

Giải thích sai số liệu thống kê

Nguyễn Thái Hà

Gải thích sai số liệu thống kê hay hiểu không đúng số liệu thống kê là hiện tượng khá phổ biến khi sử dụng thông tin thống kê. Việc giải thích sai này có thể do một trong số nguyên nhân như hiểu sai về số liệu, sử dụng các định nghĩa không so sánh được và giải thích sai thông tin mang tính chủ tâm.

Đôi khi, số liệu bị hiểu nhầm do các phương tiện thông tin. Dưới đây là ví dụ về sự đơn giản quá mức và hiểu sai số liệu. Nhật báo thống kê Canda (ngày 18/7/2000) công bố số liệu thống kê tội phạm năm 1999 theo các đề mục sau:

- Tội phạm có sử dụng bạo lực (gồm có các loại tội: giết người, giết người có chủ mưu, cướp của, hiếp dâm, hành hung, bắt cóc) giảm trong mọi phân loại¹,

- Tội cướp ngân hàng giảm mạnh,
- Buôn bán ma túy tăng,
- Tội lái xe gây ra chết người ổn định,
- Tội phạm thanh thiếu niên giảm năm thứ 7 liên tục.

Trên cơ sở thông tin công bố trên Nhật báo thống kê, một số báo Canda đưa tin với các tiêu đề như “Nền kinh tế nóng ngăn cản mức độ giết người” hay “Tỷ lệ giết người ở mức thấp trong vòng 32 năm”.

Những tiêu đề trên đều sai so với nội dung chính. Thứ nhất là báo chí đã bỏ qua phần kết quả chính mà chỉ tập trung vào một khía cạnh

của số liệu - đó là tỷ lệ tội phạm của quốc gia đã giảm 8 năm liên tục. Thứ hai, số liệu chứng tỏ rằng tỷ lệ tội phạm nghiêm trọng giảm 2,4% trong năm 1999, giảm 7 năm liên tục sau 15 năm tăng. Tất cả các loại tội phạm nghiêm trọng đều giảm trong năm 1999, trong đó tội phạm giết người giảm 4,7%, giết người có chủ mưu giảm 8,8%, hành hung giảm 2,0%, hiếp dâm giảm 7,3%, và cướp của giảm 1,5%.

Thông cáo cũng cho biết là tỷ lệ giết người nói chung đã giảm từ giữa những năm 1970. Xu thế này vẫn được duy trì trong năm 1999. Theo báo cáo của cảnh sát có 536 kẻ phạm tội giết người, ít hơn năm trước 22 người. Năm 1999, tỷ lệ giết người là 1,76 trên 100 nghìn người - là mức thấp nhất kể từ năm 1967.

Mặc dù có thể dự đoán là thịnh vượng kinh tế góp phần hạ thấp tỷ lệ tội phạm, thì số liệu do Cơ quan Thống kê Canda công bố không đo được mối quan hệ giữa số liệu thống kê kinh tế và tội phạm. Và giả sử nếu đo được mối quan hệ này thì hóa ra là tỷ lệ tội phạm đã giảm từ năm 1991- là năm Canda bắt đầu lâm vào tình trạng suy thoái.

Nhật báo Thống kê lưu ý rằng có nhiều yếu tố góp phần làm thay đổi mức độ phạm tội. Tuy nhiên, vấn đề được dẫn dắt theo quan điểm riêng của một số nhà báo này lại trở thành sự thật. Những sự liên quan được thấy trên các tit báo thực sự chỉ là tiên đoán và không thể hiện số liệu khách quan. Các nhà báo đã hiểu sai số

¹ *Crime Statistics in Canada 2007 (Catalogue no. 85-002-X, Vol. 28, no. 7)*

liệu công bố, chắc chắn là do họ hiểu sai những nguyên nhân chính và hậu quả của tội phạm.

Những người đại diện của Cơ quan Thống kê Canada đã giành nhiều thời gian xem xét việc các phương tiện truyền thông sử dụng số liệu của Cơ quan Thống kê Canada công bố từng ngày. Những người đại diện này cũng trả lời các câu hỏi của các cơ quan truyền thông đại chúng liên quan đến số liệu thống kê và đảm bảo cho số liệu được hiểu đúng. Nếu xảy ra việc hiểu sai số liệu thì những người đại diện của Cơ quan Thống kê

có thể so sánh: thời tiết ở khu vực này với một khu vực khác, các kết quả thể thao trong quá khứ với hiện tại, hoặc so sánh hoạt động nghiên cứu của nam và nữ.

Tuy nhiên, các vấn đề có thể nảy sinh trong việc so sánh các chỉ tiêu thống kê theo các định nghĩa, phân loại hoặc phương pháp thu thập số liệu khác nhau. Đây là vấn đề rất cụ thể đối với số liệu thống kê từ các nguồn khác nhau. Điều này được thể hiện rõ nhất trong thống kê hộ tịch. Ví dụ như bảng số liệu dưới đây.

