

BÀI TẬP ÔN THI

KINH TẾ LƯỢNG

Biên Soạn

ThS. LÊ TRƯỜNG GIANG

Thành phố Hồ Chí Minh, ngày 20, tháng 06, năm 2016

Mục lục

	Trang
Chương 1 Tóm tắt lý thuyết	1
1.1 Tổng quan về kinh tế lượng	1
1.1.1 Các khái niệm mở đầu	1
1.1.2 Khái niệm về hồi quy và phân tích hồi quy	2
1.2 Mô hình hồi quy đơn	3
1.2.1 Một số công thức xác suất	3
1.2.2 Xác định hàm hồi quy mẫu (SRF)	4
1.2.3 Hệ số xác định mô hình	4
1.2.4 Phương sai của sai số ngẫu nhiên	4
1.2.5 Bài toán ước lượng	5
1.2.6 Bài toán kiểm định	5
1.2.7 Bài toán dự báo	6
1.2.8 Một số lưu ý	7
1.3 Mô hình hồi quy bội	8
1.3.1 Xác định hàm hồi quy mẫu (SRF)	10
1.3.2 Hệ số xác định mô hình	11
1.3.3 Ma trận tự tương quan	11
1.3.4 Ma trận hiệp phương sai	11
1.3.5 Các bài toán	11
1.4 Hồi quy với biến giả	12
1.4.1 Khái niệm	12
1.4.2 Ý nghĩa	12
1.4.3 So sánh hai mô hình	12
1.5 Kiểm định giả thiết mô hình	14
1.5.1 Đa cộng tuyến	14
1.5.2 Phương sai thay đổi	16
1.5.3 Tự tương quan	17
Chương 2 Bài tập ứng dụng	19
2.1 Mô hình hồi quy hai biến	19

2.2	Mô hình hồi quy bội	32
2.3	Hồi quy với biến định tính	47
2.4	Bài tập tổng hợp	55
2.5	Bài tập đề nghị	68
2.6	Đề thi tham khảo	82
2.6.1	Đề 1	82
2.6.2	Đề 2	84
2.6.3	Đề 3	85
2.6.4	Đề 4	88
2.6.5	Đề 5	93
2.6.6	Đề 6	97
2.6.7	Đề 7	100

Chương 3 Thực hành Eviews 103

3.1	Cài đặt Eviews 8	103
3.2	Khởi động Eviews 8	104
3.3	Nhập dữ liệu cho Eviews 8	104
3.4	Thống kê mô tả	104
3.5	Ước lượng các hệ số của mô hình hồi quy	105
3.6	Kiểm định sự vi phạm các giả thiết của mô hình hồi quy	107
3.6.1	Hiện tượng đa cộng tuyến	107
3.6.2	Phương sai của sai số ngẫu nhiên thay đổi	107
3.6.3	Hiện tượng tự tương quan	108
3.6.4	Kiểm định biến có cần thiết trong mô hình hay không (kiểm định Wald)	108
3.6.5	Kiểm định biến bị bỏ sót trong mô hình	108
3.6.6	Sai số ngẫu nhiên không có phân phối chuẩn	109
3.6.7	Kiểm định Chow trong mô hình hồi quy với biến giả	109
3.7	Dự báo bằng mô hình hồi quy	110

Tài liệu tham khảo 123

Chương 1

Tóm tắt lý thuyết

1.1 Tổng quan về kinh tế lượng

1.1.1 Các khái niệm mở đầu

Khái niệm về kinh tế lượng

Thuật ngữ kinh tế lượng được giáo sư kinh tế học người Na-Uy là A. K. Ragnar Frisch, giải thưởng Nobel về kinh tế học (1969), sử dụng lần đầu tiên vào khoảng những năm 1930. Sau đây là một số khái niệm cho thuật ngữ này

- Econometrics = Econo + Metrics = "Đo lường kinh tế" = "Kinh tế lượng".
- Kinh tế lượng là môn khoa học sử dụng các công cụ toán học để củng cố về mặt thực nghiệm cho các lý thuyết kinh tế.
- Kinh tế lượng là một công cụ kết hợp giữa lý thuyết kinh tế hiện đại, thống kê toán và máy tính nhằm định lượng (đo lường) các mối quan hệ kinh tế, từ đó dự báo diễn biến các hiện tượng kinh tế và phân tích các chính sách kinh tế.
- Kinh tế lượng có thể được định nghĩa như là sự phân tích về lượng các vấn đề kinh tế hiện thời dựa trên việc vận dụng đồng thời lý thuyết và thực tế được thực hiện bằng các phương pháp suy đoán thích hợp.

Sơ đồ tổng quan về kinh tế lượng

Phân tích kinh tế lượng được thực hiện theo các bước sau đây

1. Nêu ra giả thiết

2. Thiết lập mô hình toán học
3. Thu thập số liệu
4. Ước lượng tham số
5. Phân tích kết quả
6. Dự báo
7. Ra quyết định

1.1.2 Khái niệm về hồi quy và phân tích hồi quy

Khái niệm

- Hồi quy là một công cụ cơ bản của đo lường kinh tế. Thuật ngữ "Hồi quy" đã được Francis Galton sử dụng vào năm 1886.
- Phân tích hồi quy nghiên cứu mối liên hệ phụ thuộc của một biến (gọi là biến phụ thuộc hay biến được giải thích) với một hay nhiều biến khác (được gọi là biến độc lập hay biến giải thích) nhằm ước lượng, dự báo giá trị trung bình của biến phụ thuộc với các giá trị đã biết của biến độc lập.
- Phân tích hồi quy giải quyết các vấn đề sau đây:
 1. Ước lượng giá trị trung bình của biến phụ thuộc với giá trị đã cho của biến độc lập
 2. Kiểm định giả thiết về bản chất của sự phụ thuộc.
 3. Dự đoán giá trị trung bình của biến phụ thuộc khi biết giá trị của các biến độc lập.
 4. Kết hợp các vấn đề trên.

Hàm hồi quy tổng thể (PRF)

$$PRF : E(Y/X_i) = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki}$$

$$PRM : Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + U_i \quad (U_i = Y_i - E(Y/X_i))$$

Trong đó:

- β_j là các hệ số hồi quy;
- β_1 là hệ số tự do;
- β_2, β_3, \dots là các hệ số hồi quy riêng;
- U_i là sai số ngẫu nhiên.

Hàm hồi quy mẫu (SRF)

$$SRF : \hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \dots + \hat{\beta}_k X_{ki}$$

$$SRM : Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \dots + \hat{\beta}_k X_{ki} + e_i \quad (e_i = Y_i - \hat{Y}_i)$$

Trong đó:

- \hat{Y}_i là ước lượng điểm của $E(Y/X_i)$
- $\hat{\beta}_j$ là các ước lượng điểm của β_j ;
- e_i là ước lượng điểm của U_i .

1.2 Mô hình hồi quy đơn

1.2.1 Một số công thức xác suất

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

$$Var(X) = E(X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} \left(\sum_{i=1}^n X_i^2 - n\bar{X}^2 \right)$$

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$$

$$Var(Y) = E(Y_i - \bar{Y})^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2 = \frac{1}{n} \left(\sum_{i=1}^n Y_i^2 - n\bar{Y}^2 \right)$$

$$Cov(X, Y) = E[(X - \bar{X})(Y - \bar{Y})] = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) = E(XY) - \bar{X}\bar{Y}$$

$$r_{XY} = \frac{Cov(X, Y)}{se(X) se(Y)}$$

1.2.2 Xác định hàm hồi quy mẫu (SRF)

$$\hat{\beta}_2 = \frac{Cov(X, Y)}{Var(X)} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{\sum_{i=1}^n X_i Y_i - n\bar{X}\bar{Y}}{\sum_{i=1}^n X_i^2 - n\bar{X}^2}$$
$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}$$

Vậy

$$\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X$$

Chú ý:

- β_1, β_2 là các hệ số hồi quy;
- β_1 là hệ số tự do (hệ số tung độ gốc);
- β_2 là hệ số góc (hệ số độ dốc).

1.2.3 Hệ số xác định mô hình

$$TSS = \sum_{i=1}^n (Y_i - \bar{Y})^2 = n \text{var}(Y)$$

$$ESS = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 = n \hat{\beta}_2^2 \text{var}(X)$$

$$RSS = TSS - ESS = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = n(1 - r^2) \text{Var}(Y)$$

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

1.2.4 Phương sai của sai số ngẫu nhiên

Ta ký hiệu phương sai của sai số ngẫu nhiên là $Var(U) = \sigma^2$. Khi đó phương sai của sai số ngẫu nhiên trên mẫu được xác định như sau:

$$\hat{\sigma}^2 = \frac{n}{n-2} (1 - r^2) \text{Var}(Y) = \frac{RSS}{n-2}$$

Trong trường hợp tổng quát ta có

$$\hat{\sigma}^2 = \frac{RSS}{n-k}$$

Với k là số biến trong mô hình (tính cả biến phụ thuộc).

1.2.5 Bài toán ước lượng

1. Ước lượng các hệ số hồi quy

Ta có

$$\text{var}(\hat{\beta}_1) = \left[\frac{1}{n} + \frac{\bar{X}^2}{n\text{var}(X)} \right] \cdot \hat{\sigma}^2 \Rightarrow \text{se}(\hat{\beta}_1) = \sqrt{\text{var}(\hat{\beta}_1)}$$

$$\text{var}(\hat{\beta}_2) = \frac{\hat{\sigma}^2}{n\text{var}(X)} \Rightarrow \text{se}(\hat{\beta}_2) = \sqrt{\text{var}(\hat{\beta}_2)}$$

Loại ước lượng	Khoảng ước lượng	giá trị C
Hai phía	$\hat{\beta}_j - C \text{se}(\hat{\beta}_j) \leq \beta_j \leq \hat{\beta}_j + C \text{se}(\hat{\beta}_j)$	$C = t_{\frac{\alpha}{2}}^{(n-k)}$
Tối đa (pt)	$\beta_j \leq \hat{\beta}_j + C \text{se}(\hat{\beta}_j)$	$C = t_{\alpha}^{(n-k)}$
Tối thiểu (pp)	$\hat{\beta}_j - C \text{se}(\hat{\beta}_j) \leq \beta_j$	$C = t_{\alpha}^{(n-k)}$

2. Ước lượng phương sai của sai số ngẫu nhiên

$$\sigma^2 \in \left[\frac{(n-2)\hat{\sigma}^2}{\chi_{\frac{\alpha}{2}}^2(n-2)}; \frac{(n-2)\hat{\sigma}^2}{\chi_{1-\frac{\alpha}{2}}^2(n-2)} \right]$$

Trong trường hợp tổng quát

$$\sigma^2 \in \left[\frac{(n-k)\hat{\sigma}^2}{\chi_{\frac{\alpha}{2}}^2(n-k)}; \frac{(n-k)\hat{\sigma}^2}{\chi_{1-\frac{\alpha}{2}}^2(n-k)} \right]$$

Với k là số biến trong mô hình (tính cả biến phụ thuộc).

1.2.6 Bài toán kiểm định

1. Kiểm định dạng t

Loại kiểm định	Giả thiết $H_0 : \beta_j = \beta_j^*$; đối thiết H_1	Bác bỏ H_0
Kiểm định 2 phía	$H_1 : \beta_j \neq \beta_j^*$	$ t > t_{\frac{\alpha}{2}}^{(n-k)}$ $P_{value} < \alpha$
Kiểm định bên trái	$H_1 : \beta_j < \beta_j^*$	$t < -t_{\alpha}^{(n-k)}$ $\frac{P_{value}}{2} < \alpha$
Kiểm định bên phải	$H_1 : \beta_j > \beta_j^*$	$t > t_{\alpha}^{(n-k)}$ $\frac{P_{value}}{2} < \alpha$
Lưu ý	$t = \frac{\widehat{\beta}_j - \beta_j^*}{se(\widehat{\beta}_j)}$	

2. Kiểm định dạng χ^2

Loại kiểm định	Gt $H_0 : \sigma^2 = \sigma_0^2$; Dt H_1	Bác bỏ H_0
Hai phía	$H_1 : \sigma^2 \neq \sigma_0^2$	$\chi^2 > \chi_{\frac{\alpha}{2}}^2(n-k)$ or $\chi^2 < \chi_{1-\frac{\alpha}{2}}^2(n-k)$
Bên trái	$H_1 : \sigma^2 < \sigma_0^2$	$\chi^2 < \chi_{1-\alpha}^2(n-k)$
Bên phải	$H_1 : \sigma^2 > \sigma_0^2$	$\chi^2 > \chi_{\alpha}^2(n-k)$
Lưu ý	$\chi^2 = \frac{(n-k)\widehat{\sigma}^2}{\sigma_0^2}$	

3. Kiểm định dạng F

Bước 1	Đặt giả thiết $H_0 : R^2 = 0$; $H_1 : R^2 \neq 0$
Bước 2	$C = F_{\alpha}(k-1; n-k)$
Bước 3	$F = \frac{(n-k)R^2}{(k-1)(1-R^2)}$
Bước 4	Nếu $F > C$ thì bác bỏ H_0 ; Nếu $F \leq C$ thì chưa có cơ sở bác bỏ H_0

1.2.7 Bài toán dự báo

Ta tính được các giá trị sau:

$$\widehat{Y}_0 = \widehat{\beta}_1 + \widehat{\beta}_2 X_0;$$

$$C = t_{\frac{\alpha}{2}}^{(n-k)}.$$

$$\text{var}(\hat{Y}_0) = \left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{n \text{var}(X)} \right] \cdot \hat{\sigma}^2 \Rightarrow \text{se}(\hat{Y}_0) = \sqrt{\text{var}(\hat{Y}_0)}$$

$$\text{var}(Y_0 - \hat{Y}_0) = \hat{\sigma}^2 + \text{var}(\hat{Y}_0) \Rightarrow \text{se}(Y_0 - \hat{Y}_0) = \sqrt{\text{var}(Y_0 - \hat{Y}_0)}$$

Khoảng dự báo cho giá trị trung bình $E(Y/X_0)$

$$\left[\hat{Y}_0 - C \text{se}(\hat{Y}_0); \hat{Y}_0 + C \text{se}(\hat{Y}_0) \right]$$

Khoảng dự báo cho giá trị cá biệt Y_0

$$\left[\hat{Y}_0 - C \text{se}(Y_0 - \hat{Y}_0); \hat{Y}_0 + C \text{se}(Y_0 - \hat{Y}_0) \right]$$

1.2.8 Một số lưu ý

Hệ số co dãn

$$\varepsilon_{\frac{Y}{X}} = \frac{dY}{dX} \frac{\bar{X}}{\bar{Y}} = \hat{\beta}_2 \frac{\bar{X}}{\bar{Y}}$$

Hệ số co dãn cho biết khi X tăng 1% thì Y sẽ tăng (hay giảm) bao nhiêu %.

Đổi đơn vị

$$\begin{cases} Y^* = k_1 Y \\ X^* = k_2 X \end{cases} \Rightarrow \begin{cases} \hat{\beta}_1^* = k_1 \hat{\beta}_1 \\ \hat{\beta}_2^* = \frac{k_1}{k_2} \hat{\beta}_2 \end{cases} \Rightarrow \hat{Y}_* = \hat{\beta}_1^* + \hat{\beta}_2^* X^*$$

Kiểm định P_{value}

+ α : mức ý nghĩa \rightarrow xác suất mắc sai lầm loại 1 \rightarrow xác suất bác bỏ giả thiết H_0 trong khi H_0 đúng.

$$\alpha = P(g \in W_\alpha / H_0 \text{ đúng})$$

+ P_{value} : mức xác suất nhỏ nhất mà tại đó giả thiết H_0 bị bác bỏ.

- $\alpha > P_{value}$: bác bỏ giả thiết H_0
- $\alpha \leq P_{value}$: chưa có cơ sở bác bỏ giả thiết H_0

Các dạng hàm đặc biệt

1. Hồi quy tuyến tính Logarit (log-log)

Hàm hồi quy mẫu (SRF): $\widehat{\ln Y_i} = \widehat{\beta}_1 + \widehat{\beta}_2 \ln X_i$

\Rightarrow Nếu X tăng lên 1% thì Y thay đổi một tỷ lệ là $\widehat{\beta}_2\%$

2. Hồi quy tuyến tính bán Logarit

+ Hàm hồi quy mẫu (SRF): $\widehat{\ln Y_i} = \widehat{\beta}_1 + \widehat{\beta}_2 X_i$

\Rightarrow Nếu X tăng lên 1 đơn vị thì Y thay đổi một lượng là $\widehat{\beta}_2 \cdot 100(\%)$

+ Hàm hồi quy mẫu (SRF): $\widehat{Y_i} = \widehat{\beta}_1 + \widehat{\beta}_2 \ln X_i$

\Rightarrow Nếu X tăng lên 1% thì Y thay đổi một lượng là $\widehat{\beta}_2 \cdot 0,01$

1.3 Mô hình hồi quy bội

Hàm hồi quy tổng thể (PRF) và mô hình hồi quy tổng thể (PRM):

$$PRF : E(Y/X_i) = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki}$$

$$PRM : Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + U_i \quad (U_i = Y_i - E(Y/X_i))$$

Giả sử ta có n quan sát, mỗi quan sát gồm k giá trị $(X_{2i}, X_{3i}, \dots, X_{ki}, Y_i), i = \overline{1, n}$.
Thay vào PRM, ta có

$$\begin{cases} Y_1 = \beta_1 + \beta_2 X_{21} + \dots + \beta_k X_{k1} + U_1 \\ Y_2 = \beta_1 + \beta_2 X_{22} + \dots + \beta_k X_{k2} + U_2 \\ \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \\ Y_n = \beta_1 + \beta_2 X_{2n} + \dots + \beta_k X_{kn} + U_n \end{cases}$$

Ta đặt

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_n \end{pmatrix}; U = \begin{pmatrix} U_1 \\ U_2 \\ \dots \\ U_n \end{pmatrix}; \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_k \end{pmatrix};$$

$$X = \begin{pmatrix} 1 & X_{21} & \dots & X_{k1} \\ 1 & X_{22} & \dots & X_{k2} \\ \dots & \dots & \dots & \dots \\ 1 & X_{2n} & \dots & X_{kn} \end{pmatrix};$$

Khi đó, mô hình có thể viết lại dưới dạng ma trận như sau

$$Y = X\beta + U$$

Hàm hồi quy mẫu (SRF) và mô hình hồi quy mẫu (SRM):

$$SRF : \hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \dots + \hat{\beta}_k X_{ki}$$

$$SRM : Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \dots + \hat{\beta}_k X_{ki} + e_i \quad (e_i = Y_i - \hat{Y}_i)$$

Mô hình được viết dưới dạng ma trận như sau

$$\hat{Y} = X\hat{\beta} + e$$

trong đó $\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \dots \\ \hat{\beta}_k \end{pmatrix}; e = \begin{pmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{pmatrix}$

Ta đi xác định các hệ số hồi quy bằng phương pháp OLS. Có nghĩa là các $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$ phải thỏa mãn $\frac{\partial (e^T e)}{\partial \hat{\beta}} = 0 \Leftrightarrow (X^T X) \hat{\beta} = X^T Y$. Suy ra

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

trong đó

$$X^T X = \begin{pmatrix} n & \sum_{i=1}^n X_{2i} & \dots & \sum_{i=1}^n X_{ki} \\ \sum_{i=1}^n X_{2i} & \sum_{i=1}^n X_{2i}^2 & \dots & \sum_{i=1}^n X_{2i} X_{ki} \\ \dots & \dots & \dots & \dots \\ \sum_{i=1}^n X_{ki} & \sum_{i=1}^n X_{ki} X_{2i} & \dots & \sum_{i=1}^n X_{ki}^2 \end{pmatrix};$$

$$X^T Y = \begin{pmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n X_{2i} Y_i \\ \dots \\ \sum_{i=1}^n X_{ki} Y_i \end{pmatrix}$$

Sau đây sẽ trình bày một số công thức cụ thể cho trường hợp hồi quy ba biến, đối với các hàm hồi quy có k biến thì việc tính toán khá phức tạp, khi đó chúng ta cần sử dụng một số phần mềm hỗ trợ như Eviews, Stata,...

1.3.1 Xác định hàm hồi quy mẫu (SRF)

$$X^T X = \begin{pmatrix} n & \sum X_{2i} & \sum X_{3i} \\ \sum X_{2i} & \sum X_{2i}^2 & \sum X_{2i} X_{3i} \\ \sum X_{3i} & \sum X_{2i} X_{3i} & \sum X_{3i}^2 \end{pmatrix}$$

$$(X^T Y) = \begin{pmatrix} \sum Y_i \\ \sum X_{2i} Y_i \\ \sum X_{3i} Y_i \end{pmatrix}$$

$$\hat{\beta} = (X^T X)^{-1} (X^T Y) = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix}$$

Vậy

$$\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3$$

1.3.2 Hệ số xác định mô hình

$$TSS = Y^T Y - n\bar{Y}^2 = \sum Y_i^2 - n\bar{Y}^2;$$

$$ESS = \hat{\beta}^T (X^T Y) - n\bar{Y}^2;$$

$$RSS = TSS - ESS = Y^T Y - \hat{\beta}^T (X^T Y);$$

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS};$$

$$\bar{R}^2 = 1 - (1 - R^2) \frac{n-1}{n-k}.$$

1.3.3 Ma trận tự tương quan

Giả sử chúng ta có mô hình hồi quy ba biến:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + U_i$$

Ký hiệu r_{tj} là hệ số tương quan giữa biến thứ t và biến thứ j . Nếu $t = 1$ thì r_{1j} là hệ số tương quan giữa biến Y và biến X_j . Ta có ma trận hệ số tương quan

$$R = \begin{pmatrix} 1 & r_{12} & r_{13} \\ r_{21} & 1 & r_{23} \\ r_{31} & r_{32} & 1 \end{pmatrix}$$

1.3.4 Ma trận hiệp phương sai

$$\hat{\sigma}^2 = \frac{RSS}{n-k}$$

$$Cov(\hat{\beta}) = \hat{\sigma}^2 (X^T X)^{-1} = \begin{pmatrix} Var(\hat{\beta}_1) & Cov(\hat{\beta}_1, \hat{\beta}_2) & Cov(\hat{\beta}_1, \hat{\beta}_3) \\ Cov(\hat{\beta}_2, \hat{\beta}_1) & Var(\hat{\beta}_2) & Cov(\hat{\beta}_2, \hat{\beta}_3) \\ Cov(\hat{\beta}_3, \hat{\beta}_1) & Cov(\hat{\beta}_3, \hat{\beta}_2) & Var(\hat{\beta}_3) \end{pmatrix}$$

1.3.5 Các bài toán

Việc giải quyết các bài toán ước lượng, kiểm định, dự báo được thực hiện tương tự như mục 1.2; sau đây là một số công thức cần chú ý:

$$Var(\hat{\beta}_2 - \hat{\beta}_3) = Var(\hat{\beta}_2) + Var(\hat{\beta}_3) - 2Cov(\hat{\beta}_2, \hat{\beta}_3)$$

$$Var(\hat{Y}_0) = \hat{\sigma}^2 (X^0)^T (X^T X)^{-1} X^0 \quad \Rightarrow \quad se(\hat{Y}_0) = \sqrt{Var(\hat{Y}_0)}$$

$$Var(Y_0 - \hat{Y}_0) = Var(\hat{Y}_0) + \hat{\sigma}^2 \quad \Rightarrow \quad se(Y_0 - \hat{Y}_0) = \sqrt{Var(Y_0 - \hat{Y}_0)}$$

1.4 Hồi quy với biến giả

1.4.1 Khái niệm

Biến giả là biến định tính, không đo được. Ví dụ: giới tính, màu sắc, khu vực,...

1.4.2 Ý nghĩa

- + Dùng để so sánh các phạm trù khác nhau trong mô hình (muốn so sánh m phạm trù ta sử dụng $m - 1$ biến giả, phạm trù gán giá trị 0 là phạm trù cơ sở).
- + Dùng để so sánh hai hàm hồi quy.
- + Phân tích mùa.

1.4.3 So sánh hai mô hình

Để kiểm định sự khác nhau của hai mô hình ta có 2 phương pháp:

1. Phương pháp kiểm định Chow

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i \rightarrow RSS$$

$$\hat{Y}_j = \hat{\lambda}_1 + \hat{\lambda}_2 X_j \rightarrow RSS_1$$

$$\hat{Y}_k = \hat{\gamma}_1 + \hat{\gamma}_2 X_k \rightarrow RSS_2$$

$$\overline{RSS} = RSS_1 + RSS_2$$

Các bước cho bài toán kiểm định

- + Đặt giả thiết

H_0 : hai mô hình là như nhau; H_1 : hai mô hình khác nhau

$$+ F = \frac{(RSS - \overline{RSS})(n_1 + n_2 - 2k)}{\overline{RSS}.k}$$

$$+ C = F_\alpha(k; n_1 + n_2 - 2k).$$

+ Kết luận

Nếu $F > C$: bác bỏ H_0 . Nếu $F < C$: chưa có cơ sở bác bỏ H_0 .

2. Phương pháp sử dụng biến giả

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 D_i + \beta_4 (D_i X_i) + U_i \quad (*)$$

$$\rightarrow \begin{cases} E(Y/D_i = 0, X_i) = \beta_1 + \beta_2 X_i \\ E(Y/D_i = 1, X_i) = (\beta_1 + \beta_3) + (\beta_2 + \beta_4) X_i \end{cases}$$

Chú ý: Xét hai mô hình hồi quy

$$Y_i = \lambda_1 + \lambda_2 X_i + U_i$$

$$Y_j = \gamma_1 + \gamma_2 X_j + U_j$$

Ta có 4 trường hợp:

$$\text{i) } \begin{cases} \lambda_1 = \gamma_1 \\ \lambda_2 = \gamma_2 \end{cases} : \text{ hai hàm hồi quy đồng nhất.}$$

$$\text{ii) } \begin{cases} \lambda_1 \neq \gamma_1 \\ \lambda_2 = \gamma_2 \end{cases} : \text{ hai hàm hồi quy cùng hệ số góc.}$$

$$\text{iii) } \begin{cases} \lambda_1 = \gamma_1 \\ \lambda_2 \neq \gamma_2 \end{cases} : \text{ hai hàm hồi quy cùng hệ số chặn.}$$

$$\text{iv) } \begin{cases} \lambda_1 \neq \gamma_1 \\ \lambda_2 \neq \gamma_2 \end{cases} : \text{ hai hàm hồi quy hoàn toàn khác nhau..}$$

Như vậy từ (*) ta suy ra để xét xem 2 mô hình có khác nhau hay không, ta tiến hành kiểm định các giả thiết sau:

$$+ H_0 : \beta_3 = 0; H_1 : \beta_3 \neq 0$$

$$+ H_0 : \beta_4 = 0; H_1 : \beta_4 \neq 0$$

1.5 Kiểm định giả thiết mô hình

1.5.1 Đa cộng tuyến

Khái niệm về đa cộng tuyến

Đa cộng tuyến là hiện tượng các biến giải thích (biến độc lập) trong mô hình phụ thuộc tuyến tính lẫn nhau. Hay

$$\text{Cov}(X_i, X_j) \neq 0, \quad \forall i \neq j$$

Nguyên nhân của đa cộng tuyến

Do sự tác động của biến độc lập X_i $\xleftrightarrow{\text{tác động}}$ X_j

- Do bản chất quan hệ giữa các biến độc lập
- Do chọn biến độc lập trong mô hình

Xét hàm hồi quy k biến $Y_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + U_i$

- Nếu tồn tại các tham số $\lambda_2, \lambda_3, \dots, \lambda_k$ không đồng thời bằng không sao cho

$$\lambda_2 X_2 + \lambda_3 X_3 + \dots + \lambda_k X_k = 0$$

thì mô hình xảy ra hiện tượng đa cộng tuyến hoàn hảo.

- Nếu tồn tại các tham số $\lambda_2, \lambda_3, \dots, \lambda_k$ không đồng thời bằng không sao cho

$$\lambda_2 X_2 + \lambda_3 X_3 + \dots + \lambda_k X_k + V = 0$$

thì mô hình xảy ra hiện tượng đa cộng tuyến không hoàn hảo.

Hậu quả của đa cộng tuyến

Khi hiện tượng đa cộng tuyến xảy ra, chúng ta sẽ gặp phải một số hậu quả sau

- Các hệ số hồi quy không xác định (hoàn hảo), khó xác định hay là các ước lượng bị chệch.
- Phương sai và hiệp phương sai của các ước lượng hồi quy lớn .
- Khoảng tin cậy rộng dẫn đến không có ý nghĩa.
- Giá trị thống kê T nhỏ dẫn đến luôn bác bỏ giả thuyết kiểm định ý nghĩa của biến độc lập trong mô hình.
- Các ước lượng rất nhạy cảm khi dữ liệu thay đổi nhỏ.
- Dấu của các hệ số hồi quy thay đổi so với thực tế.

Cách phát hiện đa cộng tuyến

Để nhận biết hiện tượng đa cộng tuyến, ta thường áp dụng một số cách sau đây

1. **Hệ số xác định** trong mô hình hồi quy gốc có giá trị rất cao, trong khi các giá trị t quan sát lại nhỏ.
2. **Tìm ma trận tương quan** giữa các biến giải thích trong mô hình. Theo kinh nghiệm, nếu hệ số tương quan giữa hai biến giải thích > 0.8 mà giá trị t quan sát thấp thì mô hình có đa cộng tuyến cao.
3. **Thực hiện hồi quy phụ**: lần lượt chọn một trong số các biến giải thích làm biến phụ thuộc rồi hồi quy theo tất cả các biến giải thích còn lại trong mô hình. Theo kinh nghiệm, nếu hệ số xác định của các mô hình hồi quy phụ (R_j^2) có giá trị từ 0.8 trở lên thì được coi là có đa cộng tuyến cao.
4. **Sử dụng nhân tử phóng đại phương sai** $VIF_j = \frac{1}{1 - R_j^2}$. Nếu $VIF_j \geq 10$ (tương đương $R_j^2 \geq 0.9$) thì kết luận mô hình gốc có đa cộng tuyến cao.

Biện pháp khắc phục đa cộng tuyến

Một số biện pháp nhằm khắc phục hiện tượng đa cộng tuyến

1. Sử dụng thông tin tiên nghiệm
2. Thu thập thêm số liệu hoặc lấy thêm mẫu mới
3. Bỏ biến
4. Sử dụng sai phân cấp một
5. Giảm tương quan trong hồi quy đa thức
6. Hồi quy thành phần chính

Ta lưu ý rằng các biện pháp được liệt kê trên đây có thể làm giải pháp cho vấn đề đa cộng tuyến như thế nào còn phụ thuộc vào tập số liệu và tính nghiêm trọng của vấn đề đa cộng tuyến.

