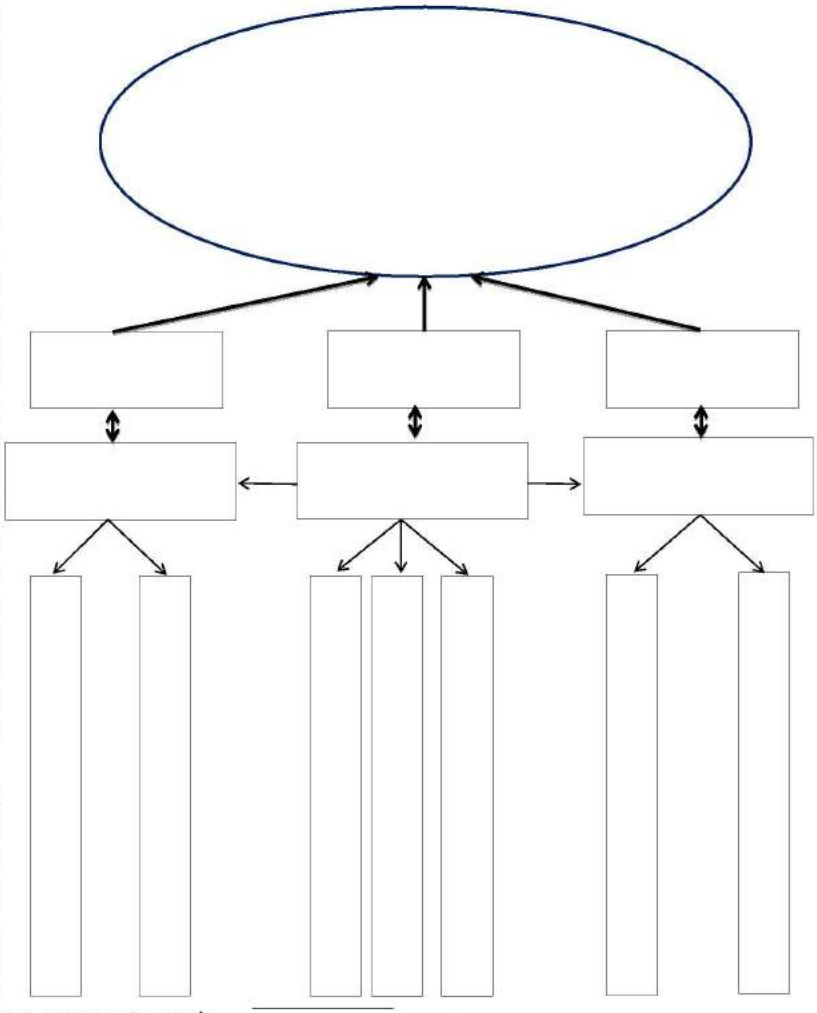
XÁC ĐỊNH MẪU TRONG THU THẬP THÔNG TIN ĐỊNH LƯỢNG KHI NGHIÊN CỨU LIÊN QUAN TỚI CÔNG ĐOÀN, NGƯỜI LAO ĐỘNG

ThS. TRƯƠNG NGỌC THẮNG*

1. Đặt vấn đề

Thực tế cho thấy, trong các nghiên cứu khoa học nói chung và trong nghiên cứu liên quan đến Công đoàn, người lao động nói riêng, để có cơ sở khoa học cho việc khái quát vấn đề nghiên cứu, trên cơ sở đó đưa ra những khuyến nghị, những giải pháp có hiệu quả cần phải có thông tin từ nhiều nguồn khác nhau. Tuy nhiên, để khái quát thực trạng về vấn đề nghiên cứu rất cần có thông tin định lượng. Nhưng cũng cần nhận thức được, nếu những khuyến nghị, những giải pháp căn cứ vào nguồn thông tin không đại diện cho tổng thể tất nhiên khuyến nghị đó cũng không có tính khả thi. Vậy, thông tin định lượng cần đáp ứng yêu cầu gì? Làm thế nào để có thông tin định lượng phù hợp? Đặc biệt, làm thế nào để số lượng đơn vị thu thập thông tin mặc dù chỉ là một phần thuộc tổng thể nhưng vẫn có cơ sở để khái quát cho toàn bộ khách thể nghiên cứu? Bài viết được trình bày dưới đây hy vọng trả lời một phần những câu hỏi này.

