

1. การวัดค่ากลางของข้อมูล

1. ค่าเฉลี่ยเลขคณิต ใช้สัญลักษณ์ \bar{X}

ข้อมูลไม่แจกแจงความถี่

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{N} \quad \text{หรือ} \quad \bar{X} = \frac{\sum x}{N}$$

วิธีตรง
$$\bar{X} = \frac{\sum_{i=1}^k f_i x_i}{\sum_{i=1}^k f_i} \quad \text{หรือ} \quad \bar{X} = \frac{\sum fx}{N}$$

วิธีตัด
$$\bar{X} = a + I \left(\frac{\sum_{i=1}^k f_i d_i}{N} \right) \quad \text{หรือ} \quad \bar{X} = a + Id$$

ค่าเฉลี่ยเลขคณิตถ่วงน้ำหนัก

ใช้ในกรณีที่ข้อมูลแต่ละค่ามีความสำคัญไม่เท่ากัน

ให้ $w_1, w_2, w_3, \dots, w_n$ เป็นความสำคัญหรือน้ำหนักถ่วงของค่าจากการสังเกต $x_1, x_2, x_3, \dots, x_n$ ตามลำดับ

ค่าเฉลี่ยเลขคณิตถ่วงน้ำหนัก
$$\bar{X} = \frac{\sum_{i=1}^N w_i x_i}{\sum_{i=1}^N w_i} \quad \text{หรือ} \quad \bar{X} = \frac{\sum wx}{\sum w}$$

ค่าเฉลี่ยเลขคณิตรวม
$$\bar{X}_{รวม} = \frac{\sum_{i=1}^k N_i \bar{X}_i}{\sum_{i=1}^k N_i} \quad \text{หรือ} \quad \bar{X}_{รวม} = \frac{\sum N\bar{X}}{\sum N}$$

\bar{X}_i คือ ค่าเฉลี่ยเลขคณิตของข้อมูลชุดที่ i

N_i คือ จำนวนข้อมูลในชุดที่ i

สมบัติของค่าเฉลี่ยเลขคณิต

1. $\sum_{i=1}^N x_i = N\bar{X}$
2. $\sum_{i=1}^N (x_i - \bar{X}) = 0$
3. $\sum_{i=1}^N (x_i - \bar{X})^2$ มีค่าน้อยที่สุด
4. $x_{\min} < \bar{X} < x_{\max}$
5. ถ้า $y_i = ax_i + b$, $i = 1, 2, 3, \dots, N$ เมื่อ a, b เป็นค่าคงตัวใดๆ
แล้ว $\bar{Y} = a\bar{X} + b$

ค่าเฉลี่ยเลขคณิตเปลี่ยนแปลงได้ง่ายเมื่อข้อมูลเปลี่ยน ในกรณีที่มีการแจกแจงความถี่เป็นอันตรภาคชั้นเปิด หรือข้อมูลที่มีบางค่าสูงหรือต่ำผิดปกติไม่ควรใช้ \bar{X}

2. มัชฐาน ใช้สัญลักษณ์ Med

มัชฐาน คือ ค่าที่มีตำแหน่งอยู่กึ่งกลางของข้อมูลทั้งหมด เมื่อเรียงข้อมูลจากน้อยไปมาก

ข้อมูลไม่แจกแจงความถี่

1. เรียงข้อมูลจากน้อยไปมาก
2. ตำแหน่งของมัชฐาน คือ $\frac{N+1}{2}$

ข้อมูลแจกแจงความถี่

1. หาคความถี่สะสม
2. หาค่า $\frac{N}{2}$
3. ถ้า $\frac{N}{2}$ เท่ากับความถี่สะสมชั้นใดชั้นหนึ่ง

มัชฐาน = ขอบบนของอันตรภาคชั้นนั้น

ถ้า $\frac{N}{2}$ ไม่เท่ากับความถี่สะสมชั้นใดเลย หาค่ามัชฐานได้

จากการเทียบบัญญัติไตรยางศ์ หรือใช้สูตร

$$\text{Med} = L + \left(\frac{\frac{N}{2} - \sum f_L}{f_m} \right) I$$

สมบัติของมัชฐาน

$$\sum_{i=1}^N |x_i - \text{มัชฐาน}| \text{ มีค่าน้อยที่สุด}$$

สำหรับข้อมูลที่มีค่าสูงหรือต่ำผิดปกติ ควรหาค่ากลางโดยใช้มัชฐาน

3. ฐานนิยม ใช้สัญลักษณ์ Mo

ฐานนิยม คือ ค่าของข้อมูลที่มีความถี่สูงสุด

4. ค่าเฉลี่ยเรขาคณิต ใช้สัญลักษณ์ G.M.

