Chương 2

MÔ HÌNH HỒI QUY HAI BIẾN

I. HỒI TUYẾN TÍNH 2 BIẾN

1. <u>Hàm hồi quy tuyến tính 2 biến của tổng thể</u>

Trong quan hệ hồi quy , <u>*một*</u> biến phụ thuộc có thể được giải thích bởi <u>*nhiều*</u> biến độc lập

Nếu chỉ nghiên cứu ***một*** biến phụ thuộc bị ảnh hưởng bởi ***một*** biến độc lập => ***Mô hình hồi quy hai biến***

Nếu mối quan hệ giữa hai biến này là tuyến tính => ***Mô hình hồi quy tuyến tính hai biến***

I. HỒI TUYẾN TÍNH 2 BIẾN

Hàm hồi quy tổng thể (PRF) của mô hình hồi quy hai biến

$$PRF : Y_i = \beta_1 + \beta_2 X_i + U_i$$

Hay:

$$E(Y \mid X_i) = \beta_1 + \beta_2 X_i$$

Trong đó

- Y : Biến phụ thuộc
- Y_i : Giá trị cụ thể của biến phụ thuộc
- X : Biến độc lập
- X_i : Giá trị cụ thể của biến độc lập
- U_i : Sai số ngẫu nhiên ứng với quan sát thứ i

I. HỒI TUYẾN TÍNH 2 BIẾN

Hàm hồi quy tổng thể (PRF) của mô hình hồi quy hai biến

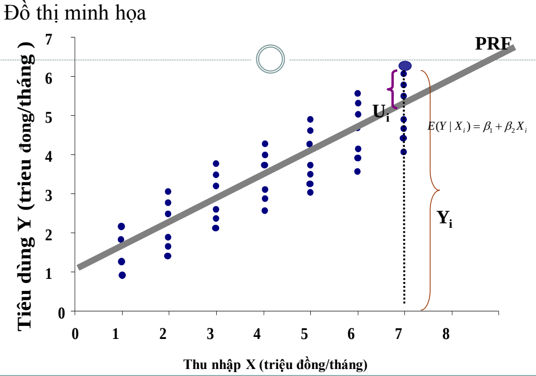
$$PRF : Y_i = \beta_1 + \beta_2 X_i + U_i$$

Trong đó

β_1,β_2 là các tham số của mô hình với ý nghĩa :

- β_1 : Tung độ gốc của hàm hồi quy tổng thể, là giá trị trung bình của biến phụ thuộc Y khi biến độc lập X nhận giá trị bằng 0
- β_2 : Độ dốc của hàm hồi quy tổng thể , là lượng thay đổi trung bình của Y khi X thay đổi 1 đơn vị

Đồ thị minh họa

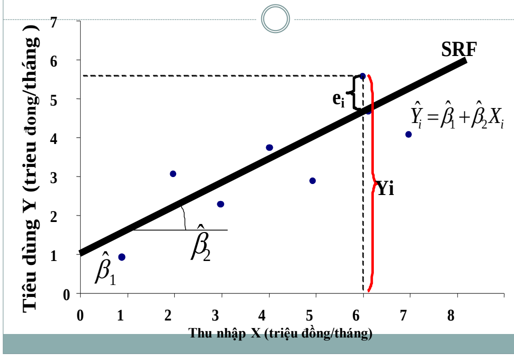


I. HỒI TUYẾN TÍNH 2 BIẾN

2. <u>Hàm hồi quy mẫu của hồi quy 2 biến</u>

Trong thực tế rất khó nghiên cứu trên tổng thể nên thông thường người ta nghiên cứu xây dựng hàm hồi quy trên một mẫu => Gọi là hàm hồi quy mẫu

Đồ thị minh họa



I. HỒI TUYẾN TÍNH 2 BIẾN

2. Hàm hồi quy mẫu của hồi quy 2 biến

$$SRF : Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + e_i$$

Trong đó

$\hat{\beta}_1$ Tung độ gốc của hàm hồi quy mẫu, là ước lượng điểm của β_1

$\hat{\beta}_2$ Độ dốc của hàm hồi quy mẫu, là ước lượng điểm của β_2

e_i Sai số ngẫu nhiên , là ước lượng điểm của U_i

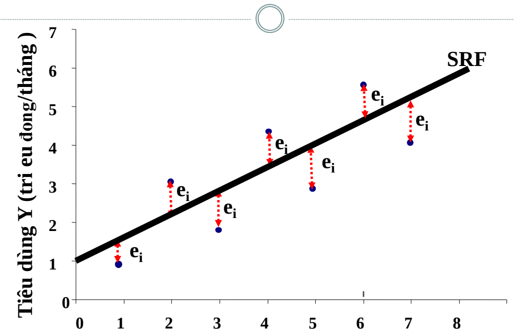
I. HỒI TUYẾN TÍNH 2 BIẾN

2. Hàm hồi quy mẫu của hồi quy 2 biến

$$SRF : Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + e_i$$

Nếu bỏ qua sai số ngẫu nhiên e_i , thì giá trị thực tế Y_i sẽ trở thành giá trị ước lượng $\hat{Y}_i$

$$SRF : \hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$



II. PHƯƠNG PHÁP BÌNH PHƯƠNG NHỎ NHẤT (OLS)

1. Ước lượng các tham số của mô hình

Giá trị thực tế $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + e_i$

Giá trị ước lượng $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$

Sai số $e_i = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i$

Tìm $\hat{\beta}_1, \hat{\beta}_2$ sao cho tổng bình phương sai số là nhỏ nhất

Tức là $\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i\right)^2 \to \min$

Tại sao chúng ta không tìm Σe_i nhỏ nhất ?

