บทที่ 4
การกระจาย
(Variation Or Dispersion)

4.1 คำนำ

มัชฌิมเลขคณิต มัธยฐานและมัชฌิมอื่น ๆ เป็นลักษณะตัวแทนของข้อมูลกลุ่มหนึ่งเพียงลักษณะเดียว ส่วนลักษณะอื่น ๆ ของข้อมูล เช่น ความแตกต่างไม่อาจแสดงได้ด้วยมัชฌิม ลักษณะของข้อมูลที่มัชฌิมไม่อาจแสดงได้นี้ มีความสำคัญไม่น้อยกว่าความโน้มเอียงเข้าสู่ส่วนกลาง ดังตัวอย่างในการตัดสินว่าควรจะส่งนักกีฬาคนใดไปแข่งขันต่างประเทศ เราก็อาจคัดโดยพิจารณาเลือกผู้เล่นอยู่ในระดับดีปานกลางและสม่ำเสมอมากกว่าผู้ที่เล่นได้ดีมากเป็นครั้งคราว หรือเราต้องการเปรียบเทียบพันธบัตรรัฐบาล 2 ชนิด ชนิด x และชนิด y ในช่วง 6 ปีที่ผ่านมา พันธบัตรชนิด x ให้ดอกเบี้ยสำหรับแต่ละปี คิดเป็นเปอร์เซ็นต์ได้

6.0, 5.7, 5.6, 5.9, 6.1 และ 5.5 เปอร์เซ็นต์

ในช่วงเวลาเดียวกัน พันธบัตรชนิด y ให้

7.2, 7.7, 4.9, 3.1, 3.4 และ 8.5 เปอร์เซ็นต์

เราจะเห็นได้ว่าดอกเบี้ยเฉลี่ยของพันธบัตร 2 ชนิดให้ค่าออกมาเหมือนกันคือ 5.8 เปอร์เซ็นต์ เราก็อาจเชื่อว่าใน 6 ปีที่ผ่านมาพันธบัตรทั้งสองชนิดให้ดอกเบี้ยดีเหมือนกัน

แต่ถ้าเราวิเคราะห์ให้ละเอียดและระมัดระวังขึ้น จะพบว่าพันธบัตรชนิด x ให้ดอกเบี้ยแต่ละปีใกล้เคียงกับดอกเบี้ยเฉลี่ย มีการเปลี่ยนแปลงเล็กน้อย ไม่เหมือนกับพันธบัตรชนิด y ซึ่งมีความยืดหยุ่นมากจากดอกเบี้ย 3.1 เปอร์เซ็นต์ในปีที่สี่มาเป็น 8.5 เปอร์เซ็นต์ในปีที่หก จากเหตุผลนี้เอง สามารถทำให้เราตัดสินได้ว่าพันธบัตรชนิด x จะต้องดีกว่าชนิด y

จากสาเหตุนี้ ในการเปรียบเทียบข้อมูลกลุ่มหนึ่งกับอีกกลุ่มหนึ่งก็ดี จำเป็นต้องพิจารณาถึงการกระจายควบคู่ไปกับค่าเฉลี่ยของข้อมูลเสมอไป การวัดการกระจายอาจวัดเป็นหน่วยเดียวกันกับค่าของข้อมูล จัดว่าเป็นการกระจายสมบูรณ์ ส่วนการวัดเป็นตัวเลขที่ปราศจากหน่วย เช่นเป็นเปอร์-เซ็นต์ จัดว่าเป็นการกระจายสัมพัทธ์

4.2 การกระจายสมบูรณ์ (Absolute variability)

4.2.1 พิสัย (Range)

การวัดการกระจายที่ง่ายที่สุดและรวดเร็วที่สุดได้แก่แบบพิสัย ค่าของพิสัยก็คือ ผลต่างระหว่างค่าสูงสุดกับค่าต่ำสุดของข้อมูลในกลุ่ม ดังตัวอย่าง จากตารางที่ 2.1 ค่าของพิสัยของคะแนนสอบนักศึกษา 50 คน ได้เท่ากับ 98-33 = 65 คะแนน

การวัดการกระจายแบบพิสัย ใช้ค่าเพียงสองค่าเท่านั้น ไม่มีประโยชน์เท่าที่ควร จริงอยู่อาจจะใช้วัดในขั้นปฐมภูมิเท่านั้นเพื่อความรวดเร็วแต่ได้ค่าที่ไม่ถูกต้อง เพราะค่าของพิสัยที่ได้นี้ไม่ได้บอกถึงลักษณะของการกระจายข้อมูลส่วนที่เหลือ ดังนั้นเรามาพิจารณาถึงข้อมูล 3 กลุ่ม เช่น

5, 17, 17, 17, 17, 17, 17, 17, 17, 17,

5, 5, 5, 5, 5, 17, 17, 17, 17, 17,

5, 6, 8, 10, 11, 14, 14, 15, 16, 17,

เราจะเห็นได้ว่าในแต่ละกลุ่มก็มีพิสัยเท่ากับ 12 แต่การกระจายของจำนวนข้อมูลไม่เหมือนกัน กลุ่มแรกมีข้อมูลตัวแรกเท่านั้นที่มีค่าเท่ากับ 5 นอกนั้นมีค่าเท่ากับ 17 ส่วนในกลุ่มที่ 3 จำนวนข้อมูลแต่ละจำนวนกระจายไปจาก 5 ถึง 17

4.2.2 ส่วนเบี่ยงเบนเฉลี่ย (Average deviation)

เป็นวิธีการจัดการกระจายของข้อมูลรอบ ๆ มัชฌิม การกระจายจะน้อยถ้าข้อมูลอยู่ใกล้รอบ ๆ มัชฌิม และจะมากถ้าข้อมูลอยู่ห่างจากมัชฌิมมาก คำจำกัดความส่วนเบี่ยงเบนเฉลี่ยก็คือ มัชฌิมแบบเลขคณิตของระยะข้อมูลทุกตัวในกลุ่มเบี่ยงเบนไปจากค่ามัชฌิมของข้อมูลกลุ่มนั้นโดยไม่คำนึงถึงเครื่องหมาย ถ้าเรามีข้อมูลอยู่กลุ่ม 1 $x_1, x_2, x_3 \dots\dots\dots x_N$ ซึ่งมีมัชฌิมเลขคณิต μ ส่วนเบี่ยงเบนเฉลี่ยก็จะคำนวณได้จากสูตร

$$\text{A.D.} = \frac{\sum_{i=1}^{N} |x_i - \mu|}{N} \quad \text{(ข้อมูลที่ไม่มีมีการแจกแจงความถี่)} \qquad \dots\dots\dots(4.2.2.1)$$

ในเมื่อ $|x_i - \mu|$ คือ (Absolute value) ของ x_i เบี่ยงเบนไปจาก μ (Absolute value เป็นค่าที่ไม่คำนึงถึงเครื่องหมาย)

ตัวอย่าง 4.2.2.1 จงคำนวนหาส่วนเบี่ยงเบนเฉลี่ยของเลข 5 จำนวน 2, 3, 6, 8, 11

$$\text{มัชฌิมเลขคณิต } \mu = \frac{2+3+6+8+11}{5} = 6$$

ส่วนเบี่ยงเบนเฉลี่ย A.D. $= \dfrac{|2-6|+|3-6|+|6-6|+|8-6|+|11-6|}{5}$

$$= \frac{|-4|+|-3|+|0|+|2|+|5|}{5}$$

$$= \frac{4+3+0+2+5}{5} = 2.8$$

ในการคำนวณหาส่วนเบี่ยงเบนเฉลี่ยตามที่กล่าวมานี้จะทำได้เร็วขึ้น ถ้าแบ่งข้อมูลในกลุ่มออกเป็น 2 พวก พวกหนึ่งมีค่าสูงกว่ามัชฌิมและอีกพวกหนึ่งต่ำกว่ามัชฌิม หาผลรวมของค่าข้อมูลในแต่ละพวก เพื่อคำนวณค่าของ A.D. จากสูตร

$$A.D. = \frac{A-B-(a-b)\mu}{N}$$

ในเมื่อ A = ผลรวมของค่าของข้อมูลในพวกที่สูงกว่า μ

B = ผลรวมของค่าของข้อมูลในพวกที่ต่ำกว่า μ

a = จำนวนข้อมูลในพวกแรก

b = จำนวนข้อมูลในพวกหลัง

ตามตัวอย่างข้างต้น

$A = 8+11 = 19$

$B = 2+3 = 5$

$a = 2$

$b = 2$

แทนค่าลงในสูตรได้

$$A.D. = \frac{19-5-(2-2)6}{5} = \frac{19-5-0}{5}$$

$$= \frac{14}{5} = 2.8$$

ในกรณีที่ข้อมูลมีการแจกแจงความถี่สูตรที่ใช้

$$A.D. = \frac{\sum_{i=1}^{k} f_i|x_i - \mu|}{\sum_{i=1}^{k} f_i \text{ หรือ } N} \qquad \text{..........(4.2.2.2)}$$

ตัวอย่าง 4.2.2.2 จากตารางที่ 2.2 ให้หาส่วนเบี่ยงเบนเฉลี่ย (μ = 65.1 เป็นค่าที่คำนวณเรียบร้อยแล้วในหัวข้อมัชฌิมเลขคณิต)

