บทที่ 6
การถดถอยแบบไม่เป็นเส้นตรง
(Nonlinear Regression)

1. ฟังก์ชันของตัวแปร

ฟังก์ชันของตัวแปรตัวเดียว f(x) เรียกว่าเป็นฟังก์ชันเส้นตรง ถ้าสามารถเขียนในรูปสมการดังนี้

$$f(x) = a + bx$$

เมื่อค่าของ a และ b ไม่ขึ้นกับค่าของ x

และฟังก์ชันของตัวแปร k ตัว คือ $f(x_1, \ldots, x_k)$ เรียกว่าเป็นฟังก์ชันเส้นตรงของตัวแปร $X_1, \ldots, X_k$ ถ้าสามารถเขียนในรูปสมการดังนี้

$$f(x_1, \ldots, x_k) = a_0 + a_1 x_1 + \ldots + a_k x_k$$

เมื่อค่าของ $a_0, \ldots, a_k$ ไม่ขึ้นกับค่าของ $x_1, \ldots, x_k$

ตัวอย่างเช่น

ฟังก์ชันเส้นตรง

$$\mu_y(x) = \beta_0 + \beta_1 x$$

ฟังก์ชันไม่เป็นเส้นตรง

$$\mu_y(x_1, x_2, \beta_0, \beta_1, \beta_2) = \beta_0 + \beta_1 x_1 + e^{\beta_2 x_2}$$

ข้อตกลงเบื้องต้น

1. ค่าเฉลี่ยของประชากรย่อยของค่า Y เมื่อ $X_1 = x_1, \ldots, X_k = x_k$ แทนด้วย $\mu_y(x_1, \ldots, x_k)$ หรือ $\mu_y(x_1, \ldots, x_k; \beta_1, \ldots, \beta_p)$ ขึ้นกับพารามิเตอร์ $\beta_1, \ldots, \beta_p$
2. ส่วนเบี่ยงเบนมาตรฐานของค่า Y ของแต่ละประชากรย่อย มีค่าเท่ากันเท่ากับ $\sigma_{y|x, \ldots, x}$ หรือ σ
3. แต่ละประชากรย่อยของค่า Y มีการแจกแจงแบบปกติ
4. ตัวอย่างขนาด n ได้มาโดยการสุ่ม
5. ตัวอย่างทุกกลุ่มของค่า $y_i, x_{i,1}, \ldots, x_{i,k}$ สำหรับ $i = 1, \ldots, n$ ไม่มีความคลาดเคลื่อน

2. การประมาณค่าพารามิเตอร์

ในทางปฏิบัติสมมติว่าเราทราบฟังก์ชันการถดถอยที่ไม่ใช่เส้นตรง $\mu_y\,(x_1, \ldots, x_k)$ แต่ไม่ทราบค่าพารามิเตอร์ $\beta_0, \ldots, \beta_p$ ตัวอย่างเช่น จากประสบการณ์เราทราบว่า $\mu_y\,(x)$ อยู่ในรูปแบบ $\beta_1 + \beta_2\, e^{\beta_3 x}$ แต่เราไม่ทราบค่า β_1, β_2 และ β_3 ในสมการนี้มีตัวแปรอิสระ 1 ตัว คือ $k = 1$ มีพารามิเตอร์ 3 ตัวคือ $p = 3$

2.1 การประมาณค่าพารามิเตอร์โดยใช้วิธีกำลังสองน้อยที่สุด

การประมาณค่าพารามิเตอร์ $\beta_1, \ldots, \beta_p$ ใช้วิธีกำลังสองน้อยที่สุด ทำการประมาณค่าพารามิเตอร์ $\beta_1, \ldots, \beta_p$ โดยที่จะได้ค่าผลบวกกำลังสองของความคลาดเคลื่อนของการทำนายคือ $\sum_{i=1}^{n} e_i^2$ น้อยที่สุด

เมื่อ $$e_i = y_i - \mu_y\,(x_{i,1}, \ldots, x_{i,k})$$

ค่าประมาณของพารามิเตอร์ $\beta_1, \ldots, \beta_p$ แทนด้วย $\hat{\beta}_1, \ldots, \hat{\beta}_p$

ค่าประมาณของค่าเฉลี่ยของประชากรย่อย $\mu_y\,(x_1, \ldots, x_k)$ แทนด้วย $\hat{\mu}_y\,(x_1, \ldots, x_k)$ โดยที่

$$\hat{\mu}_y\,(x_1, \ldots, x_k) = \mu_y\,(x_1, \ldots, x_k\,;\, \hat{\beta}_1, \ldots, \hat{\beta}_p)$$

