Chöông 2

MÔ HÌNH HỒI QUY HAI BIẾN
ƯỚC LƯỢNG VÀ KIỂM ĐỊNH GIẢ THIẾT

1- PHƯƠNG PHÁP OLS (Ordinary Least Square)

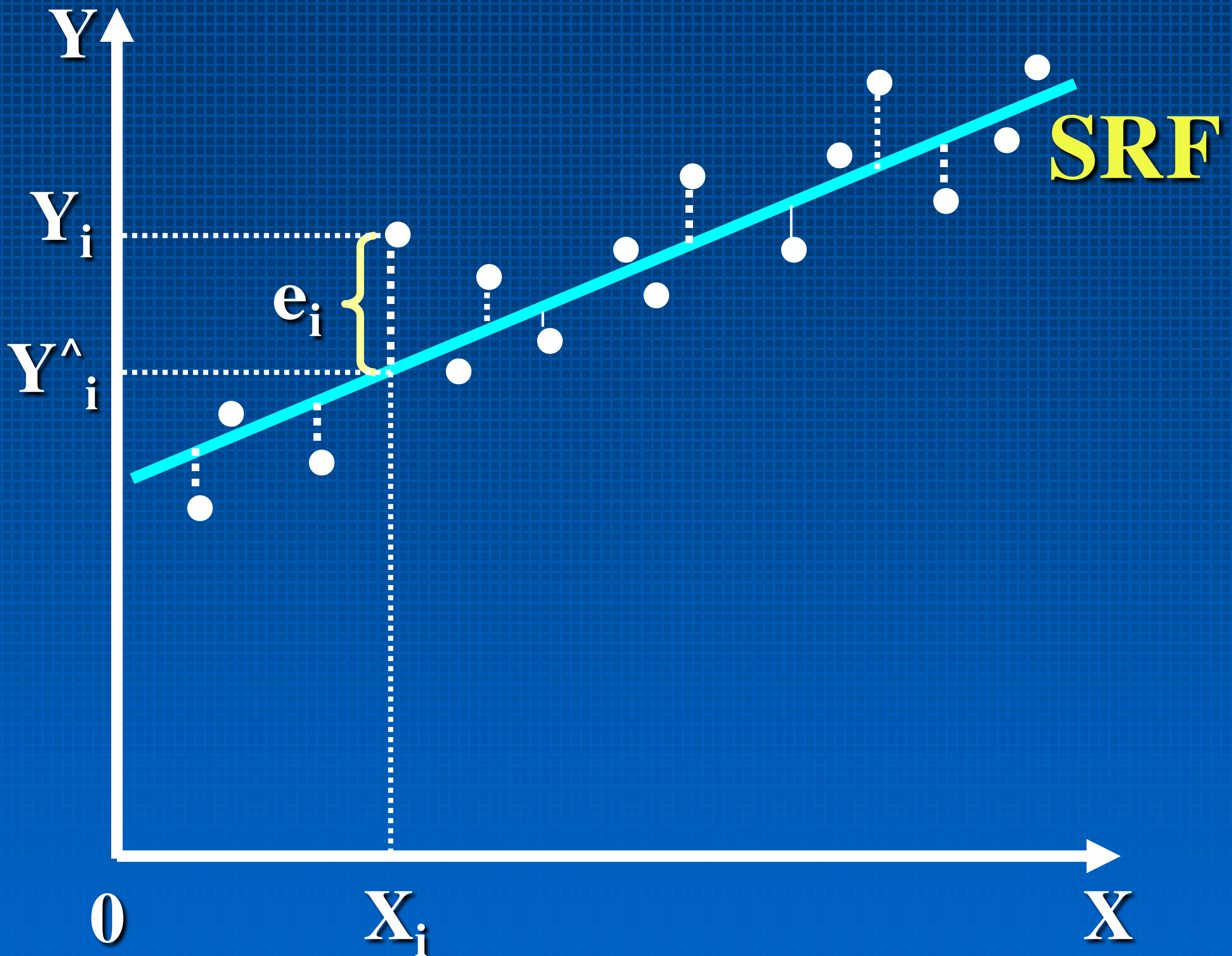
Giả sử có một mẫu gồm n quan sát (Y_i, X_i), (i = 1, 2, . . . , n)

Theo pp OLS, ta phải tìm $\hat{Y}_i$ sao cho nó càng gần với giá trị thực (Y_i) càng tốt, tức phần dư:

$$e_i = Y_i - \hat{Y}_i$$

$$= Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i$$

càng nhỏ càng tốt


Y
Y_i
e_i
$\hat{Y}_i$
SRF
0
X_i
X

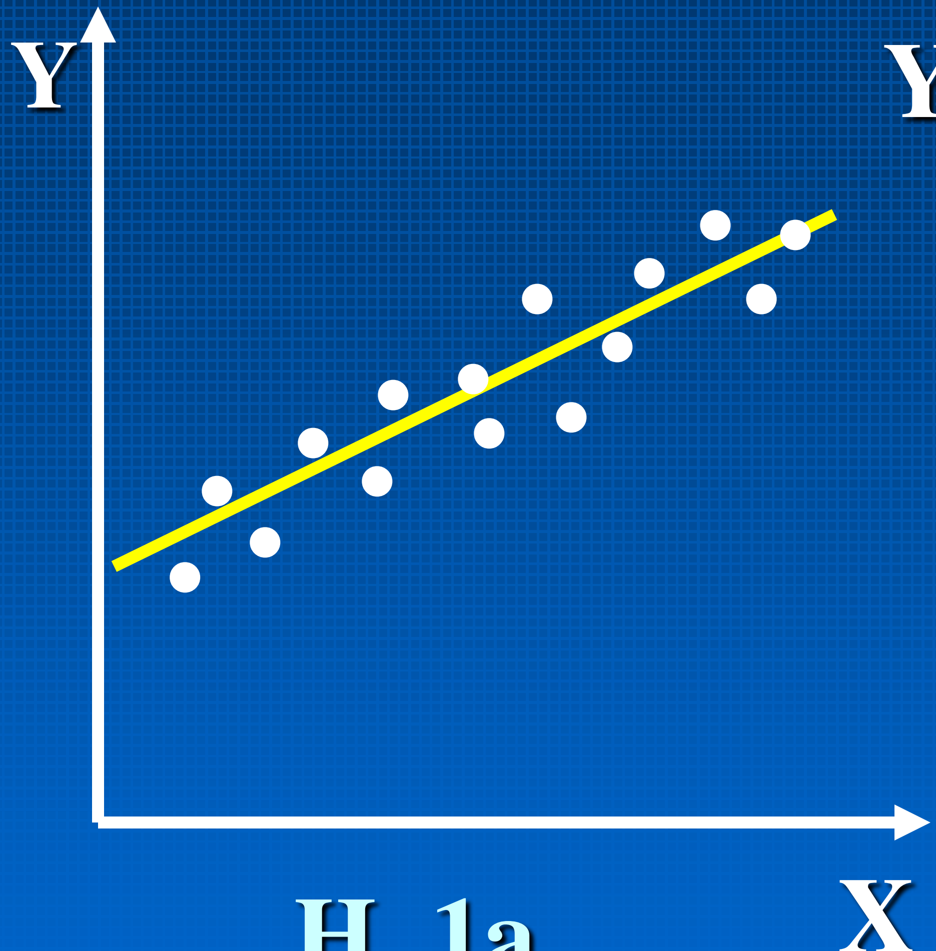
Do e_i có thể dương, có thể âm, nên ta cần tìm SRF sao cho tổng bình phương của các phần dư đạt cực tiểu.

Tức $\hat{\beta}_1, \hat{\beta}_2$ phải thoả mãn điều kiện:

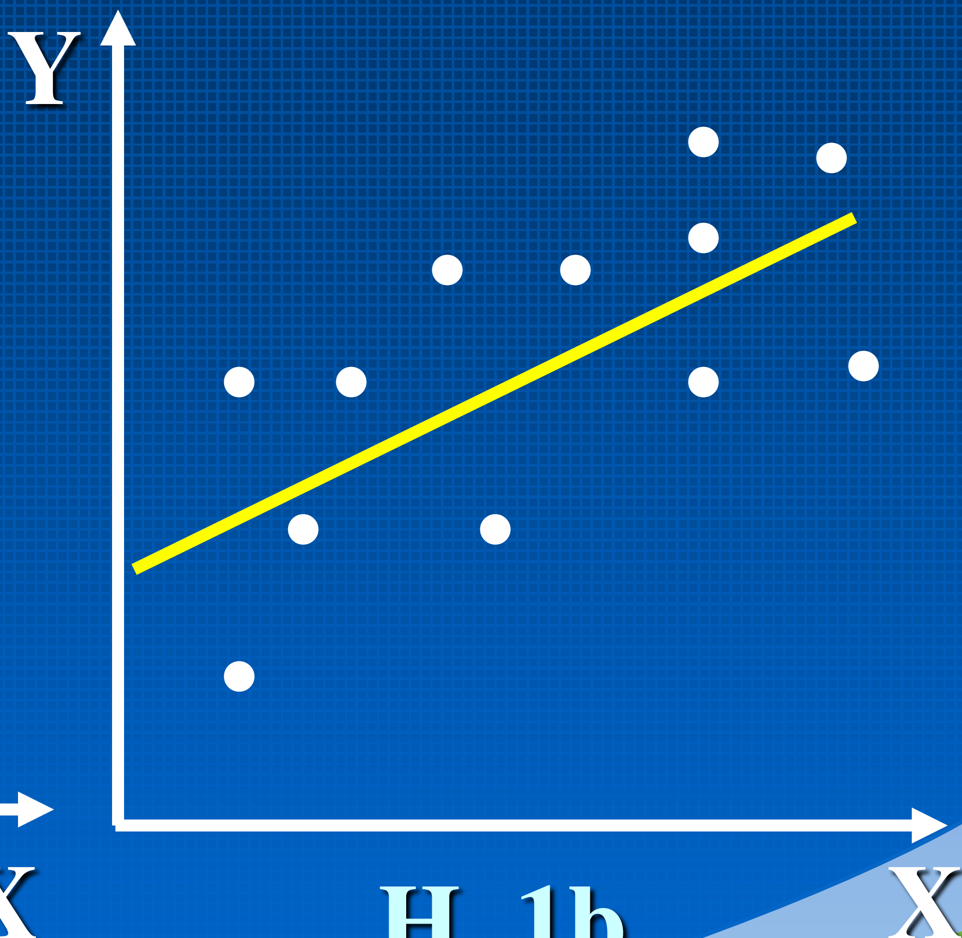
$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}\left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i\right)^2 \Rightarrow \min \qquad (*)$$

