บทที่ 5
การวิเคราะห์การถดถอยเชิงเส้นด้วยวิธีเมทริกซ์

การพยากรณ์ค่าของตัวแปรตาม y จากค่าของตัวแปรอิสระ x ที่มีความสัมพันธ์กันในรูปแบบเชิงเส้นตรงอย่างง่ายนั้น สมการถดถอยของกลุ่มตัวอย่างที่ใช้การพยากรณ์ค่า y คือ

$$\hat{y} \quad = \quad b_0 + b_1 x$$

เมื่อ $\hat{y}$ แทนค่าตัวแปรตามหรือค่าพยากรณ์จากสมการถดถอย

x แทนค่าของตัวแปรอิสระซึ่งทราบค่า

b_0 แทนค่าของ y เมื่อ $x = 0$ หรือจุดตัดของเส้นถดถอยบนแกน y

b_1 แทนค่าของความชันของสมการถดถอยหรือค่าที่แสดงให้รู้ว่า เมื่อ x เปลี่ยนไป 1 หน่วย ค่า y จะเปลี่ยนไปกี่หน่วย ค่า b_1 นี้ เมื่อนำไปคูณกับค่า x_1 จะช่วยให้ค่าที่พยากรณ์ได้ใกล้ค่าจริงมากที่สุดเนื่องจากเป็นค่าน้ำหนักความสำคัญของตัวแปร x ซึ่งจะทำให้เกิดความคลาดเคลื่อนในการพยากรณ์น้อยลง

สมการถดถอยเชิงเส้นอย่างง่าย ที่ใช้ในการคาดคะเนค่า y นั้น จะให้ $\sum(y - \hat{y})^2$ มีค่าน้อยที่สุด และ ค่า $y - \hat{y}$ หรือ ผลต่างระหว่างค่าจริง y กับค่าคาดคะเน $\hat{y}$ เรียกว่าค่าความคลาดเคลื่อน ในการคาดคะเน ใช้แทนด้วยสัญลักษณ์ e ซึ่ง $e = y - \hat{y}$ ตามหลักของวิธีกำลังสองน้อยที่สุด สมการถดถอยที่ใช้พยากรณ์ค่า y ต่างๆ จะต้องมีความคลาดเคลื่อนในการคาดคะเนน้อยที่สุด นั่นคือ $\sum e^2 = \sum(y - \hat{y})^2$ มีค่าน้อยที่สุดนั่นเอง

จากที่ได้กล่าวถึงการหาสมการถดถอยเชิงเส้นตรงด้วยวิธีทางพีชคณิตในบทที่ผ่านมา บทนี้จะกล่าวถึงการหาสมการถดถอยเชิงเส้นตรงด้วยวิธีเมทริกซ์ ซึ่งสามารถประยุกต์ใช้ในการวิเคราะห์ข้อมูลและการวิเคราะห์การถดถอยที่มีตัวแปรอิสระหลายตัวโดยใช้เครื่องคอมพิวเตอร์

5.1 ตัวแบบสมการถดถอยเชิงเส้นตรงอย่างง่ายในรูปของเมทริกซ์

การใช้ทฤษฎีของเมทริกซ์ในการวิเคราะห์ข้อมูล มีประโยชน์มากมายในกิจกรรมหลายอย่าง เช่น การวิเคราะห์ระบบขนส่งสื่อสาร การวิเคราะห์ข้อมูลในด้านสังคมวิทยาและจิตวิทยา การแก้ปัญหาต่างๆ ด้านการวางแผน ระบบควบคุมในโรงงานอุตสาหกรรม และกิจการอุตสาหกรรมอย่างกว้างขวาง

เมทริกซ์ ประกอบด้วยจำนวน หรือ ฟังก์ชันที่จัดเรียงอยู่ในรูปแถว หรือ แถวนอน (row) และสดมภ์ หรือ แถวตั้ง (column) อาจจะเขียนได้ดังนี้

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots\cdots & a_{1n} \\ a_{21} & a_{22} & \cdots\cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots\cdots & a_{mn} \end{bmatrix}$$

เรียกจำนวนหรือฟังก์ชัน a_{ij} ว่าองค์ประกอบ หรือ สมาชิก (element) ของ เมทริกซ์ A

เราอาจจะแทนเมทริกซ์ A ข้างบนโดยใช้ $\| \quad \|$ หรือ ( ) ก็ได้ และบางทีจะแทน A ด้วย $[a_{ij}]$ หรือ $\| a_{ij} \|$ หรือ (a_{ij})

เราเรียกเมทริกซ์ที่มี m แถว และ n สดมภ์ว่า เมทริกซ์ที่มีขนาด (m,n) หรือ m x n

$$\begin{array}{cl} \begin{matrix} \text{สดมภ์ที่ 1} & \text{สดมภ์ที่ 2} \\ \Downarrow & \Downarrow \end{matrix} & \\ \begin{bmatrix} 4 & 5 \\ 6 & -1 \\ 0 & 8 \end{bmatrix} & \begin{matrix} \Leftarrow \text{แถวที่ 1} \\ \Leftarrow \text{แถวที่ 2} \\ \Leftarrow \text{แถวที่ 3} \end{matrix} \end{array}$$

ถ้า $m = n$ เรียกเมทริกซ์นี้ว่า เมทริกซ์จตุรัส (square matrix) ที่มีขนาด m x n เมื่อ $m \neq n$ เราเรียกเมทริกซ์นี้ว่า เมทริกซ์สี่เหลี่ยมพื้นผ้า (rectangular matrix) ที่มีขนาด m x n

ในการนำเมทริกซ์มาประยุกต์ใช้ในการวิเคราะห์การถดถอยนั้น

ถ้า Y แทนตัวแปรตาม

X แทนตัวแปรอิสระ หรือตัวพยากรณ์

ตัวแบบสมการถดถอยเชิงเส้นตรงอย่างง่ายของประชากรคือ

$$Y_i = \beta_0 + \beta_i X_i + \varepsilon_i$$

เขียนในรูปเมทริกซ์ได้ดังนี้

$$[Y] = [X][\beta] + [\varepsilon]$$

ในที่นี้ประชากรมีสมาชิกทั้งหมด N หน่วย คือ

ประชากร	ตัวแปร (X_i)
1	X_1
2	X_2
3	X_3
....	....
....	....
N	X_N

ตัวแบบสมการถดถอยเชิงเส้นตรงอย่างง่ายของประชากร เขียนในรูปเมทริกซ์จะได้ ดังนี้

$$\begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_N \end{bmatrix} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ . & . \\ . & . \\ . & . \\ 1 & X_N \end{bmatrix} \begin{bmatrix} \beta_0 \\ \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ . \\ \varepsilon_N \end{bmatrix}$$

