CHƯƠNG V

Hồi quy với biến định tính

Nội dung chính

- I. Biến giả trong PTHQ
- II. Biến định tính là biến phụ thuộc

I. Biến giả trong PTHQ

- **1. Giới thiệu**
  - Các biến độc lập trong mô hình có thể là các biến định lượng hoặc định tính.
  - Biến giả có đặc điểm sau: chỉ nhận một trong hai giá trị 0 hoặc 1.
  - D = 1 nếu thuộc tính xuất hiện
  - D= 0 nếu thuộc tính không xuất hiện
  - Ví dụ D là biến giới tính,
    - D=1 nếu là nam
    - D= 0 nếu không phải là nam (là nữ)

2. Sử dụng biến giả làm thay đổi hệ số cắt trong mô hình

a. Biến định tính có hai thuộc tính

Ví dụ: Xem xét ảnh hưởng của số năm kinh nghiệm X2, Giới tính (D) tới mức lương của công nhân (Y)

Chúng ta quy định

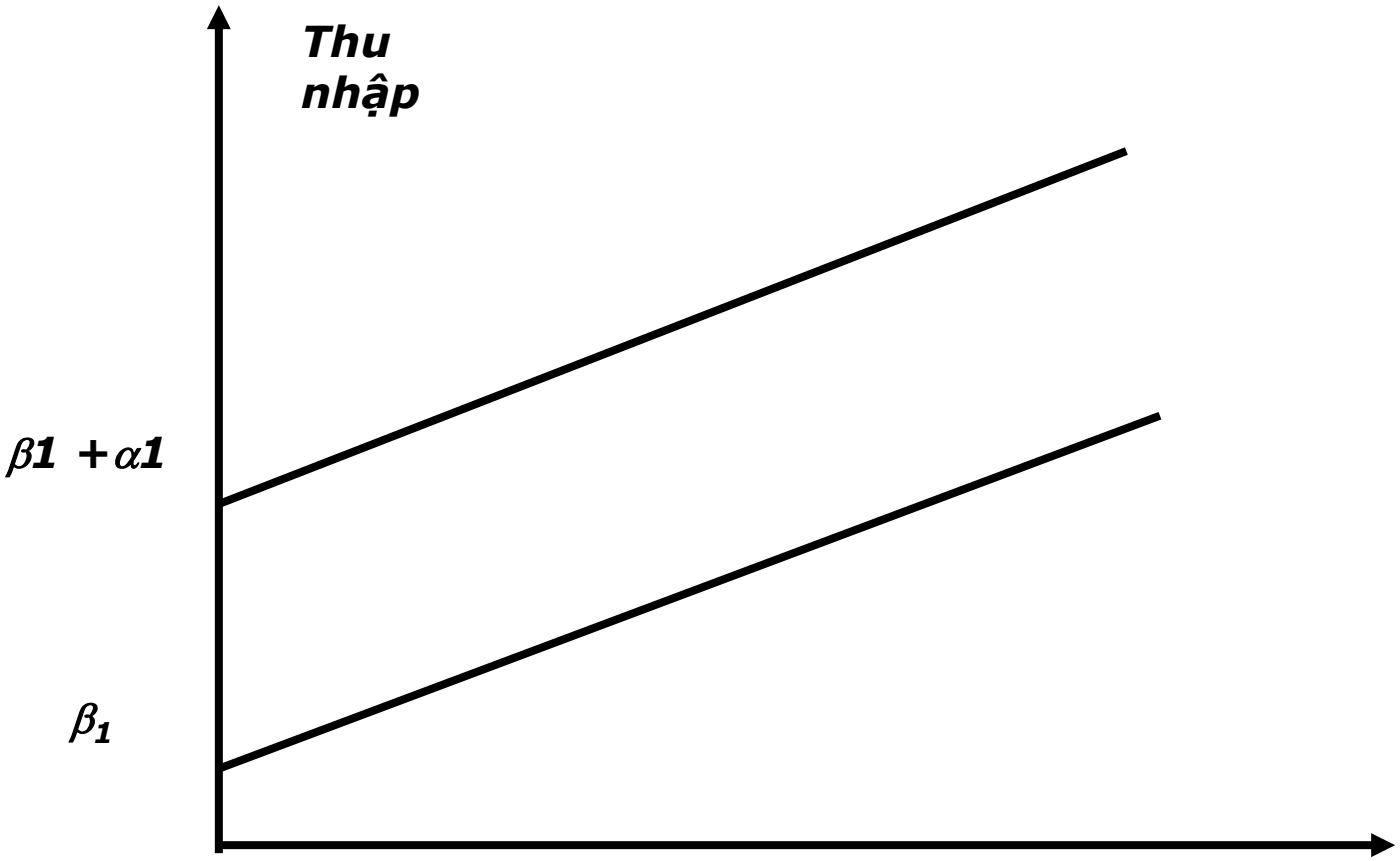
- Di = 1 nếu công nhân là nam
- Di = 0 nếu công nhân là nữ