II. PHƯƠNG PHÁP BÌNH PHƯƠNG NHỎ NHẤT (OLS)

Giải bài toán cực trị hàm hai biến , ta được

$$\boxed{\begin{aligned} \hat{\beta}_2 &= \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{\sum_{i=1}^{n} Y_i X_i - n.\bar{X}.\bar{Y}}{\sum_{i=1}^{n} X_i^2 - n.(\bar{X})^2} = \frac{\sum x_i y_i}{\sum x_i^2} \\ \hat{\beta}_1 &= \bar{Y} - \hat{\beta}_2 \bar{X} \end{aligned}}$$

Với

$\bar{X} = \frac{\sum X_i}{n}$ là giá trị trung bình của X và $x_i = X_i - \bar{X}$

$\bar{Y} = \frac{\sum Y_i}{n}$ là giá trị trung bình của Y và $y_i = Y_i - \bar{Y}$

Câu hỏi

1. Hàm hồi quy mẫu có luôn đi qua điểm trung bình của mẫu$(\bar{X},\bar{Y})$ không? Vì sao?
2. Nếu X tăng 10 lần, Y không đổi thì $\hat{\beta}_1, \hat{\beta}_2$ sẽ thay đổi như thế nào ?
3. Nếu X tăng 10 lần, Y tăng 100 lần thì $\hat{\beta}_1, \hat{\beta}_2$ sẽ thay đổi như thế nào ?

Ví dụ áp dụng

Quan sát về thu nhập (X – triệu đồng/năm) và chi tiêu (Y – triệu đồng/năm) của 10 người, ta được các số liệu sau :

X	100	80	98	95	75	79	78	69	81	88
Y	90	75	78	88	62	69	65	55	60	70

Xây dựng hàm hồi quy mẫu $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$

II. PHƯƠNG PHÁP BÌNH PHƯƠNG NHỎ NHẤT (OLS)

2. Các giả thiết của OLS

Giả thiết 1 : Quan hệ giữa Y và X là tuyến tính
Các giá trị X_i cho trước và không ngẫu nhiên

Giả thiết 2 : Các sai số U_i là đại lượng ngẫu nhiên có giá trị trung bình bằng 0

$$E(U_i \mid X_i) = 0$$

II. PHƯƠNG PHÁP BÌNH PHƯƠNG NHỎ NHẤT (OLS)

2. Các giả thiết của OLS

Giả thiết 3 : Các sai số U_i là đại lượng ngẫu nhiên có phương sai không thay đổi

$$Var(U_i \mid X_i) = \sigma^2 = const$$

Giả thiết 4 : Không có sự tương quan giữa các U_i

$$Cov(U_i, U_j \mid X_i, X_j) = 0, i \neq j$$

Giả thiết 5 : Không có sự tương quan giữa U_i và X_i

$$Cov(U_i, X_i) = 0$$

II. PHƯƠNG PHÁP BÌNH PHƯƠNG NHỎ NHẤT (OLS)

2. Các giả thiết của OLS

Định lý Guass – Markov :
Khi các giả thiết này được đảm bảo thì các ước lượng tính được bằng phương pháp OLS là các **ước lượng tuyến tính không chệch, hiệu quả nhất** của hàm hồi quy tổng thể

ước lượng OLS là **BLUE**
(***B**est **L**inear **U**nbias **E**stimator*)

II. PHƯƠNG PHÁP BÌNH PHƯƠNG NHỎ NHẤT (OLS)

2. Các giả thiết của OLS

Giả thiết 6 : các sai số U_i có phân phối chuẩn

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

II. PHƯƠNG PHÁP BÌNH PHƯƠNG NHỎ NHẤT (OLS)

3. Hệ số xác định của mô hình

Tổng bình phương toàn phần TSS (Total Sum of Squares)

$$TSS = \sum (Y_i - \bar{Y})^2 = \sum Y_i^2 - n(\bar{Y})^2$$

Tổng bình phương hồi quy ESS (Explained Sum of Squares)