โดย Y , X , β และ ε แทน เมทริกซ์ของค่าต่างๆ ดังนี้

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_N \end{bmatrix}$$

$$X = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ . & . \\ . & . \\ . & . \\ 1 & X_N \end{bmatrix}$$

$$\beta = \begin{bmatrix} \beta_0 \\ \\ \beta_1 \end{bmatrix}$$

$$\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ . \\ \varepsilon_N \end{bmatrix}$$

ตัวแบบสมการถดถอยเชิงเส้นตรงอย่างง่ายของกลุ่มตัวอย่าง คือ

$$y_i = b_0 + b_i x_i + e_i$$

เขียนในรูปเมทริกซ์จากสมการ $[y] = [x][b] + [e]$ ตัวแบบสมการถดถอยเชิงเส้นตรงอย่างง่ายของกลุ่มตัวอย่างเขียนในรูปเมทริกซ์จะได้ ดังนี้

$$\begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ . & . \\ . & . \\ . & . \\ 1 & x_n \end{bmatrix} \begin{bmatrix} b_0 \\ \\ b_1 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ . \\ . \\ . \\ e_n \end{bmatrix}$$

เมื่อสุ่มตัวอย่างจากประชากรมา n หน่วย เมทริกซ์ Y, X, b และ e เขียนได้ดังนี้

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{bmatrix}$$

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ .. & \\ . & . \\ . & . \\ 1 & x_n \end{bmatrix}$$

$$b = \begin{bmatrix} b_0 \\ \\ b_1 \end{bmatrix}$$

$$e = \begin{bmatrix} e_1 \\ e_2 \\ . \\ . \\ . \\ e_n \end{bmatrix}$$

5.2 การหาสมการถดถอยเชิงเส้นอย่างง่ายด้วยวิธีเมทริกซ์

เมื่อสุ่มตัวอย่างจากประชากรมาจำนวน n หน่วย เพื่อใช้ในการหาสมการถดถอยเชิงเส้นตรงของกลุ่มตัวอย่าง ในกรณีที่มีตัวแปรตาม y 1 ตัว และตัวแปรอิสระ x 1 ตัวหรือที่เรียกว่า การถดถอยเชิงเส้นตรงอย่างง่ายนั้น จะต้องหาค่า b_0 และ b_1 ของสมการ $\hat{y} = b_0 + b_1x$ ที่ทำให้การพยากรณ์ค่า y มีความคลาดเคลื่อนน้อยที่สุด หรือสามารถพยากรณ์ค่า y ได้ใกล้ความเป็นจริงมากที่สุดนั่นเอง ซึ่งการหาค่า b_0 และ b_1 ดังกล่าวด้วยวิธีเมทริกซ์ทำได้ดังนี้

ให้ Y แทน คอลัมน์เวคเตอร์ของตัวแปรตาม ที่ประกอบด้วนสมาชิก $y_1, y_2, ..., y_n$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{bmatrix}$$

X แทน เมทริกซ์ของตัวแปรอิสระที่มีเลข 1 จำนวน n ตัว อยู่ข้างหน้าตัวแปรอิสระ $x_1, x_2, ..., x_n$ แต่ละตัว

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ .. & \\ . & . \\ . & . \\ 1 & x_n \end{bmatrix}$$

b แทน คอลัมน์เวคเตอร์ของสัมประสิทธิ์การถดถอย ที่ประกอบด้วยสมาชิก b_0 และ b_1

$$b = \begin{bmatrix} b_0 \\ \\ b_1 \end{bmatrix}$$

การหาค่าสัมประสิทธิ์ถดถอย b_0 และ b_1 หรือในที่นี้ก็คือ คอลัมน์เวคเตอร์ b มีวิธีการหาโดยอาศัยจากสมการปกติ

$$nb_0 + (\sum x) b_1 = \sum y$$

$$(\sum x) b_0 + (\sum x^2) b_1 = \sum xy$$

มีวิธีวิเคราะห์ดังนี้

1. หาค่า $X^T X$ ซึ่งจะได้

$$X^T X = \begin{bmatrix} 1 & 1 & . & . & . & 1 \\ & & & & & \\ x_1 & x_2 & . & . & . & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ & . \\ . & . \\ . & . \\ 1 & x_n \end{bmatrix}$$

$$= \begin{bmatrix} n & \sum x \\ \sum x & \sum x^2 \end{bmatrix}$$

2. หาค่า $[X^T X]\,[b]$ จะได้

$$[X^T X]\,[b] = \begin{bmatrix} n & \sum x \\ \sum x & \sum x^2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$$

$$= \begin{bmatrix} nb_0 + (\sum x) b_1 \\ (\sum x) b_0 + (\sum x^2) b_1 \end{bmatrix}$$

3. หาค่า $X^T Y$ จะได้

$$X^T Y = \begin{bmatrix} 1 & 1 & . & . & . & 1 \\ \\ \\ x_1 & x_2 & . & . & . & x_N \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{bmatrix}$$

$$= \begin{bmatrix} \sum y \\ \\ \sum xy \end{bmatrix}$$

ซึ่งจะสอดคล้องกับสมการ ปกติ

$$nb_0 + (\sum x)\, b_1 = \sum y$$

$$(\sum x)\, b_0 + (\sum x^2) b_1 = \sum xy$$

พิจารณาร่วมกับ $[X^T X]\,[b]$ และ $[X^T Y]$ แล้วจะได้ $[X^T X]\,[b] = [X^T Y]$
ใช้ค่าอินเวอร์สของ $X^T X$ คือ $[X^T X]^{-1}$ คูณด้านหน้าทั้งสองข้างของสมการ จะได้

$$[X^T X]^{-1}\ [X^T X]\,[b] = [X^T X]^{-1}\ [X^T Y]$$

แต่เนื่องจากอินเวอร์สของเมทริกซ์ใด คูณกับเมทริกซ์นั้น จะได้ไอเดนติตี้เมทริกซ์ (identity matrics) นั่นคือ $[X^T X]^{-1}\ [X^T X] = [I]$

$$\therefore\ [I]\,[b] = [X^T X]^{-1}\ [X^T Y]$$

และ ไอเดนติตี้เมทริกซ์คูณกับเมทริกซ์ใด จะได้เมทริกซ์นั้น

$$\therefore\ [I]\,[b] = [b]$$

ดังนั้น จะได้ $[b] = [X^T X]^{-1}\ [X^T Y]$

แต่ $[b] = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$

$\therefore \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = [X^T X]^{-1} [X^T Y]$ นั่นเอง

