Chương 2:
HỒI QUI HAI BIẾN

HỒI QUI HAI BIẾN

1. Ước lượng các hệ số hồi qui bằng phương pháp bình phương bé nhất thông thường (OLS)
2. Phương sai và sai số chuẩn của các ước lượng
3. Hệ số xác định và hệ số tương quan
4. Khoảng tin cậy của các hệ số hồi qui & phương sai
5. Kiểm định giả thiết
6. Ứng dụng

Lưu ý về ký hiệu

- Y_i : giá trị quan sát thực tế mẫu,
- $\hat{Y}_i$: giá trị tính toán (lý thuyết) – Mẫu
- $X_{k,\,i}$: k- thứ tự biến trong mô hình; i – thứ tự quan sát

$$x_i = X_i - \bar{X} \;; \quad y_i = Y_i - \bar{Y} \;; \; \hat{y}_i = \hat{Y}_i - \bar{Y}$$

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 . X_i \;; \quad Y_i = \hat{\beta}_1 + \hat{\beta}_2 . X_i + \hat{U}_i$$

$$\hat{u}_i = y_i - \hat{y}_i = Y_i - \hat{Y}_i = \hat{U}_i = e_i$$

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 . X_i + \hat{u}_i$$

Dự đoán dấu β_2 → dựa vào bản chất kinh tế của quan hệ hồi qui :

VD: Quan hệ chi tiêu với thu nhập → **$\beta_2 > 0$.**

Lãi suất cho vay với mức cầu vay vốn → **$\beta_2 < 0$**

Cách giải thích ý nghĩa β_1

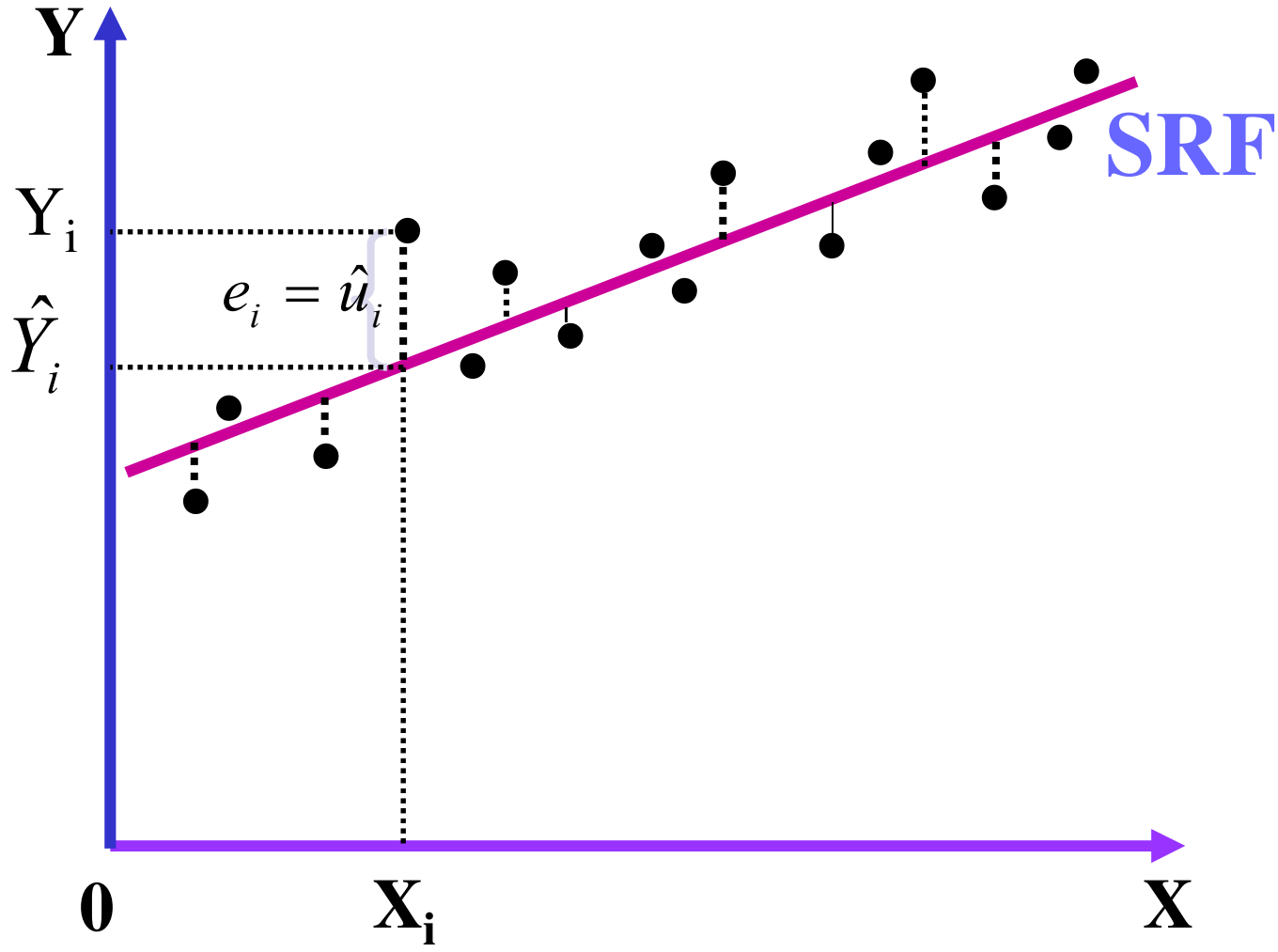
❖ Nếu **$\beta_2 < 0$** → biến X và biến Y nghịch biến

* Khi X = 0 → Y = **$\beta_1 = Y_{max}$**

❖ Nếu **$\beta_2 > 0$** → biến X và biến Y đồng biến: X tăng thì Y tăng ; X giảm thì Y giảm

* Khi X = 0 → Y = **$\beta_1 = Y_{min}$**

Ước lượng tham số hồi qui bằng phương pháp bình phương bé nhất



Ước lượng tham số hồi qui bằng phương pháp bình phương bé nhất

a/ Nguyên tắc - Ước lượng hệ số của hàm $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i$

