

Chương 4

HỒI QUY VỚI BIẾN GIẢ

By Tuan Anh(UEH)

I. BẢN CHẤT CỦA BIẾN GIẢ

Biến định lượng : giá trị thể hiện bằng những con số

Ví dụ : *Thu nhập, chi tiêu, chi phí, doanh thu, v.v...*

Biến định tính: giá trị không thể hiện bằng những con số

Ví dụ : *Giới tính, màu sắc, tôn giáo, chất liệu, v.v...*

By Tuan Anh(UEH)

I. BẢN CHẤT CỦA BIẾN GIẢ

Biến định tính thường biểu thị có hay không có một tính chất hoặc là các mức độ khác nhau của một tiêu thức thuộc tính nào đó

Để lượng hoá các biến định tính, trong phân tích hồi quy người ta dùng biến giả (***dummy variables***)

By Tuan Anh(UEH)

II. 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 : giới tính : - *Nam*
- *Nữ*

Ngôi nhà : - *Mặt tiền*
- *Không phải mặt tiền*

Khu vực bán hàng : - *Thành thị*
- *Nông thôn*

By Tuan Anh(UEH)

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

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

Giả sử : Chúng ta muốn nghiên cứu tiền lương tại một doanh nghiệp có bị ảnh hưởng bởi vấn đề giới tính hay không ? (Tức là có sự khác biệt tiền lương giữa nhân viên nam và nữ hay không ?)

Giới tính là biến định tính nên ta dùng biến giả D_i

Với $D_i = 1$: Nam

$D_i = 0$: Nữ

By Tuan Anh(UEH)

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

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

Hàm hồi qui có dạng :
 $PRF : Y_i = \beta_1 + \beta_2 D_i + U_i$
 $SRF : \hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 D_i$

Thu thập số liệu :

| Y_i (trđ/tháng) | D_i |
|-------------------|-------|
| 5,0 | 1 |
| 4,0 | 0 |
| 3,8 | 0 |
| 3,5 | 1 |
| ... | ... |

By Tuan Anh(UEH)

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

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

Tiến hành hồi quy nghư hàm hai biến, giả sử ta được ước lượng của hàm hồi quy sau :

$$Y_i = \beta_1 + \beta_2 D_i + U_i$$

Tạm thời bỏ qua sai số U_i

Đối với nữ: $D = 0 \Rightarrow Y = \beta_1$

Đối với nam: $D = 1 \Rightarrow Y = \beta_1 + \beta_2$

By Tuan Anh(UEH)

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

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

Lưu ý: Lựa chọn được gán với giá trị $D_i = 0$ trở thành "lựa chọn cơ sở" hay còn gọi là "nhóm điều khiển"

Tóm lại :

β_1 là

β_2 là

By Tuan Anh(UEH)

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

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

- Vậy làm thế nào để xét xem tại doanh nghiệp này có sự khác biệt về tiền lương giữa nhân viên nam và nữ hay không ?

Ta kiểm định giả thiết

$$H_0: \beta_2 = 0 \quad (\text{độ tin cậy } 1-\alpha)$$

$$H_1: \beta_2 \neq 0.$$

- Kiểm định bằng cách nào?

- Nếu ta đặt $D_i = 1$ là nữ thì có được không? Mô hình thay đổi như thế nào ?

By Tuan Anh(UEH)

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

1. TH 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. Có hai cách :

- Dùng biến giả có nhiều giá trị, số giá trị bằng với số lựa chọn
- Dùng nhiều biến giả, mỗi biến có giá trị 0 và 1.

Cách 2 được khuyến khích hơn

Chú ý: Để không rơi vào bẫy biến giả thì

$$\text{số các biến giả} = \text{số lựa chọn} - 1$$

By Tuan Anh(UEH)

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

1. TH biến định tính có nhiều hơn hai lựa chọn

Ví dụ : Nghiên cứu tiền lương khi ra trường của sinh viên có phụ thuộc vào kết quả tốt nghiệp hay không

Kết quả tốt nghiệp gồm :

- Xuất sắc
- Giỏi
- Khá
- Trung bình
- Yếu kém

Sẽ có bao nhiêu biến giả được đưa vào mô hình ?

By Tuan Anh(UEH)

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

1. TH biến định tính có nhiều hơn hai lựa chọn

Ta đưa 4 biến giả như sau:

$$D_{2i} = \begin{cases} 1 & \text{SV xuất sắc} \\ 0 & \text{khác} \end{cases} \quad D_{4i} = \begin{cases} 1 & \text{SV khá} \\ 0 & \text{khác} \end{cases}$$

$$D_{3i} = \begin{cases} 1 & \text{SV giỏi} \\ 0 & \text{khác} \end{cases} \quad D_{5i} = \begin{cases} 1 & \text{SV yếu kém} \\ 0 & \text{khác} \end{cases}$$

