

Chương 1. MỞ ĐẦU

Chương 1

MỞ ĐẦU

1.1 Kinh tế lượng là gì

1.2 Các khái niệm cơ bản của kinh tế lượng

Chương 1

§1.1 Kinh tế lượng là gì

1.1.1 Khái niệm

Tiếng anh: econometrics – đo lường kinh tế

Là môn học được hình thành và phát triển trên cơ sở 3 ngành khoa học khác: kinh tế học, thống kê học và toán học

Chương 1

§1.1 Kinh tế lượng là gì

1.1.2 Nội dung nghiên cứu của Kinh tế lượng

Thiết lập các mô hình toán học mô tả mối quan hệ giữa các đại lượng kinh tế (biến kinh tế)

Đo lường mức độ ảnh hưởng của các biến kinh tế này đến các biến kinh tế khác

Dựa vào các mô hình toán học để dự báo các hiện tượng kinh tế

Chương 1

§1.1 Kinh tế lượng là gì

1.1.3 Phương pháp luận của Kinh tế lượng

1. Dựa vào lý thuyết kinh tế để đưa ra giả thiết về mối quan hệ giữa các biến kinh tế quan tâm
2. Thiết lập các mô hình toán học để mô tả mối quan hệ giữa các biến kinh tế
3. Ước lượng các tham số của mô hình đã đưa ra

Chương 1

§1.1 Kinh tế lượng là gì

4. Phân tích kết quả: đánh giá độ tin cậy và kiểm định tính đúng đắn, chính xác của các ước lượng đã nhận được

Chương 1

§1.1 Kinh tế lượng là gì

5. Dự báo: sử dụng các mô hình đã xây dựng được để dự báo các hiện tượng kinh tế hoặc giá trị của các biến kinh tế mà ta quan tâm dưới ảnh hưởng của các biến kinh tế khác

Chương 1

§1.1 Kinh tế lượng là gì

6. Đề ra các chính sách mới phù hợp nhằm đạt được mục tiêu đã định

Chương 1

§1.2 Các KN cơ bản của kinh tế lượng

1.2.1 Phân tích hồi quy

Nghiên cứu mối liên hệ phụ thuộc giữa giá trị của một biến Y - gọi là biến phụ thuộc hay biến được giải thích với giá trị của một hoặc nhiều biến khác X_j ($j=1,...,m$) – các biến này gọi là các biến độc lập hay biến giải thích

Chương 1

§1.2 Các KN cơ bản của kinh tế lượng

Ta thường giả thiết

Biến phụ thuộc Y là biến ngẫu nhiên, có quy luật phân phối xác suất xác định

Các biến độc lập X_j không phải là biến ngẫu nhiên, giá trị của chúng là xác định

Chương 1

§1.2 Các KN cơ bản của kinh tế lượng

Phân tích hồi quy giúp ta:

- Ước lượng giá trị của biến phụ thuộc Y khi đã biết giá trị của (các) biến độc lập X_j
- Kiểm định giả thiết về sự phụ thuộc
- Dự báo giá trị trung bình hoặc cá biệt của biến phụ thuộc khi đã biết giá trị của (các) biến độc lập

Chương 1

§1.2 Các KN cơ bản của kinh tế lượng

1.2.2 Mô hình hồi quy tổng thể và mô hình hồi quy mẫu

Mô hình hồi quy tổng thể (hàm tổng thể - PRF) là hàm có dạng tổng quát

$$E(Y / X_{ji}) = f(X_{ji}) \quad (1.1)$$

Chương 1

§1.2 Các KN cơ bản của kinh tế lượng

Nếu (1.1) biểu diễn mối quan hệ giữa biến phụ thuộc Y và một biến giải thích X thì (1.1) được gọi là mô hình hồi quy đơn hay mô hình hồi quy 2 biến

Nếu số biến giải thích nhiều hơn 1 thì (1.1) được gọi là mô hình hồi quy bội (hồi quy nhiều biến)

Chương 1

§1.2 Các KN cơ bản của kinh tế lượng

Mô hình hồi quy mẫu (hàm hồi quy mẫu - SRF) có thể được biểu diễn như sau

$$\hat{Y}_i = \hat{F}(X_{ji}) \quad (1.2)$$

$\hat{Y}_i$ là ước lượng của $E(Y / X_{ji})$
 $\hat{F}$ là ước lượng của f

Chương 1

§1.2 Các khái niệm cơ bản của kinh tế lượng

1.2.3 Sai số ngẫu nhiên

$$U_i = Y_i - E(Y / X_{ji}), \quad j=1,\dots,m; \quad i=1,\dots,n$$

U_i được gọi là sai số ngẫu nhiên (nhiều ngẫu nhiên), biểu thị ảnh hưởng của các yếu tố khác ngoài các biến giải thích X_j tới giá trị của biến Y

Chương 1

§1.2 Các khái niệm cơ bản của kinh tế lượng

1.2.3 Sai số ngẫu nhiên

Khi đó hàm hồi quy tổng thể (1.1) có thể biểu diễn dưới dạng

$$Y_i = f(X_{ji}) + U_i$$

Chương 2

MÔ HÌNH HỒI QUY HAI BIẾN

Chương 2

MÔ HÌNH HỒI QUY HAI BIẾN

2.1 Phương pháp bình phương nhỏ nhất

2.2 Các giả thiết cơ bản của MHHQ hai biến

2.3 Ước lượng và kiểm định GT về hệ số HQ

2.4 Phân tích phương sai và sự phù hợp của MH

2.5 Phân tích hồi quy và dự báo

Chương 2

§2.1 Mô hình hồi quy hai biến và phương pháp bình phương nhỏ nhất

2.1.1 Mô hình hồi quy hai biến

$$Y_i = \beta_1 + \beta_2 X_i + U_i \quad (2.1)$$

Trong đó:

Y_i : giá trị của biến phụ thuộc Y ($i = \overline{1, n}$)

β_1 hệ số chặn

β_2 hệ số góc của biến giải thích

U_i : sai số ngẫu nhiên

Chương 2

§2.1 Mô hình hồi quy hai biến và phương pháp bình phương nhỏ nhất

Mô hình hồi quy mẫu xây dựng dựa trên mẫu ngẫu nhiên kích thước n : $\{(Y_i, X_i), i = \overline{1, n}\}$

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i \quad (2.2)$$

Trọng đó:

$\hat{Y}_i$ ước lượng của Y_i hoặc $E(Y/X_i)$ ($i = \overline{1, n}$)

$\hat{\beta}_j$ ước lượng của hệ số hồi quy tổng thể ($j = 1, 2$)

Chương 2

§2.1 Mô hình hồi quy hai biến và phương pháp bình phương nhỏ nhất

2.1.2 Phương pháp bình phương nhỏ nhất (OLS)

Từ hàm hồi qui mẫu và hàm hồi qui tổng thể

Đặt:
$$e_i = Y_i - \hat{Y}_i$$

e_i : phần dư của hàm hồi qui mẫu

Chương 2

§2.1 Mô hình hồi quy hai biến và phương pháp bình phương nhỏ nhất

Phương pháp OLS đòi hỏi các hệ số hồi qui được xác định sao cho:

$$\sum e_i^2 \rightarrow \min \quad (2.3)$$

Các hệ số $\hat{\beta}_1, \hat{\beta}_2$ nhận được từ (2.3) gọi là các ước lượng bình phương nhỏ nhất của β_1, β_2

Chương 2

§2.1 Mô hình hồi quy hai biến và phương pháp bình phương nhỏ nhất

Khai triển tổng bình phương các phần dư ta có:

$$\sum e_i^2 = \sum (y_i - \hat{y}_i)^2 = \sum (y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2$$

$$\text{Đặt : } f = f(\hat{\beta}_1, \hat{\beta}_2) = \sum e_i^2 = \sum (y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2$$

Chương 2

§2.1 Mô hình hồi quy hai biến và phương pháp bình phương nhỏ nhất

Khi đó $f(\hat{\beta}_1, \hat{\beta}_2)$ nhỏ nhất khi $\hat{\beta}_1, \hat{\beta}_2$ là nghiệm của hệ phương trình sau:

$$\begin{cases} \frac{\partial f}{\partial \hat{\beta}_1} = 0 \\ \frac{\partial f}{\partial \hat{\beta}_2} = 0 \end{cases} \quad (2.4)$$

Chương 2

§2.1 Mô hình hồi quy hai biến và phương pháp bình phương nhỏ nhất

Đạo hàm và khai triển ta được:

$$\begin{cases} \sum 2(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)(-1) = 0 \\ \sum 2(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)(-X_i) = 0 \end{cases}$$

Hay:

$$\begin{cases} n\hat{\beta}_1 + \hat{\beta}_2 \sum X_i = \sum Y_i \\ \hat{\beta}_1 \sum X_i + \hat{\beta}_2 \sum X_i^2 = \sum X_i Y_i \end{cases} \quad (2.5)$$

Chương 2

§2.1 Mô hình hồi quy hai biến và phương pháp bình phương nhỏ nhất

Khi đó nếu: $\Delta = \begin{vmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{vmatrix} = n \sum X_i^2 - (\sum X_i)^2 \neq 0$

Đặt $\bar{Y} = \frac{1}{n} \sum Y_i$; $\bar{X} = \frac{1}{n} \sum X_i$

Hệ (2.5) có nghiệm:

$$\hat{\beta}_2 = \frac{n \sum Y_i X_i - \sum Y_i \sum X_i}{n \sum X_i^2 - (\sum X_i)^2}$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}$$

Chương 2

§2.1 Mô hình hồi quy hai biến và phương pháp bình phương nhỏ nhất

Đặt
$$\begin{cases} x_i = X_i - \bar{X} \\ y_i = Y_i - \bar{Y} \end{cases}$$

Ta được:

$$\hat{\beta}_2 = \frac{\sum y_i x_i}{\sum x_i^2} \quad (2.6)$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} \quad (2.7)$$

VÍ DỤ 2.1

Theo dõi thu nhập hàng tháng và mức chi về hàng thực phẩm của 8 gia đình có số thành viên như nhau, ta có số liệu sau (đơn vị: triệu đồng)

X_i	1,2	3,1	5,3	7,4	9,6	11,8	14,5	18,7
Y_i	0,9	1,2	1,8	2,2	2,6	2,9	3,3	3,8

VÍ DỤ 2.1

Trong đó:

X_i : thu nhập hàng tháng của gia đình thứ i

Y_i : mức chi cho hàng thực phẩm của gia đình thứ i .

Dựa vào bảng số liệu trên và phương pháp OLS hãy xây dựng hàm hồi qui mẫu:

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

Bảng kết quả:

i	X_i	Y_i	x_i	x_i^2	y_i	y_i^2	$x_i y_i$
1	1,20	0,90	-7,75	60,06	-1,44	2,07	11,16
2	3,10	1,20	-5,85	34,22	-1,14	1,30	6,67
3	5,30	1,80	-3,65	13,32	-0,54	0,29	1,97
4	7,40	2,20	-1,55	2,40	-0,14	0,02	0,22
5	9,60	2,60	0,65	0,42	0,26	0,07	0,17
6	11,80	2,90	2,85	8,12	0,56	0,31	1,60
7	14,50	3,30	5,55	30,80	0,96	0,92	5,33
8	18,70	3,80	9,75	95,06	1,46	2,13	14,24
Σ	71,60	18,70	0,00	244,42	-0,02	7,12	41,35

$$\bar{X} = \frac{1}{n} \sum X_i = \frac{71,6}{8} = 8,95$$

$$\bar{Y} = \frac{1}{n} \sum Y_i = \frac{18,7}{8} \approx 2,34$$

Theo công thức (2.6), (2.7) ta thu được:

$$\hat{\beta}_2 = \frac{\sum y_i x_i}{\sum x_i^2} = \frac{41,35}{244,42} = 0,169$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} = 2,34 - 0,169 * 8,95 = 0,827$$

Ta có hàm hồi qui mẫu:

$$\hat{Y}_i = 0,827 + 0,169X_i$$

Ý nghĩa của hệ số hồi qui:

$\hat{\beta}_2 = 0,169$: Khi thu nhập của gia đình tăng lên 1 triệu đồng thì mức chi trung bình hàng tháng cho hàng thực phẩm của gia đình tăng lên khoảng 169 ngàn đồng.

Chương 2

§2.1 Mô hình hồi quy hai biến và phương pháp bình phương nhỏ nhất

2.1.3 Các tính chất của ước lượng BPNN

1. Đường hồi quy mẫu đi qua điểm trung bình mẫu $(\bar{Y}, \bar{X})$, tức là:

$$\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}$$

2. Giá trị trung bình của các giá trị $\hat{Y}_i$ được xác định theo hàm hồi quy mẫu bằng giá trị trung bình của biến phụ thuộc, tức là:

$$\bar{\hat{Y}} = \frac{1}{n} \sum \hat{Y}_i = \bar{Y}$$

Chương 2

§2.1 Mô hình hồi quy hai biến và phương pháp bình phương nhỏ nhất

3. Tổng các phần dư của hàm hồi quy mẫu bằng 0

$$\sum e_i = 0$$

4. Các phần dư e_i không tương quan với $\hat{Y}_i$

$$\sum e_i \hat{Y}_i = 0$$

5. Các phần dư e_i không tương quan với X_i

$$\sum e_i X_i = 0$$

Chương 2

§2.2 Các giả thiết cơ bản của mô hình hồi quy hai biến

2.2.1 Các giả thiết cơ bản của phương pháp BPNN

Giả thiết 1. Biến giải thích X là phi ngẫu nhiên, giá trị của nó là xác định.