Thông tin nói chung và thông tin định lượng nói riêng để phục vụ phân tích và khái quát thực trạng vấn đề nghiên cứu liên quan đến đoàn viên công đoàn, người lao động phải đáp ứng được 3 yêu cầu cơ bản, đó là: tính đại diện, tính hiệu lực và tính chính xác. Nếu không đáp ứng một trong 3 yêu cầu này, những phân tích, lý giải, khuyến nghị đều không phù hợp. Những yêu cầu đối với thông tin có thể trình bày khái quát trong sơ đồ dưới đây:



* Trường Đại học Công đoàn

2. Chọn mẫu trong nghiên cứu liên quan đến Công đoàn, người lao động

Khi thu thập thông tin liên quan đến người lao động trong các doanh nghiệp phần lớn chúng ta chỉ tiến hành thu thập một số lượng nhất định người lao, nhưng những kết luận của chúng ta lại không chỉ dừng lại ở những người cung cấp thông tin đó mà luôn mong muốn hoặc hướng tới kết luận cho toàn bộ người lao động. Vậy để những thông tin từ mẫu đảm bảo độ tin cậy khái quát cho tổng thể chúng ta cần quan tâm đến hai nội dung cơ bản đó là xác định số lượng đơn vị cần thu thập thông tin (dung lượng mẫu) và cách chọn số lượng đã xác định đó.

2.1. Xác định số lượng đơn vị để thu thập thông tin (dung lượng mẫu)

Để có được thông tin định lượng về những nội dung liên quan đến đề tài nghiên cứu, đương nhiên chúng ta phải sử dụng phương pháp thu thập thông tin phù hợp. Vậy, liệu có cần phải thu thập thông tin tổng thể đơn vị nghiên cứu hay không? Nếu thu thập thông tin từ mẫu, cần phải điều tra bao nhiêu đơn vị là đủ? Với số đơn vị đã xác định đó, chọn như thế nào để đảm bảo đại diện nhất? Đó là những câu hỏi thường xuyên xuất hiện trong quá trình triển khai nghiên cứu đối với những người tổ chức điều tra.

Nghiên cứu tổng thể có một số hạn chế nhất định như kinh phí nhiều, thời gian kéo dài, cần nhiều cộng tác viên,... Nhưng, nếu chỉ tiến hành chọn một số đối tượng để thu thập thông tin mà vẫn đảm bảo có thể khái quát cho tổng thể nghiên cứu trong một đề tài nghiên cứu cụ thể nhà nghiên cứu cần quan tâm đến một số yêu cầu cơ bài đối với chọn mẫu.

Trong quá trình thu thập thông tin liên quan đến Công đoàn, người lao động để phục vụ đề tài nghiên cứu đã chọn, việc chọn bao nhiêu đơn vị mẫu để tiến hành khảo sát cũng là một vấn đề phải được tính toán hết sức kỹ lưỡng và khoa học. Vì số lượng đơn vị mẫu sẽ ảnh hưởng đến các vấn đề như: sai số trong điểu tra, kinh phí điều tra, thời gian thực hiện,... Chính vì thế trước khi chọn mẫu cần phải căn cứ vào tình hình thực tế và các yêu cầu đặt ra của người nghiên cứu để quyết định số lượng cụ thể.

Thực tế ở Việt Nam hiện nay việc tính toán dung lượng mẫu chưa được chú ý đúng mức. Nhiều cuộc điều tra đưa ra số lượng người cần khảo sát nhưng không đưa ra cơ sở khoa học để giải thích là tại sao phải điều tra số lượng như vậy mà không phải nhiều hơn hay ít hơn. Hay nói một cách khác, dung lượng mẫu của một cuộc điều tra nhiều khi được xác định khá cảm tính. Vậy khi tính toán dung lượng mẫu cho cuộc điều tra (thu thập thông tin) về Công đoàn, người lao động chúng ta cần chú ý đến điều gì?

