

CHƯƠNG 1: GIỚI THIỆU VÀ PHƯƠNG PHÁP

1.1 GIỚI THIỆU

Cuộc Tổng điều tra dân số và nhà ở năm 2009 được thực hiện vào thời điểm 0 giờ ngày 1 tháng 4 năm 2009 theo Quyết định số 94/2008/QĐ-TTg của Thủ tướng Chính phủ ban hành ngày 10 tháng 7 năm 2008.

Kết quả sơ bộ của Tổng điều tra được công bố vào tháng 8 năm 2009. Tiếp sau đó kết quả điều tra mẫu được công bố vào tháng 12 năm 2009. Kết quả toàn bộ được phát hành vào tháng 7 năm 2010, tiếp sau đó là các báo cáo phân tích và chuyên khảo được soạn thảo và lần lượt được công bố.

Trong Tổng điều tra năm 2009 có lồng ghép điều tra mẫu 15%, thu thập nhiều thông tin chi tiết, trong đó có thông tin để ước lượng các số đo về mức độ sinh và chết của dân số Việt Nam.

Kết quả của cuộc Tổng điều tra năm 2009 và các cuộc Tổng điều tra trước đó cho thấy mức sinh của nước ta tiếp tục giảm và đã đạt mức sinh thay thế. Mức độ chết (tử vong) của trẻ sơ sinh cũng giảm liên tục và tuổi thọ trung bình tính từ khi sinh tăng.

1.2 MỤC ĐÍCH NGHIÊN CỨU VÀ NỘI DUNG

Một số nét chính về mức độ sinh và mức độ chết thu thập được trong Tổng điều tra đã được tính toán và công bố¹. Chuyên khảo này nhằm đưa ra bức tranh chi tiết hơn về mức độ sinh và mức độ chết của dân số Việt Nam trong những năm qua, sự thay đổi và khác biệt của các số đo nói trên theo các đặc trưng nhân khẩu học, đồng thời đưa ra các yếu tố có thể ảnh hưởng đến các số đo đó.

Nội dung của chuyên khảo này gồm 4 chương, như sau:

- Chương 1: Giới thiệu tóm tắt về các phương pháp sử dụng để ước lượng các số đo về mức sinh và mức chết;
- Chương 2: Trình bày các ước lượng của các số đo về mức độ sinh, thay đổi và khác biệt của các số đo đó theo các đặc trưng nhân khẩu học;
- Chương 3: Trình bày các ước lượng của các số đo về mức độ chết, sự thay đổi và khác biệt của chúng theo các đặc trưng nhân khẩu học;
- Chương 4: Các khuyến nghị chính sách để tiếp tục giảm bền vững mức sinh, giảm mức chết trẻ em, tăng tuổi thọ nhằm tạo một cơ sở cơ bản quan trọng về dân số cho công cuộc phát triển bền vững

1 Xem: Tổng điều tra dân số và nhà ở Việt Nam năm 2009, Các Kết quả Chủ yếu, Ban Chỉ đạo Tổng điều tra dân số và nhà ở Trung ương, Hà Nội, 6/2010.

1.3 NGUỒN SỐ LIỆU

Chuyên khảo này sử dụng số liệu về sinh và chết của điều tra mẫu 15% thu thập được trong Tổng điều tra 2009 và các phương pháp ước lượng gián tiếp để tính các số đo về mức độ sinh và chết, phản ánh cho thời kỳ 12 tháng trước thời điểm điều tra. Ngoài ra, số liệu đã được công bố của Tổng điều tra 1989 và 1999, và các số liệu, tài liệu liên quan của nước ta hoặc của các nước/khu vực khác cũng được dùng để so sánh đối chiếu. Các phương pháp ước lượng gián tiếp để tính các số đo về mức độ sinh và chết sử dụng trong chuyên khảo này được trình bày dưới đây:

1.4 PHƯƠNG PHÁP ƯỚC LƯỢNG

Do các dữ liệu về sinh và chết trong các cuộc điều tra mẫu, Tổng điều tra thường không cao do bị bỏ sót nên các nhà nhân khẩu học đã nghiên cứu xây dựng các kỹ thuật gián tiếp để ước lượng các chỉ tiêu phản ánh mức sinh và mức chết. Các kỹ thuật này cũng luôn luôn được làm mới và tin học hóa. Liên hợp quốc đưa ra khuyến nghị và cung cấp các kỹ thuật gián tiếp dưới dạng phần mềm.