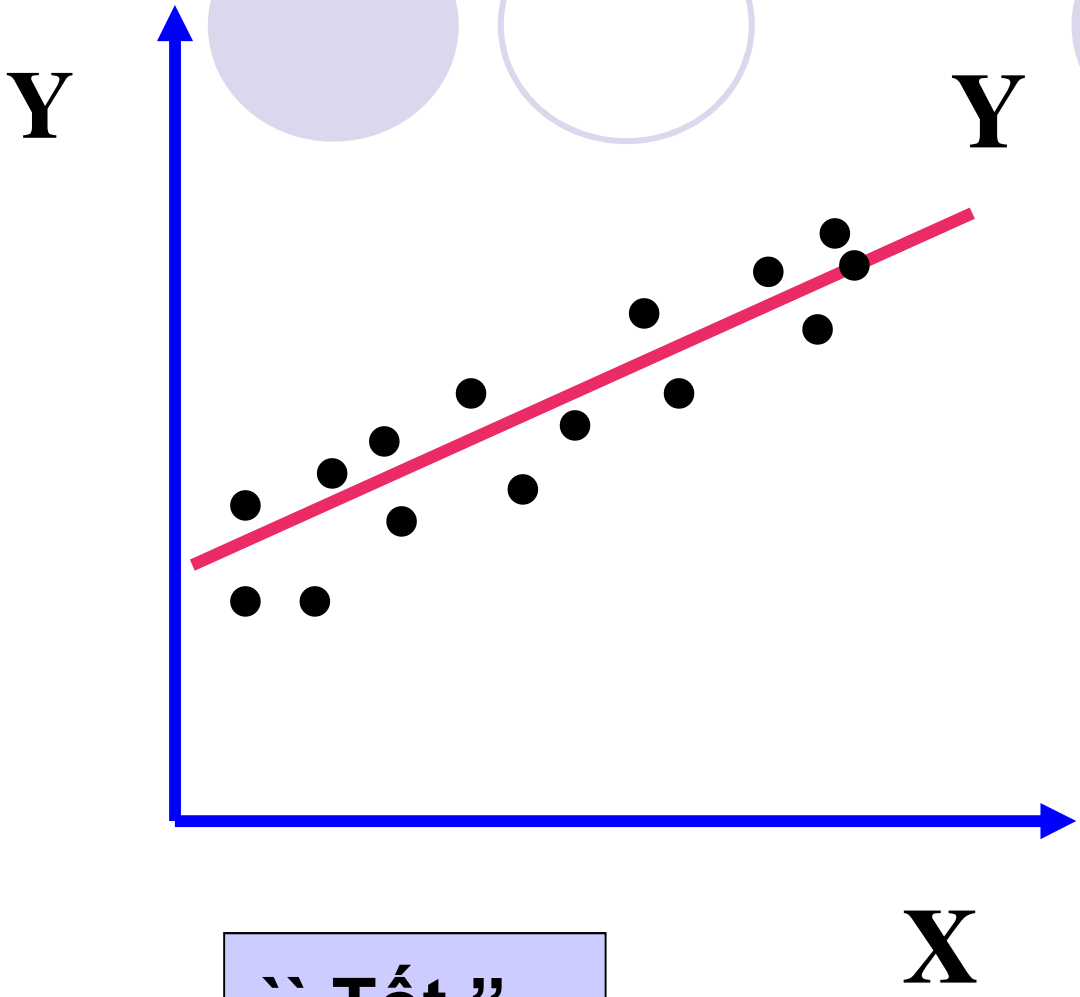
$$sao\ cho\ \sum \hat{u}_i^2 = \sum (Y_i - \hat{Y}_i)^2 \to \min$$

b. Công thức

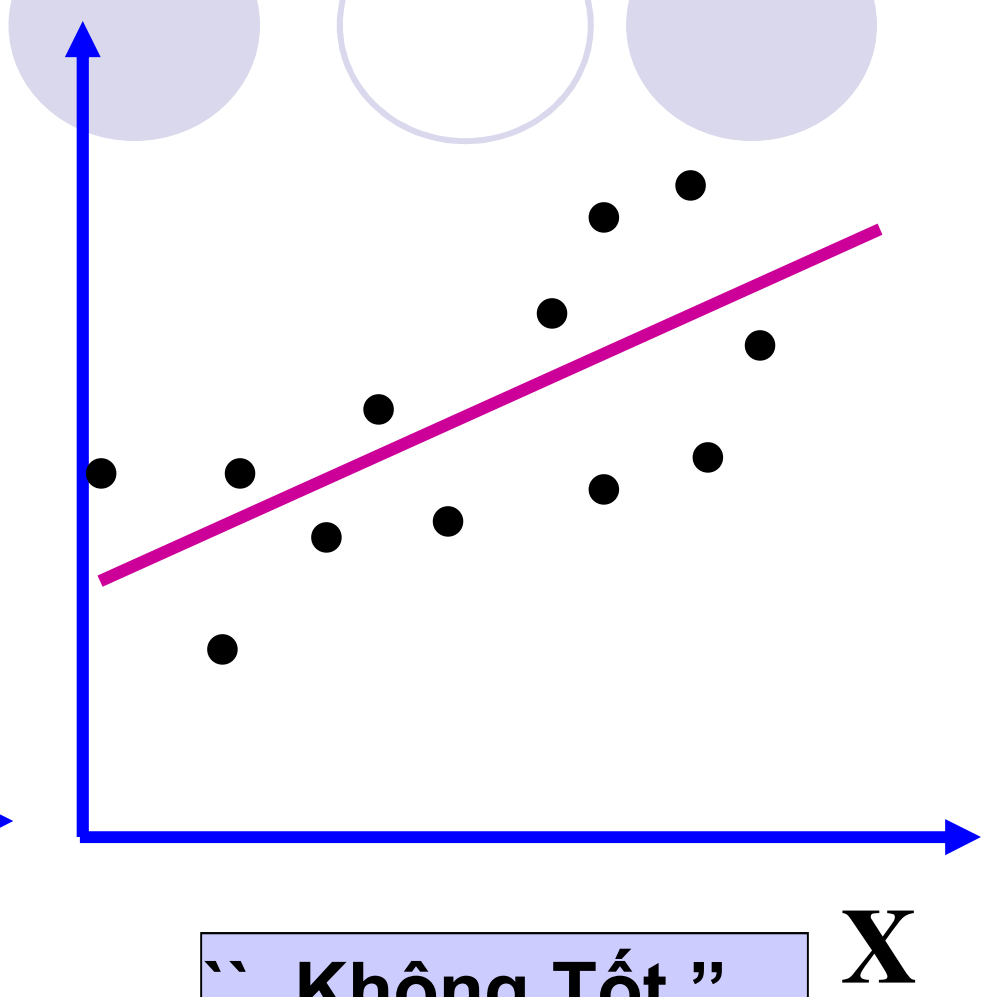
$$\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n} (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2 = f(\hat{\beta}_1, \hat{\beta}_2) \to \min$$

$$\Leftrightarrow \begin{cases} f'(\hat{\beta}_1) = 0 \\ f'(\hat{\beta}_2) = 0 \end{cases}$$

$$\Leftrightarrow \hat{\beta}_1 = \overline{Y} - \hat{\beta}_2 \overline{X} \qquad \hat{\beta}_2 = \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{X}\overline{Y}}{\sum_{i=1}^{n} X_i^2 - n(\overline{X})} = \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{X}\overline{Y}}{\sum_{i=1}^{n} x_i^2}$$



`` Tốt ”



`` Không Tốt ”

Hệ số xác định R^2

a. TSS (Total Sum of Squares) = Tổng bình phương độ lệch của Y

$$TSS = \sum_{i=1}^{n} y_i^2 = \sum_{i=1}^{n} (Y_i - \bar{Y})^2 = \sum_{i=1}^{n} Y_i^2 - n(\bar{Y})^2$$

b. ESS (Explained Sum of Squares) = Tổng bình phương độ lệch của Y được giải thích bởi SRF

$$ESS = \sum_{i=1}^{n} \hat{y}_i^2 = \sum_{i1=}^{n} (\hat{Y}_i - \bar{Y})^2 = \hat{\beta}_2^2 \sum_{i1=}^{n} x_i^2 = \hat{\beta}_2^2 (\sum_{i1=}^{n} X_i^2 - n\bar{X}^2)$$

