CHƯƠNG III. MÔ HÌNH HỒI QUY TUYẾN TÍNH ĐA BIẾN

Th.S. Vũ Thị Phương Mai- Khoa Kinh Tế Quốc Tế-
Đại Học Ngoại Thương- Hà Nội

Mô hình hồi quy tuyến tính đa biến

- Trong thực tế, các mối quan hệ kinh tế thường phức tạp, một số biến số kinh tế có thể chịu tác động của nhiều biến số kinh tế khác → mô hình hồi quy hai biến (hồi quy đơn) tỏ ra không thỏa đáng.
- VD khi n/c nhu cầu về một loại hàng hóa nào đó thì nhu cầu phụ thuộc đồng thời vào nhiều yếu tố: thu nhập của người tiêu dùng, giá bán của hàng hóa, thị hiếu người tiêu dùng…
- Vì vậy cần thiết phải mở rộng mô hình hồi quy hai biến bằng cách đưa thêm nhiều biến vào mô hình → n/c hồi quy nhiều biến (hồi quy bội hay hồi quy đa biến)
- Các ý tưởng và kết quả nghiên cứu của hồi quy hai biến được khái quát cho mô hình hồi quy nhiều biến.

Nội Dung

- 1. Thiết lập mô hình
- 2. Ước lượng các tham số
- 3. Khoảng tin cậy và kiểm định các giả thiết thống kê
- 4. Phân tích phương sai và kiểm định sự phù hợp của mô hình
- 5. Một số dạng mô hình quan trọng

1. Thiết lập mô hình

- Mô hình hồi quy k biến trình bày dưới dạng đại số như sau:

PRF $$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_{k-1} X_{k-1,i} + u_i \qquad [3.01]$$

SRF $$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \dots + \hat{\beta}_{k-1} X_{k-1,i} + \hat{u}_i \qquad [3.02]$$

- β_0: hệ số tự do, β_j (j=1,…,k-1): hệ số hồi quy riêng
- $\hat{\beta}_j$, $\hat{u}_i$ là các ước lượng điểm của β_j, u_i

Mô hình PRF dưới dạng ma trận

- Giả sử ta có n bộ giá trị quan sát của $(Y, X_1, X_2,\ldots,X_{k-1})$ là $(Y_i, X_{1i}, X_{2i},\ldots,X_{k-1,i})$ với $i = \overline{1, n}$.
- Như vậy hàm PRF ngẫu nhiên ứng với từng quan sát sẽ là:

$$Y_1 = \beta_0 + \beta_1 X_{11} + \beta_2 X_{21} + \ldots + \beta_{k-1} X_{k-1,1} + u_1$$

$$Y_2 = \beta_0 + \beta_1 X_{12} + \beta_2 X_{22} + \ldots + \beta_{k-1} X_{k-1,2} + u_2$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$Y_n = \beta_0 + \beta_1 X_{1n} + \beta_2 X_{2n} + \ldots + \beta_{k-1} X_{k-1,n} + u_n$$

Ta định nghĩa các ma trận tương ứng như sau :

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_n \end{pmatrix}_{n*1}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_{k-1} \end{pmatrix}_{k*1}, \quad u = \begin{pmatrix} u_1 \\ u_2 \\ \dots \\ u_n \end{pmatrix}_{n*1}, \quad X = \begin{pmatrix} 1 & X_{11} & X_{21} & \dots & X_{k-1,1} \\ 1 & X_{12} & X_{22} & \dots & X_{k-1,2} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & X_{1n} & X_{2n} & \dots & X_{k-1,n} \end{pmatrix}_{n*k}$$

- Khi đó PRF được viết dưới dạng ma trận như sau:

$$Y = X.\beta + u \qquad [3.03]$$

Mô hình SRF dưới dạng ma trận

- Kí hiệu:

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_{k-1} \end{pmatrix}_{k*1}, \quad \hat{u} = \begin{pmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \dots \\ \hat{u}_n \end{pmatrix}_{n*1}$$

- Ta có:

$$\text{SRF} \quad Y_i = X\beta + \hat{u} \qquad [3.04]$$

2. Ước lượng các tham số

- 2.1. Phương pháp bình phương nhỏ nhất OLS
- 2.2. Các giả thiết của PP OLS dưới dạng ma trận
- 2.3. Các tính chất của hệ số ước lượng OLS
- 2.4. Độ chính xác của các hệ số ước lượng OLS
- 2.5. Tiêu chuẩn của các hệ số ước lượng OLS

2.1. Phương pháp OLS

- Áp dụng phương pháp OLS, ta cần tìm β_j sao cho :

$$\sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} \left[Y_i - (\beta_0 + \beta_1 X_{1i} + ... + \beta_{k-1} X_{k-1}) \right]^2 \to \min$$

- Như đã được học trong chương II, ta biết rằng $\sum_{i=1}^{n} \hat{u}_i^2$ đạt cực tiểu

$$\leftrightarrow \begin{cases} f'(u) = 0 \\ f''(u) > 0 \end{cases}$$

- Ta sẽ lần lượt xét các điều kiện này bằng phương pháp ma trận.