และ $$\hat{e}_i = y_i - \hat{\mu}_y\,(x_{i,1}, \ldots, x_{i,k})$$

เรียกว่าความคลาดเคลื่อน (residual) ของกลุ่มตัวอย่างที่ i

ค่าที่น้อยที่สุดของผลบวกกำลังสองของความคลาดเคลื่อนของการทำนายคือ

$$SSE = \sum_{i=1}^{n} \hat{e}_i^2 =$$

และ $$MSE = \frac{SSE}{(n-p)}$$

เรียกว่า mean square error และเป็นค่าประมาณที่ไม่เอนเอียงของ σ^2

$$\hat{\sigma} = \sqrt{\frac{SSE}{n-p}} = \sqrt{MSE}$$

2.2 การประมาณค่าพารามิเตอร์สำหรับการถดถอยแบบเป็นเส้นตรง

วิธีคำนวณในกรณีการถดถอยพหุเชิงเส้นตรง เราประมาณค่าพารามิเตอร์จากสมการปกติ (Normal equation) สมการปกติคือ

$$X^T X\hat{\beta} = X^T Y$$

ประมาณค่าพารามิเตอร์จากสมการปกติได้ดังนี้

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

2.3 การประมาณค่าพารามิเตอร์สำหรับการถดถอยแบบไม่เป็นเส้นตรง

สำหรับกรณีที่ไม่ใช่เชิงเส้นตรง เรามักอาศัยการคำนวณโดยคอมพิวเตอร์ เราต้องตั้งค่าเริ่มต้นหรือเดาค่าเริ่มต้นของ $\beta_1, \ldots, \beta_p$ จากการพิจารณาทฤษฎี หรือโดยการพล็อตข้อมูล

ตัวอย่างที่ 1 ให้ X เป็นความเข้มข้นของสารเคมี หน่วยเป็นมิลลิกรัมต่อลิตร มี 5 ระดับ คือ 1.0, 2.0, 3.0, 4.0, 5.0 และให้ Y เป็นความสว่างของลำแสงที่ผ่านในสารละลายที่มีสารเคมีความเข้มข้นต่าง ๆ หน่วยความสว่างของลำแสงกำหนดขึ้นเอง ได้ข้อมูลดังตาราง (Franklin A. Graybill and Hariharan K. Iyer ‘Regression Analysis : Concepts and Aplications’ Duxbury Press, California, 1994. pp. 608-610.

ตาราง 6.1 ข้อมูลความสว่างของลำแสงที่ผ่านในสารละลายที่มีสารเคมีความเข้มข้นต่าง ๆ

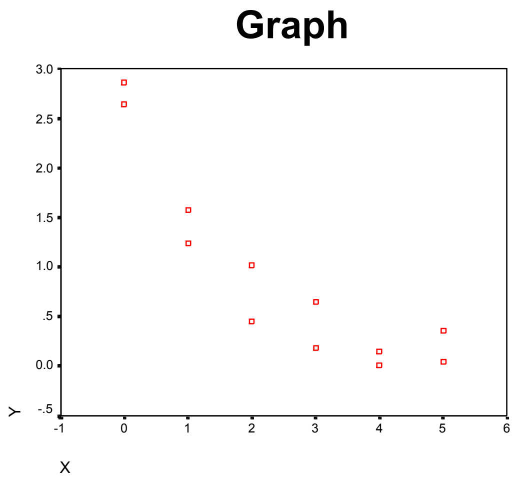
ค่าสังเกต	Y	X
1	2.86	0.0
2	2.64	0.0
3	1.57	1.0
4	1.24	1.0
5	0.45	2.0
6	1.02	2.0
7	0.65	3.0
8	0.18	3.0
9	0.15	4.0
10	0.01	4.0
11	0.04	5.0
12	0.36	5.0

$$\mu_Y(x) = \beta_1 + \beta_2\, e^{-\beta_3 x}$$

เมื่อ $\beta_1, \beta_2, \beta_3$ เป็นพารามิเตอร์ที่ต้องการประมาณค่าจากการเดาเริ่มต้นของค่าพารามิเตอร์คือ $\beta_1 = 0.0$, $\beta_2 = 2.0$, และ $\beta_3 = 0.5$

วิธีทำ

1. พล็อตกราฟของข้อมูลจะได้ภาพดังต่อไปนี้



ภาพ 6.1 กราฟของข้อมูลความสว่างของลำแสงที่ผ่านในสารละลายที่มีสารเคมีความเข้มข้นต่าง ๆ

2. การใช้โปรแกรมสำเร็จรูป SPSS วิเคราะห์การถดถอยแบบไม่ใช่เชิงเส้นตรงเพื่อประมาณค่าพารามิเตอร์ได้ผลลัพธ์ดังต่อไปนี้

Non-linear Regression

All the derivatives will be calculated numerically.

