CHƯƠNG 3

HỒI QUY ĐA BIẾN

HỒI QUY ĐA BIẾN

MỤC TIÊU

1. Biết được phương pháp ước lượng bình phương nhỏ nhất để ước lượng hàm hồi quy đa biến tổng thể dựa trên số liệu mẫu
2. Hiểu các cách kiểm định những giả thiết

NỘI DUNG

1 Mô hình hồi quy 3 biến

2 Mô hình hồi quy k biến

3 Dự báo

3.1 Mô hình hồi quy 3 biến

❑Mô hình hồi quy tổng thể PRF

$$E(Y/X_2, X_3) = \beta_1 + \beta_2 X_2 + \beta_3 X_3$$

Ý nghĩa: PRF cho biết trung bình có điều kiện của Y với điều kiện đã biết các giá trị cố định của biến X_2 và X_3.

- Y: biến phụ thuộc
- X_2 và X_3: biến độc lập
- β_1 : hệ số tự do
- β_2 , β_3 : hệ số hồi quy riêng

3.1 Mô hình hồi quy 3 biến

Ý nghĩa hệ số hồi quy riêng*: cho biết ảnh hưởng của từng biến độc lập lên giá trị trung bình của biến phụ thuộc khi các biến còn lại được giữ không đổi.*

- Mô hình hồi quy tổng thể ngẫu nhiên:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

u_i: sai số ngẫu nhiên của tổng thể

Các giả thiết của mô hình

1. Giá trị trung bình của U_i bằng 0

$$E(U_i / X_{2i,}\ X_{3i})=0$$

2. Phương sai của các U_i là không đổi

$$Var(U_i)=\sigma^2$$

3. Không có hiện tượng tự tương quan giữa các U_i

$$Cov(U_i\ ,U_j\ )=0;\ \ i\neq j$$

4. Không có hiện tượng cộng tuyến giữa X_2 và X_3

5. U_i có phân phối chuẩn: $U_i \sim N(0, \sigma^2)$

3.1.1 Ước lượng các tham số

Hàm hồi quy mẫu:

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + e_i$$

sai số của mẫu ứng với quan sát thứ i

$$e_i = Y_i - \hat{Y}_i$$

Sử dụng phương pháp bình phương nhỏ nhất để ước lượng các tham số $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$

3.1.1 Ước lượng các tham số

$$Q = \sum e_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i})^2 \to \min$$

$$\frac{dQ}{d\hat{\beta}_1} = -2\sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i}) = 0$$

$$\frac{dQ}{d\hat{\beta}_2} = 2\sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i})(-X_{2i}) = 0$$

$$\frac{dQ}{d\hat{\beta}_3} = 2\sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i})(-X_{3i}) = 0$$

3.1.1 Ước lượng các tham số

$$\hat{\beta}_2 = \frac{\sum y_i x_{2i} \sum x_{3i}^2 - \sum y_i x_{3i} \sum x_{2i} x_{3i}}{\sum x_{2i}^2 \sum x_{3i}^2 - (\sum x_{2i} x_{3i})^2}$$

$$\hat{\beta}_3 = \frac{\sum y_i x_{3i} \sum x_{2i}^2 - \sum y_i x_{2i} \sum x_{2i} x_{3i}}{\sum x_{2i}^2 \sum x_{3i}^2 - (\sum x_{2i} x_{3i})^2}$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}_{2i} - \hat{\beta}_3 \bar{X}_{3i}$$

$$x_i = X_i - \bar{X} \qquad y_i = Y_i - \bar{Y}$$

3.1.2 Phương sai của các ước lượng

$$Var(\hat{\beta}_1) = \left(\frac{1}{n} + \frac{\bar{X}_2^2 \sum x_{3i}^2 + \bar{X}_3^2 \sum x_{2i}^2 - 2\bar{X}_2 \bar{X}_3 \sum x_{2i}x_{3i}}{\sum x_{2i}^2 \sum x_{3i}^2 - (\sum x_{2i}x_{3i})^2}\right)\sigma^2$$

$$Var(\hat{\beta}_2) = \frac{\sum x_{3i}^2}{\sum x_{2i}^2 \sum x_{3i}^2 - (\sum x_{2i}x_{3i})^2}\sigma^2$$

$$Var(\hat{\beta}_3) = \frac{\sum x_{2i}^2}{\sum x_{2i}^2 \sum x_{3i}^2 - (\sum x_{2i}x_{3i})^2}\sigma^2$$