Nhắc lại về lý thuyết ma trận

- Giả sử A là một ma trận. Khi đó, theo lý thuyết về ma trận ta có $A^T.A = A^2$ với A^T là ma trận chuyển vị của ma trận A.
- Ma trận chuyển vị là ma trận ở đó các hàng được thay thể bằng các cột và ngược lại. Ma trận chuyển vị của ma trận A được ký hiệu là A^T. Ví dụ :
- $A = \begin{bmatrix} X_1 & X_2 & X_3 & X_4 \end{bmatrix} \rightarrow A^T = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix}$

- Khi đó,

$$\sum_{i=1}^{n} \hat{u}_i^2 = \hat{u}^T.\hat{u}$$
$$= (Y - X\beta)^T.(Y - X\beta)$$
$$= (Y^T - \beta^T X^T).(Y - X\beta)$$
$$= Y^T Y - Y^T X\beta - \beta^T X^T Y + \beta^T X^T X\beta$$
$$= Y^T Y - 2\beta^T X^T Y + \beta^T X^T X\beta$$

(Do $Y^T X\beta = (Y^T X\beta)^T = \beta^T X^T Y$)

- Vậy, $\sum_{i=1}^{n} \hat{u}_i^2 = f(\beta) = Y^T Y - 2\beta^T X^T Y + \beta^T X^T X\beta$

- Ta cần tìm β sao cho f(β) → min, tương đương với tìm β sao cho :

$$\begin{cases} f'(\beta) = 0 \\ f''(\beta) > 0 \end{cases}$$

Xét điều kiện cần: $f'(\beta) = 0$

$$f'(\beta) = \frac{\partial f(\beta)}{\partial \beta} = -2X^T Y + 2X^T X\beta = 0 \rightarrow X^T X\beta = X^T Y$$

- Tương đương với:

$$\begin{pmatrix} n & \sum X_{1i} & \sum X_{2i} & ... & \sum X_{k-1,i} \\ \sum X_{1i} & \sum X_{1i}^2 & \sum X_{1i}X_{2i} & ... & \sum X_{1i}X_{k-1,i} \\ ... & ... & ... & ... & ... \\ \sum X_{k-1,i} & \sum X_{k-1,i}X_{1i} & \sum X_{k-1,i}X_{2i} & ... & \sum X_{k-1,i}^2 \end{pmatrix}_{k*k} \begin{pmatrix} \beta_0 \\ \beta_1 \\ ... \\ \beta_{k-1} \end{pmatrix}_{k*1} = \begin{pmatrix} 1 & 1 & ... & 1 \\ X_{11} & X_{12} & ... & X_{1n} \\ ... & ... & ... & ... \\ X_{k-1,1} & X_{k-1,2} & ... & X_{k-1,n} \end{pmatrix}_{k*n} \begin{pmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ ... \\ \hat{Y}_n \end{pmatrix}_{n*1} = \begin{pmatrix} \sum Y_i \\ \sum Y_i X_{1i} \\ ... \\ \sum Y_i X_{k-1,i} \end{pmatrix}_{k*n}$$

$X^T X \qquad \beta \qquad X^T \qquad Y \qquad X^T Y$

Xét điều kiện cần: $f'(\beta) = 0$

- Dễ thấy rằng X^TX là ma trận đối xứng qua đường chéo chính, do đó, về mặt thực hành ta chỉ cần xác định nửa trên hoặc nửa dưới của ma trận là đủ.
- Ta giả định là tồn tại ma trận nghịch đảo $(X^TX)^{-1}$ của ma trận X^TX, với điều kiện cần và đủ là ma trận X có hạng bằng k.

