



HỒI QUY ĐA BIẾN

GV : Đinh Công Khải – Chương trình Fulbright
Môn: Các Phương Pháp Định Lượng – MPP3

Giới thiệu mô hình hồi qui tuyến tính đa biến

□ Hàm hồi qui tổng thể tuyến tính (PRF)

$$E(Y|X_k's) = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_K X_{Ki}$$

- $E(Y|X's)$ là *trung bình* (tổng thể) của phân phối của Y với điều kiện các biến X_{ki} ($k = 1 - K$)
- β_1 là *tung độ góc*; $\beta_2, \dots, \beta_K$ là *hệ số hồi qui riêng*.

$$\beta_k = \frac{\partial E[Y_i | X's]}{\partial X_k}$$

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_K X_{Ki} + u_i$$

Giới thiệu mô hình hồi qui tuyến tính đa biến

□ Ví dụ:

- ❖ $Q^D = f(\text{giá, thu nhập, giá của SP thay thế, quy mô thị trường, ...})$
- ❖ $Q^S = f(\text{vốn, lao động, công nghệ})$
- ❖ $\text{Lương nhân viên} = f(\text{trình độ, kinh nghiệm, giới tính, độ tuổi, ...})$
- ❖ $\text{Giá nhà} = f(\text{diện tích, số phòng ngủ, số phòng tắm, ...})$

Mô hình hồi qui tuyến tính đa biến

□ Hàm hồi qui mẫu (SRF)

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \dots + \hat{\beta}_K X_{Ki}$$

trong đó:

$\hat{Y}_i$ là ước lượng của $E(Y_i|X_i)$

$\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_K$ là các ước lượng của $\beta_1, \beta_2, \dots, \beta_K$.

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{\beta}_K X_{Ki} + \hat{u}_i = \hat{Y}_i + \hat{u}_i$$

Phương pháp bình phương tối thiểu thông thường (OLS)

□ Phương pháp OLS

$$\min_{\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_K} \sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \dots - \hat{\beta}_K X_{Ki})^2$$

$$\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_1} = -2 \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i} - \dots - \hat{\beta}_K X_{Ki}) = 0$$

$$\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_2} = -2 \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i} - \dots - \hat{\beta}_K X_{Ki}) X_{2i} = 0$$

⋮

$$\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_K} = -2 \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i} - \dots - \hat{\beta}_K X_{Ki}) X_{Ki} = 0$$

Phương pháp bình phương tối thiểu thông thường (OLS)

□ Giả sử chúng ta có hàm hồi qui

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

$$\hat{\beta}_2 = \frac{\sum Y_i X_{2i} \sum X_{3i}^2 - \sum Y_i X_{3i} \sum X_{2i} X_{3i}}{\sum X_{2i}^2 \sum X_{3i}^2 - (\sum X_{2i} X_{3i})^2}$$

$$\hat{\beta}_3 = \frac{\sum Y_i X_{3i} \sum X_{2i}^2 - \sum Y_i X_{2i} \sum X_{2i} X_{3i}}{\sum X_{2i}^2 \sum X_{3i}^2 - (\sum X_{2i} X_{3i})^2}$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3$$

Ý nghĩa của các hệ số ước lượng trong mô hình hồi qui tuyến tính đa biến

- Được gọi là **hệ số hồi qui riêng** hay **hệ số độ dốc riêng**.
- Ý nghĩa: *Nếu như các biến giải thích khác không đổi, khi một biến giải thích X_{ki} thay đổi một đơn vị thì biến phụ thuộc sẽ thay đổi trung bình là $\hat{\beta}_k$ đơn vị.*
- $\hat{\beta}_k$ phản ánh sự tác động trực tiếp của biến giải thích X_{ki} lên biến phụ thuộc sau khi đã loại trừ ảnh hưởng các biến hồi qui khác.