ให้ $x_1, x_2, x_3, \dots, x_N$ เป็นข้อมูล N จำนวน

$$\text{ค่าเฉลี่ยเรขาคณิต (G.M.)} = \sqrt[N]{x_1 x_2 x_3 \dots x_N}$$

5. ค่าเฉลี่ยฮาร์โมนิก ใช้สัญลักษณ์ H.M.

ให้ $x_1, x_2, x_3, \dots, x_N$ เป็นข้อมูล N จำนวน

$$\text{ค่าเฉลี่ยฮาร์โมนิก (H.M.)} = \frac{N}{\sum_{i=1}^N \frac{1}{x_i}}$$

6. ค่ากึ่งกลางพิสัย

$$\text{ค่ากึ่งกลางพิสัย} = \frac{x_{\max} + x_{\min}}{2}$$

2. ควอร์ไทล์ เดซิซัล และเปอร์เซ็นไทล์

ควอร์ไทล์ เดซิซัล และเปอร์เซ็นไทล์ คือ ค่าของข้อมูลที่แบ่งจำนวนข้อมูลออกเป็น 4 ส่วน 10 ส่วน และ 100 ส่วน ตามลำดับ

ข้อควรทราบ

- $Q_2 = D_5 = P_{50} = \text{มัธยฐาน}$
- $Q_1 = P_{25}$, $Q_2 = P_{50}$, $Q_3 = P_{75}$
- $D_1 = P_{10}$, $D_2 = P_{20}$, $D_3 = P_{30}$, \dots , $D_9 = P_{90}$

การหาค่าของควอร์ไทล์ เดซิซัล และ เปอร์เซ็นไทล์

1. ข้อมูลไม่แจกแจงความถี่

- เรียงข้อมูลจากน้อยไปมาก
- หาดำแหน่งที่ของควอร์ไทล์ เดซิซัล และเปอร์เซ็นไทล์ จากสูตรต่อไปนี้

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

$$\text{ตำแหน่งที่ของ } D_r = \frac{r(N+1)}{10} , r = 1, 2, 3, \dots, 9$$

$$\text{ตำแหน่งที่ของ } P_r = \frac{r(N+1)}{100} , r = 1, 2, 3, \dots, 99$$

- ตำแหน่งที่ได้จากข้อ 2 ตรงกับข้อมูลตัวใด จะได้ว่า ข้อมูลตัวนั้นเป็นคำตอบที่ต้องการ แต่ถ้าไม่ตรงกับข้อมูลตัวใดเลย ให้ใช้วิธีเทียบบัญญัติไตรยางค์ในการหาคำตอบ

2. ข้อมูลแจกแจงความถี่

1. สร้างตารางความถี่สะสม
2. หาคำแหน่งที่ของควอร์ไทล์ เดซิซัล และเปอร์เซ็นไทล์ โดยใช้สูตร

$$\text{ตำแหน่งที่ของ } Q_r = \frac{rN}{4}, \quad r = 1, 2, 3$$

$$\text{ตำแหน่งที่ของ } D_r = \frac{rN}{10}, \quad r = 1, 2, 3, \dots, 9$$

$$\text{ตำแหน่งที่ของ } P_r = \frac{rN}{100}, \quad r = 1, 2, 3, \dots, 99$$

3. ถ้าตำแหน่งที่ของ Q_r , D_r หรือ P_r เท่ากับความถี่สะสมชั้นใดชั้นหนึ่ง

จะได้ว่า Q_r , D_r หรือ P_r = ขอบบนของอันตรภาคชั้นนั้น

แต่ถ้าตำแหน่งที่ของ Q_r , D_r หรือ P_r ไม่เท่ากับความถี่สะสมชั้นใดเลย ให้ใช้วิธีเทียบ