σ^2 là phương sai của u_i chưa biết nên dùng ước lượng không chệch:

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n-3} = \frac{(1-R^2)\sum y_i^2}{n-3}$$

Hệ số xác định

Hệ số xác định R²

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n} y_i^2}$$

Mô hình hồi quy 3 biến $R^2 = \dfrac{\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}}{\sum y_i^2}$

Hệ số xác định hiệu chỉnh

Với *k* là tham số của mô hình, kể cả hệ số tự do

$$\bar{R}^2 = \frac{\sum e_i^2 / (n-k)}{\sum y_i^2 / (n-1)}$$

Hệ số xác định hiệu chỉnh

$$\overline{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-k}$$

Dùng $\overline{R}^2$ để xét việc đưa thêm 1 biến vào mô hình. Biến mới đưa vào mô hình phải thỏa 2 điều kiện:

- Làm $\overline{R}^2$ tăng
- Hệ số hồi quy biến mới thêm vào mô hình khác 0 có ý nghĩa

3.1.4 Khoảng tin cậy

Với mức ý nghĩa α hay độ tin cậy 1- α

$$\beta_i \in (\hat{\beta}_i - \varepsilon_i ; \hat{\beta}_i + \varepsilon_i)$$

Với

$$\varepsilon_i = SE(\hat{\beta}_i) t_{(n-3, \alpha/2)}$$

3.1.5 Kiểm định giả thiết

1. Kiểm định giả thiết H_0: $\beta_i = \beta_i^*$

B1. Tính $$t_i = \frac{\hat{\beta}_i - \beta_i^*}{SE(\hat{\beta}_i)}$$

B2. Nguyên tắc quyết định

✓Nếu $|t_i| > t_{(n-3,\alpha/2)}$: bác bỏ H_0

✓Nếu $|t_i| \leq t_{(n-3,\alpha/2)}$: chấp nhận H_0

2. Kiểm định giả thiết đồng thời bằng không:

H_0: $\beta_2 = \beta_3 = 0$;

(H_1: ít nhất 1 tham số khác 0)

B1. Tính

$$F = \frac{R^2(n-3)}{(1-R^2)2}$$

B2. Nguyên tắc quyết định

- ✓ $F > F_\alpha(2, n-3)$: Bác bỏ H_0: Mô hình phù hợp
- ✓ $F \le F_\alpha(2, n-3)$: Chấp nhận H_0: Mô hình không phù hợp

3.2 Mô hình hồi quy k biến

Mô hình hồi quy tổng thể

$$E(Y/X_2,\ldots X_k)=\beta_1+\beta_2X_{2i}+\ldots+\beta_kX_{ki}$$

Mô hình hồi quy mẫu ngẫu nhiên:

$$\hat{Y}_i=\hat{\beta}_1+\hat{\beta}_2X_{2i}+\ldots+\hat{\beta}_kX_{ki}+e_i$$

sai số của mẫu ứng với quan sát thứ i

$$e_i=Y_i-\hat{Y}_i=Y_i-\hat{\beta}_1-\hat{\beta}_2X_{2i}-\hat{\beta}_3X_{3i}-\ldots-\hat{\beta}_kX_{ki}$$

3.2.1 Ước lượng các tham số

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i} - \ldots - \hat{\beta}_k X_{ki}\right)^2 \rightarrow \min$$

$$\frac{\partial \sum_{i=1}^{n} e_i^2}{\partial \beta_1} = -2\sum_{i=1}^{n} \left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i} - \ldots - \hat{\beta}_k X_{ki}\right) = 0$$

$$\frac{\partial \sum_{i=1}^{n} e_i^2}{\partial \beta_2} = -2\sum_{i=1}^{n} \left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i} - \ldots - \hat{\beta}_k X_{k,i}\right) X_{2i} = 0$$

...

