第 5 章

簡單線性迴歸之矩陣方法

Matrices Approach to Simple Linear Regression

- 在複迴歸中，由於矩陣方法可以透過較為精簡的表達方式，來表示出廣大的聯立方程組，以及龐大的資料陣列，所以經常是必須使用的工具。
- 本章首先介紹矩陣代數，然後介紹如何將矩陣方法應用至簡單線性迴歸模型中，雖然在簡單線性迴歸模型中，尚不需使用到矩陣代數，不過透過矩陣代數的應用，將有助於進入第二與第三部分關於複迴歸模型的討論。
- 將焦點放在如何應用於迴歸分析的部分。

5.1 矩陣 Matrices

- **矩陣之定義**

  矩陣是將元素依照行與列排成一矩形之形式，如下例：

$$
\begin{array}{c c}
 & \begin{array}{cc} \text{第一行} & \text{第二行} \end{array} \\
\begin{array}{c} \text{第一列} \\ \text{第二列} \\ \text{第三列} \end{array} & \begin{pmatrix} 16{,}000 & 23 \\ 33{,}000 & 47 \\ 21{,}000 & 35 \end{pmatrix}
\end{array}
$$

- 例如第一列第二行之數字(23)表示年齡為23歲，其所得表示在第一列第一行之數字（16,000），上述矩陣之階數（*dimension*）為3×2，亦即三列二行，所以如果想透過矩陣表示1,000個人的所得與年齡，則所需矩陣之階數為1000×2。

- 一般而言，矩陣的階數先寫出它的列，再表示它的行，接下來我們定義一些符號來代表下面這個2x3之矩陣中的元素：

$$\begin{array}{c|ccc} & j=1 & j=2 & j=3 \\ \hline i=1 & a_{11} & a_{12} & a_{13} \\ i=2 & a_{21} & a_{22} & a_{23} \end{array}$$

這裡第一個下標示指該元素的列，第二個下標則是行，所以符號 a_{ij} 表示第 i 列第 j 行之元素。在上面的例子中，i =1,2 以及 j＝1,2,3。

- 習慣上用粗黑體的樣式符號來代表一個矩陣 ，因此上述矩陣可以寫成：

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}$$

或是 $\mathbf{A}=[\, a_{ij} \,]$ i = 1,2；j = 1,2,3

所以對於一個具有 r 列 c 行 之矩陣可以表是為:

$$\mathbf{A}=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1j} & \cdots & a_{1c} \\ a_{21} & a_{22} & \cdots & a_{2j} & \cdots & a_{2c} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ij} & \cdots & a_{ic} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{r1} & a_{r2} & \cdots & a_{rj} & \cdots & a_{rc} \end{bmatrix} \tag{5.1}$$

或是採用簡略之形式

$$\mathbf{A}=〔a_{ij}〕 \qquad i = 1,\ldots,r；j = 1,\ldots,c$$

甚至直接以**A**作為表示方式。

- 方陣

當矩陣之列數等於行數時，稱之為方陣(*square matrix*)

$$\begin{bmatrix} 4 & 7 \\ 3 & 9 \end{bmatrix} \quad \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

- 向量

當矩陣之行數僅為1，我們稱為行向量(*column vector*)或簡稱向量(*vector*)，當矩陣之列數僅為1，我們稱為列向量(*row vector*)

$$A = \begin{bmatrix} 4 \\ 7 \\ 10 \end{bmatrix} \qquad C = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{bmatrix} \qquad B' = [15\ 25\ 50]$$

- 向量A是一個3x1之矩陣、向量C是一個5x1之矩陣，列向量的符號右上方加一撇之表示方式稱之為轉置(*transpose*)

轉置

- 矩陣A之轉置我們習慣上用符號A´來表示。

例如

$$\underset{3\times 2}{A} = \begin{bmatrix} 2 & 5 \\ 7 & 10 \\ 3 & 4 \end{bmatrix}$$

所以轉置矩陣A´為：

$$\underset{2\times 3}{A'} = \begin{bmatrix} 2 & 7 & 3 \\ 5 & 10 & 4 \end{bmatrix}$$

- 轉置結果是行向量轉置成列向量，反之，則是列向量轉置成行向量

轉置

在一般的情形下，**A** 矩陣轉置的規則為：

$$\underset{r\times c}{\mathbf{A}}=\begin{bmatrix} a_{11} & \cdots & a_{1c} \\ \vdots & & \vdots \\ a_{r1} & \cdots & a_{rc} \end{bmatrix}=\left[a_{\underset{\text{列下標}}{i}\,\underset{\text{行下標}}{j}}\right] \quad i=1,\ldots,r; \quad j=1,\ldots,c \tag{5.2}$$

其轉置矩陣為：

$$\underset{c\times r}{\mathbf{A}'}=\begin{bmatrix} a_{11} & \cdots & a_{r1} \\ \vdots & & \vdots \\ a_{1c} & \cdots & a_{rc} \end{bmatrix}=\left[a_{\underset{\text{列下標}}{j}\,\underset{\text{行下標}}{i}}\right] \quad j=1,\ldots,c; \quad i=1,\ldots,r \tag{5.3}$$

所以在矩陣 **A** 中的第 i 列第 j 行的元素，轉置後出現在矩陣 **A'** 中的第 j 列第 i 行。

矩陣相等

$$A=\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \qquad B=\begin{pmatrix} 4 \\ 7 \\ 3 \end{pmatrix}$$

表示 $a_1=4$， $a_2=7$， $a_3=3$

在迴歸分析中由反應變數的n個觀測值所組成的向量$\mathbf{Y}$，可以表示成基本的矩陣形式如下：

$$\underset{n\times 1}{\mathbf{Y}}=\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \tag{5.4}$$

其轉置矩陣$\mathbf{Y}'$為：

$$\underset{1\times n}{\mathbf{Y}'}=\begin{bmatrix} Y_1 & Y_2 & \cdots & Y_n \end{bmatrix} \tag{5.5}$$

另外一個基本矩陣是由一行全為1，以及另外一行是預測變數的n個觀測值所組成的$\mathbf{X}$矩陣：

$$\underset{n\times 2}{\mathbf{X}}=\begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \tag{5.6}$$

其轉置矩陣為：

$$\underset{2\times n}{\mathbf{X}'}=\begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \end{bmatrix} \tag{5.7}$$

矩陣$\mathbf{X}$有時稱為**設計矩陣**(*design matrix*)。

5.2 矩陣之加法與減法

Matrix Addition and Subtraction

- 當兩個矩陣可進行相加或相減時，前提條件是兩個矩陣需具備有相同的階數。

$$\underset{3\times2}{A} = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix} \qquad \underset{3\times2}{B} = \begin{pmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 4 \end{pmatrix}$$

則

$$\underset{3\times2}{A+B} = \begin{pmatrix} 1+1 & 4+2 \\ 2+2 & 5+3 \\ 3+3 & 6+4 \end{pmatrix} = \begin{pmatrix} 2 & 6 \\ 4 & 8 \\ 6 & 10 \end{pmatrix}$$

- 一般而言，如果

$$\underset{r\times c}{\mathbf{A}}=\left[a_{ij}\right] \quad \underset{r\times c}{\mathbf{B}}=\left[b_{ij}\right] \quad i=1,\ldots,r; \quad j=1,\ldots,c$$

則

$$\underset{r\times c}{\mathbf{A}+\mathbf{B}}=\left[a_{ij}+b_{ij}\right] \quad 或 \quad \underset{r\times c}{\mathbf{A}-\mathbf{B}}=\left[a_{ij}-b_{ij}\right] \tag{5.8}$$

(5.8)式可以一般化推廣至多於兩個矩陣之加減運算，而且與一般代數規則有相同的性質：$\mathbf{A}+\mathbf{B}=\mathbf{B}+\mathbf{A}$。

