CHƯƠNG II. HỒI QUY TUYẾN TÍNH ĐƠN (PHẦN 3)

ThS. Vũ Thị Phương Mai
Khoa Kinh Tế Quốc Tế- Đại học Ngoại Thương

4. Phương pháp hợp lý tối đa (MLE)

- Ngoài phương pháp bình phương nhỏ nhất (OLS), người ta còn hay sử dụng một công cụ khác để ước lượng các tham số của mô hình kinh tế lượng, đó là **phương pháp ước lượng hợp lý tối đa** (maximum likelihood estimation).
- Phương pháp này được đánh giá là mạnh hơn so với phương pháp OLS về một số điểm lý thuyết. Chúng ta sẽ không đi sâu vào nghiên cứu phương pháp này nhưng việc nắm được bản chất của nó sẽ giúp ta trong việc đọc và hiểu các kết quả hồi quy được chạy trên các phần mềm kinh tế lượng.
- Điều cơ bản nhất chúng ta cần nắm được ở đây là:

4. Phương pháp hợp lý tối đa (MLE)

- Nếu u_i tuân theo quy luật phân phối chuẩn thì hệ số hồi quy của các ước lượng theo phương pháp ML và OLS (các β_i) là như nhau. Điều này luôn đúng trong cả hàm hồi quy đơn lẫn hàm hồi quy bội.
- Ước lượng ML của $\sigma^2 = \sum_{i=1}^{n} \hat{u}_i^2 / n$ là ước lượng chệch còn ước lượng OLS của $\sigma^2 = \sum_{i=1}^{n} \hat{u}_i^2 /(n-2)$ là ước lượng không chệch.
- Tuy nhiên do kích thước mẫu n theo phương pháp ML lớn hơn kích thước mẫu theo phương pháp OLS, nên giá trị ước lượng của σ^2 theo cả hai phương pháp trên có xu hướng bằng nhau.
- Do vậy, một cách tiệm cận, ước lượng của σ^2 theo phương pháp ML cũng được đánh giá là ước lượng không chệch.

4. Phương pháp hợp lý tối đa (MLE)

- Trên thực tế, người ta ưa chuộng phương pháp OLS hơn phương pháp ML bởi vì : phương pháp OLS cùng với giả thiết về phân phối chuẩn của ui cung cấp các công cụ cần thiết dùng để ước lượng và kiểm định các giả thiết thống kê của mô hình hồi quy tuyến tính trong khi đó nếu sử dụng phương pháp ML, ta sẽ phải đối mặt với các lý thuyết toán phức tạp hơn.

5. Ước lượng khoảng và kiểm định giả thiết thống kê

- 5.1. Ước lượng khoảng: một vài tư tưởng
- 5.2. Khoảng tin cậy của hệ số hồi quy β_1 và β_2
- 5.3. Khoảng tin cậy của phương sai
- 5.4. Kiểm định giả thiết thống kê

5.1. Ước lượng khoảng: một vài tư tưởng

- Ta biết rằng β_1 và β_2 là ước lượng điểm (point estimators) của β_1 và β_2 nhưng do các dao động của việc lấy mẫu lặp lại nên các ước lượng điểm có thể khác với giá trị thực mặc dù trung bình giá trị của các ước lượng β_1 và β_2 bằng với giá trị thực β_1 và β_2.
- Do đó người ta muốn xây dựng một khoảng xung quanh giá trị ước lượng điểm với lòng tin rằng giá trị thực sẽ nằm trong khoảng đó với một độ tin cậy nhất định.

→ Cách làm này gọi là ước lượng khoảng.

5.1. Ước lượng khoảng: một vài tư tưởng

- Giả sử ta muốn tìm $\hat{\beta}_2$ sao cho giá trị của nó gần với giá trị của β_2 nhất. Muốn vậy, ta phải tìm hai số dương δ và α, nằm trong khoảng (0,1) sao cho xác suất để khoảng ngẫu nhiên ($\hat{\beta}_2$ -δ, $\hat{\beta}_2$ +δ) chứa giá trị thực của β_2 là 1- α :

$$P(\hat{\beta}_2 - \delta \leq \beta_2 \leq \hat{\beta}_2 + \delta) = 1 - \alpha$$

  - Một khoảng ngẫu nhiên (random interval) như trên gọi là **khoảng tin cậy** (confidence interval);
  - (1- α) được gọi là **hệ số tin cậy** (confidence coefficent);
  - α (0 < α <1) là **mức ý nghĩa** (level of significance).
  - Các điểm tận cùng của khoảng tin cậy được gọi là các **giá trị tới hạn** (critical values) → ($\hat{\beta}_2$ - δ) là giá trị tới hạn dưới còn ($\hat{\beta}_2$ + δ) là giá trị tới hạn trên.

