บทที่ 5

การวิเคราะห์การถดถอยแบบโพลิโนเมียล

โมเดลถดถอยเชิงเส้นตรงอย่างง่าย $\mathbf{Y} = \mathbf{X}\beta + \epsilon$ เป็นรูปแบบของโมเดลทั่วไปที่ใช้แสดงความสัมพันธ์เชิงเส้นตรงในเทอมของพารามิเตอร์ β ซึ่งรวมถึงรูปแบบโมเดลถดถอยแบบโพลิโนเมียล (Polynomial regression models) ด้วยเช่นกัน เช่น

โมเดลถดถอยแบบโพลิโนเมียลกำลังสองกรณีที่มีตัวแปรอิสระ 1 ตัว

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \epsilon_i, \qquad i = 1, 2, \ldots, n$$

โมเดลถดถอยแบบโพลิโนเมียลกำลังสองกรณีที่มีตัวแปรอิสระ 2 ตัว

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_{11} X_{i1}^2 + \beta_{22} X_{i2}^2 + \beta_{12} X_{i1} X_{i2} + \epsilon_i, \qquad i = 1, 2, \ldots, n$$

โดยโมเดลทั้งสองข้างต้นจัดเป็นโมเดลถดถอยเชิงเส้นตรงเช่นกัน

โมเดลถดถอยแบบโพลิโนเมียลนิยมใช้กันอย่างแพร่หลาย โดยเฉพาะเมื่อความสัมพันธ์มีลักษณะเชิงเส้นโค้ง (Curvilinear) หรือมีลักษณะที่ไม่ใช่เชิงเส้นตรง (Nonlinear) โมเดลถดถอยแบบโพลิโนเมียลอาจแสดงความสัมพันธ์ดังกล่าวได้อย่างเหมาะสมภายในช่วงของ X ที่แคบ ๆ นอกจากนี้การนำสมการถดถอยแบบ โพลิโนเมียลไปใช้ในการพยากรณ์นอกขอบเขตของข้อมูลที่พิจารณา ซึ่งเรียกว่า *Extrapolation* เป็นสิ่งที่ควรระมัดระวัง โดยเฉพาะเมื่อสมการถดถอยแบบโพลิโนเมียลมีกำลังสูง ๆ เนื่องจากสมการถดถอยแบบโพลิโนเมียลอาจใช้ในการพยากรณ์สำหรับข้อมูลที่พิจารณาได้ดี แต่อาจมีทิศทางที่แตกต่างไปจากเดิมเมื่อใช้พยากรณ์นอกขอบเขตของข้อมูลที่ใช้ในการสร้างสมการ ในบทนี้จะกล่าวถึงการสร้างสมการถดถอยแบบโพลิโนเมียลกำลังสองเท่านั้น ซึ่งสามารถประยุกต์ใช้กับสมการถดถอยแบบโพลิโนเมียลที่มีกำลังมากกว่าสองได้เช่นกัน

5.1 การวิเคราะห์การถดถอยแบบโพลิโนเมียลกรณีที่มีตัวแปรอิสระ 1 ตัว

พิจารณาโมเดลถดถอยแบบโพลิโนเมียลกรณีที่มีตัวแปรอิสระ 1 ตัวต่อไปนี้

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \epsilon, \tag{5.1}$$

โดยจะเรียกโมเดล (5.1) ว่า *โมเดลโพลิโนเมียลกำลังสองในหนึ่งตัวแปร* (Second-order polynomial model in one variable) ซึ่งมีค่าคาดหวังเป็น

$$E(Y \mid X) = \beta_0 + \beta_1 X + \beta_2 X^2 \tag{5.2}$$

เมื่อ

β_0 แทน ค่าเฉลี่ยของ Y เมื่อ $X = 0$ หากพิสัยของข้อมูลไม่คลุม 0 จะได้ว่า β_0 ไม่มีความหมาย

β_1 แทน พารามิเตอร์ของอิทธิพลเชิงเส้น (Linear effect)

β_2 แทน พารามิเตอร์ของอิทธิพลกำลังสอง (Quadratic effect)

ซึ่งบางครั้งอาจเรียกสมการ (5.2) ว่า *สมการกำลังสอง (Quadratic model)* เนื่องจากอยู่ในรูปของฟังก์ชันกำลังสอง โดยทั่วไปโมเดลโพลิโนเมียลกำลัง k ในหนึ่งตัวแปร มีรูปแบบดังนี้

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \cdot + \beta_k X^k + \epsilon, \qquad i = 1, 2, \ldots, n \tag{5.3}$$

ถ้าให้ $X_1 = X$, $X_2 = X^2$, $X_3 = X^3$, $\ldots$, $X_k = X^k$ สามารถเขียนสมการ (5.3) ใหม่ได้เป็น

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdot + \beta_k X_k + \epsilon, \qquad i = 1, 2, \ldots, n \tag{5.4}$$

ซึ่งก็คือรูปแบบของสมการถดถอยเชิงเส้นตรงแบบพหุที่ประกอบด้วยตัวแปรอิสระ k ตัวนั่นเอง ดังนั้นการสร้างสมการถดถอยแบบโพลิโนเมียลยังคงใช้เทคนิคเดียวกับการสร้างสมการถดถอยเชิงเส้นตรงแบบพหุนั่นเอง

หลังจากประมาณพารามิเตอร์ในสมการ (5.2) ด้วยวิธี OLS จะได้สมการถดถอยแบบโพลิโนเมียลกำลังสองจากข้อมูลตัวอย่าง ดังนี้

$$\hat{Y} = b_0 + b_1 X + b_2 X^2 \tag{5.5}$$

เนื่องจาก X และ X^2 มีความสัมพันธ์กันค่อนข้างสูง จึงก่อให้เกิดความยุ่งยากต่อการคำนวณเมตริกซ์ $(\mathbf{X'X})^{-1}$ ในสมการปกติเพื่อใช้ในการประมาณค่าสัมประสิทธิ์ถดถอย ดังนั้นจึงนิยม Center ตัวแปรอิสระ

ก่อนที่จะประมาณค่าสัมประสิทธิ์ถดถอย เพื่อลดปัญหา Multicollinearity และได้โมเดลถดถอยแบบโพลิโน-เมียลกำลังสองของข้อมูลที่ Center แล้วดังนี้

$$Y = \beta_0' + \beta_1' x + \beta_2' x^2 + \epsilon \tag{5.6}$$

เมื่อ $x = X - \bar{X}$

ทำการประมาณพารามิเตอร์ในสมการ (5.6) ด้วยวิธีกำลังสองน้อยที่สุดได้ดังนี้

$$\mathbf{X'Xb'} = \mathbf{X'Y} \tag{5.7}$$

โดยที่ $\underset{(n\times1)}{\mathbf{Y}} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$, $\underset{(n\times3)}{\mathbf{X}} = \begin{bmatrix} 1 & (X_{11}-\bar{X}_1) & (X_{12}-\bar{X}_2) \\ 1 & (X_{21}-\bar{X}_1) & (X_{22}-\bar{X}_2) \\ 1 & (X_{31}-\bar{X}_1) & (X_{32}-\bar{X}_2) \\ \vdots & \vdots & \vdots \\ 1 & (X_{n1}-\bar{X}_1) & (X_{n2}-\bar{X}_2) \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ 1 & x_{31} & x_{32} \\ \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} \end{bmatrix}$,

$$\underset{(3\times1)}{\mathbf{b}'} = \begin{bmatrix} b_0' \\ b_1' \\ b_2' \end{bmatrix}$$

และได้สมการถดถอยแบบโพลิโนเมียลกำลังสองของข้อมูลที่ Center แล้วดังนี้

$$\hat{Y} = b_0' + b_1' x + b_2' x^2 \tag{5.8}$$

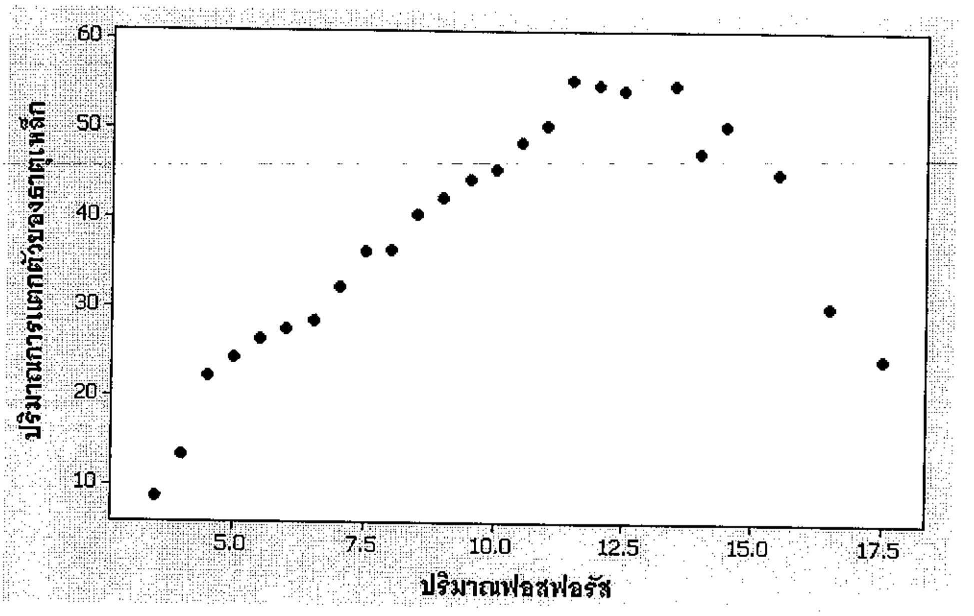
จากหลักพีชคณิตจะได้ว่า $b_2' = b_2$ ดังนั้นสมการ (5.5) และสมการ (5.8) ให้ค่าพยากรณ์ $\hat{Y}$ ค่าเดียวกัน

X	3.5	4.0	4.5	5.0	5.5	6.0	6.5	7.0	7.5	8.0	8.5	9.0	9.5
Y	8.63	13.43	22.33	24.33	26.33	27.53	28.43	32.33	36.13	36.33	40.43	42.23	44.33
X	10.0	10.5	11.0	11.5	12.0	12.5	13.5	14.0	14.5	15.5	16.5	17.5	
Y	45.53	48.43	50.33	55.43	54.83	54.33	54.83	47.33	50.33	45.13	30.13	24.23	

ตารางที่ 5.1: ข้อมูลปริมาณฟอสฟอรัสและปริมาณธาตุเหล็กที่แตกตัว

ตัวอย่างที่ 5.1 จากการศึกษาพบว่าการสกัดธาตุเหล็ก (Iron) ออกจากน้ำมันพืชจะช่วยยืดอายุการใช้งานของ

น้ำมันพืช ซึ่งวิธีที่ใช้เพื่อลดปริมาณธาตุเหล็กทำได้โดยเติมกรดฟอสฟอริก (X) ลงในสารละลายน้ำ โดยจะทำให้ธาตุเหล็ก (Y) แตกตัว ข้อมูลแสดงในตารางที่ 5.1