Giả thiết 2. Kỳ vọng toán của các sai số ngẫu nhiên U_i bằng không

$$E(U_i) = E(U / X_i) = 0 \quad (\forall i)$$

Chương 2

§2.2 Các giả thiết cơ bản của mô hình hồi quy hai biến

Giả thiết 3. Phương sai của các sai số U_i là không đổi (phương sai thuần nhất)

$$Var(U_i) = Var(U / X_i) = \sigma^2 = const \quad (\forall i)$$

Từ GT3 ta thấy :

$$Var(Y_i / X_i) = \sigma^2 \quad (\forall i)$$

Chương 2

§2.2 Các giả thiết cơ bản của mô hình hồi quy hai biến

Giả thiết 4. Các sai số U_i không tương quan với nhau

$$Cov(U_i, U_j) = 0 \quad (\forall i \neq j)$$

Giả thiết 5. Các sai số U_i và X_i không tương quan với nhau

$$Cov(U_i, X_i) = 0 \quad (\forall i)$$

Chương 2

§2.2 Các giả thiết cơ bản của mô hình hồi quy hai biến

Xét hàm HQ 2 biến và hàm HQ mẫu :

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

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

Hệ số hồi qui mẫu được xác định bằng OLS :

$$\hat{\beta}_2 = \frac{\sum y_i x_i}{\sum x_i^2}$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}$$

Chương 2

§2.2 Các giả thiết cơ bản của mô hình hồi quy hai biến

Với các giả thiết cơ bản 1-5 của OLS được thỏa mãn, ta có:

$$Var(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} \quad (2.8)$$

$$Var(\hat{\beta}_1) = \frac{\sigma^2 \sum X_i^2}{n \sum x_i^2} \quad (2.9)$$

Với : $Var(U_i) = \sigma^2 \quad (\forall i)$

Chương 2

§2.2 Các giả thiết cơ bản của mô hình hồi quy hai biến

Độ lệch chuẩn của các hệ số hồi quy mẫu :

$$se(\hat{\beta}_2) = \sqrt{\frac{\sigma^2}{\sum x_i^2}} = \frac{\sigma}{\sqrt{\sum x_i^2}} \quad (2.10)$$

$$se(\hat{\beta}_1) = \sqrt{\frac{\sigma^2 \sum X_i^2}{n \sum x_i^2}} = \sigma \sqrt{\frac{\sum X_i^2}{n \sum x_i^2}} \quad (2.11)$$

Chương 2

§2.2 Các giả thiết cơ bản của mô hình hồi quy hai biến

Do $Var(U_i) = \sigma^2$ chưa biết nên khi áp dụng công thức (2.8) – (2.11) ta thường lấy bởi ước lượng không chệch của nó:

$$\sigma^2 \approx \hat{\sigma}^2 = \frac{\sum e_i^2}{n-2} \quad (2.12)$$

Chương 2

§2.2 Các giả thiết cơ bản của mô hình hồi quy hai biến

Định lý Gauss – Markov: Với các giả thiết của phương pháp bình phương nhỏ nhất thì các ước lượng bình phương nhỏ nhất $\hat{\beta}_j$ là các ước lượng tuyến tính, không chệch và có phương sai nhỏ nhất trong lớp các ước lượng tuyến tính, không chệch của β_j ($j = \overline{1,2}$) .

Chương 2

§2.2 Các giả thuyết cơ bản của mô hình hồi quy hai biến

Cụ thể ta có: $\ddot{\beta}_2 = \sum_{i=1}^n k_i Y_i \quad (k_i = \text{const})$

$$E(\ddot{\beta}_2) = \beta_2$$

Với β_2^* là ước lượng tuyến tính, không chệch bất kì của β_2 ta có:

$$\text{Var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum_{i=1}^n x_i^2} \leq \text{Var}(\beta_2^*)$$

Chương 2

§2.2 Các giả thiết cơ bản của mô hình hồi quy hai biến

2.2.3 Giả thiết về phân phối xác suất của U_i

Giả thiết 6. Sai số ngẫu nhiên U_i có phân phối chuẩn, tức là:

$$U_i \sim N(0, \sigma^2)$$

Với các giả thiết 1-6 mô hình hồi quy 2 biến (2.1) được gọi là MHHQTT cổ điển

Chương 2

§2.2 Các giả thiết cơ bản của mô hình hồi quy hai biến

Với giả thiết 1-6 của MHHQTT cổ điển, ta có:

$$1. \hat{\beta}_1 \sim N(\beta_1, \text{Var}(\hat{\beta}_1)) \quad 3. \chi^2 = \frac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-2)$$

$$2. \hat{\beta}_2 \sim N(\beta_2, \text{Var}(\hat{\beta}_2)) \quad 4. Y_i \sim N(\beta_1 + \beta_2 X_i, \sigma^2)$$

5. Các ước lượng OLS $\hat{\beta}_1, \hat{\beta}_2$ của β_1, β_2 là các ước lượng hiệu quả

Chương 2

§2.3 Khoảng tin cậy và kiểm định giả thuyết về các hệ số hồi quy

Xét MHHQTT cổ điển và hàm hồi quy mẫu :

$$Y_i = \beta_1 + \beta_2 X_i + U_i \quad (2.1)$$

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i \quad (2.2)$$

Từ giả thiết 6 về phân phối chuẩn của sai số ngẫu nhiên, có thể suy ra:

$$\hat{\beta}_j \sim N(\beta_j, \text{Var}(\hat{\beta}_j)) \quad (j = \overline{1,2})$$

Chương 2

§2.3 Khoảng tin cậy và kiểm định giả thuyết về các hệ số hồi quy

2.3.1 Khoảng tin cậy của các hệ số hồi quy

Do σ^2 ta chưa biết mà phải thay bằng ước lượng không chệch của nó là $\hat{\sigma}^2$, nên :

$$T = \frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \sim T(n-2) \quad (j = \overline{1,2})$$

Chương 2

§2.3 Khoảng tin cậy và kiểm định giả thuyết về các hệ số hồi quy

$$P\left(\left|\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)}\right| < t_{\frac{\alpha}{2}}(n-2)\right) = 1 - \alpha$$

$$P\left(\hat{\beta}_j - t_{\frac{\alpha}{2}}(n-2).se(\hat{\beta}_j) < \beta_j < \hat{\beta}_j + t_{\frac{\alpha}{2}}(n-2).se(\hat{\beta}_j)\right) = 1 - \alpha = \gamma$$

Khoảng tin cậy của hệ số hồi quy :

$$\left(\hat{\beta}_j - t_{\frac{\alpha}{2}}(n-2).se(\hat{\beta}_j) ; \hat{\beta}_j + t_{\frac{\alpha}{2}}(n-2).se(\hat{\beta}_j)\right)$$

Chương 2

§2.3 Khoảng tin cậy và kiểm định giả thuyết về các hệ số hồi quy

2.3.2 Kiểm định giả thuyết về các hệ số HQ

Giả sử với mức ý nghĩa α cho trước ta cần kiểm định giả thuyết:

$$\begin{cases} H_0 : \beta_j = \beta_j^* \\ H_1 : \beta_j \neq \beta_j^* \end{cases} \quad (\beta_j < \beta_j^*, \quad \beta_j > \beta_j^*)$$

Chương 2

§2.3 Khoảng tin cậy và kiểm định giả thuyết về các hệ số hồi quy

Ta xây dựng tiêu chuẩn kiểm định:

$$T = \frac{\hat{\beta}_j - \beta_j^*}{se(\hat{\beta}_j)}$$

Nếu H_0 đúng thì $T \sim T(n-2)$

Chương 2

§2.3 Khoảng tin cậy và kiểm định giả thuyết về các hệ số hồi quy

Loại gt	H_0	H_1	W_α
Hai phía	$\beta_j = \beta_j^*$	$\beta_j \neq \beta_j^*$	$W_\alpha = \{ t : t > t_{\alpha/2}(n-2) \}$
Trái	$\beta_j = \beta_j^*$	$\beta_j < \beta_j^*$	$W_\alpha = \{ t : t < -t_\alpha(n-2) \}$
Phải	$\beta_j = \beta_j^*$	$\beta_j > \beta_j^*$	$W_\alpha = \{ t : t > t_\alpha(n-2) \}$

Chương 2

§2.3 Khoảng tin cậy và kiểm định giả thuyết về các hệ số hồi quy

2.3.3 Khoảng tin cậy và kiểm định giả thuyết về phương sai của sai số ngẫu nhiên

Ta có: $\chi^2 = \frac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-2)$

$$P\left(\chi^2_{1-\frac{\alpha}{2}}(n-2) < \chi^2 < \chi^2_{\frac{\alpha}{2}}(n-2)\right) = 1 - \alpha = \gamma$$

$$P\left(\frac{(n-2)\hat{\sigma}^2}{\chi^2_{\frac{\alpha}{2}}} < \sigma^2 < \frac{(n-2)\hat{\sigma}^2}{\chi^2_{1-\frac{\alpha}{2}}}\right) = \gamma$$

Chương 2

§2.3 Khoảng tin cậy và kiểm định giả thuyết về các hệ số hồi quy

Khoảng tin cậy của σ^2 :

$$\left(\frac{(n-2)\hat{\sigma}^2}{\chi^2_{\frac{\alpha}{2}}}; \frac{(n-2)\hat{\sigma}^2}{\chi^2_{1-\frac{\alpha}{2}}} \right)$$
$$\Leftrightarrow \left(\frac{\sum e_i^2}{\chi^2_{\frac{\alpha}{2}}}; \frac{\sum e_i^2}{\chi^2_{1-\frac{\alpha}{2}}} \right)$$

Chương 2

§2.3 Khoảng tin cậy và kiểm định giả thuyết về các hệ số hồi quy

Với mức ý nghĩa α ta cần kiểm định bài toán:

$$\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 \neq \sigma_0^2 \quad (\sigma^2 < \sigma_0^2, \quad \sigma^2 > \sigma_0^2) \end{cases}$$

XDTCKĐ :

$$\chi^2 = \frac{(n-2)\sigma^2}{\sigma_0^2} = \frac{\sum e_i^2}{\sigma_0^2}$$

Nếu giả thuyết H_0 đúng thì : $\chi^2 \sim \chi^2(n-2)$

Chương 2

§2.3 Khoảng tin cậy và kiểm định giả thuyết về các hệ số hồi quy

Loại gt	H_0	H_1	W_α
Hai phía	$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$\left\{ \chi_{tn}^2 : \chi_{tn}^2 < \chi_{1-\frac{\alpha}{2}}^2(n-2) \right\} \cup \left\{ \chi_{tn}^2 : \chi_{tn}^2 > \chi_{\frac{\alpha}{2}}^2(n-2) \right\}$
Phía trái	$\sigma^2 = \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$\left\{ \chi_{tn}^2 : \chi_{tn}^2 > \chi_{\alpha}^2(n-2) \right\}$
Phía phải	$\sigma^2 = \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$\left\{ \chi_{tn}^2 : \chi_{tn}^2 < \chi_{1-\alpha}^2(n-2) \right\}$

Chương 2

§2.4 Phân tích phương sai và sự phù hợp của mô hình

2.4.1 Hệ số xác định của hàm hồi qui:

Xét hàm hồi quy tổng thể và hàm hồi quy mẫu :

$$Y_i = \beta_1 + \beta_2 X_i + U_i \quad (2.1)$$

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i \quad (2.2)$$

Chương 2

§2.4 Phân tích phương sai và sự phù hợp của mô hình

$$TSS = \sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n y_i^2$$

$$ESS = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 = \hat{\beta}_2^2 \sum_{i=1}^n x_i^2$$

$$RSS = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n e_i^2$$

Ta có : $\boxed{TSS = ESS + RSS} \quad (2.13)$

Chương 2

§2.4 Phân tích phương sai và sự phù hợp của mô hình

Định nghĩa: Hệ số xác định r^2 được định nghĩa như sau:

$$r^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

Chương 2

§2.4 Phân tích phương sai và sự phù hợp của mô hình

Tính chất: $0 \leq r^2 \leq 1$

- Nếu $r^2 = 1$, hàm HQ có thể coi là hoàn hảo
- Nếu $r^2 = 0$, hàm HQ đưa ra là không phù hợp

Vì thế r^2 được dùng làm thước đo mức độ phù hợp của hàm hồi quy

Chương 2

§2.4 Phân tích phương sai và sự phù hợp của mô hình

Hệ số xác định có thể biến đổi :

$$r^2 = \frac{ESS}{TSS} = \frac{\hat{\beta}_2^2 \sum x_i^2}{\sum y_i^2} = \left(\frac{\sum x_i y_i}{\sum x_i^2} \right)^2 \frac{\sum x_i^2}{\sum y_i^2}$$

$$\Leftrightarrow r^2 = \frac{(\sum x_i y_i)^2}{\sum x_i^2 \sum y_i^2} \quad (2.14)$$

Chương 2

§2.4 Phân tích phương sai và sự phù hợp của mô hình

Định nghĩa 2: Hệ số tương quan r được xác định:

$$r = \sqrt{r^2} = \sqrt{\frac{ESS}{TSS}} = \sqrt{1 - \frac{RSS}{TSS}}$$

Chương 2

§2.4 Phân tích phương sai và sự phù hợp của mô hình

Tính chất của hệ số tương quan :

1. $-1 \leq r \leq 1$ (dấu của r chính là dấu của β_2)
2. $r(X,Y) = r(Y,X)$ (tính đối xứng)
3. Nếu $ac > 0$ và $\begin{cases} X^* = aX + b \\ Y^* = cY + d \end{cases}$ thì $r(X^*, Y^*) = r(X, Y)$
4. Nếu X, Y độc lập thì $r(X, Y) = 0$
5. Hệ số tương quan chỉ mức độ phụ thuộc tuyến tính giữa X và Y

Chương 2

§2.4 Phân tích phương sai và sự phù hợp của mô hình

2.4.2 Kiểm định sự phù hợp của mô hình

Xét giả thuyết
$$\begin{cases} H_0 : \beta_2 = 0 \\ H_1 : \beta_2 \neq 0 \end{cases}$$

hay giả thuyết tương đương
$$\begin{cases} H_0 : r^2 = 0 \\ H_1 : r^2 > 0 \end{cases}$$

Chương 2

§2.4 Phân tích phương sai và sự phù hợp của mô hình

Để kiểm định giả thuyết này ta chọn TCKĐ:

$$F = \frac{r^2}{1 - r^2} \cdot \frac{n - 2}{1}$$

Nếu giả thuyết H_0 đúng thì $F \sim F(1, n-2)$

Chương 2

§2.4 Phân tích phương sai và sự phù hợp của mô hình

Theo nguyên lý xác suất nhỏ ta có :

$$W_{\alpha} = \{ f_{tn} : f_{tn} > f_{\alpha}(1, n-2) \}$$

Trong trường hợp H_0 bị bác bỏ, chấp nhận H_1 ta có thể nói rằng hàm hồi quy đưa ra là phù hợp

Chương 2

§2.5 Phân tích hồi quy và dự báo

Xét MHHQTT cổ điển và hàm hồi quy mẫu :

$$Y_i = \beta_1 + \beta_2 X_i + U_i \quad (2.1)$$

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i \quad (2.2)$$

Vấn đề đặt ra: cần dự báo giá trị trung bình $E(Y/X_0)$ và giá trị cá biệt Y_0 khi $X=X_0$

