

Chương trình Giảng dạy Kinh tế Fulbright

Học kỳ Thu năm 2014

Các Phương Pháp Phân Tích Định Lượng

GỢI Ý GIẢI BÀI TẬP 8

HỒI QUY ĐA BIẾN

Ngày Phát: Thứ Ba 02/12/2014

Ngày Nộp: Thứ Ba 09/12/2014

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: (30 điểm)

Từ dữ liệu cho ở file Excel *câu 1* về mối quan hệ giữa thu nhập của người lao động và các nhân tố liên quan, các biến được định nghĩa:

- AHE: thu nhập bình quân 1 giờ lao động (USD/giờ/lao động).
- AGE: Tuổi của người lao động.
- GENDER: Giới tính của người lao động (Nam = 1, Nữ = 0)
- EDU: Số năm đi học của người lao động.

- a. Tóm tắt dữ liệu của tất cả các biến bằng cách sử dụng các đại lượng thống kê mô tả.

Sử dụng chức năng Group Views\Descriptive Statistic\Common Sample ta có bảng kết quả:

	AHE	AGE	GENDER	EDU
Mean	16.74272	29.49763	0.592542	13.55017
Median	14.61539	29.00000	1.000000	13.00000
Maximum	97.50000	30.00000	1.000000	18.00000
Minimum	2.136752	29.00000	0.000000	6.000000
Std. Dev.	9.402076	0.500079	0.491445	2.314100
Skewness	2.303880	0.009492	-0.376678	-0.265782
Kurtosis	13.17822	1.000090	1.141886	3.646864
Jarque-Bera	15343.40	491.6667	494.1412	86.16374
Probability	0.000000	0.000000	0.000000	0.000000
Sum	49391.02	87018.00	1748.000	39973.00
Sum Sq. Dev.	260688.8	737.4834	712.2359	15792.07
Observations	2950	2950	2950	2950

	Biến thu nhập bình quân/1h lao động (AHE)	Biến tuổi (AGE)	Biến số năm đi học (EDU)	Giới tính (GENDER)
Trung bình	16.74272	29.49763	13.55017	0.592542
Trung vị	14.61539	29.00000	13.00000	1.000000
Lớn nhất	97.50000	30.00000	18.00000	1.000000

Nhỏ nhất	2.136752	29.00000	6.000000	0.000000
Độ lệch chuẩn	9.402076	0.500079	2.314100	0.491445
Tổng	49391.02	87018.00	39973.00	1748.000

Sử dụng chức năng One-Way Tabulation trong Eviews ta có các bảng kết quả:

Tabulation of GENDER
 Included observations: 2950
 Number of categories: 2

Value	Count	Percent
0	1202	40.75
1	1748	59.25
Total	2950	100.00

- Biến AHE: Với mẫu gồm 2.950 quan sát, thu nhập bình quân 1 giờ của người lao động trung bình ở mức 16,74 USD/giờ/lao động, biến thiên lớn trong khoảng từ 2,14 đến 97,5 USD/giờ/lao động với giá trị trung vị là 14.62 USD/giờ/lao động, độ lệch chuẩn là 9,4 USD/giờ/lao động. Hệ số Skewness = 2,3 có giá trị dương cho thấy phân phối của AHE bị lệch xiên về bên phải.

- Biến AGE: trong 2.950 lao động được chọn ngẫu nhiên trong mẫu, độ tuổi của người lao động xoay quanh mức 29-30 tuổi với độ tuổi trung bình là 29,5 tuổi, giá trị trung vị là 29 tuổi, độ lệch chuẩn 0,5 tuổi và hệ số Skewness = 0,0095 cho thấy phân phối độ tuổi của người lao động gần với phân phối đối xứng (số lượng người ở độ tuổi 29 xấp xỉ số lượng người ở độ tuổi 30).

- Biến EDU: với mẫu ngẫu nhiên gồm 2.950 lao động, số năm đi học của người lao động dao động trong khoảng 6-18 năm với giá trị trung bình là 13,55 năm, giá trị trung vị là 13 năm, độ lệch chuẩn là 2,31 năm. Hệ số Skewness = -0,2658 có giá trị âm cho thấy phân phối của EDU có xu hướng hơi lệch xiên về bên trái.