Dung lượng mẫu được xác định căn cứ vào những đòi hỏi về khoa học, những điều kiện về tài chính, nhân lực, vật lực... Tuy nhiên cần đáp ứng những đặc điểm sau:

- Số lượng người trả lời không nhỏ hơn 30 (mẫu tối thiểu). Đây là giá trị được các nhà thống kê xem coi là số lớn khi tính toán các giá trị trung bình;
- Kích thước mẫu lớn tới mức nào mà ngân quỹ, thời hạn và yếu tố nhân sự cho phép;
- Bảo đảm sai số chọn mẫu nhỏ hợp lý (xác định sai số do chủ nhiệm đề tài xác định và được cơ quan giao thực hiện đề tài đồng ý).

Để xác định dung lượng mẫu cần căn cứ vào chỉ báo thực nghiệm cần thu thập thông tin, trong quá trình thu thập thông tin định lượng liên quan đến công đoàn, người lao động có rất nhiều các chỉ báo cần thu thập. Ví dụ trình độ đào tạo của cán bộ Công đoàn, thời gian làm ngoài giờ của người lao động, thực trạng giao kết hợp đồng của người lao động... Tuy nhiên, nhà nghiên cứu có thể gộp lại để tính toán tỉ lệ phần trăm. Ví dụ: Tỉ lệ phần trăm cán bộ Công đoàn có chứng chỉ Đại học phần Công đoàn và xác định giá trị trung bình (thu nhập trung bình của người lao động, thời gian trung bình làm ngoài giờ trong 1 tháng).Sau đó xác định một số nhiệm vụ (thông tin) chính cần thu thập và vận dụng công thức dưới đây để xác định dung lượng mẫu đáp ứng các yêu cầu của nghiên cứu.

Trong trường hợp chúng ta cần ước tính tỷ lệ phần trăm, ví dụ xác định phần trăm người lao động biết về luật lao động, dung lượng mẫu có thể được xác định theo công thức sau:

$$n = \frac{Nt^2pq}{N\varepsilon^2 + t^2pq}$$

Trong đó: *N - Kích thước của tổng thể*
t - hệ số tin cậy
p/q -tỉ lệ tán thành/phản đối
ε - Phạm vi sai số tối đa cho phép

Ở đây p và q được đo lường bằng số thập phân và xác suất của các tiêu thức thay phiên, tức là những câu trả lời loại trừ nhau trong câu hỏi nhị phân. Thí dụ, nếu p là xác suất câu trả lời có gặp khó khăn về vốn, thì q là xác suất câu trả lời không gặp khó khăn

về vốn. Vì p+q=1 cho nên tích pq lớn nhất là 0,25. Cho nên, chúng ta có thể thay tích của pq bằng giá trị 0,25 vào công thức trên để thuận lợi cho việc tính toán.

Khi chúng ta ước lượng giá trị trung bình, thí dụ, xác định thu nhập trung bình một tháng của người lao động, chúng ta có thể tính dung lượng mẫu n theo công thức sau:

$$n = \frac{Nt^2\sigma^2}{N\varepsilon^2 + t^2\sigma^2}$$

Trong đó: *N - Kích thước của tổng thể*

t - hệ số tin cậy

σ^2 (sigma bình phương) - phương sai

ε (epsilon)- phạm vi sai số tối đa cho phép

Trong các công thức nêu trên, giá trị của N - kích thước của tổng thể, chúng ta có thể tìm thấy trong kết quả tổng điều tra dân số, hoặc trong tài liệu của các cơ quan đơn vị hành chính. Hệ số tin cậy t và sai số chọn mẫu e thường do người đặt hàng nghiên cứu hoặc bản thân các nhà nghiên cứu đặt ra ở mức độ hợp lý và chấp nhận được. Thí dụ, độ tin cậy tối thiểu là 95%, hệ số tin cậy t = 1,96, sai số dưới 5%, ε = 0,05.

Tham số khó xác định hơn cả là phương sai của tổng thể σ^2. Trong thực tế, khi tính dung lượng mẫu điều tra chúng ta thường không có tài liệu để tính phương sai chung. Tuy nhiên chúng ta được phép sử dụng những giải pháp dưới dây và vẫn đảm bảo tính đại diện.

1) Sử dụng phương sai lớn nhất từ các lần nghiên cứu trước đó.

2) Lấy phương sai của các tiêu chí tương tự ở những địa bàn tương tự.

