

XSTK
N.T. M. Ngọc

Chương 8: Tương quan và hồi quy

Nguyễn Thị Mộng Ngọc
University of Science, VNU - HCM
ngtmngoc@hcmus.edu.vn

XSTK
N.T. M. Ngọc

Phân tích hồi quy

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPNB

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

5. Phân tích thặng dư

6. Tương quan

Bài toán: trong các hoạt động về khoa học - kỹ thuật, kinh tế, xã hội, ... ta có nhu cầu xác định mối liên giữa hai hay nhiều biến ngẫu nhiên với nhau. Ví dụ:

- Mối liên hệ giữa chiều cao và cỡ giày của một người, từ đó một cửa hàng bán giày dép có thể xác định chính xác cỡ giày của một khách hàng khi biết chiều cao,
- Độ giãn nở của một loại vật liệu theo nhiệt độ môi trường,
- Doanh thu khi bán 1 loại sản phẩm và số tiền chi cho quảng cáo và khuyến mãi,
- ...

Để giải quyết các vấn đề trên, ta sử dụng kỹ thuật **phân tích hồi quy** (Regression Analysis).

XSTK
N.T. M. Ngọc

Phân tích hồi quy

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPNB

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

5. Phân tích thặng dư

6. Tương quan

Phân tích hồi quy được sử dụng để xác định mối liên hệ giữa:

- một biến phụ thuộc Y (biến đáp ứng), và
- một hay nhiều biến độc lập X_1, X_2, \dots, X_p ; các biến này còn được gọi là biến giải thích.
 - Biến phụ thuộc Y phải là biến liên tục,
 - Các biến độc lập X_1, X_2, \dots, X_p có thể là biến liên tục, rời rạc hoặc phân loại.

XSTK
N.T. M. Ngọc

Phân tích hồi quy

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPNB

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

5. Phân tích thặng dư

6. Tương quan

Phân tích hồi quy

- Mối liên hệ giữa X_1, \dots, X_p và Y được biểu diễn bởi một hàm tuyến tính.
- Sự thay đổi trong Y được giả sử do những thay đổi trong X_1, \dots, X_p gây ra.
- Trên cơ sở xác định mối liên hệ giữa biến phụ thuộc Y và các biến giải thích X_1, X_2, \dots, X_p , ta có thể:
 - dự đoán, dự báo giá trị của Y ,
 - giải thích tác động của sự thay đổi trong các biến giải thích lên biến phụ thuộc.

XSTK

N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPBN

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

Mô hình hồi quy tuyến tính đơn

Định nghĩa

Một mô hình hồi quy tuyến tính đơn liên quan đến một biến ngẫu nhiên Y và một biến giải thích x là phương trình:

$$Y = \beta_0 + \beta_1 x + \text{Varepsilon}$$

trong đó,

- β_0 và β_1 là các tham số chưa biết (được gọi lần lượt là hệ số chặn (intercept) và hệ số góc (slope) của đường thẳng hồi quy);
- Y là biến phụ thuộc và x là biến độc lập;
- Varepsilon là thành phần sai số, Varepsilon được giả sử có phân phối chuẩn $\mathcal{N}(0, \sigma^2)$.

5. Phân tích

thặng dư

6. Tương quan

XSTK

N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPBN

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

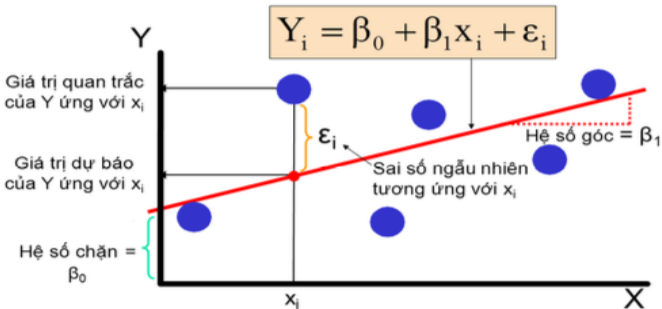
Mô hình hồi quy tuyến tính đơn

Từ mẫu ngẫu nhiên $((X_1, Y_1), \dots, (X_n, Y_n))$ ta có n cặp giá trị quan trắc $(x_1, y_1), \dots, (x_n, y_n)$. Mô hình hồi quy tuyến tính đơn của mẫu là

$$Y_i = \beta_0 + \beta_1 x_i + \text{Varepsilon}_i, \quad \text{với } i = 1, 2, \dots, n \quad (2)$$

trong đó, (Varepsilon_i) i.i.d. $\sim \mathcal{N}(0, \sigma^2)$.

Đồ thị phân tán của n cặp giá trị quan sát $(x_i, y_i), i = 1, 2, \dots, n$.



5. Phân tích

thặng dư

XSTK

N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPBN

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

Mô hình hồi quy tuyến tính đơn

Trong mô hình (1), sự thay đổi của Y được sử dụng bởi:

- Mỗi liên hệ tuyến tính của X và Y: $\beta_0 + \beta_1 x$;
- Tác động của các yếu tố khác (không phải X): thành phần sai số Varepsilon.

5. Phân tích

thặng dư

6. Tương quan

XSTK

N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPBN

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

Các giả định về sai số ngẫu nhiên

- Các sai số ngẫu nhiên Varepsilon_i trong mô hình (2) được giả sử thỏa các điều kiện sau:
 - Các sai số Varepsilon_i độc lập với nhau,
 - $\mathbb{E}(\text{Varepsilon}_i) = 0$ và $\mathbb{V}ar(\text{Varepsilon}_i) = \sigma^2$,
 - Các sai số có phân phối chuẩn: $\text{Varepsilon}_i \sim \mathcal{N}(0, \sigma^2)$ và có phương sai không đổi.
- Với quan trắc x đã biết,

$$\mathbb{E}(Y|x) = \beta_0 + \beta_1 x \quad (3)$$

- Từ (??) ta có

$$Y \sim \mathcal{N}(\beta_0 + \beta_1 x, \sigma^2) \quad (4)$$

5. Phân tích

thặng dư

XSTK

N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPBN

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

5. Phân tích thặng dư

6. Tương quan

Ước lượng các hệ số hồi quy

- Gọi $\hat{\beta}_1$ và $\hat{\beta}_0$ là các ước lượng của β_0 và β_1 .
- Đường thẳng hồi quy với các hệ số ước lượng (fitted regression line):
$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x \quad (5)$$
- Một đường thẳng ước lượng tốt phải "gần với các điểm dữ liệu".
- Tìm $\hat{\beta}_0$ và $\hat{\beta}_1$: dùng phương pháp bình phương bé nhất (method of least squares).