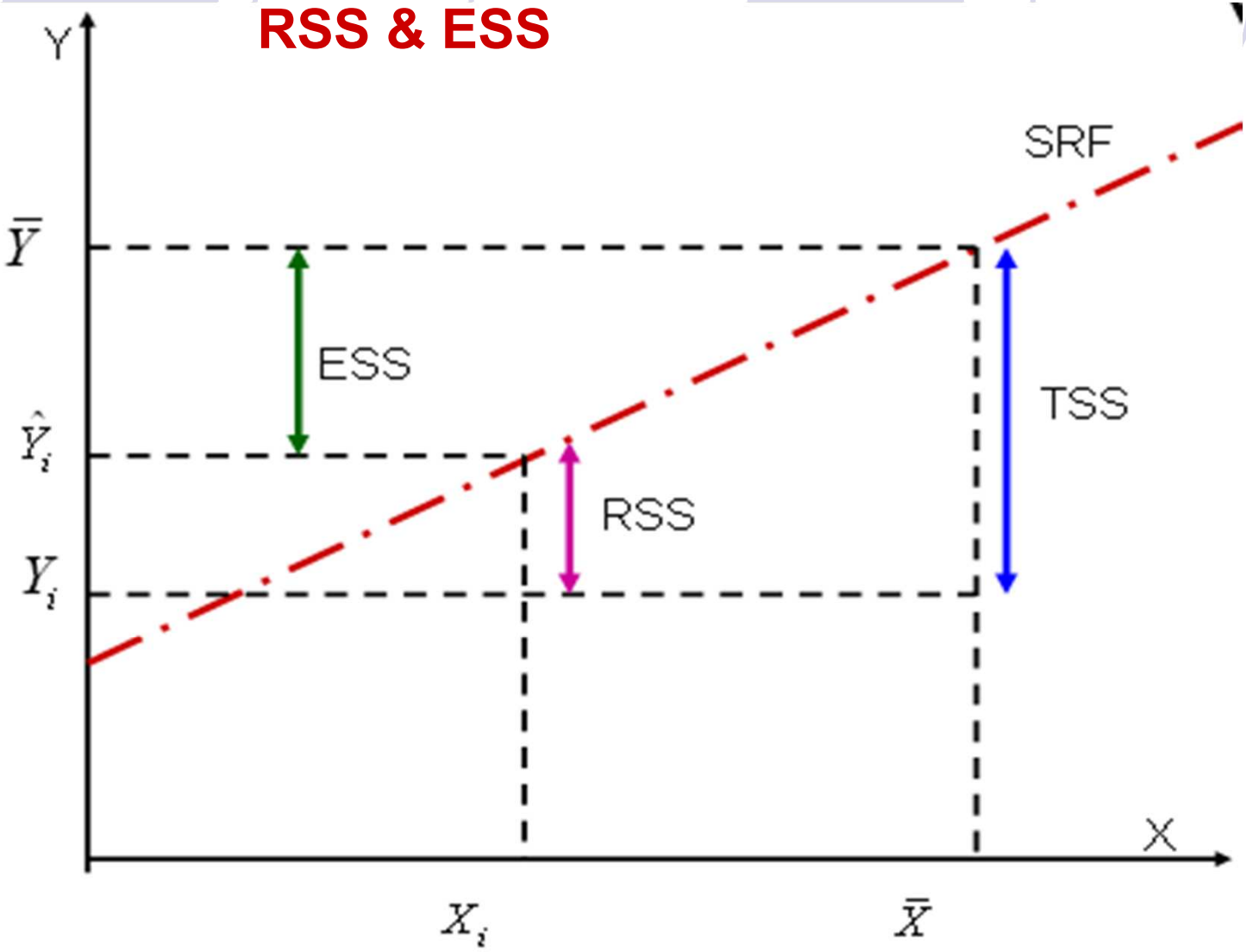
c. RSS (Residual Sum of Squares) = Tổng bình phương độ lệch giữa giá trị quan sát thực tế và giá trị tính toán = tổng bình phương độ lệch Y không được giải thích bởi SRF ➔ RSS do yếu tố ngẫu nhiên gây ra

$$RSS = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \Rightarrow TSS = ESS + RSS \rightarrow 1 = \frac{ESS}{TSS} + \frac{RSS}{TSS}$$

d. R^2 : Hệ số xác định (Coefficient of Determination) – Đo mức độ phù hợp của hàm HQ

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

Ý nghĩa hình học của TSS, RSS & ESS



Tính chất TSS & R^2

a. Tính chất của TSS

* Với 01 mẫu cho trước thì TSS cố định, nhưng ESS, RSS thay đổi tùy theo dạng hàm ta chọn

* Khi ESS càng lớn so với RSS ➔ Hàm SRF phù hợp tốt với số liệu quan sát

  Đặc biệt nếu tất cả Y_i nằm trên SRF ➔ ESS = TSS (RSS = 0) và ngược lại $\hat{Y}_i \equiv Y_i, \forall i$

* ESS càng nhỏ hơn RSS ➔ Hàm SRF kém phù hợp số liệu quan sát

Tính chất TSS & R^2

b. R^2

- $0 \le R^2 \le 1$

* Nếu $R^2 = 1$ → RSS = 0 ; TSS=ESS

→ đường hồi qui phù hợp hoàn hảo, tất cả sai lệch của Yi đều được giải thích bởi SRF

→ $\hat{Y}_i \equiv \overline{Y}, \forall i$

* Nếu $R^2 = 0$ → ESS=0; TSS = RSS

→ SRF không thích hợp, tất cả sai lệch của Y_i không được giải thích bởi hàm SRF

Phương sai & sai số chuẩn của các ước lượng → đánh giá biến động các HSHQ

1. $$\mathrm{var}(\hat{\beta}_2) = \frac{\hat{\sigma}^2}{\sum_{i=1}^{n} x_i^2} \qquad se(\hat{\beta}_2) = \sqrt{\mathrm{var}(\hat{\beta}_2)}$$

2. $$\mathrm{var}(\hat{\beta}_1) = \frac{\sum_{i=1}^{n} X_i^2}{n\sum_{i=1}^{n} x_i^2}\hat{\sigma}^2 \qquad se(\hat{\beta}_1) = \sqrt{\mathrm{var}(\hat{\beta}_1)}$$

Với:

$$\sigma^2 = \mathrm{var}(Ui) \qquad se: \text{sai số chuẩn} (standard\ error)$$

Nếu phương sai nhiễu tổng thể σ^2 chưa biết, thay nó bằng ước lượng không chệch của nó:

$$s^2 = \hat{\sigma}^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-2} = \frac{RSS}{n-2} \quad ; \quad \hat{\sigma} = \sqrt{\hat{\sigma}^2} \quad ; \quad e = Y_i - \hat{Y}_i \quad ; \quad e_i = \hat{U}_i$$

Khoảng tin cậy của β_1, β_2

→Chọn mẫu khác nhau, β_1, β_2 sẽ như thế nào?

→ Giới hạn biến động của Y khi biến X thay đổi 1 đv?

$$K.T.C.\beta = (\hat{\beta} - t_{\alpha/2;(n-2)} . se(\hat{\beta});$$

$$\hat{\beta} + t_{\alpha/2;(n-2)} . se(\hat{\beta}))$$

$t_{\alpha/2;(n-2)}$ tra bảng phân phối t hoặc dùng hàm TINV trong Excel

Khoảng tin cậy của phương sai của nhiễu

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

$$Hay: \frac{RSS}{\chi^2_{\alpha/2;(n-2)}} < \sigma^2 < \frac{RSS}{\chi^2_{1-\alpha/2;(n-2)}}$$

tra bảng phân phối χ^2 bậc tự do $(n-2)$

Ví dụ 1 : từ một mẫu 8 quan sát sau đây, hãy thiết lập hàm hồi qui

X: giá vé xe bus (ngàn đồng/vé)

Y: Tổng lượng người thường xuyên đi xe bus (triệu người)

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

$$\bar{X}=\frac{32}{8}=4\ ;\ \bar{Y}=\frac{40}{8}=5$$

$$\hat{\beta}_2=\frac{139-8*4*5}{156-8*4^2}=-0,75$$

$$\hat{\beta}_1=5-(-0,75)*4=8$$

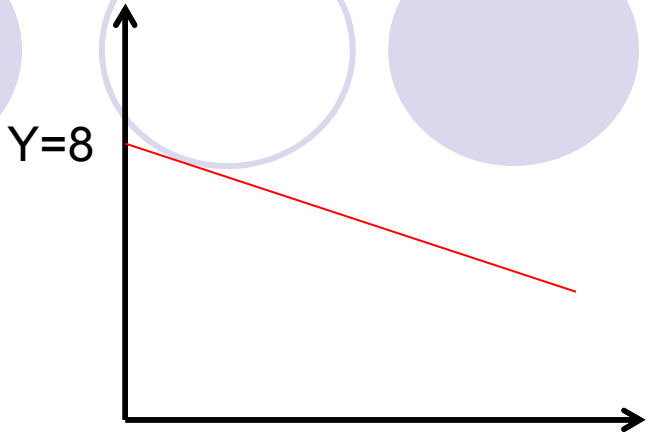
$$\hat{Y}_i=8-0,75X_i \rightarrow SRF$$

$$Y_i=8-0,75X_i+u_i \rightarrow PRF$$

Nhận xét $Y = 8 - 0,75X$

mà X: giá vé xe bus → $X \geq 0$

➔ $Y \leq 8$ hay $Y_{max} = 8$



X tăng 1 đơn vị → Y giảm 0,75 đơn vị

→ X & Y nghịch biến

→ $Y_{max} = 8$ (Khi X = 0)

