

SỬ DỤNG MÔ HÌNH LIÊN TỤC TRONG DỮ LIỆU HÀNH CHÍNH ĐỂ XÁC ĐỊNH DÂN THƯỜNG TRÚ TẠI Ý

Tóm tắt:

Những nỗ lực mạnh mẽ ở cấp độ vi mô trong việc tích hợp các nguồn số liệu thống kê khác nhau cùng với sự gia tăng số lượng các kho lưu trữ hành chính dẫn đến sự thay đổi lớn trong các quy trình mà Viện Thống kê quốc gia áp dụng để sản xuất ra các số liệu thống kê dân số. Viện Thống kê Quốc gia Ý (Istat) đang lên kế hoạch thiết kế mới cho cuộc Tổng điều tra dân số tiếp theo dựa trên sự kết hợp thuận tiện giữa các dữ liệu hành chính và điều tra.

Một cơ sở dữ liệu chuyên đề đã được tạo ra để nghiên cứu cách sử dụng các nguồn hành chính có thể cải thiện chất lượng và thông tin về đăng ký dân số: Các nguồn được tích hợp là các Sổ đăng ký dân số chính thức của thành phố cùng với các tài liệu lưu trữ hành chính từ thị trường lao động, giáo dục, dữ liệu về thu nhập và thuế.

Nghiên cứu chỉ ra cách tích hợp dữ liệu này vào sổ đăng ký, từ đó cho phép khám phá những thông tin liên quan mới về dân số như: Các nhóm cá nhân được xác định bởi những mẫu nổi bật khi phân tích "hồ sơ sẵn có" ở nhiều nguồn khác nhau và theo thời gian có thể là mối quan tâm rất lớn đối với các nghiên cứu dân số.

I. Giới thiệu

1. Trong nhiều năm, một mặt số lượng dân số dựa trên điều tra nhân khẩu học và Tổng điều tra dân số, mặt khác, số lượng dân số đô thị dựa trên Sổ đăng ký dân số. Trong quá khứ, sự tích hợp giữa các nguồn số liệu này được thiết lập ở mức tổng hợp và kết quả của Tổng điều tra được sử dụng để tăng tính chính xác của Sổ đăng ký dân số đô thị.

2. Ngày nay, những nỗ lực mạnh mẽ trong việc tích hợp cấp vi mô giữa các nguồn thống kê khác nhau cùng với sự sẵn có của một số lượng dữ liệu hành chính ngày càng tăng đưa ra sự thay đổi lớn trong sản xuất số liệu dân số.

3. Tổng điều tra dân số vẫn là bộ sưu tập số liệu thống kê lớn nhất và quan trọng nhất để cung cấp số liệu dân số ở các đơn vị địa lý nhỏ nhất. Trong khi hầu hết các nước tiên tiến thống kê vẫn sử dụng kế hoạch

truyền thống thông qua việc thống kê toàn bộ các đơn vị dân cư và nhà ở (Mỹ và Canada), thì cũng có một số lượng ngày càng tăng các quốc gia dựa vào sổ đăng ký thống kê. Điều tra dựa trên cơ sở đăng ký có thể sử dụng dữ liệu đăng ký độc quyền, như trường hợp khu vực Xcăng-đi-na-vi (Hà Lan, Thụy Điển, Đan Mạch, Phần Lan và Na Uy) hoặc sử dụng kết hợp cả hai nguồn là sổ đăng ký và dữ liệu điều tra mẫu trong Khung "Tổng điều tra dân số" (tức là Tổng điều tra dân số năm 2011 của Tây Ban Nha).

4. Istat đang lên kế hoạch Tổng điều tra dân số cho vòng điều tra dân số tiếp theo, thuận tiện trong việc tích hợp dữ liệu hành chính và điều tra và khai thác tiềm năng thông tin mới này.

5. Thêm các vấn đề quan trọng (sinh đẻ, tử vong, di cư trong nước và quốc tế) vào số liệu thống kê Tổng điều tra dân số, Istat đã tính toán sổ đăng ký thống kê, gọi là

➤ ➤ ➤ THÔNG KÊ QUỐC TẾ VÀ HỘI NHẬP

"ANagrafe Virtuale Statistica" (ANVIS). Sổ đăng ký thống kê đảm bảo chất lượng cao hơn Sổ đăng ký hành chính của thành phố và là một phần quan trọng cho khung sản xuất dựa trên sổ đăng ký.

