

Chương 4:  
PHÂN TÍCH HỒI QUY VỚI BIẾN ĐỊNH TÍNH

Th.S NGUYỄN PHƯƠNG

Bộ môn Toán kinh tế  
Trường Đại học Ngân hàng TPHCM  
Blog: <https://nguyenphuongblog.wordpress.com>  
Email: [nguyenphuong0122@gmail.com](mailto:nguyenphuong0122@gmail.com)

Ngày 18 tháng 9 năm 2016

# NỘI DUNG

- 1 Khái niệm biến giả
- 2 Mô hình có chứa biến độc lập là biến giả
- 3 Mô hình có chứa biến định lượng và biến giả
- 4 Ứng dụng của biến giả

- Thu nhập, giá cả, chi tiêu cho một loại hàng, ...  $\rightarrow$  giá trị quan sát của các biến đó là những con số  $\rightarrow$  biến định lượng.
- Giá trị quan sát của biến không phải là số  $\rightarrow$  biến định tính

Biến định tính biểu thị các mức độ, các phạm trù khác nhau của một tiêu thức, một thuộc tính nào đó.

- ✓ Giới tính (nam, nữ);
  - ✓ Vùng miền (Bắc, Trung, Nam);
  - ✓ Khu vực sống (thành thị, nông thôn);...
- Để lượng hóa những biến định tính, trong phân tích hồi quy người ta sử dụng biến giả (dummy variable).
- Biến giả chỉ nhận hai giá trị là 0 và 1. Các con số này chỉ dùng để phản ánh hai nhóm quan sát mang tính chất khác nhau.

$$D = \begin{cases} 1 & \text{nếu là phạm trù A;} \\ 0 & \text{nếu không phải là phạm trù A} \end{cases}$$

## Ví dụ 1.1

- ✓ Giới tính (nam, nữ)  $\rightarrow D = \begin{cases} 1 & \text{nếu là nam;} \\ 0 & \text{nếu là nữ} \end{cases}$
  - ✓ Khu vực sống (thành thị, nông thôn)  $\rightarrow D = \begin{cases} 0 & \text{nếu là thành thị;} \\ 1 & \text{nếu là nông thôn} \end{cases}$
  - ✓ Vùng miền (Bắc, Trung, Nam)  $\rightarrow ?$
- ☞ Để phân biệt 2 mức độ (2 phạm trù)  $\rightarrow$  dùng 1 biến giả.
  - ☞ Để phân biệt 3 mức độ (3 phạm trù)  $\rightarrow$  dùng 2 biến giả.
  - ☞ Tổng quát, để phân biệt m mức độ (m phạm trù)  $\rightarrow$  dùng m - 1 biến giả.
  - ☞ Trạng thái cơ sở là trạng thái ứng với trường hợp mà tất cả các biến giả nhận giá trị 0  $\rightarrow$  Trạng thái cơ sở dùng để so sánh với các trạng thái khác.

## Ví dụ 2.1

Hồi quy thu nhập của công chức ( $Y$ ) phụ thuộc vào giới tính ( $D$ )

$$D_i = \begin{cases} 0 & \text{nếu công chức } i \text{ là nữ;} \\ 1 & \text{nếu công chức } i \text{ là nam} \end{cases}$$

Mô hình hồi quy tổng thể:  $Y_i = \beta_1 + \beta_2 D_i + U_i$

- $E(Y_i | D_i = 0) = \beta_1 \leftarrow$  Thu nhập trung bình của công chức nữ
- $E(Y_i | D_i = 1) = \beta_1 + \beta_2 \leftarrow$  Thu nhập trung bình của công chức nam
- $\beta_2 = E(Y_i | D_i = 1) - E(Y_i | D_i = 0) \rightarrow$  mức chênh lệch về thu nhập trung bình giữa nam và nữ
- Có sự phân biệt giới tính trong thu nhập?  $\rightarrow$  Kđgt  $\begin{cases} H_0 : \beta_2 = 0 \\ H_1 : \beta_2 \neq 0 \end{cases}$
- Thu nhập trung bình của nam có cao hơn nữ?  $\rightarrow$  Kđgt  $\begin{cases} H_0 : \beta_2 = 0 \\ H_1 : \beta_2 > 0 \end{cases}$

Hệ số của các biến giả được dùng để so sánh trạng thái đang xét với trạng thái cơ sở.

## Ví dụ 2.2

Giả sử hàm hồi quy tổng thể thu nhập của công chức ( $Y$ ) theo giới tính như sau:

$$Y_i = 5,4 + 1,2D_i + U_i$$

Hãy giải thích ý nghĩa các hệ số.

## Ví dụ 2.3

Hồi quy thu nhập của công chức ( $Y$ ) phụ thuộc vào khu vực làm việc (nông thôn; thành thị và miền núi).

$$D_{2i} = \begin{cases} 1 & \text{nếu công chức } i \text{ làm việc ở nông thôn;} \\ 0 & \text{nếu công chức } i \text{ làm việc ở khu vực khác;} \end{cases}$$

$$D_{3i} = \begin{cases} 1 & \text{nếu công chức } i \text{ làm việc ở thành thị;} \\ 0 & \text{nếu công chức } i \text{ làm việc ở khu vực khác} \end{cases}$$

Mô hình hồi quy tổng thể:  $Y_i = \beta_1 + \beta_2 D_{2i} + \beta_3 D_{3i} + U_i$

	D <sub>2</sub>	D <sub>3</sub>
Nông thôn	1	0
Thành thị	0	1
Miền núi	0	0

- ▶  $E(Y_i | D_{2i} = D_{3i} = 0) = \beta_1$  ← Thu nhập trung bình của công chức làm việc ở miền núi
- ▶  $E(Y_i | D_{2i} = 1, D_{3i} = 0) = \beta_1 + \beta_2$  ← Thu nhập trung bình của công chức làm việc ở nông thôn
- ▶  $E(Y_i | D_{2i} = 0, D_{3i} = 1) = \beta_1 + \beta_3$  ← Thu nhập trung bình của công chức làm việc ở thành thị
- ▶  $\beta_2$  ?
- ▶  $\beta_3$  ?
- ▶ Có sự khác biệt về thu nhập giữa công chức làm việc ở các khu vực khác nhau? → Kđgt  $\begin{cases} H_0 : \beta_2 = \beta_3 = 0 \\ H_1 : \beta_2 \neq 0 \vee \beta_3 \neq 0 \end{cases}$
- ▶ Kđgt  $\begin{cases} H_0 : \beta_j = 0 \\ H_1 : \beta_j \neq 0 \end{cases}$  ← ?