Ý nghĩa kinh tế của các hệ số hồi quy

(1). $\hat{\beta}_1 = 8 = Y_{max}$

Khi giá vé xe bus giảm đến tối đa, tổng lượng người thường xuyên đi xe bus bình quân đạt đến mức tối đa là triệu người

(2). $\hat{\beta}_2 = -0,75 <$ ⇔ X và Y nghịch biến

→ Giá vé tăng (giảm) 1 ngàn đồng/vé, tổng lượng người thường xuyên đi xe bus bình quân giảm (tăng) 0,75 triệu người

Phương sai của nhiễu & sai số chuẩn các hệ số hồi quy

$$\hat{\sigma}^2 = \frac{RSS}{n-2} = \frac{2,25}{6} = 0,375$$

$$\rightarrow \text{var}(\hat{\beta}_2) = \frac{\hat{\sigma}^2}{\sum x_i^2} = 0,375 / 28 = 0,0134$$

$$\rightarrow se(\hat{\beta}_2) = \sqrt{\text{var}(\hat{\beta}_2)} = 0,1158$$

$$\rightarrow \text{var}(\hat{\beta}_1) = \frac{\sum X_i^2}{n.\sum x_i^2}.\hat{\sigma}^2 = \text{var}(\hat{\beta}_2).\frac{\sum X_i^2}{n}$$

$$= 0,0134 * 156 / 8 = 0,2613$$

$$\rightarrow se(\hat{\beta}_1) = \sqrt{\text{var}(\hat{\beta}_1)} = 0,5111$$

Khoảng tin cậy các HSHQ

* Kết quả hồi qui mẫu cho:

$$\hat{\beta}_1 = 8 \; ; \; \hat{\beta}_2 = -0,75$$

$$se(\hat{\beta}_1) = 0,5111 \; ; \; se(\hat{\beta}_2) = 0,1158$$

*

$$1-\alpha = 95\% \rightarrow \alpha = 5\% \quad ; \alpha / 2 = 0,025$$

$$\text{tra bảng } t-student \rightarrow \; t_{\alpha/2}^{n-2} = t_{0,025}^{6} = 2,447$$

* Suy ra:

$$*K.T.C\,\beta_1 = 8 \pm 2,447 * 0,511039 = (6,749488 \; ; 9,250512)$$

$$*K.T.C\,\beta_2 = -0,75 \pm 2,447 * 0,115728 = (-1,03319 \; ; -0,46681)$$

Ý nghĩa khoảng tin cậy các hệ số hồi qui:

(1). KTC β_1= (6,75 ; 9,25) → Tổng lượng người thường xuyên đi xe bus bình quân tối đa từ 6,75 đến 9,25 triệu người

(2). KTC β_2 = (-1,03 ; -0,4668). → Khi giá vé tăng 1 ngàn đồng/vé, tổng lượng người thường xuyên đi xe bus bình quân sẽ giảm ít nhất là 0,4668 triệu người đến cao nhất là 1,03 triệu người.

Khoảng tin cậy của phương sai

- Biết: $1-\alpha = 0,95\ ;\ \alpha = 0,05\ ;\ \alpha/2 = 0,025$

$$1-\alpha/2 = 0,975 \rightarrow tra\ bang\ \chi^2$$

$$\chi^2_{0,025}(6) = 14,4497\ ;\ \chi^2_{0,975}(6) = 1,2373$$

- Suy ra: $$K.T.C\ 95\%(\sigma^2) = \left[\frac{(n-2)\hat{\sigma}^2}{\chi^2_{\alpha/2;(n-2)}}; \frac{(n-2)\hat{\sigma}^2}{\chi^2_{1-\alpha/2;(n-2)}}\right]$$

$$Lower \qquad Upper$$

$$Hay: K.T.C\ 95\%(\sigma^2) = \left[\frac{RSS}{\chi^2_{\alpha/2;(n-2)}}; \frac{RSS}{\chi^2_{1-\alpha/2;(n-2)}}\right]$$

$$Do:\ \hat{\sigma}^2 = \frac{RSS}{(n-2)} \Rightarrow \hat{\sigma}^2(n-2) = RSS$$

$$= \left[\frac{6.(0,375)}{14,4494}; \frac{6.(0,375)}{1,2373}\right] = (0,155716\ ;1,818414)$$

Thiết lập bảng tính trung gian:

X_i	Y_i	X_i^2	Y_i^2	X_iY_i
1	8	1	64	8
2	6	4	36	12
3	6	9	36	18
4	5	16	25	20
4	4	16	16	16
5	4	25	16	20
6	4	36	16	24
7	3	49	9	21
$\Sigma X = 32$	**$\Sigma Y = 40$**	**$\Sigma X^2 = 156$**	**$\Sigma Y^2 = 218$**	**$\Sigma XY = 139$**

$$\bar{X} = \sum_{i=1}^{n} X_i / n = 32 / 8 = 4$$

$$\bar{Y} = \sum_{i=1}^{n} Y_i / n = 40 / 8 = 5$$

$$\sum x_i^2 = (\sum X^2) - n.\bar{X}^2 = 28$$

$$TSS = \sum y_i^2 = (\sum Y^2) - n.\bar{Y}^2 = 18$$

$$ESS = \hat{\beta}_2^2 \sum x_i^2 = (-0,75)^2.28 = 15,75$$

$$RSS = TSS - ESS = 2,25$$

$$R^2 = \frac{ESS}{TSS} = 0,875$$

Hệ số tương quan (r – coefficient of correlation)

a. r đo mức độ chặt chẽ trong quan hệ tuyến tính giữa X &Y

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 . \sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

$$Hay \quad r = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2 . \sum_{i=1}^{n} y_i^2}} \quad \Rightarrow r = \pm\sqrt{R^2}$$

b. Tính chất

- Dấu của r phụ thuộc dấu cov(X/Y)

$$-1 \le r \le 1$$

- r có tính đối xứng:

  $r_{xy} = r_{yx}$

- X & Y độc lập → r = 0; nhưng r = 0 → không có nghĩa là 2 biến này độc lập

Kiểm định hệ số hồi qui

a/ Khái niệm

- ***Kiểm định (test) giả thiết*** → xem xét **kết quả** từ số liệu khảo sát thực tế có **phù hợp** với giả thiết nêu ra?

("phù hợp" nghĩa là kết quả đó "đủ" sát với giá trị được giả thiết để không bác bỏ *giả thiết phát biểu*)

- Giả thuyết không $\mathbf{H_0}$ (Null hypothesis): Giả thiết phát biểu (không chứng minh)= giả thiết cần kiểm định;
- Giả thiết đối $\mathbf{H_1}$ (Alternative hypothesis) là giả thiết đối lập với $\mathbf{H_0}$

* Cơ sở: dựa trên luật phân phối xác suất của đại lượng ngẫu nhiên, xây dựng quy tắc để **bác bỏ** (khi chứng minh được $\mathbf{H_1}$) hay **không bác bỏ** $\mathbf{H_0}$.

