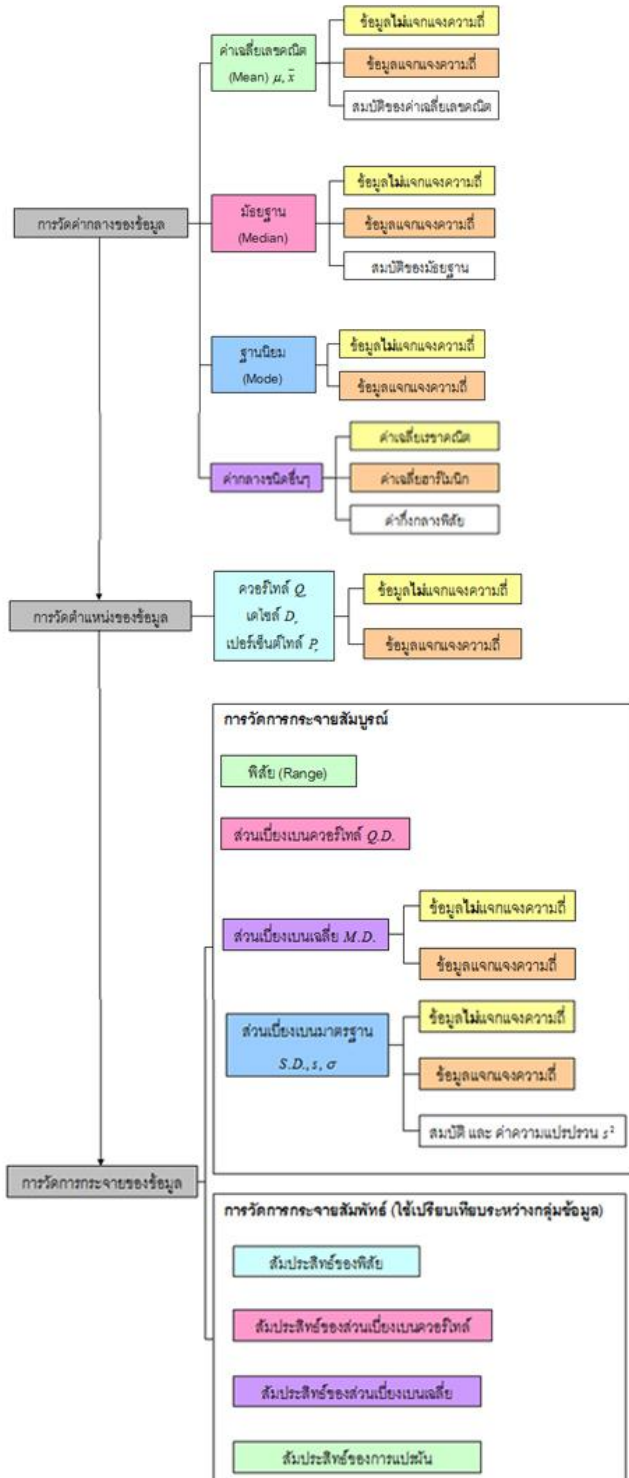


การวิเคราะห์ข้อมูลเบื้องต้น ม.6 เทอมต้น สารการเรียนรู้เพิ่มเติม



สมบัติของ Σ

ความสัมพันธ์ระหว่าง การแจกแจงความถี่, ค่ากลาง, การกระจาย

การวัดค่ากลางของข้อมูล

ค่าเฉลี่ยเลขคณิต (Mean)

ค่าเฉลี่ยเลขคณิต μ, \bar{x} คือ ผลรวมของค่าของข้อมูลทั้งหมด หารด้วยจำนวนของข้อมูล

