

# **2**

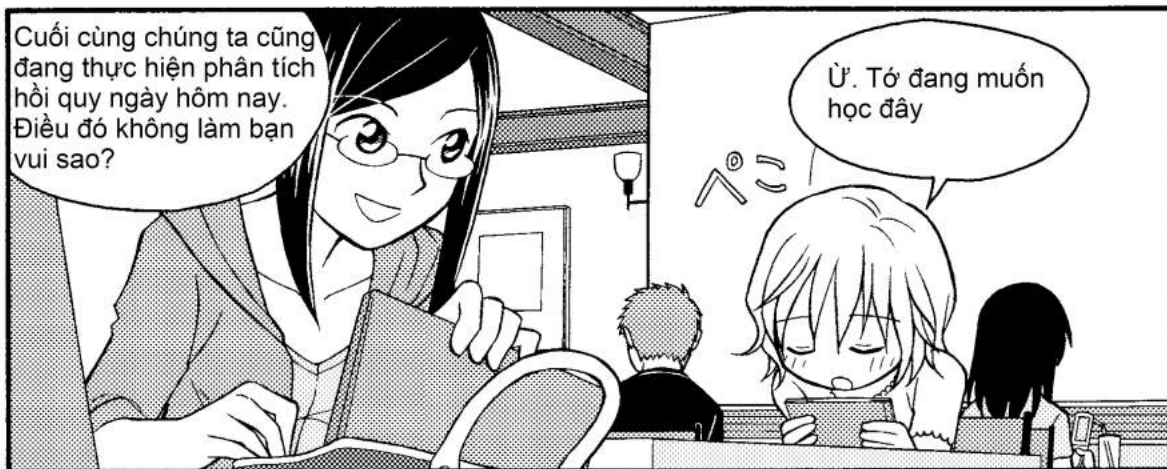
## ***Phân tích hồi quy đơn giản***





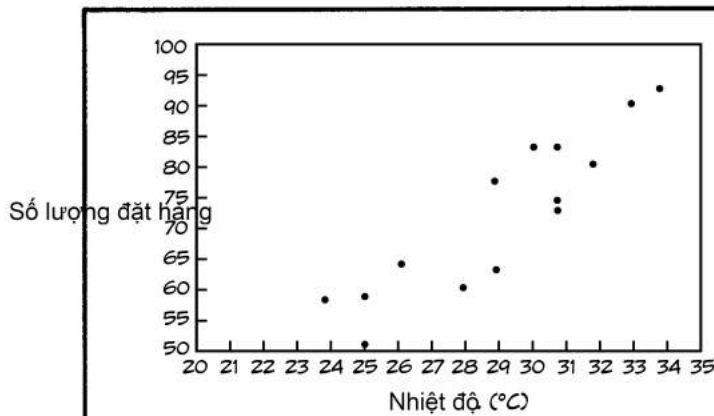


➤➤➤ HỌC THỐNG KÊ QUA TRUYỆN TRANH



Được rồi, chúng ta học thôi. Đây là bảng chỉ nhiệt độ và số lượng trà đã được đặt hàng trong 2 tuần.

Ngày	Nhiệt độ (°C)	Số lượng đặt hàng
22nd (Mon.)	29	77
23rd (Tues.)	28	62
24th (Wed.)	34	93
25th (Thurs.)	31	84
26th (Fri.)	25	59
27th (Sat.)	29	64
28th (Sun.)	32	80
29th (Mon.)	31	75
30th (Tues.)	24	58
31st (Wed.)	33	91
1st (Thurs.)	25	51
2nd (Fri.)	31	73
3rd (Sat.)	26	65
4th (Sun.)	30	84



... như này

Tớ biết

Xem cách các dấu chấm thẳng hàng như thế nào? Điều đó cho thấy các biến này có mối tương quan với nhau. Hệ số tương quan, được gọi là R, cho biết mức độ tương quan mạnh mẽ như thế nào.

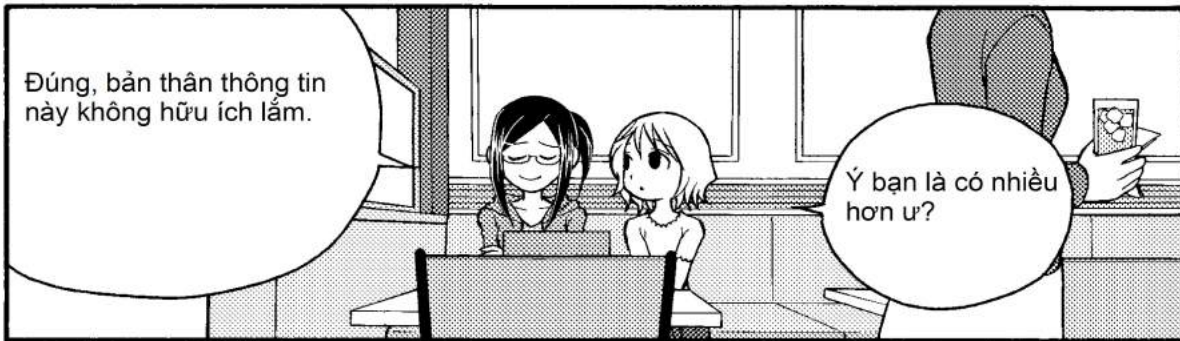
$$R = 0.9069$$

R nằm trong khoảng từ +1 đến -1 và càng xa 0 thì mối tương quan càng mạnh. \* Tớ sẽ chỉ cho cậu cách tính hệ số tương quan trên sau.

