บทที่ 5
การวิเคราะห์การถดถอยเชิงเส้นพหุ

ในการหาความสัมพันธ์ของตัวแปรอิสระกับตัวแปรตามนั้นบ่อยครั้งที่จำนวนตัวแปรอิสระที่สนใจในการศึกษามีมากกว่าหนึ่งตัวแปร เช่น เวลาที่ใช้ในการหมักขนมปังขึ้นอยู่กับปริมาตรของเหลวที่ใช้ อุณหภูมิในการหมัก ปริมาณยีสต์ที่ใช้ เป็นต้น ความสัมพันธ์เช่นนี้ไม่สามารถใช้การถดถอยโดยใช้การถดถอยเชิงเส้นอย่างง่ายในการวิเคราะห์ได้ การวิเคราะห์การถดถอยเชิงเส้นพหุ (multiple linear regression analysis) เป็นเทคนิคการวิเคราะห์การถดถอยที่เกี่ยวข้องกับตัวแปรอิสระที่มากกว่าหนึ่งตัวแปร การเพิ่มตัวแปรอิสระที่เกี่ยวข้องเข้าในการวิเคราะห์จะทำให้ความถูกต้องของการวิเคราะห์เพิ่มมากขึ้นและค่าคลาดเคลื่อนมาตรฐานของตัวประมาณค่า (standard error of estimates) ลดลง

5.1 ตัวแบบการถดถอยเชิงเส้นพหุ

จากบทที่ผ่านมานั้นสมการถดถอยเชิงเส้นอย่างง่ายมีตัวแปรอิสระเพียง 1 ตัวและความสัมพันธ์ระหว่างตัวแปรสามารถจัดอยู่ในรูปเส้นตรงได้ สำหรับกรณีที่มีตัวแปรอิสระ 2 ตัว (X_1 และ X_2) ที่มีความสัมพันธ์เชิงเส้นกับตัวแปรตามแล้วสมการถดถอยสามารถวาดอยู่ในรูประนาบโดยสามารถเขียนให้อยู่ในรูปของสมการดังนี้

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i \tag{5.1}$$

โดย

Y_i = ค่าของตัวแปรตามที่ในการเก็บข้อมูลครั้งที่ i

β_0 = ค่าเฉลี่ยของตัวแปรตามที่ตัวแปรอิสระทั้งสองมีค่าเป็นศูนย์

β_1, β_2 = ค่าสัมประสิทธิ์การถดถอยของ X_1 และ X_2

ε_i = ค่าความคลาดเคลื่อนในการเก็บข้อมูลครั้งที่ i

ค่า β_1 เป็นค่าคาดหวังของ Y ที่เปลี่ยนแปลงไปเมื่อ X_1 เปลี่ยนแปลงไป 1 หน่วยโดยที่ X_2 มีค่าเท่าเดิมในทำนองเดียวกันค่า β_2 เป็นค่าคาดหวังของ Y ที่เปลี่ยนแปลงไปเมื่อ X_2 เปลี่ยนแปลงไป

1 หน่วยโดยที่ X_1 มีค่าเท่าเดิม เช่น หาก β_1 เท่ากับ 2.5 หมายถึงค่า Y จะเพิ่มขึ้น 2.5 หน่วยเมื่อตัวแปร X_1 มีค่าเพิ่มขึ้น 1 หน่วยโดยที่ตัวแปร X_2 ไม่เปลี่ยนแปลงค่าหรือหาก β_2 เท่ากับ -3.0 หมายถึงค่า Y จะลดลง 3.0 หน่วยเมื่อตัวแปร X_2 มีค่าเพิ่มขึ้น 1 หน่วยโดยที่ตัวแปร X_1 ไม่เปลี่ยนแปลงค่า เป็นต้น

หากมีตัวแปรอิสระ p - 1 ตัวแล้วตัวแบบการถดถอยเชิงเส้นพหุในรูปทั่วไปสามารถแสดงได้ด้วยสมการดังนี้

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \ldots + \beta_{p-1} X_{ip-1} + \varepsilon_i \quad (5.2)$$

โดย $\beta_0, \beta_1, \ldots, \beta_{p-1}$ = ค่าสัมประสิทธิ์การถดถอย

$X_{i1}, X_{i2}, \ldots, X_{ip-1}$ = ค่าของตัวแปรอิสระ p - 1 ตัวในการเก็บข้อมูลครั้งที่ i

ฟังก์ชันระหว่างตัวแปรอิสระ p - 1 ตัวกับตัวแปรตามจะเป็นแนวระนาบหลายระดับ (hyperplane)

หากความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตามไม่เป็นเชิงเส้นกันเช่น

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i2}^2 + \varepsilon_i \quad (5.3)$$

ในกรณีเช่นนี้ความสัมพันธ์ระหว่างตัวแปร X_2 กับตัวแปร Y จะมีลักษณะเป็นพาราโบลา หรือสมการรูปแบบนี้ว่าสมการกำลังสอง (quadratic model) หากให้ $X_{i2}^2 = X_{i3}$ แล้วจะสามารถเขียนสมการได้ดังนี้

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i \quad (5.4)$$

ในกรณีของสมการ (5.1) เป็นสมการที่ตัวแปรอิสระทั้งสองไม่มีปฏิสัมพันธ์กัน (interaction) หรือเรียกว่ามีอิทธิพลแบบรวมกัน (additive effect) คือการที่ตัวแปรอิสระตัวหนึ่งมีผลต่อการเปลี่ยนแปลงค่าของตัวแปรตามโดยไม่ได้รับอิทธิพลของตัวแปรอิสระอีกตัว หากตัวแปรอิสระทั้งสองมีปฏิสัมพันธ์กันแล้วการเปลี่ยนแปลงค่าของตัวแปรตามที่เกิดจากตัวแปรอิสระตัวหนึ่งจะเปลี่ยนค่าไปมากหรือน้อยขึ้นอยู่กับระดับของตัวแปรอิสระตัวที่สองซึ่งแสดงได้ด้วยสมการดังนี้

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \varepsilon_i \quad (5.5)$$

โดย β_3 = ค่าสัมประสิทธิ์การถดถอยที่เกิดจากปฏิสัมพันธ์

หากให้ $X_{i1} X_{i2} = X_{i3}$ แล้วจะสามารถเขียนสมการได้ดังนี้

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i \quad (5.6)$$

จะเห็นว่าสมการ (5.5) และ (5.6) เป็นสมการถดถอยเชิงเส้นพหุนั่นเอง ดังนั้นสมการกำลังสองและสมการที่มีปฏิสัมพันธ์คือรูปแบบหนึ่งของสมการถดถอยเชิงเส้นพหุนั่นเอง

สมการถดถอยเชิงเส้นพหุมีข้อตกลงที่เกี่ยวข้องกับความคลาดเคลื่อนเช่นเดียวกันกับสมการถดถอยเชิงเส้นอย่างง่ายดังนี้คือ

(1) ε_i มีการแจกแจงแบบปกติโดยมีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนคงที่เท่ากับ σ^2 หรือ $\varepsilon_i \sim N(0,\sigma^2)$

(2) ε_i แต่ละค่าเป็นอิสระต่อกัน

จากข้อตกลงดังกล่าวทำให้ค่าของตัวแปรตามมีการแจกแจงแบบปกติหรือสามารถเขียนได้เป็น $Y_i \sim N(\beta_0 + \beta_1 X_{i1} + ... + \beta_p X_{ip-1}, \sigma^2)$

5.2 สมการถดถอยในรูปของเมตริกซ์

การคำนวณสมการถดถอยพหุนั้นโดยการใช้เมตริกซ์จะช่วยให้การคำนวณง่ายและรวดเร็ว โดยเมตริกซ์ที่ใช้นั้นจะคล้ายคลึงกับในบทที่ 4 แต่เพิ่มจำนวนตัวแปรอิสระขึ้นดังนั้นขนาดของเมตริกซ์ในการคำนวณจะใหญ่ขึ้นตามจำนวนตัวแปรอิสระดังนี้

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{5.7}$$

หรือ

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1p-1} \\ 1 & X_{21} & X_{22} & \cdots & X_{2p-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{np-1} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

พบว่าเมตริกซ์ $\mathbf{Y}$ และ $\boldsymbol{\varepsilon}$ นั้นเหมือนกับในสมการถดถอยเชิงเส้นอย่างง่ายในขณะที่เมตริกซ์ $\mathbf{X}$ จะมีจำนวนหลักจะเท่ากับจำนวนตัวแปรอิสระบวก 1 หรือ p และเวคเตอร์ $\boldsymbol{\beta}$ จะมีจำนวนแถวเท่ากับ p

นอกจากนี้ $\boldsymbol{\varepsilon}$ มีเมตริกซ์ของค่าเฉลี่ยเป็น **0** และเมตริกซ์ของความแปรปรวนคือ

$$\boldsymbol{\sigma}^2 = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2\mathbf{I}$$

5.3 การประมาณค่าพารามิเตอร์

การประมาณค่าพารามิเตอร์ของสมการถดถอยพหุใช้วิธีการเช่นเดียวกับสมการถดถอยเชิงเส้นอย่างง่ายคือการใช้วิธีกำลังสองน้อยที่สุด หากกำหนดให้ Q เป็นค่าผลรวมกำลังสองของความคลาดเคลื่อนที่น้อยที่สุดหรือ

$$Q = \sum_{i=1}^{n}\left(Y_i - \beta_0 - \beta_1 X_{i1} - \ldots - \beta_{p-1} X_{ip-1}\right)^2$$

$$= \sum_{i=1}^{n}\left(Y_i - \beta_0 - \sum_{j=1}^{p-1}\beta_j X_{ij}\right)^2 \quad (5.8)$$

ในการคำนวณเพื่อให้ได้ค่า Q ที่น้อยที่สุดนั้นจะต้องทำการหาอนุพันธ์ย่อยเทียบกับค่า β_j แต่ละค่าโดยสามารถเขียนสมการทั้งสองได้ดังนี้

$$\left.\frac{\partial Q}{\partial \beta_0}\right|_{b_0, b_1, \ldots, b_{p-1}} = -2\sum_{i=1}^{n}\left(Y_i - b_0 - \sum_{j=1}^{p-1} b_j X_{ij}\right) = 0 \quad (5.9)$$