ชั้นคะแนน	f	จุดกลาง	$\lvert X - \mu \rvert$	$f\lvert X - \mu \rvert$
30 – 39	4	34.5	30.6	122.4
40 – 49	6	44.5	20.6	123.6
50 – 59	8	54.5	10.6	84.8
60 – 69	12	64.5	.6	7.2
70 – 79	9	74.5	9.4	84.6
80 – 89	7	84.5	19.4	135.8
90 – 99	4	94.5	29.4	117.6
	50			676.0

$$\text{เพราะฉะนั้น A.D.} = \frac{676}{50}$$

$$= 13.52 \text{ คะแนน}$$

การหาส่วนเบี่ยงเบนเฉลี่ยของข้อมูล นอกจากจะหาโดยใช้ส่วนเบี่ยงเบนที่วัดจาก μ ก็อาจใช้ส่วนเบี่ยงเบนที่วัดจากค่ามัชฌิมอื่น ๆ เช่น มัธยฐาน หรือ ฐานนิยมได้เช่นกัน

4.2.3 ส่วนเบี่ยงเบนควอร์ไทล์ (Quartile Deviation)

การวัดการกระจายอีกแบบหนึ่ง ก็คือ ส่วนเบี่ยงเบนควอร์ไทล์ซึ่งคำนวณหาได้จากผลต่างระหว่าง ควอร์ไทล์ที่ 3 กับควอร์ไทล์ที่ 1 (เรียกว่าควอร์ไทล์เร้นส์) หารด้วย 2 เขียนเป็นสูตรได้

$$\text{Q.D.} = \frac{Q_3 - Q_1}{2} \qquad \text{..........(4.2.3.1)}$$

ตัวอย่าง 4.2.3.1 หาส่วนเบี่ยงเบนควอร์ไทล์ของคะแนนนักศึกษา 50 คน (ในตารางที่ 3.4)

ในเมื่อ $Q_1 = 52.625$, $Q_3 = 77.822$ (เอาค่าของ Q_1, Q_3 มาจากหัวข้อควอร์ไทล์)

แทนค่า Q_1, Q_3 ลงในสูตรข้างต้นก็จะได้ส่วนเบี่ยงเบนควอร์ไทล์

$$\text{Q.D.} = \frac{77.822 - 52.625}{2}$$

$$= \frac{25.197}{2} = 12.598 \text{ คะแนน}$$

ถ้าให้ M เป็นจุดกึ่งกลางระหว่างควอร์ไทล์ที่ 1 และที่ 3 ($M = \frac{1}{2}(Q_1+Q_3)$) จะเห็นได้ว่าข้อมูลครึ่งหนึ่งจะมีค่าอยู่ระหว่าง M − Q.D. และ M + Q.D. หรือ M ± Q.D. ตัวอย่างเช่นจุดกึ่งกลาง M ในที่นี้มีค่าเท่า (77.822 + 52.625)/2 = 65.224 คะแนน เพราะฉะนั้นกล่าวได้ว่านักเรียนชายในกลุ่มนี้ จำนวน 25 คน มีคะแนนอยู่ระหว่าง 65.224 ± 12.598 คะแนน

จะเห็นได้ว่า การวัดการกระจายแบบส่วนเบี่ยงเบนควอร์ไทล์ ก็ใช้ค่าของมูลในกลุ่มเพียง 2 ค่า เช่นเดียวกับพิสัย จึงไม่เป็นที่นิยมกันมาก

4.2.4 ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation)

การวัดการกระจายแบบส่วนเบี่ยงเบนมาตรฐานเป็นมาตรการวัดการกระจายที่ดีที่สุดและใช้มากที่สุดในทางสถิติ มีประโยชน์ในการที่จะอนุมานค่าบางอย่างของข้อมูลประชากรจากตัวอย่าง

ส่วนเบี่ยงเบนมาตรฐาน เป็นการวัดส่วนเบี่ยงเบนของข้อมูลแต่ละตัวจากมัชฌิมเลขคณิตเหมือนกับส่วนเบี่ยงเบนเฉลี่ย ต่างกันก็แต่เราไม่ต้องกังวลตัดเครื่องหมายทิ้งก่อนที่จะหาผลบวกของส่วนเบี่ยงเบน (deviations) เพราะเมื่อนำเอาค่าลบมายกกำลังสองก็กลายเป็นบวก เหตุผลที่ว่าทำไมส่วนเบี่ยงเบนมาตรฐานจึงเป็นที่นิยมมากกว่าส่วนเบี่ยงเบนเฉลี่ย นั่นเป็นคุณสมบัติทางคณิตศาสตร์ที่จะได้กล่าวในบทต่อไป

ส่วนเบี่ยงเบน (Deviation)

ส่วนเบี่ยงเบน หมายถึงผลต่างระหว่างค่าของข้อมูลแต่ละข้อมูล X_i กับมัชฌิม μ นั่นคือ ส่วนเบี่ยงเบน $D = X_i - \mu$

ความแปรปรวน (Variance)

ค่าความแปรปรวน หมายถึงค่าของมัชฌิมเลขคณิตของกำลังสองของผลต่างระหว่างค่าของข้อมูลแต่ละค่า กับค่ามัชฌิมเลขคณิต ดังแสดงตามสูตร

$$\text{ความแปรปรวนของ } X = \text{Var}(X) = \frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N} \qquad \text{..........(4.2.4.1)}$$

เราอาจจะใช้สัญลักษณ์เป็นตัวอักษรกรีก σ (เรียกว่าซิกม่า) ยกกำลังสองใช้แสดงความแปรปรวนในสูตรได้ดังนี้

$$\text{Var}(X) = \sigma^2 \qquad \text{..........(4.2.4.2)}$$

ส่วนปริมาณ $\sum_{i=1}^{N} (X_i - \mu)^2$ ซึ่งนักศึกษาจะต้องพบเห็นในบทต่อไปอีก เรียกว่าผลบวกกำลังสองของส่วนเบี่ยงเบน (sum of the squared deviations)

ค่าของความแปรปรวนใช้สำหรับวัดค่าการกระจายของกลุ่มข้อมูล ซึ่งมีหน่วยเป็นกำลังสองของหน่วยข้อมูลนั้น เพื่อให้มีหน่วยถูกต้องเป็นไปตามหน่วยข้อมูล จึงจำเป็นต้องถอดรู้ทของความแปรปรวน รู้ทของความแปรปรวนเรียกว่าส่วนเบี่ยงเบนมาตรฐาน

ตัวอย่าง 4.2.4.1 จงหาความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานของจำนวนเลขต่อไปนี้

12, 6, 7, 3, 15, 10, 18, 5

วิธีทำ เราหาค่ามัชฌิมเลขคณิตก่อนได้

$$\mu = \frac{12 + 6 + 7 + 3 + 15 + 10 + 18 + 5}{8} = 9.5$$

$$\text{ความแปรปรวน} = \sigma^2 = \frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N}$$

$$= \frac{(12-9.5)^2 + (6-9.5)^2 + (7-9.5)^2 + (3-9.5)^2 + (15-9.5)^2 + (10-9.5)^2 + (18-9.5)^2 + (5-9.5)^2}{8}$$

$$= \frac{(2.5)^2 + (-3.5)^2 + (-2.5)^2 + (-6.5)^2 + (5.5)^2 + (.5)^2 + (8.5)^2 + (-4.5)^2}{8}$$

$$= \frac{190}{8} = 23.75$$

ส่วนเบี่ยงเบนมาตรฐาน $\sigma = \sqrt{23.75} = 4.87$

(หมายเหตุ : ประชากรกลุ่มใดที่มีค่าความแปรปรวนหรือส่วนเบี่ยงเบนมาตรฐานมากย่อมมีการกระจายมากเมื่อเปรียบเทียบกับประชากรกลุ่มที่มีค่าความแปรปรวนหรือส่วนเบี่ยงเบนน้อย ดูตัวอย่างที่ 4.2.4.2)

ตัวอย่างที่ 4.2.4.2 จากตารางที่ 4.1 มีประชากรอยู่ 4 กลุ่ม ซึ่งมีมัชฌิมเลขคณิตเท่ากันแต่ค่าของความแปรปรวนหรือส่วนเบี่ยงเบนไม่เท่ากัน

ตารางที่ 4.1

ประชากรที่มีมัชฌิมเลขคณิตเท่ากันแต่ส่วนเบี่ยงเบนมาตรฐานต่างกัน

ประชากร	ก	ข	ค	ง
ข้อสังเกตหรือข้อมูล	4	2	1	0
	4	3	2	1
	4	7	9	11
ผลบวก	12	12	12	12
ส่วนเฉลี่ย	4	4	4	4
ความแปรปรวน	0	$\frac{14}{3}$	$\frac{38}{3}$	$\frac{74}{3}$
ส่วนเบี่ยงเบนมาตรฐาน	0	$\sqrt{\frac{14}{3}} = 2.16$	$\sqrt{\frac{38}{3}} = 3.55$	$\sqrt{\frac{74}{3}} = 4.96$