XSTK

N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPBN

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

5. Phân tích thặng dư

6. Tương quan

Ước lượng bình phương bé nhất

Từ (??), lấy đạo hàm theo β_0 và β_1 và giải hệ phương trình, ta tìm được các ước lượng BPBN của β_0 và β_1 là

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - \frac{(\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n}}{\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}} = \frac{S_{xy}}{S_{xx}} \quad (8)$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad (9)$$

với S_{xx} và S_{xy} xác định bởi

$$S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n} \quad (10)$$

$$S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n x_i y_i - \frac{(\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n} \quad (11)$$

XSTK

N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPBN

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

5. Phân tích thặng dư

6. Tương quan

Phương pháp bình phương bé nhất (PPBPBN)

Với dữ liệu $(x_i, y_i), i = 1, \dots, n$, ta có

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

Ta định nghĩa thặng dư (residual) thứ i :

$$e_i = y_i - \hat{y}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) \quad (6)$$

Tổng bình phương sai số (Sum of Squares for Errors - SSE) hay tổng bình phương thặng dư cho n điểm dữ liệu được định nghĩa như sau

$$SSE = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2 \quad (7)$$

Nội dung của PPBPBN là tìm các ước lượng $\hat{\beta}_0$ và $\hat{\beta}_1$ sao cho SSE đạt giá trị bé nhất.

XSTK

N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPBN

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

5. Phân tích thặng dư

6. Tương quan

Ước lượng bình phương bé nhất

- Các ước lượng $\hat{\beta}_0$ và $\hat{\beta}_1$ tìm được gọi là các ước lượng BPBN.
- Đường thẳng $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ gọi là đường thẳng BPBN, thỏa các tính chất sau:
$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (1)$$

đạt giá trị bé nhất,

$$SE = \sum_{i=1}^n (y_i - \hat{y}_i) = \sum_{i=1}^n e_i = 0 \quad (2)$$

với SE là tổng các thặng dư (Sum of Errors).

XSTK

N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPBN

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

Ví dụ :

Chi phí (triệu đồng) nghiên cứu và phát triển sản phẩm (R&D) và doanh số (tỉ đồng) của 10 công ty trong ngành công nghiệp như sau:

Công ty	A	B	C	D	E	G	H	L	M	N
R&D	280	300	250	400	500	150	600	300	200	250
Doanh số	8	5	4	12	20	3	18	6	4	5

Tìm phương trình đường thẳng hồi quy biểu diễn mối liên hệ giữa chi phí (X) và doanh số (Y). Giải thích ý nghĩa của hệ số góc.

5. Phân tích
thặng dư

6.Tương quan

XSTK

N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPBN

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

Giải:

- $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{3230}{10} = 323$; $\bullet \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = \frac{85}{10} = 8,5$.
- $\bullet \sum_{i=1}^n x_i y_i = 34640$ và $\bullet \sum_{i=1}^n x_i^2 = 1215900$.
- $\bullet \hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n (\bar{x})^2} = \frac{34640 - 10 * 323 * 8,5}{1215900 - 10 * (323)^2} = 0,0416$.
- $\bullet \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 8,5 - 0,0416 * 323 = -4,9368$.

Như vậy, phương trình đường thẳng hồi quy biểu diễn mối liên hệ giữa chi phí (X) và doanh số (Y) là $\hat{Y} = -4,9368 + 0,0416X$. Với hệ số góc $\hat{\beta}_1 = 0,0416$ (hệ số góc dương) nên ta có thể nói rằng khi chi phí R&D tăng 1 triệu đồng thì doanh số tăng 0,0416 tỉ đồng.

5. Phân tích

XSTK

N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPBN

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

Giải:

Giả sử, chi phí (triệu đồng) nghiên cứu và phát triển sản phẩm (R&D) và doanh số (tỉ đồng) có mối liên hệ tuyến tính. Khi đó, mô hình hồi quy tuyến tính đơn thể hiện mối liên hệ tuyến tính giữa chi phí X và doanh số Y là

$$Y = \beta_0 + \beta_1 X + \text{Varepsilon}.$$

Dùng phương pháp bình phương bé nhất (OLS) để ước lượng β_0 và β_1 từ mẫu dữ liệu đã cho, ta có được phương trình đường thẳng hồi quy biểu diễn mối liên hệ giữa chi phí (X) và doanh số (Y) như sau:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$$

trong đó, $\hat{\beta}_0$ và $\hat{\beta}_1$ được tính như sau:

5. Phân tích
thặng dư

6.Tương quan

XSTK

N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPBN

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

Ví dụ khác

Một nhà thực vật học khảo sát mối liên hệ giữa tổng diện tích bề mặt (đv: cm²) của các lá cây đậu nành và trọng lượng khô (đv: g) của các cây này. Nhà thực vật học trồng 13 cây trong nhà kính và đo tổng diện tích lá và trọng lượng của các cây này sau 16 ngày trồng, kết quả cho bởi bảng sau

X	411	550	471	393	427	431	492	371	470	419	407	489	439
Y	2.00	2.46	2.11	1.89	2.05	2.30	2.46	2.06	2.25	2.07	2.17	2.32	2.12

(a) Vẽ biểu đồ phân tán biểu diễn diện tích lá X và trọng lượng khô Y của cây đậu nành với mẫu quan sát đã cho.

(b) Tìm đường thẳng hồi quy biểu diễn mối liên hệ giữa trọng lượng cây Y theo diện tích lá X. Vẽ đường thẳng hồi quy tìm được trên đồ thị phân tán.

5. Phân tích

XSTK
N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPBN

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

Độ đo sự biến thiên của dữ liệu

Gọi

- **SST**: Tổng bình phương toàn phần (Total Sum of Squares)

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

- **SSR**: Tổng bình phương hồi quy (Regression Sum of Squares)

$$SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

- **SSE**: Tổng bình phương sai số (Error Sum of Squares)

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

XSTK
N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPBN

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

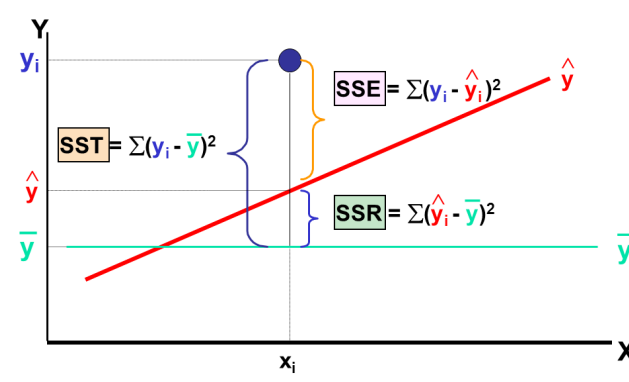
3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

Độ đo sự biến thiên của dữ liệu



XSTK
N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPBN

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

Độ đo sự biến thiên của dữ liệu

- **SST**: đo sự biến thiên của các giá trị y_i xung quanh giá trị trung tâm của dữ liệu \bar{y} ,
- **SSR**: giải thích sự biến thiên liên quan đến mối quan hệ tuyến tính của X và Y ,
- **SSE**: giải thích sự biến thiên của các nhân tố khác (không liên quan đến mối quan hệ tuyến tính của X và Y).

Ta có:

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (12)$$
$$SST = SSR + SSE$$

XSTK
N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPBN

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

Hệ số xác định

Hệ số xác định (Coefficient of Determination) là tỷ lệ của tổng sự biến thiên trong biến phụ thuộc gây ra bởi sự biến thiên của các biến độc lập (biến giải thích) so với tổng sự biến thiên toàn phần. Hệ số xác định thường được gọi là R - bình phương (R-squared), ký hiệu là R^2 .

