MÔ HÌNH
HỒI QUY ĐA BIẾN

ThS Nguyễn Thị Kim Dung

Ví dụ:

Thu nhập ảnh hưởng đến chi tiêu. (+)

Địa điểm sinh sống ảnh hưởng đến chi tiêu.

Số thành viên gia đình ảnh hưởng đến chi tiêu. (+)

…

Vậy:

Chi tiêu ←Thu nhập, Địa điểm, Số thành viên

Chi tiêu = *f* (Thu nhập, Địa điểm, Số thành viên)

1. MÔ HÌNH HỒI QUY

1.1 MÔ HÌNH HỒI QUY TỔNG THỂ

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + ... + \beta_k X_{ki} + u_i \qquad (1.1)$$

$$\text{Với:} \begin{cases} \beta_1 : \text{hệ số tự do} \\ \beta_2, \beta_3, ..., \beta_k : \text{các hệ số hồi quy riêng} \\ u_i \quad (i = 2, ..., k) : \text{sai số ngẫu nhiên} \end{cases}$$

$$\text{Đặt: } X_i' = \left(1, X_{2i}, X_{3i}, ..., X_{ki}\right); \beta = \begin{pmatrix} \beta_1 \\ ... \\ \beta_k \end{pmatrix}$$

$$\Rightarrow Y_i = X_i' \beta + u_i$$

1. MÔ HÌNH HỒI QUY
1.1 MÔ HÌNH HỒI QUY TỔNG THỂ

Nếu ta có n quan sát, mỗi quan sát gồm k giá trị $(Y_i, X_{2i}, X_{3i}, \ldots, X_{ki})$, $(i = 1, \ldots n)$, thì ta có hệ n phương trình:

$$\begin{cases} Y_1 = \beta_1 + \beta_2 X_{21} + \beta_3 X_{31} + \ldots + \beta_k X_{k1} + u_1 \\ Y_2 = \beta_1 + \beta_2 X_{22} + \beta_3 X_{32} + \ldots + \beta_k X_{k2} + u_2 \\ \ldots \\ Y_n = \beta_1 + \beta_2 X_{2n} + \beta_3 X_{3n} + \ldots + \beta_k X_{kn} + u_n \end{cases} \quad (1.2)$$

1. MÔ HÌNH HỒI QUY
1.1 MÔ HÌNH HỒI QUY TỔNG THỂ

$$\text{Đặt } Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \ldots \\ Y_n \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \ldots \\ \beta_k \end{pmatrix},$$

$$X = \begin{pmatrix} 1 & X_{21} & X_{31} & \ldots & X_{k1} \\ 1 & X_{22} & X_{32} & \ldots & X_{k2} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & X_{2n} & X_{3n} & \ldots & X_{kn} \end{pmatrix}, \quad u = \begin{pmatrix} u_1 \\ u_2 \\ \ldots \\ u_n \end{pmatrix}$$

$\Rightarrow$ Dạng ma trận của phương trình (1.2):

$$Y = X\beta + u$$

1. MÔ HÌNH HỒI QUY
1.2 MÔ HÌNH HỒI QUY MẪU

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

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \ldots + \hat{\beta}_k X_{ki} + e_i$$

$e_i = Y_i - \hat{Y}_i$ gọi là phần dư

1. MÔ HÌNH HỒI QUY
1.2 MÔ HÌNH HỒI QUY MẪU

Nếu ta có n quan sát, mỗi quan sát gồm k giá trị $(Y_i, X_{2i}, X_{3i}, \ldots, X_{ki}), (i = 1, \ldots n)$. Ta đặt:

$$\hat{Y} = \begin{pmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \ldots \\ \hat{Y}_n \end{pmatrix}, \hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \ldots \\ \hat{\beta}_k \end{pmatrix}, X = \begin{pmatrix} 1 & X_{21} & X_{31} & \ldots & X_{k1} \\ 1 & X_{22} & X_{32} & \ldots & X_{k2} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & X_{2n} & X_{3n} & \ldots & X_{kn} \end{pmatrix}, e = \begin{pmatrix} e_1 \\ e_2 \\ \ldots \\ e_n \end{pmatrix}$$

⇨Dạng ma trận của phương trình: $\hat{Y} = X\hat{\beta}$

⇨Dạng ma trận của mô hình: $Y = X\hat{\beta} + e$

2. PHƯƠNG PHÁP BÌNH PHƯƠNG NHỎ NHẤT (OLS)

2.1. Nội dung phương pháp bình phương nhỏ nhất

- Ta có mô hình hồi quy mẫu là

$$Y = X\hat{\beta} + e$$

- Ta cần tìm $\hat{Y}_i$ sao cho nó gần với giá trị thực Y_i nhất, tức là phần dư $e_i = Y_i - Y_i$ càng nhỏ càng tốt

Tìm $\hat{Y}_i$ sao cho $\sum_{i=1}^{n} e_i^2 \to \min$

$$\Leftrightarrow \sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 = \sum_{i=1}^{n}\left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \ldots - \hat{\beta}_k X_{ki}\right)^2 \to \min$$

$$\Leftrightarrow \varphi\left(\hat{\beta}_j\right) \to \min \quad (j = 1, \ldots, k)$$

Nghĩa là cần tìm $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k$ sao cho $\varphi \to \min$