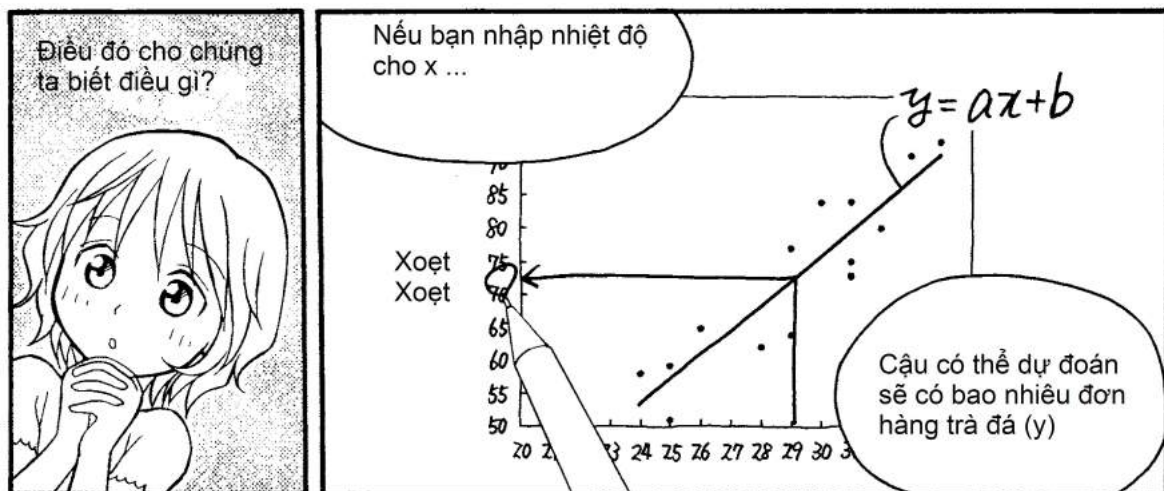
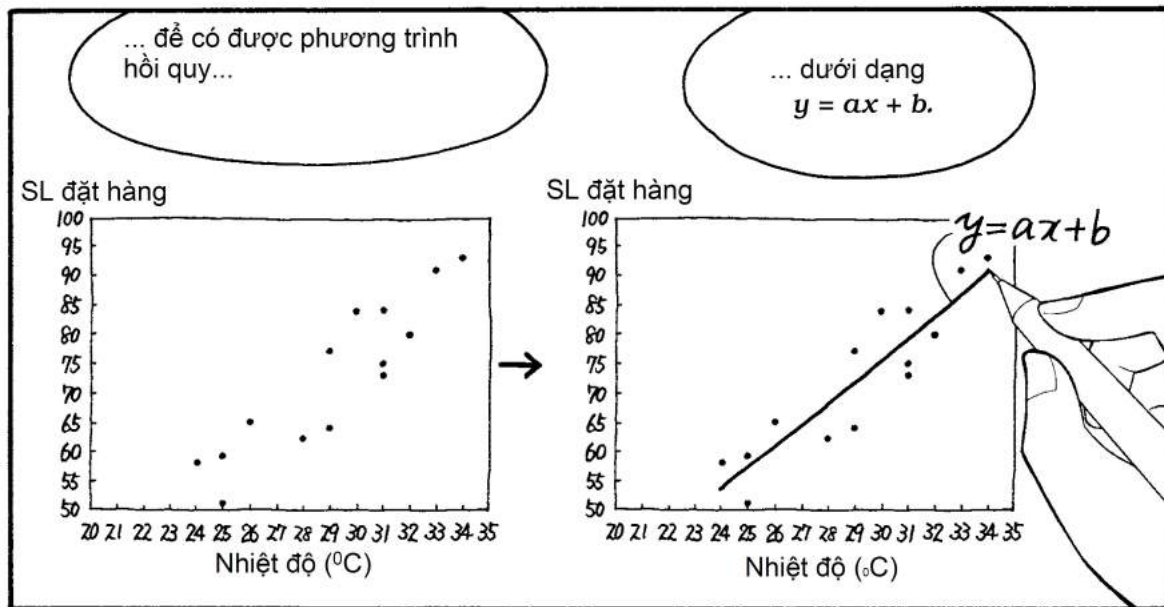
\* Giá trị R dương chỉ ra mối quan hệ thuận, nghĩa là khi x tăng thì y cũng tăng. Giá trị R âm có nghĩa là khi giá trị x tăng, giá trị y giảm.



➤➤➤ HỌC THỐNG KÊ QUA TRUYỆN TRANH



Phương trình hồi quy





➤➤➤ HỌC THỐNG KÊ QUA TRUYỆN TRANH



Tớ biết! Phân tích hồi quy đường như không quá khó.

Như tớ đã nói trước đó,  $y$  là biến phụ thuộc (hoặc kết quả) và  $x$  là biến độc lập (hoặc dự đoán).

$$y = ax + b$$

↑  
Biến độc lập

↑  
Biến phụ thuộc

$a$  là hệ số hồi quy, cho chúng ta biết độ dốc của đường chúng ta thực hiện



Hệ số  $b$  được gọi là hằng số hồi quy. Điều này cho chúng ta biết vị trí đường thẳng cắt qua trục  $y$ .

Được rồi, đã rõ.

Vậy làm cách nào để lấy được phương trình hồi quy?



Cố lên, Miu!

