

# PHƯƠNG PHÁP PHÂN TÍCH TƯƠNG QUAN

PGS.TS. Tăng Văn Khiên

## 1. Liên hệ tương quan và phương pháp phân tích tương quan

Mối liên hệ ràng buộc lẫn nhau giữa các chỉ tiêu hoặc tiêu thức của hiện tượng (từ đây chỉ dùng từ "chỉ tiêu" đặc trưng cho cả hai), trong đó sự biến động của một chỉ tiêu này (chỉ tiêu kết quả) là do tác động của nhiều chỉ tiêu khác (các chỉ tiêu nguyên nhân) gọi là liên hệ tương quan - một hình thức liên hệ không chặt chẽ.

Một phương pháp toán học áp dụng vào việc phân tích thống kê nhằm biểu hiện và nghiên cứu mối liên hệ tương quan giữa các chỉ tiêu của hiện tượng kinh tế xã hội gọi là phân tích tương quan.

Quá trình phân tích tương quan gồm các công việc cụ thể sau:

- Phân tích định tính về bản chất của mối quan hệ, đồng thời dùng phương pháp phân tổ hoặc đồ thị để xác định tính chất và xu thế của mối quan hệ đó.

- Biểu hiện cụ thể mối liên hệ tương quan bằng phương trình hồi quy tuyến tính hoặc phi tuyến tính và tính các tham số của các phương trình.

- Đánh giá mức độ chặt chẽ của mối liên hệ tương quan bằng các hệ số tương quan hoặc tỉ số tương quan.

## 2. Phân tích mối liên hệ tương quan giữa các chỉ tiêu biến đổi theo không gian

Liên hệ tương quan giữa các chỉ tiêu biến đổi theo không gian, nghĩa là mối liên hệ của các chỉ tiêu được nghiên cứu trên góc độ ở các không gian khác nhau và được sắp

xếp theo một thứ tự nào đó. Ví dụ, nghiên cứu mối liên hệ giữa tuổi nghề của công nhân với năng suất lao động của họ.

Với liên hệ tương quan không gian, thường nghiên cứu 3 trường hợp: liên hệ tương quan tuyến tính giữa hai chỉ tiêu, liên hệ tương quan phi tuyến tính giữa 2 chỉ tiêu và liên hệ tương quan tuyến tính giữa nhiều chỉ tiêu.

### 2.1. Liên hệ tương quan tuyến tính giữa 2 chỉ tiêu

#### a. Phương trình hồi quy tuyến tính (đường thẳng)

Nếu gọi y và x là các trị số thực tế của chỉ tiêu kết quả và chỉ tiêu nguyên nhân có thể xây dựng được phương trình hồi quy đường thẳng như sau:

$$\tilde{y}_x = a + bx \quad ; \quad (1a)$$

Trong đó:  $\tilde{y}_x$  - trị số lý thuyết (điều chỉnh) của chỉ tiêu kết quả;

a và b là các hệ số của phương trình

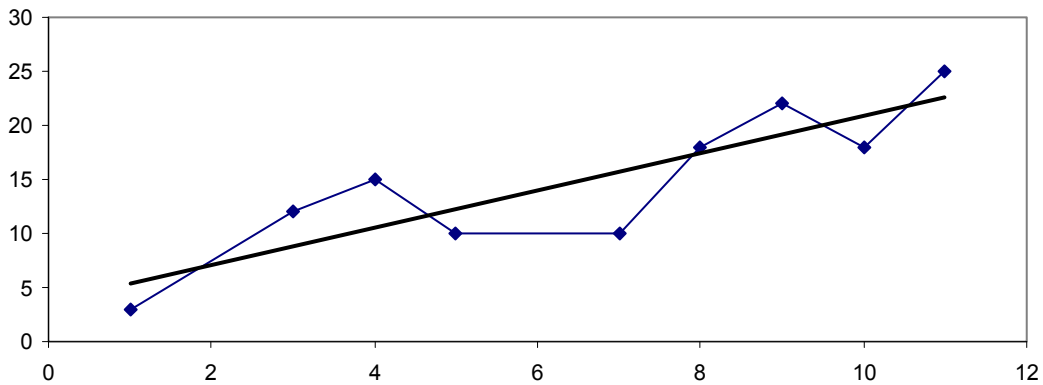
Có thể biểu diễn giá trị thực tế và giá trị lý thuyết của chỉ tiêu kết quả (qua trục tung) trong quan hệ với chỉ tiêu nguyên nhân (qua trục hoành) theo đồ thị 1:

Bằng phương pháp bình phương nhỏ nhất xây dựng được hệ phương trình chuẩn tắc xác định các hệ số a và b của phương trình đường thẳng:

$$\begin{cases} na + b\Sigma x = \Sigma y \\ a\Sigma x + b\Sigma x^2 = \Sigma xy \end{cases} \quad ; \quad (1b)$$

Ví dụ, có số liệu về tuổi nghề và năng suất lao động của các công nhân như cột 1 và 2 bảng 1:

ĐỒ THỊ 1: ĐẶC TRƯNG MỐI QUAN HỆ GIỮA CHỈ TIÊU KẾT QUẢ (Y) VÀ CHỈ TIÊU NGUYÊN NHÂN (X)



BẢNG 1. BẢNG TÍNH TOÁN CÁC HỆ SỐ CỦA PHƯƠNG TRÌNH ĐƯỜNG THẲNG

STT công nhân	Tuổi nghề - x (năm)	Năng suất lao động - y (triệu đồng)	xy	x <sup>2</sup>	y <sup>2</sup>
A	1	2	3=1x2	4=(1) <sup>2</sup>	5=(2) <sup>2</sup>
A	1	3	3	1	9
B	3	12	36	9	144
C	4	9	36	16	81
D	5	16	84	49	144
E	7	12	84	49	144
F	8	21	168	64	441
G	9	21	189	81	441
H	10	24	240	100	576
I	11	19	209	121	361
K	12	27	324	144	729
<b>Tổng</b>	<b>70</b>	<b>164</b>	<b>1369</b>	<b>610</b>	<b>3182</b>
<b>Trung bình</b>	<b>7</b>	<b>16,4</b>	<b>137,3</b>	<b>x</b>	<b>x</b>

Từ số liệu đã cho của x và y ở bảng 1, ta tính toán các đại lượng xy, x<sup>2</sup> và y<sup>2</sup> như cột 3, 4 và 5 của bảng.

Thay số liệu tính được ở bảng 2 vào hệ phương trình 1b, tính được: a = 3,52, b=1,84

Gán giá trị a và b vào phương trình tổng quát có dạng cụ thể của phương trình đường thẳng là:  $\tilde{y}_x = 3,52 + 1,84x$

b. Hệ số tương quan tuyến tính giữa hai chỉ tiêu (ký hiệu là r)

Công thức tính hệ số tương quan:

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\delta_x \cdot \delta_y} \quad ; \quad (2a)$$

$$\text{hoặc } r = b \cdot \frac{\delta_x}{\delta_y} \quad ; \quad (2b)$$

$$\text{Trong đó: } \overline{xy} = \frac{\sum xy}{n} \quad ; \quad \bar{x} = \frac{\sum x}{n} \quad ; \quad \bar{y} = \frac{\sum y}{n}$$

$$\delta_x = \sqrt{\frac{(x - \bar{x})^2}{n}} = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

$$\delta_y = \sqrt{\frac{(y - \bar{y})^2}{n}} = \sqrt{\frac{\sum y^2}{n} - \left(\frac{\sum y}{n}\right)^2}$$

Hệ số tương quan lấy giá trị trong khoảng từ -1 đến 1 ( $-1 \leq r \leq 1$ ):

Khi r càng gần 0 thì quan hệ càng lỏng lẻo, ngược lại khi r càng gần 1 hoặc -1 thì quan hệ càng chặt chẽ (r > 0 có quan hệ thuận và r < 0 có quan hệ nghịch). Trường hợp r=0 thì giữa x và y không có quan hệ.

