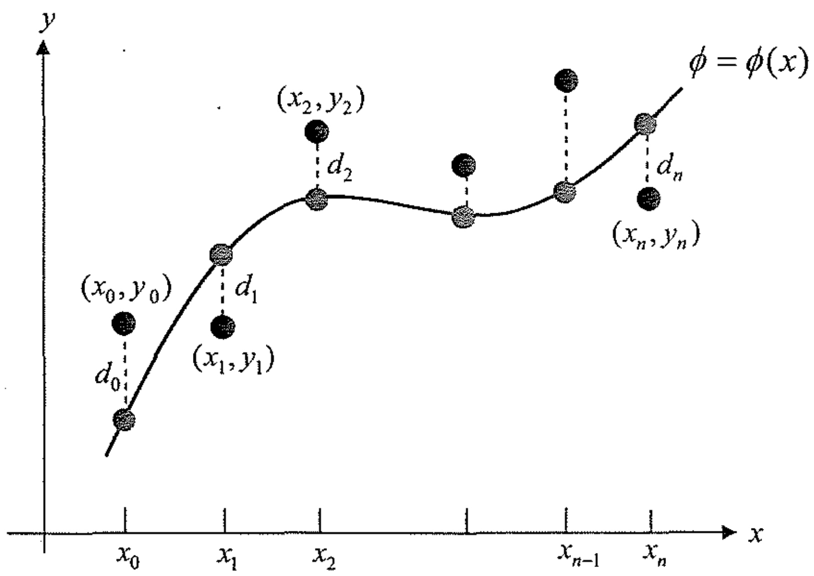
สำหรับเซตของจุดข้อมูล

$$\{(x_0, y_0), (x_1, y_1), \ldots, (x_n, y_n)\}$$

ให้ $\phi = \phi(x)$ เป็นฟังก์ชันที่สร้างขึ้นเพื่อเชื่อมโยงค่า $x_k$ กับ $y_k$ หรือฟิตข้อมูลชุดนี้ ดังนั้น ค่าคลาดเคลื่อนที่แต่ละ $x_0, x_1, \ldots, x_n$ คือ

$$\begin{aligned} d_0 &= y_0 - \phi_0 = y_0 - \phi(x_0) \\ d_1 &= y_0 - \phi_1 = y_1 - \phi(x_1) \\ &\vdots \\ d_n &= y_n - \phi_n = y_n - \phi(x_n) \end{aligned} \tag{5.1.1}$$

ค่า $d_0, d_1, d_2, \ldots, d_n$ เป็นระยะในแนวตั้งจากจุดข้อมูลไปยังจุดที่อยู่บนกราฟของ $\phi = \phi(x)$ ดังแสดงในรูปที่ 5.1.1




**รูปที่ 5.1.1**

เพื่อให้ค่าคลาดเคลื่อน $d_0, d_1, d_2, \ldots, d_n$ น้อยที่สุด มีแนวทางในการวัดค่าคลาดเคลื่อนเหล่านี้ในรูปผลบวกได้ เช่น

$$S = \sum_{k=0}^{n} d_k \tag{5.1.2}$$

$$S = \sum_{k=0}^{n} |d_k| \tag{5.1.3}$$

$$S = \sum_{k=0}^{n} d_k^2 \tag{5.1.4}$$

สำหรับค่า $S$ ในสมการ (5.1.2) อาจมีการหักล้างกันของค่า $d_0, d_1, d_2, \ldots, d_n$ ได้ ในขณะที่ $S$ ในสมการ (5.1.3) ในทางคณิตศาสตร์เป็นฟังก์ชันที่ไม่เรียบ ไม่เหมาะที่จะหาอนุพันธ์ สำหรับ $S$ ในสมการ (5.1.4) มีความเหมาะสมที่จะนำมาใช้เป็นเงื่อนไขในการสร้างฟังก์ชัน $\phi(x)$ โดยให้ค่า $S$ ในสมการ (5.1.4) มีค่าน้อยที่สุด ทำให้การฟิตข้อมูลด้วยเงื่อนไขนี้มีชื่อเรียกว่า วิธีกำลังสองน้อยสุด

