

NGHIÊN CỨU CĂN CỨ NGUYÊN TẮC VÀ QUY TRÌNH XÁC ĐỊNH CỖ MẪU, PHÂN BỐ MẪU CHO CÁC CUỘC ĐIỀU TRA CHỌN MẪU

ThS. Hoàng Thu Hiền, ThS. Vũ Thị Vân Anh*

Giới thiệu:

Xác định cỡ mẫu và phân bố mẫu điều tra là một trong những nội dung được nhiều chuyên gia thống kê trên khắp thế giới nghiên cứu hàng vài chục năm trước trong hầu hết các chủ đề liên quan chọn mẫu điều tra. Kết quả là đã có nhiều công trình xuất bản thành sách, nhiều tài liệu trình bày về việc vận dụng lý thuyết về chọn mẫu áp dụng đối với từng cuộc điều tra cụ thể, ví dụ: Tài liệu “Chọn mẫu điều tra: Lý thuyết và phương pháp – tái bản lần 2” được trình bày bởi tác giả Arijit Chaudhuri, Viện Thống kê Ấn Độ Calcutta, và Horst Stenger, Đại học Mannheim, Đức năm 2005; Tài liệu “Sổ tay quốc tế về phương pháp luận điều tra” được viết bởi nhóm tác giả Edith D. de Leeuw, Joop Hox, Don Dillman năm 2008; Tài liệu “Sổ tay khung lấy mẫu chính cho thống kê nông nghiệp”, ấn phẩm được FAO biên soạn trong khuôn khổ Chiến lược toàn cầu nhằm cải thiện Thống kê nông nghiệp và nông thôn xuất bản năm 2015. Cuốn “Sổ tay chỉ số giá sản xuất”, xuất bản năm 2004 của Quỹ tiền tệ quốc tế (IMF). Khối lượng nghiên cứu quốc tế viết về chọn mẫu tương đối đồ sộ và đa dạng,..

Ở Việt Nam, nhiều cuốn sách biên soạn về điều tra chọn mẫu hay và được áp dụng hiệu

quả trong các cuộc điều tra của Tổng cục Thống kê, ví dụ: (1) Sách “Điều tra chọn mẫu và ứng dụng trong công tác thống kê”, chủ biên PGS.TS Tăng Văn Khiên, Nhà xuất bản Thống kê ấn hành năm 2003; (2) Sách “Điều tra chọn mẫu, lý luận và ứng dụng trong công tác dân số và kế hoạch hóa gia đình” do Trung tâm nghiên cứu thông tin và tư liệu dân số chủ trì và được xuất bản năm 1998 và nhiều đề tài nghiên cứu khoa học xung quanh chủ đề này.

Năm 2023-2024, Viện Khoa học Thống kê chủ trì nghiên cứu đề tài “Nghiên cứu xác định cỡ mẫu và phân bố mẫu cho một số loại hình điều tra thống kê tại Tổng cục Thống kê, Ban chủ nhiệm giới thiệu một phần kết quả nghiên cứu đề tài về căn cứ, nguyên tắc xác định cỡ mẫu, phân bố mẫu để độc giả tham khảo và góp ý bổ sung.

1. Một số khái niệm chung liên quan điều tra chọn mẫu

1.1. Cỡ mẫu

Cỡ mẫu hay còn gọi là quy mô mẫu là số lượng các đơn vị của tổng thể chung được chọn vào tổng thể mẫu để tiến hành điều tra thực tế.

Mỗi cuộc điều tra chọn mẫu cần xác định một cỡ mẫu vừa đủ. Nếu cỡ mẫu quá lớn sẽ gây tốn kém và không bảo đảm được yêu cầu về thời gian. Ngược lại, nếu cỡ mẫu quá nhỏ sẽ không đảm bảo tính đại diện cho tổng thể

* Viện Khoa học Thống kê

chung, sai số chọn mẫu sẽ lớn dẫn đến kết quả suy rộng cho tổng thể chung có độ tin cậy không cao. Việc xác định cỡ mẫu tùy thuộc vào chỉ tiêu cần nghiên cứu, quy mô của đối tượng điều tra, điều kiện về nguồn lực, cách chọn mẫu, phương pháp tổ chức chọn mẫu.

1.2. Phân bố mẫu

Phân bố mẫu là việc phân chia số đơn vị mẫu của tổng thể mẫu cho các tổ, các địa bàn điều tra trong tổng thể chung.

Có nhiều phương pháp phân bố mẫu khác nhau, trong đó các phương pháp chủ yếu thường được sử dụng là: (1) Phân bố mẫu tỷ lệ thuận với quy mô tổng thể chung; (2) Phân bố mẫu tỷ lệ thuận với căn bậc 2 của quy mô tổng thể chung; (3) Phân bố Neyman; (4) Phân bố mẫu tối ưu; (5) Phân bố mẫu có ưu tiên cho các tổ được đánh giá là quan trọng. Phần này được trình bày chi tiết ở mục 1.3.

1.3. Tỷ lệ chọn mẫu

Tỷ lệ chọn mẫu là tỷ lệ phần trăm số đơn vị được chọn vào tổng thể mẫu so với tổng số đơn vị của tổng thể chung.

1.4. Hệ số tin cậy và khoảng tin cậy

Hệ số tin cậy (tương ứng với độ tin cậy ϕ_t) là xác suất để giá trị thực tế của chỉ tiêu nghiên cứu (\bar{X}) còn nằm trong khoảng tin cậy ($\bar{x} - t \cdot \mu_x$ đến $\bar{x} + t \cdot \mu_x$).

