

Hướng Dẫn Giải Đề Thi Kinh Tế Lượng Lần 1 & Các Vấn Đề Liên Quan

Mục Lục

(Bấm giữ Ctrl rồi bấm chuột vô mục cần xem để đến nhanh mục đó)

PHẦN LÝ THUYẾT	2
1. Mô hình và ý nghĩa các hệ số	2
2. Kiểm định giả thiết	2
3. Đa cộng tuyến	2
4. Công dụng của các loại kiểm định:	3
PHẦN BÀI TẬP	3
Đề 1	3
Câu 1:	3
Câu 2:	4
Đề 2	7
Câu 1:	7
Câu 2:	8

Phần Lý Thuyết

Theo mình thì câu hỏi bài tập thì có thể đề thi lần 2 sẽ cho dạng tương tự, chứ còn câu hỏi lý thuyết thì không thể nào hỏi lại rồi. Vậy mình không trả lời mấy câu hỏi đó ra đây nha (thực ra mình cũng không biết trả lời đâu he). Một số vấn đề mà đề có thể ra là mấy cái này:

1. Mô hình và ý nghĩa các hệ số

a) Mô hình: mô hình tuyến tính theo biến là mô hình mà biến ở dạng bậc nhất, mô hình tuyến tính ở tam số là mô hình mà tham số ở dạng bậc nhất, ví dụ:

Mô hình tuyến tính theo biến $Y = a + b^2X + c^3T$.

Mô hình tuyến tính theo tham số: $Y = a + b \cdot \log X + c\sqrt{T}$

b) Ý nghĩa các hệ số hồi quy trong mô hình tuyến tính theo biến: các tham số đứng riêng lẻ thì được gọi là hệ số chặn hay hệ số tự do. Các tham số đi liền với các biến gọi là hệ số góc. Ví dụ trong mô hình $Y = a + bX$, a là giá trị của Y khi X và T bằng 0, b là độ thay đổi của Y khi X tăng 1 đơn vị.

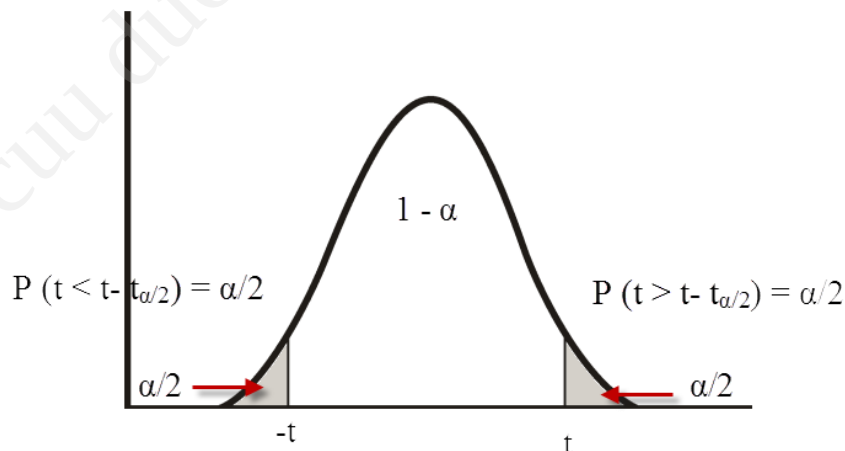
2. Kiểm định giả thiết

Ví dụ có một đại lượng thống kê K cần kiểm định, khi đó, với xác suất α cho trước (thường gọi là mức ý nghĩa), ta sẽ tìm được các hệ số K_1 và K_2 sao cho xác suất $P(K_2 < K < K_1) = 1 - \alpha$. Hay:

$$P(K < K_1) = P(K > K_2) = \alpha/2$$

Do kiểm định tham số thường dùng kiểm định t nên ghi thế này cho dễ hiểu:

$$P(t < -t_{\alpha/2}) = P(t > t_{\alpha/2}) = \alpha/2$$



Mức ý nghĩa α nghĩa là khả năng phạm phải sai lầm (xác suất $t_{kđ} < -t$ hoặc $t_{kđ} > t$). Ví dụ nếu chọn $\alpha = 0,05$ thì ta chỉ có 5% cơ hội là kết luận sai. Khoảng nằm giữa $-t$ và t gọi là khoảng tin cậy, với $\alpha = 0,05$ thì khả năng xảy ra chính xác (xác suất $-t < t_{kđ} < t$) lên tới 95% (gọi là độ tin cậy).

