

## **Chương 5**

### **HỒI QUY - TƯƠNG QUAN**

Hồi quy và tương quan là phương pháp của toán học được vận dụng trong thống kê để biểu hiện và phân tích mối liên hệ tương quan giữa các hiện tượng kinh tế - xã hội. Vì hai phương pháp này có liên quan chặt chẽ với nhau, nên để cho tiện ta gọi tắt phương pháp này là phương pháp tương quan.

Phương pháp phân tích tương quan được vận dụng khi các hiện tượng hoặc giữa nhiều tiêu thức nghiên cứu, tuy có mối liên hệ với nhau nhưng mối liên hệ này không hoàn toàn chặt chẽ. Một tiêu thức nào đó biến động do ảnh hưởng của một loạt các tiêu thức khác có liên quan, trong đó một số có ảnh hưởng đáng kể và một số khác có ảnh hưởng kém hơn. Tùy theo mục đích nghiên cứu người ta có thể tách riêng ra thành hai (hoặc có thể ba, bốn) tiêu thức đáng chú ý nhất và xác định cụ thể mối liên hệ giữa chúng với nhau, các tiêu thức còn lại được tạm loại trừ coi như không thay đổi. Những tiêu thức được chọn ra để nghiên cứu bao giờ cũng có một tiêu thức kết quả số còn lại là tiêu thức nguyên nhân.

Quá trình phân tích tương quan:

- Dựa trên cơ sở phân tích lý luận, phân tích bản chất của mối liên hệ giữa các hiện tượng (tiêu thức) đồng thời căn cứ vào tài liệu điều tra thực tế, dùng phương pháp phân tổ hay phương pháp đồ thị để xác định sự tồn tại thực tế của mối liên hệ tương quan để xác định tính chất và hình thức của mối liên hệ (xét xem mối liên hệ giữa các tiêu thức nghiên cứu là tuyến tính hay phi tuyến tính, phi tuyến tính dạng nào).

- Biểu hiện cụ thể của mối liên hệ tương quan bằng phương trình hồi quy (dạng phương trình tuyến tính hay phi tuyến tính).

- Tính các tham số của phương trình nói trên và giải thích ý nghĩa của chúng.

Muốn lập phương trình đúng phải căn cứ vào số tiêu thức được chọn, hình thức và chiều hướng của mối liên hệ giữa các tiêu thức này. Đây là vấn đề khá phức tạp và đòi hỏi phải giải quyết chính xác vì có ảnh hưởng trực tiếp đến kết quả nghiên cứu.

- Đánh giá trình độ chặt chẽ của mối liên hệ tương quan bằng các chỉ tiêu: hệ số tương quan, tỷ số tương quan...đây là nhiệm vụ quan trọng của phân tích tương quan vì căn cứ vào các hệ số này có thể kết luận về tầm quan trọng của mối liên hệ, đánh giá đúng đắn vai trò của từng nguyên nhân, trên cơ sở đó có căn cứ để xây dựng các mô hình sản xuất và tiêu dùng, xây dựng các định mức kinh tế - kỹ thuật và dự đoán mức độ của hiện tượng trong tương lai.

Chúng ta sẽ vận dụng cụ thể phương pháp này trong một số trường hợp tiêu biểu nhất.

## **1.NHỮNG VẤN ĐỀ CHUNG VỀ PHƯƠNG PHÁP HỒI QUY – TƯƠNG QUAN**

### **1.1. Khái niệm hồi quy – tương quan**

Hồi quy – tương quan là phương pháp phân tích dựa trên mối liên hệ phụ thuộc của một biến kết quả (biến phụ thuộc) vào một hay nhiều biến nguyên nhân (biến độc lập), mối liên hệ phụ thuộc này được xây dựng dựa trên phương trình hồi quy có thể là tuyến tính hay phi tuyến tính, trên cơ sở phương trình hồi quy có thể ước lượng và giải thích được sự biến động của biến phụ thuộc dựa vào sự biến động của các biến độc lập.

### **1.2. Liên hệ hàm số và liên hệ tương quan**

Theo quan điểm của duy vật biện chứng thì các hiện tượng tồn tại trong mối liên hệ phổ biến và nhiều vẻ, giữa chúng có mối quan hệ mật thiết với nhau, tác động qua lại lẫn nhau. Không có một hiện tượng nào lại phát sinh, phát triển một cách cô lập, tách rời các hiện tượng khác. Vì vậy, việc nghiên cứu mối liên hệ là một trong những nhiệm vụ quan trọng của thống kê.

Khi nghiên cứu mối liên hệ, nếu xét theo trình độ chặt chẽ thì có thể phân thành hai loại là liên hệ hàm số và liên hệ tương quan.

**Liên hệ hàm số:** Là mối liên hệ hoàn toàn chặt chẽ và được biểu hiện dưới dạng một hàm số, ví dụ  $y = f(x)$ . Điều đó có nghĩa là khi đại lượng  $x$  biến đổi thì theo một quy tắc nào đó, có thể xác định được giá trị tương ứng của đại lượng  $y$ . Ví dụ như mối liên hệ giữa chu vi đường tròn với bán kính của nó:  $C = 2\pi R$ .

**Liên hệ tương quan:** Là mối liên hệ không hoàn toàn chặt chẽ và được biểu hiện ở chỗ khi một hiện tượng biến đổi thì làm cho hiện tượng có liên quan biến đổi theo nhưng nó không có ảnh hưởng hoàn toàn quyết định đến sự biến đổi này. Đây là mối liên hệ mà với một giá trị của tiêu thức nguyên nhân sẽ tương ứng với nhiều giá trị của tiêu thức kết quả.