1.5. Sai số chọn mẫu, phạm vi sai số chọn mẫu, tỷ lệ sai số chọn mẫu

Sai số chọn mẫu (SSCM) là sự khác nhau giữa giá trị ước lượng của mẫu và giá trị của tổng thể chung. Sai số chọn mẫu còn gọi là sai số do tính đại biểu và chỉ xảy ra trong điều tra chọn mẫu vì chỉ điều tra một số ít đơn vị mà kết quả lại suy rộng cho cả tổng thể. Sai số chọn mẫu được dùng để tính tỷ lệ sai số chọn mẫu đánh giá độ tin cậy về số liệu điều tra và ước lượng số liệu điều tra theo phương pháp ước lượng khoảng của các chỉ tiêu nghiên cứu. Mỗi phương pháp tổ chức

chọn mẫu sẽ có công thức tính sai số chọn mẫu riêng.

Phạm vi sai số chọn mẫu

Phạm vi sai số chọn mẫu (Δ_x) là đại lượng phản ánh sự khác biệt giữa trị số ước lượng của mẫu và trị số tổng thể chung tương ứng với xác suất tin cậy nhất định $\phi_{(t)}$. Mỗi xác suất tin cậy $\phi_{(t)}$ có hệ số tin cậy (t).

Phạm vi sai số chọn mẫu (Δ) là giá trị dương nếu trị số tổng thể chung lớn hơn trị số ước lượng và là âm nếu có quan hệ ngược lại. Vì vậy trị số tổng thể chung của chỉ tiêu nghiên cứu (θ) sẽ nằm trong khoảng (từ $\theta' - t \cdot \mu$ đến $\theta' + t \cdot \mu$).

Tỷ lệ sai số chọn mẫu (H):

$$H = \frac{\mu}{x} ;$$

Tỷ lệ sai số để đánh giá độ tin cậy về số liệu của chỉ tiêu nghiên cứu. H càng lớn thì độ tin cậy của số liệu càng kém và ngược lại.

1.6. Phương pháp tổ chức chọn mẫu

Phương pháp tổ chức chọn mẫu là cách thức lựa chọn đơn vị từ tổng thể vào tổng thể mẫu để tiến hành điều tra. Có nhiều phương pháp tổ chức chọn mẫu, nhưng trong thực tế công tác thống kê thường áp dụng các phương pháp tổ chức chọn mẫu sau: Phương pháp tổ chức chọn mẫu ngẫu nhiên đơn giản, phương pháp tổ chức chọn mẫu phân tổ; phương pháp tổ chức chọn mẫu nhiều cấp và phương pháp tổ chức chọn mẫu chùm. Với các phương pháp tổ chức chọn mẫu khác nhau ta có công thức xác định cỡ mẫu, tính tham số mẫu khác nhau.

- Phương pháp tổ chức chọn mẫu ngẫu nhiên đơn giản

Khi tiến hành chọn mẫu ngẫu nhiên đơn giản chỉ việc lập danh sách các đơn vị trong tổng thể kèm theo số thứ tự từ 1 đến hết.

➤➤➤ NGHIÊN CỨU • TRAO ĐỔI

Sau đó dùng bảng số ngẫu nhiên hoặc rút thăm chọn ngẫu nhiên không lặp lại từ danh sách được lập trong bảng để được các đơn vị mẫu cần điều tra.

- Phương pháp tổ chức chọn mẫu phân tổ

Trước khi tiến hành chọn mẫu, ta cần phân tổng thể thành các tổ khác nhau với những đặc điểm chung nhất định. Ví dụ, cuộc điều tra toàn quốc, ta phân ra mỗi tỉnh là 1 tổ để tiến hành chọn mẫu.

- Phương pháp tổ chức chọn mẫu nhiều cấp (thông thường sử dụng chọn mẫu 2 cấp)

Chọn mẫu nhiều cấp là tiến hành điều tra theo nhiều giai đoạn. Trong đó mỗi giai đoạn là một cấp chọn mẫu. Chọn mẫu có bao nhiêu cấp thì có bấy nhiêu loại đơn vị mẫu. Ví dụ, điều tra năng suất sản lượng cây trồng nông nghiệp sử dụng phương pháp chọn mẫu 2 cấp. Mẫu cấp I là mẫu thôn, mẫu cấp II là mẫu hộ. Nghĩa là mỗi tỉnh lựa chọn X thôn mẫu để điều tra, trong mỗi thôn mẫu đó lại tiếp tục lựa chọn hộ mẫu để điều tra.

- Phương pháp tổ chức chọn mẫu chùm

Nếu điều tra chia thành nhiều cấp, các cấp tiến hành trước lựa chọn từng đơn vị mẫu, nhưng ở cấp cuối cùng không chọn ra từng đơn vị mà chọn tất cả các đơn vị để điều tra.

2. Căn cứ, nguyên tắc và quy trình xác định cỡ mẫu

Xác định cỡ mẫu (số đơn vị mẫu) chính là xác định số lượng đơn vị điều tra trong tổng thể mẫu để tiến hành thu thập số liệu. Yêu cầu của cỡ mẫu là vừa đủ để vừa đảm bảo độ tin cậy cần thiết của số liệu điều tra vừa đảm bảo phù hợp với điều kiện về nhân lực và kinh phí và có thể thực hiện được, tức là có tính khả thi.

