

**Đại Học Bách Khoa Hà Nội**  
**Khoa Toán-Tin**



## **SUY LUẬN THỐNG KÊ**

### **BÀI TẬP THỰC HÀNH III**

**Giảng viên hướng dẫn:** PGS.TS.Nguyễn Thị Thu Thủy  
**Lớp:** 155362  
**Nhóm:** 12  
**Sinh viên:** Mai Thành Duy-20227225  
Nguyễn Minh Anh-20227212  
Nguyễn Phương Nhi-20227251

# MỤC LỤC

Lời nói đầu . . . . .	2
1 Giới thiệu bộ dữ liệu . . . . .	3
1.1 Đặt vấn đề . . . . .	3
1.2 Bộ dữ liệu . . . . .	4
2 Xây dựng mô hình hồi quy tuyến tính bội . . . . .	5
2.1 Mô hình hồi quy dự đoán . . . . .	5
2.2 Ước lượng các hệ số hồi quy . . . . .	5
2.3 Nhận xét . . . . .	8
3 Kiểm tra ý nghĩa mô hình hồi quy tuyến tính bội . . . . .	9
3.1 Phân tích phương sai ANNOVA . . . . .	9
3.2 Kiểm tra hệ số xác định $R^2$ . . . . .	13
4 Kiểm tra ý nghĩa hệ số hồi quy riêng . . . . .	14
4.1 Ý nghĩa $\beta_1$ . . . . .	14
4.2 Ý nghĩa $\beta_2$ . . . . .	15
4.3 Ý nghĩa $\beta_3$ . . . . .	17
4.4 Ý nghĩa $\beta_4$ . . . . .	18
5 Kết luận mô hình hồi quy bội . . . . .	22

## Lời nói đầu

Hồi quy tuyến tính bội là một trong những phương pháp phân tích thống kê phổ biến và quan trọng trong việc nghiên cứu mối quan hệ giữa một biến phụ thuộc và nhiều biến độc lập. Với khả năng mô tả và giải thích sự biến thiên của biến phụ thuộc dựa trên các biến độc lập, hồi quy tuyến tính bội không chỉ được sử dụng rộng rãi trong các lĩnh vực như kinh tế, tài chính, y học, khoa học xã hội mà còn trở thành công cụ mạnh mẽ trong các ngành công nghệ hiện đại như học máy và phân tích dữ liệu lớn.

Phương pháp hồi quy tuyến tính bội không chỉ giúp chúng ta hiểu rõ hơn về mối quan hệ giữa các yếu tố trong dữ liệu mà còn cung cấp cơ sở để đưa ra các quyết định, dự đoán, và giải pháp dựa trên thông tin thu thập được.

Báo cáo của nhóm chúng em bao gồm các nội dung chính sau:

- Xây dựng mô hình hồi quy tuyến tính bội
- Kiểm tra ý nghĩa mô hình hồi quy tuyến tính bội
- Kiểm tra ý nghĩa hệ số hồi quy riêng

Chúng em xin gửi lời cảm ơn sâu sắc đến cô Nguyễn Thị Thu Thủy vì sự tận tình giảng dạy và hướng dẫn trong học phần này. Trong quá trình thực hiện báo cáo, dù đã nỗ lực tìm hiểu và hoàn thiện, nhóm không tránh khỏi thiếu sót. Chúng em rất mong nhận được ý kiến đóng góp từ cô để hoàn thiện báo cáo tốt hơn.

Chúng em xin chân thành cảm ơn!

**NHÓM 12**

# 1 Giới thiệu bộ dữ liệu

## 1.1 Đặt vấn đề

Bảng dữ liệu cung cấp thông tin liên quan đến các phiên giao dịch cổ phiếu của Netflix (mã cổ phiếu: NFLX) trong khoảng thời gian từ ngày 5-2-2018 đến 5-2-2022. Đây là tập dữ liệu quan trọng để phân tích biến động giá cổ phiếu và xây dựng các mô hình dự đoán.

Các biến trong dữ liệu:

- **X1: Open (Giá mở cửa):USD**

Giá cổ phiếu tại thời điểm bắt đầu phiên giao dịch.