迴歸模型

$$Y_i = E\{Y_i\} + \varepsilon_i \qquad i = 1, \ldots, n$$

可以透過矩陣符號精簡地表示出。首先定義平均反應向量與誤差項向量：

$$\underset{n\times 1}{\mathbf{E}\{\mathbf{Y}\}} = \begin{bmatrix} E\{Y_1\} \\ E\{Y_2\} \\ \vdots \\ E\{Y_n\} \end{bmatrix} \tag{5.9}$$

$$\underset{n\times 1}{\boldsymbol{\varepsilon}} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{5.10}$$

根據(5.4)式，n 個觀測值所組成的向量 $\mathbf{Y}$ 可以表示為：

$$\underset{n\times 1}{\mathbf{Y}} = \underset{n\times 1}{\mathbf{E}\{\mathbf{Y}\}} + \underset{n\times 1}{\boldsymbol{\varepsilon}}$$

因為，

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} E\{Y_1\} \\ E\{Y_2\} \\ \vdots \\ E\{Y_n\} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} = \begin{bmatrix} E\{Y_1\} + \varepsilon_1 \\ E\{Y_2\} + \varepsilon_2 \\ \vdots \\ E\{Y_n\} + \varepsilon_n \end{bmatrix}$$

所以觀測值向量 $\mathbf{Y}$ 可以等於平均反應向量與誤差項向量兩個向量之和。

5.3 矩陣相乘 Matrix Multiplication

- 純量與矩陣之相乘

一般而言，當$\mathbf{A} = [a_{ij}]$，而 k 為一個純量，則

$$k\mathbf{A} = \mathbf{A}k = \left[ka_{ij}\right] \tag{5.11}$$

- 兩矩陣之相乘

一般而言，當矩陣A之階數為$r \times c$，而矩陣B之階數為$c \times s$，則乘積AB之階數為$r \times s$，且其第i列第j行的元素為：

$$\sum_{k=1}^{c} a_{ik} b_{kj}$$

所以我們可以得到：

$$\underset{r \times s}{\mathbf{AB}} = \left[\sum_{k=1}^{c} a_{ik} b_{kj}\right] \quad i = 1, \cdots, r;\ j = 1, \cdots, s \tag{5.12}$$

更多的例題

1. $$\begin{bmatrix} 4 & 2 \\ 5 & 8 \end{bmatrix}\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 4a_1 + 2a_2 \\ 5a_1 + 8a_2 \end{bmatrix}$$

2. $$\begin{bmatrix} 2 & 3 & 5 \end{bmatrix}\begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix} = \begin{bmatrix} 2^2 + 3^2 + 5^2 \end{bmatrix} = [38]$$

在此乘積結果為一個 1×1 之矩陣，也是一個純量，所以此二矩陣之乘積為純量 38 。

3. $$\begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ 1 & X_3 \end{bmatrix}\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 X_1 \\ \beta_0 + \beta_1 X_2 \\ \beta_0 + \beta_1 X_3 \end{bmatrix}$$

迴歸例題

考慮反應變數觀測值所組成的向量 $\mathbf{Y}$ 被其本身的轉置矩陣 $\mathbf{Y}'$ 進行乘積運算：

$$\underset{1\times1}{\mathbf{Y}'\mathbf{Y}}=\begin{bmatrix}Y_1 & Y_2 & \cdots & Y_n\end{bmatrix}\begin{bmatrix}Y_1\\Y_2\\\vdots\\Y_n\end{bmatrix}=\begin{bmatrix}Y_1^2+Y_2^2+\cdots+Y_n^2\end{bmatrix}=\begin{bmatrix}\sum Y_i^2\end{bmatrix} \qquad (5.13)$$

上式中 $\mathbf{Y}'\mathbf{Y}$ 是一個純量，也就是反應變數觀測值之平方和：$\boxed{\mathbf{Y}'\mathbf{Y}=\sum Y_i^2}$，另外關於 $\mathbf{X}'\mathbf{X}$ 與 $\mathbf{X}'\mathbf{Y}$ 可以分別計算如下：

$$\underset{2\times2}{\mathbf{X}'\mathbf{X}}=\begin{bmatrix}1 & 1 & \cdots & 1\\X_1 & X_2 & \cdots & X_n\end{bmatrix}\begin{bmatrix}1 & X_1\\1 & X_2\\\vdots & \vdots\\1 & X_n\end{bmatrix}=\begin{bmatrix}n & \sum X_i\\\sum X_i & \sum X_i^2\end{bmatrix} \qquad (5.14)$$

同時階數 2 × 1 的矩陣 $\mathbf{X}'\mathbf{Y}$ 為：

$$\underset{2\times1}{\mathbf{X}'\mathbf{Y}}=\begin{bmatrix}1 & 1 & \cdots & 1\\X_1 & X_2 & \cdots & X_n\end{bmatrix}\begin{bmatrix}Y_1\\Y_2\\\vdots\\Y_n\end{bmatrix}=\begin{bmatrix}\sum Y_i\\\sum X_iY_i\end{bmatrix} \qquad (5.15)$$

5.4 特殊矩陣 Special Types of Matrices

- 對稱矩陣

➢ 當 $\mathbf{A} = \mathbf{A}'$，則矩陣**A**為對稱矩陣

➢ 所有的對稱矩陣均為方陣。在迴歸分析中，對稱矩陣經常是出現在矩陣 **X** 前面乘上其轉置 **X′**，**X X′** 必為對稱矩陣，這結果由（5.14）得知

- 對角線矩陣

當矩陣之主對角線以外的元素均為零，則此矩陣為一個對角線矩陣（diagonal matrix），

- 單位矩陣

單位矩陣(diagonal matrix)（或unit matrix）之符號習慣上用 **I** 來表示，該矩陣主對角線之元素均為1，而主對角線以外的元素均為零，任意矩陣乘上或被乘上相同階數的單位矩陣，其結果不變，

所以對於一個$r \times r$之矩陣**A**，我們有

$$\mathbf{AI} = \mathbf{IA} = \mathbf{A} \tag{5.16}$$

- 純量矩陣：純量矩陣(scalar matrix)為主對角線元素均相同之對角線矩陣
- **元素全為一之向量與矩陣：**當行向量內所有的元素均為1，我們用符號 **1** 來表示，

$$\underset{r\times 1}{\mathbf{1}} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \tag{5.17}$$

- 而對於方陣內所有的元素均為1，我們用符號 **J** 來表示，

$$\underset{r\times r}{\mathbf{J}} = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{bmatrix} \tag{5.18}$$

對於一個 n×1 之向量 I，我們有

$$I'I = \begin{bmatrix} 1 \cdots 1 \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} n \end{bmatrix} = n$$

而

$$I\,I' = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \begin{bmatrix} 1 \cdots 1 \end{bmatrix} = \begin{pmatrix} 1 \cdots 1 \\ \vdots \quad \vdots \\ 1 \cdots 1 \end{pmatrix} = \underset{n \times n}{J}$$

- 零向量

零向量是指向量內所有的元素均為0，用符號 **0** 來表示：

$$\underset{r \times 1}{\mathbf{0}} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad (5.19)$$

5.5　線性相依與矩陣的秩

Linear Dependence and Rank of Matrix

- 線性相依

➢ 當矩陣中存在有某一個行向量，可以等於其他行向量之線性組合時，該矩陣之各行為 線性相依 (*linear dependent*) 。若矩陣中不存在有任何一個行向量，可以等於其他行向量之線性組合時，該矩陣之各向量行為 線性獨立 (*linear independent*)

➢ 如果存在c個不全為 0 之純量 $k_1,\ldots,k_c$，可以滿足：

$$K_1\mathbf{C}_1 + K_2\mathbf{C}_2 + \ldots + K_c\mathbf{C}_c = \mathbf{0} \tag{5.20}$$

則 c 個向量 $C_1, \ldots, C_c$為 **線性相依**；反之，若滿足 $k_1C_1 + k_2C_2 + \ldots + k_cC_c = 0$ 之唯一條件為 $k_1 = k_2 = \ldots = k_c = 0$ ，則 c 個向量$C_1, \ldots, C_c$為 **線性獨立**。