2.1. Căn cứ xác định cỡ mẫu

Thống kê có nền tảng là định luật số lớn, các con số thống kê phản ánh tính quy luật của hiện tượng nghiên cứu, vì vậy thay vì điều tra tổng thể, dựa vào quy luật của hiện

tượng nghiên cứu mà lựa chọn 1 số lượng vừa đủ (cỡ mẫu) để tiến hành điều tra từ đó suy rộng cho tổng thể. Bài toán xác định cỡ mẫu gắn với mức độ biến thiên của hiện tượng nghiên cứu. Đây chính là lý luận, căn cứ cơ bản của lý thuyết điều tra chọn mẫu. Về lý thuyết, cỡ mẫu căn cứ vào mức độ biến thiên của chỉ tiêu nghiên cứu để tạo ra một bài toán cơ bản.

Trong điều tra chọn mẫu, có hai chỉ tiêu thống kê cơ bản cần được ước lượng (tìm ra số liệu của tổng thể dựa vào số liệu mẫu) là giá trị trung bình và giá trị tổng của tiêu thức nghiên cứu. Hai chỉ tiêu này có quan hệ chặt chẽ với nhau (giá trị tổng bằng giá trị trung bình nhân với số lượng của tổng thể). Nên đỡ phức tạp, khi nghiên cứu ta chỉ xem xét đến giá trị trung bình.

Bài toán thống kê đặt ra là tìm cỡ mẫu nhỏ nhất sao cho giá trị trung bình của tổng thể nằm trong khoảng $(\bar{x} - \Delta_x; \bar{x} + \Delta_x)$, với Δ_x là phạm vi sai số cho phép.

2.2. Nguyên tắc xác định cỡ mẫu

Trong thực tế áp dụng không nhất thiết (và không phải lúc nào) cũng sử dụng các công thức xác định cỡ mẫu như đã trình bày ở trên. Vì vậy, người ta sử dụng các nguyên tắc khác nhau để xác định cỡ mẫu. Như đã trình bày ở trên, mỗi một phương pháp chọn mẫu có thể có cách xác định cỡ mẫu riêng, song nguyên tắc để xác định cỡ mẫu được trình bày ở đây có thể áp dụng cho tất cả các phương pháp chọn mẫu khác nhau. Hai nguyên tắc cơ bản thường được sử dụng khi xác định cỡ mẫu trong thực tế của các cuộc điều tra thống kê:

- Phải dựa vào mức độ biến động của chỉ tiêu cần được nghiên cứu có trong điều tra. Mức độ biến động này được biểu hiện thông qua phương sai của chỉ tiêu đó. Mức độ biến động này càng lớn thì cỡ mẫu phải càng lớn và ngược lại, mức độ biến động của chỉ tiêu

nghiên cứu càng nhỏ, cỡ mẫu cần thiết của cuộc điều tra đó cũng nhỏ theo.

- Dựa vào mức độ sai số cho phép. Khi sử dụng thông tin từ điều tra mẫu, người dùng tin thường đòi hỏi số liệu với một sai số nhất định (ví dụ, chính xác 99%, hay 95%,...). Điều này có tính chất quyết định đến cỡ mẫu cần cho cuộc điều tra. Mức độ sai số cho phép càng lớn thì cỡ mẫu càng nhỏ và ngược lại, mức độ sai số cho phép nhỏ (độ chính xác cao) thì cỡ mẫu phải càng lớn.

Hai nguyên tắc trên, thể hiện trực tiếp thông qua công thức xác định cỡ mẫu. Ngoài ra, trong thực tế xác định cỡ mẫu còn dựa vào nguyên tắc sau:

- Dựa vào số lượng các chỉ tiêu thu thập trong điều tra đó. Số lượng các chỉ tiêu của cuộc điều tra mà ít thì cỡ mẫu có thể phải lớn, và ngược lại, số lượng các chỉ tiêu thống kê cần thu thập trong cuộc điều tra nhiều thì cỡ mẫu cần phải nhỏ. Nếu làm ngược lại sẽ ảnh hưởng đến chất lượng trong quá trình triển khai điều tra. Cũng cần ghi nhớ rằng, con số thống kê phản ánh quy luật số lớn do vậy phải có một mức độ nhất định các đơn vị cá thể mới phản ánh được quy luật của tổng thể.

- Dựa vào cấp độ, mức độ chi tiết của thông tin ảnh hưởng đến cỡ mẫu của cuộc điều tra. Ví dụ thông tin đó chi tiết cho toàn quốc, hoặc cho cấp tỉnh hay cấp huyện, hoặc mức độ chi tiết đến đâu thì cỡ mẫu càng tăng lên theo mức độ chi tiết đó có vậy mới đáp ứng được yêu cầu đặt ra. Nếu thông tin chỉ cần phản ánh tình hình của toàn quốc thì cỡ mẫu sẽ nhỏ còn nếu thông tin cần phải phản ánh cả tình hình của các tỉnh thì cỡ mẫu phải rất lớn. Dường như có bao nhiêu tỉnh thì số lượng đơn vị cần chọn vào mẫu sẽ lớn gấp bấy nhiêu lần so với cỡ mẫu chỉ đại diện cho toàn quốc.

- Cỡ mẫu phụ thuộc vào kinh phí của cuộc điều tra. Để thấy là nếu kinh phí lớn ta sẽ có điều kiện hơn để điều tra với cỡ mẫu tối đa (có tính đại diện cao), còn kinh phí điều tra

hạn hẹp cỡ mẫu chỉ có thể ở mức tối thiểu (có tính đại diện thấp). Kinh phí thấp mà cỡ tình điều tra với cỡ mẫu lớn sẽ làm cho chất lượng của cuộc điều tra bị hạn chế.