ดังนั้น โดยสรุปแล้ว การฟิตข้อมูลโดยวิธีกำลังสองน้อยสุดคือ การสร้างฟังก์ชัน $\phi = \phi(x)$ สำหรับข้อมูล $\{(x_0, y_0), (x_1, y_1), \ldots, (x_n, y_n)\}$ โดยทำให้ผลบวกของค่าคลาดเคลื่อนกำลังสอง

$$S = \sum_{k=0}^{n} d_k^2 \quad \text{มีค่าน้อยที่สุด}$$

เมื่อ $d_k = y_k - \phi(x_k),\ \ k = 0, 1, 2, \ldots, n$ การหาเงื่อนไขที่ทำให้ $S$ นี้มีค่าต่ำสุด คือการหาอนุพันธ์แล้วหาค่าวิกฤต กรณีเฉพาะของฟังก์ชัน $\phi = \phi(x)$ จะเสนอในหัวข้อ 5.2 - 5.4

## 5.2 การถดถอยเชิงเส้น (Linear Regression)

การถดถอยเชิงเส้นคือ วิธีกำลังสองน้อยสุดสำหรับฟิตข้อมูล

$$\{(x_0, y_0), (x_1, y_1), \ldots, (x_n, y_n)\}$$

ด้วยฟังก์ชันเชิงเส้น

$$\phi(x) = a_0 + a_1 x \tag{5.2.1}$$

โดยเงื่อนไขของวิธีกำลังสองน้อยสุดในหัวข้อ 5.1 ต้องหาค่า $a_0$ และ $a_1$ ที่ทำให้

$$S = \sum_{k=0}^{n} d_k^2 \tag{5.2.2}$$

มีค่าต่ำสุด โดยที่ $d_k$ ในกรณีนี้คือ ระยะในแนวดิ่งจากจุดข้อมูล $(x_k, y_k)$ ไปยังจุด $(x_k, \phi(x_k))$ ซึ่งอยู่บนกราฟของเส้นตรง $\phi(x) = a_0 + a_1 x$ ดังแสดงในรูปที่ 5.1.2 แทนค่า $d_k$ ได้

$$S = \sum_{k=0}^{n} d_k^2 = \sum_{k=0}^{n} \left(y_k - \phi(x_k)\right)^2 = \sum_{k=0}^{n} \left(y_k - (a_0 + a_1 x_k)\right)^2 \tag{5.2.3}$$

ซึ่งแสดงว่า $S$ เป็นฟังก์ชันของ $a_0$ และ $a_1$ ต่อไปหาจุดวิกฤตของ $S$ ซึ่งในกรณีนี้เป็นจุดต่ำสุดสัมบูรณ์ โดยหาอนุพันธ์ย่อยของ $S$

$$\frac{\partial S}{\partial a_0} = -2\sum_{k=0}^{n} \left(y_k - a_0 - a_1 x_k\right) \tag{5.2.4}$$

$$\frac{\partial S}{\partial a_1} = -2\sum_{k=0}^{n} x_k \left(y_k - a_0 - a_1 x_k\right) \tag{5.2.5}$$

ให้อนุพันธ์ย่อยทั้งสองเท่ากับศูนย์ ทำให้ได้สมการ

$$(n+1)a_0 + \left(\sum_{k=0}^{n} x_k\right)a_1 = \sum_{k=0}^{n} y_k \tag{5.2.6}$$

$$\left(\sum_{k=0}^{n} x_k\right)a_0 + \left(\sum_{k=0}^{n} x_k^2\right)a_1 = \sum_{k=0}^{n} x_k y_k \tag{5.2.7}$$

สมการ (5.2.6) และ (5.2.7) มีชื่อเรียกว่า สมการปรกติ (normal equations) สำหรับการถดถอยเชิงเส้น เอา $n+1$ หารสมการทั้งสอง แก้สมการได้ค่า $a_1$ และ $a_0$ ดังนี้

$$a_1 = \frac{\overline{x}\,\overline{y} - \overline{xy}}{(\overline{x})^2 - \overline{x^2}} \tag{5.2.8}$$