$$\rightarrow \beta = \frac{X^TY}{X^TX} = (X^TX)^{-1}.(X^TY)$$

Xét điều kiện đủ: $f''(\beta) > 0$

$$f''(\beta) = \frac{\partial f'(\beta)}{\partial \beta} = 2X^T X = 2X^2 > 0$$

2.2. Các giả thiết của PP OLS dưới dạng ma trận

- **Giả thiết 1** : Kỳ vọng có điều kiện của nhiễu bằng 0 : E(u)= 0. Trong đó :

$$E(u) = E\begin{pmatrix} u_1 \\ u_2 \\ \ldots \\ u_n \end{pmatrix} = \begin{pmatrix} E(u_1 / X_{11}, X_{21}, \ldots, X_{k-1,1}) \\ E(u_2 / X_{12}, X_{22}, \ldots, X_{k-1,2}) \\ \ldots \\ E(u_n / X_{1n}, X_{2n}, \ldots, X_{k-1,n}) \end{pmatrix}$$

2.2. Các giả thiết của PP OLS dưới dạng ma trận

- **<u>Giả thiết 2</u>** : E(u.u^T)= σ^2.I, trong đó I là ma trận đơn vị:

$$E(u.u^T) = E\left(\begin{bmatrix} u_1 \\ u_2 \\ \ldots \\ u_n \end{bmatrix} \begin{bmatrix} u_1 & u_2 & \ldots & u_n \end{bmatrix} \right) = E \begin{pmatrix} u_1^2 & u_1 u_2 & \ldots & u_1 u_n \\ u_2 u_1 & u_2^2 & \ldots & u_2 u_n \\ \ldots & \ldots & \ldots & \ldots \\ u_n u_1 & u_n u_2 & \ldots & u_n^2 \end{pmatrix}$$

$$= \begin{pmatrix} E(u_1^2) & E(u_1 u_2) & \ldots & E(u_1 u_n) \\ E(u_2 u_1) & E(u_2^2) & \ldots & E(u_2 u_n) \\ \ldots & \ldots & \ldots & \ldots \\ E(u_n u_1) & E(u_n u_2) & \ldots & E(u_n^2) \end{pmatrix}$$

2.2. Các giả thiết của PP OLS dưới dạng ma trận

- Giả thiết 2 có được do sử dụng giả thiết 1 kết hợp với các giả thiết sau :
  - Phương sai thuần nhất: $var(u_i) = E(u_i^2) = \sigma^2$, với mọi i.
  - Không có tự tương quan : $cov(u_i, u_j) = E(u_i, u_j) = 0$, với mọi $i \neq j$
- Khi đó,

$$E(u.u^T) = \begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma^2 \end{pmatrix} = \sigma^2 \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix} = \sigma^2 I$$

2.2. Các giả thiết của PP OLS dưới dạng ma trận

- **<u>Giả thiết 3</u>** : Ma trận X đã được xác định.
- **<u>Giả thiết 4</u>**: Hạng của ma trận X bằng k, là số tham số trong mô hình hồi quy.

Kí hiệu cho hạng của ma trận là rank(X) = k, có nghĩa là k cột trong ma trận X là độc lập tuyến tính, mà mỗi cột tương ứng với một biến độc lập nên không có tương quan tuyến tính chính xác giữa các biến độc lập, hay nói cách khác không có hiện tượng đa cộng tuyến (multicollinearity) xảy ra.

- **<u>Giả thiết 5</u>**: Véc tơ nhiễu có phân phối chuẩn nhiều chiều, tức là: $u \sim N(0, \sigma^2 I)$

2.3. Các tính chất của hệ số ước lượng OLS

- Với các giả thiết của mô hình, hàm hồi quy mẫu ƯL theo PP OLS có các tính chất tương tự như trong trường hợp hồi quy hai biến, bao gồm các tính chất sau:
  - SRF đi qua điểm ứng với các giá trị trung bình ($\overline{Y}$, $\overline{X}_1$,…, $\overline{X}_{k-1}$)
  - $\overline{Y} = \overline{\hat{Y}}$
  - $\overline{\hat{u}} = 0$
  - $\hat{u}_i$ không tương quan với X_{1i},…,$X_{k-1,i}$, tức là cov($\hat{u}$,X) = 0
  - $\hat{u}_i$ không tương quan với $\hat{Y}_i$, tức là cov($\hat{u}$, $\hat{Y}$) = 0

2.3. Các tính chất của hệ số ước lượng OLS

- Các hệ số ước lượng là thành phần của véc tơ $\hat{\beta}$ và là véc tơ ước lượng điểm của β, tìm bằng PP OLS, có các tính chất sau đây:
  - $\hat{\beta}$ được xác định một cách duy nhất với một mẫu cụ thể.
  - $\hat{\beta}$ là véc tơ ngẫu nhiên. Với các mẫu khác nhau, giá trị cụ thể của chúng sẽ khác nhau.
  - Với giả thiết u có phân phối chuẩn, véc tơ $\hat{\beta}$ tuân theo quy luật chuẩn.