ĐK (*) có nghĩa là tổng bình phương các sai lệch giữa giá trị thực tế q.sát được (Y_i) và giá trị tính theo hàm hồi qui mẫu ($\hat{Y}_i$) là nhỏ nhất.

Tức đường hồi qui mẫu với $\hat{\beta}_1, \hat{\beta}_2$ thỏa mãn điều kiện (*) sẽ là đường thẳng "gần nhất" với tập hợp các điểm quan sát, do vậy nó được coi là đường thẳng "tốt nhất", "phù hợp nhất" trong lớp các đường hồi qui mẫu có thể dùng để ước lượng hàm (2.2).



H. 1a



H. 1b

Do Y_i, X_i (i = 1, 2, . . . , n) đã biết, nên

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

là hàm của $\hat{\beta}_1, \hat{\beta}_2$

Vì vậy ta cần tìm $\hat{\beta}_1, \hat{\beta}_2$ sao cho:

$$f(\hat{\beta}_1, \hat{\beta}_2) = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2 \to \min$$

Tức $\hat{\beta}_1, \hat{\beta}_2$ là nghiệm của hệ p.t:

$$
\begin{cases}
\dfrac{\partial f(\hat{\beta}_1,\hat{\beta}_2)}{\partial \hat{\beta}_1} = \sum_{i=1}^{n} 2(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)(-1) = 0 \\
\dfrac{\partial f(\hat{\beta}_1,\hat{\beta}_2)}{\partial \hat{\beta}_2} = \sum_{i=1}^{n} 2(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)(-X_i) = 0
\end{cases}
$$

Hay:

(2.6)

$$
\begin{cases}
n\hat{\beta}_1 + \hat{\beta}_2 \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} Y_i \\
\hat{\beta}_1 \sum_{i=1}^{n} X_i + \hat{\beta}_2 \sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} X_i . Y_i
\end{cases}
$$

Hệ phương trình (2.6) gọi là hệ phương trình chuẩn.

Giải hệ p.tr này ta được:

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

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

Có thể tính $\hat{\beta}_2$ theo công thức:

$$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2}$$

Trong đó: $x_i = X_i - \overline{X}$; $y_i = Y_i - \overline{Y}$

- ***Xét điều kiện đủ:***

Ta có ma trận Hessian như sau:

$$H = \begin{pmatrix} f''_{\hat{\beta}_1\hat{\beta}_1} & f''_{\hat{\beta}_1\hat{\beta}_2} \\ f''_{\hat{\beta}_2\hat{\beta}_1} & f''_{\hat{\beta}_2\hat{\beta}_2} \end{pmatrix} = \begin{pmatrix} 2n & 2\sum X_i \\ 2\sum X_i & 2\sum X_i^2 \end{pmatrix}$$

Với:

$$|H| = 4\left[n\sum X_i^2 - \left(\sum X_i\right)^2\right]$$

$$= 4n\left(\sum X_i^2 - n\overline{X}^2\right)$$

$$= 4n\sum x_i^2 > 0$$

Vậy ma trận H xác định dương nên $\hat{\beta}_1, \hat{\beta}_2$ xác định bằng các công thức trên là điểm cực tiểu của hàm $f(\hat{\beta}_1, \hat{\beta}_2)$.

Thí dụ 2:

Bảng sau cho số liệu về lượng bán được (Y- tấn/tháng) và đơn giá của hàng A (X- ngàn đồng/kg)

Y_i	10	6	9	5	4	2
X_i	1	4	2	5	5	7

Giả sử Y, X có q.hệ t.quan t.t. Hãy ước lượng hàm h.qui của Y theo X.

Giải: **Từ số liệu q.sát của X và Y cho ở bảng trên ta tính được:**

$\sum Y_i = 36 \rightarrow \bar{Y}=6$ $\sum X_i = 24 \rightarrow \bar{X}=6$

$\sum X_i^2 = 120 \rightarrow \sum x_i^2 = 24$ $\sum X_i Y_i = 111;$

$\sum x_i y_i = -33$

$$\hat{\beta}_2 = \frac{-33}{24} = -1{,}375$$

$$\hat{\beta}_1 = 6 - (-1{,}375) \times 4 = 11{,}5$$

Hàm hồi qui tt mẫu của *chi tiêu* theo *thu nhập* là:

$$\hat{Y}_i = 11{,}5 - 1{,}375 X_i$$

CÁC GIẢ THIẾT CỦA PHƯƠNG PHÁP OLS

❶ **Biến giải thích là phi ng.n**

❷ **Kỳ vọng toán của U_i bằng 0,**

tức: $E(U_i/X_i) = 0$

❸ **Các U_i có p.sai bằng nhau**

④ **Không có t.quan giữa các U_i, tức**

$$\mathrm{cov}(U_i, U_j) = 0 \qquad (i \neq j)$$

⑤ **U_i và X_i không t.quan với nhau, tức**

$$\mathrm{cov}(U_i, X_i) = 0$$

ĐỊNH LÝ GAUSS-MARKOV

Với các giả thiết 1-5 của MH hồi qui tt cổ điển, các ước lượng của PP OLS sẽ là các ước lượng *<u>tuyến tính</u>*, *<u>không chệch</u>* và *<u>có p.sai nhỏ nhất</u>*.

Đối với hàm hai biến,

$\hat{\beta}_1$, $\hat{\beta}_2$ tương ứng là các ước lượng t.tính, không chệch, có p.sai nhỏ nhất của β_1, β_2.

2- Phương sai và sai số chuẩn của các ước lượng

$$\operatorname{var}(\hat{\beta}_1) = \frac{\sum_{i=1}^{n} X_i^2}{n \sum_{i=1}^{n} x_i^2} \sigma^2$$

$$\mathrm{se}(\hat{\beta}_1) = \sqrt{\operatorname{var}(\hat{\beta}_1)}$$

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2}$$

$$\operatorname{se}(\hat{\beta}_2) = \sqrt{\operatorname{var}(\hat{\beta}_2)}$$

Trong đó: $\sigma^2 = var(U_i)$

se: sai số chuẩn *(Standard Erorr)*

σ^2 được ước lượng bằng ước lượng không chệch $\hat{\sigma}^2$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-2}$$

$\hat{\sigma} = \sqrt{\hat{\sigma}^2}$ là sai số chuẩn

3. HỆ SỐ XÁC ĐỊNH

$$TSS = \sum_{i=1}^{n} \left(Y_i - \overline{Y}\right)^2 = \sum Y_i^2 - n.\left(\overline{Y}\right)^2$$

TSS (*Total Sum of Squares*)

$$ESS = \sum_{i=1}^{n} \left(\hat{Y}_i - \overline{Y}\right)^2 = (\hat{\beta}_2)^2 \sum_{i=1}^{n} x_i^2$$

ESS (*Explained Sum of Squares*)