5.1. Ước lượng khoảng: một vài tư tưởng

- Trong thực hành, người ta hay sử dụng α và $(1- \alpha)$ ở dạng phần trăm.
- Ví dụ: nếu $\alpha = 0,05$ (hoặc 5%) thì xác suất để khoảng tin cậy chứa giá trị thực của β_2 là 0,95 (hay 95%).

5.2. Khoảng tin cậy của hệ số hồi quy β_1 và β_2

- 5.2.1. Khoảng tin cậy của hệ số β_2
- 5.2.2. Khoảng tin cậy của hệ số β_1

5.2.1. Khoảng tin cậy của hệ số β_2

- Như đã học trong phần 3.3, với giả thiết u_i tuân theo quy luật phân phối chuẩn, ta có các ược lượng OLS β_1 và β_2 cũng tuân theo quy luật này.
- Bởi vậy, biến Z được gọi là biến phân phối chuẩn hóa với:

$$Z = \frac{\beta_2 - \beta_2}{se(\beta_2)} = \frac{(\beta_2 - \beta_2)\sqrt{\sum x_i^2}}{\sigma} \qquad [5.01]$$

- Trong thực tế, ít khi ta biết được giá trị thực của σ^2 mà chỉ có được giá trị ước lượng không chệch của nó là σ^2. Khi đó, nếu thay σ bằng σ thì [5.01] có thể được viết lại như sau :

5.2.1. Khoảng tin cậy của hệ số β_2

$$t = \frac{\hat{\beta}_2 - \beta_2}{se(\hat{\beta}_2)} = \frac{(\hat{\beta}_2 - \beta_2)\sqrt{\sum x_i^2}}{\sigma} \quad [5.02]$$

- trong đó se ($\hat{\beta}_2$) là ước lượng của sai số tiêu chuẩn.
- Người ta cũng chứng minh được rằng biến t tuân theo quy luật phân phối Student với (n-2) bậc tự do.
- Do đó, thay vì sử dụng quy luật phân phối chuẩn, chúng ta có thể sử dụng phân phối Student để xây dựng khoảng tin cậy cho β_2 như sau :

5.2.1. Khoảng tin cậy của hệ số β_2

$$P(-t_{(n-2),\alpha/2} \le t \le t_{(n-2),\alpha/2}) = 1-\alpha$$

$$\leftrightarrow \quad P\left(-t_{(n-2),\alpha/2} \le \frac{\hat{\beta}_2 - \beta_2}{se(\hat{\beta}_2)} \le t_{(n-2),\alpha/2}\right) = 1-\alpha$$

$$\leftrightarrow \quad P[\hat{\beta}_2 - t_{(n-2),\alpha/2}\, se(\hat{\beta}_2) \le \beta_2 \le \hat{\beta}_2 + t_{(n-2),\alpha/2}\, se(\hat{\beta}_2)] = 1-\alpha \qquad [5.03]$$

- Như vậy, với độ tin cậy là 100 (1- α) % thì khoảng tin cậy của β_2 là

$$\hat{\beta}_2 \pm t_{(n-2),\alpha/2}\, se(\hat{\beta}_2)$$

5.2.2. Khoảng tin cậy của hệ số β_1

- Tương tự như trên ta có thể xây dựng được khoảng tin cậy cho hệ số β_1 như sau:

$$P[\hat{\beta}_1 - t_{(n-2),\alpha/2}\, se(\hat{\beta}_1) \le \beta_1 \le \hat{\beta}_1 + t_{(n-2),\alpha/2}\, se(\hat{\beta}_1)] = 1 - \alpha \qquad [5.04]$$

- Như vậy, với độ tin cậy là 100 (1- α) % thì khoảng tin cậy của β_1 là:

$$\hat{\beta}_1 \pm t_{(n-2),\alpha/2}\, se(\hat{\beta}_1)$$

5.3. Khoảng tin cậy của phương sai

- Phương sai của tổng thể chính là phương sai của thành phần nhiễu ui mà ta kí hiệu là σ^2.
- Với giả thiết về phân phối chuẩn của số hạng nhiễu, người ta chứng minh được đại lượng ngẫu nhiên:

$$\chi^2 = (n-2)\frac{\hat{\sigma}^2}{\sigma^2}$$

tuân theo quy luật phân phối xác suất khi-đơ χ^2 với (n-2) bậc tự do.

