บทที่ 6

การวิเคราะห์ความถดถอยและสหสัมพันธ์เชิงซ้อน

(Multiple Regression and Correlation Analysis)

ในบทที่ 5 ได้ศึกษาถึงความสัมพันธ์ของตัวแปร Y กับตัวแปรอิสระ X เพียงตัวเดียว นั่นคือ สนใจศึกษาปัจจัยที่มีอิทธิพลหรือผลต่อตัวแปรตาม Y เพียงปัจจัยเดียว แต่โดยทั่วๆ ไป ปัจจัยที่มีอิทธิพลต่อ Y จะมีหลายปัจจัยหรือกล่าวได้ว่ามีตัวแปรอิสระหลายตัวที่มีอิทธิพลต่อ Y เช่น ยอดขายของรถยนต์ยี่ห้อ A จะพบว่ายอดขายอาจมีความความสัมพันธ์กับราคาขายของรถยนต์ยี่ห้อ A สมรรถภาพของรถยนต์ยี่ห้อ A ค่าโฆษณา ราคาขายของรถยนต์ยี่ห้ออื่นๆ และสภาพเศรษฐกิจ ฯลฯ โดยที่ราคาขายของรถยนต์ A สมรรถภาพของรถยนต์ยี่ห้อ A ค่าโฆษณา ราคาขายของรถยนต์ยี่ห้ออื่นๆ และสภาพเศรษฐกิจ เป็นตัวแปรอิสระ หรือตัวแปรต้น ส่วนยอดขายรถยนต์ยี่ห้อ A เป็นตัวแปรตาม (Y) การที่จะพยากรณ์ยอดขายของรถยนต์ยี่ห้อ A จะต้องทราบตัวแปรอิสระต่างๆ ดังกล่าวข้างต้น

<u>6.1 รูปแบบของสมการความถดถอยเชิงซ้อน</u>

ถ้ามีตัวแปรอิสระ k ตัว ($X_1, X_2,\ldots, X_k$) ที่มีความสัมพันธ์กับตัวแปรตาม Y โดยที่ความสัมพันธ์อยู่ในรูปเชิงเส้น จะได้สมการความถดถอยเชิงเส้น ซึ่งแสดงความสัมพันธ์ระหว่าง Y และ $X_1, X_2,\ldots, X_k$ ดังนี้

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + e$$

โดยที่ β_0 = ส่วนตัดแกน Y เมื่อกำหนดให้ $X_1 = X_2 = \ldots = X_k = 0$

$\beta_1, \beta_2,\ldots, \beta_k$ เป็นสัมประสิทธิ์ความถดถอยเชิงส่วน (Partial Regression Coefficient) โดยที่ค่า β_i เป็นค่าที่แสดงถึงการเปลี่ยนแปลงของตัวแปรตาม Y เมื่อตัวแปรอิสระ X_i เปลี่ยนไป 1 หน่วย โดยที่ตัวแปรอิสระ X ตัวอื่นๆ มีค่าคงที่

เช่น ถ้า X_1 เปลี่ยนไป 1 หน่วย ค่า Y จะเปลี่ยนไป β_1 หน่วย โดยที่ $X_2, X_3,\ldots, X_k$ มีค่าคงที่

จากตัวอย่างที่ 5.6 ซึ่งเป็นการศึกษาหาความสัมพันธ์ของยอดขายแชมพูสระผมซาซ่ากับค่าโฆษณาทางวิทยุ เพื่อประมาณยอดขายเมื่อทราบ/กำหนดค่าโฆษณา แต่จะพบว่าในความเป็นจริงนั้นยอดขายยาสระผมซาซ่าไม่ได้ขึ้นอยู่กับค่าโฆษณาทางวิทยุเพียงอย่างเดียว ยังอาจจะขึ้นอยู่กับราคาแชมพูสระผมซาซ่า คุณภาพของแชมพูสระผมซาซ่า ราคาขายของแชมพูยี่ห้ออื่นๆ ซึ่งเป็นคู่แข่ง เป็นต้น ดังนั้นถ้าต้องการประมาณยอดขายแชมพู สระผมซาซ่าให้ใกล้เคียงกับยอดขายจริงจึงควรพิจารณานำตัวแปรอิสระอื่นๆ ที่มีความสัมพันธ์กับยอดขายเข้ามาเพิ่มในสมการความถดถอยด้วย แทนที่จะพิจารณาเฉพาะค่าในโฆษณาเพียงอย่างเดียว ถ้าสนใจประมาณยอดขายโดยพิจารณาจากความสัมพันธ์ระหว่างค่าโฆษณากับราคาขายต่อหน่วย จะกำหนดให้

Y : ยอดขายแชมพูสระผมซาซ่า (หน่วย : 10,000 บาท)

X_1: ค่าโฆษณาทางวิทยุ (หน่วย : 10,000 บาท)

X_2: ราคาขายต่อหน่วยของแชมพูสระผมซาซ่า (หน่วย : 10บาท)

สมการความถดถอยซึ่งแสดงความสัมพันธ์ระหว่าง Y กับ X_1และ X_2 ในรูปเชิงเส้นจะเป็น

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e$$

โดยที่ β_0 = ส่วนตัดแกน Y เมื่อ $X_1 = X_2 = 0$

β_1 เป็นค่าที่แสดงถึงยอดขาย (Y) ที่เปลี่ยนแปลงไปเมื่อค่าโฆษณา (X_1) เปลี่ยนไป 10,000 บาท (1หน่วย) ในขณะที่ราคาขายต่อหน่วย (X_2) คงเดิม

β_2 เป็นยอดขายที่เปลี่ยนไป เมื่อให้ราคาขายต่อหน่วย (X_2) เปลี่ยนไป 10 บาท (1หน่วย)ในขณะที่ค่าโฆษณา (X_1) คงเดิม

สำหรับความหมายของเครื่องหมายของ β_1และ β_2 จะเหมือนกับในบทที่ 12 การวิเคราะห์ความถดถอยอย่างง่ายและหน่วยของ β_1และ β_2 จะเหมือนกับหน่วยของ Y

<u>สมมติฐานหรือเงื่อนไขของการวิเคราะห์ความถดถอยเชิงซ้อน</u>

สมมติฐานของการวิเคราะห์ความถดถอยเชิงซ้อนจะเหมือนกับสมมติฐานของการวิเคราะห์ความถดถอยเชิงเส้นอย่างง่าย โดยที่สมการความถดถอยเชิงซ้อนเป็น

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + e$$

สมมติฐานมีดังนี้

1. ความคลาดเคลื่อน e เป็นตัวแปรที่มีการแจกแจงแบบปกติ
2. ค่าเฉลี่ยของความคลาดเคลื่อนเป็นศูนย์ นั่นคือ $E(e) = 0$
3. ค่าแปรปรวนของความคลาดเคลื่อนเป็นค่าคงที่ที่ไม่ทราบค่า $V(e) = \sigma_e^2$
4. e_i และ e_j เป็นอิสระต่อกัน ; $i \neq j$ นั่นคือ covariance $(e_i , e_j) = 0$

<u>6.2 การประมาณค่าพารามิเตอร์ของสาการความถดถอยเชิงซ้อน</u>

จากสมการความถดถอยเชิงซ้อน ซึ่งมีพารามิเตอร์ k+1 ตัวคือ $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ การประมาณค่า $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ จะต้องใช้ข้อมูลตัวอย่างของตัวแปร $Y, X_1, X_2, \ldots, X_k$ โดยใช้ตัวอย่างขนาด n จากสมการความถดถอยเชิงซ้อน

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + e_i \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(6.1)$$

จะประมาณค่า Y หรือประมาณสมการที่ (13.1) ด้วยสมการที่ (13.2)

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \ldots + \hat{\beta}_k X_{ki} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(6.2)$$

หรือ $\hat{Y}_i = a + b_1X_{1i}+b_2X_{2i}+\ldots+b_kX_{ki}$ ………………………(6.3)

โดยที่ $\hat{\beta}_0 = a$, $\hat{\beta}_1 = b_1$, $\hat{\beta}_2 = b_2$, …., $\hat{\beta}_k = b_k$

ดังนั้นค่าคลาดเคลื่อนในการประมาณค่า Y_i ด้วย $\hat{Y}_i$ คือ $Y_i-\hat{Y}_i = e_i$ (สมการที่ (6.3)- (6.1))

การประมาณค่า $\beta_0, \beta_1, \beta_2,\ldots, \beta_k$ ด้วยค่า $a, b_1 , b_2 , \ldots, b_k$ ตามลำดับนั้นยังคงมีเป้าหมายเหมือนกับความถดถอยเชิงเส้นอย่างง่าย คือ เพื่อทำให้ผลบวกของค่าคลาดเคลื่อนยกกำลังสองมีค่าน้อยที่สุด โดยใช้วิธีกำลังสองน้อยที่สุด (Least Square) นั่นคือหาค่า $a, b_1 , b_2 , \ldots, b_k$ ที่ทำให้ $\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2$ มีค่าต่ำสุด

ในกรณีที่มีตัวแปรอิสระ 2 ตัว (X_1, X_2) ที่มีความสัมพันธ์กับ Y สมการความถดถอย คือ

$Y_i = \beta_0+ \beta_1X_{1i}+ \beta_2X_{2i} +e_i$ ………………………(6.4)

ค่าประมาณของ Y_i คือ $\hat{Y}_i = a + b_1X_{1i}+b_2X_{2i}$ ………………………(6.5)

และความคลาดเคลื่อน $e_i = Y_i- \hat{Y}_i$

ต้องการ $\min\sum_{i=1}^{n} e_i^2 = \min \sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2$ จึงใช้อนุพันธ์เชิงส่วน (Partial Derivative) เทียบกับ a, b_1 และ b_2 แล้วให้เท่ากับศูนย์ดังนี้

$$\frac{\partial}{\partial a}\left[\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2\right] = \frac{\partial}{\partial a}\left[\sum_{i=1}^{n}(Y_i - a - b_1X_{1i} - b_2X_{2i})^2\right]$$

$$= -2\sum_{i=1}^{n}(Y_i - a - b_1X_{1i} - b_2X_{2i}) = 0$$

หรือ $-2 \sum_{i=1}^{n} Y_i + 2na + 2 b_1\sum_{i=1}^{n} X_{1i} + 2 b_2 \sum_{i=1}^{n} X_{2i} = 0$

$na + b_1\sum_{i=1}^{n} X_{1i} + b_2 \sum_{i=1}^{n} X_{2i} = \sum_{i=1}^{n} Y_i$ ………………………(6.6)

$$\frac{\partial}{\partial b_1}\left[\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2\right] = \frac{\partial}{\partial b_1}\left[\sum_{i=1}^{n}(Y_i - a - b_1X_{1i} - b_2X_{2i})^2\right]$$

$$= -2X_{1i}\sum_{i=1}^{n}(Y_i - a - b_1X_{1i} - b_2X_{2i}) = 0$$

หรือ $-2\sum X_{1i}Y_i+2a\sum X_{1i}+2 b_1\sum X_{1i}^2+2 b_2\sum X_{1i} X_{2i} = 0$

$a\sum X_{1i}+ b_1\sum X_{1i}^2+ b_2\sum X_{1i} X_{2i} =\sum X_{1i}Y_i$ ………………………(6.7)

ในทำนองเดียวกัน $\frac{\partial}{\partial b_2}\left[\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2\right] = 0$ จะได้

$2\sum X_{2i}Y_i+2a\sum X_{2i}+2 b_1\sum X_{1i}X_{2i}+2 b_2\sum X_{2i}^2 = 0$

$a\sum X_{2i}+ b_1\sum X_{1i}X_{2i}+ b_2\sum X_{2i}^2 = \sum X_{2i}Y_i$ ………………………(6.8)

สมการ (5.6), (5.7) และ (5.8) เรียกว่าชุดของสมการปกติ ดังนี้

$na + b_1\sum_{i=1}^{n} X_{1i} + b_2 \sum_{i=1}^{n} X_{2i} = \sum_{i=1}^{n} Y_i$

$a\sum X_{1i}+ b_1\sum X_{1i}^2+ b_2\sum X_{1i} X_{2i} =\sum X_{1i}Y_i$ ………………………(6.9)

$a\sum X_{2i} + b_1\sum X_{1i}X_{2i} + b_2\sum X_{2i}^2 \qquad = \sum X_{2i}Y_i$

การคำนวณหาค่า a, b_1 และ b_2 อาจทำได้ 3 วิธีดังนี้

1. การหาค่า a, b_1 และ b_2 จากชุดของสมการปกติ 3 สมการ (สมการที่ (6.9))

ทำได้โดยการแทนค่า $\sum X_{1i}, \sum X_{2i}, \sum X_{1i}^2, \sum X_{2i}^2, \sum X_{1i}X_{2i}$ ในชุดสมการปกติแล้วจึงหาค่า a, b_1 และ b_2 ได้โดยแก้สมการปกติ 3 สมการ

