

1 ข้อมูล ตารางความถี่ และการนำเสนอ

1.1 ขอบเขตและชนิดของข้อมูล

ขอบเขตศึกษา	ค่าสรุปที่ใช้
ประชากร: หน่วยทั้งหมด	พารามิเตอร์ เช่น μ, σ
ตัวอย่าง: ส่วนที่เลือกศึกษา	ค่าสถิติ เช่น $\bar{x}, s$

- **เชิงคุณภาพ** บอกประเภท เช่น วิธีเดินทาง
 - **เชิงปริมาณ** แบบไม่ต่อเนื่องเกิดจากการนับ ส่วนแบบต่อเนื่องเกิดจากการวัด
 - **ปฐมภูมิ** เก็บโดยตรงตามวัตถุประสงค์ **ทุติยภูมิ** นำข้อมูลที่มีอยู่แล้วมาใช้
 - **ภาคตัดขวาง** หลายหน่วยในช่วงเดียว **อนุกรมเวลา** ติดตามค่าตามลำดับเวลา
- ตัวอย่างขนาดใหญ่ไม่รับประกันความเป็นตัวแทน ต้องพิจารณาวิธีเลือกและขอบเขตการอ้างอิง

1.2 ความถี่ สหสัมพันธ์ และสัมพัทธ์

ให้ f_j เป็นความถี่ของค่าหรือชั้นที่ j และ N เป็นจำนวนข้อมูลทั้งหมด

$$N = \sum_{j=1}^k f_j, \quad F_j = \sum_{i=1}^j f_i, \quad f_j = F_j - F_{j-1}$$

F_j คือความถี่สะสมถึงชั้นที่ j โดยกำหนด $F_0 = 0$

$$\text{สัมพัทธ์} = \frac{f_j}{N}, \quad \text{สะสมสัมพัทธ์} = \frac{F_j}{N}$$

ร้อยละเท่ากับความถี่สัมพัทธ์คูณ 100

ตัวอย่างย่อ ข้อมูล 1, 1, 2, 2, 2, 3, 3, 4 มี $N = 8$

ค่า	f_j	F_j	$\frac{f_j}{N}$	$\frac{F_j}{N}$
1	2	2	0.250	0.250
2	3	5	0.375	0.625
3	2	7	0.250	0.875
4	1	8	0.125	1.000

1.3 อันตรภาคชั้นและขอบชั้น

ถ้า L_j, U_j เป็นขอบล่างและขอบบนของชั้นที่ j

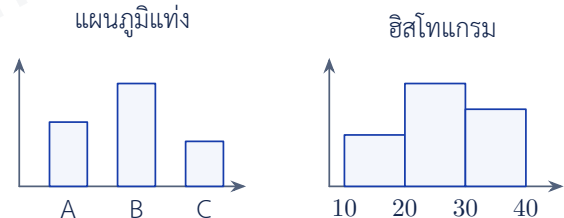
$$I_j = U_j - L_j, \quad m_j = \frac{L_j + U_j}{2}$$

I_j คือความกว้างชั้น และ m_j คือจุดกึ่งกลางชั้น ข้อมูลบันทึกละเอียด d หน่วย ใช้ขอบชั้นห่างชิดจำกัดชั้นออกไปด้านละ $\frac{d}{2}$

ตัวอย่างย่อ ชั้น 10–19 เมื่อบันทึกเป็นจำนวนเต็ม มีขอบ 9.5, 19.5 จึงได้ $I = 10, m = 14.5$
ชั้นต้องครอบคลุมข้อมูล ไม่ทับซ้อนกัน และกำหนดการรวมค่าที่ขอบให้ชัดเจน

1.4 แผนภูมิแท่งและฮิสโทแกรม

แผนภูมิแท่ง เปรียบเทียบหมวดหมู่หรือค่าที่แยกจากกัน แท่งมีช่องว่าง ส่วน**ฮิสโทแกรม** แสดงข้อมูลเชิงปริมาณที่จัดเป็นช่วง แท่งของชั้นติดกัน



