

PHƯƠNG PHÁP ĐÁNH GIÁ CHẤT LƯỢNG KHAI BÁO CHỈ TIÊU ĐỘ TUỔI TRONG THỐNG KÊ DÂN SỐ

Nguyễn Văn Phái

Cơ cấu dân số theo giới tính và độ tuổi là một trong những đặc trưng quan trọng nhất của dân số. Các số liệu dân số theo độ tuổi được sử dụng rộng rãi trong công tác kế hoạch hoá, xây dựng chính sách và trong phân tích nhân khẩu học. Rất nhiều chỉ tiêu trong thống kê dân số, nhất là trong tổng điều tra chỉ có ý nghĩa đáng kể nếu chúng được quan sát cùng với độ tuổi. Bởi vậy, việc đánh giá chất lượng khai báo chỉ tiêu tuổi trong thống kê dân số có ý nghĩa rất quan trọng.

Một trong những sai sót thường xảy ra khi khai báo chỉ tiêu độ tuổi trong thống kê dân số là hiện tượng “tập trung tuổi”. Trong thực tế, thường có sự khai báo tập trung vào một số độ tuổi ưa thích nào đó, thường là các độ tuổi (hoặc năm sinh) có số tận cùng là 0 và 5. Kết quả là, có sự tập trung bất bình thường về số dân vào những độ tuổi “ưa thích” này. Có nhiều phương pháp xác định mức độ tập trung tuổi nhưng phổ biến nhất là chỉ số Whipple và chỉ số Myer.

Chỉ số Whipple

Chỉ số Whipple được sử dụng để xác định mức độ tập trung tuổi vào các độ tuổi có tận cùng là 0 và 5 (hiện tượng làm tròn

$$k_{wp} = \frac{P_{25} + P_{30} + P_{35} + \dots + P_{55} + P_{60}}{P_{23} + P_{24} + P_{25} + P_{26} + \dots + P_{61} + P_{62}} \times 100$$

tuổi). Chỉ số này được biểu thị bằng tỷ lệ phần trăm của tổng số dân từ 23 đến 62 tuổi có tuổi tận cùng là 0 và 5 và một phần năm tổng số dân trong khoảng tuổi này. Chỉ số Whipple được tính theo công thức:

Giá trị của chỉ số Whipple càng gần với 100, mức độ tập trung khi khai báo vào các tuổi có tận cùng là 0 và 5 càng thấp và ngược lại, nếu giá trị này càng xa so với 100, sự khai báo tập trung vào những độ tuổi này càng nhiều.

Ví dụ: Theo số liệu của cuộc Tổng điều tra dân số Việt Nam 1999, phân bố dân số theo từng độ tuổi của dân số cả nước như sau (xem Biểu 1):

Theo số liệu của Biểu 1, tổng số dân số các nhóm tuổi có tận cùng là 0 và 5 trong khoảng tuổi từ 23 đến 62 là:

$$\begin{aligned} &= 126295 + 1301966 + 1203805 + 1024334 \\ &+ 680372 + 495385 + 317047 + 341422 \\ &= 6657342 \end{aligned}$$

$$\begin{aligned} &\text{Tổng số dân từ 23 đến 62 tuổi là:} \\ &= 1381830 + 1316510 + 1292651 + 1368806 + \\ &\dots + 341422 + 382345 + 359088 \\ &= 33548336 \end{aligned}$$

Như vậy, chỉ số Whipple của dân số Việt Nam trong cuộc Tổng điều tra dân số và nhà ở năm 1999 bằng:

$$k_{wp} = 5 \cdot 6657342 / 33548336 \cdot 100 = 99,2$$

Chỉ số Whipple tính được là 99,2 rất gần với 100 chứng tỏ rằng, không có sự khai báo tập trung vào các độ tuổi có tận cùng là 0 và 5 trong cuộc Tổng điều tra dân số và nhà ở Việt Nam năm 1999.

Biểu 1: Tính chỉ số Whipple cho dân số Việt Nam năm 1999, Chung hai giới

Tuổi	Số dân	Tuổi	Số dân	Tuổi	Số dân	Tuổi	Số dân
0	1263599	20	1448504	40	1024334	60	341422
1	1316569	21	1371183	41	993318	61	382345
2	1377357	22	1407360	42	869570	62	359088
3	1555036	23	1381830	43	817613	63	344292
4	1659681	24	1316510	44	845225	64	320161
5	1682306	25	1292651	45	680732	65	328315
6	1757852	26	1368806	46	701489	66	337390
7	1762242	27	1294503	47	531105	67	286139
8	1955994	28	1345956	48	645480	68	373566
9	1874768	29	1266258	49	578452	69	321365
10	1776096	30	1301966	50	495385	70	373205
11	1841307	31	1133512	51	472643	71	262276
12	1771464	32	1170555	52	389460	72	233441
13	1800719	33	1188407	53	387360	73	233261
14	1876976	34	1239266	54	359468	74	208921
15	1834600	35	1203805	55	317047	75+	1531212
16	1800168	36	1182822	56	355893		
17	1569297	37	960075	57	323041	Cộng	76323173
18	1523543	38	1174125	58	416025		
19	1494672	39	1065793	59	375001		

Nguồn: Tổng điều tra dân số và nhà ở Việt Nam 1999: Kết quả điều tra toàn bộ.- Hà Nội, Thống kê, 2001.- Trang 17.