Với β_0 là giá trị cho trước (giá trị kiểm định)

β_1 , β_2 : các hệ số hồi quy của hàm hồi quy tổng thể

Kiểm định 2 phía (đuôi)

1. Kiểm định tung độ gốc

$$\begin{cases} H_0 : \beta_1 = \beta_0 \\ H_1 : \beta_1 \neq \beta_0 \end{cases}$$

Trường hợp $\beta_0 = 0$: Kiểm định giả thiết H_0 là hàm có đường biểu diễn qua gốc tọa độ

2. Kiểm định hệ số góc

$$\begin{cases} H_0 : \beta_2 = \beta_0 \\ H_1 : \beta_2 \neq \beta_0 \end{cases}$$

Trường hợp $\beta_0 = 0$: Kiểm định giả thiết H_0 là biến giải thich X không có tác động lên biến phụ thuộc Y

- **Giả thiết :**

$$\begin{cases} H_0 : \beta_2 = \beta_0 \\ H_1 : \beta_2 \neq \beta_0 \end{cases}$$

Kết quả Kiểm định :

Bác bỏ H_0 : β_2 tồn tại → $\beta_2 \neq 0$ có ý nghĩa thống kê → biến X thực sự có ảnh hưởng lên biến Y

Không bác bỏ H_0 : β_2 không tồn tại → biến X không ảnh hưởng lên biến Y

Với β_0 là giá trị cho trước (giá trị giả thiết)

β : hệ số hồi quy β_1 hoặc β_2 của hàm HQ tổng thể

Kiểm định một phía (đuôi):

*Kiểm định bên phải

$$\begin{cases} H_0 : \beta = \beta_0 \\ H_1 : \beta > \beta_0 \end{cases}$$

* Kiểm định bên trái

$$\begin{cases} H_0 : \beta = \beta_0 \\ H_1 : \beta < \beta_0 \end{cases}$$

Phương pháp kiểm định dựa vào Khoảng tin cậy

Bước 1: Tính khoảng tin cậy của β:

* KĐ 2 phía : khoảng xem xét là

$$KTC(\beta):\ \ (\hat{\beta} - t_{\alpha/2;(n-2)}.se(\hat{\beta})\ ;\ \ \hat{\beta} + t_{\alpha/2;(n-2)}.se(\hat{\beta}))$$

* KĐ phải : khoảng xem xét là

$$\left(\hat{\beta} - t_{\alpha/2;(n-2)}.se(\hat{\beta}), +\infty\right)$$

* KĐ trái : khoảng xem xét là

$$\left(-\infty, \hat{\beta} + t_{\alpha/2;(n-2)}.se(\hat{\beta})\right)$$

Bước 2: Kết luận

Bác bỏ H_0 nếu giá trị kiểm định β_0 **không** thuộc khoảng xem xét.

Phương pháp dựa vào giá trị tới hạn
(Kđ mức ý nghĩa; kđịnh t, do dựa vào phân phối t)

- Bước 1: tính trị thống kê t (t-statistic)

$$t_0 = \frac{\hat{\beta} - \beta_0}{se(\hat{\beta})}$$

- Bước 2: tra bảng t – student tìm giá trị tới hạn t* =:

$$t_{\alpha;(n-2)}(1\ phía) \quad hoac \quad t_{\alpha/2;(n-2)}(2\ phía)$$

- Bước 3: so sánh t_0 với t*

- Kđ 2 phía: $|t_0| > t_{\alpha/2;(n-2)} \rightarrow bac\ bo\ H_0$

- Kđ phải: $t_0 > t_{\alpha;(n-2)} \rightarrow bac\ bo\ H_0$

- Kđ trái: $t_0 < -t_{\alpha;(n-2)} \rightarrow bac\ bo\ H_0$

Phương pháp kiểm định P - Value

- **Tính t_0:** $$t_0 = \frac{\hat{\beta}_{1,2} - \beta_0}{se(\hat{\beta}_{1,2})}$$
- **Tính P – value**: P – Value = $$P(|t| > |t_0|)$$

Với t – ĐLNN có phân phối t – Student, bậc tự do (n – 2)

- **Quy tắc:**

  * Kđ 2 phía: p – value < α → bác bỏ H_0

  * Kđịnh 1 phía: p – value < 2α → bác bỏ H_0

Ví dụ : Kiểm định β2 .
H0 : β2 = 0 ; H1 : β2 ≠ 0 . Độ tin cậy 95%

1. Kiểm định KTC: KTC β_2 = (- 1,03319; - 0,46681).

Vì 0 không thuộc KTC → bác bỏ H_0

2. Kiểm định t:

$$t_0 = \frac{\hat{\beta}_2 - 0}{se(\hat{\beta}_2)} = \frac{-0,75}{0,115728} = -6,48071 \; ; \; |t_0| > t_{0,025\,;\,6} = 2,447 \rightarrow bac\ bo\ H_0$$

3. Kiểm định p-value

$$t_0 = -6,48071 \; ; \; p - value = p(|t| > 6,48071) < 0,001 < \alpha = 0,05$$

$$\rightarrow bac\ bo\ H_0$$

- Ý nghĩa: Kiểm định giả thiết H_0 : β_2 = 0 → giả thiết cho rằng X không ảnh hưởng lên Y. Bác bỏ H_0 nghĩa là thừa nhận X thực sự có ảnh hưởng lên Y

Kiểm định phương sai của nhiễu

KĐ	H_0	H_1	P.P	Bác bỏ H_0
2 phía	$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	**K.T.C**	$\sigma^2 \notin KTC$
			G.T.T.H	$\chi_0^2 > \chi_{\alpha/2}^2(n-2)\ hoac\ \chi_0^2 < \chi_{1-\alpha/2}^2(n-2)$
			P-value	$p-value < \alpha/2\ \ hoac\ \ p-value > 1-\alpha/2$
Phải	$\sigma^2 = \sigma_0^2$	$\sigma^2 > \sigma_0^2$	**K.T.C**	σ_0^2 ≠ ½ KTC phải
			G.T.T.H	
			P-value	P value < α $\chi_0^2 > \chi_\alpha^2(n-2)$
Trái	$\sigma^2 = \sigma_0^2$	$\sigma^2 < \sigma_0^2$	**K.T.C**	σ_0^2 ≠ ½ KTC trái
			G.T.T.H	
			P-value	P value > 1 – α $\chi_0^2 < \chi_{1-\alpha}^2(n-2)$

Kiểm định sự phù hợp của mô hình

- **Mục đích**: đánh giá mức độ thích hợp của mô hình HQ

Giả thiết: $\begin{cases} H_0 : R^2 = 0 \\ H_1 : R^2 > 0 \end{cases}$

trong HQ 2 biến tương đương với $\begin{cases} H_0 : \beta_2 = 0 \\ H_1 : \beta_2 \neq 0 \end{cases}$

PP giá trị tới hạn:

B1: Tính $F_0 = \dfrac{R^2(n-2)}{1-R^2}$

B2: Tra bảng tìm GTTH $F_{\alpha\,(1,\,n-2)}$

B3: So sánh F_0 và $F_{\alpha\,(1,\,n-2)}$:

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

Lưu ý: R^2 trong công thức tính F_0 là của mẫu;
R^2 trong giả thiết là của tổng thể.

Kiểm định sự phù hợp của mô hình

Phương pháp p-value:

B1: Tính F_0

$$F_0 = \frac{R^2(n-2)}{1-R^2}$$

B2: Tính p-value= $P(F>F_0)$ với F: phân phối Fisher (1, n-2)

B3: So sánh p-value α

Nếu p-value < α: bác bỏ H_0

Nếu p value > α: không bác bỏ H_0

- $H_0 : R^2 = 0$

- Bác bỏ H_0 : Thừa nhận $R^2 > 0$. Mô hình phù hợp. Biến X giải thích được sự thay đổi của biến Y.

