

Chương 1

NHẬP MÔN KINH TẾ LƯỢNG

by Tuan Anh (UEH)

1. LỊCH SỬ MÔN HỌC

Thuật ngữ “*Econometrics*” được sử dụng đầu tiên bởi Pawel Ciompa vào năm 1910

Tuy nhiên, mãi đến năm 1930, với các công trình nghiên cứu của [Ragnar Frisch](#) (Na Uy) thì thuật ngữ “*Econometrics*” mới được dùng đúng ý nghĩa như ngày hôm nay

Cùng khoảng thời gian này thì [Jan Tinbergen](#) (Hà Lan) cũng độc lập xây dựng các mô hình kinh tế lượng đầu tiên

Hai ông cùng được trao giải Nobel năm 1969 – giải Nobel kinh tế đầu tiên - với những nghiên cứu của mình về kinh tế lượng

by Tuan Anh (UEH)

1. LỊCH SỬ MÔN HỌC

Từ năm 1969 đến nay đã có 5 giải Nobel trao cho các nhà kinh tế lượng

- Jan Tinbergen, Ragnar Frisch - Năm 1969
- Lawrence Klein – năm 1980
- Trygve Haavelmo – năm 1989
- Daniel McFadden , James Heckman – năm 2000
- Robert Engle , Clive Granger - năm 2003
- Lars P. Hansen, Eugene Fama, Robert J Shiller (2013)

by Tuan Anh (UEH)

2. NỘI DUNG NGHIÊN CỨU

Econometrics – Kinh tế lượng

- Ước lượng, đo lường các mối quan hệ kinh tế
- Đối chiếu lý thuyết kinh tế với thực tiễn, qua đó **kiểm định** sự phù hợp của các lý thuyết kinh tế.
- **Dự báo** các biến số kinh tế.

by Tuan Anh (UEH)

3. CÁC MÔN HỌC LIÊN QUAN

- Kinh tế vi mô và kinh tế vĩ mô
- Toán học
- Xác suất
- Thống kê
- Tin học

by Tuan Anh (UEH)

4. HỒI QUY TRONG KINH TẾ LƯỢNG

a) Quan hệ hồi quy

Hồi quy nghiên cứu sự phụ thuộc của một đại lượng kinh tế này (*biến phụ thuộc*) vào một hay nhiều đại lượng kinh tế khác (*biến độc lập, biến giải thích*) dựa trên ý tưởng là *ước lượng giá trị trung bình* của biến phụ thuộc trên cơ sở các *giá trị biết trước* của các biến độc lập

Như vậy:

- *Biến độc lập* có giá trị xác định trước
- *Biến phụ thuộc* là đại lượng ngẫu nhiên tuân theo các quy luật phân bố xác suất

by Tuan Anh (UEH)

4. HỒI QUY TRONG KINH TẾ LƯỢNG

b) Phân biệt quan hệ hồi quy với các quan hệ khác

- Quan hệ hồi quy với quan hệ nhân quả
- Quan hệ hồi quy với quan hệ tương quan
- Quan hệ hồi quy với quan hệ hàm số

Hàm số : $Y = f(X)$

Hàm hồi quy : $Y = f(X) + U$

Với U là sai số

by Tuan Anh (UEH)

Vì sao sai số U luôn tồn tại trong mô hình hồi quy ?

- Vì không biết hết các yếu tố ảnh hưởng đến biến phụ thuộc Y
- Vì không thể đưa hết các yếu tố ảnh hưởng đến Y vào mô hình (sẽ làm mô hình phức tạp)
- Vì không có tất cả các số liệu cần thiết
- Vì sai sót và sai số trong quá trình thu thập số liệu

by Tuan Anh (UEH)

4. HỒI QUY TRONG KINH TẾ LƯỢNG

c) Hàm hồi quy tổng thể - PRF (Population Regression Function)

$$PRF : Y_i = f(X_{2i}, X_{3i}, \dots, X_{ki}) + U_i$$

Y : Biến phụ thuộc

Y_i : Giá trị thực tế cụ thể của biến phụ thuộc

X_2, X_3, \dots, X_k : Các biến độc lập

$X_{2i}, X_{3i}, \dots, X_{ki}$: Giá trị cụ thể của biến độc lập

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

by Tuan Anh (UEH)

4. HỒI QUY TRONG KINH TẾ LƯỢNG

c) Hàm hồi quy tổng thể - PRF (Population Regression Function)

$$PRF : Y_i = f(X_{2i}, X_{3i}, \dots, X_{ki}) + U_i$$

Hoặc :

$$E(Y | X_{2i}, X_{3i}, \dots, X_{ki}) = f(X_{2i}, X_{3i}, \dots, X_{ki})$$

Lưu ý : gần như không bao giờ có được hàm hồi quy tổng thể

by Tuan Anh (UEH)

4. HỒI QUY TRONG KINH TẾ LƯỢNG

d) Hàm hồi quy mẫu - SRF (Sample Regression Function)

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

$$SRF : Y_i = f(X_{2i}, X_{3i}, \dots, X_{ki}) + e_i$$

Với e_i là sai số trong mẫu, là phần dư, là ước lượng của U_i .

$$SRF : \hat{Y}_i = f(X_{2i}, X_{3i}, \dots, X_{ki})$$

by Tuan Anh (UEH)

Chương 2

MÔ HÌNH HỒI QUY HAI BIẾN

I. HỒI TUYẾN TÍNH 2 BIẾN

1. Hàm hồi quy tuyến tính 2 biến của tổng thể

Trong quan hệ hồi quy, một biến phụ thuộc có thể được giải thích bởi nhiều biến độc lập

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

Nếu mối quan hệ giữa hai biến này là tuyến tính \Rightarrow *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:

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

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

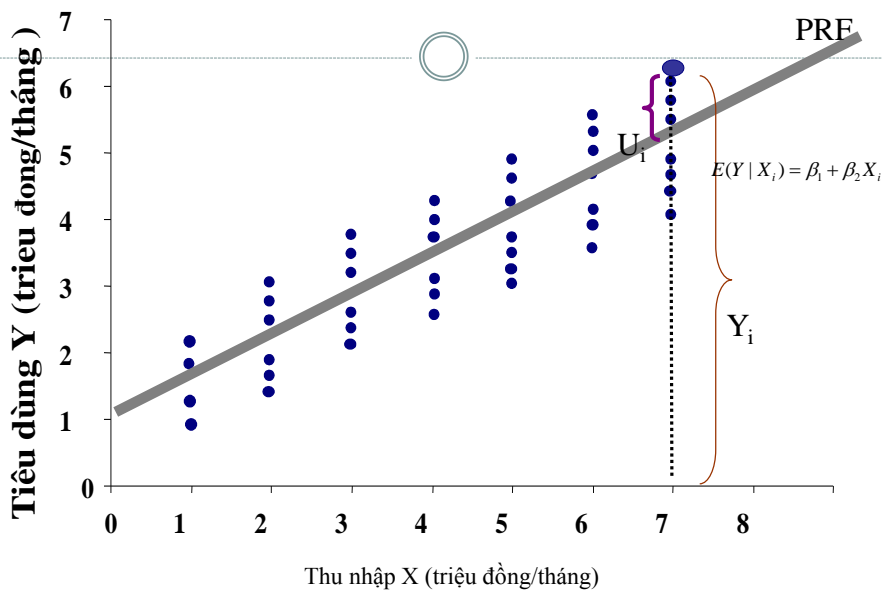
$$PRF : Y_i = \beta_1 + \beta_2 X_i + U_i$$