Chương 2

§2.5 Phân tích hồi quy và dự báo

Khi $X = X_0$ ta có: $\hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_0$

$$Var(\hat{Y}_0) = \sigma^2 \left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2} \right] = \hat{\sigma}^2 \left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2} \right]$$

$$Var(Y_0 - \hat{Y}_0) = \sigma^2 \left[1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2} \right] = \hat{\sigma}^2 \left[1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2} \right]$$

Chương 2

§2.5 Phân tích hồi quy và dự báo

2.5.1 Dự báo giá trị trung bình :

Xây dựng thống kê:

$$T = \frac{\hat{Y}_0 - E(Y / X_0)}{se(\hat{Y}_0)} \sim T(n - 2)$$

Chương 2

§2.5 Phân tích hồi quy và dự báo

Chọn phân vị $t_{\alpha/2}(n-2)$

$$P(|T| < t_{\alpha/2}(n-2)) = 1 - \alpha = \gamma$$

Ta có khoảng tin cậy $(1-\alpha)$ của $E(Y/X_0)$:

$$\left(\hat{Y}_0 - t_{\alpha/2}(n-2).se(\hat{Y}_0) \ ; \ \hat{Y}_0 + t_{\alpha/2}(n-2).se(\hat{Y}_0) \right)$$

Chương 2

§2.5 Phân tích hồi quy và dự báo

2.5.2 Dự báo giá trị cá biệt :

Xây dựng thống kê:

$$T = \frac{Y_0 - \hat{Y}_0}{se(Y_0 - \hat{Y}_0)} \sim T(n-2)$$

Chương 2

§2.5 Phân tích hồi quy và dự báo

Chọn phân vị $t_{\alpha/2}(n-2)$

$$P(|T| < t_{\alpha/2}(n-2)) = 1 - \alpha = \gamma$$

Ta có khoảng tin cậy $(1-\alpha)$ của Y_0 :

$$\left(\hat{Y}_0 - t_{\alpha/2}(n-2).se(Y_0 - \hat{Y}_0) \ ; \ \hat{Y}_0 + t_{\alpha/2}(n-2).se(Y_0 - \hat{Y}_0) \right)$$

Chương 3

MÔ HÌNH HỒI QUY NHIỀU BIẾN

Chương 3

MÔ HÌNH HỒI QUY NHIỀU BIẾN

3.1 Mô hình hồi quy nhiều biến

3.2 Phương pháp bình phương nhỏ nhất

3.3 Ước lượng và kiểm định giả thiết

3.4 Phân tích hồi quy và dự báo

Chương 3

§3.1 Mô hình hồi quy nhiều biến và phương pháp bình phương nhỏ nhất

3.1.1 Mô hình hồi quy nhiều biến

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + U_i \quad (3.1)$$

Trong đó:

Y_i : giá trị của biến phụ thuộc Y ($i = \overline{1, n}$)

β_1 hệ số chặn (hệ số tự do)

β_j hệ số góc (hệ số hồi quy riêng) của biến giải thích ($j = \overline{2, k}$)

U_i : sai số ngẫu nhiên

Chương 3

§3.1 Mô hình hồi quy nhiều biến và phương pháp bình phương nhỏ nhất

Mô hình hồi quy mẫu xây dựng dựa trên mẫu ngẫu nhiên kích thước n $\{(Y_i, X_{2i}, X_{3i}, \dots, X_{ki}), i = \overline{1, n}\}$

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \dots + \hat{\beta}_k X_{ki} \quad (3.2)$$

Trong đó:

$\hat{Y}_i$ ước lượng của Y_i ($i = \overline{1, n}$)

$\hat{\beta}_j$ ước lượng của hệ số hồi quy tổng thể
($j = \overline{1, k}$)

Chương 3

§3.1 Mô hình hồi quy nhiều biến và phương pháp bình phương nhỏ nhất

Ta ký hiệu

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_n \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_k \end{pmatrix} \quad U = \begin{pmatrix} U_1 \\ U_2 \\ \dots \\ U_n \end{pmatrix} \quad X = \begin{pmatrix} 1 & X_{21} & X_{31} & \dots & X_{k1} \\ 1 & X_{22} & X_{32} & \dots & X_{k2} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & X_{2n} & X_{3n} & \dots & X_{kn} \end{pmatrix}$$

Thì mô hình hồi quy tổng thể (3.1) có thể biểu diễn dưới dạng ma trận:

$$Y = X\beta + U \quad (3.3)$$

Chương 3

§3.1 Mô hình hồi quy nhiều biến và phương pháp bình phương nhỏ nhất

Tương tự, nếu ta ký hiệu

$$\hat{Y} = \begin{pmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \dots \\ \hat{Y}_n \end{pmatrix} \quad \hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \dots \\ \hat{\beta}_k \end{pmatrix}$$

Thì mô hình hồi quy mẫu (3.2) có thể biểu diễn dưới dạng ma trận như sau:

$$\hat{Y} = X\hat{\beta} \tag{3.4}$$

Chương 3

§3.1 Mô hình hồi quy nhiều biến và phương pháp bình phương nhỏ nhất

3.1.2 Các giả thiết cơ bản của MHHQ nhiều biến

Giả thiết 1. Các biến giải thích X_j ($j = \overline{2, k}$) không phải biến ngẫu nhiên, giá trị của chúng là xác định

Chương 3

§3.1 Mô hình hồi quy nhiều biến và phương pháp bình phương nhỏ nhất

Giả thiết 2. Kỳ vọng toán của các sai số ngẫu nhiên U_i bằng không

$$E(U_i) = E(U / X_i) = 0 \quad (\forall i)$$

Giả thiết 3.

$$\text{cov}(U_i, U_j) = E(U_i \cdot U_j) = \begin{cases} \sigma^2 & (\forall i = j) \\ 0 & (\forall i \neq j) \end{cases}$$

Chương 3

§3.1 Mô hình hồi quy nhiều biến và phương pháp bình phương nhỏ nhất

Giả thiết 4. Hạng ma trận X bằng k

$$\text{rg}(X) = k$$

Giả thiết này có nghĩa giữa các biến X_j không có hiện tượng cộng tuyến hay các cột của ma trận X độc lập tuyến tính

Giả thiết 5. $U_i \sim N(0, \sigma^2) \quad (\forall i)$

Chương 3

§3.1 Mô hình hồi quy nhiều biến và phương pháp bình phương nhỏ nhất

3.1.3 Phương pháp bình phương nhỏ nhất

Xét hàm hồi quy tổng thể và hàm hồi quy mẫu

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + U_i \quad (3.1)$$

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \dots + \hat{\beta}_k X_{ki} \quad (3.2)$$

hoặc ở dạng ma trận

$$Y = X\beta + U \quad (3.3)$$

$$\hat{Y} = X\hat{\beta} \quad (3.4)$$

Chương 3

§3.1 Mô hình hồi quy nhiều biến và phương pháp bình phương nhỏ nhất

Ta ký hiệu các phần dư e_i :

$$e_i = Y_i - \hat{Y}_i$$

Các phần dư này cũng có thể biểu diễn dưới dạng ma trận như sau:

$$\mathbf{e} = \begin{pmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{pmatrix} = \begin{pmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_n \end{pmatrix} - \begin{pmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \dots \\ \hat{Y}_n \end{pmatrix} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}$$

Chương 3

§3.1 Mô hình hồi quy nhiều biến và phương pháp bình phương nhỏ nhất

Theo phương pháp bình phương nhỏ nhất, khi xây dựng hàm hồi quy mẫu, các hệ số hồi quy mẫu $\hat{\beta}_j$ phải được xác định sao cho tổng bình phương các phần dư đạt giá trị nhỏ nhất, tức là:

$$\sum e_i^2 \rightarrow \min$$

Chương 3

§3.1 Mô hình hồi quy nhiều biến và phương pháp bình phương nhỏ nhất

Ta có $\sum e_i^2 = e^T e$

$$\sum e_i^2 \rightarrow \min \Leftrightarrow \frac{\partial(e^T e)}{\partial \hat{\beta}} = 0$$

Giải phương trình trên ta được:

$$\hat{\beta} = (X^T X)^{-1} \cdot X^T Y \quad (3.5)$$

Chương 3

§3.1 Mô hình hồi quy nhiều biến và phương pháp bình phương nhỏ nhất

$$\hat{\beta} = (X^T X)^{-1} \cdot X^T Y \quad (3.5)$$

Công thức (3.5) là công thức xác định hệ số hồi quy mẫu $\hat{\beta}_j$ theo phương pháp bình phương nhỏ nhất và các ước lượng $\hat{\beta}_j$ được xác định theo công thức (3.5) được gọi là các ước lượng bình phương nhỏ nhất.

Chương 3

§3.1 Mô hình hồi quy nhiều biến và phương pháp bình phương nhỏ nhất

Ma trận $X^T X$ được xác định như sau:

$$\begin{aligned} X^T X &= \begin{pmatrix} 1 & 1 & \dots & 1 \\ X_{21} & X_{22} & \dots & X_{2n} \\ \dots & \dots & \dots & \dots \\ X_{k1} & X_{k2} & \dots & X_{kn} \end{pmatrix} \begin{pmatrix} 1 & X_{21} & \dots & X_{k1} \\ 1 & X_{22} & \dots & X_{k2} \\ \dots & \dots & \dots & \dots \\ 1 & X_{2n} & \dots & X_{kn} \end{pmatrix} \\ &= \begin{pmatrix} n & \sum X_{2i} & \sum X_{3i} & \dots & \sum X_{ki} \\ \sum X_{2i} & \sum X_{2i}^2 & \sum X_{2i} X_{3i} & \dots & \sum X_{2i} X_{ki} \\ \dots & \dots & \dots & \dots & \dots \\ \sum X_{ki} & \sum X_{ki} X_{2i} & \sum X_{ki} X_{3i} & \dots & \sum X_{ki}^2 \end{pmatrix} \end{aligned}$$

Chương 3

§3.1 Mô hình hồi quy nhiều biến và phương pháp bình phương nhỏ nhất

Ma trận $X^T Y$ cũng được xác định tương tự:

$$X^T Y = \begin{pmatrix} 1 & 1 & \dots & 1 \\ X_{21} & X_{22} & \dots & X_{2n} \\ \dots & \dots & \dots & \dots \\ X_{k1} & X_{k2} & \dots & X_{kn} \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_n \end{pmatrix} = \begin{pmatrix} \sum Y_i \\ \sum Y_i X_{2i} \\ \dots \\ \sum Y_i X_{ki} \end{pmatrix}$$

VÍ DỤ 3.1

Nghiên cứu mối quan hệ phụ thuộc giữa doanh số bán ra với chi phí dành cho quảng cáo và giá bán, người ta thu thập được các số liệu sau đây tại 10 cửa hàng cùng kinh doanh một loại mặt hàng:

Y_i	84	90	92	96	100	108	120	126	130	136
X_i	8	9	10	9	10	12	13	14	14	15
Z_i	9	8	8	7	7	8	7	7	6	6

Trong đó:

Y_i : doanh số bán ra trong một tháng của cửa hàng thứ i (triệu đồng)

X_i : chi phí dành cho quảng cáo trong một tháng của cửa hàng thứ i (triệu đồng)

Z_i : giá bán của cửa hàng thứ i
(ngàn đồng/1 đv sp)

Bằng phương pháp bình phương nhỏ nhất và dựa vào số liệu trên, hãy xây dựng hàm hồi quy mẫu dưới dạng sau:

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{\beta}_3 Z_i$$

Đáp số:

$$X^T Y = \begin{pmatrix} \sum Y_i \\ \sum Y_i X_i \\ \sum Y_i Z_i \end{pmatrix} = \begin{pmatrix} 1082 \\ 12746 \\ 7766 \end{pmatrix}$$

$$X^T X = \begin{pmatrix} n & \sum X_i & \sum Z_i \\ \sum X_i & \sum X_i^2 & \sum X_i Z_i \\ \sum Z_i & \sum Z_i X_i & \sum Z_i^2 \end{pmatrix} = \begin{pmatrix} 10 & 114 & 73 \\ 114 & 1356 & 816 \\ 73 & 816 & 541 \end{pmatrix} \quad |X^T X| = 1944$$

$$X^T X = \begin{pmatrix} A_{11} & A_{21} & A_{31} \\ A_{12} & A_{22} & A_{32} \\ A_{13} & A_{23} & A_{33} \end{pmatrix} = \begin{pmatrix} 67740 & -2106 & -5964 \\ -2106 & 81 & 162 \\ -5964 & 162 & 564 \end{pmatrix}$$

$$\left(X^T X\right)^{-1} = \frac{1}{\left|X^T X\right|} X^T X = \frac{1}{1944} \begin{pmatrix} 67740 & -2106 & -5964 \\ -2106 & 81 & 162 \\ -5964 & 162 & 564 \end{pmatrix}$$

$$\hat{\beta} = \left(X^T X\right)^{-1} . X^T Y = \frac{1}{1944} \begin{pmatrix} 67740 & -2106 & -5964 \\ -2106 & 81 & 162 \\ -5964 & 162 & 564 \end{pmatrix} \begin{pmatrix} 1082 \\ 12746 \\ 7766 \end{pmatrix} = \begin{pmatrix} 69,53704 \\ 6,08333 \\ -4,20370 \end{pmatrix}$$

$$\hat{Y}_i = 69,53704 + 6,08333 X_i - 4,20370 Z_i$$

Ý nghĩa của các hệ số hồi quy

$\hat{\beta}_2 = 6.08333$: Khi giá bán không đổi, chi phí dành cho quảng cáo tăng lên 1 triệu đồng, thì doanh số bán ra trung bình của cửa hàng tăng lên 6.08333 triệu đồng.

$\hat{\beta}_3 = -4.2037$: Khi chi phí dành cho quảng cáo không đổi, giá bán tăng lên 1 ngàn đồng/ 1 đv sản phẩm, thì doanh số bán ra trung bình của cửa hàng giảm xuống 4.2037 triệu đồng.

