CHƯƠNG 2: HỒI QUY 2 BIẾN

2.1 ƯỚC LƯỢNG THAM SỐ HỒI QUY BẰNG PHƯƠNG PHÁP BÌNH PHƯƠNG BÉ NHẤT THÔNG THƯỜNG *OLS*.

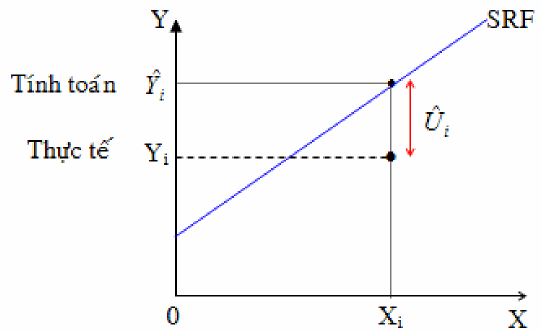
Mô hình hồi quy hai biến dạng quan hệ tuyến tính, như đã đề cập trong chương 1, có thể biểu diễn dưới dạng sau:

$$PRF\begin{cases} E(Y/X_i) = \alpha + \beta X_i \\ Y_i = E(Y/X) + U_i = \alpha + \beta X_i + U_i \end{cases}$$

$$SRF\begin{cases} \hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i \\ Y_i = \hat{Y}_i + \hat{U}_i = \hat{\alpha} + \hat{\beta} X_i + \hat{U}_i \end{cases} \qquad (2.1)$$

Bài toán đặt ra là làm cách nào để có được giá trị các tham số hồi quy? Như đã phân tích trong chương 1, chúng ta không khảo sát tổng thể nên trong thực tế, không thể biết chính xác α và β là bao nhiêu, mà chỉ có thể ước lượng chúng dựa vào mẫu ngẫu nhiên nào đó mà thôi. Để tìm các hệ số ước lượng cho α và β (ký hiệu là $\hat{\alpha}$, $\hat{\beta}$ tương ứng), ta có thể áp dụng phương pháp bình phương bé nhất *OLS* được mô tả như sau.

Giả sử chúng ta có một mẫu gồm *n* cặp giá trị quan sát của *X* và *Y*, ký hiệu là (X_i ,Y_i), trong đó các giá trị của *X* *không đồng nhất*. Ta phải tìm giá trị tính toán $\hat{Y}_i$ sao cho $\hat{Y}_i$ càng gần giá trị quan sát thực tế Y_i càng tốt, tức là $|\hat{U}_i| = |Y_i - \hat{Y}_i|$ càng nhỏ càng tốt. Để khảo sát tất cả các quan sát cùng lúc thì người ta xét $\sum_{i=1}^{n} \hat{U}_i^2$ (không xét $\sum_{i=1}^{n} |\hat{U}_i|$, tại sao?) và mong muốn $\sum_{i=1}^{n} \hat{U}_i^2 \rightarrow$ min.



$Y_i = \hat{Y}_i + \hat{U}_i (SRF) \Rightarrow |\hat{U}_i| = |Y_i - \hat{Y}|$ (càng nhỏ càng tốt)

Cách tìm các hệ số ước lượng

Bài toán trở thành tìm $\hat{\alpha}$, $\hat{\beta}$ sao cho $f \rightarrow$ *min.*

❖ *Xét điều kiện cần:*

$$\frac{\partial f}{\partial \hat{\alpha}} = f'_{\hat{\alpha}} = \sum 2(Y_i - \hat{\alpha} - \hat{\beta} X_i)(-1) = 0$$

$\rightarrow$ $n\hat{\alpha} + \hat{\beta}\sum X_i = \sum Y_i$ (2.2)

$$\frac{\partial f}{\partial \hat{\beta}} = f'_{\hat{\beta}} = \sum 2(Y_i - \hat{\alpha} - \hat{\beta} X_i)(-X_i) = 0$$

$\rightarrow$ $\hat{\alpha}\sum X_i + \hat{\beta}\sum X_i^2 = \sum X_i Y_i$ (2.3)

Giải hệ (2.2), (2.3), ta được:

$$\hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{\sum X_i Y_i - n\overline{X}\,\overline{Y}}{\sum X_i^2 - n(\overline{X})^2} \qquad (2.4)$$

$$\hat{\alpha} = \overline{Y} - \hat{\beta}\,\overline{X} \qquad (2.5)$$

Với:

$\overline{Y} = \frac{\sum Y_i}{n}$; $\overline{X} = \frac{\sum X_i}{n}$: được gọi là trung bình mẫu (theo biến) (2.6)

$x_i = X_i - \overline{X}$; $y_i = Y_i - \overline{Y}$: biểu thị độ lệch giá trị của biến so với giá trị trung bình.