矩陣的秩

- 矩陣中可以滿足線性獨立最多數目之行向量，該數稱為矩陣的秩(rank)，舉前例而言，由於四個行向量不為線性獨立，所以矩陣 A 之秩不是 4，不過第一、第二、第四行為線性獨立，所以矩陣 A 之秩為 3。
- 一個矩陣的秩是固定的，而且可以透過最大線性獨立之列數來做等價之定義，亦即一個 rxc 之矩陣，矩陣的秩不會大於 r 與 c 兩者之最小值 min(r，c)。
- 當矩陣**A**與**B**相乘後得到矩陣**C**，則矩陣**C**的秩不會大於矩陣A的秩與矩陣B的秩兩者之最小值min(rankA,rankB)。

5.6 反矩陣 Inverse of Matrix

- 一個數的反元素(inverse)是指該數的倒數，所以6的反元素是。例如：

$$6 \cdot \frac{1}{6} = \frac{1}{6} \cdot 6 = 1$$

$$\chi \cdot \frac{1}{x} = \chi \cdot \chi^{-1} = \chi^{-1} \cdot \chi = 1$$

- 在矩陣代數中，矩陣**A**的反元素用符號 $\mathbf{A}^{-1}$來表示，同時有性質：

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I} \tag{5.21}$$

- 唯有方陣中才有所謂反矩陣之定義，不過有些方陣其反矩陣並不存在，當方陣之反矩陣存在時，必定是唯一一個

例題

1. 矩陣

$$\underset{2\times 2}{\mathbf{A}} = \begin{bmatrix} 2 & 4 \\ 3 & 1 \end{bmatrix}$$

其反矩陣為

$$\underset{2\times 2}{\mathbf{A}}^{-1} = \begin{bmatrix} -.1 & .4 \\ .3 & -.2 \end{bmatrix}$$

因為

$$\mathbf{A}^{-1}\mathbf{A} = \begin{bmatrix} -.1 & .4 \\ .3 & -.2 \end{bmatrix}\begin{bmatrix} 2 & 4 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I}$$

或

$$\mathbf{A}\mathbf{A}^{-1} = \begin{bmatrix} 2 & 4 \\ 3 & 1 \end{bmatrix}\begin{bmatrix} -.1 & .4 \\ .3 & -.2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I}$$

2. 矩陣

$$\underset{3\times 3}{\mathbf{A}} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

其反矩陣為

$$\underset{3\times 3}{\mathbf{A}} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 2 \end{bmatrix} \qquad \underset{3\times 3}{\mathbf{A}}^{-1} = \begin{bmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{2} \end{bmatrix}$$

因為

$$\mathbf{A}^{-1}\mathbf{A} = \begin{bmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \mathbf{I}$$

所以對角線矩陣之反矩陣，僅需將原矩陣中之主對角線元素取倒數，而成為一個新的對角線矩陣。

反矩陣之求法

一個 $r \times r$ 之矩陣當其秩為 r 時，則存在反矩陣，這種矩陣我們稱之為非奇異(*nonsingular*)或全秩(*full rank*)，其反矩陣的秩也是r；反之，當 $r \times r$ 之矩陣當其秩小於 r 時，則不存在反矩陣，這種矩陣我們稱之奇異(*singular*)或非全秩(*not of full rank*)。

- **反矩陣之求法** (對 2×2 與3×3 之矩陣)

1.矩陣

$$\underset{2\times 2}{\mathbf{A}}=\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

則

$$\underset{2\times 2}{\mathbf{A}}^{-1}=\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1}=\begin{bmatrix} \dfrac{d}{D} & \dfrac{-b}{D} \\ \dfrac{-c}{D} & \dfrac{a}{D} \end{bmatrix} \qquad (5.22)$$

其中，

$$D=ad-bc \qquad (5.22a)$$

D稱為矩陣**A**之行列式值，當矩陣**A**為奇異，則$D=0$，而矩陣**A**之反矩陣不存在。

2.矩陣

$$\underset{3\times3}{\mathbf{B}}=\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & k \end{bmatrix}$$

則

$$\underset{3\times3}{\mathbf{B}}^{-1}=\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & k \end{bmatrix}^{-1}=\begin{bmatrix} A & B & C \\ D & E & F \\ G & H & K \end{bmatrix} \tag{5.23}$$

其中，

$$\begin{aligned} &A=(ek-fh)/Z && B=-(bk-ch)/Z && C=(bf-ce)/Z \\ &D=-(dk-fg)/Z && E=(ak-cg)/Z && F=-(af-cd)/Z \\ &G=(dh-eg)/Z && H=-(ah-bg)/Z && K=(ae-bd)/Z \end{aligned} \tag{5.23a}$$

且

$$Z=a(ek-fh)-b(dk-fg)+c(dh-eg) \tag{5.23b}$$

Z 稱為矩陣 $\mathbf{B}$ 之行列式值

迴歸分析中常用的矩陣為(5.14)的 $\mathbf{X}'\mathbf{X}$：

$$\underset{2\times 2}{\mathbf{X}'\mathbf{X}} = \begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix}$$

利用(5.22)的方法，

$$a = n \qquad b = \sum X_i$$
$$c = \sum X_i \quad d = \sum X_i^2$$

於是，

$$D = n\sum X_i^2 - \left(\sum X_i\right)\left(\sum X_i\right) = n\left[\sum X_i^2 - \frac{\left(\sum X_i\right)^2}{n}\right] = n\sum\left(X_i - \bar{X}\right)^2$$

所以，

$$\underset{2\times 2}{\left(\mathbf{X}'\mathbf{X}\right)^{-1}} = \begin{bmatrix} \dfrac{\sum X_i^2}{n\sum\left(X_i - \bar{X}\right)^2} & \dfrac{-\sum X_i}{n\sum\left(X_i - \bar{X}\right)^2} \\ \dfrac{-\sum X_i}{n\sum\left(X_i - \bar{X}\right)^2} & \dfrac{n}{n\sum\left(X_i - \bar{X}\right)^2} \end{bmatrix} \tag{5.24}$$

所以，

$$\underset{2\times 2}{(\mathbf{X}'\mathbf{X})}^{-1} = \begin{bmatrix} \dfrac{\sum X_i^2}{n\sum(X_i-\bar{X})^2} & \dfrac{-\sum X_i}{n\sum(X_i-\bar{X})^2} \\ \dfrac{-\sum X_i}{n\sum(X_i-\bar{X})^2} & \dfrac{n}{n\sum(X_i-\bar{X})^2} \end{bmatrix} \tag{5.24}$$

因為 $\sum X_i = n\bar{X}$ 以及 $\sum(X_i-\bar{X})^2 = \sum X_i^2 - n\bar{X}^2$，(5.24)式也可以簡化成：

$$\boxed{\underset{2\times 2}{(\mathbf{X}'\mathbf{X})}^{-1} = \begin{bmatrix} \dfrac{1}{n}+\dfrac{\bar{X}^2}{\sum(X_i-\bar{X})^2} & \dfrac{-\bar{X}}{\sum(X_i-\bar{X})^2} \\ \dfrac{-\bar{X}}{\sum(X_i-\bar{X})^2} & \dfrac{1}{\sum(X_i-\bar{X})^2} \end{bmatrix}} \tag{5.24a}$$

- **反矩陣之使用**

在一般代數裡，方程式：　$5y = 20$

其解法為等號兩邊均乘上5之反元素，亦即，$\frac{1}{5} = (5y) = \frac{1}{5}(20)$

則可以解出：

$$y = \frac{1}{5}(20) = 4$$

在矩陣代數中，方程式：　**AY=C**

假設矩陣**A**之反矩陣存在，則在等號兩邊乘上**A**之反矩陣：

$\mathbf{A^{-1}AY=A^{-1}C}$ ，因為 $\mathbf{A^{-1}AY=IY=Y}$，所以可以解出：$\boxed{\mathbf{Y= A^{-1}C}}$