Từ số liệu bảng 1, ta tính được:

$$\delta_x = \sqrt{\frac{610}{10} - \left(\frac{70}{10}\right)^2} = 3,464 \quad ;$$

$$\delta_y = \sqrt{\frac{3182}{10} - \left(\frac{164}{10}\right)^2} = 7,017$$

Hệ số tương quan (theo công thức 2a):

$$r = \frac{136,9 - (7 \times 16,4)}{3,464 \times 7,017} = 0,909$$

Theo kết quả tính toán có r = 0,909, chứng tỏ giữa tuổi nghề và năng suất lao động của công nhân có mối liên hệ thuận khá chặt chẽ.

## 2.2. Liên hệ tương quan phi tuyến tính giữa 2 chỉ tiêu

### a. Các phương trình hồi quy

Phương trình hồi quy phi tuyến tính thường được sử dụng:

\* Phương trình parabol bậc 2:

$$\tilde{y}_x = a + bx + cx^2 \quad ; \quad (3)$$

Phương trình parabol bậc 2 thường được áp dụng trong trường hợp các trị số của chỉ tiêu nguyên nhân tăng lên thì trị số của

chỉ tiêu kết quả tăng (hoặc giảm), việc tăng (hoặc giảm) đạt đến trị số cực đại (hoặc cực tiểu) rồi sau đó lại giảm (hoặc tăng).

\* Phương trình hyperbol

$$\tilde{y}_x = a + \frac{b}{x} \quad ; \quad (4)$$

Phương trình hyperbol được áp dụng trong trường hợp các trị số của chỉ tiêu nguyên nhân tăng lên thì trị số của chỉ tiêu kết quả giảm nhưng mức độ giảm nhỏ dần và đến một giới hạn nào đó ( $\tilde{y}_x = a$ ) thì hầu như không giảm.

\* Phương trình hàm số mũ

$$\tilde{y}_x = a.b^x \quad ; \quad (5)$$

Phương trình hàm số mũ được áp dụng trong trường hợp cùng với sự tăng lên của chỉ tiêu nguyên nhân thì trị số của các chỉ tiêu kết quả thay đổi theo cấp số nhân, nghĩa là có tốc độ tăng xấp xỉ nhau.

Bằng phương pháp bình phương nhỏ nhất ta xây dựng được các hệ phương trình chuẩn tắc phù hợp để xác định các hệ số của các phương trình tương ứng (3, 4 và 5)

### b. Tỷ số tương quan

Đối với liên hệ tương quan phi tuyến tính giữa 2 chỉ tiêu sẽ dùng tỷ số tương quan (ký hiệu  $\eta = \text{eta}$ ) để đánh giá mức độ chặt chẽ của mối liên hệ. Công thức tính tỷ số tương quan như sau:

$$\eta = \sqrt{\frac{\delta_{\tilde{y}_x}^2}{\delta_y^2}} = \frac{\delta_{\tilde{y}_x}}{\delta_y} \quad ; \quad (6)$$

Trong đó:  $\delta_{\tilde{y}_x}^2 = \frac{(\tilde{y}_x - \bar{y})^2}{n}$ : Phương sai đo độ biến thiên của chỉ tiêu y do ảnh hưởng riêng của chỉ tiêu x; với  $\tilde{y}_x$  là giá trị lý thuyết của đường hồi quy phi tuyến tính giữa y và x được xác định;

$\delta_y^2 = \frac{(y - \bar{y})^2}{n}$ : Phương sai đo độ biến thiên của chỉ tiêu y do ảnh hưởng của tất cả các chỉ tiêu nguyên nhân.

Tỉ số tương quan có một số tính chất sau:

(1) Tỉ số tương quan lấy giá trị trong khoảng [0;1], tức là  $0 \leq \eta \leq 1$ .

- Nếu  $\eta = 0$  thì giữa x và y không có liên hệ tương quan;

- Nếu  $\eta = 1$  thì giữa x và y có liên hệ hàm số;

- Nếu  $\eta$  càng gần 1 thì giữa x và y liên hệ tương quan càng chặt chẽ, và càng gần 0 thì liên hệ tương quan càng lỏng lẻo.

(2) Tỉ số tương quan lớn hơn hoặc bằng giá trị tuyệt đối của hệ số tương quan, tức là  $\eta \geq |r|$ . Nếu  $\eta = |r|$  thì giữa x và y có mối liên hệ tương quan tuyến tính.

