

Chủ đề 4: BIẾN ĐỘC LẬP ĐỊNH TÍNH (HOẶC BIẾN GIẢ)

Lê Kim Long
Phạm Thành Thái
Khoa Kinh tế - NTU

I. Hồi qui với biến độc lập đều là biến định tính.

1. Trường hợp các biến định tính chỉ có hai lựa chọn

Ví dụ, giữa hai ngôi nhà có cùng các đặc trưng, một có hồ bơi trong khi ngôi nhà còn lại không có. Tương tự, giữa hai nhân viên của một công ty có cùng tuổi, học vấn, kinh nghiệm,...một người là nam và người kia là nữ...

Để phát triển về mặt lý thuyết, chúng ta lấy ví dụ về lương và đặt Y_i là tiền lương hàng tháng của nhân viên thứ i trong công ty. Để đơn giản về mặt sự phạm, ở đây chúng ta bỏ qua các biến khác có ảnh hưởng đến lương và chỉ tập trung vào giới tính. Vì biến giới tính không phải là một biến định lượng một cách trực tiếp được nên chúng ta định nghĩa một biến giả gọi là D (Dummy variables), biến giả này là biến nhị nguyên chỉ nhận giá trị 1 với nam nhân viên và 0 với nữ nhân viên. **Lưu ý là cách định nghĩa này là hoàn toàn ngẫu nhiên. Nhóm mà giá trị D bằng 0 gọi là nhóm điều khiển (Control group).**

I. Hồi qui với biến độc lập đều là biến định tính.

1. Trường hợp các biến định tính chỉ có hai lựa chọn

Bây giờ chúng ta sẽ thiết lập và ước lượng một mô hình sử dụng biến giả như một biến giải thích. Dạng đơn giản nhất của mô hình như sau:

$$\text{[Blank Box]} \quad (6.1)$$

Chúng ta giả sử là sai số ngẫu nhiên thỏa mãn các giả thiết của mô hình hồi qui tuyến tính cổ điển. Chúng ta có thể lấy kỳ vọng có điều kiện của Y với D cho trước và được các phương trình sau:

Đối với nam: [Blank Box] (6.2)

Đối với nữ: $E(Y / D = 0) = \beta_1$ (6.3)

I. Hồi qui với biến độc lập đều là biến định tính.

1. Trường hợp các biến định tính chỉ có hai lựa chọn

Vậy, β_1 là lương trung bình của nhóm điều khiển (nhân viên nữ) và β_2 là khác biệt kỳ vọng của lương trung bình của hai nhóm cho cả tổng thể (chênh lệch về lương trung bình của một nhân viên nam so với nhân viên nữ).

Để xét xem giữa hai nhân viên có sự phân biệt về giới hay không ta tiến hành kiểm định giả thiết $H_0: \beta_2=0$ và $H_1: \beta_2 \neq 0$. Kiểm định thích hợp là kiểm định t với bậc tự do $df = n-2$.

Lưu ý: Thủ tục ước lượng phương trình (6.1) được tiến hành bình thường như những mô hình ở các chương trước bằng phương pháp OLS.

I. Hồi qui với biến độc lập đều là biến định tính.

2. Trường hợp các biến định tính có nhiều hơn hai lựa chọn

Số các lựa chọn có thể có của một biến định tính có thể nhiều hơn hai. Xét ví dụ sau đây:

Gọi Y_i là tiền tiết kiệm của một hộ gia đình thứ i . Chúng ta kỳ vọng rằng các hộ gia đình thuộc các nhóm tuổi khác nhau sẽ có mức tiết kiệm khác nhau. Nếu chúng ta có tuổi chính xác của người chủ hộ, biến này có thể đưa vào mô hình như là biến định lượng. Tuy nhiên, nếu chúng ta chỉ có nhóm tuổi (ví dụ người chủ hộ thuộc nhóm tuổi dưới 25, từ 25 đến 55 và trên 55), chúng ta xem xét biến định tính "nhóm tuổi của chủ hộ" như thế nào?. Thủ tục ở đây là chọn một trong những nhóm này làm nhóm kiểm soát và xác định các biến giả cho hai nhóm còn lại. Cụ thể hơn, chúng ta định nghĩa:

I. Hồi qui với biến độc lập đều là biến định tính.

2. Trường hợp các biến định tính có nhiều hơn hai lựa chọn

$$D_{1i} = \begin{cases} 1 & \text{nếu chủ hộ từ 25 đến 55 tuổi} \\ 0 & \text{nếu chủ hộ thuộc nhóm tuổi khác} \end{cases}$$

$$D_{2i} = \begin{cases} 1 & \text{nếu chủ hộ trên 55 tuổi} \\ 0 & \text{nếu chủ hộ thuộc nhóm tuổi khác} \end{cases}$$

Nhóm điều khiển là nhóm mà cả D_{1i} và D_{2i} đều bằng 0, có nghĩa là tất cả những hộ gia đình mà chủ hộ dưới 25 tuổi. Một mô hình đơn giản mô tả quan hệ giữa tiền tiết kiệm và nhóm tuổi như sau:



I. Hồi qui với biến độc lập đều là biến định tính.

2. Trường hợp các biến định tính có nhiều hơn hai lựa chọn

Giả sử mô hình thỏa mãn các giả thiết của mô hình hồi qui tuyến tính cổ điển. Chúng ta lấy kỳ vọng có điều kiện của Y với điều kiện của các biến D cho trước ta được các mô hình sau:

Đối với hộ gia đình dưới 25 tuổi:  (6.4)

Đối với hộ gia đình từ 25 đến 55 tuổi:  (6.5)

Đối với hộ gia đình trên 55 tuổi:  (6.6)

I. Hồi qui với biến độc lập đều là biến định tính.

Như vậy, (6.4) cho chúng ta biết tiền tiết kiệm trung bình một tháng của một hộ gia đình mà chủ hộ dưới 25 tuổi. Tương tự, (6.5) cho chúng ta biết tiền tiết kiệm trung bình một tháng của một hộ gia đình mà chủ hộ từ 25 đến 55 tuổi và (6.6) cho chúng ta biết tiền tiết kiệm trung bình một tháng của một hộ gia đình mà chủ hộ trên 55 tuổi.

- β_2 biểu thị chênh lệch về tiền tiết kiệm trung bình một tháng của một hộ gia đình mà chủ hộ thuộc nhóm tuổi từ 25 đến 55 tuổi so với nhóm tuổi dưới 25.

- β_3 biểu thị chênh lệch về tiền tiết kiệm trung bình một tháng của một hộ gia đình mà chủ hộ thuộc nhóm tuổi trên 55 so với nhóm tuổi dưới 25.

Chú ý: Để tránh trường hợp bẫy biến giả (dummy variables trap), số các biến giả luôn luôn ít hơn một biến so với số các lựa chọn.

II. Hồi qui với các biến độc lập định lượng và các biến định tính.

1. Trường hợp có một biến định tính với 2 lựa chọn:

Bước tiếp theo trong phân tích là thêm các biến độc lập có thể định lượng được. Để minh họa, ta lấy lại ví dụ về tiền lương ở trên.

- Đặt Y_i : là tiền lương hàng tháng của nhân viên thứ i .
- Gọi X_i : là kinh nghiệm của nhân viên thứ i
- Với $D_i = 1$: nhân viên nam
- Với $D_i = 0$: nhân viên nữ

Bây giờ, ta bỏ qua yếu tố giới tính, chỉ xét mối quan hệ giữa tiền lương hàng tháng và kinh nghiệm, một mô hình hồi quy đơn cho quan hệ này là:

$$\text{[Blank Box]}$$

(6.7)