6. Từ năm 2015, nhiều cuộc thử nghiệm đã được thực hiện tại Istat về việc sử dụng các nguồn hành chính để tuân thủ định nghĩa về dân số thường trú trong Quy chế Liên minh châu Âu (số 1260/2013). Theo quy định này, "nhân khẩu thường trú" là nơi người đó sống 12 tháng trước ngày tham chiếu. Những người ở địa điểm ít hơn 12 tháng trước điều tra và có ý định ở lại đó ít nhất 12 tháng cũng được coi là nhân khẩu thường trú.

7. Mục đích của nghiên cứu này là để đánh giá liệu hồ sơ từ các nguồn hành chính có thể đáp ứng các yêu cầu của định nghĩa quốc tế và cho phép tính toán dân số thường trú tại Ý hay không. Mục tiêu này đòi hỏi phải lựa chọn và đánh giá dữ liệu hành chính với mục đích để nhận diện các mô hình và sự liên kết.

8. Một vấn đề khác có liên quan là xác minh liệu việc tích hợp dữ liệu vào sổ đăng ký thích hợp có thể đánh giá các nhóm người được xác định bằng các mô hình đang nổi lên khi phân tích "hồ sơ sẵn có" ở các nguồn khác nhau.

II. Hệ thống tích hợp các hồ sơ hành chính

9. Để quản lý số lượng các bộ dữ liệu hành chính ngày càng tăng và để tối đa hóa lợi ích, Istat đã xây dựng một hệ thống tích hợp các nguồn hành chính có sẵn, tên là SIM (Integrated Microdata). Khi một tập tin hành chính mới được nhập vào hệ thống này, quá trình nhận dạng sẽ xác định bất kỳ cá thể hoặc đơn vị kinh tế nào có trong dữ liệu và chỉ định mã số nhận dạng vĩnh viễn và duy

nhất (ID): Nếu đơn vị đã có trong cơ sở dữ liệu Istat, ID này là cùng một đơn vị đã được chỉ định trong quá khứ (diBella và Ambroselli, 2014). Sau đó, bắt đầu từ cơ sở này, có thể xây dựng các cấu trúc dữ liệu cụ thể cho các quy trình thống kê và để tạo cơ sở dữ liệu theo từng chủ đề.

10. Trong tất cả các kho lưu trữ, SIM bao gồm dữ liệu từ ANVIS, giấy phép lưu trú, dữ liệu liên quan đến công nhân và người lao động tự do, sinh viên có trình độ học vấn, sinh viên đại học, người về hưu, hồ sơ không trợ cấp lương hưu và dữ liệu cá nhân về thu nhập và thuế. Các dữ liệu tổng hợp này đã được sử dụng để tạo ra một cơ sở dữ liệu theo chủ đề để phân tích các nguồn hành chính có thể cải thiện chất lượng và thông tin các Sổ đăng ký dân số như thế nào (Chieppa et al, 2016). Mã số nhận dạng vĩnh viễn cho phép liên kết các biến số xã hội và kinh tế của cá nhân, hộ gia đình và các đơn vị kinh tế đến nơi mọi người thực hiện các hoạt động hay việc sử dụng thời gian của mình.

11. Liên kết Sổ đăng ký dân số chính thức với các nguồn hành chính cụ thể (Sổ đăng ký lao động và đăng ký kinh doanh, đăng ký thuế, sổ tiết kiệm thu nhập, nghỉ hưu và không hưởng lương hưu, giấy phép lưu trú) có thể giúp xác định các nhóm tương ứng với định nghĩa quốc gia hay quốc tế về "dân số thường trú". Các sai số thiếu phạm vi có thể được điều chỉnh bằng cách sử dụng hồ sơ sẵn có cá nhân trên lãnh thổ Ý; trong khi sự vắng mặt của hồ sơ cho người dân trong Sổ đăng ký số dân có thể làm cản trở cho việc quá phạm vi.