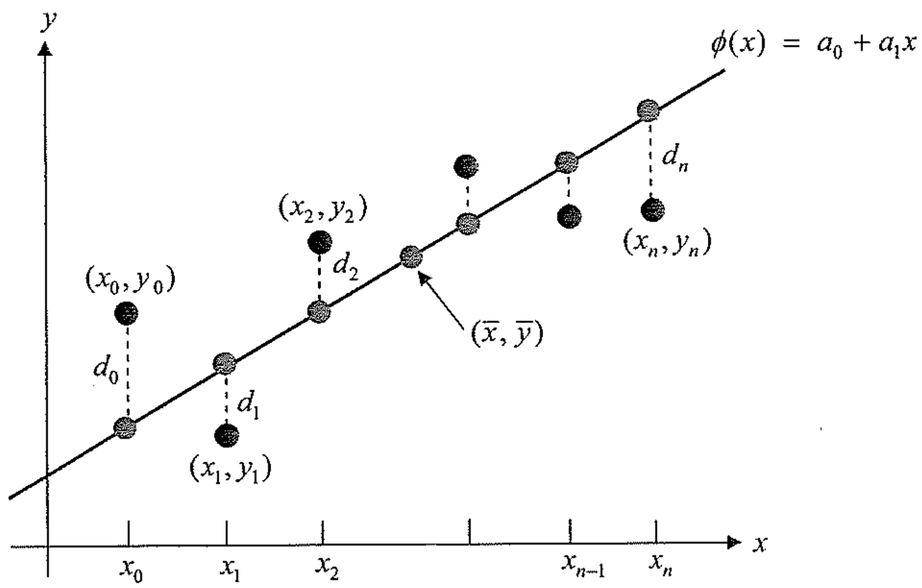
$$a_0 = \overline{y} - a_1\overline{x} \tag{5.2.9}$$

โดยที่

$$\overline{x} = \frac{1}{n+1}\sum_{k=0}^{n} x_k, \qquad \overline{y} = \frac{1}{n+1}\sum_{k=0}^{n} y_k,$$

$$\overline{xy} = \frac{1}{n+1}\sum_{k=0}^{n} x_k y_k, \qquad \overline{x^2} = \frac{1}{n+1}\sum_{k=0}^{n} x_k^2,$$

นั่นคือ ค่าเหล่านี้คือค่าเฉลี่ยนั่นเอง จากสมการ (5.2.9) ยังได้ว่า จุด $(\overline{x}, \overline{y})$ อยู่บนเส้นตรง $\phi(x) = a_0 + a_1 x$ ซึ่งเรียกว่า เส้นตรงกำลังสองน้อยสุด (least-squares line) หรือ เส้นตรงถดถอย (regression line) ดูรูปที่ 5.2.1




**รูปที่ 5.2.1** เส้นตรงถดถอย

การคำนวณค่า $a_1$ และ $a_0$ สำหรับเส้นตรงถดถอย $y = a_0 + a_1 x$ เพื่อฟิตข้อมูล $\{(x_0, y_0), (x_1, y_1), \ldots, (x_n, y_n)\}$ สามารถจัดระเบียบการคำนวณในรูปดังตารางที่ 5.2.1 โดยสองคอลัมน์แรกบันทึกข้อมูล $x_k$ และ $y_k$ สำหรับคอลัมน์ที่ 3 และ 4 คำนวณค่า $x_k^2$ และผลคูณ $x_k y_k$ ตามลำดับ

| | $x_k$ | $y_k$ | $x_k^2$ | $x_k y_k$ |
|---|---|---|---|---|
| | $x_0$<br>$x_1$<br>$\vdots$<br>$x_n$ | $y_0$<br>$y_1$<br>$\vdots$<br>$y_n$ | $x_0^2$<br>$x_1^2$<br>$\vdots$<br>$x_n^2$ | $x_0 y_0$<br>$x_1 y_1$<br>$\vdots$<br>$x_n y_n$ |
| ผลบวก | $\sum_{k=0}^{n} x_k$ | $\sum_{k=0}^{n} y_k$ | $\sum_{k=0}^{n} x_k^2$ | $\sum_{k=0}^{n} x_k y_k$ |
| ค่าเฉลี่ย | $\bar{x}$ | $\bar{y}$ | $\overline{x^2}$ | $\overline{xy}$ |