และ

$$\left.\frac{\partial Q}{\partial \beta_j}\right|_{b_0, b_1, \ldots, b_{p-1}} = -2\sum_{i=1}^{n}\left(Y_i - b_0 - \sum_{j=1}^{p-1} b_j X_{ij}\right)X_{ij} = 0 \quad (5.10)$$

ดังนั้นสมการปกติสามารถเขียนในรูปของเมตริกซ์ดังนี้

$$\mathbf{(X'X)\,b} = \mathbf{X'\,Y} \quad (5.11)$$

หรือ

$$\begin{bmatrix} n & \sum_{i=1}^{n} X_{i1} & \sum_{i=1}^{n} X_{i2} & \cdots & \sum_{i=1}^{n} X_{ip-1} \\ \sum_{i=1}^{n} X_{i1} & \sum_{i=1}^{n} X_{i2}^2 & \sum_{i=1}^{n} X_{i1}X_{i2} & \cdots & \sum_{i=1}^{n} X_{i1}X_{ip-1} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum_{i=1}^{n} X_{ip-1} & \sum_{i=1}^{n} X_{ip-1}X_{i1} & \sum_{i=1}^{n} X_{ip-1}X_{i2} & \cdots & \sum_{i=1}^{n} X_{ip-1}^2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{p-1} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n} X_{i1}Y_i \\ \vdots \\ \sum_{i=1}^{n} X_{ip-1}Y_i \end{bmatrix}$$

โดย $\mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{p-1} \end{bmatrix}$

$$\mathbf{X'X} = \begin{bmatrix} n & \sum_{i=1}^{n} X_{i1} & \sum_{i=1}^{n} X_{i2} & \cdots & \sum_{i=1}^{n} X_{ip-1} \\ \sum_{i=1}^{n} X_{i1} & \sum_{i=1}^{n} X_{i2}^2 & \sum_{i=1}^{n} X_{i1}X_{i2} & \cdots & \sum_{i=1}^{n} X_{i1}X_{ip-1} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum_{i=1}^{n} X_{ip-1} & \sum_{i=1}^{n} X_{ip-1}X_{i1} & \sum_{i=1}^{n} X_{ip-1}X_{i2} & \cdots & \sum_{i=1}^{n} X_{ip-1}^2 \end{bmatrix} \quad (5.12)$$

$$\mathbf{X'\,Y} \quad = \quad \begin{bmatrix} \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n} X_{i1}Y_i \\ \vdots \\ \sum_{i=1}^{n} X_{ip-1}Y_i \end{bmatrix} \tag{5.13}$$

จากสมการที่ (5.11) สามารถหาเวกเตอร์ **b** ได้ดังนี้

$$\mathbf{b} \quad = \quad \mathbf{(X'X)^{-1}\,X'\,Y} \tag{5.14}$$

โดย **X'X** ต้องสามารถหาเมตริกซ์ผกผันได้และเป็นเมตริกซ์สมมาตรที่มีขนาด $p \times p$ และค่าของสมาชิกในแนวเฉียงเป็นผลรวมกำลังสองของค่าในแต่ละหลัก

ตัวอย่างที่ 5.1 นักเคมีต้องการศึกษาความสัมพันธ์ระหว่างเวลาที่ใช้การเกิดสารชนิดหนึ่งมีหน่วยเป็นกรัม (Y) กับปริมาณตัวเร่งปฏิกิริยาชนิดที่ 1 มีหน่วยเป็นมิลลิกรัม (X_1) และปริมาณตัวเร่งปฏิกิริยาชนิดที่ 2 มีหน่วยเป็นมิลลิกรัม (X_2) จงประมาณค่าพารามิเตอร์การถดถอยโดยวิธีเมตริกซ์

Y	15.4	10.2	19.8	9.7	12.3	10.8	16.4	15.8	21.0	17.5
X_1	3.0	7.5	4.5	4.0	5.5	7.0	5.0	3.5	3.0	6.0
X_2	45.0	55.0	43.5	50.0	47.0	49.0	53.5	61.0	45.5	55.0

วิธีทำ

จากข้อมูลดังกล่าวสามารถเขียนอยู่ในรูปของเมตริกซ์ได้ดังนี้

$$\mathbf{X} = \begin{bmatrix} 1 & 3.0 & 45.0 \\ 1 & 7.5 & 55.0 \\ \vdots & \vdots & \vdots \\ 1 & 6.0 & 55.0 \end{bmatrix} \qquad \mathbf{Y} = \begin{bmatrix} 15.4 \\ 10.2 \\ \vdots \\ 17.5 \end{bmatrix}$$

ดังนั้น

$$\mathbf{X'X} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 3.0 & 7.5 & 4.5 & \cdots & 6.0 \\ 45.0 & 55.0 & 43.5 & \cdots & 55.0 \end{bmatrix} \begin{bmatrix} 1 & 3.0 & 45.0 \\ 1 & 7.5 & 55.0 \\ \vdots & \vdots & \vdots \\ 1 & 6.0 & 55.0 \end{bmatrix}$$

$$= \begin{bmatrix} 10 & 49 & 504.5 \\ 49 & 263 & 2492.25 \\ 504.5 & 2492.25 & 25730.75 \end{bmatrix}$$

$$(\mathbf{X'X})^{-1} = \begin{bmatrix} 9.30375 & -0.0580211 & -0.176798 \\ -0.0580211 & 0.0466504 & -0.00338089 \\ -0.176798 & -0.00338089 & 0.00383279 \end{bmatrix}$$

และ

$$\mathbf{X'Y} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 3.0 & 7.5 & 4.5 & \cdots & 6.0 \\ 45.0 & 55.0 & 43.5 & \cdots & 55.0 \end{bmatrix} \begin{bmatrix} 15.4 \\ 10.2 \\ \vdots \\ 17.5 \end{bmatrix}$$

$$= \begin{bmatrix} 148.90 \\ 699.15 \\ 7466.80 \end{bmatrix}$$

ดังนั้น

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

$$= \begin{bmatrix} 9.30375 & -0.0580211 & -0.176798 \\ -0.0580211 & 0.0466504 & -0.00338089 \\ -0.176798 & -0.00338089 & 0.00383279 \end{bmatrix} \begin{bmatrix} 148.90 \\ 699.15 \\ 7466.80 \end{bmatrix}$$

$$= \begin{bmatrix} 24.649 \\ -1.268 \\ -0.070 \end{bmatrix}$$

หรือ $b_0 = 24.649$, $b_1 = -1.268$ และ $b_2 = -0.070$

สามารถอธิบายค่า b_1 ได้ว่าหากกำหนดปริมาณตัวเร่งปฏิกิริยาตัวที่ 2 คงที่เมื่อเพิ่มปริมาณตัวเร่งปฏิกิริยาตัวที่ 1 ขึ้น 1 มิลลิกรัมแล้วเวลาที่ใช้ในการเกิดสารจะลดลง 1.268 นาทีในทำนองเดียวกันสามารถอธิบายค่า b_2 ได้ว่าหากกำหนดปริมาณตัวเร่งปฏิกิริยาตัวที่ 1 คงที่เมื่อเพิ่มปริมาณตัวเร่งปฏิกิริยาตัวที่ 2 ขึ้น 1 มิลลิกรัมแล้วเวลาที่ใช้ในการเกิดสารจะลดลง 0.070 นาที ดังนั้นสมการถดถอยคือ

$$\hat{Y}_i = 24.649 - 1.268X_{i1} - 0.070X_{i2}$$

5.4 ค่าพยากรณ์

ค่าพยากรณ์ที่ได้จากสมการถดถอยพหุ ($\hat{Y}$) สามารถเขียนอยู่ในรูปเมตริกซ์ดังนี้

$$\hat{\mathbf{Y}} = \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{bmatrix}$$

โดยสามารถคำนวณค่าพยากรณ์โดยใช้เมตริกซ์ได้ดังนี้

$$\hat{\mathbf{Y}} = \mathbf{X\,b} \tag{5.15}$$

$$= \mathbf{X(X'X)^{-1}\,X'\,Y} \tag{5.16}$$

หรือ

$$\hat{\mathbf{Y}} = \mathbf{HY} \tag{5.17}$$

ตัวอย่างที่ 5.2 จากตัวอย่างที่ 5.1 จงหาค่าพยากรณ์โดยการใช้เมตริกซ์

วิธีทำ

$$\hat{\mathbf{Y}} = \mathbf{X\,b}$$

$$= \begin{bmatrix} 1 & 3.0 & 45.0 \\ 1 & 7.5 & 55.0 \\ \vdots & \vdots & \vdots \\ 1 & 6.0 & 55.0 \end{bmatrix} \begin{bmatrix} 24.649 \\ -1.268 \\ -0.070 \end{bmatrix}$$

$$= \begin{bmatrix} 17.683 \\ 11.273 \\ \vdots \\ 13.175 \end{bmatrix}$$

5.5 ส่วนเหลือ

ส่วนเหลือ (e) สามารถเขียนได้อยู่ในรูปเวกเตอร์ดังนี้

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

หากเขียนส่วนเหลือในรูปของเวกเตอร์สามารถเขียนได้ดังนี้

$$\boldsymbol{e} = \mathbf{Y - Xb} \tag{5.18}$$

หรือเขียนอยู่ในรูปของ **H** ได้ดังนี้

$$\boldsymbol{e} = \mathbf{Y - HY}$$

$$= \mathbf{(I - H)Y} \tag{5.19}$$

และสามารถเขียนเมตริกซ์ความแปรปรวนของส่วนเหลือได้ดังนี้

$$\boldsymbol{\sigma}^2(\mathrm{e}) = \boldsymbol{\sigma}^2\mathbf{(I - H)} \tag{5.20}$$

เนื่องจากเมตริกซ์ความแปรปรวนนี้เป็นของพารามิเตอร์ซึ่งสามารถประมาณได้โดยการใช้ความแปรปรวนที่ได้จากตัวอย่างคือ

$$s^2(\mathrm{e}) \quad = \quad \boldsymbol{MSE}(\mathbf{I} - \mathbf{H}) \tag{5.21}$$

