

SỰ KHÁC NHAU GIỮA DỮ LIỆU ĐIỀU TRA THỐNG KÊ VÀ DỮ LIỆU HÀNH CHÍNH

(trường hợp tính chỉ số tiền lương và việc làm)

Rezarta Myrtollari, Viện Thống kê Quốc gia Ý

Giới thiệu

Trong quá trình sản xuất số liệu thống kê nhà nước, đối với một hiện tượng nhất định, dữ liệu có sẵn có thể đến từ các điều tra thống kê và các nguồn hành chính. Ngày nay, sự kết hợp của hai nguồn này là một chiến lược đầy hứa hẹn và sáng tạo, tác động đến chất lượng và số lượng nghiên cứu và làm tăng tiềm năng của dữ liệu (Künn, 2015). Tuy nhiên, việc sử dụng kết hợp này thường đi kèm với những thách thức rất lớn, vì thực tế đơn giản là mục đích thiết kế hai nguồn dữ liệu là khác nhau. Dữ liệu hành chính được định nghĩa là bộ dữ liệu được thu thập bởi các tổ chức chính phủ hoặc cơ quan cho các mục đích thuế, lợi ích hoặc hành chính công (UNECE, 2011). Mặt khác, dữ liệu từ các cuộc điều tra được thu thập cụ thể cho mục đích thống kê. Bài viết này xem xét sự khác biệt giữa các nguồn hành chính và điều tra tập trung cụ thể vào các biến số chung của bảng lương là nguồn hành chính và các điều tra được thực hiện trong Viện Thống kê quốc gia Ý (Istat).

Theo Penneck (2007) điều tra khác với dữ liệu hành chính theo nghĩa là chúng được thiết kế đặc biệt cho mục đích phân tích, do đó, phạm vi tổng thể, định nghĩa, phương pháp và thời gian có thể được thiết kế để đáp ứng các nhu cầu phân tích này. Tuy nhiên, cỡ mẫu có thể là một vấn đề nếu nó nhỏ vì các điều tra quy mô lớn là đắt đỏ và các điều tra quy mô nhỏ có hạn chế sử dụng. Các mẫu cũng có thể bị sai số và chệch do không phản hồi. Ngoài ra, theo Penneck, chúng ta không thể chắc chắn về tính chính xác của các phản hồi điều tra kinh doanh, ví dụ, so với dữ liệu hành chính được thu thập cho mục đích thuế. Các hệ thống hành chính cũng yêu cầu dữ liệu từ các cá nhân, nhưng sau này thường xem đây là một phần không thể thiếu trong quy trình hành chính hơn là một gánh nặng thống kê bổ sung. Các phần sau đây sẽ xem xét chi tiết một số vấn đề này dựa trên công trình của Johnson và Moore (2008), minh họa chúng bằng các ví dụ cụ thể từ trường hợp ước tính mức lương trung bình và chỉ số việc làm từ hai nguồn khác nhau trong Istat.

1. Vấn đề phạm vi tổng thể

Một hệ thống hồ sơ hành chính xác định phạm vi tổng thể bởi pháp luật dựa trên phạm vi của chương trình dự định đăng ký. Tổng thể này thường bị giới hạn bởi các đặc điểm nhân khẩu học hoặc kinh tế cụ thể.

Theo Johnson và Moore (2008), trong một số trường hợp, cá nhân có thể cần phải thực hiện một số hành động để trở thành một phần của hệ thống hành chính (ví dụ, đăng ký nông dân vào hệ thống thuế và an sinh xã hội bằng cách lấy NIPT để hưởng lợi từ các chương trình hỗ trợ của chính phủ). Do đó,

➤ ➤ ➤ THÔNG KÊ QUỐC TẾ VÀ HỘI NHẬP

điều quan trọng, theo các tác giả, để xem xét những gì khuyến khích các đơn vị cá thể là một phần của kế hoạch. Có thể có một số yếu tố có lợi cho một số cá nhân để tránh đăng ký, đặc biệt là nếu hoàn cảnh của họ đặt gần ngưỡng yêu cầu tham gia bắt buộc hoặc liên quan đến chi phí tài chính, chẳng hạn như đặt mức lương tối thiểu mà các khoản đóng góp được tính. Một yếu tố khác là sự thay đổi của các chính sách có thể làm dao động tổng thể được nghiên cứu từ năm này sang năm khác.

