บทที่ 3
การแบ่งส่วนของชุดข้อมูลและการวัดการกระจาย

1. ความหมายของการแบ่งข้อมูลออกเป็นส่วนต่าง ๆ

เมื่อนำชุดข้อมูลมาเรียงลำดับตามค่าของข้อมูลแล้วแบ่งข้อมูลที่เรียงแล้วเป็นส่วนต่าง ๆและระบุค่าของข้อมูล ณ จุดที่แบ่งนั้นจะพบว่า จำนวนค่าของข้อมูลบนจุดแบ่ง มีน้อยกว่าจำนวนส่วนที่ถูกแบ่ง เช่น เมื่อนำข้อมูลมาเรียงลำดับ แล้วแบ่งข้อมูลออกเป็น 2 ส่วนเท่า ๆกัน ค่าของข้อมูล ณ ตำแหน่งที่แบ่งมีเพียงค่าเดียว คือ ค่า ณ ตำแหน่งตรงกลาง ซึ่งก็คือ ค่าของมัธยฐาน

เมื่อนำชุดของข้อมูลชุดหนึ่งมาเรียงค่าจากน้อยไปมาก มาแบ่งข้อมูลออกเป็น 4 ส่วนเท่า ๆกันจะได้ค่า ณ จุดแบ่ง เพียง 3 ค่าโดยที่แต่ละส่วนมีจำนวนข้อมูลเท่า ๆกัน เรียกว่าควอไทล์ที่หนึ่ง (Q_1) ควอไทล์ที่สอง (Q_2) และควอไทล์ที่สาม (Q_3) ตามลำดับ

เมื่อนำชุดของข้อมูลชุดหนึ่งมาเรียงค่าจากน้อยไปมาก ค่าที่ตรงกับจุด 9 จุดที่แบ่งข้อมูลออกเป็น 10 ส่วน โดยที่แต่ละส่วนมีจำนวนข้อมูลเท่า ๆกันเรียกว่าเดไซล์ที่หนึ่ง (D_1) เดไซล์ที่สอง (D_2) เดไซล์ที่สาม (D_3) ... เดไซล์ที่เก้า (D_9) ตามลำดับ

ตัวอย่าง เดไซด์ที่ 6 บอกให้ทราบว่ามีจำนวนข้อมูลน้อยกว่าค่านี้อยู่ประมาณหกในสิบของจำนวนข้อมูลทั้งหมด

เมื่อนำชุดของข้อมูลชุดหนึ่งมาเรียงค่าจากน้อยไปมาก ค่าที่ตรงกับจุด 99 จุดที่แบ่งข้อมูลออกเป็น 100 ส่วน โดยที่แต่ละส่วนมีจำนวนข้อมูลเท่า ๆกันเรียกว่าเปอร์เซ็นไทล์ที่หนึ่ง (P_1) เปอร์เซ็นไทล์ที่สอง (P_2) เปอร์เซ็นไทล์ที่สาม (P_3) ... เปอร์เซ็นไทล์ที่เก้าสิบเก้า (P_{99}) ตามลำดับ

ตัวอย่าง เปอร์เซ็นไทล์ที่ 68 บอกให้ทราบว่ามีจำนวนข้อมูลน้อยกว่าค่านี้อยู่ประมาณหกสิบแปดในร้อยของจำนวนข้อมูลทั้งหมด

2. การหาควอไทล์ เดไซล์ และเปอร์เซ็นไทล์ของข้อมูลที่ไม่ได้แจกแจงความถี่

ในการหาต้องเรียงข้อมูลจากน้อยไปมาก แล้วหาค่าของคะแนนที่ตรงกับตำแหน่งของควอไทล์ เดไซล์ และเปอร์เซ็นไทล์ที่ต้องการ ถ้า N เป็นจำนวนข้อมูลทั้งหมดของประชากร (ถ้าเป็นกลุ่มตัวอย่างให้เปลี่ยนจาก N เป็น n และหาทำนองเดียวกัน) ตำแหน่งต่าง ๆหาได้ดังนี้

Q_1 อยู่ในตำแหน่งที่ $\frac{N+1}{4}$

Q_2 อยู่ในตำแหน่งที่ $\frac{2(N+1)}{4}$

Q_3 อยู่ในตำแหน่งที่ $\frac{3(N+1)}{4}$

นั่นคือ Q_r อยู่ในตำแหน่งที่ $\frac{r(N+1)}{4}$ เมื่อ r = 1,2,3

D_1 อยู่ในตำแหน่งที่ $\frac{N+1}{10}$

D_2 อยู่ในตำแหน่งที่ $\frac{2(N+1)}{10}$

D_3 อยู่ในตำแหน่งที่ $\frac{3(N+1)}{10}$

.

.

.

D_9 อยู่ในตำแหน่งที่ $\frac{9(N+1)}{10}$

นั่นคือ D_r อยู่ในตำแหน่งที่ $\frac{r(N+1)}{10}$ เมื่อ r = 1,2,3,…,9

P_1 อยู่ในตำแหน่งที่ $\frac{N+1}{100}$

P_2 อยู่ในตำแหน่งที่ $\frac{2(N+1)}{100}$

P_3 อยู่ในตำแหน่งที่ $\frac{3(N+1)}{100}$

.

.

.

นั่นคือ P_r อยู่ในตำแหน่งที่ $\frac{r(N+1)}{100}$ เมื่อ r = 1,2,3,...,99

เมื่อหาตำแหน่งที่ได้แล้วจึงหาค่าของข้อมูลที่ตรงกับตำแหน่งควอไทล์ เดไซล์ และเปอร์เซ็นไทล์ที่ต้องการ

หมายเหตุ ในทางสถิติไม่นิยมหาเปอร์เซ็นไทล์ของข้อมูลที่มีจำนวนน้อย เพราะค่าเปอร์เซ็นไทล์ที่คำนวณได้จะมีความถูกต้องน้อย

ตัวอย่างที่ 1 ในการสอบซ่อมนักเรียน 9 คนได้คะแนนดังนี้

8, 11, 6, 14, 15, 7, 18, 19, 16

จงหา Q_1 และ Q_3

วิธีทำ เรียงคะแนนจากน้อยไปมาก จะได้

6, 7, 8, 11, 14, 15, 16, 18, 19

Q_1 อยู่ในตำแหน่ง $\frac{9+1}{4} = 2.5$

เนื่องจาก Q_1 อยู่ในตำแหน่ง 2.5 ซึ่งเป็นจุดกึ่งกลางระหว่าง 7 และ 8

ดังนั้น $Q_1 = \frac{7+8}{2} = 7.5$

และ Q_3 อยู่ในตำแหน่ง $3\left(\frac{9+1}{4}\right) = 7.5$

เนื่องจาก Q_3 อยู่ในตำแหน่ง 7.5 ซึ่งเป็นค่ากึ่งกลางระหว่าง 16 และ 18

ดังนั้น $Q_3 = 17$

3. การหาควอไทล์ เดไซล์และเปอร์เซ็นไทล์ของข้อมูลที่แจกแจงความถี่

ในกรณีที่ข้อมูลที่จะนำมาหาควอไทล์ เดไซล์ และเปอร์เซ็นไทล์ เป็นข้อมูลที่แจกแจงความถี่ จะหาตำแหน่งที่ของควอไทล์ เดไซล์ และเปอร์เซ็นไทล์ได้จากสูตรต่อไปนี้

Q_1 อยู่ในตำแหน่งที่ $\frac{N}{4}$

Q_2 อยู่ในตำแหน่งที่ $\frac{2N}{4}$

Q_3 อยู่ในตำแหน่งที่ $\frac{3N}{4}$

นั่นคือ Q_r อยู่ในตำแหน่งที่ $\frac{rN}{4}$ เมื่อ $r = 1,2,3$

D_1 อยู่ในตำแหน่งที่ $\frac{N}{10}$

D_2 อยู่ในตำแหน่งที่ $\frac{2N}{10}$

D_3 อยู่ในตำแหน่งที่ $\frac{3N}{10}$

.

.

.

D_9 อยู่ในตำแหน่งที่ $\frac{9N}{10}$

นั่นคือ D_r อยู่ในตำแหน่งที่ $\frac{rN}{10}$ เมื่อ $r = 1, 2, 3, \ldots, 9$

P_1 อยู่ในตำแหน่งที่ $\frac{N}{100}$

P_2 อยู่ในตำแหน่งที่ $\frac{2N}{100}$

P_3 อยู่ในตำแหน่งที่ $\frac{3N}{100}$

.

.

.

