HỒI QUY VÀ TƯƠNG QUAN

Nguyễn Văn Thìn

BỘ MÔN THỐNG KÊ TOÁN HỌC
KHOA TOÁN - TIN HỌC
ĐẠI HỌC KHOA HỌC TỰ NHIÊN TP.HCM

Tháng 2 năm 2016

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

Outline

1 Giới thiệu
2 Mô hình hồi quy tuyến tính đơn
3 Khoảng tin cậy cho mô hình hồi quy
4 Kiểm định giả thuyết cho mô hình hồi quy
5 Tương quan
6 Mô hình hồi quy tuyến tính bội
7 Kiểm định giả thuyết trong mô hình hồi quy bội

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

Outline

1 Giới thiệu
2 Mô hình hồi quy tuyến tính đơn
3 Khoảng tin cậy cho mô hình hồi quy
4 Kiểm định giả thuyết cho mô hình hồi quy
5 Tương quan
6 Mô hình hồi quy tuyến tính bội
7 Kiểm định giả thuyết trong mô hình hồi quy bội

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

Nội dung chính

Mô hình hồi quy tuyến tính đơn

- Giới thiệu bài toán hồi quy và hồi quy tuyến tính đơn
- Các giả định về sai số ngẫu nhiên
- Ước lượng các hệ số hồi quy bằng phương pháp bình phương bé nhất (BPBN)
- Hệ số xác định
- Các tính chất của ước lượng BPBN
- Khoảng tin cậy cho hệ số hồi quy, trung bình biến đáp ứng
- Dự đoán giá trị quan trắc mới
- Kiểm định giả thuyết cho β_0, β_1
- Phân tích thặng dư, tương quan
- Kiểm định giả thuyết cho hệ số tương quan

Nội dung chính

Mô hình hồi quy tuyến tính bội [1]

- Biểu diễn mô hình theo dạng ma trận
- Ước lượng bình phương bé nhất
- Ước lượng phương sai của sai số
- Tính chất của các ước lượng bình phương bé nhất
- Kiểm tra sự phù hợp của mô hình
- Kiểm định ý nghĩa của mô hình
- Hệ số xác định và hệ số xác định hiệu chỉnh
- Kiểm định giả thuyết cho từng hệ số hồi quy
- Khoảng tin cậy cho trung bình biến đáp ứng
- Khoảng tin cậy cho giá trị dự báo mới

[1] Phần đọc thêm

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

Phân tích hồi quy

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 giầy của một người, từ đó một cửa hàng bán giầy dép có thể xác định chính xác cỡ giầy 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).

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

Phân tích hồi quy

- 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 $X_1, X_2, \ldots, X_p$; các biến này còn được gọi là biến giải thích.
    - Biến phụ thuộc Y phải là biến liên tục,
    - Các biến độc lập $X_1, X_2, \ldots, X_p$ có thể là biến liên tục, rời rạc hoặc phân loại.

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

Phân tích hồi quy

- Mối liên hệ giữa $X_1, \ldots, X_p$ và Y được biểu diễn bởi một hàm tuyến tính.
- Sự thay đổi trong Y được giả sử do những thay đổi trong $X_1, \ldots, X_p$ gây ra.

- Trên cơ sở xác định mối liên hệ giữa biến phụ thuộc Y và các biến giải thích X_1, X_2, ..., 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.

Outline

1 Giới thiệu
2 Mô hình hồi quy tuyến tính đơn
3 Khoảng tin cậy cho mô hình hồi quy
4 Kiểm định giả thuyết cho mô hình hồi quy
5 Tương quan
6 Mô hình hồi quy tuyến tính bội
7 Kiểm định giả thuyết trong mô hình hồi quy bội

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

Mô hình hồi quy tuyến tính

Định nghĩa 1

Một mô hình thống kê tuyến tính bội (Multiple linear regression model) liên quan đến một biến ngẫu nhiên Y và tập các biến giải thích $x_1, \ldots, x_p$ là phương trình có dạng

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \epsilon \qquad (1)$$

với

- $\beta_0, \ldots, \beta_p$ là các tham số chưa biết, gọi là các hệ số hồi quy,
- $x_1, \ldots, x_p$ là các biến độc lập, không ngẫu nhiên,
- ϵ là thành phần sai số, ϵ được giả sử có phân phối chuẩn với $\mathbb{E}(\epsilon) = 0$ và $\mathbb{V}ar(\epsilon) = \sigma^2$.

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

Mô hình hồi quy tuyến tính đơn

Định nghĩa 2

Trong phương trình (1), thay $p = 1$ ta được

$$Y = \beta_0 + \beta_1 x + \epsilon \qquad (2)$$

(2) được gọi là một mô hình hồi quy tuyến tính đơn (simple linear regression model); mô hình chỉ có một biến phụ thuộc Y và một biến giải thích X. β_0 và β_1 lần lượt là hệ số chặn (intercept) và hệ số góc (slope) của đường thẳng hồi quy.

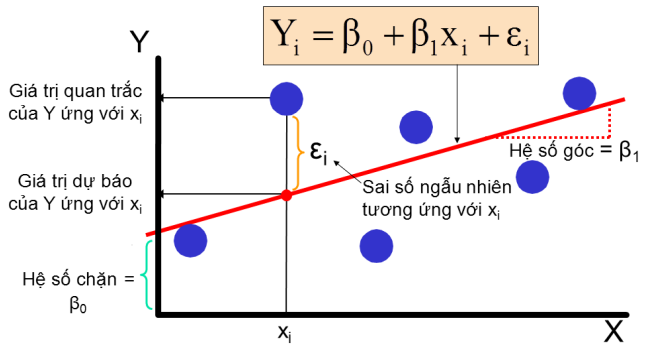
Với $(x_1, y_1), \ldots, (x_n, y_n)$ là n cặp giá trị quan trắc của một mẫu ngẫu nhiên cỡ n, từ (2) ta có

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, 2, \ldots, n \qquad (3)$$

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

Mô hình hồi quy tuyến tính đơn

- Sử dụng đồ thị phân tán (Scatter plot) để biểu diễn các cặp giá trị quan trắc (x_i, y_i) trên hệ trục tọa độ Oxy.



