บทที่ 2
ความรู้เบื้องต้นทางสถิติสำหรับงานวิจัย

หลังจากเก็บรวบรวมข้อมูลมาแล้วไม่ว่าจะเป็นข้อมูลเชิงปริมาณหรือข้อมูลเชิงคุณภาพ ขั้นต่อไปจะต้องทำการวิเคราะห์ข้อมูลขั้นต้น โดยใช้สถิติเบื้องต้นเรียกว่า สถิติพรรณนา เพื่อเป็นการสรุปลักษณะเบื้องต้นของข้อมูล เริ่มจากการแจกแจงความถี่ การวัดแนวโน้มเข้าสู่ส่วนกลางและการวัดการกระจายของข้อมูล ซึ่งจะได้กล่าวถึงในบทนี้ เมื่อทำการวิเคราะห์ข้อมูลขั้นต้นแล้ว จึงจะนำผลที่ได้จากการวิเคราะห์ไปทำการวิเคราะห์ขั้นสูงต่อไปเพื่อนำผลจากการวิเคราะห์ขั้นสูงนี้ไปช่วยใช้ในการตัดสินใจให้ถูกต้องยิ่งขึ้น

การแจกแจงความถี่ (frequency distribution)

ข้อมูลที่เก็บรวบรวมมาได้และยังมิได้มีการจัดหมวดหมู่ ข้อมูลชนิดนี้เรียกว่าข้อมูลดิบ (raw data) ข้อมูลดิบเหล่านี้ถ้ามีจำนวนมากเราจะไม่สามารถมองเห็นลักษณะของข้อมูลได้จึงต้องมีการจัดเตรียมข้อมูลดิบให้เป็นหมวดหมู่ ถ้าข้อมูลดิบมีจำนวนน้อย ก็ใช้วิธีเรียงข้อมูลจากมากไปหาน้อย หรือจากน้อยไปหามาก ข้อมูลที่เรียงลำดับแบบนี้ เรียกว่า ungrouped data แต่ถ้าข้อมูลดิบมีจำนวนมาก ต้องใช้วิธีการแจกแจงความถี่ (frequency distribution)

การแจกแจงความถี่ เป็นการจัดเรียงลำดับข้อมูลดิบที่เก็บรวบรวมมาได้ โดยจัดให้เป็นหมวดหมู่ แล้วหาจำนวนของข้อมูลในแต่ละหมู่ ข้อมูลที่หาได้โดยวิธีการนี้ เรียกว่า ข้อมูลที่เป็นหมวดหมู่ (grouped data)

รูปแบบของการแจกแจงความถี่ สามารถทำได้ 3 รูปแบบ คือ

1. ตารางแจกแจงความถี่ (frequency distribution)
2. ตารางแจกแจงความถี่สะสม (cumulative frequency distribution)
3. แผนภูมิหรือกราฟ
   - 3.1 ฮิสโทแกรม (histogram)
   - 3.2 รูปหลายเหลี่ยมแห่งความถี่ (frequency polygon)
   - 3.3 โค้งความถี่ (frequency curves)
   - 3.4 โค้งความถี่สะสม (cumulative frequency หรือ ogive curve)

1. ตารางแจกแจงความถี่ (frequency distribution)

ลักษณะของตารางแจกแจงความถี่โดยทั่วไป ประกอบด้วย ข้อมูล รอยขีด และความถี่ดังตัวอย่างต่อไปนี้

ตัวอย่างที่ 2.1 คะแนนจากการสอบของนักศึกษา 60 คน เป็นดังนี้

28	22	20	17	16	25	18	22	28	17
19	22	22	21	19	27	27	25	23	24
28	26	21	18	24	21	24	22	20	22
24	28	16	23	22	25	24	22	25	21
17	28	24	27	23	22	22	29	16	20
21	21	26	27	28	24	28	16	23	22

จากข้อมูลข้างต้นสามารถสรุปให้กระทัดรัดได้โดยใช้ตารางแจกแจงความถี่ ดังนี้

คะแนน	รอยขีด (รอยคะแนน)	ความถี่
16	IIII	4
17	III	3
18	II	2
19	II	2
20	III	3
21	~~IIII~~ I	6
22	~~IIII~~ ~~IIII~~ I	11
23	IIII	4
24	~~IIII~~ II	7
25	III	4
26	II	2
27	IIII	4
28	~~IIII~~ II	7
29	I	1
รวม		60

นอกจากนี้ข้อมูลดิบสามารถสร้างตารางแจกแจงความถี่ได้อีกลักษณะหนึ่งโดยการจัดข้อมูลที่ใกล้เคียงกันเป็นหมู่หรือเป็นชั้น (group or class) ช่วงกว้างของข้อมูลในแต่ละชั้น เรียกว่า อันตรภาคชั้น (class interval)

ตาราง 2.1 คะแนนการทดสอบวิชาคณิตศาสตร์ของนักศึกษาชั้นปีที่ 1

คะแนน	รอยขีด (รอยคะแนน)	ความถี่
45 – 49	II	2
50 – 54	IIII	4
55 – 59	~~IIII~~ III	8
60 – 64	~~IIII~~ ~~IIII~~	10
65 – 69	~~IIII~~ ~~IIII~~ ~~IIII~~	15
70 – 74	~~IIII~~ ~~IIII~~ I	11
75 – 79	~~IIII~~ I	6
80 – 84	III	3
85 – 89	I	1
รวม		60

จากตาราง 2.1 จะได้ว่าข้อมูลชุดนี้มี 9 อันตรภาคชั้น โดยอันตรภาคชั้นที่ 2 มีช่วงคะแนน 50 – 54 เรียกค่าทั้ง 2 นี้ว่าขีดจำกัดล่างและขีดจำกัดบน ในชั้นนี้มีความถี่ หรือจำนวนนักศึกษา 4 คน โดยมีคะแนนตั้งแต่ 49.5 ถึง 54.5 คะแนน เราเรียกค่าทั้ง 2 นี้ว่า ขอบเขตล่างและขอบเขตบน

วิธีการสร้างตารางแจกแจงความถี่ มีขั้นตอนการทำดังนี้

1. หาค่าพิสัย โดยคำนวณจาก พิสัย = ข้อมูลสูงสุด – ข้อมูลต่ำสุด
2. กำหนดจำนวนชั้นของข้อมูลตามความเหมาะสม
3. คำนวณหาค่าความกว้างของอันตรภาคชั้น จากสูตร

$$\text{ความกว้างของอันตรภาคชั้น} = \frac{\text{พิสัย}}{\text{จำนวนชั้น}}$$

4. ตีตาราง 3 ช่อง ประกอบด้วย ข้อมูลหรืออันตรภาคชั้น รอยขีด และความถี่
5. เขียนอันตรภาคชั้นของแต่ละชั้น โดยอาศัยความกว้างของอันตรภาคชั้นเป็นเครื่องช่วยจะเริ่มจากชั้นข้อมูลค่าต่ำสุด หรือชั้นข้อมูลค่าสูงสุดก็ได้ โดยมีหลักว่าขีดจำกัดล่างของชั้นข้อมูลค่าต่ำสุดและขีดจำกัดบนของชั้นข้อมูลค่าสูงสุดจะต้องคลุมค่าของข้อมูลทั้งหมด

6. ตรวจดูว่าข้อมูลแต่ละตัวอยู่ในอันตรภาคชั้นใด แล้วขีดรอยขีดให้ตรงกับชั้นนั้นตัวละขีด
7. นับจำนวนรอยขีดใส่ในช่องความถี่

ตัวอย่างที่ 2.2 จงสร้างตารางแจกแจงความถี่ให้มีจำนวนชั้น 5 ชั้น เมื่อกำหนดคะแนนจากการสอบของนักศึกษา 60 คน ดังนี้

38	22	20	17	16	25	18	22	28	17
19	22	22	31	19	27	37	25	23	44
28	36	41	18	24	21	24	22	20	32
24	28	16	23	22	35	24	42	25	31
17	28	24	27	23	22	22	39	16	20
21	21	26	27	28	24	28	16	23	22

วิธีทำ ข้อมูลสูงสุด = 44 และ ข้อมูลต่ำสุด = 16

$\therefore$ พิสัย = 44 – 16 = 28

โจทย์ต้องการสร้างตารางแจกแจงความถี่ โดยให้มีจำนวนชั้น 5 ชั้น

จากสูตร ความกว้างของอันตรภาคชั้น $= \dfrac{\text{พิสัย}}{\text{จำนวนชั้น}} = \dfrac{28}{5} = 5.6 \approx 6$

แล้วจึงสร้างตารางแจกแจงความถี่ ประกอบด้วย 3 ช่อง คือ ช่องคะแนน รอยขีด และจำนวนนักศึกษา ดังนี้คือ

คะแนน	รอยขีด	จำนวนคน
16 – 21	𝍸 𝍸 𝍸 II	17
22 – 27	𝍸 𝍸 𝍸 𝍸 𝍸 I	26
28 – 33	𝍸 IIII	9
34 – 39	𝍸	5
40 – 45	III	3
รวม		60

2. ตารางแจกแจงความถี่สะสม (Cumulative Frequency Distribution)

ความถี่สะสมของชั้นใด หมายถึง ความถี่ในชั้นนั้นรวมกับความถี่ของชั้นอื่นที่มีข้อมูลน้อยหรือมากกว่าชั้นนั้น ซึ่งความถี่สะสมอาจจะนับจากค่ามากไปหาน้อย หรือจากค่าน้อยไปหามากก็ได้ แต่ในทางปฏิบัติมักนิยมความถี่สะสมที่นับจากค่าน้อยไปหามาก

ตัวอย่างที่ 2.3 จากตัวอย่างที่ 2.2 สามารถสร้างตารางแจกแจงความถี่สะสมแบบน้อยกว่า (less than) ได้ดังนี้

คะแนน	ความถี่ (f)	ความถี่สะสมแบบน้อยกว่า (less than)
16 – 21	17	17
22 – 27	26	43
28 – 33	9	52
34 – 39	5	57
40 – 45	3	60