3. Đa cộng tuyến

Giả sử có mô hình hồi quy $Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3$. Nếu xảy ra hiện tượng các biến độc lập có mối quan hệ tuyến tính (ví dụ: $X_1 = a_0X_2 + a_1X_3$) thì mô hình đã xảy ra đa cộng tuyến. Trường hợp các

biến độc lập có quan hệ phi tuyến (ví dụ: $X_1 = a_0X_2^2 + a_1X_3^4$) thì... chưa thấy tài liệu nào nói hết ^^). Tuy nhiên, với mô hình dạng mô hình hàm tổng chi phí: $TC = \beta_1 + \beta_2q + \beta_3q^2 + \beta_4q^3$ thì ta có thể kết luận thẳng là không có đa cộng tuyến, vì tất cả các biến độc lập ở đây đều là từ một biến q mà ra cả.

4. Công dụng của các loại kiểm định:

Kiểm định White: kiểm tra hiện tượng phương sai thay đổi.

Kiểm định BG: kiểm tra hiện tượng tự tương quan.

Kiểm định Wald: kiểm định loại bỏ các biến không cần thiết.

Phần Bài Tập

Đề 1

Câu 1:

Trong một nghiên cứu về các yếu tố ảnh hưởng đến giá nhà, từ 14 quan sát người ta thu được mô hình sau:

$$Y = 52,351 - 0,264X - 21,345D + 1,649(XD)$$

$$\text{t-statistic} \quad (1,404) \quad (-0,006) \quad (-0,498) \quad (2,639)$$

$$R^2 = 0,836$$

Trong đó:

Y: Giá bán nhà (triệu đồng/m²)

X: Số phòng tắm trong nhà (phòng)

D: bằng 1 nếu nhà có phòng tập thể thao, bằng 0 nếu nhà không có phòng tập thể thao

a. Với mức ý nghĩa thống kê 1%. Bạn hãy kiểm định xem liệu số lượng phòng tắm có thực sự ảnh hưởng tới giá bán nhà hay không?

b. Nêu ý nghĩa của từng hệ số hồi quy gắn với biến giả? Theo bạn sự tồn tại của biến tương tác trong mô hình có cần thiết không? Giải thích kết luận của bạn với xác suất 99%?

c. Hệ số xác định mô hình nói lên điều gì?

$$\text{Với } t_{10}^{0,01} = 2,764$$

Hướng dẫn giải:

Ta đặt như sau:

$$Y = 52,351 - 0,264X - 21,245D + 1,649(XD)$$

$$\Leftrightarrow \bar{Y} = \bar{\beta}_0 + \bar{\beta}_1X + \bar{\beta}_2D + \bar{\beta}_3(XD)$$

a) Ở câu này, bạn dùng phương pháp kiểm định tham số:

Ngoài lề:

Kiểm định tham số nghĩa là kiểm tra xem các hệ số hồi quy có có ý nghĩa thống kê hay không, nói cách khác là chúng có khác 0 hay không. Giả sử mô hình của ta sẽ có hình thức như thế này: $Y = \beta_1 + \beta_2X$.

Đặt giả thiết $H_0: \beta_2=0$ (hệ số không có ý nghĩa thống kê, X chả liên can gì đến Y), $H_1: \beta_2 \neq 0$

Nếu $t_{kd} \geq t_{\alpha}^{(n-k)}$, bác bỏ H_0

Nếu $p\text{-value} \leq \alpha$, bác bỏ H_0

Người ta thường đặt H_0 trái với giả thiết cần kiểm chứng.