ถ้าชั้นกว้างต่างกัน ต้องใช้ความหนาแน่นความถี่

$$\text{ความสูงแท่ง} = \frac{f_j}{I_j}, \quad \text{พื้นที่แท่ง} = f_j$$

ถ้าชั้นกว้างเท่ากัน ใช้ความถี่เป็นความสูงได้ เพราะพื้นที่แท่งเป็นสัดส่วนกับความถี่

1.5 แผนภาพจุดและต้นและใบ



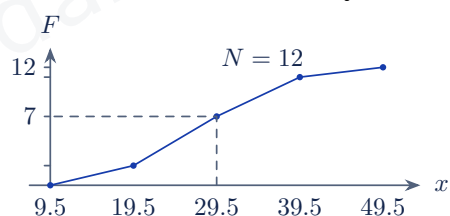
หนึ่งจุดแทนหนึ่งข้อมูล จุด 3 จุดเหนือเลข 2 หมายถึงมีข้อมูล 2, 2, 2 ไม่ใช่ค่าข้อมูลเท่ากับ 3

ต้น	ใบ
1	2 5 5
2	0 3

ต้นแทนหลักสิบ ใบแทนหลักหน่วย จึงอ่านเป็น 12, 15, 15, 20, 23 ต้องระบุหลักหรือหน่วยเสมอ

1.6 เส้นความถี่สะสมและจำนวนในช่วง

เส้นความถี่สะสมของข้อมูลจัดขึ้นผ่านขอบชั้นบนกับความถี่สะสม เริ่มที่ขอบล่างแรกด้วยสะสมศูนย์



ถ้า $G(x)$ คือจำนวนข้อมูลที่ไม่เกิน x จะได้

$$\text{จำนวนข้อมูลใน } (a, b] = G(b) - G(a)$$

การอ่านภายในชั้นด้วยเส้นตรงเป็นค่าประมาณ หากเกณฑ์อยู่กลางชั้น จำนวนที่ผ่านจริงยังไม่แน่นอน ต้องแยกขอบเขตที่เป็นไปได้ออกจากค่าประมาณ

2 ค่ากลางและตำแหน่งของข้อมูล

2.1 ค่าเฉลี่ย มัธยฐาน และฐานนิยม

ให้มีข้อมูล N ค่า และ $x_{(j)}$ เป็นค่าลำดับที่ j หลังเรียงจากน้อยไปมาก

$$\bar{x} = \frac{\sum_{i=1}^N x_i}{N} = \frac{\sum f_j x_j}{N}$$

$$Me = \begin{cases} x_{(\frac{N+1}{2})}, & N \text{ คี่} \\ \frac{x_{(\frac{N}{2})} + x_{(\frac{N}{2}+1)}}{2}, & N \text{ คู่} \end{cases}$$

ฐานนิยม Mo คือค่าหรือประเภทที่มีความถี่สูงสุด อาจมีหลายค่า หรือไม่มีค่าที่เด่นกว่า

ค่ากลาง ลักษณะสำคัญ

ค่าเฉลี่ย ใช้ทุกค่า ไวต่อค่าสุดปลาย

มัธยฐาน ใช้ลำดับ ทนต่อค่าสุดปลายกว่า

ฐานนิยม ใช้ได้กับข้อมูลเชิงคุณภาพ

2.2 ถ่วงน้ำหนัก รวมกลุ่ม และค่ากลางอื่น

ถ้า $w_i \geq 0$ และ $\sum w_i > 0$ ค่าเฉลี่ยถ่วงน้ำหนักคือ

$$\bar{x}_w = \frac{\sum w_i x_i}{\sum w_i}$$

กลุ่มที่ j มีจำนวน n_j และค่าเฉลี่ย $\bar{x}_j$

$$\bar{x}_{รวม} = \frac{\sum n_j \bar{x}_j}{\sum n_j}$$

ห้ามเฉลี่ยค่าเฉลี่ยกลุ่มโดยตรง เว้นแต่แต่ละกลุ่มมีจำนวนเท่ากัน สำหรับข้อมูลบวก N ค่า