$\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k$ là nghiệm của hệ sau:

$$\begin{cases} \varphi'_{\hat{\beta}_j} = 0 \\ j = 1, 2, \ldots, k \end{cases}$$

Theo dạng ma trận, ta có:

$$e = Y - Y = \left(Y - X\beta\right)$$

$$\Rightarrow \varphi = \sum_{i=1}^{n} e_i^{\,2} = \left(e_1\ e_2\ \ldots e_n\right).\begin{pmatrix} e_1 \\ e_2 \\ \ldots \\ e_n \end{pmatrix}$$

$$= e^T e$$

$$= \left(Y - X\beta\right)^T \left(Y - X\beta\right)$$

Nhớ lại tính chất ma trận:

$$\left(A+B\right)^T = A^T + B^T$$

$$\left(A.B\right)^T = B^T A^T$$

$$\Rightarrow \varphi = \left(Y - X\beta\right)^T \left(Y - X\beta\right)$$

$$= \left(Y^T - \beta^T X^T\right)\left(Y - X\beta\right)$$

$$= Y^T Y - Y^T X\beta - \beta^T X^T Y + \beta^T X^T X\beta$$

Nhận xét:

$$Y^T X\beta = \left(Y^T\right)_{1\times n}\left(X\right)_{n\times k}\left(\beta\right)_{k\times 1} = \left(\quad\right)_{1\times 1}$$

$$\beta^T X^T Y = \left(\beta^T\right)_{1\times k}\left(X^T\right)_{k\times n}\left(Y\right)_{n\times 1} = \left(\quad\right)_{1\times 1}$$

Mà: $\left(Y^T X\beta\right)^T = \beta^T X^T Y$

$$\Rightarrow Y^T X\beta = \beta^T X^T Y$$

$$\Rightarrow \varphi = Y^T Y - 2\beta^T X^T Y + \beta^T X^T X\beta$$

$$\varphi \to \min$$

$$\varphi'_{\beta} = 0$$

$$\Leftrightarrow -2X^T Y + 2X^T X\beta = 0$$

$$\Leftrightarrow X^T Y = X^T X\beta$$

$$\Leftrightarrow \boxed{\beta = \left(X^T X\right)^{-1}\left(X^T Y\right)}$$

Vì Y=AX $\Rightarrow$ X=A^{-1} Y

2.2. Các tính chất của hàm hồi quy mẫu (SRF) tìm được bằng phương pháp OLS

- Các ước lượng của hồi quy đa biến có đầy đủ các tính chất của ước lượng hồi quy đơn biến

3. ĐO LƯỜNG MỨC ĐỘ PHÙ HỢP CỦA ƯỚC LƯỢNG THEO OLS

3.1. Hệ số xác định

$$Y_i = \hat{Y}_i + e_i$$

$$\Leftrightarrow Y_i - \overline{Y} = \hat{Y}_i - \overline{Y} + e_i$$

$$\Rightarrow \underbrace{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2}_{\text{Tổng dao động của Y so với giá trị trung bình}} = \underbrace{\sum_{i=1}^{n}\left(\hat{Y}_i - \overline{Y}\right)^2}_{\text{Dao động được giải thích bởi mô hình}} + \underbrace{\sum_{i=1}^{n} e_i^2}_{\text{Dao động chưa được giải thích bởi mô hình – sai số}}$$

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

TSS= Total sum of square

$$\text{ESS} = \sum\left(Y_i - \overline{Y}\right)^2 = \sum\left(\left(X_{1i}\ X_{2i}\ \ldots\ X_{ki}\right) - \overline{X}\right)^2 \beta^2$$

= Explained sum of square

$$\text{RSS} = \sum_{i=1}^{n}\left(e_i\right)^2 = \text{Residual sum of square}$$

Đặt $R^2 = \dfrac{ESS}{TSS} = 1 - \dfrac{RSS}{TSS}$ $\qquad 0 \le R^2 \le 1$

Nhận xét: $$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

Khi thêm biến vào mô hình thì k tăng.
TSS không phụ thuộc *k* nên không đổi,
ESS phụ thuộc *k* nên tăng $\Rightarrow$RSS giảm → R^2 tăng
Vậy cứ thêm biến vào mô hình thì R^2 tăng, do đó không thể dùng R^2 để xem xét việc có nên đưa thêm biến vào mô hình không.

3.2. Hệ số xác định hiệu chỉnh

Đặt $\bar{R}^2 = 1 - \dfrac{RSS/(n-k)}{TSS/(n-1)} = 1 - \dfrac{n-1}{n-k}\left(1-R^2\right)$

Nhận xét:

Khi thêm biến giải thích vào mô hình thì k tăng

- TSS và (n – 1) không bị ảnh hưởng bởi *k*
- (n – k) giảm
- Khi thêm biến có ý nghĩa vào mô hình thì RSS (sai số) giảm
- Khi thêm biến không có ý nghĩa vào mô hình thì RSS (sai số) không giảm hoặc giảm ít

- Vậy khi thêm biến vào mô hình, nếu biến này có ý nghĩa thì $\overline{R}^2$ tăng, ngược lại, $\overline{R}^2$ không tăng. Do đó ta chỉ thêm biến vào mô hình khi nào $\overline{R}^2$ còn tăng

- $\overline{R}^2 < R^2, \overline{R}^2$ tăng chậm hơn R^2

- Nếu R^2 đủ nhỏ, $\overline{R}^2$ có thể mang giá trị âm

Ý nghĩa thực hành của hệ số xác định hiệu chỉnh

Khi ta thêm càng nhiều biến vào mô hình thì R^2 tăng → ta sẽ đưa quá nhiều biến vào mô hình (kể cả các biến không cần thiết).

