

HỒI QUY TUYẾN TÍNH ĐƠN

GV : Đinh Công Khải – Chương trình Fulbright
Môn: Các Phương Pháp Định Lượng – MPP3

Kinh tế lượng là gì?

- “Kinh tế lượng được quan tâm với việc xác định các qui luật kinh tế bằng thực nghiệm” (Theil, 1971)
- “ Kinh tế lượng là việc phân tích định lượng các hiện tượng kinh tế thực tế dựa trên sự phát triển đồng thời của lý thuyết và quan sát, có liên quan bởi các phương pháp suy diễn thích hợp” (Samuelson et al., 1954)

Kinh tế lượng là gì?

- Ví dụ:
 - Quy luật cung cầu
 - Lạm phát càng cao thì tỷ lệ của thu nhập mà người dân muốn giữ dưới dạng tiền càng thấp
 - Mức cầu trung bình đối với hàng hóa của công ty sẽ tăng như thế nào theo mức tăng chi phí quảng cáo.
 - Sự phụ thuộc của sản lượng vụ mùa vào giống lúa, lượng mưa, phân bón

Phương pháp luận của kinh tế lượng

- Phát biểu một lý thuyết hoặc giả thuyết
- Xác định đặc trưng mô hình toán học của lý thuyết
- Xác định đặc trưng mô hình kinh tế lượng của lý thuyết
- Thu thập dữ liệu
- Ước lượng các tham số của mô hình kinh tế lượng
- Kiểm định giả thiết
- Dự báo hay tiên đoán
- Sử dụng mô hình để kiểm soát hoặc cho mục đích chính sách

Phương pháp luận của kinh tế lượng

- Ví dụ: Một cách trung bình, người ta có xu hướng tăng chi tiêu tiêu dùng khi thu nhập của họ tăng lên, nhưng không nhiều như gia tăng trong thu nhập của họ (Keynes)
- Mô hình toán học: $Y = \beta_1 + \beta_2 X$ (Y= tiêu dùng; X= thu nhập; $0 < \beta_2 < 1$)
- Mô hình KTL : $Y = \beta_1 + \beta_2 X + u$ (u là sai số ngẫu nhiên)
- Thu thập dữ liệu
- Ước lượng mô hình KTL: $\hat{Y} = -184,08 + 0,70 X$
- Kiểm định giả thiết
- Dự báo $\hat{Y}_{(X=6000)} = -184,08 + 0,70 * 6000 = \$4015,92$

Mô hình hồi qui tuyến tính

- **Hàm hồi qui tổng thể tuyến tính (PRF)**

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

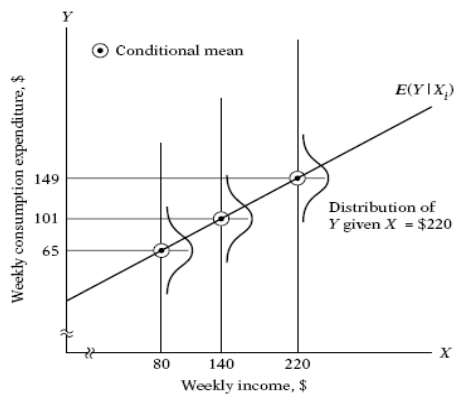
- $E(Y|X_i)$ là *trung bình* (tổng thể) của phân phối của Y với điều kiện X_i
- β_1, β_2 là các tham số của mô hình còn được gọi là *hệ số hồi qui*
- β_1 là *tung độ góc*; β_2 là *hệ số góc (độ dốc)* của đường hồi qui
- Phân tích hồi qui là nghiên cứu sự phụ thuộc của một biến, *biến phụ thuộc*, vào một hay nhiều biến khác, *biến độc lập (biến giải thích)*, với ý tưởng ước lượng giá trị trung bình (tổng thể) của biến phụ thuộc trên cơ sở các giá trị biết trước (trong mẫu lặp lại) của các biến giải thích.

