



Bài giảng 9

HỒI QUI TUYỂN TÍNH ĐƠN BIẾN

HỒI QUI TUYẾN TÍNH ĐƠN BIẾN

- Mô hồi qui tuyến tính đơn
- Phương pháp bình phương tối thiểu
- Hệ số xác định
- Các giả định của mô hình
- Kiểm định mức ý nghĩa
- Sử dụng mô hình hồi qui ước lượng để ước lượng và dự đoán
- Lời giải trên máy tính
- Phân tích phần dư: Hiệu lực của các giả định của mô hình

MÔ HÌNH HỒI QUI TUYẾN TÍNH ĐƠN BIẾN

- Phương trình mô tả y liên hệ với x như thế nào và một số hạng sai số được gọi là mô hình hồi qui.
- Mô hình hồi qui tuyến tính đơn biến là:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

Với:

β_0 và β_1 được gọi là các tham số của mô hình,
 ε là biến ngẫu nhiên được gọi là số hạng sai số

PHƯƠNG TRÌNH HỒI QUI TUYẾN TÍNH ĐƠN BIẾN

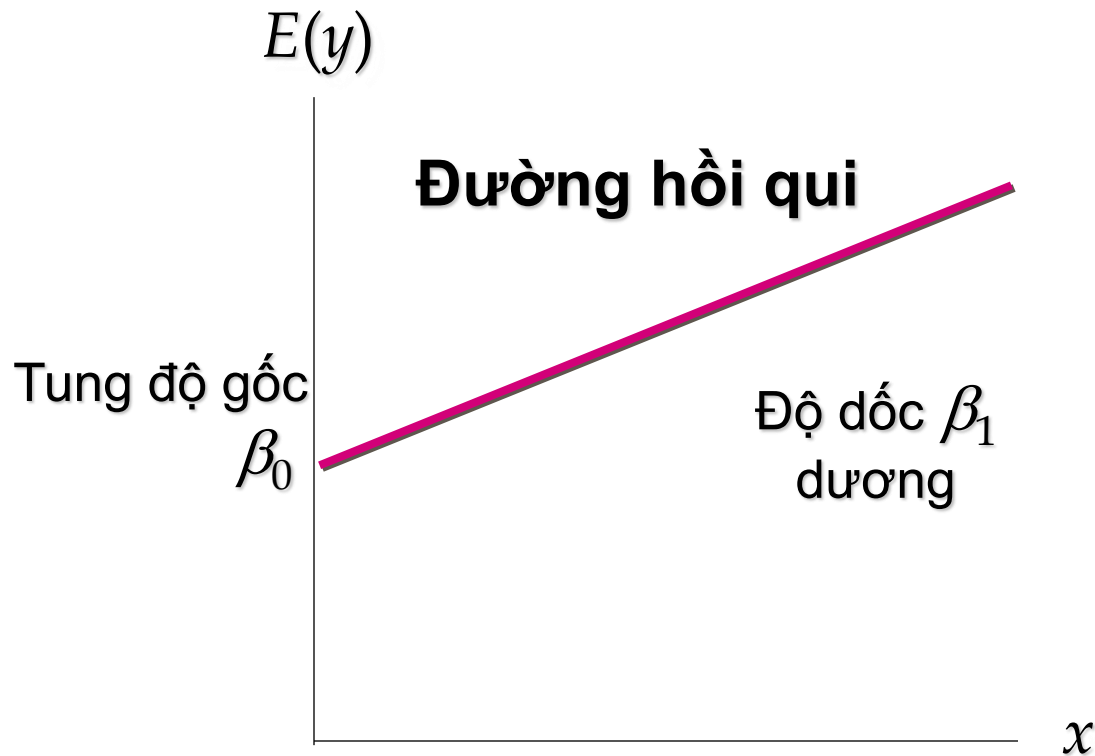
- Phương trình hồi qui tuyến tính đơn là:

$$E(y) = \beta_0 + \beta_1 x$$

- Đồ thị của phương trình hồi qui là đường thẳng.
- β_0 là tung độ gốc của đường hồi qui
- β_1 là độ dốc của đường hồi qui
- $E(y)$ là giá trị kỳ vọng của y đối với giá trị x cho trước.

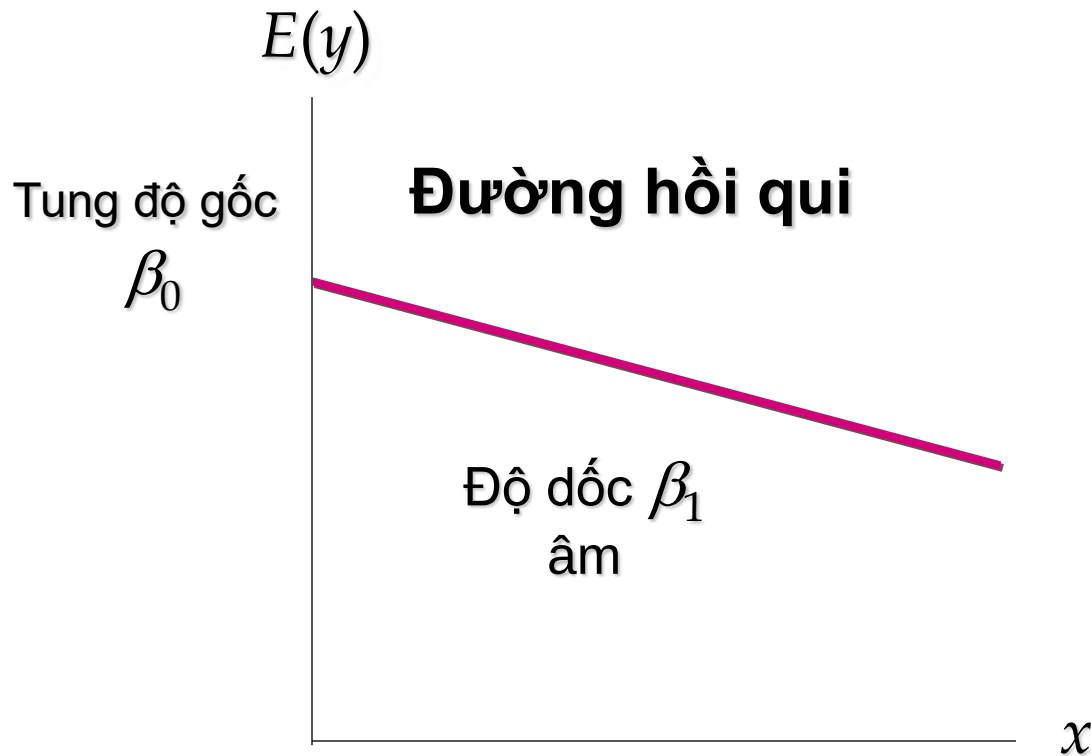
PHƯƠNG TRÌNH HỒI QUI TUYẾN TÍNH ĐƠN BIẾN