$$RSS = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2$$

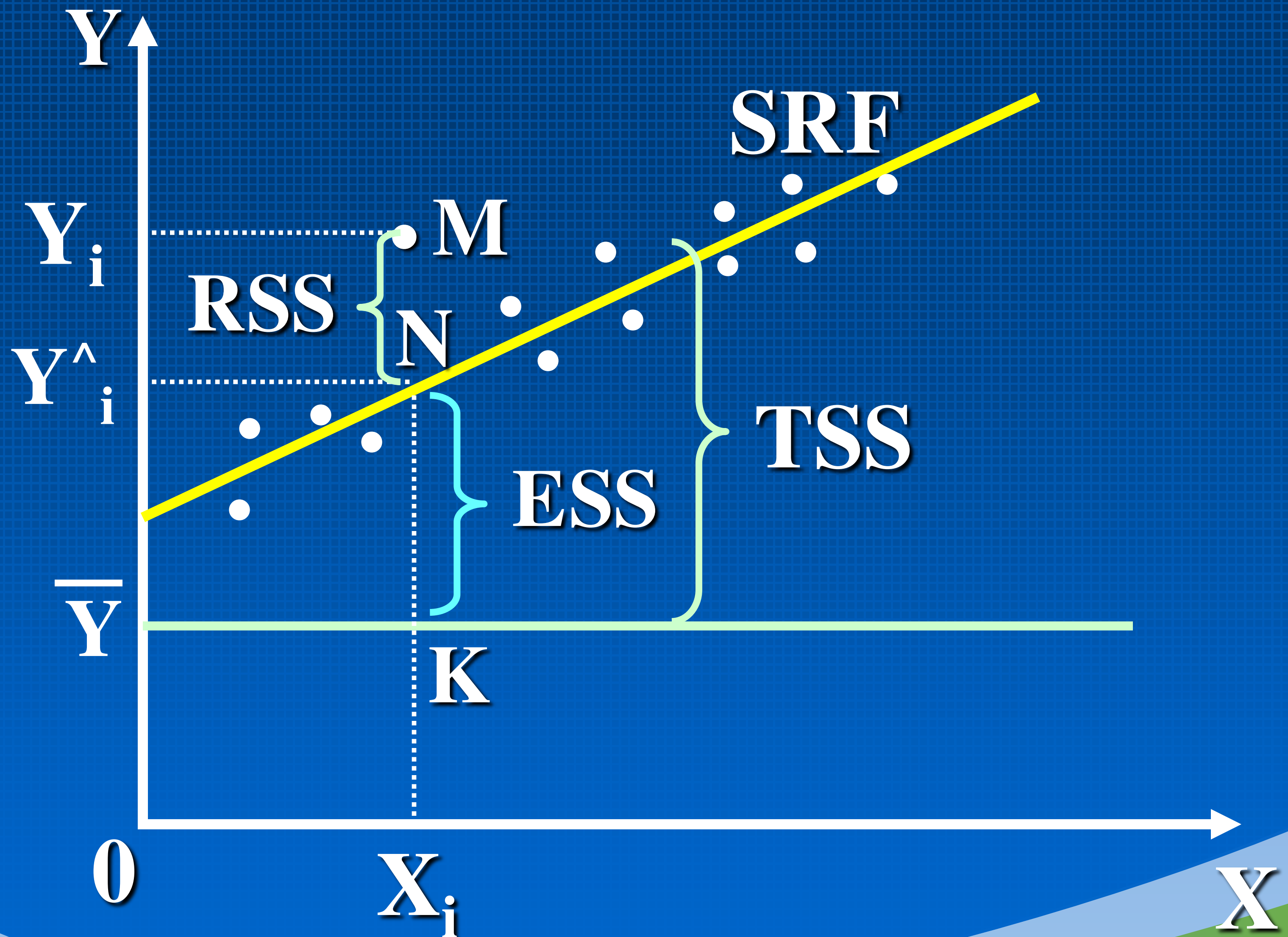
RSS (*Residual Sum of Squares*)

TSS = ESS + RSS

Nếu hàm hồi qui mẫu phù hợp tốt với các số liệu quan sát thì ESS sẽ càng lớn hơn RSS.

Nếu tất cả các giá trị q.sát của Y đều nằm trên SRF thì ESS sẽ bằng TSS và do đó RSS = 0.

Ngược lại, nếu hàm hồi qui mẫu kém phù hợp với các giá trị quan sát thì RSS sẽ càng lớn hơn ESS.



Y
SRF
Y_i
M
RSS
N
$Y^\wedge_i$
TSS
ESS
$\bar{Y}$
K
0
X_i
X

$$R^2 = \frac{ESS}{TSS}$$

R^2 - hệ số xác định

(*coefficient of determination*)

$$\mathbf{0 \le R^2 \le 1}$$

$R^2 = 1$ thì đường h.q phù hợp "*hoàn hảo*", tất cả các sai lệch của Y (so với giá trị TB) đều giải thích được bởi MH hồi quy.

Khi $R^2 = 0$ chứng tỏ X và Y không có quan hệ.

Với số liệu ở thí dụ 2:

$$\sum Y_i^2 = 132100$$

$$TSS = 132100 - 10(111)^2 = 8890$$

$$ESS = (0,5091)^2 \, 33000 = 8552,73$$

$$R^2 = (8552,73/8890) = 0,9621$$

Trong hàm hồi qui mẫu, biến X (thu nhập) giải thích được 96,21% sự thay đổi của biến Y (chi tiêu). Vậy mức độ phù hợp của SRF là khá cao.

4. HỆ SỐ TƯƠNG QUAN

Hệ số tương quan r là số đo mức độ chặt chẽ của q.hệ tuyến tính giữa X và Y

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

Hay:

$$r = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2 \cdot \sum y_i^2}}$$

Có thể chứng minh được:

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

Trong trường hợp này dấu cuả r trùng với dấu của $\hat{\beta}_2$

CÁC TÍNH CHẤT CỦA HỆ SỐ TƯƠNG QUAN TUYẾN TÍNH

❶ r có thể âm hoặc dương, dấu của r phụ thuộc vào dấu của hệ số góc.

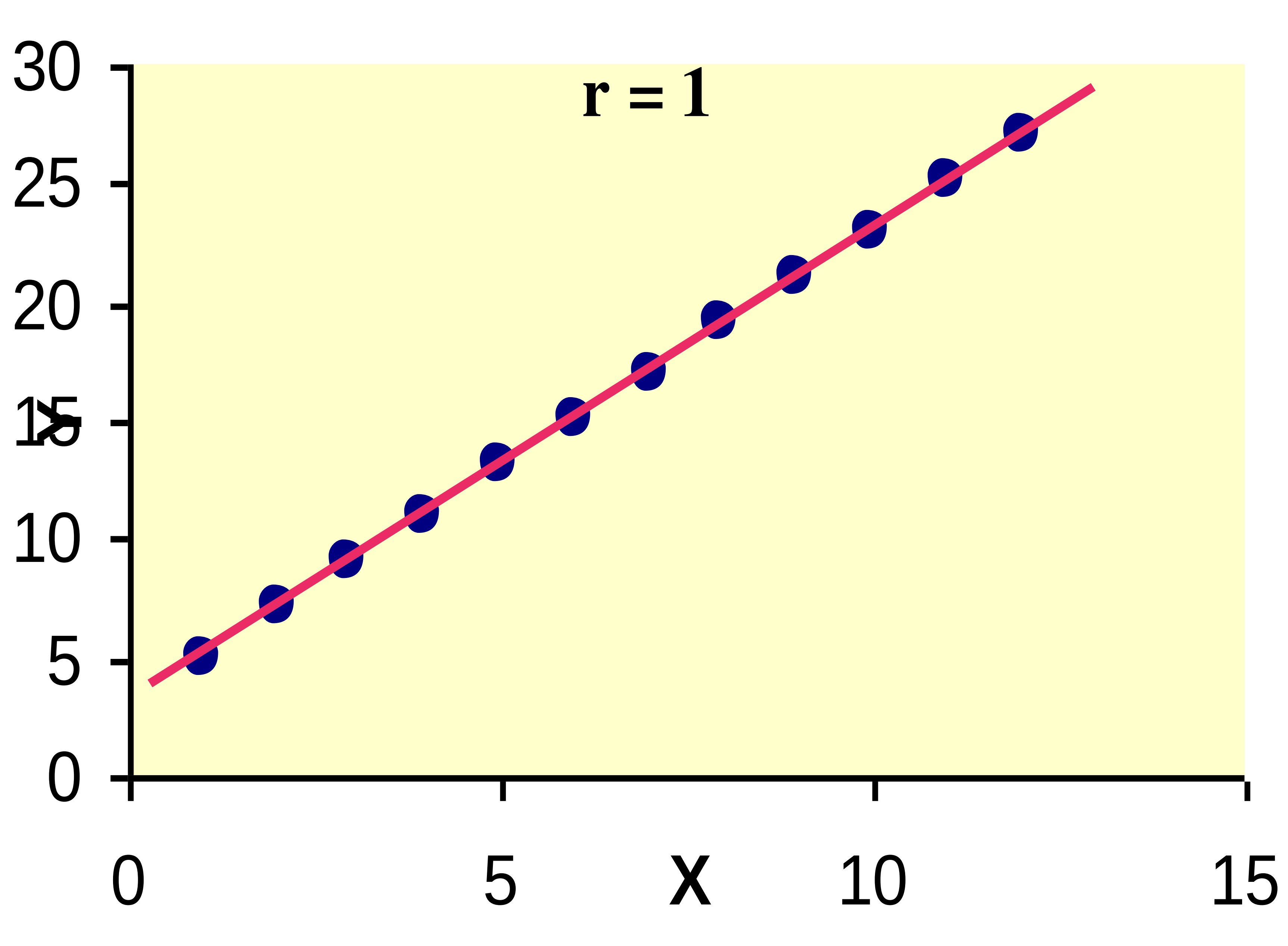
❷ r lấy giá trị trong khoảng (-1; +1)