1.4.1 Kỹ thuật gián tiếp ước lượng mức độ sinh

Kỹ thuật ước lượng gián tiếp là ước lượng tỷ suất sinh đặc trưng theo tuổi của phụ nữ trong tuổi sinh đẻ 15-49 từ số liệu về số con đã sinh tổng hợp theo tuổi người mẹ và mô hình của mức sinh theo tuổi của người mẹ, ghi nhận được trong 12 tháng trước điều tra.

Tỷ suất sinh đặc trưng theo tuổi thu thập được luôn thấp hơn (hoặc cao hơn) mức sinh thực bởi vì các sự kiện sinh thường bị bỏ sót (hoặc ghi trùng) trong điều tra. Nhà nhân khẩu học Brass² đã xây dựng một phương pháp, thường gọi là phương pháp P/F, để đánh giá và chỉnh tỷ suất sinh đã ghi nhận được bằng cách so sánh các tỷ suất đã ghi nhận được với số liệu số con sinh trung bình tính cho nhóm 5 độ tuổi của phụ nữ. Phương pháp P/F giả thiết rằng mức sinh là không đổi trong thời kỳ trước đây, mô hình của tỷ suất sinh đặc trưng theo tuổi ghi chép được (ký hiệu là ASFR) là đúng, và mức sinh tích lũy đối với nhóm phụ nữ trẻ theo số con đã sinh (CEB) là chính xác. Brass đơn giản cộng dồn và làm trơn số liệu ASFR ghi chép được dưới dạng số liệu số con sinh. Với giả thiết mức sinh không thay đổi, các số liệu đã được làm trơn (ký hiệu là ${}_nF_x$) là so sánh được với số con đã sinh ghi chép được (${}_nCEB_x$). Tỷ lệ ${}_nCEB_x / {}_nF_x$ đối với nhóm tuổi trẻ cung cấp hệ số điều chỉnh chấp nhận được để điều chỉnh các tỷ suất sinh đã ghi nhận.

Nhà nhân khẩu học Arriaga sau đó đã cải tiến phương pháp nói trên và mở rộng ra cho cả những trường hợp mức sinh đang thay đổi. Mô tả đầy đủ cách tiếp cận Brass và Arriaga về ước lượng tỷ suất sinh đặc trưng theo tuổi từ số liệu số con đã sinh ghi chép được và từ mô hình sinh theo tuổi trình bày trong công trình nghiên cứu của Arriaga.

Cơ quan Tổng điều tra của Hoa Kỳ đã tin học hóa kỹ thuật này bằng một bảng tính excel có tên gọi là bảng tính PFRATIO trong bộ các bảng tính phân tích dân số³.

2 Cẩm nang số 10: Các kỹ thuật gián tiếp về ước lượng nhân khẩu học, Nhà Xuất bản Khoa học Kỹ thuật, Hà Nội 1976.

3 Eduardo E. Arriaga, Phân tích dân số với máy vi tính, Cơ quan Tổng điều tra Hoa Kỳ, tháng 11 năm 1994 (Population Analysis with Microcomputer, Bureau of the Census, November 1994).