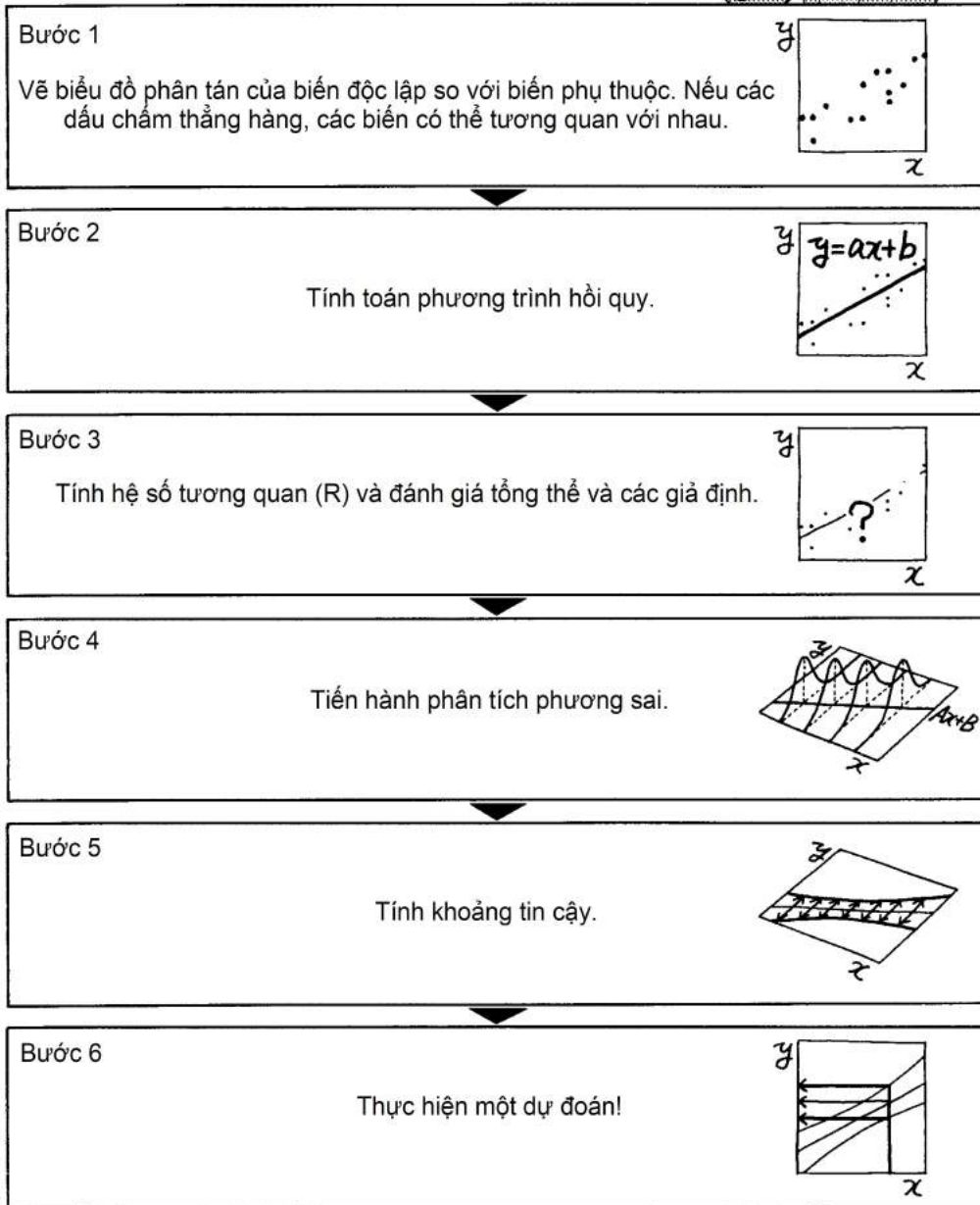


Tìm phương trình chỉ là một phần của câu chuyện

Cậu cũng cần học cách xác minh tính chính xác của phương trình bằng cách thử nghiệm trong một số trường hợp nhất định. Hãy xem xét toàn bộ quá trình.

Quy trình phân tích hồi quy chung

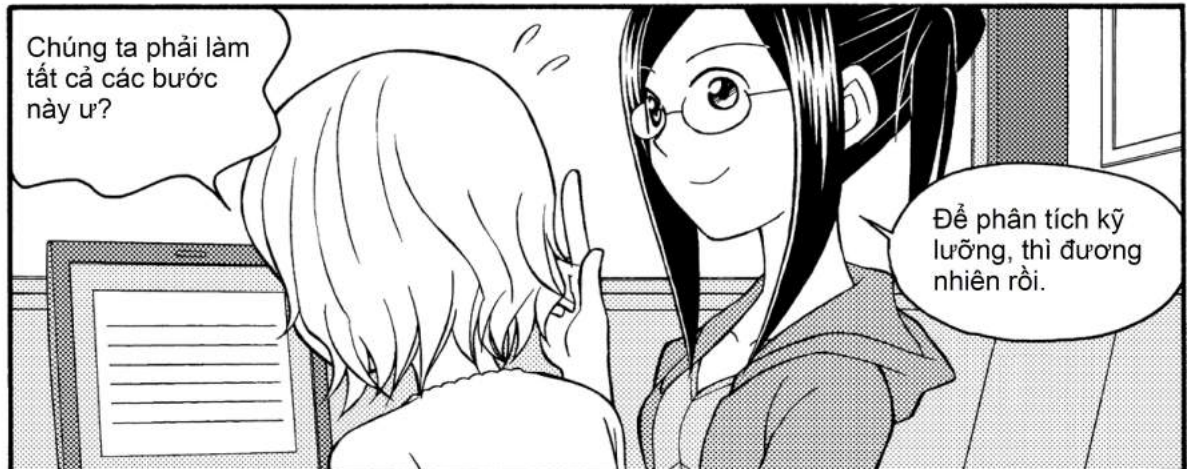
Đây là tổng quan về phân tích hồi quy.