2. ใช้เมทริกซ์

ในกรณีที่ k > 2 การคำนวณหาค่า a, b_1 , b_2 , …., b_k จากชุดของสมการปกติ k + 1 สมการอาจจะยุ่งยาก ดังนั้นอาจเขียนชุดสมการปกติ (สมการที่ (13.9)) ในรูปของเมทริกซ์ ในกรณีที่ k=2 ได้ดังนี้

$$\begin{bmatrix} n & \sum x_1 & \sum x_2 \\ \sum x_1 & \sum x_1^2 & \sum x_1x_2 \\ \sum x_2 & \sum x_1x_2 & \sum x_2^2 \end{bmatrix} \begin{bmatrix} a \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} \sum Y \\ \sum X_1Y \\ \sum X_2Y \end{bmatrix} \qquad \text{……………………….(6.10)}$$

หรือใช้สัญลักษณ์ $X'Xb \quad = \quad X'Y$ ……………………….(6.11)

โดยที่เมทริกซ์

$$X = \begin{bmatrix} 1 & X_{11} & X_{21} \\ 1 & X_{12} & X_{22} \\ \vdots & \vdots & \vdots \\ 1 & X_{1n} & X_{2n} \end{bmatrix} \quad \text{และ} \quad b = \begin{bmatrix} a \\ b_1 \\ b_2 \end{bmatrix}$$

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \quad \text{และ} \quad X'X = \underset{(k+1)x(k+1)}{\begin{bmatrix} n & \sum x_1 & \sum x_2 \\ \sum x_1 & \sum x_1^2 & \sum x_1x_2 \\ \sum x_2 & \sum x_1x_2 & \sum x_2^2 \end{bmatrix}}$$

$$X'Y = \underset{(k+1)x1}{\begin{bmatrix} \sum Y \\ \sum X_1Y \\ \sum X_2Y \end{bmatrix}} \quad \text{และ} \quad b = \begin{bmatrix} a \\ b_1 \\ b_2 \end{bmatrix} = (X'X)^{-1} X'Y$$

กรณีที่มีตัวแปรอิสระในสมการความถดถอยเพียงตัวเดียวคือ X จะได้

เวคเตอร์ $X = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}$ $\qquad b = \begin{bmatrix} a \\ b \end{bmatrix}$ $\therefore X'X = \begin{bmatrix} n & \sum X \\ \sum X & \sum X^2 \end{bmatrix}$

$$b = (X'X)^{-1} X'Y \qquad \ldots\ldots\ldots\ldots\ldots\ldots(6.12)$$

หรือ $b = \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} n & \sum X \\ \sum X & \sum X^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum Y \\ \sum XY \end{bmatrix}$

ในรูปทั่วๆไป ถ้าเมทริกซ์ $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$

จะได้ $A^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \begin{bmatrix} \frac{d}{D} & -\frac{b}{D} \\ -\frac{c}{D} & \frac{a}{D} \end{bmatrix}$ โดยที่ D เป็นดีเทอร์มิแนนท์ของเมทริกซ์ A และ D= ad-bc

กรณีที่มีตัวแปรอิสระในสมการความถดถอย 2 ตัว คือ X_1 และ X_2

จากสมการที่ (6.12) จะได้

$$\therefore b = \begin{bmatrix} a \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} n & \sum x_1 & \sum x_2 \\ \sum x_1 & \sum {x_1}^2 & \sum x_1 x_2 \\ \sum x_2 & \sum x_1 x_2 & \sum {x_2}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum Y \\ \sum X_1 Y \\ \sum X_2 Y \end{bmatrix}$$

การคำนวณหาเมทริกซ์ผกผันของเมทริกซ์ $X'X$ กรณีที่ $k \geq 2$ ค่อนข้างจะยุ่งยาก ในบทนี้จะแสดงการหาเมทริกซ์ผกผันของเมทริกซ์ $X'X$ กรณีที่ k = 2 ในรูปสูตรสำเร็จดังนี้

ถ้าให้ $X'X = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & k \end{bmatrix}$ จะได้ $(X'X)^{-1} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & k \end{bmatrix}^{-1} = \begin{bmatrix} A & B & C \\ D & E & F \\ G & H & K \end{bmatrix}$

โดยที่ $A = \frac{(ek-fh)}{Z}$; $\quad B = -\frac{(bk-ch)}{Z}$; $\quad C = \frac{(bf-ce)}{Z}$

$D = -\frac{(dk-fg)}{Z}$; $\quad E = \frac{(ak-cg)}{Z}$; $\quad F = -\frac{(af-cd)}{Z}$

$G = \frac{(dh-eg)}{Z}$; $\quad H = -\frac{(ah-bg)}{Z}$; $\quad K = \frac{(ae-bd)}{Z}$

และ Z เป็นดีเทอร์มิแนนท์ของเมทริกซ์ $X'X$ โดยที่ Z = a(ek-fh) – b(dk-fg) +c(dh-eg)

ดังนั้น $\begin{bmatrix} a \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} A & B & C \\ D & E & F \\ G & H & K \end{bmatrix} \begin{bmatrix} \sum Y \\ \sum X_1 Y \\ \sum X_2 Y \end{bmatrix}$

3. การหาค่าสัมประสิทธิ์ความถดถอยโดยใช้วิธีลัด

วิธีนี้จะใช้ค่า X_1 , X_2 และ Y กรณีที่มีตัวแปรอิสระในสมการความถดถอย 2 ตัวที่เบี่ยงเบนจากค่าเฉลี่ย โดยกำหนดให้

$$y = Y - \bar{Y}$$

$$x_1 = X_1 - \bar{X}_1$$

และ

$$x_2 = X_2 - \bar{X}_2$$

ซึ่งทำให้

$$\sum y = \sum(Y - \bar{Y}) = \sum Y - n\bar{Y} = 0$$

$$\sum x_1 = \sum(X_1 - \bar{X}_1) = \sum X_1 - n\bar{X}_1 = 0 \quad \text{.........................(6.13)}$$

$$\sum x_2 = \sum(X_2 - \bar{X}_2) = \sum X_2 - n\bar{X}_2 = 0$$

จากชุดสมการปกติ (สมการที่ (5.9)) ซึ่งมี 3 สมการแทนค่า $(X_1 - \bar{X}_1)$, $(X_2 - \bar{X}_2)$ และ $(Y - \bar{Y})$ ด้วย x_1 , x_2 และ y ตามลำดับ และจากเงื่อนไขของชุดสมการที่ (5.13) จะทำให้ชุดสมการปกติเหลือเพียง 2 สมการคือ

$$b_1\sum {x_1}^2 + b_2\sum x_1 x_2 = \sum x_1 y \quad \text{.........................(6.14)}$$

$$b_1\sum x_1 x_2 + b_2\sum {x_1}^2 = \sum x_2 y_i \quad \text{.........................(6.15)}$$

โดยที่

$$\sum x_1^2 = \sum(X_1 - \bar{X}_1)^2 = \sum {X_1}^2 - n{\bar{X}_1}^2$$

$$\sum x_1 y = \sum(X_1 - \bar{X}_1)(Y - \bar{Y}) = \sum X_1 Y - n\bar{X}_1\bar{Y}$$

$$\sum {x_2}^2 = \sum(X_2 - \bar{X}_2)^2 = \sum {X_2}^2 - n{\bar{X}_2}^2 \quad \text{.............(6.16)}$$

$$\sum x_1 x_2 = \sum(X_1 - \bar{X}_1)(X_2 - \bar{X}_2) = \sum X_1 X_2 - n\bar{X}_1\bar{X}_2$$

$$\sum x_2 y = \sum(X_2 - \bar{X}_2)(Y - \bar{Y}) = \sum X_2 Y - n\bar{X}_2\bar{Y}$$

ดังนั้นเมื่อแทนค่า $\sum x_1^2, \sum x_1 y, \sum {x_2}^2, \sum x_1 x_2, \sum x_2 y$ ลงในสมการที่ (6.14)-(6.15) จะทำให้สามารถคำนวณหาค่า b_1 และ b_2 ได้ แล้วจึงหาค่า a ได้จากสมการ

$$a = \bar{Y} - b_1\bar{X}_1 - b_2\bar{X}_2 \quad \text{.........................(6.17)}$$

ตัวอย่างที่ 6.1 บริษัทซึ่งสร้างคอนโดมิเนียมขายแห่งหนึ่ง ได้สร้างคอนโดมิเนียมขายมาแล้วหลายแห่งในสถานที่ต่างๆกัน ฝ่ายวิเคราะห์ของบริษัทต้องการศึกษาถึงความสัมพันธ์ระหว่างราคาขายกับพื้นที่และต้นทุนที่ดินในแต่ละแห่ง จึงสุ่มข้อมูลในอดีตของบริษัทได้ดังนี้

ราคาขาย (หน่วย : 10,000 บาท)	พื้นที่คอนโดมิเนียม (หน่วย : 10 ตารางเมตร)	ราคาที่ดินต่อตารางเมตร (หน่วย : 100 บาท)
36	9	8
80	15	7
44	10	9
55	11	10
35	10	6

<u>วิธีทำ</u> ให้ Y = ราคาขายคอนโดมิเนียม หน่วย : 10,000 บาท

X_1= พื้นที่คอนโดมิเนียม หน่วย : 10 ตารางเมตร

X_2 = ราคาที่ดินต่อตารางเมตร หน่วย : 100 บาท

สมการความถดถอยเชิงเส้นซึ่งแสดงความสัมพันธ์ระหว่าง Y กับ X_1และ X_2 คือ

$$Y= \beta_0+ \beta_1X_1+ \beta_2X_2 + e$$

ค่าประมาณของยอดขายคือ

$$\widehat{Y}_i = a + b_1X_1+b_2X_2$$

ในตัวอย่างนี้จะแสดงการหาค่า a, b_1 และ b_2 ทั้ง 3 วิธี

<u>วิธีที่ 1</u> จากชุดสมการปกติ

$$na + b_1\sum X_{1i}+b_2\sum X_{2i} = \sum Y_i$$

$$a\sum X_{1i}+ b_1\sum X_{1i}^2+ b_2\sum X_{1i}X_{2i} = \sum X_{1i}Y_i$$

$$a\sum X_{2i}+ b_1\sum X_{1i}X_{2i} + b_2\sum X_{2i}^2 = \sum X_{2i}Y_i$$

ตารางที่ 13.1

	Y	X_1	X_2	X_1^2	X_2^2	X_1X_2	X_1Y	X_2Y
	36	9	8	81	64	72	324	288
	80	15	7	225	49	105	1,200	560
	44	10	9	100	81	90	440	396
	55	11	10	121	100	110	605	550
	35	10	6	100	36	60	350	210
ผลรวม	250	55	40	627	330	437	2,919	2,004

แทนค่า $n=5$ $\sum X_{1i}$, $\sum X_{2i}$, $\sum Y_i$, $\sum X_{1i}Y_i$, $\sum X_{2i}Y_i$, $\sum X_{1i}X_{2i}$, $\sum X_{1i}^2$, $\sum X_{2i}^2$

จากตารางที่ 6.1 แทนค่าลงในชุดสมการปกติได้ดังนี้

$5a + 55b_1 + 40b_2 \quad = \quad 250$ ……………………….(6.18)

$55a + 627b_1 + 437b_2 \quad = \quad 2,919$ ……………………….(6.19)

$40a + 437b_1 + 330b_2 \quad = \quad 2,004$ ……………………….(6.20)

แก้สมการที่ (6.18)(6.19) และ (6.20) ซึ่งมี 3 สมการ และค่าที่ต้องการหา 3 ค่า คือ a,b_1,b_2 ได้ผลลัพธ์เป็น $a = -61.29$, $b_1= 8.07$ และ$b_2 = 2.82$

<u>วิธีที่ 2</u> สามารถหาค่า a,b_1 และ b_2 ได้จากสมการ

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

โดยที่ $\begin{bmatrix} n & \sum x_1 & \sum x_2 \\ \sum x_1 & \sum {x_1}^2 & \sum x_1x_2 \\ \sum x_2 & \sum x_1x_2 & \sum {x_2}^2 \end{bmatrix} = \begin{bmatrix} 5 & 55 & 40 \\ 55 & 627 & 437 \\ 40 & 437 & 330 \end{bmatrix}$