Ví dụ: Thu nhập của hộ nông dân thay đổi thì có thể làm cho đầu tư thay đổi, thu nhập không phải là nhân tố duy nhất và hoàn toàn quyết định sự thay đổi của đầu tư. Hơn nữa ứng với một sự thay đổi của giá bán thì không phải lúc nào doanh thu cũng thay đổi theo một tỷ lệ nhất định mà có thể có nhiều giá trị tương ứng. Đây chính là liên hệ không hoàn toàn chặt chẽ.

### 1.3. Một số dạng liên hệ

Các hiện tượng kinh tế xã hội luôn có mối liên hệ và tác động qua lại lẫn nhau, các mối liên hệ này có thể được biểu hiện dưới dạng tuyến tính (đường thẳng), hay liên hệ tương quan phi tuyến tính (đường cong).

Dựa vào phân tích đặc điểm của mỗi liên hệ hay từ số liệu được điều tra thực tế mà xây dựng các phương trình biểu diễn mỗi liên hệ một cách phù hợp nhất.

### 1.4. Nhiệm vụ của phương pháp hồi quy – tương quan

- Xây dựng phương trình hồi quy và giải thích ý nghĩa các tham số trong phương trình hồi quy. Từ phân tích định hướng dựa trên mối liên hệ giữa các hiện tượng kinh tế – xã hội để tìm ra các mối liên hệ tức là xác định tiêu thức kết quả và các tiêu thức nguyên nhân để từ đó xác định phương trình hồi quy để biểu diễn mối liên hệ giữa các tiêu thức. Từ đó giải thích ý nghĩa kinh tế các tham số trong phương trình hồi quy.

- Đánh giá chiều hướng và cường độ của mối liên hệ tương quan. Vì nó không hoàn toàn chặt chẽ nên khi tìm được mối liên hệ và xây dựng được phương trình hồi quy cần đánh giá trình độ chặt chẽ và chiều hướng của mối liên hệ cũng như đánh giá xem tiêu thức nguyên nhân giải thích được bao nhiêu phần sự biến động của tiêu thức kết quả.

## **2. HỒI QUY TƯƠNG QUAN ĐƠN**

### **2.1. Phương trình hồi quy tổng thể chung và phương trình hồi quy tổng thể mẫu**

Như ta đã biết giữa tiêu thức nguyên nhân và tiêu thức kết quả có thể tồn tại mối liên hệ tuyến tính hay phi tuyến tính. Trong phạm vi của phần này chúng ta xem xét mối liên hệ tương quan tuyến tính đơn. Tức là mối liên hệ giữa một biến nguyên nhân  $X$  và một biến kết quả  $Y$  để xem có tồn tại mối liên hệ tương quan bằng phương trình tuyến tính – phương trình hệ bậc nhất theo biến số

#### ***2.1.1. Phương trình tổng thể chung***

Là phương trình được xây dựng dựa trên kết quả nghiên cứu tổng thể. Do các số liệu thường được thu thập trên số liệu điều tra mẫu, cho nên chúng ta chỉ nghiên cứu phương trình tổng thể mẫu.

#### ***2.1.2. Phương trình tổng thể mẫu***

Hai tiêu thức có mối liên hệ tương quan tuyến tính với nhau trong trường hợp tiêu thức nguyên nhân biến thiên với những lượng tương đối đều đặn. Quan sát trên đồ thị đường hồi quy thực nghiệm vượt theo một chiều hướng rõ rệt.

Ta xét thí dụ sau: Giả sử có tài liệu về tuổi nghề (năm) và năng suất lao động (sản phẩm) của 10 công nhân tại một xí nghiệp như sau:

| Tên công nhân | Tuổi nghề (năm) | Năng suất lao động<br>(sản phẩm) |
|---------------|-----------------|----------------------------------|
| A             | 1               | 3                                |
| B             | 3               | 12                               |
| C             | 4               | 9                                |
| D             | 5               | 16                               |
| E             | 7               | 12                               |
| G             | 8               | 21                               |
| H             | 9               | 21                               |
| I             | 10              | 24                               |
| K             | 11              | 19                               |
| L             | 12              | 27                               |

Tài liệu trên cho thấy giữa tuổi nghề của công nhân và năng suất lao động của họ có mối liên hệ với nhau: Nhìn chung, với sự tăng lên của tuổi nghề (tiêu thức nguyên nhân) thì năng suất lao động (tiêu thức kết quả) cũng tăng lên. Song mối liên hệ này không hoàn toàn chặt chẽ, tức là không phải cứ tuổi nghề tăng lên thì năng suất lao động cũng tăng theo một cách tương ứng. Hay nói cách khác là mối liên hệ giữa tuổi nghề và năng suất lao động là mối liên hệ tương quan.

