

KINH NGHIỆM THIẾT KẾ MẪU CHỦ CỦA PHILIPPINES

Mẫu chủ là mẫu chung mà từ đó có thể rút ra được mẫu cho nhiều cuộc điều tra chọn mẫu. Với lợi ích là sử dụng lâu dài, đáp ứng cho nhiều cuộc điều tra, hạn chế việc làm trùng lặp và sử dụng chung các nguồn cho nhiều cuộc điều tra, nâng cao chất lượng kết quả điều tra, linh động hơn trong việc đáp ứng nhanh chóng các nhu cầu số liệu về nhiều lĩnh vực và đặc biệt nâng cao khả năng sử dụng các mặt quan trọng của số liệu (thông tin từ các cuộc điều tra có thể hòa được với nhau), mẫu chủ được xem là rất quan trọng để thiết kế các cuộc điều tra chọn mẫu. Tuy nhiên việc thiết kế mẫu không hề đơn giản, phức tạp hơn nhiều so với mẫu của cuộc điều tra một lần.

Tổng cục Thống kê đã triển khai dàn mẫu chủ từ năm 2001, dàn mẫu chủ này được xây dựng trên số liệu dân số thu được từ cuộc Tổng điều tra dân số 1999 và chủ yếu dùng cho cuộc điều tra mức sống hộ gia đình Việt Nam. Đến năm 2009, trong khuôn khổ dự án Tăng cường phát triển Thống kê Việt Nam, Sida hỗ trợ tài chính và nhân lực là các chuyên gia từ Cơ quan thống kê Thụy Điển phối hợp cùng Tổng cục Thống kê xây dựng dàn mẫu chủ mới dựa trên dữ liệu cuộc Tổng điều tra dân số 2009. Vào tháng 10/2013, Dự án 0006992 cũng dành nội dung hỗ trợ Thống kê Việt Nam nghiên cứu để xây dựng mẫu chủ phục vụ các cuộc điều tra chọn mẫu của Tổng cục Thống kê.

Tại khóa “Tập huấn về kỹ thuật chọn mẫu cho các cuộc điều tra chọn mẫu quốc gia”, chuyên gia

SIAP cũng giới thiệu những nét cơ bản và cần thiết khi xây dựng dàn mẫu chủ và trình bày kinh nghiệm khi lập dàn mẫu chủ của Philippines năm 2003. Bài viết này, trình bày một số kinh nghiệm cơ bản của quá trình xây dựng mẫu chủ của Philippines năm 2003.

1. Mục tiêu thiết kế dàn mẫu chủ

Thường thì dàn mẫu chủ được xây dựng để sử dụng trong khoảng thời gian 10 năm, khoảng cách giữa hai cuộc Tổng điều tra, các cuộc điều tra chọn mẫu trong khoảng thời gian đó dựa vào dàn mẫu chủ để chọn thiết kế mẫu phù hợp. Dàn mẫu chủ được thiết kế năm 2003 của Philippines nhằm mục tiêu: (1) Cập nhật cơ cấu dân số hiện tại được sử dụng từ kết quả của Tổng điều tra năm 2000; (2) Thể hiện trong thiết kế mẫu tái cấu trúc địa lý mới của Philippines (theo quyết định số 36 và 103 của chính phủ Philippines); (3) Kết hợp những bài học kinh nghiệm của những thiết kế mẫu chủ cũ để xây dựng thiết kế mẫu chủ mới hiệu quả hơn; (4) Có đủ mẫu để sử dụng cho các cuộc điều tra chọn mẫu đến năm 2012. Như vậy, tính cập nhật của dàn mẫu chủ luôn được đảm bảo.

2. Phạm vi tổng thể đưa vào dàn mẫu chủ

Xác định tổng thể để lập dàn mẫu chủ tại Philippines bao gồm toàn bộ các hộ gia đình trong cả nước, trừ những hộ thuộc làng khó tiếp cận (Least Accessible Barangays - viết tắt là LABS). Làng được

đưa vào nhóm LABS nếu thuộc một trong những trường hợp sau:

+ Là làng tồn kém khi liệt kê và khó khăn khi tiếp cận;

+ Có các phương tiện giao thông tới làng ít hơn 3 lần/tuần và chi phí một chiều là nhiều hơn P.500 (tương đương 200.000 VNĐ);

+ Thời gian đi bộ từ ga xe cuối cùng đến làng nhiều hơn 8 tiếng.