- Vì vậy, sử dụng quy luật phân phối χ^2, ta xây dựng được khoảng tin cậy cho σ^2 như sau:

5.3. Khoảng tin cậy của phương sai

$$P(\chi^2_{(n-2),1-\alpha/2} \le \chi^2 \le \chi^2_{(n-2),\alpha/2}) = 1-\alpha$$

$$\leftrightarrow \quad P(\chi^2_{(n-2),1-\alpha/2} \le \frac{(n-2)\hat{\sigma}^2}{\sigma^2} \le \chi^2_{(n-2),\alpha/2}) = 1-\alpha$$

$$\leftrightarrow \quad P[(n-2)\frac{\hat{\sigma}^2}{\chi^2_{(n-2),\alpha/2}} \le \sigma^2 \le (n-2)\frac{\hat{\sigma}^2}{\chi^2_{(n-2),1-\alpha/2}}] = 1-\alpha \qquad [5.05]$$

- Như vậy, với độ tin cậy là 100(1- α) % thì khoảng tin cậy của σ^2 là:

$$[(n-2)\frac{\hat{\sigma}^2}{\chi^2_{(n-2),\alpha/2}}, (n-2)\frac{\hat{\sigma}^2}{\chi^2_{(n-2),1-\alpha/2}}]$$

- Ví dụ 3: Với kết quả tìm được trong ví dụ 2, hãy tính khoảng tin cậy của hệ số hồi quy β_2 và của phương sai của mô hình ($\hat{u}_i$).

5.4. Kiểm định giả thiết thống kê

- Vấn đề kiểm định giả thiết thống kê đã được tóm lược ngắn gọn trong phần kiến thức bổ trợ (chương 0) về xác suất thống kê.
- Ở mục này chúng ta chỉ trình bày các dạng kiểm định liên quan đến hệ số hồi quy và phương sai của nhiễu trong hồi quy tổng thể, kiểm định sự phù hợp của SRF cùng với các phương pháp tiếp cận để thực hiện các kiểm định này.

5.4. Kiểm định giả thiết thống kê

- **5.4.1. Kiểm định giả thiết về hệ số hồi quy**
  - 5.4.1.1. Phương pháp khoảng tin cậy
  - 5.4.1.2. Phương pháp giá trị tới hạn
  - 5.4.1.3. Phương pháp giá trị p-value
- **5.4.2. Kiểm định giả thiết về phương sai của nhiễu**

5.4.1. Kiểm định giả thiết về hệ số hồi quy

- Có ba dạng giả thiết kiểm định như sau về hệ số hồi quy:

- Hai phía: $\begin{cases} H_0: \beta_i = \beta_i^* \\ H_1: \beta_i \neq \beta_i^* \end{cases}$

- Phía phải: $\begin{cases} H_0: \beta_i \leq \beta_i^* \\ H_1: \beta_i > \beta_i^* \end{cases}$

- Phía trái: $\begin{cases} H_0: \beta_i \geq \beta_i^* \\ H_1: \beta_i < \beta_i^* \end{cases}$

- Trong đó, β_i nhận giá trị là β_1 hoặc β_2 (trong phạm vi mô hình hồi quy đơn mà ta đang xét).
- β_i^* là giả thiết về giá trị thực của β_i, nhận giá trị là β_1^* hoặc β_2^*.

5.4.1. Kiểm định giả thiết về hệ số hồi quy

- Có ba cách để xây dựng quy tắc quyết đinh xem là chấp nhận hay bác bỏ giá thiết H_0, đó là:

  5.4.1.1. Phương pháp khoảng tin cậy

  5.4.1.2. Phương pháp giá trị tới hạn

  5.4.1.3. Phương pháp giá trị p-value

5.4.1.1. Phương pháp khoảng tin cậy

- **Theo mục 5.2, ta có khoảng tin cậy đối xứng của β_i (tương ứng với kiểm định hai phía) là:**

$$[\beta_i - t_{(n-2),\alpha/2}\, se(\beta_i),\ \beta_i + t_{(n-2),\alpha/2}\, se(\beta_i)]$$

→ Nếu giá trị β_i^* không rơi vào khoảng này thì ta bác bỏ giả thiết H_0.

- **Đối với kiểm định phía phải, khoảng tin cậy bên phải của β_i là:**

$$[\beta_i - t_{(n-2),\alpha}\, se(\beta_i),\ +\infty\ ]$$

→ Nếu giá trị β_i^* không rơi vào khoảng này thì ta bác giả thiết H_0.