1.5.2 Phương sai thay đổi

Khái niệm về phương sai thay đổi

Phương sai thay đổi là hiện tượng mà phương sai của các sai số ngẫu nhiên (U_i) trong mô hình không cố định (thay đổi). Hay

$$Var(U_i) = \sigma_i^2 \quad \forall i$$

Nguyên nhân của phương sai thay đổi

Do sự tác động của các biến độc lập X_i $\xleftrightarrow{\text{tác động}}$ U_i

- Do bản chất các mối quan hệ kinh tế.
- Do công cụ và kỹ thuật thu thập, xử lý số liệu được cải tiến thì sai số đo lường và sai số khi tính toán có xu hướng giảm dần, dẫn đến σ_i^2 có khả năng giảm.
- Phương sai thay đổi có thể xảy ra khi trong mẫu có các giá trị đột biến.
- Phương sai thay đổi có thể xảy ra khi mô hình hồi quy xác định sai (dạng hàm sai, thiếu biến quan trọng).
- Trường hợp phương sai không đồng đều thường gặp khi thu thập số liệu theo không gian (số liệu chéo).
- Do ảnh hưởng của các yếu tố khác làm tăng hay giảm sai số trong mô hình.

Hậu quả của phương sai thay đổi

Khi mô hình xảy ra hiện tượng phương sai thay đổi, ta sẽ gặp phải một số vấn đề sau

- Các ước lượng OLS vẫn là ước lượng tuyến tính, không chệch nhưng không còn hiệu quả nữa.
- Ước lượng của phương sai bị chệch nên các kiểm định mức ý nghĩa và khoảng tin cậy theo phân phối Student và Fisher không còn đáng tin cậy nữa.
- Kết quả bài toán dự báo không hiệu quả khi sử dụng các ước lượng OLS.

Cách phát hiện phương sai thay đổi

Một số phương pháp thông dụng để nhận biết mô hình xảy ra hiện tượng phương sai thay đổi

1. Phương pháp đồ thị
2. Kiểm định Park
3. Kiểm định Glejser
4. Kiểm định White

Biện pháp khắc phục

Về mặt thực tế khi mô hình bị phát hiện là có hiện tượng phương sai thay đổi thì ta có thể tiến hành như sau:

- Điều tra lại số liệu (tốn kém chi phí, khó khăn).
- Sử dụng mô hình nhưng có kiểm soát về hiện tượng phương sai thay đổi.

1.5.3 Tự tương quan

Khái niệm về tự tương quan

Tự tương quan là hiện tượng sai số ngẫu nhiên ở các thời điểm khác nhau có quan hệ với nhau. Hay

$$\text{Cov}(U_i, U_j) \neq 0, \quad \forall i \neq j$$

- + Nếu $U_i \leftrightarrow U_{i-1}$: hiện tượng tự tương quan bậc 1.
- + Nếu $U_i \leftrightarrow U_{i-1} + U_{i-2} + \dots + U_{i-p}$: hiện tượng tự tương quan bậc p.

Nguyên nhân của tự tương quan

Do sự tác động của các biến độc lập lên sai số ngẫu nhiên $U_i \overset{\text{tác động}}{\longleftrightarrow} U_j$

- Luật quán tính
- Do mô hình: không đưa đủ biến vào mô hình hoặc chọn mô hình không đúng.
- Số liệu: thu thập số liệu và xử lý số liệu.

Hậu quả của tự tương quan

Khi mô hình xảy ra hiện tượng tự tương quan, ta sẽ gặp phải một số vấn đề sau

- Các ước lượng OLS vẫn là ước lượng tuyến tính, không chệch nhưng không còn hiệu quả nữa.
- Ước lượng của phương sai bị chệch nên các kiểm định mức ý nghĩa và khoảng tin cậy theo phân phối Student và Fisher không còn đáng tin cậy nữa.
- Kết quả bài toán dự báo không hiệu quả khi sử dụng các ước lượng OLS.

Cách phát hiện tự tương quan

Một số phương pháp thông dụng để nhận biết mô hình xảy ra hiện tượng tự tương quan hay không

1. Phương pháp đồ thị
2. Kiểm định Durbin - Watson
3. Kiểm định B-G

Biện pháp khắc phục

Chương 2

Bài tập ứng dụng

2.1 Mô hình hồi quy hai biến

Bài 2.1. Cho bảng số liệu sau về tỷ lệ lạm phát ($X : \%$) và lãi suất ngân hàng ($Y : \%$)

X	7.2	4.0	3.1	1.6	4.8	51	2.0	6.6	4.4
Y	11.9	9.4	7.5	4.0	11.3	66.3	2.2	10.3	7.6

1. Tìm hàm hồi quy mẫu và giải thích ý nghĩa kinh tế của các hệ số hồi quy?
2. Tìm hệ số xác định mô hình và cho biết ý nghĩa của nó?
3. Với mức ý nghĩa 5%, hãy tìm khoảng tin cậy cho các hệ số hồi quy?
4. Kiểm định ý nghĩa của biến X trong mô hình (lạm phát có ảnh hưởng đến lãi suất không)?
5. Kiểm định sự phù hợp của mô hình? (mô hình có phù hợp với thực tế không?)
6. Với mức ý nghĩa 5%, hãy cho biết khoảng dự báo trung bình và cá biệt của lãi suất ngân hàng với mức lạm phát $X_0 = 5\%$.
7. Tính hệ số co giãn của tỷ lệ lạm phát đối với lãi suất ngân hàng tại điểm (\bar{x}, \bar{y}) và nêu ý nghĩa kinh tế.

Giải

Sử dụng máy tính bỏ túi (570ES, 570ES Plus, 570ES Plus II,...) hoặc phần mềm thống kê (Eviews, SPSS, STATA,...) ta tính được các giá trị sau đây từ bảng số liệu:

$$\begin{aligned}\hat{\beta}_2 &= \frac{\sum X_i Y_i - n \overline{XY}}{\sum X_i^2 - n(\overline{X})^2} = 1,249406687; \\ \hat{\beta}_1 &= \overline{Y} - \hat{\beta}_2 \overline{X} = 2,74169485;\end{aligned}$$

$$TSS = n \text{var}(Y) = 3102,04;$$

$$ESS = n \hat{\beta}_2^2 \text{var}(X) = 3081,211806;$$

$$RSS = TSS - ESS = n(1 - r^2) \text{var}(Y) = 20,82819405;$$

$$R^2 = \frac{ESS}{TSS} = 0,9932856462;$$

$$\hat{\sigma}^2 = \frac{n}{n-2} (1 - r^2) \text{var}(Y) = \frac{RSS}{n-2} = 2,975456293.$$

$$\text{var}(\hat{\beta}_1) = \left[\frac{1}{n} + \frac{\overline{X}^2}{n \cdot \text{var}(X)} \right] \hat{\sigma}^2 = 0,4641186156;$$

$$se(\hat{\beta}_1) = \sqrt{\text{var}(\hat{\beta}_1)} = 0,681263;$$

$$\text{var}(\hat{\beta}_2) = \frac{\hat{\sigma}^2}{n \text{var}(X)} = 0,001507433;$$

$$se(\hat{\beta}_2) = \sqrt{\text{var}(\hat{\beta}_2)} = 0,038826;$$

$$\text{var}(\hat{Y}_0) = \left[\frac{1}{n} + \frac{(X_0 - \overline{X})^2}{n \text{var}(X)} \right] \hat{\sigma}^2 = 0,359937849;$$

$$se(\hat{Y}_0) = \sqrt{\text{var}(\hat{Y}_0)} = 0,599948275;$$

$$\text{var}(Y_0 - \hat{Y}_0) = \hat{\sigma}^2 + \text{var}(\hat{Y}_0) = 3,335394142;$$

$$se(Y_0 - \hat{Y}_0) = \sqrt{\text{var}(Y_0 - \hat{Y}_0)} = 1,826306147.$$

1. Tìm mô hình hồi quy

$$\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X$$

$$\Rightarrow \widehat{LS} = 2,7417 + 1,2494LP$$

Ý nghĩa: khi tỷ lệ lạm phát tăng 1% thì lãi suất ngân hàng tăng 1,2494%.

2. Tính hệ số xác định mô hình và nêu ý nghĩa. Tính hệ số xác định có hiệu chỉnh

$R^2 = \frac{ESS}{TSS} = 0,9932856462$. Ý nghĩa: cho biết sự biến thiên của lạm phát giải thích được 99,33% sự biến thiên của lãi suất ngân hàng.

3. Tìm khoảng tin cậy cho các hệ số hồi quy

Áp dụng: $\widehat{\beta}_i - C.se(\widehat{\beta}_i) \leq \beta_i \leq \widehat{\beta}_i + C.se(\widehat{\beta}_i)$. Trong đó

$$C = t_{\frac{\alpha}{2}}(n - k) = t_{0,025}(9 - 2) = 2,365$$

+ Khoảng tin cậy của β_1

$$2,7417 - 2,365.0,6813 \leq \beta_1 \leq 2,7417 + 2,365.0,6813$$

$$\Rightarrow 1,1304 \leq \beta_1 \leq 4,353$$

+ Khoảng tin cậy của β_2

$$1,2494 - 2,365.0,0388 \leq \beta_2 \leq 1,2494 + 2,365.0,0388$$

$$\Rightarrow 1,1576 \leq \beta_2 \leq 1,3412$$

4. Kiểm định ý nghĩa của biến X trong mô hình

+ Đặt giả thiết $H_0 : \beta_2 = 0$; $H_1 : \beta_2 \neq 0$.

+ $C = t_{\frac{\alpha}{2}}^{(n-k)} = t_{0,025}(9 - 2) = 2,365$.

+ $T = \frac{\widehat{\beta}_2}{se(\widehat{\beta}_2)} = \frac{1,2494}{0,03883} = 32,2$.

+ $|T| > C$ suy ra bác bỏ H_0 . Vậy lạm phát có ảnh hưởng đến lãi suất.

5. Mô hình có phù hợp với thực tế không

+ Đặt giả thiết $H_0 : R^2 = 0$; $H_1 : R^2 \neq 0$.

+ Với $\alpha = 0,05$, $C = F_{\alpha}(k - 1; n - k) = F_{0,05}(2 - 1; 9 - 2) = 5,59$.

+ $F = \frac{(n - k) R^2}{(k - 1)(1 - R^2)} = \frac{(9 - 2) 0,993285647}{(2 - 1)(1 - 0,993285647)} = 1035,543$

+ $F > C$ nên bác bỏ H_0 . Vậy mô hình phù hợp.

6. Khoảng dự báo cho biến phụ thuộc

$$+ \hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_0 = 2,7417 + 1,2494.5 = 8,9887.$$

$$+ C = t_{\frac{\alpha}{2}}^{(n-k)} = t_{0,025} (9 - 2) = 2,365.$$

Khoảng dự báo cho giá trị trung bình $E(Y/X_0 = 5)$

$$\begin{aligned} \hat{Y}_0 - Cse(\hat{Y}_0) &\leq E(Y/X_0 = 5) \leq \hat{Y}_0 + Cse(\hat{Y}_0) \\ \Rightarrow 8,9887 - 2,365.0,5999 &\leq E(Y/X_0 = 5) \leq 8,9887 + 2,365.0,5999 \\ \Rightarrow 7,5699 &\leq E(Y/X_0 = 5) \leq 10,4074 \end{aligned}$$

Khoảng dự báo cho giá trị cá biệt Y_0

$$\begin{aligned} \hat{Y}_0 - Cse(Y_0 - \hat{Y}_0) &\leq Y_0 \leq \hat{Y}_0 + Cse(Y_0 - \hat{Y}_0) \\ \Rightarrow 8,9887 - 2,365.1,8263 &\leq Y_0 \leq 8,9887 + 2,365.1,8263 \\ \Rightarrow 4,6695 &\leq Y_0 \leq 13,30789 \end{aligned}$$

7. Tính hệ số có dãn

Hệ số co giãn của tỷ lệ lạm phát theo lãi suất tại điểm (\bar{x}, \bar{y}) là

$$\varepsilon_{\frac{Y}{X}} = \frac{dY}{dX} \frac{\bar{X}}{\bar{Y}} = \hat{\beta}_2 \frac{\bar{X}}{\bar{Y}} = 1,2494 \cdot \frac{9,411}{14,5} = 0,8109$$

Ý nghĩa: khi tỷ lệ lạm phát tăng 1% thì lãi suất ngân hàng tăng 0,8109% .

Bài 2.2. Giả sử có số liệu về chi tiêu mặt hàng A (Y triệu đồng/tháng) và thu nhập của người tiêu dùng(X triệu đồng/tháng) như sau:

Y	0.1	0.15	0.18	0.2	0.25
X	1.0	1.5	2.0	2.5	4.0

1. Hãy ước lượng mô hình hồi quy tuyến tính mô tả quan hệ giữa chi tiêu mặt hàng A và thu nhập của người tiêu dùng. Nêu ý nghĩa kinh tế của các hệ số hồi quy được ước lượng?

2. Tìm hệ số xác định mô hình và cho biết ý nghĩa của nó?
3. Xét xem thu nhập có ảnh hưởng đến chi tiêu mặt hàng A hay không với mức ý nghĩa 1%.
4. Dự đoán mức chi tiêu trung bình và cá biệt cho mặt hàng A khi thu nhập là 3 triệu đồng/tháng với độ tin cậy 99%.
5. Tính hệ số co giãn của chi tiêu loại hàng A đối với thu nhập tại điểm (\bar{x}, \bar{y}) và nêu ý nghĩa kinh tế.
6. Hãy viết hàm hồi quy mẫu khi đơn vị tính của chi tiêu là đồng/tháng và đơn vị tính của thu nhập là ngàn đồng/tháng?

Giải

Sử dụng máy tính bỏ túi (570ES, 570ES Plus, 570ES Plus II,...) hoặc phần mềm thống kê (Eviews, SPSS, STATA,...) ta tính được các giá trị sau đây từ bảng số liệu:

$$\hat{\beta}_2 = \frac{\sum X_i Y_i - n \bar{X} \bar{Y}}{\sum X_i^2 - n (\bar{X})^2} = 0,047;$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} = 0,0726;$$

$$TSS = n \text{var}(Y) = 0,01252;$$

$$ESS = n \hat{\beta}_2^2 \text{var}(X) = 0,01171;$$

$$RSS = TSS - ESS = n(1 - r^2) \text{var}(Y) = 0,00081;$$

$$R^2 = \frac{ESS}{TSS} = 0,935;$$

$$\hat{\sigma}^2 = \frac{n}{n-2} (1 - r^2) \text{var}(Y) = \frac{RSS}{n-2} = 0,00027.$$

$$\begin{aligned} \text{var}(\hat{\beta}_1) &= \left[\frac{1}{n} + \frac{\bar{X}^2}{n \cdot \text{var}(X)} \right] \hat{\sigma}^2 = 0,0003; \\ \text{se}(\hat{\beta}_1) &= \sqrt{\text{var}(\hat{\beta}_1)} = 0,0173; \\ \text{var}(\hat{\beta}_2) &= \frac{\hat{\sigma}^2}{n \text{var}(X)} = 0,00005; \\ \text{se}(\hat{\beta}_2) &= \sqrt{\text{var}(\hat{\beta}_2)} = 0,0071; \\ \text{var}(\hat{Y}_0) &= \left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{n \text{var}(X)} \right] \hat{\sigma}^2 = 8,66038 \cdot 10^{-5}; \\ \text{se}(\hat{Y}_0) &= \sqrt{\text{var}(\hat{Y}_0)} = 0,009306; \end{aligned}$$

1. Tìm mô hình hồi quy

$$\hat{Y} = 0,0726 + 0,047X$$

Ý nghĩa: khi thu nhập của người tiêu dùng tăng 1 triệu đồng/tháng thì mức chi tiêu mặt hàng A trung bình tăng 0,047 triệu đồng/tháng (tương ứng giảm).

2. Tính hệ số xác định mô hình và nêu ý nghĩa.

$$R^2 = \frac{ESS}{TSS} = 0,935$$

Ý nghĩa: cho biết thu nhập giải thích được 93,5% sự thay đổi chi tiêu của mặt hàng A.

3. Kiểm định ý nghĩa của biến X trong mô hình

+ Đặt giả thiết $H_0 : \beta_2 = 0; H_1 : \beta_2 \neq 0$.

$$+ C = t_{\frac{\alpha}{2}}^{(n-k)} = t_{0,025}^{(5-2)} = 5,841.$$

$$+ T = \frac{\hat{\beta}_2}{\text{se}(\hat{\beta}_2)} = \frac{0,047}{0,0071} = 6,6197.$$

+ $|T| > C$ suy ra bác bỏ H_0 . Vậy thu nhập có ảnh hưởng đến chi tiêu.

4. Khoảng dự báo cho biến phụ thuộc

$$+ \hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_0 = 0,2136.$$

$$+ C = t_{\frac{\alpha}{2}}^{(n-k)} = t_{0,025} (5 - 2) = 5,841.$$

Khoảng dự báo cho giá trị trung bình Y_0 ($E(Y/X_0 = 3)$)

$$\begin{aligned} \widehat{Y}_0 - Cse(\widehat{Y}_0) &\leq E(Y/X_0 = 3) \leq \widehat{Y}_0 + Cse(\widehat{Y}_0) \\ \Rightarrow 0,2136 - 5,841 \cdot 0,009306 &\leq E(Y/X_0 = 3) \leq 0,2136 + 5,841 \cdot 0,009306 \\ \Rightarrow 0,1592 &\leq E(Y/X_0 = 3) \leq 0,268 \end{aligned}$$

5. Tính hệ số có dẫn

Hệ số co giãn của chi tiêu theo thu nhập tại điểm (\bar{x}, \bar{y}) là

$$\varepsilon_{\frac{Y}{X}} = \frac{dY}{dX} \frac{\bar{X}}{\bar{Y}} = \widehat{\beta}_2 \frac{\bar{X}}{\bar{Y}} = 0,047 \cdot \frac{2,2}{0,176} = 0,5875$$

Ý nghĩa: khi thu nhập trung bình của người tiêu dùng tăng 1% thì mức chi tiêu trung bình về mặt hàng A tăng 0,59% (tương ứng giảm).

6. Đổi đơn vị

+ Đơn vị tính của chi tiêu là đồng/tháng

$$\Rightarrow Y^* = 1000000Y$$

$$\Rightarrow k_1 = 1000000$$

$$\Rightarrow \widehat{\beta}_1^* = k_1 \widehat{\beta}_1 = 72600$$

+ Đơn vị của thu nhập là ngàn đồng/tháng

$$\Rightarrow X^* = 1000X$$

$$\Rightarrow k_2 = 1000$$

$$\Rightarrow \widehat{\beta}_2^* = \frac{k_1}{k_2} \widehat{\beta}_2 = 47$$

Vậy

$$\widehat{Y}^* = \widehat{\beta}_1^* + \widehat{\beta}_2^* X^* = 72600 + 47X^*$$

Bài 2.3. Người ta muốn phân tích và đánh giá kết quả về năng suất lúa của đồng bằng sông cửu long trong thời gian 10 năm từ 1988 - 1997 đã tiến hành thu thập một mẫu số liệu gồm các giá trị quan sát về 2 đại lượng Y, X như sau

Y	40	44	46	48	52	58	60	68	74	80
X	6	10	12	14	16	18	22	24	26	32

1. Ước lượng hàm hồi quy tuyến tính mẫu $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$?
2. Nêu ý nghĩa của các hệ số hồi quy đã ước lượng được. Các giá trị đó có phù hợp với lý thuyết kinh tế hay không?
3. Tìm khoảng tin cậy của β_2 với độ tin cậy 95% và nêu ý nghĩa?
4. Với mức ý nghĩa 5%, hãy cho biết mức phân bón có thực sự ảnh hưởng đến năng suất lúa hay không?
5. Với mức ý nghĩa 5%, hãy cho biết hệ số góc của mô hình hồi quy bằng 2 được không?
6. Tính R^2 . Kiểm định sự phù hợp của hàm hồi quy với mức ý nghĩa 1%?
7. Dự báo năng suất lúa trung bình của đồng bằng sông cửu long khi mức phân bón là 20 tạ/ha với độ tin cậy 95%. Câu hỏi tương tự cho năng suất lúa cá biệt.

Giải

Ta có:

$$\hat{\beta}_2 = \frac{\sum X_i Y_i - n \bar{X} \bar{Y}}{\sum X_i^2 - n (\bar{X})^2} = 1,6597;$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} = 27,125;$$

$$TSS = n \text{var}(Y) = 1634;$$

$$ESS = n \hat{\beta}_2^2 \text{var}(X) = 1586,6519;$$

$$RSS = TSS - ESS = n(1 - r^2) \text{var}(Y) = 47,3480;$$

$$R^2 = \frac{ESS}{TSS} = 0,971;$$

$$\hat{\sigma}^2 = \frac{n}{n-2} (1 - r^2) \text{var}(Y) = \frac{RSS}{n-2} = 5,9185.$$

$$\begin{aligned} \text{var}(\widehat{\beta}_2) &= \frac{\widehat{\sigma}^2}{n\text{var}(X)} = 0,0103; \\ \text{se}(\widehat{\beta}_2) &= \sqrt{\text{var}(\widehat{\beta}_2)} = 0,1014; \\ \text{var}(\widehat{Y}_0) &= \left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{n\text{var}(X)} \right] \widehat{\sigma}^2 = 0,6329; \\ \text{se}(\widehat{Y}_0) &= \sqrt{\text{var}(\widehat{Y}_0)} = 0,7956; \\ \text{var}(Y_0 - \widehat{Y}_0) &= \widehat{\sigma}^2 + \text{var}(\widehat{Y}_0) = 6,5514; \\ \text{se}(Y_0 - \widehat{Y}_0) &= \sqrt{\text{var}(Y_0 - \widehat{Y}_0)} = 2,5596. \end{aligned}$$

1. Ước lượng hàm hồi quy tuyến tính mẫu

$$\widehat{Y} = 27,125 + 1,6597X$$

2. Ý nghĩa của các hệ số hồi quy đã ước lượng được

+ $\widehat{\beta}_1 = 27,125$: với số liệu của mẫu khi mức phân bón bằng 0, thì năng suất trung bình của lúa tối thiểu là 27,125 (tạ/ha).

+ $\widehat{\beta}_2 = 1,6597 > 0$: với mẫu số liệu trên, mức phân bón và năng suất lúa có quan hệ đồng biến. Với điều kiện các yếu tố khác không đổi, nếu mức phân bón tăng 1 (tạ/ha) thì năng suất trung bình của lúa tăng 1,6597 (tạ/ha).

+ Ý nghĩa các hệ số trên là phù hợp với lý thuyết kinh tế.

3. Khoảng tin cậy của β_2 với độ tin cậy 95%

Áp dụng: $\widehat{\beta}_i - C.se(\widehat{\beta}_i) \leq \beta_i \leq \widehat{\beta}_i + C.se(\widehat{\beta}_i)$. Trong đó

$$C = t_{\frac{\alpha}{2}}^{(n-k)} = t_{0,025}^{(10-2)} = 2,306$$

Khoảng tin cậy của β_2

$$\widehat{\beta}_2 - C.se(\widehat{\beta}_2) \leq \beta_2 \leq \widehat{\beta}_2 + C.se(\widehat{\beta}_2)$$

$$\Rightarrow 1,4259 \leq \beta_2 \leq 1,8935$$

Ý nghĩa: khi mức phân bón tăng lên 1 (tạ/ha), với điều kiện các yếu tố khác không đổi, năng suất trung bình của lúa tăng lên trong khoảng (1,4259; 1,8935) (tạ/ha) với độ tin cậy 95%.

4. Mức phân bón có thực sự ảnh hưởng đến năng suất lúa hay không?

+ Đặt giả thiết $H_0 : \beta_2 = 0; H_1 : \beta_2 \neq 0$.

$$+ C = t_{\frac{\alpha}{2}}^{(n-k)} = t_{0,025} (10 - 2) = 2,306.$$

$$+ T = \frac{\widehat{\beta}_2}{se(\widehat{\beta}_2)} = 16,3679.$$

+ $|T| > C$ suy ra bác bỏ H_0 . Vậy với mức ý nghĩa 5%, mức phân bón thực sự ảnh hưởng đến năng suất lúa.

5. Với $\alpha = 5\%$, hãy cho biết hệ số góc của mô hình hồi quy bằng 2 được không?

+ Đặt giả thiết $H_0 : \beta_2 = 2; H_1 : \beta_2 \neq 2$.

$$+ C = t_{\frac{\alpha}{2}}^{(n-k)} = t_{0,025} (10 - 2) = 2,306.$$

$$+ T = \frac{\widehat{\beta}_2 - 2}{se(\widehat{\beta}_2)} = 3,356.$$

+ $|T| > C$ suy ra bác bỏ H_0 . Vậy ý kiến trên là không đúng.

6. Tính R^2 và \bar{R}^2 . Kiểm định sự phù hợp của hàm hồi quy với mức ý nghĩa 1%?

$$* R^2 = \frac{ESS}{TSS} = 0,971$$

Ý nghĩa: mức phân bón giải thích 97,1% sự biến động về năng suất lúa. Mức phù hợp của mô hình cao.

* Kiểm định

+ Đặt giả thiết $H_0 : R^2 = 0; H_1 : R^2 \neq 0$.

+ Với $\alpha = 0,05$, $C = F_{\alpha}(k - 1; n - k) = F_{0,01}(2 - 1; 10 - 2) = 11,3$.

$$+ F = \frac{(n - k) R^2}{(k - 1)(1 - R^2)} = \frac{(10 - 2) 0,971}{(2 - 1)(1 - 0,971)} = 267,8621$$

+ $F > C$ nên bác bỏ H_0 . Vậy mô hình phù hợp với mức ý nghĩa 1%.

7. Dự báo năng suất lúa trung bình và cá biệt.

$$+ \widehat{Y}_0 = \widehat{\beta}_1 + \widehat{\beta}_2 X_0 = 27,125 + 1,6597.20 = 60,3194.$$

$$+ C = t_{\frac{\alpha}{2}}^{(n-k)} = t_{0,025} (10 - 2) = 2,306.$$

Khoảng dự báo cho giá trị trung bình Y_0 ($E(Y/X_0 = 20)$)

$$\begin{aligned} \hat{Y}_0 - C_{se}(\hat{Y}_0) &\leq E(Y/X_0 = 20) \leq \hat{Y}_0 + C_{se}(\hat{Y}_0) \\ \Rightarrow 60,3194 - 2,306.0,7956 &\leq E(Y/X_0 = 20) \leq 60,3194 + 2,306.0,7956 \\ \Rightarrow 58,4847 &\leq E(Y/X_0 = 20) \leq 62,1541 \end{aligned}$$

Khoảng dự báo cho giá trị cá biệt Y_0

$$\begin{aligned} \hat{Y}_0 - C_{se}(Y_0 - \hat{Y}_0) &\leq Y_0 \leq \hat{Y}_0 + C_{se}(Y_0 - \hat{Y}_0) \\ \Rightarrow 60,3194 - 2,306.2,5596 &\leq Y_0 \leq 60,3194 + 2,306.2,5596 \\ \Rightarrow 54,4169 &\leq Y_0 \leq 66,2218 \end{aligned}$$

Bài 2.4. Bảng sau cho số liệu về chi tiêu cho tiêu dùng (Y-USD/tuần) và thu nhập hàng tuần (X-USD/tuần) của một mẫu gồm 10 hộ gia đình

Y	70	65	90	95	110	115	120	140	155	150
X	80	100	120	140	160	180	200	220	240	260

1. Tìm mô hình hồi quy mẫu và cho biết ý nghĩa của các hệ số hồi quy?
2. Tìm khoảng tin cậy cho các hệ số hồi quy với độ tin cậy 95%?
3. Thu nhập có ảnh hưởng đến chi tiêu hay không với mức ý nghĩa 5%? (kiểm định ý nghĩa của biến X trong mô hình)
4. Mô hình có phù hợp với thực tế không? (kiểm định sự phù hợp của mô hình)
5. Dự báo khi thu nhập ở mức 100 USD/tuần với độ tin cậy 95%?

Giải

Sử dụng máy tính bỏ túi (570ES, 570ES Plus, 570ES Plus II,...) hoặc phần mềm thống kê (Eviews, SPSS, STATA,...) ta tính được các giá trị sau đây từ bảng số

liệu:

$$\hat{\beta}_2 = \frac{\sum X_i Y_i - n\bar{X}\bar{Y}}{\sum X_i^2 - n(\bar{X})^2} = 0,5091;$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} = 24,4545;$$

$$TSS = n\text{var}(Y) = 8890;$$

$$ESS = n\hat{\beta}_2^2 \text{var}(X) = 8552,73;$$

$$RSS = TSS - ESS = n(1 - r^2) \text{var}(Y) = 337,27;$$

$$R^2 = \frac{ESS}{TSS} = 0,9621;$$

$$\hat{\sigma}^2 = \frac{n}{n-2} (1 - r^2) \text{var}(Y) = \frac{RSS}{n-2} = 42,15875.$$

$$\text{var}(\hat{\beta}_1) = \left[\frac{1}{n} + \frac{\bar{X}^2}{n \cdot \text{var}(X)} \right] \hat{\sigma}^2 = 41,13672;$$

$$se(\hat{\beta}_1) = \sqrt{\text{var}(\hat{\beta}_1)} = 6,4138;$$

$$\text{var}(\hat{\beta}_2) = \frac{\hat{\sigma}^2}{n \text{var}(X)} = 0,0012775;$$

$$se(\hat{\beta}_2) = \sqrt{\text{var}(\hat{\beta}_2)} = 0,035742;$$

$$\text{var}(\hat{Y}_0) = \left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{n \text{var}(X)} \right] \hat{\sigma}^2 = 10,4758;$$

$$se(\hat{Y}_0) = \sqrt{\text{var}(\hat{Y}_0)} = 3,2366;$$

$$\text{var}(Y_0 - \hat{Y}_0) = \hat{\sigma}^2 + \text{var}(\hat{Y}_0) = 52,6345;$$

$$se(Y_0 - \hat{Y}_0) = \sqrt{\text{var}(Y_0 - \hat{Y}_0)} = 7,25497.$$

1. Mô hình hồi quy mẫu và cho biết ý nghĩa của các hệ số hồi quy

$$\hat{Y} = 24,4545 + 0,5091X$$

Ý nghĩa: khi thu nhập tăng 1 USD/tuần thì chi tiêu của người tiêu dùng tăng 0,5091 USD.