Để giúp đỡ xác định những làng nhóm LABS, Văn phòng địa phương được yêu cầu gửi danh sách các làng như vậy lên trên và được xác nhận của văn phòng trung ương (viết tắt NSO). Thông thường, chuyên gia phải kiểm tra xác suất xem danh sách đưa lên có đúng không với lý do danh sách đưa lên thường khá lớn. Và tổng số có 338 làng được xếp vào nhóm LABS, chiếm 0,81% tổng số làng của Philippines (Philippines có khoảng 41.946 làng). Tổng số dân và số hộ được đếm vào danh sách không đưa vào dàn mẫu chủ chiếm 0,38% và 0,4% tổng số tương ứng của toàn quốc.

Ngoài ra, Philippines đã xét đến một số khu vực được xem xét để không đưa vào dàn mẫu, những làng có vấn đề về trật tự và hòa bình (viết tắt là POPS). Nhưng khi tiến hành rà soát, báo cáo của các cuộc điều tra dân số mới nhất cho thấy tổng cộng có 1.101 làng thuộc nhóm POPS, chiếm khoảng 2,62 % tổng số làng, tương ứng với tổng số dân và số hộ chiếm khoảng 2,23% của toàn quốc. Xem xét thì thấy rằng số lượng này rất lớn so với dân số cả nước và chỉ mang tính chất tạm thời không ổn định, và họ quyết định rằng không nên loại trừ những làng này ra khỏi nhóm đối tượng đích. Thay vào đó đưa ra cách sáng tạo khi phỏng vấn khu vực này

(NSO đã có nhiều thành công về vấn đề này trong những lần triển khai trước đây).

3. Một số cuộc điều tra tham chiếu để thiết kế dàn mẫu chủ

Philippines sử dụng những cuộc điều tra sẽ tiến hành trong khoảng thời gian giữa hai cuộc tổng điều tra để xây dựng dàn mẫu chủ, mục đích của việc xây dựng dàn mẫu cũng chính để phục vụ cho các cuộc điều tra này, một số cuộc điều tra được lựa chọn bao gồm: (1) Điều tra lực lượng lao động; (2) Điều tra thu nhập và chi tiêu gia đình; (3) Điều tra người Philippines ở nước ngoài; (4) Điều tra chỉ số nghèo hàng năm; (5) Điều tra khả năng biết đọc, viết và giao tiếp; (6) Điều tra kế hoạch hóa gia đình; (7) Điều tra sức khỏe bà mẹ và trẻ em; (8) Điều tra trẻ em từ 5-17 tuổi; (9) Điều tra dân số và sức khỏe quốc gia; Và các cuộc điều tra hộ gia đình khác.

4. Các vấn đề cần lưu ý khi thiết kế mẫu chủ

(1). Chọn đơn vị vào mẫu (gọi là Domain)

Câu hỏi đầu tiên cho việc thiết kế là chọn Domain là gì? Một Domain được định nghĩa như một phân khu của tổng thể đích, từ đó những ước lượng chính xác được tạo ra. Xác định Domain rất quan trọng khi thiết kế vì nó ảnh hưởng đến chọn cỡ mẫu và tổng chi phí điều tra cuối cùng cũng như các vấn đề như độ tin cậy khác (sai số phi chọn mẫu). Số Domain càng lớn thì cỡ mẫu càng lớn. Luôn có mâu thuẫn giữa các ước tính “khu vực nhỏ” có được với yêu cầu về nguồn lực và việc đảm bảo chất lượng. Cơ sở khách quan để lựa chọn Domain là dựa trên sai số chuẩn được tính toán từ các cuộc điều tra trước đây. Ngoài ra còn dựa vào người dùng tin, những thông tin cần đến cấp xã, huyện hay tỉnh... cho ta sự lựa chọn Domain phù hợp hơn.

Bảng 1: Bảng phân phối giá trị của hệ số biến thiên

	Phân phối của giá trị hệ số biến thiên					
	5%	10%	15%	20%	25%	> 25%
Tỷ lệ hộ nghèo (Mẫu chủ năm 1997 – MS 97-domain)						
Năm 1997	0	31	69	86	101	26
Năm 2000	3	37	73	94	104	23
Tỷ lệ hộ nghèo (MS 97-tỉnh)						
Năm 1997	0	33	69	78	82	0
Năm 2000	3	39	72	80	81	1
Tỷ lệ hiện hành biện pháp tránh thai (MS 97-tỉnh)						
Năm 1996	13	51	65	75	82	0
Năm 2000	7	46	64	76	82	0

Mẫu chủ MS-97 domain – Những tỉnh/thành phố có dân số >150.000. Chọn cỡ mẫu n=127.

