

Chương 2:
MÔ HÌNH HỒI QUY BỘI

Th.S NGUYỄN PHƯƠNG

Bộ môn Toán kinh tế
Trường Đại học Ngân hàng TP HCM
Blog: <https://nguyenphuongblog.wordpress.com>
Email: nguyenphuong0122@gmail.com

Ngày 18 tháng 9 năm 2016

NỘI DUNG

- 1 Sự cần thiết của mô hình hồi quy bội
- 2 Mô hình hồi quy bội và Phương pháp ước lượng OLS
 - Mô hình và phương pháp OLS
 - Mô hình hồi quy sử dụng ngôn ngữ ma trận
 - Các giả thiết
 - Độ phù hợp của hàm hồi quy
 - Tính tốt nhất của ước lượng OLS
- 3 Tính vững của ước lượng OLS
- 4 Khoảng tin cậy cho hệ số hồi quy
 - Quy luật phân phối xác suất của một số thống kê mẫu
 - Khoảng tin cậy cho một hệ số hồi quy
 - Khoảng tin cậy cho biểu thức của hai hệ số hồi quy
 - Khoảng tin cậy cho phương sai sai số ngẫu nhiên
- 5 Kiểm định giả thuyết về hệ số hồi quy
 - Kiểm định giả thuyết về một hệ số hồi quy
 - Kiểm định về một ràng buộc giữa các hệ số hồi quy
 - Kiểm định về nhiều ràng buộc -kiểm định F
 - Kiểm định sự phù hợp của hàm hồi quy

- Một biến phụ thuộc Y thường chịu tác động của nhiều yếu tố.
- Mô hình hồi quy bội thường có chất lượng dự báo tốt hơn.
- Mô hình hồi quy bội cho phép sử dụng dạng hàm phong phú hơn.
- Mô hình hồi quy bội thực hiện các phân tích phong phú hơn.

Ví dụ: Ngoài thu nhập, thì có nhiều yếu tố khác cũng tác động lên tiêu dùng, chẳng hạn như độ tuổi, giới tính, nghề nghiệp, địa bàn sinh sống, vật giá, thói quen chi tiêu, ...

Hàm hồi quy tổng thể – PRF: $E(Y|X) = \beta_1 + \beta_2 X_2 + \dots + \beta_k X_k.$

Mô hình hồi quy tổng thể – PRM: $Y_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + u_i, i = \overline{1, n}$

hoặc: $Y = \beta_1 + \beta_2 X_2 + \dots + \beta_k X_k + u.$

β_1 : hệ số chặn/hệ số tự do (intercept).

β_j : hệ số góc hay hệ số hồi quy riêng, $j = \overline{2, k}$.

u : sai số ngẫu nhiên.

Câu hỏi: Ý nghĩa của các hệ số $\beta_1, \beta_2, \dots, \beta_k$.

$$\Delta E(Y|X) = \beta_2 \Delta X_2 + \dots + \beta_k \Delta X_k.$$

Ví dụ 2.1

Mô hình hồi quy tổng thể về lạm phát:

$$LP = 0,02 + 0,3m - 0,15gdp + u$$

trong đó LP, m và gdp lần lượt là tỷ lệ lạm phát, mức tăng trưởng cung tiền và mức tăng trưởng GDP (đơn vị %). Hãy giải thích ý nghĩa của các hệ số.

Hàm hồi quy mẫu – SRF: $\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X_2 + \cdots + \hat{\beta}_k X_k.$

Mô hình hồi quy mẫu – SRM: $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_k X_{ki} + e_i, \quad i = \overline{1, n};$

hoặc: $Y = \hat{\beta}_1 + \hat{\beta}_2 X_2 + \cdots + \hat{\beta}_k X_k + e.$

trong đó \hat{Y} là ước lượng cho $E(Y|X)$; $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$ tương ứng là ước lượng cho $\beta_1, \beta_2, \dots, \beta_k$; e_i là phần dư, ước lượng cho u_i .

$$\Delta \hat{Y} = \hat{\beta}_2 \Delta X_2 + \cdots + \hat{\beta}_k \Delta X_k.$$

Ví dụ 2.2

Ước lượng hàm hồi quy tuyến tính của doanh số bán hàng (Y , đv: triệu đồng) theo chi phí chào hàng (X_2 , triệu đồng) và chi phí quảng cáo (X_3 , triệu đồng), ta được: $\hat{Y} = 328,1383 + 4,6495X_2 + 2,5602X_3$. Nêu ý nghĩa của các hệ số hồi quy.

