

# Trường Chính sách Công và Quản lý Fulbright

## Nhập môn Kinh tế lượng

### Bài tập 3

Ngày Phát: Thứ sáu 10/12/2021 - Ngày Nộp: 8h20 Thứ ba 21/12/2021

#### Câu 1.

Số liệu khảo sát ghi nhận về số lượng thuốc lá tiêu thụ trung bình trong 01 ngày của người dân tại 44 bang của Hoa Kỳ được tổng hợp trong bảng sau.

State	LUNG <sup>1</sup>	CIG <sup>(2)</sup>		State	LUNG	CIG
1	17.05	18.20		23	19.50	23.75
2	19.80	25.82		24	16.70	23.32
3	15.98	18.24		25	23.03	42.40
4	22.07	28.6		26	25.95	28.64
5	22.83	31.10		27	14.59	21.16
6	24.55	33.60		28	25.02	29.14
7	27.27	40.46		29	12.12	19.96
8	23.57	28.27		30	21.98	26.38
9	13.58	20.10		31	19.45	23.44
10	22.80	27.91		32	12.11	23.78
11	20.30	26.18		33	23.68	29.18
12	16.59	22.12		34	17.45	18.06
13	16.84	21.84		35	14.11	20.94
14	17.71	23.44		36	17.60	20.08
15	25.45	21.58		37	20.74	22.57
16	20.94	28.92		38	12.01	14,00
17	26.48	25.91		39	21.22	25.89
18	22.04	26.92		40	20.34	21.17
19	22.72	24.96		41	20.55	21.25
20	14.20	22.06		42	15.53	22.86
21	15.60	16.08		43	15.92	28.04
22	20.98	27.56		44	25.88	30.34

(1): CIG – Số điều thuốc hút/người/ngày

(2): LUNG – Số người chết/100 ngàn dân được xác định là do ung thư phổi

Nguồn: .F. Fraumeni, "Cigarette Smoking and Cancers of the Urinary Tract: Geographic Variations in the United States," *Journal of the National Cancer Institute*

- a. Kỳ vọng của bạn về mối quan hệ giữa 2 biến trên là gì? Xác định đâu là biến phụ thuộc, đâu là biến độc lập (hay còn gọi là biến giải thích) dựa trên phát biểu về kỳ vọng? Viết phương trình hàm hồi quy đơn biến và giải thích các giả định của mô hình OLS được thể hiện như thế nào?

Chúng ta kỳ vọng có mối quan hệ tương quan dương giữa số điều thuốc trung bình/người/ngày và tỷ lệ chết do ung thư phổi, do đó hàm hồi quy đơn biến được xây dựng như sau:

$$LUNG_i = \beta_0 + \beta_1 CIG_i + u_i$$

Do đó,  $\beta_1$  dương.  $\beta_0$  là số người chết do ung thư phổi nếu không ai hút thuốc ( $CIG = 0$ ). Do đó, kỳ vọng  $\beta_0$  cũng dương.

- b. Vẽ đồ thị phân phối điểm về mối quan hệ giữa 2 biến trên? Tính hệ số tương quan giữa 2 biến. Nhận xét kết quả và so sánh với kỳ vọng của bạn ở câu a.

Học viên tự thực hiện. Kỳ vọng có mối quan hệ tương quan dương giữa hai biến. Hệ số tương quan Pearson được tính theo công thức:

$$\rho = \frac{\sum_i (LUNG_i - \overline{LUNG})(CIG_i - \overline{CIG})}{n - 1}$$

Lưu ý với sample nhỏ thì chúng ta dùng  $n-1$  thay vì  $n$ .

- c. Bằng công thức thủ công, bạn hãy ước lượng các tham số của mô hình. Viết phương trình hàm hồi quy vừa ước lượng. Việc trình bày cần ngắn gọn nhưng đảm bảo người đọc hiểu được cách tính ra các kết quả.

Học viên áp dụng công thức như trình bày trong bài giảng:

$$\hat{\beta}_1 = \frac{Cov(LUNG, CIG)}{Var(CIG)}$$

và

$$\hat{\beta}_0 = \overline{LUNG} - \hat{\beta}_1 \overline{CIG}$$

- d. Giải thích ý nghĩa của hệ số ước lượng từ phương trình hồi quy.