$$GM = \sqrt[N]{\prod_{i=1}^N x_i}, \quad HM = \frac{N}{\sum_{i=1}^N \frac{1}{x_i}}$$

$$HM \leq GM \leq AM \quad (AM = \bar{x})$$

เท่ากันเมื่อทุกค่าเท่ากัน ใช้ GM กับตัวคูณการเติบโต และ HM กับอัตราเร็วเมื่อระยะทางแต่ละช่วงเท่ากัน

2.3 ค่ากลางของข้อมูลจัดชั้น

ใช้จุดกึ่งกลาง m_j แทนค่าทั้งชั้น จึงเป็นค่าประมาณ

ชั้น	m_j	f_j	F_j	$f_j m_j$
10-19	14.5	2	2	29
20-29	24.5	5	7	122.5
30-39	34.5	4	11	138
40-49	44.5	1	12	44.5

$$\bar{x} \approx \frac{\sum f_j m_j}{N} = \frac{334}{12} \approx 27.83$$

2.4 มัธยฐานและฐานนิยมของข้อมูลจัดชั้น

$$Me \approx L + \frac{\frac{N}{2} - F}{f} I$$

เลือกชั้นที่มีตำแหน่ง $\frac{N}{2}$ ให้ L, F, f, I คือขอบล่าง สะสมก่อนชั้น ความถี่ และความกว้างของชั้นนั้น

$$Mo \approx L_m + \frac{d_1}{d_1 + d_2} I$$

ใช้เมื่อชั้นกว้างเท่ากันและมีชั้นฐานนิยมเด่น L_m เป็นขอบล่างชั้นนั้น d_1, d_2 คือความถี่ชั้นฐานนิยมลบความถี่ชั้นก่อนและชั้นถัดไปตามลำดับ

2.5 ควอร์ไทล์ เดซิส์ และเปอร์เซ็นต์ไทล์

เรียงข้อมูลรายตัว N ค่า แล้วใช้ข้อตกลงตำแหน่งดังนี้

$$\text{pos}(Q_r) = \frac{r(N+1)}{4} \quad r = 1, 2, 3$$

$$\text{pos}(D_r) = \frac{r(N+1)}{10} \quad r = 1, \dots, 9$$

$$\text{pos}(P_r) = \frac{r(N+1)}{100} \quad r = 1, \dots, 99$$

$$Q_1 = P_{25}, \quad Q_2 = P_{50} = Me, \quad Q_3 = P_{75}$$

ตำแหน่ง $j + \alpha$ เมื่อ $0 \leq \alpha < 1$ ใช้การแทรกค่า

$$x_{(j)} + \alpha(x_{(j+1)} - x_{(j)})$$

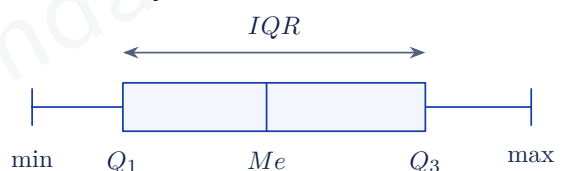
ใช้เมื่ออันดับที่ต้องการมีอยู่ และระวางการนับข้อมูลที่มีค่าซ้ำ **ข้อมูลจัดชั้น** ใช้ตำแหน่ง N แทน $N+1$

$$P_r \approx L + \frac{\frac{rN}{100} - F}{f} I$$

L, F, f, I คือขอบล่าง สะสมก่อนชั้น ความถี่ และความกว้างชั้นเป้าหมาย สูตร Q_r, D_r เปลี่ยนตำแหน่งเป็น $\frac{rN}{4}, \frac{rN}{10}$ ตามลำดับ