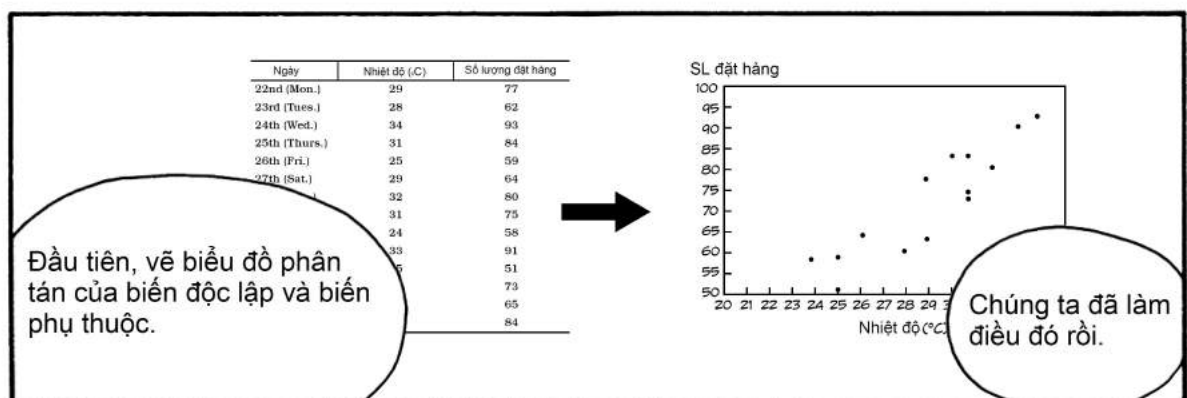
Chuẩn đoán mô hình hồi quy

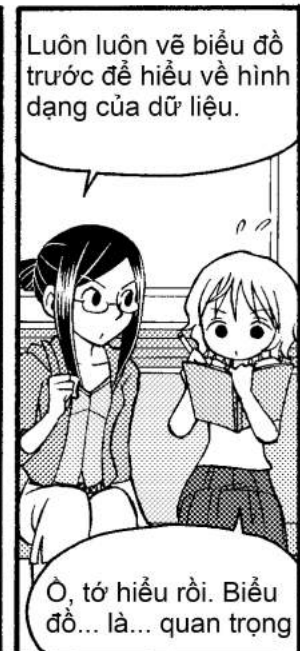
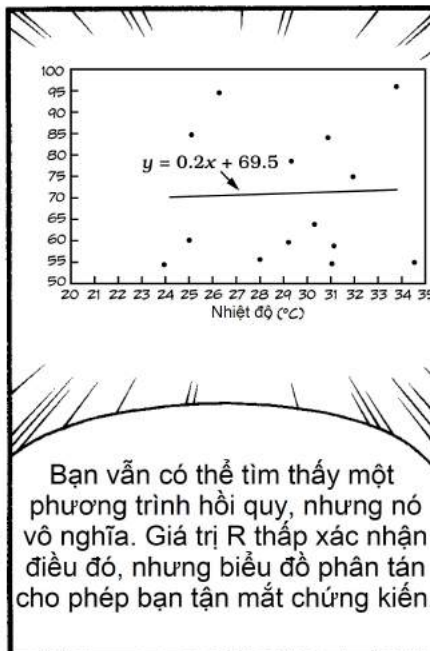
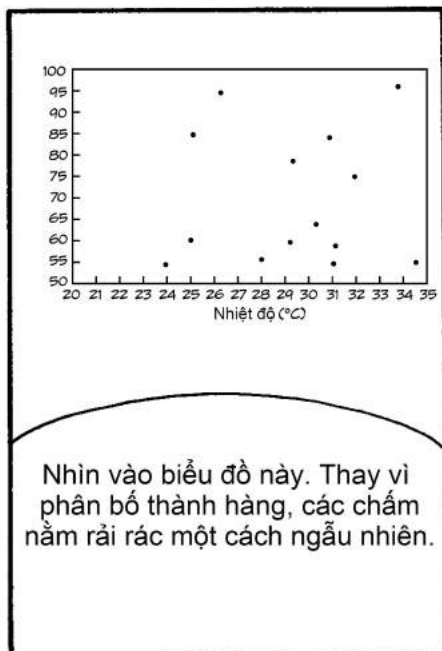
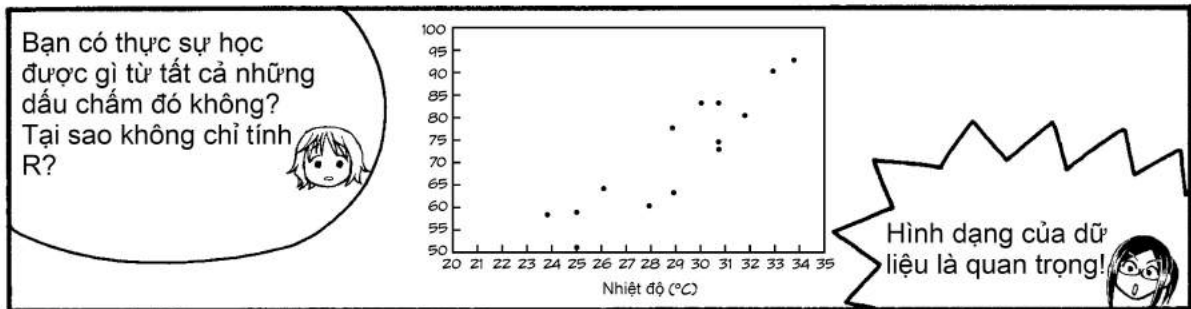


## ➤➤➤ HỌC THỐNG KÊ QUA TRUYỆN TRANH



**BƯỚC 1: VẼ BIỂU ĐỒ PHÂN TÁN CỦA BIẾN ĐỘC LẬP SO VỚI BIẾN PHỤ THUỘC. NẾU CÁC DẤU CHẤM THẲNG HÀNG, CÁC BIẾN CÓ THỂ TƯƠNG QUAN VỚI NHAU.**