Lưu ý: Nhóm ứng với giá trị $D_{2i}=D_{3i}=D_{4i}=D_{5i}= 0$ là

nhóm điều khiển

By Tuan Anh(UEH)

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

1. TH biến định tính có nhiều hơn hai lựa chọn

Thu thập số liệu, ví dụ :

| Y_i (trđ/tháng) | D_{2i} | D_{3i} | D_{4i} | D_{5i} |
|----------------------|----------|----------|----------|----------|
| 5,0 | 1 | 0 | 0 | 0 |
| 4,0 | 0 | 1 | 0 | 0 |
| 3,8 | 0 | 0 | 1 | 0 |
| 3,5 | 0 | 0 | 0 | 0 |
| ... | | | | ... |

By Tuan Anh(UEH)

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

1. TH biến định tính có nhiều hơn hai lựa chọn

Một mô hình đơn giản mô tả quan hệ giữa tiền lương và loại tốt nghiệp như sau :

$$Y_i = \beta_1 + \beta_2 D_{2i} + \beta_3 D_{3i} + \beta_4 D_{4i} + \beta_5 D_{5i} + U_i$$

- Ý nghĩa của β_1 là gì?

- Ý nghĩa của $\beta_2, \beta_3, \beta_4, \beta_5$ là gì?

By Tuan Anh(UEH)

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

1. Một biến định tính và một biến định lượng

Quay lại ví dụ về tiền lương, ta muốn kiểm tra xem liệu doanh nghiệp có tăng lương cho nhân viên theo thâm niên, đồng thời cũng muốn kiểm tra xem có phân biệt tiền lương theo giới tính hay không ?

Ta lập mô hình hồi quy với các biến như sau

- Y : là tiền lương hàng tháng của nhân viên
- X : Số năm kinh nghiệm
- Biến giả D với $D_i = 1$: nhân viên nam
 $D_i = 0$: nhân viên nữ

By Tuan Anh(UEH)

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

1. Một biến định tính và một biến định lượng

Hàm hồi quy: $Y_i = \beta_1 + \beta_2 X_i + \beta_3 D_i + U_i$

| Y_i (trđ/tháng) | X_i | D_i |
|-------------------|-------|-------|
| 5,0 | 10 | 1 |
| 4,0 | 8 | 0 |
| 3,8 | 5 | 0 |
| 3,5 | 5 | 1 |
| ... | | ... |

Tiến hành hồi quy *nghư* hàm ba biến

By Tuan Anh(UEH)

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

1. Một biến định tính và một biến định lượng

Hàm hồi quy: $Y_i = \beta_1 + \beta_2 X_i + \beta_3 D_i + U_i$

- Ý nghĩa của β_1 là gì?

- Ý nghĩa của β_2 là gì?

- Ý nghĩa của β_3 là gì?

By Tuan Anh(UEH)

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

1. Một biến định tính và một biến định lượng

Làm thế nào để kiểm tra tiền lương có bị ảnh hưởng bởi số năm kinh nghiệm hay không?

Làm thế nào để kiểm tra tiền lương có bị ảnh hưởng bởi giới tính hay không?

By Tuan Anh(UEH)

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

1. Một biến định tính và một biến định lượng

Hàm hồi quy: $Y_i = \beta_1 + \beta_2 X_i + \beta_3 D_i + U_i$

$D_i = 0 \Rightarrow$ Hàm hồi quy của nhân viên nữ

$$Y_i = \beta_1 + \beta_2 X_i + U_i$$

$D_i = 1 \Rightarrow$ Hàm hồi quy của nhân viên nam

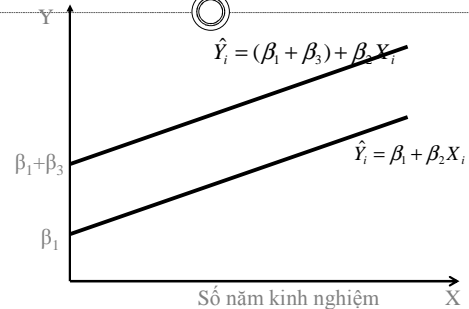
$$Y_i = (\beta_1 + \beta_3) + \beta_2 X_i + U_i$$

Có thể nhận xét gì từ hai hàm hồi quy trên? (xem đồ thị)

By Tuan Anh(UEH)

Hàm hồi quy của NV nữ $Y_i = \beta_1 + \beta_2 X_i + U_i$

Hàm hồi quy của NV nam $Y_i = (\beta_1 + \beta_3) + \beta_2 X_i + U_i$



By Tuan Anh(UEH)

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

1. Một biến định tính và một biến định lượng

Làm sao để biết tốc độ tăng lương có khác nhau giữa nam và nữ hay không?