นั่นคือ P_r อยู่ในตำแหน่งที่ $\frac{rN}{100}$ เมื่อ $r = 1, 2, 3, \ldots, 99$

ตัวอย่างที่ 2 ผลสอบวัดคณิตศาสตร์พื้นฐานนักเรียนชั้นมัธยมศึกษาปีที่ 6 จำนวน 60 คน นำคะแนนมาสร้างตารางแจกแจงความถี่ได้ดังนี้

คะแนน	ความถี่
26 - 30	4
31 - 35	16
36 - 40	11
41 - 45	12
46 - 50	8
51 - 55	6
56 - 60	3
	60

จงหา Q_1 D_5 และ P_{80}

วิธีทำ หาความถี่สะสม และตำแหน่งและค่าของ Q_1, D_5 และ P_{80} ตามลำดับ

คะแนน	ความถี่	ความถี่สะสม
26 - 30	4	4
31 - 35	16	20
36 - 40	11	31
41 - 45	12	43
46 - 50	8	51
51 - 55	6	57
56 - 60	3	60

หา Q_1

Q_1 อยู่ที่ตำแหน่ง $\frac{N}{4} = \frac{60}{4} = 15$

ดังนั้น Q_1 อยู่ในอันตรภาคชั้น 31 - 35 ซึ่งมีความถี่ 16

เทียบสัดส่วนเพื่อหา Q_1 ได้ดังนี้

ความถี่เพิ่มขึ้น 16 ในขณะที่คะแนนเพิ่มขึ้น 5 คะแนน

ความถี่เพิ่มขึ้น 15 - 4 = 11 ในขณะที่คะแนนเพิ่มขึ้น $\frac{5\times 11}{16} \approx 3.44$ คะแนน

ดังนั้น Q_1 ตรงกับคะแนน 30.5 + 3.44 = 33.94

หา D_5

D_5 อยู่ที่ตำแหน่ง $5\left(\frac{N}{10}\right) = 5\left(\frac{60}{10}\right) = 30$

ดังนั้น D_5 อยู่ในอันตรภาคชั้น 36-40 ซึ่งมีความถี่ 11

เทียบสัดส่วนเพื่อหา D_5 ได้ดังนี้

ความถี่เพิ่มขึ้น 11 ในขณะที่คะแนนเพิ่มขึ้น 5 คะแนน

ความถี่เพิ่มขึ้น 30 - 20 = 10 ในขณะที่คะแนนเพิ่มขึ้น $\frac{5\times 10}{11} \approx 4.55$ คะแนน

ดังนั้น D_5 ตรงกับคะแนน 35.5 + 4.55 = 40.05

หา P_{80}

P_{80} อยู่ที่ตำแหน่ง $80\left(\frac{N}{100}\right) = 80\left(\frac{60}{100}\right) = 48$

ดังนั้น P_{80} อยู่ในอันตรภาคชั้น 46-50 ซึ่งมีความถี่ 8

เทียบสัดส่วนเพื่อหา P_{80} ได้ดังนี้

ความถี่เพิ่มขึ้น 8 ในขณะที่คะแนนเพิ่มขึ้น 5 คะแนน

ความถี่เพิ่มขึ้น 48 - 43 = 5 ในขณะที่คะแนนเพิ่มขึ้น $\frac{5\times 5}{8} \approx 3.13$ คะแนน

ดังนั้น P_{80} ตรงกับคะแนน 45.5 + 3.13 = 48.63

ในกรณีที่ข้อมูลมีตารางแจกแจงความถี่จากอันตรภาคชั้นที่สูงกว่าไปยังอันตรภาคชั้นที่ต่ำกว่าก็หาควอไทล์ เดไซล์และเปอร์เซนไทล์ของตำแหน่ง r ต่าง ๆได้ทำนองเดียวกันโดยอาจจัดเรียงอันตรภาคชั้นใหม่จากอันตรภาคชั้นที่ต่ำกว่าไปยังอันตรภาคชั้นที่สูงกว่า หรือใช้ตารางเดิมแต่ หาความถี่สะสมจากอันตรภาคชั้นที่ต่ำกว่าไปยังอันตรภาคชั้นที่สูงกว่าแล้วเทียบสัดส่วนดังที่กล่าวแล้ว เช่นจากตารางในตัวอย่างที่ 2 หากเรียงอันตรภาคชั้นจากอันตรภาคชั้นที่สูงกว่าไปต่ำกว่า จะต้องเรียงคะแนนความถี่สะสมจากอันตรภาคชั้นที่ต่ำกว่าไปอันตรภาคชั้นที่สูงกว่าดังสดมภ์ที่สาม

คะแนน	ความถี่	ความถี่สะสม
56 - 60	3	60
51 - 55	6	57
46 - 50	8	51
41 - 45	12	43
36 - 40	11	31
31 - 35	16	20
26 - 30	4	4
N	60	

P_{80} อยู่ที่ตำแหน่ง $80\left(\frac{N}{100}\right) = 80\left(\frac{60}{100}\right) = 48$

ดังนั้น P_{80} อยู่ในอันตรภาคชั้น 46-50 ซึ่งมีความถี่ 8

เทียบสัดส่วนเพื่อหา P_{80} ได้ดังนี้

ความถี่เพิ่มขึ้น 8 ในขณะที่คะแนนเพิ่มขึ้น 5 คะแนน

ความถี่เพิ่มขึ้น 48 - 43 = 5 ในขณะที่คะแนนเพิ่มขึ้น $\frac{5 \times 5}{8} \approx 3.13$ คะแนน

ดังนั้น P_{80} ตรงกับคะแนน 45.5 + 3.13 = 48.63

หากวิเคราะห์จากขั้นตอนการคิดคำนวณจะพบว่า สูตรหา Q_r, D_r หรือ P_r เป็นดังนี้

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

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

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

เมื่อ L คือขอบล่างของอันตรภาคชั้นที่ Q_r, D_r หรือ P_r ตกอยู่

I คือความกว้างของอัตรภาคชั้นที่ Q_r, D_r หรือ P_r ตกอยู่

$\frac{rN}{4}, \frac{rN}{10}$ และ $\frac{rN}{100}$ เป็นตำแหน่งของ Q_r, D_r และ P_r ตามลำดับ

$\sum f_L$ คือผลรวมของความถี่อันตรภาคชั้นที่ต่ำกว่าอันตรภาคชั้นที่ Q_r, D_r หรือ P_r ตกอยู่

f คือความถี่ในอันตรภาคชั้นที่ Q_r, D_r หรือ P_r ตกอยู่

การหาควอไทล์ เดไซล์ และเปอร์เซ็นไทล์จากกราฟ

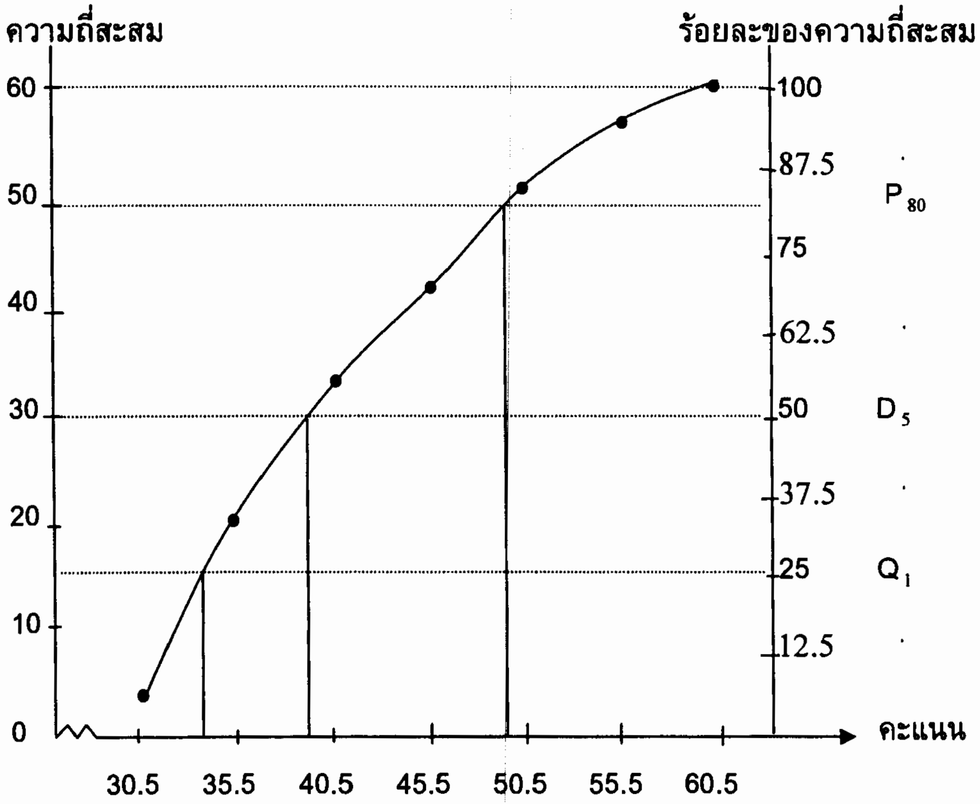
การหาควอไทล์ เดไซล์ และเปอร์เซ็นไทล์จากกราฟทำได้โดยเขียนกราฟของข้อมูลที่แจกแจงความถี่โดยเขียนเส้นโค้งของความถี่สะสมที่มีแกนตั้งแทนความถี่สะสมและแกนนอนแทนข้อมูล เช่น ถ้าต้องการหาค่า P_r ในขั้นแรกต้องหาจุดบนแกนตั้งที่แทน $\frac{rN}{100}$ แล้วลากส่วนของเส้นตรงขนานกับแกนนอน ให้ตัดเส้นโค้งของความถี่สะสมจากจุดตัดบนโค้งของความถี่สะสมให้ลากเส้นขนานกับแกนตั้งไปตัดแกนนอนที่ใด ค่าของ x บนแกนนอนนั้นจะเป็นค่า P_r ตามต้องการ ในกรณีของการหาค่า D_r หรือค่า Q_r จะต้องหา $\frac{rN}{10}$ หรือ $\frac{rN}{4}$ ตามแต่กรณีและสร้างโค้งความถี่และดำเนินการทำนองเดียวกับการหาค่า P_r