Iteration	Residual SS	B1	B2	B3
1	1.688464589	.000000000	2.00000000	.500000000
1.1	.4774066351	.119726690	2.63155692	.751137730
2	.4774066351	.119726690	2.63155692	.751137730
2.1	.4611079582	.032219820	2.71883657	.677259853
3	.4611079582	.032219820	2.71883657	.677259853
3.1	.4604272021	.028941506	2.72314902	.682907173
4	.4604272021	.028941506	2.72314902	.682907173
4.1	.4604271370	.028749738	2.72327864	.682758804
5	.4604271370	.028749738	2.72327864	.682758804
5.1	.4604271369	.028753645	2.72327641	.682762657

Run stopped after 10 model evaluations and 5 derivative evaluations.
Iterations have been stopped because the relative reduction between successive
residual sums of squares is at most SSCON = 1.000E-08

```
Nonlinear Regression Summary Statistics     Dependent Variable Y

  Source                 DF  Sum of Squares  Mean Square

  Regression              3       20.54287        6.84762
  Residual                9         .46043         .05116
  Uncorrected Total      12       21.00330

  (Corrected Total)      11       10.60589

  R squared = 1 - Residual SS / Corrected SS =     .95659

                                               Asymptotic 95 %
                              Asymptotic     Confidence Interval
  Parameter    Estimate       Std. Error     Lower          Upper

  B1          .028753645   .171520707  -.359253150   .416760441
  B2         2.723276411   .210540736  2.247000177  3.199552646
  B3          .682762657   .141658379   .362309140  1.003216174

  Asymptotic Correlation Matrix of the Parameter Estimates

                  B1         B2         B3

  B1          1.0000     -.6775      .8459
  B2          -.6775     1.0000     -.3954
  B3           .8459     -.3954     1.0000
```

ได้ค่าประมาณของพารามิเตอร์ดังนี้

$$\hat{\beta}_1 = 0.0288 \qquad \hat{\beta}_2 = 2.7233 \qquad \hat{\beta}_3 = 0.6828$$

และค่าประมาณของความคลาดเคลื่อนมาตรฐานของการทดลองคือ $\hat{\sigma} = \sqrt{0.0512} = 0.2263$

3. ช่วงความเชื่อมั่นและการทดสอบสมมติฐาน

สำหรับตัวแบบการถดถอยที่ไม่ใช่เชิงเส้นตรง เราไม่สามารถหาช่วงความเชื่อมั่นที่แน่นอน หรือการทดสอบสมมติฐานที่แน่นอนได้ แต่สามารถใช้วิธีการประมาณได้ โดยอาศัยการคำนวณจากโปรแกรมคอมพิวเตอร์และจะให้ผลดีถ้าจำนวนตัวอย่างมีขนาดใหญ่ จำนวนชั้นอิสระเท่ากับ $n - p$ โดยที่ p คือจำนวนพารามิเตอร์ β ในตัวแบบสถิติ

3.1 100 (1−α) % ช่วงความเชื่อมั่นโดยประมาณสำหรับแต่ละ β_j คือ

$$C\,[\,\hat{\beta}_j - t_{1-\frac{\alpha}{2},n-p}\ ASE(\hat{\beta}_j) \le \beta_j \le \hat{\beta}_j + t_{1-\frac{\alpha}{2},n-p}\ ASE(\hat{\beta}_j)\,] \approx 1-\alpha$$

เมื่อ $ASE(\hat{\beta}_j)$ คือ ความคลาดเคลื่อนมาตรฐานโดยประมาณ (approximate standard error) ของ $\hat{\beta}_j$

3.2 สมมติฐานที่ต้องการทดสอบ คือ

$$H_0 : \beta_j = q \quad vs \quad H_1 : \beta_j \neq q$$

เมื่อ q คือ ค่าที่กำหนดให้

คำนวณค่าสถิติ $$t_C = \frac{\hat{\beta}_j - q}{ASE(\hat{\beta}_j)}$$

การสรุปผล เราจะปฏิเสธ H_0 ถ้า $|t_C|$ มากกว่า $t_{1-\frac{\alpha}{2}, n-p}$

ตัวอย่างที่ 2 อาศัยข้อมูลจากตัวอย่างที่ 1 เราได้ ASE $(\hat{\beta}_j)$ จากตารางผลลัพธ์การคำนวณของโปรแกรมสำเร็จรูป SPSS ดังนี้

$ASE(\hat{\beta}_1) = 0.1715$

$ASE(\hat{\beta}_2) = 0.2105$

$ASE(\hat{\beta}_3) = 0.1416$

วิธีทำ

1. สมมติว่าเราต้องการทดสอบสมมติฐาน

$$H_0 : \beta_3 = 0 \quad vs \quad H_1 : \beta_3 \neq 0$$

กำหนดระดับนัยสำคัญ $\alpha = .05$

ซึ่งถ้า $\beta_3 = 0$ แล้ว $\mu_Y(x)$ จะไม่ขึ้นกับตัวแปร X

คำนวณค่าสถิติ $$t_C = \frac{0.6828 - 0}{0.1416} = 4.82$$

เปิดค่า t จากตาราง $t_{.975, 9} = 2.262$

เปรียบเทียบค่าสถิติที่คำนวณได้กับค่าวิกฤติที่เปิดได้จากตารางได้ผลคือ $|t_C|$ มากกว่า $t_{.975, 9}$