Chương 3

§3.1 Mô hình hồi quy nhiều biến và phương pháp bình phương nhỏ nhất

3.1.4 Các tính chất của ước lượng BPNN

1. Đường hồi quy mẫu đi qua điểm trung bình mẫu $(\bar{Y}, \bar{X}_2, \dots, \bar{X}_k)$, tức là:

$$\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}_2 + \dots + \hat{\beta}_k \bar{X}_k$$

trong đó:

$$\bar{Y} = \frac{1}{n} \sum Y_i \qquad \bar{X}_j = \frac{1}{n} \sum X_{ji} \qquad (j = \overline{2, k})$$

Chương 3

§3.1 Mô hình hồi quy nhiều biến và phương pháp bình phương nhỏ nhất

2. Giá trị trung bình của các giá trị $\hat{Y}_i$ được xác định theo hàm hồi quy mẫu bằng giá trị trung bình của biến phụ thuộc, tức là:

$$\bar{\hat{Y}} = \frac{1}{n} \sum \hat{Y}_i = \bar{Y}$$

3. Tổng các phần dư của hàm hồi quy mẫu bằng 0:

$$\sum e_i = 0$$

Chương 3

§3.1 Mô hình hồi quy nhiều biến và phương pháp bình phương nhỏ nhất

4. Các phần dư e_i không tương quan với $\hat{Y}_i$:

$$\sum e_i \hat{Y}_i = 0$$

5. Các phần dư e_i không tương quan với X_{ji} :

$$\sum e_i X_{ji} = 0 \quad (j = \overline{2, k})$$

Chương 3

§3.1 Mô hình hồi quy nhiều biến và phương pháp bình phương nhỏ nhất

6. Với các giả thiết của mô hình hồi quy tuyến tính cổ điển thì các ước lượng bình phương nhỏ nhất $\hat{\beta}_j$ là các ước lượng hiệu quả của β_j ($j = \overline{1, k}$).

Chương 3

§3.2 Khoảng tin cậy và kiểm định giả thuyết về các hệ số hồi quy

Xét hàm hồi quy tổng thể và hàm hồi quy mẫu

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + U_i \quad (3.1)$$

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \dots + \hat{\beta}_k X_{ki} \quad (3.2)$$

Chương 3

§3.2 Khoảng tin cậy và kiểm định giả thuyết về các hệ số hồi quy

3.2.1 Ma trận hiệp phương sai của hệ số hồi quy mẫu

Ma trận hiệp phương sai của hệ số hồi quy mẫu, kí hiệu $\text{cov}(\hat{\beta})$, là ma trận được xác định như sau:

$$\text{cov}(\hat{\beta}) = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T] \quad (3.6)$$

Chương 3

§3.2 Khoảng tin cậy và kiểm định giả thuyết về các hệ số hồi quy

$$\text{cov}(\hat{\beta}) = \begin{pmatrix} \text{Var}(\hat{\beta}_1) & \text{cov}(\hat{\beta}_1, \hat{\beta}_2) & \dots & \text{cov}(\hat{\beta}_1, \hat{\beta}_k) \\ \text{cov}(\hat{\beta}_2, \hat{\beta}_1) & \text{Var}(\hat{\beta}_2) & \dots & \text{cov}(\hat{\beta}_2, \hat{\beta}_k) \\ \dots & \dots & \dots & \dots \\ \text{cov}(\hat{\beta}_k, \hat{\beta}_1) & \text{cov}(\hat{\beta}_k, \hat{\beta}_2) & \dots & \text{Var}(\hat{\beta}_k) \end{pmatrix} \quad (3.7)$$

Ma trận hiệp phương sai của các hệ số hồi quy mẫu là ma trận vuông cấp k , đối xứng qua đường chéo chính và phần tử thứ j trên đường chéo chính là phương sai của $\hat{\beta}_j$

Chương 3

§3.2 Khoảng tin cậy và kiểm định giả thuyết về các hệ số hồi quy

Có thể chứng minh được rằng:

$$\text{cov}(\hat{\beta}) = \sigma^2 (X^T X)^{-1} \quad (3.8)$$

Do vậy ta có:

$$\text{Var}(\hat{\beta}_j) = \frac{\sigma^2}{|X^T X|} A_{jj} \quad (3.9)$$

Chương 3

§3.2 Khoảng tin cậy và kiểm định giả thuyết về các hệ số hồi quy

Trong thực hành khi sử dụng công thức (3.8) và (3.9), do phương sai chưa biết, nên người ta thường thay σ^2 bằng ước lượng không chệch của nó là:

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n - k} \quad (3.10)$$

Chương 3

§3.2 Khoảng tin cậy và kiểm định giả thuyết về các hệ số hồi quy

Trong thực hành người ta thường sử dụng công thức sau đây để xác định $\sum e_i^2$:

$$\sum e_i^2 = e^T e = Y^T Y - \hat{\beta}^T X^T Y \quad (3.11)$$

Nếu khai triển công thức (3.11) ta được:

$$\sum e_i^2 = \sum Y_i^2 - \left(\hat{\beta}_1 \sum Y_i + \hat{\beta}_2 \sum Y_i X_{2i} + \dots + \hat{\beta}_k \sum Y_i X_{ki} \right) \quad (3.12)$$

Chương 3

§3.2 Khoảng tin cậy và kiểm định giả thuyết về các hệ số hồi quy

3.2.2 Khoảng tin cậy của các hệ số hồi quy

Từ giả thiết 5 về phân phối chuẩn của sai số ngẫu nhiên, có thể suy ra:

$$\hat{\beta}_j \sim N(\beta_j, \text{Var}(\hat{\beta}_j)) \quad (j = \overline{1, k})$$

Do σ^2 ta chưa biết mà phải thay bằng ước lượng không chệch của nó là $\hat{\sigma}^2$, nên

$$T = \frac{\hat{\beta}_j - \beta_j}{\text{se}(\hat{\beta}_j)} \sim T(n - k) \quad (j = \overline{1, k})$$

Chương 3

§3.2 Khoảng tin cậy và kiểm định giả thuyết về các hệ số hồi quy

$$T = \frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \sim T(n-k) \quad (j = \overline{1, k})$$

$$P\left(\left|\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)}\right| < t_{\frac{\alpha}{2}}(n-k)\right) = 1 - \alpha$$

$$P\left(\hat{\beta}_j - t_{\frac{\alpha}{2}}(n-k).se(\hat{\beta}_j) < \beta_j < \hat{\beta}_j + t_{\frac{\alpha}{2}}(n-k).se(\hat{\beta}_j)\right) = 1 - \alpha = \gamma$$

$$\left(\hat{\beta}_j - t_{\frac{\alpha}{2}}(n-k).se(\hat{\beta}_j) ; \hat{\beta}_j + t_{\frac{\alpha}{2}}(n-k).se(\hat{\beta}_j)\right)$$

Chương 3

§3.2 Khoảng tin cậy và kiểm định giả thuyết về các hệ số hồi quy

3.2.3 Kiểm định giả thuyết về các hệ số HQ

Giả sử với mức ý nghĩa α cho trước ta cần kiểm định giả thuyết:

$$\begin{cases} H_0 : \beta_j = \beta_j^* \\ H_1 : \beta_j \neq \beta_j^* \end{cases} \quad (\beta_j < \beta_j^*, \quad \beta_j > \beta_j^*)$$

Chương 3

§3.2 Khoảng tin cậy và kiểm định giả thuyết về các hệ số hồi quy

3.2.3 Kiểm định giả thuyết về các hệ số HQ

Ta xây dựng tiêu chuẩn kiểm định

$$T = \frac{\hat{\beta}_j - \beta_j^*}{se(\hat{\beta}_j)}$$

Nếu H_0 đúng thì $T \sim T(n-k)$

Chương 3

§3.2 Khoảng tin cậy và kiểm định giả thuyết về các hệ số hồi quy

Loại gt	H0	H1	W_α
Hai phía	$\beta_j = \beta_j^*$	$\beta_j \neq \beta_j^*$	$W_\alpha = \{ t_{tn} : t_{tn} > t_{\alpha/2}(n-k) \}$
Trái	$\beta_j = \beta_j^*$	$\beta_j < \beta_j^*$	$W_\alpha = \{ t_{tn} : t_{tn} < -t_\alpha(n-k) \}$
Phải	$\beta_j = \beta_j^*$	$\beta_j > \beta_j^*$	$W_\alpha = \{ t_{tn} : t_{tn} > t_\alpha(n-k) \}$

Chương 3

§3.3 Phân tích phương sai và kiểm định giả thuyết đồng thời

3.3.1 Hệ số xác định bội

Xét hàm hồi quy tổng thể và hàm hồi quy mẫu:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + U_i \quad (3.1)$$

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \dots + \hat{\beta}_k X_{ki} \quad (3.2)$$

Chương 3

§3.3 Phân tích phương sai và kiểm định giả thuyết đồng thời

1. $TSS = \sum (Y_i - \bar{Y})^2$
2. $ESS = \sum (\hat{Y}_i - \bar{\hat{Y}})^2 = \sum (\hat{Y}_i - \bar{Y})^2$
3. $RSS = \sum (Y_i - \hat{Y}_i)^2$

Tương tự trường hợp hồi quy 2 biến ta có hệ thức sau:

$$\boxed{TSS = ESS + RSS}$$

Chương 3

§3.3 Phân tích phương sai và kiểm định giả thuyết đồng thời

Định nghĩa 1: Hệ số xác định bội R^2 được định nghĩa như sau:

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

Chương 3

§3.3 Phân tích phương sai và kiểm định giả thuyết đồng thời

Trong thực hành ta có thể sử dụng công thức:

$$R^2 = \frac{\hat{\beta}^T X^T Y - n\bar{Y}^2}{Y^T Y - n\bar{Y}^2}$$

Nếu khai triển ta được

$$R^2 = \frac{\hat{\beta}_1 \sum Y_i + \hat{\beta}_2 \sum Y_i X_{2i} + \dots + \hat{\beta}_k \sum Y_i X_{ki} - n\bar{Y}^2}{\sum Y_i^2 - n\bar{Y}^2}$$

Chương 3

§3.3 Phân tích phương sai và kiểm định giả thuyết đồng thời

1. $0 \leq R^2 \leq 1$

- Nếu $R^2 = 1$, hàm hồi quy có thể coi là hoàn hảo
 - Nếu $R^2 = 0$, hàm hồi quy đưa ra là không phù hợp
- Vì thế R^2 được dùng làm thước đo mức độ phù hợp của hàm hồi quy

2. R^2 là hàm không giảm, phụ thuộc vào số biến giải thích có trong mô hình

Chương 3

§3.3 Phân tích phương sai và kiểm định giả thuyết đồng thời

Tuy nhiên không thể dùng R^2 làm tiêu chuẩn để xét việc đưa thêm hay không đưa thêm biến độc lập mới vào mô hình mà phải dùng hệ số xác định bội đã điều chỉnh

Chương 3

§3.3 Phân tích phương sai và kiểm định giả thuyết đồng thời

Định nghĩa 2: Hệ số xác định bội đã điều chỉnh, ký hiệu $\overline{R}^2$ được định nghĩa như sau:

$$\overline{R}^2 = 1 - (1 - R^2) \frac{n - 1}{n - k}$$

Chương 3

§3.3 Phân tích phương sai và kiểm định giả thuyết đồng thời

$\overline{R}^2$ có các tính chất:

1. Nếu $k > 1$ thì $\overline{R}^2 \leq R^2 \leq 1$ và $\overline{R}^2$ cũng là hàm không giảm đối với số biến giải thích có trong mô hình
2. $\overline{R}^2$ có thể nhận giá trị âm dù R^2 luôn dương

Chương 3

§3.3 Phân tích phương sai và kiểm định giả thuyết đồng thời

Vậy khi nào cần đưa thêm biến độc lập mới vào mô hình? Có thể chứng minh được rằng việc đưa thêm biến giải thích mới vào mô hình là cần thiết chừng nào $\bar{R}^2$ còn tăng lên và hệ số hồi quy của biến mới X_j là $\beta_j \neq 0$

Chương 3

§3.3 Phân tích phương sai và kiểm định giả thuyết đồng thời

3.3.2 Kiểm định giả thuyết đồng thời

Xét giả thuyết

$$\begin{cases} H_0 : \beta_2 = \beta_3 = \dots = \beta_k = 0 \\ H_1 : \exists \text{ ít nhất một hệ số } \beta_j \neq 0 \quad (j = \overline{2, k}) \end{cases}$$

Giả thuyết tương đương là
$$\begin{cases} H_0 : R^2 = 0 \\ H_1 : R^2 > 0 \end{cases}$$

Chương 3

§3.3 Phân tích phương sai và kiểm định giả thuyết đồng thời

Ta xây dựng tiêu chuẩn kiểm định:

$$F = \frac{R^2}{1 - R^2} \cdot \frac{n - k}{k - 1}$$

Nếu H_0 đúng thì $F \sim F(k-1, n-k)$

$$P(F > f_{\alpha}(k-1, n-k)) = \alpha$$

$$W_{\alpha} = \{f_{tn} : f_{tn} > f_{\alpha}(k-1, n-k)\}$$

Chương 3

§3.4 Phân tích hồi quy và dự báo

Xét hàm hồi quy tổng thể và hàm hồi quy mẫu

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + U_i \quad (3.1)$$

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \dots + \hat{\beta}_k X_{ki} \quad (3.2)$$

hoặc ở dạng ma trận

$$Y = X\beta + U \quad (3.3)$$

$$\hat{Y} = X\hat{\beta} \quad (3.4)$$

Chương 3

§3.4 Phân tích hồi quy và dự báo

Bài toán đặt ra: với các giá trị cho trước của biến giải thích $X_2=X_{20}$, $X_3=X_{30}$, ..., $X_k=X_{k0}$ hoặc có thể ký hiệu

$$X_0 = \begin{pmatrix} 1 \\ X_{20} \\ X_{30} \\ \dots \\ X_{k0} \end{pmatrix}$$

cần dự báo giá trị trung bình $E(Y/X_0)$ hoặc giá trị cá biệt $Y=Y_0$ khi $X=X_0$

Chương 3

§3.4 Phân tích hồi quy và dự báo

3.4.1 Dự báo giá trị trung bình

Với độ tin cậy $\gamma = 1 - \alpha$ cần dự báo $E(Y/X_0)$

Ước lượng điểm của $E(Y/X_0)$ là:

$$\hat{Y}_0 = X_0^T \cdot \hat{\beta} = \hat{\beta}_1 + \hat{\beta}_2 X_{20} + \hat{\beta}_3 X_{30} + \dots + \hat{\beta}_k X_{k0}$$