Các giả định về sai số ngẫu nhiên

- Các sai số ngẫu nhiên $\epsilon_i, i = 1, \ldots, n$ trong mô hình (3) giả sử thỏa các điều kiện sau
  - Các sai số ϵ_i độc lập với nhau,
  - $\mathbb{E}(\epsilon_i) = 0$ và $\mathbb{V}ar(\epsilon_i) = \sigma^2$,
  - Các sai số có phân phối chuẩn: $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$
- Với quan trắc x đã biết,

$$\mathbb{E}(Y|x) = \beta_0 + \beta_1 x \tag{4}$$

- Từ (3) ta có

$$Y \sim \mathcal{N}(\beta_0 + \beta_1 x, \sigma^2) \tag{5}$$

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

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):

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x \tag{6}$$

- 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).

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

Phương pháp bình phương bé nhất (PPBPBN)

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

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

Ta định nghĩa thặng dư thứ i:

$$\epsilon_i = y_i - (\beta_0 + \beta_1 x_i) \tag{7}$$

Định nghĩa 3

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

$$SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 x_i)]^2 \tag{8}$$

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

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

Phương pháp bình phương bé nhất

Từ (8), lấy đạo hàm theo β_0 và β_1,

$$\frac{\partial SSE}{\partial \beta_0} = -2\sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 x_i)] = 0$$

$$\frac{\partial SSE}{\partial \beta_1} = -2\sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 x_i)]\, x_i = 0$$

ta thu được hệ phương trình

$$\begin{aligned} n\beta_0 + \beta_1 \sum_{i=1}^{n} x_i &= \sum_{i=1}^{n} y_i \\ \beta_0 \sum_{i=1}^{n} x_i + \beta_1 \sum_{i=1}^{n} x_i^2 &= \sum_{i=1}^{n} x_i y_i \end{aligned} \tag{9}$$

Ước lượng bình phương bé nhất

Giả hệ (14), ta tìm được các ước lượng BPBN của β_0 và β_1 là

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - \frac{\left(\sum_{i=1}^n x_i\right)\left(\sum_{i=1}^n y_i\right)}{n}}{\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n}} = \frac{S_{xy}}{S_{xx}} \tag{10}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \tag{11}$$

với S_{xx} và S_{xy} xác định bởi

$$S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n} \tag{12}$$

$$S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n x_i y_i - \frac{\left(\sum_{i=1}^n x_i\right)\left(\sum_{i=1}^n y_i\right)}{n} \tag{13}$$

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

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:

(1)

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

đạt giá trị bé nhất,

(2)

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

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

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

Ví dụ

Ví dụ 4

Một nhà thực vật học khảo sát mối liên hệ giữa tổng diện tích bề mặt (đv: cm^2) 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.

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

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)

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

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).

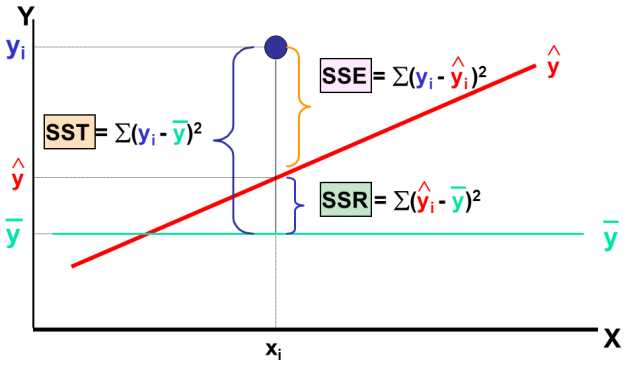
Ta có:

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \qquad (14)$$

$$SST = SSR + SSE$$

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

Độ đo sự biến thiên của dữ liệu



HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

Hệ số xác định

Định nghĩa 5

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:

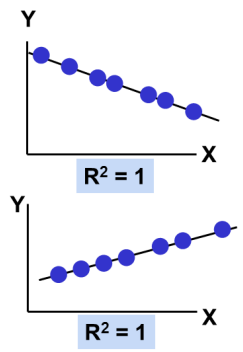
$$R^2 = \frac{SSR}{SST} \qquad (15)$$

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.

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

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

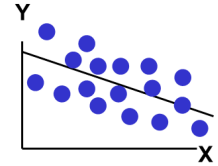


$R^2 = 1$

X và Y có mối liên hệ tuyến tính mạnh:

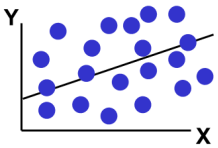
100% sự biến thiên của Y được giải thích bởi sự biến thiên của X

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



$0 < R^2 < 1$

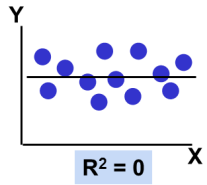
X và Y có mối liên hệ tuyến tính yếu:



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

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

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)

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

Ước lượng phương sai σ^2 của sai số

Xét mô hình

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, \ldots, n$$

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

Từ (5), 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,

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

Nên,

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

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

Ước lượng phương sai σ^2 của sai số

Ta kết luận rằng $\dfrac{SSE}{n-2}$ là một ước lượng không chệch cho σ^2.

Suy ra ước lượng $\hat{\sigma}^2$ của σ^2 được tính bởi

$$\hat{\sigma}^2 = \frac{SSE}{n-2} \tag{16}$$

- Tìm SSE:

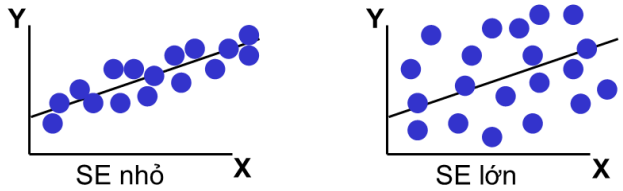
$$SSE = SST - \hat{\beta}_1 S_{xy}$$

- 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.