**ตารางที่ 5.2.1**

**ตัวอย่างที่ 5.2.1** จงฟิตข้อมูลในตารางด้วยเส้นตรงกำลังสองน้อยสุด แล้วประมาณค่าของ $g(0.05)$

| $x_k$ | 0.0 | 0.1 | 0.3 | 0.4 | 0.5 |
|---|---|---|---|---|---|
| $g(x_k)$ | 0.5 | 1.4 | 2.0 | 2.5 | 3.1 |

**วิธีทำ**

บันทึกข้อมูล $x_k$ และ $y_k = g(x_k)$ จากตารางที่กำหนดมาให้ คำนวณค่า $x_k^2$ และผลคูณ $x_k y_k$ ดังรูปแบบในตารางที่ 5.2.1 แล้วคำนวณผลบวกและค่าเฉลี่ยในแต่ละคอลัมน์ ได้ผลดังนี้

| | $x_k$ | $y_k$ | $x_k^2$ | $x_k y_k$ |
|---|---|---|---|---|
| | 0.0 | 0.5 | 0.00 | 0.00 |
| | 0.1 | 1.4 | 0.01 | 0.14 |
| | 0.3 | 2.0 | 0.09 | 0.60 |
| | 0.4 | 2.5 | 0.16 | 1.00 |
| | 0.5 | 3.1 | 0.25 | 1.55 |
| ผลบวก | 1.3 | 9.5 | 0.51 | 3.29 |
| ค่าเฉลี่ย | 0.26 | 1.9 | 0.102 | 0.658 |

แทนค่าในสมการ (5.2.8) และ (5.2.9) ได้

$$a_1 = \frac{(0.26)(1.9) - 0.658}{(0.26)^2 - 0.102} = 4.8$$

$$a_0 = 1.9 - (0.26)(4.8) = 0.65$$

ดังนั้น เส้นตรงกำลังสองน้อยสุด คือ

$$\phi(x) = 0.65 + 4.8x$$

ประมาณค่าของ $g(0.05)$ ได้

$$g(0.05) \cong \phi(0.05) = 0.65 + 4.8(0.05) = 0.89$$ □

**ตัวอย่างที่ 5.2.2** จงฟิตข้อมูลในตารางด้วยเส้นตรงกำลังสองน้อยสุด แล้วคำนวณผลบวกของค่าคลาดเคลื่อนกำลัง

| $x_k$ | 0.25 | 0.50 | 0.75 | 1.00 | 1.25 | 1.50 | 1.75 |
|---|---|---|---|---|---|---|---|
| $y_k$ | 3.10 | 1.70 | 1.00 | 0.68 | 0.42 | 0.26 | 0.14 |

**วิธีทำ**

บันทึกข้อมูล $x_k$ และ $y_k$ จากตารางที่กำหนดมาให้ คำนวณค่า $x_k^2$ และผลคูณ $x_k y_k$ ดังรูปแบบในตารางที่ 5.2.1 แล้วคำนวณผลบวกและค่าเฉลี่ยในแต่ละคอลัมน์ ได้ผลดังนี้

| | $x_k$ | $y_k$ | $x_k^2$ | $x_k y_k$ |
|---|---|---|---|---|
| | 0.25 | 3.10 | 0.0625 | 0.7750 |
| | 0.50 | 1.70 | 0.2500 | 0.8500 |
| | 0.75 | 1.00 | 0.5625 | 0.7500 |
| | 1.00 | 0.68 | 1.0000 | 0.6800 |
| | 1.25 | 0.42 | 1.5625 | 0.5250 |
| | 1.50 | 0.26 | 2.2500 | 0.3900 |
| | 1.75 | 0.14 | 3.0625 | 0.2450 |
| ผลบวก | 7.0000 | 7.3000 | 8.7500 | 4.2150 |
| ค่าเฉลี่ย | 1.0000 | 1.0429 | 1.2500 | 0.6021 |