2. Khoảng tin cậy cho các hệ số hồi quy

Áp dụng: $\widehat{\beta}_i - C.se(\widehat{\beta}_i) \leq \beta_i \leq \widehat{\beta}_i + C.se(\widehat{\beta}_i)$. Trong đó

$$C = t_{\frac{\alpha}{2}}^{(n-k)} = t_{0,025}(10 - 2) = 2,306$$

+ Khoảng tin cậy của β_1

$$\widehat{\beta}_1 - C.se(\widehat{\beta}_1) \leq \beta_1 \leq \widehat{\beta}_1 + C.se(\widehat{\beta}_1)$$

$$\Rightarrow 9,6643 \leq \beta_1 \leq 39,2447$$

+ Khoảng tin cậy của β_2

$$\widehat{\beta}_2 - C.se(\widehat{\beta}_2) \leq \beta_2 \leq \widehat{\beta}_2 + C.se(\widehat{\beta}_2)$$

$$\Rightarrow 0,4267 \leq \beta_2 \leq 0,5915$$

3. Thu nhập có ảnh hưởng đến chi tiêu hay không

+ Đặt giả thiết $H_0 : \beta_2 = 0; H_1 : \beta_2 \neq 0$.

+ $C = t_{\frac{\alpha}{2}}^{(n-k)} = t_{0,025}(10 - 2) = 2,306$.

$$+ T = \frac{\widehat{\beta}_2}{se(\widehat{\beta}_2)} = 14,243.$$

+ $|T| > C$ suy ra bác bỏ H_0 . Vậy thu nhập có ảnh hưởng đến chi tiêu.

4. Mô hình có phù hợp với thực tế không

+ Đặt giả thiết $H_0 : R^2 = 0; H_1 : R^2 \neq 0$.

+ Với $\alpha = 0,05, C = F_{\alpha}(k - 1; n - k) = F_{0,05}(2 - 1; 10 - 2) = 5,32$.

$$+ F = \frac{(n - k) R^2}{(k - 1)(1 - R^2)} = \frac{(10 - 2) 0,9621}{(2 - 1)(1 - 0,9621)} = 203,08$$

+ $F > C$ nên bác bỏ H_0 . Vậy mô hình phù hợp với thực tế.

5. Dự báo khi thu nhập ở mức 100 USD/tuần với độ tin cậy 95%

$$+ \widehat{Y}_0 = \widehat{\beta}_1 + \widehat{\beta}_2 X_0 = 24,4545 + 0,5091.100 = 75,3645.$$

$$+ C = t_{\frac{\alpha}{2}}^{(n-k)} = t_{0,025} (10 - 2) = 2,306.$$

Khoảng dự báo cho giá trị trung bình Y_0 ($E(Y/X_0 = 100)$)

$$\begin{aligned} \hat{Y}_0 - Cse(\hat{Y}_0) &\leq E(Y/X_0 = 100) \leq \hat{Y}_0 + Cse(\hat{Y}_0) \\ \Rightarrow 67,9 &\leq E(Y/X_0 = 100) \leq 82,8 \end{aligned}$$

Khoảng dự báo cho giá trị cá biệt Y_0

$$\begin{aligned} \hat{Y}_0 - Cse(Y_0 - \hat{Y}_0) &\leq Y_0 \leq \hat{Y}_0 + Cse(Y_0 - \hat{Y}_0) \\ \Rightarrow 58,6 &\leq Y_0 \leq 92,1 \end{aligned}$$

2.2 Mô hình hồi quy bội

Bài 2.5. Bảng dưới đây cho các số liệu về doanh số bán (Y), chi phí chào hàng (X_2) và chi phí quảng cáo (X_3) trong năm 2011 ở 12 khu vực bán hàng của một công ty. Hãy ước lượng hàm hồi quy tuyến tính của doanh số bán theo chi phí chào hàng và chi phí quảng cáo (đơn vị: triệu đồng).

Y	X_2	X_3
1270	100	180
1490	106	248
1060	60	190
1626	160	240
1020	70	150
1800	170	260
1610	140	250
1280	120	160
1390	116	170
1440	120	230
1590	140	220
1380	150	150

Giải

$$\begin{aligned}n &= 12; \\ \sum Y_i &= 16956; & \sum X_{2i} &= 1452; & \sum X_{3i} &= 2448; \\ \sum Y_i^2 &= 24549576; & \sum X_{2i}^2 &= 188192; & \sum X_{3i}^2 &= 518504; \\ \sum X_{2i}X_{3i} &= 303608; & \sum Y_iX_{2i} &= 2128740; & \sum Y_iX_{3i} &= 3542360.\end{aligned}$$

$$X^T X = \begin{pmatrix} n & \sum X_{2i} & \sum X_{3i} \\ \sum X_{2i} & \sum X_{2i}^2 & \sum X_{2i}X_{3i} \\ \sum X_{3i} & \sum X_{3i}X_{2i} & \sum X_{3i}^2 \end{pmatrix} = \begin{pmatrix} 12 & 1452 & 2448 \\ 1452 & 188192 & 303608 \\ 2448 & 303608 & 518504 \end{pmatrix};$$

$$X^T Y = \begin{pmatrix} \sum Y_i \\ \sum X_{2i}Y_i \\ \sum X_{3i}Y_i \end{pmatrix} = \begin{pmatrix} 16956 \\ 2128740 \\ 3542360 \end{pmatrix};$$

$$\hat{\beta} = (X^T X)^{-1} (X^T Y) = \begin{pmatrix} 328, 1383 \\ 4, 64951 \\ 2, 560152 \end{pmatrix} \Rightarrow \begin{cases} \hat{\beta}_1 = 328, 1383 \\ \hat{\beta}_2 = 4, 64951 \\ \hat{\beta}_3 = 2, 560152 \end{cases}$$

Vậy $\hat{Y}_i = 328, 1383 + 4, 64951X_{2i} + 2, 560152X_{3i}$.

Bài 2.6. Số liệu quan sát của một mẫu cho ở bảng sau. Trong đó Y là lượng hàng bán được của một loại hàng hóa (tấn/tháng); X_2 là thu nhập của người tiêu dùng (triệu/năm) và X_3 là giá bán của loại hàng này (ngàn đồng/tháng).

Y	X_2	X_3
20	8	2
18	7	3
19	8	4
18	8	4
17	6	5
17	6	5
16	5	6
15	5	7
13	4	8
12	3	8

1. Tìm hàm hồi quy mẫu?
2. Tìm hệ số xác định của mô hình?

3. Tìm ma trận hiệp phương sai của $\hat{\beta}$?

Giải

1. Hàm hồi quy mẫu

$$\begin{aligned}n &= 10; \quad \bar{Y}^2 = 16,5^2 \\ \sum Y_i &= 165; \quad \sum X_{2i} = 60; \quad \sum X_{3i} = 52; \\ \sum Y_i^2 &= 2781; \quad \sum X_{2i}^2 = 388; \quad \sum X_{3i}^2 = 308; \\ \sum X_{2i}X_{3i} &= 282; \quad \sum Y_iX_{2i} = 1029; \quad \sum Y_iX_{3i} = 813.\end{aligned}$$

$$\begin{aligned}X^T X &= \begin{pmatrix} n & \sum X_{2i} & \sum X_{3i} \\ \sum X_{2i} & \sum X_{2i}^2 & \sum X_{2i}X_{3i} \\ \sum X_{3i} & \sum X_{3i}X_{2i} & \sum X_{3i}^2 \end{pmatrix} = \begin{pmatrix} 10 & 60 & 52 \\ 60 & 388 & 282 \\ 52 & 282 & 308 \end{pmatrix}; \\ X^T Y &= \begin{pmatrix} \sum Y_i \\ \sum X_{2i}Y_i \\ \sum X_{3i}Y_i \end{pmatrix} = \begin{pmatrix} 165 \\ 1029 \\ 813 \end{pmatrix}; \\ \hat{\beta} &= (X^T X)^{-1} (X^T Y) = \begin{pmatrix} 14,99215 \\ 0,76178 \\ -0,58901 \end{pmatrix} \Rightarrow \begin{cases} \hat{\beta}_1 = 14,99215 \\ \hat{\beta}_2 = 0,76178 \\ \hat{\beta}_3 = -0,58901 \end{cases}\end{aligned}$$

$$\text{Vậy } \hat{Y}_i = 14,99215 + 0,76178X_{2i} - 0,58901X_{3i}.$$

2. Hệ số xác định của mô hình

$$\begin{aligned}TSS &= \sum Y_i^2 - n\bar{Y}^2 = 58,5; \\ ESS &= \hat{\beta}^T (X^T Y) - n\bar{Y}^2 = 56,211; \\ R^2 &= \frac{ESS}{TSS} = 0,96088\end{aligned}$$

3. Ma trận hiệp phương sai

Ta có $\hat{\sigma}^2 = \frac{RSS}{n - k} = 0,327$

$$Cov(\hat{\beta}) = \hat{\sigma}^2(X^T X)^{-1} = \begin{pmatrix} 8,55593 & -0,81664 & -0,6968 \\ -0,81664 & 0,080466 & 0,0642 \\ -0,6968 & 0,0642 & 0,05992 \end{pmatrix}$$

$$\Rightarrow \begin{cases} se(\hat{\beta}_1) = \sqrt{\text{var}(\hat{\beta}_1)} = \sqrt{8,55593} = 2,925 \\ se(\hat{\beta}_2) = \sqrt{\text{var}(\hat{\beta}_2)} = \sqrt{0,080466} = 0,28366 \\ se(\hat{\beta}_3) = \sqrt{\text{var}(\hat{\beta}_3)} = \sqrt{0,05992} = 0,24478 \end{cases}$$

Bài 2.7. Số liệu về sản lượng Y , phân hóa học X_2 , thuốc trừ sâu X_3 , tính trên một đơn vị diện tích ha, cho trong bảng sau

Y	X_2	X_3
40	6	4
44	10	4
46	12	5
48	14	7
52	16	9
58	18	12
60	22	14
68	24	20
74	26	21
80	32	24

Với mức ý nghĩa 5%, hãy trả lời các câu hỏi sau:

1. Kết quả ước lượng có phù hợp với thực tế không? Hãy giải thích ý nghĩa kinh tế của các hệ số nhận được.
2. Phân bón có ảnh hưởng đến năng suất của loại cây trồng trên hay không? câu hỏi tương tự cho thuốc trừ sâu.
3. Hãy tìm khoảng tin cậy cho các hệ số hồi quy riêng?
4. Hãy giải thích ý nghĩa của hệ số R^2 nhận được? tính hệ số xác định hiệu chỉnh?

5. Có phải cả phân bón lẫn thuốc trừ sâu đều không ảnh hưởng đến năng suất?
6. Bạn có thể bỏ biến X_3 ra khỏi mô hình được không? Vì sao?
7. Phải chăng phân bón và thuốc trừ sâu đều có ảnh hưởng như nhau đến năng suất cây trồng trên?
8. Hãy dự báo giá trị trung bình và cá biệt khi $X_2 = 20$; $X_3 = 15$.

Giải

Sử dụng máy tính bỏ túi (570ES, 570ES Plus, 570ES Plus II,...) hoặc phần mềm thống kê (Eviews, SPSS, STATA,...) ta tính được các giá trị sau từ bảng số liệu:

$$\begin{aligned}
 n &= 10; \quad \bar{Y}^2 = 57^2 \\
 \sum Y_i &= 570; \quad \sum X_{2i} = 180; \quad \sum X_{3i} = 120; \\
 \sum Y_i^2 &= 34124; \quad \sum X_{2i}^2 = 3816; \quad \sum X_{3i}^2 = 1944; \\
 \sum X_{2i}X_{3i} &= 2684; \quad \sum Y_iX_{2i} = 11216; \quad \sum Y_iX_{3i} = 7740.
 \end{aligned}$$

$$X^T X = \begin{pmatrix} n & \sum X_{2i} & \sum X_{3i} \\ \sum X_{2i} & \sum X_{2i}^2 & \sum X_{2i}X_{3i} \\ \sum X_{3i} & \sum X_{3i}X_{2i} & \sum X_{3i}^2 \end{pmatrix} = \begin{pmatrix} 10 & 180 & 120 \\ 180 & 3816 & 2684 \\ 120 & 2684 & 1944 \end{pmatrix};$$

$$X^T Y = \begin{pmatrix} \sum Y_i \\ \sum X_{2i}Y_i \\ \sum X_{3i}Y_i \end{pmatrix} = \begin{pmatrix} 570 \\ 11216 \\ 7740 \end{pmatrix};$$

$$\hat{\beta} = (X^T X)^{-1} (X^T Y) = \begin{pmatrix} 31,98067 \\ 0,65005 \\ 1,10987 \end{pmatrix} \Rightarrow \begin{cases} \hat{\beta}_1 = 31,98067 \\ \hat{\beta}_2 = 0,65005 \\ \hat{\beta}_3 = 1,10987 \end{cases}$$

$$TSS = \sum Y_i^2 - n\bar{Y}^2 = 1634;$$

$$ESS = \hat{\beta}^T (X^T Y) - n\bar{Y}^2 = 1620,3365;$$

$$RSS = TSS - ESS = 13,6635;$$

$$R^2 = \frac{ESS}{TSS} = 0,99164;$$

$$\hat{\sigma}^2 = \frac{RSS}{n-k} = 1,95193.$$

$$\text{Cov}(\hat{\beta}) = \hat{\sigma}^2 (X^T X)^{-1} = \begin{pmatrix} 2,6614 & -0,3455 & 0,3127 \\ -0,3455 & 0,0625 & -0,065 \\ 0,3127 & -0,065 & 0,0715 \end{pmatrix}$$

$$\Rightarrow \begin{cases} se(\hat{\beta}_1) = \sqrt{\text{var}(\hat{\beta}_1)} = \sqrt{2,6614} = 1,6314 \\ se(\hat{\beta}_2) = \sqrt{\text{var}(\hat{\beta}_2)} = \sqrt{0,0625} = 0,25 \\ se(\hat{\beta}_3) = \sqrt{\text{var}(\hat{\beta}_3)} = \sqrt{0,0715} = 0,2674 \end{cases}$$

$$\text{var}(\hat{Y}_0) = \hat{\sigma}^2 (X^0)^T (X^T X)^{-1} X^0 = 0,308377 \Rightarrow se(\hat{Y}_0) = 0,555;$$

$$\text{var}(Y_0 - \hat{Y}_0) = \text{var}(\hat{Y}_0) + \hat{\sigma}^2 = 2,26 \Rightarrow se(Y_0 - \hat{Y}_0) = 1,5.$$

1. Tìm mô hình hồi quy

$$\hat{Y}_i = 31,98067 + 0,65005X_{2i} + 1,10987X_{3i}$$

- Do $\hat{\beta}_2 > 0$ và $\hat{\beta}_3 > 0$, nên khi tăng lượng phân bón và thuốc trừ sâu thì năng suất cây trồng sẽ tăng. Suy ra kết quả ước lượng là phù hợp với thực tế (chú ý rằng điều này không có nghĩa là khi tăng phân bón và thuốc trừ sâu ra vô cùng thì năng suất cũng tăng như vậy).
- $\hat{\beta}_1 = 31,98067$ có nghĩa là nếu không dùng phân bón và thuốc trừ sâu thì năng suất trung bình/ha sẽ là 31,98067 tấn.
- $\hat{\beta}_2 = 0,65005$ có nghĩa là trong điều kiện lượng thuốc trừ sâu không đổi, nếu tăng lượng phân bón lên 1 tấn/ha thì năng suất trung bình/ha sẽ tăng 0,65005.
- $\hat{\beta}_3 = 1,10987$ có nghĩa là trong điều kiện lượng phân bón không đổi, nếu tăng lượng thuốc trừ sâu lên 1 tấn/ha thì năng suất trung bình/ha sẽ tăng 1,10987.

2. Kiểm định ảnh hưởng của biến X trong mô hình

* Xét ảnh hưởng của phân bón

+ Đặt giả thiết $H_0 : \beta_2 = 0$; $H_1 : \beta_2 \neq 0$.

$$+ C = t_{\frac{\alpha}{2}}^{(n-k)} = t_{0,025}^{(10-3)} = 2,365.$$

$$+ T_2 = \frac{\hat{\beta}_2}{se(\hat{\beta}_2)} = \frac{0,65005}{0,25} = 2,6002.$$

+ $|T_2| > C$ suy ra bác bỏ H_0 . Vậy phân bón ảnh hưởng đến năng suất.

* Xét ảnh hưởng của thuốc trừ sâu

+ Đặt giả thiết $H_0 : \beta_3 = 0; H_1 : \beta_3 \neq 0$.

+ $C = t_{\frac{\alpha}{2}}^{(n-k)} = t_{0,025}^{(10-3)} = 2,365$.

+ $T_3 = \frac{\hat{\beta}_3}{se(\hat{\beta}_3)} = \frac{1,10987}{0,2674} = 4,1506$.

+ $|T_3| > C$ suy ra bác bỏ H_0 . Vậy thuốc trừ sâu ảnh hưởng đến năng suất.

3. Tìm khoảng tin cậy cho các hệ số hồi quy

Áp dụng: $\hat{\beta}_i - C.se(\hat{\beta}_i) \leq \beta_i \leq \hat{\beta}_i + C.se(\hat{\beta}_i)$. Trong đó

$$C = t_{\frac{\alpha}{2}}^{(n-k)} = t_{0,025}(10-3) = 2,365$$

+ Khoảng tin cậy của β_2

$$0,65005 - 2,365 \cdot 0,25 \leq \beta_2 \leq 0,65005 + 2,365 \cdot 0,25$$

$$\Rightarrow 0,0588 \leq \beta_2 \leq 1,2413$$

+ Khoảng tin cậy của β_3

$$1,10987 - 2,365 \cdot 0,2674 \leq \beta_3 \leq 1,10987 + 2,365 \cdot 0,2674$$

$$\Rightarrow 0,477469 \leq \beta_3 \leq 1,742271$$

4. Tính hệ số xác định mô hình và nêu ý nghĩa. Tính hệ số xác định có hiệu chỉnh

+ $R^2 = \frac{ESS}{TSS} = 0,99164$. Ý nghĩa: cho biết sự biến thiên của phân bón và thuốc trừ sâu giải thích được 99,164% sự biến thiên của năng suất.

$$+ \bar{R}^2 = 1 - (1 - R^2) \frac{n-1}{n-k} = 1 - (1 - 0,99164) \frac{10-1}{10-3} = 0,98925$$

5. Mô hình có phù hợp với thực tế không

+ Đặt giả thiết $H_0 : R^2 = 0; H_1 : R^2 \neq 0$.

+ Với $\alpha = 0,05$, $C = F_\alpha(k-1; n-k) = F_{0,05}(3-1; 10-3) = 4,74$.

$$+ F = \frac{(n-k)R^2}{(k-1)(1-R^2)} = \frac{(10-3)0,99164}{(3-1)(1-0,99164)} = 415,1603$$

+ $F > C$ nên bác bỏ H_0 . Suy ra mô hình phù hợp hay phân bón và thuốc trừ sâu có ảnh hưởng đến năng suất.

6. Kiểm định loại bỏ biến ra khỏi mô hình

• Hệ số xác định của mô hình gốc: $R^2 = 0,99164$

• Hệ số xác định của mô hình đã loại bỏ biến X_3 : $R_X^2 = 0,971$

+ Đặt giả thiết: $H_0 : \beta_3 = 0$; $H_1 : \beta_3 \neq 0$

+ Với $\alpha = 5\%$ suy ra $C = F_\alpha(m, n-k) = F_{0,05}(1, 10-3) = 5,59$.

$$+ F = \frac{(n-k)(R^2 - R_X^2)}{m(1-R^2)} = \frac{(10-3)(0,99164 - 0,971)}{1 \cdot (1-0,99164)} = 17,2823.$$

+ $F > C$ ta bác bỏ H_0 . Vậy không thể loại bỏ X_3 ra khỏi mô hình.

7. Kiểm định về sự ảnh hưởng như nhau của các biến giải thích

+ Đặt giả thiết: $H_0 : \beta_2 - \beta_3 = 0$; $H_1 : \beta_2 - \beta_3 \neq 0$

+ Với $\alpha = 5\%$ suy ra $C = t_{\frac{\alpha}{2}}(n-k) = t_{0,025}(10-3) = 2,365$.

+ Giá trị quan sát

$$\begin{aligned} T &= \frac{\hat{\beta}_2 - \hat{\beta}_3}{se(\hat{\beta}_2 - \hat{\beta}_3)} = \frac{\hat{\beta}_2 - \hat{\beta}_3}{\sqrt{\text{var}(\hat{\beta}_2 - \hat{\beta}_3)}} \\ &= \frac{\hat{\beta}_2 - \hat{\beta}_3}{\sqrt{\text{var}(\hat{\beta}_2) + \text{var}(\hat{\beta}_3) - 2\text{cov}(\hat{\beta}_2, \hat{\beta}_3)}} \\ &= \frac{0,65005 - 1,10987}{\sqrt{0,0625 + 0,0715 - 2(-0,065)}} \\ &= \frac{-0,45982}{0,51381} = -0,895 \end{aligned}$$

+ $|T| < C$ suy ra chưa có cơ sở bác bỏ H_0 . Vậy phân bón và thuốc trừ sâu ảnh hưởng như nhau đến năng suất cây trồng.

8. Khoảng dự báo cho biến phụ thuộc

$$+ \hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_2^0 + \hat{\beta}_3 X_3^0 = 61,63.$$

$$+ C = t_{\frac{\alpha}{2}}^{(n-k)} = t_{0,025}^{(10-3)} = 2,365.$$

Khoảng dự báo cho giá trị trung bình Y_0 ($E(Y/X^0)$)

$$\begin{aligned} \hat{Y}_0 - Cse(\hat{Y}_0) &\leq E(Y/X^0) \leq \hat{Y}_0 + Cse(\hat{Y}_0) \\ \Rightarrow 61,63 - 2,365 \cdot 0,555 &\leq E(Y/X^0) \leq 61,63 + 2,365 \cdot 0,555 \\ \Rightarrow 60,317 &\leq E(Y/X^0) \leq 62,943 \end{aligned}$$

Khoảng dự báo cho giá trị cá biệt Y_0

$$\begin{aligned} \hat{Y}_0 - Cse(Y_0 - \hat{Y}_0) &\leq Y_0 \leq \hat{Y}_0 + Cse(Y_0 - \hat{Y}_0) \\ \Rightarrow 61,63 - 2,365 \cdot 1,5 &\leq Y_0 \leq 61,63 + 2,365 \cdot 1,5 \\ \Rightarrow 58,0825 &\leq Y_0 \leq 65,1775 \end{aligned}$$

Bài 2.8. Bảng số liệu sau đây điều tra ở một hộ gia đình. X_2 là thu nhập từ lương và các khoản có tính chất lương, X_3 là thu nhập ngoài lương, Y là chi tiêu. Đơn vị của các biến đều là triệu đồng

Y	X_2	X_3
9	10	2
9	12	0
11	13	3
12	14	4
12.5	15	4
13	16	6
16	16	8
17	17	9
14	18	5
13.5	20	3

1. Hãy ước lượng mô hình hồi quy

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + U_i$$

2. Tính các giá trị thống kê khác?

3. Giải thích ý nghĩa kinh tế các hệ số hồi quy và hệ số xác định?
4. Kiểm định sự phù hợp của hàm hồi quy ở mức ý nghĩa 5%?
5. Các hệ số hồi quy của mô hình có ý nghĩa thực tế không, với $\alpha = 0.5\%$?
6. Với kết quả hồi quy từ mẫu số liệu trên, nếu cho rằng: cùng một mức tăng thu nhập như nhau, thu nhập ngoài lương tăng sẽ dẫn đến chi tiêu cao hơn so với thu nhập từ lương tăng. Với độ tin cậy 95%, ý kiến này có đúng không?
7. Hãy dự báo giá trị trung bình và cá biệt của chi tiêu khi $X_2 = 19$ và $X_3 = 7$.

Giải

Sử dụng máy tính bỏ túi (570ES, 570ES Plus, 570ES Plus II,...) hoặc phần mềm thống kê (Eviews, SPSS, STATA,...) ta tính được các giá trị sau từ bảng số liệu:

$$\begin{aligned}
 n &= 10; \quad \bar{Y}^2 = 12,7^2 \\
 \sum Y_i &= 127; \quad \sum X_{2i} = 151; \quad \sum X_{3i} = 44; \\
 \sum Y_i^2 &= 1675,5; \quad \sum X_{2i}^2 = 2359; \quad \sum X_{3i}^2 = 260; \\
 \sum X_{2i}X_{3i} &= 702; \quad \sum Y_iX_{2i} = 1971,5; \quad \sum Y_iX_{3i} = 618,5.
 \end{aligned}$$

$$X^T X = \begin{pmatrix} n & \sum X_{2i} & \sum X_{3i} \\ \sum X_{2i} & \sum X_{2i}^2 & \sum X_{2i}X_{3i} \\ \sum X_{3i} & \sum X_{3i}X_{2i} & \sum X_{3i}^2 \end{pmatrix} = \begin{pmatrix} 10 & 151 & 44 \\ 151 & 2359 & 702 \\ 44 & 702 & 260 \end{pmatrix};$$

$$X^T Y = \begin{pmatrix} \sum Y_i \\ \sum X_{2i}Y_i \\ \sum X_{3i}Y_i \end{pmatrix} = \begin{pmatrix} 127 \\ 1971,5 \\ 618,5 \end{pmatrix};$$

$$\hat{\beta} = (X^T X)^{-1} (X^T Y) = \begin{pmatrix} 4,368007 \\ 0,347067 \\ 0,702565 \end{pmatrix} \Rightarrow \begin{cases} \hat{\beta}_1 = 4,368007 \\ \hat{\beta}_2 = 0,347067 \\ \hat{\beta}_3 = 0,702565 \end{cases}$$

$$TSS = \sum Y_i^2 - n\bar{Y}^2 = 62,6;$$

$$ESS = \hat{\beta}^T (X^T Y) - n\bar{Y}^2 = 60,589732;$$

$$RSS = TSS - ESS = 2,010268;$$

$$R^2 = \frac{ESS}{TSS} = 0,96789;$$

$$\hat{\sigma}^2 = \frac{RSS}{n - k} = 0,28718.$$

$$\text{Cov}(\hat{\beta}) = \hat{\sigma}^2 (X^T X)^{-1} = \begin{pmatrix} 0,9049 & -0,06285 & 0,01656 \\ -0,06285 & 0,004985 & -0,002822 \\ 0,01656 & -0,002822 & 0,005923 \end{pmatrix}$$

$$\Rightarrow \begin{cases} se(\hat{\beta}_1) = \sqrt{\text{var}(\hat{\beta}_1)} = \sqrt{0,9049} = 0,95126 \\ se(\hat{\beta}_2) = \sqrt{\text{var}(\hat{\beta}_2)} = \sqrt{0,004985} = 0,0706 \\ se(\hat{\beta}_3) = \sqrt{\text{var}(\hat{\beta}_3)} = \sqrt{0,005923} = 0,07696 \end{cases}$$

$$\text{var}(\hat{Y}_0) = \hat{\sigma}^2 (X^0)^T (X^T X)^{-1} X^0 = 0,0873357 \Rightarrow se(\hat{Y}_0) = 0,295526;$$

$$\text{var}(Y_0 - \hat{Y}_0) = \text{var}(\hat{Y}_0) + \hat{\sigma}^2 = 0,3745157 \Rightarrow se(Y_0 - \hat{Y}_0) = 0,611977.$$

1. Tìm mô hình hồi quy

$$\hat{Y}_i = 4,368007 + 0,347067X_{2i} + 0,702565X_{3i}$$

2. Các giá trị thống kê

$$TSS = 62,6; \quad \hat{\sigma}^2 = 0,28718;$$

$$ESS = 60,589732; \quad se(\hat{\beta}_1) = 0,95126;$$

$$RSS = 2,010268; \quad se(\hat{\beta}_2) = 0,0706$$

$$R^2 = 0,96789; \quad se(\hat{\beta}_3) = 0,07696$$

3. Nêu ý nghĩa kinh tế của các hệ số hồi quy và hệ số xác định

- $\hat{\beta}_1 = 4,368007$: khi không có thu nhập thì mức chi tiêu tối thiểu trung bình khoảng 4,368007 triệu đồng/tháng.
- $\hat{\beta}_2 = 0,347067$: khi thu nhập ngoài lương không đổi, nếu thu nhập từ lương tăng (giảm) 1 triệu đồng/tháng thì chi tiêu bình quân tăng (giảm) 0,347067 triệu đồng /tháng.
- $\hat{\beta}_3 = 0,702565$: khi thu nhập từ lương không đổi, nếu thu nhập ngoài lương tăng (giảm) 1 triệu đồng /tháng thì chi tiêu bình quân tăng (giảm) 0,702565 triệu đồng/tháng.
- Nếu cả thu nhập từ lương và thu nhập ngoài lương cùng tăng như nhau 1 triệu đồng/tháng thì chi tiêu tổng cộng tăng (0,347067+0,702565).

- $R^2 = 0,96789$: sự biến thiên của thu nhập từ lương và thu nhập ngoài lương giải thích được 96,789% sự biến thiên của chi tiêu. Còn lại (1-0,96789)% là do các yếu tố ngẫu nhiên khác giải thích.

4. Mô hình có phù hợp với thực tế không

+ Đặt giả thiết $H_0 : R^2 = 0$; $H_1 : R^2 \neq 0$.

+ Với $\alpha = 0,05$, $C = F_{\alpha}(k-1; n-k) = F_{0,05}(3-1; 10-3) = 4,74$.

$$+ F = \frac{(n-k) R^2}{(k-1)(1-R^2)} = \frac{(10-3) 0,96789}{(3-1)(1-0,96789)} = 105,5$$

+ $F > C$ nên bác bỏ H_0 . Suy ra mô hình phù hợp.

5. Kiểm định ảnh hưởng của biến X trong mô hình

* Kiểm định β_1

+ Đặt giả thiết $H_0 : \beta_1 = 0$; $H_1 : \beta_1 \neq 0$.

$$+ C = t_{\frac{\alpha}{2}}^{(n-k)} = t_{0,025}^{(10-3)} = 2,365.$$

$$+ T_1 = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)} = \frac{4,368007}{0,95126} = 4,592.$$

+ $|T_1| > C$ suy ra bác bỏ H_0 . Vậy hệ số chặn β_1 có ý nghĩa thống kê.