- **Đối với kiểm định phía trái, khoảng tin cậy bên trái của β_i là:**

$$[-\infty\ ,\ \beta_i + t_{(n-2),\alpha}\, se(\beta_i)]$$

→ Nếu giá trị β_i^* không rơi vào khoảng này thì ta bác giả thiết H_0.

5.4.1.2. Phương pháp giá trị tới hạn

- **Bước 1**: tính giá trị $t_{qs} = \dfrac{\beta_i - \beta_i}{se(\beta_i)}$

- **Bước 2**: tra bảng t-student với mức ý nghĩa α/2 (nếu là kiểm định hai phía) hoặc mức ý nghĩa α (nếu là kiểm định một phía) để có giá trị tới hạn $t_{(n-2),\alpha/2}$ hoặc $t_{(n-2),\alpha}$.

- **Bước 3**: so sánh t_{qs} với giá trị tới hạn. Quy tắc quyết định:

Loại giả thiết	Giá thiết H_0	Giả thiết đối H_1	Miền bác bỏ H_0
Hai phía	$\beta_i = \beta_i^*$	$\beta_i \neq \beta_i^*$	$\lvert t_{qs}\rvert > t_{(n-2),\alpha/2}$
Phía phải	$\beta_i \leq \beta_i^*$	$\beta_i > \beta_i^*$	$t_{qs} > t_{(n-2),\alpha}$
Phía trái	$\beta_i \geq \beta_i^*$	$\beta_i < \beta_i^*$	$t_{qs} < -\ t_{(n-2),\alpha}$

5.4.1.3. Phương pháp giá trị p-value

- **Bước 1**: tính giá trị $t_{qs} = \dfrac{\hat{\beta}_i - \beta_i}{se(\hat{\beta}_i)}$
- **Bước 2**: tính p-value = P (|t| > |t_{qs}|), trong đó t là đại lượng ngẫu nhiên có phân phối t-student với (n-2) bậc tự do.
- **Bước 3**: nếu cho trước mức ý nghĩa α, quy tắc quyết định sẽ là:
  - Kiểm định hai phía: p-value < α: bác bỏ H_0
  - Kiểm định một phía: p-value/2 < α: bác bỏ H_0

- Ví dụ 4: Hãy cho biết hệ số trong mô hình hồi quy ở ví dụ 1 có ý nghĩa thống kê hay không ?

5.4.2. Kiểm định giả thiết về phương sai của nhiễu

- Phương pháp tiến hành kiểm định giả thiết tương tự như kiểm định giả thiết về hệ số hồi quy. Bảng 2.06 trình bày một cách tóm tắt các loại giả thiết, phương pháp kiểm định và quy tắc quyết định.
- Trong giả thiết H_0, σ_0^2 là giá trị số cho trước và:

$$\chi_0^2 = \frac{(n-2)\sigma^2}{\sigma_0^2}$$

$$\text{p-value} = P(\chi^2 > \chi_0^2)$$

Bảng 5.01. Kiểm định giả thiết về phương sai của nhiễu

Loại giả thiết	Giả thiết H_0	Giả thiết đối H_1	Phương pháp	Miền bác bỏ H_0
Hai phía	$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	Khoảng tin cậy	$\sigma_0^2 \notin [(n-2)\dfrac{\sigma^2}{\chi^2_{(n-2),\alpha/2}}, (n-2)\dfrac{\sigma^2}{\chi^2_{(n-2),1-\alpha/2}}]$
			Giá trị tới hạn	$\chi_0^2 > \chi^2_{(n-2),\alpha/2}$ **hoặc** $\chi_0^2 < \chi^2_{(n-2),1-\alpha/2}$
			Giá trị *p-value*	*p-value* $< \alpha/2$ **hoặc** *p-value* $> 1- \alpha/2$
Phía phải	$\sigma^2 = \sigma_0^2$	$\sigma^2 > \sigma_0^2$	Khoảng tin cậy	$\sigma_0^2 \notin [(n-2)\dfrac{\sigma^2}{\chi^2_{(n-2),\alpha}}, +\infty]$
			Giá trị tới hạn	$\chi_0^2 > \chi^2_{(n-2),\alpha}$
			Giá trị *p-value*	*p-value* $< \alpha$
Phía trái	$\sigma^2 = \sigma_0^2$	$\sigma^2 < \sigma_0^2$	Khoảng tin cậy	$\sigma_0^2 \notin [-\infty, (n-2)\dfrac{\sigma^2}{\chi^2_{(n-2),1-\alpha}}]$
			Giá trị tới hạn	$\chi_0^2 < \chi^2_{(n-2),1-\alpha}$
			Giá trị *p-value*	*p-value* $> 1- \alpha$

