

槓桿值 Leverage

意義

衡量某一個觀察值，遠離其他觀察值的程度。在回歸模型中，槓桿值越大，代表該觀察值是離群值，影響回歸線擬合的程度也越大。

公式

$$h_{ii} = [H]_{ii}$$
$$H = X(X^T X)^{-1} X^T$$

h_{ii} : 第 i 個觀察值中，第 i 列，第 i 行中的矩陣值。

$[H]_{ii}$: 所有樣本觀察值的矩陣，矩陣對角線就是 h_{ii} 的值。

X : 回歸矩陣。

範例

下表為喝酒與壽命的調查，回歸模型為 $y = 86.28 - 0.73x$ ，求各觀察值的預測值、殘差與槓桿值。

x (alcohol)	y (life)	predict y	residuals	leverage
23	65	69.54	-4.54	0.20664
4	92	83.37	8.63	0.68009
29	58	65.18	-7.18	0.30631
35	72	60.81	11.19	0.52559
14	68	76.09	-8.09	0.28138

根據回歸模型 $y = 86.28 - 0.73x$ ，將 x 帶入求預測 y 值：

$$y_1 = 86.28 - 0.73 \times 23 = 69.54$$

$$y_2 = 86.28 - 0.73 \times 4 = 83.37$$

$$y_3 = 86.28 - 0.73 \times 29 = 65.18$$

$$y_4 = 86.28 - 0.73 \times 35 = 60.81$$

$$y_5 = 86.28 - 0.73 \times 14 = 76.09$$

實際的樣本 y 值，減去回歸模型預測的 y 值，計算出殘差：

$$65 - 69.54 = -4.54$$

$$92 - 83.37 = 8.63$$

$$58 - 65.18 = -7.18$$

$$72 - 60.81 = 11.19$$

$$68 - 76.09 = -8.09$$

依據 $H = X(X^T X)^{-1} X^T$ 公式，計算所有樣本觀察值的矩陣 $[H]$ ：

已知

$$X \text{ 回歸矩陣} = \begin{bmatrix} 23 & 69.54 \\ 04 & 83.37 \\ 29 & 65.18 \\ 35 & 60.81 \\ 14 & 76.09 \end{bmatrix} \text{-----} \textcircled{1}$$

$$X^T \text{ 轉置矩陣} = \begin{bmatrix} 23 & 4 & 29 & 35 & 14 \\ 69.54 & 83.37 & 65.18 & 60.81 & 76.09 \end{bmatrix} \text{-----} \textcircled{2}$$

$$X^T X = \begin{bmatrix} 23 & 4 & 29 & 35 & 14 \\ 69.54 & 83.37 & 65.18 & 60.81 & 76.09 \end{bmatrix} \times \begin{bmatrix} 23 & 69.54 \\ 04 & 83.37 \\ 29 & 65.18 \\ 35 & 60.81 \\ 14 & 76.09 \end{bmatrix}$$
$$= \begin{bmatrix} 2807 & 7017 \\ 7017 & 25522 \end{bmatrix}$$

$$(X^T X)^{-1} \text{ 反方陣} = \frac{1}{2807 \times 25522 - 7017 \times 7017} \begin{bmatrix} 25522 & -7017 \\ -7017 & 2807 \end{bmatrix} =$$
$$\begin{bmatrix} 0.001139 & -0.00313 \\ -0.00313 & 0.000125 \end{bmatrix} \text{-----} \textcircled{3}$$

由 $\textcircled{1} \times \textcircled{3} \times \textcircled{2}$ 可以求得槓桿值

$$[H] = \begin{bmatrix} 0.20664 & 0.14350 & 0.22659 & 0.24652 & 0.17673 \\ 0.14350 & 0.68009 & -0.02588 & -0.19536 & 0.39767 \\ 0.22659 & -0.02588 & 0.30631 & 0.38604 & 0.10700 \\ 0.24652 & -0.19536 & 0.38604 & 0.52559 & 0.03721 \\ 0.17673 & 0.39767 & 0.10700 & 0.03721 & 0.28138 \end{bmatrix}$$

每一個樣本觀察值的槓桿值，就是 H 矩陣的對角線元素。