### 2.3. Liên hệ tương quan tuyến tính giữa nhiều chỉ tiêu

Để dễ theo dõi dưới đây chỉ trình bày nội dung và phương pháp phân tích mối liên hệ tương quan giữa 3 chỉ tiêu.

#### a. Phương trình hồi quy tuyến tính giữa 3 chỉ tiêu

Nếu gọi y là chỉ tiêu kết quả và  $x_1, x_2$  là các chỉ tiêu nguyên nhân, ta có phương trình hồi quy tuyến tính giữa 3 chỉ tiêu như sau:

$$\tilde{y}_{x_1, x_2} = a_0 + a_1 x_1 + a_2 x_2 \quad ; \quad (7)$$

Bằng phương pháp bình phương nhỏ nhất, xây dựng được hệ phương trình chuẩn tắc để tính các tham số  $a_0, a_1$  và  $a_2$  của phương trình hồi quy 7:

#### b. Hệ số tương quan

Để đánh giá trình độ chặt chẽ mối liên hệ tương quan tuyến tính nhiều chỉ tiêu, người

ta thường tính toán hệ số tương quan: hệ số tương quan bội và hệ số tương quan riêng.

\* Hệ số tương quan bội (Ký hiệu là R) được dùng để đánh giá độ chặt chẽ giữa chỉ tiêu kết quả với tất cả các chỉ tiêu nguyên nhân được nghiên cứu. Công thức tính như sau:

$$R = \sqrt{\frac{r_{yx_1}^2 + r_{yx_2}^2 - 2r_{yx_1} r_{yx_2} r_{x_1 x_2}}{1 - r_{x_1 x_2}^2}} \quad ; \quad (8)$$

Trong đó:  $r_{yx_1}, r_{yx_2}$  và  $r_{x_1 x_2}$  là các hệ số tương quan tuyến tính giữa các cặp chỉ tiêu y với  $x_1, x_2$  và  $x_1$  với  $x_2$  và được tính như các công thức 2a hoặc 2b.

Hệ số tương quan bội nhận giá trị trong khoảng [0;1], tức là  $0 \leq R \leq 1$ .

Như vậy, R càng gần 0 thì quan hệ tương quan càng lỏng lẻo và R càng gần 1 thì quan hệ càng chặt chẽ.

Nếu  $R=0$  thì không có quan hệ tương quan và nếu  $R=1$  thì quan hệ tương quan trở thành quan hệ hàm số.

\* Hệ số tương quan riêng được dùng để đánh giá mức độ chặt chẽ của mối liên hệ giữa tiêu thức kết quả với từng tiêu thức nguyên nhân với điều kiện loại trừ ảnh hưởng của các tiêu thức nguyên nhân khác. Trong trường hợp mối liên hệ giữa y với  $x_1$  và  $x_2$  ở trên có thể tính:

- Hệ số tương quan riêng giữa y và  $x_1$  (loại trừ ảnh hưởng của  $x_2$ ):

$$r_{yx_1(x_2)} = \frac{r_{yx_1} - r_{yx_2} \times r_{x_1 x_2}}{\sqrt{(1 - r_{yx_2}^2) \cdot (1 - r_{x_1 x_2}^2)}} \quad ; \quad (9a)$$

- Hệ số tương quan riêng giữa y và  $x_2$  (loại trừ ảnh hưởng của  $x_1$ ):

$$r_{yx_2(x_1)} = \frac{r_{yx_2} - r_{yx_1} \times r_{x_1x_2}}{\sqrt{(1 - r_{yx_1}^2) \cdot (1 - r_{x_1x_2}^2)}} ; \quad (9b)$$

### 3. Phân tích mối liên hệ tương quan giữa hai chỉ tiêu biến động theo thời gian

Mối liên hệ tương quan theo thời gian là mối liên hệ giữa các dãy số biến động theo thời gian; trong đó có một số dãy số biểu hiện biến động của các chỉ tiêu nguyên nhân (sự biến động của nó sẽ ảnh hưởng đến biến động của chỉ tiêu kia) và một dãy số biểu hiện biến động của chỉ tiêu kết quả (sự biến động của nó phụ thuộc vào biến động của các chỉ tiêu nguyên nhân).