Mẫu chủ có 97 tỉnh – cỡ mẫu n=82

Như vậy, kết quả từ bảng phân bố hệ số biến thiên trong Bảng 1, ta thấy rằng: Với mẫu chủ năm 1997 hiện tại, nếu Domain là tỉnh thì có thể cung cấp ước lượng chính xác cho nhiều tỉnh nhưng với điều kiện không ước lượng cho biến cụ thể nào cả. Việc ước lượng theo một biến không còn chính xác, ví dụ: Đối với tỷ lệ nghèo đói, chỉ ước lượng chính xác một nửa số tỉnh hoặc đối với tỷ lệ thất nghiệp, gần khoảng 20% số tỉnh cho ước lượng chính xác. Nhưng thời điểm hiện tại bây giờ, điều tra hộ gia đình không cho phép đưa ra các ước lượng chính xác cho cấp tỉnh. Tuy nhiên về điều kiện kinh phí thì Domain không phải là tỉnh mà vùng mới được chọn là đơn vị điều tra.

Đơn vị chọn mẫu cấp 1 (PSU)

Philippines sử dụng thiết kế nhiều tầng để thiết kế dàn mẫu, bởi vì có một số hộ trong danh sách cập nhật thiếu vắng và những nguyên nhân liên quan đến vấn đề hiệu quả. Một trong những bước đầu tiên trong việc thực hiện thiết kế như vậy là sự lựa chọn mẫu cấp một (PSU). Các tiêu chuẩn sau đây đã được sử dụng trong việc lựa chọn các PSU phù hợp: (1) Phải bao quát được tổng thể đích; (2) Có đường ranh giới rõ ràng và bản đồ thích hợp; (3) Phải đủ để đáp ứng nhu cầu cho các cuộc điều tra tiến hành trong khoảng thời gian 2 cuộc tổng điều tra, 10 năm (hoặc cho đến khi một cuộc điều tra mới được thực hiện); (4) Số đơn vị chọn mẫu cấp 1 phải đủ lớn.

Đáp ứng những yêu cầu cần thiết của những cuộc điều tra giữa kỳ thì 1 PSU nên có cỡ tối thiểu là 500 hộ. Có nghĩa là, những hộ mẫu sẽ được luân chuyển trong khoảng thời gian là 10 năm là 25 hộ

mỗi quý, khi đó mỗi hộ không thể chọn quá 3 lần, mẫu PSU sẽ không cần thay đổi mới. Thông thường, số hộ cho 1 PSU nằm trong khoảng (550 - 1000). Do hạn chế về số lượng hộ, nên làng có số lượng hộ nhỏ không được chọn làm PSU, nên gộp những làng nhỏ với nhau thành PSU. Về tiêu chí ranh giới rõ ràng giữa PSU, nhóm làng tiếp giáp nhau về mặt địa lý với nhau ghép vào cùng PSU. Tóm lại, 1 PSU có thể là 1 làng hoặc 1 nhóm từ 2 làng gần nhau trở lên với số hộ ít nhất là 500 hộ.

Chú ý:

- Số lượng PSU càng lớn, càng cần nhiều chi phí bởi vì chúng ta chỉ cần đầu tư 1 khoảng thời gian trong việc chuẩn bị lập bản đồ PSU để lựa chọn PSU mới thường xuyên hơn;

- Số PSU lớn cho ta ước lượng chính xác. Chú ý rằng tương quan trong nội bộ tăng thấp hơn so với trong nhiều PSU. Điều này có nghĩa là thiết kế nhiều tầng cải tiến về mặt hiệu quả.

- Thực tế, Philippines đã chọn 16,579 PSU cho toàn quốc, không bao gồm phòng thí nghiệm.

Đơn vị chọn mẫu cấp 2 (SSU)

Vì số hộ trong mỗi PSU lớn, nên có thể nảy sinh vấn đề trong việc chuẩn bị và cập nhật (Danh sách hoạt động) của khung chọn mẫu. Ngoài ra, vì một số PSU được tạo thành từ 2 hay nhiều làng, việc di chuyển trong PSU có thể đòi hỏi nhiều công sức (ngày công) và chi phí (vận chuyển). Vì các lý do trên, cần tận dụng sự phân vùng các PSU. Trong trường hợp này, địa bàn điều tra (EA) trở thành lựa chọn tốt. Đối với dàn mẫu chủ năm 2003 của Philippines, đơn vị chọn mẫu cấp 2 là địa bàn (EA). Chú ý rằng, trong trường hợp này EA có thể là 1 làng duy nhất.