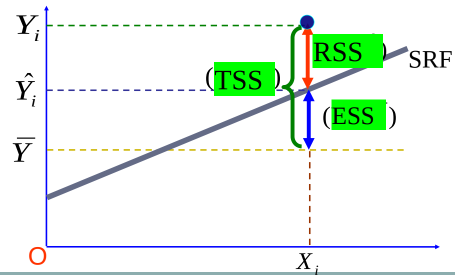
$$ESS = \sum (\hat{Y}_i - \bar{Y})^2 = \hat{\beta}_2^2 (\sum X_i^2 - n\bar{X}^2)$$

Tổng bình phương phần dư RSS (Residual Sum of Squares)

$$RSS = \sum (Y_i - \hat{Y}_i)^2 = \sum e_i^2$$

II. PHƯƠNG PHÁP BÌNH PHƯƠNG NHỎ NHẤT (OLS)

3. Hệ số xác định của mô hình



II. PHƯƠNG PHÁP BÌNH PHƯƠNG NHỎ NHẤT (OLS)

3. Hệ số xác định của mô hình

$$TSS = ESS + RSS$$ (Tại sao? -> Bài tập)

Hệ số xác định $R^2 = 1 - \frac{RSS}{TSS} = \frac{ESS}{TSS}$

- $0 \leq R^2 \leq 1$
- $R^2 = 1$: mô hình phù hợp hoàn toàn với mẫu nghiên cứu
- $R^2 = 0$: mô hình hoàn toàn không phù hợp với mẫu nghiên cứu

Ví dụ áp dụng

Từ số liệu đã cho của ví dụ trước , yêu cầu tính hệ số xác định của mô hình

III. KiỂM ĐỊNH MÔ HÌNH HỒI QUY

1. Các đại lượng ngẫu nhiên

*a. **Đại lượng ngẫu nhiên U_i***

Theo giả thiết của phương pháp OLS, U_i là đại lượng ngẫu nhiên có giá trị trung bình bằng 0 và phương sai không thay đổi

Ui ~ N(0,σ^2)

Khi đó σ^2 được gọi là phương sai của tổng thể , được ước lượng bằng phương sai mẫu

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n-2} = \frac{\sum (Y_i - \hat{Y}_i)^2}{n-2} = \frac{RSS}{n-2}$$

Vì sao chia n-2 ? => Bài tập

III. KiỂM ĐỊNH MÔ HÌNH HỒI QUY

1. Các đại lượng ngẫu nhiên

*a. **Đại lượng ngẫu nhiên U_i***

Ta có $Y_i = \beta_1 + \beta_2 X_i + U_i$

Vì **U_i ~ N(0 , σ^2)**

Nên **Y_i ~ N($\beta_1 + \beta_2 X_i$, σ^2)**

III. KiỂM ĐỊNH MÔ HÌNH HỒI QUY
1. Các đại lượng ngẫu nhiên

b. Đại lượng ngẫu nhiên $\hat{\beta}_1, \hat{\beta}_2$

Vì sao $\hat{\beta}_1, \hat{\beta}_2$ *là các đại lượng ngẫu nhiên ?*

$$\hat{\beta}_1 \sim N(\beta_1, \sigma^2_{\hat{\beta}_1})$$

$$\hat{\beta}_2 \sim N(\beta_2, \sigma^2_{\hat{\beta}_2})$$

Vì sao $\hat{\beta}_1, \hat{\beta}_2$ *có phân phối chuẩn ? => Bài tập*

Trong đó $\sigma^2_{\hat{\beta}_1}$ **là phương sai của** $\hat{\beta}_1$

$\sigma^2_{\hat{\beta}_2}$ **là phương sai của** $\hat{\beta}_2$

III. KiỂM ĐỊNH MÔ HÌNH HỒI QUY
1. Các đại lượng ngẫu nhiên

Với

$$\sigma^2_{\hat{\beta}_1} = \frac{\sum X_i^2}{n(\sum X_i^2 - n\bar{X}^2)}\sigma^2 \approx \frac{\sum X_i^2}{n(\sum X_i^2 - n\bar{X}^2)}\hat{\sigma}^2$$

$$\sigma^2_{\hat{\beta}_2} = \frac{\sigma^2}{\sum X_i^2 - n\bar{X}^2} \approx \frac{\hat{\sigma}^2}{\sum X_i^2 - n\bar{X}^2}$$

$se(\hat{\beta}_1) = \sqrt{\sigma^2_{\hat{\beta}_1}}$ sai số chuẩn của $\hat{\beta}_1$

$se(\hat{\beta}_2) = \sqrt{\sigma^2_{\hat{\beta}_2}}$ Sai số chuẩn của $\hat{\beta}_2$

III. KiỂM ĐỊNH MÔ HÌNH HỒI QUY
1. Các đại lượng ngẫu nhiên

Vì : $\hat{\beta}_1 \approx N(\beta_1, \sigma^2_{\hat{\beta}_1})$, $\hat{\beta}_2 \approx N(\beta_2, \sigma^2_{\hat{\beta}_2})$

Nên : $\frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} \approx N(0,1)$, $\frac{\hat{\beta}_2 - \beta_2}{se(\hat{\beta}_2)} \approx N(0,1)$

Nhưng do σ^2 ước lượng bằng $\hat{\sigma}^2$ dẫn đến

$$\frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} \approx T(n-2)$$

$$\frac{\hat{\beta}_2 - \beta_2}{se(\hat{\beta}_2)} \approx T(n-2)$$

Với *T(n-2)* là phân phối T-Student với bậc tự do *(n-2)*

Vì sao lại là phân phối t-Student?