- Dựa vào khả năng tổ chức thực hiện cuộc điều tra. Trình độ tổ chức cũng là một yếu tố cần được cân nhắc khi xác định cỡ mẫu cho cuộc điều tra. Nếu trình độ tổ chức và thực hiện điều tra thấp thì không được phép lấy cỡ mẫu lớn. Bởi làm như vậy sẽ dễ dẫn đến thất bại.

2.3. Quy trình xác định cỡ mẫu

* Nếu sử dụng công thức tính cỡ mẫu đã đưa ra phần trên, để có thể tính cỡ mẫu thì ngay từ khi chuẩn bị điều tra để xây dựng và quyết định phương án điều tra ta đã phải có được số liệu, cụ thể:

- Xác định số đơn vị tổng thể chung (N) lấy từ số liệu thống kê;

- Xác suất tin cậy (pt) và phạm vi SSCM (Δ) do người tổ chức điều tra yêu cầu cho từng cuộc điều tra.

- Phương sai của tổng thể chung (S^2 và $p(1-p)$) thì phải dựa vào kết quả của các cuộc điều tra trước đó. Trường hợp không có các cuộc điều tra trước tương tự hoặc có nhưng không tính được phương sai thì sẽ xử lý như sau:

+ Khi điều tra nghiên cứu chỉ tiêu bình quân thì phải điều tra mẫu nhỏ để xác định phương sai.

+ Khi điều tra nghiên cứu chỉ tiêu tỷ lệ thì sẽ lấy phương sai lớn nhất: $p(1-p) = 0,5 \times (1-0,5) = 0,25$.

* Nếu xác định cỡ mẫu dựa vào kinh nghiệm

Trong thực tế công tác điều tra người ta thường dựa vào kinh nghiệm để xác định sẵn cỡ mẫu cần thiết. Thực chất đây là dựa vào kinh nghiệm đúc rút ra từ nhiều lần điều tra để quyết định cỡ mẫu chuẩn dùng cho các lần điều tra tiếp theo. Ví dụ, qua nhiều năm tiến hành điều tra tình hình sinh đẻ của các nước trên thế giới người ta thấy rằng chỉ cần một

➤➤➤ NGHIÊN CỨU • TRAO ĐỔI

mẫu khoảng 5.000 phụ nữ thuộc nhóm tuổi 15-49 là đủ cho phép nghiên cứu về mức độ sinh, mức độ chết của trẻ sơ sinh, mức độ chết của trẻ em, tình hình áp dụng các biện pháp tránh thai cũng như tình trạng nuôi con bằng sữa mẹ của chị em phụ nữ. Từ đây suy ra: nếu lấy hộ làm đơn vị điều tra và bình quân một hộ có khoảng 1,2 phụ nữ trong độ tuổi sinh đẻ thì cỡ mẫu cần thiết là 4.200 hộ.

* Xác định cỡ mẫu trên cơ sở cân đối giữa chi phí và sai số chọn mẫu cho phép

Khi thiết kế điều tra chọn mẫu người ta luôn quan tâm xem mức chi phí (kinh phí) được phép sử dụng cho cuộc điều tra là bao nhiêu và đòi hỏi về mức độ chính xác của kết quả điều tra ra sao. Chi phí ở đây bao gồm số tiền phải bỏ ra và số người phải huy động, còn mức độ chính xác của kết quả điều tra thể hiện ở sai số chọn mẫu mắc phải. Số tiền bỏ ra nhiều và được sử dụng đúng mục đích của cuộc điều tra, sai số chọn mẫu mắc phải sẽ thấp, vì vậy nếu muốn sai số chọn mẫu thấp kinh phí bỏ ra thường phải cao. Với lý do đó khi thiết kế mẫu cần phải cân đối giữa kinh phí bỏ ra và mức độ sai số mắc phải. Trong điều kiện số tiền và số người đã cố định thì cố gắng làm sao cho sai số chọn mẫu thấp nhất, mặt khác khi đã cố định sai số cho phép, lại phải cố gắng làm sao để chi phí bỏ ra ít nhất.

* Xác định cỡ mẫu dựa vào mức độ chi tiết của nhu cầu thông tin

Trong thực tế, cỡ mẫu còn phụ thuộc vào nhu cầu chi tiết của thông tin cần thu thập. Nếu thông tin chỉ cần phản ánh tình hình chung của tổng thể cần nghiên cứu thì cỡ mẫu cần thiết sẽ nhỏ, còn nếu thông tin cần thu thập lại phải phản ánh tình hình của các bộ phận cấu thành nên tổng thể thì cỡ mẫu phải nhiều hơn. Thật vậy, lúc này từng bộ phận con của tổng thể chung có thể coi là một tổng thể thống kê mới người điều tra phải tìm hiểu quy luật thống kê của tổng thể này. Nhưng như đã trình bày ở trên, để tìm

hiểu được quy luật của hiện tượng số lớn cần phải điều tra một số lượng đơn vị nhất định tùy thuộc vào mức độ biến thiên của chỉ tiêu cần nghiên cứu. Trong trường hợp thông tin cần phản ánh quy luật của các bộ phận của tổng thể ta phải xác định cỡ mẫu tối thiểu cho bộ phận có quy mô nhỏ nhất rồi từ đó nhân với số bộ phận của tổng thể để xác định cỡ mẫu chung. Thí dụ, ta muốn điều tra chọn mẫu nghiên cứu về tỷ lệ sinh của một tỉnh. Như vậy dù tỉnh lớn hay nhỏ ta cũng chỉ cần điều tra khoảng 4.200 hộ là đủ. Giả sử bây giờ các nhà lãnh đạo yêu cầu phải cung cấp thông tin chi tiết cho từng huyện của một tỉnh. Theo kinh nghiệm người ta thấy để nghiên cứu tỷ lệ sinh riêng cho một huyện chỉ cần cỡ mẫu khoảng 840 hộ là đủ do vậy khi xác định cỡ mẫu cho tỉnh có 6 huyện ta nhân 6 với 840 hộ. Kết quả cỡ mẫu cần điều tra cho tỉnh này (trong trường hợp cần thông tin chi tiết cho từng huyện một) là 5.040 hộ.

