**Chương 2**

# Mô hình hồi qui hai biến
# Ước lượng và kiểm định giả thiết

## 1. Phương pháp bình phương bé nhất

Giả sử : $Y_i = \beta_1 + \beta_2 X_i + U_i$ (PRF)

và có một mẫu n quan sát ($Y_i$, $X_i$).

Cần ước lượng (PRF).

Ta có : $Y_i = \hat{Y}_i + e_i$ (SRF)

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

Theo phương pháp OLS, để

$\hat{Y}_i$ càng gần với $Y_i$ thì $\hat{\beta}_1, \hat{\beta}_2$ cần thỏa mãn :

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2 \rightarrow \min$$

Suy ra $\hat{\beta}_1, \hat{\beta}_2$ cần thỏa mãn :

$$\begin{cases} \dfrac{\partial \sum_{i=1}^{n} e_i^2}{\partial \hat{\beta}_1} = \sum_{i=1}^{n} 2(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)(-1) = 0 \\ \dfrac{\partial \sum_{i=1}^{n} e_i^2}{\partial \hat{\beta}_2} = \sum_{i=1}^{n} 2(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)(-X_i) = 0 \end{cases}$$

giải hệ, ta có :

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{X}\,\overline{Y}}{\sum_{i=1}^{n} X_i^2 - n(\overline{X})^2} \qquad \hat{\beta}_1 = \overline{Y} - \hat{\beta}_2 \overline{X}$$

Có thể chứng minh được :

$$\sum_{i=1}^{n} x_i y_i = \sum_{i=1}^{n} X_i Y_i - n\overline{X}\,\overline{Y}$$

$$\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} X_i^2 - n(\overline{X})^2$$

với

$$x_i = X_i - \overline{X}$$

$$y_i = Y_i - \overline{Y}$$

Nên có thể biểu diễn :

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

**Ví dụ 1**: Giả sử cần nghiên cứu chi tiêu tiêu dùng của hộ gia đình phụ thuộc thế nào vào thu nhập của họ, người ta tiến hành điều tra, thu được một mẫu gồm 10 hộ gia đình với số liệu như sau :

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Y** | **70** | **65** | **90** | **95** | **110** | **115** | **120** | **140** | **155** | **150** |
| **X** | **80** | **100** | **120** | **140** | **160** | **180** | **200** | **220** | **240** | **260** |

Trong đó : Y – chi tiêu hộ gia đình (USD/tuần)

X – thu nhập hộ gia đình (USD/tuần)

Giả sử Y và X có quan hệ tuyến tính. Hãy ước lượng mô hình hồlI qui của Y theo X.

# 2. Các giả thiết cổ điển của mô hình hồi qui tuyến tính

- **Giả thiết 1** : Biến độc lập $X_i$ là phi ngẫu nhiên, các giá trị của chúng phải được xác định trước.
- **Giả thiết 2** : Kỳ vọng có điều kiện của sai số ngẫu nhiên bằng 0 :

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

- **Giả thiết 3** : (Phương sai thuần nhất ) Các sai số ngẫu nhiên có phương sai bằng nhau :

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

- **Giả thiết 4** : Không có hiện tượng tương quan giữa các sai số ngẫu nhiên :

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

- **Giả thiết 5** : Không có hiện tượng tương quan giữa biến độc lập $X_i$ và sai số ngẫu nhiên $U_i$ : $Cov\ (X_i\ ,\ U_i\ ) = 0 \qquad \forall\ i$

- **Định lý Gauss – Markov** : Với các giả thiết từ 1 đến 5 của mô hình hồi qui tuyến tính cổ điển, các ước lượng OLS là các ước lượng ***tuyến tính, không chệch*** và có ***phương sai bé nhất*** trong lớp các ước lượng tuyến tính, không chệch.

# 3. Phương sai và sai số chuẩn của các ước lượng

| Phương sai | Sai số chuẩn |
| --- | --- |
| $Var(\hat{\beta}_1) = \sigma^2_{\hat{\beta}_1} = \dfrac{\sum X_i^2}{n\sum x_i^2}\sigma^2$ | $se(\hat{\beta}_1) = \sigma_{\hat{\beta}_1} = \sqrt{\sigma^2_{\hat{\beta}_1}}$ |
| $Var(\hat{\beta}_2) = \sigma^2_{\hat{\beta}_2} = \dfrac{1}{\sum x_i^2}\sigma^2$ | $se(\hat{\beta}_2) = \sigma_{\hat{\beta}_2} = \sqrt{\sigma^2_{\hat{\beta}_2}}$ |

Trong đó : $\sigma^2 = var(U_i)$. Do $\sigma^2$ chưa biết nên dùng ước lượng của nó là

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n-2}$$

# 4. Hệ số xác định và hệ số tương quan

**a. Hệ số xác định :** Dùng để đo mức độ phù hợp của hàm hồi qui.

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

Trong đó : TSS = ESS + RSS

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

$$ESS = \sum_{i=1}^{n} (\hat{Y}_i - \overline{Y})^2$$

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

Miền xác định của $R^2$ :

$$0 \le R^2 \le 1$$

$R^2 \rightarrow 1$ : hàm hồi qui càng phù hợp.

