

GIỚI THIỆU MỘT SỐ PHƯƠNG PHÁP Đánh giá kết quả Điều tra

Trần Mạnh Hùng – Viện KHTK

Đánh giá kết quả điều tra là công việc thận trọng, phức tạp nhưng rất quan trọng và cần thiết, là bộ phận cấu thành không thể thiếu được trong chương trình thu thập số liệu, nó đóng vai trò quan trọng trong công tác cải tiến phương pháp thu thập, xử lý và ước lượng số liệu. Tuy nhiên, không phải ai trong chúng ta khi tham gia vào các cuộc điều tra hay sử dụng số liệu điều tra cũng hiểu vấn đề này. Để giúp cho bạn đọc hiểu biết thêm về công việc đánh giá kết quả các cuộc điều tra, bài viết giới thiệu một số vấn đề cốt yếu trong vấn đề này. Do mục tiêu và các lĩnh vực tiến hành các điều tra có khác nhau nên ít nhiều trong phương pháp và nội dung đánh giá cũng khác nhau. Tuy nhiên, những khâu chủ yếu để tiến hành trong công việc đánh giá cơ bản như nhau. Sau đây là những nội dung cơ bản đó.

Độ chính xác của số liệu

Như chúng ta đã biết, trong bất cứ tập số liệu về kết quả điều tra về kinh tế xã hội nào đều có sai sót. Cho dù các chương trình thu thập số liệu đã được thiết kế cẩn thận để giảm thiểu tới mức có thể, nhưng chắc chắn các tập số liệu thu thập được vẫn còn một số sai sót. Do vậy, để phân tích kết quả điều tra, chúng ta cần có một số hiểu biết nhất định về chất lượng số liệu. Đánh giá chất lượng số liệu cũng đóng vai trò quan trọng trong công tác cải tiến phương pháp thu thập, xử lý và ước lượng số liệu sau này...Như vậy, ở một chừng mực nào đó, phương pháp đánh giá chất lượng số liệu là bộ phận cấu thành trong chương trình thu thập số liệu. Tất cả những số liệu thống kê từ bất cứ nguồn

nào và bằng phương pháp thu thập ra sao, thì vẫn tồn tại tiềm tàng trong đó những sai số với những mức độ khác nhau. Để người sử dụng sử dụng tốt số liệu điều tra, khi đưa ra kết quả điều tra nên có những trình bày về chất lượng và hạn chế của số liệu, kể cả sai số chọn mẫu và sai số không chọn mẫu trong đó.

Khi sử dụng số liệu, cần lưu ý một số vấn đề:

Thứ nhất, hiểu biết về chất lượng số liệu và cách sử dụng cũng như thể hiện chúng một cách có hiệu quả là một điều không đơn giản. Hiểu biết này được hiểu theo nghĩa là biết xác định phương pháp quan sát, thu thập thực tế đã tiến hành và độ tin cậy của chúng ra sao để có những kết quả đó. Làm được như vậy, không chỉ tạo điều kiện thay đổi và khắc phục tình trạng không đầy đủ mang tính cố hữu về số liệu, mà thông tin về các chủng loại sai số và mức độ sai số giúp ích cho công việc sửa đổi hợp lý dữ liệu hay hiệu chỉnh khi thể hiện chúng.

Thứ hai, đánh giá chất lượng số liệu có tầm quan trọng đối với việc đo lường và cải tiến trong khâu thiết kế và phương pháp điều tra sau đó. Cần kiểm tra chi tiết về các nguồn số liệu, mức độ và ảnh hưởng của sai số để có thể nhận biết và cải tiến thiết kế và phương pháp điều tra cũng như thay đổi các nguồn tài chính thông qua các khâu điều hành các cuộc điều tra khác nhau sao cho có hiệu quả hơn. Công việc giám sát và tiếp tục hoàn thiện chất lượng số liệu thu thập được càng có tầm quan trọng đặc biệt hơn khi