จากตัวอย่างที่ 2.3 ความถี่ 17 หมายความว่ามีนักศึกษา 17 คนที่สอบได้คะแนนระหว่าง 15.5 และ 21.5 คะแนน ส่วนความถี่ 26 หมายความว่ามีนักศึกษา 26 คน สอบได้คะแนนระหว่าง 21.5 และ 27.5 คะแนน

ส่วนความถี่สะสม 43 หมายความว่า มีนักศึกษา 43 คน ที่สอบได้คะแนนต่ำกว่า 27.5 คะแนนและความถี่สะสม 52 หมายความว่า มีนักศึกษา 52 คนที่สอบได้คะแนนต่ำกว่า 33.5 คะแนน

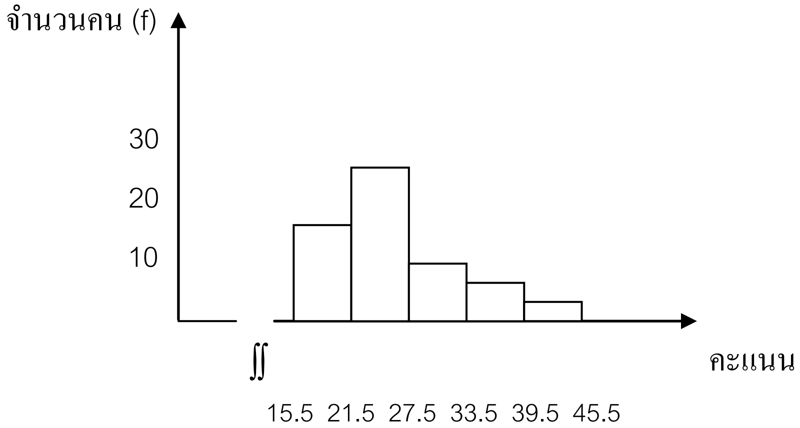
3. แผนภูมิหรือกราฟ

การแจกแจงความถี่นอกจากจะแสดงความถี่ของข้อมูลด้วยตารางทั้ง 3 แบบแล้วยังสามารถแสดงได้ด้วยแผนภูมิ หรือกราฟ ดังนี้

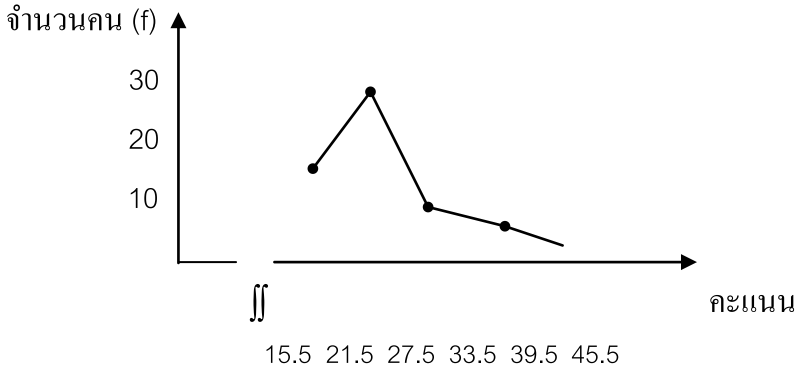
3.1 **ฮิสโทแกรม (histogram)** หรือรูปสี่เหลี่ยมผืนผ้าของความถี่ เป็นการแสดงความถี่โดยอาศัยพื้นที่สี่เหลี่ยมผืนผ้าคล้ายกับการนำเสนอข้อมูลแบบกราฟแท่ง โดยที่แต่ละแท่งจะติดกัน เพราะใช้ขอบเขตล่างและขอบเขตบนของข้อมูลแต่ละชั้น มาแบ่งหน่วยบนแกนแนวนอน (แกน X) ส่วนแกนแนวตั้ง (แกน Y) แสดงความถี่ (f)

ตัวอย่างที่ 2.5 จากตัวอย่างที่ 2.2 จงเขียนฮิสโทแกรม

คะแนน	ขีดจำกัดชั้น	จำนวน (ความถี่)
16 – 21	15.5 – 21.5	17
22 – 27	21.5 – 27.5	26
28 – 33	27.5 – 33.5	9
34 – 39	33.5 – 39.5	5
40 – 45	39.5 – 45.5	3

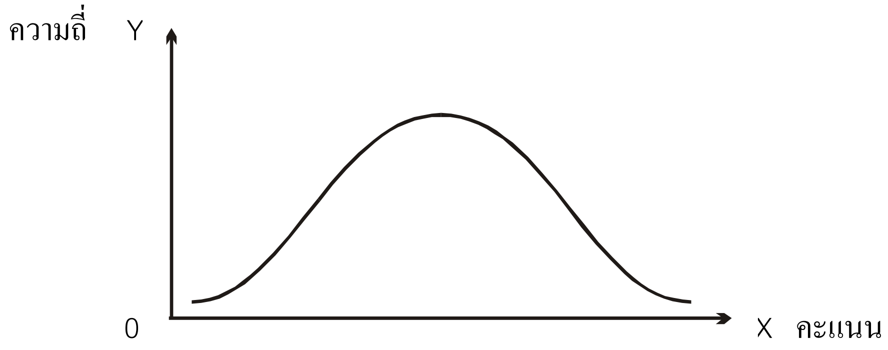


3.2 **รูปหลายเหลี่ยมแห่งความถี่ (frequency polygon)** เป็นการแสดงความถี่โดยจุดกึ่งกลางชั้น ของข้อมูลและความถี่ของข้อมูลในแต่ละชั้น แบ่งหน่วยบนแกนแนวนอนแสดงจุดกึ่งกลางและแกนแนวตั้งแสดงความถี่จะได้ตำแหน่งของจุดกึ่งกลางและความถี่ของข้อมูลในแต่ละชั้นเมื่อลากเส้นเชื่อมระหว่างตำแหน่งจากข้อมูลน้อยไปหาข้อมูลมาก ภาพที่ได้เป็นรูปหลายเหลี่ยมแห่งความถี่หรือสร้างต่อจากการทำฮิสโทแกรมโดยการแบ่งกึ่งกลางที่ยอดของแต่ละแท่งแล้วลากเส้นเชื่อมของจุดแบ่ง จะได้รูปหลายเหลี่ยมของความถี่ จากตัวอย่าง 2.2 สามารถแสดงข้อมูลโดยใช้รูปหลายเหลี่ยมแห่งความถี่ได้ดังนี้



3.3 **โค้งความถี่ (frequency curves)** เป็นโค้งที่เกิดจากการปรับเส้นของรูปหลายเหลี่ยมแห่งความถี่ให้เรียบขึ้น โดยการปรับจะต้องให้พื้นที่ภายใต้เส้นโค้งที่ปรับใหม่มีขนาดใกล้เคียงกับพื้นที่ของรูปหลายเหลี่ยมแห่งความถี่ โดยแบ่งออกเป็น 5 ชนิด คือ

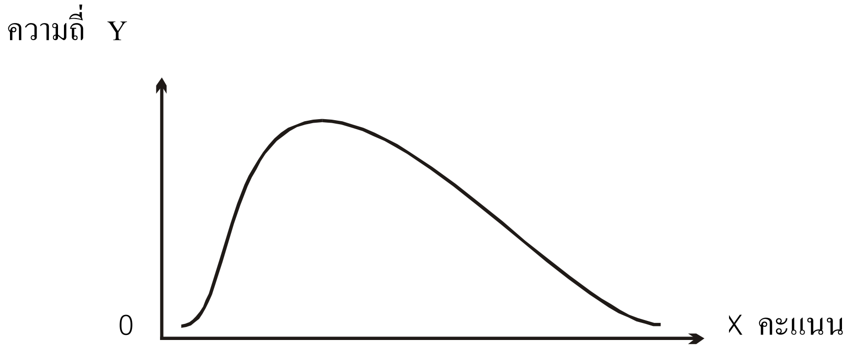
3.3.1 **โค้งปกติ (normal curves)**



รูปที่ 2.1 โค้งปกติ

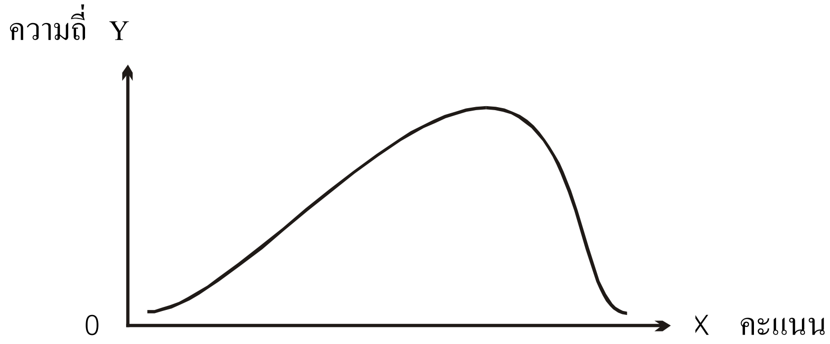
3.3.2 **โค้งเบ้ (moderately asymetrical or skewed)** มี 2 ลักษณะ คือ

3.3.2.1 โค้งเบ้ทางขวา (positively skewed) เป็นโค้งที่แสดงให้เห็นว่านักศึกษาส่วนใหญ่ได้คะแนนน้อย มีจำนวนมาก ส่วนนักศึกษาที่ได้คะแนนมาก มีจำนวนน้อย