Công thức tính:

$$R^2 = \frac{SSR}{SST} \quad (13)$$

Chú ý: $0 \leq R^2 \leq 1$.

- Hệ số xác định của một mô hình hồi quy cho phép ta đánh giá mô hình tìm được có giải thích tốt cho mối liên hệ giữa biến phụ thuộc Y và biến phụ thuộc X hay không.

XSTK

N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPBN

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

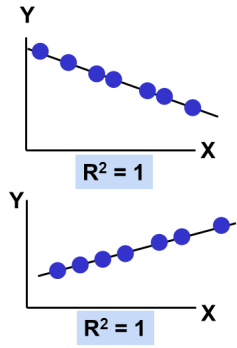
3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

Hệ số xác định và mối liên hệ giữa X và Y



$R^2 = 1$

X và Y có mối liên hệ tuyến tính mạnh:

100% sự biến thiên của Y được giải thích bởi sự biến thiên của X

5. Phân tích thặng dư

6. Tương quan

XSTK

N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPBN

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

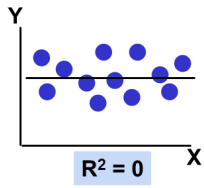
3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

Hệ số xác định và mối liên hệ giữa X và Y



$R^2 = 0$

Không có mối liên hệ tuyến tính giữa X và Y:

Giá trị của Y không phụ thuộc vào X. (Không có sự biến thiên nào của Y được giải thích bởi sự biến thiên của X)

5. Phân tích thặng dư

6. Tương quan

XSTK

N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPBN

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

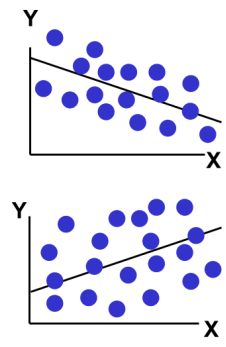
3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

Hệ số xác định và mối liên hệ giữa X và Y



$0 < R^2 < 1$

X và Y có mối liên hệ tuyến tính yếu:

Một vài nhưng không phải tất cả sự biến thiên trong Y được giải thích bởi sự biến thiên trong X

5. Phân tích thặng dư

6. Tương quan

XSTK

N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPBN

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

Ước lượng phương sai σ^2 của sai số

Xét mô hình

$$Y_i = \beta_0 + \beta_1 x_i + \text{Varepsilon}_{i}, \quad i = 1, \dots, n$$

Thành phần sai số thứ i : $\text{Varepsilon}_{i} \sim \mathcal{N}(0, \sigma^2)$. Ta cần ước lượng phương sai σ^2 .

Từ (??), ta có: $Y_i \sim \mathcal{N}(\beta_0 + \beta_1 x_i, \sigma^2)$. Do đó,

$$\frac{Y_i - (\beta_0 + \beta_1 x_i)}{\sigma} \sim \mathcal{N}(0, 1)$$

Ta có,

$$\sum_{i=1}^n \frac{[Y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2}{\sigma^2} = \frac{SSE}{\sigma^2} \sim \chi^2(n-2)$$

Nên,

$$\mathbb{E} \left[\frac{SSE}{\sigma^2} \right] = n-2 \quad \text{hay} \quad \mathbb{E} \left[\frac{SSE}{n-2} \right] = \sigma^2$$

5. Phân tích thặng dư

6. Tương quan

XSTK

N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPNB

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

5. Phân tích thặng dư

6. Tương quan

Ước lượng phương sai σ^2 của sai số

Ta kết luận rằng $\frac{SSE}{n-2}$ là một ước lượng không chệch cho σ^2 . Suy ra ước lượng $\hat{\sigma}^2$ của σ^2 được tính bởi

$$\hat{\sigma}^2 = \frac{SSE}{n-2} \quad (14)$$

Trung bình bình phương sai số (Mean Squares Error – MSE) của mô hình hồi quy tuyến tính đơn:

$$MSE = \frac{SSE}{n-2}$$

Trung bình bình phương sai số chính là ước lượng không chệch cho phương sai σ^2 của thành phần sai số của mô hình.

XSTK

N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPNB

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

5. Phân tích thặng dư

6. Tương quan

Tính chất của các ước lượng BPNB

Định lý

Xét $Y = \beta_0 + \beta_1 x + \text{Varepsilon}$ là một mô hình hồi quy tuyến tính đơn với $\text{Varepsilon} \sim \mathcal{N}(0, \sigma^2)$; với n quan trắc độc lập $y_i, i = 1, \dots, n$ ta có tương ứng các sai số Varepsilon_i . Gọi $\hat{\beta}_0$ và $\hat{\beta}_1$ là các ước lượng của β_0 và β_1 tìm được từ phương pháp bình phương bé nhất, khi đó

(a) $\hat{\beta}_0$ và $\hat{\beta}_1$ tuân theo luật phân phối chuẩn.

(b) Kỳ vọng và phương sai của $\hat{\beta}_0$ và $\hat{\beta}_1$ lần lượt là

$$\mathbb{E}(\hat{\beta}_0) = \beta_0, \text{Var}(\hat{\beta}_0) = \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right) \sigma^2, \quad (15)$$

$$\mathbb{E}(\hat{\beta}_1) = \beta_1, \text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}} \quad (16)$$

XSTK

N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPNB

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

5. Phân tích thặng dư

6. Tương quan

Ước lượng phương sai σ^2 của sai số

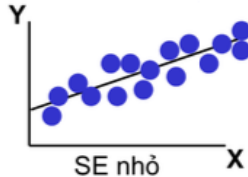
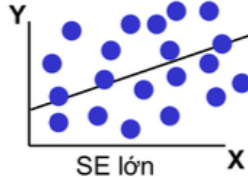
• Tìm SSE:

$$SSE = SST - \hat{\beta}_1 S_{xy}$$

• Sai số chuẩn (Standard Error) của $\hat{\sigma}^2$

$$SE(\hat{\sigma}) = \sqrt{\frac{SSE}{n-2}}$$

Sử dụng $SE(\hat{\sigma})$ để đo sự biến thiên của các giá trị quan trắc y với đường thẳng hồi quy.

XSTK

N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPNB

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

5. Phân tích thặng dư

6. Tương quan

Tính chất của các ước lượng BPNB

Trong mô hình hồi quy tuyến tính đơn, sai số chuẩn (SE) của các ước lượng $\hat{\beta}_0$ và $\hat{\beta}_1$ là

$$SE(\hat{\beta}_0) = \sqrt{\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right) \hat{\sigma}^2} \quad (17)$$

$$SE(\hat{\beta}_1) = \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} \quad (18)$$

XSTK

N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPBN

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

5. Phân tích thặng dư

6. Tương quan

Tính chất của các ước lượng BPBN

Định lý Gauss - Markov

Xét mô hình hồi quy tuyến tính đơn

$$Y = \beta_0 + \beta_1 x + \text{Varepsilon}$$

có $\hat{\beta}_0$ và $\hat{\beta}_1$ là các ước lượng BPBN cho β_0 và β_1 , khi đó $\hat{\beta}_0$ và $\hat{\beta}_1$ là các ước lượng không chệch tốt nhất.