- Quan hệ tuyến tính đồng biến



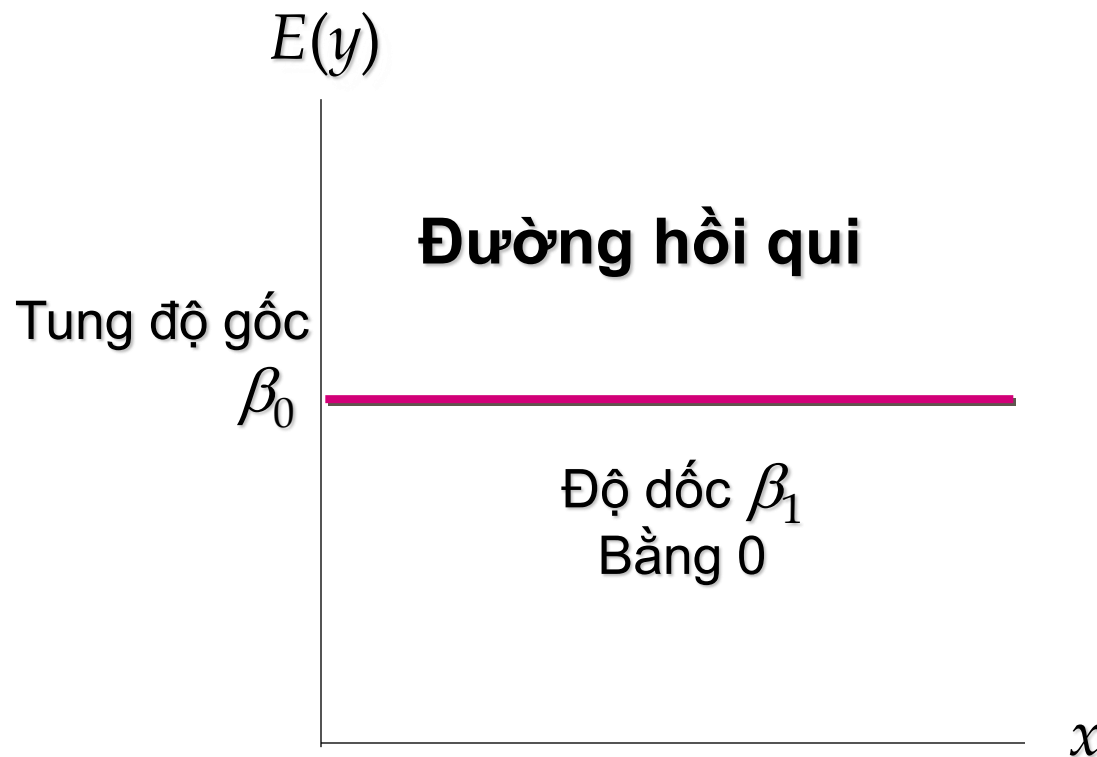
PHƯƠNG TRÌNH HỒI QUI TUYẾN TÍNH ĐƠN BIẾN

- Quan hệ tuyến tính nghịch biến



PHƯƠNG TRÌNH HỒI QUI TUYẾN TÍNH ĐƠN BIẾN

- Không quan hệ



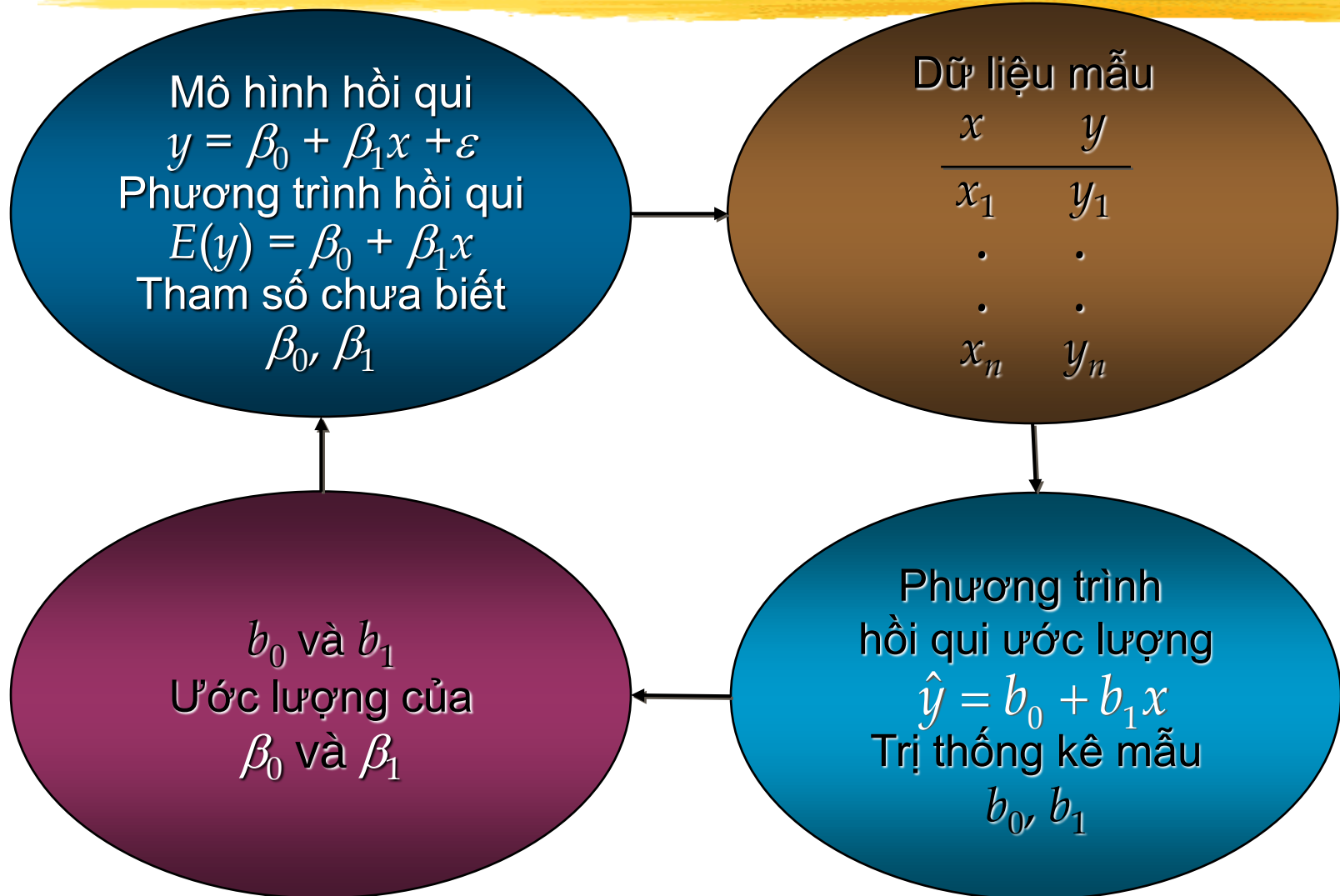
PHƯƠNG TRÌNH HỒI QUI TUYẾN TÍNH ĐƠN BIẾN ƯỚC LƯỢNG