- **X2: High (Giá cao nhất):USD**

Mức giá cao nhất của cổ phiếu đạt được trong suốt phiên giao dịch.

- **X3: Low (Giá thấp nhất):USD**

Mức giá thấp nhất của cổ phiếu đạt được trong suốt phiên giao dịch.

- **X4: Volume (Khối lượng giao dịch):USD**

Số lượng cổ phiếu được mua và bán trong phiên giao dịch.

- **Y: Adjusted Close (Giá đóng cửa điều chỉnh):USD**

Giá cổ phiếu tại thời điểm kết thúc phiên giao dịch, đã được điều chỉnh để tính đến các yếu tố như chia cổ tức, chia tách cổ phiếu.

**Ý nghĩa và mục tiêu phân tích:**

- **Biến phụ thuộc (Y):**

Adjusted Close được chọn làm biến phụ thuộc vì đây là giá trị phản ánh chính xác giá trị thực tế của cổ phiếu sau khi loại bỏ các yếu tố gây nhiễu như chia cổ tức hoặc chia tách cổ phiếu.

- **Biến độc lập (X):**

Các biến Open, High, Low và Volume được chọn làm biến độc lập vì chúng đại diện cho các yếu tố chính trong giao dịch hàng ngày có khả năng ảnh hưởng trực tiếp đến giá đóng cửa của cổ phiếu.

## 1.2 Bộ dữ liệu

Bộ dữ liệu mẫu với kích cỡ  $n = 1009$  quan sát

X1	X2	X3	X4	Y
262	267,9	250	11896100	254,6
247,7	266,7	245	12595800	265,7
266,6	272,5	264,3	8981500	264,6
...	...	...	...	...
407,3	412,8	396,6	7782400	410,2

**Table 1.1** Bảng dữ liệu mẫu

## 2 Xây dựng mô hình hồi quy tuyến tính bội

### 2.1 Mô hình hồi quy dự đoán

Mô hình hồi quy lý thuyết:

$$\mathbf{Y} = \beta_0 + \sum_{j=1}^k \beta_j x_j + \epsilon \quad (2.1)$$

Mô hình hồi quy thực nghiệm

$$\hat{\mathbf{Y}} = \hat{\beta}_0 + \sum_{j=1}^k \hat{\beta}_j x_j \quad (2.2)$$

Với bộ dữ liệu mẫu , ta có hệ  $n = 1009$  phương trình ,  $k+1 = 5$  ẩn:

$$\mathbf{Y} = \beta \mathbf{X} + \epsilon \quad (2.3)$$

trong đó :

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{(n \times 1)} ; X = \begin{bmatrix} 1 & x_{11} & x_{12} & x_{13} & x_{14} \\ 1 & x_{21} & x_{22} & x_{23} & x_{24} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & x_{n3} & x_{n4} \end{bmatrix}_{(n \times 5)} ; \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}_{(5 \times 1)} ; \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{bmatrix}_{(5 \times 1)} .$$

### 2.2 Ước lượng các hệ số hồi quy

Phương pháp bình phương cực tiểu

$$\mathcal{L} = \sum_{i=1}^n \varepsilon_i^2 = \varepsilon^T \varepsilon = (Y - X\hat{\beta})^T (Y - X\hat{\beta}) = Y^T Y - 2\hat{\beta}^T X^T Y + \hat{\beta}^T X^T X \hat{\beta} \quad (2.4)$$

Tìm cực tiểu của hàm  $\mathcal{L}$  bằng đạo hàm riêng tại  $\hat{\beta}$

$$\frac{\partial \mathcal{L}}{\partial \hat{\beta}} = 0 \quad (2.5)$$

$$\Leftrightarrow \hat{\beta} = (X^T X)^{-1} X^T Y \quad (2.6)$$

**Đoạn chương trình tính toán:**

```

1 import pandas as pd
2 import numpy as np
3
4 # Doc du lieu tu file Excel, voi ten cot o hang dau tien
5 df = pd.read_excel("NFLX.xlsx", sheet_name="NFLX")
6
7 # In ra cac ten cot de kiem tra
8 print("Ten cac cot trong du lieu:")
9 print(df.columns)
10
11 # X la ma tran du lieu, Y la vector ket qua
12 # X1 = Open, X2 = High, X3 = Low, X4 = Volume
13 # Y = Adj Close
14 X = df[["X1= Open(Gia mo cua)", "X2= High(Gia cao nhat)", "X3=
    Low(Gia thap nhat)", "X4= Volume(Khoi luong giao dich)"]].
    values # Chon cac cot Open, High, Low, Volume
15 X = np.hstack((np.ones((X.shape[0], 1)), X)) # Them cot 1 vao
    X de tinh he so chech