- Biến GENDER: trong 2.950 lao động được chọn ngẫu nhiên vào mẫu, số lượng lao động nữ là 1.202 người tương đương 40,75% và số lao động nam là 1.748 người chiếm 59,25%.

b. Nêu dấu kỳ vọng:

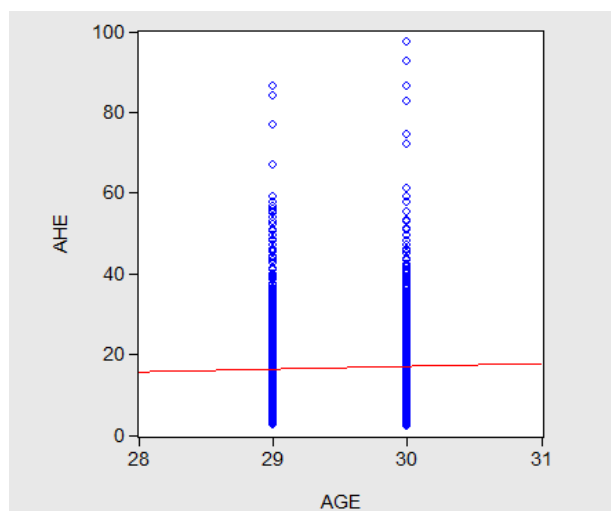
Quan hệ giữa biến thu nhập AHE với biến tuổi AGE là đồng biến, do kỳ vọng một người lớn tuổi sẽ có kinh nghiệm nhiều hơn nên được trả lương cao hơn.

Quan hệ giữa biến thu nhập AHE với biến số năm đi học EDU là đồng biến. Số năm đi học nhiều hơn trình độ cao hơn mức trả lương cao hơn.

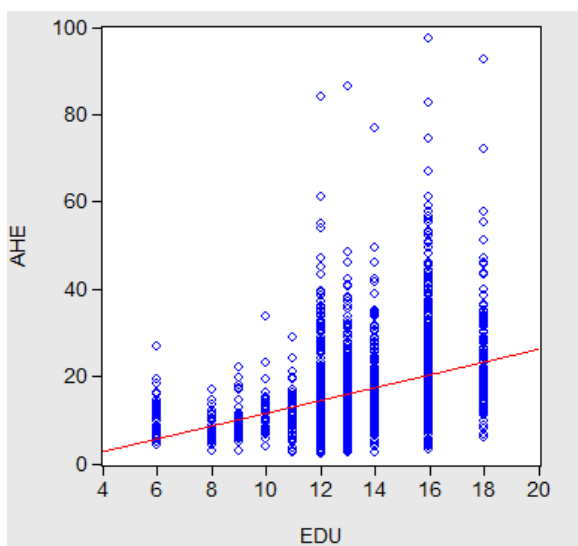
Quan hệ giữa biến thu nhập AHE với biến giới tính GENDER là đồng biến, với quy định nam = 1 và nữ = 0, thì biến thu nhập cao hơn ứng với trường hợp nam và thấp hơn đối với nữ vì kỳ vọng nam giới có sức khỏe tốt có thể có thu nhập cao hơn, nữ giới vì hạn chế về sức khỏe và thời gian chăm sóc gia đình, con cái nên kỳ vọng thu nhập có thể thấp hơn nam giới.

c. Vẽ biểu đồ phân tán:

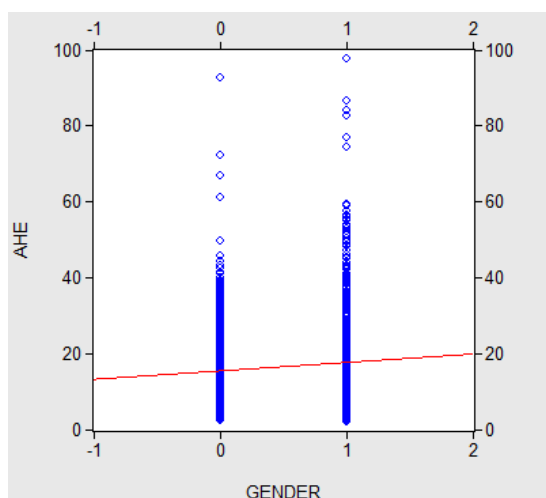
- Biểu đồ phân tán của AHE theo biến tuổi AGE cho thấy hơi có sự chênh lệch về thu nhập, tuổi cao hơn có khuynh hướng lương cao hơn.