- Phương trình hồi qui tuyến tính đơn ước lượng

$$\hat{y} = b_0 + b_1x$$

- Đồ thị được gọi là đường hồi qui ước lượng.
- b_0 là tung độ gốc của đường.
- b_1 là độ dốc của đường
- $\hat{y}$ là giá trị ước lượng của y đối với giá trị x cho trước.

QUI TRÌNH ƯỚC LƯỢNG



PHƯƠNG PHÁP BÌNH PHƯƠNG TỐI THIỂU

- Tiêu chí bình phương tối thiểu

$$\min \sum (y_i - \hat{y}_i)^2$$

Với:

y_i = giá trị quan sát của biến phụ thuộc
đối với quan sát thứ i

$y_i^{\wedge}$ = giá trị ước lượng của biến phụ thuộc
đối với quan sát thứ i

PHƯƠNG PHÁP BÌNH PHƯƠNG TỐI THIỂU

- Độ dốc của phương trình hồi qui ước lượng

$$b_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

PHƯƠNG PHÁP BÌNH PHƯƠNG TỐI THIỂU

- Tung độ gốc của phương trình hồi qui ước lượng

$$b_0 = \bar{y} - b_1 \bar{x}$$

Với:

x_i = giá trị của biến độc lập đối với quan sát thứ i

y_i = giá trị của biến phụ thuộc đối với quan sát thứ i

$\bar{x}$ = giá trị trung bình của biến độc lập

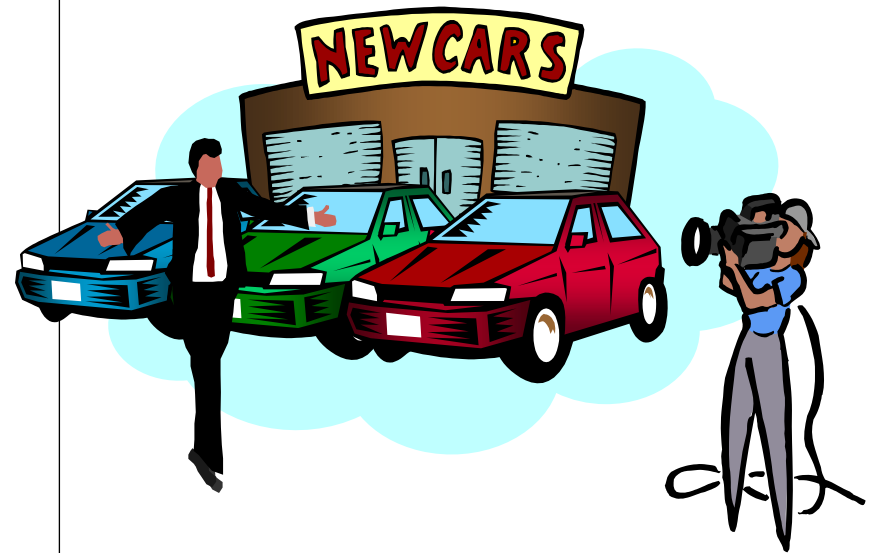
$\bar{y}$ = giá trị trung bình của biến phụ thuộc

