

CHỦ ĐỀ 4. MÔ HÌNH HỒI QUY ĐƠN BIẾN

1. TÌM ĐƯỜNG THẲNG HỒI QUY

1.1. DÙNG CÔNG THỨC

B1. Từ đề bài, ta có được $n, \sum_{i=1}^n x_i; \sum_{i=1}^n y_i; \sum_{i=1}^n x_i y_i; \sum_{i=1}^n x_i^2; \sum_{i=1}^n y_i^2$.

B2. Tính

$$\begin{aligned} S_{xy} &= \sum_{i=1}^n x_i y_i - \frac{(\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n} \\ S_{xx} &= \sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n} \\ \Rightarrow \hat{\beta}_1 &= \frac{S_{xy}}{S_{xx}} \end{aligned}$$

B3. Tính $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ và $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$, ta tính được

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

B5. Thay giá trị $\hat{\beta}_0, \hat{\beta}_1$ vào đường thẳng hồi quy

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x.$$

$\Rightarrow$ tung độ góc $\hat{\beta}_0 = \dots$

$\Rightarrow$ hệ số góc $\hat{\beta}_1 = \dots$

BÀI 4.1 (Câu 5 - Đề 1 CKI 20-21). Các phương pháp hồi quy đã được sử dụng để phân tích dữ liệu từ một nghiên cứu điều tra mối quan hệ giữa nhiệt độ bề mặt đường (x) và độ lún mặt đường (y). Số liệu được tóm tắt như sau

$$\begin{aligned} n = 20; \quad \sum_{i=1}^n y_i &= 12.75; \quad \sum_{i=1}^n y_i^2 = 8.86; \quad \sum_{i=1}^n x_i = 1478; \\ \sum_{i=1}^n x_i^2 &= 143215.8; \quad \sum_{i=1}^n x_i y_i = 1083.67 \end{aligned}$$

- a) Tìm phương trình hồi quy tuyến tính đơn y theo x . Giải thích ý nghĩa của $\hat{\beta}_1$ nhận được.
- b) Hãy tiên đoán lượng độ lún mặt đường sẽ quan trắc được khi nhiệt độ bề mặt đường là 85?

LỜI GIẢI.

- a) **Tìm phương trình hồi quy tuyến tính đơn y theo x .** Giải thích ý nghĩa của $\hat{\beta}_1$ nhận được.
Ta có $n = 20$ và

☑ $S_{xy} = \sum_{i=1}^n x_i y_i - \frac{(\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n} = 1083.67 - \frac{1478 \cdot 12.75}{20} = 141.445.$

☑ $S_{xx} = \sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n} = 143215.8 - \frac{1478^2}{20} = 33991.6.$

$$\Rightarrow \hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{141.445}{33991.6} \approx 0.0042.$$

$$\checkmark \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{20} \cdot 1478 = 73.9$$

$$\checkmark \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = \frac{1}{20} \cdot 12.75 = 0.6375$$

$$\Rightarrow \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 0.6375 - 0.0042 \cdot 73.9 \approx 0.3271.$$

$$\Rightarrow \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = 0.3271 + 0.0042x.$$

Vậy $\hat{y} = 0.3271 + 0.0042x$.

Giải thích ý nghĩa của $\hat{\beta}_1$ nhận được.

Vì $\hat{\beta}_1 \approx 0.0042 > 0$ nên khi nhiệt độ bề mặt đường tăng 1 đơn vị thì độ lún mặt đường sẽ tăng gần 0.0042 đơn vị.

b) **Hãy tiên đoán lượng độ lún mặt đường sẽ quan trắc được khi nhiệt độ bề mặt đường là 85?**

Khi nhiệt độ bề mặt đường là 85 $\Rightarrow x_0 = 85$

$$\Rightarrow \hat{y}_0 = 0.3271 + 0.0042 \cdot 85 = 0.6841.$$

Vậy nhiệt độ bề mặt đường là 85 thì độ lún mặt đường sẽ là khoảng 0.6841 đơn vị.