12. Để khai thác các nguồn hành chính, Cơ quan Thống kê Ý sử dụng lược đồ Kiến thức khám phá từ Cơ sở dữ liệu dựa trên kỹ thuật "học-từ-các ví dụ" (learning-from-examples), bao gồm lưu trữ và truy cập dữ

liệu, các thuật toán mở rộng cho các bộ dữ liệu lớn. Quá trình này dẫn đến xác định các mẫu và nhóm.

III. Xác định mô hình liên tục từ các hồ sơ hành chính

13. Với mục tiêu khám phá hồ sơ sẵn có phù hợp với định nghĩa quốc tế về dân số thường trú ở Ý, dữ liệu hành chính đã được lựa chọn và đánh giá với mục đích để nhận ra các mẫu và kết hợp các hồ sơ. Trước khi phân tích các bản ghi từ dữ liệu hành chính, cần lựa chọn khoảng thời gian tham chiếu, sự hiện diện của một cá nhân trên lãnh thổ liên kết chắc chắn với một khoảng thời gian. Việc chọn thời kỳ tham chiếu là một điểm rất quan trọng khi xác định và phân tích hồ sơ sẵn có ở Ý, khoảng thời gian càng dài thì việc đánh giá trọng lượng của bản ghi càng dài, liên tục hoặc gián đoạn theo thời gian, số lượng và sự lặp lại quy trình di cư, di chuyển việc làm, học tập, v.v...

14. Để đáp ứng các yêu cầu về định nghĩa dân số thường trú (thậm chí xem xét khía cạnh có ý định ở lại ít nhất 12 tháng), Istat đã nghiên cứu khoảng thời gian là 24 tháng. Sau đó, xem xét 1 ngày tham chiếu, ví dụ ngày 31/12/2012, phân tích hồ sơ được thực hiện vào 12 tháng trước và 12 tháng sau ngày tham chiếu.

15. Mỗi dấu hiệu có thể được quy cho một cá nhân cụ thể và đến một khu vực địa lý nhất định. Trong thời gian tham chiếu, nếu một hồ sơ được phát hiện trong Sổ đăng ký lao động và một trong Sổ đăng ký giáo dục, nhưng cả hai đều được đặt trong đô thị, chúng tôi sẽ có một hồ sơ duy nhất từ đô thị cụ thể đó. Bản ghi này được gắn nhãn với một thuộc tính cho phép theo dõi cá nhân trong cả hai nguồn ban đầu, và cũng là một thuộc tính liên quan đến thời gian hiện diện

được thể hiện dưới dạng công việc hoặc hoạt động học tập.

16. Hồ sơ liên quan đến lao động và giáo dục được xem là đáng tin cậy nhất, bao gồm thời gian và loại hình hoạt động. Họ cho phép biết từng cá nhân, theo tháng, và những gì họ đang làm. Gom cụm thành chuỗi các bản ghi này, Bảng 1 cho thấy các mô hình có hồ sơ liên tục và những người có sự hiện diện không liên tục ở Ý. Mô hình "Sự sẵn có liên tục trong năm 2012 và 2013" (1) tương ứng với sự liên tục trong suốt thời kỳ này.

17. Sự sẵn có liên tục, chủ yếu vào năm 2012 hoặc 2013 (mẫu 2 và 3), đặc trưng cho những người có mặt trong ít nhất 12 tháng, và những người đã thay đổi, bắt đầu, hoặc ngừng hoạt động, hoặc đã di chuyển. Sự sẵn có liên tục trong tháng 12 năm 2012 nhóm lại những người có ít nhất 12 tháng có mặt liên tục từ năm 2012 đến năm 2013 (mẫu 4). Mẫu 5 nhóm các hồ sơ không liên tục kéo dài hơn 12 tháng; mẫu 6 bao gồm "hồ sơ theo mùa" chỉ xuất hiện mỗi năm một lần; mẫu 7 kết hợp các bản ghi rải rác ít hơn 12 tháng. Các hồ sơ từ mẫu 8-12 không thể khai thác, bởi vì họ không thuộc phạm vi định nghĩa về nơi ở thường trú. Hồ sơ của họ chỉ hiển thị 1 tháng có mặt (mẫu 8), biến mất trước tháng 12/2012 (mẫu 9), xuất hiện sau tháng 12/2012 (mẫu 11) hoặc lặp lại ngẫu nhiên vào năm 2012 hoặc năm 2013, có mặt ít hơn 12 tháng (mẫu 10, 11 và 12).