Để tránh hiện tượng này, ta dùng hệ số xác định hiệu chỉnh, vì khi thêm biến vào mô hình, $\overline{R}^2$ có thể tăng hoặc không tăng.

Vậy $\overline{R}^2$ được dùng để xác định xem có nên thêm 1 biến mới vào mô hình hay không.

(Với cùng 1 bộ số liệu, khi thêm 1 biến mới vào mô hình thì mô hình nào có hệ số $\overline{R}^2$ lớn hơn được xem là tốt hơn)

4. BẢN CHẤT THỐNG KÊ CỦA MÔ HÌNH HỒI QUY ĐA BIẾN

- Tương tự hồi quy đơn biến, ta có:

$$\hat{\beta}_k = \beta_k + \sum_{i=1}^{n} C_{ki} U_i$$

$\hat{\beta}_k$ phụ thuộc vào **yếu tố ngẫu nhiên U_i** , nên
$\hat{\beta}_k$ cũng là 1 **yếu tố ngẫu nhiên**

4.1. GIẢ ĐỊNH CỦA CÁC YẾU TỐ NGẪU NHIÊN

Định lý Gauss-Markov

- Với các giả thiết sau đây thì các ước lượng tìm được bằng PP OLS sẽ là các ước lượng tuyến tính, không chệch, có phương sai nhỏ nhất.

Giả thiết A1: $E(U_i) = 0 \quad \forall i$

Giả thiết A2: $\mathrm{Var}(U_i) = \sigma^2 \quad \forall i$

Giả thiết A3: $U_i \overset{iid}{\sim} \mathrm{N}(0, \sigma^2) \quad \forall i$

Giả thiết A4: $E(Y_i / X'_i) = X'_i\beta \ \forall i$

Giả thiết A5: $(X_2, X_3, \ldots, X_k)$ độc lập tuyến tính

4.2 ĐẶC TRƯNG THỐNG KÊ CỦA ƯỚC LƯỢNG OLS

$$\boxed{E\left(\hat{\beta}_k\right) = \beta_k}$$

$$\text{Ta có}: \hat{\beta}_k = \beta_k + \sum_{i=1}^{n} C_{ki} U_i$$

$$E\left(\hat{\beta}_k\right) = E\left(\beta_k + \sum_{i=1}^{n} C_{ki} U_i\right) = E\left(\beta_k\right) + E\left(\sum_{i=1}^{n} C_{ki} U_i\right)$$

$$= \beta_k + \sum_{i=1}^{n} C_{ki} \underbrace{E\left(U_i\right)}_{=0(A1)} = \beta_k$$

$$\operatorname{Var}\left(\hat{\beta}_k\right) = \frac{\sigma^2}{S_{kk}}$$

$$\operatorname{Var}\left(\hat{\beta}_k\right) = \operatorname{Var}\left(\hat{\beta}_k - E\hat{\beta}_k\right) = \operatorname{Var}\left(\hat{\beta}_k - \beta_k\right)$$

$$= \operatorname{Var}\left(\sum_{i=1}^{n} C_{ki} U_i\right) = \sum_{i=1}^{n} C_{ki}{}^2 \operatorname{Var}\left(U_i\right) \quad \left(\mathrm{A}_3\right)$$

$$= \sigma^2 \sum_{i=1}^{n} C_{ki}{}^2 \quad \left(\mathrm{A}_2\right)$$

$$= \frac{\sigma^2}{S_{kk}}$$

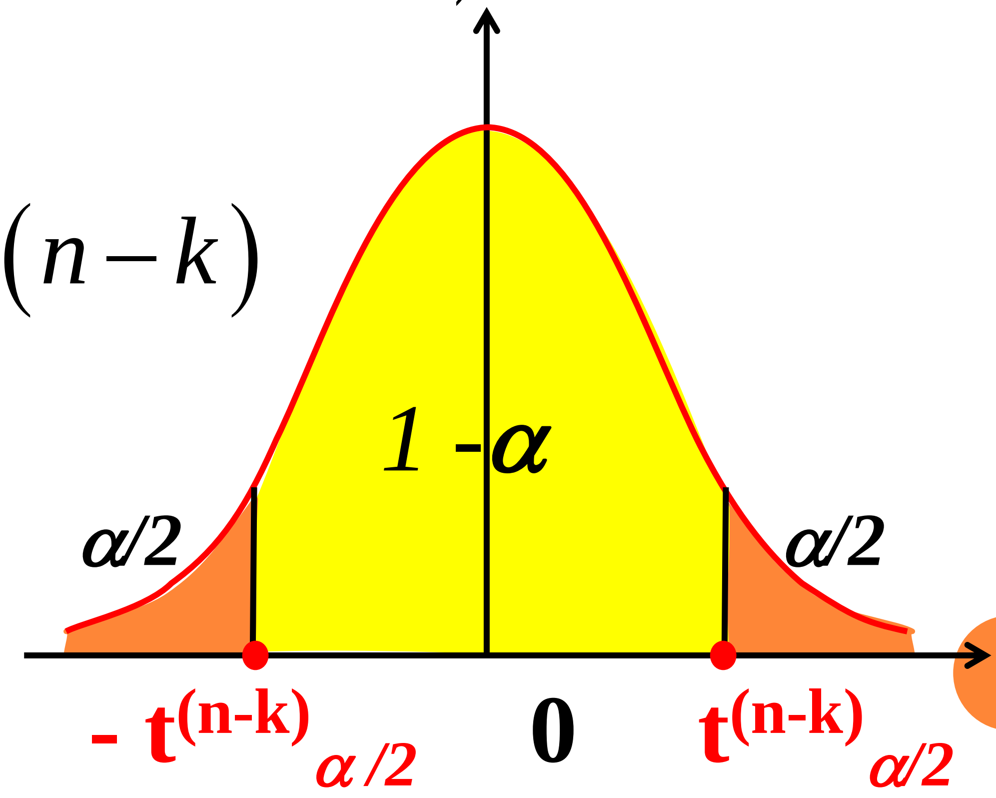
$$\Rightarrow \hat{\beta}_k \sim N\left(\beta_k, \frac{\sigma^2}{S_{kk}}\right)$$