- Biểu đồ phân tán của AHE theo nhân tố năm đi học EDU xu hướng đồng biến rõ nét hơn biến tuổi tác.



- Biểu đồ phân tán của AHE theo biến giới tính GENDER khuynh hướng nam giới có thu nhập cao hơn.



- Xác định ma trận tương quan cặp giữa tất cả các biến có trong dữ liệu và thực hiện kiểm định:

Ma trận tương quan cặp giữa tất cả các biến số định lượng:

Covariance Analysis: Ordinary
Date: 11/26/14 Time: 01:29
Sample: 1 2950
Included observations: 2950

Correlation Probability	AHE	AGE	EDU
AHE	1.000000 -----		
AGE	0.037400 0.0422	1.000000 -----	
EDU	0.361049 0.0000	0.000982 0.9575	1.000000 -----

- Đa phần các biến không có tương quan mạnh, chỉ có 2 biến AHE và EDU là có tương quan tuyến tính lớn hơn các tương quan khác nhưng vẫn ở mức dưới trung bình
- Xét về ý nghĩa thống kê, các biến đều có mối quan hệ có ý nghĩa thống kê với biến AHE ở mức 5% (do p-value nhỏ hơn 5%); Tương quan giữa AGE và EDU không có ý nghĩa ở mức 5% .

Bài 2

$$Y = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \beta_5 X_{5t} + \beta_6 X_{6t} + u_t \text{ (Mô hình A)}$$

Y: Lượng cáp điện thoại lắp đặt hàng năm (triệu feet/năm)

X₂: GNP (tỉ đô la)

X₃: Số lượng nhà dân cư xây mới (1000 nhà/năm)

X₄: Tỷ lệ thất nghiệp (%)

X₅: lãi suất cơ bản khách hàng được vay (%)

X₆: tỉ lệ hiệu quả đường truyền (%)

- Bảng lý thuyết kinh tế vĩ mô, nêu dấu kỳ vọng của các hệ số β trong mô hình
 - β_2 : dương. Kỳ vọng khi thu nhập tăng, nhu cầu lắp đặt điện thoại cũng sẽ tăng.
 - β_3 : dương. Kỳ vọng khi số lượng nhà dân cư xây mới càng tăng, quy mô thị trường đối với điện thoại cũng tăng.
 - β_4 : âm. Kỳ vọng khi tỉ lệ thất nghiệp tăng, nhu cầu sử dụng điện thoại kỳ vọng sẽ giảm do thu nhập của người tiêu dùng giảm, nhu cầu liên lạc cũng giảm đáng kể.
 - β_5 : âm. Kỳ vọng khi lãi suất cơ bản khách hàng được vay càng cao, nhu cầu đầu tư lắp đặt điện thoại cũng sẽ giảm.
 - β_6 : âm. Kỳ vọng khi tỉ lệ hiệu quả đường truyền cao, với một số lượng nhà dân cư không đổi, thì sẽ cần ít đường truyền để cung cấp tín hiệu, hay lượng cáp lắp đặt sẽ ít hơn.
- Ước lượng các hệ số trong Mô hình A bằng Eviews, nhận xét và giải thích các ý nghĩa thống kê và tác động biến cho từng hệ số hồi quy.