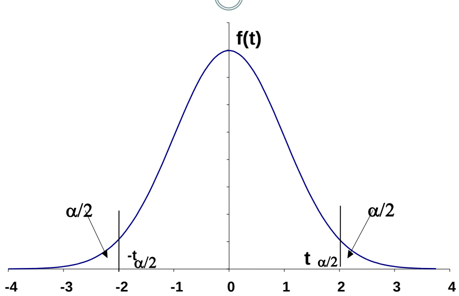
III. KiỂM ĐỊNH MÔ HÌNH HỒI QUY
2. Các khoảng tin cậy

a. Khoảng tin cậy của β_2

$$Ta \quad có \quad t = \frac{\hat{\beta}_2 - \beta_2}{se(\hat{\beta}_2)} \approx T(n-2)$$

Giả sử ta muốn xây dựng một khoảng giá trị của β_2 với độ tin cậy (1-α) .
Ví dụ (1-α) = 95% hay 0,95

Đồ thị phân phối của thống kê t



III. KiỂM ĐỊNH MÔ HÌNH HỒI QUY
2. Các khoảng tin cậy

a. Khoảng tin cậy của β_2

$$Vì \quad P\left(-t_{\alpha/2} \le \frac{\hat{\beta}_2 - \beta_2}{se(\hat{\beta}_2)} \le t_{\alpha/2}\right) = 1-\alpha$$

Nên khoảng tin cậy của β_2 với độ tin cậy 1-α là

$$\boxed{\left(\hat{\beta}_2 - t_{\frac{\alpha}{2}} \times se(\hat{\beta}_2); \hat{\beta}_2 + t_{\frac{\alpha}{2}} \times se(\hat{\beta}_2)\right)}$$

Với $t_{\alpha/2}$ *có được khi tra bảng t-Student với bậc tự do (n-2), mức ý nghĩa α/2*

III. KiỂM ĐỊNH MÔ HÌNH HỒI QUY

2. Các khoảng tin cậy

b. Khoảng tin cậy của β_1

Vì $t = \dfrac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} \approx T(n-2)$

Lập luận tương tự, khoảng tin cậy của β_1 với độ tin cậy 1-α là

$$\left(\hat{\beta}_1 - t_{\frac{\alpha}{2}} \times se(\hat{\beta}_1);\ \hat{\beta}_1 + t_{\frac{\alpha}{2}} \times se(\hat{\beta}_1)\right)$$

Giải thích ý nghĩa của độ tin cậy (1- α), ví dụ (1- α) =95%?

III. KiỂM ĐỊNH MÔ HÌNH HỒI QUY

2. Các khoảng tin cậy

c. Khoảng tin cậy của σ^2

Vì $\hat{\sigma}^2$ là ước lượng của σ^2 và người ta chứng minh được rằng

$$\frac{\hat{\sigma}^2(n-2)}{\sigma^2} \approx \chi^2(n-2)$$

Nên khoảng tin cậy của σ^2 với độ tin cậy 1-α là

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

Với $\chi^2_{\alpha/2}$ có được khi tra bảng χ^2 với bậc tự do (n-2), mức ý nghĩa α/2

Ví dụ áp dụng

Từ số liệu đã cho của ví dụ trước , yêu cầu tính khoảng tin cậy của β_1, β_2 và σ^2 với độ tin cậy 95%

III. KiỂM ĐỊNH MÔ HÌNH HỒI QUY

Nhắc lại về giả thiết H_0

Trong thống kê, giả thiết phát biểu cần được kiểm định được gọi là *giả thiết không* (ký hiệu : H_0). *Giả thiết đối* được ký hiệu là giả thiết H_1

	Báo bỏ H_0	Chấp nhận H_0
H_0 sai	*Đúng*	*Sai lầm loại II*
H_0 đúng	*Sai lầm loại I*	*Đúng*

Người ta thường đặt giả thiết H_0 sao cho sai lầm loại I là nghiêm trọng (nguy hiểm) hơn sai lầm loại II

III. KiỂM ĐỊNH MÔ HÌNH HỒI QUY

Đặt α là khả năng mắc sai lầm loại I

⇒ α là mức ý nghĩa của kiểm định

⇒ 1- α là độ tin cậy của kiểm định

Chú ý

- **Khi nói "chấp nhận giả thiết H_0", không có nghĩa H_0 đúng.**
- **Lựa chọn mức ý nghĩa α : α có thể tùy chọn, thường người ta chọn mức 1%, 5%, hoặc 10%.**