รูปที่ 5.1: แผนภาพการกระจายระหว่างปริมาณฟอสฟอรัสและปริมาณธาตุเหล็กที่แตกตัว

วิธีทำ พิจารณาแผนภาพการกระจายในรูปที่ 5.1 พบว่าความสัมพันธ์มีลักษณะเป็นเส้นโค้ง ซึ่งชี้ว่าโมเดลโพลิโนเมียลกำลังสองอาจแสดงความสัมพันธ์ระหว่างปริมาณฟอสฟอรัสและปริมาณธาตุเหล็กที่แตกตัวได้อย่างเหมาะสม เพื่อลดปัญหา Multicollinearity ทำการ Center ค่าของตัวแปรอิสระ เมื่อ $\bar{X} = 9.74$ ดังแสดงในตารางที่ 5.2 ซึ่งการสร้างสมการโพลิโนเมียลกำลังสองของข้อมูลที่ Center แล้ว จะเหมือนกับการสร้างสมการถดถอยเชิงเส้นตรงแบบพหุเมื่อมีตัวแปรอิสระ 2 ตัวนั่นเอง โดยสามารถเขียนโมเดลถดถอยเชิงเส้นตรงแบบพหุของตัวแปรที่ Center แล้วได้ดังนี้

$$\hat{Y} = \beta_0' + \beta_1' x_1 + \beta_2' x_2 + \epsilon$$

เมื่อ

$x_1 = X - 9.74$

$x_2 = (X - 9.74)^2$

X	Y	$x_1 = X - 9.74$	$x_2 = (X - 9.74)^2$
3.5	8.63	-6.24	38.9376
4.0	13.43	-5.74	32.9476
4.5	22.33	-5.24	27.4576
5.0	24.33	-4.74	22.4676
5.5	26.33	-4.24	17.9776
6.0	27.53	-3.74	13.9876
6.5	28.43	-3.24	10.4976
7.0	32.33	-2.74	7.5076
7.5	36.13	-2.24	5.0176
8.0	36.33	-1.74	3.0276
8.5	40.43	-1.24	1.5376
9.0	42.23	-0.74	0.5476
9.5	44.33	-0.24	0.0576
10.0	45.53	0.26	0.0676
10.5	48.43	0.76	0.5776
11.0	50.33	1.26	1.5876
11.5	55.43	1.76	3.0976
12.0	54.83	2.26	5.1076
12.5	54.33	2.76	7.6176
13.5	54.83	3.76	14.1376
14.0	47.33	4.26	18.1476
14.5	50.33	4.76	22.6576
15.5	45.13	5.76	33.1776
16.5	30.13	6.76	45.6976
17.5	24.23	7.76	60.2176

ตารางที่ 5.2: การคำนวณข้อมูลที่ Center แล้ว สำหรับสมการโพลิโนเมียลกำลังสอง

จากนั้นประมาณค่าสัมประสิทธิ์ถดถอยในโมเดลข้างต้นด้วยวิธีกำลังสองน้อยที่สุดได้เป็น

$$\mathbf{b}' = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \begin{bmatrix} 47.728 \\ 2.5803 \\ 0.6333 \end{bmatrix}$$

ดังนั้นได้สมการถดถอยแบบโพลิโนเมียลกำลังสองดังนี้

$$\hat{Y} = 47.728 + 2.5803x_1 + 0.6333x_2$$

$$\hat{Y} = \underset{(1.145)}{47.728} + \underset{(0.2076)}{2.5803}(X - 9.74) - \underset{(0.05181)}{0.6333}(X - 9.74)^2$$

โดยที่ $R^2 = 0.918$

การทดสอบนัยสำคัญของโมเดลถดถอยแบบโพลิโนเมียลกำลังสองดังกล่าวทำได้โดยกำหนดสมมติฐานของการ

ทดสอบเป็น

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

$$H_1: \ \beta_i \neq 0 \quad \text{อย่างน้อย 1 ตัว} \qquad (i = 1, 2)$$

สร้างตารางวิเคราะห์ความแปรปรวนได้ดังนี้

ANOVA

Source of variation	df	SS	MS	F
Regression	2	3978.8	1989.4	123.65
Error	22	354	16.1	
Total	24	4332.7		

ค่าวิกฤติคือ $F_{2,22\,(.05)} = 3.44$
ดังนั้นปฏิเสธ H_0 เนื่องจาก F ที่คำนวณได้มีค่ามากกว่าค่าวิกฤติ นั่นคือเทอมกำลังหนึ่ง หรือเทอมกำลังสอง หรือทั้งสองเทอมมีอิทธิพลต่อโมเดล ที่ระดับนัยสำคัญ 0.05

จากการตรวจสอบความเหมาะสมของโมเดลข้างต้น โดยพิจารณารูปที่ 5.2 พบว่าข้อมูลมีลักษณะที่สอดคล้องกับข้อตกลงเบื้องต้นของการวิเคราะห์โดยประมาณ ถึงแม้ว่าความคลาดเคลื่อนจะมีการแจกแจงที่มีหางหนากว่าการแจกแจงแบบปกติเล็กน้อย

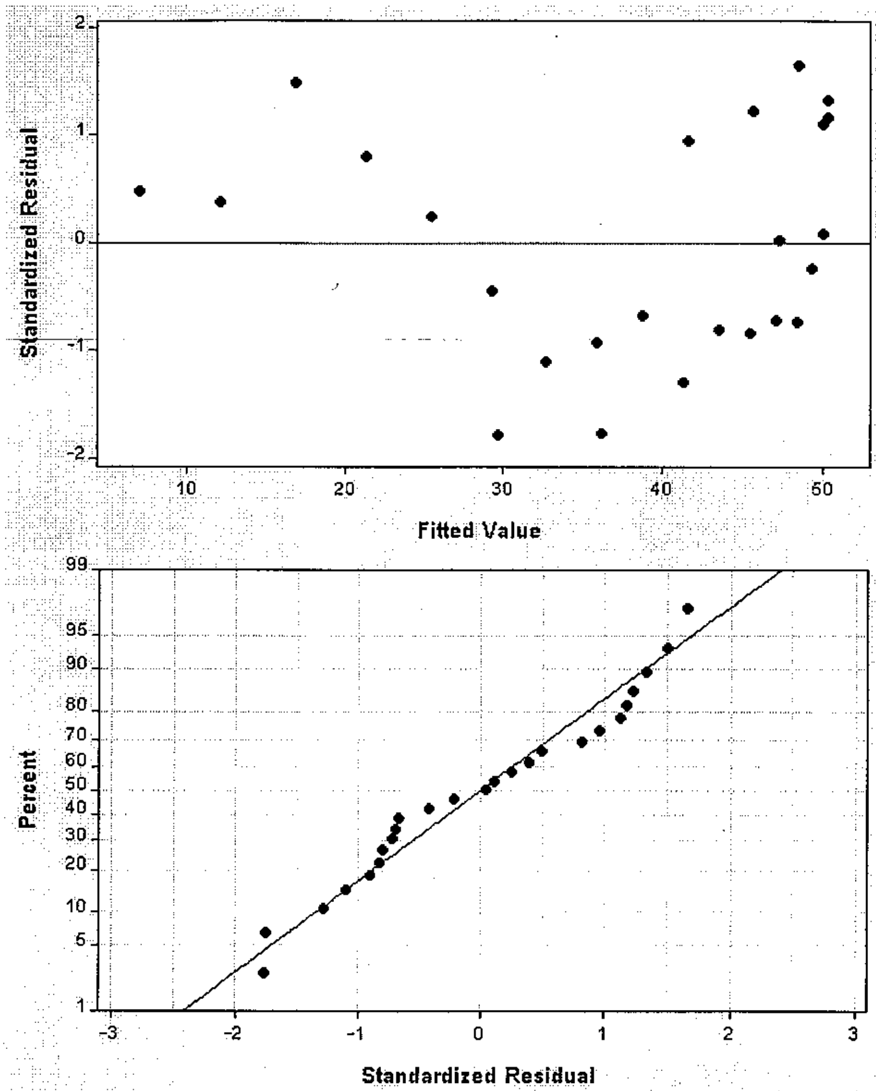
5.2 การตรวจสอบกำลังที่เหมาะสมของสมการถดถอยแบบโพลิโนเมียล

การเลือกสมการถดถอยแบบโพลิโนเมียลที่เหมาะสมมีหลักการดังนี้ พยายามเลือกสมการที่มีกำลังสูงสุดต่ำที่สุดเท่าที่จะทำได้ แต่ยังคงสามารถอธิบายความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรอิสระได้ดี ซึ่งวิธีการตรวจสอบเพื่อค้นหากำลัง ที่เหมาะสมของสมการถดถอยแบบโพลิโนเมียลทำได้หลายวิธี โดยอาจพิจารณาจากแผนภาพการกระจายระหว่างตัวแปรตามและตัวแปรอิสระแต่ละตัว หรือทดสอบสมมติฐานเกี่ยวกับค่าสัมประสิทธิ์ถดถอยของตัวแปรที่มีกำลังสูงสุดโดยใช้ t-test หรือ Partial F-test ก็ได้

พิจารณาโมเดลถดถอยแบบโพลิโนเมียลกำลังสองของข้อมูลที่มีการ Center แล้ว
Full model:

$$Y = \beta_0' + \beta_1' x_1 + \beta_2' x_2 + \epsilon \tag{5.9}$$

หากต้องการตรวจสอบเทอมกำลังสองว่ามีส่วนช่วยอธิบายลักษณะความสัมพันธ์ระหว่างตัวแปรหรือไม่ ทำได้โดย



รูปที่ 5.2: กราฟระหว่างค่า Standardized residual และค่าพยากรณ์ (บน) และ Normal probability plot ของ Standardized residuals ในโมเดลของตัวอย่างที่ 5.1 (ล่าง)

กำหนดสมมติฐานเพื่อทดสอบนัยสำคัญของเทอมกำลังสองเป็น

$$H_0 : \beta'_2 = 0 \quad vs. \quad H_1 : \beta'_2 \neq 0$$

การทดสอบสมมติฐานทำได้โดยใช้สถิติทดสอบ t

$$t_c = \frac{b'_2}{S_{b'_2}} \tag{5.10}$$

โดยจะปฏิเสธ H_0 เมื่อ $t_c > t_{\alpha/2, error\,df}$

หรือใช้วิธี Extra-sum-of-squares โดยถ้า H_0 เป็นจริง โมเดล (5.10) จะลดรูปเป็น
Reduced model:

$$Y = \beta_0' + \beta_1' x_1 + \epsilon \tag{5.11}$$

การทดสอบสมมติฐานทำได้โดยใช้ Partial F-test ดังนี้

$$F_c = \frac{SSR(x_2 \mid x_1)/1}{MSE_{Full}} \tag{5.12}$$

โดยจะปฏิเสธ H_0 เมื่อ $F_c > F_{1, error\, df\, (\alpha)}$

ตัวอย่างที่ 5.2 จากข้อมูลในตัวอย่างที่ 5.1 จงตรวจสอบว่าเทอมกำลังสองควรอยู่ในโมเดลหรือไม่ กำหนดระดับนัยสำคัญของการทดสอบเป็น 0.05
วิธีทำ จากโมเดล (5.9) กำหนดสมมติฐานของการทดสอบได้ดังนี้

$$H_0 : \beta_2' = 0 \quad vs. \quad H_1 : \beta_2' \neq 0$$

ระดับนัยสำคัญ 0.05
คำนวณสถิติทดสอบ:

$$t_c = \frac{b_2'}{S_{b_2'}} = \frac{-0.6333}{0.0518} = -12.22$$

บริเวณวิกฤติ: $t_c > t_{0.025, 22} = 2.074$
สรุปผล: ปฏิเสธ H_0 นั่นคือ เมื่อให้เทอมกำลังหนึ่งอยู่ในโมเดลแล้ว เทอมกำลังสองมีอิทธิพลต่อโมเดลอย่างมีนัยสำคัญ ที่ระดับ 0.05 ดังนั้นเทอมกำลังสองควรอยู่ในโมเดลด้วย