II. Hồi qui với các biến độc lập định lượng và các biến định tính.

1. Trường hợp có một biến định tính với 2 lựa chọn:

Lưu ý là bây giờ ta có thể kiểm soát được kinh nghiệm và có thể hỏi "Giữa hai nhân viên có cùng kinh nghiệm, có sự khác biệt trong tiền lương tháng do giới tính không?". Một cách đơn giản để trả lời câu hỏi này là đặt tung độ gốc β_1 trong phương trình (6.7) khác nhau đối với nam và nữ. Thực hiện việc này bằng cách giả sử là: $\beta_1 = \beta_3 + \beta_4 D_i$ và thay vào (6.7) ta có mô hình kinh tế lượng:

$$\text{[Blank Box]} \quad (6.8)$$

Lưu ý là β_2 , β_3 và β_4 được ước lượng bằng cách hồi quy Y theo một hằng số, biến D và X . Các quan hệ được ước lượng cho hai nhóm là:

II. Hồi qui với các biến độc lập định lượng và các biến định tính.

1. Trường hợp có một biến định tính với 2 lựa chọn:

Đối với Nữ:



(6.9)

Đối với Nam:



(6.10)

Một giả thiết tự nhiên cần kiểm định là "không có sự khác biệt trong quan hệ giữa hai nhóm". So sánh phương trình (6.9) và (6.10), chúng ta thấy là các quan hệ sẽ như nhau nếu $\beta_4 = 0$. Vì vậy, chúng ta kiểm định giả thiết $H_0: \beta_4 = 0$ và $H_1: \beta_4 \neq 0$. Kiểm định thích hợp nhất là kiểm định t với bậc tự do là $df = n - 3$.

II. Hồi qui với các biến độc lập định lượng và các biến định tính.

2. Trường hợp có một biến định tính với nhiều hơn 2 lựa chọn:

Lấy lại ví dụ về tiền tiết kiệm của các hộ gia đình ở trên, bây giờ ta thêm một biến định lượng vào mô hình đó là thu nhập của hộ (được ký hiệu là X). Từ đó, chúng ta có thể xây dựng một mô hình kinh tế lượng như sau:



Giả sử mô hình thỏa mãn các giả thiết của mô hình hồi qui tuyến tính cổ điển. Chúng ta lấy kỳ vọng có điều kiện của Y với điều kiện đã cho của thu nhập X và của các biến D cho trước ta được các mô hình sau:

II. Hồi qui với các biến độc lập định lượng và các biến định tính.

2. Trường hợp có một biến định tính với nhiều hơn 2 lựa chọn:

Đối với hộ gia đình dưới 25 tuổi:

$$\text{[Blank Box]} \quad (6.11)$$

Đối với hộ gia đình từ 25 đến 55 tuổi:

$$\text{[Blank Box]} \quad (6.12)$$

Đối với hộ gia đình trên 55 tuổi:

$$\text{[Blank Box]} \quad (6.13)$$

II. Hồi qui với các biến độc lập định lượng và các biến định tính.

2. Trường hợp có một biến định tính với nhiều hơn 2 lựa chọn:

Một số giả thuyết rất thú vị. Để kiểm định giả thuyết gia đình ở nhóm tuổi cao hơn có hành vi giống gia đình ở nhóm tuổi trẻ hơn, chúng ta đơn giản chỉ tiến hành kiểm định t đối với các hệ số β_2 hoặc β_3 bằng 0. Để kiểm định giả thuyết *"không có sự khác biệt trong hàm tiết kiệm theo độ tuổi"*, giả thuyết là $H_0: \beta_2 = \beta_3 = 0$ và giả thuyết đối là H_1 : ít nhất một trong các hệ số khác không. Giả thuyết này được kiểm định bằng kiểm định Wald. Hoặc để kiểm định giả thuyết *"không có sự khác biệt trong hành vi giữa hai nhóm tuổi trung niên và cao tuổi"*, nghĩa là $\beta_2 = \beta_3$. Giả thuyết này có thể được kiểm định bằng nhiều phương pháp khác nhau.

