

Chương Trình Giảng Dạy Kinh tế Fulbright

Học kỳ Thu năm 2012

Các Phương Pháp Phân Tích Định Lượng

LỜI GIẢI ĐỀ NGHỊ BÀI TẬP 7 HỒI QUY ĐƠN BIẾN (TIẾP THEO)

Ngày Phát: Thứ hai 26/11/2012

Ngày Nộp: Thứ hai 03/12/2012

Bản in nộp lúc **8h20 sáng**, tại Hộp nộp bài tập trong phòng Lab

Bản điện tử gửi lên <http://intranet.fetp.edu.vn:81>

BÀI 1:

Tiến hành hồi quy thực nghiệm biến ngẫu nhiên CHI_TIEU_Y_TE (ngàn đồng/năm) theo biến ngẫu nhiên THU_NHAP (ngàn đồng/năm), kết quả phân tích hồi quy được thể hiện ở bảng dưới đây:

Dependent Variable: CHI_TIEU_Y_TE
Method: Least Squares
Date: 11/26/12 Time: 17:05
Sample: 1 65
Included observations: 65

Variable	Coefficient	Std. Error	t-Statistic	Prob.
THU_NHAP	0.019107	0.005652	B	D
C	297.7579	332.7996	C	E
R-squared	0.153567	Mean dependent var		1145.923
Adjusted R-squared	0.140131	S.D. dependent var		1901.220
S.E. of regression	A	Akaike info criterion		17.81769
Sum squared resid	1.96E+08	Schwarz criterion		17.88459
Log likelihood	-577.0749	Hannan-Quinn criter.		17.84409
F-statistic	F	Durbin-Watson stat		1.346597
Prob(F-statistic)	G			

Dựa vào bảng trên trả lời các câu hỏi sau:

a. Điền các số liệu còn thiếu (A,B,C,D,E,G,F) ở bảng trên và trình bày cách tính toán.

$$- A = 1762,984 = \sqrt{\frac{\sum \hat{u}_i^2}{n-2}} = \sqrt{\frac{1,96.10^8}{65-2}}$$

$$- B = 3,380823 = \frac{\hat{\beta}_2 - \beta_2}{SE_{\beta_2}} = \frac{0,019107-0}{0,005652}$$

$$- C = 0,894706 = \frac{\hat{\beta}_1 - \beta_1}{SE_{\beta_1}} = \frac{297,7579 - 0}{332,7996}$$

$$- D = 0,0012 - \text{Tra bảng t với bậc tự do 63 và trị thống kê } 3,380823$$

$$- E = 0,3743 - \text{Tra bảng t với bậc tự do 63 và trị thống kê } 0,894706$$

$$- F = 11,42996 = B^2$$

$$- G = 0,001245 - \text{Tra bảng Chi-square với bậc tự do 63 và trị thống kê } 11,42996.$$

b. Viết phương trình đường hồi quy mẫu?

gọi biến CHI_TIEU_Y_TE là Y

biến THU_NHAP là X

SRF có dạng: $\hat{Y}_i = 297,7579 + 0,019107X_i$

c. Bao nhiêu phần trăm sự biến thiên về CHI_TIEU_Y_TE được giải thích bởi THU_NHAP?

Với hệ số xác định $R^2 = 0.153567$ ở mô hình hồi quy trên, cho ta biết nếu trong điều kiện các yếu tố khác không đổi thì biến THU_NHAP giải thích được khoảng 15,35% sự biến thiên về mặt trung bình của biến CHI_TIEU_Y_TE.

d. Sử dụng kiểm định dựa trên giá trị thống kê t, hãy xác định xem có mối quan hệ tuyến tính có ý nghĩa về mặt thống kê với mức ý nghĩa 5% giữa hai biến CHI_TIEU_Y_TE và THU_NHAP hay không.

Đặt - $H_0 : \beta_2 = 0$: THU_NHAP không ảnh hưởng đến CHI_TIEU_Y_TE

- $H_a: \beta_2 \neq 0$: THU_NHAP có ảnh hưởng đến CHI_TIEU_Y_TE.