<u>Điều kiện đủ</u>: tự xem trang 22-23

Thí dụ:

X_i	1	2	3	4	4	5	6	7
Y_i	8	6	6	5	4	4	4	3

Tìm hàm hồi quy mẫu SRF?

X_i	Y_i	X_iY_i	X_i^2
1	8	8	1
2	6	12	4
3	6	18	9
4	5	20	16
4	4	16	16
5	4	20	25
6	4	24	36
7	3	21	49
Tổng: *32*	*40*	*139*	*156*

$$\bar{X} = \frac{\sum X_i}{n} = \frac{32}{8} = 4 \ ; \ \bar{Y} = \frac{\sum Y_i}{n} = \frac{40}{8} = 5$$

$$\hat{\beta} = \frac{\sum X_i Y_i - n\bar{X}\bar{Y}}{\sum X_i^2 - n(\bar{X})^2} = \frac{139 - 8(4)(5)}{156 - 8(4)^2} = \frac{-21}{28} = -0.75$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\,\bar{X} = 5 - (-0.75)(4) = 8$$

$$SRF \begin{cases} \hat{Y}_i = 8 - 0.75X_i \\ Y_i = 8 - 0.75X_i + \hat{U}_i \end{cases}$$

Ý nghĩa của hệ số hồi quy:

1) Y: lượng hàng A bán được (tấn)
   X: giá hàng A (ngàn đ/kg)
   $\hat{Y}_i = 8 - 0.75\,X_i$

2) Y: lượng hàng bán được (tấn)
   X: chi phí quảng cáo (triệu đ)
   $\hat{Y}_i = 2 + 0.67\,X_i$

3) Y: doanh số bán (triệu đ)
   X: chi phí quảng cáo (triệu đ)
   $\hat{Y}_i = 4 + 3X_i$

Bảng sau đây minh họa cách tính $\hat{Y}_i$ và $\hat{U}_i$

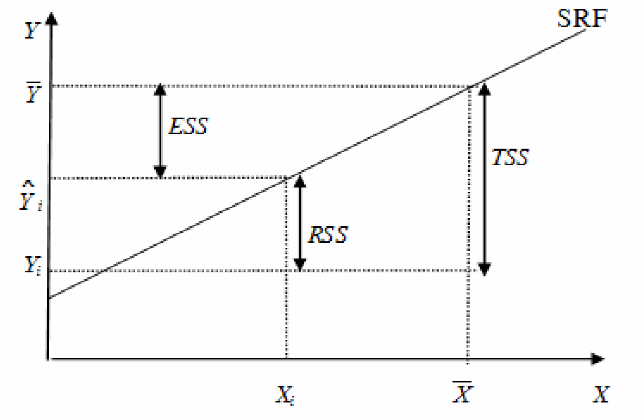
X_i	Y_i	$\hat{Y}_i = 8 - 0.75X_i$	$\hat{U}_i = Y_i - \hat{Y}_i$
1	8	8-0.75(1)= 7.25	8-7.25= 0.75
2	6	8-0.75(2)= 6.5	6-6.5 = -0.5
3	6	8-0.75(3)= 5.75	6-5.75= 0.25
4	5	8-0.75(4)= 5	5-5 = 0
4	4	8-0.75(4)= 5	4-5 = -1
5	4	8-0.75(5)= 4.25	4-4.25= -0.25
6	4	8-0.75(6)= 3.5	4-3.5 = 0.5
7	3	8-0.75(7)= 2.75	3-2.75= 0.25

Tính chất của SRF tìm được theo OLS

- Hàm hồi quy mẫu luôn đi qua điểm trung bình mẫu $(\bar{X}, \bar{Y})$, nghĩa là thỏa: $\bar{Y} = \hat{\alpha} + \hat{\beta}\bar{X}$.
- Giá trị trung bình của $\hat{Y}_i$ bằng giá trị trung bình của quan sát Y_i, nghĩa là $\bar{Y} = \bar{\hat{Y}}$ (với $\bar{\hat{Y}} = \frac{1}{n}\sum \hat{Y}_i$).

- Giá trị trung bình của các phần dư bằng 0, nghĩa là $\overline{\hat{U}} = \frac{1}{n}\sum \hat{U}_i = 0$.
- $\hat{U}_i$ và $\hat{Y}_i$ không tương quan với nhau, nghĩa là $cov(\hat{U}_i, \hat{Y}_i) = 0$ (do $\sum(\hat{U}_i \hat{Y}_i) = 0$).
- $\hat{U}_i$ và X_i không tương quan với nhau, nghĩa là $cov(\hat{U}_i, X_i) = 0$ (do $\sum(\hat{U}_i X_i) = 0$).