III. KiỂM ĐỊNH MÔ HÌNH HỒI QUY

Các giả thiết cần kiểm định gồm

- ***Các giả thiết về hệ số hồi quy***
- ***Các giả thiết về phương sai của U_i***
- ***Các giả thiết về sự phù hợp của mô hình***

Các loại giả thiết

- ***Giả thiết 2 phía , giả thiết phía trái và giả thiết phía phải***

Các cách kiểm định cơ bản :

- ***Phương pháp khoảng tin cậy***
- ***Phương pháp giá trị tới hạn***
- ***Phương pháp p-value (dùng máy vi tính)***

III. KiỂM ĐỊNH MÔ HÌNH HỒI QUY
3. Kiểm định giả thiết về hệ số hồi quy

a. Kiểm định giả thiết về β_2

Giả thiết 2 phía $H_o: \beta_2 = \beta_o$, $H_1: \beta_2 \neq \beta_o$ độ tin cậy là *1-α*

Giả thiết phía trái $H_o: \beta_2 = \beta_o$, $H_1: \beta_2 < \beta_o$

Giả thiết phía phải $H_o: \beta_2 = \beta_o$, $H_1: \beta_2 > \beta_o$

III. KiỂM ĐỊNH MÔ HÌNH HỒI QUY
3. Kiểm định giả thiết về hệ số hồi quy

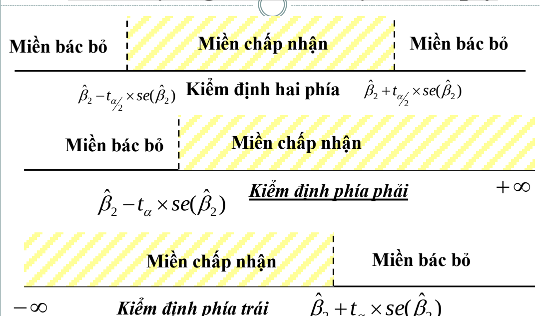
a. Kiểm định giả thiết về β_2

Phương pháp khoảng tin cậy

Bước 1 : Lập khoảng tin cậy của β_2

Bước 2 : Nếu β_0 thuộc khoảng tin cậy thì chấp nhận H_0.
Nếu β_0 không thuộc khoảng tin cậy thì bác bỏ H_0

III. KiỂM ĐỊNH MÔ HÌNH HỒI QUY
3. Kiểm định giả thiết về hệ số hồi quy



III. KiỂM ĐỊNH MÔ HÌNH HỒI QUY
2. Kiểm định giả thiết về hệ số hồi quy

a. Kiểm định giả thiết về β_2

Phương pháp giá trị tới hạn (kiểm định t)

Bước 1 : tính giá trị tới hạn $t = \dfrac{\hat{\beta}_2 - \beta_0}{se(\hat{\beta}_2)}$

Bước 2 : tra bảng t-Student với bậc tự do (n-2) tìm $t_{\alpha/2}$

Bước 3 :

Nếu $-t_{\alpha/2} \le t \le t_{\alpha/2}$: chấp nhận giả thiết H_0

Nếu $t < -t_{\alpha/2}$ hoặc $t > t_{\alpha/2}$: bác bỏ giả thiết H_0

SV tự suy luận điều kiện cho kiểm định phía trái và phải

III. KiỂM ĐỊNH MÔ HÌNH HỒI QUY
2. Kiểm định giả thiết về hệ số hồi quy

a. Kiểm định giả thiết về β_2

Phương pháp p-value

Bước 1 : tính giá trị tới hạn $t = \dfrac{\hat{\beta}_2 - \beta_0}{se(\hat{\beta}_2)}$

Bước 2 : Tính p_value = $P(|t| > |t_{\alpha/2}|)$
(tức là khả năng giả thiết H_0 bị bác bỏ)

Bước 3 :

Nếu p_value ≥ α : chấp nhận giả thiết H_0

Nếu p_value < α : bác bỏ giả thiết H_0

III. KiỂM ĐỊNH MÔ HÌNH HỒI QUY
2. Kiểm định giả thiết về hệ số hồi quy

b. Kiểm định giả thiết về β_1

$H_o: \beta_1 = \beta_o$, $H_1: \beta_1 \neq \beta_o$ Với độ tin cậy là *1-α*

Tương tự kiểm định giả thiết về β_2 nhưng giá trị tới hạn lúc này là

$$t = \frac{\hat{\beta}_1 - \beta_0}{se(\hat{\beta}_1)}$$

III. KiỂM ĐỊNH MÔ HÌNH HỒI QUY

2. Kiểm định giả thiết về hệ số hồi quy

c. ***Kiểm định giả thiết về σ^2***

$H_o: \sigma^2 = \sigma_0^2$
$H_1: \sigma^2 \neq \sigma_0^2$ Với độ tin cậy là *1-α*

Bước 1 : Lập khoảng tin cậy của σ^2

Bước 2 :
- Nếu σ_0^2 thuộc khoảng tin cậy thì chấp nhận H_0.
- Nếu σ_0^2 không thuộc khoảng tin cậy thì bác bỏ H_0