So sánh sai số chuẩn



HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

Tính chất của các ước lượng BPBN

Định lí 6

Xét $Y = \beta_0 + \beta_1 x + \epsilon$ là một mô hình hồi quy tuyến tính đơn với $\epsilon \sim \mathcal{N}(0, \sigma^2)$; với n quan trắc độc lập $y_i, i = 1, \ldots, n$ ta có tương ứng các sai số ϵ_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*

$$\mathbb{E}(\hat{\beta}_0) = \beta_0, \ \mathbb{V}ar(\hat{\beta}_0) = \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)\sigma^2, \tag{17}$$

$$\mathbb{E}(\hat{\beta}_1) = \beta_1, \ \mathbb{V}ar(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}} \tag{18}$$

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

Tính chất của các ước lượng BPBN

Định nghĩa 7

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

$$SE(\hat{\beta}_0) = \sqrt{\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)\hat{\sigma}^2} \tag{19}$$

$$SE(\hat{\beta}_1) = \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} \tag{20}$$

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

Tính chất của các ước lượng BPBN

Định lí 8 (Gauss - Markov)

Xét mô hình hồi quy tuyến tính đơn

$$Y = \beta_0 + \beta_1 x + \epsilon$$

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.

Outline

1 Giới thiệu
2 Mô hình hồi quy tuyến tính đơn
3 Khoảng tin cậy cho mô hình hồi quy
4 Kiểm định giả thuyết cho mô hình hồi quy
5 Tương quan
6 Mô hình hồi quy tuyến tính bội
7 Kiểm định giả thuyết trong mô hình hồi quy bội

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

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

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

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

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

$$T_{\beta_1} = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{MSE}{S_{xx}}}} \tag{21}$$

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

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

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

$$T_{\beta_0} = \frac{Z_0}{\sqrt{\frac{(SSE/\sigma^2)}{n-2}}} = \frac{\hat{\beta}_0 - \beta_0}{\sqrt{MSE\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}} \tag{22}$$

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

Khoảng tin cậy cho hệ số hồi quy

- Khoảng tin cậy $100(1-\alpha)\%$ cho β_1:

$$\hat{\beta}_1 - t^{n-2}_{1-\alpha/2}\sqrt{\frac{MSE}{S_{xx}}} \le \beta_1 \le \hat{\beta}_1 + t^{n-2}_{1-\alpha/2}\sqrt{\frac{MSE}{S_{xx}}} \qquad (23)$$

- Khoảng tin cậy $100(1-\alpha)\%$ cho β_0:

$$\hat{\beta}_0 - t^{n-2}_{1-\alpha/2}\sqrt{MSE\left(\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}}\right)} \le \beta_0 \le \hat{\beta}_0 + t^{n-2}_{1-\alpha/2}\sqrt{MSE\left(\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}}\right)} \qquad (24)$$

với

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

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

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

(1) $\hat{\mu}_{Y|x_0}$ tuân theo luật phân phối chuẩn.

(2) Kỳ vọng và phương sai của $\hat{\mu}_{Y|x_0}$ lần lượt là

$$\mathbb{E}(\hat{\mu}_{Y|x_0}) = \beta_0 + \beta_1 x_0$$
$$\mathbb{V}ar(\hat{\mu}_{Y|x_0}) = \left[\frac{1}{n} + \frac{(\bar{x}-x_0)^2}{S_{xx}}\right]\sigma^2$$

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

Khoảng tin cậy cho trung bình biến đáp ứng

- Ta có

$$\frac{\hat{\mu}_{Y|x_0} - \mathbb{E}(\hat{\mu}_{Y|x_0})}{\sqrt{\mathbb{V}ar(\hat{\mu}_{Y|x_0})}} = \frac{(\hat{\beta}_0 + \hat{\beta}_1 x_0) - (\beta_0 + \beta_1 x_0)}{\sigma\sqrt{\frac{1}{n} + \frac{(\bar{x}-x_0)^2}{S_{xx}}}} \sim \mathcal{N}(0,1)$$

- Vì $\hat{\mu}_{Y|x_0}$ độc lập với $SSE/\sigma^2 \sim \chi^2(n-2)$ nên

$$\frac{\hat{\mu}_{Y|x_0} - (\beta_0 + \beta_1 x_0)}{\sqrt{\frac{1}{n} + \frac{(\bar{x}-x_0)^2}{S_{xx}}}\sqrt{\frac{SSE}{n-2}}} = \frac{\hat{\mu}_{Y|x_0} - (\beta_0 + \beta_1 x_0)}{\sqrt{\mathrm{MSE}\left[\frac{1}{n} + \frac{(\bar{x}-x_0)^2}{S_{xx}}\right]}} \sim t(n-2) \qquad (25)$$

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

Khoảng tin cậy cho trung bình biến đáp ứng

- Khoảng tin cậy $100(1-\alpha)\%$ cho trung bình biến đáp ứng là

$$\hat{\mu}_{Y|x_0} - t^{n-2}_{1-\alpha/2}\sqrt{\mathrm{MSE}\left[\frac{1}{n} + \frac{(\bar{x}-x_0)^2}{S_{xx}}\right]} \le \mu_{Y|x_0} \le \hat{\mu}_{Y|x_0} + t^{n-2}_{1-\alpha/2}\sqrt{\mathrm{MSE}\left[\frac{1}{n} + \frac{(\bar{x}-x_0)^2}{S_{xx}}\right]} \qquad (26)$$

với

- $\hat{\mu}_{Y|x_0} = \hat{\beta}_0 + \hat{\beta}_1 x_0$, và
- $t^{n-2}_{1-\alpha/2}$: phân vị mức $1-\alpha/2$ của biến ngẫu nhiên $t(n-2)$.
- $\mathrm{MSE} = \mathrm{SSE}/(n-2)$: trung bình bình phương sai số.