ข้อมูลไม่แจกแจงความถี่

1) ทั่วไป

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

x_i คือ ค่าของข้อมูลแต่ละตัว

N คือ จำนวนของข้อมูล

2) ถ่วงน้ำหนัก

$$\bar{x} = \frac{\sum_{i=1}^N w_i x_i}{\sum_{i=1}^N w_i}$$

x_i คือ ค่าของข้อมูลแต่ละตัว

w_i คือ ค่าถ่วงน้ำหนักของข้อมูลแต่ละตัว

N คือ จำนวนของข้อมูล

3) ค่าเฉลี่ยเลขคณิตรวม

$$\bar{x} = \frac{\sum_{i=1}^k N_i \bar{x}_i}{\sum_{i=1}^k N_i}$$

\bar{x}_i คือ ค่าเฉลี่ยเลขคณิตของแต่ละกลุ่ม

N_i คือ จำนวนข้อมูลของแต่ละกลุ่ม

k คือ จำนวนกลุ่ม

ข้อมูลแจกแจงความถี่ (ตาราง)

$$\bar{x} = \frac{\sum_{i=1}^k f_i x_i}{\sum_{i=1}^k f_i}$$

f_i คือ ความถี่ของชั้นที่ i

x_i คือ จุดกึ่งกลางชั้นของชั้นที่ i

$\sum_{i=1}^k f_i$ หรือ N คือ จำนวนของข้อมูลทั้งหมด

k คือ จำนวนชั้นของข้อมูล

สมบัติของค่าเฉลี่ยเลขคณิต

$$1) \sum_{i=1}^N (x_i - \bar{x}) = 0$$

$$2) \sum_{i=1}^N (x_i - a)^2 \text{ มีค่าน้อยที่สุด เมื่อ } a = \bar{x}$$

$$3) x_{\min} \leq \bar{x} \leq x_{\max}$$

ถ้า $x_{\min} = \bar{x} = x_{\max}$ แสดงว่า ข้อมูลทุกตัวมีค่าเท่ากัน

$$4) y_i = a \cdot x_i + b \text{ จะได้ว่า } \bar{y} = a \cdot \bar{x} + b \text{ โดยที่ } a, b \text{ เป็นค่าคงตัวใดๆ}$$

มัธยฐาน (Median)

มัธยฐาน คือ ค่าที่อยู่ตำแหน่งตรงกลางของข้อมูลทั้งหมด
ที่เรียงลำดับจากน้อยไปมาก หรือมากไปน้อย

ข้อมูลไม่แจกแจงความถี่

ขั้นตอนการหาค่ามัธยฐาน

- 1) เรียงข้อมูลทั้งหมดจากน้อยไปมาก
- 2) หาดำแหน่งมัธยฐานได้จากสูตร

$$\text{ตำแหน่งมัธยฐาน} = \frac{N+1}{2}$$

- 3) หาค่ามัธยฐาน

ค่าของข้อมูลที่เรียงจากน้อยไปมากที่มีตำแหน่งตรงกับตำแหน่งมัธยฐาน

ค่านั้นคือค่ามัธยฐาน

ข้อมูลแจกแจงความถี่ (ตาราง)

