

PHƯƠNG PHÁP THU THẬP THÔNG TIN THỐNG KÊ THEO HỘ GIA ĐÌNH

Trần Sinh
Viện Khoa học Thống kê

1. Vị trí của các hộ gia đình (HGD) trong điều tra thống kê

Ở các nước đang phát triển, HGD là một đơn vị chung nhất trong lĩnh vực sản xuất và tiêu dùng. Chính đó là những yếu tố cơ bản quyết định vị trí của các HGD trong điều tra thống kê.

Ở nước ta những năm gần đây điều tra HGD để thu thập thông tin trong nhiều lĩnh vực sản xuất kinh doanh và tiêu dùng... được đặt ra như là điều hiển nhiên và đã thu thập được nhiều thông tin đáng tin cậy, phong phú bao gồm nhiều chủ đề kinh tế xã hội khác nhau.

Nhưng để có được những thông tin đáng tin cậy thoả mãn được mục đích nghiên cứu đặt ra, chúng ta phải vận dụng

những loại hình điều tra thích hợp với cấu trúc mẫu hợp lý.

2. Các loại điều tra hộ gia đình

Có nhiều loại điều tra HGD có thể vận dụng ở nước ta. Sau đây là một số loại điều tra thường gặp:

2.1. Điều tra thường xuyên và không thường xuyên

- Điều tra không thường xuyên là loại điều tra chỉ cần thu thập thông tin đột xuất phục vụ cho yêu cầu quản lý điều tra của các cấp lãnh đạo trong một thời gian nhất định. Mặc dù thoả mãn một số mục đích nghiên cứu, song các cuộc điều tra này thường tốn kém sức người sức của và thời gian lại gấp nên khó có thể đạt được hiệu quả cao trong việc phát triển kĩ thuật điều

tra và nâng cao độ tin cậy của số liệu. Vì thế loại điều tra này, chỉ khi nào thực sự cần thiết mới tiến hành điều tra, chẳng hạn như điều tra thiệt hại bão lụt, điều tra lương thực khi xảy ra đói kém...

- Điều tra thường xuyên là loại điều tra mang tính ổn định lặp đi lặp lại theo thời gian và thường không thay đổi. Tùy theo mục đích nghiên cứu nội dung điều tra có thể là một hoặc nhiều chủ đề kết hợp lại. Do đó, một đặc điểm chung của chương trình điều tra thường xuyên là kỹ thuật điều tra được đảm bảo và do tiến hành thường xuyên theo những chủ đề khác nhau nên cơ quan điều tra có điều kiện tổng kết được kinh nghiệm về phương pháp và tổ chức điều tra. Mặt khác, áp dụng loại điều tra này có điều kiện nâng cao từng bước chất lượng của số liệu và hạn chế được những lãng phí không cần thiết.

2.2. Điều tra đa mục tiêu theo nhiều chủ đề

Loại điều tra này có nhiều hình thức:

- Gọi là kết hợp toàn phần nếu tất cả các chủ đề đều được thu thập trên toàn bộ các HGD trong mẫu.
- Gọi là kết hợp từng phần nếu chỉ một số chủ đề được thu thập trên các HGD trong mẫu.
- Gọi là lồng nhau nếu chia mẫu điều tra HGD thành nhiều mẫu nhỏ theo từng chủ đề độc lập (những mẫu nhỏ này thường là thôn, ấp của các xã thuộc khu vực nông thôn; tổ dân phố hoặc cụm dân phố thuộc khu vực thành thị).

Đặc trưng nổi bật của cấu trúc mẫu kết hợp từng phần hoặc lồng nhau là không sử dụng toàn bộ mẫu, điều tra tất cả các chủ

đề mà chia thành những mẫu con phù hợp với từng chủ đề của đề tài nghiên cứu.

Cuộc Tổng điều tra nông thôn do Tổng cục Thống kê tiến hành năm 1989 là một ví dụ của loại điều tra kết hợp nhiều chủ đề ở cả hai mức độ từng phần và toàn phần. Nếu xét theo chủ đề lớn thì đó là cuộc điều tra kết hợp toàn phần: kinh tế, đời sống, xã hội các chủ đề này được điều tra trong tất cả các HGD trong mẫu. Nhưng nếu xét riêng chủ đề kinh tế thì chủ đề này được chia thành nhiều chủ đề con trong đó có chủ đề hiệu quả sản xuất của một số cây con như: cây lúa, cây cà phê, cây chè, con lợn, con bò.

Các chủ đề con này không thực hiện điều tra trên tất cả các vùng mà chỉ được tiến hành điều tra ở một số HGD và trên một số vùng thích hợp.