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

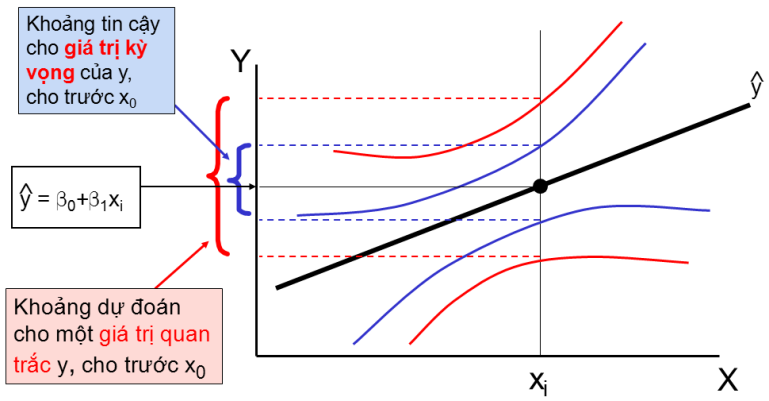
$$\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0 \tag{27}$$

$\hat{Y}_0$ 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.

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

Dự đoán giá trị quan trắc mới



HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

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

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

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

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

Do đó,

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

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

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 $\hat{\sigma}^2 = \text{MSE} = \dfrac{SSE}{n-2}$, thu được

$$T = \frac{\hat{Y}_0 - Y_0}{\sqrt{\text{MSE}\left[1 + \frac{1}{n} + \frac{(\bar{x} - x_0)^2}{S_{xx}}\right]}} \sim t(n-2) \tag{28}$$

Dự đoán giá trị quan trắc mới

- Khoảng tin cậy $100(1-\alpha)\%$ cho giá trị dự báo mới Y_0 ứng với một giá trị x_0 cho trước là

$$\hat{Y}_0 - t^{n-2}_{1-\alpha/2}\sqrt{\text{MSE}\left[1+\frac{1}{n}+\frac{(\bar{x}-x_0)^2}{S_{xx}}\right]} \le Y_0 \le \hat{Y}_0 + t^{n-2}_{1-\alpha/2}\sqrt{\text{MSE}\left[1+\frac{1}{n}+\frac{(\bar{x}-x_0)^2}{S_{xx}}\right]} \quad (29)$$

với

- $t^{n-2}_{1-\alpha/2}$ là phân vị mức $1-\alpha/2$ của $t(n-2)$,
- $\text{MSE} = \dfrac{SSE}{n-2}$.

Ví dụ

Ví dụ 9

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 $x = 5$.

Outline

1 Giới thiệu
2 Mô hình hồi quy tuyến tính đơn
3 Khoảng tin cậy cho mô hình hồi quy
4 Kiểm định giả thuyết cho mô hình hồi quy
5 Tương quan
6 Mô hình hồi quy tuyến tính bội
7 Kiểm định giả thuyết trong mô hình hồi quy bội

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:

$$(a)\begin{cases} H_0: \beta_0 = b_0 \\ H_1: \beta_0 \ne b_0 \end{cases} \qquad (b)\begin{cases} H_0: \beta_0 = b_0 \\ H_1: \beta_0 < b_0 \end{cases} \qquad (c)\begin{cases} H_0: \beta_0 = b_0 \\ H_1: \beta_0 > b_0 \end{cases}$$

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

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:

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

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

Kiểm định giả thuyết cho β_0

4. Xác định miền bác bỏ

Đối thuyết	**Miền bác bỏ**	**p - giá trị**
$H_1 : \beta_0 \neq b_0$	$\lvert t_{\beta_0}\rvert > t^{n-2}_{1-\alpha/2}$	$p = 2\mathbb{P}(T_{n-2} \geq \lvert t_{\beta_0}\rvert)$
$H_1 : \beta_0 < b_0$	$t_{\beta_0} < -t^{n-2}_{1-\alpha}$	$p = \mathbb{P}(T_{n-2} \leq t_{\beta_0})$
$H_1 : \beta_0 > b_0$	$t_{\beta_0} > t^{n-2}_{1-\alpha}$	$p = \mathbb{P}(T_{n-2} \geq t_{\beta_0})$

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

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

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:

$$(a)\begin{cases} H_0 : \beta_1 = b_1 \\ H_1 : \beta_1 \neq b_1 \end{cases} \qquad (b)\begin{cases} H_0 : \beta_1 = b_1 \\ H_1 : \beta_1 < b_1 \end{cases} \qquad (c)\begin{cases} H_0 : \beta_1 = b_1 \\ H_1 : \beta_1 > b_1 \end{cases}$$

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

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

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:

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

Kiểm định giả thuyết cho β_1

4. Xác định miền bác bỏ

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

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

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

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: $Y = \beta_0 + \beta_1 x + \epsilon$ 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 $Var(Y|x) = \sigma^2$.
  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.

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

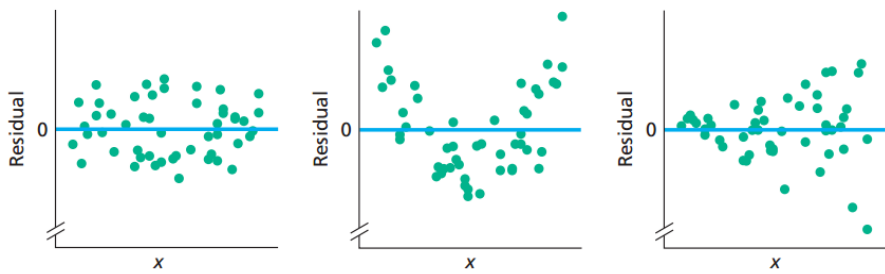
Phân tích thặng dư

- 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).

