

MỘT SỐ VẤN ĐỀ CHỦ YẾU CỦA MẪU LUÂN PHIÊN

Phạm Sơn - Viện KHTK

Mẫu luân phiên (nguyên bản tiếng Anh - Rotated Sample Survey) do một nhà thống kê người Anh đề xuất và đưa vào áp dụng trong điều tra các trang trại ở Anh trong thời kỳ thế chiến thứ II. Hiện nay, mẫu luân phiên được hầu hết các nước đưa vào ứng dụng trong công tác thống kê, nhất là các cuộc điều tra thường xuyên như điều tra thu thập chi tiêu của hộ gia đình, điều tra sinh tử trong lĩnh vực thống kê dân số... Với mục đích cung cấp tư liệu cho các nhà thống kê đang làm công tác thực tế tham khảo, bài báo này sẽ giới thiệu ngắn gọn một số vấn đề lý luận chủ yếu của mẫu luân phiên.

1. Khái niệm về mẫu luân phiên

Giả sử chúng ta cần nghiên cứu một hiện tượng kinh tế xã hội trong một quãng thời gian dài và xác định được cỡ mẫu theo các tham số chọn mẫu là W . Trong lần quan sát thứ nhất, chúng ta tiến hành thu thập thông tin ở n đơn vị được chọn. Lần quan sát tiếp theo chúng ta chỉ giữ lại n_1 mẫu ($n_1 < n$) và chọn bổ sung tiếp $n_2 = n - n_1$ từ tổng thể nghiên cứu để tiến hành thu thập thông tin. Mẫu điều tra được thiết kế như vậy gọi là mẫu luân phiên, còn $p = n_2/n_1$ gọi là tỷ lệ luân phiên.

Với cách xác định tỷ lệ luân phiên như trình bày trên đây thì P chỉ nhận giá trị từ 0 đến 1 trong trường hợp $p = 0$, người ta không gọi mẫu luân phiên, mà đó là mẫu cố định theo thời gian, còn trong trường hợp $p = 1$, người ta gọi đó là mẫu [M1] liên tiếp theo thời gian.

Ưu điểm của mẫu luân phiên là không cần tăng cỡ mẫu, nhưng mở rộng được phạm vi bao quát của mẫu.

Tuy nhiên để lợi dụng được ưu việt trên của mẫu luân phiên mà vẫn đảm bảo tính chất khoa học, đòi hỏi người thiết kế mẫu phải nắm vững những vấn đề cơ bản của mẫu luân phiên.

2. Mẫu luân phiên đơn một cấp

Mẫu luân phiên đơn một cấp là mẫu đơn không hoàn lại, trong đó theo từng chu kỳ thời gian có thể đổi mới (thay phiên) một số đơn vị. Việc thiết kế mẫu luân phiên một cấp được tiến hành như sau:

Bước 1: Lấy mẫu ngẫu nhiên không hoàn lại trên tổng thể nghiên cứu theo khối lượng mẫu xác định cho chu kỳ 1 là n trong tổng thể N đơn vị tổng thể. Khối lượng mẫu này không thay đổi trong suốt thời gian nghiên cứu và chu kỳ kế tiếp sẽ thay thế n_2 ($n_2 < n$) đơn vị trong tổng thể còn lại ($N_1 = N - n$).

Bước 2: Lấy mẫu ngẫu nhiên không hoàn lại n_2 ($n_2 < n$) đơn vị cần thay thế từ tổng thể mẫu (Có n đơn vị)

Bước 3: Lấy bổ sung n_2 đơn vị trong tổng thể $N_1 = (N - n)$ đơn vị

Ở đây có một vấn đề lý luận cần giải quyết. Đó là khi sử dụng mẫu luân phiên chỉ áp dụng duy nhất một phương pháp chọn, đó là chọn không hoàn lại do mấy lý do sau:

Trước hết, nếu áp dụng chọn có lặp lại sẽ xảy ra trường hợp 1 đơn vị có khả năng rơi vào mẫu nhiều lần, còn chọn hệ thống sau một vài lần luân phiên chúng ta sẽ không đảm bảo khoảng cách chọn và điều đó sẽ gây ra sai số chọn mẫu lớn.

Chính vì lẽ đó, các nhà chọn mẫu đã thống nhất trong mẫu luân phiên chỉ sử dụng duy nhất một phương pháp chọn không hoàn lại.

Tuy nhiên khi áp dụng qui trình thiết kế mẫu luân phiên để nghiên cứu các hiện tượng kinh tế xã hội cụ thể: Chẳng hạn như điều tra mức sống dân cư, cần lưu ý thêm.

a) Trường hợp các đơn vị được chọn lần trước, nhưng đến thời điểm này đã chuyển đi nơi khác. Ký hiệu số đơn vị này là n_3 . Cần xét 2 khía cạnh.