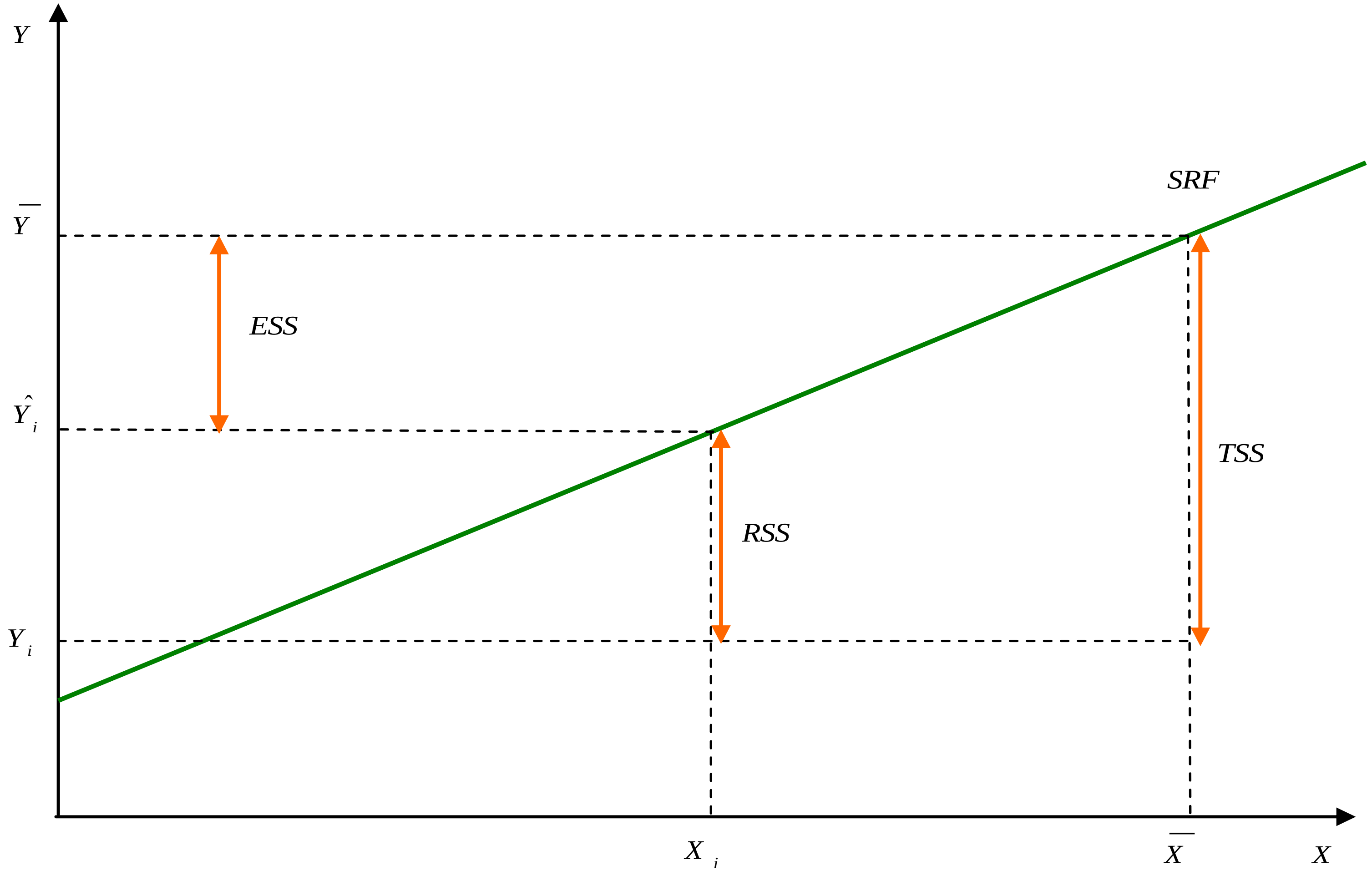
Ví dụ 5 : Giá trị phương sai của nhiễu trong mô hình hồi quy ở ví dụ 1 có lớn hơn 20500 hay không?