Dependent Variable: Y
Method: Least Squares
Date: 11/26/14 Time: 01:32
Sample: 1 16

Included observations: 16

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5962.656	2507.724	2.377716	0.0388
X2	4.883663	2.512542	1.943714	0.0806
X3	2.363956	0.843559	2.802361	0.0187
X4	-819.1287	187.7072	-4.363863	0.0014
X5	12.01048	147.0496	0.081676	0.9365
X6	-851.3927	292.1447	-2.914284	0.0155
R-squared	0.822750	Mean dependent var		7543.125
Adjusted R-squared	0.734125	S.D. dependent var		1217.152
S.E. of regression	627.6005	Akaike info criterion		16.00168
Sum squared resid	3938824.	Schwarz criterion		16.29140
Log likelihood	-122.0134	Hannan-Quinn criter.		16.01652
F-statistic	9.283507	Durbin-Watson stat		2.484497
Prob(F-statistic)	0.001615			

Nhận xét và giải thích

- Các kết quả phù hợp với kỳ vọng ban đầu ngoại trừ $\hat{\beta}_5$ có dấu ngược với kỳ vọng
- Mô hình có thể giải thích tương đối tốt do R^2 hiệu chỉnh có giá trị tương đối cao (73.41%)

Ý nghĩa thống kê của từng hệ số:

- Các hệ số ước lượng $\hat{\beta}_2$, $\hat{\beta}_3$, $\hat{\beta}_4$ và $\hat{\beta}_6$ đều có P-value nhỏ hơn 10% nên mỗi quan hệ giữa biến Y và các biến X_2 , X_3 , X_4 và X_6 có ý nghĩa thống kê tại mức 10%.
- Hệ số ước lượng $\hat{\beta}_5$ có P-value = 93.65% nên mỗi quan hệ giữa Y và X_5 không có ý nghĩa về mặt thống kê ở mức 10%.

Ý nghĩa các hệ số ước lượng:

- $\hat{\beta}_2 = 4.884$, ta có, trong điều kiện các yếu tố khác không đổi, khi X_2 (GNP) tăng 1 tỉ đô la, lượng cáp điện thoại lắp đặt hàng năm trung bình sẽ tăng 4.884 triệu feet/năm.
- $\hat{\beta}_3 = 2.364$, ta có, trong điều kiện các yếu tố khác không đổi, nếu X_3 (số lượng nhà dân cư) tăng 1000 nhà/năm, lượng cáp điện thoại lắp đặt hàng năm trung bình sẽ tăng 2.364 triệu feet/năm.
- $\hat{\beta}_4 = -819.129$, ta có, trong điều kiện các yếu tố khác không đổi, nếu X_4 (tỉ lệ thất nghiệp) tăng 1 điểm % thì lượng cáp điện thoại lắp đặt hàng năm trung bình sẽ giảm 819.129 triệu feet/năm.
- $\hat{\beta}_6 = -851.393$ trong điều kiện các yếu tố khác không đổi, nếu X_6 (tỉ lệ hiệu quả đường truyền) tăng 1 điểm % thì lượng cáp điện thoại lắp đặt hàng năm trung bình sẽ giảm 851.393 triệu feet/năm.

c. Độ co giãn trung bình của Y theo X_4 và X_6 , giải thích kết quả.

$$E_{YX_i} = \frac{\Delta Y/Y}{\Delta X_i/X_i} = \frac{\partial Y}{\partial X_i} * \frac{X_i}{Y} = \hat{\beta}_i * \frac{X_i}{Y}$$

Độ co giãn của Y theo X_4

$$E_{YX_4} = \hat{\beta}_4 * \frac{\bar{X}_4}{\bar{Y}} = -819.129 * \frac{6.388}{7543.15} = -0.6936$$

- Giải thích: Độ co giãn trung bình của Y theo X4 bằng -0.6936 nghĩa là trong điều kiện các yếu tố khác không đổi, nếu tỉ lệ thất nghiệp tăng thêm 1% thì lượng cấp laptop sẽ giảm trung bình 0.6936%.

Tương tự:

$$E_{YX_6} = -0.4395 \text{ với } \bar{X}_6 = 3.894$$

- Giải thích: Độ co giãn trung bình của Y theo X6 bằng -0.4395 nghĩa là trong điều kiện các yếu tố khác không đổi, nếu tỉ lệ hiệu quả đường truyền tăng thêm 1% thì lượng cấp laptop sẽ giảm trung bình 0.4395%.

d. Chọn mô hình hồi quy tốt hơn mô hình ở câu b

Từ nhận xét ở câu b cho thấy biến X₅ không đạt mức ý nghĩa kiểm định 10% vì vậy loại biến X₅ ra khỏi mô hình. Ta tiến hành chạy lại mô hình trên Eviews với các biến còn lại

Mô hình mới cho giá trị P_{value} của các biến X₂, X₃, X₄, X₆ đều nhỏ hơn 5%. Các biến này đều có ý nghĩa về mặt lý thuyết cũng như thống kê. Hệ số R²_{adjust} tăng lên đạt 0.758 cho thấy mô hình mới giải thích tốt sự biến động của Y Chọn mô hình này.

