

# ỨNG DỤNG HỒI QUI LOGISTIC NGHIÊN CỨU TÌNH TRẠNG BIẾT ĐỌC BIẾT VIẾT CỦA DÂN SỐ VIỆT NAM

Phan Ngọc Trâm

Mô hình hồi qui logistic là một công cụ mạnh được ứng dụng nhiều trong công tác thống kê các nước. Để giúp độc giả hiểu rõ hơn ưu điểm của mô hình, bài báo này sẽ giới thiệu kết quả nghiên cứu ứng dụng mô hình hồi qui logistic trong nghiên cứu tình trạng biết đọc biết viết của dân số Việt Nam theo số liệu của Tổng điều tra dân số và nhà ở năm 1999.

## 1. Phương pháp luận

### 1.1. Định nghĩa biết đọc biết viết

Theo Tổng điều tra dân số và nhà ở Việt Nam, một người được gọi là biết đọc biết viết nếu người ấy “biết đọc, viết và hiểu những câu đơn giản bằng ngôn ngữ quốc gia hay ngôn ngữ dân tộc hoặc tiếng nước ngoài”.

Định nghĩa này có khác với định nghĩa sử dụng trong các nghiên cứu khác được tiến hành trên cơ sở các báo cáo hành chính của Bộ Giáo dục và Đào tạo.

### 1.2. Phương pháp luận

Thông tin dùng trong bài viết này về dân số Việt Nam từ 15 tuổi trở lên, gồm các chỉ tiêu được ước lượng từ số liệu điều tra mẫu 3% sau khi đã quyền số hoá. Việc đánh giá các yếu tố tác động đến tình trạng biết đọc biết viết được thực hiện bằng cách sử dụng hồi qui logistic.

Trong khi phân tích về tình trạng biết đọc biết viết, chúng tôi có phân tổ chi tiết cho bốn yếu tố là giới tính, nơi cư trú, tuổi và dân tộc cho tất cả 61 tỉnh thành phố. Trong

đó: giới tính: nam và nữ; nơi cư trú: thành thị, nông thôn; tuổi: từ 15 đến 35 (trẻ), và từ 35 tuổi trở lên (già)<sup>(1)</sup>; dân tộc: dân tộc Kinh và các dân tộc khác (thiểu số)<sup>(2)</sup>.

### Mô hình hồi qui logistic

Mô hình này có biến phụ thuộc là biết đọc biết viết ( $y$ ) biết đọc, biết viết = 1, không = 0), và các biến giải thích là: giới tính ( $x_1$ ) Nam = 1, Nữ = 0), nơi cư trú ( $x_2$ ) thành thị = 1, Nông thôn = 0, nhóm tuổi ( $x_3$ ) nhóm tuổi 15-34 (trẻ) = 1, nhóm tuổi 35 + (già) = 0 và dân tộc ( $x_4$ ) dân tộc Kinh = 1, dân tộc khác = 0). Hàm logistic trong trường hợp này có dạng cụ thể như sau:

$$P(y) = \frac{e^{a + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4}}{1 + e^{a + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4}}$$

$P(y)$  là xác suất để biến  $y$  nhận giá trị là 1.

Kết quả tính toán được dựa trên phần mềm SPSS và Extat. Sau khi tính toán thu được các hệ số của mô hình cùng giá trị kiểm định P tương ứng. Trường hợp hệ số được ước lượng có giá trị P tương ứng thấp hơn một mức  $\alpha$  nào đó thì giá trị thực của hệ số đó có xác suất cao là không bằng 0 (thường chọn  $\alpha = 0,05$ , ở báo cáo này do qui mô số liệu đủ lớn nên chúng tôi chọn  $\alpha = 0,01$ ). Trường hợp ngược lại, khi hệ số tương ứng với giá trị P của nó lớn hơn 0,01 thì không thể bác bỏ giả thuyết hệ số thực là bằng 0 và người ta nói hệ số của biến đó là không có ý nghĩa thống kê; trong một số trường hợp người ta sẽ đưa biến đó ra khỏi mô hình.