Ủy ban Phương pháp Thống kê Liên bang chỉ ra sự khác biệt về đơn vị quan tâm. Đơn vị nghiên cứu cần thiết cho mục đích thống kê thường tập trung vào đặc điểm của các nhóm được hình thành bởi các đơn vị (ví dụ, doanh nghiệp hoạt động trong một hoạt động cụ thể hoặc doanh nghiệp lớn), trong khi dữ liệu hành chính tập trung vào việc xác định các đơn vị cụ thể để dựa trên đặc điểm cá nhân của họ (ví dụ, nhân viên làm việc toàn thời gian hoặc việc làm kép) một số hành động nhất định có thể được thực hiện. Do đó, sự khác biệt trong thực thể được báo cáo trong báo cáo thuế giới hạn tính hữu ích của dữ liệu đối với một số loại nghiên cứu.

Johnson và Moore giải thích rằng tổng thể mục tiêu của một cuộc điều tra được xác định bởi mục đích của nghiên cứu, tính sẵn có của dàn mẫu và chi phí của mẫu. Tổng thể cho hầu hết các điều tra được lấy từ các nguồn hiện có, chẳng hạn như dữ liệu dân số dựa trên địa lý, danh sách địa chỉ hoặc các nguồn hành chính khác. Thông thường, một trong những vấn đề khó khăn nhất trong việc thiết kế điều tra là tìm ra một tổng thể phù hợp (Lessler và Kalsbeek, 1992).

Nếu dàn mẫu tổng thể được chọn để chọn mẫu là không phù hợp, nó có thể dẫn đến các vấn đề thiếu phạm vi và có thể ảnh hưởng đến kết quả thu được từ dữ liệu điều tra. Một vấn đề khác phát sinh nếu một cuộc

điều tra nhằm vào một tổng thể khó tìm kiếm hoặc đo lường. Liên quan trực tiếp đến sự sẵn có của tổng thể là chi phí tiềm năng để nhận dữ liệu tổng thể và chi phí phỏng vấn một mẫu có cỡ mẫu mong muốn. Đối với các tổng thể mục tiêu khó tìm, chỉ đơn giản là tăng chi phí để tăng cỡ mẫu nhằm cung cấp độ bao phủ tốt hơn có thể gây cản trở cho việc thực hiện một sáng kiến như vậy (Johnson và Moore, 2008, tr.13).

2. Vấn đề nội dung

Johnson và Moore liệt kê một số vấn đề nội dung cần được xem xét trong khi làm việc với dữ liệu hành chính và điều tra. Một trong số đó là mục đích thu thập dữ liệu hành chính, có thể có tác động đáng kể đến tính hữu ích của chúng cho mục đích thống kê liên quan đến lượng dữ liệu có sẵn, định nghĩa dữ liệu, tính nhất quán giữa các khoảng thời gian khác nhau và chất lượng dữ liệu. Các tác giả cho rằng nhiều lần tính hữu dụng của các hệ thống đăng ký hành chính bị hạn chế vì chỉ những biến số cần thiết để quản lý chương trình nộp thuế và thuế được thu thập. Các biến này chỉ có thể là một phần nhỏ của dữ liệu được báo cáo ở dạng hành chính.

Ngoài ra, do các yêu cầu của chương trình được xác định bởi luật pháp, các khái niệm và định nghĩa về các biến được sử dụng để đáp ứng nhu cầu của chương trình không nhất thiết phải khớp với các yêu cầu cho phân tích kinh tế hoặc xã hội (Brackstone, 1987). Ví dụ, một trong những vấn đề gặp phải trong dữ liệu hành chính của chúng tôi là do sử dụng khái niệm ngày làm việc thay vì giờ làm việc trong việc khai báo bảng lương của người nộp thuế. Sự khác biệt này làm cho việc so sánh dữ liệu việc làm với tài khoản quốc gia trở nên khó khăn. Các yếu tố như vậy có thể đặt ra những hạn chế nghiêm trọng về tính hữu ích chung của hệ thống dữ liệu hành chính hoặc yêu cầu cơ quan hành chính thực hiện thu thập và/hoặc chỉnh sửa

dữ liệu bổ sung, gây ra chi phí tài chính và trì hoãn tính khả dụng của dữ liệu.

Một khía cạnh quan trọng của nội dung dữ liệu là tính liên tục theo thời gian bao gồm các biến và định nghĩa. Phạm vi và nội dung trong các hệ thống dữ liệu hành chính có thể bị gián đoạn do những thay đổi về luật pháp, quy định, thực tiễn hành chính hoặc phạm vi của chương trình (Brackstone, 1987). Ví dụ, các sửa đổi của luật về mức lương tối thiểu và tối đa làm cho mức lương cơ bản hàng tháng tối thiểu đối với người lao động được yêu cầu bởi bất kỳ pháp nhân hoặc thể nhân nào, trong nước hoặc nước ngoài. Những thay đổi như vậy có tác động đáng kể đến việc sử dụng dữ liệu thống kê để so sánh giữa các thời kỳ.