Định nghĩa: Phương pháp OLS nhằm xác định các giá trị $\hat{\beta}_j, j = 1, 2, \dots, k$ sao cho tổng bình phương các phần dư là nhỏ nhất. (Tương tự như mô hình 2 biến)

Xét mô hình k biến: $Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i, \quad i = 1, 2, \dots, n.$

Đặt

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, X = \begin{bmatrix} 1 & X_{21} & X_{31} & \cdots & X_{k1} \\ 1 & X_{22} & X_{32} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{2n} & X_{3n} & \cdots & X_{kn} \end{bmatrix}, \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix}, u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}.$$

Khi đó mô hình hồi quy tổng thể dưới dạng ma trận như sau:

$$Y = X\beta + u.$$

Từ mẫu quan sát ta có ước lượng cho Y và β như sau:

$$\hat{Y} = \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{bmatrix}, \hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_n \end{bmatrix}.$$

Ta có hàm hồi quy mẫu

$$\hat{Y} = X\hat{\beta}.$$

Véc tơ phần dư $e = Y - \hat{Y} = Y - X\hat{\beta}.$

Phương pháp OLS đi tìm $\hat{\beta}$ sao cho $e^T e \rightarrow \min$. Áp dụng phương pháp này tìm được kết quả:

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{pmatrix} = (X^T X)^{-1} X^T Y$$

$$\text{cov}(\hat{\beta}) = \begin{pmatrix} \text{var}(\hat{\beta}_1) & \text{cov}(\hat{\beta}_1, \hat{\beta}_2) & \dots & \text{cov}(\hat{\beta}_1, \hat{\beta}_k) \\ \text{cov}(\hat{\beta}_2, \hat{\beta}_1) & \text{var}(\hat{\beta}_2) & \dots & \text{cov}(\hat{\beta}_2, \hat{\beta}_k) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(\hat{\beta}_k, \hat{\beta}_1) & \text{cov}(\hat{\beta}_k, \hat{\beta}_2) & \dots & \text{var}(\hat{\beta}_k) \end{pmatrix} = \sigma^2 (X^T X)^{-1}.$$

Ước lượng của phương sai sai số ngẫu nhiên σ^2

$$\hat{\sigma}^2 = \frac{e_1^2 + e_2^2 + \dots + e_n^2}{n - k}$$

Ví dụ 2.3

Sử dụng tập số liệu ch2vd5.wf1. Hãy ước lượng hàm hồi quy tuyến tính của CT theo TN và TS, trong đó CT là chi tiêu (triệu đồng/năm), TN là thu nhập từ lao động (triệu đồng/năm) và TS là giá trị tài sản (tỷ đồng) của hộ gia đình.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	18.86018	8.832144	2.135402	0.0410
TN	0.791224	0.015991	49.48055	0.0000
TS	0.015818	0.003984	3.970277	0.0004

R-squared	0.999545	Mean dependent var	1610.415
Adjusted R-squared	0.999515	S.D. dependent var	557.2878
S.E. of regression	12.27498	Akaike info criterion	7.939512
Sum squared resid	4520.257	Schwarz criterion	8.075558
Log likelihood	-128.0019	Hannan-Quinn criter.	7.985287
F-statistic	32963.98	Durbin-Watson stat	1.912696
Prob(F-statistic)	0.000000		

(a) Kết quả hồi quy

	C	TN	TS
C	78.00677	0.073754	-0.023956
TN	0.073754	0.000256	-6.23E-05
TS	-0.023956	-6.23E-05	1.59E-05

(b) Ma trận hiệp phương sai

- $\widehat{\beta}_1 = 18,8601 \rightarrow$ với các hộ không có thu nhập và tài sản thì mức chi tiêu trung bình của họ vào khoảng 18,8601 triệu đồng/năm.
- $\widehat{\beta}_2 = 0,7912 \rightarrow$ khi thu nhập hộ gia đình tăng 1 triệu đồng/năm và giá trị tài sản không thay đổi thì mức chi tiêu trung bình tăng khoảng 0,7912 triệu đồng/năm.
- $\widehat{\beta}_3 = 0,0158 \rightarrow$ khi giá trị tài sản tăng 1 tỷ đồng và thu nhập hộ gia đình không thay đổi thì mức chi tiêu trung bình tăng khoảng 0,0158 triệu đồng/năm.

Các giả thiết của mô hình

- ✓ Giả thiết 1: Mô hình được ước lượng trên cơ sở mẫu ngẫu nhiên kích thước $n : \{(X_i, Y_i), i = 1, 2, \dots, n\}$.
- ✓ Giả thiết 2: Kỳ vọng của sai số ngẫu nhiên tại mỗi giá trị (X_{2i}, \dots, X_{ki}) bằng 0, tức là

$$E(u_i) = E(u|X_{2i}, \dots, X_{ki}) = 0.$$

- ✓ Giả thiết 3: Phương sai của sai số ngẫu nhiên tại mỗi giá trị (X_{2i}, \dots, X_{ki}) đều bằng nhau, tức là

$$\text{var}(u|X_{2i}, \dots, X_{ki}) = \sigma^2, \forall i.$$

- ✓ Giả thiết 4: Giữa các biến độc lập X_2, X_3, \dots, X_k không có đa cộng tuyến.