Theo đề bài, kiểm định số lượng phòng tắm (X) có ảnh hưởng đến giá bán nhà (Y) thực sự hay không, nghĩa là kiểm định hệ số góc đi liền với X . Ở đây là -0,264.

Vậy ta kiểm định như sau:

Đặt $H_0: \beta_1 = 0$, $H_1: \beta_1 \neq 0$

Do $|t_{kd}(\beta_1)| = 0,006 < t_{10}^{0,01} = 2,764$ nên ta chấp nhận H_0 , nghĩa là số lượng phòng tắm không ảnh hưởng đến giá bán nhà với độ tin cậy là 99%

b) Ý nghĩa các hệ số hồi quy gắn với biến giả:

$\beta_2 = -21,345$: chênh lệch về giá bán nhà giữa nhà có phòng tập thể thao với nhà không có phòng tập thể thao nếu số phòng tắm trong nhà là 0.

$\beta_3 = 1,649$: chênh lệch về mức độ tăng giá bán nhà giữa nhà có phòng tập thể thao với nhà không có phòng tập thể thao khi số phòng tắm trong nhà tăng thêm 1.

Kiểm định sự tồn tại của biến tương tác trong mô hình (biến XD):

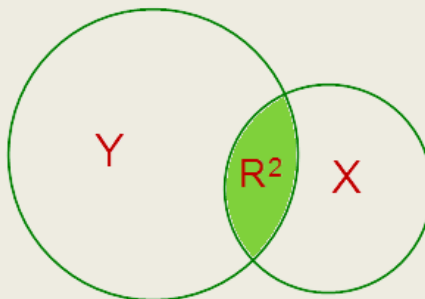
Tương tự câu a, bạn dùng phương pháp kiểm định tham số với β_3 , đặt $H_0: \beta_3 = 0$, không cần thiết có biến tương tác trong mô hình và $H_1: \beta_3 \neq 0$

Do $|t_{kd}(\beta_3)| = 2,639 < t_{10}^{0,01} = 2,764$ nên ta chấp nhận H_0 , vậy với độ tin cậy 99% ta có thể kết luận là biến tương tác không cần thiết có trong mô hình.

c) Hệ số xác định $R^2 = 0,836$ nói lên tỷ lệ phần biến động của giá bán nhà được giải thích từ các đối tượng nghiên cứu là 83,6%

Ngoài lề:

R^2 là tỷ lệ phần biến động của Y được giải thích từ X , là sự phụ thuộc của Y vào X , mức độ giải thích của X cho Y . R^2 có thể được hiểu dễ dàng hơn bằng hình minh họa như sau:



Câu 2:

Để nghiên cứu các yếu tố ảnh hưởng tới tỉ lệ nghèo đói, người ta sử dụng 58 quan sát thiết lập mô hình hồi quy như sau:

Dependent Variable: Y

Included observations: 58

Variable	Coefficient	Std. Error	t-statistic	Prob.
C	19.172	7.826		0.017
X2	5.418	1.758		0.003
X3	-0.138	0.062		0.031
X4	-0.191	0.089		0.033
R-squared		0.881	F-statistic	
			Prob(F-statistic)	0.000

Trong đó:

Y: tỷ lệ hộ nghèo (%)

X₂: số nhân khẩu/hộ (người)

X₃: tỷ lệ dân số tốt nghiệp cao đẳng, đại học (%)

X₄: thu nhập trung bình của hộ gia đình (nghìn đô-la)

a. Giải thích ý nghĩa của các hệ số hồi quy góc? Dấu của các hệ số hồi quy góc có phù hợp với những gì bạn thấy trên thực tế hay không? Tại sao?

b. Với mức ý nghĩa thống kê 5%, hãy tìm khoảng ước lượng cho các hệ số hồi quy của X₂, X₃, và X₄?

c. Có hiện tượng đa cộng tuyến xảy ra với mô hình trên hay không? Giải thích kết luận của bạn với mức độ tin cậy 95%?