ขั้นตอนการหาค่ามัธยฐาน

- 1) หาความถี่สะสมของแต่ละชั้น
- 2) หาดำแหน่งมัธยฐานได้จากสูตร

$$\text{ตำแหน่งมัธยฐาน} = \frac{N}{2} \leftarrow \text{ตำแหน่ง}$$

- 3) หาชั้นที่ตำแหน่งอยู่

ถ้าตำแหน่งมีค่าอยู่ภายในความถี่สะสมของชั้นใด ตำแหน่งจะอยู่ในชั้นนั้น

- 4) หาค่าของตำแหน่ง (ที่อยู่ในรูปตาราง) จากสูตร

$$\text{ค่าของตำแหน่ง} = L + \left(\frac{T - \sum f_L}{f_T} \right) I$$

L คือ ขอบล่างของชั้นที่ตำแหน่งอยู่

T คือ ตำแหน่ง

$\sum f_L$ คือ ผลรวมของความถี่ของชั้นที่มีค่าต่ำกว่าชั้นที่ตำแหน่งอยู่

หรือ เท่ากับความถี่สะสมของชั้นก่อนหน้าของชั้นที่ตำแหน่งอยู่

f_T คือ ความถี่ของชั้นที่ตำแหน่งอยู่

I คือ ความกว้างของอันตรภาคชั้น

Note

ถ้าตำแหน่งไปตรงกับความถี่สะสมของชั้นใด

ค่าของตำแหน่งจะมีค่าเท่ากับขอบบนของชั้นนั้น

สมบัติของมัธยฐาน

$\sum_{i=1}^N |x_i - a|$ มีค่าน้อยที่สุด เมื่อ $a = \text{ค่ามัธยฐาน}$

ฐานนิยม (Mode)

ฐานนิยม คือ ค่าของข้อมูลที่มีความถี่สูงสุด

Note

ข้อมูลชุดหนึ่งมีฐานนิยมได้มากที่สุด 2 ค่า

ถ้ามากกว่า 2 ค่า จะถือว่าข้อมูลชุดนั้นไม่มีฐานนิยม

ข้อมูลไม่แจกแจงความถี่

ข้อมูลค่าใดมีจำนวนตัวมากที่สุด ข้อมูลค่านั้นเป็นฐานนิยม

ข้อมูลแจกแจงความถี่ (ตาราง)

ฐานนิยมมีค่าอยู่ในชั้นที่มีความถี่สูงสุด

(ตามหลักสูตรปัจจุบัน ฐานนิยม \approx จุดกึ่งกลางชั้นของชั้นที่มีความถี่สูงสุด)

ค่ากลางชนิดอื่นๆ

ค่าเฉลี่ยเรขาคณิต

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

ค่าเฉลี่ยฮาร์โมนิก

$$H.M. = \frac{N}{\sum_{i=1}^N \frac{1}{x_i}}$$

ค่ากึ่งกลางพิสัย

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

การวัดตำแหน่งของข้อมูล

ควอร์ไทล์, เดไซล์, เปอร์เซ็นไทล์

ควอร์ไทล์ที่ r (Q_r) คือ ค่าที่อยู่ตำแหน่งที่ r ส่วน ใน 4 ส่วน ของข้อมูลทั้งหมด
ที่เรียงลำดับจากน้อยไปมาก

เดไซล์ที่ r (D_r) คือ ค่าที่อยู่ตำแหน่งที่ r ส่วน ใน 10 ส่วน ของข้อมูลทั้งหมด
ที่เรียงลำดับจากน้อยไปมาก

เปอร์เซ็นไทล์ที่ r (P_r) คือ ค่าที่อยู่ตำแหน่งที่ r ส่วน ใน 100 ส่วน ของข้อมูลทั้งหมด
ที่เรียงลำดับจากน้อยไปมาก

Note

$$\text{มัธยฐาน} = Q_2 = D_5 = P_{50}, \quad Q_1 = D_{2.5} = P_{25}, \quad Q_3 = D_{7.5} = P_{75}$$

ข้อมูลไม่แจกแจงความถี่

ขั้นตอนการหาค่า

- 1) เรียงข้อมูลทั้งหมดจากน้อยไปมาก
- 2) หาดำแหน่งได้จากสูตร

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

$$\text{ตำแหน่ง } D_r = \frac{r(N+1)}{10}$$

$$\text{ตำแหน่ง } P_r = \frac{r(N+1)}{100}$$

- 3) หาค่า

ค่าของข้อมูลที่เรียงจากน้อยไปมากที่มีตำแหน่งตรงกับตำแหน่งที่เราต้องการหา
ค่า นั่นคือค่าของตำแหน่ง

ข้อมูลแจกแจงความถี่ (ตาราง)

