

Chương 5

Kiểm định giả thuyết thống kê với phương trình hồi qui đa biến

TS. Đinh Thị Thanh Bình
Khoa Kinh Tế Quốc Tế- Đại học Ngoại thương

1. Phân bố xác suất của các ước lượng OLS

Giả thiết 9: Sai số u độc lập với các biến X và có phân phối chuẩn:

$$u \sim N(0, \sigma^2)$$

Định lý 4.1: Với giả thiết từ 1-9,

$$\beta_j \sim \text{Normal}[(\beta_j, \text{Var}(\beta_j))]$$

$$\Rightarrow (\beta_j - \beta_j) / \text{sd}(\beta_j) \sim \text{Normal}(0, 1)$$

Định lý 4.2: Với giả thiết từ 1-9,

$$(\hat{\beta}_j - \beta_j) / se(\hat{\beta}_j) \leq t_{n-k-1}$$

trong đó k là số lượng biến độc lập

5.1. Khoảng tin cậy

- Với cỡ mẫu n và k biến độc lập, xác định thống kê T cho kiểm định hệ số hồi qui và cho kiểm định phương sai:

$$T = \frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \sim t_{n-k-1}$$

$$T = (n-k-1) \frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-k-1}$$

5.1. Khoảng tin cậy

- Khoảng tin cậy $(1-\alpha)$ của hệ số hồi quy :

$$\hat{\beta}_j - c_{\alpha/2} se(\hat{\beta}_j) \leq \beta_j \leq \hat{\beta}_j + c_{\alpha/2} se(\hat{\beta}_j)$$

- Khoảng tin cậy $(1-\alpha)$ của phương sai nhiều :

$$\frac{(n-k-1)\hat{\sigma}^2}{C_{\alpha/2}} \leq \sigma^2 \leq \frac{(n-k-1)\hat{\sigma}^2}{C_{1-\alpha/2}}$$

5.2. Kiểm định giả thiết về hệ số hồi quy

- Nhắc lại rằng thống kê T xác định bằng biểu thức :

$$T = \frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \sim t_{n-k-1}$$

- Và giá trị: $p\text{-value} = P(|T| > |t_0| \mid H_0)$

Bảng 1. Kiểm định giả thiết về hệ số hồi quy

Giả thiết	H_0	H_1	Phương pháp	Miền bác bỏ H_0
Hai phía	$\beta_j = \beta_j^*$	$\beta_j \neq \beta_j^*$	Khoảng tin cậy	$\beta^* \notin [\hat{\beta}_j \pm c_{\alpha/2} se(\hat{\beta}_j)]$
			Giá trị tới hạn	$ T > c_{\alpha/2}$
			p -value	$p\text{-value} < \alpha$

Bảng 1. Kiểm định giả thiết về hệ số hồi quy

Giả thiết	H_0	H_1	Phương pháp	Miền bác bỏ H_0
Phía phải	$\beta_j \leq \beta_j^*$	$\beta_j > \beta_j^*$	Khoảng tin cậy	$\beta^* \notin [\hat{\beta}_j - c_\alpha se(\hat{\beta}_j), +\infty]$
			Giá trị tới hạn	$T > c_\alpha$
			<i>p-value</i>	$p\text{-value}/2 < \alpha$
Phía trái	$\beta_j \geq \beta_j^*$	$\beta_j < \beta_j^*$	Khoảng tin cậy	$\beta^* \notin [-\infty, \hat{\beta}_j + c_\alpha se(\hat{\beta}_j)]$
			Giá trị tới hạn	$T < -c_\alpha$
			<i>p-value</i>	$p\text{-value}/2 < \alpha$

5.3. Kiểm định giả thiết về phương sai của nhiều

- Thống kê:

$$T = (n - k - 1) \frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-k-1}$$

- Và

$$p\text{-value} = P(|T| > |t_0| \mid H_0)$$

Bảng 2. Kiểm định giả thiết về phương sai của nhiều

giả thiết	H_0	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	$\sigma_0^2 \notin [(n-k-1)\frac{\hat{\sigma}^2}{c_{\alpha/2}}, (n-k-1)\frac{\hat{\sigma}^2}{c_{1-\alpha/2}}]$
			Giá trị tới hạn	$T > c_{\alpha/2}$ hoặc $T < c_{1-\alpha/2}$
			<i>p-value</i>	$p\text{-value} < \alpha/2$ hoặc $p\text{-value} > 1 - \alpha/2$
Phía phải	$\sigma^2 = \sigma_0^2$	$\sigma^2 > \sigma_0^2$	Khoảng tin cậy	$\sigma_0^2 \notin [(n-k-1)\frac{\hat{\sigma}^2}{c_\alpha}, +\infty]$
			Giá trị tới hạn	$T > c_\alpha$
			<i>p-value</i>	$p\text{-value} < \alpha$
Phía trái	$\sigma^2 = \sigma_0^2$	$\sigma^2 < \sigma_0^2$	Khoảng tin cậy	$\sigma_0^2 \notin [-\infty, (n-k-1)\frac{\hat{\sigma}^2}{c_{1-\alpha}}]$
			Giá trị tới hạn	$T < c_{1-\alpha}$
			<i>p-value</i>	$p\text{-value} > 1 - \alpha$

5.4. Kiểm định đa ràng buộc tuyến tính

- Xét hai mô hình sau :

- (UR) : $Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + u$

- (R) : $Y = \beta_0 + \beta_1 X_1 + \dots + \beta_{k-q} X_{k-q} + v$

→ **q biến độc lập bị loại khỏi mô hình**

- (UR) gọi là mô hình không bị ràng buộc (Unrestricted model)
- (R) gọi là mô hình bị ràng buộc (Restricted model).