บัญญัติไตรยางศ์ หรือใช้สูตร

$$Q_r = L + \left(\frac{\frac{rN}{4} - \sum f_L}{f_{Q_r}} \right) I$$

$$D_r = L + \left(\frac{\frac{rN}{10} - \sum f_L}{f_{D_r}} \right) I$$

$$P_r = L + \left(\frac{\frac{rN}{100} - \sum f_L}{f_{P_r}} \right) I$$

3. การวัดการกระจายของข้อมูล

1. การวัดการกระจายสัมบูรณ์ คือ การวัดการกระจายของข้อมูลเพียงชุดเดียว มี 4 ชนิด

1.1 พิสัย ใช้สัญลักษณ์ R

$$\text{พิสัย} = x_{\max} - x_{\min}$$

1.2 ส่วนเบี่ยงเบนควอร์ไทล์ ใช้สัญลักษณ์ Q.D.

$$Q.D. = \frac{Q_3 - Q_1}{2}$$

1.3 ส่วนเบี่ยงเบนเฉลี่ย ใช้สัญลักษณ์ M.D.

เมื่อข้อมูลไม่แจกแจงความถี่

ให้ x_1, x_2, \dots, x_n เป็นข้อมูล N จำนวน

$$M.D. = \frac{\sum_{i=1}^N |x_i - \bar{X}|}{N}$$

เมื่อข้อมูลแจกแจงความถี่

ให้ x_i เป็นจุดกึ่งกลางของอันตรภาคชั้นที่ i

f_i เป็นความถี่ของอันตรภาคชั้นที่ i

$$M.D. = \frac{\sum_{i=1}^N f_i |x_i - \bar{X}|}{N}$$

1.4 ส่วนเบี่ยงเบนมาตรฐาน ใช้สัญลักษณ์ S.D. หรือ s

เมื่อข้อมูลไม่แจกแจงความถี่

ให้ x_1, x_2, \dots, x_n เป็นข้อมูล N จำนวน

$$S.D. = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{X})^2}{N}}$$

หรือ

$$S.D. = \sqrt{\frac{\sum_{i=1}^N x_i^2}{N} - (\bar{X})^2}$$

เมื่อข้อมูลแจกแจงความถี่

$$S.D. = \sqrt{\frac{\sum_{i=1}^N f_i (x_i - \bar{X})^2}{N}}$$

หรือ

$$S.D. = \sqrt{\frac{\sum_{i=1}^N f_i x_i^2}{N} - (\bar{X})^2}$$

ความแปรปรวน (Variance) = (ส่วนเบี่ยงเบนมาตรฐาน)² ใช้สัญลักษณ์ $S.D.^2$

หรือ s^2 หรือ σ^2 แทนความแปรปรวน

ความแปรปรวนรวม (combined or pooled variance)

ให้ N_1, N_2 เป็นจำนวนข้อมูล ของข้อมูลชุดที่ 1 และชุดที่ 2 ตามลำดับ

ให้ s_1^2, s_2^2 เป็นความแปรปรวนของข้อมูลชุดที่ 1 และชุดที่ 2 ตามลำดับ

\bar{X}_1, \bar{X}_2 เป็นค่าเฉลี่ยเลขคณิตของข้อมูลชุดที่ 1 และชุดที่ 2 ตามลำดับ

1. ถ้า $\bar{X}_1 = \bar{X}_2$ แล้ว

$$s_{รวม}^2 = \frac{N_1 s_1^2 + N_2 s_2^2}{N_1 + N_2}$$

2. ถ้า $N_1 = N_2$ และ $\bar{X}_1 = \bar{X}_2$

$$s_{รวม}^2 = \frac{s_1^2 + s_2^2}{2}$$

3. ถ้า $\bar{X}_1 \neq \bar{X}_2$ แล้ว

$$s_{รวม}^2 = \frac{N_1 s_1^2 + N_2 s_2^2 + N_1 (\bar{X}_1 - \bar{X}_{รวม})^2 + N_2 (\bar{X}_2 - \bar{X}_{รวม})^2}{N_1 + N_2}$$