6. Phân tích phương sai và kiểm định sự phù hợp của mô hình hồi quy

- 6.1. Các tổng bình phương độ lệch
- 6.2. Hệ số xác định (đơn)
- 6.3. Kiểm định sự phù hợp của mô hình hồi quy

6.1. Các tổng bình phương độ lệch

Hình 6.01. Ý nghĩa hình học của TSS, ESS và RSS



6.1. Các tổng bình phương độ lệch

$$TSS = \sum y_i^2 = \sum (Y_i - \overline{Y})^2 \quad \textbf{[6.01]}$$

là tổng bình phương các độ sai lệch giữa giá trị quan sát thực tế Y_i và giá trị trung bình của chúng (TSS= Total Sum of Squares - còn gọi là tổng bình phương độ lệch của Y).

$$ESS = \sum \hat{y}_i^2 = \sum (\hat{Y}_i - \overline{Y})^2 = \beta_2^2 \sum x_i^2 \quad \textbf{[6.02]}$$

là tổng bình phương các độ sai lệch giữa giá trị của biến phụ thuộc Y tính theo hàm hồi quy mẫu với giá trị trung bình của chúng (ESS= Explained Sum of Squares- còn gọi là tổng bình phương độ lệch của Y được giải thích bởi SRF).

$$RSS = \sum \hat{u}_i^2 = \sum (Y_i - \hat{Y}_i)^2 \quad \textbf{[6.03]}$$

là tổng bình phương các độ sai lệch giữa giá trị quan sát Yi và giá trị tính toán (RSS= Residual Sum of Squares- còn gọi là tổng bình phương phần dư. RSS do các yếu tố ngẫu nhiên gây ra.)

6.1. Các tổng bình phương độ lệch

- Ta có hàm hồi quy mẫu SRF có dạng : $Y_i = \hat{Y}_i + \hat{u}_i$
- Phương trình trên có thể viết lại dưới dạng hàm phương sai là :

$$y_i = \hat{Y}_i + \hat{u}_i$$

$$\rightarrow \qquad \sum y_i^2 = \sum (\hat{y}_i + \hat{u}_i)^2 = \sum \hat{y}_i^2 + \sum \hat{u}_i^2 + 2\sum \hat{y}_i \hat{u}_i$$

- Dựa vào các tính chất của hàm SRF , ta lại có :
- $\sum_{i=1}^{n} \hat{y}_i \hat{u}_i = 0$ và $\hat{y}_i = \beta_2 x_i$

$$\rightarrow \qquad \sum y_i^2 = \sum \hat{y}_i^2 + \sum \hat{u}_i^2 = \beta_i^2 \sum x_i^2 + \sum \hat{u}_i^2$$

- Do đó: **TSS = ESS + RSS** **[6.04]**

- Ví dụ 6 : Tính các tổng bình phương độ lệch (TSS, ESS, RSS) của mô hình hồi quy trong ví dụ 1.

6.2. Hệ số xác định (đơn)

- Ta có **TSS = ESS + RSS** → $1 = \frac{ESS}{TSS} + \frac{RSS}{TSS} = \frac{\sum(\hat{Y}_i - \overline{Y})^2}{\sum(Y_i - \overline{Y})^2} + \frac{\sum \hat{u}_i^2}{\sum(Y_i - \overline{Y})^2}$
- Ta đặt $r^2 = \frac{\sum(\hat{Y}_i - \overline{Y})^2}{\sum(Y_i - \overline{Y})^2} = \frac{ESS}{TSS}$ hay $r^2 = 1 - \frac{\sum \hat{u}_i^2}{\sum(Y_i - \overline{Y})^2} = 1 - \frac{RSS}{TSS}$
- Từ $r^2 = \frac{ESS}{TSS} = \frac{\sum(\hat{Y}_i - \overline{Y})^2}{\sum(Y_i - \overline{Y})^2}$
- Ta có $r^2 = \frac{\sum \hat{y}_i^2}{\sum y^2} = \frac{\beta_2^2 \sum x_i^2}{\sum y_i^2} = \beta_2^2 \left(\frac{\sum x_i^2}{\sum y_i^2}\right)$ [6.05]
- Nếu chia cả tử và mẫu của phân số trên cho mẫu n (hoặc (n-1) nếu là mẫu nhỏ) thì ta sẽ được :

6.2. Hệ số xác định (đơn)

$$r^2 = \beta_2^2 \left[\frac{\sum x_i^2 / (n-1)}{\sum y_i^2 / (n-1)} \right] = \beta_2^2 \left(\frac{S_x^2}{S_y^2} \right) \qquad [6.06]$$

- trong đó, s_x^2 và s_y^2 là phương sai mẫu của X và Y.
- Từ công thức tính r^2 ta có thể thấy r^2 đo tỷ lệ hay số phần trăm của toàn bộ sai lệch của Y với giá trị trung bình của chúng được giải thích bằng mô hình (hay biến độc lập).

