บทที่ 5

การวัดการกระจาย (Measures of Dispersion)

ในบทที่ 4 เราได้พูดถึงการวัดแนวโน้มเข้าสู่ส่วนกลางของข้อมูลชุดหนึ่ง ๆ ในบทนี้เราจะหาที่ตั้งค่ากลางรอบ ๆ ซึ่งชี้ให้เห็นการกระจายของข้อมูลรอบ ๆ ค่ามัชฌิม ค่าสังเกตของข้อมูลชุดหนึ่ง ๆ จะกระจายไปตลอดทั้งหมดในพิสัยระหว่างค่าสูงสุด และค่าต่ำสุดของข้อมูลชุดนั้น ๆ ซึ่งบางค่าจะเข้าใกล้กับค่ากลาง บางค่าอาจจะไกลจากค่ากลาง ถ้าพิจารณาจากข้อมูลดังต่อไปนี้

5, 5, 5, 5, 5, 5, 5, 5, 5, 5

ค่ามัชฌิมเลขคณิต $= \frac{50}{10} = 5$

และค่าเบี่ยงเบนจากมัชฌิมเลขคณิตของทุก ๆ ค่าจะเท่ากับ 0

ดังนั้น จะเห็นได้ว่าข้อมูลชุดนี้ไม่มีการกระจาย

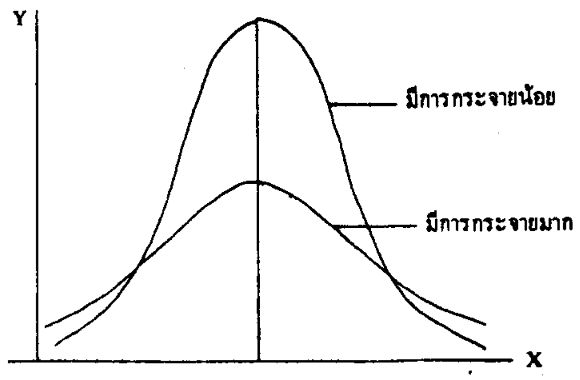
ตัวอย่าง ถ้ามีข้อมูลดังต่อไปนี้

4, 3, 6, 5, 7, 4, 8, 8, 2, 3

ค่ามัชฌิมเลขคณิต $= \frac{50}{10} = 5$ เช่นเดียวกับตัวอย่างแรก แต่ลักษณะการแจกแจงของแต่ละค่าแตกต่างกัน

ดังนั้น การวัดแนวโน้มเข้าสู่ส่วนกลางอย่างเดียวไม่เพียงพอที่จะบอกลักษณะการแจกแจงของค่าสังเกตต่าง ๆ ในข้อมูลนั้นได้ เพราะว่าข้อมูลบางชุดอาจจะมีค่ากระจายที่ห่างจากค่ากลางมาก ในขณะที่ข้อมูลบางชุดอาจจะเข้าใกล้ค่ากลาง ซึ่งถ้าเขียน Curves แสดงการแจกแจงความถี่จะเห็นได้ชัดเจนยิ่งขึ้น

จากรูปแสดงให้เห็นว่าข้อมูลจากตัวอย่างทั้ง 2 ข้างต้นนั้น มีมัชฌิมเลขคณิตเท่ากัน แต่ข้อมูลชุดแรกมีการกระจายน้อยกว่าข้อมูลชุดที่ 2 ดังรูป



การวัดการกระจายที่สำคัญมี 5 ชนิดด้วยกันคือ

1. พิสัย (Ranges)
2. ส่วนเบี่ยงเบนเฉลี่ย (Mean Deviation)
3. ส่วนเบี่ยงเบนควอร์ไทล์ (Quartile Deviation)
4. พิสัยเดไซล์ (Decile Range)
5. ความแปรปรวน และส่วนเบี่ยงเบนมาตรฐาน (Variance and Standard deviation)

ซึ่งจะกล่าวถึงวิธีวัดแต่ละวิธีดังต่อไปนี้

5.1 **พิสัย** (Ranges)

เป็นวิธีวัดการกระจายของข้อมูลที่ง่ายที่สุด โดยการหาความแตกต่างของข้อมูลที่สูงที่สุด และข้อมูลที่ต่ำที่สุด

ถ้าให้ R เป็นพิสัย

$\therefore$ R = ค่าสูงที่สุด – ค่าต่ำที่สุด

หรือ $R = X_{max} - X_{min}$

เมื่อ X_{max} = ค่าสูงสุดของข้อมูล

X_{min} = ค่าต่ำสุดของข้อมูล

ดังนั้นจะเห็นว่าค่าพิสัยเป็นค่าที่วัดการกระจายของข้อมูลอย่างหยาบ ๆ เพราะนำเอาค่าสองค่าของข้อมูล (คือค่าสูงสุดกับค่าต่ำสุด) มาใช้ในการคำนวณเท่านั้น ส่วนค่าอื่น ๆ ในข้อมูลไม่ได้นำมาใช้ในการวัดการกระจายเลย เพราะฉะนั้นถ้าข้อมูลชุดหนึ่งมีค่าใกล้เคียงกันหมด ยกเว้นค่าหนึ่ง ซึ่งมีค่าสูงกว่าปกติหรือต่ำกว่าปกติจะทำให้พิสัยที่ได้มีค่าผิดปกติไปด้วย ตัวอย่างเช่น การชั่งน้ำหนักหมู 2 คอก ปรากฏผลดังนี้