Mô hình hồi qui tuyến tính

TABLE 2.1 WEEKLY FAMILY INCOME X, \$

Y ↓ \ X →	80	100	120	140	160	180	200	220	240	260
Weekly family consumption expenditure Y, \$	55	65	79	80	102	110	120	135	137	150
	60	70	84	93	107	115	136	137	145	152
	65	74	90	95	110	120	140	140	155	175
	70	80	94	103	116	130	144	152	165	178
	75	85	98	108	118	135	145	157	175	180
	-	88	-	113	125	140	-	160	189	185
	-	-	-	115	-	-	-	162	-	191
Total	325	462	445	707	678	750	685	1043	966	1211
Conditional means of Y, $E(Y X)$	65	77	89	101	113	125	137	149	161	173

Mô hình hồi qui tuyến tính



Mô hình hồi qui tuyến tính

- Độ lệch giữa mức chi tiêu tiêu dùng của một gia đình cá thể và mức chi tiêu trung bình là

$$u_i = Y_i - E(Y | X_i) \quad \text{hay}$$

$$Y_i = E(Y | X_i) + u_i \quad (u_i \text{ là sai số ngẫu nhiên})$$

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

- Đặc trưng ngẫu nhiên của PRF

$$E(Y | X_i) = E[E(Y | X_i)] + E(u_i | X_i)$$

$$\rightarrow E(u_i | X_i) = 0$$

Mô hình hồi qui tuyến tính

- Ý nghĩa của sai số ngẫu nhiên (u_i)

- Sự mơ hồ của lý thuyết
- Dữ liệu không có sẵn
- Các biến cốt lõi và những biến ngoại vi
- Bản chất ngẫu nhiên của con người
- Các biến thay thế kém
- Nguyên tắc chi li
- Dạng hàm sai

Mô hình hồi qui tuyến tính

□ Hàm hồi qui mẫu (SRF)

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

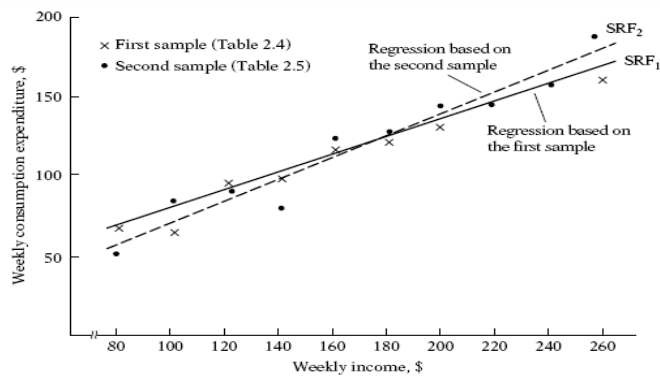
trong đó:

\hat{Y}_i là ước lượng của $E(Y_i|X_i)$

$\hat{\beta}_1$ và $\hat{\beta}_2$ là các ước lượng của β_1 và β_2 .

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

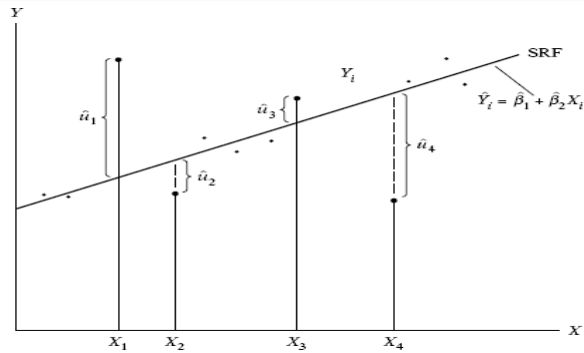
Mô hình hồi qui tuyến tính



Phương pháp bình phương tối thiểu thông thường (OLS)

- Phương pháp OLS

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



Phương pháp bình phương tối thiểu thông thường (OLS)

- Kết quả hồi qui

$$\hat{\beta}_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

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

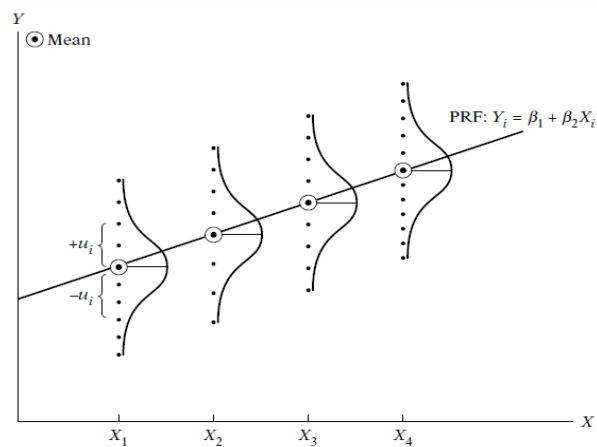
Mô hình hồi qui tuyến tính cổ điển Gauss (CLRM): Các giả thiết của OLS