จึงสรุปว่า ปฏิเสธ H_0 ที่ระดับนัยสำคัญ .05

2. ทดสอบสมมติฐาน

$$H_0 : \beta_2 \geq 3 \quad vs \quad H_1 : \beta_2 < 3$$

กำหนดระดับนัยสำคัญ $\alpha = .05$

คำนวณค่าสถิติ $$t_C = \frac{2.7233 - 3}{0.2105} = -1.314$$

เปิดค่า t จากตาราง $t_{.95, 9} = 1.833$

เปรียบเทียบค่าสถิติที่คำนวณได้กับค่าวิกฤติที่เปิดได้จากตารางได้ผลคือ t_C น้อยกว่า $t_{.95, 9}$

จึงสรุปได้ว่า ไม่สามารถปฏิเสธ H_0 ที่ระดับนัยสำคัญ .05

3. คำนวณหาช่วงความเชื่อมั่นโดยการประมาณ

ประมาณ 95% ของช่วงความเชื่อมั่นของ β_3 คือ

$$\hat{\beta}_3 \pm t_{0.975,9}\ \mathrm{ASE}\,(\hat{\beta}_3) = 0.6828 \pm 2.262\,(0.1416)$$

$$= (0.3625,\ 1.0031)$$

4. คำนวณหาช่วงความเชื่อมั่นโดยการประมาณ

ประมาณ 90% ของช่วงความเชื่อมั่นของ β_2 คือ

$$\hat{\beta}_2 \pm t_{.95,9}\ \mathrm{ASE}\,(\hat{\beta}_2) = 2.7233 \pm 1.833\,(0.2105)$$

$$= (2.3375,\ 3.1091)$$

และถ้าสนใจเฉพาะลิมิตบนของ β_2 จะได้ว่า

ช่วงความเชื่อมั่นที่ $\beta_2 \leq 3.10911 \approx 0.95$

5. คำนวณช่วงความเชื่อมั่นของ σ โดยการประมาณ

ประมาณ 95% ของช่วงความเชื่อมั่นของ σ คือ

$$C\left[\sqrt{\frac{SSE}{X^2_{.975,\,9}}} \leq \sigma \leq \sqrt{\frac{SSE}{X^2_{0.025,\,9}}}\right] \approx 0.95$$

$$C\left[\sqrt{\frac{0.4604}{19.023}} \leq \sigma \leq \frac{0.4604}{2.700}\right] \approx 0.95$$

$$C\left[0.156 \leq \sigma \leq 0.413\right] \approx 0.95$$

4. สมการถดถอยแบบโพลีโนเมียล (Polynomial Regressions)

มีการวิจัยจำนวนมากที่ต้องการหาความสัมพันธ์ระหว่าง 2 ตัวแปร คือ ตัวแปรต้นและตัวแปรตาม โดยตัวแปรต้นมีลักษณะเป็นปริมาณ ยกตัวอย่างการทดลองหนึ่งต้องการวัดการเจริญเติบโตของสัตว์เมื่อเพิ่มจำนวน nutrient ในอาหาร ทรีทเมนต์คือ แฟคเตอร์ตัวหนึ่งที่มีลักษณะเป็นปริมาณ ได้แก่ ปริมาณของ nutrient ในอาหาร 3 ระดับ คือ 0, 500, 1000 parts per million (ppm) ตัวแปรตาม คือ ลักษณะการเจริญเติบโตของสัตว์ ซึ่งเป็นฟังก์ชันของปริมาณของ nutrient ในอาหาร

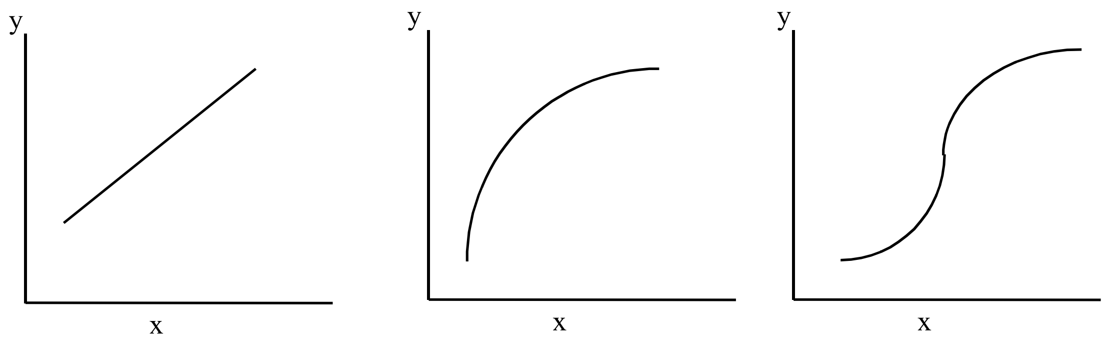
4.1 ฟังก์ชันโพลีโนเมียล

ความสัมพันธ์ระหว่างตัวแปรตาม Y และตัวแปรต้น X ซึ่งมีลักษณะเป็นปริมาณสามารถอธิบายได้ด้วยตัวแบบโพลีโนเมียล