Dựa vào hệ số điều chỉnh “k” chọn được, tỷ suất sinh đặc trưng theo tuổi (ASFR), Tổng tỷ suất sinh (TFR) và Tỷ suất sinh thô (CBR) cũng được điều chỉnh một cách tương tự, cụ thể là:

$$ASFR_i^* = k \times ASFR_i, \text{ i là các nhóm 5 độ tuổi của khoảng tuổi 15-49,}$$

$$TFR^* = k \times TFR,$$

$$CBR^* = k \times CBR,$$

trong đó, $ASFR_i$, TFR , CBR là chỉ tiêu chưa điều chỉnh, còn $ASFR_i^*$, TFR^* và CBR^* là các chỉ số đã được điều chỉnh.

1.4.2 Kỹ thuật gián tiếp ước lượng mức độ chết

(i) Gián tiếp ước lượng tỷ suất chết trẻ em dưới 1 tuổi

Nhà nhân khẩu học Brass đã chứng minh rằng xác suất chết trong khoảng thời gian từ khi sinh đến độ tuổi a (ký hiệu là $q(a)$) có thể ước lượng theo công thức: $q(a) = {}_5M_x \cdot {}_5D_x$, trong đó ${}_5D_x$ là tỷ trọng trẻ em chết đối với phụ nữ nhóm tuổi (x,x+5) và ${}_5M_x$ hệ số đặc trưng theo tuổi, gọi là hệ số nhân, nó phụ thuộc vào chỉ số mô hình tuổi của mức sinh. Tỷ trọng trẻ em chết đối với phụ nữ nhóm tuổi 15-19, 20-24, 25-29, ..., 45-49 được sử dụng để tính $q(a)$ cho các giá trị tương ứng bằng 1, 2, 3, 5, 10, 15 và 20. Sau đó Sullivan⁴ chứng minh rằng kiểu quan hệ như thế cũng tồn tại khi số liệu được tính toán theo độ dài hôn nhân. Trong trường hợp này, độ dài hôn nhân 0-4 năm, 5-9 năm, ..., 30-34 năm tương ứng với $q(a)$ đối với các độ tuổi tương ứng là 2, 3, 5, 10, 15, 20 và 25. Để tái hiện, phương trình hồi quy được xây dựng liên quan đến hệ số nhân ${}_5M_x$ đối với các chỉ số của trật tự sinh. Có 9 tập riêng biệt phương trình hồi quy được ước lượng, 5 tập đầu cho từng mô hình Liên hợp quốc⁵ và 4 tập cuối cho từng mô hình của Coale và Demeny (hồi quy Trussell⁶). Thông qua tập thứ hai, các phương trình hồi quy cũng đã được xây dựng từ cùng một tập các biến độc lập, chúng ước lượng thời gian tương ứng với các giá trị $q(a)$. Các biến độc lập mà chúng ước lượng giá trị $q(a)$, cũng như thời gian tham chiếu, được tính toán từ số liệu đầu vào. Hơn nữa, đối với tỷ trọng trẻ em chết theo nhóm tuổi hoặc theo độ dài hôn nhân của phụ nữ, các biến cần được tính bằng tỷ lệ giữa số trẻ em sinh bình quân của phụ nữ ở nhóm tuổi đầu tiên hoặc nhóm độ dài hôn nhân đầu tiên và số liệu đó của nhóm tuổi thứ hai hoặc nhóm độ dài hôn nhân thứ hai, tỷ lệ giữa số trẻ em sinh bình quân của phụ nữ ở nhóm tuổi thứ hai hoặc nhóm độ dài hôn nhân thứ hai và số liệu đó của nhóm tuổi thứ ba hoặc nhóm độ dài hôn nhân thứ ba, và tuổi sinh con trung bình. Biến cuối cùng chỉ được sử dụng để tính toán dựa vào các mô hình của Liên hợp quốc; ước lượng gần đúng tuổi sinh con trung bình có thể được tính từ các trường hợp sinh trong 12 tháng trước điều tra và tuổi của người mẹ. Phương trình hồi quy được sử dụng để tính toán các ước lượng tỷ suất chết của trẻ em dưới 1 tuổi (${}_1q_0$), xác suất chết giữa 1 và 5 tuổi (${}_4q_1$), và tuổi thọ trung bình tính từ lúc sinh tương ứng với các giá trị $q(a)$ trong từng mô hình bảng sống mẫu (đối với cả hai giới). Cách ước lượng này đã được Liên hợp quốc tin học hóa bằng phần mềm có tên là QFIVE để ước lượng ${}_1q_0$, ${}_4q_1$ và tuổi thọ trung bình tính từ lúc sinh (viết tắt là e_0).