คอกที่ 1 52, 53, 55, 57, 60, 62

คอกที่ 2 52, 54, 55, 56, 61, 95

พิสัยของน้ำหนักหมูคอกที่ 1 คือ 62-52 = 10

พิสัยของน้ำหนักหมูคอกที่ 2 คือ 95 − 52 = 43

จะเห็นว่าน้ำหนักของหมูทั้ง 2 คอก คล้ายคลึงกัน แต่ในคอกที่ 2 น้ำหนักหมูมีผิดปกติอยู่ 1 ตัว ซึ่งชั่งน้ำหนักได้ 95 จึงทำให้พิสัยของน้ำหนักของหมูทั้ง 2 คอก ต่างกันมาก

ถึงแม้ว่าพิสัยเป็นค่าวัดการกระจายอย่างหยาบ แต่ก็เป็นค่าที่ใช้กันเสมอสำหรับคนทั่ว ๆ ไป เพราะค่าพิสัยเป็นค่าที่คำนวณง่าย สังเกตได้เร็ว จึงเหมาะที่จะนำไปใช้ในกรณีที่ต้องการทราบการกระจายของข้อมูลโดยรวดเร็ว

การคำนวณหาพิสัย

ก. **สำหรับข้อมูลที่ไม่ได้จัดกลุ่ม** (Ungrouped data)

หาได้จาก

$$R = x_{max} - x_{min}$$

ตัวอย่าง แสดงได้ดังตารางต่อไปนี้

จำนวนข้อมูล	พิสัย	
	ผลต่าง	ค่าต่ำสุดถึงค่าสูงสุด
1, 5, 7, 13	13 − 1 = 12	1 ถึง 13
14, 3, 17, 4, 8, 73, 36, 48	73 − 3 = 70	3 ถึง 73
3.2, 4.7, 5.6, 2.1, 1.9, 10.3	10.3 − 1.9 = 8.4	1.9 ถึง 10.3

ข. **สำหรับข้อมูลที่จัดกลุ่ม** (grouped data)

ในกรณีที่ข้อมูลจัดกลุ่มในตารางแจกแจงความถี่ การหาพิสัยนั้นทำได้ 2 วิธี คือ

1. หาผลต่างระหว่างขีดจำกัดบนของชั้นสูงสุดกับขีดจำกัดล่างของชั้นต่ำสุด

$\therefore$ R = ขีดจำกัดบนของชั้นสูงสุด − ขีดจำกัดล่างของชั้นต่ำสุด

2. หาผลต่างระหว่างจุดกึ่งกลางของชั้นสูงสุดกับชั้นต่ำสุด

$\therefore$ R = ค่ากึ่งกลางของชั้นสูงสุด − ค่ากึ่งกลางของชั้นต่ำสุด

ซึ่งทั้ง 2 วิธีจะให้ค่าพิสัยออกมาไม่เท่ากัน ในทางปฏิบัติเรามักจะใช้ทั้ง 2 วิธี

ตัวอย่าง การวัดความสูงของนักศึกษา 106 คน เป็นดังนี้

ความสูง	ความถี่
120-123	1
124-127	0
128-131	2
132-135	7
136-139	21
140-143	41
144-147	19
148-151	12
152-155	2
156-159	1
รวม	106

หาพิสัยวิธีที่ 1

ในที่นี้ขีดจำกัดบนของชั้นสูงสุด (156-159) เท่ากับ 159

และขีดจำกัดล่างของชั้นต่ำสุด (120-123) เท่ากับ 120

$\therefore$ พิสัย $= 159 - 120 = 39$

หาพิสัยโดยวิธีที่ 2

จากตารางจุดกึ่งกลางของชั้นสูงสุด (156 – 159) เท่ากับ 157.5

และจุดกึ่งกลางของชั้นต่ำสุด (120 – 123) เท่ากับ 121.5

$\therefore$ พิสัย $= 157.5 - 121.5 = 36$

5.2 ส่วนเบี่ยงเบนเฉลี่ย (Mean Deviation or Average Deviation)

เป็นการวัดการกระจายของข้อมูล โดยวัดจากค่าเฉลี่ยของส่วนเบี่ยงเบนของแต่ละข้อมูลจากมัชฌิมเลขคณิต (ไม่คิดเครื่องหมาย)

ถ้ามีข้อมูลชุดหนึ่ง มีมัชฌิมเลขคณิตเป็น $\overline{X}$

$\therefore$ ส่วนเบี่ยงเบนจากมัชฌิมเลขคณิต ($\overline{X}$) ของข้อมูลทั้งหมด

คือ $(X_1 - \overline{X}), (X_2 - \overline{X}), (X_3 - \overline{X}), \ldots, (X_n - \overline{X})$

