

TRAO ĐỔI NGHIỆP VỤ

PHƯƠNG PHÁP CHỌN MẪU TRONG NGHIÊN CỨU XÃ HỘI HỌC

TÔN THIỆN CHIẾU

Trong xã hội học cũng như trong nhiều ngành khoa học xã hội khác, khi nghiên cứu một đề tài nào đó tốt nhất là chúng ta nghiên cứu toàn bộ tổng thể cần nghiên cứu. Các cuộc điều tra dân số là một ví dụ về cách nghiên cứu như vậy. Phương pháp này cho phép phản ánh đúng đắn và chính xác thực trạng của khách thể, song nó lại gặp phải nhiều hạn chế như điều kiện tài chính, nhân lực, vật tư, thời gian, v.v... Vì vậy các cuộc nghiên cứu xã hội học thường được tiến hành bằng phương pháp chọn mẫu.

Mẫu là một phần của khách thể nghiên cứu (tổng thể), mà trên đó chúng ta tiến hành thu thập thông tin cần thiết cho vấn đề cần nghiên cứu. Phương pháp nghiên cứu dựa trên các kết quả ở mẫu để rồi suy ra cho toàn bộ tổng thể với một độ chính xác nhất định được gọi là phương pháp nghiên cứu chọn mẫu: Ví dụ, khi nghiên cứu tính tích cực lao động của công nhân một xí nghiệp, nếu chúng ta tiến hành thu thập thông tin ở tất cả các công nhân của xí nghiệp, thì đó là nghiên cứu toàn bộ. Nếu chúng ta thu thập thông tin ở một số công nhân tại các phân xưởng, hoặc toàn bộ công nhân của một phân xưởng, thì đó là phương pháp nghiên cứu chọn mẫu của tổng thể xí nghiệp.

Yêu cầu cơ bản của phương pháp mẫu là tính đại diện của nó. Tính đại diện là sự tái tạo những đặc trưng của tổng thể trong mẫu. Nói cách khác, đó là sự phù hợp giữa kết cấu của tổng thể với kết cấu mẫu được hiểu theo quan điểm thống kê toán. Nghĩa là có thể tìm được ước lượng của các tham số của tổng thể nhờ các tham số của mẫu.

Trong xã hội học, khách thể nghiên cứu là tập hợp các cá thể, có những đặc điểm khác nhau.

Số các đặc điểm này là rất nhiều. Vì thế, một tập hợp nhỏ các cá thể (mẫu) không thể đại diện cho tổng thể theo tất cả các đặc điểm. Vậy, trong nghiên cứu xã hội học, thế nào là một mẫu đại diện cho khách thể nghiên cứu?

Ở đây, tính đại diện của mẫu được hiểu từ quan điểm của các nhiệm vụ nghiên cứu. Nghĩa là cấu trúc của mẫu phải tương ứng với cấu trúc của tổng thể trên cơ sở các giả thiết cơ bản. Ví dụ, khi nghiên cứu thị hiếu thời trang của thanh niên, người nghiên cứu có thể đưa ra giả thiết về sự phụ thuộc của thị hiếu vào trình độ học vấn của họ. Khi đó nếu biết trong tổng thể nghiên cứu có 20% thanh niên có trình độ đại

học, 45% thanh niên có trình độ cấp 3 và 35% thanh niên có trình độ văn hóa cấp 2 trở xuống, thì ta phải giữ đúng tỷ lệ này trong mẫu. Tất nhiên nếu giả thuyết nghiên cứu còn cho rằng: thị hiếu thời trang còn phụ thuộc vào nơi ở (nội thành và ngoại thành) và nghề nghiệp, thì ta cũng phải đảm bảo tỷ lệ trong mẫu phù hợp với tỷ lệ trong tổng thể nghiên cứu theo các chỉ báo này.

Như vậy, vì không thể đại diện cho tổng thể theo tất cả tính chất của nó, nên cấu trúc của mẫu hoàn toàn do đặc tính của các giả thuyết nghiên cứu quy định. Có nghĩa là, trong mẫu ta chỉ tính đến các đặc trưng là tính chất của tổng thể nghiên cứu có ý nghĩa cần bản với các nhiệm vụ nghiên cứu được đặt ra.

Các sai lệch về cấu trúc thống kê giữa mẫu và tổng thể sẽ dẫn đến những sai số trong kết quả thu được. Khi sử dụng phương pháp chọn mẫu, chúng ta phải chú ý đến hai loại sai số ngẫu nhiên và sai số hệ thống.

Sai số ngẫu nhiên là sai số thống kê đặc thù cho phương pháp chọn mẫu do sự vi phạm ngẫu nhiên các thủ tục thu thập thông tin gây ra. Độ lớn của sai số loại này có thể tính được nên nó không ảnh hưởng nhiều đến kết quả nghiên cứu.

Sai số hệ thống là sự tái tạo lại không đầy đủ trong mẫu những đặc tính của tổng thể. Các sai số hệ thống có thể làm cho các kết quả nghiên cứu không phản ánh đúng bản chất của hiện tượng, làm mất ý nghĩa thực tiễn của cuộc nghiên cứu. Nghĩa là các kết quả chỉ đúng cho những người cá thể riêng lẻ, mà không có ý nghĩ gì đối với khách thể nghiên cứu. Với những kết quả như vậy thì không thể đề xuất được gì nhằm giải quyết nhiệm vụ được đặt ra.