nó là cuộc điều tra về lực lượng lao động toàn quốc. Kết quả của cuộc điều tra như vậy thường tạo điều kiện cho việc đưa ra những quyết định về chính sách quan trọng; những kết quả điều tra đó cũng đưa ra một chuẩn mực hay là tài liệu tham khảo với một phạm vi thông tin rộng lớn, nảy sinh từ những nguồn khác nhau. Đánh giá hoạt động kinh tế, việc làm, thất nghiệp và hiện tượng có liên quan ở những trường hợp khác nhau là một công việc phức tạp. Những kết quả như vậy có thể là rất nhạy cảm đối với những khái niệm và định nghĩa đã được sử dụng, nói một cách chính xác là những kết quả đó được thao tác ra sao và số liệu được thu thập và xử lý trong những điều kiện như thế nào.

Các loại sai số: Mục đích của một cuộc điều tra là từ những quan sát hữu hạn (một mẫu) của tổng thể, ta ước lượng hay suy rộng khả năng sử dụng chung cho một tổng thể nghiên cứu, chúng ta có thể phân biệt hai nhóm sai số tác động đến quá trình này như sau:

(a) Sai số phát sinh trong quá trình ngoại suy kết quả từ các đơn vị quan sát cho toàn bộ tổng thể nghiên cứu. Những sai số này bao gồm cả những sai số phát sinh do diện điều tra mẫu (diện điều tra không đầy đủ), do chọn mẫu, do khâu triển khai điều tra, do không trả lời câu hỏi cũng như do tính biến động của việc chọn mẫu và cuối cùng là do có độ lệch trong ước lượng. Nhóm sai số này tập trung vào quá trình thiết kế mẫu, vào khâu thực hiện điều tra và ước lượng.

(b) Sai số nảy sinh sau khi thực hiện điều tra, là những gì đo lường được lại sai lệch so với dự kiến ở những đơn vị điều tra. Những sai số về đo lường như vậy tập trung vào nội dung trọng yếu của cuộc điều tra. Chúng nảy sinh khi:

- Xác định đối tượng điều tra

- Vận hành chúng trong các tập bảng hỏi (có mang tính mạch lạc và phù hợp không)

- Truyền đạt câu hỏi đến người trả lời, tính tự nguyện cần có và khả năng cung cấp thông tin của người trả lời.

- Khi ghi chép, đánh mã, hiệu chỉnh và xử lý những câu trả lời.

Dựa vào những xem xét mang tính thao tác chúng ta đưa ra các phân loại trên chủ yếu, mang tính chất nặng về ý nghĩa hơn là đưa ra ở đây một sự khác biệt thông thường giữa những sai số chọn mẫu và sai số không chọn mẫu. Nhóm (a) nói về khả năng tạo lập tổng thể mục tiêu từ những đơn vị quan sát, kể cả những khả năng thay đổi việc lấy mẫu, cũng như đề cập đến nhiều loại độ lệch khác nhau có liên quan đến chọn mẫu và sai số về phạm vi và sai số do không trả lời. Tất cả những loại này có quan hệ trực tiếp đối với các chọn lựa thiết kế mẫu. Trong nhiều trường hợp một số cuộc điều tra hay những vòng điều tra có chung sườn chọn mẫu, mẫu chủ, thiết kế mẫu và thậm chí đôi khi mẫu chung là của các đơn vị tổng thể. Trong điều kiện như vậy những sai số liên quan đến quá trình chọn mẫu có chiều hướng là những sai số chung của những cuộc điều tra này, ít hay nhiều cũng độc lập với các chi tiết mang tính chủ quan và các phương pháp đánh giá và kiểm soát thông thường đối với sai số sử dụng cho suốt cuộc điều tra. Tất nhiên, phương pháp này cũng áp dụng cho các vòng điều tra khác nhau của một cuộc điều tra lực lượng lao động tiến hành thường xuyên.