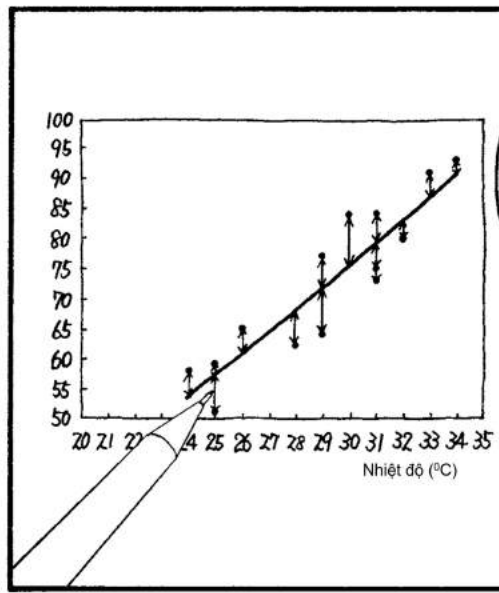
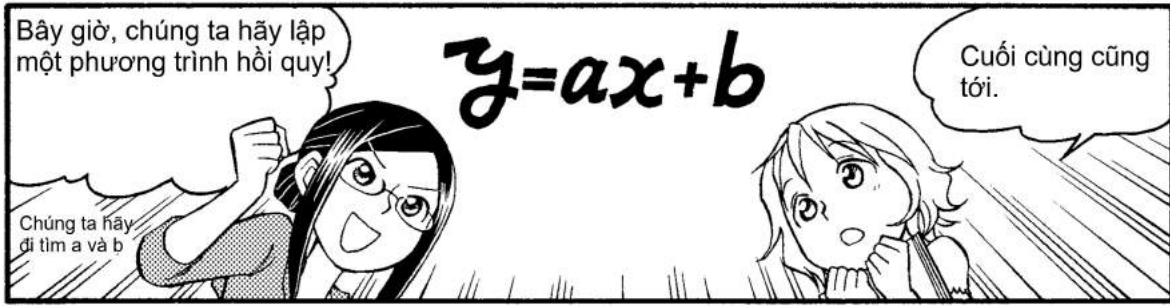







# ➤➤➤ HỌC THỐNG KÊ QUA TRUYỆN TRANH

BƯỚC 2: TÍNH TOÁN PHƯƠNG TRÌNH HỒI QUY.  $y = ax + b$



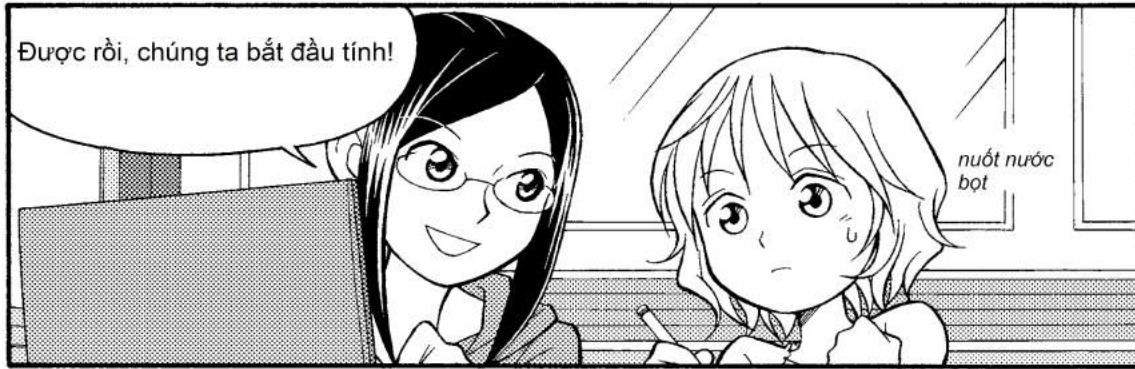
Các mũi tên nhỏ là khoảng cách từ đường thẳng, đại diện cho các giá trị ước tính của mỗi dấu chấm, là các giá trị đo được thực tế. Các khoảng cách được gọi là phần dư. Mục đích là để tìm đường có giá trị nhỏ tốt nhất cho tất cả các phần dư.

Đây được gọi là hồi quy bình phương nhỏ nhất tuyến tính



- Step 1** Tính  $S_x$  (tổng bình phương của x),  $S_y$  (tổng bình phương của y) và  $S_{xy}$  (tổng các tích của x và y).
- Step 2** Tính  $S_e$  (tổng bình phương phần dư).
- Step 3** Lấy vi phân  $S_e$  theo a và b, và đặt nó bằng 0.
- Step 4** Tách a và b ra.
- Step 5** Cố định a.
- Step 6** Tìm phương trình hồi quy.





Ta có:  
 Tổng bình phương của  $x$ ,  $S_{xx}: (x-\bar{x})^2$   
 Tổng bình phương của  $y$ ,  $S_{yy}: (y-\bar{y})^2$   
 Tổng tích của  $x$  và  $y$ ,  $S_{xy}: (x-\bar{x})(y-\bar{y})$

Lưu ý: Dấu ngang trên một biến (như  $x$ ) là một ký hiệu có nghĩa là trung bình. Chúng ta có thể gọi biến này là  $x$ -ngang.