$\hat{\beta}_1$  là tỷ lệ chết tăng thêm (người/100,000 dân) nếu số điều thuốc trung bình tăng thêm một điều/ngày.

- e. Dựa vào các dữ liệu cung cấp như trên, bạn hãy cho biết biến độc lập đã giải thích được bao nhiêu % sự thay đổi của biến phụ thuộc?

Học viên cần ước lượng

$$R^2 = \frac{SSE}{SST} = \frac{Var(\hat{Y})}{Var(Y)} = \frac{\sum(\widehat{LUNG}_i - \overline{LUNG})}{\sum(LUNG_i - \overline{LUNG})}$$

Với  $\widehat{LUNG}_i = \hat{\beta}_0 + \hat{\beta}_1 \times CIG_i$ .  $\hat{\beta}_0$  và  $\hat{\beta}_1$  đã được tính từ câu c.

- f. Kiểm chứng kết quả ước lượng thủ công với kết quả ước lượng bằng một phần mềm bất kỳ.

Học viên tự kiểm chứng để cho thấy kết quả tính thủ công tương đương với kết quả đầu ra của phần mềm.

## Câu 2.

Bạn được giao nhiệm vụ phân tích tác động của chính sách giáo dục (thông qua số năm đi học) đến thu nhập cá nhân để báo cáo cho chính phủ. Học viên được yêu cầu xây dựng và ước lượng một mô hình hồi quy đa biến để giải thích các nhân tố ảnh hưởng đến thu nhập cá nhân hàng năm của nhóm người không học quá phổ thông trung học. Học viên được cung cấp bộ dữ liệu khảo sát mức sống hộ gia đình 2006, 2008, và 2010 đã xử lý thô. Đây là bài tập cần sự tư duy và các lập luận phù hợp nên không có đáp án duy nhất cho các câu hỏi. Học viên cần vận dụng kiến thức, sự sáng tạo cũng như kinh nghiệm thực tế khi đưa ra các giả định và giải thích các giả định một cách hợp lý.

- a. Đề xuất một mô hình hồi quy đa biến mô tả các nhân tố ảnh hưởng đến thu nhập cá nhân, tập trung vào yếu tố vốn con người. Thông thường các mô hình này sử dụng hàm  $\log(\text{income})$  làm biến phụ thuộc. Giải thích việc lựa chọn các biến độc lập và kỳ vọng về tác động của biến đó lên thu nhập.

Dựa theo lý thuyết về vốn con người chúng ta giả định mô hình hồi quy đa biến giải thích tỷ suất thu nhập của đi học như sau:

$$\log(\text{income}) = \beta_0 + \beta_1 \text{yoeduc} + \beta_2 \text{exper} + \beta_3 \text{exper}^2 + \beta_4 \text{sex} + \beta_5 \text{married} + u$$

Biến số năm kinh nghiệm có thể được tính gián tiếp qua tuổi và số năm đi học như sau:

$$exper = age - 6 - yoeduc$$

Học viên cũng thể dùng biến tuổi thay cho biến số năm kinh nghiệm, và các cấu trúc hàm khác. Tham số  $\beta_1$  là tỷ suất thu nhập của đi học, kỳ vọng là dương.  $\beta_2$  và  $\beta_3$  thể hiện tác động phi tuyến của số năm kinh nghiệm đến tiền lương. Dựa trên các bằng chứng thực nghiệm ở các quốc gia, chúng ta kỳ vọng kinh nghiệm tác động đến thu nhập phi tuyến, tăng sau đó giảm, do đó  $\beta_2$  dương và  $\beta_3$  âm. Kỳ vọng thu nhập của nam cao hơn của nữ, và người có gia đình có thu nhập cao hơn người chưa có gia đình, trong điều kiện các nhân tố khác không đổi. Do đó,  $\beta_4$  dương (giả định thông thường biến sex được mã hóa 0 là nữ 1 là nam), và  $\beta_5$  dương (married = 1 với người có gia đình, 0 nếu không phải, bao gồm độc thân, góa và ly thân).