ซึ่งถ้าเราจะหาผลรวมของส่วนเบี่ยงเบนจากมัชฌิมเลขคณิตของข้อมูลทั้งหมด จะได้ $\Sigma (X - \overline{X})$

$$\text{จาก } \Sigma (X - \overline{X}) = \Sigma X - \Sigma \overline{X}$$

$$= N\overline{X} - N\overline{X} = 0$$

และ $\Sigma (X - \overline{X})$ จะเท่ากับศูนย์เสมอ ไม่ว่าข้อมูลจะกระจายมากหรือน้อย ซึ่งจะทำให้ผลรวมของส่วนเบี่ยงเบนจากมัชฌิมเลขคณิตใช้ประโยชน์ไม่ได้ ดังนั้นจึงต้องหาผลรวมของส่วนเบี่ยงเบนโดยไม่คำนึงถึงเครื่องหมายบวกและลบ นั่นก็คือ การหาค่าสัมบูรณ์ (Aboolute value) ของผลรวมของส่วนเบี่ยงเบน

$$\therefore \text{ ส่วนเบี่ยงเบนเฉลี่ย (MD หรือ AD)} = \frac{\Sigma |X - \overline{X}|}{N}$$

วิธีคำนวณหาส่วนเบี่ยงเบนเฉลี่ย

ก. สำหรับข้อมูลที่ไม่ได้จัดกลุ่ม (Ungrouped data)

$$\text{MD (หรือ AD)} = \frac{\Sigma |X - \overline{X}|}{N}$$

ตัวอย่างเช่น อายุของบ้าน 5 หลัง เป็นดังนี้คือ 2, 2, 4, 5 และ 2 ปี ตามลำดับ จงหาส่วนเบี่ยงเบนเฉลี่ยของอายุของบ้าน 5 หลังนี้

จากข้อมูล คำนวณหา $\overline{X}$ ได้เท่ากับ

$$\overline{X} = \frac{2+2+4+5+2}{5} = \frac{15}{5} = 3$$

X	$\|X - \overline{X}\|$
2	$2-3 = 1$
2	$2-3 = 1$
4	$4-3 = 1$
5	$5-3 = 2$
2	$2-3 = 1$
รวม	6

$$\therefore \text{ AD} = \frac{\Sigma |X - \overline{X}|}{N} = \frac{6}{5} = 1.2 \text{ ปี}$$

สรุปขั้นตอนในการหา AD **สำหรับ** Ungrouped data

1. หามัชฌิมเลขคณิตของข้อมูลชุดนั้น
2. หาผลต่างของค่าสังเกตทุกตัวกับค่ามัชฌิมเลขคณิตโดยไม่คิดเครื่องหมาย
3. หาผลรวมในข้อ 2
4. หารผลรวมที่ได้ในข้อ 3 ด้วยจำนวนข้อมูลทั้งหมดก็จะได้ค่า AD ตามที่ต้องการ

ข. สำหรับข้อมูลที่จัดกลุ่ม (Grouped data)

สำหรับข้อมูลที่แจกแจงด้วยตารางแจกแจงความถี่หา AD ได้จากสูตร

$$AD = \frac{\Sigma f\,|X-\overline{X}|}{N}$$

เมื่อ f = ความถี่ของแต่ละชั้น

X = ค่ากึ่งกลางของแต่ละชั้น

ตัวอย่าง รายได้ต่อวันของคนงาน 100 คน เป็นดังนี้จงหา AD

รายได้	ความถี่ (จำนวนคนงาน)
60-62	5
63-65	18
66-68	42
69–71	27
72–74	8
รวม	100

จากตาราง คำนวณหา $\overline{X}$ โดยสร้างช่องค่ากึ่งกลาง (X), $|X-\overline{X}|$ และ $f\,|X-\overline{X}|$ หา $\overline{X}$ โดยวิธีลัดได้ดังนี้

รายได้	ค่ากึ่งกลาง	f	d	fd
60-62	61	5	- 2	- 10
63-65	64	18	- 1	- 18
66-68	67	42	0	0
69-71	70	27	1	27
72-74	73	8	2	16
รวม		loo		15

จาก $\bar{X} = a + i\bar{d}$

$$\bar{X} = 67 + 3\left(\frac{\Sigma fd}{\Sigma f}\right)$$

$$= 67 + 3\left(\frac{15}{100}\right)$$

$$= 67 + \frac{45}{100} = 61.45$$

เมื่อหา $\bar{X}$ ได้แล้วสร้างตารางหา AD

รายได้	ค่ากึ่งกลาง	f	$\lvert X-\bar{X}\rvert = \lvert X-67.45\rvert$	$f\lvert X-\bar{X}\rvert$
60-62	61	5	6.45	32.25
63-65	64	18	3.45	62.10
66-68	67	42	0.45	18.90
69-71	70	27	2.55	68.85
72+74	73	8	5.55	44.40
รวม		loo		226.50

$$\therefore \quad AD = \frac{\Sigma f\lvert X-\bar{X}\rvert}{N}$$

$$= \frac{226.50}{100} = 2.265$$

สรุปขั้นตอนในการหา AD สำหรับ Grouped data

1. หา $\overline{X}$ โดยวิธีตรงหรือวิธีลัด
2. สร้างช่องค่ากึ่งกลาง
3. สร้างช่อง $|X-\overline{X}|$
4. สร้างช่อง $f\,|X-\overline{X}|$
5. หาผลรวมของ $f\,|X-\overline{X}|$
6. หารผลรวมข้อ 5 ด้วยจำนวนข้อมูลทั้งหมด (N) ก็จะได้ AD ตามที่ต้องการ

5.3 ส่วนเบี่ยงเบนควอร์ไทล์ (Quartile Deviation)