ในการหาตำแหน่งเปอร์เซ็นไทล์ ตำแหน่งเดไซล์หรือตำแหน่งควอไทล์ ของค่าใด ๆก็ใช้วิธีการกลับกันกล่าวคือจากค่าที่ต้องการหาตำแหน่งเปอร์เซ็นไทล์ ตำแหน่งเดไซน์หรือตำแหน่งควอไทล์ บนแกนนอนให้ลากเส้นขนานกับแกนตั้งไปตัดกราฟของความถี่สะสมแล้วลากจากจุดตัดบนกราฟของความถี่สะสมขนานกับแกนนอนไปตัดแกนตั้งที่จุดใดจุดนั้นคือค่าของ $\frac{rN}{100}, \frac{rN}{10}$ หรือ $\frac{rN}{4}$ ตามต้องการ แล้วนำไปคำนวณหาค่า r โดยแทนค่า N เพื่อให้สะดวกในการอ่านค่า เราควรหาร้อยละของความถี่สะสมสัมพัทธ์ดังตัวอย่าง

ตัวอย่างที่ 3 จากโจทย์ตัวอย่างที่สอง จงหา Q_1, D_5 และ P_{80} จากกราฟ

วิธีทำ หาความถี่สะสมและร้อยละของความถี่สะสมได้ดังนี้

คะแนน	ความถี่	ความถี่สะสม	ร้อยละของความถี่สะสมสัมพัทธ์
26-30	4	4	6.67
31-35	16	20	33.33
36-40	11	31	51.66
41-45	12	43	71.66
46-50	8	51	85
51-55	6	57	95
56-60	3	60	100



รูปเส้นโค้งของความถี่สะสม

เนื่องจาก Q_1 จะอยู่ในตำแหน่งเดียวกับ P_{25} ดังนั้นจากร้อยละของความถี่สะสมที่ 25% ลากเส้นขนานกับแกน X ไปตัดกราฟ และที่จุดตัดของกราฟลากเส้นขนานกับแกน Y ตัดแกน Y อ่านค่าที่แกน Y จะได้คะประมาณ 34

นั่นคือ Q_1 มีค่าประมาณ 34

ทำนองเดียวกัน

อ่านค่า D_5 จากกราฟได้ประมาณ 40

อ่านค่า P_{80} จากกราฟได้ประมาณ 48.5

หมายเหตุ

1. ค่าที่ได้จากการใช้กราฟจะมีความคลาดเคลื่อนไปบ้าง
2. ค่าวัดตำแหน่งของควอไทล์ เดไซล์ และเปอร์เซ็นไทล์ที่คำนวณได้เป็นค่าโดยประมาณ

ชวนศึกษา

มีตำราบางเล่ม เช่น Statistical Applications for the Behavioral Sciences ของ Laurence G. Grimm ได้เสนอวิธีหาค่าลำดับชั้นเปอร์เซ็นไทล์ของคะแนน (percentile Rank of a Score) สำหรับข้อมูลที่ไม่ได้แจกแจงความถี่โดยใช้สูตรดังนี้

สูตรสำหรับแปลงคะแนนเป็นตำแหน่งเปอร์เซ็นไทล์

สูตรสำหรับหาตำแหน่งเปอร์เซ็นไทล์ของ X

$$\text{เปอร์เซ็นไทล์ของ X} = \left(\frac{\mathrm{B} + \frac{1}{2}E}{N}\right)100 \qquad \text{..........(3.1)}$$

เมื่อ B คือจำนวนของคะแนนที่ต่ำกว่าจำนวน X ที่กำหนดให้

E คือจำนวนคะแนนที่เท่ากับคะแนน X ถ้ามีเพียงX ตัวเดียว แล้ว E = 1

N คือจำนวนทั้งหมดของคะแนนในการแจกแจง

ตัวอย่าง5 สำหรับการแจงแจงต่อไปนี้ จงหาลำดับเปอร์เซ็นไทล์ของคะแนน 14, 10,11,11,12,14,16,20

วิธีทำ N = 7, B = 4 , E = 1 ใช้สูตร 3.1 จะได้

$$\text{ลำดับเปอร์เซ็นไทล์ของคะแนน 14} = \left(\frac{4 + \frac{1}{2}(1)}{7}\right)100$$

$$= 64$$

ลำดับเปอร์เซ็นไทล์ของคะแนน 14 = 64 %

ดังนั้น 64%ของคะแนนของการแจกแจงนี้อยู่ต่ำกว่าคะแนน 16 คะแนน 16 เป็นเปอร์เซ็นไทล์ที่ 64

แต่สำหรับการใช้สูตรนี้หาลำดับชั้นเปอร์เซ็นไทล์ของข้อมูลจากตารางแจกแจงความถี่ในหนังสือดังกล่าวดูจะมีความคลาดเคลื่อนอยู่ จึงไม่นำมากล่าวในที่นี้ นักศึกษาอาจพิจารณาหาโดยเทียบกับสูตรที่กล่าวมาแล้วก่อนหน้านี้

4. การวัดการกระจายของข้อมูล (measures of dispersion)

แบ่งออกเป็น 2 วิธี คือ

การกระจายสัมบูรณ์ (absolute variation) คือ การวัดการกระจายของข้อมูลเพียงชุดเดียว เพื่อดูว่าข้อมูลชุดนั้นแต่ละค่ามีความแตกต่างกันมากหรือน้อยเพียงใด การวัดการกระจายสัมบูรณ์ที่นิยมใช้มี 4 ชนิดคือ

(1) พิสัย (range)

(2) ส่วนเบี่ยงเบนควอไทล์ หรือกึ่งช่วงควอไทล์(quartile deviation หรือ semi-interquatile range)

(3) ส่วนเบี่ยงเบนเฉลี่ย (mean deviation หรือ average deviation)

(4) ส่วนเบี่ยงเบนมาตรฐาน (standard deviation)

การกระจายสัมพัทธ์(relative variation) คือการหาค่าเพื่อเปรียบเทียบการกระจายระหว่างข้อมูลมากกว่าหนึ่งชุด โดยใช้อัตราส่วน เช่นอัตราส่วนระหว่างค่าการกระจายสัมบูรณ์กับค่ากลางของข้อมูลชุดนั้นๆ การวัดการกระจายสัมพัทธ์ของข้อมูลแต่ละชุดเพื่อใช้เปรียบเทียบการกระจายของข้อมูลระหว่างชุดของข้อมูลเหล่านั้น การวัดการกระจายสัมพัทธ์มี 4 ชนิด

(1) สัมประสิทธิ์ของพิสัย (coefficient of range)

(2) สัมประสิทธิ์ของส่วนเบี่ยงเบนควอไทล์ (coefficient of quartile deviation)

(3) สัมประสิทธิ์ของส่วนเบี่ยงเบนเฉลี่ย (coefficient of average deviation)

(4) สัมประสิทธิ์ของความแปรผัน (coefficient of variation)

4.1 การวัดการกระจายสัมบูรณ์ (measures of absolute variation)

พิสัย(range)

พิสัยคือค่าผลต่างของข้อมูลที่มีค่าสูงสุดและข้อมูลที่มีค่าต่ำสุด ถ้า x_1, x_2, x_3, ..., x_N คือค่าของข้อมูลชุดหนึ่งพิสัยของข้อมูลชุดนี้เท่ากับ $x_{max} - x_{min}$

เมื่อ x_{max} เป็นค่าสูงสุดของข้อมูล และ x_{min} เป็นค่าต่ำสุดของข้อมูล

ตัวอย่างที่ 1 ผลการสอบวิชาภาษาไทยและภาษาอังกฤษของนักเรียน 10 คน ได้ผลคะแนนดังนี้

วิชาภาษาไทย : 3, 4, 5, 5, 6, 6, 6, 8, 8, 9

วิชาภาษาอังกฤษ : 3, 4, 4, 5, 5, 6, 6, 8, 8, 10

จากข้อมูลจะพบว่า

พิสัยของวิชาภาษาไทยเท่ากับ 9 - 3 = 6

พิสัยของวิชาภาษาอังกฤษเท่ากับ 10 - 3 = 7

กรณีนี้เราอาจสรุปโดยดูจากพิสัยว่า ผลคะแนนวิชาภาษาอังกฤษมีการกระจายมากกว่าผลคะแนนวิชาภาษาไทย

