CHAPTER 5

การพยากรณ์ (Forecasting)

การพยากรณ์ (Forecasting) คือการใช้ข้อมูลของสิ่งที่สนใจที่เกิดขึ้นในอดีตเพื่อสร้างแบบจำลองคณิตศาสตร์ที่ใช้ในการคำนวณสิ่งที่สนใจนั้นในอนาคต และในบางตำราจะยังรวมถึงการใช้สภาวะการณ์ที่สำรวจได้ในปัจจุบันรวมไปถึงในอดีตเพื่อทำนาย (Prediction) ค่าที่สนใจ ตัวอย่างเช่น

- ฝ่ายขายใช้ยอดขายย้อนหลัง 12 เดือนเพื่อทำนายยอดขายในเดือนถัดไป
- ทีมการตลาดต้องการทำนายความต้องการซื้อสินค้าบางอย่างของลูกค้าโดยอาศัยคุณลักษณะต่าง ๆ เช่นเพศ อายุ ฐานเงินเดือน สินค้าที่ซื้อในเดือนที่แล้ว เป็นต้น

ในรายวิชานี้ เราจะศึกษา 2 รูปแบบการทำนายหลัก ๆ ดังนี้

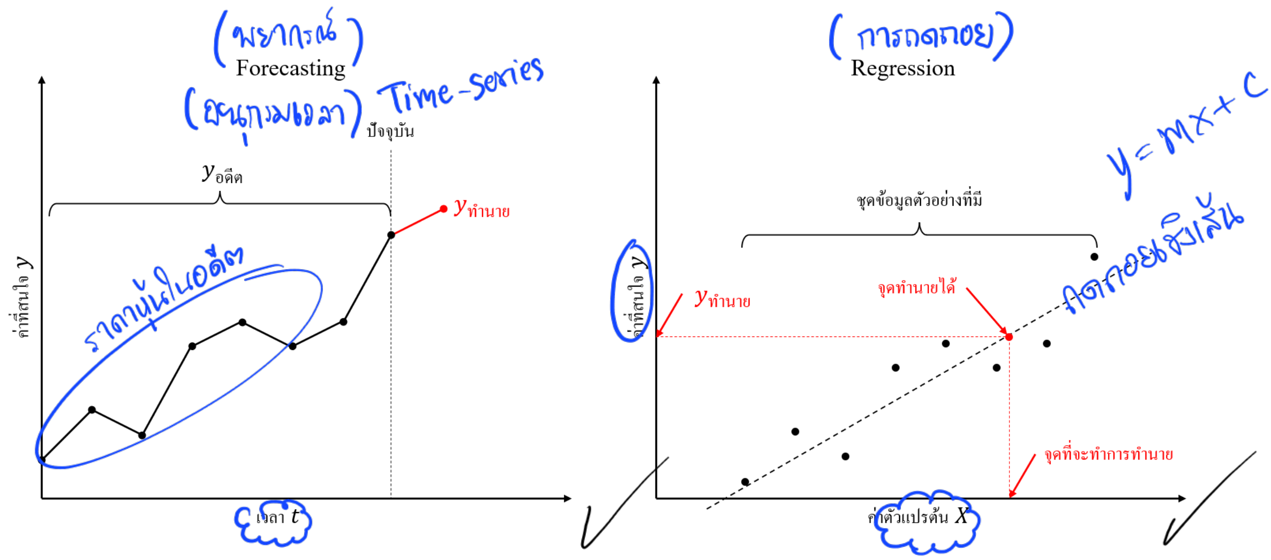


Figure 5.1. ความ แตก ต่าง ระหว่าง ตัว แบบ อนุกรม เวลา และ ตัว แบบ ถดถอย

1. **ตัวแบบอนุกรมเวลา** (time series forecasting): เป็นตัวแบบที่อยู่บนสมมติฐานว่าตัวแปรที่เราสนใจมีค่าขึ้นอยู่กับตัวแปรเดียวกันที่เกิดขึ้นในอดีต

$$\hat{y}_t \sim y_{t-1}, y_{t-2}, \ldots, y_{t-k}$$

2. **ตัวแบบการถอดถอย** (regression model): เป็นตัวแบบที่อยู่บนสมมติฐานว่าตัวแปรที่เราสนใจมีค่าขึ้นอยู่กับตัวแปรอื่น ๆ ที่เกิดขึ้นพร้อมกัน (หรือเป็นอยู่)

$$\hat{y}_t \sim x_{1,t}, x_{2,t}, \ldots, x_{n,t}$$

ทั้งนี้ เราไม่สามารถบอกได้ว่าวิธีการใดเป็นวิธีการที่ดีที่สุด เพราะแต่ละวิธีการ (ที่กำลังจะกล่าวในหัวข้อถัดไป) มีสมมติฐานตั้งต้นที่แตกต่างกัน ขึ้นอยู่กับลักษณะของข้อมูลว่าจะเหมาะสมกับตัวแบบไหน แต่ในทางปฏิบัติ ถ้าไม่มีได้ติดขัดเรื่องปัญหาด้านทรัพยากรในการทำการคำนวณ เรามักจะลองทุกวิธีการและวัดผลเพื่อตรวจสอบเปรียบเทียบความสามารถของแต่ละตัวแบบ (เรียนในหัวข้อสุดท้าย)

y = ค่าจริง

$\hat{y}$ (y hat) = ค่าทำนาย

5.1 ตัวแบบอนุกรมเวลา

สมมติเป็นเดือน

$$\hat{y}_{\text{กันยา}} = \frac{y_{\text{มกรา}} + y_{\text{กุมภา}} + \cdots + y_{\text{สิงหา}}}{8}$$