## 2. Kết quả nghiên cứu

Theo tính toán từ số liệu TĐTDS 1999, thì tỷ lệ biết đọc biết viết của dân số 15 tuổi trở lên của Việt Nam là 90,3%. Phân tích chi tiết theo bốn yếu tố là giới tính, nơi cư trú, nhóm tuổi và dân tộc: nam có tỷ lệ biết đọc biết viết là 94,0% trong khi đó tỷ lệ này ở nữ chỉ là 86,9%. Dân số nông thôn có tỷ lệ này là 88,7% thấp hơn tỷ lệ 94,8% của thành thị. Dân tộc Kinh có tỷ lệ 92,8% biết đọc biết viết cao hơn rất nhiều so với dân tộc thiểu số 72,2%. Tỷ lệ biết đọc biết viết của dân số trẻ là 94,1% cao hơn tỷ lệ này của dân số già (85,8%).

### 2.1. Sự khác nhau giữa các tỉnh

Giữa các tỉnh có khác nhau đáng kể về tỷ lệ biết đọc biết viết. Nếu lấy mục tiêu quốc gia là tỷ lệ biết đọc biết viết phải đạt hoặc vượt mức 94%, thì thấy không một tỉnh nào trong ba vùng Tây bắc, Tây nguyên và Đồng bằng sông Cửu long đạt được mục tiêu quốc gia. Ở ba vùng khác là Bắc Trung Bộ, Duyên hải Miền Trung và Đông Nam Bộ mỗi vùng cũng chỉ có một tỉnh đạt mục tiêu quốc gia.

Ở miền Bắc, Đồng bằng sông Hồng và vùng Đông Bắc đạt tỷ lệ biết đọc biết viết tương đối cao mặc dầu ở hai vùng này vẫn còn có một số tỉnh có tỷ lệ biết đọc biết viết thấp hơn 94%.

### 2.2. Các yếu tố ảnh hưởng đến tình trạng biết đọc biết viết của từng tỉnh

Để khảo sát xem các yếu tố nào tác động đến tình trạng biết đọc biết viết của từng tỉnh, chúng tôi đã sử dụng mô hình logistic với bốn yếu tố kể trên. Mô hình hồi qui cho biết, liệu trên thực tế các yếu tố đó có tác động đến tình trạng biết đọc biết viết

hay không căn cứ vào giá trị P tương ứng. Nhìn chung các hệ số trong 61 mô hình biết đọc biết viết của 61 tỉnh phần lớn có giá trị P nhỏ hơn mức 0,01 trừ một số tỉnh có giá trị P tương ứng với biến dân tộc có giá trị lớn hơn nhiều so với mức 0,01.

Kết quả sử dụng mô hình logistic cho thấy:

a. Về yếu tố giới tính: sự khác biệt giữa nam và nữ về tỷ lệ biết đọc biết viết có ý nghĩa thống kê ở tất cả các tỉnh: nam có tỷ lệ biết đọc biết viết cao hơn nữ. Tuy nhiên, sự khác biệt này ở các tỉnh phía Nam thấp.

Đối với nam giới, tỷ lệ biết đọc biết viết thấp nhất là ở Lai Châu chỉ đạt 63,1%. Tỷ lệ biết đọc biết viết cao nhất là ở Hà Nội đạt mức 99,1%; có 33 tỉnh có tỷ lệ này của nam giới cao hơn 94%.

Đối với nữ giới, tỷ lệ cao nhất là ở Hà Nội đạt 94,9%, và đây là tỉnh duy nhất có tỷ lệ này cao hơn 94%. Những vùng cả nam và nữ đều có tỷ lệ đặc biệt thấp là Tây Bắc và Tây nguyên.

b. Về yếu tố nơi cư trú khu vực thành thị có tỷ lệ biết đọc biết viết cao hơn nông thôn, sự khác biệt này có ý nghĩa thống kê ở tất cả các tỉnh, trừ hai tỉnh Bình Thuận và Cà Mau. Sự khác biệt giữa thành thị và nông thôn cũng là nhỏ nhất ở các tỉnh khu vực phía Nam. Sự khác biệt này ở hai tỉnh Cao Bằng và Bắc Kạn thuộc loại lớn nhất.