2.4. Độ chính xác của các hệ số ước lượng OLS

- Để đo mức độ dao động và tương quan giữa các hệ số ước lượng được, ta sử dụng ma trận hiệp phương sai của hệ số hồi quy dạng tổng quát như sau:

$$\mathrm{cov}(\beta) = \begin{bmatrix} \mathrm{var}(\beta_0) & \mathrm{cov}(\beta_0, \beta_1) & \dots & \mathrm{cov}(\beta_0, \beta_{k-1}) \\ \mathrm{cov}(\beta_1, \beta_0) & \mathrm{var}(\beta_1) & \dots & \mathrm{cov}(\beta_1, \beta_{k-1}) \\ \dots\dots\dots\ \ \dots\dots\dots & \dots\dots\dots\ \ \dots\dots\dots & \dots & \dots\dots\dots\ \ \dots\dots\dots \\ \mathrm{cov}(\beta_{k-1}, \beta_0) & \mathrm{cov}(\beta_{k-1}, \beta_1) & \dots & \mathrm{var}(\beta_{k-1}) \end{bmatrix} = \sigma^2.(X^TX)^{-1} \quad [3.05]$$

- Vì σ^2 là phương sai tổng thể chưa biết, nên ta dùng σ^2 thay thế:

$$\sigma^2 = \frac{\sum_{i=1}^{n} \hat{u}_i^2}{n - k} \qquad [3.06]$$

2.5. Tiêu chuẩn của các hệ số ước lượng OLS

- Với các giả thiết của mô hình hồi quy tuyến tính cổ điển, $\hat{\beta}$ là ước lượng tuyến tính, không chệch, có phương sai nhỏ nhất trong lớp tất cả các ước lượng tuyến tính không chệch của β (tiêu chuẩn **BLUE**).
- Đây chính là nội dung của **định lý Gauss-Markov** mà ta đã biết trong chương II hồi quy hai biến.

3. Khoảng tin cậy và kiểm định các giả thiết thống kê

- 3.1. Khoảng tin cậy
- 3.2. Kiểm định giả thiết về hệ số hồi quy
- 3.3. Kiểm định giả thiết về phương sai của nhiễu
- nh
- 3.5. Kiểm định tổ hợp tuyến tính về hệ số hồi quy

3.1. Khoảng tin cậy

- Đối với mô hình hồi quy nhiều biến, việc xác định khoảng tin cậy cho các hệ số hồi quy và phương sai nhiễu của tổng thể được tiến hành tương tự như trong hồi quy hai biến. Với cỡ mẫu n và k tham số, cùng với giả thiết nhiễu u có phân phối chuẩn, khi sử dụng $\hat{\sigma}^2$ thay thế cho σ^2, ta có :

- $$t = \frac{\hat{\beta}_i - \beta_i}{se(\hat{\beta}_i)} \sim t_{(n-k)} \quad [3.07]$$

- $$\chi^2 = (n-k)\frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{(n-k)} \quad [3.08]$$

3.1. Khoảng tin cậy

- Với độ tin cậy (1-α), ta có công thức xác định khoảng tin cậy như sau :
  - Khoảng tin cậy của hệ số hồi quy :

$$\hat{\beta}_i - t_{(n-k),\alpha/2}\, se(\hat{\beta}_i) \le \beta_i \le \hat{\beta}_i + t_{(n-k),\alpha/2}\, se(\hat{\beta}_i) \quad [3.09]$$

  - Khoảng tin cậy của phương sai nhiễu :

$$(n-k)\frac{\hat{\sigma}^2}{\chi^2_{(n-k),\alpha/2}} \le \sigma^2 \le (n-k)\frac{\hat{\sigma}^2}{\chi^2_{(n-k),1-\alpha/2}} \quad [3.10]$$

3.2. Kiểm định giả thiết về hệ số hồi quy

- PP tiến hành kiểm định giả thiết về hệ số hồi quy trong mô hình hồi quy bội hoàn toàn giống hoàn toàn như trong mô hình hai biến. Điều khác biệt ở đây cần chú ý đó là bậc tự do. Đối với mô hình hai biến, có hai tham số, bậc tự do tương ứng của thống kê t là (n-2), trong khi đối với mô hình hồi quy bội tổng quát có k tham số, bậc tự do tương ứng của thống kê t là (n-k).
- Nhắc lại rằng thống kê t_{qs} xác định bằng biểu thức :

$$t_{qs} = \frac{\beta_i - \beta_i}{se(\beta_i)}$$

- Và giá trị: p-value = P (|t| > |tqs|)