Để khắc phục các sai số kể trên, nghĩa là để tăng tính đại diện của mẫu, ta thường tăng độ lớn của mẫu lên. Độ lớn của mẫu, hay còn gọi là dung lượng của mẫu, là số lượng người được đưa vào trong mẫu để thu thập thông tin. Dung lượng mẫu phụ thuộc vào tích đồng nhất các đặc điểm của khách thể nghiên cứu và nhiệm vụ nghiên cứu. Mức độ đồng nhất của các đặc điểm của khách thể càng ít, yêu cầu độ chính xác càng cao thì dung lượng mẫu phải càng lớn. Ngược lại tính đồng nhất của khách thể càng cao thì dung lượng mẫu càng bé. Thông thường dung lượng mẫu được xác định theo công thức sau đây :

$$n = \frac{t^2 \sigma^2}{\epsilon^2} \text{ hay } n = \frac{t^2 (1-f)f}{\epsilon^2}$$

Trong công thức này, n và dung lượng mẫu

- độ chính xác được cho trước theo yêu cầu của cuộc nghiên cứu ;

σ^2 - phương sai của dấu hiệu cần tìm ;

t là một ẩn số phụ mà ta có thể tính được qua xác suất tin cậy cho trước :

$$\gamma = 2\phi(t)$$

Tra bảng phụ lục và hàm số laplace, ta tính được t với f cho trước.

f là tỷ lệ của dấu hiệu được xét “có” hoặc “không có” trong mẫu.

Theo công thức trên, thì dung lượng của mẫu (n) sẽ không được xác định nếu không biết phương sai σ^2 hoặc f. Vì vậy, trước khi tính toán dung lượng mẫu, người nghiên cứu phải tiến hành điều tra thử để xác định σ^2 và f. Đôi khi, để đơn giản, người ta còn xác định dung lượng mẫu theo công thức sau đây .

$$n = \frac{t^2}{4\varepsilon^2}$$

Các công thức này giúp ta tính được số lượng người cần thiết để thu thập thông tin với độ chính xác và xác suất tin cậy cho trước.

Trong nhiều cuộc nghiên cứu, một vấn đề được đặt ra là làm sao lấy được mẫu đại diện khi chưa biết kết cấu của tổng thể. Các phương pháp lựa chọn ngẫu nhiên xác suất mà cơ sở của chúng là lý thuyết xác suất và thống kê toán sẽ cho phép ta giải quyết vấn đề này. Có nhiều phương pháp để lựa chọn các cá thể vào mẫu như:

- Mẫu ngẫu nhiên đơn giản.
- Mẫu ngẫu nhiên hệ thống (máy móc).
- Mẫu tỷ lệ.
- Mẫu hạn ngạch.
- Mẫu ỏ (loạt).
- Mẫu phân lớp
- Mẫu tổ hợp nhiều bậc...

Tùy theo từng cuộc nghiên cứu cụ thể mà ta chọn phương pháp nào cho thích hợp với khách thể nghiên cứu. Song có hai phương pháp được sử dụng nhiều nhất trong các cuộc nghiên cứu là mẫu ngẫu nhiên (bao gồm cả ngẫu nhiên hệ thống lẫn ngẫu nhiên đơn giản) và mẫu phân lớp.

Trong các cuộc nghiên cứu xã hội học có tính chất địa phương (huyện, quận hoặc nhà máy, xí nghiệp), người ta thường sử dụng mẫu ngẫu nhiên đơn giản. Để tổ chức lấy mẫu ngẫu nhiên đơn giản phải có một lượng thông tin rất lớn: đó là bảng kê tất cả các đơn vị của tổng thể. Dựa trên bảng kê này ta lần lượt lấy ra một cách hù họa từng người cho đến khi đạt được dung lượng mẫu cần thiết. Phương pháp chọn mẫu ngẫu nhiên hệ thống cho phép ta đơn giản hóa cách lấy mẫu ngẫu nhiên đơn giản. Dựa vào dung lượng mẫu cần thiết và số lượng người của tổng thể, ta quy định cứ bao nhiêu người lại lấy ra một người để nghiên cứu. Ví dụ tổng thể có 2.500 người, dung lượng mẫu là 2.500 người thì cứ 10 người ta lấy ra 1 người. Đơn vị đầu tiên phép lấy mẫu này thường được xác định một cách ngẫu nhiên (có thể bằng các bảng số ngẫu nhiên).

Trong những cuộc nghiên cứu có quy mô lớn hơn, các phương pháp lấy mẫu như trên rất khó thực hiện. Khi đó người ta sử dụng mẫu phân lớp. Bước đầu tiên của việc lấy mẫu phân lớp là phân chia tổng thể thành các lớp thuần nhất. Các giả thuyết nghiên cứu là cơ sở để phân chia các lớp. Sau đó, ta có thể lấy mẫu theo phương pháp ngẫu nhiên đơn giản, hệ thống hoặc mẫu ỏ theo các lớp đã được phân chia.

Phương pháp chọn mẫu không cho ta một lời giải chung, duy nhất về cách thức lấy mẫu cũng như dung lượng của mẫu trong mỗi trường hợp. Tùy từng trường hợp cụ thể, yêu cầu và mục đích của cuộc nghiên cứu mà ta sử dụng một phương pháp lựa chọn thích hợp. Có hệ giả thuyết nghiên cứu chính xác, có phương pháp luận thống kê, chúng ta có thể vận dụng một cách có hiệu quả phương pháp chọn mẫu trong các nghiên cứu xã hội học cụ thể.