ข้อสังเกต ในการเปรียบเทียบการกระจายโดยใช้พิสัย คะแนนเต็มของสองกลุ่มจะต้องเท่ากัน จึงจะทำให้การเปรียบเทียบนี้พอจะเปรียบเทียบกันได้ แต่ก็เป็นการเปรียบเทียบเพียงคร่าว ๆ เมื่อใช้วิธีการเปรียบเทียบแบบอื่นอาจพบว่าผลการเปรียบเทียบต่างไปจากการเปรียบเทียบโดยพิสัยก็ได้ ดังนั้นการวัดการกระจายโดยใช้พิสัยจึงเป็นเพียงการพิจารณาในเบื้องต้นเท่านั้น

ส่วนเบี่ยงเบนควอไทล์ หรือกึ่งช่วงควอไทล์(quatile deviation หรือ semi-interquatile)

ส่วนเบี่ยงเบนควอไทล์ คือ ค่าที่ใช้วัดการกระจายที่หาได้จากครึ่งหนึ่งของผลต่างระหว่างค่าควอไทล์ที่สาม(Q_3) และค่าควอไทล์ที่หนึ่ง (Q_1) ใช้สัญลักษณ์ Q.D.แทนส่วนเบี่ยงเบนควอไทล์

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

ตัวอย่างที่ 2 จงหาส่วนเบี่ยงเบนควอไทล์ของข้อมูลต่อไปนี้

12, 23, 25, 18, 7, 16, 9, 10, 15, 24, 29, 30, 6

วิธีทำ ข้อมูลมีทั้งหมด 13 ค่า เรียงคะแนนจากน้อยไปมากจะได้

6, 7, 9, 10, 12, 15, 16,18, 23, 24, 25, 29, 30

หา Q_1 และ Q_3 ดังนี้

Q_1 อยู่ในตำแหน่ง $\frac{13+1}{4}$ = 3.5 ดังนั้น Q_1 = 9.5

Q_3 อยู่ในตำแหน่ง $3\left(\frac{13+1}{4}\right)$ = 10.5 ดังนั้น Q_3 = 24.5

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

$$= \frac{24.5 - 9.5}{2}$$

$$= \frac{15}{2}$$

$$= 7.5$$

ตัวอย่างที่ 3 จงหาส่วนเบี่ยงเบนควอไทล์ของข้อมูลต่อไปนี้

คะแนน	**ความถี่**
26 - 30	6
31 - 35	13
36 - 40	24
41 - 45	35
46 - 50	22
51 - 55	16
56 - 60	4
รวม	120

วิธีทำ หาความถี่สะสมดังนี้

คะแนน	**ความถี่**	**ความถี่สะสม**
26 - 30	6	6
31 - 35	13	19
36 - 40	24	43
41 - 45	35	78
46 - 50	22	100
51 - 55	16	116
56 - 60	4	120

หา Q_1

Q_1 อยู่ในตำแหน่ง $\frac{120}{4} = 30$

Q_1 ตกอยู่ในอันตรภาคชั้น 36-40

ความถี่เพิ่มขึ้น 24 เมื่อคะแนนเพิ่มขึ้น 5

ความถี่เพิ่มขึ้น 30 - 19 = 11 เมื่อคะแนนเพิ่มขึ้น $\frac{5 \times 11}{24} \approx 2.29$

ดังนั้น Q_1 = 35.5 + 2.29 = 37.79

หา Q_3

Q_3 อยู่ในตำแหน่ง $3\left(\frac{120}{4}\right)$ = 90

Q_3 ตกอยู่ในอันตรภาคชั้น 46-50

ความถี่เพิ่มขึ้น 22 เมื่อคะแนนเพิ่มขึ้น 5

ความถี่เพิ่มขึ้น 90 - 78 = 12 เมื่อคะแนนเพิ่มขึ้น $\frac{5\times 12}{22}$ = 2.73

ดังนั้น Q_3 = 45.5 + 2.73 = 48.23

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

$$= \frac{48.23 - 37.79}{2}$$

$$= \frac{10.44}{2}$$

$$= 5.22$$

ส่วนเบี่ยงเบนเฉลี่ย (mean deviation หรือ average deviation)

ส่วนเบี่ยงเบนเฉลี่ยคือค่าที่ได้จากการเฉลี่ยค่าสัมบูรณ์ของความแตกต่างระหว่างค่าของข้อมูลแต่ละค่าและค่ากลางของข้อมูลชุดนั้น ค่ากลางที่ใช้อาจเป็นค่ามัชฌิมเลขคณิตหรือมัธยฐานก็ได้ แต่ส่วนมากนิยมใช้ค่ามัชฌิมเลขคณิต

การหาส่วนเบี่ยงเบนเฉลี่ย

ถ้า $x_1, x_2, x_3, \ldots, x_n$ เป็นข้อมูลของตัวอย่าง n จำนวน มีค่ามัชฌิมเลขคณิตเป็น $\overline{X}$ เมื่อใช้สัญลักษณ์ M.D. แทนส่วนเบี่ยงเบนเฉลี่ยแล้ว

$$M.D. = \frac{/x_1 - \overline{X}/ + /x_2 - \overline{X}/ + /x_3 - \overline{X}/ + \ldots + /x_n - \overline{X}/}{n}$$

$$= \frac{\sum_{i=1}^{n} /x_i - \overline{X}/}{n}$$

ตัวอย่างที่ 4 จงหาส่วนเบี่ยงเบนเฉลี่ยของการแจกแจง A และการแจกแจง B

การแจกแจง A : 11, 12, 13, 14, 15, 16, 17

การแจกแจง B : 5, 8, 11, 14, 17, 20, 23

วิธีทำ มัชฌิมเลขคณิตของการแจกแจง A คือ

$$\overline{X} = \frac{11+12+13+14+15+16+17}{7} = \frac{98}{7} = 14$$

ดังนั้นส่วนเบี่ยงเบนเฉลี่ยของการแจกแจง A คือ

$$\text{M.D.} = \frac{/11-14/+/12-14/+/13-14/+/14-14/+/15-14/+/16-14/+/17-14/}{7}$$

$$= \frac{3+2+1+0+1+2+3}{7}$$

$$= \frac{12}{7}$$

$$\approx 1.71$$

วิธีทำ หาค่ามัชฌิมเลขคณิตของการแจกแจง B คือ

$$\overline{X} = \frac{5+8+11+14+17+20+23}{7} = \frac{98}{7} = 14$$

ดังนั้นส่วนเบี่ยงเบนเฉลี่ยของการแจกแจง B คือ

$$\text{M.D.} = \frac{/5-14/+/8-14/+/11-14/+/14-14/+/17-14/+/20-14/+/23-14/}{7}$$

$$= \frac{9+6+3+0+3+6+9}{7}$$

$$= \frac{36}{7} \approx 5.14$$

ดังนั้นการแจกแจง A มีการกระจายน้อยกว่าการแจกแจง B

การหาส่วนเบี่ยงเบนเฉลี่ยของข้อมูลที่แจกแจงความถี่แล้ว

ถ้า $x_1, x_2, x_3, \ldots, x_k$ เป็นจุดกึ่งกลางของอันตรภาคชั้นต่างๆ k อันตรภาคชั้น ซึ่งมีความถี่ $f_1, f_2, f_3, \ldots, f_k$ ตามลำดับ n เป็นจำนวนข้อมูลของตัวอย่างทั้งหมด และมีค่ามัชฌิมเลขคณิตเป็น $\overline{X}$ แล้ว

$$\text{M.D.} = \frac{f_1 / x_1 - \overline{X} / + f_2 / x_2 - \overline{X} / + f_3 / x_3 - \overline{X} / + \ldots + f_n / x_n - \overline{X} /}{n}$$

$$= \frac{\sum_{i=1}^{n} f_i / x_i - \overline{X} /}{n}$$

เมื่อ k แทนจำนวนอันตรภาคชั้น

f_i แทนความถี่ของอันตรภาคชั้นที่ i

X_i แทนจุดกึ่งกลางของอันตรภาคชั้นที่ i

หมายเหตุ การหาส่วนเบี่ยงเบนเฉลี่ยของประชากรก็หาได้เช่นเดียวกันโดยการหาจากประชากรทั้งหมด ซึ่งจำนวนของข้อมูลจะเปลี่ยนจาก n เป็น N