→ r^2 được gọi là hệ số xác định hay hệ số xác định đơn ((sample) coefficient of determination)

→ r^2 nằm trong đoạn [0,1]

→ r^2 càng tiệm cận 1 thì sai lệch của Y so với giá trị trung bình càng nhỏ, do đó, mô hình hồi quy (hay biến độc lập) có ý nghĩa càng lớn

6.2. Hệ số xác định (đơn)

- Mặt khác, theo tính chất của SRF ta lại có: $y_i = \hat{\beta}_2 x_i + \hat{u}_i$

$$\rightarrow \hat{\beta}_2 = \frac{y_i - \hat{u}_i}{x_i} = \frac{\sum x_i (y_i - \hat{u}_i)}{\sum x_i x_i} = \frac{\sum x_i y_i - \sum x_i \hat{u}_i}{\sum x_i^2}$$

- Do $\sum x_i \hat{u}_i = 0 \rightarrow \hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2}$ [2.28]

- Thay kết quả của [2.28] vào [2.26] ta được: $r^2 = \frac{(\sum x_i y_i)^2}{\sum x_i^2 \sum y_i^2}$ [6.07]

- Căn bậc hai cả hai vế của [2.29] ta được:

$$r = \pm \frac{\sum x_i y_i}{\sqrt{(\sum x_i^2)(\sum y_i^2)}} = \frac{n \sum X_i Y_i - (\sum X_i)(\sum Y_i)}{\sqrt{\left[n \sum X_i^2 - (\sum X_i)^2\right]\left[n \sum Y_i^2 - (\sum Y_i)^2\right]}} \quad [6.08]$$

→ CT tính của hệ số tương quan đơn (sample correlation coefficient).

- Ví dụ 7 : Tính hệ số xác định r^2 của mô hình hồi quy trong ví dụ 5.

6.3. Kiểm định sự phù hợp của mô hình hồi quy

- **6.3.1. Phân tích phương sai**
- **6.3.2. Kiểm định mô hình**
  - 6.3.2.1. Phương pháp giá trị tới hạn
  - 6.3.2.2. Phương pháp giá trị p-value

6.3.1. Phân tích phương sai

- Như ta đã biết:

$$\sum y_i^2 = \beta_i^2 \sum x_i^2 + \sum \hat{u}_i^2$$

$$\leftrightarrow \textbf{TSS = ESS + RSS}$$

- Từ góc độ của nghiên cứu hồi quy, việc nghiên cứu các thành phần ESS và RSS của tổng thể TSS được gọi chung là phân tích phương sai (analysis of variance- ANOVA).
- Ứng với mỗi giá trị tổng bình phương (Sum of Squares- SS) sẽ có một hệ số bậc tự do được xây dựng dựa trên cơ sở số lượng các quan sát. Chẳng hạn như:

6.3.1. Phân tích phương sai

- **TSS** $=\sum y_i^2$ có (n-1) bậc tự do vì ta đã mất một bậc tự do để tính giá trị của trung bình mẫu $\bar{Y}$.
- **ESS** $=\sum \hat{y}_i^2 = \hat{\beta}_2^2 \sum x_i^2$ có 1 bậc tự do vì:

$$\hat{\beta}_2 \sim N(\beta_2, \sigma^2_{\hat{\beta}_2})$$

$$\leftrightarrow \frac{(\hat{\beta}_2 - \beta_2)}{\sigma_{\hat{\beta}_2}} \sqrt{\sum x_i^2} \sim N(0,1)$$

$$\leftrightarrow \frac{(\hat{\beta}_2 - \beta_2)^2}{\sigma^2_{\hat{\beta}_2}} \sum x_i^2 \sim \chi^2(1)$$

- **RSS** $=\sum \hat{u}_i^2$ có (n-2) bậc tự do vì $\frac{(n-2)\hat{\sigma}^2}{\sigma^2} = \frac{\sum \hat{u}_i^2}{\sigma^2} \sim \chi^2$ (do $\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-2}$)

6.3.1. Phân tích phương sai

Bảng 6.02. Phân tích phương sai cho mô hình hồi quy hai biến

Nguồn biến thiên (1)	Tổng bình phương (SS) (2)	Bậc tự do (df) (3)	Phương sai (MSS)* (4) = (2)/(3)
Từ hàm hồi quy (ESS)	$\sum \hat{y}_i^2 = \beta_2^2 \sum x_i^2$	1	$\beta_2^2 \sum x_i^2$
Từ các yếu tố ngẫu nhiên (RSS)	$\sum \hat{u}_i^2$	(n-2)	$\frac{\sum \hat{u}_i^2}{n-2} = \sigma^2$
Tổng biến thiên (TSS)	$\sum y_i^2$	(n-1)	