ก.) $\sigma^2 = \frac{(4-4)^2 + (4-4)^2 + (4-4)^2}{3}$

$= 0$

ข.) $\sigma^2 = \frac{(2-4)^2 + (3-4)^2 + (7-4)^2}{3}$

$= \frac{4+1+9}{3} = \frac{14}{3}$

ค.) $\sigma^2 = \dfrac{(1-4)^2 + (2-4)^2 + (9-4)^2}{3}$

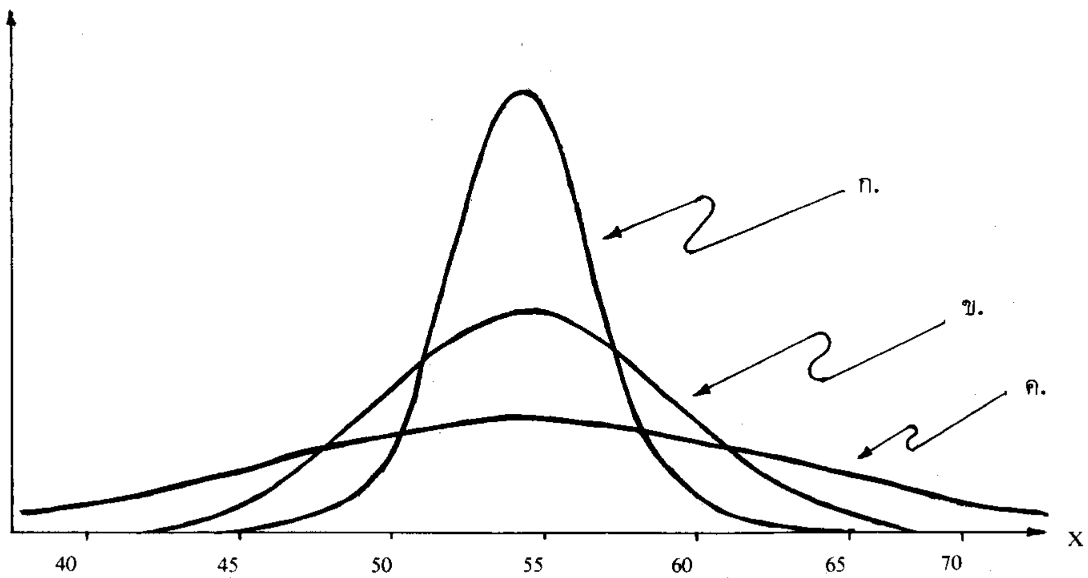
$= \dfrac{9+4+25}{3} = \dfrac{38}{3}$

ง.) $\sigma^2 = \dfrac{(0-4)^2 + (1-4)^2 + (11-4)^2}{3}$

$= \dfrac{74}{3}$

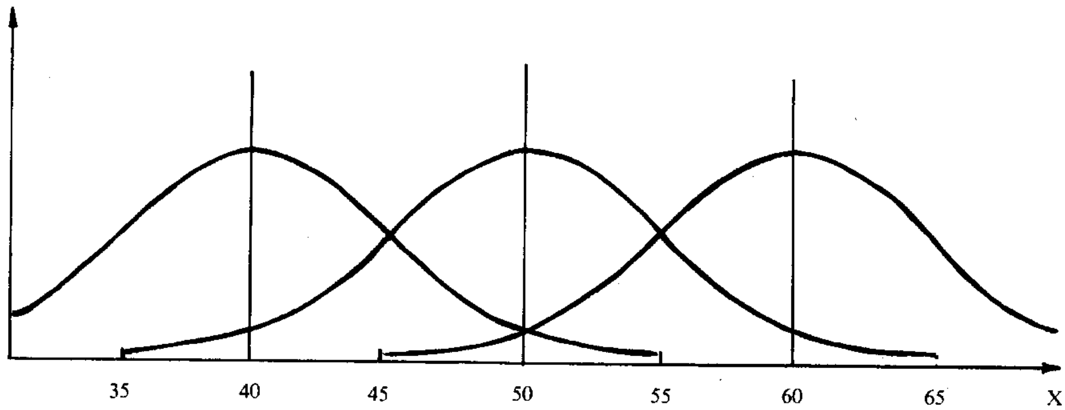
จากตารางข้างต้นประชากร ก. ไม่มีค่าความแปรปรวนหรือส่วนเบี่ยงเบนมาตรฐาน ประชากร ข. มีค่าความแปรปรวนมากขึ้น ประชากร ง. มีค่าความแปรปรวนมากที่สุด

รูปที่ 4.1 ในกรณีการแจกแจงปกติซึ่งมีมัชฌิมเลขคณิตเท่ากันแต่ส่วนเบี่ยงเบนมาตรฐานไม่เท่ากัน



รูปที่ 4.1

เส้นโค้งปกติที่มีส่วนเบี่ยงเบนมาตรฐานต่างกัน แต่มัชฌิมเท่ากัน รูปที่ 4.1



รูปที่ 4.2

รูปที่ 4.2 แสดงการแจกแจงปกติซึ่งมีมัชฌิมต่างกันแต่ส่วนเบี่ยงเบนมาตรฐานเหมือนกัน เส้นโค้งปกติ 3 เส้นที่มีมัชฌิมต่างกันแต่ส่วนเบี่ยงเบนมาตรฐานเหมือนกัน

ในกรณีที่ข้อมูลมีการแจกแจงความถี่ ค่าของความแปรปรวนก็คำนวณได้จากสูตร

$$\mathrm{Var}(X) = \sigma^2 = \frac{\sum_{i=1}^{K} f_i(X_i - \mu)^2}{N} \qquad \text{..........(4.2.4.3)}$$

ในเมื่อ X_i เป็นค่าของจุดกลางของชั้น, μ เป็นค่ามัชฌิม f_i เป็นความถี่ของแต่ละชั้น, N เป็นจำนวนความถี่ทั้งหมด $\left(\sum_{i=1}^{K} f_i = n \right)$; และ K จำนวนชั้น

ถอดรู้ทของความแปรปรวนก็จะได้ส่วนเบี่ยงเบนมาตรฐาน

$$\sigma = \sqrt{\frac{\sum_{i=1}^{K} f_i (X_i - \mu)^2}{N}} \qquad \text{..........(4.2.4.4)}$$

ตัวอย่างที่ 4.2.4.3 หาส่วนเบี่ยงเบนมาตรฐานของนักศึกษา 50 คน (μ = 65.1 ได้มาจากตัวอย่างของบทที่ 3)

ชั้นคะแนน	f	จุดกึ่งกลาง X	$(X-\mu)$	$(X-\mu)^2$	$f(X-\mu)^2$
30 – 39	4	34.5	– 30.6	936.36	3745.44
40 – 49	6	44.5	– 20.6	424.36	2546.16
50 – 59	8	54.5	– 10.6	112.36	898.88
60 – 69	12	64.5	.6	.36	4.32
70 – 79	9	74.5	9.4	88.36	715.24
80 – 89	7	84.5	19.4	376.36	2634.52
90 – 99	4	94.5	29.4	864.36	3457.44
	50				14,082.00

แทนค่าลงในสูตร $\sigma = \sqrt{\frac{14082}{50}}$

$= \sqrt{281.64}$

$=$ 16.78 คะแนน

การหาส่วนเบี่ยงเบนข้างต้น เราจะต้องคำนวณมัชฌิมก่อน จึงคำนวณหาส่วนเบี่ยงเบนมาตรฐานได้ ตัวเลขที่จะต้องเอามาคำนวณมีมากอาจผิดพลาดได้ง่าย เพื่อหลีกเลี่ยงความยุ่งยากวิธีนี้ จึงมีอีกวิธีหนึ่งซึ่งสะดวกและง่ายกว่าดังคำนวณได้จากสูตร

$$\sigma = I\sqrt{\frac{\sum_{i=1}^{K} f_i d_i^2}{N} - \left[\frac{\sum_{i=1}^{K} f_i d_i}{N}\right]^2} \qquad \text{..........(4.2.4.5)}$$

ในเมื่อ I คืออันตรภาคชั้น, f_i คือความถี่ของชั้นที่ i

d_i คือส่วนเบี่ยงเบนของชั้นที่ i จากค่าที่สมมุติขึ้น

N คือจำนวนทั้งหมดของข้อมูล

ตัวอย่าง 4.2.4.4 จงคำนวณหาส่วนเบี่ยงเบนมาตรฐานของนักศึกษา 50 คน (จาก ตารางข้างต้น)

ชั้นคะแนน	f	d	fd	d^2	fd^2
30 – 39	4	–3	–12	9	36
40 – 49	6	–2	–12	4	24
50 – 59	8	–1	– 8	1	8
60 – 69	12	0	0	0	0
70 – 79	9	1	9	1	9
80 – 89	7	2	14	4	28
90 – 99	4	3	12	9	36

คำอธิบาย เริ่มแรกเราสมมุติให้ชั้น 60-69 เป็นจุดเริ่มมีค่าเท่ากับ 0 ชั้นที่อยู่ก่อนชั้น 60-69 เป็น -1, -2 และ -3 ส่วนชั้นที่อยู่ถัดจากจุดเริ่มก็เป็น 1, 2, 3, เหตุผลในการเลือกชั้น 60-69 ก็เพื่อสะดวกในการคำนวณ แต่ถ้าสังเกตเห็นว่าการคำนวณจะง่ายขึ้นเมื่อให้จุดเริ่มต้นอยู่ที่ชั้น 70-79 เราก็เลือกชั้นนั้น