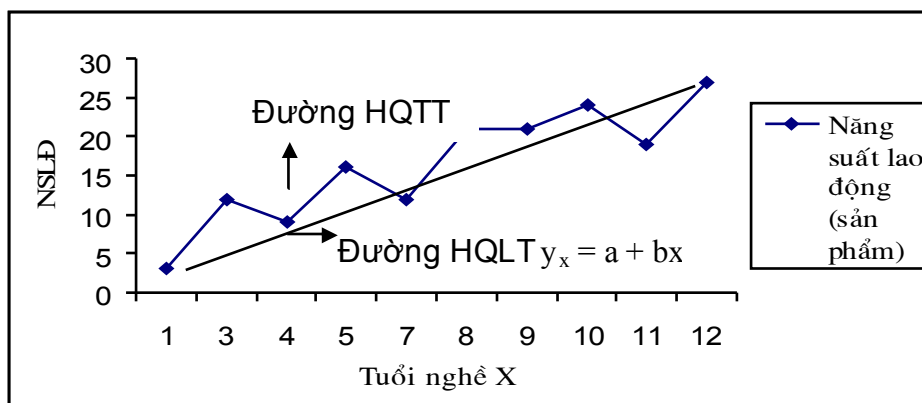
Để thấy rõ chiều hướng của mối liên hệ này, có thể dùng đồ thị với trục hoành là tuổi nghề (x) và trục tung là năng suất lao động (y). Trong mặt phẳng hệ trục tọa độ có 10 chấm nói lên tuổi nghề và năng suất lao động của 10 công nhân ở trên. Nối các chấm lại với nhau sẽ được một đường gấp khúc được gọi là đường hồi quy thực tế. Căn cứ vào hình dáng của đường hồi quy thực tế, người ta điều tiết đường hồi quy lý thuyết, được biểu hiện dưới dạng một hàm số. Ở đây ta có thể thay thế đường hồi quy thực tế bằng đường hồi quy lý thuyết là một phương trình đường thẳng:

$$\bar{y}_x = a + bx(1)$$

Trong đó x: Trị số của tiêu thức nguyên nhân (tuổi nghề)

$\bar{y}_x$  : Trị số điều chỉnh của tiêu thức kết quả (NSLĐ)

a, b: Các tham số



Phương trình (1) được gọi là phương trình hồi quy. Sau khi đã xác định được dạng của phương trình hồi quy, dựa vào tài liệu thực tế để xác định giá trị cụ thể của các tham số. Thường thường các tham số của phương trình hồi quy được tính toán bằng phương pháp bình phương nhỏ nhất.

## 2.2. Phương pháp bình phương nhỏ nhất

Để xác định các hệ số hồi quy trong phương trình hồi quy tổng thể mẫu là phương pháp bình phương nhỏ nhất. Tức là các tham số được tính toán trên cơ sở tổng bình phương các độ lệch giữa trị số thực tế và trị số lý thuyết đạt cực tiểu, tức là:

Tìm giá trị nhỏ nhất của chênh lệch giữa giá trị thực tế và giá trị từ phương trình hồi quy lý thuyết

$$S = \sum (y - \bar{y}_x)^2 = \min$$

$$\text{Hay } S = \sum (y - a - bx)^2 = \min$$

Muốn vậy, đạo hàm riêng của các tham số phải triệt tiêu, tức là ta có hệ phương trình sau đây - được gọi là hệ phương trình chuẩn:

$$\begin{cases} \frac{\partial S}{\partial a} = 0 \\ \frac{\partial S}{\partial b} = 0 \end{cases}$$

Giải hệ phương trình này, ta được:

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

Để giải hệ phương trình ta lập bảng

| Tuổi nghề (x) | NSLĐ (y) | xy   | x <sup>2</sup> | y <sup>2</sup> |
|---------------|----------|------|----------------|----------------|
| 1             | 3        | 3    | 1              | 9              |
| 3             | 12       | 36   | 9              | 144            |
| 4             | 9        | 36   | 16             | 81             |
| 5             | 16       | 80   | 25             | 256            |
| 7             | 12       | 84   | 49             | 144            |
| 8             | 21       | 168  | 64             | 441            |
| 9             | 21       | 189  | 81             | 441            |
| 10            | 24       | 240  | 100            | 576            |
| 11            | 19       | 209  | 121            | 361            |
| 12            | 27       | 324  | 144            | 729            |
| 70            | 164      | 1369 | 610            | 3182           |

Thay các số liệu tính được trong bảng vào hệ phương trình(2):

$$\begin{cases} 164 = 10a + 70b \\ 1369 = 70a + 610b \end{cases}$$

Giải ra ta được: a = 3,508 ; b = 1,84

Do đó dạng cụ thể của (1) là:

$$\bar{y}_x = 3,508 + 1,84x$$

Đây là phương trình hồi quy phản ánh mối liên hệ giữa tuổi nghề (x) và năng suất lao động (y) theo tài liệu từ 10 công nhân đã nói ở trên. Trong phương trình này giá trị của a = 3,508 có thể được xem như mức độ ảnh hưởng của các nguyên nhân khác đến năng suất lao động, ngoài tuổi nghề; b = 1,84 nói lên khi tăng thêm 1 năm tuổi nghề thì năng suất lao động tăng bình quân là 1,84 sản phẩm.

Từ hệ phương trình (2) bằng phép biến đổi đơn giản ta được công

thức sau đây:

$$b = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x^2}$$

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

### 2.3. Ý nghĩa các hệ số trong phương trình hồi quy

- a là hệ số tự do (hệ số chặn) thể hiện giá trị của y khi  $x=0$ , tuy nhiên sự giải thích này chỉ có ý nghĩa nếu như trong tổng thể x có nhận giá trị 0, còn nếu x không nhận giá trị 0 thì điều giải thích này không có ý nghĩa. Hoặc x nhận giá trị 0 thì hệ số tự do không có ý nghĩa và khi đó chỉ có thể coi là ảnh hưởng trung bình của tất cả các tiêu thức nguyên nhân (biến nguyên nhân) khác ngoài tiêu thức x trong phương trình tới tiêu thức kết quả y (biến phụ thuộc).