$R^2 \rightarrow 0$ : hàm hồi qui càng ít phù hợp

Ví dụ : …

**b. Hệ số tương quan :** Là số đo mức độ chặt chẽ của quan hệ tuyến tính giữa X và Y.

$$r = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum (X_i - \overline{X})^2 \sum (Y_i - \overline{Y})^2}} = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2 \sum y_i^2}}$$

Chứng minh được : $|r| = \sqrt{R^2}$

Và dấu của r trùng với dấu của hệ số của X trong hàm hồi qui ($\hat{\beta}_2$).

**<u>Tính chất của hệ số tương quan</u>** :

1. Miền giá trị của r : $-1 \le r \le 1$

   | r| → 1 : quan hệ tuyến tính giữa X và Y càng chặt chẽ.

2. r có tính đối xứng : $r_{XY} = r_{YX}$

3. Nếu X, Y độc lập thì r = 0. Điều ngược lại không đúng.

# 5. Phân phối xác suất của các ước lượng

**Giả thiết 6** : $U_i$ có phân phối N (0, $\sigma^2$),

Với giả thiết 6, các ước lượng có thêm các tính chất sau :

1. Khi số quan sát đủ lớn thì các ước lượng xấp xỉ với giá trị thực của phân phối :

$$\hat{\beta}_1 \xrightarrow{n \to \infty} \beta_1, \ \hat{\beta}_2 \xrightarrow{n \to \infty} \beta_2$$

2. $\hat{\beta}_1 \sim N(\beta_1, \sigma^2_{\hat{\beta}_1}) \Rightarrow Z = \dfrac{\hat{\beta}_1 - \beta_1}{\sigma_{\hat{\beta}_1}} \sim N(0,1)$

$\hat{\beta}_2 \sim N(\beta_2, \sigma^2_{\hat{\beta}_2}) \Rightarrow Z = \dfrac{\hat{\beta}_2 - \beta_2}{\sigma_{\hat{\beta}_2}} \sim N(0,1)$

3. $\dfrac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-2)$

4. $\mathbf{Y_i \sim N(\beta_1 + \beta_2 X_i, \sigma^2)}$

# 6. Khoảng tin cậy của các hệ số hồi qui

- Sử dụng phân phối của thống kê t :

$$\boxed{t = \frac{\hat{\beta}_j - \beta_j}{\hat{se}(\hat{\beta}_j)} \sim t(n-2) \qquad j = 1,2}$$

Ta có khoảng tin cậy của $\beta_1$ :

$$\hat{\beta}_1 - \hat{se}(\hat{\beta}_1).t_{\alpha/2}(n-2) \le \beta_1 \le \hat{\beta}_1 + \hat{se}(\hat{\beta}_1).t_{\alpha/2}(n-2)$$

Ta có khoảng tin cậy của $\beta_2$ :

$$\hat{\beta}_2 - \hat{se}(\hat{\beta}_2).t_{\alpha/2}(n-2) \le \beta_2 \le \hat{\beta}_2 + \hat{se}(\hat{\beta}_2).t_{\alpha/2}(n-2)$$

# 7. Kiểm định giả thiết về các hệ số hồi qui

- Giả sử $H_0 : \beta_2 = a \; (a = \text{const})$

  $H_1 : \beta_2 \neq a$

Có 2 cách kiểm định :

1. Dùng khoảng tin cậy :

   Khoảng tin cậy của $\beta_2$ là $[\alpha, \beta]$

   - Nếu $a \notin [\alpha, \beta] \Rightarrow$ bác bỏ $H_0$
   - Nếu $a \in [\alpha, \beta] \Rightarrow$ chấp nhận $H_0$

2. Dùng kiểm định t :

   Thống kê sử dụng : $t = \dfrac{\hat{\beta}_2 - \beta_2}{\hat{se}(\hat{\beta}_2)} \sim t(n-2)$

Có hai cách đọc kết quả kiểm định t :

Cách 1 : dùng giá trị tới hạn.

- Tính $t = \dfrac{\hat{\beta}_2 - a}{\hat{se}(\hat{\beta}_2)}$

- Tra bảng t tìm $t_{\alpha/2}(n-2)$

- Nếu $|t| > t_{\alpha/2}(n-2) \Rightarrow$ bác bỏ $H_0$.

