--Alan Lee

迴歸模式

REGRESSION MODEL

目錄

重點 1. 迴歸模式 Regression Model：

考慮 X 與 Y 二個隨機變數，其中的 X 表示「自變數」independent variables， Y 表示「依變數」dependent variables， ε 表示誤差項，迴歸方程式表示如下：

$Y = \alpha + \beta X + \varepsilon$

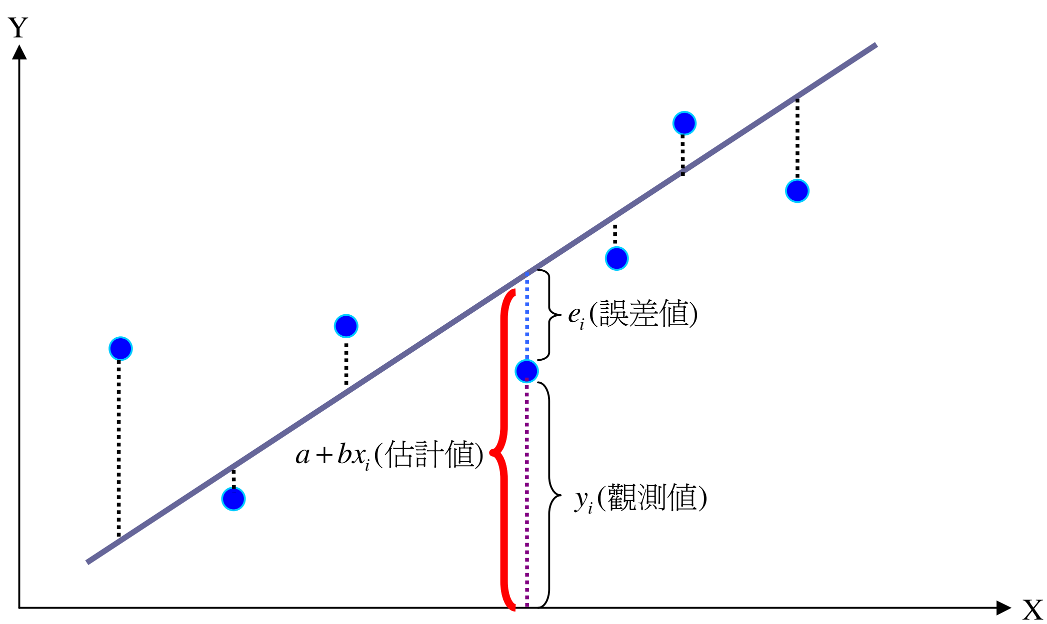
上式必須假設以下基本條件:

(1). Y_i 是獨立的常態分佈 $N\left(\alpha + \beta X_i, \sigma^2\right), i = 1, 2, \ldots, n$

(2). ε_i 是獨立的常態分佈 $N\left(0, \sigma^2\right), i = 1, 2, \ldots, n$

即 Y 的估計值爲 $\hat{y} = a + bx$ ，其中^發音爲 hat，而估計值的誤差，參圖 1 之說明 $e_i =$ 觀測值(實際值) － 估計值 = $y_i - \hat{y}_i$， 一般估計的目標之一是讓估計值誤差愈小愈好，但要達成 Min e_i for i = 1 to n 不可行，故考慮 $Min\left\{\sum_{i=1}^{n} e_i\right\}$ ，並盡量達成 $\sum_{i=1}^{n} e_i \to 0$ 即考慮 $Min\left\{\sum_{i=1}^{n} e_i^2\right\}$ ，以下說明最小平方法找出 α, β 的估計量

圖 1 迴歸模式之估計誤差說明圖



參考書籍：

(1).Richard A. Johnson, Probability and statistics for Engineers, Prentice Hall, 2000.

(2).賀力行,林淑萍,蔡春明, 統計學－觀念、理論與方法, 前程企業管理(有).

重點 2. 最小平方法 Least Squares Method：

考慮 $Min\left\{\sum_{i=1}^{n} e_i^2\right\} = Min\left\{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2\right\} = Min\left\{\sum_{i=1}^{n}\left(y_i - a - b\,x_i\right)^2\right\}$，將左式分別對 a,b 取微分，並令上式微分等於零，參考以下說明，即可解出 a,b

令 $w=\sum_{i=1}^{n}\left(y_i - a - b\,x_i\right)^2$

$$\frac{dw}{da} = 2\left(\sum_{i=1}^{n}\left(y_i - a - b\,x_i\right)\right)\times(-1) = 0$$

$$\frac{dw}{db} = 2\left(\sum_{i=1}^{n}\left(y_i - a - b\,x_i\right)\right)\times(-x_i) = 0$$

改寫上式可整理以下 Normal equations

$$\sum_{i=1}^{n} y_i = an + b\sum_{i=1}^{n} x_i \text{..........} (1)$$

$$\sum_{i=1}^{n} x_i y_i = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 \text{..........} (2)$$

將(1)式 $\times \sum_{i=1}^{n} X_i \rightarrow \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right) = an\sum_{i=1}^{n} x_i + b\left(\sum_{i=1}^{n} x_i\right)^2$ ....(3)

將(2)式 $\times n \quad \rightarrow n\sum_{i=1}^{n} x_i y_i = an\sum_{i=1}^{n} x_i + bn\sum_{i=1}^{n} x_i^2$ .... (4)

將(4)式 - (3)式 $\rightarrow n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right) = b\left[n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right]$

$b = \dfrac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$，左式與生管課本第105頁公式(3−9)相同，

另將分子與分母同除以 n 可得 $b = \dfrac{\sum_{i=1}^{n} x_i y_i - \dfrac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n}}{\sum_{i=1}^{n} x_i^2 - \dfrac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}$