- Mô hình hồi quy tổng thể được viết như sau:
- $Y_i = \beta_1 + \alpha D_i + \beta_2 X_{2i} + u_i$

Mô hình có biến giả cho phép xem xét sự khác biệt trong thu nhập giữa nam giới và phụ nữ:

Phụ nữ: $E(Y|X_2, Di=0) = \beta_1 + \beta_2 X_2$

Nam giới: $E(Y|X_2, Di=1) = \beta_1 + \alpha + \beta_2 X_2$


Thu nhập
$\beta 1 + \alpha 1$
β_1

D=1, nam

Regression Statistics	
Multiple R	0.321607
R Square	0.103431
Adjusted R Square	0.10221
Standard Error	170.1109
Observations	1472

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	4904035	2452017	84.7343	1.49E-35
Residual	1469	42509509	28937.72		
Total	1471	47413544			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	335.5799	9.753061	34.4076	8.9E-19	316.4485	354.711
X2	5.154614	0.441921	11.6641	3.95E-3	4.28775	6.02147
D	35.36062	9.194312	3.84592	0.00012	17.32524	53.396

b. Biến định tính có m thuộc tính

- Với biến định tính có m thuộc tính, chúng ta đưa m-1 biến giả vào trong mô hình
- Ví dụ: xét ảnh hưởng của số năm kinh nghiệm (X) , trình độ giáo dục tới mức lương công nhân (Y).
- Giả sử có 3 mức trình độ học vấn: Dưới trung học cơ sở, từ THCS tới PTTH, và từ đại học trở lên.
- D1i = 1 nếu là trình độ dưới THCS
- D1i = 0 nếu là trình độ khác
- D2i = 1 nếu là trình độ từ THCS-PTTH
- D2i = 0 nếu là trình độ khác
- Vậy khi D1i = D2i = 0, đó là trình độ trên PTTH

- Mô hình hồi quy :
- $Y_i = \beta_1 + \alpha_1 D_{1i} + \alpha_2 D_{2i} + \beta_2 X_{2i} + u_i$
- Mô hình hồi quy viết cho 3 cấp sẽ là:
- $E(Y|X_2, D_{1i}=1, D_{2i}=0) = \beta_1 + \alpha_1 + \beta_2 X_2$
- $E(Y|X_2, D_{1i}=0, D_{2i}=1) = \beta_1 + \alpha_2 + \beta_2 X_2$
- $E(Y|X_2, D_{1i}=D_{2i}=0) = \beta_1 + \beta_2 X_2$
- Kỳ vọng về dấu của các hệ số hồi quy
- Ví dụ minh họa

Regression Statistics	
Multiple R	0.48299
R Square	0.23328
Adjusted R Square	0.23171
Standard Error	157.363
Observations	1472

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	11060932	368697	148.88	
Residual	1468	36352612	24763.		
Total	1471	47413544			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	356.678	8.119	43.934	0.000	340.753	372.60
exper	7.446	0.423	17.615	0.000	6.617	8.275
D1	-209.27	17.067	-12.262	0.000	-242.748	-175.79
D2	-139.00	10.973	-12.668	0.000	-160.525	-117.47

3. Sử dụng biến giả làm thay đổi hệ số góc

VD trên: ảnh hưởng của số năm kinh nghiệm tới mức lương công nhân còn phụ thuộc vào giới tính.

