

ข้อมูล ความถี่ แผนภาพ และค่าเฉลี่ย
ข้อมูลและการแจกแจงความถี่

ข้อมูลเชิงคุณภาพแสดงประเภท ข้อมูลเชิงปริมาณแสดงจำนวนหรือค่าที่วัดได้

ความถี่ f คือจำนวนครั้งที่ค่าปรากฏ ความถี่สะสม F คือผลบวกความถี่ถึงค่านั้น

ข้อมูล 1, 1, 2, 2, 2, 3 มี $N = 6$ ค่า

ค่า	f	F	f/N	F/N
1	2	2	1/3	1/3
2	3	5	1/2	5/6
3	1	6	1/6	1

f/N คือความถี่สัมพัทธ์ F/N คือความถี่สะสมสัมพัทธ์ คุณ 100 ได้ร้อยละ

ผลรวมความถี่เท่ากับ N ผลรวมความถี่สัมพัทธ์เท่ากับ 1

อันตรภาคชั้นและขอบชั้น

ข้อมูลจำนวนเต็มทีวัดละเอียดหนึ่งหน่วย ชั้น 10–19 มีขอบล่าง 9.5 ขอบบน 19.5

$$I = \text{ขอบบน} - \text{ขอบล่าง}, \quad x_i = \frac{\text{ขอบล่าง} + \text{ขอบบน}}{2}$$

I คือความกว้างชั้น x_i คือจุดกึ่งกลาง

เลือกชั้นให้ครอบคลุมค่าต่ำสุดถึงสูงสุด ไม่ซ้ำหรือขาดช่วง แล้วนับข้อมูลลงชั้นตามขอบที่กำหนด

พิสัย $R = x_{\max} - x_{\min}$

ฮิสโทแกรมกับแผนภูมิแท่ง

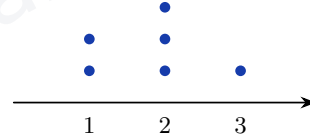
แผนภูมิแท่งใช้เปรียบเทียบประเภทหรือค่าที่แยกเป็นหมวด แท่งเว้นช่องได้

ฮิสโทแกรมใช้ข้อมูลเชิงปริมาณที่จัดเป็นช่วง แท่งต่อเนื่องกัน และพื้นที่แท่งแทนความถี่

ชั้นกว้างเท่ากัน ใช้ความสูงตามความถี่ได้ ถ้ากว้างไม่เท่ากัน ใช้ความสูง f/I เพื่อให้พื้นที่แท่งเท่ากับความถี่

แผนภาพจุดและต้น-ใบ

จุดสามจุดเหนือค่า 2 หมายถึงข้อมูล 2, 2, 2



ต้น-ใบสำหรับจำนวนสองหลัก: ต้นเป็นหลักสิบ ใบเป็นหลักหน่วย

2 | 1 4 4 8 อ่านเป็น 21, 24, 24, 28

ใบที่ซ้ำต้องแสดงทุกตัว จึงนับจำนวนข้อมูลและหามัธยฐานหรือฐานนิยมได้ถูกต้อง

ค่าเฉลี่ยและค่าเฉลี่ยรวม

ข้อมูล $x_1, \dots, x_N$:

$$\bar{x} = \frac{\sum x_i}{N}$$

จากตารางความถี่: $\bar{x} = \frac{\sum f_i x_i}{\sum f_i}$

ค่าเฉลี่ยถ่วงน้ำหนัก ใช้น้ำหนัก w_i

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

รวมหลายกลุ่มที่มีจำนวน N_i และค่าเฉลี่ย $\bar{x}_i$

$$\bar{x}_{\text{รวม}} = \frac{\sum N_i \bar{x}_i}{\sum N_i}$$

ห้ามเฉลี่ยค่าเฉลี่ยกลุ่มโดยตรงถ้าขนาดกลุ่มไม่เท่ากัน

ตารางจัดชั้นและค่าเฉลี่ยประมาณ

ชั้น	f_i	F_i	x_i	$f_i x_i$
10–19	2	2	14.5	29
20–29	5	7	24.5	122.5
30–39	3	10	34.5	103.5
รวม	10			255

ใช้จุดกึ่งกลางแทนค่าทั้งชั้น

$$\bar{x} \approx \frac{255}{10} = 25.5$$

เป็นค่าประมาณ เพราะไม่ทราบข้อมูลแต่ละตัวในชั้น

ค่ากลาง การกระจาย และควอร์ไทล์

มัธยฐานของข้อมูลรายตัว

เรียงข้อมูลจากน้อยไปมากก่อน ให้ $x_{(j)}$ หมายถึงค่าลำดับที่ j
 N คือ: $M = x_{((N+1)/2)}$
 N คู่: $M = [x_{(N/2)} + x_{(N/2+1)}]/2$
 ในตารางความถี่ ใช้ความถี่สะสมค้นหาค่าที่ครอบคลุมตำแหน่งกลาง