```

```

1 Y = df["Y= Adj Close( Gia dong phien dieu chinh)"].values
2
3 # Tinh Beta mu theo cong thuc: Beta_hat = (X^T X)^-1 X^T Y
4 XT = X.T
5 Beta_hat = np.linalg.inv(XT @ X) @ XT @ Y
6
7 # In ra ket qua
8 print("Beta mu (uoc luong):")
9 print(Beta_hat)
10
11 # Hien thi mo hinh hoi quy
12 feature_names = ["1"] + ["X1", "X2", "X3", "X4"]
13 equation_terms = [f"{Beta_hat[i]:.11f} * {feature_names[i]}"
14                   for i in range(len(Beta_hat))]
15 model_equation = "y = " + " + ".join(equation_terms)
16
17 print("Mo hinh hoi quy (y_hat):")
18 print(model_equation)

```

Kết quả :

$$\hat{\beta} = \begin{bmatrix} -0,455323324 \\ -0,597693885 \\ 0,772316330 \\ 0,826377970 \\ 0,880383531 \times 10^{-7} \end{bmatrix}$$

$$\hat{y} = -0,4553 - 0,5977x_1 + 0,7723x_2 + 0,8264x_3 + (0,8804 \times 10^{-7}) x_4$$



## 2.3 Nhận xét

Dựa trên các hệ số ước lượng ( $\hat{\beta}_i$ ) trong mô hình hồi quy bội, ta có các nhận xét như sau:

- $\beta_0 = -0,4553$ : Đây là giá trị dự đoán của Giá đóng phiên điều chỉnh Y khi tất cả các biến độc lập ( $x_1, x_2, x_3, x_4$ ) đều bằng 0. Ý nghĩa của hệ số này phụ thuộc vào bối cảnh, nhưng nó thường là điểm khởi đầu cho dự đoán.
- $\beta_1 = -0,5977$ : Nếu Giá mở cửa  $x_1$  tăng lên 1 đơn vị (và các biến khác giữ nguyên), thì Giá đóng phiên điều chỉnh Y trung bình sẽ **giảm 0,5977 đơn vị**.
- $\beta_2 = 0,7723$ : Nếu Giá cao nhất  $x_2$  tăng lên 1 đơn vị, thì  $y$  trung bình sẽ **tăng 0,7723 đơn vị**, với giả thiết các biến khác không đổi.
- $\beta_3 = 0,8264$ : Nếu Giá thấp nhất  $x_3$  tăng lên 1 đơn vị, thì  $y$  trung bình sẽ **tăng 0,8264 đơn vị**, các biến khác giữ nguyên.
- $\beta_4 = 0,8804 \times 10^{-7}$ : Nếu Khối lượng giao dịch  $x_4$  tăng lên 1 đơn vị, thì  $y$  trung bình sẽ **tăng  $0,8804 \times 10^{-7}$  đơn vị**, các biến khác giữ nguyên.

### 3 Kiểm tra ý nghĩa mô hình hồi quy tuyến tính bội

#### 3.1 Phân tích phương sai ANNOVA

Bước 1: Cặp giả thuyết

$$\begin{cases} H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0 \\ H_1 : \beta_j \neq 0; j = 1, 2, 3, 4 \end{cases}$$

Bước 2: Chọn Tiêu chuẩn kiểm định

$$F_0 = \frac{MS_R}{MS_E} \sim f_{(k, n-(k+1))} \quad (3.1)$$

Với  $k = 4$ ;  $n = 1009$

Bước 3: Tính toán từ bộ dữ liệu giá trị quan sát

- Tổng bình phương sai số hồi quy :

$$MS_R = \frac{SS_R}{k} = 2.951.350,5489 \quad (3.2)$$

Trong đó:

$$SS_R = \hat{\beta}^T X^T y - \frac{(\sum_{i=1}^n y_i)^2}{n} = 11.805.402,1958$$