Với $t_{54}^{1,674} = 1,674$

Hướng dẫn giải:

Ta đặt như sau:

$$Y = 19.172 + 5.418X_2 - 0.138X_3 - 0.191X_4$$

$$\Leftrightarrow \bar{Y} = \bar{\beta}_0 + \bar{\beta}_1X_2 + \bar{\beta}_2X_3 + \bar{\beta}_3X_4$$

a) Ý nghĩa các hệ số hồi quy góc:

$\bar{\beta}_1 = 5.418$: tỷ lệ hộ nghèo tăng thêm khi số nhân khẩu trong gia đình tăng thêm 1 người. Trong điều kiện các yếu tố khác không đổi.

$\bar{\beta}_2 = -0.138$: tỷ lệ hộ nghèo giảm xuống khi tỷ lệ dân số tốt nghiệp cao đẳng, đại học tăng 1%. Trong điều kiện các yếu tố khác không đổi.

$\bar{\beta}_3 = -0.191$: tỷ lệ hộ nghèo giảm xuống khi thu nhập trung bình của hộ gia đình tăng 1 nghìn đô-la. Trong điều kiện các yếu tố khác không đổi.

b) Tìm khoảng ước lượng (tin cậy) của các hệ số hồi quy góc:

Ngoài lề:

Công thức tìm khoảng ước lượng (tin cậy) cho các tham số:

$$\bar{\beta}_j - t_{\alpha}^{(n-k)} Se(\hat{\beta}_j) \leq \beta_j \leq \bar{\beta}_j + t_{\alpha}^{(n-k)} Se(\hat{\beta}_j)$$

Với mức ý nghĩa thống kê 5% ta có $t_{54}^{1,674} = 1,674$:

Khoảng ước lượng cho β_1 (hệ số hồi quy góc của X_2):

$$\begin{aligned} \hat{\beta}_1 - t_{54}^{1,674} \text{Se}(\hat{\beta}_1) &\leq \beta_1 \leq \hat{\beta}_1 + t_{54}^{1,674} \text{Se}(\hat{\beta}_1) \\ \Leftrightarrow 5,418 - 1,674 \times 1,758 &\leq \beta_1 \leq 5,418 + 1,674 \times 1,758 \\ \Leftrightarrow 2,475 &\leq \beta_1 \leq 8,361 \end{aligned}$$

Khoảng ước lượng cho β_2 (hệ số hồi quy góc của X_3):

$$\begin{aligned} \hat{\beta}_2 - t_{54}^{1,674} \text{Se}(\hat{\beta}_2) &\leq \beta_2 \leq \hat{\beta}_2 + t_{54}^{1,674} \text{Se}(\hat{\beta}_2) \\ \Leftrightarrow -0,138 - 1,674 \times 0,062 &\leq \beta_2 \leq -0,138 + 1,674 \times 0,062 \\ \Leftrightarrow -0,2417 &\leq \beta_2 \leq 0,034 \end{aligned}$$

Khoảng ước lượng cho β_3 (hệ số hồi quy góc của X_4):

$$\begin{aligned} \hat{\beta}_3 - t_{54}^{1,674} \text{Se}(\hat{\beta}_3) &\leq \beta_3 \leq \hat{\beta}_3 + t_{54}^{1,674} \text{Se}(\hat{\beta}_3) \\ \Leftrightarrow -0,191 - 1,674 \times 0,089 &\leq \beta_3 \leq -0,191 + 1,674 \times 0,089 \\ \Leftrightarrow 0,339 &\leq \beta_3 \leq 0,042 \end{aligned}$$

c) Từ bảng trên ta tính được t-statistic như sau:

Coefficient	Std. Error	t-statistic
19.172	7.826	2.449
5.418	1.758	3.082
-0.138	0.062	-2.225
-0.191	0.089	-2.146

Với mức ý nghĩa 5%, ta thấy các giá trị |t-statistic| đều lớn hơn giá trị $t_{54}^{1,674} = 1,674$ nghĩa là tất cả các hệ số đều khác không một cách có ý nghĩa thống kê, hệ số xác định $R^2 = 0.881$ khá cao và p-value(F-statistic) = 0,000 < $\alpha = 0,05$ nên R^2 cũng đáng tin cậy. Vậy mô hình là phù hợp và không có hiện tượng đa cộng tuyến xảy ra, độ tin cậy là 95%.