หรืออาจทดสอบสมมติฐานข้างต้นโดยใช้ Partial F-test ทำได้โดยสร้างสมการถดถอยแบบโพลิโนเมียลที่ไม่มีเทอมกำลังสองได้ดังนี้

$$\hat{Y} = \underset{(2.190)}{37.746} + \underset{(0.5516)}{1.9991} x_1$$

โดยที่ $R^2 = 0.363, \quad SSR(x_1) = 1574.8$ และ $SSE(x_1) = 2758.0$

คำนวณ Extra-sum-of-squares ได้เป็น

$$
\begin{aligned}
SSR(x_2 \mid x_1) &= SSR(x_1, x_2) - SSR(x_1) \\
&= 3978.8 - 1574.8 \\
&= 2404
\end{aligned}
$$

สร้างตารางวิเคราะห์ความแปรปรวนได้ดังนี้

ANOVA

Source of variation	df	SS	MS	F
Regression	2	3978.8	1989.4	123.65
x_1	1	1574.8	1574.8	
$x_2 \mid x_1$	1	2404.0	2404.0	149.32
Error	22	354.0	16.1	
Total	24	4332.7		

บริเวณวิกฤติ: $F_c > F_{1,22\,(.05)} = 4.30$

ดังนั้นปฏิเสธ H_0 เนื่องจาก $F_c = 149.32$ มีค่ามากกว่าค่าวิกฤติ ซึ่งสอดคล้องกับผลที่ได้จากการใช้สถิติทดสอบ t

หมายเหตุ การทดสอบสมมติฐาน $H_0 : \beta_2' = 0 \quad vs. \quad H_1 : \beta_2' \neq 0$ จะเหมือนกับการทดสอบสมมติฐาน

$$
\begin{aligned}
H_0 : &\quad E(\mathbf{Y} \mid \mathbf{X}) = \beta_0' + \beta_1' x_1 \\
H_1 : &\quad E(\mathbf{Y} \mid \mathbf{X}) = \beta_0' + \beta_1' x_1 + \beta_2 x_2
\end{aligned}
$$

หากต้องการตรวจสอบว่าเทอมกำลังหนึ่งควรอยู่ในโมเดลหรือไม่ สามารถทำได้เช่นเดียวกัน โดยกำหนดสมมติฐานของการทดสอบเป็น $H_0 : \beta_1' = 0 \quad vs. \quad H_1 : \beta_1' \neq 0$ ซึ่งการทดสอบดังกล่าวสามารถทำได้โดยใช้ t-test หรือ Partial F-test

ตัวอย่างที่ 5.3 จากข้อมูลในตัวอย่างที่ 5.1 จงตรวจสอบว่าเทอมกำลังหนึ่งควรอยู่ในโมเดลหรือไม่ กำหนดระดับนัยสำคัญของการทดสอบเป็น 0.05

วิธีทำ กำหนดสมมติฐานของการทดสอบเป็น

$$H_0 : \beta_1' = 0 \quad vs. \quad H_1 : \beta_1' \neq 0$$

ระดับนัยสำคัญ 0.05

คำนวณสถิติทดสอบ:

$$t_c = \frac{b_1'}{S_{b_1'}} = \frac{2.5803}{0.2076} = 12.43$$

บริเวณวิกฤติ: $t_c > t_{0.025,22} = 2.074$

สรุปผล: ปฏิเสธ H_0 นั่นคือ เมื่อให้เทอมกำลังสองอยู่ในโมเดลแล้ว เทอมกำลังหนึ่งมีอิทธิพลต่อโมเดล ที่ระดับนัยสำคัญ 0.05

หรือหากทดสอบสมมติฐานข้างต้นโดยใช้ Partial F-test ทำได้โดยสร้างสมการถดถอยแบบโพลิโนเมียลที่ไม่มีเทอมกำลังหนึ่งได้ดังนี้

$$\hat{Y} = \underset{(3.129)}{45.403} - \underset{(0.1397)}{0.4858}x_2$$

โดยที่ $R^2 = 0.345$, $SSR(x_2) = 1492.8$ และ $SSE(x_2) = 2840.0$

คำนวณ Extra-sum-of-squares ได้เป็น

$$\begin{aligned} SSR(x_1 \mid x_2) &= SSR(x_1, x_2) - SSR(x_2) \\ &= 3978.8 - 1492.8 \\ &= 2486.0 \end{aligned}$$

สร้างตารางวิเคราะห์ความแปรปรวนได้ดังนี้

ANOVA

Source of variation	df	SS	MS	F
Regression	2	3978.8	1989.4	123.65
x_2	1	1492.8	1574.8	
$x_1 \mid x_2$	1	2486.0	2486.0	154.41
Error	22	354.0	16.1	
Total	24	4332.7		

บริเวณวิกฤติ: $F_c > F_{1,22\,(.05)} = 4.30$

ดังนั้นปฏิเสธ H_0 เนื่องจาก $F_c = 154.41$ มีค่ามากกว่าค่าวิกฤติ ซึ่งให้ผลสรุปที่สอดคล้องกับการใช้สถิติทดสอบ t

การใช้สมการถดถอยแบบโพลิโนเมียลกำลัง k เมื่อ $k > 2$ ควรทำด้วยความระมัดระวัง เพราะนอกจากจะมีความยุ่งยากต่อการแปลผลแล้ว การใช้สมการเพื่อการพยากรณ์ทั้งภายในและภายนอกขอบเขตของข้อมูลที่ใช้ในการสร้างสมการอาจผิดพลาดสูง ถึงแม้ว่าสมการถดถอยแบบโพลิโนเมียลที่มีกำลังสูง ๆ มักเหมาะสมกับข้อมูลที่ใช้ในการสร้างสมการเป็นอย่างดี แต่อาจจะไม่ได้แสดงความสัมพันธ์ที่ชัดเจนระหว่างตัวแปรตามและตัวแปรอิสระ

โมเดลถดถอยแบบโพลิโนเมียลมีประโยชน์ในกรณีที่ทราบลักษณะความสัมพันธ์ที่แท้จริงว่าเป็นเส้นโค้ง หรือในกรณีที่ไม่ทราบความสัมพันธ์ที่แท้จริง อาจใช้ประมาณความสัมพันธ์ดังกล่าวได้ โดยเฉพาะเมื่อรูปแบบของความสัมพันธ์ไม่ใช่เชิงเส้นตรงและมีรูปแบบที่ซับซ้อน สิ่งที่ควรพิจารณาสำหรับการสร้างสมการถดถอยแบบโพลิโนเมียลในหนึ่งตัวแปรมีดังนี้

- พยายามหลีกเลี่ยงสมการโพลิโนเมียลที่มีกำลังสูง ๆ ($k > 2$) ควรสร้างสมการโพลิโนเมียลให้มีกำลังต่ำเท่าที่เป็นไปได้ ถึงแม้ว่าสมการโพลิโนเมียลที่มีกำลังสูง ๆ มักจะใช้ในการพยากรณ์ได้ดี แต่อาจไม่สามารถอธิบายรูปแบบฟังก์ชันที่ไม่ทราบค่าได้ ซึ่งโดยทั่วไปมักจะจำกัดความสนใจอยู่ที่สมการโพลิโนเมียลกำลังหนึ่งหรือกำลังสองเท่านั้น

- พิจารณาวิธีในการเลือกกำลังที่เหมาะสมของสมการโพลิโนเมียลหลาย ๆ วิธี เช่น วิธี Forward selection ซึ่งได้จากการสร้างสมการโพลิโนเมียลตามลำดับขั้น โดยเพิ่มกำลังของสมการทีละขั้น จนกระทั่งสถิติทดสอบทีที่ใช้ในการทดสอบอิทธิพลของเทอมที่มีกำลังสูงสุดไม่มีนัยสำคัญทางสถิติ หรือวิธี Backward elimination ซึ่งได้จากการสร้างสมการโพลิโนเมียลที่มีกำลังสูงสุดที่เหมาะสม จากนั้นตัดตัวแปรออกไปทีละตัว เริ่มจากตัวที่มีกำลังสูงสุด จนกระทั่งเทอมที่มีกำลังสูงสุดมีนัยสำคัญทางสถิติ แต่อย่างไรก็ตามวิธีในการเลือกสมการโพลิโนเมียลที่เหมาะสมแต่ละวิธีอาจนำไปสู่สมการสุดท้ายที่แตกต่างกัน ซึ่งการเลือกสมการควรคำนึงถึงรูปแบบที่ง่ายต่อการอธิบายด้วย

- การนำโมเดลถดถอยแบบโพลิโนเมียลไปใช้ในการพยากรณ์นอกขอบเขตของข้อมูลที่ใช้ในการสร้างสมการ เป็นสิ่งที่ควรพึงระวัง เนื่องจากโมเดลถดถอยแบบโพลิโนเมียลอาจเปลี่ยนทิศทางหรือเปลี่ยนรูปแบบ

- เมื่อกำลังของโมเดลถดถอยแบบโพลิโนเมียลเพิ่มขึ้น เมตริกซ์ $\mathbf{X}'\mathbf{X}$ จะมีค่าใกล้เคียง Singular matrix มากขึ้น ทำให้การคำนวณ $(\mathbf{X}'\mathbf{X})^{-1}$ ไม่ถูกต้อง และส่งผลให้การประมาณพารามิเตอร์มีความคลาดเคลื่อนเกิดขึ้น

- ถ้าพิสัยของ X อยู่ในช่วงแคบ ๆ อาจก่อให้เกิดปัญหา Multicollinearity ระหว่างคอลัมน์ของเมตริกซ์ $\mathbf{X}$ เช่น ถ้า $1 \leq X \leq 3$ จะได้ว่า $1 \leq X^2 \leq 9$ ซึ่งทำให้เกิดปัญหา Multicollinearity ที่ค่อนข้างรุนแรงระหว่างตัวแปร X และ X^2

