

PHƯƠNG PHÁP ĐÁNH GIÁ ĐỘ CHÍNH XÁC CỦA PHÂN BỐ TUỔI

VŨ TUẤN HUY

Trong nghiên cứu dân số cũng như nghiên cứu xã hội học thực nghiệm, việc thu nhập, xử lý thông tin và kết quả được đưa ra dưới dạng tương quan. Đó là những tương quan giữa giới tính và tuổi với các đặc trưng khác như số con, trình độ văn hóa, mức thu nhập, tình cơ động xã hội v.v... Động thái của những mặt, những hiện tượng của đời sống xã hội được biểu hiện qua những chỉ số này. Tuy nhiên trên bình diện rộng, chẳng hạn một bước, chỉ riêng với số liệu về tuổi và giới tính, bằng phương pháp mô hình hóa, chúng ta có thể đánh giá một cách khái quát mức độ phát triển kinh tế- xã hội là: phát triển hoặc đang phát triển. Điều này bắt nguồn từ một thực tế là cơ cấu kinh tế của một vùng, một nước là kết quả tổng hợp của những yếu tố văn hóa, xã hội, nguồn tài nguyên, cơ cấu dân số v.v... Trong đó tỷ lệ cao của dân cư ở độ tuổi lao động (45-60) và đặc biệt là tỷ trọng phụ nữ tham gia lao động xã hội ngày càng tăng là yếu tố thúc đẩy sự tăng trưởng kinh tế. Vì vậy cơ cấu về tuổi và giới tính của dân cư là những chỉ báo nhân khẩu xã hội quan trọng, là cơ sở dữ liệu ban đầu cho việc xác lập độ lớn và cơ cấu mẫu điều tra trong nghiên cứu xã hội học. Thông thường nguồn số liệu về tuổi và giới tính dựa vào các cuộc điều tra dân số và các số liệu tke. Hệ thống thu thập số liệu này của chúng ta hiện nay còn nhiều nhược điểm, thiếu tính hệ thống, những sai sót về mặt kỹ thuật, phân loại, tính toán. Vì vậy những số liệu này cần được đánh giá và hiệu đính trước khi sử dụng. Ở đây chúng tôi chỉ đề cập đến cách đánh giá và hiệu đính trước khi sử dụng. Ở đây chúng tôi chỉ đề cập đến cách đánh giá độ chính xác về tuổi mà những nguyên nhân dẫn đến sai sót đó thuộc về người khai báo.

Số liệu về tuổi thiếu chính xác do nhiều nguyên nhân khác nhau. Chẳng hạn, phụ nữ thường khai tuổi thấp hơn, còn người già khai tăng tuổi lên so với tuổi thực của họ. Cư dân ở vùng nông thôn có thói quen tính tuổi theo Âm lịch, vì vậy thường tăng lên 1 tuổi so với tuổi thực tế. Một nguyên nhân khác phổ biến hơn, cư dân thường làm tròn tuổi, với con số tận cùng là “0” hoặc “5”. Ví dụ: 15, 20, 25, 30,...

Hiện tượng này trong dân số học gọi là sự xếp chồng tuổi (heaping on age) được đánh giá bằng những phương pháp toán học khác nhau. Những phương pháp này căn cứ vào dạng phân bố dân số theo tuổi qua một phần hoặc toàn bộ của miền tuổi. Dạng phân bố này được giả định là đúng và dùng nó để so sánh với phân bố tuổi cần kiểm tra.

Dạng phân bố đơn giản và hiệu quả nhất là phân bố tuổi hình chữ (nghĩa là phân bố mà trong đó số lượng người mới trong mỗi tuổi được coi là bằng nhau). Phân bố này xác định trong miền tuổi, ví dụ: 3 năm, 5 năm hoặc 11 năm. Căn cứ vào giả định này, chúng ta có thể đánh giá mức độ chính xác của số liệu thu được về tuổi qua hai chỉ số: chỉ số làm tròn tuổi và chỉ số Whipple

Chỉ số làm tròn tuổi: hiện trạng làm tròn ở một tuổi nào đó dẫn đến việc giảm tương ứng ở những độ tuổi lân cận. Căn cứ vào giả định đã nêu trên, người ta đã đưa ra chỉ số làm tròn tuổi có dạng sau:

- Với miền xác định tuổi 3 năm

$$I = \frac{P_x}{1/3 (P_{x-1} + P_x + P_{x+1})} \cdot 100$$

- Với miền xác định tuổi 5 năm

$$I = \frac{P_x}{1/5 (P_{x-2} + P_{x-1} + P_x + P_{x+1} + P_{x+2})} \cdot 100$$

Trong đó: I là chỉ số làm tròn tuổi

P_x là số người trong độ tuổi x mà ta cần đánh giá xem có hiện tượng chồng tuổi hay không?