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

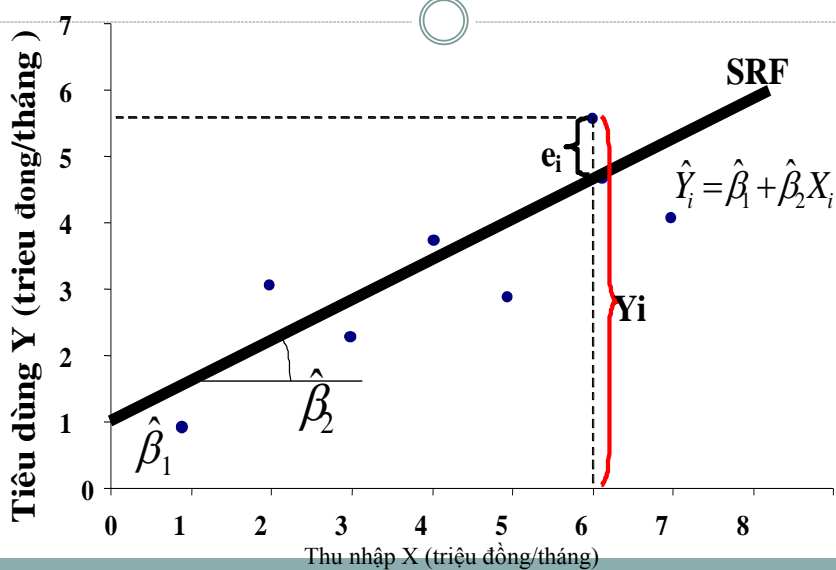


I. HỒI TUYẾN TÍNH 2 BIẾN

2. Hàm hồi quy mẫu của hồi quy 2 biến

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

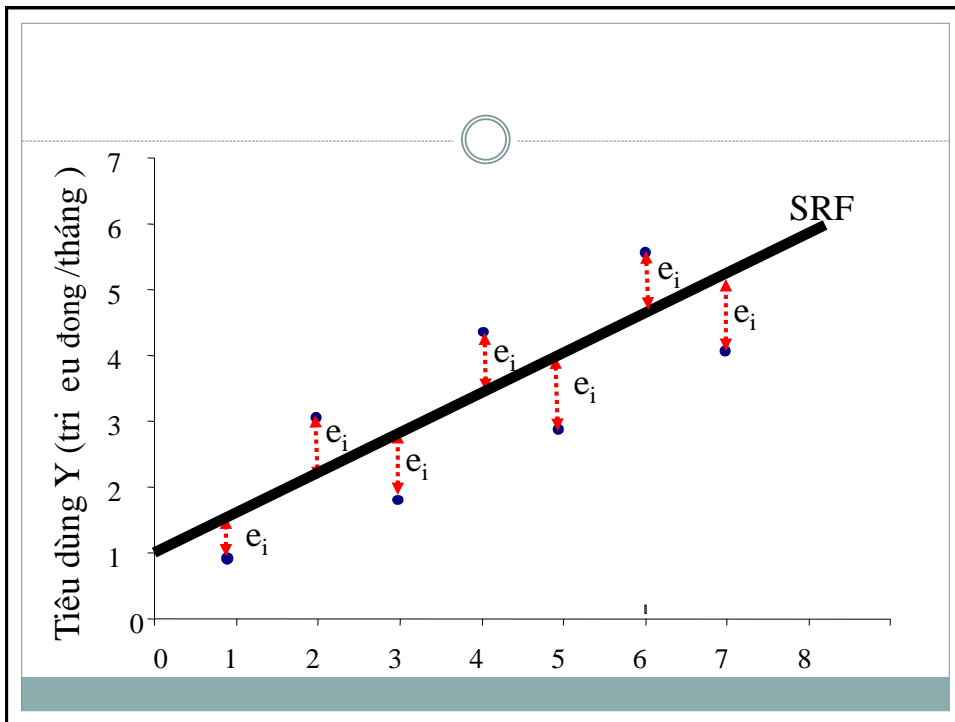


2. Hàm hồi quy mẫu của hồi quy 2 biến

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

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 $Y_i = \beta_1 + \beta_2 X_i + e_i$

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 $\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2 \rightarrow \min$

Tại sao chúng ta không tìm $\sum e_i$ nhỏ nhất ?

II. PHƯƠNG PHÁP BÌNH PHƯƠNG NHỎ NHẤT (OLS)

Giải bài toán cực trị hàm hai biến , ta được

$$\hat{\beta}_2 = \frac{\sum_{i=1}^n X_i Y_i - n.\bar{X}.\bar{Y}}{\sum_{i=1}^n X_i^2 - n.(\bar{X})^2}$$
$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}$$

Với

$$\bar{X} = \frac{\sum X_i}{n}$$

là giá trị trung bình của X

$$\bar{Y} = \frac{\sum Y_i}{n}$$

là giá trị trung bình của Y

II. PHƯƠNG PHÁP BÌNH PHƯƠNG NHỎ NHẤT (OLS)

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

$$E(U_i | X_i) = 0$$

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

$$\text{Var}(U_i | X_i) = \sigma^2 = \text{const}$$

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

$$\text{Cov}(U_i, U_j | X_i, X_j) = 0, i \neq j$$

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

$$\text{Cov}(U_i, X_i) = 0$$

II. PHƯƠNG PHÁP BÌNH PHƯƠNG NHỎ NHẤT (OLS)

2. Các giả thiết của OLS

Định lý Gauss – 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**
(*Best Linear Unbiased Estimator*)

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)

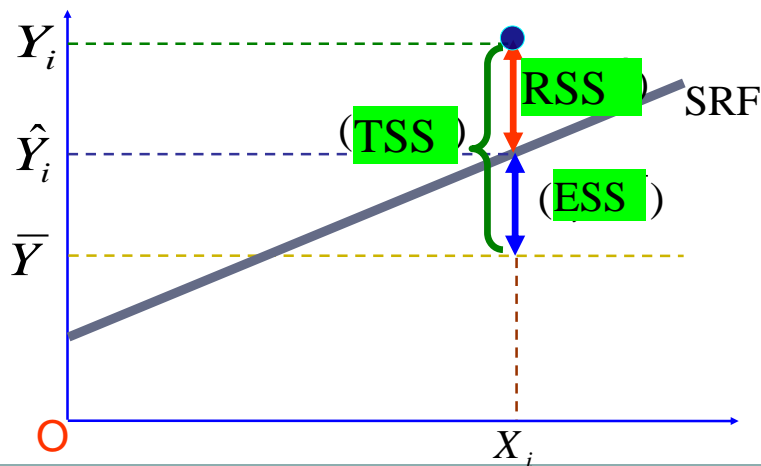
$$ESS = \sum (\hat{Y}_i - \bar{Y})^2 = \hat{\beta}_2^2 (\sum X_i^2 - n\bar{X}^2)$$

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

$$TSS = ESS + RSS$$

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

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

II. PHƯƠNG PHÁP BÌNH PHƯƠNG NHỎ NHẤT (OLS)

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

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

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

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

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

Nên
$$Y_i \sim N(\beta_1 + \beta_2 X_i, \sigma^2)$$

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 ?

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

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

Trong đó $\sigma_{\hat{\beta}_1}^2$ là phương sai của $\hat{\beta}_1$

$\sigma_{\hat{\beta}_2}^2$ là phương sai của $\hat{\beta}_2$

III. KIỂM ĐỊNH MÔ HÌNH HỒI QUY

1. Các đại lượng ngẫu nhiên

Với

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

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

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

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

III. KIỂM ĐỊNH MÔ HÌNH HỒI QUY

2. Các khoảng tin cậy

a. Khoảng tin cậy của β_2

Khoảng tin cậy của β_2 với độ tin cậy $1-\alpha$ là

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

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

III. KIỂM ĐỊNH MÔ HÌNH HỒI QUY

2. Các khoảng tin cậy

b. Khoảng tin cậy của β_1

Khoảng tin cậy của β_1 với độ tin cậy $1-\alpha$ 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 $(1-\alpha)$, ví dụ $(1-\alpha)=95\%$?