18. Vì vậy, hồ sơ có thể được sử dụng để lấy các biến mới có liên quan đến các cá nhân có liên quan và loại điều kiện sống của họ ở Ý. Cụ thể hơn đối với số lượng dân số, thông tin mới này có thể xác định các trường hợp có sự sẵn có thường trú tương ứng với định nghĩa dân số thường trú và khái niệm theo quy định quốc tế, không phải lúc nào cũng tương ứng với nội dung ghi trong Sổ đăng ký số dân.

➤ ➤ ➤ THÔNG KÊ QUỐC TẾ VÀ HỘI NHẬP

19. Các biến số nhân khẩu học, đặc biệt là giới tính, tuổi và quốc gia về quyền công dân cũng như vị trí của hồ sơ trên lãnh thổ

đã chứng tỏ là các biến số rất quan trọng trong việc xác định các tiểu vùng phụ.

Bảng 1: Sơ đồ mô hình liên tục trong hoạt động làm việc và nghiên cứu

Thời gian từ tháng 1/2012 đến tháng 12/2013																								Loại sẵn có trong Sổ đăng ký lao động và giáo dục			
J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D				
																									1	liên tục trong năm 2012-2013	} Hồ sơ liên tục
																									2	liên tục, chủ yếu năm 2012	
																									3	liên tục, chủ yếu năm 2013	
																									4	liên tục trong tháng 12 năm 2012	
																									5	liên tục cùng với tạm dừng	
																									6	theo mùa	} Hồ sơ gián đoạn
																									7	rải rác trong năm 2012 và 2013	
																									8	chỉ trong tháng 12 năm 2012	} Hồ sơ không sử dụng được
																									9	khoảng thời gian ngắn năm 2012	
																									10	rải rác năm 2012	
																									11	khoảng thời gian ngắn năm 2013	
																									12	rải rác năm 2013	

IV. Dân số được phân loại tính nhất quán của sổ đăng ký

20. Chúng tôi phân loại các trường hợp được ghi trong cơ sở dữ liệu thử nghiệm dựa trên tính nhất quán của các bản ghi (Bảng 2). Đầu tiên chúng ta liên kết Sổ đăng ký dân số, hồ sơ hoạt động từ Sổ đăng ký lao động và giáo dục và giấy phép lưu trú, sau đó chúng tôi cô lập những người về hưu hoặc những người được hưởng các quyền lợi khác từ Sổ đăng ký lao động và giáo dục; thứ ba, tờ khai thuế trả về những người có thể biện minh cho sự sẵn có của họ ở Ý.

21. Bước đầu tiên, sự liên kết này cho phép chúng tôi mô tả 4 nhóm: Nhóm A gồm 36,6 triệu người được ghi trong Sổ đăng ký dân số và Sổ đăng ký lao động và giáo dục. Nhóm B gồm 24,5 triệu người được ghi trong Sổ đăng ký dân số nhưng không có trong Sổ đăng ký lao động và giáo dục. Nhóm C gồm 1,1 triệu người không có trong Sổ đăng ký dân số nhưng lại được ghi trong Sổ đăng ký lao động và giáo dục. Nhóm D gồm 351 nghìn người chỉ đăng ký trong giấy phép để lưu trữ.