- b là hệ số hồi quy nói lên ảnh hưởng trực tiếp của tiêu thức nguyên nhân x tới tiêu thức kết quả y. Cụ thể, khi x thay đổi một đơn vị thì làm cho y thay đổi trung bình là b đơn vị. Giá trị b có thể âm, dương hay bằng 0, tùy thuộc vào mối liên hệ giữa x và y.

### 2.4. Hệ số tương quan (r)

Hệ số tương quan là chỉ tiêu tương đối dùng để đánh giá trình độ chặt chẽ của các mối liên hệ tương quan tuyến tính giữa hai tiêu thức.

Hệ số tương quan có tác dụng quan trọng đối với lý luận và công tác thực tế, chỉ tiêu này giúp ta xác định được cường độ của mối liên hệ, xem xét giữa tiêu thức nguyên nhân và tiêu thức kết quả có mối liên hệ với nhau đến chừng mực nào đó. Trong những điều kiện khác nhau mối liên hệ tương quan giữa cùng một số hiện tượng hay tiêu thức cũng có thể có trình độ chặt chẽ khác nhau. Qua việc đánh giá này giúp ta chọn ra những tiêu thức nguyên nhân có tác dụng chủ yếu hay thứ yếu đến hiện tượng nghiên cứu.

Hệ số tương quan cũng giúp ta xác định phương hướng cụ thể của mối liên hệ, nếu chỉ tiêu này mang dấu dương thì đó là mối liên hệ tương quan thuận, nếu chỉ tiêu này mang dấu âm, thì đó là mối liên hệ tương quan nghịch.

- Với tài liệu không phân tổ, hệ số tương quan có thể được tính theo công thức sau đây:

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}} \quad (1)$$



Từ công thức (1) bằng những phép biến đổi đơn giản có thể có những công thức khác nhau để tính hệ số tương quan. Sau đây là một số công thức thường được sử dụng:

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y} \quad (2)$$

$$r = b \cdot \frac{\sigma_x}{\sigma_y} \quad (3)$$

$$r = \frac{n \sum xy - (\sum x) \cdot (\sum y)}{\sqrt{[n \sum x^2 - (\sum x)^2] \cdot [n \sum y^2 - (\sum y)^2]}} \quad (4)$$

Ví dụ:

- Tính theo công thức (2) với:

$$\overline{xy} = \frac{\sum xy}{n} = \frac{1369}{10} = 136,9$$

$$\bar{x} = \frac{\sum x}{n} = \frac{70}{10} = 7$$

$$\bar{y} = \frac{\sum y}{n} = \frac{164}{10} = 16,4$$

$$\sigma_x = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2} = \sqrt{\frac{610}{10} - \left(\frac{70}{10}\right)^2} = 3,464$$

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

Do đó:

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

- Tính theo công thức (3):

$$r = 1,84 \times \frac{3,464}{7,017} = 0,909$$

- Tính theo công thức (4):

$$r = \frac{10 \times 1369 - 70 \times 164}{\sqrt{[10 \times 610 - (70)^2] \cdot [10 \times 3182 - (164)^2]}} = 0,909$$

- Với tài liệu phân tổ thì  $r = \frac{\sum (x - \bar{x}) \cdot (y - \bar{y}) n_{xy}}{\sqrt{\sum (x - \bar{x})^2 n_x \cdot \sum (y - \bar{y})^2 n_y}}$

### Tính chất của hệ số tương quan:

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

Khi  $r = 1$  (hoặc  $r = -1$ ) thì giữa  $x$  và  $y$  có liên hệ hàm số.

Khi  $r$  càng gần 1 (hoặc -1) thì liên hệ tương quan càng chặt chẽ.

Khi  $r = 0$  thì giữa  $x$  và  $y$  không có liên hệ tuyến tính.

Tính chất hệ số tương quan được mô tả như sau:

| Càng chặt chẽ         |                      | Càng chặt chẽ        |
|-----------------------|----------------------|----------------------|
| -1                    | 0                    | +1                   |
| Liên hệ hàm số        | Không có mối liên hệ | Liên hệ hàm số       |
| <i>Liên hệ nghịch</i> |                      | <i>Liên hệ thuận</i> |

Cụ thể nếu  $|r| < 0,3$  thì mối liên hệ giữa  $x$  và  $y$  rất lỏng lẻo (kém ý nghĩa)

$0,3 \leq |r| < 0,5$  thì mối liên hệ giữa  $x$  và  $y$  lỏng lẻo (ý nghĩa hạn chế)

$0,5 \leq |r| < 0,7$  thì mối liên hệ giữa  $x$  và  $y$  là khá chặt chẽ

$0,7 \leq |r| < 0,9$  thì mối liên hệ giữa  $x$  và  $y$  là chặt chẽ

$0,9 \leq |r| < 1$  thì mối liên hệ giữa  $x$  và  $y$  là vô cùng (rất) chặt chẽ

Theo kết quả tính toán ở trên  $r = 0,909$  cho phép kết luận mối liên hệ giữa tuổi nghề và năng suất lao động là mối liên hệ thuận và rất chặt chẽ.