2.6 แผนภาพกล่องและค่าพิตปกติ

ควอร์ไทล์แบ่งตำแหน่งเป็นสี่ส่วน ส่วนละประมาณ 25% ค่าสรุปหาค่าแสดงในรูป



$$IQR = Q_3 - Q_1$$

IQR คือพิสัยระหว่างควอร์ไทล์ กำหนดไว้

$$L_f = Q_1 - 1.5IQR, \quad U_f = Q_3 + 1.5IQR$$

ค่าพิตปกติคือ $x < L_f$ หรือ $x > U_f$ ค่าเท่าไรก็ไม่พิตปกติ
หมวดแบบปรับปรุงจบที่ค่าจริงสุดปลายที่ไม่พิตปกติ ไม่ใช่รั้วเสมอไป

3 การกระจาย การแปลงข้อมูล และการรวมกลุ่ม

3.1 พิสัยและส่วนเบี่ยงเบนเฉลี่ย

สำหรับข้อมูล $x_1, \dots, x_N$ ที่มีค่าเฉลี่ย $\bar{x}$

$$R = x_{\max} - x_{\min}, \quad MD = \frac{\sum |x_i - \bar{x}|}{N}$$

R คือพิสัย และ MD คือส่วนเบี่ยงเบนเฉลี่ยรอบค่าเฉลี่ย ทั้งสองมีหน่วยเดียวกับข้อมูล กรณีที่มีความถี่แทนด้วย $\sum f_j |x_j - \bar{x}|$

3.2 ความแปรปรวนและส่วนเบี่ยงเบนมาตรฐาน

ประชากร ขนาด N ค่าเฉลี่ย μ

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N} = \frac{\sum x_i^2}{N} - \mu^2$$

$$\sigma = \sqrt{\sigma^2}$$

ตัวอย่าง ขนาด $n \geq 2$ ค่าเฉลี่ย $\bar{x}$

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{\sum x_i^2 - n\bar{x}^2}{n-1}$$

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

ความแปรปรวนมีหน่วยเป็นกำลังสอง ส่วนเบี่ยงเบนมาตรฐานมีหน่วยเดียวกับข้อมูล ไม่สลับตัวหาร N กับ $n-1$

ตัวอย่างย่อ ประชากร 2, 4, 6 มี $\mu = 4$, $\sigma^2 = \frac{8}{3}$ หากเป็นตัวอย่างชุดเดียวกัน จะมี $s^2 = 4$

3.3 ข้อมูลที่มีความถี่และข้อมูลจัดชั้น

ข้อมูลรายค่ามีความถี่ f_j และ $N = \sum f_j$

$$\mu = \frac{\sum f_j x_j}{N}, \quad \sigma^2 = \frac{\sum f_j x_j^2}{N} - \mu^2$$

ข้อมูลจัดชั้นแทน x_j ด้วยจุดกึ่งกลาง m_j และใช้เครื่องหมายประมาณ สำหรับตัวอย่างให้ใช้

$$s^2 = \frac{\sum f_j m_j^2 - n\bar{x}^2}{n-1}, \quad n = \sum f_j$$

ความถี่ f_j เป็นจำนวนครั้งที่ข้อมูลปรากฏ ไม่ใช่จำนวนทุกหน่วย นำสูตร $n-1$ ไปใช้ได้เสมอ

3.4 การกระจายสัมพัทธ์

$$CV = \frac{\sigma}{\mu}, \quad CV\% = \frac{\sigma}{\mu} \times 100\%$$

CV คือสัมประสิทธิ์การแปรผัน สำหรับตัวอย่างใช้ $\frac{s}{\bar{x}}$ ใช้กับข้อมูลอัตราส่วนที่มีศูนย์แท้และค่าเฉลี่ยบวก ค่า CV น้อย แสดงการกระจายเทียบค่าเฉลี่ยน้อย $QD = \frac{Q_3 - Q_1}{2}$ คือส่วนเบี่ยงเบนควอร์ไทล์

$$C_R = \frac{x_{\max} - x_{\min}}{x_{\max} + x_{\min}}, \quad C_Q = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