2.2 CÁC TỔNG BÌNH PHƯƠNG ĐỘ LỆCH



Câu hỏi:
- ý nghĩa TSS
- Ý nghĩa ESS
- Ý nghĩa RSS
- TSS=ESS+RSS?

- Hệ số xác định: mục đích định nghĩa? Tính chất, công thức tính?

2.3 HỆ SỐ XÁC ĐỊNH

$$TSS = ESS + RSS \Rightarrow 1 = \frac{ESS}{TSS} + \frac{RSS}{TSS} \qquad (2.18)$$

Với một mẫu cụ thể và sử dụng phương pháp ước lượng *OLS*, *TSS* là giá trị cố định, nhưng *ESS* và *RSS* có giá trị thay đổi tùy theo dạng hàm hồi quy. *Hàm SRF gọi là phù hợp tốt với các số liệu quan sát (mẫu) khi $\hat{Y}_i$ gần Y_i.*

Ta đưa ra một đại lượng để đo mức độ phù hợp của *SRF* với các số liệu quan sát, gọi là hệ số xác định R^2:

$$R^2 = 1 - \frac{RSS}{TSS} = \frac{ESS}{TSS} \qquad (2.19)$$

Hệ số xác định R^2 có các tính chất sau đây:

- R^2 nhận giá trị trong [0, 1]. Điều này dễ dàng thấy được vì R^2 được tính dựa theo các tổng bình phương và thỏa biểu thức (2.18).
- Với $R^2 = 1$ thì đường hồi quy thích hợp "hoàn hảo", tất cả các sai lệch của Y_i (so với giá trị trung bình) đều được giải thích bởi *SRF* (vì khi đó RSS=0, nghĩa là $\hat{Y}_i \equiv Y_i, \forall i$).
- Với $R^2 = 0$ thì *SRF* không thích hợp, tất cả các sai lệch của Y_i không được giải thích bởi hàm *SRF* (khi đó $RSS=TSS$, nghĩa là $\hat{Y}_i \equiv \overline{Y}, \forall i$).

Trong thực nghiệm, đối với mô hình hồi quy đơn :

Cách 1: $R^2 = 1 - \dfrac{RSS}{TSS} = \dfrac{ESS}{TSS} = \dfrac{\hat{\beta}^2 \sum x_i^2}{\sum y_i^2}$ (2.20)

Cách 2: $R^2 = \dfrac{(\sum x_i y_i)^2}{(\sum x_i^2).(\sum y_i^2)}$ (2.21)

Cách 3: $R^2 = (r_{Y\hat{Y}})^2$, với $r_{Y\hat{Y}}$ là hệ số tương quan giữa Y và $\hat{Y}$. (2.22)

Thí dụ 2.1 (tiếp theo): Tính các tổng bình phương độ lệch và hệ số xác định của hàm *SRF* ứng với mẫu đã cho như sau:

$$TSS = \sum y_i^2 = \sum Y_i^2 - n(\overline{Y})^2 = 218 - (8.5)^2 = 18$$

$$ESS = \hat{\beta}^2 \sum x_i^2 = (-0.75)^2 .(28) = 15.75$$

$$RSS = TSS - ESS = 18\text{-}15.75 = 2.25$$

$$R^2 = ESS / TSS = 15.75/18 = 0.875$$

$$1\text{-}R^2 = 0.125$$

<u>Hệ số tương quan</u>:

- Y=α+βX+U. ngoài tuyến tính theo tham số còn tuyến tính theo biến → đo mức độ tương quan tuyến tính giữa 2 biến X,Y.

  ta đã có công thức tính:

Cách 1: $r = \dfrac{\sum x_i y_i}{\sqrt{\sum x_i^2 . \sum y_i^2}}$ (phần ôn tập)

- Công thức tính theo hệ số xác định?

Cách 2: $r = \pm\sqrt{R^2}$, và r cùng dấu $\hat{\beta}$. (2.24)

Thí dụ 2.1 (tiếp theo): Hệ số tương quan sẽ là:

$$r = \pm\sqrt{R^2} = \pm\sqrt{0.875} = \pm 0.9354$$

Vì $\hat{\beta} = -0.75 < 0$, chứng tỏ X và Y có quan hệ nghịch biến, vậy $r = -0.9354$

2.4 CÁC GIẢ THIẾT CỦA PHƯƠNG PHÁP *OLS*

Trong phân tích hồi quy, mục đích của chúng ta là ước lượng, dự báo về tổng thể, tức là ước lượng $E(Y/X_i)=\alpha+\beta X_i$. Hàm *SRF* xây dựng ở trên nhằm mục đích này. Chúng ta không biết chất lượng của ước lượng này như thế nào, vì nó phụ thuộc vào:

- Dạng hàm của mô hình được lựa chọn
- Phương pháp ước lượng được sử dụng
- Phụ thuộc vào các X_i , U_i
- Phụ thuộc vào kích thước mẫu

Các giả thiết của phương pháp OLS

Các giả thiết của mô hình hồi quy tuyến tính cổ điển sẽ giúp chúng ta xác định các tính chất của các ước lượng tìm được, bao gồm:

Giả thiết 1: Các giá trị của X được xác định trước và không phải là đại lượng ngẫu nhiên.