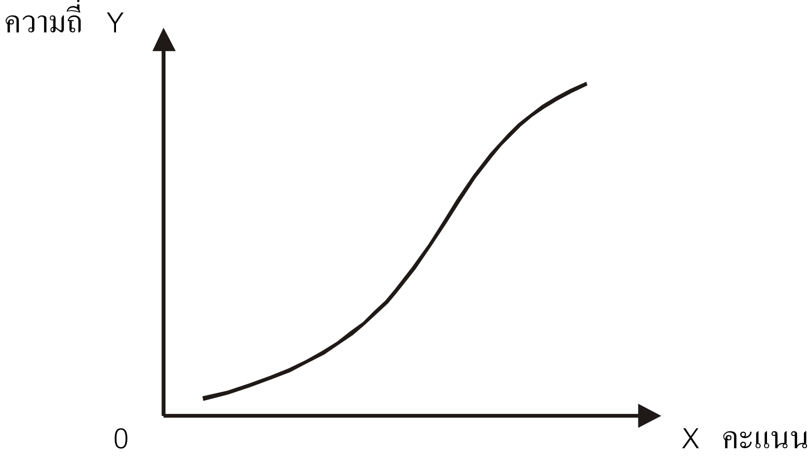
รูปที่ 2.2 โค้งเบ้ทางขวา

3.3.2.2 โค้งเบ้ทางซ้าย (negatively skewed) เป็นโค้งที่มีลักษณะตรงกันข้ามกันโค้งเบ้ทางขวา กล่าวคือ ส่วนใหญ่นักศึกษาที่สอบได้คะแนนมากมีจำนวนมาก นักศึกษาที่สอบได้คะแนนน้อยจะมีจำนวนน้อย



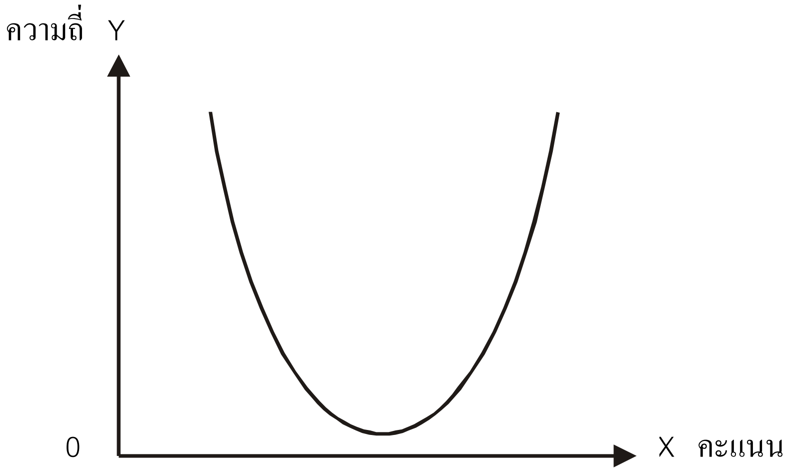
รูปที่ 2.3 โค้งเบ้ทางซ้าย

3.3.3 **โค้งรูปเจ (J - shaped)** เป็นโค้งความถี่ที่มีความถี่ต่ำทางด้านคะแนนน้อยและค่อย ๆ เพิ่มขึ้นเมื่อคะแนนมาก ซึ่งถ้าเป็นผลการศึกษาแสดงว่า มีการพัฒนาขึ้น



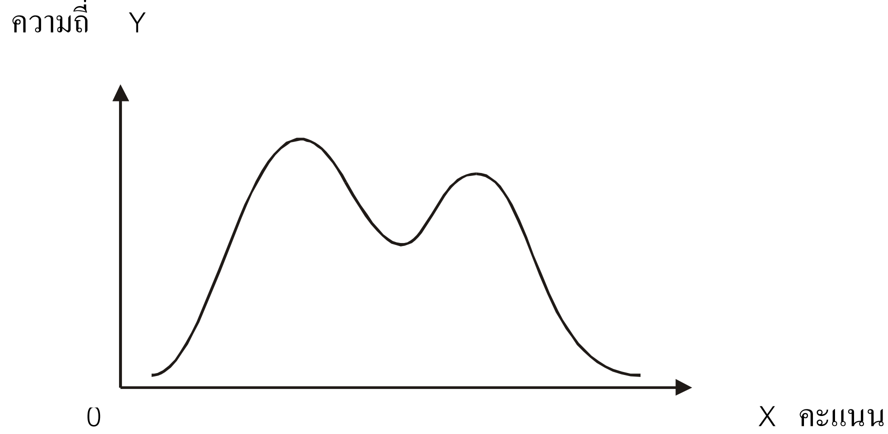
รูปที่ 2.4 โค้งรูปตัวเจ

3.3.4 **โค้งรูปถ้วย (U - shaped)** เป็นการแจกแจงที่มีความถี่มากที่ตำแหน่งคะแนนมากและคะแนนน้อย แต่ตรงกลางมีความถี่น้อย



รูปที่ 2.5 โค้งรูปถ้วย

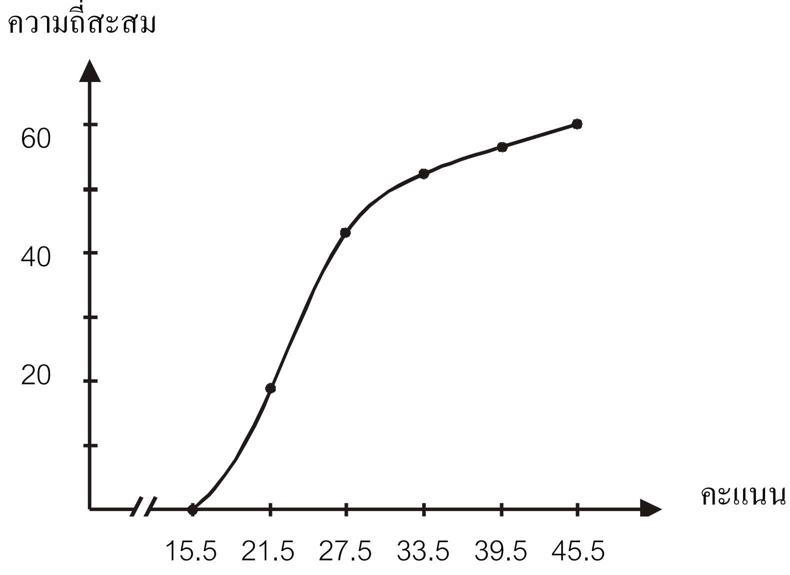
3.3.5 **โค้งสองยอด (bimodel shaped)** โค้งลักษณะนี้จะมีนักศึกษาที่ได้คะแนนมากที่สุดในแต่ละช่วง แยกเป็น 2 แห่ง



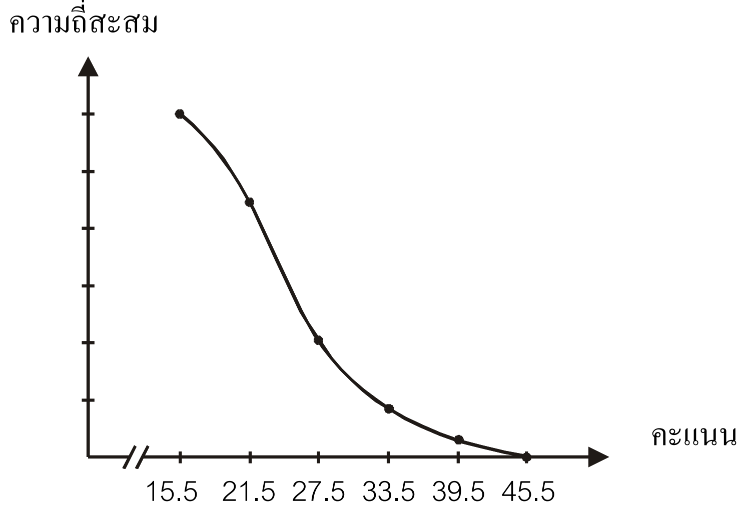
รูปที่ 2.6 โค้งสองยอด

3.4 **โค้งความถี่สะสม (cumulative frequency หรือ ogive curve)** เป็นโค้งที่แสดงความถี่สะสมของข้อมูลตั้งแต่ค่าต่ำสุด ก่อนสร้างโค้งความถี่สะสม ควรสร้างตารางแจกแจงความถี่สะสมก่อน เมื่อได้ตารางแจกแจงความถี่สะสมแล้วจึงเขียนกราฟความถี่สะสมโดยใช้แกน Y เป็นความถี่สะสมและแกน X เป็นขีดจำกัดชั้นที่แท้จริงของข้อมูลแต่ละชั้น ต่อจากนั้นจึงลากเส้นโยงจุดแต่ละจุดเหล่านั้น

จากตัวอย่างที่ 2.3 และตัวอย่างที่ 2.4 สามารถสร้างโค้งความถี่สะสมแบบน้อยกว่า (รูปที่ 2.7) และโค้งความถี่สะสมแบบมากกว่า (รูปที่ 2.8) ได้ดังนี้



รูปที่ 2.7 โค้งความถี่สะสมแบบน้อยกว่าของคะแนนการสอบของนักศึกษาจำนวน 60 คน



รูปที่ 2.8 โค้งความถี่สะสมแบบมากกว่าของคะแนนการสอบของนักศึกษาจำนวน 60 คน

การวัดแนวโน้มเข้าสู่ส่วนกลาง (measures of central tendency)

การวัดแนวโน้มเข้าสู่ส่วนกลาง เป็นระเบียบวิธีการทางสถิติพรรณาที่ใช้ในการหาค่ากลางหรือค่าเฉลี่ยเพื่อใช้เป็นตัวแทนแสดงขนาดและลักษณะของข้อมูลแต่ละชุด ทำให้ได้ตัวแทนของข้อมูลแต่ละชุดโดยไม่จำเป็นต้องนำข้อมูลทั้งชุดมาพิจารณา

การวัดแนวโน้มเข้าสู่ส่วนกลางที่นิยมใช้กันทั่วไป มี 3 วิธี คือ

1. ค่าเฉลี่ยเลขคณิต (arithmetic mean)
2. มัธยฐาน (median)
3. ฐานนิยม (mode)

1. **ค่าเฉลี่ยเลขคณิต** เป็นการวัดแนวโน้มเข้าสู่ส่วนกลางที่ใช้กันมากที่สุด ใช้สัญลักษณ์ $\bar{X}$ อ่านว่า เอกซ์บาร์ (X – bar) สำหรับข้อมูลที่ได้จากกลุ่มตัวอย่างและใช้สัญลักษณ์ μ อ่านว่า มิว (mu) สำหรับข้อมูลที่ได้จากประชากร ค่าเฉลี่ยเลขคณิตเป็นค่าที่หาได้ โดยนำผลรวมของข้อมูลทั้งหมดหารด้วยจำนวนข้อมูลทั้งหมด