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_p x^p + e$$

ตัวอย่างตัวแบบโพลีโนเมียล

$y = \beta_0 + \beta_1 x$ $\quad$ $y = \beta_0 + \beta_1 x + \beta_2 x^2$ $\quad$ $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$



(a) เทอมที่มีกำลังหนึ่ง (linear) (b) เทอมที่มีกำลังสอง (quadratic) (c) เทอมที่มีกำลังสาม (cubic)

ภาพ 6.2 ตัวอย่างของเส้นโค้งโพลีโนเมียล

ตัวอย่างที่ 3 ต้องการหาความสัมพันธ์ระหว่างผลผลิตและความหนาแน่นของปริมาณต้นพืช ซึ่งคาดว่าความสัมพันธ์จะมีลักษณะโพลีโนเมียล

ทรีทเมนต์คือ ความหนาแน่นของต้นพืชที่สนใจศึกษา 5 ระดับ คือ 10, 20, 30, 40 และ 50 เซ็นติเมตร หน่วยทดลองคือแปลงพืช สุ่มทรีทเมนต์ให้แปลงพืช ทรีทเมนต์ละ 3 แปลง ออกแบบการทดลองแบบสุ่มสมบูรณ์ เก็บข้อมูลปริมาณผลผลิตต่อแปลงดังแสดงในตารางต่อไปนี้

ตารางที่ 6.2 ข้อมูลปริมาณผลผลิตต่อแปลง (กิโลกรัมต่อไร่) จากความหนาแน่นของต้นพืช 5 ระดับ

	ความหนาแน่นของต้นพืช (x)				
	10	20	30	40	50
	12.2	16.0	18.6	17.6	18.0
	11.4	15.5	20.2	19.3	16.4
	12.4	16.5	18.2	17.1	16.6
ค่าเฉลี่ย	12.0	16.0	19.0	18.0	17.0

4.2 หาออธอกอนอลโพลีโนเมียล

ในตัวอย่างปัจจัยความหนาแน่นของต้นพืชมี $t = 5$ ระดับ คิดเป็นออธอกอนอลคอนทรัสได้ $(t - 1) = 4$ ทำการแปลงตัวแบบโพลีโนเมียลในรูปของ x ให้อยู่ในรูปออธอกอนอลโพลีโนเมียล คือ

$$y = \mu + \alpha_1 P_{1i} + \alpha_2 P_{2i} + \alpha_3 P_{3i} + \alpha_4 P_{4i} + e_{ij}$$

เมื่อ μ คือ ค่าเฉลี่ย

P_{ci} คือ ออธอกอนอลโพลีโนเมียลลำดับที่ c สำหรับทรีทเมนต์ระดับที่ i

การแปลงกำลังของ x เป็นออธอกอนอลโพลีโนเมียล

ค่าเฉลี่ย (mean) : $P_0 = 1$

เทอมที่มีกำลังหนึ่ง (linear) : $P_1 =$

เทอมที่มีกำลังสอง (quadratic) : $P_2 =$

เทอมที่มีกำลังสาม (cubic) : $P_3 =$

เมื่อ t คือ จำนวนระดับของปัจจัย

x คือ ค่าของระดับของปัจจัย

คือ ค่าเฉลี่ยของปัจจัยระดับต่าง ๆ

d คือ ระยะห่างระหว่างระดับต่าง ๆ ของปัจจัย

ตารางค่าของออธอกอนอลโพลีโนเมียลสำหรับ $t = 3$ ถึง $t = 10$ ดูจากตาราง Coefficients of Orthogonal Polynomials อยู่ในภาคผนวก

การแปลงออธอกอนอลโพลีโนเมียลสำหรับความหนาแน่นของต้นพืชด้วย $d = 10$, $t = 5$ $= 30$ และ $x = 10, 20, 30, 40$ หรือ 50 แต่ละเซตของสัมประสิทธิ์ P_1 ถึง P_4 อยู่ในรูปความแตกต่างระหว่างทรีทเมนต์ซึ่งผลบวกของสัมประสิทธิ์ในแต่ละคอลัมน์เท่ากับ 0