Giả thiết 2: Đại lượng sai số ngẫu nhiên (nhiễu) U_i có kỳ vọng bằng 0, nghĩa là: $E(U_i/X_i) = 0$.

Giả thiết 3: U_i có phương sai là hằng số, nghĩa là $var(U_i/X_i) = \sigma^2$ với mọi i. Giả thiết này còn được gọi là giả thiết về hiện tượng phương sai không đổi (đồng đều, thuần nhất).

Giả thiết 4: Không có sự tương quan giữa các sai số ngẫu nhiên U_i. Điều này cũng đồng nghĩa với: $cov(U_i,U_j)=E\{[U_i-E(U_i)].[U_j-E(U_j)]\}=0$, với $i\neq j$

Giả thiết 5: U_i và X_i không tương quan với nhau, nghĩa là

$$cov(U_i, X_i)=0, \forall i$$

Giả thiết 6: Đại lượng sai số ngẫu nhiên có phân phối chuẩn, ký hiệu $U_i \sim N(0,\sigma^2)$

2.5 CÁC TÍNH CHẤT CỦA HỆ SỐ HỒI QUY

$\hat{\alpha}$, $\hat{\beta}$ là các ước lượng điểm của α và β, tìm được bằng phương pháp bình phương nhỏ nhất và có các tính chất sau:

- $\hat{\alpha}$, $\hat{\beta}$ được xác định một cách duy nhất với n cặp giá trị quan sát (X_i ,Y_i).
- $\hat{\alpha}$, $\hat{\beta}$ là các đại lượng ngẫu nhiên, với các mẫu khác nhau, giá trị của chúng sẽ khác nhau.

Để đánh giá mức độ biến động của $\hat{\alpha}$, $\hat{\beta}$, ta có công thức xác định phương sai (*var-variance*) và sai số chuẩn (*se–standard error*) của các tham số ước lượng như sau:

$$var(\hat{\alpha}) = \sigma^2_{\hat{\alpha}} = \frac{\sum X_i^2}{n} \cdot \frac{\sigma^2}{\sum x_i^2}; \quad se(\hat{\alpha}) = \sqrt{\operatorname{var}(\hat{\alpha})},$$

$$var(\hat{\beta}) = \sigma^2_{\hat{\beta}} = \frac{\sigma^2}{\sum x_i^2}; \; se(\hat{\beta}) = \sqrt{\operatorname{var}(\hat{\beta})} \quad .$$

Trong đó $\sigma^2 = var(U_i/X_i)$ là phương sai nhiễu của tổng thể mà trong thực tế hiếm khi biết được giá trị thực của nó.

ta thường dùng ước lượng điểm của σ^2 là $\hat{\sigma}^2$:

$$\hat{\sigma}^2 = \frac{RSS}{n-2} \qquad (2.27)$$

$\hat{\sigma} = \sqrt{\hat{\sigma}^2}$ được gọi là sai số chuẩn của hồi quy (*Standard Error of the Regression – SER*).

Khi thay $\hat{\sigma}^2$ vào công thức $var(\hat{\alpha})$, $var(\hat{\beta})$ thì thực chất ta thu được ước lượng điểm của chúng, nghĩa là $\hat{var}(\hat{\alpha})$, $\hat{var}(\hat{\beta})$:

$$\hat{var}(\hat{\alpha}) = \frac{\sum X_i^2}{n} \cdot \frac{\hat{\sigma}^2}{\sum x_i^2} \qquad (2.28)$$

$$\hat{var}(\hat{\beta}) = \frac{\hat{\sigma}^2}{\sum x_i^2} \qquad (2.29)$$

Thí dụ 2.1 (tiếp theo):

$$\hat{\sigma}^2 = \frac{RSS}{n-2} = \frac{2.25}{6} = 0.375$$

$$\hat{se}(\hat{\alpha}) = \sqrt{\frac{\sum X_i^2}{n} \cdot \frac{\hat{\sigma}^2}{\sum x_i^2}} = \sqrt{\frac{156}{8} \cdot \frac{0.375}{28}} = 0.511039$$

$$\hat{se}(\hat{\beta}) = \sqrt{\frac{\hat{\sigma}^2}{\sum x_i^2}} = \sqrt{\frac{0.375}{28}} = 0.115728$$