5. ƯỚC LƯỢNG VÀ KIỂM ĐỊNH GIẢ THIẾT CÁC HỆ SỐ HỒI QUY

5.1. KHOẢNG TIN CẬY

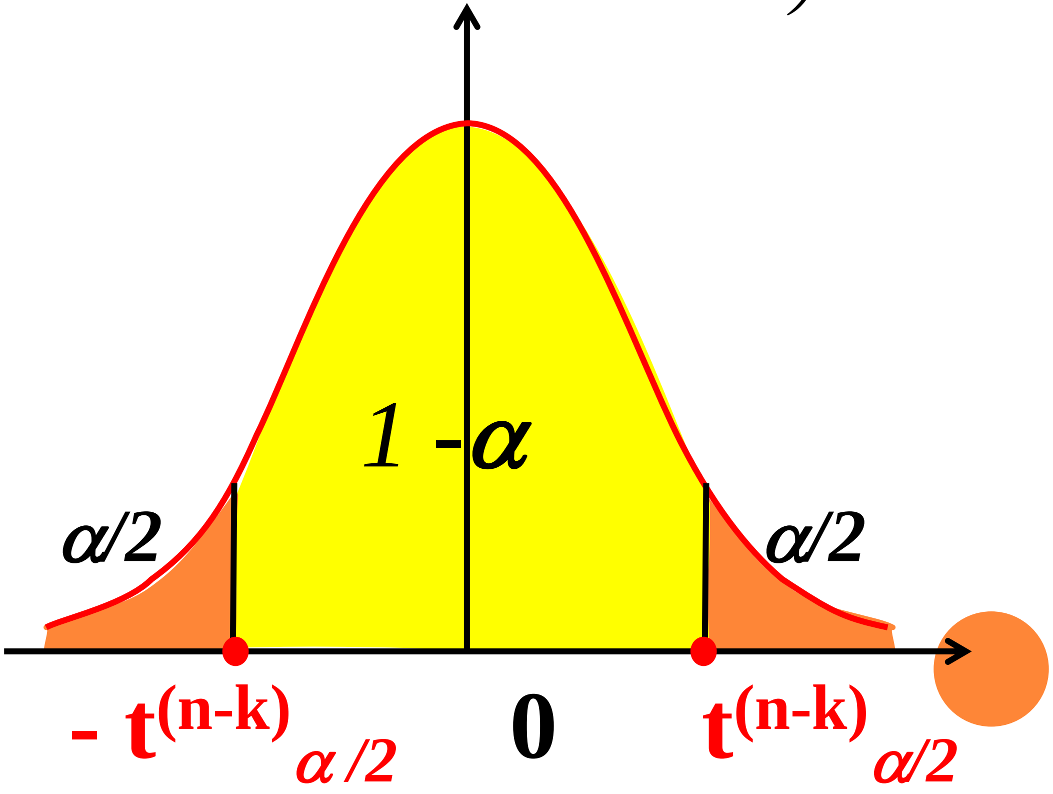
Ta có: $\hat{\beta}_i \sim N\left(\beta_i, \operatorname{Var}\left(\hat{\beta}_i\right)\right) \forall i = 1, \ldots, k$