ขั้นตอนการหาค่า

- 1) หาความถี่สะสมของแต่ละชั้น
- 2) หาค่าแห่งได้จากสูตร

$$\text{ตำแหน่ง } Q_r = \frac{rN}{4} \leftarrow \text{ตำแหน่ง}$$

$$\text{ตำแหน่ง } D_r = \frac{rN}{10} \leftarrow \text{ตำแหน่ง}$$

$$\text{ตำแหน่ง } P_r = \frac{rN}{100} \leftarrow \text{ตำแหน่ง}$$

- 3) หาชั้นที่ตำแหน่งอยู่

ถ้าตำแหน่งมีค่าอยู่ภายในความถี่สะสมของชั้นใด ตำแหน่งจะอยู่ในชั้นนั้น

- 4) หาค่าของตำแหน่ง (ที่อยู่ในรูปตาราง) จากสูตร

$$\text{ค่าของตำแหน่ง} = L + \left(\frac{T - \sum f_L}{f_T} \right) I$$

L คือ ขอบล่างของชั้นที่ตำแหน่งอยู่

T คือ ตำแหน่ง

$\sum f_L$ คือ ผลรวมของความถี่ของชั้นที่มีค่าต่ำกว่าชั้นที่ตำแหน่งอยู่

หรือ เท่ากับความถี่สะสมของชั้นก่อนหน้าของชั้นที่ตำแหน่งอยู่

f_T คือ ความถี่ของชั้นที่ตำแหน่งอยู่

I คือ ความกว้างของอันตรภาคชั้น

Note

ถ้าตำแหน่งไปตรงกับความถี่สะสมของชั้นใด

ค่าของตำแหน่งจะมีค่าเท่ากับขอบบนของชั้นนั้น

การวัดการกระจายของข้อมูล

การวัดการกระจายสัมบูรณ์

พิสัย (Range)

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

ส่วนเบี่ยงเบนควอร์ไทล์ $Q.D.$

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

ส่วนเบี่ยงเบนเฉลี่ย $M.D.$

ข้อมูลไม่แจกแจงความถี่

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

\bar{x} คือ ค่าเฉลี่ยเลขคณิต

x_i คือ ค่าของข้อมูลแต่ละตัว

N คือ จำนวนของข้อมูล

ข้อมูลแจกแจงความถี่ (ตาราง)

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

\bar{x} คือ ค่าเฉลี่ยเลขคณิต

f_i คือ ความถี่ของชั้นที่ i

x_i คือ จุดกึ่งกลางชั้นของชั้นที่ i

$\sum_{i=1}^k f_i$ หรือ N คือ จำนวนของข้อมูลทั้งหมด

k คือ จำนวนชั้นของข้อมูล

ส่วนเบี่ยงเบนมาตรฐาน $S.D.$, s , σ

ข้อมูลไม่แจกแจงความถี่

$$s = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N}} \quad \text{หรือ} \quad s = \sqrt{\frac{\sum_{i=1}^N x_i^2}{N} - (\bar{x})^2}$$

\bar{x} คือ ค่าเฉลี่ยเลขคณิต

x_i คือ ค่าของข้อมูลแต่ละตัว

N คือ จำนวนของข้อมูล

ข้อมูลแจกแจงความถี่ (ตาราง)

$$s = \sqrt{\frac{\sum_{i=1}^k f_i (x_i - \bar{x})^2}{N}} \quad \text{หรือ} \quad s = \sqrt{\frac{\sum_{i=1}^k f_i x_i^2}{N} - (\bar{x})^2}$$

\bar{x} คือ ค่าเฉลี่ยเลขคณิต

f_i คือ ความถี่ของชั้นที่ i

x_i คือ จุดกึ่งกลางชั้นของชั้นที่ i

N หรือ $\sum_{i=1}^k f_i$ คือ จำนวนของข้อมูลทั้งหมด

k คือ จำนวนชั้นของข้อมูล

สมบัติของส่วนเบี่ยงเบนมาตรฐาน และ ค่าความแปรปรวน s^2

- 1) s ไม่มีค่าเป็นลบ
- 2) ถ้า $s = 0$ แสดงว่าข้อมูลทุกตัวมีค่าเท่ากัน

- 3) $\sqrt{\frac{\sum_{i=1}^N (x_i - a)^2}{N}}$ มีค่าน้อยที่สุด เมื่อ $a = \bar{x}$

