

Chương 4: PHÂN TÍCH HỒI QUY VỚI BIẾN ĐỊNH TÍNH

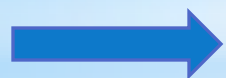
1. KHÁI NIỆM BIẾN GIẢ
2. MÔ HÌNH CÓ CHỨA BIẾN ĐỘC LẬP LÀ BIẾN GIẢ
3. MÔ HÌNH VỚI BIẾN GIẢ VÀ BIẾN TƯƠNG TÁC
4. BIẾN ĐỊNH TÍNH CÓ NHIỀU PHẠM TRÙ

4.1 KHÁI NIỆM BIẾN GIẢ

Trong phân tích kinh tế, nhiều tình huống biến phụ thuộc chịu tác động của các biến định lượng và các yếu tố định tính.

Ví dụ 1:

- 1) Chi tiêu vào thời trang của nam và nữ khác nhau.
- 2) Lạm phát ở các quốc gia chịu ảnh hưởng bởi các bất ổn chính trị của quốc gia đó.
- 3) Chi tiêu cho y tế của các hộ gia đình phụ thuộc vào mùa trong năm.



Thông tin định tính được đưa vào mô hình hồi quy thông qua biến giả (dummy variable), chỉ nhận hai giá trị là 0 và 1.

4.1 KHÁI NIỆM BIẾN GIẢ

Ví dụ 1:

1) $GT = \begin{cases} 1 & \text{nếu quan sát là nữ.} \\ 0 & \text{nếu quan sát là nam} \end{cases}$

2) $D = \begin{cases} 1 & \text{nếu quốc gia có bất ổn chính trị.} \\ 0 & \text{nếu quốc gia không có bất ổn chính trị.} \end{cases}$

3) $X = \begin{cases} 1 & \text{nếu quan sát vào mùa xuân.} \\ 0 & \text{nếu quan sát không vào mùa xuân.} \end{cases}$

$H = \begin{cases} 1 & \text{nếu quan sát vào mùa hè.} \\ 0 & \text{nếu quan sát không vào mùa hè.} \end{cases}$

$T = \begin{cases} 1 & \text{nếu quan sát vào mùa thu.} \\ 0 & \text{nếu quan sát không vào mùa thu} \end{cases}$

4.1 KHÁI NIỆM BIẾN GIẢ

Biến định tính Z nào đó có 2 phạm trù A và A'.

Đặt biến giả :

$$D = \begin{cases} 1 & \text{nếu quan sát thuộc phạm trù A.} \\ 0 & \text{nếu quan sát thuộc phạm trù A'.} \end{cases}$$

Ví dụ 2: Ước lượng mô hình hồi quy chi tiêu vào thời trang phụ thuộc vào giới tính như sau: $CT = 1,02 + 0,32 GT + e$.

Trong đó, CT : Chi tiêu vào thời trang – triệu đồng/tháng.

GT = 1 nếu một người là nữ.

Và 0 nếu một người là nam.

- a) Hãy giải thích ý nghĩa của các hệ số ước lượng.
- b) Bằng cách nào để khẳng định có sự khác nhau về chi tiêu vào thời trang giữa hai giới tính?
- c) Nếu đặt $GT = 0$ nếu một người là nữ và bằng 1 nếu là nam thì mô hình thay đổi như thế nào?

4.2 MÔ HÌNH CÓ CHỨA BIẾN ĐỘC LẬP LÀ BIẾN GIẢ

Xét mô hình hồi quy tuyến tính k biến có chứa biến giả D thể hiện biến định tính Z: $Y = \beta_1 + \beta_2 D + \beta_3 X_3 + \dots + \beta_k X_k + U$

- Hệ số β_2 thể hiện sự khác biệt giữa giá trị trung bình của Y trong nhóm các quan sát thuộc các phạm trù khác nhau của yếu tố định tính Z, khi các biến X_j là như nhau.
- Đây là mô hình hồi quy tuyến tính k biến, vẫn đòi hỏi các giả thiết GT1,2,3,4,5 được thỏa mãn để cho ra các ước lượng tốt nhất.