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

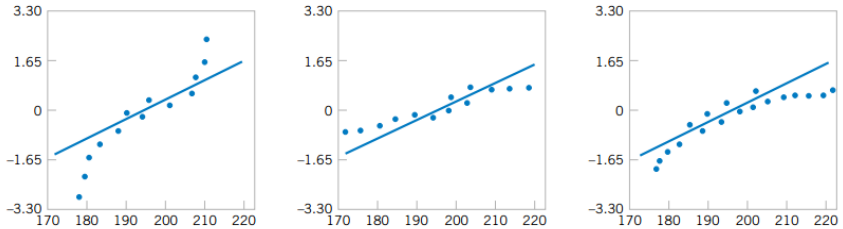
Phân tích thặng dư

- Đồ thị thặng dư:



Phân tích thặng dư

- Đồ thị xác suất chuẩn:



Outline

1 Giới thiệu
2 Mô hình hồi quy tuyến tính đơn
3 Khoảng tin cậy cho mô hình hồi quy
4 Kiểm định giả thuyết cho mô hình hồi quy
5 Tương quan
6 Mô hình hồi quy tuyến tính bội
7 Kiểm định giả thuyết trong mô hình hồi quy bội

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.

Định nghĩa 10

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

$$\mathbb{C}\text{ov}(X, Y) = \mathbb{E}\left[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right] = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) \tag{30}$$

Phân tích tương quan

Định nghĩa 11

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

$$\rho_{XY} = \frac{\mathbb{C}\text{ov}(X, Y)}{\sqrt{\mathbb{V}\text{ar}(X)\mathbb{V}ar(Y)}} \tag{31}$$

Với hai biến ngẫu nhiên X và Y bất kỳ

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

Phân tích tương quan

Định nghĩa 12

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

$$r_{XY} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}} = \frac{S_{XY}}{\sqrt{S_{XX} SST}} \quad (32)$$

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

Phân tích tương quan

Chú ý rằng:

$$\hat{\beta}_1 = \sqrt{\frac{SST}{S_{XX}}} r_{XY}$$

suy ra,

$$r_{XY}^2 = \hat{\beta}_1^{\,2} \frac{S_{XX}}{SST} = \hat{\beta}_1 \frac{S_{XY}}{SST} = \frac{SSR}{SST}$$

• Hệ số xác định, R^2, của mô hình hồi quy tuyến tính đơn bằng với bình phương của hệ số tương quan mẫu

$$R^2 = r_{XY}^2$$

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

Đánh giá hiệp phương sai

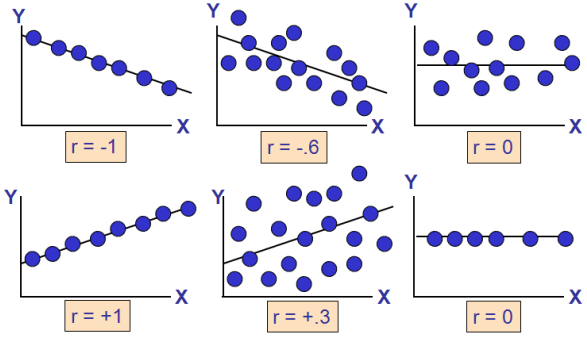
- $\mathbb{C}\text{ov}(X, Y) > 0$: X và Y có xu hướng thay đổi cùng chiều.
- $\mathbb{C}\text{ov}(X, Y) < 0$: X và Y có xu hướng thay đổi ngược chiều.
- $\mathbb{C}\text{ov}(X, Y) = 0$: X và Y độc lập (tuyến tính).

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

Đánh giá hệ số tương quan

- Miền giá trị: $-1 \leq r_{XY} \leq 1$,
- $-1 \leq r_{XY} < 0$: tương quan âm. r_{XY} càng gần -1 biểu thị mối liên hệ tuyến tính nghịch giữa X và Y càng mạnh.
- $0 < r_{XY} \leq 1$: tương quan dương. r_{XY} càng gần 1 biểu thị mối liên hệ tuyến tính thuận giữa X và Y càng mạnh.
- r_{XY} càng gần 0, biểu thị mối liên hệ tuyến tính yếu. $r_{XY} = 0$: không có mối liên hệ tuyến tính giữa X và Y.

Đánh giá hệ số tương quan



HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

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

$$H_0 : \rho = 0$$

- Thống kê kiểm định

$$T = r_{XY}\sqrt{\frac{n-2}{1-r_{XY}^2}} \tag{33}$$

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

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

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:

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

với mức ý nghĩa α cho trước.

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

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.

Kiểm định giả thuyết cho hệ số tương quan

4. Xác định miền bác bỏ

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

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

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

Outline

1 Giới thiệu
2 Mô hình hồi quy tuyến tính đơn
3 Khoảng tin cậy cho mô hình hồi quy
4 Kiểm định giả thuyết cho mô hình hồi quy
5 Tương quan
6 Mô hình hồi quy tuyến tính bội
7 Kiểm định giả thuyết trong mô hình hồi quy bội

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

Giới thiệu [1]

Định nghĩa 13

Một mô hình thống kê tuyến tính bội (Multiple linear regression model) liên quan đến một biến ngẫu nhiên Y và tập các biến giải thích $x_1, \ldots, x_k$ là phương trình có dạng

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon \tag{34}$$

với

- $\beta_0, \ldots, \beta_p$ là các tham số chưa biết, gọi là các hệ số hồi quy,
- $x_1, \ldots, x_k$ là các biến độc lập, không ngẫu nhiên,
- ϵ là thành phần sai số, ϵ được giả sử có phân phối chuẩn với $\mathbb{E}(\epsilon) = 0$ và $\mathbb{V}ar(\epsilon) = \sigma^2$.