舉例來說，假設聯立方程組：

$$2y_1 + 4y_2 = 20$$
$$3y_1 + y_2 = 10$$

我們可以用矩陣表示成：

$$\begin{pmatrix} 2 & 4 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 20 \\ 10 \end{pmatrix}$$

於是，

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 2 & 4 \\ 3 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 20 \\ 10 \end{pmatrix}$$

所以解出：

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} -0.1 & 0.4 \\ 0.3 & -0.2 \end{pmatrix} \begin{pmatrix} 20 \\ 10 \end{pmatrix} = \begin{pmatrix} 2 \\ 4 \end{pmatrix}$$

因此 $y_1 = 2$，$y_2 = 4$ 可以滿足原來的聯立方程組。

5.7 矩陣之基本定理

- 一些常用的矩陣基本定理：

$$\mathbf{A}+\mathbf{B}=\mathbf{B}+\mathbf{A} \tag{5.25}$$

$$(\mathbf{A}+\mathbf{B})+\mathbf{C}=\mathbf{A}+(\mathbf{B}+\mathbf{C}) \tag{5.26}$$

$$(\mathbf{AB})\mathbf{C}=\mathbf{A}(\mathbf{BC}) \tag{5.27}$$

$$\mathbf{C}(\mathbf{A}+\mathbf{B})=\mathbf{CA}+\mathbf{CB} \tag{5.28}$$

$$k(\mathbf{A}+\mathbf{B})=k\mathbf{A}+k\mathbf{B} \tag{5.29}$$

$$(\mathbf{A}')' = \mathbf{A} \tag{5.30}$$

$$(\mathbf{A}+\mathbf{B})' = \mathbf{A}' + \mathbf{B}' \tag{5.31}$$

$$(\mathbf{AB})' = \mathbf{B}'\mathbf{A}' \tag{5.32}$$

$$(\mathbf{ABC})' = \mathbf{C}'\mathbf{B}'\mathbf{A}' \tag{5.33}$$

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1} \tag{5.34}$$

$$(\mathbf{ABC})^{-1} = \mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1} \tag{5.35}$$

$$(\mathbf{A}^{-1})^{-1} = \mathbf{A} \tag{5.36}$$

$$(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})' \tag{5.37}$$

5.8 隨機向量與矩陣 Random Vertors and Matrices

- 一個隨機向量或隨機矩陣是指的其內部元素是隨機變數，所以在（5.4）中的觀測值向量Y即為一個隨機向量
- **隨機向量與矩陣之期望值**

假設有n=3個觀測值，構成隨機向量Y：

$$\underset{3\times1}{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix}$$

則Y的期望值為一個向量，用符號E{Y}表示，其定義如下

$$\underset{3\times1}{E\{\mathbf{Y}\}} = \begin{pmatrix} E\{Y_1\} \\ E\{Y_2\} \\ E\{Y_3\} \end{pmatrix}$$

一般情形下，一個隨機向量**Y**之期望值我們可以表示為：

$$\underset{n\times 1}{\mathbf{E}}\{\mathbf{Y}\}=\left[E\{Y_i\}\right] \quad i=1,\cdots,n \tag{5.38}$$

而對於一個階數為$n\times p$之隨機矩陣**Y**，我們可以將其期望值表示為：

$$\underset{n\times p}{\mathbf{E}}\{\mathbf{Y}\}=\left[E\{Y_{ij}\}\right] \quad i=1,\cdots,n;\ j=1,\cdots,p \tag{5.39}$$

例題 假設迴歸應用中的觀測值 $n = 3$，三個誤差項$\varepsilon_1, \varepsilon_2, \varepsilon_3$之期望值均為零，則誤差向量：

$$\underset{3\times1}{\boldsymbol{\varepsilon}} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix}$$

因為

$$\begin{bmatrix} E\{\varepsilon_1\} \\ E\{\varepsilon_2\} \\ E\{\varepsilon_3\} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

我們得到：

$$\underset{3\times1}{\mathbf{E}\{\boldsymbol{\varepsilon}\}} = \underset{3\times1}{\mathbf{0}}$$

隨機向量之共變異矩陣

- 假設一隨機向量Y之三個隨機變數為Y_1，Y_2，Y_3，若其變異數分別為$\sigma^2\{Y_i\}$，另外還存在有三個隨機變數中的任一兩個變數間之共變異數$\sigma^2\{Y_i，Y_j\}$，則可以 $\sigma^2\{Y_i\}$ 來表示**Y** 共變異矩陣(*variance-covarance matrix*)：

- $$\boldsymbol{\sigma}^2\{\mathbf{Y}\}=\begin{bmatrix}\sigma^2\{Y_1\} & \sigma\{Y_1,Y_2\} & \sigma\{Y_1,Y_3\}\\ \sigma\{Y_2,Y_1\} & \sigma^2\{Y_2\} & \sigma\{Y_2,Y_3\}\\ \sigma\{Y_3,Y_1\} & \sigma\{Y_3,Y_2\} & \sigma^2\{Y_3\}\end{bmatrix} \qquad (5.40)$$

由於

$$\boldsymbol{\sigma}^2\{\mathbf{Y}\}=\mathbf{E}\left\{\left[\mathbf{Y}-\mathbf{E}\{\mathbf{Y}\}\right]\left[\mathbf{Y}-\mathbf{E}\{\mathbf{Y}\}\right]'\right\} \qquad (5.41)$$

在上述的例子中，有：

$$\boldsymbol{\sigma}^2\{\mathbf{Y}\}=\mathbf{E}\left\{\begin{bmatrix}Y_1-E\{Y_1\}\\ Y_2-E\{Y_2\}\\ Y_3-E\{Y_3\}\end{bmatrix}\begin{bmatrix}Y_1-E\{Y_1\} & Y_2-E\{Y_2\} & Y_3-E\{Y_3\}\end{bmatrix}\right\}$$

經過矩陣之相乘運算，並取期望值後，我們可以得到下面結果：

乘積位置	項	期望值
第一列，第一行	$(Y_1 - E\{Y_1\})^2$	$\sigma^2\{Y_1\}$
第一列，第二行	$(Y_1 - E\{Y_1\})(Y_2 - E\{Y_2\})$	$\sigma\{Y_1, Y_2\}$
第一列，第三行	$(Y_1 - E\{Y_1\})(Y_3 - E\{Y_3\})$	$\sigma\{Y_1, Y_3\}$
第二列，第一行	$(Y_2 - E\{Y_2\})(Y_1 - E\{Y_1\})$	$\sigma\{Y_2, Y_1\}$
等等	等等	等等

將上述結果推廣至一個$n \times 1$的隨機向量**Y**，其共變異矩陣為：

$$\boxed{\underset{n\times n}{\boldsymbol{\sigma}^2\{\mathbf{Y}\}} = \begin{bmatrix} \sigma^2\{Y_1\} & \sigma\{Y_1,Y_2\} & \cdots & \sigma\{Y_1,Y_n\} \\ \sigma\{Y_2,Y_1\} & \sigma^2\{Y_2\} & \cdots & \sigma\{Y_2,Y_n\} \\ \vdots & \vdots & & \vdots \\ \sigma\{Y_n,Y_1\} & \sigma\{Y_n,Y_2\} & \cdots & \sigma^2\{Y_n\} \end{bmatrix}} \quad (5.42)$$

此處的σ^2(Y) 是一個對稱矩陣

同樣是觀測值 $n = 3$ 的例子，三個誤差項$\varepsilon_1, \varepsilon_2, \varepsilon_3$之變異數均為常數 $\sigma^2\{\varepsilon_i\} = \sigma^2$，同時誤差項間均為獨立，所以對於所有的 $i \neq j$，我們有 $\sigma\{\varepsilon_i, \varepsilon_j\} = 0$，則隨機向量$\boldsymbol{\varepsilon}$之共變異矩陣為：

$$\underset{3\times3}{\boldsymbol{\sigma}^2\{\boldsymbol{\varepsilon}\}} = \begin{bmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix}$$