Việc lập danh sách và cập nhật số hộ sẽ chỉ được tiến hành với những EA được chọn. Nếu số hộ trong EA không đủ, EA mới được lựa chọn trong PSU đó. Trong trường hợp này một thao tác lựa chọn danh sách mới nên thực hiện. Điều này có nghĩa là, lập danh sách hộ là hoạt động thường xuyên trong chu kỳ vận hành mẫu chủ.

Đơn vị lấy mẫu cuối cùng (USU)

Hộ gia đình ở mỗi địa bàn là các đơn vị lấy mẫu cuối cùng.

(2). Xác định cỡ mẫu ban đầu

Thông thường, cỡ mẫu thích hợp được tính toán cho mỗi Domain để đáp ứng độ chính xác nhất định. Với dàn mẫu chủ 2003, sử dụng kết quả dữ liệu năm 2000, độ tin cậy là tỉ lệ nghèo ước tính theo vùng nhỏ hơn 5% để ước tính cỡ mẫu vùng theo công thức:

$$n_h = \frac{deff(1-\hat{p})}{(CV(\hat{p}))^2 \times \hat{p}} \quad (1)$$

Trong đó: deff: Hệ số thiết kế

$\hat{p}$: Tỷ lệ nghèo hộ gia đình;

CV: Hệ số biến thiên.

Ban đầu, tổng cỡ mẫu được tính toán là 56.836 hộ. Nếu chọn CV là 10% thì cỡ mẫu giảm xuống còn 43.882. Cỡ mẫu này được xem xét lại sau khi cân nhắc số câu không trả lời dự kiến, sai số toàn bộ và dự kiến số hộ gia đình tăng lên từ năm 2000 đến năm 2003.

(3). Phân bổ cỡ mẫu

Trong khi cỡ mẫu ban đầu cho mỗi vùng/địa bàn được tính toán dựa trên các điều kiện về độ tin cậy được định sẵn, cỡ mẫu tổng thể được phân bổ lại giữa các vùng khi cân nhắc tính toán đến các mục tiêu sau:

(1) Giảm số vùng nhưng vẫn có ước lượng toàn phần chính xác;

(2) So sánh độ chính xác của các ước tính giữa các vùng.

Phân bổ mẫu theo tỷ lệ đảm bảo mục tiêu thứ nhất, nhưng không lúc nào đảm bảo mục tiêu thứ 2. Phân bổ tỷ lệ bằng nhau đáp ứng mục tiêu thứ 2. Vì vậy, phương án phân bổ khác được xem như là cân bằng giữa hai mục đích trái ngược nhau.

- Những phương pháp phân bổ mẫu khác được xem xét và nghiên cứu như là:

+ Phân bổ tỷ lệ căn bậc hai của tổng thể:

$$n_h = n \times \frac{\sqrt{N_h}}{\sum_h \sqrt{N_h}} \quad (2)$$

Trong đó:

n_h : Số mẫu hộ gia đình của vùng h ;

n : Tổng số mẫu;

N_h : Tổng số hộ gia đình (HHs) của vùng h ;

+ Phân bổ Kish:

$$n_h = \frac{\sqrt{L^2 + IW_h^2}}{\sum_h (\sqrt{L^2 + IW_h^2})} \quad (3)$$

Trong đó:

n_h : Số mẫu hộ gia đình của vùng h ;

N_h : Tổng số hộ gia đình (HHs) của vùng h ;

$$_h = \frac{N_h}{N}$$

L : Là tổng số địa bàn/vùng (=17)

I : Đại lượng đo tầm quan trọng của phân bổ tỷ lệ với phân bổ đều. Nếu $I=1$, chứng tỏ hai phân bổ này như nhau.

Ở Philippines, để chọn phân bổ nào phụ thuộc vào phương sai đặt ra, họ chọn phân bổ Kish, với $I=1$. Khi đó có được danh sách cỡ mẫu cho từng vùng.

Tài liệu tham khảo:

1. Arturo Y.Pacificador, JR, "The 2003 Master Sample Design", Bài trình bày tại khóa tập huấn chọn mẫu điều tra cho cán bộ TCTK Việt Nam năm 2013;
2. Hans Pettersson, "Báo cáo Mẫu tổng thể cho điều tra hộ gia đình tại Việt Nam", Báo cáo từ chuyên công tác tới Tổng cục Thống kê Việt Nam năm 2009;
3. David J. Megill, International Sampling Consultant; Nguyen Manh The, National Consultant; Nguyen Thi Thanh Mai, National Consultant GSO, UNDP, "Assessment of Sampling Frames for GSO National Household Surveys in Viet Nam, and Initial Planning for New Master Sample", Thực hiện năm 2013.

(Còn nữa)

Vũ Văn Anh (Lược dịch và tổng hợp)