[1] Từ slide này trở về sau dùng để đọc thêm

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

Giới thiệu

Một số dạng khác của mô hình:

- Mô hình đa thức

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \epsilon$$

Nếu đặt $x_1 = x, x_2 = x^2, x_3 = x^3$, ta có

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$$

- Mô hình với tương tác

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \epsilon$$

Nếu đặt $x_3 = x_1 x_2$ và $\beta_3 = \beta_{12}$, ta có

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$$

Biểu diễn mô hình theo dạng ma trận

Xét $y_1, \ldots, y_n$ là n giá trị quan trắc độc lập của Y. Khi đó, mỗi y_i có thể biểu diễn dưới dạng

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \epsilon_i$$

với x_{ij} là biến độc lập thứ j của quan trắc thứ i, $i = 1, 2, \ldots, n$ và các sai số ϵ_i độc lập với nhau tương tự như trong mô hình hồi quy tuyến tính.
Đặt $x_0 = 1$, định nghĩa các ma trận sau:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}; \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1k} \\ 1 & x_{21} & x_{22} & \ldots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \ldots & x_{nk} \end{bmatrix}; \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}; \quad \text{và} \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_k \end{bmatrix}$$

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

Biểu diễn mô hình theo dạng ma trận

Mô hình hồi quy bội dưới dạng ma trận

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{35}$$

Tổng quát, $\mathbf{y}$ là véc-tơ cỡ $(n \times 1)$ chứa các giá trị quan trắc, $\mathbf{X}$ là ma trận cỡ $(n \times p)$ chứa các biến độc lập, $\boldsymbol{\beta}$ là véc-tơ cỡ $(p \times 1)$ chứa các hệ số hồi quy và ϵ là véc-tơ cỡ $(n \times 1)$ của các sai số ngẫu nhiên. (chú ý: $p = k + 1$)
Tổng bình phương thặng dư trong mô hình hồi quy bội được định nghĩa như sau

$$L = \sum_{i=1}^{n} \epsilon^2 = \boldsymbol{\epsilon}'\boldsymbol{\epsilon} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \tag{36}$$

Ta đi tìm véc-tơ ước lượng bình phương bé nhất $\hat{\boldsymbol{\beta}}$ mà làm cho L đạt giá trị nhỏ nhất, là nghiệm của phương trình

$$\frac{\partial L}{\partial \boldsymbol{\beta}} = \mathbf{0} \tag{37}$$

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

Ước lượng bình phương bé nhất

Kết quả thu được khi lấy đạo hàm (37) là

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y} \tag{38}$$

Giải hệ (38) ta tính được véc-tơ các ước lượng bình phương bé nhất

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{39}$$

Với $\hat{\boldsymbol{\beta}} = [\beta_0, \beta_1, \ldots, \beta_k]'$, mô hình hồi quy ước lượng có dạng

$$\hat{y}_i = \hat{\beta}_0 + \sum_{j=1}^{k} \hat{\beta}_j x_{ij}, \quad i = 1, 2 \ldots, n \tag{40}$$

Biểu diễn dưới dạng ma trận: $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$
Véc-tơ giá trị thặng dư:

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} \tag{41}$$

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

Ví dụ

Ví dụ 14

Cho dữ liệu về các sợi dây làm chất bán dẫn (tụ): gồm y = lực tác động (dùng để phát vỡ tụ), x_1 = chiều dài sợi dây và x_2 = chiều cao của khối tụ. Hãy thiết lập mô hình hồi quy tuyến tính bội biểu diễn mối quan hệ y và x_1, x_2. Dữ liệu chứa trong file Wire-bond-data.csv.

Xem code **R** trong file: ex-multiple-regression.R.

Ước lượng phương sai của sai số

- Sai số trong mô hình hồi quy

$$\epsilon \sim \mathcal{N}(0, \sigma^2)$$

- Ước lượng của σ^2 cho bởi công thức

$$\hat{\sigma^2} = \frac{\sum_{i=1}^{n} e_i^2}{n-p} = \frac{SSE}{n-p} \tag{42}$$

Với $p = k + 1$.

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

Tính chất của các ước lượng bình phương bé nhất

Gọi $\hat{\boldsymbol{\beta}} = [\beta_0, \beta_1, \ldots, \beta_k]'$ là véc-tơ chứa các ước lượng bình phương bé nhất, $\hat{\boldsymbol{\beta}}$ thỏa các tính chất sau:

- $\mathbb{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$, tức là $\hat{\boldsymbol{\beta}}$ là một ước lượng không chệch (unbiased estimator) cho $\boldsymbol{\beta}$.
- Đặt $\mathbf{C} = (\mathbf{X}'\mathbf{X})^{-1}$ ($\mathbf{C}$ là ma trận đối xứng cỡ $p \times p$), ta có

$$\mathbb{V}ar(\hat{\beta}_j) = \sigma^2 C_{jj}, \quad j = 0, 1, \ldots, p$$
$$Cov(\hat{\beta}_i, \hat{\beta}_j) = \sigma^2 C_{ij}, \quad i \neq j$$

Tổng quát, ma trận hiệp phương sai của $\hat{\boldsymbol{\beta}}$ là một ma trận đối xứng cỡ $p \times p$ trong đó thành phần thứ jj là phương sai của $\hat{\beta}_j$ và thành phần thứ ij là hiệp phương sai giữa $\hat{\beta}_i$ và $\hat{\beta}_j$ có dạng như sau

$$Cov(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2\mathbf{C}$$

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

Tính chất của các ước lượng bình phương bé nhất

Ví dụ, trường hợp $p = 2$, ta có ma trận $\mathbf{C}$ là

$$\mathbf{C} = \begin{bmatrix} C_{00} & C_{01} & C_{0,2} \\ C_{10} & C_{11} & C_{12} \\ C_{20} & C_{21} & C_{22} \end{bmatrix}$$

Phương sai của $\hat{\beta}_j$ bằng $\sigma^2 C_{jj}$, $j = 0, 1, 2$. Sai số chuẩn của ước lượng bình phương bé nhất xác định như sau

$$SE(\hat{\beta}_j) = \sqrt{\hat{\sigma}^2 C_{jj}} \tag{43}$$

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

Kiểm tra sự phù hợp của mô hình

Để kiểm tra sự phù hợp của mô hình và kiểm tra các giả định, ta sử dụng các dạng đồ thị sau và nhận xét:

- Đồ thị xác suất.
- Đồ thị phân tán: giữa biến phụ thuộc và các biến độc lập; đồ thị thặng dư.