Dependent Variable: Y
Method: Least Squares
Date: 11/26/14 Time: 01:39
Sample: 1 16
Included observations: 16

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6031.919	2250.907	2.679773	0.0214
X2	5.052408	1.363741	3.704816	0.0035
X3	2.309170	0.487881	4.733064	0.0006
X4	-824.3777	168.2115	-4.900841	0.0005
X6	-864.4400	233.2971	-3.705318	0.0035
R-squared	0.822632	Mean dependent var		7543.125
Adjusted R-squared	0.758134	S.D. dependent var		1217.152
S.E. of regression	598.5931	Akaike info criterion		15.87735
Sum squared resid	3941451.	Schwarz criterion		16.11878
Log likelihood	-122.0188	Hannan-Quinn criter.		15.88971
F-statistic	12.75448	Durbin-Watson stat		2.485635
Prob(F-statistic)	0.000408			

e. Viết kết quả mô hình sau cùng.

$$\hat{Y} = 6031.919 + 5.052X_2 + 2.309X_3 - 824.38X_4 - 864.44X_6$$

$$P_{\text{value}} \quad 0.0035 \quad 0.0006 \quad 0.0005 \quad 0.0035$$

$$R^2_{\text{adjust}} = 0.758134$$

f. Khuyến cáo với nhà làm chính sách:

Trên quan điểm của một người muốn tăng lượng cấp điện thoại laptop hàng năm, nhà hoạch định chính sách cần phải xem xét các giải pháp nhằm nâng cao thu nhập người dân, giảm tỉ lệ thất nghiệp, và khuyến khích người dân tách hộ để tăng số lượng nhà ở (điều này có thể được

do xu hướng sống thành đại gia đình của người Á Đông). Tuy nhiên vấn đề hiệu quả đường truyền không nên bàn đến trong trường hợp này vì không thể kỳ vọng làm giảm hiệu quả đường truyền.

Bài 3

- a. Giải thích mối quan hệ kỳ vọng có thể giữa CM và 2 biến FLFP, PGN:
- Mối quan hệ giữa FLFP và CM là nghịch biến. Khi tỉ lệ nữ biết chữ trong lực lượng lao động FLFP tăng, thì phụ nữ có kiến thức về sinh đẻ và chăm sóc trẻ sơ sinh cao hơn, nhờ vậy sẽ có ít trẻ tử vong.
 - Mối quan hệ giữa PGNP và CM là nghịch biến. Khi thu nhập bình quân đầu người PGNP tăng, các điều kiện chăm sóc cho trẻ sơ sinh sẽ tốt hơn vì vậy mà tỉ lệ tử vong của trẻ sẽ giảm.
- b. Thực hiện hồi quy:

Dependent Variable: CM
Method: Least Squares
Date: 11/26/14 Time: 01:41
Sample: 1 64
Included observations: 64

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	263.6416	11.59318	22.74109	0.0000
FLR	-2.231586	0.209947	-10.62927	0.0000
PGNP	-0.005647	0.002003	-2.818703	0.0065
R-squared	0.707665	Mean dependent var		141.5000
Adjusted R-squared	0.698081	S.D. dependent var		75.97807
S.E. of regression	41.74780	Akaike info criterion		10.34691
Sum squared resid	106315.6	Schwarz criterion		10.44811
Log likelihood	-328.1012	Hannan-Quinn criter.		10.38678
F-statistic	73.83254	Durbin-Watson stat		2.186159
Prob(F-statistic)	0.000000			

Nhận xét:

Dấu các hệ số hồi quy giống như kỳ vọng, cả 2 biến độc lập đều có hệ số thống kê có ý nghĩa ở mức 5% (giá trị P-value rất thấp).