Ngoài lề:

Có 3 cách để phát hiện đa cộng tuyến, đó là:

1. Hệ số xác định R^2 cao nhưng tỷ số |t| thấp: khi R^2 cao, kinh nghiệm cho thấy nếu $R^2 > 0,8$ thì giả thiết các hệ số hồi quy đồng thời bằng 0 bị bác bỏ. Tuy nhiên tỷ số |t| thấp thì ta chấp nhận giả thiết các hệ số hồi quy bằng 0. Khi mâu thuẫn này xảy ra, người ta thường nghĩ đến hiện tượng đa cộng tuyến.
2. Hệ số tương quan giữa các biến độc lập cao: khi hệ số tương quan giữa các biến độc lập lớn hơn 0,8 thì ta có thể chắc chắn có đa cộng tuyến.
3. Dùng hồi quy phụ: Giả sử có mô hình hồi quy như sau: $Y = \beta_0 + \beta_1 X + \beta_2 Y + \beta_3 Z + U$. Ta ước lượng mô hình hồi quy phụ giữa biến độc lập X với các biến độc lập còn lại như sau: $X = \alpha_0 + \alpha_1 Y + \alpha_2 Z + U$. Tiếp theo ta dùng phương pháp kiểm định tham số để kiểm tra sự ảnh hưởng của biến Y và Z đến X (xem câu a, bài 1 đề 1).

(*) Trong 3 cách này, 2 cách đầu chỉ là tương đối, chỉ có cách thứ 3 mới có thể đi đến kết luận chắc chắn nhất. Nhưng theo dữ kiện đề bài cho, ta chỉ có thể dùng cách 1 mà thôi.

(**) Thực ra không cần phải tính t-statistic, bạn có thể dùng p-value. Chỉ là vấn đề kiểm định tham số mà.

Đề 2

Câu 1:

Trong một nghiên cứu về các yếu tố ảnh hưởng đến sản lượng ô tô bán ra, từ 40 quan sát người ta thu được mô hình như sau:

$$Y = -31,852 - 1,751X + 0,092D + 4,698(XD)$$

$$t\text{-statistic} \quad (-5,8) \quad (-7,6) \quad (2,8) \quad (-3,969)$$

$$R^2 = 0,787$$

Trong đó:

Y: số ô tô bán ra/người (chiếc)

X: chỉ số giá xe hơi

D: bằng 1 nếu là mùa xuân, bằng 0 mùa khác

a. Với mức ý nghĩa thống kê 1%, bạn hãy kiểm định xem liệu chỉ số giá xe hơi có ảnh hưởng tới số ô tô bán ra/người hay không?

b. Nêu ý nghĩa của từng hệ số hồi quy gắn với biến giả? Có người cho rằng vào mùa xuân, khi chỉ số giá xe hơi tăng 1 đơn vị thì số ô tô bán ra tăng 2,5 chiếc/người có đúng hay không? Bạn hãy giải thích kết luận của mình với độ tin cậy 99%?

c. Hệ số xác định mô hình nói lên điều gì?

$$\text{Với } t_{36}^{0,01} = 2,434$$

Hướng dẫn giải:

Ta đặt như sau:

$$Y = -31,852 - 1,751X + 0,092D + 4,698(XD)$$

$$\Leftrightarrow \bar{Y} = \bar{\beta}_0 + \bar{\beta}_1 X + \bar{\beta}_2 D + \bar{\beta}_3 (XD)$$

a) Tương tự câu a, bài 1 đề 1, ta đặt giả thiết $H_0: \bar{\beta}_1 = 0$, $H_1: \bar{\beta}_1 X \neq 0$. Với mức ý nghĩa thống kê 1%, do $|\text{tkđ}(\bar{\beta}_1)| = 7.6 > t_{36}^{0,01} = 2,434$ nên ta bác bỏ H_0 , vậy chỉ số giá xe có ảnh hưởng đến lượng ô tô bán ra/người.

b) Ý nghĩa của hệ số hồi quy gắn với biến giả:

$\bar{\beta}_2 = 0.092$: chênh lệch về số ô tô bán ra/người vào mùa xuân so với các mùa khác khi chỉ số giá xe bằng 0.

$\bar{\beta}_3 = 4,698$: chênh lệch về mức độ tăng số lượng ô tô bán ra/người vào mùa xuân so với các mùa khác khi chỉ số giá xe tăng thêm 1 đơn vị.

Kiểm định hệ số hồi quy của biến tương tác: bó tay ^^

c) Hệ số xác định $R^2 = 0,787$ nói lên tỷ lệ phần biến động của giá bán nhà được giải thích từ các đối tượng nghiên cứu là 78,7% (xem lại câu c, bài 1 đề 1).

Câu 2:

Để nghiên cứu các yếu tố ảnh hưởng tới tỉ lệ nghèo đói, người ta sử dụng 58 quan sát thiết lập mô hình hồi quy như sau:

Dependent Variable: Y
Included observations: 58

Variable	Coefficient	Std. Error	t-statistic	Prob.
C	19.172	7.826		0.017
X2	5.418	1.758		0.003
X3	-0.138	0.062		0.031
X4	-0.191	0.089		0.033
R-squared		0.881	F-statistic	
			Prob(F-statistic)	0.000

Trong đó:

Y: tỷ lệ hộ nghèo (%)

X₂: số nhân khẩu/hộ (người)

X₃: tỷ lệ dân số tốt nghiệp cao đẳng, đại học (%)

X₄: thu nhập trung bình của hộ gia đình (nghìn đô-la)

a. Giải thích ý nghĩa của các hệ số hồi quy góc? Dấu của các hệ số hồi quy góc có phù hợp với những gì bạn thấy trên thực tế hay không? Tại sao?

b. Với mức ý nghĩa thống kê 5%, hãy tìm khoảng ước lượng cho các hệ số hồi quy của X₂, X₃, và X₄?

c. Có người cho rằng với xác suất 95%, khi nhân khẩu của hộ tăng thêm 1 người làm cho tỷ lệ hộ nghèo tăng lên 3%? Bạn cho ý kiến này có thỏa đáng hay không?

Với $t_{54}^{1.674} = 1.674$

Hướng dẫn giải:

Ta đặt như sau:

$$Y = 19.172 + 5.418X_2 - 0.138X_3 - 0.191X_4$$

$$\Leftrightarrow \bar{Y} = \bar{\beta}_0 + \bar{\beta}_1X_2 + \bar{\beta}_2X_3 + \bar{\beta}_3X_4$$

a) Là câu a bài 2 đề 1.

b) Là câu b bài 2 đề 1.

c) Đặt giả thiết $H_0: \bar{\beta}_1 = 3, H_1: \bar{\beta}_1 \neq 3$.

$$\text{Ta có } t_{kd}(\bar{\beta}_1) = \frac{\bar{\beta}_1 - 3}{\text{Se}(\bar{\beta}_1)} = \frac{5.418 - 3}{1.758} = 1.375$$

Với xác suất 95%, do $t_{kd}(\bar{\beta}_1) = 1.375 < t_{54}^{1.674} = 1.674$ nên ta chấp nhận H_0 . Vậy giả thiết khi nhân khẩu của hộ tăng thêm 1 người thì tỷ lệ hộ nghèo tăng lên 3% là thỏa đáng.