- Nếu $|t| \le t_{\alpha/2}(n-2) \Rightarrow$ chấp nhận $H_0$.

Cách 2 : Dùng p-value (mức ý nghĩa chính xác)

$$p = P(|T| > t_a)$$

$$\text{với } t_a = t = \frac{\hat{\beta}_2 - a}{s\hat{e}(\hat{\beta}_2)}$$

- Nếu $p \le \alpha \Rightarrow$ bác bỏ $H_0$.

- Nếu $p > \alpha \Rightarrow$ chấp nhận $H_0$.

# 8. Kiểm định sự phù hợp của hàm hồi qui. Phân tích hồi qui và phân tích phương sai

- Giả thiết $H_0 : \beta_2 = 0$ ( hàm hồi qui không phù hợp)

  $H_1 : \beta_2 \neq 0$ (hàm hồi qui phù hợp)

Sử dụng phân phối của thống kê F :

$$F = \frac{(\hat{\beta}_2 - \beta_2)^2 \sum x_i^2 / 1}{\sum e_i^2 /(n-2)} \sim F(1, n-2)$$

Khi $\beta_2 = 0$ , F có thể viết :

$$F = \frac{\hat{\beta}_2^2 \sum x_i^2}{\sum e_i^2 /(n-2)} = \frac{ESS/1}{RSS/(n-2)} = \frac{R^2/1}{(1-R^2)/(n-2)} \quad (*)$$

<u>Nên có thể dùng qui tắc kiểm định sau</u> :

- Tính

$$F = \frac{R^2/1}{(1-R^2)/(n-2)}$$

- Nếu $F > F_\alpha(1, n\text{-}2) \Rightarrow$ bác bỏ $H_0$

$\Rightarrow$ hàm hồi qui phù hợp.

Mặt khác, cũng từ (*) cho thấy :

Phân tích phương sai cho phép đưa ra các phán đoán thống kê về độ thích hợp của hồi qui ( xem bảng phân tích phương sai).

* **<u>Một số chú ý khi kiểm định giả thiết</u>** :

- Khi nói "chấp nhận giả thiết $H_0$", không có nghĩa $H_0$ đúng.
- Lựa chọn mức ý nghĩa $\alpha$ : $\alpha$ có thể tùy chọn, thường người ta chọn mức 1%, 5%, nhiều nhất là 10%.

# 9. Dự báo

a. Dự báo giá trị trung bình :

Cho X =$X_0$ , tìm $E(Y/X_0)$.

- Dự báo điểm của $E(Y/X_0)$ là :

$$\hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_0$$

- Dự báo khoảng của $E(Y/X_0)$ là :

$$\boxed{\hat{Y}_0 - s\hat{e}(\hat{Y}_0).t_{\alpha/2}^{(n-2)} \le E(Y/X_0) \le \hat{Y}_0 + s\hat{e}(\hat{Y}_0).t_{\alpha/2}^{(n-2)}}$$

Trong đó :

$$v\hat{a}r(\hat{Y}_0) = \left[\frac{1}{n} + \frac{(X_0 - \overline{X})^2}{\sum x_i^2}\right] \times \hat{\sigma}^2$$

## b. Dự báo giá trị cá biệt :

Cho $X = X_0$ , tìm $Y_0$.