หรือ $\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & k \end{bmatrix} = \begin{bmatrix} 5 & 55 & 40 \\ 55 & 627 & 437 \\ 40 & 437 & 330 \end{bmatrix}$

$$X'Y = \begin{bmatrix} 59 \\ 2919 \\ 2004 \end{bmatrix}$$

จะได้ a=5, b=55, c= 40, d=55, e=627, f=437, g=40, h =437 และ k =330

และดีเทอร์มิแนนท์ของ $X'X = a(ek\text{-}fh) – b(dk\text{-}fg) + c(dh\text{-}eg)$

$$Z = 5[(627)(330)-(437)(437)]-55[(55)(350)-(437)(40)]+ 40[55(437)-627(40)] = 1,055$$

และ

$$A = \frac{(ek-fh)}{Z} = \frac{(627)(330)-(437)(437)}{1055} = 15.11$$

$$B = -\frac{(bk-ch)}{Z} = \frac{-(55(330)-40(437))}{1055} = -.635$$

$$C = \frac{(bf-ce)}{Z} = \frac{(55)(437)-(40)(627)}{1055} = -.991$$

$$D = -\frac{(dk-fg)}{Z} = \frac{-(55(330)-437(40))}{1055} = -.635$$

$$E = \frac{(ak-cg)}{Z} = \frac{(5)(330)-(40)(40))}{1055} = .047$$

$$F= -\frac{(af-cd)}{Z} = \frac{-(5(437)-40(55))}{1055} = .014$$

$$G = \frac{(dh-eg)}{Z} = \frac{(55)(437)-(627)(40)}{1055} = -.991$$

$$H = -\frac{(ah-bg)}{Z} = \frac{(5)(437)-(55)(40)}{1055} = .014$$

$$K = \frac{(ac-bd)}{Z} = \frac{(5)(627)-(55)(55)}{1055} = .104$$

$\therefore (X'X)^{-1} = \begin{bmatrix} A & B & C \\ D & E & F \\ G & H & K \end{bmatrix} = \begin{bmatrix} 15.11 & -.635 & -.991 \\ -.635 & .047 & .014 \\ -.991 & .014 & .104 \end{bmatrix}$

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

$\therefore b = \begin{bmatrix} a \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 15.11 & -.635 & -.991 \\ -.635 & .047 & .014 \\ -.991 & .014 & .104 \end{bmatrix} \begin{bmatrix} 250 \\ 2919 \\ 2004 \end{bmatrix} = \begin{bmatrix} -61.29 \\ 8.07 \\ 2.82 \end{bmatrix}$

นั่นคือ $a = -61.29$, $b_1 = 8.07$, $b_2 = 2.82$

<u>วิธีที่ 3</u> คำนวณ $\bar{Y} = \sum Y/n = 250/5 = 50$, $\bar{X}_1 = 55/5 = 11$, $\bar{X}_2 = 40/5 = 8$

จากชุดสมการที่ (13.16) คำนวณหาค่าต่างๆ โดยใช้ค่าจากตารางที่ 13.1 ได้ดังนี้

$\sum x_1^2 = \sum X_1^2 - n\bar{X}_1^2 = 627-5(11)^2 = 20$

$\sum x_1 y = \sum X_1 Y - n\bar{X}_1\bar{Y} = 2919-5(11)(50) = 169$

$\sum x_2^2 = \sum X_2^2 - n\bar{X}_2^2 = 330 - 5(8)^2 = 10$

$\sum x_1 x_2 = \sum X_1 X_2 - n\bar{X}_1\bar{X}_2 = 437-5(11)(8) = -3$

$\sum x_2 y = \sum X_2 Y - n\bar{X}_2\bar{Y} = 2004-5(8)(50) = 4$

นำค่า $\sum x_1^2, \sum x_1 y, \sum x_2^2, \sum x_1 x_2, \sum x_2 y$ แทนค่าในสมการที่ (13.14) และ (13.15) ได้ดังนี้

$22b_1 - 3b_2 = 169$ ……………………….(6.21)

$-3 b_1 - 10b_2 = 4$ ……………………….(6.22)

3X(13.21) : $66b_1 - 9b_2 = 507$ ……………………….(6.23)

22X(13.22) : $-66b_1 + 220b_2 = 88$ ……………………….(6.24)

(6.23) + (6.24) ได้ : $211 b_2 = 595$

$b_2 = 595/211 = 2.82$

แทนค่า $b_2 = 2.82$ ลงในสมการที่ (6.22) จะได้

$-3b_1 + 10(2.82) = 4$

$b_1 = 8.07$

หาค่า a โดยการแทนค่า $b_1 = 8.07$ และ $b_2 = 2.82$ ลงในสมการที่ (6.17) จะได้

$a = \bar{Y} - b_1 \bar{X}_1 - b_2 \bar{X}_2 = 50 - 8.07(11) - 2.82(8) = -61.29$

$\therefore$ ค่าประมาณของสมการความถดถอยคือ

$\hat{Y} = -61.29 + 8.07X_1 + 2.82X_2$

<u>6.3 ความหมายของสัมประสิทธิ์ความถดถอยเชิงส่วน</u>

ถ้ามีตัวแปรอิสระที่มีความสัมพันธ์กับตัวแปรตาม (Y) 3 ตัว คือ X_1, X_2 และ X_3 สมการความถดถอยเชิงซ้อนคือ

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + e$$

ค่าประมาณของ Y คือ

$$\hat{Y} = a + b_1X_1 + b_2X_2 + b_3X_3 \qquad \text{……………………….(6.25)}$$

จากสมการที่ (6.25) ค่าประมาณของพารามิเตอร์คือ a, b_1, b_2 และ b_3 ตามลำดับโดยที่

a คือ ส่วนหรือระยะตัดแกน Y ซึ่ง β_0, β_1 และ β_2 หมายถึงเมื่อกำหนดให้ $X_1 = X_2 = X_3 = 0$

b_1, b_2 และ b_3 เป็นค่าประมาณของสัมประสิทธิ์การตัดสินใจเชิงส่วนซึ่งมีหน่วยเหมือน Y และมีความหมายดังนี้

b_1 เป็นค่าซึ่งแสดงถึงความสัมพันธ์ระหว่าง Y และ X_1 หมายถึง ถ้า X_1 เพิ่มขึ้น 1 หน่วย จะทำให้ Y เปลี่ยนไป b_1 หน่วย (ขึ้นอยู่กับเครื่องหมายของ b_1) โดยที่กำหนดให้ตัวแปรอิสระอื่นๆ คือ X_2 และ X_3 มีค่าคงที่

b_2 เป็นค่าซึ่งแสดงถึงความสัมพันธ์ระหว่าง Y และ X_2 หมายถึง ถ้า X_2 เพิ่มขึ้น 1 หน่วย จะทำให้ Y เปลี่ยนไป b_2 หน่วย โดยที่กำหนดให้ X_1 และ X_3 มีค่าคงที่ในทำนองเดียวกัน

b_3 จะแสดงถึงความสัมพันธ์ระหว่าง Y กับ X_3 โดยที่ถ้า X_3 เพิ่มขึ้น 1 หน่วย จะทำให้ Y เปลี่ยนไป b_3 หน่วย โดยที่กำหนดให้ X_1 และ X_2 มีค่าคงที่

<u>ตัวอย่างที่ 6.2</u> จากตัวอย่างที่ 6.1 จงอธิบายความหมายของค่า a, b_1 และ b_2

<u>วิธีทำ</u> จากตัวอย่างที่ 6.1 ซึ่งมีตัวแปรอิสระ 2 ตัว คือ X_1 (พื้นที่คอนโดมิเนียม) และ X_2 (ราคาที่ดิน) ได้สมการความถดถอยเป็น

$$\hat{Y} = -61.29 + 8.07X_1 + 2.82X_2$$

โดยที่ a = -61.29 (10,000 บาท) = -612,900 บาท ซึ่งเป็นค่าที่ $X_1 = X_2 = 0$

$b_1 = 8.07$ หมายถึงถ้าเพิ่มพื้นที่คอนโดมิเนียมขึ้น 1 หน่วย หรือ 10 ตารางเมตร จะทำให้ราคาขายเพิ่มขึ้น 8.07 หน่วย หรือ 80,700 บาท โดยที่ให้ราคาที่ดิน (X_2) คงที่เช่น ให้ X_2 เป็น 9 หน่วย หรือ 900 บาทต่อตารางเมตร

b_2= 2.82 หมายถึงถ้าราคาที่ดิน (ต้นทุน) เพิ่มขึ้น 1 หน่วยหรือ 100 บาท จะมีผลทำให้ราคาขาย เพิ่มขึ้น 2.82 หน่วย หรือ 28,200 บาท โดยให้ขนาดของคอนโดมิเนียม (X_1) คงที่ เช่นให้ X_1 =10 หน่วยหรือ 100 ตารางเมตร

หมายเหตุ เครื่องหมายของ b_1, …, b_k จะมีความหมายเหมือนเครื่องหมายของ b ในความหมายถดถอยอย่างง่ายในบทที่ 12 คือถ้า b_i มีค่าติดลบ หมายความว่า ถ้า X_i เพิ่มขึ้น 1 หน่วย จะทำให้ Y ลดลง b_i หน่วย โดยกำหนดให้ $X_1,X_2,\dots,X_{i-1},\dots,X_k$ มีค่าคงที่ แต่ถ้า b_i มีค่าบวก หมายความว่า ถ้า X_i เพิ่มขึ้น 1 หน่วย จะทำให้ Y เพิ่มขึ้น b_i หน่วย โดยที่ให้ $X_1,X_2,\dots,X_{i-1},\dots,X_k$ มีค่าคงที่

6.4 การทดสอบสมการความถดถอยเชิงซ้อนโดยใช้การวิเคราะห์ความแปรปรวนแบบทางเดียว

จากสมการความถดถอยเชิงซ้อน

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + e$$

ค่าแปรปรวนของ Y = ค่าแปรปรวนที่เกิดจากอิทธิพลของ $X_1, X_2,\dots,X_k$ + ค่าแปรปรวนอย่างสุ่ม

หรือ SST = SSR + SSE

โดยที่ SST (Sum Square of Total) คือค่าแปรปรวนทั้งหมดของ Y หรือ SST = $\sum_{i=1}^{n}(Y_i - \bar{y})^2$

SSR (Sum Square of Regression) คือ ค่าแปรปรวนของ Y เนื่องจากอิทธิพลของ $X_1,\dots,X_k$

SSE (Sum Square of Error) คือ ค่าแปรปรวนของ Y เนื่องจากอิทธิพลอื่นๆ หรือเรียกว่า ค่าแปรปรวนอย่างสุ่ม

ตารางที่ 6.2 : การวิเคราะห์ความแปรปรวนของการวิเคราะห์ความถดถอยเชิงซ้อน

แหล่งแปรปรวน (SV)	คงค่าอิสระ (DF)	ผลบวกกำลังสอง (SS)	ผลบวกกำลังสองเฉลี่ย (MS)	F
ความถดถอย (Regression)	k	SSR	MSR = $\frac{SSR}{k}$	$\frac{MSR}{MSE}$
ความคลาดเคลื่อน (Error)	n-k-1	SSE	MSE= $\frac{SSE}{n-k-1}$	
ผลรวม (Total)	n-1	SST		

โดยที่ SSR = $\underline{b}'\, X'\underline{Y} - n\bar{y}^2$

$$SST = \sum_{i=1}^{n}(Y_i - \bar{y})^2 = \underline{Y}'\,\underline{Y} - n\bar{y}^2$$

$SSE = \sum_{i=1}^{n}[Y_i - (a + b_1X_1 + b_2X_2 + \cdots + b_kX_k)]^2$

หรือ $SSE = SST - SSR = \underline{Y}'\,\underline{Y} - \underline{b}'\,X'\underline{Y}$

จากตารางการวิเคราะห์ความแปรปรวนจะใช้ในการทดสอบสมมติฐานเกี่ยวกับความสัมพันธ์ระหว่าง Y และ $X_1, X_2, \ldots, X_k$ โดยตั้งสมมติฐานไว้ดังนี้

> $H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$
>
> H_1 : มี β_i อย่างน้อย 1 ค่าที่ $\neq 0$; i = 1,2,…,k
>
> สถิติทดสอบ $F = \frac{MSR}{MSE} = \frac{(b'X'Y - n\bar{y}^2)/k}{(Y'Y - b'X'Y)/(n-k-1)}$
>
> เขตปฏิเสธ จะปฏิเสธสมมติฐาน H_0 ถ้า $F > F_{k,n-k-1;1-\alpha}$

ผลของการทดสอบสมมติฐานอาจจะเป็น

ก. ยอมรับสมมติฐาน $H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$ ซึ่งสรุปได้ว่า Y ไม่มีความสัมพันธ์กับ $X_1, X_2, \ldots, X_k$ ในรูปเชิงเส้น

ข. ปฏิเสธสมมติฐาน H_0 หรือยอมรับสมมติฐาน H_1 ซึ่งสรุปได้ว่า มี X_i อย่างน้อย 1 ตัว ที่มีความสัมพันธ์กับ Y ในรูปเชิงเส้น จึงต้องทดสอบต่อไปว่า X_i ตัวใดที่มีความสัมพันธ์กับ Y โดยใช้สถิติทดสอบ t ซึ่งจะได้กล่าวถึงในหัวข้อ 6.8

<u>ตัวอย่างที่ 6.3</u> จากตัวอย่างที่ 6.1 จงทดสอบความสัมพันธ์ระหว่างราคาขายคอนโดมิเนียม (Y) กับพื้นที่คอนโดมิเนียม (X_1) และราคาที่ดิน (X_2)

<u>วิธีทำ</u> จากตัวอย่างที่ 6.1 ซึ่งมีความสัมพันธ์ระหว่าง Y กับ X_1และ X_2 เป็น

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + e$$

จะได้ $\underline{Y}'\,\underline{Y} = [36\ \ 80\ \ 44\ \ 55\ \ 35]\begin{bmatrix}36\\80\\44\\55\\35\end{bmatrix} = 13{,}882$