Bảng 2: Quy trình thực hiện và số lượng dân số theo điều kiện cư trú hợp lệ hoặc có thể tại Ý vào ngày 1/1/2013

Đơn vị tính: Nghìn người

Bước 1					Bước 2				Bước 3		
Số đăng ký dân số	Số đăng ký lao động và giáo dục	Giấy phép lưu trú	Nhóm	Số lượng	Đăng ký nghỉ không lương	Nhóm	Số lượng	Số đăng ký thuế	Nhóm	Số lượng	
61.068	37.704	3.378			20.764			26.649			
Các loại hồ sơ					Các loại hồ sơ			Các loại hồ sơ			
Có	Có	-	A	36.618							
Có	Không	-	B	24.450	Có	E	14.485	Có	G	6.939	
					Không	F	9.965				Không
Không	Có	-	C	1.086							
Không	Không	Có	D	351							

Trong đó: là số dân đủ điều kiện hợp lệ, là số dân có tiềm năng

22. Trong Bước 2, chúng tôi liên kết nhóm B với Số đăng ký người đã nghỉ hưu, người nhận trợ cấp thai sản hay trợ cấp thất nghiệp. Từ mỗi liên kết này, chúng tôi phân biệt liệu các cá nhân có được ghi ở một trong hai sổ đăng ký (Nhóm E gồm 14,5 triệu người) hay không (Nhóm F gồm 10 triệu người).

23. Trong Bước 3, chúng tôi liên kết nhóm F với Số đăng ký thuế, nhằm xác định những người có tài sản được hỗ trợ bởi các cá nhân có thu nhập chịu thuế (Nhóm G gồm 6,9 triệu người) hoặc những người được ghi trong Sổ đăng ký dân số mà không có trong nguồn hồ sơ nào khác có sẵn (Nhóm H gồm 3 triệu người).

24. Các thành viên thuộc nhóm A, E, và G là những người thường trú, trong khi những người thuộc nhóm C, D, và H là những người "có tiềm năng" ở Ý. Do đó, từ thời điểm này, chúng tôi tập trung vào nhóm C (1,1 triệu người), kết hợp khả năng có thể bao phủ các Sổ đăng ký dân số. Bảng 3 phân biệt nhóm C theo loại hồ sơ và thời gian:

409.157 người (Nhóm C1, C2, C3) có hồ sơ liên tục, đa số là người nước ngoài (90%), và có hoạt động chuyên môn (78%), trong khi những người khác đang ở trường (58.000 người) hoặc sinh viên (33.000 người).

Bảng 3: Nhóm C - Những người không có trong Sổ đăng ký dân số nhưng có ở các nguồn hành chính khác ngày 1/1/2013

Đơn vị tính: Người

Nhóm	Các nguồn và hồ sơ	Giá trị
C1	Hồ sơ liên tục về người lao động	318.159
C2	Hồ sơ liên tục về sinh viên đại học	32.671
C3	Hồ sơ liên tục về học sinh tiểu học và trung học	58.327
C4	Hồ sơ không liên tục về sự sẵn có	266.763
C5	Hồ sơ không có khả năng sử dụng hoặc không có thông tin hàng tháng	410.242
Tổng		1.086.162

➤ ➤ ➤ THÔNG KÊ QUỐC TẾ VÀ HỘI NHẬP

25. Hồ sơ không liên tục nhưng vẫn có thể khai thác (Nhóm C4) là 266.763 người, chủ yếu là người nước ngoài (90%). Hồ sơ không liên tục và không thể khai thác được do thiếu thông tin là khoảng 410 nghìn người (Nhóm C5).

V. Tiếp cận theo chiều dọc và kết quả sơ bộ

26. Một phân tích chi tiết hơn về các nhóm có nguy cơ thiếu phạm vi trong Sổ đăng ký dân số đã nêu bật một số chủ đề quan trọng khi chúng ta sử dụng cách tiếp cận theo chiều dọc. Phân tích tất cả các cá nhân không đăng ký trong Sổ đăng ký dân số vào năm 2012 và phân tích họ trong ba năm tới, có thể nhận thấy 3 nhóm sau (Bảng 4)¹.

