



BIẾN ĐỘC LẬP ĐỊNH TÍNH (BIẾN GIẢ)

GV : Đinh Công Khải – Chương trình Fulbright
Môn: Các Phương Pháp Định Lượng – MPP4

Giới thiệu chung

- Các biến độc lập có thể là những biến định tính được dùng để giải thích biến Y ví dụ như giới tính, chủng tộc, tôn giáo, khu vực địa lý, bất ổn kinh tế hay chính trị, sự thay đổi chính sách,...
- Biến định tính hay còn được gọi biến giả, biến chỉ định, biến nhị phân, biến phân loại hay phạm trù (giá trị 1 biểu thị sự xuất hiện một tính chất; giá trị 0 khi không có tính chất đó).

Giới thiệu chung

$$\square Y_i = \alpha_1 + \alpha_2 D_i + u_i \quad (\text{phân tích phương sai ANOVA})$$

Y = mức lương năm của một giáo sư đại học

$D_i = 1$ nếu là nam; 0 nếu là nữ.

→ Mức lương trung bình của một giáo sư đại học là nữ: $E(Y_i | D_i = 0) = \alpha_1$;

→ Mức lương trung bình của một giáo sư đại học là nam: $E(Y_i | D_i = 1) = \alpha_1 + \alpha_2$;

$$\square \hat{Y}_i = 18,0 + 3,28 D_i$$

t (57,54) (7,44) $R^2 = 0,87$

Hồi qui một biến định lượng và một biến định tính có 2 phạm trù/đặc tính

$$\square Y_i = \alpha_1 + \alpha_2 D_i + \beta X_i + u_i \quad (\text{phân tích tích sai ANCOVA})$$

Y = mức lương năm của một giáo sư đại học

X = số năm kinh nghiệm giảng dạy

$D_i = 1$ nếu là nam; 0 nếu là nữ.

→ Lương trung bình của một giáo sư đại học là nữ:

$$E(Y_i | X_i, D_i = 0) = \alpha_1 + \beta X_i;$$

→ Lương trung bình của một giáo sư đại học là nam:

$$E(Y_i | X_i, D_i = 1) = (\alpha_1 + \alpha_2) + \beta X_i;$$

Các thức xây dựng biến giả

- Giả sử, chúng ta cần xây dựng biến giả để phân biệt giới tính nam và nữ
 - $D_{2i} = 1$ nếu giáo sư là nam;
= 0 nếu khác.
 - $D_{3i} = 1$ nếu giáo sư là nữ;
= 0 nếu khác.
 - $Y_i = \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \beta X_i + u_i$
 - D_2 và D_3 sẽ có hiện tượng đa cộng tuyến hoàn hảo
 - Nếu biến định tính có m phạm trù, chỉ cần đưa (m-1) biến giả vào mô hình

Các thức xây dựng biến giả

- Việc giải thích kết quả hồi qui của biến giả phụ thuộc vào giá trị 1 và 0 được gán cho biến giả như thế nào.

$$Y_i = \alpha'_1 + \alpha'_2 D_i + \beta X_i + u_i$$

$$D_i = 1 \text{ nếu là nữ; } 0 \text{ nếu là nam.}$$
- Lương trung bình của một giáo sư đại học là nam:

$$E(Y_i | X_i, D_i = 0) = \alpha'_1 + \beta X_i;$$
- Lương trung bình của một giáo sư đại học là nữ:

$$E(Y_i | X_i, D_i = 1) = (\alpha'_1 + \alpha'_2) + \beta X_i; \quad (\alpha'_2 < 0)$$

Các thức xây dựng biến giả

- Nhóm phạm trù hay phân loại được gán cho giá trị 0 thường được coi là phạm trù cơ sở/mốc/kiểm soát/ tham chiếu.
- α_2 được gọi là hệ số tung độ gốc chênh lệch (sự khác biệt giữa giá trị tung độ gốc của phạm trù nhận giá trị 1 và giá trị tung độ gốc của phạm trù nhận giá trị 0)

Hồi qui một biến định lượng và một biến định tính có nhiều phạm trù/đặc tính

- $Y_i = \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \beta X_i + u_i$

Y = chi tiêu y tế hàng năm

X = thu nhập hàng năm

$D_{2i} = 1$ nếu là có trình độ trung học; 0 nếu khác.

$D_{3i} = 1$ nếu là có trình độ từ đại học; 0 nếu khác.

- $E(Y_i | X_i, D_{2i} = 0, D_{3i} = 0) = \alpha_1 + \beta X_i;$
- $E(Y_i | X_i, D_{2i} = 1, D_{3i} = 0) = (\alpha_1 + \alpha_2) + \beta X_i;$
- $E(Y_i | X_i, D_{2i} = 0, D_{3i} = 1) = (\alpha_1 + \alpha_2 + \alpha_3) + \beta X_i$

Hồi qui một biến định lượng và 2 biến định tính

$$\square Y_i = \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \beta X_i + u_i$$

Y = lương hàng năm

X = số năm kinh nghiệm giảng dạy

$D_{2i} = 1$ nếu là nam; 0 nếu khác.

$D_{3i} = 1$ nếu là da trắng; 0 nếu khác.