Chất lượng của các ước lượng tìm được bằng phương pháp *OLS* phụ thuộc vào những giả thiết cổ điển của mô hình. Chúng ta có định lý phản ánh tính chất của các ước lượng *OLS* như sau:

Định lý Gauss–Markov: Với các giả thiết (từ 1 đến 5) của mô hình hồi quy tuyến tính cổ điển, các ước lượng bình phương bé nhất là ước lượng tuyến tính, không chệch, có phương sai nhỏ nhất trong lớp các ước lượng tuyến tính, không chệch của α, β.

2.6 Khoảng tin cậy

- 2.6.1 Khoảng tin cậy của hệ số hồi quy. Trang 33
- Tại sao ta phải xét ước lượng khoảng (khoảng tin cậy)?

Với độ tin cậy 95%, tức là 1-α= 0.95⇒ α= 0.05 ⇒ $\alpha/2$ = 0.025 . Tra bảng thống kê *t-student* với phân vị 0.025 và bậc tự do là *n*-2= 8-2= 6, ta có $t_{0.025}^{(6)}$ = 2.447. Vậy khoảng tin cậy của các tham số α, β là:

$$\hat{\alpha} \pm t_{\alpha/2}^{(n-2)}.\hat{se}(\hat{\alpha}) \sim 8 \pm (2.447).(0.511039) \sim (6.749488\ ;\ 9.250512)$$

$$\hat{\beta} \pm t_{\alpha/2}^{(n-2)}.\hat{se}(\hat{\beta}) \sim -0.75 \pm (2.447).(0.115728)$$

$$\sim (-1.03319\ ;\ -0.46681)$$

2.6.2 khoảng tin cậy của phương sai

Xem trang 35

2.6.2 Khoảng tin cậy của phương sai

Phương sai tổng thể chính là phương sai của thành phần nhiễu *U*, mà ta ký hiệu là σ². Như đã biết, $\hat{\sigma}^2$ là ước lượng điểm của σ². Với độ tin cậy 1–α, tra bảng số của phân phối χ^2 để tìm các giá trị tới hạn $\chi^2_{1-\alpha/2}$, $\chi^2_{\alpha/2}$.

khoảng tin cậy cho σ²:

$$\frac{(n-2)\hat{\sigma}^2}{\chi^2_{\alpha/2}(n-2)} < \sigma^2 < \frac{(n-2)\hat{\sigma}^2}{\chi^2_{1-\alpha/2}(n-2)} \qquad (2.36)$$

Thí dụ 2.1 (tiếp theo): Cho độ tin cậy 95%, ta có α=0.05 ⇒ α/2 = 0.025 ;
1–α/2=0.975 . Tra bảng thống kê χ^2 với hai mức phân vị 0.025 và 0.975, bậc tự do tương ứng là *n*-2=6, ta được $\chi^2_{0.025}(6)= 14.4494$; $\chi^2_{0.975}(6)= 1.2373$. Vậy khoảng tin cậy 95% của σ² là:

$$\left(\frac{(n-2)\hat{\sigma}^2}{\chi^2_{\alpha/2}(n-2)};\frac{(n-2)\hat{\sigma}^2}{\chi^2_{1-\alpha/2}(n-2)}\right) \sim \left(\frac{6.(0.375)}{14.4494};\frac{6.(0.375)}{1.2373}\right)$$

$$\sim (0.155716 \ ; \ 0.818414)$$

2.7 Kiểm định giả thiết

- Ta xét các kiểm định giả thiết về các vấn đề gì?
- Các loại giả thiết kiểm định?
- Có bao nhiêu cách kiểm định cơ bản?

Trang 36

- Các kiểm định cần có:
  - Kiểm định giả thiết về hệ số hồi quy
  - Kiểm định giả thiết về phương sai của nhiễu
  - Kiểm định giả thiết về sự phù hợp của hàm SRF
- Có 3 loại giả thiết: 2 phía, phía phải, phía trái
- Có 3 cách kiểm định cơ bản:
  - Khoảng tin cậy (ít thông dụng)
  - Giá trị tới hạn (tính tay)
  - P-value (máy tính)

2.7.1 Kiểm định giả thiết về hệ số hồi quy

Ta sử dụng ký hiệu θ để chỉ α hoặc β, $\hat{\theta}$ là $\hat{\alpha}$ hoặc $\hat{\beta}$. Có ba dạng giả thiết kiểm định như sau về hệ số hồi quy:

- *Hai phía*: giả thiết thống kê H_0: $\theta=\theta_0$; giả thiết đối H_1: $\theta\neq\theta_0$. Với θ_0 là một giá trị số cho trước.
- *Phía phải*: giả thiết thống kê H_0: $\theta=\theta_0$; giả thiết đối H_1: $\theta>\theta_0$.
- *Phía trái*: giả thiết thống kê H_0: $\theta=\theta_0$; giả thiết đối H_1: $\theta<\theta_0$.