Giá trị nhỏ nhất chỉ số $I = 100$, điều đó chứng tỏ không có hiện tượng xếp chồng ở độ tuổi x. Ngược lại khi giá trị của I càng lớn hơn 100 thì số người tập trung vào độ tuổi x càng lớn.

Chỉ số Whipple: chỉ số này không dùng để đánh giá một độ tuổi cụ thể nào mà vào một nhóm tuổi có số tận cùng đặc biệt “0” và “5”.

Sử dụng giả định phân bố tuổi hình chữ nhật ở trên chỉ số Whipple dùng để đánh giá sự xếp chồng tuổi ở tận cùng là “0”, trong miền tuổi từ 23 đến 62 được tính theo công thức sau:

$$I = \frac{P_{30} + P_{35} + P_{40} + P_{45}}{1/10 (P_{23} + P_{24} + P_{25} + \dots + P_{60} + P_{61} + P_{62})} \cdot 100$$

Tương tự như vậy, trong miền tuổi từ 23 đến 62 chỉ số Whipple dùng để đánh giá sự xếp chồng tuổi ở tận cùng là số “5”, được tính như sau:

$$I = \frac{P_{25} + P_{30} + P_{35} + \dots + P_{55} + P_{60}}{1/5 (P_{23} + P_{24} + P_{25} + \dots + P_{60} + P_{61} + P_{62})} \cdot 100$$

Miền giá trị của I từ 100 đến 500. Nếu $I = 100$, chứng tỏ rằng không có hiện tượng xếp chồng tuổi ở các độ tuổi có số tận cùng là “0” và “5”. Ngược lại, khi giá trị của I càng gần 500 chỉ ra có hiện tượng xếp chồng tuổi. Trên đây là hai chỉ số đơn giản và hiệu quả mà nhờ đó ta có thể đánh giá mức độ chính xác phân bố cơ cấu tuổi của dân cư. Nó có ý nghĩa thực tế về mặt phương pháp trong nghiên cứu xã hội hoặc là căn cứ để hiệu đính số liệu phân bố tuổi được sử dụng khi xác định mẫu điều tra cũng như khi phân nhóm tương quan. Để giúp bạn đọc hiểu được ý nghĩa của hai chỉ số trên, chúng tôi đưa ra một ví dụ sau:

Giả sử chúng ta có số liệu về phân bố tuổi của dân cư ở vùng như sau: (xem bảng)

Đây là dân số đóng không bị ảnh hưởng của yếu tố di cư, và không bị ảnh hưởng đột biến của yếu tố tử vong.

Áp dụng chỉ số làm tròn tuổi, chúng ta kiểm tra xem có hiện tượng chồng tuổi ở tuổi 35 hay không. Trong miền tuổi 3 năm chỉ số này là:

$$I = \frac{402}{1/3 (233 + 402 + 243)} \cdot 100 = 137$$

Tuổi	Số lượng	Tuổi	Số lượng
19	492	41	127
20	566	42	218
21	485	43	169
22	516	44	152
23	457	45	319
24	425	46	160
25	522	47	160
26	358	48	237
27	377	49	155
28	386	50	313
29	300	51	78
30	536	52	128
31	221	53	93
32	318	54	95
33	246	55	164
34	233	56	88
35	402	57	72
36	243	58	93
37	424	59	72
38	316	60	267
39	225	61	32
40	434	62	50

Kết quả chỉ ra có sự tăng đáng kể số người trong độ tuổi 35 và giảm tương ứng ở các độ tuổi lân cận.

Sử dụng chỉ số Whipple, chúng ta kiểm tra tất cả các tuổi với số tận cùng là “0” và “5”. Ta có:

$$I = \frac{522 + 536 + 402 + 434 + 319 + 313 + 164 + 267}{1/5 (457 + 425 + \dots + 32 + 50)} \cdot 100$$

$$I = \frac{2957}{1802} \cdot 100 = 164$$

Kết quả khi áp dụng chỉ số này chỉ ra có hiện tượng chồng tuổi ở tất cả các độ tuổi với số tận cùng là các số “0” và “5”. Như vậy, số liệu phân bố tuổi này không chính xác có thể do người khai báo đã làm tròn tuổi.

Tóm lại, tuổi là một trong những dấu hiệu mà cuộc điều tra xã hội học cần đến. Nó không những cần thiết để đánh giá cơ cấu nhóm người được điều tra theo độ tuổi, mà còn là một dấu hiệu phân nhóm khi tính tương quan. Song vì những lí do nào đó mà số liệu về tuổi thường mắc sai số. Vì vậy chúng ta cần đánh giá nó trước khi sử dụng. Hai chỉ số được nêu ra ở trên giúp chúng ta kiểm tra được những sai số có thể xảy ra về độ tuổi.