

# Phương pháp quy đổi số liệu bị thiếu trong dãy số liệu

Nguyễn Thị Việt Hồng

**T**rong công tác thống kê, chúng ta thường gặp nhiều trường hợp số liệu không thu thập được một cách đầy đủ. Điều này có thể gây ảnh hưởng đến những khâu tiếp theo của quá trình tổng hợp và phân tích thống kê. Để khắc phục tình trạng này, các nhà kinh tế đã sử dụng nhiều phương pháp khác nhau để bổ sung số liệu bị thiếu trên cơ sở những số liệu đã có. Mỗi phương pháp được ứng dụng tùy thuộc vào đặc tính riêng của từng chỉ tiêu, từng dãy số liệu. Trong bài viết này, chúng tôi xin giới thiệu hai phương pháp khắc phục những số liệu bị thiếu trong dãy số liệu thường được áp dụng trong công tác thống kê tổng hợp và phân tích số liệu.

## 1. Quy đổi số liệu của tổng số

Phương pháp này thường được áp dụng đối với các chỉ tiêu tổng hợp theo

phạm vi không gian (toàn quốc, vùng, tỉnh, ngành...) và số liệu trong dãy số bị thiếu tương đối nhiều. Tuy nhiên, nếu số liệu bị thiếu càng nhiều thì độ tin cậy của số liệu tổng hợp càng kém. Vì vậy, chúng ta chỉ sử dụng phương pháp quy đổi số liệu trong trường hợp những đơn vị bị thiếu số liệu không có tầm ảnh hưởng quyết định trong dãy số liệu và giới hạn tối đa về giá trị của những số liệu bị thiếu không vượt quá 1/3 so với tổng số nếu tính cho kỳ gốc hoặc kỳ liền trước. Giả thiết này được đặt ra để đảm bảo mức sai số cho phép dưới 5% đối với số liệu được quy đổi.

Để dễ theo dõi, qui trình tính được thể hiện qua ví dụ giả định về “Số lao động đang làm việc trong các doanh nghiệp thời điểm 31/12 hàng năm của vùng Tây nguyên”.

### Trường hợp 1: Số liệu bị thiếu chỉ xảy ra ở năm 2006

Đơn vị tính: Người

Tên tỉnh	Năm 2002 <sup>(1)</sup>	Năm 2005	Năm 2006
Kon Tum	14291	17238	...
Gia Lai	44725	44816	47883
Đắk Lắk	46307	63309	64915
Đắk Nông	4108	7239	...
Lâm Đồng	25926	34282	35555
<b>Tổng số</b>	<b>135358</b>	<b>166884</b>	<b>173852</b> (166884 x 1,041754)

**Bước 1:** Tính tỷ trọng của tỉnh bị thiếu số liệu năm 2006 qua số liệu năm 2005 và tham khảo năm 2002.

Tỷ trọng của năm 2005:

$$(17238 + 7239)/166884 = 14,7\% < 1/3$$

Tỷ trọng của năm 2002:

$$(14291 + 4108)/135358 = 13,6\% < 1/3$$

Như vậy có thể tính số liệu tổng số của năm 2006.

**Bước 2:** tính tốc độ phát triển của những tỉnh có số liệu năm 2006 so với 2005.

$$(47883 + 64915 + 35555) / (44816 + 63309 + 34282) = \mathbf{1,041754}$$

động của các tỉnh thiếu số liệu là tương đương các tỉnh có số liệu trong vùng.

Bước 3: Tính số liệu tổng số của năm 2006 với giả thiết tốc độ phát triển về số lao

$$\text{Tổng lao động năm 2006} = 166884 \times 1,041754 = \mathbf{173852}$$

**Trường hợp 2:** Số liệu bị thiếu xảy ra ở cả năm 2005 và 2006

Đơn vị tính: Người

Tên tỉnh	Năm 2002 <sup>(2)</sup>	Năm 2005	Năm 2006
Kon Tum	14291	...	...
Gia Lai	44725	44816	47883
Đắc Lắc	46307	63309	64915
Đắc Nông	4108	7239	...
Lâm Đồng	25926	...	35555
<b>Tổng số</b>	<b>135358</b>	<b>164131</b> (135358 x 1,212571)	<b>171225</b> (164131 x 1,043218)

Bước 1: Tính tỷ trọng về giá trị của tỉnh bị thiếu số liệu năm 2005 và 2006 qua số liệu năm 2002

- Đối với các tỉnh thiếu số liệu năm 2002:

$$(14291 + 25926) / 135358 = 29,7\% < 1/3$$

- Đối với các tỉnh thiếu số liệu năm 2006:

$$(14291 + 4108) / 135358 = 13,6\% < 1/3$$

Như vậy có thể ước tính số liệu tổng số của năm 2006.