Phân tích mối liên hệ tương quan giữa các dãy số theo thời gian chính là xác định mức độ chặt chẽ của mối liên hệ giữa các dãy số. Do đặc điểm nghiên cứu tương quan theo dãy số thời gian là rất phức tạp nên ở đây chỉ trình bày tương quan tuyến tính giữa hai dãy số.

Đặc điểm của dãy số biến động theo thời gian là tồn tại cái gọi là tự tương quan (TTQ). Để kiểm tra các dãy số biến động theo thời gian có đặc điểm này hay không, ta tiến hành tính hệ số liên hệ tương quan tuyến tính giữa các mức độ của dãy số đã cho ( $x_t$  hoặc  $y_t$ ) với mức độ của dãy số đó nhưng lệch đi thời gian 1 năm ( $t=1$ ). Khi nghiên cứu riêng cho từng dãy (đại lượng  $x$  hay  $y$ ) về bản chất đều có công thức tính giống nhau, chỉ khác nhau (hoặc là theo  $x$  hoặc là theo  $y$ ). Từ đây các trường hợp nghiên cứu riêng của từng dãy thống nhất chỉ ký hiệu chung là  $x$ ).

Công thức hệ số TTQ riêng cho từng dãy số chẳng hạn  $x$  như sau:

$$r_{x_t, x_{t+1}} = \frac{\overline{x_t \cdot x_{t+1}} - \bar{x}_t \cdot \bar{x}_{t+1}}{\sigma_t \cdot \sigma_{t+1}} ; \quad (10)$$

Trong đó:

$t$  chỉ thứ tự thời gian theo từng năm;

$x_t, x_{t+1}$  - là mức độ thực tế của dãy thuộc năm  $t$  và của năm sau năm  $t$  ( $t+1$ );

$\sigma_t$  và  $\sigma_{t+1}$  - là các độ lệch chuẩn tương ứng.

$r_{x_t, x_{t+1}}$  là hệ số phản ánh mức độ TTQ. Hệ số này càng gần 1 thì đặc điểm TTQ càng mạnh, và ngược lại càng gần 0 thì đặc điểm TTQ càng yếu.

Khi kiểm tra đặc điểm TTQ của dãy số:

- Nếu thấy đặc điểm này yếu ( $r_{x_t, x_{t+1}}$  gần 0) thì hệ số tương quan tuyến tính giữa hai dãy  $x_t$  và  $y_t$  ( $r_{x,y}$ ) vẫn tính trực tiếp theo các mức độ thực tế ( $x_t$  và  $y_t$ ) như tương quan tuyến tính giữa hai chỉ tiêu biến động theo không gian (xem công thức 2a và 2b đã trình bày ở trên).

- Nếu thấy đặc điểm TTQ của hai dãy số mạnh ( $r_{x_t, x_{t+1}}$  gần +1) thì hệ số tương quan giữa 2 dãy  $x_t$  và  $y_t$  không thể tính trực tiếp theo các mức độ thực tế ( $x_t$  và  $y_t$ ) mà theo các độ lệch giữa mức độ thực tế ( $x_t, y_t$ ) và mức độ lý thuyết tương ứng ( $\hat{x}_t, \hat{y}_t$ ). Công thức tính hệ số tương quan ( $R_{xy}$ ) như sau:

$$R_{xy} = \frac{\sum d_{x_t} \cdot d_{y_t}}{\sqrt{\sum d_{x_t}^2 \cdot \sum d_{y_t}^2}} ; \quad (11)$$

Trong đó:  $d_{x_t}, d_{y_t}$  là các độ lệch giữa mức độ thực tế ( $x_t, y_t$ ) và các mức độ lý thuyết tương ứng ( $\hat{x}_t, \hat{y}_t$ ), tức là  $d_{x_t} = x_t - \hat{x}_t$  và  $d_{y_t} = y_t - \hat{y}_t$ .

Các mức độ lý thuyết  $\hat{x}_t$  và  $\hat{y}_t$  có thể xác định được bằng nhiều phương pháp, nhưng phổ biến và có ý nghĩa nhất là theo

phương trình toán học (phương trình hồi quy).