由於所有的變異數均為相同常數σ^2，而非主對角線之元素均為零，所以該共變異矩陣可以視為一個純量為σ^2之純量矩陣，可以用一個較為簡單之形式來表示：

$$\underset{3\times3}{\boldsymbol{\sigma}^2\{\boldsymbol{\varepsilon}\}} = \sigma^2 \underset{3\times3}{\mathbf{I}}$$

因為

$$\sigma^2\mathbf{I} = \sigma^2 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix}$$

- **基本定理**

有時我們經常會遇到一個常數矩陣**A**（元素固定之矩陣），乘上一個隨機向量**Y**，而形成一個隨機向量**W**如下：

$$\mathbf{W} = \mathbf{A}\mathbf{Y} \tag{5.43}$$

則可能會用到一些基本定理：

$$\mathbf{E}\{\mathbf{A}\} = \mathbf{A}$$

$$\mathbf{E}\{\mathbf{W}\} = \mathbf{E}\{\mathbf{A}\mathbf{Y}\} = \mathbf{A}\mathbf{E}\{\mathbf{Y}\} \tag{5.44}$$

$$\boldsymbol{\sigma}^2\{\mathbf{W}\} = \boldsymbol{\sigma}^2\{\mathbf{A}\mathbf{Y}\} = \mathbf{A}\boldsymbol{\sigma}^2\{\mathbf{Y}\}\mathbf{A}' \tag{5.45}$$

(5.46)

其中，$\boldsymbol{\sigma}^2\{\mathbf{Y}\}$ 為**Y**的共變異矩陣。

例題

我們考慮：

$$\underset{2\times 1}{\mathbf{W}} = \begin{bmatrix} W_1 \\ W_2 \end{bmatrix} = \underset{2\times 2}{\mathbf{A}}\,\underset{2\times 1}{\mathbf{Y}} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} Y_1 - Y_2 \\ Y_1 + Y_2 \end{bmatrix}$$

則根據(5.45)，可以得到：

$$\underset{2\times 1}{\mathbf{E}\{\mathbf{W}\}} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} E\{Y_1\} \\ E\{Y_2\} \end{bmatrix} = \begin{bmatrix} E\{Y_1\} - E\{Y_2\} \\ E\{Y_1\} + E\{Y_2\} \end{bmatrix}$$

又從(5.46)，我們有結果：

$$\underset{2\times 2}{\boldsymbol{\sigma}^2\{\mathbf{W}\}} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \sigma^2\{Y_1\} & \sigma\{Y_1, Y_2\} \\ \sigma\{Y_2, Y_1\} & \sigma^2\{Y_2\} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \sigma^2\{Y_1\} + \sigma^2\{Y_2\} - 2\sigma\{Y_1, Y_2\} & \sigma^2\{Y_1\} - \sigma^2\{Y_2\} \\ \sigma^2\{Y_1\} - \sigma^2\{Y_2\} & \sigma^2\{Y_1\} + \sigma^2\{Y_2\} + 2\sigma\{Y_1, Y_2\} \end{bmatrix}$$

所以，

$$\sigma^2\{W_1\} = \sigma^2\{Y_1 - Y_2\} = \sigma^2\{Y_1\} + \sigma^2\{Y_2\} - 2\sigma\{Y_1, Y_2\}$$

$$\sigma^2\{W_2\} = \sigma^2\{Y_1 + Y_2\} = \sigma^2\{Y_1\} + \sigma^2\{Y_2\} + 2\sigma\{Y_1, Y_2\}$$

$$\sigma\{W_1, W_2\} = \sigma\{Y_1 - Y_2, Y_1 + Y_2\} = \sigma^2\{Y_1\} - \sigma^2\{Y_2\}$$

多元常態分配

- **密度函數**　對於一個多元常態分配的密度函數，最好是透過矩陣的方式來表示，先定義一些向量與矩陣，假設觀測值向量**Y**是由p個Y變數所組成，則習慣上可以寫成：

$$\underset{p\times 1}{\mathbf{Y}} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_p \end{bmatrix} \tag{5.47}$$

用符號μ表示向量**Y**之期望值**E{Y}**，所以

$$\underset{p\times 1}{\boldsymbol{\mu}} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} \tag{5.48}$$

最後，$\mathbf{Y}$的共變異矩陣 $\boldsymbol{\sigma}^2\{\mathbf{Y}\}$ 用符號 $\boldsymbol{\Sigma}$ 來表示：

$$\boxed{\underset{p\times p}{\boldsymbol{\Sigma}} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2p} \\ \vdots & \vdots & & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_p^2 \end{bmatrix}} \tag{5.49}$$

在這裡，σ_1^2為Y_1的變異數，而σ_{12}則是Y_1與Y_2共變係數。對於一個多元常態分配的密度函數，可以寫成：

$$f(\mathbf{Y}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{Y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{Y}-\boldsymbol{\mu})\right] \tag{5.50}$$

其中，$|\boldsymbol{\Sigma}|$ 表示共變異矩陣$\boxed{\boldsymbol{\Sigma}}$之行列式值

5.9 簡單線性迴歸模型的矩陣表示

Simple Liner Regression Model in Matrix Terms

- 在本節中將利用矩陣來處理簡單線性迴歸，首先考慮的是常態誤差迴歸模型(2.1)：

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad i = 1, \cdots, n \tag{5.51}$$

亦即，

$$\begin{aligned} Y_1 &= \beta_0 + \beta_1 X_1 + \varepsilon_1 \\ Y_2 &= \beta_0 + \beta_1 X_2 + \varepsilon_2 \\ &\vdots \\ Y_n &= \beta_0 + \beta_1 X_n + \varepsilon_n \end{aligned} \tag{5.51a}$$

- 另外還需增加迴歸係數所組成的向量：

$$\underset{n\times 1}{\mathbf{Y}}=\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \quad \underset{n\times 2}{\mathbf{X}}=\begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \quad \underset{2\times 1}{\boldsymbol{\beta}}=\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \quad \underset{n\times 1}{\boldsymbol{\varepsilon}}=\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{5.52}$$

於是可以將(5.51a)利用 $\boldsymbol{\beta}$ 矩陣做一個較為精簡的表達方式：

$$\boxed{\underset{n\times 1}{\mathbf{Y}}=\underset{n\times 2}{\mathbf{X}}\ \underset{2\times 1}{\boldsymbol{\beta}}+\underset{n\times 1}{\boldsymbol{\varepsilon}}} \tag{5.53}$$

因為

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}=\begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}+\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}=\begin{bmatrix} \beta_0+\beta_1 X_1 \\ \beta_0+\beta_1 X_2 \\ \vdots \\ \beta_0+\beta_1 X_n \end{bmatrix}+\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}=\begin{bmatrix} \beta_0+\beta_1 X_1+\varepsilon_1 \\ \beta_0+\beta_1 X_2+\varepsilon_2 \\ \vdots \\ \beta_0+\beta_1 X_n+\varepsilon_n \end{bmatrix}$$

- 上式中 $\mathbf{X\beta}$ 表示各個觀測值Y_i之期望值所組成之向量，而$E\{Y_i\} = \beta_0 + \beta_1 X_i$，所以

$$\boxed{\underset{n\times 1}{\mathbf{E}\{\mathbf{Y}\}} = \underset{n\times 1}{\mathbf{X\beta}}} \tag{5.54}$$

在 X 矩陣中有一全為1，我們可以將其視為一個虛擬變數 $X_0=1$，用以作為替代之迴歸模型(1.5)

$$Y_i = \beta_0 X_0 + \beta_1 X_i + \varepsilon_i \quad \text{其中} X_0=1$$

故 X 矩陣包含了虛擬變數 $X_0=1$ 之行向量及預測變數觀測值X_i 之行向量。

- 在誤差項部分中，迴歸模型(2.1)假設 E{ ε_i}＝0，σ^2{ ε_i} ＝σ^2且 ε_i之間為獨立之常態隨機變數，我們可以將條件寫成矩陣之形式：

$$\underset{n\times 1}{\mathbf{E}\{\boldsymbol{\varepsilon}\}} = \underset{n\times 1}{\mathbf{0}} \tag{5.55}$$