XSTK

N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPBN

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

5. Phân tích thặng dư

6. Tương quan

Khoảng tin cậy cho hệ số hồi quy

- Đặt

$$MSE = \frac{SSE}{n-2} = \frac{1}{n-2} \sum_{i=1}^n \left[y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) \right]^2$$

MSE gọi là trung bình bình phương sai số (Mean Square Error). Khi đó

$$T_{\beta_1} = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{MSE}{S_{xx}}}} \tag{19}$$

T_{β_1} có phân phối Student với $n - 2$ bậc tự do.

XSTK

N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPBN

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

5. Phân tích thặng dư

6. Tương quan

3. 1 Khoảng tin cậy cho hệ số hồi quy

- Xét đường thẳng hồi quy:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

- Vì $\hat{\beta}_1 \sim \mathcal{N}\left(\beta_1, \frac{\sigma^2}{S_{xx}}\right)$, đặt

$$Z_1 = \frac{\hat{\beta}_1 - \beta_1}{\sigma / \sqrt{S_{xx}}} \sim \mathcal{N}(0, 1)$$

- Do $\frac{SSE}{\sigma^2}$ độc lập với $\hat{\beta}_1$ và $\frac{SSE}{\sigma^2} \sim \chi^2(n-2)$ nên

$$T_{\beta_1} = \frac{Z_1}{\sqrt{\frac{(SSE/\sigma^2)}{n-2}}} \sim t(n-2)$$

XSTK

N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPBN

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

5. Phân tích thặng dư

6. Tương quan

Khoảng tin cậy cho hệ số hồi quy

- Tương tự, vì $\hat{\beta}_0 \sim \mathcal{N}\left(\beta_0, \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right) \sigma^2\right)$, đặt

$$Z_0 = \frac{\hat{\beta}_0 - \beta_0}{\sigma \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}} \sim \mathcal{N}(0, 1)$$

- Do $\hat{\beta}_0$ và SSE độc lập nên ta có

$$T_{\beta_0} = \frac{Z_0}{\sqrt{\frac{(SSE/\sigma^2)}{n-2}}} = \frac{\hat{\beta}_0 - \beta_0}{\sqrt{MSE \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)}} \tag{20}$$

T_{β_0} có phân phối Student với $n - 2$ bậc tự do.

XSTK
N.T. M. Ngọc

1. Giới thiệu
2. Mô hình hồi quy tuyến tính đơn
2.1 Mô hình hồi quy tuyến tính đơn
2.2 Ước lượng các hệ số hồi quy
2.3 Độ đo sự biến thiên của dữ liệu
2.4 Hệ số xác định
Ước lượng phương sai σ^2 của sai số
2.5 Tính chất của các ước lượng BPBN
3. Khoảng tin cậy cho mô hình hồi quy
3.1 Khoảng tin cậy cho hệ số hồi quy
3.2 Khoảng tin cậy cho trung bình biến đáp ứng
3.3 Dự đoán giá trị quan trắc mới
4. Kiểm định giả thuyết cho mô hình hồi quy
4.1 Kiểm định giả thuyết cho β_0
4.2 Kiểm định giả thuyết cho β_1

Khoảng tin cậy cho hệ số hồi quy

- Khoảng tin cậy $100(1 - \alpha)\%$ cho β_1 :
$$\hat{\beta}_1 - t_{1-\alpha/2}^{n-2} \sqrt{\frac{MSE}{S_{xx}}} \leq \beta_1 \leq \hat{\beta}_1 + t_{1-\alpha/2}^{n-2} \sqrt{\frac{MSE}{S_{xx}}} \quad (21)$$
- Khoảng tin cậy $100(1 - \alpha)\%$ cho β_0 :
$$\hat{\beta}_0 - t_{1-\alpha/2}^{n-2} \sqrt{MSE \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)} \leq \beta_0 \leq \hat{\beta}_0 + t_{1-\alpha/2}^{n-2} \sqrt{MSE \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)} \quad (22)$$

với

- n = số cặp giá trị quan trắc (x_i, y_i) ;
- $t_{1-\alpha/2}^{n-2}$ là phân vị mức $1 - \alpha/2$ của biến ngẫu nhiên $t(n - 2)$.

5. Phân tích thặng dư
6. Tương quan

XSTK
N.T. M. Ngọc

1. Giới thiệu
2. Mô hình hồi quy tuyến tính đơn
2.1 Mô hình hồi quy tuyến tính đơn
2.2 Ước lượng các hệ số hồi quy
2.3 Độ đo sự biến thiên của dữ liệu
2.4 Hệ số xác định
Ước lượng phương sai σ^2 của sai số
2.5 Tính chất của các ước lượng BPBN
3. Khoảng tin cậy cho mô hình hồi quy
3.1 Khoảng tin cậy cho hệ số hồi quy
3.2 Khoảng tin cậy cho trung bình biến đáp ứng
3.3 Dự đoán giá trị quan trắc mới
4. Kiểm định giả thuyết cho mô hình hồi quy
4.1 Kiểm định giả thuyết cho β_0
4.2 Kiểm định giả thuyết cho β_1

Khoảng tin cậy cho trung bình biến đáp ứng

- Ta có
$$\frac{\hat{\mu}_{Y|x_0} - \mathbb{E}(\hat{\mu}_{Y|x_0})}{\sqrt{\text{Var}(\hat{\mu}_{Y|x_0})}} = \frac{(\hat{\beta}_0 + \hat{\beta}_1 x_0) - (\beta_0 + \beta_1 x_0)}{\sigma \sqrt{\frac{1}{n} + \frac{(\bar{x} - x_0)^2}{S_{xx}}}} \sim \mathcal{N}(0, 1)$$
- Vì $\hat{\mu}_{Y|x_0}$ độc lập với $SSE/\sigma^2 \sim \chi^2(n - 2)$ nên
$$\frac{\hat{\mu}_{Y|x_0} - (\beta_0 + \beta_1 x_0)}{\sqrt{\frac{1}{n} + \frac{(\bar{x} - x_0)^2}{S_{xx}}} \sqrt{\frac{SSE}{n-2}}} = \frac{\hat{\mu}_{Y|x_0} - (\beta_0 + \beta_1 x_0)}{\sqrt{MSE \left[\frac{1}{n} + \frac{(\bar{x} - x_0)^2}{S_{xx}} \right]}} \sim t(n - 2) \quad (23)$$

5. Phân tích thặng dư

XSTK
N.T. M. Ngọc

1. Giới thiệu
2. Mô hình hồi quy tuyến tính đơn
2.1 Mô hình hồi quy tuyến tính đơn
2.2 Ước lượng các hệ số hồi quy
2.3 Độ đo sự biến thiên của dữ liệu
2.4 Hệ số xác định
Ước lượng phương sai σ^2 của sai số
2.5 Tính chất của các ước lượng BPBN
3. Khoảng tin cậy cho mô hình hồi quy
3.1 Khoảng tin cậy cho hệ số hồi quy
3.2 Khoảng tin cậy cho trung bình biến đáp ứng
3.3 Dự đoán giá trị quan trắc mới
4. Kiểm định giả thuyết cho mô hình hồi quy
4.1 Kiểm định giả thuyết cho β_0
4.2 Kiểm định giả thuyết cho β_1