$$\frac{\hat{\beta}_i - \beta_i}{Se\left(\hat{\beta}_i\right)} \sim T(n-k)$$



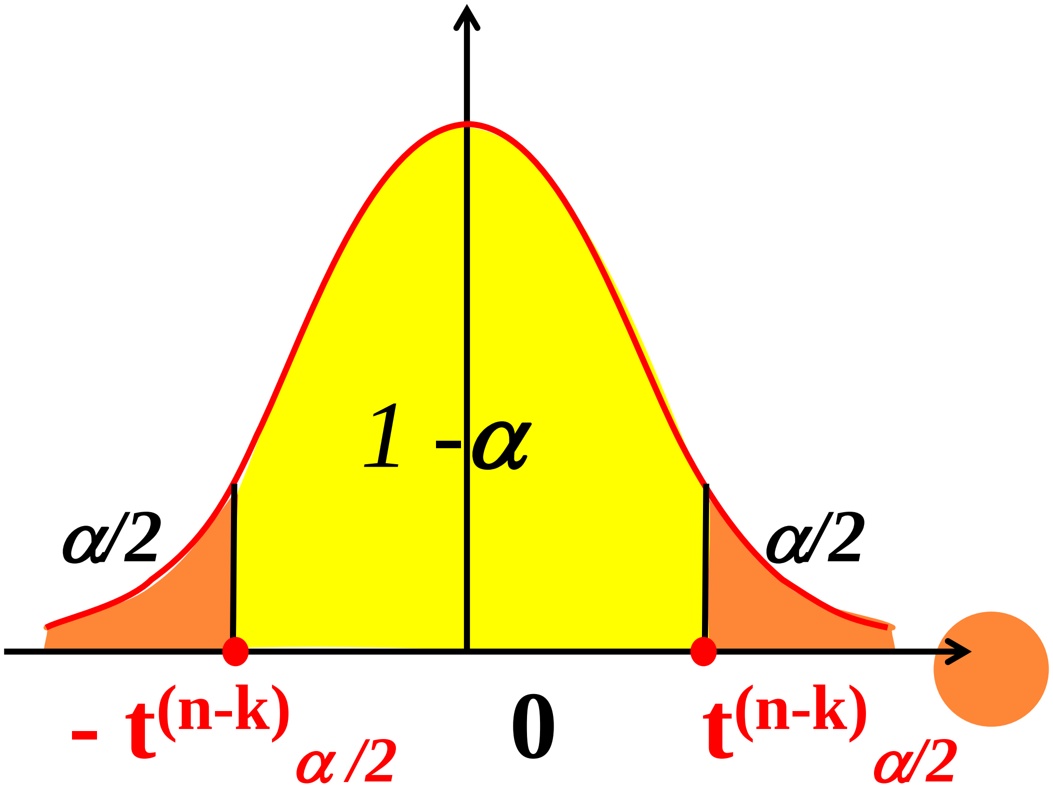
$$P\left(-t_{\alpha/2}^{(n-k)} \le \frac{\hat{\beta}_i - \beta_i}{Se\left(\hat{\beta}_i\right)} \le t_{\alpha/2}^{(n-k)}\right) = 1-\alpha$$

$$P\left(\hat{\beta}_i - t_{\alpha/2}^{(n-k)} Se\left(\hat{\beta}_i\right) \le \beta_i \le \hat{\beta}_i + t_{\alpha/2}^{(n-k)} Se\left(\hat{\beta}_i\right)\right) = 1-\alpha$$



vậy :

$$\beta_i \in \left(\hat{\beta}_i \pm t_{\alpha/2}^{(n-k)} Se\left(\hat{\beta}_i\right) \right) \text{với do tin cậy } 1-\alpha$$