$$\frac{\partial \sum_{i=1}^{n} e_i^2}{\partial \beta_k} = -2\sum_{i=1}^{n} \left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i} - \ldots - \hat{\beta}_k X_{ki}\right) X_{ki} = 0$$

3.2.2 Khoảng tin cậy

Với mức ý nghĩa α hay độ tin cậy 1- α

$$\beta_i \in (\hat{\beta}_i - \varepsilon_i ; \hat{\beta}_i + \varepsilon_i)$$

Với

$$\varepsilon_i = SE(\hat{\beta}_i).t_{(n-k,\alpha/2)}$$

Hệ số xác định

$$R^2 = \frac{\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i} + \ldots + \hat{\beta}_k \sum y_i x_{ki}}{\sum y_i^2}$$

Hệ số xác định hiệu chỉnh

$$\overline{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-k}$$

Với *k* là tham số của mô hình, kể cả hệ số tự do

Hệ số xác định hiệu chỉnh

$$\overline{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-k}$$

Dùng $\overline{R}^2$ để xem xét việc đưa thêm biến vào mô hình. Biến mới đưa vào mô hình phải thỏa 2 điều kiện:

- Làm $\overline{R}^2$ tăng
- Biến mới có ý nghĩa thống kê trong mô hình mới

1. Kiểm định giả thiết về hệ số hồi quy

Kiểm định giả thiết H_0: $\beta_i = \beta_i^*$

B1.Tính

$$t_i = \frac{\hat{\beta}_i - \beta_i^*}{SE(\hat{\beta}_i)}$$

B2. Nguyên tắc quyết định

- ✓Nếu $|t_i| > t_{(n-k,\alpha/2)}$: bác bỏ H_0
- ✓Nếu $|t_i| \le t_{(n-k,\alpha/2)}$: chấp nhận H_0

2. Kiểm định sự phù hợp của mô hình: kiểm định giả thiết đồng thời bằng không:

H_0: $\beta_2 = \beta_3 = \ldots = \beta_k = 0$;

(H_1: ít nhất 1 trong k tham số khác 0)

$$F = \frac{R^2(n-k)}{(1-R^2)(k-1)}$$

B1. Tính

B2. Nguyên tắc quyết định:

✓Nếu $F > F_\alpha(k-1, n-k)$: Bác bỏ H_0: Mô hình phù hợp

✓Nếu $F \le F_\alpha(k-1, n-k)$: Chấp nhận H_0: Mô hình không phù hợp

3.3 DỰ BÁO

Mô hình hồi quy

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_2 + ... + \hat{\beta}_k X_k$$

Cho trước giá trị

$$X^0 = \begin{bmatrix} 1 \\ X_2^0 \\ .... \\ X_k^0 \end{bmatrix}$$

Dự báo giá trị trung bình và giá trị cá biệt của Y với mức ý nghĩa α hay độ tin cậy 1 - α.

3.3 DỰ BÁO

*** Ước lượng điểm**

$$\hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_2^0 + \ldots + \hat{\beta}_k X_k^0$$

*** Dự báo giá trị trung bình của Y**

$$E(Y / X_0) \in (\hat{Y}_0 - \varepsilon_0 ; \hat{Y}_0 + \varepsilon_0)$$

Với:

$$\varepsilon_0 = SE(\hat{Y}_0) t_{(n-k, \alpha/2)}$$

$$SE(\hat{Y}_0) = \sqrt{Var(\hat{Y}_0)}$$

$$Var(\hat{Y}_0) = \hat{\sigma}^2 X^{0T} (X^T . X)^{-1} . X^0$$

3.3 DỰ BÁO

*** Dự báo giá trị cá biệt của Y**

$$Y_0 \in (\hat{Y}_0 - \varepsilon_0^{'} ; \hat{Y}_0 + \varepsilon_0^{'})$$

Với:

$$\varepsilon_0^{'} = SE(Y_0 - \hat{Y}_0) t_{(n-k, \alpha/2)}$$

$$SE(Y_0 - \hat{Y}_0) = \sqrt{Var(Y_0 - \hat{Y}_0)}$$

$$Var(Y_0 - \hat{Y}_0) = Var(\hat{Y}_0) + \hat{\sigma}^2$$