- 上式所代表之意義為：

$$\begin{bmatrix} E\{\varepsilon_1\} \\ E\{\varepsilon_2\} \\ \vdots \\ E\{\varepsilon_n\} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

- 由於ε_i具備有常數變異數，而且ε_i與ε_j彼此間無相關性，亦即共變異數為零(換句話說，對於所有的i與j，i≠j，$\sigma\{\varepsilon_i，\varepsilon_j\}=0$)，誤差項之共變異矩陣可以表示如下：

$$\underset{n\times n}{\boldsymbol{\sigma}^2\{\boldsymbol{\varepsilon}\}}=\begin{bmatrix}\sigma^2 & 0 & 0 & \cdots & 0\\ 0 & \sigma^2 & 0 & \cdots & 0\\ \vdots & \vdots & \vdots & & \vdots\\ 0 & 0 & 0 & \cdots & \sigma^2\end{bmatrix} \tag{5.56}$$

而此一共變異矩陣為一個純量矩陣，所以可以進一步表示為：

$$\underset{n\times n}{\boldsymbol{\sigma}^2\{\boldsymbol{\varepsilon}\}}=\sigma^2\underset{n\times n}{\mathbf{I}} \tag{5.56a}$$

- 因此常態誤差迴歸模型(2.1)寫成矩陣形式為：

$$\mathbf{Y}=\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon} \tag{5.57}$$

其中ε為獨立常態隨機變數之向量，且$\mathbf{E}\{\varepsilon\}=0$，$\sigma^2\{\varepsilon\}=\sigma^2\mathbf{I}$

5.10 迴歸參數的最小平方估計

Least Squares Estimation of Regression Parameters

- **標準方程式**

在第1章中的(1.9)式之標準方程式：

$$\begin{aligned} nb_0 + b_1\sum X_i &= \sum Y_i \\ b_0\sum X_i + b_1\sum X_i^2 &= \sum X_iY_i \end{aligned} \tag{5.58}$$

用矩陣形式表示則為：

$$\boxed{\underset{2\times2}{\mathbf{X'X}}\ \underset{2\times1}{\mathbf{b}} = \underset{2\times1}{\mathbf{X'Y}}} \tag{5.59}$$

其中**b**為最小平方法估計下的迴歸係數向量：

$$\underset{2\times1}{\mathbf{b}} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} \tag{5.59a}$$

$$nb_0 + b_1 \sum X_i = \sum Y_i$$

$$b_0 \sum X_i + b_1 \sum X_i^2 = \sum X_i Y_i$$

$$\boxed{\underset{2\times 2}{\mathbf{X}'\mathbf{X}} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \end{bmatrix} \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} = \begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix}}$$

$$\boxed{\underset{2\times 1}{\mathbf{X}'\mathbf{Y}} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix}}$$

$$\underset{2\times 2}{\mathbf{X}'\mathbf{X}} \ \underset{2\times 1}{\mathbf{b}} = \underset{2\times 1}{\mathbf{X}'\mathbf{Y}}$$

- 迴歸係數之估計

透過(5.59)的矩陣形式之標準方程式運算出所估計的迴歸係數，必須先假設 $\mathbf{X'X}$ 的反矩陣存在，然後在等號兩邊同時乘上 $\mathbf{X'X}$ 的反矩陣：

$$(\mathbf{X'X})^{-1}\mathbf{X'Xb} = (\mathbf{X'X})^{-1}\mathbf{X'Y}$$

由於 $(\mathbf{X'X})^{-1}\mathbf{X'X} = \mathbf{I}$，而 $\mathbf{Ib} = \mathbf{b}$，所以

$$\boxed{\underset{2\times 1}{\mathbf{b}} = \underset{2\times 2}{(\mathbf{X'X})^{-1}}\underset{2\times 1}{\mathbf{X'Y}}} \qquad (5.60)$$

其結果與先前的估計量 b_0(1.10a)及估計量 b_1 (1.10b) 相同。

例題 我們在此透過矩陣運算形式重新計算一次Toluca公司的案例，案例中變數 Y 與 X 的資料均列出在表 1.1 中，於是我們可以定義觀測向量 $\mathbf{Y}$ 與 $\mathbf{X}$ 之矩陣如下：

$$\textbf{(5.61a)}\quad \mathbf{Y}=\begin{bmatrix}399\\121\\\vdots\\323\end{bmatrix}\qquad \textbf{(5.61b)}\quad \mathbf{X}=\begin{bmatrix}1 & 80\\1 & 30\\\vdots & \vdots\\1 & 70\end{bmatrix}\tag{5.61}$$

同時可以計算矩陣乘積：

$$\mathbf{X}'\mathbf{X}=\begin{bmatrix}1 & 1 & \cdots & 1\\80 & 30 & \cdots & 70\end{bmatrix}\begin{bmatrix}1 & 80\\1 & 30\\\vdots & \vdots\\1 & 70\end{bmatrix}=\begin{bmatrix}25 & 1{,}750\\1{,}750 & 142{,}300\end{bmatrix}\tag{5.62}$$

$$\mathbf{X}'\mathbf{Y}=\begin{bmatrix}1 & 1 & \cdots & 1\\80 & 30 & \cdots & 70\end{bmatrix}\begin{bmatrix}399\\121\\\vdots\\323\end{bmatrix}=\begin{bmatrix}7{,}807\\617{,}180\end{bmatrix}\tag{5.63}$$

並且應用(5.22)得出 $\mathbf{X}'\mathbf{X}$ 的反矩陣：

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} .287475 & -.003535 \\ -.003535 & .00005051 \end{bmatrix} \tag{5.64}$$

最後代入(5.60)，可以得到：

$$\begin{aligned} \mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} &= \begin{bmatrix} .287475 & -.003535 \\ -.003535 & .00005051 \end{bmatrix} \begin{bmatrix} 7,807 \\ 617,180 \end{bmatrix} \\ &= \begin{bmatrix} 62.37 \\ 3.5702 \end{bmatrix} \end{aligned} \tag{5.65}$$

於是得出結果 $b_0 = 62.37$ 與 $b_1 = 3.5702$，自然與第 1 章之結果吻合。

2. 比較標準方程式以及 $\mathbf{X}'\mathbf{X}$，可以發現當 $\mathbf{X}'\mathbf{X}$ 之各行線性相依，則標準方程式之各行也會線性相依，此時 b_0 與 b_1 不存在唯一解，不過對於大部分的迴歸應用而言，$\mathbf{X}'\mathbf{X}$ 之各行總是線性獨立的，因此可以得到 b_0 與 b_1 的唯一解。■

5.11 配適值與殘差 Fitted Values and Residuals

- **配適值**

  在矩陣向量的形式中，我們可以用向量 $\hat{\mathbf{Y}}$來表示各個配適值 $\hat{Y}_i$：

$$\underset{n\times 1}{\hat{\mathbf{Y}}} = \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{bmatrix} \tag{5.70}$$

  然後用矩陣代數：

$$\underset{n\times 1}{\hat{\mathbf{Y}}} = \underset{n\times 2}{\mathbf{X}}\ \underset{2\times 1}{\mathbf{b}} \tag{5.71}$$

  因為：

$$\begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{bmatrix} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} b_0 + b_1 X_1 \\ b_0 + b_1 X_2 \\ \vdots \\ b_0 + b_1 X_n \end{bmatrix}$$

例題 在Toluca公司的案例中，利用矩陣(5.61b)與(5.65)可以得出配適值向量：

$$\hat{\mathbf{Y}} = \mathbf{X}\mathbf{b} = \begin{bmatrix} 1 & 80 \\ 1 & 30 \\ \vdots & \vdots \\ 1 & 70 \end{bmatrix} \begin{bmatrix} 62.37 \\ 3.5702 \end{bmatrix} = \begin{bmatrix} 347.98 \\ 169.47 \\ \vdots \\ 312.28 \end{bmatrix} \tag{5.72}$$