4 Sullivan, J. M. 1972. “Mô hình ước lượng xác suất chết trong thời gian từ khi sinh đến độ tuổi nào đó của trẻ em”, Nghiên cứu dân số, tập 26, số 1 (tháng 3 năm 1972), trang 77-99.

5 Palloni, A và L. Heliman, 1985, “Ước lượng lại các tham số cấu trúc để ước lượng mức độ chết của các nước đang phát triển”, Bản tin dân số của Liên hợp quốc, số 18, trang 10-33.

6 Cẩm nang số 10: Các kỹ thuật gián tiếp về ước lượng nhân khẩu học, Nhà Xuất bản Khoa học Kỹ thuật, Hà Nội 1976.

(ii) Ước lượng tỷ suất chết trẻ em dưới 5 tuổi

Tỷ suất chết trẻ em dưới 5 tuổi thường được viết tắt là ${}_5q_0$ và tính theo công thức sau:

$${}_5q_0 = {}_1q_0 + {}_4q_1 - {}_1q_0 \times {}_4q_1,$$

trong đó: ${}_1q_0$ là tỷ suất chết trẻ em dưới 1 tuổi; ${}_4q_1$ là xác suất chết giữa 1 và 5 tuổi được ước lượng gián tiếp như ở đã trình bày ở trên bằng phần mềm QFIVE.

(iii) Gián tiếp ước lượng mức độ sót của khai báo chết của toàn bộ dân số

Ngoài kỹ thuật ước lượng gián tiếp tỷ suất chết của trẻ em, hai phương pháp cân bằng tăng trưởng chung (GGB) và thể hệ chết giả định (SEG) cũng đã được áp dụng kết hợp để đánh giá mức độ đầy đủ của thông tin về chết của toàn bộ dân số.

Phương pháp cân bằng tăng trưởng chung: Năm 1975 nhà nhân khẩu học Brass đã đưa ra phương pháp GGB, thu được từ dân số ổn định được biểu thị bằng biểu thức quan hệ trực giác đối với mỗi một nhóm tuổi mở a+ của dân số đóng, tỷ lệ tham gia vào nhóm tuổi (b(a+)) bằng tỷ lệ tăng trưởng của nhóm (r(a+)) cộng với tỷ lệ ra khỏi nhóm tuổi (tỷ lệ chết) (d(a+)). Tất nhiên, điều tất yếu xảy ra đối với toàn bộ dân số đóng là tỷ lệ tăng trưởng bằng tỷ lệ sinh trừ tỷ lệ chết. Vì vậy,

$$r(a+) = b(a+) - d(a+) \quad \text{hay} \quad b(a+) = r(a+) + d(a+) \quad (1)$$

Đối với dân số đóng, tỷ lệ tăng trưởng là hằng số đối với tất cả các nhóm tuổi, do đó tỷ lệ tham gia vào nhóm và tỷ lệ chết phải quan hệ tuyến tính với nhau. Nếu biểu thị N(a) và N(a+) tương ứng là số tham gia vào nhóm (tức là số sinh vào tuổi a) và dân số của nhóm tuổi a, r là tỷ lệ tăng trưởng của dân số ổn định, và D(a+) là số chết từ tuổi a trở lên, ta có:

$$N(a)/N(a+) = r + (D(a)/N(a+)) \quad (2)$$

Nếu tỷ lệ tham gia vào nhóm được tính toán phân bố dân số chỉ sử dụng cách tiếp cận tương đối đơn giản, thí dụ N(a) bằng một phần năm của trung bình nhóm 5 độ tuổi nhỏ hơn và lớn hơn tuổi a, bất kỳ sai số phạm vi nào bất biến với tuổi phải được xoá bỏ, ngược lại, tỷ lệ chết, tính từ số chết theo tuổi và dân số theo tuổi, sẽ bị ảnh hưởng bởi tất cả khác biệt giữa dân số và số chết. Hệ số góc của đường thẳng giữa tỷ lệ tham gia vào nhóm tuổi đối và tỷ lệ ra khỏi nhóm tuổi sẽ ước lượng mức độ đầy đủ của số chết ghi nhận được và cung cấp hệ số điều chỉnh số chết.

$$N^0(a)/N^0(a+) = r + ((1/c).(D^0(a+)/N^0(a+))) \quad (3)$$

trong đó, chỉ số trên "0" dùng để biểu thị giá trị quan sát được, $N^0(a)/N^0(a+)$ là tỷ lệ tham gia vào nhóm tuổi, $D^0(a+)/N^0(a+)$ là tỷ lệ chết quan sát được, r là tỷ lệ tăng trưởng của dân số ổn định và c là mức độ đầy đủ của số chết ghi nhận được đối với dân số ghi nhận được (giả thiết rằng là hằng số đối với tuổi).

Năm 1987, các nhà nhân khẩu học đã mở rộng phương pháp đơn giản này cho dân số không ổn định với biểu thức sau:

$$N^0(a)/N^0(a+) - r^0(a+) = k + ((1/c).(D^0(a+)/N^0(a+))) \quad (4)$$

trong đó, $r^0(a+)$ là tỷ lệ tăng trưởng quan sát được của dân số từ tuổi a trở lên, và k là sai số trong tỷ lệ tăng trưởng (giả thiết rằng là hằng số đối với tuổi).

Phương pháp GGB yêu cầu 3 giả thiết chính: i). Dân số đóng; ii). Dân số và số chết thay đổi theo thời gian, nhưng cùng một nguồn; và iii). Việc ghi chép tuổi của dân số và người chết là chính xác.

Phương pháp thể hệ chết giả định: Phương pháp SEG được hai nhà nhân khẩu học Bennett và Horiuchi đưa ra vào năm 1984 dựa vào đề xuất của Vicent (1951) rằng trong dân số đóng với đăng ký đầy đủ số chết, dân số độ tuổi a vào thời điểm t có thể được ước lượng bằng cách cộng tích lũy số chết đối với các thể hệ sau thời điểm t cho đến khi thể hệ đó chết hết. Điều này tương đương với quan hệ của bảng sống rằng

$$l(a) = \sum_{x=a}^{\omega} d(x) \quad (5)$$

Vì vậy, đối với dân số ổn định, số chết cả thời kỳ từ độ tuổi a trở đi tương đương với dân số vào đúng độ tuổi a . Bennett và Horiuchi đã phát triển phương pháp này cho dân số đóng không ổn định bằng cách sử dụng tỷ lệ tăng trưởng theo độ tuổi. Dân số độ tuổi a có thể ước lượng từ số chết thời kỳ của tất cả tuổi x lớn hơn a bằng cách cộng dồn hàm mũ của các tỷ lệ chết đặc trưng theo tuổi từ a đến x cho phép biết được lịch sử nhân khẩu học của dân số:

$$N(a) = \int_{x=a}^{\omega} D(x) e^{\int_a^x r(y) dy} dx \quad (6)$$

Tỷ số giữa dân số độ tuổi a được ước lượng theo cách này từ số chết quan sát được và dân số độ tuổi a cho phép ước lượng mức độ đầy đủ của số chết ghi nhận được (giả thiết là không đổi đối với tất cả các độ tuổi) trong điều tra:

$$\check{c}(a) = \check{N}(a) / N^0(a) = \left(\int_a^{\omega} D^0(x) e^{\int_a^x r(y) dy} dx \right) / N^0(a) \quad (7)$$