ตัวอย่างที่ 5.3 จากตัวอย่างที่ 5.1 จงหาค่าส่วนเหลือโดยการใช้เมตริกซ์

วิธีทำ

$$\begin{aligned} \boldsymbol{e} &= \mathbf{Y} - \hat{\mathbf{Y}} \\ &= \begin{bmatrix} 15.4 \\ 10.2 \\ \vdots \\ 17.5 \end{bmatrix} - \begin{bmatrix} 17.683 \\ 11.273 \\ \vdots \\ 13.175 \end{bmatrix} \\ &= \begin{bmatrix} -2.282 \\ -1.073 \\ \vdots \\ 4.325 \end{bmatrix} \end{aligned}$$

5.6 คุณสมบัติของตัวประมาณค่าพารามิเตอร์

ตัวประมาณค่าพารามิเตอร์ของตัวแบบสมการถดถอยเชิงเส้นพหุมีคุณสมบัติเช่นเดียวกับตัวประมาณค่าพารามิเตอร์ที่ได้สมการถดถอยเชิงเส้นอย่างง่ายคือมีคุณสมบัติเป็นตัวประมาณค่าที่ดีหรือ BLUE (best linear unbiased estimator) ของพารามิเตอร์ $\boldsymbol{\beta}$ กล่าวคือเป็นตัวประมาณค่าเชิงเส้นที่ไม่เอนเอียงและมีความแปรปรวนที่น้อยที่สุดในบรรดาตัวประมาณค่าทั้งหมดที่ได้จากวิธีอื่นค่าคาดหวังของตัวประมาณค่าจะมีค่าเท่ากับค่าพารามิเตอร์หรือ $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$

เมตริกซ์ความแปรปรวนและความแปรปรวนร่วมของตัวประมาณค่าพารามิเตอร์เป็นเมตริกซ์ที่มีขนาด $p \times p$ โดยที่ค่าในแนวที่ j ของแนวเฉียงคือค่าความแปรปรวนของตัวประมาณค่า b_j เมตริกซ์สามารถเขียนได้ดังนี้

$$\boldsymbol{\sigma}^2\{b\} = \begin{bmatrix} \sigma^2\{b_0\} & \sigma\{b_0, b_1\} & \cdots & \sigma\{b_0, b_{p-1}\} \\ \sigma\{b_0, b_1\} & \sigma^2\{b_1\} & \cdots & \sigma\{b_1, b_{p-1}\} \\ \vdots & \vdots & & \vdots \\ \sigma\{b_1, b_0\} & \sigma\{b_{p-0}, b_1\} & \cdots & \sigma^2\{b_{p-1}\} \end{bmatrix} \tag{5.22}$$

หรือ

$$\boldsymbol{\sigma}^2\{b\} = \sigma^2\left(\boldsymbol{X'X}\right)^{-1} \tag{5.23}$$

เนื่องจาก σ^2 สามารถประมาณได้โดยการใช้ *MSE* ดังนั้นเมตริกซ์ของความแปรปรวนและความแปรปรวนร่วมคือ

$$\boldsymbol{S}^2\{b\} = MSE(\boldsymbol{X}'\boldsymbol{X})^{-1} \tag{5.24}$$

5.7 การวิเคราะห์ความแปรปรวน

การวิเคราะห์ความแปรปรวนเป็นการทดสอบสัมประสิทธิ์การถดถอยทั้งหมดมีค่าเท่ากับ 0 หรือไม่ หรือ $\mathrm{H}_0 : \beta_1 = \beta_2 = \ldots = \beta_{p-1} = 0$ หรือ $\boldsymbol{\beta} = \mathbf{0}$ และ H_1 : อย่างน้อย 1 ตัวที่มีค่าไม่เท่ากับ 0 และใช้การทดสอบเอฟเช่นเดียวกับการวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย หากปฏิเสธสมมติฐานหลักแสดงว่ามีตัวแปรอิสระอย่างน้อย 1 ตัวที่มีความสัมพันธ์เชิงเส้นตรงกับตัวแปรตาม การวิเคราะห์ความแปรปรวนสามารถคำนวณในรูปของเมตริกซ์ได้โดยในที่นี้จะกล่าวถึงเฉพาะผลรวมกำลังสองดังนี้

5.7.1 ผลรวมกำลังสองทั้งหมด

ผลรวมกำลังสองทั้งหมด (*SST*) เขียนในรูปของเมตริกซ์ได้ดังนี้

$$SST = \mathbf{Y}'\mathbf{Y} - \left(\frac{1}{n}\right)\mathbf{Y}'\mathbf{J}\mathbf{Y} \tag{5.25}$$

หรือ

$$= \mathbf{Y}'\left[\mathbf{I} - \left(\frac{1}{n}\right)\mathbf{J}\right]\mathbf{Y} \tag{5.26}$$

โดย *SST* มีองศาเสรีเท่ากับ $n-1$

5.7.2 ผลรวมกำลังสองความคลาดเคลื่อน

ผลรวมกำลังสองความคลาดเคลื่อน (*SSE*) เขียนในรูปของเมตริกซ์ได้ดังนี้

$$SSE = \mathbf{Y}'\mathbf{Y} - \mathbf{b}'\mathbf{X}'\mathbf{Y} \tag{5.27}$$

หรือ

$$= \mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y} \tag{5.28}$$

โดย *SSE* มีองศาเสรีเท่ากับ $n-p$

5.7.3 ผลรวมกำลังสองถดถอย

ผลรวมกำลังสองถดถอย (*SSR*) เขียนในรูปของเมตริกซ์ได้ดังนี้

$$SSR = \mathbf{b}'\mathbf{X}'\mathbf{Y} - \left(\frac{1}{n}\right)\mathbf{Y}'\mathbf{J}\mathbf{Y} \tag{5.29}$$

หรือ

$$= \mathbf{Y}'\left[\mathbf{H} - \left(\frac{1}{n}\right)\mathbf{J}\right]\mathbf{Y} \qquad (5.30)$$

โดย *SSR* มีองศาเสรีเท่ากับ $p-1$ เนื่องจากมีการประมาณค่าพารามิเตอร์ทั้งหมด p ค่า

ตารางวิเคราะห์ความแปรปรวนสามารถเขียนได้ดังนี้

Source of variation	*SS*	*df*	*MS*	*F*
Regression	$SSR = \mathbf{b}'\mathbf{X}'\mathbf{Y} - \left(\frac{1}{n}\right)\mathbf{Y}'\mathbf{J}\mathbf{Y}$	$p-1$	$MSR = \frac{SSR}{p-1}$	$F = \frac{MSR}{MSE}$
Error	$SSE = \mathbf{Y}'\mathbf{Y} - \mathbf{b}'\mathbf{X}'\mathbf{Y}$	$n-p$	$MSE = \frac{SSE}{n-p}$	
Total	$SST = \mathbf{Y}'\mathbf{Y} - \left(\frac{1}{n}\right)\mathbf{Y}'\mathbf{J}\mathbf{Y}$	$n-1$		

ตัวอย่างที่ 5.4 จากตัวอย่างที่ 5.1 จงสร้างตารางวิเคราะห์ความแปรปรวนโดยการใช้เมตริกซ์

วิธีทำ

เนื่องจาก
$$\mathbf{Y}'\mathbf{Y} = \begin{bmatrix}15.4 & 10.2 & \cdots & 17.5\end{bmatrix}\begin{bmatrix}15.4\\10.2\\\vdots\\17.5\end{bmatrix}$$

$$= 2,361.11$$

$$\left(\frac{1}{n}\right)\mathbf{Y}'\mathbf{J}\mathbf{Y} = \left(\frac{1}{10}\right)\begin{bmatrix}15.4 & 10.2 & \cdots & 17.5\end{bmatrix}\begin{bmatrix}1 & 1 & \cdots & 1\\1 & 1 & \cdots & 1\\\vdots & \vdots & & \vdots\\1 & 1 & \cdots & 1\end{bmatrix}\begin{bmatrix}15.4\\10.2\\\vdots\\17.5\end{bmatrix}$$

$$= 2,217.121$$

$$\mathbf{b}'\mathbf{X}'\mathbf{Y} = \begin{bmatrix}24.649 & -1.268 & -0.070\end{bmatrix}\begin{bmatrix}148.90\\699.15\\7466.80\end{bmatrix}$$

$$= 2,258.909$$

ดังนั้น
$$SST = \mathbf{Y}'\mathbf{Y} - \left(\frac{1}{n}\right)\mathbf{Y}'\mathbf{J}\mathbf{Y}$$