Trong trường hợp các nhà nghiên cứu đòi hỏi thông tin phản ánh cơ cấu chi tiết của hiện tượng, chẳng hạn như khi nghiên cứu về mức sinh của dân số người ta không chỉ muốn biết tỷ lệ sinh thô mà còn muốn biết cả tỷ lệ sinh đặc trưng theo từng nhóm tuổi. Để xác định cỡ mẫu cho phép đáp ứng nhu cầu thông tin này, trước tiên cần xác định mức độ phân tổ chi tiết cần thiết cũng như phân bố của hiện tượng sau đó theo lý thuyết thì mỗi một tổ ở cấp chi tiết nhất cần có ít nhất 5 sự kiện. Như vậy để xác định số các đơn vị cần được điều tra đầu tiên ta phải xác định số các sự kiện cần thiết phải có bằng cách lấy 5 chia cho tỷ trọng của tổ có tỷ trọng nhỏ nhất của cấp chi tiết nhất rồi nhân với tỷ trọng của các tổ khác sau đó cộng tất cả các kết quả lại. Tổng tính được chính là số các sự kiện cần thiết phải thu được trong cuộc điều tra. Từ tổng này, dựa vào điều kiện cụ thể ta tính được cỡ mẫu cần thiết.

Nhiều khi người ta có thể nghiên cứu các chỉ tiêu thống kê chi tiết theo đơn vị lãnh thổ.

Ví dụ, ở các tỉnh người ta không chỉ muốn ước lượng mức thu nhập bình quân cho một hộ trên phạm vi toàn tỉnh mà còn muốn ước lượng mức thu nhập bình quân của một hộ cho từng huyện trong tỉnh nữa. Trong trường hợp như vậy cỡ mẫu cũng cần được xác định sao cho đáp ứng được yêu cầu ấy. Cũng cần lưu ý là khi cỡ mẫu đáp ứng được yêu cầu nghiên cứu cho từng đơn vị hành chính nhỏ thì tất yếu sẽ đáp ứng được yêu cầu nghiên cứu chung của toàn tỉnh. Vì vậy trong trường hợp xác định cỡ mẫu cho từng đơn vị hành chính nhỏ thì xác định cỡ mẫu chung cho toàn tỉnh bằng cách cộng tất cả cỡ mẫu của các đơn vị nhỏ lại.

Tuy nhiên, cũng có thể xác định cỡ mẫu chung cho toàn bộ tổng thể sau đó mới phân bổ cỡ mẫu cho từng tổ nhỏ. Trong trường hợp này, khi phân bổ cỡ mẫu cần phải lưu ý tới nhu cầu nghiên cứu theo lãnh thổ của địa phương. Việc xác định và phân bổ cỡ mẫu cho từng tổ là một khâu quan trọng của điều tra chọn mẫu phân tổ, do vậy sẽ được trình bày sâu hơn trong chương phương pháp điều tra chọn mẫu phân tổ.

3. Căn cứ, nguyên tắc và quy trình phân bổ mẫu

3.1. Căn cứ phân bổ mẫu

Về phương diện lý thuyết, việc phân bổ mẫu cho các tổ phụ thuộc vào mức biến động của chỉ tiêu cần nghiên cứu ở từng tổ. Mặt khác, còn phải căn cứ vào khối lượng thông tin của từng tổ và từng đơn vị điều tra.

3.2. Nguyên tắc phân bổ mẫu

Với căn cứ nêu ra trên thì phân bổ mẫu vào các tổ cần tuân thủ những nguyên tắc cơ bản:

- Dựa vào mức độ biến động của chỉ tiêu cần thu thập thông tin trong từng tổ. Với nguyên tắc này, tổ nào có biến động lớn thì tổ đó cần được bố trí phân bổ nhiều mẫu hơn và ngược lại.

- Dựa vào tầm quan trọng của từng cá thể, từng nhóm cá thể đối với thông tin cần cho ra để phân bổ mẫu.

Với một nội dung thông tin cần thu thập, các đơn vị điều tra có khối lượng thông tin khác nhau. Có đơn vị điều tra mang nhiều thông tin hơn, ngược lại có đơn vị có rất ít thông tin. Trong điều kiện như vậy, người ta dựa vào lượng thông tin để phân tổ tổng thể ra thành các tổ có dung lượng thông tin khác nhau. Trên cơ sở đó, phân bổ cỡ mẫu cho các tổ theo nguyên tắc các tổ có dung lượng thông tin phong phú tỷ lệ mẫu phải cao, thậm chí có khi phải chọn hết (ví dụ, điều tra doanh nghiệp, những doanh nghiệp nằm trong nhóm có doanh thu lớn phải điều tra hết)

- Dựa vào nhu cầu thông tin của các cấp lãnh đạo.

- Dựa vào số lượng các đơn vị cá thể và tính đồng đều của các đơn vị cá thể trong mỗi tổ. Số lượng càng lớn thì số lượng đơn vị mẫu có thể giảm. Tính đồng đều cao thì số lượng mẫu có thể ít.