Ví dụ áp dụng

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

III. KIỂM ĐỊNH MÔ HÌNH HỒI QUY

Nhắc lại về giả thuyết H_0

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

	Bảo bỏ H_0	Chấp nhận H_0
H_0 sai	<i>Đúng</i>	<i>Sai lầm loại II</i>
H_0 đúng	<i>Sai lầm loại I</i>	<i>Đúng</i>

Người ta thường đặt giả thuyết 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

$\Rightarrow \alpha$ là mức ý nghĩa của kiểm định

$\Rightarrow 1 - \alpha$ là độ tin cậy của kiểm định

Chú ý

- Khi nói “chấp nhận giả thuyết 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 thuyết cần kiểm định gồm

- Các giả thuyết về hệ số hồi quy
- Các giả thuyết về sự phù hợp của mô hình

Các cách kiểm định cơ bản :

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

III. KIỂM ĐỊNH MÔ HÌNH HỒI QUY

3. Kiểm định giả thuyết về hệ số hồi quy

a. Kiểm định giả thuyết về β_2

Giả thuyết $H_0: \beta_2 = \beta_0$ $H_1: \beta_2 \neq \beta_0$ mức ý nghĩa α

Cách 1: 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 thuyết về hệ số hồi quy

a. Kiểm định giả thuyết về β_2

Cách 2 : Phương pháp giá trị tới hạn (kiểm định t)

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

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

Bước 3 :

Nếu $|t| \geq t_{\alpha/2}$: chấp nhận giả thuyết H_0

Nếu $|t| < t_{\alpha/2}$: bác bỏ giả thuyết H_0

III. KIỂM ĐỊNH MÔ HÌNH HỒI QUY

2. Kiểm định giả thuyết về hệ số hồi quy

a. Kiểm định giả thuyết về β_2

Cách 3 : Phương pháp p-value

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

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

Bước 3 :
Nếu p_value $\geq \alpha$: chấp nhận giả thuyết H_0
Nếu p_value $< \alpha$: bác bỏ giả thuyết H_0

III. KIỂM ĐỊNH MÔ HÌNH HỒI QUY

2. Kiểm định giả thuyết về hệ số hồi quy

b. Kiểm định giả thuyết về β_1

$$\begin{aligned} H_0: \beta_1 &= \beta_0 \\ H_1: \beta_1 &\neq \beta_0 \end{aligned} \quad \text{Với độ tin cậy là } 1-\alpha$$

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

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

Ví dụ áp dụng

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

- a) $H_0: \beta_2 = 0$
 $H_1: \beta_2 \neq 0$ Với độ tin cậy là 95%
- b) $H_0: \beta_1 = 0$
 $H_1: \beta_1 \neq 0$ Với độ tin cậy là 95%

III. KIỂM ĐỊNH MÔ HÌNH HỒI QUY

3. Kiểm định sự phù hợp của mô hình

Kiểm định giả thuyết

$$\begin{array}{l} H_0: R^2 = 0 \\ H_1: R^2 \neq 0 \end{array} \quad \text{Với độ tin cậy là } 1 - \alpha$$

Phương pháp kiểm định F

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

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 $F \leq F(1, n-2)$, chấp nhận H_0

Câu hỏi



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

Việc kiểm định giả thuyết $H_o: R^2 = 0$
 $H_1: R^2 \neq 0$ độ tin cậy là $(1-\alpha)$
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%

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: $Y_i^* = k_1 Y_i$ Khi đó

$$X_i^* = k_2 X_i$$

$$\hat{\beta}_1^* = k_1 \hat{\beta}_1$$

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

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

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

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

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

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

IV. SỬ DỤNG MÔ HÌNH HỒI QUY

3. Vấn đề dự báo

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

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

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

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

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

$$\hat{\beta}_2 = \frac{\sum X_i Y_i}{\sum X_i^2} \quad \text{và} \quad \sigma_{\hat{\beta}_2}^2 = \frac{\sigma^2}{\sum X_i^2}$$

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

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_{tho} :

$$R^2_{tho} = \frac{(\sum X_i Y_i)^2}{\sum X_i^2 \sum Y_i^2}$$

- Không thể so sánh R^2 với R^2_{tho}

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

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

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 1 đơn vị thì Y thay đổi $(100.\beta_2) \%$*

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 $(\beta_2/100)$ 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

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

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

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

- Nêu ý nghĩa kinh tế của các hệ số hồi quy
- Xét xem giá bán có ảnh hưởng đến doanh số bán không ?(với mức ý nghĩa 1%)
- Nếu giá bán là 8,5 ngàn đồng /kg thì doanh số bán trung bình là bao nhiêu?
- Hãy viết lại SRF ở trên nếu đơn vị tính của Y là triệu đồng/năm
- Kiểm định giả thuyết $H_0: \beta_2 = -1$; $H_1: \beta_2 \neq -1$; với mức ý nghĩa $\alpha=1\%$
- Tính hệ số co giãn của Y theo X tại điểm (\bar{X}, \bar{Y})

Chương 3

HỒI QUY TUYẾN TÍNH BỘI

1. HÀM HỒI QUY TỔNG THỂ (PRF)

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + U_i$$

Trong đó

- Y là biến phụ thuộc
- X_2, X_3, \dots, X_k là các biến độc lập
- U_i là các sai số ngẫu nhiên
- β_1 : Hệ số tự do
- $\beta_2, \beta_3, \dots, \beta_k$ là các hệ số hồi quy riêng

2. CÁC GIẢ THIẾT CỦA MÔ HÌNH

Giả thiết 1 : Các biến độc lập X_2, X_3, \dots, X_k không ngẫu nhiên

Giả thiết 2 : Các sai số ngẫu nhiên U_i có giá trị trung bình bằng 0 và có phương sai không thay đổi

$$E(U_i | X) = 0 \quad \text{Var}(U_i | X) = \sigma^2$$

Giả thiết 3: Không có sự tương quan giữa các sai số U_i

$$\text{Cov}(U_i, U_j | X) = 0, i \neq j$$

2. CÁC GIẢ THIẾT CỦA MÔ HÌNH

Giả thiết 4 : Không có hiện tượng cộng tuyến giữa các biến độc lập X_2, X_3, \dots, X_k

Giả thiết 5 : Không có tương quan giữa các biến độc lập X_2, X_3, \dots, X_k với các sai số ngẫu nhiên U_i

$$Cov(U, X) = 0$$

3. ƯỚC LƯỢNG CÁC HỆ SỐ HỒI QUY

Hàm hồi quy mẫu :

$$SRF: Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \dots + \hat{\beta}_k X_{ki} + e_i$$

$$hoặc: \hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \dots + \hat{\beta}_k X_{ki}$$

3. ƯỚC LƯỢNG CÁC HỆ SỐ HỒI QUY

$$\text{SRF: } Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \dots + \hat{\beta}_k X_{ki} + e_i$$

$$\text{hoặc: } \hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \dots + \hat{\beta}_k X_{ki}$$

Khi đó

$$\begin{aligned} e_i &= (Y_i - \hat{Y}_i) \\ &= Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i} - \dots - \hat{\beta}_k X_{ki} \end{aligned}$$

3. ƯỚC LƯỢNG CÁC HỆ SỐ HỒI QUY

Theo nguyên lý của phương pháp OLS thì các tham số

$\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \dots, \hat{\beta}_k$ được chọn sao cho

$$\sum e_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i} - \dots - \hat{\beta}_k X_{ki})^2$$

$\rightarrow \min$

Các tính toán sẽ được hỗ trợ bằng Eviews

3. ƯỚC LƯỢNG CÁC HỆ SỐ HỒI QUY



Ví dụ minh họa

Bảng dưới đây cho các số liệu về doanh số bán (Y), chi phí chào hàng (X_2) và chi phí quảng cáo (X_3) của một công ty