Trị kiểm định: $t_{n-2, \alpha/2} = t_{63; 0,025} = 1,9982$

Giá quan sát $t_{qs} = 3,380573248$

So sánh ta có: $t_{qs} > t_{n-2, \alpha/2}$ vậy ta có thể bác bỏ H_0 ở mức ý nghĩa 5% tức là mối quan hệ tuyến tính giữa hai biến CHI_TIEU_Y_TE và THU_NHAP có ý nghĩa về mặt thống kê với mức ý nghĩa 5% hay biến THU_NHAP có ảnh hưởng đến biến CHI_TIEU_Y_TE ở độ tin cậy 95%.

e. Sử dụng kiểm định dựa trên khoảng tin cậy của độ dốc trong phương trình hồi qui của tổng thể, hãy xác định xem có mối quan hệ tuyến tính có ý nghĩa về mặt thống kê với mức ý nghĩa 5% giữa hai biến CHI_TIEU_Y_TE và THU_NHAP hay không. So sánh với kết quả câu d.

Đặt - $H_0 : \beta_2 = 0$: THU_NHAP không ảnh hưởng đến CHI_TIEU_Y_TE

- $H_a: \beta_2 \neq 0$: THU_NHAP có ảnh hưởng đến CHI_TIEU_Y_TE.

Với: $t_{n-2, \alpha/2} = t_{63; 0,025} = 1,9982$

Khoảng tin cậy $(1 - \alpha) * 100\%$ của β_2

$$\hat{\beta}_2 \pm t_{\alpha/2} s_{\hat{\beta}_2} = 0,019107 \pm 1,9982 * 0,005652 = 0,019107 \pm 0,0112927$$

Tính toán ta có với mức ý nghĩa 5% thì khoảng tin cậy của ước lượng β_2 là [0,007813; 0,030400]

Nhận xét: Ở mức ý nghĩa 5% thì khoảng tin cậy của ước lượng β_2 không chứa giá trị 0 cho nên ta có cơ sở để bác bỏ H_0 . Như vậy có mối quan hệ tuyến tính có ý nghĩa về mặt thống kê với mức ý nghĩa 5% giữa hai biến CHI_TIEU_Y_TE và THU_NHAP.

Kết quả này giống kết quả đã kết luận ở câu d.

f. Sử dụng kiểm định dựa trên giá trị p-value, hãy xác định xem có mối quan hệ tuyến tính có ý nghĩa về mặt thống kê với mức ý nghĩa 5% và 1% giữa hai biến CHI_TIEU_Y_TE và THU_NHAP hay không. Giải thích ý nghĩa của giá trị p-value.

Đặt - $H_0 : \beta_2 = 0$: THU_NHAP không ảnh hưởng đến CHI_TIEU_Y_TE

- $H_a: \beta_2 \neq 0$: THU_NHAP có ảnh hưởng đến CHI_TIEU_Y_TE.

So sánh giá trị p-value của ước lượng β_2 lần lượt với mức ý nghĩa 5% và 1% như sau:

+ $P_{\text{value}} = 0,001246 < \alpha = 5\%$ nên bác bỏ giả thuyết H_0 .

+ $P_{\text{value}} = 0,001246 < \alpha = 1\%$ nên bác bỏ giả thuyết H_0 .

Như vậy với cả mức ý nghĩa 5% và 1% thì giữa hai biến CHI_TIEU_Y_TE và THU_NHAP có mối quan hệ tuyến tính có ý nghĩa về mặt thống kê.

Giải thích ý nghĩa của giá trị p-value: cho biết xác suất mắc sai lầm loại 1 tối đa khi bác bỏ giả thuyết H_0 . Như vậy với giá trị p-value của ước lượng β_2 có nghĩa là ta chỉ mắc sai lầm 0,1246% khi bác bỏ giả thuyết H_0 trong khi đó sai lầm cho phép lần lượt là 5% và 1%.

g. Nếu thu nhập của một cá nhân tăng thêm 1 triệu đồng/năm thì chi tiêu bình quân cho y tế tăng bao nhiêu?

Từ phương trình hồi quy mẫu: $\hat{Y}_i = 297,7579 + 0,019107X_i$

Ta có: nếu thu nhập của một cá nhân tăng thêm 1 triệu đồng/năm thì chi tiêu bình quân cho y tế sẽ tăng : $0,019107 * 1.000 = 19.107$ ngàn đồng/năm với các điều kiện khác không đổi

h. Giả sử CHI_TIEU_Y_TE và THU_NHAP đo bằng đơn vị triệu đồng/năm thay vì đo bằng ngàn đồng/năm. Theo Anh/Chị việc thay đổi này có ảnh hưởng gì đến giá trị độ dốc và tung độ gốc đã được xác định ở bảng trên hay không. Theo Anh/Chị việc thay đổi đơn vị này có ảnh hưởng đến mối quan hệ tuyến tính có ý nghĩa về mặt thống kê với mức ý nghĩa 5% giữa hai biến CHI_TIEU_Y_TE và THU_NHAP hay không.

