Chương 2

MÔ HÌNH HỒI QUY HAI BIẾN

I. HÀM HỒI QUY TỔNG THỂ VÀ HÀM HỒI QUY MẪU

1. <u>Hàm hồi quy tổng thể của hồi quy 2 biến</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. HÀM HỒI QUY TỔNG THỂ VÀ HÀM HỒI QUY MẪU

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

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

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 biến độc lập
U_i : Sai số ngẫu nhiên ứng với quan sát thứ i

I. HÀM HỒI QUY TỔNG THỂ VÀ HÀM HỒI QUY MẪU

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

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

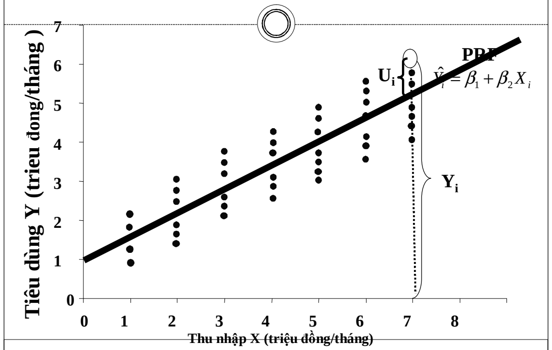
Trong đó

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

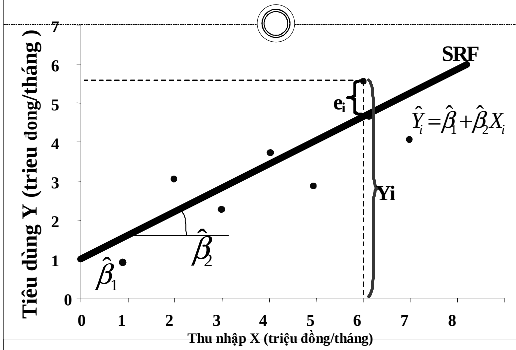


I. HÀM HỒI QUY TỔNG THỂ VÀ HÀM HỒI QUY MẪU

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. HÀM HỒI QUY TỔNG THỂ VÀ HÀM HỒI QUY MẪU

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

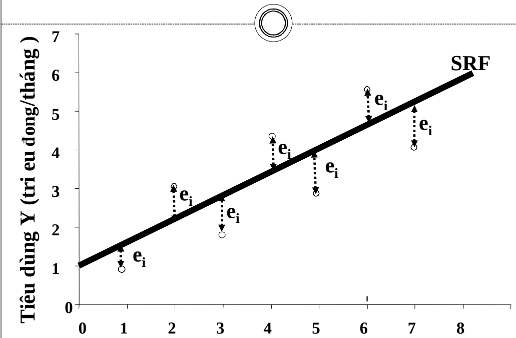
I. HÀM HỒI QUY TỔNG THỂ VÀ HÀM HỒI QUY MẪU

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

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

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}$$

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 $y_i = Y_i - \bar{Y}$

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_i	31	50	47	45	39	50	35	40	45	50
Y_i	29	42	38	30	29	41	23	36	42	48

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 mô hình

Giả thiết 1 : 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

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

II. PHƯƠNG PHÁP BÌNH PHƯƠNG NHỎ NHẤT (OLS)

2. Các giả thiết của mô hình

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

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

Khi các giả thiết này được đảm bảo thì các ước lượng tính được bằng phương pháp OLS là các ước lượng tốt nhất và hiệu quả nhất của hàm hồi quy tổng thể

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

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)

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

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

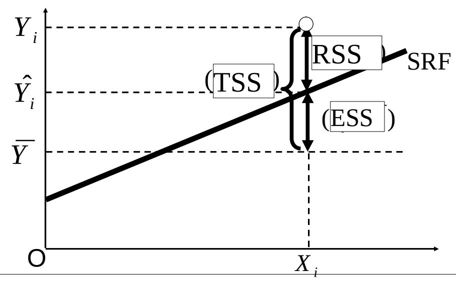
$$ESS = \sum (\hat{Y}_i - \overline{Y})^2 = \hat{\beta}_2^2 (\sum X_i^2 - n\overline{X}^2)$$

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

Người ta chứng minh được $TSS = ESS + RSS$

Hệ số xác định $R^2 = \dfrac{ESS}{TSS}$

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

II. PHƯƠNG PHÁP BÌNH PHƯƠNG NHỎ NHẤT (OLS)

4. Hệ số tương quan

Là số đo mức độ chặt chẽ của quan hệ tuyến tính giữa X và Y

$$r = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum (X_i - \overline{X})^2 \sum (Y_i - \overline{Y})^2}}$$

Ta chứng minh được : $|r| = \sqrt{R^2}$

Và dấu của r trùng với dấu của $\hat{\beta}_2$

II. PHƯƠNG PHÁP BÌNH PHƯƠNG NHỎ NHẤT (OLS)

Tính chất của hệ số tương quan :

- $-1 \leq r \leq 1$

  $|r| \rightarrow 1$: quan hệ tuyến tính giữa X và Y càng chặt chẽ.

- r có tính đối xứng : $r_{XY} = r_{YX}$
- Nếu X, Y độc lập thì r = 0. Điều ngược lại không đúng.

***Lưu ý* : ý nghĩa của hệ số tương quan khác xa ý nghĩa của R^2**