- Nếu $n_3 < n_2$ tức là số đơn vị mẫu bị mất ngẫu nhiên nhỏ hơn số đơn vị cần thay thế, thì xem n_3 đơn vị mất đó như những đơn vị mẫu cần thay thế và chúng ta chỉ chọn $n_4 = n_2 - n_3$ đơn vị trong tổng thể $N - n_2$ đơn vị để cho đủ n_2 đơn vị cần thay thế. Sau đó từ tổng thể $N - n_3$ chọn ngẫu nhiên không hoàn lại n_2 đơn vị để luân phiên.

- Nếu $n_3 > n_2$ tức là số đơn vị ngẫu nhiên ra khỏi mẫu nhiều hơn số cần luân phiên, cách xử lý như sau:

+ Chọn ngẫu nhiên trong tổng thể $N - n$ đơn vị $n_5 = n_3 - n_2$ và thu thập thông tin ở các đơn vị này theo thời kỳ trước.

+ Chọn n_2 đơn vị trong tổng thể $N - n - n_5$ đơn vị để luân phiên.

b) Nếu trong trường hợp tổng thể nghiên cứu có thêm các đơn vị mới chẳng hạn trong điều tra mức sống, có thể có hộ mới chuyển về sinh sống hoặc mới tách hộ ... phải bổ sung các hộ này vào tổng thể $(N - n)$ hộ sau đó mới tiến hành chọn n_2 đơn vị mới để luân phiên.

3. Mẫu luân phiên đơn nhiều cấp

Mẫu luân phiên nhiều cấp là mẫu nhiều cấp, những đơn vị mẫu của cấp cuối cùng được thiết kế theo mẫu luân phiên một cấp.

Chẳng hạn trong điều tra mức sống dân cư ở trên địa bàn một tỉnh theo mẫu 3 cấp.

Cấp 1: Chọn các huyện, thị xã, quận

Cấp 2: Chọn các địa bàn đại diện cho cho mẫu cấp 1

Cấp 3: Chọn các hộ đại diện từ mẫu cấp cấp 2

Như vậy, việc luân phiên chỉ xảy ra ở đơn vị mẫu cấp 3 và do đó các đơn vị mẫu cấp 1 và cấp 2 có thể áp dụng các phương pháp chọn khác nhau. Riêng đơn vị mẫu cấp 3 (hộ) phải tuân thủ nguyên tắc chọn mẫu luân phiên một cấp như trình bày trên đây.

4. Mẫu luân phiên chùm

Mẫu luân phiên chùm cũng được thiết kế tương tự như mẫu luân phiên đơn đã trình bày trên đây, chỉ khác ở chỗ là việc luân phiên xảy ra cả chùm đơn vị chứ không riêng lẻ từng đơn vị, có hai loại mẫu luân phiên chùm:

Mẫu luân phiên chùm một cấp tương tự như mẫu luân phiên đơn một cấp và mẫu luân phiên chùm nhiều cấp tương tự mẫu luân phiên đơn nhiều cấp.

5. Mẫu luân phiên so sánh được

Một trong các yêu cầu của nghiên cứu thống kê là số lượng giữa hai thời kỳ liên tiếp có thể so sánh với nhau. Để đáp ứng yêu cầu đó, các nhà điều tra chọn mẫu đã định ra nguyên tắc "quá bán tuyệt đối" như sau:

Muốn thông tin thu thập được ở hai kỳ liên tiếp có thể so sánh được thì tỷ lệ luân phiên với $p \leq 25\%$. Nguyên tắc này được lý giải như sau: Nếu chúng ta có số liệu

của một mẫu luân phiên với tỷ lệ luân phiên $p=25\%$. Như vậy, phần thông tin thu thập trên cùng một đơn vị mẫu là 75% , còn phần thông tin chỉ có một trong hai mẫu là 50% . Như vậy tỷ lệ thông tin ở những đơn vị so sánh được so với tổng số đơn vị có thông tin là: $75\%/125\% = 3/5$ và con số $3/5$ người ta gán cho thuật ngữ "quá bán tuyệt đối".

Đây là nguyên tắc hết sức quan trọng trong thực hành điều tra chọn mẫu, là giới hạn trên cho phép khi chọn tỷ lệ luân phiên.

6. Ước lượng số bình quân và tỷ lệ tăng bình quân khi sử dụng mẫu luân phiên

Có hai phương pháp ước lượng hai tham số trên đối với mẫu luân phiên.