5.1.1 วิธีการค่าเฉลี่ยรวม

ผลรวม / จำนวนตัว

- ◇ เหมาะกับข้อมูลที่มีลักษณะที่ค่อนข้างคงที่ในภาพรวม (ไม่ได้มีแนวโน้มที่เปลี่ยนแปลงไปเช่นตลาดโตขึ้นเรื่อย ๆ)
- ◇ วิธีการคำนวณ:

$$\hat{y}_{t+1} = \frac{\sum_{i=1}^{t} y_i}{t} = \frac{y_1 + y_2 + \cdots + y_t}{t}$$

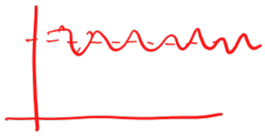
ตัวอย่าง 5.1.1: วิธีการค่าเฉลี่ยรวม

บริษัทหนึ่งมีความต้องการทำนายยอดขายในเดือนที่ 7 โดยที่มียอดขาย 6 เดือนที่ผ่านมาตามตารางด้านล่างนี้ ทั้งนี้ลองหาค่าทำนายของแต่ละเดือนก่อนหน้าด้วย

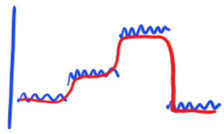
เดือน	ยอดขาย
1	800
2	900
3	800
4	1000
5	1000
6	1300

$$\hat{y}_7 = \frac{800 + 900 + 800 + 1000 + 1000 + 1300}{6} = 967$$

เฉลี่ยรวม



5.1.2 วิธีค่าเฉลี่ยเคลื่อนที่ (Moving Average)



- ถ้าข้อมูลระยะยาวไม่คงที่ วิธีการหาค่าเฉลี่ยทั้งหมดอาจจะเอาผลที่ไกลเกินไปมารวม
- แต่ถ้าพบว่ามีความคงที่ในระยะสั้น ๆ เช่น ในช่วง 6 เดือนไม่ได้มีการเปลี่ยนแปลงมากนัก เราก็ควรนำแค่ 6 เดือนย้อนหลังมาคิด ซึ่งจะเรียกว่าการหาค่าเฉลี่ยเคลื่อนที่ย้อนหลัง 6 เดือน
- วิธีการคำนวณ:

$$\hat{y}_{t+1} = \frac{\sum_{i=t-n+1}^{t} y_i}{n} = \frac{y_{t-n+1} + y_{t-n+2} + \cdots + y_t}{n}$$

ตัวอย่าง 5.1.2: วิธีการค่าเฉลี่ยเคลื่อนที่

จากตารางเดิม จงหาค่าเฉลี่ยเคลื่อนที่ 3 เดือน และค่าเฉลี่นเคลื่อนที่ 4 เดือนของเดือนทั้งหมดที่เป็นไปได้จนถึงเดือนที่ 7

$\frac{\square + \square + \square}{3}$ $\frac{\square + \square + \square + \square}{4}$

เดือน	ยอดขาย
1	800
2	900
3	800
4	1000
5	1000
6	1300

เฉลี่ยเคลื่อนที่ 3 เดือน

$$\hat{y}_7 = \frac{y_4 + y_5 + y_6}{3} = \frac{1000 + 1000 + 1300}{3} = 1100$$

เดือน	ยอดขาย
1	800
2	900
3	800
4	1000
5	1000
6	1300

เฉลี่ยเคลื่อนที่ 4 เดือน

$$\hat{y}_7 = \frac{y_3 + y_4 + y_5 + y_6}{4} = \frac{800 + 1000 + 1000 + 1300}{4} = 1025$$

หน่วยกิต × เกรด①   เกรด②
1   A (4.00)   B (3.00)
3   B (3.00)   A (4.00)

5.1.3 วิธีค่าเฉลี่ยเคลื่อนที่ถ่วงน้ำหนัก (Weighted Moving Average)

$\frac{\text{ผลรวม(น้ำหนัก} \times \text{แต้ม)}}{\text{น้ำหนักรวม}}$

- เป็นวิธีการที่ต่อยอดมาจากการทำค่าเฉลี่ยนเคลื่อนที่ แต่มีแนวคิดว่าผลกระทบที่ยิ่งห่างออกไปยิ่งควรมีความสำคัญน้อยลง แต่ในขณะที่เหตุการณ์ล่าสุดควรมีผลกระทบมากที่สุด
- วิธีการถ่วงน้ำหนักที่ง่ายที่สุดคือไล่ 1, 2, 3, ... จากอดีตสุดมาปัจจุบันสุด
- เรียกอีกชื่อว่า **วิธีปรับเรียบแบบเชิงเส้น**
- วิธีการคำนวณ:

ค่อยลดน้ำหนักทีละ 1   น้ำหนัก

$$\hat{y}_{t+1} = \frac{\sum_{i=t-n+1}^{t} (i-t+n)\, y_i}{\sum_{i=1}^{n} i} = \frac{1 \cdot y_{t-n+1} + 2 \cdot y_{t-n+2} + \cdots + n \cdot y_t}{\underbrace{1+2+\cdots+n}_{\text{ผลรวมน้ำหนัก}}}$$

ค่าล่าสุดมีน้ำหนักสูงสุด

ตัวอย่าง 5.1.3: วิธีการค่าเฉลี่ยเคลื่อนที่ถ่วงน้ำหนัก

จากตารางเดิม จงหาค่าเฉลี่ยเคลื่อนที่ถ่วงน้ำหนัก 3 เดือน และค่าเฉลี่นเคลื่อนที่ถ่วงน้ำหนัก 4 เดือนของเดือนทั้งหมดที่เป็นไปได้จนถึงเดือนที่ 7

แรกสุด $\frac{1 \times \boxed{} + 2 \times \boxed{} + 3 \times \boxed{6}}{1+2+3}$ ล่าสุด

เดือน	ยอดขาย
1	800
2	900
3	800
4	1000
5	1000
6	1300

(3 เดือน)

$$\hat{y}_7 = \frac{1 \times y_4 + 2 \times y_5 + 3 \times y_6}{1+2+3} = \frac{1 \times 1000 + 2 \times 1000 + 3 \times 1300}{1+2+3} = 1150$$

เดือน	ยอดขาย
1	800
2	900
3	800
4	1000
5	1000
6	1300

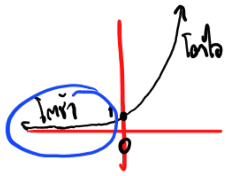
(4 เดือน)

$$\hat{y}_7 = \frac{1 \times y_3 + 2 \times y_4 + 3 \times y_5 + 4 \times y_6}{1+2+3+4} = \frac{1 \times 800 + 2 \times 1000 + 3 \times 1000 + 4 \times 1300}{1+2+3+4}$$

$$= 1100$$

$\frac{}{10}$

$2^0 = 1$



ตัวเลข^x

$y = 2^x$ $y = 10^x$

5.1.4 วิธีปรับเรียบแบบเอกซ์โพเนเชียล (exponential smoothing)

10:20

- เป็นอีกวิธีในการถ่วงน้ำหนัก โดยให้ความสำคัญของค่าล่าสุดเริ่มที่ 1 และลดค่าความสำคัญลงไปแบบ exponential โดยที่ยังมีการนำค่าตั้งแต่จุดเริ่มต้นมาพิจารณา
- แต่สูตรการคำนวณถูกจัดให้อยู่ในรูปที่คำนวณได้ง่าย (แค่ต้องคำนวณไต่ลำดับขึ้นมาเรื่อย ๆ) รูปแบบ exponential จึงไม่เห็นอยู่ในสูตร
- วิธีการคำนวณ:

ปัจจุบัน: เดือนที่ t

ค่าทำนายที่เดือนถัดไป

ค่าทำนายของเดือนปัจจุบัน

$$\hat{y}_{t+1} = \hat{y}_t + \boxed{\alpha}(y_t - \hat{y}_t)$$

ค่าจริงของเดือนปัจจุบัน

ถ้า $\alpha = 0 \rightarrow \hat{y}_{t+1} = \hat{y}_t$ (ค่าทำนายทุกเดือนเท่ากัน)

ถ้า $\alpha = 1 \rightarrow \hat{y}_{t+1} = \hat{y}_t + (y_t - \hat{y}_t) = \cancel{\hat{y}_t} + y_t - \cancel{\hat{y}_t}$

$\hat{y}_{t+1} = y_t$ (ใช้ค่าจริงเดือนนี้เป็นค่าทำนายของเดือน)

$0 < \alpha \leq 1$

น้ำหนักระหว่างค่าจริงและค่าทำนาย ยิ่ง $\alpha \to 1$ ยิ่งให้ความสำคัญกับค่าจริงมากขึ้น

หรือ

$$\hat{y}_{t+1} = \alpha y_t + (1 - \alpha)\,\hat{y}_t$$

โดยที่ $0 \leq \alpha \leq 1$

- ค่า α เป็นค่าที่ต้องกำหนดขึ้นมาตั้งแต่เริ่มตัดสินใจ (เหมือนกับที่เราต้องเลือกว่าจะ moving average หรือ weighted moving average ของกี่เดือน)

ตัวอย่าง 5.1.4: วิธีปรับเรียบแบบเอกซ์โพเนนเชียล

จากตารางเดิม จงหาค่าทำนายจากวิธีการปรับเรียบแบบเอกซ์โพเนนเชียล โดยใช้ $\alpha = 0.3$ และ $\alpha = 0.8$

$\alpha = 0.3$

เดือน	ยอดขาย
1	800
2	900
3	800
4	~~1000~~
5	~~1000~~
6	~~1300~~

$\hat{y}_{t+1} = \hat{y}_t + \alpha(y_t - \hat{y}_t)$ โดยที่ $y_0 = \hat{y}_0 = y_1 = 800$