C_R, C_Q คือสัมประสิทธิ์พิสัยและส่วนเบี่ยงเบนควอร์ไทล์ ใช้กับข้อมูลอัตราส่วนและตัวหารบวก

3.5 ผลของการแปลงเชิงเส้น

ให้ทุกค่าเปลี่ยนตาม $Y = aX + b$ โดย a, b คงตัว

ค่าสรุปเดิม	ค่าสรุปใหม่
μ_X	$a\mu_X + b$
σ_X^2	$a^2\sigma_X^2$
σ_X	$ a \sigma_X$
R_X, MD_X, IQR_X	คูณด้วย $ a $

ใช้หลักเดียวกันกับค่าเฉลี่ยและความแปรปรวนตัวอย่าง การเลื่อนด้วย b ไม่เปลี่ยนการกระจายแบบสัมบูรณ์ แต่เปลี่ยน CV ได้

$$P_r(Y) = \begin{cases} aP_r(X) + b, & a > 0 \\ aP_{100-r}(X) + b, & a < 0 \end{cases}$$

เมื่อ $a < 0$ ลำดับข้อมูลกลับด้าน ถ้า $a = 0$ ทุกค่าเท่ากับ b และการกระจายเป็นศูนย์

3.6 เพิ่ม ลบ และแก้ข้อมูลผิด

สำหรับประชากรขนาด N ค่าเฉลี่ย μ และความแปรปรวน σ^2 ให้เก็บผลรวมสองชนิด

$$S_1 = N\mu, \quad S_2 = N(\sigma^2 + \mu^2)$$

แก้ u เป็น v ใช้ $S'_1 = S_1 - u + v$ และ $S'_2 = S_2 - u^2 + v^2$ หากเพิ่มหรือลบข้อมูลต้องปรับจำนวนเป็น N' ด้วย

$$\mu' = \frac{S'_1}{N'}, \quad \sigma'^2 = \frac{S'_2}{N'} - (\mu')^2$$

สำหรับข้อมูลตัวอย่างเก็บ $S_1 = n\bar{x}$ และ

$S_2 = (n-1)s^2 + n\bar{x}^2$ แล้วใช้สูตรตัวอย่างกับจำนวนใหม่

3.7 การรวมกลุ่มและการกระจายระหว่างกลุ่ม

กลุ่มที่ j มีขนาด n_j ค่าเฉลี่ย μ_j และความแปรปรวนประชากร σ_j^2 ให้ $N = \sum n_j$

$$\mu = \frac{\sum n_j \mu_j}{N}$$

$$\sigma^2 = \frac{\sum n_j [\sigma_j^2 + (\mu_j - \mu)^2]}{N}$$

รวมทั้งความแปรปรวนภายในกลุ่มและความต่างของค่าเฉลี่ยกลุ่ม ไม่เฉลี่ยความแปรปรวนอย่างเดียวก่อน กรณีตัวอย่างแต่ละกลุ่มมี $n_j \geq 2$ และไม่ซ้อนกัน ให้ $\bar{x}$ เป็นค่าเฉลี่ยรวม

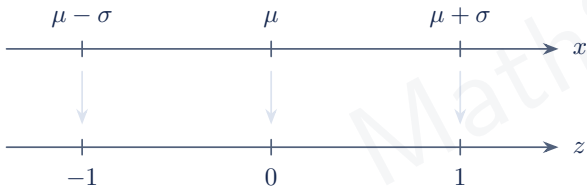
$$s_{\text{รวม}}^2 = \frac{\sum (n_j - 1)s_j^2 + \sum n_j (\bar{x}_j - \bar{x})^2}{N - 1}$$

4 ค่ามาตรฐานและการแจกแจงปกติ

4.1 ค่ามาตรฐานและการเปรียบเทียบ

ให้ประชากรมีค่าเฉลี่ย μ และส่วนเบี่ยงเบนมาตรฐาน $\sigma > 0$
ค่ามาตรฐานของข้อมูล x คือ

$$z = \frac{x - \mu}{\sigma}, \quad x = \mu + z\sigma$$



$z > 0$ สูงกว่าค่าเฉลี่ย $z < 0$ ต่ำกว่าค่าเฉลี่ย โดย $|z|$ บอกระยะห่างเป็นจำนวนเท่าของส่วนเบี่ยงเบนมาตรฐาน