จาก $[X^T X] = \begin{bmatrix} n & \sum x \\ \sum x & \sum x^2 \end{bmatrix}$

$$[X^T X]^{-1} = \begin{bmatrix} n & \sum x \\ \sum x & \sum x^2 \end{bmatrix}^{-1}$$

$$= \frac{1}{|X^T X|} \text{ . adj } [X^T X]$$

เนื่องจาก $|X^T X| = \begin{vmatrix} n & \sum x \\ \sum x & \sum x^2 \end{vmatrix}$

$$= n \sum x^2 - (\sum x)^2$$

$$= n \sum (x - \bar{x})^2$$

และ adj $[X^T X]$ ก็คือ ทรานสโพสของเมทริกซ์โคแฟคเตอร์ เมื่อ

เมทริกซ์โคแฟคเตอร์ คือ $\begin{bmatrix} \sum x^2 & -\sum x \\ -\sum x & n \end{bmatrix}$

$$\therefore \quad \text{adj } [X^T X] = \begin{bmatrix} \sum x^2 & -\sum x \\ -\sum x & n \end{bmatrix}^T = \begin{bmatrix} \sum x^2 & -\sum x \\ -\sum x & n \end{bmatrix}$$

ดังนั้นจะได้

$$[X^T X]^{-1} = \frac{1}{n\sum (x - \bar{x})^2} \begin{bmatrix} \sum x^2 & -\sum x \\ -\sum x & n \end{bmatrix}$$

จากสมการ (4.11) ข้างต้น

$$[X^T Y] = \begin{bmatrix} \sum y \\ \sum xy \end{bmatrix}$$

แทนค่า $[X^T X]^{-1}$ และ $[X^T Y]$ ในสมการ

$$\therefore \quad \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = [X^T X]^{-1} [X^T Y]$$ เราจะได้

$$\begin{bmatrix} b_0 \\ \\ b_1 \end{bmatrix} = \begin{bmatrix} \dfrac{\sum x^2}{n\sum\left(x-\bar{x}\right)^2} & \dfrac{-\sum x}{n\sum\left(x-\bar{x}\right)^2} \\ \\ \dfrac{-\sum x}{n\sum\left(x-\bar{x}\right)^2} & \dfrac{n}{n\sum\left(x-\bar{x}\right)^2} \end{bmatrix} \begin{bmatrix} \sum y \\ \\ \sum xy \end{bmatrix}$$

$$= \begin{bmatrix} \dfrac{\sum x^2 \sum y}{n\sum\left(x-\bar{x}\right)^2} - \dfrac{\sum x \sum xy}{n\sum\left(x-\bar{x}\right)^2} \\ \\ \dfrac{-\sum x \sum y}{n\sum\left(x-\bar{x}\right)^2} + \dfrac{n\sum xy}{n\sum\left(x-\bar{x}\right)^2} \end{bmatrix}$$

$$= \begin{bmatrix} \dfrac{\sum x^2 \bar{y}}{\sum\left(x-\bar{x}\right)^2} - \dfrac{\bar{x}\sum xy}{\sum\left(x-\bar{x}\right)^2} \\ \\ \dfrac{\sum xy}{\sum\left(x-\bar{x}\right)^2} - \dfrac{n\overline{xy}}{\sum\left(x-\bar{x}\right)^2} \end{bmatrix}$$

$$= \begin{bmatrix} \dfrac{\bar{y}\sum x^2 - \bar{y}n\bar{x}^2 + \bar{y}n\bar{x}^2 - \bar{x}\sum xy}{\sum\left(x-\bar{x}\right)^2} \\ \\ \dfrac{\sum xy - n\overline{xy}}{\sum\left(x-\bar{x}\right)^2} \end{bmatrix}$$

$$\begin{bmatrix} b_0 \\ \\ b_1 \end{bmatrix} = \begin{bmatrix} \dfrac{\bar{y}\left(\sum x^2 - n\bar{x}^2\right) - \bar{x}\left(\sum xy - n\bar{x}\bar{y}\right)}{\sum\left(x - \bar{x}\right)^2} \\ \\ \dfrac{\sum xy - n\bar{x}\bar{y}}{\sum\left(x - \bar{x}\right)^2} \end{bmatrix}$$

$$= \begin{bmatrix} \bar{y} \cdot \dfrac{\sum\left(x - \bar{x}\right)^2}{\sum\left(x - \bar{x}\right)^2} - \dfrac{\sum xy - n\bar{x}\bar{y}}{\sum\left(x - \bar{x}\right)^2} \cdot \bar{x} \\ \\ \dfrac{\sum xy - n\bar{x}\bar{y}}{\sum\left(x - \bar{x}\right)^2} \end{bmatrix}$$

$$[b] = \begin{bmatrix} \bar{y} - b_1\bar{x} \\ \\ \dfrac{\sum xy - n\bar{x}\bar{y}}{\sum\left(x - \bar{x}\right)^2} \end{bmatrix}$$

นั่นคือ

$$b_0 = \bar{y} - b_1\bar{x}$$

$$b_1 = \frac{\sum xy - n\bar{x}\bar{y}}{\sum\left(x - \bar{x}\right)^2} = \frac{\sum xy}{\sum x^2}$$

ดังนั้น จึงสรุปได้ว่า การหาค่า b_0 และ b_1 ด้วยวิธีเมทริกซ์ หาได้จาก

$$\begin{bmatrix} b \end{bmatrix} = \begin{bmatrix} b_0 \\ \\ b_1 \end{bmatrix} = [X^T X]^{-1} \; [X^T Y]$$

เมื่อหาค่า b_0 และ b_1 ได้แล้ว ก็สามารถสร้างสมการถดถอยเชิงเส้นตรงอย่างง่ายของกลุ่มตัวอย่าง $\hat{y}_1 = b_0 + b_1X_1$ ที่จะใช้คาดคะเนค่า y เมื่อรู้ค่า x ซึ่งจะให้ค่าคาดคะเน $\hat{y}$ มีความคาดเคลื่อนน้อยที่สุด

สมการถดถอยเชิงเส้นตรงอย่างง่ายของกลุ่มตัวอย่างดังกล่าว เมื่อเขียนในรูปเมทริกซ์จะได้ดังนี้

$$[\hat{Y}] = [X][b]$$

หรือ
$$\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ . \\ . \\ . \\ \hat{y}_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ . & . \\ . & . \\ . & . \\ 1 & x_n \end{bmatrix} \begin{bmatrix} b_0 \\ \\ b_1 \end{bmatrix}$$

$$= \begin{bmatrix} b_0 + b_1x_1 \\ b_0 + b_1x_2 \\ . \\ . \\ . \\ b_0 + b_1x_n \end{bmatrix}$$

