

# Chương 3: SUY DIỄN THỐNG KÊ VÀ DỰ BÁO TỪ MÔ HÌNH HỒI QUY

1. QUY LUẬT PHÂN PHỐI XÁC SUẤT CỦA MỘT SỐ THỐNG KÊ MẪU
2. BÀI TOÁN TÌM KHOẢNG TIN CẬY CHO CÁC HỆ SỐ HỒI QUY
3. BÀI TOÁN KIỂM ĐỊNH GIẢ THUYẾT THỐNG KÊ VỀ HỆ SỐ HỒI QUY
4. DỰ BÁO GIÁ TRỊ CỦA BIẾN PHỤ THUỘC VÀ SAI SỐ DỰ BÁO

# 3.1 QUY LUẬT PHÂN PHỐI XÁC SUẤT CỦA MỘT SỐ THỐNG KÊ MẪU

Xét mô hình hồi quy tuyến tính k biến

$$Y = \beta_1 + \beta_2 X_2 + \dots + \beta_k X_k + U$$

**GT5:** Sai số ngẫu nhiên tuân theo quy luật phân phối chuẩn  $U_i \sim N(0, \sigma^2)$ .

Khi GT1 – 2, 3, 4, 5 thỏa thì các ước lượng OLS là ước lượng không chệch tốt nhất BUE (Best Unbiased Estimator) kể cả trong lớp các ước lượng tuyến tính và phi tuyến.

Khi các GT1, 2, 3, 4, 5 thỏa mãn thì :

- $\hat{\beta}_j \sim N(\beta_j, \text{var}(\hat{\beta}_j))$

- Do đó  $\frac{\hat{\beta}_j - \beta_j}{\sqrt{\text{var}(\hat{\beta}_j)}} \sim N(0, 1)$

- Và  $\frac{\hat{\beta}_j - \beta_j}{\text{se}(\hat{\beta}_j)} \sim T_{n-k}$

- $\frac{a\hat{\beta}_j + b\hat{\beta}_s - a\beta_j - b\beta_s}{\text{se}(a\hat{\beta}_j + b\hat{\beta}_s)} \sim T_{n-k}$

## 3.2 BÀI TOÁN TÌM KHOẢNG TIN CẬY CHO CÁC HỆ SỐ HỒI QUY

Xét mô hình hồi quy tuyến tính k biến

$$Y = \beta_1 + \beta_2 X_2 + \dots + \beta_k X_k + U$$

Khoảng tin cậy đối xứng với độ tin cậy  $1 - \alpha$  cho hệ số hồi quy  $\beta_j$  ( $j=1, \dots, k$ ) :

$$\beta_j - t_{\alpha/2, n-k} \text{se}(\beta_j) ; \beta_j + t_{\alpha/2, n-k} \text{se}(\beta_j)$$

Khoảng tin cậy đối xứng với độ tin cậy  $1 - \alpha$  cho giá trị  $a\beta_j + b\beta_s$  (với  $a, b$  bất kỳ,  $j$  và  $s = 1, \dots, k$ ) :

$$a\beta_j + b\beta_s - t_{\alpha/2, n-k} \text{se}(a\beta_j + b\beta_s) ; a\beta_j + b\beta_s + t_{\alpha/2, n-k} \text{se}(a\beta_j + b\beta_s)$$

Trong đó:

$$\text{se}(a\beta_j + b\beta_s) = \sqrt{a^2 \text{se}^2(\beta_j) + b^2 \text{se}^2(\beta_s) + 2ab \text{cov}(\beta_j, \beta_s)}$$

Khoảng tin cậy hai phía với độ tin cậy  $1 - \alpha$  cho phương sai của sai số ngẫu nhiên.

$$\frac{(n - k) \sigma^2}{\chi^2_{\alpha/2}(n - k)} < \sigma^2 < \frac{(n - k) \sigma^2}{\chi^2_{1 - \alpha/2}(n - k)}$$

**Ví dụ 1:**(Đơn vị : triệu đồng)

Sử dụng mẫu gồm 30 quan sát thu được kết quả ước lượng:

$$CT = 4,53 + 0,88 TN + 0,14 TS + e$$

$$(se) \quad (2,113) \quad (0,023) \quad (0,004)$$

$$Cov(\hat{\beta}_2, \hat{\beta}_3) = 0,0001$$

Với độ tin cậy 95%.

a) Hãy tìm các khoảng tin cậy cho các hệ số hồi quy.