- **Giả thiết 1: Mô hình hồi qui tuyến tính.** Mô hình hồi qui là tuyến tính theo các tham số của mô hình

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

- **Giả thiết 2: Các giá trị của X được cố định trong việc lấy mẫu lặp lại.** Giá trị lấy ra từ biến X được coi là cố định trong các mẫu lặp lại. X được cho là không ngẫu nhiên
- **Giả thiết 3: $E(u_i|X_i) = 0$**

Mô hình hồi qui tuyến tính cổ điển Gauss (CLRM): Các giả thiết của OLS



Mô hình hồi qui tuyến tính cổ điển Gauss (CLRM): Các giả thiết của OLS

- **Giả thiết 4:** Đồng phương sai giữa u_i và X_i bằng 0, $\text{cov}(u_i, X_i) = 0$.

→ Với GT 3 và 4, các tham số ước lượng theo OLS là không thiên lệch

$$E(\hat{\beta}_1) = \beta_1 \text{ và } E(\hat{\beta}_2) = \beta_2$$

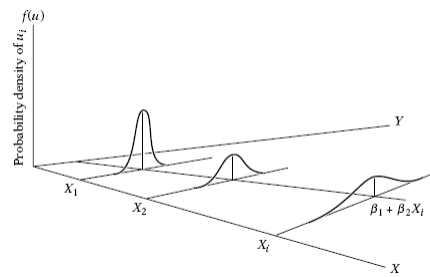
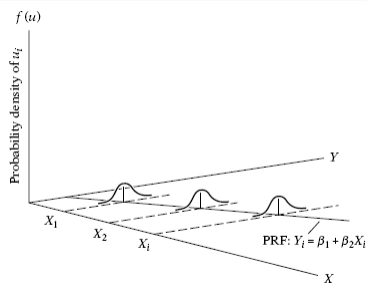
- **Giả thiết 5:** Sự biến thiên trong các giá trị của X . Các giá trị X_i trong mẫu cho trước không thể tất cả đều bằng nhau, $\text{var}(X_i) \neq 0$.

→ Với GT 3, 4, và 5, các tham số ước lượng theo OLS có tính nhất quán

Mô hình hồi qui tuyến tính cổ điển Gauss (CLRM): Các giả thiết của OLS

- **Giả thiết 6:** Phương sai của sai số không đổi.

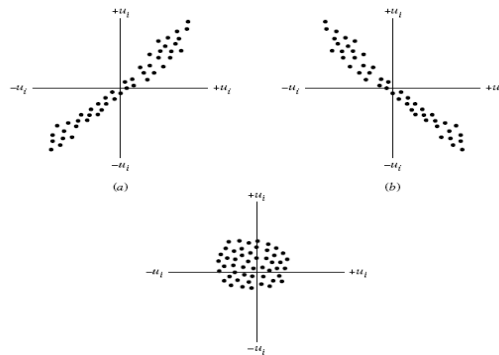
$$\text{var}(u_i | X_i) = E[u_i - E(u_i | X_i)]^2 = E(u_i^2 | X_i) = \sigma^2$$



Mô hình hồi qui tuyến tính cổ điển Gauss (CLRM): Các giả thiết của OLS

- **Giả thiết 7: Độc lập theo chuỗi.** Không có tương quan giữa các sai số

$$\text{cov}(u_i, u_j | X_i, X_j) = 0$$



Mô hình hồi qui tuyến tính cổ điển Gauss (CLRM): Các giả thiết của OLS

- **Giả thiết 8:** Mô hình hồi qui được xác định một cách đúng đắn (không có độ thiên lệch hoặc sai số đặc trưng)
 - **Giả thiết 9:** Không có tính đa cộng tuyến hoàn toàn
- **Định lý Gauss-Markov:** Ước lượng của OLS là ước lượng tuyến tính không thiên lệch, có tính nhất quán, và có hiệu quả nhất, BLUE.