4) $y_i = ax_i + b$ จะได้ว่า $s_y = |a|s_x$ โดยที่ a, b เป็นค่าคงตัวใดๆ

5) s^2 คือค่าความแปรปรวน

สูตรเสริมพิเศษ (ใช้ได้เฉพาะกรณีที่มีค่าเฉลี่ยเลขคณิตของข้อมูลทุกกลุ่มมีค่าเท่ากัน)

ค่าความแปรปรวนรวม

$$s^2 = \frac{\sum_{i=1}^k N_i (s_i^2)}{\sum_{i=1}^k N_i}$$

s_i^2 คือ ค่าความแปรปรวนของแต่ละกลุ่ม

N_i คือ จำนวนข้อมูลของแต่ละกลุ่ม

k คือ จำนวนกลุ่ม

การวัดการกระจายสัมพัทธ์

สัมประสิทธิ์ของพิสัย

$$\text{สัมประสิทธิ์ของพิสัย} = \frac{x_{\max} - x_{\min}}{x_{\max} + x_{\min}}$$

สัมประสิทธิ์ของส่วนเบี่ยงเบนควอร์ไทล์

$$\text{สัมประสิทธิ์ของส่วนเบี่ยงเบนควอร์ไทล์} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

สัมประสิทธิ์ของส่วนเบี่ยงเบนเฉลี่ย

$$\text{สัมประสิทธิ์ของส่วนเบี่ยงเบนเฉลี่ย} = \frac{M.D.}{x}$$

สัมประสิทธิ์ของการแปรผัน

$$\text{สัมประสิทธิ์ของการแปรผัน} = \frac{s}{x} \text{ (ถ้าจะทำเป็นเปอร์เซ็นต์ก็คูณด้วย 100)}$$

สมบัติของ Σ

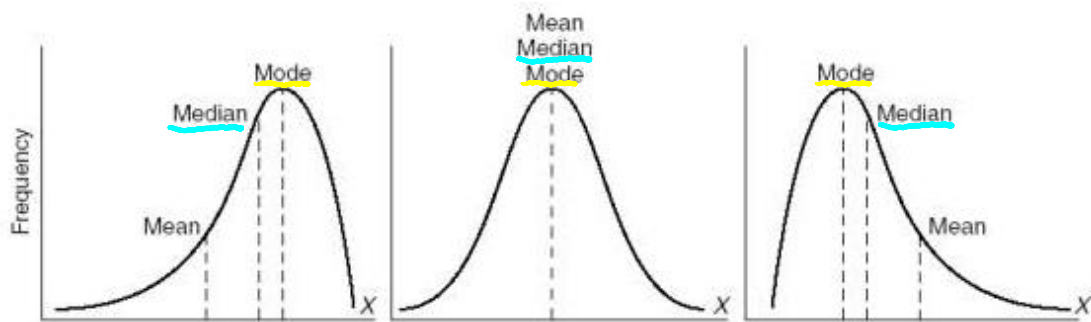
ให้ c เป็นค่าคงตัว

$$1) \sum_{i=1}^N c = Nc$$

$$2) \sum_{i=1}^N c x_i = c \sum_{i=1}^N x_i$$

$$3) \sum_{i=1}^N (x_i \pm y_i) = \sum_{i=1}^N x_i \pm \sum_{i=1}^N y_i$$

ความสัมพันธ์ระหว่างการแจกแจงความถี่, ค่ากลาง, การกระจาย



เบ้ซ้าย (ซ้ายหาง)

แจกแจงปกติ

เบ้ขวา (ขวาหาง)

Note

ถ้าแจกแจงปกติ จะได้ว่า Mean = Median = Mode

Median อยู่ระหว่าง Mean กับ Mode เสมอ

Mode อยู่ที่ตำแหน่งความถี่สูงสุด (โต่งสุด) เสมอ