Sự thay đổi đơn vị tính từ ngàn đồng/năm thành triệu đồng/năm chỉ làm thay đổi tung độ góc – giảm 1.000 lần còn giá trị độ dốc không thay đổi. (có thể chứng minh bằng việc biến đổi hàm hồi quy mẫu)

Việc thay đổi này không làm ảnh hưởng đến mối quan hệ tuyến tính có ý nghĩa về mặt thống kê với mức ý nghĩa 5% giữa hai biến chi tiêu y tế và thu nhập vì 2 biến cùng các đơn vị đo lường và các giá trị kiểm định sẽ không thay đổi

BÀI 2:

a. Tính toán các thống kê mô tả cho biến tổng chi tiêu cho y tế (Y) và biến tuổi chủ hộ (X3).
 Tính toán các thống kê mô tả cho biến tổng chi tiêu cho y tế (Y)

<i>Tổng chi tiêu cho y tế của hộ Y (1000đ/năm)</i>	
Mean	1145.923077
Standard Error	235.817275
Median	500
Mode	0
Standard Deviation	1901.219652
Sample Variance	3614636.166
Kurtosis	12.37948821
Skewness	3.301999451
Range	10530
Minimum	0
Maximum	10530
Sum	74485
Count	65
Confidence Level(95.0%)	471.0991632

Tính toán các thống kê mô tả cho biến tuổi chủ hộ (X3).

Tuổi của chủ hộ X3

Mean	51.76923077
Standard Error	1.832895097
Median	48
Mode	43
Standard Deviation	14.7772727
Sample Variance	218.3677885
Kurtosis	-0.717803321
Skewness	0.229495525
Range	69
Minimum	18
Maximum	87
Sum	3365
Count	65
Confidence Level(95.0%)	3.661628889

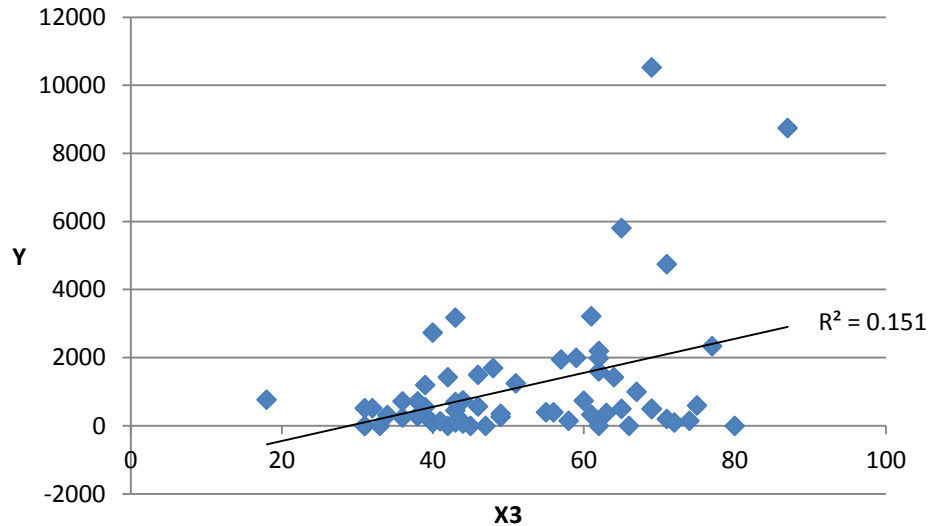
b. Theo những hiểu biết của bạn thì tuổi của chủ hộ có ảnh hưởng tới việc chi tiêu cho y tế hay không, bạn có kỳ vọng gì về dấu của mối quan hệ này.

Khi tuổi càng cao thì sức khỏe có xu hướng giảm dần cho nên sẽ chi tiêu cho y tế nhiều hơn. Sự thay đổi việc chi tiêu cho y tế phụ thuộc vào từng độ tuổi của chủ hộ, chẳng hạn: nếu chủ hộ ở tuổi thanh niên thì ngoài chi tiêu y tế cho bản thân thì hộ vẫn còn chi tiêu cho con trẻ nên việc chi tiêu y tế của hộ có thể sẽ cao hơn chủ hộ ở độ tuổi trung niên. Nếu chủ hộ là cao

niên thì rõ ràng chi tiêu y tế cũng sẽ tăng thêm do chi tiêu y tế cho người cao tuổi. Như vậy, việc chi tiêu y tế sẽ là khác nhau tùy vào từng độ tuổi của chủ hộ

Tuy nhiên xét theo tổng thể thì kỳ vọng biến tuổi của chủ hộ sẽ có ảnh hưởng đồng biến tới việc chi tiêu cho y tế và mối tương quan giữa hai biến này là mối tương quan thuận. Do đó dấu của mối quan hệ này là dấu dương (+).

c. Vẽ biểu đồ phân tán của Y theo X3, tính hệ số tương quan giữa 2 biến này, cho biết nhận xét của bạn.