Ví dụ áp dụng

Từ số liệu đã cho của ví dụ trước , yêu cầu kiểm định các giả thiết sau

a) $H_o: \beta_2 = 0$, $H_1: \beta_2 \neq 0$ Với độ tin cậy là 95%

b) $H_o: \beta_1 = 0$, $H_1: \beta_1 \neq 0$ Với độ tin cậy là 95%

c) $H_o: \sigma^2 = 16$, $H_1: \sigma^2 \neq 16$ Với độ tin cậy là 95%

III. KiỂM ĐỊNH MÔ HÌNH HỒI QUY

4. Kiểm định sự phù hợp của mô hình

Kiểm định giả thiết

$H_o: R^2 = 0$
$H_1: R^2 \neq 0$ Với độ tin cậy là 1- α

Phương pháp kiểm định F

Bước 1 : tính $F = \dfrac{R^2(n-2)}{(1-R^2)}$

Bước 2 : Tra bảng tìm F(1,n-2), mức ý nghĩa là *α*

Bước 3 : Nếu F>F(1,n-2) , bác bỏ H_0
Nếu F≤F(1,n-2) , chấp nhận H_0

Câu hỏi

Việc kiểm định giả thiết $H_o: \beta_2 = 0$, $H_1: \beta_2 \neq 0$ ***độ tin cậy là (1-α) có ý nghĩa như thế nào?***

Việc kiểm định giả thiết $H_o: R^2 = 0$, $H_1: R^2 \neq 0$ ***độ tin cậy là (1-α) có ý nghĩa như thế nào?***

Ví dụ áp dụng

Từ số liệu đã cho của ví dụ trước , yêu cầu kiểm định sự phù hợp của mô hình với độ tin cậy 95%

5. Đánh giá kết quả hồi quy

- Dấu của các hệ số hồi qui ước lượng được phù hợp với lý thuyết hay tiên nghiệm không.
- Các hệ số hồi qui ước lượng được có ý nghĩa về mặt thống kê hay không ?
- Mức độ phù hợp của mô hình (R^2) và mô hình có thực sự phù hợp?
- Kiểm tra xem mô hình có thỏa mãn *các giả thiết của mô hình* hồi qui tuyến tính cổ điển hay không.

IV. SỬ DỤNG MÔ HÌNH HỒI QUY

1. Trình bày kết quả hồi quy

Kết quả hồi quy được trình bày như sau :

$\hat{Y}_i$	$=$	$\hat{\beta}_1$	$+$	$\hat{\beta}_2 X_i$	R^2
se		$se(\hat{\beta}_1)$		$se(\hat{\beta}_2)$	df
t		$t(\hat{\beta}_1)$		$t(\hat{\beta}_2)$	F_0
p_value		$p(\hat{\beta}_1)$		$p(\hat{\beta}_2)$	$p(F_0)$

IV. SỬ DỤNG MÔ HÌNH HỒI QUY

1. **Trình bày kết quả hồi quy**

Kết quả hồi quy trong ví dụ trước :

$\hat{Y}_i$	$=$	$-5{,}4517$	$+$	$0{,}9549 X_i$	$0{,}672$
se					
t					
p_value					

IV. SỬ DỤNG MÔ HÌNH HỒI QUY

2. **Vấn đề đổi đơn vị tính trong hàm hồi quy**

Trong hàm hồi quy hai biến , nếu đơn vị tính của X và Y thay đổi thì ta không cần hồi quy lại mà chỉ cần áp dụng công thức đổi đơn vị tính

Hàm hồi quy theo đơn vị tính cũ $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$

Hàm hồi quy theo đơn vị tính mới $\hat{Y}_i^* = \hat{\beta}_1^* + \hat{\beta}_2^* X_i^*$

Trong đó : $Y_i^* = k_1 Y_i$, $X_i^* = k_2 X_i$ Khi đó

$$\hat{\beta}_1^* = k_1 \hat{\beta}_1$$
$$\hat{\beta}_2^* = \frac{k_1}{k_2} \hat{\beta}_2$$

IV. SỬ DỤNG MÔ HÌNH HỒI QUY

2. Vấn đề đổi đơn vị tính trong hàm hồi quy

Ngoài ra :

$$\hat{\sigma}^{*2} = k_1^2 \hat{\sigma}^2$$
$$\sigma^2_{\hat{\beta}_1^*} = k_1^2 \sigma^2_{\hat{\beta}_1} \Rightarrow se(\hat{\beta}_1^*) = k_1 se(\hat{\beta}_1)$$
$$\sigma^2_{\hat{\beta}_2^*} = \frac{k_1^2}{k_2^2} \sigma^2_{\hat{\beta}_2} \Rightarrow se(\hat{\beta}_2^*) = \frac{k_1}{k_2} se(\hat{\beta}_2)$$

Tuy nhiên, việc thay đổi đơn vị tính của các biến không làm thay đổi tính BLUE của mô hình

