

Xác suất
thống kê - Hồi
quy tuyến tính
đơn

Dẫn nhập

Mô hình hồi
quy tuyến tính
đơn

Ước lượng các
tham số hồi
quy

Hệ số R^2

Các tính chất
của các ước
lượng

Dự đoán

Xác suất thống kê - Hồi quy tuyến tính đơn

Ngày 12 tháng 12 năm 2024

Ví dụ dẫn nhập

Xác suất
thống kê - Hồi
quy tuyến tính
đơn

Dẫn nhập

Mô hình hồi
quy tuyến tính
đơn

Ước lượng các
tham số hồi
quy

Hệ số R^2

Các tính chất
của các ước
lượng

Dự đoán

Ví dụ dẫn nhập

Cho bảng giá trị sau:

x	1	3	4	5	7	9	10
y	2	6	8	10	14	18	20

Ví dụ dẫn nhập

Xác suất
thống kê - Hồi
quy tuyến tính
đơn

Dẫn nhập

Mô hình hồi
quy tuyến tính
đơn

Ước lượng các
tham số hồi
quy

Hệ số R^2

Các tính chất
của các ước
lượng

Dự đoán

Ví dụ dẫn nhập

Cho bảng giá trị sau:

x	1	3	4	5	7	9	10
y	2	6	8	10	14	18	20

Dễ thấy giữa x và y có một mối liên hệ tuyến tính, cụ thể: $y = 2x$.

Ví dụ dẫn nhập

Xác suất
thống kê - Hồi
quy tuyến tính
đơn

Dẫn nhập

Mô hình hồi
quy tuyến tính
đơn

Ước lượng các
tham số hồi
quy

Hệ số R^2

Các tính chất
của các ước
lượng

Dự đoán

Ví dụ dẫn nhập

Cho bảng giá trị sau:

x	1	3	4	5	7	9	10
y	2	6	8	10	14	18	20

Dễ thấy giữa x và y có một mối liên hệ tuyến tính, cụ thể: $y = 2x$.
Tuy nhiên, nếu bảng giá trị này có thay đổi nhỏ, chẳng hạn:

x	1	3	4	5	7	9	10
y	3	5	7	9	15	17	21

Ví dụ dẫn nhập

Xác suất
thống kê - Hồi
quy tuyến tính
đơn

Dẫn nhập

Mô hình hồi
quy tuyến tính
đơn

Ước lượng các
tham số hồi
quy

Hệ số R^2

Các tính chất
của các ước
lượng

Dự đoán

Ví dụ dẫn nhập

Cho bảng giá trị sau:

x	1	3	4	5	7	9	10
y	2	6	8	10	14	18	20

Dễ thấy giữa x và y có một mối liên hệ tuyến tính, cụ thể: $y = 2x$.
Tuy nhiên, nếu bảng giá trị này có thay đổi nhỏ, chẳng hạn:

x	1	3	4	5	7	9	10
y	3	5	7	9	15	17	21

thì mối quan hệ tuyến tính này không còn nữa.

Ví dụ dẫn nhập

Xác suất
thống kê - Hồi
quy tuyến tính
đơn

Dẫn nhập

Mô hình hồi
quy tuyến tính
đơn

Ước lượng các
tham số hồi
quy

Hệ số R^2

Các tính chất
của các ước
lượng

Dự đoán

Ví dụ dẫn nhập

Cho bảng giá trị sau:

x	1	3	4	5	7	9	10
y	2	6	8	10	14	18	20

Dễ thấy giữa x và y có một mối liên hệ tuyến tính, cụ thể: $y = 2x$.
Tuy nhiên, nếu bảng giá trị này có thay đổi nhỏ, chẳng hạn:

x	1	3	4	5	7	9	10
y	3	5	7	9	15	17	21

thì mối quan hệ tuyến tính này không còn nữa. Trong trường hợp này, ta có thể viết mối quan hệ giữa x và y như sau: $y = 2x + \epsilon$, với ϵ là sai số nhỏ.