Chương 3

§3.4 Phân tích hồi quy và dự báo

Xây dựng thống kê

$$T = \frac{\hat{Y}_0 - E(Y / X_0)}{se(\hat{Y}_0)} \sim T(n-k)$$

ta tìm giá trị phân vị $t_{\alpha/2}(n-k)$ sao cho:

$$P(|T| < t_{\alpha/2}(n-k)) = 1 - \alpha = \gamma$$

Chương 3

§3.4 Phân tích hồi quy và dự báo

Bằng phép biến đổi tương đương ta suy ra khoảng tin cậy cho $E(Y/X_0)$:

$$\left(\hat{Y}_0 - t_{\alpha/2}(n-k).se(\hat{Y}_0) ; \hat{Y}_0 + t_{\alpha/2}(n-k).se(\hat{Y}_0) \right)$$

Chương 3

§3.4 Phân tích hồi quy và dự báo

Trong đó

$$Var(\hat{Y}_0) = X_0^T \cdot \text{cov}(\hat{\beta}) \cdot X_0 = \sigma^2 \cdot X_0^T \cdot (X^T X)^{-1} X_0$$

$$se(\hat{Y}_0) = \sqrt{Var(\hat{Y}_0)} = \sigma \sqrt{X_0^T \cdot (X^T X)^{-1} X_0}$$

Chương 3

§3.4 Phân tích hồi quy và dự báo

3.4.2 Dự báo giá trị cá biệt

Với độ tin cậy γ cần dự báo giá trị $Y=Y_0$ khi $X=X_0$

Ước lượng điểm của Y_0 vẫn là:

$$\hat{Y}_0 = X_0^T \cdot \hat{\beta} = \hat{\beta}_1 + \hat{\beta}_2 X_{20} + \hat{\beta}_3 X_{30} + \dots + \hat{\beta}_k X_{k0}$$

Chương 3

§3.4 Phân tích hồi quy và dự báo

Hoàn toàn tương tự ta xây dựng thống kê

$$T = \frac{Y_0 - \hat{Y}_0}{se(Y_0 - \hat{Y}_0)} \sim T(n-k)$$

Bằng phép biến đổi tương đương ta cũng suy ra được khoảng tin cậy của Y_0 là

$$\left(\hat{Y}_0 - t_{\alpha/2}(n-k) se(Y_0 - \hat{Y}_0); \hat{Y}_0 + t_{\alpha/2}(n-k) se(Y_0 - \hat{Y}_0) \right)$$

Chương 3

§3.4 Phân tích hồi quy và dự báo

Trong đó

$$\text{Var}(Y_0 - \hat{Y}_0) = \text{Var}(\hat{Y}_0) + \sigma^2$$

$$\text{se}(Y_0 - \hat{Y}_0) = \sqrt{\text{Var}(Y_0 - \hat{Y}_0)}$$

Chương 3

§3.4 Phân tích hồi quy và dự báo

Ví dụ: Xét tiếp ví dụ 2. Với độ tin cậy $\gamma = 0,98$ hãy dự báo doanh số bán ra trung bình trong một tháng của các cửa hàng có chi phí dành cho quảng cáo là 10 triệu đồng/ tháng và giá bán là 8 ngàn đồng/ đvsp.

Chương 3

§3.4 Phân tích hồi quy và dự báo

Xét tiếp ví dụ 2. Với độ tin cậy $\gamma = 0,98$ hãy dự báo doanh số bán ra trong một tháng của cửa hàng có chi phí dành cho quảng cáo là 10 triệu đồng/ tháng và giá bán là 8 ngàn đồng/ đ.vị.

Chương 4

MÔ HÌNH HỒI QUY VỚI BIẾN GIẢ

Chương 4

MÔ HÌNH HỒI QUY VỚI BIẾN GIẢ

4.1 Mô hình hồi quy với biến giả

4.2 Ứng dụng của mô hình hồi quy với biến giả

Chương 4

§4.1 Mô hình hồi quy với biến giả

4.1.1 Khái niệm về biến giả

Biến số lượng: Giá trị của các biến đó được biểu thị bằng số (ví dụ: thu nhập, doanh số...)

Biến chất lượng: Biểu thị những thuộc tính nào đó (ví dụ: giới tính, nghề nghiệp...)

Chương 4

§4.1 Mô hình hồi quy với biến giả

Để biểu thị mức độ ảnh hưởng của các biến chất lượng tới biến phụ thuộc, ta cần lượng hóa các tiêu thức, thuộc tính này bằng cách sử dụng biến giả

Chương 4

§4.1 Mô hình hồi quy với biến giả

Vậy biến giả là gì? Là biến chất lượng đã được lượng hóa, các giá trị có thể có của biến giả chỉ là 2 giá trị 0 và 1. Nó chỉ ra có hay không có một thuộc tính nào đó.

VD: Để biểu thị giới tính, ta sử dụng biến giả Z và quy ước:

- $Z = 0$ Nam
- $Z = 1$ Nữ

Chương 4

§4.1 Mô hình hồi quy với biến giả

4.1.2 MHHQ với biến chất lượng có 2 phạm trù

Giả sử một xí nghiệp sản xuất có thể áp dụng 2 công nghệ sản xuất A và B, năng suất của mỗi công nghệ là ĐLNN phân phối theo quy luật chuẩn có phương sai bằng nhau, nhưng kỳ vọng toán có thể khác nhau

Chương 4

§4.1 Mô hình hồi quy với biến giả

$$Y_i = \beta_1 + \beta_2 Z_i + U_i \quad (4.1)$$

Trong đó:

Y_i : năng suất của xí nghiệp

U_i : sai số ngẫu nhiên

Z_i : biến giả biểu thị công nghệ sản xuất được áp dụng và có thể quy ước:

- $Z_i = 0$ công nghệ sản xuất A

- $Z_i = 1$ công nghệ sản xuất B

Chương 4

§4.1 Mô hình hồi quy với biến giả

Để ước lượng MHHQ (4.1) ta cũng sẽ tiến hành tương tự đối với MHHQ 2 biến thông thường

Y_i	22	19	18	21	18.5	21	20.5	17	17.5	21.5
Z_i	1	0	0	1	0	1	1	0	0	1

Chương 4

§4.1 Mô hình hồi quy với biến giả

Để ước lượng MHHQ (4.1) ta cũng sẽ tiến hành tương tự đối với MHHQ 2 biến thông thường

$$\hat{\beta}_2 = \frac{\sum y_i z_i}{\sum z_i^2} = \frac{8}{2,5} = 3,2$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{Z} = 19,6 - 3,2 \times 0,5 = 18$$

$$\hat{Y}_i = 18 + 3,2 Z_i$$

Chương 4

§4.1 Mô hình hồi quy với biến giả

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

$$E(Y_i / Z_i = 0) = \beta_1 \quad (4.2)$$

$$E(Y_i / Z_i = 1) = \beta_1 + \beta_2 \quad (4.3)$$

-(4.2) cho biết khi áp dụng công nghệ sản xuất A, năng suất trung bình của xí nghiệp là β_1

-(4.3) cho biết khi áp dụng công nghệ sản xuất B, năng suất trung bình của xí nghiệp là $\beta_1 + \beta_2$

Chương 4

§4.1 Mô hình hồi quy với biến giả

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

$$E(Y_i / Z_i = 0) = \beta_1 \quad (4.2)$$

$$E(Y_i / Z_i = 1) = \beta_1 + \beta_2 \quad (4.3)$$

- Như vậy β_2 là sự chênh lệch (khác nhau) về năng suất khi chuyển từ công nghệ sản xuất A sang công nghệ sản xuất B

Chương 4

§4.1 Mô hình hồi quy với biến giả

Trở lại ví dụ vừa rồi:

$$\hat{Y}_i = 18 + 3,2Z_i$$

Năng suất trung bình khi áp dụng công nghệ sản xuất A là 18 (đvsp)

Năng suất trung bình khi áp dụng công nghệ sản xuất B là $18 + 3,2 = 21,2$ (đvsp)

Chương 4

§4.1 Mô hình hồi quy với biến giả

4.1.3 MHHQ với biến chất lượng có nhiều hơn 2 phạm trù

Nếu kí hiệu số phạm trù là m thì số biến giả cần đưa vào mô hình (để lượng hóa biến chất lượng) sẽ là $m - 1$

Chương 4

§4.1 Mô hình hồi quy với biến giả

Giả sử xí nghiệp nọ ngoài công nghệ sản xuất A và B còn có thể áp dụng công nghệ sản xuất C, khi đó ta cần sử dụng 2 biến giả là Z_{1i} và Z_{2i} và mô hình hồi quy tổng thể sẽ có dạng sau:

$$Y_i = \beta_1 + \beta_2 Z_{1i} + \beta_3 Z_{2i} + U_i \quad (4.4)$$

Chương 4

§4.1 Mô hình hồi quy với biến giả

$$Y_i = \beta_1 + \beta_2 Z_{1i} + \beta_3 Z_{2i} + U_i \quad (4.4)$$

Y_i : biến phụ thuộc (năng suất)

U_i : sai số ngẫu nhiên

Z_{1i}, Z_{2i} : biến giả biểu thị các công nghệ sản xuất được áp dụng

	Z_{1i}	Z_{2i}
A	0	0
B	1	0
C	0	1

Chương 4

§4.1 Mô hình hồi quy với biến giả

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

$$E(Y_i / Z_{1i} = Z_{2i} = 0) = \beta_1$$

$$E(Y_i / Z_{1i} = 1, Z_{2i} = 0) = \beta_1 + \beta_2$$

$$E(Y_i / Z_{1i} = 0, Z_{2i} = 1) = \beta_1 + \beta_3$$

Chương 4

§4.1 Mô hình hồi quy với biến giả

- Năng suất trung bình của xí nghiệp khi áp dụng công nghệ sản xuất A là β_1
- Năng suất trung bình của xí nghiệp khi áp dụng công nghệ sản xuất B là $\beta_1 + \beta_2$
- Năng suất trung bình của xí nghiệp khi áp dụng công nghệ sản xuất C là $\beta_1 + \beta_3$

Chương 4

§4.1 Mô hình hồi quy với biến giả

$$E(Y_i / Z_{1i} = Z_{2i} = 0) = \beta_1$$

$$E(Y_i / Z_{1i} = 1, Z_{2i} = 0) = \beta_1 + \beta_2$$

$$E(Y_i / Z_{1i} = 0, Z_{2i} = 1) = \beta_1 + \beta_3$$

- β_2 : chênh lệch (khác nhau) về năng suất khi chuyển từ CNSX A sang CNSX B
- β_3 : chênh lệch (khác nhau) về năng suất khi chuyển từ CNSX A sang CNSX C

Chương 4

§4.1 Mô hình hồi quy với biến giả

$$Y_i = \beta_1 + \beta_2 Z_{1i} + \beta_3 Z_{2i} + U_i \quad (4.4)$$

$$H_0: \beta_2 = \beta_3 = 0$$

Nếu ở mức ý nghĩa α nào đó ta không bác bỏ được H_0 thì điều này có nghĩa (ở mức ý nghĩa α đó) các công nghệ sản xuất khác nhau cho năng suất như nhau

Chương 4

§4.1 Mô hình hồi quy với biến giả

Chú ý

Phạm trù ứng với các giá trị bằng 0 của các biến giả được gọi là phạm trù cơ sở.

Phạm trù cơ sở hiểu theo nghĩa là việc so sánh được tiến hành với phạm trù này

Chương 4

§4.1 Mô hình hồi quy với biến giả

4.1.4 MHHQ với nhiều biến chất lượng

Giả sử MHHQ có k biến giải thích là biến chất lượng, số phạm trù của X_j là m_j

Tổng số biến giả cần thiết để biểu thị các biến chất lượng cần đưa vào mô hình sẽ là:

$$\sum (m_j - 1)$$

Chương 4

§4.1 Mô hình hồi quy với biến giả

Xét mô hình hồi quy biểu thị nhu cầu về may mặc của nam, nữ thanh niên thuộc các nghề nghiệp khác nhau: công nhân, nông dân, tiểu thương, sinh viên.

Số biến giả cần đưa vào mô hình sẽ là:

$$(2 - 1) + (4 - 1) = 4$$

Chương 4

§4.1 Mô hình hồi quy với biến giả

	Z_{1i}	Z_{2i}	Z_{3i}	Z_{4i}
Nam SV	0	0	0	0
Nữ SV	1	0	0	0
Nam CN	0	1	0	0
Nữ CN	1	1	0	0
Nam ND	0	0	1	0
Nữ ND	1	0	1	0
Nam TT	0	0	0	1
Nữ TT	1	0	0	1

Chương 4

§4.1 Mô hình hồi quy với biến giả

4.1.5 MHHQ hỗn hợp

Là mô hình hồi quy mà trong đó có cả biến số lượng lẫn biến chất lượng, biến chất lượng sẽ được biểu thị bằng biến giả theo quy tắc như đã nói ở trên.