การประมาณสัมประสิทธิ์ α_c สำหรับสมการออธอกอนอลโพลีโนเมียล และผลบวกกำลังสองของออธอกอนอลโพลีโนเมียลคอนทรัส แสดงในตารางที่ 6.3 แล้วประมาณสมการ

ออธอกอนอลโพลีโนเมียลได้จากค่าประมาณของ μ , α_1 , α_2 , α_3 , α_4 ในตารางที่ 6.3

ตารางที่ 6.3 การคำนวณออธอกอนอลคอนทรัสและผลบวกกำลังสองของออธอกอนอลโพลีโนเมียลคอนทรัส

ความหนาแน่นของต้นพืช (x)		สัมประสิทธิของออธอกอนอลโพลีโนเมียล (P) mean P_0	linear P_1	quadratic P_2	cubic P_3	quartic P_4
10	12	1	−2	2	−1	1
20	16	1	−1	−1	2	−4
30	19	1	0	−2	0	6
40	18	1	1	−1	−2	−4
50	17	1	2	2	1	1
λ_c		−	1	1	5/6	35/12
ผลรวม =		82	12	−14	1	7
ตัวหาร =		5	10	14	10	70
SSP =		−	43.2	42.0	0.3	2.1
$\hat{\alpha}_c$ =	$\sum P_{ci}\bar{y}_i / \sum P_{ci}^2$	16.4	1.2	−1.0	0.1	0.1

จากตารางจะได้ค่าประมาณของสมการออธอกอนอลโพลีโนเมียล คือ

$$\hat{y}_i = 16.4 + 1.2\ P_{1i} - 1.0\ P_{2i} + 0.1\ P_{3i} + 0.1\ P_{4i}$$

4.3 การประมาณเส้นโค้ง

4.3.1 สมการออธอกอนอลโพลีโนเมียล

การประมาณเส้นโค้งควอดราทิกของปริมาณผลผลิต ซึ่งเป็นฟังก์ชันของความหนาแน่นของต้นพืช ประมาณได้จากสมการออธอกอนอลโพลีโนเมียลคือ

$$\hat{y}_i = \bar{y}\ldots + \hat{\alpha}_1 P_{1i} + \hat{\alpha}_2 P_{2i}$$
$$= 16.4 + 1.2\ P_{1i} - 1.0\ P_{2i}$$

4.3.2 การประมาณค่า y_i

แทนค่า P_1 และ P_2 ของสมการข้างต้นที่ระดับต่าง ๆ ของความหนาแน่นของต้นพืช (x) เพื่อใช้ในการประมาณค่าปริมาณผลผลิต $\hat{y}_i$ ผลการคำนวณแสดงไว้ในตารางที่ 6.4

ตารางที่ 6.4 ค่าสังเกตและค่าประมาณของปริมาณผลผลิตซึ่งได้จากสมการควอดราทิกออธอกอนอลโพลีโนเมียล

ความหนาแน่นของต้นพืช	สัมประสิทธิ์		ค่าประมาณ	ค่าสังเกต
x	P_1	P_2	$\hat{y}_i$	$\bar{y}_{i\cdot}$
10	−2	2	12.0	12
20	−1	−1	16.2	16
30	0	−2	18.4	19
40	1	−1	18.6	18
50	2	2	16.8	17

จากตารางที่ 6.4 ที่ความหนาแน่นของต้นพืช $x = 10$, $P_1 = -2$ และ $P_2 = 2$ แทนค่าในสมการได้

$$\hat{y} = 16.4 + 1.2(-2) - 1.0(2) = 12.0$$

4.3.3 วิธีการประมาณสมการเส้นโค้ง

โดยการแปลงออธอกอนอลโพลีโนเมียล P ในสมการออธอกอนอลโพลีโนเมียลให้เป็นสมการโพลีโนเมียลในรูปของ x จากตารางที่ 6.3 ซึ่งได้ค่า $\lambda_1 = 1$, $\lambda_2 = 1$, $d = 10$, $\bar{x} = 30$, และ $t = 5$ ดังนั้น แทนค่าในสมการสำหรับ P_i คือ

$$\hat{y} = 16.4 + 1.2\,P_1 - 1.0\,P_2$$

$$= 16.4 + 1.2\,(1)\left[\frac{x-30}{10}\right] - 1.0\,(1)\left[\left(\frac{x-30}{10}\right)^2 - \left(\frac{5^2-1}{12}\right)\right]$$

