HỒI QUY ĐA BIẾN

GV : Đinh Công Khải – FETP
Môn: Các Phương Pháp Định Lượng – MPP5

Giới thiệu mô hình hồi qui tuyến tính đa biến

- **Hàm hồi qui tổng thể tuyến tính (PRF)**

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

  - E(Y|X's) là *trung bình* (tổng thể) của phân phối của Y với điều kiện các biến X_{ki} (k = 2 - K)
  - β_1 là *tung độ gốc*; $\beta_{2,\ldots,}\beta_K$ là *hệ số hồi qui riêng.*

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

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

Giới thiệu mô hình hồi qui tuyến tính đa biến

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

Mô hình hồi qui tuyến tính đa biến

- **Hàm hồi qui mẫu (SRF)**

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

trong đó:

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

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

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

Phương pháp bình phương tối thiểu thông thường (OLS)

- Phương pháp OLS

$$\min_{\hat\beta_1,\hat\beta_2,\ldots,\hat\beta_K} \sum \hat u_i^2 = \sum (Y_i - \hat\beta_1 - \hat\beta_2 X_{2i} - \ldots - \hat\beta_K X_{Ki})^2$$

$$\frac{\partial \sum \hat u_i^2}{\partial \hat\beta_1} = -2\sum\left(\mathrm{Y_i} - \hat\beta_1 - \hat\beta_2 X_{2i} - \hat\beta_3 X_{3i} - \cdots - \hat\beta_{\mathrm{K}} X_{Ki}\right) = 0$$

$$\frac{\partial \sum \hat u_i^2}{\partial \hat\beta_2} = -2\sum\left(\mathrm{Y_i} - \hat\beta_1 - \hat\beta_2 X_{2i} - \hat\beta_3 X_{3i} - \cdots - \hat\beta_{\mathrm{K}} X_{Ki}\right)\mathrm{X_{2i}} = 0$$

$$\vdots$$

$$\frac{\partial \sum \hat u_i^2}{\partial \hat\beta_K} = -2\sum\left(\mathrm{Y_i} - \hat\beta_1 - \hat\beta_2 X_{2i} - \hat\beta_3 X_{3i} - \cdots - \hat\beta_{\mathrm{K}} X_{Ki}\right)\mathrm{X_{Ki}} = 0$$

Phương pháp bình phương tối thiểu thông thường (OLS)

- Giả sử chúng ta có hàm hồi qui

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

$$\hat\beta_2 = \frac{\left(\sum y_i x_{2i}\right)\left(\sum x_{3i}^2\right) - \left(\sum y_i x_{3i}\right)\left(\sum x_{2i} x_{3i}\right)}{\left(\sum x_{2i}^2\right)\left(\sum x_{3i}^2\right) - \left(\sum x_{2i} x_{3i}\right)^2}$$

$$\hat\beta_3 = \frac{\left(\sum y_i x_{3i}\right)\left(\sum x_{2i}^2\right) - \left(\sum y_i x_{2i}\right)\left(\sum x_{2i} x_{3i}\right)}{\left(\sum x_{2i}^2\right)\left(\sum x_{3i}^2\right) - \left(\sum x_{2i} x_{3i}\right)^2}$$

$$\hat\beta_1 = \overline{\mathrm{Y}} - \hat\beta_2 \overline{X}_2 - \hat\beta_3 \overline{X}_3$$

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

- $\hat{\beta}_k$ (k = 2-K) 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: $E(u_i \mid X's) = 0$
- Không có tương quan chuỗi: $cov(u_i, u_j \mid X's) = 0$ 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 có thiên lệch đặc trưng (thiếu biến quan trọng, dạng mô hình sai)
- Không có hiện tượng đa cộng tuyến

$\lambda_2 X_{2i} + \lambda_3 X_{3i} + \cdots + \lambda_K X_{Ki} = 0$ ➔ 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***

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

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

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

trong đó

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-3}$$ (mẫu số sẽ bằng n-K 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**

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

Độ thích hợp của mô hình

12

Mối liên hệ giữa TSS, ESS, và RSS

TSS = ESS + 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)

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

- $0 \le R^2 \le 1$
- $R^2 = 1$, 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

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

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

- Khi so sánh 2 mô hình dựa trên tiêu chí R^2 hay R^2 điều chỉnh cần lưu ý rằng *cỡ mẫu n* và *biến phụ thuộc của 2 mô hình phải giống nhau (các biến giải thích có thể ở bất kỳ dạng gì)*.