### 3. Hồi quy tương quan phi tuyến tính

Phương trình biểu diễn mối liên hệ giữa các hiện tượng kinh tế - xã hội được trình bày ở trên đó là các phương trình tuyến tính tức là phương trình bậc nhất. Tuy nhiên trong thực tế các hiện tượng thường không tuân theo mối liên hệ tuyến tính mà theo một đường cong, liên hệ phi tuyến, chẳng hạn như khi tuổi của công nhân sản xuất ban đầu có thể liên hệ đồng biến, tuyến tính đến năng suất lao động (NSLĐ) nhưng theo thời gian thì mối quan hệ này không còn là tuyến tính và đồng biến nữa mà sẽ là liên hệ nghịch biến, do vậy mối liên hệ của tuổi công nhân và NSLĐ theo thời gian có thể theo đường cong parabol. Hoặc khi sản xuất số lượng sản phẩm nhiều thì làm giá thành giảm xuống tuy nhiên sự giảm của giá thành không thể theo một đường tuyến tính vì số lượng sản phẩm chỉ là một nhân tố tác động đến giá thành nên sản xuất đến một số lượng nào đó thì giá thành sẽ giảm nhưng tốc độ giảm này sẽ giảm dần dần do đó mối liên hệ này tốt nhất là theo đường Hypebol... Nội dung cơ bản của việc xây dựng phương trình và ước lượng các tham số cơ bản tương tự như với phương trình hồi quy tuyến tính.

## BÀI TẬP CHƯƠNG 5

### Bài 1:

Có số liệu như sau:

|                                    |         |         |         |         |         |         |         |         |          |          |
|------------------------------------|---------|---------|---------|---------|---------|---------|---------|---------|----------|----------|
| Chi phí<br>quảng cáo<br>(1000 USD) | 21      | 27      | 38      | 48      | 57      | 60      | 68      | 72      | 75       | 85       |
| Doanh Thu<br>(1000 USD)            | 22<br>0 | 32<br>0 | 40<br>0 | 52<br>0 | 70<br>0 | 70<br>0 | 75<br>0 | 90<br>0 | 100<br>0 | 111<br>0 |

- Xác định phương trình hồi qui biểu diễn mối quan hệ giữa chi phí quảng cáo và doanh thu.
- Nêu ý nghĩa các tham số trong phương trình.

Bài 2 :

Có số liệu như sau:

|                           |    |    |    |    |    |    |    |    |    |    |
|---------------------------|----|----|----|----|----|----|----|----|----|----|
| Cầu về mặt hàng A (kg)    | 20 | 18 | 18 | 17 | 15 | 14 | 13 | 13 | 12 | 8  |
| Giá mặt hàng A (1000đ/kg) | 30 | 35 | 36 | 38 | 40 | 42 | 45 | 48 | 50 | 60 |

- Xác định phương trình hồi qui biểu diễn mối liên hệ giữa 2 tiêu thức trên. Nêu ý nghĩa các tham số trong phương trình.
- Đánh giá trình độ chặt chẽ của mối liên hệ.

Bài 3:

Có số liệu sau :

Đ/v tính : cm

|                   |     |     |     |     |     |     |     |     |     |     |
|-------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Chiều cao của bố  | 158 | 160 | 163 | 165 | 167 | 170 | 167 | 172 | 177 | 181 |
| Chiều cao của con | 163 | 158 | 167 | 170 | 165 | 180 | 170 | 175 | 180 | 175 |

- Xác định phương trình hồi qui biểu hiện mối liên hệ giữa chiều cao của bố và chiều cao của con. Nêu ý nghĩa của các tham số trong phương trình.
- Đánh giá trình độ chặt chẽ của mối liên hệ.

## Chương 6

### DÃY SỐ THỜI GIAN VÀ DỰ ĐOÁN

#### 1. KHÁI NIỆM VÀ CÁC THÀNH PHẦN CỦA DÃY SỐ THỜI GIAN

##### 1.1. Khái niệm

Mặt lượng của hiện tượng thường xuyên biến động qua thời gian. Trong thống kê, để nghiên cứu sự biến động này, người ta thường dựa vào dãy số thời gian.

*Dãy số thời gian trong thống kê là dãy các trị số của một chỉ tiêu thống kê được sắp xếp theo thứ tự thời gian, dùng để phản ánh quá trình phát triển của hiện tượng.*

**Ví dụ:** Có sản lượng điện Việt Nam (tỷ kw/h) từ năm N đến năm N+7 như sau: 14,7; 17,0; 19,3; 21,7; 23,6; 26,6; 30,7; 35,6

Một dãy số thời gian có dạng chung nhất :

|                     |       |       |       |     |           |       |
|---------------------|-------|-------|-------|-----|-----------|-------|
| Thời gian ( $t_i$ ) | $t_1$ | $t_2$ | $t_3$ | ... | $t_{n-1}$ | $t_n$ |
| Mức độ ( $y_i$ )    | $y_1$ | $y_2$ | $y_3$ | ... | $y_{n-1}$ | $y_n$ |

Trong đó:

$t_i$  : thời gian thứ i

$y_i$  : mức độ thứ i tương ứng với thời gian thứ i

Qua dãy số thời gian có thể nghiên cứu các đặc điểm về sự biến động của hiện tượng, vạch rõ xu hướng và tính quy luật của sự phát triển, đồng thời để dự đoán các mức độ của hiện tượng trong tương lai.