Độ chính xác của ước lượng

- **Phương sai và độ lệch chuẩn của ước lượng**

$$\begin{aligned}\text{var}(\hat{\beta}_2) &= \frac{\sigma^2}{\sum x_i^2} \\ \text{se}(\hat{\beta}_2) &= \frac{\sigma}{\sqrt{\sum x_i^2}} \\ \text{var}(\hat{\beta}_1) &= \frac{\sum X_i^2}{n \sum x_i^2} \sigma^2 \\ \text{se}(\hat{\beta}_1) &= \sqrt{\frac{\sum X_i^2}{n \sum x_i^2}} \sigma\end{aligned}$$

trong đó
$$\hat{\sigma}^2 = \frac{\sum u_i^2}{n-2} = \frac{\sum y_i^2 - [(\sum x_i y_i)^2 / \sum x_i^2]}{n-2} \quad (n > 2)$$

Độ chính xác của ước lượng

- **Điều kiện:** Số lượng các quan sát n phải lớn hơn số lượng các tham số được ước lượng ($n > k$)
- **Đồng phương sai giữa 2 ước lượng**

$$\begin{aligned}\text{cov}(\hat{\beta}_1, \hat{\beta}_2) &= -\bar{X} \text{var}(\hat{\beta}_2) \\ &= -\bar{X} \left(\frac{\sigma^2}{\sum x_i^2} \right)\end{aligned}$$

Độ thích hợp của mô hình

23

Mối liên hệ giữa TSS, ESS, và RSS

$$TSS = ESS + RSS$$

$$\sum (Y_i - \bar{Y})^2 = \sum (\hat{Y}_i - \bar{Y})^2 + \sum (Y_i - \hat{Y}_i)^2$$

TSS = Tổng bình phương toàn phần

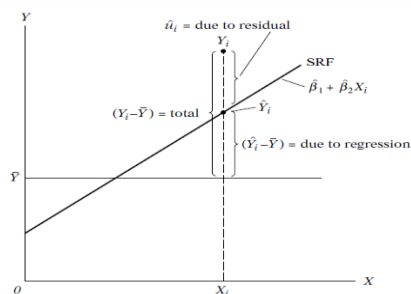
ESS = Tổng bình phương giải thích được

RSS = Tổng bình phương phần dư

Độ thích hợp của mô hình (goodness of fit)

- Hệ số xác định (coefficient of determination)

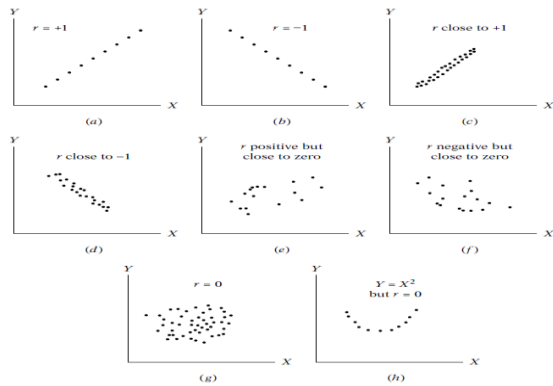
$$r^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$



Độ thích hợp của mô hình (goodness of fit)

- Hệ số tương quan mẫu

$$r = \pm \sqrt{r^2} \quad (\text{dấu của } r \text{ phụ thuộc vào dấu của } \hat{\beta}_2)$$



Phân phối xác suất của các sai số

- Với các giả thiết $E(u_i) = 0$, $cov(u_i, X_i) = 0$, $var(u_i/X_i) = \sigma^2$, $cov(u_i, u_j) = 0$, ước lượng OLS $\hat{\beta}_1$ và $\hat{\beta}_2$ là BLUE.
- Để kiểm định giả thuyết chúng ta cần biết phân phối xác suất của các sai số u_i
- Giả định u_i tuân theo phân phối chuẩn: $u_i \sim N(0, \sigma^2)$
- u_i chứa các biến độc lập không được đề cập trong mô hình hồi qui (tác động của những biến này là nhỏ và có thuộc tính ngẫu nhiên)
- Theo Định lý giới hạn trung tâm, một lượng lớn các biến ngẫu nhiên có phân phối giống nhau và độc lập thì phân phối của tổng các biến đó sẽ có phân phối chuẩn nếu số lượng các biến tăng lên vô hạn