5.3 การวิเคราะห์การถดถอยแบบโพลิโนเมียลกรณีที่มีตัวแปรอิสระ 2 ตัว

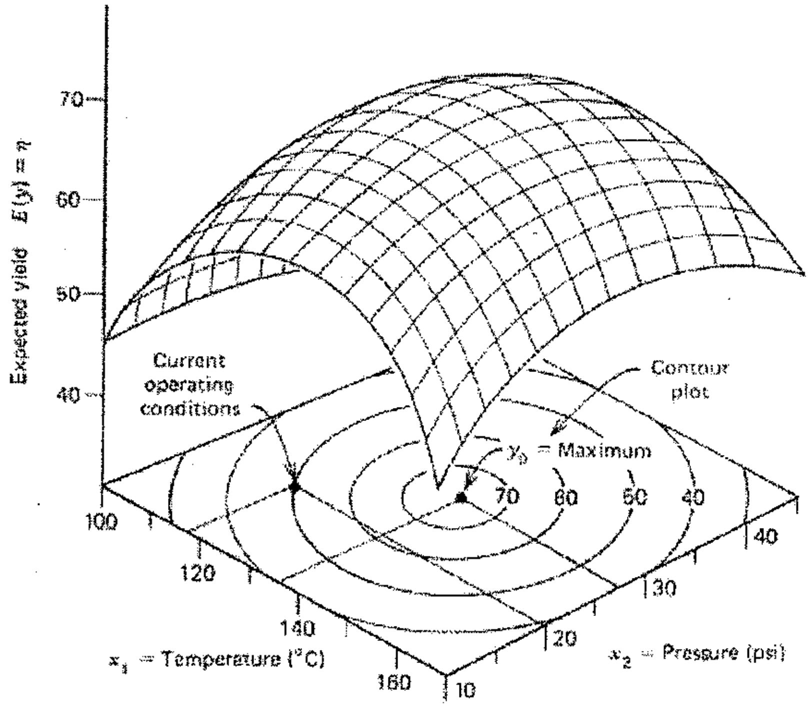
โมเดลถดถอยแบบโพลิโนเมียลกรณีที่มีตัวแปรอิสระ 2 ตัว ของข้อมูลที่ Center แล้ว มีรูปแบบดังนี้

$$Y = \beta_0' + \beta_1' x_1 + \beta_2' x_2 + \beta_{11}' x_1^2 + \beta_{22}' x_2^2 + \beta_{12}' x_1 x_2 + \epsilon \qquad (5.13)$$

โดยที่ $x_1 = X_1 - \bar{X}_1$ และ $x_2 = X_2 - \bar{X}_2$

จะเรียกโมเดล (5.13) ว่า *โมเดลโพลิโนเมียลกำลังสองในสองตัวแปร* (Second-order polynomial model in two variables) ซึ่งมีค่าคาดหวังเป็น

$$E(\mathbf{Y} \mid \mathbf{X}) = \beta_0' + \beta_1' x_1 + \beta_2' x_2 + \beta_{11}' x_1^2 + \beta_{22}' x_2^2 + \beta_{12}' x_1 x_2 \tag{5.14}$$



รูปที่ 5.3: ตัวอย่างของ Quadratic response surface (จาก Neter et al. (1999))

เมื่อ

β_0' แทน ค่าเฉลี่ยของ Y เมื่อ $x_1 = 0$ และ $x_2 = 0$

β_1', β_2' แทน พารามิเตอร์ของอิทธิพลเชิงเส้น (Linear effect) ของตัวแปรอิสระ x_1 และ x_2 ตามลำดับ

β_{11}', β_{22}' แทน พารามิเตอร์ของอิทธิพลกำลังสอง (Quadratic effect) ของตัวแปรอิสระ x_1 และ x_2 ตามลำดับ

β_{12}' แทน พารามิเตอร์ของอิทธิพลร่วม (Interaction effect) ระหว่างตัวแปรอิสระ x_1 และ x_2

และนิยมเรียกฟังก์ชัน (5.14) ว่า *Response surface* รูปที่ 5.3 แสดง Response surface และ Contour curve ของฟังก์ชันโพลิโนเมียลกำลังสองใน 2 ตัวแปร ซึ่งจะเห็นถึงระดับของตัวแปรตามที่แตกต่างกัน เนื่องมาจากค่าของตัวแปรอิสระ 2 ตัวที่เกิดร่วมกัน

หมายเหตุ เทอม $\beta'_{12}x_1x_2$ จะถือว่าเป็นเทอมกำลังสองเช่นเดียวกับ $\beta'_{11}x_1^2$ และ $\beta'_{22}x_2^2$ เนื่องจาก

$$\beta'_{11}x_1^2 = \beta'_{11}x_1x_1 \quad \text{และ} \quad \beta'_{22}x_2^2 = \beta'_{22}x_2x_2$$

เมื่อมีอิทธิพลร่วมอยู่ในโมเดล ค่าสัมประสิทธิ์ถดถอย β'_1 และ β'_2 จะมีความหมายที่แตกต่างไปจากเดิม โดย β'_1 และ β'_2 ไม่ได้หมายถึงการเปลี่ยนแปลงของตัวแปรตามเฉลี่ย เมื่อตัวแปรอิสระเพิ่มขึ้นหนึ่งหน่วย ขณะที่ตัวแปรอิสระอื่นมีค่าคงที่อีกต่อไป ซึ่งในที่นี้ถ้า x_1 เพิ่มขึ้นหนึ่งหน่วย เมื่อให้ x_2 มีค่าคงที่แล้ว ค่าของตัวแปรตามเฉลี่ยจะเปลี่ยนไป $\beta'_1 + \beta'_3x_2$ หน่วย ทำนองเดียวกันถ้า x_2 เพิ่มขึ้นหนึ่งหน่วย เมื่อให้ x_1 มีค่าคงที่แล้ว ตัวแปรตามเฉลี่ยจะเปลี่ยนไป $\beta'_2 + \beta'_3x_1$ หน่วย จะเห็นได้ว่าจากโมเดล (5.13) อิทธิพลของ x_1 ในแต่ละระดับของ x_2 และอิทธิพลของ x_2 ในแต่ละระดับของ x_1 จะขึ้นอยู่กับระดับของตัวแปรอื่น

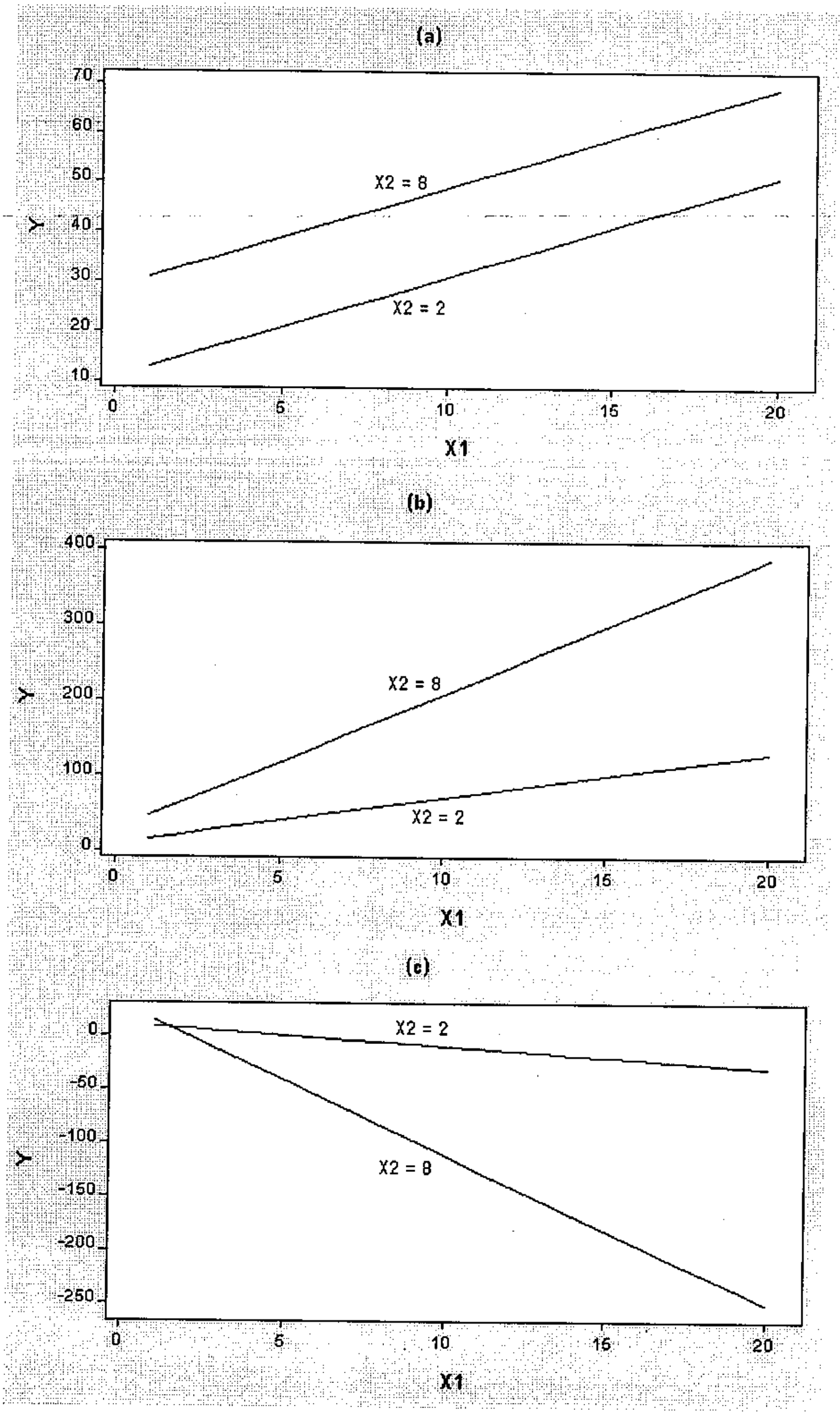
รูปที่ 5.4 แสดงผลของ X_1 ที่มีต่อ $E(Y)$ เมื่อกำหนดระดับของ X_2 ที่แตกต่างกัน ซึ่งเรียกว่า *Conditional effects plot* โดยรูปที่ 5.4 (a) แสดงค่าของ $E(Y) = 5+2X_1+3X_2$ เมื่อ $X_2 = 2$ และ $X_2 = 8$ ซึ่งจะเห็นได้ว่า X_1 และ X_2 มีความสัมพันธ์เชิงบวกต่อกันเท่านั้น จะเรียกโมเดลลักษณะนี้ว่า *Additive model* จากรูปจะเห็นได้ว่าฟังก์ชันทั้งสองขนานกัน เนื่องจากเมื่อ X_1 เพิ่มขึ้น 1 หน่วย อัตราการเปลี่ยนแปลงของตัวแปรตามเฉลี่ยมีค่าเท่ากับ $\beta_1 = 2$ หน่วยเท่ากัน ไม่ว่า $X_2 = 2$ หรือ $X_2 = 8$ รูปที่ 5.4 (b) แสดงค่าของ $E(Y) = 5+2X_1+3X_2+2X_1X_2$ เมื่อ $X_2 = 2$ และ $X_2 = 8$ จะเห็นได้ว่าเมื่อ X_1 เพิ่มขึ้น 1 หน่วย จะส่งผลกระทบต่อตัวแปรตามที่ $X_2 = 8$ มากกว่าที่ $X_2 = 2$ นั่นคือ ความชันระหว่าง $E(Y)$ กับ X_1 มีค่าเพิ่มขึ้น เมื่อ X_2 มีค่าสูงขึ้น (β_3 มีค่าเป็นบวก) จะเรียกอิทธิพลร่วมระหว่างตัวแปร เชิงปริมาณ 2 ตัวนี้ว่า *Reinforcement or synergistic type* รูปที่ 5.4 (c) แสดงค่าของ $E(Y) = 5+2X_1+3X_2-2X_1X_2$ เมื่อ $X_2 = 2$ และ $X_2 = 8$ จะเห็นได้ว่าความชันระหว่าง $E(Y)$ กับ X_1 มีค่าลดลง เมื่อ X_2 มีค่าสูงขึ้น (β_3 มีค่าเป็นลบ) จะเรียกอิทธิพลร่วมระหว่างตัวแปรเชิงปริมาณ 2 ตัวนี้ว่า *Interference or antagonistic type*