Biết  $t_{0,025; 27} = 2,052$ ;  $t_{0,05; 27} = 1,703$ ;  $t_{0,025; 28} = 2,048$ .

b) Khi TN tăng 2 triệu đồng và TS giảm 1 triệu đồng thì CT trung bình thay đổi như thế nào?

# 3.3 BÀI TOÁN KIỂM ĐỊNH GIẢ THUYẾT THỐNG KÊ VỀ HỆ SỐ HỒI QUY

Xét mô hình hồi quy tuyến tính k biến

$$Y = \beta_1 + \beta_2 X_2 + \dots + \beta_k X_k + U$$

3.3.1 Kiểm định so sánh  $\beta_j$  với  $\beta_j^*$  với mức ý nghĩa  $\alpha$ .

Loại giả thuyết	$H_0$	$H_1$	Miền bác bỏ $H_0$
Hai phía	$\beta_j = \beta_j^*$	$\beta_j \neq \beta_j^*$	$ t  > t_{\alpha/2, n-k}$
1 phía trái	$\beta_j \geq \beta_j^*$	$\beta_j < \beta_j^*$	$t < -t_{\alpha, n-k}$
1 phía phải	$\beta_j \leq \beta_j^*$	$\beta_j > \beta_j^*$	$t > t_{\alpha, n-k}$

Trong đó

$$t = \frac{\beta_j - \beta_j^*}{se(\beta_j)}$$



**Ví dụ 2:** W: lương - triệu đồng/tháng; GD: số năm đi học; KN: số năm đi làm. Sử dụng mẫu gồm 30 quan sát từ doanh nghiệp tư nhân, thu được kết quả hồi quy:

$$\begin{array}{ccccccc} W & = & 2,23 & + & 0,28GD & + & 0,2KN & + & e \\ (se) & & (0,15) & & (0,02) & & (0,03) & & \end{array}$$

**a)** Giả sử mỗi năm lương của viên chức nhà nước tăng 0,16 triệu. Có thể cho rằng mức lương tăng theo số năm kinh nghiệm của doanh nghiệp tư nhân là cao hơn so với viên chức nhà nước không? Biết  $t_{0,05; 27} = 1,703$ .

**b)** Giả sử nhà nước có chính sách hỗ trợ người lao động vùng cao với mức tăng lương 0,18 triệu mỗi năm. Mức tăng lương này có khác biệt với doanh nghiệp tư nhân không?

Khi kết luận kiểm định, không nên phát biểu **“thừa nhận  $H_0$ ”**, chỉ nên phát biểu :

- \* Chưa đủ chứng cứ để bác bỏ  $H_0$ , do đó không thừa nhận  $H_1$ .**
- \* Hoặc đủ chứng cứ để bác bỏ  $H_0$ , do đó thừa nhận  $H_1$ .**

**Ví dụ 3:** Sử dụng mẫu gồm 40 quan sát thu được ước lượng sản lượng (Q) phụ thuộc vào vốn (K) và lao động (L) như sau:

$$\text{Ln}(Q) = 0,2 + 0,35\text{Ln}(K) + 0,48\text{Ln}(L) + e$$

(se)      (1,8)    (0,01)                      (0,13)

Cho biết  $\text{Cov}(\beta_2, \beta_3) = 0,002$

- a) Hãy giải thích ý nghĩa các hệ số hồi quy của biến độc lập.
- b) Có ý kiến cho rằng hệ số co giãn của sản lượng theo lao động là 0,6. Với mức ý nghĩa 5%, hãy cho nhận xét về ý kiến đó.
- c) Khi vốn và lao động cùng tăng 1% thì sản lượng trung bình thay đổi trong khoảng nào với độ tin cậy 95%?
- d) Với mức ý nghĩa 5%, hãy cho nhận xét về ý kiến: "Mức độ ảnh hưởng của lao động đến sản lượng cao hơn mức độ ảnh hưởng của vốn". Cho biết

$$t_{0,025;37} = 2,026; t_{0,025;38} = 2,024;$$

$$t_{0,05;37} = 1,687; t_{0,05;38} = 1,686$$

### 3.3.2 Giá trị xác suất P (P – value) của các thống kê kiểm định

Giá trị xác suất P là mức ý nghĩa nhỏ nhất mà giả thuyết  $H_0$  bị bác bỏ tương ứng với giá trị quan sát của thống kê kiểm định này.

