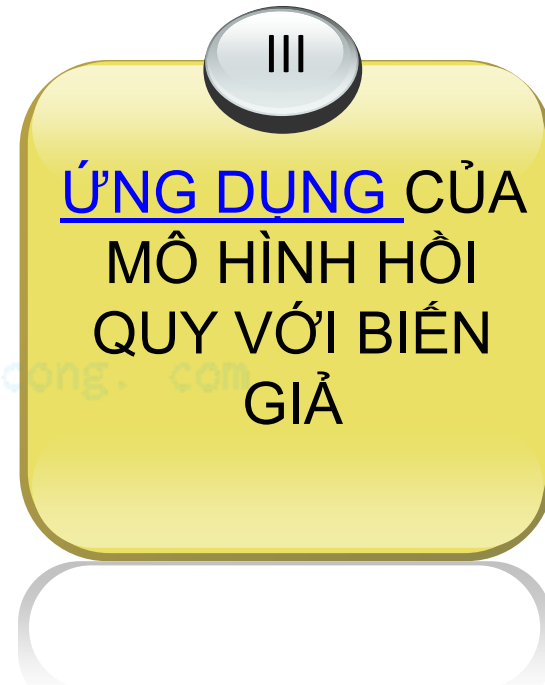


MÔ HÌNH HỒI QUY VỚI BIẾN GIẢ



I. MÔ HÌNH HỒI QUY VỚI BIẾN GIẢ

I.1 Khái niệm biến giả

I.2 Mô hình với biến chất lượng có 2 phạm trù

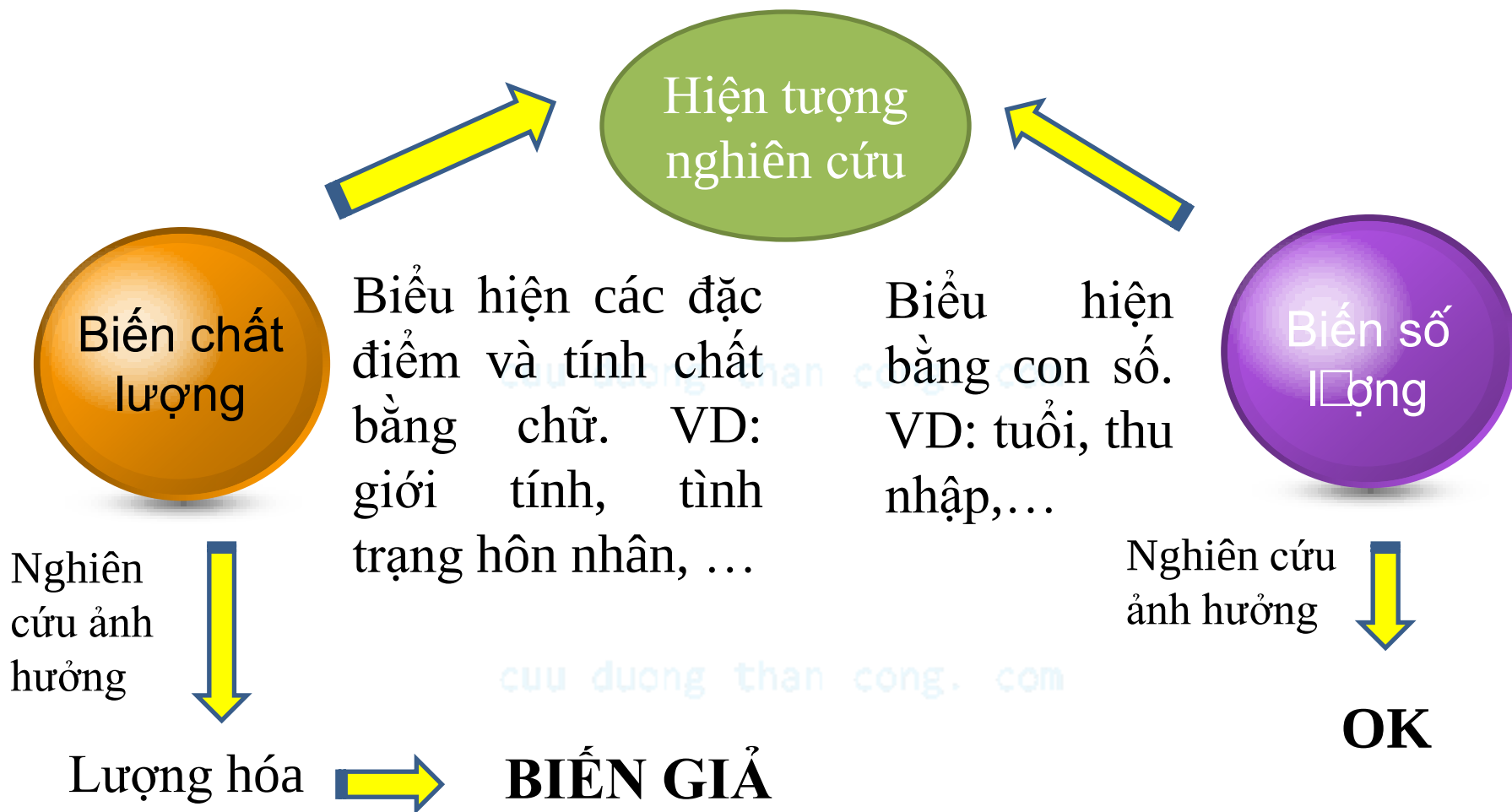
I.3 Mô hình với biến chất lượng có nhiều hơn 2 phạm trù