*MSS= Mean sum of squares

6.3.1. Phân tích phương sai

- Bây giờ, dựa vào các dữ kiện ở bảng 2.10, ta có thể xây dựng biến F sao cho:

$$F = \frac{ESS/1}{RSS/(n-2)} = \frac{\hat{\beta}_2^2 \sum x_i^2 / 1}{\sum \hat{u}_i^2/(n-2)} = \frac{\hat{\beta}_2^2 \sum x_i^2}{\hat{\sigma}^2} \qquad [6.09]$$

Hay $$F = \frac{TSS.r^2/1}{(1-r^2)TSS/(n-2)} = \frac{r^2}{(1-r^2)} \cdot \frac{(n-2)}{1} \qquad [6.10]$$

→ Dễ dàng thấy F tuân theo quy luật phân phối Fisher :

$$F \sim F_{\alpha, (1, n-2)}$$

6.3.2. Kiểm định mô hình

- Như đã phân tích ở phần hệ số xác định r^2, để đánh giá mức độ thích hợp của mô hình hồi quy, nghĩa là mô hình hồi quy giải thích được bao nhiêu % sự thay đổi của biến phụ thuộc Y, thì ta sử dụng hệ số xác định r^2. Hệ số r^2 càng gần 1 bao nhiêu thì mô hình hồi quy càng có ý nghĩa bấy nhiêu.

6.3.2. Kiểm định mô hình

- Do đó, với kết quả hồi quy của mẫu cụ thể, người ta quan tâm đến việc đánh giá xem giá trị của r^2 khác 0 có ý nghĩa thống kê hay không. Nghĩa là ta tiến hành kiểm định giả thiết:

$$\begin{cases} H_0 : r^2 = 0 \\ H_1 : r^2 \neq 0 \end{cases}$$

- Đối với mô hình hồi quy hai biến, giả thiết H_0 còn có ý nghĩa là biến độc lập không ảnh hưởng đến biến phụ thuộc Y, hay nói cách khác, $\beta_2 = 0$. Vì vậy ta có giả thiết trên tương đương với giả thiết:

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

- Ta sẽ tiến hành kiểm định giả thiết này dựa vào giá trị của F được tính theo công thức [2.32].

6.3.2. Kiểm định mô hình

- Thông thường người ta tiến hành kiểm định H_0 bằng hai phương pháp sau:
- **6.3.2.1. Phương pháp giá trị tới hạn**
- **6.3.2.2. Phương pháp giá trị p-value**

6.3.2.1. Phương pháp giá trị tới hạn

- **Bước 1**: Tính $F_0 = \frac{r^2(n-2)}{(1-r^2)}$
- **Bước 2**: Tra bảng F với mức ý nghĩa α và hai bậc tự do (1, n-2) ta được giá trị tới hạn $F_{\alpha, (1, n-2)}$
- **Bước 3**: So sánh F0 và $F_{\alpha, (1, n-2)}$
  - Nếu $F_0 > F_{\alpha, (1, n-2):}$ bác bỏ H0
  - Nếu $F_0 < F_{\alpha, (1, n-2):}$ không có cơ sở để bác bỏ H_0

6.3.2.2. Phương pháp giá trị p-value

- **Bước 1**: Tính $F_0 = \frac{r^2(n-2)}{(1-r^2)}$
- **Bước 2**: Tính p-value = P(F > F_0) với F là phân phối Fisher có hai bậc tự do là (1, n-2)
- **Bước 3**: So sánh p-value và mức ý nghĩa α
  - Nếu p-value < α : bác bỏ H_0
  - Nếu p-value > α : không có cơ sở để bác bỏ H_0

6.3.2.2. Phương pháp giá trị p-value

Lưu ý:

- Thông thường giá trị p-value sẽ được tính sẵn trên các phần mềm KTL. Nhiệm vụ của chúng ta chỉ là đọc các kết quả này.
- Khi áp dụng phương pháp kiểm định mức ý nghĩa cho kiểm định t, F, χ^2, nếu không biết α, ta quyết định chấp nhận hay bác bỏ H_0 dựa vào mức ý nghĩa thực p-value và quy tắc kinh nghiệm là so sánh với 0,05.
- Ví dụ như đối với kiểm định hai phía của t, ta có quy tắc quyết định như sau :
  - p-value > 0,05 : chấp nhận H_0
  - p-value < 0,05 : có cơ sở bác bỏ H_0

6.3.2.2. Phương pháp giá trị p-value

- Ví dụ 8: Hãy kiểm định sự phù hợp của mô hình hồi quy trong ví dụ 1.