## 專題: 從不同的角度估計簡單迴歸模式的參數

**目的**:

---

本練習從簡單的迴歸模式出發, 利用線性代數的技巧, 從不同的角度估計模型的參數, 做爲以 MATLAB 寫作程式的第一個實際案例。雖然簡單但極具代表性, 請細心完成。

**怎麼做**:

---

**問題描述**: 假設女人的身高( $X$ ) 體重 ( $Y$ ) 的關係爲簡單迴歸模式

$$Y = \beta_0 + \beta_1 X \tag{1}$$

已知 20 個女人的身高 ( $x_1, x_2, \cdots, x_{20}$ ) 體重 ( $y_1, y_2, \cdots, y_{20}$ ) 資料。

將已知資料帶入迴歸模式, 可以得到下列的矩陣資料模式

$$A\mathbf{x} = \mathbf{b} \tag{2}$$

其中 $A = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_{20} \end{bmatrix}, \mathbf{x} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{20} \end{bmatrix}$

我們希望估計式 (2) 中的參數向量 $\mathbf{x}$ 。很明顯的, 這是個無解的方程式組, 只能依據不同的法則找出適當的解, 以下介紹三種法則, 請分別利用程式設計計算最佳的解。

**解決方法**:

法則一: 最小平方法 (Least Squared Errors)

$$\min_{\mathbf{x}} ||A\mathbf{x} - \mathbf{b}||_2^2 \tag{3}$$

解:

$$\hat{\mathbf{x}}_{LS} = (A^T A)^{-1} A^T \mathbf{b} \tag{4}$$

改寫式 (2) 的的資料模式爲

$$D\theta = \mathbf{0} \tag{5}$$

其中 $D = [-\mathbf{b} \vdots A]$ 爲 $20 \times 3$ 的資料矩陣, $\theta = \begin{bmatrix} 1 \\ \cdots \\ \mathbf{x} \end{bmatrix}$ 是 $3 \times 1$ 的參數向量。

法則二: 限制式最小平方法

$$\min_{\theta, \theta(1)=1} \theta^T D^T D\theta \tag{6}$$

解:

$$\hat{\theta}_{LS} = \lambda (D^T D)^{-1} \mathbf{e}_1 \tag{7}$$

其中 $\lambda = \frac{1}{\mathbf{g}(1)}, \mathbf{g} = (D^T D)^{-1}\mathbf{e}_1$, $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$

改寫式 (5) 的資料模式成爲

$$D^T D\theta = \mathbf{0} \tag{8}$$

法則三: Near Null Space

$$D^T D\theta \approx \mathbf{0}$$

解:

$$\hat{\theta} = \mathbf{v}/\mathbf{v}(1) \tag{9}$$

其中 $\mathbf{v}$ 代表矩陣 $D^T D$ 的特徵向量中, 其特徵值接近 0 者。在本題中, 只有一個這樣的特徵向量存在。

利用下列的方式比較 (4)(7)(9) 三種法則下的最佳解:

- 計算殘差 (residuals) 均方根 (Root Mean Squared Error):

$$RMSE = \sqrt{\frac{\sum_{k=1}^{n}(y_k - \hat{y}_k)^2}{N}}$$

  其中 $\hat{y}_k$ 代表擬合值。

- 在散佈圖上畫出迴歸線: $y = \hat{\beta}_0 + \hat{\beta}_1 x$

- 其他資料: 僅從本單元提供的資料判斷不同方法的優劣, 似乎粗糙了些, 不同的資料來源或許會產生不一樣的結果, 從實驗的精神來看, 應該多試幾種不同的資料, 做綜合的判斷。

- 模擬資料: 既然假設資料具式 (1) 的關係, 我們可以藉此產生模擬資料, 藉此操控資料的型態與相關性, 如

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

  藉由設定不同的 $\beta_0, \beta_1$, 給予 $X$ 與 $\epsilon$ 亂數型態的資料, 例如 $X \sim N(\mu, \sigma_1^2), \epsilon \sim N(0, \sigma_2^2)$, 可以產生變異情況不同的相關資料。