หมายเหตุ

1 จากฟังก์ชัน $E(Y) = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3$

1.1 หาก $\beta_1, \beta_2 > 0$ แล้ว ถ้า $\beta_3 > 0$ จะได้อิทธิพลร่วมแบบ Reinforcement และถ้า $\beta_3 < 0$ จะได้อิทธิพลร่วมแบบ Interference

1.2 หาก $\beta_1, \beta_2 < 0$ แล้ว ถ้า $\beta_3 > 0$ จะได้อิทธิพลร่วมแบบ Interference และถ้า $\beta_3 < 0$ จะได้อิทธิพลร่วมแบบ Reinforcement



รูปที่ 5.4: (a) Additive model (b) Reinforcement interaction effect (c) Interference interaction effect

2 การหาอนุพันธ์ของ $E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$ เทียบกับ X_1 และ X_2 จะได้ว่า

$$\frac{\partial E(Y)}{\partial X_1} = \beta_1 + \beta_3 X_2 \tag{5.15}$$

$$\frac{\partial E(Y)}{\partial X_2} = \beta_2 + \beta_3 X_2 \tag{5.16}$$

ซึ่งแทนอัตราการเปลี่ยนแปลงของตัวแปรตามเฉลี่ย เมื่อ X_1 เพิ่มขึ้น 1 หน่วย และ X_2 คงที่ และอัตราการเปลี่ยนแปลงของตัวแปรตามเฉลี่ย เมื่อ X_2 เพิ่มขึ้น 1 หน่วย และ X_1 คงที่ ตามลำดับ

พิจารณา *โมเดลโพลิโนเมียลกำลังสองในสามตัวแปร* (Second-order polynomial model in three variables) ต่อไปนี้

$$Y = \beta'_0 + \beta'_1 x_1 + \beta'_2 x_2 + \beta'_3 x_3 + \beta'_{11} x_1^2 + \beta'_{22} x_2^2 + \beta'_{33} x_3^2 + \beta'_{12} x_1 x_2 + \beta'_{13} x_1 x_3 + \beta'_{23} x_2 x_3 + \epsilon$$

เมื่อ $x_i = X_i - \bar{X}_i,\ i = 1, 2, 3$ โดยมีค่าคาดหวังเป็น

$$E(\mathbf{Y} \mid \mathbf{X}) = \beta'_0 + \beta'_1 x_1 + \beta'_2 x_2 + \beta'_3 x_3 + \beta'_{11} x_1^2 + \beta'_{22} x_2^2 + \beta'_{33} x_3^2 + \beta'_{12} x_1 x_2 + \beta'_{13} x_1 x_3 + \beta'_{23} x_2 x_3$$

เมื่อ β'_{12}, β'_{13} และ β'_{23} แทน อิทธิพลร่วมระหว่างตัวแปรอิสระแต่ละคู่

ซึ่งการประมาณค่าพารามิเตอร์ของสมการข้างต้น จะเหมือนกับการประมาณค่าพารามิเตอร์ในสมการถดถอยเชิงเส้นตรงแบบพหุเมื่อมีตัวแปรอิสระ 9 ตัวนั่นเอง

ตัวอย่างที่ 5.4 ข้อมูลต่อไปนี้เป็นผลที่ได้จากการทดลองเพื่อศึกษาอิทธิพลของอุณหภูมิและระยะเวลาที่เหมาะสมต่อการทำ Popcorn โดยให้ X_1 แทน อุณหภูมิของเตาอบ X_2 แทน ระยะเวลาที่ใช้ และ Y แทน จำนวน Popcorn ที่ไม่สามารถรับประทานได้ (ไม่สุกหรือไหม้) เพื่อที่จะลดความผันแปรของการทดลอง ผู้ทดลองใช้เตาอบเตาเดียวและ Popcorn ถุงเดียวกัน โดยใช้ปริมาณ Popcorn เท่า ๆ กันในแต่ละซ้ำของการทดลอง เพื่อขจัดปัญหาความร้อนของเตาอบที่ยังตกค้างอยู่ จึงทำการทดลองหนึ่งครั้งในแต่ละวัน และผู้ทดลองคนเดียวกันประเมินผลของการทดลองในแต่ละครั้ง ได้ข้อมูลดังนี้

การทดลองที่	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X_1	6	4	6	6	5	5	4	5	4	5	4	6	5	5	5
X_2	105	120	120	105	120	105	90	120	105	90	105	90	105	105	90
Y	29	33	45	47	16	21	131	39	37	37	39	22	35	22	55

ตารางที่ 5.3: ข้อมูลระหว่างอุณหภูมิ (X_1) ระยะเวลา (X_2) และจำนวน Popcorn ที่ไม่สามารถรับประทานได้ (Y)

เมื่อ $\bar{X}_1 = 5$ และ $\bar{X}_2 = 105$

จากประสบการณ์ในอดีต ผู้ทดลองเชื่อว่าโมเดลโพลิโนเมียลกำลังสองเป็นรูปแบบโมเดลที่เหมาะสมกับข้อมูลชุดนี้ จงสร้างสมการแสดงความสัมพันธ์ของข้อมูลดังกล่าว

วิธีทำ พิจารณาโมเดลโพลิโนเมียลกำลังสองของข้อมูลที่ Center แล้ว

$$Y = \beta'_0 + \beta'_1 x_1 + \beta'_2 x_2 + \beta'_{11} x_1^2 + \beta'_{22} x_2^2 + \beta'_{12} x_1 x_2 + \epsilon$$

เมื่อ $x_1 = X_1 - 5$ และ $x_2 = X_2 - 105$

	อุณหภูมิ	เวลา	Popcorn	x_1	x_2	x_1^2	x_2^2	$x_1 x_2$
	6	105	29	1	0	1	0	0
	4	120	33	-1	15	1	225	-15
	6	120	45	1	15	1	225	15
	6	105	47	1	0	1	0	0
	5	120	16	0	15	0	225	0
	5	105	21	0	0	0	0	0
	4	90	131	-1	-15	1	225	15
	5	120	39	0	15	0	225	0
	4	105	37	-1	0	1	0	0
	5	90	37	0	-15	0	225	0
	4	105	39	-1	0	1	0	0
	6	90	22	1	-15	1	225	-15
	5	105	35	0	0	0	0	0
	5	105	22	0	0	0	0	0
	5	90	55	0	-15	0	225	0
รวม	75	1505	608	0	0	8	1800	0

ตารางที่ 5.4: ตารางแสดงการคำนวณค่าต่าง ๆ ของข้อมูลในตัวอย่างที่ 5.4

การประมาณค่าพารามิเตอร์ของโมเดลข้างต้นจะใช้วิธีกำลังสองน้อยที่สุด เช่นเดียวกับการประมาณค่าพารามิเตอร์ในโมเดลถดถอยเชิงเส้นตรงแบบพหุที่มีตัวแปรอิสระ 5 ตัว ตารางที่ 5.4 แสดงการคำนวณค่าต่าง ๆ และได้ค่าประมาณของสัมประสิทธิ์ถดถอยเป็น

$$\mathrm{b}' = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \begin{bmatrix} 23.231 \\ -12.125 \\ -0.9333 \\ 16.846 \\ 0.0693 \\ 2.0167 \end{bmatrix}$$

ดังนั้นสามารถเขียนสมการโพลิโนเมียลกำลังสองมีได้ดังนี้

$$\hat{Y} = \underset{(7.414)}{23.231} + \underset{(5.457)}{-12.125}x_1 + \underset{(0.3638)}{-0.9333}x_2 + \underset{(8.008)}{6.846}x_1^2 + \underset{(0.0356)}{0.0693}x_2^2 + \underset{(0.5145)}{2.0167}x_1x_2$$

โดยที่ $R^2 = 0.793$

หากต้องการทดสอบนัยสำคัญของสมการข้างต้น ทำได้โดยกำหนดสมมติฐานของการทดสอบเป็น

H_0 : $\beta'_1 = \beta'_2 = \beta'_{11} = \beta'_{22} = \beta'_{12} = 0$

H_1 : อย่างน้อยมีค่าสัมประสิทธิ์ถดถอย 1 ค่าที่ไม่เป็นศูนย์

สร้างตารางวิเคราะห์ความแปรปรวน

ANOVA

Source of variation	***df***	***SS***	***MS***	***F***
Regression	5	8231.8	1646.4	6.91
Error	9	2143.9	238.2	
Total	14	10375.7		

บริเวณวิกฤติ: $F_c > F_{5,9\,(.10)} = 2.61$

เนื่องจาก $F_c = 6.91$ มีค่ามากกว่าค่าวิกฤติ ดังนั้นปฏิเสธ H_0 นั่นคือ อย่างน้อยมีหนึ่งเทอมที่มีอิทธิพลต่อจำนวน Popcorn ที่ระดับนัยสำคัญ .10

จากนั้นตรวจสอบอิทธิพลของเทอมกำลังสองต่อโมเดลที่กำลังพิจารณาโดยใช้วิธี Extra-sum-of-squares ซึ่งกำหนดสมมติฐานของการทดสอบเป็น

H_0 : $\beta'_{11} = \beta'_{22} = \beta'_{12} = 0$

H_1 : $\beta'_{ij} \neq 0$ อย่างน้อย 1 ค่า, $\quad i, j = 1, 2, 3$

เมื่อ H_0 เป็นจริง จะได้โมเดลกำลังหนึ่งดังนี้

$$Y = \beta_0' + \beta_1' x_1 + \beta_2' x_2 + \epsilon$$

ประมาณพารามิเตอร์ของโมเดลข้างต้นโดยใช้วิธีกำลังสองน้อยที่สุด จะได้สมการถดถอยกำลังหนึ่งเป็น

$$\hat{Y} = \underset{(6.511)}{40.533} - \underset{(8.916)}{12.125}x_1 - \underset{(0.5944)}{0.9333}x_2$$

โดยที่ $R^2 = 0.264$, $SSR(x_1, x_2) = 2744.1$ $SSE(x_1, x_2) = 7631.6$
จะได้ว่า

$$\begin{aligned} SSR(x_1^2, x_2^2, x_1x_2 \mid x_1, x_2) &= SSR(x_1, x_2, x_1^2, x_2^2, x_1x_2) - SSR(x_1, x_2) \\ &= 8231.8 - 2744.1 \\ &= 5487.7 \end{aligned}$$

คำนวณสถิติทดสอบ:

$$\begin{aligned} F_c &= \frac{SSR(x_1^2, x_2^2, x_1x_2 \mid x_1, x_2)/(df_{Full} - df_{Reduced})}{MSE_{Full}} \\ &= \frac{5487.7/3}{238.2} \\ &= 7.68 \end{aligned}$$