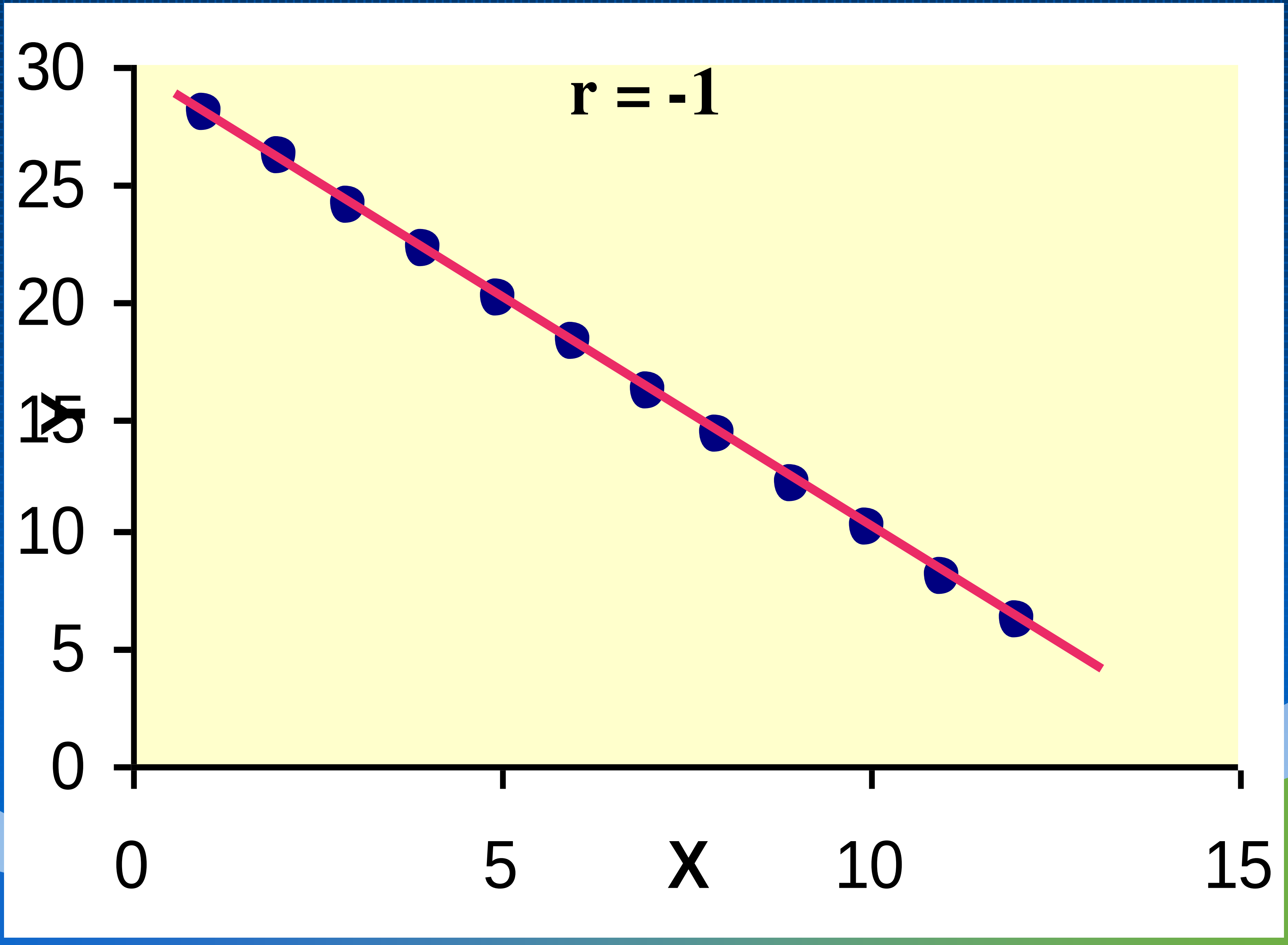
❸ r có tính chất đối xứng $r_{XY} = r_{YX}$

❹ r độc lập với gốc tọa độ và các tỷ lệ.

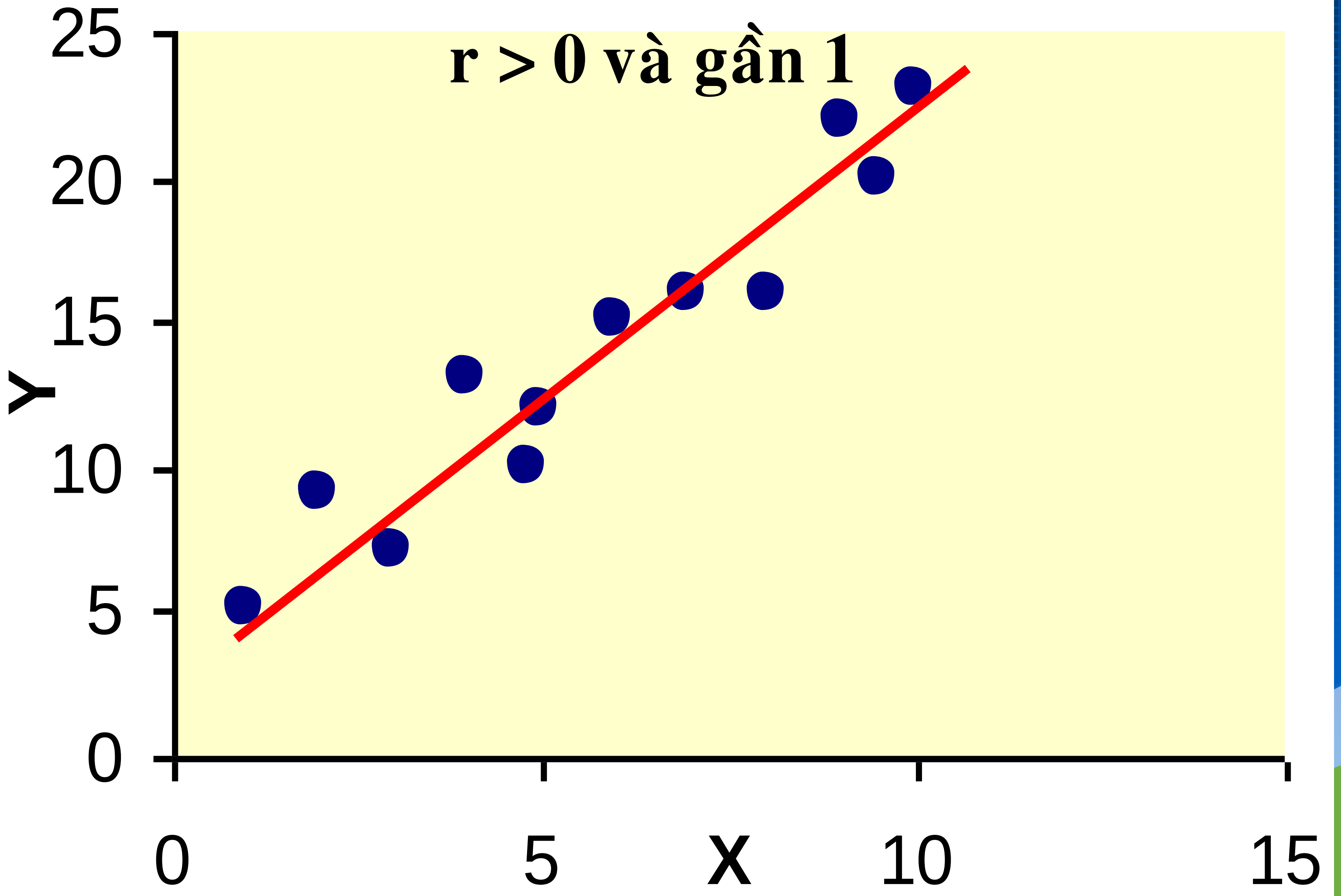
⑤ Nếu X, Y độc lập thì $r_{XY} = 0$; nhưng khi $r_{XY} = 0$ thì điều đó không có nghĩa là hai biến này độc lập.

⑥ r chỉ đo mức độ phụ thuộc tuyến tính, r không có ý nghĩa khi mô tả quan hệ phi tuyến.