I.4 Mô hình có nhiều biến chất lượng

I.5 Mô hình hỗn hợp

I.1

Khái niệm biến giả



Khái niệm:

Biến giả là biến chất lượng đã được lượng hóa, các giá trị có thể có của biến giả chỉ là 2 giá trị 0 và 1. Nó chỉ ra có hay không có một thuộc tính nào đó.

Ví dụ:

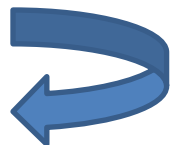
cuu duong than cong. com

Để biểu thị giới tính, ta sử dụng biến giả Z và quy ước:

$Z = 1$: Nữ, $Z = 0$: Nam

Để biểu thị tình trạng hôn nhân có thể quy ước:

$Z = 0$: Độc thân; $Z = 1$: Có gia đình



I.2

Mô hình với biến chất lượng có 2 phạm trù

Giả sử một xí nghiệp sản xuất có thể áp dụng 2 công nghệ sản xuất A và B, năng suất của mỗi công nghệ là đại lượng ngẫu nhiên phân phối theo quy luật chuẩn có phương sai bằng nhau, nhưng kỳ vọng toán có thể khác nhau.

Năng suất (kg)	22	19	18	21	18,5	21	20,5	17	17,5	21,5
Công nghệ	B	A	A	B	A	B	B	A	A	B

Mô hình hồi quy cho bởi: $Y_i = \beta_1 + \beta_2 Z_i + U_i$

Trong đó:

Y_i : năng suất của xí nghiệp U_i : sai số ngẫu nhiên

Z_i : biểu thị công nghệ sản xuất được áp dụng, quy ước:

$Z_i = 0$: công nghệ sản xuất A; $Z_i = 1$: công nghệ sản xuất B



Y_i	22	19	18	21	18,5	21	20,5	17	17,5	21,5
Z_i	1	0	0	1	0	1	1	0	0	1

→ Trở lại mô hình hồi quy 2 biến thông thường:

$$\hat{Y}_i = 18 + 3,2 * Z_i$$

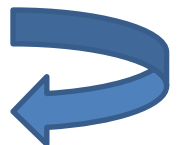
Ý nghĩa các hệ số

Với giả thiết $E(U_i) = 0$ được thỏa mãn thì:

$E(Y/Z_i = 0) = \beta_1$: khi áp dụng công nghệ sản xuất A, năng suất trung bình của xí nghiệp là $\beta_1 = 18$ kg

$E(Y/Z_i = 1) = \beta_1 + \beta_2$: khi áp dụng công nghệ sản xuất B, năng suất trung bình của xí nghiệp là $\beta_1 + \beta_2 = 21,2$ kg

→ β_2 là sự chênh lệch về năng suất khi chuyển từ công nghệ sản xuất A sang công nghệ sản xuất B



I.3

Mô hình với biến chất lượng có nhiều hơn 2 phạm trù

Giả sử xí nghiệp nọ ngoài công nghệ sản xuất A và B còn có thể áp dụng công nghệ sản xuất C (năng suất của chúng có phương sai như nhau), khi đó, chúng ta cần dùng 2 biến giả Z_{1i} , Z_{2i} biểu thị công nghệ sản xuất được áp dụng cụ thể như sau:

- $Z_{1i} = 1$: CNXS B; $Z_{1i} = 0$: CNXS khác
- $Z_{2i} = 1$: CNXS C; $Z_{2i} = 0$: CNXS khác

Việc cho giá trị các biến giả này tương đương với bảng sau:

Công nghệ	A	B	C
Z_{1i}	0	1	0
Z_{2i}	0	0	1

→ Nếu biến chất lượng có m phạm trù thì số biến giả cần đưa vào mô hình là m-1 biến.