SST $= \underline{Y}'\,\underline{Y} - n\bar{y}^2 = 13{,}882 - 5(50)^2 = 1{,}382$

$\underline{b}' = [-61.29 \quad 8.07 \quad 2.82]$ $X'\,\underline{Y} = \begin{bmatrix}250\\2919\\2004\end{bmatrix}$

SSR $= \underline{b}'\,X'\,\underline{Y} - n\bar{y}^2$

$= [-61.29 \quad 8.07 \quad 2.82]\begin{bmatrix}250\\2919\\2004\end{bmatrix} - 5(50)^2$

$= 13,874.49 - 12,500 = 1,374.49$

$\therefore$ SSE $= SST - SSR$

$= 1,382 - 1,374.49 = 7.51$ นำค่าต่างๆ ใส่ในตาราง ANOVA

ตารางที่ 6.3 : ANOVA

แหล่งแปรปรวน	องศาอิสระ	ผลบวกกำลังสอง	ผลบวกกำลังสองเฉลี่ย (MS)	F
ความถดถอย (พื้นที่และราคาที่ดิน)	2	1,374.49	$\frac{1,374.49}{2} = 684.25$	$\frac{684.25}{3.75} = 183.09$
ความคลาดเคลื่อน	5-2-1 = 2	7.51	$\frac{7.51}{2} = 3.75$	
ผลรวม	5-1 = 4	1,382		

<u>สมมติฐานที่ต้องการทดสอบ</u>

$H_0 : \beta_1 = \beta_2 = 0$

H_1 : มี β_i อย่างน้อย 1 ค่าที่ $\neq 0$; $i = 1, 2$

<u>สถิติทดสอบ</u> $F = \frac{MSR}{MSE} = \frac{684.25}{3.75} = 183.09$

<u>เขตปฏิเสธ H_0</u>

เปิดตาราง F ในภาคผนวกที่ 7 ได้ค่า $F_{2,2;.95} = 19.0$ แต่ F ที่คำนวณได้ = 183.09 ซึ่งมากกว่า 19.0 จึงปฏิเสธสมมติฐาน H_0 นั่นคือมี X_i อย่างน้อย 1 ค่าที่มีความสัมพันธ์กับราคาขายคอนโดมิเนียม (Y) จึงควรทำการทดสอบต่อไปว่า ขนาดพื้นที่ (X_1) หรือ ราคาที่ดิน (X_2) ที่สัมพันธ์กับ Y หรือทั้งขนาดพื้นที่และราคาที่ดินสัมพันธ์กับ Y ทั้งคู่ซึ่งจะได้กล่าวถึงให้หัวข้อที่ 6.8

6.5 การประมาณค่าคลาดเคลื่อนของความถดถอย (Estimation of Standard Deviation of Regression)

การประมาณค่าคลาดเคลื่อนของความถดถอย หรือเรียกกันว่าการประมาณค่าความคลาดเคลื่อนมาตรฐานของการประมาณ โดยที่ความคลาดเคลื่อนซึ่งเกิดจากการพยากรณ์หรือประมาณค่า Y ด้วย $\hat{Y}$คือ e ในกรณีที่มีตัวแปรอิสระ k ตัว จะได้ค่าแปรปรวนของการประมาณ คือ

$S_e^2 = S^2_{Y.12...k} = S^2$ โดยที่ $S^2 = \frac{SSE}{n-k-1} = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n-k-1}$

ดังนั้นความคลาดเคลื่อนมาตรฐานของค่าประมาณคือ

$S = \sqrt{S^2} = \sqrt{SSE/(n-k-1)} = \sqrt{MSE}$

ในกรณีที่มีตัวแปรอิสระ 2 ตัว คือ X_1 และ X_2 จะได้ $S^2 = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n-2-1} = MSE$

ตัวอย่างที่ 6.4 จากตัวอย่างที่ 6.3 จงหาค่าความคลาดเคลื่อนของการประมาณ

$S^2 = MSE = 3.75$

$\therefore \quad S = S_{y.12} = \sqrt{3.75} = 1.936$

6.6 การประมาณค่าเบี่ยงเบนมาตรฐานของ b_i

ในการประมาณค่า β_i ด้วยค่า b_i แบบช่วงนั้น เราจะต้องทราบค่าเบี่ยงเบนมาตรฐานของ b_i; i = 1, 2, …,k โดยที่ค่าแปรปรวนของ b_i คือ $\hat{V}(\underline{b}) = S^2(\underline{b}) = (X'X)^{-1}S^2$

ในกรณีที่ k =2 หรือมีตัวแปรอิสระ 2 ตัว คือ X_1 และ X_2

$$\therefore \quad \underline{b} = \begin{bmatrix} a \\ b_1 \\ b_2 \end{bmatrix} \qquad S_b^{\ 2} = \begin{bmatrix} S_a^2 & S_{a,b_1} & S_{a,b_2} \\ S_{a,b_1} & S_{b_1}^2 & S_{b_1b_2} \\ S_{a,b_2} & S_{b_1b_2} & S_b^2 \end{bmatrix}$$

และค่าเบี่ยงเบนมาตรฐานของ b_i คือ

$$S_{b1} = \sqrt{S_{b_1}^2} = \frac{S}{\sqrt{\sum x_1^2(1-r_{12}^2)}} \qquad \text{……………………….(6.26)}$$

$$S_{b2} = \sqrt{S_{b_2}^2} = \frac{S}{\sqrt{\sum x_2^2(1-r_{12}^2)}} \qquad \text{……………………….(6.27)}$$

โดยที่ $x_i = (X_{ij} - \overline{X}_i)$

r_{12} = สัมประสิทธิ์สหสัมพันธ์อย่างง่ายระหว่าง X_1 และ X_2

$$r_{12} = \frac{\sum x_1x_2}{\sqrt{\left(\sum x_1^2\right)\left(\sum x_2^2\right)}} \qquad \text{……………………….(6.28)}$$

6.7 การประมาณค่าสัมประสิทธิ์ความถดถอยแบบช่วง

ในการประมาณค่า β_i แบบช่วง จะต้องทราบค่าเบี่ยงเบนมาตรฐานของ b_i โดยที่แบ่งเป็น

ก. ตัวอย่างมีขนาดเล็ก (n< 30)

ค่าประมาณแบบช่วงของ β_i ที่ระดับความเชื่อมั่น (1- α) 100% คือ

$$\boxed{b_i \pm t_{1-\alpha/2;\, n-k-1}\, S_{bi}}$$

ข. ตัวอย่างมีขนาดใหญ่ (n $\geq$ 30)

ค่าประมาณแบบช่วงของ β_i ที่ระดับความเชื่อมั่น (1- α) 100% คือ

$$\boxed{b_i \pm Z_{1-\alpha/2} S_{bi}}$$

ตัวอย่างที่ 6.5 จากตัวอย่างที่ 6.1จงหาค่า $S^2(\underline{b})$ และจงประมาณค่า β_1และ β_2 แบบช่วงที่ระดับความเชื่อมั่น 95%

วิธีทำ จากตัวอย่างที่ 6.1 ที่กำหนดให้

Y = ราคาขายคอนโดมิเนียม หน่วย : 10,000 บาท

X_1 = พื้นที่คอนโดมิเนียม หน่วย : 10 ตารางเมตร

X_2 = ราคาที่ดินต่อตารางเมตร หน่วย : 100 บาท

และได้สมการความถดถอยดังนี้

$$\widehat{Y} = -61.29 + 8.07X_1 + 2.82X_2$$

$$\therefore\ b = \begin{bmatrix} a \\ b_1 \\ b_2 \end{bmatrix} \text{ และได้ } (X'X)^{-1} = \begin{bmatrix} 15.11 & -.635 & -.991 \\ -.635 & .047 & .014 \\ -.991 & .014 & .104 \end{bmatrix}$$

และ $S^2_{Y.12}$ = MSE = 3.75 $\therefore\ S^2_{Y.12} = (X'X)^{-1}$

$$= (3.75)\begin{bmatrix} 15.11 & -.635 & -.991 \\ -.635 & .047 & .014 \\ -.991 & .014 & .104 \end{bmatrix} = \begin{bmatrix} 56.66 & -2.38 & -.991 \\ -2.38 & .176 & .014 \\ -3.72 & .053 & .104 \end{bmatrix}$$

ดังนั้น $S^2_{b1} = .176$ และ $S^2_{b2} = .39$

$\therefore\ S_{b1} = \sqrt{.176} = .42 \qquad S_{b2} = \sqrt{.39} = .625$

หรืออาจใช้สมการที่ (13.25) และ (13.26) หาค่า S_{b1} และ S_{b2} ได้ดังนี้

$$S_{bi} = \frac{S}{\sqrt{\sum x_i^2(1-r_{12}^2)}}\ ;\ i = 1,2$$

จากตารางที่ 13.1 แทนค่าต่างๆ ได้ดังนี้

$$\sum x_1^2 = \sum x_1^2 - n\overline{X}_1{}^2 = 627 - 5(11)^2 = 22$$

$$\sum x_2^2 = \sum x_2^2 - n\overline{X}_2{}^2 = 330 - 5(8)^2 = 10$$

$$\sum x_1 x_2 = \sum x_1 x_2 - n\overline{X}_1\overline{X}_2 = 437 - 5(11)(8) = -3$$

$$r_{12} = \frac{\sum x_1 x_2}{\sqrt{\left(\sum x_1^2\right)\left(\sum x_2^2\right)}} = \frac{-3}{\sqrt{(22)(10)}} = -.202 \text{ ดังนั้น } r^2_{12} = (-.202)^2 = .0408$$

$$S = \sqrt{S^2} = \sqrt{3.75} = 1.936$$

$$\therefore \qquad S_{b1} = \frac{1.936}{\sqrt{22(1-.0408)}} = .42 \qquad S_{b2} = \frac{1.936}{\sqrt{10(1-.0408)}} = .625$$

เนื่องจาก n= 5 ดังนั้นค่าประมาณแบบช่วงของสัมประสิทธิ์ความถดถอยระหว่างราคาขายกับขนาดพื้นที่ (β_1) ที่ระดับความเชื่อมั่น 95% คือ

$b_i \pm t_{.975,2}S_{b1}$ = $(8.07 \pm 4.30(.42)$ = (6.27, 9.87)

ดังนั้น 6.27 $\leq \beta_1 \leq$ 9.87 (หน่วย : 10,000 บาท)

หรือ 62,7000 $\leq \beta_1 \leq$ 98,700 บาท

นั่นคือถ้าเพิ่มพื้นที่คอนโดมิเนียมขึ้น 10 ตารางเมตร (1 หน่วย) จะทำให้ราคาขายเพิ่มขึ้นในช่วง 62,700 บาท ถึง 98,700 บาท ที่ระดับความเชื่อมั่น 95%

ในทำนองเดียวกัน ค่าประมาณแบบช่วงของสัมประสิทธิ์ความถดถอยระหว่างราคาขายกับ ราคาที่ดิน (β_2) ที่ระดับความเชื่อมั่น 95%คือ

$b_2 \pm t_{.975,2}S_{b2}$

$2.82 \pm 4.30(.42)$

ดังนั้น 1.02 $\leq \beta_2 \leq$ 4.62 (หน่วย : 10,000 บาท)

หรือ 10,200 $\leq \beta_2 \leq$ 46,200 บาท

นั่นคือถ้าราคายที่ดินเพิ่มขึ้น (1หน่วย) จะทำให้ราคาขายคอนโดมิเนียมเพิ่มขึ้น 10,200 บาทถึง 46,200 บาท ที่ระดับความเชื่อมั่น95%

<u>6.8 การทดสอบสมมติฐานเกี่ยวกับสัมประสิทธิ์ความถดถอย</u>

ถ้ามีตัวแปรอิสระ k ตัว ($X_1, X_{2,\ldots} X_k$) ที่มีความสัมพันธ์กับตัวแปรตาม Y และเมื่อได้ทดสอบ F-test จากตารางการวิเคราะห์ความแปรปรวนเพื่อทดสอบสมมติฐาน

$H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$

H_1 : มี β_i อย่างน้อย 1 ค่าที่ $\neq 0$; i = 1,2,…,k

ผลของการทดสอบสมมติฐานข้างต้นโดยการใช้สถิติ F จะเป็น

ก.ยอมรับสมมติฐาน H_0 ถ้า $F < F_{k,n-k-1}$ แสดงว่าตัวแปร Y ไม่มีความสัมพันธ์กับตัวแปรอิสระทั้ง k ตัว ($X_1, X_{2,\ldots} X_k$) หรือ

ข. ปฏิเสธสมมติฐาน H_0 ถ้า $F < F_{k,n-k-1}$ แสดงว่ามีตัวแปรอิสระ (X's) อย่างน้อย 1ตัวที่มีความสัมพันธ์กับ Y ในกรณีที่เกิด ข. คือ ปฏิเสธ H_0 จะต้องทำการทดสอบต่อไปว่ามี β_i ตัวใดบ้างที่ไม่เท่ากับศูนย์ หรือมี X ตัวใดบ้างที่สัมพันธ์กับ Y โดยการทดสอบสมมติฐานดังต่อไปนี้