แทนค่าในสมการ (5.2.8) และ (5.2.9) ได้

$$a_1 = \frac{(1.0000)(1.0429) - (0.6021)}{(1.0000)^2 - (1.2500)} = -1.7629$$

$$a_0 = 1.0429 - (-1.7629)(1.0000) = 2.8057$$

ดังนั้น เส้นตรงกำลังสองน้อยสุด คือ

$$\phi(x) = 2.8057 - 1.7629x$$

ค่าคลาดเคลื่อนกำลังสองที่จุดต่อแต่ละจุดคือ

$$d_k^2 = \left(y_k - \phi(x_k)\right)^2 = \left(y_k - (2.8057 - 1.7629x_k)\right)^2$$

ได้ผลการคำนวณดังตาราง

| $x_k$ | 0.25 | 0.50 | 0.75 | 1.00 | 1.25 | 1.50 | 1.75 |
|---|---|---|---|---|---|---|---|
| $d_k^2$ | 0.5402 | 0.0503 | 0.2338 | 0.1317 | 0.0332 | 0.0097 | 0.1758 |

และผลบวกของค่าคลาดเคลื่อนกำลังสองคือ

$$\sum_{k=0}^{6} d_k^2 = 1.1747 \qquad \square$$

## 5.3 การแปลงให้เป็นเชิงเส้น (Linearization)

การฟิตข้อมูล $\{(x_0, y_0), (x_1, y_1), \ldots, (x_n, y_n)\}$ ในกรณีที่ความสัมพันธ์ของ $y_k$ และ $x_k$ ไม่เป็นแบบเชิงเส้น เช่น กรณีที่อัตราการเปลี่ยนแปลงของ $y$ แปรตามปริมาณของ $y$ ตัวมันเอง ซึ่งเกิดขึ้นในปัญหาทางวิศวกรรม ในกรณีเช่นนี้ ฟังก์ชันที่ใช้ฟิตข้อมูลเป็นแบบฟังก์ชันเลขชี้กำลัง(exponential model)

$$\phi(x) = Ae^{Bx} \tag{5.3.1}$$

ซึ่งสามารถแปลงให้เป็นเชิงเส้นได้ โดยใช้ฟังก์ชันลอการิทึมดังนี้

$$\begin{aligned} \ln(\phi(x)) &= \ln\left(Ae^{Bx}\right) \\ &= \ln A + Bx \end{aligned}$$

โดยให้

$$Y = \ln(\phi(x)), \quad X = x, \quad a_0 = \ln A, \quad a_1 = B$$

ผลที่ตามมาคือ การแปลงฟังก์ชันเลขชี้กำลังในสมการ (5.3.1) เป็นรูปแบบฟังก์ชันเชิงเส้น

$$Y = a_0 + a_1 X \tag{5.3.2}$$

ซึ่งทำให้สามารถคำนวณค่า $a_0$ และ $a_1$ จากสูตรของเส้นตรงกำลังสองน้อยสุดได้ เมื่อได้ค่าทั้งสองแล้ว ก็สามารถหาค่า $A$ และ $B$ แล้วแปลงกลับไปที่ฟังก์ชันเลขชี้กำลังในสมการ (5.3.1) ได้

ในกรณีฟิตข้อมูลด้วยฟังก์ชันกำลังหรือกรณีอื่นๆ พิจารณาการแปลงให้เป็นแบบเชิงเส้น ดังสรุปไว้ในตารางที่ 5.3.1

| ฟังก์ชันไม่เชิงเส้น | การแปลงให้เป็นฟังก์ชันเชิงเส้น |
|---|---|
| $y = Ae^{Bx}$ | $Y = a_0 + a_1X$<br>$X = x, \quad Y = \ln y, \quad a_0 = \ln A, \quad a_1 = B$ |
| $y = Ax^n$ | $Y = a_0 + a_1X$<br>$X = \ln x, \quad Y = \ln y, \quad a_0 = \ln A, \quad a_1 = n$ |
| $y = A + \dfrac{B}{x}$ | $Y = a_0 + a_1X$<br>$X = \dfrac{1}{x}, \quad Y = y, \quad a_0 = A, \quad a_1 = B$ |
| $y = \dfrac{1}{A + Bx}$ | $Y = a_0 + a_1X$<br>$X = x, \quad Y = \dfrac{1}{y}, \quad a_0 = A, \quad a_1 = B$ |
| $y = A\dfrac{x}{B + x}$ | $Y = a_0 + a_1X$<br>$X = \dfrac{1}{x}, \quad Y = \dfrac{1}{y}, \quad a_0 = \dfrac{1}{A}, \quad a_1 = \dfrac{B}{A}$ |