ตัวอย่างย่อ $\mu = 60, \sigma = 8, x = 72$ ให้ $z = 1.5$
การแปลงเป็น z ทำให้ข้อมูลไม่แจกแจงปกติ แต่ไม่ทำให้ข้อมูลกลายเป็นปกติโดยอัตโนมัติ

4.2 สมบัติค่ามาตรฐาน

สำหรับประชากรทั้ง N ค่า ซึ่งใช้ μ, σ ของชุดนั้น

$$\sum z_i = 0, \quad \sum z_i^2 = N$$

กรณีตัวอย่างขนาด n ใช้ $s > 0$ จะได้ผลรวมค่ามาตรฐานเป็น 0 และผลรวมกำลังสองเป็น $n - 1$ เมื่อ $Y = aX + b$ และ $a \neq 0$ จะมี

$$z_Y = \frac{a}{|a|} z_X$$

ค่าเดิมที่เปรียบเทียบกับสองกลุ่มต้องให้ค่าจริงตรงกัน

$$x = \mu_1 + z_1\sigma_1 = \mu_2 + z_2\sigma_2$$

4.3 การแจกแจงปกติและตารางพื้นที่ซ้าย

$$X \sim N(\mu, \sigma^2) \Rightarrow Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

โค้งปกติสมมาตรรอบ μ มีค่าเฉลี่ย มัธยฐาน และฐานนิยมเท่ากัน พื้นที่รวมเท่ากับ 1

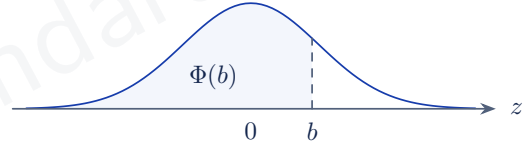
$$\Phi(z) = P(Z \leq z), \quad \Phi(-z) = 1 - \Phi(z)$$

z	$\Phi(z)$	z	$\Phi(z)$
0	0.5000	1.645	0.9500
0.5	0.6915	1.960	0.9750
1.0	0.8413	2.000	0.9772
1.282	0.9001	2.326	0.9900

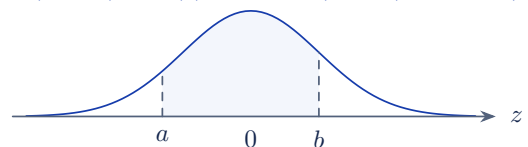
ค่าตารางเป็นค่าประมาณ ตรวจสอบว่าตารางที่ใช้ให้พื้นที่ด้านซ้ายหรือพื้นที่จากศูนย์กลางก่อนแทนค่า

4.4 พื้นที่ใต้โค้งและความน่าจะเป็น

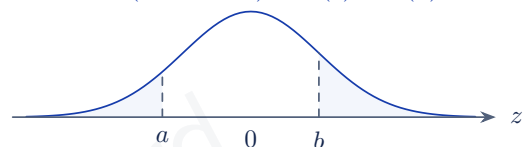
สูตรต่อไปนี้จะใช้กับ $Z \sim N(0, 1)$ และ $a < b$



$$P(Z \leq b) = \Phi(b), \quad P(Z > b) = 1 - \Phi(b)$$



$$P(a < Z < b) = \Phi(b) - \Phi(a)$$



$$P(Z < a \text{ หรือ } Z > b) = \Phi(a) + 1 - \Phi(b)$$

ถ้าขอบสมมาตรและ $c > 0$

$$P(-c < Z < c) = 2\Phi(c) - 1$$

จุดเดียวมีความน่าจะเป็นศูนย์ เครื่องหมาย $<$ กับ $\leq$ จึงให้พื้นที่เท่ากัน ต้องมีสมมติฐานปกติก่อนใช้ตารางนี้