Trong kinh tế thường dùng một số dạng, phương trình toán học chủ yếu sau đây để điều chỉnh các dãy số:

- Phương trình tuyến tính (bậc nhất):

$$\hat{y} = a_0 + a_1 t \quad ; \quad (12a)$$

- Phương trình parabol bậc hai:

$$\hat{y} = a_0 + a_1 t + a_2 t^2 \quad ; \quad (12b)$$

- Phương trình parabol bậc ba:

$$\hat{y} = a_0 + a_1 t + a_2 t^2 + a_3 t^3 \quad ; \quad (12c)$$

- Phương trình hyperbol:

$$\hat{y} = a_0 + \frac{a_1}{t} \quad ; \quad (12d)$$

- Phương trình hàm số mũ:

$$\hat{y} = a_0 \cdot a_1^t \quad ; \quad (12e)$$

Các hệ số theo từng dạng phương trình 12a, 12b, 12c, 12d và 12e tính được bằng

cách giải các hệ phương trình chuẩn tắc tương ứng được xây dựng theo phương pháp bình phương nhỏ nhất.

Để xác định quy luật phát triển của từng dãy số theo loại phương trình này, trước tiên phải đưa số liệu lên đồ thị để chọn một số loại phương trình nào đó tiến hành điều chỉnh dãy số. Sau đó ứng với mỗi phương trình đã được điều chỉnh chúng ta tính toán các sai số mô tả:

$V_x = \frac{\sigma_x}{\bar{x}}$  và  $V_y = \frac{\sigma_y}{\bar{y}}$  rồi chọn phương trình nào có hệ số mô tả nhỏ nhất.

Dưới đây là ví dụ tính toán hệ số tương quan tuyến tính phản ánh mối liên hệ giữa: mức trang bị vốn (MTBV) cho người lao động và năng suất lao động (NSLĐ) của công nghiệp Việt Nam từ 1990 đến 2003.

BẢNG 2: MỨC TRANG BỊ VỐN VÀ NĂNG SUẤT LAO ĐỘNG CỦA CÔNG NGHIỆP VN

Đơn vị: triệu đồng

Năm	Thứ tự	MTBV	NSLĐ	Năm	Thứ tự	MTBV	NSLĐ
	t	$x_i$	$y_i$		t	$x_i$	$y_i$
A	B	1	2	A	B	1	2
1990	1	25,18	12,97	1997	8	58,97	28,65
1991	2	30,96	15,61	1998	9	64,30	29,96
1992	3	35,44	18,71	1999	10	69,72	30,40
1993	4	41,33	21,69	2000	11	75,30	32,60
1994	5	46,37	24,50	2001	12	83,35	35,21
1995	6	50,45	25,78	2002	13	85,14	35,58
1996	7	53,75	26,84	2003	14	87,28	36,45

Từ số liệu bảng 2 ta lần lượt tính toán như sau:

1. Kiểm tra tính chất TTQ của 2 dãy số trên

Áp dụng công thức 10 ta tính được các hệ số TTQ:

$$\text{Của dãy } x_t: R_{x_t, x_{t+1}} = 0,9965$$

$$\text{Của dãy } y_t: R_{y_t, y_{t+1}} = 0,9942$$

Kết quả tính toán trên chứng tỏ cả 2 dãy số đều có tính chất TTQ rất mạnh.

2. Tiến hành hồi quy hai dãy số về mức NSLĐ và MTBV cho lao động theo các dạng hàm: tuyến tính, hàm bậc hai và hàm số mũ. Kết quả tính toán cho thấy cả hai dãy số

NSLĐ và MTBV của lao động hồi quy theo hàm Parabol bậc hai có hệ số mô tả nhỏ nhất, tức là có hệ số xác định lớn nhất.

Vậy hàm số được lựa chọn để điều chỉnh biến động của hai dãy số như sau:

- Đối với dãy số  $x_t$ :

$$\hat{x}_t = 20,6536 + 4,9791t + 0,0044 t^2; \quad (13a)$$

- Đối với dãy số  $y_t$ :

$$\hat{y}_t = 10,71973 + 2,86166t - 0,0745t^2; \quad (13b)$$

3. Từ các dạng hàm lý thuyết 13a và 13b, lần lượt thay giá trị t nhận từ 1 đến 14 vào tính được các giá trị lý thuyết về MTBV ( $\hat{x}_t$ ) và NSLĐ ( $\hat{y}_t$ ) như số liệu cột 3 và 4 bảng 3:

BẢNG 3: ĐỘ LỆCH GIỮA GIÁ TRỊ THỰC TẾ VÀ LÝ THUYẾT CỦA MTBV VÀ NSLĐ

Đơn vị tính: triệu đồng

Năm	Giá trị thực tế (TT)		Giá trị lý thuyết (LT)		Độ lệch giữa TT và LT	
	MTBV	NSLĐ	MTBV	NSLĐ	MTBV	NSLĐ
	$x_t$	$y_t$	$\hat{x}_t$	$\hat{y}_t$	$d_{x_t}$	$d_{y_t}$
A	1	2	3	4	5=1-3	6=2-4
1990	25,18	12,97	25,6284	13,5069	-0,4460	-0,5391
1991	30,96	15,61	30,5944	16,1450	0,3668	-0,5318
1992	35,44	18,71	35,5517	18,6342	-0,1164	0,0718
1993	41,33	21,69	40,5003	20,9744	0,8344	0,7203
1994	46,37	24,50	45,4402	23,1655	0,9268	1,3301
1995	50,45	25,78	50,3714	25,2077	0,0802	0,5701
1996	53,75	26,84	55,2938	27,1009	-1,5480	-0,2574
1997	58,97	28,65	60,2076	28,8450	-1,2368	-0,1996
1998	64,30	29,96	65,1126	30,4402	-0,8163	-0,4850
1999	69,72	30,40	70,0089	31,8864	-0,2882	-1,4899
2000	75,30	32,60	74,8965	33,1835	0,4010	-0,5811
2001	83,35	35,21	79,7754	34,3317	3,5736	0,8736
2002	85,14	35,58	84,6456	35,3309	0,4912	0,2454
2003	87,28	36,45	89,5071	36,1810	-2,2223	0,2725

Từ số liệu theo giá trị thực tế và giá trị lý thuyết của MTBV và NSLĐ ta tính được các độ lệch tương ứng ở cột 5 và 6 bảng 3.

#### 4. Tính hệ số tương quan giữa NSLĐ và MTBV

Từ số liệu về các giá trị  $d_{xi}$  và  $d_{yi}$  của bảng 3, ta tiếp tục lập bảng 4 để xác định các đại lượng tính hệ số tương quan.

BẢNG 4: XÁC ĐỊNH CÁC ĐẠI LƯỢNG ĐỂ TÍNH HỆ SỐ TƯƠNG QUAN

STT	$d_{xi}$	$d_{yi}$	$d_{xi}^2$	$d_{yi}^2$	$d_{xi} \cdot d_{yi}$
1	-0,4460	-0,5391	0,1989	0,2907	0,2405
2	0,3668	-0,5318	0,1345	0,2828	-0,1950
3	-0,1164	0,0718	0,0135	0,0051	-0,0083
4	0,8344	0,7203	0,6962	0,5189	0,6010
5	0,9268	1,3301	0,8590	1,7692	1,2328
6	0,0802	0,5701	0,0064	0,3250	0,0457
7	-1,5480	-0,2574	2,3965	0,0662	0,3984
8	-1,2368	-0,1996	1,5297	0,0398	0,2468
9	-0,8163	-0,4850	0,6663	0,2352	0,3959
10	-0,2882	-1,4899	0,0831	2,2197	0,4294
11	0,4010	-0,5811	0,1608	0,3377	-0,2330
12	3,5736	0,8736	12,7707	0,7632	3,1219
13	0,4912	0,2454	0,2412	0,0602	0,1205
14	-2,2223	0,2725	4,9384	0,0743	-0,6057
Tổng cộng	x	x	<b>24,6953</b>	<b>6,9879</b>	<b>5,7909</b>

Theo số liệu bảng 4, áp dụng công thức 11 ta tính được hệ số tương quan:

$$R_{xy} = \frac{5,7909}{\sqrt{24,6953 \cdot 6,9879}} = 0,4408$$

Hệ số tương quan bằng 0,4408 chứng tỏ mối quan hệ giữa năng suất lao động và mức trang bị vốn cố định cho lao động của ngành công nghiệp ở mức trung bình ■