**Quy tắc:**

**Giá trị xác suất  $P < \alpha \rightarrow$  Bác bỏ  $H_0$ .**

**Ví dụ 4:** Kết quả ước lượng giá cổ phiếu của ngân hàng ACB theo các yếu tố **lãi suất** (%/năm), **tốc độ tăng trưởng cung tiền M2** (%/năm) và **tỷ giá hối đoái** (VND/USD) như sau:

Dependent Variable: LOG(GIA)				
Method: Least Squares				
Included observations: 39 after adjustments				
Variable	Coef	Std. Error	t-Stat	Prob.
C	-58.61522	26.21639	-2.23582	0.0318
LAI_SUAT	0.05687	0.01041	5.46104	0.0000
M2	-0.00939	0.00373	-2.52005	0.0164
LOG(TI_GIA)	6.13076	2.62832	2.33257	0.0255

a) Hãy giải thích ý nghĩa các ước lượng của các hệ số hồi quy của biến độc lập.

b) Với mức ý nghĩa 5%, hãy nhận xét về sự ảnh hưởng của các yếu tố lãi suất, M2, tỷ giá đến giá cổ phiếu của ngân hàng ACB.

### 3.3.3 Kiểm định giả thuyết về nhiều ràng buộc của các hệ số hồi quy – Kiểm định F.

Xét mô hình hồi quy tuyến tính k biến:  $Y = \beta_1 + \beta_2 X_2 + \dots + \beta_k X_k + U$

**Ví dụ : Kiểm định cặp giả thuyết**

$$H_0: \beta_2 = 0, \beta_3 = 0; H_1: \beta_2^2 + \beta_3^2 \neq 0$$

**B1:** Lập cặp giả thuyết thống kê.

**B2:** Ước lượng:

$Y = \beta_1 + \beta_2 X_2 + \dots + \beta_k X_k + U$  thu được  $RSS_U$ .

$Y = \beta_1 + \beta_4 X_4 + \dots + \beta_k X_k + V$  thu được  $RSS_R$ .

**B3:** Tính giá trị quan sát của thống kê kiểm định  
(m là số ràng buộc trong  $H_0$ .)

$$F = \frac{RSS_R - RSS_U}{RSS_U / (n - k)}$$

**B4:** Nếu  $F > f_\alpha(m, n - k)$  thì bác bỏ giả thuyết  $H_0$ .

Nếu  $F \leq f_\alpha(m, n - k)$  thì chưa có cơ sở bác bỏ giả thuyết  $H_0$ .

## Ví dụ : Kiểm định cặp giả thuyết

$$H_0: \beta_2 = 0, \beta_3 = 0; H_1: \beta_2^2 + \beta_3^2 \neq 0$$

$$\begin{aligned} F &= \frac{RSS_R - RSS_U}{m} \\ &= \frac{R_U^2 - R_R^2}{1 - R_U^2} \end{aligned}$$

Kiểm định F được thực hiện hoàn toàn tương tự cho các ràng buộc tuyến tính giữa các hệ số hồi quy.



Dependent Variable: LOG(ACB)			
Method: Least Squares			
Included observations: 39			
Variable	Coef	t-Stat	Prob.
C	-58.615	-2.236	0.0318
LAISUAT	0.057	5.461	0.0000
M2	-0.009	-2.520	0.0164
LOG(TIGIA)	6.131	2.333	0.0255
$R^2 = 0.5629$		RSS= 0.41698	

### Ví dụ 5:

Dependent Variable: LOG(ACB)			
Method: Least Squares			
Included observations: 39			
Variable	Coef	t-Stat	Prob.
C	2.449	37.216	0.0000
LAISUAT	0.041	4.995	0.0000
$R^2=0.4027$		RSS=0.569780	

Có ý kiến cho rằng cung tiền và tỉ giá cùng không có ảnh hưởng đến giá cổ phiếu của ngân hàng ACB với mức ý nghĩa 1%. Hãy nhận xét ý kiến đó. Cho biết  $f_{0,01}(2; 35)=5,268$



### 3.3.4 Kiểm định về sự phù hợp của hàm hồi quy.

Xét mô hình hồi quy tuyến tính k biến:  $Y = \beta_1 + \beta_2 X_2 + \dots + \beta_k X_k + U$

#### Kiểm định cặp giả thuyết

$$H_0: \beta_2 = 0, \beta_3 = 0; \dots \beta_k = 0$$

$$H_1: \beta_2^2 + \beta_3^2 + \dots + \beta_k^2 > 0$$

Đây là trường hợp đặc biệt của kiểm định F.