n = tổng số quan sát

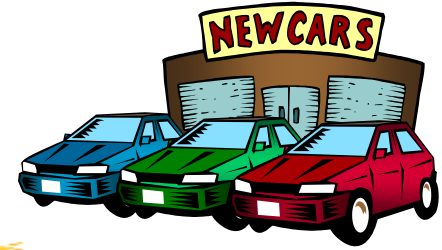
HỒI QUI TUYỂN TÍNH ĐƠN BIẾN

□ Ví dụ: Doanh số xe hơi

Quảng cáo TV	Doanh số xe hơi
1	14
3	24
2	18
1	17
3	27



PHƯƠNG TRÌNH HỒI QUI ƯỚC LƯỢNG



- Độ dốc của phương trình hồi qui ước lượng

$$b_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{20}{4} = 5$$

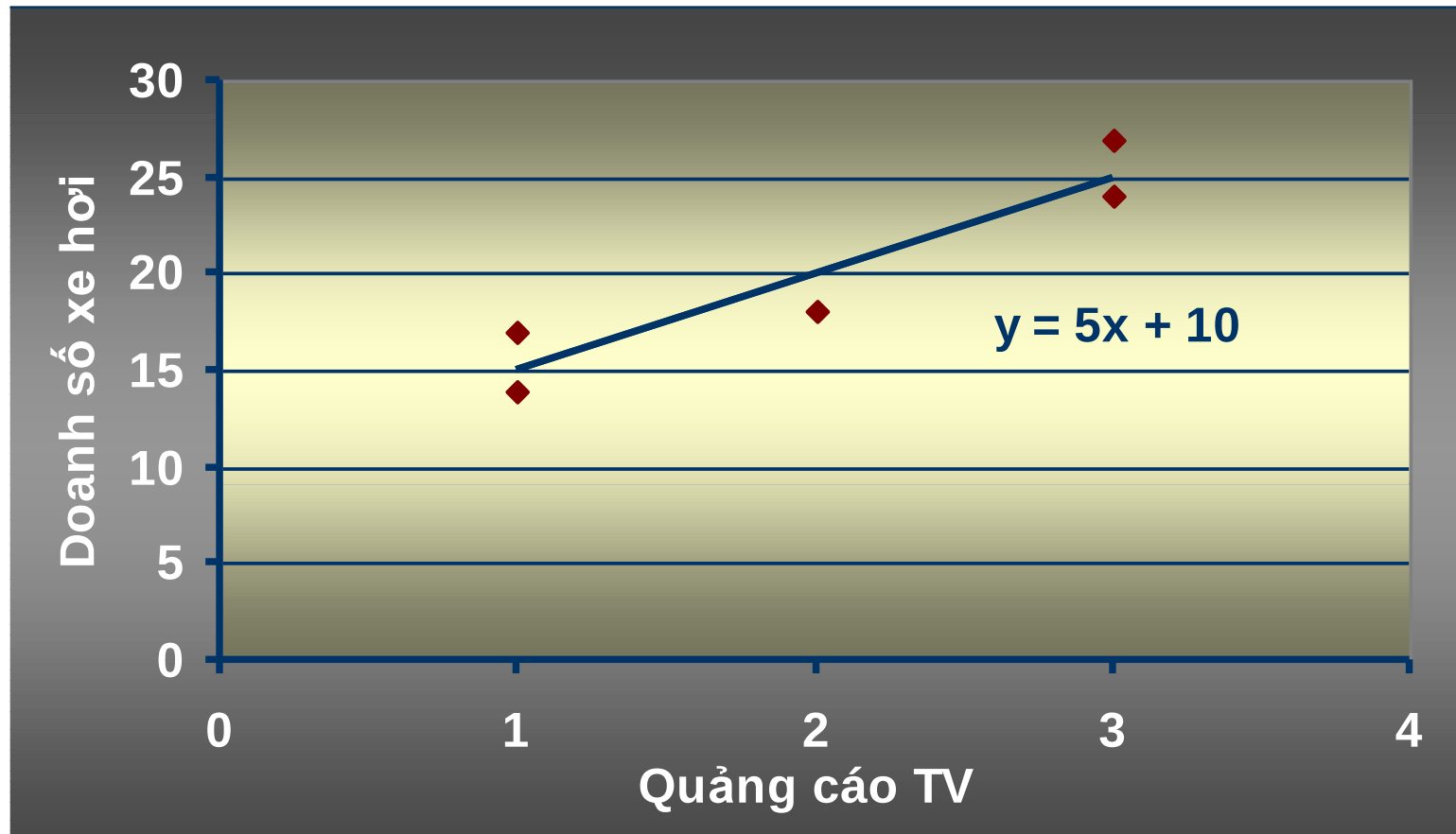
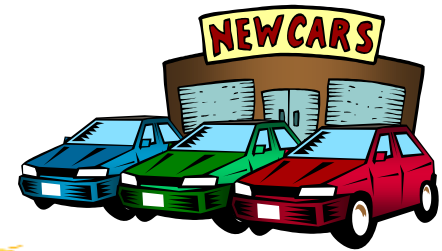
- Tung độ gốc của phương trình hồi qui ước lượng

$$b_0 = \bar{y} - b_1 \bar{x} = 20 - 5(2) = 10$$

- Phương trình hồi qui ước lượng

$$\hat{y} = 10 + 5x$$

ĐỒ THỊ PHÂN TÁN ĐIỂM VÀ ĐƯỜNG XU HƯỚNG



HỆ SỐ XÁC ĐỊNH

- Mỗi liên hệ giữa SST, SSR, SSE

$$\begin{array}{ccccc} \text{SST} & = & \text{SSR} & + & \text{SSE} \\ \swarrow & & \downarrow & & \searrow \\ \sum (y_i - \bar{y})^2 & = & \sum (\hat{y}_i - \bar{y})^2 & + & \sum (y_i - \hat{y}_i)^2 \end{array}$$

Với:

SST = Tổng bình phương toàn phần

SSR = Tổng bình phương hồi qui

SSE = Tổng bình phương sai số

HỆ SỐ XÁC ĐỊNH



- Hệ số xác định là:

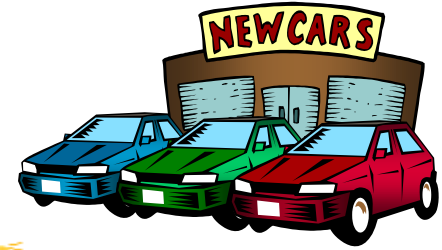
$$r^2 = SSR/SST$$

Với:

SSR = Tổng bình phương hồi qui

SST = Tổng bình phương toàn phần

HỆ SỐ XÁC ĐỊNH



$$r^2 = SSR/SST = 100/114 = .8772$$

Mối quan hệ hồi qui rất mạnh; 88% sự biến thiên của doanh số xe hơi có thể được giải thích bởi mối quan hệ tuyến tính giữa số quảng cáo trên TV và doanh số xe hơi.

HỆ SỐ TƯƠNG QUAN MẪU



$$r_{xy} = (\text{sign of } b_1) \sqrt{\text{Coefficient of Determination}}$$

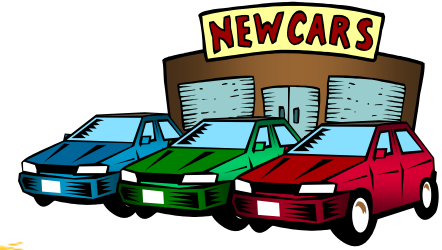
$$r_{xy} = (\text{sign of } b_1) \sqrt{r^2}$$

Với

b_1 = độ dốc của phương trình hồi qui ước lượng

$$\hat{y} = b_0 + b_1 x$$

HỆ SỐ TƯƠNG QUAN MẪU



$$r_{xy} = (\text{sign of } b_1) \sqrt{r^2}$$

Dấu của b_1 trong phương trình $\hat{y} = 10 + 5x$ là “+”.

$$r_{xy} = +\sqrt{.8772}$$

$$r_{xy} = +.9366$$

CÁC GIẢ ĐỊNH VỀ SỐ HẠNG SAI SỐ ε



1. Sai số ε là biến ngẫu nhiên với trung bình bằng 0
2. Phương sai của ε , ký hiệu σ^2 , sẽ giống nhau đối với tất cả các giá trị của biến độc lập.
3. Các giá trị của ε là độc lập.
4. Sai số ε là biến ngẫu nhiên tuân theo phân phối chuẩn

KIỂM ĐỊNH Ý NGHĨA

Để kiểm định ý nghĩa của mối quan hệ hồi qui, chúng Ta phải tiến hành kiểm định giả thuyết xác định xem giá trị của β_1 có bằng 0 hay không.