1.1 **ข้อมูลที่ไม่ได้แบ่งกลุ่ม (ungrouped data)**

นิยามที่ 2.1 ให้ $X_1, X_2, X_3, \ldots, X_n$ เป็นข้อมูลชุดหนึ่งมี n จำนวน ค่าเฉลี่ยเลขคณิต คือ $\bar{X}$ หาได้

จาก $\bar{X} = \dfrac{X_1 + X_2 + X_3 + \cdots + X_n}{n} = \dfrac{\sum_{i=1}^{n} X_i}{n}$ เมื่อ

$\bar{X}$ = ค่าเฉลี่ยเลขคณิตของกลุ่มตัวอย่าง

n = จำนวนข้อมูลทั้งหมด

X_i = ข้อมูลแต่ละตัว i = 1, 2, 3, . . ., n

หรือเขียนย่อ ๆ ว่า $\bar{X} = \frac{\sum X}{n}$

สำหรับค่าเฉลี่ยเลขคณิตที่เป็นข้อมูลจากประชากร (μ) หาได้จากสูตร $\mu = \frac{\sum X}{N}$ เมื่อ N เป็นจำนวนข้อมูลทั้งหมดของประชากร

ตัวอย่างที่ 2.6 จงหาค่าเฉลี่ยเลขคณิตของน้ำหนักของเด็ก 10 คน เป็นกิโลกรัม ดังนี้

47.3, 42.5, 49, 49.1, 52, 50, 55.2, 47.5, 52.2, 49.5

วิธีทำ ให้ $\bar{X}$ เป็นค่าเฉลี่ยเลขคณิตของน้ำหนักของเด็ก 10 คน

จากสูตร $\bar{X} = \frac{\sum X}{n}$

$$= \frac{47.3 + 42.5 + 49 + \cdots + 49.5}{10} = 49.43$$

นั่นคือ ค่าเฉลี่ยเลขคณิตของน้ำหนักของเด็ก 10 คน เท่ากับ 49.43 กิโลกรัม

1.2 **ข้อมูลที่แบ่งกลุ่ม (grouped data)** ค่าเฉลี่ยเลขคณิตหาได้โดยรวมผลคูณของความถี่กับจุดกึ่งกลางของข้อมูลในแต่ละชั้นหารด้วยจำนวนข้อมูลทั้งหมด

ให้จุดกึ่งกลางของข้อมูลแต่ละชั้นเป็น $X_1, X_2, X_3, \ldots, X_k$ และมีความถี่ $f_1, f_2, f_3, \ldots, f_k$ ตามลำดับ

1.2.1 **ค่าเฉลี่ยเลขคณิตของประชากร** (μ) หาได้จาก

$$\mu = \frac{\sum_{i=1}^{k} f_i X_i}{N} = \frac{\sum fX}{N} \qquad \text{เมื่อ } N = \sum_{i=1}^{k} f_i$$

1.2.2 **ค่าเฉลี่ยเลขคณิตของกลุ่มตัวอย่าง** ($\bar{X}$) หาได้จาก

$$\bar{X} = \frac{\sum_{i=1}^{k} f_i X_i}{n} = \frac{\sum fX}{n}$$

ตัวอย่างที่ 2.7 สมาคมนักกีฬาแห่งประเทศไทยได้เก็บรวบรวมข้อมูลเกี่ยวกับเงินที่มีผู้บริจาค ดังตาราง

เงินบริจาค (บาท)	จำนวนคน
0 – 50	2
50 – 100	6
100 – 150	7
150 – 200	6
200 - 250	4
250 - 300	5

จงหาเงินบริจาคโดยเฉลี่ยของสมาคมนักกีฬาแห่งประเทศไทย

วิธีทำ ให้ $\bar{X}$ เป็นเงินบริจาคโดยเฉลี่ยของสมาคมนักกีฬาแห่งประเทศไทย

จากสูตร $\bar{X} = \frac{\sum fX}{n}$

เงินบริจาค (บาท)	จำนวน (f)	X	fX
0 – 50	2	25	50
50 – 100	6	75	450
100 – 150	7	125	875
150 – 200	6	175	1050
200 - 250	4	225	900
250 - 300	5	275	1375
รวม	30		4700

แทนค่าในสูตร $\bar{X} = \frac{4700}{30} = 156.67$

นั่นคือ เงินบริจาคโดยเฉลี่ยของสมาคมนักกีฬาแห่งประเทศไทย เท่ากับ 156.67 บาท

1.3 **ค่าเฉลี่ยเลขคณิตแบบถ่วงน้ำหนัก (weighted arithmetic mean)**

การหาค่าเฉลี่ยเลขคณิตบางครั้งให้ผลไม่ดี ถ้าข้อมูลแต่ละตัวมิได้มีน้ำหนักทัดเทียมกัน ผลลัพธ์ก็จะได้ค่ากลางที่ไม่ยุติธรรม คือเป็นค่ากลางที่สมมติว่าข้อมูลทุกตัวมีน้ำหนักเท่ากัน

ให้ข้อมูล $X_1, X_2, X_3, \ldots, X_n$ มีค่าถ่วงน้ำหนักเป็น $W_1, W_2, W_3, \ldots, W_n$ ตามลำดับ จะได้ว่า

$$\bar{X} = \frac{W_1X_1 + W_2X_2 + W_3X_3 + \cdots + W_nX_n}{W_1 + W_2 + W_3 + \cdots + W_n}$$

$$= \frac{\sum W_iX_i}{\sum W_i}$$

ตัวอย่างที่ 2.8 ผลการเรียนของนาย ก. ดังตาราง จงหาคะแนนเฉลี่ยของ นาย ก.

วิชา	หน่วยกิต	เกรด
คณิตศาสตร์	4	1
ภาษาไทย	3	2
พลศึกษา	1	4
ศิลปศึกษา	2	2

วิธีทำ ให้ $\bar{x}$ เป็นคะแนนเฉลี่ยของนาย ก.

จากสูตร $$\bar{X} = \frac{\sum W_i X_i}{\sum W_i}$$

จากโจทย์ $$X_1 = 1, X_2 = 2, X_3 = 4, X_4 = 2$$

$$W_1 = 4, W_2 = 3, W_3 = 1, W_4 = 2$$

แทนค่า $$\bar{X} = \frac{(4)(1) + (3)(2) + (1)(4) + (2)(2)}{4 + 3 + 1 + 2}$$

$$= \frac{4+6+4+4}{10} = \frac{18}{10} = 1.8$$

นั่นคือ คะแนนเฉลี่ยของการเรียนของนาย ก. เท่ากับ 1.8

1.4 **ค่าเฉลี่ยเลขคณิตรวม (combination arithmetic mean)**

ถ้ามีข้อมูลหลาย ๆ ชุดและทราบค่าเฉลี่ยเลขคณิตของข้อมูลแต่ละชุด จะสามารถหาค่าเฉลี่ยเลขคณิตรวมได้ ให้ $\bar{x}_1, \bar{x}_2, \bar{x}_3, \ldots, \bar{x}_k$ เป็นค่าเฉลี่ยเลขคณิตของข้อมูลชุดที่ 1, 2, 3, . . .,k ตามลำดับ จะได้ว่า

$$\bar{X}_{\text{รวม}} = \frac{n_1\bar{X}_1 + n_2\bar{X}_2 + n_3\bar{X}_3 + \cdots + n_k\bar{X}_k}{n_1 + n_2 + n_3 + \cdots + n_k}$$

$$= \frac{\sum n_i \bar{X}_i}{\sum n_i}$$

ตัวอย่างที่ 2.9 ผลการสอบวิชาสถิติธุรกิจของนักศึกษา 3 ห้อง ปรากฏผลดังนี้ ห้องที่ 1 มีนักศึกษา 50 คน คะแนนเฉลี่ย 67 คะแนน ห้องที่ 2 มีนักศึกษา 60 คน คะแนนเฉลี่ย 65 คะแนน ห้องที่ 3 มีนักศึกษา 45 คน คะแนนเฉลี่ย 80 คะแนน จงหาค่าเฉลี่ยรวมของผลการสอบของนักศึกษาทั้งหมดของ 3 ห้องรวมกัน

วิธีทำ ให้ $\bar{X}_{\text{รวม}}$ เป็นค่าเฉลี่ยรวมของผลการสอบของนักศึกษาทั้งหมดของ 3 ห้องรวมกัน

จากสูตร $$\bar{X}_{\text{รวม}} = \frac{\sum n_i \bar{X}_i}{\sum n_i}$$

จากโจทย์ $$\bar{X}_1 = 67, n_1 = 50,\ \bar{X}_2 = 65, n_2 = 60, \bar{X}_3 = 80, n_3 = 45$$

แทนค่า $$\bar{X}_{\text{รวม}} = \frac{(50)(67) + (60)(65) + (45)(80)}{50 + 60 + 45}$$

$$= \frac{3350 + 3900 + 3600}{155} = \frac{10850}{155} = 70$$

นั่นคือ คะแนนเฉลี่ยรวมของผลการสอบของนักศึกษาทั้งหมด 3ห้องเท่ากับ70 คะแนน

1.5 **คุณสมบัติของค่าเฉลี่ยเลขคณิต**

1.5.1 **ทฤษฎีบทที่ 2.1** ถ้าข้อมูลแต่ละตัวมีค่าลดลง หรือเพิ่มขึ้นเท่ากันทุกค่าสมมติให้เท่ากับค่า k ซึ่งเป็นค่าคงที่แล้วค่าเฉลี่ยเลขคณิตใหม่จะมีค่าลดลง หรือเพิ่มขึ้นเท่ากับจำนวนค่าคงที่ k นั้น

ให้ $X_1, X_2, X_3, \ldots, X_n$ เป็นข้อมูลชุดเดิม จะได้ $X_1 \pm k, X_2 \pm k, X_3 \pm k, \cdots, X_n \pm k$ เป็นข้อมูลชุดใหม่ $\therefore \quad \bar{X}_{\text{ใหม่}} = \bar{X}_{\text{เดิม}} \pm k$