บริเวณวิกฤติ: $F_c > 2.81$ เมื่อ $F_{3,9\,(.10)} = 2.81$
เนื่องจากค่า F_c ที่คำนวณได้มีค่ามากกว่าค่าวิกฤติ ดังนั้นปฏิเสธ H_0 นั่นคือ อย่างน้อยมีเทอมกำลังสองหนึ่งเทอมที่มีอิทธิพลต่อโมเดล ที่ระดับนัยสำคัญ .10

พารามิเตอร์	ค่าประมาณ	ความคลาดเคลื่อนมาตรฐาน	สถิติทดสอบ t
β_0	23.231	7.414	3.13
β_1	-12.125	5.457	-2.22
β_2	-0.9333	0.3638	-2.57
β_{11}	16.846	8.008	2.10
β_{22}	0.0693	0.0356	1.95
β_{12}	2.0167	0.5145	3.92

ตารางที่ 5.5: แสดงการทดสอบนัยสำคัญของค่าสัมประสิทธิ์ถดถอยแต่ละค่า (ค่าวิกฤติ: $t_{.05,9} = 1.833$)

การทดสอบนัยสำคัญของค่าสัมประสิทธิ์ถดถอยแต่ละค่า สามารถทำได้โดยใช้สถิติทดสอบ t ดังแสดงในตาราง 5.5 ซึ่งจะเห็นได้ว่าค่าสัมบูรณ์ของสถิติทดสอบ t ที่คำนวณได้ทุกตัวมีค่ามากกว่าค่าวิกฤติ นั่นคือค่าสัมประสิทธิ์ถดถอยทุกตัวแตกต่างจากศูนย์อย่างมีนัยสำคัญ

การตรวจสอบความเหมาะสมของโมเดลโพลิโนเมียลกำลังสองจะพิจารณาจากรูปที่ 5.5 ซึ่งจะเห็นได้ว่าข้อมูลมีการแจกแจงแบบปกติโดยประมาณ และมีค่าสังเกตหนึ่งค่าที่มีความคลาดเคลื่อนมากกว่าค่าอื่น ซึ่งจะต้องมีการตรวจสอบต่อไปว่าเป็นค่าที่ผิดปกติหรือไม่

5.4 โมเดลโพลิโนเมียลแบบออโธโกนอล (Orthogonal Polynomial Model)

เป็นที่ทราบแล้วว่าการ Center ข้อมูลก่อนที่จะสร้างโมเดลโพลิโนเมียลกรณีที่มีตัวแปรอิสระเพียงตัวเดียว ก็เพื่อลดปัญหา Multicollinearity แต่บางครั้งการ Center ข้อมูล อาจไม่สามารถขจัดปัญหา Multicollinearity ให้หมดไปได้ ซึ่งทางเลือกอีกทางก็คือการสร้างสมการโพลิโนเมียลแบบออโธโกนอล
พิจารณาโมเดลโพลิโนเมียลกำลัง k

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_k x_i^k + \epsilon, \qquad i = 1, 2, \ldots, n \tag{5.17}$$

โดยทั่วไปแล้วคอลัมน์ของเมตริกซ์ $\mathbf{X}$ ไม่เป็นอิสระกัน ซึ่งถ้ากำลังสูงสุดของโมเดลโพลิโนเมียลเพิ่มขึ้น โดยเพิ่มเทอม $\beta_{k+1}x^{k+1}$ เข้าไปในโมเดล จะต้องคำนวณ $(\mathbf{X}'\mathbf{X})$ ใหม่ และส่งผลให้ค่าประมาณของสัมประสิทธิ์ถดถอยของเทอมที่มีกำลังต่ำกว่า ($b_1, b_2, \ldots, b_k$) มีค่าแตกต่างไปจากเดิม
รูปแบบโมเดลโพลิโนเมียลแบบออโธโกนอล:

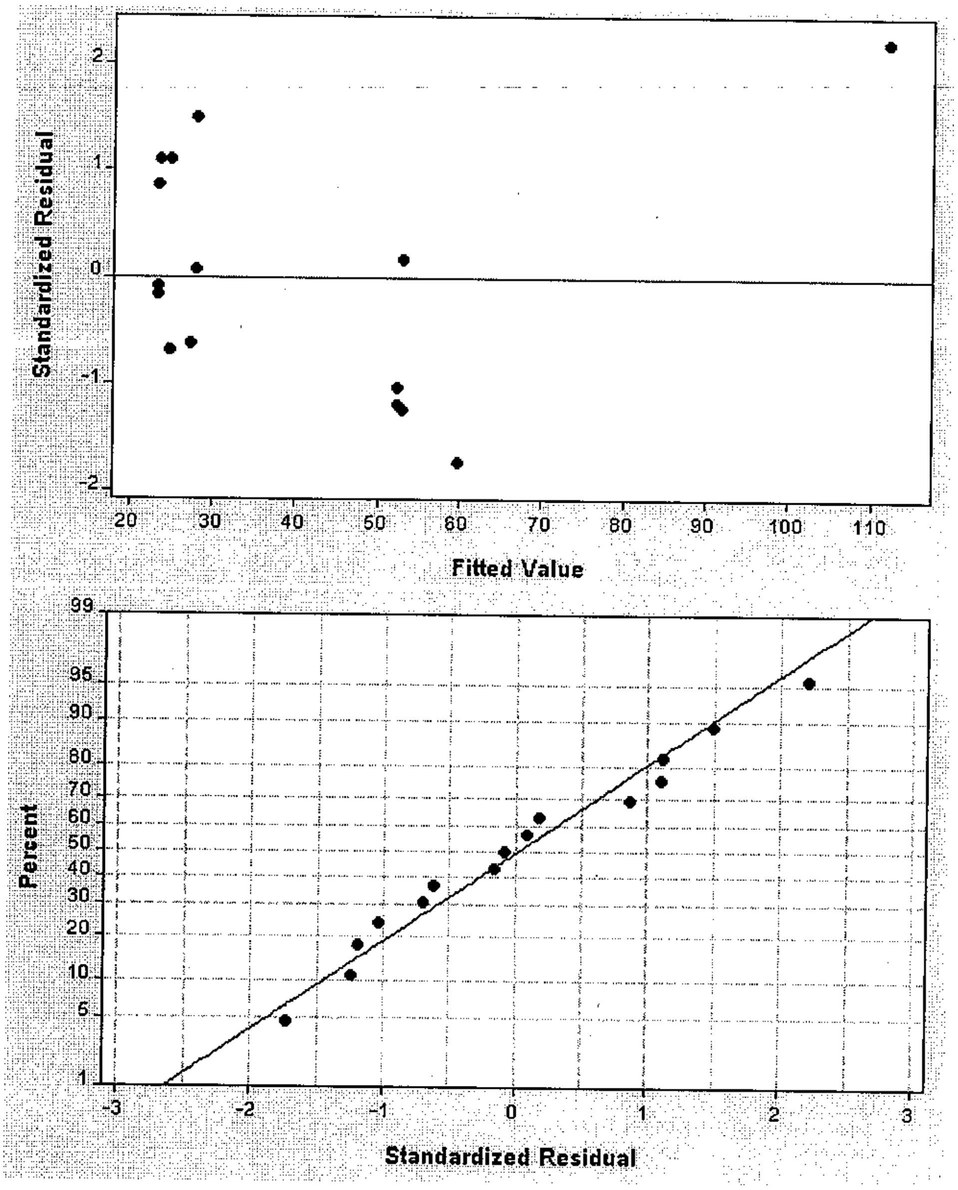
$$Y = \alpha_0 P_0(X_i) + \alpha_1 P_1(X_i) + \alpha_2 P_2(X_i) + \cdots + \alpha_k P_k(X_i) + \epsilon_i, \qquad i = 1, 2, \ldots, n \tag{5.18}$$

เมื่อ $P_u(X_i)$ แทน โพลิโนเมียลแบบออโธโกนอลกำลัง u โดยที่

$$\begin{aligned} \sum_{i=1}^{n} P_s(X_i)\, P_t(X_i) &= 0, \qquad s \neq t, \quad s, t = 0, 1, \ldots, k; \\ P_0(X_i) &= 1 \end{aligned}$$

สมการ (5.18) สามารถเขียนให้อยู่ในรูปของเมตริกซ์ได้เป็น

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\alpha} + \boldsymbol{\epsilon} \tag{5.19}$$



รูปที่ 5.5: กราฟระหว่างค่า Standardized residuals และค่าพยากรณ์ (บน) และ Normal probability plot ของ Standardized residuals ที่ได้จากสมการโพลิโนเมียลกำลังสองในตัวอย่างที่ 5.4 (ล่าง)

เมื่อ

$$\mathbf{X} = \begin{bmatrix} P_0(X_1) & P_1(X_1) & \cdots & P_k(X_1) \\ P_0(X_2) & P_1(X_2) & \cdots & P_k(X_2) \\ \vdots & \vdots & & \vdots \\ P_0(X_n) & P_1(X_n) & \cdots & P_k(X_n) \end{bmatrix}$$

เนื่องจากเมตริกซ์ $\mathbf{X}$ ข้างต้นมีคอลัมน์ที่เป็นอิสระกัน ดังนั้น

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} \sum_{i=1}^{n} P_0^2(X_i) & 0 & \cdots & 0 \\ 0 & \sum_{i=1}^{n} P_1^2(X_i) & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sum_{i=1}^{n} P_k^2(X_i) \end{bmatrix}$$

และ

$$\mathbf{X}'\mathbf{Y} = \begin{bmatrix} \sum_{i=1}^{n} P_0(X_i)\,Y_i \\ \sum_{i=1}^{n} P_1(X_i)\,Y_i \\ \vdots \\ \sum_{i=1}^{n} P_k(X_i)\,Y_i \end{bmatrix}$$

ประมาณพารามิเตอร์ α ในสมการ (5.19) ด้วยวิธีกำลังสองน้อยที่สุด จะได้ตัวประมาณของพารามิเตอร์ α ดังนี้

$$\hat{\alpha} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \tag{5.20}$$

นั่นคือ ตัวประมาณพารามิเตอร์ α_j คือ

$$\hat{\alpha}_j = \frac{\sum_{i=1}^{n} P_j(X_i)\,Y_i}{\sum_{i=1}^{n} P_j^2(X_i)}, \qquad j = 0, 1, \ldots, k \tag{5.21}$$

เนื่องจากโพลิโนเมียลกำลังศูนย์มีค่าเท่ากับ 1 ($P_0(X_i) = 1$) จะได้ว่า

$$\hat{\alpha}_0 = \bar{Y} \tag{5.22}$$

คำนวณผลรวมกำลังสองเนื่องจากความคลาดเคลื่อนและสมการถดถอยได้ดังนี้

$$
\begin{aligned}
SSE &= S_{yy} - \sum_{j=1}^{k} \hat{\alpha}_j \left(\sum_{i=1}^{n} P_j(X_i)\, Y_i \right) &(5.23)\\
SSR(\alpha_j) &= \hat{\alpha}_j \sum_{i=1}^{n} P_j(X_i)\, Y_i &(5.24)
\end{aligned}
$$