Những cuộc điều tra đa mục tiêu kết hợp nhiều chủ đề thường đạt hiệu quả kinh tế hơn các cuộc điều tra riêng theo từng chủ đề, mặt khác loại điều tra này rất thuận tiện cho việc xây dựng và sử dụng các cấu trúc mẫu tổng hợp cũng như tập huấn cán bộ và nhân viên điều tra, tiết kiệm được thời gian và kinh phí.

Một ưu điểm nổi bật của điều tra đa mục tiêu thường có điều kiện phân bổ các hộ gia đình theo nhiều chủ đề khác nhau làm cho quá trình phân tích kết quả điều tra đa dạng và có điều kiện đi sâu nghiên cứu tổng hợp tình hình phát triển thuộc phạm vi và đối tượng điều tra.

Nhược điểm cơ bản của loại điều tra này là nội dung thông tin cần thu nhập khá phức tạp, thời gian phỏng vấn dài, khối lượng chỉ tiêu tổng hợp nhiều, kinh phí đầu tư cho điều tra lớn. Vì vậy, việc tổ chức và chỉ đạo điều

tra phải kết hợp chặt chẽ và khoa học thì mới đáp ứng được yêu cầu đặt ra.

Do vậy loại điều tra này ở nước ta chỉ nên tiến hành từ 5-10 năm một lần theo một số chủ đề nào đó, tương tự như cuộc Tổng điều tra nông thôn hay Tổng điều tra kinh tế vừa qua.

Về mức độ và chủ đề kết hợp theo chúng tôi cần tuân theo nguyên tắc: Các chỉ tiêu phản ánh nghiêng về mặt lượng nhiều hơn tiến hành điều tra đa mục tiêu và các chỉ tiêu phản ánh nghiêng về mặt chất nhiều hơn thì nên tiến hành điều tra chuyên đề hoặc lồng nhau. Chẳng hạn các nhóm chỉ tiêu phản ánh kết quả sản xuất, các tư liệu sản xuất chủ yếu và các khoản đóng góp cho nhà nước... nên kết hợp toàn phần còn các chỉ tiêu phản ánh hiệu quả sản xuất, trình độ sản xuất, trình độ thâm canh, đầu tư vốn nên kết hợp từng phần hoặc lồng nhau.

Thực hiện nguyên tắc này nhằm đạt hiệu quả cao của cấu trúc mẫu mà mục đích nghiên cứu cũng ít bị ảnh hưởng.

2.3. Điều tra chuyên đề

Đây là loại điều tra quen dùng ở nước ta đi sâu vào một vấn đề theo mục đích nghiên cứu. Chúng có thể là những cuộc điều tra đặc biệt nằm trong chương trình điều tra thường xuyên hoặc không thường xuyên nhưng được tiến hành trên những mẫu độc lập đáp ứng một yêu cầu nghiên cứu nào đó. Chúng cũng có thể được tiến hành định kỳ, không định kì hoặc một lần. Trên thực tế, loại điều tra này được vận dụng một cách linh hoạt với những cỡ mẫu khác nhau và trên các khu vực cũng rất khác nhau. Điều tra chuyên đề cũng có thể vận dụng để điều tra hiệu quả sản xuất của

các ngành, nghề, cây, con; điều tra trình độ thâm canh cây trồng; điều tra đầu tư phát triển sản xuất của các HGD, điều tra trình độ phát triển hàng hoá của các HGD cũng như của từng vùng. Trong các trường hợp này không cần sử dụng các mẫu lớn và chỉ nên thực hiện trên những phạm vi thích hợp và những vùng cần thiết.

Khi tiến hành điều tra chuyên đề cần xem xét kĩ hiệu quả của nó. Trong một loạt các cuộc điều tra chuyên đề tiến hành để thu thập những thông tin, mà những thông tin đó có thể thu được bằng một vài cuộc điều tra chuyên đề là không hiệu quả. Mặt khác chúng không có khả năng so sánh, phân tổ theo nhiều chủ đề, phân tích mối quan hệ giữa chúng với nhau như trong các cuộc điều tra mục tiêu. Do vậy cần hạn chế sử dụng loại điều tra này, khi cần thiết mới sử dụng và cần cân nhắc kĩ đến hiệu quả của nó.