1.5.2 **ทฤษฎีบทที่ 2.2** ถ้าข้อมูลแต่ละตัวมีค่าเพิ่มขึ้น k เท่าของค่าเดิม ค่าเฉลี่ยเลขคณิตใหม่จะมีค่าเป็น k เท่าของค่าเดิม

ให้ $X_1, X_2, X_3, \ldots, X_n$ เป็นข้อมูลชุดเดิม จะได้ $kX_1, kX_2, kX_3, \ldots, kX_n$ เป็นข้อมูลชุดใหม่

$\therefore \quad \bar{X}_{\text{ใหม่}} = k\bar{X}_{\text{เดิม}}$

ให้ $X_1, X_2, X_3, \ldots, X_n$ เป็นข้อมูลชุดเดิม จะได้ $\frac{\bar{X}_1}{k}, \frac{\bar{X}_2}{k}, \frac{\bar{X}_3}{k}, \cdots, \frac{\bar{X}_n}{k}$ เป็นข้อมูลชุดใหม่

$\therefore \quad \bar{X}_{\text{ใหม่}} = \frac{\bar{X}_{\text{เดิม}}}{k}$

1.5.3 ผลรวมของความแตกต่างระหว่างข้อมูลแต่ละตัวจากค่าเฉลี่ยเลขคณิตของข้อมูลชุดนั้นมีค่าเท่ากับ 0

1.5.4 ผลรวมของความแตกต่างกำลังสองของข้อมูลแต่ละตัวจากจำนวนใด ๆ จะน้อยที่สุดเมื่อจำนวนนั้นมีค่าเท่ากับค่าเฉลี่ยเลขคณิตของข้อมูลชุดนั้น

$$\sum_{i=1}^{N}(X_i - A)^2 \quad \text{มีค่าน้อยที่สุด เมื่อ } A = \bar{X}$$

หรือเขียนว่า
$$\sum_{i=1}^{N}(X_i - \bar{X})^2 \le \sum_{i=1}^{N}(X_i - A)^2 \quad \text{เมื่อ A เป็นจำนวนใด ๆ}$$

1.5.5 ค่าเฉลี่ยเลขคณิตของข้อมูลชุดใด จะต้องมีค่าอยู่ระหว่างข้อมูลตัวที่มีค่าน้อยที่สุดกับข้อมูลตัวที่มีค่ามากที่สุดของข้อมูลชุดนั้น $\quad X_{min} < \bar{X} < X_{max}$

เมื่อ X_{min} เป็นข้อมูลตัวที่มีค่าน้อยที่สุด และ X_{max} เป็นข้อมูลตัวที่มีค่ามากที่สุด

2. **มัธยฐาน** หมายถึงข้อมูลที่อยู่ในตำแหน่งกลางระหว่างข้อมูลทั้งหมด หรือข้อมูลตัวกลางที่มีค่าต่ำกว่าข้อมูลนั้น เป็นจำนวนเท่ากับข้อมูลที่มีค่าสูงกว่าข้อมูลนั้น ค่ามัธยฐานอาจเป็นข้อมูลแท้ ๆ ตัวใดตัวหนึ่งก็ได้ หรืออาจเป็นเลขจำนวนหนึ่งซึ่งแสดงว่า 50% ของข้อมูลทั้งหมดมีค่ามากกว่าอีก 50% มีค่าต่ำกว่า

2.1 **ข้อมูลที่ไม่ได้แบ่งกลุ่ม (ungrouped data)**

นิยามที่ 2.2 ให้ N เป็นจำนวนข้อมูลทั้งหมดที่เรียงลำดับจากน้อยไปหามาก หรือมากไปหาน้อย มัธยฐาน คือ ข้อมูลที่อยู่ในตำแหน่งที่ $\frac{N+1}{2}$ เมื่อ N เป็นเลขคี่และเท่ากับค่าเฉลี่ยของข้อมูลในตำแหน่งที่ $\frac{N}{2}$ และ $\frac{N+2}{2}$ เมื่อ N เป็นเลขคู่

ตัวอย่างที่ 2.10 จงหาค่ามัธยฐานของจำนวนวันที่คนไข้นอนพักรักษาตัวอยู่ที่โรงพยาบาล โดยจากการสำรวจรายงานของโรงพยาบาลแห่งหนึ่ง พบว่า จำนวนวันที่คนไข้นอนพักรักษาตัวเป็นดังนี้ 7, 4, 9, 2, 5, 10, 8

วิธีทำ เรียงข้อมูลจากน้อยไปมาก 2, 4, 5, 7, 8, 9, 10

มัธยฐานอยู่ในตำแหน่งที่ $\frac{N+1}{2} = \frac{7+1}{2} = 4$

$\therefore$ มัธยฐาน = 7

นั่นคือ มัธยฐานของจำนวนวันที่คนไข้นอนพักรักษาตัวอยู่ที่โรงพยาบาลเป็นจำนวน 7 วัน

ตัวอย่างที่ 2.11 จงหามัธยฐานของข้อมูลต่อไปนี้ 18.3, 19.4, 20.6, 22.2, 20.4, 18.7

วิธีทำ เรียงข้อมูลจากน้อยหาไปมาก 18.3, 18.7, 19.4, 20.4, 20.6, 22.2

N = 6, มัธยฐาน คือ ค่าเฉลี่ยของข้อมูลที่อยู่ตำแหน่งที่ $\frac{N}{2}$ และ $\frac{N+2}{2}$ คือ ตำแหน่งที่ 3 และ 4

$\therefore$ มัธยฐาน $= \frac{19.4+20.4}{2}$ = 19.9

นั่นคือ มัธยฐาน เท่ากับ 19.9

2.2 **ข้อมูลที่แบ่งกลุ่ม (grouped data)**

คำนวณได้โดยใช้สูตร $\text{Med.} = L + \frac{i(F_N - F_1)}{F_2 - F_1}$

เมื่อ L = ขอบเขตล่างของชั้นที่มี F_N อยู่

i = ความกว้างของอันตรภาคชั้น

F_N = ตำแหน่งที่ต้องการหา = $\frac{N}{2}$

F_1 = ความถี่สะสมของชั้นที่อยู่ถัดจากชั้น F_N ไปทางชั้นที่น้อยกว่า

F_2 = ความถี่สะสมของชั้น F_N

ตัวอย่างที่ 2.12 ตารางต่อไปนี้แสดงถึงอายุการใช้งานของหลอดไฟ (หน่วยเป็นเดือน) ยี่ห้อหนึ่ง จำนวน 1000 หลอด จงหาค่ามัธยฐานของอายุการใช้งานของหลอดไฟจำนวน 1000 หลอด

อายุการใช้งาน (เดือน)	จำนวนหลอดไฟ
3.5 – 3.9	50
4.0 – 4.4	150
4.5 – 4.9	430
5.0 – 5.4	200
5.5 – 5.9	110
6.0 – 6.4	60
รวม	1000

วิธีทำ จากสูตร $Med. = L + \dfrac{i(F_N - F_1)}{F_2 - F_1}$

อายุการใช้งาน (เดือน)	จำนวนหลอดไฟ (f_i)	F
3.5 – 3.9	50	50
4.0 – 4.4	150	200
4.5 – 4.9	430	630
5.0 – 5.4	200	830
5.5 – 5.9	110	940
6.0 – 6.4	60	1000
รวม	1000	

จากตารางได้ว่า $i = 0.5, F_N = \dfrac{N}{2} = \dfrac{1000}{2} = 500, F_1 = 200, F_2 = 630$ และ $L = 4.45$

แทนค่า $Med. \quad = 4.45 + \dfrac{0.5(500-200)}{630-200} \quad = 4.45 + 0.35 = 4.8$

นั่นคือ ค่ามัธยฐานของอายุการใช้งานของหลอดไฟจำนวน 1000 หลอด เท่ากับ 4.8 เดือน

3. **ฐานนิยม** เป็นการวัดแนวโน้มเข้าสู่ส่วนกลางอีกแบบหนึ่ง หาได้โดยการพิจารณาว่าข้อมูลตัวใดซ้ำกันมากที่สุด หรือมีความถี่มาก ข้อมูลตัวนั้น คือ ฐานนิยม

3.1 **ข้อมูลที่ไม่ได้แบ่งกลุ่ม (ungrouped data)** หาได้โดยการพิจารณาตรวจดูว่าข้อมูลตัวใดเป็นที่ซ้ำกันมากที่สุด ข้อมูลตัวนั้น คือ ฐานนิยม เช่น

4, 5, 7, 5, 2, 5 ฐานนิยม คือ 5

3, 5, 9, 8, 6, 4	ฐานนิยม	คือ	ไม่มีฐานนิยม
2, 3, 7, 8, 2, 8, 1	ฐานนิยม	คือ	2, 8

3.2 **ข้อมูลที่แบ่งกลุ่ม (grouped data)**

ฐานนิยมของข้อมูลที่แจกแจงความถี่แล้วจะอยู่ในชั้นที่มีความถี่สูงสุด โดยมีสูตรในการคำนวณ ดังนี้ $\text{Mode} = L + \frac{id_1}{d_1 + d_2}$

เมื่อ L = ขอบเขตล่างของชั้นที่มีความถี่มากที่สุด

i = ความกว้างของอันตรภาคชั้น

d_1 = ผลต่างระหว่างความถี่ของชั้นที่มากที่สุดกับชั้นที่ติดกับชั้นที่มีข้อมูลน้อยกว่า

d_2 = ผลต่างระหว่างความถี่ของชั้นที่มากที่สุดกับชั้นที่ติดกับชั้นที่มีข้อมูลมากกว่า

ตัวอย่างที่ 2.13 คะแนนสอบวิชาคณิตศาสตร์ของนักศึกษาจำนวน 100 คน ดังนี้

คะแนนสอบ	จำนวน
35 – 39	8
40 – 44	12
45 – 49	20
50 – 54	35
55 – 59	15
60 – 64	10

จงหาฐานนิยมของคะแนนสอบวิชาคณิตศาสตร์ของนักศึกษากลุ่มนี้

วิธีทำ จากสูตร $\text{Mode} = L + \frac{id_1}{d_1 + d_2}$

จากโจทย์นักศึกษากลุ่มนี้ได้คะแนนสอบมากที่สุดที่อยู่ในช่วงคะแนน 50 – 54

ดังนั้น $L = 49.5$, $i = 5$, $d_1 = 35 - 20 = 15$, $d_2 = 35 - 15 = 20$

$\therefore \quad \text{Mode} = 49.5 + \frac{5(15)}{15 + 20} = 49.5 + 2.14 = 51.64$