3) Tổ chức điều tra thử để tính phương sai mẫu s^2 và dùng nó để thay thế phương sai chung. (Thông thường trong quy trình nghiên cứu luôn có bước nghiên cứu thử nghiệm).

4) Căn cứ vào ý kiến các chuyên gia hiểu biết về địa bàn và vấn đề nghiên cứu để ước đoán phương sai của tiêu chí được xem xét.

5) Lấy mẫu theo hai bước. Trước hết lấy một mẫu n_1 để tính toán phương sai. Trên cơ sở phương sai này tính dung lượng mẫu n. Nếu như $n > n_1$ thì chúng ta cần phải chọn thêm một lượng mẫu bổ sung n_2. Còn nếu như $n_1 \geq n$ thì sử dụng n_1 làm dung lượng mẫu chuẩn.

Với một số công thức xác định dung lượng mẫu trên đây, nhóm nghiên cứu có thể căn cứ vào yêu cầu của đề tài, căn cứ vào thực tế kinh phí, thời gian, nhân lực, vật lực để xác định dung lượng mẫu cụ thể. Nếu có nhiều chỉ báo cần thu thập, mỗi chỉ báo sẽ xác định được dung lượng mẫu cụ thể, sau đó sử dụng dung lượng mẫu lớn nhất trong số các dung lượng mẫu đã xác định và tiến hành thu thập thông tin với dung lượng đó.

Ví dụ đối với chỉ báo thứ nhất chúng ta xác định dung lượng mẫu là 500, chỉ báo 2 dung lượng mẫu là 800, tương tự các mẫu khác là 600, 700, 1000. Trong trường hợp này chúng ta sẽ lựa chọn dung lượng mẫu để thu thập thông tin chính là 1000.

Tuy nhiên, ngoài cách xác định dung lượng mẫu trên, khi thu thập thông tin liên quan đến công đoàn, người lao động, chúng ta có thể vận dụng công thức chung dưới đây vẫn có thể đáp ứng được yêu cầu của thu thập thông tin:

$$n = \frac{Nt^2 0.25}{N\varepsilon^2 + t^2 0.25}$$

Trong đó: *N - Kích thước của tổng thể*

t - hệ số tin cậy (xác định theo bảng dưới

ε (epsilon)- phạm vi sai số tối đa cho phép

Nếu sử dụng công thức này chúng ta xác định được dung lượng mẫu lớn nhất so với các công thức còn lại, như vậy có thể đảm bảo yêu cầu về sai số chọn mẫu cho nhiều chỉ báo đo lường khác nhau.

Trong trường hợp không hoặc khó xác định được kích thước của tổng thể (N) nhóm nghiên cứu có thể chia cả tử số và mẫu số cho N để triệt tiêu N. Cụ thể, như công thức dưới đây:

$$n = \frac{0.25t^2}{\varepsilon^2}$$

Trong các công thức trên chúng ta sử dụng giá trị của t phụ thuộc và độ tin cậy và được tính sẵn theo hàm $\Phi_{(t)}$ của Lia-pu-nốp dưới đây.

Một số giá trị của hệ số tin cậy t được tính sẵn theohàm $\Phi_{(t)}$ của Lia-pu-nốp:

T	$\Phi_{(t)}$	T	$\Phi_{(t)}$	t	$\Phi_{(t)}$
1.0	0.6827	1.7	0.9109	2.4	0.9836
1.1	0.7287	1.8	0.9281	2.5	0.9876
1.2	0.7699	1.9	0.9426	2.6	0.9907
1.3	0.8064	2.0	0.9545	2.7	0.9931
1.4	0.8385	2.1	0.9643	2.8	0.9949
1.5	0.8684	2.2	0.9722	2.9	0.9963
1.6	0.8904	2.3	0.9766	3.0	0.9973

Trong bảng trên, nếu chúng ta muốn độ tin cậy trong quá trình khái quát từ mẫu cho tổng thể là

95,45% thì giá trị t để xác định dung lượng mẫu là 2 (Thực tế, nếu chúng ta thường chọn độ tin cậy 95% và t được xác định là 1,96).