ตัวอย่างที่ 5 จงเปรียบเทียบส่วนเบี่ยงเบนเฉลี่ยของการกระจายสองชุด

การกระจาย A

คะแนน	ความถี่
41-45	6
46-50	8
51-55	18
56-60	20
61-65	15

การกระจาย B

คะแนน	ความถี่
36-40	3
41-45	5
46-50	11
51-55	14
56-60	22

66-70	5		61-65	13
71-75	2		66-70	6
รวม	74		รวม	74

วิธีทำ หาค่ามัชฌิมเลขคณิต และหา M.D. จากตาราง A ดังนี้

คะแนน	จุดกึ่งกลาง x_i	ความถี่ f_i	ความถี่คูณจุดกึ่งกลาง $f_i x_i$	$/x_i - \bar{X}/$	$f_i /x_i - \bar{X}/$
41-45	43	6	258	13.58	81.48
46-50	48	8	384	8.58	68.64
51-55	53	18	954	3.58	64.44
56-60	58	20	1160	1.42	28.4
61-65	63	15	945	6.42	96.3
66-70	68	5	340	11.42	57.1
71-75	73	2	146	16.42	32.84
		74	4187		429.2

$$\bar{X} = \frac{4187}{74} = 56.58$$

$$\text{M.D.} = \frac{429}{74}$$

$$\approx 5.80$$

หาค่ามัชฌิมเลขคณิต และหา M.D. จากตาราง B ดังนี้

คะแนน	จุดกึ่งกลาง x_i	ความถี่ f_i	ความถี่คูณจุดกึ่งกลาง $f_i x_i$	$/x_i - \bar{X}/$	$f_i /x_i - \bar{X}/$
36-40	38	3	114	15.99	47.97
41-45	43	5	215	10.99	54.95
46-50	48	11	528	5.99	65.89

51-55	53	14	742	0.99	13.86
56-60	58	22	1276	4.01	88.22
61-65	63	13	819	9.01	117.13
66-70	68	6	408	14.01	84.06
		74	4102		472.08

$$\overline{X} = \frac{4102}{74} = 53.99$$

$$\text{M.D.} = \frac{472.08}{74}$$

$$\approx 6.38$$

จากส่วนเบี่ยงเบนเฉลี่ยที่คำนวณได้แสดงว่ากลุ่ม B ข้อมูลมีการกระจายออกจากค่าเฉลี่ยมากกว่ากลุ่ม A

ส่วนเบี่ยงเบนมาตรฐาน (standard deviation)

ส่วนเบี่ยงเบนมาตรฐานของประชากรคือรากที่สองที่เป็นบวกของค่าเฉลี่ยของผลบวกของกำลังสองของผลต่างของข้อมูลแต่ละจำนวนและค่ามัชฌิมเลขคณิต นั่นคือ ถ้า $x_1, x_2, x_3, \ldots, x_N$ เป็นข้อมูลของประชากร N จำนวน มีค่ามัชฌิมเลขคณิตเป็น μ

ถ้าใช้สัญลักษณ์ σ (อ่านว่า sigma) แทนส่วนเบี่ยงเบนมาตรฐานของประชากร จะได้ว่า

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}}$$

จากสูตรส่วนเบี่ยงเบนมาตรฐานข้างต้นนี้ในทางคณิตศาสตร์สามารถจัดรูปใหม่ได้เป็น

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} x_i^2}{N} - \mu^2}$$

ในการหาส่วนเบี่ยงเบนมาตรฐานจากกลุ่มตัวอย่าง เราจะใช้สัญลักษณ์ s หรือ S.D. แทนส่วนเบี่ยงเบนมาตรฐานของกลุ่มตัวอย่าง นั่นคือ

ถ้า x_1, x_2, x_3, ..., x_n เป็นข้อมูลของกลุ่มตัวอย่าง n จำนวน มีค่ามัชฌิมเลขคณิตเป็น $\bar{X}$ แล้วส่วนเบี่ยงเบนมาตรฐานของกลุ่มตัวอย่างคือ

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n-1}}$$

จากสูตรส่วนเบี่ยงเบนมาตรฐานข้างต้นนี้ในทางคณิตศาสตร์สามารถจัดรูปใหม่ได้เป็น

$$s = \sqrt{\frac{\sum_{i=1}^{n}x_i^2 - n\bar{X}}{n-1}}$$

ตัวอย่างที่ 1 ผลการสอบซ่อมวิชาคณิตศาสตร์ของนักเรียนสองห้อง ปรากฏผลซึ่งเรียงคะแนนจากน้อยไปมากได้ดังนี้

ห้อง 4/1 : 7, 7, 8, 8, 11,12, 14,14, 15, 15, 15,16

ห้อง 4/2 : 6, 7, 8, 8, 8,12, 14, 14, 14, 14, 15, 17, 17

จงหาส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบซ่อมของนักเรียนแต่ละห้อง

แล้วพิจารณาว่าคะแนนสอบซ่อมของนักเรียนห้องใดมีการกระจายมากกว่าคะแนนของอีกห้องหนึ่ง

ในการคิดคำนวณนี้เราจะใช้สูตร $\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}}$

เนื่องจากเป็นการเปรียบเทียบผลการสอบซ่อมของประชากรสองกลุ่ม

วิธีทำ หาค่ามัชฌิมเลขคณิตและส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบซ่อมห้อง 4/1 ได้ดังนี้

$$\mu = \frac{7+7+8+8+11+12+14+14+15+15+15+16}{12}$$

$$= \frac{142}{12}$$

$$\approx 11.83$$

$$\sum_{i=1}^{12}(x-\mu)^2 = 2(4.83)^2 + 2(3.83)^2 + (0.83)^2 + (0.17)^2 + 2(2.17)^2 + 3(3.17)^2 + (4.17)^2$$

$$\approx 46.66+29.34+0.69+0.03+9.42+30.15+17.39$$

$$= 133.68$$

$$\sigma = \sqrt{\frac{133.68}{12}}$$

$$= \sqrt{11.14}$$

$$\approx 3.34$$

หาค่ามัชฌิมเลขคณิตและส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบซ่อมห้อง 4/2 ได้ดังนี้

$$\mu = \frac{6+7+8+8+8+12+14+14+14+14+15+17+17}{13}$$

$$= \frac{154}{13}$$

$$\approx 11.85$$

$$\sum_{i=1}^{13}(x-\mu)^2 = (5.85)^2 + (4.85)^2 + 3(3.85)^2 + (0.15)^2 + 4(2.15)^2 + (3.15)^2 + 2(5.15)^2$$

$$\approx 34.22 + 114.08 + 44.47 + 0.02 + 18.49 + 9.92 + 53.05$$

$$= 274.25$$

$$\sigma = \sqrt{\frac{274.25}{13}}$$

$$\approx \sqrt{21.10}$$

$$\approx 4.59$$

เนื่องจากส่วนเบี่ยงเบนมาตรฐานของห้อง 4/2 มากกว่าห้อง 4/1 ดังนั้นคะแนนของห้อง 4/2 มีการกระจายมากกว่าห้อง 4/1

การหาส่วนเบี่ยงเบนมาตรฐานของข้อมูลประชากรที่แจกแจกความถี่แล้ว

ถ้า $x_1, x_2, x_3, \ldots, x_k$ เป็นค่ากึ่งกลางของข้อมูลของประชากรในอันตรภาคชั้นต่าง ๆ k อันตรภาคชั้น ที่มีความถี่ $f_1, f_2, f_3, \ldots, f_k$ ตามลำดับ โดยที่ $f_1 + f_2 + f_3 + \ldots + f_k = N$ และข้อมูลประชากรมีค่ามัชฌิมเลขคณิตเป็น μ แล้วส่วนเบี่ยงเบนมาตรฐานของข้อมูลประชากรคือ

$$\sigma = \sqrt{\frac{\sum_{i=1}^{k} f_i (x_i - \mu)^2}{\sum_{i=1}^{k} f_i}}$$

หรือ

$$\sigma = \sqrt{\frac{\sum_{i=1}^{k} f_i x_i^2 - N\mu}{N}}$$

สูตรส่วนเบี่ยงเบนมาตรฐานของกลุ่มตัวอย่างเป็นดังนี้

ถ้า $x_1, x_2, x_3, \ldots, x_k$ เป็นค่ากึ่งกลางของข้อมูลของกลุ่มตัวอย่างในอันตรภาคชั้นต่าง ๆ k อันตรภาคชั้น ที่มีความถี่ $f_1, f_2, f_3, \ldots, f_k$ ตามลำดับ โดยที่ $f_1 + f_2 + f_3 + \ldots + f_k = n$ และข้อมูลตัวอย่างมีค่ามัชฌิมเลขคณิตเป็น $\overline{X}$ แล้วส่วนเบี่ยงเบนมาตรฐานของข้อมูลตัวอย่างคือ

$$S = \sqrt{\frac{\sum_{i=1}^{k} f_i (x_i - \overline{X})^2}{n-1}}$$

จากสูตรส่วนเบี่ยงเบนมาตรฐานข้างต้นนี้สามารถจัดรูปใหม่ได้เป็น

$$S = \sqrt{\frac{\sum_{i=1}^{k} f_i x_i^2 - n\overline{X}}{n-1}}$$

ตัวอย่างที่ 2 จากการศึกษารายได้ของผู้ปกครองนักเรียนชั้นมัธยมศึกษาปีที่ 3 ห้อง 1 และ ห้อง 2 มีหน่วยเป็นพันบาทปรากฏผลดังนี้

ห้อง 3/1

รายได้	จำนวน
20-24	2
25-29	4
30-34	8
35-39	12
40-44	7
45-49	4
50-54	2
55-59	1
รวม	40

ห้อง 3/2

รายได้	จำนวน
20-24	1
25-29	4
30-34	6
35-39	9
40-44	13
45-49	4
50-54	3
55-59	2
รวม	42

