

II. ỨNG DỤNG CỦA MÔ HÌNH HỒI QUY VỚI BIẾN GIẢ

II.1

So sánh hai hồi quy

II.2

Phân tích yếu tố mùa vụ

II.3

Hồi quy tuyến tính từng đoạn

cuu duong than cong. com

II.1 So sánh hai hồi quy

Có 2 tập số liệu khác nhau tương ứng với 2 mẫu khác nhau nhưng đều cùng liên quan đến 1 biến phụ thuộc và 1 số biến giải thích nào đó.

Câu hỏi: có thể xây dựng 1 mô hình hồi quy duy nhất dựa trên cả 2 tập số liệu đó hay không hay phải xây dựng 2 mô hình hồi quy khác nhau, mỗi hồi quy dựa trên 1 tập số liệu?

Để trả lời câu hỏi này, chúng ta gộp 2 tập số liệu làm một và ước lượng các mô hình hồi quy.

Trường hợp chỉ có 1 biến số lượng là biến giải thích

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

Trong đó: X_i : biến giải thích là biến số lượng

Z_i : biến giả, được quy ước: $Z_i = 0$: Số liệu thuộc mẫu I

$Z_i = 1$: Số liệu thuộc mẫu II

Với giả thiết $E(U_i) = 0$ được thỏa mãn thì

$$E(Y/Z_i = 0; X_i) = \beta_1 + \beta_2 X_i$$

$$E(Y/Z_i = 1; X_i) = \beta_1 + \beta_3 + (\beta_2 + \beta_4) X_i$$

- β_3 : chênh lệch (khác nhau) về hệ số chặn
- β_4 : chênh lệch (khác nhau) về hệ số góc

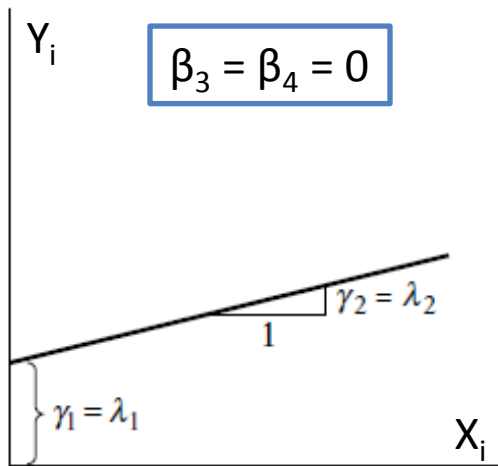
Xét MHHQ 2 biến

Giả sử mẫu I có PRF1:

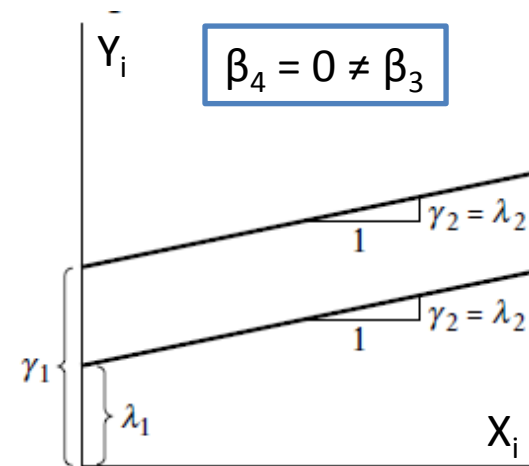
$$Y_i = \lambda_1 + \lambda_2 X_i + U_i$$

Mẫu II có PRF2:

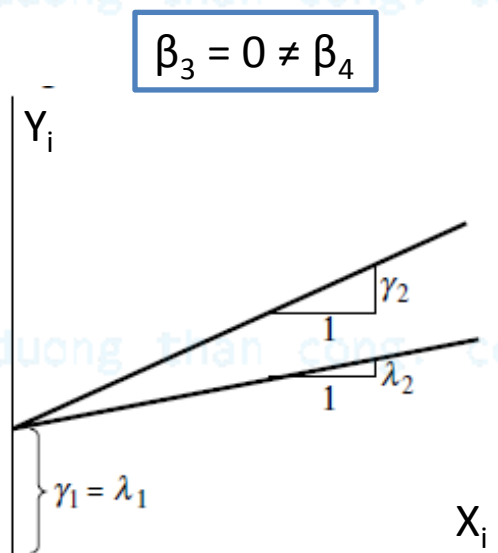
$$Y_i = \gamma_1 + \gamma_2 X_i + U_i$$



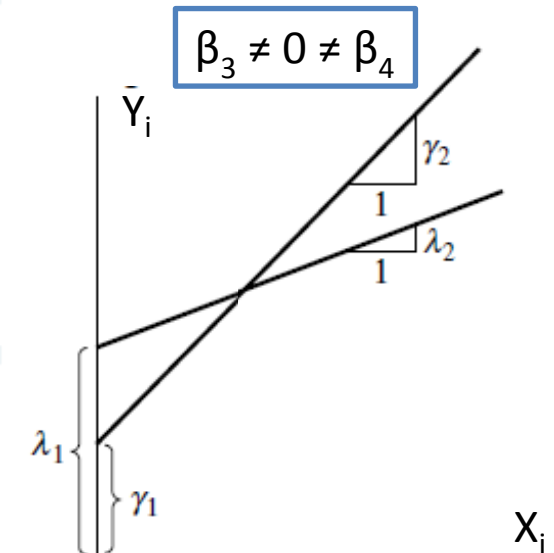
Gộp được



Không gộp được



Không gộp được



Không gộp được



Hai tập số liệu gộp được với nhau (tương đồng nhau) nếu:

$$E(Y/Z_i = 0; X_i) = E(Y/Z_i = 1; X_i) \forall X_i \Leftrightarrow \begin{cases} \beta_3 = 0 \\ \beta_4 = 0 \end{cases}$$

Như vậy so sánh tập số liệu (hay 2 hồi quy) thực chất là kiểm định giả thuyết:

$$\begin{cases} H_0: \beta_3 = \beta_4 = 0 \\ H_1: \beta_3 \neq 0 \text{ hoặc } \beta_4 \neq 0 \end{cases}$$

-> Kiểm định giả thuyết đồng thời của mô hình hồi quy bội.

Nếu ở mức ý nghĩa α nào đó ta không có cơ sở bác bỏ H_0 thì điều này có nghĩa (ở mức ý nghĩa α đó) có thể gộp 2 tập số liệu được.

Trường hợp có nhiều hơn 1 biến số lượng là biến giải thích:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \beta_{k+1} Z_i \\ \beta_{k+2} X_{2i} Z_i + \dots + \beta_{2k} X_{ki} Z_i + U_i$$

Trong đó: X_{ji} : các biến số lượng, $j=1..k$; Z_i : biến giả, quy ước:

$Z_i = 0$: Số liệu thuộc mẫu I; $Z_i = 1$: Số liệu thuộc mẫu II

Với giả thiết $E(U_i) = 0$ được thỏa mãn thì:

$$E(Y/Z_i = 0; X_{2i}; \dots; X_{ki}) = \beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki}$$

$$E(Y/Z_i = 1; X_{2i}; \dots; X_{ki}) = \beta_1 + \beta_{k+1} + (\beta_2 + \beta_{k+2})X_{2i} + \\ (\beta_3 + \beta_{k+3})X_{3i} + \dots + (\beta_k + \beta_{2k})X_{ki}$$

- β_{k+1} : chênh lệch về hệ số chặn
- β_{k+j} : chênh lệch (khác nhau) về các hệ số góc, $j=2..k$

Hai tập số liệu gộp được với nhau (tương đồng nhau) nếu:

$$\forall (X_{2i}; \dots; X_{ki}) E(Y/Z_i = 0; X_{2i}; \dots; X_{ki}) = E(Y/Z_i = 1; X_{2i}; \dots; X_{ki})$$

$$\Leftrightarrow \beta_{k+1} = \beta_{k+2} = \dots = \beta_{2k} = 0$$

Như vậy so sánh tập số liệu (hay 2 hồi quy) thực chất là kiểm định giả thuyết:

$$\begin{cases} H_0: \beta_{k+1} = \beta_{k+2} = \dots = \beta_{2k} = 0 \\ H_1: \exists j \in \{1, \dots, k\} | \beta_{k+j} \neq 0 \end{cases}$$

-> Kiểm định giả thuyết đồng thời của mô hình hồi quy bội.