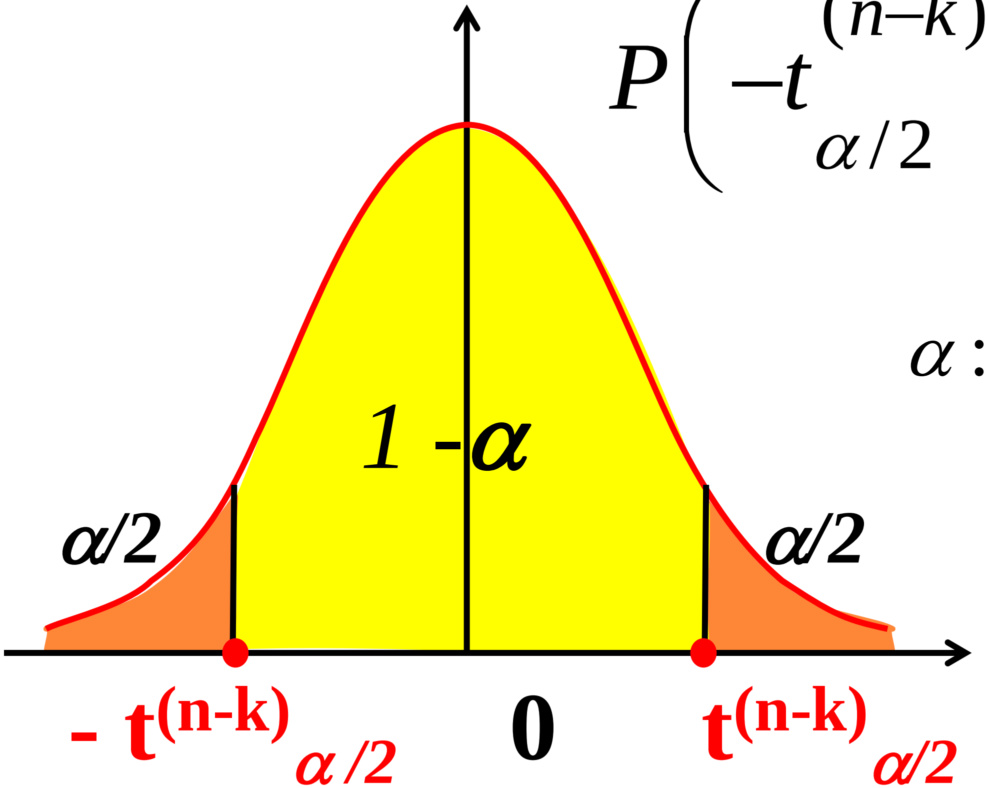


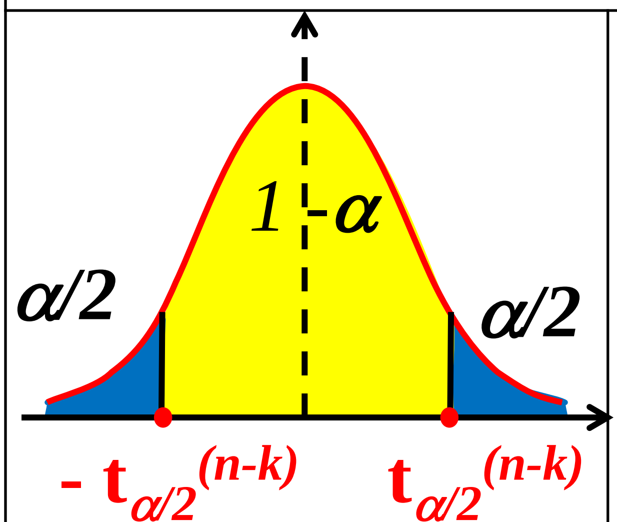
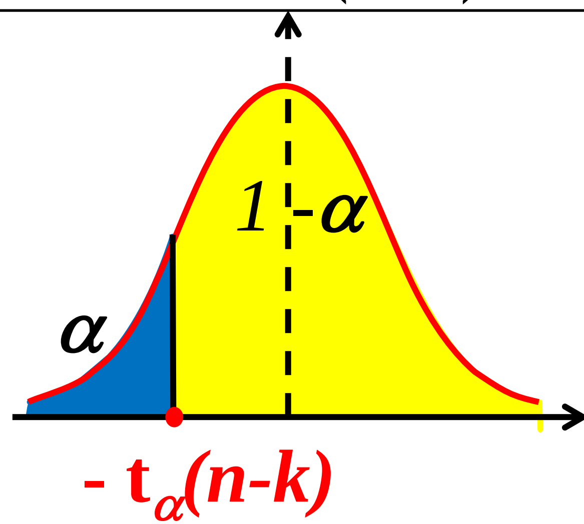
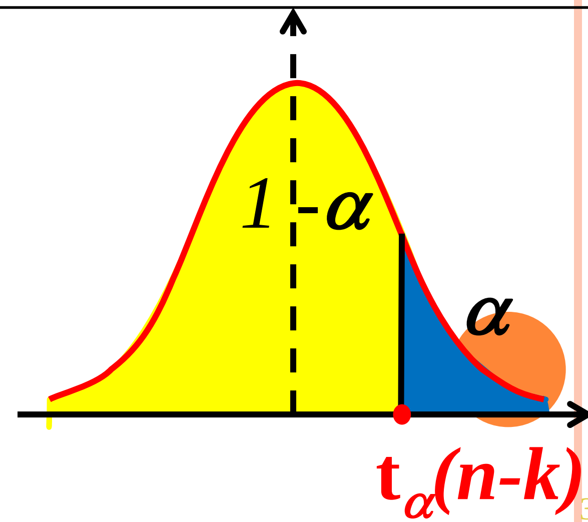
5.2. KIỂM ĐỊNH GIẢ THUYẾT

$$t = \frac{\hat{\beta}_i - \beta_i^*}{se\left(\hat{\beta}_i\right)} \sim t_{n-k}$$

$$P\left(-t_{\alpha/2}^{(n-k)} \le t \le t_{\alpha/2}^{(n-k)}\right) = 1-\alpha$$