จงหาส่วนเบี่ยงเบนมาตรฐานของเงินเดือนผู้ปกครองของนักเรียนสองห้องนี้และพิจารณาว่าเงินเดือนของผู้ปกครองกลุ่มใดมีการกระจายมากกว่า

$$\sigma = \sqrt{\frac{\sum_{i=1}^{k} f_i (x_i - \mu)^2}{N}}$$

วิธีทำ หาค่ามัชฌิมเลขคณิตและส่วนเบี่ยงเบนมาตรฐานของรายได้ผู้ปกครองนักเรียนห้อง 3/1 จากตารางแจกแจงความถี่จะได้

รายได้	x_i	f_i	$f_i x_i$	$(x_i-\mu)$	$(x_i-\mu)^2$	$f_i(x_i-\mu)^2$
20-24	22	2	44	-15.4	237.16	474.32
25-29	27	4	108	-10.4	108.16	432.64
30-34	32	8	256	-5.4	29.16	233.28
35-39	37	12	444	-0.4	0.16	1.92
40-44	42	7	294	4.6	21.16	148.12
45-49	47	4	188	9.6	92.16	368.64
50-54	52	2	104	14.6	213.16	426.32
55-59	57	1	57	19.6	384.16	384.16
รวม		40	1495			2469.4

$$\mu = \frac{1495}{40}$$

$$\approx 37.4$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{k} f_i (x_i - \mu)^2}{N}}$$

$$= \sqrt{\frac{2469.4}{40}}$$

$$\approx 6.32$$

หาค่ามัชฌิมเลขคณิตและส่วนเบี่ยงเบนมาตรฐานของรายได้ผู้ปกครองนักเรียนห้อง 3/2 จากตารางแจกแจง ดังนี้

รายได้	x_i	f_i	$f_i x_i$	$(x_i-\mu)$	$(x_i-\mu)^2$	$f_i(x_i-\mu)^2$
20-24	22	1	22	-17.7	313.29	313.29
25-29	27	4	108	-12.7	161.29	615.16
30-34	32	6	192	-7.7	59.29	355.74
35-39	37	9	333	-2.7	7.29	65.61
40-44	42	13	546	2.3	5.29	68.77
45-49	47	4	188	7.3	53.29	213.16
50-54	52	3	165	12.3	151.29	453.87
55-59	57	2	114	17.3	299.29	598.58
รวม		42	1668			2684.18

$$\mu = \frac{1668}{42}$$

$$\approx 39.7$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{k} f_i(x_i - \mu)^2}{N}}$$

$$= \sqrt{\frac{2684.18}{42}}$$

$$\approx 7.99$$

จากผลการวิเคราะห์ข้อมูลจะพบว่า ค่าเฉลี่ยของรายได้ของผู้ปกครองนักเรียนห้อง 3/2 สูงกว่าห้อง 3/1 และส่วนเบี่ยงเบนมาตรฐานของรายได้ผู้ปกครองห้อง 3/2 สูงกว่าห้อง 3/1 ด้วย ซึ่งในส่วนของส่วนเบี่ยงเบนมาตรฐานบอกให้ทราบว่ารายได้ของผู้ปกครองนักเรียนของห้อง 3/2 กระจายมากกว่ารายได้ของผู้ปกครองห้อง 3/1

สมบัติของส่วนเบี่ยงเบนมาตรฐาน

- ส่วนเบี่ยงเบนมาตรฐานมีค่ามากกว่าหรือเท่ากับศูนย์เสมอ

- ถ้าคำนวณหาส่วนเบี่ยงเบนมาตรฐานโดยใช้ค่ากลางของข้อมูลชนิดอื่นๆ ที่ไม่ใช่ค่ามัชฌิมเลขคณิต ส่วนเบี่ยงเบนมาตรฐานที่ได้จะมีค่ามากกว่าส่วนเบี่ยงเบนมาตรฐานที่ใช้ค่ามัชฌิมเลขคณิตเสมอ นั่นคือ

$$\sqrt{\frac{\sum_{i=1}^{N}(x_i - A)^2}{N}} > \sqrt{\frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}}$$

อสมการนี้เป็นจริงเสมอเมื่อ A เป็นจำนวนจริงใดๆที่ไม่เท่ากับ μ

4.2. การวัดการกระจายสัมพัทธ์ (measures of relative variation)

ในหัวข้อนี้จะแสดงให้เห็นว่าข้อมูลที่มาจากกลุ่มตัวอย่างเดียวกันหรือประชากรเดียวกันแต่เปลี่ยนแปลงไปเนื่องจากข้อมูลชุดหนึ่งเป็น c เท่าของข้อมูลอีกชุดหนึ่ง แล้วการตีความการกระจายโดยวิธีการที่จะกล่าวต่อไปนี้จะเท่ากัน

การวัดการกระจายสัมพัทธ์ประกอบด้วย

(1) สัมประสิทธิ์ของพิสัย = $\dfrac{x_{\max} - x_{\min}}{x_{\max} + x_{\min}}$

(2) สัมประสิทธิ์ของส่วนเบี่ยงเบนควอไทล์ = $\dfrac{Q_3 - Q_1}{Q_3 + Q_1}$

(3) สัมประสิทธิ์ของส่วนเบี่ยงเบนเฉลี่ยของตัวอย่าง = $\dfrac{M.D.}{\bar{X}}$

(4) สัมประสิทธิ์ของส่วนเบี่ยงเบนเฉลี่ยของประชากร = $\dfrac{M.D.}{\mu}$

(5) สัมประสิทธิ์ของการแปรผันของตัวอย่าง = $\dfrac{s}{\bar{X}}$

(6) สัมประสิทธิ์ของการแปรผันของประชากร = $\dfrac{\sigma}{\mu}$

พิจารณาข้อมูลสองชุดต่อไปนี้

A: 2, 3, 5, 6, 7, 8

B: 20, 30, 50, 60, 70, 80

ถ้าเปรียบเทียบข้อมูลจากพิสัยจะเห็นว่าข้อมูล A มีพิสัย 8 − 2 = 6 ข้อมูล B มีพิสัยเท่ากับ 80 − 20 = 60 เราอาจสรุปว่าข้อมูล B มีการกระจายมากกว่าข้อมูล A เนื่องจากพิสัยของข้อมูล B สูงกว่าพิสัยของข้อมูล A

แต่ถ้ากล่าวว่า ข้อมูล B ได้จากข้อมูล A โดยการคูณข้อมูลแต่ละตัวด้วย 10 เช่น เมื่อให้คะแนนนักเรียนดังข้อมูล A โดยมีคะแนนเต็ม 10 แต่เมื่อครูเปลี่ยนคะแนนเต็มเป็น 100 นักเรียนจะได้คะแนนเป็นข้อมูล B ดังนั้นข้อมูลทั้งสองชุดจึงต้องมีค่าการกระจายที่เท่ากัน

ถ้าเราหาค่าการกระจายดังกล่าวด้วย สัมประสิทธิ์ของพิสัยเท่ากับ $\dfrac{x_{\max} - x_{\min}}{x_{\max} + x_{\min}}$ จะพบว่าข้อมูล

A มีการกระจาย $\dfrac{8-2}{8+2}$ = 0.6

B มีการกระจาย $\frac{80-20}{80+20}$ = 0.6

ข้อมูล A และข้อมูล B มีค่าการกระจายเท่ากัน

พิจารณาข้อมูลสองชุดต่อไปนี้

A: 2, 3, 5, 6, 7, 8

B: 20, 30, 50, 60, 70, 80

ข้อมูล A มี

ตำแหน่ง Q_1 คือ $\frac{6+1}{4}$ = 1.75 ดังนั้น Q_1 = 2.75

ตำแหน่ง Q_3 คือ $3(\frac{6+1}{4})$ = 5.25 ดังนั้น Q_3 = 7.25

ข้อมูล B มี

ตำแหน่ง Q_1 คือ $\frac{6+1}{4}$ = 1.75 ดังนั้น Q_1 = 27.5

ตำแหน่ง Q_3 คือ $3(\frac{6+1}{4})$ = 5.25 ดังนั้น Q_3 = 72.5

ถ้าเปรียบเทียบข้อมูลโดยใช้ส่วนเบี่ยงเบนควอไทล์

Q.D. $=\frac{Q_3-Q_1}{2}$ จะพบว่ามีค่าต่างกัน กล่าวคือส่วนเบี่ยงเบนควอไทล์ของข้อมูล A เท่ากับ 2.25 แต่ส่วนเบี่ยงเบนควอไทล์ของข้อมูล B เท่ากับ 22.5 ซึ่งมีค่าต่างกัน หากจะตีความการกระจายจากค่า Q.D. จะตีความว่าข้อมูล B กระจายมากกว่าข้อมูล A ทำให้การตีความผิดไป