* Kiểm định β_2

+ Đặt giả thiết $H_0 : \beta_2 = 0$; $H_1 : \beta_2 \neq 0$.

$$+ C = t_{\frac{\alpha}{2}}^{(n-k)} = t_{0,025}^{(10-3)} = 2,365.$$

$$+ T_2 = \frac{\hat{\beta}_2}{se(\hat{\beta}_2)} = \frac{0,347067}{0,0706} = 4,916.$$

+ $|T_2| > C$ suy ra bác bỏ H_0 . Vậy hệ số β_2 có ý nghĩa thống kê.

* Kiểm định β_3

+ Đặt giả thiết $H_0 : \beta_3 = 0$; $H_1 : \beta_3 \neq 0$.

$$+ C = t_{\frac{\alpha}{2}}^{(n-k)} = t_{0,025}^{(10-3)} = 2,365.$$

$$+ T_2 = \frac{\hat{\beta}_3}{se(\hat{\beta}_3)} = \frac{0,702565}{0,07696} = 9,12896.$$

+ $|T_2| > C$ suy ra bác bỏ H_0 . Vậy hệ số β_3 có ý nghĩa thống kê.

6. Kiểm định về sự ảnh hưởng như nhau của các biến giải thích

Yêu vầu bài toán tương đương β_3 có thực sự lớn hơn β_2 không

+ Đặt giả thiết: $H_0 : \beta_3 - \beta_2 = 0$; $H_1 : \beta_3 - \beta_2 > 0$

+ Với $\alpha = 5\%$ suy ra $C = t_{\frac{\alpha}{2}}(n - k) = t_{0,025}(10 - 3) = 2,365$.

+ Giá trị quan sát

$$\begin{aligned} T &= \frac{\hat{\beta}_3 - \hat{\beta}_2}{se(\hat{\beta}_3 - \hat{\beta}_2)} = \frac{\hat{\beta}_3 - \hat{\beta}_2}{\sqrt{\text{var}(\hat{\beta}_3 - \hat{\beta}_2)}} \\ &= \frac{\hat{\beta}_3 - \hat{\beta}_2}{\sqrt{\text{var}(\hat{\beta}_2) + \text{var}(\hat{\beta}_3) - 2\text{cov}(\hat{\beta}_3, \hat{\beta}_2)}} \\ &= \frac{0,702565 - 0,347067}{\sqrt{0,005923 + 0,004985 - 2(-0,002822)}} \\ &= 2,7632 \end{aligned}$$

+ $|T| > C$ suy ra bác bỏ H_0 . Vậy β_3 thực sự lớn hơn β_2 .

7. Khoảng dự báo cho biến phụ thuộc

$$+ \hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_2^0 + \hat{\beta}_3 X_3^0 = 15,880235.$$

$$+ C = t_{\frac{\alpha}{2}}^{(n-k)} = t_{0,025}^{(10-3)} = 2,365.$$

Khoảng dự báo cho giá trị trung bình Y_0 ($E(Y/X^0)$)

$$\begin{aligned} \hat{Y}_0 - C se(\hat{Y}_0) &\leq E(Y/X^0) \leq \hat{Y}_0 + C se(\hat{Y}_0) \\ \Rightarrow 15,880235 - 2,365 \cdot 0,295526 &\leq E(Y/X^0) \leq 15,880235 + 2,365 \cdot 0,295526 \\ \Rightarrow 15,1813 &\leq E(Y/X^0) \leq 16,57915 \end{aligned}$$

Khoảng dự báo cho giá trị cá biệt Y_0

$$\begin{aligned}\widehat{Y}_0 - C_{se} (Y_0 - \widehat{Y}_0) &\leq Y_0 \leq \widehat{Y}_0 + C_{se} (Y_0 - \widehat{Y}_0) \\ \Rightarrow 15,880235 - 2,365.0,611977 &\leq Y_0 \leq 15,880235 + 2,365.0,611977 \\ \Rightarrow 14,4334 &\leq Y_0 \leq 17,32707\end{aligned}$$

Bài 2.09. Người ta cho rằng tổng vốn đầu tư (Y : tỉ đồng) không chỉ phụ thuộc vào lãi suất ngân hàng (X_2 : %) mà còn phụ thuộc vào tốc độ tăng trưởng GDP (X_3 : %). Với số liệu gồm có 20 quan sát, người ta ước lượng được mô hình:

$$\begin{aligned}\widehat{Y} &= 40.815 - 1.012X_2 + 2.123X_3 & R^2 &= 0,901 \\ t &= (2.748) \quad (-2,842) \quad (3.485)\end{aligned}$$

1. Hãy nêu ý nghĩa kinh tế của các hệ số hồi quy?
2. Tìm khoảng tin cậy của các hệ số hồi quy tổng thể với độ tin cậy 95%?
3. Tính hệ số xác định có hiệu chỉnh.
4. Kiểm định sự phù hợp của mô hình với mức ý nghĩa 5%.

Giải

a) Ý nghĩa của các hệ số hồi quy

+ Khi tốc độ tăng trưởng GDP không đổi, lãi suất tăng (hoặc giảm) 1% thì vốn đầu tư trung bình giảm (hoặc tăng) 1,012 tỉ đồng.

+ Khi lãi suất không đổi, tốc độ tăng trưởng của GDP tăng (hoặc giảm) 1% thì vốn đầu tư trung bình tăng (hoặc giảm) 2,123 tỉ đồng.

b) Khoảng tin cậy của các hệ số hồi quy

Áp dụng: $\widehat{\beta}_i - C.se(\widehat{\beta}_i) \leq \beta_i \leq \widehat{\beta}_i + C.se(\widehat{\beta}_i)$. Trong đó

$$C = t_{\frac{\alpha}{2}}^{(n-k)} = t_{0,025}^{(20-3)} = 2,11$$

$$t_i = \frac{\hat{\beta}_i}{se(\hat{\beta}_i)} \Rightarrow se(\hat{\beta}_i) = \frac{\hat{\beta}_i}{t_i}$$

$$\Rightarrow \begin{cases} se(\hat{\beta}_1) = \frac{\hat{\beta}_1}{t_1} = \frac{40,815}{2,748} = 14,8526 \\ se(\hat{\beta}_2) = \frac{\hat{\beta}_2}{t_2} = \frac{-1,012}{-2,842} = 0,3561 \\ se(\hat{\beta}_3) = \frac{\hat{\beta}_3}{t_3} = \frac{2,123}{3,485} = 0,6092 \end{cases}$$

+ Khoảng tin cậy của β_1

$$40,815 - 2,11.14,8526 \leq \beta_1 \leq 40,815 + 2,11.14,8526$$

$$\Rightarrow 9,476 \leq \beta_1 \leq 72,154$$

+ Khoảng tin cậy của β_2

$$-1,012 - 2,11.0,3561 \leq \beta_2 \leq -1,012 + 2,11.0,3561$$

$$\Rightarrow -1,763 \leq \beta_2 \leq -0,26063$$

+ Khoảng tin cậy của β_3

$$2,123 - 2,11.0,6092 \leq \beta_3 \leq 2,123 + 2,11.0,6092$$

$$\Rightarrow 0,838 \leq \beta_3 \leq 3,408$$

c) Hệ số xác định có hiệu chỉnh

$$\bar{R}^2 = 1 - (1 - R^2) \frac{n-1}{n-k} = 1 - (1 - 0,901) \frac{19}{17} = 0,8894$$

d) Kiểm định sự phù hợp của mô hình

+ Đặt giả thiết $H_0 : R^2 = 0; H_1 : R^2 \neq 0$.

+ Với $\alpha = 0,05, C = F_{\alpha}(k-1; n-k) = F_{0,05}(2; 17) = 3,59$.

$$+ F = \frac{(n-k)R^2}{(k-1)(1-R^2)} = \frac{(20-3)0,901}{(3-1)(1-0,901)} = 77,36$$

+ $F > C$ nên bác bỏ H_0 . Vậy mô hình phù hợp.

Bài 2.10. Từ một mẫu gồm 10 quan sát, người ta tiến hành hồi quy Y theo X_2 và X_3 ta có kết quả sau:

$$\begin{aligned}\widehat{Y} &= 99.09 - 0.39X_2 - 13.195X_3 & R^2 &= 0,8766 \\ Se &= (70.4) & (2.11) & (13.33)\end{aligned}$$

1. Tìm khoảng tin cậy cho hệ số hồi quy của biến X với độ tin cậy 98%?
2. Kiểm định giả thiết cho rằng hệ số hồi quy của biến X_3 trong hàm hồi quy tổng thể là -12 với mức ý nghĩa 5%.

Giải

1. Khoảng tin cậy cho hệ số hồi quy

+ Áp dụng: $\widehat{\beta}_i - C.se(\widehat{\beta}_i) \leq \beta_i \leq \widehat{\beta}_i + C.se(\widehat{\beta}_i)$. Trong đó

$$C = t_{\frac{\alpha}{2}}^{(n-k)} = t_{0,01}^{(10-3)} = 2,998$$

+ Khoảng tin cậy của β_1

$$\begin{aligned}-0,39 - 2,998.2,11 &\leq \beta_1 \leq -0,39 + 2,998.2,11 \\ \Rightarrow -6,71578 &\leq \beta_1 \leq 5,93578\end{aligned}$$

2. Kiểm định giả thiết

+ Đặt giả thiết $H_0 : \beta_3 = -12; H_1 : \beta_3 \neq -12$.

+ $C = t_{\frac{\alpha}{2}}^{(n-k)} = t_{0,025}^{(10-3)} = 2,365$.

$$+ T = \frac{\widehat{\beta}_3 - (-12)}{se(\widehat{\beta}_3)} = \frac{-13,195 - (-12)}{13,33} = -0,0896.$$

+ $|T| < C$ suy ra chấp nhận H_0 .

2.3 Hồi quy với biến định tính

Bài 2.11. Khảo sát về năng suất của hai công nghệ sản xuất, người ta thu được số liệu cho ở bảng sau

D	0	1	1	0	0	1	0	1	1	0
Y	28	32	35	27	25	37	29	34	33	30

Trong đó $Y_i (i = 1, 2, \dots, 10)$ là năng suất 1 ngày (đơn vị: tấn). $D = 1$ nếu là công nghệ A, $D = 0$ nếu là công nghệ B. Ước lượng mô hình hồi quy tuyến tính và nêu ý nghĩa.

Giải

Ta có $\hat{\beta}_1 = 27,8$; $\hat{\beta}_2 = 6,4$.

Vậy $\hat{Y}_i = 27,8 + 6,4D$

Ý nghĩa:

+ Nếu $D = 0$, năng suất trung bình của công nghệ B là 27,8 (tấn/ngày)

$$E(\hat{Y}_i/D = 0) = 27,8$$

+ Nếu $D = 1$, năng suất trung bình của công nghệ B là $(27,8 + 6,4) = 34,2$ (tấn/ngày)

$$E(\hat{Y}_i/D = 1) = 34,2$$

Bài 2.12. Số liệu về tiết kiệm và thu nhập cá nhân ở nước Anh từ năm 1946 đến 1963 (đơn vị Pound) cho ở bảng sau

Thời kỳ I	Y	X	Thời kỳ II	Y	X
1946	0.36	8.8	1955	0.59	15.5
1947	0.21	9.4	1956	0.9	16.7
1948	0.08	10	1957	0.95	17.7
1949	0.2	10.6	1958	0.82	18.6
1950	0.1	11	1959	1.04	19.7
1951	0.12	11.9	1960	1.53	21.1
1952	0.41	12.7	1961	1.94	22.8
1953	0.5	13.5	1962	1.75	23.9
1954	0.43	14.3	1963	1.99	25.2

Trong đó Y tiết kiệm; X thu nhập. Xét xem tiết kiệm ở hai thời kỳ có như nhau hay không với mức ý nghĩa 5%.

Giải

+ Mô hình hồi quy gốc

$$\widehat{Y}_i = -1,082071 + 0,117845X_i \Rightarrow RSS = n(1 - r^2) \text{ var}(Y) = 0,572226$$

+ Mô hình hồi quy ở thời kỳ I

$$\widehat{Y}_i = -0,266249 + 0,047028X_i \Rightarrow RSS_1 = n_1(1 - r_1^2) \text{ var}(Y) = 0,13965$$

+ Mô hình hồi quy ở thời kỳ II

$$\widehat{Y}_i = -1,750172 + 0,15045X_i \Rightarrow RSS_2 = n_2(1 - r_2^2) \text{ var}(Y) = 0,193121$$

$$+ \overline{RSS} = RSS_1 + RSS_2 = 0,332771.$$

Kiểm định giả thiết

+ Đặt giả thiết: H_0 tiết kiệm ở 2 thời kỳ là như nhau; H_1 tiết kiệm ở 2 thời kỳ là khác nhau.

$$+ C = F_\alpha(k; n_1 + n_2 - 2k) = F_{0,05}(2; 14) = 3,74$$

+ Giá trị quan sát

$$\begin{aligned} F &= \frac{(RSS - \overline{RSS})(n_1 + n_2 - 2k)}{k \cdot \overline{RSS}} \\ &= \frac{(0,572226 - 0,332771)(9 + 9 - 2 \cdot 2)}{2 \cdot 0,332771} \\ &= 5,03705 \end{aligned}$$

+ $F > C$ suy ra bác bỏ H_0 .

Vậy tiết kiệm ở 2 thời kỳ là khác nhau với mức ý nghĩa 5%.

Bài 2.13. Người ta cho rằng chi tiêu mặt hàng A (Y - triệu đồng/tháng) không chỉ phụ thuộc vào thu nhập của người tiêu dùng (X - triệu đồng/tháng) mà còn phụ thuộc vào giới tính của người tiêu dùng (D = 1 nếu là nam và D = 0 nếu là nữ). Với số liệu của một mẫu gồm 20 quan sát, người ta đã ước lượng được mô hình

$$\begin{aligned} \widehat{Y} &= 6.426 + 0.098X + 2.453D - 0.025XD \\ Se &= (3.628) \quad (0.032) \quad (0.988) \quad (0.011) \end{aligned}$$

1. Hãy nêu ý nghĩa của hệ số hồi quy của biến D và biến XD?

2. Tìm khoảng tin cậy của các hệ số hồi quy trong hàm hồi quy tổng thể với độ tin cậy 95%?
3. Hãy cho biết chi tiêu về mặt hàng A của nam và nữ có giống nhau không? (với mức ý nghĩa 5%)

Giải

a) Ý nghĩa của các hệ số hồi quy của biến D và biến XD

+ Hệ số hồi quy của biến D là 2,453 cho biết mức chênh lệch của hệ số **tung độ góc** giữa hai hàm hồi quy phản ánh mối quan hệ giữa chi tiêu mặt hàng A đối với thu nhập của nam và nữ.

+ Hệ số hồi quy của biến XD là -0,025 cho biết mức chênh lệch của **hệ số góc** giữa hai hàm hồi quy phản ánh mối quan hệ giữa chi tiêu mặt hàng A đối với thu nhập của nam và nữ.

b) Khoảng tin cậy của các hệ số hồi quy

Áp dụng: $\hat{\beta}_i - C.se(\hat{\beta}_i) \leq \beta_i \leq \hat{\beta}_i + C.se(\hat{\beta}_i)$. Trong đó

$$C = t_{\frac{\alpha}{2}}^{(n-k)} = t_{0,025}^{(20-4)} = 2,12$$

+ Khoảng tin cậy của β_1

$$6,426 - 2,12 \cdot 3,628 \leq \beta_1 \leq 6,426 + 2,12 \cdot 3,628$$

$$\Rightarrow 1,26536 \leq \beta_1 \leq 14,11736$$

+ Khoảng tin cậy của β_2

$$0,098 - 2,12 \cdot 0,032 \leq \beta_2 \leq 0,098 + 2,12 \cdot 0,032$$

$$\Rightarrow 0,03016 \leq \beta_2 \leq 0,16584$$

+ Khoảng tin cậy của β_3

$$2,453 - 2,12 \cdot 0,988 \leq \beta_3 \leq 2,453 + 2,12 \cdot 0,988$$

$$\Rightarrow 0,35844 \leq \beta_3 \leq 4,54756$$

+ Khoảng tin cậy của β_4

$$-0,025 - 2,12 \cdot 0,011 \leq \beta_4 \leq -0,025 + 2,12 \cdot 0,011$$

$$\Rightarrow -0,048 \leq \beta_4 \leq -0,00168$$

c) Chi tiêu của nam và nữ có giống nhau hay không?

• Xét β_3

+ Đặt giả thiết $H_0 : \beta_3 = 0; H_1 : \beta_3 \neq 0$.

+ $C = t_{\frac{\alpha}{2}}^{(n-k)} = t_{0,025} (20 - 4) = 2,12$.

+ $T_3 = \frac{\hat{\beta}_3}{se(\hat{\beta}_3)} = \frac{2,453}{0,988} = 2,4828$.

+ $|T_3| > C$ suy ra bác bỏ H_0 (1)

• Xét β_4

+ Đặt giả thiết $H_0 : \beta_4 = 0; H_1 : \beta_4 \neq 0$.

+ $C = t_{\frac{\alpha}{2}}^{(n-k)} = t_{0,025} (20 - 4) = 2,12$.

+ $T_4 = \frac{\hat{\beta}_4}{se(\hat{\beta}_4)} = \frac{2,453}{0,988} = -2,273$.

+ $|T_4| > C$ suy ra bác bỏ H_0 (2)

Từ (1) và (2) ta suy ra chi tiêu của nam và nữ là khác nhau.

Bài 2.14. Người ta cho rằng chi tiêu mặt hàng A (Y - ngàn đồng/tháng) không chỉ phụ thuộc vào thu nhập của người tiêu dùng (X - triệu đồng/tháng) mà còn phụ thuộc vào giới tính của người tiêu dùng ($D = 1$ nếu là nam và $D = 0$ nếu là nữ). Với số liệu của một mẫu gồm 20 quan sát, người ta đã ước lượng được mô hình

$$\begin{aligned}\hat{Y} &= 96.458 + 38.928X - 8.415D - 6.525XD \\ Se &= (33.228) \quad (11.312) \quad (4.207) \quad (1.812)\end{aligned}$$

1. Hãy nêu ý nghĩa của các hệ số hồi quy?
2. Tìm khoảng tin cậy của các hệ số hồi quy trong hàm hồi quy tổng thể với độ tin cậy 95%?
3. Hãy cho biết chi tiêu về mặt hàng A của nam và nữ có giống nhau không? (với mức ý nghĩa 5%) Vì sao?

Giải

a) Ý nghĩa của các hệ số hồi quy

+ Người tiêu dùng là nữ: khi thu nhập của nữ tăng 1 triệu đồng/tháng thì mức chi tiêu cho mặt hàng A trung bình tăng 38,928 ngàn đồng/tháng

+ Người tiêu dùng là nam: khi thu nhập của nam tăng 1 triệu đồng/tháng thì mức chi tiêu cho mặt hàng A trung bình tăng $(38,928 - 6,525) = 32,403$ ngàn đồng/tháng

+ Với cùng một mức thu nhập thì chi tiêu trung bình của về mặt hàng A của nữ cao hơn nam $(8,415 + 6,525)$ ngàn đồng/tháng.

b) Khoảng tin cậy của các hệ số hồi quy

Áp dụng: $\widehat{\beta}_i - C.se(\widehat{\beta}_i) \leq \beta_i \leq \widehat{\beta}_i + C.se(\widehat{\beta}_i)$. Trong đó

$$C = t_{\frac{\alpha}{2}}^{(n-k)} = t_{0,025}^{(20-4)} = 2,12$$

+ Khoảng tin cậy của β_1

$$\begin{aligned} 96,458 - 2,12.33,228 &\leq \beta_1 \leq 96,458 + 2,12.33,228 \\ \Rightarrow 26,01464 &\leq \beta_1 \leq 166,90136 \end{aligned}$$

+ Khoảng tin cậy của β_2

$$\begin{aligned} 38,928 - 2,12.11,312 &\leq \beta_2 \leq 38,928 + 2,12.11,312 \\ \Rightarrow 14,94656 &\leq \beta_2 \leq 62,90944 \end{aligned}$$

+ Khoảng tin cậy của β_3

$$\begin{aligned} 8,415 - 2,12.4,207 &\leq \beta_3 \leq 8,415 + 2,12.4,207 \\ \Rightarrow -17,334 &\leq \beta_3 \leq 0,50384 \end{aligned}$$

+ Khoảng tin cậy của β_4

$$\begin{aligned} -6,525 - 2,12.1,812 &\leq \beta_4 \leq -6,525 + 2,12.1,812 \\ \Rightarrow -10,366 &\leq \beta_4 \leq -2,68356 \end{aligned}$$

c) Chi tiêu của nam và nữ có giống nhau hay không?

• Xét β_3

+ Đặt giả thiết $H_0 : \beta_3 = 0; H_1 : \beta_3 \neq 0$.

+ $C = t_{\frac{\alpha}{2}}^{(n-k)} = t_{0,025} (20 - 4) = 2, 12$.

$$+ T_3 = \frac{\hat{\beta}_3}{se(\hat{\beta}_3)} = \frac{-8, 415}{4, 207} = -2, 0002.$$

+ $|T_3| < C$ suy ra chưa có cơ sở bác bỏ H_0 (1)

• Xét β_4

+ Đặt giả thiết $H_0 : \beta_4 = 0; H_1 : \beta_4 \neq 0$.

+ $C = t_{\frac{\alpha}{2}}^{(n-k)} = t_{0,025} (20 - 4) = 2, 12$.

$$+ T_4 = \frac{\hat{\beta}_4}{se(\hat{\beta}_4)} = \frac{-6, 525}{1, 812} = -3, 6009.$$

+ $|T_4| > C$ suy ra bác bỏ H_0 (2)

Từ (1) và (2) ta suy ra chi tiêu của nam và nữ là khác nhau.

Bài 2.15. Xét hàm hồi quy mẫu $\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X + \hat{\beta}_3 Z$ trong đó: Y là mức chi tiêu cho mặt hàng A (đơn vị 100 ngàn đồng/tháng); Z là giới tính ($Z = 1$ nếu là nam, $Z = 0$ nếu là nữ). Từ số liệu của một mẫu (kích thước $n = 20$) người ta tìm được kết quả như sau

$$\begin{aligned} \hat{Y} &= -4.1365 + 0.5133X + 0.2053Z + 0.325XZ \\ t &= (-4.889) \quad (11.35) \quad (0.557) \quad (2.42) \\ R^2 &= 0.7485; \quad d = 2.07 \end{aligned}$$

1. Với mức ý nghĩa 5%, hãy xét xem mô hình trên có xảy ra hiện tượng tự tương quan hay không?
2. Kiểm định giả thiết $H_0 : \beta_2 = 0.6$ với mức ý nghĩa 5%.
3. Xét xem chi tiêu về mặt hàng A của nam và nữ có khác nhau không? ($\alpha = 5\%$).

Giải

1. Ta có $d = 2,07$, nhận thấy $1 < d < 3$ nên theo quy tắc kiểm định Durbin Watson giản đơn ta có thể kết luận là mô hình không xảy ra hiện tượng tự tương quan.

2. Kiểm định giả thiết

+ Đặt giả thiết $H_0 : \beta_2 = 0,6$; $H_1 : \beta_2 \neq 0,6$.

+ $C = t_{\frac{\alpha}{2}}^{(n-k)} = t_{0,025}(20 - 4) = 2,12$.

+ Chú ý rằng

$$t = \frac{\hat{\beta}_2}{se(\hat{\beta}_2)} \Rightarrow se(\hat{\beta}_2) = \frac{\hat{\beta}_2}{t} = \frac{0,5133}{11,35} = 0,0452$$

$$+ T = \frac{\hat{\beta}_2 - (0,6)}{se(\hat{\beta}_2)} = \frac{0,5133 - (0,6)}{0,0452} = -1,92.$$

+ $|T| < C$ suy ra chấp nhận H_0 .

3. Chi tiêu về mặt hàng A của nam và nữ có giống nhau hay không?

- Xét β_3

+ Đặt giả thiết $H_0 : \beta_3 = 0$; $H_1 : \beta_3 \neq 0$.

+ $C = t_{\frac{\alpha}{2}}^{(n-k)} = t_{0,025}(20 - 4) = 2,12$.

+ $T_3 = 0,557$.

+ $|T_3| < C$ suy ra chưa có cơ sở bác bỏ H_0 (1)

- Xét β_4

+ Đặt giả thiết $H_0 : \beta_4 = 0$; $H_1 : \beta_4 \neq 0$.

+ $C = t_{\frac{\alpha}{2}}^{(n-k)} = t_{0,025}(20 - 4) = 2,12$.

+ $T_4 = 2,42$.

+ $|T_4| > C$ suy ra bác bỏ H_0 (2)

Từ (1) và (2) ta suy ra chi tiêu về mặt hàng A của nam và nữ là khác nhau.

2.4 Bài tập tổng hợp

Bài 2.16. Khảo sát về nhu cầu tiêu thụ Cafe thông qua số tách Cafe 1 người dùng mỗi ngày Y (tách/người/ngày) và giá bán lẻ trung bình của Cafe X (USD/pao), người ta thu được bảng số liệu

Năm	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980
Y	2.57	2.5	2.35	2.3	2.25	2.2	2.11	1.94	1.97	2.06	2.02
X	0.77	0.74	0.72	0.73	0.76	0.76	1.08	1.81	1.39	1.2	1.17

- Ước lượng mô hình (1): $Y_i = \beta_1 + \beta_2 X_i + U_i$. Nêu ý nghĩa của hệ số góc hồi quy được?
- Tìm khoảng tin cậy cho β_2 với độ tin cậy 95%.
- Kiểm định giả thiết $H_0 : \beta_2 = 0$ với mức ý nghĩa 5%.
- Với mức ý nghĩa 5%, hãy cho biết $\beta_2 < 0.4$ hay không?
- Dự báo nhu cầu tiêu thụ Cafe trung bình (hoặc cá biệt) khi giá bán lẻ trung bình là 1 USD/pao với độ tin cậy 95%.
- Tính R^2 , nêu ý nghĩa. Kiểm định sự phù hợp của mô hình với mức ý nghĩa 5%.
- Nêu ý nghĩa của các hệ số góc trong các mô hình hồi quy sau:
 - $\ln(Y_i) = 0.7774 - 0.2530 \ln(X_i)$.
 - $Y_i = 2.1848 - 0.5520 \ln(X_i)$.
 - $\ln(Y_i) = 1.0100 - 0.2202 X_i$.
- Xét thêm yếu tố khuyến mãi Z được quy ước như sau ($Z = 1$: có khuyến mãi, $Z = 0$: không có khuyến mãi). Kết quả hồi quy mô hình (2):

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 Z_i + U_i$$

Ta được

$$Y_i = 2.6144 - 0.4789 X_i + 0.2093 Z_i + e_i$$

$$R^2 = 0.915$$

- Kiểm định sự phù hợp của mô hình (2) với mức ý nghĩa 5%.
- Tính \bar{R}^2 của hai mô hình và cho biết mô hình nào phù hợp hơn (so sánh hai mô hình).

c. Nêu ý nghĩa của hệ số góc của biến giả Z hồi quy được?

Giải

$$\hat{\beta}_2 = \frac{\sum X_i Y_i - n \bar{X} \bar{Y}}{\sum X_i^2 - n(\bar{X})^2} = -0,479529;$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} = 2,691124;$$

$$TSS = n \text{var}(Y) = 0,4421;$$

$$ESS = n \hat{\beta}_2^2 \text{var}(X) = 0,293;$$

$$RSS = TSS - ESS = n(1 - r^2) \text{var}(Y) = 0,1491;$$

$$R^2 = \frac{ESS}{TSS} = 0,6627;$$

$$\hat{\sigma}^2 = \frac{n}{n-2} (1 - r^2) \text{var}(Y) = \frac{RSS}{n-2} = 0,128703.$$

$$\text{var}(\hat{\beta}_2) = \frac{\hat{\sigma}^2}{n \text{var}(X)} = 0,013;$$

$$se(\hat{\beta}_2) = \sqrt{\text{var}(\hat{\beta}_2)} = 0,11402;$$

$$\text{var}(\hat{Y}_0) = \left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{n \text{var}(X)} \right] \hat{\sigma}^2 = 0,0015;$$

$$se(\hat{Y}_0) = \sqrt{\text{var}(\hat{Y}_0)} = 0,0388;$$

$$\text{var}(Y_0 - \hat{Y}_0) = \hat{\sigma}^2 + \text{var}(\hat{Y}_0) = 0,01806;$$

$$se(Y_0 - \hat{Y}_0) = \sqrt{\text{var}(Y_0 - \hat{Y}_0)} = 0,1344.$$

1. Ước lượng mô hình

$$\hat{Y} = 2,6911 - 0,4795X$$

Ý nghĩa: khi giá bán trung bình của cafe tăng 1 (USD/pao) thì nhu cầu tiêu thụ cafe trung bình giảm 0,4795 (tách/người/ngày).

2. Khoảng tin cậy cho β_2 với độ tin cậy 95%

Áp dụng: $\hat{\beta}_i - C.se(\hat{\beta}_i) \leq \beta_i \leq \hat{\beta}_i + C.se(\hat{\beta}_i)$. Trong đó

$$C = t_{\frac{\alpha}{2}}^{(n-k)} = t_{0,025}(11-2) = 2,262$$

Khoảng tin cậy của β_2

$$-0,479529 - 2,262.0,11402 \leq \beta_2 \leq -0,479529 + 2,262.0,11402$$

$$\Rightarrow -0,7374 \leq \beta_2 \leq -0,2216$$

3. Kiểm định giả thiết $H_0 : \beta_2 = 0$ với mức ý nghĩa 5%.

♠ Cách 1: khoảng tin cậy 95% của β_2 là $(-0,7374; -0,2216)$ không chứa 0, nên ta có cơ sở để bác bỏ H_0 với mức ý nghĩa 5%. Vậy $\beta_2 \neq 0$ với mức ý nghĩa 5%.

♠ Cách 2 + Đặt giả thiết $H_0 : \beta_2 = 0; H_1 : \beta_2 \neq 0$.

$$+ C = t_{\frac{\alpha}{2}}^{(n-k)} = t_{0,025} (11 - 2) = 2,262.$$

$$+ T = \frac{\hat{\beta}_2}{se(\hat{\beta}_2)} = -4,205592.$$

+ $|T| > C$ suy ra bác bỏ H_0 . Vậy $\beta_2 \neq 0$ với mức ý nghĩa 5%.

4. Với mức ý nghĩa 5%, hãy cho biết $\beta_2 < 0.4$ hay không?

+ Đặt giả thiết $H_0 : \beta_2 = -0,4; H_1 : \beta_2 < -0,4$.

$$+ C = t_{\alpha} (n - k) = t_{0,05} (11 - 2) = 1,833.$$

$$+ T = \frac{\hat{\beta}_2}{se(\hat{\beta}_2)} = \frac{1,2494}{0,03883} = 32,2.$$

+ $T > -C$ suy ra chưa có cơ sở bác bỏ H_0 . Vậy $\beta_2 < -0,4$ với mức ý nghĩa 5%.