จะเห็นได้ว่าผลรวมกำลังสองเนื่องจากสมการถดถอยของพารามิเตอร์ใด ๆ ไม่ขึ้นอยู่กับพารามิเตอร์อื่นในโมเดล นอกจากนี้การทดสอบนัยสำคัญของเทอมที่มีกำลังสูงสุด สามารถกำหนดสมมติฐานของการทดสอบได้เป็น

$$H_0 : \alpha_k = 0 \quad vs. \quad H_1 : \alpha_k \neq 0$$

ซึ่งการทดสอบ $H_0 : \alpha_k = 0$ จะเหมือนกับการทดสอบ $H_0 : \beta_k = 0$ นั่นเอง คำนวณสถิติทดสอบ

$$F_c = \frac{SSR(\alpha_j)/1}{SSE/(n-k-1)} = \frac{\hat{\alpha}_j \sum_{i=1}^{n} P_j(X_i)\, Y_i}{SSE/(n-k-1)} \qquad (5.25)$$

โดยจะปฏิเสธ H_0 ถ้า $F_c > F_{1, n-k-1\,(\alpha)}$

หมายเหตุ ถ้ากำลังของโพลิโนเมียลแบบออโธโกนอลเพิ่มจาก k เป็น $k+r$ จะต้องประมาณค่าสัมประสิทธิ์ถดถอยใหม่เพียง r ค่า ส่วนค่าประมาณสัมประสิทธิ์ถดถอย $\hat{\alpha}_0, \hat{\alpha}_1, \ldots, \hat{\alpha}_k$ เดิมที่มีอยู่แล้วยังคงไม่เปลี่ยนแปลง เนื่องจากคุณสมบัติออโธโกนอลของโมเดล

การคำนวณค่า $P_j(X_i)$ ในโพลิโนเมียลแบบออโธโกนอล สามารถทำได้ง่าย ถ้า X มีระยะห่างเท่า ๆ กัน โดย $P_j(X_i)$ 5 อันดับแรกคือ

$$
\begin{aligned}
P_0(X_i) &= 0\\
P_1(X_i) &= \lambda_1 \left[\frac{X_i - \bar{X}}{d} \right]\\
P_2(X_i) &= \lambda_2 \left[\left(\frac{X_i - \bar{X}}{d} \right)^2 - \frac{n^2-1}{12} \right]\\
P_3(X_i) &= \lambda_3 \left[\left(\frac{X_i - \bar{X}}{d} \right)^3 - \left(\frac{X_i - \bar{X}}{d} \right) \left(\frac{3n^2-7}{20} \right) \right]\\
P_4(X_i) &= \lambda_4 \left[\left(\frac{X_i - \bar{X}}{d} \right)^4 - \left(\frac{X_i - \bar{X}}{d} \right)^2 \left(\frac{3n^2-13}{14} \right) + \frac{3(n^2-1)(n^2-9)}{560} \right]
\end{aligned}
$$

เมื่อ

d แทน ระยะห่างระหว่างระดับของ X

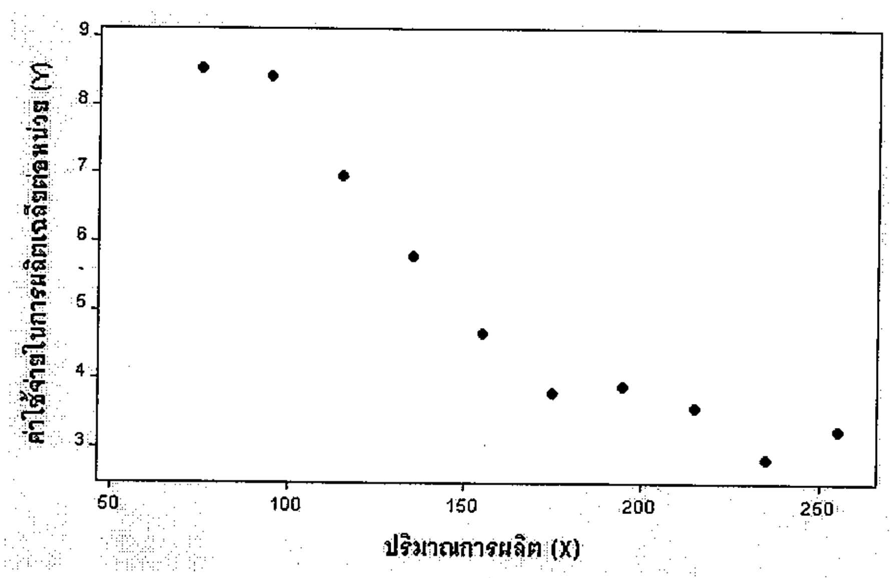
λ_j แทน ค่าคงที่ที่เลือกมาเพื่อให้โพลิโนเมียลมีค่าเป็นจำนวนเต็ม

โดยค่าสัมประสิทธิ์ของโพลิโนเมียลแบบออโธโกนอลแสดงในตารางที่ 6 (ภาคผนวก) และการสร้างโมเดลโพลิโนเมียลแบบออโธโกนอล เมื่อตัวแปรอิสระมีระยะห่างไม่เท่ากันก็สามารถทำได้เช่นกัน

ตัวอย่างที่ 5.5 นักวิจัยดำเนินงานคนหนึ่งต้องการศึกษาความสัมพันธ์ระหว่างปริมาณการผลิต (X) ของสินค้าชนิดหนึ่ง กับค่าใช้จ่ายในการผลิตเฉลี่ยต่อหน่วย (Y) ได้ข้อมูลดังนี้

ครั้งที่ของการผลิต	1	2	3	4	5	6	7	8	9	10
X	75	95	115	135	155	175	195	215	235	255
Y	8.53	8.41	6.95	5.78	4.67	3.78	3.89	3.59	2.82	3.26

ตารางที่ 5.6: ข้อมูลระหว่างปริมาณการผลิต (X) กับค่าใช้จ่ายในการผลิตเฉลี่ยต่อหน่วย (Y)



รูปที่ 5.6: กราฟระหว่างปริมาณการผลิต (X) กับค่าใช้จ่ายในการผลิตเฉลี่ยต่อหน่วย ของข้อมูลในตัวอย่างที่ 5.5

จากแผนภาพการกระจายในรูปที่ 5.6 พบว่าความสัมพันธ์มีลักษณะเป็นเส้นโค้ง จึงเชื่อว่าโมเดลโพลิโนเมียลกำลังสองน่าจะเป็นรูปแบบโมเดลที่เหมาะสมกับข้อมูลชุดนี้ จงสร้างสมการแสดงความสัมพันธ์ของข้อมูลดังกล่าว

วิธีทำ พิจารณาโมเดลโพลิโนเมียลแบบออโธโกนอลกำลังสอง

$$Y_i = \alpha_0 P_0(X_i) + \alpha_1 P_1(X_i) + \alpha_2 P_2(X_i) + \epsilon_i, \qquad i = 1, 2, \ldots, 10$$

Y	$P_0(X_i)$	$P_1(X_i)$	$P_2(X_i)$
8.53	1	-9	6
8.41	1	-7	2
6.95	1	-5	-1
5.78	1	-3	-3
4.67	1	-1	-4
3.78	1	1	-4
3.89	1	3	-3
3.59	1	5	-1
2.82	1	7	2
3.26	1	9	6
$\sum_{i=1}^{10} Y_i = 51.68$	$\sum_{i=1}^{10} P_0^2(X_i) = 10$	$\sum_{i=1}^{10} P_1^2(X_i) = 330$	$\sum_{i=1}^{10} P_2^2(X_i) = 132$
		$\lambda_1 = 2$	$\lambda_2 = 1/2$

ตารางที่ 5.7: ค่าสัมประสิทธิ์ของโพลิโนเมียลแบบออโธโกนอลกำลังสอง

ตารางที่ 5.7 แสดงค่าสัมประสิทธิ์ของโพลิโนเมียลแบบออโธโกนอล $P_0(X_i)$, $P_1(X_i)$ และ $P_2(X_i)$ ซึ่งได้จากตารางที่ 6 (ภาคผนวก) และได้ว่า

$$\mathbf{X'X} = \begin{bmatrix} \sum_{i=1}^{10} P_0^2(X_i) & 0 & 0 \\ 0 & \sum_{i=1}^{10} P_1^2(X_i) & 0 \\ 0 & 0 & \sum_{i=1}^{10} P_2^2(X_i) \end{bmatrix} = \begin{bmatrix} 10 & 0 & 0 \\ 0 & 330 & 0 \\ 0 & 0 & 132 \end{bmatrix}$$

$$\mathbf{X'Y} = \begin{bmatrix} \sum_{i=1}^{10} P_0(X_i)\,Y_i \\ \sum_{i=1}^{10} P_1(X_i)\,Y_i \\ \sum_{i=1}^{10} P_k(X_i)\,Y_i \end{bmatrix} = \begin{bmatrix} 51.68 \\ -109.92 \\ 19.85 \end{bmatrix}$$

หาตัวประมาณของพารามิเตอร์ α ได้ดังนี้

$$\hat{\alpha} = (\mathbf{X'X})^{-1}\mathbf{X'Y} = \begin{bmatrix} 5.1680 \\ -0.3331 \\ 0.1504 \end{bmatrix}$$

ได้สมการโพลิโนเมียลแบบออโธโกนอลดังนี้

$$\hat{Y} = 5.1680 - 0.3331\,P_1(X) + 0.1504\,P_2(X)$$

คำนวณผลรวมกำลังสองเนื่องจากสมการถดถอยได้เป็น

$$\begin{aligned} SSR(\alpha_1, \alpha_2) &= \sum_{j=1}^{2} \hat{\alpha_j} \left[\sum_{i=1}^{10} P_j(X_i)\,Y_i \right] \\ &= (-0.3331)(-109.92) + (0.1504)(19.85) \\ &= 39.5998 \end{aligned}$$

กำหนดสมมติฐานสำหรับทดสอบนัยสำคัญของสมการ

$$H_0: \quad \alpha_1 = \alpha_2 = 0$$

$H_1: \quad \alpha_i \neq 0$ อย่างน้อย 1 ค่า, $\quad i = 1, 2$

ANOVA

Source of variation	df	SS	MS	F
Regression	2	39.5998	19.7999	114.0547
Linear (α_1)	1	36.6144	36.6144	210.9124
Quadratic (α_2)	1	2.9854	2.9854	17.1970
Error	7	1.2152	0.1736	
Total	9	40.8150		

ค่าวิกฤติ: $F_{2,7(.05)} = 4.74$ และ $F_{1,7(.05)} = 5.59$
เนื่องจาก $F_c = 114.0547$ มีค่ามากกว่าค่าวิกฤติ ดังนั้นปฏิเสธ H_0 นั่นคือ อย่างน้อยมีหนึ่งเทอมที่มีอิทธิพลต่อโมเดลอย่างมีนัยสำคัญ ที่ระดับนัยสำคัญ .05

หากต้องการทดสอบอิทธิพลของเทอมกำลังหนึ่งและเทอมกำลังสอง กำหนดสมมติฐานของการทดสอบได้ดังนี้

$$H_0: \alpha_i = 0 \quad vs. \quad H_1: \alpha_i \neq 0, \qquad i = 1, 2$$

เนื่องจากค่า F ที่คำนวณได้ทั้งสองตัว $F_1 = 210.9124$ และ $F_2 = 17.1970$ มีค่ามากกว่าค่าวิกฤติ ดังนั้นปฏิเสธ H_0 นั่นคือ ทั้งสองเทอมมีอิทธิพลต่อโมเดลอย่างมีนัยสำคัญที่ระดับ 0.05