แต่ถ้าดูการกระจายจากสัมประสิทธิ์ของส่วนเบี่ยงเบนควอไทล์จะได้ว่า

ข้อมูล A

สัมประสิทธิ์ของส่วนเบี่ยงเบนควอไทล์ $= \frac{Q_3-Q_1}{Q_3+Q_1}$

$= \frac{72.5-17.5}{72.5+17.5}$

= 0.61

ข้อมูล B

สัมประสิทธิ์ของส่วนเบี่ยงเบนควอไทล์ $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$

$= \frac{7.25 - 1.75}{7.25 + 1.75}$

$= 0.61$

จะเห็นได้ว่าข้อมูล A และข้อมูล B มีค่าสัมประสิทธิ์ของส่วนเบี่ยงเบนควอไทล์เท่ากัน นั่นคือข้อมูล A และข้อมูล B มีการกระจายเท่ากัน

พิจารณาข้อมูลสองชุดต่อไปนี้

A: 2, 3, 5, 7, 8

B: 20, 30, 50, 70, 80

ข้อมูล A มี $\overline{X}$ = 5

$$\text{M.D.} = \frac{/2-5/+/3-5/+/5-5/+/7-5/+/8-5/}{5}$$

$= 2$

ข้อมูล B มี $\overline{X}$ = 50

$$\text{M.D.} = \frac{/20-50/+/30-50/+/50-50/+/70-50/+/80-50/}{5}$$

$= 20$

ถ้าเปรียบเทียบข้อมูลโดยใช้ส่วนเบี่ยงเบนเฉลี่ยจะพบว่ามีค่าต่างกัน กล่าวคือส่วนเบี่ยงเบนเฉลี่ยของข้อมูล A เท่ากับ 2 และส่วนเบี่ยงเบนเฉลี่ยของข้อมูล B เท่ากับ 20 ซึ่งมีค่าต่างกัน หากจะตีความการกระจายจากค่า M.D. จะตีความว่าข้อมูล B กระจายมากกว่าข้อมูล A ทำให้การตีความผิดไป

แต่ถ้าดูการกระจายจากสัมประสิทธิ์ของส่วนเบี่ยงเบนเฉลี่ยจะได้ว่า

ข้อมูล A

$$\frac{M.D.}{\overline{X}} = \frac{2}{5}$$

ข้อมูล B

$$\frac{M.D.}{\overline{X}} = \frac{20}{50} = \frac{2}{5}$$

จะเห็นได้ว่า ข้อมูล A และข้อมูล B มีค่าสัมประสิทธิ์ของส่วนเบี่ยงเบนเฉลี่ยเท่ากัน

นั่นคือข้อมูล A และข้อมูล B มีการกระจายเท่ากัน

พิจารณาข้อมูลสองชุดต่อไปนี้

A: 2, 3, 5, 7, 8

B: 20, 30, 50, 70, 80

ข้อมูล A มี $\overline{X}$ = 5

$$s = \sqrt{\frac{(2-5)^2 + (3-5)^2 + (5-5)^2 + (7-5)^2 + (8-5)^2}{5}}$$

$$= \sqrt{\frac{26}{5}}$$

$$= 2.28$$

ข้อมูล B มี $\overline{X}$ = 50

$$s = \sqrt{\frac{(20-50)^2 + (30-50)^2 + (50-50)^2 + (70-50)^2 + (80-50)^2}{5}}$$

$$= \sqrt{\frac{2600}{5}}$$

$$= 22.8$$

ถ้าเปรียบเทียบข้อมูลโดยใช้ส่วนเบี่ยงเบนมาตรฐานจะพบว่ามีค่าต่างกัน กล่าวคือ ส่วนเบี่ยงเบนมาตรฐานของข้อมูล A เท่ากับ 2.28 และส่วนเบี่ยงเบนมาตรฐานข้อมูล B เท่ากับ 22.8 ซึ่งมีค่าต่างกัน หากจะตีความการกระจายจากค่า s จะตีความว่าข้อมูล B กระจายมากกว่าข้อมูล A ทำให้การตีความผิดไป

แต่ถ้าดูการกระจายจากสัมประสิทธิ์ของการแปรผันจะได้ว่า

ข้อมูล A

$$\frac{s}{\overline{X}} = \frac{2.28}{5} = 0.456$$

ข้อมูล B

$$\frac{s}{\overline{X}} = \frac{22.8}{50} = 0.456$$

จะเห็นได้ว่าข้อมูล A และข้อมูล Bมีค่าสัมประสิทธิ์ของการแปรผันเท่ากัน นั่นคือข้อมูล A และข้อมูล B มีการกระจายเท่ากัน

สัมประสิทธิ์ของการแปรผันเป็นวิธีวัดการกระจายสัมพัทธ์ที่นิยมใช้มากที่สุด โดยนิยมตอบในรูปเปอร์เซ็นต์โดยการคูณด้วย 100

ตัวอย่างที่ 1 จงเปรียบเทียบการกระจายของข้อมูลสองชุดต่อไปนี้โดยใช้การวัดการกระจายสัมพัทธ์ทั้งสี่แบบ

A : 10, 11, 11, 13, 14, 14, 16, 17, 19

B : 41, 46, 54, 68, 74, 78, 86, 88, 94

วิธีทำ (1) หาสัมประสิทธิ์ของพิสัย

ข้อมูล A มีสัมประสิทธิ์ของพิสัย

$$= \frac{x_{max} - x_{min}}{x_{max} + x_{min}} = \frac{19-10}{19+10} = \frac{9}{29} \approx 0.31$$

ข้อมูล B มีสัมประสิทธิ์ของพิสัย

$$= \frac{x_{max} - x_{min}}{x_{max} + x_{min}} = \frac{94-41}{94+41}$$

$$= \frac{53}{135}$$

$$\approx 0.39$$

จากการหาสัมประสิทธิ์ของพิสัยแสดงว่าข้อมูล B มีการกระจายของข้อมูลมากกว่าข้อมูล A

(2) หาสัมประสิทธิ์ของส่วนเบี่ยงเบนควอไทล์

A : 10, 11, 11, 13, 14, 14, 16, 17, 19

B : 41, 46, 54, 68, 74, 78, 86, 88, 94

ข้อมูล A มี Q_1 อยู่ในตำแหน่ง $\frac{9+1}{4}$ = 2.5 ดังนั้น Q_1 = 11

Q_3 อยู่ในตำแหน่ง $3(\frac{9+1}{4})$ = 7.5 ดังนั้น Q_3 = 16.5

ข้อมูล A มีสัมประสิทธิ์ของส่วนเบี่ยงเบนควอไทล์ $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$

$$= \frac{16.5-11}{16.5+11}$$

$$= 0.2$$

ข้อมูล B มี Q_1 อยู่ในตำแหน่ง $\frac{9+1}{4}$ = 2.5 ดังนั้น $Q_1 = \frac{46+54}{2} = 50$

Q_3 อยู่ในตำแหน่ง $3(\frac{9+1}{4})$ = 7.5 ดังนั้น Q_3 = 87

ข้อมูล B มีสัมประสิทธิ์ของส่วนเบี่ยงเบนควอไทล์ $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$

$$= \frac{87-50}{87+50}$$

$$= 0.27$$

จากการหาสัมประสิทธิ์ของของส่วนเบี่ยงเบนควอไทล์ ข้อมูล B มีการกระจายมากกว่าข้อมูล A

(3) หาสัมประสิทธิ์ของส่วนเบี่ยงเบนเฉลี่ย

A : 10, 11, 11, 13,14 ,14, 16, 17, 19

B : 41, 46, 54, 68, 74, 78, 86, 88, 94

ข้อมูล A มี $\overline{X} = \dfrac{10+11+11+13+14+14+16+17+19}{9}$

$\approx$ 13.9

M.D. $= \dfrac{3.9+2.9+2.9+0.9+0.1+0.1+2.1+3.1+5.1}{9}$

$\approx$ 2.3

$\dfrac{M.D.}{\overline{X}} = \dfrac{2.3}{13.9} \approx 0.17$

ข้อมูล B มี $\overline{X} = \dfrac{41+46+54+68+74+78+86+88+94}{9}$

$\approx$ 69.9

M.D. $= \dfrac{28.9+23.9+15.9+1.9+4.1+8.1+16.1+18.1+24.1}{9}$

$\approx$ 15.7

$\dfrac{M.D.}{\overline{X}} = \dfrac{15.7}{69.9}$

$\approx$ 0.22

จากการหาสัมประสิทธิ์ของส่วนเบี่ยงเบนเฉลี่ย ข้อมูล B มีการกระจายมากกว่าข้อมูล A

(4) หาสัมประสิทธิ์ของการแปรผัน

A : 10, 11, 11, 13, 14, 14, 16, 17, 19

B : 41, 46, 54, 68, 74, 78, 86, 88, 94

ข้อมูล A มี $\overline{X} \approx 13.9$

$$s = \sqrt{\frac{(3.9)^2+(2.9)^2+(2.9)^2+(0.9)^2+(0.1)^2+(0.1)^2+(2.1)^2+(3.1)^2+(5.1)^2}{9}}$$

$$= \sqrt{\frac{68.48}{9}}$$

$$\approx 2.76$$

$$\text{สัมประสิทธิ์ของการแปรผัน} = \frac{s}{\overline{X}} = \frac{2.76}{13.9} \approx 0.20$$