Bước 2: Tính tốc độ phát triển của những tỉnh có số liệu:

- Năm 2005 so với 2002:

$$(44816 + 63309 + 7239) / (44725 + 46307 + 4108) = \mathbf{1,212571}$$

- Năm 2006 so với 2005:

$$(47883 + 64915) / (44816 + 63309) = \mathbf{1,043218}$$

Bước 3: Tính số liệu tổng số của năm 2005 và 2006 với giả thiết tốc độ phát triển về số lao động của các tỉnh thiếu số liệu là tương đương các tỉnh có số liệu.

$$\text{Tổng lao động năm 2005} = 135358 \times 1,212571 = \mathbf{164131}$$

$$\text{Tổng lao động năm 2006} = 164131 \times 1,043218 = \mathbf{171225}$$

Lưu ý: Phương pháp này chỉ sử dụng để qui đổi số liệu của tổng số. Nếu muốn qui đổi số liệu chi tiết cần phải xét đến mối liên hệ mật thiết nhất của chỉ tiêu cần qui đổi số liệu với chỉ tiêu khác.

## 2. Qui đổi số liệu chi tiết

Trong công tác thống kê, có nhiều trường hợp chúng ta phải tự tính những số liệu chi tiết (cá thể) để có thể đạt được mục đích nghiên cứu. Song phải căn cứ vào những trường hợp cụ thể của những số liệu bị thiếu để tìm ra phương pháp phù hợp cho việc ước lượng. Ở đây, chúng tôi giới thiệu phương pháp tính qui đổi số liệu bị thiếu của thời kỳ (t+1) và (t+2) dựa vào số liệu của thời kỳ (t) và (t+3).

Phương pháp này thường được ứng dụng để tính số liệu tháng đối với các chỉ tiêu không thể thu thập được số liệu chính thức theo tháng do hạn chế về nguồn lực và tính hiệu quả trong khâu thu thập, tổng hợp. Nó được sử dụng với giả thiết là số liệu qui đổi có xu hướng chịu ảnh hưởng của thời kỳ liền kề nhiều hơn; số liệu sẵn có đã được xử lý các ảnh hưởng của yếu tố mùa vụ, chu kỳ

và các yếu tố bất thường, rủi ro. Đồng thời nếu sử dụng số liệu để phân tích và dự báo thì số chỉ tiêu cần quy đổi số liệu phải chiếm tỷ trọng nhỏ trong tổng số các chỉ tiêu cần đưa vào phân tích (khoảng dưới 10%) để đảm bảo mức sai số cho phép. Nội dung của phương pháp như sau:

Giả sử số liệu của thời kỳ "t" =  $\alpha$  ; số liệu của thời kỳ "t+3" =  $\beta$ .

Ta sẽ quy đổi số liệu cho thời kỳ (t+1) và (t+2) như sau:

- Số liệu của thời kỳ "t+1" =  $(2\alpha + \beta) / 3$

- Số liệu của thời kỳ "t+2" =  $(\alpha + 2\beta) / 3$

Để giúp người đọc dễ theo dõi, chúng ta xét ví dụ cụ thể dưới đây :

**Ví dụ** : Giả sử có số liệu về giá trị sản xuất công nghiệp 2006 như sau:

Quý I: 37358 tỷ đồng ; Quý III: 52169 tỷ đồng ; Quý II: 45107 tỷ đồng ; Quý IV: 64789 tỷ đồng.

Ta lập bảng tính dưới đây:

**Giá trị sản xuất công nghiệp theo giá thực tế năm 2006**

Đơn vị tính: Tỷ đồng

Tháng	Giá trị sản xuất CN theo quý	Số liệu được quy đổi	Giá trị sản xuất CN theo tháng
<b>3.</b>	<b>37358</b>		<b>37358</b>
4.	...	$(37358 \times 2 + 45107)/3 = 39941$	39941
5.	...	$(37358 + 45107 \times 2)/3 = 42524$	42524
<b>6.</b>	<b>45107</b>		<b>45107</b>
7.	...	$(45107 \times 2 + 52169) /3 = 47461$	47461
8.	...	$(45107 + 52169 \times 2) /3 = 49815$	49815
<b>9.</b>	<b>52169</b>		<b>52169</b>
10.	...	$(52169 \times 2 + 64789)/3 = 56376$	56376
11.	...	$(52169 + 64789 \times 2) /3 = 60582$	60582
<b>12.</b>	<b>64789</b>		<b>64789</b>

Trên đây là hai phương pháp được sử dụng trong việc quy đổi những số liệu bị thiếu trên cơ sở những số liệu đã có. Song việc lựa chọn phương pháp ứng dụng vào những trường hợp cụ thể cần phải cân nhắc đến đặc tính riêng có của từng chỉ tiêu và điều kiện thực tế cho phép, đồng thời phải đối chiếu so sánh kết quả của các phương pháp sử dụng khác nhau để có thể lựa chọn được phương pháp thực hành tốt nhất■

(1) Giả thiết là: số liệu của năm 2002 được coi là năm cơ bản để so sánh đối chiếu số liệu cho các năm điều tra mẫu 2005 và 2006. Số liệu năm 2002 được điều chỉnh trên cơ sở số liệu của

Tổng điều tra Cơ sở kinh tế, Hành chính, Sự nghiệp năm 2002.

(2) Giả sử số liệu năm 2002 là năm tiến hành Tổng điều tra Cơ sở kinh tế, Hành chính, Sự nghiệp được sử dụng làm năm cơ sở để so sánh đối chiếu số liệu cho các năm điều tra mẫu 2005 và 2006.

**Tài liệu tham khảo**

1. Economic and Social Research Institute - Cabinet Office, Government of Japan - Training in Conducting and Analysis of Business Statistics, March 2003.
2. World Bank: Statistical Manual - Method of Gap-filling.