Outline

1 Giới thiệu
2 Mô hình hồi quy tuyến tính đơn
3 Khoảng tin cậy cho mô hình hồi quy
4 Kiểm định giả thuyết cho mô hình hồi quy
5 Tương quan
6 Mô hình hồi quy tuyến tính bội
7 Kiểm định giả thuyết trong mô hình hồi quy bội

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

Kiểm định ý nghĩa của mô hình hồi quy

Kiểm định ý nghĩa của mô hình hồi quy là kiểm định dùng để xác định xem liệu có tồn tại mối liên hệ tuyến tính giữa biến phụ thuộc y và các biến độc lập $x_1, x_2, \ldots, x_k$ hay không. Giả thuyết và đối thuyết được phát biểu như sau

$$H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$$
$$H_1 : \beta_j \neq 0 \text{ với ít nhất một giá trị } j \tag{44}$$

Việc bác bỏ H_0 suy ra rằng có ít nhất một biến độc lập $x_1, x_2, \ldots, x_k$ có mối liên hệ tuyến tính với y, tức là mô hình có ý nghĩa.
Thống kê kiểm định cho giả thuyết H_0 là

$$F_0 = \frac{SSR/k}{SSE/(n-p)} = \frac{MSR}{MSE} \tag{45}$$

với MSR là trung bình bình phương hồi quy (Mean Squared Regression) và MSE là trung bình bình phương sai số (Mean Squared Error).

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

Kiểm định ý nghĩa của mô hình hồi quy

Nhắc lại rằng,

$$SST = SSR + SSE$$

với

$$SST = \sum_{i=1}^{n} y_i^2 - \frac{(\sum_{i=1}^{n} yi)^2}{n} = \mathbf{y}'\mathbf{y} - \frac{(\sum_{i=1}^{n} yi)^2}{n}$$

và

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

Thay thế $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\hat{\beta}$, ta thu được

$$SSE = \mathbf{y}'\mathbf{y} - \hat{\beta}'\mathbf{X}'\mathbf{y}$$

suy ra,

$$SSR = \hat{\beta}'\mathbf{X}'\mathbf{y} - \frac{(\sum_{i=1}^{n} yi)^2}{n}$$

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

Kiểm định ý nghĩa của mô hình hồi quy

Bảng phân tích phương sai cho kiểm định ý nghĩa của mô hình hồi quy bội:

Nguồn gốc của sự biến thiên	Tổng bình phương	Bậc tự do	Trung bình bình phương	F_0
Hồi quy	SSR	k	MSR	MSR/MSE
Sai số	SSE	n-p	MSE	
Tổng cộng	SST	n-1		

Bác bỏ H_0 khi:

$$f_0 > f_{\alpha;k,n-p}$$

Với $f_{\alpha;k,n-p}$ là phân vị trên mức α của biến ngẫu nhiên $F(k, n-p)$.

Hệ số xác định và Hệ số xác định hiệu chỉnh

- Tương tự như mô hình hồi quy tuyến tính đơn, để đánh giá sự phù hợp của mô hình hồi quy bội, ta sử dụng hệ số xác định $R-$ bình phương R-squared:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \tag{46}$$

- Ta có nhận xét rằng nếu R^2 càng lớn thì mô hình càng phù hợp; tuy nhiên, một nhược điểm là khi ta thêm một biến mới vào trong mô hình thì R^2 luôn tăng lên; điều này sẽ gây ra khó khăn trong việc đánh giá nếu ta muốn xác định biến mới thêm vào có phù hợp với mô hình hay không. Để khắc phục điều này, ta sử dụng hệ số xác định hiệu chỉnh

$$R^2_{adj} = 1 - \frac{SSE/(n-p)}{SST/(n-1)} \tag{47}$$

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

Kiểm định giả thuyết cho từng hệ số hồi quy

- Khi ta thêm một biến mới vào mô hình hồi quy sẽ làm tăng tổng bình phương hồi quy (SSR) và giảm tổng bình phương sai số (SSE) (Đó là lý do hệ số xác định R^2 luôn tăng khi thêm biến),
- Do vậy, ta phải quyết định xem mức độ tăng của R^2 như thế nào là đủ lớn để đưa biến mới vào mô hình. Hơn thế nữa, việc đưa một biến không quan trọng vào mô hình có thể làm tăng trung bình bình phương sai số, điều này sẽ làm mô hình kém phù hợp với dữ liệu.
- Ta cần thực hiện kiểm định ý nghĩa cho từng hệ số hồi quy, từ đó sẽ đưa ra quyết định nên thêm biến vào mô hình hay không, với giả thuyết và đối thuyết

$$\begin{aligned} H_0 &: \beta_j = 0 \\ H_1 &: \beta_j \neq 0 \end{aligned} \tag{48}$$

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

Kiểm định giả thuyết cho từng hệ số hồi quy

- Thống kê kiểm định cho giả thuyết H_0 là

$$T_0 = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j} = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 C_{jj}}} \tag{49}$$

Thống kê T_0 có phân phối Student với $n-p$ bậc tự do.

- Bác bỏ H_0 khi:

$$|t_0| > t^{n-p}_{1-\alpha/2}$$

với $t^{n-p}_{1-\alpha/2}$ là phân vị mức $1-\alpha/2$ của biến ngẫu nhiên $T_0 \sim t(n-p)$.