Mô hình hồi quy tuyến tính đơn

Xác suất
thống kê - Hồi
quy tuyến tính
đơn

Dẫn nhập

Mô hình hồi
quy tuyến tính
đơn

Ước lượng các
tham số hồi
quy

Hệ số R^2

Các tính chất
của các ước
lượng

Dự đoán

Mô hình hồi quy tuyến tính đơn

Cho bộ dữ liệu gồm n điểm dữ liệu $\{(x_1, Y_1), (x_2, Y_2), \dots, (x_n, Y_n)\}$.
Mô hình tuyến tính đơn cho biết mối quan hệ tuyến tính giữa x_i và Y_i qua đẳng thức:

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad \forall i \in \overline{1, n}$$

trong đó β_0, β_1 là các hằng số chưa biết, $(\epsilon_1, \epsilon_2, \dots, \epsilon_n)$ là các biến ngẫu nhiên độc lập, cùng phân phối, và thỏa

$$\mathbb{E}(\epsilon_i) = 0, \text{Var}(\epsilon_i) = \sigma^2 > 0$$

Trong chương này, ta có thể giả sử rằng $(\epsilon_1, \dots, \epsilon_n) \sim \mathcal{N}(0, \sigma^2)$.

Biến hồi quy và biến đáp ứng

Trong mô hình hồi quy, ta gọi (x_1, x_2, \dots, x_n) là các **biến hồi quy** và (Y_1, Y_2, \dots, Y_n) là các biến đáp ứng.

Ước lượng các tham số hồi quy

Xác suất
thống kê - Hồi
quy tuyến tính
đơn

Dẫn nhập

Mô hình hồi
quy tuyến tính
đơn

Ước lượng các
tham số hồi
quy

Hệ số R^2

Các tính chất
của các ước
lượng

Dự đoán

Ước lượng tham số hồi quy

Cho các điểm quan trắc $\{(x_1, y_1), \dots, (x_n, y_n)\}$, quan sát được từ mô hình hồi quy tuyến tính.

Ước lượng các tham số hồi quy

Xác suất
thống kê - Hồi
quy tuyến tính
đơn

Dẫn nhập

Mô hình hồi
quy tuyến tính
đơn

Ước lượng các
tham số hồi
quy

Hệ số R^2

Các tính chất
của các ước
lượng

Dự đoán

Ước lượng tham số hồi quy

Cho các điểm quan trắc $\{(x_1, y_1), \dots, (x_n, y_n)\}$, quan sát được từ mô hình hồi quy tuyến tính.

- **Mục tiêu:** Tìm các giá trị ước lượng $\hat{\beta}_0, \hat{\beta}_1$ của β_0 và β_1 sao cho sai số là nhỏ nhất.

Ước lượng các tham số hồi quy

Xác suất
thống kê - Hồi
quy tuyến tính
đơn

Dẫn nhập

Mô hình hồi
quy tuyến tính
đơn

Ước lượng các
tham số hồi
quy

Hệ số R^2

Các tính chất
của các ước
lượng

Dự đoán

Ước lượng tham số hồi quy

Cho các điểm quan trắc $\{(x_1, y_1), \dots, (x_n, y_n)\}$, quan sát được từ mô hình hồi quy tuyến tính.

- **Mục tiêu:** Tìm các giá trị ước lượng $\hat{\beta}_0, \hat{\beta}_1$ của β_0 và β_1 sao cho sai số là nhỏ nhất.
- **Sai số:** Với hai giá trị $\hat{\beta}_0, \hat{\beta}_1$, sai số (phần dư) thứ i được định nghĩa bởi $e_i = y_i - \hat{y}_i$ với $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$, được gọi là giá trị phù hợp thứ i .

Phương pháp bình phương bé nhất

Phương pháp bình phương bé nhất là phương pháp tìm các giá trị ước lượng $\hat{\beta}_0, \hat{\beta}_1$ của β_0 và β_1 sao cho **tổng bình phương các sai số là nhỏ nhất**, tức đại lượng

$$SSR = \sum_{k=1}^n e_k^2 = \sum_{k=1}^n (y_k - \hat{\beta}_0 - \hat{\beta}_1 x_k)^2$$

là nhỏ nhất

Ước lượng các tham số hồi quy

Xác suất
thống kê - Hồi
quy tuyến tính
đơn

Dẫn nhập

Mô hình hồi
quy tuyến tính
đơn

Ước lượng các
tham số hồi
quy

Hệ số R^2

Các tính chất
của các ước
lượng

Dự đoán

Ước lượng cho β_0 và β_1

Phương pháp bình phương bé nhất cho ta hai giá trị ước lượng cho β_0 và β_1 được cho bởi công thức sau:

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

$$\hat{\beta}_1 = \frac{\sum_{k=1}^n (x_k - \bar{x})(y_k - \bar{y})}{\sum_{k=1}^n (x_k - \bar{x})^2}$$

hay

$$\hat{\beta}_1 = \frac{n \sum_{k=1}^n x_k y_k - (\sum_{k=1}^n x_k) (\sum_{k=1}^n y_k)}{n \sum_{k=1}^n x_k^2 - (\sum_{k=1}^n x_k)^2}$$

Ước lượng các tham số hồi quy

Xác suất
thống kê - Hồi
quy tuyến tính
đơn

Dẫn nhập

Mô hình hồi
quy tuyến tính
đơn

Ước lượng các
tham số hồi
quy

Hệ số R^2

Các tính chất
của các ước
lượng

Dự đoán

Ước lượng cho σ^2

Ước lượng cho tham số σ^2 được cho bởi:

$$s^2 = \frac{SSR}{n-2} = \frac{1}{n-2} \sum_{k=1}^n (y_i - \hat{y}_i)^2$$

Ước lượng các tham số hồi quy

Xác suất
thống kê - Hồi
quy tuyến tính
đơn

Dẫn nhập

Mô hình hồi
quy tuyến tính
đơn

Ước lượng các
tham số hồi
quy

Hệ số R^2

Các tính chất
của các ước
lượng

Dự đoán

Tổng kết

Đặt $S_{xx} = \sum_{k=1}^n (x_k - \bar{x})^2$, $S_{yy} = \sum_{k=1}^n (y_k - \bar{y})^2$ và
 $S_{xy} = \sum_{k=1}^n (x_k - \bar{x})(y_k - \bar{y})$. Khi đó,

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$$

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

$$SSR = S_{yy} - \hat{\beta}_1 S_{xy}$$

$$s^2 = \frac{SSR}{n-2}$$

Hệ số R^2

Xác suất
thống kê - Hồi
quy tuyến tính
đơn

Dẫn nhập

Mô hình hồi
quy tuyến tính
đơn

Ước lượng các
tham số hồi
quy

Hệ số R^2

Các tính chất
của các ước
lượng

Dự đoán

Hệ số R^2

Hệ số R^2 được cho bởi công thức:

$$R^2 = 1 - \frac{SSR}{SST}$$

trong đó,

$$SSR = \sum_{k=1}^n (y_k - \hat{y}_k)^2$$

$$SST = \sum_{k=1}^n (y_k - \bar{y})^2$$

Hệ số R^2

Xác suất
thống kê - Hồi
quy tuyến tính
đơn

Dẫn nhập

Mô hình hồi
quy tuyến tính
đơn

Ước lượng các
tham số hồi
quy

Hệ số R^2

Các tính chất
của các ước
lượng

Dự đoán

Ý nghĩa của hệ số R^2

- Hệ số R^2 được dùng để đánh giá mô hình hồi quy.
- Nhận xét: $R^2 \leq 1$. Khi $R^2 \approx 1$ thì bộ dữ liệu có quan hệ tuyến tính ***mạnh***.
- Khi $R^2 \approx 0$, thì bộ dữ liệu có quan hệ tuyến tính ***yếu***, và $\hat{\beta}_0 \approx \bar{y}, \hat{\beta}_1 \approx 0$.
- **Lưu ý:** Hệ số R^2 có thể nhận giá trị âm.

Các tính chất của các ước lượng

Xác suất
thống kê - Hồi
quy tuyến tính
đơn

Dẫn nhập

Mô hình hồi
quy tuyến tính
đơn

Ước lượng các
tham số hồi
quy

Hệ số R^2

Các tính chất
của các ước
lượng

Dự đoán

Các tính chất của ước lượng cho β_0

- $\mathbb{E}(\hat{\beta}_0) = \beta_0$
- $\text{Var}(\hat{\beta}_0) = \frac{\sigma^2}{S_{xx}} \left(\frac{1}{n} \sum_{k=1}^n x_k^2 \right)$
- Khoảng tin cậy $(1 - \alpha) \cdot 100\%$ cho β_0 được cho bởi
 $[\hat{\beta}_0 - \epsilon, \hat{\beta}_0 + \epsilon]$

với dung sai

$$\epsilon = t_{1-\alpha/2}^{n-2} \frac{s}{\sqrt{S_{xx}}} \sqrt{\frac{1}{n} \sum_{k=1}^n x_k^2}$$

trong đó, $t_{1-\alpha/2}^{n-2}$ là phân vị thứ $1 - \frac{\alpha}{2}$ của phân phối Student với bậc tự do $n - 2$.