Bảng 3.01. Kiểm định giả thiết về hệ số hồi quy

Loại giả thiết	Giả thiết H_0	Giả thiết đối H_1	Phương pháp	Miền bác bỏ H_0
Hai phía	$\beta_i = \beta_i^*$	$\beta_i \neq \beta_i^*$	Khoảng tin cậy	$\beta_i^* \notin [\beta_i - t_{(n-k),\alpha/2}\, se(\beta_i),\ \beta_i + t_{(n-k),\alpha/2}\, se(\beta_i)]$
			Giá trị tới hạn	$\lvert t_{qs}\rvert > t_{(n-k),\alpha/2}$
			Giá trị *p-value*	*p-value* $< \alpha$
Phía phải	$\beta_i \leq \beta_i^*$	$\beta_i > \beta_i^*$	Khoảng tin cậy	$\beta_i^* \notin [\beta_i - t_{(n-k),\alpha}\, se(\beta_i), +\infty\ ]$
			Giá trị tới hạn	$t_{qs} > t_{(n-k),\alpha}$
			Giá trị *p-value*	*p-value/2* $< \alpha$
Phía trái	$\beta_i \geq \beta_i^*$	$\beta_i < \beta_i^*$	Khoảng tin cậy	$\beta_i^* \notin [-\infty,\ \beta_i + t_{(n-k),\alpha}\, se(\beta_i)]$
			Giá trị tới hạn	$t_{qs} < -\ t_{(n-k),\alpha}$
			Giá trị *p-value*	*p-value/2* $< \alpha$

3.3. Kiểm định giả thiết về phương sai của nhiễu

- Bảng dưới đây tóm tắt các cách kiểm định giả thiết về phương sai của nhiễu cùng với quy tắc tương ứng để bác bỏ giả thiết H_0, với :

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

- Và

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

Bảng 3.02. Kiểm định giả thiết 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	$\sigma_0^2 \notin \left[(n-k)\dfrac{\sigma^2}{\chi^2_{(n-k),\alpha/2}}, (n-k)\dfrac{\sigma^2}{\chi^2_{(n-k),1-\alpha/2}}\right]$
			Giá trị tới hạn	$\chi_0^2 > \chi^2_{(n-k),\alpha/2}$ **hoặc** $\chi_0^2 < \chi^2_{(n-k),1-\alpha/2}$
			Giá trị *p-value*	*p-value* < α/2 **hoặc** *p-value* > 1- α/2
Phía phải	$\sigma^2 = \sigma_0^2$	$\sigma^2 > \sigma_0^2$	Khoảng tin cậy	$\sigma_0^2 \notin \left[(n-k)\dfrac{\sigma^2}{\chi^2_{(n-k),\alpha}}, +\infty\right]$
			Giá trị tới hạn	$\chi_0^2 > \chi^2_{(n-k),\alpha}$
			Giá trị *p-value*	*p-value* < α
Phía trái	$\sigma^2 = \sigma_0^2$	$\sigma^2 < \sigma_0^2$	Khoảng tin cậy	$\sigma_0^2 \notin \left[-\infty, (n-k)\dfrac{\sigma^2}{\chi^2_{(n-k),1-\alpha}}\right]$
			Giá trị tới hạn	$\chi_0^2 < \chi^2_{(n-k),1-\alpha}$
			Giá trị *p-value*	*p-value* > 1- α

3.4. Kiểm định tính có ý nghĩa của mô hình

- Xét hai mô hình sau :
- (U) : Y= $\beta_0 + \beta_1 X_1 + \ldots + \beta_{m-1} X_{m-1} + \beta_m X_m + \ldots + \beta_{k-1} X_{k-1} + u$
- (R) : Y = $\beta_0 + \beta_1 X_1 + \ldots + \beta_{m-1} X_{m-1} + v$
- (U) gọi là mô hình không bị ràng buộc (Unrestricted model)
- (R) gọi là mô hình bị ràng buộc (Restricted model).