(a) Những người có hồ sơ sẵn có vào năm 2012² ghi trong Sổ đăng ký dân số trong hai năm tới. Nhóm này bao gồm hơn 165.000 người cho thấy hồ sơ có mặt liên tục ở Ý

(khoảng 34% số người có khả năng bao phủ), và 76.000 người khác có hồ sơ không liên tục;

(b) Những người có hồ sơ liên tục có mặt từ năm 2012 đến năm 2014 nhưng chưa bao giờ ghi trong Sổ đăng ký dân số. Nhóm này gồm hơn 180.000 người (khoảng 37%) với hồ sơ liên tục, và khoảng 90.000 người với hồ sơ không liên tục. Cả hai loại này đại diện cho "cốt lõi" trong thiếu phạm vi của Sổ đăng ký dân số;

(c) Những người có hồ sơ chỉ có mặt vào năm 2012 và 2013 (không có năm 2014) mà không có trong Sổ đăng ký dân số là hơn 140.000 người (khoảng 29%) với hồ sơ liên tục và 317.000 người khác (khoảng 66%) không có hồ sơ liên tục. Cả hai loại này có thể chắc chắn đã rời khỏi đất nước hoặc có thể đang trong tình trạng bấp bênh vì mất việc làm.

Bảng 4: Số lượng người phân tích theo chiều dọc có trong hồ sơ sẵn có tại Ý giai đoạn 2012-2014

Đơn vị tính: Nghìn người

Hồ sơ trong Sổ đăng ký lao động và giáo dục			Hồ sơ trong Sổ đăng ký dân số			Hồ sơ liên tục	Hồ sơ gián đoạn	Thông tin các nhóm
1/1/2012	1/1/2013	1/1/2014	1/1/2012	1/1/2013	1/1/2014	Giá trị	Giá trị	
Có	Có	Có	Không	Có	Có	165,4	76,0	Những người có hồ sơ sẵn có vào năm 2012 được ghi trong Sổ đăng ký dân số trong 2 năm tới
						33,8%	15,7%	
Có	Có	Có	Không	Không	Không	182,5	89,7	Những người có hồ sơ sẵn có từ năm 2012-2014 không được ghi trong Sổ đăng ký dân số
						37,2%	18,6%	
Có	Có	Không	Không	Không	Không	142,0	317,6	Những người có hồ sơ sẵn có chỉ trong năm 2012 và 2013 mà không có trong Sổ đăng ký dân số
						29,0%	65,7%	
						489,9	483,3	Tổng số

¹ Các cá nhân không khai thác được đã bị loại khỏi phân tích theo chiều dọc (nhóm C5 của Bảng 3).

² Nên xem xét đối với mỗi năm (2012, 2013, 2014), chúng tôi quan sát hồ sơ dữ liệu hành chính từ 12 tháng trước và 12 tháng sau ngày tham chiếu 1/1.

THỐNG KÊ VÀ CUỘC SỐNG ◀◀◀

27. Khi xem xét thiếu phạm vi bằng phương pháp tiếp cận theo chiều dọc, phân tích cho thấy các kết quả có liên quan với nhau. Trước hết, các tiêu chí quan sát nguồn hồ sơ hành chính trong khoảng thời gian 24 tháng để đáp ứng được định nghĩa về dân số thường trú đang được thực hiện. Hơn nữa, việc phân cụm các cá nhân theo mô hình liên tục hoặc gián đoạn là một công cụ phân loại hữu ích để xác định sự tồn tại ổn định trên lãnh thổ, đặc biệt là đối với người nước ngoài.

28. Vị trí địa lý và quốc tịch cụ thể sẽ là điều cần thiết để xác định những người làm việc ở nước ngoài, có thể chấp nhận việc vắng mặt trong Sổ đăng ký dân số.

29. Tuy nhiên, phân tích theo chiều dọc cho thấy một số cá nhân không có hồ sơ liên tục có thể liên quan đến sự sẵn có ổn định trên lãnh thổ (khoảng từ 15-18%) và do đó cần phải cải thiện hồ sơ.

30. Hồ sơ trong Sổ đăng ký hành chính nâng cao nhận thức về dân số và cho thấy khi dân số đăng ký số liệu đã được liên kết với các nguồn hành chính khác, Istat có thể đánh giá chất lượng và tính chính xác của nguồn. Một số người nước ngoài hoặc người di cư không ghi vào Sổ đăng ký dân số, nhưng được ghi trong Sổ đăng ký lao động.

Lan Phương (dịch)

Nguồn: The Conference of European Statisticians, Group of Experts on Population and Housing Censuses, Nineteenth Meeting, Geneva, 4-6 October 2017;

https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2017/Meeting-Geneva-Oct/ECE_CES_GE.41_2017_13Rev1_ENG.pdf