Có ba cách để xây dựng quy tắc quyết định xem là chấp nhận hay bác bỏ giả thiết H_0, đó là: phương pháp khoảng tin cậy, phương pháp giá trị tới hạn, phương pháp giá trị *p-value*.

Phương pháp khoảng tin cậy

- Tự xem, trang 36.
- Xem pp giá trị tới hạn, phương pháp p-value

2.7.1.2 Phương pháp giá trị tới hạn.

Bước 1: tính giá trị $t_0 = \dfrac{\hat{\theta}-\theta_0}{\hat{se}(\hat{\theta})}$.

Bước 2: tra bảng số *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 $t_{\alpha/2}^{(n-2)}$ hoặc $t_{\alpha}^{(n-2)}$.

Bước 3: So sánh t_0 với giá trị tới hạn. Quy tắc quyết định như sau:

- Kiểm định hai phía: $|t_0| \succ t_{\alpha/2}^{(n-2)}$: bác bỏ H_0.
- Kiểm định phía phải: $t_0 \succ t_{\alpha}^{(n-2)}$: bác bỏ H_0.
- Kiểm định phía trái: $t_0 \prec -t_{\alpha}^{(n-2)}$: bác bỏ H_0.

Thí dụ 2.1 (tiếp theo): Kiểm định giả thiết H_0: β=0 *(X không ảnh hưởng Y)*, H_1: β≠0 *(X có ảnh hưởng Y)*, với mức ý nghĩa 5%, ta áp dụng một trong các cách làm sau:

- *Phương pháp giá trị tới hạn*:

$$t_0 = \frac{\hat{\beta}-0}{\hat{se}(\hat{\beta})} = \frac{-0.75}{0.115728} = -6.48071$$

$|t_0| = 6.48071 > t_{0.025}^{(6)} = 2.447 \Rightarrow$ bác bỏ H_0.

- *Phương pháp giá trị p-value*:

$$t_0 = \frac{\hat{\beta}-0}{\hat{se}(\hat{\beta})} = \frac{-0.75}{0.115728} = -6.48071$$

p-value=$P(|t| > 6.48071) = 0.000641 < \alpha = 0.05 \Rightarrow$ bác bỏ H_0.

2.7.1.3 Phương pháp giá trị p-value.

- Bước 1: tính giá trị $t_0 = \dfrac{\hat{\theta} - \theta_0}{\hat{se}(\hat{\theta})}$.
- Bước 2: tính $p\_value = P(|t| > |t_0|)$, trong đó t là *ĐLNN* có phân phối *t-student* với (n-2) bậc tự do [2].
- 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.

<u>Nếu chưa biết α</u>:

p–value> 0.05: chấp nhận H_0.

p–value < 0.05: có cơ sở bác bỏ H_0.

- Phương pháp giá trị p-value:

$$t_0 = \frac{\hat{\beta} - 0}{\hat{se}(\hat{\beta})} = \frac{-0.75}{0.115728} = -6.48071$$

$p\text{-}value = P(|t| > 6.48071) = 0.000641 < \alpha = 0.05 \Rightarrow$ bác bỏ H_0.

Con số 0.000641 này dùng phần mềm để tính, hổng phải tính bằng tay.

2.7.2 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 2.1 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:

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

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

Loại giả thiết	Giả thiết	Giả thiết đối	Phương pháp tiến hành	Quy tắc quyết định
Hai phía	$H_0: \sigma^2 = \sigma_0^2$	$H_1: \sigma^2 \neq \sigma_0^2$	Giá trị tới hạn	$\chi_0^2 > \chi^2_{\alpha/2}(n-2)$ hoặc $\chi_0^2 < \chi^2_{1-\alpha/2}(n-2)$: bác bỏ H_0
			Giá trị p–value	*p-value* < α/2 hoặc *p-value* > 1 – α/2 : bác bỏ H_0
Phía phải	$H_0: \sigma^2 = \sigma_0^2$	$H_1: \sigma^2 > \sigma_0^2$	Giá trị tới hạn	$\chi_0^2 > \chi^2_{\alpha}(n-2)$: bác bỏ H_0
			Giá trị p–value	*p-value* < α: bác bỏ H_0
Phía trái	$H_0: \sigma^2 = \sigma_0^2$	$H_1: \sigma^2 < \sigma_0^2$	Giá trị tới hạn	$\chi_0^2 < \chi^2_{1-\alpha}(n-2)$: bác bỏ H_0
			Giá trị p–value	*p-value* > 1- α: bác bỏ H_0