4.5 เปอร์เซ็นไทล์ เกณฑ์ และจำนวนคาดหวัง

ถ้า $X \sim N(\mu, \sigma^2)$ เปอร์เซ็นไทล์ที่ r หาได้จาก

$$\Phi(z_r) = \frac{r}{100}, \quad P_r = \mu + z_r\sigma$$

คัดกลุ่มบนสัดส่วน p ใช้พื้นที่ด้านซ้าย $1 - p$ หากประชากรมี N หน่วย และผ่านเกณฑ์ด้วยโอกาส p

$$\text{ร้อยละ} = 100p, \quad \text{จำนวนคาดหวัง} = Np$$

จำนวนคาดหวังไม่รับประกันจำนวนที่เกิดขึ้นจริง

4.6 ย้อนหาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน

ค่าจริง x_1, x_2 ตรงกับค่ามาตรฐาน z_1, z_2 บนโค้งเดียวกัน โดย $z_1 \neq z_2$

$$\sigma = \frac{x_2 - x_1}{z_2 - z_1}, \quad \mu = x_1 - z_1\sigma$$

ต้องตรวจ $\sigma > 0$ และแทนกลับทั้งสองเงื่อนไข

ตัวอย่างย่อ $x_1 = 50, z_1 = -1$ และ $x_2 = 80, z_2 = 2$ จะได้ $\sigma = 10, \mu = 60$

5 ความสัมพันธ์ การถดถอย และแบบจำลอง: เนื้อหาเสริม

5.1 ข้อมูลคู่และสหสัมพันธ์เพียร์สัน

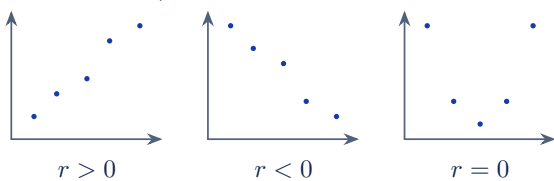
มีข้อมูลคู่ (x_i, y_i) จำนวน n คู่ ให้ $\bar{x}, \bar{y}$ เป็นค่าเฉลี่ย และกำหนดผลรวมปรับ

$$S_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$$

$$S_{yy} = \sum y_i^2 - \frac{(\sum y_i)^2}{n}$$

$$S_{xy} = \sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}$$

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} \quad (S_{xx}, S_{yy} > 0)$$



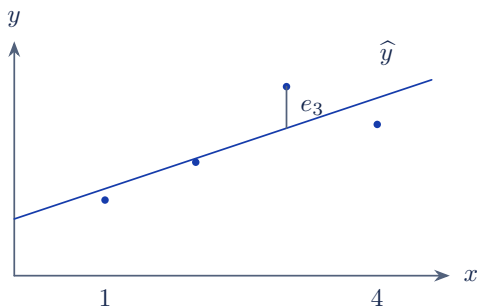
$-1 \leq r \leq 1$ เครื่องหมายแสดงทิศทางเชิงเส้น และ $|r|$ ใกล้ 1 แสดงความสัมพันธ์เชิงเส้นสูง ค่า $r = 0$ ไม่แสดงความสัมพันธ์แบบโค้ง และสหสัมพันธ์ไม่ยืนยันเหตุและผล

5.2 เส้นถดถอยและกำลังสองน้อยที่สุด

เส้นทำนาย y จาก x เขียน $\hat{y} = c + mx$ โดยเลือก m, c ให้ $\sum (y_i - \hat{y}_i)^2$ ต่ำสุด

$$m = \frac{S_{xy}}{S_{xx}}, \quad c = \bar{y} - m\bar{x} \quad (S_{xx} > 0)$$

$$\hat{y} = \bar{y} + m(x - \bar{x})$$



เส้นผ่าน $(\bar{x}, \bar{y})$ ค่าคลาดเคลื่อนแนวตั้งคือ $e_i = y_i - \hat{y}_i$ และ $\sum e_i = 0$ ความชันมีหน่วย y ต่อ x เมื่อ x เปลี่ยน Δx ค่าทำนายเปลี่ยน $m\Delta x$