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

Khoảng tin cậy cho trung bình biến đáp ứng

- Xét mô hình hồi quy bội

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

- Trung bình của biến đáp ứng $\mathbf{y}$ bằng tại một điểm cho trước $(x_{01}, x_{02}, \ldots, x_{0k})$ có dạng

$$\mathbb{E}(Y|\mathbf{x}_0) = \boldsymbol{\mu}_{Y|\mathbf{x}_0} = \mathbf{x}'_0\boldsymbol{\beta}$$

Với

$$\mathbf{x}_0 = \begin{bmatrix} 1 \\ x_{01} \\ \vdots \\ x_{0k} \end{bmatrix}$$

- Ước lượng của $\boldsymbol{\mu}_{Y|\mathbf{x}_0}$ là

$$\hat{\boldsymbol{\mu}}_{Y|\mathbf{x}_0} = \mathbf{x}'_0\hat{\boldsymbol{\beta}} \tag{50}$$

Khoảng tin cậy cho trung bình biến đáp ứng

$\hat{\boldsymbol{\mu}}_{Y|\mathbf{x}_0}$ là ước lượng không chệch cho $\boldsymbol{\mu}_{Y|\mathbf{x}_0}$ vì $\mathbb{E}(\mathbf{x}'_0\hat{\boldsymbol{\beta}}) = \mathbf{x}'_0\boldsymbol{\beta} = \mu_{Y|\mathbf{x}_0}$ và phương sai của $\hat{\boldsymbol{\mu}}_{Y|\mathbf{x}_0}$ là

$$\mathbb{V}ar(\hat{\boldsymbol{\mu}}_{Y|\mathbf{x}_0}) = \sigma^2\mathbf{x}'_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0 \tag{51}$$

Định nghĩa 15

Trong mô hình hồi quy tuyến tính bội, một khoảng tin cậy với độ tin cậy $100(1-\alpha)\%$ cho trung bình biến đáp ứng tại điểm $(x_{01}, x_{02}, \ldots, x_{0k})$ có dạng

$$\hat{\boldsymbol{\mu}}_{Y|\mathbf{x}_0} - t^{n-p}_{1-\alpha/2}\sqrt{\hat{\sigma}^2\mathbf{x}'_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0} \le \boldsymbol{\mu}_{Y|\mathbf{x}_0} \le \hat{\boldsymbol{\mu}}_{Y|\mathbf{x}_0} + t^{n-p}_{1-\alpha/2}\sqrt{\hat{\sigma}^2\mathbf{x}'_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0} \tag{52}$$

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

Khoảng tin cậy cho giá trị dự báo mới

Cho trước điểm $\mathbf{x}_0 = [1, x_{01}, x_{02}, \ldots, x_{0k}]'$, một ước lượng điểm cho giá trị dự báo mới của Y_0 là

$$\hat{\mathbf{y}}_0 = \mathbf{x}'_0\hat{\boldsymbol{\beta}} \tag{53}$$

Định nghĩa 16

Một khoảng tin cậy với độ tin cậy $100(1-\alpha)\%$ cho giá trị dự báo mới có dạng

$$\hat{\mathbf{y}}_0 - t^{n-p}_{1-\alpha/2}\sqrt{\hat{\sigma}^2(1+\mathbf{x}'_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0)} \le \mathbf{Y}_0 \le \hat{\mathbf{y}}_0 + t^{n-p}_{1-\alpha/2}\sqrt{\hat{\sigma}^2(1+\mathbf{x}'_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0)} \tag{54}$$

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

Ví dụ

Ví dụ 17

Với dữ liệu về các sợi dây làm chất bán dẫn (tụ): gồm y = lực tác động (dùng để phát vỡ tụ), x_1 = chiều dài sợi dây và x_2 = chiều cao của khối tụ.

(a) Xây dựng mô hình hồi quy bội biểu diễn mối quan hệ của y theo x_1 và x_2.
(b) Tính hệ số xác định R^2 và hệ số xác định hiệu chỉnh R^2_{adj}.
(c) Kiểm định ý nghĩa cho véc-tơ hệ số hồi quy $\boldsymbol{\beta}$ và từng hệ số hồi quy. ($\alpha = 5\%$)
(d) Tìm khoảng tin cậy 95% cho trung bình biến đáp ứng tại $(x_1 = 8, x_2 = 275)$.
(e) Tìm khoảng tin cậy 95% cho giá trị dự báo mới tại $(x_1 = 8, x_2 = 275)$.

HỒI QUY VÀ TƯƠNG QUAN
Nguyễn Văn Thìn
Giới thiệu
Mô hình hồi quy tuyến tính đơn
Khoảng tin cậy cho mô hình hồi quy
Kiểm định giả thuyết cho MHHQ
Tương quan
Mô hình hồi quy tuyến tính bội
Kiểm định giả thuyết trong mô hình hồi quy bội

Qua bài giảng ta hiểu/có thể áp dụng

Mô hình hồi quy tuyến tính đơn

- Bài toán hồi quy và hồi quy tuyến tính đơn
- Các giả định về sai số ngẫu nhiên
- Ước lượng các hệ số hồi quy bằng phương pháp bình phương bé nhất (BPBN)
- Hệ số xác định
- Các tính chất của ước lượng BPBN
- Khoảng tin cậy cho hệ số hồi quy, trung bình biến đáp ứng
- Dự đoán giá trị quan trắc mới
- Kiểm định giả thuyết cho β_0, β_1
- Phân tích thặng dư, tương quan
- Kiểm định giả thuyết cho hệ số tương quan

Qua bài giảng ta hiểu/có thể áp dụng

Mô hình hồi quy tuyến tính bội

- Biểu diễn mô hình theo dạng ma trận
- Ước lượng bình phương bé nhất
- Ước lượng phương sai của sai số
- Tính chất của các ước lượng bình phương bé nhất
- Kiểm tra sự phù hợp của mô hình
- Kiểm định ý nghĩa của mô hình
- Hệ số xác định và hệ số xác định hiệu chỉnh
- Kiểm định giả thuyết cho từng hệ số hồi quy
- Khoảng tin cậy cho trung bình biến đáp ứng
- Khoảng tin cậy cho giá trị dự báo mới