Như vậy hệ số tương quan giữa hai biến là $r_{x_3,y} = \sqrt{R^2} = 0,38858 \Rightarrow X_3$ và Y tương quan đồng biến nhưng tương quan yếu điều này là đúng theo kỳ vọng do việc chi tiêu y tế thay đổi theo từng độ tuổi của chủ hộ.

d. Hãy ước lượng hàm hồi quy thể hiện mối quan hệ tuyến tính $Y = f(X_3)$.

Sử dụng các phần mềm hồi quy sẽ cho hàm SRF như sau:

$$\hat{Y}_i = -1.442,234 + 49,99414 * X_{3i}$$

e. Hãy cho biết tác động biên của tuổi chủ hộ lên chi tiêu cho y tế? Bạn có khuyến nghị chính sách gì dựa trên kết quả này.

Tác động biên của tuổi chủ hộ lên chi tiêu cho y tế chính là hệ số ước lượng của β_2 :

$\hat{\beta}_2 = 49,99$ cho ta biết: Trong khi điều kiện các yếu tố khác không đổi, khi tuổi chủ hộ tăng thêm 1 tuổi thì về mặt trung bình, chi tiêu cho y tế tăng thêm 49,99 ngàn đồng.

Khuyến nghị chính sách: Một số lý thuyết cho thấy tuổi của chủ hộ cũng ảnh hưởng đến thu nhập của hộ, đặc biệt là đối với các hộ mà chủ hộ đang ở độ tuổi thanh niên và cao niên thì thu nhập của hộ là thấp hơn so với độ tuổi khác trong khi chi tiêu y tế lại cao hơn. Do đó khuyến nghị nhằm vào 2 đối tượng mà chủ hộ đang ở trong độ tuổi này như sau: Chính sách BHYT và khám chữa bệnh miễn phí cho trẻ em (con của chủ hộ trong độ tuổi thanh niên) và cho người già

f. Thay vì tìm mối quan hệ tuyến tính giữa Y và X3, bây giờ, tìm mối quan hệ tuyến tính giữa Y và (X3)2. Hãy ước lượng hàm hồi quy tuyến tính $Y = f(X_3^2)$.

Sử dụng các phần mềm để hồi quy!(các bạn tự chạy hồi quy)

Hàm hồi quy thể hiện mối quan hệ tuyến tính $Y = f(X_3^2)$:

$$\hat{Y}_i = -281,0661 + 0,492905(X_{3i})^2$$

g. Theo bạn thì hàm ước lượng nào tốt hơn? Vì sao?

Để xem xét hàm ước lượng nào tốt hơn, cần quan tâm hệ số xác định R^2 (cho biết phần trăm sự biến thiên về mặt trung bình của Y có thể giải thích từ mô hình) của hai mô hình. Mặc dù hệ số xác định của hai mô hình đều thấp, nhưng mô hình 2 có R^2 lớn hơn nên mô hình này có khả năng giải thích sự biến thiên về mặt trung bình của Y cao hơn.

Do đó kết luận mô hình 2 ước lượng tốt hơn.

h. Theo bạn có thể dùng hàm ước lượng $Y = f(X_3, X_3^2)$ để chỉ ra tác động của tuổi chủ hộ lên chi tiêu y tế hay không? Vì sao?

Sử dụng phần mềm hồi quy để chạy mô hình (các bạn tự chạy hồi quy)

Ta có hàm hồi quy tuyến tính $Y = f(X, X_3^2)$, như sau:

$$\hat{Y}_i = 2771,55 - 120,8235X_{3i} + 1,599035(X_{3i})^2$$

Theo kết quả hồi quy ta có: P-value (X_3) = 22,72% và P-value (X_3^2) = 8,62% là tương đối cao, tuy nhiên theo như kỳ vọng ban đầu và một số lý thuyết đã nghiên cứu trước thì mối quan hệ giữa biến tuổi chủ hộ và chi tiêu y tế có mối quan hệ phi tuyến và gần tương đương với hàm bậc 2 cho nên ta vẫn có thể dùng hàm này để chỉ ra tác động của tuổi chủ hộ lên chi tiêu y tế.

Mặt khác mô hình hồi quy này không vi phạm các giả thuyết OLS (không có đa cộng tuyến hoàn hảo).