Hai kiểm định được dùng phổ biến là:

t Test và F Test

Cả kiểm định t và F đều yêu cầu một ước lượng của σ^2 , phương sai của ε trong mô hình hồi qui

KIỂM ĐỊNH Ý NGHĨA

- Ước lượng của σ

Trung bình của sai số bình phương (MSE) sẽ cung cấp Ước lượng của σ^2 , và s^2 cũng được sử dụng.

$$s^2 = \text{MSE} = \text{SSE} / (n - 2)$$

Với

$$\text{SSE} = \sum (y_i - \hat{y}_i)^2 = \sum (y_i - b_0 - b_1 x_i)^2$$

KIỂM ĐỊNH Ý NGHĨA

- Ước lượng của σ
 - Để ước lượng σ chúng ta lấy căn bậc hai của σ^2 .
 - Kết quả s được gọi là sai số chuẩn của ước lượng

$$s = \sqrt{\text{MSE}} = \sqrt{\frac{\text{SSE}}{n-2}}$$

KIỂM ĐỊNH Ý NGHĨA: KIỂM ĐỊNH t

- Giả thuyết

$$H_0: \beta_1 = 0$$

$$H_a: \beta_1 \neq 0$$

- Trị thống kê

$$t = \frac{b_1}{s_{b_1}}$$

KIỂM ĐỊNH Ý NGHĨA: KIỂM ĐỊNH t

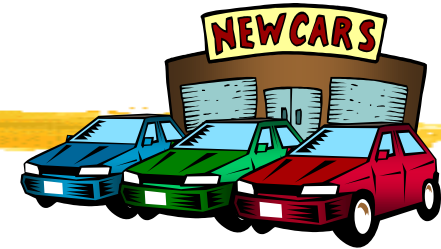
- Quy tắc bác bỏ

Bác bỏ H_0 nếu $p\text{-value} \leq \alpha$
hay $t \leq -t_{\alpha/2}$ hay $t \geq t_{\alpha/2}$

Với

$t_{\alpha/2}$ được dựa trên phân phối t
với bậc tự do là $n - 2$

KIỂM ĐỊNH Ý NGHĨA: KIỂM ĐỊNH t



1. Xác định giả thuyết.

$$H_0: \beta_1 = 0$$

$$H_a: \beta_1 \neq 0$$

2. Xác định mức ý nghĩa.

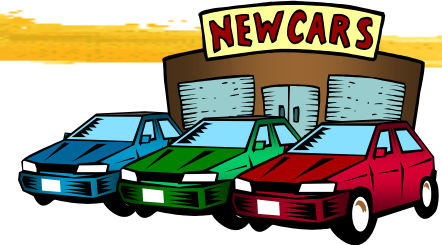
$$\alpha = .05$$

3. Lựa chọn trị thống kê kiểm định. $t = \frac{b_1}{s_{b_1}}$

4. Quy tắc bác bỏ.

Bác bỏ H_0 nếu $p\text{-value} \leq .05$
hay $|t| > 3.182$ (với độ tự do là 3)

KIỂM ĐỊNH Ý NGHĨA: KIỂM ĐỊNH t



5. Tính giá trị của trị thống kê kiểm định.

$$t = \frac{b_1}{s_{b_1}} = \frac{5}{1.08} = 4.63$$

6. Xác định có bác bỏ H_0 hay không

$t = 4.541$ cho một diện tích .01 ở phần đuôi phía trên. Vì vậy, p -value nhỏ hơn .02. (Cũng vậy, $t = 4.63 > 3.182$.) chúng ta có thể bác bỏ H_0 .

KHOẢNG TIN CẬY CỦA β_1

- Chúng ta có thể dùng một khoảng tin cậy 95% của β_1 để kiểm định các giả thuyết vừa mới dùng trong kiểm định t.
- H_0 bị bác bỏ nếu giá trị được giả thuyết của β_1 không nằm trong khoảng tin cậy của β_1 .

KHOẢNG TIN CẬY CỦA β_1

- Công thức của khoảng tin cậy đối với β_1 là:

The diagram shows the confidence interval formula $b_1 \pm t_{\alpha/2} s_{b_1}$ in the center. Two callout boxes point to parts of the formula: one on the left points to b_1 and contains the text " b_1 là ước lượng điểm", and one on the right points to $t_{\alpha/2} s_{b_1}$ and contains the text " $t_{\alpha/2} s_{b_1}$ là biên của sai số".

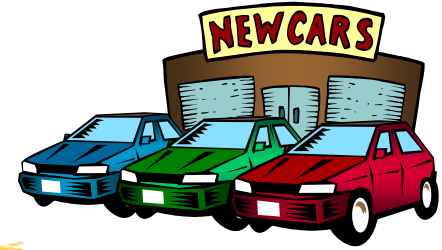
$$b_1 \pm t_{\alpha/2} s_{b_1}$$

b_1 là ước lượng điểm

$t_{\alpha/2} s_{b_1}$ là biên của sai số

Với $t_{\alpha/2}$ là giá trị t cho một diện tích là $\alpha/2$ trong phần đuôi phía trên của phân phối t
Với độ tự do là $n - 2$

KHOẢNG TIN CẬY CỦA β_1



- Quy tắc bác bỏ

Bác bỏ H_0 nếu 0 không nằm trong khoảng tin cậy của β_1 .

- Khoảng tin cậy 95% của β_1

$$b_1 \pm t_{\alpha/2} s_{b_1} = 5 + / - 3.182(1.08) = 5 + / - 3.44$$

or 1.56 to 8.44

- Kết luận

0 không nằm trong khoảng tin cậy .

Bác bỏ H_0

KIỂM ĐỊNH Ý NGHĨA: KIỂM ĐỊNH F

- Giả thuyết

$$H_0: \beta_1 = 0$$

$$H_a: \beta_1 \neq 0$$

- Trị thống kê

$$F = \text{MSR} / \text{MSE}$$

KIỂM ĐỊNH Ý NGHĨA: KIỂM ĐỊNH F

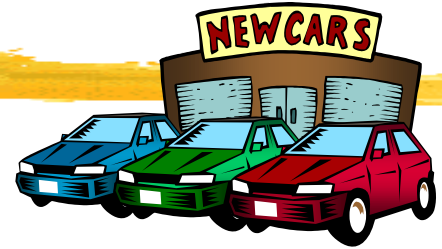
- Quy tắc bác bỏ

Bác bỏ H_0 nếu
 $p\text{-value} \leq \alpha$
hay $F \geq F_\alpha$

Với:

F_α dựa trên phân phối F với độ tự do là 1 ở tử số và độ tự do là $n - 2$ ở mẫu số