3.4. Kiểm định tính có ý nghĩa của mô hình

- Điều kiện ràng buộc trong mô hình (R) chính là hệ số hồi quy của các biến độc lập X_{m-1}, $X_{m+1}, \ldots, X_{k-1}$ đồng thời bằng 0.
- Để kiểm định điều kiện ràng buộc trên, ta xây dựng giả thiết :

$$H_0 : \beta_m = \ldots = \beta_{k-1} = 0$$

$$H_1 : \text{có ít nhất một } \beta_j \neq 0$$

- a a mô nh tiến hành qua các bước:
  - **<u>Bước 1</u>** : Hồi quy (U) gồm k tham số, tính RSS_U, (n-k) bậc tự do.
  - **<u>Bước 2</u>** : Hồi quy (R) gồm m tham số, tính RSS_R, (n-m) bậc tự do.
  - **<u>Bước 3</u>** : Sử dụng thống kê F như sau :

$$F_w = \frac{(RSS_R - RSS_U)/(k-m)}{RSS_U/(n-k)} = \frac{(R_U^2 - R_R^2)/(k-m)}{(1-R_U^2)/(n-k)} \sim F_{(k-m),(n-k)}$$

3.4. Kiểm định tính có ý nghĩa của mô hình

- Với mức ý nghĩa α, tra bảng giá trị tới hạn $F_{\alpha, (k-m), (n-k)}$

  → Nếu $\mathbf{F_w} > \mathbf{F_{\alpha,(k-m),(n-k)}}$ thì bác bỏ H_0
- Thông

  a F_w

  F-stat.
- Đi ng cho ra p-value a F-stat, và người sử dụng có thể áp dụng quy tắc quyết định dựa trên giá trị tới hạn hay mức ý nghĩa để bác bỏ hay chấp nhận H_0.
- Ngoài ra, cũng lưu ý rằng, nếu giả thiết là $H_0 : \beta_j = 0$ thì kết luận của kiểm định Wald tương đương với kết luận kiểm định t.

3.5. Kiểm định tổ hợp tuyến tính về hệ số hồi quy

- Phần này trình bày cách thức kiểm định giả thiết về sự liên hệ mang đặc tính tuyến tính giữa các hệ số hồi quy. Xét tình huống cụ thể như sau :
- Giả sử ta có mô hình hồi quy (U) : Y= $\beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$. Ta muốn kiểm định giả thiết :

$$\begin{cases} H_0 : \beta_1 = \beta_2 \\ H_1 : \beta_1 \neq \beta_2 \end{cases}$$

- Giả thiết $\beta_1 = \beta_2$ có thể biến đổi về dạng $\beta_1 - \beta_2 = 0$ → Dạng tương đương là cách biểu diễn tổ hợp tuyến tính của hai hệ số hồi quy.
- Ta có thể áp dụng ba phương pháp để kiểm định.

3.5. Kiểm định tổ hợp tuyến tính về hệ số hồi quy

Cách 1 : Kiểm định Wald

- Biến đổi mô hình bằng cách sử dụng giả thiết H_0, ta được :

$$Y = \beta_0 + \beta_1 X_1 + \beta_1 X_2 + u = \beta_0 + \beta_1 (X_1 + X_2) + u$$

- Đặt $Z = X_1 + X_2$, ta có mô hình (R) : $Y = \beta_0 + \beta_1 Z + u$
- Áp dụng kiểm định Wald :

$$F_w = \frac{(R_U^2 - R_R^2)/(3-2)}{(1 - R_U^2)/(n-3)} \sim F_{(1),(n-3)}$$

- Nếu $F_w > F_{\alpha, 1, (n-3)}$ thì bác bỏ H_0.

3.5. Kiểm định tổ hợp tuyến tính về hệ số hồi quy

ch p

- t $\delta = \beta_1 - \beta_2$, t H_0 nh: $H_0 : \beta_1 - \beta_2 \leftrightarrow H_0 : \delta = 0$
- Thay $\beta_2 = \delta - \beta_1$ c :

  $Y = \beta_0 + \beta_1 X_1 + (\delta - \beta_1) X_2 + u = \beta_0 + \beta_1 (X_1 + X_2) - \delta X_2 + u$
- t $Z = X_1 + X_2$ nh (R) : $Y = \beta_0 + \beta_1 Z - \delta X_2 + u$
- nh GT: $\mathbf{H_0: \delta = 0}$

  t :

$$t = \frac{\hat{\delta} - 0}{se(\hat{\delta})} \sim t_{(n-3)}$$

- trong a δ. u $|t| > t\alpha/2$, (-

  H_0.