α : mức ý nghĩa

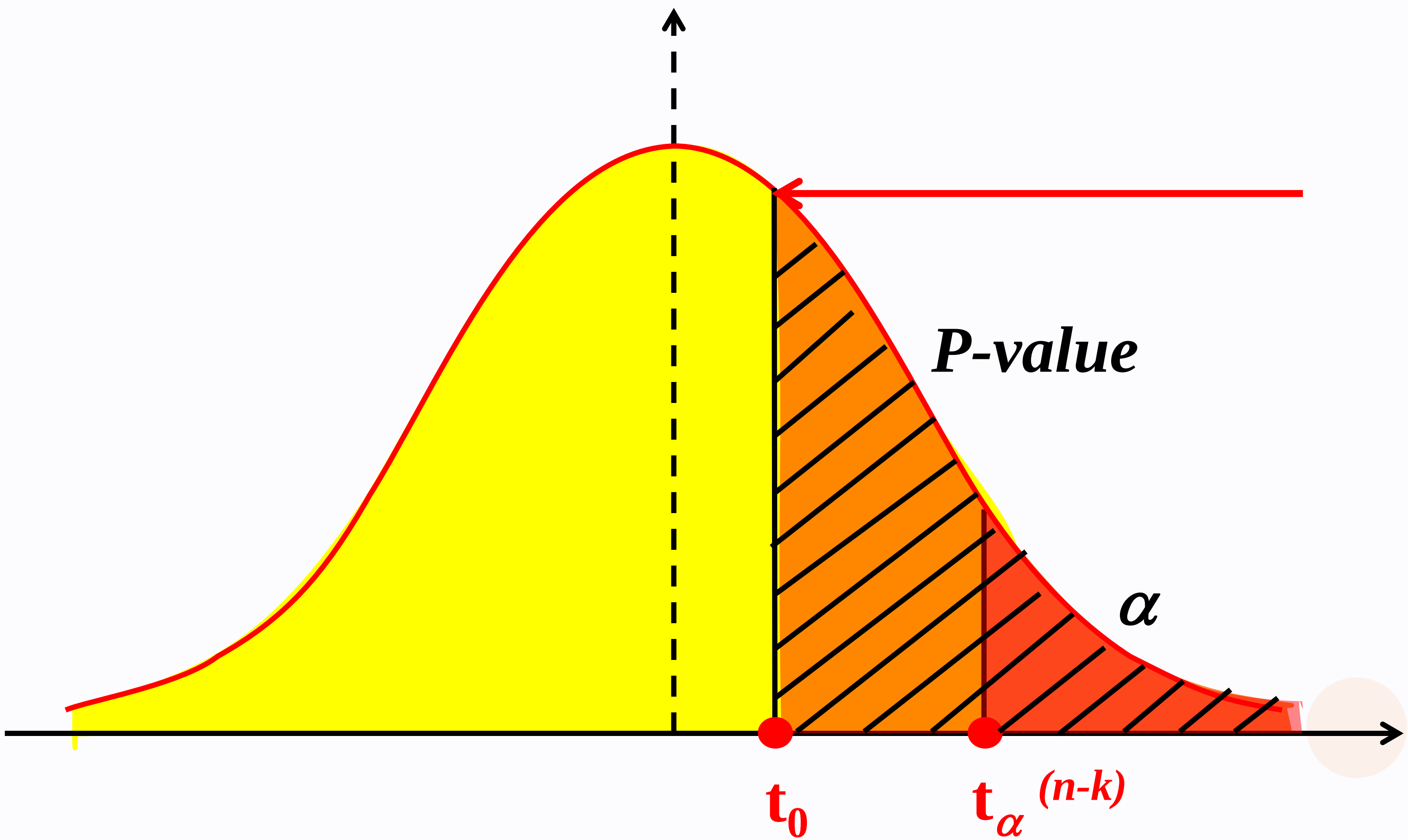


Kiểm định hai bên	Kiểm định bên trái	Kiểm định bên phải
$\begin{cases} H_0: \beta_i = \beta_i^* \\ H_1: \beta_i \neq \beta_i^* \end{cases}$	$\begin{cases} H_0: \beta_i \geq \beta_i^* \\ H_1: \beta_i < \beta_i^* \end{cases}$	$\begin{cases} H_0: \beta_i \leq \beta_i^* \\ H_1: \beta_i > \beta_i^* \end{cases}$
$t = \dfrac{\hat{\beta}_i - \beta_i^*}{se\left(\hat{\beta}_i\right)}$		
		

Kiểm định hai bên	Kiểm định bên trái	Kiểm định bên phải
$\begin{cases} H_0: \beta_i = \beta_i^* \\ H_1: \beta_i \neq \beta_i^* \end{cases}$	$\begin{cases} H_0: \beta_i \geq \beta_i^* \\ H_1: \beta_i < \beta_i^* \end{cases}$	$\begin{cases} H_0: \beta_i \leq \beta_i^* \\ H_1: \beta_i > \beta_i^* \end{cases}$
$$t = \frac{\hat{\beta}_i - \beta_i^*}{se\left(\hat{\beta}_i\right)}$$		
Bác bỏ Ho khi: $\lvert t_0 \rvert > \mathbf{t}_{\alpha/2}^{(n-k)}$	Bác bỏ Ho khi: $t_0 < -\mathbf{t}_{\alpha}^{(n-k)}$	Bác bỏ Ho khi: $t_0 > \mathbf{t}_{\alpha}^{(n-k)}$