3.3. Quy trình phân bổ mẫu

Như vậy, để phân bổ mẫu tốt, cần có nhiều thông tin, trước khi phân tổ phải phân tích kỹ nội dung điều tra cũng như tìm hiểu đơn vị điều tra, dựa vào ưu nhược điểm của từng phương pháp phân bổ mẫu để lựa chọn phương pháp phân bổ phù hợp.

4. Khái quát thực trạng xác định cỡ mẫu và phân bổ mẫu ở các cuộc điều tra của Tổng cục Thống kê

4.1. Thực trạng xác định cỡ mẫu

Trong Luật Thống kê năm 2015 quy định có 02 loại điều tra thống kê, gồm: (1) Tổng điều tra thống kê quốc gia và điều tra thống kê trong chương trình điều tra thống kê quốc gia và (2) Điều tra thống kê ngoài chương trình điều tra thống kê quốc gia. Chương trình điều tra thống kê quốc gia hiện nay đang áp dụng theo Quyết định số 03/2023/QĐ-TTg

➤➤➤ NGHIÊN CỨU • TRAO ĐỔI

ngày 15/02/2023. Theo đó, có 45 cuộc điều tra, khảo sát trong đó có 03 cuộc tổng điều tra thống kê quốc gia (Tổng điều tra dân số và nhà ở; Tổng điều tra nông thôn, nông nghiệp; Tổng điều tra kinh tế) và 42 cuộc điều tra thống kê ở các lĩnh vực khác nhau. Trong đó có 21 cuộc điều tra chọn mẫu, 19 cuộc kết hợp điều tra toàn bộ và điều tra chọn mẫu. Tổng cục Thống kê chủ trì thực hiện 31 cuộc điều tra chọn mẫu và điều tra toàn bộ kết hợp với điều tra chọn mẫu (gọi chung là điều tra chọn mẫu). Trên cơ sở phương án điều tra BCN đã tổng hợp, xác định cỡ mẫu phân thành các loại sau:

+ Dạng thứ nhất: Cỡ mẫu được xác định theo một tỷ lệ chọn mẫu cụ thể, ví dụ:

(i) Tổng điều tra dân số và nhà ở ấn định cỡ mẫu điều tra là 10% tổng số hộ;

(ii) Điều tra dân số và nhà ở giữa kỳ, cỡ mẫu là 20% địa bàn điều tra của cả nước (tương đương 39.340 địa bàn) để ước lượng quy mô dân số đại diện đến cấp huyện, cỡ mẫu 6,5% địa bàn điều tra của cả nước để đại diện cho các chỉ tiêu cấp tỉnh;

(iii) Điều tra nông thôn, nông nghiệp giữa kỳ, cơ mẫu địa bàn ở nông thôn là 10% tổng số địa bàn khu vực nông thôn của cả nước và ở thành thị là 20% tổng số địa bàn khu vực thành thị;

(iv) Điều tra lâm nghiệp, cỡ mẫu điều tra sản lượng gỗ khai thác tùy thuộc vào số lượng hộ có diện tích rừng trồng cho khai thác trong 12 tháng qua của mỗi nhóm gỗ trên địa bàn tỉnh. Từ 1000 hộ trở lên, cỡ mẫu 5,5% số hộ nhưng không quá 60 hộ/nhóm gỗ; từ 500 hộ đến 999 hộ: 6%; từ 300 hộ đến 499 hộ: 9%; từ 100 hộ đến 299 hộ: 12%; từ 50 hộ đến 99 hộ: 20%; dưới 50 hộ, cỡ mẫu là 30% số hộ; nếu tổng số hộ có rừng trồng được khai thác dưới 10 hộ thì điều tra toàn bộ.

Có khoảng 13/31 cuộc điều tra mẫu có dạng cỡ mẫu được xác định cỡ mẫu cụ thể.

+ Dạng thứ hai: Cỡ mẫu được cố định bằng 1 số lượng cụ thể, ví dụ:

(i) Điều tra biến động dân số và kế hoạch gia đình (2024) ấn định cụ thể điều tra mẫu 7640 địa bàn, mỗi địa bàn chọn 40 hộ;

(ii) Điều tra chăn nuôi, căn cứ số lượng địa bàn có hoạt động chăn nuôi trong từng huyện để xác định số địa bàn chọn mẫu, cụ thể: Ở khu vực Thành thị huyện có từ 50 địa bàn trở lên: 15 địa bàn; có 30-49 địa bàn: điều tra 12 địa bàn mẫu; 10-29 địa bàn: 8 địa bàn mẫu; 5-10 địa bàn: 5 địa bàn; dưới 5 địa bàn điều tra toàn bộ. Ở khu vực nông thôn: Từ 500 địa bàn trở lên: 45 địa bàn; có 400-499 địa bàn: điều tra 35 địa bàn mẫu; 300-399 địa bàn: 25 địa bàn mẫu; 200-299 địa bàn: 18 địa bàn; 100-199 địa bàn: 12 địa bàn; dưới 100 địa bàn: điều tra 8 địa bàn;

(iii) Điều tra năng suất, sản lượng cây nông nghiệp, cỡ mẫu 2 cấp cố định: Số thôn mẫu/phường mẫu trong 1 huyện cho cây lúa là 6 thôn, nếu diện tích trồng từ 100ha-5.000ha; diện tích từ 5.000ha - 10.000 số mẫu là 8 thôn; diện tích từ 10.000 ha trở lên số mẫu là 10 thôn. Mỗi thôn mẫu/phường mẫu: 10 hộ. Tương tự các cây khác, dựa vào quy mô diện tích trồng cố định số lượng mẫu để điều tra;