3. Cấu trúc mẫu tổng quát

3.1. Phân cấp và xác định đơn vị mẫu của từng cấp

Để có một thiết kế mẫu tối ưu cho các cuộc điều tra HGD, trong điều kiện một nước có nhiều vùng địa lý tập quán sinh sống, sản xuất tiêu dùng như nước ta cần phải phân cấp và phân tổ kết hợp trước khi chọn các HGD vào mẫu điều tra. Quá trình này được tiến hành theo các bước sau đây:

Bước 1: Phân cấp theo vùng địa lý và xác định mẫu đơn vị cấp 1

Dựa trên cơ sở các vùng địa lý đã được hình thành ở nước ta, cấp 1 bao gồm 8 vùng:

1. Đồng bằng sông Hồng
2. Đông Bắc

3. Tây Bắc
4. Bắc Trung Bộ
5. Duyên hải Nam Trung Bộ
6. Tây Nguyên
7. Đông Nam Bộ
8. Tây Nam Bộ

Tất cả các huyện, quận (hoặc tương đương) thuộc các vùng địa lý lập thành các đơn vị tổng thể cấp 1. Căn cứ vào mục đích và chủ đề nghiên cứu điều kiện kinh phí cho phép và phương pháp điều tra thích hợp chọn ra các quận huyện (hoặc tương đương) cần điều tra ở mỗi vùng hoặc tất cả các vùng, các huyện, quận (hoặc tương đương) được chọn này gọi là đơn vị cấp 1.

Về mặt kỹ thuật có thể sử dụng phương pháp chọn phân tổ. Tiêu thức phân tổ có thể là diện tích canh tác, dân số, lao động, hoặc kết quả sản xuất kinh doanh... tùy theo mục đích và chủ đề nghiên cứu để xác định tiêu thức thích hợp. Tuy vậy những đặc điểm địa lý của vùng, đặc điểm về nông thôn thành thị phải đặc biệt coi trọng việc xác định tiêu thức phân tổ các đơn vị mẫu cấp 1. Xác suất chọn cũng cần được quan tâm nhất là đối với cách chọn theo tỷ lệ với quy mô dân số (lao động) diện tích đất canh tác hoặc kết quả sản xuất...

Bước 2: Xác định đơn vị mẫu cấp 2

Các đơn vị cấp 2 là tất cả xã, phường thuộc đơn vị mẫu (huyện, quận hoặc tương đương) đã được chọn ở cấp 1. Trên cơ sở những đơn vị (xã phường) này theo phương pháp và xác suất thích hợp để chọn ra các đơn vị mẫu cấp 2.

Ở cấp này đối với các huyện có đặc trưng về mặt địa lý tập quán canh tác không thuần nhất theo chúng tôi cũng nên phân tổ trước khi chọn các đơn vị mẫu. Chẳng hạn một huyện có các xã mang những đặc trưng của vùng núi, những xã còn lại mang đặc trưng của đồng bằng hoặc trung du hay ven biển. Thậm chí có huyện tuy là huyện đồng bằng hoặc miền núi nhưng vẫn có sự khác biệt về nhất định về địa lý, tập quán sinh sống và canh tác trong những trường hợp đó nên phân tổ thành những tiểu vùng gồm các xã có đặc trưng tương tự với nhau để chọn các đơn vị mẫu của cấp 2. Xác suất chọn thích hợp hơn cả là chọn tỷ lệ với nhân khẩu (lao động) của các xã đó. Sở dĩ sử dụng cách phân tổ và phương pháp chọn như vậy là vì ở nước ta khả năng sản xuất và tiêu dùng (hai yếu tố quyết định vị trí của HGD trong các cuộc điều tra thống kê) khu vực nông thôn phụ thuộc rất lớn vào quy mô nhân khẩu (lao động) của các HGD và những đặc trưng địa lý tập quán sinh sống và canh tác của vùng mà họ cư trú.

Mặt khác căn cứ vào đặc điểm địa lý để phân tổ và chọn theo xác suất tỷ lệ với nhân khẩu (lao động) của các xã đảm bảo tính khách quan hơn và như vậy cũng có nghĩa là loại trừ được những yếu tố chủ quan can thiệp không cần thiết vào quá trình chọn mẫu.

Bước 3: Xác định các đơn vị mẫu cấp 3

Các HGD của tất cả các đơn vị mẫu (xã, phường) đã chọn ở cấp 2 lập thành các đơn vị tổng thể cấp 3. Căn cứ vào những đơn vị (HGD) này theo những phương pháp và xác suất thích hợp chọn ra các đơn vị mẫu (HGD) cấp 3.

Do các đơn vị mẫu cấp 1 và đơn vị mẫu cấp 2 được chọn theo phương pháp có phân tổ với xác suất hợp lý, nên cách chọn các đơn vị mẫu của cấp này trở nên đơn giản hơn. Chúng ta có thể sử dụng phương pháp chọn mẫu ngẫu nhiên đơn giản hoặc chọn máy móc. Trong trường hợp chọn theo phương pháp ngẫu nhiên máy móc cần lập danh sách các HGD xếp độ dốc theo nhân khẩu (lao động) của các HGD để chọn. Theo cách sắp xếp này có thể tạo thuận lợi cho việc chọn các HGD theo tỷ lệ với nhân khẩu (lao động) tức là các HGD có số nhân khẩu (lao động) nhiều hơn khả năng rơi vào mẫu hơn.