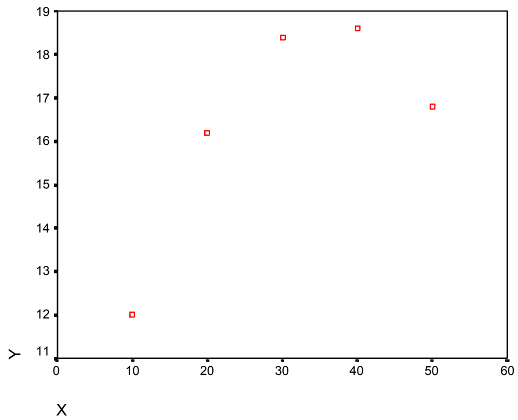
ดังนั้น $\hat{y} = 5.8 + 0.72\,x - 0.01\,x^2$

4.3.4 พล็อตกราฟของสมการโพลีโนเมียล

เราสามารถประมาณเส้นโค้งของปริมาณผลผลิตได้จากสมการข้างต้นและสามารถพล็อตกราฟได้โดยการแทนค่า x ในสมการโพลีโนเมียล

x	$y = 5.8 + 0.72x - 0.01x^2$

10	12.0
20	16.2
30	18.4
40	18.6
50	16.8



ภาพ 6.3 เส้นโค้งของปริมาณผลผลิตซึ่งเป็นฟังก์ชันของความหนาแน่นของต้นพืช

เส้นโค้งที่ได้นี้ใช้แสดงความสัมพันธ์ระหว่าง y และ x ที่มีค่าระหว่าง 10 ถึง 50

4.4 การวิเคราะห์ความแปรปรวน

การวิเคราะห์ความแปรปรวนสำหรับตัวอย่างที่ 3 คำนวณด้วยโปรแกรม SPSS ได้ผลการวิเคราะห์ดังตารางที่ 6.5

ตารางที่ 6.5 การวิเคราะห์ความแปรปรวนสำหรับตัวแบบออธอกอนอลโพลีโนเมียลแสดงความสัมพันธ์ระหว่างปริมาณผลผลิตกับความหนาแน่นของต้นพืช

	Sum of Squares	df	Mean Square	F	Sig.

Between Groups	(Combined)		87.600	4	21.900	29.278	.000
	Linear Term	Contrast	43.200	1	43.200	57.754	.000
		Deviation	44.400	3	14.800	19.786	.000
	Quadratic	Contrast	42.000	1	42.000	56.150	.000
	Term	Deviation	2.400	2	1.200	1.604	.249
	Cubic Term	Contrast	.300	1	.300	.401	.541
		Deviation	2.100	1	2.100	2.807	.125
Within Groups			7.480	10	.748		
Total			95.080	14			

4.5 การทดสอบสมมติฐานทางสถิติ

4.5.1 สมมติฐานที่ต้องการทดสอบคือ

$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ VS $H_1 : \alpha_i \neq \alpha_j$ สำหรับบาง i , j ที่ระดับนัยสำคัญ $\alpha = .05$

ค่าสถิติทดสอบคือ $$F_0 = \frac{MSTr}{MSE} = \frac{21.90}{0.75} = 29.28$$

เขตวิกฤติ คือ $F_0 > F_{.05,4,10} = 3.48$

เนื่องจาก F_0 ที่คำนวณได้มากกว่าค่าวิกฤติ $F_{.05,4,10}$ จึงสรุปว่าปฏิเสธ H_0

4.5.2 หลังจากนั้นเราจะทดสอบว่าสมการของผลการทดลองมีลักษณะเป็นแบบใด โดยทดสอบเรียงลำดับจาก linear , quadratic , cubic , และต่อไปเรื่อย ๆ

สมมุติฐานที่ใช้ทดสอบต่อไป เรียงตามลำดับคือ

$H_0 : \alpha_1 = 0$, $H_0 : \alpha_2 = 0$, $H_0 : \alpha_3 = 0$ และ $H_0 : \alpha_4 = 0$

สถิติทดสอบ คือ F test

ค่าสถิติทดสอบคือ $$F_0 = \frac{MSC}{MSE}$$

ค่าสถิติของโพลีโนเมียลคอนทรัสแต่ละตัวคำนวณให้แล้วอยู่ในตารางที่ 6.5

สรุปผลการทดสอบสมมติฐานได้ว่า ปฏิเสธ H_0 สำหรับตัวแบบเชิงเส้นตรง และตัวแบบโพลีโนเมียลที่มีเทอมยกกำลังสอง (Sig of F = .000) แต่ไม่ปฏิเสธ H_0 สำหรับตัวแบบโพลีโนเมียลที่มีเทอมยกกำลังสาม (Sig of F = .541) หมายความว่า ความสัมพันธ์ระหว่างปริมาณผลผลิตและความหนาแน่นของต้นพืช อธิบายได้ด้วยตัวแบบโพลีโนเมียลที่มีเทอมยกกำลังสอง

