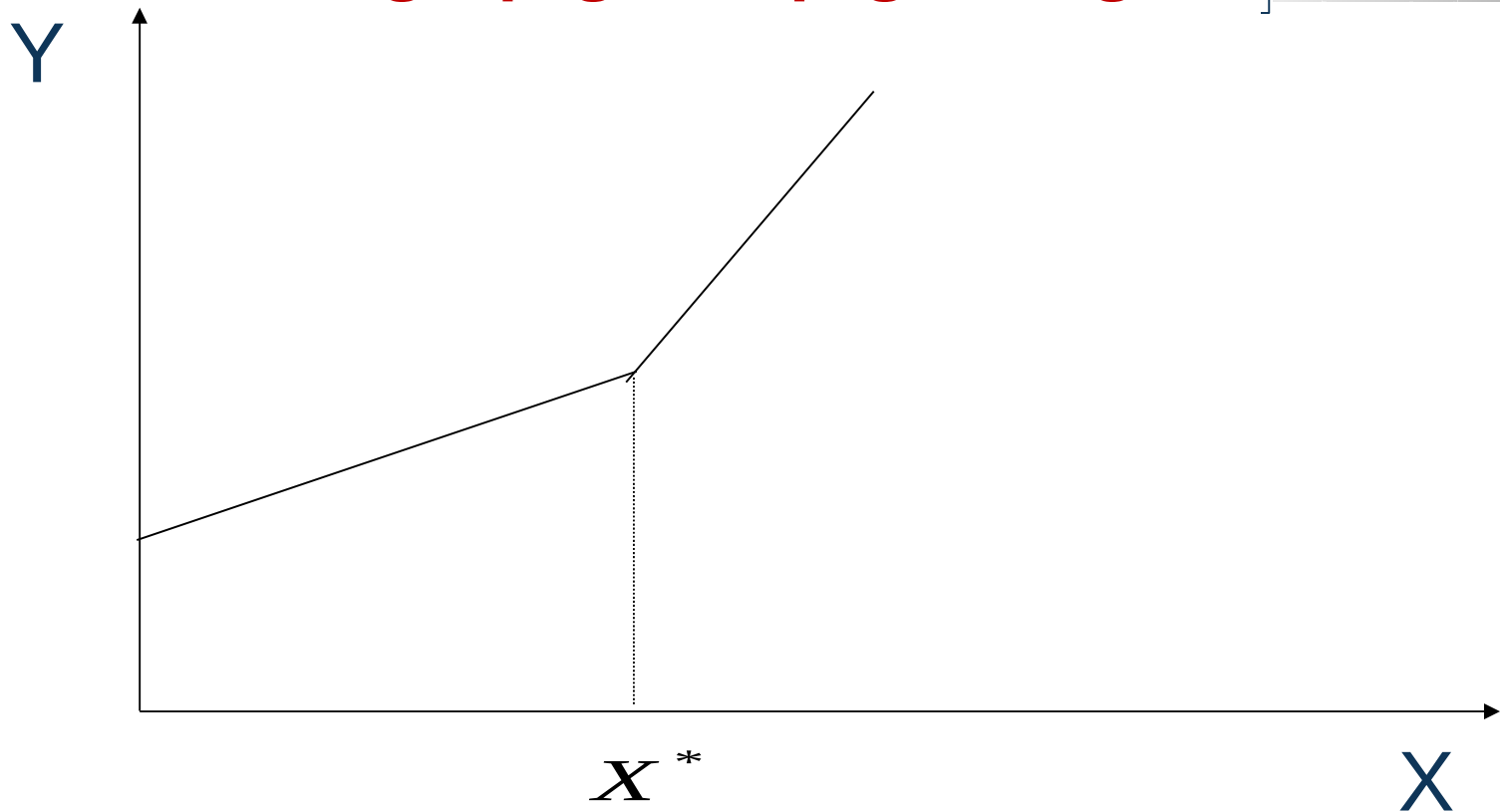
X Doanh thu

X^* Giá trị ngưỡng sản lượng

$Z_i = 1$ nếu $X_i > X^*$

$Z_i = 0$ nếu $X_i \leq X^*$

5.3 Ứng dụng sử dụng biến giả



Hình 5.7 Hàm tuyến tính từng khúc

- Kiểm định giả thiết $H_0: \beta_3=0$

Nếu bác bỏ H_0 : hàm hồi quy thay đổi cấu trúc