$\hat{y}_1 = \hat{y}_0 + 0.3(y_0 - \hat{y}_0) = 800 + 0.3(800 - 800) = 800$ $(\hat{y}_1 = y_1$ เสมอ$)$

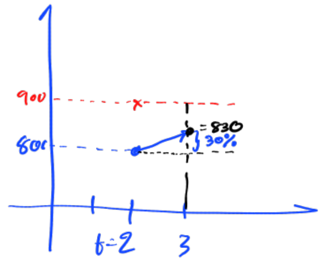
$\hat{y}_2 = \hat{y}_1 + 0.3(y_1 - \hat{y}_1) = 800 + 0.3(800 - 800) = 800$ $(\hat{y}_2 = y_1$ เสมอ$)$

$\hat{y}_3 = \hat{y}_2 + 0.3(y_2 - \hat{y}_2) = 800 + 0.3(900 - 800) = 830$

$\hat{y}_4 = \hat{y}_3 + 0.3(y_3 - \hat{y}_3) = 830 + 0.3(800 - 830) = 821$ (-30)

$\hat{y}_5 = \hat{y}_4 + 0.3(y_4 - \hat{y}_4) =$ ______

$\hat{y}_6 = \hat{y}_5 + 0.3(y_5 - \hat{y}_5) =$ ______



เดือน	ยอดขาย
1	800
2	900
3	800
4	1000
5	1000
6	1300

$800 + 0.3(900 - 800)$
$= 800 + 0.3 \times 100$
$= 800 + 30$
$= 830$

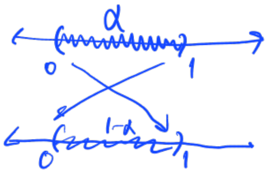
$830 + 0.3(800 - 830)$
$= 830 + 0.3 \times (-30)$
$= 830 - 9$
$= 821$

หมายเหตุ 5: สาเหตุที่เรียกว่าการปรับเรียบแบบเอกซ์โพเนนเชียล

รูปแบบสมการที่นิยามมาในด้านบนเรียกว่าการนิยามแบบเวียนเกิด คือการจะหาพจน์ที่ใด ๆ ได้ จะต้องคำนวณให้ทราบค่าพจน์ก่อนหน้าก่อน ดังนั้นจึงจำเป็นต้องคำนวณไล่จากขั้นที่ 1 ขึ้นมาจนถึงขั้นที่ต้องการ แต่ทั้งนี้ เรายังสามารถจัดรูปให้อยู่ในรูปที่ไม่ขึ้นกับพจน์ก่อนหน้าได้ดังนี้

$$
\begin{aligned}
\hat{y}_{t+1} &= \alpha y_t + (1-\alpha)\,\hat{y}_t \\
&= \alpha y_t + (1-\alpha)\left(\alpha y_{t-1} + (1-\alpha)\,\hat{y}_{t-1}\right) \\
&= \alpha y_t + \alpha(1-\alpha)\,y_{t-1} + (1-\alpha)^2\,\hat{y}_{t-1} \\
&= \alpha y_t + \alpha(1-\alpha)\,y_{t-1} + \alpha(1-\alpha)^2\,y_{t-2} + (1-\alpha)^3\,\hat{y}_{t-3} \\
&= \alpha y_t + \alpha(1-\alpha)\,y_{t-1} + \alpha(1-\alpha)^2\,y_{t-2} + \alpha(1-\alpha)^3\,y_{t-3} + (1-\alpha)^4\,\hat{y}_{t-4} \\
&\ \ \vdots \\
&= \alpha y_t + \alpha(1-\alpha)\,y_{t-1} + \alpha(1-\alpha)^2\,y_{t-2} + \cdots + \alpha(1-\alpha)^{t-2}\,y_2 + (1-\alpha)^{t-1}\,y_1
\end{aligned}
$$

หรือถ้าเขียนเป็นตัวอย่างแบบตัวเลขชัดเจน จะมีดังนี้



$$
\begin{aligned}
\hat{y}_1 &= y_1 \\
\hat{y}_2 &= (1-\alpha)^0\,y_1 = y_1 \\
\hat{y}_3 &= \alpha y_2 + (1-\alpha)\,y_1 \\
\hat{y}_4 &= \alpha y_3 + \alpha(1-\alpha)\,y_2 + (1-\alpha)^2\,y_1 \\
\hat{y}_5 &= \alpha \boxed{y_4} + \alpha(1-\alpha)\boxed{y_3} + \alpha(1-\alpha)^2\boxed{y_2} + (1-\alpha)^3\boxed{y_1}
\end{aligned}
$$