其結果自然與表 1.2 相同。

帽子矩陣

- 應用(5.60)中的**b**公式可以將(5.71)的 $\hat{\mathbf{Y}}$ 成 $\underset{2\times1}{\mathbf{b}} = \underset{2\times2}{(\mathbf{X}'\mathbf{X})^{-1}}\underset{2\times1}{\mathbf{X}'\mathbf{Y}}$

$$\underset{n\times1}{\hat{\mathbf{Y}}} = \underset{n\times2}{\mathbf{X}}\,\underset{2\times1}{\mathbf{b}} \qquad \hat{\mathbf{Y}}=\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

亦即

$$\boxed{\underset{n\times1}{\hat{\mathbf{Y}}} = \underset{n\times n}{\mathbf{H}}\,\underset{n\times1}{\mathbf{Y}}} \tag{5.73}$$

其中，

$$\boxed{\underset{n\times n}{\mathbf{H}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'} \tag{5.73a}$$

透過(5.73)可知適配值 $\hat{\mathbf{Y}}$可以表示為反應變數觀測值Y_i的線性組合，而係數則為矩陣**H**中的元素，同時矩陣**H**僅僅牽涉到預測變數**X**。$n\times n$之方陣**H**稱為帽子矩陣(hat matrix)，該矩陣使用於迴歸分析時是否會受到少數觀測值之影響

- 矩陣**H**為一個對稱矩陣，並且具備一個**冪等性**的特殊性質：

$$\mathbf{HH}=\mathbf{H} \tag{5.74}$$

在矩陣代數中，當矩陣 **M** 滿足**MM** = **M** 時，則稱該矩陣為**冪等**的。

- 殘差

將殘差 $e_i = Y_i - \hat{Y}_i$ 所組成之向量用符號**e**表示：

$$\underset{n\times 1}{\mathbf{e}} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \tag{5.75}$$

透過矩陣代數，我們有：

$$\underset{n\times 1}{\mathbf{e}} = \underset{n\times 1}{\mathbf{Y}} - \underset{n\times 1}{\hat{\mathbf{Y}}} = \underset{n\times 1}{\mathbf{Y}} - \underset{n\times 1}{\mathbf{Xb}} \tag{5.76}$$

在 Toluca 公司的案例中，應用(5.61a)與(5.72)可以得出殘差向量：

$$\mathbf{e} = \begin{bmatrix} 399 \\ 121 \\ \vdots \\ 323 \end{bmatrix} - \begin{bmatrix} 347.98 \\ 169.47 \\ \vdots \\ 312.28 \end{bmatrix} = \begin{bmatrix} 51.02 \\ -48.47 \\ \vdots \\ 10.72 \end{bmatrix} \tag{5.77}$$

其結果自然與表 1.2 相同。

殘差之共變異矩陣　一如配適值$\hat{Y}_i$，殘差e_i可以表示為反應變數觀測值Y_i的線性組合，應用(5.73)中$\hat{\mathbf{Y}}$的結果：

$$\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{HY} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$$

我們可以得到：

$$\boxed{\underset{n\times 1}{\mathbf{e}} = \underset{n\times n}{(\mathbf{I} - \mathbf{H})}\underset{n\times 1}{\mathbf{Y}}} \tag{5.78}$$

其中 **H** 為(5.73a)所定義之帽子矩陣，而矩陣 **I** – **H** 與 **H** 都是對稱且為冪等的。

殘差向量 **e** 的共變異矩陣，與矩陣 **I** – **H** 有關：

$$\underset{n\times n}{\boldsymbol{\sigma}^2\{\mathbf{e}\}} = \sigma^2(\mathbf{I} - \mathbf{H}) \tag{5.79}$$

其估計式為：

$$\boxed{\underset{n\times n}{\mathbf{s}^2\{\mathbf{e}\}} = MSE(\mathbf{I} - \mathbf{H})} \tag{5.80}$$

5.12 變異數分析 Analysis of Variance Results

- **平方和**

  首先從(2.43)有關於*SSTO*（總平方和）之定義開始，進行有關於如何利用矩陣來表達變異數分析：

$$SSTO = \sum\left(Y_i - \bar{Y}\right)^2 = \sum Y_i^2 - \frac{\left(\sum Y_i\right)^2}{n} \tag{5.81}$$

而(5.13)為：
$$\boxed{\mathbf{Y}'\mathbf{Y} = \sum Y_i^2}$$

透過(5.18)所定義之矩陣 **J**，可以將 $(\sum Y_i)^2/n$ 表示為：

$$\frac{\left(\sum Y_i\right)^2}{n} = \left(\frac{1}{n}\right)\mathbf{Y}'\mathbf{J}\mathbf{Y} \tag{5.82}$$

舉例來說，若n＝2，則可得：
$(Y_1+Y_2)Y_1+(Y_1+Y_2)Y_2$

$$\left(\frac{1}{2}\right)\begin{pmatrix}Y_1 & Y_2\end{pmatrix}\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}\begin{pmatrix}Y_1\\ Y_2\end{pmatrix} = \frac{(Y_1+Y_2)\cdot(Y_1+Y_2)}{2} = \frac{1}{2}(Y_1+Y_2)^2$$

因此$SSTO$可以表示為：

$$\boxed{SSTO = \mathbf{Y}'\mathbf{Y} - \left(\frac{1}{n}\right)\mathbf{Y}'\mathbf{J}\mathbf{Y}} \tag{5.83}$$

同理，$SSE = \sum e_i^2 = \sum \left(Y_i - \hat{Y}_i\right)^2$，也可以透過矩陣來表達：

$$SSE = \mathbf{e}'\mathbf{e} = \left(\mathbf{Y} - \mathbf{X}\mathbf{b}\right)'\left(\mathbf{Y} - \mathbf{X}\mathbf{b}\right) \tag{5.84}$$

而且可以證明出下面的結果：

$$\boxed{SSE = \mathbf{Y}'\mathbf{Y} - \mathbf{b}'\mathbf{X}'\mathbf{Y}} \tag{5.84a}$$

以及SSR：

$$\boxed{SSR = \mathbf{b}'\mathbf{X}'\mathbf{Y} - \left(\frac{1}{n}\right)\mathbf{Y}'\mathbf{J}\mathbf{Y}} \tag{5.85}$$

例題 在 Toluca 公司的案例中，應用矩陣(5.61a)與(5.84a)可以得到：

$$\mathbf{Y}'\mathbf{Y} = \begin{bmatrix} 399 & 121 & \cdots & 323 \end{bmatrix} \begin{bmatrix} 399 \\ 121 \\ \vdots \\ 323 \end{bmatrix} = 2{,}745{,}173$$

另外，根據(5.65)與(5.63)，我們可以發現：

$$\mathbf{b}'\mathbf{X}'\mathbf{Y} = \begin{bmatrix} 62.37 & 3.5702 \end{bmatrix} \begin{bmatrix} 7{,}807 \\ 617{,}180 \end{bmatrix} = 2{,}690{,}348$$

所以可以計算出 SSE：

$$SSE = \mathbf{Y}'\mathbf{Y} - \mathbf{b}'\mathbf{X}'\mathbf{Y} = 2{,}745{,}173 - 2{,}690{,}348 = 54{,}825$$

其結果自然與第 1 章相同。

- 平方和的二次式表示

對於ANOVA中的平方和，事實上均可以證明它們為**二次式**，例如$n=2$時，觀測值 Y_i 的二次式：

$$5Y_1^2+6Y_1Y_2+4Y_2^2 \tag{5.86}$$

就是一個二階的多項式，該多項式之各項均為觀測值之平方或是兩兩之乘積，用矩陣表示(5.86)如下：

$$\begin{bmatrix} Y_1 & Y_2 \end{bmatrix}\begin{bmatrix} 5 & 3 \\ 3 & 4 \end{bmatrix}\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}=\mathbf{Y'AY} \tag{5.86a}$$