- Không bác bỏ H_0 : Mô hình không phù hợp. Biến X không giải thích cho biến Y.

Ví dụ: Kiểm định $H_0 : R^2 = 0$; $H_1: R^2 > 0$

- **Kiểm định theo PP giá trị tới hạn.**

$$F_0 = \frac{R^2(n-2)}{1-R^2} = \frac{(0,875).6}{1-0,875} = 42$$

$\alpha = 0,05$. Tra bảng F, ta có: $F_{\alpha\ ;\ (n-2)} = F_{0,05;(1,6)} = 5,987$

$F_0 > F_{0,05;(1,6)} \Rightarrow$ bác bỏ H_0, thừa nhận $R^2 > 0$ và có ý nghĩa thống kê': X giải thích được 87,5% sự thay đổi của Y, 12,5% thay đổi còn lại do các yếu tố ngẫu nhiên gây ra

Trình bày kết quả hồi qui

$\hat{Y}_i$ = 8 - 0,75 X_1 ; n = 8

se = (0,511) (0,115) ; R^2 = 0,875

t = (15,654) (-6,4807) ; F_0 = 42

p – value = (0,0000) (0,0006)

TSS = 18; ESS = 15,75; RSS = 2,25;

$$\hat{\sigma}^2 = 0,375$$

Một ứng dụng của phân tích hồi qui: Dự báo

1. Cơ sở

- Từ số liệu của mẫu → hàm hồi qui mẫu.
- Dùng hàm HQ mẫu để dự báo Y trong tương lai ứng với một giá trị của X cho trước

2. a. Dự báo giá trị trung bình

$$\left[\hat{Y}_0 - t_{\alpha/2\ ;\ (n-2)}.se(\hat{Y}_0) < E\left(Y / X_0\right) < \hat{Y}_0 + t_{\alpha/2\ ;\ (n-2)}.se(\hat{Y}_0)\right]$$

$$\text{var}(\hat{Y}_0) = \sigma^2 \left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum_{i=1}^{n} x_i^2}\right] \quad ; \quad se(\hat{Y}_0) = \sqrt{\text{var}(\hat{Y}_0)}$$

2.b. Dự báo giá trị riêng biệt (Y_0) khi $X = X_0$

$$Y_0 = \hat{Y}_0 \pm t_{\alpha/2\ ;\ (n-2)} . se(Y_0 - \hat{Y}_0) \quad Hay:$$

$$\hat{Y}_0 - t_{\alpha/2\ ;\ (n-2)} . se(Y_0 - \hat{Y}_0) \le Y_0 \le \hat{Y}_0 + t_{\alpha/2\ ;\ (n-2)} . se(Y_0 - \hat{Y}_0)$$

$$Var(Y_0 - \hat{Y}_0) = \sigma^2 \left[1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum_{i=1}^{n} x_i^2} \right] \quad ; \quad se(Y_0 - \hat{Y}_0) = \sqrt{\text{var}(Y_0 - \hat{Y}_0)}$$

Lưu ý: khoảng dự báo của giá trị cá biệt Y_0 rộng hơn khoảng dự báo của giá trị trung bình của Y

Các giả thiết của phương pháp OLS

GT1. Biến giải thích: đại lượng xác định trước (không ngẫu nhiên): thu nhập

GT2. $E(u_i/X_i) = 0$

Kỳ vọng yếu tố ngẫu nhiên $u_i = 0$. Các yếu tố không có trong mô hình không ảnh hưởng đến Y ($u_i > 0 = u_i < 0$).

VD: chênh lệch chi tiêu trung bình giữa các nhóm nghề khác nhau nhưng cùng thu nhập → bù trừ nhau

GT3. $Var(u_i/X_i) = \sigma^2$

Các u_i có phương sai bằng nhau (đều, thuần nhất)

VD: Chi tiêu Nhóm thu nhập thấp và cao có khác nhau

→ Gỉa thiết này có thể bị vi phạm

GT4. Cov(u_i,u_j) = 0 Với i≠ j

Không có tương quan giữa các u_i

VD: Chi tiêu các thành viên cùng gia đình, thu nhập khác nhau nhưng các yếu tố khác có thể cùng tác động → Gỉa thiết có thể bị vi phạm

GT5. Cov(ui,Xi) = 0

u_i và X_i không tương quan nhau . Nếu u và X tương quan, tức là đã không tách rời ảnh hưởng X & u lên Y.

VD: Nếu hoàn cảnh gia đình (trong u) có ảnh hưởng đến thu nhập cá nhân → giả thiết có thể bị vi phạm

GT6. U_i~ N(0,σ^2)

Bài tập

X – Thu nhập (tr $ tháng)

Y: chi tiêu (tr $ tháng)

Yêu cầu:

1/ Ước lượng hàm hồi quy $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$

2/ Nêu ý nghĩa kinh tế của các hệ số hồi quy

n	X_i	Y_i	X_i^2	X_iY_i
1	5	4,5		
2	4	3,7		
3	6	6,2		
4	8	7,3		
5	9	8,0		
6	3	3,3		
7	4	3,9		
8	7	6,3		
9	10	8,2		
10	12	10,1		
Tổng				

$$\hat{Y}_i = 0,936082 + 0,76675\ X_i$$

Bài tập 1

1/ Một khảo sát về lãi suất (X - %năm) và tổng vốn đầu tư (Y – tỷ đ) tại tỉnh KCT qua 10 năm như sau:

X	7,0	6,5	6,5	6,0	6,0	6,0	5,5	5,5	5,0	4,5
Y	28	32	30	34	32	35	40	42	48	50

Yêu cầu:

a/ Lập mô hình HQTT mẫu có dạng $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$

b/ Nêu ý nghĩa kinh tế các HSHQ

c/ Kiểm định giả thiết HSHQ trong hàm PRF bằng 0 ($\beta_2=0$, $\beta_1=0$) với mức ý nghĩa 2% và nêu ý nghĩa của kết quả.

d/ Đánh giá mức phù hợp của mô hình với độ tin cậy 99%

e/ Dự báo giá trị trung bình và giá trị cá biệt của tổng vốn đầu tư khi lãi suất 8% năm với độ tin cậy 95%.

f/ Cho rằng khi lãi suất tăng 1% năm thì tổng vốn đầu tư giảm 12 tỷ $, bạn nhận xét như thế nào về ý kiến này, độ tin cậy 95%.

(1). Ý nghĩa kinh tế của β_1

β_1= 0,93 → Khi X = 0 (không có thu nhập) thì **Y_{min} = 0,93 triệu**. Nghĩa là, khi không có thu nhập, chi tiêu bình quân sẽ ở mức tối thiểu là 930.000 đồng / tháng

(2). Ý nghĩa kinh tế của β_2

β_2 = 0,76 > 0 → X và Y đồng biến. Khi tăng(hay giảm) thu nhập 1 triệu đồng / tháng → Chi tiêu bình quân sẽ tăng (giảm) 0,76 triệu đồng /tháng.

Nói cách khác, khuynh hướng chi tiêu biên là 0,76.

Lưu ý: Nhận xét kết quả ước lượng phù hợp thực tế và lý thuyết kinh tế:

(1). Dù không thu nhập, vẫn phải chi tiêu.

(2). Khuynh hướng chi tiêu biên là MPC đạt: $0 < MPC = \hat{\beta}_2 < 1$

Khi thu nhập tăng 1 $, chi tiêu tăng nhưng ít hơn 1 $.