□

BÀI 4.2 (Câu 6 - Đề 3 CKII 19-20). Bệnh tiểu đường và béo phì là những vấn đề sức khỏe nghiêm trọng ở Hoa Kỳ và phần lớn các nước phát triển. Đo lường mỡ cơ thể của một người là một cách để theo dõi tiến độ kiểm soát cân nặng, nhưng đo chính xác nó phải sử dụng đến thiết bị X-quang đắt tiền hoặc nhúng cơ thể xuống một hồ bơi. Thay vào đó, chỉ số khối cơ thể (BMI) thường được sử dụng làm đại diện cho mỡ cơ thể vì nó dễ đo: $\text{BMI} = \text{khối lượng (kg)} / (\text{chiều cao (m)})^2 = 703 \text{ khối lượng (lb)} / (\text{chiều cao (in)})^2$. Trong một nghiên cứu của 250 người đàn ông tại Đại học Bingham Young, cả BMI (X) và mỡ cơ thể (Y) được đo lường. Các nhà nghiên cứu đã tìm thấy các thống kê tóm tắt sau:

$$\sum_{i=1}^n x_i = 6322.28; \quad \sum_{i=1}^n x_i^2 = 162674.18; \quad \sum_{i=1}^n y_i = 4757.90;$$

$$\sum_{i=1}^n y_i^2 = 107679.27; \quad \sum_{i=1}^n x_i y_i = 125471.10$$

a) Viết phương trình hồi quy biểu diễn mối liên hệ của Y theo X có dạng như sau: $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$ và nêu ý nghĩa của hệ số $\hat{\beta}_1$.

b) Sử dụng đường thẳng hồi quy, hãy tiên đoán lượng mỡ cơ thể của một người đàn ông sẽ được quan trắc nếu có chỉ số BMI là 30?

LỜI GIẢI.

Ta có

$$n = 250; \quad \sum_{i=1}^n x_i = 6322.28; \quad \sum_{i=1}^n x_i^2 = 162674.18; \quad \sum_{i=1}^n y_i = 4757.90;$$

$$\sum_{i=1}^n y_i^2 = 107679.27; \quad \sum_{i=1}^n x_i y_i = 125471.10$$

a) **Tính những ước lượng bình phương tối thiểu cho hệ số góc và tung độ góc.**

$$\checkmark S_{xy} = \sum_{i=1}^n x_i y_i - \frac{(\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n} = 125471.10 - \frac{6322.28 \cdot 4757.90}{250} = 5147.99595.$$

$$\checkmark S_{xx} = \sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n} = 162674.18 - \frac{6322.28^2}{250} = 2789.28241.$$

$$\Rightarrow \hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{5147.99595}{2789.28241} \approx 1.84563.$$

$$\checkmark \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{250} \cdot 6322.28 = 25.28912$$

$$\checkmark \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = \frac{1}{250} \cdot 4757.90 = 19.0316$$

$$\Rightarrow \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 19.0316 - 25.28912 \cdot 1.84563 \approx -27.64276.$$

$$\Rightarrow \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = -27.64276 + 1.84563x.$$

$$\text{Vậy } \hat{y} = -27.64276 + 1.84563x$$

- b) **Sử dụng đường thẳng hồi quy, hãy tiên đoán lượng mỡ cơ thể của một người đàn ông sẽ được quan trắc nếu có chỉ số BMI là 30?**

$$\text{Chỉ số BMI là 30, tức } x = 30 \Rightarrow \hat{y} = -27.64276 + 1.84563 \cdot 30 = 27.72614.$$

□

BÀI 4.3 (Câu 6 - Đề 2 HKI 19-20). Một bài báo trong Nghiên cứu Bê tông “Đặc tính bề mặt gần bê tông: Tính thấm nội tại” (1989, Tập 41) trình bày dữ liệu về cường độ nén x và độ thấm nội tại y của các hỗn hợp bê tông và phương pháp xử lý khác nhau. Số liệu được tóm tắt như sau:

$$n = 14; \quad \sum_{i=1}^n y_i = 572; \quad \sum_{i=1}^n y_i^2 = 23530; \quad \sum_{i=1}^n x_i = 43;$$

$$\sum_{i=1}^n x_i^2 = 157.42; \quad \sum_{i=1}^n x_i y_i = 1697.80$$

- a) Tính những ước lượng bình phương tối thiểu cho hệ số góc và tung độ góc. Vẽ đồ thị của đường thẳng hồi quy.
- b) Sử dụng đường thẳng hồi quy, hãy tiên đoán lượng độ thấm nội sẽ quan trắc được khi cường độ nén là $x = 4.3$?

LỜI GIẢI.