นั่นคือ ฐานนิยมของคะแนนสอบวิชาคณิตศาสตร์ของนักศึกษากลุ่มนี้ เท่ากับ 51.64 คะแนน

การวัดการกระจาย (measures of variation)

การพิจารณาลักษณะของข้อมูลด้วยการใช้การวัดแนวโน้มเข้าสู่ส่วนกลางเพียงอย่างเดียว อาจทำให้ไม่ทราบถึงลักษณะของข้อมูลเพียงพอ จำเป็นต้องมีการวัดกระจายควบคู่กันไปด้วย ข้อมูลชุดใดประกอบด้วยค่าที่มีความมากน้อยแตกต่างกันมาก แสดงว่าข้อมูลนั้นมีการกระจายมาก ข้อมูลใดที่ประกอบด้วยค่าที่ใกล้เคียงกัน ข้อมูลชุดนั้นจะมีการกระจายน้อย ดังนั้นตัวเลขที่ใช้วัดการกระจายของข้อมูลกับค่าของตัวกลางจะเป็นตัวแทนของข้อมูลทั้งชุดในการวิเคราะห์ข้อมูลต่อไป

วิธีการวัดการกระจายมีหลายวิธี แต่ที่นิยมกันมากมี 5 วิธี คือ

1. พิสัย (range)
2. ส่วนเบี่ยงเบนควอไทล์ (quartile deviation)
3. ส่วนเบี่ยงเบนเฉลี่ย (mean deviation or average deviation)
4. ส่วนเบี่ยงเบนมาตรฐาน (standard deviation) และ ความแปรปรวน (variance)
5. สัมประสิทธิ์แห่งการกระจาย(the coefficient of variation)

1. พิสัย

นิยามที่ 2.3 พิสัยของข้อมูล หาได้จากความแตกต่างระหว่างค่าสูงสุดกับค่าต่ำสุดของข้อมูลชุดหนึ่งๆ

พิสัย = ค่าสูงสุด – ค่าต่ำสุด

1.1 **ข้อมูลที่ไม่ได้แบ่งกลุ่ม (ungrouped data)**

ตัวอย่างที่ 2.14 จงหาพิสัยของข้อมูลต่อไปนี้

ข้อมูลชุดที่ 1 ประกอบด้วย 3, 15, 15, 15, 15, 15, 15, 15

ข้อมูลชุดที่ 2 ประกอบด้วย 3, 3, 3, 3, 15, 15, 15, 15

ข้อมูลชุดที่ 3 ประกอบด้วย 3, 4, 7, 10, 11, 12, 13, 15

วิธีทำ พิสัยของข้อมูลทั้ง 3 ชุด = ค่าสูงสุด – ค่าต่ำสุด = 15 – 3 = 12

จากตัวอย่างที่ 2.14 จะเห็นว่าข้อมูลทั้ง 3 ชุดมีพิสัยเท่ากัน คือ 12 แต่การกระจายของข้อมูลต่างกัน ข้อมูลชุดที่ 1 จำนวนแรกเท่านั้นเท่ากับ 3 นอกนั้นมีค่าเท่ากับ 15 ทั้งหมด ข้อมูลชุดที่ 2 สี่จำนวนแรกเท่ากับ 3 นอกนั้นมีค่าเท่ากับ 15 ส่วนข้อมูลชุดที่ 3 มีการกระจายตั้งแต่ 3 ถึง 15

การวัดการกระจายโดยใช้พิสัยเป็นวิธีที่คำนวณได้ง่าย แต่มีข้อจำกัด คือ ใช้ค่าเพียง 2 ค่าเท่านั้นในการคำนวณ ส่วนค่าอื่น ๆ มิได้นำมาใช้ในการคำนวณเลยจึงเป็นการคิดเพียงคร่าว ๆ เท่านั้น เราจึงใช้พิสัยในการวัดการกระจาย ก็ต่อเมื่อต้องการความรวดเร็วและไม่คำนึงถึงความละเอียดถูกต้องมากนัก

1.2 **ข้อมูลที่แบ่งกลุ่ม (grouped data)**

ในกรณีที่ข้อมูลที่แบ่งกลุ่ม ค่าพิสัยของข้อมูลชุดนั้นหาได้จาก

พิสัย = ค่าสูงสุดของอันตรภาคชั้นที่มีค่ามากที่สุด – ค่าต่ำสุดของอันตรภาคชั้นที่มีค่าน้อยที่สุด

ตัวอย่างที่ 2.15 จงหาค่าพิสัยของค่าใช้จ่ายต่อวันของนักศึกษา 100 คน

ค่าใช้จ่าย (บาท)	ความถี่
50 – 54	5
55 – 59	17
60 – 64	42
65 – 69	28
70 – 74	8

วิธีทำ จากสูตร

พิสัย = ค่าสูงสุดของอันตรภาคชั้นที่มีค่ามากที่สุด – ค่าต่ำสุดของอันตรภาคชั้นที่มีค่าน้อยที่สุด

= 74.5 – 49.5 = 25

นั่นคือ ค่าพิสัยของค่าใช้จ่ายต่อวันของนักศึกษา 100 คน = 25 บาท

2. ส่วนเบี่ยงเบนควอไทล์ ใช้สัญลักษณ์เป็น Q.D. ส่วนเบี่ยงเบนควอไทล์เป็นการวัดการกระจายซึ่งคำนวณได้จากครึ่งหนึ่งของระยะระหว่างควอไทล์ที่ 3 และควอไทล์ที่ 1 โดยมีสูตร ดังนี้

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

เมื่อ Q.D. คือ ส่วนเบี่ยงเบนควอไทล์ Q_1 คือ ควอไทล์ที่ 1 และ Q_3 คือ ควอไทล์ที่ 3

2.1 ข้อมูลที่ไม่ได้แบ่งกลุ่ม (ungrouped data)

ตัวอย่างที่ 2.16 จงหาค่าส่วนเบี่ยงเบนควอไทล์ของข้อมูลต่อไปนี้

60, 65, 72, 50, 63, 68, 74, 56, 67, 61, 79

วิธีทำ จัดเรียงข้อมูลใหม่ โดยเรียงจากค่าน้อยไปค่ามากตามลำดับ จะได้

50, 56, 60, 61, 63, 65, 67, 68, 72, 74, 79

หาค่า Q_1 โดยการหาค่าตำแหน่งของ Q_1 จาก $\frac{1}{4}(N+1) = \frac{1}{4}(11+1) = 3 \quad \therefore Q_1 = 60$

ในทำนองเดียวกับหาค่า Q_3 โดยการหาค่าตำแหน่งของ Q_3 จาก $\frac{3}{4}(11+1) = 9 \quad \therefore Q_3 = 72$

$$\therefore \quad \text{Q.D.} = \frac{72-60}{2} = 6$$

นั่นคือ ค่าส่วนเบี่ยงเบนควอไทล์ของข้อมูลชุดนี้ คือ 6

ตัวอย่างที่ 2.17 จงหาค่าส่วนเบี่ยงเบนควอไทล์ของข้อมูลต่อไปนี้ 16, 18, 8, 12, 14, 20, 24, 20, 26, 28

วิธีทำ จัดเรียงข้อมูลใหม่ โดยเรียงจากค่าน้อยไปค่ามากตามลำดับ จะได้

8, 12, 14, 16, 18, 20, 20, 24, 26, 28

หาค่า Q_1 โดยการหาค่าตำแหน่งของ Q_1 จาก $\frac{1}{4}(N+1) = \frac{1}{4}(10+1) = 2\frac{3}{4}$

จะได้ว่า Q_1 มีค่าเท่ากับค่าในตำแหน่งที่ 2 รวมกับ $\frac{3}{4}$ ของระยะห่างระหว่างตำแหน่งที่ 2 กับ ที่ 3

$$\therefore \quad Q_1 = 12 + \frac{3}{4}(14-12) = 12+1.5$$
$$= 13.5$$

หาค่า Q_3 โดยการหาค่าตำแหน่งของ Q_3 จาก $\frac{3}{4}(N+1) = \frac{3}{4}(10+1) = 8\frac{1}{4}$

จะได้ว่า Q_3 มีค่าเท่ากับค่าในตำแหน่งที่ 8 รวมกับ $\frac{1}{4}$ ของระยะห่างระหว่างตำแหน่งที่ 8 กับ ที่ 9

$$\therefore \quad Q_3 = 24 + \frac{1}{4}(26-24) = 24.5$$

$$\therefore \quad Q.D. = \frac{24.5-13.5}{2} = 5.5$$

นั่นคือ ส่วนเบี่ยงเบนควอไทล์ของข้อมูลชุดนี้ เท่ากับ 5.5

2.2 ข้อมูลที่แบ่งกลุ่ม (Grouped Data)

ตัวอย่างที่ 2.18 จงหาค่าส่วนเบี่ยงเบนควอไทล์ของผลการสอบวิชาภาษาไทยของนักศึกษาปีที่ 2 ดังข้อมูลต่อไปนี้

คะแนน	จำนวนนักศึกษา
50 – 59	8
60 – 69	7
70 – 79	18
80 – 89	8
90 – 99	9
รวม	50

วิธีทำ จากตารางให้หาค่าความถี่สะสม

คะแนน	จำนวนนักศึกษา	ความถี่สะสม (F)
50 – 59	8	8
60 – 69	7	15
70 – 79	18	33
80 – 89	8	41
90 – 99	9	50

หาค่า Q_1 โดยการหาค่าตำแหน่งของ Q_1 จาก $\frac{1}{4} \cdot N = \frac{1}{4}(50) = 12.5$

$\therefore$ Q_1 จะมีค่าอยู่ระหว่าง 59.5 และ 69.5

จากการเทียบบัญญัติไตรยางค์

$\therefore$ $Q_1 = 59.5 + \frac{10}{7}(4.5) = 65.93$

หาค่า Q_3 โดยการหาค่าตำแหน่งของ Q_3 จาก $\frac{3}{4} \cdot N = \frac{3}{4}(50) = 37.5$

$\therefore$ Q_3 จะมีค่าอยู่ระหว่าง 79.5 และ 89.5

จากการเทียบบัญญัติไตรยางค์

$\therefore$ $Q_3 = 79.5 + \frac{10}{8}(4.5) = 85.13$

$\therefore$ $Q.D. = \frac{85.13 - 65.93}{2} = 9.6$

นั่นคือ ส่วนเบี่ยงเบนควอไทล์ของผลการสอบวิชาภาษาไทยของนักศึกษาปีที่ 2 เท่ากับ 9.6 คะแนน

3.ส่วนเบี่ยงเบนเฉลี่ย หาได้จากค่าเฉลี่ยของความแตกต่างของข้อมูลแต่ละค่าที่แตกต่างไปจากค่าเฉลี่ยของข้อมูลชุดนั้น โดยไม่คำนึงถึงเครื่องหมาย สัญลักษณ์ที่ใช้ คือ M.D.หรือ A.D.