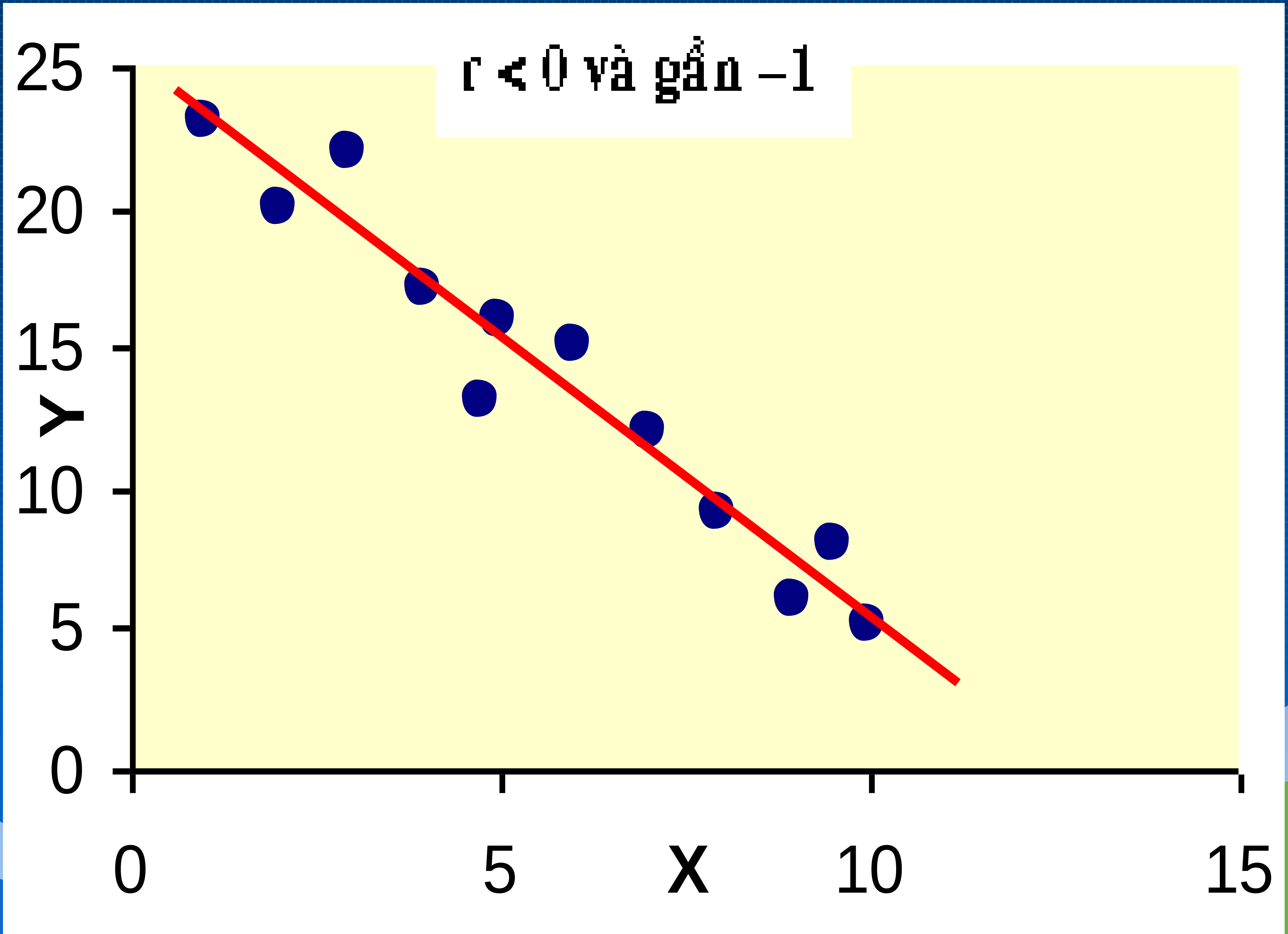

r = 1
30
25
20
15
10
5
0
0
5
10
15
X
Y



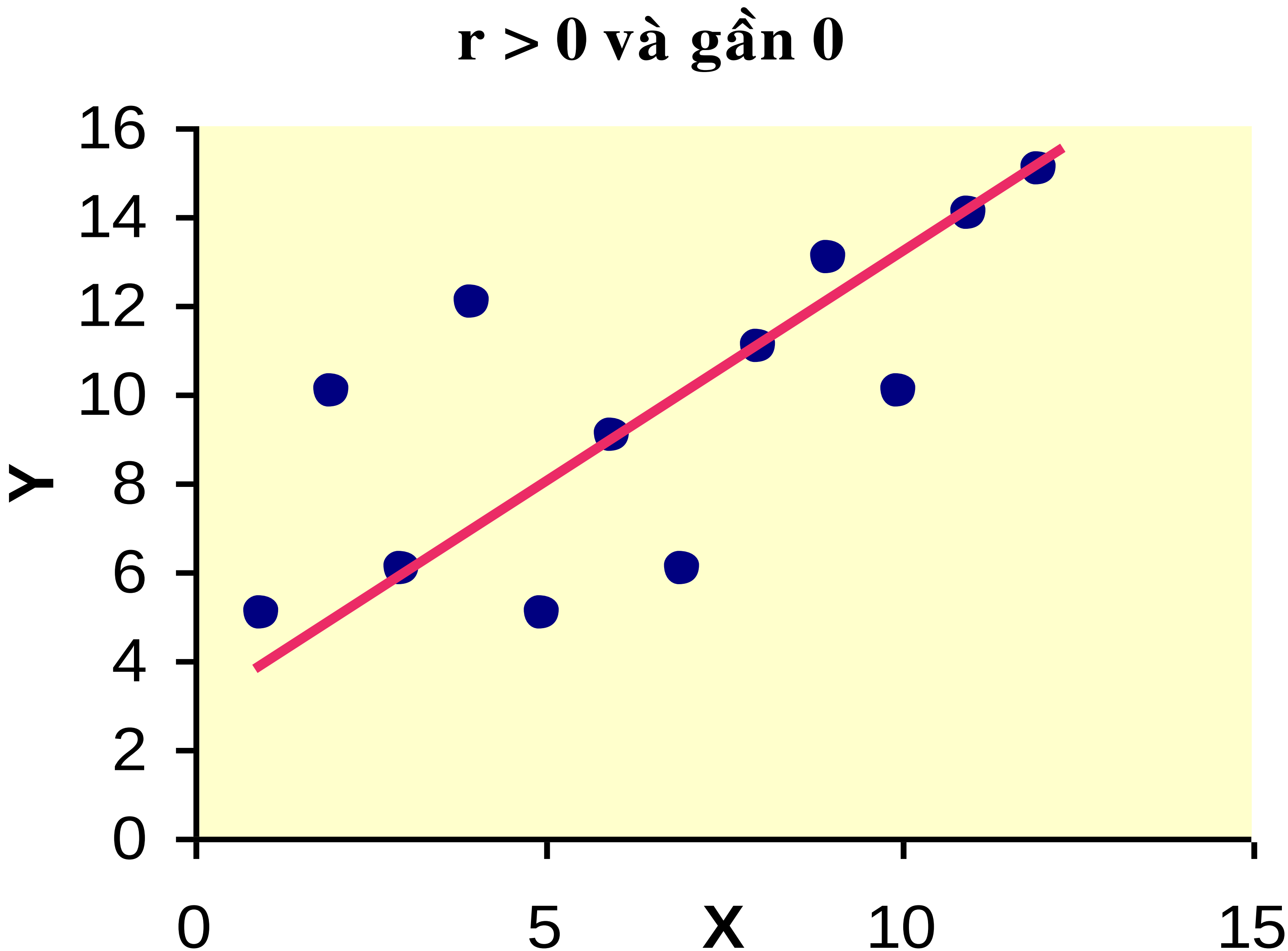
r = -1
30
25
20
15
10
5
0
Y
0
5
X
10
15

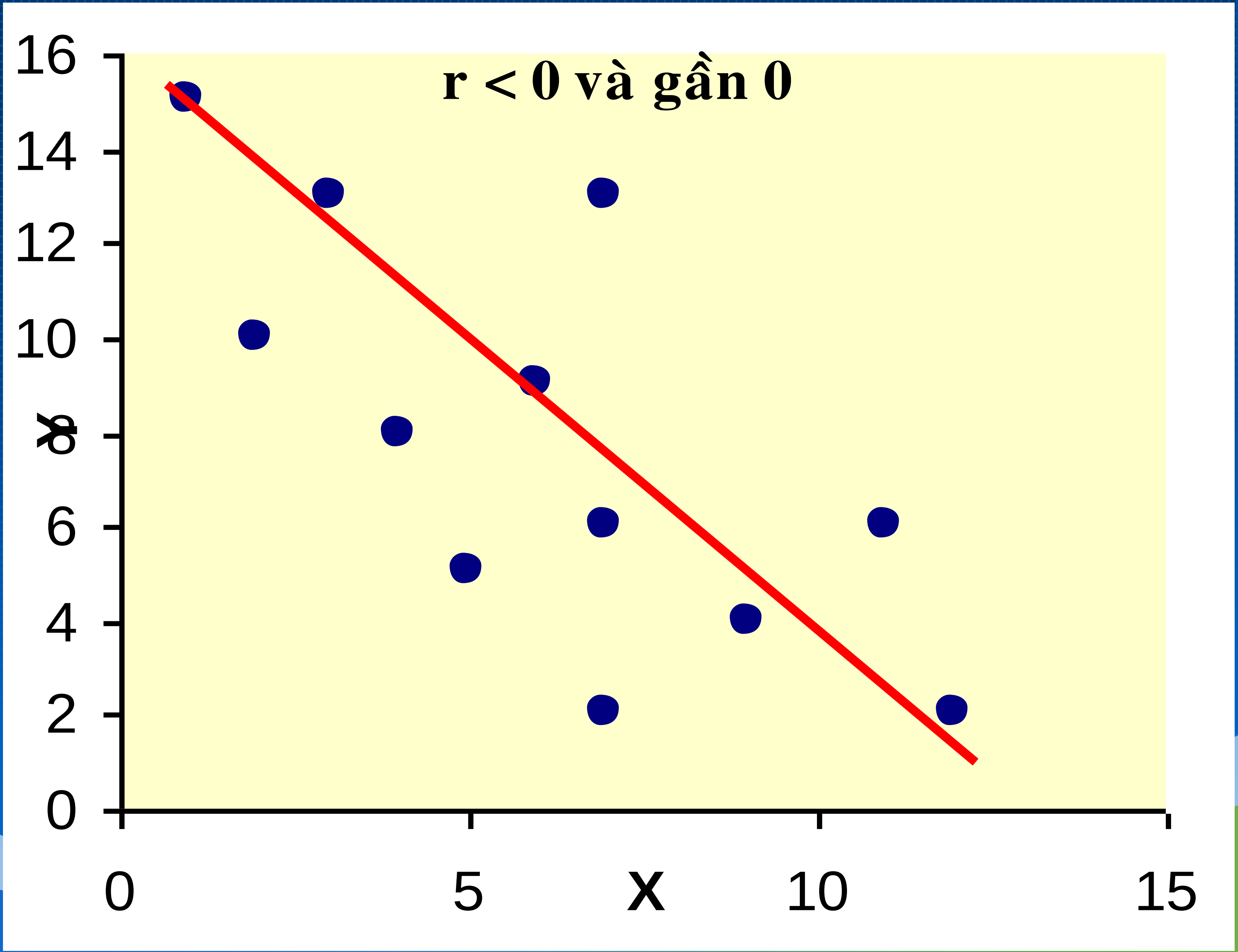


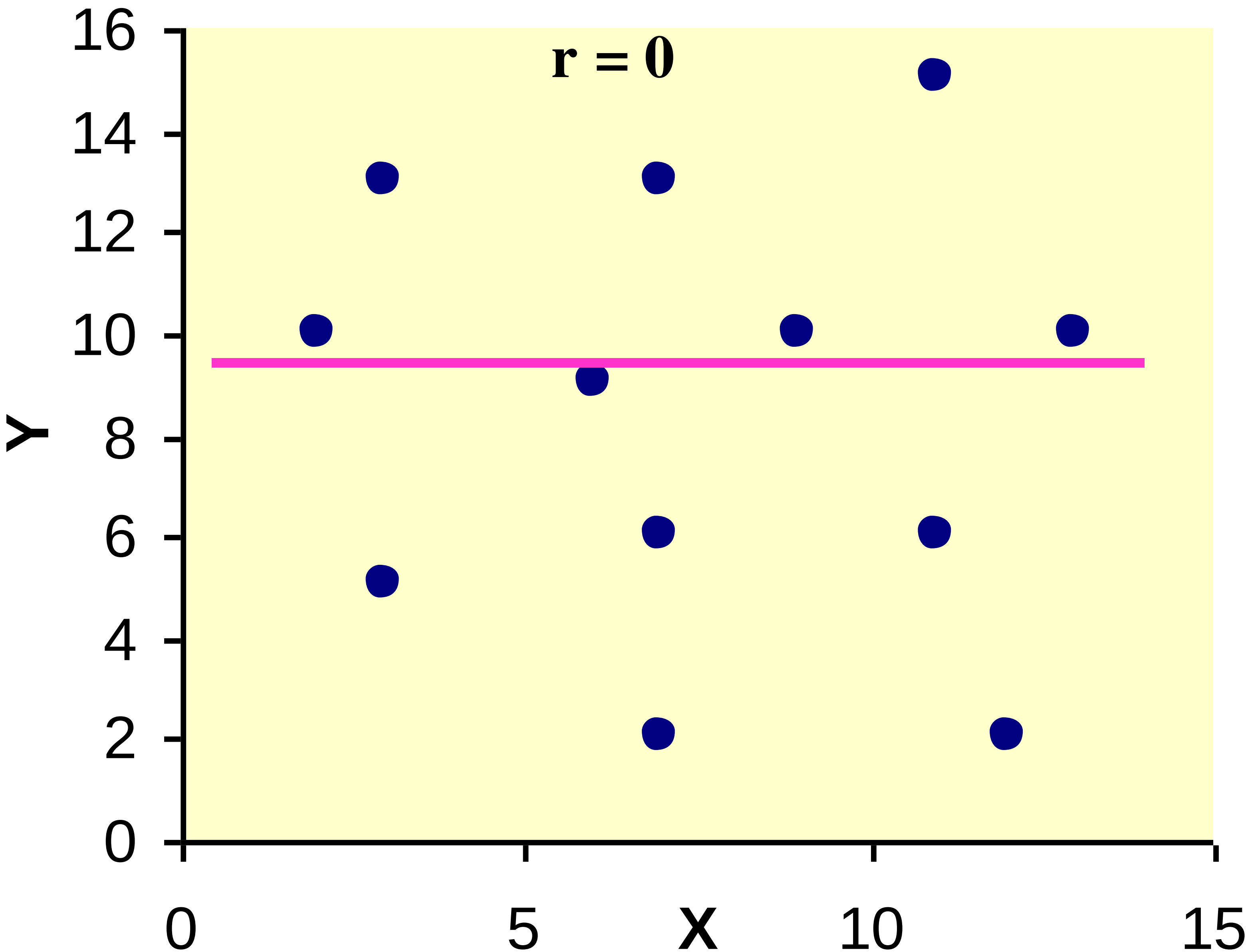
r > 0 và gần 1
25
20
15
10
5
0
Y
0
5
X
10
15

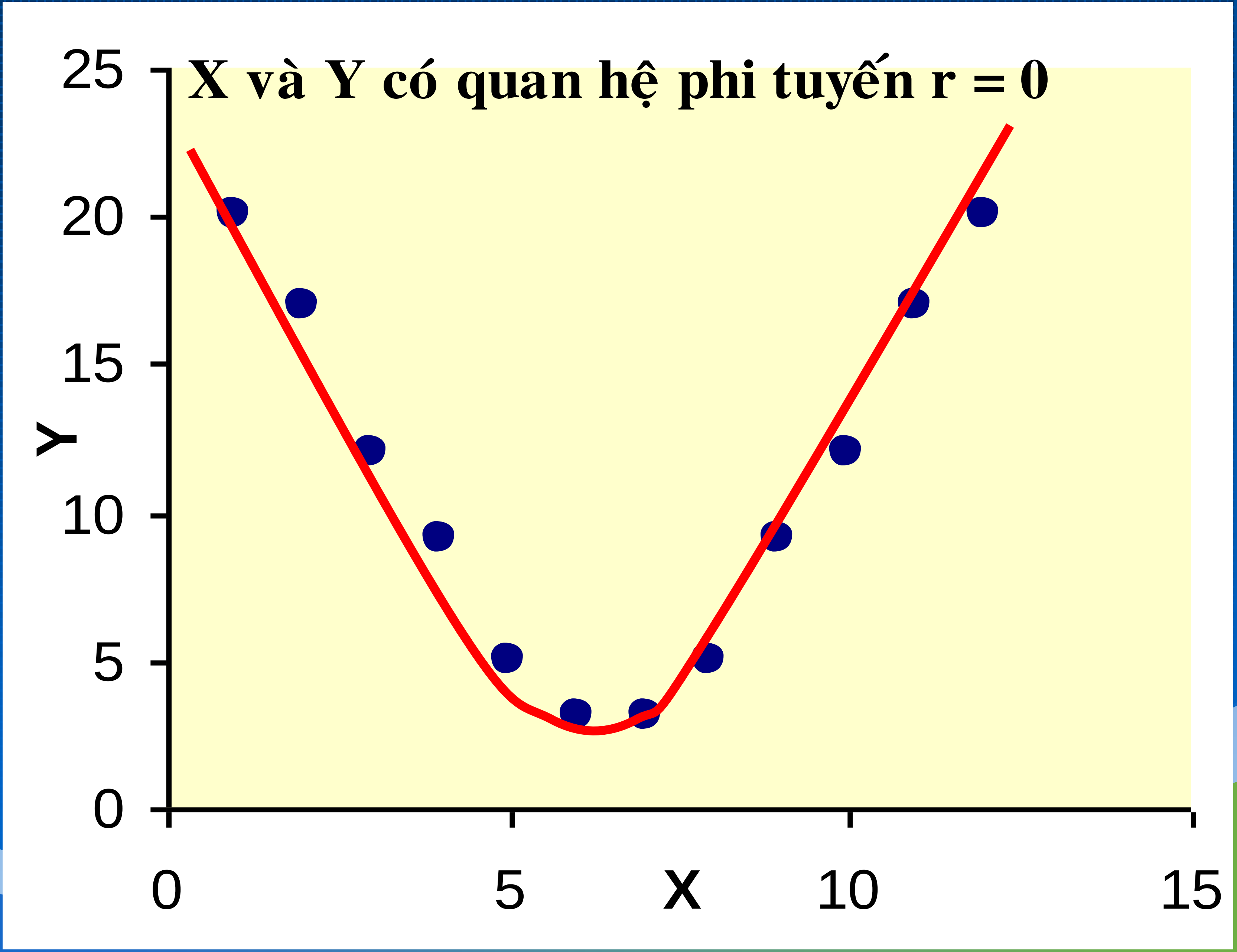


r < 0 và gần –1
25
20
15
10
5
0
Y
0
5
X
10
15


r > 0 và gần 0
16
14
12
10
8
6
4
2
0
Y
0
5
X
10
15


r < 0 và gần 0
Y
X
16
14
12
10
8
6
4
2
0
0
5
10
15


r = 0
Y
X
16
14
12
10
8
6
4
2
0
0
5
10
15


X và Y có quan hệ phi tuyến r = 0
25
20
15
10
5
0
Y
0
5
X
10
15

- r > 0 thì X ,Y có tương quan thuận (tương quan dương). Tức X tăng thì giá trị trung bình của Y tăng; X giảm thì giá trị trung bình của Y giảm

- **r < 0 thì X ,Y có tương quan nghịch (tương quan âm). Tức X tăng thì giá trị trung bình của Y giảm; X giảm thì giá trị trung bình của Y tăng.**

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

5- PHÂN PHỐI XÁC SUẤT CỦA CÁC ƯỚC LƯỢNG

Giả thiết 6:

U_i có p.phối chuẩn $N(0, \sigma^2)$

Với các g.thiết trên, các ước lượng $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\sigma}^2$ có các t/chất sau đây:

1. Chúng là các ước lượng không chệch.
2. Có phương sai cực tiểu.
3. Khi số quan sát đủ lớn thì các ước lượng này xấp xỉ với giá trị thực của phân phối.