Trong dãy số biến động theo thời gian có hai yếu tố:

- **Yếu tố thời gian:** Chỉ rõ thời gian mà dãy số phản ánh và ít nhất phải có 2 số hạng thời gian được sắp xếp theo thứ tự và có thể là ngày, tháng, năm,...tùy mục đích nghiên cứu.

- **Yếu tố chỉ tiêu:** Nêu rõ mặt lượng của hiện tượng nghiên cứu có thể biểu hiện bằng số tuyệt đối, số tương đối hay số bình quân, trị số của chỉ tiêu nghiên cứu được gọi là mức độ của dãy số thời gian. Trong phạm vi giáo trình này, chúng ta chỉ đi sâu nghiên cứu về dãy số tuyệt đối

Dãy số tuyệt đối dựa vào đặc điểm của các mức độ (phản ánh quy mô, khối lượng của hiện tượng qua thời gian), có thể phân thành hai loại: **Dãy số thời kỳ** và **dãy số thời điểm**.

### **1.1.1. *Dãy số biến động theo thời kỳ (gọi tắt là dãy số thời kỳ).***

Là dãy số mà các mức độ của nó là những số tuyệt đối thời kỳ, phản ánh quy mô (khối lượng) của hiện tượng được tích lũy trong những khoảng thời gian nhất định.

**Ví dụ:** Có tài liệu về mức doanh của một số năm tại DN X như sau:

| Năm                       | N    | N +1 | N +2 | N +3 | N +4 |
|---------------------------|------|------|------|------|------|
| Doanh thu<br>(triệu đồng) | 4000 | 4600 | 5000 | 5500 | 6200 |

Dãy số thời kỳ thường dùng để tổng kết thành quả hoạt động sản xuất nghiệp vụ của đơn vị trong một thời gian nhất định và nêu lên xu hướng phát triển của hiện tượng. Khoảng cách thời gian trong dãy số càng dài thì trị số càng lớn. Do vậy ta có thể cộng các trị số này lại với nhau để phản ánh mức độ của hiện tượng trong thời kỳ dài hơn.

### **1.1.2. *Dãy số biến động theo thời điểm (gọi tắt là dãy số thời điểm).***

Là dãy số mà trong đó các mức độ của chỉ tiêu là những số tuyệt đối thời điểm, phản ánh quy mô (khối lượng) của hiện tượng tại những thời điểm nhất định.

**Ví dụ:** Có tài liệu về vốn lưu động còn lại đầu các quý (triệu đồng) của doanh nghiệp X năm N như sau:

| Thời điểm    | 1/1 | 1/4 | 1/7 | 1/10 | 31/12 |
|--------------|-----|-----|-----|------|-------|
| Vốn lưu động | 920 | 900 | 880 | 890  | 850   |

Trị số trong dãy số thời điểm đặc trưng cho trạng thái hiện tượng vào một thời điểm nhất định, vì vậy không thể cộng các mức được với nhau vì các con số cộng sẽ không có ý nghĩa kinh tế.

Yêu cầu cơ bản khi xây dựng một dãy số thời gian phải đảm bảo tính chất có thể so sánh được giữa các mức độ trong dãy số. Muốn vậy thì nội dung và phương pháp tính toán chỉ tiêu qua thời gian phải thống nhất, phạm vi của hiện tượng nghiên cứu trước sau phải nhất trí, các khoảng cách thời gian trong dãy số nên bằng nhau (nhất là đối với dãy số thời kỳ).

## 1.2. Các thành phần của dãy số thời gian

Thông thường một dãy số thời gian có 4 thành phần:

- **Xu thế:** Phản ánh xu hướng biến động cơ bản của hiện tượng qua thời gian. Mặc dù các mức độ của hiện tượng có thể dao động thay đổi tại các mốc thời gian khác nhau nhưng nếu quan sát hiện tượng trong cả một quá trình dài (thường trên một năm) ta sẽ thấy hiện tượng có một xu hướng biến động cơ bản hoặc là tăng hoặc là giảm

- **Biến động chu kỳ:** Phản ánh quy luật lặp lại của dãy số trong những khoảng thời gian nhất định thường là vài năm. Chu kỳ tác động lên dãy số thời gian thường là chu kỳ kinh tế hay kinh doanh như chu kỳ suy thoái kinh tế, lạm phát, hay vòng đời của sản phẩm. Nghiên cứu biến động chu kỳ cũng đòi hỏi dãy số thời gian có độ dài trên một năm.

- **Biến động thời vụ:** hay mùa vụ là những biến động của hiện tượng có tính chất lặp đi lặp lại trong từng thời gian nhất định của năm. Nguyên nhân gây ra biến động thời vụ là do ảnh hưởng của điều kiện tự nhiên như: thời tiết, khí hậu, hoặc phong tục tập quán sinh hoạt của dân cư. Ví dụ lượng khách du lịch biến động vào mùa hè tăng lên, mùa thu giảm xuống.