Ghi Chú: Dân số 75+ được cộng dân số theo các độ tuổi từ 75 trở lên.

Chỉ số Myer

Một chỉ số khác thường được dùng để đánh giá mức độ chính xác trong khai báo tuổi là chỉ số Myer. Chỉ số này cho biết, khi khai báo tuổi, dân số có xu hướng thích khai báo vào những tuổi tận cùng là con số nào, còn những độ tuổi nào không được ưa thích. Thực chất, đây là tổng số dân số có tuổi tận cùng là một con số giống nhau trong một khoảng tuổi nào đó, thường là từ 10 đến 89, được biểu thị bằng phần trăm so với tổng số dân trong khoảng đó. Để tránh những sai sót tạo ra bởi một thực tế là dân số phân bố không đồng đều trong tất cả các lứa tuổi, cần lặp lại việc tính toán 10 lần, mỗi lần cho một con số và lấy kết quả trung bình. Chênh lệch giữa tỷ lệ phần trăm trung bình của mỗi con số và giá trị lẽ ra là 10% chính là thước đo mức độ ưa thích hoặc không ưa thích khi khai báo tuổi.

Chỉ số Myer (chỉ số chung) dao động từ 0 đến 90. Chỉ số này càng nhỏ, mức độ

chính xác trong khai báo tuổi càng cao và ngược lại.

Quy trình tính chỉ số Myer như sau:

1/ Tính dân số từ 10 tuổi trở lên có tuổi kết thúc cùng bằng một con số;

2/ Tính dân số từ 20 tuổi trở lên có tuổi kết thúc cùng bằng một con số;

3/ Tính dân số pha trộn;

4/ Tính tỷ lệ phần trăm của dân số pha trộn;

5/ Tính giá trị chênh lệch giữa tỷ lệ phần trăm của dân số pha trộn với 10;

6/ Tính tổng các giá trị tuyệt đối các giá trị chênh lệch giữa tỷ lệ phần trăm của dân số pha trộn với 10.

Biểu 2 trình bày chi tiết ví dụ tính chỉ số Myer cho dân số chia theo từng độ tuổi thu được trong cuộc Tổng điều tra dân số và nhà ở 1999 (chung cả hai giới).

Biểu 2: Tính chỉ số Myer cho dân số Việt Nam năm 1999, Chung hai giới

Tuổi có số tận cùng	Tổng dân số các tuổi có cùng số cuối cùng từ 10+	Tổng dân số các tuổi có cùng số cuối cùng từ 10+	Quyền số		Dân số pha trộn (cột2*cột4+cột3*cột5)	Phân bố phần trăm	Chênh lệch với 10
			Cho cột 2	Cho cột 3			
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
0	6783302	5007206	1	9	51848156	10,0	0,0
1	6565028	4723721	2	8	50919824	9,8	-0,2
2	6292131	4520667	3	7	50521062	9,7	-0,3
3	6237036	4436317	4	6	51566046	9,9	-0,1
4	6238372	4361396	5	5	52998840	10,2	0,2
5	5902274	4067674	6	4	51684340	9,9	-0,1
6	5978926	4178758	7	3	54388756	10,5	0,5
7	5162615	3593318	8	2	48487556	9,3	-0,7

8	5687386	4163843	9	1	55350317	10,6	0,6
9	5267000	3772328	10	0	52670000	10,1	0,1
Cộng					520434897	100,0	2,8

Nguồn: Tổng cục Thống kê, Tổng điều tra dân số và nhà ở Việt Nam 1999: Kết quả điều tra toàn bộ.- Hà Nội: Thống kê, 2001.- Trang 17.

Các kết quả tính toán được từ Biểu 2 cho thấy, mức độ chính xác trong khai báo tuổi trong cuộc Tổng điều tra dân số và nhà ở năm 1999 của Việt Nam rất cao (2,8). Sự tập trung vào các tuổi ưa thích không đáng kể. Những tuổi có tận cùng là 6 và 8 nhiều hơn bình thường còn những độ tuổi có tận cùng là 7 lại ít hơn bình thường. Lý do chủ yếu dẫn đến hiện tượng

có nhiều người khai báo có độ tuổi tận cùng là 8 là do thời điểm của cuộc Tổng điều tra là 1 tháng 4 năm 1999 và câu hỏi về tuổi được hỏi và ghi theo tháng năm sinh. Thực tế dân số có xu hướng khai báo các năm sinh vào những năm có số tận cùng là 0 (1930, 1940, 1950,...) và những người này, nếu sinh từ tháng 4 đến hết tháng 12 khi tính tuổi sẽ có số tận cùng là 8.