Mức lương trung bình của giáo sư nữ da đen:

$$\rightarrow E(Y_i | X_i, D_{2i} = 0, D_{3i} = 0) = \alpha_1 + \beta X_i;$$

Hồi qui một biến định lượng và 2 biến định tính

Mức lương trung bình của giáo sư nam da đen:

$$\rightarrow E(Y_i | X_i, D_{2i} = 1, D_{3i} = 0) = (\alpha_1 + \alpha_2) + \beta X_i;$$

Mức lương trung bình của giáo sư nữ da trắng:

$$\rightarrow E(Y_i | X_i, D_{2i} = 0, D_{3i} = 1) = (\alpha_1 + \alpha_3) + \beta X_i$$

Mức lương trung bình của giáo sư nam da trắng:

$$\rightarrow E(Y_i | X_i, D_{2i} = 1, D_{3i} = 1) = (\alpha_1 + \alpha_2 + \alpha_3) + \beta X_i$$

Kiểm định tính ổn định cấu trúc của các mô hình hồi qui

- $Y_i = \alpha_1 + \alpha_2 X_i + u_{1i}$ (thời kỳ tái thiết)
- $Y_i = \beta_1 + \beta_2 X_i + u_{2i}$ (thời kỳ hậu tái thiết)
- Y = tiết kiệm; X = thu nhập.
- Các trường hợp:
 - $\alpha_1 = \beta_1$ và $\alpha_2 = \beta_2$; hồi qui trùng khớp.
 - $\alpha_1 \neq \beta_1$ và $\alpha_2 = \beta_2$; hồi qui song song.
 - $\alpha_1 = \beta_1$ và $\alpha_2 \neq \beta_2$; hồi qui đồng quy.
 - $\alpha_1 \neq \beta_1$ và $\alpha_2 \neq \beta_2$; hồi qui không giống nhau.

Kiểm định tính ổn định cấu trúc của các mô hình hồi qui

- $Y_i^{\wedge} = -0,27 + 0,047 X_i$ (thời kỳ tái thiết)
- $Y_i^{\wedge} = -1,75 + 0,15 X_i$ (thời kỳ hậu tái thiết)
- Y = tiết kiệm; X = thu nhập.
- **So sánh 2 hồi qui: Phương pháp biến giả**

$$Y_i = \alpha_1 + \alpha_2 D_i + \beta_1 X_i + \beta_2 (X_i D_i) + u_i$$

$D_i = 1$ nếu là thời kỳ tái thiết; 0 nếu là thời kỳ hậu tái thiết.

 - ❖ Tiết kiệm trung bình thời kỳ tái thiết: $E(Y_i | X_i, D_i = 1) = (\alpha_1 + \alpha_2) + (\beta_1 + \beta_2) X_i$;
 - ❖ Tiết kiệm trung bình thời kỳ hậu tái thiết: $E(Y_i | X_i, D_i = 0) = \alpha_1 + \beta_1 X_i$;

Kiểm định tính ổn định cấu trúc của các mô hình hồi qui

- Ví dụ

$$Y_i^{\wedge} = -1,75 + 1,48 D_i + 0,15 X_i - 0,1(X_i D_i)$$

$$t \quad (-5,27) \quad (3,15) \quad (9,22) \quad (-3,11) \quad R^2 = 0,94$$

- Thời kỳ tái thiết

$$Y_i^{\wedge} = (-1,75 + 1,48) + (0,15 - 0,1)X_i$$

$$Y_i^{\wedge} = -0,27 + 0,05X_i$$

- Thời kỳ hậu tái thiết

$$Y_i^{\wedge} = -1,75 + 0,15 X_i$$

Biến giả tương tác

- $Y_i = \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \beta X_i + u_i$

Y = chi tiêu may mặc hàng năm

X = thu nhập

$D_{2i} = 1$ nếu là nữ; 0 nếu nam.

$D_{3i} = 1$ nếu đã tốt nghiệp đại học; 0 nếu khác.

- Tương tác giữa 2 biến giả:

$$Y_i = \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \alpha_4 (D_{2i} D_{3i}) + \beta X_i + u_i$$

- $E(Y_i | X_i, D_{2i} = 1, D_{3i} = 1) = (\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4) + \beta X_i$

Sử dụng biến giả trong phân tích vụ mùa

$$\square Y_t = \alpha_1 + \alpha_2 D_{2t} + \alpha_3 D_{3t} + \alpha_4 D_{4t} + \beta X_t + u_t$$

Y = lợi nhuận; X = doanh thu

$D_{2t} = 1$ nếu là quý II; 0 nếu khác.

$D_{3t} = 1$ nếu là quý III; 0 nếu khác.

$D_{4t} = 1$ nếu là quý IV; 0 nếu khác.

$$\square Y_t = 6688 + 1323 D_{2t} - 218 D_{3t} + 184 D_{4t} + 0,038 X_t$$

t (3,9) (2,07) (-0,34) (0,28) (3,33) $R^2 = 0,52$

Sử dụng biến giả trong hồi qui tuyến tính từng khúc

$$\square Y_i = \alpha_1 + \beta_1 X_i + \beta_2 (X_i - X^*) D_i + u_i$$

Y = hoa hồng; X = doanh thu

$D_i = 1$ nếu $X_i > X^*$; 0 nếu $X_i < X^*$

→ Mức hoa hồng trung bình khi doanh thu thấp hơn hay bằng X^*

$$E(Y_i | X_i, X^*, D_i = 0) = \alpha_1 + \beta_1 X_i;$$

→ Mức hoa hồng trung bình khi doanh thu cao hơn X^*

$$E(Y_i | X_i, X^*, D_i = 1) = (\alpha_1 - \beta_2 X^*) + (\beta_1 + \beta_2) X_i;$$