มัธยฐานของข้อมูลจัดชั้น

ชั้นมัธยฐานคือชั้นแรกที่มีความถี่สะสมถึงหรือเกิน $N/2$

$$M \approx L + \frac{N/2 - F}{f} I$$
 L คือขอบล่างชั้นมัธยฐาน F คือความถี่สะสมก่อนชั้น f คือความถี่ของชั้น และ I คือความกว้างชั้น
 จากตารางหน้าแรก $N/2 = 5$ อยู่ในชั้น 20–29

$$M \approx 19.5 + \frac{5 - 2}{5}(10) = 25.5$$
 สมมติข้อมูลกระจายสม่ำเสมอภายในชั้นที่ประมาณ

ฐานนิยมและการเลือกค่ากลาง

ข้อมูลรายตัว: ค่าที่มีความถี่สูงสุด อาจมีหลายค่าหรือไม่มีค่าที่เด่นกว่า
 ข้อมูลจัดชั้นกว้างเท่ากัน มีชั้นฐานนิยมเด่นชัด

$$Mo \approx L + \frac{D_1}{D_1 + D_2} I$$
 L คือขอบล่างชั้นฐานนิยม I คือความกว้างชั้น
 D_1 คือความถี่ชั้นฐานนิยมลบชั้นก่อนหน้า
 D_2 คือความถี่ชั้นฐานนิยมลบชั้นถัดไป
 จากตาราง $D_1 = 5 - 2 = 3$, $D_2 = 5 - 3 = 2$

$$Mo \approx 19.5 + \frac{3}{3 + 2}(10) = 25.5$$

ค่าเฉลี่ยใช้ทุกค่าแต่ไวต่อค่าสุดโต่ง มัธยฐานทนต่อค่าสุดโต่งกว่า ฐานนิยมใช้กับข้อมูลประเภทได้

ความแปรปรวนและส่วนเบี่ยงเบนมาตรฐาน

ประชากร N ค่า ค่าเฉลี่ย μ :

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

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

เมื่อข้อมูลมีความถี่ ใช้ $\sum f_i x_i^2$ แทน $\sum x_i^2$ และ $N = \sum f_i$
 ความแปรปรวนมีหน่วยกำลังสอง ส่วนเบี่ยงเบนมาตรฐานมีหน่วยเดียวกับข้อมูล

ถ้าแปลง $y = ax + b$ ค่าเฉลี่ยใหม่ $a\mu + b$

ส่วนเบี่ยงเบนมาตรฐาน $|a|\sigma$ และความแปรปรวน $a^2\sigma^2$

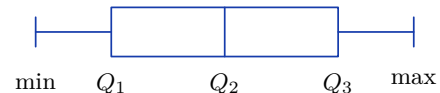
ควอร์ไทล์และแผนภาพกล่อง

เรียงข้อมูล แล้วใช้ตำแหน่งตามข้อตกลงของเล่มหลัก

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

ตำแหน่งไม่เต็มจำนวนให้แทรกค่าตามสัดส่วนระหว่างลำดับที่ติดกัน โดย Q_2 คือมัธยฐาน

ควอร์ไทล์แบ่งตำแหน่งข้อมูลเป็นสี่ส่วน ส่วนละประมาณ 25%



$$IQR = Q_3 - Q_1$$

IQR คือพิสัยระหว่างควอร์ไทล์ ช่วง Q_1 ถึง Q_3 ครอบคลุมตำแหน่งข้อมูลกลางประมาณ 50%

ความยาวช่วงในกล่องบอกการกระจาย ไม่ได้บอกจำนวนข้อมูลมากกว่าช่วงอื่น และข้อมูลซ้ำทำให้สัดส่วนที่นับได้ต่างจาก 25%

การเปลี่ยนข้อมูลและการตีความ

ถ้าแก้ค่าผิด u เป็น v โดยจำนวนข้อมูลเดิม N คงที่ ค่าเฉลี่ยเปลี่ยนเป็น $\bar{x} + (v - u)/N$

หากโจทย์กำหนดการแจกแจงปกติ ช่วง $\mu \pm \sigma$, $\mu \pm 2\sigma$, $\mu \pm 3\sigma$ มีข้อมูลประมาณ 68%, 95%, 99.7% ตามลำดับ