5. Dự báo nhu cầu tiêu thụ Cafe

$$+ \hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_0 = 2,6911 + 0,4795.1 = 2,2116.$$

$$+ C = t_{\frac{\alpha}{2}}^{(n-k)} = t_{0,025} (11 - 2) = 2,262.$$

Khoảng dự báo cho giá trị trung bình Y_0 ($E(Y/X_0 = 5)$)

$$\hat{Y}_0 - C se(\hat{Y}_0) \leq E(Y/X_0 = 1) \leq \hat{Y}_0 + C se(\hat{Y}_0)$$

$$\Rightarrow 2,1238 \leq E(Y/X_0 = 1) \leq 2,2994$$

Khoảng dự báo cho giá trị cá biệt Y_0

$$\begin{aligned}\widehat{Y}_0 - C_{se} (Y_0 - \widehat{Y}_0) &\leq Y_0 \leq \widehat{Y}_0 + C_{se} (Y_0 - \widehat{Y}_0) \\ \Rightarrow 1,9076 &\leq Y_0 \leq 2,5156\end{aligned}$$

6. Tính R^2 , nêu ý nghĩa. Kiểm định sự phù hợp của mô hình với mức ý nghĩa 5%

♠ $R^2 = \frac{ESS}{TSS} = 0,6627$. Ý nghĩa: giá bán lẻ trung bình của cafe giải thích được 66,28% sự thay đổi giá trị của nhu cầu tiêu thụ cafe theo mô hình hồi quy tuyến tính, còn lại 33,73% do các yếu tố khác ngoài mô hình tác động.

♠ Kiểm định:

+ Đặt giả thiết $H_0 : R^2 = 0$; $H_1 : R^2 > 0$.

+ Với $\alpha = 0,05$, $C = F_{\alpha}(k-1; n-k) = F_{0,05}(2-1; 11-2) = 5,12$.

$$+ F = \frac{(n-k)R^2}{(k-1)(1-R^2)} = 17,687$$

+ $F > C$ nên bác bỏ H_0 . Vậy mô hình phù hợp với mức ý nghĩa 5%.

7. a. Khi giá bán lẻ trung bình của cafe tăng 1% thì nhu cầu tiêu thụ cafe trung bình giảm 0,253%.
- b. Khi giá bán lẻ trung bình của cafe tăng 1% thì nhu cầu tiêu thụ cafe trung bình giảm 0,00552 (tách/người/ngày).
- c. Khi giá bán lẻ trung bình của cafe tăng 1USD/pao thì nhu cầu tiêu thụ cafe trung bình giảm 22,02%.

8. a. Kiểm định sự phù hợp của mô hình (2) với mức ý nghĩa 5%

+ Đặt giả thiết $H_0 : R^2 = 0$; $H_1 : R^2 > 0$.

+ Với $\alpha = 0,05$, $C = F_{\alpha}(k-1; n-k) = F_{0,05}(2-1; 11-3) = 4,46$.

$$+ F = \frac{(n-k)R^2}{(k-1)(1-R^2)} = 43,0823$$

+ $F > C$ nên bác bỏ H_0 . Vậy mô hình (2) phù hợp với mức ý nghĩa 5%.

b. Tính $\overline{R^2}$ của hai mô hình và cho biết mô hình nào phù hợp hơn

$$+ \overline{R_1^2} = 1 - (1 - R^2) \frac{n-1}{n-k} = 1 - (1 - 0,6628) \frac{11-1}{11-2} = 0,6253.$$

$$+ \overline{R_2^2} = 1 - (1 - R^2) \frac{n-1}{n-k} = 1 - (1 - 0,915) \frac{11-1}{11-3} = 0,8938.$$

Ta nhận thấy $\overline{R_2^2} > \overline{R_1^2}$ nên mô hình (2) phù hợp hơn so với mô hình (1).

c. Ý nghĩa của hệ số góc của biến giả Z hồi quy được?

$\hat{\beta}_3 = 0,2093$: khi có khuyến mãi thì nhu cầu tiêu thụ cafe trung bình tăng 0,2093 (tách/người/ngày) so với khi không có khuyến mãi, với điều kiện giá bán lẻ trung bình của cafe không đổi.

Bài 2.17. Cho một mẫu thống kê như sau:

Y_i	30	50	40	55	50	60	58	62	60	65
X_i	7	8	9	9	11	12	13	13	14	15
G. tính	Nam	Nữ	Nam	Nữ	Nam	Nữ	Nam	Nữ	Nam	Nữ

Trong đó: Y là chi tiêu về mặt hàng A (đơn vị: 100 ngàn đồng/tháng); X là thu nhập của người tiêu dùng (triệu đồng/tháng).

Câu 1. Hồi quy Y theo X ta được kết quả cho ở bảng sau:

Dependent Variable: Y					
Variable	Coefficient	Std. Error	t-Statistic	Prob.	
C	15.33632	8.501202	1.804018	0.1089	
X	3.393124	0.745892	4.549084	0.0019	
R-squared	0.721198	Mean dependent var	53.00000		
Adjusted R-squared	0.686347	Akaike info criterion	6.631583		
S.E. of regression	6.100828	Schwarz criterion	6.692100		
Log likelihood	-31.15791	F-statistic	20.69417		
Durbin-Watson stat	3.155164	Prob(F-statistic)	0.001877		

- Viết mô hình hồi quy tuyến tính mẫu của Y theo X và nêu ý nghĩa của hệ số góc.
- Kiểm định hệ số hồi quy của biến X trong hàm hồi quy tổng thể bằng 0 với mức ý nghĩa 5% và cho biết ý nghĩa của kết quả kiểm định.

- c) Viết hàm hồi quy khi đơn vị của chi tiêu và thu nhập đều là triệu đồng/năm.
- d) Tính hệ số co dẫn tại điểm (\bar{X}, \bar{Y}) và nêu ý nghĩa.
- e) Dự báo mức chi tiêu trung bình của một người có thu nhập là 10 triệu đồng/tháng với độ tin cậy 95%.

Câu 2. Đặt $Z_i = 0$ nếu là nam; $Z_i = 1$ nếu là nữ. Hồi quy Y theo X và Z ta được kết quả cho ở bảng sau:

Dependent Variable: Y					
Variable	Coefficient	Std. Error	t-Statistic	Prob.	
C	13.10545	5.383313	2.434459	0.0451	
X	3.193939	0.472439	6.760539	0.0003	
Z	8.883636	2.443928	3.634983	0.0083	
R-squared	0.903448	Mean dependent var	53.00000		
Adjusted R-squared	0.875862	Akaike info criterion	5.771162		
S.E. of regression	3.838109	Schwarz criterion	5.861937		
Log likelihood	-25.85581	F-statistic	32.74988		
Durbin-Watson stat	2.235416	Prob(F-statistic)	0.000280		

- a) Viết mô hình hồi quy mẫu và nêu ý nghĩa của các hệ số hồi quy riêng.
- b) Kiểm định giả thiết: hệ số hồi quy của biến X trong hàm hồi quy tổng thể bằng 3.5 với mức ý nghĩa 5%.
- c) Để dự báo Y ta nên dùng mô hình ở câu 1 hay mô hình ở câu 2? Vì sao? (Với mức ý nghĩa 5%).
- d) Từ bảng kết quả dưới đây, bạn hãy cho biết mô hình hồi quy ở câu 2 có xảy ra hiện tượng phương sai thay đổi hay không, với mức ý nghĩa 5%?

While Heteroskedasticity			
F-statistic	3.553680	Probability	0.087187
Obs*R-squared	6.398784	Probability	0.093741

- e) Dựa trên kết quả hồi quy dưới đây, với mức ý nghĩa 5%, theo bạn có thể kết luận rằng mô hình hồi quy ở câu 2 bỏ sót biến thích hợp hay không?

Ramsey RESET Test			
F-statistic	29.41432	Probability	0.001717
Log likelihood ratio	25.46764	Probability	0.000003

Câu 3 .

- a. Hồi quy $\ln Y$ theo X ta được kết quả: $\widehat{\ln Y} = 3.155 + 0.071X$. Nếu ý nghĩa hệ số hồi quy của biến X .
- b. Hồi quy Y theo $\ln X$ ta được kết quả: $\widehat{Y} = -33.858 + 36.526 \ln X$. Nếu ý nghĩa hệ số hồi quy của biến $\ln X$.

Giải

Câu 1 .

- a) Hàm hồi quy tuyến tính mẫu

$$\widehat{Y} = 15,33632 + 3,393124X$$

Ý nghĩa hệ số góc β_2 : khi thu nhập của người tiêu dùng tăng 1 (triệu/tháng) thì chi tiêu của mặt hàng A tăng trung bình 3,393124 (100 ngàn đồng) trong điều kiện các yếu tố khác không đổi.

- b) Ta có thể giải theo 2 cách

✚ Cách 1

+ Đặt giả thiết $H_0 : \beta_2 = 0; H_1 : \beta_2 \neq 0$.

+ $C = t_{\frac{\alpha}{2}}^{(n-k)} = t_{0,025} (10 - 2) = 2,306$.

+ $T = \frac{\widehat{\beta}_2}{se(\widehat{\beta}_2)} = 4,549084$.

+ $|T| > C$ suy ra bác bỏ H_0

✚ Cách 2

+ Đặt giả thiết $H_0 : \beta_2 = 0; H_1 : \beta_2 \neq 0$.

+ $P_{value} = 0,0019 < \alpha = 0,05$ suy ra bác bỏ H_0

Ý nghĩa: thu nhập của người tiêu dùng thực sự có ảnh hưởng đến chi tiêu cho mặt hàng A.

- c) Đổi đơn vị

+ Đơn vị của chi tiêu là triệu đồng/năm

$$\Rightarrow Y^* = 1,2Y$$

$$\Rightarrow k_1 = 1,2$$

$$\Rightarrow \hat{\beta}_1^* = k_1 \hat{\beta}_1 = 18,403584$$

+ Đơn vị của thu nhập là triệu đồng/năm

$$\Rightarrow X^* = 12X$$

$$\Rightarrow k_2 = 12$$

$$\Rightarrow \hat{\beta}_2^* = \frac{k_1}{k_2} \hat{\beta}_2 = 0,3393124$$

Vậy

$$\hat{Y}^* = \hat{\beta}_1^* + \hat{\beta}_2^* X^* = 18,403584 + 0,3393124 X^*$$

d) Hệ số co dãn

Ta có $\bar{X} = 11,1$; $\bar{Y} = 53$. Hệ số co dãn của chi tiêu theo thu nhập tại điểm $(\bar{X}; \bar{Y})$ là

$$\varepsilon_{\frac{Y}{X}} = \frac{dY}{dX} \frac{\bar{X}}{\bar{Y}} = \hat{\beta}_2 \frac{\bar{X}}{\bar{Y}} = 0,7106$$

Ý nghĩa: khi thu nhập của người tiêu dùng tăng 1% thì chi tiêu cho mặt hàng A tăng trung bình 0,7106%.

e) Dự báo

$$+ \hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_0 = 15,33632 + 3,393124 \cdot 10 = 49,267.$$

$$+ C = t_{\frac{\alpha}{2}}^{(n-k)} = t_{0,025}^{(10-2)} = 2,306.$$

$$+ \text{var}(\hat{Y}_0) = \left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{n \cdot \text{var}(X)} \right] \hat{\sigma}^2 = 4,3952 \Rightarrow \text{se}(\hat{Y}_0) = 2,0965$$

Khoảng dự báo cho giá trị trung bình Y_0 ($E(Y/X_0 = 10)$)

$$\hat{Y}_0 - C \text{se}(\hat{Y}_0) \leq E(Y/X_0 = 10) \leq \hat{Y}_0 + C \text{se}(\hat{Y}_0)$$

$$\Rightarrow 44,4325 \leq E(Y/X_0 = 10) \leq 54,102$$

Câu 2 .

a) Hàm hồi quy tuyến tính mẫu

$$\hat{Y} = 13,10545 + 3,193939X + 8,883636Z$$

Ý nghĩa hệ số góc β_2 : khi thu nhập của người tiêu dùng bất kể nam hay nữ tăng lên 1 (triệu/tháng) thì chi tiêu của mặt hàng A tăng trung bình 3,193939 (100 ngàn đồng) trong điều kiện các yếu tố khác không đổi.

Ý nghĩa hệ số góc β_3 : khi có cùng một mức thu nhập, chi tiêu trung bình cho mặt hàng A của người nữ sẽ cao hơn 8,8836 (100 ngàn đồng) so với chi tiêu trung bình cho mặt hàng A của người nam (trong điều kiện các yếu tố khác không đổi).

b) Kiểm định giả thiết

+ Đặt giả thiết $H_0 : \beta_2 = 3,5$; $H_1 : \beta_2 \neq 3,5$.

$$+ C = t_{\frac{\alpha}{2}}^{(n-k)} = t_{0,025}^{(10-3)} = 2,365.$$

$$+ T = \frac{\hat{\beta}_2 - 3,5}{se(\hat{\beta}_2)} = -0,6478.$$

+ $|T| < C$ suy ra chưa có cơ sở bác bỏ H_0

c) Lựa chọn mô hình

$$\overline{R}_1^2 = 0,686347 < \overline{R}_2^2 = 0,87586$$

Kiểm định:

+ Đặt giả thiết $H_0 : \beta_3 = 0$; $H_1 : \beta_3 \neq 0$.

+ $P_{value} = 0,0083 < \alpha = 0,05$ suy ra bác bỏ H_0

Vậy ta nên chọn mô hình 2.

d) Đặt giả thiết H_0 : mô hình không có phương sai thay đổi.

$P_{value} = 0,087187 > \alpha = 0,05$ suy ra chưa có cơ sở bác bỏ H_0

e) Đặt giả thiết H_0 : mô hình không bỏ sót biến.

$P_{value} = 0,001717 < \alpha = 0,05$ suy ra bác bỏ H_0

Câu 3 .

- a) Ý nghĩa của $\beta_2 = 0,0713$: khi thu nhập của người tiêu dùng tăng 1 (triệu đồng/tháng) thì chi tiêu trung bình cho mặt hàng A tăng 7,31%.
- b) Ý nghĩa của $\beta_2 = 36,5266$: khi thu nhập của người tiêu dùng tăng 1% thì chi tiêu trung bình cho mặt hàng A tăng 0,365266 (trăm ngàn đồng /tháng).

Bài 2.18. Giả sử có mẫu thống kê của 10 tháng trong một năm như sau:

Tháng	1	2	3	4	5	6	7	8	9	10
Y	47	52	72	52	32	52	67	66	42	37
X_2	58	57	56	59	58	58	57	57	59	60
X_3	74	73	71	72	75	72	71	70	73	73

Trong đó: Y là lượng cà phê tiêu thụ của một cá nhân (tách/tháng); X_2 là giá bán lẻ trung bình của cà phê (ngàn đồng/kg) và X_3 là giá bán lẻ trung bình của đường (ngàn đồng/kg).

Câu 1. Hồi quy Y theo X ta được kết quả cho ở bảng sau:

Dependent Variable: Y					
Variable	Coefficient	Std. Error	t-Statistic	Prob.	
C	541.5814	145.1367	3.731526	0.0058	
X2	-8.457364	2.506197	-3.374581	0.0097	
R-squared	0.587369	Mean dependent var	51.90000		
Adjusted R-squared	0.535790	S.D. dependent var	13.21153		
S.E. of regression	9.001400	Akaike info criterion	7.409494		
Sum squared resid	648.2016	Schwarz criterion	7.470011		
Log likelihood	-35.04747	F-statistic	11.38780		
Durbin-Watson stat	2.298401	Prob(F-statistic)	0.009719		

- a) Hãy viết mô hình hồi quy tuyến tính mẫu mô tả quan hệ giữa lượng cà phê tiêu thụ của một cá nhân theo giá cà phê. Nêu ý nghĩa kinh tế của hệ số góc được ước lượng.
- b) Tìm khoảng tin cậy của hệ số góc tổng thể với độ tin cậy 99%. Xét xem giá cà phê có ảnh hưởng đến lượng cà phê được tiêu thụ hay không với mức ý nghĩa 1%.
- c) Dự đoán lượng cà phê tiêu thụ trung bình khi giá cà phê là 55 (ngàn đồng/kg) với độ tin cậy 95%.

- d) Hãy viết hàm hồi quy khi đơn vị tính của lượng cà phê tiêu thụ là tách/năm và giá cà phê là đồng/kg.
- e) Xét tại mức giá 55 (ngàn đồng/kg). Hãy cho biết để tăng doanh thu thì nên tăng hay giảm giá bán của cà phê.

Câu 2. Với số liệu đã cho ở câu 1. Hồi quy Y theo X_2 và X_3 ta có kết quả cho ở bảng sau:

Dependent Variable: Y					
Variable	Coefficient	Std. Error	t-Statistic	Prob.	
C	772.0533	67.32520	11.46752	0.0000	
X2	-5.083493	1.117926	-4.547255	0.0026	
X3	-5.881478	0.888982	-6.615973	0.0003	
R-squared	0.943109	Mean dependent var	51.90000		
Adjusted R-squared	0.926855	Akaike info criterion	5.628077		
S.E. of regression	3.573113	Schwarz criterion	5.718852		
Log likelihood	-25.14038	F-statistic	58.02123		
Durbin-Watson stat	2.079731	Prob(F-statistic)	0.000044		

- a) Hãy viết kết quả hồi quy theo quy ước, nêu ý nghĩa của các hệ số hồi quy riêng.
- b) Kiểm định sự phù hợp của mô hình với mức ý nghĩa 1%.
- c) Cho biết mô hình trên có xảy ra hiện tượng tự tương quan hay không?

Câu 3. a) Từ bảng kết quả dưới đây, bạn hãy cho biết mô hình hồi quy ở câu 2 có xảy ra hiện tượng phương sai thay đổi hay không, với mức ý nghĩa 5%?

White Heteroskedasticity Test:			
F-statistic	1.792213	Probability	0.267601
Obs*R-squared	5.891150	Probability	0.207426

- b) Dựa trên kết quả hồi quy dưới đây, với mức ý nghĩa 5%, theo bạn có thể kết luận rằng mô hình hồi quy ở câu 2 bỏ sót biến thích hợp hay không?

Ramsey RESET Test:			
F-statistic	0.558796	Probability	0.603912
Log likelihood ratio	2.017305	Probability	0.364710

- c) Để dự báo Y , bạn sẽ chọn mô hình nào trong hai mô hình: mô hình ở câu 1 và mô hình ở câu 2? vì sao? (với mức ý nghĩa 5%).

Giải

Câu 1 .

- a) Hàm hồi quy tuyến tính mẫu

$$\hat{Y} = 541,5814 - 8,457364X$$

Ý nghĩa hệ số góc $\beta_2 = -8,457364$: khi giá bán lẻ trung bình cafe tăng 1 (ngàn đồng/kg) thì lượng tiêu thụ cafe của một cá nhân giảm 8,457364 (tách/tháng) trong điều kiện các yếu tố khác không đổi.

- b) Khoảng tin cậy

Áp dụng: $\hat{\beta}_i - C.se(\hat{\beta}_i) \leq \beta_i \leq \hat{\beta}_i + C.se(\hat{\beta}_i)$. Trong đó

$$C = t_{\frac{\alpha}{2}}^{(n-k)} = t_{0,005}^{(10-2)} = 3,3554$$

Khoảng tin cậy của hệ số góc

$$\begin{aligned} -8,457364 - 3,3554 \cdot 2,506197 &\leq \beta_2 \leq -8,457364 + 3,3554 \cdot 2,506197 \\ \Rightarrow -16,867 &\leq \beta_2 \leq -0,048 \end{aligned}$$

Ta nhận thấy $0 \notin (-16,867; -0,048)$ nên ta bác bỏ giả thiết $H_0 : \beta_2 = 0$. Vậy với mức ý nghĩa 1%, giá bán lẻ của cafe thực sự có ảnh hưởng đến lượng cafe tiêu thụ của một cá nhân.

- c) Dự báo

$$+ \hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_0 = 541,5814 - 8,457364 \cdot 55 = 76,42638.$$

$$+ C = t_{\frac{\alpha}{2}}^{(n-k)} = t_{0,025}^{(10-2)} = 2,306.$$

+ $\text{var}(\hat{Y}_0) = \left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{n \text{var}(X)} \right] \hat{\sigma}^2 = 60,9259 \Rightarrow se(\hat{Y}_0) = 7,8055$. Khoảng dự báo cho giá trị trung bình Y_0 ($E(Y/X_0 = 55)$)

$$\begin{aligned} \hat{Y}_0 - C se(\hat{Y}_0) &\leq E(Y/X_0 = 55) \leq \hat{Y}_0 + C se(\hat{Y}_0) \\ \Rightarrow 76,42638 - 2,306 \cdot 7,8055 &\leq E(Y/X_0 = 55) \leq 76,42638 + 2,306 \cdot 7,8055 \\ \Rightarrow 58,4269 &\leq E(Y/X_0 = 55) \leq 94,4259 \end{aligned}$$

d) Đổi đơn vị

+ Đơn vị tính của Y là đồng/năm

$$\Rightarrow Y^* = 12Y$$

$$\Rightarrow k_1 = 12$$

$$\Rightarrow \hat{\beta}_1^* = k_1 \hat{\beta}_1 = 6498,9768$$

+ Đơn vị của X là triệu đồng/năm

$$\Rightarrow X^* = 1000X$$

$$\Rightarrow k_2 = 1000$$

$$\Rightarrow \hat{\beta}_2^* = \frac{k_1}{k_2} \hat{\beta}_2 = -0,1015$$

Vậy

$$\hat{Y}^* = \hat{\beta}_1^* + \hat{\beta}_2^* X^* = 6498,9768 - 0,1015 X^*$$

e) Đặt T là doanh thu, khi đó ta có

$$T = XY \Rightarrow T'_X = Y + XY'_X = Y \left(1 + Y'_X \frac{X}{Y} \right) = Y (1 + \varepsilon_{Y/X})$$

$$X_2 = 55 \Rightarrow Y = 76,42638$$

$$\varepsilon_{Y/X_2} = Y'_{X_2} \frac{X_2}{Y} = -8,457364 \frac{55}{76,42638} = -6,0863$$

$$\Rightarrow 1 + \varepsilon_{Y/X_2} = -5,0863 < 0$$

$$\Rightarrow T'_{X_2} < 0$$

Vậy X_2 giảm thì T tăng.

Câu 2 .

a) Hàm hồi quy tuyến tính mẫu

$$\hat{Y} = 772,0533 - 5,083493X_{2i} - 5,881478X_{3i}$$

Ý nghĩa hệ số góc β_2 : khi giá bán lẻ trung bình của cafe tăng lên 1 (ngàn đồng/kg) thì lượng cafe tiêu thụ của một cá nhân giảm 5,083493 (tách/tháng) trong điều kiện các yếu tố khác không đổi.

Ý nghĩa hệ số góc β_3 : khi giá bán lẻ trung bình của đường tăng lên 1 (ngàn đồng/kg) thì lượng cafe tiêu thụ của một cá nhân giảm 5,881478 (tách/tháng) trong điều kiện các yếu tố khác không đổi.

b) Kiểm định sự phù hợp của mô hình

¶ Cách 1.

+ Đặt giả thiết $H_0 : R^2 = 0; H_1 : R^2 > 0$.

+ Với $\alpha = 1\% \Rightarrow C = F_{\alpha}(k-1; n-k) = F_{0,01}(2; 7) = 9,55$

+ $F = 58,02123$

+ Ta có $F > C$ nên suy ra bác bỏ H_0 . Vậy mô hình (2) phù hợp với mức ý nghĩa 1%.

¶ Cách 2.

+ Đặt giả thiết $H_0 : R^2 = 0; H_1 : R^2 > 0$.

+ $P_{value} = 0,000044 < \alpha = 0,01$ nên bác bỏ H_0 . Vậy mô hình (2) phù hợp với mức ý nghĩa 1%.

c) Ta có $d = 2,079731 \in (1, 3)$, suy ra mô hình (2) không xảy ra hiện tượng tự tương quan.

Câu 3 .

a) Đặt giả thiết H_0 : mô hình không có phương sai thay đổi.

$P_{value} = 0,207426 > \alpha = 0,05$ suy ra chưa có cơ sở bác bỏ H_0

b) Đặt giả thiết H_0 : mô hình không bỏ sót biến.

$P_{value} = 0,603912 > \alpha = 0,05$ suy ra chưa có cơ sở bác bỏ H_0

c) Lựa chọn mô hình

$$\overline{R}_1^2 = 0,5358 < \overline{R}_2^2 = 0,9269$$

Kiểm định:

+ Đặt giả thiết $H_0 : \beta_3 = 0; H_1 : \beta_3 \neq 0$.

+ $P_{value} = 0,0003 < \alpha = 0,05$ suy ra bác bỏ H_0

Vậy ta nên chọn mô hình 2.

2.5 Bài tập đề nghị

Bài 2.19. Cho bảng số liệu quan sát về X, Y như sau:

X	23	19,5	24	21	25	22	26,5	23,1	25	28	29,5	26
Y	3	2	4	2	5	4	7	6	8	9	10	8

Trong đó Y là thu nhập của giảng viên đại học (triệu đồng/năm), X là thâm niên giảng dạy (năm)

1. Tìm hàm hồi quy tuyến tính mẫu của Y theo X và phát biểu ý nghĩa của các hệ số hồi quy?
2. Tính hệ số tương quan tuyến tính r và đánh giá mức độ phụ thuộc tương quan tuyến tính?
3. Dự báo thu nhập trung bình của một giảng viên có thâm niên giảng dạy là 6 năm với độ tin cậy 95%?

Đáp số

1. $\hat{Y} = 18,8972 + 0,968145X$
2. $r = 0,91875$
3. $E(Y/X_0 = 6) \in (23, 43; 25, 48)$

Bài 2.20. Dữ liệu về giá nhà (Y : triệu đồng), số phòng (X_1 : phòng) và diện tích (X_2 m^2) được cho ở bảng sau

Y	X_2	X_3
562	5	86.5
279.5	2	76.6
653	5	96.5
240	3	43
285	3	42
415	4	57
270	3	32
332.5	4	45
110	2	25
212	2	29

1. Tìm các hệ số hồi quy mẫu?
2. Tìm hệ số xác định mô hình và phương sai của sai số ngẫu nhiên?

3. Tìm ma trận hiệp phương sai và từ đó cho biết độ lệch chuẩn của các hệ số hồi quy mẫu?
4. Với $X_2 = 4, X_3 = 50$, hãy cho biết dự báo điểm của giá nhà (\widehat{Y}_0) và từ đó cho biết $Var(\widehat{Y}_0)$ và $Var(Y_0 - \widehat{Y}_0)$?

Đáp số

1.

$$\begin{cases} \widehat{\beta}_1 = -106,84553 \\ \widehat{\beta}_2 = 83,82841 \\ \widehat{\beta}_3 = 3,11888 \end{cases}$$

2. $R^2 = 0,95671; \widehat{\sigma}^2 = 1506,80003$.

3.

$$\begin{aligned} Cov(\widehat{\beta}) &= \begin{pmatrix} 1518,21117 & -376,17964 & -2,36835 \\ -376,17964 & 230,4484 & -7,21555 \\ -2,36835 & -7,21555 & 0,49154 \end{pmatrix} \\ \Rightarrow \begin{cases} Se(\widehat{\beta}_1) = 38,96423 \\ Se(\widehat{\beta}_2) = 15,18053 \\ Se(\widehat{\beta}_3) = 0,7011 \end{cases} \end{aligned}$$

4. $\widehat{Y}_0 = 384,41211; Var(\widehat{Y}_0) = 301,75542; Var(Y_0 - \widehat{Y}_0) = 1808,55545$.

Bài 2.21. Bảng dưới đây cho các giá trị quan sát về thu nhập (Y - USD/người), tỷ lệ lao động nông nghiệp (X_2 - %) và số năm trung bình được đào tạo đối với những người trên 25 tuổi (X_3 - năm)

Y	X_2	X_3
6	9	8
8	10	13
8	8	11
7	7	10
7	10	12
12	4	16
9	5	10
8	5	10
9	6	12
10	8	14
10	7	12
11	4	16
9	9	14
10	5	10
11	8	12

1. Tìm hàm hồi quy mẫu và giải thích ý nghĩa kinh tế của các hệ số hồi quy?
2. Tìm ước lượng phương sai của sai số ngẫu nhiên?
3. Tìm ước lượng sai số chuẩn của các hệ số hồi quy?
4. Tìm khoảng tin cậy đối xứng của các hệ số hồi quy với độ tin cậy 95%.
5. Với mức ý nghĩa 5% hãy cho biết tỷ lệ lao động nông nghiệp có thực sự ảnh hưởng đến thu nhập hay không?
6. Tính hệ số R^2 và \overline{R}^2 .
7. Phải chăng cả hai yếu tố "tỷ lệ lao động nông nghiệp" và "số năm được đào tạo" đều không ảnh hưởng đến thu nhập?

Đáp số

1. $\hat{Y} = 6,20298 - 0,3761639X_{2i} + 0,452514X_{3i}$
2. $\hat{\sigma}^2 = 1,02265673$
3. $se(\hat{\beta}_1) = 1,862253; se(\hat{\beta}_2) = 0,1195112; se(\hat{\beta}_3) = 0,1327238$

4. $\beta_1 \in (2, 145478; 10, 26048)$; $\beta_2 \in (-0, 6653441; -0, 0869837)$;
 $\beta_3 \in (0, 1921215; 0, 7129064)$
5. Bác bỏ H_0 . Tỷ lệ lao động nông nghiệp thực sự ảnh hưởng đến thu nhập.
6. $R^2 = 0,6932$ và $\bar{R}^2 = 0,6421$.
7. Như vậy không phải cả hai yếu tố "tỷ lệ lao động nông nghiệp" và "số năm được đào tạo" đều không ảnh hưởng đến thu nhập.