- Tổng bình phương sai số đo lường sự thay đổi của mô hình từ những ảnh khác ngoài các ảnh  $x_1, x_2, x_3, x_4$ :

$$MS_E = \frac{SS_E}{n - (k + 1)} = 15,0751 \quad (3.3)$$

Trong đó:

$$SS_E = y^T y - \hat{\beta}^T X^T y = 15.135,4286$$

- Giá trị quan sát:

$$f_0 = \frac{MS_R}{MS_E} = 195.776,1513 \quad (3.4)$$

#### Bước 4: Miền bác bỏ

$$W_\alpha = \{f_0 \mid f_0 > f_\alpha(k, n - (k + 1))\}$$

Trong khuôn khổ bài thực hành này, mức ý nghĩa là  $\alpha = 0,05$

$$W_\alpha = \{f_0 \mid f_0 > f_{0,05}(4, 1004)\} = (2,3808; +\infty)$$

#### Bước 5: Kết luận & Vẽ bảng ANNOVA

$$\Rightarrow f_0 \in W_\alpha$$

Chúng ta hoàn toàn có thể bác bỏ giả thuyết  $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ , tức là mô hình này có ý nghĩa hồi quy

#### Bảng phân tích phương sai ANNOVA

Nguồn	Tổng bình phương	Bậc tự do	Bình phương trung bình	$f_0$
Hồi quy	$SS_R = 11.805.402,1958$	4	$MS_R = 2.951.350,5489$	
Sai số	$SS_E = 15.135,4286$	1004	$MS_E = 15,0751$	
Tổng	$SS_T = 11.820.537,6244$	1008		$\frac{MS_R}{MS_E} = 195.776,1513$

### Đoạn mã tính toán

```

1 from scipy.stats import f
2 # So luong mau (n) va so bien doc lap (k)
3 n = X.shape[0]
4 k = X.shape[1] - 1 # Tru di cot he so chan
5 # Tinh tong binh phuong sai so hoi quy (SS_R)
6 SS_R = Beta_hat.T @ XT @ Y - (np.sum(Y) ** 2) / n
7 # Tinh tong binh phuong sai so do luong su thay doi ngoai mo
   hinh (SS_E)
8 SS_E = Y.T @ Y - Beta_hat.T @ XT @ Y
9 SS_T = SS_R + SS_E
10 # Tinh MS_R va MS_E
11 MS_R = SS_R / k
12 MS_E = SS_E / (n - (k + 1))
13 # Tinh gia tri quan sat f_0
14 f_0 = MS_R / MS_E

```

```
1 # Tim gia tri toi han f_{alpha}(4, 1004)
2 alpha = 0.05
3 df1 = k # Bac tu do cua MS_R
4 df2 = n - (k + 1) # Bac tu do cua MS_E
5 f_critical = f.ppf( 1 - alpha, df1, df2)
6 # Kiem tra f_0 co thuoc W_alpha hay khong
7 if f_0 > f_critical:
8     result = "f_0 thuoc W_alpha"
9 else:
10     result = "f_0 khong thuoc W_alpha"
11 # In ket qua
12 print("\nKet qua tinh toan:")
13 print(f"Tong binh phuong sai so hoi quy (SS_R): {SS_R}")
14 print(f"Tong binh phuong sai so do luong su thay doi ngoai mo
15     hinh (SS_E): {SS_E}")
16 print(f"Tong binh phuong sai so tong the mo hinh(SS_T): {SS_T}"
17     )
18 print(f"MS_R (trung binh binh phuong hoi quy): {MS_R}")
19 print(f"MS_E (trung binh binh phuong sai so): {MS_E}")
20 print(f"Gia tri quan sat f_0: {f_0}")
21 print("\nKet qua kiem tra Fisher:")
22 print(f"Gia tri toi han f_{alpha}(4, 1004): {f_critical}")
23 print(f"Ket luan: {result}")
```

### 3.2 Kiểm tra hệ số xác định $R^2$

- Hệ số xác định:

$$R^2 = \frac{SS_R}{SS_T} = 0.9987$$

- Giá trị  $R^2 = 0.9987$  nghĩa là 99.87% phương sai của Y được giải thích bởi mô hình hồi quy, chỉ có 0.13% phương sai không được giải thích (phần nhiều hoặc do các yếu tố ngoài mô hình).
- Mô hình có vẻ rất phù hợp với dữ liệu.