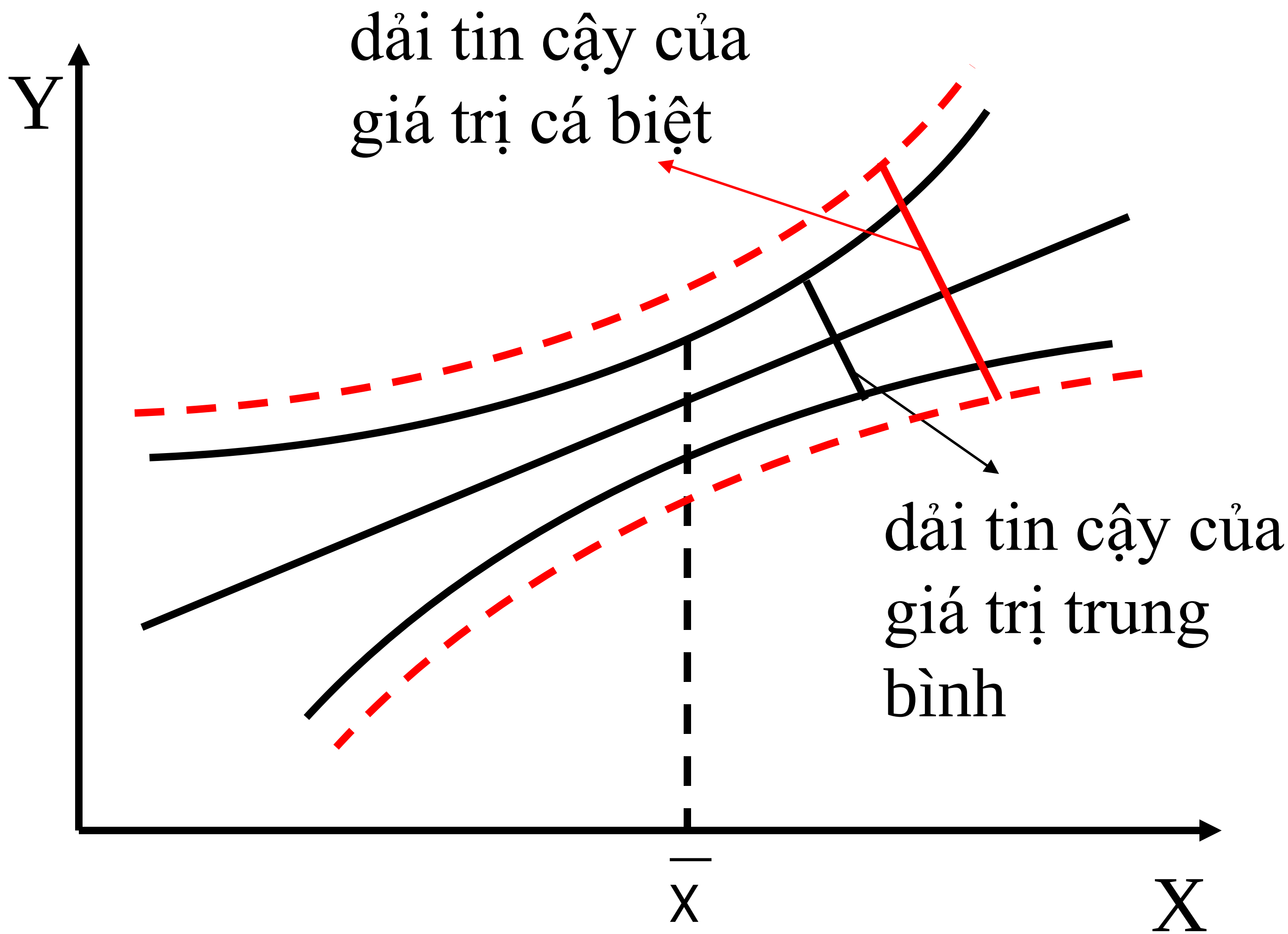
$$\boxed{\hat{Y}_0 - \hat{se}(Y_0 - \hat{Y}_0).t_{\alpha/2}^{(n-2)} \le Y_0 \le \hat{Y}_0 + \hat{se}(Y_0 - \hat{Y}_0).t_{\alpha/2}^{(n-2)}}$$

Trong đó :

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

nên

$$\widehat{\operatorname{var}}(Y_0 - \hat{Y}_0) = \widehat{\operatorname{var}}(\hat{Y}_0) + \hat{\sigma}^2$$




# * Đặc điểm của dự báo khoảng

# 10. Trình bày kết quả hồi qui

$$
\begin{array}{llll}
\hat{Y}_i & = \hat{\beta}_1 & + \hat{\beta}_2 X_i & R^2 = \\
se & = \hat{se}(\hat{\beta}_1) & \hat{se}(\hat{\beta}_2) & n = \\
t & = t_1 & t_2 & F = \\
p & = p(>t_1) & p(>t_2) & p(>F) =
\end{array}
$$

**Trong đó :**

$$t_1 = \frac{\hat{\beta}_1 - 0}{\hat{se}(\hat{\beta}_1)} \qquad t_2 = \frac{\hat{\beta}_2 - 0}{\hat{se}(\hat{\beta}_2)}$$

---

$$
\begin{array}{lllll}
\hat{Y}_i & = & 24,4545 & + \; 0,5091\, X_i & R^2 = 0,9621 \\
se & = & (6,4138) & (0,0357) & n = 10 \\
t & = & (3,813) & (14,243) & F = 202,87 \\
p & = & (0,005) & (0,000) & p = (0,000)
\end{array}
$$

# 11. Đánh giá kết quả của phân tích hồi qui

- Dấu của các hệ số hồi qui ước lượng được phù hợp với lý thuyết hay tiên nghiệm không.
- Các hệ số hồi qui ước lượng được có ý nghĩa về mặt thống kê hay không.
- Mức độ phù hợp của mô hình ($R^2$).
- Kiểm tra xem mô hình có thỏa mãn các giả thiết của mô hình hồi qui tuyến tính cổ điển hay không.