Ví dụ áp dụng

Cho hàm hồi quy giữa lượng tiêu thụ cà phê (Y – *ly/ngày*) với giá bán cà phê (X – *ngàn đồng/kg*) như sau

$$\hat{Y}_i = 9 - 0{,}2 X_i$$

Viết lại hàm hồi quy nếu đơn vị tính của Y là *ly/tuần*

Ví dụ áp dụng

Từ số liệu đã cho của ví dụ trước về chi tiêu và thu nhập , yêu cầu viết lại hàm hồi quy với đơn vị tính như sau

a) Y – triệu đồng/tháng ; X – triệu đồng/năm

b) Y – triệu đồng/ tháng ; X – triệu đồng / tháng

c) Y – ngàn đồng/tháng ; X – ngàn đồng /tháng

IV. SỬ DỤNG MÔ HÌNH HỒI QUY
3. Vấn đề dự báo

Giả sử $SRF: \hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$

Khi $X=X_0$ thì ước lượng trung bình của Y_0 sẽ là

$$\hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_0$$

$\hat{Y}_0$ là đại lượng ngẫu nhiên có phân phối chuẩn

$$\hat{Y}_0 \sim N(\beta_1 + \beta_2 X_0, \sigma^2_{\hat{Y}_0})$$

Vì sao $\hat{Y}_0$ là đại lượng nhẫu nhiên ?
Tại sao có phân phối chuẩn ?

IV. SỬ DỤNG MÔ HÌNH HỒI QUY
3. Vấn đề dự báo

Với

$$\sigma^2_{\hat{Y}_0} = \sigma^2 \left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum X_i^2 - n(\bar{X})^2} \right]$$

$$se(\hat{Y}_0) = \sqrt{\sigma^2_{\hat{Y}_0}}$$

Khoảng tin cậy giá trị trung bình của Y_0 với độ tin cậy (1-α) là

$$\boxed{\left(\hat{Y}_0 - t_{\frac{\alpha}{2}} \times se(\hat{Y}_0); \hat{Y}_0 + t_{\frac{\alpha}{2}} \times se(\hat{Y}_0) \right)}$$

Ví dụ áp dụng

Từ số liệu đã cho của ví dụ trước , yêu cầu dự báo khoảng giá trị của Y khi $X_0 = 60$ (triệu đồng/năm) với độ tin cậy 95%

V. MỞ RỘNG MÔ HÌNH HỒI QUY HAI BiẾN
1. Hồi quy qua gốc tọa độ

Khi tung độ gốc bằng 0 thì mô hình trở thành mô hình hồi quy qua gốc tọa độ , khi đó hàm hồi quy như sau

$$PRF: Y_i = \beta_2 X_i + U_i$$

$$SRF: Y_i = \hat{\beta}_2 X_i + e_i$$

Với

$$\hat{\beta}_2 = \frac{\sum X_i Y_i}{\sum X_i^2} \qquad \text{Và} \qquad \sigma^2_{\hat{\beta}_2} = \frac{\sigma^2}{\sum X_i^2}$$

σ^2 được ước lượng bằng $\hat{\sigma}^2 = \frac{RSS}{n-1}$

V. MỞ RỘNG MÔ HÌNH HỒI QUY HAI BiẾN
1. Hồi quy qua gốc tọa độ

*Lưu ý :

- R^2 có thể âm đối với mô hình này, nên không dùng R^2 mà thay bởi $R^2_{thô}$:

$$R^2_{thô} = \frac{\left(\sum X_i Y_i\right)^2}{\sum X_i^2 \sum Y_i^2}$$

- Không thể so sánh R^2 với $R^2_{thô}$

Trên thực tế ít khi dùng đến mô hình hồi quy qua gốc tọa độ

V. MỞ RỘNG MÔ HÌNH HỒI QUY HAI BiẾN
2. Mô hình tuyến tính logarit

Hay còn gọi là mô hình *log-log* hay mô hình *log kép*

$$PRF: \ln Y_i = \beta_1 + \beta_2 \ln X_i + U_i$$

Mô hình không tuyến tính theo các biến nhưng có thể chuyển về dạng tuyến tính bằng cách đặt :

$$Y_i^* = \ln Y_i$$

$$X_i^* = \ln X_i$$

Khi đó $PRF: Y_i^* = \beta_1 + \beta_2 X_i^* + U_i$

Đây là dạng hồi quy tuyến tính đã biết

V. MỞ RỘNG MÔ HÌNH HỒI QUY HAI BiẾN

2. **Mô hình tuyến tính logarit**

Lấy đạo hàm 2 vế của hàm hồi quy log-log, ta được

$$\frac{Y'}{Y} = \beta_2 \frac{1}{X} \quad \Rightarrow \quad \beta_2 = Y'.\frac{X}{Y} = \frac{dY}{dX}.\frac{X}{Y}$$

Ý nghĩa của hệ số β_2 : ***khi X thay đổi 1% thì Y thay đổi β_2 % (Đây chính là hệ số co giãn của Y đối với X)***