#### Đoạn mã tính toán

```
1 # Tính R^2
2 R_squared = SS_R / SS_T
3 print(f"R^2 (he so xac dinh): {R_squared}")
```

## 4 Kiểm tra ý nghĩa hệ số hồi quy riêng

### 4.1 Ý nghĩa $\beta_1$

Bước 1: Cặp giả thuyết

$$\begin{cases} H_0 : \beta_1 = \beta_{1,0} = 0 \\ H_1 : \beta_1 \neq \beta_{1,0} = 0 \end{cases}$$

Bước 2: Chọn Tiêu chuẩn kiểm định

$$T_1 = \frac{\hat{\beta}_1 - \beta_{1,0}}{se(\hat{\beta}_1)} \sim t_{(n-(k+1))} \quad (4.1)$$

Bước 3: Tính toán từ bộ dữ liệu giá trị quan sát

- Sai số chuẩn của ước lượng :

$$se(\hat{\beta}_1) = \sqrt{\sigma^2 C_{22}} \quad (4.2)$$

Trong đó:

$\sigma^2$  chưa biết nên cần ước lượng bởi:  $s^2 = \frac{SS_E}{n-(k+1)} = 15,0751$

$C_{22}$  là phần tử thứ 2 trên đường chéo chính của ma trận  $(X^T X)^{-1}$  ứng với  $\hat{\beta}_1$

- Giá trị quan sát :

$$t_1 = \frac{-0,5977 - 0}{\sqrt{15,0751 \times 4,2439 \times 10^{-5}}} = -23,6299 \quad (4.3)$$

**Bước 4: Miền bác bỏ**

$$W_\alpha = \{t_1 \mid t_1 < t_{\frac{-\alpha}{2}; n-(k+1)} \text{ hoặc } t_1 > t_{\frac{\alpha}{2}; n-(k+1)}\}$$

Trong khuôn khổ bài thực hành này, mức ý nghĩa là  $\alpha = 0,05$

$$W_\alpha = \{t_1 \mid t_1 > t_{0,025; 1004} \text{ hoặc } t_1 < -t_{0,025; 1004}\} = (-\infty, -1,9623) \cup (1,9623, \infty)$$

**Bước 5: Kết luận**

$$\Rightarrow t_1 \in W_\alpha$$

Chúng ta hoàn toàn có thể bác bỏ giả thuyết  $H_0 : \beta_1 = 0$ , tức là biến  $x_1$  có ý nghĩa trong mô hình hồi quy.

**4.2 Ý nghĩa  $\beta_2$** **Bước 1: Cặp giả thuyết**

$$\begin{cases} H_0 : \beta_2 = \beta_{2,0} = 0 \\ H_1 : \beta_2 \neq \beta_{2,0} = 0 \end{cases}$$

**Bước 2: Chọn Tiêu chuẩn kiểm định**

$$T_2 = \frac{\hat{\beta}_2 - \beta_{2,0}}{se(\hat{\beta}_2)} \sim t_{(n-(k+1))} \quad (4.4)$$



**Bước 3:Tính toán từ bộ dữ liệu giá trị quan sát**

- Sai số chuẩn của ước lượng :

$$se(\hat{\beta}_2) = \sqrt{\sigma^2 C_{33}} \quad (4.5)$$

Trong đó:

$\sigma^2$  chưa biết nên cần ước lượng bởi:  $s^2 = \frac{SS_E}{n-(k+1)} = 15,0751$

$C_{33}$  là phần tử thứ 3 trên đường chéo chính của ma trận  $(X^T X)^{-1}$  ứng với  $\hat{\beta}_2$

- Giá trị quan sát :

$$t_2 = \frac{0,7723 - 0}{\sqrt{15,0751 \times 4,6059 \times 10^{-5}}} = 29,3094 \quad (4.6)$$