Nhóm (b) đề cập đến tính chính xác của độ đo ở cấp đơn vị tổng thể được tiến hành trong điều tra, người trả lời cung cấp những thông tin ra sao, người điều tra viên ghi lại như thế nào, công việc đánh mã và

hiệu chỉnh do cán bộ xử lý ra sao; sửa chữa, nhập máy và cuối cùng máy tính kẻ bảng biểu, tất cả những công việc trên khi tiến hành có thể làm chệch so với thực tế từng loại số liệu liên quan. Nhóm sai số này có thể được nghiên cứu trong từng giai đoạn khác nhau trong công tác vận hành điều tra, như thu thập số liệu, đánh mã, hiệu chỉnh, nhập máy và xử lý bằng máy tính.

Phân loại rõ ràng trên đây cần được bổ sung thông qua việc nhận biết nguồn gốc và chủng loại sai số cụ thể càng chi tiết càng tốt, để tạo điều kiện kiểm soát những sai số và đánh giá ảnh hưởng của chúng. Thí dụ, trong các sai số nảy sinh khi thực hiện chọn mẫu, lợi ích ở đây không chỉ là phân biệt giữa sai số do việc điều tra thiếu với do không trả lời, mà còn phân biệt thêm giữa sai số do điều tra thiếu hộ với điều tra thiếu thành viên trong hộ; hay giữa sai số về không trả lời do từ chối và do không tìm thấy đối tượng để phỏng vấn, và giữa những câu trả lời định sẵn hay những câu trả lời thông qua tiếp xúc với đối tượng điều tra mà có.

Sai số biến phân và độ chệch: Nói chung ảnh hưởng của bất kỳ nguồn sai số cá biệt nào trong ước lượng điều tra có thể chia ra làm 2 loại: Sai số biến phân và độ chệch. Môi trường là một trong những điều kiện quan trọng để các cuộc điều tra tiến hành, đó là những điều kiện xã hội nói chung, chất lượng của dàn điều tra mẫu được sử dụng, phương pháp thu thập số liệu được tiến hành, loại thông tin cần đến, bảng hỏi thiết kế, khả năng cũng như công tác đào tạo cán bộ điều tra. Điều kiện công việc của họ bao hàm cả việc nắm giữ chức vụ và mức lương và hình thức trả công lao động, loại người được phỏng vấn trong đó v.v...đều tác động đến kết quả điều tra. Ngoài những điều kiện quan trọng đó ra thì kết quả điều tra cũng bị các

yếu tố nhất thời và ngẫu nhiên ảnh hưởng, các yếu tố đó là những đơn vị đặc thù được chọn vào mẫu, những điều tra viên và những người đánh mã cẩn thận được trưng dụng, những điều kiện mà các cuộc phỏng vấn đặc biệt có thể xảy ra một cách ngẫu nhiên v.v... Chúng ta có thể nhận thức được rằng sau này cuộc điều tra có thể tiến hành trở lại trong cùng những điều kiện cơ bản như vậy. Tuy nhiên, những cuộc điều tra lặp lại như vậy sẽ cho ta những kết quả khác nhau, tùy thuộc vào những yếu tố nhất thời hay ngẫu nhiên. Sai số biến phân đo lường tính biến thiên giữa các ước lượng khác nhau từ các lần điều tra lại mang tính giả thiết. Ước lượng trung bình của tất cả những lần điều tra lại khả dĩ là giá trị dự kiến hay kỳ vọng chịu tác động của những điều kiện điều tra chính yếu đã được qui định. Sự khác nhau giữa kỳ vọng hay giá trị trung bình này và giá trị thực tế đối với một tổng thể mục tiêu là độ lệch.

Sai số về phạm vi điều tra và những sai số có liên quan trong chọn mẫu và thực hiện điều tra, cùng với sai số do không trả lời và sai số hệ thống trong những khâu trả lời nhận được trong điều tra, nhìn chung là những độ lệch nảy sinh từ những điều kiện chính yếu khi chỉ đạo hoạt động điều tra. Chúng ta có thể kỳ vọng về những sai số với mức độ tương tự nếu như tiến hành lại theo cùng những điều kiện như vậy. Kích cỡ của độ lệch phụ thuộc nhiều vào việc so sánh với thông tin bên ngoài theo từng cuộc điều tra riêng. Thông qua những phương pháp khác nhau, ta nhận được độ đo này với độ chính xác cao hơn.