3.1 ข้อมูลที่ไม่ได้แบ่งกลุ่ม (ungrouped data)

นิยามที่ 2.4 กำหนดให้ $X_1, X_2, \ldots, X_n$ เป็นข้อมูลทั้งหมด N ค่าจากประชากรที่มีค่าเฉลี่ยเป็น μ จะได้ว่า

$$M.D. = \frac{\sum_{i=1}^{N} |X_i - \mu|}{N}$$

ในกรณีที่เป็นกลุ่มตัวอย่าง จะได้ค่าเบี่ยงเบนเฉลี่ยของกลุ่มตัวอย่าง คือ

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

ตัวอย่างที่ 2.19 ข้อมูลต่อไปนี้เป็นส่วนสูง (เซนติเมตร) ของคน 8 คน 157, 156, 160, 158, 175, 170, 164, 167 จงหาส่วนเบี่ยงเบนมาตรฐานของส่วนสูงของคนทั้ง 8 คนนี้

วิธีทำ ให้ M.D. เป็นส่วนเบี่ยงเบนมาตรฐานของส่วนสูงคนทั้ง 8 คน

จากสูตร $M.D. = \frac{\sum_{i=1}^{n} |X_i - \bar{X}|}{n}$

และ $\bar{X} = \frac{\sum X}{n}$

$= \frac{157 + 156 + \cdots + 167}{8}$

$= 163.38$

หาค่า M.D.ได้จากตาราง

X	157	156	160	158	175	170	164	167	รวม
$\lvert X-\bar{X}\rvert$	6.38	7.38	3.38	5.38	11.62	6.62	0.62	3.62	45

$$\therefore \quad \text{M.D.} = \frac{45}{8} = 5.625$$

นั่นคือ ส่วนเบี่ยงเบนมาตรฐานของส่วนสูงของคนทั้ง 8 คนนี้ เท่ากับ 5.625 คะแนน

3.2 ข้อมูลที่แบ่งกลุ่ม (grouped data)

ประชากรที่มีข้อมูลทั้งหมด N ค่าและมีค่าเฉลี่ยเป็น μ หาค่าส่วนเบี่ยงเบนเฉลี่ยของประชากรได้จากสูตร

$$\text{M.D.} = \frac{\sum_{i=1}^{N} f_i \lvert X_i - \mu\rvert}{N}$$

ในกรณีที่เป็นส่วนเบี่ยงเบนเฉลี่ยของกลุ่มตัวอย่างจะได้ว่า

$$\text{M.D.} = \frac{\sum_{i=1}^{n} f_i \lvert X_i - \bar{X}\rvert}{n}$$

ตัวอย่างที่ 2.20 จงหาค่าส่วนเบี่ยงเบนเฉลี่ยจากข้อมูลในตาราง

คะแนน	จำนวน
8 – 12	2
13 – 17	5
18 – 22	9
23 – 27	12
28 – 32	8
33 – 37	3
38 – 42	1
รวม	40

วิธีทำ หาค่า M.D. จากสูตร $\text{M.D.} = \dfrac{\sum f_i \lvert X_i - \bar{X}\rvert}{n}$

คะแนน	จำนวน (f_i)	X_i	f_iX_i	$\lvert X_i - \bar{X} \rvert$	$f_i \lvert X_i - \bar{X} \rvert$
8 – 12	2	10	20	14	28
13 – 17	5	15	75	9	45
18 – 22	9	20	180	4	36
23 – 27	12	25	300	1	12
28 – 32	8	30	240	6	48
33 – 37	3	35	105	11	33
38 – 42	1	40	40	16	16
	40		960		218

จากสูตร $\bar{X} = \dfrac{\sum f_i X_i}{n}$

$$= \frac{960}{40} = 24$$

$$\therefore \quad \text{M.D.} = \frac{218}{40} = 5.45$$

นั่นคือ ส่วนเบี่ยงเบนเฉลี่ยของข้อมูลชุดนี้ คือ 5.45

4. ส่วนเบี่ยงเบนมาตรฐานและความแปรปรวน การวัดการกระจายด้วยส่วนเบี่ยงเบนมาตรฐานและความแปรปรวน เป็นที่นิยมใช้กันมากที่สุด มีลักษณะคล้ายกับส่วนเบี่ยงเบนเฉลี่ยมาก ปรับปรุงโดยการยกกำลังสองผลต่างระหว่างคะแนนกับค่าเฉลี่ยของข้อมูลชุดนั้น แล้วนำมาหาค่าเฉลี่ย เราเรียกว่าความแปรปรวน เขียนสัญลักษณ์แทนด้วย σ^2 และค่ารากที่สองของความแปรปรวน เรียกว่าส่วนเบี่ยงเบนมาตรฐาน เขียนสัญลักษณ์แทนด้วย σ

4.1 ความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานของข้อมูลที่ไม่ได้แบ่งกลุ่ม

4.1.1 ข้อมูลประชากร

นิยามที่ 2.5 ให้ $X_1, X_2, X_3, \ldots, X_N$ เป็นข้อมูลทั้งหมด N ค่า ของประชากรที่มีค่าเฉลี่ยเป็น μ ความแปรปรวน σ^2 คือ

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

แต่เนื่องจาก

$$\begin{aligned}\sum_{i=1}^{N}(X_i - \mu)^2 &= \sum_{i=1}^{N}(X_i^2 - 2X_i\mu + \mu^2) \\ &= \sum_{i=1}^{N} X_i^2 - 2N\mu^2 + N\mu^2 \\ &= \sum_{i=1}^{N} X_i^2 - N\mu^2\end{aligned}$$

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

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

ส่วนเบี่ยงเบนมาตรฐาน (σ) คือ รากที่สองที่เป็นบวกของความแปรปรวน จะได้ว่า

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

ตัวอย่างที่ 2.21 ข้อมูลชุดหนึ่งประกอบด้วย 13 , 10 , 12 , 16 , 13 , 14 จงหาค่าความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุดนี้

วิธีทำ ให้ σ^2 เป็นค่าความแปรปรวนของข้อมูลชุดนี้

จากสูตร $\qquad \sigma^2 = \frac{\sum(X_i-\mu)^2}{N}$

หาค่า μ จากสูตร $\qquad \mu = \frac{\sum X_i}{N}$

$$= \frac{13+10+12+16+13+14}{6} = 13$$

หาค่า σ^2 จากตาราง

ข้อมูล	13	10	12	16	13	14	รวม
$(X_i-\mu)^2$	0	9	1	9	0	1	20

$$\therefore \qquad \sigma^2 = \frac{20}{6} = 3.33$$

และ $\qquad \sigma = \sqrt{3.3} = 1.83$

นั่นคือ ความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุดนี้ เท่ากับ 3.33 และ 1.8

4.1.2 ข้อมูลตัวอย่าง

นิยามที่ 2.6 ให้ $x_1, x_2, x_3, \cdots, x_n$ เป็นข้อมูลทั้งหมด n ค่าของกลุ่มตัวอย่าง ที่มีค่าเฉลี่ยเป็น $\bar{x}$ ค่าความแปรปรวนของกลุ่มตัวอย่าง แทนด้วย σ^2 โดยที่

$$S^2 = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1} \qquad \text{หรือ} \qquad S^2 = \frac{\sum_{i=1}^{n} X_i^2}{n-1} - \frac{n\bar{X}^2}{n-1}$$

ส่วนเบี่ยงเบนมาตรฐานของกลุ่มตัวอย่าง แทนด้วย S คือ การหารากที่สองของ S^2

ตัวอย่างที่ 2.22 จากการสุ่มตัวอย่างเด็ก 10 คน ชั่งน้ำหนัก (กิโลกรัม) ได้ข้อมูลดังนี้ 34, 49, 43, 46, 51, 45, 52, 49, 54, 47 จงหาค่าส่วนเบี่ยงเบนมาตรฐานของน้ำหนักของเด็ก 10 คนนี้

วิธีทำ ให้ S เป็นส่วนเบี่ยงเบนมาตรฐานของน้ำหนักของเด็ก 10 คน

จากสูตร $S = \sqrt{\dfrac{\sum (X_i - \bar{X})^2}{n-1}}$

และ $\bar{X} = \dfrac{\sum X_i}{n} = \dfrac{34+49+43+\cdots+47}{10} = 47$

หาค่า S จากตาราง

น้ำหนัก	34	49	43	46	51	45	52	49	54	47	รวม
$(X_i - \bar{X})^2$	169	4	16	1	16	4	25	4	49	0	288

$\therefore \quad S = \sqrt{\dfrac{288}{9}} = 5.66$

นั่นคือ ส่วนเบี่ยงเบนมาตรฐานของน้ำหนักของเด็ก 10 คนนี้ เท่ากับ 5.66 กิโลกรัม

4.2 ความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานของข้อมูลที่แบ่งกลุ่ม