所以 $a = \overline{y} - b\overline{x}$， 其中 $\overline{y} = \dfrac{\sum_{i=1}^{n} y_i}{n}, \overline{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$

針對 b 的表示法，可利用 Sums of squares , Sums of Cross-products

(1)Sums of squares

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

$$S_{yy}=\sum_{i=1}^{n}(y_i-\overline{y})^2=\sum_{i=1}^{n}y_i^2-\frac{\left(\sum_{i=1}^{n}y_i\right)^2}{n}$$

(2)Sums of Cross-Products

$$S_{xy}=\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})=\sum_{i=1}^{n}x_iy_i-\frac{\left(\sum_{i=1}^{n}x_i\right)\left(\sum_{i=1}^{n}y_i\right)}{n},\text{即}b=\frac{S_{xy}}{S_{xx}}$$

重點 3. 如何估算殘差的平方和

Residual Sum of Squares (即 Sum of Squares Error, SSE)

Residual Sum of Squares = $\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2$

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

$$=\sum_{i=1}^{n}(y_i-a-bx_i)^2$$

$$=\sum_{i=1}^{n}\left[(y_i-\overline{y})-b(x_i-\overline{x})+(\overline{y}-a-b\overline{x})\right]^2$$

$$=\sum_{i=1}^{n}\left[\begin{array}{l}(y_i-\overline{y})^2+b^2(x_i-\overline{x})^2+(\overline{y}-a-b\overline{x})^2\\-2b(y_i-\overline{y})(x_i-\overline{x})+2(y_i-\overline{y})(\overline{y}-a-b\overline{x})-2b(x_i-\overline{x})\end{array}\right]$$

$$=S_{yy}+b^2S_{xx}+n(\overline{y}-a-b\overline{x})^2-2bS_{xy}$$

$$=n(\overline{y}-a-b\overline{x})^2+\left(b^2S_{xx}-2bS_{xy}+\frac{S_{xy}^2}{S_{xx}}\right)+S_{yy}-\frac{S_{xy}^2}{S_{xx}}$$

$$=n(\overline{y}-a-b\overline{x})^2+\left(b\sqrt{S_{xx}}-\frac{S_{xy}}{\sqrt{S_{xx}}}\right)^2+S_{yy}-\frac{S_{xy}^2}{S_{xx}}$$

$$=n\left[\overline{y}-(\overline{y}-b\overline{x})-b\overline{x}\right]^2+\left(\frac{S_{xy}}{S_{xx}}\sqrt{S_{xx}}-\frac{S_{xy}}{\sqrt{S_{xx}}}\right)^2+S_{yy}-\frac{S_{xy}^2}{S_{xx}}$$

$$=0+0+S_{yy}-\frac{S_{xy}^2}{S_{xx}}$$

$$=S_{yy}-\frac{S_{xy}^2}{S_{xx}},\text{ 結論：}SSE=S_{yy}-\frac{S_{xy}^2}{S_{xx}}$$

重點 4. 如何估算σ^2,即找出$\hat{\sigma}^2$

$$\hat{\sigma}^2 = Se^2 \frac{\sum_{i=1}^{n}\left[y_i - (a + bx_i)\right]^2}{n-2} = \frac{\text{Residual Sum of Squares}}{n-2} = \frac{S_{yy} - \frac{S_{xy}^2}{S_{xx}}}{n-2}$$

其中 Se 稱爲 Standard error of estimate ,

另 $\frac{(n-2)Se^2}{\sigma^2}$ 爲自由度 $n-2$的卡方分佈

Chi-Square distribution, 以$\chi^2(n-2)$表示

重點 5. 如何估算$\hat{\alpha},\hat{\beta}$之信賴區間 confidence interval

$$\hat{\alpha} \sim N\left(a, \frac{\sigma^2 \sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n}(x_i - \overline{x})^2}\right)$$

$$\hat{\beta} \sim N\left(b, \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}\right)$$

$$\hat{\alpha}\text{之}(1-\alpha)\%\text{信賴區間}: a \pm t_{\frac{\alpha}{2}} \times Se \times \sqrt{\frac{1}{n} + \frac{(\overline{x})^2}{S_{xx}}}$$

$$\hat{\beta}\text{之}(1-\alpha)\%\text{信賴區間}: b \pm t_{\frac{\alpha}{2}} \times Se \times \frac{1}{\sqrt{S_{xx}}}$$

其中$t_{\frac{\alpha}{2}}$分配的自由度爲$n-2$

重點 6. 非線性轉換模式

在現實工作中，不一定會符合線性模式，但經由簡單的函數轉換功能即可將非線性（Non-linear）模式轉變成線性模式

1. 指數函數 Exponential function

$y = \alpha\beta^x$

左右兩邊同時取log

$\to \log y = \log(\alpha\beta^x) = \log\alpha + \log\beta^x = \log\alpha + x\log\beta$

$\to$ 將 $(x_i, \log y_i)$ 以 $\hat{y} = a + bx$處理

2. 倒數函數 Reciprocal function

$$y = \frac{1}{\alpha + \beta x}$$

$$\rightarrow \frac{1}{y} = \alpha + \beta x$$

$\rightarrow$ 將 $(x_i, \frac{1}{y_i})$ 以 $\hat{y} = a + bx$ 處理

3. 次方函數 Power function

$$y = \alpha x^{\beta}$$

左右兩邊同時取log

$$\rightarrow \log y = \log(\alpha x^{\beta}) = \log \alpha + \log x^{\beta} = \log \alpha + \beta \log x$$

$\rightarrow$ 將 $(\log x_i, \log y_i)$ 以 $\hat{y} = a + bx$ 處理

全篇完^_^