Các công thức tính dung lượng mẫu nêu trên mới chỉ cho chúng ta biết số lượng đơn vị nghiên cứu của *khối mẫu chính* và là dung lượng mẫu tối thiểu cần thu thập. Kinh nghiệm thực tế cho thấy các điều tra viên có thể gặp thái độ bất hợp tác từ phía các cá nhân như bị từ chối trả lời, vì vậy khi tính toán dung lượng mẫu cần phải chọn thêm khoảng 10% mẫu chính.

Việc cộng thêm một tỷ lệ nhất định vào khối mẫu chính thì *số* lượng người được hỏi trong đợt điều tra vẫn hoàn toàn không đổi. Chúng ta chỉ khảo sát cho đủ số lượng các đơn vị nghiên cứu thì chấm dứt đợt khảo sát. Việc tính thêm khối mẫu dự phòng chỉ nhằm phòng ngừa những rủi ro ngẫu nhiên khi tìm gặp các đơn vị nghiên cứu. Nó giúp việc chọn mẫu không phải làm đi làm lại nhiều lần. Do vậy nâng cao được tính khoa học và giảm được sai số mẫu.

Để thông tin liên quan đến công đoàn, người lao động có khả năng khái quát cho tổng thể (đại diện cho tổng thể) ngoại nhiệm vụ xác định số lượng các đơn vị thu thập thông tin (dung lượng mẫu) chúng ta cũng cần phải xác định cách chọn phù hợp.

2.2. Xác định cách chọn mẫu (Chọn như thế nào)

Trong nghiên cứu khoa học nói chung và cụ thể là nghiên cứu liên quan đến Công đoàn, người lao động, nhiệm vụ xác định chọn bao nhiêu đơn vị là điều kiện bắt buộc để đảm bảo khả năng khái quát cho tổng thể nghiên cứu; tuy nhiên việc xác định số đơn vị nghiên cứu chỉ đáp ứng được điều kiện cần của tính đại diện. Bên cạnh đó, xác định cách chọn cũng là yếu tố ảnh hưởng đến tính đại diện của thông tin. Trong phương pháp chọn mẫu có nhiều cách chọn khác nhau như chọn chọn mẫu ngẫu nhiên đơn giản, Chọn mẫu ngẫu nhiên hệ thống, chọn mẫu phân tầng - ngẫu nhiên, Chọn mẫu cụm, chọn ngẫu nhiên cụm nhiều giai đoạn,...

Tuy nhiên đa số những trường hợp nghiên cứu định lượng, để đảm bảo tính đại diện, theo chúng tôi nên sử dụng cách chọn cụm nhiều giai đoạn là phù hợp hơn.

Các cách chọn mẫu khác thường rất tốn kém khi thiết kế, thậm chí nhiều khi không thể thực hiện được bởi vì không có đủ thông tin, do vậy không thể lập được khung mẫu. Trên thực tế việc lập khung mẫu đối với các đơn vị nghiên cứu lớn như các tỉnh, thành trong cả nước, các khu công nghiệp, các loại hình doanh nghiệp,.. là có tính khả thi, Nhưng việc lập khung mẫu bao gồm toàn bộ các cá nhân hoặc các hộ gia đình hoặc số lượng công nhân trong tổng thể là việc làm hết sức khó khăn, thậm chí không khả thi.

Mẫu cụm rất hay được thiết kế theo khu vực địa lý. Thí dụ một cụm có thể là 1 tỉnh, thành, 1 huyện. Những đơn vị tạo thành cụm mẫu sẽ nằm gọn trong một khu vực địa lí nhất định; các khu công nghiệp, các loại hình doanh nghiệp,...

Quy trình chọn mẫu cụm:

Bước 1. Phân chia tổng thể thành các nhóm riêng biệt theo những tiêu chí nhất định, thường là dấu hiệu về địa lý. Các nhóm này được gọi là các cụm. Lập danh sách của tất cả cụm này. Dùng danh sách nay làm khung mẫu.

Thí dụ: xác định mỗi tỉnh, thành phố của Việt nam là một cụm mẫu. Lập danh sách 63 tỉnh thành của Việt Nam làm khung mẫu. Hoặc xác định mỗi loại ngành nghề là một đơn vị trong cụm.