KIỂM ĐỊNH Ý NGHĨA: KIỂM ĐỊNH F



1. Xác định giả thuyết.

$$H_0: \beta_1 = 0$$

$$H_a: \beta_1 \neq 0$$

2. Xác định mức ý nghĩa.

$$\alpha = .05$$

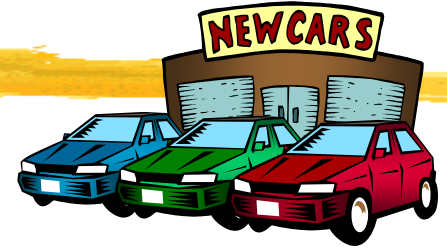
3. Lựa chọn trị thống kê kiểm định.

$$F = \text{MSR} / \text{MSE}$$

4. Quy tắc bác bỏ.

Bác bỏ H_0 nếu $p\text{-value} \leq .05$
hay $F \geq 10.13$ (với d.f ở tử số
là 1 và d.f ở mẫu số là 3)

KIỂM ĐỊNH Ý NGHĨA: KIỂM ĐỊNH F



5. Tính giá trị của trị thống kê kiểm định.

$$F = \text{MSR} / \text{MSE} = 100 / 4.667 = 21.43$$

6. Xác định có bác bỏ H_0 hay không

$F = 17.44$ cho một diện tích là $.025$ ở phần đuôi phía trên. Vì vậy, p -value tương ứng với $F = 21.43$ sẽ nhỏ hơn $2(.025) = .05$. Vì vậy, chúng ta bác bỏ H_0 .

Chứng cứ thống kê đủ để kết luận có mối quan hệ có ý nghĩa giữa số lượng quảng cáo TV và doanh số xe hơi

MỘT VÀI LƯU Ý VỀ GIẢI THÍCH CÁC KIỂM ĐỊNH Ý NGHĨA



- Bác bỏ $H_0: \beta_1 = 0$ và kết luận rằng mối quan hệ giữa x và y có ý nghĩa, chúng ta không thể kết luận có mối quan hệ nhân quả giữa x và y

SỬ DỤNG PHƯƠNG TRÌNH HỒI QUI ƯỚC LƯỢNG ĐỂ ƯỚC LƯỢNG VÀ DỰ ĐOÁN

- Khoảng tin cậy của $E(y_p)$

$$\hat{y}_p \pm t_{\alpha/2} s_{\hat{y}_p}$$

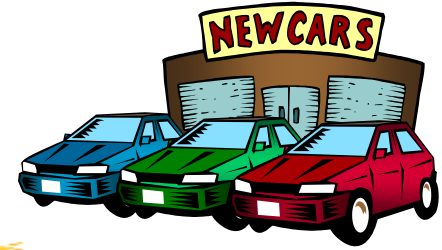
- Dự đoán khoảng tin cậy của y_p

$$y_p \pm t_{\alpha/2} s_{\text{ind}}$$

Với:

hệ số tin cậy là $1 - \alpha$ và
 $t_{\alpha/2}$ dựa trên phân phối t
với độ tự do là $n - 2$

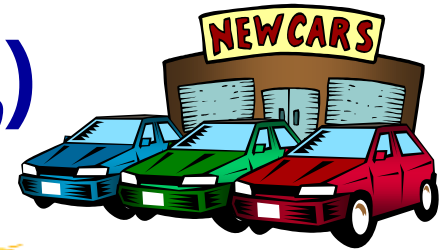
ƯỚC LƯỢNG ĐIỂM



Nếu 3 quảng cáo TV được tiến hành trước khi bán, chúng ta kỳ vọng trung bình doanh số xe hơi là:

$$\hat{y} = 10 + 5(3) = 25 \text{ xe hơi}$$

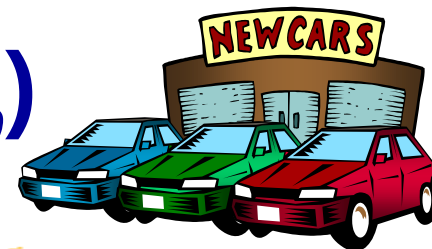
KHOẢNG TIN CẬY CỦA $E(y_p)$



- Kết quả khoảng tin cậy từ Excel

	D	E	F	G
1		CONFIDENCE INTERVAL		
2		x_p	3	
3		$\bar{x}$	2.0	
4		$x_p - \bar{x}$	1.0	
5		$(x_p - \bar{x})^2$	1.0	
6		$\Sigma (x_p - \bar{x})^2$	4.0	
7		Variance of $\hat{y}$	2.1000	
8		Std. Dev of $\hat{y}$	1.4491	
9		t Value	3.1824	
10		Margin of Error	4.6118	
11		Point Estimate	25.0	
12		Lower Limit	20.39	
13		Upper Limit	29.61	

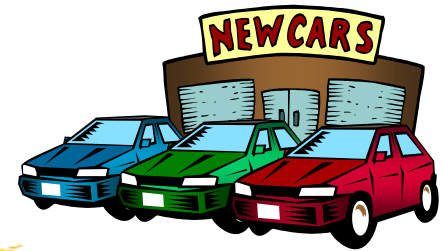
KHOẢNG TIN CẬY CỦA $E(y_p)$



Ước lượng khoảng tin cậy 95% của trung bình doanh số xe hơi khi 3 quảng cáo TV được thực hiện là:

$$25 \pm 4.61 = 20.39 \text{ to } 29.61 \text{ xe hơi}$$

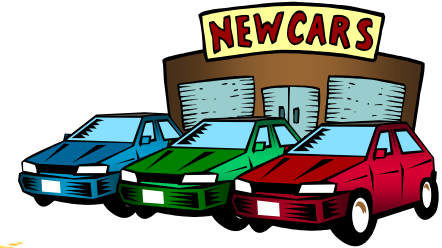
KHOẢNG DỰ ĐOÁN CỦA y_p



- Kết quả khoảng dự đoán từ Excel

	H	I
1	PREDICTION INTERVAL	
2	Variance of yind	6.76667
3	Std. Dev. of yind	2.60128
4	Margin of Error	8.27845
5	Lower Limit	16.72
6	Upper Limit	33.28
7		