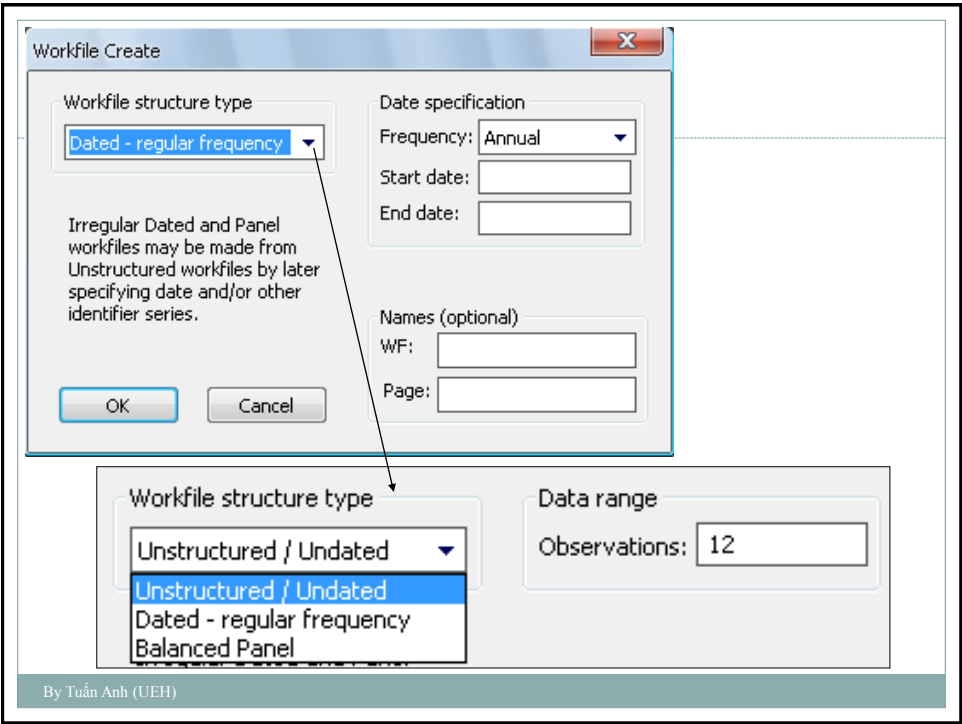
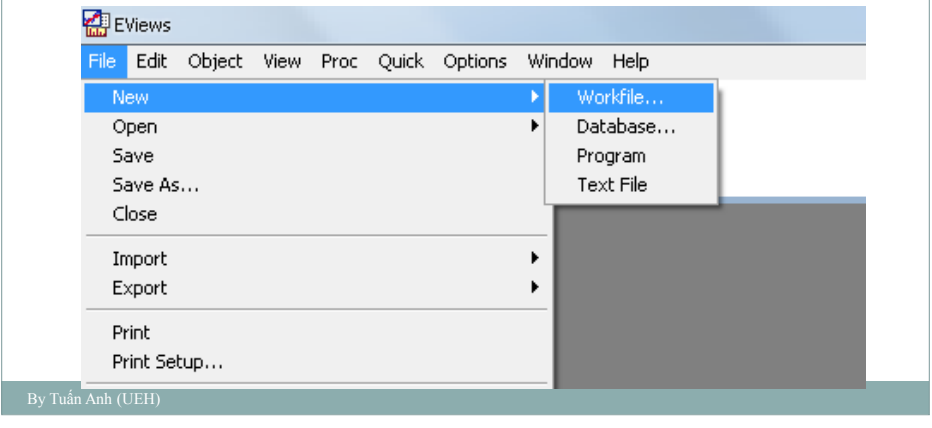
Hãy ước lượng hàm hồi quy tuyến tính của doanh số bán theo chi phí chào hàng và chi phí quảng cáo

Doanh số bán Y_i (trđ)	Chi phí chào hàng X_2	Chi phí quảng cáo X_3
1270	100	180
1490	106	248
1060	60	190
1626	160	240
1020	70	150
1800	170	260
1610	140	250
1280	120	160
1390	116	170
1440	120	230
1590	140	220
1380	150	150

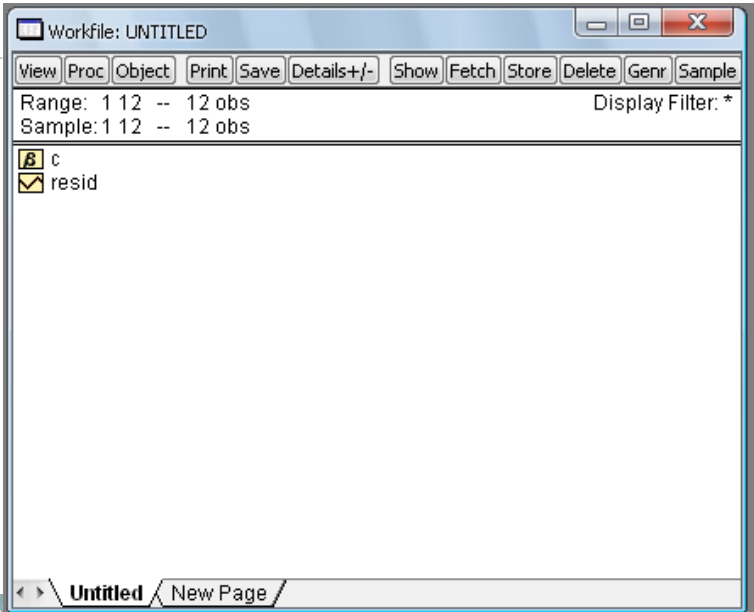
Hướng dẫn hồi quy bằng Eviews

Tập tin chứa số liệu trong Eviews được gọi là Workfile và có phần mở rộng là “.wfl”

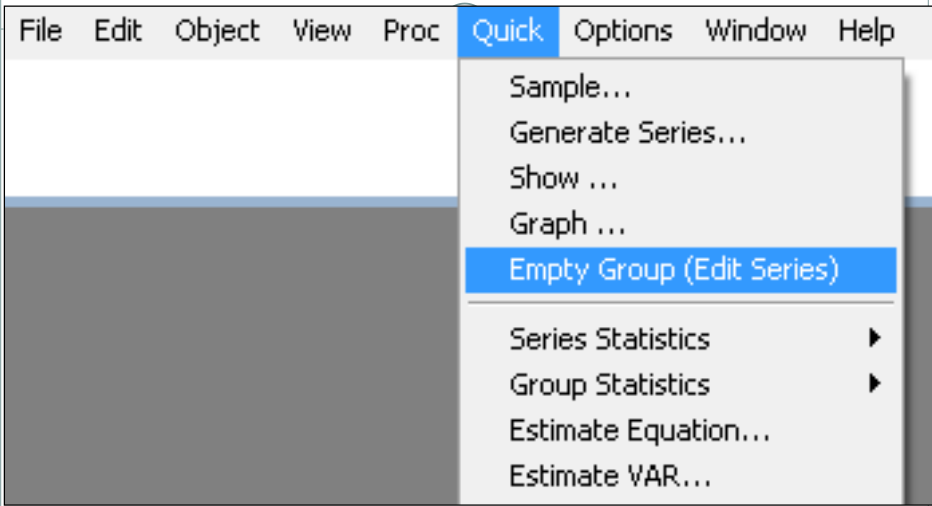
Thao tác tạo file như sau : **File\New\Workfile**



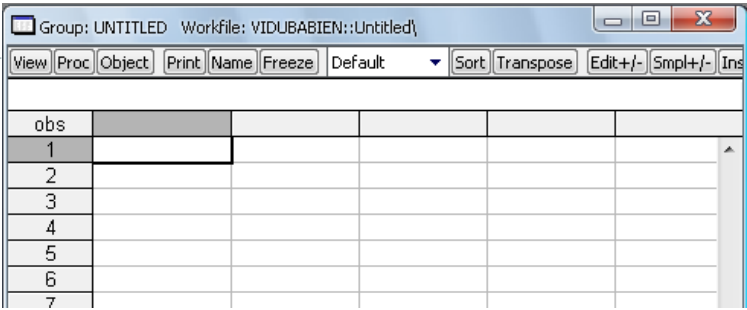
Sao khi tạo thì workfile sẽ hiện ra như sau



Nhập số liệu thì vào **Quick\Empty Group(Edit series)**



Màn hình hiện ra



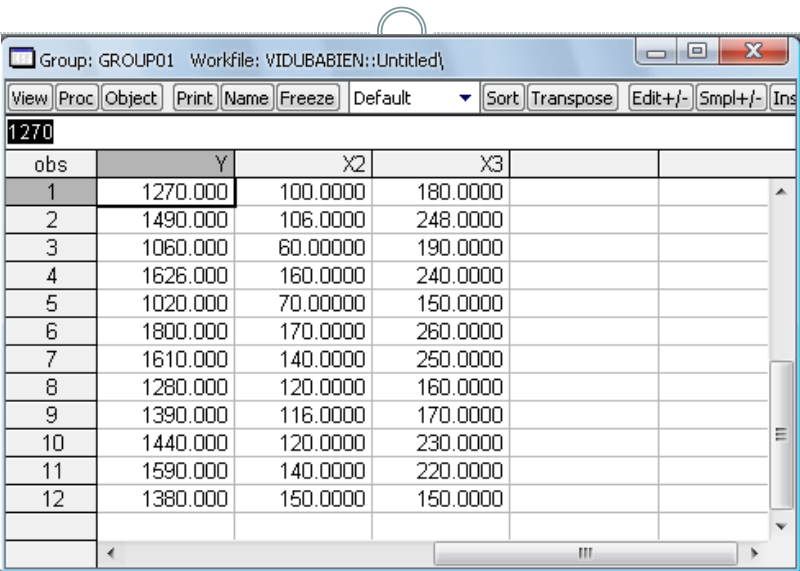
The screenshot shows the EViews software interface. The title bar indicates 'Group: UNTITLED' and 'Workfile: VIDUBABIEN::Untitled'. The menu bar includes 'View', 'Proc', 'Object', 'Print', 'Name', 'Freeze', 'Default', 'Sort', 'Transpose', 'Edit+/-', 'Smpl+/-', and 'Ins'. The main workspace displays a table with 7 rows and 6 columns. The first column is labeled 'obs' and contains the numbers 1 through 7. The other columns are empty.