Ta có

$$n = 14; \quad \sum_{i=1}^n y_i = 572; \quad \sum_{i=1}^n y_i^2 = 23530; \quad \sum_{i=1}^n x_i = 43;$$

$$\sum_{i=1}^n x_i^2 = 157.42; \quad \sum_{i=1}^n x_i y_i = 1697.80$$

- a) **Tính những ước lượng bình phương tối thiểu cho hệ số góc và tung độ góc.**

$$\checkmark S_{xy} = \sum_{i=1}^n x_i y_i - \frac{(\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n} = 1697.80 - \frac{43 \cdot 572}{14} = -\frac{2067}{35}.$$

$$\checkmark S_{xx} = \sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n} = 157.42 - \frac{43^2}{14} = \frac{4436}{175}.$$

$$\Rightarrow \hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{-\frac{2067}{35}}{\frac{4436}{175}} \approx -2.32980.$$

$$\checkmark \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{14} \cdot 43 = \frac{43}{14}$$

$$\checkmark \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = \frac{1}{14} \cdot 572 = \frac{286}{7}$$

$$\Rightarrow \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \frac{286}{7} - \frac{43}{14} \cdot (-2.32980) \approx 48.01296.$$

$$\Rightarrow \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = 48.01296 - 2.32980x.$$

$$\text{Vậy } \hat{y} = 48.01296 - 2.32980x.$$

- b) **Sử dụng đường thẳng hồi quy, hãy tiên đoán lượng độ thấm nội sẽ quan trắc được khi cường độ nén là $x = 4.3$?**

$$\text{Cường độ nén là } x = 4.3 \Rightarrow \hat{y} = 48.01296 - 2.32980 \cdot 4.3 = 37.99482.$$

□

1.2. DÙNG MÁY TÍNH CẦM TAY

Nếu muốn ước lượng đường thẳng hồi quy từ dữ liệu có dạng bảng như sau

x	77	50	71	72	81	94	96	99	67
y	82	66	78	34	47	85	99	99	68

Khi đó, ta tiến hành thao tác trên máy tính cầm tay như sau

Fx 570	Fx 580
Mode → 3:STAT → 2: A + BX	MENU → 6: Stat → 2: y = a + bX
→ Nhập dữ liệu → AC	→ Nhập dữ liệu → AC
SHIFT → 1 → 5: Reg	OPTN → ↓ → 4: Regression
1: A 2: B 3: r	1: a 2: b 3: r
SHIFT → 1 → 4: Var → 1:n	OPTN → ↓ → 2: Variable → 6:n
Tính các tổng để tính R^2	Tính các tổng để tính R^2
SHIFT → 1 → 3: Sum	OPTN → ↓ → 1: Summation
1: $\sum x^2$ 2: $\sum x$ 3: $\sum y^2$ 4: $\sum y$ 5: $\sum xy$	1: $\sum x$ 2: $\sum x^2$ 3: $\sum y$ 4: $\sum y^2$ 5: $\sum xy$

trong đó

☑ Hệ số chặn (tung độ góc): $\hat{\beta}_0 = A(=a) = 12.06232$

☑ Hệ số góc: $\hat{\beta}_1 = B(=b) = 0.77714$
 $\Rightarrow \hat{y} = 12.06232 + 0.77714x.$

2. DỰ ĐOÁN GIÁ TRỊ QUAN TRẮC MỚI

Bài toán: Cho trước giá trị $x = x_0$, tính $\hat{y}_0$.

Thay giá trị $x = x_0$ đề bài cho vào biểu thức vừa tìm được ở trên

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x.$$

ta tính được giá trị $\hat{y}_0$ tương ứng.

3. Ý NGHĨA HỆ SỐ HỒI QUY

☑ $\hat{\beta}_0$: Khi các yếu tố bị loại khỏi mô hình ($x = 0$), thì trung bình của Y là $\hat{\beta}_0$.

☑ $\hat{\beta}_1$: khi x thay đổi 1 đơn vị (của x), các yếu tố khác không đổi thì Y thay đổi $\hat{\beta}_1$ đơn vị (của y).

Thông thường sẽ hỏi về ý nghĩa hệ số góc $\hat{\beta}_1$

BÀI 4.4 (Câu 5 - Đề 1 HKI 22-23). Một người sử dụng một sợi dây thun (hoặc một lò xo) có độ đàn hồi cao để làm một cái cân đơn giản. Anh ta treo các vật nặng lên dây và ghi lại chiều dài sợi dây cho mỗi lần cân. Dưới đây là dữ liệu của một số lần cân

Khối lượng g (x)	50	100	150	200	250	300	350	400
Chiều dài mm (y)	37	48	60	71	80	90	102	109

a) Tìm phương trình hồi quy tuyến tính y theo x . Giải thích ý nghĩa của $\hat{\beta}_1$ nhận được.

b) Bạn dự đoán chiều dài sợi dây là bao nhiêu nếu một vật có trọng lượng 375 g được treo lên?