Số người đã kết hôn chia theo giới tính, năm 1998 - 2002

	1998	1999	2000	2001	2002
Cả hai giới	14.630.173	14.711.793	14.806.694	14.913.766	15.018.130
Nam	7.299.132	7.337.226	7.381.266	7.431.522	7.476.537
Nữ	7.331.041	7.374.567	7.425.428	7.482.244	7.541.593

Canada sẽ cố gắng để sửa cho đúng.

Sử dụng những định nghĩa không thể so sánh

Vấn đề quan trọng là hiểu các định nghĩa và khái niệm thống kê sau những thông tin mà bạn sử dụng. Nếu bạn đang xem xét các vấn đề về lực lượng lao động, thì bạn phải biết rõ các định nghĩa liên quan đến lực lượng lao động như thất nghiệp, việc làm, và tỷ lệ tham gia lực lượng lao động. Nếu bạn đang xem xét số liệu về vấn đề môi trường, thì cần xem xét định nghĩa và các khái niệm liên quan đến các từ ngữ như rừng, đất rừng, sự tuyệt chủng, các loài bị đe dọa tuyệt chủng, và vườn quốc gia.

Một lợi thế của thông tin thống kê đó là nó có thể so sánh được, theo đó các xu thế và các đặc điểm được phát hiện. Ví dụ như, một người

Nguồn: CANSIM table 051-0010

Theo Thống kê Canada, định nghĩa người đã kết hôn bao gồm những người đã kết hôn hợp pháp và sống cùng nhau, những người đã kết hôn và ly thân hợp pháp, và những người kết hôn theo thông luật. Nếu so sánh số lượng những người đã kết hôn ở bảng số liệu trên với số liệu từ một cuộc điều tra (không tính đến những người kết hôn theo thông luật) trong phân loại “đã kết hôn”, thì kết quả sẽ rất khác nhau vì các định nghĩa không giống nhau. Trong bảng số liệu trên, tại sao số liệu lại chứng tỏ là có nhiều phụ nữ đã kết hôn hơn nam giới? về logic thì hai số liệu này phải bằng nhau. Vấn đề xảy ra ở đây là do các cá nhân đã tự gán định nghĩa riêng của họ theo thuật ngữ đã kết hôn, và việc tự gán này đã làm cho số liệu có phần khác nhau.

Có ý hiểu sai thông tin

Thông tin thực tế phải trung thực, khách quan và chính xác. Tuy nhiên, vấn đề quan trọng để nhận ra thông tin đó là thông tin bị hiểu sai do xu hướng cá nhân, số liệu thống kê thiếu chính xác, và thậm chí là hư cấu số liệu.

Đối với vấn đề này, bạn luôn phải quyết định về thông tin đã được cung cấp tới bạn. Đảm bảo rằng bạn biết được thông tin này từ đâu và tìm hiểu xem nguồn thông tin này có tin cậy hay không. Cũng như vậy, cần tìm hiểu xem phương pháp nào đã được sử dụng để thu thập thông tin và xử lý số liệu.

Vấn đề quan trọng cuối cùng là cần phải biết độ chính xác của số liệu thống kê là bao nhiêu. Nguyên nhân là do các cuộc điều tra có hai loại sai số: sai số chọn mẫu và sai số phi chọn mẫu. Sai số chọn mẫu xuất hiện trong các cuộc điều tra chọn mẫu (vì chỉ có một phần tổng thể được nghiên cứu mà không phải là toàn bộ tổng thể). Sai số phi chọn mẫu xảy ra trong cả điều tra chọn mẫu và tổng điều tra.

Liên hệ với Thống kê Việt Nam

Tổng cục Thống kê đã có một số lần giải thích với các phương tiện truyền thông đại chúng về thông tin thống kê. Sau đây là ví dụ về chỉ số giá (CPI):

Năm 2007 khi Tổng cục Thống kê trình Thủ tướng xem xét thay đổi trong cách tính CPI thì một số phương tiện thông tin đại chúng cho rằng *phương pháp tính chỉ số giá của Tổng cục Thống kê có vấn đề* và đã dẫn đến sự nghi ngờ các chỉ số giá mà Tổng cục Thống kê đã công bố. Theo Bà Trần Thị Hằng, Vụ trưởng Vụ

Thương mại, Dịch vụ và Giá cả, Tổng cục Thống kê, sự thay đổi trong cách tính CPI (đã được trình Thủ tướng xem xét) *không phải là sự thay đổi về phương pháp tính, mà là bổ sung gốc so sánh* khi công bố CPI cũng như khi xác định chỉ tiêu kế hoạch về tỷ lệ lạm phát theo nội dung của CPI [2].

Cũng liên quan đến chỉ số giá tiêu dùng, Thứ trưởng Tài chính Trần Văn Tá nói “theo cách tính mới, CPI 11 tháng chỉ tăng 7,92%” chứ không phải là 9,45% như Tổng cục thống kê công bố.