สมบัติของการวัดการกระจายสัมบูรณ์

1. การวัดการกระจายสัมบูรณ์จะมีค่าไม่เป็นลบ
2. การวัดการกระจายมีค่าเป็นศูนย์ก็ต่อเมื่อ ข้อมูลทุกตัวมีค่าเท่ากันหมด
3. ถ้าข้อมูลชุดหนึ่งเปลี่ยนเป็นข้อมูลอีกชุดหนึ่ง โดยใช้ค่าคงที่ใดๆ ไปบวก หรือลบข้อมูลทุกๆตัว แล้วการวัดการกระจายมีค่าเท่าเดิม
4. ถ้าข้อมูลชุดหนึ่งเปลี่ยนเป็นข้อมูลอีกชุดหนึ่ง โดยใช้ค่าคงที่ k ใดๆ ($k \neq 0$) ไปคูณ หรือหารข้อมูลทุกๆตัว แล้วการวัดการกระจายของข้อมูลชุดใหม่ เท่ากับการวัดการกระจายของข้อมูลชุดเดิมคูณด้วย $|k|$

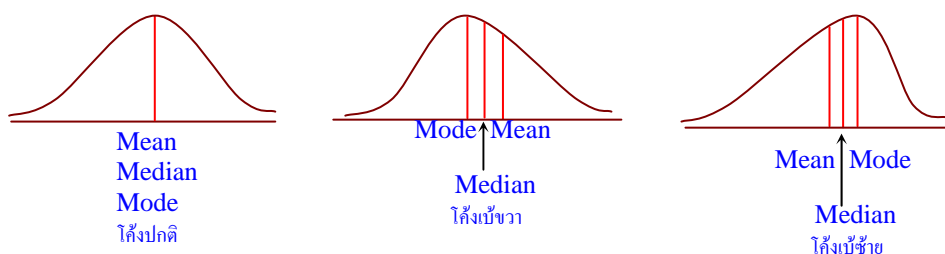
2. การวัดการกระจายสัมพัทธ์

การวัดการกระจายสัมพัทธ์ ใช้ในการเปรียบเทียบข้อมูลตั้งแต่สองชุดขึ้นไป เพื่อตัดสินว่าข้อมูลชุดใดมีการกระจายมากกว่ากัน ค่าของการวัดการกระจายสัมพัทธ์ไม่มีหน่วย จึงสามารถวัดการกระจายของข้อมูลที่มีหน่วยต่างกันได้

การวัดการกระจายสัมพัทธ์ที่นิยมใช้มี 4 ชนิด คือ

1. สัมประสิทธิ์ของพิสัย (Coefficient of range) $= \frac{X_{\max} - X_{\min}}{X_{\max} + X_{\min}}$
2. สัมประสิทธิ์ของส่วนเบี่ยงเบนควอร์ไทล์ (Coefficient of quartile) $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$
3. สัมประสิทธิ์ของส่วนเบี่ยงเบนเฉลี่ย (coefficient of mean deviation) $= \frac{M.D.}{\bar{X}}$
4. สัมประสิทธิ์ของการแปรผัน (coefficient of variation : V) $= \frac{S.D.}{\bar{X}}$

ความสัมพันธ์ระหว่างการแจกแจงความถี่ ค่ากลาง และการกระจายของข้อมูล



4. ค่ามาตรฐานหรือคะแนนมาตรฐาน

$$z_i = \frac{x_i - \bar{X}}{s}$$

เมื่อ z_i = ค่ามาตรฐาน

x_i = ค่าของข้อมูลดิบที่ต้องการแปลงให้เป็นค่ามาตรฐาน

\bar{X} = ค่าเฉลี่ยเลขคณิตของข้อมูลชุดนั้น

s = ส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุดนั้น

ข้อควรจำเกี่ยวกับค่ามาตรฐาน

1. ค่ามาตรฐานไม่มีหน่วย
2. ค่ามาตรฐานอาจมีค่าเป็นบวก ลบ หรือศูนย์ก็ได้
3. ถ้า $z_1, z_2, z_3, \dots, z_n$ เป็นค่ามาตรฐานของ $x_1, x_2, x_3, \dots, x_n$