Nếu ở mức ý nghĩa α nào đó ta không có cơ sở bác bỏ H_0 thì điều này có nghĩa (ở mức ý nghĩa α đó) có thể gộp 2 tập số liệu được.



II.2

Phân tích yếu tố mùa vụ

Yếu tố thời vụ (mùa) là yếu tố có nhiều ảnh hưởng đến một số các đại lượng kinh tế như nhu cầu về một số mặt hàng, doanh số bán ra, sản lượng một số loại nông sản ...

Thí dụ, nhu cầu tiêu dùng hàng may mặc chịu ảnh hưởng của yếu tố thu nhập và yếu tố mùa mà ta có thể coi là thay đổi theo các quý trong năm. Mô hình hồi quy có dạng:

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 Z_{1i} + \beta_4 Z_{2i} + \beta_5 Z_{3i} + U_i$$

Y_i : tiêu dùng hàng may mặc

X_i : thu nhập

Z_{1i} , Z_{2i} , Z_{3i} : các biến giả

Quý	I	II	III	IV
Z_{1i}	0	1	0	0
Z_{2i}	0	0	1	0
Z_{3i}	0	0	0	1

Với giả thuyết $E(U_i)$ được thỏa mãn thì:

$$E(Y|Z_{1i} = Z_{2i} = Z_{3i} = 0; X_i) = \beta_1 + \beta_2 X_i$$

→ chỉ tiêu trung bình cho may mặc quý I với mức thu nhập X_i

$$E(Y|Z_{1i} = 1; Z_{2i} = Z_{3i} = 0; X_i) = (\beta_1 + \beta_3) + \beta_2 X_i$$

→ chỉ tiêu trung bình cho may mặc quý II với mức thu nhập X_i

$$E(Y|Z_{2i} = 1; Z_{1i} = Z_{3i} = 0; X_i) = (\beta_1 + \beta_4) + \beta_2 X_i$$

→ chỉ tiêu trung bình cho may mặc quý III với mức thu nhập X_i

$$E(Y|Z_{1i} = Z_{2i} = 0; Z_{3i} = 1; X_i) = (\beta_1 + \beta_5) + \beta_2 X_i$$

→ chỉ tiêu trung bình cho may mặc quý IV với mức thu nhập X_i

Như vậy, $\beta_3, \beta_4, \beta_5$ là chênh lệch về nhu cầu may mặc khi chuyển từ quý I sang quý II, quý III và quý IV ở cùng mức thu nhập X_i

Trường hợp có ảnh hưởng tương tác giữa biến giải thích X_i và yếu tố mùa vụ (giả sử có 4 phạm trù) lên biến phụ thuộc Y_i , khi đó ta sử dụng mô hình:

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 Z_{1i} + \beta_4 Z_{2i} + \beta_5 Z_{3i} \\ + \beta_6 X_i Z_{1i} + \beta_7 X_i Z_{2i} + \beta_8 X_i Z_{3i} + U_i$$

→ mô hình thường hay sử dụng.

Với giả thuyết $E(U_i) = 0$ được thỏa mãn thì:

$$E(Y|Z_{1i} = Z_{2i} = Z_{3i} = 0; X_i) = \beta_1 + \beta_2 X_i$$

$$E(Y|Z_{1i} = 1; Z_{2i} = Z_{3i} = 0; X_i) = (\beta_1 + \beta_3) + (\beta_2 + \beta_6)X_i$$

$$E(Y|Z_{2i} = 1; Z_{1i} = Z_{3i} = 0; X_i) = (\beta_1 + \beta_4) + (\beta_2 + \beta_7)X_i$$

$$E(Y|Z_{1i} = Z_{2i} = 0; Z_{3i} = 1; X_i) = (\beta_1 + \beta_5) + (\beta_2 + \beta_8)X_i$$