Trao đổi với Báo Thanh niên, ông Đào Ngọc Lâm - nguyên phó vụ trưởng Vụ Phương pháp chế độ Tổng cục Thống kê đã giải thích 2 con số này được *tính theo 2 cách khác nhau*: 9,45% là chỉ số CPI thời điểm (tháng 11 của năm 2007 so với tháng 11 năm 2006) và 7,92% là CPI bình quân của 11 tháng năm 2007 so với bình quân của 11 tháng năm 2006.

Phòng vấn của phóng viên Vietnamconomy với ông Nguyễn Đức Hòa, Tổng cục trưởng Tổng cục Thống kê: “... nhưng đến thời điểm này nhiều người dân vẫn còn chưa hiểu rõ cách tính CPI, thậm chí nhiều người còn *nghi ngờ tính sát thực của những con số về giá cả vì Bộ Tài chính cho rằng có hai cách tính CPI?*”[3]

... và trả lời của ông Nguyễn Đức Hòa “Tôi xin khẳng định rằng, từ trước đến nay, Ngành Thống kê chúng tôi vẫn *chỉ áp dụng một phương pháp tính CPI* ... Chúng tôi vẫn thường công bố giá theo các cấp so sánh, đó là: tốc độ tăng giá so với cùng kỳ năm trước, tốc độ tăng giá so với tháng trước và tốc độ tăng giá bình

quân của cả năm so với năm trước. Đây là một phương pháp tính đã được quốc tế công nhận và đánh giá cao tính sát thực. Vì vậy, trong thời gian qua không thể có chuyện chúng tôi thay đổi phương pháp tính CPI như một số phương tiện thông tin đại chúng đã nêu”.

Phần mở đầu của bài *9,3 triệu hay 3,4 triệu doanh nghiệp hộ gia đình ở Việt Nam năm 2004* đăng trên Thông tin Khoa học thống kê số 2 năm 2008 viết: “9,3 triệu doanh nghiệp hộ gia đình phi nông nghiệp ở Việt Nam (năm 2004) là con số mà Wim Vijverberg, tác giả bài viết “Đề xuất do hoạt động của doanh nghiệp hộ gia đình ở Việt Nam” đã ước lượng từ kết quả cuộc Khảo sát mức sống hộ gia đình năm 2004. Trong khi đó, Điều tra cơ sở sản xuất kinh doanh cá thể phi nông, lâm nghiệp và thủy sản năm 2004 chỉ cho kết quả là 2,9 triệu doanh nghiệp (cơ sở), bằng 1/3 so với số liệu tác giả bài viết nói trên đưa ra. Từ đó, tác giả khẳng định cuộc điều tra cơ sở SXKD cá thể phi nông nghiệp năm 2004 không thành công do bỏ sót lượng lớn doanh nghiệp hộ gia đình có thể làm cho GDP Việt Nam thấp hơn một cách đáng kể so với thực tế” [4].

Tác giả bài viết, Thạc sỹ Nguyễn Văn Đoàn đã lý giải nguyên nhân của sự khác biệt lớn về số lượng doanh nghiệp hộ gia đình trong hai cuộc điều tra Khảo sát Mức sống hộ gia đình năm 2004 và Điều tra cơ sở sản xuất kinh doanh cá thể phi nông nghiệp năm 2004 là do mục đích điều tra khác nhau, đối tượng, đơn vị, phạm vi, phương pháp, nội dung khác nhau nên hai số liệu này đương nhiên là khác nhau.

Hai ví dụ trên cho thấy tầm quan trọng của

việc hiểu được những kiến thức thống kê chủ yếu và sử dụng các số liệu thống kê theo các phương pháp thích hợp. Đặc biệt là việc hiểu và nhận thức đúng của những người đưa thông tin và số liệu thống kê trên các phương tiện thông tin đại chúng. Và điều này cũng cho thấy Tổng cục Thống kê cần phải theo dõi thường xuyên việc thông tin về số liệu trên các phương tiện thông tin đại chúng để định hướng cho độc giả hiểu đúng bản chất của thông tin, số liệu công bố và sửa cho đúng khi có những nhìn nhận đánh giá không đúng về tình hình kinh tế xã hội và số liệu thống kê. ■

TÀI LIỆU THAM KHẢO:

1. *Misunderstanding the data*

<http://www.statcan.gc.ca/edu/power-pouvoir/ch6/misinterpretation-mauvaiseinterpretation/5214805-eng.htm>

2. *Không thay đổi phương pháp tính chỉ số giá.*

Thời báo kinh tế điện tử ngày 20/09/2007 (<http://vneconomy.vn/66410POC19/khong-thay-doi-phuong-phap-tinh-chi-so-gia.htm>)

3. *Không có chuyện thay cách tính CPI”.* *Việt báo.vn ngày 02/01/2008.*

<http://vietbao.vn/Kinh-te/Khong-co-chuyen-thay-cach-tinh-CPI/65116977/87/>

4. *Nguyễn Văn Đoàn. 9,3 triệu hay 3,4 triệu*

doanh nghiệp hộ gia đình ở Việt Nam năm 2004. Thông tin Khoa học thống kê số 2/2008.