3.5. Kiểm định tổ hợp tuyến tính về hệ số hồi quy

ch p

$$t = \frac{\hat{\delta} - 0}{se(\hat{\delta})} = \frac{\hat{\beta}_1 - \hat{\beta}_2}{se(\hat{\beta}_1 - \hat{\beta}_2)}$$

- i $se(\hat{\beta}_1 - \hat{\beta}_2) = \sqrt{\mathrm{var}(\hat{\beta}_1 - \hat{\beta}_2)}$
- $\sqrt{\mathrm{var}(\hat{\beta}_1 - \hat{\beta}_2)} = \mathrm{var}(\hat{\beta}_1) + \mathrm{var}(\hat{\beta}_2) - 2\mathrm{cov}(\hat{\beta}_1, \hat{\beta}_2)$
- i quy $\hat{\beta}_i$ c 3.05).

4. Phân tích phương sai và kiểm định sự phù hợp của mô hình

- ch
- 4.2. Hệ số xác định bội và hệ số tương quan bội
- 4.3. Kiểm định sự phù hợp của mô hình
- 4.4. Bảng phân tích phương sai (ANOVA)

4.1. Các tổng bình phương độ lệch

- ch như sau :

$$TSS = \sum (Y_i - \overline{Y})^2 = \sum Y_i^2 - n(\overline{Y})^2 = Y^T Y - n(\overline{Y})^2 \qquad [3.11]$$

$$ESS = \sum (\hat{Y}_i - \overline{Y})^2 = \beta^T (X^T Y) - n(\overline{Y})^2 \qquad [3.12]$$

$$RSS = \sum (Y_i - \hat{Y}_i)^2 = \sum \hat{u}_i^2 = TSS - ESS \qquad [3.13]$$

- Y

n.

4.2. Hệ số xác định bội và hệ số tương quan bội

- 4.2.1. Hệ số xác định bội
- 4.2.2. Hệ số xác định bội đã hiệu chỉnh (adjusted R-squared)
- 4.2.3. Hệ số tương quan bội (coefficient of correlation)
- 4.2.4. Hệ số tương quan riêng phần (partial correlation coefficients)

4.2.1. Hệ số xác định bội

- n, r^2

ch X_2, $X_3,\ldots,X_k$ u R^2.

- nh như sau :

$$R^2 = 1 - \frac{RSS}{TSS} = \frac{ESS}{TSS} \quad [3.14]$$

- Y a R^2 a R^2 i.

4.2.1. Hệ số xác định bội

- ng R^2 :
  - TSS = $\sum (Y_i - \bar{Y})^2$ không

    (n-1).
  - i c:

$$RSS = \sum (Y_i - \hat{Y}_i)^2 = \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - .... - \hat{\beta}_{k-1} X_{k-1,i})^2$$

- nh.

- y R^2

R^2

p.

4.2.2. Hệ số xác định bội đã hiệu chỉnh

-

c:

$$\bar{R}^2 = 1 - \frac{RSS/(n-k)}{TSS/(n-1)} = 1 - (1-R^2)\frac{n-1}{n-k} = R^2 + (1-R^2)\frac{1-k}{n-k}$$

- $\bar{R}^2$ c t sau :
  - Khi $\bar{R}^2 \leq R^2 \leq 1$
  - k n $\bar{R}^2$ hơn R^2
  - $\bar{R}^2$ c $\bar{R}^2 = 0$.

4.2.2. Hệ số xác định bội đã hiệu chỉnh

- ng $\overline{R}^2$ cho R^2
  nh không.
- $\overline{R}^2$
  ng kê.

4.2.3. Hệ số tương quan bội

-

:

- X_j :

$$r_{0j} = \frac{\sum y_i x_{ji}}{\sqrt{\sum y_i^2 \sum x_{ji}^2}} \quad [3.15]$$

- :

$$r_{tj} = \frac{\sum x_{ti} x_{ji}}{\sqrt{\sum x_{ti}^2 \sum x_{ji}^2}} \quad [3.16]$$

- i: $y_i = Y_i - \overline{Y}$; $x_{ji} = X_{ji} - \overline{X}_j$

4.2.3. Hệ số tương quan bội

- ng:

$$R = \begin{bmatrix} 1 & r_{01} & \dots & r_{0,k-1} \\ r_{10} & 1 & \dots & r_{1,k-1} \\ \dots & \dots & \dots & \dots \\ r_{k-1,0} & r_{k-1,1} & \dots & 1 \end{bmatrix}$$

-

$r_{tj} = r_{jt}$

nên $r_{jj} = 1$.