<u>สมมติฐาน</u> $H_0 : \beta_i = 0$

$H_1 : \beta_i \neq 0$; i = 1,2,…,k

<u>สถิติทดสอบ</u> $t = \frac{b_i - 0}{S_{bi}}$ หรือใช้สถิติทดสอบ Z ถ้า n มีค่ามาก

เขตปฏิเสธสมมติฐาน H_0

จะปฏิเสธสมมติฐาน H_0 เมื่อ $t > t_{1-\alpha;n-k-1}$ หรือ $t < -t_{1-\alpha;n-k-1}$ หรือกล่าวว่าจะปฏิเสธ H_0 ถ้า $|t| > t_{1-\alpha;n-k-1}$

สรุปหลักเกณฑ์ในการทดสอบสัมประสิทธิ์ความถดถอย (β_i ; i=1,2,…,k)

การทดสอบ 2 ด้าน	การทดสอบด้านซ้าย	การทดสอบด้านขวา
$H_0 : \beta_i = 0$ $H_1 : \beta_i \neq 0$	$H_0 : \beta_i \geq 0$ $H_1 : \beta_i < 0$	$H_0 : \beta_i \leq 0$ $H_1 : \beta_i > 0$
สถิติทดสอบ $t = \frac{b_i}{S_{bi}}$	สถิติทดสอบ $t = \frac{b_i}{S_{bi}}$	สถิติทดสอบ $t = \frac{b_i}{S_{bi}}$
เขตปฏิเสธ H_0 $\lvert t\rvert > t_{1-\alpha/2;n-k-1}$	เขตปฏิเสธ H_0 $t < -t_{1-\alpha;n-k-1}$	เขตปฏิเสธ H_0 $t > t_{1-\alpha;n-k-1}$

ตัวอย่างที่ 6.6 จากตัวอย่างที่ 6.3 จงทดสอบความสัมพันธ์ระหว่างราคาขายกับพื้นที่คอนโดมิเนียม และราคาขายกับราคาที่ดิน ที่ระดับนัยสำคัญ .05

วิธีทำ จากตัวอย่างที่ 6.3 ซึ่งทดสอบความสัมพันธ์ระหว่างราคาขายกับพื้นที่และราคาที่ดิน โดยใช้สถิติ F ซึ่งผลการทดสอบคือปฏิเสธสมมติฐาน H_0: $\beta_1 = \beta_2 = 0$ นั่นคือมี β_i อย่างน้อย 1 ตัวไม่เท่ากับศูนย์ หรือกล่าวได้ว่าพื้นที่คอนโดมิเนียมมีความสัมพันธ์กับราคาขายหรือราคาที่ดินมีความสัมพันธ์กับราคาขายหรือทั้งพื้นที่และราคาที่ดินมีความสัมพันธ์กับราคาขายคอนโดมิเนียม จึงต้องทำการทดสอบต่อไปโดยใช้สถิติทดสอบ t

สมการความถดถอยซึ่งแสดงความสัมพันธ์ระหว่างราคาขายกับพื้นที่และราคาที่ดินเป็น

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e$

ก. การทดสอบความสัมพันธ์ระหว่างราคาขายกับขนาดของคอนโดมิเนียม

$H_0 : \beta_1 = 0$ หรือ H_0 : ราคาขายและพื้นที่คอนโดมิเนียมไม่มีความสัมพันธ์กัน

$H_1 : \beta_1 \neq 0$ หรือ H_1 : ราคาขายและพื้นที่คอนโดมิเนียมมีความสัมพันธ์กัน

สถิติทดสอบ $t = \frac{b_i}{S_{bi}} = 8.07/.42 = 19.26$

เขตปฏิเสธสมมติฐาน H_0 $t_{1-\alpha/2;2} = t_{.975;2} = 4.30$

จะปฏิเสธ H_0 ถ้า $t > 4.30$ หรือ $t < -4.30$ หรือจะปฏิเสธ H_0 ถ้า $|t| > 4.30$

เนื่องจาก $t = 19.26 > 4.30$ จึงปฏิเสธ H_0

สรุปผลการทดสอบ ราคาขายคอนโดมิเนียมมีความสัมพันธ์กับพื้นที่คอนโดมิเนียมที่ความเชื่อมั่น 95% หรือระดับนัยสำคัญ .05

ข.การทดสอบความสัมพันธ์ระหว่างราคาขายกับราคาที่ดิน

$H_0 : \beta_2 = 0$ หรือ H_0 : ราคาขายและพื้นที่คอนโดมิเนียมไม่มีความสัมพันธ์กัน

$H_1 : \beta_2 \neq 0$ หรือ H_1 : ราคาขายและพื้นที่คอนโดมิเนียมมีความสัมพันธ์กัน

สถิติทดสอบ $t = \frac{b_i}{S_{bi}} = 2.82/.62 = 4.55$

เขตปฏิเสธสมมติฐาน H_0

จะปฏิเสธ H_0 ถ้า $t > 4.30$ หรือ $t < -4.30$ เนื่องจาก $t = 4.55 > 4.30$ จึงปฏิเสธ H_0 ที่ระดับนัยสำคัญ .05

สรุปผลการทดสอบ ราคาขายคอนโดมิเนียมมีความสัมพันธ์กับราคาที่ดินที่ระดับนัยสำคัญ .05 ดังนั้นจากผลการทดสอบทั้ง ก. และ ข. สรุปได้ว่า ราคาขายคอนโดมิเนียมมีความสัมพันธ์กับพื้นที่คอนโดมิเนียมและราคาที่ดินที่ระดับนัยสำคัญ .05

ข้อสังเกต นอกจากจะสามารถสรุปถึงความสัมพันธ์ระหว่าง Y กับ X_1 และ Y กับ X_2 โดยการทดสอบสมมติฐานดังแสดงในตัวอย่างที่ 13.6 แล้ว ยังอาจใช้ค่าประมาณแบบช่วงของ β_1 และ β_2 ในการสรุปความสัมพันธ์ระหว่าง Y กับ X_1 และ Y กับ X_2 ได้เช่นกัน จากตัวอย่างที่ 13.5 ได้ค่าประมาณแบบช่วงของ β_1 เป็น

$62,700 \leq \beta_1 \leq 98,700$ บาท

ซึ่งค่าต่ำสุดและค่าสูงสุดมากกว่าศูนย์ นั่นคือ $\beta_1 \neq 0$ จึงสรุปได้ว่าราคาขายกับพื้นที่คอนโดมิเนียมมีความสัมพันธ์กันและสัมพันธ์ในทิศทางเดียวกัน คือถ้าพื้นที่เพิ่ม ราคาขายจะเพิ่มขึ้นด้วย

และ $10,200 \leq \beta_2 \leq 46,200$ บาท

ซึ่งหมายถึง $\beta_2 \neq 0$ นั่นคือ ราคาขายคอนโดมิเนียมกับราคาที่ดินมีความสัมพันธ์กันในทางเดียวกันเช่นกัน

6.9 สัมประสิทธิ์การตัดสินใจเชิงซ้อน (Multiple Coefficient of Determination : R^2 หรือ r^2)

สัมประสิทธิ์การตัดสินใจเชิงซ้อนจะมีความหมายเหมือนกับความหมายของสัมประสิทธิ์การตัดสินใจในบทที่ 12 คือเป็นสัดส่วนหรือเปอร์เซ็นต์ที่ตัวแปรอิสระ ($X_1, X_2, \ldots X_k$) สามารถอธิบายการเปลี่ยนแปลงของ Y ได้ หรือกล่าวได้ว่าสัมประสิทธิ์การตัดสินใจเชิงซ้อนเป็นสัดส่วนหรือเปอร์เซ็นต์ของความผันแปร Y ที่มีสาเหตุเนื่องจากความผันแปรขง ($X_1, X_2, \ldots X_k$) โดยที่สัมประสิทธิ์การตัดสินใจเชิงซ้อนจะให้สัญลักษณ์ $R^2_{Y.123\ldots k}$ แต่โดยทั่วไปจะใช้ R^2

$$r^2 = R^2 = \frac{\text{ความผันแปรของ Y เนื่องจากอิทธิพลของ } X_1, X_2, \ldots, X_k}{\text{ความผันแปรทั้งหมด}}$$

$$= SSR/SST$$

หรือ $$r^2 = R^2 = \frac{SST-SSE}{SST} = 1 - \frac{SSE}{SST}$$ ....(6.29)

โดยที่ $0 \leq R^2 \leq 1$

ถ้าค่า R^2 ที่ใกล้ 1 จะหมายถึง $X_1, X_2, \ldots, X_k$ มีความสัมพันธ์กับ Y มาก แต่ถ้า R^2 เข้าใกล้ศูนย์ หมายถึง ค่า$X_1, X_2, \ldots, X_k$มีความสัมพันธ์กับ Y น้อย

เนื่องจาก SSR จะเพิ่มขึ้นถ้าเพิ่มตัวแปรอิสระ เช่น เดิมมี X_1 และ X_2ที่มีความสัมพันธ์กับ Y แต่ถ้าเพิ่มตัวแปรอิสระ X_3 เข้าในสมการความถดถอย จะได้ว่า

$$SSR\ (X_1, X_2, X_3) > SSR\ (X_1, X_2)$$

โดยที่ SSR $(X_1, X_2, , X_3)$ หมายถึง SSR ของสมการความถดถอยที่มีตัวแปรอิสระ $X_1, X_2, , X_3$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + e$$

และ SSR (X_1, X_2) หมายถึง SSR ของสมการความถดถอยที่มีตัวแปรอิสระเพียง 2 ตัว คือ X_1, X_2

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e$$

ดังนั้นเมื่อเพิ่มตัวแปรอิสระเข้าสมการความถดถอยจะทำให้ค่า R^2 มากขึ้นทั้งที่ตัวแปรอิสระ X ที่เพิ่มเข้าไปใหม่อาจจะไม่มีความสัมพันธ์กับ Y เลยก็ได้ จึงมีการปรับค่า R^2 ให้ถูกต้องขึ้นเรียกว่า Adjusted R^2 โดยที่

$$R^2_a = \text{Adjusted } R^2 = 1 - \frac{SSE/(n-k-1)}{SST/(n-1)}$$ ....(6.30)

หรือ $R^2_a = 1 + \frac{(n-1)}{n-k-1}(R^2 - 1)$

<u>ตัวอย่างที่ 6.7</u> จากตัวอย่างที่ 6.3 จงหาสัมประสิทธิ์การตัดสินใจเชิงซ้อน

<u>วิธีทำ</u> จากตารางที่ 6.3 จะได้

SSR = 1,374.49 SSE = 7.51 และ SST = 1,382

$$\therefore\ R^2 = \frac{1{,}374.49}{1{,}382} = 0.9946$$

หรือ $R^2 = 99.46\%$

$$\therefore\ R^2_a = 1 - \frac{SSE/(n-k-1)}{SST/(n-1)} = 1 - \frac{7.51/(5-2-1)}{1{,}382/(5-1)} = 0.989$$

หมายความว่า ความผันแปรทั้งหมดของราคาขายคอนโดมิเนียม (Y) มีสาเหตุจากความผันแปรของ

ขนาดพื้นที่ของคอนโดมิเนียม (X_1) และราคาที่ดิน (X_2) 99.46% (R^2) หรือ 98.9% (R^2_a) ส่วนความผันแปรของราคาขายคอนโดมิเนียมที่เหลืออีก 54% (หรือ 1.1% ถ้าใช้ R^2_a) เกิดจากสาเหตุอื่นๆ

6.10 สัมประสิทธิ์สหสัมพันธ์เชิงซ้อน (Multiple Coefficient of Correlation)

ค่าของสัมประสิทธิ์สหสัมพันธ์เชิงซ้อนได้จากการถอดรากที่สองของสัมประสิทธิ์การตัดสินใจเชิงซ้อนสัมประสิทธิ์เชิงซ้อน = $R_{Y.12...k} = R = \sqrt{R^2_{Y.12...k}}$ โดยที่ $0 \leq R \leq 1$

โดยที่สัมประสิทธิ์สหสัมพันธ์เชิงซ้อนแสดงถึงความสัมพันธ์ระหว่าง Y กับ $X_1, X_2, \dots, X_k$ ดังนี้

1. R มีค่าเข้าใกล้ศูนย์ แสดงว่า Y มีความสัมพันธ์กับ $X_1, X_2, \dots, X_k$ น้อยมาก
2. ถ้า R = 0 แสดงว่า Y ไม่มีความสัมพันธ์กับ $X_1, X_2, \dots, X_k$เลย
3. R มีค่าเข้าใกล้ 1 แสดงว่า Y มีความสัมพันธ์กับตัวแปรอิสระทั้ง k ตัวมีมาก

ตัวอย่างที่ 6.8 จากตัวอย่างที่ 6.7 จงหาค่าสัมประสิทธิ์สหสัมพันธ์เชิงซ้อน

วิธีทำ จากตัวอย่างที่ 6.7 ได้ $R^2_{Y.12} = R^2 = .9946$

$\therefore$ สัมประสิทธิ์สหสัมพันธ์เชิงซ้อน $R_{Y.12} = R = \sqrt{.9946} = .9972$ หรือ R = 99.72%