5.4. Kiểm định đa ràng buộc tuyến tính

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

$$H_0 : \beta_{k-q+1} = \dots = \beta_k = 0$$

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

5.4. Kiểm định đa ràng buộc tuyến tính

- **Bước 1** : Hồi quy (UR) gồm k tham số, tính SSR_{UR} ,
(n-k-1) bậc tự do $\leftarrow df_{ur}$
- **Bước 2** : Hồi quy (R) gồm k-q tham số, tính SSR_R ,
[(n-(k-q))-1] bậc tự do. $\leftarrow df_r$
- **Bước 3** : Sử dụng thống kê F như sau :

$$F = \frac{(SSR_r - SSR_{ur}) / q}{SSR_{ur} / (n - k - 1)} \square F_{q, n-k-1}$$

$$F = \frac{(R_{ur}^2 - R_r^2) / q}{(1 - R_{ur}^2) / (n - k - 1)} \square F_{q, n-k-1}$$

5.4. Kiểm định đa ràng buộc tuyến tính

- Với mức ý nghĩa α , tra bảng F tìm giá trị tới hạn c_α

Nếu $F > c_\alpha$ thì bác bỏ H_0



Kiểm định **F** hay được gọi là kiểm định **Wald**

5.4. Kiểm định đa ràng buộc tuyến tính

- Thông thường các phần mềm ứng dụng về KTL sẽ cho ra thông báo về việc kiểm định giả thuyết về tính có ý nghĩa chung của cả mô hình (overall significance). Giá trị của F_w lúc này được gọi là **F-stat**.
- Đi kèm theo nó, các phần mềm cũng cho ra p-value của 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.

5.5. Kiểm định F cho toàn bộ hệ số hồi quy

- Giả sử ta có mô hình hồi quy mẫu với k biến độc lập:
- Ta muốn kiểm định giả thiết :

$$H_0 : \beta_1 = \beta_2 \dots = \beta_k = 0$$

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

5.5. Kiểm định F cho toàn bộ hệ số hồi quy

- Áp dụng kiểm định Wald :

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

- Nếu $F > c_\alpha$ thì bác bỏ H_0 .

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

- Giả sử ta có mô hình hồi quy mẫu với 4 biến độc lập:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + u$$

- Ta muốn kiểm định giả thiết :

$$H_0 : \beta_1 = 1, \beta_2 = 0, \beta_3 = 0, \beta_4 = 0$$

$$H_1 : H_0 \text{ không đúng}$$

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

UR:
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + u \quad [1]$$

R:
$$Y = \beta_0 + X_1 + u \quad [2]$$

$\Downarrow$

$$Y - X_1 = \beta_0 + u \quad [3]$$

Bước 1: Ước lượng [1] $\rightarrow SSR_{ur}$

Bước 2: Ước lượng [3] $\rightarrow SSR_r$

Bước 3: Tính thống kê F với $q=4$ bậc tự do ở tử số và $n-5$ bậc tự do ở mẫu số

Bước 4: $F > c_\alpha \rightarrow$ bác bỏ H_0

5.7. Kiểm định sự phù hợp của mô hình

- Để kiểm định sự phù hợp của mô hình hồi quy, ta xây dựng giả thiết như sau:

$$\mathbf{H}_0 : R^2 = 0 \leftrightarrow \mathbf{H}_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$\mathbf{H}_1 : R^2 \neq 0 \leftrightarrow \mathbf{H}_1 : \text{Có ít nhất một } \beta_i \neq 0$$

- Các biến độc lập đồng thời không ảnh hưởng đến biến phụ thuộc $\rightarrow$ hàm hồi quy mẫu không giải thích được sự giao động của biến phụ thuộc $\rightarrow$ SRF không phù hợp.
- $\rightarrow$ Giống phần 5.5. Kiểm định F cho toàn bộ hệ số hồi qui
- **Lưu ý:** Giả thiết $H_0: \beta_1 = 0$ và $H_0: \beta_2 = 0$ (kiểm định riêng) không tương đương với $H_0: \beta_1 = \beta_2 = 0$ (kiểm định đồng thời)

•	Source		SS	df	MS	Number of obs =	1191
•	-----+	-----				F(5, 1185) =	9.55
•	Model		18705.5567	5	3741.11135	Prob > F =	0.0000
•	Residual		464041.135	1185	391.595895	R-squared =	0.0387
•	-----+	-----				Adj R-squared =	0.0347
•	Total		482746.692	1190	405.669489	Root MSE =	19.789

•	-----						
•	bwght		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
•	-----+	-----					
•	cigs		-.5959362	.1103479			
•	faminc		.0560414	.0365616			
•	fatheduc		.4723944	.2826433			
•	motheduc		-.3704503	.3198551			
•	parity		1.787603	.6594055			
•	_cons		114.5243	3.728453			
•	-----						

Source	SS	df	MS	Number of obs =	1191
-----+-----				F(5, 1185) =	9.55
Model	18705.5567	5	3741.11135	Prob > F =	0.0000
Residual	464041.135	1185	391.595895	R-squared =	0.0387
-----+-----				Adj R-squared =	0.0347
Total	482746.692	1190	405.669489	Root MSE =	19.789

bwght	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
-----+-----						
cigs	-.5959362	.1103479	-5.40	0.000	-.8124352	-.3794373
faminc	.0560414	.0365616	1.53	0.126	-.0156913	.1277742
fatheduc	.4723944	.2826433	1.67	0.095	-.0821426	1.026931
motheduc	-.3704503	.3198551	-1.16	0.247	-.9979957	.2570951
parity	1.787603	.6594055	2.71	0.007	.4938709	3.081336
_cons	114.5243	3.728453	30.72	0.000	107.2092	121.8394
-----+-----						