Bước 2. Trên cơ sở khung mẫu lựa chọn ra một hoặc một số cụm mẫu đủ cho nghiên cứu bằng phương pháp ngẫu nhiên đơn giản hoặc ngẫu nhiên hệ thống.

Bước 3. Từ cụm mẫu đã chọn được nhà nghiên cứu có thể điều tra toàn bộ nếu số lượng các đơn vị nghiên cứu trong đó không quá nhiều, và những điều kiện về tài chính, tổ chức của nghiên cứu cũng cho phép. Ngược lại nếu số lượng các đơn vị nghiên cứu quá lớn, các điều kiện về tổ chức, về tài chính của nghiên cứu không cho phép thì có thể tiếp tục phân chia cụm mẫu đã xác định được thành những cụm nhỏ hơn, lập khung mẫu cho các cụm này và tiếp tục chọn. Lúc này chọn mẫu cụm đã chuyển thành chọn mẫu cụm nhiều giai đoạn.

Lợi thế cơ bản của mẫu cụm là dùng hữu ích trong những trường hợp không thể có đầy đủ danh sách của tổng thể; điều tra viên thường không mất nhiều công sức và thời gian thu thập thông tin từ các đơn vị nghiên cứu trong cụm. Vì khoảng cách không gian giữa các đơn vị khảo sát rất gần nhau.

Tuy nhiên, mẫu cụm có nhược điểm lớn là sai số chọn mẫu lớn, vì trong các cụm các đơn vị được khảo sát có những tính chất, đặc điểm khá giống nhau về nơi cư trú, về học vấn, về nghề nghiệp, về thu nhập v.v. và từ đó là về ý kiến xung quanh chủ đề nghiên cứu. Do vậy, dù có khảo sát hàng trăm, hàng ngàn người trong một cụm thì thông tin thu

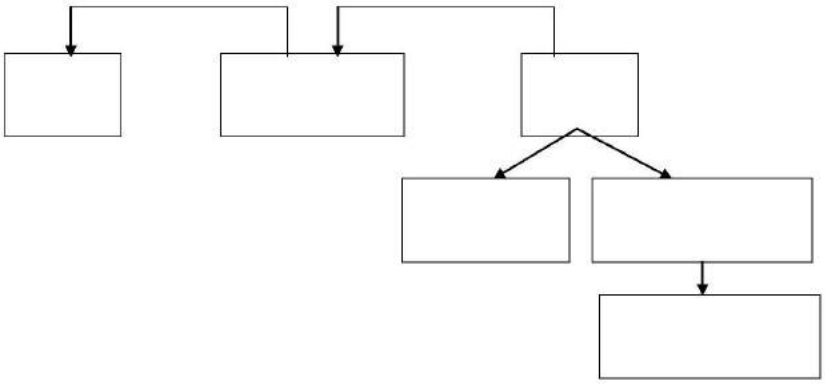
được từ đó vẫn có tính đại diện thấp cho toàn bộ tổng thể. Có thể giả quyết vấn đề này bằng cách chọn một mẫu có số lượng lớn các cụm và số lượng các đơn vị trong cụm đó nhỏ. Thí dụ thay vì chọn một khu công nghiệp (cụm lớn) và tiến hành nghiên cứu toàn bộ, tốt nhất chúng ta chọn nhiều khu công nghiệp (cụm nhỏ) ở nhiều tỉnh để khảo sát.

Khi chọn mẫu cụm, vấn đề quan trọng là lập được danh sách chọn mẫu (khung chọn mẫu). Vì tính đại diện của thông tin phụ thuộc rất lớn vào khung chọn mẫu này.

Khung chọn mẫu chính là danh sách, hoặc bảng liệt kê tất cả các đơn vị cấu thành tổng thể. Thí dụ danh sách các liên đoàn lao động, danh cách các ngành nghề v.v... Tuy nhiên cần lưu ý rằng các danh sách trên có thể được dùng làm khung chọn mẫu (khung mẫu) nếu như nó hội đủ những tính chất: tính đầy đủ, tính chính xác, tính thích hợp, tính đơn nhất (không trùng lặp), tính thuận tiện.