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

- Cho trước giá trị x_0 , cần tìm khoảng tin cậy cho $\mu_{Y|x_0} = \mathbb{E}(Y|x_0) = \beta_0 + \beta_1 x_0$, gọi là trung bình biến đáp ứng. Ước lượng của $\mu_{Y|x_0}$ từ đường thẳng hồi quy là
$$\hat{\mu}_{Y|x_0} = \hat{\beta}_0 + \hat{\beta}_1 x_0$$
- $\hat{\mu}_{Y|x_0}$ có các tính chất sau
 - $\hat{\mu}_{Y|x_0}$ tuân theo luật phân phối chuẩn.
 - Kỳ vọng và phương sai của $\hat{\mu}_{Y|x_0}$ lần lượt là
$$\mathbb{E}(\hat{\mu}_{Y|x_0}) = \beta_0 + \beta_1 x_0$$
$$\text{Var}(\hat{\mu}_{Y|x_0}) = \left[\frac{1}{n} + \frac{(\bar{x} - x_0)^2}{S_{xx}} \right] \sigma^2$$

5. Phân tích thặng dư
6. Tương quan

XSTK
N.T. M. Ngọc

1. Giới thiệu
2. Mô hình hồi quy tuyến tính đơn
2.1 Mô hình hồi quy tuyến tính đơn
2.2 Ước lượng các hệ số hồi quy
2.3 Độ đo sự biến thiên của dữ liệu
2.4 Hệ số xác định
Ước lượng phương sai σ^2 của sai số
2.5 Tính chất của các ước lượng BPBN
3. Khoảng tin cậy cho mô hình hồi quy
3.1 Khoảng tin cậy cho hệ số hồi quy
3.2 Khoảng tin cậy cho trung bình biến đáp ứng
3.3 Dự đoán giá trị quan trắc mới
4. Kiểm định giả thuyết cho mô hình hồi quy
4.1 Kiểm định giả thuyết cho β_0
4.2 Kiểm định giả thuyết cho β_1

Khoảng tin cậy cho trung bình biến đáp ứng

- Khoảng tin cậy $100(1 - \alpha)\%$ cho trung bình biến đáp ứng là
$$\hat{\mu}_{Y|x_0} - t_{1-\alpha/2}^{n-2} \sqrt{MSE \left(\frac{1}{n} + \frac{(\bar{x} - x_0)^2}{S_{xx}} \right)} \leq \mu_{Y|x_0} \leq \hat{\mu}_{Y|x_0} + t_{1-\alpha/2}^{n-2} \sqrt{MSE \left[\frac{1}{n} + \frac{(\bar{x} - x_0)^2}{S_{xx}} \right]} \quad (24)$$

với

- $\hat{\mu}_{Y|x_0} = \hat{\beta}_0 + \hat{\beta}_1 x_0$,
- $t_{1-\alpha/2}^{n-2}$: phân vị mức $1 - \alpha/2$ của biến ngẫu nhiên T tuân theo phân phối Student với bậc tự do $(n - 2)$,
- $MSE = SSE/(n - 2)$: trung bình bình phương sai số.

5. Phân tích thặng dư

XSTK

N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPBN

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

5. Phân tích thặng dư

6. Tương quan

Dự đoán giá trị quan trắc mới

Giả sử với giá trị x_0 , ta cần dự đoán giá trị quan trắc Y_0 trong tương lai tương ứng với x_0 bằng bao nhiêu. Từ mô hình hồi quy, ta có

$\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$

(25)

\hat{Y}_0 là một ước lượng điểm của giá trị quan trắc mới Y_0 .

Cần tìm khoảng tin cậy cho Y_0 .

Cho trước giá trị x_0 , cần phân biệt rõ khoảng tin cậy giữa trung bình của biến ngẫu nhiên Y là $\mu_{Y|x_0}$ và khoảng tin cậy của giá trị quan trắc thực sự của Y tương ứng với x_0 .

XSTK

N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPBN

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

5. Phân tích thặng dư

6. Tương quan

Dự đoán giá trị quan trắc mới

Đặt

$\eta = Y_0 - \hat{Y}_0$

Vì Y_0 và \hat{Y}_0 có phân phối chuẩn nên η có phân phối chuẩn với kỳ vọng và phương sai là

$\mathbb{E}(\eta) = \mathbb{E}(Y_0) - \mathbb{E}(\hat{Y}_0) = 0$

$\text{Var}(\eta) = \text{Var}(Y_0) + \text{Var}(\hat{Y}_0) = \sigma^2 + \sigma^2 \left[\frac{1}{n} + \frac{(\bar{x} - x_0)^2}{S_{xx}} \right]$

$= \left[1 + \frac{1}{n} + \frac{(\bar{x} - x_0)^2}{S_{xx}} \right] \sigma^2$

Do đó,

$\eta \sim \mathcal{N} \left(0, \left[1 + \frac{1}{n} + \frac{(\bar{x} - x_0)^2}{S_{xx}} \right] \sigma^2 \right)$

XSTK

N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPBN

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

5. Phân tích thặng dư

6. Tương quan

Dự đoán giá trị quan trắc mới

Khoảng tin cậy cho giá trị kỳ vọng của y , cho trước x_0

$\hat{y} = \beta_0 + \beta_1 x_i$

Khoảng dự đoán cho một giá trị quan trắc y , cho trước x_0

XSTK

N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPBN

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

5. Phân tích thặng dư

6. Tương quan

Dự đoán giá trị quan trắc mới

Và,

$Z = \frac{\hat{Y}_0 - Y_0}{\sqrt{\sigma^2 \left[1 + \frac{1}{n} + \frac{(\bar{x} - x_0)^2}{S_{xx}} \right]}} \sim \mathcal{N}(0, 1)$

Nếu ta thay thế σ^2 bởi $\hat{\sigma}^2 = MSE = \frac{SSE}{n-2}$, thu được

$T = \frac{\hat{Y}_0 - Y_0}{\sqrt{MSE \left[1 + \frac{1}{n} + \frac{(\bar{x} - x_0)^2}{S_{xx}} \right]}} \sim t(n-2)$

(26)

$$\hat{y} = 60 + 5x.$$

XSTK

N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPBN

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

5. Phân tích thặng dư

6. Tương quan

Gợi ý giải:

Với hệ số góc $\beta_1 = 5 > 0$, có nghĩa là khi số sinh viên tăng lên, doanh số cũng tăng lên. Chúng ta có thể kết luận rằng, sự gia tăng số lượng sinh viên thêm 1 đơn vị (1000 người) sẽ tương ứng với sự gia tăng doanh số kỳ vọng là 5 000 đô la.

XSTK

N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPBN

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

5. Phân tích thặng dư

6. Tương quan

Ví dụ khác

Xét mẫu ngẫu nhiên gồm 10 cặp giá trị (x_i, y_i) cho bởi bảng

x	-1	0	2	-2	5	6	8	11	12	-3
y	-5	-4	2	-7	6	9	13	21	20	-9

(a) Vẽ biểu đồ phân tán cho dữ liệu, tìm đường thẳng hồi quy.