นั่นคือ
$$\hat{y}_1 = b_0 + b_1x_1$$
$$\hat{y}_2 = b_0 + b_1x_2$$
$$.$$
$$.$$
$$.$$
$$\hat{y}_n = b_0 + b_1x_n$$

ตัวอย่างที่ 5.1 สมมุติให้ x เป็นตัวแปรอิสระ และ y เป็นตัวแปรตาม ดังตารางที่ 5.1

ตารางที่ 5.1 ตารางแสดงข้อมูลระหว่างตัวแปรอิสระ (x) และตัวแปรตาม (y)

x	y
8	12
10	16
5	10
7	8
14	20
4	6

ซึ่งเมื่อจัดอยู่ในรูปเมทริกซ์ X และ Y ได้ดังนี้

$$X = \begin{bmatrix} 1 & 8 \\ 1 & 10 \\ 1 & 5 \\ 1 & 7 \\ 1 & 14 \\ 1 & 4 \end{bmatrix} \quad \text{และ} \quad Y = \begin{bmatrix} 12 \\ 16 \\ 10 \\ 8 \\ 20 \\ 6 \end{bmatrix}$$

จงหาค่า b_0 และ b_1 ด้วยวิธีเมทริกซ์

วิธีการวิเคราะห์ ให้ b เป็นคอลัมน์เวคเตอร์ ประกอบด้วยสมาชิก b_0 และ b_1

$$b = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$$

หา $X^T X$ จะได้

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 8 & 10 & 5 & 7 & 14 & 4 \end{bmatrix} \begin{bmatrix} 1 & 8 \\ 1 & 10 \\ 1 & 5 \\ 1 & 7 \\ 1 & 14 \\ 1 & 4 \end{bmatrix}$$

$$X^TX = \begin{bmatrix} 6 & 48 \\ 48 & 450 \end{bmatrix}$$

หาอินเวอร์สของ X^TX หรือ $[X^TX]^{-1}$

$$|X^TX| = 6(450) - 48(48) = 396$$

เมื่อเมทริกซ์โคแฟคเตอร์ คือ $\begin{bmatrix} 450 & -48 \\ -48 & 6 \end{bmatrix}$

หาทรานสโพสของเมทริกซ์โคแฟคเตอร์ จะได้ adj $[X^TX]$

$$\therefore \ \text{adj}\ [X^TX] = \begin{bmatrix} 450 & -48 \\ -48 & 6 \end{bmatrix}$$

$$\text{ดังนั้น} \quad [X^TX]^{-1} = \frac{1}{|X^TX|} \cdot \text{adj}\,[X^TX]$$

$$= \frac{1}{396} \cdot \begin{bmatrix} 450 & -48 \\ -48 & 6 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{450}{396} & \frac{-48}{396} \\ \frac{-48}{396} & \frac{6}{396} \end{bmatrix}$$

หา X^TY

$$X^TY = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 8 & 10 & 5 & 7 & 14 & 4 \end{bmatrix} \begin{bmatrix} 12 \\ 16 \\ 10 \\ 8 \\ 20 \\ 6 \end{bmatrix}$$

$$= \begin{bmatrix} 72 \\ 666 \end{bmatrix}$$

เนื่องจาก $b = [X^TX]^{-1}\,[X^TY]$

$$= \begin{bmatrix} \frac{450}{396} & \frac{-48}{396} \\ \frac{-48}{396} & \frac{6}{396} \end{bmatrix} \begin{bmatrix} 72 \\ 666 \end{bmatrix}$$

$$[b] = \begin{bmatrix} 1.09 \\ 1.36 \end{bmatrix}$$

นั่นคือ $b_0 = 1.09$ และ $b_1 = 1.36$

ดังนั้น สมการถดถอยเชิงเส้นตรงอย่างง่ายคือ $\hat{y} = 1.09 + 1.36\,x$

5.3 การวิเคราะห์ความแปรปรวนของการถดถอยเชิงเส้นตรงอย่างง่ายด้วยวิธีเมทริกซ์

ให้ Y แทน คอลัมน์เวคเตอร์ของตัวแปรตาม

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{bmatrix}$$

X แทนเมทริกซ์ของตัวแปรอิสระ ที่มีเลข 1 อยู่หน้า x

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ . & . \\ . & . \\ . & . \\ 1 & x_n \end{bmatrix}$$

b แทน คอลัมน์เวคเตอร์ ที่ประกอบด้วยสมาชิก b_0 และ b_1

$$b = \begin{bmatrix} b_0 \\ \\ b_1 \end{bmatrix}$$

$$\therefore \quad Y^TY = \begin{bmatrix} y_1 & y_2 & . & . & . & y_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{bmatrix} = \sum y^2$$

$$X^TY = \begin{bmatrix} 1 & 1 & . & . & . & 1 \\ x_1 & x_2 & . & . & . & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{bmatrix} = \begin{bmatrix} \sum y \\ \sum xy \end{bmatrix}$$

และจาก X^TY นี้ สามารถหาค่าของ $\overline{Y} = \dfrac{\sum y}{n}$ ได้

เนื่องจาก $b^T[X^TY] = \begin{bmatrix} b_0 & b_1 \end{bmatrix} \begin{bmatrix} \sum y \\ \sum xy \end{bmatrix}$

$$\therefore \quad b^T[X^TY] = b_0\sum y + b_1\sum xy$$

ค่าของ *SST*

$$\begin{aligned} SST &= \sum(y - \bar{y})^2 \\ &= \sum(y^2 - 2\bar{y}\,y + \bar{y}^2) \\ &= \sum y^2 - 2\bar{y}\sum y + n\bar{y}^2 \\ &= \sum y^2 - 2n\bar{y}^2 + n\bar{y}^2 \\ &= \sum y^2 - n\bar{y}^2 \end{aligned}$$

จากสมการ

$$Y^TY = \begin{bmatrix} y_1 & y_2 & . & . & . & y_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{bmatrix} = \sum y^2$$

$$Y^TY = \sum y^2$$

$$\therefore \quad SST = Y^TY - n\bar{y}^2$$

ค่าของ SSR

$$SSR = \sum(\hat{y} - \bar{y})^2$$

แต่ $\hat{y} = b_0 + b_1 x$

$$\therefore \quad SSR = \sum(b_0 + b_1 x - \bar{y})^2$$

$$= \sum(b_0 + 2b_0 b_1 x + b^2_1 x^2 - 2b_0\bar{y} - 2b_1\bar{y}x + \bar{y}^2)$$