ส่วนเบี่ยงเบนควอร์ไทล์เราใช้สัญลักษณ์แทนด้วย "Q" หรือ "QD" หรือมักจะเรียกกันว่า กึ่งพิสัยควอร์ไทล์ (Semi-interquartile Range) ในบทที่แล้ว ค่าของควอร์ไทล์ 2 ค่า ที่เราได้กล่าวถึงมาแล้วคือ Q_1 และ Q_3 ส่วนเบี่ยงเบนควอร์ไทล์ หรือกึ่งพิสัยควอร์ไทล์คือครึ่งหนึ่งของระยะทางระหว่างควอร์ไทล์ทั้ง 2 นี้ ซึ่งเขียนเป็นสูตรได้ดังนี้

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

ตัวอย่างเช่น จากตารางดังต่อไปนี้ จงคำนวณส่วนเบี่ยงเบนควอร์ไทล์

คะแนน	ความถี่	ความถี่สะสม
25 – 29	1	1
30 – 34	0	1
35 – 39	3	4
40 – 44	6	10
45 – 49	6	16
50 – 54	6	22
55 – 59	7	29
60 – 64	4	33
65 – 69	4	37

คะแนน	ความถี่	ความถี่สะสม
70 – 74	1	38
75 – 79	1	39
80 – 84	1	40
รวม	40	

จากตาราง ต้องคำนวณหา Q_1 และ Q_3 ก่อน โดยใช้วิธีที่เรียนมาแล้วข้างต้น

$\therefore$ Q_1 คือคะแนนที่บอกให้ทราบว่ามี $\frac{N}{4}$ คะแนนที่ต่ำกว่าและอีก $\frac{3N}{4}$ คะแนนที่สูงกว่า

$$\therefore \quad \frac{N}{4} = \frac{40}{4} = 10$$

ดูจากช่องความถี่สะสมจะเห็นว่าความถี่ 10 อยู่ในชั้น 40 – 44 (39.5 – 44.5) พอดี (ไม่ต้องเทียบบัญญัติไตรยางค์)

$$\therefore \quad Q_1 = 44.5$$

Q_3 คือคะแนนที่บอกให้ทราบว่ามี $\frac{3N}{4}$ คะแนนที่ต่ำกว่า และอีก $\frac{N}{4}$ คะแนนที่สูงกว่า

$$\frac{3N}{4} = \frac{3 \times 40}{4} = 30$$

ดูช่องความถี่สะสมจะเห็นว่าความถี่ 30 อยู่ในชั้น 60 – 64 (59.5 – 64.5)

จากความถี่ต่างกัน 4 คะแนนต่างกัน 5

ความถี่ต่างกัน 1 คะแนนต่างกัน $\frac{5 \times 1}{4} = 1.25$

$$\therefore \quad Q_3 = 59.5 + 1.25 = 60.75$$

เมื่อได้ Q_1 และ Q_3 แล้วนำไปแทนค่าในสูตรก็จะได้ Q

จาก $$Q = \frac{Q_3 - Q_1}{2}$$

$$\therefore \quad Q = \frac{60.75 - 44.5}{2}$$

$$= \frac{16.25}{2}$$

$$= 8.125$$

$$\approx 8.13$$

5.4 พิสัยเดไซล์ (Decile Range)

คือผลต่างระหว่างเดไซล์ที่ 9 และเดไซล์ที่ 1 หรือเปอร์เซ็นไทล์ที่ 90 กับเปอร์เซ็นไทล์ที่ 10

$\therefore$ พิสัยเดไซล์ $= D_9 - D_1$

หรือ พิสัยเดไซล์ $= P_{90} - P_{10}$

5.5 ความแปรปรวนและส่วนเบี่ยงเบนมาตรฐาน (Variance and Standard deviation)

เพื่อให้ได้การวัดการกระจายที่สะดวกกว่า ส่วนเบี่ยงเบนเฉลี่ย เราจึงใช้วิธียกกำลังสองของส่วนเบี่ยงเบน เพื่อให้เป็นบวกหมด แล้วหาค่าเฉลี่ยของค่าที่ยกกำลังสองนี้ ค่าที่ได้นี้เรียกว่า ความแปรปรวน (Variance)

ค่าความแปรปรวนที่ได้ออกมาจะเป็นเลขคงที่จำนวนหนึ่งและเลขจำนวนนี้จะมีค่าเป็นบวกเสมอ

ความแปรปรวนเป็นการวัดการกระจายที่สำคัญ เชื่อถือได้ดีและเป็นการวัดการกระจายที่ดีที่สุด สามารถนำไปใช้ในการคำนวณทางสถิติขั้นสูงต่อไปได้

การคำนวณหาความแปรปรวนเราอาจจะหาความแปรปรวนจากประชากรทั้งหมด (Population) หรือจากตัวอย่าง (Sample) ก็ได้ ถ้าคำนวณจากประชากรทั้งหมด ความแปรปรวนที่ได้ก็จะเป็นความแปรปรวนของประชากร ซึ่งเราใช้สัญลักษณ์แทนด้วย "σ^2" (อ่านว่า ซิกม่ากำลัง 2) แต่ถ้าคำนวณจากตัวอย่าง ความแปรปรวนที่ได้ก็จะเป็นความแปรปรวนของตัวอย่าง ซึ่งเราใช้สัญลักษณ์แทนด้วย "S^2"