KHOẢNG DỰ ĐOÁN CỦA y_p



Ước lượng khoảng dự đoán 95% của doanh số xe hơi trong 1 tuần cụ thể khi 3 quảng cáo TV được thực hiện là:

$$25 \pm 8.28 = 16.72 \text{ to } 33.28 \text{ xe hơi}$$

PHÂN TÍCH PHẦN DƯ

- Nếu các giả định về số hạng sai số ε không đảm bảo thì các kiểm định giả thuyết về ý nghĩa của mối quan hệ hồi qui và các kết quả ước lượng khoảng không còn hiệu lực
- Các phần dư sẽ cho thông tin tốt nhất về ε .
- Phần dư của quan sát thứ i

$$y_i - \hat{y}_i$$

- Rất nhiều phân tích phần dư dựa trên việc khảo sát đồ thị phần dư

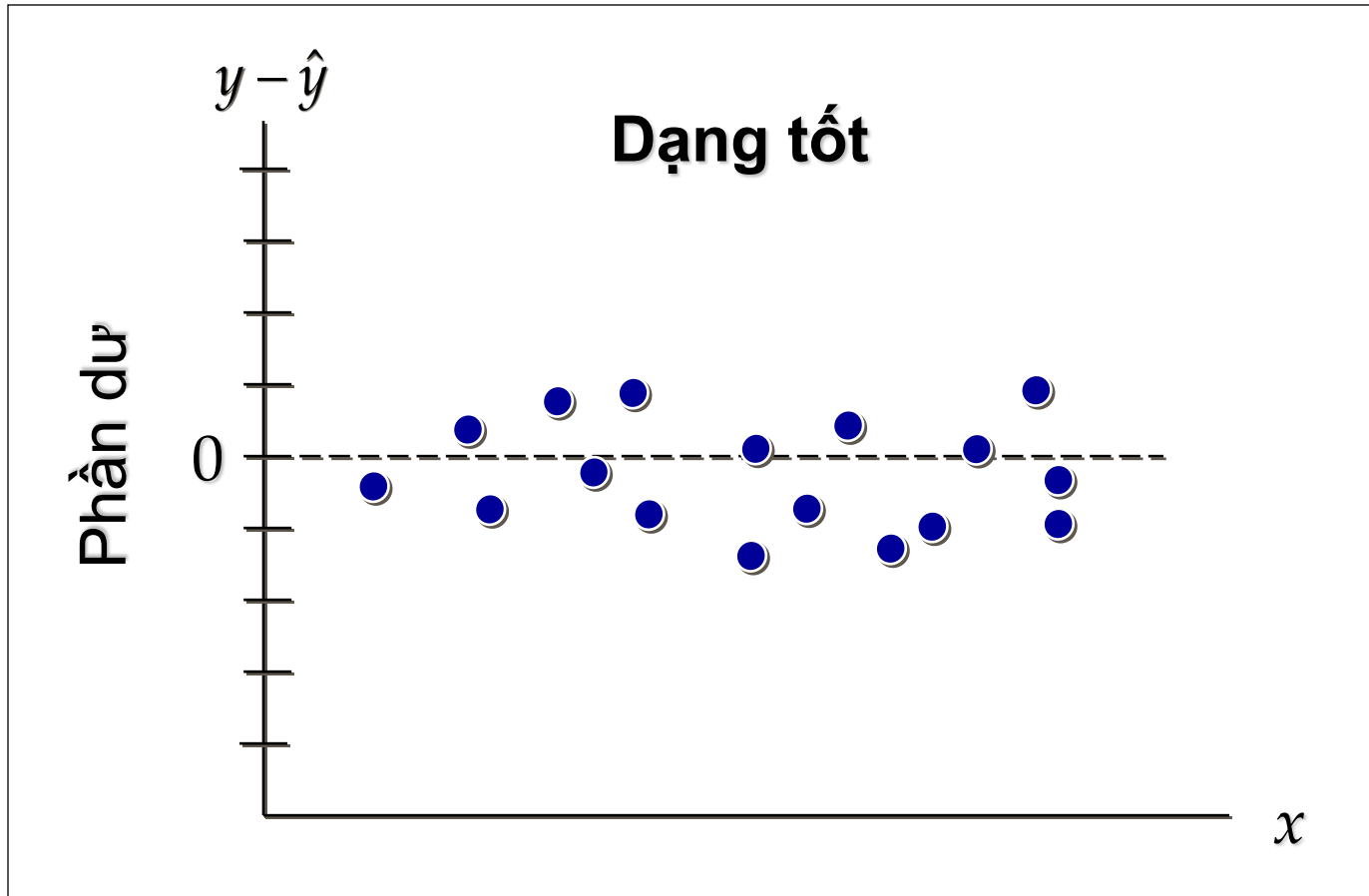
ĐỒ THỊ PHẦN DƯ THEO X



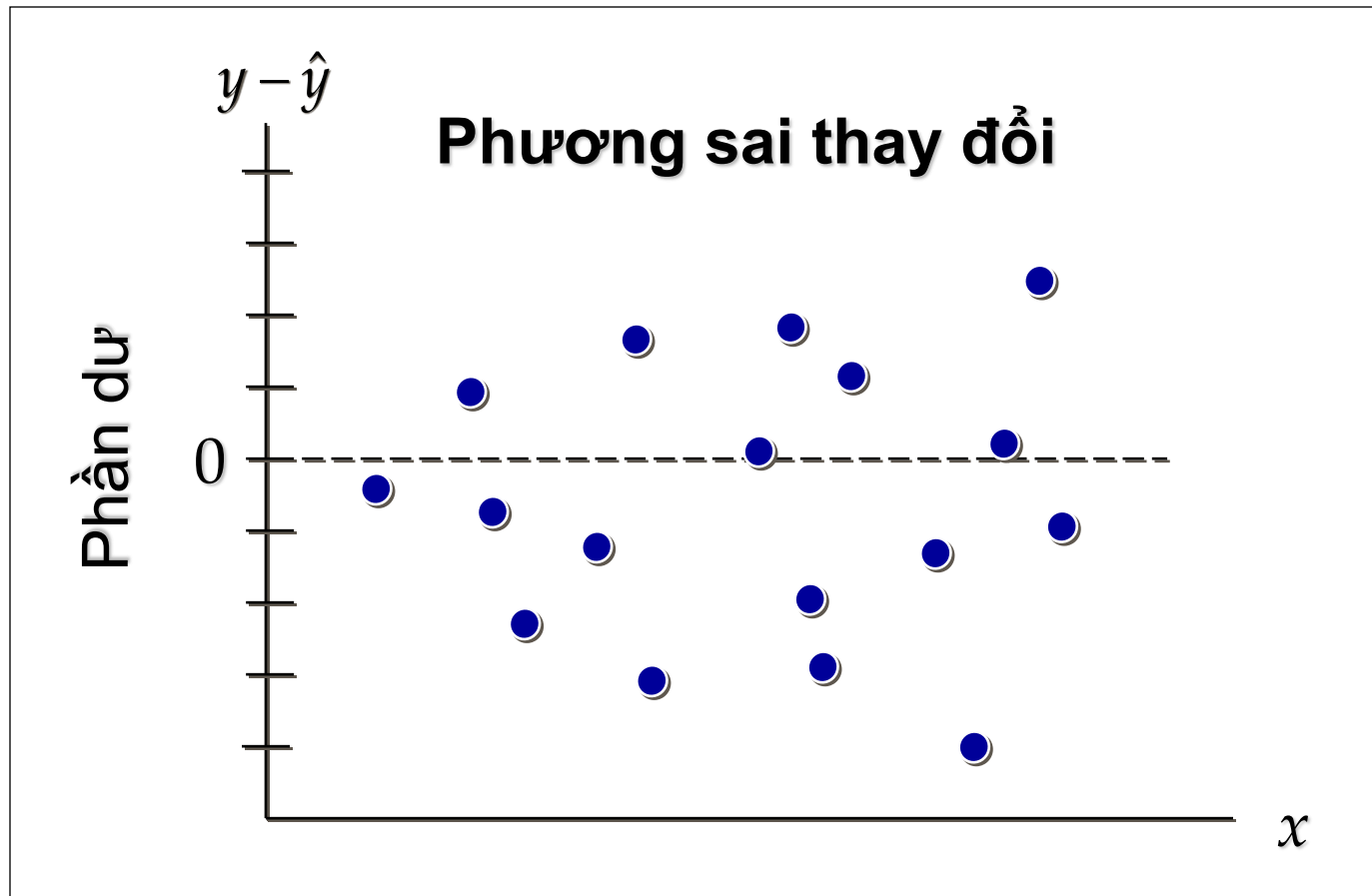
- Nếu giả định phương sai của ε giống nhau đối với tất cả các giá trị của x được thỏa, và mô hình hồi qui giả định là một biểu diễn đầy đủ của mối quan hệ giữa các biến, thì