Ở thành thị, tỷ lệ biết đọc biết viết dao động từ 89,3% (An Giang) đến 98,4% (Hà Nội). Ở khu vực nông thôn, tỷ lệ biết đọc biết viết dao động từ 44% (Lai Châu) đến 94,7% (Hà Nội). Vùng có dân số nông thôn có tỷ lệ biết đọc biết viết đặc biệt thấp vẫn là vùng Tây bắc và Tây nguyên.

c. Về yếu tố nhóm tuổi: giữa nhóm tuổi trên 35 tuổi và dưới 35 có sự khác biệt có ý nghĩa thống kê ở tất cả các tỉnh. Sự khác biệt này lớn nhất ở các tỉnh vùng Đồng bằng sông Hồng. Điều này cho thấy sự tiến bộ theo thế hệ về tình trạng biết đọc biết viết. Một số tỉnh thuộc vùng Tây bắc, vùng Đông bắc và một số tỉnh thuộc khu vực phía Nam sự khác biệt giữa thế hệ trẻ và thế hệ già không nhiều.

Tỷ lệ biết đọc của nhóm dân số trẻ dao động từ 52,8% đến 99,5%; có 38 tỉnh có tỷ lệ này của dân số trẻ cao hơn 94%. Tuy nhiên có một số tỉnh thuộc vùng Đông bắc và 2 tỉnh vùng Tây bắc, 2 tỉnh vùng Tây nguyên có tỷ lệ này của dân số trẻ thấp hơn 80%. Lai Châu có tỷ lệ này thấp nhất là 52,3%.

d. Về yếu tố dân tộc: có 16 tỉnh sự khác biệt giữa dân tộc Kinh và dân tộc thiểu số là không có ý nghĩa thống kê, điều này thể hiện ở giá trị P tương ứng với yếu tố dân tộc lớn hơn rất nhiều so với mức  $\alpha = 0,01$ , cho thấy đối với các tỉnh này, yếu tố dân tộc không có tác động đến tình trạng biết đọc biết viết. Đó là các tỉnh, thành phố: Hà Nội, Hải Phòng, Hải Dương, Hưng Yên, Hà Nam, Nam Định, Thái Bình, Vĩnh Phúc, Bắc Ninh, Hà Tĩnh, Đà Nẵng, Bình Định, Đồng Tháp, Tiền Giang, Bến Tre. Đây cũng là các tỉnh có tỷ lệ đồng bào dân tộc thiểu số rất nhỏ, thấp hơn 1% dân số (trừ tỉnh Vĩnh phúc - tỷ lệ dân tộc thiểu số là 2,2%). Ở đây chúng tôi tạm qui ước đó là các tỉnh thuộc nhóm 1 là các tỉnh có tình trạng biết đọc biết viết không bị tác động bởi yếu tố dân tộc, còn nhóm 2 là các tỉnh còn lại.

Ở các tỉnh thuộc nhóm 2, sự khác biệt giữa dân tộc Kinh và dân tộc thiểu số là lớn nhất ở các tỉnh thuộc vùng Tây Bắc, Tây Nguyên và một số tỉnh thuộc vùng Đông

Bắc. Ở các tỉnh này, dân tộc thiểu số có tỷ lệ biết đọc biết viết thấp hơn rất nhiều so với dân tộc Kinh. Trong khi nhóm dân tộc Kinh tỷ lệ biết đọc biết viết dao động từ 85,7% đến 97,4%, thì nhóm dân tộc thiểu số tỷ lệ biết đọc biết viết dao động từ 36,1% đến 96,9%.

Đối với tất cả 16 tỉnh thuộc nhóm 1, yếu tố nhóm tuổi có tác động lớn nhất đối với tình trạng biết đọc biết viết, kế đó là yếu tố giới tính, sau cùng là yếu tố nơi cư trú.