Ngày	Nhiệt độ (°C) $x$	Số lượng đặt hàng $y$	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})^2$	$(y - \bar{y})^2$	$(x - \bar{x})(y - \bar{y})$
22nd (Mon.)	29	77	-0.1	4.4	0.0	19.6	-0.6
23rd (Tues.)	28	62	-1.1	-10.6	1.3	111.8	12.1
24th (Wed.)	34	93	4.9	20.4	23.6	417.3	99.2
25th (Thurs.)	31	84	1.9	11.4	3.4	130.6	21.2
26th (Fri.)	25	59	-4.1	-13.6	17.2	184.2	56.2
27th (Sat.)	29	64	-0.1	-8.6	0.0	73.5	1.2
28th (Sun.)	32	80	2.9	7.4	8.2	55.2	21.2
29th (Mon.)	31	75	1.9	2.4	3.4	5.9	4.5
30th (Tues.)	24	58	-5.1	-14.6	26.4	212.3	74.9
31st (Wed.)	33	91	3.9	18.4	14.9	339.6	71.1
1st (Thurs.)	25	51	-4.1	-21.6	17.2	465.3	89.4
2nd (Fri.)	31	73	1.9	0.4	3.4	0.2	0.8
3rd (Sat.)	26	65	-3.1	-7.6	9.9	57.8	23.8
4th (Sun.)	30	84	0.9	11.4	0.7	130.6	9.8
	408	1016	0	0	129.7	2203.4	484.9
<b>Tổng</b>	29.1	72.6					
Trung bình	$\bar{x}$	$\bar{y}$			$S_{xx}$	$S_{yy}$	$S_{xy}$

\* Một số số liệu trong chương này được làm tròn để tiện cho việc in ấn, nhưng các phép tính được thực hiện bằng cách sử dụng các giá trị đầy đủ, không làm tròn dữ liệu thô, trừ khi có quy định khác.



**Step2** Tìm tổng bình phương phần dư,  $S_e$ .

- $y$  là giá trị quan sát.
- $\hat{y}$  là giá trị ước tính dựa trên phương trình hồi quy của chúng ta.
- $y - \hat{y}$  được gọi là phần dư và được viết là  $e$ .

Lưu ý: Dấu mũ trong  $\hat{y}$  được gọi trừu mến là mũ, vì vậy chúng ta gọi ước lượng tham số này là  $y$ -mũ.

Ngày	Nhiệt độ (°C)	Số lượng đặt hàng	Ước tính số lượng đặt hàng	Phần dư ( $e$ )	Bình phương phần dư
	$x$	$y$	$\hat{y} = ax + b$	$y - \hat{y}$	$(y - \hat{y})^2$
22nd (Mon.)	29	77	$a \times 29 + b$	$77 - (a \times 29 + b)$	$[77 - (a \times 29 + b)]^2$
23rd (Tues.)	28	62	$a \times 28 + b$	$62 - (a \times 28 + b)$	$[62 - (a \times 28 + b)]^2$
24th (Wed.)	34	93	$a \times 34 + b$	$93 - (a \times 34 + b)$	$[93 - (a \times 34 + b)]^2$
25th (Thurs.)	31	84	$a \times 31 + b$	$84 - (a \times 31 + b)$	$[84 - (a \times 31 + b)]^2$
26th (Fri.)	25	59	$a \times 25 + b$	$59 - (a \times 25 + b)$	$[59 - (a \times 25 + b)]^2$
27th (Sat.)	29	64	$a \times 29 + b$	$64 - (a \times 29 + b)$	$[64 - (a \times 29 + b)]^2$
28th (Sun.)	32	80	$a \times 32 + b$	$80 - (a \times 32 + b)$	$[80 - (a \times 32 + b)]^2$
29th (Mon.)	31	75	$a \times 31 + b$	$75 - (a \times 31 + b)$	$[75 - (a \times 31 + b)]^2$
30th (Tues.)	24	58	$a \times 24 + b$	$58 - (a \times 24 + b)$	$[58 - (a \times 24 + b)]^2$
31st (Wed.)	33	91	$a \times 33 + b$	$91 - (a \times 33 + b)$	$[91 - (a \times 33 + b)]^2$
1st (Thurs.)	25	51	$a \times 25 + b$	$51 - (a \times 25 + b)$	$[51 - (a \times 25 + b)]^2$
2nd (Fri.)	31	73	$a \times 31 + b$	$73 - (a \times 31 + b)$	$[73 - (a \times 31 + b)]^2$
3rd (Sat.)	26	65	$a \times 26 + b$	$65 - (a \times 26 + b)$	$[65 - (a \times 26 + b)]^2$
4th (Sun.)	30	84	$a \times 30 + b$	$84 - (a \times 30 + b)$	$[84 - (a \times 30 + b)]^2$
Tổng	408	1016	$408a + 14b$	$1016 - (408a + 14b)$	$S_e$
Trung bình	29.1	72.6	$29.1a + b$ $= \bar{x}a + b$	$72.6 - (29.1a + b)$ $= \bar{y} - (\bar{x}a + b)$	$= \frac{S_e}{14}$

↓  
 $\bar{x}$

↓  
 $\bar{y}$

$$S_e = [77 - (a \times 29 + b)]^2 + \dots + [84 - (a \times 30 + b)]^2$$

Tổng bình phương phần dư còn được gọi là tổng bình phương dư. Nó được viết là  $S_e$  hoặc RSS.