**ตารางที่ 5.3.1**

**ตัวอย่างที่ 5.3.1** จงฟิตข้อมูลในตารางด้วยฟังก์ชันเลขชี้กำลัง $\phi(x) = Ae^{Bx}$ และคำนวณผลบวกของค่าคลาดเคลื่อนกำลังสอง

| $x_k$ | 0.25 | 0.50 | 0.75 | 1.00 | 1.25 | 1.50 | 1.75 |
|---|---|---|---|---|---|---|---|
| $y_k$ | 3.10 | 1.70 | 1.00 | 0.68 | 0.42 | 0.26 | 0.14 |

ซึ่งเป็นข้อมูลชุดเดียวกับตัวอย่างที่ 5.2.2 โดยเปรียบเทียบผลบวกของค่าคลาดเคลื่อนกำลังสองกับผลของการฟิตข้อมูลชุดเดียวกันนี้ด้วยฟังก์ชันเชิงเส้น

**วิธีทำ**

ฟังก์ชันที่ใช้ฟิตข้อมูลคือ

$$\phi(x) = Ae^{Bx}$$

แปลงเป็นแบบเชิงเส้นโดย

$$\ln\phi(x) = \ln\left(Ae^{Bx}\right) = \ln A + Bx$$

ให้ $Y = \ln\phi(x)$, $X = x$, $a_0 = \ln A$ และ $a_1 = B$ ดังนั้น

$$Y = a_0 + a_1X$$

จากตารางที่กำหนดมาให้ คำนวณ $Y_k = \ln y_k$ แล้วบันทึก $X_k = x_k$ และ $Y_k$ ในคอลัมน์ที่ 1 และ 2 ของตาราง คำนวณค่า $X_k^2$ และผลคูณ $X_kY_k$ บันทึกในตารางในคอลัมน์ที่ 3 และ 4 แล้วคำนวณผลบวกและค่าเฉลี่ยในแต่ละคอลัมน์ได้ผลดังแสดงในตาราง

| | $X_k = x_k$ | $Y_k = \ln(y_k)$ | $X_k^2$ | $X_kY_k$ |
|---|---|---|---|---|
| | 0.25 | 1.131402 | 0.062500 | 0.282851 |
| | 0.50 | 0.530628 | 0.250000 | 0.265314 |
| | 0.75 | 0.000000 | 0.562500 | 0.000000 |
| | 1.00 | -0.385662 | 1.000000 | -0.385662 |
| | 1.25 | -0.867501 | 1.562500 | -1.084376 |
| | 1.50 | -1.347074 | 2.250000 | -2.020610 |
| | 1.75 | -1.966113 | 3.062500 | -3.440697 |
| ผลบวก | 7.00 | -2.904319 | 8.750000 | -6.383182 |
| ค่าเฉลี่ย | 1.00 | -0.414903 | 1.250000 | -0.911883 |

แทนค่าในสมการ (5.2.8) และ (5.2.9) ได้

$$a_1 = \frac{(1.00)(-0.414903) - (-0.911883)}{(1.00)^2 - (1.250000)} = -1.987921$$

$$a_0 = -0.414903 - (-1.987921)(1.00) = 1.573019$$

ดังนั้น เส้นตรงกำลังสองน้อยสุด $Y = a_0 + a_1X$ คือ

$$Y = 1.573019 - 1.987921X$$

เพราะว่า $a_0 = \ln A$ ดังนั้น

$$A = e^{a_0} = e^{1.573019} = 4.821179$$

และ

$$B = a_1 = -1.987921$$