เนื่องจาก $b_0 = \bar{y} - b_1\bar{x}$

$$\therefore \quad SSR = \sum\left[(\bar{y}^2 - 2b_1\bar{x}\,\bar{y} + b^2_1\bar{x}^2) + 2(\bar{y} - b_1\bar{x})b_1 x + b^2_1 x^2 - 2(\bar{y} - b_1\bar{x})\bar{y} - 2b_1\bar{y}x + \bar{y}^2\right]$$

$$= \sum\left[\bar{y}^2 - 2b_1\bar{x}\,\bar{y} + b^2_1\bar{x}^2 + 2b_1\bar{y}x - 2b^2_1\bar{x}x + b^2_1 x^2 - 2\bar{y}^2 + 2b_1\bar{x}\,\bar{y} - 2b_1\bar{y}x + \bar{y}^2\right]$$

$$= \sum\left[b^2_1\bar{x}^2 - 2b^2_1\bar{x}x + b^2_1 x^2\right]$$

$$= nb^2_1\bar{x}^2 - 2b^2_1\bar{x}\sum x + b^2_1\sum x^2$$

$$= nb^2_1\bar{x}^2 - 2b^2_1 n\bar{x}^2 + b^2_1\sum x^2$$

$$= b^2_1\sum x^2 - nb^2_1\bar{x}^2$$

$$= b^2_1(\sum x^2 - n\bar{x}^2)$$

แต่ $b_1 = \dfrac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2}$

แทนค่า b_1 หนึ่งค่า จะได้

$$SSR = b_1 \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2} \left(\sum x^2 - n\bar{x}^2\right)$$

แต่ $(\sum x - \bar{x}) = \sum x^2 - n\bar{x}^2$

$$\begin{aligned}
\therefore \quad SSR &= b_1 \sum (x - \bar{x})(y - \bar{y}) \\
&= b_1 \sum (xy - x\bar{y} - \bar{x}y + \bar{x}\bar{y}) \\
&= b_1 \sum (xy - \bar{y}\sum x - \bar{x}\sum y + n\bar{x}\bar{y}) \\
&= b_1 \sum (xy - n\bar{x}\bar{y} - n\bar{x}\bar{y} + n\bar{x}\bar{y}) \\
&= b_1 \sum xy - n\bar{y}b_1\bar{x}
\end{aligned}$$

แต่ $b_1\bar{x} = \bar{y} - b_0$

$$\begin{aligned}
\therefore \quad SSR &= b_1 \sum xy - n\bar{y}(\bar{y} - b_0) \\
&= b_1 \sum xy - n\bar{y}^2 + n\bar{y}b_0 \\
&= b_1 \sum xy - n\bar{y}^2 + b_0 \sum y \\
&= b_0 \sum y + b_1 \sum xy - n\bar{y}^2
\end{aligned}$$

จากสมการ $b^T [X^TY] = b_0\sum y + b_1\sum xy$

$= b_0 \sum y + b_1 \sum xy$

$$\therefore \quad SSR = b^T [X^TY] - n\bar{y}^2$$

ค่า *SSE* หาได้จาก

$$\begin{aligned}
SSE &= SST - SSR \\
&= \{[Y^TY] - n\bar{y}^2\} - \{b^T [X^TY] - n\bar{y}^2\} \\
&= [Y^TY] - b^T [X^TY]
\end{aligned}$$

ตารางที่ 5.2 ตารางวิเคราะห์ความแปรปรวนของการถดถอยเชิงเส้นตรงอย่างง่ายในรูปของเมทริกซ์

Source of variation	df	SS	MS
Regression	1	$b^T[X^TY] - n\bar{y}^2$	$b^T[X^TY] - n\bar{y}^2$
Residual (Error)	n - 2	$[Y^TY] - b^T[X^TY]$	$\dfrac{[Y^TY]-b^T[X^TY]}{n-2}$
Total	n - 1	$[Y^TY] - n\bar{y}^2$	

สถิติที่ใช้ทดสอบสมมุติฐาน $H_0 : \beta_1 = 0$ และ $H_1 : \beta_1 \neq 0$ คือ

$$F = \frac{MSR}{MSE}$$

$$= \frac{\left\{ b^T \left[X^TY\right] - n\bar{y}^2 \right\}(n-2)}{\left[Y^TY\right] - b^T\left[X^TY\right]}$$

5.4 สัมประสิทธิ์สหสัมพันธ์ในรูปของเมทริกซ์

ค่าสัมประสิทธิ์สหสัมพันธ์ เขียนแทนด้วยสัญลักษณ์ r อาจจะพิจารณาจาก r^2

เนื่องจาก $r^2 = \dfrac{SSR}{SST}$

ดังนั้น ค่า r^2 เมื่อเขียนในรูปของเมทริกซ์ จะได้ดังนี้

$$r^2 = \frac{b^T\left[X^TY\right] - n\bar{y}^2}{\left[Y^TY\right] - n\bar{y}^2}$$

ตัวอย่างที่ 5.2 สมมุติตัวแปร y และ ตัวแปร x มีค่าดังนี้

y	12	16	10	8	20	6
x	8	10	5	7	14	4

จงหาค่า SST , SSR , SSE , และ r^2 ด้วยวิธีเมทริกซ์

ให้ $$Y = \begin{bmatrix} 12 \\ 16 \\ 10 \\ 8 \\ 20 \\ 6 \end{bmatrix}$$

$$[Y^TY] = \begin{bmatrix} 12 & 16 & 10 & 8 & 20 & 6 \end{bmatrix} \begin{bmatrix} 12 \\ 16 \\ 10 \\ 8 \\ 20 \\ 6 \end{bmatrix} = 1{,}000$$