(iv) Điều tra thu thập thông tin lập bảng cân đối Liên ngành và tính hệ số chi phí trung gian (Điều tra IO), cỡ mẫu là 20.000 doanh nghiệp;

(v) Khảo sát mức sống dân cư Việt Nam, cỡ mẫu gồm 72.054 hộ từ 4.177 địa bàn của Mẫu chủ 2019, chia làm 2 nhóm. Nhóm mẫu 1: gồm 46.995 hộ dân cư được chọn từ 3.133 địa bàn của Khảo sát mức sống dân cư năm 2024, gồm 2 phần: phần 1- có 37.596 hộ thu nhập và quyền sở chỉ số giá tiêu dùng và phần 2- Có 9.399 hộ thu nhập chi tiêu. Nhóm mẫu 2, gồm 25.059 hộ quyền sở chỉ số giá

tiêu dùng, gồm 2 phần: phần 1- có 9.399 hộ quyền số chỉ số giá tiêu dùng được chọn trong 3.133 địa bàn của Khảo sát mức sống dân cư 2024, mỗi địa bàn chọn 3 hộ ngoài các hộ đã được chọn cho Mẫu 1 và phần 2- có 15.660 hộ quyền số chỉ số giá tiêu dùng được chọn trong 1.044 địa bàn (ngoài mẫu Khảo sát mức sống dân cư năm 2024) từ Mẫu chủ 2019, mỗi địa bàn chọn 15 hộ;

(vi) Điều tra người khuyết tật năm 2023, là mẫu phân tầng được chọn theo 2 bước: Bước 1 chọn địa bàn điều tra; bước 2 chọn hộ điều tra. Số lượng địa bàn là 2.925 địa bàn. Mỗi địa bàn chọn điều tra 20 hộ mẫu.

Có khoảng 9/31 cuộc điều tra mẫu có dạng cỡ mẫu được xác định cỡ mẫu cụ thể.

+ Dạng thứ 3: Cỡ mẫu dành cho các cuộc điều tra giá hiện nay không thuộc 2 dạng trên, do phải xây dựng mạng lưới điều tra bao gồm các khu vực điều tra, điểm điều tra, mặt hàng điều tra. Việc này không cố định số lượng mẫu hay theo tỷ lệ cụ thể.

4.2. Thực trạng phân bổ mẫu

Xác định cỡ mẫu và phân bổ mẫu là hai công việc thực hiện trong quá trình thiết kế mẫu, chúng có mối quan hệ mật thiết không thể tách rời. Tương tự như phần xác định cỡ mẫu, phân bổ mẫu ở các cuộc điều tra chọn mẫu của Tổng cục Thống kê tạm phân thành 02 dạng cơ bản. Dạng thứ nhất, khi biết cỡ mẫu của tổng thể, dùng các phương pháp phân bổ mẫu chia số đơn vị mẫu cho các tổ. Dạng thứ 2, ta xác định được các tổ trước, từng tổ xác định số lượng mẫu điều tra, sau đó mới tính toán cỡ mẫu tổng thể (gọi là phân bổ mẫu trong quá trình xác định cỡ mẫu của tổng thể). Cụ thể:

+ Dạng thứ nhất: Cỡ mẫu của tổng thể phân bổ cho các tổ theo phương pháp phân bổ cụ thể.

(i) Trong Tổng điều tra dân số và nhà ở, có 2 lần phân bổ mẫu: Phân bổ mẫu địa bàn

điều tra cho mỗi huyện bằng phương pháp tỷ lệ căn bậc 2 của quy mô; phân bổ mẫu hộ cho mỗi địa bàn điều tra;

(ii) Trong Tổng điều tra nông thôn, nông nghiệp, có 2 lần phân bổ mẫu: Phân bổ mẫu địa bàn điều tra cho mỗi tỉnh bằng phương pháp tỷ lệ căn bậc 2 của quy mô là số hộ; phân bổ mẫu hộ cho mỗi địa bàn điều tra, mỗi địa bàn 9 hộ;

(iii) Trong Tổng điều tra kinh tế, đơn vị điều tra là cơ sở sản xuất kinh doanh cá thể. Tổng thể mẫu là 3,5% tổng số cơ sở SXKD. Phân bổ mẫu cho từng tỉnh dựa vào phương pháp tỷ lệ căn bậc 2 của quy mô cơ sở của tỉnh. Trong mỗi tỉnh, phân bổ mẫu cho từng ngành bằng phương pháp tỷ lệ căn bậc 2 quy mô từng ngành;

(iv) Điều tra dân số và nhà ở: cũng phân bổ mẫu theo cách này và sử dụng phương pháp phân bổ mẫu tỷ lệ căn bậc 2 của quy mô.

Như vậy, với các cuộc điều tra mà xác định cỡ mẫu dạng 1 và dạng 2 thì thường đi kèm phân bổ mẫu theo dạng 1. Hầu hết phân bổ mẫu theo dạng 1 thì đều lựa chọn phương pháp phân bổ mẫu tỷ lệ căn bậc 2 của quy mô.

+ Dạng thứ 2: Phân bổ mẫu trong quá trình xác định cỡ mẫu của tổng thể.