$$= 2,361.11 - 2,217.121$$

$$= 143.989$$

$$SSE = \mathbf{Y}'\mathbf{Y} - \mathbf{b}'\mathbf{X}'\mathbf{Y}$$

$$= 2{,}361.11 - 2{,}258.909$$

$$= 102.201$$

$$SSR = SST - SSE$$

$$= 143.989 - 102.201$$

$$= 41.788$$

$$MSR = \frac{SSR}{p-1}$$

$$= \frac{41.788}{3-1} = 20.894$$

$$MSE = \frac{SSE}{n-p}$$

$$= \frac{102.201}{10-3} = 14.600$$

ดังนั้น

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

$$= \frac{20.894}{14.600} = 1.431$$

ดังนั้นตารางวิเคราะห์ความแปรปรวนคือ

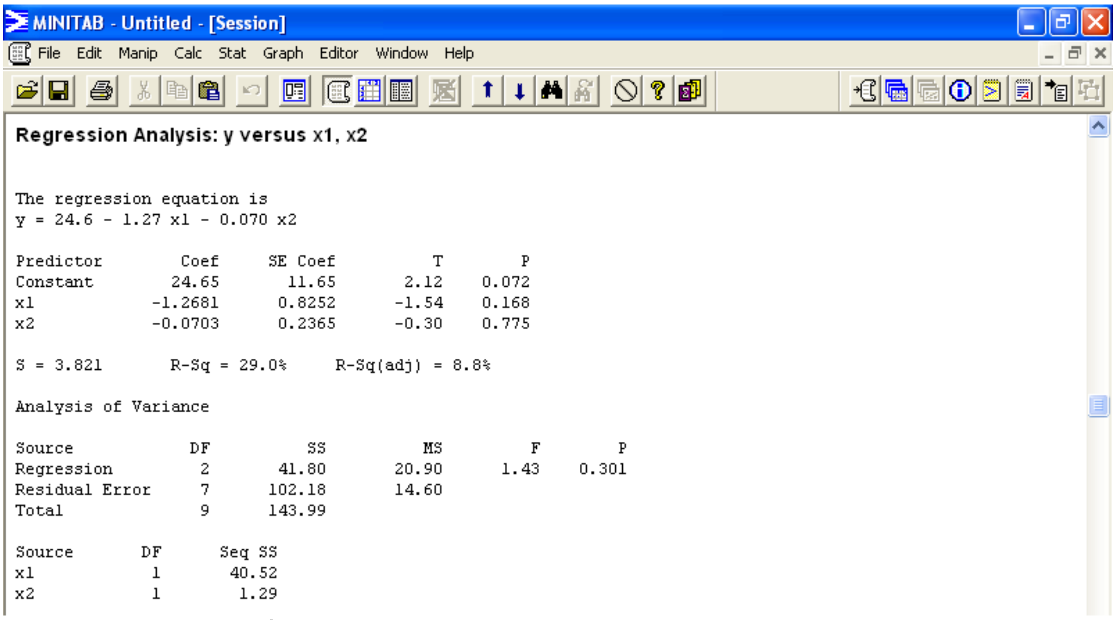
Source of variation	***SS***	***df***	***MS***	***F***
Regression	$SSR = 41.788$	2	$MSR = 20.894$	$F = 1.431$
Error	$SSE = 102.201$	7	$MSE = 14.600$	
Total	$SST = 143.989$	9		

เนื่องจากค่า $F = 1.431 < F_{.05,\,2,\,22} = 4.74$ ดังนั้นสรุปได้ว่าปริมาณตัวเร่งปฏิกิริยาทั้งสองชนิดไม่มีผลต่อเวลาที่ใช้ในการเกิดสารชนิดนี้ที่ระดับนัยสำคัญ 0.05

หมายเหตุ การใช้โปรแกรม MINITABช่วยในการคำนวณจะเหมือนกับการวิเคราะห์การถดถอยเชิงเส้นอย่างง่ายแต่เพิ่มตัวแปรอิสระเข้าไปตามที่ต้องการ

จากตัวอย่างที่ 5.4 หากใช้โปรแกรม MINITAB ช่วยในการคำนวณจะได้ผลลัพธ์ดังภาพที่ 5.1 โดยในหน้าต่าง “Session” แสดงสมการถดถอยในบรรทัดแรกและส่วนถัดมาแสดงค่า b_0, b_1 และ b_2 พร้อมทั้งการทดสอบและค่า p-value ส่วนถัดมาเป็นค่าส่วนเบี่ยงเบนมาตรฐาน (S)

สัมประสิทธิ์การตัดสินใจของตัวแปรอิสระหลายตัว (R-Sq) และ สัมประสิทธิ์การตัดสินใจที่ปรับแล้วของตัวแปรอิสระหลายตัว (R-Sq(adj)) จากนั้นเป็นตาราง ANOVA พร้อมทั้งคำนวณค่าผลรวมกำลังสองแบบต่อเนื่อง (sequential sum of squares)




```
Regression Analysis: y versus x1, x2

The regression equation is
y = 24.6 - 1.27 x1 - 0.070 x2

Predictor        Coef     SE Coef          T        P
Constant        24.65       11.65       2.12    0.072
x1            -1.2681      0.8252      -1.54    0.168
x2            -0.0703      0.2365      -0.30    0.775

S = 3.821       R-Sq = 29.0%      R-Sq(adj) = 8.8%

Analysis of Variance

Source          DF          SS          MS         F        P
Regression       2       41.80       20.90      1.43    0.301
Residual Error   7      102.18       14.60
Total            9      143.99

Source       DF      Seq SS
x1            1       40.52
x2            1        1.29
```

ภาพที่ 5.1 หน้าจอผลลัพธ์การวิเคราะห์การถดถอยเชิงเส้นพหุ

5.8 สัมประสิทธิ์การตัดสินใจสำหรับตัวแปรอิสระหลายตัว

ในสมการถดถอยเชิงเส้นอย่างง่ายนั้นค่าสัมประสิทธิ์การตัดสินใจ (R^2) มาจากตัวแปรอิสระเพียง 1 ตัวแต่ในสมการถดถอยเชิงเส้นพหุที่มีตัวแปรอิสระมากกว่า 1 ตัวแล้วค่าสัมประสิทธิ์การตัดสินใจจะถูกเรียกว่าค่าสัมประสิทธิ์การตัดสินใจสำหรับตัวแปรอิสระหลายตัว (coefficient of multiple determination) หรือ R^2 ซึ่งยังคงเป็นการวัดสัดส่วนของความแปรผันทั้งหมดของตัวแปรตามที่สามารถอธิบายได้โดยการใช้ตัวแปรอิสระในสมการถดถอยสามารถแสดงได้ดังสมการ

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \tag{5.31}$$

ค่า R^2 มีค่าตั้งแต่ 0 ถึง 1 หรือ $0 \le R^2 \le 1$ โดยที่กำหนดว่าค่า R^2 เท่ากับ 0 เมื่อค่าสัมประสิทธิ์การถดถอยทุกตัวมีค่าเป็น 0 และมีค่าเท่ากับ 1 เมื่อค่าของตัวแปรตามทุกค่ามีค่าเท่ากับค่าพยากรณ์หรือ $Y_i = \hat{Y}_i$ สำหรับข้อมูลทั้ง n ค่า

เนื่องจากค่า R^2 ได้มาจากการใช้ข้อมูลในการคำนวณดังนั้นช่วงพิสัยของตัวแปรอิสระจะมีผลต่อค่า R^2 การที่ค่า R^2 มีค่าสูงนั้นมิได้หมายความว่าสมการถดถอยนั้นเหมาะสมกับทุกช่วงค่า

ของตัวแปรอิสระ หากพยากรณ์นอกช่วงค่าของตัวแปรอิสระที่เก็บข้อมูลมา (extrapolation) แล้วสมการถดถอยอาจไม่เหมาะสมก็ได้นั่นคือค่าพยากรณ์ที่ได้อาจมีความคลาดเคลื่อนได้

การเพิ่มจำนวนตัวแปรอิสระเข้าไปในสมการอาจทำให้ค่า R^2 เพิ่มขึ้นเนื่องจากค่า SSE อาจมีค่าลดลงได้แต่ค่า SST มีค่าเท่าเดิมไม่เปลี่ยนแปลงโดยที่ตัวแปรอิสระที่เพิ่มเข้าไปนั้นอาจไม่มีความสัมพันธ์กับตัวแปรตามก็ได้ ดังนั้นจึงมีการปรับค่า R^2 โดยการหารค่า SSE และ SST ด้วยองศาเสรีของมันเรียกค่า R^2 นี้ว่า สัมประสิทธิ์การตัดสินใจที่ปรับแล้วสำหรับตัวแปรอิสระหลายตัว (adjusted coefficient of multiple determination) หรือ R^2_{adj} สามารถแสดงได้โดยสมการดังนี้

$$R^2_{adj} = 1 - \frac{\frac{SSE}{n-p}}{\frac{SST}{n-1}} = 1 - \left(\frac{n-1}{n-p}\right)\frac{SSE}{SST} \qquad (5.32)$$

ค่า R^2_{adj} จะไม่เพิ่มขึ้นหากตัวแปรอิสระที่เพิ่มเข้าไปไม่มีความสัมพันธ์ต่อตัวแปรตามแต่ส่วนใหญ่แล้วจะมีค่าลดลงเมื่อเพิ่มตัวแปรอิสระที่ไม่มีความสัมพันธ์ต่อตัวแปรตามเข้าไปในสมการเนื่องจาก SSE อาจลดลงมากกว่าค่าองศาเสรีของมัน

ตัวอย่างที่ 5.5 จากตัวอย่างที่ 5.1 จงคำนวณค่าสัมประสิทธิ์การตัดสินใจสำหรับตัวแปรอิสระหลายตัว และค่าสัมประสิทธิ์การตัดสินใจที่ปรับแล้วสำหรับตัวแปรอิสระหลายตัว

วิธีทำ

จากตัวอย่างที่ 5.4 ได้ค่า SST = 143.989 และ SSE = 102.201

ดังนั้น

$$\begin{aligned} R^2 &= 1 - \frac{SSE}{SST} \\ &= 1 - \frac{102.201}{143.989} \\ &= 0.2902 \end{aligned}$$

และ

$$\begin{aligned} R^2_{adj} &= 1 - \left(\frac{n-1}{n-p}\right)\frac{SSE}{SST} \\ &= 1 - \left(\frac{9}{7}\right)\frac{102.201}{143.989} \\ &= 0.0874 \end{aligned}$$

พบว่าค่าที่ได้ทั้งสองต่ำมากแสดงว่าตัวเร่งปฏิกิริยาทั้งสองมีความสัมพันธ์เชิงเส้นกับเวลาที่ใช้ในการเกิดสารนี้น้อยมากเนื่องจากความแปรผันที่เกิดขึ้นทั้งหมดสามารถใช้ตัวแปรทั้งสองในการอธิบายได้เพียง 8.74%

5.9 การอนุมานค่าพารามิเตอร์ของสมการถดถอยเชิงเส้นพหุ

การอนุมานค่าพารามิเตอร์สำหรับสมการถดถอยเชิงเส้นพหุสามารถแบ่งได้เป็นการสร้างช่วงความเชื่อมั่นของตัวประมาณค่าพารามิเตอร์และการทดสอบสมมติฐานของพารามิเตอร์แต่ละค่าดังนี้

5.9.1 ช่วงความเชื่อมั่นของตัวประมาณค่า β_j

ในการสร้างช่วงความเชื่อมั่นของตัวประมาณค่าพารามิเตอร์ β_j นั้นจะคล้ายกับการสร้างช่วงความเชื่อมั่นของสมการถดถอยเชิงเส้นอย่างง่ายโดยใช้สถิติทีที่องศาเสรี $n - p$ เนื่องจากค่าประมาณของพารามิเตอร์แต่ละตัวมีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเท่ากับ β_j และส่วนเบี่ยงเบนมาตรฐานเท่ากับ $s\{b_j\}$ ดังนั้นช่วงความเชื่อมั่น 100(1- α)% ของ β_j คือเท่ากับ

$$b_j - t_{\alpha/2,n-p}s\{b_j\} \le \beta_j \le b_j + t_{\alpha/2,n-p}s\{b_j\} \qquad \text{เมื่อ } j = 0,\ldots,p-1 \tag{5.33}$$