4.2.4. Hệ số tương quan riêng phần

- i.
- i p>0.
- p X_j

  -

  c (k-2).

4.2.4. Hệ số tương quan riêng phần

- n:

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

- ng a:
- $r_{12,3}$ X_2 trong khi X_3 i.
- $r_{13,2}$ X_3 trong khi X_2 i.
- $r_{23,1}$ n X_2 X_3 i.

4.2.4. Hệ số tương quan riêng phần

- ng:

$$r_{12,3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}}; \quad r_{13,2} = \frac{r_{13} - r_{12}r_{23}}{\sqrt{(1 - r_{12}^2)(1 - r_{23}^2)}}; \quad r_{3,2} = \frac{r_{13} - r_{12}r_{23}}{\sqrt{(1 - r_{12}^2)(1 - r_{23}^2)}}$$

- $r_{12,34}$ n r_{12}, r_{13}

  c không.

- sau:

$$R^2 = \frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2}; \quad R^2 = r_{12}^2 + (1 - r_{12}^2)r_{13,2}^2 \quad R^2 = r_{13}^2 + (1 - r_{13}^2)r_{12,3}^2$$

- TQ p 0.

4.3. Kiểm định sự phù hợp của mô hình

- t như sau:

$$\mathbf{H_0} : R^2 = 0 \leftrightarrow \mathbf{H_0} : \beta_1 = \beta_2 = \ldots = \beta_{k-1} = 0$$

$$\mathbf{H_1} : R^2 \neq 0 \leftrightarrow \mathbf{H_1} \quad t\ \beta_i \neq 0$$

- p.

4.3. Kiểm định sự phù hợp của mô hình

- t $\mathbf{H_0 : \beta_1 = \beta_2 = \ldots = \beta_{k-1} = 0}$

  :

$$F = \frac{ESS/(k-1)}{RSS/(n-k)} = \frac{R^2}{(1-R^2)} \cdot \frac{(n-k)}{(k-1)} \sim F_{(k-1,\, n-k)}$$

$$F_0 = \frac{R^2}{(1-R^2)} \cdot \frac{(n-k)}{(k-1)}$$

- Ta ng PP QĐ.
- Nguyên c QĐ: u $F_0 > F_{\alpha,\,(k-1;\, n-k)}$ c p-value = $P(F > F_0) < \alpha$ H_0.
-

  a tương đương nhau.
- n Y= $\beta_0 + \beta_1 X_1 + \beta_1 X_2 + u$
- t $\mathbf{H_0: \beta_1 = 0}$ $\mathbf{H_0: \beta_2 = 0}$ không tương đương i

  $\mathbf{H_0: \beta_1 = \beta_2 = 0}$

4.4. Bảng phân tích phương sai (ANOVA)

-

ch phương sai ANOVA (Analysis of variance).

ch phương sai ANOVA

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)	$\beta^T(X^TY) - n(\overline{Y})^2$	k - 1	ESS/(k-1)
Từ các yếu tố ngẫu nhiên (RSS)	$Y^TY - \beta^T(X^TY)$	(n-k)	RSS/(n-k)
Tổng biến thiên (TSS)	$Y^TY - n(\overline{Y})^2$	(n-1)	TSS/(n-1)

t

$Y_1,\ldots,Y_n$

c Y_i nh $\overline{Y} = \sum Y_i / n$

a $TSS = \sum (Y_i - \overline{Y})^2 = Y^TY - n(\overline{Y})^2$ (n-1).

4.4. Bảng phân tích phương sai (ANOVA)

- :
- t $Y_1,\ldots,Y_n$
  c Y_i
  nh $\overline{Y} = \sum Y_i / n$. (n-1),
  với TSS = $\sum (Y_i - \overline{Y})^2 = Y^T Y - n(\overline{Y})^2$
- i RSS = $\sum (Y_i - \hat{Y}_i)^2 = Y^T Y - \beta^T (X^T Y)$
  u nhiên Y_i : $\hat{Y}_i = \sum \beta_0 + \beta_1 X_{1i} + \ldots + \beta_{k-1} X_{k-1,i}$

  (n-k).
- -
  RSS.

4.4. Bảng phân tích phương sai (ANOVA)

-

.

-

c.

-

o MSS.

5. Một số dạng mô hình quan trọng

-

m.

-

không còn

i.

5. Một số dạng mô hình quan trọng

- p (double- log model)
- n logarit (semi-log model)
- o (hyperbol) (reciprocal model)