obs					
1					
2					
3					
4					
5					
6					
7					

Lưu ý : nhấp chuột vào ô đầu tiên dòng 1, nhấn phím mũi tên lên trên bàn phím trước khi paste dữ liệu vào Eviews

By Tuấn Anh (UEH)

Nhập số liệu vào theo cột

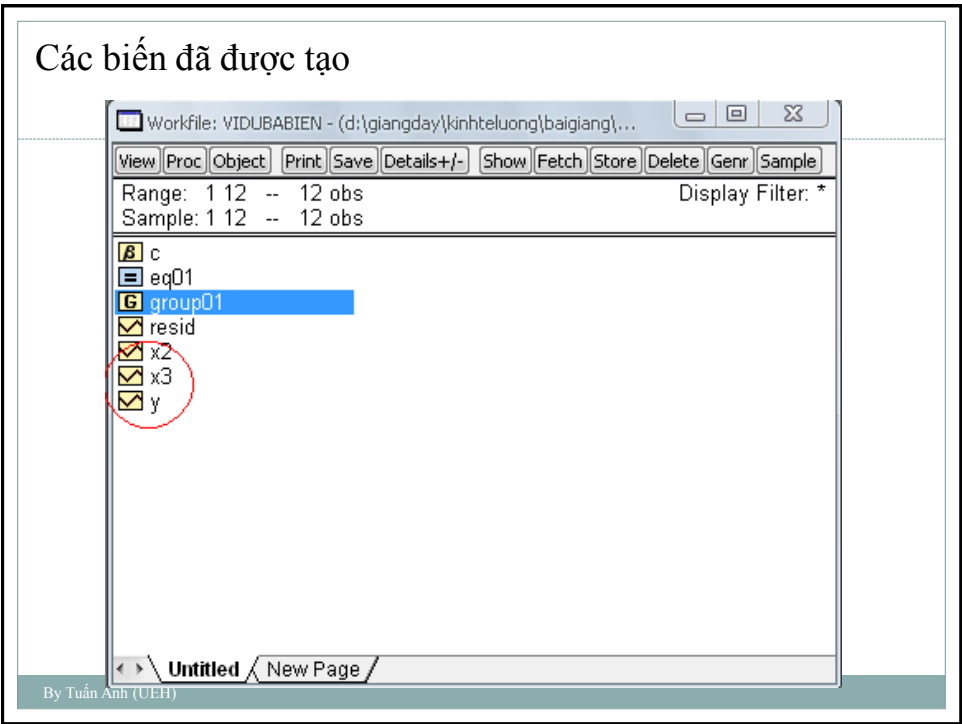


The screenshot shows the EViews software interface with data entered into the workspace. The title bar indicates 'Group: GROUP01' and 'Workfile: VIDUBABIEN::Untitled'. The menu bar is the same as in the previous screenshot. The main workspace displays a table with 12 rows and 6 columns. The first column is labeled 'obs' and contains the numbers 1 through 12. The second column is labeled 'Y' and contains values ranging from 1060.000 to 1626.000. The third column is labeled 'X2' and contains values ranging from 60.00000 to 180.0000. The fourth column is labeled 'X3' and contains values ranging from 150.0000 to 250.0000. The other columns are empty.

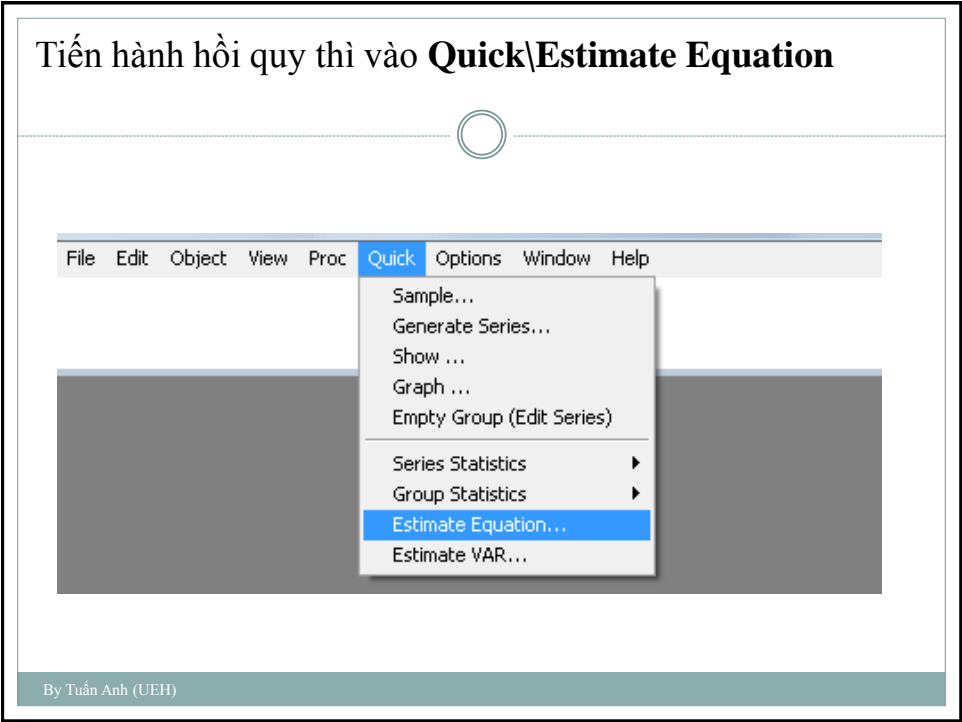
obs	Y	X2	X3		
1	1270.000	100.0000	180.0000		
2	1490.000	106.0000	248.0000		
3	1060.000	60.00000	190.0000		
4	1626.000	160.0000	240.0000		
5	1020.000	70.00000	150.0000		
6	1800.000	170.0000	260.0000		
7	1610.000	140.0000	250.0000		
8	1280.000	120.0000	160.0000		
9	1390.000	116.0000	170.0000		
10	1440.000	120.0000	230.0000		
11	1590.000	140.0000	220.0000		
12	1380.000	150.0000	150.0000		

By Tuấn Anh (UEH)

Các biến đã được tạo



Tiến hành hồi quy thì vào **Quick\Estimate Equation**



Nhập dạng hàm cần hồi quy

Equation Estimation

Specification Options

Equation specification
Dependent variable followed by list of regressors including ARMA and PDL terms, OR an explicit equation like $Y=c(1)+c(2)*X$.

y c x2 x3

Estimation settings
Method: LS - Least Squares (NLS and ARMA)
Sample: 1 12

OK Cancel

By Tuấn Anh (UEH)

Kết quả hồi quy

Equation: UNTITLED Workfile: VIDUBABIEN::Untitled

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: Y
Method: Least Squares
Date: 03/31/09 Time: 08:22
Sample: 1 12
Included observations: 12

	Coefficient	Std. Error	t-Statistic	Prob.
C	328.1383	71.99136	4.558023	0.0014
X2	4.649510	0.469146	9.910592	0.0000
X3	2.560152	0.379411	6.747707	0.0001

R-squared	0.967693	Mean dependent var	1413.000
Adjusted R-squared	0.960514	S.D. dependent var	231.7420
S.E. of regression	46.04989	Akaike info criterion	10.70965
Sum squared resid	19085.33	Schwarz criterion	10.83087
Log likelihood	-61.25787	Hannan-Quinn criter.	10.66476
F-statistic	134.7884	Durbin-Watson stat	2.493309
Prob(F-statistic)	0.000000		

By Tuấn Anh (UEH)

4. HỆ SỐ XÁC ĐỊNH CỦA MÔ HÌNH



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

Hệ số xác định hiệu chỉnh: $\bar{R}^2 = 1 - (1 - R^2) \frac{n-1}{n-k}$

Vì sao cần tính hệ số xác định hiệu chỉnh ?