Dưới đây là những thành phần chủ yếu, khái quát về các sai số trong điều tra.

Nhóm thứ nhất, nhóm “các sai số quan sát” đề cập đến chênh lệch giữa giá

trị quan sát và giá trị thực tế đối với từng đơn vị trong tổng thể điều tra. Nhóm này được chia ra làm ba loại theo thời kỳ mà sai số phát sinh

(1) Sai số nhận thức (khái niệm), nảy sinh trong giai đoạn thiết kế điều tra;

(2) Sai số trong khâu trả lời nảy sinh trong giai đoạn thu thập số liệu;

(3) Sai số xử lý, có thể chia nhỏ ra thành các sai số hiệu chỉnh, sai số đánh mã, sai số nhập số liệu v.v...

Sai số do trả lời đặc biệt nảy sinh từ nhiều nguồn và theo những dạng khác nhau. Một số trong đó là độ lệch hệ thống, loại khác lại là những sai số biến phân, mang nặng tính ngẫu nhiên hơn (như thay đổi điều tra viên).

Nhóm thứ 2, nhóm về các sai số ngoại suy, liên quan đến việc suy rộng kết quả từ các đơn vị điều tra được quan sát thực tế ra toàn bộ tổng thể nghiên cứu hay tổng thể mục tiêu. Các phân loại nhỏ bao gồm:

(4) Sai số diện điều tra và sai số có liên quan (như các sai số trong chọn mẫu);

(5) Sai số do không quan sát hay không trả lời và

(6) Sai số mẫu.

Những sai số về diện điều tra xuất hiện phổ biến nhất theo dạng lệch hệ thống do điều tra thiếu. Trong các sai số do không trả lời như từ chối hay vắng nhà v.v... nếu ta phân biệt được các loại lỗi như vậy thì rất có lợi. Toàn bộ từ (1) - (5) được gọi là sai số không chọn mẫu, khác biệt với sai số chọn mẫu (6). Sai số chọn mẫu có thể chia ra độ lệch do lấy mẫu, độ lệch do ước lượng (độ lệch này thường là nhỏ trong các mẫu thiết kế thích hợp theo cơ mẫu hợp lý) và phương sai mẫu.

Sai số quan sát:

(1) Sai số nhận thức (khái niệm)

(2) Sai số trả lời hay sai số thu thập:

- Độ lệch do trả lời
- Phương sai trả lời

(3) Sai số xử lý:

- Sai số hiệu chỉnh
- Sai số đánh mã
- Sai số nhập dữ liệu

Sai số ngoại suy:

(4) Sai số do diện điều tra

- Điều tra thiếu
- Điều tra thừa
- Sai số chọn mẫu

(5) Sai số do không quan sát (không trả lời)

- Do từ chối
- Do vắng nhà
- Vì lý do khác

(6) Sai số chọn mẫu

- Độ chệch lấy mẫu hay độ chệch do ước lượng
- Phương sai lấy mẫu

(Sai số không chọn mẫu từ (1) đến (5))

Sai số chọn mẫu và sai số ngẫu nhiên nói đến trong các giai đoạn thu thập và xử lý số liệu là nằm trong các sai số biến phân quan trọng nhất. Những sai số này cho thấy sự không chắc chắn trong các ước lượng điều tra, không chắc chắn như ta mong đợi về độ đo mà ta thường đo được trong chính cuộc điều tra (thông qua việc so sánh những mẫu con nội bộ bên trong, độc lập nhau và mỗi loại cho ta một ước lượng đúng), hay so sánh với những mẫu điều tra độc lập với nhau và trong cùng một điều kiện như nhau.

(Còn nữa)