Bài 2.22. Bảng dưới đây là số liệu giả thiết về mức lương giảng viên đại học (Y - ngàn USD/năm), số năm kinh nghiệm giảng dạy (X - năm) và giới tính (nam: $Z_i = 1$; nữ: $Z_i = 0$)

Y_i	X_i	Z_i
23	11	1
19.5	9	0
24	10	1
21	12	0
25	13	1
22	12	0
26.5	14	1
23.1	14	0
25	15	0
28	15	1
29.5	16	1
26	16	0
27.5	17	0
31.5	18	1
29	18	0

1. Tìm hàm hồi quy mẫu và giải thích ý nghĩa kinh tế của các hệ số hồi quy?
2. Tìm hệ số xác định mô hình và cho biết ý nghĩa của nó?
3. Dự báo mức lương của một giảng viên nam có số năm kinh nghiệm giảng dạy 17 năm với độ tin cậy 95%?
4. Dự báo mức lương của một giảng viên nữ có số năm kinh nghiệm giảng dạy 19 năm với độ tin cậy 98%?

Đáp số

1. $\hat{Y} = 8,993523 + 1,0714X + 2,935395Z$
2. $R^2 = 0,9594$
3. (29,38427; 30,92633)
4. (28,0552; 30,6732)

Bài 2.23. Cho bảng số liệu về X, Y, Z như sau

Y_i	X_i	Z_i
23	15	1
19	10	0
24	20	1
21	20	0
25	30	1
22	30	0
26	35	1
23	25	0
25	20	0
28	25	1

Trong đó Y là thu nhập của hộ gia đình (triệu đồng/năm); X là tỷ lệ thu nhập chi cho giáo dục (%); Z là biến giả ($Z = 1$ nếu hộ gia đình ở thành phố; $Z = 0$ nếu hộ gia đình ở nông thôn)

1. Tìm hàm hồi quy tuyến tính mẫu của X theo Y và nêu ý nghĩa của các hệ số hồi quy?
2. Tính hệ số tương quan tuyến tính và đánh giá mức độ phụ thuộc tương quan tuyến tính giữa X và Y ?
3. Kiểm định giả thiết hệ số hồi quy của Y trong hàm hồi quy tổng thể bằng 0 với mức ý nghĩa 5% và nêu ý nghĩa của kết quả?
4. Viết hàm hồi quy tuyến tính mẫu của X theo Y khi đơn vị tính của Y là ngàn đồng/tháng?

Đáp số

1. $\hat{X} = -16,8543 + 1,68874Y$

2. $r = 0,581$. Mức độ phụ thuộc tương quan tuyến tính không chặt chẽ.
3. Chưa có cơ sở bác bỏ H_0 . Thu nhập không ảnh hưởng đến tỷ lệ thu nhập chi cho giáo dục.
4. $\widehat{X}^* = -16,8543 + 0,020265Y^*$

Bài 2.24. Từ các số liệu thu thập ở 30 doanh nghiệp may ở TP Hồ Chí Minh người ta đã ước lượng được mô hình sau

$$\begin{aligned} \widehat{Y}_i &= -5.54 + 0.112X_i - 14.48D_i + 0.25X_iD_i & R^2 &= 0.57 \\ Se &= (10.85) \quad (0.08) \quad (15.47) \quad (0.12) & d &= 2.16 \end{aligned}$$

Trong đó Y là lợi nhuận (tỉ VND); X là doanh thu (tỉ VND); $D_i = 1$ nếu giám đốc đã tốt nghiệp đại học và $D_i = 0$ nếu giám đốc chưa tốt nghiệp đại học

1. Mô hình trên có bị hiện tượng tự tương quan không (với mức ý nghĩa 5%)? Giả sử rằng các giả định khác đều đúng.
2. Có thể dùng kiểm định t để kiểm định giả thiết khác không của các hệ số hồi quy ở mô hình được ước lượng bởi mô hình trên hay không? vì sao? Nếu được hãy tiến hành kiểm định với mức ý nghĩa 5% và cho biết có sự khác biệt về lợi nhuận giữa doanh nghiệp có giám đốc đã tốt nghiệp đại học với chưa tốt nghiệp đại học không?
3. Khi doanh nghiệp có doanh thu tăng 1 tỉ đồng thì lợi nhuận của doanh nghiệp thay đổi như thế nào?
4. Nếu đơn vị tính của X và Y đổi thành triệu đồng thì các hệ số của mô hình trên thay đổi thế nào?

Đáp số

1. Với mức ý nghĩa 5%, mô hình không xảy ra hiện tượng tự tương quan.
2. Có thể dùng kiểm định t để kiểm định giả thiết khác không của các hệ số hồi quy ở mô hình được ước lượng bởi mô hình trên. Có sự khác biệt về lợi nhuận giữa doanh nghiệp có giám đốc đã tốt nghiệp đại học với chưa tốt nghiệp đại học.
3. Nếu là doanh nghiệp có giám đốc chưa tốt nghiệp đại học ($D=0$) thì lợi nhuận sẽ tăng 0,12 tỉ đồng. Nếu là doanh nghiệp có giám đốc tốt nghiệp đại học ($D=1$) thì lợi nhuận sẽ tăng 0,37 tỉ đồng.
4. $\widehat{Y} = -5540 + 0,12X - 14480D + 0,25XD$

Bài 2.25. Một công ty thu thập dữ liệu trong vòng 14 tháng, gồm các biến sau:

- + Doanh thu (Y - triệu đồng/tháng)
- + Chi phí quảng cáo trên báo (X_1 - trăm ngàn đồng/tháng)
- + Chi phí quảng cáo trên radio (X_2 - trăm ngàn đồng/tháng)

Câu 1. Hàm hồi quy nghiên cứu sự phụ thuộc doanh thu của công ty vào chi phí quảng cáo trên báo có dạng: $Y = \alpha + \beta X_1 + U$ (MH1). Kết quả hồi quy như sau:

Method: Least Squares				
Included observations: 14				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	74.98143	16.28961	4.603022	0.0006
X_1	2.106169	0.515583	4.085021	0.0015
R-squared	0.581698	Mean dependent var		137.4143
Adjusted R-squared	0.546839	S.D. dependent var		31.32914
S.E. of regression	21.08990	F-statistic		16.68740
Durbin-Watson stat	1.057844	Prob(F-statistic)		0.001512

- a) Viết hàm hồi quy mẫu ngẫu nhiên tương ứng và cho biết ý nghĩa kinh tế của hệ số hồi quy riêng?
- b) Nếu trong tháng công ty chi thêm 1 trăm ngàn đồng cho quảng cáo trên báo thì doanh thu của công ty trong tháng đó sẽ tăng 3 triệu đồng. Có thể chấp nhận ý kiến trên không, với độ tin cậy 98%?
- c) Viết hàm hồi quy mới với X_1^* có đơn vị tính: triệu đồng/tháng.
- d) Mô hình hồi quy có phù hợp không, mức ý nghĩa 1%
- e) Dự báo doanh thu trung bình khi công ty chi 6 triệu đồng cho quảng cáo trên báo, độ tin cậy 95%.
- f) Từ dữ liệu ban đầu, có hàm hồi quy như sau: $\widehat{\ln(Y)} = 4.4053 + 0.0166X_1$. Nêu ý nghĩa của hệ số hồi quy đứng trước biến X_1 .

Câu 2. Hàm hồi quy Y theo $X_1; X_2$ có dạng: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + V$ (MH2). Kết quả hồi quy như sau:

Method: Least Squares					
Included observations: 14					
Variable	Coefficient	Std. Error	t-Statistic	Prob.	
X_1	1.751702	0.455891	3.842372	0.0027	
X_2	0.774569	0.314865	2.460005	0.0317	
C	43.55144	18.70765	2.328001	0.0400	
R-squared	0.730153	Mean dependent var		137.4143	
Adjusted R-squared	0.681090	S.D. dependent var		31.32914	
S.E. of regression	3443.162	F-statistic		14.88195	

- Viết hàm hồi quy mẫu ngẫu nhiên ứng với bảng kết quả trên? Nêu ý nghĩa của các hệ số hồi quy riêng.
- Tìm khoảng tin cậy của hệ số hồi quy đứng trước biến X_1 , độ tin cậy 99%?
- Để dự đoán doanh thu của công ty, bạn sẽ chọn (MH1) hay (MH2), với mức ý nghĩa 5%?

Đáp số

- Câu 1.
- $\hat{Y} = 74,9814 + 2,1062X_1$
 - Chưa có cơ sở bác bỏ H_0
 - $\hat{Y} = 74,9814 + 0,21062X_1^*$
 - $P_{value} = 0,0015 < 0,01$ suy ra bác bỏ H_0 . Hàm hồi quy phù hợp.
 - $Var(\hat{Y}_0) = 276,7592$; (156,1232; 246,5836)
 - Khi chi phí quảng cáo trên báo tăng 1 (100000 đồng/tháng) thì doanh thu của công ty tăng trung bình 1,66%.
- Câu 2.
- $\hat{Y} = 43,5514 + 1,7517X_1 + 0,7746X_2$
 - Chọn MH2.

Bài 2.28. Cho một mẫu gồm các giá trị quan sát sau:

Y	34	54	44	57	53	64	62	66	63	69
X	27	26	25	25	24	23	21	20	22	19
Z	0	1	0	1	0	1	0	0	0	1

Trong đó: Y là lượng hàng A bán được (đơn vị: tấn/tháng); X là giá của hàng A (đơn vị: 100 ngàn đồng/kg); $Z = 1$: có chương trình quảng cáo, $Z = 0$: không có chương trình quảng cáo.

Câu 1. Hồi quy Y theo X ta được kết quả cho ở bảng sau:

Dependent Variable: Y				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	137.2893	17.52906	7.832099	0.0001
X	-3.477987	0.751138	-4.630291	0.0017
R-squared	0.728257	Mean dependent var		56.60000
Adjusted R-squared	0.694289	S.D. dependent var		10.83410
S.E. of regression	5.990296	Akaike info criterion		6.595015
Sum squared resid	287.0692	Schwarz criterion		6.655532
Log likelihood	-30.97508	F-statistic		21.43959
Durbin-Watson stat	2.967667	Prob(F-statistic)		0.001687

- Hãy viết mô hình hồi quy tuyến tính mẫu biểu diễn mối quan hệ của Y theo X . Nêu ý nghĩa kinh tế của hệ số góc của hàm hồi quy tìm được. (MH1)
- Kiểm định sự phù hợp của mô hình với mức ý nghĩa 1%.
- Dự báo mức lượng hàng bán được trung bình khi giá của hàng A là 2450 ngàn đồng/kg, với độ tin cậy 95%.
- Hãy viết hàm hồi quy mẫu khi đơn vị tính của Y là tấn/năm, đơn vị tính của X là triệu đồng/tấn.

Câu 2. Từ số liệu trên, hồi quy Y theo $\ln(X)$ ta được:

Dependent Variable: Y				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	301.4422	55.77150	5.404951	0.0006
LOG(X)	-78.02251	17.76134	-4.392829	0.0023
R-squared	0.706927	Mean dependent var		56.60000
Adjusted R-squared	0.670293	S.D. dependent var		10.83410
S.E. of regression	6.220956	Akaike info criterion		6.670581
Sum squared resid	309.6024	Schwarz criterion		6.731098
Log likelihood	-31.35290	F-statistic		19.29695
Durbin-Watson stat	2.820120	Prob(F-statistic)		0.002309

- Nêu ý nghĩa kinh tế của hệ số góc trong hàm hồi quy trên (MH2).

- b) Từ bảng kết quả dưới đây, bạn hãy cho biết mô hình hồi quy ở câu 2 có xảy ra hiện tượng phương sai thay đổi hay không, với mức ý nghĩa 5%?

While Heteroskedasticity Test:			
F-statistic	8.853394	Probability	0.012106
Obs*R-squared	7.166771	Probability	0.027781

- c) Dựa vào bảng kết quả hồi quy dưới đây, với mức ý nghĩa 5%, theo bạn có thể kết luận rằng mô hình hồi quy ở câu 2 bỏ sót biến thích hợp hay không?

Ramsey RESET Test:			
F-statistic	1.581611	Probability	0.280743
Log likelihood ratio	4.234385	Probability	0.120369

Câu 3. Với số liệu đã cho đã cho ở trên, hồi quy Y theo X và Z ta có kết quả như sau:

Dependent Variable: Y					
Variable	Coefficient	Std. Error	t-Statistic	Prob.	
C	134.7955	13.47027	10.00689	0.0000	
X	-3.501966	0.575794	-6.081979	0.0005	
Z	7.625164	2.964080	2.572532	0.0369	
R-squared	0.860316	Mean dependent var	56.60000		
Adjusted R-squared	0.820406	S.D. dependent var	10.83410		
S.E. of regression	4.591331	Akaike info criterion	6.129542		
Sum squared resid	147.5623	Schwarz criterion	6.220318		
Log likelihood	-27.64771	F-statistic	21.55654		
Durbin-Watson stat	2.084056	Prob(F-statistic)	0.001019		

- a) Viết hàm hồi quy mẫu và nêu ý nghĩa các hệ số hồi quy riêng. (MH3)
 b) Kiểm định sự phù hợp của mô hình với mức ý nghĩa 5%.
 c) Để dự báo lượng hàng A bán được trung bình, bạn sẽ chọn mô hình nào trong hai mô hình: MH1 và MH3? Vì sao? (với mức ý nghĩa 5%).
 d) Dự báo lượng hàng bán được trung bình khi có chương trình quảng cáo và với mức giá là 2500 ngàn đồng/kg.

Đáp số

Câu 1. a) $\hat{Y} = 137,2893 - 3,477987X$

- b) $P_{value} = 0,001687 < 0,01$ suy ra bác bỏ H_0 . Mô hình phù hợp.
 c) $Var(\widehat{Y}_0) = 4,5419; (47,1641; 56,9931)$
 d) $\widehat{Y}^* = 1647,4716 - 0,4174X^*$

- Câu 2. a) $\widehat{\beta}_2 = -78,02251$: khi X tăng 1% thì Y giảm trung bình 0,7802 (tấn/tháng)
 b) Bác bỏ H_0 . Vậy có hiện tượng phương sai thay đổi.
 c) Chưa có cơ sở bác bỏ H_0 . Vậy mô hình không bỏ sót biến.

- Câu 3. a) $\widehat{Y}_i = 134,7955 - 3,501966X + 7,625164Z_i$
 b) $P_{value} = 0,001019 < 0,05$ suy ra bác bỏ H_0 . Vậy mô hình phù hợp.
 c) Chọn mô hình ở câu 3.
 d) $\widehat{Y}_0 = 54,8715$ (tấn/tháng).

Bài 2.27. Giám đốc của công ty vận tải công cộng muốn kiểm tra mối quan hệ giữa chi phí bảo trì hàng năm của 1 chiếc xe buýt (Y - triệu đồng) và thời gian hoạt động của xe buýt (X - năm). Công ty này thuê cả tài xế nam và nữ. Họ nghi ngờ rằng chi phí bảo trì hàng năm phụ thuộc cả vào kỹ thuật lái khác nhau giữa hai nhóm tài xế nam và nữ. Dữ liệu thu thập như sau:

Xe buýt	1	2	3	4	5	6	7	8	9	10
Chi phí bảo trì	20	17	13	17	25	7	9	16	24	13
TG hoạt động	10	6	4	11	12	3	2	9	14	8
Tài xế	Nữ	Nữ	Nữ	Nam	Nữ	Nam	Nam	Nam	Nữ	Nam

- Câu 1. Giả sử hàm hồi quy tuyến tính nghiên cứu sự phụ thuộc của chi phí bảo trì hàng năm theo thời gian hoạt động của xe buýt có dạng: $Y = \beta_0 + \beta_1X + U$ (MH1). Hồi quy Y theo X ta được kết quả cho ở bảng sau

Dependent Variable: Y					
Variable	Coefficient	Std. Error	t-Statistic	Prob.	
C	5.769231	1.978930	2.915328	0.0194	
X	1.307692	0.225374	5.802328	0.0004	
R-squared	0.808002	Mean dependent var	16.10000		
Adjusted R-squared	0.784002	Akaike info criterion	5.024495		
S.E. of regression	2.731582	Schwarz criterion	5.085012		
Log likelihood	-23.12248	F-statistic	33.66701		
Durbin-Watson stat	2.424564	Prob(F-statistic)	0.000404		

- a) Viết hàm hồi quy mẫu $\widehat{Y}_i = \widehat{\beta}_1 + \widehat{\beta}_2 X_i$ và cho biết ý nghĩa kinh tế của hệ số góc.
- b) Kiểm định sự phù hợp của mô hình, với độ tin cậy 95%.
- c) Có ý kiến cho rằng: "Khi thời gian hoạt động của xe buýt tăng thêm 1 năm thì chi phí bảo trì trung bình hàng năm tăng thêm 2 triệu đồng". Bạn có tin điều này hay không? Với mức ý nghĩa 10%.
- d) Dự đoán chi phí bảo trì của chiếc xe buýt đã hoạt động được 7 năm, với độ tin cậy 98%.
- e) Khi thực hiện hồi quy (MH1), người ta nghi ngờ mô hình có thể xảy ra hiện tượng phương sai thay đổi và tự tương quan. Kết quả các kiểm định như sau:

Breusch-Godfrey Serial Correlation LM Test AR(1):			
F-statistic	1.142480	Probability	0.320596
Obs*R-squared	1.403110	Probability	0.236203

White Heteroskedasticity Test:			
F-statistic	0.922421	Probability	0.440989
Obs*R-squared	2.085783	Probability	0.352434

Với mức ý nghĩa 5%, hãy nêu kết luận của bạn.

Câu 2. Xét mô hình $\ln Y = \beta_0 + \beta_1 X + U$ (MH2). Kết quả hồi quy như sau:

$$\widehat{\ln Y} = 2.0071 + 0.0890X \quad R^2 = 0.7853$$

$$t = (13.8881) \quad (5.4094) \quad d = 1.9098$$

- a) Nêu ý nghĩa kinh tế của hệ số hồi quy đứng trước X .
- b) Mô hình có hiện tượng tự tương quan không, với mức ý nghĩa 5%?
- c) Có thể so sánh R^2 của (MH1) và (MH2) không, tại sao?

Câu 3. Đặt $Z = 0$ nếu là nữ, $Z = 1$ nếu là nam. Ước lượng mô hình: $Y = \beta_0 + \beta_1 X + \beta_2 Z + V$ (MH3). Với số liệu đã cho, ta có kết quả hồi quy như sau:

Dependent Variable: Y					
Variable	Coefficient	Std. Error	t-Statistic	Prob.	
C	9.609231	1.327407	7.239098	0.0002	
X	1.107692	0.126197	8.777505	0.0001	
Z	-4.520000	0.967362	-4.672502	0.0023	
R-squared	0.953386	Mean dependent var	16.10000		
Adjusted R-squared	0.940068	Akaike info criterion	3.808910		
S.E. of regression	1.438864	Schwarz criterion	3.899686		
Log likelihood	-16.04455	F-statistic	71.58466		
Durbin-Watson stat	2.640242	Prob(F-statistic)	0.000022		

- a) Viết mô hình hồi quy mẫu và nêu ý nghĩa của các hệ số hồi quy riêng.
b) Tìm khoảng tin cậy của hệ số hồi quy của biến X trong tổng thể với độ tin cậy 98%.
c) Để dự báo Y bạn sẽ chọn mô hình nào trong hai mô hình: MH1 và MH3? Vì sao? (với mức ý nghĩa 5%)

Đáp số

- Câu 1. a) $\hat{Y} = 5,7692 + 1,3077X$
b) $P_{value} = 0,000404 < 0,05$ suy ra bác bỏ H_0 . Mô hình phù hợp.
c) Bác bỏ H_0
d) $Var(\hat{Y}_0) = 0,7873; (12,5861; 17,2601)$
e) Mô hình không có tự tương quan bậc 1. Mô hình có phương sai không thay đổi.
- Câu 2. a) ý nghĩa (SV tự nêu)
b) Không có hiện tượng tự tương quan
c) Không thể so sánh R^2 của hai mô hình (vì biến phụ thuộc ở dạng khác nhau).
- Câu 3. a) $\hat{Y} = 9,6092 - 1,1077X - 4,52Z$
b) $(0,7294; 1,486)$
c) Chọn mô hình ở câu MH3.

2.6 Đề thi tham khảo

2.6.1 Đề 1

Câu 1. Bảng sau đây cho chuỗi thời gian về mức tiêu dùng (Y : đơn vị 100000 VNĐ) và thu nhập (X : đơn vị 100000 VNĐ). Tính theo đầu người và tính theo giá cố định năm 1980 trong thời kỳ 1971 - 1990 ở một khu vực:

Năm	Y	X	Năm	Y	X
1971	48,34	52,02	1981	52,17	63,36
1972	48,54	52,41	1982	60,84	67,42
1973	47,44	51,55	1983	60,73	67,86
1974	54,58	58,88	1984	76,04	83,39
1975	55,00	59,66	1985	76,42	84,26
1976	63,49	68,42	1986	69,34	77,41
1977	59,22	64,27	1987	61,75	70,08
1978	57,77	63,01	1988	68,78	77,44
1979	60,22	65,61	1989	67,07	75,79
1980	55,40	61,05	1990	72,94	81,89

Giả sử Y và X có quan hệ tuyến tính

- Hãy ước lượng mô hình hồi quy của mức tiêu dùng phụ thuộc vào thu nhập và nêu ý nghĩa kinh tế của các hệ số hồi quy tìm được.
- Tính hệ số xác định mô hình và giải thích ý nghĩa.
- Tìm khoảng tin cậy của các hệ số hồi quy tổng thể, với độ tin cậy 95%.
- Dự báo giá trị cá biệt của mức tiêu dùng khi thu nhập là 8 triệu đồng, với độ tin cậy 95%.
- Có người cho rằng "kiếm mười thì tiêu một". Hãy kiểm chứng nhận định trên với mức ý nghĩa 5%?

Câu 2. Trong một nghiên cứu về các yếu tố ảnh hưởng đến giá nhà, từ 14 quan sát người ta thu được mô hình sau:

$$\hat{Y} = 39,057 + 0,147X + 6,188D + 5,634XD$$
$$t \quad (0,436) \quad (4,869) \quad (2,486) \quad (3,769)$$
$$R^2 = 0,806$$

Trong đó Y : giá bán nhà (triệu đồng/ m^2); X : diện tích nhà (m^2); D bằng 1 nếu nhà có lò sưởi, bằng 0 nếu nhà không có lò sưởi

- Với mức ý nghĩa thống kê 5%, bạn hãy kiểm định xem liệu diện tích nhà có thực sự ảnh hưởng tới giá bán nhà hay không?
- 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 hay không? Hãy giải thích kết luận của mình với mức ý nghĩa 5%?
- Hệ số xác định mô hình nói lên điều gì? Tính hệ số xác định mô hình hiệu chỉnh?

Câu 3. Cho mức ý nghĩa 5%

- Với kết quả sau, hãy phân tích xem có hiện tượng gì xảy ra trong mô hình?

White Heteroskedasticity Test:			
F-statistic	1.461	Probability	0.263
Obs*R-squared	6.859	Probability	0.76

- Với kết quả sau, hãy phân tích xem có hiện tượng gì xảy ra trong mô hình?

Breusch-Godfrey Serial Correlation LM Test:			
F-statistic	2.415	Probability	0.123
Obs*R-squared	4.872	Probability	0.007

Đáp số

Câu 1.

a) $\hat{Y} = 3,22923 + 0,85563X$

b) $R^2 = 0,9714$

c) Khoảng tin cậy

$$\beta_1 \in (-1,71577; 8,17423)$$

$$\beta_2 \in (0,78294; 0,92832)$$

d) $\hat{Y}_0 = 71,67963; Y_0 \in (68,2516; 75,10766)$.

e) $|T| = |21,839| > C = t_{0,025}^{18} = 2,101$ suy ra bác bỏ giả thiết $H_0 : \beta_2 = 0, 1$

Câu 2.

- $|T| = |4,869| > C = t_{0,025}^{10} = 2,228$ suy ra bác bỏ giả thiết H_0 . Vậy diện tích nhà thực sự ảnh hưởng đến giá bán.

- b) Sinh viên tự nêu ý nghĩa. Sự tồn tại của biến tương tác trong mô hình là cần thiết.
- c) $R^2 = 0,806$; $\overline{R}^2 = 0,7478$.

Câu 3.

- a) Mô hình không xảy ra hiện tượng phương sai thay đổi.
- b) Mô hình xảy ra hiện tượng tự tương quan.

2.6.2 Đề 2

Câu 1. Quan sát về thu nhập (X - USD/tuần) và chi tiêu (Y - USD/tuần) của 10 người, người ta thu được các số liệu sau

X	31	50	47	45	39	50	35	40	45	50
Y	29	42	38	30	29	41	23	36	42	48

- a) Tìm hàm hồi quy tuyến tính mẫu của Y theo X và phát biểu ý nghĩa của các hệ số hồi quy?
- b) Tìm hệ số xác định mô hình và cho biết ý nghĩa của nó?
- c) Tìm khoảng tin cậy của β_1 và β_2 với độ tin cậy 95%?
- d) Kiểm định giả thiết $H_0 : \beta_2 = 0$; $H_1 : \beta_2 \neq 0$ với mức ý nghĩa 5%.
- e) Dự báo điểm cho chi tiêu của một người khi mức thu nhập 40 USD/tuần?
- Câu 2. Người ta cho rằng chi tiêu mặt hàng A (Y ngàn đồng/tháng) không chỉ phụ thuộc vào thu nhập của người tiêu dùng (X triệu đồng/tháng) mà còn phụ thuộc vào giới tính của người tiêu dùng ($D = 1$ nếu là nam và $D = 0$ nếu là nữ). Với một mẫu gồm 20 quan sát, người ta đã ước lượng được mô hình

$$\hat{Y} = 86.542 + 35.976X - 7.438D - 5.845XD$$

$$Se = (32.866) \quad (10.654) \quad (4.117) \quad (1.698)$$

- a) Hãy nêu ý nghĩa của các hệ số hồi quy.
- b) Tìm khoảng tin cậy của các hệ số hồi quy trong hàm hồi quy tổng thể với độ tin cậy 90%.
- c) Với mức ý nghĩa 1%, hãy cho biết có sự khác nhau về chi tiêu đối với mặt hàng A của nam và nữ hay không.
- d) Một nhận định cho rằng hệ số của biến X trong hàm hồi quy tổng thể là 36.5; với mức ý nghĩa 1% hãy cho nhận xét về nhận định này.

Đáp số

Câu 1.

a) $\hat{Y} = -5,45193 + 0,954906X$

b) $R^2 = 0,672$

c) $\beta_1 \in (-29,1983; 18,2944); \beta_2 \in (0,411; 1,4987)$

d) Bác bỏ H_0 . Thu nhập thực sự có ảnh hưởng đến chi tiêu.

e) $\hat{Y}_0 = 32,7443$

Câu 2.

a) Ta nhận xét

$$D = 0, \hat{Y} = 86,542 + 35,976X$$

$$D = 1, \hat{Y} = 79,104 + 30,131X$$

Nếu X như nhau thì chi tiêu mặt hàng A của nữ cao hơn nam: 13.283

b) Khoảng tin cậy

$$\beta_1 \in (29,15796; 143,92604)$$

$$\beta_2 \in (17,374116; 54,57788)$$

$$\beta_3 \in (-14,62628; -0,24972)$$

$$\beta_4 \in (-8,80971; -2,88029)$$

c) Chi tiêu về mặt hàng A của nam và nữ là không giống nhau

d) $|T_2| = |-0,04918| < C = t_{0,005}^{16} = 2,921$ suy ra chưa có cơ sở bác bỏ H_0 . Vậy nhận định trên là phù hợp.

2.6.3 Đề 3

Câu 1. Cho số liệu của một mẫu gồm 15 quan sát, thu thập tại 15 cửa hàng khác nhau thuộc cùng một công ty kinh doanh cùng loại sản phẩm. Trong đó Y là lượng hàng bán được (tấn/tháng); X_2 là chi phí quảng cáo (triệu đồng/tháng) và X_3 giá bán loại hàng này (ngàn đồng/kg)

Y	X_2	X_3
14	5	4
21	9	22
20	8	2.4
18	7	2.8
19	8	2.8
18	8	3
17	6	3.1
17	6	3.3
16	5.7	3.7
15	5.5	3.9
13	4	4.1
12	3	4.3
18.5	7	2.7
19	8.2	2.5
22	9.5	2

Với mức ý nghĩa 5%, hãy trả lời các câu hỏi sau:

1. Tìm hàm hồi quy mẫu và giải thích ý nghĩa kinh tế của các hệ số hồi quy?
2. Tìm hệ số xác định mô hình và nêu ý nghĩa? tính hệ số xác định hiệu chỉnh?
3. Hãy tìm khoảng tin cậy cho các hệ số hồi quy?
4. Chi phí quảng cáo có ảnh hưởng đến lượng hàng bán được hay không?
5. Mô hình có phù hợp với thực tế không? (Có phải cả chi phí quảng cáo lẫn giá bán đều không ảnh hưởng đến lượng hàng bán được?)
6. Hãy dự báo giá trị trung bình và cá biệt của lượng hàng bán được khi mà chi phí quảng cáo là 10 triệu đồng/tháng và giá bán 1kg sản phẩm là 9 nghìn đồng.

Câu 2. Từ các số liệu thu thập ở 30 doanh nghiệp may của thành phố Hồ Chí Minh người ta đã ước lượng được mô hình sau

$$\hat{Y} = -5.54 + 0.112X - 14.48D + 0.25XD$$

$$Se = (10.85) \quad (0.08) \quad (15.47) \quad (0.12)$$

$$R^2 = 0.729$$

$$d = 2.1511$$

Trong đó Y là lợi nhuận (tỉ VND); X là doanh thu (tỉ VND); $D = 1$ nếu giám đốc đã tốt nghiệp cao học và $D = 0$ nếu giám đốc chưa tốt nghiệp cao học.

- Hãy nêu ý nghĩa của các hệ số hồi quy; Cho biết mô hình có xảy ra hiện tượng tự tương quan hay không? (giả sử các giả định khác đều đúng).
- Tìm khoảng tin cậy của các hệ số hồi quy trong hàm hồi quy tổng thể với độ tin cậy 95%.
- Với mức ý nghĩa 5%, hãy cho biết có sự khác nhau về lợi nhuận của doanh nghiệp khi giám đốc đã tốt nghiệp và chưa tốt nghiệp cao học hay không?
- Với mức ý nghĩa 1%, hãy cho biết mô hình trên có phù hợp với thực tế không?

Đáp số

Câu 1.

- $\hat{Y} = 7,18471 + 1,51606X_{2i} + 0,00412X_{3i}$.
- $R^2 = 0,95285$; $\bar{R}^2 = 0,94499$.
- $\beta_1 \in (5,71907; 8,65035)$; $\beta_2 \in (1,29816; 1,73396)$; $\beta_3 \in (-0,07798; 0,08622)$.
- $T_2 = 15,1606$; $|T| > C = 2,179$ suy ra bác bỏ H_0 . Vậy chi phí quảng cáo có ảnh hưởng đến lượng hàng bán được.
- $F = 121,25345 > C$ suy ra bác bỏ H_0 . Vậy mô hình có phù hợp với thực tế.
- $(12,62253; 14,28737)$; $(11,76526; 15,14462)$.