Phân phối xác suất của các sai số

- Nếu $u_i \sim N(0, \sigma^2)$ thì phân phối xác suất của các ước lượng OLS cũng sẽ có phân phối chuẩn (bất kỳ hàm tuyến tính nào của các biến tuân theo phân phối chuẩn thì tự nó cũng sẽ có phân phối chuẩn).
- Các tính chất của ước lượng OLS theo giả định phân phối chuẩn

$$E(\hat{\beta}_2) = \beta_2$$

$$\text{Var}(\hat{\beta}_2) = \sigma_{\hat{\beta}_2}^2 = \frac{\sigma^2}{\sum x_i^2}$$

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

Kiểm định giả thuyết

- Phương pháp kiểm định ý nghĩa: **Kiểm định t**
- **Kiểm định 2 phía**

$$H_0: \beta_2 = a$$

$$H_a: \beta_2 \neq a$$

Trị kiểm định thống kê

$$t = \frac{\hat{\beta}_2 - \beta_2}{s_{\hat{\beta}_2}}$$

$$s_{\hat{\beta}_2} = \frac{\hat{\sigma}}{\sqrt{\sum x_i^2}}$$

Kiểm định giả thuyết

Quy tắc bác bỏ

- ❖ Bác bỏ nếu $|t| > t_{\alpha/2}$ với $t_{\alpha/2}$ dựa trên phân phối t với **bậc tự do là $(n-2)$**
- ❖ Hoặc $p_{\text{value}} < \alpha$.

▪ Quy tắc kinh nghiệm “2-t”

Nếu **bậc tự do lớn hơn hay bằng 20** và $\alpha = 5\%$ thì H_0 có thể bị bác bỏ nếu **giá trị thống kê t lớn hơn 2** theo giá trị tuyệt đối

Kiểm định giả thuyết

▪ Kiểm định 1 phía

$$H_0: \beta_2 \geq a$$

$$H_0: \beta_2 \leq a$$

$$H_a: \beta_2 < a$$

$$H_a: \beta_2 > a$$

Quy tắc bác bỏ

- ❖ Bác bỏ nếu $t < -t_\alpha$ $t > t_\alpha$
- ❖ Hoặc $p_{\text{value}} < \alpha$ $p_{\text{value}} < \alpha$

Kiểm định giả thuyết

*** Trường hợp kiểm định giả thuyết đối với β_1 .

$$t = \frac{\hat{\beta}_1 - \beta_1}{s_{\hat{\beta}_1}}$$

$$s_{\hat{\beta}_1} = \frac{\sum X_i^2}{n\sqrt{\sum x_i^2}} \hat{\sigma}$$

Kiểm định giả thuyết

- Phương pháp kiểm định dựa trên khoảng tin cậy $(1-\alpha)100\%$

$$\hat{\beta}_2 \pm t_{\alpha/2} s_{\hat{\beta}_2}$$

Quy tắc bác bỏ

Bác bỏ H_0 nếu 0 không nằm trong khoảng tin cậy của β_2

Kiểm định giả thuyết

- Phương pháp kiểm định ý nghĩa: Kiểm định F (Phân tích ANOVA)

$$F = \frac{MSE}{MSR}$$

Nguồn biến thiên	Công thức	Bậc tự do	MSS
ESS	$\sum \hat{y}_i^2 = \hat{\beta}_2^2 \sum x_i^2$	1	$\hat{\beta}_2^2 \sum x_i^2$
RSS	$\sum \hat{u}_i^2$	n-2	$\sum \hat{u}_i^2 / (n-2) = \hat{\sigma}^2$
TSS	$\sum y_i^2$	n-1	

Quy tắc bác bỏ

Bác bỏ H_0 nếu $F \geq F_\alpha$ (phân phối F với bậc tự do bằng 1) hoặc $p_{\text{value}} \leq \alpha$

Sử dụng phân tích hồi qui để ước lượng và dự báo

- Ước lượng khoảng tin cậy của giá trị trung bình $E(Y | X = X_0) = \hat{Y}_0$

$$\hat{Y}_0 \pm t_{\alpha/2} s_{\hat{Y}_0}$$

$$s_{\hat{Y}_0} = \hat{\sigma} \sqrt{\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2}}$$