(Thiết lập bảng tính các đại lượng trung gian)

X_i	Y_i	$X_i Y_i$	X_i^2	Y_i^2
$\sum_{i=1}^{n} X_i$	$\sum_{i=1}^{n} Y_i$	$\sum_{i=1}^{n} X_i Y_i$	$\sum_{i=1}^{n} X_i^2$	$\sum_{i=1}^{n} Y_i^2$

(Thiết lập bảng tính các đại lượng trung gian)

X_i	Y_i	X_iY_i	X_i^2	Y_i^2
7	28	196	49	784
6,5	32	208	42,25	1024
6,5	30	195	42,25	900
6	34	204	36	1156
6	32	192	36	1024
6	35	210	36	1225
5,5	40	220	30,25	1600
5,5	42	231	30,25	1764
5	48	240	25	2304
4,5	50	225	20,25	2500
58,50	371	2.121	347,25	14.281

$$\overline{X} = (\sum_{i=1}^{n} X_i) / n = 58,5 / 10 = 5,85 \; ; \; \overline{Y} = (\sum_{i=1}^{n} Y_i) / n = 371 / 10 = 37,1$$

$$\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} X_i^2 - n(\overline{X})^2 = 347,25 - 10 * (5,85)^2 = 5,025$$

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{X}\overline{Y}}{\sum_{i=1}^{n} x_i^2} = \frac{2121 - 10 * 5,85 * 37,1}{5,025} = -9,8209$$

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

$$= 37,1 + 9,8209 * 5,85 = 94,552$$

1. $\hat{Y}_i = 94,5523 - 9,8209 X_i$

2. Ý nghĩa kinh tế của các hệ số hồi quy

$\beta_1 = 94,5523$; $\beta_2 = -9,8209 < 0$

Nhận xét:

1. X & Y nghịch biến → $Y_{max} = \beta_1 = 94,5523$ (khi X=0): **Mức vốn đầu tư bình quân tối đa khoảng 94,5523 tỷ đồng**

2. **Lãi suất tăng (giảm) 1%năm → Tổng vốn đầu tư bình quân giảm (tăng) 9,8209 tỷ đ/năm**

$$TSS = \sum_{i=1}^{n} y_i^2 = \sum_{i=1}^{n} Y_i^2 - n(\bar{Y})^2 = 14281 - 10 * (371)^2 = 516,9$$

$$ESS = \hat{\beta}^2 . \sum_{i=1}^{n} x_i^2 = (-9,8209)^2 * 5,025 = 484,6616$$

$$RSS = TSS - ESS = 516,9 - 484,6616 = 32,2384$$

3. Kiểm định các HSHQ

a/ Kiểm định β_2

$$H_0 : \beta_2 = 0$$

$$H_1 : \beta_2 \neq 0$$

$$\hat{\sigma}^2 = \frac{RSS}{(n-2)} = \frac{32,2384}{10-2} = 4,0298$$

$$\Rightarrow Var(\hat{\beta}_2) = \frac{\hat{\sigma}^2}{\sum x_i^2} = \frac{4,0298}{5,025} = 0,802$$

$$\Rightarrow se(\hat{\beta}_2) = \sqrt{Var(\hat{\beta}_2)} = \sqrt{0,802} = 0,8955$$

$$t_0 = \frac{\hat{\beta}_2}{se(\hat{\beta}_2)} = \frac{-9,8209}{0,8955} = -10,967$$

tra bang suy ra $t_{\alpha/2\,;\,(n-2)} = t_{0,01\,;\,8} = 2,896$

$$|t_0| = 10,967 > t_{0,01\,;\,8} = 2,896$$

➔Bác bỏ H_0 → $\beta_2 \neq 0$ có ý nghĩa thống kê

Nói cách khác, biến X thực sự có ảnh hưởng lên biến Y, nghĩa là lãi suất ngân hàng có ảnh hưởng lên tổng vốn đầu tư

b/ Kiểm định β_1

$$\begin{cases} H_0 : \beta_1 = 0 \\ H_1 : \beta_1 \end{cases}$$

$$Var(\hat{\beta}_1) = \frac{\sum X_i^2}{n\sum x_i^2}\sigma^2 = \frac{347,25}{10 * 5,025}4,0298 = 27,84$$

$$=> se(\hat{\beta}_1) = \sqrt{27,84} = 5,27$$

$$=> t_0 = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)} = \frac{94,5523}{5,27} = 17,94$$

$$\text{tra bang} => t_{0,01\,;\,8} = 2,896 < t_0 = 17,94$$

→**Bác bỏ H_0**

→ **Hệ số chặn có ý nghĩa thống kê**

→ **khi lãi suất giảm cực thấp, mức đầu tư vẫn là số >0**

4/ Kiểm định giả thiết $R^2 = 0$

$$\begin{cases} H_0 : R^2 = 0 \\ H_1 : R^2 > 0 \end{cases}$$

$$F_0 = \frac{R^2(n-2)}{1-R^2} = \frac{(0,9375).8}{1-0,9375} = 120 \qquad (R^2 = 0,9375)$$

$$\text{Chọn } \alpha = 0,01, \text{ tra bảng F ta có: } F_{\alpha\,;\,(1;n-2)} = F_{0,01;(1,8)} = 11,3 < F_0$$

➔Bác bỏ H_0 , thừa nhận kết luận $R^2>0$ là có ý nghĩa thống kê

➔ X giải thích được 93,75% sự thay đổi của Y.

Nói cách khác, lãi suất giải thích được 93,75% sự thay đổi tổng vốn đầu tư, 6,25% thay đổi còn lại do các yếu tố ngẫu nhiên gây ra

Y = 94,5523 – 9,8209 X Với X_0 = 8% năm →

*** Dự báo giá trị trung bình của Y**

$$\hat{Y}_0 = 94,5523 - 9,8209 * 8 = 15,987$$

$$\text{var}(\hat{Y}_0) = \sigma^2 \left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum_{i=1}^{n} x_i^2} \right] = 4,03 \left[\frac{1}{10} + \frac{(8 - 5,85)^2}{5,025} \right] = 4,1101$$

$$se(\hat{Y}_0) = 2,0273 \quad ; \quad t_{0,025\,;\,8} = 2,306$$

$$\hat{Y}_0 - t_{\alpha/2\,;\,(n-2)}.se(\hat{Y}_0) < E\left(Y / X_0\right) < \hat{Y}_0 + t_{\alpha/2\,;\,(n-2)}.se(\hat{Y}_0)$$

$$15,987 - 2,306 * 2,0273 < E\left(Y / X_0\right) < 15,987 + 2,306 * 2,0273$$

$$\Rightarrow \; 11,312 < E\left(Y / X_0\right) < 20,662$$

Dự báo giá trị cá biệt của Y

Ta tính

$$Var(Y_0 - \hat{Y}_0) = \sigma^2 \left[1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum_{i=1}^{n} x_i^2} \right]$$

$$Var(Y_0 - \hat{Y}_0) = 4,03 \left[1 + \frac{1}{10} + \frac{(8 - 5,85)^2}{5,025} \right] = 8,14$$

$$se(Y_0 - \hat{Y}_0) = 2,8531$$

$$\hat{Y}_0 - t_{\alpha/2\ ;\ (n-2)}.se(Y_0 - \hat{Y}_0) < Y_0 < \hat{Y}_0 + t_{\alpha/2\ ;\ (n-2)}.se(Y_0 - \hat{Y}_0)$$

$$15,987 - 2,306 * 2,8513 < Y_0 < 15,987 + 2,306 * 2,8513$$

$$9,4078 < Y_0 < 22,5643$$

f/ $H_0 : \beta_2 = -12$

$H_1 : \beta_2 < -12$

$$t_o = \frac{\hat{\beta}_2 - (-12)}{se(\hat{\beta}_2)} = \frac{-9,82 + 12}{0,8955} = \frac{2,18}{0,8955} = 2,4334$$

$$t_{\alpha\,;\,(n-2)} = t_{0,05\,;\,8} = 1,86$$

$$t_0 = 2,4334 > -t_{0,05\,;\,8} = -1,86$$

$\Rightarrow$ Chấp nhận H_0

Vậy, khi lãi suất tăng 1 % năm, tổng vốn đầu tư có thể giảm 12 tỷ $

Bài tập 2

Một mẫu khảo sát về tổng cầu vay vốn (Y – tỷ $) với lãi suất cho vay (X - % năm) của ngân hàng tại tỉnh LGC qua 12 năm liền như sau:

X	5,0	5,5	5,5	6,0	6,2	6,5	6,5	6,8	7,0	7,0	7,5	7,5
Y	80	76	80	74	72	70	71	69	70	67	64	62

1/ Lập mô hình HQTT có dạng $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + e_i$

2/ Nêu ý nghĩa kinh tế các HSHQ

3/ Kiểm định giả thiết $H_0 : \beta_2 = 0$; $H_1 : \beta_2 \neq 0$ với mức ý nghĩa 5% và nêu ý nghĩa của kết quả.