Đối với các tỉnh thuộc nhóm 2 (45 tỉnh), trong phần lớn các tỉnh (32/45 tỉnh) yếu tố dân tộc có tác động lớn nhất đến tình trạng biết đọc biết viết, sau đó là yếu tố nơi cư trú. Đối với các tỉnh phía Nam, sau yếu tố dân tộc yếu tố mạnh thứ hai là nhóm tuổi.

e. Nghiên cứu riêng về nhóm dân số dưới 35 tuổi

Vào những năm 1990, phong trào xoá nạn mù chữ đặt mục tiêu vào nhóm dân số dưới 35 tuổi, nên ở đây chúng tôi nghiên cứu chi tiết hơn cho nhóm tuổi này. Như trên đã nêu tỷ lệ biết đọc biết viết của nhóm dân số trẻ đều cao hơn một cách đáng kể so với nhóm dân số già ở tất cả các tỉnh, điều đó đã khẳng định thành công của chương trình phổ cập tiểu học và phong trào xoá mù trong vòng 20 năm trở lại đến 1999. Tuy nhiên khi phân tích chi tiết thì thấy đó không phải là thành công hoàn toàn, còn có những nhóm dân số trẻ có tỷ lệ biết đọc biết viết rất thấp.

Xét theo giới tính: trong nhóm dân số trẻ, chỉ có sự khác biệt nhỏ giữa nam và nữ trên hầu hết các tỉnh trên toàn quốc; đặc biệt là vùng Đồng bằng sông Hồng, Đồng bằng sông Cửu Long, một số tỉnh vùng Đông Nam Bộ và vùng Đông Bắc không thấy có sự khác biệt giữa nam và nữ. Sự khác biệt lớn

giữa nam và nữ thường tồn tại ở những tỉnh có tỷ lệ biết đọc, biết viết thấp.

Xét theo nơi cư trú, dân số trẻ nông thôn vẫn có tỷ lệ biết đọc biết viết thấp hơn của dân thành thị. Trong khi thành thị có 54 tỉnh có tỷ lệ này cao hơn 94%, thì mới có 32 tỉnh có khu vực nông thôn đạt tiêu chuẩn này. Không có tỉnh nào có tỷ lệ biết đọc, biết viết thấp hơn 90% ở khu vực thành thị, trong khi đó khu vực nông thôn có 6 tỉnh, 10 tỉnh khác có tỷ lệ này ở khu vực nông thôn nằm trong khoảng từ 80-90%.

### 3. Kết luận và khuyến nghị

Năm 1999, tỷ lệ biết đọc biết viết của dân số 15 tuổi trở lên 90,3%, riêng đối với nhóm dân số trẻ, tỷ lệ này là 94,1%, cho thấy có sự tiến bộ theo thế hệ về tình trạng biết đọc biết viết. Tuy nhiên, việc phân tích chi tiết cho các tỉnh và cho các nhóm dân số cho thấy ngay cả với nhóm dân số trẻ, còn tồn tại sự khác biệt giữa hai khu vực nông thôn và thành thị, điều này cho thấy chương trình giáo dục cho toàn dân đã có hiệu quả với dân thành thị nhiều hơn là với dân nông thôn.

Trong nhóm dân số trẻ, chỉ có sự khác biệt nhỏ giữa nam và nữ trên hầu hết các tỉnh trên toàn quốc. Sự khác biệt lớn giữa nam và nữ tồn tại ở những tỉnh có tỷ lệ biết đọc biết viết thấp.

Sự khác biệt giữa dân tộc Kinh và dân tộc thiểu số thể hiện chủ yếu là ở những vùng nông thôn miền núi. Theo kết quả nghiên cứu chi tiết có thể thấy phần lớn các nhóm dân số trẻ có tỷ lệ biết đọc biết viết rất thấp là nhóm đồng bào các dân tộc thiểu số sống ở nông thôn miền núi. Điều này cho

thấy các chương trình xóa nạn mù chữ trong thập kỷ vừa qua cho dù đã có nhiều nỗ lực vẫn chưa tạo được những đột phá để giải quyết tốt vấn đề biết đọc biết viết của đồng bào dân tộc thiểu số sống ở các vùng nông thôn miền núi.

Như vậy, nguyên nhân của tỷ lệ biết đọc biết viết thấp là do nơi cư trú, đặc biệt là nông thôn miền núi, vùng sâu vùng xa nơi cơ sở hạ tầng còn chưa được phát triển và do phương thức sống của đồng bào dân tộc thiểu số ở những vùng rừng núi và dân cư thưa thớt.