(b) Tìm ước lượng $\hat{\sigma}^2$ cho phương sai σ^2 của sai số ngẫu nhiên.

(c) Thiết lập khoảng tin cậy 95% cho các hệ số β_0 và β_1 .

(d) Thiết lập khoảng dự đoán 95% tại $x = 5$.

XSTK

N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPBN

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

5. Phân tích thặng dư

6. Tương quan

Gợi ý giải:

Với hệ số góc $\beta_1 = 5 > 0$, có nghĩa là khi số sinh viên tăng lên, doanh số cũng tăng lên. Chúng ta có thể kết luận rằng, sự gia tăng số lượng sinh viên thêm 1000 người (1 đơn vị) sẽ tương ứng với sự gia tăng doanh số kỳ vọng là 5 000 đô la.

Chúng ta muốn dự đoán doanh số bán theo quý cho một nhà hàng được đặt gần trường với 16 000 sinh viên, chúng ta tính toán: Ta có, $x_{n+1} = 16$, nên ta thay vào phương trình đường thẳng hồi quy tuyến tính (ước lượng) vừa tìm ta được:

$$\hat{y}_{n+1} = \hat{\beta}_0 + \hat{\beta}_1 x_{n+1} = 60 + 5x_{n+1} = 60 + 5 * 16 = 140$$

Như vậy, chúng ta dự đoán doanh số bán theo quý cho một nhà hàng được đặt gần trường với 16 000 sinh viên là 140 000 đô la.

XSTK

N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPBN

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

5. Phân tích thặng dư

6. Tương quan

Kiểm định giả thuyết cho β_0

• Bài toán kiểm định giả thuyết cho hệ số chặn β_0 trong mô hình hồi quy tuyến tính đơn gồm các trường hợp sau:

$(a) \begin{cases} H_0 : \beta_0 = b_0 \\ H_1 : \beta_0 \neq b_0 \end{cases}$

$(b) \begin{cases} H_0 : \beta_0 = b_0 \\ H_1 : \beta_0 < b_0 \end{cases}$

$(c) \begin{cases} H_0 : \beta_0 = b_0 \\ H_1 : \beta_0 > b_0 \end{cases}$

với giá trị b_0 và mức ý nghĩa α cho trước. Thông thường $b_0 = 0$.

XSTK
N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPNB

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

Kiểm định giả thuyết cho β_0

Các bước kiểm định

1 Phát biểu giả thuyết H_0 và đối thuyết,

2 Xác định mức ý nghĩa α ,

3 Tính giá trị thống kê kiểm định:

$$t_{\beta_0} = \frac{\hat{\beta}_0 - b_0}{\sqrt{MSE \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)}}$$

XSTK
N.T. M. Ngọc

5. Phân tích thặng dư

6. Tương quan

XSTK
N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPNB

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

Kiểm định giả thuyết cho β_1

Bài toán kiểm định giả thuyết cho hệ số góc β_1 trong mô hình hồi quy tuyến tính đơn gồm các trường hợp sau:

$(a) \begin{cases} H_0 : \beta_1 = b_1 \\ H_1 : \beta_1 \neq b_1 \end{cases}$

$(b) \begin{cases} H_0 : \beta_1 = b_1 \\ H_1 : \beta_1 < b_1 \end{cases}$

$(c) \begin{cases} H_0 : \beta_1 = b_1 \\ H_1 : \beta_1 > b_1 \end{cases}$

với giá trị b_1 và mức ý nghĩa α cho trước. Thông thường $b_1 = 0$.

XSTK
N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPNB

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

Kiểm định giả thuyết cho β_0

4. Xác định miền bác bỏ

Đối thuyết

Miền bác bỏ

p - giá trị

$H_1 : \beta_0 \neq b_0$

$|t_{\beta_0}| > t_{1-\alpha/2}^{n-2}$

$p = 2\mathbb{P}(T_{n-2} \geq |t_{\beta_0}|)$

$H_1 : \beta_0 < b_0$

$t_{\beta_0} < -t_{1-\alpha}^{n-2}$

$p = \mathbb{P}(T_{n-2} \leq t_{\beta_0})$

$H_1 : \beta_0 > b_0$

$t_{\beta_0} > t_{1-\alpha}^{n-2}$

$p = \mathbb{P}(T_{n-2} \geq t_{\beta_0})$

5. Kết luận: Bác bỏ H_0 /Chưa đủ cơ sở để bác bỏ H_0 .

XSTK
N.T. M. Ngọc

5. Phân tích thặng dư

6. Tương quan

XSTK
N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPNB

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

Kiểm định giả thuyết cho β_1

Các bước kiểm định

1 Phát biểu giả thuyết H_0 và đối thuyết,

2 Xác định mức ý nghĩa α ,

3 Tính thống kê kiểm định:

$$t_{\beta_1} = \frac{\hat{\beta}_1 - b_1}{\sqrt{\frac{MSE}{S_{xx}}}}$$

XSTK
N.T. M. Ngọc

1. Giới thiệu
2. Mô hình hồi quy tuyến tính đơn
2.1 Mô hình hồi quy tuyến tính đơn
2.2 Ước lượng các hệ số hồi quy
2.3 Độ đo sự biến thiên của dữ liệu
2.4 Hệ số xác định
Ước lượng phương sai σ^2 của sai số
2.5 Tính chất của các ước lượng BPNB
3. Khoảng tin cậy cho mô hình hồi quy
3.1 Khoảng tin cậy cho hệ số hồi quy
3.2 Khoảng tin cậy cho trung bình biến đáp ứng
3.3 Dự đoán giá trị quan trắc mới
4. Kiểm định giả thuyết cho mô hình hồi quy
4.1 Kiểm định giả thuyết cho β_0
4.2 Kiểm định giả thuyết cho β_1

Kiểm định giả thuyết cho β_1

4. Xác định miền bác bỏ

Đối thuyết	Miền bác bỏ	p - giá trị
$H_1: \beta_1 \neq b_1$	$ t_{\beta_1} > t_{1-\alpha/2}^{n-2}$	$p = 2\mathbb{P}(T_{n-2} \geq t_{\beta_1})$
$H_1: \beta_1 < b_1$	$t_{\beta_1} < -t_{1-\alpha}^{n-2}$	$p = \mathbb{P}(T_{n-2} \leq t_{\beta_1})$
$H_1: \beta_1 > b_1$	$t_{\beta_1} > t_{1-\alpha}^{n-2}$	$p = \mathbb{P}(T_{n-2} \geq t_{\beta_1})$

5. Kết luận: Bác bỏ H_0 /Chưa đủ cơ sở để bác bỏ H_0 .

XSTK
N.T. M. Ngọc

1. Giới thiệu
2. Mô hình hồi quy tuyến tính đơn
2.1 Mô hình hồi quy tuyến tính đơn
2.2 Ước lượng các hệ số hồi quy
2.3 Độ đo sự biến thiên của dữ liệu
2.4 Hệ số xác định
Ước lượng phương sai σ^2 của sai số
2.5 Tính chất của các ước lượng BPNB
3. Khoảng tin cậy cho mô hình hồi quy
3.1 Khoảng tin cậy cho hệ số hồi quy
3.2 Khoảng tin cậy cho trung bình biến đáp ứng
3.3 Dự đoán giá trị quan trắc mới
4. Kiểm định giả thuyết cho mô hình hồi quy
4.1 Kiểm định giả thuyết cho β_0
4.2 Kiểm định giả thuyết cho β_1

Phân tích thặng dư

• Với n cặp quan trắc (x_i, y_i) , gọi $e_i = y_i - \hat{y}_i$ là thặng dư thứ i . Để kiểm tra các giả định của mô hình có thỏa hay không, nhận xét trên đồ thị của các giá trị thặng dư.