Ví dụ 3: Ước lượng hàm hồi quy chi tiêu vào thời trang phụ thuộc vào giới tính và thu nhập như sau:

$$CT = 1,02 + 0,25GT + 0,33TN + e.$$

Trong đó, CT: Chi tiêu vào thời trang – triệu đồng/tháng.

TN: thu nhập – triệu đồng/tháng

GT = 1 nếu quan sát là nữ.

Và 0 nếu quan sát là nam.

Hãy giải thích ý nghĩa của các hệ số ước lượng.

4.3 MÔ HÌNH VỚI BIẾN GIẢ VÀ BIẾN TƯƠNG TÁC

Nếu có sự khác biệt về hệ số chặn và hệ số góc của hai nhóm quan sát ứng với hai phạm trù A và A' của biến định tính Z, ta đưa thêm biến tương tác giữa biến giả và biến độc lập vào mô hình.

Ví dụ 4: Tiếp theo ví dụ 3, giả sử nhận thấy chi tiêu biên cho thời trang theo thu nhập phụ thuộc vào giới tính. Khi đó, ta có mô hình hồi quy:

$$CT = \beta_1 + \beta_2 GT + \beta_3 TN + \beta_4 GT * TN + U$$

Biến
tương tác

Xét mô hình hồi quy: $CT = \beta_1 + \beta_2 GT + \beta_3 TN + \beta_4 GT * TN + U$

Làm thế nào để kết luận có sự khác biệt trong chi tiêu về thời trang giữa hai nhóm giới tính?

Nếu có ít nhất một trong hai hệ số β_2 và β_4 khác 0 thì kết luận có sự khác biệt.



Thực hiện kiểm định cặp giả thuyết:

$$H_0: \beta_2 = \beta_4 = 0$$

$$H_1: \beta_2^2 + \beta_4^2 > 0$$

Kiểm định Wald (kiểm định đồng thời nhiều hệ số)

Xét hai mô hình

$$(F): Y = \beta_1 + \beta_2 X_2 + \dots + \beta_k X_k + U$$

$$(R): Y = \beta_1 + \beta_{m+1} X_{m+1} + \dots + \beta_k X_k + U$$

Thực hiện kiểm định cặp giả thuyết:

$$H_0: \beta_2 = \dots = \beta_m = 0$$

$$H_1: \beta_2^2 + \dots + \beta_m^2 > 0$$

$$F = \frac{C1}{C2} = \frac{\frac{RSS_R - RSS_F}{m - 1}}{\frac{RSS_F}{n - k}} = \frac{\frac{R_F^2 - R_R^2}{m - 1}}{\frac{1 - R_F^2}{n - k}} \quad F > F_\alpha(m - 1; n - k) \Rightarrow$$

Bác bỏ giả thuyết
 H_0

Thực hiện nghiên cứu giá một căn hộ chung cư (GIA - tỷ đồng) phụ thuộc vào diện tích (DT - m²) của căn hộ và nơi tọa lạc của chung cư (TT, TT = 1 nếu chung cư ở trung tâm thành phố và TT = 0 nếu không ở trung tâm) trên mẫu 40 quan sát thu được kết quả như sau:

$$\text{GIA} = 1,042 + 0,025\text{DT} + 0,037\text{TT} + 0,019\text{TT}*\text{DT} + e$$

$$(\text{se}) \quad (0,257) \quad (0,003) \quad (0,009) \quad (0,013)$$

$$R^2 = 0,7925$$

- a) Hãy giải thích ý nghĩa hệ số ước lượng của biến TT*DT.
- b) Có ý kiến cho rằng, giá căn hộ chung cư không phụ thuộc vào việc có ở trung tâm thành phố hay không. Với mức ý nghĩa 5%, hãy

cho kết luận về nhận xét đó. Biết rằng hàm hồi quy mẫu giá căn hộ theo diện tích có hệ số xác định là 0,7048.
Cho biết

$$f_{0,05}(2,36) = 3,26; f_{0,05}(2,37) = 3,25$$

4.4 BIẾN ĐỊNH TÍNH CÓ NHIỀU PHẠM TRÙ

Mô hình hồi quy cần $(n - 1)$ biến giả $0 - 1$.