Các tính chất của các ước lượng

Xác suất
thống kê - Hồi
quy tuyến tính
đơn

Dẫn nhập

Mô hình hồi
quy tuyến tính
đơn

Ước lượng các
tham số hồi
quy

Hệ số R^2

Các tính chất
của các ước
lượng

Dự đoán

Các tính chất của ước lượng cho β_1

- $\mathbb{E}(\hat{\beta}_1) = \beta_1$
- $\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}}$
- Khoảng tin cậy $(1 - \alpha) \cdot 100\%$ cho β_1 được cho bởi
 $[\hat{\beta}_1 - \epsilon, \hat{\beta}_1 + \epsilon]$

với dung sai

$$\epsilon = t_{1-\alpha/2}^{n-2} \frac{s}{\sqrt{S_{xx}}}$$

trong đó, $t_{1-\alpha/2}^{n-2}$ là phân vị thứ $1 - \frac{\alpha}{2}$ của phân phối Student với bậc tự do $n - 2$.

Dự đoán trung bình hay biến đáp ứng

Cho mô hình hồi quy

$$Y = \beta_0 + \beta_1 x + \epsilon, \epsilon \sim \mathcal{N}(0, \sigma^2)$$

Với điểm dữ liệu mới $x = x_0$, ta có thể thực hiện 2 cách sau:

- Dự đoán biến đáp ứng tương ứng: $Y_0 = \beta_0 + \beta_1 x_0 + \epsilon$
- Dự đoán trung bình: $\mu_{Y|x_0} = \beta_0 + \beta_1 x_0$

Khoảng tin cậy cho dự đoán

Xác suất
thống kê - Hồi
quy tuyến tính
đơn

Dẫn nhập

Mô hình hồi
quy tuyến tính
đơn

Ước lượng các
tham số hồi
quy

Hệ số R^2

Các tính chất
của các ước
lượng

Dự đoán

Khoảng tin cậy cho dự đoán của trung bình

Khoảng tin cậy $(1 - \alpha) \cdot 100\%$ cho dự đoán của trung bình $\mu_{Y|x_0}$ được cho bởi

$$[\hat{y}_0 - \epsilon, \hat{y}_0 + \epsilon]$$

với dung sai $\epsilon = t_{1-\alpha/2}^{n-2} s \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}$, và $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$.

Khoảng tin cậy cho dự đoán

Xác suất
thống kê - Hồi
quy tuyến tính
đơn

Dẫn nhập

Mô hình hồi
quy tuyến tính
đơn

Ước lượng các
tham số hồi
quy

Hệ số R^2

Các tính chất
của các ước
lượng

Dự đoán

Khoảng tin cậy cho dự đoán của biến đáp ứng

Khoảng tin cậy $(1 - \alpha) \cdot 100\%$ cho dự đoán của biến đáp ứng
 $Y_0 = \beta_0 + \beta_1 x_0 + \epsilon$ được cho bởi

$$[\hat{y}_0 - \epsilon, \hat{y}_0 + \epsilon]$$

với dung sai $\epsilon = t_{1-\alpha/2}^{n-2} s \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}$, và $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$.

Ví dụ minh họa

Xác suất
thống kê - Hồi
quy tuyến tính
đơn

Dẫn nhập

Mô hình hồi
quy tuyến tính
đơn

Ước lượng các
tham số hồi
quy

Hệ số R^2

Các tính chất
của các ước
lượng

Dự đoán

Ví dụ: Bài tập 6.1

Điểm thi giữa kì (x) và cuối kì (y) của một lớp có 9 sinh viên là như sau:

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

- (a) Ước lượng đường hồi quy tuyến tính.
- (b) Ước lượng điểm bài thi cuối kì của một sinh viên có điểm giữa kì là 85.
- (c) Tính s^2 .
- (d) Xây dựng khoảng tin cậy 95% cho β_0 .
- (e) Xây dựng khoảng tin cậy 95% cho β_1 .
- (f) Sử dụng giá trị s^2 , xây dựng khoảng tin cậy 95% cho $\mu_{Y|85}$