ดังนั้น เราจะหา σ^2 ได้จาก

$$\sigma^2 = \frac{\Sigma (X-\mu)^2}{N}$$

เมื่อ X = ค่าของข้อมูล

μ = ค่าเฉลี่ยของประชากร

N = จำนวนประชากรทั้งหมด

ในทางปฏิบัติค่าของ σ^2 เราไม่สามารถจะคำนวณได้เพราะจำนวนประชากรมักจะมีขนาดใหญ่มาก ดังนั้น เรามักจะใช้การสุ่มตัวอย่างแทน ซึ่งเราจะหาความแปรปรวนของตัวอย่าง (S^2) ได้จากสูตร

$$S^2 = \frac{\Sigma (X-\bar{X})^2}{n-1}$$

เมื่อ X = ค่าสังเกตของตัวอย่าง

$\bar{X}$ = ค่าเฉลี่ยของตัวอย่าง

n = ขนาดของตัวอย่างที่สุ่มมา

ตัวอย่าง จงหาความแปรปรวนของตัวอย่างดังต่อไปนี้ 2, 4, 6, 8, 10

X	$\bar{X}$	$(X-\bar{X})$	$(X-\bar{X})^2$
2	6	−4	16
4	6	−2	4
6	6	0	0
8	6	2	4
10	6	4	16
รวม		0	40

จาก $$S^2 = \frac{\Sigma (X-\bar{X})^2}{n-1}$$

$$\therefore \quad S^2 = \frac{40}{5-1} = 10$$

สรุปขั้นตอน ในการคำนวณหาความแปรปรวนมีดังนี้

1. คำนวณหาค่าเฉลี่ย
2. นำค่าเฉลี่ยไปลบออกจากค่าสังเกตแต่ละค่า

3. หากำลังสองของส่วนเบี่ยงเบนแต่ละตัว

4. หาผลรวมของกำลัง 2 ของส่วนเบี่ยงเบน

5. หารผลรวมที่ได้ด้วย $n-1$ สำหรับตัวอย่าง หรือหารผลรวมที่ได้ด้วย N สำหรับประชากร ก็จะได้ความแปรปรวนตามที่ต้องการ

ในกรณีที่มีข้อมูลจำนวนมาก ๆ จะเห็นว่าการคำนวณหา σ^2 หรือ S^2 โดยใช้สูตรดังกล่าวข้างต้นจะยุ่งยากมาก จึงไม่เป็นที่นิยมส่วนใหญ่แล้วในทางปฏิบัติจะใช้คำนวณโดยวิธีใช้สูตรดังนี้ (วิธีลัด)

$$\sigma^2 = \frac{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}{N}$$

และ $$S^2 = \frac{\Sigma X^2 - \frac{(\Sigma X)^2}{n}}{n-1}$$

ค่า σ^2 และ S^2 ที่คำนวณออกมาได้ โดยวิธีใช้สูตรนี้ (วิธีลัด) จะเท่ากับการคำนวณโดยวิธีใช้สูตรแรก (วิธีตรง)

$$\because \quad \Sigma(X-X)^2 = \Sigma(X^2 - 2X\,\overline{X} + \overline{X}^2)$$

$$= \Sigma X^2 - 2\overline{X}\,\Sigma X + \Sigma\,\overline{X}^2$$

จาก $$\Sigma X = n\overline{X}$$

$$\therefore \quad \Sigma(X-\overline{X})^2 = \Sigma X^2 - 2\overline{X}(n\overline{X}) + \Sigma\,\overline{X}^2$$

$$= \Sigma X^2 - 2n.\overline{X}^2 + n.\overline{X}^2$$

$$= \Sigma X^2 - n\,\overline{X}^2$$

$$= \Sigma X^2 - n\left(\frac{\Sigma X}{n}\right)^2$$

$$= \Sigma X^2 - n\frac{(\Sigma X)^2}{n^2}$$

$$= \Sigma X^2 - \frac{(\Sigma X)^2}{n}$$

$\therefore$ จาก $$S^2 = \frac{\Sigma(X-\overline{X})^2}{n-1}$$

แทนค่า $\Sigma(X-\bar{X})^2$ จะได้

$$S^2 = \frac{\Sigma X^2 - \frac{(\Sigma X)^2}{n}}{n-1}$$

จากตัวอย่างเดิม มีข้อมูลดังต่อไปนี้คือ 2, 4, 6, 8, 10 จงหา S^2

X	X^2
2	4
4	16
6	36
8	64
10	100
$\Sigma X = 30$	$\Sigma X^2 = 220$

$$\therefore \quad S^2 = \frac{220 - \frac{(30)^2}{5}}{5-1}$$

$$= \frac{220-180}{4} = \frac{40}{4} = 10$$

ซึ่งค่าที่ได้ออกมาจะเท่ากับค่าที่คำนวณโดยใช้สูตรแรก (วิธีตรง)

สรุปขั้นตอนในการคำนวณหาความแปรปรวนโดยวิธีลัด

1. ยกกำลัง 2 ของค่าสังเกตทุก ๆ ค่า
2. หาผลรวมของค่าสังเกต
3. หาผลรวมของกำลังสองของค่าสังเกตที่ได้ในข้อ 1
4. แทนค่าในสูตร ก็จะได้ความแปรปรวนตามที่ต้องการ