Trường hợp

biến định

tính có n

phạm trú

$(n > 2)$,

Phạm trú cơ sở là phạm trú ứng với trường hợp tất cả các biến giả nhận giá trị 0.

Hệ số của biến giả dùng để so sánh hệ số chặn của nhóm với nhóm cơ sở.

Hệ số của biến tương tác dùng để đánh giá sự khác biệt trong tác động của biến độc lập lên biến phụ thuộc giữa một nhóm với nhóm cơ sở.

Ví dụ 6: Khi nghiên cứu chi tiêu cho chăm sóc sức khỏe của hộ gia đình, ước lượng được hàm hồi quy mẫu sau đây:

$$CT = 2,9 + 0,31TN + 0,12X + 0,22H + 0,17T + 0,11X*TN + e$$


Trong đó: CT: chi tiêu (triệu đồng/tháng);

TN: thu nhập (triệu đồng/tháng);

$X = \begin{cases} 1 & \text{nếu quan sát vào mùa xuân.} \\ 0 & \text{nếu quan sát không vào mùa xuân.} \end{cases}$

$H = \begin{cases} 1 & \text{nếu quan sát vào mùa hè.} \\ 0 & \text{nếu quan sát không vào mùa hè.} \end{cases}$

$T = \begin{cases} 1 & \text{nếu quan sát vào mùa thu.} \\ 0 & \text{nếu quan sát không vào mùa thu} \end{cases}$




Hãy giải thích ý nghĩa của các hệ số hồi quy?

4.4 BIẾN ĐỊNH TÍNH CÓ NHIỀU PHẠM TRÙ

Khi biến định tính là biến chỉ thứ bậc

Ví dụ:

- 1) Trình độ học vấn của người dân: dưới phổ thông trung học, phổ thông trung học, đại học và trên đại học.
- 2) Mức độ hài lòng khi sử dụng dịch vụ: rất hài lòng, hài lòng, bình thường, không hài lòng.



Sử dụng biến giả đại diện cho biến định tính nhiều phạm trù.

Ví dụ 7

Khảo sát 40 nông dân để nghiên cứu thu nhập (TN – triệu đồng/tháng) phụ thuộc vào diện tích đất canh tác (DT – ha) và nơi sinh sống (Miền Bắc, Trung và Nam), thu được hàm hồi quy mẫu như sau: **$TN = 2,93 - 0,46TRUNG + 0,84NAM + 9,62DT + e$** .

(se) (0,87) (0,18)

(0,35)

(3,43)

Trong đó:

*TRUNG=1 nếu nông dân đó sinh sống ở miền Trung và TRUNG =0 nếu sinh sống ở miền khác.

*NAM =1 nếu nông dân đó sinh sống ở miền Nam và NAM =0 nếu sinh sống ở miền khác.

a) Hãy giải thích ý nghĩa các hệ số ước lượng của biến TRUNG và NAM.

b) Khi không có đất để canh tác, thu nhập trung bình của nông dân miền Bắc vào khoảng bao nhiêu với độ tin cậy 95%?

*Cho biết: $t_{0,025;37} = 2,026$; $t_{0,025;36} = 2,028$

Ví dụ 8

- a) Hãy đề xuất mô hình nghiên cứu thu nhập (TN –triệu đồng/tháng) của người lao động phụ thuộc vào số năm đi làm (THAMNIEN – năm) và trình độ học vấn (Sau đại học, đại học, phổ thông trung học và dưới phổ thông trung học).
- b) Nhận thấy rằng, khi số năm đi làm tăng lên, mức tăng thu nhập của người lao động phụ thuộc vào trình độ học vấn. Hãy thiết kế mô hình nghiên cứu và kiểm định vấn đề đó.
- c) Làm thế nào để kiểm tra sự phụ thuộc của thu nhập vào trình độ học vấn?