Đồ thị phần dư sẽ cho một ấn tượng tổng thể về giải bằng các điểm nằm ngang

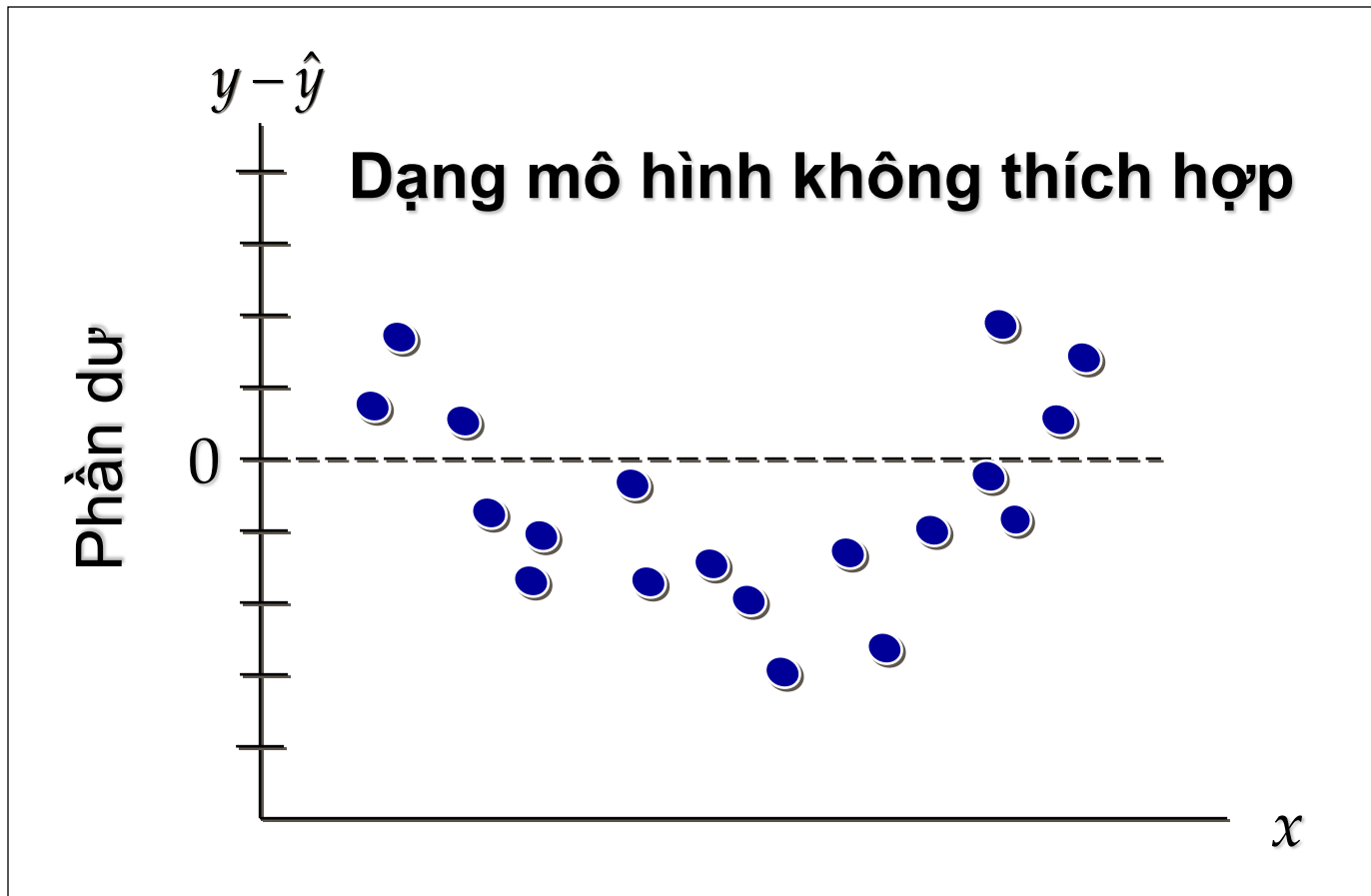
ĐỒ THỊ PHẦN DƯ THEO X



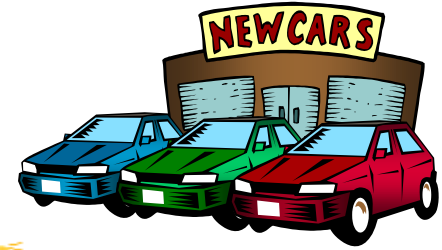
ĐỒ THỊ PHẦN DƯ THEO X



ĐỒ THỊ PHẦN DƯ THEO X



ĐỒ THỊ PHẦN DƯ THEO X



- Phần dư

Observation	Predicted Cars Sold	Residuals
1	15	-1
2	25	-1
3	20	-2
4	15	2
5	25	2

ĐỒ THỊ PHẦN DƯ THEO X