🔗 LỜI GIẢI.

Giả sử X, Y thỏa mãn mô hình hồi quy tuyến tính đơn biến.

a) **Tìm phương trình hồi quy tuyến tính y theo x .**

Từ bảng dữ liệu, dùng máy cầm tay, ta ước lượng được: $\hat{\beta}_0 \approx 27.8571$; $\hat{\beta}_1 \approx 0.2079$.

Phương trình hồi quy tuyến tính đơn y theo x là $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = 27.8571 + 0.2079x$.

✳ **Giải thích ý nghĩa của $\hat{\beta}_1$ nhận được:**

Vì $\hat{\beta}_1 = 0.2079 > 0$ nên khi khối lượng lên 1 g thì chiều dài của sợi dây tăng 0.2079 mm.

b) **Bạn dự đoán chiều dài sợi dây là bao nhiêu nếu một vật có trọng lượng 375 g được treo lên?**

Nếu vật có trọng lượng 375 g $\Rightarrow x_0 = 375$

$$\Rightarrow \hat{y}_0 = 27.8571 + 0.2079 \cdot 375 = 105.8196.$$

Vậy nếu một vật có trọng lượng 375 g được treo lên thì chiều dài của sợi dây gần bằng 105.8196 mm.

□

BÀI 4.5 (Câu 5 - Đề 2 HKI 22-23). Một thí nghiệm đã được thực hiện để nghiên cứu sự thay đổi độ hòa tan của hóa chất Y trong nước. Khối lượng hòa tan (tính bằng kg) trong 1 lít nước ở các nhiệt độ nước khác nhau được cho trong bảng dưới đây

Nhiệt độ $^{\circ}\text{C}$ (x)	15	20	25	30	35	50	70
Khối lượng của Y (y)	2.1	2.6	2.9	3.3	4.0	5.1	7.0

- Tìm phương trình hồi quy tuyến tính y theo x . Giải thích ý nghĩa của $\hat{\beta}_1$ nhận được.
- Bạn dự đoán khối lượng hòa tan là bao nhiêu nếu nhiệt độ nước là 42°C ?

LỜI GIẢI.

Giả sử X, Y thỏa mãn mô hình hồi quy tuyến tính đơn biến.

- Tìm phương trình hồi quy tuyến tính y theo x .**

Từ bảng dữ liệu, dùng máy cầm tay, ta ước lượng được: $\hat{\beta}_0 \approx 0.7549$; $\hat{\beta}_1 \approx 0.0886$.

Phương trình hồi quy tuyến tính đơn y theo x là $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = 0.7549 + 0.0886x$.

*** Giải thích ý nghĩa của $\hat{\beta}_1$ nhận được:**

Vì $\hat{\beta}_1 = 0.0886 > 0$ nên khi tăng nhiệt độ lên 1°C thì khối lượng hòa tan tăng 0.0886 kg.

- Bạn dự đoán khối lượng hòa tan là bao nhiêu nếu nhiệt độ nước là 42°C .**

Khi nhiệt độ nước là $42^{\circ}\text{C} \Rightarrow x_0 = 42$

$$\Rightarrow \hat{y}_0 = 0.7549 + 0.0886 \cdot 42 = 4.4761.$$

Vậy nếu nhiệt độ nước là 42°C thì khối lượng hòa tan là 4.4761 kg.

□

BÀI 4.6 (Câu 4 - Đề 1 CKII 21-22). Dữ liệu bên dưới mô tả về trọng lượng của các điều thuốc lá x (g) được sản xuất từ các nhà máy khác nhau và hàm lượng nicotin y (mg) trong mỗi điều thuốc:

x	15.8	14.9	9.0	4.5	15.0	17.0	8.6	12.0	4.1	16.0
y	0.957	0.886	0.852	0.911	0.889	0.919	0.969	1.118	0.946	1.094

- Tìm phương trình hồi quy tuyến tính đơn y theo x . Giải thích ý nghĩa của $\hat{\beta}_1$ nhận được.
- Dự đoán hàm lượng nicotin của một điều thuốc có trọng lượng 11 g.