แทนค่าซึ่งคำนวณได้จากตารางลงในสูตรข้างต้นได้

$$\sigma = I\sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} = 10\sqrt{\frac{141}{50} - \left(\frac{3}{50}\right)^2}$$

$$\sigma = 10\sqrt{2.8164} = 16.78 \text{ คะแนน}$$

ความแปรปรวนผสม (Combined Variance)

ในกรณีที่เราต้องหาค่าของความแปรปรวนของข้อมูลในกลุ่มผสม เมื่อทราบแต่เพียงค่าของความแปรปรวนของข้อมูลในกลุ่มย่อย ๆ ได้ และมัชฌิมเลขคณิตของกลุ่มย่อย ๆ จะต้องเท่ากันด้วย

สูตรคำนวณหาความแปรปรวนร่วมหรือของข้อมูลในกลุ่มผสมได้ (สองกลุ่ม)

$$\sigma_{12}^2 = \frac{N_1\sigma_1^2 + N_2\sigma_2^2}{N_1 + N_2} \qquad \text{..........(4.2.4.6)}$$

ในเมื่อ σ_1^2 = ความแปรปรวนของกลุ่มแรก

N_1 = จำนวนของข้อมูลกลุ่มแรก

σ_2^2 = ความแปรปรวนของกลุ่มที่สอง

N_2 = จำนวนข้อมูลกลุ่มที่สอง

σ_{12}^2 = ความแปรปรวนผสม

ตัวอย่าง 4.2.4.5 ผลการสอบวิชาสถิติของนักศึกษาชั้นปีที่ 2 ปรากฏว่าคะแนนส่วนเฉลี่ยของนักศึกษาหญิง 400 คน เท่ากับคะแนนส่วนเฉลี่ยของนักศึกษาชาย 600 คน แต่ส่วนเบี่ยงเบนมาตรฐานของนักศึกษาหญิงและชายมีค่าเท่ากับ 1 และ 1.5 ตามลำดับจงคำนวณหาส่วนเบี่ยงเบนมาตรฐานของนักศึกษาชั้นปีที่สอง

วิธีทำ ให้ μ_1 = มัชฌิมของนักศึกษาหญิง

μ_2 = มัชฌิมของนักศึกษาชาย

N_1 = จำนวนนักศึกษาหญิง = 400 คน

σ_1 = ส่วนเบี่ยงเบนมาตรฐานของนักศึกษาหญิง = 1

σ_2 = ส่วนเบี่ยงเบนมาตรฐานของนักศึกษาชาย เท่ากับ 1.5

N_2 = จำนวนนักศึกษาชายเท่ากับ 600 คน

แต่ $\mu_1 = \mu_2$

$$\text{ความแปรปรวนผสม} = \sigma_{12}^2 = \frac{400 \times (1)^2 + 600\,(1.5)^2}{400 + 600} = \frac{400 + (600 \times 2.25)}{1000}$$

$$= 1.750\ (\text{คะแนน})^2$$

ส่วนเบี่ยงเบนมาตรฐานผสมนักศึกษาชั้นปีที่ 2 = $\sqrt{1.75}$ = 1.32 คะแนน

4.3 คุณสมบัติของพิสัย, ส่วนเบี่ยงเบนควอร์ไทล์ ส่วนเบี่ยงเบนมาตรฐาน

คุณสมบัติของพิสัย

พิสัยเป็นมาตราวัดการกระจายที่ง่ายและหยาบ ในทุก ๆ สภาวะไม่ว่าจะเป็นตัวอย่างที่เลือกจากการแจกแจงปกติ พิสัยจะแปรมากกว่ามาตราวัดอื่น ๆ สำหรับการสุ่มตัวอย่างที่มีการเปลี่ยนแปลง เพราะเป็นค่าที่คำนวณได้จากค่าผลต่างระหว่างค่าสูงสุดกับต่ำสุดของการแจกแจงเท่านั้น พิสัยขึ้นอยู่กับขนาดตัวอย่าง พิสัยจะมีค่ามากขึ้นต่อเมื่อขนาดตัวอย่างใหญ่ขึ้น พิสัยคำนวณหาได้ง่ายกว่ามาตราวัดการกระจายอื่น ๆ แต่ความแม่นยำน้อยกว่า

คุณสมบัติของส่วนเบี่ยงเบนควอร์ไทล์

เป็นระยะทางครึ่งหนึ่งระหว่าง Q_1 กับ Q_3 หรือระยะทางเฉลี่ยระหว่างมัธยฐานกับ Q_1 กับ Q_3 ส่วนเบี่ยงเบนควอร์ไทล์มีความสัมพันธ์ใกล้ชิดกับมัธยฐานมาก เนื่องจากว่าทั้งสองค่านิยามในรูปของควอร์ไทล์ของการแจกแจง มัธยฐานมีผลตอบสนองต่อจำนวนคะแนนหรือค่าสังเกตที่วางอยู่ล่างมัธยฐานค่อนข้างมากกว่าตำแหน่งที่แน่นอนของคะแนน Q_1 และ Q_3 ก็เป็นค่าที่ถูกนิยามในหนทางเดียวกัน เพราะฉะนั้นเราอาจคาดหวังได้ว่ามัธยฐานและส่วนเบี่ยงเบนควอร์ไทล์มีคุณสมบัติร่วมกันได้ ตัวอย่างเช่น Q_3 มีความไวต่อจำนวนค่าสังเกตที่มีค่ามากกว่า Q_3 เท่านั้น เพราะฉะนั้นค่าสังเกตทั้งหมดที่มีค่ามากกว่า Q_3 มีความสำคัญเท่ากันในการกำหนดส่วนเบี่ยงเบนควอร์ไทล์ ค่าสังเกตที่มีค่าน้อย ๆ และมาก ๆ จะไม่นับรวมเข้าด้วยในการคำนวณ ดังนั้น ส่วนเบี่ยงเบนควอร์ไทล์จึงมีความไวต่อค่าสังเกตที่มีค่าน้อย ๆ และมาก ๆ น้อยกว่าส่วนเบี่ยงเบนมาตรฐาน ถ้าหากว่าการแจกแจงหนึ่งที่เบ้ไม่ค่อยดี หรือมีค่าสังเกตจำนวนน้อยที่อยู่ปลายสองข้างมาก ๆ แล้ว ส่วนเบี่ยงเบนควอร์ไทล์จะสนองตอบต่อค่าสังเกตเช่นนั้น แต่จะไม่ให้ค่าสังเกตเหล่านี้ถูกถ่วงน้ำหนักเกินควร

ถ้าหากว่าการแจกแจงเป็นแบบปลายเปิด การคำนวณหาส่วนเบี่ยงเบนมาตรฐานและพิสัยไม่ได้นอกจากมีเงื่อนไขเพิ่มเติม แต่ส่วนเบี่ยงเบนควอร์ไทล์เป็นมาตราวัดที่สมเหตุสมผลในกรณีนี้

ส่วนเบี่ยงเบนควอร์ไทล์เหมาะสมดีสำหรับการสุ่มตัวอย่างที่ไม่มีการเปลี่ยนแปลง แต่ก็ไม่ดีเท่ากับส่วนเบี่ยงเบนมาตรฐาน

การคำนวณหาได้ง่าย ถ้าหากว่าจัดลำดับค่าสังเกตหรือจัดเป็นกลุ่มในรูปอันตรภาคชั้น

คุณสมบัติของส่วนเบี่ยงเบนมาตรฐาน

ส่วนเบี่ยงเบนมาตรฐานมีคุณสมบัติคล้ายกับมัชฌิมเลขคณิต คือ มีผลสนองตอบต่อตำแหน่งที่แน่นอนของทุก ๆ ค่าสังเกตของการแจกแจงเมื่อไรค่าสังเกตเปลี่ยนตำแหน่งไปจากมัชฌิมเลขคณิตมากขึ้น ส่วนเบี่ยงเบนมาตรฐานจะมีค่ามากขึ้น ถ้าหากว่าการเปลี่ยนตำแหน่งเข้าหามัชฌิมเลขคณิตมากขึ้น ขนาดของส่วนเบี่ยงเบนมาตรฐานจะลดลง เพราะฉะนั้นส่วนเบี่ยงเบนมาตรฐานจึงมีความไวต่อเงื่อนไขที่แน่นอนของการแจกแจงมากกว่ามาตราวัดอื่น ๆ ดังตัวอย่าง พิจารณาเจ็ดค่าสังเกตดังนี้ 2, 5, 7, 8, 9, 11, 14 ตารางที่ 4.2 แสดงค่าส่วนเบี่ยงเบนมาตรฐาน (σ) ส่วนเบี่ยงเบนควอร์ไทล์ (QD) และพิสัย (R) สำหรับค่าสังเกตของกลุ่มนี้ สำหรับค่าสังเกตกลุ่มใหม่ (เพิ่มค่าสังเกตที่มีค่ามากที่สุด) เป็น 2, 5, 7, 8, 9, 11, 14, 24 ตารางที่ 4.2 แสดงมาตราวัดการกระจายแบบ