4. HỆ SỐ XÁC ĐỊNH CỦA MÔ HÌNH



\bar{R}^2 có các đặc điểm sau :

- ❖ Khi $k > 1$ thì $\bar{R}^2 \leq R^2 \leq 1$
- ❖ \bar{R}^2 có thể âm, và khi nó âm, coi như bằng 0

Đọc kết quả hồi quy về hệ số xác định

Equation: UNTITLED Workfile: VIDUBABIEN::Untitled\

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: Y
 Method: Least Squares
 Date: 03/31/09 Time: 08:22
 Sample: 1 12
 Included observations: 12

	Coefficient	Std. Error	t-Statistic	Prob.
C	328.1383	71.99136	4.558023	0.0014
X2	4.649510	0.469146	9.910592	0.0000
X3	2.560152	0.379411	6.747707	0.0001

R-squared	0.967693	Mean dependent var	1413.000
Adjusted R-squared	0.960514	S.D. dependent var	231.7420
S.E. of regression	46.04989	Akaike info criterion	10.70965
Sum squared resid	19085.33	Schwarz criterion	10.83087
Log likelihood	-61.25787	Hannan-Quinn criter.	10.66476
F-statistic	134.7884	Durbin-Watson stat	2.493309
Prob(F-statistic)	0.000000		

By Tuấn Anh (UEH)

5. KIỂM ĐỊNH GIẢ THUYẾT VỀ HỆ SỐ HỒI QUY



Giả thuyết

$$H_0: \beta_j = \beta_0$$

$$H_1: \beta_j \neq \beta_0$$

mức ý nghĩa α

Cách 1: Phương pháp khoảng tin cậy

Bước 1: Lập khoảng tin cậy của β_j

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

$t_{\alpha/2}$ tra bảng t-Student với bậc tự do $(n-k)$, mức ý nghĩa $\alpha/2$

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

5. KIỂM ĐỊNH GIẢ THUYẾT VỀ HỆ SỐ HỒI QUY



Cách 2 : Phương pháp giá trị tới hạn (kiểm định t)

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

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

Bước 3 :

Nếu $|t| \geq t_{\alpha/2}$: chấp nhận giả thuyết H_0

Nếu $|t| < t_{\alpha/2}$: bác bỏ giả thuyết H_0

5. KIỂM ĐỊNH GIẢ THUYẾT VỀ HỆ SỐ HỒI QUY



Cách 3 : Phương pháp p-value

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

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

Bước 3 :

Nếu $p_value \geq \alpha$: chấp nhận giả thuyết H_0

Nếu $p_value < \alpha$: bác bỏ giả thuyết H_0

Một vài kết quả hồi quy khác bằng Eviews

Dependent Variable: Y_CNHN
Method: Least Squares
Date: 04/01/09 Time: 10:23
Sample: 1 120
Included observations: 120

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.477215	2.207579	-1.575126	0.1180
X2_CNMS	0.080708	0.444005	0.181772	0.8561
X3_COM	0.543629	0.382117	1.422675	0.1576
X4_SUA	0.005008	0.001267	3.951200	0.0001
X5_CD	0.086614	0.009635	8.989518	0.0000
X6_RAU	0.011712	0.015035	0.779036	0.4376
X7_CB	0.143078	0.019623	7.291392	0.0000
R-squared	0.702570	Mean dependent var	17.42500	
Adjusted R-squared	0.686778	S.D. dependent var	3.364902	
S.E. of regression	1.883210	Akaike info criterion	4.160395	
Sum squared resid	400.7523	Schwarz criterion	4.322999	
Log likelihood	-242.6237	Hannan-Quinn criter.	4.226430	
F-statistic	44.48696	Durbin-Watson stat	1.881293	
Prob(F-statistic)	0.000000			

Theo kết quả bài tập của nhóm 13 lớp KK1_05 trường Đại học Hồng Bàng

Các yếu tố ảnh hưởng đến giá bán 1 căn nhà

Dependent Variable: Y
Method: Least Squares
Date: 04/01/09 Time: 10:32
Sample: 1 225
Included observations: 225

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.485774	0.648155	3.835155	0.0002
X2	0.003751	0.004468	0.839525	0.4021
D1	0.402730	0.294885	1.365721	0.1734
D2	0.480869	0.241677	1.989719	0.0479
D3	-1.285296	0.563680	-2.280187	0.0236
D4	1.122022	0.532487	2.107134	0.0362
D5	2.560528	0.366850	6.979768	0.0000
R-squared	0.669341	Mean dependent var	3.515511	
Adjusted R-squared	0.660240	S.D. dependent var	2.844026	
S.E. of regression	1.657752	Akaike info criterion	3.879419	
Sum squared resid	599.0949	Schwarz criterion	3.985698	
Log likelihood	-429.4346	Hannan-Quinn criter.	3.922313	
F-statistic	73.54820	Durbin-Watson stat	1.764107	
Prob(F-statistic)	0.000000			

X2 : diện tích
D1 : môi trường
D2 : khu vực
kinh doanh
D3 : nhu cầu bán
D4 : an ninh khu
vực
D5 : vị trí nhà
D6 : thị trường
đồng bằng

Theo kết quả bài tập của nhóm 4 lớp KK2_05 trường Đại học Hồng Bàng

Chương 4

HỘI QUY VỚI BIẾN GIÁ

By Tuan Anh(UEH)

I. BẢN CHẤT CỦA BIẾN GIÁ

Biến định lượng : giá trị thể hiện bằng những con số

Ví dụ : Thu nhập, chi tiêu, chi phí, doanh thu, v.v...

Biến định tính: giá trị không thể hiện bằng những con số

Ví dụ : Giới tính, màu sắc, tôn giáo, chất liệu, v.v...

By Tuan Anh(UEH)

I. BẢN CHẤT CỦA BIẾN GIẢ



Biến định tính thường biểu thị có hay không có một tính chất hoặc là các mức độ khác nhau của một tiêu thức thuộc tính nào đó

Để lượng hoá các biến định tính, trong phân tích hồi quy người ta dùng biến giả (***dummy variables***)

By Tuan Anh(UEH)

II. Hồi qui với biến độc lập đều là biến định tính.

1. Trường hợp các biến định tính chỉ có hai lựa chọn



Ví dụ : giới tính : - *Nam*
- *Nữ*

Ngôi nhà : - *Mặt tiền*
- *Không phải mặt tiền*

Khu vực bán hàng : - *Thành thị*
- *Nông thôn*

By Tuan Anh(UEH)

II. Hồi qui với biến độc lập đều là biến định tính.

1. Trường hợp các biến định tính chỉ có hai lựa chọn

Giả sử : Chúng ta muốn nghiên cứu tiền lương tại một doanh nghiệp có bị ảnh hưởng bởi vấn đề giới tính hay không ? (Tức là có sự khác biệt tiền lương giữa nhân viên nam và nữ hay không ?)

Giới tính là biến định tính nên ta dùng biến giả D_i

Với $D_i = 1$: Nam

$D_i = 0$: Nữ

By Tuan Anh(UEH)

II. Hồi qui với biến độc lập đều là biến định tính.

1. Trường hợp các biến định tính chỉ có hai lựa chọn

Hàm hồi quy có dạng : $PRF : Y_i = \beta_1 + \beta_2 D_i + U_i$

Thu thập số liệu : $SRF : \hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 D_i$

Y_i (trđ/tháng)	D_i
5,0	1
4,0	0
3,8	0
3,5	1
...	...