Câu 2.

- Ta nhận xét

$$D = 0, \hat{Y} = -5,54 + 0,112X$$

$$D = 1, \hat{Y} = -20,02 + 0,362X$$

$d = 2,1511 \in (1, 3)$: không xảy ra

- Khoảng tin cậy

$$\beta_1 \in (-27,8476; 16,7676)$$

$$\beta_2 \in (-0,05248; 0,27648)$$

$$\beta_3 \in (-46,28632; 17,32632)$$

$$\beta_4 \in (0,00328; 0,49672)$$

c) Khác nhau

d) $F = 23,31365 > C = 4,64$ suy ra bác bỏ H_0 . Vậy mô hình phù hợp với thực tế.

2.6.4 Đề 4

Câu 1. Dữ liệu sau mô tả tổng chi phí sản xuất (Y : triệu đồng) và lượng sản phẩm (X : ngàn đơn vị) được sản xuất tại một nhà máy trong 11 tháng quan sát như sau:

Y	25	11	34	23	32	20	33	29	25	37	41
X	5	2	8	4	6	4	7	6	5	8	9

Giả sử Y và X có quan hệ tuyến tính

- Hãy ước lượng hàm hồi quy của tổng chi phí theo số lượng sản phẩm. Nếu ý nghĩa của hệ số góc.
- Tính hệ số xác định mô hình? Kiểm định sự phù hợp của mô hình với mức ý nghĩa 5%.
- Với mức ý nghĩa 5%, hãy cho biết nếu số lượng sản phẩm tăng lên 1 ngàn đơn vị thì tổng chi phí sản xuất tăng tối đa là bao nhiêu? tăng tối thiểu là bao nhiêu?
- Hãy ước lượng phương sai của sai số ngẫu nhiên tổng thể với độ tin cậy 95%.
- Dự báo giá trị trung bình và cá biệt của tổng chi phí khi số lượng sản phẩm là 3 ngàn đơn vị, với độ tin cậy 95%.

Câu 2. Cho kết quả xuất ra từ phần mềm Eview như sau

Dependent Variable: Y

Included observations: 40

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1278.580	1332.039	-0.959867	0.3437
X2		0.057329	0.731081	0.4696
X3	-0.993928		-1.940140	0.0605
X4		61.25178	2.052273	0.0477
X5	29.40612		1.956261	0.0585
R-squared	0.322177	Mean dependent var		1598.150
Adjusted R-squared	0.244711	S.D. dependent var		1370.032
F-statistic		Durbin-Watson stat		1.963835

Trong đó:

Y: Số giấy phép xây dựng nhà thực tế,

X_2 : Mật độ dân số tính theo dặm vuông,

X_3 : Giá trị trung bình của căn nhà,

X_4 : Thu nhập trung bình của chủ hộ,

X_5 : Phần trăm tăng dân số từ 1980 đến 1992.

- Viết hàm SRF. Nêu ý nghĩa các hệ số hồi quy.
- Mô hình trên có phù hợp không? kết luận với mức ý nghĩa 5%.
- Hãy ước lượng các hệ số hồi quy tổng thể của biến X_3 và X_5 với độ tin cậy 95%.

Câu 3. Nêu mục đích và kết quả của các kiểm định sau

Kiểm định 1.

Heteroskedasticity Test: White

F-statistic	1.301898	Prob. F(9,30)	0.2769
Obs*R-squared	11.23480	Prob. Chi-Square(9)	0.2600
Scaled explained SS	9.268554	Prob. Chi-Square(9)	0.4129

Kiểm định 2.

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	0.028905	Prob. F(2,34)	0.9715
Obs*R-squared	0.067896	Prob. Chi-Square(2)	0.9666

Đáp số

Câu 1. Ta tính được

$$n = 11;$$

$$\bar{X} = 5,818181; \text{var}(X) = 3,966942;$$

$$\bar{Y} = 28,181818; \text{var}(Y) = 65,785124;$$

$$\hat{\beta}_2 = \frac{\sum X_i Y_i - n \bar{X} \bar{Y}}{\sum X_i^2 - n \bar{X}^2} = 3,995833; \hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} = 4,933333.$$

a) Hàm hồi quy mẫu:

$$\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X = 4,933333 + 3,995833X$$

ý nghĩa hệ số góc: khi lượng sản phẩm tăng lên 1 (ngàn đơn vị) thì tổng chi phí sản xuất trung bình tăng 3,995833 (triệu đồng).

b) Hệ số xác định mô hình:

$$TSS = n \text{Var}(Y) = 723,636363;$$

$$ESS = n \hat{\beta}_2^2 \text{Var}(X) = 696,7280303;$$

$$RSS = TSS - ESS = 26,908333;$$

$$R^2 = \frac{ESS}{TSS} = 0,962815$$

Kiểm định:

+ Đặt giả thiết $H_0 : R^2 = 0; H_1 : R^2 \neq 0$.

+ Với $\alpha = 0,05, C = F_{\alpha}(k-1; n-k) = F_{0,05}(2-1; 11-2) = 5,12$.

$$+ F = \frac{(n-k) R^2}{(k-1)(1-R^2)} = \frac{(11-2) 0,962815}{(2-1)(1-0,962815)} = 233,0331$$

+ $F > C$ nên bác bỏ H_0 . Vậy mô hình phù hợp.

c) Ta có $C = t_{\alpha}^{(n-k)} = t_{0,05}^9 = 1,833$

Ước lượng tối đa cho β_2 :

$$\beta_2 \leq \hat{\beta}_2 + C.se(\hat{\beta}_2) = 3,995833 + 1,833.0,261757 = 4,475634$$

Ước lượng tối thiểu cho β_2 :

$$\beta_2 \geq \hat{\beta}_2 - C.se(\hat{\beta}_2) = 3,995833 - 1,833.0,261757 = 3,516032$$

Vậy với mức ý nghĩa 5%, nếu số lượng sản phẩm tăng lên 1 ngàn đơn vị thì tổng chi phí sản xuất tăng tối đa 4,475634 triệu đồng và tăng tối thiểu là 3,516032 triệu đồng.

d) Ta có

$$\chi_{\frac{\alpha}{2}}^2(n-2) = \chi_{0,025}^2(9) = 19,023; \quad \chi_{1-\frac{\alpha}{2}}^2(n-2) = \chi_{0,975}^2(9) = 2,7$$

Khoảng ước lượng cho phương sai của sai số ngẫu nhiên

$$\sigma^2 \in \left[\frac{(n-2)\hat{\sigma}^2}{\chi_{\frac{\alpha}{2}}^2(n-2)}; \frac{(n-2)\hat{\sigma}^2}{\chi_{1-\frac{\alpha}{2}}^2(n-2)} \right] = [1,414516; 9,96605]$$

e) Ta có

$$\begin{aligned} X_0 &= 3; \quad \hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_0 = 16,920832; \quad C = t_{\frac{\alpha}{2}}^{(n-k)} = t_{0,025}^9 = 2,262; \\ Var(\hat{Y}_0) &= \left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{nVar(X)} \right] \hat{\sigma}^2 = 0,81597; \quad \Rightarrow se(\hat{Y}_0) = 0,903311; \\ Var(Y_0 - \hat{Y}_0) &= \hat{\sigma}^2 + Var(\hat{Y}_0) = 3,805785; \quad \Rightarrow se(Y_0 - \hat{Y}_0) = 1,950842 \end{aligned}$$

Khoảng dự báo trung bình cho tổng chi phí

$$\left[\hat{Y}_0 - C.se(\hat{Y}_0); \hat{Y}_0 + C.se(\hat{Y}_0) \right] = [14,877543; 18,964121]$$

Khoảng dự báo cá biệt cho tổng chi phí

$$\left[\hat{Y}_0 - C.se(Y_0 - \hat{Y}_0); \hat{Y}_0 + C.se(Y_0 - \hat{Y}_0) \right] = [12,508027; 21,333637]$$

Câu 2. Ta tính được

$$\begin{aligned} \hat{\beta}_2 &= T_2.se(\hat{\beta}_2) = 0,041912; \quad \hat{\beta}_4 = T_4.se(\hat{\beta}_4) = 125,705374; \\ se(\hat{\beta}_3) &= \frac{\hat{\beta}_3}{T_3} = 0,512297; \quad se(\hat{\beta}_5) = \frac{\hat{\beta}_5}{T_5} = 15,031798; \\ F &= \frac{(n-k)R^2}{(k-1)(1-R^2)} = 4,158975 \end{aligned}$$

a) Hàm hồi quy mẫu:

$$\widehat{Y} = -1278,58 + 0,041912X_2 - 0,993928X_3 + 125,705374X_4 + 29,40612X_5$$

ý nghĩa các hệ số hồi quy: (sinh viên dựa theo bài giảng trên lớp)

b) Kiểm định sự phù hợp của mô hình

+ Đặt giả thiết $H_0 : R^2 = 0; H_1 : R^2 \neq 0$.

+ Với $\alpha = 0,05, C = F_{\alpha}(k-1; n-k) = F_{0,05}(5-1; 40-5) \approx 2,65$.

$$+ F = \frac{(n-k)R^2}{(k-1)(1-R^2)} = \frac{(40-5)0,322177}{(5-1)(1-0,322177)} = 4,158975$$

+ $F > C$ nên bác bỏ H_0 . Vậy mô hình trên là phù hợp.

c) Ta có $C = t_{\frac{\alpha}{2}}^{(n-k)} = t_{0,025}^{35} \approx 2,03$

Khoảng ước lượng cho β_3 :

$$\beta_3 \in \left[\widehat{\beta}_3 - C.se(\widehat{\beta}_3); \widehat{\beta}_3 + C.se(\widehat{\beta}_3) \right] = [-2,033891; 0,046035]$$

Khoảng ước lượng cho β_5 :

$$\beta_5 \in \left[\widehat{\beta}_5 - C.se(\widehat{\beta}_5); \widehat{\beta}_5 + C.se(\widehat{\beta}_5) \right] = [-1,10843; 59,92067]$$

Câu 3. Mục đích và kết quả của các kiểm định

Kiểm định 1.

Mục đích: kiểm định xem mô hình có xảy ra hiện tượng phương sai thay đổi hay không.

Kết quả: do $P_{value} = 0,26 > \alpha = 0,05$ nên suy ra chưa có cơ sở bác bỏ giả thiết. Vậy mô hình không xảy ra hiện tượng phương sai thay đổi.

Kiểm định 2.

Mục đích: kiểm định xem mô hình có xảy ra hiện tượng tự tương quan hay không.

Kết quả: do $P_{value} = 0,9666 > \alpha = 0,05$ nên suy ra chưa có cơ sở bác bỏ giả thiết. Vậy mô hình không xảy ra hiện tượng tự tương quan.

2.6.5 Đề 5

Câu 1. Giám đốc của cửa hàng bán đồ điện máy cho rằng kinh nghiệm là nhân tố quan trọng nhất quyết định sự thành công của một nhân viên kinh doanh. Để chứng minh cho nhận định của mình, anh ta thu thập dữ liệu gồm doanh số hàng tháng của một nhân viên (Y : triệu đồng/tháng) và kinh nghiệm làm việc của nhân viên đó (X : năm), dữ liệu của 14 nhân viên được cho trong bảng sau

Y	28	30	36	39	57	45	70	47	50	60	66	32	48	51
X	1	3	3	4	10	6	14	8	9	10	12	7	7	8

Giả sử Y và X có quan hệ tuyến tính

- Hãy ước lượng hàm hồi quy của doanh số hàng tháng theo kinh nghiệm. Nếu ý nghĩa của hệ số góc.
- Tính hệ số xác định mô hình? Kiểm định sự phù hợp của mô hình với mức ý nghĩa 5%.
- Hãy ước lượng các hệ số hồi quy tổng thể với độ tin cậy 95%.
- Với mức ý nghĩa 5%, hãy cho biết nếu kinh nghiệm tăng lên 1 năm thì doanh số hàng tháng tăng tối đa là bao nhiêu? tăng tối thiểu là bao nhiêu?
- Dự báo giá trị trung bình và cá biệt của doanh số hàng tháng khi kinh nghiệm là 5 năm, với độ tin cậy 95%.

Câu 2. Cho kết quả xuất ra từ phần mềm Eview như sau

Dependent Variable: Y

Included observations: 392

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	44.37096	1.480685	29.96651	0.0000
X2		0.413067	-0.648313	0.5172
X3	-0.012674		-1.536225	0.1253
X4		0.000714	-7.995143	0.0000
R-squared	0.699293	Mean dependent var		23.44592
S.E. of regression	4.296531	Akaike info criterion		5.763645
F-statistic		Durbin-Watson stat		0.834161

Trong đó:

Y : Số ki lô mét đi được trên một lít nhiên liệu (km),

X_2 : Số động cơ trong xe,

X_3 : Dung tích xy-lanh (m^3),

X_4 : Trọng lượng của xe (kg)

- Viết hàm SRF. Nêu ý nghĩa các hệ số hồi quy.
- Mô hình trên có phù hợp không? kết luận với mức ý nghĩa 5%.
- Hãy ước lượng các hệ số hồi quy tổng thể của biến X_2 và X_4 với độ tin cậy 95%.

Câu 3. Nêu mục đích và kết quả của các kiểm định sau

Kiểm định 1.

Heteroskedasticity Test: Glejser			
F-statistic	13.28274	Prob. F(3,388)	0.0000
Obs*R-squared	36.50944	Prob. Chi-Square(3)	0.0000
Scaled explained SS	41.99027	Prob. Chi-Square(3)	0.0000

Kiểm định 2.

Breusch-Godfrey Serial Correlation LM Test:			
F-statistic	200.1447	Prob. F(1,387)	0.0000
Obs*R-squared	133.6242	Prob. Chi-Square(1)	0.0000

Đáp số

Câu 1. Ta tính được

$$n = 14;$$

$$\bar{X} = 7,285714; \text{Var}(X) = 12,489796;$$

$$\bar{Y} = 47,07143; \text{Var}(Y) = 160,637755;$$

$$\hat{\beta}_2 = \frac{\sum X_i Y_i - n\bar{X}\bar{Y}}{\sum X_i^2 - n\bar{X}^2} = 3,355392; \hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} = 22,625;$$

a) Hàm hồi quy mẫu:

$$\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X = 22,625 + 3,355392X$$

ý nghĩa hệ số góc: khi kinh nghiệm tăng lên 1 năm thì doanh số hàng tháng trung bình tăng 3,355392 triệu đồng/tháng.

b) Hệ số xác định mô hình: $TSS = n \cdot Var(Y) = 2248,928571$;

$$ESS = n \cdot \hat{\beta}_2^2 \cdot Var(X) = 1968,656513;$$

$$RSS = TSS - ESS = 280,2721;$$

$$R^2 = \frac{ESS}{TSS} = 0,875375;$$

Kiểm định:

+ Đặt giả thiết $H_0 : R^2 = 0$; $H_1 : R^2 \neq 0$.

+ Với $\alpha = 0,05$, $C = F_{\alpha}(k-1; n-k) = F_{0,05}(2-1; 14-2) = 4,75$.

$$+ F = \frac{(n-k)R^2}{(k-1)(1-R^2)} = \frac{(14-2)0,875375}{(2-1)(1-0,875375)} = 84,288867$$

+ $F > C$ nên bác bỏ H_0 . Vậy mô hình phù hợp.

c) Ta có

$$\hat{\sigma}^2 = \frac{RSS}{n-2} = 23,356;$$

$$Var(\hat{\beta}_1) = \left[\frac{1}{n} + \frac{\bar{X}^2}{n \cdot Var(X)} \right] \hat{\sigma}^2 = 8,7585 \Rightarrow se(\hat{\beta}_1) = 2,95947;$$

$$Var(\hat{\beta}_2) = \frac{\hat{\sigma}^2}{n \cdot Var(X)} = 0,133572 \Rightarrow se(\hat{\beta}_2) = 0,365475;$$

$$C = t_{\frac{\alpha}{2}}^{(n-k)} = t_{0,025}^{12} = 2,179$$

Khoảng ước lượng cho β_1 :

$$\beta_1 \in \left[\hat{\beta}_1 - C \cdot se(\hat{\beta}_1); \hat{\beta}_1 + C \cdot se(\hat{\beta}_1) \right] = [16,176315; 29,073685]$$

Khoảng ước lượng cho β_2 :

$$\beta_2 \in \left[\hat{\beta}_2 - C \cdot se(\hat{\beta}_2); \hat{\beta}_2 + C \cdot se(\hat{\beta}_2) \right] = [2,559022; 4,151762]$$

d) Ta có $C = t_{\alpha}^{(n-k)} = t_{0,05}^{12} = 1,782$

Ước lượng tối đa cho β_2 :

$$\beta_2 \leq \hat{\beta}_2 + C.se(\hat{\beta}_2) = 3,355392 + 1,782.0,365475 = 4,006668$$

Ước lượng tối thiểu cho β_2 :

$$\beta_2 \geq \hat{\beta}_2 - C.se(\hat{\beta}_2) = 3,355392 - 1,782.0,365475 = 2,704116$$

Vậy với mức ý nghĩa 5%, nếu kinh nghiệm tăng lên 1 năm thì doanh số hàng tháng tăng tối đa 4,006668 triệu đồng/tháng và tăng tối thiểu là 2,704116 triệu đồng/tháng.

e) Ta có

$$X_0 = 5; \hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_0 = 39,40196; C = t_{\frac{\alpha}{2}}^{(n-k)} = t_{0,025}^{12} = 2,179;$$

$$Var(\hat{Y}_0) = \left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{n.Var(X)} \right] \hat{\sigma}^2 = 3,033445 \Rightarrow se(\hat{Y}_0) = 1,741679;$$

$$Var(Y_0 - \hat{Y}_0) = \hat{\sigma}^2 + Var(\hat{Y}_0) = 26,389445 \Rightarrow se(Y_0 - \hat{Y}_0) = 5,137066$$

Khoảng dự báo trung bình cho doanh số hàng tháng

$$\left[\hat{Y}_0 - C.se(\hat{Y}_0); \hat{Y}_0 + C.se(\hat{Y}_0) \right] = [35,606841; 43,197079]$$

Khoảng dự báo cá biệt cho doanh số hàng tháng

$$\left[\hat{Y}_0 - C.se(Y_0 - \hat{Y}_0); \hat{Y}_0 + C.se(Y_0 - \hat{Y}_0) \right] = [28,208293; 50,595627]$$

Câu 2. Ta tính được

$$\hat{\beta}_2 = T_2.se(\hat{\beta}_2) = -0,267797; \hat{\beta}_4 = T_4.se(\hat{\beta}_4) = -0,005709;$$

$$se(\hat{\beta}_3) = \frac{\hat{\beta}_3}{T_3} = 0,00825;$$

$$F = \frac{(n-k)R^2}{(k-1)(1-R^2)} = 300,7642$$

a) Hàm hồi quy mẫu:

$$\hat{Y} = 44,37096 - 0,267797X_2 - 0,012674X_3 - 0,005709X_4$$

ý nghĩa các hệ số hồi quy: (sinh viên dựa theo bài giảng trên lớp)

b) Kiểm định sự phù hợp của mô hình

+ Đặt giả thiết $H_0 : R^2 = 0; H_1 : R^2 \neq 0$.

+ Với $\alpha = 0,05, C = F_{\alpha}(k-1; n-k) = F_{0,05}(4-1; 392-4) \approx 2,60$.

$$+ F = \frac{(n-k) R^2}{(k-1)(1-R^2)} = \frac{(392-4) 0,699293}{(4-1)(1-0,699293)} = 300,7642$$

+ $F > C$ nên bác bỏ H_0 . Vậy mô hình trên là phù hợp.

c) Ta có $C = t_{\frac{\alpha}{2}}^{(n-k)} = t_{0,025}^{388} \approx 1,96$

Khoảng ước lượng cho β_2 :

$$\beta_2 \in \left[\widehat{\beta}_2 - C.se(\widehat{\beta}_2); \widehat{\beta}_2 + C.se(\widehat{\beta}_2) \right] = [-1,077408; 0,541814]$$

Khoảng ước lượng cho β_4 :

$$\beta_4 \in \left[\widehat{\beta}_4 - C.se(\widehat{\beta}_4); \widehat{\beta}_4 + C.se(\widehat{\beta}_4) \right] = [-0,007108; -0,00431]$$

Câu 3. Mục đích và kết quả của các kiểm định

Kiểm định 1.

Mục đích: kiểm định xem mô hình có xảy ra hiện tượng phương sai thay đổi hay không.

Kết quả: do $P_{value} = 0,0000 < \alpha = 0,05$ nên suy ra bác bỏ giả thiết. Vậy mô hình xảy ra hiện tượng phương sai thay đổi.

Kiểm định 2.

Mục đích: kiểm định xem mô hình có xảy ra hiện tượng tự tương quan hay không.

Kết quả: do $P_{value} = 0,0000 < \alpha = 0,05$ nên suy ra bác bỏ giả thiết. Vậy mô hình xảy ra hiện tượng tự tương quan.

2.6.6 Đề 6

Cho một mẫu gồm các giá trị quan sát sau:

Y_i	10	6	5	8	7	8	7	7	8	9
X_i	2	3	4	2	3	3	4	3	3	2
Z_i	12	9	9	10	9	11	10	9	11	11

Trong đó:

Y là lượng khách đi xe buýt (đơn vị: 100 ngàn người),

X là giá vé (đơn vị: ngàn đồng),

Z là giá xăng (đơn vị ngàn đồng/lít).

Câu 1. Giả sử Y và X có quan hệ tuyến tính

- Hãy ước lượng hàm hồi quy của Y theo X . Nêu ý nghĩa của hệ số góc.
- Tính hệ số xác định mô hình và kiểm định sự phù hợp của mô hình với mức ý nghĩa 5%.
- Tìm khoảng tin cậy cho các hệ số hồi quy tổng thể, với độ tin cậy 95%.
- Với mức ý nghĩa 5%. Nếu giá vé tăng 1 ngàn đồng thì số lượng khách đi xe buýt tăng tối đa là bao nhiêu?
- Dự báo lượng khách đi xe buýt trung bình khi giá vé là 4,5 ngàn đồng, với độ tin cậy 95%.

Câu 2. Với số liệu đã cho, ước lượng mô hình $Y_i = \beta_1 + \beta_2 X_i + \beta_3 Z_i + \varepsilon_i$ bằng EVIEWS ta được

Dependent Variable: Y					
Variable	Coefficient	Std. Error	t-Statistic	Prob.	
C	1.544503	2.185040	0.706853	0.5025	
X		0.251340	-3.437074	0.0109	
Z	0.837696	0.168518			
R-squared	0.916230	Mean dependent var		7.500000	
Adjusted R-squared	0.892296	S.D. dependent var		1.433721	
S.E. of regression	0.470522	Akaike info criterion		1.573378	
Sum squared resid	1.549738	Schwarz criterion		1.664153	
Log likelihood	-4.866890	F-statistic			

- Viết hàm hồi quy mẫu (SRF) và nêu ý nghĩa các hệ số hồi quy riêng.
- Với mức ý nghĩa 5%. Xét xem giá xăng có ảnh hưởng đến lượng khách đi xe buýt không?
- Mô hình trên có phù hợp không? Kết luận với mức ý nghĩa 5%.

Câu 3. Nêu mục đích và kết quả của các kiểm định sau

Kiểm định 1.

Wald Test:

Equation: Untitled

Test Statistic	Value	df	Probability
t-statistic	-7.995143	388	0.0000
F-statistic	63.92231	(1, 388)	0.0000
Chi-Square	63.92231	1	0.0000

Null Hypothesis: $C(4) = 0$

Null Hypothesis Summary:

Normalized Restriction (=0)	Value	Std. Err.
C(4)	-0.005708	0.000714

Kiểm định 2.

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	0.028905	Prob. F(2,34)	0.9715
Obs*R-squared	0.067896	Prob. Chi-Square(2)	0.9666

Đáp số

Sinh viên tiến hành giải theo các bước tương tự như đã hướng dẫn trong các bài tập phía trên, sau đây là một số kết quả tính toán để áp dụng cho Câu 1.

$$n = 10;$$

$$\bar{X} = 2,9; \text{Var}(X) = 0,49;$$

$$\bar{Y} = 7,5; \text{Var}(Y) = 1,85;$$

$$\hat{\beta}_2 = \frac{\sum X_i Y_i - n \bar{X} \bar{Y}}{\sum X_i^2 - n \bar{X}^2} = -1,530612; \hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} = 11,93878;$$

$$TSS = n \cdot \text{Var}(Y) = 18,5;$$

$$ESS = n \cdot \hat{\beta}_2^2 \cdot \text{Var}(X) = 11,479592;$$

$$RSS = TSS - ESS = 7,020408;$$

$$R^2 = \frac{ESS}{TSS} = 0,620518;$$

$$\hat{\sigma}^2 = \frac{RSS}{n-2} = 0,877551;$$

$$Var(\hat{\beta}_1) = \left[\frac{1}{n} + \frac{\bar{X}^2}{n \cdot Var(X)} \right] \hat{\sigma}^2 = 1,593919 \Rightarrow se(\hat{\beta}_1) = 1,262505;$$

$$Var(\hat{\beta}_2) = \frac{\hat{\sigma}^2}{n \cdot Var(X)} = 0,179092 \Rightarrow se(\hat{\beta}_2) = 0,423193;$$

$$Var(\hat{Y}_0) = \left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{n \cdot Var(X)} \right] \hat{\sigma}^2 = 0,546231 \Rightarrow se(\hat{Y}_0) = 0,739074;$$

$$Var(Y_0 - \hat{Y}_0) = \hat{\sigma}^2 + Var(\hat{Y}_0) = 1,423782 \Rightarrow se(Y_0 - \hat{Y}_0) = 1,193223$$

2.6.7 Đề 7

Câu 1. Cho một mẫu cụ thể, trong đó X là thu nhập, Y là chi tiêu (đơn vị là triệu đồng)

X	7,2	8,1	8,3	9,1	10,4	11,2	12,3	13,4	16,2	17,3
Y	5,7	6,2	5,8	7,2	8,1	9,1	9,1	10,3	12,1	13,2

Hồi quy tuyến tính Y theo X

- Hãy ước lượng các hệ số hồi quy và giải thích ý nghĩa kinh tế của hệ số góc.
- Tính hệ số xác định và giải thích ý nghĩa. Với mức ý nghĩa 5%, mô hình có phù hợp không?

Câu 2. Cho bảng kết quả (tóm tắt) chạy trên Eviews của một mô hình kinh tế lượng như sau

Dependent Variable: Y

Method: Least Squares

Included observations: 17

Variable	Coefficient	Se	t	P-value
C	5432		26,5432	0,0001
X_2	12,4563	0,2132		0,0001
X_3	23,1234		6,3212	0,0002

- Hãy xác định hàm hồi quy mẫu và giải thích ý nghĩa của các hệ số hồi quy riêng.

- b) Với độ tin cậy 95%, hãy tìm khoảng ước lượng tối thiểu cho hệ số của X_2 và giải thích ý nghĩa kết quả.
- c) Với độ tin cậy 99%, nếu X_3 thay đổi một đơn vị trong điều kiện giá trị của X_2 không đổi thì giá trị trung bình của Y sẽ thay đổi trong khoảng nào?
- d) Với mức ý nghĩa 5%, X_2 có thực sự tác động lên biến động của Y không?

Câu 3. Cho kết quả kiểm định sau, của một mô hình hồi qui

Heteroskedasticity Test: White			
F-statistic	0.519117	Prob. (F-statistic)	0.0565
Obs*R-squared	4.174768	Prob. (Chi-Square)	0.0545

Với mức ý nghĩa 5%, mô hình gốc có phương sai sai số thay đổi không?

Đáp số

Sinh viên tiến hành giải theo các bước tương tự như đã hướng dẫn trong các bài tập phía trên, sau đây là một số kết quả tính toán để áp dụng cho Câu 1.

$$n = 10;$$

$$\bar{X} = 11,35; \text{Var}(X) = 10,7105;$$

$$\bar{Y} = 8,68; \text{Var}(Y) = 6,0956;$$

$$\hat{\beta}_2 = \frac{\sum X_i Y_i - n \bar{X} \bar{Y}}{\sum X_i^2 - n \bar{X}^2} = 0,749078; \hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} = 0,177965;$$

$$TSS = n \cdot \text{Var}(Y) = 60,956;$$

$$ESS = n \cdot \hat{\beta}_2^2 \cdot \text{Var}(X) = 60,098529;$$

$$RSS = TSS - ESS = 0,857471;$$

$$R^2 = \frac{ESS}{TSS} = 0,985933;$$

$$\hat{\sigma}^2 = \frac{RSS}{n-2} = 0,107184;$$

$$\text{Var}(\hat{\beta}_1) = \left[\frac{1}{n} + \frac{\bar{X}^2}{n \cdot \text{Var}(X)} \right] \hat{\sigma}^2 = 0,139636 \Rightarrow se(\hat{\beta}_1) = 0,373679;$$

$$\text{Var}(\hat{\beta}_2) = \frac{\hat{\sigma}^2}{n \cdot \text{Var}(X)} = 0,001 \Rightarrow \text{se}(\hat{\beta}_2) = 0,031634;$$

Chương 3

Thực hành Eviews

- ✘ Có rất nhiều phần mềm hỗ trợ cho phân tích kinh tế lượng, một trong số đó là **Eviews**. Eviews cung cấp các công cụ phân tích dữ liệu phức tạp, hồi quy và dự báo chạy trên Windows. Với Eviews ta có thể nhanh chóng xây dựng các mối quan hệ kinh tế từ dữ liệu có sẵn và sử dụng mối quan hệ này để dự báo cho các vấn đề trong tương lai.
- ✘ Workfile là một tập tin làm việc của Eviews. Workfile chứa các đối tượng của Eviews.
- ✘ Mỗi đối tượng bao gồm một tập hợp các thông tin có liên quan đến một lĩnh vực phân tích cụ thể. Làm việc trên Eviews chủ yếu liên quan đến các đối tượng chứa trong một Workfile.

3.1 Cài đặt Eviews 8

Sử dụng thư mục "Eviews 8 Enterprise Edition" đã được cung cấp, quá trình cài đặt Eviews 8 gồm ba bước

- ✘ Chạy file **Eviews8Installer.exe** để cài đặt, gõ vào password: demo.
- ✘ Sau khi cài đặt xong chép file **Eviews.8-patch (32-bit).exe** trong thư mục **patch** vào thư mục cài đặt trong ổ **C:\programfile\eview 8**.
- ✘ Click chuột phải vào file **Eviews.8-patch (32-bit).exe** chạy bằng Administrator nhấn Patch là xong.