V. MỞ RỘNG MÔ HÌNH HỒI QUY HAI BiẾN

3. **Mô hình log-lin**

$$PRF : \ln Y_i = \beta_1 + \beta_2 X_i + U_i$$

Mô hình không tuyến tính theo các biến nhưng có thể chuyển về dạng tuyến tính bằng cách đặt :

$$Y_i^* = \ln Y_i$$

Khi đó $PRF : Y_i^* = \beta_1 + \beta_2 X_i + U_i$

Biến phụ thuộc xuất hiện dưới dạng log và biến độc lập xuất hiện dưới dạng tuyến tính (linear) nên mô hình có tên gọi là log-lin

V. MỞ RỘNG MÔ HÌNH HỒI QUY HAI BiẾN

3. **Mô hình log-lin**

Ý nghĩa của hệ số β_2 : ***khi X thay đổi 1đơn vị thì Y thay đổi (100.β_2) %***

V. MỞ RỘNG MÔ HÌNH HỒI QUY HAI BiẾN

4. **Mô hình lin-log**

$$PRF : Y_i = \beta_1 + \beta_2 \ln X_i + U_i$$

Mô hình không tuyến tính theo các biến nhưng có thể chuyển về dạng tuyến tính bằng cách đặt :

$$X_i^* = \ln X_i$$

Khi đó $PRF : Y_i = \beta_1 + \beta_2 X_i^* + U_i$

V. MỞ RỘNG MÔ HÌNH HỒI QUY HAI BiẾN

4. **Mô hình lin-log**

Ý nghĩa của hệ số β_2 : ***khi X thay đổi 1 % thì Y thay đổi (β_2/100) đơn vị***

V. MỞ RỘNG MÔ HÌNH HỒI QUY HAI BiẾN

5. **Mô hình nghịch đảo**

$$PRF : Y_i = \beta_1 + \beta_2 \frac{1}{X_i} + U_i$$

Mô hình không tuyến tính theo các biến nhưng có thể chuyển về dạng tuyến tính bằng cách đặt :

$$X_i^* = \frac{1}{X_i}$$

Khi đó $PRF : Y_i = \beta_1 + \beta_2 X_i^* + U_i$

Ví dụ áp dụng

Từ số liệu đã cho của ví dụ trước , yêu cầu ước lượng hàm hồi quy

$$PRF : \ln Y_i = \beta_1 + \beta_2 \ln X_i + U_i$$

X_i	Y_i	$X_i^*=\ln X_i$	$Y_i^*=\ln Y_i$	$X_i^*Y_i^*$	X_i^{*2}
31	29	3.4340	3.3673	11.5633	11.7923
50	42	3.9120	3.7377	14.6218	15.3039
47	38	3.8501	3.6376	14.0052	14.8236
45	30	3.8067	3.4012	12.9472	14.4907
39	29	3.6636	3.3673	12.3363	13.4217
50	41	3.9120	3.7136	14.5276	15.3039
35	23	3.5553	3.1355	11.1478	12.6405
40	36	3.6889	3.5835	13.2192	13.6078
45	42	3.8067	3.7377	14.2280	14.4907
50	48	3.9120	3.8712	15.1442	15.3039
tổng	**cộng**	**37.5413**	**35.5525**	**133.7406**	**141.1791**
trung	**bình**	**3.7541**	**3.5553**		

Ví dụ áp dụng

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n} X_i^* - n.\overline{X}^*.\overline{Y}^*}{\sum_{i=1}^{n} X_i^{*2} - n.(\overline{X}^*)^2} = 1{,}1142$$

$$\hat{\beta}_1 = \overline{Y}^* - \hat{\beta}_2 \overline{X}^* = -0{,}6278$$

Kết quả hồi quy:

$$\hat{Y}_i^* = -0{,}6217 + 1{,}1142 X_i^*$$

$$\ln \hat{Y} = -0{,}6217 + 1{,}1142.\ln X_i$$

Cho kết quả hồi quy giữa Y – doanh số bán (trđ/tấn) và X - giá bán (ngàn đồng/kg) như sau :

$\hat{Y}$	$=$	$18{,}8503$	$-\ 1{,}0958 X_i$	$0{,}8681$
se		$1{,}5729$	$0{,}1743$	$df = 6$
t		$11{,}9837$	$-6{,}2842$	$39{,}49$

a) Nêu ý nghĩa kinh tế của các hệ số hồi quy
b) Xét xem giá bán có ảnh hưởng đến doanh số bán không ?(với mức ý nghĩa 1%)
c) Nếu giá bán là 8,5 ngàn đồng /kg thì doanh số bán trung bình là bao nhiêu?
d) Hãy viết lại SRF ở trên nếu đơn vị tính của Y là triệu đồng/năm
e) Kiểm định giả thiết H_0:β_2 = -1; H1 :$\beta_2 \neq$ -1; với mức ý nghĩa α=1%
f) Tính hệ số co giãn của Y theo X tại điểm $(\overline{X}, \overline{Y})$