**Bước 4:Miền bác bỏ**

$$W_\alpha = \{t_2 \mid t_2 < t_{\frac{-\alpha}{2}; n-(k+1)} \text{ hoặc } t_2 > t_{\frac{\alpha}{2}; n-(k+1)}\}$$

Trong khuôn khổ bài thực hành này, mức ý nghĩa là  $\alpha = 0,05$

$$W_\alpha = \{t_2 \mid t_2 > t_{0,025;1004} \text{ hoặc } t_2 < -t_{0,025;1004}\} = (-\infty, -1,9623) \cup (1,9623, \infty)$$

**Bước 5:Kết luận**

$$\Rightarrow t_2 \in W_\alpha$$

Chúng ta hoàn toàn có thể bác bỏ giả thuyết  $H_0 : \beta_2 = 0$  , tức là biến  $x_2$  có ý nghĩa trong mô hình hồi quy.

### 4.3 Ý nghĩa $\beta_3$

**Bước 1:**Cặp giả thuyết

$$\begin{cases} H_0 : \beta_3 = \beta_{3,0} = 0 \\ H_1 : \beta_3 \neq \beta_{3,0} = 0 \end{cases}$$

**Bước 2:**Chọn Tiêu chuẩn kiểm định

$$T_3 = \frac{\hat{\beta}_3 - \beta_{3,0}}{se(\hat{\beta}_3)} \sim t_{(n-(k+1))} \quad (4.7)$$

**Bước 3:**Tính toán từ bộ dữ liệu giá trị quan sát

- Sai số chuẩn của ước lượng :

$$se(\hat{\beta}_2) = \sqrt{\sigma^2 C_{44}} \quad (4.8)$$

Trong đó:

$\sigma^2$  chưa biết nên cần ước lượng bởi:  $s^2 = \frac{SS_E}{n-(k+1)} = 15,0751$

$C_{44}$  là phần tử thứ 3 trên đường chéo chính của ma trận  $(X^T X)^{-1}$  ứng với  $\hat{\beta}_3$

- Giá trị quan sát :

$$t_3 = \frac{0,8264 - 0}{\sqrt{15,0751 \times 4,5649 \times 10^{-5}}} = 31,5018 \quad (4.9)$$

**Bước 4: Miền bác bỏ**

$$W_\alpha = \{t_3 \mid t_3 < t_{\frac{-\alpha}{2}; n-(k+1)} \text{ hoặc } t_3 > t_{\frac{\alpha}{2}; n-(k+1)}\}$$

Trong khuôn khổ bài thực hành này, mức ý nghĩa là  $\alpha = 0,05$

$$W_\alpha = \{t_3 \mid t_3 > t_{0,025;1004} \text{ hoặc } t_3 < -t_{0,025;1004}\} = (-\infty, -1,9623) \cup (1,9623, \infty)$$

**Bước 5: Kết luận**

$$\Rightarrow t_3 \in W_\alpha$$

Chúng ta hoàn toàn có thể bác bỏ giả thuyết  $H_0 : \beta_3 = 0$ , tức là biến  $x_3$  có ý nghĩa trong mô hình hồi quy.

**4.4 Ý nghĩa  $\beta_4$** **Bước 1: Cặp giả thuyết**

$$\begin{cases} H_0 : \beta_4 = \beta_{4,0} = 0 \\ H_1 : \beta_4 \neq \beta_{4,0} = 0 \end{cases}$$

**Bước 2: Chọn Tiêu chuẩn kiểm định**

$$T_4 = \frac{\hat{\beta}_4 - \beta_{4,0}}{se(\hat{\beta}_4)} \sim t_{(n-(k+1))} \quad (4.10)$$

**Bước 3: Tính toán từ bộ dữ liệu giá trị quan sát**

- Sai số chuẩn của ước lượng :

$$se(\hat{\beta}_2) = \sqrt{\sigma^2 C_{55}} \quad (4.11)$$

Trong đó:

$\sigma^2$  chưa biết nên cần ước lượng bởi:  $s^2 = \frac{SS_E}{n-(k+1)} = 15,0751$

$C_{55}$  là phần tử thứ 3 trên đường chéo chính của ma trận  $(X^T X)^{-1}$  ứng với  $\hat{\beta}_4$