Ta sử dụng dạng hàm hồi quy:

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 D_i + \beta_4 X_i D_i + U_i$$

Khi đó biến $X_i D_i$ được gọi là *biến tương tác* giữa X và D

By Tuan Anh(UEH)

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 D_i + \beta_4 X_i D_i + U_i$$

$D_i = 0 \Rightarrow$ Hàm hồi quy của nhân viên nữ

$D_i = 1 \Rightarrow$ Hàm hồi quy của nhân viên nam

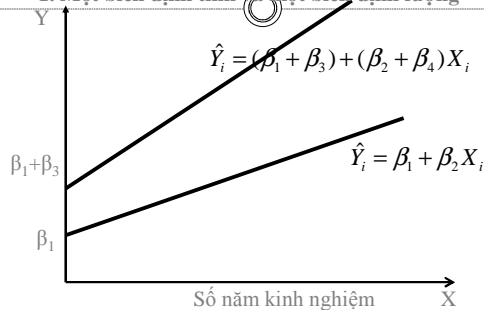
- Ý nghĩa của β_1 là gì? - Ý nghĩa của β_3 là gì?

- Ý nghĩa của β_2 là gì? - Ý nghĩa của β_4 là gì?

By Tuan Anh(UEH)

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

1. Một biến định tính và một biến định lượng



By Tuan Anh(UEH)

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

1. Một biến định tính và một biến định lượng

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 D_i + \beta_4 X_i D_i + U_i$$

Từ hàm hồi quy này làm sao để biết tốc độ tăng lương có khác nhau giữa nam và nữ hay không?

By Tuan Anh(UEH)

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

2. Nhiều biến định tính và nhiều biến định lượng

Nếu mô hình có nhiều biến định tính, chúng ta có thể xác định số biến giả được đưa vào mô hình như sau:

$$n = \sum_{i=1}^k (n_i - 1)$$

Trong đó: n - là số biến giả cần thiết đưa vào mô hình
 k - là số biến định tính
 n_i - là số lựa chọn của biến định tính thứ i

By Tuan Anh(UEH)

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

2. Nhiều biến định tính và nhiều biến định lượng

Ví dụ : Khảo sát các yếu tố ảnh hưởng đến kết quả học tập của sinh viên.

By Tuan Anh(UEH)

Ví dụ minh họa

Cho số liệu giả thiết về mức lương của nhân viên (Y-trđ/năm), số năm kinh nghiệm giảng dạy (X) và giới tính (D=1:nam; D=0:nữ)

| Y _i | X _i | D _i | Y _i | X _i | D _i |
|----------------|----------------|----------------|----------------|----------------|----------------|
| 115 | 11 | 1 | 125 | 15 | 0 |
| 95 | 9 | 0 | 140 | 15 | 1 |
| 120 | 10 | 1 | 147 | 16 | 1 |
| 105 | 12 | 0 | 130 | 16 | 0 |
| 125 | 13 | 1 | 128 | 17 | 0 |
| 110 | 12 | 0 | 158 | 18 | 1 |
| 132 | 14 | 1 | 145 | 18 | 0 |
| 116 | 14 | 0 | | | |

By Tuan Anh(UEH)

Nhận xét kết quả hồi quy sau :

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|----------|-------------|------------|-------------|--------|
| C | 45.28511 | 5.538866 | 8.175881 | 0.0000 |
| X | 5.236452 | 0.379310 | 13.80520 | 0.0000 |
| Z | 16.00976 | 2.058934 | 7.775752 | 0.0000 |

| | | | |
|--------------------|-----------|-----------------------|----------|
| R-squared | 0.952576 | Mean dependent var | 126.0667 |
| Adjusted R-squared | 0.944672 | S.D. dependent var | 16.89238 |
| S.E. of regression | 3.973393 | Akaike info criterion | 5.773974 |
| Sum squared resid | 189.4542 | Schwarz criterion | 5.915584 |
| Log likelihood | -40.30481 | Hannan-Quinn criter. | 5.772466 |
| F-statistic | 120.5192 | Durbin-Watson stat | 1.919503 |
| Prob(F-statistic) | 0.000000 | | |

Và kết quả hồi quy này giúp rút ra kết luận gì ?

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|----------|-------------|------------|-------------|--------|
| C | 46.29423 | 7.521975 | 6.154532 | 0.0001 |
| X | 5.165010 | 0.522340 | 9.888206 | 0.0000 |
| Z | 13.67223 | 11.37238 | 1.202231 | 0.2545 |
| X*Z | 0.167307 | 0.799342 | 0.209306 | 0.8380 |

| | | | |
|--------------------|-----------|-----------------------|----------|
| R-squared | 0.952765 | Mean dependent var | 126.0667 |
| Adjusted R-squared | 0.939882 | S.D. dependent var | 16.89238 |
| S.E. of regression | 4.141834 | Akaike info criterion | 5.903333 |
| Sum squared resid | 188.7027 | Schwarz criterion | 6.092146 |
| Log likelihood | -40.27500 | Hannan-Quinn criter. | 5.901321 |
| F-statistic | 73.95857 | Durbin-Watson stat | 1.911440 |
| Prob(F-statistic) | 0.000000 | | |