- b. Lập bảng thống kê mô tả dữ liệu, so sánh thay đổi thu nhập qua thời gian và giữa các nhóm và giải thích ngắn gọn. Học viên có thể sử dụng biểu đồ và đồ thị phân phối nếu cần thiết.

Học viên tự thực hiện, bao gồm bảng biểu đồ thị, diễn giải các thông số như trung bình, độ lệch chuẩn, min/max, nhận định thay đổi theo thời gian (ví dụ tiền lương), khác biệt nhóm giữa nam/nữ, vẽ đồ thị minh họa...

Lưu ý kiểm tra cách mã hóa từng biến để đảm bảo diễn giải kết quả đúng với cách mã hóa biến. Ví dụ biến marital = 2 là có gia đình, = 1 là tình trạng khác. Chúng ta tạo biến married = 1 là có gia đình, = 0 nếu tình trạng khác. Việc tạo biến mới giúp diễn giải kết quả dễ hơn. Nếu học viên không tạo biến mới thì kết quả cũng không sai, tuy nhiên khi diễn giải thì lưu ý cách mã hóa biến để diễn giải cho đúng.

- c. Ước lượng mô hình đề xuất và diễn giải kết quả ngắn gọn. Thực hiện các kiểm định cần thiết.

Chúng ta lần lượt ước lượng các mô hình hồi quy đa biến trên. Kết quả cho thấy tỷ suất thu nhập của đi học từ 9-10%/năm, và đều có ý nghĩa thống kê. Tất cả các biến khác trong mô hình đều có dấu giống với kỳ vọng, đó là hàm phi tuyến parabol úp xuống theo số năm kinh nghiệm, nam thu nhập cao hơn nữ, và người có gia đình có thu nhập cao hơn người không gia đình (với hồi quy cho dữ liệu 2010). Với dữ liệu 2006 và 2008 thì tình trạng hôn nhân không có ý nghĩa thống kê. Học viên cũng có thể kiểm chứng số năm kinh nghiệm tối ưu từ 25-27 năm, tương tự như các nghiên cứu thực nghiệm khác đã thực hiện trên thế giới.

Các kết quả này cho thấy tỷ suất thu nhập của đi học ước lượng được vào khoảng 9-10% là có thể tin cậy được. Giai đoạn 2006-2010, kinh tế Việt Nam còn ở mức độ đang phát

triển nên những con số này không quá chênh lệch so với kỳ vọng ở các nước đang phát triển. Với những nước đã phát triển, tỷ suất thu nhập của đi học thấp hơn.

Ngoài ra, trong bài này chúng ta không xét đến vấn đề mô hình thiếu biến quan trọng có tương quan với số năm đi học, chẳng hạn như tố chất cá nhân vừa có tương quan với số năm đi học vừa tương quan dương với thu nhập. Khi này, mô hình trên gặp phải vấn đề thiếu biến quan trọng có tương quan với biến giải thích, vi phạm điều kiện 4.2, và ước lượng OLS của  $\beta_1$  bị chệch lên (và không nhất quán).

- d. Kiểm tra độ vững của kết quả bằng cách ước lượng mô hình cho các nhóm đối tượng khác nhau (ví dụ nam hoặc nữ, nhóm tuổi, nhóm thu nhập, tác động của quan sát ngoại vi...) và giải thích ngắn gọn kết quả.

Học viên có thể kiểm tra độ vững của kết quả cho các nhóm nam và nữ (kiểm tra liệu tỷ suất thu nhập của đi học của nhóm nữ có đúng là cao hơn nhóm nam?), nhóm tuổi, so sánh giữa các năm, loại quan sát ngoại vi theo các quy tắc thực nghiệm.

### **Lưu ý với bài tập nộp:**

Học viên nộp bài viết, chất lọc các thông tin quan trọng và trình bày kết quả theo yêu cầu của đề bài, tránh in tất cả dữ liệu đầu ra từ phần mềm để nộp bài. Học viên cần gửi kèm do-file và log-file (bản điện tử) để trợ giảng đối chiếu chấm điểm.