Chương 4

§4.1 Mô hình hồi quy với biến giả

4.1.5 MHHQ hỗn hợp

Việc phân tích hồi quy mối quan hệ giữa biến phụ thuộc với các biến giải thích (bao gồm cả biến số lượng và biến giả), được tiến hành tương tự như đối với mô hình hồi quy nhiều biến

VÍ DỤ 3

Điều tra về mức chi tiêu cho may mặc của nam, nữ công nhân ở một nhà máy, người ta thu được số liệu sau:

Y_i	3.0	4.0	3.5	4.5	4.0	4.5	4.0	5.5
X_i	2.5	2.5	3.0	3.0	3.5	3.5	4.0	4.0
Z_i	0	1	0	1	0	1	0	1

Trong đó:

Y_i : mức chi tiêu cho may mặc trong một năm (triệu đồng)

X_i : mức thu nhập trong một tháng (triệu đồng)

Z_i : là biến giả biểu thị giới tính, trong đó:

$Z_i = 0$: nam $Z_i = 1$: nữ

Bằng phương pháp bình phương nhỏ nhất và dựa vào số liệu trên, hãy xây dựng hàm hồi quy mẫu dưới dạng sau:

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{\beta}_3 Z_i$$

Đáp số: ($Y_1=14.56$)

$$X^T Y = \begin{pmatrix} \sum Y_i \\ \sum Y_i X_i \\ \sum Y_i Z_i \end{pmatrix} = \begin{pmatrix} 7,28 \\ 92,016 \\ 7,64 \end{pmatrix}$$

$$X^T X = \begin{pmatrix} n & \sum X_i & \sum Z_i \\ \sum X_i & \sum X_i^2 & \sum X_i Z_i \\ \sum Z_i & \sum Z_i X_i & \sum Z_i^2 \end{pmatrix} = \begin{pmatrix} 8 & 50,4 & 4 \\ 5,4 & 321,2 & 25,2 \\ 4 & 25,2 & 4 \end{pmatrix} \quad |X^T X| = 57,6$$

$$X^T X = \begin{pmatrix} A_{11} & A_{21} & A_{31} \\ A_{12} & A_{22} & A_{32} \\ A_{13} & A_{23} & A_{33} \end{pmatrix} = \begin{pmatrix} 649,44 & -100,8 & -14,4 \\ -100,8 & 16 & 0 \\ -14,4 & 0 & 28,8 \end{pmatrix}$$

$$\left(\mathbf{X}^T \mathbf{X}\right)^{-1} = \frac{1}{\left|\mathbf{X}^T \mathbf{X}\right|} \mathbf{X}^T \mathbf{X} = \frac{1}{57,6} \begin{pmatrix} 649,44 & -100,8 & -14,4 \\ -100,8 & 16 & 0 \\ -14,4 & 0 & 28,8 \end{pmatrix}$$

$$\hat{\beta} = \left(\mathbf{X}^T \mathbf{X}\right)^{-1} . \mathbf{X}^T \mathbf{Y} = \frac{1}{57,6} \begin{pmatrix} 649,44 & -100,8 & -14,4 \\ -100,8 & 16 & 0 \\ -14,4 & 0 & 28,8 \end{pmatrix} \begin{pmatrix} 7,28 \\ 92,016 \\ 7,64 \end{pmatrix} = \begin{pmatrix} 1,226 \\ 0,08 \\ 0,18 \end{pmatrix}$$

$$\hat{\mathbf{Y}}_i = 1,226 + 0,08 \mathbf{X}_i + 0,18 \mathbf{Z}_i$$

Ý nghĩa của các hệ số hồi quy

$\hat{\beta}_2 = 0.08$: Nếu cùng giới tính, khi thu nhập tăng thêm 1 triệu đồng/ tháng thì chi tiêu trung bình cho may mặc sẽ tăng thêm 0.08 triệu đồng/ năm.

$\hat{\beta}_3 = 0.18$: Nếu cùng thu nhập, thì nữ công nhân chi tiêu cho may mặc trung bình nhiều hơn nam là 0.18 triệu đồng/ năm.

Chương 4

§4.2 Ứng dụng của MHHQ với biến giả

4.2.1 So sánh 2 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.

Chương 4

§4.2 Ứng dụng của MHHQ với biến giả

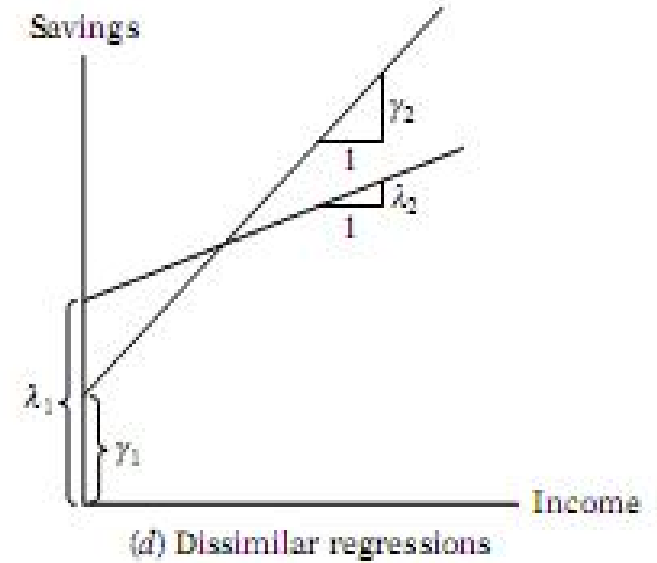
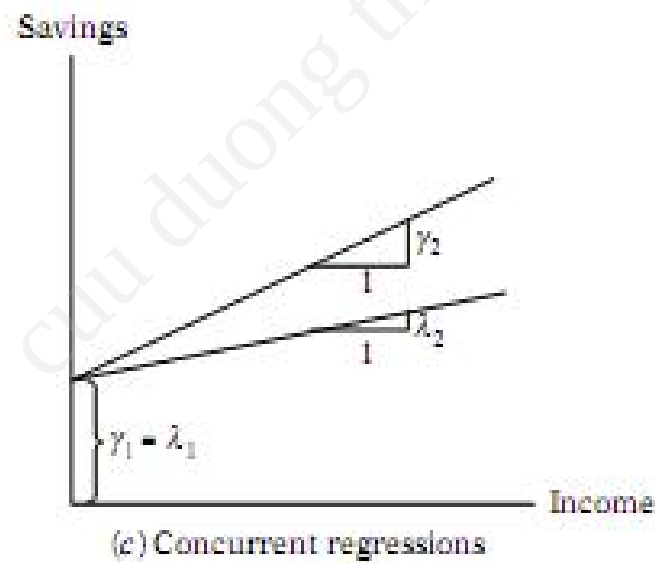
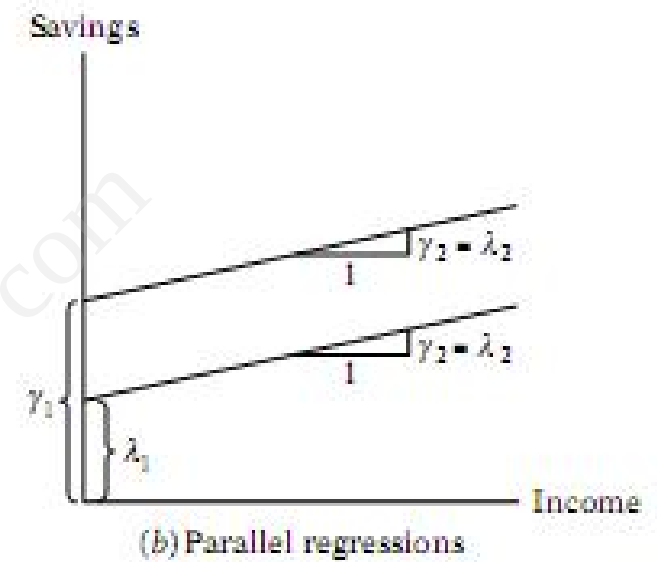
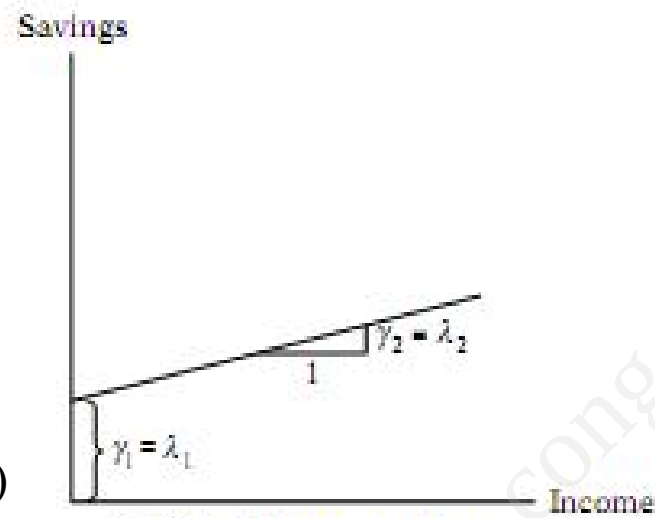
Bài toán đặt ra ở đây: có thể xây dựng 1 mô hình hồi quy duy nhất dựa trên cả 2 tập số liệu đó hay không hay phải xây dựng 2 mô hình hồi quy khác nhau, mỗi hồi quy dựa trên 1 tập số liệu?

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

(*)

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

(**)



Chương 4

§4.2 Ứng dụng của MHHQ với biến giả

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 Z_i + \beta_4 (X_i Z_i) + U_i \quad (4.5)$$

Y_i : biến phụ thuộc

X_i : biến giải thích là biến số lượng

U_i : sai số ngẫu nhiên

Z_i : biến giả được quy ước

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

Chương 4

§4.2 Ứng dụng của MHHQ với biến giả

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

$$E(Y_i / Z_i = 0, X_i) = \beta_1 + \beta_2 X_i$$

$$E(Y_i / Z_i = 1, X_i) = (\beta_1 + \beta_3) + (\beta_2 + \beta_4) X_i$$

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

Chương 4

§4.2 Ứng dụng của MHHQ với biến giả

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 Z_i + \beta_4 (X_i Z_i) + U_i \quad (4.5)$$

$$H_0: \beta_3 = \beta_4 = 0$$

Nếu ở mức ý nghĩa α nào đó ta không bác bỏ được H_0 thì điều này có nghĩa (ở mức ý nghĩa α đó) không có sự chênh lệch cả về hệ số chặn và hệ số góc khi chuyển từ mẫu I sang mẫu II, tức là có thể gộp 2 tập số liệu được.

Chương 4

§4.2 Ứng dụng của MHHQ với biến giả

4.2.2 Phân tích thời vụ (mùa)

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 ...

Chương 4

§4.2 Ứng dụng của MHHQ với biến giả

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

Chương 4

§4.2 Ứng dụng của MHHQ với biến giả

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 Z_{1i} + \beta_4 Z_{2i} + \beta_5 Z_{3i} + U_i \quad (4.7)$$

Y_i : nhu cầu tiêu dùng
hàng may mặc

X_i : thu nhập

Z_{1i}, Z_{2i}, Z_{3i} : biến giả được quy ước

	Z_1 i	Z_2 i	Z_3 i
I	0	0	0
II	1	0	0
III	0	1	0
IV	0	0	1

Chương 4

§ 4.2 Ứng dụng của MHHQ với biến giả

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

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

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

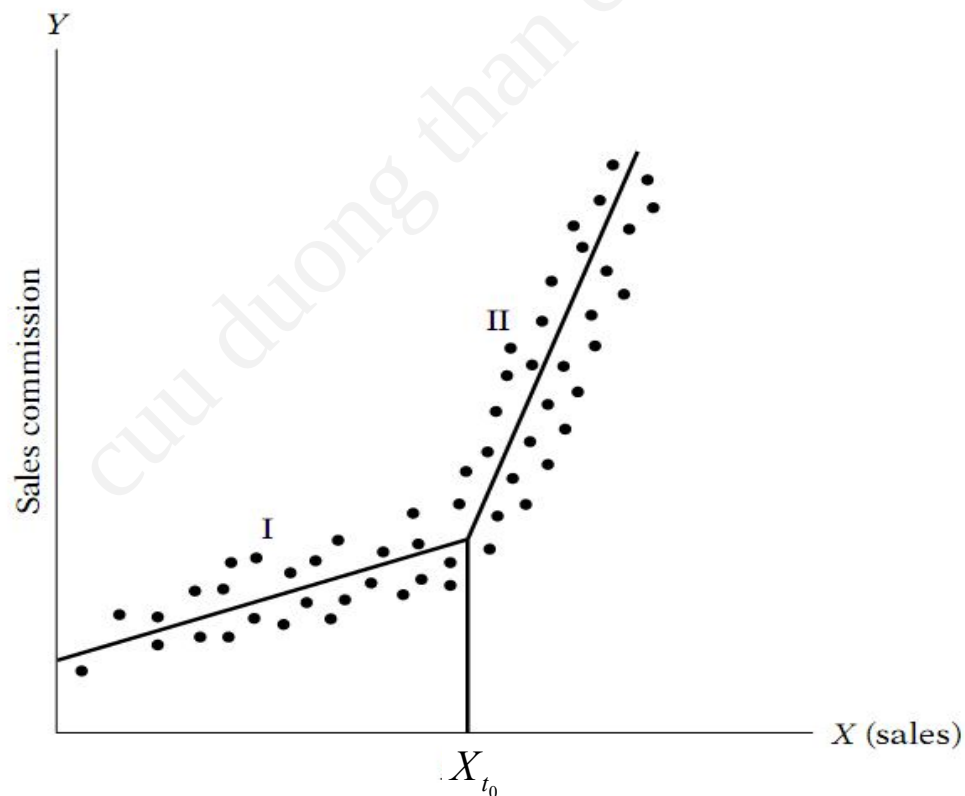
$$E(Y / Z_{2i} = 1, Z_{1i} = Z_{3i} = 0; X_i) = (\beta_1 + \beta_4) + \beta_2 X_i$$

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

Chương 4

§4.2 Ứng dụng của MHHQ với biến giả

4.2.3 Hồi quy tuyến tính từng đoạn



Chương 4

§ 4.2 Ứng dụng của MHHQ với biến giả

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

Y_t : biến phụ thuộc (năng suất)

U_t : sai số ngẫu nhiên

X_{t_0} : giá trị biến X tại thời điểm $t = t_0$

Z : biến giả được quy ước

$$Z_t = \begin{cases} 0 & t \leq t_0 \\ 1 & t > t_0 \end{cases}$$

Chương 4

§ 4.2 Ứng dụng của MHHQ với biến giả

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

$$E(Y / Z_t = 0, X_t) = \beta_1 + \beta_2 X_t$$

$$E(Y / Z_t = 1, X_t) = (\beta_1 - \beta_3 X_{t_0}) + (\beta_2 + \beta_3) X_t$$

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

Chương 4

§ 4.2 Ứng dụng của MHHQ với biến giả

$$E(Y / Z_t = 0, X_t) = \beta_1 + \beta_2 X_t$$

$$E(Y / Z_t = 1, X_t) = (\beta_1 - \beta_3 X_{t_0}) + (\beta_2 + \beta_3) X_t$$

$$H_0: \beta_3 = 0$$

Nếu ở mức ý nghĩa α nào đó ta không bác bỏ được H_0 thì điều này có nghĩa (ở mức ý nghĩa α đó) tốc độ biến thiên của Y phụ thuộc X không có gì khác trước và sau thời điểm chuyển đổi

Chương 4

§ 4.2 Ứng dụng của MHHQ với biến giả

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:

Chương 4

§ 4.2 Ứng dụng của MHHQ với biến giả

$$Y_t = \beta_1 + \beta_2 X_t + \beta_3 (X_t - X_{t_0}) Z_{1t} + \beta_4 (X_t - X_{t_1}) Z_{2t} + U_t \quad (4.8)$$

$$Z_{1t} = \begin{cases} 0 & t \leq t_0 \\ 1 & t > t_0 \end{cases}$$

$$Z_{2t} = \begin{cases} 0 & t \leq t_1 \\ 1 & t > t_1 \end{cases}$$

Chương 5

PHƯƠNG SAI CỦA SAI SỐ THAY ĐỔI

Chương 5

PHƯƠNG SAI CỦA SAI SỐ THAY ĐỔI

5.1 Phương sai của sai số thay đổi –
Nguyên nhân và hậu quả

5.2 Phát hiện sự tồn tại phương sai của
sai số thay đổi

5.3 Khắc phục hiện tượng phương sai của
sai số thay đổi (nội dung thảo luận)

Chương 5

§5.1 Phương sai của sai số thay đổi – Nguyên nhân và hậu quả

5.1.1 Hiện tượng phương sai của sai số thay đổi và nguyên nhân

Xảy ra khi giả thiết $\text{Var}(U_i) = \sigma^2$ ($\forall i$) bị vi phạm, tức là

$$\text{Var}(U_i) = \sigma_i^2$$

Chương 5

§5.1 Phương sai của sai số thay đổi – Nguyên nhân và hậu quả

- Do bản chất của các mối liên hệ giữa các đại lượng kinh tế
- Do kỹ thuật thu thập và xử lý số liệu, dạng hàm sai

Chương 5

§5.1 Phương sai của sai số thay đổi – Nguyên nhân và hậu quả

5.1.2 Hậu quả của hiện tượng phương sai của sai số thay đổi

Các ước lượng BPNN $\hat{\beta}_j$ vẫn là các ước lượng tuyến tính, không chệch nhưng không còn là hiệu quả

Chương 5

§5.1 Phương sai của sai số thay đổi – Nguyên nhân và hậu quả

Các ước lượng của các phương sai sẽ là các ước lượng chệch, thống kê T và F không còn có ý nghĩa. Do đó khoảng tin cậy và các kiểm định dựa trên thống kê T và F không còn đáng tin cậy nữa

Chương 5

§5.2 Phát hiện phương sai của sai số thay đổi

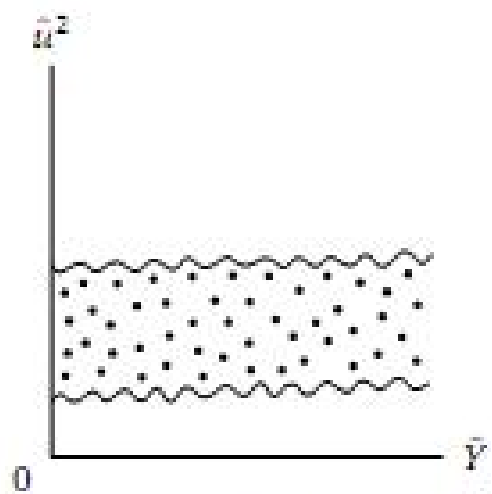
5.2.1 Phương pháp đồ thị

Vì phần dư e_i của hàm hồi quy mẫu chính là ước lượng của sai số ngẫu nhiên U_i nên dựa vào đồ thị phần dư (hoặc bình phương phần dư) đối với biến giải thích X ta có kết luận:

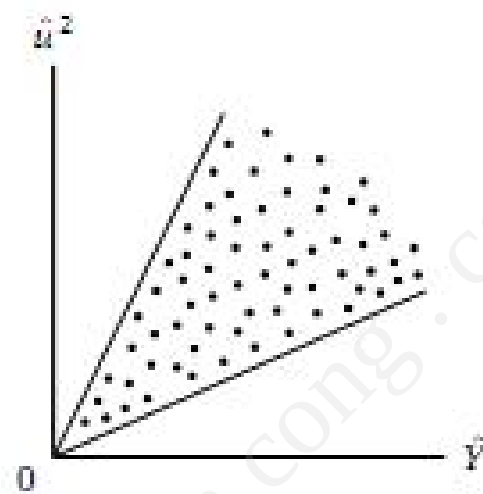
Chương 5

§5.2 Phát hiện phương sai của sai số thay đổi

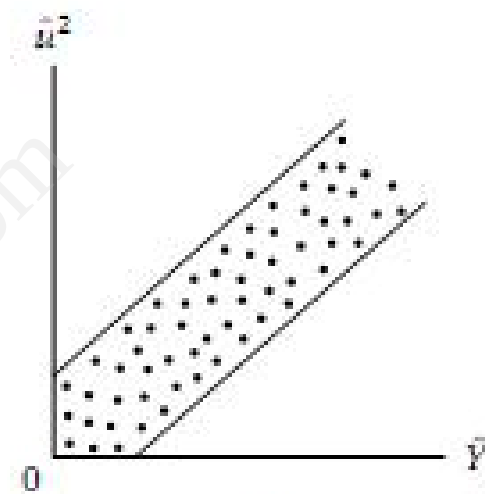
Nếu độ rộng của phần dư e (hay e^2) tăng hay giảm khi X tăng thì có thể nghi ngờ phương sai của sai số thay đổi. Trong trường hợp nhiều hơn 1 biến giải thích, có thể dùng đồ thị e (hoặc e^2) đối với $\hat{Y}_i$



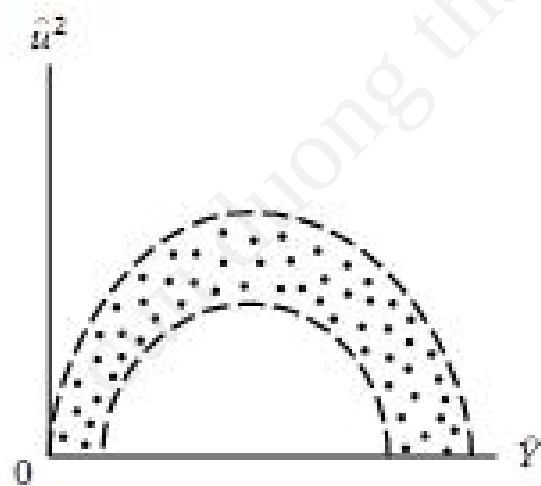
(a)



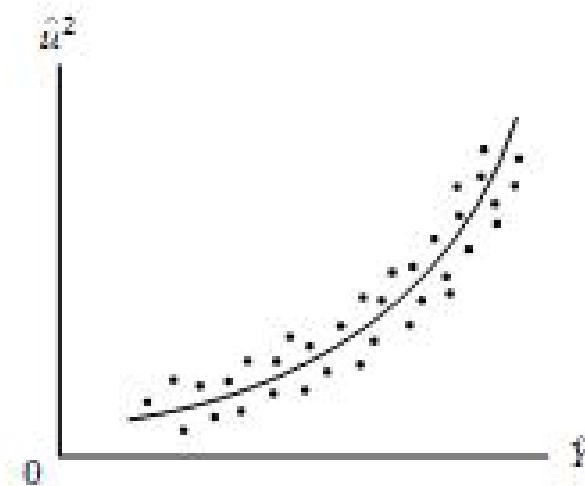
(b)



(c)



(d)



(e)

Chương 5

§5.2 Phát hiện phương sai của sai số thay đổi

5.2.2 Kiểm định Goldfield – Quant (G - Q)

Bước 1. Sắp xếp các giá trị quan sát theo chiều tăng của biến X_j

Bước 2. Bỏ c quan sát ở giữa theo quy tắc:

Nếu $n = 30$: lấy $c = 4$ hoặc 6 .

Nếu $n = 60$: lấy $c = 10$

Các quan sát còn lại chia 2 nhóm, mỗi nhóm có $(n-c)/2$ quan sát

Chương 5

§5.2 Phát hiện phương sai của sai số thay đổi

Bước 3. Ước lượng mô hình với $(n-c)/2$ quan sát đầu và cuối thu được RSS_1 và RSS_2 tương ứng với bậc tự do là:

$$d = \frac{n-c}{2} - k = \frac{n-c-2k}{2}$$

Chương 5

§5.2 Phát hiện phương sai của sai số thay đổi

Bước 4. Xây dựng TCKĐ:

$$F = \frac{RSS_1}{d} \bigg/ \frac{RSS_2}{d}$$

Nếu giả thiết H_0 : phương sai của sai số ngẫu nhiên không đổi được thỏa mãn thì $F \sim F(d, d)$

$$W_\alpha = \{f_{tn}, f_{tn} > f_\alpha(d, d)\}$$

Chương 5

§5.2 Phát hiện phương sai của sai số thay đổi

5.2.3 Kiểm định Park

Park đưa ra giả thiết

$$\sigma_i^2 = \sigma^2 X_i^{\beta_2} e^{v_i}$$

$$\Leftrightarrow \ln \sigma_i^2 = \ln \sigma^2 + \beta_2 \ln X_i + v_i$$

Vì thường σ_i^2 chưa biết nên thay thế bởi ước lượng của nó là e_i^2

$$\ln e_i^2 = \ln \sigma^2 + \beta_2 \ln X_i + v_i$$

Chương 5

§5.2 Phát hiện phương sai của sai số thay đổi

Bước 1. Ước lượng hồi quy gốc để thu được các phần dư e_i

Bước 2. Ước lượng hồi quy

$$\ln e_i^2 = \beta_1 + \beta_2 \ln X_i + v_i$$

Nếu có nhiều biến giải thích thì ước lượng hồi quy này với từng biến giải thích hoặc với $\hat{Y}_i$

Chương 5

§5.2 Phát hiện phương sai của sai số thay đổi

$$\ln e_i^2 = \beta_1 + \beta_2 \ln X_i + v_i$$

Bước 3. Kiểm định gt $H_0: \beta_2 = 0$

Nếu H_0 bị bác bỏ thì kết luận có phương sai của sai số thay đổi

Chương 5

§5.2 Phát hiện phương sai của sai số thay đổi

5.2.4 Kiểm định Glejser

$$|e_i| = \beta_1 + \beta_2 X_i + v_i \qquad |e_i| = \beta_1 + \beta_2 \sqrt{X_i} + v_i$$

$$|e_i| = \beta_1 + \beta_2 \frac{1}{X_i} + v_i \qquad |e_i| = \beta_1 + \beta_2 \frac{1}{\sqrt{X_i}} + v_i$$

Nếu giả thuyết $H_0 : \beta_2 = 0$ bị bác bỏ thì kết luận có phương sai của sai số thay đổi

Chương 6

TỰ TƯƠNG QUAN

Chương 6

TỰ TƯƠNG QUAN

6.1 Hiện tượng tự tương quan – Nguyên nhân và hậu quả

6.2 Phát hiện sự tồn tại tự tương quan

6.3 Khắc phục hiện tượng tự tương quan (nội dung thảo luận)

Chương 6

§6.1 Hiện tượng tự tương quan – Nguyên nhân và hậu quả

6.1.1 Hiện tượng TTQ và nguyên nhân

Hiện tượng tự tương quan xảy ra khi

$$\text{cov}(U_i, U_j) = E(U_i U_j) \neq 0 \quad \forall i \neq j$$

Chương 6

§6.1 Hiện tượng tự tương quan— Nguyên nhân và hậu quả

$$U_t = \rho U_{t-1} + \varepsilon_t$$

ρ : hệ số tự tương quan bậc 1 (hay hệ số tự hồi quy bậc 1)

ε_t : nhiễu ngẫu nhiên thoả mãn mọi giả thiết của MHHQT cổ điển

Chương 6

§6.1 Hiện tượng tự tương quan— Nguyên nhân và hậu quả

$$U_t = \rho U_{t-1} + \varepsilon_t$$

U_t tuân theo lược đồ tự hồi quy bậc 1, ký hiệu $AR(1)$

Chương 6

§6.1 Hiện tượng tự tương quan— Nguyên nhân và hậu quả

$$U_t = \rho_1 U_{t-1} + \rho_2 U_{t-2} + \dots + \rho_p U_{t-p} + \varepsilon_t$$

ρ_j : hệ số tự hồi quy bậc j ($j = \overline{1, p}$)

ε_t : nhiễu ngẫu nhiên thoả mãn mọi giả thiết của MHHQTT cổ điển

Chương 6

§6.1 Hiện tượng tự tương quan— Nguyên nhân và hậu quả

$$U_t = \rho_1 U_{t-1} + \rho_2 U_{t-2} + \dots + \rho_p U_{t-p} + \varepsilon_t$$

U_t tuân theo lược đồ tự hồi quy bậc p , $AR(p)$

Chương 6

§6.1 Hiện tượng tự tương quan – Nguyên nhân và hậu quả

Quán tính – tính chất phổ biến của các đại lượng kinh tế quan sát theo thời gian

Hiện tượng mạng nhện

Tính chất “trễ” của các đại lượng kinh tế

Chương 6

§6.1 Hiện tượng tự tương quan – Nguyên nhân và hậu quả

Phương pháp (kỹ thuật) thu thập và xử lý số liệu

Sai lầm khi lập mô hình: bỏ biến (không đưa biến vào mô hình), dạng hàm sai...

Chương 6

§6.1 Hiện tượng tự tương quan – Nguyên nhân và hậu quả

6.1.2 Hậu quả của hiện tượng tự tương quan

Các ước lượng BPNN $\hat{\beta}_j$ là các ước lượng tuyến tính, không chệch nhưng không phải là hiệu quả nữa

Chương 6

§6.1 Hiện tượng tự tương quan – Nguyên nhân và hậu quả

Các ước lượng của các phương sai là chệch và thông thường là thấp hơn giá trị thực của phương sai, do đó giá trị của thống kê T được phóng đại lên nhiều lần so với giá trị thực của nó

Chương 6

§6.1 Hiện tượng tự tương quan – Nguyên nhân và hậu quả

Thống kê T và F không còn có ý nghĩa về mặt thống kê nên việc kiểm định các giả thiết thống kê không còn đáng tin cậy nữa

Các dự báo dựa trên các ước lượng BPNN không còn tin cậy nữa

Chương 6

§6.2 Phát hiện tự tương quan

6.2.1 Kiểm định d (Durbin – Watson)

Thống kê d được định nghĩa:

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}$$

Chương 6

§6.2 Phát hiện tự tương quan

$$d \approx 2(1 - \hat{\rho})$$

Trong đó

$$\hat{\rho} = \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=1}^n e_t^2}$$

Chương 6

§6.2 Phát hiện tự tương quan

$$d \approx 2(1 - \hat{\rho})$$

Vì $-1 \leq \rho \leq 1$ nên $0 \leq d \leq 4$

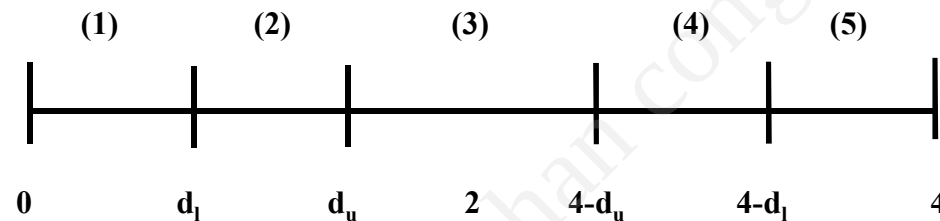
Nếu $\rho = -1$ thì $d = 4$: TTQ ngược chiều.