Thí dụ 2.1 (tiếp theo): Giả sử muốn thực hiện kiểm định giả thiết phương sai của tổng thể bằng 0.3, ta có giả thiết H_0: σ^2=0.3, H_1: $\sigma^2 \neq 0.3$. Trong mục 2.6.1, ta tính được $\hat{\sigma}^2$=0.375. Với mức ý nghĩa α=0.05 và bậc tự do tương ứng *n*-2=6, tra bảng *Chi* bình phương ta có:

$\chi^2_{\alpha/2} = \chi^2_{0.025} = 14.4494$; $\chi^2_{1-\alpha/2} = \chi^2_{0.975} = 1.2373$

- *Phương pháp giá trị tới hạn*:

$$\chi^2_0 = \frac{(n-2)\hat{\sigma}^2}{\sigma^2_0} = \frac{6.(0.375)}{0.3} = 7.5$$

Nhận thấy

$\chi^2_{0.975} = 1.2373 < \chi^2_0 = 7.5 < \chi^2_{0.025} = 14.4494$:

chấp nhận H_0.

- *Phương pháp giá trị p–value*:

Ta có: *p-value*=$P(\chi^2 > \chi^2_0 = 7.5) = 0.277068$

$\alpha/2$=0.025<p-value<1-$\alpha/2$=0.975: chấp nhận H_0

2.7.3 Kiểm định sự phù hợp của mô hình

H_0: $R^2 = 0$, H_1: $R^2 > 0$. 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, $\beta = 0$:

$$H_0: R^2 = 0 \sim H_0: \beta = 0$$

Thông thường người ta tiến hành kiểm định H_0: $R^2 = 0$ bằng hai phương pháp như sau:

* *Phương pháp giá trị tới hạn:*

Bước 1: Tính $F_0 = \dfrac{ESS/1}{RSS/(n-2)} = \dfrac{R^2(n-2)}{1-R^2}$ (2.37)

Bước 2: Tra bảng *F* với mức ý nghĩa α và hai bậc tự do (1, *n*-2) ta được giá trị tới hạn $F_\alpha(1,n\text{-}2)$.

Bước 3: So sánh F_0 và $F_\alpha(1, n\text{-}2)$.

Nếu $F_0 > F_\alpha(1, n\text{-}2)$: bác bỏ H_0.

Nếu $F_0 \leq F_\alpha(1, n\text{-}2)$: chấp nhận H_0.

Thí dụ 2.1 (tiếp theo): Để kiểm định H_0: $R^2 = 0$, ta tính

$$F_0 = \frac{R^2(n-2)}{1-R^2} = \frac{(0.875).6}{1-0.875} = 42$$

Với mức ý nghĩa α=0.05, tra bảng F ta có $F_{0.05}(1,6)$=5.987 . Vậy F_0 >$F_{0.05}(1,6)$ nên bác bỏ H_0, thừa nhận $R^2 > 0$ là có ý nghĩa thống kê.

* *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 (0.05): bác bỏ H_0.

Nếu *p-value* > α (0.05): chấp nhận H_0.

2.8 HỒI QUY VÀ ĐƠN VỊ ĐO CỦA BIẾN

Trong mô hình hồi quy hai biến $Y_i = \hat{\alpha} + \hat{\beta}X_i + \hat{U}_i$, nếu đơn vị đo của X và Y thay đổi thì ta không cần phải tiến hành hồi quy lại, mà có thể áp dụng công thức đổi đơn vị đo. Nếu ký hiệu mô hình hồi quy mới là:

$$\hat{Y}_i^* = \hat{\alpha}^* + \hat{\beta}^* X_i^* + \hat{U}_i^*$$

trong đó:

$$Y_i^* = k_1 Y_i;\ X_i^* = k_2 X_i$$

k_1, k_2 thực chất là các hệ số tỉ lệ quy đổi giữa hai hệ thống đơn vị cũ và mới. Các hệ số hồi quy của mô hình mới được quy đổi như sau:

$$\hat{\alpha}^* = k_1\hat{\alpha};\ \hat{\beta}^* = \frac{k_1}{k_2}\hat{\beta} \qquad (2.38)$$

Thí dụ: $\hat{Y}_i$ **(triệu đ) = 3+7 X_i (triệu đ)**

* $\hat{Y}_i^*$ (ngàn đ) = 3.(10^3)+ 7.(10^3) X^*_i (triệu đ)