6.11 สัมประสิทธิ์สหสัมพันธ์เชิงส่วน (Coefficients of Partial Correlation)

สัมประสิทธิ์สหสัมพันธ์เชิงส่วนเป็นค่าที่แสดงความสัมพันธ์ระหว่าง Y กับ X ตัวใดตัวหนึ่งโดยกำหนดให้ X ตัวอื่นๆ มีค่าคงที่ เช่นถ้า X ความสัมพันธ์กับตัวแปรอิสระ 3 ตัว (X_1, X_2, X_3)

สัมประสิทธิ์สหสัมพันธ์เชิงส่วนระหว่าง Y กับ X_i โดยกำหนดให้ X_j และ X_k คงที่ ($i \neq j \neq k$) จะแสดงความสัมพันธ์ระหว่าง Y กับ X_iจริงๆ โดยกำจัดอิทธิพลของ X_j และ X_k ที่มีต่อ Y

สัญลักษณ์ของสัมประสิทธิ์สหสัมพันธ์เชิงส่วนที่ใช้ คือ

$r_{Y1.23}$ = สัมประสิทธิ์สหสัมพันธ์เชิงส่วนระหว่าง Y กับ X_1 โดยกำหนดให้ X_2, X_3 มีค่าคงที่ เป็นค่าที่แสดงความสัมพันธ์ระหว่าง Y กับ X_1 โดยให้ X_2, X_3 มีค่าคงที่ จึงเป็นค่าที่แสดงความสัมพันธ์ระหว่าง Y กับ X_1เท่านั้น มิใช่ความสัมพันธ์ของ X_2, X_3 กับ Y

$r_{Y2.13}$ = สัมประสิทธิ์สหสัมพันธ์เชิงส่วนระหว่าง Y กับ X_2 โดยกำหนดให้ X_1, X_3 มีค่าคงที่

$r_{Y3.12}$ = สัมประสิทธิ์สหสัมพันธ์เชิงส่วนระหว่าง Y กับ X_3 โดยกำหนดให้ X_1, X_2 มีค่าคงที่

โดยที่ $-1 \leq r_{Yi;jk} \leq 1$

การคำนวณหาสัมประสิทธิ์สหสัมพันธ์เชิงส่วนทำได้โดยการหาสัมประสิทธิ์สหสัมพันธ์อย่างง่าย ซึ่งเป็นค่าที่แสดงความสัมพันธ์ระหว่างตัวแปร 2 ตัวดังได้กล่าวแล้วในบทที่ 5

<u>สัญลักษณ์ของสัมประสิทธิ์สหสัมพันธ์อย่างง่าย</u>

r_{Yi} = สัมประสิทธิ์สหพันธ์ระหว่างตัวแปร Y กับ X_i เป็นค่าที่แสดงความสัมพันธ์ระหว่าง Y กับ X_i

r_{ij} = สัมประสิทธิ์สหพันธ์ระหว่างตัวแปร X_i กับ X_j โดยที่ $i \neq j$ เช่น r_{12}, r_{13}, r_{23}

<u>ก. ถ้ามีตัวแปรอิสระ 2 ตัวคือ X_1, X_2 ซึ่งมีความสัมพันธ์กับตัวแปรตาม Y</u>

สูตรสำหรับคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรคู่ใดคู่หนึ่งเป็นดังนี้

$r_{12} = \frac{\sum(X_1-\bar{X}_1)(X_2-\bar{X}_2)}{\sqrt{\sum(X_1-\bar{X}_1)^2\sum(X_2-\bar{X}_2)^2}} = \frac{\sum X_1X_2}{\sqrt{(\sum X_1^2)(\sum X_2^2)}}$ โดยที่ $x_i = X_i - \bar{X}_i$ $y = Y - \bar{Y}$; i=1,2

$$r_{Y1} = \frac{\sum(Y-\bar{Y})(X_1-\bar{X}_1)}{\sqrt{\sum(Y-\bar{Y})^2\sum(X_1-\bar{X}_1)^2}} = \frac{\sum yX_1}{\sqrt{(\sum y^2)(\sum X_1^2)}}$$

$$r_{Y2} = \frac{\sum(Y-\bar{Y})(X_2-\bar{X}_2)}{\sqrt{\sum(Y-\bar{Y})^2\sum(X_2-\bar{X}_2)^2}} = \frac{\sum yX_2}{\sqrt{(\sum y^2)(\sum X_2^2)}}$$

ดังนั้นจึงสามารถคำนวณหาสัมประสิทธิ์สหสัมพันธ์เชิงส่วนได้ดังนี้

$r_{Y1.2}$ = สัมประสิทธิ์สหสัมพันธ์เชิงส่วนระหว่าง Y กับ X_1 โดยกำหนดให้ X_2 มีค่าคงที่

$$r_{Y1.2} = \frac{r_{Y1}-r_{Y2}.r_{12}}{\sqrt{(1-r^2{}_{Y2})(1-r^2{}_{12})}}$$

$r_{Y2.1}$ = สัมประสิทธิ์สหสัมพันธ์เชิงส่วนระหว่าง Y กับ X_2 โดยกำหนดให้ X_1 มีค่าคงที่

$$r_{Y2.1} = \frac{r_{Y2}-r_{Y1}.r_{12}}{\sqrt{(1-r^2{}_{Y1})(1-r^2{}_{12})}}$$

<u>ข. กรณีที่มีตัวแปรอิสระ 3 ตัว คือX_1, X_2, X_3 ซึ่งมีความสัมพันธ์กับตัวแปรตาม Y</u>

การคำนวณหาสัมประสิทธิ์สหสัมพันธ์ระหว่าง Y กับ X_1, Y กับ X_2 , Y กับ X_3 , X_1 กับ X_2, X_1กับ X_3 และ X_2กับ X_3 ทำได้โดยใช้สูตรสหสัมพันธ์อย่างง่าย

ส่วนการคำนวณหาสัมประสิทธิ์สหสัมพันธ์เชิงส่วน $r_{Y1.23}$, $r_{Y2.13}$, $r_{Y3.12}$ ทำได้ดังนี้

$r_{Y1.23}$ = สัมประสิทธิ์สหสัมพันธ์เชิงส่วนระหว่าง Y กับ X_1 โดยกำหนดให้ X_2, X_3 มีค่าคงที่

$$r_{Y1.23} = \frac{r_{Y1.3}-r_{12.3}.r_{Y2.3}}{\sqrt{(1-r^2{}_{12.3})(1-r^2{}_{Y2.3})}} \; ; \; r_{Y2.13} = \frac{r_{Y2.3}-r_{12.3}.r_{Y1.3}}{\sqrt{(1-r^2{}_{12.3})(1-r^2{}_{1.3})}}$$

$$r_{Y3.12} = \frac{r_{Y3.2}-r_{31.2}.r_{Y1.2}}{\sqrt{(1-r^2{}_{31.2})(1-r^2{}_{Y1.2})}}$$

<u>ตัวอย่างที่</u> 6.9 จากตัวอย่างที่ 6.1 จงหาค่าสัมประสิทธิ์สหสัมพันธ์เชิงส่วน

<u>วิธีทำ</u> จากตัวอย่างที่ 6.1 คำนวณค่า $\sum x_1^2$ = 22 , $\sum x_1y$ = 169 , $\sum x_2^2$= 10 , $\sum x_1x_2$ = -3 และ$\sum x_2y$= 4 และ $\sum y^2$ = 1,382

คำนวณหาสัมประสิทธิ์สหสัมพันธ์อย่างง่ายได้ดังนี้

$r_{12} = \frac{\sum X_1X_2}{\sqrt{(\sum X_1^2)(\sum X_2^2)}} = \frac{-3}{\sqrt{(22)(10)}} = -0.202$

$r_{Y1} = \frac{\sum yX_1}{\sqrt{(\sum y^2)(\sum X_1^2)}} = \frac{169}{\sqrt{(1,382)(22)}} = 0.969$

$r_{Y2} = \frac{\sum yX_2}{\sqrt{(\sum y^2)(\sum X_2^2)}} = \frac{4}{\sqrt{(1,382)(10)}} = 0.034$

คำนวณค่าสัมประสิทธิ์สหสัมพันธ์เชิงส่วนได้ดังนี้

$r_{Y1.2} = \frac{r_{Y1}-r_{Y2}.r_{12}}{\sqrt{(1-r^2{}_{Y2})(1-r^2{}_{12})}} = \frac{0.969-(.034)(-.202)}{\sqrt{(1-(.034)^2)(1-(-.202)^2)}}$

$= .9758/.9788 = .997$

เมื่อเปรียบเทียบ $r_{Y1} = .969$ ซึ่งเป็นสหสัมพันธ์อย่างง่ายระหว่าง Y กับ X_1 แสดงถึงความสัมพันธ์ระหว่างราคาขาย (Y) กับ ขนาดพื้นที่ (X_1) โดยไม่ได้พิจารณาถึงราคาที่ดิน (X_2) เลย จะพบว่ามีความสัมพันธ์กันมากและอยู่ในทิศทางเดียวกัน

และ $r_{Y1.2} = .997$ หมายถึงถ้ากำหนดให้ราคาที่ดินคงที่ (ราคาที่ดินต่อตารางเมตรคงที่ เช่น ตารางเมตรละ 1,000 บาท) ราคาขายคอนโดมิเนียมกับขนาด (พื้นที่) คอนโดมิเนียมมีความสัมพันธ์ในทิศทางเดียวกัน

$r_{Y2.1} = \frac{r_{Y2}-r_{Y1}.r_{12}}{\sqrt{(1-r^2{}_{Y1})(1-r^2{}_{12})}} = \frac{.034-(.969)(-.202)}{\sqrt{(1-(.969)^2)(1-(-.202)^2}}$

$= .2297/.2419 = .9497$

ซึ่งหมายถึงเมื่อกำหนดให้พื้นที่ของคอนโดมิเนียม (X_1) คงที่ (เช่นให้มีขนาด 100 ตารางเมตร) ราคาขายกับราคาที่ดินจะมีความสัมพันธ์กันในทิศทางเดียวกัน โดยมีสัมประสิทธิ์สหสัมพันธ์เชิงส่วน = .9497

แต่ $r_{Y2} = .034$ หมายถึง ความสัมพันธ์ระหว่างราคาขายกับราคาที่ดิน โดยไม่ได้พิจารณาถึงพื้นที่คอนโดมิเนียมเลย ซึ่ง ความสัมพันธ์ยังคงอยู่ในทิศทางเดียวกัน แต่มีความสัมพันธ์กันน้อย

6.12 การใช้โปรแกรมสำเร็จรูป

ในกรณีที่ $k \geq 2$ เช่น $k = 5$ การคำนวณค่า b ค่า MSR, MSE, จะทำได้ไม่ง่ายนัก เพราะจะต้องมีความรู้ทางคณิตสาสตร์ในเรื่องเมทริกซ์ แต่เนื่องจากในปัจจุบันนี้ได้มีการนำคอมพิวเตอร์เข้ามาช่วยในการคำนวณ และมีโปรแกรมสำเร็จรูปที่คำนวณค่าสถิติต่างๆ หลายโปรแกรม เช่น SPSS,SAS,STATPACK, BMDP. ฯลฯ ในที่นี้จะใช้ SPSS for windows ช่วยในการคำนวณ และการทดสอบสมมติฐานเกี่ยวกับความสัมพันธ์ระหว่างตัวแปรตาม (Y) กับตัวแปรอิสระ $(X_1, X_2, \ldots, X_k)$ ดังจะแสดงในตัวอย่างที่ 6.9

ตัวอย่างที่ 6.9 จากตัวอย่างที่ 6.1 จงทดสอบความสัมพันธ์ระหว่างราคาขายคอนโดมิเนียม (Y) กับขนาดพื้นที่คอนโด (X_1) และ ราคาที่ดิน (X_2) ที่ระดับนัยสำคัญ .05 โดยใช้โปรแกรมสำเร็จรูป SPSS for windows

หมายเหตุ สำหรับวิธีการใช้โปรแกรม SPSS for windows ศึกษาได้จากหนังสือ "การใช้ SPSS for windows ในการวิเคราะห์ข้อมูล" ของ ดร. กัลยา วานิชย์บัญชา โดยได้ผลลัพธ์ดังนี้

ผลลัพธ์ที่ได้จากโปรแกรม SPSS for windows แสดงในตารางที่ 6.4-6.6

ตารางที่ 6.4 : Model Summary

Model	R	❶ R Square	❷ Adjusted R Square	Std. Error of the Estimate
1	.977[a]	.995	.989	1.94

A Predictors : (Constant) , cost, area

จากผลลัพธ์ ❶และ ❷ ของตารางที่ 13.4 ได้ค่า R^2 = .995 และ R^2_{adj} = .989 หมายถึงพื้นที่คอนโดมิเนียมและราคาที่ดินสามารถอธิบายความผันแปรของราคาขายได้ 98.9% ที่เหลืออีก 1.1% มีสาเหตุมาจากปัจจัยอื่นๆที่ยังไม่ได้นำมาพิจารณา