④ $\hat{\beta}_1 \sim N(\beta_1, \sigma^2_{\hat{\beta}_1})$

Suy ra $Z = (\hat{\beta}_1 - \beta_1)/\sigma_{\hat{\beta}_1} \sim N(0, 1)$

⑤ $\hat{\beta}_2 \sim N(\beta_2, \sigma^2_{\hat{\beta}_2})$

suy ra $Z = (\hat{\beta}_2 - \beta_2)/\sigma_{\hat{\beta}_2} \sim N(0, 1)$

⑥ $(n-2)\hat{\sigma}^2/\sigma^2 \sim \chi^2(n-2)$

⑦ $Y_i \sim N(\beta_1 + \beta_2 X_i; \sigma^2)$

CÁC ĐỊNH LÝ XÁC SUẤT

Định lý 1:Nếu Z_1 ,Z_2 ,...Z_n là các đại lượng ngẫu nhiên tuân theo phân bố chuẩn thì $Z=\sum k_i Z_i$ với k_i là hằng số, thì Z cũng tuân theo phân bố chuẩn.

* Định lý 2: Nếu Z_1 ,Z_2 ,...Z_n là các đại lượng ngẫu nhiên tuân theo phân bố chuẩn N(O,1)thì $Z=\sum Z^2_i$ tuân theo **χ2 (n-2)** phân bố Chi bình phươngvới bậc tự do n.

- * Định lý 3:Nếu ĐLNN Z_1 ~ N(O,1), Z_2 ,~ **χ2 (k)** và Z_1, Z_2 độc lập thì T=Z/ $\pm\sqrt{R^2}$ $\sum k_i Z_i$ với k_i là hằng số, thì Z cũng tuân theo

6- Khoảng tin cậy của β_1; β_2; σ^2

Với độ tin cậy 1- α, KTC của β_2 là:

$$\boxed{\hat{\beta}_2 \pm t_{\alpha/2}.se(\hat{\beta}_2)}$$

Khoảng tin cậy của β_1 là:

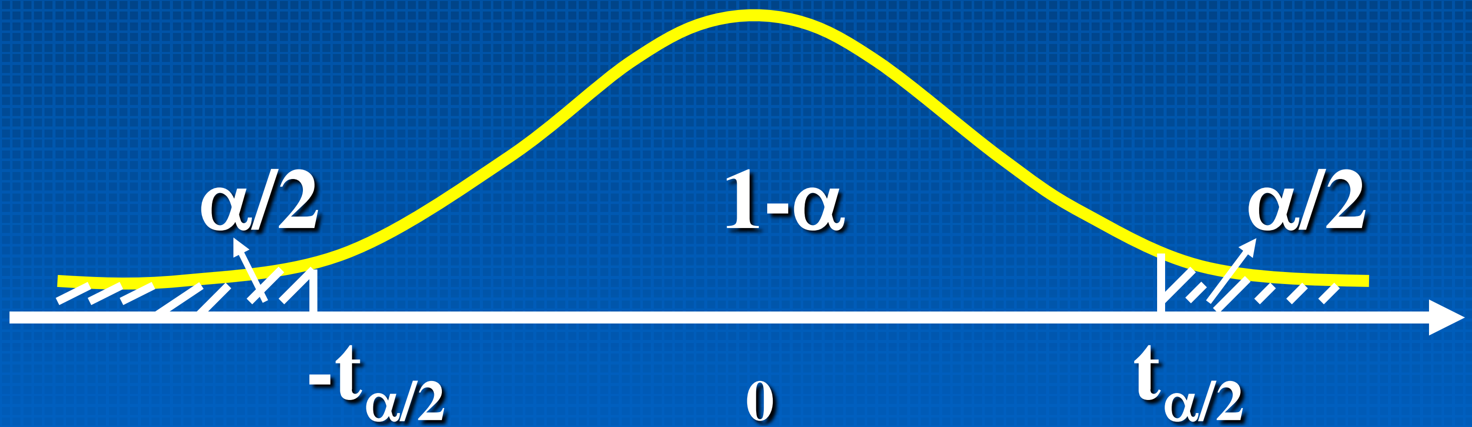
$$\hat{\beta}_1 \pm t_{\alpha/2}\, se(\hat{\beta}_1)$$

Khoảng tin cậy của σ^2 là:

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

Trong đó $t_{\alpha/2}$ là giá trị của ĐLNN T: T ~ T(n-2) thỏa ĐK:

$$P(|T| > t_{\alpha/2}) = \alpha$$



Để xác định $t_{\alpha/2}$ ta có thể tra bảng hoặc dùng hàm TINV trong Excel

7- KIỂM ĐỊNH GIẢ THIẾT VỀ CÁC HỆ SỐ HỒI QUI

7.1 Kiểm định giả thiết: phương pháp khoảng tin cậy

Kiểm định giả thiết:

$$H_0: \beta_2 = \beta^*; H_1: \beta_2 \neq \beta^*$$

Qui tắc quyết định:

Thiết lập khoảng tin cậy với độ tin cậy 1-α cho β_2.

Nếu β* thuộc khoảng tin cậy này thì chấp nhận H_0.

Nếu β* nằm ngoài khoảng này thì bác bỏ H_0.

Thí dụ: $H_0: \beta_2 = 0,3; H_1: \beta_2 \neq 0,3$

KTC của β_2 với độ tin cậy 95% là:

$$(0,4268 < \beta_2 < 0,5914)$$

$$\beta^* = 0,3 \notin (0,4268; 0,5914)$$

nên ta bác bỏ gt H_0

7.2 Kiểm định giả thiết: phương pháp mức ý nghĩa

Kiểm định giả thiết:

$$H_0: \beta_2 = \beta^*; \quad H_1: \beta_2 \neq \beta^*$$

Qui tắc quyết định:

- **Tính:** $t = (\hat{\beta}_2 - \beta^*)/se(\hat{\beta}_2)$

- **Với mức ý nghĩa α, tra bảng (hoặc dùng hàm TINV) để tìm $t_{\alpha/2}$**
- **Nếu $|t| > t_{\alpha/2}$ thì bác bỏ giả thiết H_0**
- **Nếu $|t| \leq t_{\alpha/2}$ thì chấp nhận giả thiết H_0**

* Chú ý:

❶ **Kiểm định giả thiết:**

$$H_0: \beta_2 = \beta^*$$

với giả thiết đối: $H_1: \beta_2 \neq \beta^*$

gọi là kiểm định giả thiết hai phía (miền bác bỏ nằm về hai phía của miền chấp nhận)

❷ Kiểm định giả thiết

$$H_0: \beta_2 = \beta^*$$

với giả thiết đối $H_1: \beta_2 > \beta^*$

(hoặc $H_1: \beta_2 < \beta^*$) gọi là kiểm định giả thiết một phía (miền bác bỏ nằm về một phía của miền chấp nhận)

❸ **Nếu dùng các phần mềm Kinh tế lượng thì giá trị:**

$$p = P(|T| > |t|)$$

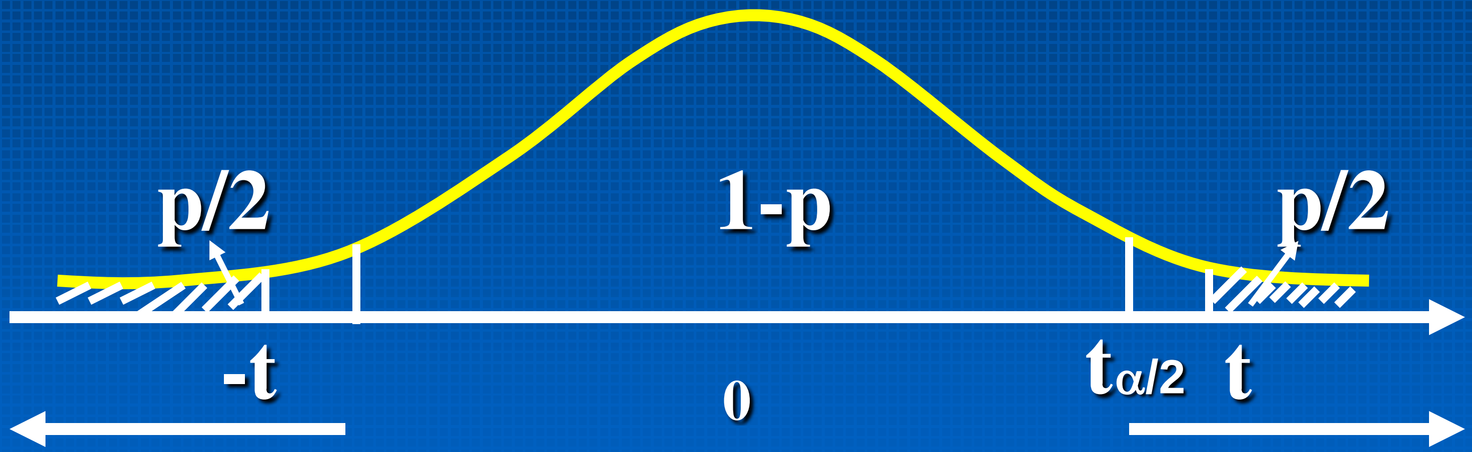
Trong đó:

$$t = \frac{\hat{\beta}_2}{se(\hat{\beta}_2)}$$

được tính sẵn và ghi ở bảng kết quả (bảng output).