ซึ่งเป็นข้อมูลชุดหนึ่งตามลำดับ จะได้ว่า

3.1 ผลรวมของค่ามาตรฐานทุกค่าในข้อมูลนั้นมีค่าเป็นศูนย์ $\left(\sum_{i=1}^N z_i = 0 \right)$

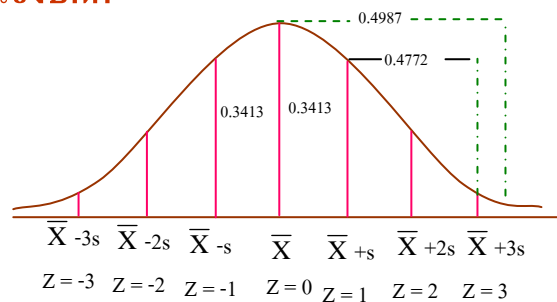
3.2 ค่าเฉลี่ยเลขคณิตของค่ามาตรฐานมีค่าเป็นศูนย์ $(\bar{z} = 0)$

3.3 ผลรวมกำลังสองของค่ามาตรฐานจะเท่ากับจำนวนข้อมูล $\left(\sum_{i=1}^N z_i^2 = N \right)$

3.4 ความแปรปรวน หรือส่วนเบี่ยงเบนมาตรฐานมีค่าเป็น 1

$$(s_z^2 = 1 \text{ หรือ } s_z = 1)$$

5. การแจกแจงปกติ



กำหนดให้พื้นที่ใต้เส้นโค้งปกติมีค่าเท่ากับ 1 หน่วย

พื้นที่ช่วง $\bar{X} - s$ ถึง $\bar{X} + s$ คิดเป็น 68.26% ของพื้นที่ทั้งหมด

พื้นที่ช่วง $\bar{X} - 2s$ ถึง $\bar{X} + 2s$ คิดเป็น 95.46% ของพื้นที่ทั้งหมด

พื้นที่ช่วง $\bar{X} - 3s$ ถึง $\bar{X} + 3s$ คิดเป็น 99.74% ของพื้นที่ทั้งหมด

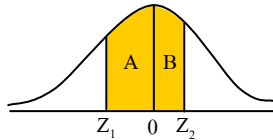
โค้งการแจกแจงปกติจะมีลักษณะสมมาตร (symmetry) คือ พื้นที่ทางด้านซ้ายของ \bar{X} กับพื้นที่ทางด้านขวาของ \bar{X} เท่ากัน คือ เท่ากับ 50% ของพื้นที่ทั้งหมด

ตารางแสดงพื้นที่ใต้เส้นโค้งปกติจึงแสดงเฉพาะค่า z ที่เป็นบวก และการคิดพื้นที่ที่จะคิดจาก $z = 0$ ไปทางซ้าย หรือ ขวาเสมอ

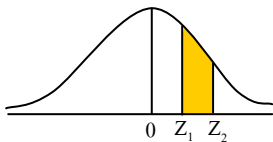
การหาพื้นที่ใต้เส้นโค้งปกติในลักษณะต่างๆ

ให้พื้นที่เส้นโค้งปกติระหว่าง $z = 0$ ถึง $z = z_1$ เป็น A

พื้นที่เส้นโค้งปกติระหว่าง $z = 0$ ถึง $z = z_2$ เป็น B



พื้นที่ใต้เส้นโค้งปกติระหว่าง z_1 ถึง z_2 เท่ากับ $A+B$



พื้นที่ใต้เส้นโค้งปกติระหว่าง z_1 ถึง z_2 เท่ากับ $B-A$

6. การวิเคราะห์ความสัมพันธ์เชิงฟังก์ชันระหว่างข้อมูล

ความสัมพันธ์เชิงฟังก์ชันระหว่างข้อมูล แบ่งออกเป็น 2 ชนิด คือ

1. ความสัมพันธ์เชิงฟังก์ชันที่เป็นแบบเส้นตรง หรือความสัมพันธ์เชิงเส้น

$$\text{มีรูปสมการเป็น } Y = mX + c$$

$$\begin{aligned} \text{สมการปกติ คือ } \sum_{i=1}^N Y_i &= m \sum_{i=1}^N X_i + cN \\ \sum_{i=1}^N X_i Y_i &= m \sum_{i=1}^N X_i^2 + c \sum_{i=1}^N X_i \end{aligned}$$