ตารางที่ 6.5 : ANOVA[b]

Model	Sum of Squares	Df	❸ Mean Square	❹ F	❺ Sig.
1 Regression	1.74.493	2	687.246	183.092	.005[a]
Residual	7.507	2	3.754		
Total	1382.000	4			

a Predictors : (Constant) , cost, area

b Dependent Variable : price

ความหมายผลลัพธ์ตารางที่ 6.5

จาก ❸ ได้ค่า MSR = 687.246 และ MSE = 3.754

การทดสอบ

$H_0 : \beta_1 = \beta_2 = 0$ หรือ H_0 : ราคาขายคอนโดมิเนียมไม่ขึ้นกับพื้นที่และราคาที่ดิน

H_1 : มี $\beta_i \neq 0$ อย่างน้อย 1 ค่า หรือ H_1 : ราคาขายคอนโดมิเนียมขึ้นกับพื้นที่หรือราคาที่ดิน หรือขึ้นกับทั้งพื้นที่และราคาขายที่ดิน

สถิติทดสอบ : $F = \frac{MSR}{MSE} = \frac{687.246}{3.754}$ = 183.092 หรือได้จากผลลัพธ์ ❹ ในตารางที่ 6.5

เขตปฏิเสธ : จะปฏิเสธ H_0 ถ้า $F > F_{.95;2,2} = 19$ หรือ ปฏิเสธ H_0 ถ้า Sig. จากตารางการทดสอบ < α ในที่นี้ $F = 183.902 > 19$ จึงปฏิเสธ H_0 หรือถ้าพิจารณาค่า Sig. ในผลลัพธ์ ❺ ได้ค่า Sig. = .005 ซึ่งน้อยกว่าค่า α = .05 จึงปฏิเสธ H_0

สรุปผลการทดสอบ β_1 หรือ β_2 หรือทั้ง β_1 และ β_2 ไม่เท่ากับศูนย์นั่นคือราคาขายคอนโดมิเนียมขึ้นกับพื้นที่หรือราคาที่ดิน หรือขึ้นกับพื้นที่และราคาที่ดิน จึงต้องทำการทดสอบต่อไปว่า β_i ตัวใดบ้างที่ $\neq 0$ โดยใช้ผลลัพธ์ในตารางที่ 6.6

I.การทดสอบว่าราคาขายคอนโดมิเนียม (Y) มีความสัมพันธ์กับพื้นที่ (X_i) หรือไม่

ตารางที่ 6.6 : Coefficients [a]

Model	Unstandardized Coefficients		Standardized	❼ T	❽ Sig.
	❻ B	Std. Error	Beta		
1 (Constant)	-61.289	7.531		-8.138	.015
Area (X_1)	8.066	.422	1.018	19.125	.003
Cost (X_2)	2.820	.626	.240	4.508	.046

a. Dependent Variable : price

$H_0 : \beta_1 = 0$ หรือ H_0 : ราคาขายคอนโดมิเนียมไม่มีความสัมพันธ์กับพื้นที่คอนโดมิเนียม

H_1 : มี $\beta_1 \neq 0$ H_1 : ราคาขายคอนโดมิเนียมมีความสัมพันธ์กับพื้นที่คอนโดมิเนียม

สถิติทดสอบ : $t = \frac{b_i}{S_{bi}} = \frac{8.066}{.422} = 19.125$ (จาก ❼ ในตารางที่ 6.6)

เขตปฏิเสธ : จะปฏิเสธ H_0 ถ้า $|t| = t_{1-\alpha/2;n-k-1} = t_{.975;2} = 4.30$

หรือ Sig. < α = .05 จากผลลัพธ์ ❽ ได้ Sig. = .003 ซึ่งน้อยกว่า α = .05 จึงปฏิเสธ H_0

II.การทดสอบว่าราคาขายคอนโดมิเนียม (Y) มีความสัมพันธ์กับราคาที่ดิน (X_2) หรือไม่

$H_0 : \beta_2 = 0$

H_1 : มี $\beta_2 \neq 0$

สถิติทดสอบ : จาก ❼ ได้ t= 4.508

เขตปฏิเสธ : จะปฏิเสธ H_0 ถ้า Sig. < α

ซึ่งจาก ❽ ได้ Sig. = .046 < .005 จึงปฏิเสธ H_0 นั่นคือ $\beta_2 \neq 0$

สรุปผลการทดสอบ ราคาขายคอนโดมิเนียมมีความสัมพันธ์กับราคาขายที่ดิน ดังนั้นจากผลการทดสอบทั้งหมดจะได้ว่า ราคาขายคอนโดมิเนียมมีความสัมพันธ์กับพื้นที่และราคาขายที่ดิน จาก ❻ ในตารางที่ 6.6 ได้ $b_1 = 8.066$ และ $b_2 = 2.82$ สมการความถดถอยจึงเป็น $\hat{Y} = -61.289 + 8.066\ X_1 + 2.82\ X_2$

<u>ตัวอย่างที่ 6.10</u> โรงงานแห่งหนึ่งต้องการวิเคราะห์ต้นทุนในการทำกล่องกระดาษว่ามีความสัมพันธ์กับประมาณกระดาษเวลาที่ใช้เครื่องจักรและต้นทุนคงที่เช่นค่าเช่าโรงงาน จึงเก็บข้อมูลตัวอย่างการทำกล่องกระดาษได้ดังนี้

ต้นทุน (1000 บาท)	ปริมาณกระดาษ (ตัน)	เวลาที่ใช้เครื่องจักร (ชั่วโมง)	ต้นทุนคงที่ (1,000 บาท)
1102	550	218	112
1008	502	199	99
1227	616	249	126
1395	701	277	143
1710	838	363	191
1881	919	399	210
1924	99	411	216
1246	622	248	124
1255	626	259	127
1314	659	266	135
1557	740	334	181
1887	901	401	216
1204	610	238	127
1211	598	246	117
1287	646	259	135
1451	732	286	155
1828	891	389	208
1903	932	404	220
1997	964	430	233
1363	680	271	129
1421	723	286	146
1543	784	317	158
1774	841	376	199
1929	922	415	228

ก.จงสร้างสมการความถดถอยเชิงเส้นแสดงความสัมพันธ์ระหว่างต้นทุนผลิตกล่องกระดาษกับประมาณกระดาษที่ใช้ เวลาที่ใช้เครื่องจักร ต้นทุนคงที่และแรงงานที่ใช้

ข.อธิบายความหมายของสมการความถดถอยในข้อ ก.

ค.จงทดสอบความสัมพันธ์ระหว่างต้นทุนผลิตกล่องกระดาษ กับปริมาณกระดาษที่ใช้ เวลาที่ใช้เครื่องจักร ต้นทุนคงที่และแรงงานที่ใช้ที่ระดับนัยสำคัญ .05

ง. จงประมาณค่าแบบช่วงของ β_i ที่ระดับความเชื่อมั่น 95%

จ. จงหาค่าสัมประสิทธิ์การตัดสินใจ

<u>วิธีทำ</u> ในที่นี้ตัวแปรตามคือต้นทุน และมีตัวแปรอิสระ 3 ตัว (k=3) ซึ่งคาดว่ามีความสัมพันธ์กับต้นทุน ตัวแปรอิสระ ได้แก่ ปริมาณกระดาษ เวลาที่ใช้เครื่องจักร ต้นทุนคงที่และแรงงาน โดยกำหนดให้

Y = ต้นทุนผลิตกล่องกระดาษ (หน่วย : 1,000 บาท)

X_1 = ปริมาณกระดาษ (หน่วย : ตัน)

X_2 = เวลาที่ใช้เครื่องจักร (หน่วย : ชั่วโมง)

X_3 = ต้นทุนคงที่ (หน่วย : 1,000 บาท)

สมการความถดถอยเชิงเส้น ซึ่งแสดงความสัมพันธ์ระหว่าง Y กับ X_1, X_2 และ X_3 เป็น

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + e$$

ค่าประมาณของ Y คือ $\hat{Y}$ โดยที่ $\hat{Y} = a + b_1X_1 + b_2X_2 + b_3X_3$

การหาค่า a, b_1,b_2 และ b_3 ค่อนข้างยุ่งยาก ในที่นี้จะใช้โปรแกรมสำเร็จรูป SPSS for windows โดยได้ผลลัพธ์ดังในตารางที่ 13.7-13.9

ตารางที่ 6.7 : Coefficients [a]

Model	Unstandardized Coefficients		Standardized	❸	❹
	❶ B	Std. Error ❷	Beta	T	Sig.
1 (Constant)	63.752	19.035		3.349	.003
X_1	.954	.104	.435	9.175	.000
X_2	2.086	.353	.495	5.906	.000
X_3	.520	.452	.073	1.152	.263

a. Dependent Variable : price

ก. จากในแถวตั้ง B ใน ❶ของตารางที่ 6.7 จะได้สมการความถดถอยเชิงเส้น ดังนี้

$$\hat{Y} = 63.752 + .954 X_1 + 2.086X_2 + .52X_3$$

ข. จากสมการความถดถอยเชิงเส้นใน ข้อ ก. จะได้ว่า

b_1= .954 นั่นคือ ถ้าใช้กระดาษเพิ่มขึ้น 1 ตัน จะทำให้ต้นทุนในการผลิต กล่องกระดาษเพิ่มขึ้น .954 หน่วย หรือ 954 บาท

b_2= 2.086 หมายความว่า ถ้าใช้เครื่องจักรในการผลิตกล่องกระดาษเพิ่มขึ้น 1 ชั่วโมงจะทำให้ต้นทุนในการผลิตกล่องกระดาษเพิ่มขึ้น 2.086 หน่วย หรือ 2086 บาท

b_3 = .52 หมายถึง ถ้าต้นทุนคงที่เพิ่มขึ้น 1,000 บาท จะทำให้ต้นทุนในการผลิต กล่องกระดาษเพิ่มขึ้น 520 บาท

ค. การทดสอบความสัมพันธ์ระหว่างต้นทุน (Y) กับ ปริมาณกระดาษที่ใช้ (X_1) เวลาเครื่องจักรที่ใช้ (X_2) และต้นทุนคงที่ (X_3) จะต้องทำการทดสอบ 2 ขั้นตอน ดังนี้

ขั้นที่ 1 $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$

H_1 : มี β_i อย่างน้อย 1 ค่า $\neq 0$; i=1,2,3,4

ตารางที่ 6.8 : ANOVA[b]

Model	Sum of Squares	Df	❺ Mean Square	❻ F	❼ Sig.
1 Regression	2179906	3	726635.4	7462.866	.000[a]
Residual	1947.336	20	97.367		
Total	2181854	23			

a Predictors : (Constant) , X_3, X_1, X_2

b Dependent Variable : price

จาก ❺ ในตารางที่ 13.8 จะได้ MSR = 726,635.4 , MSE = 97.367

<u>สถิติทดสอบ</u> : $F = \frac{MSR}{MSE}$ =7462.866 (จาก ❻ ในตารางที่ 6.8)

ในที่นี้ k=3, n= 24; n-k-1 = 24-3-1 =20

<u>เขตปฏิเสธ</u> : จะปฏิเสธ H_0 ถ้า $F > F_{.95;3,20}$ หรือ Sig. < α

ในที่นี้ α =.05 และจาก ❼ ได้ Sig. = 0.000 ซึ่งน้อยกว่า .05 จึงปฏิเสธ H_0 นั้นคือ $\beta_i \neq 0$ อย่างน้อย 1 ค่า ดังนั้น จะมีตัวแปร X_1 หรือ X_2 หรือ X_3 อย่างน้อย 1 ตัวที่มีความสัมพันธ์กับ Y จึงต้องทำการทดสอบต่อไปว่ามี X_i ตัวใดที่มีความสัมพันธ์กับ Y โดยใช้สถิติทดสอบ t

<u>ขั้นที่ 2</u> ใช้ผลลัพธ์จากตารางที่ 13.7 : Coefficients โดยพิจารณาจากผลลัพธ์ ❸และ ❹

$H_0 : \beta_i = 0$

H_1 : มี $\beta_i \neq 0$

<u>สถิติทดสอบ</u> : t ซึ่งคือค่า t ใน ❸ ของตารางที่ 6.7

<u>เขตปฏิเสธ :</u> จะปฏิเสธ H_0 ถ้า $|t| = t_{1-\alpha/2;}20$ หรือ Sig. หรือผลลัพธ์ ❹ ในตารางที่ 6.7 < α = .05

<u>1. การทดสอบเกี่ยวกับค่าคงที่ β_0</u>

$H_0 : \beta_0 = 0$

$H_1 : \beta_0 \neq 0$

<u>สถิติทดสอบ</u> t = 3.349 (จากผลลัพธ์ ❸ ในตารางที่ 6.7)

<u>เขตปฏิเสธ</u> Sig. < .05

ในที่นี้จากผลลัพธ์ ❹ ในตารางที่ 13.7 ได้ Sig. = .003 ซึ่งน้อยกว่า .05 จึงปฏิเสธ H_0 นั่นคือ $\beta_0 \neq 0$