5.9.2 การทดสอบสมมติฐานของค่าพารามิเตอร์ β_j

การเพิ่มตัวแปรอิสระเข้าในสมการถดถอยจะทำให้ SSE ลดลงหรือเพิ่ม SSR ขึ้น แต่การเพิ่มขึ้นของ SSR นั้นอาจไม่ได้หมายความว่าตัวแปรอิสระนั้นๆ จำเป็นในการพยากรณ์ตัวแปรตาม การเพิ่มตัวแปรอิสระขึ้นทำให้ความแปรปรวนของค่าพยากรณ์เพิ่มขึ้นด้วยดังนั้นจึงต้องมีการพิจารณาตัวแปรอิสระแต่ละตัวว่าตัวใดมีความสำคัญกับตัวแปรตาม (Montgomery & Peck, 1992, p. 138) ในการทดสอบสมมติฐานของค่าพารามิเตอร์แต่ละค่านั้นจะคล้ายกับการทดสอบในสมการถดถอยเชิงเส้นอย่างง่ายโดยสมมติฐานมีดังนี้

$$H_0 : \beta_j = \beta_{j0}$$

$$H_1 : \beta_j \neq \beta_{j0}$$

สถิติที่ใช้คือ

$$t = \frac{(b_j - \beta_{j0})}{s\{b_j\}} \tag{5.34}$$

โดย หาก $|t| \le t_{\alpha/2,n-p}$ แล้วไม่ปฏิเสธสมมติฐานหลัก (H_0)

หาก $|t| > t_{\alpha/2,n-p}$ แล้วปฏิเสธสมมติฐานหลัก (H_0)

หมายเหตุ

เช่นเดียวกับสมการถดถอยเชิงเส้นอย่างง่าย หากกำหนด $\beta_{j0} = 0$ แล้วการทดสอบสมมติฐานนั้นจะทดสอบว่าตัวแปรอิสระนั้นมีความสัมพันธ์เชิงเส้นกับตัวแปรตามหากไม่ปฏิเสธสมมติฐานหลักแล้วสามารถที่จะละตัวแปรอิสระนั้นออกจากตัวแบบได้

ตัวอย่างที่ 5.6 จากตัวอย่างที่ 5.1 จงสร้างช่วงความเชื่อมั่นของ β_1 และ β_2 พร้อมทั้งทดสอบว่าตัวเร่งปฏิกิริยาชนิดที่ 1 มีความสัมพันธ์เชิงเส้นกับเวลาที่ใช้ในการเกิดสารหรือไม่ที่ระดับนัยสำคัญ 0.05

วิธีทำ

(1) เนื่องจาก $\boldsymbol{S}^2\{b\} = MSE(\boldsymbol{X'X})^{-1}$

$$= 14.6\begin{bmatrix} 9.30375 & -0.0580211 & -0.176798 \\ -0.0580211 & 0.0466504 & -0.00338089 \\ -0.176798 & -0.00338089 & 0.00383279 \end{bmatrix}$$

$$= \begin{bmatrix} 135.83475 & -0.8471081 & -2.5812508 \\ -0.8471081 & 0.6810958 & -0.04936099 \\ -2.5812508 & -0.04936099 & 0.05595873 \end{bmatrix}$$

นั่นคือ

$$s^2\{b_1\} = 0.6810958 \quad \text{หรือ } s\{b_1\} = 0.825285$$

และ
$$s^2\{b_2\} = 0.05595873 \quad \text{หรือ } s\{b_2\} = 0.236556$$

ดังนั้นช่วงความเชื่อมั่นที่ 95% ของ β_1 คือ

$$b_1 - t_{0.025,7}s\{b_1\} \le \beta_1 \le b_1 - t_{0.025,7}s\{b_1\}$$

$$-1.268 - 2.365 \times 0.825285 \le \beta_1 \le -1.268 + 2.365 \times 0.825285$$

$$-3.2198 \le \beta_1 \le 0.6838$$

และช่วงความเชื่อมั่นที่ 95% ของ β_2 คือ

$$b_2 - t_{0.025,7}s\{b_2\} \le \beta_2 \le b_2 - t_{0.025,7}s\{b_2\}$$

$$-0.070 - 2.365 \times 0.236556 \le \beta_2 \le -0.070 + 2.365 \times 0.236556$$

$$-1.25945 \le \beta_2 \le -0.14055$$

(2) เนื่องจากต้องการทดสอบว่าตัวเร่งปฏิกิริยาชนิดที่ 1 มีความสัมพันธ์เชิงเส้นกับเวลาที่ใช้ในการเกิดสารหรือไม่ดังนั้นสมมติฐานจึงเป็น

$$H_0 : \beta_1 = 0$$

$$H_1 : \beta_1 \neq 0$$

และค่าสถิติคือ

$$t = \frac{(b_1 - 0)}{s\{b_1\}}$$
$$= \frac{-1.268}{0.825285}$$
$$= -1.5364$$

พบว่าค่าสถิติ t = -1.5364 น้อยกว่าค่า $t_{0.025,\ 7} = 2.365$ ดังนั้นจึงไม่ปฏิเสธสมมติฐานหลัก และสรุปว่าตัวเร่งปฏิกิริยาชนิดที่ 1 ไม่มีความสัมพันธ์เชิงเส้นกับเวลาที่ใช้ในการเกิดสารชนิดนี้ ที่ระดับนัยสำคัญ 0.05

5.10 การประมาณค่าแบบช่วงของค่าเฉลี่ยของตัวแปรตาม

การประมาณค่าพยากรณ์ของค่าคาดหวังหรือค่าเฉลี่ยของ Y หรือ $E(Y|X_0)$ เมื่อ $X_1 = X_{01}$, $X_2 = X_{02}, \ldots, X_{p-1} = X_{0p-1}$ หรือสามารถเขียนเป็นเวคเตอร์ของ X_0 ได้ดังนี้

$$\boldsymbol{X}_0 = \begin{bmatrix} 1 \\ X_{01} \\ X_{02} \\ \vdots \\ X_{0p-1} \end{bmatrix}$$

ค่าพยากรณ์ ณ จุดนี้สามารถเขียนให้อยู่ในรูปเมตริกซ์คือ

$$\hat{Y}_0 = \boldsymbol{X}_0'\boldsymbol{b} \qquad (5.35)$$

เช่นเดียวกับการวิเคราะห์การถดถอยเชิงเส้นอย่างง่ายค่า $\hat{Y}_0$ เป็นตัวประมาณค่าที่ไม่เอนเอียงและมีความแปรปรวนเท่ากับ $\sigma^2 \boldsymbol{X}_0'(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{X}_0$ หรือสามารถเขียนได้ในรูป $\hat{Y}_0 \sim N\left(\boldsymbol{X}'_0\boldsymbol{b}, \sigma^2 \boldsymbol{X}_0'(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{X}_0\right)$

ช่วงความเชื่อมั่น 100(1- α)% ของค่าคาดหวังของ Y หรือ Y_0 คือ

$$\hat{Y}_0 - t_{\alpha/2,n-p}\sqrt{MSE\boldsymbol{X}_0'(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{X}_0} \le Y_0 \le \hat{Y}_0 + t_{\alpha/2,n-p}\sqrt{MSE\boldsymbol{X}_0'(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{X}_0} \qquad (5.36)$$

ช่วงความเชื่อมั่น 100(1- α)% ของค่า Y ใหม่ ณ x_0 ที่ไม่ได้อยู่ในข้อมูลจำนวน 1 ค่าหรือ Y_0 คือ

$$\hat{Y}_0 - t_{\alpha/2,n-p}\sqrt{MSE(1+\boldsymbol{X}_0'(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{X}_0)} \le Y_0 \le \hat{Y}_0 + t_{\alpha/2,n-p}\sqrt{MSE(1+\boldsymbol{X}_0'(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{X}_0)} \qquad (5.37)$$

ช่วงความเชื่อมั่น 100(1- α)% ของค่า Y ใหม่ ณ x_0 ที่ไม่ได้อยู่ในข้อมูลจำนวน m ค่าหรือ Y_0 คือ

$$\hat{y}_0 - t_{\alpha/2,n-p}\sqrt{MSE(\frac{1}{m} + \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0)} \le y_0 \le \hat{y}_0 + t_{\alpha/2,n-p}\sqrt{MSE(\frac{1}{m} + \boldsymbol{X}_0'(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{X}_0)} \quad (5.38)$$

ตัวอย่าง 5.7 จากข้อมูลในตัวอย่าง 5.1 จงสร้างช่วงความเชื่อมั่นของเวลาที่ใช้ในการเกิดสารใหม่ที่ระดับความเชื่อมั่น 95% ของค่าพยากรณ์ที่ใช้ตัวเร่งปฏิกิริยาชนิดที่ 1 และ 2 เท่ากับ 5 และ 45 ตามลำดับโดยพยากรณ์เพียงค่าเดียว

วิธีทำ

(1) เมื่อ $x_{01} = 5$ และ $x_{02} = 45$ แล้ว $\hat{Y}_0 = 24.649 - 1.268\times 5 - 0.070\times 45 = 15.16$