4.6 ความคลาดเคลื่อนมาตรฐานของเส้นโค้ง

ตัวอย่างที่ 4 อาศัยการศึกษาจากตัวอย่างที่ 3 เกี่ยวกับการประมาณความสัมพันธ์ระหว่างปริมาณผลผลิตและความหนาแน่นของต้นพืช ด้วยสมการควอดราทิกซึ่งต้องประมาณค่า 3 ค่า คือ $\bar{y}_{\bullet\bullet}$ = 16.4 , $\bar{\alpha}_1 = 1.2$, $\bar{\alpha}_2 = -1.0$ ต้องการประมาณค่าความแปรปรวนสำหรับค่าประมาณของปริมาณผลผลิตต่อแปลง ($\hat{y}$) ที่มีความหนาแน่นของต้นพืช 5 ระดับจากสมการออธอกอนอลโพลีโนเมียล

$$\hat{y} = \bar{y}_{\bullet\bullet} + \hat{\alpha}_1 P_1 + \hat{\alpha}_2 P_2$$

วิธีทำ

1. ค่าประมาณความแปรปรวนสำหรับค่าประมาณของปริมาณผลผลิตต่อแปลงคือ

$$s^2_{\hat{y}} = s^2_{\bar{y}_{\bullet\bullet}} + s^2_{\hat{\alpha}_1} P_1^2 + s^2_{\hat{\alpha}_2} P_2^2$$

2. ค่าประมาณความแปรปรวนสำหรับโพลีโนเมียลคอนทรัส คือ

$$s^2_c = \frac{s^2}{r\sum P^2_{ci}}$$

3. ความคลาดเคลื่อนมาตรฐานสำหรับโพลีโนเมียลคอนทรัส คือ

$$s_c = \sqrt{s^2_c}$$

4. จากตารางวิเคราะห์ความแปรปรวนในตารางที่ 6.5 ได้ค่าประมาณความแปรปรวนคือ

$$s^2 = \text{MSE} = 0.75$$

5. และจากตัวหารในตารางที่ 6.3 เราสามารถหาความแปรปรวนของแต่ละเทอมได้ดังนี้

$$s^2_{\bar{y}_{\bullet\bullet}} = \frac{0.75}{(3)(5)} = 0.050$$

$$s^2_{\hat{\alpha}_1} = \frac{0.75}{(3)(10)} = 0.025$$

$$s^2_{\hat{\alpha}_2} = \frac{0.75}{(3)(14)} = 0.018$$

6. จะได้ค่าประมาณความแปรปรวนสำหรับค่าประมาณของปริมาณผลผลิตต่อแปลงคือ

$$s^2_{\hat{y}} = s^2_{\bar{y}_{\bullet\bullet}} + s^2_{\hat{\alpha}_1} P_1^2 + s^2_{\hat{\alpha}_2} P_2^2$$

$$= 0.05 + 0.025 P_1^2 + 0.018 P_2^2$$

ค่าของ P_1 และ P_2 จะมีค่าแตกต่างกันไปตามค่าของ x (ความหนาแน่นของต้นพืช) สำหรับ x = 10 จะใช้ $P_1 = -2$ และ $P_2 = 2$ แทนค่าในสมการ

$$s_{\hat{y}}^2 = 0.05 + 0.025\,(-2)^2 + 0.018\,(2)^2 = 0.222$$

7. จะได้ความคลาดเคลื่อนมาตรฐาน สำหรับ x = 10 คือ

$$s_{\hat{y}} = \sqrt{0.222}$$
$$= 0.471$$

8. 100 $(1-\alpha)$ % ช่วงความเชื่อมั่นของ $\hat{y}$ คือ

$$\hat{y} \pm t_{\alpha/2,N-t}\left(s_{\hat{y}}\right)$$

ที่ระดับนัยสำคัญ $\alpha = .05$ ความคลาดเคลื่อนมาตรฐานและ 95% ช่วงความเชื่อมั่นสำหรับค่าประมาณของปริมาณผลผลิตที่แต่ละระดับของความหนาแน่นของต้นพืชทั้ง 5 ระดับใช้ $t_{.025,10} = 2.228$ ได้ผลดังตาราง 6.6

ตารางที่ 6.6 ค่าเฉลี่ยและค่าประมาณของปริมาณผลผลิต และความคลาดเคลื่อนมาตรฐานสำหรับค่าประมาณของปริมาณผลผลิตของแต่ละระดับของความหนาแน่นของต้นพืช

ความหนาแน่นของต้นพืช x	ปริมาณผลผลิต		ความคลาดเคลื่อนมาตรฐาน	95% ช่วงความเชื่อมั่น
	ค่าเฉลี่ย $\bar{y}_{i\cdot}$	ค่าประมาณ $\hat{y}_i$		
10	12	12.0	0.471	(10.95 , 13.05)
20	16	16.2	0.305	(15.52 , 16.88)
30	19	18.4	0.349	(17.62 , 19.18)
40	18	18.6	0.305	(17.92 , 19.28)
50	17	16.8	0.471	(15.75 , 17.85)