เดียวกันภายหลังได้ทำการเปลี่ยนแปลง ตัวเลขแสดงเปอร์เซนต์เพิ่มขึ้นของขนาดแต่ละมาตราวัดการกระจายภายหลังเพิ่มค่าสังเกตใหม่ สังเกตว่าส่วนเบี่ยงเบนมาตรฐานมีผลมากกว่าส่วนเบี่ยงเบนควอร์ไทล์ สัดส่วนของการเปลี่ยนแปลงควรจะแตกต่างจากการแจกแจงหนึ่งไปยังอีกการแจกแจง ส่วนเบี่ยงเบนมาตรฐานไม่อาจจะเป็นมาตราวัดการกระจายที่ดีที่สุดได้เมื่อไรการแจกแจงประกอบด้วยค่าสังเกตที่มีค่ามาก ๆ เพียงไม่กี่ค่า หรือเมื่อไรการแจกแจงเบ้ ดังตัวอย่าง ถ้าเราต้องการ

ตาราง 4.2 ผลเกี่ยวกับมาตราวัดการกระจายสำหรับเพิ่มค่าสังเกตที่มีค่ามากที่สุดของการแจกแจง

	ส่วนเบี่ยงเบนมาตรฐาน	ส่วนเบี่ยงเบนควอร์ไทล์	พิสัย
การแจกแจงเดิม (N = 7)	3.63	2.75	12
การแจกแจงพร้อมด้วยเพิ่มค่าสังเกต	(N = 8) 6.29	3.25	22
เปอร์เซนต์เพิ่มขึ้นจากการเพิ่มค่าสังเกต	73	18	83

เปรียบเทียบการกระจายของสองการแจกแจงในเมื่อการแจกแจงหนึ่งประกอบด้วยหลาย ๆ ค่าสังเกตที่มีค่ามากอีกการแจกแจงหนึ่งไม่มี ความแตกต่างในค่าสังเกตที่มีค่ามากควรจะมีอิทธิพลเกี่ยวกับขนาดของส่วนเบี่ยงเบนมาตรฐาน ซึ่งไม่เป็นสัดส่วนสัมพันธ์กับจำนวน แต่ถ้าหากว่า N ใหญ่พอและค่าสังเกตที่มีค่ามาก ๆ มีจำนวนน้อยมากก็จะทำให้ความแตกต่างกันไม่มาก

ความไวของส่วนเบี่ยงเบนมาตรฐานต่อค่าสังเกตที่อยู่ปลาย ๆ หรือห่างจากมัชฌิมเลขคณิตเปลี่ยนแปลงบางจุด ส่วนเบี่ยงเบนมาตรฐานจะสนองตอบมากกว่าการเปลี่ยนแปลงแบบเดียวกันของค่าสังเกตที่อยู่ใกล้ชิดมัชฌิมเลขคณิตกว่า เหตุผลพื้นฐานก็คือว่าส่วนเบี่ยงเบนมาตรฐานมีความสมนัยกับค่าส่วนเบี่ยงเบนกำลังสอง ค่าส่วนเบี่ยงเบนที่มีค่าน้อยกว่าเมื่อยกกำลังสองย่อมมีค่าน้อยกว่าค่าส่วนเบี่ยงเบนที่มีค่ามากกว่ายกกำลังสอง อย่างเช่น ค่าส่วนเบี่ยงเบนเปลี่ยนจาก +2 ไปเป็น +4 ความแตกต่างของกำลังสองของค่าเหล่านี้เป็น 12 ขณะเดียวกันกับค่าส่วนเบี่ยงเบนเปลี่ยนจาก +10 ไปเป็น +12 ความแตกต่างของกำลังสองของค่าเหล่านี้เป็น 44 อิทธิพลภายใต้การพิจารณานี้เป็นเพียงหัวข้อหนึ่งที่เห็นได้อย่างเด่นชัด

เมื่อไรที่ค่าส่วนเบี่ยงเบนคำนวณหาได้จากมัชฌิมเลขคณิต ผลบวกกำลังสองของค่าเหล่านี้มีค่าน้อยกว่าค่าส่วนเบี่ยงเบนที่ได้มารอบจุดอื่น ๆ นั่นคือ

$\Sigma (X - A)^2$ มีค่าน้อยที่สุด เมื่อ $A = \mu$

พิสูจน์

ให้ A เป็นจุดซึ่งได้ค่าส่วนเบี่ยงเบนของแต่ละค่าสังเกตและนิยามให้เป็นจุดซึ่งแตกต่างไปจากมัชฌิมเลขคณิตเป็นจำนวน d: $A = \mu + d$ แล้ว

$$\begin{aligned} \Sigma (X-A)^2 &= \Sigma [X - (\mu+d)]^2 \\ &= \Sigma [(X-\mu) - d]^2 \\ &= \Sigma [(X-\mu)^2 - 2d(X-\mu) + d^2] \\ &= \Sigma (X-\mu)^2 - 2d\,\Sigma (X-\mu) + Nd^2 \end{aligned}$$

เนื่องจากว่า $\Sigma(X-\mu) = 0$

$$\Sigma (X-A)^2 = \Sigma(X-\mu)^2 + Nd^2$$

พิจารณาทางด้านขวาของสมการแสดงว่า $\Sigma(X-A)^2$ จะมีค่าน้อยที่สุด เมื่อ $d = 0$ จากนิยามของ d จะเกิดขึ้นเมื่อ $A = \mu$ เนื่องจากว่า $\sigma = \sqrt{\Sigma(X-\mu)^2/N}$ แสดงว่า σ มีค่าน้อยที่สุดเมื่อไรที่ค่าส่วนเบี่ยงเบนได้มาจาก μ

ตัวอย่าง เปรียบเทียบ $\Sigma(X-A)^2$ เมื่อค่าส่วนเบี่ยงเบนได้มาจากมัชฌิมเลขคณิต ($\mu = 5$) กับเมื่อค่าส่วนเบี่ยงเบนได้มาจากจุดอื่น ๆ ($X = 6$)

คะแนน	$(X-5)$	$(X-5)^2$	คะแนน	$(X-6)$	$(X-6)^2$
8	+3	9	8	2	4
6	+1	1	6	0	0
4	−1	1	4	−1	1
2	−3	9	2	−4	16
		$\Sigma(X-\mu)^2 = 20$			$\Sigma(X-6)^2 = 24$
ค่าส่วนเบี่ยงเบนรอบ $\mu = 5$			ค่าส่วนเบี่ยงเบนรอบ $\mu = 6$		

ส่วนเบี่ยงเบนมาตรฐานจะขจัดการเปลี่ยนแปลงหรือความไม่แน่นอนของการสุ่มตัวอย่าง ตัวอย่างสุ่มที่ได้มาซ้ำ ๆ กันจากการเลือกมาจากประชากรเดียวกัน มูลค่าตัวเลขของค่าส่วนเบี่ยงเบนมาตรฐานมีแนวทางที่จะกระโดดโดยประมาณน้อยกว่ามาตราวัดแบบอื่น ๆ ในตัวอย่างเดียวกัน คุณสมบัติข้อนี้จะเข้าใจดียิ่งขึ้นก็ต่อเมื่อได้ทำการทดลองด้วยตัวของเราเองและได้ศึกษาวิธีการ

วิเคราะห์ทางสถิติ ค่าส่วนเบี่ยงเบนมาตรฐานจะปรากฏออกมาโดยแน่นอนหรือวางตรึงอยู่ในหลาย ๆ กระบวนการทั้งทางสถิติเชิงพรรณนาและอนุมาน โดยเฉพาะสถิติเชิงอนุมาน พิสัยและส่วนเบี่ยงเบนควอร์ไทล์ใช้น้อยมาก

คุณสมบัติของส่วนเบี่ยงเบนมาตรฐานมีความสัมพันธ์กับคุณสมบัติของมัชฌิมเลขคณิตในหลาย ๆ วิธีด้วยกัน เพื่อประโยชน์ในงานสถิติ

4.4 การใช้เครื่องหมายทางพีชคณิต

ในหัวข้อนี้เสนอการพิสูจน์ทางพีชคณิตสำหรับสมการต่าง ๆ ของบทนี้ การพิสูจน์เหล่านี้บางหัวข้ออาจปรากฏยุ่งยากต่อผู้อ่านมากกว่าการเสนอในหัวข้อการใช้เครื่องหมายทางพีชคณิตของบทก่อน ๆ แต่ความรู้พื้นฐานทางพีชคณิต และกฎว่าด้วยการบวกก็ได้กล่าวไว้ในบทก่อน ๆ คงจะเป็นพื้นความรู้เพียงพอสำหรับการพิสูจน์เหล่านี้

พิสูจน์

$$\sigma = \sqrt{\frac{\Sigma (X_i-\mu)^2}{N}} = \frac{1}{N}\sqrt{N \sum_{i=1}^{N} X_i^2 - \left(\sum_{i=1}^{N} X_i\right)^2}$$