4.2.1 ข้อมูลประชากร

นิยามที่ 2.7 ให้ $X_1, X_2, X_3, \cdots, X_N$ เป็นข้อมูลทั้งหมด N ค่าที่มีความถี่เป็น $f_1, f_2, f_3, \cdots, f_N$ ของประชากรที่มีค่าเฉลี่ยเป็น μ จะได้ว่าความแปรปรวน (σ^2) คือ

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

ส่วนเบี่ยงเบนมาตรฐาน (σ) คือ รากที่สองที่เป็นบวกของความแปรปรวน

4.2.2 ข้อมูลตัวอย่าง

นิยามที่ 2.8 ให้ $X_1, X_2, X_3, \cdots, X_n$ เป็นข้อมูลทั้งหมด n ค่าที่มีความถี่เป็น $f_1, f_2, f_3, \cdots, f_n$ ของกลุ่มตัวอย่างที่มีค่าเฉลี่ยเป็น $\bar{X}$ จะได้ว่าความแปรปรวนตัวอย่าง (S^2) คือ

$$S^2 = \frac{\sum_{i=1}^{n} f_i (X_i - \bar{X})^2}{n-1} \qquad \text{หรือ} \qquad S^2 = \frac{\sum_{i=1}^{n} f_i X_i^2}{n-1} - \frac{n\bar{X}^2}{n-1}$$

ส่วนเบี่ยงเบนมาตรฐานตัวอย่าง (S) คือ รากที่สองที่เป็นบวกของความแปรปรวนตัวอย่าง

ตัวอย่างที่ 2.23 จากตัวอย่างที่ 2.20 จงหาค่าความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุดนี้

วิธีทำ จากตัวอย่างที่ 2.20 ได้ค่า $\bar{X} = 24$

คะแนน	จำนวน (f_i)	X_i	$(X_i - \bar{X})$	$(X_i - \bar{X})^2$	$f_i(X_i - \bar{X})^2$
8 – 12	2	10	14	169	338
13 – 17	5	15	9	81	405
18 – 22	9	20	4	16	144
23 – 27	12	25	1	1	12
28 – 32	8	30	6	36	288
33 – 37	3	35	11	121	363
38 – 42	1	40	16	256	256
	40				1806

จากสูตร $S^2 = \frac{\sum f_i (X_i - \bar{X})^2}{n-1} = \frac{1806}{39} = 46.31$

และ $S = \sqrt{46.31} = 6.8$

นั่นคือ ความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุดนี้ เท่ากับ 46.31 และ 6.8 คะแนน

4.3 คุณสมบัติของความแปรปรวนและส่วนเบี่ยงเบนมาตรฐาน

4.3.1 **ทฤษฎีบทที่ 2.3** ถ้าข้อมูลแต่ละตัวมีค่าลดลง หรือเพิ่มขึ้นเท่า ๆ กันทุกค่าสมมติให้เท่ากับ C ซึ่งเป็นค่าคงที่แล้ว ค่าความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุดใหม่ จะเท่ากับความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุดเดิม

พิสูจน์ ให้ $X_1, X_2, X_3, \cdots, X_N$ เป็นค่าของข้อมูล N ค่า

โดยที่ $\sigma^2 = \frac{\sum (X_i - \mu)^2}{N}$ และ $\mu = \frac{\sum X_i}{N}$

ถ้าข้อมูลแต่ละตัวมีค่าเพิ่มขึ้นเท่ากับ C

$\therefore \quad X_1 + C, X_2 + C, X_3 + C, \cdots, X_N + C$ เป็นค่าของข้อมูลชุดใหม่ N ค่า

$$\because \quad \mu_{\text{ใหม่}} = \frac{\sum (X_i + C)}{N}$$

$$= \frac{\sum X_i}{N} + \frac{\sum C}{N}$$

$$= \mu + C$$

$$\therefore \quad \sigma^2_{\text{ใหม่}} = \frac{\sum [(X_i + C) - (\mu + C)]^2}{N}$$

$$= \frac{\sum(X_i + C - \mu - C)^2}{N}$$

$$= \frac{\sum(X_i - \mu)^2}{N}$$

$$= \sigma^2$$

4.3.2 **ทฤษฎีบทที่ 2.4** ถ้าข้อมูลแต่ละตัวมีค่าเพิ่มขึ้น C เท่าของค่าเดิม ค่าความแปรปรวน ของข้อมูลชุดใหม่ เท่ากับกำลังสองของค่าคงที่ C คูณค่าความแปรปรวนของข้อมูลชุดเดิม หรือส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุดใหม่ เท่ากับ ค่าคงที่ C คูณส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุดเดิม

พิสูจน์ ให้ $X_1, X_2, X_3, \cdots, X_N$ เป็นค่าของข้อมูล N ค่า

โดยที่ $\sigma^2 = \frac{\sum(X_i - \mu)^2}{N}$ และ $\mu = \frac{\sum X_i}{N}$

ถ้าข้อมูลแต่ละตัวมีค่าเพิ่มขึ้นเท่ากับ C เท่าของข้อมูลชุดเดิม

$\therefore$ $CX_1, CX_2, CX_3, \cdots, CX_N$ เป็นค่าของข้อมูลชุดใหม่ N ค่า

$\therefore$ $\mu_{\text{ใหม่}} = \frac{\sum CX_i}{N} = C\mu$

และ $\sigma^2_{\text{ใหม่}} = \frac{\sum(CX_i - C\mu)^2}{N}$

$$= \frac{C^2\sum(X_i - \mu)^2}{N}$$

$$= C^2\sigma^2$$

และ $\sigma_{\text{ใหม่}} = C\sigma$

ตัวอย่างที่ 2.24 ในการสอบย่อยครั้งหนึ่งคะแนนเต็ม 20 คะแนน ค่าเฉลี่ยเลขคณิตและความแปรปรวนของคะแนนที่สอบได้เป็น 12.5 และ 1.2 ตามลำดับ ถ้าครูจะปรับคะแนนเต็มเป็น 60 คะแนน จงหาค่าเฉลี่ยและความแปรปรวนของคะแนนนักเรียนชุดใหม่

วิธีทำ จากโจทย์กำหนด $\bar{X} = 12.5$, $S^2 = 1.2$ คะแนนเต็ม = 20 คะแนน

เนื่องจากครูปรับคะแนนเต็มเป็น 60 คะแนน มีค่าเป็น 3 เท่าของคะแนนเดิม

จากคุณสมบัติของ $\bar{X}$ (ท.บ. 2.2) ได้ว่า

$$\begin{aligned} \bar{X}_{\text{ใหม่}} &= k\bar{X}_{\text{เดิม}} \\ &= 3(12.5) \\ &= 37.5 \end{aligned}$$

จากคุณสมบัติของ σ^2 (ท.บ. 2.4) ได้ว่า

$$\sigma^2_{\text{ใหม่}} = C^2\sigma^2_{\text{เดิม}}$$
$$= 3^2(1.2)$$
$$= 10.8$$

นั่นคือ ค่าเฉลี่ยและความแปรปรวนของคะแนนของนักเรียนชุดใหม่เป็น 37.5 และ 10.8 ตามลำดับ

.5. สัมประสิทธิ์แห่งการกระจาย ถ้าข้อมูล 2 ชุด มีค่าเฉลี่ยเท่ากัน เราอาจนำค่าส่วนเบี่ยงเบนมาตรฐานของข้อมูลทั้ง 2 ชุด มาเปรียบเทียบกันโดยตรง เพื่อเปรียบเทียบการกระจายของข้อมูลทั้ง 2 ชุด แต่ถ้าข้อมูลทั้ง 2 ชุดมีหน่วยการวัดต่างกัน หรือมีค่าเฉลี่ยไม่เท่ากันจะนำส่วนเบี่ยงเบนมาตรฐานของข้อมูลแต่ละชุดมาเปรียบเทียบกันเองโดยตรงไม่ได้จะเปรียบเทียบโดยการใช้สัมประสิทธิ์แห่งการกระจายจากสูตร $C.V. = \frac{100S}{\bar{X}}$ โดยที่ C.V.เป็นสัมประสิทธิ์แห่งการกระจาย

ตัวอย่างที่ 2.25 ถ้านาย ก. จะต้องตัดสินใจเลือกซื้อหุ้นบริษัทใดบริษัทหนึ่ง จากที่มีให้เลือก 3 บริษัทที่มีอัตราเงินปันผลดังนี้ บริษัท A เงินปันผลเฉลี่ย 15.6 ต่อปี และค่าส่วนเบี่ยงเบนมาตรฐาน 3.7 บริษัท B เงินปันผลเฉลี่ย 13.7 ต่อปี และค่าส่วนเบี่ยงเบนมาตรฐาน 2.5 และบริษัท C เงินปันผลเฉลี่ย 18.9 ต่อปี และค่าส่วนเบี่ยงเบนมาตรฐาน 5.8 ถ้าท่านเป็นนาย ก. ท่านจะตัดสินใจเลือกลงทุนซื้อหุ้นของบริษัทใด

วิธีทำ โจทย์ต้องการทราบว่านาย ก. จะตัดสินใจเลือกลงทุนซื้อหุ้นของบริษัทใด สามารถคำนวณได้จากค่าของ C.V.จากสูตร

$$C.V. = \frac{100S}{\bar{X}}$$

โจทย์กำหนด $\bar{X}_A = 15.6, S_A = 3.7$ $\bar{X}_B = 13.7, S_B = 2.5$ $\bar{X}_C = 18.9, S_C = 5.8$

$$C.V.\ \text{ของบริษัท}\ A = \frac{100(3.7)}{15.6} = 23.72$$

$$C.V.\ \text{ของบริษัท}\ B = \frac{100(2.5)}{13.7} = 18.25$$

$$C.V.\ \text{ของบริษัท}\ C = \frac{100(5.8)}{18.9} = 30.69$$

$$\therefore \quad C.V._B < C.V._A < C.V._C$$

นั่นคือ นาย ก. ตัดสินใจเลือกลงทุนซื้อหุ้นของบริษัท B