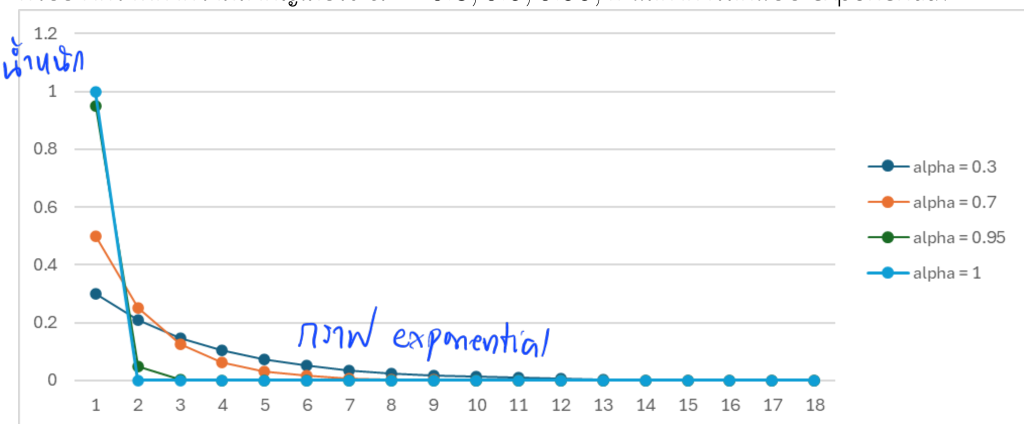
คูณด้วย 1-α เรื่อยๆ ลดแบบ exponential

สมมติ $\alpha = 0.3$ $\hat{y}_5 = 0.3y_4 + 0.3\times 0.7y_3 + 0.3\times 0.7^2 y_2 + 0.7^3 y_1$

ซึ่งจะเห็นว่า

1. จำนวนพจน์ของค่าจริงที่มาใช้คำนวณจะไม่ถูกกำหนดตายตัวเหมือนวิธีค่าเฉลี่ยเคลื่อนที่แบบอื่นๆ
2. แต่ความสำคัญก็จะถูกลดทอนลงเรื่อย ๆ จนเข้าใกล้ 0

ตัวอย่างกราฟค่าความสำคัญเมื่อใช้ $\alpha = 0.3, 0.5, 0.95, 1$ แสดงการลดแบบ exponential



5.2 ตัวแบบการถดถอยเชิงเส้น

ตัวแปร

Simple linear regression
เส้นตรง ถดถอย

- ตัวแบบการถดถอย (regression) เป็นการสร้างตัวแบบการทำนายโดยอยู่บนสมมติฐานว่าตัวแปรต้นตัวหนึ่ง (x) มีความสัมพันธ์เชิงฟังก์ชันกับค่าตัวแปรที่เราสนใจ (y)
- ในวิชานี้เราสนใจแค่การถดถอยเชิงเส้นตัวแปรเดียว กล่าวคือ มีชุดข้อมูล $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ ที่มีความสัมพันธ์

สิ่งที่อยากได้

$$\boxed{Y = a_0 + a_1 X} + \cancel{\epsilon}$$

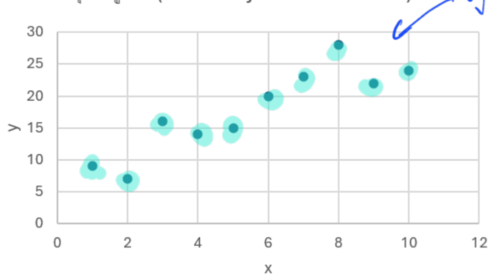
โดย a_0, a_1 เป็นค่าคงที่ และ ϵ คือพจน์ค่าคาดเคลื่อน

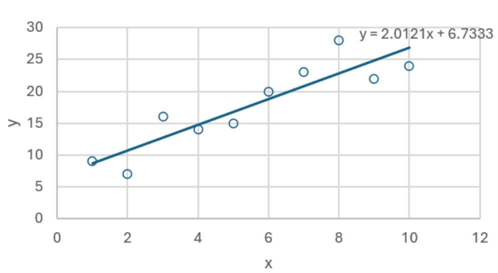
- เป้าหมายคือเราต้องการประมาณค่า $a_0 = \alpha_0, a_1 = \alpha_1$ ที่

(สิ่งที่จะคำนวณหา)

$$Y \approx \hat{Y} = \alpha_0 + \alpha_1 X$$

ดูคล้ายเส้นตรง





ขั้นตอน 5.1

ค่า a_0, a_1 ของ ตัว แบบ การ ถดถอย เชิง เส้น $Y = a_0 + a_1 X + \epsilon$ ของ ชุด ข้อมูล $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ สามารถ ประมาณ ค่า ได้ ด้วย α_0, α_1 (ด้วย วิธี การ ทาง คณิตศาสตร์ที่เรียกว่าวิธีกำลังสองต่ำสุด (Least Squared Error)) ตามสูตร

★★★

ผลรวม

$$\boxed{\alpha_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2},} \qquad \alpha_0 = \bar{y} - \alpha_1 \bar{x}$$

และจะได้ว่า $\hat{Y} = \alpha_0 + \alpha_1 X$ เป็นตัวแบบการถดถอยเชิงเส้นของ $Y = a_0 + a_1 X + \epsilon$
โดยขั้นตอนการคำนวณตามสูตรดังกล่าวคือ