<u>2. การทดสอบว่าต้นทุน (Y) มีความสัมพันธ์กับประมาณกระดาษ (X_1) หรือไม่</u>

$H_0 : \beta_1 = 0$

$H_1 : \beta_1 \neq 0$

<u>เขตปฏิเสธ</u> Sig. = .000< .05 จึงปฏิเสธ H_0 นั่นคือ $\beta_1 \neq 0$

นั่นคือต้นทุนผลิตกล่องกระดาษมีความสัมพันธ์กับประมาณกระดาษที่ใช้ในรูป เชิงเส้นที่ระดับนัยสำคัญ .05

<u>3. การทดสอบว่าต้นทุน (Y) มีความสัมพันธ์กับเวลาที่ใช้เครื่องจักร(X_2) หรือไม่</u>

$H_0 : \beta_2 = 0$

$H_1 : \beta_2 \neq 0$

<u>เขตปฏิเสธ</u> Sig. < .05

แต่ Sig T = .000 <.05 จึงยอมรับ H_0 นั่นคือ $\beta_2 = 0$ หรือต้นทุนผลิตกล่องกระดาษมีความสัมพันธ์กับเวลาที่ใช้เครื่องจักรผลิตที่ระดับนัยสำคัญ .05

<u>4. การทดสอบว่าต้นทุน (Y) มีความสัมพันธ์กับต้นทุนคงที่ (X_3) หรือไม่</u>

$H_0 : \beta_3 = 0$

$H_1 : \beta_3 \neq 0$

<u>เขตปฏิเสธ</u> Sig. < .05

แต่ Sig = .263 >.05 (จาก❹ในตารางที่ 6.7) จึงปฏิเสธ H_0 นั่นคือ

$\beta_3 \neq 0$ หรือ ต้นทุนผลิตกล่องกระดาษ ไม่มีความสัมพันธ์กับต้นทุนคงที่ ที่ระดับนัยสำคัญ .05

จากการทดสอบ 1-4 สรุปได้ว่า ต้นทุนผลิตกล่องกระดาษ (Y) มีความสัมพันธ์กับปริมาณกระดาษที่ใช้ (X_1) กับเวลาเครื่องจักรที่ใช้(X_2) ในรูปเชิงเส้น ส่วนต้นทุนคงที่ (X_3) ไม่มีความสัมพันธ์กับ Y ในรูปเชิงเส้นและค่าคงที่ $\beta_0 \neq 0$ ที่ระดับนัยสำคัญ .05

ดังนั้นสมการความถดถอยจะกลายเป็น

$$\hat{Y} = 63.752 + .954\, X_1 + 2.086X_2$$

ง. ค่าประมาณแบบช่วงของ β_i ที่ระดับความเชื่อมั่น 95% คือ $b_i \pm t_{.975;20}S_{bi}$

จาก ❷ ในตารางที่ 13.7 ในแถวตั้งของ Std.Error ซึ่งหมายถึง ค่าคลาดเคลื่อนมาตรฐานของ b_i ดังนั้น S_{b1} = .104 S_{b2} = .353 และ S_{b3} = .452 $t_{.975;20}$ = 2.09

ง.1 ค่าประมาณแบบช่วงของ β_1 ที่ระดับความเชื่อมั่น 95% คือ

$$b_1 \pm t_{.975;20}S_{b1} = .954 \pm 2.09(.104) = .954 \pm .217$$

นั่นคือ $.827 \leq \beta_1 \leq 1.171$

หมายความว่า ถ้าปริมาณกระดาษที่ใช้เพิ่มขึ้น (X_1) 1 ตัน จะทำให้ต้นทุนผลิตกล่องกระดาษเพิ่มขึ้นในช่วง 827 บาท ถึง 1171 บาท

ง.2 ค่าประมาณ แบบช่วงของ β_2 ที่ระดับความเชื่อมั่น 95% คือ

$$2.086 \pm 2.09(.353) = 2.086 \pm .738$$

หรือ $1.348 \leq \beta_2 \leq 2.824$

หมายความว่า ถ้าเวลาที่ใช้เครื่องจักรผลิตเพิ่มขึ้น 1 ชั่วโมง จะทำให้ต้นทุนผลิตกล่องกระดาษเพิ่มขึ้นในช่วง 1,348 บาท ถึง 2,824 บาท

ง.3 ค่าประมาณแบบช่วงของ β_3 ที่ระดับความเชื่อมั่น 95% คือ

$$.52 \pm 2.09(.452) = .52 \pm .945$$

หรือ $-.425 \leq \beta_3 \leq 1.465$

จะพบว่าค่าประมาณแบบช่วงของ β_3 รวมศูนย์อยู่ด้วย (ค่าต่ำสุดติดลบ (-.425) แต่ค่าสูงสุดเป็นบวก (1.465)) แสดงว่า β_3 อาจจะเป็นศูนย์ซึ่งผลเหมือนกับการทดสอบ $H_0 : \beta_3 = 0$ ซึ่งยอมรับว่า $\beta_3 = 0$

จากข้อ ง.1 – ง.3 จะได้ผลสรุปเหมือนกับข้อ ค. คือมีตัวแปร X_1 กับ X_2 เท่านั้น ที่มีความสัมพันธ์กับ Y ในรูปเชิงเส้น

<u>จ.การหาค่าสัมประสิทธิ์การตัดสินใจ</u>

ตารางที่ 6.9 : Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	1.000[a]	.999	.999	9.87

a Predictors : (Constant) , X_3,X_1,X_2

จากตารางที่ 13.9 มีค่า $R^2 = R^2_{adj}$ = .999 นั่นคือ สัมประสิทธิ์การตัดสินใจ = .999

<u>6.3 การประมาณค่า Y เมื่อกำหนดค่า X_1,X_2,... X_k</u>

เมื่อทดสอบความสัมพันธ์ระหว่าง Y กับตัวแปรอิสระ X_1,X_2,... X_k ว่า X_i ตัวใดมีความสัมพันธ์กับ Y บ้าง ถ้ามี X_i อย่างน้อย 1 ตัวที่มีความสัมพันธ์กับตัวแปรตาม เมื่อกำหนดค่า X ที่สัมพันธ์กับ Y จะทำให้สามารถประมาณ / พยากรณ์ค่า Y ได้

<u>ตัวอย่าง 6.11</u> จากตัวอย่างที่ 6.1 ถ้าพื้นที่คอนโดมิเนียมเป็น 12 ตารางเมตร และที่ดินราคาตารางเมตรละ 990 บาท ราคาขายคอนโดมิเนียมจะเป็นเท่าใด

<u>วิธีทำ</u> จากตัวอย่างที่ 13.6 ทำให้ทราบว่าสมการความถดถอยที่ถูกต้องคือ

$\widehat{Y} = -61.289 + 8.066\,X_1 + 2.82X_2$

ในที่นี้กำหนดให้ $X_1 = 12$, $X_2 = 9.9$

$\therefore \widehat{Y} = -61.289 + 8.066\,(12) + 2.82(9.9) = 63.468$

นั่นคือ ถ้าคอนโดมิเนียมมีขนาด 12 ตารางเมตร และราคาที่ดินเป็น 990 บาทต่อตารางเมตร จะขายในราคา 634,680 บาท

<u>แบบฝึกหัดที่ 6</u>

6.1 ถ้า n= 70 , k=8 และคำนวณได้ SST = 1,526.3 SSE = 1,162.2 จงหา

ก. S_e ข. R^2 ค. F

6.2 จากสมการความถดถอยได้ $Y = \beta_1 X_1 + \beta_2 X_2 + e$, n = 50 คำนวณได้ SST = 321.2, SSE = 259 จงหา

ก. SSR ข. ค่าคลาดเคลื่อนของค่าประมาณ

ค. สัมประสิทธิ์การตัดสินใจ ง. สัมประสิทธิ์การตัดสินใจแบบ Adjusted

จ. จงทดสอบความเหมาะสมของสมการความถดถอยข้างต้น ที่ระดับนัยสำคัญ .01

6.3 ห้างสรรพสินค้าแห่งหนึ่งต้องการศึกษาถึงความสัมพันธ์ระหว่างยอดขายรายเดือนกับจำนวนพนักงานขายและค่าโฆษณา โดยมีข้อมูลดังนี้

เดือนที่	ยอดขาย (แสนบาท)	จำนวนพนักงานขาย (คน)	ค่าโฆษณา (1,000 บาท)
1	20	70	125
2	24	72	130
3	16	68	120
4	28	74	135
5	32	75	140
6	19	68	120
7	24	71	130
8	29	72	135
9	35	75	140

ก.จงหาสัมประสิทธิ์การตัดสินใจ พร้อมทั้งอธิบายความหมาย

ข. จากข้อมูลข้างต้น จะสรุปได้หรือไม่ว่า ยอดขายมีความสัมพันธ์กับจำนวนพนักงานขาย และค่าโฆษณาที่ระดับนัยสำคัญ .05

6.4 สมาคมโรงแรมของพัทยาเชื่อว่า การที่ห้องพักตามโรงแรมต่างๆ ในพัทยาจะว่างในช่วงสงกรานต์มากหรือน้อยขึ้นอยู่กับอากาศร้อนหรือไม่ และขึ้นอยู่กับจำนวนวันที่ฝนตก จึงเก็บข้อมูลในช่วงสงกรานต์ของ 15 ปี ที่ผ่านมาได้ข้อมูลดังนี้

จำนวนห้องว่าง	อุณหภูมิ (F)	จำนวนวันที่ฝนตก
6	75	1
12	68	4
11	67	3
8	73	2
13	65	3
9	71	4
6	73	1
8	74	1
4	76	2
10	73	3
8	69	3
9	71	4
9	66	2
12	64	4
2	73	5

ก. จงสร้างสมการความถดถอยแสดงความสัมพันธ์ระหว่างจำนวนห้องว่าง กับอุณหภูมิ และจำนวนวันที่ฝนตก พร้อมอธิบายความหมาย

ข. จงทดสอบความสัมพันธ์ในข้อ ก.

ค. จงทดสอบว่าเมื่ออุณหภูมิเพิ่มขึ้น จะทำให้ห้องว่างน้อยลง จริงหรือไม่ ที่ระดับนัยสำคัญ .05

6.5 เครื่องสำอางยี่ห้อหนึ่งใช้ระบบขายตรงในการจำหน่ายเครื่องสำอาง ผู้จัดการบริษัทเชื่อว่า ยอดขายเครื่องสำอางของเขาขึ้นอยู่กับค่าโฆษณาและจำนวนตัวแทนขาย จึงเก็บข้อมูลได้ดังนี้

ยอดขาย (ล้านบาท)	ค่าโฆษณา (1,000 บาท)	จำนวนตัวแทนขาย (คน)
32	249	15
47	292	18
18	183	14
25	201	16
49	310	21
41	248	20
52	246	18
38	241	14
36	288	13
29	191	15
43	248	21
28	210	18
24	256	20
36	275	16
41	241	19

ก.จงเขียนสมการความถดถอยแสดงความสัมพันธ์ระหว่างยอดขายกับค่าโฆษณาและจำนวนตัวแทนขาย พร้อมทั้งประมาณค่าสัมประสิทธิ์ความถดถอย

ข.จงทดสอบความสัมพันธ์ในข้อ ก. ที่ระดับนัยสำคัญ .05

ค. จงหาสัมประสิทธิ์การตัดสินใจ พร้อมอธิบาย

ง.จงหาสัมประสิทธิ์สหสัมพันธ์เชิงส่วน

6.6 นักสะสมของเก่าคนหนึ่งต้องการศึกษาว่าราคาของเก่าจะขายได้โดยการประมูลมีความสัมพันธ์กับอายุของเก่าและจำนวนผู้ที่เข้าประมูลหรือไม่ จึงเก็บข้อมูลของเก่าที่ขายได้ 32 ชิ้น ได้ข้อมูลดังนี้

ราคาขาย (1,000 บาท)	จำนวนผู้เข้าประมูล (คน)	อายุ (ปี)
1235	13	127
1080	12	115
845	7	127
1522	9	150
1047	6	156
1979	11	182
1822	12	156
1253	10	132
1297	9	137
946	9	113
1713	15	137
1024	11	117
1147	8	137
1092	6	153
1152	13	117
1326	10	126
2131	14	170
1550	8	182
1884	11	162
2041	10	184
854	6	143
1483	9	159
1055	14	108
1545	8	175
729	6	108
1792	9	179
1175	15	111
1593	8	187
785	7	111
1356	7	115
1262	5	194
	7	168

ก.จงเขียนสมการความถดถอยแสดงความสัมพันธ์ระหว่างราคาขายกับอายุ และจำนวนผู้เข้าร่วมประมูล

ข.จงทดสอบความเหมาะสมของสมการความถดถอยในข้อ ก.

ค. ถ้ามีของเก่าอายุ 100 ปี และมีผู้เข้าร่วมประมูล 10 คน จะขายของเก่าได้ในราคาเท่าใด