其中，**A**為一個對稱矩陣。

- 一般對於二次式之定義為：

$$\underset{n\times 1}{\mathbf{Y}'\mathbf{A}\mathbf{Y}} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}Y_iY_j \qquad \text{其中} \quad a_{ij} = a_{ji} \tag{5.87}$$

而**A**為一個$n\times n$之對稱矩陣，特別稱它為二次式矩陣。

ANOVA中的平方和：*SSTO*、*SSE*、*SSR*都是二次式。從(5.71)式應用(5.32)可以得到：

$$\mathbf{b}'\mathbf{X}' = \left(\mathbf{X}\mathbf{b}\right)' = \hat{\mathbf{Y}}'$$

利用(5.73)之結果可得出：

$$\mathbf{b}'\mathbf{X}' = \left(\mathbf{H}\mathbf{Y}\right)'$$

因**H**為對稱矩陣，並利用(5.32)可得：

$$\mathbf{b}'\mathbf{X}' = \mathbf{Y}'\mathbf{H} \tag{5.88}$$

此一結果可以讓我們以下列方式表示ANOVA中之 平方和：

$$SSTO = \mathbf{Y}'\left[\mathbf{I} - \left(\frac{1}{n}\right)\mathbf{J}\right]\mathbf{Y} \tag{5.89a}$$

$$SSE = \mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y} \tag{5.89b}$$

$$SSR = \mathbf{Y}'\left[\mathbf{H} - \left(\frac{1}{n}\right)\mathbf{J}\right]\mathbf{Y} \tag{5.89c}$$

- 每個平方和均可以表示為$\mathbf{Y}'\mathbf{A}\mathbf{Y}$，相對應的$\mathbf{A}$矩陣則為：

$$\mathbf{I}-\left(\frac{1}{n}\right)\mathbf{J} \tag{5.90a}$$

$$\mathbf{I}-\mathbf{H} \tag{5.90b}$$

$$\mathbf{H}-\left(\frac{1}{n}\right)\mathbf{J} \tag{5.90c}$$

5.13　迴歸分析推論

Inference in Regression Analysis

- 區間估計的共同形式，均為利用點估計量加減所估計標準差之某一個倍數，在此進行點估計之變異數估計，其矩陣形式之討論。
- 迴歸係數

迴歸係數向量**b**，有下面的共變異矩陣：

$$\underset{2\times 2}{\boldsymbol{\sigma}^2\{\mathbf{b}\}} = \begin{bmatrix} \sigma^2\{b_0\} & \sigma\{b_0, b_1\} \\ \sigma\{b_1, b_0\} & \sigma^2\{b_1\} \end{bmatrix} \tag{5.91}$$

可以表示為：

$$\underset{2\times 2}{\boldsymbol{\sigma}^2\{\mathbf{b}\}} = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \tag{5.92}$$

或是應用(5.24a)，表示為：

$$\underset{2\times2}{\boldsymbol{\sigma}^2\{\mathbf{b}\}}=\begin{bmatrix}\dfrac{\sigma^2}{n}+\dfrac{\sigma^2\bar{X}^2}{\sum(X_i-\bar{X})^2} & \dfrac{-\bar{X}\sigma^2}{\sum(X_i-\bar{X})^2}\\ \dfrac{-\bar{X}\sigma^2}{\sum(X_i-\bar{X})^2} & \dfrac{\sigma^2}{\sum(X_i-\bar{X})^2}\end{bmatrix} \qquad (5.92a)$$

用 MSE 取代(5.92a)中的，可以得到 **b** 之估計的共變異矩陣 $\mathbf{s}^2\{\mathbf{b}\}$ ：

$$\underset{2\times2}{\mathbf{s}^2\{\mathbf{b}\}}=MSE(\mathbf{X}'\mathbf{X})^{-1}=\begin{bmatrix}\dfrac{MSE}{n}+\dfrac{\bar{X}^2 MSE}{\sum(X_i-\bar{X})^2} & \dfrac{-\bar{X}MSE}{\sum(X_i-\bar{X})^2}\\ \dfrac{-\bar{X}MSE}{\sum(X_i-\bar{X})^2} & \dfrac{MSE}{\sum(X_i-\bar{X})^2}\end{bmatrix} \qquad (5.93)$$

在(5.92a)中可以看出(2.3)中b_0的變異數與(2.9)中b_1的變異數，以及(4.5) 中b_0與b_1共變異數。

例題

在 Toluca 公司的案例中，應用圖 2.2 與(5.64)可以得到 $s^2\{b_0\}$與 $s^2\{b_1\}$：

$$\underset{2\times 2}{\mathbf{s}^2\{\mathbf{b}\}} = MSE\left(\mathbf{X}'\mathbf{X}\right)^{-1} = 2{,}384\begin{bmatrix} .287475 & -.003535 \\ -.003535 & .00005051 \end{bmatrix}$$

$$= \begin{bmatrix} 685.34 & -8.428 \\ -8.428 & .12040 \end{bmatrix} \tag{5.94}$$

所以 $s^2\{b_0\} = 685.34$，而 $s^2\{b_1\} = .12040$，結果與第 2 章相同。

平均反應

在此將針對X_h進行平均反應之估計，首先定義向量：

$$\underset{2\times1}{\mathbf{X}_h} = \begin{bmatrix} 1 \\ X_h \end{bmatrix} \quad 或 \quad \underset{1\times2}{\mathbf{X}'_h} = [1 \quad X_h] \tag{5.95}$$

用矩陣符號表示其配適值為：$\boxed{\hat{Y}_h = \mathbf{X}'_h\mathbf{b}}$

因為，

$$\mathbf{X}'_h\mathbf{b} = [1 \quad X_h]\begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = [b_0 + b_1X_h] = \left[\hat{Y}_h\right] = \hat{Y}_h \tag{5.96}$$

在式子(2.29b)中給定$\hat{Y}_h$之變異數，利用矩陣形式表 示如下：

$$\boxed{\sigma^2\{\hat{Y}_h\} = \sigma^2\mathbf{X}'_h(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_h} \tag{5.97}$$

應用式子(5.92)，(5.93)中 $\hat{Y}_h$之變異數，可以表示為估計迴係數時，其共變異矩陣 $\boldsymbol{\sigma}^2\{\mathbf{b}\}$ 之函數：

$$\boxed{\sigma^2\{\hat{Y}_h\} = \mathbf{X}'_h\boldsymbol{\sigma}^2\{\mathbf{b}\}\mathbf{X}_h} \tag{5.97a}$$

在(2.30)中所列出$\hat{Y}_h$之估計變異數，在此利用矩陣形式表示為：

$$\boxed{s^2\left\{\hat{Y}_h\right\} = MSE\left(\mathbf{X}_h'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}_h\right)} \tag{5.98}$$

例題

在 Toluca 公司的案例中，當 $X_h = 65$ 時，我們可以計算 $s^2\{\hat{Y}_h\}$，首先定義

$$\mathbf{X}_h' = \begin{bmatrix} 1 & 65 \end{bmatrix}$$

並應用(5.94)之結果，我們有：

$$\begin{aligned} s^2\left\{\hat{Y}_h\right\} &= \mathbf{X}_h'\mathbf{s}^2\{\mathbf{b}\}\mathbf{X}_h \\ &= \begin{bmatrix} 1 & 65 \end{bmatrix}\begin{bmatrix} 685.34 & -8.428 \\ -8.428 & .12040 \end{bmatrix}\begin{bmatrix} 1 \\ 65 \end{bmatrix} = 98.37 \end{aligned}$$

結果與第 2 章相同。

- **新觀測值之預測**

  在式子(2.38)中所給的估計變異數 $s^2\{\text{pred}\}$ 用矩陣形式表示為：

$$s^2\{\text{pred}\} = MSE\left(1 + \mathbf{X}_h'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_h\right) \qquad (5.100)$$