1. คำนวณหาค่าเฉลี่ย $\boxed{\bar{x} = \sum_{i=1}^{n} x/n}$ และค่าเฉลี่ย $\boxed{\bar{y} = \sum_{i=1}^{n} y/n}$
2. $(x_i - \bar{x})$: คำนวณหาผลต่างระหว่างค่า x และค่าเฉลี่ย $\bar{x}$ ของทุกข้อมูล
3. $(y_i - \bar{y})$: คำนวณหาผลต่างระหว่างค่า y และค่าเฉลี่ย $\bar{y}$ ของทุกข้อมูล
4. $(x_i - \bar{x})(y_i - \bar{y})$: นำค่าผลต่างจาก 2 ข้อก่อนหน้ามาคูณกัน

5. $\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})$: นำค่าผลคูณของทุกข้อมูลจากขั้นที่ผ่านมามาบวกกัน
6. $(x_i-\bar{x})^2$: นำค่าผลต่างที่คำนวณไว้ในขั้นที่ 2 มายกกำลังสอง
7. $\sum_{i=1}^{n}(x_i-\bar{x})^2$: นำค่ากำลังสองของผลต่างในขั้นตอนที่ผ่านมามาบวกกัน
8. $\alpha_1 =$ ค่าผลบวกจากขั้นตอนที่ 5 หารด้วยค่าผลบวกจากขั้นตอนที่ 7
9. $\alpha_0 = \bar{y} - \alpha_1\bar{x}$

นักศึกษาสามารถสร้างตารางการคำนวณตามตัวอย่างด้านล่างเพื่อใช้ประกอบการคำนวณได้

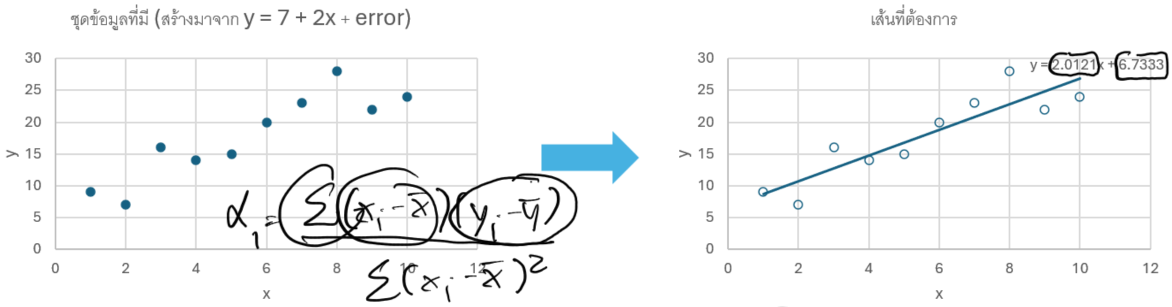
x_i	y_i	$x_i-\bar{x}$	$y_i-\bar{y}$	$(x_i-\bar{x})(y_i-\bar{y})$	$(x_i-\bar{x})^2$
x_1	y_1				
x_2	y_2				
x_n	y_n				
$\bar{x}=\frac{\sum_{i=1}^{n}x_i}{n}$	$\bar{y}=\frac{\sum_{i=1}^{n}y_i}{n}$			$\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})$	$\sum_{i=1}^{n}(x_i-\bar{x})^2$

$\alpha_1 = \frac{5}{6}$

$\alpha_0 = 2 - \alpha_1 \times 1$

ตัวอย่าง 5.2.1: การถดถอยเชิงเส้น 1 ตัวแปร

จงประมาณตัวแบบการถดถอยเชิงเส้นของชุดข้อมูลดังตารางด้านล่าง (ข้อมูลเดียวกับรูปตัวอย่าง)



$$\alpha_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

① x	② y	③ $x_i - \bar{x}$ $x_i - 5.5$	④ $y_i - \bar{y}$ $y_i - 17.8$	⑤ $(x_i-\bar{x})(y_i-\bar{y})$ ③ × ④	⑥ $(x_i - \bar{x})^2$ ③^2
1	9	(1−5.5) −4.5	−8.8	39.6	20.25
2	7	−3.5	−10.8	37.8	12.25
3	16	−2.5	−1.8	4.5	6.25
4	14	−1.5	−3.8	5.7	2.25
5	15	−0.5			
6	20	0.5			
7	23	1.5			
8	28	2.5			
9	22	3.5			
10	24	4.5			
$\bar{x} = 5.5$	$\bar{y} = 17.8$			$\Sigma = 166$	$\Sigma = 82.5$

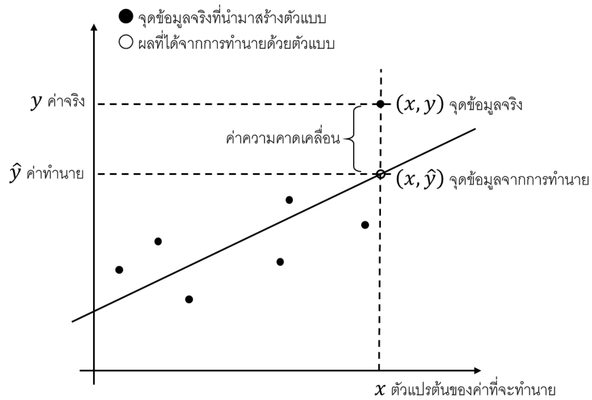
$$\therefore \alpha_1 = \frac{166}{82.5} = 2.012, \quad \alpha_0 = \bar{y} - \alpha_1 \bar{x} = 17.8 - \frac{166}{82.5} \times 5.5 = 6.733$$