ที่ระดับความเชื่อมั่น 95% จะมีค่า $t_{\alpha/2,\,n-3} = t_{0.025,\;7} = 2.365$ (t จากตาราง) และ

$$\boldsymbol{X}_0'(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{X}_0 = [1 \quad 5 \quad 45]\begin{bmatrix} 9.30375 & -0.0580211 & -0.176798 \\ -0.0580211 & 0.0466504 & -0.00338089 \\ -0.176798 & -0.00338089 & 0.00383279 \end{bmatrix}\begin{bmatrix} 1 \\ 5 \\ 45 \end{bmatrix}$$

$$= 0.21799$$

และช่วงความเชื่อมั่นของค่าพยากรณ์ของเวลาที่ใช้ในการเกิดสารใหม่คือ

$$\hat{Y}_0 - t_{\alpha/2,n-p}\sqrt{MSE(1 + \boldsymbol{X}_0'(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{X}_0)} \le Y_0 \le \hat{Y}_0 + t_{\alpha/2,n-p}\sqrt{MSE(1 + \boldsymbol{X}_0'(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{X}_0)}$$

$$15.16 - 2.365\sqrt{14.6(1 + 0.21799)} \le Y_0 \le 15.16 + 2.365\sqrt{14.6(1 + 0.21799)}$$

$$5.1869 \le Y_0 \le 25.1330$$

ดังนั้นเชื่อมั่นได้ 95%ว่าเวลาที่ใช้ในการเกิดสารใหม่อยู่ระหว่าง 5.1869 กับ 25.1330 นาทีหากใช้ตัวเร่งปฏิกิริยาชนิดที่ 1 และ 2 เท่ากับ 5 และ 45 ตามลำดับ

5.11 การแบ่งส่วนของตัวแบบ

การเลือกตัวแปรอิสระว่าตัวใดมีผลต่อตัวแปรตามบางครั้งอาจสามารถเลือกทดสอบตัวแปรอิสระทีละตัวโดยทดสอบด้วยสถิติ t หรือสามารถเลือกตัวแปรอิสระเป็นกลุ่มโดยการใช้การแบ่งส่วนของตัวแบบโดยทดสอบด้วยสถิติ F และสัมประสิทธิ์การตัดสินใจบางส่วน

5.11.1 การทดสอบเอฟบางส่วน

บางครั้งนักวิจัยอาจมีความต้องการที่จะทดสอบตัวแปรอิสระเป็นกลุ่มหรือทดสอบพารามิเตอร์เป็นกลุ่มจึงต้องมีการใช้การทดสอบเอฟบางส่วน (partial F-test) พิจารณาตัวแบบถดถอยที่มีตัวแปรอิสระ k ตัวโดยเขียนเป็นสมการได้คือ

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

โดย $\mathbf{Y}$ = เวคเตอร์ของค่าของตัวแปรตามที่มีขนาด $n \times 1$

$\mathbf{X}$ = เมตริกซ์ของค่าของตัวแปรอิสระ k ตัวที่มีขนาด $n \times p$

$\boldsymbol{\beta}$ = เวคเตอร์ของค่าพารามิเตอร์ของสมการถดถอยที่มีขนาด $p \times 1$

$\boldsymbol{\varepsilon}$ = เวคเตอร์ของค่าความคลาดเคลื่อนที่มีขนาด $n \times 1$

$p = k + 1$

หากแบ่ง $\boldsymbol{\beta}$ ออกเป็นสองส่วนคือ $\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix}$ โดย $\boldsymbol{\beta}_1$ มีขนาด $(p - r) \times 1$ และ $\boldsymbol{\beta}_2$ มีขนาด $r \times 1$ ดังนั้นสมการถดถอยสามารถเขียนได้เป็น

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon} \tag{5.39}$$

โดย $\mathbf{X}_1$ = เมตริกซ์ค่าของตัวแปรอิสระที่สอดคล้องกับ $\boldsymbol{\beta}_1$ ที่มีขนาด $n \times (p - r)$

$\mathbf{X}_2$ = เมตริกซ์ค่าของตัวแปรอิสระที่สอดคล้องกับ $\boldsymbol{\beta}_2$ ที่มีขนาด $n \times r$

สมการถดถอยที่ (5.39) เรียกสมการเต็มรูปแบบ (full model) และมี $\mathbf{b} = (\mathbf{X'X})^{-1}\mathbf{X'Y}$ และมีผลรวมกำลังสองถดถอย (*SSR*) คือ $SSR = \mathbf{b'X'Y}$ โดยมีองศาเสรีเท่ากับ p และ $MSE = \dfrac{\mathbf{Y'Y} - \mathbf{b'X'Y}}{n-p}$ หากต้องการทดสอบสมมติฐาน

$$H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$$

$$H_1 : \boldsymbol{\beta}_2 \neq \mathbf{0}$$

หากสมมติฐานหลักเป็นจริง ($\boldsymbol{\beta}_2 = \mathbf{0}$) แล้วสมการถดถอยจะสามารถเขียนได้อยู่ในรูป

$$\mathbf{Y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon} \tag{5.40}$$

สมการที่ (5.40) เรียก สมการลดรูป (reduced model) ดังนั้นเวคเตอร์ตัวประมาณค่าของ $\boldsymbol{\beta}_1$ ในสมการถดถอยลดรูปคือ

$$\mathbf{b_1} = (\mathbf{X_1'X_1})^{-1}\mathbf{X_1'Y} \tag{5.41}$$

และผลรวมกำลังสองถดถอยของสมการลดรูปหรือ $SSR(\boldsymbol{\beta}_1)$ เท่ากับ

$$SSR(\boldsymbol{\beta}_1) = \mathbf{b_1'X_1'Y} \tag{5.42}$$

โดยมีองศาเสรีเท่ากับ $p - r$ และผลรวมกำลังสองถดถอยเนื่องมาจาก $\boldsymbol{\beta}_2$ โดยกำหนดว่า $\boldsymbol{\beta}_1$ อยู่ในสมการแล้วหรือ $SSR(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1)$ เท่ากับ

$$SSR(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1) = SSR - SSR(\boldsymbol{\beta}_1) \tag{5.43}$$

ผลรวมกำลังสองถดถอยที่ได้ในสมการ (5.43) เรียกว่าผลรวมกำลังสองที่เพิ่มขึ้น (extra sum of squares) เนื่องมาจาก $\boldsymbol{\beta}_2$ ทั้งนี้เพราะค่านี้เป็นการวัดการเพิ่มขึ้นของ SSR ที่เป็นผลมาจากการเพิ่มตัวแปรอิสระ $X_{k-r+1}, X_{k-r+2},\ldots, X_k$ ของ $\mathbf{X}_2$ เข้าในตัวแบบที่มี $X_1, X_2,\ldots, X_{k-r}$ อยู่ในตัวแบบก่อนแล้ว $SSR(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1)$ ที่ได้จะเป็นอิสระต่อ MSE และการทดสอบสมมติฐาน $\boldsymbol{\beta}_2 = \mathbf{0}$ ทำได้โดยใช้สถิติ F และเรียกการทดสอบนี้ว่าการทดสอบเอฟบางส่วน (partial F-test) โดยมีสูตรดังนี้

$$F = \frac{SSR(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1)/r}{MSE} \tag{5.44}$$

การทดสอบสมมติฐานทำโดยเปรียบเทียบกับค่า F จากตารางสถิติที่องศาเสรีเท่ากับ r และ $n-p$ หากค่า F ที่คำนวณได้มีค่าน้อยกว่าหรือเท่ากับค่า F จากตารางแล้วสมมติฐานหลักจะเป็นจริงแสดงว่าพารามิเตอร์ $\boldsymbol{\beta}_2$ มีค่าเป็น 0 แต่หากปฏิเสธสมมติฐานหลักแสดงว่ามีพารามิเตอร์อย่างน้อยหนึ่งตัวใน $\boldsymbol{\beta}_2$ ที่มีค่าไม่เป็น 0 และมีตัวแปรอิสระอย่างน้อยหนึ่งตัวใน $X_{k-r+1}, X_{k-r+2},\ldots, X_k$ ที่จำเป็นในตัวแบบถดถอยนั้น หากตัวแบบคือ $Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \varepsilon$ และแยกผลรวมกำลังสองถดถอยได้เป็น $SSR(\beta_1|\beta_0, \beta_2, \beta_3)$ $SSR(\beta_2|\beta_0, \beta_1, \beta_2)$ และ $SSR(\beta_3|\beta_0, \beta_1, \beta_2)$ โดยผลรวมทั้งสามค่านี้เป็นการวัดความสำคัญของตัวแปรอิสระแต่ละตัวที่มีต่อตัวแปรตามโดยกำหนดว่ามีตัวแปรอิสระที่เหลืออยู่ในตัวแบบแล้ว หากทำการทดสอบเอฟบางส่วนกับตัวแปรอิสระทีละตัวจะมีผลเท่ากับการทดสอบโดยใช้สถิติ t นอกจากการใช้การทดสอบเอฟบางส่วนกับตัวแปรอิสระแต่ละตัวแล้วยังสามารถทดสอบเป็นกลุ่มได้อีกด้วยเพื่อหาว่ากลุ่มของตัวแปรใดให้ตัวแบบที่เหมาะสมที่สุดสำหรับตัวแบบนั้นๆ นอกจากการพิจารณาการเพิ่มตัวแปรอิสระเข้ามาทีละตัวโดยที่มีตัวแปรอื่นๆทั้งหมดอยู่ในตัวแบบแล้วยังสามารถพิจารณาโดยการเพิ่มเข้ามาทีละตัวได้เช่นกันดังสมการข้างล่าง

$$SSR(\beta_1, \beta_2, \beta_3| \beta_0) = SSR(\beta_1|\beta_0) + SSR(\beta_2|\beta_0, \beta_1) + SSR(\beta_3|\beta_0, \beta_1, \beta_2)$$

จะเห็นว่าผลรวมกำลังสองถดถอยข้างซ้ายของสมการข้างบนมีองศาเสรีเท่ากับ 3 และผลรวมกำลังสองถดถอยในด้านขวาของสมการแต่ละตัวมีองศาเสรีเท่ากับ 1