Nếu $\rho = 0$ thì $d = 2$: không có TTQ.

Nếu $\rho = 1$ thì $d = 0$: tồn tại TTQ thuận chiều

Chương 6

§6.2 Phát hiện tự tương quan



- $d \in (1)$: tồn tại tự tương quan thuận chiều.
 $d \in (2)$: không xác định.
 $d \in (3)$: không có tự tương quan.
 $d \in (4)$: không xác định.
 $d \in (5)$: tồn tại tự tương quan ngược chiều

Chương 6

§6.2 Phát hiện tự tương quan

Chú ý: Các giá trị d_L , d_U được tính sẵn phụ thuộc mức ý nghĩa α , kích thước mẫu n và số biến giải thích k' có trong mô hình ($k' = k - 1$).

Chương 6

§6.2 Phát hiện tự tương quan

6.2.2 Kiểm định BG (Breush – Godfrey)

$$Y_t = \beta_1 + \beta_2 X_t + U_t$$

Giả sử rằng:

$$U_t = \rho_1 U_{t-1} + \rho_2 U_{t-2} + \dots + \rho_p U_{t-p} + \varepsilon_t$$

$$H_0 : \rho_1 = \rho_2 = \dots = \rho_p = 0$$

Chương 6

§6.2 Phát hiện tự tương quan

Bước 1: Ước lượng mô hình ban đầu bằng phương pháp BPNN thông thường để nhận được các phần dư e_t

Bước 2: Cũng bằng phương pháp BPNN, ước lượng mô hình sau để thu được hệ số xác định bội R^2

$$e_t = \beta_1 + \beta_2 X_t + \rho_1 e_{t-1} + \rho_2 e_{t-2} + \dots + \rho_p e_{t-p} + v_t$$

Chương 6

§6.2 Phát hiện tự tương quan

Bước 3: $H_0: \rho_1 = \rho_2 = \dots = \rho_p = 0$

Nếu H_0 đúng thì

$$\chi^2 = (n - p)R^2 \sim \chi^2(p)$$

$$W_\alpha = \{\chi_{tn}^2; \chi_{tn}^2 > \chi_\alpha^2(p)\}$$

Chương 7

ĐA CỘNG TUYẾN

Chương 7 ĐA CỘNG TUYẾN

7.1 Đa cộng tuyến và hậu quả của đa cộng tuyến (nội dung thảo luận)

7.2 Phát hiện đa cộng tuyến và biện pháp khắc phục (nội dung thảo luận)

Chương 7

§7.1 Đa cộng tuyến và hậu quả

7.1.1 Bản chất của đa cộng tuyến

Xét mô hình hồi quy tuyến tính cổ điển nhiều biến

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + U_i$$

Chương 7

§7.1 Đa cộng tuyến và hậu quả

Hiện tượng đa cộng tuyến toàn phần xảy ra giữa các biến giải thích $X_2, X_3, \dots, X_k$ nếu tồn tại $\lambda_2, \lambda_3, \dots, \lambda_k$ không đồng thời bằng 0 sao cho

$$\lambda_2 X_{2i} + \lambda_3 X_{3i} + \dots + \lambda_k X_{ki} = 0 \quad \forall i$$

Chương 7

§7.1 Đa cộng tuyến và hậu quả

Hiện tượng đa cộng tuyến không toàn phần (đa cộng tuyến) xảy ra giữa các biến giải thích $X_2, X_3, \dots, X_k$ nếu tồn tại $\lambda_2, \lambda_3, \dots, \lambda_k$ không đồng thời bằng 0 sao cho

$$\lambda_2 X_{2i} + \lambda_3 X_{3i} + \dots + \lambda_k X_{ki} + v_i = 0 \quad \forall i$$

trong đó v_i là nhiễu ngẫu nhiên

Chương 7

§7.1 Đa cộng tuyến và hậu quả

Trong thực tế thường xảy ra đa cộng tuyến không toàn phần, hiếm khi xảy ra đa cộng tuyến toàn phần

Chương 7

§7.1 Đa cộng tuyến và hậu quả

7.1.2 Hậu quả của đa cộng tuyến

Trường hợp **đa cộng tuyến toàn phần**: các hệ số hồi quy mẫu là không xác định và các độ lệch tiêu chuẩn là vô hạn

Chương 7

§7.1 Đa cộng tuyến và hậu quả

7.1.2 Hậu quả của đa cộng tuyến

Trường hợp **đa cộng tuyến không toàn phần**:
Trong trường hợp này có thể xác định được các hệ số hồi quy mẫu nhưng dẫn đến các hậu quả sau

Chương 7

§7.1 Đa cộng tuyến và hậu quả

1. Phương sai và độ lệch tiêu chuẩn của các hệ số hồi quy mẫu sẽ rất lớn

Chẳng hạn

$$Var(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_{2i}^2 (1 - r_{23}^2)}$$

2. Khoảng tin cậy của các hệ số hồi quy rộng hơn

Chương 7

§7.1 Đa cộng tuyến và hậu quả

3. Tỷ số T mất ý nghĩa

4. Hệ số xác định bội R^2 cao nhưng t nhỏ

5. Dấu các ước lượng của các hệ số hồi quy sai do đó các ước lượng BPNN trở nên rất nhạy với những thay đổi nhỏ trong số liệu

Chương 7

§7.2 Phát hiện sự tồn tại đa cộng tuyến và biện pháp khắc phục

7.2.1 Phát hiện sự tồn tại đa cộng tuyến

1. Hệ số xác định bội R^2 cao nhưng tỷ số T thấp
2. Hệ số tương quan cặp giữa các biến giải thích cao
3. Xét hồi quy phụ
4. Sử dụng nhân tử phóng đại phương sai VIF

Chương 7

§7.2 Phát hiện sự tồn tại đa cộng tuyến và biện pháp khắc phục

7.2.2 Biện pháp khắc phục đa cộng tuyến

1. Bỏ biến giải thích có khả năng là tổ hợp tuyến tính của các biến còn lại
2. Thu thập số liệu và lấy mẫu mới
3. Kiểm tra lại mô hình
4. Đổi biến số

Chương 8

CHỌN MÔ HÌNH VÀ KIỂM ĐỊNH VIỆC CHỌN MÔ HÌNH

Chương 8

CHỌN MÔ HÌNH VÀ KIỂM ĐỊNH VIỆC CHỌN MÔ HÌNH

8.1 Các thuộc tính của 1 mô hình tốt

8.2 Các loại sai lầm thường mắc

8.3 Phát hiện và kiểm định các sai lầm chỉ định

8.4 Một số mô hình kinh tế thông dụng
(bài tập áp dụng)

Chương 8

§8.1 Các thuộc tính của mô hình tốt

- Tính Kiệm
- Đồng nhất
- Phù hợp
- Bền vững về mặt lý thuyết
- Có khả năng dự báo tốt

Chương 8

§8.2 Các loại sai lầm khi chọn mô hình

- Bỏ sót biến thích hợp
- Đưa vào mô hình biến không thích hợp
- Chọn dạng hàm không đúng

Chương 8

§8.2 Các loại sai lầm khi chọn mô hình

8.2.1 Bỏ sót biến giải thích

Giả sử mô hình đúng:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + U_t$$

Nhưng ta chọn mô hình:

$$Y_t = \alpha_1 + \alpha_2 X_{2t} + V_t$$

Chương 8

§8.2 Các loại sai lầm khi chọn mô hình

$$\hat{Y}_t = \hat{\alpha}_1 + \hat{\alpha}_2 X_{2t}$$

Nếu X_2 tương quan X_3 thì $\hat{\alpha}_1$, $\hat{\alpha}_2$ không phải là UL vững và là ước lượng chệch và của β_1 , β_2

Nếu X_2 không tương quan X_3 thì $\hat{\alpha}_2$ là UL vững và là ước lượng không chệch và của β_2 , nhưng $\hat{\alpha}_1$ vẫn là UL chệch của β_1

Chương 8

§8.2 Các loại sai lầm khi chọn mô hình

Phương sai của sai số ước lượng từ mô hình đúng và phương sai của sai số ước lượng của mô hình chỉ định sai sẽ không như nhau.

Khoảng tin cậy thông thường và các thủ tục kiểm định giả thiết không còn đáng tin cậy nữa.

Chương 8

§8.2 Các loại sai lầm khi chọn mô hình

8.2.2 Đưa biến không thích hợp vào mô hình

Giả sử mô hình đúng:

$$Y_t = \beta_1 + \beta_2 X_{2t} + U_t$$

Nhưng ta chọn mô hình:

$$Y_t = \alpha_1 + \alpha_2 X_{2t} + \alpha_3 X_{3t} + V_t$$

Chương 8

§8.2 Các loại sai lầm khi chọn mô hình

Hàm hồi quy mẫu của mô hình “sai”:

$$\hat{Y}_t = \hat{\alpha}_1 + \hat{\alpha}_2 X_{2t} + \hat{\alpha}_3 X_{3t}$$

Ước lượng của σ^2 là ước lượng vững

Các ước lượng BPNN $\hat{\alpha}_j$ là ước lượng không chệch và vững nhưng không hiệu quả dẫn đến khoảng tin cậy sẽ rộng hơn

Chương 8

§8.2 Các loại sai lầm khi chọn mô hình

8.2.3 Chọn dạng hàm không đúng

Các kết quả thu được từ việc phân tích hồi quy trong mô hình “sai” sẽ không đúng với thực tế và dẫn đến các kết luận sai lầm.

Chương 8

§8.3 Phát hiện và KĐ các sai lầm chỉ định

8.3.1 Phát hiện biến không cần thiết trong MH

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

$$H_0 : \beta_5 = 0$$

$$H_0 : \beta_4 = \beta_5 = 0$$

Chương 8

§8.3 Phát hiện và KĐ các sai lầm chỉ định

8.3.2 Kiểm định các biến bị bỏ sót

$$Y_t = \beta_1 + \beta_2 X_{2t} + U_t$$

Nếu đã có số liệu của Z ta chỉ cần UL mô hình

$$Y_t = \alpha_1 + \alpha_2 X_t + \alpha_3 Z_t + V_t$$

$$H_0: \alpha_3 = 0$$

Chương 8

§8.3 Phát hiện và KĐ các sai lầm chỉ định

Nếu không có số liệu của Z ta có thể sử dụng một trong các kiểm định sau

Chương 8

§8.3 Phát hiện và KĐ các sai lầm chỉ định

a. Kiểm định RESET của RAMSEY

Bước 1. Hồi quy Y_t theo X_t ta có $\hat{Y}_t$ và R^2_{old}

Bước 2. Hồi quy Y_t theo X_t , $\hat{Y}_t^2$, $\hat{Y}_t^3$ được R^2_{new}

và kiểm định các hệ số của $\hat{Y}_t^2$, $\hat{Y}_t^3$ bằng 0

Chương 8

§8.3 Phát hiện và KĐ các sai lầm chỉ định

Bước 3. Kiểm định có điều kiện ràng buộc:

$$F = \frac{(R^2_{new} - R^2_{old}) / m}{(1 - R^2_{new}) / (n - k)}$$

m : số biến mới được đưa vào MH

k : số hệ số của mô hình mới

Khi n lớn ta có $F \sim F(m, n-k)$

Chương 8

§8.3 Phát hiện và KĐ các sai lầm chỉ định

b. Kiểm định d (Durbin-Watson)

Bước 1. Ước lượng mô hình :

$$Y_i = \beta_1 + \beta_2 X_{2i} + U_i$$

Bước 2. Sắp xếp e_i theo thứ tự tăng dần của biến bỏ sót Z , nếu Z chưa có số liệu thì sắp xếp e_i theo X

Chương 8

§8.3 Phát hiện và KĐ các sai lầm chỉ định

Bước 3.

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}$$

Bước 4. H_0 : Dạng hàm đúng (không có TTQ)

Chương 8

§8.3 Phát hiện và KĐ các sai lầm chỉ định

c. Phương pháp nhân tử Lagrange (LM)

Bước 1. Hồi quy mô hình gốc thu được $\hat{Y}_t$ và e_t

Bước 2. Ước lượng MH sau để thu được R^2 :

$$e_t = \beta_1 + \beta_2 X_t + \alpha_2 \cdot \hat{Y}_t^2 + \dots + \alpha_p \cdot \hat{Y}_t^p + V_t$$

Chương 8

§8.3 Phát hiện và KĐ các sai lầm chỉ định

Với n khá lớn $\chi^2 = nR^2$ có phân phối $\chi^2(p)$ từ đó ta kết luận cho bài toán.

Chương 8

§8.3 Phát hiện và KĐ các sai lầm chỉ định

8.3.3 Kiểm định tính PP chuẩn của sai số NN

Thường dùng kiểm định Jarque-Berra (JB)

$$JB = n \left[\frac{S^2}{6} + \frac{(K - 3)^2}{24} \right]$$

S: skewness

K: kurtosis

Chương 8

§8.3 Phát hiện và KĐ các sai lầm chỉ định

$$JB = n \left[\frac{S^2}{6} + \frac{(K-3)^2}{24} \right]$$

Nếu H_0 đúng thì $JB \sim \chi^2(2)$, từ đó ta có kết luận cho bài toán

Chương 8

§8.4 Một số MH kinh tế lượng thông dụng

Hàm sản xuất Cobb-Douglas

Hàm tăng trưởng kinh tế

Mô hình Hyperbol

Mô hình hồi quy Đa thức