- Ước lượng khoảng tin cậy của giá trị cá biệt Y_0

$$Y_0 \pm t_{\alpha/2} s_{ind}$$

$$s_{ind} = \hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2}}$$

PHÂN TÍCH PHẦN DƯ (Nguồn: Cao Hào Thi)

- Nếu các giả định về số hạng sai số u 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 u .
- 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

35

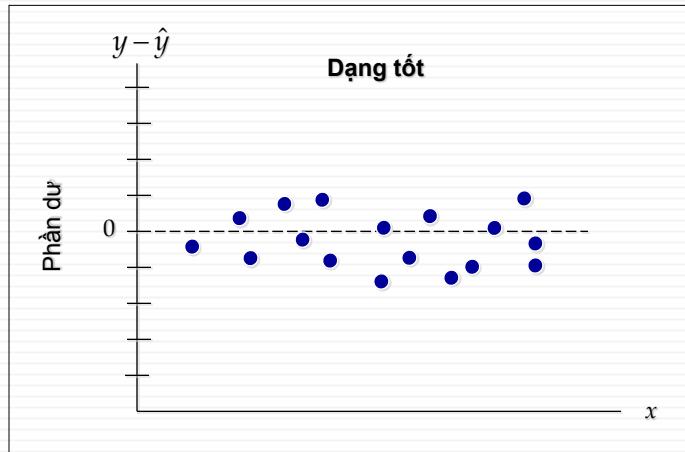
ĐỒ THỊ PHẦN DƯ THEO X (Nguồn: Cao Hào Thi)

- Nếu giả định $Var(u_i|X) = \sigma^2$ 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

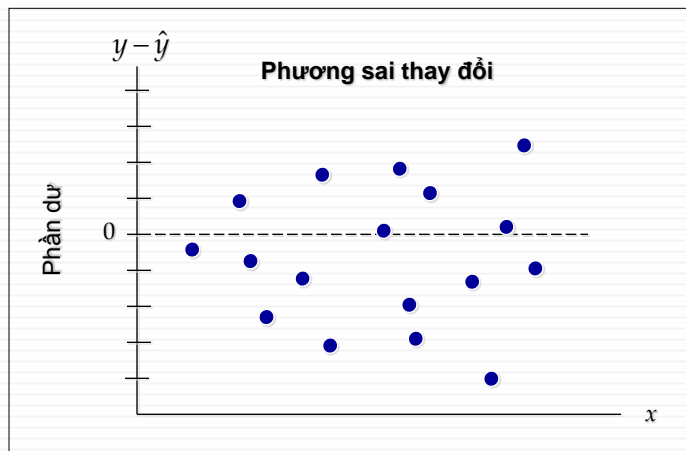
36

ĐỒ THỊ PHẦN DƯ THEO X (Nguồn: Cao Hào Thi)

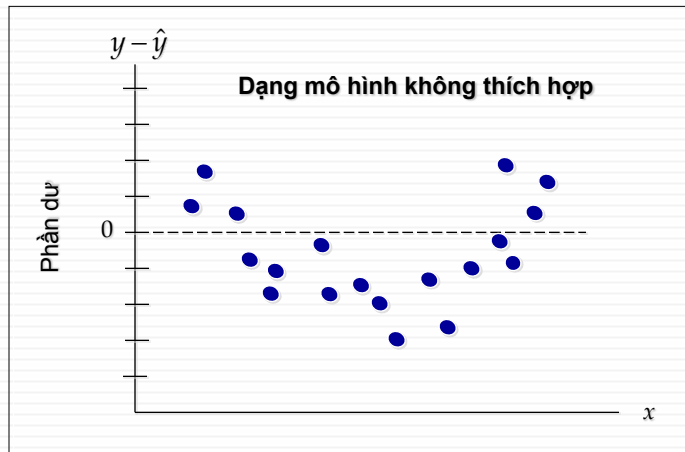


37

ĐỒ THỊ PHẦN DƯ THEO X (Nguồn: Cao Hào Thi)



38

ĐỒ THỊ PHẦN DƯ THEO X (Nguồn: Cao Hào Thi)

39