- **Biến động ngẫu nhiên:** Là do các yếu tố ngẫu nhiên gây ra, và thường rất khó dự đoán do tính chất bất thường của nó. Bên cạnh đó sự tồn tại của biến động ngẫu nhiên còn làm lu mờ tính quy luật của các thành phần còn lại trong dãy số khiến cho việc dự đoán các thành phần này khó khăn hơn. Biến động này đòi hỏi số liệu phải được thu thập trong những khoảng thời gian dài. Với hạn chế về mặt số liệu nên trong thực hành phân tích dãy số thời gian, người ta thường giả sử dãy số không có thành phần chu kỳ. Khi đó một dãy số thời gian chỉ bao gồm ba thành phần cơ bản là xu thế, biến động thời vụ, và biến động ngẫu nhiên. Chúng kết hợp theo nhiều dạng khác nhau trong đó có hai dạng phổ biến là: Mô hình kết hợp theo dạng cộng, và dạng nhân ba loại với nhau để tạo ra mức độ của dãy số. Nếu biên độ dao động đều đặn qua thời gian thì áp dụng mô hình cộng, còn nếu mạnh thì áp dụng mô hình nhân.

## 2. CÁC CHỈ TIÊU PHÂN TÍCH ĐẶC ĐIỂM BIẾN ĐỘNG CỦA HIỆN TƯỢNG QUA THỜI GIAN

### 2.1. Mức độ bình quân theo thời gian

Mức độ bình quân theo thời gian là số bình quân về các mức độ của chỉ tiêu trong dãy số thời gian, biểu hiện mức độ điển hình của hiện tượng nghiên cứu trong một khoảng thời gian dài. Tùy theo dãy số thời kỳ hoặc dãy số thời điểm mà có các công thức tính khác nhau.

#### 2.1.1. Đối với dãy số thời kỳ

Mức độ bình quân theo thời gian được tính theo công thức sau đây:

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$$

Trong đó

$\bar{y}$  - Mức độ bình quân theo thời gian

$y_i$  ( $i = \overline{1, n}$ ) - Các mức độ của chỉ tiêu trong dãy số thời gian

$n$  - Số mức độ trong dãy số

*Ví dụ:* Từ số liệu về doanh thu của doanh nghiệp X, tính được

Doanh thu bình quân bình quân qua 5 năm như sau:

$$\frac{4000 + 4600 + 5000 + 5500 + 6200}{5} = 5060 \text{ (trđ)}$$

#### 2.1.2. Đối với dãy số thời điểm

Tùy theo đặc điểm biến động của dãy số và nguồn số liệu, chỉ tiêu này được tính theo các cách sau:

##### 2.1.2.1. Dãy số thời điểm với khoảng cách thời gian bằng nhau

Theo dãy số thời điểm ở trên ta nhận thấy mỗi mức độ của dãy số chỉ phản ánh số vốn lưu động vào các ngày đầu của các quý. Trên thực tế các giá trị này luôn thay đổi, vì vậy việc tính số bình quân theo thời gian phải dựa trên cơ sở giả thiết rằng, giữa hai thời điểm điều tra mức độ của hiện tượng tăng, giảm tương đối đều đặn.



Trở lại ví dụ phần 1.1.2. Như vậy trước hết ta có thể tính được số vốn lưu động bình quân của từng quý như sau:

$$+ \text{Quý 1: } \bar{y}_1 = \frac{920+900}{2} = 910 \text{ (triệu đồng)}$$

$$+ \text{Quý 2: } \bar{y}_2 = \frac{900+880}{2} = 890 \text{ (triệu đồng)}$$

$$+ \text{Quý 3: } \bar{y}_3 = \frac{880+890}{2} = 885 \text{ (triệu đồng)}$$

$$+ \text{Quý 4: } \bar{y}_4 = \frac{890+850}{2} = 870 \text{ (triệu đồng)}$$

Số vốn lưu động bình quân của doanh nghiệp X năm 2005 là

$$\begin{aligned} \bar{y} &= \frac{\bar{y}_1 + \bar{y}_2 + \bar{y}_3 + \bar{y}_4}{4} = \frac{\frac{920+900}{2} + \dots + \frac{890+850}{2}}{5} \\ &= \frac{\frac{920}{2} + 900 + \dots + \frac{850}{2}}{5} = \frac{3555}{4} = 888,75 \text{ (triệu đồng)} \end{aligned}$$

Từ đó khái quát thành công thức sau đây để tính mức độ bình quân theo thời gian từ một dãy số thời điểm có khoảng cách thời gian bằng nhau:

$$\bar{y} = \frac{\frac{y_1}{2} + y_2 + \dots + \frac{y_n}{2}}{n-1}$$

Trong đó

$y_i$  ( $i = 1, 2, \dots, n$ ) là các mức độ của dãy số thời điểm.

$n$  - Số thời điểm trong dãy số

#### 2.1.2.2. Dãy số thời điểm có khoảng cách thời gian không đều nhau

Trường hợp này phải lấy thời gian trong mỗi khoảng cách làm quyền số. Khi đó, công thức là:

$$\bar{y} = \frac{\sum_{i=1}^n y_i t_i}{\sum_{i=1}^n t_i}$$

Trong đó :  $t_i$  là độ dài thời gian có mức độ  $y_i$

**Ví dụ:** Có tài liệu về số công nhân của một xí nghiệp trong tháng 4- N như sau:

Ngày 1 - 4 có 400 công nhân.

Ngày 10 - 4 nhận thêm 5 công nhân.

Ngày 15 - 4 nhận thêm 3 công nhân.