2. ความสัมพันธ์เชิงฟังก์ชันที่ไม่เป็นเส้นตรง

2.1 ความสัมพันธ์ที่อยู่ในรูปพาราโบลา มีรูปสมการเป็น

$$Y = aX^2 + bX + c, a \neq 0$$

$$\begin{aligned} \text{สมการปกติ คือ } \sum_{i=1}^N Y_i &= a \sum_{i=1}^N X_i^2 + b \sum_{i=1}^N X_i + cN \\ \sum_{i=1}^N X_i Y_i &= a \sum_{i=1}^N X_i^3 + b \sum_{i=1}^N X_i^2 + c \sum_{i=1}^N X_i \\ \sum_{i=1}^N X_i^2 Y_i &= a \sum_{i=1}^N X_i^4 + b \sum_{i=1}^N X_i^3 + c \sum_{i=1}^N X_i^2 \end{aligned}$$

2.2 ความสัมพันธ์ที่อยู่ในรูปเอกซ์โพเนนเชียล มีรูปสมการเป็น

$$Y = ab^x \quad \text{หรือ} \quad \log y = \log a + x \log b$$

$$\text{สมการปกติคือ} \quad \sum_{i=1}^N \log Y_i = n \log a + (\log b) \sum_{i=1}^N X_i$$

$$\sum_{i=1}^N X_i \log Y_i = (\log a) \sum_{i=1}^N X_i + (\log b) \sum_{i=1}^N X_i^2$$

โดยที่ X เป็นตัวแปรต้นหรือตัวแปรอิสระ

Y เป็นตัวแปรตาม

a, b, c และ m เป็นค่าคงที่

การวิเคราะห์ความสัมพันธ์เชิงฟังก์ชัน

1. เขียนแผนภาพจากการกระจาย
2. เขียนเส้นที่ใช้ประมาณความสัมพันธ์จากแผนภาพการกระจายในข้อ 1
3. หาสมการปกติของเส้นที่ใช้ประมาณในข้อ 2 โดยใช้ระเบียบวิธีกำลังสองน้อยที่สุด

ในการสร้างความสัมพันธ์เชิงฟังก์ชันระหว่างข้อมูลสองชุด ถ้าต้องการคาดคะเนหรือทำนายค่าของข้อมูลชุดใด ต้องให้ค่าของข้อมูลชุดนั้นเป็นตัวแปรเสมอ เราไม่สามารถทำนายค่าตัวแปร Y จากตัวแปร X และทำนายค่าของ X จากตัวแปร Y โดยใช้สมการเดียวกัน

ความสัมพันธ์เชิงฟังก์ชันของข้อมูลที่อยู่ในรูปอนุกรมเวลา

ข้อมูลที่อยู่ในรูปอนุกรมเวลา คือ ข้อมูลที่แสดงการเปลี่ยนแปลงตามลำดับเวลาที่เกิดขึ้น ซึ่งความสัมพันธ์นี้อยู่ในรูป $Y = f(X)$ โดยที่ X แทนเวลา เช่น ปี พ.ศ. เดือน หรือวัน เป็นต้น และเป็นตัวแปรอิสระ ส่วน Y แทนค่าของข้อมูลและเป็นตัวแปรตาม

การกำหนดค่า X แทนเวลา มีลักษณะดังนี้

1. ถ้าจำนวนระยะเวลาที่กำหนดมาให้เป็นจำนวนคี่ ให้ระยะเวลาที่อยู่ตรงกลางเป็น 0 โดยเรียงลำดับก่อนหลังเป็น $\dots, -3, -2, -1, 0, 1, 2, 3, \dots$
2. ถ้าจำนวนระยะเวลาที่กำหนดมาให้เป็นจำนวนคู่ ให้ระยะเวลาที่อยู่ตรงกลางเป็น 1 และ -1 โดยเรียงลำดับก่อนหลังเป็น $\dots, -5, -3, -1, 1, 3, 5, \dots$