4/ Đánh giá mức phù hợp của mô hình với độ tin cậy 95%

5/ Dự báo giá trị trung bình của tổng cầu vay vốn với mức lãi suất 7,3% năm với độ tin cậy 95%.

Bài tập 3

Khảo sát mối liên quan giữa số lượng sản phẩm A tiêu thụ (Y–nghìn SP) với giá bán đơn vị (triệu $/SP), được số liệu:

X	4,0	6,4	5,3	4,6	5,8	6,8	4,2	7,3	6,1	7,5
Y	12	10,4	11,0	11,6	10,7	10,1	12,2	9,7	10,8	9,5

1/ Lập mô hình HQTT

2/ Nêu ý nghĩa kinh tế các HSHQ

3/ Kiểm định giả thiết $H_0 : \beta_2 = 0$; $H_1 : \beta_2 \neq 0$ với mức ý nghĩa 5% và nêu ý nghĩa của kết quả.

4/ Đánh giá mức phù hợp của mô hình với độ tin cậy 95%

5/ Dự báo giá trị trung bình của tổng lượng hàng bán được với mức giá 7,0 triệu/SP với độ tin cậy 95%.

6/ Có ý kiến cho rằng xu hướng tiêu thụ biên tế trong mẫu quan sát này là 0,9, với độ tin cậy 95% bạn nhận xét ý kiến này thế nào?

Bài tập 4

Khảo sát về thu nhập (X – triệu $/tháng) và chi tiêu cá nhân (Y – triệu/tháng) của một mẫu, được kết quả như sau:

X	3,0	6,3	7,6	4,2	5,5	3,5	5,0	6,7	7,0	4,5
Y	3,1	5,5	6,5	4,0	4,8	3,2	5,0	6,4	6,2	4,2

1/ Ước lượng mô hình HQTT

2/ Nêu ý nghĩa kinh tế các HSHQ

3/ Kiểm định giả thiết $H_0 : \beta_2 = 0$; $H_1 : \beta_2 \neq 0$ và nêu ý nghĩa của kết quả.

4/ Bạn nhận xét như thế nào khi cho rằng xu hướng tiêu dùng biên trong trường hợp này không lớn hơn 0,4?

5/ Đánh giá mức phù hợp của mô hình

6/ Dự báo giá trị trung bình của mức chi tiêu hàng tháng khi thu nhập bình quân 6,0 triệu/tháng.

Cho biết: độ tin cậy 95%.

(1). **Y = 0,630877+0,799085 X2** ; R2 = 0.965726

(2). **Ý nghĩa kinh tế của các HSHQ**

• β2= 0,799085 > 0 → X và Y đồng biến.

Ymin = β1= 0,630877 (X2 = 0): Khi thu nhập bằng 0, chi tiêu trung bình tối thiểu là 0,630877 triệu đồng tháng → phù hợp với lý thuyết kinh tế

• β2= 0,799085 > 0 → X và Y đồng biến. Khi thu nhập tăng (giảm) 1 triệu đồng tháng, chi tiêu tăng (giảm) 0,799085 đồng tháng, các yếu tố khác không đổi → phù hợp lý thuyết kinh tế

Equation: UNTITLED Workfile: UNTITLED::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: Y
Method: Least Squares
Date: 04/22/10 Time: 19:15
Sample: 1 10
Included observations: 10

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.630877	0.294286	2.143750	0.0644
X2	0.799085	0.053223	15.01384	0.0000

R-squared	0.965726	Mean dependent var	4.890000
Adjusted R-squared	0.961442	S.D. dependent var	1.260908
S.E. of regression	0.247594	Akaike info criterion	0.222803
Sum squared resid	0.490422	Schwarz criterion	0.283320
Log likelihood	0.885986	F-statistic	225.4154
Durbin-Watson stat	1.465025	Prob(F-statistic)	0.000000

(3). Kiểm định β_2: $H_0 : \beta_2 = 0$; $H_1 : \beta_2 \neq 0$

Độ tin cậy 96% → $\alpha = 0,04$ → $(\alpha / 2) = 0,02$

$$t_0 = \frac{\hat{\beta}_2}{se(\beta_2)} = \frac{0,799085}{0,053223} = 15,0139 > t_{\alpha/2;(n-2)} = 2,449$$

($\alpha = 0,04$ → $\alpha/2 = 0,02$)

→ Bác bỏ giả thiết H_0

→ β_2 khác không có ý nghĩa thống kê, nghĩa là thu nhập có ảnh hưởng lên chi tiêu (biến X thực sự có ảnh hưởng lên biến Y)

(4). Kiểm định H_0: $\beta 2 = 0,4$; H_1 : $\beta 2 > 0,4$

Bài tập 5

Khảo sát về thu nhập (X – triệu $/tháng) và chi tiêu cá nhân (Y – triệu/tháng) của một mẫu, được kết quả như sau:

X	3,0	6,3	7,6	4,2	5,5	3,5	5,0	6,7	7,0	4,5
Y	3,1	5,5	6,5	4,0	4,8	3,2	5,0	6,4	6,2	4,2

1/ Ước lượng mô hình HQTT

2/ Nêu ý nghĩa kinh tế các HSHQ

3/ Kiểm định giả thiết $H_0 : \beta_2 = 0$; $H_1 : \beta_2 \neq 0$ với độ tin cậy 95% và nêu ý nghĩa của kết quả.

4/ Đánh giá mức phù hợp của mô hình với độ tin cậy 95%

5/ Dự báo giá trị trung bình của mức chi tiêu hàng tháng khi thu nhập bình quân 6,0 triệu/tháng với độ tin cậy 95%.

6/ Có ý kiến cho rằng xu hướng tiêu dùng biên là 0,8, với độ tin cậy 95%, bạn nhận xét ra sao về ý kiến trên?

Bài tập 6

Khảo sát về thu nhập (X – triệu $/tháng) và chi tiêu cá nhân (Y – triệu/tháng) của một mẫu, được kết quả như sau:

X	3,0	6,3	7,6	4,2	5,5	3,5	5,0	6,7	7,0	4,5
Y	3,1	5,5	6,5	4,0	4,8	3,2	5,0	6,4	6,2	4,2

1/ Ước lượng mô hình HQTT

2/ Nêu ý nghĩa kinh tế các HSHQ

3/ Kiểm định giả thiết $H_0 : \beta_2 = 0$; $H_1 : \beta_2 \neq 0$ với độ tin cậy 95% và nêu ý nghĩa của kết quả ($t_{\alpha/2;\,8} = 2{,}306$)

4/ Dự báo giá trị trung bình của mức chi tiêu hàng tháng khi thu nhập bình quân 6,0 triệu/tháng với độ tin cậy 95%.

5/ Có ý kiến cho rằng xu hướng tiêu dùng biên không lớn hơn 0,6, với độ tin cậy 95%, bạn nhận xét ra sao về ý kiến trên?

6/ Có ý kiến cho rằng xu hướng tiêu dùng biên là 0,8, với độ tin cậy 95%, bạn nhận xét ra sao về ý kiến trên?