- Giá trị quan sát :

$$t_4 = \frac{0,8804 \times 10^{-7} - 0}{\sqrt{15,0751 \times 6,0197 \times 10^{-17}}} = 2,9226 \quad (4.12)$$

**Bước 4: Miền bác bỏ**

$$W_\alpha = \{t_4 \mid t_4 < t_{\frac{\alpha}{2}; n-(k+1)} \text{ hoặc } t_4 > t_{\frac{\alpha}{2}; n-(k+1)}\}$$

Trong khuôn khổ bài thực hành này, mức ý nghĩa là  $\alpha = 0,05$

$$W_\alpha = \{t_4 \mid t_4 > t_{0,025; 1004} \text{ hoặc } t_4 < -t_{0,025; 1004}\} = (-\infty, -1,9623) \cup (1,9623, \infty)$$

**Bước 5: Kết luận**

$$\Rightarrow t_4 \in W_\alpha$$

Chúng ta hoàn toàn có thể bác bỏ giả thuyết  $H_0 : \beta_4 = 0$ , tức là biến  $x_4$  có ý nghĩa trong mô hình hồi quy.

Hệ số hồi quy  $\beta_4$  có giá trị nhỏ, cho thấy tác động của  $X_4$  đến  $Y$  là yếu, mặc dù ý nghĩa thống kê chỉ ra rằng tác động này không phải là ngẫu nhiên.

### Đoạn mã tính toán

In ra ma trận  $(X^T X)^{-1}$

```

1 XTX_inv = np.linalg.inv(XT @ X)
2 # In ma tran (X^T X)^-1
3 print("\nMa tran (X^T X)^-1:")
4 print(XTX_inv)

```

Tính toán  $t_j$  với  $j=1,2,3,4$

```

1 # Tinh cac gia tri T_j
2 C_diag = np.diag(np.linalg.inv(XT @ X)) # Lay duong cheo chinh
      cua ma tran (X^T X)^-1
3 T_values = Beta_hat / np.sqrt(MS_E * C_diag) # Tinh T_j theo
      cong thuc
4
5 # In ket qua
6 print("Gia tri T_j:")
7 for j, T_j in enumerate(T_values):
8     print(f"T_{j} = {T_j}")

```

Tính  $W_\alpha$  và xét xem  $t_j \in W_\alpha$  hay không

```

1 from scipy.stats import t
2
3 # Dinh nghia muc y nghia alpha va bac tu do
4 alpha = 0.05
5 df = n - (k + 1) # Bac tu do
6
7 # Tim gia tri toi han t_{alpha/2; df} cho mien hai phia
8 t_critical = t.ppf(1 - alpha / 2, df) # Vi phan phoi t doi

```

```

    xung, chi can t_critical ben phai
9
10 # Kiem tra tung t_j co thuoc mien bac bo hay khong
11 W_alpha_results = []
12 for j, T_j in enumerate(T_values):
13     if T_j > t_critical or T_j < -t_critical:
14         W_alpha_results.append(f"T_{j} thuoc W_alpha ( $|T_{j}| >$ 
15                                 {t_critical:.4f}))")
16     else:
17         W_alpha_results.append(f"T_{j} khong thuoc W_alpha ( $|T_{j}|$ 
18                                 {t_critical:.4f}))")
19
20 # In ket qua
21 print("\nKiem tra T_j thuoc mien bac bo W_alpha:")
22 print("\n".join(W_alpha_results))
```

## 5 Kết luận mô hình hồi quy bội

Dựa trên mô hình đã xây dựng và kiểm định đã thực hiện trên mô hình:  
**Mô hình hồi quy tuyến tính bội:**

$$\hat{y} = -0,4553 - 0,5977x_1 + 0,7723x_2 + 0,8264x_3 + (0,8804 \times 10^{-7}) x_4$$

**Kiểm định mô hình bằng ANNOVA:**

- Phương pháp kiểm định Fisher được sử dụng để đánh giá sự phù hợp của mô hình.
- Giá trị  $f_0$  quan sát được nằm trong miền bác bỏ, chứng tỏ rằng mô hình hồi quy tổng thể có ý nghĩa thống kê.
- Điều này khẳng định rằng ít nhất một biến độc lập trong mô hình có ảnh hưởng đến biến phụ thuộc