Phạm trù ứng với các giá trị bằng 0 của các biến giả được gọi là phạm trù cơ sở. Phạm trù cơ sở hiểu theo nghĩa là việc so sánh được tiến hành với phạm trù này.

Mô hình hồi quy ở ví dụ trên được cho dưới dạng:

$$Y_i = \beta_1 + \beta_2 Z_{1i} + \beta_3 Z_{2i} + U_i$$

Trong đó: CNSX A là phạm trù cơ sở

Y_i : năng suất của xí nghiệp U_i : sai số ngẫu nhiên

→ Trở lại mô hình hồi quy bội thông thường

Với giả thiết $E(U_i) = 0$ được thỏa mãn thì:

$E(Y/Z_{1i} = 0; Z_{2i} = 0) = \beta_1$: khi áp dụng CNSX A, năng suất trung bình của xí nghiệp là β_1 .

$E(Y/Z_{1i} = 1; Z_{2i} = 0) = \beta_1 + \beta_2$: khi áp dụng CNSX B, năng suất trung bình của xí nghiệp là $\beta_1 + \beta_2$

$E(Y/Z_{1i} = 0; Z_{2i} = 1) = \beta_1 + \beta_3$: khi áp dụng CNSX C, năng suất trung bình của xí nghiệp là $\beta_1 + \beta_3$

→ β_2 là sự chênh lệch về năng suất khi chuyển từ CNSX A sang CNSX B

→ β_3 là sự chênh lệch về năng suất khi chuyển từ CNSX A sang CNSX C

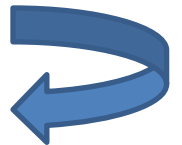
Một vài kiểm định hay gặp

$\begin{cases} H_0: \beta_2 = 0 \\ H_1: \beta_2 \neq 0 \end{cases} \longrightarrow$ Không hay có sự khác biệt giữa CNSX A và CNSX B

$\begin{cases} H_0: \beta_3 = 0 \\ H_1: \beta_3 \neq 0 \end{cases} \longrightarrow$ Không hay có sự khác biệt giữa CNSX A và CNSX C

$\begin{cases} H_0: \beta_2 = \beta_3 = 0 \\ H_1: \beta_3 \neq 0 \text{ hoặc } \beta_2 \neq 0 \end{cases} \longrightarrow$ Không hay có sự khác biệt giữa các CNSX

$\longrightarrow$ Trở lại kiểm định các hệ số hồi quy và kiểm định giả thuyết đồng thời trong mô hình hồi quy bội.



I.4

Mô hình có nhiều biến chất lượng

Giả sử mô hình có k biến giải thích là biến chất lượng, số phạm trù của X_j là m_j

Tổng số biến giả cần thiết để biểu thị các biến chất lượng cần đưa vào mô hình sẽ là: $\sum_{j=1}^k (m_j - 1)$

Xét mô hình hồi quy biểu thị nhu cầu về may mặc của nam, nữ thanh niên thuộc các nghề nghiệp khác nhau: công nhân, nông dân, tiểu thương, sinh viên.

Số biến giả cần đưa vào mô hình sẽ là: $(2 - 1) + (4 - 1) = 4$

Giá trị các biến giả ứng với các phạm trù có thể được cho như sau:

	Z_{1i}	Z_{2i}	Z_{3i}	Z_{4i}
Nam SV	0	0	0	0
Nữ SV	1	0	0	0
Nam CN	0	1	0	0
Nữ CN	1	1	0	0
Nam ND	0	0	1	0
Nữ ND	1	0	1	0
Nam TT	0	0	0	1
Nữ TT	1	0	0	1

➡ Phạm trù cơ sở là Nam SV. Mô hình hồi quy được cho dưới dạng:

$$Y_i = \beta_1 + \beta_2 Z_{1i} + \beta_3 Z_{2i} + \beta_4 Z_{3i} + \beta_5 Z_{4i} + U_i$$

Trong đó: Y_i : chỉ cho may mặc U_i : sai số ngẫu nhiên

➡ Trở lại mô hình hồi quy bội thông thường

Ý nghĩa các hệ số

Chi TB cho may mặc		Nghề nghiệp			
		SV	CN	ND	TT
Giới tính	Nam	β_1	$\beta_1 + \beta_3$	$\beta_1 + \beta_4$	$\beta_1 + \beta_5$
	Nữ	$\beta_1 + \beta_2$	$\beta_1 + \beta_2 + \beta_3$	$\beta_1 + \beta_2 + \beta_4$	$\beta_1 + \beta_2 + \beta_5$

→ β_2 : chênh lệch về chi tiêu dành cho may mặc khi chuyển từ nam sang nữ ở cùng một nghề.