- $Y_i = \beta_1 + \beta_2 X_i + \alpha D_i X_i + u_i$

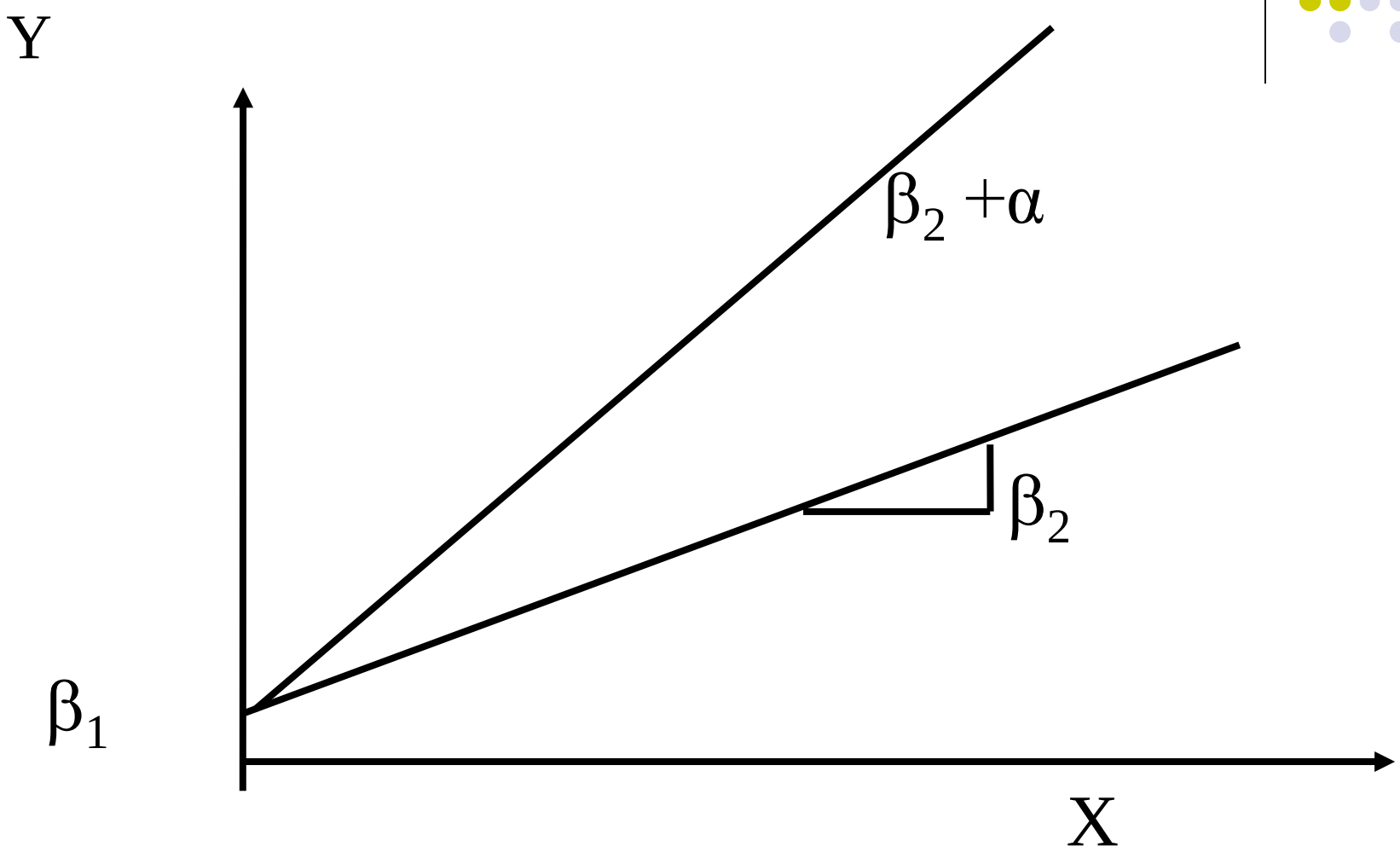
Trong đó:

- Y là mức lương công nhân
- X là số năm kinh nghiệm
- D là biến giới tính (1 nam và 0 là nữ)
- Phương trình cho nam

$$E(Y \mid D_i = 1) = \beta_1 + (\beta_2 + \alpha) X_i$$

- Phương trình cho nữ

$$E(Y \mid D_i = 0) = \beta_1 + \beta_2 X_i$$


Y
$\beta_2 + \alpha$
β_2
β_1
X

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.324
R Square	0.105
Adjusted R Square	0.104
Standard Error	169.958
Observations	1472

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	4980522	2490261	86.21101	3.96E-36
Residual	1469	42433022	28885.65		
Total	1471	47413544			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	356.000	8.758	40.650	0.000	338.821	373.179
X	3.913	0.567	6.905	0.000	2.801	5.024
DX	1.995	0.477	4.179	0.000	1.059	2.931

II. Biến giả là biến PT

- Mô hình xác suất tuyến tính
- Mô hình phản ứng nhị phân: Logit và Probit

- → tham khảo giáo trình