By Tuan Anh(UEH)

II. Hồi qui với biến độc lập đều là biến định tính.

1. Trường hợp các biến định tính chỉ có hai lựa chọn

Tiến hành hồi quy *như hàm hai biến*, giả sử ta được ước lượng của hàm hồi quy sau :

$$Y_i = \beta_1 + \beta_2 D_i + U_i$$

β_1 là tiền lương trung bình của nhân viên nữ

Tạm thời bỏ qua sai số U_i

Đối với nữ: $D = 0 \Rightarrow Y = \beta_1$

Đối với nam: $D = 1 \Rightarrow Y = \beta_1 + \beta_2$

$(\beta_1 + \beta_2)$ là tiền lương trung bình của nhân viên nam

(β_2) là chênh lệch tiền lương trung bình giữa nhân viên nam và nữ

By Tuan Anh(UEH)

II. Hồi qui với biến độc lập đều là biến định tính.

1. Trường hợp các biến định tính chỉ có hai lựa chọn

Lưu ý: Lựa chọn được gán với giá trị $D_i = 0$ trở thành “lựa chọn cơ sở” hay còn gọi là “nhóm điều khiển”

Tóm lại :

β_1 là lương trung bình của nhóm điều khiển (nhân viên nữ)

β_2 là chênh lệch về lương trung bình của một nhân viên nam so với nhân viên nữ.

By Tuan Anh(UEH)

II. Hồi qui với biến độc lập đều là biến định tính.

1. Trường hợp các biến định tính chỉ có hai lựa chọn

- Vậy làm thế nào để xét xem tại doanh nghiệp này có sự khác biệt về tiền lương giữa nhân viên nam và nữ hay không ?

Ta kiểm định giả thiết

$$H_0: \beta_2 = 0 \quad (\text{độ tin cậy } 1-\alpha)$$

$$H_1: \beta_2 \neq 0.$$

- Kiểm định bằng cách nào?

- Nếu ta đặt $D_i = 1$ là nữ thì có được không? Mô hình thay đổi như thế nào ?

By Tuan Anh(UEH)

II. Hồi qui với biến độc lập đều là biến định tính.

1. TH biến định tính có nhiều hơn hai lựa chọn

Số các lựa chọn có thể có của một biến định tính có thể nhiều hơn hai. Có hai cách :

- Dùng biến giả có nhiều giá trị, số giá trị bằng với số lựa chọn
- Dùng nhiều biến giả, mỗi biến có giá trị 0 và 1.

Cách 2 được khuyến khích hơn

**Chú ý: Để không rơi vào bẫy biến giả thì
số các biến giả = số lựa chọn - 1**

By Tuan Anh(UEH)

Dependent Variable: Y
 Method: Least Squares
 Date: 11/04/09 Time: 12:40
 Sample: 1 15
 Included observations: 15

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	24.13750	1.129105	21.37755	0.0000
Z	2.648214	1.652840	1.602220	0.1331
R-squared	0.164906	Mean dependent var	25.37333	
Adjusted R-squared	0.100668	S.D. dependent var	3.367591	
S.E. of regression	3.193591	Akaike info criterion	5.283735	
Sum squared resid	132.5873	Schwarz criterion	5.378142	
Log likelihood	-37.62801	Hannan-Quinn criter.	5.282729	
F-statistic	2.567109	Durbin-Watson stat	0.158630	
Prob(F-statistic)	0.133116			

By Tuan Anh(UEH)

II. Hồi qui với biến độc lập đều là biến định tính.

1. TH biến định tính có nhiều hơn hai lựa chọn

Ví dụ : Nghiên cứu tiền lương khi ra trường của sinh viên có phụ thuộc vào kết quả tốt nghiệp hay không

Kết quả tốt nghiệp gồm :

- Xuất sắc
- Giỏi
- Khá
- Trung bình
- Yếu kém

Sẽ có bao nhiêu biến giả được đưa vào mô hình ?

By Tuan Anh(UEH)

II. Hồi qui với biến độc lập đều là biến định tính.

1. TH biến định tính có nhiều hơn hai lựa chọn

Ta đưa 4 biến giả như sau:

$$D_{2i} = \begin{cases} 1 & \text{SV xuất sắc} \\ 0 & \text{khác} \end{cases} \quad D_{4i} = \begin{cases} 1 & \text{SV khá} \\ 0 & \text{khác} \end{cases}$$

$$D_{3i} = \begin{cases} 1 & \text{SV giỏi} \\ 0 & \text{khác} \end{cases} \quad D_{5i} = \begin{cases} 1 & \text{SV yếu kém} \\ 0 & \text{khác} \end{cases}$$

Lưu ý: Nhóm ứng với giá trị $D_{2i}=D_{3i}=D_{4i}=D_{5i}= 0$ là nhóm điều khiển

By Tuan Anh(UEH)

II. Hồi qui với biến độc lập đều là biến định tính.

1. TH biến định tính có nhiều hơn hai lựa chọn

Thu thập số liệu, ví dụ :

Y_i (trđ/tháng)	D_{2i}	D_{3i}	D_{4i}	D_{5i}
5,0	1	0	0	0
4,0	0	1	0	0
3,8	0	0	1	0
3,5	0	0	0	0
...				...

By Tuan Anh(UEH)

II. Hồi qui với biến độc lập đều là biến định tính.

1. TH biến định tính có nhiều hơn hai lựa chọn

Một mô hình đơn giản mô tả quan hệ giữa tiền lương và loại tốt nghiệp như sau :

$$Y_i = \beta_1 + \beta_2 D_{2i} + \beta_3 D_{3i} + \beta_4 D_{4i} + \beta_5 D_{5i} + U_i$$

- Ý nghĩa của β_1 là gì?

- Ý nghĩa của $\beta_2, \beta_3, \beta_4, \beta_5$ là gì?

By Tuan Anh(UEH)

II. Hồi qui với biến độc lập định tính và định lượng

1. Một biến định tính và một biến định lượng

Quay lại ví dụ về tiền lương , ta muốn kiểm tra xem liệu doanh nghiệp có tăng lương cho nhân viên theo **thâm niên**, đồng thời cũng muốn kiểm tra xem có phân biệt tiền lương theo **giới tính** hay không ?

Ta lập mô hình hồi quy với các biến như sau

- Y : là tiền lương hàng tháng của nhân viên
- X : Số năm kinh nghiệm
- Biến giả D với $D_i = 1$: nhân viên nam
 $D_i = 0$: nhân viên nữ

By Tuan Anh(UEH)

II. Hồi qui với biến độc lập định tính và định lượng

1. Một biến định tính và một biến định lượng

Hàm hồi quy: $Y_i = \beta_1 + \beta_2 X_i + \beta_3 D_i + U_i$

Y_i (trđ/tháng)	X_i	D_i
5,0	10	1
4,0	8	0
3,8	5	0
3,5	5	1
...		...

Tiến hành hồi quy *như hàm ba biến*

By Tuan Anh(UEH)

II. Hồi qui với biến độc lập định tính và định lượng

1. Một biến định tính và một biến định lượng

Hàm hồi quy: $Y_i = \beta_1 + \beta_2 X_i + \beta_3 D_i + U_i$

- Ý nghĩa của β_1 là gì?

- Ý nghĩa của β_2 là gì?

- Ý nghĩa của β_3 là gì?

By Tuan Anh(UEH)

II. Hồi qui với biến độc lập định tính và định lượng

1. Một biến định tính và một biến định lượng

Làm thế nào để kiểm tra tiền lương có bị ảnh hưởng bởi số năm kinh nghiệm hay không?

chúng ta kiểm định giả thiết

$$H_0: \beta_2 = 0$$

$$H_1: \beta_2 \neq 0. \text{ (độ tin cậy } 1-\alpha)$$

Làm thế nào để kiểm tra tiền lương có bị ảnh hưởng bởi giới tính hay không?

chúng ta kiểm định giả thiết

$$H_0: \beta_3 = 0$$

$$H_1: \beta_3 \neq 0. \text{ (độ tin cậy } 1-\alpha)$$

By Tuan Anh(UEH)