Ngày 21 - 4 cho thôi việc 2 công nhân và từ đó đến hết tháng 4 không thay đổi. Từ đó ta lập bảng sau:

| Thời gian            | Số ngày ( $t_i$ ) | Số công nhân ( $y_i$ ) | $y_i t_i$ |
|----------------------|-------------------|------------------------|-----------|
| Từ 1 - 4 đến 9 - 4   | 9                 | 400                    | 3600      |
| Từ 10 - 4 đến 14 - 4 | 5                 | 405                    | 2025      |
| Từ 15 - 4 đến 20 - 4 | 6                 | 408                    | 2448      |
| Từ 21 - 4 đến 30 - 4 | 10                | 406                    | 4060      |
| Cộng                 | 30                |                        | 12133     |

Áp dụng công thức trên ta tính được số công nhân trung bình trong tháng 4 là:

$$\bar{y} = \frac{12133}{30} = 404 \text{ (người)}$$

## 2.2. Lượng tăng (hoặc giảm) tuyệt đối

Lượng tăng (hoặc giảm) tuyệt đối là hiệu số giữa hai mức độ của chỉ tiêu trong dãy số thời gian, phản ánh sự thay đổi tuyệt đối của mức độ hiện tượng qua hai thời gian khác nhau. Nếu hướng phát triển của hiện tượng tăng thì lượng tăng tuyệt đối mang dấu dương và ngược lại. Tùy theo mục đích nghiên cứu có thể tính các lượng tăng (hoặc giảm) tuyệt đối sau đây:

### **2.2.1. Lượng tăng (hoặc giảm) tuyệt đối liên hoàn (hay từng kỳ)**

Đó là hiệu số của một mức độ bất kỳ trong dãy số ở kỳ nghiên cứu với mức độ của kỳ liền trước nó. Công thức tính như sau:

$$\bar{\delta}_i = y_i - y_{i-1}$$

Trong đó:

$\bar{\delta}_i$  - Lượng tăng (hoặc giảm) tuyệt đối liên hoàn

$y_i$  - Mức độ của chỉ tiêu trong dãy số kỳ nghiên cứu

$y_{i-1}$  - Mức độ ở kỳ liền trước mức độ kỳ nghiên cứu

### **2.2.2. Lượng tăng (hoặc giảm) tuyệt đối định gốc (hay tính dồn)**

Đó là hiệu số giữa mức độ bất kỳ ở kỳ nghiên cứu trong dãy số với mức độ được chọn làm gốc không thay đổi (thường là mức độ đầu tiên trong dãy số). Công thức tính:

$$\Delta_i = y_i - y_1$$

Trong đó:

$\Delta_i$  - Lượng tăng (hoặc giảm) tuyệt đối định gốc

$y_1$  - Mức độ của chỉ tiêu ở kỳ được chọn làm gốc so sánh

Dễ dàng nhận thấy:  $\sum_{i=2}^n \bar{\delta}_i = \Delta_n$  Tức là tổng các lượng tăng (hoặc giảm) tuyệt đối liên hoàn bằng lượng tăng (hoặc giảm) tuyệt đối định gốc.

### **2.2.3. Lượng tăng (hoặc giảm) tuyệt đối bình quân**

Đó là số bình quân của các lượng tăng (hoặc giảm) tuyệt đối từng kỳ. Công thức tính:

$$\bar{\bar{\delta}} = \frac{\sum_{i=2}^n \bar{\delta}_i}{n-1} = \frac{\Delta_n}{n-1} = \frac{y_n - y_1}{n-1}$$

Trong đó:  $\bar{\bar{\delta}}$  - Lượng tăng (hoặc giảm) tuyệt đối bình quân.

Vẫn lấy ví dụ trên ta có:

*Đơn vị (triệu đồng)*

| Năm                  | N    | N+1  | N+2  | N+3  | N+4  |
|----------------------|------|------|------|------|------|
| Doanh thu            | 4000 | 4600 | 5000 | 5500 | 6200 |
| $\bar{\delta}_i$     | -    | 600  | 400  | 500  | 700  |
| $\Delta_i$           | -    | 600  | 1000 | 1500 | 2200 |
| $\bar{\delta} = 550$ |      |      |      |      |      |

## 2.3. Tốc độ phát triển (Chỉ số phát triển)

Tốc độ phát triển là chỉ tiêu tương đối dùng để phản ánh nhịp điệu biến động của hiện tượng nghiên cứu qua hai thời kỳ/thời điểm khác nhau và được biểu hiện bằng số lần hay số phần trăm. Tốc độ phát triển được tính bằng cách so sánh giữa hai mức độ của chỉ tiêu trong dãy số biến động theo thời gian, trong đó một mức độ được chọn làm gốc so sánh. Tùy theo mục đích nghiên cứu, có thể tính các loại tốc độ phát triển sau:

### 2.3.1. Tốc độ phát triển liên hoàn (hay từng kỳ)

Dùng để phản ánh sự phát triển của hiện tượng qua từng thời gian ngắn liền nhau, được tính bằng cách so sánh một mức độ bất kỳ trong dãy số ở kỳ nghiên cứu với mức độ liền trước đó. Công thức tính:

$$t_i = \frac{y_i}{y_{i-1}}$$

với  $t_i$  là tốc độ phát triển liên hoàn

### 2.3.2. Tốc độ phát triển định gốc (hay cộng dồn)

Dùng để phản ánh sự phát triển của hiện tượng qua một thời gian dài được tính bằng cách so sánh mức độ bất kỳ của kỳ nghiên cứu trong dãy số với mức độ được chọn làm gốc không thay đổi (thường là mức độ đầu tiên trong dãy số).

Công thức tính:

$$T_i = \frac{y_i}{y_1}$$

Với  $T_i$  là tốc độ phát triển định gốc