5.3 การประเมิณผลความแม่นยำในการทำนาย

อย่างที่ได้กล่าวไปก่อนหน้านี้ว่าเราไม่สามารถระบุได้ว่าตัวแบบใดเป็นตัวแบบที่ดีที่สุด เพราะตัวแบบในการทำนายที่ดีขึ้นอยู่กับข้อมูลที่มีว่ามีลักษณะข้อมูลเป็นอย่างไร ตัวแบบเดียวกันอาจจะทำงานได้ดีในชุดข้อมูลหนึ่ง แต่อาจจะทำได้ไม่ดีในอีกชุดข้อมูลหนึ่ง เพราะฉะนั้น ในกระบวนการทำงานจริง จึงต้องมีการวัดผลเพื่อประเมิณความแม่นยำของตัวแบบเพื่อที่จะเปรียบเทียบความสามารถในการทำนายของแต่ละตัวแบบได้ ซึ่งแนวคิดหลักของการวัดผลคือการใช้ค่า**ความคลาดเคลื่อน** (error) เพื่อเป็นตัวบอกว่าสิ่งที่ตัวแบบทำนายออกมาได้คลาดเคลื่อนออกไปจากค่าจริงเท่าใด

(error)

$$\text{ค่าความคลาดเคลื่อนดิบ} = \left|\text{ค่าที่ตัวแบบทำนายได้} - \text{ค่าจริงจากชุดข้อมูล}\right|$$



ในหนังสือเล่มนี้ จะแบ่งการวัดผลออกเป็น 2 รูปแบบหลักได้แก่

1. **การวัดผลด้วยมาตรวัดของข้อมูล**: เป็นการวัดผลที่มีหน่วยออกมาเป็นหน่วยเดียวกันกับข้อมูลที่เราต้องการจะทำนาย โดยต้องการวัดระยะห่าง มีข้อดีในแง่การแสกงค่าคาดเคลื่อนจริง ๆ เช่นทำนายคลาดเคลื่อนไปกี่บาท มักถูกใช้ในการเปรียบเทียบระหว่างตัวแบบต่าง ๆ บนข้อมูลชุดเดียวกัน โดยจะกล่าวถึง

   ◇ ค่าเฉลี่ยของความผิดพลาดสัมบูรณ์ (mean absolute error: MAE)
   ◇ ค่ารากที่สองของค่าเฉลี่ยของความผิดพลาดกำลังสอง (root mean squared error: RMSE)

2. **การวัดผลเชิงสัมพัทธ์**: เป็นการวัดผลในเชิงการหาร้อยละเทียบเคียงกับค่าจริงว่าคลาดเคลื่อนไปกี่เปอร์เซนต์ ซึ่งวิธีการนี้มักใช้กับชุดข้อมูลที่ความรุนแรงของการคลาดเคลื่อนขึ้นอยู่กับขนาดของค่าจริง กล่าวคือการคาดเคลื่อนด้วยปริมาณหนึ่งตอนที่ค่าจริงมีค่าน้อย ๆ จะรุนแรงกว่าการคาดเคลื่อนขนาดเดียวกันเมื่อค่าจริงมีค่ามาก ๆ (ตัวอย่างเช่น เงินหาย 9 บาทจาก 10 บาท กับเงินหายไป 9 บาทจาก 1 ล้านบาท) อีกทั้งยังเป็นค่าความคลาดเคลื่อนที่ใช้เพื่อการเปรียบเที่ยบการทำงานของตัวแบบในต่างชุดข้อมูลที่อาจจะมีค่าที่ต้องการทำนายอยู่ในคนละมาตรวัดกัน โดยจะกล่าวถึง

   ◇ ค่าเฉลี่ยของความผิดพลาดสัมพัทธ์ (mean absolute percentage error: MAPE)
   ◇ ค่าเฉลี่ยของความผิดพลาดสัมบูรณ์ที่ปรับมาตราส่วน (mean absolute scaled error: MASE)

เวลาวัดผล → วัดแบบเฉลี่ยต่อตัว

error = $|\hat{y}_i - y_i|$

นิยาม 5.3.1: mean absolute error เฉลี่ยผิดพลาดสัมบูรณ์

$$\boxed{MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i|}$$

เฉลี่ยค่าคลาดเคลื่อน

นิยาม 5.3.2: root mean squared error

$$\boxed{RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}}$$

√ เฉลี่ย (กำลังสองของ error)

และในบางครั้ง อาจมีการใช้ MSE ซึ่งคือ

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2$$

แต่สิ่งที่ต้องระวังตอนอ่านค่าคือค่าที่ได้จะอยู่ในหน่วยกำลังสองของหน่วยเดิม (เช่นหน่วย บาท^2) ซึ่งไม่ได้มีความหมายในโลกจริง