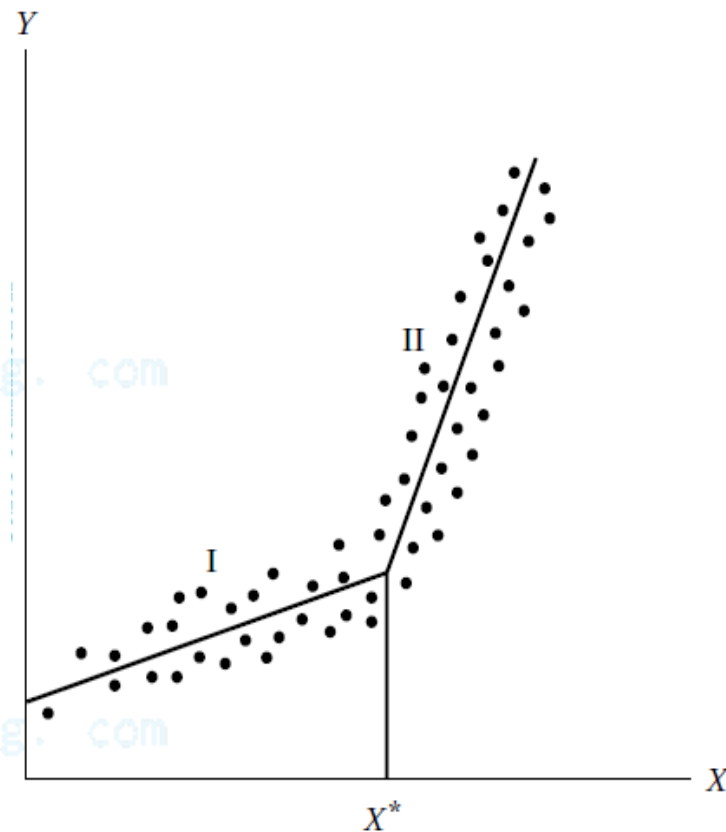
Một số lưu ý khi đưa biến vào mô hình:

- Yếu tố mùa vụ có $(m+1)$ phạm trù, đưa vào mô hình m biến giả $Z_{1i}; \dots; Z_{mi}$
 - Ứng với 1 biến giải thích X_i là biến số lượng, đưa vào mô hình m biến $X_i Z_{1i}; \dots; X_i Z_{mi}$. Các biến giải thích khác là biến số lượng làm tương tự.
- Nếu mô hình có k biến giải thích là biến số lượng, yếu tố mùa vụ có $(m+1)$ phạm trù, không kể m biến giả $Z_{1i}; \dots; Z_{mi}$; phải đưa thêm vào mô hình $k*m$ biến mới $X_{ji} Z_{ti}$ ($j=1..k; t=1..m$)



II.3 Hồi quy tuyến tính từng đoạn

Bài toán 1: Một công ty trả tiền hoa hồng cho các đại lý theo cách như sau: tiền hoa hồng là một đại lượng ngẫu nhiên phụ thuộc vào doanh số bán hàng. Tiền hoa hồng tăng tuyến tính với doanh số bán hàng cho đến một ngưỡng X^* nào đó, sau đó nó cũng tăng tuyến tính với doanh số bán hàng nhưng ở một tỷ lệ cao hơn nhiều.



Trong trường hợp này, chúng ta có một mô hình hồi quy tuyến tính từng đoạn:

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 (X_i - X^*) D_{1i} + U_i$$

Trong đó: Y_i là tiền hoa hồng; X_i là doanh số của đại lý
 D_{1i} là biến giả, quy ước: $D_{1i} = 1$ nếu $X_i \geq X^*$

$D_{1i} = 0$ nếu ngược lại.

$$\rightarrow E(Y|X_i) = \begin{cases} \beta_1 + \beta_2 X_i & \text{nếu } X_i \leq X^* \\ (\beta_1 - \beta_3 X^*) + (\beta_2 + \beta_3) X_i & \text{nếu } X_i > X^* \end{cases}$$

→ TB tiền hoa hồng trước và sau ngưỡng X^*

Kiểm định:

$\begin{cases} H_0: \beta_3 = 0 \\ H_1: \beta_3 \neq 0 \end{cases}$ Tại X^* không hay có sự thay đổi tốc độ biến thiên của Y (phụ thuộc vào X)

Giả sử biết thêm rằng, doanh thu vượt qua ngưỡng X^* , đến ngưỡng X^{**} nào đó ($X^{**} > X^*$), tiền hoa hồng vẫn tăng tuyến tính với doanh số bán hàng nhưng ở một tỷ lệ cao hơn nữa. Khi đó, mô hình của chúng ta gồm 3 phân đoạn và mô hình được đề xuất:

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 (X_i - X^*) D_{1i} + \beta_4 (X_i - X^{**}) D_{2i} + U_i$$

Trong đó: D_{2i} là biến giả, quy ước: $D_{2i} = 1$ nếu $X_i \geq X^{**}$

$D_{2i} = 0$ nếu ngược lại.

$$\rightarrow E(Y|X_i) = \begin{cases} \beta_1 + \beta_2 X_i & \text{nếu } X_i \leq X^* \\ (\beta_1 - \beta_3 X^*) + (\beta_2 + \beta_3) X_i & \text{nếu } X^{**} \geq X_i > X^* \\ (\beta_1 - \beta_3 X^* - \beta_4 X^{**}) + (\beta_2 + \beta_3 + \beta_4) X_i & \text{nếu } X^{**} < X_i \end{cases}$$

Bài toán 2: Coi chi tiêu là một đại lượng ngẫu nhiên phụ thuộc vào thu nhập. Nghiên cứu về ảnh hưởng của chuyển đổi kinh tế xảy ra ở năm t_0 đến chi tiêu, ta xét mô hình:

$$Y_t = \beta_1 + \beta_2 X_t + \beta_3 (X_t - X_{t_0}) D_{1t} + U_t$$

Trong đó: Y_t : chi tiêu ở năm t ; X_t : thu nhập ở năm t ; X_{t_0} : thu nhập ở năm bắt đầu chuyển đổi; D_{1t} là biến giả, quy ước:

$D_{1t} = 1$ nếu $t \geq t_0$; $D_{1t} = 0$ nếu ngược lại.

$$E(Y|X_t) = \begin{cases} \beta_1 + \beta_2 X_t & \text{nếu } t \leq t_0 \\ (\beta_1 - \beta_3 X_{t_0}) + (\beta_2 + \beta_3) X_t & \text{nếu } t > t_0 \end{cases}$$

TB chi tiêu trước và sau chuyển đổi

Kiểm định:

$$\begin{cases} H_0: \beta_3 = 0 \\ H_1: \beta_3 \neq 0 \end{cases} \quad \begin{array}{l} \text{Không hay có ảnh hưởng của} \\ \text{chuyển đổi kinh tế tới chi tiêu.} \end{array}$$

Trong trường hợp có nhiều hơn một thời điểm chuyển đổi, chẳng hạn ngoài thời điểm chuyển đổi t_0 còn có thời điểm chuyển đổi thứ hai $t_1 > t_0$ thì có thể đề nghị mô hình hồi quy tuyến tính từng đoạn như sau:

$$Y_t = \beta_1 + \beta_2 X_t + \beta_3 (X_t - X_{t_0}) D_{1t} + \beta_4 (X_t - X_{t_1}) D_{2t} + U_t$$

Trong đó: D_{2i} là biến giả, quy ước: $D_{2i} = 1$ nếu $t \geq t_1$;

$D_{2i} = 0$ nếu ngược lại.

$$E(Y|X_t) = \begin{cases} \beta_1 + \beta_2 X_t & \text{nếu } t \leq t_0 \\ (\beta_1 - \beta_3 X_{t_0}) + (\beta_2 + \beta_3) X_t & \text{nếu } t_0 < t \leq t_1 \\ (\beta_1 - \beta_3 X_{t_0} - \beta_4 X_{t_1}) + (\beta_2 + \beta_3 + \beta_4) X_t & \text{nếu } t_1 < t \end{cases}$$

Làm tương tự với các mốc thời gian tiếp theo