- Giá trị thống kê kiểm định

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

- Nếu  $F > f_{\alpha}(k-1, n - k)$  thì bác bỏ giả thuyết  $H_0$ .

Nếu  $F \leq f_{\alpha}(k-1, n - k)$  thì chưa có cơ sở bác bỏ giả thuyết  $H_0$ .

**Ví dụ 6:** Mô hình sau đây có phù hợp với mức ý nghĩa 5%?

Dependent Variable: DT

Method: Least Squares

Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	521.209	77.127	6.758	0.0000
CH	5.580	0.551	10.132	0.0000
QC	1.188	0.361	3.289	0.0043
R-squared	0.901	Mean dependent var		1416.700
Adjusted R-squared	0.889	S.D. dependent var		219.3150
F-statistic	77.43126	Durbin-Watson stat		1.70928
Prob(F-statistic)	0.000000			

# Chú ý:

So  
sánh

*Trường hợp kiểm định một ràng buộc:* giá trị xác suất của 2 thống kê quan sát là như nhau nên kết luận nhận được giống nhau.

kiểm  
định

*Trường hợp kiểm định đồng thời nhiều hơn một ràng buộc:* Có sự khác biệt trong kết luận.

T và  
kiểm  
định  
F

Kiểm định T có thể sử dụng cho ràng buộc dạng đẳng thức hoặc bất đẳng thức.

Kiểm định F chỉ sử dụng được cho các ràng buộc dạng đẳng thức.

Dependent Variable: WAGE				
Method: Least Squares				
Included observations: 33				
Variable	Coef	Std. E	t-Stat	Prob
C	6.78913	0.12384	54.8239	0.0000
EDU	0.06002	0.01507	3.98181	0.0002
EXPERT	0.01929	0.00589	3.32197	0.0018
R-squared	0.459472			

WAGE: tiền lương (USD/tháng),  
EDU: số năm được đào tạo sau phổ thông,  
EXPERT: số năm kinh nghiệm.

- a)** Viết mô hình hồi quy tổng thể và mô hình hồi quy mẫu.
- b)** Có thể cho rằng nếu được đào tạo sau phổ thông thêm 1 năm thì mức tăng của trung bình tiền lương sẽ bằng mức tăng trung bình tiền lương

của một nhân viên có thêm 1 năm kinh nghiệm hay không với mức ý nghĩa 5%? Biết

$$\text{cov}(\beta_2, \beta_3) = 0,00001$$

- c)** Với mức ý nghĩa 5%, mô hình có phù hợp hay không?

# 3.4 DỰ BÁO GIÁ TRỊ CỦA BIẾN PHỤ THUỘC VÀ SAI SỐ DỰ BÁO

## 3.4.1 Dự báo giá trị của biến phụ thuộc

Xét mô hình hồi quy 2 biến:  $\hat{Y} = \beta_1 + \beta_2 X$

$$var(\hat{Y}_0) = \sigma^2 \left( \frac{1}{n} + \frac{x_0^2}{\sum_{i=1}^n x_i^2} \right); \quad se(\hat{Y}_0) = \sigma \left( \frac{1}{n} + \frac{x_0^2}{\sum_{i=1}^n x_i^2} \right)^{1/2}$$

Khoảng tin cậy với độ tin cậy  $(1 - \alpha)$  cho giá trị trung bình của biến phụ thuộc khi  $X = X_i$  là:

$$\hat{Y}_0 - t_{\alpha/2}(n-2).se(\hat{Y}_0); \quad \hat{Y}_0 + t_{\alpha/2}(n-2).se(\hat{Y}_0)$$

### 3.4.2 Đánh giá sai số dự báo

Các chỉ tiêu đánh giá sai số:

- \* Căn bậc hai trung bình bình phương sai số (Root Mean Squared Error)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2}$$

- \* Sai số trung bình tuyệt đối (Mean Absolute Error)

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{Y}_i - Y_i|$$

- \* Sai số trung bình tuyệt đối theo phần trăm (Mean Absolute Percent Error)

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{Y}_i - Y_i}{Y_i} \right|$$