นิยาม 5.3.3: mean absolute percentage error

error

ค่าจริง

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \times 100\%$$

นิยาม 5.3.4: mean absolute scaled error

สำหรับการทำการถดถอยเชิงเส้นเมื่อเทียบกับการทำนายด้วยการหยิบแต่ค่าเฉลี่ยมาเป็นค่าทำนาย

$$MASE = \frac{\frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i|}{\frac{1}{n}\sum_{i=1}^{n}|\bar{y} - y_i|}$$

สำหรับอนุกรมเวลาเมื่อเทียบกับการทำนายด้วยการหยิบค่าของครั้งก่อนหน้ามาเป็นค่าทำนายของครั้งปัจจุบัน

$$MASE = \frac{\frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i|}{\frac{1}{T-1}\sum_{t=2}^{T}|y_t - y_{t-1}|}$$

สำหรับการวัดผลด้วย MASE นั้น เราสามารถปรับเปลี่ยนวิธีการประมาณค่าตัวเทียบ (ตัวส่วน) เป็นตัวแบบแบบอื่นได้เช่นกัน เพียงแต่ 2 สูตรด้านบนเป็นการเทียบจากตัวแบบที่ง่ายที่สุดที่มักจะนึกถึงกันเป็นอันดับแรกตอนทำนาย

ตัวอย่าง 5.3.1: การวัดผลอนุกรมเวลา

จากตารางการทำตัวแบบอนุกรมเวลาแบบต่าง ๆ ที่ทำมาในตัวอย่างที่ผ่าน ๆ มา จงวัดผลค่าความคลาดเคลื่อน MAE, RMSE, MAPE, MASE ของแต่ละตัวแบบ โดยสมมติเพิ่มว่าค่าจริงของเดือนที่ 7 มีค่าเท่ากับ 1200 (และเพื่อความสะดวกในการคำนวณ จึงขอปัดค่าทำนายให้เป็นจำนวนเต็ม)

วิธีค่าเฉลี่ย (รวม)

เดือน	ยอดขายจริง	ค่าทำนาย	MAE $\lvert y_i-\hat{y}_i\rvert$	RMSE $(y_i-\hat{y}_i)^2$	MAPE $\lvert(y_i-\hat{y}_i)/y_i\rvert$
1	800				
2	900	800	100	10000	0.1111
3	800	850	50	2500	0.0625
4	1000	833	167	27889	0.1670
5	1000	875	125	15625	0.1250
6	1300	900	400	160000	0.3077
7	1200	967	233	54289	0.1942

$$MAE = \frac{100+50+167+125+400+233}{6} = 179.17$$

$$RMSE = \sqrt{\frac{10000+2500+27889+15625+160000+54289}{6}} = 212.25$$

$$MAPE = \frac{0.1111+0.0625+\cdots+0.1942}{6} = 0.1612 = 16.12\%$$

วิธีเฉลี่ยเคลื่อนที่ 3 เดือน

เดือน	ยอดขายจริง	ค่าทำนาย
1	800	
2	900	
3	800	
4	1000	833
5	1000	900
6	1300	933
7	1200	1100

วิธีเฉลี่ยเคลื่อนที่ 4 เดือน

เดือน	ยอดขายจริง	ค่าทำนาย
1	800	
2	900	
3	800	
4	1000	
5	1000	875
6	1300	925
7	1200	1025

วิธีค่าเฉลี่ยถ่วงน้ำหนัก 3 เดือน

เดือน	ยอดขายจริง	ค่าทำนาย
1	800	
2	900	
3	800	
4	1000	833
5	1000	917
6	1300	967
7	1200	1150

วิธีค่าเฉลี่ยถ่วงน้ำหนัก 4 เดือน

เดือน	ยอดขายจริง	ค่าทำนาย
1	800	
2	900	
3	800	
4	1000	
5	1000	900
6	1300	950
7	1200	1100

วิธีปรับเรียบแบบเอ็กโพเนนเชียล โดย $\alpha = 0.3$

เดือน	ยอดขายจริง	ค่าทำนาย
1	800	800
2	900	800
3	800	830
4	1000	821
5	1000	875
6	1300	912
7	1200	1029

วิธีปรับเรียบแบบเอ็กโพเนนเชียล โดย $\alpha = 0.8$

เดือน	ยอดขายจริง	ค่าทำนาย
1	800	800
2	900	800
3	800	880
4	1000	806
5	1000	964
6	1300	975
7	1200	1222

ตัวอย่าง 5.3.2: การวัดผลการถดถอยเชิงเส้น

จากตัวอย่างการหาตัวแบบการถดถอยเชิงเส้นในตัวอย่าง 5.2.1 จงวัดผลความคลาดเคลื่อน MAE, RMSE, MAPE, MASE (ในทำนองเดียวกัน ถ้าเป็นการคำนวณมือให้ปัดเป็นจำนวนเต็มเพื่อคิดเลขได้ จะได้คำนวณได้สะดวก)

x	y	$\hat{y}$ ......
1	9	
2	7	
3	16	
4	14	
5	15	
6	20	
7	23	
8	28	
9	22	
10	24	

ใช้ Excel

5.4 การใช้ Excel เพื่อช่วยคำนวณหาตัวแบบต่าง ๆ