• Nếu các giả định của mô hình hồi quy thỏa, thì:

- Đồ thị của các giá trị thặng dư tương ứng với các giá trị biến độc lập X phải có hình dạng thô, không đều (roughly), trải dọc theo chiều ngang và đối xứng qua trục Ox .
- Đồ thị xác suất chuẩn (Normal Probability plot) của các giá trị thặng dư phải có dạng gần tuyến tính (roughly linear).

XSTK
N.T. M. Ngọc

1. Giới thiệu
2. Mô hình hồi quy tuyến tính đơn
2.1 Mô hình hồi quy tuyến tính đơn
2.2 Ước lượng các hệ số hồi quy
2.3 Độ đo sự biến thiên của dữ liệu
2.4 Hệ số xác định
Ước lượng phương sai σ^2 của sai số
2.5 Tính chất của các ước lượng BPNB
3. Khoảng tin cậy cho mô hình hồi quy
3.1 Khoảng tin cậy cho hệ số hồi quy
3.2 Khoảng tin cậy cho trung bình biến đáp ứng
3.3 Dự đoán giá trị quan trắc mới
4. Kiểm định giả thuyết cho mô hình hồi quy
4.1 Kiểm định giả thuyết cho β_0
4.2 Kiểm định giả thuyết cho β_1

Phân tích thặng dư

• Phân tích thặng dư (Analysis of Residuals) được sử dụng để kiểm tra các giả định của mô hình hồi quy tuyến tính.

• Các giả định của mô hình:

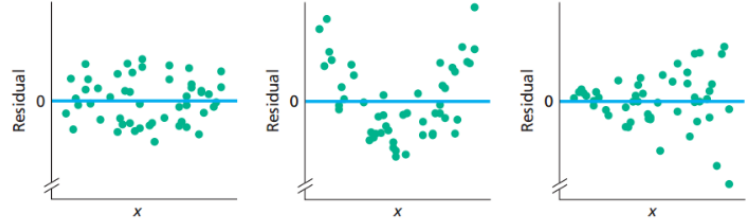
- 1 Đường thẳng hồi quy tổng thể: $Y = \beta_0 + \beta_1 x + \text{Varepsilon}$ với β_0 và β_1 là các hằng số sao cho với mỗi giá trị x , $\mathbb{E}(Y|x) = \beta_0 + \beta_1 x$.
- 2 Phương sai bằng nhau: phương sai của biến đáp ứng (biến phụ thuộc) Y bằng nhau với mọi giá trị của biến độc lập X , tức là $\text{Var}(Y|x) = \sigma^2$.
- 3 Phân phối chuẩn: với mỗi giá trị của biến độc lập, phân phối có điều kiện (cho trước giá trị x) của biến đáp ứng là phân phối chuẩn, $Y|x \sim \mathcal{N}(\beta_0 + \beta_1 x, \sigma^2)$.
- 4 Độc lập: các quan trắc của biến đáp ứng Y độc lập với nhau.

XSTK
N.T. M. Ngọc

1. Giới thiệu
2. Mô hình hồi quy tuyến tính đơn
2.1 Mô hình hồi quy tuyến tính đơn
2.2 Ước lượng các hệ số hồi quy
2.3 Độ đo sự biến thiên của dữ liệu
2.4 Hệ số xác định
Ước lượng phương sai σ^2 của sai số
2.5 Tính chất của các ước lượng BPNB
3. Khoảng tin cậy cho mô hình hồi quy
3.1 Khoảng tin cậy cho hệ số hồi quy
3.2 Khoảng tin cậy cho trung bình biến đáp ứng
3.3 Dự đoán giá trị quan trắc mới
4. Kiểm định giả thuyết cho mô hình hồi quy
4.1 Kiểm định giả thuyết cho β_0
4.2 Kiểm định giả thuyết cho β_1

Phân tích thặng dư

• Đồ thị thặng dư:



XSTK
N.T. M. Ngọc

1. Giới thiệu
2. Mô hình hồi quy tuyến tính đơn
2.1 Mô hình hồi quy tuyến tính đơn
2.2 Ước lượng các hệ số hồi quy
2.3 Độ đo sự biến thiên của dữ liệu
2.4 Hệ số xác định
Ước lượng phương sai σ^2 của sai số
2.5 Tính chất của các ước lượng BPBN
3. Khoảng tin cậy cho mô hình hồi quy
3.1 Khoảng tin cậy cho hệ số hồi quy
3.2 Khoảng tin cậy cho trung bình biến đáp ứng
3.3 Dự đoán giá trị quan trắc mới
4. Kiểm định giả thuyết cho mô hình hồi quy
4.1 Kiểm định giả thuyết cho β_0
4.2 Kiểm định giả thuyết cho β_1

Phân tích thặng dư

- Đồ thị phân vị chuẩn (Q-Q plot):

XSTK
N.T. M. Ngọc

5. Phân tích thặng dư
6. Tương quan

1. Giới thiệu
2. Mô hình hồi quy tuyến tính đơn
2.1 Mô hình hồi quy tuyến tính đơn
2.2 Ước lượng các hệ số hồi quy
2.3 Độ đo sự biến thiên của dữ liệu
2.4 Hệ số xác định
Ước lượng phương sai σ^2 của sai số
2.5 Tính chất của các ước lượng BPBN
3. Khoảng tin cậy cho mô hình hồi quy
3.1 Khoảng tin cậy cho hệ số hồi quy
3.2 Khoảng tin cậy cho trung bình biến đáp ứng
3.3 Dự đoán giá trị quan trắc mới
4. Kiểm định giả thuyết cho mô hình hồi quy
4.1 Kiểm định giả thuyết cho β_0
4.2 Kiểm định giả thuyết cho β_1

Phân tích tương quan

Hệ số tương quan (Correlation coefficient) của hai biến ngẫu nhiên X và Y , ký hiệu ρ_{XY} , được xác định như sau

$$\rho_{XY} = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}} \quad (29)$$

Với hai biến ngẫu nhiên X và Y bất kỳ

$$-1 \leq \rho_{XY} \leq 1$$

XSTK
N.T. M. Ngọc

1. Giới thiệu
2. Mô hình hồi quy tuyến tính đơn
2.1 Mô hình hồi quy tuyến tính đơn
2.2 Ước lượng các hệ số hồi quy
2.3 Độ đo sự biến thiên của dữ liệu
2.4 Hệ số xác định
Ước lượng phương sai σ^2 của sai số
2.5 Tính chất của các ước lượng BPBN
3. Khoảng tin cậy cho mô hình hồi quy
3.1 Khoảng tin cậy cho hệ số hồi quy
3.2 Khoảng tin cậy cho trung bình biến đáp ứng
3.3 Dự đoán giá trị quan trắc mới
4. Kiểm định giả thuyết cho mô hình hồi quy
4.1 Kiểm định giả thuyết cho β_0
4.2 Kiểm định giả thuyết cho β_1

Phân tích tương quan

- Phân tích tương quan (Correlation Analysis) dùng để đo độ mạnh của mối liên hệ tuyến tính giữa hai biến ngẫu nhiên.