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

$$= \frac{1}{N}\left[\sum_{i=1}^{N} (X^2 - 2X_i\,\mu + \mu^2)\right]$$

$$= \frac{1}{N}\left[\sum_{i=1}^{N} X_i^2 - \sum_{i=1}^{N} 2X_i\mu + \sum_{i=1}^{N} \mu^2\right]$$

$$= \frac{\sum_{i=1}^{N} X_i^2}{N} - 2\mu \frac{\sum_{i=1}^{N} X_i}{N} + \frac{N\mu^2}{N}$$

แต่ $\mu = \dfrac{\sum_{i=1}^{N} X_i}{N}$

เราได้

$$\sigma^2 = \frac{\sum_{i=1}^{N} X_i^2}{N} - 2\left(\frac{\sum_{i=1}^{N} X_i}{N}\right)^2 + \left(\frac{\sum_{i=1}^{N} X_i}{N}\right)^2$$

$$= \frac{N \sum_{i=1}^{N} X_i^2 - (\sum_{i=1}^{N} X_i)^2}{N^2}$$

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

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

พิสูจน์

$$\sigma_{(x+k)} = \sigma$$

กำหนดให้ประชากรประกอบด้วย N ค่าสังเกต (X_i) พร้อมด้วยมัชฌิมเลขคณิต μ ส่วนเบี่ยงเบนมาตรฐาน σ ถ้าค่าคงที่ k บวกเข้ากับแต่ละค่าสังเกตมัชฌิมเลขคณิตของ $X_i + k$ ค่าสังเกต $\mu_{(x+k)}$ เป็น $\mu + k$ (ได้พิสูจน์มาแล้ว) ส่วนเบี่ยงเบนมาตรฐานของ $X_i + k$ ค่าสังเกต $\sigma_{(x+k)}$ อาจคำนวณหาได้โดยตรงด้วยการแทนค่า $X_i + k$ สำหรับ X_i และ $\mu + k$ สำหรับ μ ดังต่อไปนี้

$$\sigma_{(x+k)} = \sqrt{\frac{\sum_{i=1}^{N} [(X_i+k) - (\mu+k)]^2}{N}}$$

$$= \sqrt{\frac{\sum_{i=1}^{N} (X_i+k-\mu-k)^2}{N}}$$

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

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

ในทำนองเดียวกันการพิสูจน์

$$\sigma_{(x-k)} = \sigma$$

พิสูจน์

$$\sigma_{gx} = g\sigma$$

กำหนดให้ประชากรของ N ค่าสังเกต (X_i) พร้อมด้วยมัชฌิมเลขคณิต μ และส่วนเบี่ยงเบนมาตรฐาน σ ถ้าหากว่าแต่ละค่าสังเกตคูณด้วยค่าคงที่ g มัชฌิมเลขคณิตของค่าสังเกต $g X_i$ คือ $\mu_{gx} = g\mu$ (ได้พิสูจน์มาจากบทก่อน) ส่วนเบี่ยงเบนมาตรฐานของค่าสังเกต $g X_i$ คือ σ_{gx} อาจคำนวณหาได้ดังต่อไปนี้

$$\sigma_{gx} = \sqrt{\frac{\sum_{i=1}^{N} (g X_i - g\mu)^2}{N}}$$

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

$$\sigma_{gx} = \sqrt{\frac{g^2 \sum_{i=1}^{N} (X_i - \mu)^2}{N}}$$

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

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

$$= g\,\sigma$$

ในทำนองเดียวกัน เราสามารถพิสูจน์ได้ว่า

$$\sigma_{x/g} = \frac{\sigma}{g}$$

พิสูจน์

$$Z_i = \frac{X_i - \mu}{\sigma}$$

เป็นคะแนนมาตรฐานที่มีมัชฌิมเลขคณิตเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 เราคำนวณหามัชฌิมเลขคณิต μ_Z ได้

$$\mu_Z = \frac{\sum_{i=1}^{N} Z_i}{N}$$

$$= \frac{\sum_{i=1}^{N} \left(\frac{X_i - \mu}{\sigma}\right)}{N}$$

สังเกตว่า μ กับ σ เป็นค่าคงที่

$$\mu_Z = \frac{\frac{1}{\sigma}\sum_{i=1}^{N}(X_i-\mu)}{N}$$

$$= \frac{\frac{1}{\sigma}\left(\sum_{i=1}^{N}X_i - \sum_{i=1}^{N}\mu\right)}{N}$$

$$= \frac{\frac{1}{\sigma}\left(\sum_{i=1}^{N}X_i - N\mu\right)}{N}$$

$$= \frac{\frac{1}{\sigma}\left(\sum_{i=1}^{N}X_i - N\frac{\sum_{i=1}^{N}X_i}{N}\right)}{N}$$

$$= \frac{\frac{1}{\sigma}\left(\sum_{i=1}^{N}X_i - \sum_{i=1}^{N}X_i\right)}{N}$$

$$= \frac{\frac{1}{\sigma}(0)}{N} = 0$$

เราสามารถคำนวณส่วนเบี่ยงเบนมาตรฐาน σ_Z ได้ดังนี้

$$\sigma_Z = \sqrt{\frac{\sum_{i=1}^{N}(Z_i-0)^2}{N}}$$

$$= \sqrt{\frac{\sum_{i=1}^{N}Z_i^2}{N}}$$

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

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

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

$$= \frac{1}{\sigma}\sqrt{\sigma^2}$$

$$= \frac{1}{\sigma}\,\sigma$$

$$= 1$$

ค่าของมัชฌิมและความแปรปรวนคำนวณได้มาจากข้อมูล เมื่อมีการเปลี่ยนแปลงของข้อมูลจะมีผลถึงมัชฌิมและความแปรปรวน ตัวอย่างเช่น มัชฌิมของข้อมูล 3 จำนวน 11, 11, 17 คือ 13

ตารางที่ 4.3

ความแปรปรวนของ X

X	$(X-\mu)$	$(X-\mu)^2$
11	−2	4
11	−2	4
17	4	16
39		24
	$\mu = \frac{39}{3} = 13$	$\sigma^2 = \frac{24}{3} = 8$

ถ้าเอา 5 บวกเข้าแต่ละข้อมูล ข้อมูลใหม่ก็จะเป็น 16, 16, 22 และมัชฌิมใหม่เป็น 18 ซึ่งมากไปกว่ามัชฌิมเดิมอยู่ 5 อย่างไรก็ตามความแปรปรวนมีค่าเท่าเดิมดูตารางที่ 4.4

ตารางที่ 4.4

ความแปรปรวนของ X + 5

X	$(X-\mu)$	$(X-\mu)^2$
16	−2	4
16	−2	4
22	4	16
54		24
$\mu = \frac{54}{3} = 18$		$\sigma^2 = \frac{24}{3} = 8$

เหตุผลง่าย ๆ ก็คือ ค่าของความแปรปรวนคำนวณได้จากการเบี่ยงเบนของข้อมูลจากมัชฌิม เนื่องจากข้อมูลแต่ละข้อมูลและมัชฌิมเปลี่ยนไปเป็นจำนวนเท่ากัน การเบี่ยงเบน $(x-\mu)$ จึงไม่เปลี่ยนแปลงดัง ตารางที่ 4.3 และ 4.4 เพราะฉะนั้น ค่าของความแปรปรวนจึงไม่มีผลในการเปลี่ยนข้อมูลนั้น

ตัวอย่าง 4.3.1 ในปัจจุบันครอบครัวหนึ่งมีอยู่ 7 คน คุณพ่ออายุ 45 ปี คุณแม่อายุ 42 ปี บุตรคนโต อายุ 21 ปี บุตรคนกลางอายุ 17 ปี คนสุดท้องอายุ 10 ปี มัชฌิมเลขคณิตและส่วนเบี่ยงเบนมาตรฐาน ของคนทั้ง 7 เท่ากับ 24 ปี และ 3.5 ปี ตามลำดับ อยากทราบว่าใน 6 ปีข้างหน้า มัชฌิมเลขคณิต และส่วนเบี่ยงเบนมาตรฐานจะเป็นเท่าใด

วิธีทำ ในปัจจุบันมัชฌิมเลขคณิตของคนทั้ง 7 เท่ากับ 24 ปี
ส่วนเบี่ยงเบนมาตรฐาน เท่ากับ 3.5 ปี

ใน 6 ปีข้างหน้าแต่ละคนก็จะเพิ่มขึ้นอีกคนละ 6 ปี รวม 7 คน เป็น 7 คูณ 6 เท่ากับ 42 ปี แต่มัชฌิมเลขคณิตเพิ่มขึ้นของคนทั้ง 7 เท่ากับ $\frac{42}{7}$ เท่ากับ 6 ปี

เพราะฉะนั้นมัชฌิมเลขคณิตของคนทั้ง 7 ใน 6 ปีข้างหน้าก็จะมีค่าเท่ากับมัชฌิมเลขคณิต ในปัจจุบันบวกด้วยมัชฌิมเลขคณิตที่เพิ่มขึ้นของคนทั้ง 7 เท่ากับ 24+6 เท่ากับ 30 ปี แต่ส่วนเบี่ยงเบน ของคนทั้ง 7 ไม่เปลี่ยนแปลง