Thời kỳ 2000-2004, nhiều nỗ lực đã được thực hiện trong việc xóa nạn mù chữ và phổ cập tiểu học. Số liệu điều tra mức sống hộ gia đình 2002 cho thấy tỷ lệ biết đọc biết viết chung của dân số 10 tuổi trở lên là 92,1% tăng 1% so với năm 1999. Do với số liệu này không có điều kiện để phân tổ chi tiết như chúng tôi đã thực hiện với số liệu 3% của tổng điều tra dân số và nhà ở 1999, nhưng khi xem xét số liệu phân tổ cho 8 vùng địa lý của số liệu điều tra mức sống hộ gia đình năm 2002, chúng tôi thấy, sự khác biệt giữa các vùng vẫn giống như ở năm 1999, chưa thấy có sự đột phá nào.

Tuy chỉ còn 6% dân số trẻ trong toàn quốc chưa biết đọc biết viết nhưng số dân này lại tập trung chủ yếu ở vùng nông thôn miền núi và là dân tộc thiểu số, vì vậy việc hạ thấp tỷ lệ này không đơn giản. Thành công thực sự chỉ có thể có được khi Bộ Giáo dục và Đào tạo trong khi thực hiện chương trình xóa nạn mù chữ có sự phối hợp chặt chẽ của các chương trình xóa đói giảm nghèo và các (tiếp theo trang 9)

## ỨNG DỤNG HỒI QUI LOGISTIC NGHIÊN CỨU.... (tiếp theo trang 25)

chương trình phát triển cơ sở hạ tầng cho các vùng sâu vùng xa đặc biệt khó khăn.

Dân số trẻ là dân tộc thiểu số nông thôn có tỷ lệ biết đọc biết viết quá thấp (ví dụ: Lào Cai: 49,3%, Lai Châu: 38,6%, Quảng Trị: 45,0%; Quảng Ngãi: 56%; Gia Lai: 51,7%; Lâm Đồng: 68,3%; Ninh Thuận: 52%) cho thấy có thể do đồng bào dân tộc thiểu số thường sống thành các buôn bản và họ giao tiếp với nhau bằng tiếng dân tộc cho nên mặc dù đã đi học tiếng phổ thông nhưng do ít có điều kiện đọc sách báo nên lại thành mù chữ trở lại. Để giúp cho đồng bào các dân tộc ở những vùng này biết đọc biết viết bền vững, Bộ Giáo dục và Đào tạo nên phát triển mạnh mẽ các chương trình dạy tiếng dân tộc kết hợp với tiếng phổ thông có kết hợp phát hành các tài liệu thiết thực với đời sống của đồng bào viết bằng tiếng dân tộc và bằng tiếng phổ thông ■

---

*(1) Tuổi 35 được chọn làm điểm cắt bởi vì chiến dịch xóa nạn mù chữ hiện nay nhằm vào các đối tượng có độ tuổi từ 15-35 tuổi*

*(2) Dân tộc được phân thành nhóm dân tộc Kinh và nhóm các dân tộc khác do người Kinh chiếm đại bộ phận (86,3 % - Tổng Điều tra dân số 1999) dân số Việt nam. Số các dân tộc còn lại (bao gồm trên 50 dân tộc) chỉ chiếm có 13,5% dân số.*

### Tài liệu tham khảo:

1. Dữ liệu và kết quả điều tra mẫu 3% Tổng điều tra Dân số và Nhà ở 1/4/1999. Trung tâm tính toán Thống kê Trung ương, version 1.0, 8/2000.
2. Applied Logistic Regression, David W Hosmer & Stanley Lemeshow, second Edittion. John Wiley & Sons, Inc
3. Literacy in Vietnam-an Atlas, Tram Phan, Ayse Bilgin, Ann Eyland, Pamela Shaw, 2004. [http://www.stat.mq.edu.au/research/res\\_paper.htm](http://www.stat.mq.edu.au/research/res_paper.htm)
4. Education for all in Vietnam (1990-2000), Phạm Minh Hạc, National Committee Literacy, Hanoi-2000.
5. Kết quả điều tra mức sống hộ gia đình năm 2002, Nhà xuất bản Thống kê, 2002