**Step3**

Lấy vi phân  $S_e$  theo  $a$  và  $b$ , và đặt nó bằng 0. Khi vi phân  $y = (ax + b)^n$  theo  $x$ , kết quả là:

$$\frac{dy}{dx} = n(ax + b)^{n-1} \times a.$$

$$\frac{dS_e}{da} = 2[77 - (29a + b)] \times (-29) + \dots + 2[84 - (30a + b)] \times (-30) = 0 \quad \text{①}$$

$$\frac{dS_e}{db} = 2[77 - (29a + b)] \times (-1) + \dots + 2[84 - (30a + b)] \times (-1) = 0 \quad \text{②}$$

**Step4**

Sắp xếp lại 1 và 2 từ bước trước.  
Sắp xếp lại 1.

$$2[77 - (29a + b)] \times (-29) + \dots + 2[84 - (30a + b)] \times (-30) = 0$$

$$[77 - (29a + b)] \times (-29) + \dots + [84 - (30a + b)] \times (-30) = 0 \quad \text{Chia 2 về cho 2}$$

$$29[(29a + b) - 77] + \dots + 30[(30a + b) - 84] = 0 \quad \text{nhân 2 về cho -1}$$

$$(29 \times 29a + 29 \times b - 29 \times 77) + \dots + (30 \times 30a + 30 \times b - 30 \times 84) = 0 \quad \text{Nhân ra}$$

$$\text{③ } (29^2 + \dots + 30^2)a + (29 + \dots + 30)b - (29 \times 77 + \dots + 30 \times 84) = 0 \quad \text{Tách a và b ra}$$

Sắp xếp lại 2.

$$2[77 - (29a + b)] \times (-1) + \dots + 2[84 - (30a + b)] \times (-1) = 0$$

$$[77 - (29a + b)] \times (-1) + \dots + [84 - (30a + b)] \times (-1) = 0 \quad \text{Chia 2 về cho 2}$$

$$[(29a + b) - 77] + \dots + [(30a + b) - 84] = 0 \quad \text{nhân 2 về cho -1}$$

$$(29 + \dots + 30)a + \underbrace{b + \dots + b}_{14} - (77 + \dots + 84) = 0 \quad \text{Tách a và b ra}$$

$$(29 + \dots + 30)a + 14b - (77 + \dots + 84) = 0$$

$$14b = (77 + \dots + 84) - (29 + \dots + 30)a \quad \text{Trừ 14b cho cả hai vế và nhân với -1.}$$

$$\text{④ } b = \frac{77 + \dots + 84}{14} - \frac{29 + \dots + 30}{14}a \quad \text{Tách b ở vế trái của phương trình.}$$

$$\text{⑤ } b = \bar{y} - \bar{x}a \quad \text{Các thành phần trong 4 là trung bình của y và x.}$$



**Step 5** Gán giá trị của b tìm được ở 4 vào dòng 3 (3 và 4 là kết quả từ Bước 4).

$$\textcircled{3} (29^2 + \dots + 30^2)a + (29 + \dots + 30) \left( \frac{77 + \dots + 84}{14} - \frac{29 + \dots + 30}{14} a \right) - (29 \times 77 + \dots + 30 \times 84) = 0$$

Giờ đây a là biến duy nhất.

$$(29^2 + \dots + 30^2)a + \frac{(29 + \dots + 30)(77 + \dots + 84)}{14} - \frac{(29 + \dots + 30)^2}{14}a - (29 \times 77 + \dots + 30 \times 84) = 0$$

$$\left[ (29^2 + \dots + 30^2) - \frac{(29 + \dots + 30)^2}{14} \right] a + \frac{(29 + \dots + 30)(77 + \dots + 84)}{14} - (29 \times 77 + \dots + 30 \times 84) = 0$$

Kết hợp biểu thức

$$\left[ (29^2 + \dots + 30^2) - \frac{(29 + \dots + 30)^2}{14} \right] a = (29 \times 77 + \dots + 30 \times 84) - \frac{(29 + \dots + 30)(77 + \dots + 84)}{14}$$

Chuyển vế

Sắp xếp lại vế trái của phương trình.

$$\begin{aligned} & (29^2 + \dots + 30^2) - \frac{(29 + \dots + 30)^2}{14} \\ &= (29^2 + \dots + 30^2) - 2 \times \frac{(29 + \dots + 30)^2}{14} + \frac{(29 + \dots + 30)^2}{14} \end{aligned}$$

Chúng ta cộng và trừ  $\frac{(29 + \dots + 30)^2}{14}$ .

$$\begin{aligned} &= (29^2 + \dots + 30^2) - 2 \times (29 + \dots + 30) \times \frac{29 + \dots + 30}{14} + \left( \frac{29 + \dots + 30}{14} \right)^2 \times 14 \end{aligned}$$

Số hạng cuối cùng được nhân với  $\frac{14}{14}$ .

$$\begin{aligned} &= (29^2 + \dots + 30^2) - 2 \times (29 + \dots + 30) \times \bar{x} + (\bar{x})^2 \times 14 \end{aligned}$$

$\bar{x} = \frac{29 + \dots + 30}{14}$

$$\begin{aligned} &= (29^2 + \dots + 30^2) - 2 \times (29 + \dots + 30) \times \bar{x} + \underbrace{(\bar{x})^2 + \dots + (\bar{x})^2}_{14} \\ &= [29^2 - 2 \times 29 \times \bar{x} + (\bar{x})^2] + \dots + [30^2 - 2 \times 30 \times \bar{x} + (\bar{x})^2] \\ &= (29 - \bar{x})^2 + \dots + (30 - \bar{x})^2 \\ &= S_{xx} \end{aligned}$$

Sắp xếp lại vế phải của phương trình.

$$\begin{aligned} & (29 \times 77 + \dots + 30 \times 84) - \frac{(29 + \dots + 30)(77 + \dots + 84)}{14} \\ &= (29 \times 77 + \dots + 30 \times 84) - \frac{29 + \dots + 30}{14} \times \frac{77 + \dots + 84}{14} \times 14 \\ &= (29 \times 77 + \dots + 30 \times 84) - \bar{x} \times \bar{y} \times 14 \end{aligned}$$

Chúng ta cộng và trừ  $\bar{x} \times \bar{y} \times 14$ .