ให้ $$X = \begin{bmatrix} 1 & 8 \\ 1 & 10 \\ 1 & 5 \\ 1 & 7 \\ 1 & 14 \\ 1 & 4 \end{bmatrix}$$

$$\left[X^T Y \right] = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 8 & 10 & 5 & 7 & 14 & 4 \end{bmatrix} \begin{bmatrix} 12 \\ 16 \\ 10 \\ 8 \\ 20 \\ 6 \end{bmatrix}$$

$$[X^T Y] = \begin{bmatrix} 72 \\ 666 \end{bmatrix}$$

ให้ $b = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$

จากตัวอย่างที่ 5.1 หาค่า b_0 และ b_1 ได้ดังนี้

$b_0 = 1.09$

$b_1 = 1.36$

$$\therefore \; b = \begin{bmatrix} 1.09 \\ 1.36 \end{bmatrix}$$

$$\bar{y} = \frac{12 + 16 + 10 + 8 + 20 + 6}{6} = 12$$

จากสูตร

$$\begin{aligned} SST &= \left[Y^T Y \right] - n\bar{y}^2 \\ &= 1{,}000 - 6(12)^2 \\ &= 136 \end{aligned}$$

$$SSR = b^T [X^T Y] - n\bar{y}^2$$

$$= \begin{bmatrix} 1.09 & 1.36 \end{bmatrix} \begin{bmatrix} 72 \\ 666 \end{bmatrix} - 6(12)^2$$

$$= 984.24 - 864$$

$$= 120.24$$

$$SSE = SST - SSR$$

$$= 136 - 120.24 = 15.76$$

ค่า $r^2 = \dfrac{SSR}{SST}$

$$= \frac{120.24}{136} = 0.884$$

หรือ $r^2 = \dfrac{b^T [X^T Y] - n\bar{y}^2}{[Y^T Y] - n\bar{y}^2}$

$$= \frac{984.24 - 864}{1{,}000 - 864}$$

$$= \frac{120.24}{136} = 0.884$$

นั่นคือ เมื่อรู้ค่า x จะคาดคะเนค่า y ได้ถูกประมาณ 0.884 X 100 = 88.40 % และหากทำการวิเคราะห์จากตารางวิเคราะห์ความแปรปรวนของการถดถอยเชิงเส้นตรงอย่างง่ายในรูปของเมทริกซ์ อาจจะทำได้โดยการสร้างตารางวิเคราะห์ความแปรปรวนได้ดังนี้

ตารางวิเคราะห์ความแปรปรวน

Source of variation	df	SS	MS
Regression	1	120.24	120.24
Residual (Error)	4	15.76	3.94
Total	5	136.0	

สถิติที่ใช้ทดสอบสมมุติฐาน $H_0 : \beta_1 = 0$ และ $H_1 : \beta_1 \neq 0$ คือ

$$F = \frac{MSR}{MSE}$$

$$= \frac{120.24}{3.94}$$

$$= 30.52$$

จากการใช้ตารางวิเคราะห์ความแปรปรวนพบว่า เราปฏิเสธ $H_0 : \beta_1 = 0$ แสดงว่า ความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตามมีความสัมพันธ์กัน สอดคล้องกับการหาค่า r^2 ที่มีค่าเท่ากับ 0.884 หรือ ตัวแปรอิสระอธิบายตัวแปรตามได้ 88.4 % นั่นเอง

ตัวอย่างที่ 5.3 จากข้อมูลเกี่ยวกับค่าใช้จ่าย(บาท)กับพื้นที่ใช้สอย(ตารางเมตร) ปรากฏดังตาราง จงใช้กระบวนการทางเมทริกซ์ วิเคราะห์หารูปแบบสมการถดถอย

ค่าใช้จ่าย	19583	20263	20325	26800	29470	26610
พื้นที่ใช้สอย	3346	3114	3554	4642	4669	4888

จัดรูปแบบ X และ Y ได้ดังนี้

$$
\begin{array}{cc}
X & Y \\
\begin{bmatrix} 1 & 3346 \\ 1 & 3114 \\ 1 & 3554 \\ 1 & 4642 \\ 1 & 4669 \\ 1 & 4888 \end{bmatrix} & \begin{bmatrix} 19583 \\ 20263 \\ 20325 \\ 26800 \\ 29470 \\ 26610 \end{bmatrix}
\end{array}
$$

คำนวณหาค่า $[X^TY]$ โดยนำมาจัดรูปแบบดังนี้

$$
\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 3346 & 3114 & 3554 & 4642 & 4669 & 4888 \end{bmatrix}
\begin{bmatrix} 19583 \\ 20263 \\ 20325 \\ 26800 \\ 29470 \\ 26610 \end{bmatrix}
$$

คำนวณหาค่า $[X^TY]$

$$
[X^TY] = \begin{bmatrix} 143051 \\ 592929460 \end{bmatrix}
$$

คำนวณหาค่า b

$$
b = \begin{bmatrix} 3154.699888 \\ 5.126287559 \end{bmatrix}
$$

กล่าวโดยสรุปจะได้ $b_0 = 3154.699888$

$b_1 = 5.126287559$

สมการถดถอยคือ $\hat{y} = 3154.699888 + 5.126287559x$

หากพิจารณาเปรียบเทียบกับโปรแกรมสำเร็จรูป EXCEL พบว่ามีความสอดคล้องกัน ดังนี้

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.93678
R Square	0.877556
Adjusted R Square	0.846945
Standard Error	1672.707
Observations	6

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	80211788	80211788	28.66807	0.005869
Residual	4	11191795	2797949		
Total	5	91403583			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	3154.7	3923.563	0.80404	0.46644
X Variable 1	5.126288	0.957423	5.354257	0.005869

ตัวอย่างที่ 5.4 จากข้อมูลต่อไปนี้ จงหาสมการถดถอยโดยใช้กระบวนการทางเมทริกซ์

x	8	10	12	15	16	5	7	14	4
y	12	16	16	20	20	10	8	20	6

จัดรูปแบบของ x และ y คำนวณหาค่า $[X^TY]$ โดยนำมาจัดรูปแบบดังนี้

$$\begin{array}{cc} X & Y \\ \begin{bmatrix} 1 & 8 \\ 1 & 10 \\ 1 & 12 \\ 1 & 15 \\ 1 & 16 \\ 1 & 5 \\ 1 & 7 \\ 1 & 14 \\ 1 & 4 \end{bmatrix} & \begin{bmatrix} 12 \\ 16 \\ 16 \\ 20 \\ 20 \\ 10 \\ 8 \\ 20 \\ 6 \end{bmatrix} \end{array} \qquad [X^T Y] = \begin{bmatrix} 128 \\ 1478 \end{bmatrix}$$

คำนวณหาค่า b

$$b = \begin{bmatrix} 2.225 \\ 1.187 \end{bmatrix}$$

กล่าวโดยสรุปจะได้ $b_0 = 2.225$

$b_1 = 1.187$

สมการถดถอยคือ $y = 2.225 + 1.187x$

จากตัวอย่างที่กล่าวมาจะเห็นว่าหากตัวแปรอิสระมีจำนวนมากอาจจะใช้วิธีทางเมทริกซ์ในการวิเคราะห์หาความสัมพันธ์ของตัวแปรได้ ซึ่งยังมีวิธีการวิเคราะห์ที่เหมาะสมกับข้อมูลแต่ละประเภทอีกมาก โดยจะได้กล่าวในตอนต่อไป