Còn chọn theo phương pháp ngẫu nhiên đơn giản thì dựa hoàn toàn vào bản danh sách HGD mà công an xã phường cung cấp để chọn.

3.2. Phạm vi điều tra, cỡ mẫu và các sai số

Trên cơ sở những chủ đề và mục đích nghiên cứu các đơn vị mẫu - HGD không nhất thiết phải trải ra trên phạm vi cả nước tức là ở tất cả các vùng địa lý. Chẳng hạn chủ đề nghiên cứu là cây cà phê hoặc cây chè, rõ ràng các đơn vị mẫu chỉ có thể chọn ở các vùng mà ở đó các HGD có trồng chè hoặc trồng cà phê là phổ biến. Ngược lại với chủ đề nghiên cứu cây lúa hoặc chăn nuôi lợn của cả nước thì mẫu có thể phải rải ra trên tất cả các vùng địa lý hoặc ít nhất cũng phải rải ra ở những huyện (đơn vị mẫu cấp 1) có quy mô diện tích trồng lúa khá.

Cỡ mẫu, sai số chọn mẫu và sai số không chọn mẫu có quan hệ mật thiết với nhau và cùng tác động đến sai số chung của các ước lượng. Nếu cỡ mẫu tăng lên nói chung sẽ làm cho sai số chọn mẫu giảm

xuống nhưng sai số không chọn mẫu lại có xu hướng tăng lên. Ngược lại, muốn giảm sai số không chọn mẫu ta có thể tiến hành điều tra trên một mẫu nhỏ, nhưng trong trường hợp này sai số chọn mẫu lại tăng lên. Vậy phải xác định cỡ mẫu là bao nhiêu để có một sai số chung có thể chấp nhận được.

Giải đáp vấn đề này cần đưa vào hai yếu tố cơ bản: yêu cầu về độ chính xác của các ước lượng và mức độ kinh phí cho cuộc điều tra:

- Trong trường hợp kinh phí không bị giới hạn nghiêm ngặt chúng ta có thể nâng độ chính xác lên bằng cách tăng kinh phí cho cuộc điều tra sao cho cỡ mẫu tăng nhưng chi phí bình quân cho một đơn vị mẫu cũng phải tăng lên. Điều này có nghĩa chúng ta vừa tăng cỡ mẫu để giảm sai số chọn mẫu vừa tăng kinh phí để huấn luyện cho cán bộ và nhân viên điều tra và trả lương cao cho họ... để giảm sai số không chọn mẫu.

- Trường hợp kinh phí bị giới hạn nghiêm ngặt mà ta lại muốn nâng cao độ chính xác của các ước lượng cần phải có phương pháp chọn mẫu tối ưu. Đây là một vấn đề phức tạp nhưng rất cơ bản của lý thuyết mẫu cần được xem xét kỹ lưỡng khi thiết kế những lược đồ mẫu cụ thể. Ở đây cần đặc biệt lưu ý đến cách chọn mẫu phân cấp có kết hợp phân bố. Vì sai số phụ thuộc rất nhiều vào độ chênh lệch của các lượng biến giữa các đơn vị của tổng thể khi tiến hành phân cấp kết hợp phân tổ có nghĩa là ta đã làm cho độ chênh lệch đó trong nội bộ của từng cấp, từng tổ giảm xuống.

- Trong trường hợp kinh phí bị giới hạn nghiêm ngặt và một lược đồ chọn mẫu tối ưu

nhưng vẫn không thoả mãn được yêu cầu về độ chính xác của số liệu thì buộc chúng ta phải hạ yêu cầu chính xác hoặc độ tin cậy của các ước lượng xuống.

Từ những phân tích trên cho phép chúng tôi đi đến kết luận: kinh phí và kĩ thuật chọn mẫu là hai yếu tố cơ bản nhất quyết định đến độ chính xác của ước lượng trong các cuộc điều tra mẫu. Do vậy khi tiến hành bất cứ một cuộc điều tra nào, chủ đề gì, điều trước tiên cần phải quan tâm đến hai yếu tố này, đặc biệt là kinh phí cho cuộc điều tra.

Kinh phí ít độ chính xác của các ước lượng lại muốn cao đó là điều không tưởng.

Qua những ý tưởng vừa được trình bày của bài viết này chúng tôi nghĩ cần đưa ra một mẫu quốc gia phân tổ có phân cấp kết hợp. Đối với điều kiện địa lý và kinh tế - xã hội ở nước ta, một cấu trúc như thế có thể sử dụng làm cơ sở thực hiện những loại hình điều tra thích hợp để thu thập thông tin theo HGD không những trên phạm vi toàn quốc mà còn ở từng vùng, thậm chí tỉnh thành phố■