Nếu khung mẫu không đại diện thực sự cho tổng thể mà nó liệt kê, thì mẫu không thể là đại diện của tổng thể. Mẫu chỉ đại diện cho khung mẫu, cho nên trong quá trình thiết kế mẫu, chúng ta cần phải xem xét đến khả năng không phù hợp (không tương xứng) có thể có giữa khung mẫu và tổng thể.

Sơ đồ dưới đây là mối quan hệ giữa tổng thể - khung mẫu - mẫu và sai số khi khung mẫu không đại diện cho tổng thể:



Kết luận

Trong khi thu thập thông tin phục vụ cho đề tài nghiên cứu liên quan đến Công đoàn, người lao động, chúng ta hoàn toàn có thể chỉ tiến hành thu thập thông tin một số nhất định mẫu liên quan nhưng đủ cơ sở khoa học để khái quát cho toàn bộ. Tuy nhiên, để đảm bảo điều kiện cần và đủ khi khái quát cho tổng thể, nhà nghiên cứu cần xác định hai nhiệm vụ cơ bản đó là xác định dung lượng mẫu và cách chọn. Đối với việc xác định dung lượng mẫu (số lượng các đơn vị để tiến hành thu thập thông tin . Nhà nghiên cứu có thể vận dụng công thức để khi tính toán cho chúng ta một mẫu có dung lượng mẫu lớn nhất $n=[Nt^2x\ 0,25]/[Ne^2+ t^2x\ 0,25]$; Trong trường hợp không hoặc khó xác định được đơn vị của tổng thể chúng ta có thể dùng công thức $n=(t^2x\ 0,25)/^2$. Đối với cách chọn, chúng ta nên sử dụng cách chọn mẫu cụm nhiều giai đoạn. ❑

Tài liệu tham khảo

1. *Alan Agresti, Barbara Finlay. Statistical Methods for Social sciences. Prentice Hall, inc, 1997. PP.12-35*
2. *E. A. Capitonov. Xã hội học thế kỷ 20: Lịch sử và công nghệ. Biên dịch Nguyễn Quý Thanh. NXB ĐHQG Hà Nội 2000, 2002.*
3. *L. Therese Baker. Thực hành nghiên cứu xã hội. NXB CTQG,1998. Tr. 245-279. Tài liệu dịch.*
4. *Đào Hữu Hồ. Thống kê xã hội học. NXB ĐHQGHN, 1999.*
5. *Kim Ngọc Huynh chủ biên.* Bài giảng và bài tập thống kê. *NXB ĐH và GDCN,1992. Tr.269-315.*
6. *Tăng Văn Khiên. Bài giảng: Một số vấn đề điều tra chọn mẫu. NXB Thống kê, 1997.*
7. *James M. Lepkowski. Phương pháp chọn mẫu khu vực. Tài liệu tập huấn tại khoá đào tạo 1 tuần do viện Xã hội học tổ chức. Người dịch: Vũ Tuấn Huy.*
8. *Ô-xi-pốp chủ biên. Những cơ sở nghiên cứu xã hội học. NXB Tiến bộ Matxcova, 1988. Tr.258-299.*
9. *Nguyễn Quý Thanh,* Xã hội họcvề dư luận xã hội, *NXB ĐHQG Hà Nội.*
10. *Nguyễn Minh Thắng. Phương pháp chọn mẫu trong điều tra Xã hội học và dân số. NXB KHXH,1998.*
11. *Trần Hữu Quang,* Xã hội học Truyền thông đại chúng, *ĐH Mở TP HCM.*
12. *Thiết kế qui trình lấy mẫu. Tập thể tác giả.*
13. *Tăng Văn Khiên, Lý thuyết điều tra chọn mẫu, NXB Thống kê, Hà Nội - 2003.*
14. *Phạm Sơn, Một số vấn đề lý luận và thực tiễn áp dụng phương pháp điều tra chọn mẫu trong thực tiễn công tác thống kê nước ta, Thông tin Khoa học Thống kê số 3 - 1992.*
15. *Leslie Kish, Survey Sampling, P.H.John Weley, 1995.*
16. *Houston Dominique và Nguyễn Phong, Điều chỉnh cỡ mẫu điều tra mức sống dân cư Việt Nam, theo số liệu 2001-2002, Thông tin Khoa học Thống kê số 3-2003.*