ข้อมูล B มี $\overline{X} \approx 69.9$

$$\text{M.D.} = \frac{28.9+23.9+15.9+1.9+4.1+8.1+16.1+18.1+24.1}{9}$$

$$s = \sqrt{\frac{(28.9)^2+(23.9)^2+(15.9)^2+(1.9)^2+(4.1)^2+(8.1)^2+(16.1)^2+(18.1)^2+(24.1)^2}{9}}$$

$$= \sqrt{\frac{2585.28}{9}}$$

$$\approx 16.95$$

สัมประสิทธิ์ของการแปรผัน $= \frac{s}{\overline{X}}$

$= \frac{16.95}{69.9}$

≈ 0.24

จากการหาสัมประสิทธิ์ของการแปรผัน ข้อมูล B มีการกระจายมากกว่าข้อมูล A

ตัวอย่างที่ 2 ผลสอบวัดความสามารถทางคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 6 ของโรงเรียนสองโรงเรียน นำเสนอด้วยตารางแจกแจงความถี่ได้ดังนี้

โรงเรียนเสริมวิทย์ศึกษา

คะแนน	จำนวน
45-49	3
50-54	5
55-59	8
60-64	12
65-69	24
70-74	13
75-79	8
80-84	4
รวม	71

โรงเรียนเจนอักษร

คะแนน	จำนวน
40-44	2
45-49	4
50-54	6
55-59	9
60-64	13
65-69	14
70-74	21
75-79	2
รวม	70

จงเปรียบเทียบการกระจายโดยใช้สัมประสิทธิ์ของการแปรผัน

วิธีทำ หาค่ามัชฌิมเลขคณิตและส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบของโรงเรียนเสริมวิทย์ศึกษา

คะแนน	x_i	f_i	$f_i\ x_i$	$(x_i - \bar{X})$	$(x_i - \bar{X})^2$	$f_i(x_i - \bar{X})^2$
45-49	47	3	141	-24.68	209.10	627.3
50-54	52	5	260	-19.68	307.30	1536.5
55-59	57	8	456	-14.68	215.50	1724
60-64	62	12	744	-9.68	93.70	1124.4
65-69	67	24	1608	-4.68	21.90	525.6
70-74	72	13	936	0.32	0.10	1.3
75-79	77	8	616	5.32	28.30	226.4
80-84	82	4	328	10.32	106.50	426
		71	5089			6191.5

$$\bar{X} = \frac{5089}{71} \approx 71.68$$

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

$$= \sqrt{\frac{6191.5}{71}}$$

$$\approx 9.33$$

หาค่ามัชฌิมเลขคณิตและส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบของโรงเรียนเจนอักษร

คะแนน	x_i	f_i	$f_i\ x_i$	$(x_i - \bar{X})$	$(x_i - \bar{X})^2$	$f_i(x_i - \bar{X})^2$
40-44	42	2	84	-22.39	501.31	1,002.62
45-49	47	4	188	-17.39	302.41	1,209.64
50-54	52	6	312	-12.39	153.51	921.06
55-59	57	9	513	-7.39	54.61	491.49
60-64	62	13	806	-2.39	5.71	74.23
65-69	67	14	938	2.61	6.81	95.34
70-74	72	21	1,512	7.61	57.91	1,216.11
75-79	77	2	154	12.61	159.01	318.02
		70	4, 507			5,328.51

$$\bar{X} = \frac{4507}{70} \approx 64.39$$

$$s = \sqrt{\frac{\sum_{i=1}^{k} f_i(x_i - \bar{X})^2}{N}} \approx \sqrt{\frac{5328.52}{70}} = 8.72$$

หาสัมประสิทธิ์ของการแปรผันของคะแนนโรงเรียนเสริมวิทย์ศึกษา

สัมประสิทธิ์ของการแปรผัน $= \frac{s}{\bar{X}} = \frac{9.33}{71.68} \approx 0.130$

หาสัมประสิทธิ์ของการแปรผันของคะแนนโรงเรียนเจนอักษร

สัมประสิทธิ์ของการแปรผัน $= \frac{s}{\bar{X}} = \frac{8.72}{64.38} \approx 0.135$

สัมประสิทธิ์ของการแปรผันของโรงเรียนเจนอักษรสูงกว่าของโรงเรียนเสริมวิทย์ศึกษาเล็กน้อยแสดงว่าคะแนนผลการสอบของโรงเรียนเจนอักษรกระจายมากกว่าของโรงเรียนเสริมวิทย์ศึกษาเล็กน้อย นอกจากนั้นค่าเฉลี่ยของโรงเรียนเจนอักษรก็ต่ำกว่าของโรงเรียนเสริมวิทย์ศึกษาค่อนข้างมาก

แบบฝึกหัด 3

1. จงหา Q_1, Q_2 และ Q_3 ของคะแนนต่อไปนี้

    1.1 12, 3, 0, 11, 7, 8, 15, 18, 12

    1.2 11, 13, 10, 14, 17, 18, 11, 12, 13, 16

    1.3 6, 6, 8, 12, 7, 7, 14, 14, 18, 17

    1.4 16, 13, 19, 11, 17, 18, 8, 10, 12, 14

2. จงหา D_1, D_3, D_6, D_8 ของคะแนนต่อไปนี้

    2.1 12, 13, 10, 15, 7, 18, 19, 11, 19, 12, 14

    2.2 15, 24, 23, 13, 19, 14, 18, 18, 16, 12, 13, 16

    2.3 16, 26, 38, 12, 37, 47, 24, 24, 18, 17, 23, 22

    2.4 24, 28, 46, 36, 43, 29, 31, 57, 28, 8, 10, 16, 14

3. จงหา P_{25}, P_{50}, P_{75}, P_{80} ของคะแนนผลการสอบวิชาคณิตศาสตร์เพิ่มเติมของนักเรียนชั้นมัธยมศึกษาปีที่ 6 จำนวน 48 คน

22	33	10	35	17	58	9	41	19	52	14
35	24	53	23	19	14	38	18	26	12	13
16	46	26	38	12	37	47	24	24	8	17
23	42	24	38	46	36	43	29	31	57	28
48	40	36	24							

4. จากข้อมูลในข้อ 3. จงหาว่า

    4.1 นักเรียนต้องสอบได้คะแนนเท่าไรจึงจะมีประมาณครึ่งหนึ่งของนักเรียนในชั้นได้คะแนนต่ำกว่า

    4.2 นักเรียนจะต้องสอบได้คะแนนเท่าไรจึงจะมีนักเรียนประมาณหนึ่งในสี่ของชั้นได้คะแนนสูงกว่า

    4.3 นักเรียนจะต้องสอบได้คะแนนเท่าไรจึงจะมีผู้สอบได้คะแนนน้อยกว่าอยู่ 6 ใน10

5. คะแนนผลการสอบวิชาภาษาไทยของนักเรียนชั้นมัธยมศึกษาจำนวน 120 คนแสดงดังตารางแจกแจงความถี่ต่อไปนี้

คะแนน	ความถี่
21-25	6
26-30	12
31-35	16
36-40	23
41-45	30
46-50	18
51-55	10
56-60	5
รวม	120

5.1 จงหา Q_2, D_5 และ P_{50} แล้วเปรียบเทียบค่าที่ได้

5.2 จงหา $\frac{Q_3 - Q_1}{2}$

6. จากการสำรวจจำนวนเสื้อยี่ห้อต่าง ๆที่ร้านค้าที่ตัวแทนจำหน่ายแห่งหนึ่งจำหน่ายได้ในวันที่ 1 เดือนธันวาคม ปี 2550 มีดังนี้

ยี่ห้อ	ราคาต่อหน่วย	จำนวนหน่วย
A	485	24
B	550	32
C	675	22
D	875	14
E	990	31
F	1220	16
G	1320	8
H	1440	9

6.1 จงหาค่ามัชฌิมเลขคณิต มัธยฐาน และฐานนิยมของราคาเสื้อที่ขายได้ในวันดังกล่าว

6.2 จงหา Q.D. ของราคาเสื้อผ้าที่ขายได้ในวันดังกล่าว

6.3 จงหา M.D. ของราคาเสื้อที่ขายได้ในวันดังกล่าว

7. จากการศึกษารายได้ของคนในหมู่บ้านแห่งหนึ่ง จำนวน 138 คนปรากฏข้อมูลดังนี้

รายได้	จำนวน
28,000-30,999	8
31,000-33,999	14
34,000-36,999	25
37,000-39,999	19
40,000-42,999	19
43,000 -45,999	17
46,000-48,999	14
49,000-51,999	10
52,000-54,999	8
55,000-57,999	4

7.1 จงหาค่ามัชฌิมเลขคณิต มัธยฐาน และฐานนิยมของรายได้ของคนในหมู่บ้านแห่งนี้

7.2 จงหา Q.D. ของรายได้ของคนในหมู่บ้านนี้

7.3 จงหา M.D. ของรายได้ของคนในหมู่บ้านนี้