ตัวอย่างที่ 5.8 กำหนดข้อมูลดังนี้

Y	3.30	8.30	10.50	15.05	20.84	4.75	9.05	11.95	15.84	20.75
X_1	1	1	1	1	1	2	2	2	2	2
X_2	1	2	3	4	5	1	2	3	4	5

Y	9.03	10.35	15.94	16.85	20.45	10.55	11.94	15.50	18.64	21.50
X_1	3	3	3	3	3	4	4	4	4	4
X_2	1	2	3	4	5	1	2	3	4	5

Y	11.86	14.01	16.89	20.55	21.09
X_1	5	5	5	5	5
X_2	1	2	3	4	5

ต้องการทดสอบว่าตัวแปร X_2 มีผลต่อตัวแปร Y หรือไม่ที่ระดับนัยสำคัญ 0.05

วิธีทำ

เนื่องจากต้องการทดสอบว่าตัวแปร X_2 มีผลต่อตัวแปร Y หรือไม่ดังนั้นสมมติฐานคือ

$$H_0 : \beta_2 = 0$$

$$H_1 : \beta_2 \neq 0$$

เพื่อที่จะทดสอบสมมติฐานนี้จำเป็นต้องคำนวณหาผลรวมกำลังสองถดถอยที่มาจาก β_2 หรือ

$$SSR(\beta_2 | \beta_0, \beta_1) = SSR(\beta_0, \beta_1, \beta_2) - SSR(\beta_0, \beta_1)$$

$$= SSR(\beta_1, \beta_2 | \beta_0) - SSR(\beta_1 | \beta_0)$$

เนื่องจาก

$$SSR(\beta_1, \beta_2 | \beta_0) = \mathbf{b'X'Y} - \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n}$$

$$= 5683.250 - 5054.641 = 629.184 \quad \text{(องศาเสรีเท่ากับ 2)}$$

ตัวแบบลดรูปคือ $Y = \beta_0 + \beta_1 X_1 + \varepsilon$ ซึ่งมีค่าผลรวมกำลังสองถดถอยเท่ากับ

$$SSR(\beta_1 | \beta_0) = 94.150 \quad \text{(องศาเสรีเท่ากับ 1)}$$

ดังนั้น $SSR(\beta_2 | \beta_0, \beta_1) = 629.184 - 94.150 = 535.034$ (องศาเสรีเท่ากับ 1)

ค่า $SSR(\beta_2|\ \beta_0,\ \ \beta_1)$ = 535.034 เป็นค่าของผลรวมกำลังสองถดถอยที่เพิ่มขึ้นเมื่อเพิ่มตัวแปร X_2 เข้าไปในตัวแบบโดยที่มีตัวแปร X_1 อยู่แล้วดังนั้นการทดสอบสมมติฐานทำได้โดย

$$F = \frac{SSR(\beta_2|\beta_0, \beta_1)/1}{MSE} = \frac{535.034/1}{1.53} = 349.695$$

โดยที่ MSE ได้จากตัวแบบที่มีตัวแปรอิสระทั้งสองอยู่หรือตัวแบบเต็มรูปแบบ เนื่องจากค่า $F = 349.695 > F_{0.05,1,22} = 4.30$ ดังนั้นจึงปฏิเสธสมมติฐานหลักและสรุปว่าตัวแปร X_2 มีผลต่อตัวแบบที่ระดับนัยสำคัญ 0.05

หมายเหตุ

หากใช้โปรแกรม MINITAB ช่วยในการคำนวณสมการถดถอยพหุจะได้ผลลัพธ์ดังภาพข้างล่าง โดยที่สองบรรทัดสุดท้ายในผลลัพธ์เป็นค่าผลรวมกำลังสองถดถอย $SSR(\beta_1|\ \beta_0, \beta_2)$ และ $SSR(\beta_2|\ \beta_0, \beta_1)$ จาก (Seq SS) เท่ากับ 94.15 และ 535.04 ตามลำดับ

```
The regression equation is
y = 0.289 + 1.37 x1 + 3.27 x2

Predictor        Coef     SE Coef        T       P
Constant       0.2890      0.7832     0.37   0.716
x1             1.3722      0.1751     7.84   0.000
x2             3.2712      0.1751    18.68   0.000

S = 1.238       R-Sq = 94.9%     R-Sq(adj) = 94.4%

Analysis of Variance

Source            DF        SS        MS        F       P
Regression         2    629.18    314.59   205.14   0.000
Residual Error    22     33.74      1.53
Total             24    662.92

Source       DF      Seq SS
x1            1       94.15
x2            1      535.04
```

5.11.2 สัมประสิทธิ์การตัดสินใจบางส่วน

สัมประสิทธิ์การตัดสินใจบางส่วน (partial coefficient of determination) ใช้ช่วยในการพิจารณาว่าตัวแปรที่เพิ่มเข้าไปในตัวแบบนั้นช่วยลดความแปรผันที่อธิบายไม่ได้ (SSE) ลงเท่าใดโดยที่ในตัวแบบที่มีตัวแปรอิสระที่เหลืออยู่แล้วโดยมีสูตรดังนี้

$$R^2(X_{-r+1},\ldots,X_k|\ X_1,\ldots,X_{k-r}) = \frac{SSE_R - SSE_F}{SSE_R} \qquad (5.45)$$

โดย SSE_R = SSE ของสมการลดรูป

SSE_F = SSE ของสมการเต็มรูป

ตัวอย่างที่ 5.9 จากตัวอย่างที่ 5.8 จงคำนวณหาค่าสัมประสิทธิ์การตัดสินใจบางส่วน

วิธีทำ

เนื่องจาก SSE_R = 568.78 และ SSE_F = 33.74 ดังนั้น

$$R^2(x_2|\, x_1) = \frac{568.78 - 33.74}{568.78}$$

$$= 0.9407$$

นั่นคือการเพิ่มตัวแปร X_2 เข้าไปในตัวแบบเมื่อมีตัวแปร X_1 อยู่ในตัวแบบแล้วทำให้ค่าความแปรผันที่ไม่สามารถอธิบายได้ (*SSE*) ลดลงไป 94.07%

ตัวอย่างที่ 5.10 จากข้อมูลของงานวิจัยเรื่อง “ปัจจัยที่มีผลต่อการนำระบบความปลอดภัยของอาหารมาใช้ในอุตสาหกรรมแปรรูปอาหารขนาดกลางและขนาดย่อมในประเทศไทย” ของพรเลิศ อาภานุทัตและพรสิน สุภวาลย์ (2555) มีตัวแปรอิสระจำนวน 7 ตัว คือ ความคาดหวังถึงการยอมรับจากสังคมที่คาดว่าจะได้รับ (X_1) ความคาดหวังถึงความได้เปรียบทางการแข่งขันที่จะคาดว่าจะได้รับ (X_2) ระดับการรับรู้ถึงความสำคัญของผู้มีส่วนได้ส่วนเสียภายใน (X_3) ระดับการรับรู้ถึงความสำคัญของผู้มีส่วนได้ส่วนเสียภายนอก (X_4) ระดับความมุ่งมั่นของผู้บริหาร (X_5) การนำระบบการจัดการความปลอดภัยของอาหารมาใช้ของกิจการอื่น (X_6) และระดับการมีปฏิสัมพันธ์กับองค์การภายนอกที่เกี่ยวข้องกับการจัดการความปลอดภัยของอาหาร(X_7) โดยมีตัวแปรตามคือ ระดับการนำระบบการจัดการความปลอดภัยของอาหารมาใช้ (*Y*) นำมาสร้างสมการถดถอยพหุโดย MINITAB และคำนวณค่าสัมประสิทธิ์การตัดสินใจที่ปรับแล้ว

วิธีทำ

จากการใช้โปรแกรม MINITAB ในการสร้างสมการถดถอยพหุได้ผลดังนี้คือ

```
The regression equation is
y = - 0.797 + 0.164 x1 + 0.0795 x2 - 0.0208 x3 + 0.0569 x4 + 0.595 x5 + 0.127 x6
    + 0.179 x7

Predictor      Coef   SE Coef      T      P
Constant    -0.7965    0.3567  -2.23  0.027
x1          0.16387   0.07000   2.34  0.020
x2          0.07954   0.08902   0.89  0.373
x3         -0.02077   0.04914  -0.42  0.673
x4          0.05691   0.06602   0.86  0.390
x5          0.59461   0.07990   7.44  0.000
x6          0.12693   0.05641   2.25  0.026
x7          0.17879   0.06110   2.93  0.004

S = 0.496254   R-Sq = 52.9%   R-Sq(adj) = 51.1%

Analysis of Variance

Source           DF       SS      MS      F      P
Regression        7  51.7358  7.3908  30.01  0.000
Residual Error  187  46.0521  0.2463
Total           194  97.7879

Source  DF   Seq SS
x1       1  21.3330
x2       1   4.4185
x3       1   0.0609
x4       1   1.6753
x5       1  20.4838
x6       1   1.6552
x7       1   2.1090
```

จากผลลัพธ์ที่ได้ข้างต้นพบว่า ความคาดหวังถึงการยอมรับจากสังคมที่คาดว่าจะได้รับ (X_1) ระดับความมุ่งมั่นของผู้บริหาร (X_5) การนำระบบการจัดการความปลอดภัยของอาหารมาใช้ของกิจการอื่น (X_6) และระดับการมีปฏิสัมพันธ์กับองค์การภายนอกที่เกี่ยวข้องกับการจัดการความปลอดภัยของอาหาร(X_7) สามารถใช้ในการพยากรณ์ระดับการนำระบบการจัดการความปลอดภัยของอาหารมาใช้ (Y) ได้ที่ระดับนัยสำคัญทางสถิติที่ 0.05 โดยมีค่าสัมประสิทธิ์การตัดสินใจที่ปรับแล้ว R_{adj}^2 เท่ากับ 51.1%