Xét hai biến ngẫu nhiên X, Y . Hiệp phương sai (Covariance) của X và Y , ký hiệu là $Cov(X, Y)$, được định nghĩa như sau

$$Cov(X, Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y) \quad (28)$$

XSTK
N.T. M. Ngọc

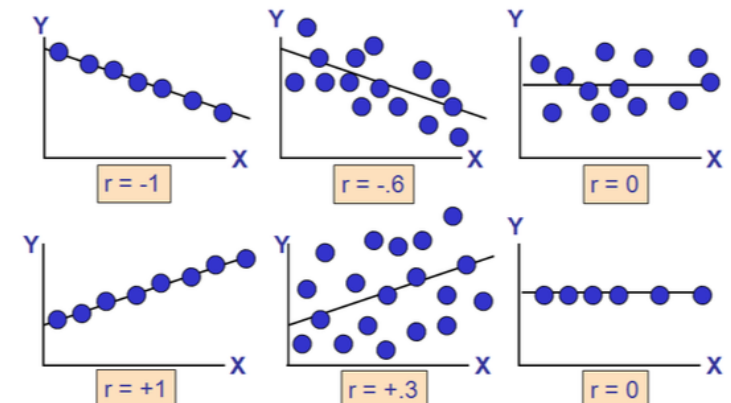
5. Phân tích thặng dư
6. Tương quan

1. Giới thiệu
2. Mô hình hồi quy tuyến tính đơn
2.1 Mô hình hồi quy tuyến tính đơn
2.2 Ước lượng các hệ số hồi quy
2.3 Độ đo sự biến thiên của dữ liệu
2.4 Hệ số xác định
Ước lượng phương sai σ^2 của sai số
2.5 Tính chất của các ước lượng BPBN
3. Khoảng tin cậy cho mô hình hồi quy
3.1 Khoảng tin cậy cho hệ số hồi quy
3.2 Khoảng tin cậy cho trung bình biến đáp ứng
3.3 Dự đoán giá trị quan trắc mới
4. Kiểm định giả thuyết cho mô hình hồi quy
4.1 Kiểm định giả thuyết cho β_0
4.2 Kiểm định giả thuyết cho β_1

Phân tích tương quan

Với mẫu ngẫu nhiên cỡ $n: (X_i, Y_i), i = 1, \dots, n$. Hệ số tương quan mẫu, ký hiệu r_{XY} , được xác định như sau

$$r_{XY} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}} = \frac{S_{XY}}{\sqrt{S_{XX}S_{YY}}} \quad (30)$$



XSTK

N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPPN

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

5. Phân tích thặng dư

6. Tương quan

Kiểm định giả thuyết cho hệ số tương quan

Ta cần kiểm định giả thuyết H_0 : không có mối liên hệ tuyến tính giữa X và Y

$$H_0 : \rho = 0$$

Tổng kê kiểm định

$$T = r_{XY} \sqrt{\frac{n-2}{1-r_{XY}^2}} \quad (31)$$

có phân phối Student với $n-2$ bậc tự do.

XSTK

N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPPN

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

5. Phân tích thặng dư

6. Tương quan

Kiểm định giả thuyết cho hệ số tương quan

Các bước kiểm định

1. Phát biểu giả thuyết H_0 và đối thuyết,

2. Xác định mức ý nghĩa α ,

3. Tính thống kê kiểm định:

$$T = r_{XY} \sqrt{\frac{n-2}{1-r_{XY}^2}}$$

T có phân phối Student với $n-2$ bậc tự do.

XSTK

N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPPN

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

5. Phân tích thặng dư

6. Tương quan

Kiểm định giả thuyết cho hệ số tương quan

Bài toán kiểm định giả thuyết cho hệ số tương quan gồm các trường hợp sau:

$$(a) \begin{cases} H_0 : \rho = 0 \\ H_1 : \rho \neq 0 \end{cases} \quad (b) \begin{cases} H_0 : \rho = 0 \\ H_1 : \rho < 0 \end{cases} \quad (c) \begin{cases} H_0 : \rho = 0 \\ H_1 : \rho > 0 \end{cases}$$

với mức ý nghĩa α cho trước.

XSTK

N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPPN

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

5. Phân tích thặng dư

6. Tương quan

Kiểm định giả thuyết cho hệ số tương quan

4. Xác định miền bác bỏ

Đối thuyết	Miền bác bỏ	p - giá trị
$H_1 : \rho \neq 0$	$ t > t_{1-\alpha/2}^{n-2}$	$p = 2\mathbb{P}(T_{n-2} \geq t)$
$H_1 : \rho < 0$	$t < -t_{1-\alpha}^{n-2}$	$p = \mathbb{P}(T_{n-2} \leq t)$
$H_1 : \rho > 0$	$t > t_{1-\alpha}^{n-2}$	$p = \mathbb{P}(T_{n-2} \geq t)$

5. Kết luận: Bác bỏ H_0 /Chưa đủ cơ sở để bác bỏ H_0 .

XSTK

N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPNB

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

5. Phân tích thặng dư

6.Tương quan

7. Mô hình hồi quy tuyến tính bội

Ví dụ

Điểm thi giữa kỳ (x) và cuối kỳ (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 kỳ của một sinh viên có điểm giữa kỳ là 85.

c

Tính $\hat{\sigma}^2 = s^2$

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 đã tìm được ở câu (??), hãy xây dựng khoảng tin cậy 95% cho $\mu_{Y|85}$.

XSTK

N.T. M. Ngọc

1. Giới thiệu

2. Mô hình hồi quy tuyến tính đơn

2.1 Mô hình hồi quy tuyến tính đơn

2.2 Ước lượng các hệ số hồi quy

2.3 Độ đo sự biến thiên của dữ liệu

2.4 Hệ số xác định

Ước lượng phương sai σ^2 của sai số

2.5 Tính chất của các ước lượng BPNB

3. Khoảng tin cậy cho mô hình hồi quy

3.1 Khoảng tin cậy cho hệ số hồi quy

3.2 Khoảng tin cậy cho trung bình biến đáp ứng

3.3 Dự đoán giá trị quan trắc mới

4. Kiểm định giả thuyết cho mô hình hồi quy

4.1 Kiểm định giả thuyết cho β_0

4.2 Kiểm định giả thuyết cho β_1

5. Phân tích thặng dư

6.Tương quan

7. Mô hình hồi quy tuyến tính bội

7. Mô hình hồi quy tuyến tính bội

Bài tập lớn