5.3 การทำนายกลับทิศทาง

ถ้าต้องการทำนาย x จาก y ต้องหาสมการอีกเส้น

$$\hat{x} = \bar{x} + \frac{S_{xy}}{S_{yy}}(y - \bar{y}) \quad (S_{yy} > 0)$$

$$m_{y|x} m_{x|y} = r^2$$

$m_{y|x}, m_{x|y}$ คือความชันของเส้นทำนายแต่ละทิศ โดยทั่วไปไม่ใช้การกลับสมการเติมแทนเส้นถดถอยกลับทิศ

5.4 สมการปกติและตัวอย่างคำนวณ

สำหรับแบบจำลอง $\hat{y} = c + mx$ เจือไขกำลังสองน้อยที่สุดให้ระบบสมการ

$$\begin{cases} nc + m \sum x_i = \sum y_i \\ c \sum x_i + m \sum x_i^2 = \sum x_i y_i \end{cases}$$

ตัวอย่างย่อ ข้อมูล $(1, 2), (2, 3), (3, 5), (4, 4)$ ให้

$\bar{x}$	$\bar{y}$	S_{xx}	S_{yy}	S_{xy}
2.5	3.5	5	4	

$$r = 0.8, \quad m = 0.8, \quad c = 1.5, \quad \hat{y} = 1.5 + 0.8x$$

5.5 อนุกรมเวลาและการกำหนดรหัสเวลา

เส้นแนวโน้ม $\hat{y} = c + mt$ ใช้รหัสเวลา t_i ที่เพิ่มตามระยะเวลาจริง ถ้าเลือกให้ $\sum t_i = 0$ จะได้

$$c = \bar{y}, \quad m = \frac{\sum t_i y_i}{\sum t_i^2} \quad (\sum t_i^2 > 0)$$

คือ	-2	-1	0	1	2
คู่	-3	-1	1	3	

จำนวนช่วงที่ใช้รหัสสมมาตรเพิ่มทีละ 1 จำนวนช่วงที่ใช้รหัสสมมาตรเพิ่มทีละ 2 ได้ แต่ต้องแปลงเวลาที่ต้องการทำนายด้วยรหัสเดียวกัน

5.6 แบบจำลองกำลังสอง

กรณีบังคับผ่านจุดกำเนิดและไม่มีพจน์เชิงเส้น ใช้ $\hat{y} = kx^2$ ซึ่งให้

$$k = \frac{\sum x_i^2 y_i}{\sum x_i^4} \quad (\sum x_i^4 > 0)$$

แบบจำลองทั่วไป $\hat{y} = c + bx + ax^2$ ใช้ระบบ

$$\begin{cases} nc + b \sum x_i + a \sum x_i^2 = \sum y_i \\ c \sum x_i + b \sum x_i^2 + a \sum x_i^3 = \sum x_i y_i \\ c \sum x_i^2 + b \sum x_i^3 + a \sum x_i^4 = \sum x_i^2 y_i \end{cases}$$

ผลรวมทุกตัวใช้ข้อมูล n คู่เดียวกัน ต้องมีค่า x ต่างกันอย่างน้อยสามค่าเพื่อให้พารามิเตอร์ทั้งสามกำหนดได้เอกเทศ

5.7 ขอบเขตการใช้แบบจำลอง

- พิจารณาแผนภาพการกระจายก่อนเลือกเส้นตรงหรือเส้นโค้ง และตรวจหน่วยทุกตัวแปร
- เลือกทิศทางการทำนายให้ตรงคำถาม
- การทำนายนอกช่วงข้อมูลมีความไม่แน่นอนสูงขึ้น การแบ่งค่ารวมเป็นรายช่วงต้องมีสมมติฐานรองรับ