6.1 Phương pháp ước lượng trực tiếp:

- Ước lượng số bình quân

Số bình quân của mẫu luân phiên chính là số bình quân số học giản đơn của các lần luân phiên.

Nếu ký hiệu : \bar{x}_1 là số bình quân của đại lượng nghiên cứu trong lần quan sát thứ nhất.

\bar{x}_2 là số bình quân của đại lượng nghiên cứu trong lần quan sát thứ hai

\bar{x}_k lần quan sát k

$$\text{Lúc đó: } \bar{x} = \frac{1}{k} \sum_{i=1}^k \bar{x}_i \quad (1)$$

là số bình quân chung của mẫu luân phiên

- Ước lượng tỷ lệ tăng bình quân

gọi $t_1 = \frac{\bar{x}_2}{\bar{x}_1}$ là tỷ lệ tăng của chu kỳ

2 so với chu kỳ 1

$$t_2 = \frac{\bar{x}_3}{\bar{x}_2} \quad \text{của kỳ 3 so với kỳ 2}$$

.....

$$t_k = \frac{\bar{x}_{k+1}}{\bar{x}_k} \quad \text{của kỳ (k + 1) so với kỳ k}$$

$$\text{Lúc đó } \bar{t} = \sqrt[k]{\prod_{i=1}^k t_i} \quad (2)$$

là tỷ lệ tăng bình quân của đại lượng (chỉ tiêu) nghiên cứu trong mẫu luân phiên.

6.2 Phương pháp ước lượng có điều chỉnh tỷ lệ tăng

Việc ước lượng có điều chỉnh tỷ lệ tăng bình quân từ số lượng mẫu luân phiên khá phức tạp và qua một số bước sau:

Bước 1: Tính các số bình quân chung, bình quân 2 kỳ của các đơn vị mẫu không luân phiên và bình quân của các đơn vị loại ra và các đơn vị mới C luân phiên vào theo công thức sau:

Bình quân chung của thời kỳ 1:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (3)$$

Bình quân chung của thời kỳ 2:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad (4)$$

Bình quân của các đơn vị bị loại ra sau chu kỳ 1

$$\bar{x}_1 = \frac{1}{n_2} \sum_{j=1}^{n_2} x_j' \quad (5)$$

x_j' là giá trị của chỉ tiêu nghiên cứu ở chu kỳ 1 nhưng bị loại ở chu kỳ 2

$$\bar{x}_2 = \frac{1}{n_1} \sum_{j=1}^{n_1} x_j'' \quad (6)$$

x_j'' là giá trị của chỉ tiêu nghiên cứu ở chu kỳ 1 và các đơn vị này tiếp tục được quan sát ở chu kỳ 2

Tương tự như vậy chúng ta tính được y_1 và y_2 (Số liệu ở chu kỳ 2)

Bước 2: Nội suy và ngoại suy số liệu của các đơn vị chỉ mới có số liệu ở 1 trong 2 thời kỳ

Ngoại suy số liệu của các đơn vị có ở thời kỳ 1 nhưng không có ở thời kỳ 2 theo công thức sau:

$$\bar{x}_1' = \frac{\bar{x}_2}{\bar{y}_2} \bar{x}_1 \quad (7)$$

Nội suy số liệu của các đơn vị mới bổ sung ở chu kỳ 2 nhưng chưa có ở chu kỳ 1.

$$\bar{y}_1' = \frac{\bar{y}_2}{\bar{x}_2} \bar{y}_1 \quad (8)$$

Bước 3: Tính các số bình quân cho từng thời kỳ theo số liệu mới bổ sung:

$$\bar{x}' = \frac{n_2 \cdot \bar{x}_1 + n \cdot \bar{x}}{n_2 + n} \quad (\text{Thời kỳ 1}) \quad (9)$$

$$\bar{y}' = \frac{n_2 \cdot \bar{y}_1 + n \cdot \bar{y}}{n_2 + n} \quad (\text{Thời kỳ 2}) \quad (10)$$

Bước 4: Tính tỷ lệ tăng có điều chỉnh giữa hai thời kỳ

$$t = \frac{\bar{y}'}{\bar{x}'}$$

Muốn tính tỷ lệ tăng bình quân có điều chỉnh, có thể áp dụng công thức như khi ước lượng tỷ lệ tăng bình quân chưa điều chỉnh (Xem công thức 2)

Trên đây là một số vấn đề lý luận hết sức cơ bản và đơn giản của mẫu luân phiên. Còn những vấn đề lý luận phức tạp hơn sẽ có dịp giới thiệu với bạn đọc trong những số tiếp theo ■