II. Hồi qui với biến độc lập định tính và định lượng

1. Một biến định tính và một biến định lượng

Hàm hồi quy: $Y_i = \beta_1 + \beta_2 X_i + \beta_3 D_i + U_i$

$D_i = 0 \Rightarrow$ Hàm hồi quy của nhân viên nữ

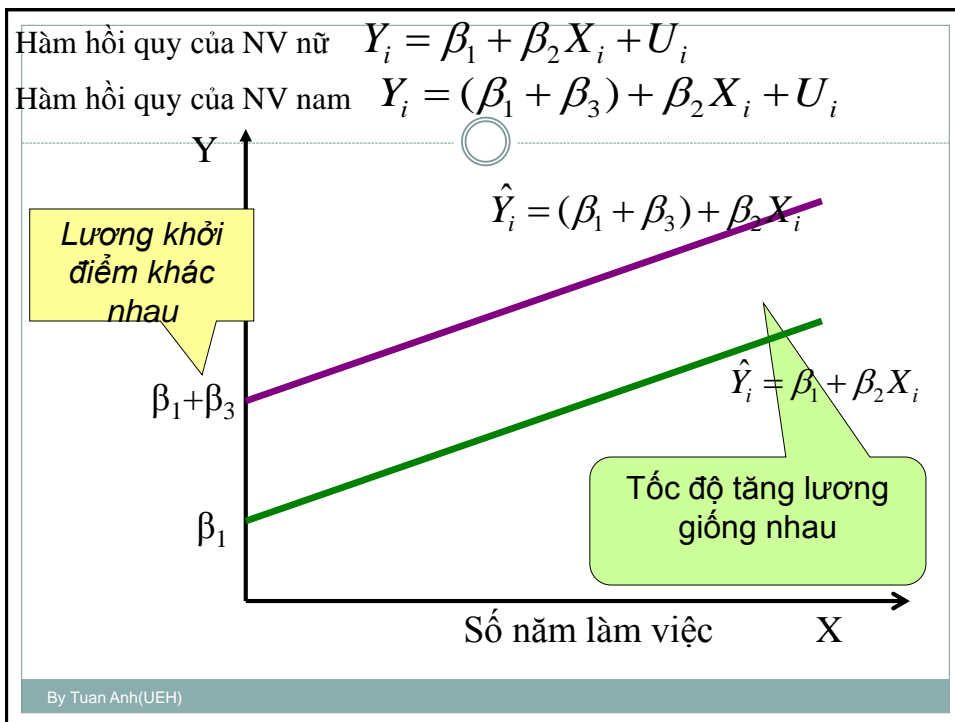
$$Y_i = \beta_1 + \beta_2 X_i + U_i$$

$D_i = 1 \Rightarrow$ Hàm hồi quy của nhân viên nam

$$Y_i = (\beta_1 + \beta_3) + \beta_2 X_i + U_i$$

Có thể nhận xét gì từ hai hàm hồi quy trên? (xem đồ thị)

By Tuan Anh(UEH)



II. Hồi qui với biến độc lập định tính và định lượng

1. Một biến định tính và một biến định lượng

Làm sao để biết tốc độ tăng lương có khác nhau giữa nam và nữ hay không?

Ta sử dụng dạng hàm hồi quy:

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 D_i + \beta_4 X_i D_i + U_i$$

Khi đó biến $X_i \cdot D_i$ được gọi là **biến tương tác** giữa X và D

By Tuan Anh(UEH)

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 D_i + \beta_4 X_i D_i + U_i$$

$D_i = 0 \Rightarrow$ Hàm hồi quy của nhân viên nữ

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

$D_i = 1 \Rightarrow$ Hàm hồi quy của nhân viên nam

$$Y_i = (\beta_1 + \beta_3) + (\beta_2 + \beta_4) X_i + U_i$$

- Ý nghĩa của β_1 là gì?

- Ý nghĩa của β_3 là gì?

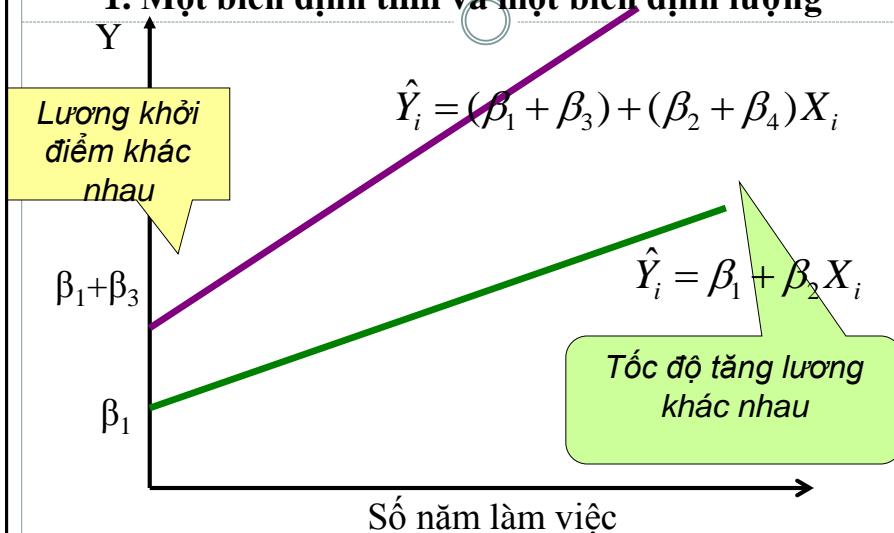
- Ý nghĩa của β_2 là gì?

- Ý nghĩa của β_4 là gì?

By Tuan Anh(UEH)

II. Hồi qui với biến độc lập định tính và định lượng

1. Một biến định tính và một biến định lượng



By Tuan Anh(UEH)

II. Hồi qui với biến độc lập định tính và định lượng

1. Một biến định tính và một biến định lượng

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 D_i + \beta_4 X_i D_i + U_i$$

Từ hàm hồi qui này làm sao để biết tốc độ tăng lương có khác nhau giữa nam và nữ hay không?

chúng ta kiểm định giả thiết

$$H_0: \beta_4 = 0$$

$$H_1: \beta_4 \neq 0. \text{ (độ tin cậy } 1-\alpha\text{)}$$

By Tuan Anh(UEH)

II. Hồi qui với biến độc lập định tính và định lượng

2. Nhiều biến định tính và nhiều biến định lượng

Nếu mô hình có nhiều biến định tính, chúng ta có thể xác định số biến giả được đưa vào mô hình như sau:

Trong đó: n - là số biến giả cần thiết đưa vào mô hình

k - là số biến định tính

n_i - là số lựa chọn của biến định tính thứ i

By Tuan Anh(UEH)

II. Hồi qui với biến độc lập định tính và định lượng

2. Nhiều biến định tính và nhiều biến định lượng

Ví dụ : Khảo sát các yếu tố ảnh hưởng đến kết quả học tập của sinh viên.

By Tuan Anh(UEH)

Ví dụ minh họa
Cho số liệu giả thiết về mức lương của nhân viên (Y-trở/năm), số năm kinh nghiệm (X) và giới tính ($D_i=1$:nam; $D_i=0$:nữ)

Y_i	X_i	D_i
115	11	1
95	9	0
120	10	1
105	12	0
125	13	1
110	12	0
132	14	1
116	14	0

Y_i	X_i	D_i
125	15	0
140	15	1
147	16	1
130	16	0
128	17	0
158	18	1
145	18	0

By Tuan Anh(UEH)

Dependent Variable: Y

Method: Least Squares

Date: 08/13/09 Time: 00:12

Sample: 1 15

Included observations: 15

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	45.28511	5.538866	8.175881	0.0000
X	5.236452	0.379310	13.80520	0.0000
Z	16.00976	2.058934	7.775752	0.0000
R-squared	0.952576	Mean dependent var	126.0667	
Adjusted R-squared	0.944672	S.D. dependent var	16.89238	
S.E. of regression	3.973393	Akaike info criterion	5.773974	
Sum squared resid	189.4542	Schwarz criterion	5.915584	
Log likelihood	-40.30481	Hannan-Quinn criter.	5.772466	
F-statistic	120.5192	Durbin-Watson stat	1.919503	
Prob(F-statistic)	0.000000			

Và kết quả hồi quy này giúp rút ra kết luận gì ?

Dependent Variable: Y

Method: Least Squares

Date: 08/13/09 Time: 00:18

Sample: 1 15

Included observations: 15

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	46.29423	7.521975	6.154532	0.0001
X	5.165010	0.522340	9.888206	0.0000
Z	13.67223	11.37238	1.202231	0.2545
X*Z	0.167307	0.799342	0.209306	0.8380
R-squared	0.952765	Mean dependent var	126.0667	
Adjusted R-squared	0.939882	S.D. dependent var	16.89238	
S.E. of regression	4.141834	Akaike info criterion	5.903333	
Sum squared resid	188.7027	Schwarz criterion	6.092146	
Log likelihood	-40.27500	Hannan-Quinn criter.	5.901321	
F-statistic	73.95857	Durbin-Watson stat	1.911440	
Prob(F-statistic)	0.000000			