ที่กล่าวมาข้างต้นนี้เป็นการคำนวณหาความแปรปรวนของข้อมูลที่ไม่ได้จัดกลุ่ม (Ungrouped data) ถ้าข้อมูลที่ได้เรานำมาจัดกลุ่ม (grouped data) เรียบร้อยแล้ว เราคำนวณหาความแปรปรวนได้ดังนี้ (จะกล่าวถึงข้อมูลที่ได้จากตัวอย่างเท่านั้น)

การคำนวณโดยวิธีตรง

หา S^2 ได้จาก

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

เมื่อ k = **จำนวนชั้น**

f_i = **ความถี่ของชั้นที่ i**

X_i = **ค่ากึ่งกลางของชั้นที่ i**

n = **จำนวนของตัวอย่าง**

ซึ่งการคำนวณโดยวิธีตรงไม่นิยมเนื่องจากยุ่งยากมากนักจึงใช้การคำนวณโดยวิธีลัด ดังนี้

การคำนวณ S^2 โดยวิธีลัด ทำได้ 2 วิธีคือ

1.
$$S^2 = \frac{\sum_{i=1}^{k} f_i X_i^2 - \frac{\left(\sum_{i=1}^{k} f_i X_i\right)^2}{n}}{n-1}$$

2.
$$S^2 = i^2 \left[\frac{\sum_{i=1}^{k} f_i d_i^2 - \frac{\left(\sum_{i=1}^{k} f_i d_i\right)^2}{n}}{n-1} \right]$$

เมื่อ $d_i = \frac{X_i - a}{i}$

a = **มัชฌิมสมมติ**

i = **อันตรภาคชั้น**

ตัวอย่าง

จงคำนวณหา S^2 จากตารางต่อไปนี้

คะแนน	ความถี่
26-28	I
23-25	4
20-22	7
17-19	12
14-16	18
11-13	II
8-10	9
5-7	3
2-4	1
รวม	66

วิธีหา เราใช้วิธีลัด ซึ่งต้องสร้างช่อง X_i, X_i^2, f_iX_i และ $f_iX_i^2$ ดังนี้

คะแนน	ความถี่ (f)	x_i	X_i^2	f_iX_i	$f_iX_i^2$
26-28	1	27	729	27	729
23-25	4	24	576	96	2,304
20-22	7	21	441	147	3,087
17-19	12	18	324	216	3,888
14-16	18	15	225	270	4,050
11-13	11	12	144	132	1,584
8-10	9	9	81	81	729
s-7	3	6	36	18	108
2-4	1	3	9	3	9
รวม	66			990	16,488

จาก

$$S^2 = \frac{\Sigma f_i X_i^2 - \frac{(\Sigma f_i X_i)^2}{n}}{n-1}$$

แทนค่าจะได้

$$S^2 = \frac{16,488 - \frac{(990)^2}{66}}{66-1}$$

$$= \frac{16,488 - \frac{980,100}{66}}{65}$$

$$= \frac{16,488 - 14,850}{65}$$

$$= \frac{1,638}{65}$$

$$= 25.20$$

สรุปขั้นตอนในการหา S^2 โดยวิธีลัด มีดังนี้

1. สร้างช่องจุดกึ่งกลาง X_i
2. สร้างช่อง X_i^2
3. สร้างช่อง f_iX_i
4. หาผลรวมของ f_iX_i
5. สร้างช่อง $f_iX_i^2$
6. หาผลรวมของ $f_iX_i^2$
7. นำไปแทนค่าในสูตรจะได้ S^2 ตามที่ต้องการ

อีกวิธีหนึ่งหาได้จากสูตร

$$S^2 = i^2 \left[\frac{\Sigma f_i d_i^2 - \frac{(\Sigma f_i d_i)^2}{n}}{n-1}\right]$$

$\therefore$ ต้องสร้างช่อง X_i, d_i, d_i^2, f_id_i และ $f_id_i^2$ ดังตัวอย่างต่อไปนี้

จากตารางต่อไปนี้ จงหา S^2

จำนวนน้ำฝนที่ตก (นิ้ว)	f_i	X_i	d_i	d_i^2	f_id_i	$f_id_i^2$
3-5	3	4	−2	4	−6	12
6-8	5	7	−1	1	−5	5
9-11	10	10	0	0	0	0
12-14	7	13	1	1	7	7
รวม	25				−4	24

จาก $$S^2 = i^2\left[\frac{\Sigma f_id_i^2 - \frac{(\Sigma f_id_i)^2}{n}}{n-1}\right]$$

ในที่นี้ $i = 3$

$$\therefore S^2 = 3^2\left[\frac{24-\frac{(-4)^2}{25}}{25-1}\right]$$

$$= 3^2\left[\frac{24-\frac{16}{25}}{24}\right]$$

$$= 9\left[\frac{\frac{600-16}{25}}{24}\right]$$

$$= \frac{9}{24}\times\frac{584}{25}$$

$$= \frac{5256}{600} = 8.76$$

สรุปขั้นตอนในการหา S^2 โดยวิธีลัดวิธีที่ 2 มีดังนี้

1. สร้างช่อง X_i
2. สร้างช่อง d_i โดยที่ $d_i = \frac{X_i - a}{i}$
3. สร้างช่อง d_i^2
4. สร้างช่อง f_id_i แล้วหาผลรวม f_id_i