Trong đó $\check{c}(a)$ là ước lượng số chết lớn hơn tuổi a của dân số và $\check{N}(a)$ là dân số độ tuổi a ước lượng được thu được từ số chết và tỷ lệ tăng trưởng lớn hơn độ tuổi a . Trong hình thức cơ bản này, phương pháp SEG thêm vào các giả thiết bổ sung – dân số thay đổi theo thời gian – đối với 3 giả thiết của phương pháp GGB được đề cập ở trên.

Bennett và Horiuchi cũng gợi ý sử dụng kết hợp phương pháp SEG và GGB: đầu tiên ước lượng sự thay đổi phạm vi điều tra bằng cách sử dụng phương pháp GGB, sau đó điều chỉnh số liệu tổng điều tra theo sự thay đổi phạm vi đã ước lượng ở trên, cuối cùng là áp dụng phương pháp SEG; người ta gọi kiểu làm này cách tiếp cận “GGB–SEG kết hợp”. Cách tiếp cận này đã được thể hiện dưới dạng một bảng tính có thông tin đầu vào là dân số vào một thời điểm và số người chết trong 12 tháng trước thời điểm đó chia theo độ tuổi.

1.5 HẠN CHẾ

Một số hạn chế của nghiên cứu về mức độ sinh và mức độ chết trong chuyên khảo này bao gồm:

Thứ nhất, do các chỉ tiêu về mức độ sinh và mức độ chết trong tài liệu này được ước lượng gián tiếp. Những chỉ tiêu này không phải được tính toán trực tiếp từ số liệu vi mô của điều tra mẫu, vì vậy không thể tình được sai số mẫu và khoảng tin cậy.

Thứ hai, thiết kế mẫu của Tổng điều tra dân số 1989, 1999 và 2009 là nhằm cung cấp số liệu đại diện cho cấp tỉnh. Do vậy, những phân tổ nhỏ hơn, chi tiết hơn chỉ để nhằm cung cấp xu hướng thay đổi của chỉ tiêu đó, chứ không đảm bảo đó là giá trị thực.

Thứ ba, các kỹ thuật ước lượng gián tiếp luôn yêu cầu một số giả thiết, ví dụ như dân số đóng, mức sinh hoặc mức chết không đổi, tính đầy đủ của số liệu gốc... Vì vậy, tính chính xác của kết quả thu được từ kỹ thuật ước lượng gián tiếp phụ thuộc vào các giả thiết mà kỹ thuật đó yêu cầu có được đảm bảo hay không. Ví dụ, liên quan đến giả thiết dân số đóng (dân số không có biến động cơ học - di chuyển), thì một kỹ thuật ước lượng gián tiếp có thể cung cấp một kết quả rất tốt cho cả nước, nhưng đưa ra kết quả hạn chế hơn cho cấp đơn vị hành chính cấp dưới (tỉnh/thành phố); thậm chí trong cùng cấp tỉnh/thành phố, kết quả ước lượng của đơn vị này cũng có thể khả dĩ hơn so với kết quả của đơn vị khác. Điều này là do mức độ biến động cơ học của các đơn vị là khác nhau, trong phạm vi cả nước biến động cơ học không có hoặc ở mức độ không đáng kể, còn biến động cơ học của tỉnh/thành phố lớn hơn của cả nước và khác nhau giữa các tỉnh/thành phố.

Với những lý do trên, nên khi sử dụng số liệu trình bày trong tài liệu này cần có cân nhắc thận trọng đối với các chỉ tiêu phân tổ chi tiết.