5.5 การวิเคราะห์การถดถอยพหุคูณด้วยวิธีเมทริกซ์

การนำวิธีทางเมทริกซ์มาช่วยในการวิเคราะห์การถดถอยพหุคูณ โดยมีขั้นตอนดังที่กล่าวมาเช่นเดียวกับการวิเคราะห์การถดถอยอย่างง่าย โดยมี ตัวแบบสมการถดถอยเชิงพหุคูณของประชากรเขียนในรูปเมทริกซ์จะได้ ดังนี้

$$\begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_N \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & X_{13}\ldots\ldots & X_{1k} \\ 1 & X_{11} & X_{12} & X_{13}\ldots\ldots & X_{1k} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ 1 & X_{11} & X_{12} & X_{13}\ldots\ldots & X_{1k} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ . \\ . \\ . \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ . \\ \varepsilon_N \end{bmatrix}$$

โดย Y , X , β และ ε แทน เมทริกซ์ และเขียนในรูปเมทริกซ์

จากสมการ $[Y] = [X][\beta] + [\varepsilon]$ เมื่อสุ่มตัวอย่างจากประชากรมา n ชุด เมทริกซ์ Y, X, b และทำการวิเคราะห์หาสัมประสิทธิ์การถดถอยด้วยวิธีทางเมทริกซ์ เช่นเดียวกันที่ไดกล่าวมาแล้ว คือ

$$[\,b\,] = [X^T X]^{-1}\ [X^T Y]$$

$$[\hat{Y}] = [X][\mathrm{b}]$$

ตัวอย่างที่ 5.5 จากข้อมูลต่อไปนี้ จงหาสมการถดถอยโดยใช้กระบวนการทางเมทริกซ์

x_1	x_2	y
12	5	72
11	8	76
15	6	78
10	5	70
11	3	68
16	9	80
14	12	82

จัดรูปแบบของตัวแปรในรูปแบบเมทริกซ์ X และ Y ดังนี้

จัดรูปแบบเมทริกซ์ X และ Y

$$X = \begin{array}{cccc} & & x_1 & x_2 \\ \end{array} \begin{bmatrix} 1 & 12 & 5 \\ 1 & 11 & 8 \\ 1 & 15 & 6 \\ 1 & 10 & 5 \\ 1 & 11 & 3 \\ 1 & 16 & 9 \\ 1 & 14 & 12 \end{bmatrix} \qquad Y = \begin{array}{c} y \end{array} \begin{bmatrix} 72 \\ 76 \\ 78 \\ 70 \\ 68 \\ 80 \\ 82 \end{bmatrix}$$

วิเคราะห์หา $X^T X$ และ $(X^T X)^{-1}$

$$X^T X = \begin{bmatrix} 7 & 89 & 48 \\ 89 & 1163 & 633 \\ 48 & 633 & 384 \end{bmatrix}$$

$$(X^T X)^{-1} = \begin{bmatrix} 5.428 & -0.448 & 0.061 \\ -0.448 & 0.045 & -0.019 \\ 0.061 & -0.019 & 0.026 \end{bmatrix}$$

จาก $\quad [\, b \,] = [X^T X]^{-1} \; [X^T Y]$

จะได้

$$[\, b \,] = \begin{bmatrix} 54.2944 \\ 1.00816 \\ 1.17110 \end{bmatrix}$$

กล่าวโดยสรุปจะได้ $b_0 = 54.2944$

$b_1 = 1.00816$

$b_2 = 1.17110$

สมการถดถอยคือ $\hat{y} = 54.2944 + 1.00816x_1 + 1.17110x_2$

หากพิจารณาเปรียบเทียบกับโปรแกรมสำเร็จรูป EXCEL พบว่ามีความสอดคล้องกัน ดังนี้

	Coefficients
Intercept	54.2944
x_1	1.00816
x_2	1.17110

ตัวอย่างที่ 5.6 จากข้อมูลต่อไปนี้ จงหาสมการถดถอยโดยใช้กระบวนการทางเมทริกซ์

ความหนาของผนัง (นิ้ว) :x_1	อุณหภูมิภายนอก (^{0}F): x_2	ปริมาณ GAS (ลูกบาศก์ฟุต):y
6	40	20.3
12	40	26.9
8	49	22.1

จัดรูปแบบของตัวแปร ดังนี้ เมทริกซ์ X และ Y

$$X = \begin{array}{c} \begin{array}{ccc} & x_1 & x_2 \end{array} \\ \begin{bmatrix} 1 & 6 & 40 \\ 1 & 12 & 40 \\ 1 & 8 & 49 \end{bmatrix} \end{array} \qquad Y = \begin{array}{c} y \\ \begin{bmatrix} 20.3 \\ 26.9 \\ 22.1 \end{bmatrix} \end{array}$$

วิเคราะห์หา $X^T X$ และ $(X^T X)^{-1}$

$$X^T X = \begin{bmatrix} 3 & 26 & 129 \\ 26 & 244 & 1112 \\ 129 & 1112 & 5601 \end{bmatrix}$$

$$(X^T X)^{-1} = \begin{bmatrix} 44.62 & -0.747 & -0.879 \\ -0.747 & 0.056 & 0.006 \\ -0.879 & 0.006 & 0.019 \end{bmatrix}$$

จาก $\quad [\,b\,] = [X^T X]^{-1}\ [X^T Y]$ จะได้

$$[\,b\,] = \begin{bmatrix} 15.48 \\ 1.1 \\ -0.044 \end{bmatrix}$$

กล่าวโดยสรุปจะได้ $b_0 = 15.48$, $b_1 = 1.1$, $b_2 = -0.044$

สมการถดถอยคือ $\hat{y} = 15.48 + 1.1x_1 - 0.044x_2$

หากพิจารณาเปรียบเทียบกับโปรแกรมสำเร็จรูป EXCEL พบว่ามีความสอดคล้องกัน ดังนี้

	Coefficients
Intercept	15.47778
x_1	1.1
x_2	-0.04444

บทสรุป

การวิเคราะห์การถดถอยโดยใช้วิธีทางเมทริกซ์เป็นกระบวนการวิเคราะห์การถดถอยวิธีหนึ่งที่เป็นพื้นฐานในการนำไปประยุกต์ใช้และการนำเอารูปแบบการวิเคราะห์โดยวิธีเมทริกซ์มาคำนวณซึ่งทำให้ผลที่เกิดจากการวิเคราะห์มีความหมายเดียวกัน ทั้งนี้จะต้องมีความรู้พื้นฐานทางการคำนวณแบบเมทริกซ์ด้วย อาจจะใช้เครื่องคำนวณต่างๆ ที่เป็นโปรแกรมสำเร็จรูปมาช่วยในการวิเคราะห์จะทำให้การคำนวณสะดวกยิ่งขึ้น