5. สร้างช่อง $f_id_i^2$ แล้วหาผลรวม $f_id_i^2$

6. นำค่าที่ได้ไปแทนค่าในสูตรถ้าจะได้ S^2 ตามที่ต้องการ

จะเห็นได้ว่าหน่วยของความแปรปรวนที่ได้เป็นกำลังสอง (เช่น นิ้ว^2, บาท^2) ซึ่งแตกต่างจากหน่วยของข้อมูล

เนื่องจากในทางปฏิบัติมีความลำบากยุ่งยากในการนำความแปรปรวนไปใช้ 2 อย่างด้วยกัน คือ ข้อที่ 1 ความแปรปรวนที่ได้มักจะเป็นจำนวนเลขที่มาก เมื่อเปรียบเทียบกับค่าสังเกตของข้อมูลที่ได้ เช่น ถ้ามีค่าสังเกตเป็นหลัก 1,000 ความแปรปรวนที่ได้จะเป็นหลักล้าน เป็นต้น

ข้อที่ 2 หน่วยของความแปรปรวนที่ได้จะไม่เหมือนกับหน่วยของค่าสังเกตจากข้อมูล เช่น หน่วยของข้อมูลเป็นฟุต หน่วยของความแปรปรวนจะเป็นฟุต^2 เป็นต้น แต่อย่างไรก็ตามความแปรปรวนก็ยังมีความสำคัญในทางคณิตศาสตร์ หรือทฤษฎีทางสถิติมาก

ดังนั้นเพื่อไม่ให้เกิดความยุ่งยากในการตีความหมายในทางปฏิบัติ เราจึงใช้ส่วนเบี่ยงเบนมาตรฐาน (Standard deviation) มากกว่า เพราะว่าตีความหมายได้ดีกว่า

ส่วนเบี่ยงเบนมาตรฐานหาได้โดยถอดรากที่สองของความแปรปรวน หน่วยของส่วนเบี่ยงเบนมาตรฐานจึงเป็นหน่วยเดียวกับหน่วยของข้อมูล และมีหน่วยเดียวกับค่าเฉลี่ยด้วย

ส่วยเบี่ยงเบนมาตรฐานมีประโยชน์มากในทางทฤษฎีสถิติขั้นสูงต่อไป

ถ้าให้ σ = ส่วนเบี่ยงเบนมาตรฐานของประชากร

$\therefore \sigma = \sqrt{\sigma^2}$ (ใช้เฉพาะค่าบวกเท่านั้น)

และทำให้ S = ส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง

$\therefore S = \sqrt{S^2}$ (ใช้เฉพาะค่าบวกเท่านั้น)

กรณีที่ข้อมูลไม่ได้จัดกลุ่ม (Ungrouped data)

ก. **คำนวณโดยวิธีตรง**

$$\sigma = \sqrt{\frac{\Sigma(X-\mu)^2}{N}}$$

$$S = \sqrt{\frac{\Sigma(X-\overline{X})^2}{n-1}}$$

ข. คำนวณโดยวิธีลัด

$$\sigma = \sqrt{\frac{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}{N}}$$

$$S = \sqrt{\frac{\Sigma X^2 - \frac{(\Sigma X)^2}{n}}{n-1}}$$

กรณีที่ข้อมูลนำมาจัดกลุ่มแล้ว (grouped data)

(จะกล่าวถึงเฉพาะของตัวอย่าง)

ก.วิธีตรง

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

ข. วิธีลัด

1. $$S = \sqrt{\frac{\sum_{i=1}^{k} f_i X_i^2 - \frac{\left(\sum_{i=1}^{k} f_i X_i\right)^2}{n}}{n-1}}$$

2. $$S = i\sqrt{\frac{\sum_{i=1}^{k} f_i d_i^2 - \frac{\left(\sum_{i=1}^{k} f_i d_i\right)^2}{n}}{n-1}}$$

ตัวอย่าง

จากข้อมูลที่จัดเป็นกลุ่มแล้วข้างต้น

จำนวนน้ำฝน (นิ้ว)	f_i	X_i	d_i	d_i^2	$f_i d_i$	$f_i d_i^2$
3-5	3	4	−2	4	−6	12
6-8	5	7	−1	1	−5	5
9-11	10	10	0	0	0	0
12-14	7	13	1	1	7	7
รวม	25				−4	24

ซึ่งเราคำนวณหา S^2 ได้ = 8.76

$\therefore$ ส่วนเบี่ยงเบนมาตรฐาน (S) $= \sqrt{8.76}$

$= 2.959$

$= 2.96$

5.6 คุณสมบัติของความแปรปรวนและส่วนเบี่ยงเบนมาตรฐาน

ข้อมูลชุดหนึ่ง $X_1, X_2, \ldots, X_n$ ซึ่งมีส่วนเบี่ยงเบนมาตรฐานเป็น σ_x และความแปรปรวนเป็น σ_x^2