เพราะฉะนั้นส่วนเบี่ยงเบนมาตรฐานของคนทั้ง 7 คนมีค่าเท่าเดิม คือ 3.5 ปี

ในกรณีค่าสังเกตหรือข้อมูล 11, 11, 17, คูณด้วยจำนวนปริมาณคงที่ สมมุติให้เป็น 10 ค่าสังเกตหรือข้อมูลใหม่จะเป็น 110, 110, 170, จะมีค่ามากขึ้น 10 เท่าของข้อมูลเก่าและค่ามัชฌิม เลขคณิตจะเป็น 130 คือเป็น 10 เท่าของค่ามัชฌิมเดิม

ตารางที่ 4.5

ความแปรปรวนของ 10 X

X	$(X-\mu)$	$(X-\mu)^2$
110	−20	400
110	−20	400
170	40	1,600
390		2,400
$\mu = \frac{390}{3} = 130$		$\sigma^2 = \frac{2400}{3} = 800$

ความแปรปรวนเดิมดังแสดงในตารางที่ 4.3 และข้อมูลใหม่ในตารางที่ 4.5 พร้อมด้วยค่าความแปรปรวนใหม่เป็น 100 เท่าของค่าความแปรปรวนเดิม แต่ว่าส่วนเบี่ยงเบนมาตรฐานใหม่ก็จะมีค่าเป็น 10 เท่าของค่าส่วนเบี่ยงเบนมาตรฐานเดิม ส่วนเบี่ยงเบน (110-130), (110-130) และ (170-130) แต่ละจำนวนก็จะมีค่า 10 เท่าของค่าเดิมและถ้าแต่ละจำนวนของส่วนเบี่ยงเบนยกกำลังสองก็จะมีค่าเป็น 100 เท่าของค่าเดิม

ในกรณีตัวคูณจำนวนคงที่เป็นตัวหาร อย่างเช่นตัวคูณมีค่าเป็น 10 เปลี่ยนเป็น 1/10 ดังจะได้กล่าวดังต่อไปนี้

ถ้าทุก ๆ ค่าสังเกตหรือข้อมูลคูณด้วย 2 มัชฌิมใหม่ก็จะมีค่าเป็น 2 เท่าของมัชฌิมเดิม ความแปรปรวนใหม่ก็จะมีค่าเป็น 4 เท่าของค่าความแปรปรวนเดิม ส่วนเบี่ยงเบนมาตรฐานใหม่จะมีค่าเป็น 2 เท่าของค่าส่วนเบี่ยงเบนมาตรฐานเดิม แต่ถ้าทุก ๆ ข้อมูลหรือค่าสังเกตหารด้วย 2 หรือคูณด้วย 1/2 มัชฌิมเลขคณิตใหม่ก็จะมีค่าเป็น 1/2 ของค่ามัชฌิมเลขคณิตเดิม ความแปรปรวนใหม่ก็จะมีค่าเป็น 1/4 ของค่าความแปรปรวนเดิมและส่วนเบี่ยงเบนมาตรฐานใหม่จะมีค่า 1/2 ของค่าส่วนเบี่ยงเบนมาตรฐานเดิม

วัตถุประสงค์ของสองหัวข้อนี้มีความสำคัญเกี่ยวกับความรู้ในหลักสถิติมากกว่าในการวิเคราะห์ข้อมูล ตัวอย่างเช่น ชายคนหนึ่งต้องการทราบน้ำหนักเฉลี่ยและค่าความแปรปรวนสุทธิของสารเคมี 1,000 ห่อ แน่ละเขาจะต้องหักน้ำหนักของภาชนะออกก่อนที่จะดำเนินการคำนวณหาค่ามัชฌิมและค่าความแปรปรวน ถ้าภาชนะแต่ละภาชนะหนัก 2 ออนซ์ เพราะฉะนั้นภาชนะ 1,000 ห่อ แต่ละห่อก็จะต้องลบด้วย 2 ออนซ์ ซึ่งไม่จำเป็นเลย วิธีที่ง่ายในการหามัชฌิมและความแปรปรวนของน้ำหนักทั้งหมด (น้ำหนักของสารเคมีและภาชนะรวมกัน) มัชฌิมสุทธิสามารถคำนวณโดยเอา 2 ลบออกจากมัชฌิมทั้งหมด (หมายความว่ามัชฌิมที่ผิดรวมทั้งภาชนะด้วย) แต่ส่วนค่าความแปรปรวนสุทธิมีค่าคงเดิมเหมือนค่าความแปรปรวนที่คิดรวมทั้งภาชนะด้วย

ในกรณีปริมาณคงที่เป็นตัวคูณข้อมูลหรือค่าสังเกต ดังตัวอย่าง มีข้อมูลอยู่ 1,000 ข้อมูล มีหน่วยเป็นหลา แต่ถ้าเราต้องการทราบมัชฌิมเลขคณิต และค่าความแปรปรวนให้มีหน่วยเป็นฟุตและฟุตกำลังสอง (ตารางฟุต) แน่ละ แต่ละข้อมูล ในจำนวน 1,000 ข้อมูลจะต้องคูณด้วย 3 เพื่อเปลี่ยนจากหน่วยหลา เป็น ฟุต แล้วคำนวณหาค่ามัชฌิมและความแปรปรวนจากข้อมูลใหม่เหล่านี้ แต่ก็ไม่จำเป็นเลย เราสามารถคำนวณหาค่ามัชฌิมเลขคณิตและความแปรปรวนได้จากค่ามัชฌิมเลขคณิตและความแปรปรวนเดิม อย่างเช่นค่าของมัชฌิมและความแปรปรวนเดิมเป็น 20 หลา และ 10 หลากำลังสอง (ตารางหลา) ตามลำดับมัชฌิมใหม่ที่มีหน่วยฟุตก็จะมีค่าเป็น 20 คูณ 3 หรือ

60 ฟุต และความแปรปรวนก็จะมีค่าเป็น 10 คูณ 3^2 หรือ 90 ตารางฟุต

4.5 การกระจายสัมพัทธ์ (Relative Dispersion)

การเปลี่ยนแปลงการกระจายของข้อมูล 2 กลุ่ม หรือมากกว่าในกรณีที่ข้อมูลที่นำมาเปรียบเทียบอาจมีหน่วยต่างกัน อย่างเช่น ข้อมูลกลุ่มหนึ่งมีหน่วยเป็น เซ็นติเมตร แต่อีกกลุ่มหนึ่งมีหน่วยเป็นกิโลกรัม หรือว่าข้อมูลที่นำมาเปรียบเทียบจะมีหน่วยเหมือนกัน ถึงแม้ว่าข้อมูลจะมีขนาดแตกต่างกัน เราก็สามารถเปรียบเทียบการกระจายของข้อมูล 2 กลุ่มหรือมากกว่าได้ โดยดัดแปลงมาตรการที่ใช้การวัดการกระจายเป็นแบบอัตราส่วนหรือร้อยละ ดังตัวอย่าง ถ้าวัดการกระจายของข้อมูลด้วย พิสัย ก็เทียบเป็นอัตราส่วนหรือร้อยละ ของผลบวกของค่าปลายทั้งสอง ถ้าใช้ส่วนเบี่ยงเบนเฉลี่ยก็เทียบกับค่าของมัชฌิมเลขคณิต ค่าที่ได้นี้เรียกว่า **"สัมประสิทธิ์ของการกระจาย"** (Coefficient of Dispersion) แต่ที่นิยมมากที่สุดได้แก่การใช้ส่วนเบี่ยงเบนมาตรฐานเทียบกับค่าของมัชฌิมเลขคณิต ค่าที่ได้นี้เรียกว่า "สัมประสิทธิ์ของความแปรผัน" (Coefficient of Variation) และเขียนเป็นสูตรได้

$$V = \frac{\sigma}{\mu} \times 100 \qquad \text{..........4.5.1}$$

ตัวอย่าง 4.5.1 สมมุติว่าผลการสอบสองครั้งของนักศึกษากลุ่มหนึ่ง สอบครั้งแรกได้คะแนนเฉลี่ย 60 และส่วนเบี่ยงเบนมาตรฐาน 6 ของคะแนนเต็ม 100 สอบครั้งที่สองได้คะแนนเฉลี่ย 700 และส่วน เบี่ยงเบนมาตรฐาน 7 ของคะแนนเต็ม 1,000 อยากทราบว่าผลการสอบสองครั้ง ครั้งไหนจะมีการกระจายมากกว่ากัน

วิธีทำ สอบครั้งแรก $\mu = 60; \sigma = 6$

$$\because \quad V = \frac{6}{60} \times 100 = \frac{1}{10} \times 100 = 10$$

สอบครั้งที่สอง $\mu = 700; \sigma = 7$

$$\because \quad V = \frac{7}{700} \times 100 = \frac{1}{100} \times 100 = 1$$

เราจะเห็นได้ว่าการกระจายสัมพัทธ์ของผลการสอบครั้งแรกเป็น 10 เท่าของผลการสอบ ครั้งหลัง