- c. Khuyến nghị về mặt chính sách:

Do các biến FLFP và PGNP đều có tác động âm có ý nghĩa về mặt thống kê lên tỉ lệ tử vong của trẻ sơ sinh nên để giảm tỉ lệ tử vong của trẻ sơ sinh, các nhà hoạch định chính sách cần đưa ra những chính sách nhằm làm tăng các biến này. Nghĩa là cần nâng cao tỉ lệ biết chữ của lực lượng lao động nữ và tăng thu nhập bình quân đầu người.

- d. Thực hiện trình tự hồi quy đơn:

Bước 1: Tìm một hồi quy đơn giữa biến CM và FLFP

Dependent Variable: CM
Method: Least Squares
Date: 11/26/14 Time: 01:43
Sample: 1 64
Included observations: 64

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	263.8635	12.22499	21.58395	0.0000
FLR	-2.390496	0.213263	-11.20917	0.0000
R-squared	0.669590	Mean dependent var		141.5000
Adjusted R-squared	0.664261	S.D. dependent var		75.97807
S.E. of regression	44.02399	Akaike info criterion		10.43810
Sum squared resid	120163.0	Schwarz criterion		10.50556
Log likelihood	-332.0191	Hannan-Quinn criter.		10.46468
F-statistic	125.6455	Durbin-Watson stat		2.314744
Prob(F-statistic)	0.000000			

Đặt tên RESID_1=resid

Bước 2: Tìm hồi quy đơn của biến PGNP theo FLFP

Dependent Variable: PGNP
Method: Least Squares
Date: 11/26/14 Time: 01:47
Sample: 1 64
Included observations: 64

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-39.30328	734.9526	-0.053477	0.9575
FLR	28.14268	12.82111	2.195027	0.0319
R-squared	0.072108	Mean dependent var		1401.250
Adjusted R-squared	0.057142	S.D. dependent var		2725.696
S.E. of regression	2646.674	Akaike info criterion		18.63075
Sum squared resid	4.34E+08	Schwarz criterion		18.69821
Log likelihood	-594.1839	Hannan-Quinn criter.		18.65732
F-statistic	4.818143	Durbin-Watson stat		2.266593
Prob(F-statistic)	0.031915			

Đặt tên RESID_2=resid

Bước 3: Tìm hồi quy đơn của RESID_1 với RESID_2 không có hệ số tung độ gốc.

Dependent Variable: RESID_1
Method: Least Squares
Date: 11/26/14 Time: 01:50
Sample: 1 64
Included observations: 64

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2.26E-06	5.176220	-4.37E-07	1.0000
RESID_2	-0.005647	0.001987	-2.841715	0.0061
R-squared	0.115238	Mean dependent var		-2.50E-06
Adjusted R-squared	0.100968	S.D. dependent var		43.67320
S.E. of regression	41.40976	Akaike info criterion		10.31566
Sum squared resid	106315.6	Schwarz criterion		10.38313
Log likelihood	-328.1012	Hannan-Quinn criter.		10.34224
F-statistic	8.075343	Durbin-Watson stat		2.186160
Prob(F-statistic)	0.006066			

Hệ số ước lượng từ kết quả này (-0.0056) là tác động biên của PGNP lên CM

So sánh kết quả tác động biên PGNP lên CM

2 giá trị thể hiện tác động biên của biến PGNP lên biến CM ở câu b và câu d là như nhau

Bước 1: Phần dư RESID_1 chỉ bao gồm những yếu tố ảnh hưởng đến CM ngoài tác động của FLFP/ đã loại trừ tác động của biến FLFP ra khỏi phần dư

Bước 2: Phần dư RESID_2 chỉ bao gồm các yếu tố tác động lên PGNP ngoài tác động của biến FLFP/ đã loại trừ tác động của biến FLFP ra khỏi phần dư

Bước 3: Hệ số của hồi quy RESID_2 lên RESID_1 là tác động của PGNP lên CM (do cả 2 biến RESID đều đã loại bỏ tác động của FLFP ra khỏi mô hình rồi)