1. เมื่อเอาค่าคงที่ a ใด ๆ ไปบวกหรือลบกับค่าของข้อมูลทุกตัว ความแปรปรวน หรือส่วนเบี่ยงเบนมาตรฐานที่ได้จะมีค่าเท่าเดิม คือ

$$\sigma_x^2 \pm a = \sigma_x^2$$

หรือ $\sigma_x \pm a = \sigma_x$

2. เมื่อเอาค่าคงที่ a ใด ๆ ไปคูณหรือหารค่าของข้อมูลทุกตัว ความแปรปรวน หรือส่วนเบี่ยงเบนมาตรฐานที่ได้จะมีค่าเปลี่ยนแปลงไปจากเดิม คือ

$$\sigma_{ax}^2 = a^2\sigma_x^2$$

$$\sigma_{\frac{x}{a}}^2 = \frac{1}{a^2}\sigma_x^2$$

หรือ $\sigma_{ax} = |a|\sigma_x$

$$\sigma_{\frac{a}{x}} = \frac{1}{|a|}\sigma_x$$

แบบฝึกหัด

1. ค่าสัมบูรณ์ของส่วนเบี่ยงเบนเฉลี่ยเป็นลบหรือเป็นศูนย์ได้หรือไม่จงอธิบาย
2. ค่าของส่วนเบี่ยงเบนมาตรฐานจะเป็นศูนย์ได้หรือไม่จงอธิบาย

   และเป็นค่าติดลบได้หรือไม่จงอธิบาย
3. จงหาพิสัยของข้อมูลในแบบฝึกหัดท้ายบทที่ 3 ข้อ 2
4. มีเงื่อนไขอะไรบ้างที่ทำให้ค่าของความแปรปรวนของตัวแปรมีค่าเท่ากับศูนย์
5. จงหาส่วนเบี่ยงเบนมาตรฐานของรายได้จากการขายสินค้าต่อวันดังนี้

   8100, 9000, 4580, 5600, 7680, 4800, 10640
6. จากข้อมูลต่อไปนี้ 2.1, 2.5, 2.7, 2.3, 2.4, 2.0, 2.7, 3.0, 1.4,

   2.4, 2.8, จงหา

   ก. พิสัย

   ข. ส่วนเบี่ยงเบนเฉลี่ย

   ค. ส่วนเบี่ยงเบนควอร์ไทล์

   ง. พิสัยเดไซล์

   จ. ความแปรปรวนและส่วนเบี่ยงเบนมาตรฐาน
7. ถ้าข้อมูลชุดหนึ่งมีค่ามัชฌิมเลขคณิตเท่ากับ 200 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 10

   จงคำนวณส่วนเบี่ยงเบนมาตรฐานของ y เมื่อ

   ก. $y = x/10$

   ข. $y = 5x$

   ค. $y = 7x+3$
8. จากแบบฝึกหัดท้ายบทที่ 3 ข้อ 6 จงหา

   ก. พิสัย

   ข. ส่วนเบี่ยงเบนเฉลี่ย

   ค. ความแปรปรวน

   ง. ส่วนเบี่ยงเบนมาตรฐาน

9. จากแบบฝึกหัดท้ายบทที่ 3 ข้อ 7 จงหา

ก. ส่วนเบี่ยงเบนเฉลี่ย

ข. ความแปรวน

ค. ส่วนเบี่ยงเบนมาตรฐาน

10. จากแบบฝึกหัดท้ายบทที่ 3 ข้อ 3 จงหาความแปรปรวนและส่วนเบี่ยงเบนมาตรฐาน

11. จากแบบฝึกหัดท้ายบทที่ 3 ข้อ 4 จงหา

ก. พิสัย

ข. ส่วนเบี่ยงเบนเฉลี่ย

ค. ส่วนเบี่ยงเบนมาตรฐาน

12. จงหาพิสัยของความสูง (นิ้ว) ของคนงานหญิง 50 คนที่ได้ในแบบฝึกหัดท้ายบทที่ 3 ข้อ 5

13. จงคำนวณค่า S^2 และ S ของข้อมูล 3, 4, 5, 6, 7

ก. ถ้าบวกข้อมูลแต่ละตัวด้วย 2 จงหา S^2 และ S จงพิจารณาดูว่าผลจะแตกต่างกันหรือไม่ถ้าเรานำเอาเลขที่มีค่ามาก ๆ เช่น 200 มาบวกกับข้อมูลแต่ละตัว

ข. ถ้านำข้อมูลแต่ละตัวมาลบออกจาก 2 จงหา S^2 และ S และให้พิจารณาดูว่าถ้าเรานำมาลบด้วย 200 ผลจะแตกต่างกันหรือไม่

ค. ถ้านำข้อมูลแต่ละตัวมาคูณด้วย 2 จงหา S^2 และ S และจงสรุปผลที่ได้

ง. ถ้านำข้อมูลแต่ละตัวมาหารด้วย 2 จงหา S^2 และ S และจงสรุปผลที่ได้

14. จงคำนวณหาส่วนเบี่ยงเบนมาตรฐานของข้อมูลดังต่อไปนี้

ก. 10, 8, 6, 0, 8, 3, 2, 2, 8, 0

ข. 1, 3, 3, 5, 5, 5, 7, 7, 9

ค. 20, 1, 2, 5, 4, 4, 4, 0

ง. 5, 5, 5, 5, 5, 5, 5, 5, 5, 5

15. ทำไมส่วนเบี่ยงเบนมาตรฐานที่คำนวณได้ในข้อ 14 (ค) จึงมีค่ามาก
จงอธิบาย