ตัวอย่าง 4.5.2 สมมุติว่าเราประสงค์ที่จะเปรียบเทียบประสิทธิภาพของเครื่องจักรที่ผลิตลูกเทนนิสกับประสิทธิภาพของเครื่องจักรที่ผลิตลูกกอล์ฟ เส้นผ่าศูนย์กลางของลูกเทนนิสเฉลี่ย 3 นิ้ว และเส้นผ่าศูนย์กลางของลูกกอล์ฟ 1 นิ้ว ความผันแปรของเส้นผ่าศูนย์กลางในรูปของส่วนเบี่ยงเบนมาตรฐานทั้งของลูกเทนนิสและลูกกอล์ฟเป็น 0.003 นิ้ว และ 0.003 นิ้ว นั่นคือเท่ากัน ปัญหาก็คือ เครื่องจักรชนิดไหนทำงานมีประสิทธิภาพกว่ากัน ถ้าหากว่าเครื่องจักรทั้งสองผลิตลูกเทนนิสและลูกกอล์ฟให้ค่าส่วนเบี่ยงเบนมาตรฐานเท่ากัน นี้หมายความว่ามีประสิทธิภาพเท่ากัน คำตอบก็คือเปล่า เพราะว่าขนาดเฉลี่ยของเส้นผ่าศูนย์กลางของลูกเทนนิสใหญ่กว่า ด้วยเหตุนี้ความผันแปร 0.003 ของเครื่องจักรที่ผลิตลูกกอล์ฟจึงใหญ่กว่า ความผันแปรของเครื่องจักรที่ผลิตลูกเทนนิส มาตราที่ใช้วัดเปรียบเทียบความผันแปรคือสัมประสิทธิ์ของความผันแปรคือ

$$CV = \frac{\sigma}{\mu} 100$$

สัมประสิทธิ์ของความผันแปรสำหรับลูกเทนนิส

$$CV_1 = \frac{0.003}{3} = 0.001$$

และสัมประสิทธิ์ของความผันแปรสำหรับลูกกอล์ฟ

$$CV_2 = \frac{0.003}{1} = 0.003$$

เพราะฉะนั้นสัมประสิทธิ์ของความผันแปรของลูกกอล์ฟมากเป็นสามเท่าของลูกเทนนิส เครื่องจักรผลิตลูกกอล์ฟมีประสิทธิภาพน้อยกว่า

ตามที่ได้กล่าวมาข้างต้น เป็นการเปรียบเทียบการกระจายระหว่างกลุ่มข้อมูล เราอาจต้องการเปรียบเทียบการกระจายระหว่างค่า 2 ค่า หรือมากกว่าจากกลุ่มข้อมูลที่มีลักษณะต่างกันอย่างใดอย่างหนึ่งเกี่ยวกับ

1. มีมัชฌิมเลขคณิตหรือส่วนเบี่ยงเบนมาตรฐานต่างกัน
2. มีหน่วยต่างกัน

การเปรียบเทียบแบบนี้คิดเป็นหน่วยของส่วนเบี่ยงเบนมาตรฐาน หรือเรียกว่าคะแนนมาตรฐาน คะแนนมาตรฐาน คือตัวเลขที่บอกว่าส่วนเบี่ยงเบนของคะแนนดิบ จากคะแนนเฉลี่ยเป็นกี่เท่าของส่วนเบี่ยงเบนมาตรฐาน เขียนเป็นสูตรได้

$$Z = \frac{X - \mu}{\sigma} \qquad \text{..........(4.5.2)}$$

ตัวอย่าง 4.5.3 นักศึกษาคนหนึ่งสอบวิชาคณิตศาสตร์ได้ 84 คะแนน แต่คะแนนเฉลี่ยของผู้สอบคณิตศาสตร์ได้ 76 และส่วนเบี่ยงเบนมาตรฐาน 10 และสอบวิชาสถิติได้ 90 คะแนน คะแนนส่วนเฉลี่ยของผู้สอบสถิติ 82 และส่วนเบี่ยงเบนมาตรฐาน 16 อยากทราบว่าเขาสอบวิชาใดได้ดีกว่าเมื่อเปรียบเทียบกับส่วนเฉลี่ยของนักศึกษา

วิธีทำ สำหรับวิชาคณิตศาสตร์ คะแนนมาตรฐานของนักศึกษาผู้นี้

$$Z = \frac{X-\mu}{\sigma}$$

ในเมื่อ $X = 84;\ \mu = 76;\ \sigma = 10$

แทนค่าได้ $Z = \frac{84-76}{10} = \frac{8}{10} = .8$

สำหรับวิชาสถิติ ในเมื่อ $X = 90;\ \mu = 82;\ \sigma = 16$

$$Z = \frac{90-82}{16} = \frac{8}{16} = .5$$

เพราะฉะนั้นคะแนนวิชาคณิตศาสตร์ของนักศึกษาผู้นี้สูงกว่าคะแนนเฉลี่ยของผู้สอบคณิตศาสตร์ .8 เท่าของส่วนเบี่ยงเบนมาตรฐาน แต่คะแนนวิชาสถิติสูงกว่าคะแนนเฉลี่ย .5 เท่าของส่วนเบี่ยงเบนมาตรฐาน สรุปได้ว่านักเรียนผู้นี้ทำคะแนนสอบในวิชาคณิตศาสตร์ได้ดีกว่าวิชาสถิติ

แบบฝึกหัดบทที่ 4

1. ประชากรประกอบด้วยข้อมูลดังต่อไปนี้

   1.8 2.0 1.8 1.9 2.0

   ก. หามัชฌิมเลขคณิต (μ) (1.9)

   ข. หาค่าความแปรปรวน (σ^2) (ทศนิยมสามตำแหน่ง) (.008)

   ค. เอา 1.0 ลบออกทุก ๆ ข้อมูลและหามัชฌิมเลขคณิตและค่าความแปรปรวนของข้อมูลใหม่ (0.9 ; .008)

   ง. คูณข้อมูลทุก ๆ ข้อมูลด้วย 5 และหามัชฌิมเลขคณิตและค่าความแปรปรวนของข้อมูลใหม่ (9.5 ; .2)

2. กำหนดให้จำนวนเลข 2 กลุ่ม คือ 2, 5, 8, 11, 14, และ 2, 8, 14 จงคำนวณหา

   ก. มัชฌิมเลขคณิตและค่าความแปรปรวนของแต่ละกลุ่ม ($\mu_1 = 8, \mu_2 = 8$, $\sigma_1^2 = 18$, $\sigma_2^2 = 24$)

   ข. มัชฌิมรวมหรือผสม Combined mean และค่าความแปรปรวนร่วมหรือผสม (Combined variance) ($\mu_{12} = 8$, $\sigma_2^2 = 20.2$)

3. ในเดือนหนึ่ง นาย ก. ขายผลิตภัณฑ์เคมีได้ 15,000 บาท ขณะที่นาย ข. ขายเครื่องตกแต่งสำนักงานได้ 6,250 บาท ถ้าคิดถัวเฉลี่ยแล้ว ก. และ ข. ขายผลิตภัณฑ์เคมีและเครื่องตกแต่ง สำนักงานได้ 13,000 และ 5,000 และส่วนเบี่ยงเบนมาตรฐาน 2,000 และ 500 บาทตามลำดับ อยากทราบว่าใครมีอัตราการขายได้สูงกว่า ($Z_{ก} = 1$, $Z_{ข} = 2.5$)

4. สำนักงานการลงทุนมีรายงานการขายประจำวันดังนี้ สินค้าชนิด ก. ขายได้ราคา 110 บาท แต่ราคาขายเฉลี่ย 80 บาท ส่วนเบี่ยงเบนมาตรฐาน 20 บาท สินค้าชนิด ข. ขายได้ราคา 18 บาท แต่ราคาขายเฉลี่ย 10 บาท ส่วนเบี่ยงเบนมาตรฐาน 2 บาท อยากทราบว่า นายทุนควรจะเลือกลงทุนสินค้าชนิดใด ($Z_{ก} = 1.5$, $Z_{ข} = 4$)

5. จงคำนวณหาส่วนเบี่ยงเบนมาตรฐาน ส่วนเบี่ยงเบนควอร์ไทล์ของคะแนนสอบไล่ของนักศึกษาจำนวน 600 คน ($\sqrt{139.7775}$, 7.45)

คะแนน	ความถี่ (f)
20 – 29	4
30 – 39	48
40 – 49	117
50 – 59	235
60 – 69	140
70 – 79	39
80 – 89	10
90 – 99	7

6. จากตารางแจกแจงความถี่ข้อ 5 ถ้าจุดกึ่งกลาง (X) ของแต่ละชั้นลบด้วย 55.35 และหารด้วย $\sqrt{139.7775}$ ความสัมพันธ์ระหว่างจุดกลางใหม่ (U) และเก่า (X) คือ ($\mu_u = 0, \sigma_u^2 = 1$)

$$U = \frac{X - 55.35}{\sqrt{139.7775}}$$ จงคำนวณหามัชฌิมเลขคณิตและความแปรปรวน

7. เปรียบเทียบการกระจายของมูลค่าการจำหน่ายสินค้าเป็นรายเดือนว่าร้านไหนจำหน่ายต่อเดือนแปรเปลี่ยนมากกว่ากัน ($cv_ก = 16.7\%$, $cv_ข = 1.4\%$)

	ร้าน ก.	ร้าน ข.
มัชฌิมเลขคณิต	30,000 บาท	50,000 บาท
ส่วนเบี่ยงเบนมาตรฐาน	5,000 บาท	700 บาท