แบบฝึกหัดที่ 5

1. กำหนดให้

พื้นที่อาศัย(ตร.ม.)	ราคา(พันบาท)
15	145
38	228
23	150
16	130
16	160
13	114
20	142
24	265

1.1 กำหนดให้ X แทน พื้นที่อาศัย(ตร.ม.) Y แทน ราคา(พันบาท) โดยใช้วิธีทางเมทริกซ์

ก. จงหาค่า X^TX และ X^TY

ข. จงหาค่า b_0 และ b_1

ค. สมการถดถอยของข้อมูลชุดนี้

ง. จงหาความแปรปรวนของ b_0 และ b_1

จ. จงหาค่า SS_R

1.2 จงเปรียบเทียบวิธีทางเมทริกซ์และวิธีการในบทที่ 2 จงอธิบาย

2. ถ้า $X^TX = \begin{bmatrix} 5 & 15 \\ 15 & 55 \end{bmatrix}$ และ $X^TY = \begin{bmatrix} 40 \\ 130 \end{bmatrix}$

และ $\sum Y^2 = 330$

2.1 จงหาสมการถดถอยเชิงเส้นตรงอย่างง่ายของข้อมูลชุดนี้

2.2 จงหาความแปรปรวนของ b_0 และ b_1

2.3 จงหาค่า SS_R

3. ถ้า Y เป็นตัวแปรเกณฑ์ X เป็นตัวแปรอิสระ มีค่าต่างๆดังนี้

$\sum X = 450$ $\sum X^2 = 7\}020$ $\sum XY = 9,900$

$\sum Y = 660$ $\sum Y^2 = 15,270$ $n = 30$

3.1 จงหาค่า b_0 และ b_1 ด้วยวิธีเมทริกซ์

3.2 ด้วยความเชื่อมั่น 95 % β_1 มีค่าอยู่ระหว่างเท่าใด

3.3 ถ้า X = 18 ด้วยความเชื่อมั่น 95 % ช่วงระยะของค่าคาดคะเนค่าเฉลี่ยของ Y จะเป็นเท่าใด

3.4 จงหาค่า SS_E และ SS_R

4. กำหนดให้

$$X = \begin{bmatrix} 1 & 3 \\ 1 & 5 \\ 1 & 8 \\ 1 & 1 \\ 1 & 10 \end{bmatrix} \qquad Y = \begin{bmatrix} 15 \\ 20 \\ 16 \\ 10 \\ 22 \end{bmatrix}$$

4.1 จงหาค่า X^TX และ X^TY

4.2 จงหาค่า b_0 และ b_1

5. เมื่อ Y เป็นตัวแปรตาม X เป็นตัวแปรอิสระ n = 10 และหาค่าต่างๆได้ดังนี้

$$X^TY = \begin{bmatrix} 38 \\ 420 \end{bmatrix}$$

$$Y^TY = [\,1,504\,]$$

$$b = \begin{bmatrix} 1.5 \\ 1.0 \end{bmatrix}$$

และ $\sum X = 240$

5.1 จงหาสมการถดถอยที่ใช้คาดคะเนค่า Y เมื่อรู้ค่า X

5.2 จงหาค่า SS_R และ SS_E

5.3 จงหาค่า r^2

6. จากข้อมูลต่อไปนี้ **จงหาสมการถดถอยโดยใช้กระบวนการทางเมทริกซ์**

Datafile Name: Clouds

Reference: Chambers, Cleveland, Kleiner, and Tukey. (1983). *Graphical Methods for Data Analysis.* Wadsworth International Group, Belmont, CA, 351. Original Source: Simpson, Alsen, and Eden. (1975). A Bayesian analysis of a multiplicative treatment effect in weather modification. *Technometrics 17*, 161-166.

Description: Rainfall from Cloud-Seeding. The rainfall in acre-feet from 52 clouds 26 of which were chosen at random and seeded with silver nitrate.

Number of cases: 9

Variable Names:

Unseeded_Clouds: Amount of rainfall from unseeded clouds (in acre-feet)

Seeded_Clouds: Amount of rainfall from seeded clouds with silver nitrate (in acre-feet)

The Data:

Unseeded_Clouds	Seeded_Clouds	Unseeded_Clouds	Seeded_Clouds
47.3	242.5	41.1	200.7
36.6	198.6	29.0	129.6
28.6	119.0	26.3	118.3
26.1	115.3	24.4	92.4
21.7	40.6		

7. จงหาสมการถดถอยของ Alcohol and Tobacco โดยมีตัวแปรดังนี้ 1) **Region**: Region of Great Britain 2) **Alcohol**: Average weekly household spending on alcoholic beverages in pounds 3) **Tobacco**: Average weekly household spending on tobacco products in pounds

Region	Alcohol	Tobacco	Region	Alcohol	Tobacco
North	6.47	4.03	Yorkshire	6.13	3.76
Northeast	6.19	3.77	East Midlands	4.89	3.34
West Midlands	5.63	3.47	East Anglia	4.52	2.92

Reference: Moore, David S., and George P. McCabe (1989). *Introduction to the Practice of Statistics*, p. 179. Original source: *Family Expenditure Survey*, Department of Employment, 1981 (British official statistics)

8. จากรูปแบบ X และ Y นี้ จงหาสมการถดถอย

$$
X = \begin{bmatrix} 1 & 20 \\ 1 & 12 \\ 1 & 15 \\ 1 & 22 \\ 1 & 14 \\ 1 & 17 \end{bmatrix} \quad Y = \begin{bmatrix} 7 \\ 6 \\ 7 \\ 8 \\ 7 \\ 6 \end{bmatrix}
$$

9. หากคำนวณหาค่าต่อไปนี้แล้ว สมการถดถอยมีความหมายว่าอย่างไร

$$X^TX = \begin{bmatrix} 6 & 100 \\ 100 & 1738 \end{bmatrix} \qquad \text{Determinant of } X^TX = 428$$

$$(X^TX)^{-1} = \begin{bmatrix} 4.061 & -0.234 \\ -0.234 & 0.014 \end{bmatrix} \qquad (X^TX)^{-1}(X^TY) = \begin{bmatrix} 41 \\ 693 \end{bmatrix}$$

10. จงเปรียบเทียบการวิเคราะห์หาสมการถดถอยในรูปแบบที่แตกต่างกันอย่างน้อย 2 รูปแบบ พร้อมทั้งอภิปรายผล