$k_1 = 10^3$, $k_2 = 1$

* $\hat{Y}_i^*$ (triệu đ) = 3.(1)+ 7.($\frac{1}{10^3}$) X^*_i (ngàn đ)

$k_1 = 1$, $k_2 = 10^3$

* $\hat{Y}_i^*$ (ngàn đ) = 3.(10^3)+ 7.(1) X^*_i (ngàn đ)

$k_1 = 10^3$, $k_2 = 10^3$

câu hỏi (không trả lời, thì sẽ có):

1) $\hat{Y}_i$ **(triệu đ) = 2-6 X_i (năm)**

* $\hat{Y}_i^*$ (triệu đ) = ?+ ? X^*_i (tháng)

Với 1 năm = 12 tháng

* $\hat{Y}_i^*$ (ngàn đ) = ?+ ? X^*_i (tháng)

2) $\hat{Y}_i$ **(triệu đ/tháng) = 4+9 X_i (tấn)**

* $\hat{Y}_i^*$ (triệu đ/tháng) = ?+ ? X^*_i (kg)

* $\hat{Y}_i^*$ (triệu đ/năm) = ?+ ? X^*_i (kg)

Với 1 năm = 12 tháng

3) $\hat{Y}_i$ **(triệu đ/năm) = 4+9 X_i (kg)**

* $\hat{Y}_i^*$ (triệu đ/tháng) = ?+ ? X^*_i (tấn)

2.9 TRÌNH BÀY KẾT QUẢ HỒI QUY

Thí dụ 2.1 (tiếp theo): chúng ta có thể trình bày các kết quả hồi quy tìm được của thí dụ này như sau:

$$\hat{Y}_i = 8 \quad - 0.75X_i \quad ; \quad n=8$$
$$se = (0.511) \quad (0.115) \quad ; \quad R^2 = 0.875$$
$$t = (15.654) \; (-6.4807) \; ; \quad F_0 = 42$$
$$p\text{-}value = (0.0000) \; (0.0006) \quad ; \quad (0.000641)$$
$$TSS = 18 \; ; \; ESS = 15.75 \; ; \; RSS = 2.25 \; ; \; \hat{\sigma}^2 = 0.375$$

Lưu ý rằng giá trị t được tính dựa trên phần sai lệch của tham số ước lượng so với 0, nghĩa là: $t = \dfrac{\hat{\theta} - 0}{\hat{se}(\hat{\theta})}$

Thí dụ 2.2: xem liền ngay tại lớp

Câu hỏi: trình bày kết quả hồi quy của thí dụ này?

Bắt chước thí dụ 2.1, một sự bắt chước thú vị!

TRÌNH BÀY KẾT QUẢ HỒI QUY của *Thí dụ 2.2*

$$\hat{Y}_i = 94.5522 \quad - 9.8209X_i \quad ; \quad n=10$$
$$se = (5.2771) \quad (0.8955) \quad ; \quad R^2 = 0.9376$$
$$t = (17.9174) \quad (-10.9667) \quad ; \quad F_0 = 120.2678$$
$$p\text{-}value = (0.0000) \quad (0.0000) \quad ; \quad (0.000004)$$
$$TSS = 516.9 \; ; \; ESS = 484.6612 \; ; \; RSS = 32.2388 \; ; \; \hat{\sigma}^2 = 4.02985$$

Kết quả tính bằng phần mềm Eviews cho thí dụ 2.2

- Giải thích chi tiết ở trang 46-48

Dependent Variable: Y
Method: Least Squares
Date: 09/24/06 Time: 23:32
Sample: 1 10
Included observations: 10

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	94.55224	5.277127	17.91737	0.0000
X	-9.820896	0.895522	-10.96667	0.0000

R-squared	0.937630	Mean dependent var	37.10000
Adjusted R-squared	0.929834	S.D. dependent var	7.578478
S.E. of regression	2.007449	Akaike info criterion	4.408463
Sum squared resid	32.23881	Schwarz criterion	4.468980
Log likelihood	-20.04231	F-statistic	120.2678
Durbin-Watson stat	1.002633	Prob(F-statistic)	0.000004

Câu hỏi:

- Nhìn kết quả tính bằng tay và kết quả tính bằng máy, rùi rút ra nhận xét?
- “Cực khổ” nhận xét 1 lần, dùng cả “1 đời” Kinh tế lượng! *(Trúc hổng Xanh)*
- Các thí dụ còn lại: tự xem!

Mời ghé thăm trang web:

- http://kinhteluong.ungdung.googlepages.com
- http://xacsuatthongke.googlepages.com
- http://phamtricao.googlepages.com
- www37.websamba.com/phamtricao
- www.phamtricao.web1000.com