LỜI GIẢI.

Giả sử X, Y thỏa mãn mô hình hồi quy tuyến tính đơn biến.

- Tìm phương trình hồi quy tuyến tính đơn y theo x :**

Từ bảng dữ liệu, dùng máy cầm tay, ta ước lượng được: $\hat{\beta}_0 \approx 0.9184$; $\hat{\beta}_1 \approx 3.0563 \cdot 10^{-3} \approx 0.0031$.

Phương trình hồi quy tuyến tính đơn y theo x là $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = 0.9184 + 0.0031x$.

*** Giải thích ý nghĩa của $\hat{\beta}_1$ nhận được:**

Vì $\hat{\beta}_1 = 0.0031 > 0$ nên khi trọng lượng của điều thuốc lá tăng 1 g và các yếu tố khác không đổi thì hàm lượng nicotin trong điều thuốc tăng 0.0031 mg.

- Dự đoán hàm lượng nicotin của một điều thuốc có trọng lượng 11 g.**

Điều thuốc có trọng lượng 11 g $\Rightarrow x_0 = 11$

$$\Rightarrow \hat{y}_0 = 0.9184 + 0.0031 \cdot 11 = 0.9525.$$

Vậy nếu một điều thuốc có trọng lượng 11 g thì hàm lượng nicotin là 0.9525 mg.

BÀI 4.7 (Câu 1 - Đề 1 CKI 21-22). Trong cấu tạo của một loại dây thừng, người ta quan tâm đến hàm lượng nylon X (Đv: %) có ảnh hưởng như thế nào đến lực căng Y (Đv: psi) (là lực kéo tối đa trước khi sợi dây bị đứt). Số liệu bên dưới cho kết quả đo của 10 sợi dây với hàm lượng nylon khác nhau

Hàm lượng nylon X	7	12	17	22	32	42	47	52	52	62
Lực căng Y	169	249	289	329	349	404	459	519	529	559

(Làm tròn các đáp án đến 4 chữ số sau dấu chấm thập phân)

- Tìm phương trình đường thẳng hồi quy ước lượng $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ ($\hat{\beta}_0$: hệ số chặn hay hệ số tự do; $\hat{\beta}_1$: hệ số góc).
- Tính hệ số tương quan r_{XY} giữa X và Y .
- Nếu một sợi dây có hàm lượng nylon bằng 45% thì giá trị dự báo cho lực căng của sợi dây bằng bao nhiêu?

🔗 LỜI GIẢI.

Giả sử X, Y thỏa mãn mô hình hồi quy tuyến tính đơn.

- Tìm phương trình đường thẳng hồi quy ước lượng $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ ($\hat{\beta}_0$: hệ số chặn hay hệ số tự do; $\hat{\beta}_1$: hệ số góc).

Từ bảng dữ liệu, dùng máy cầm tay, ta ước lượng được: $\hat{\beta}_0 \approx 154.4151$; $\hat{\beta}_1 \approx 6.6981$.
 Phương trình hồi quy tuyến tính đơn y theo x là $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = 154.4151 + 6.6981x$.

- Tính hệ số tương quan r_{XY} giữa X và Y .

Ta có $r_{XY} \approx 0.9836$. (bấm máy)

- Nếu một sợi dây có hàm lượng nylon bằng 45% thì giá trị dự báo cho lực căng của sợi dây bằng bao nhiêu?

Sợi dây có hàm lượng nylon bằng 45% $\Rightarrow x_0 = 45$

$$\Rightarrow \hat{y}_0 = 154.4151 + 6.6981 \cdot 45 = 455.8296.$$

Nếu một sợi dây có hàm lượng nylon bằng 45% thì giá trị dự báo cho lực căng của sợi dây bằng 455.8296 psi.

Một số câu tương tự

- ☑ Câu 4 - Đề 1 CKII 20-21: Một nghiên cứu về khối lượng đường bị biến đổi...
- ☑ Câu 4 - Đề 2 CKII 20-21: Khối lượng của một hợp chất hóa học (y)...
- ☑ Câu 5 - Đề 2 CKI 20-21: Điểm thi giữa kỳ (x) và cuối kỳ (y) của một lớp...
- ☑ Câu 4 - Đề 2 CKII 19-20: Bảng số liệu sau cung cấp thông tin về chiều cao và số nhánh của 10 cây vừng...