Johnson và Moore (2008) đã khuyến cáo, hệ thống dữ liệu hành chính cũng không thể đảm bảo chất lượng dữ liệu hoàn hảo. Thông tin có thể quan trọng đối với các nhà thống kê, nhưng ít quan trọng hơn cho các mục đích hành chính, thường được báo cáo và xử lý không hoàn hảo. Ở đây chúng ta có thể đề cập đến sự lựa chọn nghề nghiệp của người khai báo tiền lương và tiền công. Biến chỉ ra loại nghề nghiệp có tầm quan trọng thứ yếu đối với các cơ quan hành chính miễn là người đó thường xuyên tuyên bố đóng góp của mình. Tuy nhiên, biến này có tầm quan trọng đặc biệt đối với việc sản xuất số liệu thống kê về mức lương trung bình theo nhóm ngành nghề. Các biến khác được sử dụng chủ yếu làm thông tin thứ cấp hoặc bổ sung có thể có chất lượng thấp hoặc thậm chí không đầy đủ (như trong trường hợp ngày làm việc mà khai báo thường là tiêu chuẩn của 21 hoặc 22 ngày làm việc). Hiện tượng này cũng có thể xảy ra với dữ liệu được thu thập cụ thể cho mục đích thống kê sử dụng các kênh hành chính hiện có, chẳng hạn như trong trường hợp phân loại hoạt động doanh nghiệp trong Nace Rev.2 được thu thập bởi các cơ quan hành chính cho tài khoản của

Istat. Các biến này có thể có chất lượng thấp hơn nếu mức độ ưu tiên của chúng không quá cao đối với cơ quan quản lý hoặc đơn vị cung cấp thông tin (Jensen, 1987). Một vấn đề khác được Johnson và Moore (2008) chỉ ra là độ tin cậy của dữ liệu có thể bị ảnh hưởng nếu thông tin được cung cấp cho tổ chức thuế có thể gây ra lợi nhuận hoặc tổn thất cho chủ thể khai báo. Ngoài ra, do dữ liệu được thu thập và xử lý cho mục đích hành chính thường được ưu tiên hơn so với yêu cầu thống kê, lượng xử lý cần thiết để cung cấp dữ liệu hành chính phù hợp cho mục đích thống kê có thể ảnh hưởng đến thời gian các dữ liệu này được cung cấp cho các nhà thống kê, các tác giả tranh luận.

Johnson và Moore nói, nhiều vấn đề nêu trên được giải quyết tốt nhất và được giải quyết thông qua các cuộc điều tra. Tuy nhiên, các tác giả nhận thấy rằng các vấn đề nội dung và xác nhận khác của một số loại xuất hiện trong dữ liệu điều tra. Vấn đề chính ở đây là bản chất tự nguyện của các phản hồi đối với các cuộc điều tra so với nghĩa vụ pháp lý tham gia vào các chương trình dữ liệu hành chính. Người trả lời cần được thuyết phục để cung cấp thời gian của họ và thông tin cần thiết mặc dù thực tế là không có hậu quả nếu họ từ chối và không có lợi ích nếu điều tra được thực hiện. Tuy nhiên, nếu một người được hỏi đồng ý tham gia cuộc điều tra, có thể anh ta hoặc cô ta từ chối trả lời các câu hỏi theo cách thực tế của người Hồi giáo.

Johnson và Moore (2008) tuyên bố, đối với những người trả lời đồng ý tham dự và trả lời tất cả các câu hỏi điều tra, sai số đo lường vẫn là mối quan tâm đối với dữ liệu điều tra. Những người được hỏi có thể đắn đo khi trả lời các câu hỏi hoặc họ có thể gặp khó khăn khi nhớ các sự kiện trong quá khứ. Các sai số đo lường điển hình khác bao gồm làm tròn số tiền, hiểu sai câu hỏi và thay đổi câu trả lời do lo ngại về dữ liệu được tiết lộ