**Kiểm định từng hệ số góc  $\beta_j$**

- Tất cả các giá trị đều thuộc miền bác bỏ, tức là:  
 Kết quả này khẳng định rằng tất cả các hệ số góc  $\beta_j$  đều có ý nghĩa thống kê ở mức ý nghĩa

**Kết luận với mục tiêu đưa ra**

- **Kiểm tra độ phù hợp của mô hình:**
  - Mô hình hồi quy bội được xây dựng là phù hợp và có ý nghĩa thống kê để dự đoán biến phụ thuộc với  $R^2 = 0,9987$  :
  - Y (giá đóng phiên điều chỉnh) dựa trên các biến độc lập bao gồm: giá mở cửa, giá cao nhất, giá thấp nhất, và khối lượng giao dịch.

- **Phân tích ảnh hưởng của các biến độc lập:**

- Các hệ số hồi quy ( $\beta_j$ ) phản ánh đúng mối quan hệ giữa các biến độc lập và phụ thuộc, với mỗi biến đều đóng vai trò quan trọng trong mô hình.

- **Dự đoán giá cổ phiếu**

- Mô hình  $\hat{y}$  cung cấp khả năng dự đoán giá Adjusted Close dựa trên các yếu tố giao dịch hàng ngày. Kết quả mô hình cho thấy:

- Mô hình đạt được độ chính xác tốt  $R^2 = 0,9987$  khi so sánh giá trị dự đoán ( $\hat{y}$ ) với giá trị thực tế ( $y$ ).

- Mô hình có thể được sử dụng để dự báo giá cổ phiếu trong tương lai, hỗ trợ các quyết định đầu tư.

- **Hiểu biến động thị trường**

- Mô hình hồi quy đã được xây dựng dựa trên dữ liệu từ năm 2018 đến 2022, phản ánh mối quan hệ giữa các yếu tố thị trường (giá mở cửa, giá cao nhất, giá thấp nhất, và khối lượng giao dịch) và giá cổ phiếu sau điều chỉnh.

- Kết quả cho thấy:

Các yếu tố như giá mở cửa ( $X_1$ ), giá cao nhất ( $X_2$ ), và giá thấp nhất ( $X_3$ ) có ảnh hưởng đáng kể đến giá cổ phiếu. Hệ số hồi quy rất nhỏ của khối lượng giao dịch ( $X_4$ ) chỉ ra rằng yếu tố này đóng vai trò hạn chế trong việc giải thích sự biến động giá cổ phiếu, nhưng vẫn có ý nghĩa thống kê.

⇒ Điều này gợi ý rằng các nhà đầu tư nên tập trung hơn vào việc phân tích giá mở cửa, giá cao nhất và giá thấp nhất để nắm bắt xu hướng giá.

- **Ứng dụng vào quản lý rủi ro**

- Mô hình  $\hat{y}$  cho phép dự đoán giá cổ phiếu dựa trên các yếu tố đầu vào, từ đó : Giảm thiểu rủi ro giao dịch

⇒ Cung cấp thông tin đáng tin cậy giúp các nhà đầu tư nhận diện và đánh giá xu hướng biến động giá cổ phiếu.



- **Hỗ trợ ra quyết định**

- Các yếu tố có ý nghĩa thống kê cao như giá mở cửa, giá cao nhất và giá thấp nhất có thể được sử dụng để thiết lập chiến lược giao dịch tối ưu.

- **Hạn chế và hướng nghiên cứu tương lai**

- Hạn chế**

- Khối lượng giao dịch có thể ảnh hưởng gián tiếp hoặc có mối quan hệ phi tuyến với giá cổ phiếu, đòi hỏi các nghiên cứu sâu hơn hoặc mô hình phức tạp hơn như mô hình phi tuyến hoặc học máy.

- Định hướng tương lai**

- Tăng độ chính xác dự đoán: Kết hợp các yếu tố vĩ mô (như lãi suất, dữ liệu kinh tế) hoặc sử dụng các phương pháp học máy để cải thiện khả năng dự đoán.

- Dự đoán dài hạn: Áp dụng mô hình trên các khung thời gian dài hơn để phân tích xu hướng thị trường và chiến lược đầu tư.