$$\begin{aligned} &= (29 \times 77 + \dots + 30 \times 84) - \bar{x} \times \bar{y} \times 14 - \bar{x} \times \bar{y} \times 14 + \bar{x} \times \bar{y} \times 14 \\ &= (29 \times 77 + \dots + 30 \times 84) - \frac{29 + \dots + 30}{14} \times \bar{y} \times 14 - \bar{x} \times \frac{77 + \dots + 84}{14} \times 14 + \bar{x} \times \bar{y} \times 14 \\ &= (29 \times 77 + \dots + 30 \times 84) - (29 + \dots + 30) \bar{y} - \bar{x} (77 + \dots + 84) + \bar{x} \times \bar{y} \times 14 \\ &= (29 \times 77 + \dots + 30 \times 84) - (29 + \dots + 30) \bar{y} - (77 + \dots + 84) \bar{x} + \underbrace{\bar{x} \times \bar{y} + \dots + \bar{x} \times \bar{y}}_{14} \\ &= (29 - \bar{x})(77 - \bar{y}) + \dots + (30 - \bar{x})(84 - \bar{y}) \\ &= S_{xy} \end{aligned}$$

$$S_{xx} a = S_{xy}$$

$$\textcircled{6} a = \frac{S_{xy}}{S_{xx}}$$

Cô lập a ở vế trái của phương trình.

**Step 6** Tính toán phương trình hồi quy.

Từ 6 trong Bước 5,  $a = \frac{S_{xy}}{S_{xx}}$  Từ 5 trong Bước 4,  $b = \bar{y} - \bar{x}a$ .

Nếu chúng tôi gán các giá trị mà chúng ta đã tính toán ở Bước 1,

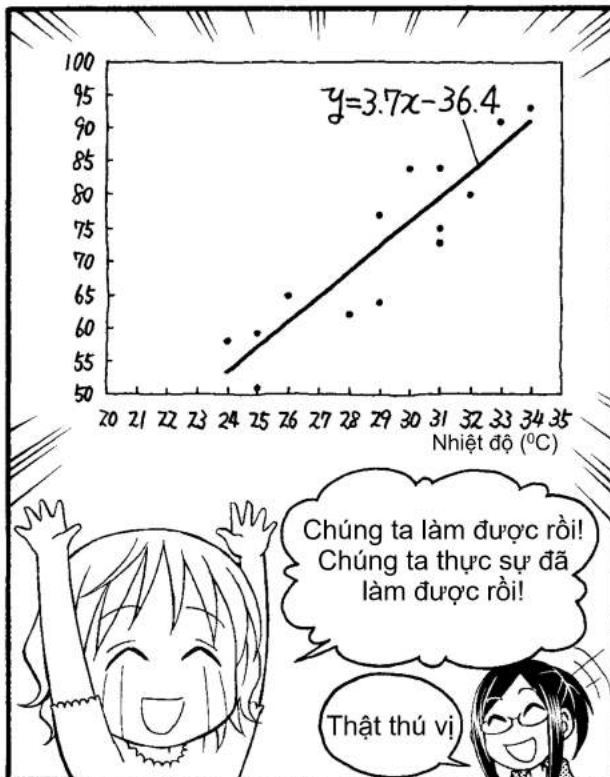
$$\begin{cases} a = \frac{S_{xx}}{S_{xy}} = \frac{484.9}{129.7} = 3.7 \\ b = \bar{y} - \bar{x}a = 72.6 - 29.1 \times 3.7 = -36.4 \end{cases}$$

thì phương trình hồi quy là

$$y = 3.7x - 36.4.$$

Thật đơn giản!

Lưu ý: Các giá trị hiển thị được làm tròn để thuận tiện cho việc in ấn, nhưng kết quả (36.4) được tính bằng cách sử dụng các giá trị đầy đủ, không làm tròn.



Mối quan hệ giữa phần dư với hệ số góc a và hệ số giao nhau b luôn là:

$$a = \frac{\text{Tổng các tích của x và y}}{\text{Tổng bình phương của x}} = \frac{S_{xy}}{S_{xx}}$$

$$b = \bar{y} - \bar{x}a$$

Điều này đúng với bất kỳ hồi quy tuyến tính nào.



➤➤➤ HỌC THỐNG KÊ QUA TRUYỆN TRANH

Miu, vậy giá trị trung bình của nhiệt độ và đơn đặt hàng trà đá là bao nhiêu?

Hãy nhớ rằng, nhiệt độ trung bình là  $x$  và số lượng đơn đặt hàng trung bình là  $y$ . Bây giờ có một phép thuật nhỏ.

Để tớ xem...

29.1°C và 72.6 đơn đặt hàng

Không cần nhìn, tớ có thể nói với cậu rằng phương trình hồi quy đi qua điểm (29, 1, 72,6).

Phương trình hồi quy có thể là...

$$y = ax + b$$

$$= a\bar{x} + (\bar{y} - \bar{x}a)$$

$$= a(x - \bar{x}) + \bar{y}$$

... sắp xếp lại như này

Đó là từ Bước 4!

Tớ hiểu!

Bây giờ, nếu chúng ta đặt  $x$  là giá trị trung bình ( $\bar{x}$ ) mà chúng ta đã tìm thấy trước đây ...

$$= a(x - \bar{x}) + \bar{y}$$

$$\rightarrow = a(\bar{x} - \bar{x}) + \bar{y}$$

$$= a \times 0 + \bar{y}$$

$$= \bar{y}$$

Xem điều gì sẽ xảy ra?

khi  $x$  là giá trị trung bình, thì  $y$  cũng vậy!

Biên dịch: Anh Tuấn  
(Còn tiếp)