3.2 Khởi động Eviews 8

Có hai cách khởi động Eviews 8

- ✘ Double click vào biểu tượng shortcut Eviews 8 trên màn hình máy tính.
- ✘ \Start\Programs\Eviews 8\Eviews 8.

3.3 Nhập dữ liệu cho Eviews 8

Có nhiều cách nhập dữ liệu cho Eviews 8, ta có thể liệt kê ba cách điển hình: nhập trực tiếp, copy-paste, lấy dữ liệu từ file có sẵn. Thông thường để tiết kiệm thời gian cũng như công sức, trên thực tế người ta thường sử dụng cách thứ ba, sau đây là một số đường dẫn để thực hiện

- ✘ Khởi động Eviews 8\File\New\Workfile (Ctrl + N). Từ cửa sổ Eviews chọn Quick\Empty Group (Edit Series), sau đó nhập dữ liệu trực tiếp vào cửa sổ Group hoặc copy từ một file dữ liệu đã có vào cửa sổ này.
- ✘ Khởi động Eviews 8\Open a Foreign file (such as Excel)\tìm đến thư mục cần lấy tập tin Excel và double click vào tập tin này \Next\Next\Finish.
- ✘ Khởi động Eviews 8\File\Open\Foreign Data as Workfile\tìm đến thư mục cần lấy tập tin Excel và double click vào tập tin này \Next\Next\Finish.

3.4 Thống kê mô tả

- ✘ **Tạo đối tượng:** Object\New Object\....
- ✘ **Hiện bảng dữ liệu:** Chọn đối tượng\Show\Ok\Name\Ok.
- ✘ **Các giá trị thống kê:** Quick\Group Statistics\Descriptive Statistics\Common sample\Chọn các đối tượng cần nghiên cứu\....

Giải thích bảng kết quả thống kê mô tả

- **Mean:** Trung bình.
- **Median:** Trung vị.
- **Maximum:** Giá trị lớn nhất.
- **Minimum:** Giá trị nhỏ nhất.
- **Std. Dev:** Độ lệch chuẩn.

- **Skewness:** Hệ số bất đối xứng.
 - **Kurtosis:** Hệ số nhọn.
 - **Jarque - Bera:** Kiểm định phân phối chuẩn.
 - **Sum:** Tổng các quan sát.
 - **Sum sq. Dev:** Độ lệch chuẩn của tổng bình phương.
 - **Observations:** Số quan sát (cỡ mẫu).
- ✘ **Ma trận tương quan:** Quick\Group Statistics\Correlations\Chọn các đối tượng cần nghiên cứu\Ok\Freeze\Name\Ok.
- ✘ **Ma trận hiệp phương sai:** Quick\Group Statistics\Covariances\Chọn các đối tượng cần nghiên cứu\Ok\Freeze\Name\Ok. Hoặc trong cửa sổ Equation chọn View\Covariance Matrix\....
- ✘ **Đồ thị:** Quick\Graph\gõ vào đối tượng cần khảo sát\Ok\....(ví dụ: Scatter, Distribution,...)
- ✘ **Định mẫu:** Workfile\Sample\điều chỉnh kích thước mẫu cần khảo sát\Ok.
- ✘ **Khảo sát từng biến:** chọn biến\Show\....
- View\ Descriptive Statistics & test\ Histogram and Stats.
 - View\ Graph\...
 - View\ Label. Trở lại chọn View\Spread Sheet.

3.5 Ước lượng các hệ số của mô hình hồi quy

- ✘ **Hàm hồi quy mẫu (SRF):** Quick\Estimate Equation\mô hình hồi quy (y c x_2 x_3)\Ok\Name\Ok. Hoặc gõ lệnh trực tiếp vào Command window (ví dụ: ls y c x_2 x_3). Để hiện hàm hồi quy chọn View\Representations\..., trở lại bảng kết quả hồi quy chọn View\Estimation Output.

Phân tích bảng kết quả hồi quy:

- **Dependent Variable:** Tên biến phụ thuộc.
- **Method: Least Squares:** Phương pháp bình phương tối thiểu.
- **Date - Time:** Ngày giờ thực hiện.
- **Sample:** Số liệu mẫu đang khảo sát (ví dụ: 1 - 10).
- **Included observations:** Cỡ mẫu (số các quan sát, ví dụ: 10).

- **Cột Variable:** Các biến giải thích có trong mô hình (trong đó C là hệ số bị chặn).
- **Cột Coefficient:** Giá trị các hệ số hồi quy $\hat{\beta}_j$.
- **Cột Std. Error:** Sai số chuẩn của các hệ số hồi quy $Se(\hat{\beta}_j)$.
- **Cột t - Statistic:** Giá trị thống kê t tương ứng $t_j = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)}$.
- **Cột Prob:** Giá trị xác suất (p-value) của thống kê t tương ứng

$$p - value_j = P(t > t_j)$$

- **R-Squared:** Hệ số xác định mô hình (R^2).
- **Adjusted R - Squared:** Hệ số xác định có hiệu chỉnh (\bar{R}^2).
- **S.E. of regression:** Giá trị ước lượng cho σ : $\hat{\sigma}$ (sai số chuẩn của hồi quy).
- **Sum squared resid:** Tổng bình phương các sai lệch (phần dư) (RSS).
- **Log likelihood:** Tiêu chuẩn ước lượng hợp lý (Logarit của hàm hợp lý).
- **Durbin - Watson stat:** Thống kê Durbin - Watson.
- **Mean dependent var:** Giá trị trung bình mẫu của biến phụ thuộc.
- **S.D. dependent var:** Độ lệch chuẩn mẫu hiệu chỉnh của biến phụ thuộc.
- **Akaike info criterion:** Tiêu chuẩn Akaike.
- **Schwarz info criterion:** Tiêu chuẩn Schwarz.
- **F - Statistic:** Giá trị của thống kê F.
- **Prob (F - Statistic):** Giá trị xác suất (p-value) của thống kê F tương ứng

$$P_{value} = P(F > F - statistic)$$

✘ **Đồ thị phần dư:** Equation\Resids\Freeze\Name\Ok. Hoặc Equation\ View\ Actual, Fitted, Residual\Actual, Fitted, Residual Graph\...

✘ **Khoảng tin cậy cho các hệ số hồi quy:** Equation\Coefficient Diagnostics\ Confidence Intervals\Ok\Freeze\Name\Ok.

3.6 Kiểm định sự vi phạm các giả thiết của mô hình hồi quy

3.6.1 Hiện tượng đa cộng tuyến

Để nhận biết hiện tượng đa cộng tuyến, ta thường áp dụng một số cách sau đây

- ✘ **Hệ số xác định** trong mô hình hồi quy gốc có giá trị rất cao, trong khi các giá trị t quan sát lại nhỏ.

Quick\Estimate Equation\mô hình hồi quy (y c x2 x3)\....

- ✘ **Tìm ma trận tương quan** giữa các biến giải thích trong mô hình. Theo kinh nghiệm, nếu hệ số tương quan giữa hai biến giải thích > 0.8 mà giá trị t quan sát thấp thì mô hình có đa cộng tuyến cao.

Quick\Group Statistics\ Correlations\ gõ vào hai biến giải thích (X2 X3)\ Ok\ Freeze\Name\Ok.

- ✘ **Thực hiện hồi quy phụ:** lần lượt chọn một trong số các biến giải thích làm biến phụ thuộc rồi hồi quy theo tất cả các biến giải thích còn lại trong mô hình. Theo kinh nghiệm, nếu hệ số xác định của các mô hình hồi quy phụ (R_j^2) có giá trị từ 0.8 trở lên thì được coi là có đa cộng tuyến cao.

Quick\Estimate Equation\mô hình hồi quy phụ (x2 c x3)\....

- ✘ **Sử dụng nhân tử phóng đại phương sai** $VIF_j = \frac{1}{1 - R_j^2}$. Nếu $VIF_j \geq 10$ (tương đương $R_j^2 \geq 0.9$) thì kết luận mô hình gốc có đa cộng tuyến cao. Sau khi thực hiện hồi quy gốc, từ cửa sổ Equation ta dùng lệnh:

View\Coefficient Diagnostics\Variance Inflation Factors.

Các giá trị VIF_j tương ứng với cột **Centered VIF**.

3.6.2 Phương sai của sai số ngẫu nhiên thay đổi

+ Đặt giả thiết:

H_0 : mô hình không xảy ra hiện tượng phương sai thay đổi;

H_1 : mô hình xảy ra hiện tượng phương sai thay đổi.

+ Thực hiện trên Eviews:

Equation\View\ Residual Diagnostics\ Heteroskedasticity Tests\ chọn kiểu kiểm

định (White, Glejser, Breusch-Pagan-Godfrey,...)

+ Kết luận

$P_{value} > \alpha$: chấp nhận H_0 ; $P_{value} < \alpha$: bác bỏ H_0 .

3.6.3 Hiện tượng tự tương quan

+ Đặt giả thiết:

H_0 : mô hình không xảy ra hiện tượng tự tương quan (bậc 2,...);

H_1 : mô hình xảy ra hiện tượng tự tương quan (bậc 2,...).

+ Thực hiện trên Eviews:

Equation\View\Residual Diagnostics\Serial Correlation LM Test\ chọn bậc tương quan (vd: 2)

+ Kết luận

$P_{value} > \alpha$: chấp nhận H_0 ; $P_{value} < \alpha$: bác bỏ H_0 .

3.6.4 Kiểm định biến có cần thiết trong mô hình hay không (kiểm định Wald)

+ Đặt giả thiết:

H_0 : biến (X2) không cần thiết cho mô hình;

H_1 : biến (X2) cần thiết cho mô hình.

+ Thực hiện trên Eviews:

Equation\View\Coefficient Diagnostics\Wald Test...\c(2) = 0.

+ Kết luận

$P_{value} > \alpha$: chấp nhận H_0 ; $P_{value} < \alpha$: bác bỏ H_0 .

3.6.5 Kiểm định biến bị bỏ sót trong mô hình

+ Đặt giả thiết:

H_0 : biến (X3) không ảnh hưởng tới Y ($\beta_3 = 0$);

H_1 : biến (X3) ảnh hưởng tới Y ($\beta_3 \neq 0$).

+ Thực hiện trên Eviews:

Tìm hàm hồi quy mẫu của Y theo X2. Từ cửa sổ Equation ta thực hiện lệnh: View\Coefficient Diagnostics\ Omitted Variables Test...\ X3.

+ Kết luận

$P_{value} > \alpha$: chấp nhận H_0 ; $P_{value} < \alpha$: bác bỏ H_0 .

3.6.6 Sai số ngẫu nhiên không có phân phối chuẩn

+ Đặt giả thiết:

H_0 : Sai số ngẫu nhiên có phân phối chuẩn;

H_1 : Sai số ngẫu nhiên không có phân phối chuẩn.

+ Thực hiện trên Eviews:

Equation\View\ Residual Diagnostics\ Histogram - Normality.

+ Kết luận

$P_{value} > \alpha$: chấp nhận H_0 ; $P_{value} < \alpha$: bác bỏ H_0 .

3.6.7 Kiểm định Chow trong mô hình hồi quy với biến giả

+ Đặt giả thiết:

H_0 : Hai mô hình hồi quy là như nhau;

H_1 : Hai mô hình hồi quy là khác nhau.

+ Thực hiện trên Eviews:

Tìm hàm hồi quy mẫu của Y theo X. Từ cửa sổ Equation ta thực hiện lệnh: View\ Stability Diagnostics\ Chow Breakpoint Test\ gõ vào năm đầu tiên của thời kỳ thứ hai (vd: 1955)\Ok.

+ Kết luận

$P_{value} > \alpha$: chấp nhận H_0 ; $P_{value} < \alpha$: bác bỏ H_0 .

3.7 Dự báo bằng mô hình hồi quy

Có nhiều phương pháp khác nhau để dự báo, một trong số đó là dự báo dựa trên mô hình hồi quy. Ta cũng lưu ý rằng mô hình hồi quy dùng để dự báo phải đảm bảo là một mô hình đủ tốt, có nghĩa là các giả thiết về mô hình hoặc không vi phạm hoặc đã được khắc phục nhằm đảm bảo độ tin cậy cho kết quả dự báo được. Các bước dự báo được trình bày ngắn gọn bằng các đường dẫn sau:

✦ Nhập dữ liệu

- File\ Open\ Eviews Workfile\....\ gdp-xnkv.nwf1 \
- Proc\ Structure\ Resize current page\ 2013 → 2014\....
- Chọn GDP CPI NK XK nhấn Show
- Group\ Edit...\....(thêm dữ liệu)

✦ Xác định các thông số

- Quick\ Estimate Equation\ GDP C CPI NK XK.
- Equation\ Forecast\....

$$\left\{ \begin{array}{l} gdpf \rightarrow gdpf \\ \quad \rightarrow Se1 \end{array} \right.$$

- Object\ New object\ Series\.... Se2.
- Workfile\Genr\....Se2 = Sqr((Se1)² - (eq01.@Se)²)

✦ Tìm khoảng dự báo

- Workfile\Genr\....

$$\text{canduoitrungbinh} = \text{Gdpf} - @qtdist(1 - \frac{\alpha}{2}, n - k) * \text{Se2}$$

$$\text{cantrentrungbinh} = \text{Gdpf} + @qtdist(1 - \frac{\alpha}{2}, n - k) * \text{Se2}$$

$$\text{canduoidacbiet} = \text{Gdpf} - @qtdist(1 - \frac{\alpha}{2}, n - k) * \text{Se1}$$

$$\text{cantrendacbiet} = \text{Gdpf} + @qtdist(1 - \frac{\alpha}{2}, n - k) * \text{Se1}$$

- Chọn 4 đối tượng sau đó nhấn Enter.

Bảng 1: Bảng phân vị Student $t_\alpha(n)$ bậc tự do $n - 1$, mức xác suất α

$$P(T > t_\alpha(n - 1)) = \alpha \text{ với } T \sim St(n).$$

$n - 1; \alpha$	0.10	0.05	0.025	0.01	0.005	0.001
1	3.078	6.314	12.706	31.821	63.675	66.619
2	1.886	2.920	4.303	6.965	9.925	22.326
3	1.638	2.353	3.182	4.541	5.841	10.213
4	1.533	2.132	2.776	3.747	4.604	7.173
5	1.476	2.015	2.571	3.365	4.032	5.893
6	1.440	1.943	2.447	3.143	3.707	5.208
7	1.415	1.895	2.365	2.998	3.499	4.785
8	1.397	1.860	2.306	2.896	3.355	4.501
9	1.383	1.833	2.262	2.821	3.250	4.297
10	1.372	1.812	2.228	2.764	3.169	4.144
11	1.363	1.796	2.201	2.718	3.106	4.025
12	1.356	1.782	2.179	2.861	3.055	3.930
13	1.350	1.771	2.160	2.650	3.012	3.852
14	1.345	1.761	2.145	2.624	2.977	3.787
15	1.341	1.753	2.131	2.602	2.947	3.733
16	1.337	1.746	2.120	2.583	2.921	3.686
17	1.333	1.740	2.110	2.567	2.898	3.646
18	1.330	1.734	2.101	2.552	2.878	3.610
19	1.328	1.719	2.093	2.539	2.861	3.579

Bảng 1 (tt): Bảng phân vị Student $t_\alpha(n)$ bậc tự do $n - 1$, mức xác suất α

$$P(T > t_\alpha(n - 1)) = \alpha \text{ với } T \sim St(n).$$

$n - 1; \alpha$	0.10	0.05	0.025	0.01	0.005	0.001
20	1.325	1.725	2.086	2.528	2.845	3.552
21	1.323	1.721	2.080	2.518	2.831	3.527
22	1.321	1.717	2.074	2.508	2.819	3.505
23	1.319	1.714	2.069	2.500	2.807	3.485
24	1.318	1.711	2.064	2.492	2.797	3.467
25	1.316	1.708	2.060	2.485	2.787	3.450
26	1.315	1.706	2.056	2.479	2.779	3.435
27	1.314	1.703	2.052	2.473	2.771	3.421
28	1.313	1.701	2.048	2.467	2.763	3.408
29	1.311	1.699	2.045	2.462	2.756	3.396
30	1.310	1.697	2.042	2.457	2.750	3.385
40	1.303	1.684	2.021	2.423	2.704	3.307
60	1.296	1.671	2.000	2.390	2.660	3.232
120	1.289	1.658	1.980	2.358	2.617	3.160
$+\infty$	1.282	1.645	1.960	2.326	2.576	3.090

Bảng 2: Bảng phân vị Khi bình phương bậc tự do n mức xác suất α

$n; \alpha$	0,995	0,99	0,975	0,95	0,05	0,025	0,01	0,005
1	0,000	0,000	0,001	0,004	3,841	5,024	6,635	7,879
2	0,010	0,020	0,051	0,103	5,911	7,378	9,210	10,597
3	0,072	0,115	0,216	0,352	7,815	9,348	11,345	12,838
4	0,207	0,297	0,484	0,711	9,488	11,143	13,277	14,860
5	0,412	0,554	0,831	1,145	10,070	12,832	15,086	16,750
6	0,676	0,872	1,237	1,635	12,592	14,449	16,812	18,548
7	0,989	1,239	1,690	2,167	14,067	16,013	18,475	20,278
8	1,314	1,646	2,180	2,733	15,507	17,535	20,090	21,995
9	1,735	2,088	2,700	3,322	16,919	19,023	21,666	23,589
10	2,156	2,558	3,247	3,940	18,307	20,483	23,209	25,188
11	2,603	3,053	3,816	4,575	19,675	21,920	24,725	26,757
12	3,074	3,571	4,404	5,226	21,026	23,337	26,217	28,300
13	3,565	4,107	5,009	5,982	22,362	24,736	27,688	29,819
14	4,075	4,660	5,629	6,571	23,685	26,119	29,141	31,319
15	4,601	5,229	5,262	7,261	24,996	27,488	30,758	32,801
16	5,142	5,812	6,908	7,962	26,296	28,845	32,000	34,267
17	5,697	6,408	7,564	8,672	27,587	30,191	33,409	35,718
18	6,265	7,015	8,231	9,390	28,869	31,526	34,805	37,156
19	6,844	7,633	8,907	10,117	30,144	32,852	36,191	38,582

Bảng 2(tt): Bảng phân vị Khi bình phương bậc tự do n mức xác suất α

$n; \alpha$	0,995	0,99	0,975	0,95	0,05	0,025	0,01	0,005
20	7,343	8,260	9,591	10,851	31,410	34,170	37,566	39,997
21	8,034	8,897	10,283	11,591	32,671	35,479	38,932	41,401
22	8,543	9,542	10,982	12,388	33,924	36,781	40,289	42,796
23	9,260	10,196	11,689	13,091	35,172	38,076	41,638	44,181
24	9,886	10,856	12,401	13,848	36,415	39,364	42,980	45,558
25	10,520	11,524	13,120	14,611	37,652	40,646	44,314	46,928
26	11,160	12,198	13,844	15,379	38,885	41,923	45,642	48,290
27	11,808	12,879	14,573	16,151	40,113	43,194	46,963	49,645
28	12,461	13,565	15,308	16,928	41,337	44,461	48,278	50,993
29	13,121	14,256	16,047	17,708	42,557	45,722	49,588	52,336
30	13,787	14,930	16,791	18,493	43,773	46,979	50,892	63,672
40	20,707	22,164	24,433	26,509	55,578	59,342	63,691	66,766
50	27,991	29,707	32,307	24,754	67,505	71,420	76,154	79,490
100	67,328	70,065	74,222	77,929	124,34	129,56	135,80	140,16

Bảng 3: Bảng giá trị $F_{\alpha}(n; m)$ của phân phối Fisher
 $P(F > F_{\alpha}(n; m)) = \alpha$ với $\alpha = 0.01$

Df mẫu (m)	Bậc tự do (df) của tử số (n)									
	1	2	3	4	5	6	7	8	9	10
1	4052	4999	5404	5624	5764	5859	5928	5981	6022	6056
2	98,50	99,00	99,16	99,25	99,30	99,33	99,36	99,38	99,39	99,40
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,34	27,23
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,55
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16	10,05
6	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87
7	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69
17	8,40	6,11	5,19	4,67	4,34	4,10	3,93	3,79	3,68	3,59
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40	3,31
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26

Bảng 3(tt): Bảng giá trị $F_\alpha(n; m)$ của phân phối Fisher
 $P(F > F_\alpha(n; m)) = \alpha$ với $\alpha = 0.01$

Df mẫu (m)	Bậc tự do (df) của tử số (n)									
	1	2	3	4	5	6	7	8	9	10
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26	3,17
25	7,77	5,57	4,68	4,18	3,85	3,63	3,46	3,32	3,22	3,13
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18	3,09
27	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26	3,15	3,06
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12	3,03
29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	3,09	3,00
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,80
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63
120	6,85	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,56	2,47
∞	6,63	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,41	2,32

Bảng 3(tt): Bảng giá trị $F_\alpha(n; m)$ của phân phối Fisher
 $P(F > F_\alpha(n; m)) = \alpha$ với $\alpha = 0.01$

Df mẫu (m)	Bậc tự do (df) của tử số (n)								
	11	12	15	20	24	30	40	60	120
1	6083	6107	6157	6209	6234	6260	6286	6313	6340
2	99,41	99,42	99,43	99,45	99,46	99,47	99,48	99,48	99,49
3	27,13	27,05	26,87	26,69	26,60	26,50	26,41	26,32	26,22
4	14,45	14,37	14,20	14,02	13,93	13,84	13,75	13,65	13,56
5	9,96	9,89	9,72	9,55	9,47	9,38	9,29	9,20	9,11
6	7,79	7,72	7,56	7,40	7,31	7,23	7,14	7,06	6,97
7	6,54	6,47	6,31	6,16	6,07	5,99	5,91	5,82	5,74
8	5,73	5,67	5,52	5,36	5,28	5,20	5,12	5,03	4,95
9	5,18	5,11	4,96	4,81	4,73	4,65	4,57	4,48	4,40
10	4,77	4,71	4,56	4,41	4,33	4,25	4,17	4,08	4,00
11	4,46	4,40	4,25	4,10	4,02	3,94	3,86	3,78	3,69
12	4,22	4,16	4,01	3,86	3,78	3,70	3,62	3,54	3,45
13	4,02	3,96	3,82	3,66	3,59	3,51	3,43	3,34	3,25
14	3,86	3,80	3,66	3,51	3,43	3,35	3,27	3,18	3,09
15	3,73	3,67	3,52	3,37	3,29	3,21	3,13	3,05	2,96
16	3,62	3,55	3,41	3,26	3,18	3,10	3,02	2,93	2,84
17	3,52	3,46	3,31	3,16	3,08	3,00	2,92	2,83	2,75
18	3,43	3,37	3,23	3,08	3,00	2,92	2,84	2,75	2,66
19	3,36	3,30	3,15	3,00	2,92	2,84	2,76	2,67	2,58
20	3,29	3,23	3,09	2,94	2,86	2,78	2,69	2,61	2,52
21	3,24	3,17	3,03	2,88	2,80	2,72	2,64	2,55	2,46
22	3,18	3,12	2,98	2,83	2,75	2,67	2,58	2,50	2,40

Bảng 3(tt): Bảng giá trị $F_\alpha(n; m)$ của phân phối Fisher
 $P(F > F_\alpha(n; m)) = \alpha$ với $\alpha = 0.01$

Df mẫu (m)	Bậc tự do (df) của tử số (n)								
	11	12	15	20	24	30	40	60	120
23	3,14	3,07	2,93	2,78	2,70	2,62	2,54	2,45	2,35
24	3,09	3,03	2,89	2,74	2,66	2,58	2,49	2,40	2,31
25	3,06	2,99	2,85	2,70	2,62	2,54	2,45	2,36	2,27
26	3,02	2,96	2,81	2,66	2,58	2,50	2,42	2,33	2,23
27	2,99	2,93	2,78	2,63	2,55	2,47	2,38	2,29	2,20
28	2,96	2,90	2,75	2,60	2,52	2,44	2,35	2,26	2,17
29	2,93	2,87	2,73	2,57	2,49	2,41	2,33	2,23	2,14
30	2,91	2,84	2,70	2,55	2,47	2,39	2,30	2,21	2,11
40	2,73	2,66	2,52	2,37	2,29	2,20	2,11	2,02	1,92
60	2,56	2,50	2,35	2,20	2,12	2,03	1,94	1,84	1,73
120	2,40	2,34	2,19	2,03	1,95	1,86	1,76	1,66	1,53
∞	2,25	2,18	2,04	1,88	1,79	1,70	1,59	1,47	1,32

Bảng 3(tt): Bảng giá trị $F_{\alpha}(n; m)$ của phân phối Fisher
 $P(F > F_{\alpha}(n; m)) = \alpha$ với $\alpha = 0.05$

Df mẫu (m)	Bậc tự do (df) của tử số (n)									
	1	2	3	4	5	6	7	8	9	10
1	161,5	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5	241,9
2	18,51	19,49	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30

Bảng 3(tt): Bảng giá trị $F_\alpha(n; m)$ của phân phối Fisher
 $P(F > F_\alpha(n; m)) = \alpha$ với $\alpha = 0.05$

Df mẫu (m)	Bậc tự do (df) của tử số (n)									
	1	2	3	4	5	6	7	8	9	10
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99
120	3,92	3,07	2,68	2,45	2,29	2,18	2,09	2,02	1,96	1,91
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83

Bảng 3(tt): Bảng giá trị $F_\alpha(n; m)$ của phân phối Fisher
 $P(F > F_\alpha(n; m)) = \alpha$ với $\alpha = 0.05$

Df mẫu (m)	Bậc tự do (df) của tử số (n)								
	11	12	15	20	24	30	40	60	120
1	243	244	246	248	249	250	251	252	253
2	19,4	19,4	19,4	19,5	19,5	19,5	19,5	19,5	19,5
3	8,76	8,74	8,70	8,66	8,64	8,62	8,59	8,57	8,55
4	5,94	5,91	5,86	5,80	5,77	5,75	5,72	5,69	5,66
5	4,70	4,68	4,62	4,56	4,53	4,50	4,46	4,43	4,40
6	4,03	4,00	3,94	3,87	3,84	3,81	3,77	3,74	3,70
7	3,60	3,57	3,51	3,44	3,41	3,38	3,34	3,30	3,27
8	3,31	3,28	3,22	3,15	3,12	3,08	3,04	3,01	2,97
9	3,10	3,07	3,01	2,94	2,90	2,86	2,83	2,79	2,75
10	2,94	2,91	2,85	2,77	2,74	2,70	2,66	2,62	2,58
11	2,82	2,79	2,72	2,65	2,61	2,57	2,53	2,49	2,45
12	2,72	2,69	2,62	2,54	2,51	2,47	2,43	2,38	2,34
13	2,63	2,60	2,53	2,46	2,42	2,38	2,34	2,30	2,25
14	2,57	2,53	2,46	2,39	2,35	2,31	2,27	2,22	2,18
15	2,51	2,48	2,40	2,33	2,29	2,25	2,20	2,16	2,11
16	2,46	2,42	2,35	2,28	2,24	2,19	2,15	2,11	2,06
17	2,41	2,38	2,31	2,23	2,19	2,15	2,10	2,06	2,01
18	2,37	2,34	2,27	2,19	2,15	2,11	2,06	2,02	1,97
19	2,34	2,31	2,23	2,16	2,11	2,07	2,03	1,98	1,93
20	2,31	2,28	2,20	2,12	2,08	2,04	1,99	1,95	1,90
21	2,28	2,25	2,18	2,10	2,05	2,01	1,96	1,92	1,87
22	2,26	2,23	2,15	2,07	2,03	1,98	1,94	1,89	1,84

Bảng 3(tt): Bảng giá trị $F_\alpha(n; m)$ của phân phối Fisher
 $P(F > F_\alpha(n; m)) = \alpha$ với $\alpha = 0.05$

Df mẫu (m)	Bậc tự do (df) của tử số (n)								
	11	12	15	20	24	30	40	60	120
23	2,24	2,20	2,13	2,05	2,01	1,96	1,91	1,86	1,81
24	2,22	2,18	2,11	2,03	1,98	1,94	1,89	1,84	1,79
25	2,20	2,16	2,09	2,01	1,96	1,92	1,87	1,82	1,77
26	2,18	2,15	2,07	1,99	1,95	1,90	1,85	1,80	1,75
27	2,17	2,13	2,06	1,97	1,93	1,88	1,84	1,79	1,73
28	2,15	2,12	2,04	1,96	1,91	1,87	1,82	1,77	1,71
29	2,14	2,10	2,03	1,94	1,90	1,85	1,81	1,75	1,70
30	2,13	2,09	2,01	1,93	1,89	1,84	1,79	1,74	1,68
40	2,04	2,00	1,92	1,84	1,79	1,74	1,69	1,64	1,58
60	1,95	1,92	1,84	1,75	1,70	1,65	1,59	1,53	1,47
120	1,87	1,83	1,75	1,66	1,61	1,55	1,50	1,43	1,35
∞	1,79	1,75	1,67	1,57	1,52	1,46	1,39	1,32	1,22

Tài liệu tham khảo

- [1] Nguyễn Thành Cả, Nguyễn Thị Ngọc Miên (2014), *Kinh tế lượng*, Nhà xuất bản Kinh tế Thành phố Hồ Chí Minh.
- [2] Nguyễn Quang Dong, Nguyễn Thị Minh (2013), *Giáo trình Kinh tế lượng*, Nhà xuất bản Đại học Kinh tế Quốc dân.
- [3] Nguyễn Quang Dong (2003), *Bài giảng Kinh tế lượng*, Nhà xuất bản Thống kê.
- [4] Huỳnh Đạt Hùng, Nguyễn Khánh Bình, Phạm Xuân Giang (2013), *Kinh tế lượng*, Nhà xuất bản Phương Đông.
- [5] Hoàng Ngọc Nhậm và các tác giả (2008), *Giáo trình Kinh tế lượng*, Nhà xuất bản Lao động Xã hội.
- [6] Nguyễn Thị Ngọc Thanh và các tác giả (2015), *Bài tập Kinh tế lượng với sự trợ giúp của Eviews*, Trường Đại học Kinh tế Thành phố Hồ Chí Minh.
- [7] Nguyễn Văn Tùng (2014), *Thực hành Kinh tế lượng cơ bản với Eviews*, Nhà xuất bản Kinh tế thành phố Hồ Chí Minh.