➤➤➤ THÔNG KÊ QUỐC TẾ VÀ HỘI NHẬP

hoặc mong muốn bảo vệ quyền riêng tư. Nhiều nghiên cứu tồn tại liên quan đến các phép đo sai số và ảnh hưởng của chúng đối với dữ liệu quan sát (Lessler và Kalsbeek, 1992). Mặc dù đúng là đối với dữ liệu hành chính, việc không phản hồi không phải là vấn đề thiết yếu, không rõ liệu hồ sơ hành chính luôn chính xác hơn dữ liệu quan sát hay không, báo cáo của các tác giả. Một ví dụ về số lượng nhân viên được doanh nghiệp khai báo; một số công ty cố ý có thể khai báo số lượng nhân viên thấp hơn vào báo cáo của họ để giảm nghĩa vụ thuế. Các cá nhân tương tự có thể báo cáo giá trị thực trong việc trả lời bảng hỏi vì không có hậu quả pháp lý nếu giá trị thực được nêu. Một vấn đề nội dung khác đối với dữ liệu điều tra là tính kịp thời của dữ liệu. Trong khi nhiều cuộc điều tra đơn giản được thực hiện với tần suất thường xuyên, hàng tháng hoặc hàng quý, hầu hết các cuộc điều tra phức tạp nhất xảy ra hàng năm hoặc thậm chí hiếm hơn. Chi phí và các hạn chế tài nguyên khác là các yếu tố chính trong việc sử dụng kịp thời dữ liệu điều tra. Một vấn đề nội dung cuối cùng cho các cuộc điều tra được xây dựng bởi Johnson và Moore (2008) là xác nhận dữ liệu. Theo các tác giả đôi khi có thể tiến hành các nghiên cứu xác nhận sau khi hoàn thành một cuộc điều tra và các nghiên cứu này thêm chi phí bổ sung cho điều tra hoặc xác thực các biến dữ liệu được chọn có thể được thực hiện bằng các nguồn bên ngoài như kiểm duyệt hoặc hồ sơ hành chính, nhưng thường không có nguồn xác nhận.

3. Vấn đề riêng tư

Trong công việc của họ, Johnson và Moore (2008) coi quyền riêng tư dữ liệu là một vấn đề rất quan trọng. Các tác giả giải thích rằng bất kỳ việc sử dụng dữ liệu hành chính cho mục đích nghiên cứu nên tính đến các luật bảo vệ sự riêng tư của dữ liệu. Việc nghiên cứu dữ liệu hành chính thường bị giới

hạn trong việc sử dụng trong phạm vi nhiệm vụ của cơ quan và chỉ nên được thực hiện bởi những người làm việc cho cơ quan với tư cách là nhân viên, nhà thầu hoặc theo Bản ghi nhớ cho phép nhân viên của các tổ chức khác nhau trao đổi dữ liệu. Cách công chúng nhận thức bảo vệ quyền riêng tư đối với dữ liệu của họ có tác động trực tiếp đến tính liên tục của các mức độ khai báo. Thông thường, do các yếu tố này, dữ liệu có sẵn không chứa các biến xác định. Ví dụ, trong trường hợp dữ liệu cá nhân từ nguồn hành chính, các biến trực tiếp xác định chủ thể bị thiếu. Tất nhiên trong một kịch bản khác, sự sẵn có của các biến này có thể dẫn đến việc sử dụng thông kê rộng hơn và kết hợp dữ liệu từ các nguồn khác nhau.

Tuy nhiên, các tác giả nhấn mạnh rằng bảo mật dữ liệu có tầm quan trọng lớn đối với thành công hiện tại và tương lai của bất kỳ kỳ quan sát và đăng ký hành chính nào. Nếu các đối tượng không tin rằng dữ liệu của họ được bảo vệ đầy đủ, tỷ lệ trả lời và chất lượng dữ liệu tổng thể sẽ bị biến dạng. Luật riêng tư và quyền riêng tư cung cấp sự bảo vệ đáng kể chống lại việc lạm dụng dữ liệu tiềm năng (Johnson và Moore, 2008).

4. Dữ liệu việc làm và tiền công

Cả hai nguồn hành chính và điều tra (như điều tra thống kê ngắn hạn hàng quý) cung cấp thông tin quan trọng với tần suất hàng quý liên quan đến số lượng nhân viên và quỹ lương. Dữ liệu này được sử dụng để tính toán tiền lương và chỉ số tiền lương và chỉ số việc làm. Hồ sơ hành chính có thông tin nhân khẩu học phong phú hơn về dữ liệu cá nhân và chi tiết về đóng góp xã hội và sức khỏe. Mặt khác, dữ liệu điều tra bị hạn chế hơn, chỉ bao gồm số lượng nhân viên và tiền lương và quỹ tiền lương của doanh nghiệp được điều tra.