Ví dụ, trong Tổng điều tra kinh tế, đơn vị điều tra là doanh nghiệp, đầu tiên chia thành 2 tổ: Tổ điều tra chuyên ngành và tổ điều tra chuyên đề. Trong các tổ điều tra mẫu xác định theo các tổ nhỏ hơn. Đối với điều tra chuyên ngành chia doanh nghiệp thành 4 tổ: Doanh nghiệp có vốn nhà nước chiếm trên 50% vốn điều lệ; Doanh nghiệp có vốn đầu tư trực tiếp nước ngoài; Doanh nghiệp ngoài nhà nước có từ 100 lao động trở lên, riêng doanh nghiệp hoạt động ngành dịch vụ có từ 50 lao động trở lên. Mỗi tổ có quy định về số lượng mẫu cụ thể. Như vậy, từ việc biết cỡ mẫu từng tổ ta tổng hợp chung lên sẽ biết cỡ doanh nghiệp ta điều tra trong tổng điều tra cơ sở kinh tế năm 2021.

Ví dụ, trong Tổng điều tra mà xác định cỡ mẫu dạng 3 trình bày ở trên có phân bố mẫu theo dạng 2. Tức là phân bố mẫu cho các tổ trước rồi mới xác định cỡ mẫu sau. Đề tài tổng kết được rằng, gần như tất cả các cuộc điều tra giá, điều tra đơn vị điều tra doanh nghiệp thì có sử dụng phương pháp xác định cỡ mẫu, phân bố mẫu này.

5. Đánh giá chung

Cả 3 dạng xác định cỡ mẫu hiện tại của các cuộc điều tra Tổng cục Thống kê đang thực hiện đều dựa vào kinh nghiệm thực tế của người thiết kế mẫu. Về mặt lý thuyết cỡ mẫu không phụ thuộc vào quy mô của tổng thể mà phụ thuộc vào mức độ biến động của chỉ tiêu lựa chọn điều tra. Cách quy định như vậy dẫn đến phân tổ nào có quy mô lớn sẽ có nhiều đơn vị được chọn vào mẫu và ngược lại phân tổ nào có quy mô nhỏ thì số lượng đơn vị chọn mẫu ít hơn (dạng xác định mẫu thứ nhất). Cỡ mẫu như vậy sẽ dẫn đến có tổ quá thừa số lượng đơn vị mẫu và có phân tổ cỡ mẫu không đủ đại diện, hay nói cách khác thiếu số lượng mẫu cần điều tra. Phần 1.2 nêu xác định cỡ mẫu cho các phương pháp chọn mẫu trên đã giúp xác định cỡ mẫu theo công thức phụ thuộc vào 2 yếu tố là mức độ biến thiên của chỉ tiêu nghiên cứu và sai số chọn mẫu cho phép đối với cuộc điều tra chứ hoàn toàn không phụ thuộc vào quy mô của tổng thể.

Tương tự, đối với dạng xác định cỡ mẫu thứ 2 cũng dựa vào cảm nhận và kinh nghiệm điều tra. Ví dụ, trong điều tra biến động dân số và kế hoạch hóa gia đình, cho rằng, để đại diện cho cấp tỉnh thì số mẫu là 305.600 hộ. Tuy nhiên, dựa vào kết quả thu được từ các tỉnh, tính toán lại cỡ mẫu khi lựa chọn 1 số chỉ tiêu để tính toán cỡ mẫu thì cỡ mẫu như vậy là quá lớn.

Lượng đồ mẫu và phương án điều tra cũng như báo cáo kết quả điều tra hầu như xác định cỡ mẫu đều không được trình bày cụ thể. Cỡ

mẫu được xác định chủ yếu dựa theo phương pháp chuyên gia, có tham vấn của chuyên gia, tổ chức nước ngoài để ấn định tỷ lệ hoặc cỡ mẫu cụ thể. Không tính toán sai số chọn mẫu, sử dụng thông tin của cuộc điều tra trước đó để làm căn cứ đánh giá đại diện mẫu và đề xuất mẫu cho cuộc điều tra tiếp theo.

Cần phải đưa ra nguyên tắc, quy trình xác định cỡ mẫu, phân bố mẫu để đảm bảo chất lượng chỉ tiêu được tính toán từ mẫu. Đề tài "Nghiên cứu xác định cỡ mẫu và phân bố mẫu cho một số loại hình điều tra thống kê tại Tổng cục Thống kê" nghiên cứu đề xuất nguyên tắc, quy trình xác định cỡ mẫu, phân bố mẫu cho một số cuộc điều tra.

Tài liệu tham khảo

1. PGS. TS Tăng Văn Khiên (2003), Điều tra chọn mẫu và ứng dụng trong công tác thống kê.
2. PGS. TS Tăng Văn Khiên (2006), Phương pháp luận thống kê.
3. NCV. Lê Văn Dục (1998), Điều tra chọn mẫu, Lý luận và ứng dụng trong công tác dân số và kế hoạch hóa gia đình.
4. Tổng Cục Thống Kê (2016), Từ điển thống kê; Nhà xuất bản Thống kê.
5. Tổng cục Thống kê, Phương án điều tra của 31 cuộc điều tra chọn mẫu, điều tra toàn bộ kết hợp chọn mẫu trong Chương trình điều tra thống kê quốc gia.
6. NCV. Lê Văn Dục (2004), Đề tài "Nghiên cứu các nguyên tắc xác định cỡ mẫu và phân bố mẫu ứng dụng vào thực tiễn điều tra thống kê ở nước ta.
7. TS. Đinh Thị Thúy Phương (2022), Đề tài "Nghiên cứu xây dựng quy trình lập dàn mẫu quốc gia và hoàn thiện phương pháp biên soạn chỉ số giá sản xuất công nghiệp và dịch vụ theo năm gốc 2020 ở Việt Nam".
8. NCV. Lê Văn Dục (2002), Đề tài "Nghiên cứu sai số trong điều tra thống kê và biện pháp khắc phục".