Trong đó t là giá trị của ĐLNN T:
T ~ T(n-2) thỏa ĐK:

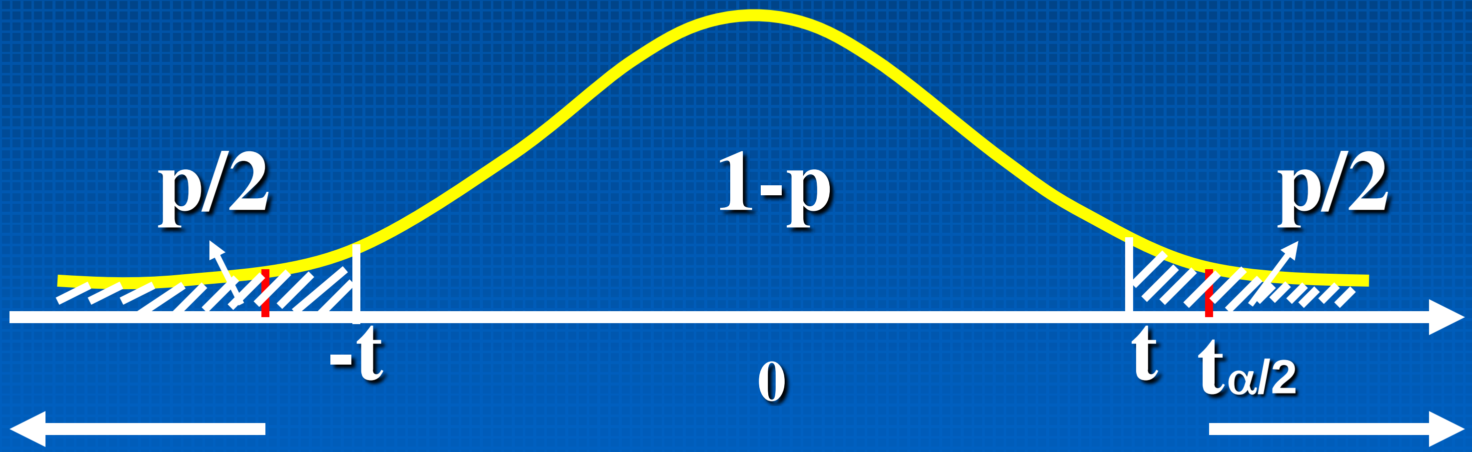
$$P(|T| > |t|) = p$$



Nếu $p < \alpha \Rightarrow (1-p) > (1-\alpha)$.
Khi đó t ở phía bên phải của $t_{\alpha/2}$.

Nếu $p > \alpha \Rightarrow (1-p) < (1-\alpha)$.

Khi đó t ở phía bên trái của $t_{\alpha/2}$.



Khi đó để kiểm định giả thiết:

$$H_0: \beta_2 = 0; H_1: \beta_2 \neq 0$$

ta áp dụng *qui tắc kiểm định bằng p – value):*

- Nếu $p < \alpha$ thì bác bỏ giả thiết H_0
- Nếu $p \geq \alpha$ thì có thể chấp nhận giả thiết H_0

(α là mức ý nghĩa)

8- KIỂM ĐỊNH SỰ PHÙ HỢP CỦA HÀM HỒI QUI

* $H_0: R^2 = 0$; $H_1: R^2 \neq 0$

$$F = R^2(n-2)/(1-R^2)$$

Với mức ý nghĩa α, tra bảng (hoặc dùng hàm FINV) để tìm $F_\alpha(1; n-2)$.

* Nếu $F > F_\alpha(1; n-2)$ thì bác bỏ H_0. Tức hàm hồi qui phù hợp.

* Nếu $F \le F_\alpha(1; n-2)$ thì có thể chấp nhận H_0. Hàm hồi qui không phù hợp.

9- DỰ BÁO BẰNG MÔ HÌNH HỒI QUI

❶ **Dự báo giá trung bình của Y khi $X = X_0$**

Giả sử $X = X_0$, cần dự báo

$$E(Y/X_0) = \beta_1 + \beta_2 X_0$$

Dự báo điểm của E(Y/X$_0$) là:

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

Dự báo khoảng của $E(Y/X_0)$ với độ tin cậy $1-\alpha$ là:

$$\boxed{\hat{Y}_0 \pm t_{\alpha/2}.se(\hat{Y}_0)}$$

Trong đó:

$$se(\hat{Y}_0) = \sqrt{var(\hat{Y}_0)}$$

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

❷ Dự báo g.trị cá biệt của Y

Giả sử $X = X_0$, cần dự báo:

$$Y_0 = \beta_1 + \beta_2 X_0 + U_i$$

Dự báo khoảng của Y_0 với độ tin cậy $1-\alpha$ là:

$$\boxed{\hat{Y}_0 \pm t_{\alpha/2}.se(Y_0 - \hat{Y}_0)}$$

Trong đó:

$$se(Y_0 - \hat{Y}_0) = \sqrt{var(Y_0 - \hat{Y}_0)}$$

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

10- TRÌNH BÀY KẾT QUẢ PHÂN TÍCH HỒI QUI

$$\hat{Y}_i = 11,5 \quad - \quad 1,375\, X_i$$

$$se = (0,3609) \qquad (0,0806)$$

$$t = (31,8697) \qquad (-17,0579)$$

$$p = (0,0000) \qquad (0,0001)$$

$$R^2 = 0,9864 \qquad F = 292,12$$

* ***Chú ý***:

- Các giá trị t được tính theo công thức:

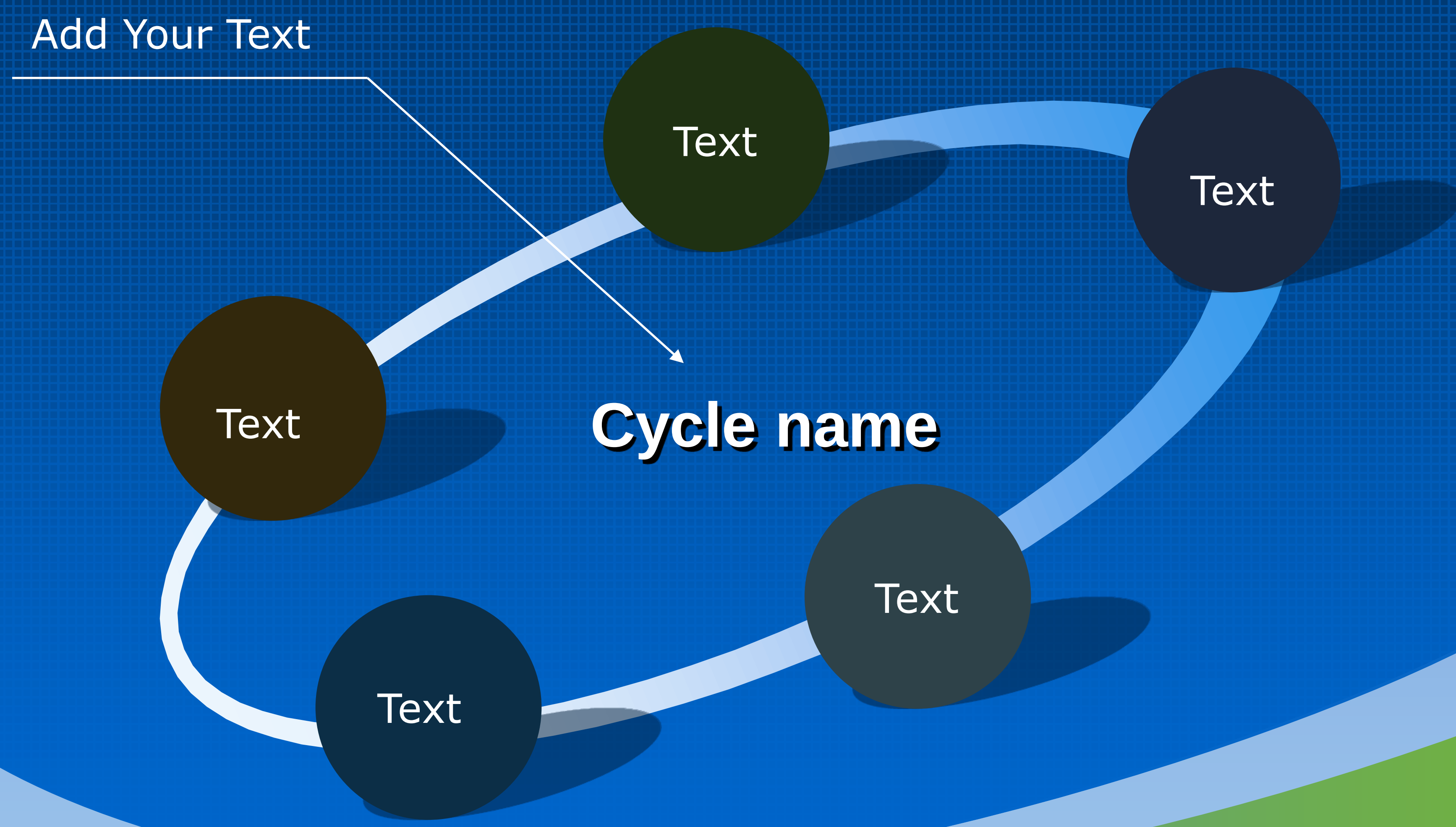
$t_1 = \hat{\beta}_1/se(\hat{\beta}_1)$; $t_2 = \hat{\beta}_2/se(\hat{\beta}_2)$

- Giá trị p:

$P(|T| > 17{,}0579) = 0{,}0001$

Hết chương 2

Cycle Diagram



3-D Pie Chart