→ β_j ($j = 3, 4$ hoặc 5): chênh lệch về chi tiêu dành cho may mặc khi chuyển từ sinh viên sang các nghề tương ứng ở cùng giới tính.

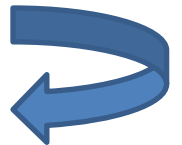
Một vài kiểm định hay gặp

$\begin{cases} H_0: \beta_2 = 0 \\ H_1: \beta_2 \neq 0 \end{cases} \longrightarrow$ Không hay có sự khác biệt về chi cho may mặc giữa giữa nam và nữ

$\begin{cases} H_0: \beta_3 = 0 \\ H_1: \beta_3 \neq 0 \end{cases} \longrightarrow$ Không hay có sự khác biệt về chi cho may mặc giữa SV và CN

$\begin{cases} H_0: \beta_3 = \beta_4 = \beta_5 = 0 \\ H_1: \exists j \in \{3, 4, 5\} | \beta_j \neq 0 \end{cases} \longrightarrow$ Không hay có sự khác biệt về chi cho may mặc giữa các nghề

$\longrightarrow$ Trở lại kiểm định các hệ số hồi quy và kiểm định giả thuyết đồng thời trong mô hình hồi quy bội.



I.5

Mô hình hỗn hợp

Là mô hình hồi quy mà trong đó có cả biến số lượng lẫn biến chất lượng, biến chất lượng sẽ được biểu thị bằng biến giả theo quy tắc như đã nói ở trên.

Việc phân tích hồi quy mối quan hệ giữa biến phụ thuộc với các biến giải thích (bao gồm cả biến số lượng và biến giả), được tiến hành tương tự như đối với mô hình hồi quy nhiều biến.

Ví dụ

Điều tra về mức chi tiêu cho may mặc của nam, nữ công nhân ở một nhà máy, người ta thu được số liệu sau:

Y_i	3	4	3,5	4,5	4	4,5	4	5,5
X_i	2,5	2,5	3	3	3,5	3,5	4	4
Z_i	0	1	0	1	0	1	0	1

Trong đó: X_i : mức thu nhập trong một tháng (triệu đồng)

Y_i : mức chi tiêu cho may mặc trong một năm (triệu đồng)

Z_i : là biến giả biểu thị giới tính, trong đó:

$Z_i = 0$: nam; $Z_i = 1$: nữ

Bằng phương pháp bình phương nhỏ nhất và dựa vào số liệu trên, hãy xây dựng hàm hồi quy mẫu dưới dạng sau:

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{\beta}_3 Z_i$$

$$X^T Y = \begin{pmatrix} 33 \\ 109,25 \\ 18,5 \end{pmatrix}; X^T X = \begin{pmatrix} 8 & 26 & 4 \\ 26 & 87 & 13 \\ 4 & 13 & 4 \end{pmatrix}; \widetilde{X^T X} = \begin{pmatrix} 179 & -52 & -10 \\ -52 & 16 & 0 \\ -10 & 0 & 20 \end{pmatrix}$$

$$\det(X^T X) = 40 \quad \hat{\beta} = \frac{1}{40} \begin{pmatrix} 179 & -52 & -10 \\ -52 & 16 & 0 \\ -10 & 0 & 20 \end{pmatrix} \begin{pmatrix} 33 \\ 109,25 \\ 18,5 \end{pmatrix} = \begin{pmatrix} 1,025 \\ 0,8 \\ 1 \end{pmatrix}$$

$$\hat{Y}_i = 1,025 + 0,8 * X_i + Z_i$$

Ý nghĩa các hệ số:

$\hat{\beta}_2 = 0,8$: Nếu cùng giới tính, khi thu nhập tăng thêm 1 triệu đồng/ tháng thì chi tiêu trung bình cho may mặc sẽ tăng thêm 0.8 triệu đồng/ năm.

$\hat{\beta}_3 = 1$: Nếu cùng thu nhập, thì nữ công nhân chi tiêu cho may mặc trung bình nhiều hơn nam là 1 triệu đồng/ năm.