Những thay đổi quan trọng nhất giữa hai nguồn, theo lý thuyết đã được thảo luận ở trên, liên quan đến đơn vị điều tra, phạm vi tổng thể và cỡ mẫu. STS là một cuộc điều tra hàng quý trong đó đơn vị được điều tra là doanh nghiệp và các biến chính là: Doanh thu thuần, sản xuất công nghiệp, sản xuất xây dựng, số lượng nhân viên trung bình, quỹ tiền lương và tiền công, giá sản xuất, giá nhập khẩu, chi phí xây dựng (Istat, 2017). Tất cả các chỉ tiêu được sản xuất được thể hiện trong các chỉ số, trong các thay đổi hàng năm và hàng quý. Đơn vị nghiên cứu cho nguồn hành chính luôn là cá nhân và các chỉ số lương trung bình được thể hiện dưới dạng giá trị tuyệt đối. Quy mô của mẫu STS bị giới hạn do tần suất và chi phí, và điều tra đó không bao gồm tất cả các hoạt động kinh tế, bỏ qua việc đánh giá các hoạt động nông nghiệp (ngành A), các dịch vụ tài chính và bảo hiểm (ngành K), bất động sản (ngành L), hành chính công (ngành O), giáo dục và y tế (ngành P&C), cũng như các hoạt động nghệ thuật, giải trí, các dịch vụ và hoạt động khác của các tổ chức quốc tế (ví dụ, ngành R, S, T, U) nằm ngoài phạm vi điều tra. Điều này có nghĩa là thông tin hàng quý từ cuộc điều tra về việc làm và tiền lương bị thiếu cho các ngành này. Mặt khác, thông tin từ nguồn hành chính bao gồm các cá nhân và doanh nghiệp trong tất cả các hoạt động kinh tế. Sự thay đổi trong phương pháp luận có tác động trực tiếp đến các ước tính xuất phát từ mỗi nguồn. Ngoài ra, ước tính STS không đặc biệt tập trung vào ước tính tiền lương trung bình và thiếu thông tin nhân viên chi tiết (ví dụ, việc làm toàn thời gian hoặc bán thời gian, việc làm kép, loại đóng góp,...) khiến không thể áp dụng phương pháp tương tự như được sử dụng để ước tính mức lương trung bình từ nguồn hành chính. Hơn nữa, dữ liệu từ các cuộc điều tra phải tuân theo quy trình, trong khi việc đánh giá từ nguồn hành chính là đơn giản.

5. Kết luận

Ngày nay, để thỏa mãn nhu cầu ngày càng tăng từ người dùng số liệu thống kê chất lượng tốt, cho phép đo lường nhanh hơn các hiện tượng mới. Đồng thời, nhu cầu của những người dùng này phù hợp với nhu cầu của môi trường hiện nay mà gánh nặng đặt lên các doanh nghiệp và công dân giảm dần (Laux, Baigorri, & Radermacher, 2009). Do đó, việc sử dụng và kết hợp dữ liệu hành chính hoặc thứ cấp của các nhà thống kê được xem là một điều cần thiết trong thời đại ngày nay, nhưng nó cũng đi kèm với một số thách thức. Một số chỉ số, chẳng hạn như các chỉ số được thảo luận ở trên, có thể được tạo bằng dữ liệu có thể được lấy từ cả hai nguồn hành chính và thống kê, nhưng sự khác biệt về cấu trúc cơ bản giữa hai nguồn này, cũng như thay đổi trong phương pháp áp dụng, dẫn đến sự khác biệt trong ước tính và, tất nhiên, kết quả cuối cùng thu được từ chúng. Những thay đổi này có mặt trong hầu hết tất cả các khía cạnh của chất lượng, chẳng hạn như tính phù hợp, tính chính xác, tính kịp thời, khả năng tiếp cận, khả năng so sánh và thời gian. Vì lý do này, người dùng nên biết về những thay đổi này tại thời điểm sử dụng các ước tính từ các nguồn khác nhau và nên hiểu nguồn gốc của dữ liệu, việc thu thập và sử dụng của họ, để tránh nhầm lẫn và hiểu lầm. Điều này cho phép họ chọn các chỉ số phù hợp nhất với mục tiêu học tập của họ (Laux, Baigorri, & Radermacher, 2009). Không chỉ là các nguồn cạnh tranh, dữ liệu hành chính và điều tra nên được coi là nguồn bổ sung. Như Kapteyn & Ypma (2007) đã nói, câu hỏi liệu các nguồn lực hành chính hay quan sát cho thấy sự thật có phải là một câu hỏi triết học hay không.

Anh Tuấn (dịch)

Nguồn: www.Istat.gov.al/media/4271/7-difference-between-survey-and-administrative-data.pdf