5.12 ขนาดตัวอย่าง

ขนาดตัวอย่างในการวิเคราะห์การถดถอยพหุจะมีผลต่ออำนาจการทดสอบและการนำผลที่ได้ไปอ้างอิงกับประชากรกลุ่มอื่น (Generalization) โดย Hair et al (2010, p. 174-176) กล่าวว่าเมื่อขนาดตัวอย่างมากเกิน 1,000 ข้อมูลแล้วจะทำให้มีโอกาสในการปฏิเสธสมมติฐานหลักของการทดสอบความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตามมากขึ้นถึงแม้ตัวแปรอิสระนั้นจะไม่มีความสัมพันธ์กับตัวแปรตามก็ตาม สัดส่วนของขนาดตัวอย่างต่อจำนวนตัวแปรอิสระที่น้อยที่สุดเท่ากับ 5 ข้อมูลต่อตัวแปรอิสระ 1 ตัวหากสัดส่วนต่ำกว่านี้จะทำให้เกิดภาวะที่เรียกว่า overfitting หรือไม่สามารถนำไปอ้างอิงใช้กับประชากรกลุ่มอื่นได้ ทั้งนี้สัดส่วนของขนาดตัวอย่างต่อจำนวนตัวแปรอิสระที่เหมาะสมในการนำผลไปอ้างอิงกับประชากรกลุ่มอื่นนั้นเท่ากับ 15 – 20 ข้อมูลต่อตัวแปรอิสระ 1 ตัว แต่หากนักวิจัยใช้การเลือกตัวแบบโดยวิธีการเพิ่มตัวแปรอิสระแบบขั้นตอน (stepwise regression) แล้วสัดส่วนของขนาดตัวอย่างต่อจำนวนตัวแปรอิสระควรเป็น 50 ข้อมูลต่อตัวแปรอิสระ 1 ตัวเนื่องจากเทคนิคนี้จะเลือกตัวแปรอิสระจากความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตามที่มากที่สุด หากตัวแปรตามมีการแจกแจงที่เบ้แล้วสัดส่วนของขนาดตัวอย่างต่อจำนวนตัวแปรต้องสูงขึ้น (Tabachnick & Fidel, 2007, p. 124)

สรุป

การวิเคราะห์การถดถอยพหุมีหลักการคล้ายกับการวิเคราะห์การถดถอยเชิงเส้นอย่างง่ายแต่มีจำนวนตัวแปรอิสระมากกว่า 1 ตัวทำให้การคำนวณยุ่งยากและซับซ้อนมากขึ้น เมตริกซ์ช่วยให้การวิเคราะห์ทำได้ง่ายและรวดเร็วยิ่งขึ้น

คำถามท้ายบท

5.1 จากข้อมูลข้างล่างจงหา

Y	400	620	640	550	500	560	520	630	650	590
X_1	120	114	132	116	121	126	124	114	125	113
X_2	3.1	4.2	3.1	3.9	2.7	3.3	4.4	4.8	3.0	4.3

(1) **X'X**

(2) **X'Y**

(3) **b**

(4) สมการถดถอย

5.2 จงอธิบายผลลัพธ์ข้างล่างที่ได้จากการใช้โปรแกรม MINITAB

Regression Analysis: Y versus X1, X2, X3, X4, X5, X6

```
The regression equation is
Y = 2.84 + 0.0039 X1 + 0.0071 X2 + 0.198 X3 + 0.044 X4 - 0.133 X5 + 1.08 X6

Predictor        Coef     SE Coef         T        P
Constant        2.838       6.551      0.43    0.672
X1            0.00390     0.08181      0.05    0.963
X2            0.00712     0.03188      0.22    0.827
X3             0.1978      0.1716      1.15    0.270
X4             0.0443      0.1440      0.31    0.763
X5            -0.1330      0.2490     -0.53    0.602
X6             1.0751      0.2995      3.59    0.003

S = 3.960       R-Sq = 78.8%     R-Sq(adj) = 69.0%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         6      758.43      126.41      8.06    0.001
Residual Error    13      203.87       15.68
Total             19      962.31

Source       DF      Seq SS
X1            1       30.83
X2            1       83.93
X3            1      368.47
X4            1       72.81
X5            1        0.26
X6            1      202.13
```

5.3 จากข้อมูลในข้อ 5.1 จงสร้างตารางวิเคราะห์ความแปรปรวนโดยการใช้เมตริกซ์ พร้อมทั้งทดสอบตัวแบบและตัวแปรอิสระทั้งสองที่ระดับนัยสำคัญ 0.05

5.4 จากข้อมูลในข้อ 5.1 จงพยากรณ์ค่า Y ที่ $X_0' = [1 \;\; 100 \;\; 2.0]$ พร้อมทั้งสร้างช่วงความเชื่อมั่น 90% ของค่าพยากรณ์นี้

5.5 จงอธิบายผลลัพธ์ข้างล่างที่ได้จากการใช้โปรแกรม MINITAB ในการวิเคราะห์สมการถดถอยพหุของปริมาณคาร์บอนไดออกไซด์โดยมีปัจจัยที่สนใจคือจำนวนรถบนถนนและความเร็วลม

Regression Analysis: CO versus Traffic, Wind

```
The regression equation is
CO = 1.27 + 0.0183 Traffic + 0.175 Wind

Predictor        Coef     SE Coef          T        P
Constant       1.2745      0.1981       6.43    0.000
Traffic      0.018290    0.001343      13.62    0.000
Wind          0.17475     0.05677       3.08    0.006

S = 0.4987      R-Sq = 95.0%     R-Sq(adj) = 94.5%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         2      98.233      49.116    197.46    0.000
Residual Error    21       5.224       0.249
Total             23     103.456

Source       DF      Seq SS
Traffic       1      95.875
Wind          1       2.357

Unusual Observations
Obs    Traffic         CO          Fit       SE Fit     Residual     St Resid
 11        200      6.600        5.334       0.112        1.266        2.60R
```

5.6 นักชีววิทยาต้องการพยากรณ์ปริมาตรของต้นเชอรี่ (ลบ.ฟุต) โดยใช้ความสูง (ฟุต) และเส้นผ่าศูนย์กลาง (นิ้ว) โดยมีข้อมูลดังนี้

เส้นผ่าศูนย์กลาง (X_1)	8.3	8.6	8.8	10.5	10.7	10.8	11.0	11.0
ความสูง (X_2)	70.0	65.0	63.0	72.0	81.0	83.0	66.0	75.0
ปริมาตร (Y)	10.3	10.3	10.2	16.4	18.8	19.7	15.6	18.2

เส้นผ่าศูนย์กลาง	11.1	11.2	11.3	11.4	11.4	11.7	12.0	12.9
ความสูง	80.0	75.0	79.0	76.0	76.0	69.0	75.0	74.0
ปริมาตร	22.6	19.9	24.2	21.0	21.4	21.3	19.1	22.2

เส้นผ่าศูนย์กลาง	12.9	13.3	13.7	13.8	14.0	14.2	14.5	16.0
ความสูง	85.0	86.0	71.0	64.0	78.0	80.0	74.0	72.0
ปริมาตร	33.8	27.4	25.7	24.9	34.5	31.7	36.3	38.3

เส้นผ่าศูนย์กลาง	16.3	17.3	17.5	17.9	18.0	18	20.6
ความสูง	77.0	81.0	82.0	80.0	80.0	80	87.0
ปริมาตร	42.6	55.4	55.7	58.3	51.5	51	77.0

ที่มา: Ryan, Joiner & Ryan, 1985, p. 328-329

จากข้อมูลดังกล่าวจงสร้างสมการถดถอยพหุพร้อมทั้งคำนวณค่าพยากรณ์และค่าคลาดเคลื่อน

5.7 จากข้อมูลในข้อ 5.6 จงสร้างตารางวิเคราะห์ความแปรปรวนโดยการใช้เมตริกซ์ พร้อมทั้งทดสอบตัวแบบและตัวแปรอิสระทั้งสองที่ระดับนัยสำคัญ 0.05 โดยเปรียบเทียบผลที่ได้จากการใช้โปรแกรม MINITAB

5.8 จากข้อมูลในข้อ 5.6 จงคำนวณค่าสัมประสิทธิ์การตัดสินใจที่ปรับแล้ว

5.9 ในการทดลองกำจัดสารเจือปนหนึ่งมีปัจจัยที่สนใจคือ อุณหภูมิและความดันโดยมีข้อมูลต่อไปนี้

ปริมาณที่กำจัดได้ (Y)	28	25	30	26	31	30	14	18	22
อุณหภูมิ (X_1)	40	40	40	50	50	50	65	65	65
ความดัน (X_2)	50	60	70	50	60	70	50	60	70

(1) จงสร้างสมการถดถอยพหุ

(2) จงทดสอบว่าตัวแบบที่ได้สามารถใช้พยากรณ์ปริมาณสารเจือปนหรือไม่ที่ระดับนัยสำคัญ 0.10

(3) จงหาว่าปัจจัยใดมีผลต่อการปริมาณสารเจือปนที่กำจัดได้ที่ระดับนัยสำคัญ 0.10 โดยใช้สถิติ t และเอฟบางส่วน พร้อมทั้งเปรียบเทียบผลที่ได้

(4) จงสร้างช่วงความเชื่อมั่น 95% ของ β_1 และ β_2

5.10 ในการทดสอบความพึงพอใจต่อน้ำผลไม้โดยมีปัจจัยที่สนใจคือ pH ปริมาณกรดแลคติกและปริมาณน้ำตาลโดยมีข้อมูลดังนี้

ความพึงพอใจ (Y)	**pH** (X_1)	**ปริมาณกรด** (X_2)	**ปริมาณน้ำตาล** (X_3)
11.0	4.5	3.14	8.6
22.5	5.2	5.04	5.3
35.7	5.4	5.44	5.7
48.2	5.6	7.54	8.1
6.7	4.6	3.87	9.9
26.3	5.8	7.01	1.9
37.7	5.9	8.76	12.9
20.4	6.1	7.69	17.8
17.1	4.8	3.85	12.1
21.0	5.2	4.14	5.7
34.8	5.7	6.42	6.2
57.2	6.4	7.08	9.5
3.7	4.5	2.96	10.6
25.6	5.2	4.92	13.0

(1) สร้างสมการถดถอยเชิงเส้นพหุ

(2) คำนวณค่าสัมประสิทธิ์การตัดสินใจสำหรับตัวแปรอิสระหลายตัวและค่าสัมประสิทธิ์การตัดสินใจที่ปรับแล้ว

5.11 จากข้อมูลในข้อ 5.10 จงหา

(1) จงสร้างช่วงความเชื่อมั่น 95% ของ β_1 β_2 และ β_3

(2) ช่วงความเชื่อมั่น 95% ของค่าคาดหวังของความพึงพอใจที่ pH = 4.9 ปริมาณกรดแลกติก = 6.80 และปริมาณน้ำตาล = 12.0

(3) ช่วงความเชื่อมั่น 95% ของค่าพยากรณ์ของความพึงพอใจจำนวน 5 ค่าที่ pH = 4.9 ปริมาณกรดแลกติก = 6.80 และปริมาณน้ำตาล = 12.0