นอกจากนี้ยังสามารถแปลงสมการโพลิโนเมียลออโธโกนอลให้อยู่ในรูปของตัวแปรเดิมเพื่อประโยชน์ในการ

พยากรณ์ได้ดังนี้

$$\begin{aligned}\hat{Y} &= 5.1680 - 0.3331\,P_1(X) + 0.1504\,P_2(X) \\ &= 5.1680 - 0.3331(2)\left[\frac{X-165}{20}\right] + 0.1504\left(\frac{1}{2}\right)\left[\left(\frac{X-165}{20}\right)^2 - \frac{(10)^2-1}{12}\right] \\ &= 4.5476 - 0.03331(X-165) + 0.0002(X-165)^2\end{aligned}$$

จะเห็นได้ว่าการสร้างโมเดลถดถอยแบบโพลิโนเมียลจัดเป็นกรณีพิเศษของการสร้างโมเดลถดถอยเชิงเส้นตรงทั่ว ๆ ไป ซึ่งการสร้างโมเดลถดถอยแบบโพลิโนเมียลเพื่อประมาณฟังก์ชันถดถอยที่แท้จริงนิยมใช้หลักของ *Hierarchical approach* กล่าวคือ ถ้าเทอมที่มีกำลังสูง ๆ มีนัยสำคัญต่อโมเดลที่พิจารณาแล้ว เทอมที่มีกำลังต่ำกว่าที่เกี่ยวข้องทั้งหมดจะต้องอยู่ในโมเดลด้วย เช่น จะไม่ตัดเทอมกำลังสองของโมเดลโพลิโนเมียลกำลังสามทิ้งไป เนื่องจากเทอมกำลังสองมีกำลังต่ำกว่า และอาจพิจารณาว่าเป็นการให้ข้อมูลพื้นฐานเกี่ยวกับรูปร่างของฟังก์ชันของตัวแปรตาม ส่วนเทอมกำลังสามจะให้ข้อมูลเกี่ยวกับรูปร่างของฟังก์ชันที่ละเอียดมากขึ้น นอกจากนี้ถ้าอิทธิพลร่วมมีนัยสำคัญต่อโมเดลแล้ว เทอมกำลังหนึ่งที่เกี่ยวข้องทั้งหมดจะต้องอยู่ในโมเดลด้วยเช่นกัน

หลังจากสร้างสมการโพลิโนเมียลแล้ว โดยทั่วไปนิยมนำเสนอสมการสุดท้ายที่อยู่ในรูปของตัวแปรเดิมมากกว่าที่จะนำเสนอในรูปของตัวแปรที่ Center แล้ว ซึ่งสามารถทำได้โดยง่าย เช่น พิจารณาโมเดลโพลิโนเมียลกำลังสองของข้อมูลที่ Center แล้ว ดังนี้

$$\hat{Y} = b_0' + b_1'x + b_2'x^2$$

แปลงให้อยู่ในรูปของตัวแปรเดิมได้เป็น

$$\begin{aligned}\hat{Y} &= b_0' + b_1'x + b_2'x^2 \\ &= b_0' + b_1'(X-\bar{X}) + b_2'(X-\bar{X})^2 \\ &= b_0' + b_1'(X-\bar{X}) + b_2'(X^2 - 2X\bar{X} + \bar{X}^2) \\ &= (b_0' - b_1'\bar{X} + b_2'\bar{X}^2) + (b_1' - 2b_2'\bar{X})X + b_2'X^2 \\ &= b_0 + b_1X + b_2X^2\end{aligned}$$

เมื่อ

$$
\begin{aligned}
b_0 &= b'_0 - b'_1\bar{X} + b'_2\bar{X}^2 \\
b_1 &= b'_1 - 2b'_2\bar{X} \\
b_2 &= b'_2
\end{aligned}
$$

ซึ่งค่าพยากรณ์และค่าความคลาดเคลื่อนที่ได้จากสมการถดถอยในเทอมของตัวแปรอิสระ X จะเหมือนกับค่าที่ได้จากสมการถดถอยของข้อมูลที่ Center แล้ว ทั้งนี้การสร้างสมการถดถอยสำหรับข้อมูลที่ Center ก็เพื่อลดปัญหา Multicollinearity ที่เกิดจาก $X, X^2, X^3, \ldots$ นั่นเอง แต่อย่างไรก็ตามค่าประมาณของความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ถดถอยที่ได้จากสมการที่อยู่ในรูปของตัวแปรเดิมและตัวแปรที่ Center แล้วมีค่าแตกต่างกัน

หมายเหตุ

- การสร้างโมเดลโพลิโนเมียลอาจทำให้จำนวนองศาความเป็นอิสระมีค่ามากกว่าจำนวนองศาความเป็นอิสระที่ได้จากโมเดลที่ไม่ใช่เชิงเส้นตรง (Nonlinear models) หรือโมเดลเชิงเส้นตรงที่มีการแปลงค่าของตัวแปร
- การสร้างโมเดลโพลิโนเมียลอาจก่อให้เกิดปัญหา Multicollinearity ที่รุนแรง ถึงแม้ว่าจะมีการ Center ค่าของตัวแปรแล้วก็ตาม
- ควรมีการตรวจสอบ Lack of fit ของโมเดลควบคู่ไปกับการทดสอบนัยสำคัญของค่าสัมประสิทธิ์ถดถอยแบบโพลิโนเมียลเสมอ

แบบฝึกหัดบทที่ 5

1. พิจารณาข้อมูลต่อไปนี้

ค่าสังเกตที่	X	Y
1	4.0	24.60
2	4.0	24.71
3	4.0	23.90
4	5.0	39.50
5	5.0	39.60
6	6.0	57.12
7	6.5	67.11
8	6.5	67.24
8	6.8	67.15
10	7.0	77.87
11	7.1	80.11
12	7.3	84.67
13	7.5	80.00
14	8.0	79.00
15	8.1	77.00
16	8.5	76.00
17	9.0	75.00
18	9.0	77.00
19	10.0	78.00
20	10.3	73.00

1.1 จงสร้างสมการถดถอยเชิงเส้นตรงแสดงความสัมพันธ์ระหว่างตัวแปรตาม (Y) และตัวแปรอิสระ (X) พร้อมทั้งทดสอบ Lack of fit ของสมการดังกล่าวที่ระดับนัยสำคัญ 0.05 สรุปผลที่ได้

1.2 จงสร้างกราฟของความคลาดเคลื่อนระหว่าง Standardized residuals และค่าพยากรณ์ (Fitted values) ของสมการในข้อ 1.1 พร้อมทั้งอธิบายผลที่ได้

1.3 จงสร้างสมการถดถอยแบบโพลิโนเมียลกำลังสองของข้อมูลที่มีการ Center ค่าของตัวแปรอิสระ X

1.4 ที่ระดับนัยสำคัญ 0.05 จงทดสอบนัยสำคัญของสมการในข้อ 1.3 และสรุปผลที่ได้

1.5 ที่ระดับนัยสำคัญ 0.05 จงตรวจสอบว่าเทอมกำลังสองควรอยู่ในสมการหรือไม่ พร้อมทั้งสรุปผลที่ได้

2. บริษัทผลิตน้ำอัดลมเชื่อว่า ระดับของ Carbonate ในน้ำอัดลม (Y) ถูกกระทบด้วยอุณหภูมิ (X_1) และความดันในระหว่างการผลิต (X_2) จึงได้รวบรวมข้อมูลจากค่าสังเกตทั้งหมด 12 ค่า ดังนี้

ค่าสังเกตที่	1	2	3	4	5	6	7	8	9	10	11	12
Y	2.60	2.40	17.32	15.60	16.12	5.36	6.19	10.17	2.62	2.98	6.92	7.06
X_1	31.0	31.0	31.5	31.5	31.5	30.5	31.5	30.5	31.0	30.5	31.0	30.5
X_2	21.0	21.0	24.0	24.0	24.0	22.0	22.0	23.0	21.5	21.5	22.5	22.5

2.1 จงสร้างสมการถดถอยแบบโพลิโนเมียลกำลังสอง (Center ตัวแปรอิสระทุกตัว ก่อนที่จะคำนวณ) โดยแสดงอิทธิพลร่วม (Interaction effect) ระหว่างตัวแปรอิสระด้วย

2.2 ที่ระดับนัยสำคัญ 0.05 จงทดสอบนัยสำคัญของสมการ พร้อมทั้งสรุปผลที่ได้

2.3 ที่ระดับนัยสำคัญ 0.05 จงทดสอบว่าเทอมกำลังสองทั้งหมดมีอิทธิพลอย่างมีนัยสำคัญต่อโมเดลหรือไม่ พร้อมทั้งสรุปผล

2.4 ที่ระดับนัยสำคัญ 0.05 จงทดสอบว่ามีอิทธิพลร่วมระหว่างตัวแปรอิสระหรือไม่ พร้อมทั้งสรุปผลที่ได้

3. ข้อมูลต่อไปนี้แสดงถึงน้ำหนักที่หายไปของจรวดขับเคลื่อนด้วยเชื้อเพลิงแข็ง (Y) หน่วยเป็นกิโลกรัม กับจำนวนเดือนหลังจากที่ผลิต (X)

ค่าสังเกตที่	1	2	3	4	5	6	7	8	9	10
X	0.25	0.50	0.75	1.00	1.25	1.50	1.75	2.00	2.25	2.50
Y	1.42	1.39	1.55	1.89	2.43	3.15	4.05	5.15	6.43	7.89

3.1 จงสร้างสมการถดถอยเชิงเส้นตรง (Center ตัวแปรอิสระ ก่อนที่จะคำนวณ) พร้อมทั้งสร้างกราฟของความคลาดเคลื่อน (Standardized residuals และค่าพยากรณ์) พร้อมทั้งอธิบายผลที่ได้ ท่านคิดว่าข้อบกพร่องของสมการดังกล่าวคืออะไร

3.2 จงสร้างสมการถดถอยแบบโพลิโนเมียลกำลังสอง (Center ตัวแปรอิสระ ก่อนที่จะคำนวณ)

3.3 ที่ระดับนัยสำคัญ 0.05 จงทดสอบนัยสำคัญของสมการในข้อ 3.2 พร้อมทั้งสรุปผลที่ได้

3.4 ที่ระดับนัยสำคัญ 0.05 จงตรวจสอบว่าเทอมกำลังสองมีอิทธิพลอย่างมีนัยสำคัญต่อโมเดลในข้อ 3.2 หรือไม่ พร้อมทั้งสรุปผลที่ได้

3.5 จงสร้างสมการโพลิโนเมียลกำลังสองแบบ Orthogonal พร้อมทั้งตรวจสอบว่าที่ระดับนัยสำคัญ 0.05 เทอมกำลังสองมีอิทธิพลอย่างมีนัยสำคัญต่อโมเดลหรือไม่ สรุปผลที่ได้

3.6 ที่ระดับนัยสำคัญ 0.05 จงตรวจสอบว่าควรเพิ่มเทอมกำลังสามเข้าไปในโมเดลโพลิโนเมียลแบบ Orthogonal กำลังสองหรือไม่ พร้อมทั้งสรุปผลที่ได้