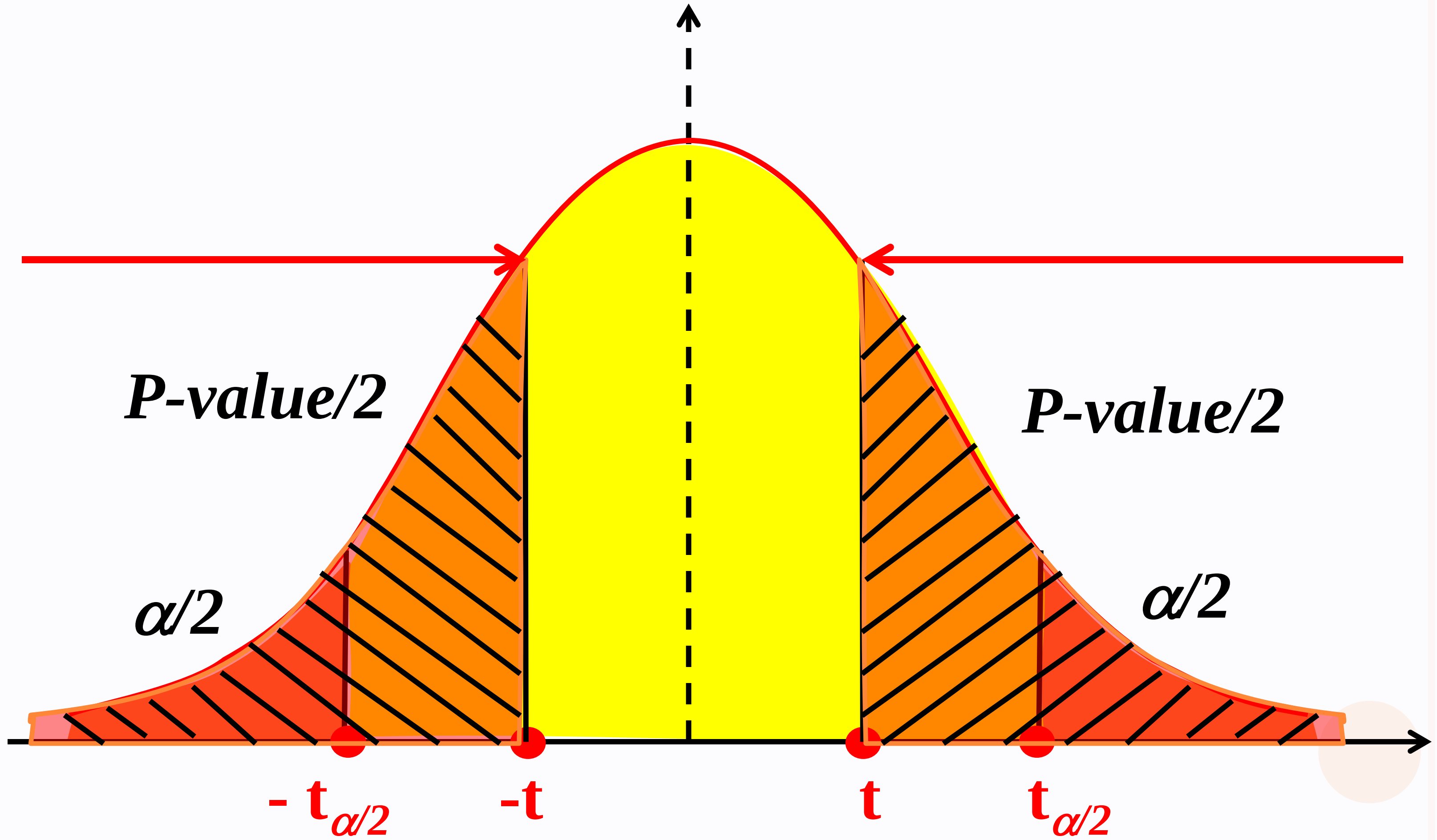
P-VALUE

$$\text{P-value} = P(|\, t_{(n-k)} \,| \geq |t|)$$



P-VALUE

$$\text{P-value} = P(|\,t_{(n-k)}\,| \geq |t|)$$



Quy luật dùng P-value:

$$\textbf{\textit{P-value}} < \alpha \Rightarrow \text{Bác bỏ Ho}$$

$$\textbf{\textit{P-value}} \geq \alpha \Rightarrow \text{Chấp nhận Ho}$$

5.3. KIỂM ĐỊNH SỰ PHÙ HỢP CỦA HÀM HỒI QUY

$$F = \frac{\mathrm{ESS}/(k-1)}{RSS/(n-k)} \sim F(k-1, n-k)$$

Kiểm định giả thiết $\begin{cases} \mathrm{H}_0 : R^2 = 0 \\ \mathrm{H}_1 : R^2 \neq 0 \end{cases}$

B1: Tính $F = \dfrac{R^2(n-k)}{(1-R^2)(k-1)}$

Tra bảng tìm $F_\alpha(k-1, n-k)$ (phụ lục 4)

B2: Kết luận: Bác bỏ H_0 nếu $F > F_\alpha(k-1, n-k)$

5.4. KIỂM ĐỊNH HỒI QUY CÓ ĐIỀU KIỆN (KIỂM ĐỊNH GIẢ THIẾT ĐỒNG THỜI)

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + u_i \; (U)$$

Nếu bỏ đi m biến thì mô hình (U) trở thành:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_{k-m} X_{(k-m)i} + u_i \; (R)$$

Vậy mô hình (R) chính là mô hình (U) với điều kiện

$$\beta_k = \beta_{k-1} = \ldots = \beta_{k-m+1} = 0$$

Việc lựa chọn mô hình nào, (U) hay (R), chính là thực hiện kiểm định

$$\begin{cases} H_0 : \beta_k = \beta_{k-1} = \ldots = \beta_{k-m+1} = 0 \\ H_1 : \exists \beta_i \neq 0 \end{cases}$$

Nếu Ho sai, nghĩa là m biến giải thích này thật sự có ảnh hưởng đến Y thì $RSS_U < RSS_R$

Vậy nếu ($RSS_R - RSS_U$) lớn thì ta sẽ bác bỏ Ho

<u>Phương pháp kiểm định:</u>

$$F = \frac{\left(RSS_R - RSS_U\right)/m}{RSS_U/\left(n-k\right)} \sim F_\alpha(m, n-k)$$

$$\Leftrightarrow F = \frac{\left(R^2{}_U - R^2{}_R\right)/m}{\left(1 - R^2{}_U\right)/\left(n-k\right)} \sim F_\alpha(m, n-k)$$

Nếu $F > F(m,n\text{-}k)$ thì bác bỏ Ho