Mô hình hồi qui tuyến tính cổ điển Gauss (CLRM): Các giả thiết của OLS

- Giá trị kỳ vọng của u_i bằng không: $E(u_i | X's) = 0$
 - Không có tương quan chuỗi: $cov(u_i, u_j | X's) = 0$ với $i \neq j$
 - Phương sai đồng nhất: $var(u_i) = \sigma^2$
 - Nhiễu ngẫu nhiên không có tương quan với các X : $cov(u_i, X_{ki}) = 0$
 - Không nhận dạng sai mô hình
 - Không có hiện tượng đa cộng tuyến
- $\lambda_2 X_{2i} + \lambda_3 X_{3i} + \dots + \lambda_K X_{Ki} = 0 \rightarrow$ Có hiện tượng đa cộng tuyến

Mô hình hồi qui tuyến tính cổ điển Gauss (CLRM): Các giả thiết của OLS

- ▣ **Định lý Gauss-Markov:** Ước lượng của OLS là ước lượng tuyến tính không thiên lệch, có tính nhất quán, và có hiệu quả nhất, BLUE.

Độ chính xác của ước lượng

- ▣ **Phương sai và độ lệch chuẩn của ước lượng**

$$Var(\hat{\beta}_2) = \frac{\sum x_{3i}^2}{\sum x_{2i}^2 \sum x_{3i}^2 - (\sum x_{2i} x_{3i})^2} \sigma^2 = \frac{1}{\sum x_{2i}^2 (1 - r_{23}^2)} \sigma^2$$

$$r_{23}^2 = \frac{(\sum x_{2i} x_{3i})^2}{\sum x_{2i}^2 \sum x_{3i}^2}$$

$$Var(\hat{\beta}_3) = \frac{\sum x_{2i}^2}{\sum x_{2i}^2 \sum x_{3i}^2 - (\sum x_{2i} x_{3i})^2} \sigma^2 = \frac{1}{\sum x_{3i}^2 (1 - r_{23}^2)} \sigma^2$$

trong đó

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n - 3} \quad (\text{mẫu số sẽ bằng } n-K \text{ trong trường hợp tổng quát})$$

Độ chính xác của ước lượng

- ❑ **Điều kiện:** Số lượng các quan sát n phải lớn hơn số lượng các tham số được ước lượng ($n > K$)
- ❑ **Đồng phương sai giữa 2 ước lượng**

$$\text{Cov}(\hat{\beta}_2, \hat{\beta}_3) = \frac{-r_{23}}{(1 - r_{23}^2)^2 \sqrt{x_{2i}^2} \sqrt{x_{3i}^2}} \sigma^2$$

Độ thích hợp của mô hình

12

Mối liên hệ giữa TSS, ESS, và RSS

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

$$\sum (Y_i - \bar{Y})^2 = \sum (\hat{Y}_i - \bar{Y})^2 + \sum (Y_i - \hat{Y}_i)^2$$

TSS = Tổng bình phương toàn phần

ESS = Tổng bình phương giải thích được

RSS = Tổng bình phương phần dư

Độ thích hợp của mô hình (goodness of fit)

- Hệ số xác định (coefficient of determination)

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum \hat{u}_i^2}{\sum y_i^2}$$

- $0 \leq R^2 \leq 1$
- $R^2 = 1$, các biến độc lập giải thích 100% sự biến thiên của biến phụ thuộc
- $R^2 = 0$, mô hình không giải thích được bất kỳ sự biến đổi nào của biến phụ thuộc

Độ thích hợp của mô hình

- Hệ số xác định có điều chỉnh

$$\bar{R}^2 = 1 - \frac{RSS/(n-K)}{TSS/(n-1)} = 1 - \frac{\sum \hat{u}_i^2/(n-K)}{\sum y_i^2/(n-1)}$$

$$\bar{R}^2 = 1 - (1 - R^2) \frac{n-1}{n-K}$$

- Khi so sánh 2 mô hình dựa trên tiêu chí R^2 cần lưu ý rằng *cỡ mẫu n* và *biến phụ thuộc của 2 mô hình phải giống nhau.*
- R^2 , R^2 điều chỉnh cao **hay** Ý nghĩa lý thuyết và ý nghĩa thống kê của X_s ?