II. Hồi qui với các biến độc lập định lượng và các biến định tính.

3. Trường hợp có nhiều hơn một biến định tính.

Phân tích biến giả dễ dàng mở rộng cho trường hợp trong đó có nhiều biến định tính. Để minh họa, hãy xem xét hàm tiết kiệm được mô tả trước đây, trong đó Y là tiền tiết kiệm của hộ gia đình và X là thu nhập của hộ gia đình. Có thể đưa ra giả thuyết là ngoài tuổi của chủ hộ, các yếu tố khác như sở hữu nhà, trình độ học vấn, tình trạng nghề nghiệp,... cũng là các yếu tố xác định tiết kiệm của hộ gia đình.

Giả sử ta có thông tin là chủ hộ có trình độ sau đại học, có trình độ đại học, chỉ tốt nghiệp trung học; Chủ hộ có thể làm một trong những nghề sau: quản lý, công nhân lành nghề, công nhân không có tay nghề, thư ký, kinh doanh tự do. Cũng tương tự, ta không biết chính xác tuổi của chủ hộ nhưng biết được ông/bà ta thuộc nhóm tuổi nào,...

II. Hồi qui với các biến độc lập định lượng và các biến định tính.

3. Trường hợp có nhiều hơn một biến định tính.

$$D_{1i} = \begin{cases} 1 & \text{nếu chủ hộ từ 25 đến 55 tuổi} \\ 0 & \text{nếu chủ hộ thuộc nhóm tuổi khác} \end{cases}$$

$$D_{2i} = \begin{cases} 1 & \text{nếu chủ hộ trên 55 tuổi} \\ 0 & \text{nếu chủ hộ thuộc nhóm tuổi khác} \end{cases}$$

$$D_{3i} = \begin{cases} 1 & \text{nếu chủ hộ sở hữu căn nhà} \\ 0 & \text{nếu điều kiện khác} \end{cases}$$

II. Hồi qui với các biến độc lập định lượng và các biến định tính.

3. Trường hợp có nhiều hơn một biến định tính.

$$D_{4i} = \begin{cases} 1 & \text{nếu chủ hộ có trình độ sau đại học} \\ 0 & \text{nếu điều kiện khác} \end{cases}$$

$$D_{5i} = \begin{cases} 1 & \text{nếu chủ hộ có trình độ đại học} \\ 0 & \text{nếu điều kiện khác} \end{cases}$$

II. Hồi qui với các biến độc lập định lượng và các biến định tính.

3. Trường hợp có nhiều hơn một biến định tính.

$$D_{6i} = \begin{cases} 1 & \text{nếu chủ hộ là nhà quản lý} \\ 0 & \text{nếu điều kiện khác} \end{cases}$$

$$D_{7i} = \begin{cases} 1 & \text{nếu chủ hộ là công nhân lành nghề} \\ 0 & \text{nếu điều kiện khác} \end{cases}$$

II. Hồi qui với các biến độc lập định lượng và các biến định tính.

3. Trường hợp có nhiều hơn một biến định tính.

$$D_{8i} = \begin{cases} 1 & \text{nếu chủ hộ là thư ký} \\ 0 & \text{nếu điều kiện khác} \end{cases}$$

$$D_{9i} = \begin{cases} 1 & \text{nếu chủ hộ kinh doanh tự do} \\ 0 & \text{nếu điều kiện khác} \end{cases}$$

II. Hồi qui với các biến độc lập định lượng và các biến định tính.

3. Trường hợp có nhiều hơn một biến định tính.

Một mô hình kinh tế lượng được xây dựng như sau:



Nên lưu ý là đặc điểm của nhóm điều khiển như sau: chủ hộ có độ tuổi dưới 25, là công nhân không có tay nghề, với trình độ học vấn chỉ ở bậc trung học và không sở hữu nhà. Một cách rất dễ dàng, chúng ta cũng có thể ước lượng mô hình trên và tính được tiền tiết kiệm trung bình của một hộ gia đình với những điều kiện của biến X và các biến D cho trước và nêu ý nghĩa của nó.